Identification of prognostic immune-related gene signature associated with tumor microenvironment of colorectal cancer

Yuanyuan Wang
Academy of Medical Engineering and Translational Medicine, Tianjin University

Wei Li
Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University

Xiaojing Jin
Departments of Emergency, The First Hospital of Hebei Medical University

Xia Jiang
Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University

Shang Guo
Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University

Fei Xu
Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University

Xingkai Su
Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University

Guiqi Wang
Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University

Zengren Zhao (✉ zzr-doctor@163.com)
Department of General Surgery, Hebei Key Laboratory of Colorectal Cancer Precision Diagnosis and Treatment, The First Hospital of Hebei Medical University

Xiaosong Gu
Academy of Medical Engineering and Translational Medicine, Tianjin University

---

Research Article

Abstract

Background

The tumor microenvironment (TME) has significantly correlation with tumor occurrence and prognosis. Our study aimed to identify the prognostic immune-related genes (IRGs)in the tumor microenvironment of colorectal cancer (CRC).

Methods

Transcriptome and clinical data of CRC cases were downloaded from TCGA and GEO databases. Stromal score, immune score, and tumor purity were calculated by the ESTIMATE algorithm. Based on the scores, we divided CRC patients from the TCGA database into low and high groups, and the differentially expressed genes (DEGs) were identified. Immune-related genes (IRGs) were selected by venn plots. To explore underlying pathways, protein-protein interaction (PPI) networks and functional enrichment analysis were used. After utilizing LASSO Cox regression analysis, we finally established a multi-IRGs signature for predicting the prognosis of CRC patients. A nomogram consists of the thirteen-IRGs signature and clinical parameters was developed to predict the overall survival (OS). We investigated the association between prognostic validated IRGs and immune infiltrates by TIMER database.

Results

Gene expression profiles and clinical information of 1635 CRC patients were collected from the TCGA and GEO databases. Higher stromal score, immune score and lower tumor purity were observed positive correlation with tumor stage and poor OS. Based on stromal score, immune score and tumor purity, 1517 DEGs, 1296 DEGs, and 1892 DEGs were identified respectively. The 948 IRGs were screened by venn plots. A thirteen-IRGs signature was constructed for predicting survival of CRC patients. Nomogram with a C-index of 0.769 (95%CI, 0.717−0.821) was developed to predict survival of CRC patients by integrating clinical parameters and thirteen-IRGs signature. The AUC for 1-, 3-, and 5-year OS were 0.789, 0.783 and 0.790, respectively. Results from TIMER database revealed that CD1B, GPX3 and IDO1 were significantly related with immune infiltrates.

Conclusions

In this study, we established a novel thirteen immune-related genes signature that may serve as a validated prognostic predictor for CRC patients, thus will be conducive to individualized treatment decisions.

Background

Colorectal cancer (CRC) is ranked as the third most common cause of cancer-related mortality globally. In the gastrointestinal tract, the incidence of CRC is one of the most frequently diagnosed (1). The heterogeneity of CRC exists with differences in molecular pathogenesis, clinical features, and prognosis (2). A significant number of early-stage CRC patients successfully undergo aggressive surgical removal of the primary tumors (2). However, among stage II/III patients, the association of the recurrence with metastases has been evident (3). Stage IV patients with high risks of metastatic relapses receive standardized cytotoxic chemotherapy (4). Generally, the treatment of CRC involves a combination of one or more cancer drugs in their regimens, such as 5-fluorouracil (5-FU), folinic acid, irinotecan and oxaliplatin (5). But in the presence of metastases, new adjuvant chemotherapy may not produce successful results as it could lead to cancer progression and the patient's resistance to the drug (6).

The key determinants of overall survival (OS) are high-risk pathological characteristics like peritoneal involvement/serosal invasion or vascular invasion. These are predominantly present in diseases that are classified as node-negative (7, 8). However, it has become evident that the profiles of other hosts and tumors could also impact the clinical prognosis. Furthermore, the fundamental characteristics of tumor cells and mechanisms within the tumor microenvironment (TME), such as tumor-infiltrating immune cells (TIICs) (9) and stromal cells linked to the tumor, affect the progression of CRC and the clinical outcomes (10).

In epithelial cancer cells, the increased presence of mesenchymal genes leads to the recommendation that the poor prognosis of CRC patients is associated with the epithelial-to-mesenchymal transition (11). However, the transcriptome of the tissue that contains the tumor characterizes the expression profiling of mesenchymal cells that establish the TME and the epithelial cancer cells (12, 13). The stroma or TME is comprised of the structural and functional features of the connective tissue in homeostasis and pathological angiogenesis in wound healing and illness. Furthermore, the stroma is made up of lymph and blood vessels, inflammatory/immune cells, extracellular matrix, and fibroblasts. The cancer cells alter the dynamic and multifaceted structure of the stroma, which eventually affects the progression of malignant cells (14).

The two main classifications of non-tumor mechanisms are the stromal and immune cells in the TME. These play a significant role when diagnosing, making a prognosis for, and assessing tumors. In the recent decade, researchers have created several algorithms to forecast the purity of tumors in different types of cancer. These algorithms are grounded on the precise gene signature of immune cells and/or stromal cells (15, 16). An example is the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE), which is an algorithm designed by Yoshihara K et al. (17). This algorithm calculates both stromal score and immune score to forecast the presence of infiltrating non-tumor cells by analyzing the signatures of specific genes.

After the founding of the ESTIMATE logarithm, research studies have applied it to prove the efficacy of big-data algorithms on evaluating prostate cancer (18), glioblastoma (19), and cutaneous melanoma (20). Nevertheless, researchers also need to extensively study the value of the ESTIMATE when evaluating

stromal and/or immune scores of CRC. In this literature, we have pioneered the extraction of the list that itemizes how the TME-associated genes forecast poor prognosis amongst CRC patients. We completed this by using the data of CRC cohorts from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA).

Methods

Gene expression data sets

Our study utilized publicly available data sets. We collected gene expression profiles and identified the equivalent material on the prognosis, as well as the tumor and normal tissues of CRC patients. We obtained the data from the TCGA (https://tcga-data.nci.nih.gov/tcga/) and the GEO (https://www.ncbi.nlm.nih.gov/geo/) web sites, which were uploaded up to 31 March 2019. Moreover, we excluded duplications and datasets that have sample sizes of less than 50 (N < 50). Then, we organized the clinical data and expression profiles of each sample manually. Our inclusion criteria indicate that we include diagnosed CRC patients who have available clinicopathological and survival information. Therefore, six datasets were included in our study (GSE12945, GSE39582, GSE41258, GSE72970, GSE103479 and TCGA)(21-26). Stromal scores, immune scores and tumor purity were calculated by the ESTIMATE algorithm (17). For the succeeding genomic analysis, we used the TCGA dataset. To authenticate the prognosis of genetic information recognized by the TCGA analysis, we selected the largest CRC dataset GSE39582 as the validation cohort from the GEO database. This study complied with the TCGA and GEO approved publication guidelines. Also, we sourced the data from both databases. Thus, this research did not require the approval of the ethics committee.

Identification of immune-related genes (IRGs)

Based on the ESTIMATE results, we classified the sample groups of the stromal scores, immune scores and tumor purity into high and low to choose the intersection genes. This study analyzed the data using the R package limma (27). The cutoff values were established at FDR < 0.05 and Fold change > 2 to filter the differentially expressed genes. Finally, we obtained intersect genes among stromal scores, immune scores and tumor purity as immune-related genes (IRGs).

Functional enrichment of IRGs

To complete the functional enrichment analysis of the IRGs and classify the GO categories according to the molecular functions (MF), the biological processes (BP), or the cellular components (CC), we used a R package called "clusterprofiler" (28). Additionally, we utilized it to make a pathway enrichment analysis while referring to the pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG). Finally, the cut-off value was seted at a false discovery rate (FDR) < 0.05.

Protein-protein interaction (PPI) network

The STRING database (29) produced the PPI network that the Cytoscape software rebuilt (30). For further examination, this study only included individual networks with 10 or more nodes and excluded those with less than 10. The connectivity degree of each network node was calculated, and then, we searched for the clusters according to their typology to trace densely connected regions using the Molecular COmplex DEtection (MCODE).

Establishment of the IRGs signature for CRC

To investigate the link between IRGs and the prognosis of CRC patients, we applied the univariate Cox regression analysis. In this analysis, the statistical significance was established at p < 0.05. We selected a panel of genes according to these outcomes through the LASSO Cox regression analysis that used the R package "glmnet". Then, we set up a multigene signature to forecast the prognosis of these patients. By conducting cross-validations 1000 times and adopting the best penalty parameter λ value's one standard error, we established the most simplified (smallest perimeter) model of immune gene expression signatures. For each patient, the sum of the corresponding coefficients and products of each gene expression level determined the risk score formula. To verify the efficiency of the signature constructed by the TCGA cohort, we validated the results in the GEO cohort (GSE39582).

TIMER Database Analysis

To methodically analyses the immune infiltrates in different cancers, TIMER is a wide-ranging resource (https://cistrome.shinyapps.io/timer/) (31). This algorithm uses a statistical deconvolution method known as the, which was recently made available (15) to determine the TIICs based on the gene expression profiles. Moreover, TIMER covers 32 types of cancer and contains 10,897 samples from TCGA to approximate the abundance of immune-infiltrating. Our analysis encompassed the identification of prognostic immune-related genes (IRGs) in CRC and its correlation with the abundance of immune-infiltrating. These infiltrates include macrophages, neutrophils, B cells, CD8+ T cells, CD4+ T cells and dendritic cells (DC) through gene modules. On the left-most panel, the graph illustrates the gene expression profiles against tumor purity (32).

