Identification of an immune-related gene signature to improve prognosis prediction of colorectal cancer patients

Siqi Dai
the second affiliated hospital of Zhejiang University school of medicine https://orcid.org/0000-0002-5729-6768

Shuang Xu
Peking University People's Hospital

Yao Ye
Zhejiang University School of Medicine Second Affiliated Hospital

Kefeng Ding (dingkefeng@zju.edu.cn)
Zhejiang University School of Medicine Second Affiliated Hospital

Primary research

Keywords: Colorectal cancer, Immune-related gene, Prognostic, Survival, Nomogram

Posted Date: August 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-56380/v1

Abstract

Background: Despite recent advance in immune therapy, great heterogeneity exists in colorectal cancer (CRC) patients' outcome. In this study, we aimed to analyze the Immune-related gene (IRG) expression profiles from two independent public databases and develop an effective signature to forecast patients' prognosis.

Method: IRGs were collected from The Cancer Genome Atlas (TCGA) database to identify a prognostic genes signature, which was verified in Gene Expression Omnibus (GEO) database. Gene function enrichment and immune cell infiltration analyses were conducted. A prognostic nomogram was built incorporating the IRG signature with clinical risk factors.

Results: The training and test datasets had 487 and 579 patients, respectively. A prognostic six-gene-signature (CCL22, LIMK1, MAPKAPK3, FLOT1, GPRC5B and IL20RB) was developed through feature selection, and yielded incredible differentiation between low and high-risk group in the training set ($P < 0.001$), which was confirmed in the test group ($P < 0.001$). The signature outperformed tumor staging regarding survival prediction. Go and KEGG functional annotation analysis suggested the signature were significantly enriched in metabolic process and regulation of immunity ($P<0.05$). When combined with clinical risk factors, the model showed robust prediction capability.

Conclusion: The immune-related six-gene signature is a reliable prognostic indicator for CRC patients and could bring insight for personalized cancer management.

Introduction

Colorectal cancer (CRC) is the third most common malignant tumor across the globe and ranks the second in tumor-related death [1]. In China, CRC is in the third place of annual incidence and is the fifth leading cause of death among cancer [2]. For operable disease, resection offers the best shot of long-term survival and potential cure [3]. While for inoperable patients, chemotherapy (mostly 5-fluorouracil or oxaliplatin based) and target therapy (epidermal growth factor receptor or vascular endothelial growth factor-targeted) have been the standard of care [4]. However, despite recent advancements in chemo-regiments and clinical managements, the overall survival of CRC remains unsatisfying: the five-year overall survival is merely over a half [5]. More discerningly, there lays great heterogeneity in individuals not only regarding tumor development but also in the response to uniformed treatment: in those receiving surgeries, while some could enjoy disease-free survival, many suffered from tumor recurrence [6]. Same is seen during non-surgical managements where tumors reactions vary stunningly: less than 60% of patients had objective treatment response [7], and adverse tumor response remains a strong predictor for unfavorable survival [8]. Thus, identifying reliable biomarkers for prediction of tumor behavior and outcome will benefit personalized modification in clinical managements.

Recently, immune checkpoint blockade therapies which provide revolutionary treatments in multiple solid tumors (melanoma, non-small cell lung cancer, head-and-neck squamous cancer, colorectal cancer etc.), has brought the community's scope to tumor-related immunology [9,10]. It's being increasingly recognized that immune condition plays a decisive role in the genesis and progressiveness of malignant tumors. The host's immune dysfunction significantly impairs body's anti-tumor surveillance, along with ill-transformed cells' immune-avoiding mechanism acquired from the accumulation of gene mutation, is a vital step towards tumor development [11-13]. The most wildly recognized prognostic biomarkers for immune therapy were programmed death ligand (PD-1), tumor mutation burden (TMB) and microsatellite instability/ mismatch repair deficiency (dMMR) [14]. However, throughout research, these solitary biomarkers only showed moderate stratification efficacy [15-18], more so in CRC [19], and a universal immune-related gene (IRG) panel as prognostic signature in CRC has not been scored

Within the decade, several limited-scale studies have made attempts developing predictive gene signature to stratify high-risk population with high-throughput technology [20-22]. However, most suffer from over-fitting due to insufficient sample pool, and an external validation is hardly presented. Besides, difference among high-throughput protocols often leads to inconsistency in expression values among studies, making it challenging for comprehensive meta-data analysis. From this perspective, the large-scale publicly available genomic databases provide sufficient among of samples, comparable gene expression at probe level and solid follow-up information, and thus were recognized as ideal platforms for gene-signature construction and validation.

In this study, we intended to identify and validate an IRG signature to stratify significantly ill-survival CRC patients in two independent public databases. The signature was then incorporated with clinical risk factors to provide robust prediction efficacy regarding long-term survival.

Material And Method

The schematic of this study is in Fig 1.

[insert Figure 1 here]

Gene expression data acquisition of CRC patients

Two sets of colorectal cancer patients with complete survival information including survival status and survival time were retrospectively enrolled from the publicly-available The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/; https://xenabrowser.net/datapages/) and Gene Expression Omnibus (GEO, GSE30219) data pool as training and external validation set, respectively. Clinical variables and gene expression level were comprehensively extracted for further procedures. To exclude low-quality data, cases with undetermined survival information and missing genes expression values in > 20% samples were removed.

Selection of immune-related genes

To filter genes that actively participate in immune activity, a third comprehensive data set of immune genes was acquired from the the Immunology Database and Analysis Portal database (ImmPort, https://immport.niaid.nih.gov). After cross-reference with ImmPort database, a pack of 2112 IRGs was identified in the two above-mentioned sets. Duplicates, genes showing no expression value and that expressed in less than 70 presents in all samples were subsequently removed. As a result, a panel of 1684 IRGs was identified for survival analysis (Supplemental Table S1).

Development of prognostic gene models in training set

Univariate COX regression was performed for each gene regarding survival status to screen out prognostic immune-related genes. For those showing statistical significance ($p<0.05$) in COX regression, random survival forest algorithm (RSFVH) was adopted for dimensional reduction. Further, the risk score of the prognostic models were determined as follows

$$\text{Risk Score} = \sum_{i=1}^{N} (Expression_i * coefficient_i)$$

where N is the number of gene, Expression is the gene expression value and coefficient is the gene coefficient value in Cox regression analysis, while the median risk score was utilized to group the patients as Low-Risk and High-Risk population.

To rule out over-fitting, we constructed full-scale combinations of genes yielded in RSFVH. Time-dependent Receiver operating characteristic (time-ROC) analysis was used to test performance. The C-index, which, by value, equals to area under curve (AUC), was used to evaluate the concordance between prediction model and reality. The combination with the highest C-index was designated as the optimal prediction model, which was subsequently verified for performance in internal and external verification (GEO database).

Function analysis of included IRGs

Interpretation of tiled IRGs was fuzzy, and could not unveil the nature of the discovery. The co-expressing genes (CEGs) of the selected IRGs was then screened out using online library. The correlation factor was 0.6, as was consistent in other literature. To explore the high dimensional links among included IRGs, gene enrichment test, namely gene ontology (GO) analysis and KEGG analysis were executed using their co-expressing genes. Using GO method, the functions of IRGs in cellular component (CC), molecular function (MF) and biological process (BP) were systematically analyzed. The underlying overlap among IRG-related pathways was verified by KEGG analysis. Then, we used protein-protein interaction (PPI) network to discover the role of IRGs on a protein level. In addition, we conducted digital cytometry of tumor infiltrating immune cells (TIIC) between high and low-risk group with CIBERSORT [23] (http://cibersort.stanford.edu/), an web tool that analysis 22 subsets of leukocytes representation in bulk tumor.

