A Novel 23-Gene Expression Signature is Predictor of Inferior Prognosis in Melanoma

Min Du
Qilu Hospital of Shandong University Qingdao
Ling Ma
Qilu Hospital of Shandong University
Fan Zhang
Qilu Hospital of Shandong University
Dong Li (drli2021@163.com)
Qilu Hospital of Shandong University

Research Article

Keywords: melanoma, the 23-gene signature, overall survival

Posted Date: March 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-264055/v1

Abstract

Background: Melanoma is a serious form of skin cancer that begins in melanocytes. Metastasis, somatic mutations and gene expression profiles are important prognostic factors for melanoma patients. However, accurate prediction of patient prognosis remains an unsolved problem for the disease. This study was to develop a novel gene profile to accurately classify melanoma patients into subgroups with different survival probabilities.

Methods: Survival-related genes were determined by Kaplan–Meier survival analysis and multivariate analysis using the expression and clinical data of 467 melanoma patients from The Cancer Genome Atlas (TCGA) database and validated in an independent Gene Expression Omnibus (GEO) dataset. Feature selection was performed by the Least Absolute Shrinkage and Selection Operator (LASSO) method. A prognostic 23-gene score was established and compared with two known gene-expression risk scores. The stratification of melanoma patients was performed by unsupervised hierarchical clustering of 23 gene expression levels to identify clusters of melanoma patients with different survival probabilities.

Results: The LASSO model comprising 23 genes was considered as the optimal model. The 23-gene score was associated with increased mortality in melanoma patients regardless of clinicopathological characteristics. Hierarchical clustering analysis of the 23 genes revealed three subgroups of melanoma patients. The cluster3 melanoma tumours were associated with higher 23-gene score and poorer overall survival than cluster1 and cluster2 tumours. The 23-gene score had higher area under curve (0.76) than the 8-gene risk score and IRGs score (0.58 and 0.59) in the prediction of overall survival of melanoma patients.

Conclusions: The 23-gene score is superior to the two established prognostic gene signatures in the prediction of prognosis of melanoma patients.

Background

Melanoma is the most serious type of skin cancer that develops from melanocytes. Melanoma mostly occurs in the skin but rarely occurs in the eyes and inside body, such as in nose or throat[1]. Skin cutaneous melanoma accounts for only 2% of total skin cancer cases. However, it causes more than 72% of deaths in skin carcinoma, due to its high degree of malignancy and invasiveness [2]. The disease has an overall 5-year survival rate of 92%, however, once the tumor has metastasized to distant tissues, the 5-year survival rate could be reduced to only 15–20% [2]. Over the last few decades, the incidence of malignant melanoma continues to increase annually [3], with increased incidence particularly in males and young women (≤ 39 years) [4, 5].

To date, the relationships between various clinicopathological features and prognosis have been elucidated. For instance, metastasis is the major factor which affects melanoma patient prognosis, accounting for more than 90% of deaths in melanoma [6, 7]. Recent whole-genome mRNA expression [illegible] umors into distinct molecular subtypes which

are associated with clinical properties and prognosis based on gene expression patterns [8, 9]. Several gene patterns in melanoma have been reported to independently predict melanoma patients with a high risk of poor survival [10, 11]. However, these studies have focused on immune markers and their relationship with prognosis. Altogether, a better understanding of the molecular characteristics of melanoma is highly significant.

In the present study, we aimed to develop a novel gene profile to accurately classify melanoma patients into subgroups with different survival probabilities. We performed various survival analyses to screen for prognosis-associated genes using the expression and clinical data of The Cancer Genome Atlas (TCGA) database [12]and validated the results in the melanoma patients of Gene Expression Omnibus (GEO) database[13]. We created a novel 23-gene score based on a linear combination of 23 gene expression levels to accurately predict the overall survival (OS) of melanoma patients. Lastly, we performed unsupervised hierarchical clustering of 23 genes and defined three distinct subsets of melanoma patients with significant differences in overall survival.

Methods And Materials

Data acquisition

RNA-seq expression data of 19639 genes, and clinical characteristics of 467 melanoma patients were obtained from the TCGA database. Genes whose expression values were zero in more than 10% melanoma samples were removed from the study, leaving15218 genes in the study. Gene expression and clinical data of 124 melanoma patients were downloaded from the GEO database (GSE59455) to validate the associations of gene expression with OS. Detailed clinical information of melanoma patients of the TCGA and GEO datasets are shown in Table 1 and supplementary Table 1 respectively.

Survival Analyses

To investigate the association between gene expression and OS, melanoma patients were divided into the "high-expression" and "low-expression" groups according to the median expression value. Kaplan–Meier curves and log-rank methods were utilized to study the prognostic importance of gene expression using the survival package [14, 15]. Multivariate Cox regression model was performed to confirm whether gene expression was independently prognostic after adjustment of clinical factors. Survival-related genes were further classified into risk genes with odd ratio (OR) greater than 1 and protective genes with OR ranging from 0 and 1. Receiver operating characteristic (ROC) curve analysis was conducted by the R package of pROC to further validate the prognostic importance of the 23-gene score [16]. Then, we computed the area under curve (AUC) values accordingly by the R package of pROC for three different prognostic scores.

Development And Validation Of The 23-gene Score

As the gene expression units are different between the TCGA and GEO cohorts, gene expression was normalized following the z score formula z = (x-$\bar{x}$)/s, where x is the gene expression value, $\bar{x}$ is the mean, and s is the standard deviation. We performed 10-fold cross-validation of the least absolute shrinkage and selection operator (LASSO) model to select the optimal combination of genes for the prediction of OS in the TCGA dataset using the R package glmnet[17]. A prognostic 23-gene score formula was established based on a linear combination of expression levels. 23-gene score = -0.12 + expression of gene 1 × β1 + expression of gene 2 × β2 +⋯+ expression of gene n × βn. β values were the coefficients derived from the LASSO model of the TCGA dataset. We performed Kaplan–Meier survival analysis and multivariate Cox regression analysis to analyze the association between 23-gene score and OS. The associations between clinical characteristics and 23-gene score were analyzed by linear regression model. P < 0.05 was considered statistically significant.

