Establishment of an immune-related gene pair model to predict colon adenocarcinoma prognosis

Jihang Luo
Zunyi Medical University https://orcid.org/0000-0001-5351-2622
Puyu Liu
Zunyi Medical University
Leibo Wang
Zunyi Medical University
Yi Huang
Zunyi Medical University
Yuanyan Wang
Zunyi Medical University
Wenjing Geng
Zunyi Medical University
Duo Chen
Zunyi Medical University
Yuju Bai (byj6618@163.com)
https://orcid.org/0000-0003-2897-9997
Ze Yang
Zunyi Medical University

Research article

Keywords: colon adenocarcinoma, immune-related gene pairs, prognosis, TCGA, GEO

Posted Date: October 27th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-35654/v3

Version of Record: A version of this preprint was published on November 9th, 2020. See the published version at https://doi.org/10.1186/s12885-020-07532-7.

Abstract

Background. Colon cancer is the most common type of gastrointestinal cancer and has high morbidity and mortality. Colon adenocarcinoma (COAD) is the main pathological type of colon cancer, and much evidence has supported the correlation between the prognosis of COAD and the immune system. The current study aimed to develop a robust prognostic immune-related gene pair (IRGP) model to estimate the overall survival of patients with COAD.

Methods. The gene expression profiles and clinical information of patients with colon adenocarcinoma were obtained from the TCGA and GEO databases and were divided into training and validation cohorts. Immune genes were selected that showed a significant association with prognosis.

Results. Among 1,647 immune genes, a model with 17 IRGPs was built that was significantly associated with OS in the training cohort. In the training and validation datasets, the IRGP model divided patients into the high-risk group and low-risk group, and the prognosis of the high-risk group was significantly worse (P 0.001). Univariate and multivariate Cox proportional hazard analyses confirmed the feasibility of this model. Functional analysis confirmed that multiple tumor progression and stem cell growth-related pathways were upregulated in the high-risk groups. Regulatory T cells and macrophages M0 were significantly highly expressed in the high-risk group. **Conclusion.** We successfully constructed an IRGP model that can predict the prognosis of COAD, providing new insights into the treatment strategy of COAD.

Background

According to the latest GLOBOCAN[1] report, colorectal cancer (CRC) is the third most commonly diagnosed cancer worldwide (10.2%) and has the second-highest mortality rate (9.2%). Approximately 145,600 new colorectal cancer cases occur each year in the United States, among which 101,420 cases are colon cancer, and the remainder is rectal cancer[2]. In recent years, colon cancer mortality has continued to rise in many countries with limited resources and health infrastructure, particularly in South America and Eastern Europe[3]. Colon adenocarcinoma (COAD) is the primary pathological type of colon cancer. Surgery combined with postoperative chemotherapy is currently the main treatment for COAD. However, the survival of COAD has improved due to the continuous advancement of surgical technology. However, postoperative recurrence and chemotherapy resistance remain two major obstacles to the long-term survival of patients[4-6].

With the development of high-throughput omics, various omics techniques, such as whole-genome sequencing, epigenomics, and proteomics, have been applied to study COAD[7-10]. Increasing evidence has shown that COAD is not a consistent disease type but a molecularly heterogeneous disease comprising a series of genetic changes[11]. Tumor heterogeneity can alter the tumor growth rate, invasive ability, sensitivity to drugs, prognosis and other aspects, making it one of the main obstacles affecting tumor treatment[12, 13]. Therefore, dividing patients with COAD into different risk groups based on gene

expression profiles helps to predict the risk of tumor progression or metastasis and recurrence and is a necessary prerequisite for proper individualized treatment[14-16].

There is increasing evidence that the immune system plays an important role in the occurrence and development of cancer[17-19]. For example, Salem M[20] found that disrupting the cell surface receptor glycoprotein-A repetitions predominant (GARP) on activated regulatory T (Treg) cells reduces immune tolerance and the development of colon cancer. In recent years, a method based on the relative ranking of gene expression levels was proposed to eliminate the shortcomings of data standardization and scaling in gene expression data processing, achieving reliable results in various studies[21, 22]. The present study selected immune genes that are significantly associated with the prognosis of COAD. Next, we integrated these genes to construct an immune-related gene pair (IRGP) risk model and verified its feasibility as a prognostic marker for COAD.

Methods

Sources of colon adenocarcinoma patients. The data analyzed in this study were all obtained from public databases. The training cohort datasets were downloaded from TCGA (https://tcga-data.nci.nih.gov/tcga)[23], and the validation datasets were obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/). The training cohort datasets included clinical datasets (n=452), transcriptome datasets (n=449), and verification datasets from GSE39582 (n=585)[24] and GSE17538 (n=244)[25].

Data processing. The human General Transfer Format (hunman.gtf) from Ensemble (https://www.ensembl.org/index.html)[26] was downloaded, and the TCGA data were annotated using Perl language[27]. The chip data file (GSE39582 and GSE17538) was preprocessed using Perl language through the annotation file of the GPL570 platform. Using the above operations, all the gene probe IDs were converted to corresponding gene symbols. To analyze the correlation between the IRGP signature and prognosis in COAD, only patient data containing complete overall survival (OS) were selected.

Establishment of the prognostic immune-related gene pair (IRGP) model. We downloaded a list of immune-related genes (IRGs) from IMMPORT (https://www.immport.org/)[28], a website with open access to immunoassay data for translation and clinical research. Next, the R language[29] limma package (version 3.42.2) was used to control the list to screen out the IRGs in the downloaded TCGA transcriptome data. To further select valuable IRGPs, we measured and stored IRGs with a relatively high variation on all the platforms in this study (as determined by the median absolute deviation (MAD) >0.5) [30]. The expression levels of IRGs in each sample in the transcriptome, GSE39582 and GSE17538 were compared in pairs to form each IRGP according to a previously validated method[22]. Specifically, in the pairwise comparison of each sample, if the expression level of the first gene is greater than that of the second gene, the output is 1; otherwise, it is 0. Samples with a ratio of 0 and 1 less than 20% were deleted to retain gene pairs that may be related to survival. These IRGPs were merged with the survival time of the clinical data downloaded by the corresponding platform to evaluate the correlation between each

IRGP in the training dataset and overall survival rate of the patient. Based on previous reports[31, 32], we used the R language survival software package (version 3.1-11) to perform univariate Cox regression analysis and $P<0.001$ to screen the effective IRGPs in the TCGA data. From these IRGPs, we used R language for Lasso Cox proportional hazards regression (glmnet software package, version 3.0-2) to construct the risk score, and the final prognostic model was defined using 17 gene pairs. Finally, in the training cohort, we set the overall survival to 5 years and constructed the time-dependent receiver operating characteristic (ROC) curve (survivalROC, version 1.0.3) to determine the best cutoff value for the risk score and divide patients into low-risk and high-risk groups accordingly.

Further validation of the model. Using the R package survival and survminer (version 0.4.6), Kaplan–Meier plots were applied to establish survival curves for the high-risk and low-risk groups in the training and verification cohorts. The differences in the survival curves were analyzed using the log-rank test. Cox proportional hazards analysis was used for univariate and multivariate analyses to assess the effect of the risk score and other clinical factors.

Gene expression profiles (GEPs) of immune cell infiltration in tumors. We used CIBERSORT[33] to infer the relative abundance of tumor-infiltrating immune cells in different risk groups. CIBERSORT estimated the putative proportion of infiltrating immune cells using a reference set with 22 sorted immune cell subtypes for each sample in the training cohort and validation cohorts. Monte Carlo sampling was used in CIBERSORT to calculate the *P*-value of the deconvolution of each sample to provide the estimated confidence. The permutation is set to greater than 100, and the corresponding *P*-value is generated.

