

Immune Cell as a Promising Biomarker in the Diagnosis and Prognosis of Cutaneous Melanoma by Using Machine Learning

Jing Tang

People's Hospital of Leshan

Hongqun Ye

People's Hospital of Leshan

Wan qi (✉ wang9@mial2.sysu.edu.cn)

People's Hospital of Leshan

Research

Keywords: immune cell, diagnosis, prognosis, cutaneous melanoma

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-327019/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Tumor infiltration, is known to associate with various cancer initiations and progressions, is potential therapeutic target for this aggressive skin cancer.

Methods: single sample gene set enrichment analysis (ssGSEA) algorithm was applied to assess the relative expression of 24 types of immune cell from public database. Firstly, the differentially expressed immune cells between melanomas and normal samples were identified. Next, multiple machine learning algorithms were performed to evaluate the efficiency of immune cells in diagnosis of melanoma. In addition, the feature selection in machine learning methods was used to figure out the most important prognostic immune cells for developing biomarker to predict the prognosis of melanoma.

Results: In comparison with the expression of immune cells in tumors and normal controls, we built the immune diagnostic models in training dataset, which can accurately classify melanoma patients from normal (LR AUC= 0.965, RF AUC= 0.99, SVM AUC=0.963, LASSO AUC= 0.964 and NNET AUC=0.989). These diagnostic models also validated in three outside datasets and suggested over 90% sensitivity and specificity to distinguish melanomas from normal patients. Moreover, we also developed a robust immune cell biomarker which could estimate the prognosis of melanoma. This biomarker also further validated in internal and external datasets. Next, we constructed nomogram combined risk score of biomarker and clinical characteristics, which showed good accuracies in predicting 3 and 5 years' survival. The decision curve of nomogram model manifested a higher net benefit than tumor stage. In addition, melanoma patients divided into high and low risk subgroups by applied risk score system. The high risk group have a significantly shorter survival time than the low risk subgroup. Gene Set Enrichment Analysis (GSEA) analysis revealed that complement, epithelial mesenchymal transition and inflammatory response and so on significantly activated in high risk group.

Conclusions: We constructed immune cell related diagnostic and prognostic models, which could provide new clinical applications for diagnosing and predicting the survival of melanoma patients.

1. Introduction

Melanoma is the most aggressive type of skin cancer which derived from melanocyte lineage with the highest metastasis and mortality rate[1]. Despite melanoma consist of only 5% of all skin-related cancers, it's accounting for approximately 80% of deaths related to skin tumor. Like most other cancers, surgical enucleation and drug therapy are difficult to treat once it has metastasized[2–4]. In addition, it is difficult to detect at early, the majority of patients with melanoma were diagnosed at an advanced stage[5, 6]. Currently, cancer treatment guiding and prognosis predicting are largely determined by TNM staging system. However, clinical experience revealed that many patients even within the same TNM stage have differences in the overall survival[7]. The clinical limitations of TNM stage are increasingly becoming apparent. Thus, it is urgently required to explore the new biomarkers for early diagnosis and prognostic prediction.

Growing studies have recently reported that the tumor microenvironment plays crucial roles in the development and progression of many malignant tumors[8, 9]. In tumor microenvironment, the type, function and location of immune cell are intimately associated with clinical outcome[10, 11]. Tumor cells regarded as antigens which will attract immune cells and leukocytes by many chemokines to influence immune response. Moreover, immune escape of tumor cell was considered as a crucial factor in tumorigenesis[12–14]. The prognostic value of tumor microenvironment also had been demonstrated in numerous melanoma experiments. High immune infiltration in melanoma also have shown to be associated with favorable prognosis. To date, immunotherapy is pivotal for the treatment of patients with advanced melanoma, such as CTLA-4, PD-1 and PD-L1 inhibitors[15, 16]. Therefore, prognostic or predictive biomarkers related to tumor microenvironment may hold great promise for identifying novel molecular targets and guiding patient management. Besides tumor cells, melanoma also commonly includes various types of immune cell, which may regard as a potential diagnostic signature to classify tumors from suspected patients. Thus, systematically evaluated infiltrating immune cells was recognized a significant supplemental biomarker to the TNM stage for diagnosis and prognosis prediction. Fortunately, the availability of public, large-scale datasets like the cancer genome atlas (TCGA) and Gene Expression Omnibus (GEO) databases which provided numerous transcriptome profiles to investigate the landscape of immune cells.

Machine learning is a powerful tool to analyze and summarize complex datasets, which can provide various computational approaches to predict clinical diseases[17]. Previously, several algorithms have been successfully applied to diagnose and predict disease, including logistic regression model[18], support vector machine[19], random forests analysis[20] and artificial neural network[21]. Compared to classical methods, machine learning often reach remarkable high sensitivity and specificity. What's more, machine learning is the best choice to process the increasingly growing genomic data and clinical information in oncology research and predict the susceptibility, survival and recurrence of cancer.

Therefore, in this research, we first estimate the proportions of 24 immune cells in 944 samples (653 tumors and 291 normal controls) according to their gene expression profiling download from public databases. Next, we used multcategory machine learning to identify several important immune cell signatures as well as construct diagnostic and prognostic models, which show more effective performance for the diagnosis and prognosis of melanoma patients.

2. Materials And Methods

2.1 Melanoma collection and normal controls

The melanoma patients were collected from public datasets. The eligible datasets were downloaded from the Xena Public Data Hubs (<http://xena.ucsc.edu/public-hubs/>) and GEO database (<https://www.ncbi.nlm.nih.gov/geo>). Finally, five datasets consist of 944 samples including TCGA, GSE3189, GSE15605, GSE46517 and GSE54467 were screened out for this study. The raw transcriptome

Loading [MathJax]/jax/output/CommonHTML/jax.js

annotated according to the annotation information of platform. For the same gene corresponding to multiple IDs, the median expression value will be calculated to represent the gene expression level. Next, genes with a variance of 0 will be excluded for its tiny expression level. Finally, the raw matrix data were normalized by $\log_2(x + 1)$ conversion.

2.2 Estimation of immune cell types

To estimate the proportions of immune cells, single sample gene set enrichment analysis (ssGSEA) algorithm were applied to transform the normalized gene expression data into 24 human immune cell types expression which including natural killer (NK) cells, CD56bright NK cells, CD56dim NK cells, dendritic cells (DCs), activated DCs (aDC), immature DCs (iDC), plasmacytoid DCs (pDC), neutrophils, macrophages, eosinophils, mast cells, T cells, B cells and cytotoxic cells. Moreover, the T cells consist of T effector memory (Tem), T central memory cells (Tcm), CD8 T cells, Tgd cells, regulatory T cells (Treg), T helper cells and T follicular helper cells (TFH), Th1, Th2, Th17[22, 23].

2.3 Diagnostic analysis

Firstly, these samples were classified into tumor tissue groups and normal tissue groups. Next, the difference analysis of immune cells in tumor tissues and normal tissues was performed and p values < 0.05 considered as the differentially expressed immune cells (DEICs). Then, the Upset plot analysis was used to explore the overlap of DEICs among these datasets. To develop a diagnostic model with selected DEICs, five machine learning algorithms including Logistic Regression (LR), Random Forests (RF), Support Vector Machines (SVM), Least Absolute Shrinkage and Selection Operator (LASSO), neural network (NNET) and five-fold cross-validation were systematically used to construct the models in TCGA dataset and the sensitivity and specificity of the diagnostic models was evaluated by the receiver operating characteristic (ROC) curves. Principal component analysis was performed to determine whether these DEICs could definitely classify tumors form normal controls. The diagnostic model also validated in another three independent datasets. Besides, the coefficients of LR were used to calculate the diagnostic score of each sample and the formula is $\text{Diagnosis score} = \sum_{i=1}^N (\text{coef}_i \times \text{expr}_i)$, which could well distinguish normal and tumor tissues.

