Identification and Validation of an Anoikis-Related Prognostic Signature for Oesophageal Carcinoma Based on Integrated TCGA and GTEx Data

Hailun Li*
Second Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou 510405, China
*Corresponding email: helenli200549@foxmail.com

Abstract

Oesophageal carcinoma (ESCA) is a highly lethal malignancy for which reliable prognostic biomarkers remain limited. Anoikis resistance is recognised as a pivotal mechanism driving tumor progression and metastatic dissemination. This study aimed to develop a novel anoikis-related gene signature to predict overall survival in patients with ESCA. Ribonucleic acid (RNA)-sequencing data from The Cancer Genome Atlas (TCGA) oesophageal cancer cohort and normal oesophageal tissue samples from the Genotype-Tissue Expression (GTEx) project were integrated. Differentially expressed genes (DEGs) between tumor and normal tissues were intersected with curated anoikis-related gene sets. Univariate Cox regression, least absolute shrinkage and selection operator (LASSO) regression, and multivariate Cox regression were sequentially applied to construct a prognostic signature. Model performance was validated using Kaplan-Meier survival analysis, Time-dependent Receiver Operating Characteristic (tdROC) curves, and calibration plots. A total of 263 anoikis-related DEGs were identified. Following LASSO and multivariate Cox analyses, a 16-gene signature was established, comprising growth arrest and deoxyribonucleic acid (DNA) damage inducible beta (GADD45B), bombesin receptor subtype 3 (BRS3), paired box 4 (PAX4), folate receptor gamma (FOLR3), interleukin 17A (IL17A), proline rich and Gla domain 3 (PRRG3), nuclear factor I C (NFIC), von Willebrand factor D and EGF domains (VWDE), G protein-coupled receptor 26 (GPR26), hyperpolarization-activated cyclic nucleotide-gated potassium channel 1 (HCN1), GRAM domain containing 1C (GRAMD1C), calpain-12 (CAPN12), sodium/potassium transporting ATPase interacting 3 (NKAIN3), haptoglobin-related protein (HPR), cTAGE family member 15 (CTAGE15), and Rho GTPase Activating Protein 11B (ARHGAP11B). The risk score derived from this signature significantly stratified patients into high-risk and low-risk groups ($p<0.0001$). TdROC analysis demonstrated robust predictive accuracy, with area under the curve (AUC) values of 0.924, 0.935, and 0.952 for 1-year, 3-year, and 5-year overall survival, respectively. Calibration curves indicated excellent agreement between nomogram-predicted and observed survival probabilities. The 16-gene anoikis-related signature represents a robust and independent prognostic tool for ESCA and may facilitate personalised therapeutic decision-making.

Keywords

Oesophageal carcinoma, Anoikis, Prognostic signature, The cancer genome atlas, Genotype-Tissue Expression, Least absolute shrinkage and selection operator regression

Introduction

Oesophageal carcinoma ranks among the most aggressive malignancies worldwide and is typically characterised by late-stage diagnosis and dismal five-year survival rates [1]. Despite notable advances in multimodal therapeutic strategies, the prognosis for patients with ESCA remains unfavourable, underscoring the critical need for reliable prognostic biomarkers and individualised risk-stratification approaches [2].

Anoikis, a specialised form of programmed cell death triggered by the loss of cell-matrix attachment, constitutes a crucial intrinsic barrier against metastatic dissemination. During malignant progression, tumor cells frequently acquire resistance to anoikis, thereby facilitating their survival within the circulatory system and promoting subsequent colonisation of distant organs. Accumulating evidence has highlighted the prognostic

relevance of anoikis-related genes across multiple cancer types. Nevertheless, a comprehensive anoikis-based prognostic signature for oesophageal carcinoma has yet to be fully characterized [3,4].

In the present study, we integrated transcriptomic data from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) project to systematically identify anoikis-associated differentially expressed genes. A 16-gene prognostic signature was established through rigorous machine learning algorithms and Cox regression analyses, and its predictive performance was thoroughly assessed. In addition, a nomogram incorporating relevant clinical parameters was developed to enhance translational applicability. Collectively, our findings provide a novel molecular framework for prognostic evaluation in ESCA and may inform future individualised therapeutic strategies.

Materials and methods

Data acquisition and preprocessing

Ribonucleic acid (RNA)-sequencing data (counts and transcripts per million [TPM] values) for oesophageal carcinoma were obtained from TCGA database. Normal oesophageal tissue expression profiles were retrieved from GTEx project [5,6]. Corresponding clinical information, including overall survival time and vital status, was extracted from the TCGA clinical annotation files.

Batch effect correction and differential expression analysis

To mitigate potential platform-specific batch effects between the TCGA and GTEx datasets, the ComBat algorithm implemented in the "sva" R package was applied to log2-transformed TPM values. Differentially expressed genes (DEGs) between tumor and normal samples were identified using the "DESeq2" package with the following thresholds: |log2 fold change| > 1 and false discovery rate (FDR) <0.05.