Gene set enrichment analysis (GSEA). The chemical and genetic perturbation analysis-related documents involved in the study were downloaded from the Molecular Signature Database (MSigDB C2, version 7.1) (https://www.gsea-msigdb.org/gsea/datasets.jsp). GSEA[34] was performed using the R package fgsea (version 1.12.0) with default parameters. A log 2-fold change was made between GEPs in the high-risk vs low-risk groups. The difference in the gene sets between the high- and low-risk groups was compared. Differences with an FDR-adjusted $P<0.05$ were defined as significant.

Statistical Analysis. For all the above tests, a *P*-value less than 0.05 denoted the presence of a statistically significant difference. Statistical significance was indicated as follows: $*P<0.05$, $**P<0.01$, $***P<0.001$.

Results

Construction of the Prognostic IRGP model. The TCGA transcriptome data were used as a training cohort. From the list of immune-related genes (IRGs) obtained by IMMPORT, the genes in the transcriptome data were searched in turn, and 1,647 IRGs were identified. To ensure relatively high variation in the genes of the two platforms, we retained 325 IRGs with a median absolute deviation (MAD)>0.5. In total, 40,375 pairs were deleted with a ratio of 0 and 1 less than 20%. Next, 12,275 immune-related gene pairs (IRGPs) were built based on these 325 IRGs. After univariate Cox regression analysis of these IRGPs in the training group, 28 potential prognostic IRGPs remained. Using Lasso Cox proportional hazards regression

to define the model on the training set, 17 IRGPs were retained to form the final prognostic risk model. These IRGPs comprised 26 unique IRGs, most of which are antibiotics, cytokine receptors and cytokine-related molecules (Table 1). Next, the risk score for each patient in the TCGA dataset was calculated based on the model. Finally, we used a time-dependent ROC curve analysis to classify patients into high- or low-immune risk groups. The optimal cutoff value for the risk score was set to -0.576 (Figure 1). This value successfully stratified the patients in the training cohort into high- and low-risk groups. In other words, the overall survival (OS) of the low-risk group was significantly higher than that of the high-risk group (Figure 2A). We further performed univariate and multivariate Cox proportional hazards analyses to test whether the IRGP model predicted survival independently of other prognostic factors in the TCGA cohort. Among these analyses, the risk score of the model can be used as an independent prognostic factor (Figure 3A, Figure 3B).

Verification of the feasibility of the IRGP model to predict survival. To determine whether the model had consistent prognostic value in different risk groups, we applied the model to GSE39582 and GSE17538 as external validation. The patients in the verification cohort were divided into two groups according to the risk score. The OS of subgroups in the low-risk group increased significantly (Figure 2B, Figure S1A). After performing univariate and multivariate Cox proportional hazards analyses in the validation group, we found that the results were similar to those of the training group, and the high risk score of this model suggests a poor prognostic factor (Figure 3C, Figure 3D, Figure S1B and Figure S1C).

Immune cell infiltration in different risk groups. Previous studies have revealed that tumor-infiltrating immune cells are related to prognosis[35]. To determine the infiltration of specific tumor immune cell subsets, we used CIBERSORT to estimate the relative proportion of 22 different immune cells per patient in different risk groups. Three radar charts depict a comparative summary of various immune cells in these two risk groups (Figure 4, Figure S2 and Figure S6). In the training cohort, we found that activated dendritic cells, resting dendritic cells, eosinophils, M0 macrophages, monocytes, resting CD4 memory T cells and regulatory T cells (Tregs) were enriched in different risk groups. Among them, regulatory T cells (Tregs) and M0 macrophages were significantly and highly expressed in the high-risk group, and the rest were highly expressed in the low-risk group (Figure 5). The high-risk group of GSE39582 highly expressed M0 macrophages, M1 macrophages, monocytes, neutrophils, CD8 T cells and follicular helper T cells (Figure S3). The high-risk population in GSE17538 also highly expressed monocytes and Tregs (Figure S7).

Functional evaluation of the IRGP model. To investigate the expression signatures of genetic perturbations that were significantly altered by the IRGP model, GSEA was performed in the high-risk and low-risk groups in the TCGA cohort. The bubble chart revealed that genes in the high-risk populations were enriched in stem cells and various advanced tumors(Figure 6). The top five genetic perturbations in the high-risk group were enriched stem cells, increased breast cancer ductal invasion, a multicancer invasiveness signature, increased advanced vs early-stage gastric cancer and enriched mammary stem cells (Figure 7). We also obtained similar results when performing the above analysis on GSE39582 and GSE17538 (Figure S4, Figure S8). The high-risk group genes in GSE39582 were significantly enriched in

breast cancer ductal invasion and stem cells (Figure S5). The GSEA results obtained in GSE17538 also showed that the high-risk group genes are enriched in tumor cell growth and invasion (Figure S9).

Discussion

Colon cancer is the most common type of gastrointestinal cancer and has high morbidity and mortality. Approximately 95% of colon cancer is colon adenocarcinoma (COAD). In recent years, immunotherapy has been a hotspot in the research of major tumor types. In the COAD field, studies on the high-level microsatellite instability (MSI-H) population have been performed successively since 2015. The Keynote 016, Keynote 164, Checkmate 142, and NICHE clinical trial results all indicate the extraordinary efficacy of immunotherapy[36-39]. Patients with MSI-H have a better prognosis than those with microsatellite stability (MSS). However, the MSI-H population accounts for only approximately 10% of COAD. Most patients still face the dilemma of not having an effective prognostic indicator. Thus, the determination of new prognostic biomarkers is urgent to predict the survival of colon adenocarcinoma patients.

To obtain the robustness of the prognosis prediction in this study, we adopted a method for data analysis without considering the technical deviation of different platforms. The newly established prognostic model is based on the ranking and pairing comparison of relative gene expression values; thus, data preprocessing, such as scaling and normalization, is not required. This method has reliable results in many studies[40, 41].

In this study, we identified an immune-related gene pair model to predict the overall survival for colon adenocarcinoma. The prognostic model comprises 17 immune-related gene pairs containing 26 unique immune-related genes. Most genes in this immune model are cytokine receptors and cytokines, which play a vital role in the adaptive immune response. Among these IRGs, no evidence supports that the overexpression of IL17RB can enhance the invasion and metastasis of thyroid cancer cells[42]. STC2 overexpression is associated with a poor prognosis in patients with nasopharyngeal carcinoma (NPC) and can be used as a predictor of NPC responses to radiation[43]. The increase in IL-7 in colorectal cancer (CRC) is related to metastatic disease and tumor location[44]. Decreased CXCL14 expression indicates a poor prognosis and causes metastasis in colon cancer[45]. GRP signaling alters the invasion of colon cancer through heterochromatin protein $1^{Hs\beta}$ and can improve the prognosis of patients with colon cancer[46]. Moreover, regulatory T cells (Tregs) and M0 macrophages are related to the poor clinical prognosis of many patients with cancer[47, 48]. Dendritic cells are associated with cancer immunity and a favorable prognosis[49]. At the same time, the immune cell types M0 macrophages, M1 macrophages, monocytes, neutrophils, CD8 T cells and follicular helper T cells in the high-risk group of GSE39582 are all related to tumor progression and poor prognosis[50-53]. These findings are consistent with our results. In this study, we also found that several expression characteristics of genetic perturbations, such as increased stem cells, increased breast cancer ductal invasion, a multicancer invasiveness signature, increased advanced vs early gastric cancer and increased mammary stem cells, were related to the IRGP model. These results were verified by corresponding experiments[54-58], confirming their importance in

tumor development and cell growth. These findings indicate that the IRGP model may play an essential role in tumor invasiveness and progression in COAD.

