A Novel Signature of 23 Immunity-related Gene Pairs is Prognostic of Cutaneous Melanoma

Yanan Xue
zhejiang university

Yinan Xue
Zhejiang normal university

Zhengcai Wang
Zhejiang University

Yongzhen Mo
Xiangya Hospital Central South University

Pinyan Wang
Central South University Third Xiangya Hospital

Weiqiang Tan (✉ imgrout@163.com)
Zhejiang University https://orcid.org/0000-0002-1315-4629

Primary research

Keywords: Immune related gene pairs (IRGPs), CM, immune micro-environment, prognosis

DOI: https://doi.org/10.21203/rs.3.rs-38060/v1

License: ☕️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background: We aimed to identify immune-related signature for predicting cutaneous melanoma (CM) prognosis.

Methods: We used TCGA samples (n=471) to develop the best 23 Immune related gene pairs (23-IRGP) prognostic signature and divided patients into high- and low-immune risk group in TCGA dataset and validation datasets: GSE65904 (n=214), GSE59455 (n=141), and GSE22153 (n=79).

Results: 23-IRGP presented precise ability in cutaneous melanoma (CM) which high-risk groups showed poor prognosis and indicated significant predict power in immune micro-environment and biological analysis as well.

Conclusions: we established a novel promising prognostic model in CM and built the bridge between immune micro-environment and CM patient results. This approach can be applied to discover the signatures in other diseases without technical bias from different platforms.

Introduction

The incidence of cutaneous melanoma (CM) continues to increase more rapidly than that of any other cancer, accounting for about 132,000 new case and 65,000 deaths worldwide annually[1]. CM is recognized as the most lethal form of skin cancer and constitutes a serious public health concern. The primary clinical feature of CM is early stage metastasis and most diagnoses are made in the terminal stage, thus surgical results are often poor. At present, chemotherapy is the first-line treatment for CM[2], although many cases respond poorly to such regimens due to the high prevalence of adverse drug reactions and resistance to chemotherapeutic agents.

CM is associated with strong immunogenicity, thus immunotherapy presents a promising treatment alternative[3]. Initial trials of immunotherapies against CM with the use of interferon-α[4]and high dose interleukin-2 for advanced cases[5] reported successful results. At present, immune checkpoint inhibitors, such monoclonal antibodies against cytotoxic T-lymphocyte-associated protein (CTLA-4)[6] and programmed cell death protein 1 (PD1)[7], have resulted in meaningful clinical outcomes against advanced melanoma, as demonstrated by improved survival and a greater possibility of a curative effect for a continuously increasing proportion of patients.

Broad efforts have strived to identify novel prognostic biomarkers, but the clinical behavior of CM is unpredictably rendering the prediction of survival time and tumor stage particularly difficult[8]. Therefore, novel biomarkers and patient-tailored treatments and are needed, especially for patients at higher risks of recurrence. Although the immune system plays essential roles in the development and progression of CM[9], few immunity-related genes (IRGs) have been identified for use as markers of prognosis. Moreover, many studies have reported that genetic mutations as well as interactions between the tumor and its microenvironment will impact the biological features and malignant potential of CM. Considering that
many immunity-related treatments have shown significant clinical therapeutic effects, identification of the interactions between cancer cells and the host immune response is especially important[10, 11].

In this study, 23 immunity-related genes related to CM were identified from whole transcriptome sequencing (RNA-seq) data retrieved from The Cancer Genome Atlas (TCGA)[12] and the ImmPort dataset[13]. Then, three microarray datasets (GSE65904, GSE59455, and GSE22153) were selected from the Gene Expression Omnibus (GEO) database[14] to verify the usefulness of a signature of 23 immunity-related gene pairs (IRGPs) for the prognosis of CM. Also, possible relationships between clinical pathological factors and the prognostic signature were explored in order to validate the predictive effectiveness and accuracy of the IRGPs. Finally, immune cell infiltration, tumor micro-environment and biological function analyses of different risk groups based on the 23 identified IRGPs were performed.

Materials And Methods

Data sources of CM patients

Four independent RNA-seq datasets and clinical data from diverse high-throughput sequencing platforms were comprehensively analyzed in this retrospective study. A CM dataset (n = 471) was downloaded from the TCGA data portal (https://portal.gdc.cancer.gov) and randomly assigned to a training dataset (n = 236) or a testing dataset (n = 235). Also, the datasets GSE65904 (n = 214), GSE59455 (n = 141), and GSE22153 (n = 79) were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) for validation of the IRGP signature. In total, 905 samples were available for analysis. A file of 1534 IRGs that was downloaded from the ImmPort database (https://immport.niaid.nih.gov) and the CIBERSORT analytical tool (https://cibersort.stanford.edu/) were used to provide an estimation of the abundances of 22 distinct cell types in a mixed cell population based on gene expression data.

Data preprocessing

All data were preprocessed using R software (version 3.6.2). If more than one probe matched the same target gene, the average value of the probe was calculated to replace the single gene expression level. If multiple samples were from the same patient, the average value of each gene was calculated to replace the level of gene expression.