Unsupervised Hierarchical Clustering Analysis

Hierarchical clustering analysis of the 23 genes revealed three subgroups of melanoma patients in the TCGA dataset (Fig. 4A). The cluster3 melanoma tumours were associated with higher 23-gene score and poorer OS than cluster1 and cluster2 tumours (P values < 0.05 for all cases, student t test or log-rank test, Fig. 4B and supplementary Table 6). We also performed the classification of 124 melanoma patients using the 23 gene panel and found three clusters of melanoma patients in the GEO dataset (Fig. 4C). Cluster3 tumours exhibited significantly higher 23-gene score and poorer OS than cluster1 tumours (P values < 0.05 for all cases, student t test or log-rank test, Fig. 4D and supplementary Table 7). The other clinical factors didn't exhibit significant difference between subgroups of melanoma patients in the TCGA and GEO cohorts (P values > 0.05 for all cases, fisher exact test or student t test, supplementary Table 6 and 7). These results suggest the 23-gene panel successfully classifies melanoma patients into clusters of patients with different survival probabilities independently of clinical characteristics.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) [19] was implemented using a java software developed and the default parameters were used by comparing the high-risk (higher than median) and low-risk group (lower than median). P < 0.05 was considered statistically significant.

Results

Survival analyses between patient mortality and gene expression in melanoma

To evaluate the prognostic importance of gene expression for patients' OS, the 467 melanoma patients of

on groups based on the median expression

values. Kaplan-Meier survival analysis showed that high expression levels of 558 genes and 1067 genes were associated with longer or shorter OS respectively ($P < 0.05$ for all cases, log rank test, Fig. 1). Then multivariate Cox regression analysis were performed between patients' OS and clinical features, and 1625 gene expression levels. Multivariate Cox regression analysis confirmed 249 protective genes and 529 risk genes after the adjustment of clinical features. In order to validate the findings above, the association between 778 gene expression and mortality was analyzed in 124 melanoma samples of the GEO dataset. Kaplan-Meier survival analysis and multivariate Cox regression analysis confirmed 56 genes and 11 genes were positively and negatively correlated with OS in the GEO cohort respectively ($P < 0.05$ for all cases, log rank test, Fig. 1).

The twenty-three score is a negative prognostic factor in melanoma

We performed 10-fold cross validation of LASSO model to determine the optimal model for the prediction of OS in the TCGA dataset. When the log (lambda) was equal to -4.2 and the number of genes with non-zero coefficients was 23, the AUC value of LASSO model was maximum and the mean squared error was minimum (Fig. 2A-B). Therefore, the LASSO model comprising 23 genes was considered as the optimal model. The association of 23 genes with OS, intercept and coefficients of 23 genes were presented in the supplementary Table 2,3 4 and supplementary Fig. 1. A prognostic 23-gene score formula was established based on a linear combination of the expression levels of 23 genes weighted with the coefficients derived from the optimal LASSO model. 23-gene scores were computed for melanoma patients and then they were divided into high and low risk groups based on the median risk score. The patients with high 23-gene scores showed higher mortality rates than those with low 23-gene scores in the TCGA cohort ($P < 0.001$, log rank test, Fig. 2C). Following adjustment of clinical factors, multivariate Cox regression analysis confirmed that the 23-gene score was associated with increased mortality rate in melanoma patients ($P < 0.001$, OR: 1.93, 95% confidence interval [CI]: 1.37–2.71, Table 2). To validate the findings above, the 23-gene score was calculated following the formula in the GEO dataset. The negative correlation was validated between OS and the 23-gene score in the GEO dataset (Table 2, supplementary Table 5 and Fig. 2D).

Correlations between the twenty-three gene score and clinical factors in melanoma

Linear regression model was used to analyze the association between the 23-gene score and clinical factors in the TCGA and GEO cohorts. In the TCGA cohort, the 23-gene score was significantly positively correlated with gender, *NRAS* mutation and negatively correlated with pathologic nodal stage ($p < 0.05$ for all cases, Fig. 3A). Furthermore, the 23-gene score also showed negative correlation with metastasis and *NRAS* mutation in the GEO cohort ($p < 0.05$ for all cases, Fig. 3B). Next, we stratified melanoma patients into subgroups based on the clinicopathological characteristics and performed the Kaplan-Meier survival analysis to analyze whether the 23-gene risk score predicts overall survival independently of clinical characteristics. Overall, the high 23-gene risk score was significantly correlated with inferior OS in age > 60, age < = 60, male, female, stage I-II, stage III-IV, NRAS-mutant, NRAS-wildtype, BRAF-mutant, BRAF-wildtype, prior malignancy-negative subgroups in the TCGA cohort ($P < 0.05$ for all cases, log rank test,

supplementary Fig. 2–3). The GEO cohort exhibited a similar distribution (supplementary Fig. 4–5). These results confirm that the 23-gene signature accurately predicts prognosis regardless of clinicopathological characteristics.