The difference between this study and previously published studies[59] is that the IRGP model was established based on the TCGA database. Second, our strategy to establish a prognostic model was different. To screen out immune-related gene pairs that are significantly related to OS in patients with colon cancer, we used univariate Cox regression analysis before determining the final model using Lasso regression analysis. Finally, we conducted GSEA in the training and validation cohorts to further analyze the specific differences between the high- and low-risk groups. We found that the high-risk group genes were significantly enriched in tumor cell invasion and growth.

Similar to all RNA-seq and microarray analyses, our study had limitations. First, the training dataset to build the immune model was obtained from a retrospective study, which included fresh frozen samples; the stability and efficiency of formalin-fixed and paraffin-embedded (FFPE) samples remain questionable. Therefore, it may be necessary to add more datasets with different sample attributes for more extensive verification. Second, because the prognostic model was based on TCGA and other databases, it required proficiency in bioinformatics. Additionally, the gene expression profiles produced by RNA-seq or microarray platforms require high prices and long conversion cycles. Therefore, this method is challenging to popularize in daily clinical applications.

Conclusions

In summary, our immune-related gene pair model can provide an evaluation reference for the prognostic risk of patients with colon adenocarcinoma. The immune-related model was associated with the prognosis of patients with COAD. The tumor-infiltrating immune cells and genetic perturbations distinguished by this model in the high- and low-risk groups can further elucidate the role of our prognostic model in the development of colon adenocarcinoma. Therefore, the risk model will be a useful tool to better evaluate patients who may benefit from immunotherapy.

Abbreviations

COAD colon adenocarcinoma

CRC colorectal cancer

FFPE formalin-fixed and paraffin-embedded

GARP glycoprotein-A repetitions predominant

GEPs gene expression profiles

GSEA gene set enrichment analysis

hunman.gtf human general transfer format

IRGPs immune-related gene pairs

IRGs immune-related genes

MAD median absolute deviation

MSI-H high-level microsatellite instability

MSS microsatellite stability

NPC nasopharyngeal carcinoma

OS overall survival

ROC receiver operating characteristic

Tregs regulatory T cells

Declarations

Ethics approval and consent to participate

Not applicable. All data in this study are publicly available.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed in this study can be found in the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) and TCGA (https://portal.gdc.cancer.gov/).

Competing interests

The authors declare that they have no conflicts of interest.

Funding

This study was supported by the National Natural Science Foundation of China (No. 81760548). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

JHL, PYL, LBW, YH, YYW, WJG, DC, YJB and ZY contributed to the study conception and design. Material preparation, data collection and analysis were performed by JHL, YJB and ZY. JHL wrote the majority of the first draft of the manuscript text; and YJB and ZY revised the manuscript. All authors read and approved the manuscript.

Acknowledgment

Not Applicable.

References

1. 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. doi:10.3322/caac.21492.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34. doi:10.3322/caac.21551.
3. Center MM, Jemal A, Smith RA, Ward E. Worldwide variations in colorectal cancer. CA Cancer J Clin. 2009;59(6):366-78. doi:10.3322/caac.20038.
4. Bertelsen CA, Larsen HM, Neuenschwander AU, Laurberg S, Kristensen B, Emmertsen KJ. Long-term Functional Outcome After Right-Sided Complete Mesocolic Excision Compared With Conventional Colon Cancer Surgery: A Population-Based Questionnaire Study. Dis Colon Rectum. 2018;61(9):1063-72. doi:10.1097/DCR.0000000000001154.
5. Brungs D, Aghmesheh M, de Souza P, Carolan M, Clingan P, Rose J et al. Safety and Efficacy of Oxaliplatin Doublet Adjuvant Chemotherapy in Elderly Patients With Stage III Colon Cancer. Clin Colorectal Cancer. 2018;17(3):e549-e55. doi:10.1016/j.clcc.2018.05.004.
6. Grothey A, Sobrero AF, Shields AF, Yoshino T, Paul J, Taieb J et al. Duration of Adjuvant Chemotherapy for Stage III Colon Cancer. N Engl J Med. 2018;378(13):1177-88. doi:10.1056/NEJMoa1713709.
7. Guo M, Xu E, Ai D. Inferring Bacterial Infiltration in Primary Colorectal Tumors From Host Whole Genome Sequencing Data. Front Genet. 2019;10:213. doi:10.3389/fgene.2019.00213.
8. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. Gastroenterology. 2015;149(5):1204-25 e12. doi:10.1053/j.gastro.2015.07.011.
9. Allen J, Sears CL. Impact of the gut microbiome on the genome and epigenome of colon epithelial cells: contributions to colorectal cancer development. Genome Med. 2019;11(1):11. doi:10.1186/s13073-019-0621-2.
10. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell. 2019;177(4):1035-49 e19. doi:10.1016/j.cell.2019.03.030.

11. Choi MR, Gwak M, Yoo NJ, Lee SH. Regional Bias of Intratumoral Genetic Heterogeneity of Apoptosis-Related Genes BAX, APAF1, and FLASH in Colon Cancers with High Microsatellite Instability. Dig Dis Sci. 2015;60(6):1674-9. doi:10.1007/s10620-014-3499-2.
12. Sugai T, Eizuka M, Takahashi Y, Fukagawa T, Habano W, Yamamoto E et al. Molecular subtypes of colorectal cancers determined by PCR-based analysis. Cancer Sci. 2017;108(3):427-34. doi:10.1111/cas.13164.
13. Mamlouk S, Childs LH, Aust D, Heim D, Melching F, Oliveira C et al. DNA copy number changes define spatial patterns of heterogeneity in colorectal cancer. Nat Commun. 2017;8:14093. doi:10.1038/ncomms14093.
14. Hsu YL, Lin CC, Jiang JK, Lin HH, Lan YT, Wang HS et al. Clinicopathological and molecular differences in colorectal cancer according to location. Int J Biol Markers. 2019;34(1):47-53. doi:10.1177/1724600818807164.
15. Liu T, Li C, Jin L, Li C, Wang L. The Prognostic Value of m6A RNA Methylation Regulators in Colon Adenocarcinoma. Med Sci Monit. 2019;25:9435-45. doi:10.12659/MSM.920381.
16. Missiaglia E, Jacobs B, D'Ario G, Di Narzo AF, Soneson C, Budinska E et al. Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. Ann Oncol. 2014;25(10):1995-2001. doi:10.1093/annonc/mdu275.
17. Patel SA, Minn AJ. Combination Cancer Therapy with Immune Checkpoint Blockade: Mechanisms and Strategies. Immunity. 2018;48(3):417-33. doi:10.1016/j.immuni.2018.03.007.
18. Woo SR, Corrales L, Gajewski TF. Innate immune recognition of cancer. Annu Rev Immunol. 2015;33:445-74. doi:10.1146/annurev-immunol-032414-112043.
19. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21(8):938-45. doi:10.1038/nm.3909.
20. Salem M, Wallace C, Velegraki M, Li A, Ansa-Addo E, Metelli A et al. GARP Dampens Cancer Immunity by Sustaining Function and Accumulation of Regulatory T Cells in the Colon. Cancer Res. 2019;79(6):1178-90. doi:10.1158/0008-5472.CAN-18-2623.
21. Heinaniemi M, Nykter M, Kramer R, Wienecke-Baldacchino A, Sinkkonen L, Zhou JX et al. Gene-pair expression signatures reveal lineage control. Nat Methods. 2013;10(6):577-83. doi:10.1038/nmeth.2445.
22. Li B, Cui Y, Diehn M, Li R. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. JAMA Oncol. 2017;3(11):1529-37. doi:10.1001/jamaoncol.2017.1609.
23. Wei HT, Guo EN, Liao XW, Chen LS, Wang JL, Ni M et al. Genomescale analysis to identify potential prognostic microRNA biomarkers for predicting overall survival in patients with colon adenocarcinoma. Oncol Rep. 2018;40(4):1947-58. doi:10.3892/or.2018.6607.
24. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS

Med. 2013;10(5):e1001453. doi:10.1371/journal.pmed.1001453.

25. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology. 2010;138(3):958-68. doi:10.1053/j.gastro.2009.11.005.
26. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM et al. Ensembl 2019. Nucleic Acids Res. 2019;47(D1):D745-D51. doi:10.1093/nar/gky1113.
27. Liu W, Islamaj Dogan R, Kwon D, Marques H, Rinaldi F, Wilbur WJ et al. BioC implementations in Go, Perl, Python and Ruby. Database (Oxford). 2014;2014. doi:10.1093/database/bau059.
28. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data. 2018;5:180015. doi:10.1038/sdata.2018.15.
29. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. 2015;12(2):115-21. doi:10.1038/nmeth.3252.
30. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350-6. doi:10.1038/nm.3967.
31. Wu M, Li X, Zhang T, Liu Z, Zhao Y. Identification of a Nine-Gene Signature and Establishment of a Prognostic Nomogram Predicting Overall Survival of Pancreatic Cancer. Front Oncol. 2019;9:996. doi:10.3389/fonc.2019.00996.
32. Wan B, Liu B, Huang Y, Yu G, Lv C. Prognostic value of immune-related genes in clear cell renal cell carcinoma. Aging (Albany NY). 2019;11(23):11474-89. doi:10.18632/aging.102548.
33. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-7. doi:10.1038/nmeth.3337.
34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. doi:10.1073/pnas.0506580102.
35. Domingues P, Gonzalez-Tablas M, Otero A, Pascual D, Miranda D, Ruiz L et al. Tumor infiltrating immune cells in gliomas and meningiomas. Brain Behav Immun. 2016;53:1-15. doi:10.1016/j.bbi.2015.07.019.
36. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz H-J, Morse MA et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. The Lancet Oncology. 2017;18(9):1182-91. doi:10.1016/s1470-2045(17)30422-9.
37. Chalabi M, Fanchi LF, Dijkstra KK, Van den Berg JG, Aalbers AG, Sikorska K et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. Nat Med. 2020;26(4):566-76. doi:10.1038/s41591-020-0805-8.
38. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 2015;372(26):2509-20. doi:10.1056/NEJMoa1500596.

39. Le DT, Kim TW, Van Cutsem E, Geva R, Jager D, Hara H et al. Phase II Open-Label Study of Pembrolizumab in Treatment-Refractory, Microsatellite Instability-High/Mismatch Repair-Deficient Metastatic Colorectal Cancer: KEYNOTE-164. J Clin Oncol. 2020;38(1):11-9. doi:10.1200/JCO.19.02107.
40. Eddy JA, Sung J, Geman D, Price ND. Relative expression analysis for molecular cancer diagnosis and prognosis. Technol Cancer Res Treat. 2010;9(2):149-59. doi:10.1177/153303461000900204.
41. Popovici V, Budinska E, Tejpar S, Weinrich S, Estrella H, Hodgson G et al. Identification of a poor-prognosis BRAF-mutant-like population of patients with colon cancer. J Clin Oncol. 2012;30(12):1288-95. doi:10.1200/JCO.2011.39.5814.
42. Ren L, Xu Y, Liu C, Wang S, Qin G. IL-17RB enhances thyroid cancer cell invasion and metastasis via ERK1/2 pathway-mediated MMP-9 expression. Mol Immunol. 2017;90:126-35. doi:10.1016/j.molimm.2017.06.034.
43. Lin S, Guo Q, Wen J, Li C, Lin J, Cui X et al. Survival analyses correlate stanniocalcin 2 overexpression to poor prognosis of nasopharyngeal carcinomas. J Exp Clin Cancer Res. 2014;33:26. doi:10.1186/1756-9966-33-26.
44. Krzystek-Korpacka M, Zawadzki M, Neubauer K, Bednarz-Misa I, Gorska S, Wisniewski J et al. Elevated systemic interleukin-7 in patients with colorectal cancer and individuals at high risk of cancer: association with lymph node involvement and tumor location in the right colon. Cancer Immunol Immunother. 2017;66(2):171-9. doi:10.1007/s00262-016-1933-3.
45. Liu J, Wang D, Zhang C, Zhang Z, Chen X, Lian J et al. Identification of liver metastasis-associated genes in human colon carcinoma by mRNA profiling. Chin J Cancer Res. 2018;30(6):633-46. doi:10.21147/j.issn.1000-9604.2018.06.08.
46. Tell R, Rivera CA, Eskra J, Taglia LN, Blunier A, Wang QT et al. Gastrin-releasing peptide signaling alters colon cancer invasiveness via heterochromatin protein 1Hsbeta. Am J Pathol. 2011;178(2):672-8. doi:10.1016/j.ajpath.2010.10.017.
47. Najafı M, Farhood B, Mortezaee K. Contribution of regulatory T cells to cancer: A review. J Cell Physiol. 2019;234(6):7983-93. doi:10.1002/jcp.27553.
48. Liu X, Wu S, Yang Y, Zhao M, Zhu G, Hou Z. The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. Biomed Pharmacother. 2017;95:55-61. doi:10.1016/j.biopha.2017.08.003.
49. Veglia F, Gabrilovich DI. Dendritic cells in cancer: the role revisited. Curr Opin Immunol. 2017;45:43-51. doi:10.1016/j.coi.2017.01.002.
50. Giese MA, Hind LE, Huttenlocher A. Neutrophil plasticity in the tumor microenvironment. Blood. 2019;133(20):2159-67. doi:10.1182/blood-2018-11-844548.
51. Olingy CE, Dinh HQ, Hedrick CC. Monocyte heterogeneity and functions in cancer. J Leukoc Biol. 2019;106(2):309-22. doi:10.1002/JLB.4RI0818-311R.
52. Reading JL, Galvez-Cancino F, Swanton C, Lladser A, Peggs KS, Quezada SA. The function and dysfunction of memory CD8(+) T cells in tumor immunity. Immunol Rev. 2018;283(1):194-212.

doi:10.1111/imr.12657.

