

Identification of a Prognostic Signature of Malignant Melanoma Based on Tumor Immune Microenvironment Exploration

lingyan yuan

The First Clinical Medical College of Lanzhou University

Zhitong Bing

Department of Computational Physics

Jianshu Wang

Gansu Provincial Cancer Hospital

Jing Li

Gansu Provincial Cancer Hospital

Xiaodong Jin

Key Laboratory of Heavy Ion Radiation Biology and Medicine of Chinese Academy of Sciences, Lanzhou 730000, China

Rui Li

School of Public Health

Caiyun Wang

The First Clinical Medical College of Lanzhou University

Xueqiang Sun

The First Clinical Medical College of Lanzhou University

Lei Yang

Department of Computational Physics

Kuhu Yang (✉ kehuyangebm2006@126.com)

Lanzhou University

Research

Keywords: Malignant melanoma (MM), observer survival (OBS), immune-checkpoint genes (ICGs), immune-related gene signature (IRGs), tumor-infiltrating lymphocytes (TILs)

Posted Date: October 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-94428/v1>

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

[Read Full License](#)

Abstract

Background:

In contrast to identification of well-defined oncogenic alterations like BRAF mutations for malignant melanoma (MM) patient stratification, effective selection of predictive biomarkers remains a challenge in the era of checkpoint blockade.

Methods:

The differentially expressed genes (DEGs) related to the TME were identified using The Cancer Genome Atlas (TCGA) dataset by Wilcoxon rank sum test. The prognostic effects of immune-related genes (IRGs) were analyzed using univariate Cox regression. Next, the prognostic model was constructed by step multivariate Cox regression and risk score of each sample was calculated. Then, survival and Receiver Operating Characteristic (ROC) analyses were conducted to validate the model using TCGA and the Gene Expression Omnibus (GEO) datasets, respectively. Finally, the overall immune status, tumor purity of high- and low-risk groups was further analyzed to reveal the potential mechanisms of prognostic effects of the model.

Results:

Twenty eight IRGs were identified, the univariate cox analysis indicated the hazard ratio ranged from 0.796 to 2.621 (p -value < 0.05). 6 genes (SLPI, S100A7, LYZ, CCL19, CXCR4 and CD79A) were screened out by step multivariate cox regression and a 6-IRGs, which can be used as an independent prognostic factor, was constructed. The MM patients in both training (TCGA) and testing (GEO) datasets can be well stratified as high-risk and low-risk groups with the 6-IRGs signature, and the 3-year and 5-year area under curve (AUC) of ROC curves of GEO set were 0.681 and 0.678 (GSE19234).

Conclusions:

In sum, we identified and constructed a 6-IRGs, which can be used to predict the prognosis of metastasis in MM patients.

Background

Melanoma is a malignant neoplasm of melanocytes that accounts for the majority of skin cancer deaths despite comprising less than 5% of all cutaneous malignancies. Its incidence has increased faster than that of any other cancer over the past half-century and the annual costs of treatment in the United States alone have risen rapidly. The average 5-year survival rate is approximately 90, 70, 40, and less than 10% for patients with stage I, II, III, and IV malignant melanoma, respectively [1-4]. One of the standard methods of diagnosing and treating melanoma is represented by surgical resection (melanoma is frequently located in the skin, so excision is still widely applied at present for patients with early-stage melanoma due to its facile accessibility).

Although the majority of primary melanomas are cured with local excision, metastatic melanoma historically carries a grim prognosis with a median survival of 9 months and a long-term survival rate of 10% [1, 5]. Except for melanoma in situ, the surgical procedure is generally not enough when it is performed as monotherapy [4], local recurrence, and / or metastases develop in most patients after a variable period of time.

In contrast to identification of well-defined oncogenic alterations like BRAF mutations for MM patient stratification, effective selection of predictive biomarkers remains a challenge in the era of checkpoint blockade. Tumor cellular diversity poses both challenges and opportunities for cancer therapy. This is exemplified by the varied clinical efficacy achieved in malignant melanoma with targeted therapies and immunotherapies. Immune checkpoint inhibitors (ICIs) can produce clinical responses in some patients with metastatic melanomas [6, 7], but the genomic and molecular determinants of response to these agents remain incompletely understood. Although tumor neoantigens and PD-L1 expression have been demonstrated clearly correlated with this response[8, 9], other factors from subsets of malignant cells, the microenvironment, and tumor-infiltrating lymphocytes (TILs) may also play essential roles[10]. Classification of the immune landscape in tumors depends on both PD-L1 expression and the presence of T cell infiltration, referred to as tumor immunity in microenvironment (TIME) classification, which have important implications for the success of anti PD-1/PD-L1 therapy [2]. Therefore, it is critical to define and validate an effective gene signature model for the prognosis of MM. Because of the differences in genetic backgrounds and therapeutic efficacy of immunotherapy, specialized therapeutic strategies for these subtypes of melanoma should be established in the future.

In this study, we selected significant mRNAs for the prognosis of metastasis MM patients via data mining, and the data were validated in independent datasets. This may provide effective candidate genes for the prediction and treatment of metastasis group in MM.

Materials And Methods

Data source and preprocessing

Publicly available dataset of MM cohort, including Level 3 data of gene expression profile and relevant clinical information, was downloaded from TCGA data portal (<https://portal.gdc.cancer.gov/>, accessed DEC 21, 2019). The 448 patients with MM consisted of 352 metastasis samples and 96 primary samples. The clinical information includes age, gender, pathologic stage, Clark level value, pathologic TNM, Breslow depth value, and survival time. Normalized gene expression was measured as fragments per kilobase of transcript per million mapped reads (FPKM) and log₂-based transformation.

To validate the effectiveness of extrapolation of the signatures obtained by the above mentioned approaches. The validation dataset, containing 44 metastatic melanoma samples with RNA sequencing information from three groups were included for validation, including GSE19234[11].

Identification of genes

The gene expression differences between the primary group and metastasis group compared using the “limma” package in the R software 3.3.1[12], and genes with $|\log \text{fold change (FC)}| > 1$ and Benjamini-Hochberg (BH)-adjusted $P < 0.01$ were considered to be significant DEGs.

Acquisition of immune-related genes

A comprehensive list of IRGs was downloaded from the Immunology Database and Analysis Portal (ImmPort) database[13]. The list comprised a total of 2,498 IRGs, covering 17 immune categories.

Construction of MM immune-related prognostic model

Using the “survival” package in R, we employed univariate cox regression on IRGs and Observed Survival, OBS[14] of metastasis MM in the TCGA database to identify survival-associated IRGs. OBS was defined as the time interval from TCGA sampling to patient death or last follow-up. After testing for collinearity, multivariate cox regression analysis was performed to establish the IRG-derived risk signature in metastasis MM. The following formula based on a combination of cox coefficient and gene expression was used to calculate the risk score[15]:

$$\text{Riskscore} = \sum_{i=1}^k \beta_i S_i \quad (1)$$

where k , β_i , S_i represent the number of signature genes, the coefficient index, and the gene expression level, respectively.

To stratify metastasis patients into low- and high-risk groups, the optimum cutoff value for the risk score was determined using the “survminer” package in R. Next, the Kaplan Meier survival curve and log-rank test was performed to evaluate the survival rates between low- and high-risk groups. The area under the receiver operating characteristic (ROC) curve (AUC) was calculated using the “survival ROC” package in R[16], and the AUC was calculated to measure the predictive accuracy of this prognostic signature for time-dependent cancer death. For ROC curve analysis, the AUC was calculated and for Kaplan-Meier (KM) survival analysis, the p-value of a two-side log-rank test was calculated by the “survminer” package[17].

Functional enrichment analysis of DEGs

Gene ontology (GO) analysis was performed via an R clusterProfiler package[18] for screening of differential methylation of genes was conducted to search the vital functions related to the differential gene methylation in MM [19]. The p-value was adjusted into false discovery rate (FDR) by Benjamini and Hochberg study[20]. $FDR < 0.01$ was considered as the cutoff value. $P\text{-value} < 0.05$ was regarded as the cut off threshold value. By this strategy, the screened genes with methylation differences were then assigned to different gene functions including cellular component (CC), molecular function (MF) and biological process (BP).

Infiltrating immune cells of MM analysis in CIBERSORTx database

CIBERSORTx is a suite of machine learning tools for the assessment of cellular abundance and cell type-specific gene expression patterns from bulk tissue transcriptome profiles[21]. The gene expression data was uploaded to CIBERSORTx's online platform (<https://cibersortx.stanford.edu>). A validated leukocyte gene signature matrix (LM22) was used to identify 22 human hematopoietic cell subsets in bulk tissues, including tumors. We utilized Bulk-mode batch correction and 1000 permutations for the analysis in relative mode. Quantile normalization was disabled as the dataset was generated from RNA-seq. The filter criteria of each sample is set as the CIBERSORTx calculation of P-value < 0.05 (statistical significance of the deconvolution result across all cell subsets to ensure adequate goodness-of-fit).

Calculation of Stromal and Immune Scores

ESTIMATE is an algorithm that infers the infiltration situation of immune cells and stromal cells in tumor tissue according to the transcriptome data of TME-related genes which contain a set of immune and stromal signature genes[22]. The expression data of these TME-related genes were extracted from TCGA MM dataset. ESTIMATE outputs stromal and immune scores by performing single-sample gene set-enrichment analysis[23, 24]. For tumor samples of MM cohort, first, gene expression values were rank-normalized and rank ordered. Then, the empirical cumulative distribution functions were calculated for genes in the signature and the remaining genes. Finally, a statistic was calculated by an integration of the difference between the empirical cumulative distribution function.

Validation of the gene signature in the testing dataset

The predictive performance of the IRGs was further validated in GEO dataset. Samples in the testing datasets were divided into high- and low-risk group according to the formula of risk score derived from the training dataset, respectively. KM survival analysis and ROC curve were used to evaluate the predicting power of the gene signature.

Evaluation of immune status between high-risk and low-risk groups stratified by prognostic model

To explore the potential relationship between immune system and IRGs, we analyzed the immune status of the high-risk and low-risk samples. First, using six- IRGs, we quantified the immune activities between high-risk and low-risk samples by the ESTIMATE algorithm was used to calculate corresponding immune scores, stromal scores and tumor purities[22], and the difference of tumor purities between high-risk and low-risk samples was further analyzed.

Results

Design of workflow

For investigating the biomarkers in MM, a hypothesis was proposed, which assumes that differential genes related to immunity may be related to the prognosis of metastatic patients. Thus, the prognosis was explored in MM patients with OBS from TCGA database. The workflow was listed in Figure1

Patient demographics and clinical factors

Our study sample comprised 448 patients. Patient characteristics of the discovery and validation set are shown in Table 1 and Table 2. The TCGA sample from 448 patients as the discovery set (96 patients of primary, 352 patients of metastasis) was analyzed, and the GSE19234 (44 patients with above stage III of MM) samples of MM patients as the validation set (Table 2) (44 patients of metastasis) was analyzed.

The analysis of immune-related genes in primary and metastasis tissue of MM patients

Compared to primary tissues, a total of 172 DEGs consisting of 21 up-regulated and 151 down-regulated genes were identified. The heatmap and volcano plot of the DEGs are shown in Figures. 2 A. To further verify whether these 172 DEGs were related to immune microenvironment reconstruction, the immune-related genes (IRGs) from the Immport database <https://immport.niaid.nih.gov/home> (The Immunology Database and Analysis Portal) were applied to validate. Finally, 28 IRGs were obtained by matched them, the heatmap of 28 differentially expressed IRGs was shown in Figure. 2B (Supplement file. 1).

Establishment of the immune-related genes risk signature

Seven predominant clinical and prognostic factors, including age, gender, race, clark level value, pathologic TNM, stage, Breslow depth value and IRGs were evaluated using univariate and after testing for collinearity, the results showed that 16 IRGs and two clinical factors (Clerk scores and pathologic T) were selected (Table. 3), for univariable Cox regression, the coefficients of clerk scores (HR was 1.261 with 95% CI [1.026, 1.55], and p-value = 0.0275 *) and pathologic T stage (HR was 1.101 with 95% CI [1.013, 1.197], and p-value = 0.024 *) were statistically significant. Then, the stepwise multivariate Cox regression analysis was performed to establish the IRG-derived risk signature in metastasis MM. After stepwise multivariate cox proportional hazard regression, only 6 IRGs for OBS were included in the model (Table 4). The data showed that the pathologic T stage continued to be influencing factors of the prognosis of MM patients in TCGA, with p-values of 0.025. For every MM patient, a PI as a risk score was calculated according to equation (1). The PI was significantly associated with OBS in metastasis group. As, the PI was the most significant with the smallest p-value ($1.921 \times 10^{-13}^{**}$) and the largest HR (1.614 [1.420, 1.833]), which thereby implied that the PI was superior to conventional prognostic markers, including T stage in predicting patient outcome.

The time-dependent ROC curve was created in Figure. 3B. The ability and efficiency of each model to predict metastasis patient's outcome was estimated by calculating the AUC of ROC curve, which was conducted using the survival ROC package in R software. The time-dependent ROC curve was created, the AUC of ROC curve were 0.700 (5 year), 0.692 (3 year) and 0.709 (1 year) larger than 0.50 for OBS indicated that the model performed well. Above results demonstrated that the integrative model could effectively classify patients. Thus, if the PI of a patient was larger than 0.228 (1 year) and 0.239 (3 year) or -0.153(5 year), this patient was assigned to a high-risk group. On the contrary, a patient would be considered low-risk (Figure. 3C). K-M curves (Figure. 3A) show that the p-value of the log-rank test was

2.01×10^{-10} , which implied that the PI could significantly separate patients in a high-risk and low-risk group. Therefore, it can be concluded that the PI calculated via 6 IRGs had a good predictive value of prognosis of MM patients in the TCGA database.

GO functional enrichment analysis

We analyzed the GO enrichment and pathways in which gene-signature involved in. Of these IRGs models, the number of genes in a gene signature model is so small that it is difficult to enrich in GO analysis. ClusterProfiler package was employed to analyze the models. The above package can compare the results of biological process, cellular component and molecular function (Figure. 4). The results of biological process suggested that the 172 differential genes share many processes (Figure. 4A), include epidermis development, skin development, epidermal cell differentiation. These biological processes are related to the occurrence and development of MM.

Correlation of the risk score with tumor-infiltrating immune cells in MM

By applying the CIBERSORTx algorithm to RNA-seq data, the relative proportions of 22 immune cell subsets of MM were acquired. Consecutively, 96 cases of primary MM and 352 cases of metastasis MM in the TCGA dataset were enrolled for further analysis after the filter criteria with p-value < 0.05 via CIBERSORTx algorithms.

The boxplot showed that in TCGA datasets, the immune status of the primary and metastasis samples showed a certain degree of heterogeneity. Overall, we noticed that the majority of infiltrating immune cells within primary MM tissue were comprised of naïve B cells followed by resting CD8 + T cells and plasma cells (Figure. 5A, Supplement file.2).

