JBUON 2019; 24(5): 2161-2167
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
Email: editorial_office@jbuon.com

ORIGINAL ARTICLE

A four-gene expression-based signature predicts the clinical outcome of melanoma

Liping Sun[1], Ping Li[1], He Ren[1], Gang Liu[1], Lining Sun[2]

[1]Shanghai University of Medicine & Health Sciences, Pudong New Area, Shanghai, P.R.China; [2]School of Mechanical and Electric Engineering, Soochow University, Suzhou, China

Summary

Purpose: *Although clinical indicators provide effective prognostic information, the prognosis of melanoma is difficult due to its genomic and biological complexity. Our goal was to elucidate the impact of genes on survival.*

Methods: *Public cohorts of melanoma gene expression and machine learning were used to develop a model for prognosis. A four-gene model was developed to predict the clinical outcome of melanoma in TCGA datasets. The performance was further validated in four independent cohorts. The relationship between clinical indicators and melanoma score was assayed and the correlated pathways were identified.*

Results: *The samples with high melanoma scores had a significantly better survival rate than those with low melanoma scores in the training cohort. This observation was confirmed in four independent cohorts, GSE22138, GSE54467, GSE65904 and E-MTAB-4725. In addition, the melanoma score was independent of most clinically used indicators. Cox univariate regression showed that the melanoma score was significantly associated with survival. Multiple significantly enriched pathways were identified between the high-score and low-score groups.*

Conclusion: *The melanoma score model was robust and effective for melanoma prognosis.*

Key words: *gene expression, prognosis, model, melanoma*

Introduction

Melanoma is among the most prevalent cancers worldwide due to its rapid progression. According to cancer statistics in the U.S. in 2015, there were an estimated 73,870 new melanoma cases and 9,940 related deaths [1, 2]. Classical prognostic factors, including age and AJCC (American Joint Committee on Cancer) stage, proved to be effective indicators for melanoma [3-5]. Melanoma-specific indicators, including Clark level [6], were also reported to indicate prognosis. However, these clinicopathological indicators reflect the prognosis of melanoma based on cancer behavior instead of cancer molecular subtypes. Thus, it is reasonable these indicators often failed to predict clinical outcome.

To investigate the prognosis of melanoma at the molecular level, efforts have been devoted to find biomarkers on multiple biological levels. For example, HIF expression was shown to stimulate melanoma cell migration, and its high expression level was associated with poor prognosis [7]. Methylation of TNFRSF10D was reported to predict clinical outcome in terms of both overall survival and recurrence-free survival rates [8]. Copy number variation of the interferon cluster was associated with T cell infiltration and thus affected overall survival [9]. The expression level of miRNA-203 was also reported as an indicator of prognosis [10]. MIA, S100 and LDH were indicated to impact on the survival of melanoma[11]. Similarly, siRNA

Corresponding author: Liping Sun, PhD. Shanghai University of Medicinal & Health Sciences, Room 503, Building 18, No.257 Tianxiong Rd, Pudong New Area, Shanghai, P.R.China.
Tel: + 86 021 33755214, Email: sunlp@sumhs.edu.cn
Received: 08/12/2018; Accepted: 21/01/2019

knock down of Notch1 suppresses melanoma cell proliferation[12]. Despite these findings, none of these biomarkers are used clinically. One of the most important reasons is that the performance of these biomarkers is unstable across cohorts due to the heterogeneity of melanoma at the genetic level. Currently, optimized panels of gene expression based-models are emphasized due to their robustness across datasets [13-16]. For example, Mamaprint and OncotypeDX have been used clinically because of their high performance [17-19].

The purpose of this study was to develop a model for melanoma by utilizing the gene expression data and assay the robustness of the model across cohorts.

Methods

Sample enrollment and data preprocessing

Samples that were not primary melanoma were excluded from this study, as were samples without complete follow-up information (n=2). The expression matrix of the TCGA cohort was downloaded from the TCGA website (https://cancergenome.nih.gov). The RPKM values were normalized using the upper-quantile method [20], and the relative expression values were log 2 transformed. The normalized GSE22138 and GSE54467 data were downloaded from the GEO website (https://www.ncbi.nlm.nih.gov/geo) using the function affy::rma. To eliminate the batch effect and platform differences, expression values were transformed to z-scores for further analysis. If a single gene had more than one unique probe, the average value was calculated as the expression level.

