Prognostic Modeling of Patients with Metastatic Melanoma Based on Tumor Immune Microenvironment Characteristics

Jing Liu  
Guangdong Pharmaceutical University
Ting Ye  
Guangdong Pharmaceutical University
Xue fang Zhang  
Dongguan People's Hospital
Yong jian Dong  
Guangdong Pharmaceutical University
Wen feng Zhang  
Guangdong Pharmaceutical University
Feng lin Wu  
Guangdong Pharmaceutical University
Hua ben Bo  
Guangdong Pharmaceutical University
Hong wei Shao  
Guangdong Pharmaceutical University
Rong xin Zhang  
Guangdong Pharmaceutical University
Han Shen (shenhanbc@163.com)  
Guangdong Pharmaceutical University

Research Article

Keywords: Metastatic melanoma, Tumor microenvironment, ESTIMATE algorithm, Prognostic model, Bioinformatics

Posted Date: July 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-730205/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Version of Record: A version of this preprint was published at Mathematical Biosciences and Engineering on January 1st, 2021. See the published version at https://doi.org/10.3934/mbe.2022067.
Abstract

Most of the malignant melanomas are already in the middle and advanced stages when they are diagnosed, which is often accompanied by the metastasis and spread of other organs. Besides, the prognosis of patients is bleak. The characteristics of the local immune microenvironment in metastatic melanoma have important implications for both tumor progression and tumor treatment. In this study, data on patients with metastatic melanoma from the TCGA and GEO datasets were selected for immune, stromal, and estimate scores, and overlapping differentially expressed genes (DEGs) were screened. A nine-IRGs prognostic model (ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22) was established by univariate COX regression, LASSO and multivariate COX regression. Receiver operating characteristic (ROC) curves were used to test the predictive accuracy of the model. Immune infiltration was analyzed by using CIBERSORT, Xcell and ssGSEA in high-risk and low-risk groups. The immune infiltration of the high-risk group was significantly lower than that of the low-risk group. Immune checkpoint analysis revealed that the expression of PDCD1, CTLA4, TIGIT, CD274, HAVR2 and LAG3 were significantly different in groups with different levels of risk scores. WGCNA analysis found that the yellow-green module contained seven genes (ALOX5AP, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22) from the nine-IRG prognostic model, of which the yellow-green module had the highest correlation with risk scores. The results of GO and KEGG suggested that the genes in the yellow-green module were mainly enriched in immune-related biological processes. Finally, we analyzed the prognostic ability and expression characteristics of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 in metastatic melanoma. Overall, a prognostic model for metastatic melanoma based on the characteristics of the tumor immune microenvironment was established, which was helpful for further studies. It could function well in helping people to understand the characteristics of the immune microenvironment in metastatic melanoma and to find possible therapeutic targets.

1. Introduction

Melanoma is a highly lethal skin malignancy with low morbidity, easily early metastasis [1]. There are approximately 200,000 new cases diagnosed each year worldwide, and the mortality rate is showing a rapid increase each year[2]. For melanoma in the early stage, complete resection of the primary tumor can achieve a good prognosis and be cured [3]. However, for melanoma in the late stage, especially stage IV melanoma, the risk of recurrence after surgery is high and the prognosis is poor [4]. Stage I and II involve localized disease, stage III is characterized by metastasis to the local lymphnodes, and stage IV represents distant metastasis [5]. In recent years, the management of patients with metastatic melanoma has evolved considerably through targeted therapies and immunotherapy [6]. These findings underscore the importance of discovering new prognostic biomarkers and immunotherapy for patients with stage III or IV melanoma who have developed metastases.

Immunotherapy has become a new therapeutic approach to fight cancer by activating the immune system, which is widely used in various tumors, especially melanoma, lung cancer and renal cell carcinoma[7]. Significant progress has been made in prolonging patient survival and improving the
patients’ living quality. The research of immunotherapy focuses on the recruitment of tumor-infiltrating T cells which are important lymphocytes in tumor immunity and play a leading role in tumor immunity by not only playing an immunosurveillance role in the body but also correcting the body's tumor immune deficiency and directly killing tumor cells [8, 9]. As a malignant tumor with a high mutation rate, melanoma can produce a large number of tumor antigens to be recognized by the immune system and has high immunogenicity [10]. Immune checkpoint inhibitors are an immunomodulator, which has completely changed the treatment strategy of tumors and has become the standard therapy for a variety of tumors, including SKCM [11]. For example, CTLA-4 monoclonal antibody is a protein expressed on the surface of T cells. After binding to CD80/CD86 of antigen-presenting cells, it can negatively regulate the immune response of T cells and block CTLA-4 to enhance the anti-tumor effect [12]. Effective immunotherapy depends largely on the immune status of the tumor microenvironment [13]. What's more, immunotherapy would be more effective if the patients contain a high density of tumor-infiltrating lymphocytes. The tumor microenvironment is the cellular environment where the tumor is located, which composes of immune cells, mesenchymal cells, endothelial cells, inflammatory mediators and extracellular matrix molecules [14]. The cells and molecules of the tumor microenvironment are in a dynamic process that reflects the evolutionary nature of cancer and jointly promotes immune escape, tumor growth, and metastasis [15]. Immune and stromal cells are two major non-tumor components that are considered to be of great value in the diagnosis and prognosis of tumors [16, 17]. There is growing evidence that the tumor microenvironment is strongly correlated with the survival of patients with metastatic melanoma treated with immunotherapy [18]. Therefore, understanding the composition and function of the tumor microenvironment and screening for immune-related biological targets are essential to effectively control the progression of metastatic melanoma and improve the overall survival of patients with metastatic melanoma.

