Development of immune gene pair-based signature predictive of prognosis and immunotherapy in esophageal cancer

Kui Cao[1,2#], Tianjiao Ma[3#], Xiaodong Ling[4#], Mingdong Liu[2], Xiangyu Jiang[4], Keru Ma[4], Jinhong Zhu[1^], Jianqun Ma[4^]

[1]Department of Clinical Laboratory, Biobank, Harbin Medical University Cancer Hospital, Harbin, China; [2]Department of Clinical Oncology, Harbin Medical University Cancer Hospital, Harbin, China; [3]Department of Cardiovascular Surgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China; [4]Department of Thoracic Surgery, Harbin Medical University Cancer Hospital, Harbin, China

Contributions: (I) Conception and design: K Cao, J Ma, J Zhu, T Ma, X Ling; (II) Administrative support: J Ma, J Zhu; (III) Provision of study materials or patients: K Cao, J Ma; (IV) Collection and assembly of data: K Cao, T Ma, X Ling, M Liu, X Jiang, K Ma; (V) Data analysis and interpretation: K Cao, J Ma, T Ma, X Ling, J Zhu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.
[#]These authors contributed equally to this work.

Correspondence to: Jianqun Ma. Department of Thoracic Surgery, Harbin Medical University Cancer Hospital, 150 Haping Road, Harbin 150040, China. Email: jianqunma@hrbmu.edu.cn; Jinhong Zhu. Department of Clinical Laboratory, Biobank, Harbin Medical University Cancer Hospital, Harbin, China. Email: zhujinhong@hrbmu.edu.cn.

Background: Esophageal cancer (EC) is one of the deadliest solid malignancies, mainly consisting of esophageal squamous cell carcinoma (ESCC) and adenocarcinoma (EAC). Robust biomarkers that can improve patient risk stratification are needed to optimize cancer management. We sought to establish potent prognostic signatures with immune-related gene (IRG) pairs for ESCC and EAC.

Methods: We obtained differentially expressed IRGs by intersecting the Immunology Database and Analysis Portal (ImmPort) with the transcriptome data set of The Cancer Genome Atlas (TCGA)-ESCC and EAC cohorts. A novel rank-based pairwise comparison algorithm was applied to select effective IRG pairs (IRGPs), followed by constructing a prognostic IRGP signature via the least absolute shrinkage and selection operator (LASSO) regression model. We assessed the predictive power of the IRGP signatures on prognosis, tumor-infiltrating immune cells, and immune checkpoint inhibitor (ICI) efficacy in EC. Kaplan-Meier survival analysis and receiver operating characteristic curves (ROC) were used to evaluate the clinical significance of IRGPs. Univariate and multivariate Cox regression analyses were performed to investigate the association of overall survival (OS) with IRGPs and clinical characteristics.

Results: We built a 19-IRGP signature for ESCC (n=75) and a 17-IRGP signature for EAC (n=78), with an area under the ROC curve (AUC) of 0.931 and 0.803, respectively. IRGP signature-derived risk scores stratified patients into low- and high-risk groups with significantly different OS in ESCC and EAC (P<0.001). Nomogram and decision curve analysis were used to evaluate the clinical relevance of the prognostic signatures, achieving a C-index of 0.973 in ESCC and 0.880 in EAC. The risk scores were associated with immune and ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) scores and the composition of immune cells in the tumor microenvironment. The association between risk score and human leukocyte antigens (HLAs), mismatch repair (MMR) genes, and immune checkpoint molecules demonstrated its predictive value for ICI response. Differential immune characteristics and predictive value of the risk score were observed in EAC.

Conclusions: The established immune signatures showed great promise in predicting prognosis, tumor immunogenicity, and immunotherapy response in ESCC and EAC.

Keywords: Esophageal cancer (EC); immune-related gene pair; signature; prognosis

Introduction

Esophageal cancer (EC) is the sixth leading cause of death among all cancers, accounting for 5% of all cancer-related deaths in 2018 (1). EC is histologically classified into esophageal squamous cell cancer (ESCC) and adenocarcinoma (EAC). Around half of EC patients present with unresectable or metastatic conditions at diagnosis (2). Over the past few decades, modest improvements have been attained in the treatment of inoperable EC due to multidisciplinary therapies. Unfortunately, the prognosis of metastatic EC remains unsatisfying, with a median survival of shorter than 1 year.

Recently, immune checkpoint inhibitors (ICIs) targeting programmed cell death protein 1 (PD1), programmed cell death 1 ligand 1 (PDL1), or cytotoxic T lymphocyte antigen 4 (anti-CTLA4) have shown great promise in EC patients (2). However, only some EC patients benefit from ICIs and reach a sustainable clinical response, whereas an innate resistance occurs in others. Therefore, robust predictive biomarkers to identify subgroup patients suitable for ICI treatment are urgently needed.

A growing body of evidence has indicated that the tumor immune microenvironment (TIME), characterized by the density, composition, functional state, and organization of tumor-infiltrating leukocytes, is associated with prognosis and therapy response (3-6). Different types of immune signatures have been reported, including prognostic, predictive, mechanistic, and escape signatures (7,8). A review of published immune signatures revealed a frequent overlap of some core elements in prognostic and predictive signatures (6), suggesting that crucial immune features may be potentially used to guide cancer management. In EC, chronic inflammation-induced constitutive activation of pro-inflammatory signaling pathways stimulate survival and proliferation of tumor cells (9). Myeloid-derived suppressor cells and regulatory T cells suppress antitumor immunity (9). Other immune cells, such as tumor-associated macrophages and cancer-associated fibroblasts, also contribute to EC carcinogenesis (9). Moreover, the abundance of T_H17 cells were associated with improved survival in EC cancer (10). Leukocyte surface antigen CD47 was an independent predictor for overall survival (OS) and progression-free

survival (PFS) in ESCC (11). Besides, anti-PD1 and anti-PDL1 immunotherapy are approved to treat EC by the US Food and Drug Administration (FDA), and durable and objective responses to these two ICIs were achieved (6). These findings demonstrate the crucial role of TIME in EC prognosis. Therefore, it is biologically plausible that immune gene signatures capable of recapitulating TIME features may potentially be reliable cancer biomarkers for predicting prognosis and immunotherapy sensitivity in EC.

With the development of high throughput sequencing technology, various transcriptome-based signatures for survival classification or therapeutic response prediction have emerged in EC (12-18). However, few signatures can predict response to ICIs and none are translated into routine clinical practice, which is partly attributable to the critical impact of batch effects in high-throughput data and the challenges of data normalization (19). Recently, a method based on the relative ranking of gene expression levels by pairwise comparison has been developed (20). This method circumvents formidable data reprocessing, including scaling and normalization, and has been substantiated to yield potent results in risk classification of cancer, including lung cancer (21,22), osteosarcoma (23), hepatocellular carcinoma (24), and serous ovarian carcinoma (25).

With this in mind, we aimed to develop prognostic signatures with immune-related gene pairs (IRGPs) for ESCC and EAC by adopting a pairwise comparison-based gene ranking method. We retrieved the immune-related genes (IRGs) from the Immunology Database and Analysis Portal (ImmPort) and then identified differentially expressed IRGs in The Cancer Genome Atlas (TCGA)-ESCC and EAC cohorts. The expression levels of all differentially expressed IRGs underwent a pairwise comparison in each tumor sample. The top-ranking and survival-associated IRGPs were used to construct predictive gene signatures.

Overall, we developed robust IRGP signatures to predict survival in ESCC and EAC. Nomogram and decision curve analysis (DCA) were performed to verify the discrimination ability and clinical benefit of the prognostic signature, respectively. Moreover, we used Tumor Immune Estimation Resource (TIMER2.0, http://timer.cistrome.org/) to evaluate immune cell infiltration in ESCC and

EAC, in which seven different algorithms were applied to determine the proportion of immune cells in the tumor microenvironment, including CIBERSORT, CIBERSORT-ABS, EPIC, MCP-counter, quanTIseq, TIMER, and xCell. Finally, we extended the use of the signature to predict response to ICIs by evaluating the associations of the risk score with immune checkpoint genes, human leukocyte antigens (HLAs), and mismatch repair (MMR) genes.

We present the following article in accordance with the TRIPOD reporting checklist (available at https://dx.doi.org/10.21037/atm-21-5217).

Methods

Clinical sample and data collection

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

We acquired RNA sequencing (RNA-seq) expression data of EC samples from TCGA database (https://tcga-data.nci.nih.gov/tcga/), accompanied by the corresponding clinical data. The ESCC dataset consisted of 11 normal and 80 ESCC samples, while the EAC project comprised 9 normal and 78 EAC samples. General Transfer Format (GTF) files downloaded from GENCODE (http://www.gencodegenes.org/) were used to annotate genes. Genes with an average expression value of less than 1 were removed as were samples lacking survival information. The remaining 75 ESCC and 78 EAC patients were included in the study.