As shown by bar plot in Figure 5, the abundance of the 22 infiltrative immune cells by using CIBERSORTx were significantly different between primary group (Figure. 5B) and metastasis group (Figure. 5C) in MM cohorts. Among them, the macrophage M2 was the most significant enrichment of immune cells.

Subsequently, as shown in the box plots (Figure. 5D), the infiltration levels of plasma cells, naïve B cells and Tfh cells were significantly differential in primary group and metastasis group in TCGA datasets (p-value < 0.05). The naïve B cells and plasma cells were all significantly higher in metastasis group than that in primary group (p-value < 0.05). And conversely, for Tfh cells, there is a lower infiltration level in metastasis group than primary group.

Correlation of the risk score with tumor-infiltrating immune cells

As we known the immune checkpoint gene expression, which is associated with treatment responses of immune check point inhibitors, was also analyzed. We evaluated nine genes previously reported to be targets of ICIs: PDCD1 (PD-1), CD274 (PD-L1), CD276 (B7-H3), CTLA-4, IDO1, LAG3, VTCN1, HAVCR2 and TNFRSF4 (OX40). The correlation analysis showed that correlation between 22 types of tumor immune cells both and 9 immune-checkpoint genes (ICGs) and 6-IRGs were very different in primary tissues and

metastatic tissues of MM. (Figure. 6A,C and 7A,C, Supplement file. 3,5 and 7,9). Except for CXCR4 and VTCNI, the correlation between 6 genes and 9-ICGs is very similar. In the primary group, CXCR4 and VTCNI showed a negative correlation ($r = -0.18$, $p\text{-value} = 0.01$) in the metastasis group, while there was no correlation in the primary group. The transfer group CD79A ($r = -0.28$, $p\text{-value} \leq 0.001$) and LYZ ($r = -0.189$, $p\text{-value} \leq 0.001$) showed a negative correlation with CD276, but there was no between CD79A, LYZ, and CD276 in the primary group correlation ($p\text{-value} = 1$), (Figure. 6B and 7B, Supplement file.4 and 8). The correlation between 6-IRGs and tumor purity showed that (Figure. 6D and 7D, Supplement file.6 and 10), fourth IRGs were negativities correlated with the tumor purity both in primary group and metastasis group. The correlation analysis results of the 6-genes and the differential tumor infiltrating cells in the primary and metastatic groups showed that (Figure. 6E-G and 7E-G): in the metastatic group, LYZ ($r = 0.309$, $p\text{-value} < 0.0001$) and CXCR4($r = 0.331$, $p\text{-value} < 0.0001$) were highly correlated with plasma cells, and CCL19($r = 0.303$, $p\text{-value} < 0.0001$), CD79A ($r = 0.241$, $p\text{-value} < 0.0001$) and the naïve B cells were correlated. Not only that, CXCR4 is also highly correlated with Tfh ($r = -0.175$, $p\text{-value} < 0.0001$). Perhaps these genes are related to the transfer of primary tumor.

Validation set prognosis analysis for MM in GEO cohort

We screened six IRGs and verified them with the GEO data set again (GSE19234), which can distinguish the prognosis of high-risk and low-risk patients. The validation result is shown in Figure 8. From figure 8, the PI was significantly associated with OS in independent data of MM (HR= 1.457, 95% CI= 0.991-2.142) of GSE19234 dataset. The value of AUC=0.676 (5 year), AUC=0.681 (3 year) and AUC=0.845 (1 year) also indicated that the model performed well.

Validation set immune analysis for MM in GEO cohort

The correlation analysis showed that correlation between 22 types of tumor immune cells and 6 IRGs were very similarity in metastatic tissues of TCGA dataset and tissues of MM in GSE19234(Figure. 9A, supplement file.11). The correlation analysis results of 6 IRGS and 22 immune infiltrating cells showed that: CXCR4 showed a significant positive correlation with plasma cells ($r = 0.524$, $p\text{-value} = 0.03$), and a significant negative correlation with activated NK cells and resting DC cells ($r = -0.534$, $p\text{-value} = 0.02$). CCL19 showed a significant positive correlation with T cells CD4 memory activated ($r = 0.534$, $p\text{-value} = 0.02$), and a significant negative correlation with NK cells activated ($r = -0.683$, $p\text{-value} < 0.001$) and Dendritic cells resting ($r = -0.535$, $p\text{-value} = 0.02$). It is worth noting that the correlation analysis results of 6 IRGS and TME showed that 6 IRGS showed a significant negative correlation with tumor purity ($p\text{-value} < 0.001$), and a significant positive correlation with other immune microenvironment scores ($p\text{-value} < 0.001$) (except SLP1 and S100A7) (Figure 9B, supplement file.12). From Figure 9C (supplement file. 13), we can see the correlation between 6 IRGs and ITLS, 6 IRGs and PDL1 have a strong positive correlation (except SLP1 and S1007A), which also provides a certain amount for future immunotherapy Ideas. From Figure 9D, we can see the correlation between 6-IRGs and three type ITLSs, verifying the results of the TCGA dataset.

Evaluation the immune status between low-risk and high-risk groups

To further explore the relationship between six-gene signature and TME, ESTMATE method was used to assess the overall immune status of high-risk and low-risk groups by analyzing the expression profiles of the 29 immune signature gene sets. The low-risk group in TCGA and GSE19234 (Figure.10) showed more immune activities than that of high-risk group. Tumor purity of low-risk group in TCGA and GSE19234 were significantly higher than that of high-risk group, which suggested more infiltrated immune and stromal cells in the TME of low-risk group.

Discussion

Immune checkpoint blockade represents a major advancement in cancer therapy for advanced melanoma. However, durable clinical responses are seen in only a minority of patients treated with single-agent CTLA-4 (1) or PD-1 blockade [27, 28]. Although higher response rates are achieved when CTLA-4 and PD-1 inhibitors are administered concurrently, this regimen also has greatly increased toxicity [28, 29]. There is a clinical need to predict who will benefit from immunotherapy and to understand mechanisms of therapeutic resistance to improve patient management and outcomes.

The cancer immune microenvironment has been intensively studied in the past few decades, paving the way for the recent clinical application of immunotherapies targeting immune checkpoints such as cytotoxic T-lymphocyte associated protein 4 (CTLA4), programmed cell death 1 (PDCD1/PD1) and programmed cell death 1 ligand 1 (CD274/PDL1) [30, 31]. Various immunogenomic approaches have been applied to analyze tumor-immune cell interactions [32, 33] and accumulating evidence supports the impact of host immunity on cancer progression and response to immunotherapy [34–37].

Immunotherapy for the treatment of advanced melanoma and other cancer types have become a primary treatment in the clinic. Checkpoint inhibitors block natural pathways that dampen or inhibit an immune response to stimulus.

These pathways include programmed cell death 1 receptor/programmed death-ligand 1 and cytotoxic T lymphocyte antigen-4 [38]. Although the genetic and epigenetic changes in tumor cells are crucial to the oncogenesis and progress of tumors, accumulating evidence shows that the interaction between the tumor cells and its surrounding normal cells also plays an indispensable role [39].

Based on the important role of TME in the development of tumors, we identified the immune-related DEGs, constructed a six-IRGs and validated its association with OS in a GEO test dataset to identify prognostic biomarkers associated with TME in MM's samples. The results indicated that the six-IRGs can well classified the MM patients of training and testing datasets into high- and low-risk groups, and the high-risk groups were associated with poorer prognosis. The univariate and multivariate cox analysis confirmed that the six-IRGs can be used as independent prognostic factor for predicting the patients' outcome. The ROC analyses using training, testing datasets also proved the robusticity of the prognostic signature.

For there is now abundant evidence that both tumor [40–44] and host-related factors [45–48] can influence heterogeneous response and resistance to immune checkpoint blockade. GO analysis was

conducted to study the potential molecular mechanism of prognostic effects of gene signature and the results showed that the expression changes of genes in the prognostic model mainly affected the cell-cell junction and the pathways related to IL-17 signaling pathway. To explore the state of immunity in MM, ESTIMATE method was used to evaluate the overall immune status and tumor purity in metastasis TME in MM. Consistent with the KEGG pathway analysis, the overall immune activity of most low-risk groups was higher than that of the high-risk group. Correspondingly, the tumor purity was lower than that of the high-risk group, suggesting that more stromal cells and immune cells were infiltrated in the TME [38].

Immune and stromal cells infiltrated in TME are composed of many different types of cells. In this study, using transcriptome profiling data, we estimated immune cell composition in clinically tumor-primary tissue from MM patients and metastasis tissue from MM individuals. Correlations between immune infiltration, immune activity and patient survival were also investigated. As expected, significantly altered immune cell composition was seen in metastasis tumors compared to primary samples. However, we also found elevated infiltration of CD8⁺ T cells, naive B cells and CD4⁺ memory T cells in metastasis tumor compared to primary samples from MM individuals, whereas infiltration of master cells resting and activated NK cells showed a decrease.

Immune scores and stromal scores calculated based on the ESTIMATE algorithm could facilitate the quantification of the immune and stromal components in a tumor. To better understand the effects of genes involved in immune and stromal cells on prognosis, we identified the immune-related differential expression genes of MM primary samples and metastatic samples in the TCGA database, and the expression of the genes was significantly related to the prognosis of MM patients. The immune/matrix scores of patients with different prognostic risk groups were also identified. The overall immune score was also higher in primary samples compared with metastasis samples. This result has also been confirmed in other tumor studies[49].

We suggested that the level of LYZ, CCL19, CXCR4 and CD79A expression was negatively correlated to the tumor purity in metastatic MM. The expression of immune check point, a potential therapeutic target of MM, was also analyzed in this study. The immune check point signaling pathway related PD-L1, CTLA-4 and IDO1 were consistently upregulated in highly immune-activated MM. These findings indicated that expression of new immunotherapy targets was correlated to expression of PD-L1 and CTLA-4; therefore, it is suggested that the MM showed clinical response to anti-PD-1 or PD-L1 antibody might also respond well to new immune check point inhibitors, such as IDO1 inhibitor. Moreover, these results suggested that a combination of immune check point inhibitors can be an effective therapeutic strategy in metastatic MM.

Of the 6 genes, five genes(LYZ,S100A7, CCL19, CXCR4 and CD79A) have been reported to be associated with MM's progression or significant in MM survival prediction[50–54], indicating that our big data analysis based on ESTIMATE algorithm has prognostic values in the MM cohort of TCGA database. The other two genes have never been reported to correlate with MM development and prognosis before and can be perceived as novel potential biomarkers for MM.

Secretory leukocyte protease inhibitor (SLPI) is a member of whey acidic protein four-disulfide core family and plays important role in tumorigenesis and cancer progression and metastasis. SLPI expression is highly upregulated in lung cancer, ovarian cancer, pancreatic carcinoma, papillary thyroid cancer and uterine cervix cancer[55, 56], while downregulated in bladder tumors, prostatic neoplasms and some breast carcinoma [57]. As for the function of SLPI in cancers, Wagenblast et al. reported that SLPI not only drives the formation of extra-vascular networks but also acting as anticoagulants to ensure tumor perfusion[58]. In breast cancer (BC) SLPI may inhibit cell proliferation and promote cancer metastasis[59]. However, the underlying mechanisms have not been fully elucidated in melanoma.

Previously study found the expression of these S100 family genes to be significantly correlated with both lymphatic and distant melanoma metastasis, as well as with American Joint Committee on Cancer grade but not with Clark's grade, age, or sex. This suggests that expression of these genes may be related to the degree of tumor invasion. Although further validation through basic and clinical trials is needed, our results suggest that the S100 family genes have the potential to play an important role in the diagnosis of melanoma. S100 expression may be related to tumor invasion and facilitate the early diagnosis of melanoma, allowing for a more accurate prognosis. S100-targeted therapies are also potentially viable strategies in the context of melanoma[60].

The study showed that stage IV melanoma was characterized by an increased frequency of CCR7⁺CD56 bright NK cells as well as high serum concentrations of the CCR7 ligand CCL19. CCR7 expression and CCL19 secretion were also observed in melanoma cell lines[61].

Melanoma is at the forefront of development of systemic therapeutics with both molecular targeted therapies and immune checkpoint inhibitors as cornerstones of treatment. Additionally, programmed-death-1 (PD-1) inhibitors have revolutionized the treatment of melanoma and are set to pave future improvements in other solid tumors. Combinations of PD-1 inhibitors with novel immune checkpoints or with molecularly targeted therapies are under investigation and may make considerable progress. This study has identified an IRGs that may facilitate the development of therapeutic strategies for metastatic MM.

Conclusions

Our analysis has uncovered a 6-gene signature with significant difference between primary and metastasis samples in melanoma and thus can be used to predict the prognosis for patients with malignant melanoma.

Furthermore, we identified the 6-gene signature linked to IRGs to targeted therapies in metastatic melanomas and validated the outcomes in a number of melanoma samples using different datasets. Our data suggested that the 6-gene signature can be developed as potential biomarker for distinguishing prognosis and selecting patients for immune checkpoint blockade in patients with malignant melanoma.

Abbreviations

MM	Malignant melanoma
OBS	observer survival
ICGs	immune-checkpoint genes
IRGs	immune-related gene signature
TILs	tumor-infiltrating lymphocytes
PD-1	programmed-death-1
SPLI	leukocyte protease inhibitor
CTLA4	cytotoxic T-lymphocyte associated protein 4
TIME	tumor immunity in microenvironment
ROC	receiver operating characteristic
DEGs	differentially expressed genes
FPKM	per kilobase of transcript per million mapped reads
ImmPort	Immunology Database and Analysis Portal database
TCGA	The Cancer Genome Atlas
GEO	the Gene Expression Omnibus
CC	cellular component
MF	molecular function
BP	biological process
GO	Gene ontology

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have read and approved the content and agree to submit for consideration for publication in the journal.

Availability of data and materials

All data was obtained from The Cancer Genome Atlas (TCGA, <https://tcga-data.nci.nih.gov/tcga/>) and Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>).

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Author Contributions

Conceptualization, LYY and ZTB; Formal analysis, JSW, JL; Investigation, RL, XDJ and XQS; Methodology, ZTB and LYY; Project administration, ZTB; Software, LYY, LJ; Supervision, LY, KHY; Validation, RL, CYW; Visualization, LYY and ZTB; Writing – original draft, LYY and ZTB; Writing – review & editing, LY and KHY. All authors read and approved the final version of the manuscript.