Statistical analysis

Results were showed as mean ± standard deviation (SD). To contrast the immune score, stromal score and tumor purity in different groups, we used the Kruskal-Wallis test of variance. The cutoff value of high and lower groups of immune score, stromal score and tumor purity were calculated by X-tile (33). We calculated the OS using the Kaplan-Meier method, and the statistical significance was determined by the log-rank test. The specificity and sensitivity of survival prediction according to the determined risk score were obtained by time-dependent receiver operating characteristic (ROC) curves, with AUC values quantified with the survivalROC package. This paper also executed the univariate and multivariate analyses of OS to distinguish the prognostic determinants of CRC patients from the TCGA and GSE39582 datasets. Furthermore, using the R package "rms", we established a nomogram according to the TCGA CRC

cohort and incorporated different prognostic factors. Moreover, we analyzed its performance through the C-index, calibration plots and decision curve analysis (DCA). Additionally, this study utilized the R software version 3.6.0 to conduct all of the statistical analyses (http://www.rproject.org/). Finally, we set the statistical significance at a two-sided P < 0.05.

Results

Association of stromal scores, immune scores and tumor purity with CRCprognosis

Stromal scores, immune scores and tumor purity were significantly associated with CRC clinical stages and prognosis. Using a variety of technologies, we obtained the genomic profiles. The stromal and immune scores, as well as the tumor purity, were generated through the uniform algorithm. The results did not reveal any obvious cohort-bias clustering. Hence, we merged the groups from TCGA and GEO and further investigated whether there were stromal scores, immune scores and tumor purity statistically correlated with CRC clinical stages and prognosis. In the present study, we downloaded the gene expression profiles and clinical data of 1635 CRC patients from the TCGA and GEO databases. Patients were 66.06 ± 12.95 years old, with 882 males (53.9%) and 753 (46.1%) females.

For CRC, TNM stages are valuable prognostic indicators. We included them in the analysis. The stromal scores ranged from -2232.54 to 2193.09, immune scores were distributed between -899.57 and 3202.84, and tumor purity ranged from 0.28 to 0.98, respectively (Data were calculated by ESTIMATE algorithm). Figure 1 illustrates that across different stages, the distribution of stromal scores differed (p = 4.79e-10) while the distribution of immune scores had no variation (p = 0.58). Figures 1 show that tumor purity (p = 1.40e-04) were inversely associated with different stages.

To study the possible correlation of the prognosis with stromal score, as well as immune score and tumor purity, we classified the CRC patients into high and lower groups. For this purpose, we created the X-tile and Kaplan-Meier survival curves. The results suggest that the stromal scores were positive correlation with the OS (p=0.036) is statistically significant. Also, the immune scores are positively associated with OS, but it was not statistically significant (p=0.38). Conversely, tumor purity was significantly negative association with OS (p=0.024) (Figure 1).

Identification of DEGs

To uncover the correlation of gene expression profiles with stromal scores, immune scores and tumor purity, we analysed RNAseq data of all 611 CRC cases obtaining in TCGA database. Figure 2, the volcano plots highlight unique gene expression profiles of cases that we classified under high/low stromal scores, immune scores, and tumor purity groups. We based our comparative analyses on three factors: first, stromal scores through the upregulation of 1507 genes and downregulation of 10 genes in the high stromal score group; second, immune scores through the upregulation of 1235 genes and downregulation of 61 genes in the high and low groups; and third, tumor purity group through the upregulation of 1824 genes and downregulation of 68 genes in the low score group. Furthermore, Figure 2 illustrates the IRGs were the upregulated or downregulated intersection genes in the low tumor purity group and high stromal or immune groups (944 upregulated genes and 4 downregulated genes).

Functional analysis of IRGs

Using the STRING tool that included 2834 edges and 471 nodes, we generated PPI networks to study the interactions among the identified immune-related genes (Figure 3A). For further analysis, we chose three modules with at least 20 nodes. As shown in module A (Figure 3B), the network had formations of 903 edges and 43 nodes. Each gene in this model was associated with 42 other genes, indicating that this module might be the core sub-network of PPI. In module B (Figure 3C), there were 395 edges and 43 nodes that had numerous genes in the middle that were critical to immune response. These included OLR1, ITGB2, SLC2A3, ITGAM, and ATP8B4. In module C (Figure 3D), there were 198 edges and 29 nodes where the higher degree values of connectivity in FBN1, COL3A1, and COL1A2 proved their existence as the core genes in the module.

The clustering of these enriched functional-related genes is associated with the immune response, which is in line with the PPI network analysis. Furthermore, we identified statistically significant GO terms, such as the total of 1325 for biological processes, 83 for molecular function, and 77 for the cellular component (FDR<0.05). GO terms (Table S1) included regulation of leukocyte activation, leukocyte migration and T cell activation (Figure 4A), extracellular matrix and side of membrane (Figure 4B), and receptor regulator activity (Figure 4C). Additionally, the KEGG analysis (Figure 4D) generated pathways that were all linked with the immune response.

Construction and validation of the IRGs-based signature

In the context of this study's TCGA CRC patients, it utilized the univariate Cox models to investigate the association between the expression levels of their IRGs and their OS. Then, to perform the LASSO method, our selection included 171 statistically significant IRGs, which ultimately determined the fundamental IRGs that were significantly related to OS (Figure 5A and 5B). Subsequently, thirteen IRGs and their coefficient values were identified. The formula is as follows: risk score = (0.1217 × expression of A2ML1) + (0.03442 × expression of CALB2) + (-0.6693 × expression of CD1B) + (0.04806 × expression of COL22A1) + (0.4471 × expression of FCRL2) + (0.00069 × expression of GPX3) + (0.05368 × expression of HAND1) + (0.0023 × expression of IDO1) + (0.006 × expression of LAMP5) + (0.07625 × expression of MAP2) + (0.02431 × expression of MMRN1) + (0.1085 × expression of NKAIN4) + (0.3541 × expression of VAX2). Patients were classified into low- and high-risk groups according to a cutoff risk score of 1.096 (Figure 5C and 5D). Figure 5E demonstrates that the high-risk group had substantially lower levels of OS than in the low-risk at p < 0.0001. For 1-, 3-, and 5-year OS, the area under the ROC curve (AUC) was 0.713, 0.724, and 0.689, respectively (Figure 5F). The GSE39582 cohort verified the prognostic model (N=531). Patients were classified into low- and high-risk groups according to an ideal cutoff risk score of 0.973 (Figure 5G and 5H). Figure 5I illustrates that the high-risk group had substantially lower levels of OS than in the low-risk at p = 5.97e-04. For 1-, 3-, and 5-year OS, the AUC was 0.725, 0.643, and 0.606, respectively (Figure 5J).

Independent prognostic role of the thirteen-IRGs signature

The univariate and multivariate Cox regression analyses were executed in both GEO and TCGA datasets through the adjustment of clinicopathological features like the T stage, N stage, M stage, and tumor stage, as well as the age and gender. This process aimed to assess if the signature-based risk score of the thirteen-IRGs was an independent prognosis factor for OS, which the results confirmed as true (Figure 6A, 6B, 6D and 6E). We verified the clinical significance of the thirteen-IRGs signature using the Chi-square test to ascertain the signature's association with the clinical parameters. In the TCGA cohort, significant correlations were found between the higher-risk score and tumor stage (p<0.001), M stage (p<0.01), N stage (p<0.001), and T stage (p<0.001) (Figure 6C). However, no significant difference was found in gender and age. Comparable outcomes were observed in the validation cohort of GSE39582 (Figure 6F).

Establishment and assessment of the nomogram based on the thirteen-IRGs signature

A nomogram was established in terms of the smallest Akaike Information Criterion (AIC) value occurred from the multivariate Cox regression (AIC = 1132.81), which includes age, T stage, N stage, M stage and the thirteen-IRGs signature (Figure 7A). The calibration plot demonstrated that the evaluating capability of the nomogram was best in forecasting 3-year OS (Figure 7B). The C-index of the nomogram was 0.770 (95% CI = 0.718−0.823). As Figure 7C displays, the nomogram illustrated a greater net benefit that had a wider range of threshold probability on the DCA for predicting 1-, 3-, and 5- year OS. Combining thirteen-IRGs signature with TNM stage showed better net benefit for predicting 3-year OS (Figure 7D). The AUC for 1-, 3-, and 5- year OS of the nomogram in TCGA dataset was 0.789, 0.783 and 0.790, respectively (Figure 7E). Based on Figure 7F, CRC patients from the high-risk group, stratified by the nomogram, had significantly lower OS rates compared to the low-risk group (P <0.001). Using the GSE39582 cohort for validation, we obtained similar results, as displayed by Figure S1.

Survival analysis of IRGs

For additional confirmation of their prognostic value and expression, we applied the Kaplan-Meier survival analysis for the thirteen IRGs in signature in patients with CRC from TCGA (Figure 8). We found that A2ML1, CALB2, COL22A1, FCRL2, GPX3, HAND1, IDO1, LAMP5, MAP2, MMRN1, NKAIN4 and VAX2 were identified as cancer-promoting factors given their high expression correlation with shorter OS in patients with CRC. On the other hand, the high expression of CD1B exhibited a significant correlation with longer OS. This association implies the possible protective role of RNAs in CRC. Additionally, we verified these results in the GSE39582 cohort (Figure S2).