Identification of anoikis-related DEGs

A curated list of anoikis-related genes was compiled from the GeneCards database and previously published literature [7,8]. The intersection between the identified DEGs and the anoikis-related gene set was visualised using a Venn diagram.

Construction of the prognostic gene signature

Univariate Cox regression analysis was performed on the anoikis-related DEGs to identify genes significantly associated with overall survival ($p<0.05$). Least absolute shrinkage and selection operator (LASSO) regression with ten-fold cross-validation was subsequently employed using the "glmnet" R package to further refine the candidate gene set and mitigate overfitting [9]. The optimal penalty parameter (λ) was determined based on the minimum partial likelihood deviance. Multivariate Cox regression was then applied to establish the final coefficients for each gene retained in the signature. The risk score for each patient was calculated using the following formula:

$$Risk\ Score = \Sigma\ (Expression\_gene\ i \times Coef\_i) \quad (1)$$

Patients were stratified into high- and low-risk groups using the median risk score as the cutoff.

Validation and performance assessment

Kaplan-Meier survival curves were generated using the "survival" and "survminer" R packages to compare overall survival between the high- and low-risk groups, and statistical significance was evaluated using the log-rank test [10]. Time-dependent receiver operating characteristic (tdROC) curves were constructed using the "tdROC" R package to assess the predictive accuracy of the signature at 1, 3, and 5 years, and the corresponding area under the curve (AUC) values were calculated. A nomogram integrating the risk score and clinical stage was developed using the "rms" R package, and calibration curves were plotted to examine the concordance between predicted and observed survival probabilities [11,12].

Statistical analysis

All statistical analyses were performed using R software (version 4.2.0). A two-sided $p<0.05$ was considered statistically significant.

Result

Identification of differentially expressed genes in ESCA

Following batch effect correction between the TCGA and GTEx datasets, principal component analysis (PCA) demonstrated substantial overlap between samples, confirming successful removal of technical variation. Differential expression analysis identified a total of 4,041 DEGs between oesophageal tumors and normal tissues,

comprising 2,161 upregulated and 1,880 downregulated genes (Figure 1a). A heatmap of the 50 most significantly altered DEGs revealed distinct expression patterns distinguishing tumor from normal samples (Figure 1b).