Acknowledgements

This is a text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

Contributor Information

Lingyan Yuan, Email: yuanly18@lzu.edu.cn

Zhitong Bing, Email: bingzt@impcas.ac.cn

Jianshu Wang, Email: wjshu6@hotmail.com

Jing Li, Email: 1368406388@qq.com

Kehu Yang, Email: kehuyangebm2006@126.com

Lei Yang, Email: lyang@impcas.ac.cn

References

1. Ernst DS, Petrella T, Joshua AM, Hamou A, Thabane M, Vantuyghem S, Gwady-Sridhar F: **Burden of illness for metastatic melanoma in Canada, 2011-2013.** (1198-0052 (Print)).
2. Gwady-Sridhar F, Nikan S, Hamou A, Seung SJ, Petrella T, Joshua AM, Ernst S, Mittmann N: **Resource utilization and costs of managing patients with advanced melanoma: a Canadian population-based study.** (1198-0052 (Print)).

3. Szczepaniak Sloane RA-O, Gopalakrishnan V, Reddy SM, Zhang X, Reuben A, Wargo JA: **Interaction of molecular alterations with immune response in melanoma.** (1097-0142 (Electronic)).
4. Garbe C, Eigentler Tk Fau - Keilholz U, Keilholz U Fau - Hauschild A, Hauschild A Fau - Kirkwood JM, Kirkwood JM: **Systematic review of medical treatment in melanoma: current status and future prospects.** (1549-490X (Electronic)).
5. Turner NA-O, Ware O, Bosenberg MA-O: **Genetics of metastasis: melanoma and other cancers.** (1573-7276 (Electronic)).
6. Spain L, Larkin J: **Combination immune checkpoint blockade with ipilimumab and nivolumab in the management of advanced melanoma.** (1744-7682 (Electronic)).
7. Postow MA, Chesney J Fau - Pavlick AC, Pavlick Ac Fau - Robert C, Robert C Fau - Grossmann K, Grossmann K Fau - McDermott D, McDermott D Fau - Linette GP, Linette Gp Fau - Meyer N, Meyer N Fau - Giguere JK, Giguere Jk Fau - Agarwala SS, Agarwala Ss Fau - Shaheen M *et al.* **Nivolumab and ipilimumab versus ipilimumab in untreated melanoma.** (1533-4406 (Electronic)).
8. Giuffrida P, Arpa G, Grillo F, Klersy C, Pathology ADSJM: **Correction: PD-L1 in small bowel adenocarcinoma is associated with etiology and tumor-infiltrating lymphocytes, in addition to microsatellite instability.** 2020:1-1.
9. Boudewijns S, Bloemendal M, Haas Nd, Westdorp H, Immunology IJMdVJC, Immunotherapy: **Autologous monocyte-derived DC vaccination combined with cisplatin in stage III and IV melanoma patients: a prospective, randomized phase 2 trial.** 2020(1).
10. Hojabr, Kakavand, James, Wilmott, Alexander, Menzies, Ricardo, Research VJCcraojotAAfC: **PD-L1 Expression and Tumor-Infiltrating Lymphocytes Define Different Subsets of MAPK Inhibitor-Treated Melanoma Patients.** 2015.
11. Bogunovic D, O'Neill DW, Belitskaya-Levy I, Vacic V, Yu YL, Adams S, Darvishian F, Berman R, Shapiro R, Pavlick AC *et al.* **Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(48):20429-20434.
12. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** (1362-4962 (Electronic)).
13. Bhattacharya S, Andorf S Fau - Gomes L, Gomes L Fau - Dunn P, Dunn P Fau - Schaefer H, Schaefer H Fau - Pontius J, Pontius J Fau - Berger P, Berger P Fau - Desborough V, Desborough V Fau - Smith T, Smith T Fau - Campbell J, Campbell J Fau - Thomson E *et al.* **ImmPort: disseminating data to the public for the future of immunology.** (1559-0755 (Electronic)).
14. Xiong J, Bing Z, Guo S: **Observed Survival Interval: A Supplement to TCGA Pan-Cancer Clinical Data Resource.** *Cancers* 2019, **11**(3).
15. Hu F, Zeng W, Liu X: **A Gene Signature of Survival Prediction for Kidney Renal Cell Carcinoma by Multi-Omic Data Analysis.** LID - 10.3390/ijms20225720 [doi] LID - 5720. (1422-0067 (Electronic)).
16. Wu H, Zhang JJGO: **Decreased expression of TFAP2B in endometrial cancer predicts poor prognosis: A study based on TCGA data.** 2018, **149**(3):S0090825818302373.

17. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara SI, Watanabe H, Kurashina K, Hatanaka HJN: **Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer.**
18. Pathan M, Keerthikumar S, Chisanga D, Alessandro R, Mathivanan S: **A novel community driven software for functional enrichment analysis of extracellular vesicles data.**
19. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 2000, **25**(1):25-29.**
20. Benjamini Y, Hochberg YJ: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *57*(1):289-300.**
21. Newman AA-O, Steen CB, Liu CL, Gentles AA-O, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D *et al.* **Determining cell type abundance and expression from bulk tissues with digital cytometry. (1546-1696 (Electronic)).**
22. Yoshihara K, Shahmoradgoli M Fau - Martínez E, Martínez E Fau - Vegesna R, Vegesna R Fau - Kim H, Kim H Fau - Torres-Garcia W, Torres-Garcia W Fau - Treviño V, Treviño V Fau - Shen H, Shen H Fau - Laird PW, Laird Pw Fau - Levine DA, Levine Da Fau - Carter SL *et al.* **Inferring tumour purity and stromal and immune cell admixture from expression data. (2041-1723 (Electronic)).**
23. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA *et al.* **Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications* 2013, **4**:2612.**
24. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP *et al.* **Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* 2010, **17**(1):98-110.**
25. Subramanian A, Tamayo P Fau - Mootha VK, Mootha Vk Fau - Mukherjee S, Mukherjee S Fau - Ebert BL, Ebert Bl Fau - Gillette MA, Gillette Ma Fau - Paulovich A, Paulovich A Fau - Pomeroy SL, Pomeroy Sl Fau - Golub TR, Golub Tr Fau - Lander ES, Lander Es Fau - Mesirov JP *et al.* **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. (0027-8424 (Print)).**
26. Hänzelmann S, Castelo R Fau - Guinney J, Guinney J: **GSVA: gene set variation analysis for microarray and RNA-seq data. (1471-2105 (Electronic)).**
27. Topalian SL, Hodi Fs Fau - Brahmer JR, Brahmer Jr Fau - Gettinger SN, Gettinger Sn Fau - Smith DC, Smith Dc Fau - McDermott DF, McDermott Df Fau - Powderly JD, Powderly Jd Fau - Carvajal RD, Carvajal Rd Fau - Sosman JA, Sosman Ja Fau - Atkins MB, Atkins Mb Fau - Leming PD *et al.* **Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. (1533-4406 (Electronic)).**
28. Larkin J, Chiarion-Sileni V Fau - Gonzalez R, Gonzalez R Fau - Grob JJ, Grob Jj Fau - Cowey CL, Cowey Cl Fau - Lao CD, Lao Cd Fau - Schadendorf D, Schadendorf D Fau - Dummer R, Dummer R Fau - Smylie M, Smylie M Fau - Rutkowski P, Rutkowski P Fau - Ferrucci PF *et al.* **Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. (1533-4406 (Electronic)).**

29. Pagès C, Baroudjian B, Lebbé C: **[Not Available]**. (1769-6917 (Electronic)).
30. Binnewies M, Roberts EW, Kersten K, Chan VA-O, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg SA-O, Hedrick CC *et al*: **Understanding the tumor immune microenvironment (TIME) for effective therapy**. (1546-170X (Electronic)).
31. Baumeister SH, Freeman GJ, Dranoff G, Sharpe AH: **Coinhibitory Pathways in Immunotherapy for Cancer**. (1545-3278 (Electronic)).
32. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z: **Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade**. (2211-1247 (Electronic)).
33. Hackl H, Charoentong P, Finotello F, Trajanoski Z: **Computational genomics tools for dissecting tumour-immune cell interactions**. (1471-0064 (Electronic)).
34. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA *et al*: **The Immune Landscape of Cancer**. (1097-4180 (Electronic)).
35. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD *et al*: **The prognostic landscape of genes and infiltrating immune cells across human cancers**. (1546-170X (Electronic)).
36. Galon J, Mlecnik B Fau - Bindea G, Bindea G Fau - Angell HK, Angell Hk Fau - Berger A, Berger A Fau - Lagorce C, Lagorce C Fau - Lugli A, Lugli A Fau - Zlobec I, Zlobec I Fau - Hartmann A, Hartmann A Fau - Bifulco C, Bifulco C Fau - Nagtegaal ID *et al*: **Towards the introduction of the 'Immunoscore' in the classification of malignant tumours**. (1096-9896 (Electronic)).
37. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, Lugli A, Zlobec I, Rau TT, Berger MD *et al*: **International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study**. (1474-547X (Electronic)).
38. Cuevas LM, Daud AI: **Immunotherapy for melanoma**. (1085-5629 (Print)).
39. Quail DF, Joyce JA: **Microenvironmental regulation of tumor progression and metastasis**. (1546-170X (Electronic)).
40. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS *et al*: **Genetic basis for clinical response to CTLA-4 blockade in melanoma**. (1533-4406 (Electronic)).
41. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS *et al*: **Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer**. (1095-9203 (Electronic)).
42. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM *et al*: **Genomic correlates of response to CTLA-4 blockade in metastatic melanoma**. (1095-9203 (Electronic)).
43. Diaz La Jr Fau - Le DT, Le DT: **PD-1 Blockade in Tumors with Mismatch-Repair Deficiency**. (1533-4406 (Electronic)).

44. Spranger S, Bao R, Gajewski TF: **Melanoma-intrinsic β -catenin signalling prevents anti-tumour immunity.** (1476-4687 (Electronic)).
45. Hamid O, Robert C Fau - Daud A, Daud A Fau - Hodi FS, Hodi Fs Fau - Hwu W-J, Hwu Wj Fau - Kefford R, Kefford R Fau - Wolchok JD, Wolchok Jd Fau - Hersey P, Hersey P Fau - Joseph RW, Joseph Rw Fau - Weber JS, Weber Js Fau - Dronca R *et al.*: **Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma.** (1533-4406 (Electronic)).
46. Tumei PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V *et al.*: **PD-1 blockade induces responses by inhibiting adaptive immune resistance.** (1476-4687 (Electronic)).
47. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N: **Molecular and genetic properties of tumors associated with local immune cytolytic activity.** (1097-4172 (Electronic)).
48. Angelova M Fau - Charoentong P, Charoentong P Fau - Hackl H, Hackl H Fau - Fischer ML, Fischer Ml Fau - Snajder R, Snajder R Fau - Krogsdam AM, Krogsdam Am Fau - Waldner MJ, Waldner Mj Fau - Bindea G, Bindea G Fau - Mlecnik B, Mlecnik B Fau - Galon J, Galon J Fau - Trajanoski Z *et al.*: **Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy.** (1474-760X (Electronic)).
49. Jia D, Li S, Li D, Xue H, Yang D, Liu Y: **Mining TCGA database for genes of prognostic value in glioblastoma microenvironment.** *Aging* 2018, **10**(4):592-605.
50. Xiong TF, Pan FQ, Li D: **Expression and clinical significance of S100 family genes in patients with melanoma.** *Melanoma research* 2019, **29**(1):23-29.
51. Fujimura T, Tanita K, Sato Y, Lyu C, Kambayashi Y, Fujisawa Y, Uchi H, Yamamoto Y, Otsuka A, Yoshino K *et al.*: **Immune checkpoint inhibitor-induced vitiligo in advanced melanoma could be related to increased levels of CCL19.** *The British journal of dermatology* 2020, **182**(5):1297-1300.
52. Yang S, Wei J, Cui YH, Park G, Shah P, Deng Y, Aplin AE, Lu Z, Hwang S, He C *et al.*: **m(6)A mRNA demethylase FTO regulates melanoma tumorigenicity and response to anti-PD-1 blockade.** *Nature communications* 2019, **10**(1):2782.
53. Cantor AS, Moschos S, Jukic DM: **A principal case of multiple lymphoid collision tumors involving both B-cell chronic lymphocytic leukemia and metastatic malignant melanoma.** *Dermatology online journal* 2010, **16**(7):6.
54. Taube JM, Young GD, McMiller TL, Chen S, Salas JT, Pritchard TS, Xu H, Meeker AK, Fan J, Cheadle C *et al.*: **Differential Expression of Immune-Regulatory Genes Associated with PD-L1 Display in Melanoma: Implications for PD-1 Pathway Blockade.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2015, **21**(17):3969-3976.
55. Moreau T, Baranger K, Dadé S, Dallet-Choisy S, Guyot N, Zani ML: **Multifaceted roles of human elafin and secretory leukocyte proteinase inhibitor (SLPI), two serine protease inhibitors of the chelonianin family.** *Biochimie* 2008, **90**(2):284-295.
56. Stępień T, Brożyna M, Kuzdak K, Motylewska EA-O, Komorowski J, Stępień HA-O, Ławnicka HA-O: **Elevated Concentrations of SERPINE2/Protease Nexin-1 and Secretory Leukocyte Protease Inhibitor**

- in the Serum of Patients with Papillary Thyroid Cancer.** (1875-8630 (Electronic)).
57. Xuan Q, Yang X, Mo L, Huang F, Pang Y, Qin M, Chen Z, He M, Wang Q, Mo ZN: **Expression of the serine protease kallikrein 7 and its inhibitor antileukoprotease is decreased in prostate cancer.** *Archives of Pathology and Laboratory Medicine* 2008, **132**(11):1796-1801.
58. Wagenblast E, Soto M, Gutiérrez-Ángel S, Hartl CA, Gable AL, Maceli AR, Erard N, Williams AM, Kim SY, Dickopf S *et al*: **A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis.** *Nature* 2015, **520**(7547):358-362.
59. Xie W, Zhang H, Qin S, Zhang J, Fan X, Yin Y, Liang R, Long H, Yi W, Fu D *et al*: **The expression and clinical significance of secretory leukocyte proteinase inhibitor (SLPI) in mammary carcinoma using bioinformatics analysis.** (1879-0038 (Electronic)).
60. Xiong TF, Pan Fq Fau - Li D, Li D: **Expression and clinical significance of S100 family genes in patients with melanoma.** (1473-5636 (Electronic)).
61. Cristiani CM, Turdo A, Ventura V, Apuzzo T, Capone M, Madonna G, Mallardo D, Garofalo C, Giovannone ED, Grimaldi AM *et al*: **Accumulation of Circulating CCR7(+) Natural Killer Cells Marks Melanoma Evolution and Reveals a CCL19-Dependent Metastatic Pathway.** (2326-6074 (Electronic)).

Supplementary Files

Supplementary file 1: 28 IRGs for primary group and metastasis group.