Establishment of prognostic IRGPs based on the TCGA dataset

The TCGA CM dataset was divided into a training group or a testing group, while the datasets GSE65904, GSE59455, and GSE2215 were employed as the validation group. IRGs with relatively high variation (median absolute deviation >0.5) were extracted from the platforms as described previously. For pairwise comparison of a specific sample, two IRGs were paired off and if expression of the first IRG was greater than that of the second, the two IRGs were merged as an IRGP and assigned a score of 1, otherwise the score was 0. Then, IRGPs with score = 1 or 0 in more than 80% of the specimens both in the training and testing groups were selected as potential prognostic IRGPs. Based on the results of the log-rank test,
IRGPs with a false discovery rate (FDR) of < 0.001 (n = 23) were retained and entered into a least absolute shrinkage and selection operator (LASSO) penalized Cox regression model (iteration = 1000). The median value of the risk score was used as a cut-off in order to divide the patients into high and low immune risk groups. Next, a heat map, risk score map, state map of overall survival (OS), and 1-, 3-, and 5-year receiver operating characteristic (ROC) curves were created, and the concordance (C)-index was calculated. Then, the IRGPs were integrated with other clinical factors to construct a nomogram and a calibration curve. Finally, univariate and multivariate Cox regression analyses were performed to determine whether the 23 IRGPs were independent from other clinical parameters.

**Verification and assessment of the IRGP signature for prediction of OS**

The TCGA testing dataset and three microarray data files were selected to validate the signature of 23 IRGPs. Every dataset was stratified into the high or low immune risk group based on the cut-off value of the prognostic signature. Next, the log-rank test and Cox analysis were performed and a graph of OS was created in order to calculate the C-index between the high and low immune risk groups in each dataset. Finally, the signature of the 23 IRGPs was compared to the 1-, 3-, and 5-year area under the ROC curves (AUCs) and the C-indices.

**Infiltration of immune cells in CM**

The CIBERSORT analytical tool [15] was used to explore the enrichment of immune cells of two CM groups. CIBERSORT is a tool for deconvolution of the expression matrix of immune cell subtypes based on the principle of linear support vector regression, which can estimate the enrichment of various immune cell types in CM. Then, based on the CIBERSORT results, a radar chart was constructed of significantly differentially expressed immunity-related genes between the two risk groups. All procedures were performed using R software (version 3.6.2).

**Biological function analysis of the 23 IRGPs**

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analyses of two risk groups were performed using the R bioconductor package “fgsea.” GO analysis and KEGG pathways files (c5.all.v7.0.symbols.gmt and c2.cp.kegg.v7.1.symbols.gmt) were downloaded from Gene Set Enrichment Analysis (GSEA) website (https://www.gsea-msigdb.org/gsea/datasets.jsp)[16]. Gene sets with a FDR-adjusted probability (p) value of <0.05 were considered statistically significant.

**Statistical analysis**

All statistical analyses were performed using the software packages R (version 3.6.2, www.r-project.org) and Prism 8 (GraphPad Software Inc., San Diego, CA, USA). OS curves were plotted using the R packages “survival” and “survminer.” A heat map of the IRGPs, risk score map, and OS status graph were created with the R package “pheatmap.” A model of prognostic IRGPs was established with the R package “glmnet”[17]. Univariate and multivariate Cox regression analyses were performed using the R package “survival.” ROC curves were constructed using the R package “survivor.” A nomogram and calibration
curve were plotted with the R packages “rms,” “foreign,” and “survival.” The C-index was computed with the R package “Hmisc.” The infiltration of immune cells was processed with the R packages “e701,” “limma,” and “fmsb”[18]. The tumor-environment plot based on the R package “estimate”[19], and 6 single expression were present by R package “ggpubr”. GO and KEGG analyses were conducted with the R package “fgsea.”

**Results**

**Establishment and definition of the IRGP signature**

A flowchart of the establishment and validation of 23 IRGPs is presented in Figure S1. The TCGA dataset was divided into a training dataset (n = 236) and a testing dataset (n = 235). Filter analysis was applied to establish a prognostic model of 1,811 unique IRGs that were obtained from the ImmPort database (accessed on January 4, 2020). In total, 620 IRGs with a median absolute deviation of > 0.5 were common among all datasets. After removing IRGPs with a score of 0 or 1 in less than 20% or more than 80% of the samples in the TCGA training and testing datasets, a total of 74,750 IRGPs had remained. Of these, 6,800 prognostic IRGPs were significantly associated with OS (FDR < 0.001), as determined with the log-rank test. Finally, 23 IRGPs were selected for the LASSO penalized Cox regression model (Figure 1a), including 46 unique IRGs, of which most coded for molecules related to antimicrobials and cytokines (Table 1). What’s more, we analyzed the Principal Component Analysis (PCA) of 46 unique IRGs by GEPIA web (gepia.cancer-pku.cn/), and found these genes expressed in TCGA tumor samples were independent parts, compared with normal skin that had sun explored or not (Figure S2).

The 23 IRGPs were used to calculate the risk score to predict the 5-year OS rate of each patient in the training cohort. Analysis of the 5-year dependent ROC curve revealed that the best cut-off value of the 23 IRGPs to stratify patients in the training cohort into the high or low immune risk group was -0.674 (Figure 2a). The data suggested that the high risk group had a higher risk index than the low risk group. A higher risk score indicates a higher number of deaths (Figure 1b, c), indicating that OS was significantly poorer for the high risk group than the low risk group (p < 0.001) (Figure 1e and Table S1). As shown in Figure 1d, the AUC values[20] of the 1-, 3-, and 5-year OS rates of the training cohort were 0.909, 0.901, and 0.912, respectively, and the C-index was 0.775 (95% confidence interval [CI] = 0.748–0.802) (Figure 4e). A nomogram of OS was created by combining all of the clinicopathological factors such as age, sex, tumor stage and IRGPs risk score to predict the prognosis of CM (Figure 2c). As we can see, IRGPs risk score made the great contribution to the nomogram, and the 1-, 3-, and 5-year calibration curves (Figure 2b) demonstrated promising predictive ability of the nomogram.