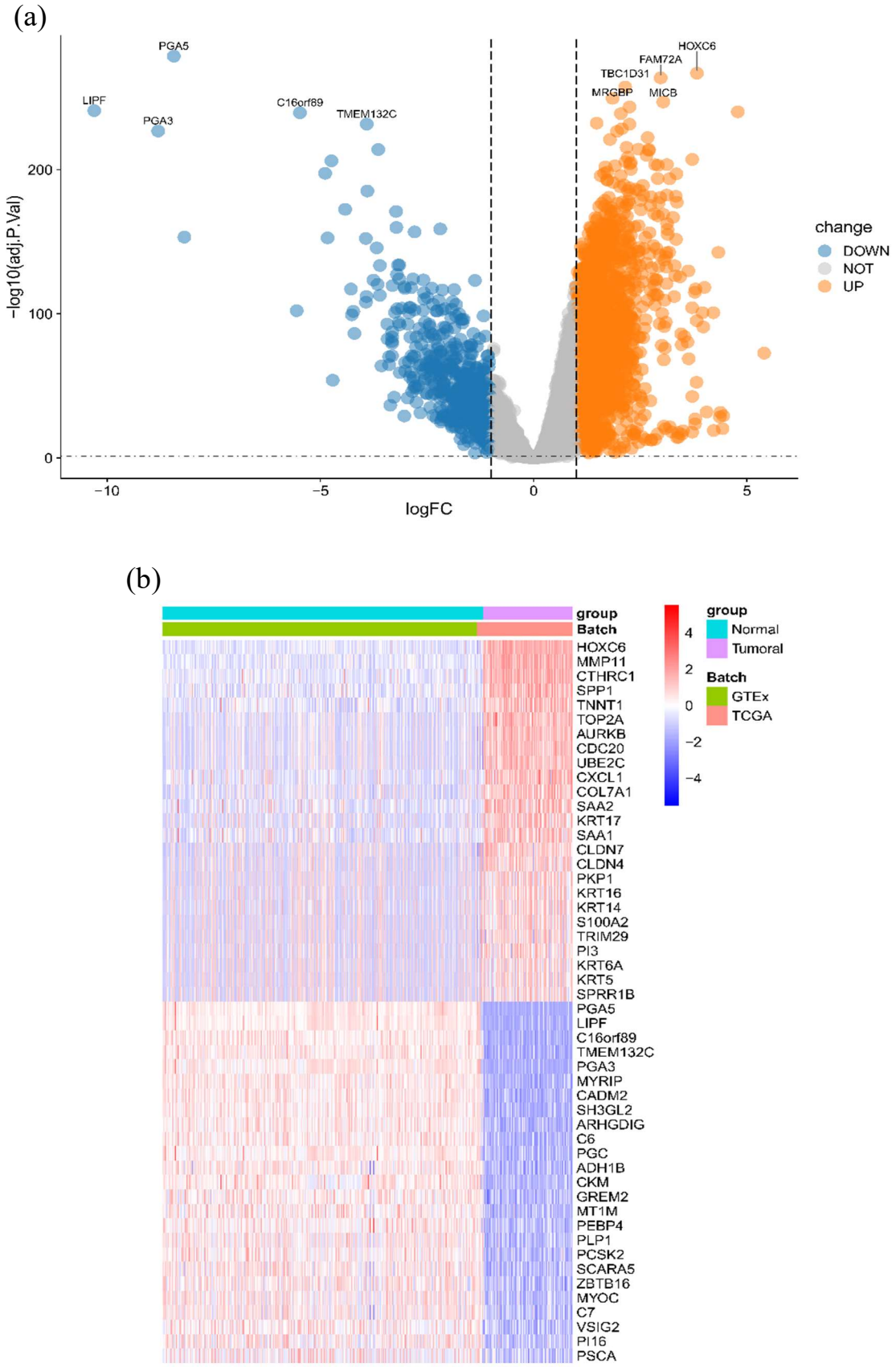


Figure 1. Identification of differentially expressed genes in esophageal carcinoma. (a) Volcano plot depicting DEGs between tumor and normal tissues. (b) Heatmap of the top 50 most significant DEGs.

Screening of anoikis-related prognostic DEGs

Intersection of the full DEGs list with the anoikis-related gene set yielded a total of 263 anoikis-related DEGs (Figure 2). These genes were subsequently subjected to univariate Cox regression analysis to evaluate their individual prognostic relevance.

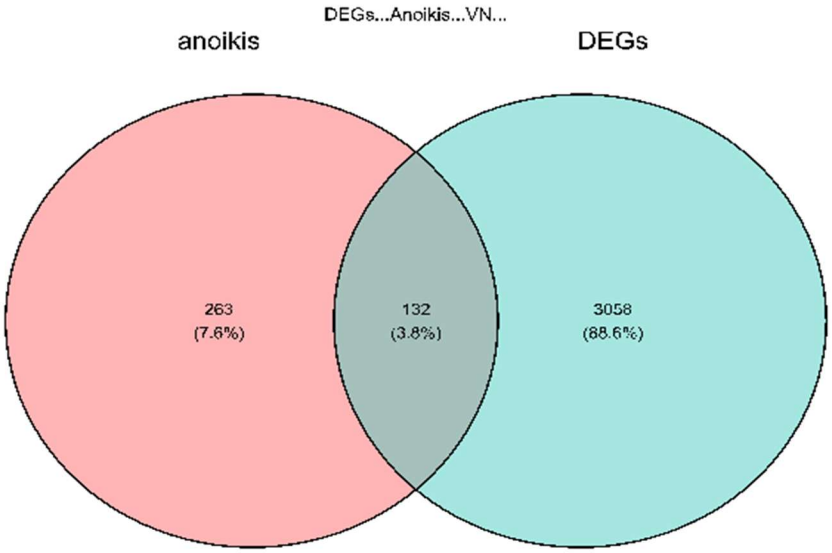


Figure 2. Venn diagram illustrating the overlap between differentially expressed genes and curated anoikis-related gene sets.

Construction of a 16-gene prognostic signature

Univariate Cox regression analysis identified 144 genes significantly associated with overall survival ($p<0.05$). LASSO regression was subsequently applied to further refine the gene set and mitigate overfitting, with the optimal lambda value selected based on the minimum partial likelihood deviance (Figures 3a and 3b). Multivariate Cox regression retained 16 genes with independent prognostic significance (Figure 3c). The corresponding hazard ratios (HRs) and 95% confidence intervals for each gene in the final model are summarised in Table 1. Notably, GADD45B (HR = 2.924, $p<0.001$) and BRS3 (HR = 1.598, $p<0.001$) exhibited the strongest associations with increased risk, whereas GPR26 (HR = 0.523, $p<0.001$) and CAPN12 (HR = 0.617, $p<0.001$) conferred protective effects.

Table 1. Multivariate Cox regression results for the 16 genes included in the prognostic signature.