2.4 Prognostic analysis

To explore the most significant immune cells in the prognosis of melanoma patients, the eligible patients in TCGA dataset were randomly separated into training and testing samples (1:1). Multicategory machine learning methods including LASSO, SVM- recursive feature elimination (SVM-RFE), RF-feature selection (RF-FS) were conducted to identified the important immune cells in training dataset. Then, cox regression analysis method was performed to develop prognostic model with selected cells for the risk formula and the risk score is generated as follows: $\text{Risk score} = \sum_{i=1}^N (\text{coef}_i \times \text{expr}_i)$, in which N means the number of feature cells, expr_i means the expression level of cells and coef_i means regression coefficient. The risk score of each sample in training dataset was estimated and the patients were accordingly classified into high- and low-risk group by optimal cutoff value. To compare the differences between high- and low-risk

group, Kaplan–Meier survival curves were drawn and significance was calculated by log-rank tests. The area under the curve (AUC) of Receiver operating characteristic curves (ROC) was used to evaluate the 5-year overall survival predictive accuracy of the model. Besides, to proven the robustness of the result, these immune cell features were further validated in testing dataset and GSE54467. Finally, nomograms were established in this study based on patients' clinical characteristics and risk score. To compare the prediction and actual survival, calibration curves were drawn. And decision curves were also plotted to discriminate the clinical usefulness of nomogram and tumor stage.

2.5 Stratified analysis

To assess the relationship between risk score distribution and clinical characteristics, the subgroup analysis of clinical variables included age, gender, race, stage, vital status and tumor status were performed. Moreover, the univariate and multivariate cox logistic regression were carried out to compare the prognostic value between the risk score and clinical variables. Next, to investigate the potential biological phenotypes between high and low risk groups, gene expression data corelated to immune checkpoint regulators and epithelial–mesenchymal transition (EMT) were analyzed[24–27]. Firstly, the expression data of these genes were extracted from the TCGA dataset and divided into high and low risk groups by applying the optimal cutoff value. Finally, subgroup analysis of these genes was conducted.

2.6 Gene set enrichment analysis

In order to explore the different signaling pathways between the low- and high-risk groups, Gene Set Enrichment Analysis (GSEA) was conducted by “clusterProfiler” package in R software. Firstly, the differential analysis of all genes between low- and high-risk groups were generated and these genes were ordered by the value of log2 fold change. Then the cancer hallmark pathway database (h.all.v7.0.symbols) and KEGG pathway database (c2.cp.kegg.v7.0.symbols) in GSEA were performed to investigate the signaling pathways correlated with different subgroups of melanoma. The random sample permutations were 1000 and the q value < 0.05 was the significance threshold.

2.7 Statistical analysis

All statistical analyses were conducted using R package (v.3.6.0) and corresponding packages. The Upset plot was drawn by the “UpSetR” package. ssGSEA method was estimated by “GSVA” package. LASSO and LR analysis were calculated by the “glmnet” package. SVM and SVM-RFE methods were conducted by “e1017” package. RF and RF-FS algorithms were applied by “randomForest” and “varSelRF” packages respectively. NNET method were performed by “nnet” package. The optimal cutoff values computed by using the “survminer” package. Kaplan–Meier survival analysis was applied by using “survival” and “survivalROC” packages. GSEA analysis was conducted by “clusterProfiler” package. The hazard ratios (HR) and 95% confidence intervals (95% CI) of the prognostic factors were calculated. $P < 0.05$ was regarded as statistically significant in all statistical tests.

3. Results

3.1 Melanoma collection and normal controls

Totally, 944 samples were selected for the subsequent analysis, which were acquired from the five datasets including TCGA of melanoma, GSE3189, GSE15605, GSE46517 and GSE54467. The TCGA of cutaneous melanoma were downloaded from the Xena Public Data Hubs, which contained 372 melanoma tumors and 233 normal controls. The GSE3189 contained 45 melanoma tumors and 25 normal controls. GSE15605 included 58 melanoma tumors and 16 normal healthy controls. GSE46517 included 121 samples which consist of 104 melanoma tumors and 17 normal controls. GSE54467 dataset only had 74 melanoma samples with no healthy control. Moreover, 870 samples obtained from TCGA of melanoma, GSE3189, GSE15605 and GSE46517 were used to diagnostic analysis. 446 melanoma samples obtained from TCGA of melanoma and GSE54467 were conducted to prognostic analysis. The complete analysis workflow in this study is illustrated in Fig. 1.

3.2 Differentially expressed immune cells (DEICs)

Firstly, the 24 immune cells expression matrix was calculated by ssGSEA. According to standard of differential analysis, 19 DEICs were identified in TCGA dataset in which 11 cells were up-regulated and 8 genes were down-regulated (Fig. 2A). 13 DEICs contained 4 up-regulated and 9 down-regulated cells were distinguished in GSE3189 (Fig. 2B). 11 DEICs were found in GSE15605, which consist of 6 up-regulated cells and 5 down-regulated cells (Fig. 2C). Besides, 7 up-regulated cells and 6 down-regulated cells were identified in GSE46517 (Fig. 2D). Eventually, 6 overlap of DEICs were found in the four datasets (Fig. 2E). These immune cells including iDC, DC, Eosinophils, NK CD56bright cells, Mast cells, Treg were selected for subsequent research.

3.3 Construct diagnostic model

In order to develop a diagnostic model by using the 6 identified DEICs and assess the effectiveness, five machine learning algorithms were comprehensively conducted to diagnose melanoma samples from normal healthy controls. Additionally, we also applied 5-fold cross-validation to estimate the accuracy of each model in TCGA dataset. The ROC curves manifested that the six DEICs can accurately classify melanoma patients from normal (LR AUC = 0.965, RF AUC = 0.99, SVM AUC = 0.963, LASSO AUC = 0.964 and NNET AUC = 0.989) (Fig. 2F). This diagnostic models also validated in GSE3189, GSE15605 and GSE46517. Similarly, the ROC curves suggested over 90% sensitivity and specificity to distinguish melanomas from normal patients in all datasets (Fig. 2G-I). Principal components analysis showed that tumors and normal controls could be well differentiated based on the expression of six DEICs (Fig. 2J). Next, we built a diagnostic score model with these DEICs by using LR method and calculated the score of each sample according to the diagnostic formula. The diagnostic score distribution of tumors and normal controls was shown in Fig. 2K. The violin plot revealed that the diagnostic values were significant difference between tumor and normal groups.

3.4 Develop prognostic model

Loading [MathJax]/jax/output/CommonHTML/jax.js

After eliminating patients with no survival information, 432 melanoma patients were obtained from the TCGA of melanoma and GSE54467. Firstly, the TCGA of melanoma dataset were randomly classified into training samples (N = 179) and testing samples (N = 179). The clinical statistical information of training and testing samples is displayed in Table 1, and results indicated that there is no difference between two datasets. Next, Combining the feature selection results of LASSO method (Fig. 3A), RF-FS method (Fig. 3B) and SVM-RFE method (Fig. 3C) showed that four overlapping immune cells were selected out (Fig. 3D). Then these immune cells were further used to construct a risk score system by applying Cox analysis in the training dataset. By applying this risk model, a risk score for each sample will be calculated. The distributions of the risk scores, overall survival (OS) time, vital status, and expression levels of corresponding immune cells in training (Fig. 3E-G), testing (Fig. 3H-J) and GSE54467 (Fig. 3K-M) datasets were respectively shown. Afterwards, melanoma patients in training dataset were classified into a high-risk group and a low-risk group by applying the optimal cut-off value of the risk scores. Kaplan-Meier curves showed that patients in high-risk group have a shorter survival time than low-risk with a log-rank test of $p = 0.003$ (Fig. 4A). To estimating the prediction power of features, the ROC curve was drawn and 5 years of AUC was 0.664 (Fig. 4B). Besides, in order to confirm the robustness of the result, verification tests were conducted in testing and GSE54467 datasets. The testing and GSE54467 datasets were divided into high-risk and low-risk groups based on training dataset. Kaplan-Meier curves showed that there is a significant difference between high-risk and low-risk group both in testing dataset (log-rank $p < 0.001$) (Fig. 4C) and GSE54467 dataset (log-rank $p = 0.002$) (Fig. 3E). The 5 years of AUC were 0.832 (Fig. 3D) and 0.729 (Fig. 4F) respectively.

3.5 Nomogram building and validating

In order to supply a simple and accuracy method for OS prediction, nomogram was built on the basis of clinical information and risk scores of patients in train dataset (Fig. 5A). Then, a total point summarized the points of each parameter, which can predict the probability of OS at 3 and 5 years. The calibration plots suggested that the nomogram performed well in comparison with ideal model (Fig. 5D). Moreover, nomograms constructed by testing (Fig. 5B) and GSE54467 (Fig. 5C) were used to validate the results. Surprisingly, calibration plots in testing (Fig. 5E) and GSE54467 (Fig. 5F) for nomogram predicting 3 and 5 years' OS also performed well with the ideal model. Similarly, by applying the decision curve analysis, the nomogram model offered a higher net benefit and a better clinical utility than tumor stage no matter in training (Fig. 5G) or testing (Fig. 5H) and GSE54467 (Fig. 5I) datasets.

