

Effect of Aging-Related Genes on the Prognosis of Colon Cancer

Yushu Liu

The First Affiliated Hospital of Nanchang University

Jiantao Gong

The First Affiliated Hospital of Nanchang University

Yanyi Huang

The First Affiliated Hospital of Nanchang University

Qunguang Jiang (✉ fbron.student@sina.com)


The First Affiliated Hospital of Nanchang University <https://orcid.org/0000-0001-7085-6312>

Research

Keywords: Colon cancer, Aging-related gene, Prognosis, Immune microenvironment, Nomogram

Posted Date: November 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-984927/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Colon cancer is a common malignant cancer with high incidence and poor prognosis. Cell senescence and apoptosis are important mechanisms related genes (ARGs) play an important role. This study aimed to establish a prognostic risk model based on ARGs for diagnosis and prognosis prediction of colon cancer.

Methods: We downloaded transcriptome data and clinical information of colon cancer patients from the Cancer Genome Atlas (TCGA) database and the microarray dataset (GSE39582) from the Gene Expression Omnibus (GEO) database. Univariate COX, least absolute shrinkage and selection operator (LASSO) regression algorithm and multivariate COX regression analysis were used to construct a 6-ARG prognosis model and calculate the riskScore. The prognostic signature is validated by internal validation cohort and external validation cohort (GSE39582). In addition, functional enrichment pathways and immune microenvironment of aging-related genes (ARGs) were also analyzed. We also analyzed the correlation between riskScore and clinical features and constructed a nomogram based on riskScore. We are the first to construct prognostic nomogram based on ARGs.

Results: Through univariate COX, LASSO regression algorithm and multivariate COX regression analysis, 6 prognostic ARGs (PDPK1, RAD52, GSR, IL7, BDNF and SERPINE1) were screened out and riskScore was constructed. We have verified that riskScore has good prognostic value in both internal validation cohort and external validation cohort. Pathway enrichment and immunoanalysis of ARGs provide a direction for the treatment of colon cancer patients. We also found that riskScore was closely related to the clinical characteristics of patients. Based on riskScore and related clinical features, we constructed a nomogram, which has good predictive performance.

Conclusion: The 6-ARG prognostic signature we constructed has a certain clinical predictive ability. Its riskScore is also closely related to clinical characteristics, and nomogram based on this has stronger predictive ability than a single indicator. ARGs and the nomogram we constructed may provide a promising treatment for colon cancer patients.

1. Introduction

Colon cancer is a common gastrointestinal malignancy, which is prone to occur at the junction of rectum and sigmoid colon. Although the survival rate of colon cancer has improved in recent decades, its mortality rate is the second in the world, while its incidence rate is 10%, which is the third in the world (1, 2). With the increase of age, the incidence rate of the colon cancer is increasing, and its prognosis is also very poor [3]. Most patients will still relapse after surgery, especially in patients with stage III and high-risk stage II diseases, the risk of recurrence is high (2-4).

At present, more and more evidence indicates that cell senescence and cell apoptosis are important mechanisms for inhibiting the occurrence and development of tumors (5-7). Cell senescence is the irreversible stop of cell proliferation, preventing damaged cells from continuing to proliferate, which can effectively reduce the risk of cancer (8, 9); Cell apoptosis refers to the programmed cell death controlled by genes. Aging-related genes (ARGs) play an important role in cell senescence and apoptosis. The expression of senescence genes can inhibit tumor cell proliferation and activate the body's specific immune response by promoting tumor cell senescence and apoptosis, and promote the killing and elimination of tumor cells (5, 10, 11). However, the high expression of some ARGs can also inhibit tumor cell senescence and apoptosis (12, 13). It is currently known that the abnormal activation of multiple signal pathways such as PI3K/AKT, TGF β /SMADs, RAS, p53/p21, P16/Rb, etc., is involved in the process of tumor senescence or apoptosis (14, 15).

Due to the tumor heterogeneity of colon cancer, the diagnosis, treatment and prognosis of patients are different. Therefore, targeted treatment plans should be given according to personal risk factors and genetic factors. Studying the molecular characteristics at the genomic level can provide more insights for treatment and prognosis. Moreover, the prognosis and molecular mechanisms of colon cancer in different sites are also different (16, 17), so understanding the changes in the molecular mechanism of colon cancer and finding new biomarkers is of great significance to the prognosis of colon cancer. In order to conduct a more in-depth study on the prognosis of colon cancer, we constructed the Aging-Related genes (ARGs) signature based on the Cancer Genome Atlas (TCGA) database and performed internal verification, and at the same time we performed external verification on the GSE39582 in the Gene Expression Omnibus (GEO) database. Finally, we also constructed a clinical risk prognosis model based on the ARG signature to improve the predictive performance of the model.

2. Methods And Materials

2.1 Data sources and processing

We downloaded transcriptome data and clinical information of colon cancer patients from the Cancer Genome Atlas (TCGA) database portal (<https://portal.gdc.cancer.gov>). The microarray dataset (GSE39582) was downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) containing 585 cases. The patients without overall survival (OS) time were excluded, and 339 cases from TCGA and 557 cases from GSE39582 were included. For analyzing the ARGs which affected OS in colon cancer patients, we used the sample function in R to randomly divide TCGA patients into the development cohort (n=170) and internal validation cohort (n=169) according to the ratio of 1:1. The cases from GSE39582 were used as external validation cohort. 307 human ARGs from the Human Aging Genomic Resources (<https://genomics.senescence.info/genes/>). The specific contents of ARGs are shown in Table S1.

2.2 Screening of prognostic ARGs

All ARGs were included in our study. For identifying the ARGs related to prognosis ($P < 0.05$), the 'survival' package was used in R by univariate COX (18) in the development cohort. Considering that the prognostic ARGs were too many, least absolute shrinkage and selection operator (LASSO) regression algorithm (19) with penalty term was used to delete ARGs with multicollinearity.

2.3 Construction the prognosis ARG signature and generation of RiskScore

Through multivariate COX regression analysis (20, 21) with bothway recursive elimination, 6-ARG prognosis model was established using the “glmnet” package. We performed an independent prognostic analysis of the risk score in development cohort. The riskScore was calculated as (22):

$$\text{riskScore} = \sum_{i=1}^n \text{Exp}_i \times \text{Coef}.$$

With n means the number of ARGs in the signature, Exp_i means the levels of ARG expression in the signature and Coef means the estimated regression coefficient value from the COX-PH algorithm. According to the median of riskScore, patients were divided into high-risk and low-risk groups.

2.4 The validation of prognostic ARG models

Kaplan–Meier(K-M) survival analysis was performed to validate the predictive power of the model. The receiver operating characteristic (ROC) curve (23) was also plotted for judging the effect of the model. For the 6 ARGs in the model, we compared their expression in the normal group and the cancer group. What's more, we analyzed their expression enrichment in the high and low risk group using ‘pheatmap’ package (24) in R. All analyses were performed in the development cohort, internal validation cohort and external validation cohort.

2.5 Gene Ontology analyses in ARGs

To explore the pathways associated with ARGs, they were enriched by biological processes of Gene Ontology(GO) enrichment analyses (25) through a free online platform (<http://www.bioinformatics.com.cn>) for data analysis and visualization. The most significantly enriched pathways in the Molecular Function(MF), Biological Process(BP) and Cellular Component(CC) were visualized.

2.6 Immune microenvironment analysis

From previous studies (26-31), immune-related gene set with 29 immune cell types and immune-related functions were obtained. We enriched 29 types of immune cells and their functions in each sample by single sample gene set enrichment analysis(ssGSEA) algorithm with ‘GSVA’ package in R (32). For further analysis of immune infiltration in the high and low risk groups, we also calculate the immune score, stromal score, ESTIMATE score and tumor purity using ‘estimate’ package. The differences between the two groups were compared through Mann-Whitney U test. We also visualized the result using ‘pheatmap’ package (24).

2.7 Nomogram based on riskScore and clinical factors

We compared the expression of ARGs of riskScore in the patients with different clinical characteristics, and judged the correlation between riskScore and clinical characteristics of patients. Nomogram was constructed to assess 1-, 3- and 5-year OS in the development cohort. In order to judge the clinical application effect of the model, decision curve analysis (DCA) was used and the calibration curves of the nomogram were plotted. The tests were performed in the development cohort, internal validation cohort and external validation cohort.