Acquisition of differentially expressed IRGs

We retrieved 1,509 IRGs from the Immunology Database and Analysis Portal (ImmPort) (https://www.immport.org/) (26). We then obtained differentially expressed genes (DEGs) between normal and tumor tissues (where $P<0.05$ and $|\log_2\text{fold change [FC]}| >1.5$) by using the R package edgeR (27), identifying 9,606 and 8,737 DEGs for ESCC and EAC, respectively. Venn diagram analysis was carried out using the DEG and IRG datasets, and genes in the intersection of the 2 datasets were identified as differentially expressed immunity-related genes (DEIRGs), which are hereafter referred to as IRGs.

Identification of prognostic IRGPs in patients with EC

We then performed a pairwise comparison of expression levels in each sample to assign a score to each IRGP. If the expression level of the first IRG in a sample was lower than that of the second IRG, this IRGP was scored as 1; otherwise, the IRGP scored 0. IRGPs that constantly received the same score (0 or 1) in more than 80% of samples were deleted since they failed to present discriminative information with respect to survival. Prognostic IRGPs were identified using univariate Cox regression analysis to check the association of each IRGP with OS in the TCGA cohort. IRGPs with $P<0.05$ were used to establish prognostic signatures.

Construction of signatures based on IRGPs

Least absolute shrinkage and selection operator (LASSO) Cox regression analysis (glmnet package) was adopted to select prognostic IRGPs and build prognostic signatures (28,29). After the establishment of the IRGP signature, a risk score was assigned for each patient by linearly integrating the patient score (0 or 1) and risk contribution of each IRGP included in the prognostic signature, using a previously described formula (30-32). $Risk\ score = \sum_{i=1}^{n} \beta i \chi i$, β and χ representing coefficient and the 0-or-1 matrix of each IRGP, respectively.

Estimation of tumor-infiltrating immune cells

The ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) tool enabled us to assess the composition of infiltrating stromal and immune cells in tumor tissues using gene expression signatures (33). Briefly, a stromal signature was developed to quantify the stroma levels, while an immune signature was used to recapitulate the infiltration of immune cells in tumor tissues. The ESTIMATE algorithm relies on single sample gene set enrichment analysis and generates 3 scores. The immune and stromal scores reflect the estimated immune and stromal cells in the tumor, respectively. The ESTIMATE score, integrating the 2 former scores, negatively correlates to tumor purity (33). In addition, we used the Tumor Immune Estimation Resource (TIMER2.0, http://timer.cistrome.org/) to evaluate immune cell infiltration in ESCC and EAC. This web tool provides seven different algorithms to infer components of immune cells in tumor microenvironment, including CIBERSORT, CIBERSORT-ABS, EPIC, MCP-counter, quanTIseq, TIMER, and xCell (34). A lollipop diagram was used to

Page 4 of 16

Cao et al. Prognostic immune-related gene pair signature in EC

exhibit the correlation between risk score and immune cell subsets (ggstatsplot package).

Analysis of the expression of potential biomarkers predictive of ICI efficacy

R packages ggpubr and ggplot2 were used to analyze the association of risk scores with HLAs and MMR genes, respectively. The ggstatsplot package was employed to study the relationship between the model and the expression level of immune checkpoint molecules, and violin plots were used to visualize results.

Gene set enrichment analysis (GSEA)

GSEA software (version 4.0.1) was used to compare high- and low-risk groups. The GSEA immunologic signature gene sets (C7) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) subset of canonical pathways from the GSEA curated gene set (C2) were applied to dissect differential immune cell subsets and signaling pathways in the high- and low-risk groups. $P<0.05$ and false discovery rate (FDR) <0.25 were considered statistically significant.

Statistical analysis

All statistical analyses were carried out using R (version 4.0.3, https://www.R-project.org). We allocated patients to a low- or high-risk group using an optimal risk score cut-off value, which was estimated through a time-dependent receiver operating characteristic (ROC) curve (survival ROC package) at 3 years (35). We plotted a Kaplan-Meier survival curve for OS, and the log-rank test was used to evaluate the difference in OS between the 2 risk groups. ROC curve analysis was carried out to evaluate the sensitivity and specificity of IRGPs, clinical characteristics alone or in combination, followed by the calculation of area under curves (AUCs). Univariate and multivariate Cox regression hazard proportional analyses were performed to investigate the association of OS with IRGPs and clinical characteristics. Significant risk factors were further subjected to multivariate analysis. We combined the clinical characteristics of the TCGA data set with the IRGP signature to construct a nomogram. A concordance index (C-index) was used to evaluate the discriminative power of the nomogram (36). We also drew calibration curves and conducted DCA (37,38) to evaluate the nomogram's accuracy and clinical net benefit, respectively. The

Wilcoxon signed-rank test was used to check statistically significant differences between groups. Pearson's chi-square tests were employed to estimate correlations between variables. Unless otherwise stated, $P<0.05$ was considered statistically significant.

Results

Study populations

We initially obtained data for 80 ESCC and 78 EAC samples. After removing samples missing survival information, 75 ESCC and 78 EAC cases were included in the study. Characteristics of participants are displayed in Table S1.

Identification of differentially expressed IRGs

A flow chart of this study is shown in *Figure 1*. A total of 1,509 IRGs were acquired from the ImmPort database (https://www.immport.org/) (26). In addition, we retrieved 9,606 DEGs in ESCC (n=75) and 8,737 DEGs in EAC (n=78). By intersecting IRGs and DEGs, we identified 298 and 289 genes present in both IRG and DEG data sets (i.e., differentially expressed IRGs) for ESCC and EAC, respectively. These genes were used for subsequent analyses.

Construction and evaluation of IRGP signatures

Gene ranking by pairwise comparison yielded 11,770 and 11,945 discriminative IRGPs for ESCC and EAC, respectively. From these, univariate Cox analysis identified 31 ESCC- and 65 EAC-related prognostic IRGPs (P<0.01). The LASSO Cox regression model was applied to survival time data sets for the TCGA-ESCC and -EAC cohorts, separately. We generated a 19-IRGP signature with 36 IRGs for ESCC (*Figure 2A*, Table S2) and a 17-IRGP signature for EAC composed of 28 IRGs (*Figure 2B*, Table S2). Every patient received a risk score derived from the IRGP signatures, with a larger risk score signifying a higher risk for poor survival. We predicated prognosis for ESCC with indicated variables and gained an AUC of 0.555, 0.700, 0.611, 0.407, 0.475, 0.510, and 0.931 for tumor (T), node (N), metastasis (M), sex, age, stage, and risk score, respectively (*Figure 3A*). Notably, the risk score of IRGP performance exceeded traditional biomarkers. *Figure 3B* shows that the risk score also exhibited the highest AUC (0.830) among all indicated variables in EAC.

To divide patients into high- and low-risk groups based
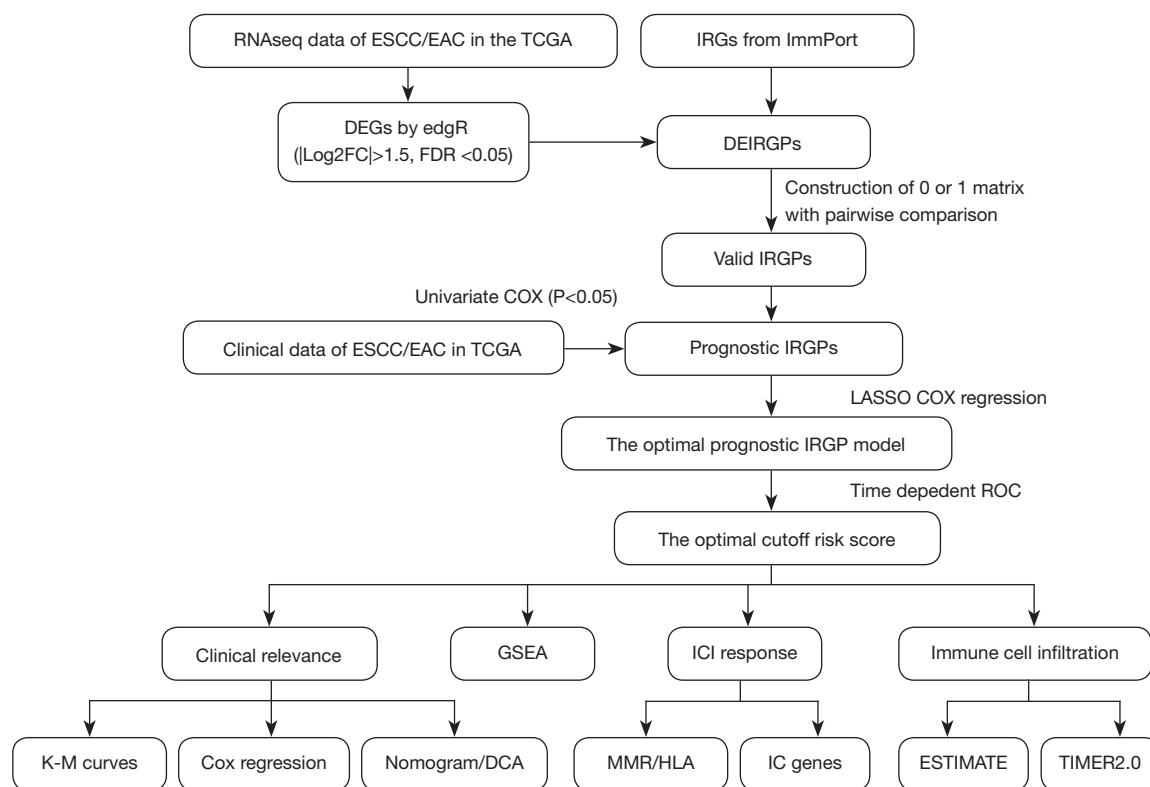
Figure 1 Flow diagram of the study. ESCC, esophageal squamous cell cancer; EAC, esophageal adenocarcinoma; TCGA, The Cancer Genome Atlas; IRGs, immune-related genes; DEGs, differentially expressed genes; DEIRGPs, differentially expressed immunity-related genes; IRGPs, immune-related gene pairs; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; GSEA, Gene set enrichment analysis; ICI, immune checkpoint inhibitors; DCA, decision curve analysis; MMR, mismatch repair; HLA, human leukocyte antigens; ESTIMATE, Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data; TIMER2.0, Tumor Immune Estimation Resource.