Univariate and multivariate Cox proportional hazards regression analyses of the TCGA dataset were performed to further assess the prognostic accuracy of the IRGPs for other clinical elements. The results of univariate and multivariate Cox analyses showed that the signature of the 23 IRGPs and clinical factors, such as tumor stage, were indeed predictive of prognosis. However, although the IRGP signature was highly predictive of prognosis, the p-value was notably low (Figure 3a, b, Table 2).
Stratified analyses of patient age, tumor stage, and sex were also conducted. First, all patients in the TCGA training dataset were stratified by age into a young dataset (age ≤ 65 years) or an old dataset (age > 65 years). OS was expected to be better for younger patients (p = 0.029, Figure 3c). Then, all patients from the TCGA training dataset were further divided into an early onset dataset (stages I and II) or a later onset dataset (stages III and IV). Similar results were observed for the late dataset (p = 0.002, Figure 3d). Finally, all patients were stratified by sex into a male dataset or a female dataset. As shown in Figure 3e, there was little difference in the OS rate between males and females (p = 0.543), possibly because of the small number of samples.

Collectively, the results showed that the predictive ability of the 23-IRGP signature was independent from other clinical parameters and was predictive of OS of CM patients.

**Verification of the 23-IRGP signature to predict OS**

To determine whether the 23-IRGP signature had consistent prognostic value in different risk groups, the TCGA test datasets GES65904 (n = 214), GSE59455 (n = 141), and GSE54467 (n = 79) were applied for external validation. The risk score of each patient was calculated with the same 23-IRGP prognostic signature. Then, based on the median risk score, the patients were assigned to the low or high immune risk group. Interestingly, OS was poorer for the high immune risk group (Figure 4a–d). The results of multivariate Cox regression analysis (Table 2) showed that the risk score of the 23-IRGP signature was an independent prognostic factor after adjustment for age, sex, and tumor stage in the testing dataset (hazard ratio [HR] = 2.08, 95% CI = 1.41–3.05, p = 0.0002), as well as the datasets GSE65904 (HR = 1.98, 95% CI = 1.49–2.62, p = 1.87E-06), GSE59455 (HR = 1.64, 95% CI = 0.94–2.82, p = 0.042), and GSE22153 (HR = 2.07, 95% CI = 1.16–3.69, p = 0.014). Finally, the C-index values for the testing, GSE65904, GSE59455, and GSE22153 datasets were 0.636 (95% CI = 0.585–0.687), 0.650 (95% CI = 0.609–0.691), 0.557 (95% CI = 0.508–0.606), and 0.610 (95% CI = 0.537–0.683), respectively (Figure 4e).

**Comparison of the IRG prognostic signatures of CM**

The 23-IRGP prognostic signature was also compared with the prognostic signatures of individual IRGs. First, as the TCGA CM data had only one normal sample, the samples from the Genotype-Tissue Expression dataset [21] and TCGA CM dataset were merged. Then, significantly different expressed immunity-related genes were selected for the next procedure. The LASSO penalized Cox regression model was applied to the TCGA clinical dataset and 24 prognostic IRGs were selected for the final risk scoring model (Figure 5a). Most of the 24 prognostic IRGs coded for molecules related to antimicrobials and cytokines. The IRGs significantly stratified the TCGA dataset patients into a low or high risk OS group. The data suggested that the high risk group had a higher risk index than the low risk group, as a higher risk score was associated with a higher risk of death (Figure 5b, c). Moreover, the high-risk group had a significantly poorer OS than the low risk group (p < 0.001) (Figure 5e). As shown in Figure 5d, the AUC values of the 1-, 3-, and 5-year OS rates were 0.731, 0.760, and 0.749, respectively, and the C-index was 0.647 (95% CI = 0.612–0.682) (Figure 5E). Collectively, these results demonstrated that the prognostic
signatures of the IRGs had predictive ability, but with a smaller AUC and lower C-index than the 23-IRGP signature, demonstrating that the 23-IRGP signature was the more precise predictive model in CM.