3.6 Associations between risk model with clinical characteristics and gene phenotypes

The association between risk model and clinical characteristics was investigated and the violin plot manifested that only vital status and tumor status were correlated with risk score (Fig. 6A). Other clinical characteristics, such as age, gender, race and stage had no relationships with risk score. Thus, the risk model was associated with two clinical variables and had on effect on other clinical characteristics.

Furthermore, in comparison the prognostic value of risk score with clinical variables, univariate and multivariate logistic regression were applied in training, testing, and GSE54467 datasets (Table 2). The univariate regression showed that age, stage, tumor status and risk score were significantly associated with OS, but the multivariate regression revealed that only the risk score were significantly correlated with OS and could be regarded as an independent risk factor in training (HR = 3.517, P = 0.005), testing (HR = 1.869, P = 0.042), and GSE54467 (HR = 2.661, P < 0.000). To evaluate the associations between risk model and selected immune checkpoint related genes, subgroup analysis of immune checkpoint related genes was performed. The violin plot revealed that CD28, CTLA4, ICOS, PDCD1, TIGIT, CD274, CD226, CD40 and CD40LG in high risk group had a higher expression value than those in low risk group (Fig. 6B). Interestingly, subgroup analysis of EMT related genes showed that majority of EMT related genes were differently expressed between high and low risk group. The high risk group had significantly higher expression levels of CTNNB1, FGF2, EGFR, SNAI2, ZEB1, CXCL12, SNAI1 and PDGFB (Fig. 6C).

3.7 Gene set enrichment analysis

To investigate the significant pathways shared by different high and low risk group, we performed cancer hallmark and KEGG pathway enrichment by GSEA analysis. Based on selection standard and q values. Multiple significant cancer hallmark pathways were enriched, such as allograft rejection, complement, EMT and inflammatory response (Fig. 6D). Additionally, KEGG enrichment showed that complement and coagulation cascades, ECM receptor interaction, natural killer cell mediated cytotoxicity and T cell receptor signaling pathway were positively enriched in high risk group (Fig. 6E).

4. Discussion

Recently, melanoma patients are growing younger and with highly metastasize and deadly threatening, which places a huge burden to thousands of people worldwide. In spite of numerous advanced therapeutic methods were used to treat melanoma, such as chemotherapy and radiotherapy and immunotherapies, their survival rate still remains low[1, 3]. Besides, the traditional classification is often ineffective and lacks clinical benefits. Therefore, researchers are struggling to explore the new biomarkers for better diagnose and predict prognosis. Huang et al identified eight immune-related genes biomarker which could predict the prognosis of melanoma[28]. An RNA sequencing-based 12-gene signature was established by applying univariate and multivariate regression models to predict the survival of malignant melanoma patients[29]. Lu et al discovered a five-miRNA signature by analyzing microarray dataset in GEO database, which could be an independent prognostic biomarker in melanoma patients[30]. Recently, the tumor immune microenvironment in melanoma has become a research hotspot and under active investigation[31]. Moreover, the immune cell types which differentially distributed in cancer tissue on diagnosis has attracted great interest in recent years. Therefore, in this study, we systematically analyzed the immune microenvironment and tried to establish a more evaluable and precise signature for advanced melanoma patients.

In spite of various differential expression of genes were analyzed to diagnose tumors. Nevertheless, little research attention looked at the effects of immune cell on the diagnosis of melanoma. Firstly, we used the ssGSEA method to calculate the relative expression of 24 human immune cells. Because of compared to normal tissues, the distribution of immune cell was significantly higher in tumor tissues. The overlapping DEICs were identified and put into machine learning analysis. The high sensitivity and specificity of multiple machine learning algorithms indicated that DEICs was an efficient indicator for diagnosis of melanoma. In addition, we built a diagnostic score model by logistic regression method, which could effectively distinguish the melanomas from the normal controls, replying that the immune system is closely associated with the tumorigenesis of melanoma. Similar results have been reported that infiltration of immune cell can be used to diagnose colon cancer, even all digestive system cancers[32, 33]. In this sense, immune infiltration opened a novel strategy for diagnosing and treating melanoma.

To subsequently investigate the prognostic value of immune infiltration in melanoma, LASSO, RF-FS and SVM-RFE methods jointly applied to select potential immune cells for building the prognostic model. Finally, four types of immune cell including Th2 cells, T helper cells, Macrophages, iDC were used to construct the risk score system by Cox regression method, which also was validated in internal and external datasets. Among these immune cells, some have been proven to be associated with melanoma. For instance, approximate 70% of melanoma metastatic lymph nodes were detected the distribution of immature DCs which may take an immunosuppressive function in melanoma[34]. Under normal immune environment, Th1 cells and Th2 cells are in a relatively balanced state. Th2 bias signifies the imbalance of Th1/Th2. Th2 could strongly inhibit Th1 responses[35]. Th2 bias is one of the mechanisms of tumor immune escape. Studies have also shown that Th2 dominance could mediate a chronic inflammation which could promote melanoma metastasis. Moreover, Falleni et al confirmed that the accumulation of Macrophages was a poor predictor for survival of melanoma patients and could be regarded as a potential therapeutic target[36]. In order to assess the accuracy of prognostic prediction, we also constructed nomogram integrate risk score and clinical information. The calibration curve for the observed 3-year, and 5-year outcomes showed that the nomogram model performed well with the ideal prediction model. What's more, compared with the tumor stage, the decision curve plots depicted that the nomogram model can acquire more benefit. The multivariate regression analysis also indicated that the risk score of immune cell related biomarker could be regard as an independent prognostic factor in melanoma.

According to the optimal cutoff value of risk score, melanoma patients classified into different risk groups. The Kaplan–Meier revealed that patients in high risk group have a poor prognosis. Thus, in order to explore the underlying mechanism with different subgroups, stratified analyses of clinical characteristics and gene phenotypes were performed. The risk score distribution of clinical features showed that the risk score was only correlated to vital status and tumor status, and had on effect on other clinical features. Presently, checkpoint blockade immunotherapies represent a promising strategy for cancer therapy and acquired extensive investigations[37, 38]. However, the efficacy of

Loading [MathJax]/jax/output/CommonHTML/jax.js individual patients and different subtypes of cancer. In our

research, the expression of immune checkpoint related genes including CD28, CTLA4, ICOS, PDCD1, TIGIT, CD274, CD226, CD40 and CD40LG were highly expressed in high risk of melanoma patients. Besides, epithelial mesenchymal transition (EMT) recognized the indicator for the invasion and progression of many cancers[39, 40]. The selected EMT related genes also including our research and the results also manifested that most of them are highly expressed in high risk group. Hence, we have enough reasons to believe that our immune cell feature closely correlated with the prognosis of melanoma.

Further investigating the potential biological mechanism in high risk phenotype, GSEA method was applied to analyze the candidate pathways. The results showed that high risk phenotype was positively associated with cancer hallmarks such as allograft rejection, complement, EMT and inflammatory response, which supported the previous findings that EMT and immune related genes are highly expressed in high risk group. The complement system, an essential constituent of innate immunity, affects tumor growth and metastasis by regulating chronic inflammation. Moreover, KEGG pathway analysis showed that complement and coagulation cascades, ECM receptor interaction, natural killer cell mediated cytotoxicity and T cell receptor signaling pathways were enriched in high risk phenotype, which largely consisted with cancer hallmarks analysis. ECM-receptor interaction pathway is important in tumor metastasis[41]. The significance of the ECM-receptor interaction pathway implied the interaction between tumor cell and environment are very dynamic[42].

5. Conclusion

In summary, our study identified several differential immune cells and proven the efficiency of immune cells in diagnosing and predicting prognosis of melanoma. The constructed diagnosis and prognosis models might provide a more simple and accurate prediction for melanoma patients in clinical application and management.

Abbreviations

ssGSEA
single sample gene set enrichment analysis
DEMIncRNAs
Differential expressed macrophage associated lncRNAs
TCGA
The Cancer Genome Atlas database

Declarations

Ethics approval and consent to participate

No permissions were required to use the repository data

Loading [MathJax]/jax/output/CommonHTML/jax.js

Consent for publication

Not applicable

Availability of data and materials

The datasets used and analysed during the current study available from the corresponding author on reasonable request

Competing interests

All authors declare that they have no competing interests.

Funding

There is no sponsorship or funding arrangements relating to our research

Authors' contributions

JT designed the study. QW and HQY wrote the paper. All authors read and approved the final manuscript.