Twenty-three Score Related Pathway Analysis

In order to understand why 23-gene score is predictive of melanoma patients' survival, we partitioned the melanoma samples into high and low risk groups according to the median 23-gene score. GSEA was performed to analyze the altered signalling pathways between the two different risk groups. Six signalling pathways were significantly up-regulated in the high 23-gene score group of the TCGA cohort, with aminoacyl tRNA biosynthesis, adherens junction, lysine degradation, wnt signaling pathway and thyroid cancer the top five most enriched pathways (Fig. 5, $p < 0.05$ for all cases). Conversely, genes in the pathways of glycosaminoglycan degradation, RNA polymerase were significantly up-regulated in the low 23-gene score group of the TCGA cohort, with natural killer cell mediated cytotoxicity, cytokine cytokine receptor interaction, cell adhesion molecules cams, chemokine signalling pathway and systemic lupus erythematosus the top ranking pathways ($p < 0.05$ for all cases, supplementary Table 8). These results suggest that the survival of melanoma patients can be accurately predicted by the 23-gene score, perhaps the above-mentioned pathways might play a critical role in the association of 23-gene score with survival.

Comparisons of prognostic significance of the 23-gene score with established prognostic gene signatures

We compared the survival impact of the 23-gene expression with other established gene expression-based prognostic signatures to further evaluate the potential of the 23-gene expression as a prognostic marker for melanoma. We performed multivariate analysis of the 23-gene score, 8-gene risk score and immune-related genes (IRGs) score as well as prognosis-associated features in the TCGA cohort. The 23-gene score and 8-gene risk score remained significant prognostic factors independently of prognosis-associated features. Notably, the 23-gene score achieved a higher OR than the 8-gene risk score and IRGs score in the multivariate survival analysis (supplementary Table 9). Furthermore, the 23-gene score had higher AUC (0.76) than the 8-gene risk score and IRGs score (0.58 and 0.59, supplementary Fig. 6). Our data suggested the 23-gene score is superior to the two established prognostic gene signatures in the prediction of prognosis of melanoma patients.

Discussion

The American Joint Committee on Cancer melanoma Staging are widely applied to the evaluation of prognostic risk and guide treatment [20]. Somatic mutations in *NRAS* are associated with a higher frequency of nodal relapse and development of metastatic disease, *BRAF*-mutant melanoma patients showed a tendency for better overall survival [21]. In recent years, several gene expression profiles have been demonstrated to be potential prognostic biomarkers in melanoma. Sha et al developed a 8-gene risk

GL3, PVRL1, ATP1B1, CDAN1, *FAU*, and

TNFSF14 expression levels and found the risk score was effectively predictive of increased mortality rate in melanoma patients [10]. Stanley developed a two-gene signature consisting of *CCL8* and *DEFB1* (IRGs score) and a nomogram integrating the IRGs score, age and TNM stage that could effectively perform prognosis prediction in melanoma [11]. Despite significant advances in the risk classification of melanoma, the accuracies of these methods are still needed to be improved.

In this study, we performed Kaplan-Meier and multivariate analyses using the mRNA expression data of two independent datasets and found the 23 gene expression levels could predict the OS of melanoma patients. Furthermore, we established the 23-gene score using a linear combination of 23 gene expression levels and β-values from the LASSO model. The 23-gene score remained significantly associated with poor OS after adjustment of prognosticators. Moreover, we have demonstrated that the 23-gene score performed better than the 8-gene risk score and the IRGs score in the estimation of patient prognosis. The mechanisms underlying poor prognostic implication of higher 23-gene score in melanoma remain to be explored. The GESA analysis revealed the aminoacyl tRNA biosynthesis, adherens junction, lysine degradation, wnt signalling pathway and thyroid cancer were significantly up-regulated in the high 23-gene score group. Cell-cell interaction in skin homeostasis is tightly controlled by adherens junctions. Alterations in such regulation lead to melanoma development. Changes in cell-cell communications by somatic mutations in AJ cadherins function as one of mechanisms to trigger melanoma development [22]. We believe the adherens junctions and the wnt signalling pathway in part contribute to the prognostic importance of 23-gene score in melanoma.

Of the 23 genes, many genes play diverse roles in the tumorigeneses of cancers. For instance, the *TRIM14* encodes the TRIM14 protein which is a member of the tripartite motif family. The gene is abnormally expressed and shows different biological functions in cancers [23–25]. Upregulated expression of *TRIM14* was observed in breast cancer specimens and cell lines. Decreased *TRIM14* expression suppressed cell proliferation but induced cell apoptosis in the breast cancer cell lines [23]. *TRIM14* is upregulated in osteosarcoma specimens and cell lines, and correlated with osteosarcoma progression and shorter survival times. *TRIM14* enhances cell proliferation and invasion via activating the AKT signaling pathway in osteosarcoma [24]. The gene also promotes the migration and invasion of gastric cancer by increasing epithelial-to-mesenchymal transition through promotion of AKT signaling [25]. These results suggest that *TRIM14* may act as an oncogene in breast cancer, osteosarcoma and gastric cancer. Conversely, the gene was reported as a tumor suppressor in Non-Small Cell Lung Cancer, as evidenced by enhanced tumor growth in xenograft mouse models, resistance to hypoxia-induced cell death and decreased interferon response after *TRIM14* knock-down [26]. *PSMB8* encodes a member of the proteasome B-type family, that is a 20S core beta subunit. Silencing *PSMB8* expression inhibits glioma cell migration, proliferation, and induces apoptosis via regulating ERK1/2 and PI3K/AKT signaling pathways [27]. The inhibition of migration and invasion by PSMB8 was also shown in gastric cancer. Increased nuclear expression of *PSMB8* and PBK was correlated with invasion depth, lymph node metastasis and shorter survival time [28]. These results demonstrate *PSMB8* functions as a tumor suppressor in cancers.

Furthermore, the 23-gene expression signature effectively stratified melanoma patients into three subgroups with different survival probabilities. Given the 23-gene expression signature is independent of clinical parameters, the 23-gene expression signature may have prognostic values for the faction of melanomas without mutations in *BRAF* and *NRAS*. Lastly, of the 23 prognosis-associated genes, some genes may become druggable targets for melanoma patients. Take the *PSMB8* for example, knockdown of the gene enabled significant inhibition of cellular proliferation, invasion, indicating targeting the gene might make it possible for potential cure of melanoma patients.