**Immune cell infiltration—tumor micro-environment and 6 key genes analysis**

Reportedly, the infiltration of immune cells is associated with the prognosis of CM patients. The CIBERSORT analytical tool can be used to estimate the abundances of immune cell subsets and has been used in many previous studies of the cancer microenvironment. Therefore, the CIBERSORT analytical tool was applied to estimate the relative abundances of 22 different immune cells for each patient within the different risk groups. A comparative summary of the CIBERSORT output resulting from these two risk groups is shown in Figure 6a. Immune cells, such as M0, M1, and M2 macrophages, plasma cells, activated CD4+ memory T cells, monocytes, CD8+ T cells, follicular helper T cells, and gamma delta T cells were enriched in different risk groups. As shown in Figure 6b and D, M0 macrophages (p = 0.004) and M2 macrophages (p = 0.003) were significantly highly expressed in the high risk group, while the abundances of M1 macrophages (p = 0.001), activated CD4+ memory T cells (p = 0.005), monocytes (p = 0.047), plasma cells (p = 0.011), CD8+ T cells (p = 0.028), follicular helper T cells (p = 0.017), and gamma delta T cells (p = 0.014) were significantly enriched in the low-risk group (Figure 6c, e–j). Then, we estimated tumor micro-environment (TME) in two different groups and found the high risk group had higher tumor purity with lower immune cells and stromal cells infiltration (Figure 7a). What's more, we were seeing that 3 key targets of immunotherapy: PD1, PDL-1, CTLA-4 were highly expressed in low immune risk group (Figure 7b,c,d), which means the low-immune risk group may have better effect with immunotherapy. Kalaora S et al reported that that the PSMB8 and PSMB9 overexpression was predictive for better survival and improved response to immune-checkpoint inhibitors of melanoma patients[22], which were expressed highly in our study’s low immune risk group (Figure 7e,f). Interestingly, PRAME which as an independent biomarker of metastasis Uveal Melanoma[23] were also significant expressed in high risk group (Figure 7g). Previous results were the same in the Immunohistochemical results downloaded from The Human Protein Atlas dataset (THPA)(Figure 7n-w), as THPA has no result about CTLA4, we only present the other 5 genes expression in melanoma tissue. What's more, in GEPIA, the patients with high PD1, PDL-1, CTLA4, PSMB8 and PSMB9 expression had better OS (Figure 7h-l). cBioPortal (https://www.cbioportal.org/) was used to investigate the mutation, according to the results, PD1, PDL-1, CTLA4, PSMB8 and PSMB9 had 5%, 1.9%, 1.6%, 5% and 4% probability to mutate(Figure S3a), respectively. Poorer OS occured if these genes with mutation (Figure S3b). in addition, in GEPIA analysis, PRAME had no significant effect in OS, which need further study to verify (Figure 7m).

**Biological function analysis in high risk group divided by 23-IRGP signature**

First, GSEA was used to investigate the results of the GO and KEGG pathway analyses between the high and low risk groups with the use of genes that were more highly expressed in the high risk group than the low risk group. According to the results of GO analysis, these genes were positively correlated with skin-related biological functions, such as keratinization, epidermal cell differentiation, keratin filament,
intermediate filament cytoskeleton, and skin development (padj < 0.05). A bubble graph of the 16 GO terms enriched in the high risk group with a padj value of < 0.05 is presented in Figure 8. Information of every GO term is shown in Table S1. As shown in Figure S4, multiple melanoma progression-related pathways, including oxidative phosphorylation[24, 25], retinol metabolism[26-28], and ribosome[29, 30], were significantly upregulated in the high risk group (padj < 0.05). Collectively, these results provided evidence of molecular mechanisms affected by the IRGP signature and, thus, predictive of the prognosis of CM patients.

Discussion

CM is a solid malignant tumor with strong immunogenicity and a rapidly increasing incidence. Since the approval of interferon-α for the treatment of CM in 1995, immunotherapies against CM have received much attention from researchers over the past 25 years[31]. Like many other tumors, immune checkpoint (PD1/PDL1 and CTLA-4) blockade therapy has received great acclaim. In 2011, ipilimumab, which targets CTLA-4, was a major breakthrough for the clinical treatment of CM and was subsequently approved for marketing by the Food and Drug Administration. Ipilimumab is the first antibody drug to prolong OS of patients with metastatic cancer[32]. However, because of the toxicity of ipilimumab, other immune surveillance sites were investigated, which led to phase III clinical studies of the anti-PD-1 antibody drugs pembrolizumab and nivolumab. These drugs were approved for use by the Food and Drug Administration in September and December 2014, respectively. With lower anti-drug resistance and higher clinical safety, anti-PD-1 antibody drugs offer hope to patients with advanced unresectable or metastatic CM[33, 34]. Given that the results of single antibody drugs are limited and there are many links between the occurrence and development of CM, a reasonable multiple-immune therapy strategy may have more prospects in the future[35], So, it is necessary to develop a prognostic signature using IRGs.

As sequencing platforms have technical bias, gene expression data must be preprocessed for standardization, which is especially significant to establish prognostic signatures. In order to achieve the robustness of prognosis prediction without technical bias from different platforms, prognostic signatures of the IRGPs were established by pairwise comparison, which does not require preprocessing for data standardization. The IRGP scores were calculated based on the expression levels of immunity-related genes in the same sample. So, the prognostic signature can not only overcome the batch effects of the different platforms, but also does not require the scaling and normalization of data. As compared with prognostic IRGs that require preprocessing for data standardization, according to the AUC and C-index values, the prediction capability of IRGPs is more promising. Moreover, this approach has been reported as robust in other cancer-related studies[36, 37].

In the present study, an IRGP signature was established with the use of a LASSO penalized Cox regression model to predict OS of CM patients. The prognostic signature of 23 IRGPs consisted of 46 unique IRGs. Most of the genes involved in the immune signature code for molecules related to antimicrobials and cytokines, which play important roles in the response to stimuli and the immune microenvironment. Of these immunity-related genes, active expression of serine/threonine kinase 1
promotes melanoma metastasis[38]. Serum levels of C-C motif chemokine ligand 17 (CCL17) is an independent prognostic marker of distant metastasis of melanoma. Among patients with high CCL17 levels, 43% were reported to survive at 3 years[39]. Singh et al. found that activation of intratumoral cluster of differentiation 40 induced T cell-mediated eradication of melanoma in the brain[40]. Smith et al. discovered that endothelin 1 was enhanced in treated melanomas and confers drug resistance through endothelin receptor type A[41]. Ribonuclease L reportedly interacts with micro RNA-146a as a sex-specific factor in melanoma[42] and emaphoring 7A can reduce pulmonary metastasis of melanoma[43].