Supplementary file 2: CIBERSORTx of SKCM in TCGA

Supplementary file 3: Cor of 9-ICGs and TILs 22 of primary group in primary in TCGA

Supplementary file 4: Cor of 9-ICGs and 6-IRGs of primary group in TCGA

Supplementary file 5: Cor of 6-IRGs and TILs 22 of primary group in primary in TCGA

Supplementary file 6: Tumor purity of primary group in TCGA by ESTIMATE

Supplementary file 7: Cor between 9-ICGs and TILs 22 in metastasis group of TCGA

Supplementary file 8: Cor between 9-ICGs and 6-IRGs in metastasis group of TCGA

Supplementary file 9: Cor between 6-IRGs and TILs 22 in metastasis group of TCGA

Supplementary file 10: Tumor purity of metastasis group in TCGA by ESTIMATE

Supplementary file 11: Cor between 6-IRGs and TILs 22 in metastasis group of GEO

Supplementary file 12: Cor between 6-IRGs and tumor purity in metastasis group of GEO

Supplementary file 13: Cor between 6-IRGs and 9-ICGs in metastasis group of GEO

Tables

Table 1 Patient demographics and clinical factors of TCGA (n=448)

TCGA	Primary	Metastasis	Total
	(N=96)	(N=352)	(N=448)
Gender			
Male	57 (59.4%)	224 (63.6%)	281 (62.7%)
Female	39 (40.6%)	128 (36.4%)	167 (37.3%)
Age (years)			
Mean (SD)	63.2 (13.3)	56.1 (15.7)	57.6 (15.5)
Median [Min, Max]	63.0 [24.0, 86.0]	56.0 [15.0, 87.0]	58.0 [15.0, 87.0]
Race			
Asian	6 (6.2%)	5 (1.4%)	11 (2.5%)
Black/African american	0 (0%)	1 (0.3%)	1 (0.2%)
Not reported	1 (1.0%)	7 (2.0%)	8 (1.8%)
white	89 (92.7%)	339 (96.3%)	428 (95.5%)
Status			
Alive	66 (68.8%)	161 (45.7%)	227 (50.7%)
Death	30 (31.2%)	191 (54.3%)	221 (49.3%)
Stage			
Mean (SD)	4.23 (0.946)	3.88 (1.44)	3.95 (1.35)
Median [Min, Max]	4.00 [0, 6.00]	4.00 [0, 6.00]	4.00 [0, 6.00]
observer survival (OBS)			
Mean (SD)	528 (279)	941 (965)	853 (881)
Median [Min, Max]	444 [39.0, 1790]	602 [8.00, 5630]	539 [8.00, 5630]
Survival time			
Mean (SD)	539 (283)	2260 (2030)	1890 (1940)
Median [Min, Max]	459 [39.0, 1790]	1620 [62.0, 11300]	1180 [39.0, 11300]
Clark			
clark level I	0 (0%)	5 (1.4%)	5 (1.1%)
clark level II	1 (1.0%)	17 (4.8%)	18 (4.0%)
clark level III	13 (13.5%)	62 (17.6%)	75 (16.7%)

clark level IV	34 (35.4%)	127 (36.1%)	161 (35.9%)
clark level V	21 (21.9%)	30 (8.5%)	51 (11.4%)
Not reported	27 (28.1%)	111 (31.5%)	138 (30.8%)
pathologic T			
pathologic T0	0 (0%)	23 (6.5%)	23 (5.1%)
pathologic T1	1 (1.0%)	40 (11.4%)	41 (9.2%)
pathologic T2	5 (5.2%)	72 (20.5%)	77 (17.2%)
pathologic T3	11 (11.5%)	79 (22.4%)	90 (20.1%)
pathologic T4	77 (80.2%)	65 (18.5%)	142 (31.7%)
pathologic Tis	0 (0%)	7 (2.0%)	7 (1.6%)
pathologic TX	0 (0%)	42 (11.9%)	42 (9.4%)
Not reported	2 (2.1%)	24 (6.8%)	26 (5.8%)
pathologic N			
pathologic N0	49 (51.0%)	173 (49.1%)	222 (49.6%)
pathologic N1	9 (9.4%)	63 (17.9%)	72 (16.1%)
pathologic N2	10 (10.4%)	39 (11.1%)	49 (10.9%)
pathologic N3	13 (13.5%)	42 (11.9%)	55 (12.3%)
pathologic NX	14 (14.6%)	18 (5.1%)	32 (7.1%)
Not reported	1 (1.0%)	17 (4.8%)	18 (4.0%)
pathologic M			
pathologic M0	90 (93.8%)	311 (88.4%)	401 (89.5%)
pathologic M1	4 (4.2%)	18 (5.1%)	22 (4.9%)
Not reported	2 (2.1%)	23 (6.5%)	25 (5.6%)

Table 2 Patient demographics and clinical factors of GEO(n=44)

GSE19234	Alive	Death	Total
	(N=20)	(N=24)	(N=44)
Gender			
Male	15 (75.0%)	13 (54.2%)	28 (63.6%)
Female	5 (25.0%)	11 (45.8%)	16 (36.4%)
Age (years)			
Mean (SD)	64.1 (16.4)	62.7 (20.0)	63.3 (18.2)
Median [Min, Max]	64.0 [30.0, 87.0]	67.0 [30.0, 92.0]	66.5 [30.0, 92.0]
Stage			
Stage IIIA	4 (20.0%)	0 (0%)	4 (9.1%)
Stage IIIB	11 (55.0%)	12 (50.0%)	23 (52.3%)
Stage IIIC	5 (25.0%)	7 (29.2%)	12 (27.3%)
Stage IV	0 (0%)	5 (20.8%)	5 (11.4%)
days to metastasis			
Mean (SD)	800 (235)	488 (261)	630 (292)
Median [Min, Max]	792 [282, 1160]	440 [68.0, 1070]	631 [68.0, 1160]
days to diagnosis			
Mean (SD)	1780 (1030)	1630 (1650)	1700 (1390)
Median [Min, Max]	1510 [282, 4230]	831 [318, 5680]	1140 [282, 5680]

Table 3. Univariate Cox regression analysis of risk scores

Vari	HR (95% CI)	p-value
pathologic T T stage	1.101(1.013-1.197)	0.022
clark	1.261(1.026-1.550)	0.028
<i>PI3</i>	1.136(1.015-1.270)	0.026
<i>SLPI</i>	1.169(1.066-1.282)	0.0009
<i>DEFB1</i>	1.199(1.011-1.423)	0.037
<i>S100A7</i>	1.167(1.052-1.294)	0.003
<i>S100A7A</i>	2.621(1.575-4.361)	0.0002
<i>S100A14</i>	1.211(1.077-1.363)	0.001
<i>MMP9</i>	0.877(0.817-0.941)	0.0002
<i>SLC40A1</i>	0.882(0.797-0.977)	0.0161
<i>CYBB</i>	0.796(0.726-0.872)	1.14E-06
<i>LYZ</i>	0.847(0.792-0.905)	1.06E-06
<i>MARCO</i>	0.846(0.775-0.923)	0.000173
<i>CCL19</i>	0.891(0.838-0.946)	0.00017
<i>CXCR4</i>	0.820(0.739-0.910)	0.000191
<i>CD79A</i>	0.829(0.771-0.892)	4.45E-07
<i>C3</i>	0.894(0.831-0.962)	0.003
<i>CBLC</i>	1.896(1.351-2.659)	0.0002

Table 4. Multivariate Cox regression analysis of risk scores

Vair	HR (95% CI)	p-value
pathologic T T stage	1.205(1.024-1.417)	0.025
<i>SLPI</i>	1.179(1.036-1.343)	0.013
<i>S100A7</i>	1.230(1.077-1.405)	0.002
<i>LYZ</i>	0.799(0.706-0.905)	0.0003
<i>CCL19</i>	1.140(0.992-1.310)	0.065
<i>CXCR4</i>	1.192(0.995-1.427)	0.056
<i>CD79A</i>	0.779(0.658-0.924)	0.004

Figures

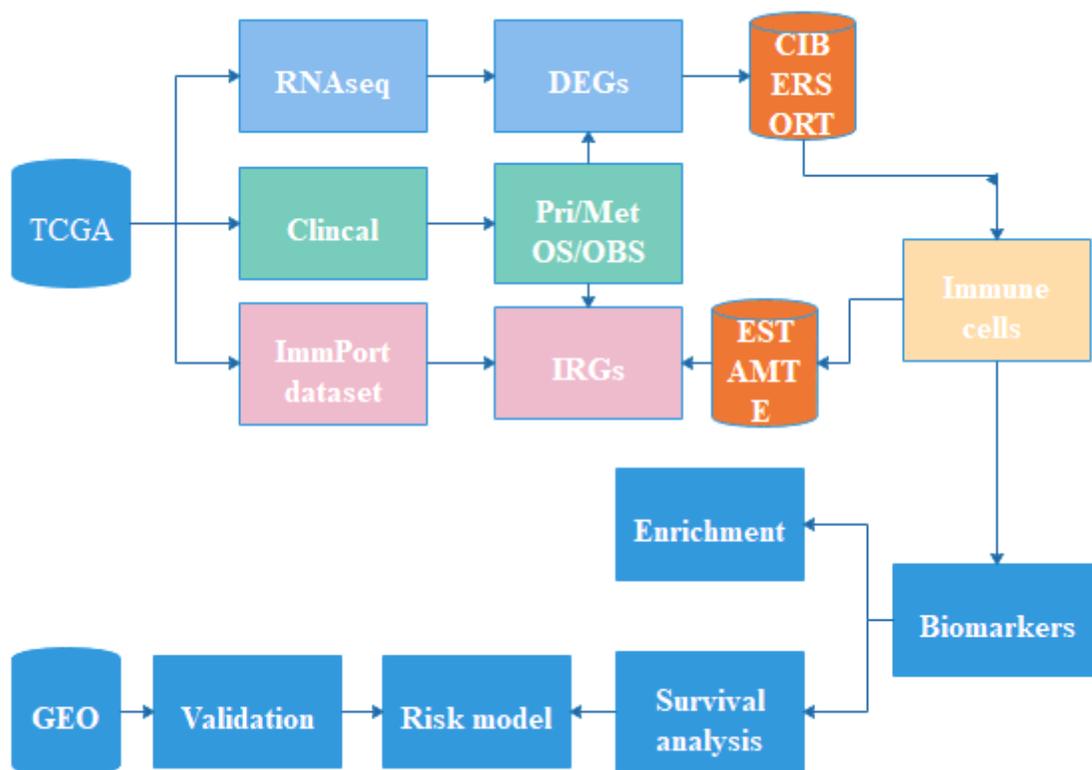


Figure 1

Flow diagram for data analysis

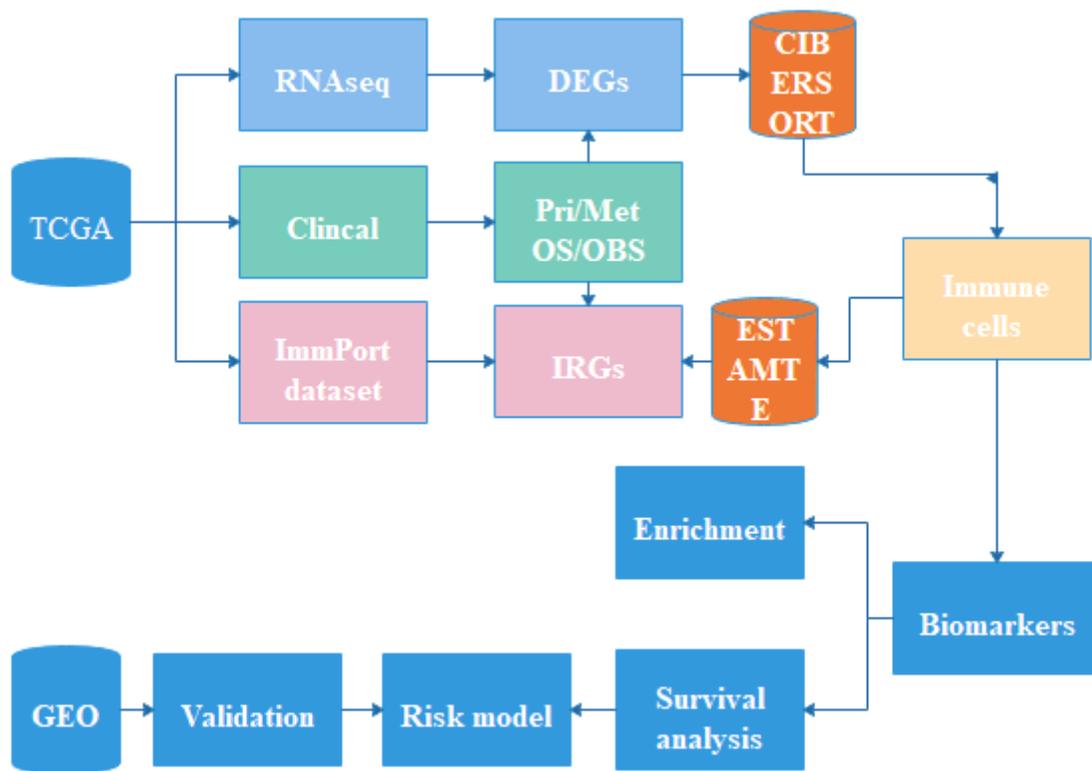


Figure 1

Flow diagram for data analysis

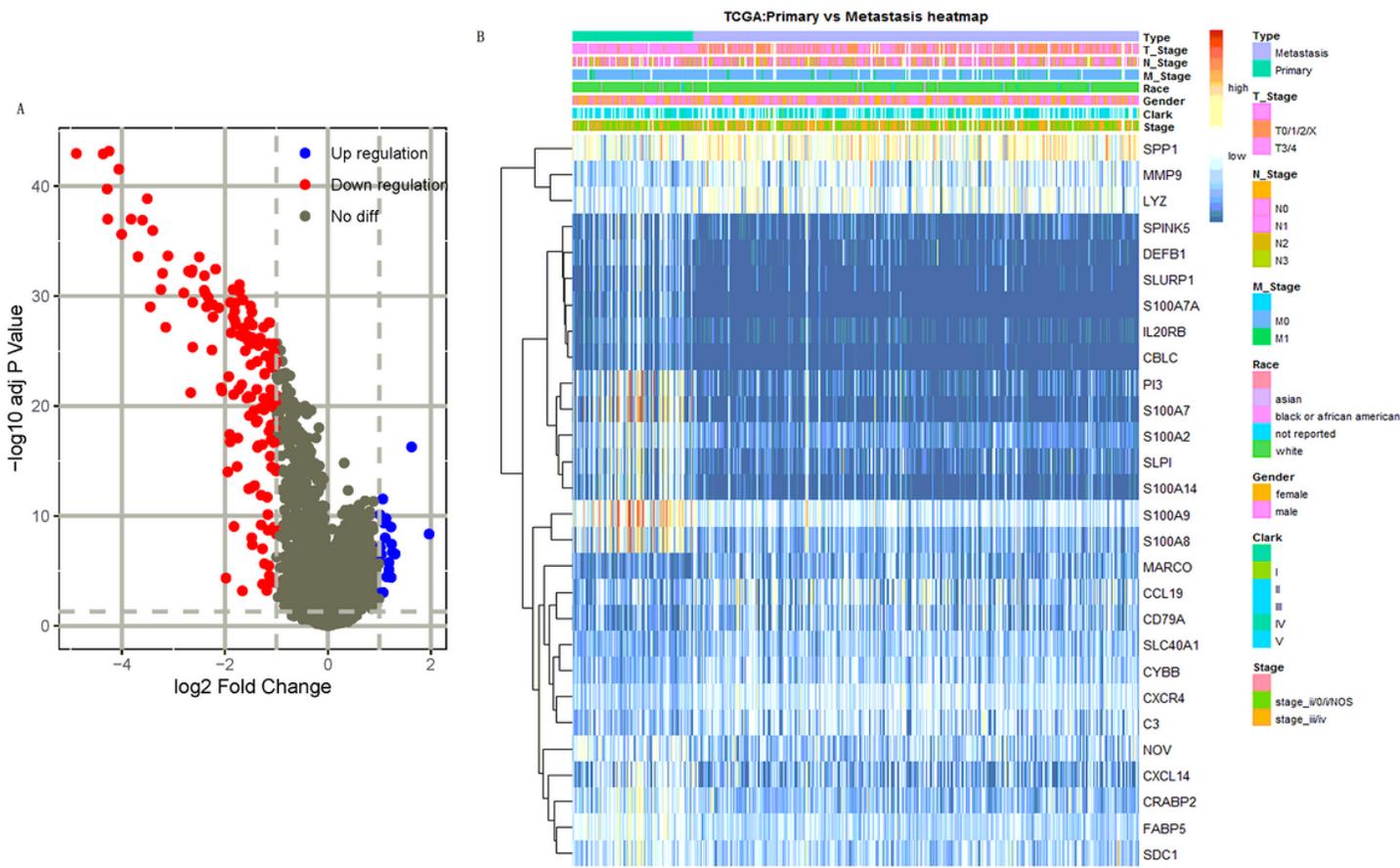


Figure 2

The volcano and heatmap of DEGs and IRGs in primary and metastasis groups (A) The volcano of DEGs. (B) The heatmap of 28 IRGs in primary and metastasis groups.

