Diagnostic and Prognostic Values of a Branched-chain Amino Acid-associated Gene Signature in Lung Cancer

Lei Jiang (ndyfy05820@ncu.edu.cn)
Nanchang University
Quanjin Li
Nanchang University
Zhi Hu
Nanchang University
Mingshan Liu
Nanchang University
Zhuofan Yang
Nanchang University

Research

Keywords: Branched-chain amino acids, Non-small-cell lung carcinoma, diagnostic signature, prognostic signature, nomogram

Posted Date: September 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-907919/v1

Abstract

Background

Lung cancer is the malignancy with the highest morbidity and mortality worldwide. Several studies have indicated that branched-chain amino acids (BCAAs) can regulate cancer progression and metabolism. In this study, a BCAA-associated prognostic signature was established and assessed to verify whether it can be used in clinical treatment of patients with non-small-cell lung carcinoma (NSCLC).

Methods

The candidate BCAA-associated genes were screened from TCGA cohort by differential expression analysis. Two datasets, the training and test sets, were divided from the lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), and LUSC-only datasets in TCGA cohort and the novel model were constructed based on the training dataset. Moreover, all prognosis-related genes were screened via univariate Cox analysis. The LASSO Cox method was used to avoid collinearity. To further probe the clinical effect, a large number of clinical characteristics were included in the prognostic risk model. The outcomes for the calibration curves and nomogram proved the capability of the prognostic signature.

Results

Based on the 366 samples in the training set, six survival-related genes (CAV1, GRIA1, LDHA, AOX1, GRIN2D, and AKR1C1) were screened using the LASSO Cox regression analysis. Patients were divided into high- and low-risk groups. The signature demonstrated good diagnostic and prognostic capabilities.

Conclusion

We developed the first-ever BCAA-related signature with both prognostic predictive power and diagnostic efficacy. The prognostic signature and diagnostic model suggested that the expression of BCAAs could serve as a powerful potential marker in cancer clinical treatment in the future.

Highlights

1. Branched-chain amino acids-associated gene signatures with both prognostic predictive power and diagnostic efficacy in non-small-cell lung carcinoma were identified.
2. 44 related candidate genes, 34 of which were upregulated while 12 were downregulated.
3. Six survival-related genes (CAV1, GRIA1, LDHA, AOX1, GRIN2D, and AKR1C1) were screened using the LASSO Cox regression analysis.
4. The BCAA-related prognostic signature showed good performance in the both test set and external verification set.
5. The diagnostic accuracy of the diagnostic signature constructed based on key genes in the prognostic model was satisfactory.

Introduction

Lung cancer, including small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC), presents the highest mortality worldwide [1]. NSCLC represents over 80% of lung cancer cases [2], and it is mainly composed of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). Many clinical strategies have been adopted for lung cancer treatment, such as surgery, chemotherapy, and targeted therapy for patients with stage I–IV cancer [3]. However, the poor 5-year overall survival rate suggests that current therapies need improvement [4]. In addition, nearly 75% of patients diagnosed are already in the terminal stages of the disease. Although it is possible to assess risk and optimize therapy for every patient by clinical and pathological indicators, such as positron emission tomography CT (PET-CT) or pathology results, these indicators cannot be quantified to differentiate a patient with low risk from a patient high risk and are not sufficient in clearly assessing individual circumstances [5, 6]. Thus, a dependable novel signature with early diagnostic efficacy and prognostic predictive power would greatly improve patient care and treatment development.

Branched-chain amino acids (BCAAs), including the essential amino acids leucine and isoleucine, can regulate cancer progression and its metabolism [7, 8]. For instance, abnormal activation of BCAAs occur in many types of human cancers, such as leukemia [9], breast cancer [10], hepatocellular carcinoma (HCC) [11], glioblastoma [12], pancreatic ductal adenocarcinoma [13], and especially in NSCLC [14]. NSCLC tumors incorporate free BCAAs into tissue proteins and use BCAAs as a nitrogen source. Moreover, the enzymes responsible for BCAAs, such as BCAT1 and BCAT2, can impair NSCLC formation.

Therefore, studies on BCAAs in cancer progression represent a growing field. To date, no study has been performed to illustrate the relationship between a BCAA-associated prognostic model and lung cancer. In this study, we first used a novel selection operator (LASSO) Cox regression analysis to establish a BCAA-associated prognostic signature in the training set of The Cancer Genome Atlas (TCGA) LUAD and LUSC cohorts. Both signatures in the test set and external verification set were verified thoroughly. Meanwhile, a feasible clinical model could make the individual application rely on various clinical characteristics. In addition, the diagnostic signature established based on the number of key genes in the prognostic signature presented satisfactory diagnostic efficacy. We proposed that the BCAA-associated prognostic signature can be adopted in clinical treatment in the future to benefit not only patients with LUAD but also those with LUSC only.

Results

Identification of the candidate genes

The expression matrix of 159 BCAA-associated genes was obtained by differential expression analysis. There were 44 related candidate genes, 34 of which were upregulated while 12 were downregulated (Figure 2; Table S1 and S2). According to the results of the functional enrichment analysis, candidate

genes were associated with the ionotropic glutamate receptor signaling pathway, glutamate receptor signaling pathway, and ionotropic glutamate receptor activity (Figure 3 and Table S3).

Establishment of the BCAA-associated prognostic signature

Based on the 366 samples in the training set, six survival-related genes (CAV1, GRIA1, LDHA, AOX1, GRIN2D, and AKR1C1) were screened using the LASSO Cox regression analysis (Table 1). Furthermore, to verify the reliability of survival-related genes, LASSO Cox analysis was performed in the training dataset, and a signature of seven key genes was established (Figure 4A and B. The formula used to compute the risk score for patients is as follows:

Risk score = $Exp_{CAV1} \times 0.0679 + Exp_{GRIA1} \times (-0.0372) + Exp_{LDHA} \times 0.0669 + Exp_{AOX1} \times 0.0039 + Exp_{GRIN2D} \times 0.0105 + Exp_{AKR1C1} \times (-0.0062)$.

Using the preceding formula for calculating the risk score, each patient was assigned a risk score. The patients were then divided into high- and low-risk groups with a cut-off value of 1.8. Several patients with a high-risk level died during follow-up, whereas most patients in the low-risk group remained alive (Figure 4C, red dots indicate patients who died, and blue dots indicate patients who remained alive). Compared with the low-risk group, the survival status of the high-risk group was remarkably worse in terms of overall survival (hazard ratio [HR] = 11; 95% CI = 4.6–28; $p < 0.001$) (Figure 4D).

Validation of the prognostic signature

Furthermore, the test and external validation set GSE30219 were used to verify the dependability and stability of this prognostic model based on the risk score calculation formula. In the test set, the patients were divided into high- and low-risk groups with a cut-off value of 1.73. The survival rates of the two groups were significantly disparate (HR = 2.8; 95% CI = 0.82–9.3, $p < 0.004$) (Figure 5A&C). Meanwhile, high-risk scores also significantly predicted poor prognosis (HR = 2.6; 95% CI = 0.97–7.2; $p < 0.001$) in high- and low-risk groups, with a cut-off value of 1.79 for GSE30219 (Figure 5B&D).