Construction and assessment of a novel nomogram incorporating IRGs and clinical factors

Via univariate and multivariate COX regression, independent clinical risk factors ($p < 0.05$) were incorporated with the IRG panel. Based on these, a comprehensive prediction nomogram was formulated. Subsequently, we used time-ROC test at different time-points to test its performance in the training and validation group. Also, the nomogram's prediction bias was evaluated via internal and external calibration.

Statistical analysis

The statistical analysis was performed with R Software (version 3.6.2, www.r-project.org) while packages of pROC, TimeROC, randomForestSRC and survival were utilized. Data distribution was validated by Kolmogorov-Smirnov test. Continuous variables' statistical significance between the training and validation set was measured by student's t-test. While Pearson's Chi square or rank-sum test was performed for layered variables. Kaplan–Meier analysis was used to assess the high-and-low-risk groups. Z-test was adopted for statistical difference between ROC curves. Function prediction of prognostic genes was analyzed by Clugo, which is a Cytoscape plug-in to perform GO and KEGG analysis. TIIC analysis was performed via CIBERSORT as previously described.

Results

Identifying the prognostic signature in the training set

Following the aforementioned inclusion criteria, a total of 487 and 579 patients with CRC was enrolled as the training (TCGA) and validation (GEO) set, respectively. The clinical characteristics was presented in Table 1. The median age of training set was 68 years. At the time of enrollment, most patients were alive (77.8% in the training and 66.6% in the validation set), the median surveillance time was 699 and 1582 days in the training and validation wing, respectively. Most patients don't have lymph node involvement (stage I to II). Most patients have tumor infiltration to the subserosa and serosa level (80.5% and 85.6%, respectively).

[insert Table 1 here]

After cross-comparing with ImmPort database, 2112 immune-related gene were selected. Then, repeated, not or inconsistently expressed genes were excluded, leaving 1684 genes as candidates. For each gene, univariate COX regression was performed and 143 suggested significant protective or risk effect (Fig 2a & Table S1). Via RSFVH, nine immune-related genes stood out as independent prognostic predictors (Fig 2b).

Next, to explore the best predictive signature and to preclude over-fitting, we formed a penal of full-size combinations of these nine genes (2^9-1=511 combinations, Table S2). Using the risk formula previous discussed, 511 possible predictive signatures were calculated. Each signature's performance was verified via time-ROC curve. The AUC value of each combination was rated. A combination of the following gene CCL22, LIMK1, MAPKAPK3, FLOT1, GPRC5B, IL20RB was screened out with the highest prediction

precision (AUC = 0.746, Fig 2c and d). Their hazard ratio (HR) and p value is listed in Table 2. Thus, the designated risk model was: Risk score = (-0.421× expression value of CCL22) + (0.402 × expression value of LIMK1) + (-0.465× expression value of MAPKAPK3) + (0.599× expression value of FLOT1) +(0.613× expression value of GPRC5B) +(0.596× expression value of IL20RB).

[insert Figure 2 here]

[insert Table 2 here]

The validation of performance in predicting overall survival

Using the algorithm, a risk score was calculated for individuals. Kaplan-Meier (KM) test was performed to verify survival difference between high (n=243) and low risk (n=244) population divided by median risk-score-value, the method was consistent with other studies[24, 25]. As is demonstrated in Fig 3a, significant longevity was observed in low-risk patients in training wing ($p<0.001$). The median survival time was 8.46 years for low-risk patients versus 4.12 years in high-risk population. To explore this in another independent population, the same methodology was then adopted in the external validation set with relatively larger sample pool (Fig 3b). Again, the model showed significant differentiation capability ($p<0.001$).

When the training and validation array were sorted by risk score, clusters in gene expression level and survival information could be observed in Figure 3c and d, respectively. Indeed, genes with adverse prognostic effect, namely LIMK1, FLOT1, GPRC5B and IL20RB, showed consistent elevation in expression in high-risk population.

[insert Figure 3 here]

The relationship between the signature and clinical characteristics

We further explored the potential relationship between gene signature and clinical characteristics in TCGA and GEO database (Table 3). Neither patients' age (stratified by 68 years) nor gender showed co-relation with gene signature via Pearson's Chi square. Rather, tumors' TNM staging was significantly advanced in high-risk population ($p<0.001$). In univariate COX regression, old age (> 68), more advanced tumor stage (stage III and IV) and immune-related gene signature revealed statistical significance and was confirmed as independent adverse predictor via the following multivariate COX regression. In both training and validation pool, the gene signature suggested great predictive potential regarding patients' clinical outcome. (High- vs. Low-risk, HR training =4.56 95% CI 2.81-7.40, $P < 0.001$, n=192; HR test = 4.53, 95% CI 2.79–7.34, $P < 0.001$, n=193, Table 4).

[insert Table 3 here]

[insert Table 4 here]

Comparison of predictive performance with other clinical variables

The model's performance was competed against tumor TNM staging in predicting clinical outcome. To this end, time-ROC curves in the training and validation database were constructed to compare both models (Fig S1). In the training set (Fig S1a), the C-index was 0.746 (95CI: 0.695-0.796). While in the test set (Fig S1b), the C-index was 0.622 (95CI: 0.574-0.670). Tumor staging, which is most used for clinical prognosis expectancy, was borrowed as contrast. As was indicated in Supplement Fig S1, in both wings, the six-gene signature yielded superior accuracy against traditional staging.

Exploring the high-dimensional functions of IRGs

To bring new perspective into the biological importance of the obtained IRG panel, we conducted GO analyses based on A total of 355 relative CEGs which were selected using online library. Via GO analyses (Fig 4a), the CEGs were most significantly involved in metabolic process and immune regulations (GO terms, such as multiple cellular components biogenesis and processing, leukocyte cell-cell adhesion, regulation of T cell proliferation and interleukin). The findings were later mirrored in KEGG analysis, where the CEGs were actively related to "Primary immunodeficiency". We then explored the CEGs on the level of proteins using PPI analysis (Fig 4b) for potential treatment targets. From the major continent of the network, we could tell that a good number of proteins (like UBA52, NHP2 and FBL) played as junctions (hubs).

[insert Figure 4 here]

Additionally, we conducted leukocyte subsets analysis using CIBERSORT. The patients were sorted by the risk-value and the detailed proportion of twenty-two immune cell in each patient was achieved. The map of subsets is in Supplement Fig S2 and the statistical significance of TIIC in the two groups were given in Supplement table 3. Unfortunately, no significantly variated TIIC subset was discovered.

Development of a predictive gene-clinical nomogram for clinical outcome

To facilitate comprehensive outcome prediction, the six-gene prediction model was combined with clinical independent risk factors, namely tumor stage and age, and transformed into a predictive nomogram (Fig 5a) to provide a straightforward estimation of survival at 1,3,5-year. For instance, an old-aged (> 68 years) advanced-staged (stage III to IV) patient with a gene-signature value of four would have a total risk score of roughly 60, and the odds of survival would be 80, 55 and 35 percent. Via time-ROC (Fig 5b), the AUC value in the training group at 1,3 and 5 years were 0.822 (95CI: 0.761-0.883), 0.835 (95CI: 0.775-0.895) and 0.798 (95CI: 0.715-0.881), respectively. The nomogram was tested for performance in the GRO database, which yielded overall comparable precision: the AUC at 1,3 and 5 years were 0.707 (95CI: 0.622-0.792), 0.692 (95CI: 0.641-0.744) and 0.681 (95CI: 0.628-0.733) (Fig 5c), respectively. It could be judged from the nomogram that the six-gene signature was the most prominent factor affecting patients' survival and the performance was consistent over various time points.