Correlation between TIICs and prognostic IRGs

The independent predicted factors of sentinel lymph node status and OS among cancer patients are the Tumor-infiltrating immune cells (TIICs) (34). As a result, we explored the correlation of those thirteen prognostic IRGs with the immune infiltration levels in CRC using TIMER (Table 1). The outcomes reveal the positive correlation of the CD1B expression level with the infiltrating levels of neutrophils (r=0.468, p=3.31E-23) and Dendritic Cells (r=0.505, P=2.05E-27) in colon adenocarcinoma (COAD) (Figure 9A). Moreover, GPX3 has the highest significant correlations with infiltrating levels of macrophages (r=0.54, p=5.93E-32), neutrophils (r=0.408, p=4.45E-17) and Dendritic Cells (r=0.434, P=7.01E-20) (Figure 9B). In addition, IDO1 expression was significantly associated with infiltrating levels of CD8+ T cells (r=0.4, P=4.66E-17), neutrophils (r=0.638, p=3.97E-47) and Dendritic Cells (r=0.564, P=3.47E-35) (Figure 9C).

Discussion

The TME affects the progression and growth of a tumor (35). Moreover, it is comprised of the tumor and non-tumor cells like fibroblasts and immune cells (36). Tumor-infiltrating immune cells are significantly linked to oncogenesis and angiogenesis, as well as to the metastasis and growth of tumor cells, which could regulate the abundance and differentiation of immune cells(37). Recent literature highlights how the discrepancies between tumor progression and host's immune response could result in tumor's growth (38). Thus, it is vital to understand the stromal and immune status in the TME to develop strategies that could enhance the patient's response rate to immunotherapy.

The cutting-edge therapeutic methods include immunotherapies and molecularly targeted therapies, as well as radiotherapies and chemotherapies. However, limitations remain prevalent and include toxicity, immunity of patients to treatments, low response rates, and financial burdens caused by high costs (39, 40). Also, modern cancer treatments still need biomarkers for prognosis. Hence, we attempted to find prognostic IRGs that impact the patients' OS by analyzing the TME.

Some studies applied the ESTIMATE to some types of cancers (19, 20), which signify the benefits of using algorithms in large data sets. For example, Vincent et al. used this algorithm to compute for the breast cancer tumor purity and prove that the immune and stromal components are nonexistent in vitro (41). To comprehend the TME of CRC, we applied the ESTIMATE to generate tumor purity, as well as the stromal and immune scores. In recent studies, (42, 43) our analysis deduced the positive correlation of stromal scores and immune scores with the stages of a tumor, and tumor purity was inversely correlated with tumor stages. Moreover, longer survival times highlighted the presence of low tumor purity and high stromal and immune scores amongst the CRC patients. This suggests the key role of TME in patient outcomes.

We classified the TCGA patients into high and low stromal scores, immune scores and tumor purity groups to detect IRGs. Then, we applied the GO analysis to confirm the involvement of those genes in the TME, such as leukocyte activation, leukocyte migration, T cell activation, extracellular matrix, side of membrane and receptor regulator activity. The outcomes have been consistent with the results of previous studies, where immune cells and ECM molecules influenced the creation of the interdependence within the CRC's tumor microenvironment (14, 35). Also, we established the PPI modules to unveil the association and purpose of IRGs. The nodes with high degree of connectivity among modules, such as C3, FBN1 and ITGB2, were linked to apoptosis, migration, angiogenesis, proliferation, and immune response (44−46).

Then, we performed LASSO Cox analyses to generate a thirteen-IRGs signature for predicting the prognosis of OS patients in TCGA dataset, and cross-validated from the GEO database GSE39582. The AUC values for the signature in predicting 1, 3, and 5-year survival were 0.703, 0.711, and 0.676, which indicates a good capability for predicting survival in CRC patients. We further evaluated the association between the signature and clinical parameters to figure out the clinical value of this thirteen-IRGs signature. It was found that high-risk patients were associated significantly with TNM stages. DCA and ROC analysis indicated that the nomogram consists of the thirteen-IRGs signature and conventional clinical parameters exhibited high performance in prognostic sensitivity for patients with CRC. These results demonstrated great applicability and stability of the thirteen-IRGs signature for predicting prognosis of patients with CRC.

To ascertain the association of IRGs with immune infiltration, we used a marker called CD1B. Guo et al. found that miR-582/CD1B, which are involved in resting and activated dendritic cells, may be potential novel biomarkers for immunotherapy (47). Similar to our study, in prostate cancer, the decreased levels of CD1B expression could be linked to the poorer disease-free survival (48). Finally, using the TIMER algorithm, we evaluated how prognosis-related genes are correlated with immune-infiltrating levels in CRC. Our results suggest that in both READ and COAD, the expression level of CD1B is significantly positive correlation with the tumor-infiltrating levels of neutrophils and Dendritic Cells.

Previous studies indicated that IDO1 suppressed the CD8 + T cell response in colon cancer, which may provide a theoretical basis for the development of new immunotherapy for the treatment of colon cancer (49). The positive correlation of the IDO1 expression with PD-L1 pathways on T cells is also evident (50). Furthermore, during the PD-L1 therapy, both CD8[+] and CD4[+] T cells had higher abundance levels in the TME. As a result, this activates the IDO1 on the T cells, which improves this process and causes improvements in tumor immunity.

To our understanding, this is a pioneering study on CRC that investigates IRGs in the TME. Thus, we took advantage of the opportunities. First, we used the GEO and TCGA databases to study the TME. Both sources gave us an extensive amount of CRC samples. We crosschecked our results using independent cohorts. Then, we acknowledged the complexities associated with the TME, which include genetic factors among many others. Furthermore, this study extensively evaluated how the TME (immune score, stromal scores and tumor purity) interacted with prognostic IRGs.

Conclusions

In conclusion, our current study established a thirteen-gene signature based on ESTIMATE algorithm-derived immune/stromal scores, which could serve as a favorable prognostic factor. Our model would be of clinical value and provide additional data for better understanding of the TME. Finally, the results of functional analysis provide a novel insight for revealing the molecular mechanism in tumor initiation and progression.

Abbreviations

TME: tumor microenvironment, IRGs:immune-related genes, CRC:colorectal cancer, TCGA:The Cancer Genome Atlas, GEO:Gene Expression Omnibus, ESTIMATE:Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data, DEGs:differentially expressed genes, PPI:protein-protein interaction, LASSO:Least absolute shrinkage and selection operator, OS:overall survival, TIICs:tumor-infiltrating immune cells, FDR:false discovery rate, GO:gene ontology, MF:molecular functions, BP:biological processes, CC:cellular components, KEGG:Kyoto Encyclopedia of Genes and Genomes, ROC:receiver operating characteristic, DCA:decision curve analysis, AIC:Akaike Information Criterion.

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 data that support the findings of this study are available in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). These data were derived from the following resources available in the public domain: [https://tcga-data.nci.nih.gov/] and [https://www.ncbi.nlm.nih.gov/geo/].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

WL, YYW, XJ and ZRZ contributed to the planning of the study. WL and XJJ conducted data analysis. WL, SG, XKS and FX drafted the manuscript and revised the manuscript. WL, YYW, and GQW prepared all the figures and tables. ZRZ and XSG contributed to interpretation of data and review of the manuscript. All the authors reviewed and approved the final manuscript.

Acknowledgements

The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov.

References

1. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2015;64(10):1637-49.

2. Bahrami A, Shahidsales S, Khazaei M, Ghayour-Mobarhan M, Maftouh M, Hassanian SM, et al. C-Met as a potential target for the treatment of gastrointestinal cancer: Current status and future perspectives. Journal of cellular physiology. 2017;232(10):2657-73.

3. Gholamin S, Fiuji H, Maftouh M, Mirhafez R, Shandiz FH, Avan A. Targeting c-MET/HGF signaling pathway in upper gastrointestinal cancers: rationale and progress. Current drug targets. 2014;15(14):1302-11.

4. Bahrami A, Khazaei M, Hasanzadeh M, ShahidSales S, Joudi Mashhad M, Farazestanian M, et al. Therapeutic Potential of Targeting PI3K/AKT Pathway in Treatment of Colorectal Cancer: Rational and Progress. Journal of cellular biochemistry. 2018;119(3):2460-9.

5. Bahrami A, Hesari A, Khazaei M, Hassanian SM, Ferns GA, Avan A. The therapeutic potential of targeting the BRAF mutation in patients with colorectal cancer. Journal of cellular physiology. 2018;233(3):2162-9.

6. Bahrami A, Amerizadeh F, ShahidSales S, Khazaei M, Ghayour-Mobarhan M, Sadeghnia HR, et al. Therapeutic Potential of Targeting Wnt/beta-Catenin Pathway in Treatment of Colorectal Cancer: Rational and Progress. Journal of cellular biochemistry. 2017;118(8):1979-83.

7. Petersen VC, Baxter KJ, Love SB, Shepherd NA. Identification of objective pathological prognostic determinants and models of prognosis in Dukes' B colon cancer. Gut. 2002;51(1):65-9.

8. Roxburgh CS, McMillan DC, Richards CH, Atwan M, Anderson JH, Harvey T, et al. The clinical utility of the combination of T stage and venous invasion to predict survival in patients undergoing surgery for colorectal cancer. Annals of surgery. 2014;259(6):1156-65.