Feature selection and model development

In the TCGA dataset, correlation analysis between survival and gene expression information was implemented using Cox univariate regression, and genes identified as significantly ($p<0.05$) associated with overall survival were selected as candidate features. The entire panel of combinations of these features was tested to develop a multivariate regression model. The model was calculated as $MS=\sum_{i}^{n}$, where b is the coefficient, and x is the relative gene expression. The survival difference between the low-score and high-score subgroups was compared using the median melanoma score as the cutoff. The panel with the smallest p value was selected as the optimized panel.

Data analysis

All data analyses were performed on the R platform (www.r-project.org) (v3.2.0) and R packages. The ROC curve was plotted using the R package "pROC" [21] (v1.8). The two-year survival nomogram was calculated with the "rms" package (v4.5.0). The Cox univariate regression and multivariate regression were carried out with "survival" package (v2.39.4). Gene Set Enrichment Analysis was performed on Java software [22] (http://software.broadinstitute.org/gsea) by labeling the high-melanoma score and low-melanoma score groups. The expression matrix was normalized in previous steps.

Results

Feature selection, panel optimization, and model development

A gene expression matrix for primary melanoma in the TCGA cohort (n=103) was used for feature selection after data cleaning and normalization. Genes significantly ($p<0.05$) correlated with overall survival of patients with melanoma were selected as features for model development according to the Cox univariate regression between gene expression level and overall survival. Twenty genes were identified. All combinations of genes identified were generated, and Cox multivariate regression was implemented for each combination (panel). The melanoma score for each combination was evaluated. For each panel, the melanoma scores were divided using the median melanoma score as the cutoff, and the survival difference of patients with a high or low melanoma score was evaluated using Kaplan-Meier survival analysis. The panel generating the smallest p value was selected as the optimized panel. A combination of four genes was identified (Table 1). The melanoma score was calculated according to the following formula:

Table 1. Univariate and multivariate regression of genes used for the melanoma score

Genes	*Univariate*			*Multivariate*		
	HR	*95%CI*	*p value*	*HR*	*95%CI*	*p value*
RHBDL3	1.7	1.2-2.5	0.00536	1.55	0.99-2.43	0.0535
GPR64	1.5	1.0-2.3	0.03036	1.13	0.68-1.89	0.6348
ANKRD30A	1.4	1.1-1.7	0.01152	1.34	1.05-1.70	0.01654
PRKCD	0.66	0.46-0.93	0.01803	0.76	0.52-1.11	0.1565

Melanoma score= (-0.277×PRKCD) + (0.293× ANKRD30A) + (0.123×GPR64) +(0.441×RHBDL3)

In the formula, the gene symbol indicates the relative mRNA level of each gene. The negative coefficient indicates that this gene is negatively correlated with survival.

Model performance in training dataset

The prognostic value of the melanoma score was evaluated by comparing the survival difference between high-score and low-score melanoma samples using the median melanoma score value as the cutoff in the TCGA cohort. The overall survival time of melanoma patients with a high melanoma score was significantly longer than that of those with a low melanoma score (Figure 1A). The disease-free survival difference of these two subgroups was also compared. As expected, the disease-free survival period for the high melanoma score group was significantly longer than that of the low melanoma score group (Figure 1B). We observed that tumor suppressor genes were highly expressed in the high melanoma score group, while oncogenes had low expression (Figure 1C). The two-year survival area under the receiving operating characteristic (AUROC) was calculated to evaluate the prognostic value of the melanoma score and other clinical indicators (Figure 1D). The two-year survival AUROC of the melanoma score, depth to skin, gender, primary tumor stage and Clark level was 0.715, 0.618, 0.571, 0.650, and 0.543, respectively. Collectively, these results indicate that the melanoma score is an important indicator of both overall survival and disease-free survival.