Clinical potential capacity of the prognostic signature

To determine whether the BCAA-related prognostic signature could serve as a reliable prognostic indicator for NSCLC, univariate and multivariate Cox regression analyses were conducted in TCGA LUAD and LUSC cohort. Age, gender, smoking, TNM stage, and risk score were included in the analysis. The results demonstrated that our prognostic risk model could work independently for the prediction of prognosis in patients with NSCLC (Table 1). To further assess the reliability of the prognostic signature for patients with different clinical conditions, we conducted a stratified analysis based on TCGA cohort. Subsequently, the relationship between the prognostic signature and TNM stage was analyzed. The risk score significantly increased in the late clinical stage (Figure 6A), larger tumor size (Figure 6B), and lymph node metastasis (Figure 6C). These results (95% CI) further illustrated the predictive prognostic value of the prognostic signature.

Furthermore, to promote a clinically feasible approach capable of predicting an individual's 3- and 5-year survival probability, we conducted a nomogram including the prognostic signature and clinical characteristics listed earlier in TCGA LUAD and LUSC cohort (Figure 7A). The calibration curves revealed that the predicted survival of the nomogram was very close to the actual survival outcomes (Figure 7B).

Construction and verification of the diagnostic model based on the signature

We further investigated whether the diagnostic signature of genes included in the prognostic signature was valuable. The diagnostic signature was built in TCGA LUAD and LUSC cohort based on six key genes in the prognostic signature, and the model was subjected to the external validation dataset GSE30219. The area under the curve values of 0.997 in TCGA and 0.807 in GSE30219 indicated acceptable quality in discriminating patients with NSCLC from normal patients (Figure 8).

Evaluation of risk scores &Immune checkpoints of the signature

Six classical immune checkpoints, namely, IL6, CXCR4, CD276, TGFB1, CCL2, and CD274, were assessed to compare the mutation differences (Figure 9A), as well as the coefficient of Pearson's and P value were recorded in Table 2. The Box plot presented the expression levels of the six immune checkpoints genes in LUAD and LUSC, which illustrated IL6, CD276, TGFB1 and CD274 were obviously higher in the high-risk group than CXCR4 and CCL2 in the low-risk group. The heatmap presents the expression levels of the six selected genes in LUAD and LUSC, demonstrating

These six genes could be differentiated from LUAD and LUSC apparently.

Discussion

In this study, we focused on BCAA-associated genes and obtained six vital genes, namely, CAV1, GRIA1, LDHA, AOX1, GRIN2D, and AKR1C1, via univariate Cox analysis and differential analysis. Then, the LASSO Cox method was used to construct the six-gene prognostic signature. Finally, we assessed the relationship of the BCAA-associated gene signature on immunity, diagnosis, and clinical potential.

BCAAs play a key role in several cancers, such as lung cancer, colorectal cancer, and liver cancer. For example, the level of circulating BCAAs in the blood was found to be significantly upregulated in pancreatic cancer compared with that in the control group owing to the breakdown of peripheral tissue protein [27]. Overexpressed BCAAs may represent a predictive marker for several cancers [28]. Furthermore, metabolism of BCAA is influenced by the solid tumor microenvironments in NSCLC [29], breast cancer [30], and melanoma [31]. For example, BCAT1 is not only able to support proliferation of normal breast tissue [32] but also supports the growth of cancer cells in a glioblastoma model [33]. Emerging evidence suggests that the regulation of BCAA catabolism can have affect the biology of some cancers [34]. According to our results, the prognostic signature and diagnostic model suggested that the expression of BCAAs could serve as a powerful potential marker in cancer clinical treatment in the future.

However, until date, there have been no studies on the BCAA-associated genes, and the potential mechanism and clinical prospects for cancer treatment should be further explored.

In this study, first, the BCAA-associated candidate genes were screened from TCGA cohort by differential expression analysis. Many of these have already been found to be involved in cancer mechanism. For example, CAV1 is regarded as a downstream effector of multiple molecules, including protein kinase B (AKT), epidermal growth factor receptor (EGFR), focal adhesion kinase (FAK), and extracellular regulated protein kinases (ERK), which mediate the vital aspects of tumor invasion and progression. CAV1 expression is particularly higher in lung cancer tissue than in normal tissue, and it has a significantly higher expression in SCLC tissue than in NSCLC tissue [35].

Second, two datasets, a training set and a test set, were established from TCGA LUAD and LUSC cohort. A novel model was constructed based on the training dataset. Moreover, all prognosis-related genes were screened via univariate Cox analysis. Cox regression analysis was performed to probe the six key genes. To ensure the reliability and accuracy of this model, the LASSO Cox method was also used to avoid collinearity. As a result, the six-gene prognostic model was developed containing CAV1, GRIA1, LDHA, AOX1, GRIN2D, and AKR1C1, these genes in the prognostic signature have already been identified for their potential functions in cancers.

According to the results of the two datasets, the prognostic signature had powerful predictive capability. To further probe its clinical capability, a large number of clinical characteristics were included for analysis. As expected, the prognostic risk model could serve as a self-reliant prognostic indicator for patients with NSCLC, according to both univariate and multivariate Cox regression analyses. At the same time, levels of higher risk were conspicuously related to large tumor size, lymph node metastasis, and late clinical stage. The outcomes of the calibration curves and nomogram proved the capability of the prognostic signature.

Overall, in TCGA cohort, the mutation rate of the high-risk population was usually higher than that of the low-risk population. In addition, the diagnostic accuracy based on the key genes in the prognosis model was satisfactory, which may mean this model has clinical applicability. Most importantly, the results demonstrated that BCAAs can influence the progression of cancer. However, this study has several limitations. Normal samples were markedly fewer than tumor samples, which may lead to a certain degree of deviation in the outcome. In addition, owing to the lack of basic experiments, the molecular mechanism cannot be explained in depth. In summary, our findings suggested that BCAAs could serve as a powerful potential marker in cancer clinical treatment in the future.

Conclusion

In this study, we developed the first-ever BCAA-related signature with both prognostic predictive power and diagnostic efficacy. The prognostic signature and diagnostic model suggested that the expression of BCAAs could serve as a powerful potential marker in cancer clinical treatment in the future.

Materials And Methods

Data sources and processing

TCGA LUAD and TCGA LUSC RNA-Seq data, including count value and clinical information, were downloaded from TCGA database obtained in the TCGAbiolinks R package [15] in R portal. There were 1,037 tumors and 108 control samples in the original data. A total of 1,016 tumor samples were selected after matching the gene expression profile with their corresponding clinical information. In addition, GSE30219, used as the external validation dataset, was extracted from Gene Expression Omnibus [16] using the GEOquery R package [17, 18], and included 293 tumor samples and 14 normal samples. Moreover, BCAAs were extracted from the National Center for Biotechnology Information–GENE (NCBI-GENE), the database of MSigDB [18], as well as from the Metabolic Atlas database [19]. After removing duplicated BCAA-associated genes, 159 genes were obtained.

Differential BCAA-associated genes expression analysis

The trimmed mean of M-values (TMM) method was performed to normalize the RNA-Seq data, using the edgeR R package [20]. Then, the gene expression matrix of BCAA-associated genes was extracted using the intended procedure. A false discovery rate (FDR) of less than 0.05 and |log2 (fold change)| greater than one as the truncation standard [21] were used as the criteria for the BCAA-associated differentially expressed genes in tumor and normal tissues. All these genes were named candidate genes. Heatmap and volcano plots were also designed with the heatmap R package and the ggplot2, respectively.

Functional enrichment analyses

Gene ontology functional analysis and genome pathway enrichment analyses for candidate genes were performed using the clusterProfiler R package in R portal [22]. An FDR < 0.05 was considered to be statistically significant.