Conclusion

In summary, this study identified a novel 23 gene expression signature that has prognostic values and effectively stratifies melanoma patients into subgroups of melanoma patients. The 23-gene score is superior to established gene-expression risk scores and indicative of an inferior prognosis in melanoma patients.

Abbreviations

The Cancer Genome Altas
TCGA
Gene Expression Omnibus
GEO
Overall survival
OS
Immune-related genes
IRGs
Receiver operating curves
ROC
Area under curve
AUC
Odd ratio
OR
Confidence interval
CI

Declarations

Acknowledgement

None

Authors' contributions

Dong Li conceived the study. Min Du and Ling Ma performed the survival analyses. Min Du and Fan Zhang conducted unsupervised hierarchical clustering analysis and development of risk score. Min Du wrote the manuscript. All authors read and approved the final manuscript.

Funding

None

Ethics approval and consent to participate

None

Consent for publication

None

Competing interests

The authors declare no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are available upon reasonable request.

References

1. Leonardi GC, Falzone L, Salemi R, Zanghì A, Spandidos DA, Mccubrey JA, et al. Cutaneous melanoma: From pathogenesis to therapy (Review). Int J Oncol. 2018;52:1071–80.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70:7–30.
3. Davies MA, Flaherty KT. Melanoma in 2017: Moving treatments earlier to move further forwards. Nat Rev Clin Oncol [Internet]. Nature Publishing Group; 2018;15:75–6. Available from: http://dx.doi.org/10.1038/nrclinonc.2017.183.
4. Reed KB, Brewer JD, Lohse CM, Bringe KE, Pruitt CN, Gibson LE. Increasing incidence of melanoma among young adults: an epidemiological study in Olmsted County, Minnesota. Mayo Clin Proc. 2012;87:328–34.
5. Robsahm TE, Bergva G, Hestvik UE, Møller B. Sex differences in rising trends of cutaneous malignant melanoma in Norway, 1954–2008. Melanoma Res England. 2013;23:70–8.
6. Zhang RD, Price JE, Schackert G, Itoh K, Fidler IJ. Malignant potential of cells isolated from lymph node or brain metastases of melanoma patients and implications for prognosis. Cancer Res United States. 1991;51:2029–35.
7. Jain BP. Prognostic model for primary melanoma. Ann. Intern. Med. United States; 1997. p. 832.

8. Harbst K, Staaf J, Lauss M, Karlsson A, Måsbäck A, Johansson I, et al. Molecular profiling reveals low- and high-grade forms of primary melanoma. Clin Cancer Res. 2012;18:4026–36.
9. Jönsson G, Busch C, Knappskog S, Geisler J, Miletic H, Ringnér M, et al. Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. Clin cancer Res an Off J Am Assoc Cancer Res United States. 2010;16:3356–67.
10. Wang J, Kong PF, Wang HY, Song D, Wu WQ, Zhou HC, et al. Identification of a Gene-Related Risk Signature in Melanoma Patients Using Bioinformatic Profiling. J Oncol. 2020;2020.
11. Liao M, Zeng F, Li Y, Gao Q, Yin M, Deng G, et al. A novel predictive model incorporating immune-related gene signatures for overall survival in melanoma patients. Sci Rep [Internet]. Nature Publishing Group UK; 2020;10:1–12. Available from: https://doi.org/10.1038/s41598-020-69330-2.
12. Tcga. Genomic Classification of Cutaneous Melanoma. Cell [Internet]. 2015;161:1681–96. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0092867415006340.
13. Budden T, Davey RJ, Vilain RE, Ashton KA, Braye SG, Beveridge NJ, et al. Repair of UVB-induced DNA damage is reduced in melanoma due to low XPC and global genome repair. Oncotarget. 2016;7:60940–53.
14. Therneau T. Survival Analysis. Cran [Internet]. 2016; Available from: https://cran.r-project.org/web/packages/survival/survival.pdf.
15. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.2264&rep=rep1&type=pdf.
16. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics [Internet]. BioMed Central Ltd; 2011;12:77. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-77.
17. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw. 2011;39:1–13.
18. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R programming tools for plotting data. R Packag. version. 2005.
19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci [Internet]. 2005;102:15545–50. Available from: http://www.pnas.org/content/102/43/15545.abstract.
20. Dickson PV, Gershenwald JE. Staging and prognosis of cutaneous melanoma. Surg Oncol Clin N Am. 2011;20:1–17.
21. Heppt MV, Siepmann T, Engel J, Schubert-Fritschle G, Eckel R, Mirlach L, et al. Prognostic significance of BRAF and NRAS mutations in melanoma: a German study from routine care. BMC Cancer BMC Cancer. 2017;17:536.
22. Korla PK, Chen C-C, Gracilla DE, Lai M-T, Chen C-M, Chen HY, et al. Somatic mutational landscapes of adherens junctions and their functional consequences in cutaneous melanoma development.

23. Hu G, Pen W, Wang M. TRIM14 Promotes Breast Cancer Cell Proliferation by Inhibiting Apoptosis. Oncol Res. 2019;27:439–47.
24. Xu G, Guo Y, Xu D, Wang Y, Shen Y, Wang F, et al. TRIM14 regulates cell proliferation and invasion in osteosarcoma via promotion of the AKT signaling pathway. Sci Rep. 2017;7:42411.
25. Wang F, Ruan L, Yang J, Zhao Q, Wei W. TRIM14 promotes the migration and invasion of gastric cancer by regulating epithelial–to–mesenchymal transition via activation of AKT signaling regulated by miR–195–5p. Oncol Rep. 2018;40:3273–84.
26. Hai J, Zhu C-Q, Wang T, Organ SL, Shepherd FA, Tsao M-S. TRIM14 is a Putative Tumor Suppressor and Regulator of Innate Immune Response in Non-Small Cell Lung Cancer. Sci Rep. 2017;7:39692.
27. Yang B-Y, Song J-W, Sun H-Z, Xing J-C, Yang Z-H, Wei C-Y, et al. PSMB8 regulates glioma cell migration, proliferation, and apoptosis through modulating ERK1/2 and PI3K/AKT signaling pathways. Biomed Pharmacother France. 2018;100:205–12.
28. Kwon CH, Park HJ, Choi YR, Kim A, Kim HW, Choi JH, et al. PSMB8 and PBK as potential gastric cancer subtype-specific biomarkers associated with prognosis. Oncotarget. 2016;7:21454–68.