Validation of melanoma score

The good performance of the melanoma score may result from over-fitness between the model and the training (TCGA) dataset. To exclude this possibility, we evaluated the performance of the

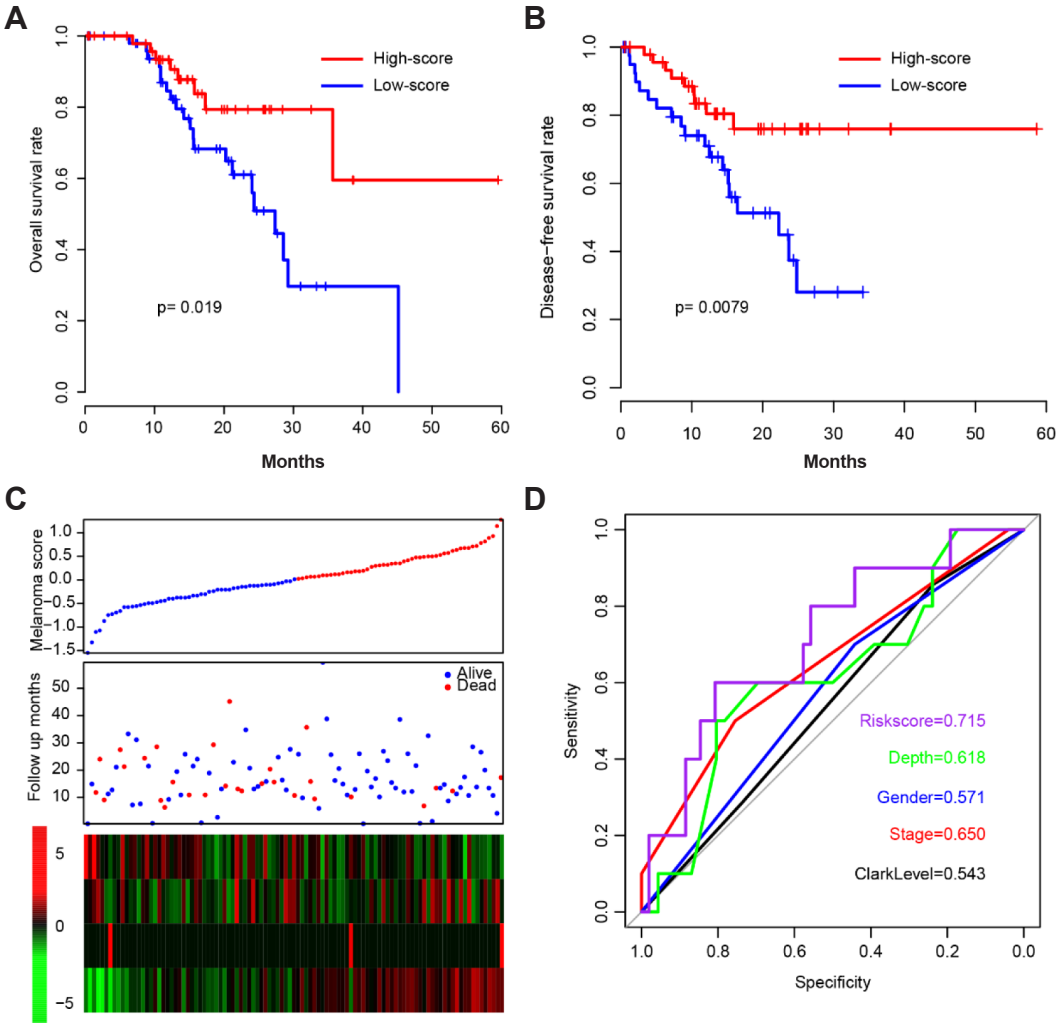

Figure 1. The performance of the melanoma score in the TCGA dataset. The overall survival **(A)** and disease-free survival **(B)** difference between the high-score and low-score groups. Detailed survival information is shown in **(C)**. The blue dots in the upper panel are the low melanoma score samples, and the red dots are the high melanoma score samples. **(D):** The two-year survival ROC of the melanoma score and clinical indicators.