Acknowledgment

Not applicable

References

1. Ascierto PA, Eggermont AMM. Neoadjuvant therapy in melanoma: the next step? *Lancet Oncol.* 2018;19(2):151–3.
2. Elder DE, Van Belle P, Elenitsas R, Halpern A, Guerry D. Neoplastic progression and prognosis in melanoma. *Semin Cutan Med Surg.* 1996;15(4):336–48.
3. Jakub JW, Racz JM, Hieken TJ, Gonzalez AB, Kottschade LA, Markovic SN, Yan Y, Block MS. Neoadjuvant systemic therapy for regionally advanced melanoma. *J Surg Oncol.* 2018;117(6):1164–9.
4. McKean MA, Amaria RN. Multidisciplinary treatment strategies in high-risk resectable melanoma: Role of adjuvant and neoadjuvant therapy. *Cancer Treat Rev.* 2018;70:144–53.
5. Cancer Genome Atlas N. Genomic Classification of Cutaneous Melanoma. *Cell.* 2015;161(7):1681–96.
6. Warner AB, Postow MA. The brim of uncertainty in adjuvant treatment of melanoma. *Lancet Oncol.* 2018;19(4):436–7.
7. Neagu M, Constantin C, Tanase C. Immune-related biomarkers for diagnosis/prognosis and therapy monitoring of cutaneous melanoma. *Expert Rev Mol Diagn.* 2010;10(7):897–919.

8. Chen F, Yang Y, Zhao Y, Pei L, Yan H. Immune Infiltration Profiling in Nonsmall Cell Lung Cancer and Their Clinical Significance: Study Based on Gene Expression Measurements. *DNA Cell Biol.* 2019;38(11):1387–401.
9. Kumpers C, Jokic M, Haase O, Offermann A, Vogel W, Gratz V, Langan EA, Perner S, Terheyden P. Immune Cell Infiltration of the Primary Tumor, Not PD-L1 Status, Is Associated With Improved Response to Checkpoint Inhibition in Metastatic Melanoma. *Front Med (Lausanne).* 2019;6:27.
10. Garcia-Salum T, Villablanca A, Matthaus F, Tittarelli A, Baeza M, Pereda C, Gleisner MA, Gonzalez FE, Lopez MN, Hoheisel JD, et al. Molecular signatures associated with tumor-specific immune response in melanoma patients treated with dendritic cell-based immunotherapy. *Oncotarget.* 2018;9(24):17014–27.
11. Zoroquiain P, Esposito E, Logan P, Aldrees S, Dias AB, Mansure JJ, Santapau D, Garcia C, Saornil MA, Belfort Neto R, et al. Programmed cell death ligand-1 expression in tumor and immune cells is associated with better patient outcome and decreased tumor-infiltrating lymphocytes in uveal melanoma. *Mod Pathol.* 2018;31(8):1201–10.
12. Bisgin A, Meng WJ, Adell G, Sun XF. Interaction of CD200 Overexpression on Tumor Cells with CD200R1 Overexpression on Stromal Cells: An Escape from the Host Immune Response in Rectal Cancer Patients. *J Oncol.* 2019;2019:5689464.
13. Walsh SR, Simovic B, Chen L, Bastin D, Nguyen A, Stephenson K, Mandur TS, Bramson JL, Lichty BD, Wan Y. Endogenous T cells prevent tumor immune escape following adoptive T cell therapy. *J Clin Invest.* 2019;129(12):5400–10.
14. Lim SA, Kim J, Jeon S, Shin MH, Kwon J, Kim TJ, Im K, Han Y, Kwon W, Kim SW, et al. Defective Localization With Impaired Tumor Cytotoxicity Contributes to the Immune Escape of NK Cells in Pancreatic Cancer Patients. *Front Immunol.* 2019;10:496.
15. Dabrosin N, Sloth Juul K, Baehr Georgsen J, Andrup S, Schmidt H, Steiniche T, Heide Ollegaard T, Bonnelykke Behrndtz L. Innate immune cell infiltration in melanoma metastases affects survival and is associated with BRAFV600E mutation status. *Melanoma Res.* 2019;29(1):30–7.
16. Ott PA, Hodi FS, Robert C. CTLA-4 and PD-1/PD-L1 blockade: new immunotherapeutic modalities with durable clinical benefit in melanoma patients. *Clin Cancer Res.* 2013;19(19):5300–9.
17. Zhou L, Zhang Z, Chen YC, Zhao ZY, Yin XD, Jiang HB. A Deep Learning-Based Radiomics Model for Differentiating Benign and Malignant Renal Tumors. *Transl Oncol.* 2019;12(2):292–300.
18. Nunes N, Yazbek J, Ambler G, Hoo W, Naftalin J, Jurkovic D. Prospective evaluation of the IOTA logistic regression model LR2 for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol.* 2012;40(3):355–9.
19. Zhang B, Liang XL, Gao HY, Ye LS, Wang YG. **Models of logistic regression analysis, support vector machine, and back-propagation neural network based on serum tumor markers in colorectal cancer diagnosis.** *Genet Mol Res* 2016, 15(2).
20. Fan Y, Murphy TB, Byrne JC, Brennan L, Fitzpatrick JM, Watson RW. Applying random forests to DIGE data for the detection and staging of prostate cancer. *J*

Proteome Res. 2011;10(3):1361–73.

21. Duan X, Yang Y, Tan S, Wang S, Feng X, Cui L, Feng F, Yu S, Wang W, Wu Y. Application of artificial neural network model combined with four biomarkers in auxiliary diagnosis of lung cancer. *Med Biol Eng Comput.* 2017;55(8):1239–48.
22. Ye L, Zhang T, Kang Z, Guo G, Sun Y, Lin K, Huang Q, Shi X, Ni Z, Ding N, et al. Tumor-Infiltrating Immune Cells Act as a Marker for Prognosis in Colorectal Cancer. *Front Immunol.* 2019;10:2368.
23. Shen Y, Peng X, Shen C. Identification and validation of immune-related lncRNA prognostic signature for breast cancer. *Genomics.* 2020;112(3):2640–6.
24. Tao C, Huang K, Shi J, Hu Q, Li K, Zhu X. Genomics and Prognosis Analysis of Epithelial-Mesenchymal Transition in Glioma. *Front Oncol.* 2020;10:183.
25. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol.* 2017;14(11):655–68.
26. Lynes J, Jackson S, Sanchez V, Dominah G, Wang X, Kuek A, Hayes CP, Benzo S, Scott GC, Chittiboina P, et al. Cytokine Microdialysis for Real-Time Immune Monitoring in Glioblastoma Patients Undergoing Checkpoint Blockade. *Neurosurgery.* 2019;84(4):945–53.
27. Lipson EJ, Velculescu VE, Pritchard TS, Sausen M, Pardoll DM, Topalian SL, Diaz LA Jr. Circulating tumor DNA analysis as a real-time method for monitoring tumor burden in melanoma patients undergoing treatment with immune checkpoint blockade. *J Immunother Cancer.* 2014;2(1):42.
28. Huang R, Mao M, Lu Y, Yu Q, Liao L. A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment. *Aging.* 2020;12(8):6966–80.
29. Song LB, Zhang QJ, Hou XY, Xiu YY, Chen L, Song NH, Lu Y. A twelve-gene signature for survival prediction in malignant melanoma patients. *Ann Transl Med.* 2020;8(6):312.
30. Lu T, Chen S, Qu L, Wang Y, Chen HD, He C. Identification of a five-miRNA signature predicting survival in cutaneous melanoma cancer patients. *PeerJ.* 2019;7:e7831.
31. Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, Zhang Y, Shen B, Qian P, Xu S, et al. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. *J Cell Physiol.* 2020;235(2):1025–35.
32. Zhou R, Zhang J, Zeng D, Sun H, Rong X, Shi M, Bin J, Liao Y, Liao W. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage III colon cancer. *Cancer Immunol Immunother.* 2019;68(3):433–42.
33. Yang S, Liu T, Cheng Y, Bai Y, Liang G. Immune cell infiltration as a biomarker for the diagnosis and prognosis of digestive system cancer. *Cancer Sci.* 2019;110(12):3639–49.
34. Polak ME, Johnson P, Di Palma S, Higgins B, Hurren J, Borthwick NJ, Jager MJ, McCormick D, Cree IA. Presence and maturity of dendritic cells in melanoma lymph node metastases. *J Pathol.* 2005;207(1):83–90.
35. Guenova E, Watanabe R, Teague JE, Desimone JA, Jiang Y, Dowlatshahi M, Schlapbach C, Schaeckel K, Rook AH, Tawa M, et al. TH2 cytokines from malignant cells suppress TH1 responses and enforce