Tables

Due to technical limitations, table 1 and 2 is only available as a download in the Supplemental Files section.

Figures

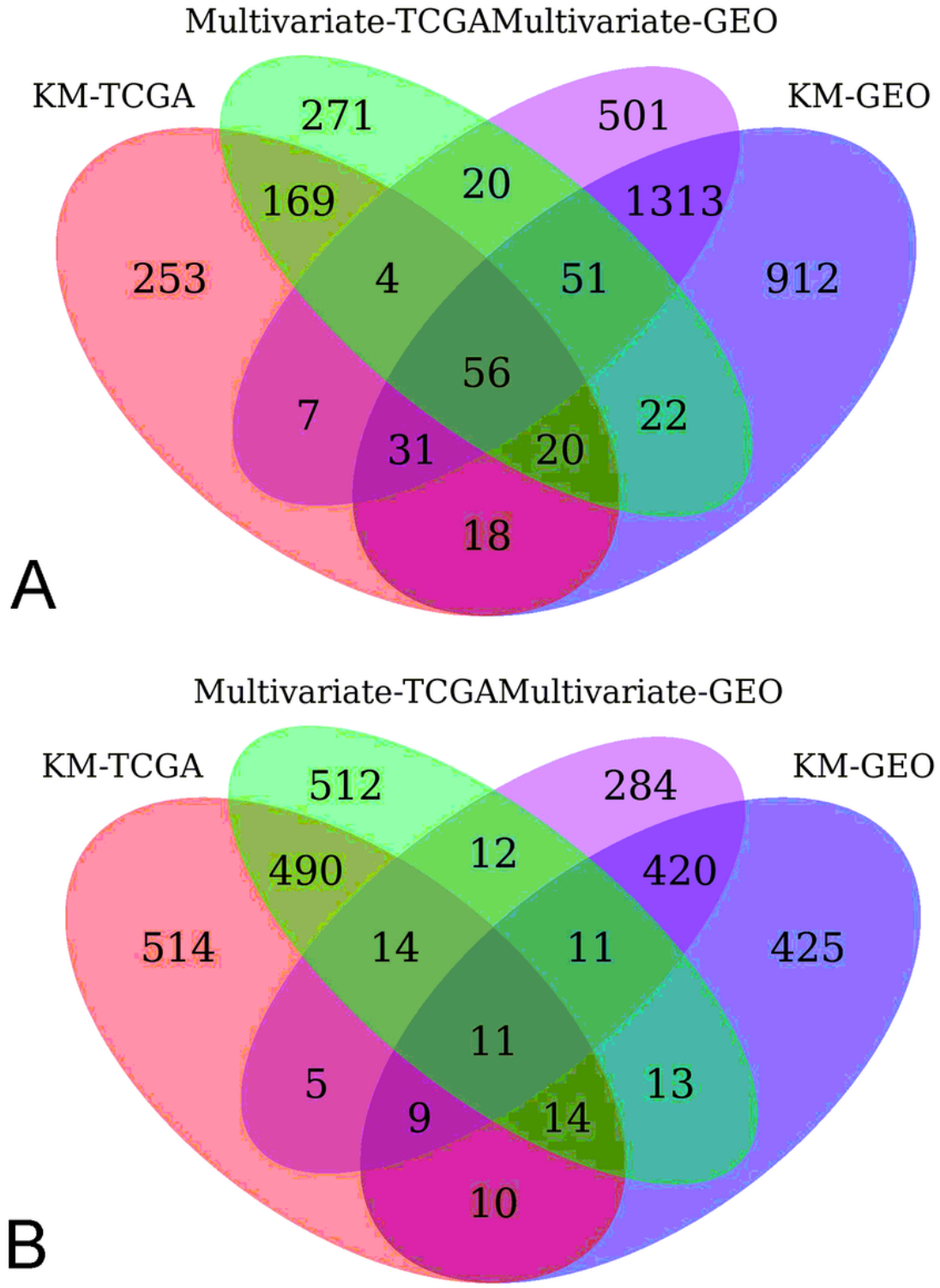


Figure 1

The overlap of prognosis-associated genes between GEO and TCGA datasets. A. The overlap of protective type genes between GEO and TCGA datasets, B. The overlap of risk type genes between GEO and TCGA datasets.