To access how this nomogram mimics real situation, calibration curves using 1000-time bootstrap test were plotted. As is shown in Fig 5d, in the training data set, the nomogram presents excellent calibration. What's more, in the external calibration (Fig 5e), the calibration curve showed minor wobble, but still in the near proximity of the 45-degree-dashed line. These suggested that our nomogram close mimics real-life situation, and via calibration in two different large-scale databases, the nomogram showed great utility.

[insert Figure 5 here]

Discussion

Resent advancements in immune checkpoint blockade therapy warrants further understanding in immune gene variation, and there are imminent needs for robust prognostic biomarkers to guide selective management strategy. In this study, we used two independent, large-scale international genome databases for the exploration and validation of a prognostic IRG panel. Through dimension-reduction of acquired IRG data and ruling out over-fitting, which is often the case in other researches, we developed a full-scale recombination of the nine figures and the most accurate six-gene prediction signature was achieved. The signature alone showed improved prognostic performance comparing to tumor stage (C-index 0.746 and 0.622, against 0.704 and 0.609 with tumor stage in the training and validation group). When the IRG signature was incorporated with independent clinical risk factors, the model unleashed outstanding performance. Calibration curve also showed good agreement among model prediction and reality in the training set. Compared to traditional clinical risk scoring, incorporating our IRG signature with clinical risk factors would benefit prognostic prediction. For those anticipated to have significantly ill survival, a more close-up surveillance strategy should be made to spot early onset of tumor recurrence after resection or tumor progression during non-surgical intervention. Also, surgeons would be more informed when making treatment decision.

There were a few freshly published studies using IRGs signature to predict prognosis in CRC patients within the past 3 years. However, not all was conducted in a reasonable sample size and only moderate performance were achieved. Zuo et al. [26] developed a 6-gene signature model forecasting patients' prognosis without an external validation. In their study, the gene selection was based on multivariate regression and no re-combination was performed to rule out overfitting. Indeed, the AUC was only 0.711 and 0.683 for the 3-year and 5-year survival, and inconsistency was seen in subgroup analysis. Bai et al. [27] also reported a 14-IRG panel using TCGA cohort in the absence of external validation. Besides, they conducted GO and KEGG analysis not with CEG of the selected IRG but instead with the CEG of the whole 676 IRGs. From this perspective, the accuracy of the analysis could be biased. Our proposed link between IRG signature and disease characteristics were further confirmed by Wang et al. earlier this year [28]. Regretfully, none of these studies incorporated IRG signature with clinical risk factors for outcome prediction so their clinical utilities were largely limited.

The included genes for signature were: CCL22, LIMK1, MAPKAPK3, FLOT1, GPRC5B, IL20RB. Through literature research, CCL22 was identified as up-stream regulator of PI3K/AKT pathway. Secreted by M2

macrophages, CCL22 regulates CRC cells' epithelial-mesenchymal transition (EMT), and promotes tumor's resistance to chemotherapy [29, 30]. FLOT1 also induces EMT and alters cell cycle by modulating Erk/Akt signaling axis [31]. Besides, the prognostic value of IL20RB was actively discussed in multiple tumors including glaucoma, anal cancer and lung adenocarcinoma [32-34]. Moreover, MAPKAPK3 is a member of stress-responsive kinase that induces autophagy in terms of stress (inflammation, infection and starvation)and thus determining cell fate[35, 36].

However, linearly discussion of the selected gene is plane and could not represent the high-dimensional connections and interactions among them. We introduced 355 CEGs of the enrolled IRGs and conducted comprehensive interpretation of these genes on the scale of cellular functions, pathway enrichments and protein-protein interaction network. Beside metabolic functions, we discovered that these genes were excessively involved in immune regulations. Immune cell adhesion, lymphocyte proliferation regulation and interleukin regulation were the most enriched function of CEGs. Cell-cell adhesion, mediated by adhesion molecules, is a classic foundation of intercellular signal transduction[37]. These adhesion molecules including receptors of the immunoglobulin superfamily and integrins regulate a variety of immune cell functions including immune cell migration into tumor microenvironment, immune function activation and the formation of immune cell-immune cell or immune cell-target cell synapse and thus is a hallmark in cancer progression [38-40]. Interleukin (IL) was secreted by a wide range of immune, cancerous or endocrine cells, belong to a sophisticated superfamily that's closely associated with immune cell proliferation, hemostasis regulation, up-or-down regulation of cell activity and expansions during cancerous and inflammation situations [41-43]. IL-2 was reported to trigger synergetic cancer-specific immune reaction at high dose [44], while IL-10 known as the tolerogenic cytokine which inhibit pro-inflammatory cytokine production and suppress the stimulation of T cells [45, 46]. On the other hand, as part of self-regulating mechanism, immune cells (eg. B cell and CD8+ T cells) actively regulate interleukins during the intercell crosstalk, thus achieving modulation of remote immune cells [43]. Via KEGG analysis, the most enriched pathway lead to immunodeficiency, which indeed is the center of cancer genesis and development. Additional PPI analysis was performed to investigate the genes on a proteinic level. Using PPI network, it could be interpreted that hubs like UBA52 and NHP2 serve as critical regulators of CRC development, and thus could be potential targets for cancer therapy. Previous reports have suggested that UBA52 can modulate ubiquitin and 60S ribosomal protein L40 (RPL40), which regulates p53 expression, thus promoting CRC genesis [47]. The cancer genesis effect of UBA52 was also discussed in non-small-cell lung cancer where UBA52 boosts cell cycle progression and proliferation [48]. On the other hand, NHP2 was a cofactor located on the telomeres at the end of the chromosomes which in turn regulate cell aging [49]. The malfunction of NHP2, thus, was closely linked to pathological aging and predisposition of cancer [49, 50]. Besides, fibrillarin (FBL) is a key influencer of ribosome's biogenesis and subsequently induces inherent or acquired chemo-resistance [51]. These findings provide insight for potential targets in the revolutionary era of immune therapy. However, further works are warranted to convey these prognostic influencers into feasible treatment markers. Regretfully, we failed to identify significantly varied TIIC subsets using CIBERSORT. As a virtual analyzing tool, CIBERSORT gives TIIC

proportions based on gene sequencing data [23, 52] and could potentially be affected by sequencing protocols.

Our study also had several limitations, first, the gene levels in different cohorts were not measured via universal sequencing protocols, which could lead to inconsistency, and the minor drift of calibration regarding validation group, to some end, might answer for the slightly decrease in C-index in test group. It's important to recognize that microarray protocols among databases was not consistent and that different ethnic and geographical variation could result in reasonable inter-cohort bias. Secondly, contrast to the volume of gene data, the clinical information in these databases were largely limited, it's best to combine the gene signature with more comprehensive clinical factors for optimal prognostic prediction.

Conclusion

Taken together, we developed a predictive IRG panel that can legitimately forecast CRC patients' outcome, the gene signature was more robust when incorporated with clinical risk factors. Our model could potentially benefit individualized clinical management for CRC patients. For instance, a shorter check-up interval should be considered for patients with adverse survival as timely medical intervention would be ideal regarding tumor progression or recurrence.