Establishment and validation of the BCAA-associated prognostic signature

A total of 1,016 tumor samples were randomly divided into the training set and test set at a ratio of 7:3, by using the caret R package in R portal [23]. Candidate genes, using LASSO regression analysis in the training dataset, with a p-value of 0.05 were identified as survival-related genes, obtained via the glmnet package in R software [24]. Moreover, 10-fold cross-validation and 1,000-times calculation were set to ascertain the λ value, as well as to filter λ, which caused minimal cross-validation loss (lambda.min) in the analysis. The BCAA-associated prognostic signature was subsequently established by conducting the multivariate Cox regression method. A BCAA-associated risk score was calculated using the LASSO Cox coefficient as follows:

Risk score = $Exp_{CAV1} \times 0.0679 + Exp_{GRIA1} \times (-0.0372) + Exp_{LDHA} \times 0.0669 + Exp_{AOX1} \times 0.0039 + Exp_{GRIN2D} \times 0.0105 + Exp_{AKR1C1} \times (-0.0062)$.

The difference was evaluated by Kaplan–Meier survival analysis. The predictive potential capability of the BCAA-associated prognostic signature was then proved in the test dataset and external dataset GSE30219. Samples in the test dataset and validation dataset were divided into two groups of high and low risk based on the risk score formula.

Clinical utility of prognostic signature

Several clinical characteristics such as age, ethnicity, gender, operation name, pathology, and stage were extracted from TCGA database. These clinical features were then subjected to univariate and multivariate Cox regression analyses to assess whether there were risk scores serving as a ponderable prognostic factor in LUAD and LUSC. A p-value < 0.05 was regarded as statistically significant. Furthermore, the outcome of nomogram, showing risk scores and clinical factors, was verified using the rms R package [25]. In addition, the predictive potential was assessed using calibration curves showing the diversity between the predicted and practical survival rates.

Evaluation of risk scores and immune checkpoints

To assess the relationship between the risk scores and tumor immunology, a number of classical immune checkpoints, such as IL6, CXCR4, CD276, TGFB1, CCL2, and CD274, were studied to compare the mutation differences. The results were visualized in a box plot with the ggpubr package in the R portal [26]. Concurrently, Pearson correlation coefficient was used to analyze the correlation between prognostic risk score and immune checkpoint expression, and a p-value < 0.05 was regarded as statistically significant.

Establishment and validation of the BCAA-associated diagnostic signature

Based on the BCAA-associated genes in the prognostic signature, a new predictive diagnostic model was established using a logistic regression method in the dataset of TCGA based on the vital genes of the prognostic signature. Additionally, the outcomes indicated that GSE30219 was a convincing dataset to verify the diagnostic potential capacity of this novel model.

Abbreviations

including small-cell lung cancer (SCLC), non-small-cell lung carcinoma (NSCLC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), Branched-chain amino acids (BCAAs), positron emission tomography CT (PET-CT), hepatocellular carcinoma (HCC), The Cancer Genome Atlas (TCGA), the National Center for Biotechnology Information–GENE (NCBI-GENE), The trimmed mean of M-values (TMM)

Declarations

Ethical Approval and Consent to participate

Not applicable.

Consent for publication

All authors have read and the final manuscript and approved the publication.

Availability of data and materials

Not applicable.

Competing interests

The authors declare no conflict of interest.

Funding

Not applicable.

Authors' contributions

LJ contributed significantly to conception of the review. QL and ZH wrote the manuscript. ML and ZY contributed to make preparations and revise the manuscript.

References

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.

2. Ye, Z., et al., *Breakthrough in targeted therapy for non-small cell lung cancer.* Biomed Pharmacother, 2021. **133**: p. 111079.

3. Wu, H., et al., *Long-term and short-term outcomes of robot- versus video-assisted anatomic lung resection in lung cancer: a systematic review and meta-analysis.* Eur J Cardiothorac Surg, 2020.

4. Suzuki, S. and T. Goto, *Role of Surgical Intervention in Unresectable Non-Small Cell Lung Cancer.* J Clin Med, 2020. **9**(12).

5. Sands, J., et al., *Lung Screening Benefits and Challenges: A Review of The Data and Outline for Implementation.* J Thorac Oncol, 2021. **16**(1): p. 37-53.

6. Oudkerk, M., et al., *Lung cancer LDCT screening and mortality reduction - evidence, pitfalls and future perspectives.* Nat Rev Clin Oncol, 2020.

7. Le Couteur, D.G., et al., *Branched chain amino acids, aging and age-related health.* Ageing Res Rev, 2020. **64**: p. 101198.

8. Peng, H., Y. Wang, and W. Luo, *Multifaceted role of branched-chain amino acid metabolism in cancer.* Oncogene, 2020. **39**(44): p. 6747-6756.

9. Wilkinson, A.C. and S. Yamazaki, *The hematopoietic stem cell diet.* Int J Hematol, 2018. **107**(6): p. 634-641.

10. Tobias, D.K., et al., *Circulating branched-chain amino acids and long-term risk of obesity-related cancers in women.* Sci Rep, 2020. **10**(1): p. 16534.

11. Hachiya, H., et al., *Effects of branched-chain amino acids on postoperative tumor recurrence in patients undergoing curative resection for hepatocellular carcinoma: A randomized clinical trial.* J Hepatobiliary Pancreat Sci, 2020. **27**(11): p. 819-829.

12. Fadhlullah, S.F.B., et al., *Pathogenic mutations in neurofibromin identifies a leucine-rich domain regulating glioma cell invasiveness.* Oncogene, 2019. **38**(27): p. 5367-5380.

13. Li, J.T., et al., *BCAT2-mediated BCAA catabolism is critical for development of pancreatic ductal adenocarcinoma.* Nat Cell Biol, 2020. **22**(2): p. 167-174.

14. Mayers, J.R., et al., *Tissue of origin dictates branched-chain amino acid metabolism in mutant Kras-driven cancers.* Science, 2016. **353**(6304): p. 1161-5.

15. Colaprico, A., et al., *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data.* Nucleic Acids Res, 2016. **44**(8): p. e71.

16. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Res, 2002. **30**(1): p. 207-10.

17. Davis, S. and P.S. Meltzer, *GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.* Bioinformatics, 2007. **23**(14): p. 1846-7.

18. Brown, G.R., et al., *Gene: a gene-centered information resource at NCBI.* Nucleic Acids Res, 2015. **43**(Database issue): p. D36-42.

19. Robinson, J.L., et al., *An atlas of human metabolism.* Sci Signal, 2020. **13**(624).

20. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

21. Korthauer, K., et al., *A practical guide to methods controlling false discoveries in computational biology.* Genome Biol, 2019. **20**(1): p. 118.

22. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* Omics, 2012. **16**(5): p. 284-7.

23. Murrell, D.S., et al., *Chemically Aware Model Builder (camb): an R package for property and bioactivity modelling of small molecules.* J Cheminform, 2015. **7**: p. 45.

24. Candia, J. and J.S. Tsang, *eNetXplorer: an R package for the quantitative exploration of elastic net families for generalized linear models.* BMC Bioinformatics, 2019. **20**(1): p. 189.

25. Huang, C., et al., *Establishment of Decision Rules and Risk Assessment Model for Preoperative Prediction of Lymph Node Metastasis in Gastric Cancer.* Front Oncol, 2020. **10**: p. 1638.

26. Mayakonda, A., et al., *Maftools: efficient and comprehensive analysis of somatic variants in cancer.* Genome Res, 2018. **28**(11): p. 1747-1756.

27. Mayers, J.R., et al., *Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development.* Nat Med, 2014. **20**(10): p. 1193-1198.