In this study, samples with metastatic melanoma were obtained from the TCGA and GSE65904 databases, and all samples were analyzed for immune, stromal and estimate scores by using the ESTIMATE algorithm. Taking differentially expressed genes as the object, the nine-IRGs prognostic model in metastatic melanoma was determined by univariate Cox regression, LASSO and multivariate Cox regression. This model could effectively distinguish the high and low risks of patients with metastatic melanoma, with patients in the low-risk group showing longer survival and more immune cell infiltration. Further analysis of the expression characteristics of the nine genes in the prognostic model revealed that our study provided a potential model and biomarker for further immune-related work and personalized drugs for the treatment of patients with metastatic melanoma.

2. Materials And Methods

2.1 Data collection and processing

RNA-seq data of skin melanoma patients, containing corresponding clinical data such as age, TNM, staging, gender, survival-time and status were downloaded from the TCGA database (https://tcga-data.nci.nih.gov/tcga/). 189 patients with metastatic melanoma in Stage III and Stage IV were screened
for subsequent analysis. The gene expression data of GSE65904 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

2.2 Survival analysis of stromal, immune and estimate scores

Stromal, immune and estimate scores for patients with metastatic melanoma were calculated by the ESTIMATE package of R software in the GSE65904 and TCGA-SKCM cohorts. Survival analysis of stromal, immune and estimate scores was performed by Kaplan-Meyer analysis. Statistical significance was tested by a log-rank test with the significance threshold for p-value set at 0.05.

2.3 Differential expression analysis

Patients with metastatic melanoma from the TCGA-SKCM and GSE65904 cohorts were divided into high and lowscores groups based on median immune score. Differentially expressed genes were analyzed by the R software limma package in the GSE65904 cohort. Differential expression analysis of the count matrix of the samples was performed in the TCGA-SKCM cohort using the DESeq2 package. Genes with |log fold change| > 1 and adjusted p-values < 0.05 were considered to be significantly differentially expressed genes (DEGs). Venn diagram (https://bioinfogp.cnb.csic.es/tools/venny/index.html) was used to identify overlapping DEGs. The ClusterProfiler package was used for gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for the overlapping DEGs. GO enrichment analysis focused on describing biological processes (BP), cellular components (CC) and molecular functions (MF) associated with DEGs. KEGG pathway analysis revealed biological pathways associated with DEGs.

2.4 Prognostic modeling

In the TCGA database, Clinical data and clinical related information were downloaded, and samples with Over Survival (OS) data were used for further survival analysis. To get survival and immune-related genes (IRGs), univariate Cox regression analysis of DEGs was performed by the "survivff " function in the survival package of R software (P < 0.05). The least absolute shrinkage and selection operator (LASSO) regression was used for dimensionality reduction analysis of survival-associated genes, and "survival" and "glmnet" packages in LASSO were used to select prognostic genes by imposing a penalty proportional to the contraction of the regression coefficients. Then, multivariate Cox regression analysis was performed to screen candidate prognostic IRGs. Finally, the regression coefficients obtained from COX regression multivariate analysis were multiplied by the expression level of each marker gene to construct the optimal prognostic model.

2.5 Validation of the prognostic model

Patients with metastatic melanoma in the training cohort and the GSE65904 test cohort were divided into high-risk and low-risk groups based on the median value of the TCGA-SKCM training cohort risk score. Kaplan-Meier curves and log-rank tests were applied to compare the survival differences between the different risk groups. The receiver operating characteristic (ROC) curves were used to assess the
predictive power of the risk score model. Forest plots were also used to show hazard ratios for selected prognostic genes.

2.6 Construction and evaluation of nomogram

To verify whether the prediction of the risk score model was independent of the traditional clinical characteristics of patients with metastatic melanoma, we performed univariate and multivariate Cox regression analyses on risk score, TNM, age, sex and stage. A nomogram integrating the prognostic signatures was constructed by the R software rms package for predicting OS at 1, 2, and 3 years in patients with metastatic melanoma, and calibration curves were plotted to validate the accuracy of the prediction model.

2.7 Gene set enrichment analysis (GSEA)

To explore the biological pathways involved between the high-risk and low-risk groups of patients, Gene Set Enrichment Analysis (GSEA) was conducted by using the "KEGG" and "GO" gene sets from the Molecular Signatures Database (http://software.broadinstitute.org/gsea/index.jsp). Gene expression profile files were used as input phenotypic data. Genomic alignments were analyzed 1000 times [19]. P-values < 0.05 were set as statistical significance.