9. Roxburgh CS, McMillan DC. The role of the in situ local inflammatory response in predicting recurrence and survival in patients with primary operable colorectal cancer. Cancer treatment reviews. 2012;38(5):451-66.

10. Freeman MR, Li Q, Chung LW. Can stroma reaction predict cancer lethality? Clinical cancer research : an official journal of the American Association for Cancer Research. 2013;19(18):4905-7.

11. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. The Journal of clinical investigation. 2009;119(6):1420-8.

12. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome biology. 2016;17(1):231.

13. Winslow S, Lindquist KE, Edsjo A, Larsson C. The expression pattern of matrix-producing tumor stroma is of prognostic importance in breast cancer. BMC cancer. 2016;16(1):841.

14. Peddareddigari VG, Wang D, Dubois RN. The tumor microenvironment in colorectal carcinogenesis. Cancer microenvironment : official journal of the International Cancer Microenvironment Society. 2010;3(1):149-66.

15. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome biology. 2016;17(1):174.

16. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature methods. 2015;12(5):453-7.

17. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature communications. 2013;4:2612.

18. Shah N, Wang P, Wongvipat J, Karthaus WR, Abida W, Armenia J, et al. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. eLife. 2017;6.

19. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. Aging (Albany NY). 2018;10(4):592-605.

20. Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, et al. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. Journal of cellular physiology. 2019.

21. Staub E, Groene J, Heinze M, Mennerich D, Roepcke S, Klaman I, et al. An expression module of WIPF1-coexpressed genes identifies patients with favorable prognosis in three tumor types. Journal of molecular medicine (Berlin, Germany). 2009;87(6):633-44.

22. 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 medicine. 2013;10(5):e1001453.

23. Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, et al. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(17):7131-6.

24. Del Rio M, Mollevi C, Bibeau F, Vie N, Selves J, Emile JF, et al. Molecular subtypes of metastatic colorectal cancer are associated with patient response to irinotecan-based therapies. European journal of cancer (Oxford, England : 1990). 2017;76:68-75.

25. Allen WL, Dunne PD, McDade S, Scanlon E, Loughrey M, Coleman H, et al. Transcriptional subtyping and CD8 immunohistochemistry identifies poor prognosis stage II/III colorectal cancer patients who benefit from adjuvant chemotherapy. JCO precision oncology. 2018;2018.

26. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.

27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015;43(7):e47.

28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.

29. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research. 2015;43(Database issue):D447-52.

30. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

31. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer research. 2017;77(21):e108-e10.

32. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nature communications. 2015;6:8971.

33. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2004;10(21):7252-9.

34. Loi S, Drubay D, Adams S, Pruneri G, Francis PA, Lacroix-Triki M, et al. Tumor-Infiltrating Lymphocytes and Prognosis: A Pooled Individual Patient Analysis of Early-Stage Triple-Negative Breast Cancers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2019;37(7):559-69.

35. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nature medicine. 2013;19(11):1423-37.

36. Lim YZ, South AP. Tumour-stroma crosstalk in the development of squamous cell carcinoma. The international journal of biochemistry & cell biology. 2014;53:450-8.

37. Taddei ML, Giannoni E, Comito G, Chiarugi P. Microenvironment and tumor cell plasticity: an easy way out. Cancer letters. 2013;341(1):80-96.

38. Galon J, Angell HK, Bedognetti D, Marincola FM. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. Immunity. 2013;39(1):11-26.

39. Zhang Y, Xu J, Zhang N, Chen M, Wang H, Zhu D. Targeting the tumour immune microenvironment for cancer therapy in human gastrointestinal malignancies. Cancer letters. 2019;458:123-35.

40. Fong W, To KKW. Drug repurposing to overcome resistance to various therapies for colorectal cancer. Cellular and molecular life sciences : CMLS. 2019;76(17):3383-406.

41. Vincent KM, Findlay SD, Postovit LM. Assessing breast cancer cell lines as tumour models by comparison of mRNA expression profiles. Breast cancer research : BCR. 2015;17:114.

42. Huijbers A, Tollenaar RA, v Pelt GW, Zeestraten EC, Dutton S, McConkey CC, et al. The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial. Annals of oncology : official journal of the European Society for Medical Oncology. 2013;24(1):179-85.

43. Danielsen HE, Hveem TS, Domingo E, Pradhan M, Kleppe A, Syvertsen RA, et al. Prognostic markers for colorectal cancer: estimating ploidy and stroma. Annals of oncology : official journal of the European Society for Medical Oncology. 2018;29(3):616-23.

44. Lind GE, Danielsen SA, Ahlquist T, Merok MA, Andresen K, Skotheim RI, et al. Identification of an epigenetic biomarker panel with high sensitivity and specificity for colorectal cancer and adenomas. Molecular cancer. 2011;10:85.

45. Wang H, Luo C, Zhu S, Fang H, Gao Q, Ge S, et al. Serum peptidome profiling for the diagnosis of colorectal cancer: discovery and validation in two independent cohorts. Oncotarget. 2017;8(35):59376-86.

46. Chang CM, Chia VM, Gunter MJ, Zanetti KA, Ryan BM, Goodman JE, et al. Innate immunity gene polymorphisms and the risk of colorectal neoplasia. Carcinogenesis. 2013;34(11):2512-20.

47. Guo J, Jin H, Xi Y, Guo J, Jin Y, Jiang D. The miR-582/CD1B Axis Is Involved in Regulation of Dendritic Cells and Is Associated with Clinical Outcomes in Advanced Lung Adenocarcinoma. Biomed Res Int. 2020;2020:4360930.

48. Lee CH, Chen LC, Yu CC, Lin WH, Lin VC, Huang CY, et al. Prognostic Value of CD1B in Localised Prostate Cancer. Int J Environ Res Public Health. 2019;16(23).

49. Lou Q, Liu R, Yang X, Li W, Huang L, Wei L, et al. miR-448 targets IDO1 and regulates CD8(+) T cell response in human colon cancer. Journal for immunotherapy of cancer. 2019;7(1):210.

50. Takada K, Kohashi K, Shimokawa M, Haro A, Osoegawa A, Tagawa T, et al. Co-expression of IDO1 and PD-L1 in lung squamous cell carcinoma: Potential targets of novel combination therapy. Lung Cancer. 2019;128:26-32.

Tables

Table 1: Correlation of the ten prognostic IRGs with the levels of immune infiltration in CRC.

|		Purity		B Cell		CD8+ T Cell		CD4+ T Cell		Macrophage		Neutrophil		Dendritic Cell	
		Cor	P	Cor	P	Cor	P	Cor	P	Cor	P	Cor	P	Cor	P
A2ML1	COAD	-0.156	0.002	-0.103	0.039	-0.016	0.743	0.136	0.006	0.048	0.339	0.019	0.705	0.025	0.624
	READ	-0.029	0.73	-0.04	0.639	-0.119	0.162	-0.101	0.239	0.014	0.866	-0.005	0.957	0.115	0.179
CALB2	COAD	-0.291	2.26E-09	0.014	0.769	0.109	0.028	0.263	9.11E-08	0.331	8.51E-12	0.284	7.19E-09	0.299	8.76E-10
	READ	-0.334	5.49E-05	-0.081	0.345	-0.002	0.984	0.286	6.37E-04	0.321	1.16E-04	0.204	0.02	0.291	5.13E-04
CD1B	COAD	-0.205	2.99E-05	0.256	1.86E-07	0.264	6.42E-08	0.315	9.91E-11	0.36	8.18E-14	0.468	3.31E-23	0.505	2.05E-27
	READ	-0.289	5.29E-04	0.237	0.005	0.016	0.853	0.29	5.34E-04	0.147	0.085	0.109	0.204	0.408	6.25E-07
COL22A1	COAD	-0.265	5.40E-08	0.022	0.664	0.131	0.008	0.396	1.62E-16	0.473	6.29E-24	0.398	1.06E-16	0.395	1.91E-16
	READ	-0.296	3.84E-04	0.099	0.242	0.0006	0.994	0.411	5.07E-07	0.399	1.16E-06	0.254	0.002	0.299	3.50E-04
FCRL2	COAD	-0.395	1.21E-16	0.484	3.74E-25	0.237	1.35E-06	0.447	3.63E-21	0.297	1.12E-09	0.378	4.34E-15	0.434	6.73E-20
	READ	-0.411	4.66E-07	0.516	8.06E-11	0.232	0.006	0.123	0.15	-0.015	0.861	0.115	0.181	0.304	2.73E-04
HAND1	COAD	-0.176	3.49E-04	-0.103	0.038	0.038	0.446	0.256	2.06E-07	0.377	4.04E-15	0.155	0.002	0.211	1.93E-05
	READ	-0.067	0.433	-0.016	0.856	-0.06	0.479	0.219	0.01	0.317	1.46E-04	-0.019	0.828	0.111	0.191
GPX3	COAD	-0.3	3.83E-10	0.022	0.648	0.177	3.45E-04	0.309	2.51E-10	0.54	5.93E-32	0.402	4.45E-17	0.434	7.01E-20
	READ	-0.317	1.36E-04	-0.068	0.429	0.029	0.732	0.313	1.78E-04	0.412	4.59E-07	0.042	0.621	0.329	7.57E-05
IDO1	COAD	-0.353	2.17E-13	0.248	4.42E-07	0.4	4.66E-17	0.286	5.44E-09	0.303	5.33E-10	0.638	3.97E-47	0.564	3.47E-35
	READ	-0.377	4.46E-06	0.25	0.003	0.342	3.82E-05	0.184	0.031	0.046	0.594	0.364	1.16E-05	0.567	3.38E-13
LAMP5	COAD	-0.315	6.93E-11	0.099	0.047	0.155	0.002	0.447	3.90E-21	0.567	9.59E-36	0.351	4.22E-13	0.451	1.32E-21
	READ	-0.29	5.01E-04	0.184	0.03	0.084	0.325	0.331	6.73E-05	0.482	1.84E-09	0.146	0.087	0.434	9.32E-08
MAP2	COAD	-0.255	1.82E-07	0.125	0.012	0.286	4.59E-09	0.441	1.38E-20	0.535	3.03E-31	0.429	2.22E-19	0.465	5.68E-23
	READ	-0.283	7.15E-04	0.074	0.389	0.311	1.92E-04	0.214	0.012	0.419	2.74E-07	0.234	0.006	0.335	5.54E-05
MMRN1	COAD	-0.395	1.09E-16	0.244	7.27E-07	0.315	8.80E-11	0.482	9.52E-25	0.583	4.47E-38	0.488	2.29E-25	0.533	7.25E-31
	READ	-0.364	9.73E-06	0.274	0.0011	0.313	1.72E-04	0.15	0.078	0.292	4.8E-04	0.256	0.002	0.36	1.37E-05
NKAIN4	COAD	-0.274	1.92E-08	-0.07	0.161	-0.007	0.887	0.301	7.29E-10	0.341	1.90E-12	0.229	3.78E-06	0.273	2.79E-08
	READ	-0.325	9.17E-05	-0.08	0.327	-0.123	0.149	0.389	2.20E-06	0.454	1.93E-08	0.113	0.188	0.197	0.02
VAX2	COAD	-0.239	1.12E-06	-0.016	0.742	0.118	0.017	0.239	1.21E-06	0.352	3.26E-13	0.284	7.00E-09	0.285	5.61E-09
	READ	-0.108	0.203	-0.157	0.07	-0.119	0.163	0.206	0.015	0.139	0.104	0.012	0.89	0.096	0.261