M2 macrophages have been shown to favor growth and are related to poorer OS of melanoma patients, while M1 macrophages support tumor destruction and antigen presentation[44]. Yamaguchi et al. found that anti-PD-1 antibody (nivolumab) therapy increased activated effector memory phenotypes of central memory T cells and subsets of CD4+ and CD8+ central memory T cells, as well as activated Th1 plus T-helper follicular 1 cells[45], which means these immune cells activated can prolong patients survival. The results of the present study revealed a significant increase in the abundance of infiltrating M0 and M2 macrophages in the high risk immune group, while the abundances of infiltrating M1 macrophages, activated memory CD4+ T cells, CD8+ T cells, follicular helper T cells, monocytes, plasma cells, and gamma delta T cells were greater in the low risk immune group which had better survival rate. In addition, both in the mRNA analysis and Immunohistochemistry results showed that high immune risk group had higher tumor purity and lower immune cells and stromal cells infiltration, and the immune checkpoint inhibitor such as PD1, PDL1 and CTLA-4 were highly expressed in low immune risk group. Meanwhile, the updated report showed PSMB8 and PSMB9 overexpressed could predictive for better survival and improved response to immune-checkpoint inhibitors, which were expressed highly in low immune risk group, this proved that the low immune risk group had better survival rate and Immunotherapy effect, and if these genes mutates, poorer OS occured. To our surprise, PRAME which acted independent biomarker in metastasis Uveal Melanoma were also significant expressed in high risk group, it means that PRAME may also be a biomarker in CM, but it needs intensive study to verify. Thus, research outcomes were closely consistent with previous public studies, demonstrating the precise predict power.

Moreover, as determined by the 23-IRGP signature, three pathways (i.e., oxidative phosphorylation, retinol metabolism, and ribosome) were highly related to the invasiveness of melanoma, suggesting that a high risk score was correlated to upregulation of melanoma metastasis and poor survive. These results suggested the capability of the IRGP signature to predict tumor invasion in CM patients.

Nevertheless, there were some limitations to this study that should be addressed. First, the 23-IRGP prognostic signature was based on a retrospective study using the TCGA CM dataset and validated with the use of three microarray datasets from the GEO dataset. Thus, these results should be further validated with more datasets with different sample attributes in a prospective cohort. Second, as 23 IRGPs were used for the prognostic signature model, different prognostic signature models are needed for comparison. Third, further validation of the 23 IRGPs by quantitative real-time polymerase chain reaction, western blotting, and immunohistochemical analyses are needed prior to clinical application.
Conclusions

In conclusion, our findings indicate a novel promising prognostic model in CM and built the bridge between immune micro-environment and CM patient results. This approach can be applied to discover the signatures in other diseases without technical bias from different platforms.

Abbreviations

CM: cutaneous melanoma; IRGPs: Immune related gene pairs; IRGs: Immune related genes;

Declarations

Acknowledgements

We would like to acknowledge the investigators and institutions involved in this study.

Author Contributions

XYN conceptualized the project, all data analysis and wrote the first draft of the manuscript. XYN, WZC, MYZ and WPY contributed to processing, analysis, and interpretation of the data. TWQ contributed to guide the data analysis, and manuscript writing. All authors read and approved the final manuscript.

Funding

No application.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1 Department of Plastic Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, 310016, China 2 Department of Chemistry and Life Sciences, Zhejiang Normal University,
References

1. Skin cancers [https://www.who.int/uv/faq/skincancer/en/index1.html]


immunotherapy: CII 2015, 64(9):1075-1082.