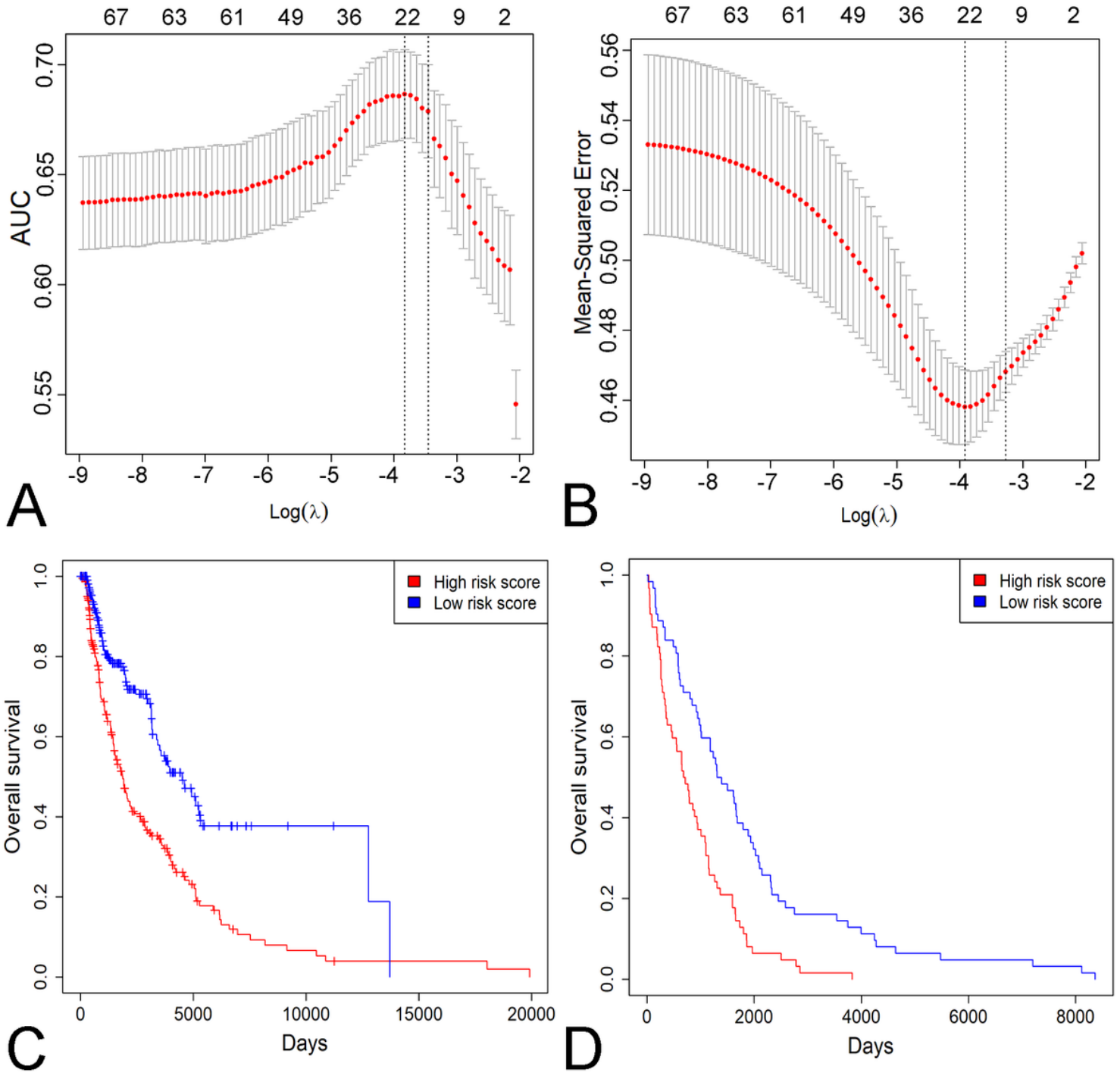
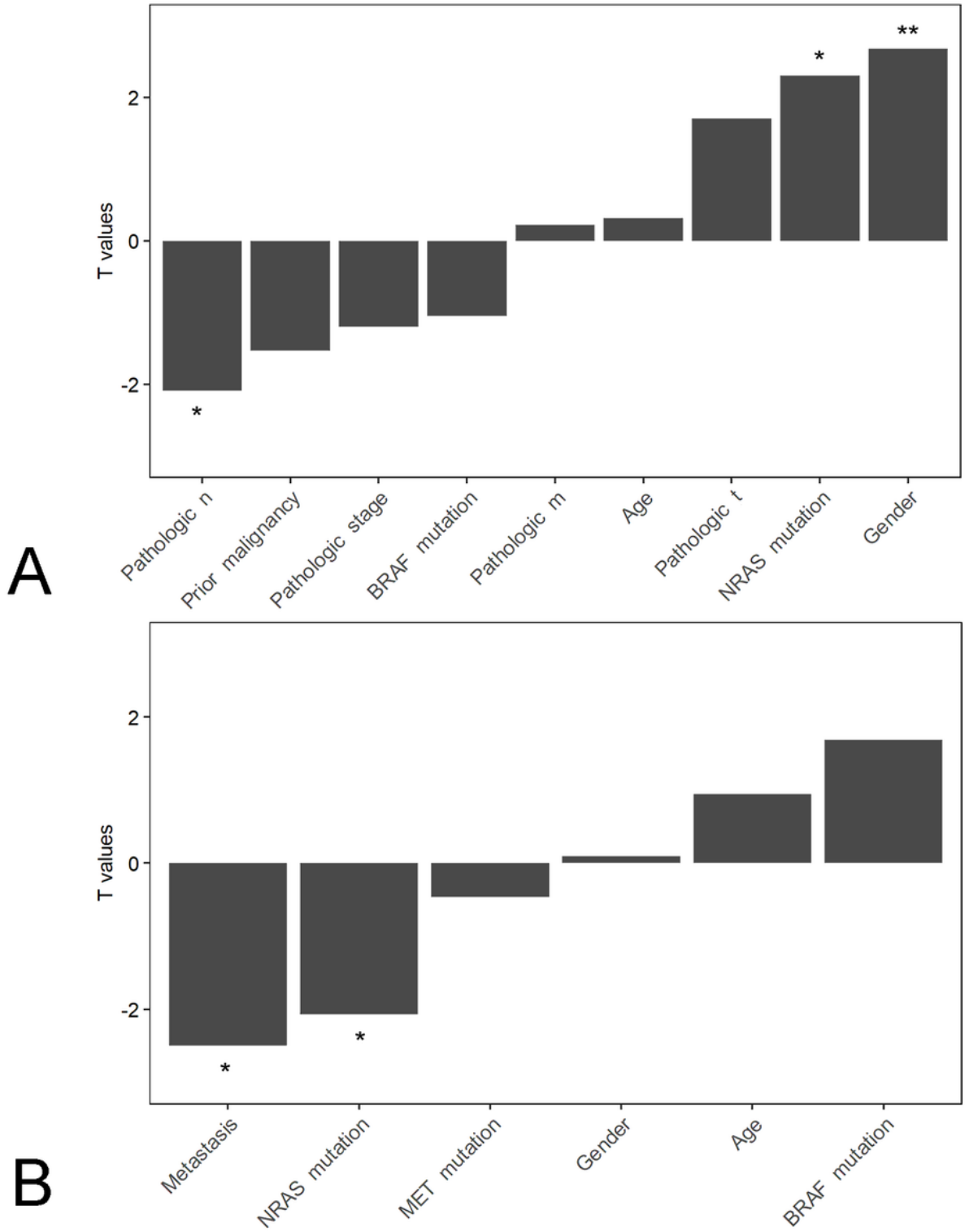