on risk scores, we plotted a time-dependent ROC curve at 3 years and used the risk score corresponding to the maximum value of the Youden index as the optimal cut-off value. Kaplan-Meier survival curves of OS for high- and low-risk patients revealed that survival of patients in the low-risk group was significantly longer than those in the high-risk group in terms of both ESCC (*Figure 3C*) and EAC (*Figure 3D*). Patient survival status is visualized in *Figure 3E*. ESCC patients with low risk were more likely to remain alive than high-risk counterparts. Similar results were observed for EAC (*Figure 3F*).

Univariate and multivariate analyses were performed with the Cox hazard proportional regression model. Significant risk factors were further subjected to multivariate analysis. Univariate analyses demonstrated the association of risk score with OS in ESCC [hazard ratio (HR) =28.592, 95% confidence interval (CI): 7.266–112.507, P<0.001] and

EAC (HR =19.861, 95% CI: 8.780–44.927, P<0.001). The risk score was an independent predictor of prognosis in ESCC (adjusted HR =30.883, 95% CI: 7.518–126.862, P<0.001) (*Figure 4A*). As shown in Figure 4B, the risk score was also independently predictive of prognosis in EAC (adjusted HR =23.117, 95% CI: 9.451–56.543, P<0.001). We constructed a nomogram for ESCC with the IRGP signature and sex (*Figure 4C*), achieving a C-index as high as 0.973. A nomogram was also built for EAC based on the IRGP signature and metastatic status (*Figure 4D*) with a C-index of 0.880. Calibration curves confirmed that actual survival fit predicted survival curves well, suggesting a high predictive accuracy (*Figure 4E*). DCA indicated that the risk score model produced decent clinical benefits across a broad range (*Figure 4F*). The prognostic value of the EAC-specific 17-IRGP signature was also further evaluated (*Figure 4G,4H*).
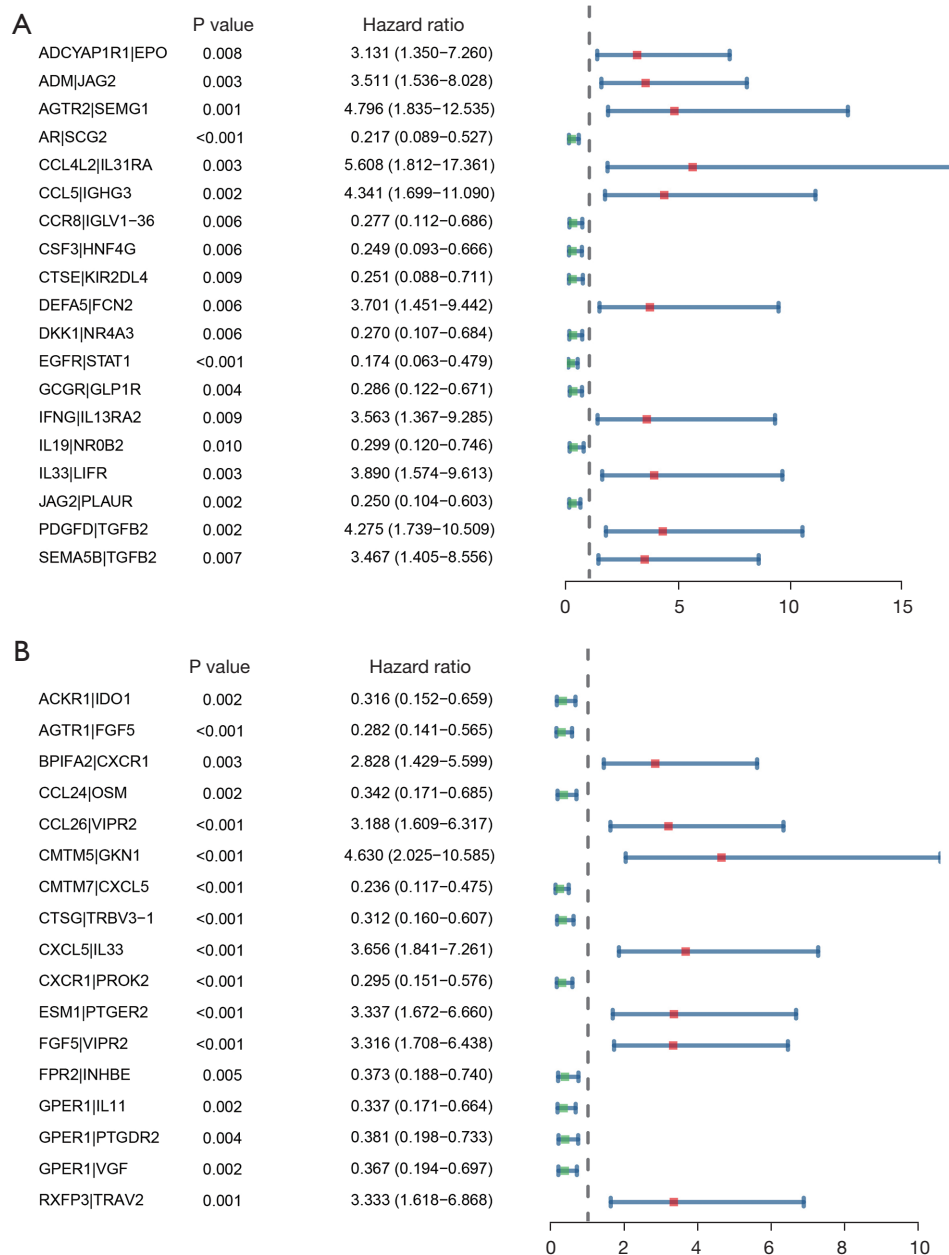
Figure 2 Association between overall survival and IRGPs included in the prognostic signatures. The forest plot displays univariate Cox regression analysis results of IRGPs in ESCC (A) and EAC (B). IRGPs, immune-related gene pairs; ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma.

Association of IRGP signatures with immune cell infiltration

TIME is frequently associated with cancer prognosis. Therefore, we examined whether there were differences in immune cell infiltration between the 2 prognostic risk groups in the ESCC and EAC cohorts, separately. As revealed by the ESTIMATE algorithm, both the immune and ESTIMATE scores were significantly enhanced in the high-risk group compared with the low-risk group (*Figure 5A,5B*). Intriguingly, the opposite results were observed for the immune/ESTIMATE score and risk score in EAC (*Figure 5C,5D*). Accordingly, the 2 types of scores were associated with risk scores in ESCC (*Figure 5E,5F*)