- a global TH2 bias in leukemic cutaneous T-cell lymphoma. *Clin Cancer Res.* 2013;19(14):3755–63.
36. Falleni M, Savi F, Tosi D, Agape E, Cerri A, Moneghini L, Bulfamante GP. M1 and M2 macrophages' clinicopathological significance in cutaneous melanoma. *Melanoma Res.* 2017;27(3):200–10.
37. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, Tian T, Wei Z, Madan S, Sullivan RJ, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med.* 2018;24(10):1545–9.
38. Prat A, Navarro A, Pare L, Reguart N, Galvan P, Pascual T, Martinez A, Nuciforo P, Comerma L, Alos L, et al. Immune-Related Gene Expression Profiling After PD-1 Blockade in Non-Small Cell Lung Carcinoma, Head and Neck Squamous Cell Carcinoma, and Melanoma. *Cancer Res.* 2017;77(13):3540–50.
39. Li CY, Wang Q, Wang X, Li G, Shen S, Wei X. Scutellarin inhibits the invasive potential of malignant melanoma cells through the suppression epithelial-mesenchymal transition and angiogenesis via the PI3K/Akt/mTOR signaling pathway. *Eur J Pharmacol.* 2019;858:172463.
40. Feng H, Jia XM, Gao NN, Tang H, Huang W, Ning N. Overexpressed VEPH1 inhibits epithelial-mesenchymal transition, invasion, and migration of human cutaneous melanoma cells through inactivating the TGF-beta signaling pathway. *Cell Cycle.* 2019;18(21):2860–75.
41. Kai F, Drain AP, Weaver VM. The Extracellular Matrix Modulates the Metastatic Journey. *Dev Cell.* 2019;49(3):332–46.
42. Zhang HJ, Tao J, Sheng L, Hu X, Rong RM, Xu M, Zhu TY. Twist2 promotes kidney cancer cell proliferation and invasion by regulating ITGA6 and CD44 expression in the ECM-receptor interaction pathway. *Onco Targets Ther.* 2016;9:1801–12.

Tables

Table 1: Clinical characteristics of train and test dataset. IQR: interquartile range; OS: overall survival.

	level	test.samples	train.samples	p
n		179	179	
age (median [IQR])		58.00 [47.00, 71.00]	58.00 [48.00, 70.00]	0.685
gender (%)	FEMALE	59 (33.0)	73 (40.8)	0.154
	MALE	120 (67.0)	106 (59.2)	
race (%)	[Not Evaluated]	4 (2.2)	1 (0.6)	0.512
	[Unknown]	1 (0.6)	2 (1.1)	
	ASIAN	4 (2.2)	3 (1.7)	
	WHITE	170 (95.0)	173 (96.6)	
stage (%)		16 (8.9)	10 (5.6)	
	I/II NOS	4 (2.2)	4 (2.2)	
	Stage I	32 (17.9)	35 (19.6)	
	Stage II	51 (28.5)	64 (35.8)	
	Stage III	67 (37.4)	58 (32.4)	
	Stage IV	9 (5.0)	8 (4.5)	
vital_status (%)	Alive	94 (52.5)	94 (52.5)	1.000
	Dead	85 (47.5)	85 (47.5)	
tumor_status (%)		2 (1.1)	4 (2.2)	0.452
	TUMOR FREE	76 (42.5)	84 (46.9)	
	WITH TUMOR	101 (56.4)	91 (50.8)	
OS.time (median [IQR])		3.16 [1.51, 6.56]	3.47 [1.39, 6.68]	0.831

Table 2: Univariate and multivariate Cox regression analyses of clinical characteristics associated with overall survival in train, test and GSE54467 datasets. IQR: interquartile range. HR: hazard ratio.

Train sample (n = 179)	Univariate analysis				Multivariate analysis			
	pvalue	HR	lower .95	upper .95	pvalue	HR	lower .95	upper .95
marker								
age	0.000	1.028	1.013	1.044	0.004	1.024	1.007	1.041
gender	0.880	0.965	0.608	1.533	0.840	1.055	0.631	1.763
race	0.002	0.157	0.047	0.521	0.074	0.329	0.097	1.114
stage	0.004	1.493	1.140	1.957	0.009	1.466	1.099	1.957
tumor_status	0.000	8.711	3.789	20.027	0.000	7.067	3.039	16.435
Riskscore	0.005	2.619	1.333	5.147	0.005	3.517	1.457	8.492
Test sample(n = 179)								
age	0.002	1.023	1.008	1.038	0.020	1.020	1.003	1.037
gender	0.187	1.367	0.859	2.175	0.570	1.160	0.695	1.935
race	0.474	0.484	0.066	3.531	0.015	0.063	0.007	0.585
stage	0.012	1.471	1.088	1.988	0.000	1.909	1.387	2.626
tumor_status	0.000	9.188	4.233	19.946	0.000	19.431	6.995	53.973
Riskscore	0.018	2.028	1.131	3.637	0.042	1.869	1.022	3.417
GSE54467 sample (n = 79)								
age	0.014	1.023	1.005	1.041	0.003	1.029	1.010	1.049
gender	0.998	1.001	0.568	1.764	0.572	1.181	0.663	2.105
stage	0.675	1.077	0.762	1.522	0.148	1.321	0.906	1.927
Riskscore	0.000	2.683	1.670	4.312	0.000	2.661	1.654	4.281

Figures

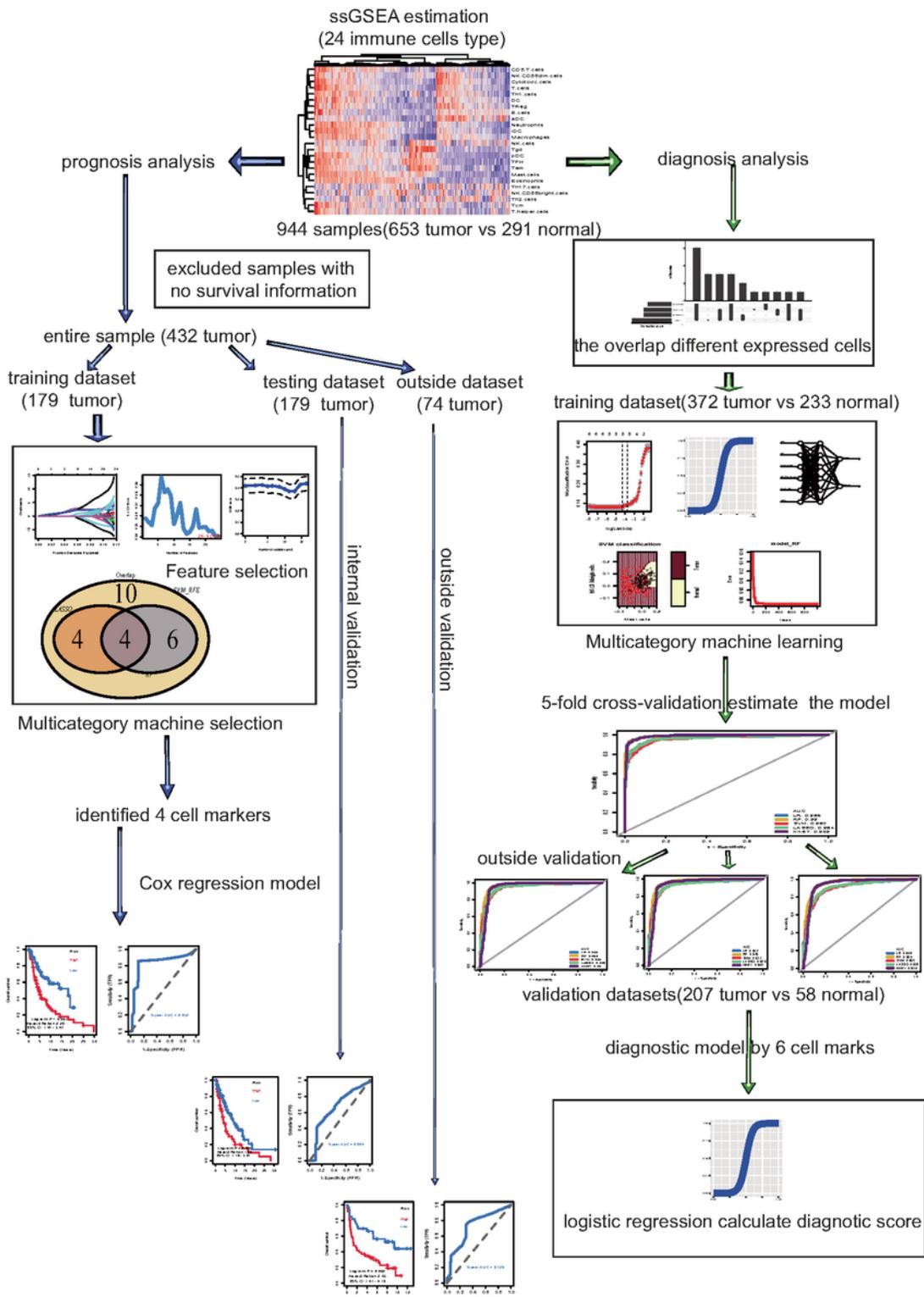


Figure 1

The complete workflow of the analysis in this study.