53. Townsend W, Pasikowska M, Yallop D, Phillips EH, Patten PEM, Salisbury JR et al. The architecture of neoplastic follicles in follicular lymphoma; analysis of the relationship between the tumor and follicular helper T cells. Haematologica. 2020;105(6):1593-603. doi:10.3324/haematol.2019.220160.
54. Taranger CK, Noer A, Sorensen AL, Hakelien AM, Boquest AC, Collas P. Induction of dedifferentiation, genomewide transcriptional programming, and epigenetic reprogramming by extracts of carcinoma and embryonic stem cells. Mol Biol Cell. 2005;16(12):5719-35. doi:10.1091/mbc.e05-06-0572.
55. Hennigs A, Fuchs V, Sinn HP, Riedel F, Rauch G, Smetanay K et al. Do Patients After Reexcision Due to Involved or Close Margins Have the Same Risk of Local Recurrence as Those After One-Step Breast-Conserving Surgery? Ann Surg Oncol. 2016;23(6):1831-7. doi:10.1245/s10434-015-5067-1.
56. Anastassiou D, Rumjantseva V, Cheng W, Huang J, Canoll PD, Yamashiro DJ et al. Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. BMC Cancer. 2011;11:529. doi:10.1186/1471-2407-11-529.
57. Vecchi M, Nuciforo P, Romagnoli S, Confalonieri S, Pellegrini C, Serio G et al. Gene expression analysis of early and advanced gastric cancers. Oncogene. 2007;26(29):4284-94. doi:10.1038/sj.onc.1210208.
58. Fu NY, Pal B, Chen Y, Jackling FC, Milevskiy M, Vaillant F et al. Foxp1 Is Indispensable for Ductal Morphogenesis and Controls the Exit of Mammary Stem Cells from Quiescence. Dev Cell. 2018;47(5):629-44 e8. doi:10.1016/j.devcel.2018.10.001.
59. Wu J, Zhao Y, Zhang J, Wu Q, Wang W. Development and validation of an immune-related gene pairs signature in colorectal cancer. Oncoimmunology. 2019;8(7):1596715. doi:10.1080/2162402X.2019.1596715.

Figures

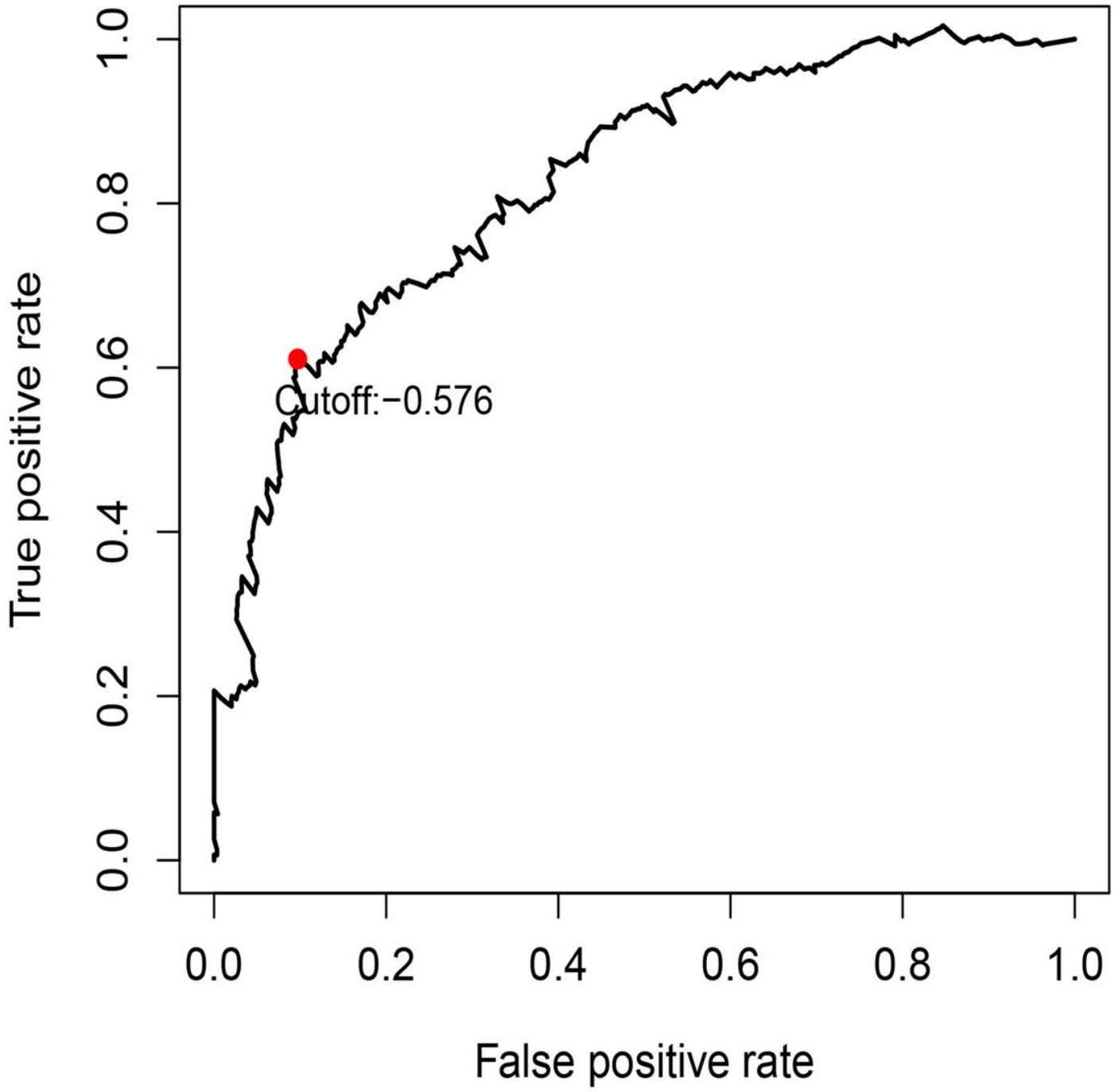

Figure 1

Time-dependent ROC curve for IRGPs risk model in the training cohort. Risk score of −0.576 which was used as cut-off value for the model to stratify patients into high risk group or low risk group. Abbreviations: ROC, receiver operating characteristic; IRGPs, immune-related gene pairs.