2.8 Exploration of the immune microenvironment based on the IRGs signature

We used the CIBERSORT method, a generalized deconvolution algorithm based on a gene expression matrix to quantify large tissue cell fractions, to estimate the proportions of 22 immune cell phenotypes in high- and low-risk cohorts [20]. Xcell is a method for inferring the proportion of 64 immune and stromal cell types in samples using gene expression signatures and determines the enrichment score of each cell type by ssGSEA [21]. In addition, ssGSEA is a method that performs ordinal normalization based on the gene expression values of samples and then calculates an enrichment score using an empirical cumulative distribution function to assess the level of immune infiltration of the sample [22]. Wilcoxon rank-sum test was used to compare the differential abundance of immune cells in high- and low-risk cohorts. To ensure the accuracy of the algorithm, the threshold was set as P value < 0.05.

2.9 Immune Checkpoint Analysis based on the IRGs signature

To investigate the expression of immune checkpoint inhibitors between high-risk and low-risk score groups, we analyzed the differences between high-risk and low-risk groups based on gene expression data of PDCD1, CTLA4, TIGIT, CD274, HAVR2 and LAG3 in the TCGA-SKCM cohort. Wilcoxon rank-sum test was used to determine the significance of the difference of results and P < 0.05 was considered to be significant. Box plot was plotted by the boxplot package of R software to indicate the differential expression of immune checkpoint inhibitors between high-risk and low-risk groups.
2.10 Construction of weighted gene co-expression network analysis (WGCNA)

The TCGA-SKCM cohort included 189 metastatic melanoma samples with detailed overall survival information, which was suitable for the construction of WGCNA. We could link the module signature genes to the survival data of the samples. WGCNA analysis was performed on the top 25% of genes with the highest variance in the tumor samples by using the "WGCNA" package in R software. The sample hierarchical clustering method was used to detect and exclude anomalous samples before selecting the appropriate soft threshold power to implement a standard scale-free network. In the next stage, the gene dendrogram and module identification were completed with dynamic tree cutting by the construction of adjacency and topological overlap matrices (TOM) and the calculation of the corresponding phase dissimilarity, and the minimum module size was 40. Then the clustering of the module characteristic genes realized the merging of highly similar modules and calculated the correlation between the module characteristic genes and the clinical phenotype of the sample. The module with the highest correlation with risk scores was selected for further analysis.

2.11 Immune cell infiltration of the nine IRGs

TIMER (https://cistrome.shinyapps.io/timer/) is a comprehensive platform for the systematic analysis of immune cell infiltration in various malignancies, which re-analyzes gene expression data to assess the level of infiltration of six immune cell subtypes, including CD4 + T cells, B cells, CD8 + T cells, neutrophils, macrophages, and dendritic cells. Thus, we analyzed the association of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2, and TRIM22 with immune cell infiltration by using the TIMER database.

2.12 Statistical analysis

The FPKM data of the TCGA-SKCM cohort was transformed into transcripts per kilobase million (TPM) values. The batch effect between different data sets was eliminated by the "sva" R package. The Log-Rank test and Kaplan-Meier method were used for survival analysis. The distributions in two sets of any continuous variable were compared using the Wilcoxon rank-sum test. Univariate and multivariate Cox regression analyses were performed using the "survival" R package. The ROC curves were generated using the "survivalroc" R package and the corresponding area under the curve was obtained. Differences were considered statistically significant at P < 0.05. All statistical analyses were performed by using R software (version 3.6.3).

3 Results
3.1 Survival analysis of stromal, immune and estimate scores

According to the ESTIMATE algorithm, for the TCGA-SKCM cohort, the stromal score ranged between −1836.9 and 1854.648, the immune score between −1002.3 and 3833.031 and the estimate score between −2696.16 and 5144.515, and for the GSE65904 cohort, the stromal score ranged between −1111.36 and 2282.686, the immune score between −570.15 and 3381.81 and estimate score between −1343.02 and 4989.343. To identify potential correlations between overall survival and immune, stromal and estimate scores, patients with metastatic melanoma based on the median score were divided into high and low score groups in the TCGA-SKCM and GSE65904 cohorts. Kaplan-Meier survival curve showed that the overall survival of the high immune, stromal and estimate score group was longer than that in the low score group (immune score, p = 0.001; stromal score, p = 0.014; estimate score, p = 0.003) in the TCGA-SKCM cohort, which was statistically significant (Fig. 1A-C). Similarly, in the GSE65904 cohort, the overall survival of the high immune score group was also longer than that of the low score group (p = 0.012), while there was no significant difference in OS between patients with high and low stromal and estimate scores (stromal score, p = 0.973; estimate score, p = 0.202) (Fig. 1D-F). The above results indicated that the good prognosis of patients with metastatic melanoma may be related to the remodeling of immune components.