3.result

3.1 Establish the prognostic model based on ARGs

For screening the ARGs associated with survival in colon cancer patients, we used univariate Cox under the condition of P < 0.05 to analyze the data from the development cohort. 26 prognostic ARGs were screened out and shown in Table 1. Then, we used LASSO algorithm for further screening of ARGs (Figure 1A,B) and obtain 8 ARGs. Finally, multivariate COX regression analysis with bothway recursive elimination was used and the 6-ARG prognosis signature was constructed (Figure 1C)(Table 2). The riskScore was calculated as: PDPK1* -1.450 + RAD52* 1.478 + GSR* -0.665 + IL7* -0.708 + BDNF* 1.413 + SERPINE1* 0.430. According to the median of riskScore, patients were assigned to the high risk or low risk group.

3.2 Prognostic value of the 6-ARG prognosis signature

To further analyze the ARGs in the model, we compared their expression in the normal and cancer groups. The results are shown in Figure 2. P < 0.05 was defined as statistically significant. The results showed that there were differences in the expression of all other ARGs except RAD52. According to Naccarati A et al. (33), the TT genotype of RAD52 rs11226 with longer survival in the colon cancer patients. Considering the limited amount of our data, there may be some deviation. Patients were divided into high risk and low risk groups based on the median riskScore in the development cohort. Kaplan–Meier(K-M) survival analysis were performed and the patients in high risk group was had significantly worse survival rate than those in low risk group (Figure 3A). What's more, the area under the curve(AUC) of the riskScore for 1-year, 2-year, 3-year and 5-year OS were 0.87, 0.81 and 0.80 respectively (Figure 3B). Scatter plots were used to show the distribution of patients in the high and low risk group (Figure 3C). The distribution of patient riskScore was also shown in Figure 3D. The heatmap showed the cluster analysis results of ARGs in the model (Figure 3E). The same analysis was performed in both internal validation cohort (Figure S1) and external validation cohort (Figure S2). In the internal validation cohort, AUC of the riskScore for 1-year, 2-year, 3-year and 5-year OS were 0.65, 0.59 and 0.67 while that in the external validation cohort were 0.61, 0.58 and 0.59 respectively.

3.3 Functional analysis of ARGs

To explore the functions and pathways associated with ARGs, GO enrichment analysis was performed (Figure 4). The pathway with the highest enrichment score in BP is aging. Transcription regulator complex obtained the highest enrichment score in CC and in MF, DNA-binding transcription factor binding obtained the highest enrichment score. Mole DJ et al.'s research(34) showed that transcription regulatory complex (TRC) regulates osteopontin that is implicated in colorectal cancer dissemination. ARGs may be a key link between TRC and colorectal cancer. DNA-binding transcription factor binding plays an important role in telomeres against cell aging which was used to activate telomerase(35).

3.5 Immune microenvironment landscape

We obtained 29 immune cell types and immune-related functions and calculate immune score, stromal score, ESTIMATE score and tumor purity to explore the difference of immune infiltration between high and low risk groups by ssGSEA algorithm (Figure 5). iDCs, NK cells and Th 2 cells showed higher levels in the low risk group. We found statistical differences in stromal score between the high and low risk groups. Studies(36) have shown that stromal score is related to the survival of colon cancer patients, and the role of ARGs in this study cannot be ignored.

3.6 Expression of model ARGs in different clinical features

The clinical characteristics we included age, gender, T, N, M, stage and tumor side. The boxplot showed differences in gene expression with different clinical features (Figure 6). The gene expression of GSR was statistically different in gender, T, N, M, stage and tumor side. The gene expression of SERPINE1 was statistically different in T. The gene expression of IL7 was statistically different in N, M and stage. The gene expression of PDPK1 was statistically different in tumor side.

3.7 Construction of nomogram

We applied univariate Cox (Figure 7A) and multivariate COX regression analysis (Figure 7B) finding that riskScore consistently showed significant statistical differences. Considering the correlation between riskScore and clinical features, we constructed nomogram (Figure 7C) and judged its predictive power. In the development cohort, the AUC for 1-year, 2-year, 3-year and 5-year OS were 0.85, 0.90 and 0.83 (Figure 7D). In the internal validation cohort, the AUC for 1-year, 2-year, 3-year and 5-year OS were 0.84, 0.79 and 0.73 (Figure 7E). In the external validation cohort, the AUC for 1-year, 2-year, 3-year and 5-year OS were 0.72, 0.67 and 0.65 (Figure 7F). Compared with riskScore, nomogram based on riskScore has stronger predictive ability, which can provide some significant guiding value for clinical work. What's more, we performed calibration curves and DCA in the development cohort (Figure 8), internal validation cohort (Figure S3) and external validation cohort (Figure S4). Compared with stage, nomogram based on riskScore has higher prediction performance, which was reflected in all three cohorts. Therefore, the nomogram we have constructed do have more predictive power than a single indicator.

4. discussion

Colon cancer is one of the most common malignant tumors in the world with high mortality and poor prognosis. In 2018, there were about 1.1 million new cases, accounting for about 6.1% of the total cancer cases, and about 550,000 deaths. Accounting for about 5.8% of the total deaths, In 2020, there were 930,000 deaths, or 10% of the total number of deaths(1, 37). Therefore, finding suitable treatment decisions and improving the prognosis of colon cancer patients is particularly important. Due to the tumor heterogeneity of colon cancer, it is difficult to predict the prognosis of patients, and the prediction of prognosis by traditional factors is difficult to meet the needs(38). At present, many ARGs can be used as good prognostic markers for colon cancer. Our research used ARGs to construct a clinical risk prognostic model for colon cancer.

In this study, we collected 41 normal samples and 473 colon cancer patient samples from the TCGA database. There were 339 cancer samples with complete clinical data. 19 normal samples from the GEO database GSE39582, 566 colon cancer patients, and 557 cancer samples retained complete clinical data. Firstly, we extracted ARG based on TCGA database, and divided the samples into training set and test set according to 1:1. There were 170 cases in the training set, 169 cases in the test set, and the test set was used for internal verification. We performed Univariate Cox analysis on the training set, then performed Lasso dimensionality reduction, and then performed multi-factor cox analysis to screen for differentially expressed genes. Finally, we constructed a clinical risk prognosis model based on the differential genes, and externally verified the model with the GSE39582 chip. We also constructed a nomogram and a nomogram based on the prognostic model, and evaluated the prediction performance through 1, 3, and 5 year correction curves and DCA.

Our results show that the constructed model has good prognostic performance and is conducive to treatment decisions. The genes involved in the construction of prognostic models include PDPK1, RAD52, GSR, IL7, BDNF, and SERPINE1. These genes have been reported in previous studies to be associated with the prognosis of colon cancer or other tumors. Among the six genes, RAD52, BDNF, and SERPINE1 are dangerous genes, and PDPK1, GSR, and IL7 are beneficial genes for prognosis. PDPK1 mainly interacts with the MAPK/AKT and PI3K/AKT/mTOR signaling pathways (39, 40). The 3-phosphoinositide-dependent protein kinase, the encoded product of PDPK1, can phosphorylate proteins downstream of these signaling pathways, which is closely related to cell proliferation and survival (40, 41). In invasive breast cancer cell lines and ovarian cancer, PDPK1 is highly expressed(41), while it is low expressed in colon cancer cells. Inhibition of PDPK1 expression can promote cell senescence and reduce cell migration, (13, 42). The expression level of GSR in colon cancer is low, and its encoded product is glutathione reductase, which participates in the reduction metabolism of glutathione. According to reports, glutathione peroxidation may be related to the malignant degree of colon cancer(43). In addition, when the expression of GSR and Fas is inhibited, it can promote the metastasis of colon cancer and is conducive to the proliferation of tumor cells(44). The coding product of IL7 gene is interleukin-7, and the combination of IL7 and its receptor IL7R is essential for T cells(45). It has been reported that IL7 can inhibit the development of colon cancer, and the level of IL7 in colon cancer samples is significantly reduced(10), which is consistent with our results. IL7 may play an anti-tumor role through apoptosis pathway[48] In addition, IL7 can also enhance the anti-tumor immunity by combining other factors to enhance the function of immune cells(46). The protein encoded by the RAD52 gene is a DNA-binding protein that participates in the DNA repair process, Double-strand breaks are the main cause of tumor cell death, and the RAD52 protein plays an important role in the repair of double-strand break damage(33, 47). In addition, the RAD52 protein is also essential for DNA synthesis during