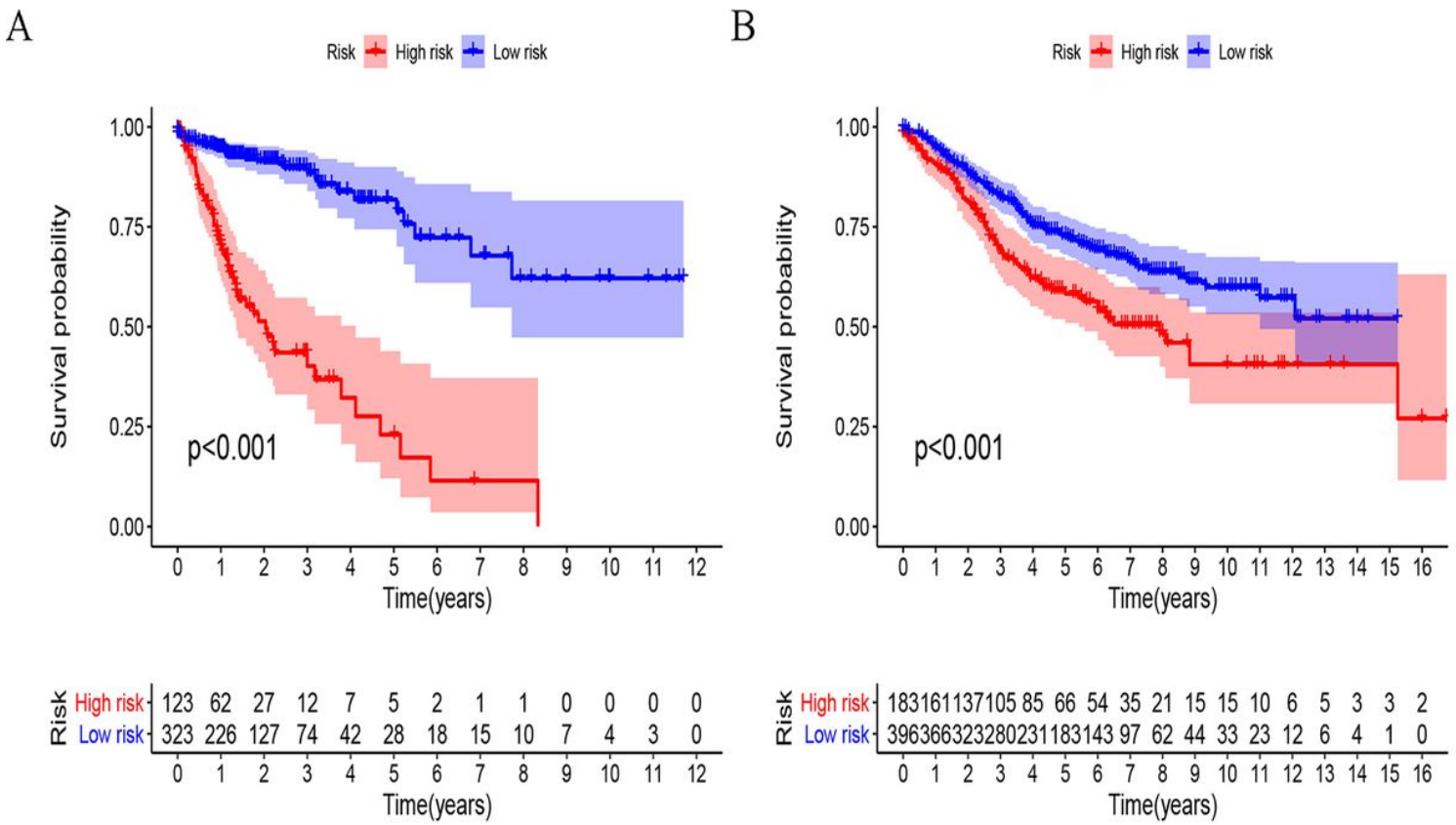

Figure 2

Kaplan-meier curves of OS among different risk groups. Patients were stratified by immune-related gene pairs model. OS among patients in the training (A) and validation cohorts(B). Abbreviation: OS, overall survival.

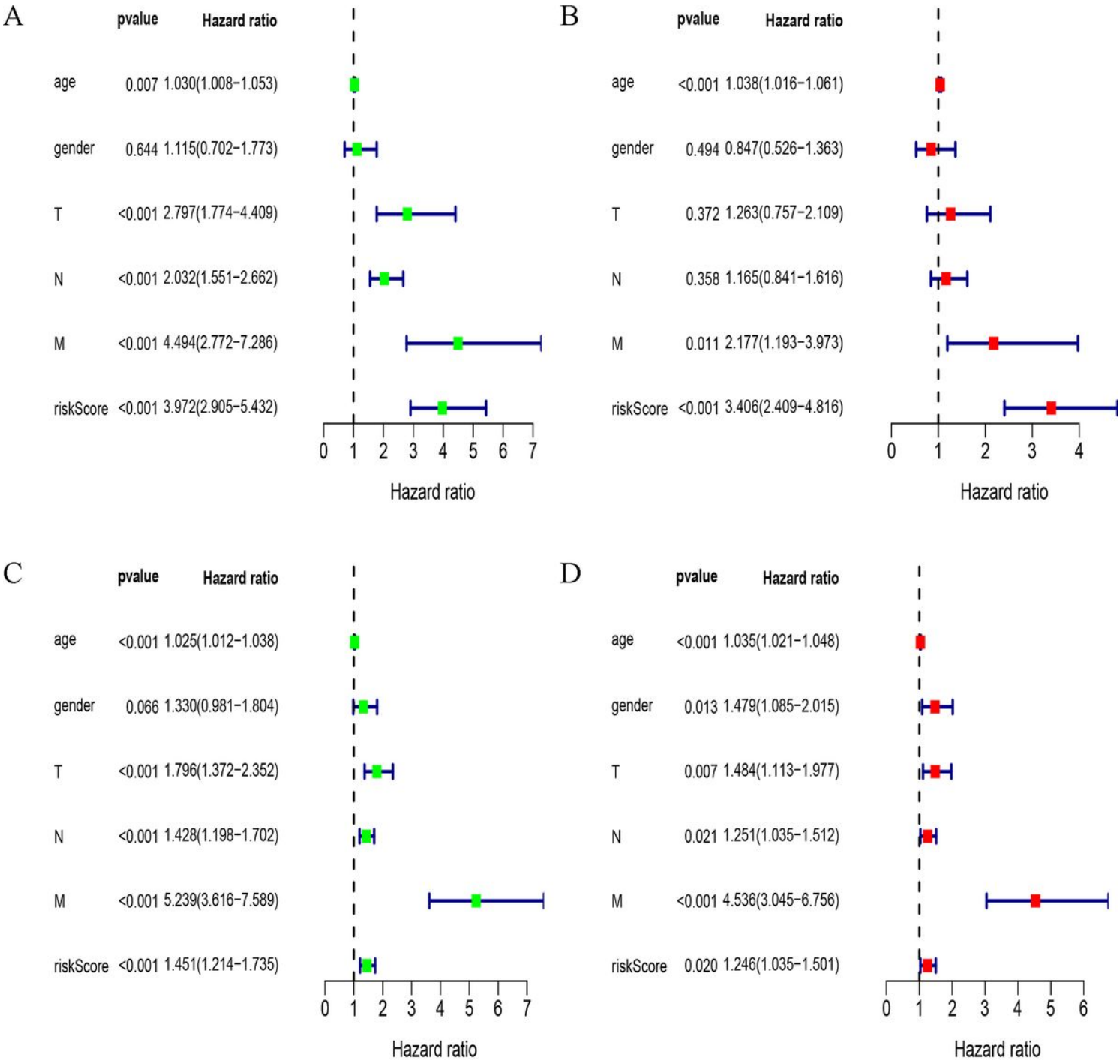

Figure 3

Univariate and multivariate analyses of prognostic factors in the training and validation cohort. (A) and (C) represent the univariate analysis of training cohort and validation cohort, respectively. (B) and (D) represent the multivariate analyses of the training cohort and the validation cohort, respectively.