Ann Transl Med 2021;9(20):1591 | https://dx.doi.org/10.21037/atm-21-5217
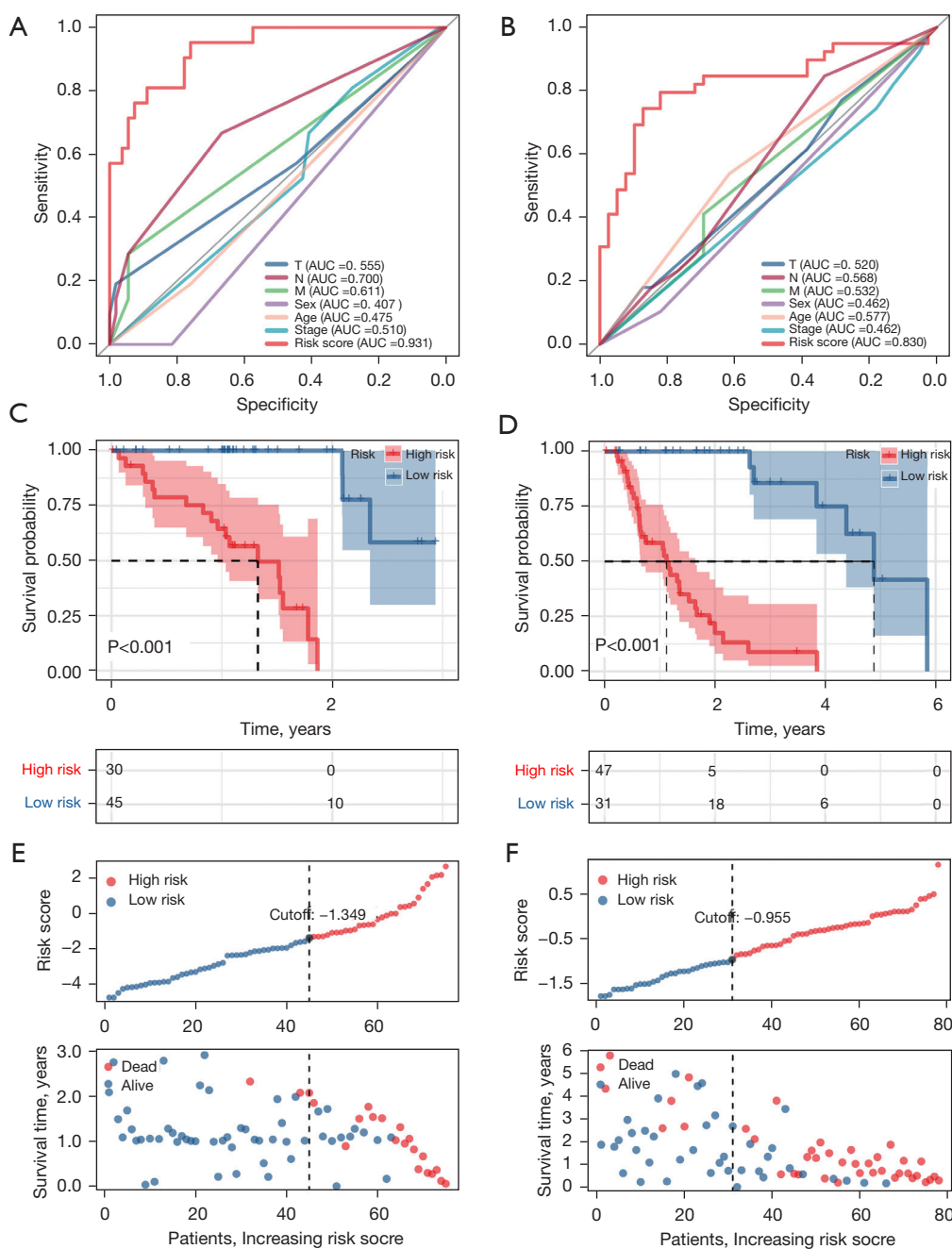
Figure 3 Screening prognosis IRGs for model construction. ROC curve of clinical characteristics and IRGP signature-derived risk score in ESCC (A) and EAC (B). Kaplan-Meier survival curves of survival time in high- and low-risk patients, stratified by the optimal risk score cut-off value of –1.349 in ESCC (C) and –0.955 in EAC (D). (E) Visualization of survival status of patients sorted by increasing risk scores in ESCC and EAC (F). IRGs, immune-related gene; ROC, receiver operating characteristic; IRGPs, immune-related gene pairs; ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma.

and EAC (*Figure 5G,5H*). We used TIMER2.0 to evaluate immune cell infiltration in ESCC and EAC (only significant results are shown) and found that the majority of significant

immune cell subpopulations were positively associated with risk scores in ESCC (*Figure 5I*). On the contrary, most of the significant immune cell subpopulations were negatively

Page 8 of 16

Cao et al. Prognostic immune-related gene pair signature in EC
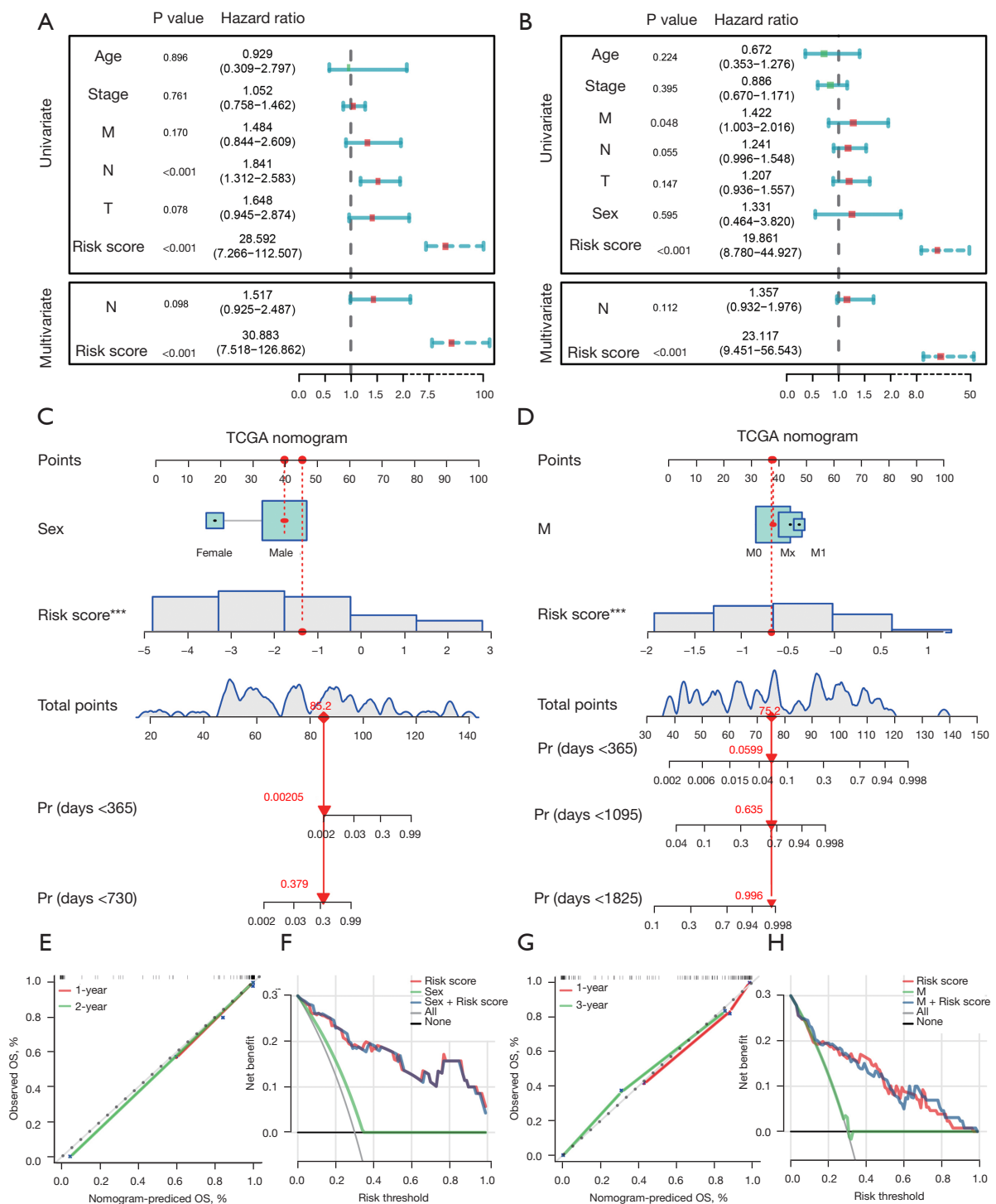
Figure 4 Evaluation of the prognostic value of IRGP signature-derived risk score and construction of nomograms in TCGA cohort. (A) Forest plot of univariate and multivariate Cox regression analysis results of IRGP signatures and clinical characteristics in ESCC and EAC (B). Nomogram predicts the probability of patient mortality based on IRGP signature and clinical variables in ESCC (C) and EAC (D), followed by respective calibration curves (E,G). Decision curve analyses of the nomograms based on IRGP signature for 3-year overall survival in ESCC (F) and EAC (H). (F) The calibration plot for internal validation of the nomogram. IRGPs, immune-related gene pairs; TCGA, The Cancer Genome Atlas; ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma.
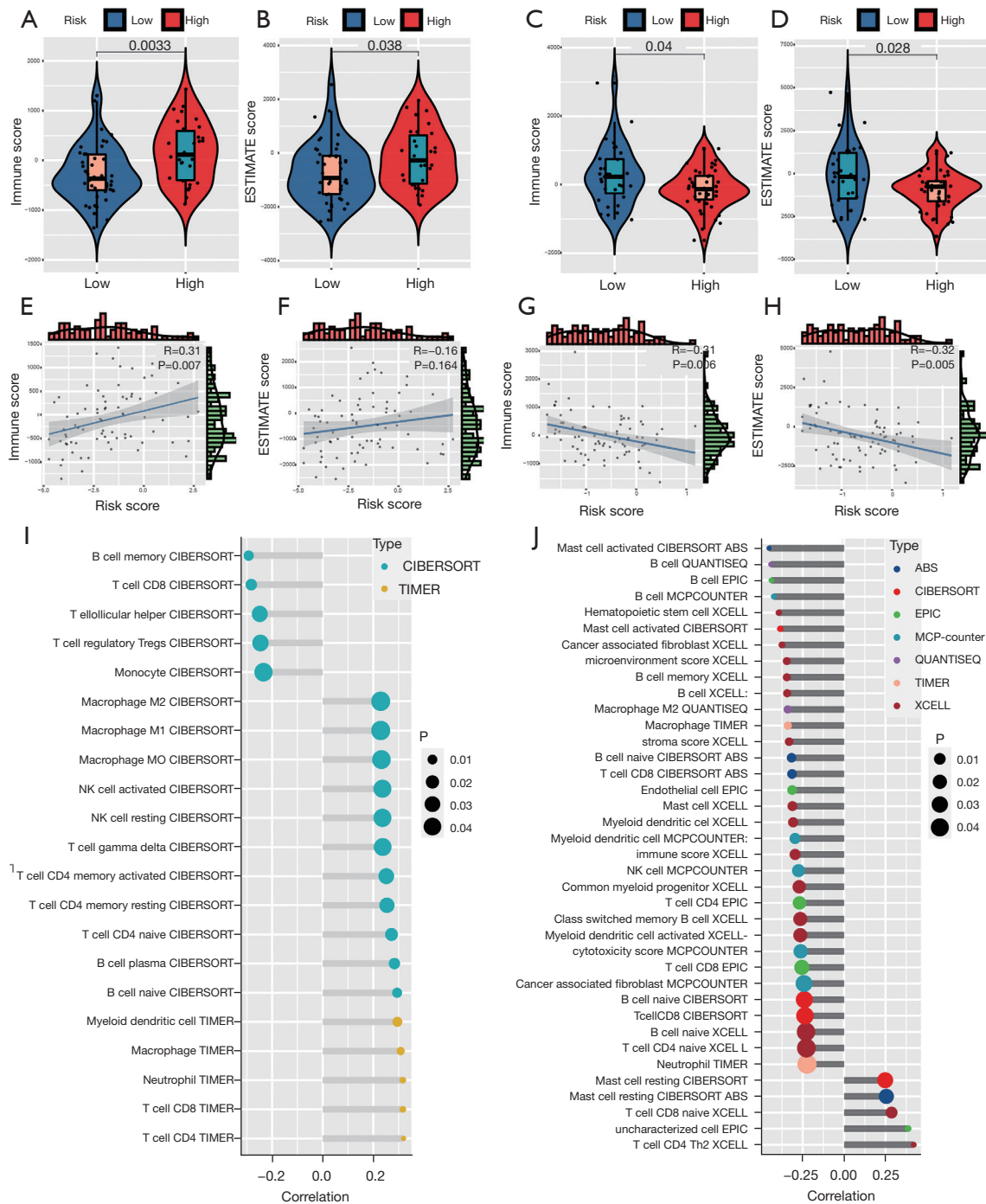
Figure 5 Correlation of the signature of IRGPs with immunity by ESTIMATE algorithm. In ESCC, the high-risk group shows a significantly elevated immune score (A) and ESTIMATE score (B). In EAC, both immune score (C) and ESTIMATE score (D) were significantly downregulated in the high-risk group. Correlation between IRGP signature-derived risk score and immune score in ESCC (E) and EAC (G). Correlation between risk score and ESTIMATE score in ESCC (F) and EAC (H). TIMER2.0 was used to reveal (I) the profile of immune cell subsets significantly associated with ESCC and (J) the profile of immune cell subsets significantly associated with EAC. IRGPs, immune-related gene pairs; ESTIMATE, Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data; ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma.