Abbreviations

TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; IRG, Immune-related gene；RSFVH, random survival forest algorithm; GO, gene ontology; PPI, protein-protein interaction；TIIC, tumor infiltrating immune cells；KM, Kaplan-Meier; ROC, Receiver operating characteristic; AUC, area under curve; PD-1, programmed death ligand, TMB, tumor mutation burden; dMMR, microsatellite instability/ mismatch repair deficiency; CEG, co-expressing gene; CC, cellular component, MF, molecular function; BP, biological process; IL; Interleukin; FBL, fibrillarin;

References

Ethics approval and consent to participate

All relative information was gathered from the publicly available international databases namely The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database. All process performed in this study involving medical information were in accordance with ethical standards of the Second Affiliated Hospital of Zhejiang University School of Medicine and was permitted by the ethical committee (2019-LYS-#394).

Consent for publication

The submission of this article is approved by all authors.

Declarations of interest:

none

Data Accessibility

The data supporting the results of this study are available from the corresponding author on reasonable request.

Acknowledgement

The authors acknowledge the phenomenal contributions of large open-access genomic and clinical database including The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/; https://xenabrowser.net/datapages/) and Gene Expression Omnibus (GEO, GSE30219) who allowed rapid progression in cancer research.

Fundings

This work was supported by National Natural Science Foundation of China (81672916, 11932017, 81802750), the Key Technology Research and Development Program of Zhejiang Province (No.2017C03017), Natural Science Foundation of Zhejiang Province (LQ20H180014 to Y Ye), China Postdoctoral Science Foundation (2019M652117 to Y Ye).

Authors' contributions

The data gathering was the joint effort of Siqi Dai and Shuang Xu. The bioinformatic analysis was done by Siqi Dai and Yao Ye. The construction of this article was accomplished by Siqi Dai and Shuang Xu. The study was conducted under the supervision of Prof. Ding as corresponding author.

Reference

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: a cancer journal for clinicians 2018;68:394-424.
2. Chen WQ, Li H, Sun KX, Zheng RS, Zhang SW, Zeng HM, Zou XN, Gu XY, He J. [Report of Cancer Incidence and Mortality in China, 2014]. Zhonghua zhong liu za zhi [Chinese journal of oncology] 2018;40:5-13.
3. Adam R, Delvart V, Pascal G, Valeanu A, Castaing D, Azoulay D, Giacchetti S, Paule B, Kunstlinger F, Ghémard O, Levi F, Bismuth H. Rescue surgery for unresectable colorectal liver metastases downstaged by chemotherapy: a model to predict long-term survival. Annals of surgery 2004;240:644-57; discussion 57-8.
4. YH Xie, YX Chen, JY Fang. Comprehensive review of targeted therapy for colorectal cancer. Signal transduction and targeted therapy 2020;5:22.

5. M. Frampton, R. S. Houlston. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. Genet Med 2017;19:314-21.
6. Stelzner S, Radulova-Mauersberger O, Zschuppe E, Kittner T, Abolmaali N, Puffer E, Zimmer J, Witzigmann H. Prognosis in patients with synchronous colorectal cancer metastases after complete resection of the primary tumor and the metastases. Journal of surgical oncology 2019;120:438-45.
7. M. Okuno, E. Hatano, H. Nishino, S. Seo, K. Taura, S. Uemoto. Does response rate of chemotherapy with molecular target agents correlate with the conversion rate and survival in patients with unresectable colorectal liver metastases?: A systematic review. Eur J Surg Oncol 2017;43:1003-12.
8. Saskia Litière, Gaëlle Isaac, Elisabeth G. E. De Vries, Jan Bogaerts, Alice Chen, Janet Dancey, Robert Ford, Stephen Gwyther, Otto Hoekstra, Erich Huang, Nancy Lin, Yan Liu, et al. RECIST 1.1 for Response Evaluation Apply Not Only to Chemotherapy-Treated Patients But Also to Targeted Cancer Agents: A Pooled Database Analysis. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2019;37:1102-10.
9. Q. Chen, T. Li, W. Yue. Drug response to PD-1/PD-L1 blockade: based on biomarkers. Onco Targets Ther 2018;11:4673-83.
10. F. Pagni, E. Guerini-Rocco, A. M. Schultheis, G. Grazia, E. Rijavec, M. Ghidini, G. Lopez, K. Venetis, G. A. Croci, U. Malapelle, N. Fusco. Targeting Immune-Related Biological Processes in Solid Tumors: We do Need Biomarkers. Int J Mol Sci 2019;20.
11. X García-Albéniz, V Alonso, P Escudero, M Méndez, J Gallego, JR Rodríguez, A Salud, J Fernández-Plana, H Manzano, M Zanui, E Falcó, J Feliu, et al. Prospective Biomarker Study in Advanced RAS Wild-Type Colorectal Cancer: POSIBA Trial (GEMCAD 10-02). The oncologist 2019;24:e1115-e22.
12. Y. Shi, Z. Li, W. Zheng, X. Liu, C. Sun, J. B. Laugsand, Z. Liu, G. Cui. Changes of immunocytic phenotypes and functions from human colorectal adenomatous stage to cancerous stage: Update. Immunobiology 2015;220:1186-96.
13. D. O. Croci, M. F. Zacarías Fluck, M. J. Rico, P. Matar, G. A. Rabinovich, O. G. Scharovsky. Dynamic cross-talk between tumor and immune cells in orchestrating the immunosuppressive network at the tumor microenvironment. Cancer Immunol Immunother 2007;56:1687-700.
14. M. J. Duffy, J. Crown. Biomarkers for Predicting Response to Immunotherapy with Immune Checkpoint Inhibitors in Cancer Patients. Clin Chem 2019;65:1228-38.
15. A. S. Mansfield, S. J. Murphy, T. Peikert, E. S. Yi, G. Vasmatzis, D. A. Wigle, M. C. Aubry. Heterogeneity of Programmed Cell Death Ligand 1 Expression in Multifocal Lung Cancer. Clin Cancer Res 2016;22:2177-82.
16. S. P. Patel, R. Kurzrock. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. Mol Cancer Ther 2015;14:847-56.
17. A. Snyder, V. Makarov, T. Merghoub, J. Yuan, J. M. Zaretsky, A. Desrichard, L. A. Walsh, M. A. Postow, P. Wong, T. S. Ho, T. J. Hollmann, C. Bruggeman, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med 2014;371:2189-99.