mitosis(48). The high expression of RAD52 gene is related to the poor prognosis of colon cancer patients. BDNF has been shown to be related to a variety of signaling pathways, promoting colon cancer cell metastasis through ERK, PI3K/Akt, and p38 signaling pathways(49); activating PI3K/Akt and ERK pathways can increase the resistance of colon cancer cells to chemotherapy drugs, BDNF-Akt-Bcl2 signal, BDNF/TrkB signal is also related to the decrease of colon cancer cell apoptosis (12, 50). In addition, BDNF can also promote the expression of endothelial growth factor (VEGF), and the VEGF pathway plays a key role in tumor-induced angiogenesis (2, 12). The encoded product of SERPINE1 is plasminogen activator inhibitor 1, which has the effect of inhibiting fibrinolysis. SERPINE is significantly related to the poor prognosis of head and neck cancer, glioma, gastric cancer, and colon cancer (51-53). In colon cancer, the expression of SERPINE1 is up-regulated, which is related to the aggressiveness of the tumor. SERPINE1 is also considered to be a member of the (epithelial cell-mesenchymal transition) EMT pathway(54). When the expression of SERPINE1 is inhibited, the EMT process is also inhibited, and the level of tumor cell apoptosis increases and the invasiveness of tumor cells decreases(52).

In addition, we analyzed the immune cell infiltration of colon cancer specimens, and the results showed that there were statistical differences in iDCs, Th2 cells, and NK cells between the high-risk group and the low-risk group, among which iDCs and NK cells were significantly different. The expression of these three immune cells in the low-risk group was higher than that in the high-risk group. Colorectal tumor antigen can recruit dendritic cells and promote cell maturation and cytokine release, which is conducive to the generation of effective Th1 immune response(55). The survival rate of patients with high DC infiltration is significantly better than that of patients with low DC infiltration, the more immature dendritic cells in the tumor stroma and the more mature DCs at the edge of the infiltration, the longer the overall survival of the patient (55, 56). Moreover, there is a significant positive correlation between immature DC and regulatory T cells (Treg), which may be related to the interaction between DC and Treg(56). NK cells are important natural immune cells in vivo, which are related to anti-tumor immunity, anti-viral infection, and immune regulation. NK cells can directly recognize and kill tumor cells, which is of great significance to tumor immunity. The infiltration degree of NK cells is positively correlated with the good prognosis of cancer patients(57). The cytokines in the body have a regulatory effect on the cytotoxicity and recruitment of NK cells, thereby regulating tumor immunity(58). Studies have shown that in patients with colon cancer, Th1/Th2 cytokines are imbalanced, which may be related to the immune evasion of tumor cells. This phenomenon promotes the occurrence, development and metastasis of tumors (59, 60).

This study discovered the ARGs related to the prognosis of colon cancer patients and constructed a prognostic model, providing a new method for the prognosis assessment of colon cancer patients. However, this study has some limitations. Firstly, the small number of samples may affect the accuracy and reliability of the prediction model. Therefore, it needs to be further verified by other independent large sample cohorts. In addition, the potential role and mechanism of these ARGs in the progression of colon cancer require further research in basic experiments.

Conclusion

Our results indicates that 6-ARG prognostic signature has a certain clinical predictive ability. Its riskScore is also closely related to clinical characteristics, and nomogram based on this has stronger predictive ability than a single indicator. ARGs may provide a promising treatment for colon cancer patients.

Abbreviations

ARG: Aging-Related gene; TCGA: The Cancer Genome Atlas; GEO: the Gene Expression Omnibus; OS: Overall survival; LASSO: least absolute shrinkage and selection operator; K-M: Kaplan-Meier; ROC: receiver operating characteristic; AUC: area under the curve; GO: Gene Ontology; MF: Molecular Function; BP: Biological Process; CC: Cellular Component; ssGSEA: single sample gene set enrichment analysis; DCA: decision curve analysis; TRC: transcription regulatory complex.

Declarations

Acknowledgements

The authors thank all the researchers who supported the The Cancer Genome Atlas and the Gene Expression Omnibus Research Network.

Authors' contributions

YSL, JTG, HHY and QGJ jointly designed this study. YSL and JTG collected the mRNA transcriptome data and clinical information from TCGA. HHY downloaded the microarray dataset (GSE39582) from GEO database and performed analyses on data. QGJ and HHY performed statistical analyses. YSL and JTG wrote the manuscript. HHY reviewed and revised the manuscript. All authors read and approved the finally manuscript.

Funding

This work was supported by a grant from National Natural Science Foundation of China (nfs:81560397 and 81660403).

Data Availability Statement

All analyzed data are accessible online, and the results of this article are included within the article as well as in additional files.

TCGA: <https://portal.gdc.cancer.gov>

GSE39582: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582>

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Conflict of Interest Statement

The authors confirm that there are no conflicts of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-49.
2. Riera P, Virgili AC, Salazar J, Sebio A, Tobeña M, Sullivan I, et al. Genetic variants in the VEGF pathway as prognostic factors in stages II and III colon cancer. *Pharmacogenomics J.* 2018;18(4):556-64.
3. Yeo SA, Chew MH, Koh PK, Tang CL. Young colorectal carcinoma patients do not have a poorer prognosis: a comparative review of 2,426 cases. *Tech Coloproctol.* 2013;17(6):653-61.
4. Al-Maghrabi J, Gomaa W, Buhmeida A, Qari Y, Al-Qahtani M, Al-Ahwal M. Prognostic significance of VEGFR1/Flt-1 immunoexpression in colorectal carcinoma. *Tumour Biol.* 2014;35(9):9045-51.
5. Eggert T, Wolter K, Ji J, Ma C, Yevsa T, Klotz S, et al. Distinct Functions of Senescence-Associated Immune Responses in Liver Tumor Surveillance and Tumor Progression. *Cancer Cell.* 2016;30(4):533-47.
6. Rodier F, Campisi J. Four faces of cellular senescence. *J Cell Biol.* 2011;192(4):547-56.
7. Lu H, Hao L, Yang H, Chen J, Liu J. miRNA-34a suppresses colon carcinoma proliferation and induces cell apoptosis by targeting SYT1. *Int J Clin Exp Pathol.* 2019;12(8):2887-97.
8. Loo TM, Miyata K, Tanaka Y, Takahashi A. Cellular senescence and senescence-associated secretory phenotype via the cGAS-STING signaling pathway in cancer. *Cancer Sci.* 2020;111(2):304-11.
9. Sakaki M, Ebihara Y, Okamura K, Nakabayashi K, Igarashi A, Matsumoto K, et al. Potential roles of DNA methylation in the initiation and establishment of replicative senescence revealed by array-based methylome and transcriptome analyses. *PLoS One.* 2017;12(2):e0171431.
10. Wei X, Zhang Y, Yang Z, Sha Y, Pan Y, Chen Y, et al. Analysis of the role of the interleukins in colon cancer. *Biol Res.* 2020;53(1):20.
11. Kang TW, Yevsa T, Woller N, Hoenicke L, Wuestefeld T, Dauch D, et al. Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature.* 2011;479(7374):547-51.
12. Yang X, Martin TA, Jiang WG. Biological influence of brain-derived neurotrophic factor (BDNF) on colon cancer cells. *Exp Ther Med.* 2013;6(6):1475-81.
13. Zhu B, Ferry CH, Blazanin N, Bility MT, Khozoe C, Kang BH, et al. PPAR β/δ promotes HRAS-induced senescence and tumor suppression by potentiating p-ERK and repressing p-AKT signaling. *Oncogene.* 2014;33(46):5348-59.
14. Zhang H. Molecular signaling and genetic pathways of senescence: Its role in tumorigenesis and aging. *J Cell Physiol.* 2007;210(3):567-74.
15. Rayess H, Wang MB, Srivatsan ES. Cellular senescence and tumor suppressor gene p16. *Int J Cancer.* 2012;130(8):1715-25.
16. Liu J, Dong C, Jiang G, Lu X, Liu Y, Wu H. Transcription factor expression as a predictor of colon cancer prognosis: a machine learning practice. *BMC Med Genomics.* 2020;13(Suppl 9):135.
17. Liang L, Zeng JH, Qin XG, Chen JQ, Luo DZ, Chen G. Distinguishable Prognostic Signatures of Left- and Right-Sided Colon Cancer: a Study Based on Sequencing Data. *Cell Physiol Biochem.* 2018;48(2):475-90.
18. Tibshirani RJ. Univariate shrinkage in the cox model for high dimensional data. *Stat Appl Genet Mol Biol.* 2009;8(1):Article21.
19. Alhamzawi R, Ali HTM. The Bayesian adaptive lasso regression. *Math Biosci.* 2018;303:75-82.
20. Scosyrev E, Glimm E. Power analysis for multivariable Cox regression models. *Stat Med.* 2019;38(1):88-99.