Figures
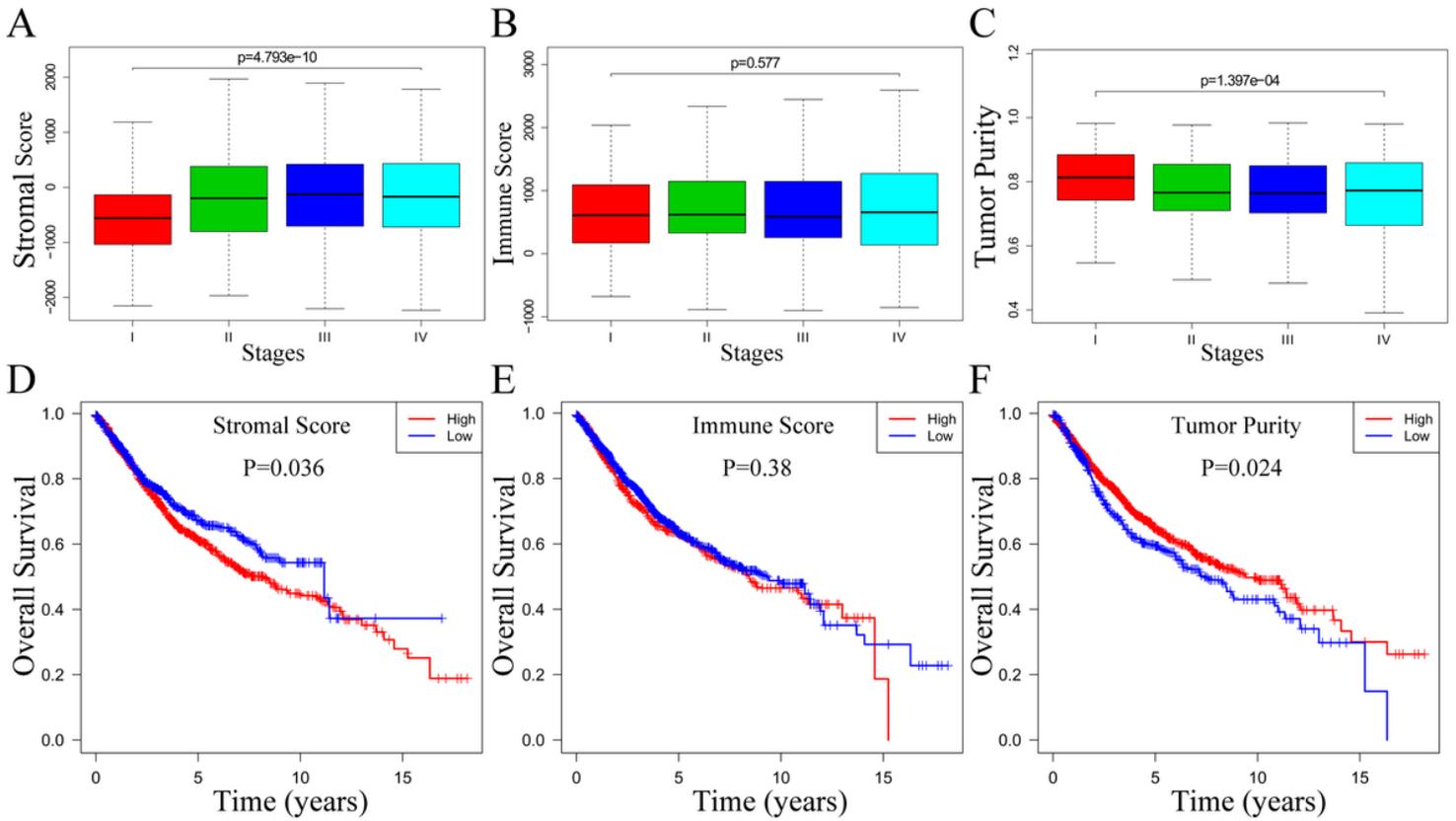
Figure 1

Immune score, stromal score and tumor purity are associated with tumor stages and patients' OS. A, Distribution of stromal scores of tumor stages. B, Distribution of immune scores of tumor stages. C, Distribution of tumor purity of tumor stages. D, Comparison of overall survival between high and low stromal scores groups. E, Comparison of overall survival between high and low immune scores groups. F, Comparison of overall survival between high and low tumor purity groups.
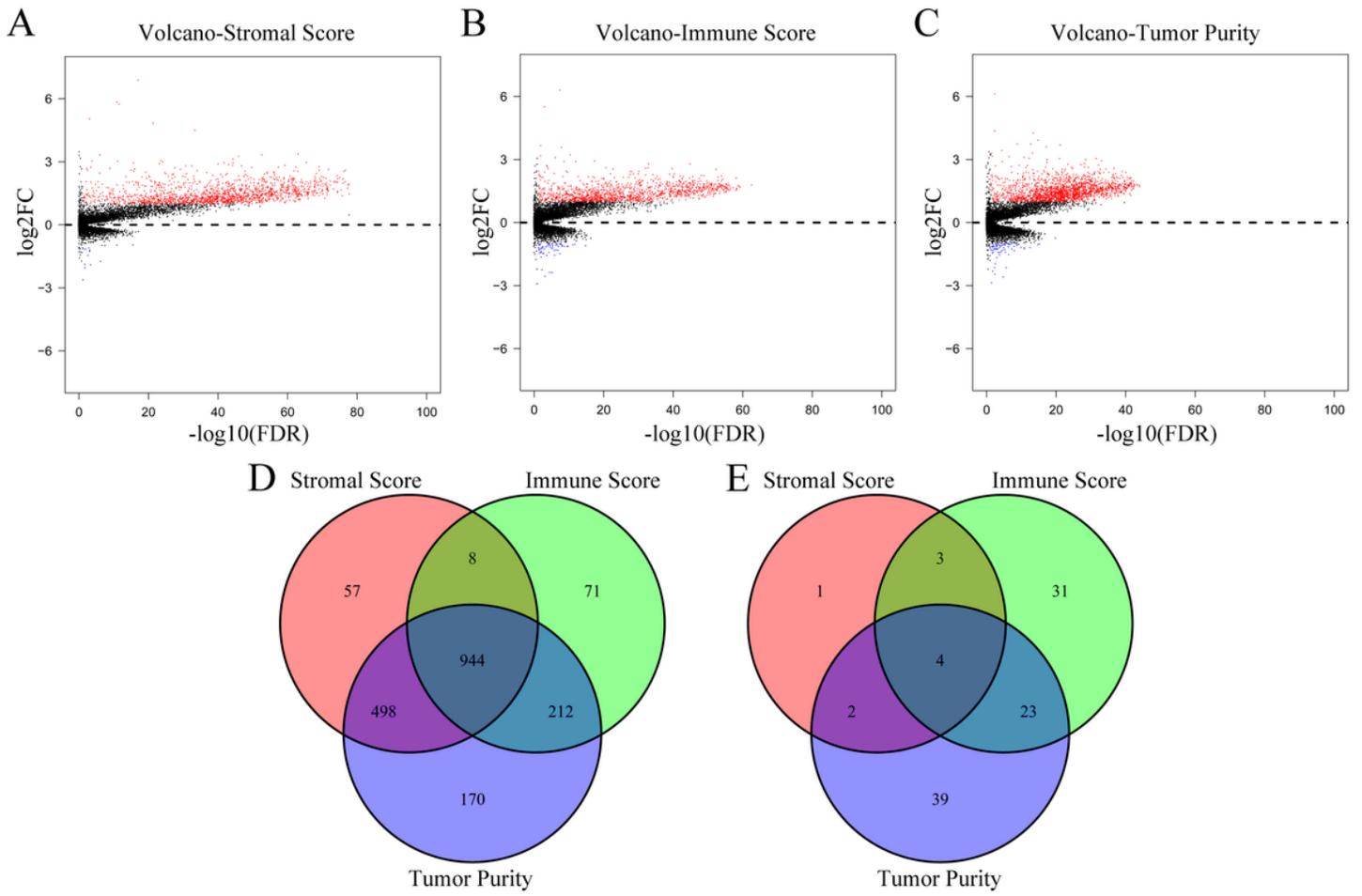
Figure 2