28. Ramzan, I., et al., *A Novel Dietary Intervention Reduces Circulatory Branched-Chain Amino Acids by 50%: A Pilot Study of Relevance for Obesity and Diabetes.* Nutrients, 2020. **13**(1).

29. Neinast, M.D., et al., *Quantitative Analysis of the Whole-Body Metabolic Fate of Branched-Chain Amino Acids.* Cell Metab, 2019. **29**(2): p. 417-429.e4.

30. Sheen, J.H., et al., *Defective regulation of autophagy upon leucine deprivation reveals a targetable liability of human melanoma cells in vitro and in vivo.* Cancer Cell, 2011. **19**(5): p. 613-28.

31. Cha, J.Y., et al., *Branched-chain amino acids complex inhibits melanogenesis in B16F0 melanoma cells.* Immunopharmacol Immunotoxicol, 2012. **34**(2): p. 256-64.

32. Thewes, V., et al., *The branched-chain amino acid transaminase 1 sustains growth of antiestrogen-resistant and ERα-negative breast cancer.* Oncogene, 2017. **36**(29): p. 4124-4134.

33. Zhang, B., et al., *Regulation of branched-chain amino acid metabolism by hypoxia-inducible factor in glioblastoma.* Cell Mol Life Sci, 2020.

34. White, P.J., et al., *The BCKDH Kinase and Phosphatase Integrate BCAA and Lipid Metabolism via Regulation of ATP-Citrate Lyase.* Cell Metab, 2018. **27**(6): p. 1281-1293.e7.

35. Shi, Y.B., et al., *Multifaceted Roles of Caveolin-1 in Lung Cancer: A New Investigation Focused on Tumor Occurrence, Development and Therapy.* Cancers (Basel), 2020. **12**(2).

Tables

Table 1: Univariate and multivariate analyses of clinicopathological variables and risk model in the TCGA cohort.

	Univariate analysis		Multivariate analysis	
	Hazard Ratio (95% CI)	*P* value	Hazard Ratio (95% CI)	*P* value
Age*	1.01 (1.00-1.02)	0.141	1.01 (1.00-1.02)	0.055
Gender (Male vs. Female)	1.13 (0.92-1.39)	0.257	1.03 (0.83-1.28)	0.767
Smoke (Ever-smoke vs. Non-smoke)	1.06 (0.83-1.36)	0.641	1.06 (0.83-1.35)	0.658
Risk (High vs. Low)	1.73 (1.41-2.12)	<0.001	1.61 (1.31-1.98)	<0.001
Stage (II vs. I)	1.59 (1.25-2.03)	<0.001	1.50 (1.17-1.92)	<0.001
Stage (III vs. I)	2.15 (1.65-2.79)	<0.001	1.99 (1.53-2.59)	<0.001
Stage (IV vs. I)	3.16 (2.00-4.99)	<0.001	3.35 (2.11-5.31)	<0.001

*Continuous variable

Table 2: The coefficient of Pearson's and P value of IL6, CXCR4, CD276, TGFB1, CCL2, and CD274.

Name	*Pearson's cor*	*P* value
CCL2	0.125	0.032
CD274	0.231	0.068
CD276	0.270	0.047
CXCR4	0.729	0.020
IL6	0.217	0.011
TGFB1	0.269	0.044

Figures

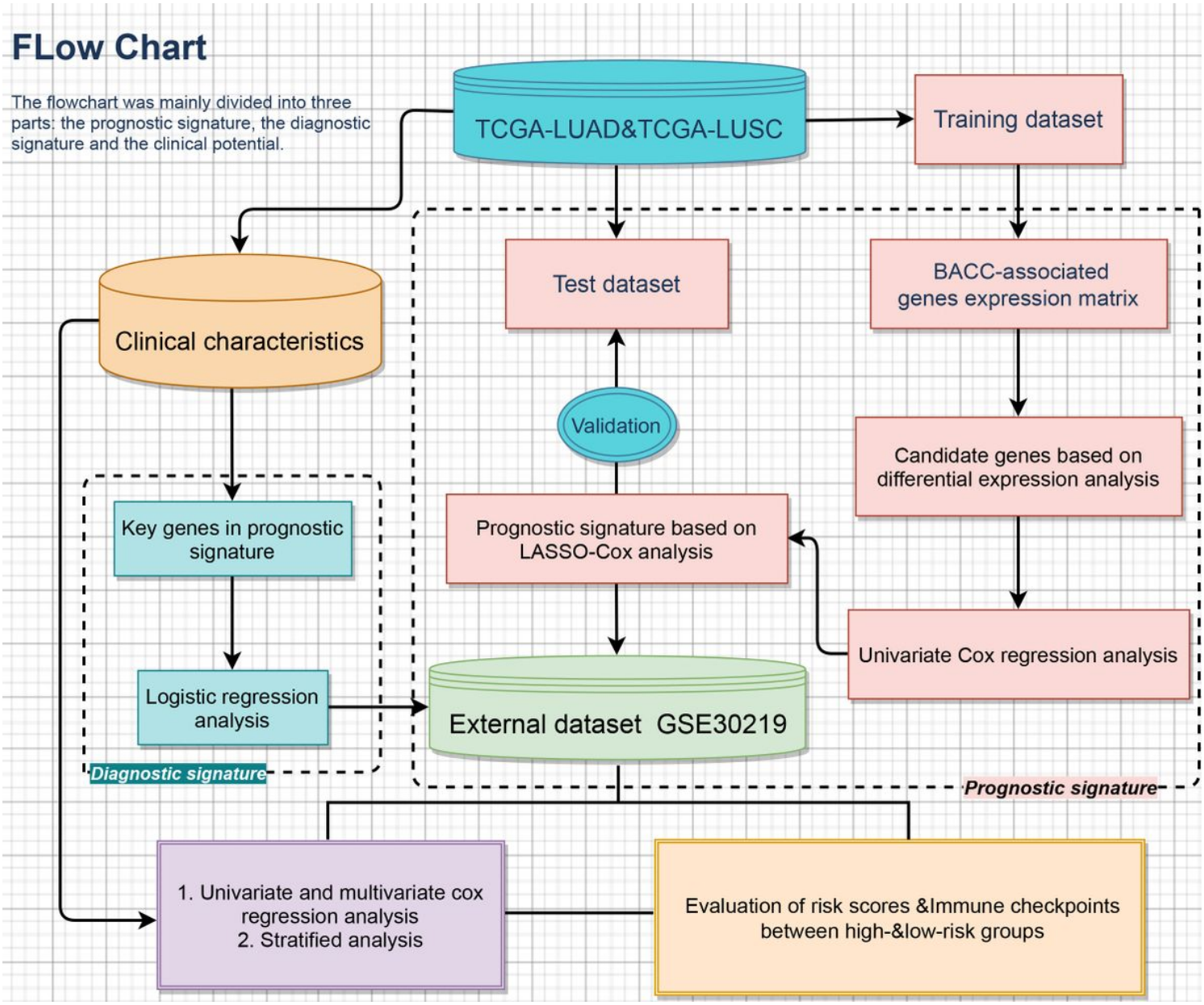

Figure 1

Flow chart of the study design. There were three parts in this study: the prognostic signature, clinical potential, and diagnostic signature. Branched-chain amino acid (BCAA)-associated signature was extracted from the datasets provided by The Cancer Genome Atlas and Gene Expression Omnibus databases. TCGA LUAD and LUSC: The Cancer Genome Atlas-lung adenocarcinoma and lung squamous cell carcinoma, and TCGA-lung squamous cell carcinoma.