21. Grant SW, Hickey GL, Head SJ. Statistical primer: multivariable regression considerations and pitfalls. *Eur J Cardiothorac Surg.* 2019;55(2):179-85.
22. Li R, Liu X, Zhou XJ, Chen X, Li JP, Yin YH, et al. Identification of a Prognostic Model Based on Immune-Related Genes of Lung Squamous Cell Carcinoma. *Front Oncol.* 2020;10:1588.
23. Blangero Y, Rabilloud M, Laurent-Puig P, Le Malicot K, Lepage C, Ecochard R, et al. The area between ROC curves, a non-parametric method to evaluate a biomarker for patient treatment selection. *Biom J.* 2020;62(6):1476-93.
24. Galili T, O'Callaghan A, Sidi J, Sievert C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics.* 2018;34(9):1600-2.
25. Chen L, Zhang YH, Wang S, Zhang Y, Huang T, Cai YD. Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PLoS One.* 2017;12(9):e0184129.
26. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160(1-2):48-61.
27. Bedognetti D, Hendrickx W, Marincola FM, Miller LD. Prognostic and predictive immune gene signatures in breast cancer. *Curr Opin Oncol.* 2015;27(6):433-44.
28. Ren Y, Cherukuri Y, Wickland DP, Sarangi V, Tian S, Carter JM, et al. HLA class-I and class-II restricted neoantigen loads predict overall survival in breast cancer. *Oncoimmunology.* 2020;9(1):1744947.
29. Liu Z, Li M, Jiang Z, Wang X. A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer. *Transl Oncol.* 2018;11(2):311-29.
30. Maibach F, Sadozai H, Seyed Jafari SM, Hunger RE, Schenk M. Tumor-Infiltrating Lymphocytes and Their Prognostic Value in Cutaneous Melanoma. *Front Immunol.* 2020;11:2105.
31. Aran D, Lasry A, Zinger A, Biton M, Pikarsky E, Hellman A, et al. Widespread parainflammation in human cancer. *Genome Biol.* 2016;17(1):145.
32. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
33. Naccarati A, Rosa F, Vymetalkova V, Barone E, Jiraskova K, Di Gaetano C, et al. Double-strand break repair and colorectal cancer: gene variants within 3' UTRs and microRNAs binding as modulators of cancer risk and clinical outcome. *Oncotarget.* 2016;7(17):23156-69.
34. Mole DJ, O'Neill C, Hamilton P, Olabi B, Robinson V, Williams L, et al. Expression of osteopontin coregulators in primary colorectal cancer and associated liver metastases. *Br J Cancer.* 2011;104(6):1007-12.
35. Avin BA, Umbricht CB, Zeiger MA. Human telomerase reverse transcriptase regulation by DNA methylation, transcription factor binding and alternative splicing (Review). *Int J Oncol.* 2016;49(6):2199-205.
36. Chen Y, Wang W, Jiang B, Yao L, Xia F, Li X. Integrating Tumor Stroma Biomarkers With Clinical Indicators for Colon Cancer Survival Stratification. *Front Med (Lausanne).* 2020;7:584747.
37. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424.
38. Chen J, He Q, Wu P, Fu J, Xiao Y, Chen K, et al. ZMYND8 expression combined with pN and pM classification as a novel prognostic prediction model for colorectal cancer: Based on TCGA and GEO database analysis. *Cancer Biomark.* 2020;28(2):201-11.
39. Kucherlapati MH, Esfahani S, Habibollahi P, Wang J, Still ER, Bronson RT, et al. Genotype directed therapy in murine mismatch repair deficient tumors. *PLoS One.* 2013;8(7):e68817.
40. Lu Z, Cox-Hipkin MA, Windsor WT, Boyapati A. 3-phosphoinositide-dependent protein kinase-1 regulates proliferation and survival of cancer cells with an activated mitogen-activated protein kinase pathway. *Mol Cancer Res.* 2010;8(3):421-32.
41. Alberobello AT, D'Esposito V, Marasco D, Doti N, Ruvo M, Bianco R, et al. Selective disruption of insulin-like growth factor-1 (IGF-1) signaling via phosphoinositide-dependent kinase-1 prevents the protective effect of IGF-1 on human cancer cell death. *J Biol Chem.* 2010;285(9):6563-72.
42. Ericson K, Gan C, Cheong I, Rago C, Samuels Y, Velculescu VE, et al. Genetic inactivation of AKT1, AKT2, and PDPK1 in human colorectal cancer cells clarifies their roles in tumor growth regulation. *Proc Natl Acad Sci U S A.* 2010;107(6):2598-603.
43. Cui Y, Han B, Zhang H, Liu H, Zhang F, Niu R. Identification of Metabolic-Associated Genes for the Prediction of Colon and Rectal Adenocarcinoma. *Oncotargets Ther.* 2021;14:2259-77.
44. Gu L, Liu Y, Jiang C, Sun L, Zhou H. Identification and clinical validation of metastasis-associated biomarkers based on large-scale samples in colon-adenocarcinoma. *Pharmacol Res.* 2020;160:105087.