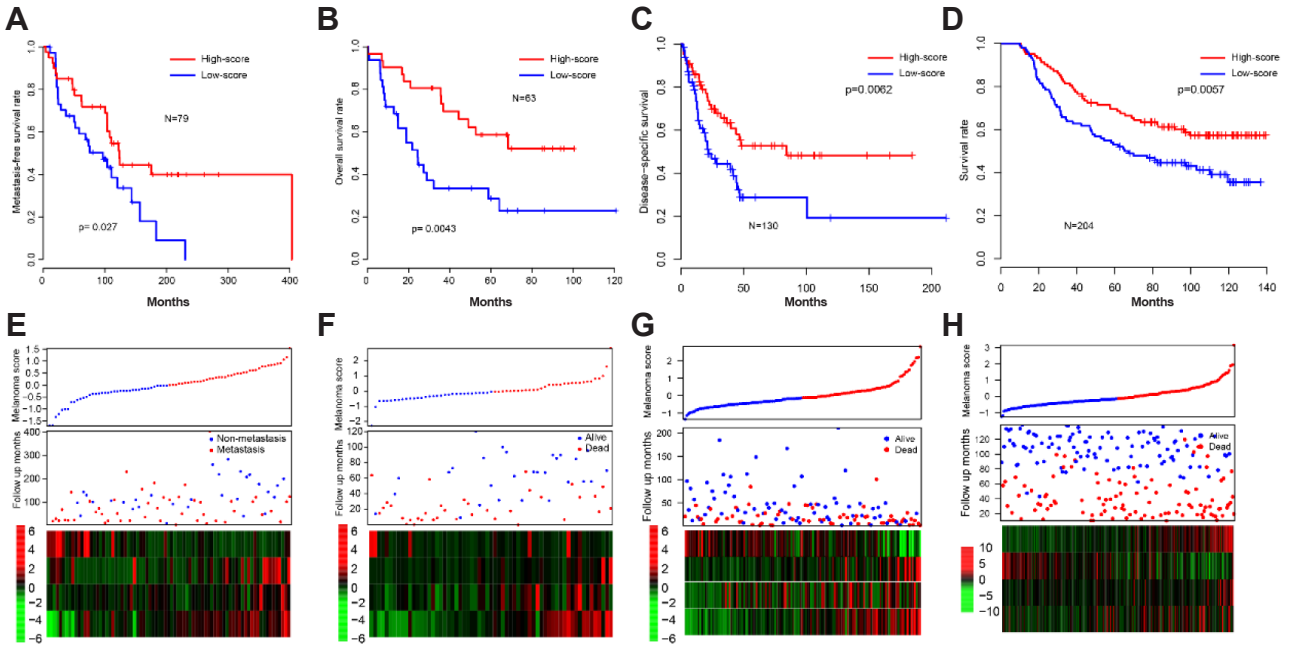

Figure 2. Prognostic value validation of the melanoma score. The survival difference of samples with high and low melanoma scores in GSE22138 **(A)** and GSE54467 **(B)**. The detailed melanoma score, survival information and gene expression of GSE22138 **(C)** and GSE54467 **(D)** are also shown. In addition, another two independent cohorts, GSE65904 **(E,G)** and E-MTAB4725 were also assayed **(F,H)**.