Comparison of gene expression profiles with stromal score, immune score and tumor purity. A, Volcano plots of the DEGs based on stromal score. B, Volcano plots of the DEGs based on immune scores. C, Volcano plots of the DEGs based on tumor purity. D, Venn diagrams shown the number of upregulated DEGs in stromal score, immune score and tumor purity groups. E, Venn diagrams shown the number of downregulated DEGs in stromal score, immune score and tumor purity groups.
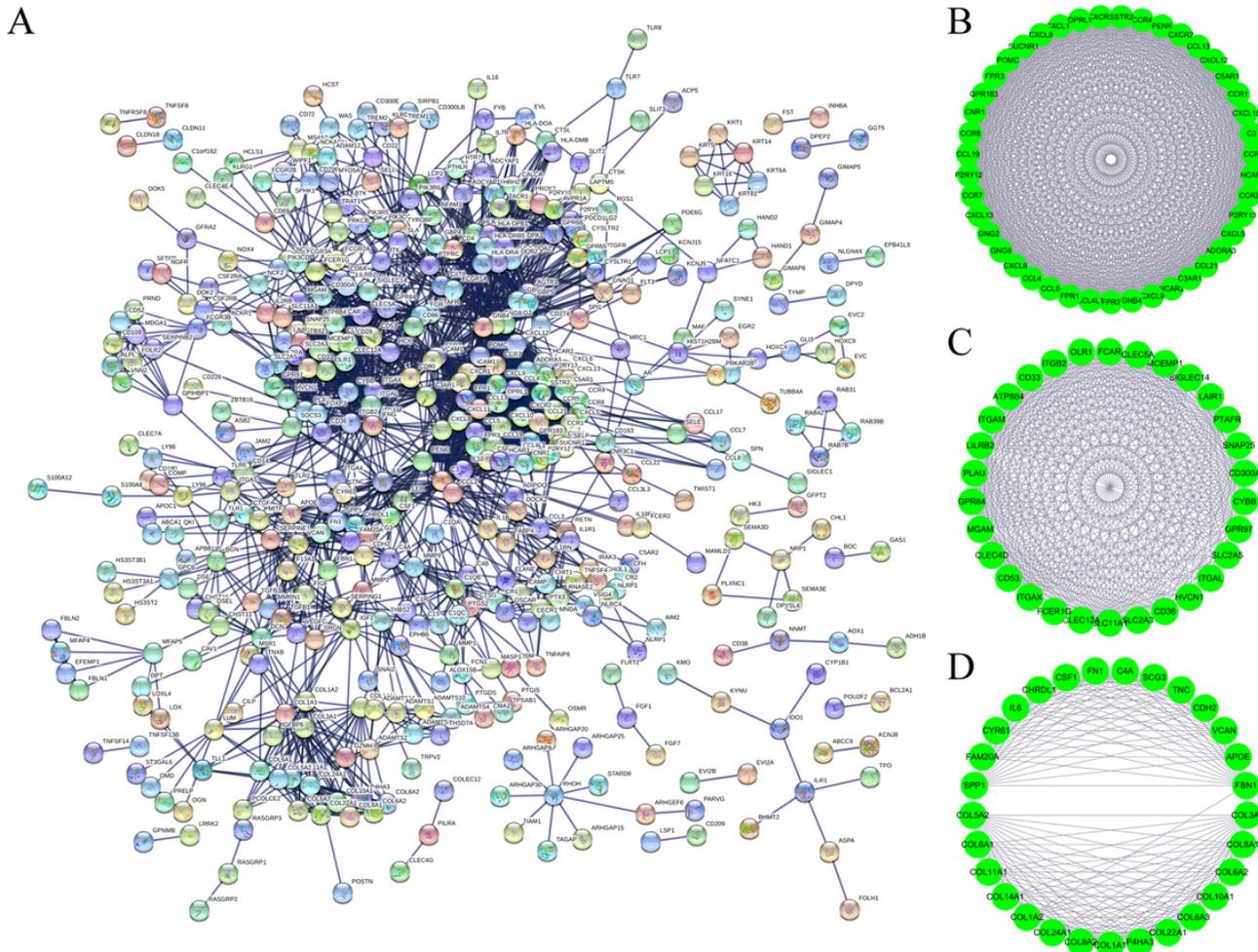
Figure 3

PPI network of immune-related genes obtained from the STRING database. A, PPI network of IRGs. B-D, The PPI networks of top 3 modules.
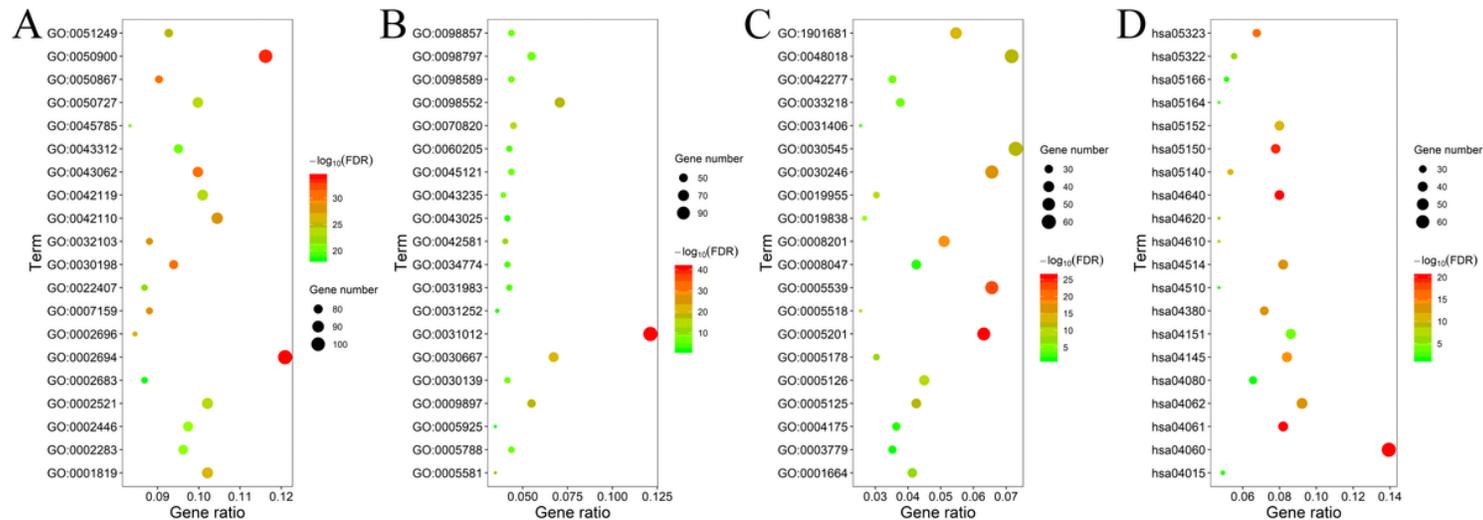
Figure 4