45. Czajka-Francuz P, Francuz T, Cisoń-Jurek S, Czajka A, Fajkis M, Szymczak B, et al. Serum cytokine profile as a potential prognostic tool in colorectal cancer patients - one center study. *Rep Pract Oncol Radiother.* 2020;25(6):867-75.
46. Gou HF, Huang J, Shi HS, Chen XC, Wang YS. Chemo-immunotherapy with oxaliplatin and interleukin-7 inhibits colon cancer metastasis in mice. *PLoS One.* 2014;9(1):e85789.
47. Ho V, Chung L, Singh A, Lea V, Abubakar A, Lim SH, et al. Aberrant Expression of RAD52, Its Prognostic Impact in Rectal Cancer and Association with Poor Survival of Patients. *Int J Mol Sci.* 2020;21(5).
48. Wang H, Li S, Oaks J, Ren J, Li L, Wu X. The concerted roles of FANCM and Rad52 in the protection of common fragile sites. *Nat Commun.* 2018;9(1):2791.
49. Huang SM, Lin C, Lin HY, Chiu CM, Fang CW, Liao KF, et al. Brain-derived neurotrophic factor regulates cell motility in human colon cancer. *Endocr Relat Cancer.* 2015;22(3):455-64.
50. de Farias CB, Heinen TE, dos Santos RP, Abujamra AL, Schwartzmann G, Roesler R. BDNF/TrkB signaling protects HT-29 human colon cancer cells from EGFR inhibition. *Biochem Biophys Res Commun.* 2012;425(2):328-32.
51. Liang Y, Su Q, Wu X. Identification and Validation of a Novel Six-Gene Prognostic Signature of Stem Cell Characteristic in Colon Cancer. *Front Oncol.* 2020;10:571655.
52. Hu B, Chen Z, Wang X, Chen F, Song Z, Cao C. MicroRNA-148a-3p Directly Targets SERPINE1 to Suppress EMT-Mediated Colon Adenocarcinoma Progression. *Cancer Manag Res.* 2021;13:6349-62.
53. Lu YC, Shi JQ, Zhang ZX, Zhou JY, Zhou HK, Feng YC, et al. Transcriptome Based System Biology Exploration Reveals Homogeneous Tumorigenicity of Alimentary Tract Malignancy. *Front Oncol.* 2020;10:580276.
54. Wang WJ, Li HT, Yu JP, Han XP, Xu ZP, Li YM, et al. A Competing Endogenous RNA Network Reveals Novel Potential lncRNA, miRNA, and mRNA Biomarkers in the Prognosis of Human Colon Adenocarcinoma. *J Surg Res.* 2019;235:22-33.
55. Gulubova MV, Ananiev JR, Vlaykova TI, Yovchev Y, Tsoneva V, Manolova IM. Role of dendritic cells in progression and clinical outcome of colon cancer. *Int J Colorectal Dis.* 2012;27(2):159-69.
56. Nagorsen D, Voigt S, Berg E, Stein H, Thiel E, Loddenkemper C. Tumor-infiltrating macrophages and dendritic cells in human colorectal cancer: relation to local regulatory T cells, systemic T-cell response against tumor-associated antigens and survival. *J Transl Med.* 2007;5:62.
57. Bigley AB, Simpson RJ. NK cells and exercise: implications for cancer immunotherapy and survivorship. *Discov Med.* 2015;19(107):433-45.
58. Bhat R, Rommelaere J. NK-cell-dependent killing of colon carcinoma cells is mediated by natural cytotoxicity receptors (NCRs) and stimulated by parvovirus infection of target cells. *BMC Cancer.* 2013;13:367.
59. Contasta I, Berghella AM, Pellegrini P, Adorno D. Passage from normal mucosa to adenoma and colon cancer: alteration of normal sCD30 mechanisms regulating TH1/TH2 cell functions. *Cancer Biother Radiopharm.* 2003;18(4):549-57.
60. Hou N, Zhang X, Zhao L, Zhao X, Li Z, Song T, et al. A novel chronic stress-induced shift in the Th1 to Th2 response promotes colon cancer growth. *Biochem Biophys Res Commun.* 2013;439(4):471-6.

Tables

Table 1. The prognostic ARGs by univariable Cox analysis in the development cohort

ARGs	HR	95%CI		P value
		Low	High	
GHRH	3.038	1.263	7.309	0.013
POU1F1	30.647	2.256	416.385	0.010
MYC	0.595	0.396	0.894	0.012
EPOR	2.047	1.090	3.844	0.026
BRCA1	0.512	0.273	0.960	0.037
PDPK1	0.465	0.233	0.930	0.030
CEBPA	0.682	0.480	0.969	0.033
SSTR3	7.857	2.068	29.846	0.002
CACNA1A	9.817	2.451	39.314	0.001
LRP2	120231.551	106.506	135725447.880	0.001
RAD52	2.630	1.195	5.785	0.016
GSR	0.531	0.331	0.851	0.009
HSPA1A	1.472	1.100	1.970	0.009
HSPA1B	1.512	1.045	2.187	0.028
IL7	0.475	0.279	0.809	0.006
MAPK9	0.287	0.089	0.926	0.037
BDNF	7.544	2.820	20.185	<0.001
DBN1	1.428	1.018	2.003	0.039
NOG	5.603	1.556	20.182	0.008
TP63	5.772	1.829	18.211	0.003
PAPPA	4.869	1.066	22.234	0.041
UCP1	88.861	2.196	3595.741	0.017
FGF23	3.107	1.526	6.324	0.002
CNR1	4.298	1.455	12.694	0.008
SERPINE1	1.522	1.194	1.941	<0.001
TRAP1	0.561	0.333	0.946	0.030

Abbreviations: ARG: aging-related Gene; CI: confidence interval; HR: hazard ratio.

Table 2. Multivariate Cox regression modeling in the development cohort

ARGs	coef	HR	95%CI		P value
			Low	High	
PDPK1	-1.450	0.235	0.102	0.540	<0.001
RAD52	1.478	4.384	1.723	11.154	0.002
GSR	-0.665	0.514	0.279	0.949	0.033
IL7	-0.708	0.493	0.288	0.844	0.010
BDNF	1.413	4.109	1.023	16.500	0.046
SERPINE1	0.430	1.537	1.138	2.075	0.005

Abbreviations: ARG: aging-related gene; CI: confidence interval; HR: hazard ratio.

Figures

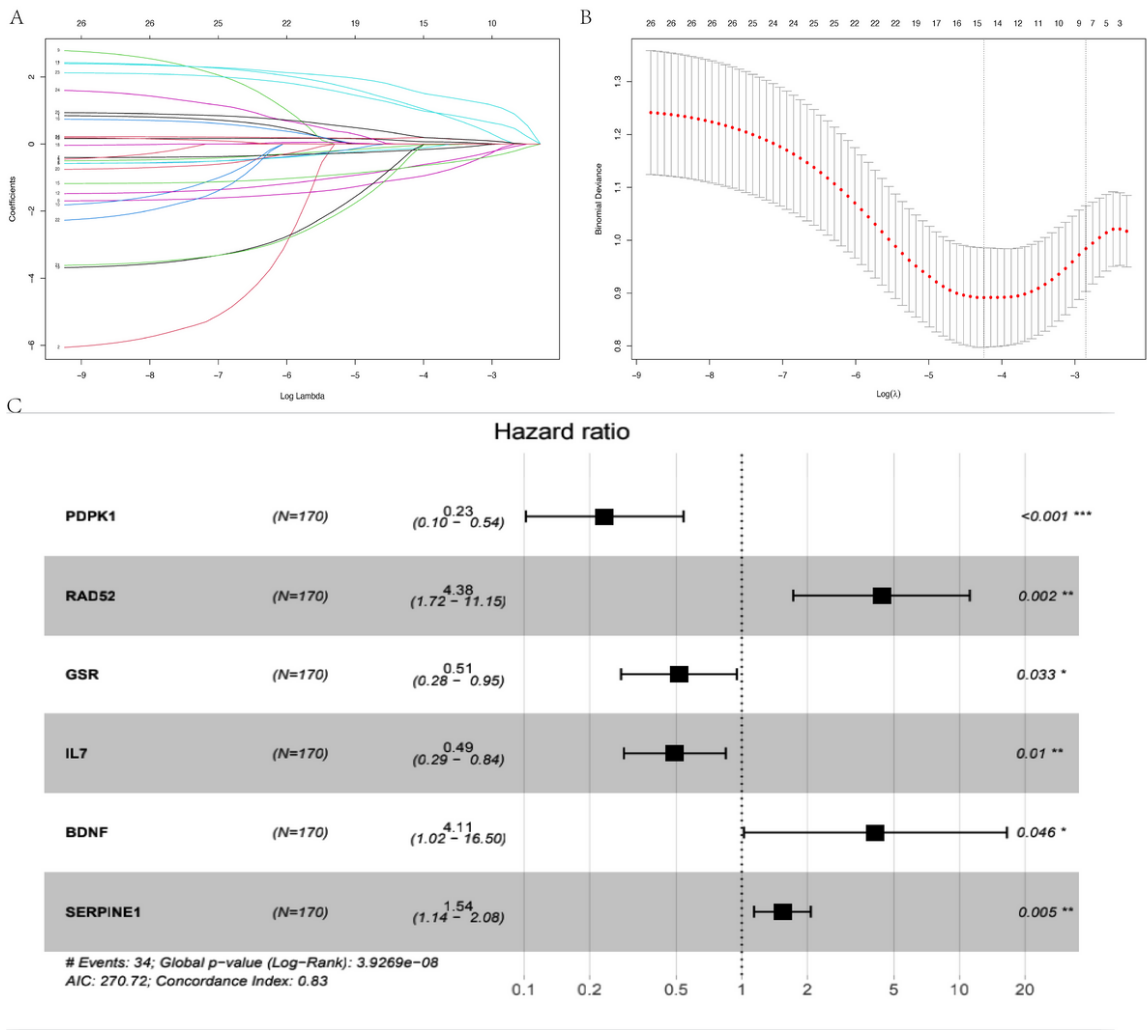


Figure 1

Screening for prognostic associated ARGs. (A) Selection of the optimal parameter (lambda) in the least absolute shrinkage and selection operator (LASSO) model and drew the dotted vertical lines at the optimal values using the minimum criteria. (B) LASSO coefficient profiles of the 8 prognosis-associated ARGs with non-zero coefficients determined by the optimal lambda. (C) Six ARGs (PDPK1,RAD52,GSR,IL7,BDNF and SERPINE1) were screened out through multivariate Cox regression analysis to construct riskScore. *P < 0.05 and **P < 0.01.