Page 10 of 16

Cao et al. Prognostic immune-related gene pair signature in EC

associated with risk scores in EAC (*Figure 5J*).

Association between IRGP signatures and potential biomarkers of ICI response

ICIs have emerged as an encouraging therapeutic option for nonresectable EC. Currently, there are several biomarkers which may help identify patients likely to benefit from ICIs, including HLAs, MMR, and immune checkpoint molecules (39,40). Defective HLA may lead to immune tolerance and resistance to ICIs (41). Solid tumors lacking MMR genes are usually highly immunogenic and exhibit extensive infiltrating T cells, which make them sensitive to ICIs (39). We observed that most known HLAs were associated with high-risk groups in ESCC (*Figure 6A*), whereas several HLAs were associated with the high-risk group in the inverse direction in EAC (*Figure 6B*). We checked the correlation between the risk score and 4 essential MMR genes (MSH6, MLH1, PMS2, and MSH2). No significant correlations were found in ESCC (*Figure 6C,6D*). However, the expression levels of MSH2, MSH6, and PMS2 in the low-risk group were significantly decreased when compared with the high-risk group (*Figure 6E*) and correlated with the risk score in EAC (*Figure 6F*). We also explored the relationship between IRGP signature-derived risk score and common immune checkpoint molecules, including cytotoxic T-lymphocyte-associated antigen 4 (CTLA4), hepatitis A virus cellular receptor 2/T-cell immunoglobulin mucin receptor 3 (HAVCR2/TIM3), lymphocyte activating 3 (LAG3), programmed cell death 1 (PDCD1), and PDCD1 ligand 1 (PDCD1L1, also known as CD274), and PDCD1 ligand 2 (PDCD1lG2) (42). As displayed in *Figure 7A*, the levels of PDCD1 and HAVCR2 were significantly upregulated in the high-risk group in ESCC. However, significantly increased CTLA4 levels were seen in the low-risk group in EAC (*Figure 7B*). Correlation matrixes were also exhibited for ESCC (*Figure 7C*) and EAC (*Figure 7D*).

### Signaling pathways associated with the risk score

In the present study, the IRGP signatures were highly correlated with prognosis and TIME. Therefore, we further investigated the crucial signaling pathways involved in ESCC and EAC. GSEA, with reference to the immunologic signature gene sets (C7), revealed the immune signaling pathways associated with either low- or high-risk groups in both ESCC (*Figure 8A*) and EAC (*Figure 8B*). For instance, differences in cell types, status,

and perturbations were found between the high- and low-risk groups in EAC, including memory CD4 T cells, NK cells, CD4 T cells, and naïve CD4 T cells, suggesting an essential role for these immune cell subsets in EAC TIME (*Figure 8B*). KEGG pathway enrichment analysis unveiled that the DEGs between the high- and low-risk groups were significantly enriched in some fundamental signaling pathways, including the insulin, notch, autophagy, and Wnt signaling pathways in ESCC (*Figure 8C*). Not surprisingly, the DEGs of the 2 prognostic risk groups were significantly enriched in different signaling pathways in EAC. Among them, the mismatch repair and base excision repair pathways were most significantly associated with the high-risk group, while the most significantly associated pathways in the low-risk group were the intestinal immune network for immunoglobulin A (IgA) production and B cell receptor signaling pathways (*Figure 8D*).

Discussion

EC is a life-threatening malignancy with a high fatality rate. In particular, the prognosis of patients with locally advanced, unresectable, and metastatic disease, which account for 50–60% of cases, is extremely poor (2). Despite the emergence of definitive concurrent chemoradiotherapy (CRT), targeted therapies, and ICIs, a better understanding of the biology of EC and optimal use of multidisciplinary treatment options are needed to help individual patients achieve the best possible outcomes. ICIs targeting PD-1, PD-L1, and CTLA4 have significantly prolonged survival time in a proportion of EC patients. However, it remains challenging to discriminate patients sensitive to ICIs from those exhibiting innate resistance. Therefore, biomarkers predicting patient subpopulations appropriate for ICIs warrant intensive investigation.