18. E. M. Van Allen, D. Miao, B. Schilling, S. A. Shukla, C. Blank, L. Zimmer, A. Sucker, U. Hillen, M. H. G. Foppen, S. M. Goldinger, J. Utikal, J. C. Hassel, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 2015;350:207-11.
19. D Ciardiello, PP Vitiello, C Cardone, G Martini, T Troiani, E Martinelli, F Ciardiello. Immunotherapy of colorectal cancer: Challenges for therapeutic efficacy. Cancer treatment reviews 2019;76:22-32.
20. J Li, J Zhang, H Hu, Y Cai, J Ling, Z Wu, Y Deng. Gene Expression Signature to Predict Prognosis and Adjuvant Chemosensitivity of Colorectal Cancer Patients. Cancer management and research 2020;12:3301-10.
21. NA Abdul Aziz, NM Mokhtar, R Harun, MM Mollah, I Mohamed Rose, I Sagap, A Mohd Tamil, WZ Wan Ngah, R Jamal. A 19-Gene expression signature as a predictor of survival in colorectal cancer. BMC medical genomics 2016;9:58.
22. H Ito, Q Mo, LX Qin, A Viale, SK Maithel, AV Maker, J Shia, P Kingham, P Allen, RP DeMatteo, Y Fong, WR Jarnagin, et al. Gene expression profiles accurately predict outcome following liver resection in patients with metastatic colorectal cancer. PloS one 2013;8:e81680.
23. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, A. A. Alizadeh. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.
24. Q. Song, J. Shang, Z. Yang, L. Zhang, C. Zhang, J. Chen, X. Wu. Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. J Transl Med 2019;17:70.
25. J. Wang, H. Lin, M. Zhou, Q. Xiang, Y. Deng, L. Luo, Y. Liu, Z. Zhu, Z. Zhao. The m6A methylation regulator-based signature for predicting the prognosis of prostate cancer. Future Oncol 2020.
26. S. Zuo, G. Dai, X. Ren. Identification of a 6-gene signature predicting prognosis for colorectal cancer. Cancer Cell Int 2019;19:6.
27. J. Bai, X. Zhang, Z. X. Xiang, P. Y. Zhong, B. Xiong. Identification of prognostic immune-related signature predicting the overall survival for colorectal cancer. Eur Rev Med Pharmacol Sci 2020;24:1134-41.
28. J. Wang, S. Yu, G. Chen, M. Kang, X. Jin, Y. Huang, L. Lin, D. Wu, L. Wang, J. Chen. A novel prognostic signature of immune-related genes for patients with colorectal cancer. J Cell Mol Med 2020.
29. C. Wei, C. Yang, S. Wang, D. Shi, C. Zhang, X. Lin, B. Xiong. M2 macrophages confer resistance to 5-fluorouracil in colorectal cancer through the activation of CCL22/PI3K/AKT signaling. Onco Targets Ther 2019;12:3051-63.
30. D. Wang, L. Yang, D. Yue, L. Cao, L. Li, D. Wang, Y. Ping, Z. Shen, Y. Zheng, L. Wang, Y. Zhang. Macrophage-derived CCL22 promotes an immunosuppressive tumor microenvironment via IL-8 in malignant pleural effusion. Cancer Lett 2019;452:244-53.
31. L. Zhang, Y. Mao, Q. Mao, W. Fan, L. Xu, Y. Chen, L. Xu, J. Wang. FLOT1 promotes tumor development, induces epithelial-mesenchymal transition, and modulates the cell cycle by regulating the Erk/Akt signaling pathway in lung adenocarcinoma. Thorac Cancer 2019;10:909-17.

32. M. Zhang, K. Zhu, H. Pu, Z. Wang, H. Zhao, J. Zhang, Y. Wang. An Immune-Related Signature Predicts Survival in Patients With Lung Adenocarcinoma. Front Oncol 2019;9:1314.
33. E. Jeannot, A. Harlé, A. Holmes, X. Sastre-Garau. Nuclear factor I X is a recurrent target for HPV16 insertions in anal carcinomas. Genes Chromosomes Cancer 2018;57:638-44.
34. M. K. Wirtz, K. E. Keller. The Role of the IL-20 Subfamily in Glaucoma. Mediators Inflamm 2016;2016:4083735.
35. Y. Wei, Z. An, Z. Zou, R. Sumpter, M. Su, X. Zang, S. Sinha, M. Gaestel, B. Levine. The stress-responsive kinases MAPKAPK2/MAPKAPK3 activate starvation-induced autophagy through Beclin 1 phosphorylation. Elife 2015;4.
36. M. B. Menon, J. Gropengießer, J. Fischer, L. Novikova, A. Deuretzbacher, J. Lafera, H. Schimmeck, N. Czymmeck, N. Ronkina, A. Kotlyarov, M. Aepfelbacher, M. Gaestel, et al. p38(MAPK)/MK2-dependent phosphorylation controls cytotoxic RIPK1 signalling in inflammation and infection. Nat Cell Biol 2017;19:1248-59.
37. H Läubli, L Borsig. Altered Cell Adhesion and Glycosylation Promote Cancer Immune Suppression and Metastasis. Frontiers in immunology 2019;10:2120.
38. H Harjunpää, M Llort Asens, C Guenther, SC Fagerholm. Cell Adhesion Molecules and Their Roles and Regulation in the Immune and Tumor Microenvironment. Frontiers in immunology 2019;10:1078.
39. H. Wurzer, C. Hoffmann, A. Al Absi, C. Thomas. Actin Cytoskeleton Straddling the Immunological Synapse between Cytotoxic Lymphocytes and Cancer Cells. Cells 2019;8.
40. M. L. Dustin. The immunological synapse. Cancer Immunol Res 2014;2:1023-33.
41. P Aparicio-Domingo, H Cannelle, MB Buechler, S Nguyen, SM Kallert, S Favre, N Alouche, N Papazian, B Ludewig, T Cupedo, DD Pinschewer, SJ Turley, et al. Fibroblast-derived IL-33 is dispensable for lymph node homeostasis but critical for CD8 T cell responses to acute and chronic viral infection. European journal of immunology 2020.
42. OS Dmitrieva, IP Shilovskiy, MR Khaitov, SI Grivennikov. Interleukins 1 and 6 as Main Mediators of Inflammation and Cancer. Biochemistry. Biokhimiia 2016;81:80-90.
43. DO Griffin, TL Rothstein. Human "orchestrator" CD11b(+) B1 cells spontaneously secrete interleukin-10 and regulate T-cell activity. Molecular medicine (Cambridge, Mass.) 2012;18:1003-8.
44. Q Wen, W Xiong, J He, S Zhang, X Du, S Liu, J Wang, M Zhou, L Ma. Fusion cytokine IL-2-GMCSF enhances anticancer immune responses through promoting cell-cell interactions. Journal of translational medicine 2016;14:41.
45. J Geginat, P Larghi, M Paroni, G Nizzoli, A Penatti, M Pagani, N Gagliani, P Meroni, S Abrignani, RA Flavell. The light and the dark sides of Interleukin-10 in immune-mediated diseases and cancer. Cytokine & growth factor reviews 2016;30:87-93.
46. P Zhang, GR Hill. Interleukin-10 mediated immune regulation after stem cell transplantation: Mechanisms and implications for therapeutic intervention. Seminars in immunology 2019;44:101322.

47. Q. Zhou, Z. Hou, S. Zuo, X. Zhou, Y. Feng, Y. Sun, X. Yuan. LUCAT1 promotes colorectal cancer tumorigenesis by targeting the ribosomal protein L40-MDM2-p53 pathway through binding with UBA52. Cancer Sci 2019;110:1194-207.
48. F. Wang, X. Chen, X. Yu, Q. Lin. Degradation of CCNB1 mediated by APC11 through UBA52 ubiquitination promotes cell cycle progression and proliferation of non-small cell lung cancer cells. Am J Transl Res 2019;11:7166-85.
49. J. W. Shay, W. E. Wright. Telomeres and telomerase: three decades of progress. Nat Rev Genet 2019;20:299-309.
50. M. Benyelles, M. F. O'Donohue, L. Kermasson, E. Lainey, R. Borie, C. Lagresle-Peyrou, H. Nunes, C. Cazelles, C. Fourrage, E. Ollivier, A. Marcais, A. S. Gamez, et al. NHP2 deficiency impairs rRNA biogenesis and causes pulmonary fibrosis and Høyeraal-Hreidarsson syndrome. Hum Mol Genet 2020;29:907-22.
51. B. El Hassouni, D. Sarkisjan, J. C. Vos, E. Giovannetti, G. J. Peters. Targeting the Ribosome Biogenesis Key Molecule Fibrillarin to Avoid Chemoresistance. Curr Med Chem 2019;26:6020-32.
52. A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S. Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, A. A. Alizadeh. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;37:773-82.