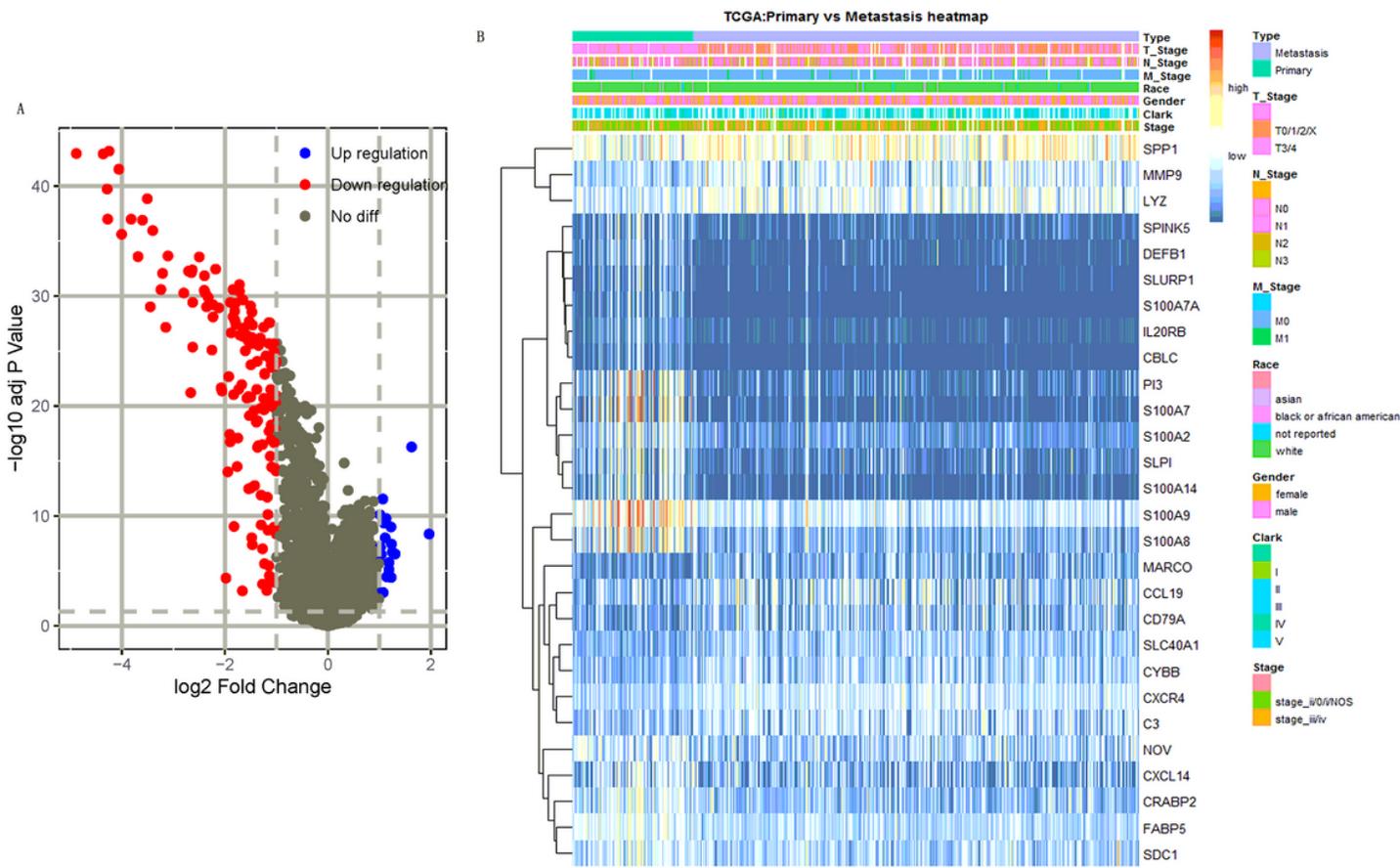


Figure 2

The volcano and heatmap of DEGs and IRGs in primary and metastasis groups (A) The volcano of DEGs. (B) The heatmap of 28 IRGs in primary and metastasis groups.

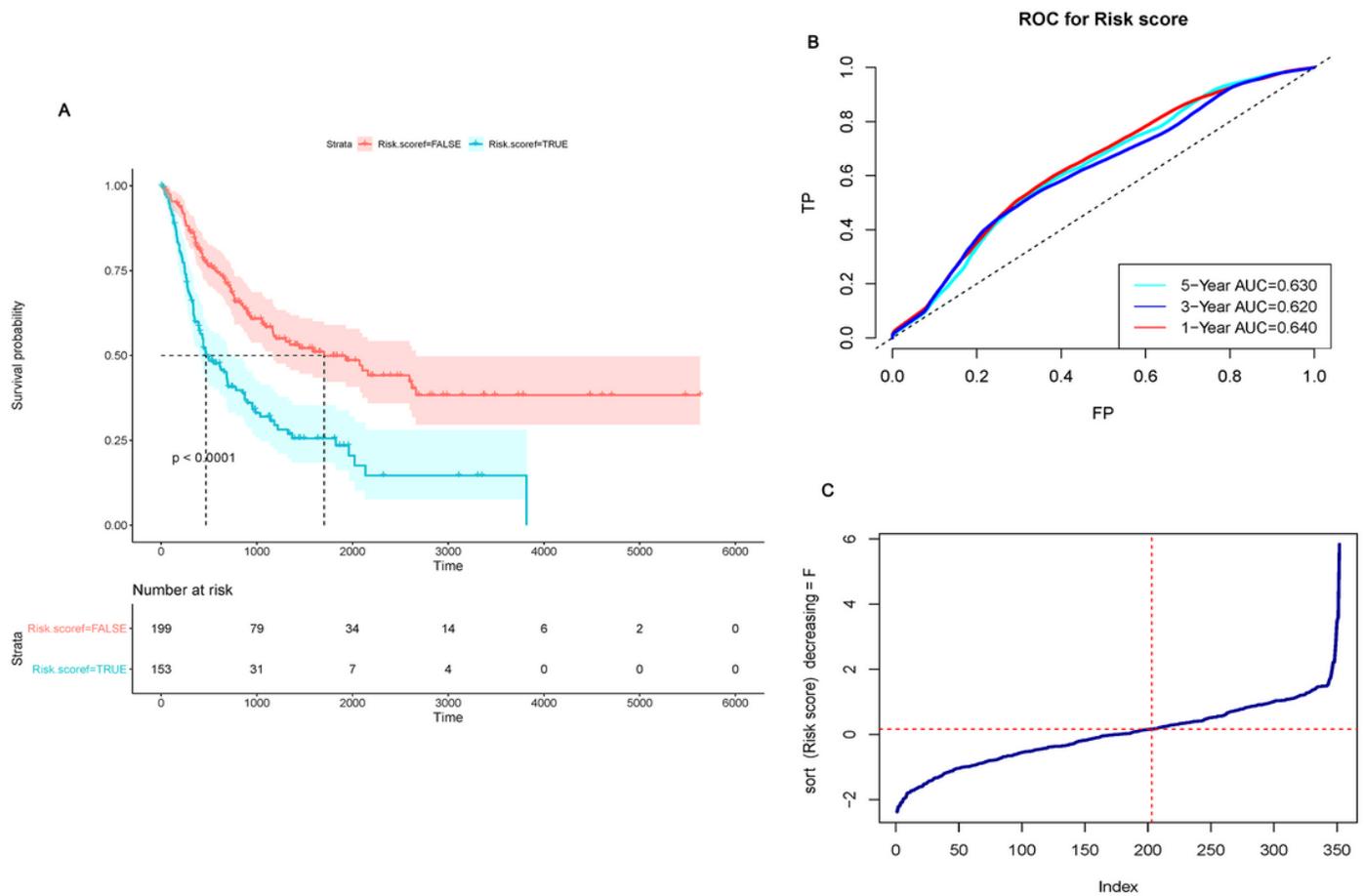


Figure 3

Data mining results of the training set. (A) K-M curves showing overall survivals of high-risk and low-risk patients.(B) Time-dependent ROC curve based on the risk score. (C) For OBS, the best cutoff point allows J to be maximal of risk score value (0.153) is as a sign for classification with metastasis MM patients into low-risk and high-risk groups

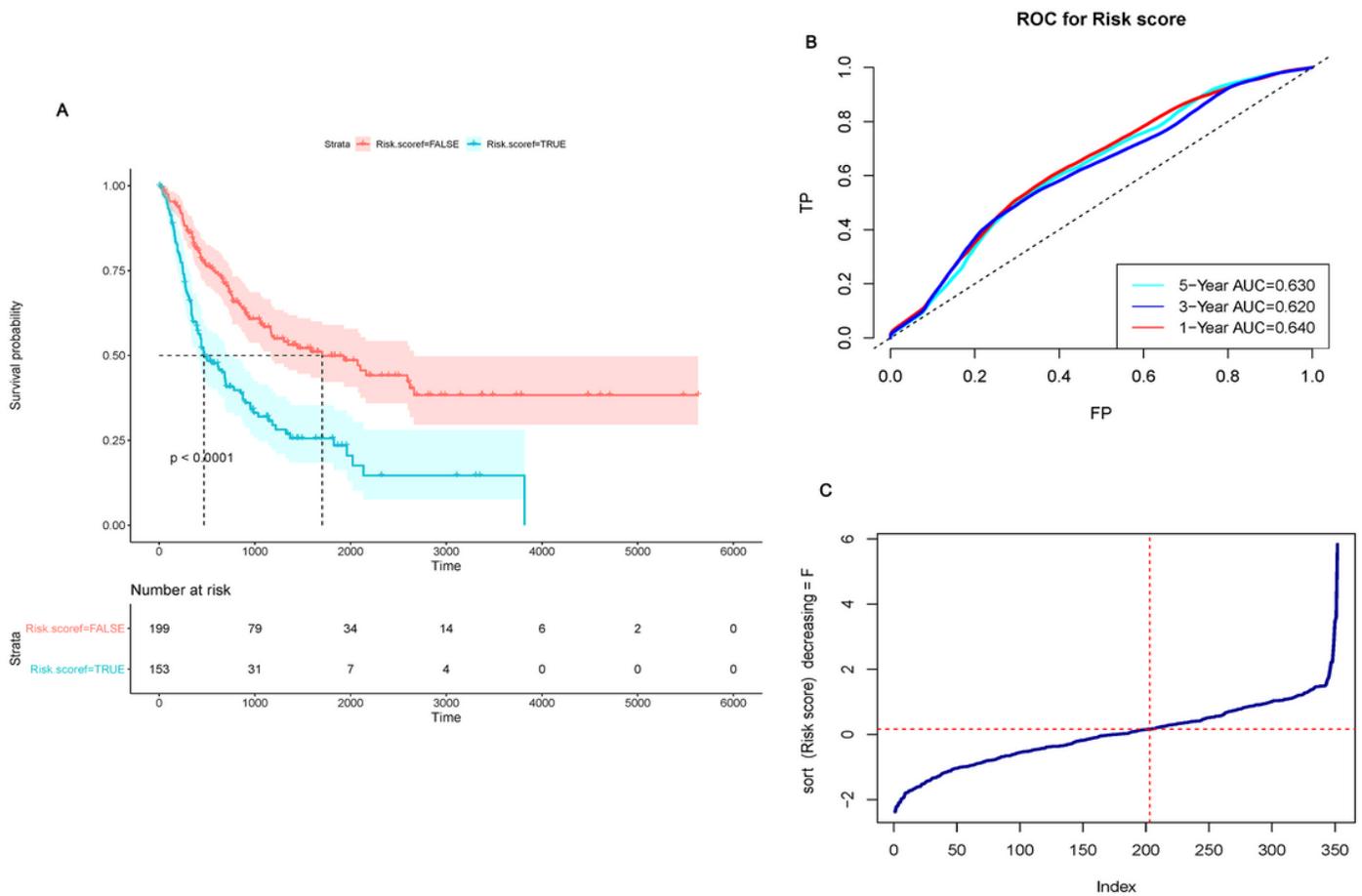


Figure 3

Data mining results of the training set. (A) K-M curves showing overall survivals of high-risk and low-risk patients.(B) Time-dependent ROC curve based on the risk score. (C) For OBS, the best cutoff point allows J to be maximal of risk score value (0.153) is as a sign for classification with metastasis MM patients into low-risk and high-risk groups