GO enrichment and KEGG pathway functional enrichment analyses of the IRGs. A, The biological function (BF) of GO classification. B, The cell component (CC) of GO classification. C, The molecular function (MF) of GO classification. D, KEGG pathway functional classification and annotation.
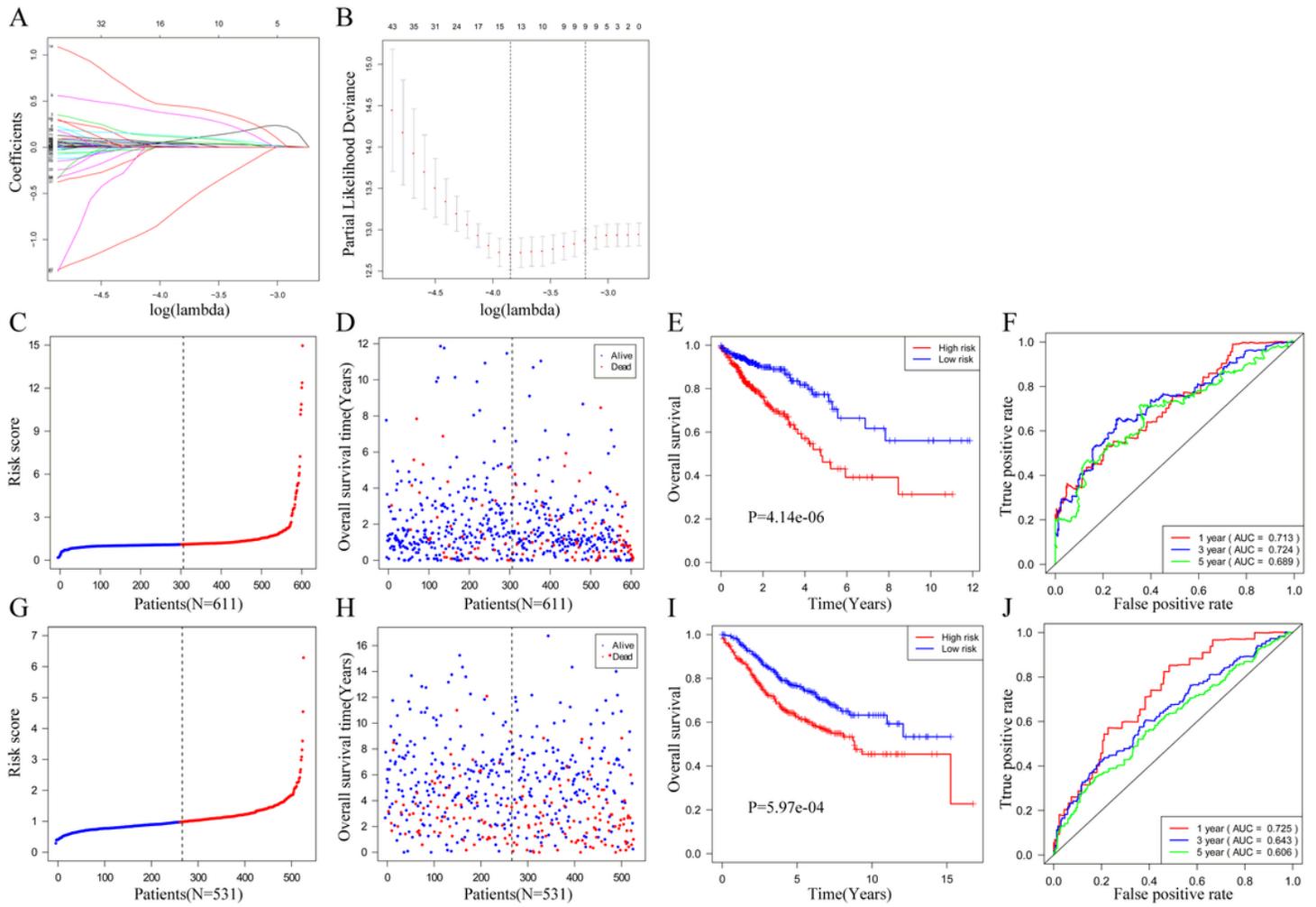
Figure 5

Construction and validation of the thirteen-IRGs signature. A, LASSO coefficient profiles of the IRGs associated with the overall survival of CRC. B, Partial likelihood deviance for the LASSO coefficient profiles. C, The risk score distribution on the basis of the thirteen-IRGs signature in TCGA cohort. D, The survival status of 611 patients with CRC in high- and low-risk groups. E, Kaplan−Meier curves for overall survival prediction by the thirteen-IRGs signature in TCGA cohort. F, Time-dependent ROC curves for overall survival prediction by the thirteen-IRGs signature in TCGA cohort. G, The risk score distribution on the basis of the thirteen-IRGs signature in GSE39582 cohort. H, The survival status of 531 patients with CRC in high- and low-risk groups. I, Kaplan−Meier curves for overall survival prediction by the thirteen-IRGs signature in GSE3958 cohort. J, Time-dependent ROC curves for overall survival prediction by the thirteen-IRGs signature in GSE3958 cohort.
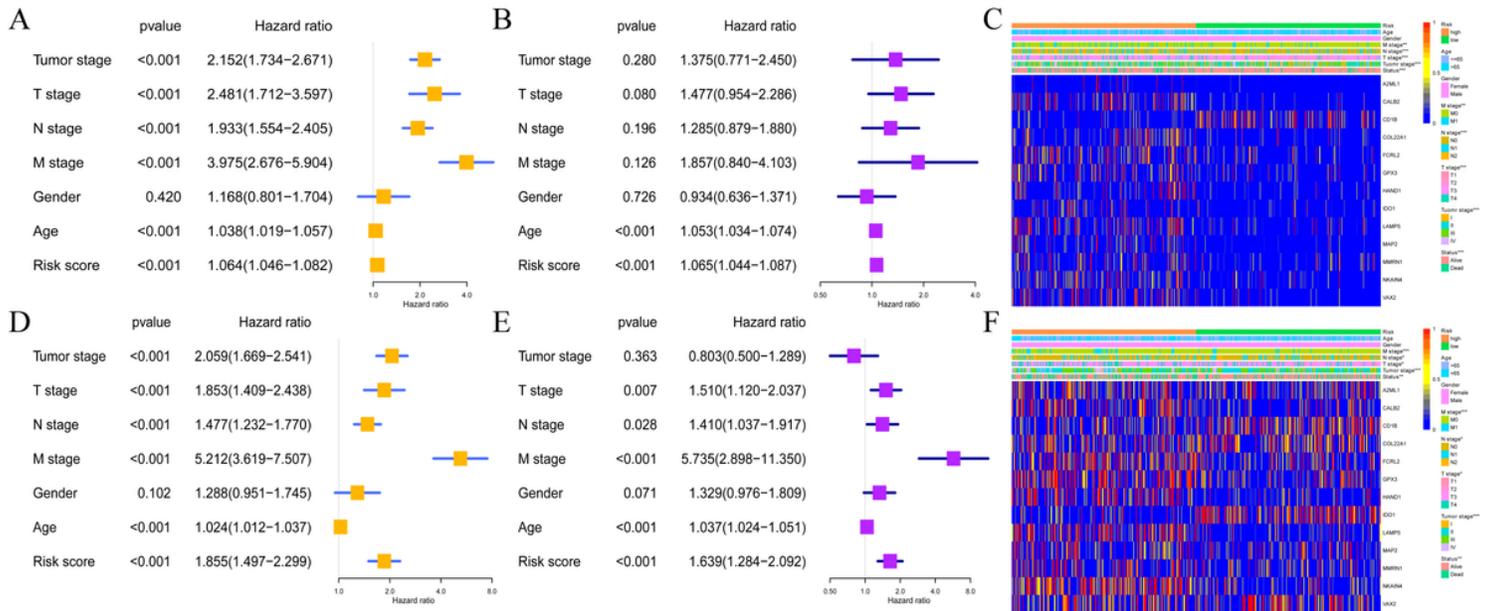
Figure 6

Independent prognostic role of the thirteen-IRGs signature. A, Univariate and analyses of clinicopathological features associated with overall survival in TCGA cohort. B, Multivariate analyses of clinicopathological features associated with overall survival in TCGA cohort. C, Association of the thirteen-IRGs signature with clinical parameters in TCGA cohort. D, Univariate and analyses of clinicopathological features associated with overall survival in GSE39582 cohort. E, Multivariate analyses of clinicopathological features associated with overall survival in GSE39582 cohort. F, Association of the thirteen-IRGs signature with clinical parameters in GSE39582 cohort.
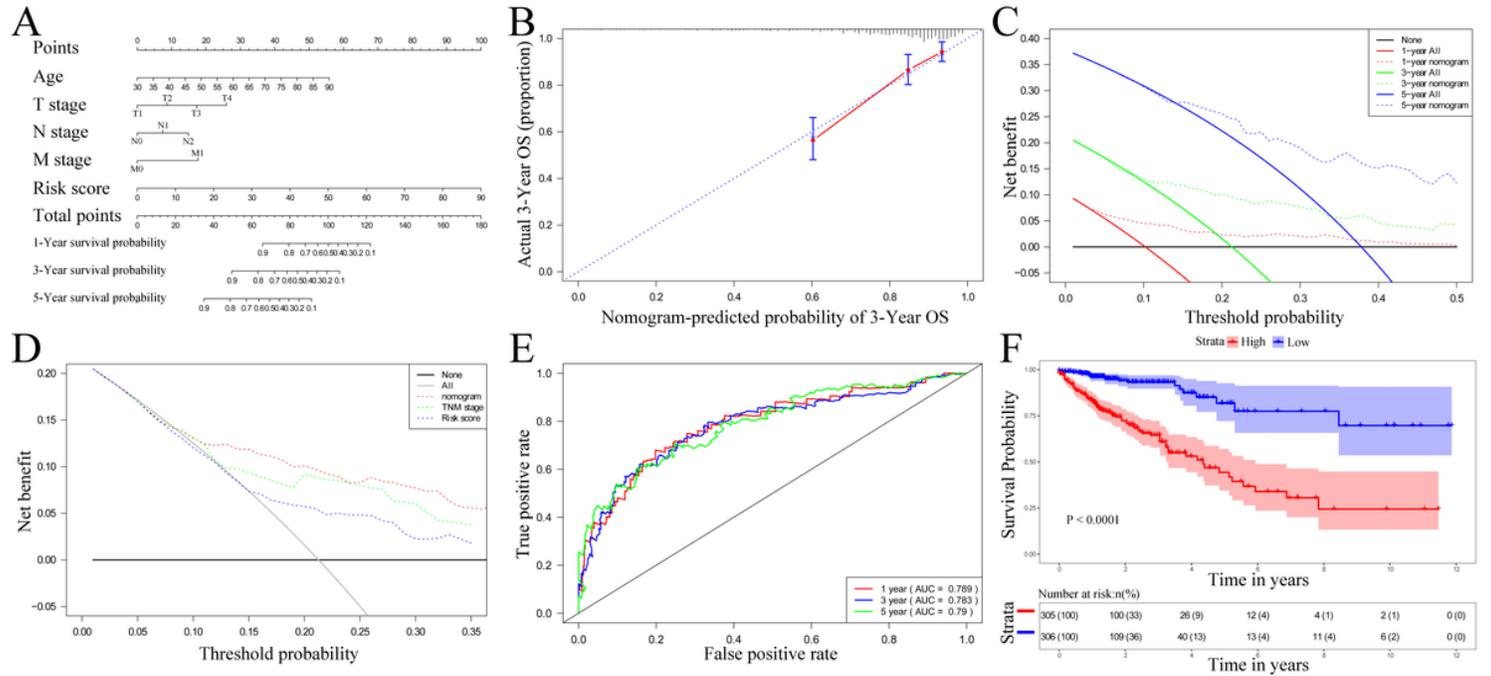