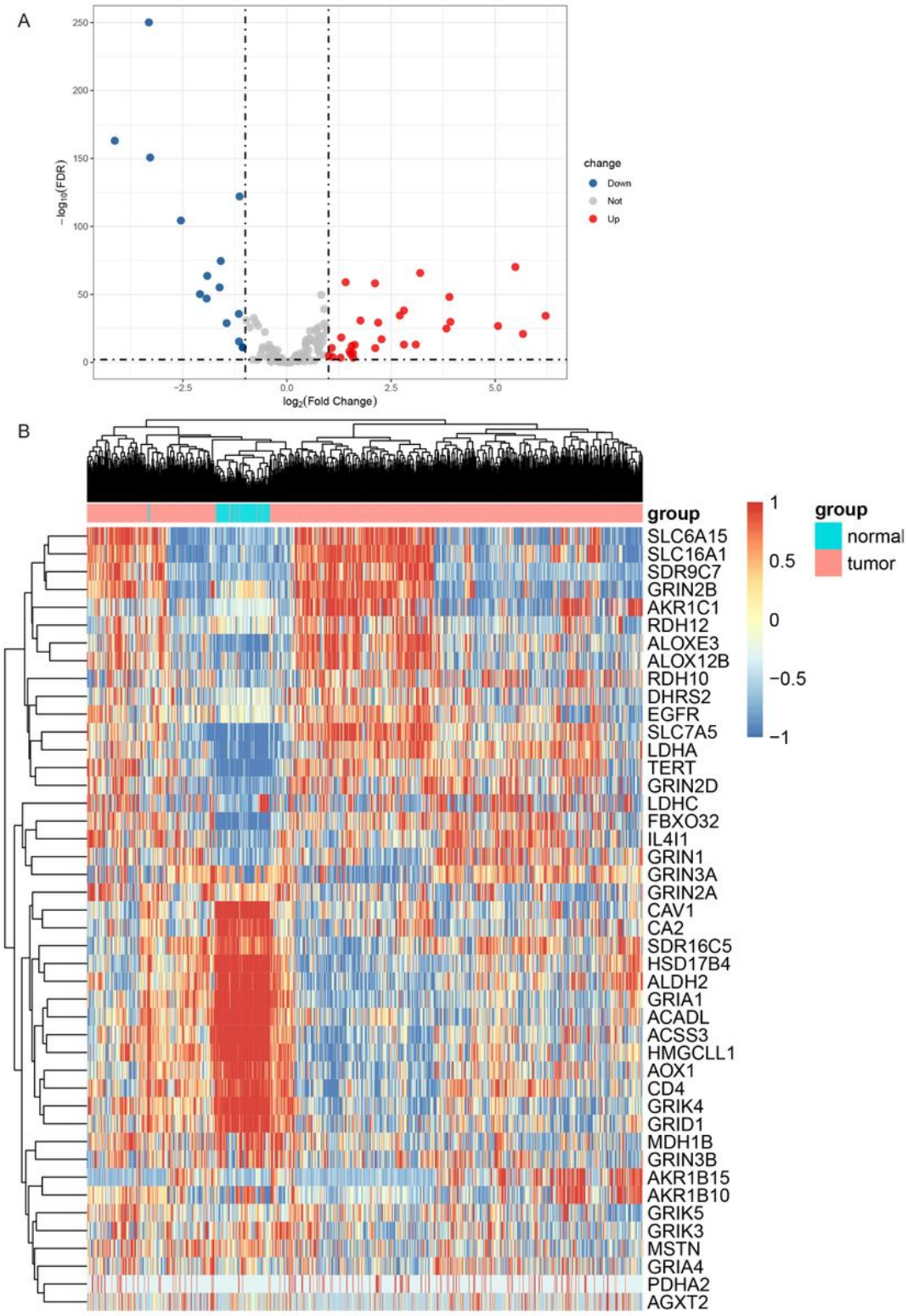

Figure 2

Identification of the candidate genes based on the differential expression analysis in The Cancer Genome Atlas cohort. (A) Volcano plot. The dotted line represents the truncation standard, the red dots represent the upregulated genes, and the blue dots represent the downregulated genes. (B) The heatmap presents the expression levels of the 44 candidate genes. FDR: False discovery rate.

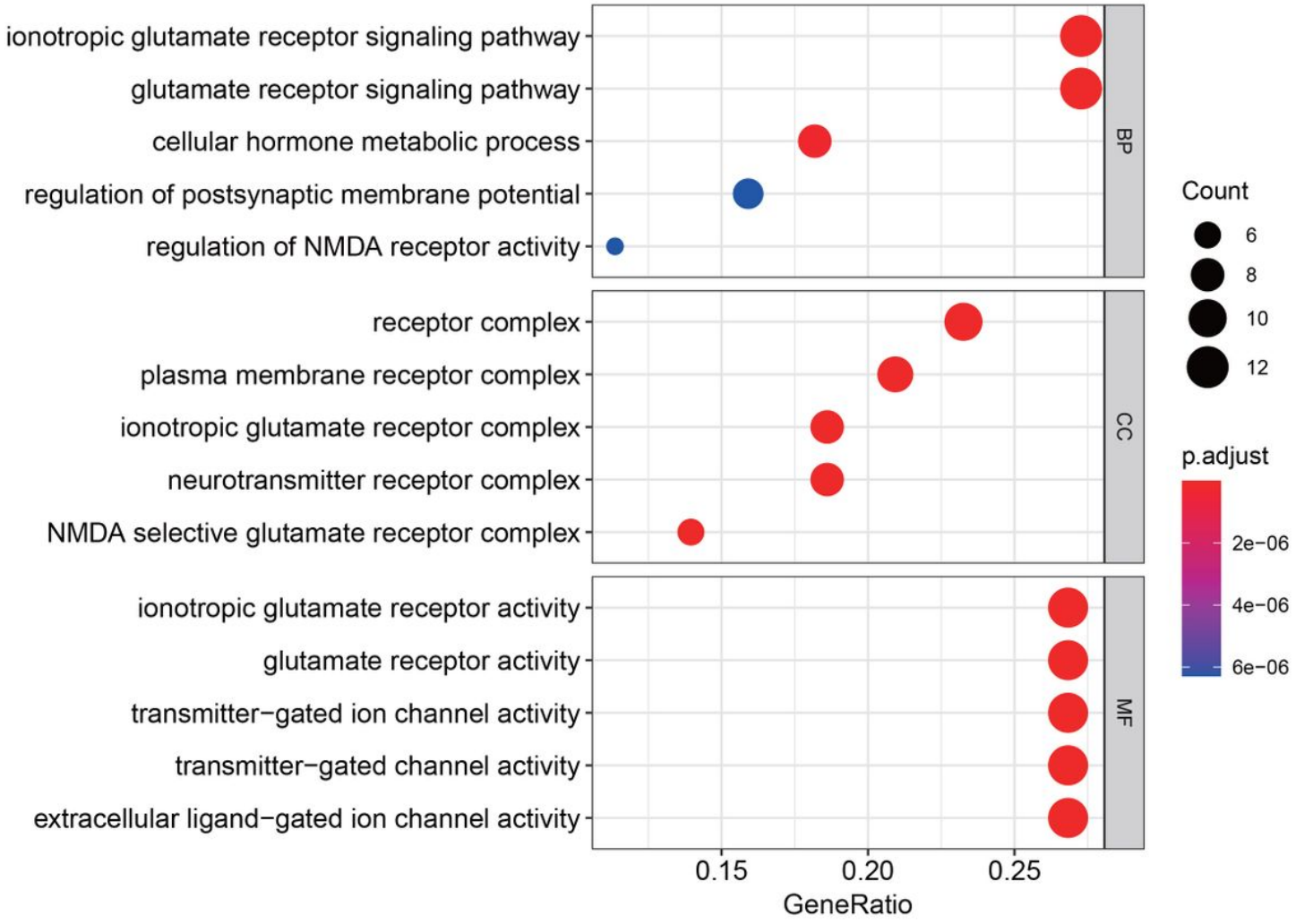

Figure 3

Gene ontology (GO) terms of three categories, biological process (BP), cellular components (CC), and molecular function (MF), which are divided into three parts.