Gene	Hazard ratio (HR) (95% confidence interval (CI))	p-value
GADD45B	2.924 (2.069-4.131)	<0.001
BRS3	1.598 (1.302-1.962)	<0.001
GRAMD1C	1.577 (1.204-2.065)	<0.001
ARHGAP11B	1.519 (1.083-2.130)	0.015
PRRG3	1.492 (1.226-1.815)	<0.001
FOLR3	1.441 (1.256-1.652)	<0.001
HPR	1.269 (0.999-1.611)	0.051
PAX4	1.200 (1.075-1.339)	0.001
IL17A	1.164 (1.038-1.304)	0.009
VWDE	0.881 (0.800-0.970)	0.010
NKAIN3	0.835 (0.693-1.004)	0.056
CTAGE15	0.748 (0.661-0.848)	<0.001
HCN1	0.690 (0.580-0.822)	<0.001
NFIC	0.622 (0.410-0.943)	0.025
CAPN12	0.617 (0.503-0.756)	<0.001
GPR26	0.523 (0.401-0.683)	<0.001

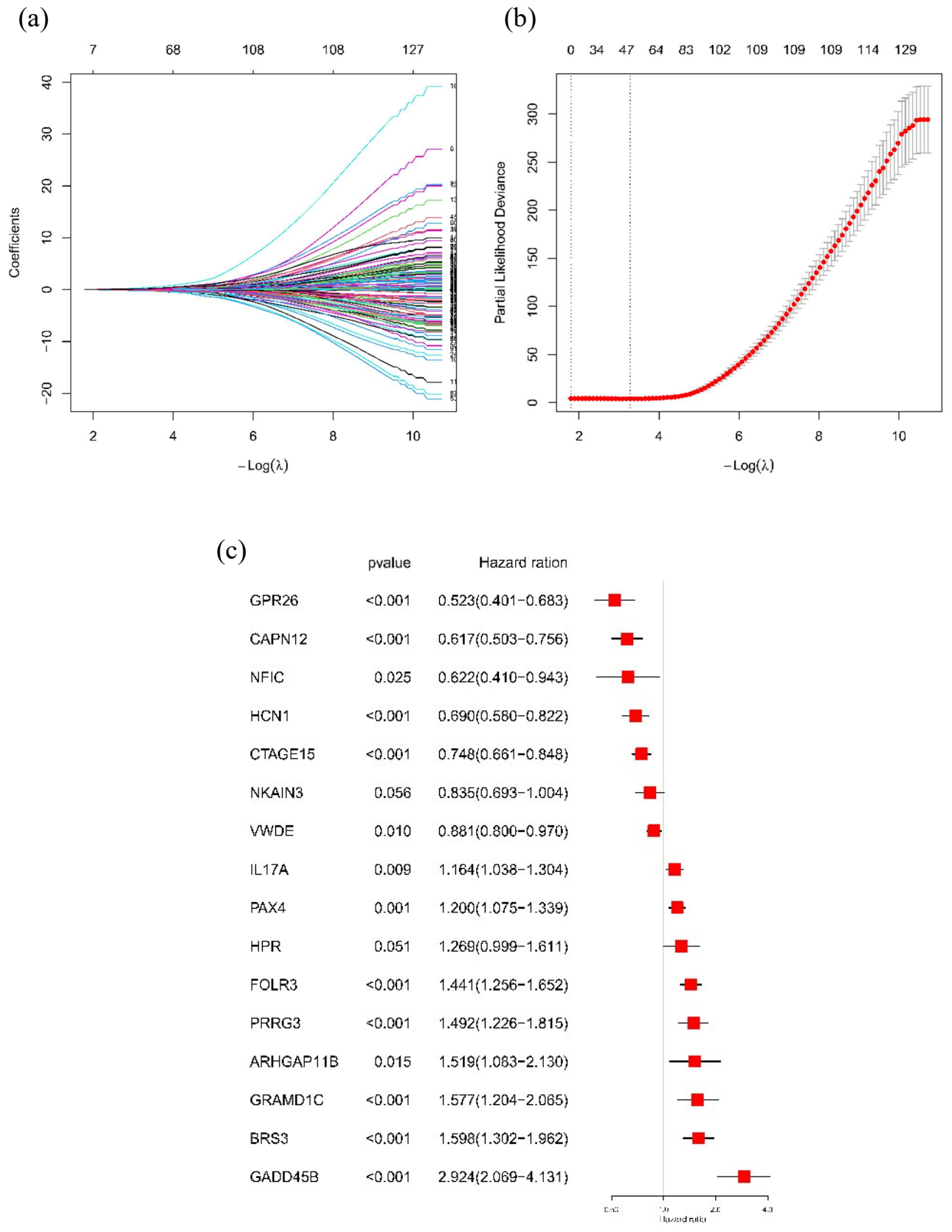
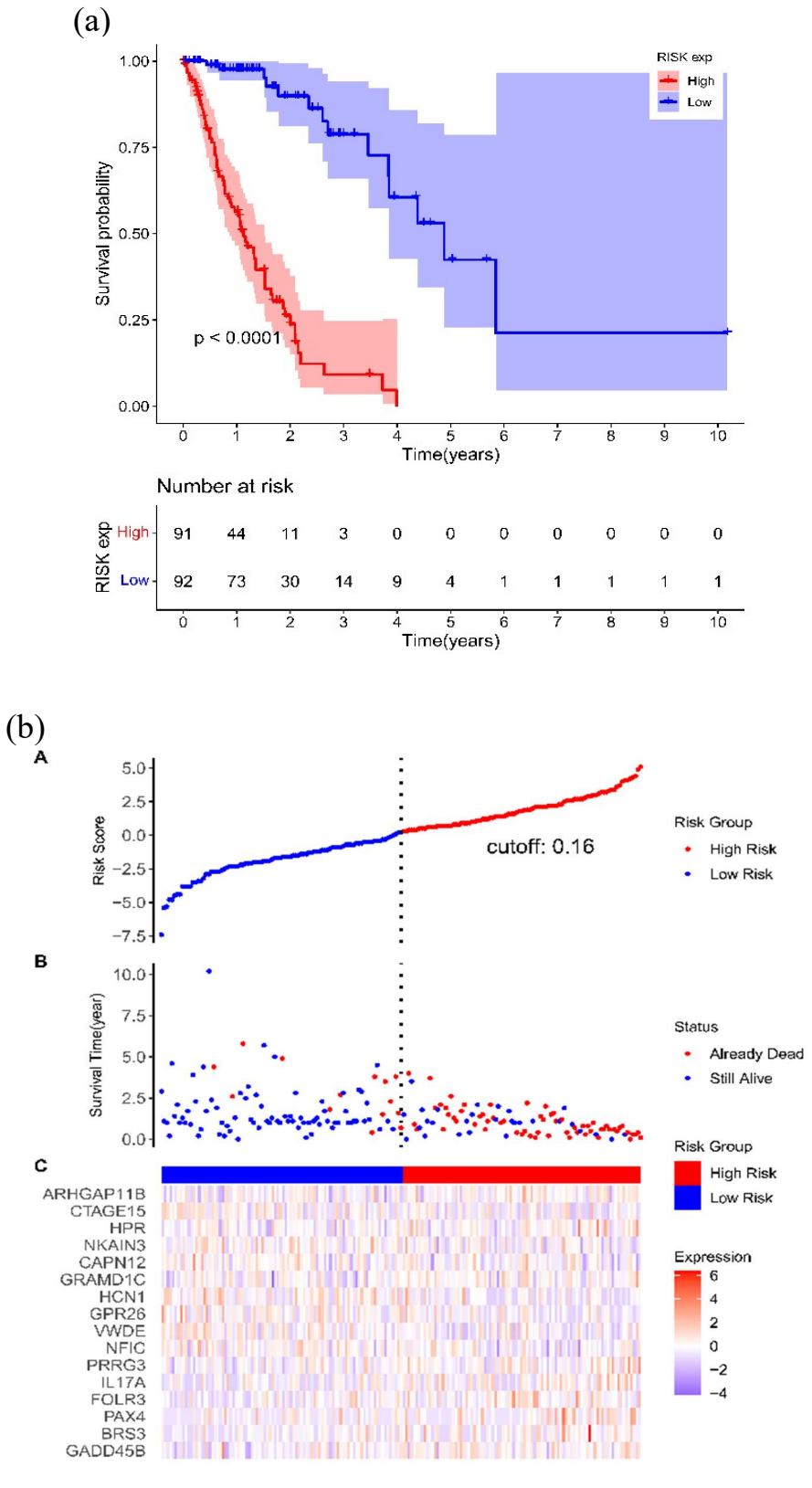