Tables
<table>
<thead>
<tr>
<th>Gene Pair1</th>
<th>Full Name</th>
<th>Immune Processes</th>
<th>Gene Pair2</th>
<th>Full Name</th>
<th>Immune Processes</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD8A</td>
<td>CD8a molecule</td>
<td>Antigen_Processing_and_Presentation, Antimicrobials, TCRsignalingPathway</td>
<td>LTBP2</td>
<td>latent transforming growth factor beta binding protein 2</td>
<td>Cytokines</td>
<td>-0.009639653</td>
</tr>
<tr>
<td>HSPA1A</td>
<td>heat shock 70kDa protein 1A</td>
<td>Antigen_Processing_and_Presentation</td>
<td>LYZ</td>
<td>lysozyme (renal amyloidosis)</td>
<td>Antimicrobials</td>
<td>0.098898359</td>
</tr>
<tr>
<td>HSPA1A</td>
<td>heat shock 70kDa protein 1A</td>
<td>Antigen_Processing_and_Presentation</td>
<td>GBP2</td>
<td>guanylate binding protein 2, interferon-inducible</td>
<td>Antimicrobials</td>
<td>0.06573857</td>
</tr>
<tr>
<td>HSPA1A</td>
<td>heat shock 70kDa protein 1A</td>
<td>Antigen_Processing_and_Presentation</td>
<td>LTBP3</td>
<td>latent transforming growth factor beta binding protein 3</td>
<td>Cytokines</td>
<td>0.00106882</td>
</tr>
<tr>
<td>ICAM1</td>
<td>intercellular adhesion molecule 1</td>
<td>Antigen_Processing_and_Presentation</td>
<td>PLXNB2</td>
<td>plexin B2</td>
<td>Chemokine_Receptors, Cytokine_Receptors</td>
<td>-0.039340917</td>
</tr>
<tr>
<td>PSME1</td>
<td>proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)</td>
<td>Antigen_Processing_and_Presentation</td>
<td>NENF</td>
<td>neuron derived neurotrophic factor</td>
<td>Cytokines</td>
<td>-0.008677831</td>
</tr>
<tr>
<td>ZC3HAV1L</td>
<td>zinc finger CCCH-type, antiviral 1-like</td>
<td>Antimicrobials</td>
<td>CMTM8</td>
<td>CKLF-like MARVEL transmembrane domain containing 8</td>
<td>Cytokines</td>
<td>0.037052249</td>
</tr>
<tr>
<td>APOBEC3G</td>
<td>apolipoprotein B</td>
<td>Antimicrobials</td>
<td>IKBKE</td>
<td>inhibitor of kappa light</td>
<td>Antimicrobials</td>
<td>-0.511310011</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>Antimicrobials</td>
<td>SIC</td>
<td>Cytokine Function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYBB</td>
<td>Cytochrome e b-245, beta polypeptide</td>
<td>Antimicrobials</td>
<td>SEMA7A</td>
<td>Semaphorin 7A, GPI membrane anchor (John Milton Hagen blood group)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRIM5</td>
<td>Tripartite motif-containing 5</td>
<td>Antimicrobials</td>
<td>SOS1</td>
<td>Son of sevenless homolog 1 (Drosophila)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFS10</td>
<td>Tumor necrosis factor (ligand) superfamily member 10</td>
<td>Antimicrobials, Cytokines, Natural Killer Cell_Cytotoxicity, TNF_Family_Members</td>
<td>STC1</td>
<td>Stanniocalcin 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNASEL</td>
<td>Ribonuclease L (2',5'-oligoisoacylate nylate synthetase-dependent)</td>
<td>Antimicrobials</td>
<td>TRIM22</td>
<td>Tripartite motif-containing 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RARRES3</td>
<td>Retinoic acid receptor responder (tazarotene induced) 3</td>
<td>Antimicrobials</td>
<td>DDX17</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD40</td>
<td>CD40 molecule, TNF receptor superfamily member 5</td>
<td>Antimicrobials, Cytokine_Receptors</td>
<td>BIRC5</td>
<td>Baculoviral IAP repeat-containing 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISG15</td>
<td>ISG15 ubiquitin-like modifier</td>
<td>Antimicrobials</td>
<td>NENF</td>
<td>Neuron derived neurotrophic factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNLY</td>
<td>Granulysin</td>
<td>Antimicrobials</td>
<td>EDNRA</td>
<td>Endothelin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Page 16/30
<table>
<thead>
<tr>
<th>Protein</th>
<th>Description</th>
<th>Type</th>
<th>Antimicrobials</th>
<th>Cytokine Receptors</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRF7</td>
<td>interferon regulatory factor 7</td>
<td>Antimicrobials</td>
<td>AKT1</td>
<td>v-akt murine thymoma viral oncogene homolog 1</td>
<td>-0.0351425</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BCR Signaling Pathway, TCR Signaling Pathway</td>
<td>72</td>
</tr>
<tr>
<td>TRIM22</td>
<td>tripartite motif-containing 22</td>
<td>Antimicrobials</td>
<td>CCL17</td>
<td>chemokine (C-C motif) ligand 17</td>
<td>-0.2700542</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Antimicrobials, Chemokines, Cytokines</td>
<td>36</td>
</tr>
<tr>
<td>BIRC5</td>
<td>baculoviral IAP repeat-containing 5</td>
<td>Antimicrobials</td>
<td>NGFR</td>
<td>nerve growth factor receptor (TNFR superfamily, member 16)</td>
<td>0.2116530</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cytokine Receptors</td>
<td>36</td>
</tr>
<tr>
<td>GBP2</td>
<td>guanylate binding protein 2, interferon-inducible</td>
<td>Antimicrobials</td>
<td>PLXNA1</td>
<td>plexin A1</td>
<td>-0.0312641</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chemokine Receptors, Cytokine Receptors</td>
<td>98</td>
</tr>
<tr>
<td>GBP2</td>
<td>guanylate binding protein 2, interferon-inducible</td>
<td>Antimicrobials</td>
<td>SHC4</td>
<td>SHC (Src homology 2 domain containing) family, member 4</td>
<td>-0.0901113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Natural Killer Cell Cytotoxicity</td>
<td>9</td>
</tr>
<tr>
<td>PTK2B</td>
<td>PTK2B protein tyrosine kinase 2 beta</td>
<td>Antimicrobials, Natural Killer Cell Cytotoxicity</td>
<td>LHB</td>
<td>luteinizing hormone beta polypeptide</td>
<td>-0.3972202</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cytokines</td>
<td>58</td>
</tr>
<tr>
<td>CD86</td>
<td>CD86 molecule</td>
<td>Antimicrobials</td>
<td>IL1RN</td>
<td>interleukin 1 receptor antagonist</td>
<td>-0.0183748</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cytokines, Interleukins</td>
<td>51</td>
</tr>
</tbody>
</table>
Table 2.
Univariate Cox and Multivariate Cox analysis of clinicopathological factors and risk signatures