3.2 Identification and analysis of DEGs

To reveal the relationship between patients with metastatic melanoma and the tumor microenvironment, we performed the differential analysis of gene expression profiles of samples from the TCGA-SKCM and GSE65904 cohorts based on the immune scores. In the TCGA-SKCM cohort, 3716 differentially expressed genes were identified using the DESeq2 package in R software, of which 744 genes were down-regulated and 2942 genes were up-regulated (Fig. 2A). In the GSE65904 cohort, 347 differentially expressed genes were identified by the limma package, of which 9 genes were down-regulated and 338 genes were up-regulated (Fig. 2B). Venn analysis of the differential genes identified in the TCGA-SKCM and GSE65904 cohorts revealed 272 overlapping genes (Fig. 2C), of which one gene was down-regulated and 271 genes were up-regulated. We performed functional enrichment analysis, including GO: BP, CC, MF, and KEGG pathway analysis on the obtained 272 overlapping genes. GO function was mainly enriched in response to T cell activation, leukocyte cell-cell adhesion, leukocyte proliferation, and regulation of lymphocyte activation (Fig. 2D), while the KEGG pathway was mainly enriched in several signaling pathways in immune diseases, such as Hematopoietic cell lineage, Allograft rejection, Graft-versus-host disease, and Type I diabetes mellitus (Fig. 2E).

3.3 Prognostic modeling

A univariate Cox proportional hazard regression model was used to analyze 272 overlapping differential genes and 177 genes related to survival were identified. In addition, to prevent overfitting of the model, we obtained a gene signature for 17 prognostic genes by employing the least absolute shrinkage selection operator (LASSO) regression method(Fig. 3A-B). We further used multivariate COX regression analysis to
establish a prognostic risk regression model containing nine genes ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 (Fig. 3C). Risk scores for each sample were calculated based on the regression coefficients and expression levels of the nine genes. Riskscore=-(0.00069*expression level of ALOX5AP)-(0.00056*expression level of ARHGAP15) - (0.00096*expression level of CCL8) - (0.00039*expression level of FCER1G) - (0.00077* expression level of GBP4)+( 0.007688*expression level of HCK)-(0.00033*expression level of MMP9) - (0.00057*expression level of RARRES2) - (0.00178 * expression level of TRIM22). Patients were divided into high-risk and low-risk groups according to the median risk score. We then generated risk curves and scatter plots of risk scores and survival status for each sample in the TCGA-SKCM cohort and found that as risk scores increased, the number of deaths also increased (Fig. 3D-E). Kaplan-Meier analysis showed that patients in the high-risk group had lower OS than those in the low-risk group (P < 0.05) (Fig. 3F). Additionally, the area under the curve (AUC) at 1, 2, and 3 years OS was 0.712, 0.838, and 0.828, respectively (Fig. 3G). Our results suggested that the nine-IRGs prognostic model for survival prediction had acceptable sensitivity and specificity.

### 3.4 Validation of the prognostic model in the GEO melanoma cohort

To determine the stability of the prognostic model, the performance of the nine-IRGs prognostic model was evaluated in the GSE65904 cohort. Based on the nine-IRGs prognostic model, the risk score and survival status of each sample in the GSE65904 validation cohort were shown in the figure (Fig. 4A-B). The receiver operating characteristic (ROC) curves were used to predict survival rates at 1, 2 and 3 years and the area under the curve was 0.586, 0.639, and 0.640, respectively (Fig. 4C). Kaplan-Meier survival curves showed that the survival rate of high-risk patients is significantly shorter than that of low-risk patients (Fig. 4D). Consistent with the results of the training cohort, in the GSE65904 validation cohort, patients with metastatic melanoma who were assigned to the high-risk group according to the prognostic model had significantly worse overall survival (OS) than those assigned to the low-risk group.

### 3.5 Independent predictive capability of prognostic models

To further explore the independent predictive ability of the prognostic model, traditional clinical variables such as age, gender, tumor stage, TNM stage, and risk score were again analyzed by univariate and multivariate Cox regression analysis to evaluate the independent predictive ability of the model. In the TCGA-SKCM cohort, age, TN stage, and risk score models were statistically significant in the univariate analysis (Fig. 5A). Multivariate analysis further showed that T N stage and risk score were independent parameters associated with survival (Fig. 5B). In addition, to extend the availability of the nine-IRGs prognostic model and clinical applications, a nomogram was made for clinical variables and risk scores that could be used to visually predict the 1-year, 2-year and 3-year survival times of patients (Fig. 5C), and calibration curves were plotted to validate the accuracy of the prediction model (Fig. 5D-F). The predicted values matched well with the actual values, indicating that our model could be used to predict the prognosis of patients with metastatic melanoma. Notably, the findings suggested that the risk score
remained independent in predictive ability and could be used as independent factors in the prognosis of patients with metastatic melanoma.