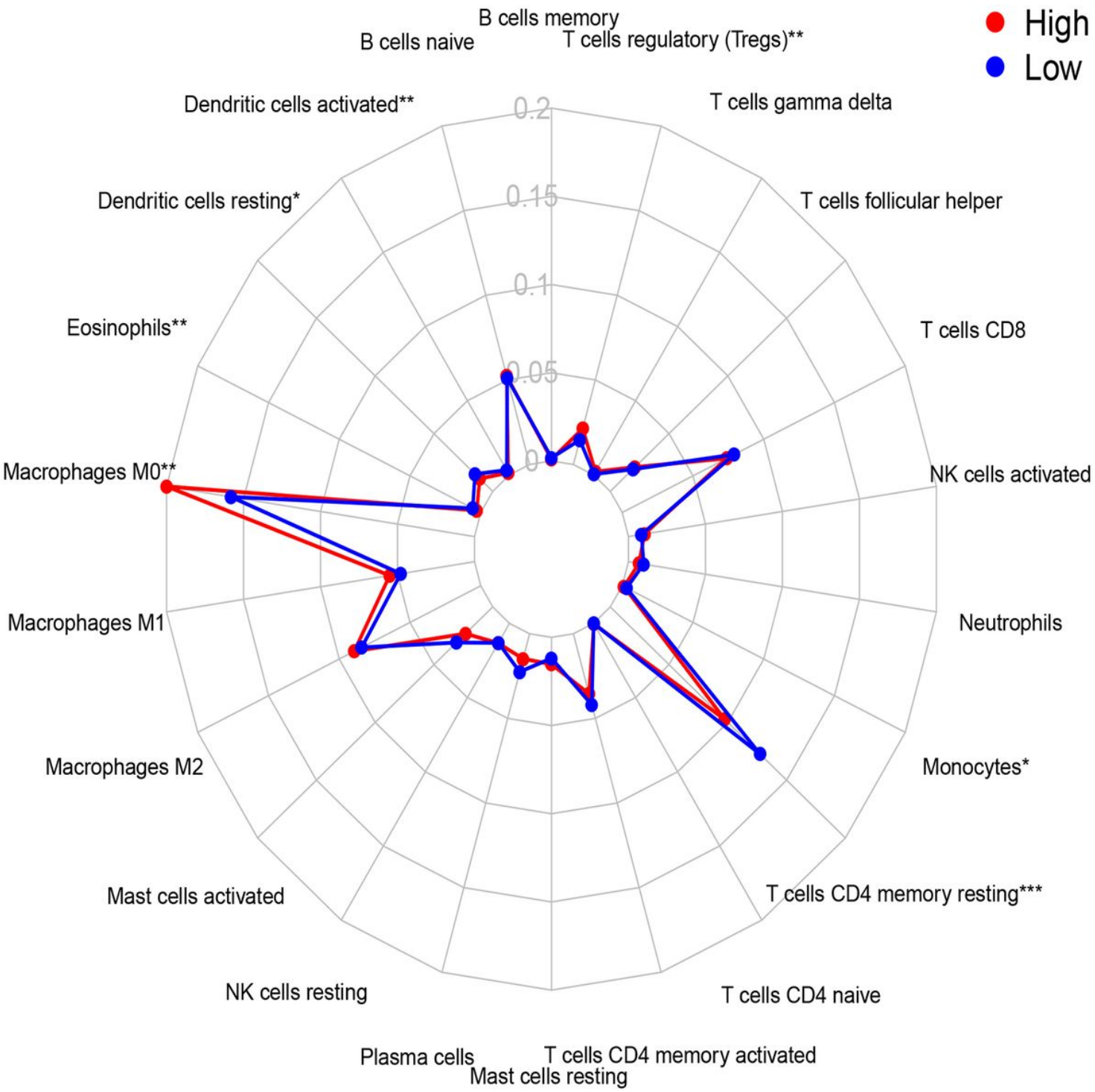

Figure 4

Summary of the 22 immune cells' abundance estimated by CIBERSORT for different risk groups. P-values are based on t-test(*P<0.05, **P<0.01, ***P<0.001).

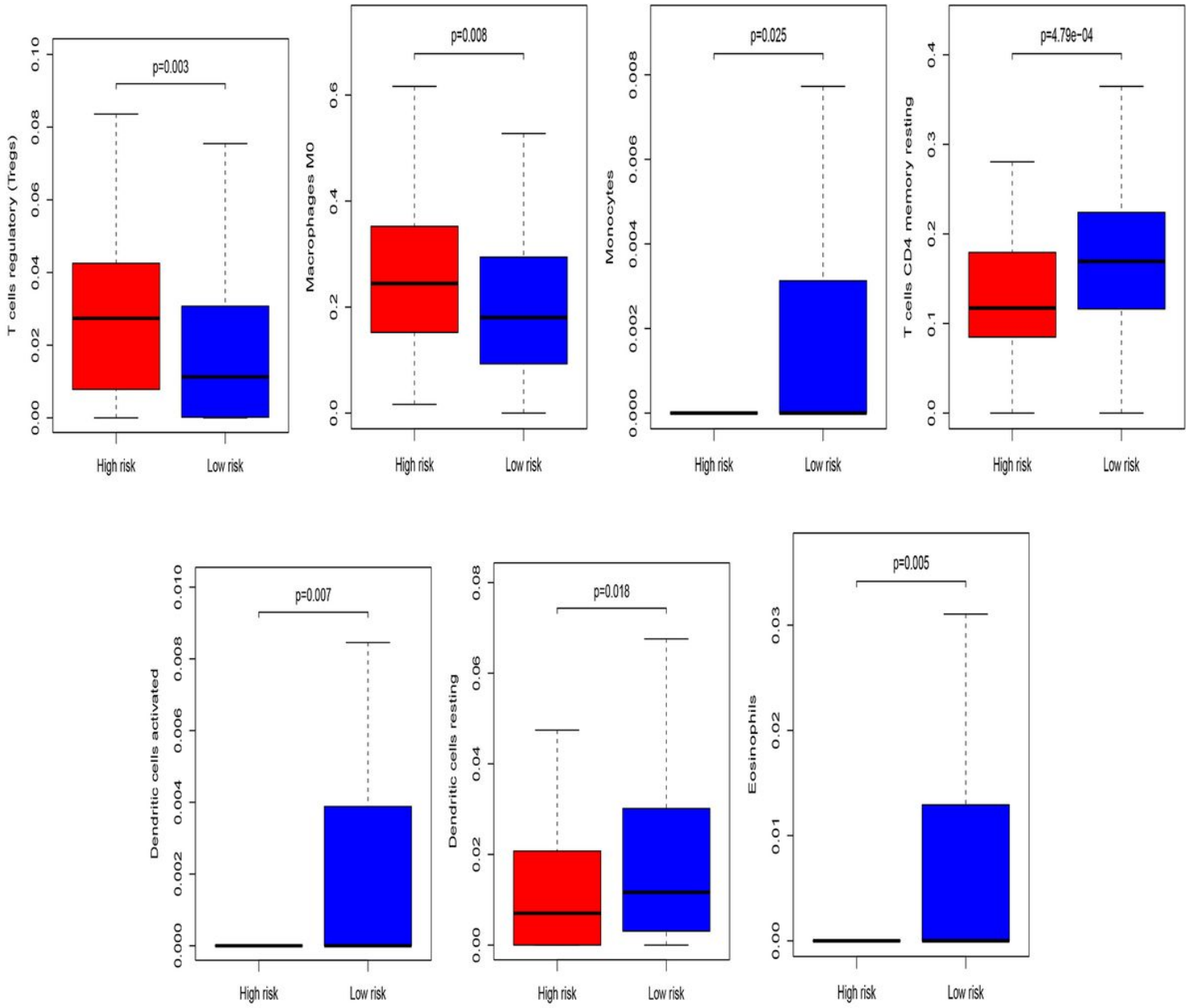

Figure 5

The abundance distribution of specific immune cells' within different risk groups. T cells regulatory and Macrophage M0 were significantly highly expressed in the high-risk group, while the rest were significantly higher in the low-risk group.