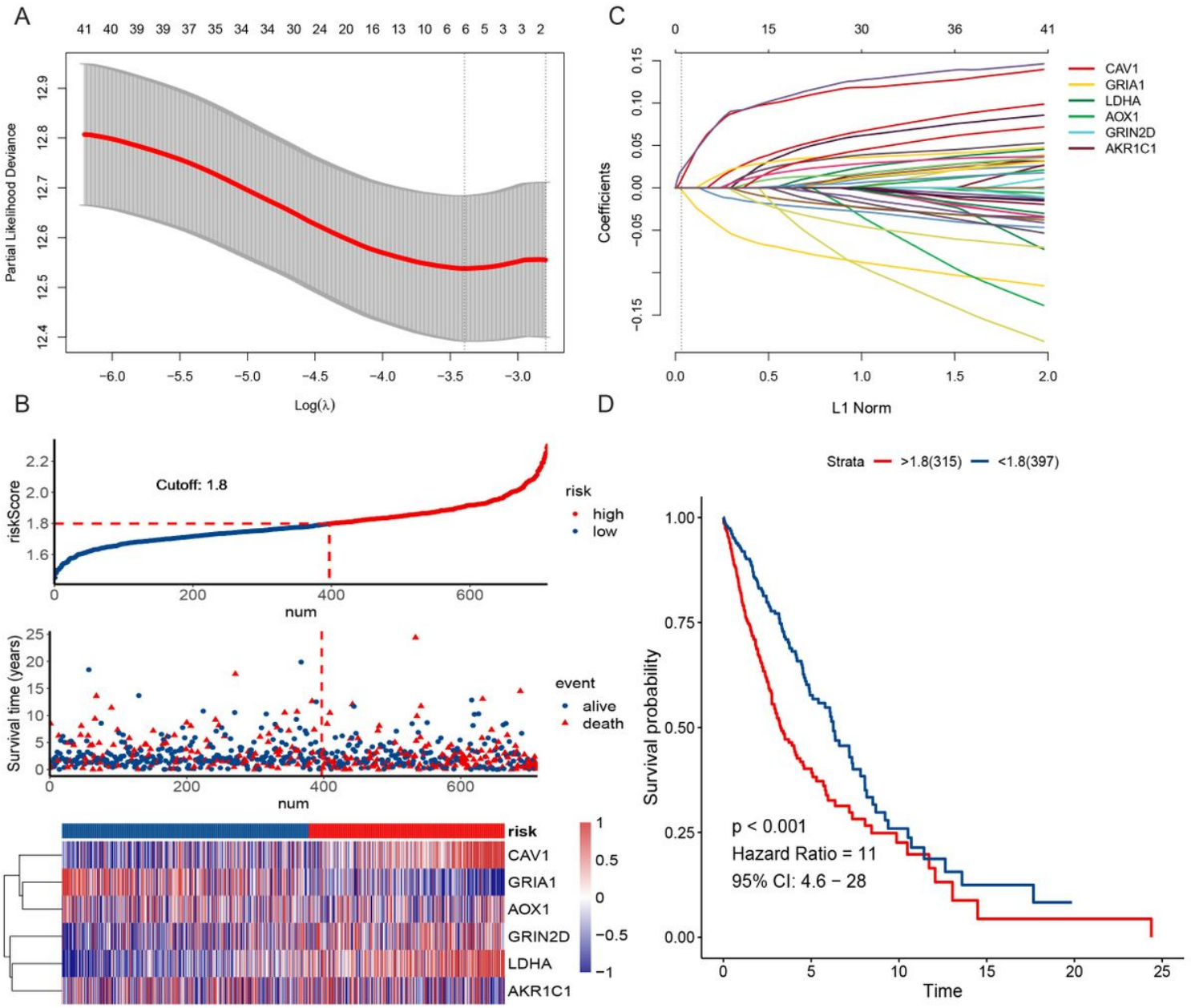

Figure 4

Result of the branched-chain amino acid (BCAA)-associated signature in the training dataset. (A & C) Least absolute shrinkage and selection operator regression identified seven key genes. (B) The patients' risk scores and survival condition, and a heatmap of key gene expression levels. (D) The survival probability analysis between risk-identified groups. CI: Confidence interval; L1: In statistics, L1 norm is one of the academic terms of L1 regularization without further explanation; Num: Number.