<table>
<thead>
<tr>
<th>Variable</th>
<th>TCGA-train dataset</th>
<th>TCGA-test dataset</th>
<th>GSE65904 validation dataset</th>
<th>GSE59455 validation dataset</th>
<th>GSE22153 validation dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>HR</td>
<td>95%CI</td>
<td>pvalue</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>age</td>
<td>1.02</td>
<td>1.00-1.03</td>
<td>0.013</td>
<td>1.01</td>
<td>0.99-1.02</td>
</tr>
<tr>
<td>gender</td>
<td>1.11</td>
<td>0.73-1.71</td>
<td>0.621</td>
<td>1.25</td>
<td>0.79-1.95</td>
</tr>
<tr>
<td>stage</td>
<td>1.54</td>
<td>1.20-1.98</td>
<td>0.001</td>
<td>1.56</td>
<td>1.20-2.04</td>
</tr>
<tr>
<td>riskScore</td>
<td>10.41</td>
<td>6.43-16.84</td>
<td>1.49E-21</td>
<td>10.64</td>
<td>6.47-17.49</td>
</tr>
<tr>
<td>id</td>
<td>HR</td>
<td>95%CI</td>
<td>pvalue</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>age</td>
<td>1.02</td>
<td>1.01-1.04</td>
<td>0.002</td>
<td>1.02</td>
<td>1.01-1.04</td>
</tr>
<tr>
<td>gender</td>
<td>0.94</td>
<td>0.60-1.49</td>
<td>0.805</td>
<td>0.84</td>
<td>0.53-1.34</td>
</tr>
<tr>
<td>stage</td>
<td>1.27</td>
<td>1.02-1.58</td>
<td>0.036</td>
<td>1.3</td>
<td>1.03-1.64</td>
</tr>
<tr>
<td>riskScore</td>
<td>2.12</td>
<td>1.44-3.13</td>
<td>0</td>
<td>2.08</td>
<td>1.41-3.05</td>
</tr>
<tr>
<td>id</td>
<td>HR</td>
<td>95%CI</td>
<td>pvalue</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>age</td>
<td>0.99</td>
<td>0.98-1.01</td>
<td>0.567</td>
<td>1</td>
<td>0.99-1.02</td>
</tr>
<tr>
<td>gender</td>
<td>1.37</td>
<td>0.90-2.08</td>
<td>0.143</td>
<td>1.31</td>
<td>0.86-1.99</td>
</tr>
<tr>
<td>stage</td>
<td>2.35</td>
<td>1.53-3.60</td>
<td>8.55E-05</td>
<td>2.77</td>
<td>1.78-4.29</td>
</tr>
<tr>
<td>riskScore</td>
<td>1.76</td>
<td>1.34-2.31</td>
<td>4.48E-05</td>
<td>1.98</td>
<td>1.49-2.62</td>
</tr>
<tr>
<td>id</td>
<td>HR</td>
<td>95%CI</td>
<td>pvalue</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>age</td>
<td>0.99</td>
<td>0.98-1.00</td>
<td>0.177</td>
<td>0.99</td>
<td>0.98-1.00</td>
</tr>
<tr>
<td>gender</td>
<td>1.24</td>
<td>0.84-1.82</td>
<td>0.283</td>
<td>1.17</td>
<td>0.80-1.73</td>
</tr>
<tr>
<td>stage</td>
<td>0.77</td>
<td>0.67-0.88</td>
<td>0</td>
<td>0.76</td>
<td>0.66-0.87</td>
</tr>
<tr>
<td>riskScore</td>
<td>1.35</td>
<td>0.78-2.33</td>
<td>0.141</td>
<td>1.64</td>
<td>0.94-2.82</td>
</tr>
<tr>
<td>id</td>
<td>HR</td>
<td>95%CI</td>
<td>pvalue</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>age</td>
<td>0.99</td>
<td>0.98-1.00</td>
<td>0.177</td>
<td>0.99</td>
<td>0.98-1.00</td>
</tr>
<tr>
<td>gender</td>
<td>1.24</td>
<td>0.84-1.82</td>
<td>0.283</td>
<td>1.17</td>
<td>0.80-1.73</td>
</tr>
<tr>
<td>stage</td>
<td>0.77</td>
<td>0.67-0.88</td>
<td>0</td>
<td>0.76</td>
<td>0.66-0.87</td>
</tr>
<tr>
<td>riskScore</td>
<td>1.35</td>
<td>0.78-2.33</td>
<td>0.141</td>
<td>1.64</td>
<td>0.94-2.82</td>
</tr>
<tr>
<td>id</td>
<td>HR</td>
<td>95%CI</td>
<td>pvalue</td>
<td>HR</td>
<td>95%CI</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>--------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>age</td>
<td>1.01</td>
<td>0.99-1.03</td>
<td>0.372</td>
<td>1.01</td>
<td>0.99-1.03</td>
</tr>
<tr>
<td>gender</td>
<td>0.99</td>
<td>0.55-1.76</td>
<td>0.967</td>
<td>1.32</td>
<td>0.71-2.43</td>
</tr>
<tr>
<td>stage</td>
<td>1.69</td>
<td>0.66-4.30</td>
<td>0.27</td>
<td>1.05</td>
<td>0.33-3.31</td>
</tr>
<tr>
<td>riskScore</td>
<td>1.95</td>
<td>1.15-3.30</td>
<td>0.013</td>
<td>2.07</td>
<td>1.16-3.69</td>
</tr>
</tbody>
</table>

**Figures**
Figure 1

Establishment and assessment of a 23-IRGP signature. (a) A heat map of the risk scores of the 23 IRGPs. (b) According to the 23 IRGPs, the training cohort was divided into high and low immune risk groups. The red and green points represent the risk scores of the high and low risk groups, respectively. (c) A plot of OS based on the 23 IRGPs. The red points represent deaths, while the blue points represent survivors. (d) The AUCs for 1-, 3-, and 5-year OS in the training cohort were 0.909, 0.901, and 0.912, respectively. (e) According to the OS curve, OS was poorer for the high risk group as compared to the low risk group in the training cohort (p < 0.001).
Figure 2