Figure 7

Establishment of the nomogram predicting overall survival for CRC patients in the TCGA cohort. A, Nomogram predicting 1-year, 3-year and 5-year OS for CRC patients. B, Calibration plot for nomogram predicted and observed 3-year overall survival rate. C, Decision curve analysis for 1-year, 3-year and 5-year overall survival predictions. D, Decision curve analysis comparing nomogram with the TNM stage and the thirteen-IRGs signature. E, Time-dependent ROC curves for overall survival prediction by the nomogram in TCGA cohort. F, Kaplan−Meier curves for overall survival prediction by the nomogram in TCGA cohort.
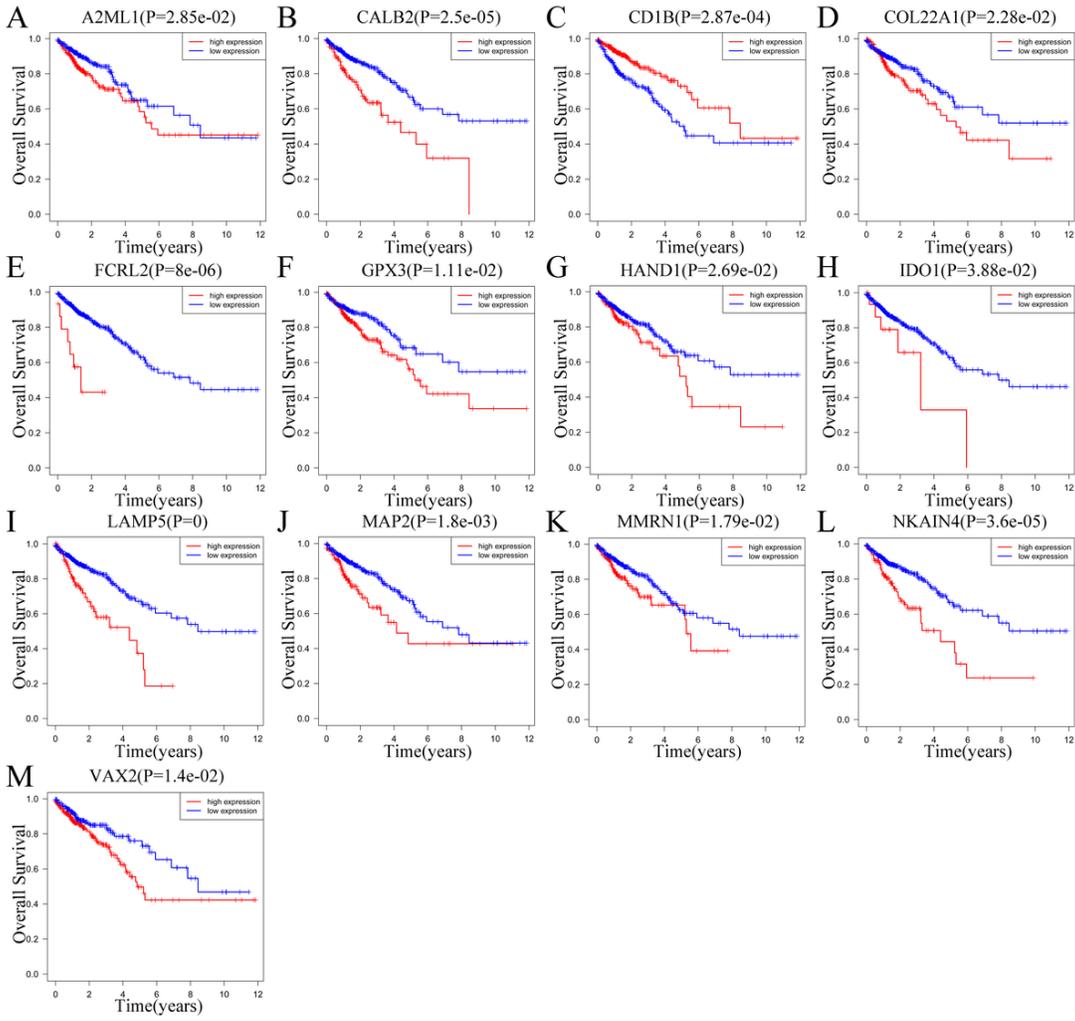
Figure 8

Correlation of expression of those thirteen IRGs in overall survival in TCGA. Kaplan-Meier survival curves were generated for selected immune-related genes extracted from the comparison of groups of high (red line) and low (blue line) gene expression. p<0.05 in Log-rank test.
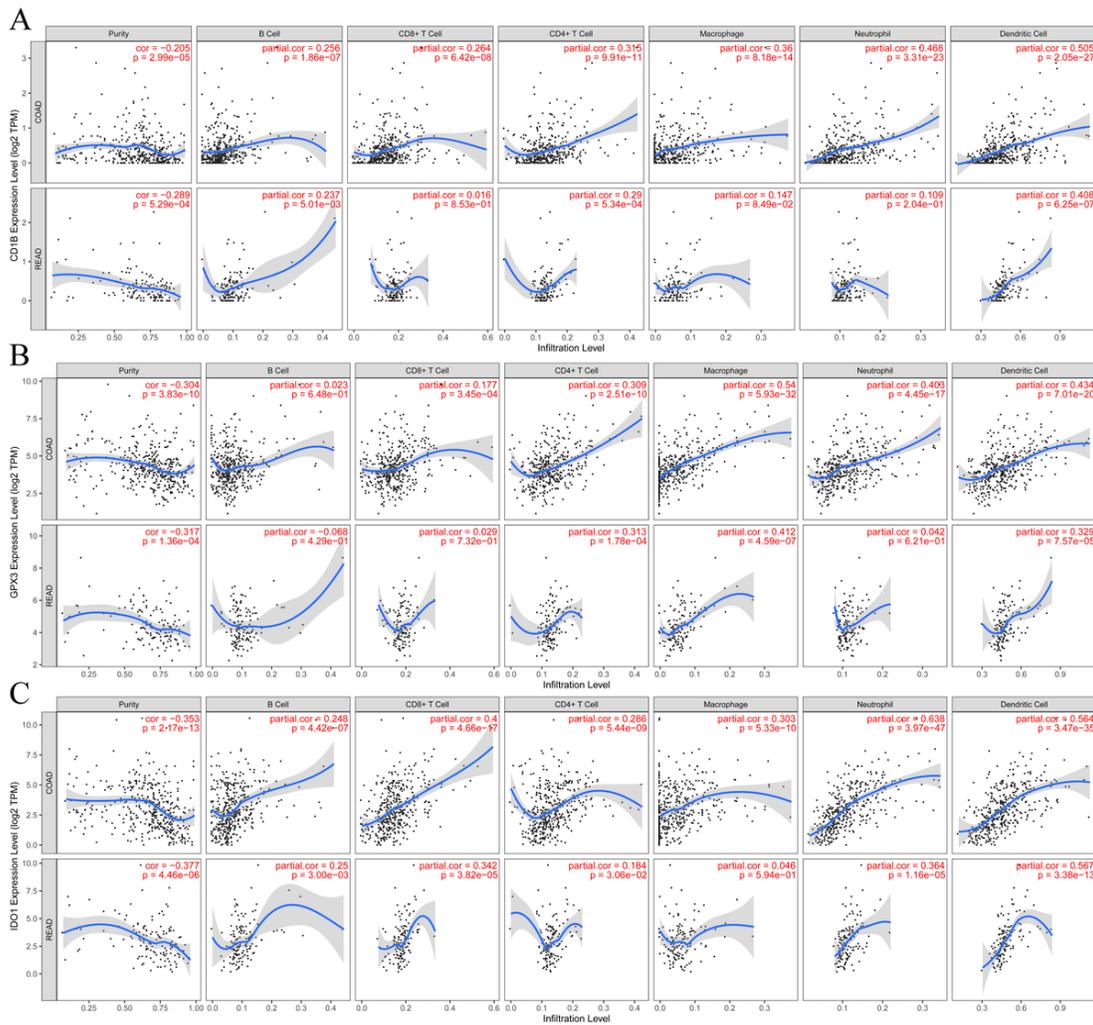
Figure 9

Integrative analysis between the IRGs with tumor infiltration immune cells in CRC. A, The correlation between CD1B gene expression and immune cells. B, The correlation between GPX3 gene expression and immune cells. C, The correlation between IDO1 gene expression and immune cells.

Supplementary Files

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

- FigureS1.tif
- FigureS2.tif
- TableS1.docx