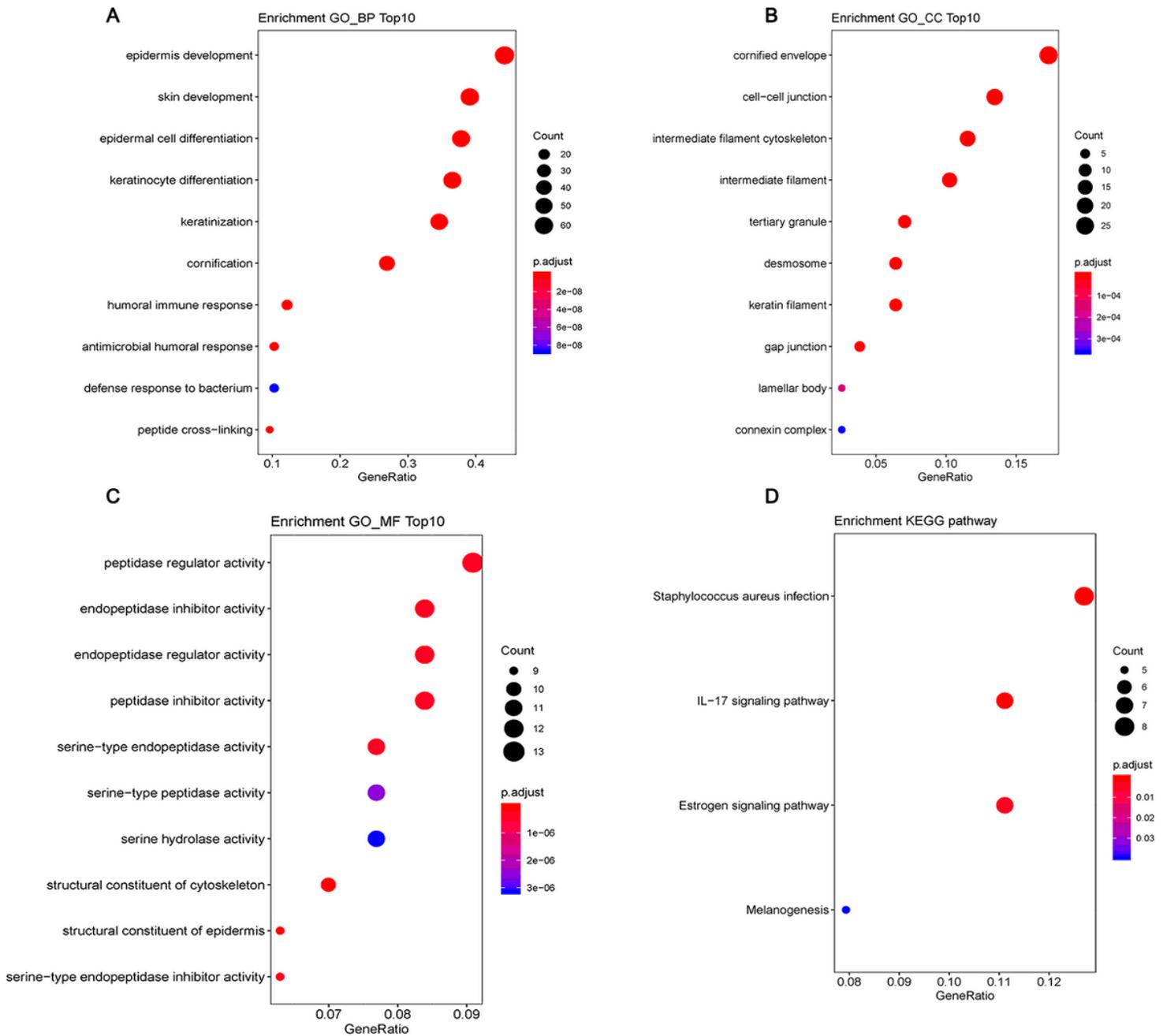


Figure 4

Enrichment pathway of gene-signature (A) Biological Process for 172 DEGs, BP. (B) Molecular Function for 172 DEGs, MF. (C) Cellular Components for 172 DEGs, CC. (D) Enrichment KEGG pathway for 172 DEGs.

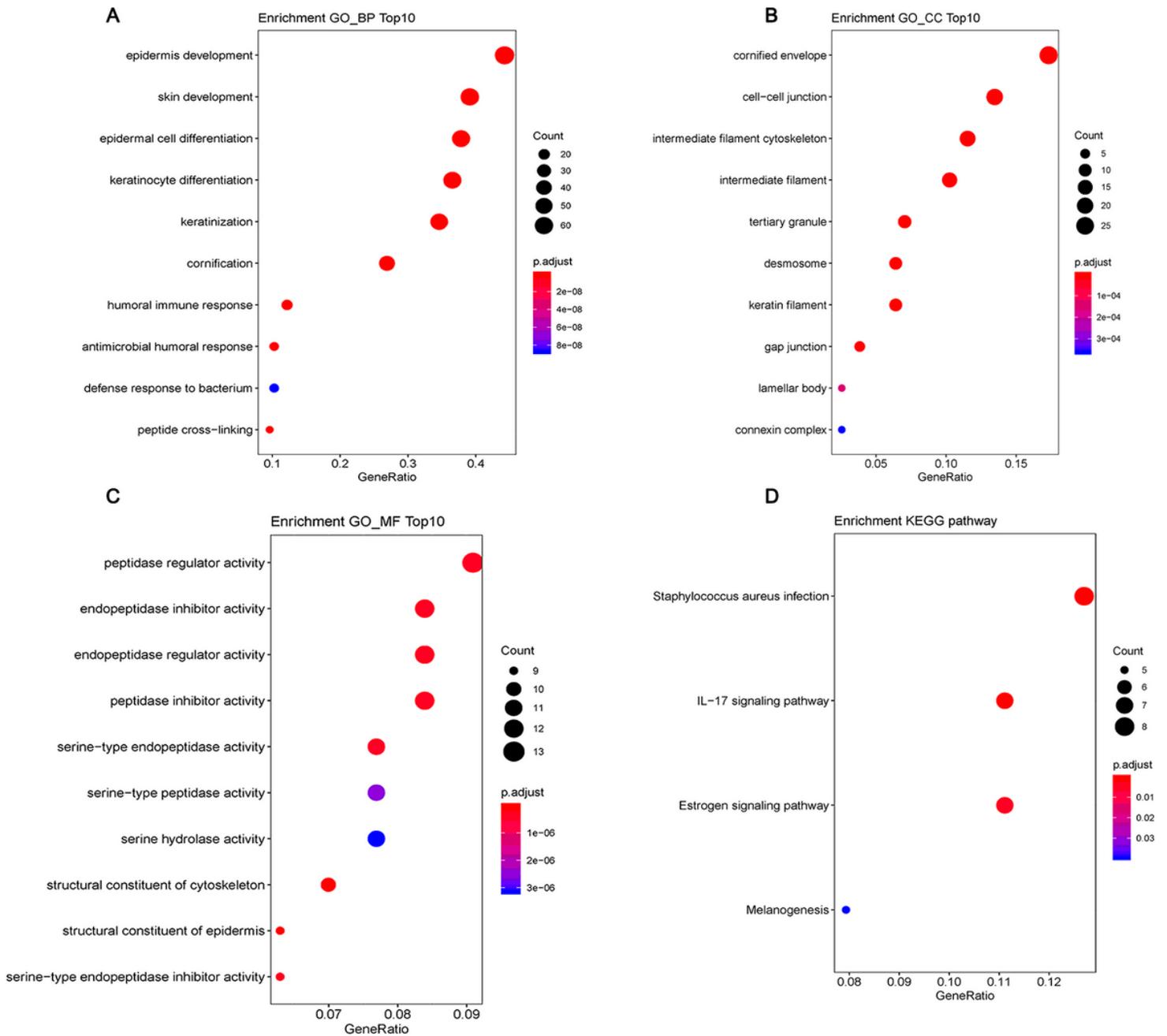


Figure 4

Enrichment pathway of gene-signature (A) Biological Process for 172 DEGs, BP. (B) Molecular Function for 172 DEGs, MF. (C) Cellular Components for 172 DEGs, CC. (D) Enrichment KEGG pathway for 172 DEGs.

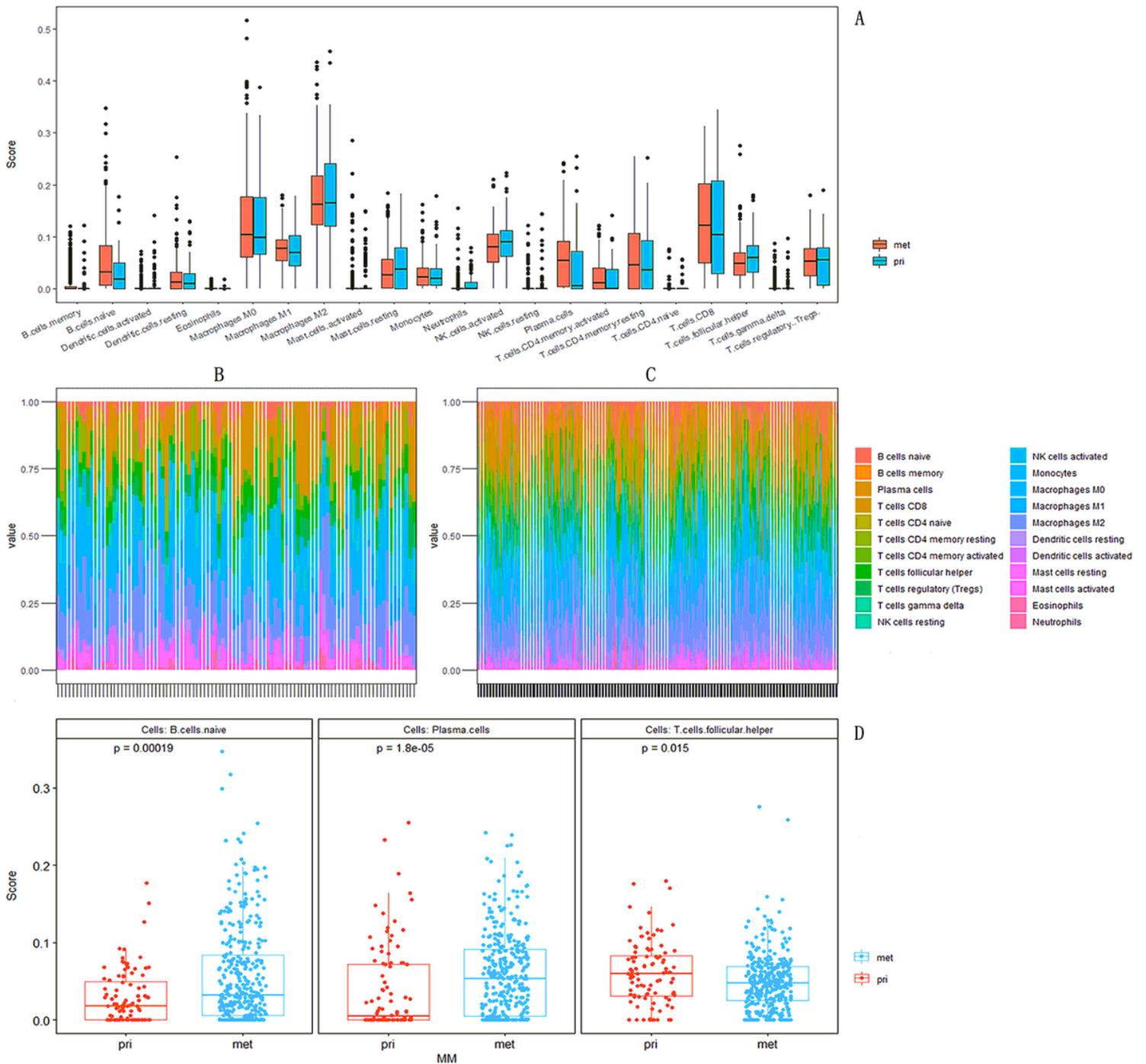


Figure 5

Boxplot showing the differential abundance of 22 infiltrative immune cells calculated by CIBERSORTx between primary group and metastasis group in MM from TCGA cohorts (A) Boxplot showing the abundance of 22 infiltrative immune cells calculated between primary group and metastasis group. (B) 22 infiltrative immune cells of primary group. (C) 22 infiltrative immune cells of metastasis group. (D) Boxplot showing the differential abundance of infiltrative immune cells between primary group and metastasis group.

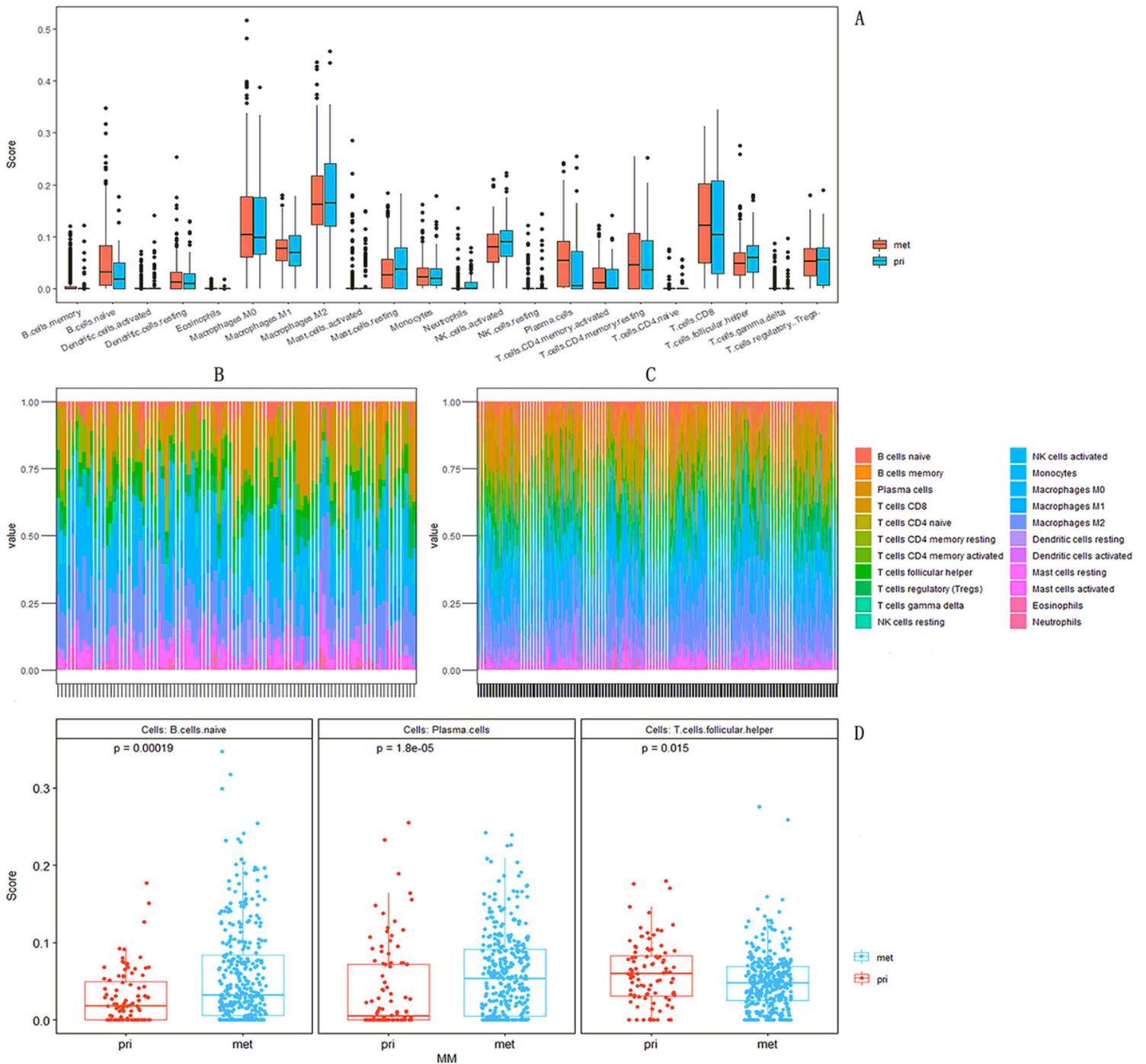


Figure 5

Boxplot showing the differential abundance of 22 infiltrative immune cells calculated by CIBERSORTx between primary group and metastasis group in MM from TCGA cohorts (A) Boxplot showing the abundance of 22 infiltrative immune cells calculated between primary group and metastasis group. (B) 22 infiltrative immune cells of primary group. (C) 22 infiltrative immune cells of metastasis group. (D) Boxplot showing the differential abundance of infiltrative immune cells between primary group and metastasis group.