Construction of a Robust nomogram in TCGA training dataset. (a) A time-dependent ROC curve for IRGPs in the training dataset. An IRGP score of -0.674 was used as a cut-off to assign patients to the high- or low-risk group. (b) The 1-, 3-, and 5-year calibration curves of the nomogram. (c) A nomogram of OS was established by 23-IRGP risk score and other clinicopathological parameters.
Cox proportional hazards model and stratified analysis of the training cohort revealed 23-IRGP was an independent prognostic factor. (a) Age, stage, and risk score were the independent prognostic factors in Univariate Cox analysis. (b) Stage and risk score were the independent prognostic factors in Multivariate Cox analysis. Stratified analyses applied by age (p = 0.029) (c), tumor stage (p = 0.002) (d), and sex (p = 0.543) (e) demonstrated the predictive of 23-IRGP in OS of CM patients.
Figure 4

Validation of the 23-IRGP signature. As indicated, OS was poorer for the high-risk group than the low-risk group of the testing cohort (a). Datasets GSE65904 (b), GSE59455 (c), and GSE54467 (d). These results showed that the 23-IRGP signature had good predictive ability ($p < 0.05$) (E). The C-index values for the
TCGA training and testing cohorts, as well as the datasets GES65904, GSE59455, and GSE54467 were 0.775 (95% CI = 0.748–0.802), 0.636 (95% CI = 0.585–0.687), 0.650 (95% CI = 0.609–0.691), 0.557 (95% CI = 0.508–0.606), and 0.610 (95% CI = 0.537–0.683), respectively.

Figure 5

Identification of IRGs and comparison with IRGPs. (a) A heat map of the patient risk scores of the 24 IRGs. (b) Based on the 24 IRGs, patients in the TCGA dataset were assigned to the high- or low-immune risk groups. The red and green points represent the risk scores of the high and low risk groups, respectively. (c) A plot of OS based on the 24 IRGs. The red points represent deaths and blue points represent survivors. (d) The AUCs for 1-, 3-, and 5-year OS in the training cohort were 0.731, 0.760, and 0.749, respectively. (e) According to the OS curve, OS was poorer for the high risk group as compared to the low risk group in the training cohort (p < 0.001). These results showed that the prognostic signature of the IRGs had good predictive power, but with a smaller AUC and lower C-index (0.610 in Figure 5e) than the 23 IRGPs signature. Therefore, the 23-IRGP were more precisely predictive for CM.
Figure 6

Immune infiltration status of the 23 IRGPs. (a) Summary of the abundances of 22 types of immune cells, as estimated with the use of the CIBERSORT analytical tool for different risk groups. (b-j) The abundance distribution of specific immune cells within different risk groups. The abundances of M0 macrophages ($p = 0.004$) and M2 macrophages ($p = 0.003$) were significantly greater in the high risk group, while the abundances of M1 macrophages ($p = 0.001$), activated CD4+ memory T cells ($p = 0.005$), monocytes ($p = 0.047$), plasma cells ($p = 0.011$), CD8+ T cells ($p = 0.028$), follicular helper T cells ($p = 0.017$), and gamma delta T cells ($p = 0.014$) were significantly enriched in the low risk group. *$p < 0.05$, **$p < 0.01$ (t-test).
Figure 7

Tumor micro-environment (TME) and key genes expression in two different groups. In TME, “TumorPurity” is the percentage of tumor cells, “ImmuneScore” is the percentage of Immune cells, “StromalScore” is the percentage of stromal cells, “EstimateScore” is the percentage that merge the “ImmuneScore” and “StromalScore”. as we can see, high risk group had higher tumor purity with lower immune cells and stromal cells infiltration (a). in low immune risk group, PD1, PDL-1, CTLA-4, PSMB8 and PSMB9 were highly expressed (b-f). PRAME were significant expressed in high risk group (g). In GEPIA,
the patients with high PD1, PDL-1, CTLA4, PSMB8 and PSMB9 expression had better OS (h-l), but PRAME had no significant effect in OS (m). (n) IHC result of PD1 protein in high risk group. Staining. not detected; intensity, negative; quantity, none; location, none. (o) IHC result of PD1 protein in low risk group. Staining, low; intensity, weak; quantity, 75-25%; location, cytoplasmic/membranous. (p) IHC result of PDL1 protein in high risk group. Staining. not detected; intensity, negative; quantity, none; location, none. (q) IHC result of PDL1 protein in low risk group. Staining, high; intensity, Strong; quantity, >75%; location, cytoplasmic/membranous. (r) IHC result of PSMB8 protein in high risk group. Staining. low; intensity, moderate; quantity, <25%; location, cytoplasmic/membranous nuclear. (s) IHC result of PSMB8 protein in low risk group. Staining, high; intensity, Strong; quantity, >75%; location, cytoplasmic/membranous nuclear. (t) IHC result of PSMB9 protein in high risk group. Staining. not detected; intensity, negative; quantity, none; location, none. (u) IHC result of PSMB9 protein in low risk group. Staining. medium; intensity, moderate; quantity, >75%; location, cytoplasmic/membranous nuclear. (v) IHC result of PRAME protein in high risk group. Staining. medium; intensity, moderate; quantity, 75-25%; location, cytoplasmic/membranous. (w) IHC result of PRAME protein in low risk group. Staining. not detected; intensity, weak; quantity, <25%; location, cytoplasmic/membranous. ***p < 0.01 (t-test).
Figure 8

GO analysis of the high-immune risk group based on the 23 IRGPs (padj < 0.05). "Normalized enrichment score" is the percentage of differentiated genes for the given GO term.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• TableS1.docx
• FigureS2.tif
• FigureS4.tif
• FigureS1.tif
• FigureS3.tif