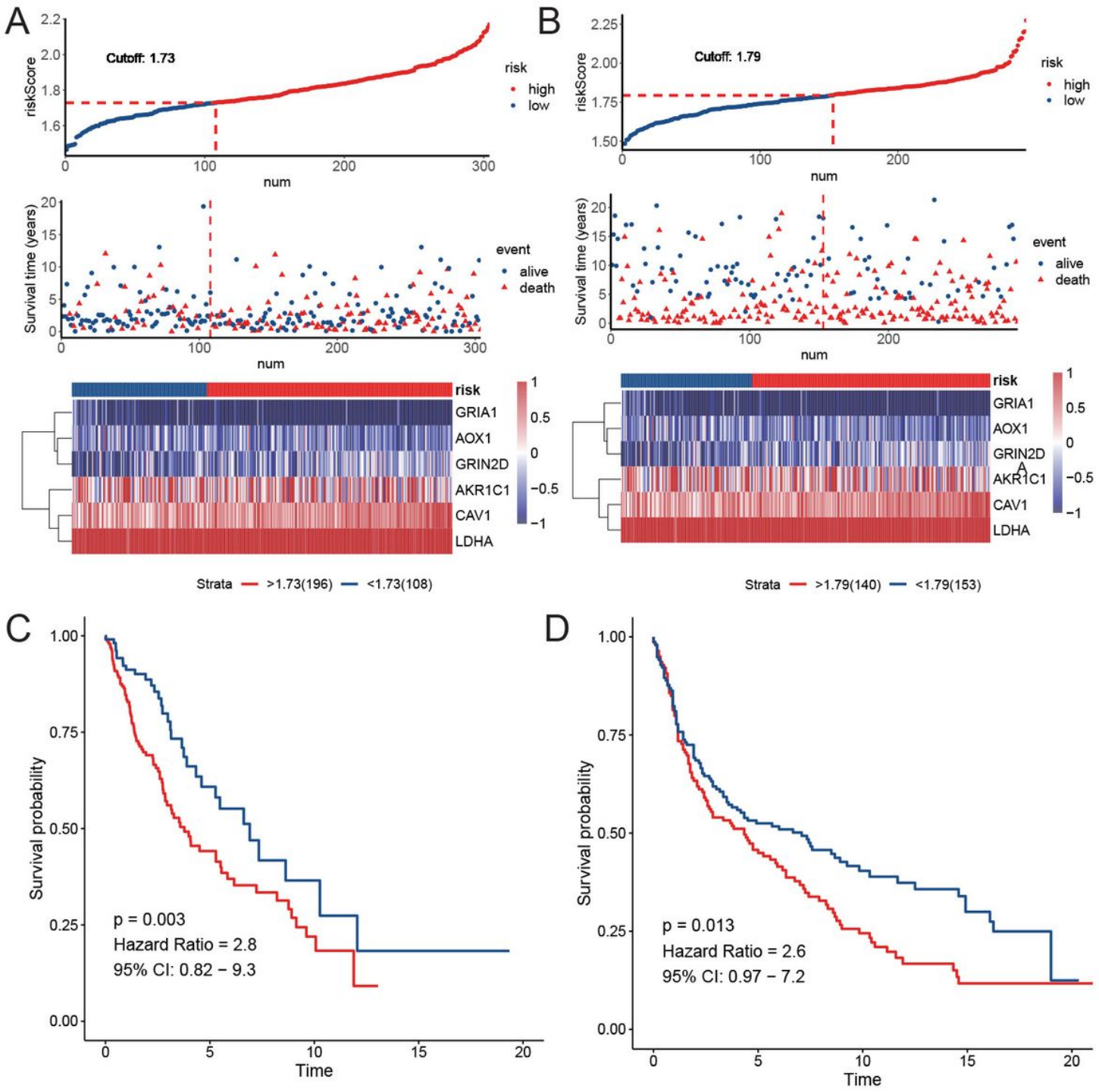

Figure 5

Verification of branched-chain amino acid (BCAA)-associated signature in the test dataset and external dataset. (A and B) The test dataset and external dataset included the patients' risk scores and survival condition, and a heatmap of key gene expression levels, respectively. (C and D) Survival analysis of time-dependent receiver operating characteristic curves of the prognostic signature in the GSE30219 dataset. CI: Confidence interval.

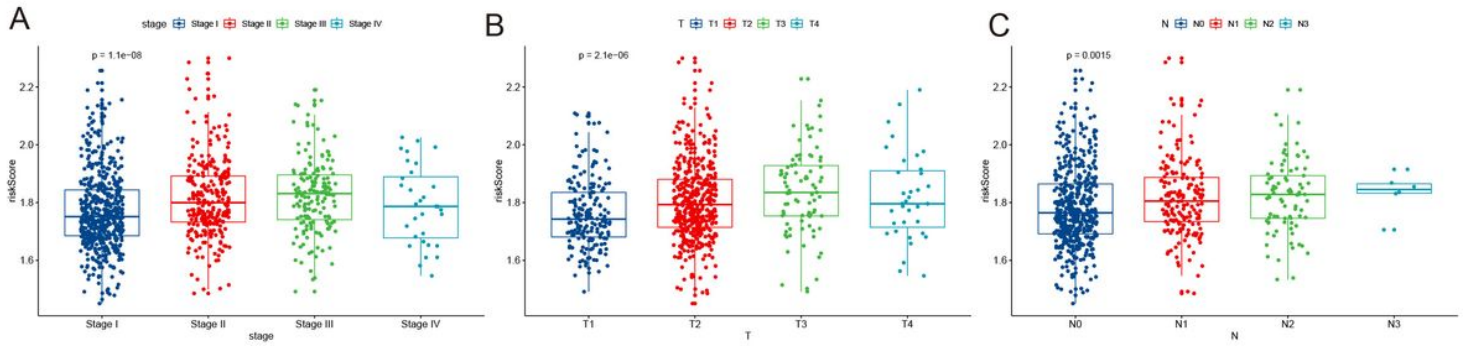

Figure 6

Stratified analysis. Relationship between the prognostic signature and TNM stage in The Cancer Genome Atlas lung cancer cohort. CI: Confidence interval; TNM: Tumor, node, and metastasis.

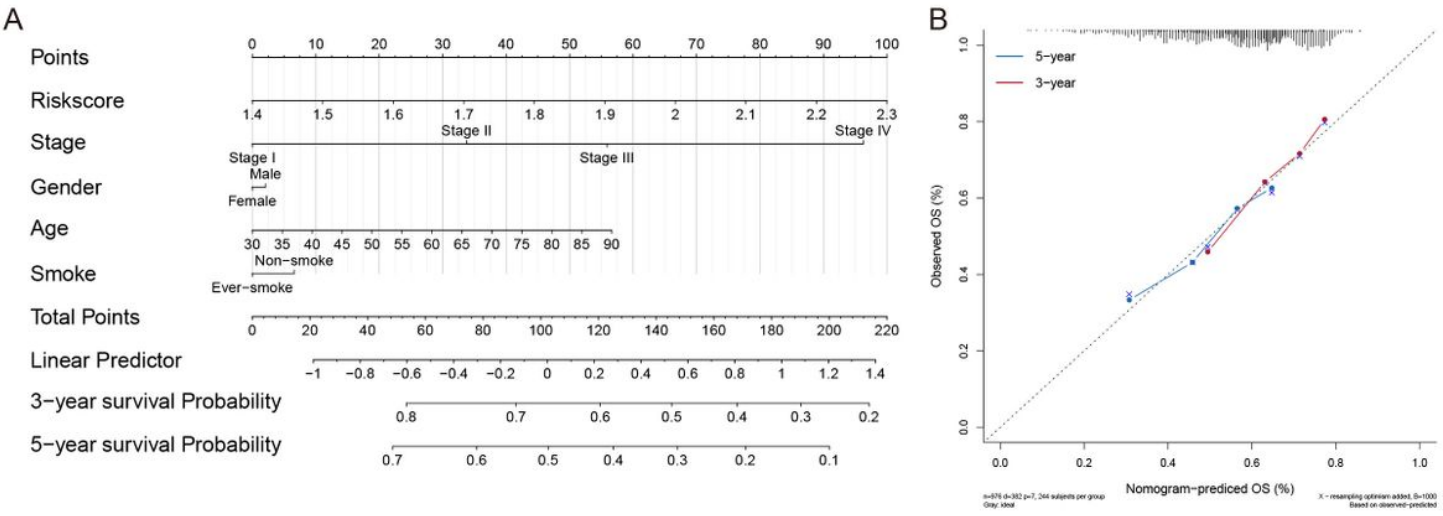

Figure 7

Structure of a nomogram on branched-chain amino acid (BCAA)-associated signature in The Cancer Genome Atlas cohort LUAD and LUSC, and LUSC only. (A) The nomogram included the signature and clinical variables. (B) Calibration curves evaluated the predictive efficiency. OS: Overall survival.