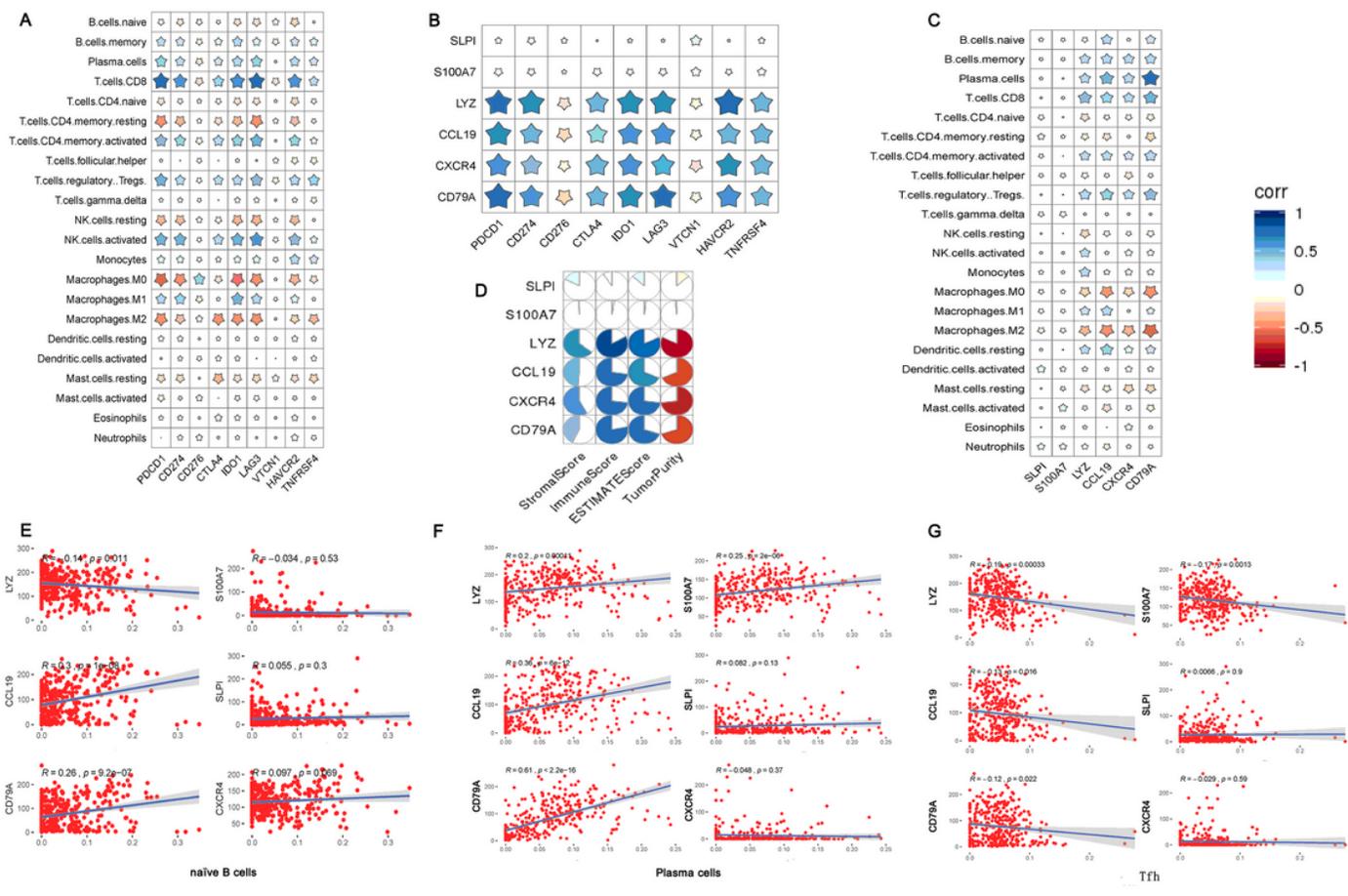


Figure 7

The correlation between IRGs and TME in metastasis tissue of MM (A) The correlation between ICGs and 22 type ITLS. (B) The correlation between ICGs and IRGs. (C) The correlation between IRGs and 22 type ITLS. (D) The correlation between IRGs and TME. (E) The correlation between IRGs and naïve B cells. (F) The correlation between IRGs and plasma cells. (G) The correlation between IRGs and Tfh

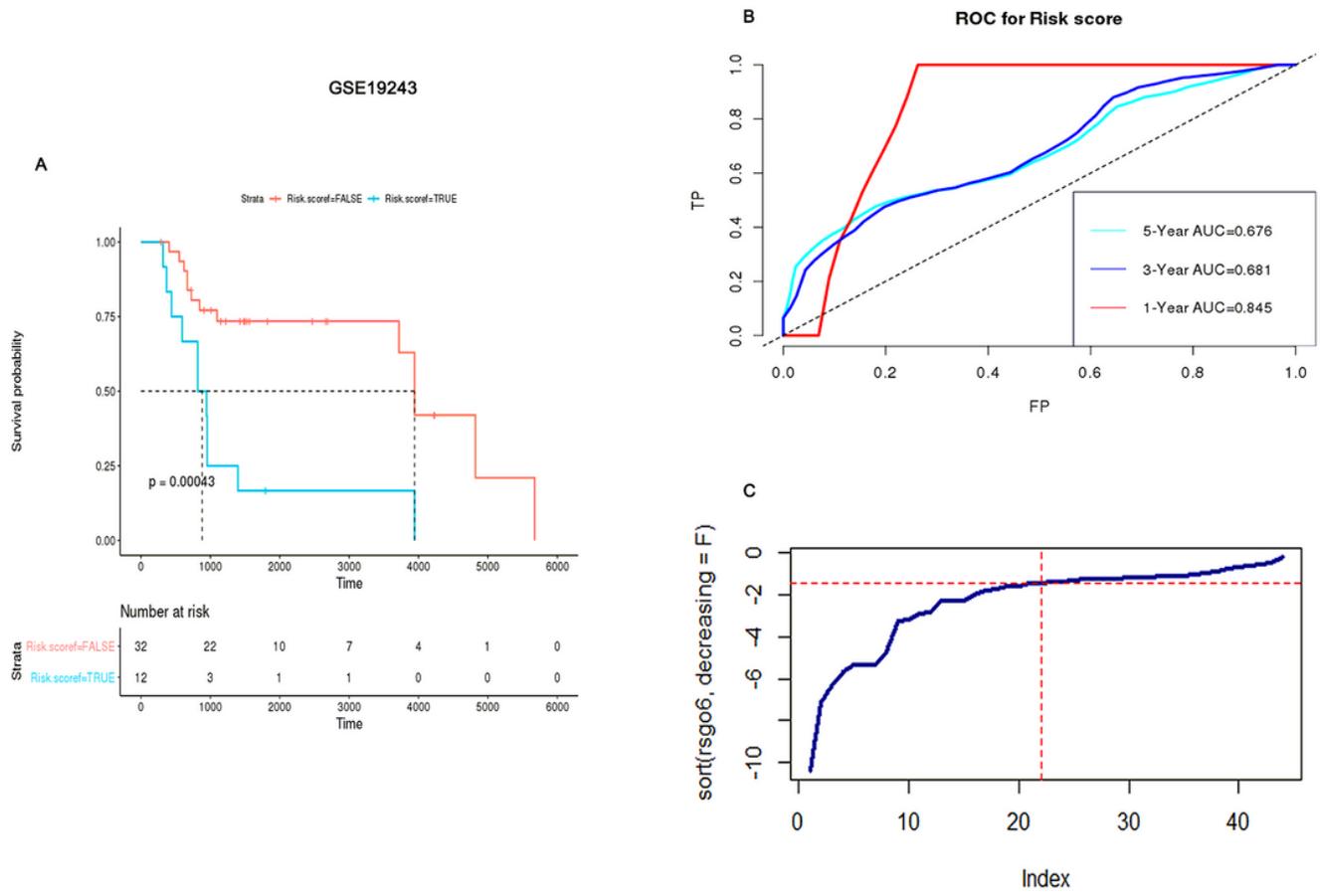


Figure 8

Data mining results of the GEO dataset. (A) Kaplan-Meier (K-M) curves showing overall survivals of high-risk and low-risk patients. (B) Time-dependent ROC curve based on the risk score. (C) For OS, the best cutoff point allows J to be maximal of risk score value (-1.133) is as a sign for classification with MM patients into low-risk and high-risk groups in GEO cohort (GSE19234).

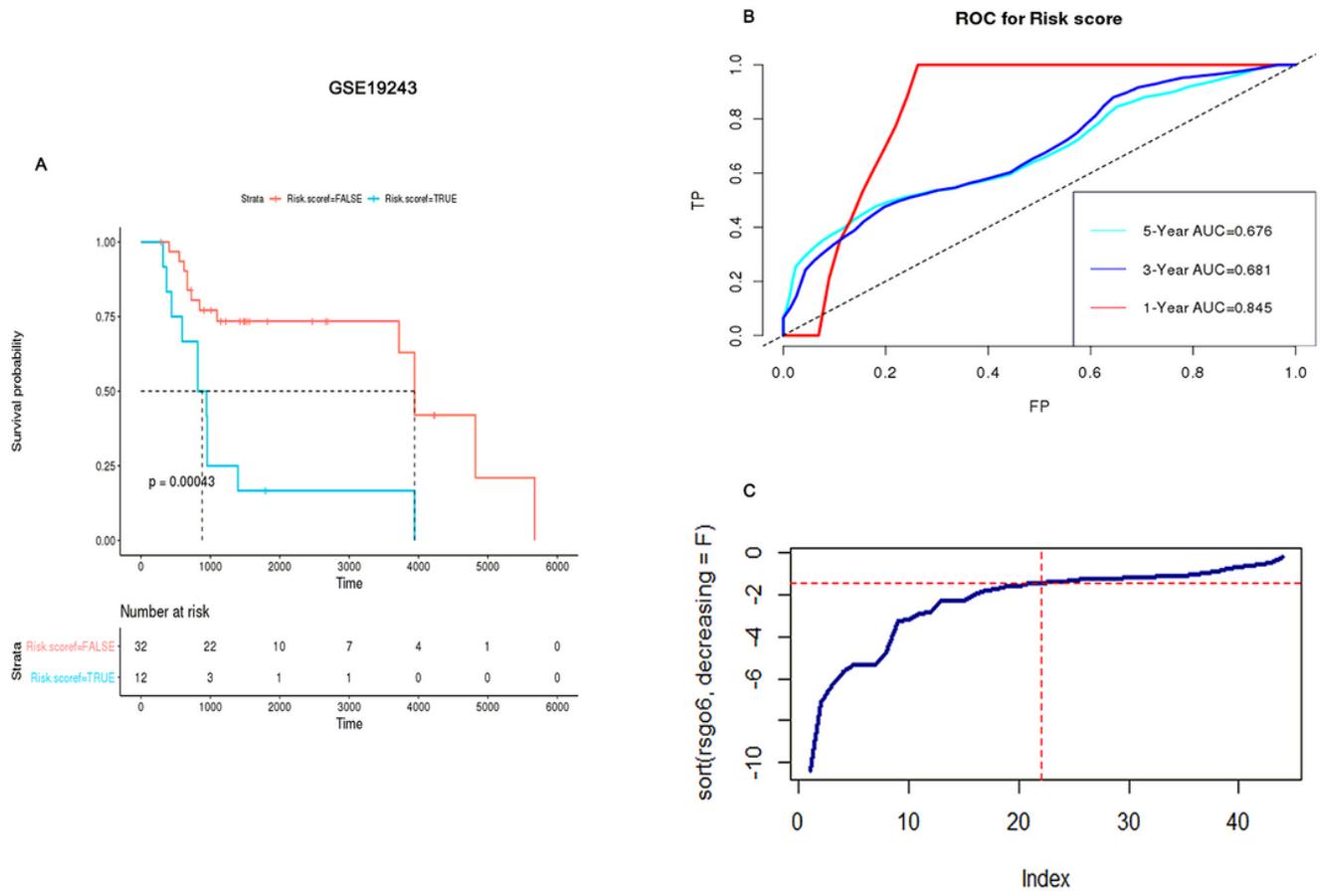


Figure 8

Data mining results of the GEO dataset. (A) Kaplan-Meier (K-M) curves showing overall survivals of high-risk and low-risk patients. (B) Time-dependent ROC curve based on the risk score. (C) For OS, the best cutoff point allows J to be maximal of risk score value (-1.133) is as a sign for classification with MM patients into low-risk and high-risk groups in GEO cohort (GSE19234).