### 3.6 Gene set enrichment analysis

To identify potential signaling pathways and biological processes associated with the nine-IRGs prognostic model, gene set enrichment analysis (GSEA) was used between high-low and low-risk groups in the TCGA-SKCM cohort. The GO results showed that the low-risk group was mainly enriched in immune receptor activity (NES = 2.229, FDR < 0.002), cytokine receptor activity (NES = 2.202, FDR < 0.002), regulation of lymphocyte-mediated immunity (NES = 2.196, FDR < 0.002), and regulation of leukocyte-mediated immunity (NES = 2.186, FDR < 0.002) (Fig. 6A). KEGG results showed that the low-risk group was mainly enriched in cytokine-cytokine receptor interaction (NES = 2.184, FDR < 0.001), natural killer cell-mediated cytotoxicity (NES = 2.034, FDR < 0.002), antigen processing and presentation (NES = 2.01, FDR < 0.002), and chemokine signaling pathway (NES = 1.960, FDR < 0.004) (Fig. 6B). These results indicated that the nine-IRGs prognostic model could be potentially valuable for inferring tumor microenvironment in patients with metastatic melanoma.

### 3.7 Immune infiltration analysis of high-risk and low-risk groups

To further elucidate the relationship between the prognostic model and tumor-infiltrating immune cells, we performed a series of analyses of cell types on the tumor microenvironment in the TCGA-SKCM cohort. First, 189 patients were divided into high-risk and low-risk groups based on the median risk score. The CIBERSORT algorithm was used to quantify the proportion of immune cells, and the difference of 22 immune cell types in the high-risk and low-risk groups was studied. The violin plot showed that plasma cells, CD8+ T cells, memory-activated CD4 T cells and M1 macrophages had higher infiltration levels in the low-risk group, while M2 macrophages were highly expressed in the high-risk group (Fig. 7A). In addition, to verify the accuracy of the results, we quantified the level of tumor immune infiltration from patients with metastatic melanoma by xCell and ssGSEA. Visual analysis revealed that the infiltration level of most immune cells in the low-risk group was significantly higher than that in the high-risk group (Fig. 7B-C). Therefore, we concluded that metastatic melanoma patients with different risk score subtypes significantly affected the tumor microenvironment.

### 3.8 Analysis of immune checkpoints in high-risk and low-risk groups

Since immune checkpoint inhibitors (ICIs) were used to treat malignant melanoma in clinical practice, we studied whether the risk model was related to ICIs-related biomarkers and found that the high-risk score group was negatively correlated with the high expression of CTLA4 (p < 0.001), HAVR2 (p < 0.001), LAG3 (p < 0.001), TIGIT (p < 0.001), CD274 (p < 0.001) and PDCD1 (p < 0.001) (Fig. 8). Therefore, patients in the low-risk group could have a better response to treatment with ICIs.
3.9 Construction of weighted gene co-expression network and identification of key modules, especially the yellow-green module that is highly correlated with risk scores

WGCNA is a systems biology method for analyzing the expression patterns of multiple genes in different samples, which can form clusters or modules containing genes with the same expression pattern[23]. If certain genes are located in one module, they may have the same biological function and the association between the module and the sample characteristics (such as clinical characteristics) can be studied. Three outlier samples were excluded from the TCGA-SKCM cohort so that a total of 186 samples with survival data were included in the WGCNA (Fig. 9A). In this study, we chose the power of $\beta = 5$ (scale-free $R^2 = 0.9$) as a soft threshold to implement the scale-free network. As a result, six gene co-expression modules were identified after excluding gray modules using merged dynamic tree cut (Fig. 9B). The heat map plotted the TOM of the 4709 selected genes in the analysis, which indicated that each module was validated independently of the other. Subsequently, we found that the yellow-green module, containing ALOX5AP, FCER1G, GBP4, HCK, MMP9, RARRES2, and TRIM22 in the nine-IRGs prognostic model, had the highest correlation with risk scores. ($R^2 = -0.3, P = 3e-05$) (Fig. 9C). We plotted the scatter plot of GS vs MM and risk scores in the yellow-green module (Cor = 0.44, P = 1.8e-37), elucidating that ALOX5AP (GS = 0.33, MM = 0.72), FCER1G (GS = 0.23, MM = 0.24), GBP4 (GS = 0.30, MM = 0.64), HCK (GS = 0.07, MM = 0.89), MMP9 (GS = 0.17, MM = 0.22), RARRES2 (GS = 0.28, MM = 0.39) and TRIM22 (GS = 0.34, MM = 0.80) seven genes were the central genes in the yellow-green module (Fig. 9D).

3.10 Analysis of biological functions and immune status of yellow-green module genes

In gene networks that conform to a scale-free distribution, genes with similar expression patterns can be co-regulated, functionally related, or pathway sharing. To further explore the potential biological functions of ALOX5AP, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 in metastatic melanoma, we performed GO and KEGG pathway analysis of genes associated with the yellow-green module using the "GO" and "KEGG" packages of R software. GO analysis showed that genes were mainly enriched in response to interferon-gamma, antigen processing and presentation, cellular response to interferon-gamma and adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains (Fig. 10A). KEGG pathway analysis was mainly associated with immunological diseases such as allograft rejection, autoimmune thyroid disease and viral myocarditis (Fig. 10B).