Figure 2

The 23-gene score is negative prognostic biomarker in melanoma. A. The relationship between AUC and log scaled lambda values and number of genes with non-zero coefficients in the LASSO model. The x and y labels denoted log scaled lambda values and AUC respectively. The numbers on the top were the number of genes with non-zero coefficients reserved in the LASSO model. The left and right vertical dotted lines indicated the lambda.min and lambda.1se for λ respectively. The former is the one which minimizes out-of-sample loss in cross validation. The latter is the one which is the largest lambda value

within 1 standard error of the minimum. B. The relationship between mean squared error and log scaled lambda values and number of genes in the LASSO model. C. Kaplan-Meier survival analysis of patients' OS with the 23-gene score in the GEO cohort, D. Kaplan-Meier survival analysis of patients' OS with the 23-gene score in the TCGA dataset. E. The difference of the 23-gene scores between deceased and living melanoma patients in the GEO and TCGA cohorts. F. The ROC curves of the 23-gene scores in the GEO and TCGA datasets.



The associations of clinical characteristics with the 23-gene score. A. The associations between clinical characteristics with the 23-gene score in the GEO cohort. B. The associations between clinical characteristics with the 23-gene score in the TCGA cohort. Of note, *, ** and *** represent P value < 0.05, <0.01 and 0.001 respectively.

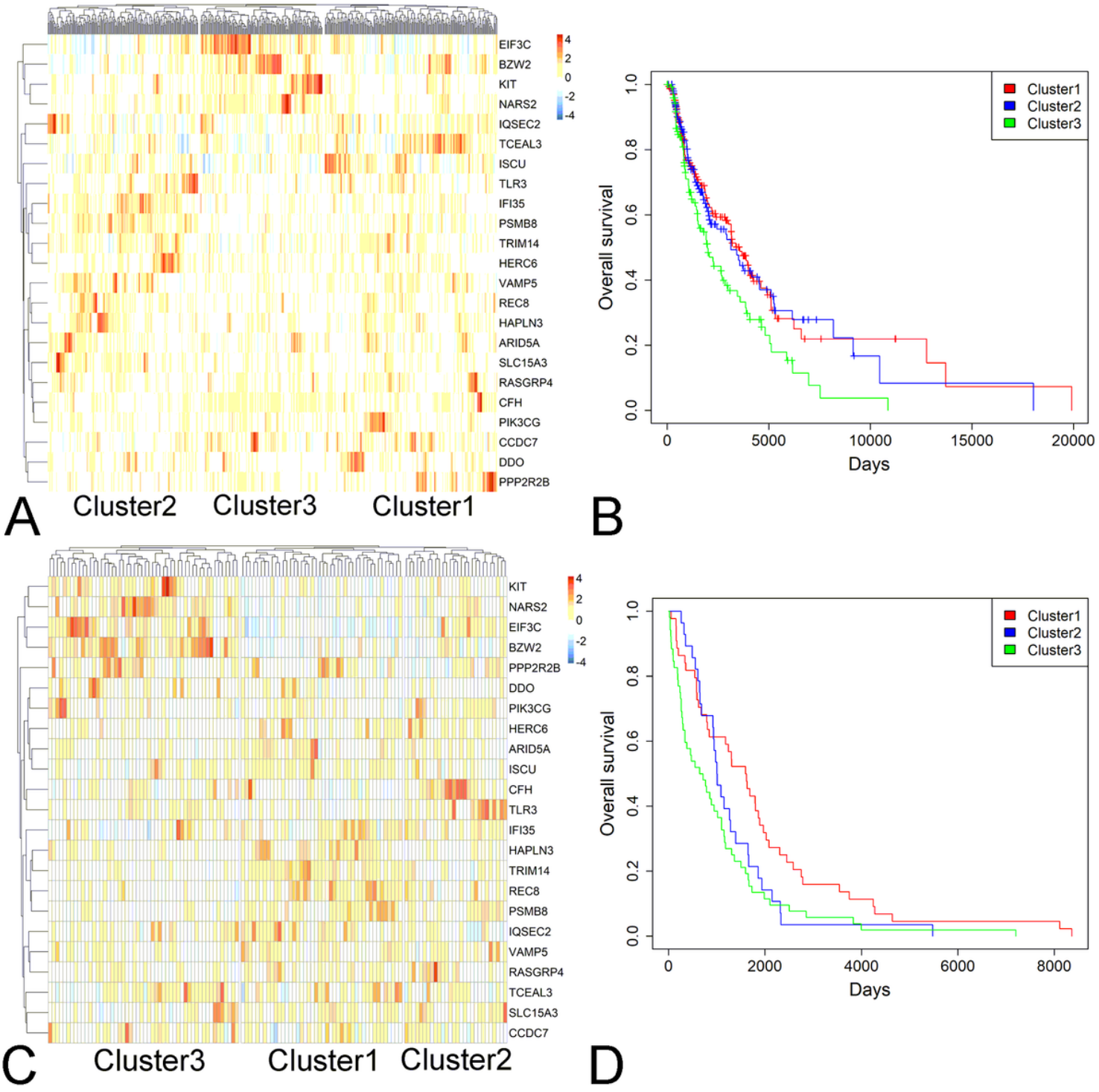


Figure 4

Unsupervised hierarchical clustering of the 23 gene panel uncovered three classes of melanoma patients. A. Unsupervised hierarchical clustering of the 23 gene panel uncovered three classes of melanoma