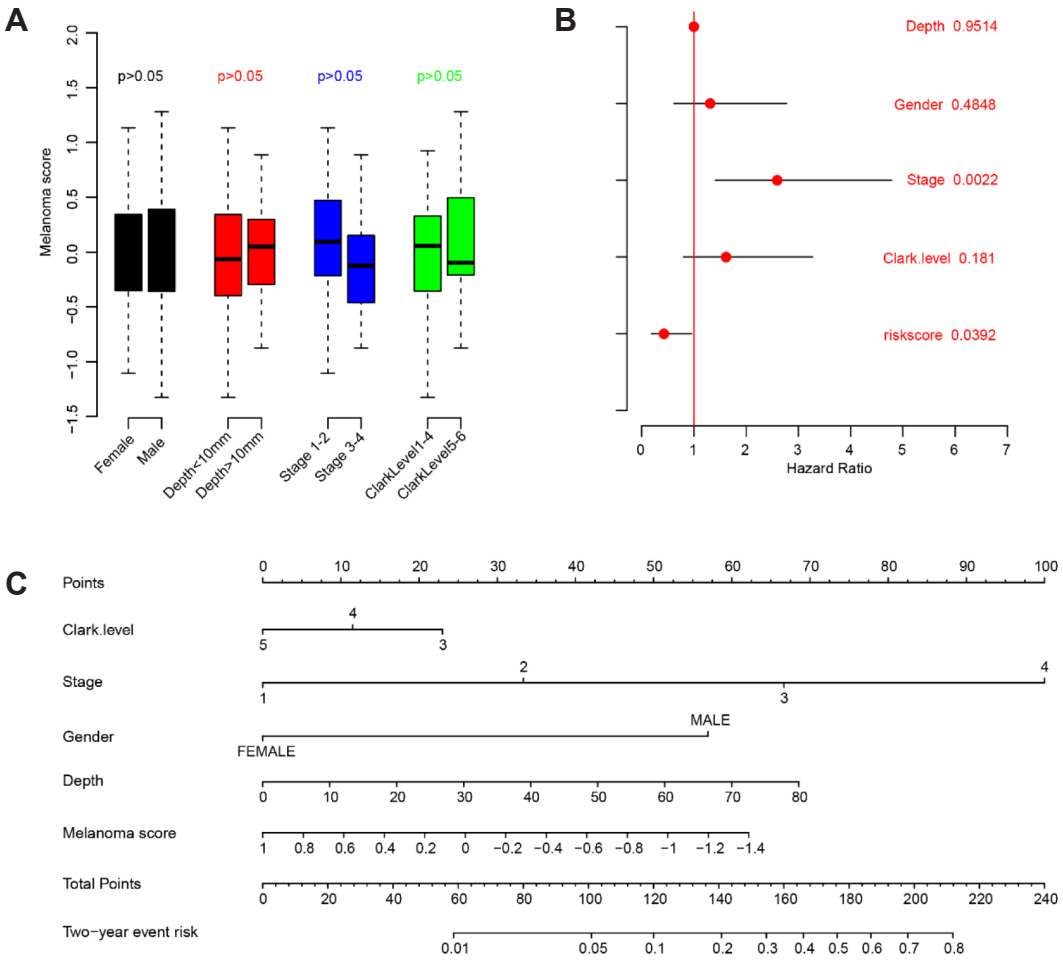

Figure 3. Melanoma score and clinical indicators. The melanoma score is independent from clinical indicators **(A)** and significantly contributes to the survival of patients with melanoma **(B)** according to the Cox regression. The line indicates the 95% confidence interval, and the red dots indicate the hazard ratio. The nomogram of two-year survival is shown **(C)**.

melanoma score in two other cohorts, GSE22138 and GSE54467, to evaluate the robustness of the model. After locking the coefficient of each gene, the melanoma score was calculated for each sample in both datasets. The samples in both datasets were divided into high-score and low-score groups using the median melanoma score as the cutoff, as we did in the TCGA cohort. As expected, the metastasis-free survival of GSE22138 and overall survival of GSE54467 were significantly better in the high-score group than in the low-score group (Figures 2A-B). We also observed that the samples with a high melanoma score tended to have high expression of tumor suppressor genes, low expression of oncogenes and early death/metastasis (Figures 2C-D), which is consistent with the training dataset. We also validated the performance in two other independent cohorts, GSE65904 and E-MTAB-4725 (Figures 2E-H, p=0.0062 and 0.0057, respectively). Taken together, the results indicate that the performance of the melanoma score was reproducible across cohorts.

Melanoma score and clinicopathological indicators

We evaluated the correlation between the melanoma score and clinical observations. As shown in Figure 3A, the melanoma score was not significantly associated with clinical observations, including gender, depth, primary tumor stage, and Clark level. We implemented univariate regression between clinical indicators and overall survival. We found that the melanoma score was significantly associated with overall survival, while the other clinical observations were not, except for primary tumor stage (Figure 3B). To facilitate the utilization of the melanoma score along with other clinico-