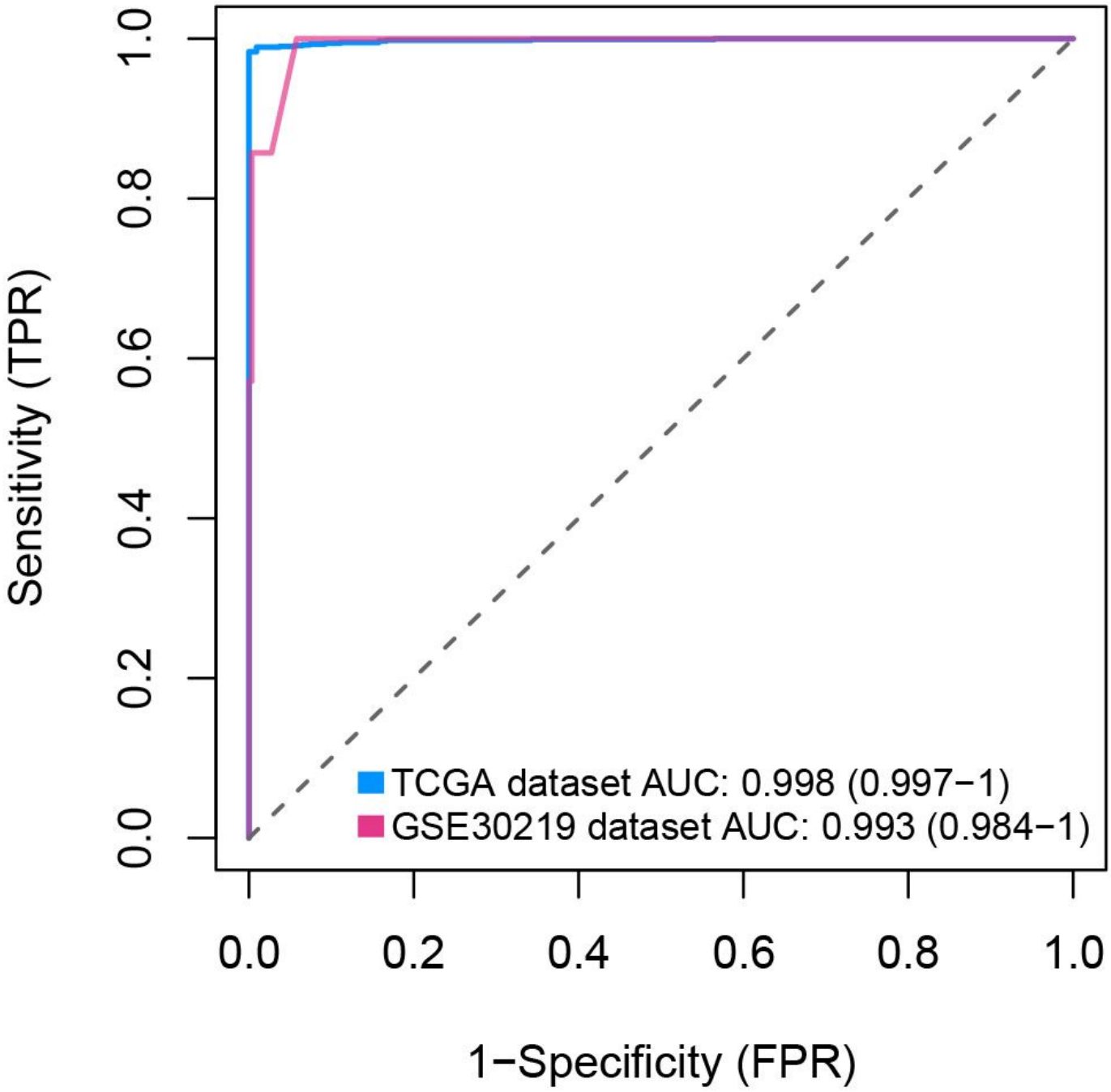

Figure 8

Receiver operating characteristic curve analysis of the efficacy of the diagnostic signature of the seven key genes in the prognostic signature. AUC, area under the curve; FPR, false positive rate; TPR, true positive rate; TCGA LUAD and LUSC, and LUSC only: The Cancer Genome Atlas-lung adenocarcinoma and lung squamous cell carcinoma, and TCGA-lung squamous cell carcinoma.

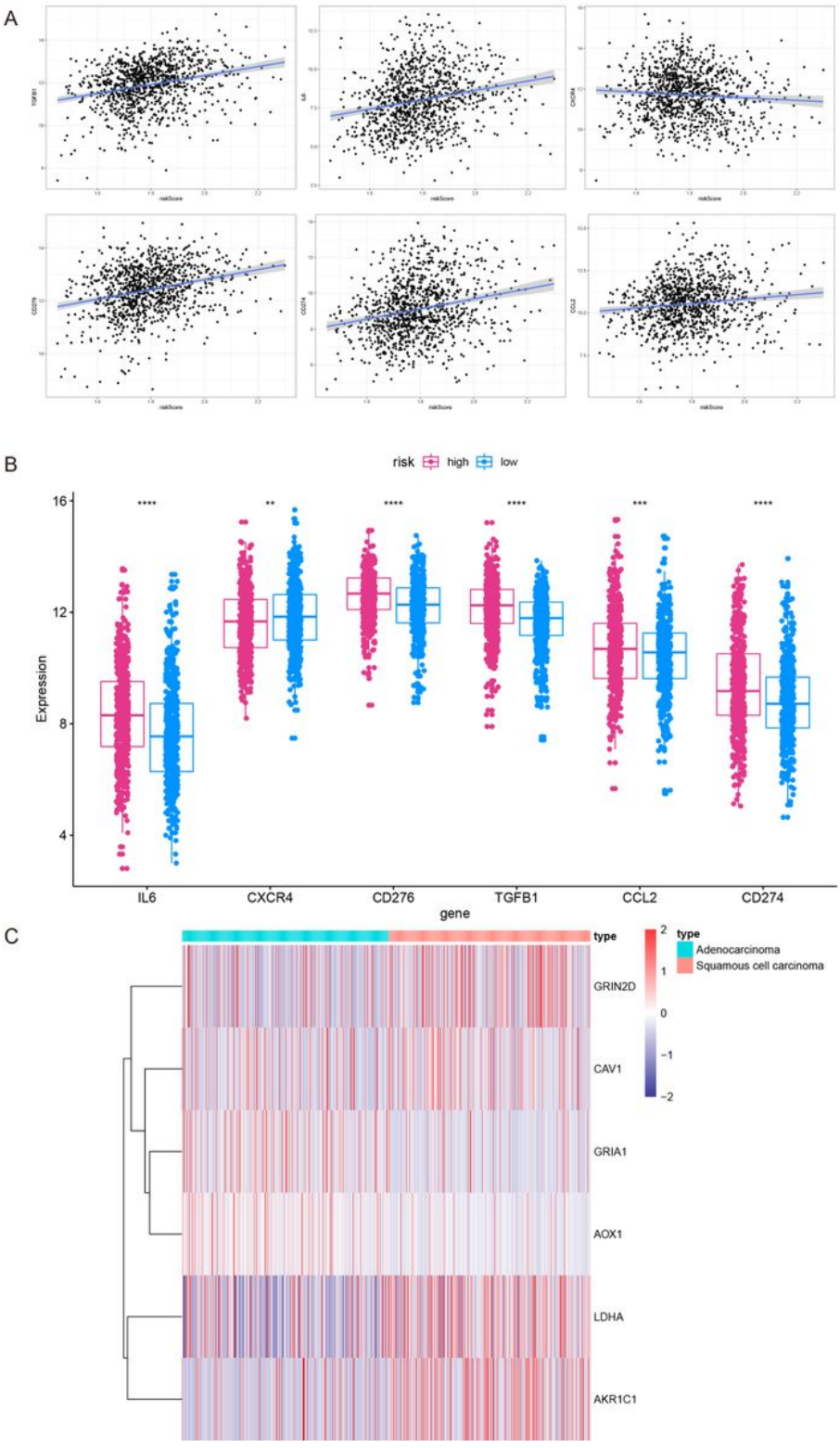

Figure 9

Verification of the diagnostic model (A) Six classical immune checkpoints, namely, IL6, CXCR4, CD276, TGFB1, CCL2, and CD274, were assessed to compare the mutation differences. (B) The Box plot presents the expression levels of the six immune checkpoints genes in LUAD and LUSC. (C) The heatmap presents the expression levels of the CAV1, GRIA1, LDHA, AOX1, GRIN2D, and AKR1C1 in LUAD and LUSC.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1BCAAassociatedgenes.csv
- TableS2differentialexpressiongenes.csv
- TableS3GOfuctionanalysis.csv