patients in the GEO dataset. B. The difference in overall survival between the three subgroups of melanoma patients in the GEO cohort. C. Unsupervised hierarchical clustering of the 23 gene panel uncovered three classes of melanoma patients in the TCGA dataset. D. The difference in overall survival between the three subgroups of melanoma patients in the TCGA cohort.

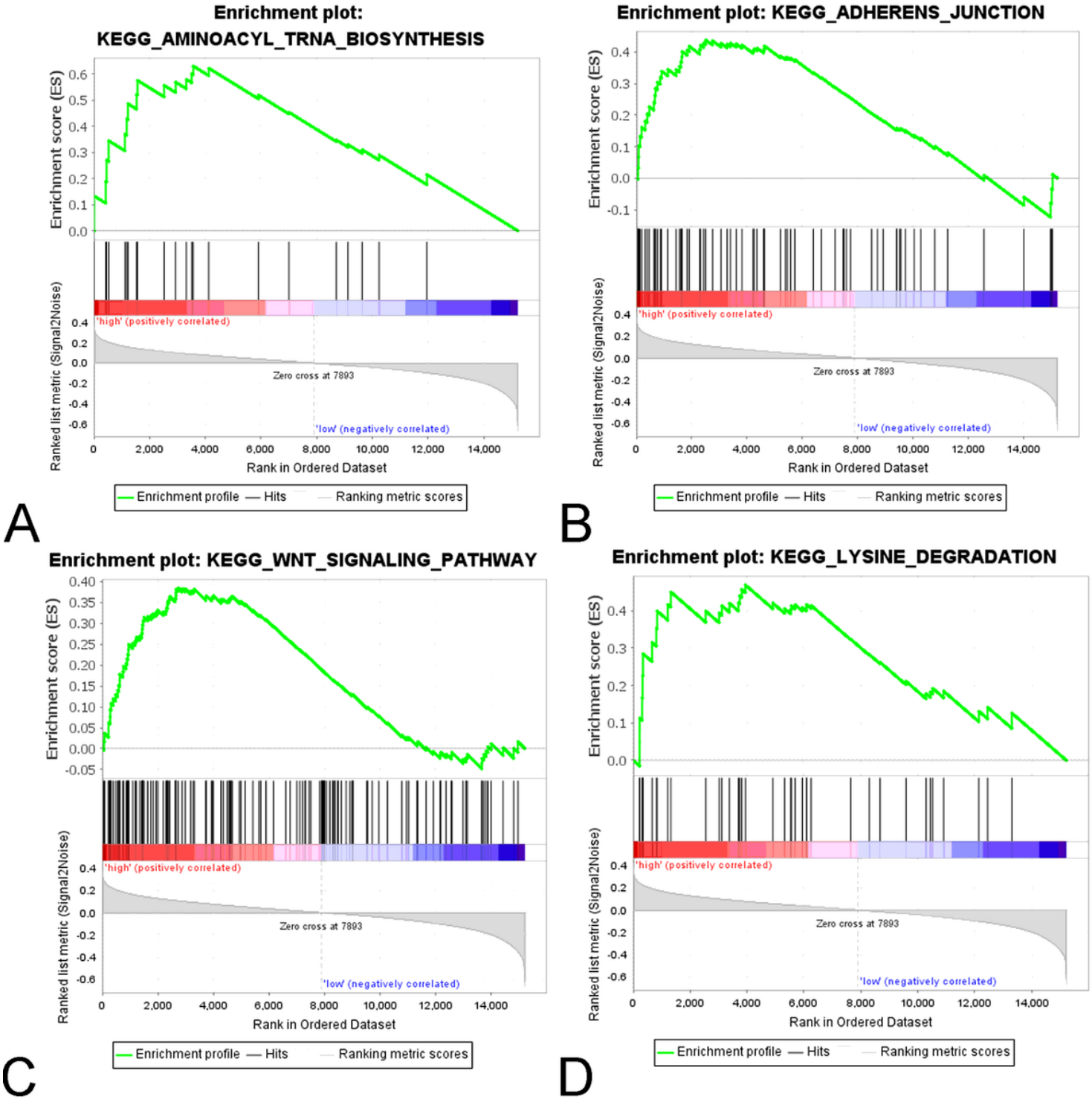


Figure 5

GSEA based on the expression of GEO dataset revealed significant pathways in the high 23-gene score group, including long term depression (A), glycerolipid metabolism(B), VEGF signalling pathway(C),

phosphatidylinositol signalling system (D) and gap junction (E). For each gene set, vertical bars along the x-axis of the GSEA plot represent the positions of genes within the ranked list. Negative GSEA enrichment score curve indicates anti-enrichment (down-regulation), and positive curve denotes enrichment (up-regulation).

Supplementary Files

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

- Tables.pdf
- STables.docx
- SupplementaryFigure1.tiff
- supplementaryFigure2.tif
- supplementaryFigure3.tif
- supplementaryFigure4.tif
- supplementaryFigure5.tif
- supplementaryFigure6.tiff