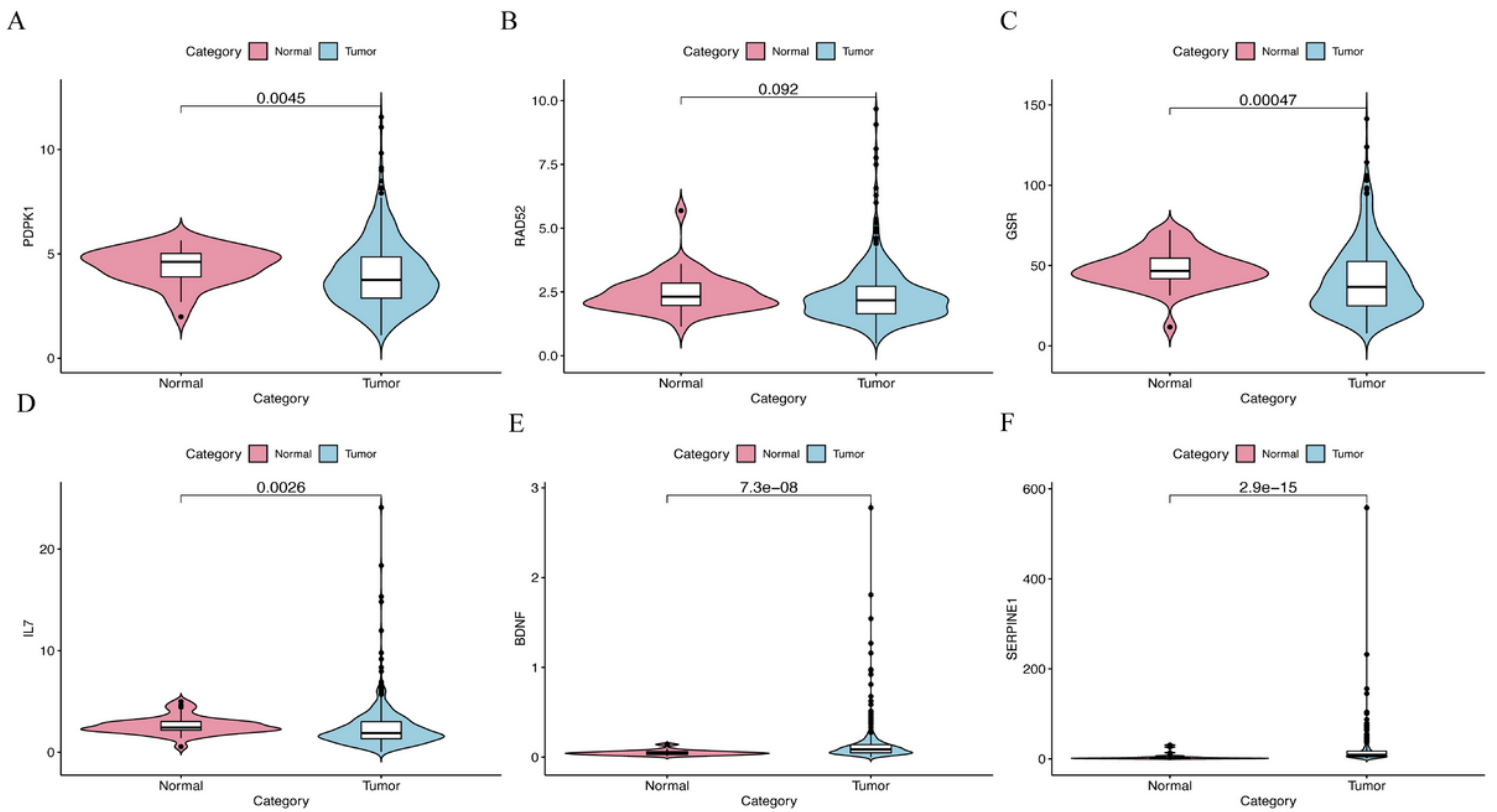


Figure 2
 The model ARGs' expression in the normal group and in the cancer group. (A)The expression of PDPK1 was statistically different between normal group and cancer group($P=0.005$). (B)The expression of RAD52 was not statistically different between normal group and cancer group($P=0.092$). (C)The expression of GSR was statistically different between normal group and cancer group($P<0.001$). (D)The expression of IL7 was statistically different between normal group and cancer group($P=0.003$). (E)The expression of BDNF was statistically different between normal group and cancer group($P<0.001$). (F)The expression of SERPINE1 was statistically different between normal group and cancer group($P<0.001$).

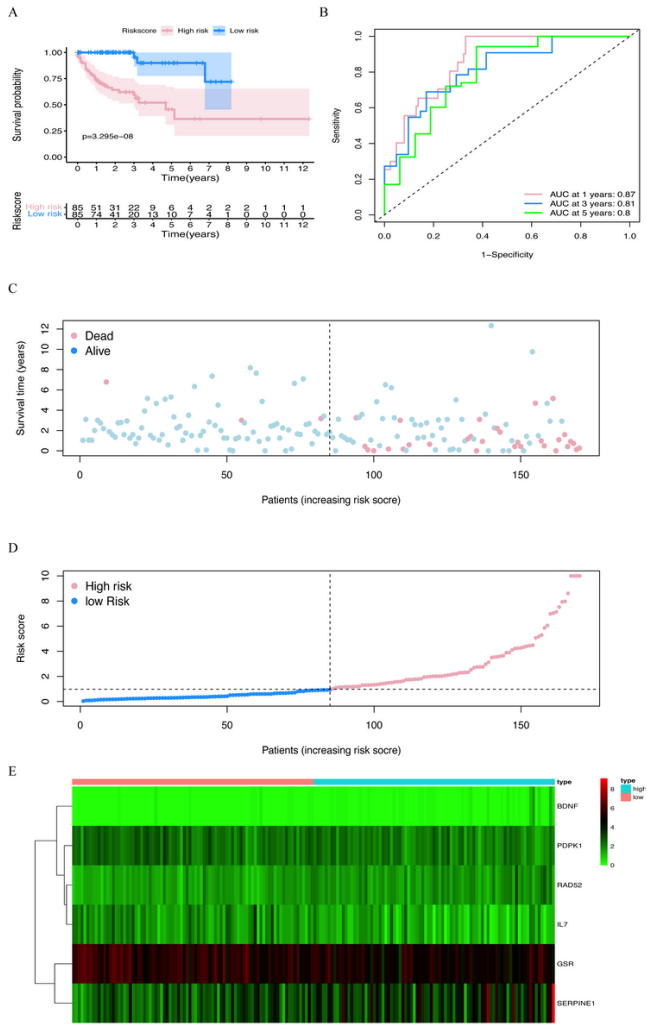


Figure 3

The prognostic value of riskScore in the development cohort. (A) Kaplan–Meier(K-M) survival analysis in high and low risk group($P<0.001$). (B) Time-dependent ROC curves of riskScore for 1-year,3-year and 5-year overall survival(OS) prediction. (C) Scatter plot showed distribution of patients in the high and low risk group. (D) Distribution of patients in the high and low risk group. (E) The heatmap showed the cluster analysis results of ARGs in the model.