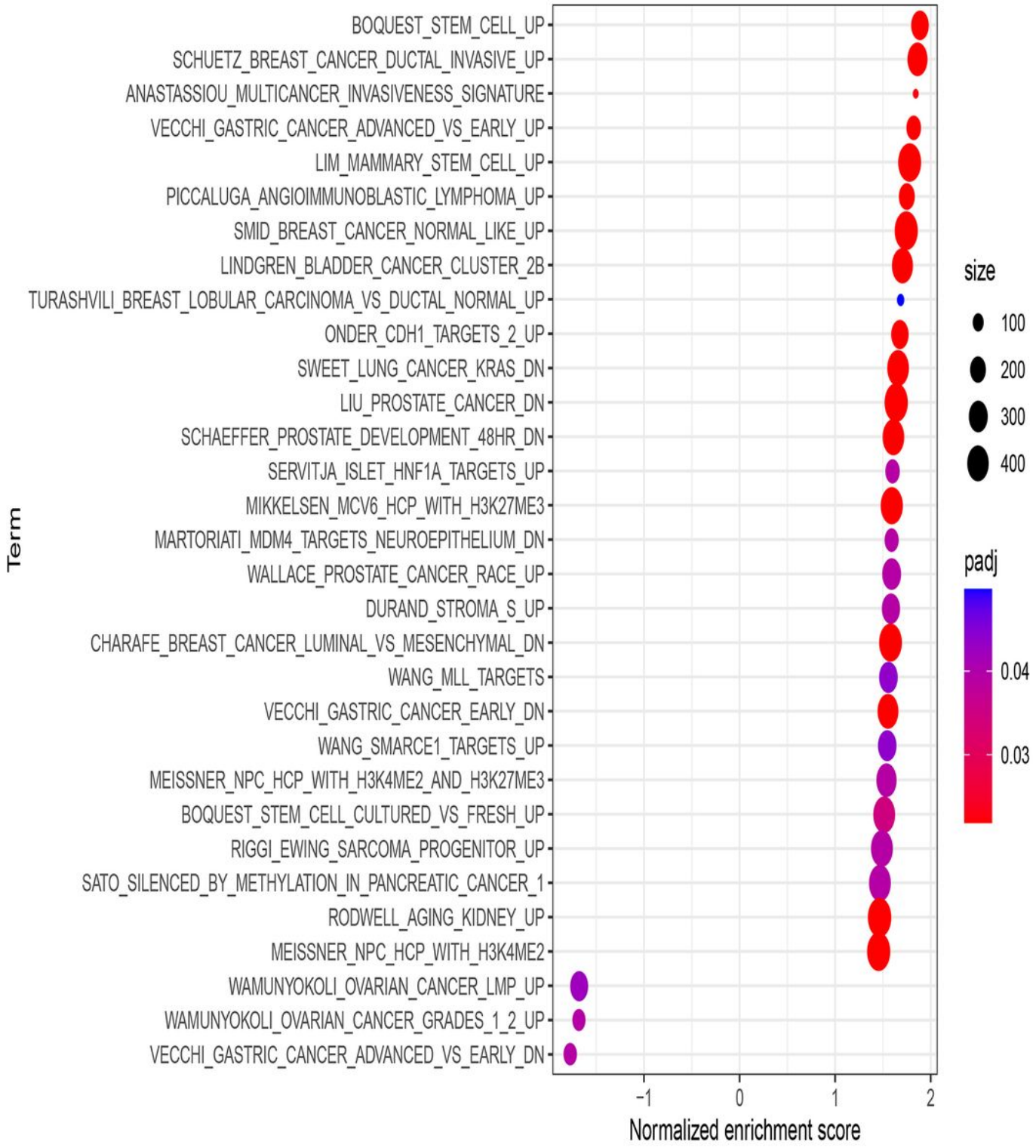

Figure 6

The expression characteristics of genetic perturbations significantly changed by the IRGPs model. A number of these gene sets come in pairs: xxx_UP (and xxx_DN) gene sets representing genes induced (and repressed) by the perturbation.

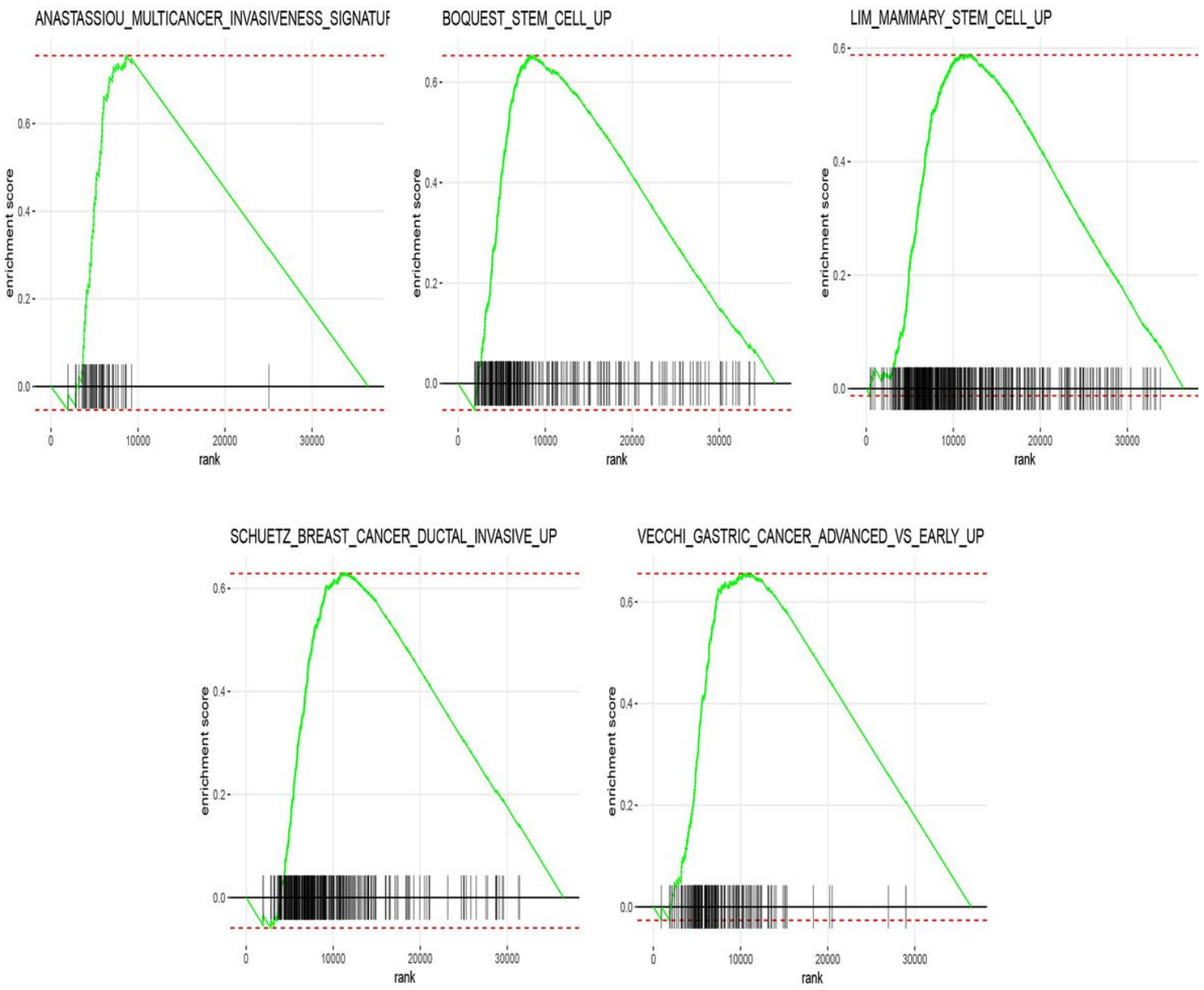

Figure 7

Gene Set Enrichment Analysis (GSEA). Gene set enrichment analysis confirmed that multiple tumor progression and stem cell growth-related pathways in high-risk groups were up-regulated.

Supplementary Files

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

- FigureS1.jpg
- FigureS2.jpg
- FigureS3.jpg

- FigureS4.jpg
- FigureS5.jpg
- FigureS6.jpg
- FigureS7.jpg
- FigureS8.jpg
- FigureS9.jpg