Tables

Table 1 Clinical characteristics of the CRC patients.		
Characteristic	**Training (n=487)**	**Test (n=578)**
Age		
≤ 68	243	311
> 68	244	267
Gender		
Female	229	260
Male	258	919
Survival status		
Living	379	385
Dead	108	194
Pathological M		
Unknown	63	22
M0	355	496
M1	69	61
Pathological N		
Unknown	20	26
N0	291	311
N1	111	136
N2	85	100
N3	0	6
Pathological T		
Unknown	1	23
T0	0	1
T1	11	12
T2	83	48
T3	334	376
T4	58	119
Tumor stage		

Unknown	12	0
Stage 0	0	4
Stage I	80	37
Stage II	193	269
Stage III	133	209
Stage IV	69	60

Abbreviation: CRC, Colorectal Cancer

Table 2. Survival analysis of the IRG in the prognostic signature

SYMBOL	Ensemble ID	HR	right	left	COX P
CCL22	ENSG00000102962	0.66	0.50	0.87	0.001***
LIMK1	ENSG00000106683	1.50	1.05	2.13	0.03*
MAPKAPK3	ENSG00000114738	0.63	0.44	0.90	0.01**
FLOT1	ENSG00000137312	1.82	1.19	2.79	0.01**
GPRC5B	ENSG00000167191	1.85	1.35	2.52	0.001***
IL20RB	ENSG00000174564	1.82	1.35	2.43	0.001***

Abbreviation: IRG, Immune-related gene；HR, Hazard ratio

*：$P < 0.05$； **：$P < 0.01$； ***：$P < 0.001$

Table 3. Association of the IRG signature with clinical characteristics in CRC patients						
Variables	**Training group**		***P***	**Test group**		***P***
	Low risk	**High risk ***		**Low risk**	**High risk**	
Age (years)			0.30			0.46
≤ 68	115	128		150	161	
> 68	128	116		138	129	
Gender			0.06			0.96
Female	125	104		129	131	
Male	118	140		160	159	
Tumour stage			0.001***			0.001***
Stage 0	3	9		85	124	
Stage I	60	20		3	1	
Stage II	112	81		27	10	
Stage III	55	78		152	117	
Stage IV	13	56		22	38	
Pathologic T			0.001***			0.001***
Unknown		1		5	18	
T0				1	0	
T1	9	2		7	5	
T2	59	24		31	17	
T3	157	177		198	178	
T4	18	40		47	72	
Pathologic N			0.001***			0.001***
Unknown				7	25	
N0	177	114		184	127	
N1	48	63		62	74	
N2	18	67		36	64	
Pathologic M			0.001***			0.001***
Unknown	26	37		3	19	

M0	204	151	263	233
M1	13	56	23	38

Abbreviation: IRG, Immune-related gene；CRC, Colorectal Cancer

*：*P* < 0.05； **：*P* < 0.01； ***：*P* < 0.001

Table 4. Cox regression analysis of the IRG signature with survival

		Univariable analysis				Multivariable analysis			
Variables		**HR**	**95% CI of HR**		***P***	**HR**	**95% CI of HR**		***P***
			lower	**upper**			**lower**	**upper**	
The Training group									
Age	**>68 vs.≤68**	1.59	1.08	2.34	0.02*	1.78	1.20	2.64	：0.001***
Gender	**Male vs. Female**	1.15	0.79	1.69	0.47	0.91	0.61	1.34	0.62
Tumor stage	**III, IV Vs. I, II**	1.04	1.01	1.06	： 0.001***	1.03	1.00	1.05	0.02*
Signature	**High risk vs. Low risk**	4.49	2.79	7.23	： 0.001***	4.56	2.81	7.40	： 0.001****
The Test group									
Age	**>68 vs.≤68**	1.89	1.42	2.51	： 0.001***	1.03	1.01	1.04	：0.001***
Gender	**Male vs. Female**	1.31	0.98	1.74	0.07	0.88	0.60	1.30	0.53
Tumor stage	**III, IV Vs.I, II**	1.94	1.59	2.36	： 0.001***	1.02	1.00	1.05	0.03*
Signature	**High risk vs. Low risk**	1.70	1.27	2.26	： 0.001***	4.53	2.79	7.34	：0.001***

Abbreviation: IRG, Immune-related gene；HR, Hazard ratio

*：*P* < 0.05； **：*P* < 0.01； ***：*P* < 0.001

Figures

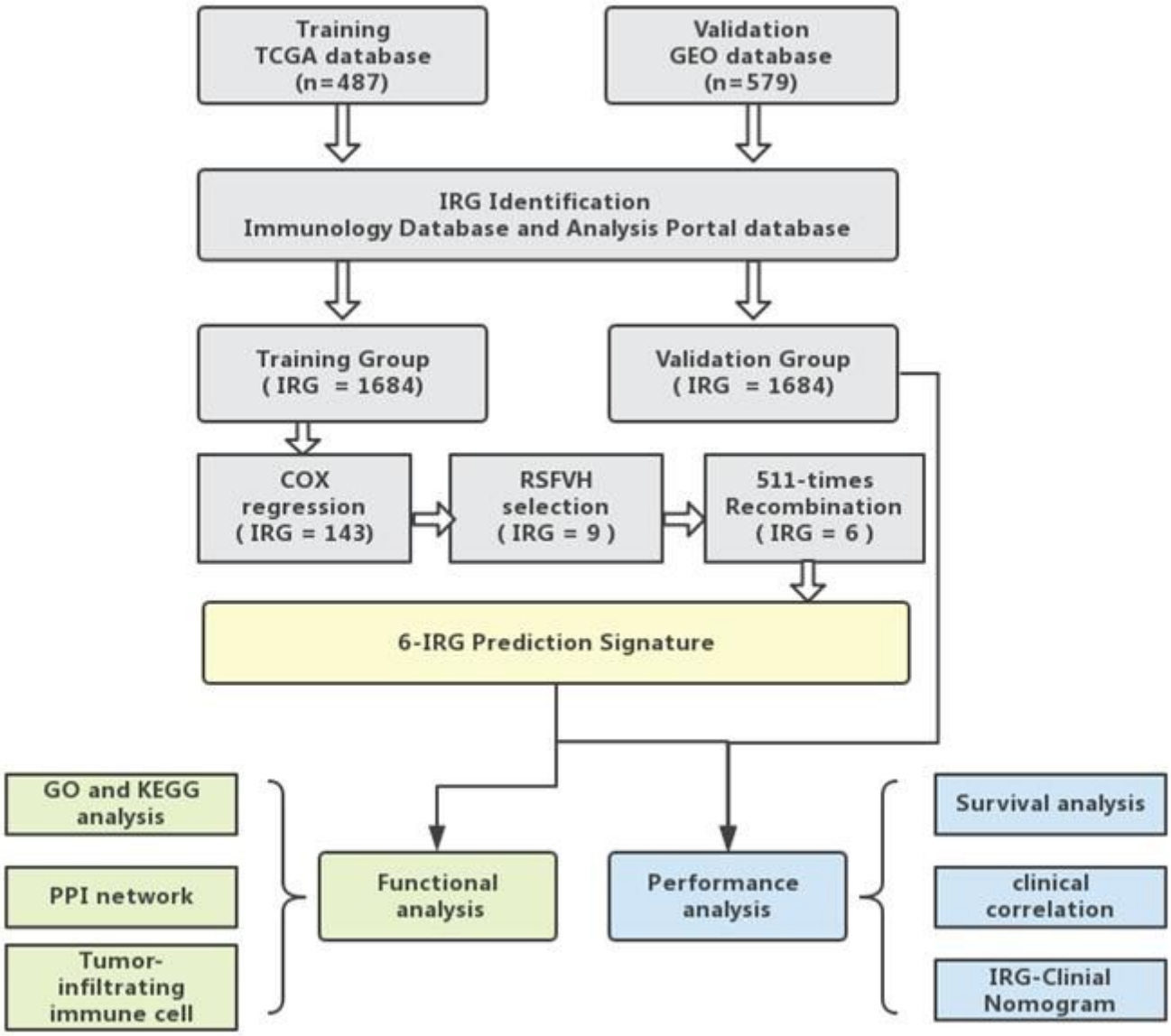

Figure 1

The schematic of this study (Abbreviation: TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; IRG, Immune-related gene；RSFVH, random survival forest algorithm; GO, gene ontology; PPI, protein-protein interaction.)