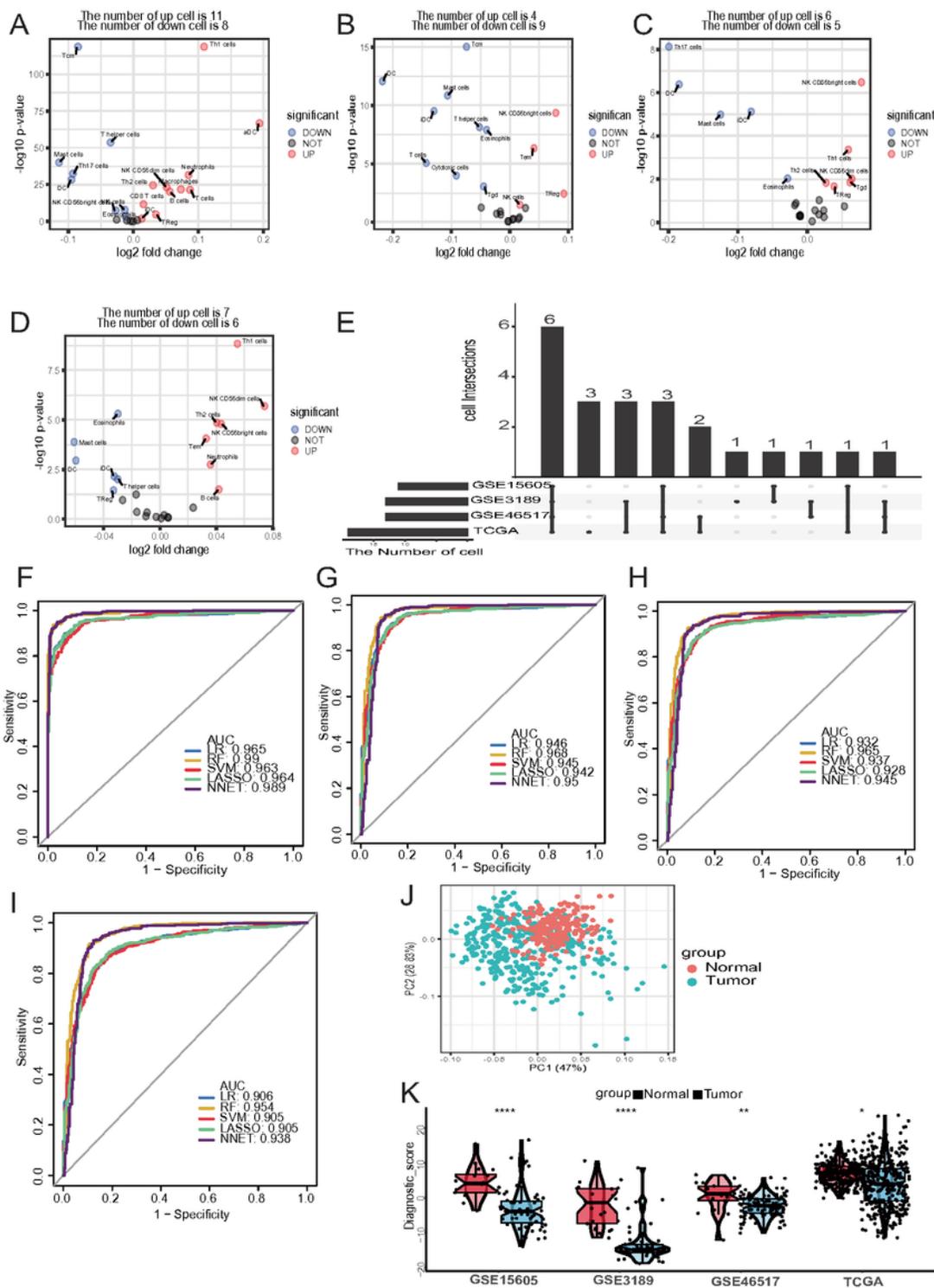


Figure 2

Construction and validation of the immune diagnostic model in melanoma patients. A-D: Volcano plots of TCGA dataset (A), GSE3189 (B), GSE15605 (C), GSE46517 (D), which visualizing the differentially infiltrated immune cells between tumors and normal controls. Red and blue plots represent differentially statistical significance ($p < 0.05$). E: Upset plot of differential immune cells in multiple datasets. The dark Loading [MathJax]/jax/output/CommonHTML/jax.js

drawing represent the intersections of immune cells. F-I: Receiver operating characteristic (ROC) curves on multiple machine learning algorithms for diagnostic model in TCGA dataset (F), GSE3189 (G), GSE15605 (H), GSE46517 (I). AUC, area under ROC curve; LR, Logistic Regression; RF, Random Forests; SVM, Support Vector Machines; LASSO, Least Absolute Shrinkage and Selection Operator; NNET, neural network. J: Principal component analysis of the expression of immune cells in the TCGA dataset. K: Distribution of diagnostic scores in different datasets. The box plots inside the violin indicate the median value and interquartile range of diagnostic scores. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