We applied a rank-based algorithm in the current study, ranking IRGs by pairwise comparison, to choose informative IRG pairs to establish prognostic signature. We generated a 19- and 17-IRGP signature for ESCC and EAC, respectively. The effectiveness of this method has been verified in a previous study (43). The prognostic value of signatures derived from gene pairs have been confirmed in several cancer types (21-23,25). Li and colleagues reported an individualized immune prognostic signature of IRGPs that divided patients into 2 risk groups with significantly different survival (21). The signature composed of 20 IRGPs was significantly associated with OS in serous ovarian carcinoma (25). Recently, a signature of 14 IRGPs was
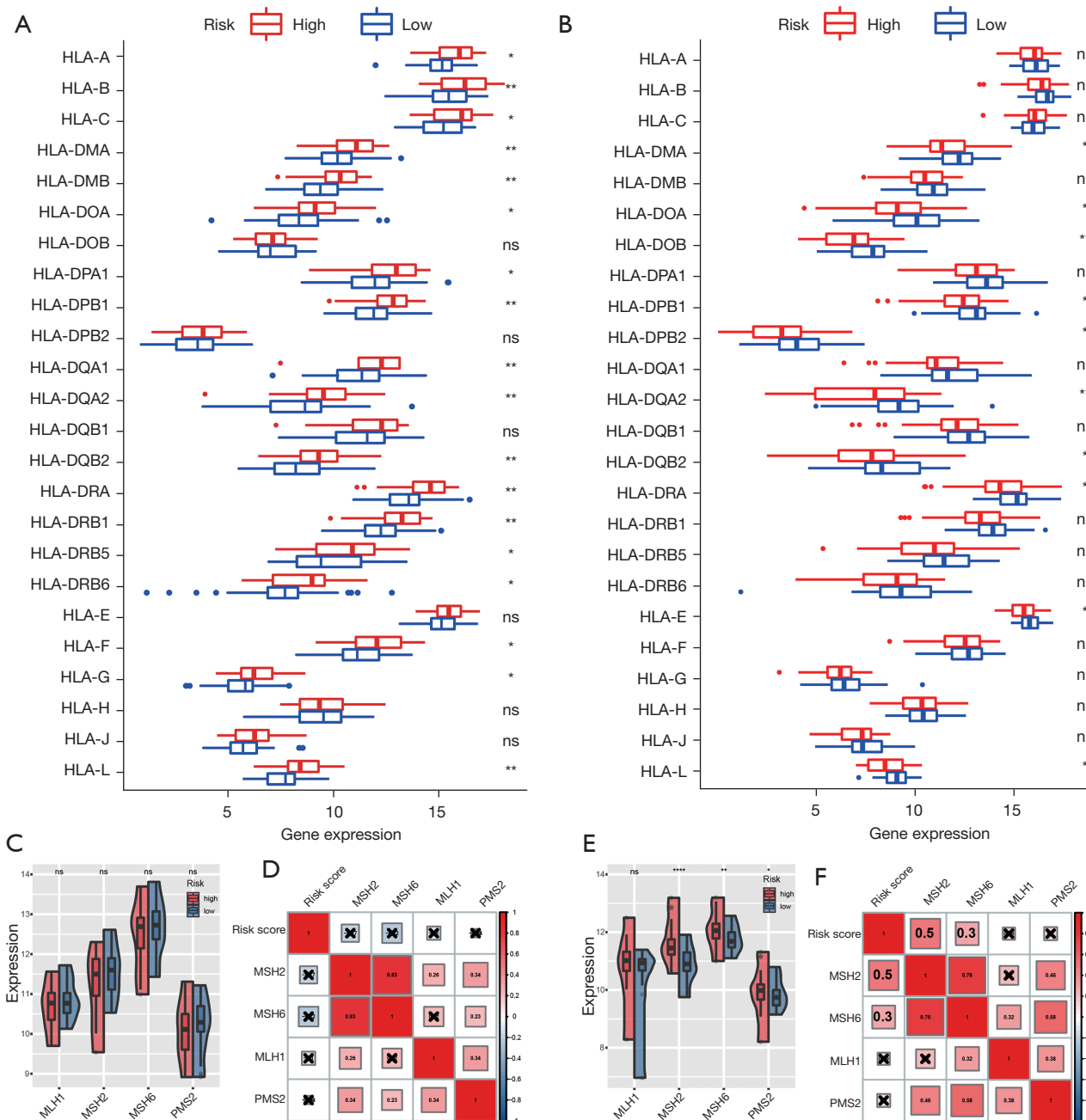
Figure 6 Association of risk score with expression levels of HLA and MMR genes. Expression of the HLA gene family in the low- and high-risk groups in ESCC (A) and EAC (B). Comparison of MMR gene expression between risk groups defined by risk score and correlation between risk score and HLA gene expression in ESCC (C,D) and EAC (E,F). *P<0.05, **P<0.01. ns, not significant; HLA, human leukocyte antigens; MMR, mismatch repair; ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma.

shown to predict prognosis accurately in osteosarcoma (23). The reliability and robustness of an IRGP-derived signature were also validated in lung adenocarcinoma (22). Our results showed that the IRGP signatures for ESCC and

EAC were different; however, it is not surprising. Although ESCC and EAC arise from the same origin, they are two different diseases regarding etiology, risk factors, molecular characterization, histology of the lesion, predilection sites,
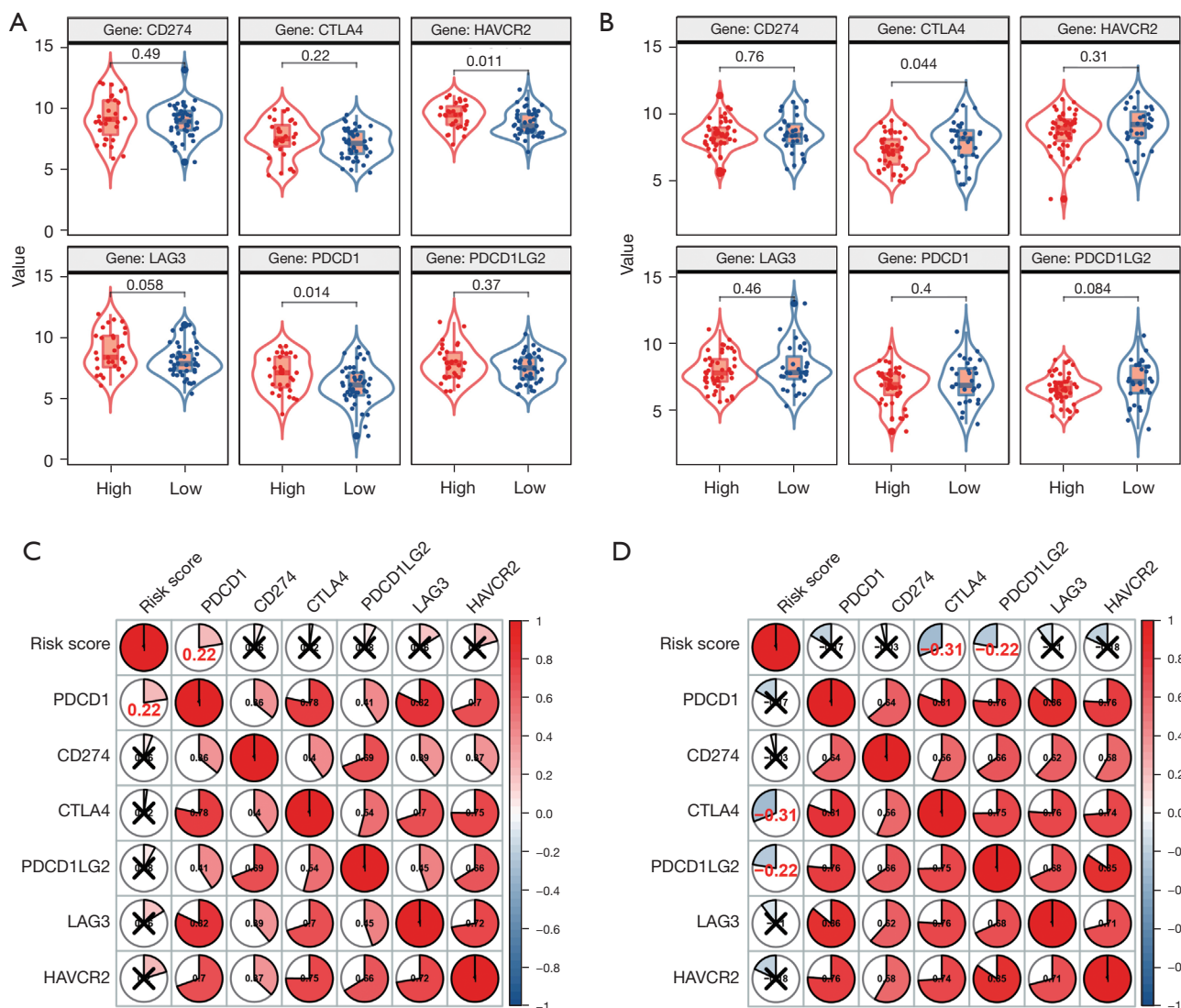
Figure 7 Association of risk score and immune checkpoint genes. The differential expression of immune checkpoint genes between high- and low-risk groups in ESCC (A) and EAC (B). Correlation matrix between immune checkpoint genes and risk score in ESCC (C) and EAC (D). ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma.

therapeutic regime, and geographical distribution (44). There were several advantages of using a gene pair-based strategy. First, as the signature was rooted in the relative ranking of gene expression levels, no data normalization was needed. Second, pairwise comparisons were performed for all genes within a sample; therefore, the signature was able to evaluate the survival of individual cancer patients. In other words, the prognostic signature did not need to account for the batch effects of different platforms and could bypass data scaling and normalization. Overall, gene pair-based prognostic signatures are robust and easy to use.