Figure 3. Construction of the prognostic gene signature using LASSO and multivariate Cox regression. (a) LASSO coefficient profiles of candidate genes. (b) Ten-fold cross-validation for optimal tuning parameter (λ) selection in the LASSO model. (c) Forest plot displaying the multivariate Cox regression results for the 16 genes retained in the final signature.

Validation of the risk signature in the TCGA cohort

Patients were dichotomised into high-risk and low-risk groups according to the median risk score. Kaplan-Meier analysis revealed significantly poorer overall survival in the high-risk group compared with the low-risk group ($p < 0.0001$; Figure 4a). The distribution of risk scores, patient survival status, and expression patterns of the 16 signature genes across the cohort are displayed in Figure 4b. Time-dependent ROC curves demonstrated robust predictive performance, with AUC values of 0.924, 0.935, and 0.952 for 1-year, 3-year, and 5-year overall survival, respectively (Figure 4c).



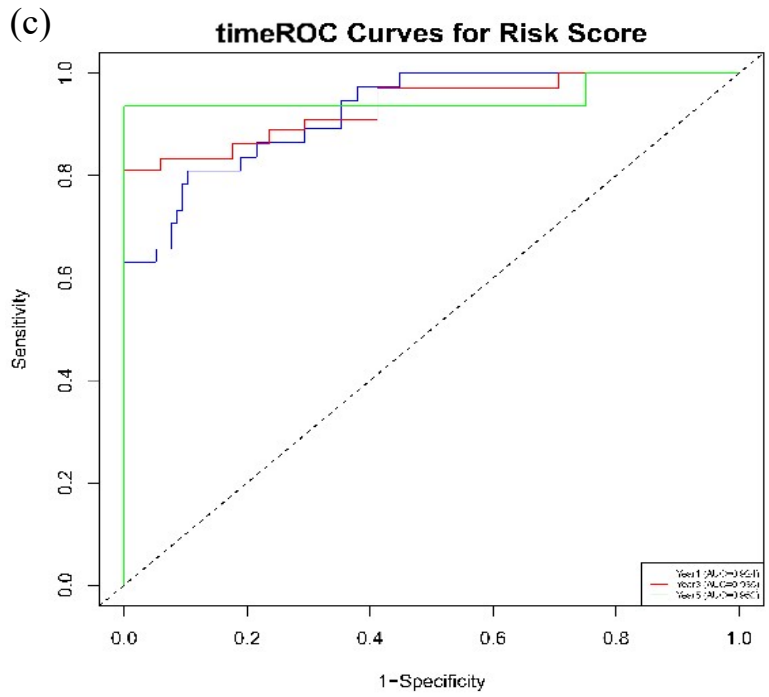
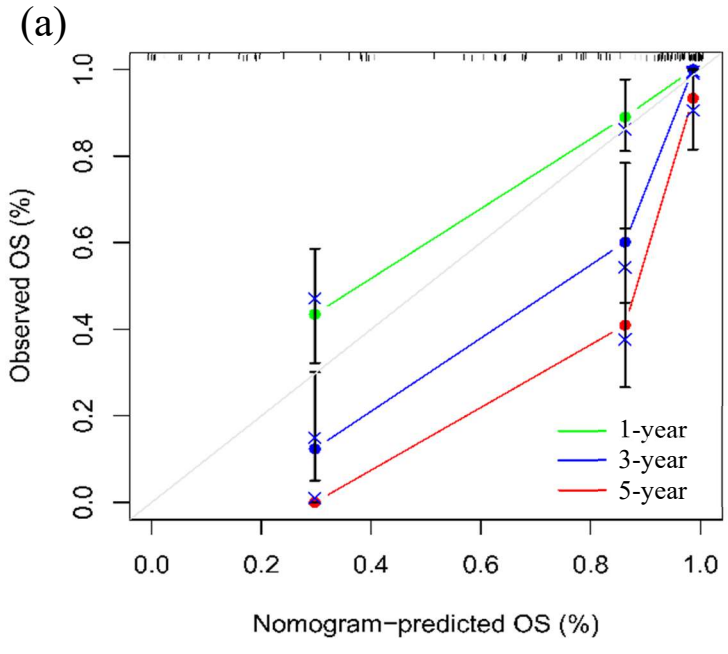


Figure 4. Validation of the 16-gene risk signature in the TCGA cohort. (a) Kaplan-Meier survival curves for patients stratified into high- and low-risk groups according to the median risk score ($p < 0.0001$, log-rank test). (b) Distribution of risk scores, survival status, and expression heatmap of the 16 signature genes across patients. (c) Time-dependent receiver operating characteristic (tdROC) curves evaluating the predictive accuracy of the signature at 1, 3, and 5 years.

Independent prognostic value and nomogram development

Multivariate Cox regression incorporating the risk score together with clinical covariates (age, gender, and tumor stage) confirmed that the risk score remained an independent prognostic factor (HR=2.680, 95% CI: 2.18-3.29, p<0.001).

A nomogram integrating the risk score and tumor stage was constructed to estimate 1-year, 3-year, and 5-year overall survival probabilities (Figure 5a). Calibration curves for these time points exhibited excellent concordance between nomogram-predicted and actual observed survival, supporting the reliability of the model (Figure 5b).