3.11 High expression of nine-IRGs is related to the progression of metastatic melanoma and tumor purity

To explore the prognostic ability of the nine-IRGs in different subgroups of metastatic melanoma, the Kaplan-Meier survival curves of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and
TRIM22 were plotted by the Survival package of R software. The results showed that the high expression (median) of ALOX5AP (p < 0.001), ARHGAP15 (p < 0.001), CCL8 (p < 0.001), FCER1G (p = 0.003), GBP4 (p < 0.001), HCK (p = 0.015), RARRES2 (p = 0.01) and TRIM22 (p < 0.001) has a good prognosis in metastatic melanoma, which was statistically significant (Fig. 11A). We next used gene expression data to assess the role of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 in the progression of metastatic melanoma. Compared with the high-risk group, the expression level of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 in the low-risk group was higher (Fig. 11B).

### 3.12 TIMER analysis

To understand whether ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2, and TRIM22 accurately reflected the status of the immune microenvironment in metastatic melanoma, the TIMER database was used to analyze the relationship between ALOX5AP, ARHGAP15, CCL8, FCER1G GBP4, HCK, MMP9, RARRES2 and TRIM22 with immune cell infiltration. The results showed that the expression of ALOX5AP, ARHGAP15, CCL8, FCER1G GBP4, HCK, MMP9, RARRES2 and TRIM22 showed a positive correlation with B cells, CD8 + T cells, CD4 + T cells, macrophages, neutrophils and dendritic cells and a negative correlation with tumor purity (Fig. 12).

### 4. Discussion

Melanoma is a highly malignant tumor that is prone to invasive metastasis and has a poor prognosis, which is also a highly immunogenic tumor and immunotherapy has outstanding advantages in improving the efficacy of melanoma[24]. In recent years, more and more evidence has shown that biomarkers based on tumor immunity can be used for the diagnosis, prognosis and treatment of tumor patients[25]. Therefore, it is meaningful and necessary to identify accurate biomarkers to construct prognostic characteristics to better predict patients with refractory diseases and poor survival.

To establish a suitable and simple protocol to observe the immune status of patients with metastatic melanoma and to suggest clinical outcomes, we used the ESTIMATE algorithm to calculate the stromal, immune and estimate scores of samples in the TCGA-SKCM and GSE65904 cohorts. Survival analysis showed that the survival time of patients in the low immune score group in the TCGA-SKCM cohort and GSE65904 cohort was significantly shorter than that in the high immune score group. Samples in the TCGA-SKCM and GSE65904 cohorts were divided into high and low scoring groups based on median immune score values and screened for 272 overlapping differential genes (DEGs) that were significantly associated with immune responses (such as T cell activation, Leukocyte adhesion and leukocyte proliferation). We constructed a nine-IRGs prognostic model by univariate COX regression, LASSO and multivariate COX regression and demonstrated its validity in the GSE65904 cohort. Patients with high-risk scores had a poor prognosis and survival time decreased with the increase of high-risk scores. In addition, univariate COX and multivariate COX regression analyses of risk scores and other clinical
factors found that risk scores could be used as independent prognostic factors. A nomograph consisting of risk scores and other clinical factors was used to predict survival. Our study suggested that the nine-IRGs can be used as prognostic markers and indicators for patients with metastatic melanoma.