Next, we evaluated the association between the IRGP signature-based risk scores and immune activity and immune cell infiltration in the tumor microenvironment. Interestingly, the 2 types of EC exhibited distinct immune features. Both the immune score and ESTIMATE score were positively associated with risk scores in ESCC, but a negative association was observed in EAC. Moreover, a number of immune cell subsets were correlated with risk scores in ESCC. However, infiltration of most immune subsets dropped while the risk score increased in EAC. Differential immune cell infiltration patterns in the 2 EC
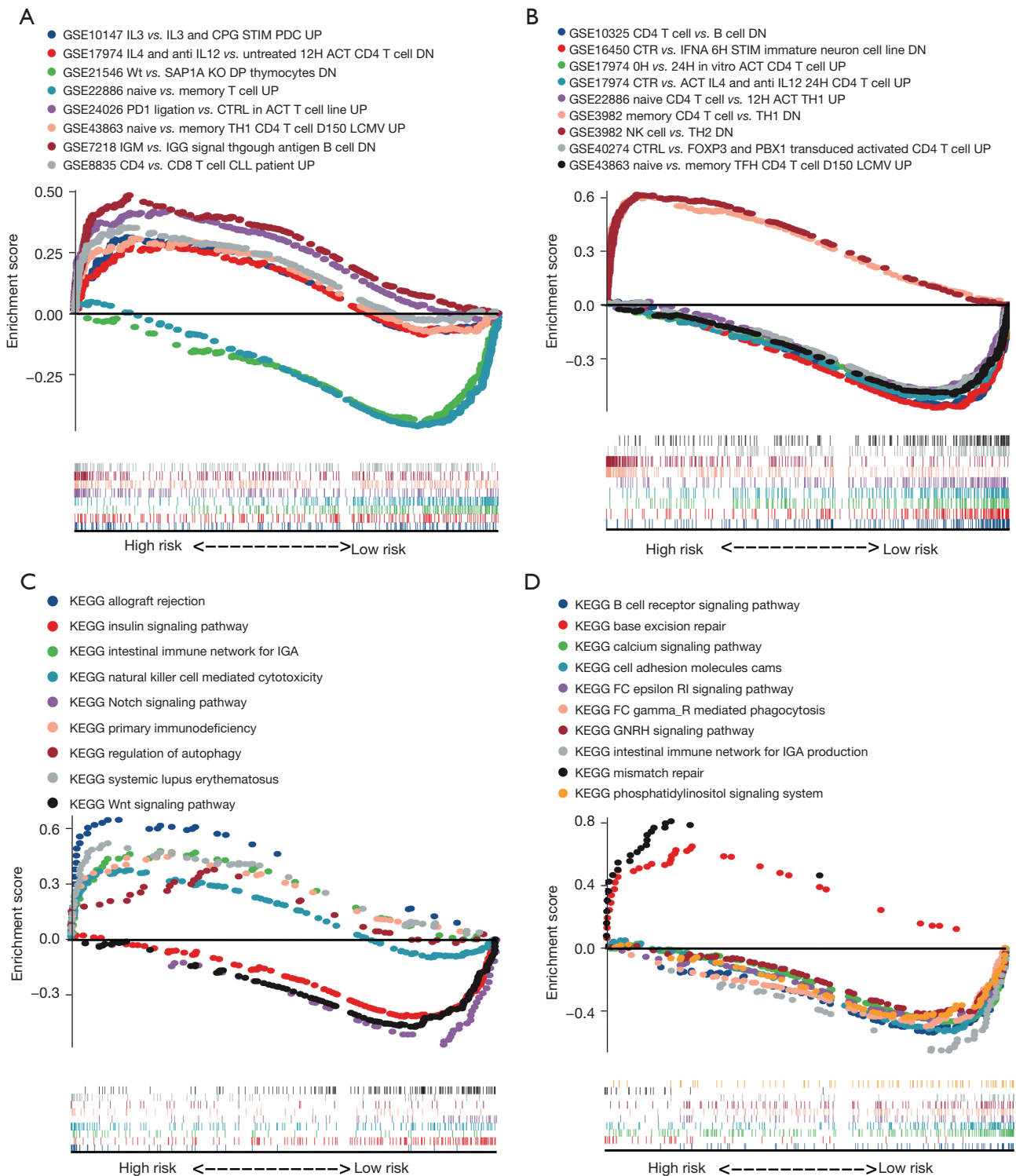
Figure 8 Gene set enrichment analysis in the TCGA cohort. The significantly enriched immune cells subset in ESCC (A) and EAC (B). The significantly enriched KEGG signaling pathways in ESCC (C) and EAC (D). TCGA, The Cancer Genome Atlas; ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma; KEGG, Kyoto Encyclopedia of Genes and Genomes.

types may help explain their different response to ICIs. ESCC patients receiving a selective PD-1 inhibitor (i.e., nivolumab) showed improved disease-free survival in the adjuvant therapy setting (CHECKMATE-557) (45). In first-line treatment, PD-1-inhibitor (pembrolizumab) combined with chemotherapy significantly prolonged OS time for ESCC patients (KEYNOTE-590) (45). Similar encouraging results were observed in second-line treatment (45). However, EAC patients benefited from the same ICIs to a lesser extent (45).

We were also interested in whether the IRGP signatures had the potential to predict immunotherapy response. To answer this question, we examined the relationship between the risk scores and HLA, MMR, and immune checkpoint genes, which have been considered potential biomarkers of response to ICIs. Regarding ESCC, we found that the majority of HLA types were significantly upregulated in the high-risk group compared to the low-risk group. No significant differences in MMR were found between the two groups. A significant increase in HAVCR2 and PDCD1 was also observed in the high-risk group. These results suggested that the ESCC patients in the high-risk group with high HLA levels and increased expression of immune checkpoint molecules were more likely to respond to ICIs than those in the low-risk group. HLA encodes intracellular peptides on the cell surface to be recognized by T cell receptors, taking responsibility for neoantigen presentation and cytolytic T cell activity. Lack of HLA may be detrimental to the ability of cells to present neoantigens and cause immune tolerance (41). HLA-I loss of heterogeneity (LOH) was significantly related to poor outcomes in patients with non-squamous non-small cell lung carcinoma receiving ICI treatment (40). Moreover, HLA-I LOH showed a linear correlation with PD-L1 expression (40). In EAC, significantly increased immune score and HLA genes, decreased levels of MSH2, MSH6, and PMS2, and increased CTLA4 were seen in the low-risk group when compared with the high-risk group. These results collectively suggested that the low-risk group defined by the IRGP signature was highly immunogenic and more likely to benefit from immunotherapy in EAC.

Our study contains some limitations that should be noted. First, as a retrospective study, the sample size was relatively small. Second, sampling bias might have been an issue due to intratumor molecular heterogeneity. Third, external validation by other independent EC cohorts is warranted, although this ranking-based method was validated in ESCC and EAC. Forth, all the findings were based on bioinformatic analysis. Experiments should be performed to validate the predictive accuracy of IRGPs and their association with the infiltration of immune cells in the future. Finally, due to the unavailability of therapeutic regimes on patients, we failed to integrate this important information into the nomogram. Therefore, our findings should be explained cautiously.

Overall, we successfully developed IRGP-based signatures using a novel rank-based pairwise comparison algorithm, showing remarkable performance in predicting prognosis in ESCC and EAC. The IRGPs were tightly associated with the components of immune cells in the tumor microenvironment and held great promise in predicting sensitivity to ICIs in ESCC and EAC. Prospective studies are indispensable to validate the predictive accuracy of the signatures further before clinical application.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://dx.doi.org/10.21037/atm-21-5217

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/atm-21-5217). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons