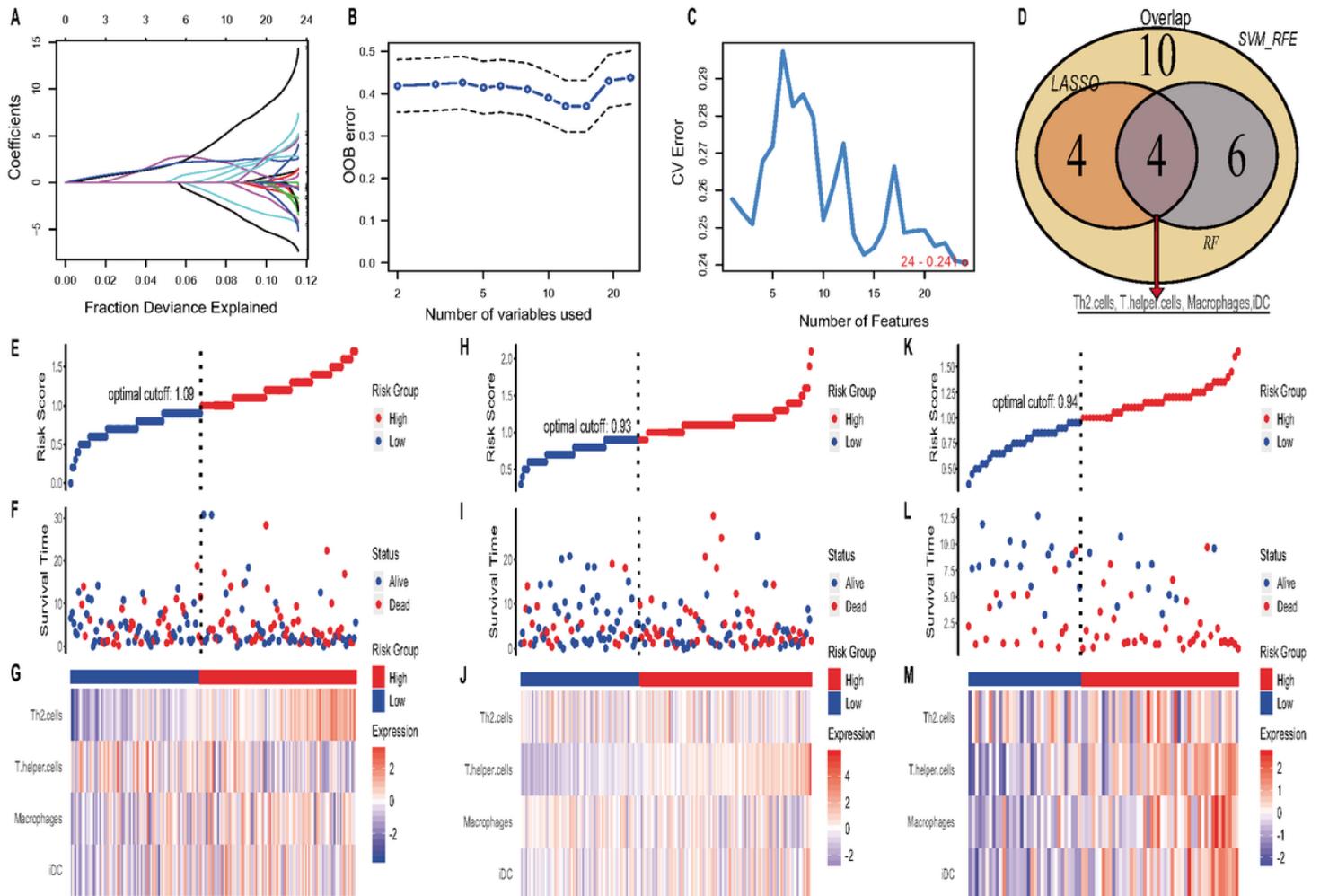


Figure 3

Prognostic feature selection and risk score model. A: The Least Absolute Shrinkage and Selection Operator (LASSO) algorithm coefficient profiles of the 24 immune cells. B: Random forests feature selection (RF-FS) algorithm. The lowest point of the curve indicates the lowest out of bag (OOB) error, and the corresponding immune cells at this point are the best signature selected by RF-FS C: Support Vector Machine-Recursive Feature Elimination (SVM-RFE) algorithms. The point highlighted indicates the lowest error rate, and the corresponding immune cells at this point are the best signature selected by SVM-RFE. D: The Venn plot of overlapping immune cells selected by LASSO, RF-FS and SVM-RFE algorithms. E-G: The risk model in training dataset: the distribution of risk score which are arranged in ascending order

Loading [MathJax]/jax/output/CommonHTML/jax.js

from left to right (E); Overall survival (OS) time and life status (F); The prognostic immune cells expression patterns for melanoma patients (G). H-J: The risk model in testing dataset: the distribution of risk score which are arranged in ascending order from left to right (H); Overall survival (OS) time and life status (I); The prognostic immune cells expression patterns for melanoma patients (J). K-M: The risk model in GSE54467 dataset: the distribution of risk score which are arranged in ascending order from left to right (K); Overall survival (OS) time and life status (L); The prognostic immune cells expression patterns for melanoma patients (M).

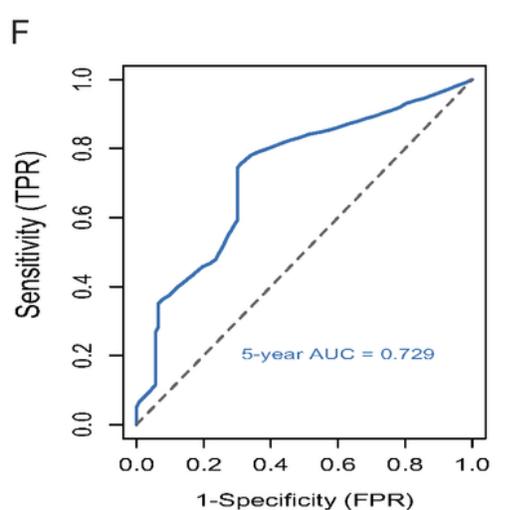
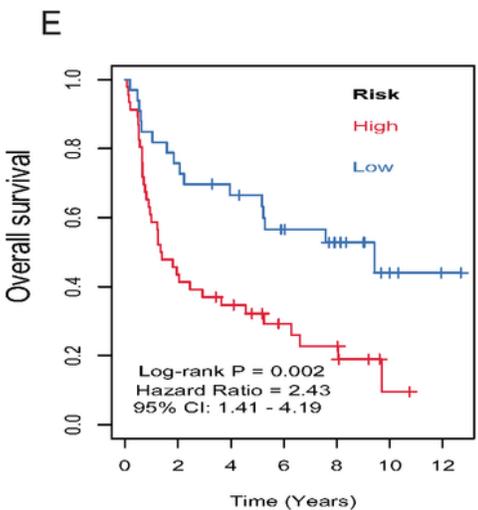
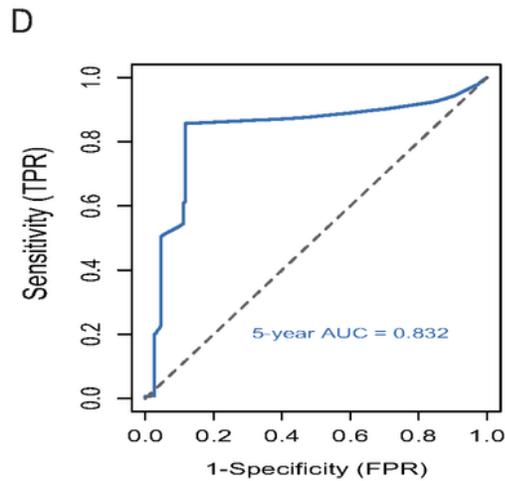
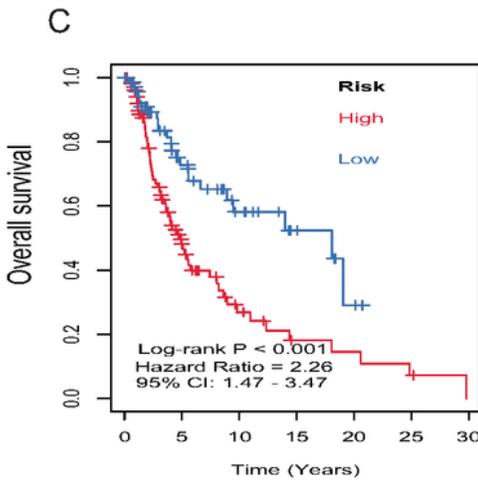
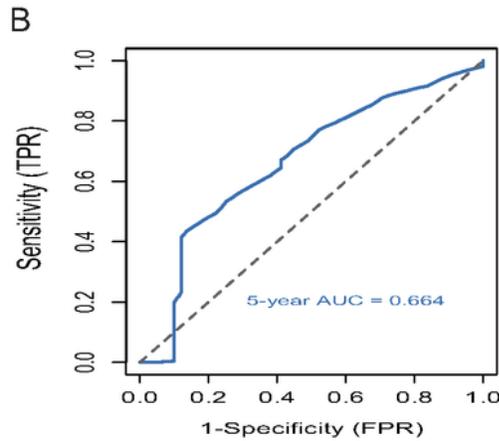
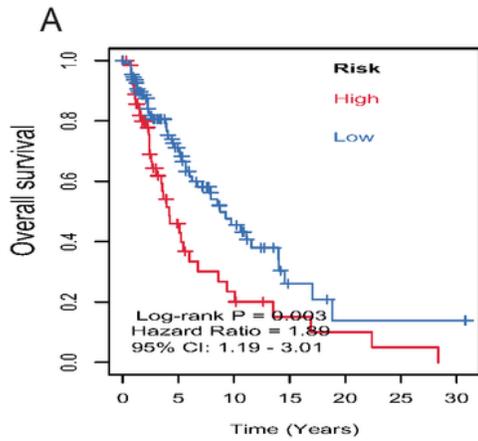


Figure 4

Identification and validation of immune cells biomarker for survival prediction. A: Kaplan–Meier analysis of immune cells biomarker and 5 years of the receiver operating characteristic (ROC) curve in training dataset. E: Kaplan–Meier analysis of immune cells biomarker and 5 years of the receiver operating characteristic (ROC) curve in testing dataset. F: Kaplan–Meier analysis of immune cells biomarker and 5 years of the receiver operating characteristic (ROC) curve in GSE54467 dataset.