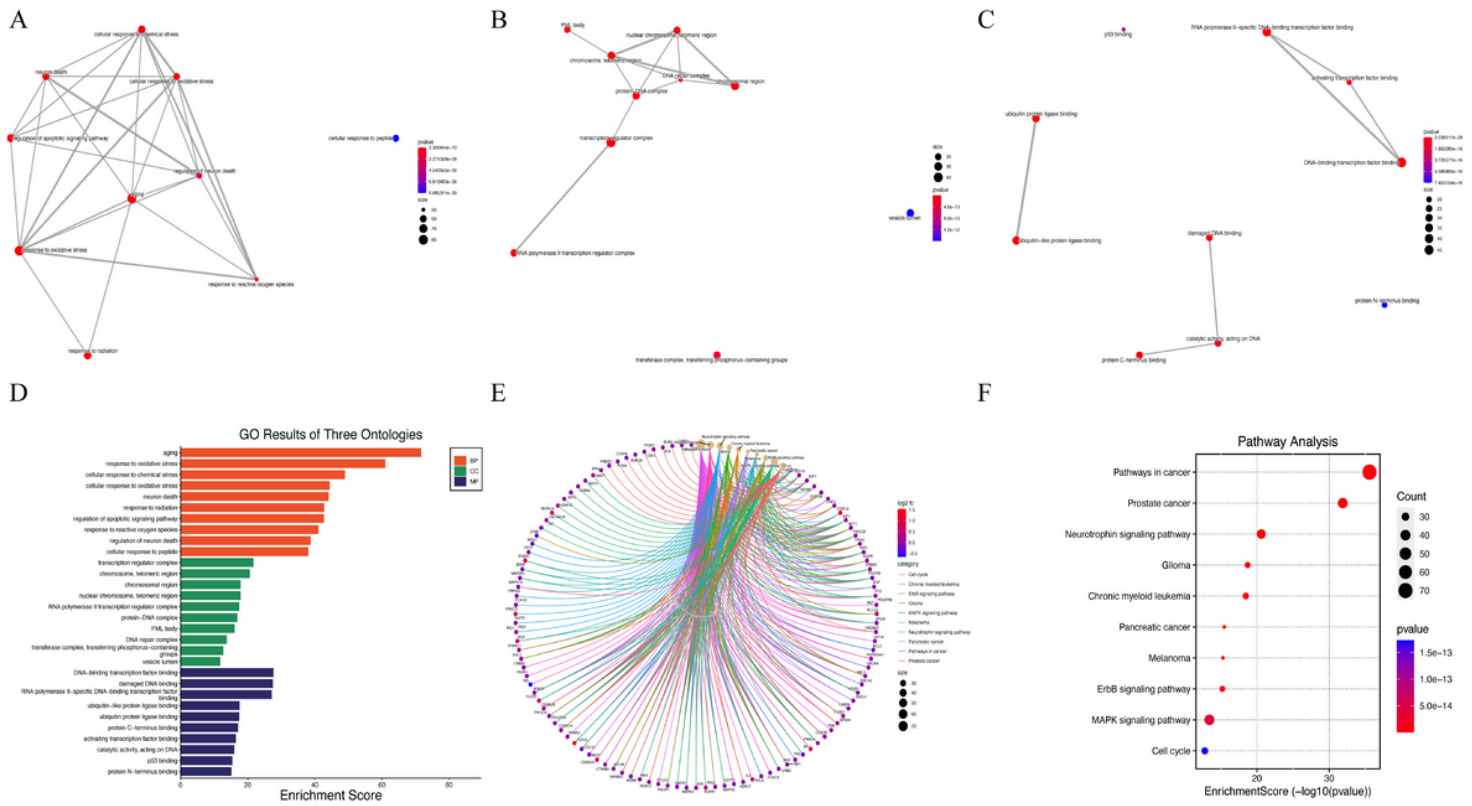


Figure 4
 Gene Ontology(GO) enrichment analyses Molecular Function(MF),Biological Process(BP) and Cellular Component(CC) (A)Demonstrates overlapping relationships between enrichment pathways in Biological Process(BP). (B)Demonstrates overlapping relationships between enrichment pathways in Cellular Component(CC). (C)Demonstrates overlapping relationships between enrichment pathways in Molecular Function(MF). (D)GO results of BP,CC and MF. (E)Enriched genes. (F)GO pathway analysis.

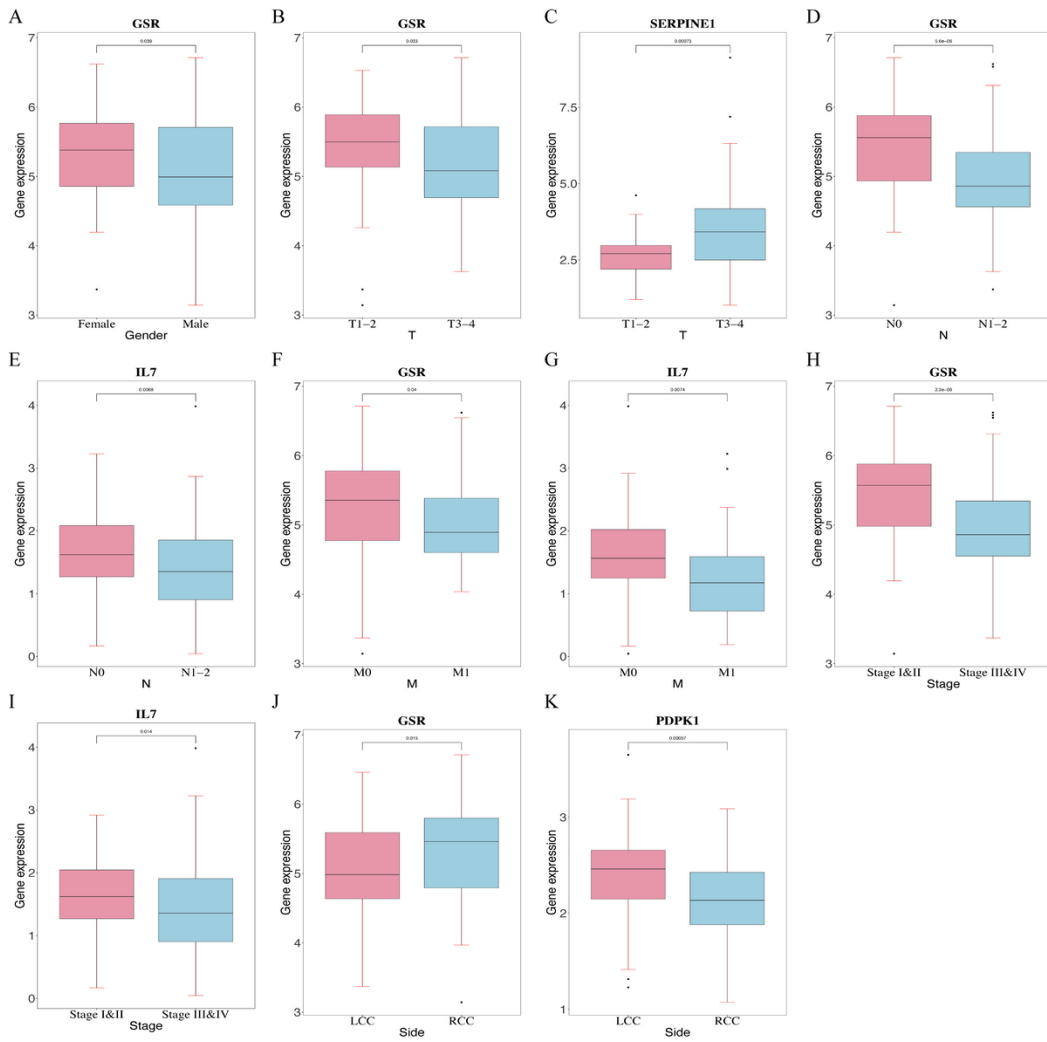


Figure 6

The differences in model ARGs expression with different clinical features. The gene expression of GSR was statistically different in gender(A), T(B), N(D), M(F), stage(H), and tumor side(J). The gene expression of SERPINE1 was statistically different in T(C). The gene expression of IL7 was statistically different in N(E), M(G) and stage(I). The gene expression of PDPK1 was statistically different in tumor side(K).