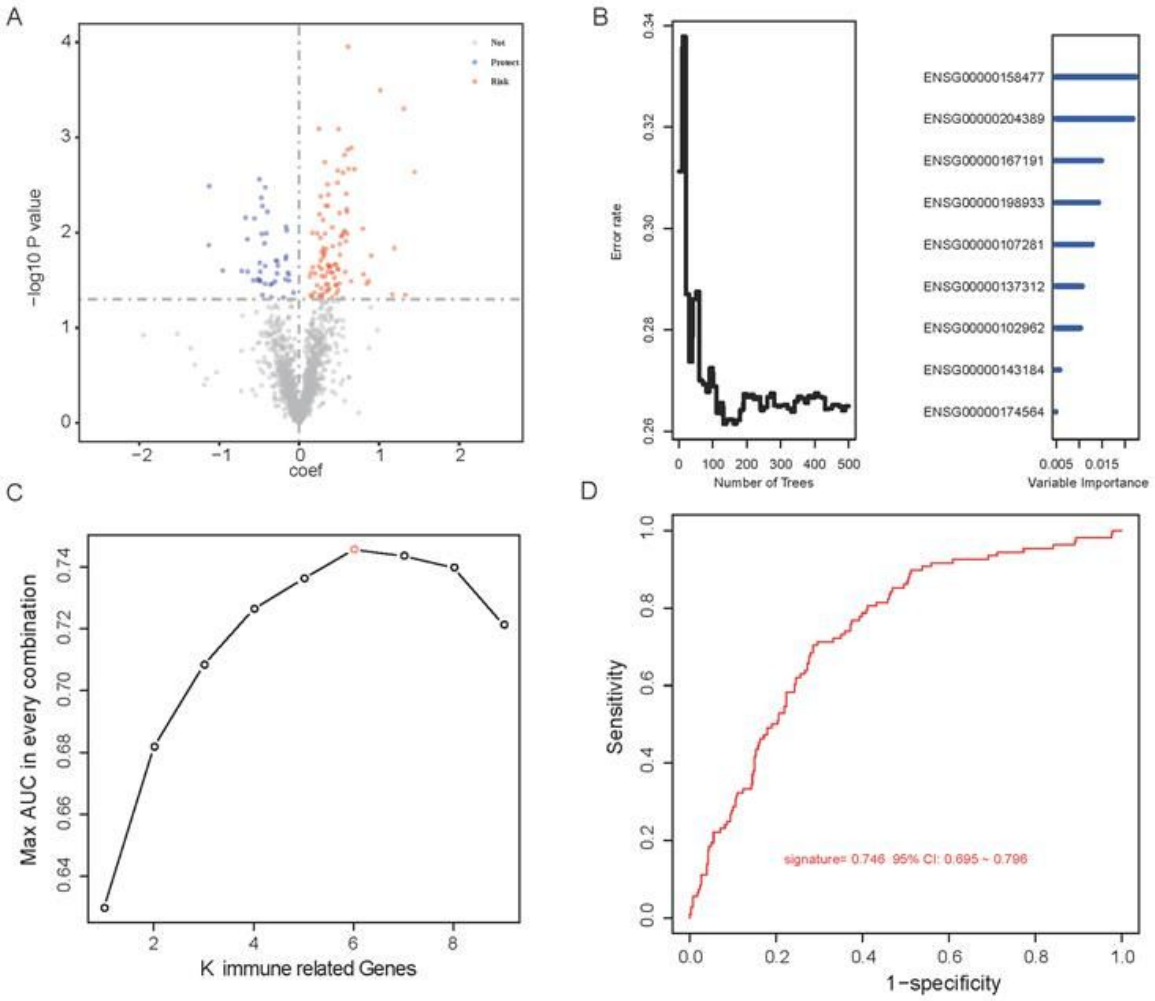

Figure 2

Identification of a six-IRG panel for prognosis prediction (1a: Univariate COX regression of 1684 IRGs regarding patients' survival. IRGs with statistical significance (n=143) was marked blue if it's a protective factor or red as risk factor. 1b: RSFVH was used to select highly survival-correlated IRGs. And nine genes were enrolled. 1c: Full-scale recombination of the nine IRGs was conducted to rule out overfitting (combinations, n=511). The six-gene panel with an AUC of 0.746 was selected as the optimal prediction panel. 1d: The ROC curve of the designated IRG panel.)