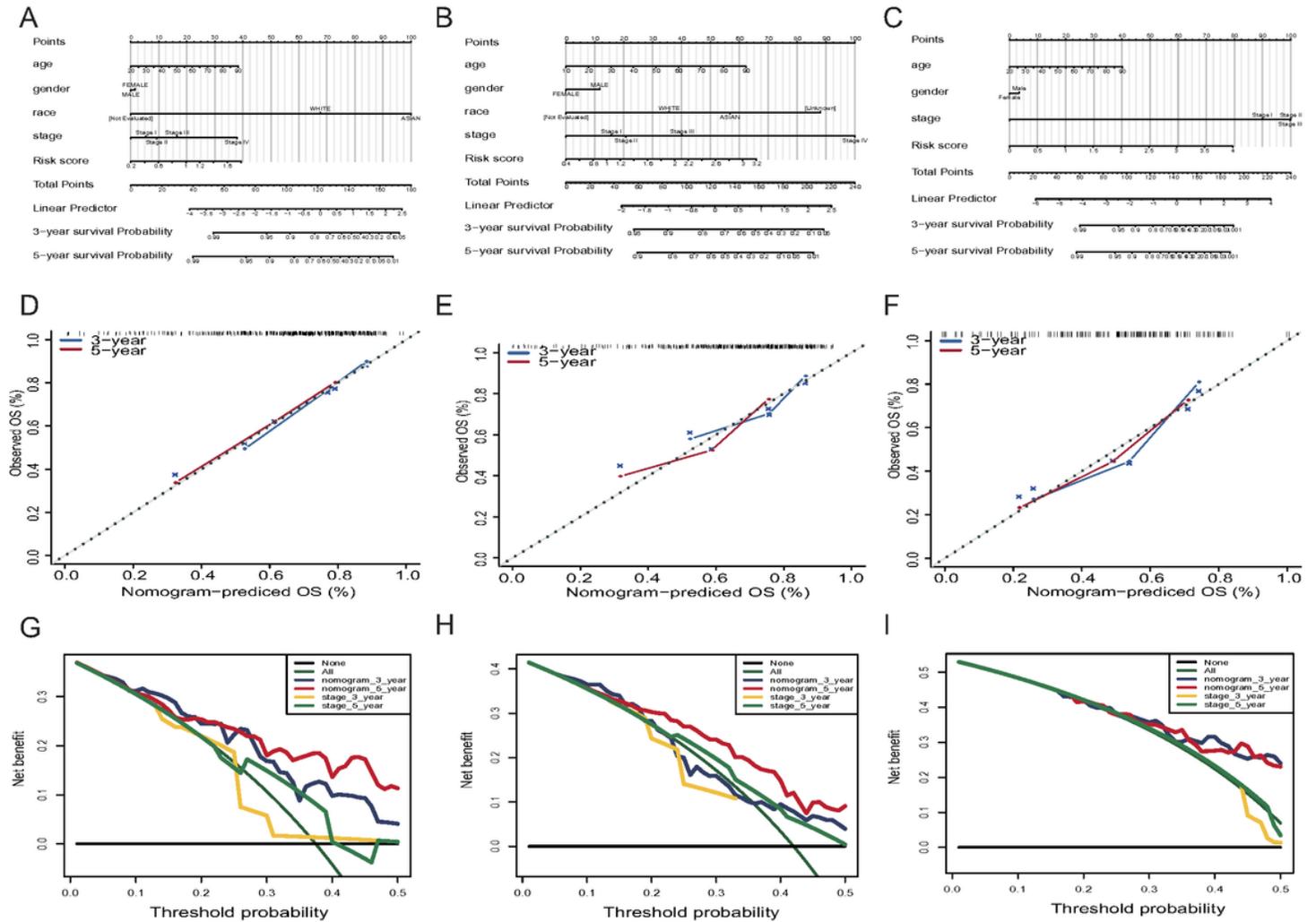


Figure 5

Nomogram construction and validation. A-C: Nomogram for predicting 3-, and 5-year Overall survival (OS) for melanoma patients in training dataset (A), testing dataset (B) and GSE54467 dataset (C) based on immune risk score and clinical characteristics. D-F: Calibration curves of nomograms in terms of agreement between predicted and observed 3-, and 5-year outcomes in training dataset (D), testing dataset (E) and GSE54467 dataset (F). The dashed line of 45° represents perfect prediction, and the actual performances of our nomogram are shown by blue and red lines. G-I: Decision curve analyses of the nomogram and tumor stage for 3-, and 5-year risk in training dataset (G), testing dataset (H) and GSE54467 dataset (I).

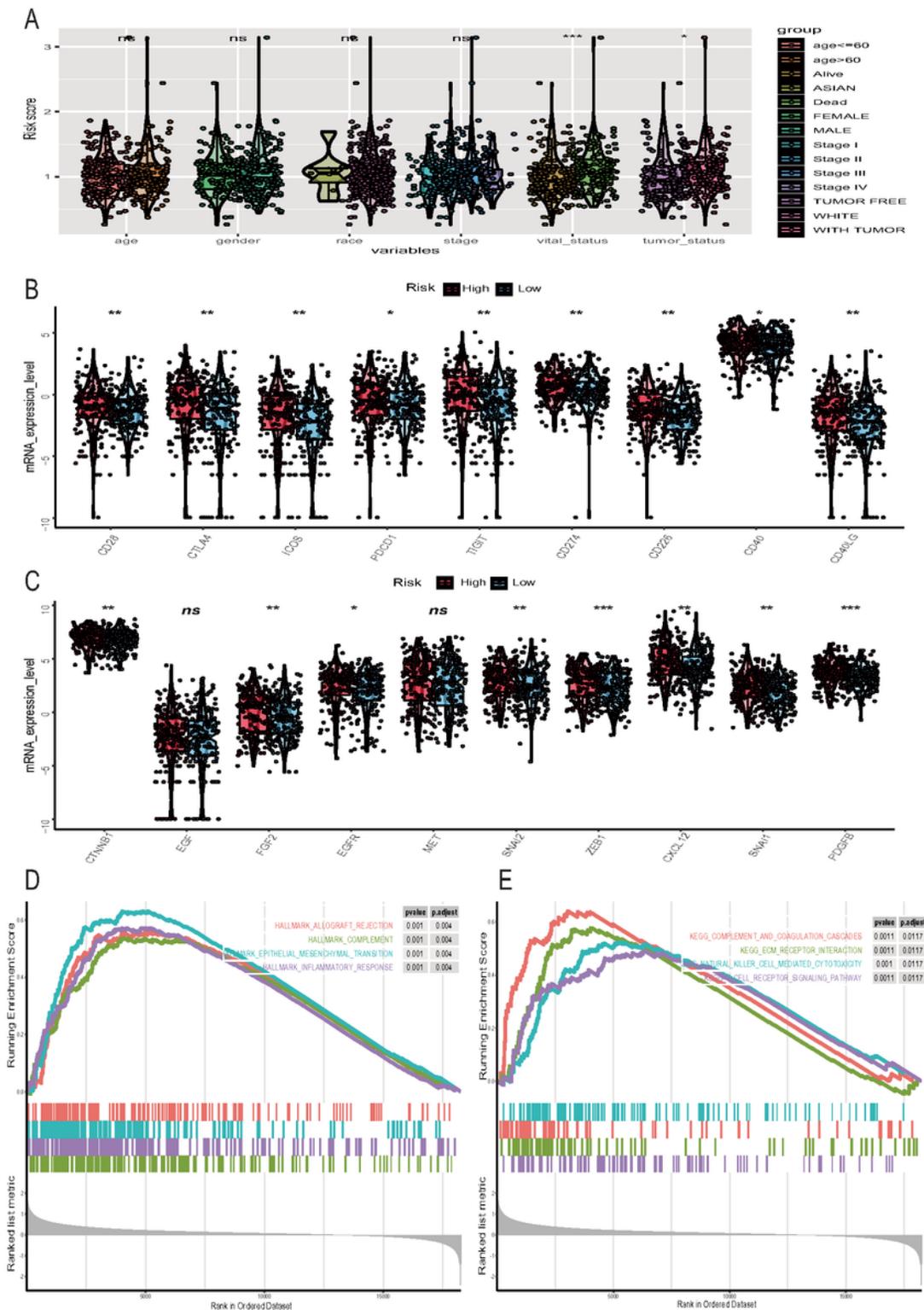


Figure 6

Stratified analysis and GSEA analysis. A: The relationship between risk score distribution and clinical variables which including age, gender, race, stage, vital status and tumor status. B: Expression of immune checkpoint related genes between the high and low risk melanoma patients. C: Expression of epithelial mesenchymal transition (EMT) related genes between the high and low risk melanoma patients. * $p < 0.05$;

Loading [MathJax]/jax/output/CommonHTML/jax.js e set enrichment analysis (GSEA) of high vs. low risk score

groups by using gene sets of the cancer hallmark pathway (h.all.v7.0.symbols) (D) and KEGG pathway (c2.cp.kegg.v7.0.symbols) (E) downloaded from the MSigDB database. Samples were classified into high- and low-risk groups. Each run was performed with 1000 permutations. Enrichment results with significant associated pathways are shown.