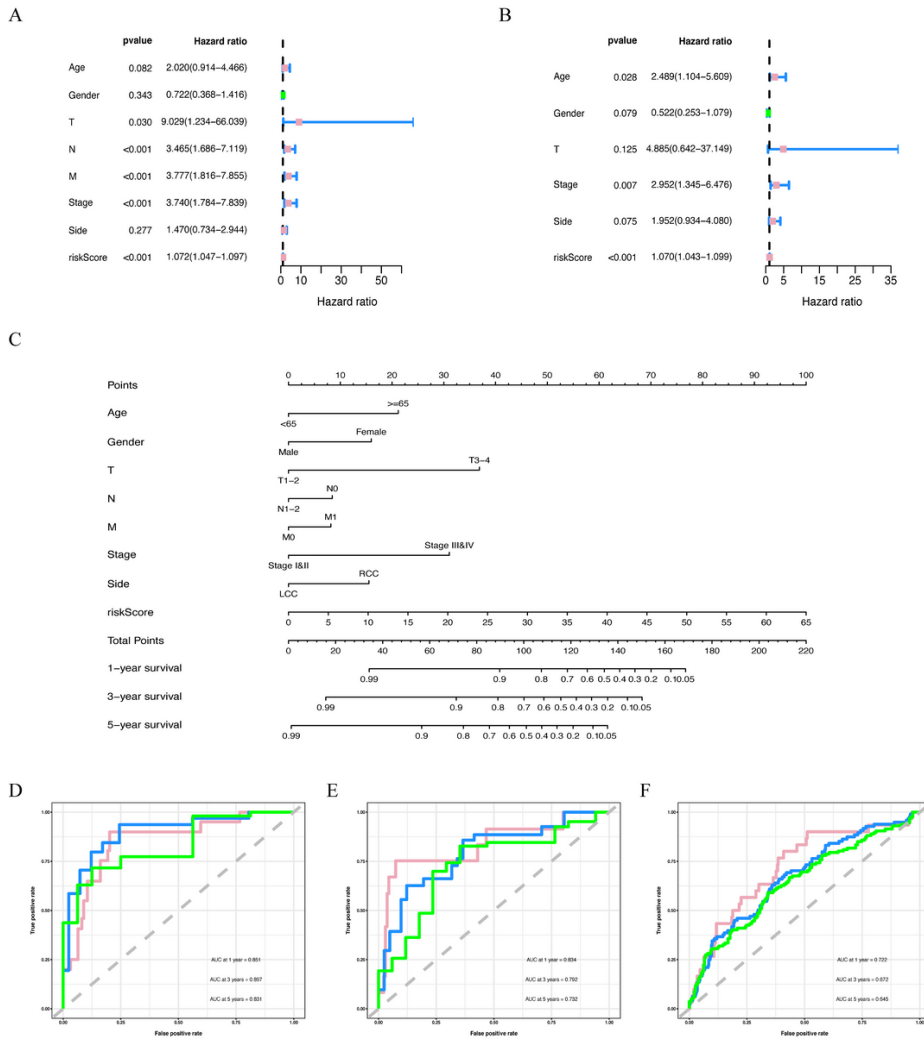


Figure 7
 Construction of nomogram based on riskScore. (A)Univariate Cox regression analysis of riskScore and clinical features in the development cohort. (B)Multivariate COX regression analysis of riskScore and clinical features in the development cohort. (C)Nomogram based on riskScore and clinical features in the development cohort. (D)Time-dependent ROC curves of nomogram for 1-year,3-year and 5-year overall survival(OS) prediction in the development cohort. (E)Time-dependent ROC curves of nomogram for 1-year,3-year and 5-year overall survival(OS) prediction in the internal validation cohort. (F)Time-dependent ROC curves of nomogram for 1-year,3-year and 5-year overall survival(OS) prediction in the external validation cohort.

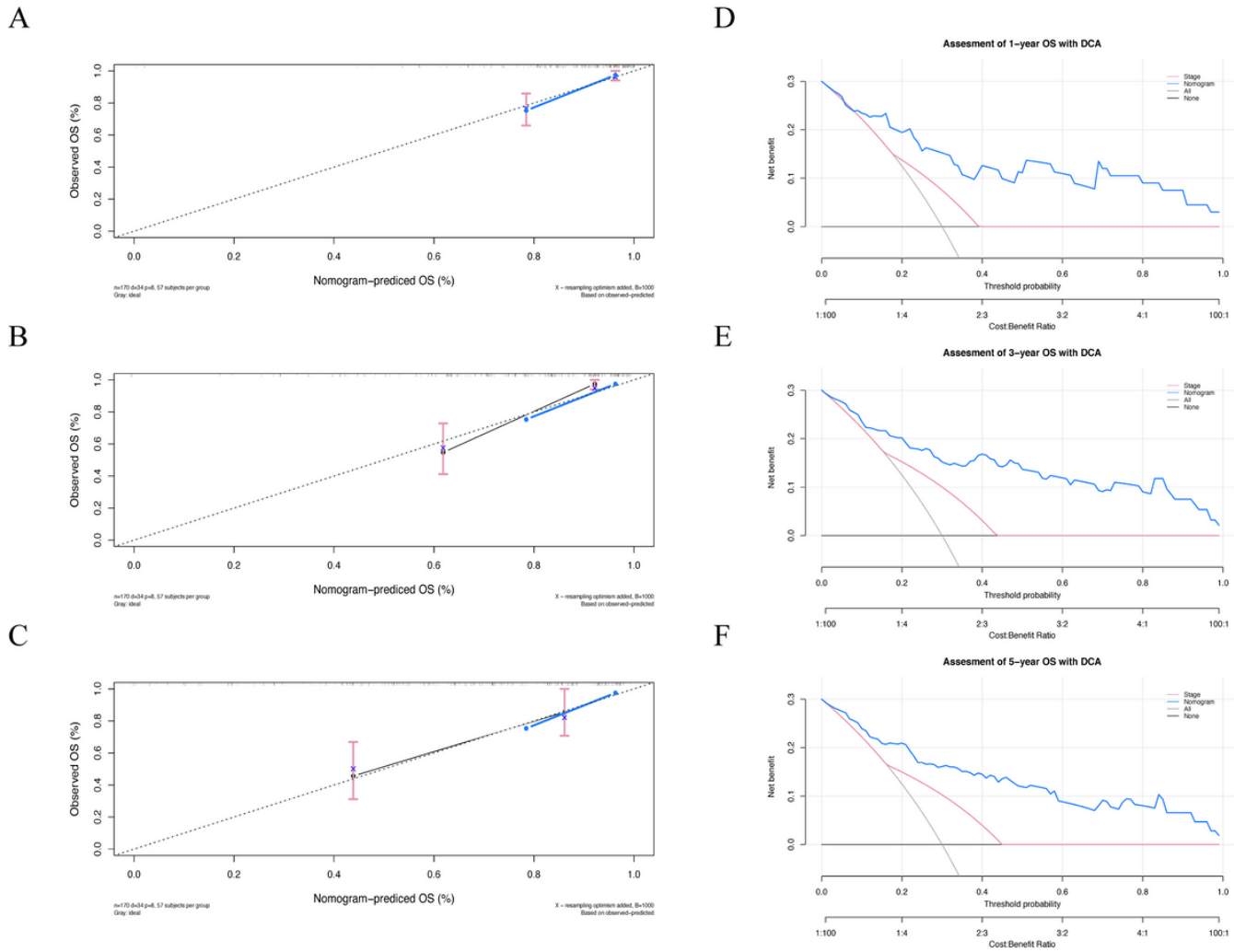


Figure 8
 Model discrimination and performance in the development cohort. (A) Calibration plot for 1-year overall survival(OS) rate. (B) Calibration plot for 3-year overall survival(OS) rate. (C) Calibration plot for 5-year overall survival(OS) rate. (D) Decision curve analysis(DCA) of the nomogram for 1-year overall survival(OS) rate. (E) Decision curve analysis(DCA) of the nomogram for 3-year overall survival(OS) rate. (F) Decision curve analysis(DCA) of the nomogram for 5-year overall survival(OS) rate.

Supplementary Files

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

- [TableS1.docx](#)
- [renameddec47.tif](#)
- [renamed42bba.tif](#)
- [renameded2ce.tif](#)
- [renamed39054.tif](#)