References

1. Abrams JA, Sharaiha RZ, Gonsalves L, et al. Dating the rise of esophageal adenocarcinoma: analysis of Connecticut Tumor Registry data, 1940-2007. Cancer Epidemiol Biomarkers Prev 2011;20:183-6.
2. van Rossum PSN, Mohammad NH, Vleggaar FP, et al. Treatment for unresectable or metastatic oesophageal cancer: current evidence and trends. Nat Rev Gastroenterol Hepatol 2018;15:235-49.
3. Fridman WH, Zitvogel L, Sautes-Fridman C, et al. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol 2017;14:717-34.
4. Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett 2017;387:61-8.
5. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nature Medicine 2018;24:541-50.
6. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. Nat Rev Cancer 2020;20:662-80.
7. Galon J, Angell HK, Bedognetti D, et al. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. Immunity 2013;39:11-26.
8. Blank CU, Haanen JB, Ribas A, et al. Cancer Immunology. The "cancer immunogram". Science 2016;352:658-60.
9. Lin EW, Karakasheva TA, Hicks PD, et al. The tumor microenvironment in esophageal cancer. Oncogene 2016;35:5337-49.
10. Lv L, Pan K, Li XD, et al. The accumulation and prognosis value of tumor infiltrating IL-17 producing cells in esophageal squamous cell carcinoma. PLoS One 2011;6:e18219.
11. Wang JH, Huang ST, Zhang L, et al. Combined prognostic value of the cancer stem cell markers CD47 and CD133 in esophageal squamous cell carcinoma. Cancer Med 2019;8:1315-25.
12. Zhang Z, Chen C, Fang Y, et al. Development of a prognostic signature for esophageal cancer based on nine immune related genes. BMC Cancer 2021;21:113.
13. Zheng W, Chen C, Yu J, et al. An energy metabolism-based eight-gene signature correlates with the clinical outcome of esophagus carcinoma. BMC Cancer 2021;21:345.
14. Li W, Liu J, Zhao H. Identification of a nomogram based on long non-coding RNA to improve prognosis prediction of esophageal squamous cell carcinoma. Aging (Albany NY) 2020;12:1512-26.
15. Wang L, Wei Q, Zhang M, et al. Identification of the prognostic value of immune gene signature and infiltrating immune cells for esophageal cancer patients. Int Immunopharmacol 2020;87:106795.
16. Hao D, He S, Harada K, et al. Integrated genomic profiling and modelling for risk stratification in patients with advanced oesophagogastric adenocarcinoma. Gut 2021;70:2055-65.
17. Gao J, Tang T, Zhang B, et al. A Prognostic Signature Based on Immunogenomic Profiling Offers Guidance for Esophageal Squamous Cell Cancer Treatment. Front Oncol 2021;11:603634.
18. Mao Y, Fu Z, Zhang Y, et al. A seven-lncRNA signature predicts overall survival in esophageal squamous cell carcinoma. Sci Rep 2018;8:8823.
19. Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 2010;11:733-9.
20. Heinäniemi M, Nykter M, Kramer R, et al. Gene-pair expression signatures reveal lineage control. Nat Methods 2013;10:577-83.
21. Li B, Cui Y, Diehn M, et al. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. JAMA Oncol 2017;3:1529-37.
22. Jiang X, Gao Y, Zhang N, et al. Establishment of Immune-related Gene Pair Signature to Predict Lung Adenocarcinoma Prognosis. Cell Transplant 2020;29:963689720977131.
23. Li LQ, Zhang LH, Zhang Y, et al. Construction of immune-related gene pairs signature to predict the overall survival of osteosarcoma patients. Aging (Albany NY) 2020;12:22906-26.
24. Hong W, Liang L, Gu Y, et al. Immune-Related lncRNA to Construct Novel Signature and Predict the Immune Landscape of Human Hepatocellular Carcinoma. Mol Ther Nucleic Acids 2020;22:937-47.
25. Zhang L, Zhu P, Tong Y, et al. An immune-related gene pairs signature predicts overall survival in serous ovarian

Page 16 of 16

Cao et al. Prognostic immune-related gene pair signature in EC

carcinoma. Onco Targets Ther 2019;12:7005-14.

26. Bhattacharya S, Andorf S, Gomes L, et al. ImmPort: disseminating data to the public for the future of immunology. Immunol Res 2014;58:234-9.

27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.

28. Jiang Y, He Y, Zhang H. Variable Selection with Prior Information for Generalized Linear Models via the Prior LASSO Method. J Am Stat Assoc 2016;111:355-76.

29. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385-95.

30. Liu Y, Wu L, Ao H, et al. Prognostic implications of autophagy-associated gene signatures in non-small cell lung cancer. Aging (Albany NY) 2019;11:11440-62.

31. Liu Y, Guo X, Zhao M, et al. Contributions and prognostic values of m(6) A RNA methylation regulators in non-small-cell lung cancer. J Cell Physiol 2020;235:6043-57.

32. Zhu J, Liu Y, Zhao M, et al. Identification of downstream signaling cascades of ACK1 and prognostic classifiers in non-small cell lung cancer. Aging (Albany NY) 2021;13:4482-502.

33. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.

34. Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res 2020;48:W509-W14.

35. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000;56:337-44.

36. Iasonos A, Schrag D, Raj GV, et al. How To Build and Interpret a Nomogram for Cancer Prognosis. Journal of Clinical Oncology 2008;26:1364-70.

37. Vickers AJ, Cronin AM, Elkin EB, et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 2008;8:53.

38. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Medical Decision Making 2006;26:565-74.

39. So A. Bladder cancer, ESMO 2016. Can Urol Assoc J 2016;10:S224-S6.

40. Montesion M, Murugesan K, Jin DX, et al. Somatic HLA Class I Loss Is a Widespread Mechanism of Immune Evasion Which Refines the Use of Tumor Mutational Burden as a Biomarker of Checkpoint Inhibitor Response. Cancer Discov 2021;11:282-92.

41. McGranahan N, Rosenthal R, Hiley CT, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell 2017;171:1259-71.e11.

42. Khalil DN, Smith EL, Brentjens RJ, et al. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. Nat Rev Clin Oncol 2016;13:273-90.

43. Popovici V, Budinska E, Tejpar S, et al. Identification of a poor-prognosis BRAF-mutant-like population of patients with colon cancer. J Clin Oncol 2012;30:1288-95.

44. Smyth EC, Lagergren J, Fitzgerald RC, et al. Oesophageal cancer. Nat Rev Dis Primers 2017;3:17048.

45. Högner A, Thuss-Patience P. Immune Checkpoint Inhibition in Oesophago-Gastric Carcinoma. Pharmaceuticals (Basel) 2021;14:151.

Table S1 Characteristics of ESCC and EAC patients

Characteristics	ESCC, n (%)	EAC, n (%)
Total	75 (100.0)	78 (100.0)
Survival		
Alive	54 (72.0)	39 (50.0)
Dead	21 (28.0)	39 (50.0)
Age		
≤65 years	58 (77.3)	36 (46.2)
>65 years	17 (22.7)	42 (53.8)
Sex		
Male	65 (86.7)	67 (85.9)
Female	10 (13.3)	11 (14.1)
Stage		
Stage I	1 (1.3)	2 (2.6)
Stage II	18 (24.0)	3 (3.8)
Stage III	10 (13.3)	7 (9.0)
Stage IV	4 (5.3)	5 (6.4)
Stage X	42 (56.0)	61 (78.2)
Tumor (T)		
T1	7 (9.3)	20 (25.6)
T2	26 (34.7)	10 (12.8)
T3	37 (49.3)	35 (44.9)
T4	3 (4.0)	1 (1.3)
Tx	2 (2.7)	12 (15.4)
Metastasis (M)		
M0	66 (88.0)	50 (64.1)
M1	3 (4.0)	5 (6.4)
Mx	6 (8.0)	23 (29.5)
Lymph node (N)		
N0	43 (57.3)	19 (24.4)
N1	23 (30.7)	37 (47.4)
N2	5 (6.7)	4 (5.1)
N3	1 (1.3)	5 (6.4)
N4	3 (4.0)	13 (16.7)

Table S2 Lasso regression coefficients of IRGPs in ESCC and EAC

Gene pairs	Coefficient
ESCC	
ADCYAP1R1|EPO	0.39890181
ADM|JAG2	0.60740626
AGTR2|SEMG1	0.39046504
AR|SCG2	−0.7027797
CCL4L2|IL31RA	0.40931681
CCL5|IGHG3	0.45963646
CCR8|IGLV1-36	−0.3796451
CSF3|HNF4G	−0.3826402
CTSE|KIR2DL4	−0.2659804
DEFA5|FCN2	0.78567696
DKK1|NR4A3	−0.3713388
EGFR|STAT1	−0.6977208
GCGR|GLP1R	−1.5148802
IFNG|IL13RA2	0.13426929
IL19|NR0B2	−1.1132949
IL33|LIFR	0.23463465
JAG2|PLAUR	−0.3432173
PDGFD|TGFB2	0.52901538
SEMA5B|TGFB2	0.53657821
EAC	
ACKR1|IDO1	−0.1194319
AGTR1|FGF5	−0.1304876
BPIFA2|CXCR1	0.03742268
CCL24|OSM	−0.0312613
CCL26|VIPR2	0.35557189
CMTM5|GKN1	0.08567827
CMTM7|CXCL5	−0.2460852
CTSG|TRBV3-1	−0.3191845
CXCL5|IL33	0.10684079
CXCR1|PROK2	−0.2091523
ESM1|PTGER2	0.2462026
FGF5|VIPR2	0.10719466
FPR2|INHBE	−0.2394095
GPER1|IL11	−0.1534362
GPER1|PTGDR2	−0.312505
GPER1|VGF	−0.0160096
RXFP3|TRAV2	0.32288024