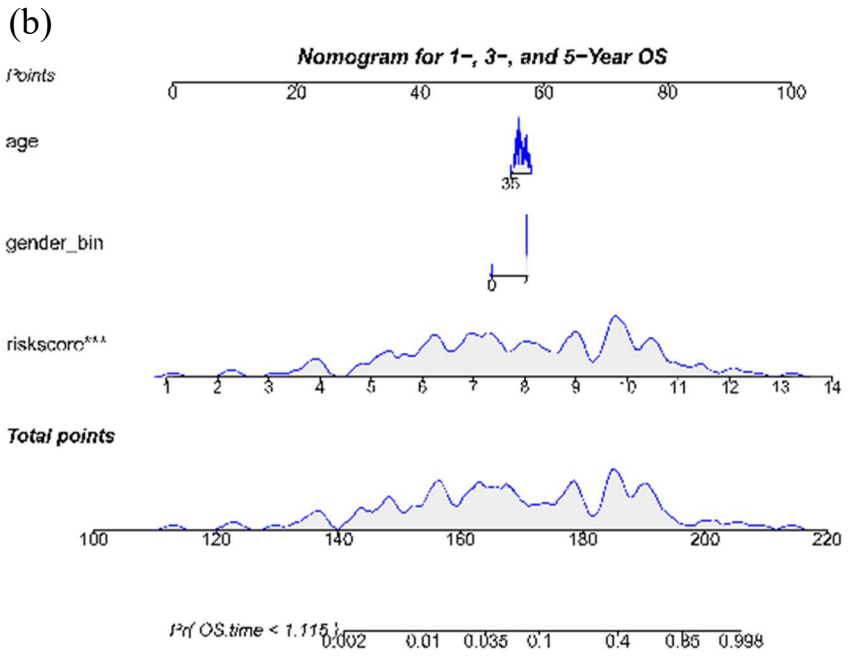


Figure 5. Nomogram construction and calibration. (a) Nomogram integrating age, gender, and the anoikis-related risk score. (b) Calibration curves of the nomogram at 1, 3, and 5 years.

Clinical characteristics of risk groups

Comparison of baseline clinical features between the high- and low-risk groups revealed no statistically significant differences in age, gender, or tumor stage distribution (all $p>0.05$; Table 2), indicating that the prognostic value of the signature is independent of these potential confounding factors.

Table 2. Baseline clinical characteristics of patients in the high- and low-risk groups.

Variable	High-risk (n=92.00)	Low-risk (n=91.00)	p-value
Age (mean ± SD)	61.50 ± 11.20	60.80 ± 10.50	0.63
Gender (Male/Female)	78.00/14.00	76.00/15.00	0.82
Stage (I-II / III-IV)	80.00/12.00	80.00/11.00	0.85

Note: SD, standard deviation. P-values were calculated using the student's t-test for continuous variables and the chi-square test or Fisher's exact test for categorical variables.

Discussion

In the present study, we successfully construct and validated a 16-gene anoikis-related prognostic signature for oesophageal carcinoma by integrating transcriptomic data from TCGA and GTEx. The signature effectively stratified patients into distinct risk categories and exhibited robust predictive performance, as evidenced by time-dependent tdROC analysis and calibration curves. Moreover, the risk score was identified as an independent prognostic factor, underscoring its potential utility beyond traditional clinicopathological parameters.

The 16 genes comprising the signature encompass both established oncogenic drivers and relatively under-characterised players in cancer biology. GADD45B, the strongest risk factor in our model (HR=2.920), encodes a stress sensor involved in the regulation of apoptosis, cell cycle arrest, and DNA repair [13]. Its overexpression has been associated with poor prognosis in multiple malignancies, including gastric and colorectal cancers, which aligns with our findings in ESCA [14,15]. BRS3 is an orphan G protein-coupled receptor that has been implicated in tumor cell proliferation and metastatic dissemination [16]. In addition, IL17A, a pro-inflammatory cytokine, has been shown to promote tumor progression and immune evasion in oesophageal squamous cell carcinoma, which may partially account for its risk-conferring role (HR=1.164) in our model

[17,18]. FOLR3 and PAX4 have also been reported to exhibit oncogenic properties in certain cancer types, although their functions in oesophageal carcinoma warrant further exploration [19].

Conversely, several genes in the signature exhibited protective effects. GPR26, an orphan G protein-coupled receptor, has been described as a tumor suppressor in glioblastoma and breast cancer, yet its role in oesophageal carcinoma remains virtually unexplored. CAPN12, a member of the calcium-dependent cysteine protease family, has been linked to modulation of apoptosis and cell migration, but its specific function in ESCA has not been characterized. The divergent prognostic implications of these genes highlight the complexity of anoikis-related pathways and underscore the need for functional studies to elucidate the underlying molecular mechanisms.

Several limitations of this study should be acknowledged. First, the analysis is entirely retrospective and relies on publicly available datasets. External validation in independent cohorts is therefore warranted to confirm the generalizability of the signature. Second, in the absence of experimental validation, the mechanistic contributions of the identified genes to anoikis resistance and oesophageal carcinogenesis remain speculative. Future investigations employing vitro and vivo models are necessary to dissect the biological functions of these genes. Third, the sample size of the TCGA oesophageal cohort is relatively modest, and larger prospective studies will be required to evaluate the clinical applicability of the signature.

Conclusion

In summary, we develop and validate a novel 16-gene anoikis-related prognostic signature that reliably predicts overall survival in patients with oesophageal carcinoma. The signature, in combination with a nomogram integrating clinical variables, offers a practical tool for individualised risk stratification and may inform future therapeutic decision-making. Further studies are needed to elucidate the underlying molecular mechanisms and to validate the signature in prospective clinical settings.