Our prognostic model consisted of nine genes, ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2, and TRIM22. ALOX5AP (arachidonic acid 5-lipoxygenase-activating protein) encodes a protein that is required for the synthesis of leukotrienes which are metabolites of arachidonic acid and are involved in various types of inflammatory reactions. ALOX5AP has been reported to be involved in serous ovarian cancer progression through M2 macrophage recruitment and polarization and to mediate immunosuppression in the tumor immune microenvironment[26]. ARHGAP15 is a member of the RhoGAP family and is involved in a variety of biological processes including cell proliferation and migration. Shengli Pan et al. suggested that it may play an oncogenic role in colorectal cancer progression and metastasis through the PTEN/ AKT/ foxo1 signaling pathway[27]. CCL8, an antimicrobial gene, is one of several chemokine genes clustered on the q-arm of chromosome 17. Chemokines form a superfamily of secreted proteins involved in immunoregulatory and inflammatory processes. Some studies have reported that elevated levels of chemokine CCL8 may contribute to T-cell infiltration[28]. FCER1G, also known as FcRγ, is a key molecule involved in allergic reactions. Liang Chen et al. believe that FCER1G may affect the progression of clear cell renal cell carcinoma by affecting immune-related pathways[29]. GBP4, guanylate-binding protein, is induced by interferon and hydrolyze GTP to both GDP and GMP. It is a guanosine binding protein that is essential for the activation of inflammasomes and can mediate anti-tumor effects through the inflammasome adaptor ASC. Yan Qu et al. and BaohuiHu et al. found that it correlated with cancer prognosis and could be used as a prognostic biomarker for melanoma[13, 30]. HCK is a member of the Src family of tyrosine kinases and is mainly expressed in the myeloid and B-lymphocyte lineages. It has been reported that the expression of HCKis controlled by the toll-like receptor (TLR) adapter protein MYD88, which is aberrantly expressed in MCL cell lines and primary MCL via the TLR agonist, leading to the development of lymphoma[31]. MMP9 (matrix metallopeptidase 9) is the most important family of proteases associated with tumorigenesis. There is a growing number of studies reporting a link between the MMP family and cancer immunotherapy. For example, Youqiong Ye et al. reported that MMP9 inhibitors could improve the efficacy of anti-PD -1 and anti-CTLA4 treatment in mouse models of melanoma and lung cancer, as well as in pulmonary metastatic melanoma[32]. RARRES2 (retinoic acid receptor responder 2) encodes a secreted chemotactic protein that initiates chemotaxis via the ChemR23 G protein-coupled seven-transmembrane domain ligand. Endocrine Oncology Branch et al. found that RARRES2 was overexpressed in adrenocortical carcinomas and exerts oncogenic effects by promoting βcatenin phosphorylation/degradation and inhibiting p38 mitogen-activated protein kinase phosphorylation[33]. The protein encoded by TRIM22 is a member of the tripartite motif (TRIM) family. TRIM22 has been reported to inhibit endometrial cancer progression through the NOD2/NFκappaB signaling pathway and provide a good prognosis[34]. The role and mechanism of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 in the response to anti-tumor immunity and immunotherapy need to be further studied.
To explore the relationship between prognostic models and tumor-infiltrating immune cells, we used three commonly used methods for assessing immune infiltrating cells, including CIBERSORT, Xcell, and ssGSEA. Due to the complexity and shortcomings of these algorithms, they are rarely compared with each other. CIBERSORT results indicated that plasma cells, T cells CD8, T cells CD4 memory activated and macrophages M1 had higher infiltration levels in the low-risk group, while M2 macrophages were highly expressed in the high-risk group. The results of Xcell and ssGSEA showed that the infiltration level of most immune cell types in the low-risk group was significantly higher than that in the high-risk group. Therefore, the immunotherapy effect of patients in the low-risk group may be better.

Interestingly, we found that the risk score correlated not only with immune cell infiltration, but also with immune checkpoint inhibitors. Immune checkpoint inhibitors (ICIs) have become an important therapeutic option for many patients with malignant melanoma over the past few years, aiming to restore and promote the function of effector T cells to specifically recognize and kill tumor cells and systemically enhance the systemic anti-tumor immune response[35]. It is a better treatment option for patients with high risk of recurrence after surgical resection or for those with advanced disease (unresectable or with metastases)[36]. Although immune checkpoint inhibitors (ICIs) have been written into the clinical application guidelines for various recurrent malignant tumors and advanced cancers, they are only effective for a long time in a part of the population, and drug resistance or immune-related adverse events (irAEs) may occur in another part of the population[3]. Currently, there are two main classes of immune checkpoint inhibitors: antibodies targeting cytotoxic T-lymphocyte antigen-4 (CTLA-4) and antibodies targeting programmed cell death protein-1 (PD-1) and its ligand programmed death receptor ligand-1 (PD-L1)[37]. Ipilimumab, the first ICIs approved for marketing by the FDA, is an antibody targeting CTLA-4 and is primarily used for the treatment of melanoma[38]. The high expression of PD-L1 protein in tumor cells or tumor immune microenvironment is considered to be a biomarker for predicting the treatment of melanoma by PD-1 inhibitors, and has been approved by the FDA as a predictive indicator for selecting patients with potential for ICI therapy[39, 40]. Our study found the expression of CTLA4, HAVR2, LAG3, TIGIT, CD274 and PDCD1 is higher in the low-risk group, so it is possible that the low-risk group had a better response to immune checkpoint inhibitor therapy, which is consistent with the poor prognosis of patients in the high-risk group.

To further understand the relationship between prognostic models and genes, the top 25% of the genes with variance ranking were selected for WGCNA analysis. The results showed that the yellow-green module had the highest correlation with the risk score model, and ALOX5AP, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 of the nine-IRGs prognostic model were located in the yellow-green module. GO and KEGG analysis revealed that the genes of the yellow-green module were mainly enriched in immune-related functions and pathways. In addition, our analysis revealed that ALOX5AP, ARHGAP15, CCL8, FCEER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 had better overall survival and were good prognostic genes in the high (median) expression group, and ALOX5AP, ARHGAP15, CCL8, FCEER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 were also more highly expressed in the low-risk score group.
Despite the remarkable results of this study, there were still shortcomings. Although our study provided new insights into the immune microenvironment of metastatic melanoma and related therapeutic targets, it had its limitations because it was retrospective. Therefore, further prospective studies were needed to confirm our findings.

Conclusion

A nine-IRGs prognostic model was identified based on the tumor microenvironment, which can effectively predict survival outcomes in patients with metastatic melanoma, and the model was validated in both internal and external datasets. This model could also predict the infiltration of immune cells in metastatic melanoma. Our research provided a potential model and biomarker for further immune-related work and personalized medicine for the treatment of metastatic melanoma.