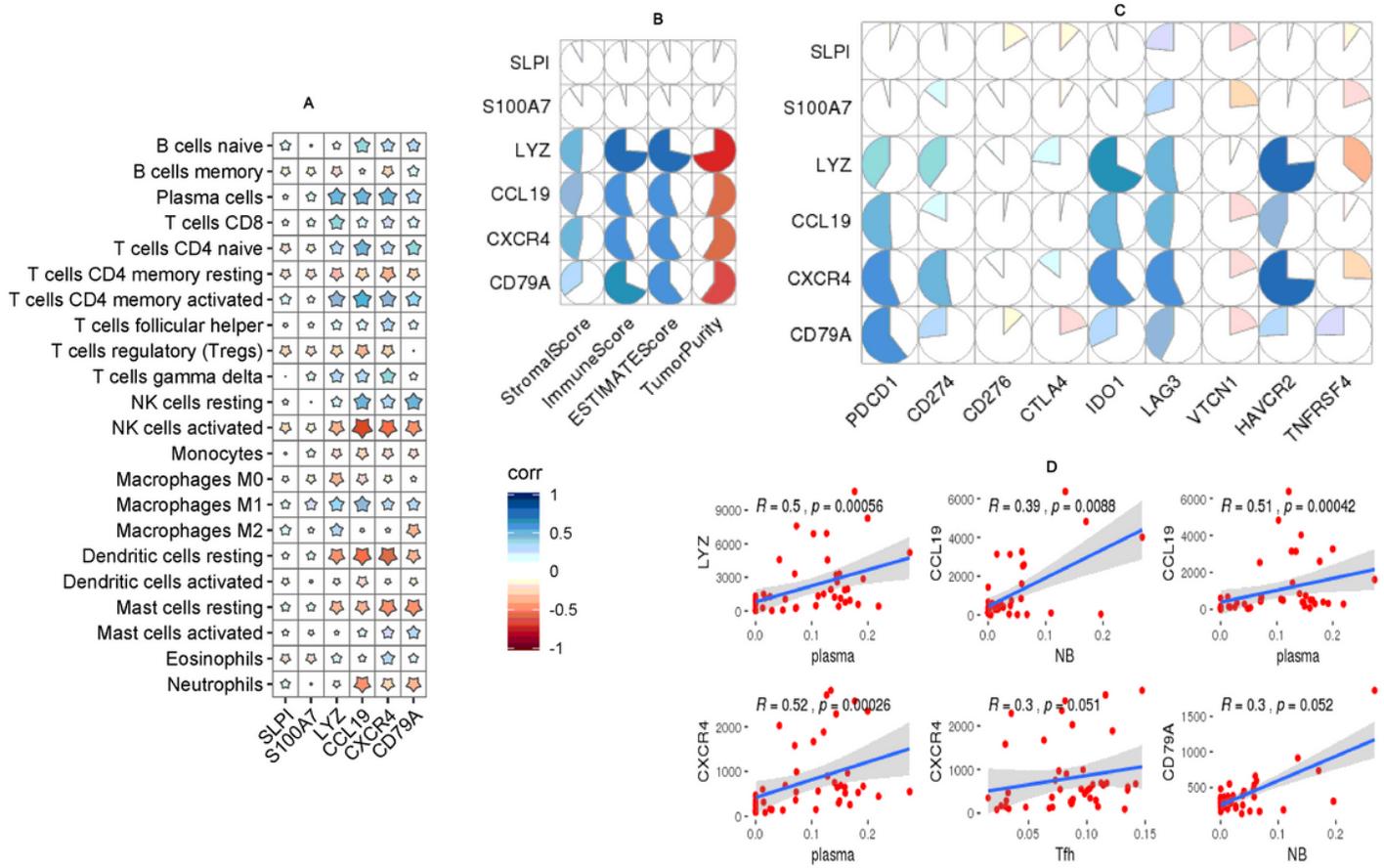


Figure 9

The correlation between IRGs and TME in metastasis tissue of MM in GEO (A) The correlation between IRGs and 22 type ITLS. (B) The correlation between IRGs and TME. (C) The correlation between IRGs and ICGs. (D) The correlation between IRGs and 3 type ITLs.

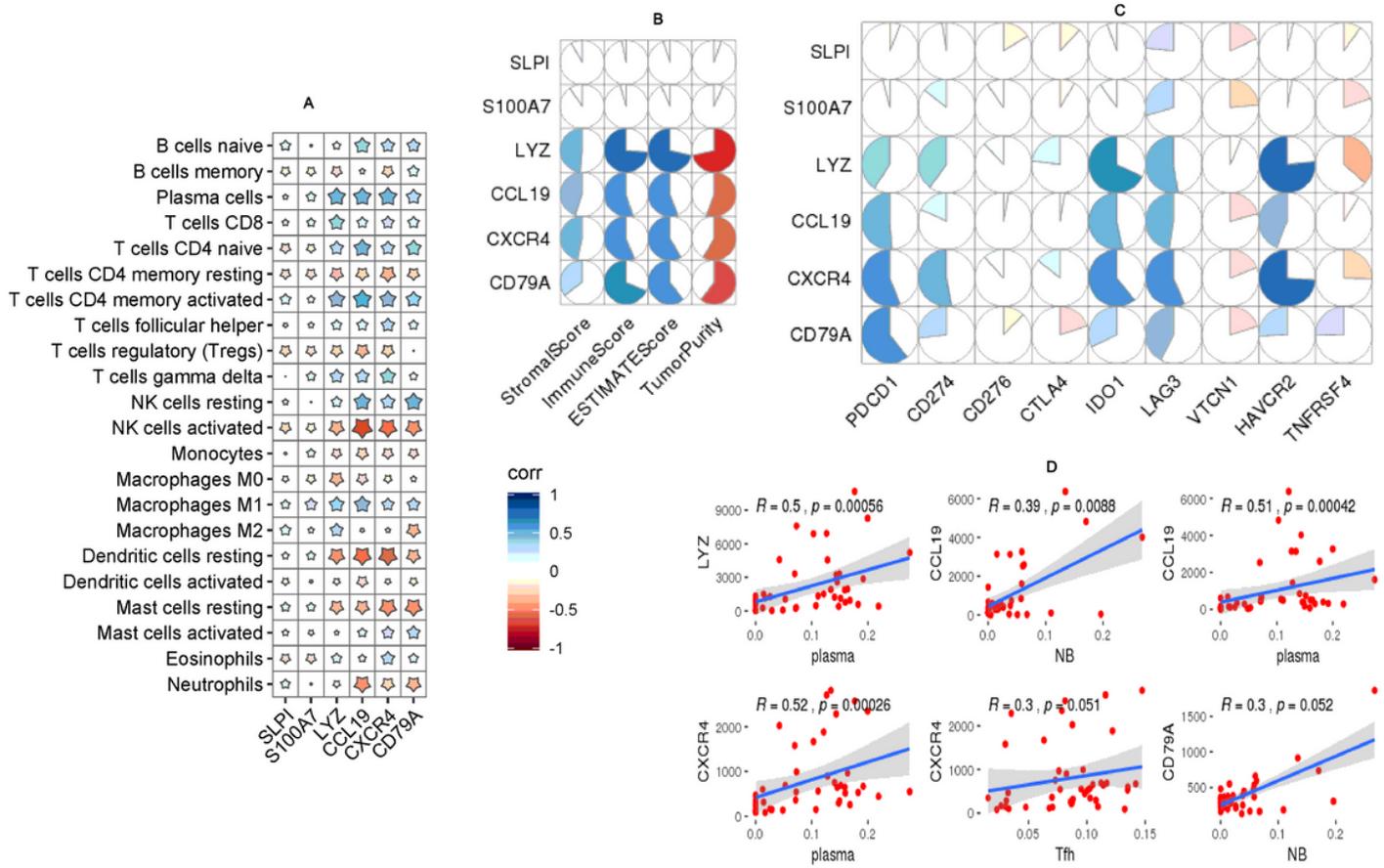


Figure 9

The correlation between IRGs and TME in metastasis tissue of MM in GEO (A) The correlation between IRGs and 22 type ITLs. (B) The correlation between IRGs and TME. (C) The correlation between IRGs and ICGs. (D) The correlation between IRGs and 3 type ITLs.

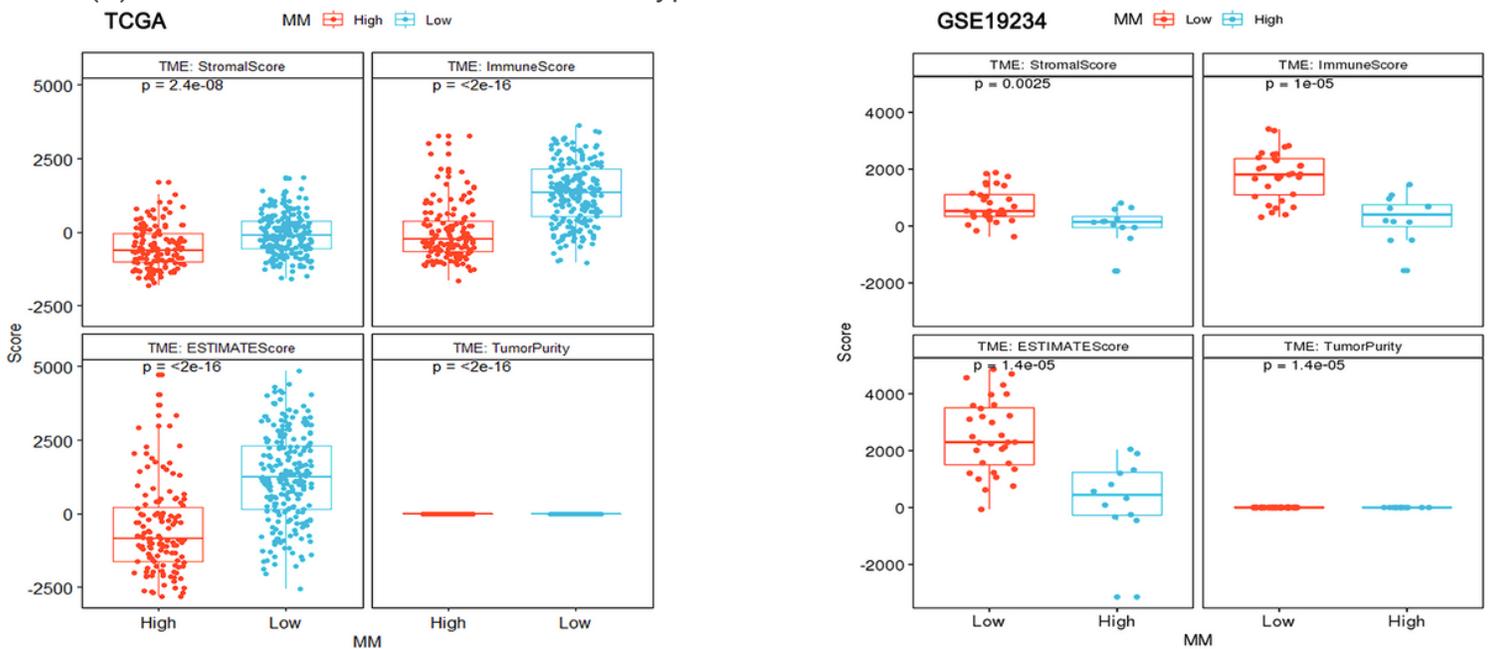


Figure 10

The box plots of tumor purity in low- and high-risk groups.

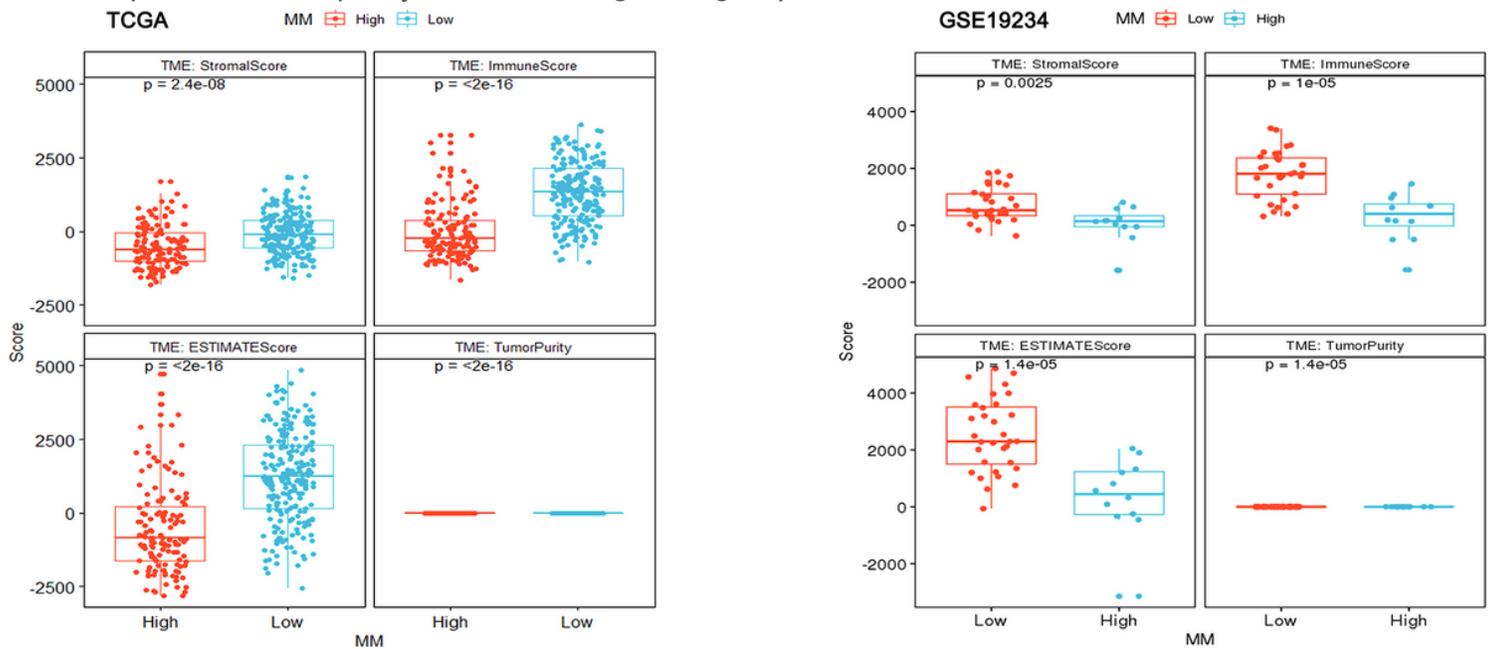


Figure 10

The box plots of tumor purity in low- and high-risk groups.

Supplementary Files

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

- [Supplementfile13.xlsx](#)
- [Supplementfile13.xlsx](#)
- [Supplementfile12.xlsx](#)
- [Supplementfile12.xlsx](#)
- [Supplementfile11.xlsx](#)
- [Supplementfile11.xlsx](#)
- [Supplementfile10.xlsx](#)
- [Supplementfile10.xlsx](#)
- [Supplementfile9.xlsx](#)
- [Supplementfile9.xlsx](#)
- [Supplementfile8.xlsx](#)
- [Supplementfile8.xlsx](#)
- [Supplementfile7.xlsx](#)

- [Supplementfile7.xlsx](#)
- [Supplementfile6.xlsx](#)
- [Supplementfile6.xlsx](#)
- [Supplementfile5.xlsx](#)
- [Supplementfile5.xlsx](#)
- [Supplementfile4.xlsx](#)
- [Supplementfile4.xlsx](#)
- [Supplementfile3.xlsx](#)
- [Supplementfile3.xlsx](#)
- [Supplementfile2CIBERSORTxTCGAResults.pdf](#)
- [Supplementfile2CIBERSORTxTCGAResults.pdf](#)
- [Supplementfile128IRGs.xlsx](#)
- [Supplementfile128IRGs.xlsx](#)