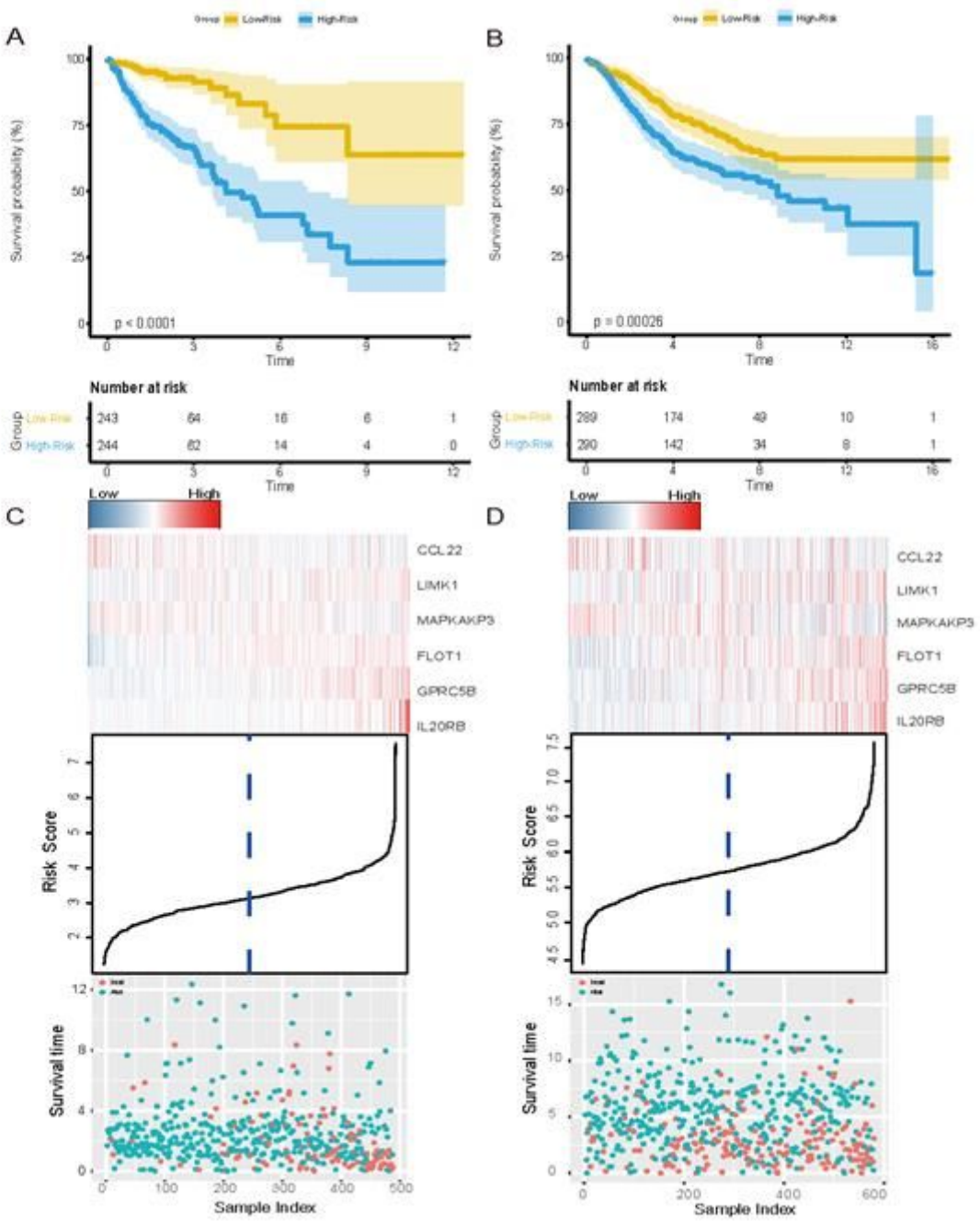

Figure 3

Survival analysis and IRG expression information (2a and 2b: In the training and validation set, Kaplan-Meier (KM) test was performed to verify survival between high and low risk population. The p value was less than 0.001 in both groups. 2c and 2d: Clusters in gene expression level and survival information in training and validation group, respectively. Genes with adverse prognostic effect, namely LIMK1, FLOT1, GPRC5B and IL20RB, showed consistent elevation in expression in high-risk population.)

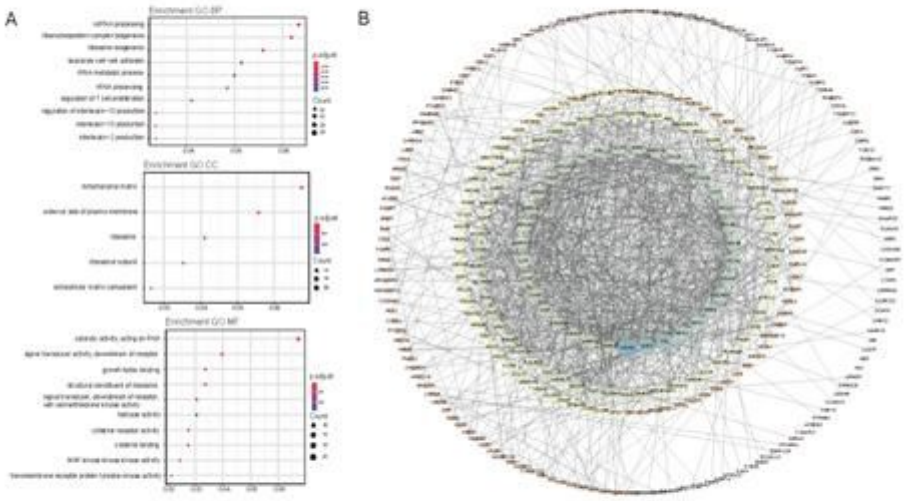

Figure 4

Gene function analysis and proteomic interaction network using co-expressing genes (3a: GO analysis of the CEGs consists cellular component (CC), molecular function (MF) and biological process (BP). 3b: Protein-protein interaction network using the CEGs. Large-sized hubs in the inner circle represent great importance in the network.)

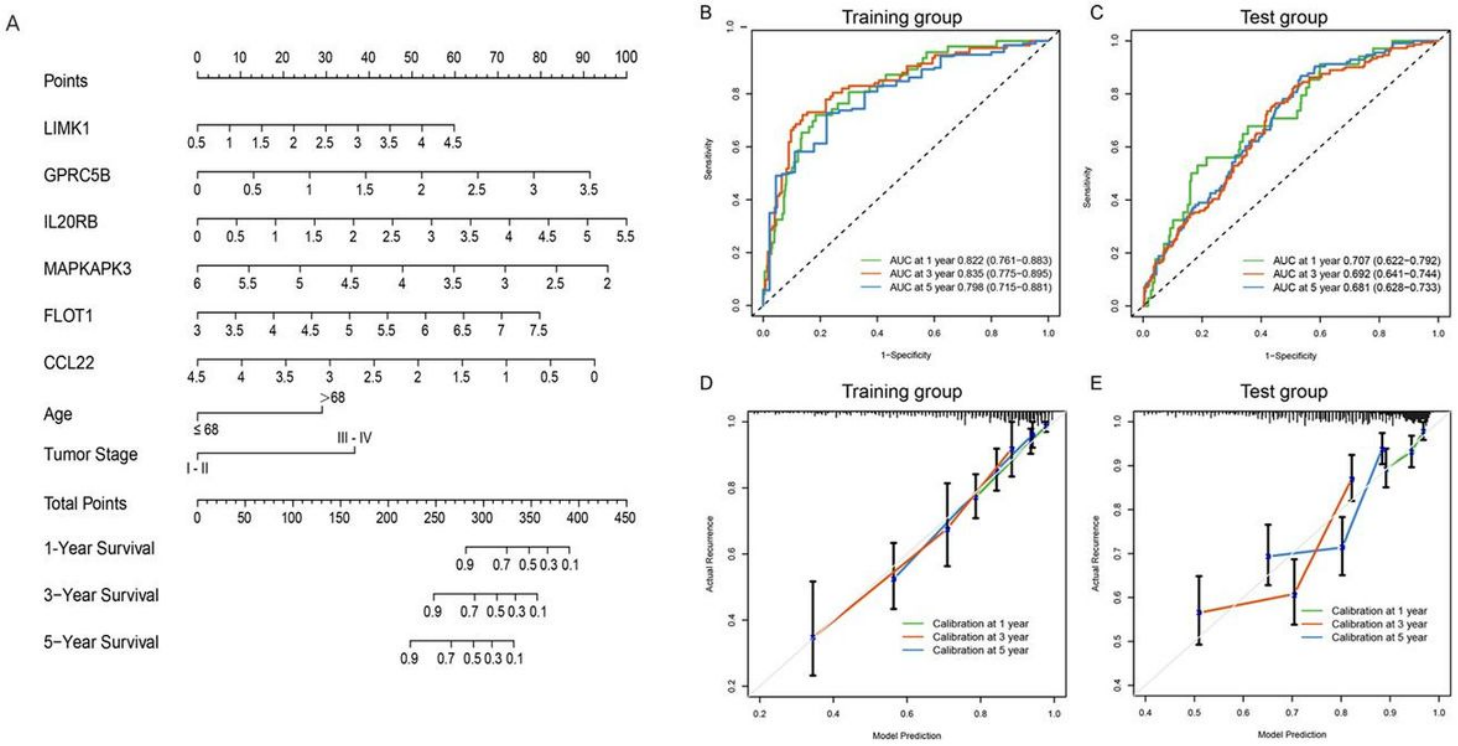

Figure 5

Nomogram incorporating clinical factors and IRG signature (4a: A predictive nomogram was constructed using clinical risk factors and the designated IRG panel. The probability of 1, 3 and 5-year survival could be calculated. 4b and 4c: Time-ROC curves of the nomogram in the training and validation group. The AUC values were consistent at different time points. 4d and 4e: Calibration curves were plotted to reveal concordance between prediction and reality. In both training and validation set, the model showed good calibration.)

Supplementary Files

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

- FigureS1.pdf
- supplement2.pdf
- TableS3.doc
- TableS2.xls
- TableS1.doc