Abbreviations

TCGA: The cancer genome atlas; GEO: Gene expression omnibus; DEGs: Differentially expressed genes; IRGs: Immune-related genes; ROC: Receiver operating characteristic; WGCNA: Construction of weighted gene co-expression network analysis; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; SKCM: Skin cutaneous melanoma; LASSO: The least absolute shrinkage and selection operator; BP: Biological processes; CC: Cellular components; MF: Molecular functions; OS: Over Survival; GSEA: Gene set enrichment analysis; TOM: Topological overlap matrices; FPKM: fragments per kilobase million; TPM: transcripts per kilobase million; ICIs :Immune checkpoint inhibitors.

Declarations

Disclosure of potential conflicts of interest All authors declare that they have no competing interests.

Research involving Human Participants and/or Animals This article is not involved in any studies with human participants or animals performed by any of the authors.

Informed consent All authors consent to the publication of this research.

Funding This study was supported by the National Natural Science Foundation of China (grant No. 31300737), the Natural Science Foundation of Guangdong Province (grant Nos. 2018A030313114, 2018A030313860 and 2020A1515010889) and the Research Development Fund of Dongguan People's Hospital (grant No. 2021-5-02).

Availability of data and materials The gene expression datas used in this study were downloaded from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/) and Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

Code availability Not applicable.
Authors' contributions Jing Liu and Han Shen designed the study. Jing Liu and Ting Ye performed the bioinformatics analysis. Xuefang Zhang and Yongjian Dong contributed to the clinical information analysis. WenFeng Zhang, Hongwei Shao and Fenglin Wu wrote the manuscript. Rongxin Zhang, and HuaBen Bao drew the figures. All authors reviewed and proofed the final manuscript.

References


Figures
Figure 1

Analysis of stromal, immune and estimate scores with the survival of samples in the TCGA-SKCM and GSE65904 cohorts. A-C Survival curves of immune, stromal and estimate score in the TCGA-SKCM cohort. D-F Survival curves of immune, stromal and estimate score in the GSE65904 cohort.
Figure 2

Figure 3

Establishment of the nine-IRGs prognostic model in the TCGA-SKCM cohort. A-B LASSO regression identified 17 IRGs most associated with OS. C Hazard ratio forest plots showed the predictive value of IRGs. D-E Distribution of risk scores and survival status of patients in high-risk and low-risk groups. F Kaplan-Meier analysis of OS in low-risk and high-risk groups. G ROC curves of the nine-IRGs prognostic model for 1, 2, and 3 years.
Figure 4

Validation of the nine-IRGs prognostic model in the GSE65904 cohort. A-B Distribution of risk scores and survival status in the high-risk and low-risk groups in the GSE65904 cohort. C Kaplan-Meier analysis of OS in the low-risk and high-risk groups in the GSE65904 cohort. D ROC curves of the nine-IRGs prognostic model at 1, 2, and 3 years in the GSE65904 cohort.
Figure 5

The risk score combined with other clinical factors predicted the prognosis of patients with metastatic melanoma in the TCGA-SKCM cohort. A Univariate Cox regression analysis of clinical characteristics in the TCGA-SKCM cohort. B Multivariate Cox regression analysis of clinical characteristics in the TCGA-SKCM cohort. C A nomogram for predicting the survival rate of patients with metastatic melanoma in 1,
2, and 3 years. D Nomogram calibration curve for predicting 1-year OS. E Nomogram calibration curve for predicting 2-year OS. F Nomogram calibration curve for predicting 3-year OS.

Figure 6

GSEA analysis of the high-risk and low-risk groups of the TCGA-SKCM cohort according to the nine-IRGs prognostic model. A Four typical GO enrichments in the low-risk group. B Four typical KEGG pathways in the low-risk group.
Figure 7

Figure 8

Correlation between high-risk and low-risk groups and immune checkpoints CTLA4, HAVR2, LAG3, TIGIT, CD274 and PDCD1.
Figure 9

Construction of weighted gene co-expression networks and identification of key modules with high correlation to risk scores. A Hierarchical clustering dendrogram of TCGA-SKCM samples. Clinical features of survival status, survival time, age, sex, stage, TNM classification and risk score were shown at the bottom. B Hierarchical clustering dendrogram of different similarity genes based on topological overlap. Modules were branches of the clustering tree. C Correlation of module feature genes with clinical traits. The rows correspond to the characteristic genes of the module and the columns represented the clinical characteristics. Each cell contained correlations and p-values. D Scatter plot of yellow-green module genes.
Figure 10

Figure 11

The relationship between the expression of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 and the prognosis and development of metastatic melanoma. A Kaplan-Meyer analysis of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22. B Association between the expression of ALOX5AP, ARHGAP15, CCL8, FCER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22 and risk scores.
Figure 12

Immune infiltration of ALOX5AP, ARHGAP15, CCL8, FCLER1G, GBP4, HCK, MMP9, RARRES2 and TRIM22