Funding

This work was not supported by any fundings.

Acknowledgments

The author would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interests

The author declares no conflict of interest.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021) Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209-249.

[2] Thrift, A. P. (2021) Global burden and epidemiology of Barrett oesophagus and oesophageal cancer. *Nature Reviews Gastroenterology and Hepatology*, 18(6), 432-443.

[3] Adeshakin, F. O., Adeshakin, A. O., Afolabi, L. O., Yan, D., Zhang, G., Wan, X. (2021) Mechanisms for modulating anoikis resistance in cancer and the relevance of metabolic reprogramming. *Frontiers in Oncology*, 11, 626577.

[4] Li, C., Weng, J., Yang, L., Gong, H., Liu, Z. (2024) Development of an anoikis-related gene signature and prognostic model for predicting the tumor microenvironment and response to immunotherapy in colorectal cancer. *Frontiers in Immunology*, 15, 1378305.

[5] Wu, L. W., Deshmukh, S. K., Wu, S., Xiu, J., Jang, S. J., Park, J., Moy, R. H. (2025) Multi-omic characterization of early-onset esophagogastric cancer. *Precision Oncology*, 9(1), 241.

[6] Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Lappalainen, T. (2020) Cell type-specific genetic regulation of gene expression across human tissues. *Science*, 369(6509), eaaz8528.

[7] Fishilevich, S., Barshir, R., Iny-Stein, T., Zelig, O., Guan-Golan, Y., Safran, M., Lancet, D. (2021) Disease interpretation of non-coding genomic elements with the GeneCards Suite. *Molecular Genetics and Metabolism*, 132, S123-S123.

[8] Wang, Q., Sun, N., Zhang, M. (2023) Identification and validation of anoikis-related signatures for predicting prognosis in lung adenocarcinoma with machine learning. *International Journal of General Medicine*, 16, 1833-1844.

[9] Ma, X., Lin, L., Gai, Y. (2023) A general framework of online updating variable selection for generalized linear models with streaming datasets. *Journal of Statistical Computation and Simulation*, 93(3), 325-340.

[10] Abd ElHafeez, S., D'Arrigo, G., Leonardis, D., Fusaro, M., Tripepi, G., Roumeliotis, S. (2021) Methods to analyze time-to-event data: the cox regression analysis. *Oxidative Medicine and Cellular Longevity*, 2021(1), 1302811.

[11] Beyene, K. M., Chen, D. G. (2024) Time-dependent receiver operating characteristic curve estimator for correlated right-censored time-to-event data. *Statistical Methods in Medical Research*, 33(1), 162-181.

[12] Wang, J., Luo, Z., Lin, L., Sui, X., Yu, L., Xu, C., Wu, Q. (2022) Anoikis-associated lung cancer metastasis: mechanisms and therapies. *Cancers*, 14(19), 4791.

[13] Chen, S., Sun, L., Zhou, X., Guo, Y., Song, J., Qian, S., You, Z. (2020) Mechanically and biologically skin-like elastomers for bio-integrated electronics. *Nature Communications*, 11(1), 1107.

[14] de Jong, F. C., Laajala, T. D., Hoedemaeker, R. F., Jordan, K. R., van der Made, A. C., Boevé, E. R., Zuiverloon, T. C. (2023) Non-muscle-invasive bladder cancer molecular subtypes predict differential response to intravesical Bacillus Calmette-Guérin. *Science Translational Medicine*, 15(697), eabn4118.

[15] Han, S., Wang, Y., Ma, J., Wang, Z., Wang, H. M. D., Yuan, Q. (2020) Sulforaphene inhibits esophageal cancer progression via suppressing SCD and CDH3 expression, and activating the GADD45B-MAP2K3-p38-p53 feedback loop. *Cell Death and Disease*, 11(8), 713.

[16] Su, P. F., Lin, C. C. K., Hung, J. Y., Lee, J. S. (2022) The proper use and reporting of survival analysis and cox regression. *World Neurosurgery*, 161, 303-30.

[17] Yu, M., Qian, X. X., Li, G., Cheng, Z., Lin, Z. (2022) Prognostic biomarker IL17A correlated with immune infiltrates in head and neck cancer. *World Journal of Surgical Oncology*, 20(1), 243.

[18] Qiu, L., Ding, F. (2021) Understanding single-walled carbon nanotube growth for chirality controllable synthesis. *Accounts of Materials Research*, 2(9), 828-841.

[19] Meng, L., Jiang, Y. P., Zhu, J., Li, B. (2020) MiR-188-3p/GPR26 modulation functions as a potential regulator in manipulating glioma cell properties. *Neurological Research*, 42(3), 222-227.