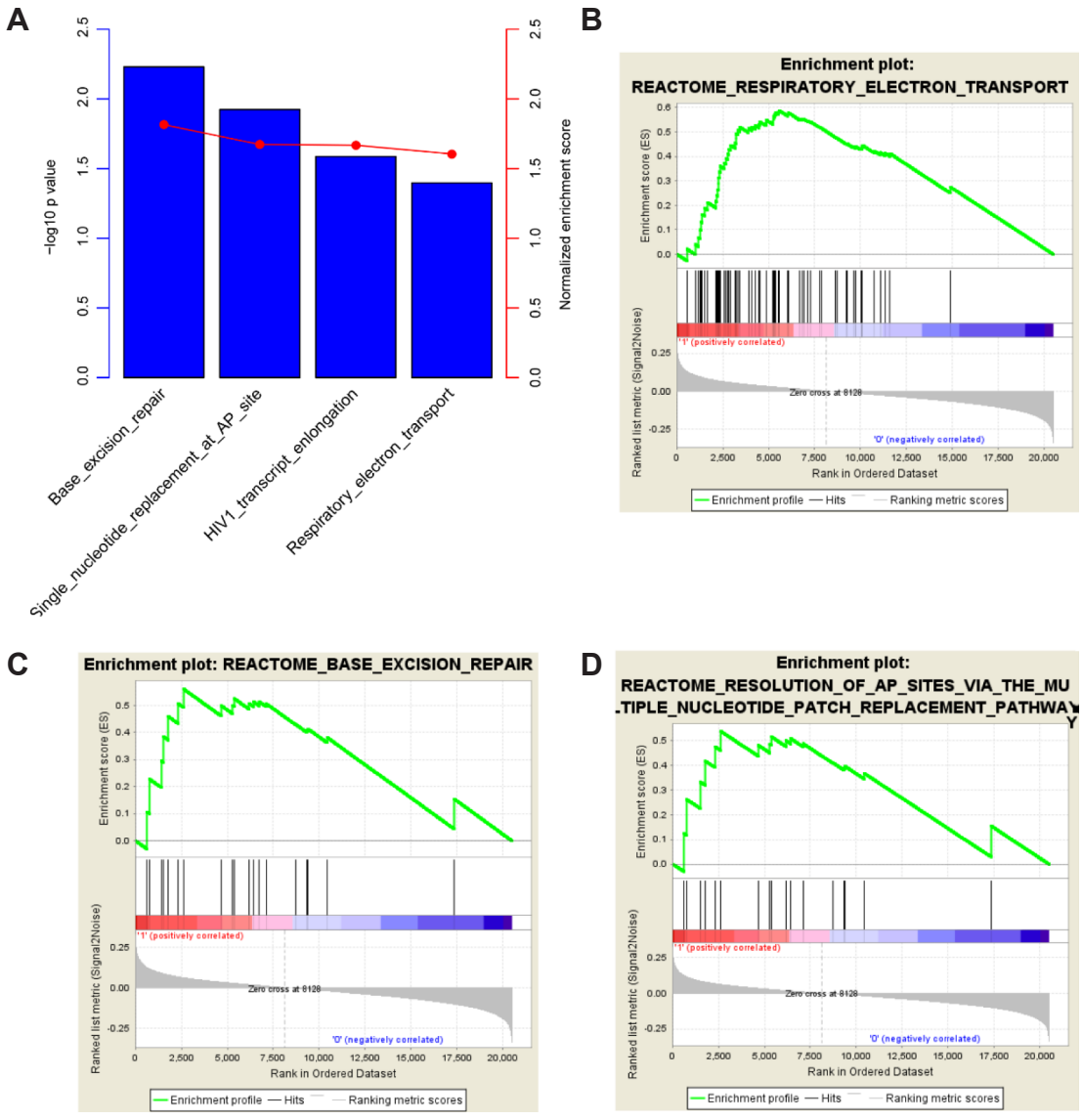

Figure 4. Pathways associated with the melanoma score. The pathways significantly associated with the melanoma score were identified with GSEA **(A)**. Several melanoma-related pathways are noted **(B-D)**.

pathological indicators, a two-year event (overall survival) nomogram was plotted and showed that the melanoma score is one of the most important indicators for melanoma. Collectively, these results suggest that the melanoma score is an important and independent clinical indicator of survival.

Pathways associated with the melanoma score

The melanoma score was developed based on the expression of four genes. Together with the fact that the score is independent of clinical indicators, this suggests that the melanoma score predicts the clinical outcome of melanoma by reflecting the biological heterogeneity of melanoma. To gain insight into the melanoma score, significantly different pathways were identified using Gene Set Enrichment Analysis (GSEA) by comparing the high-score and low-score subgroups in the TCGA dataset. Pathways including base excision repair, single nucleotide replacement at AP site, and respiratory transport were identified (Figures 4A-D). We found that the melanoma score reflects the cell response to mutation and metabolism. Taken together, these results suggest that the melanoma score predicts the clinical outcome of melanoma as a reflection of multiple biological conditions in melanoma cells.

Discussion

The prognosis of melanoma is complex, due to the fact that it is influenced by a lot of reasons, including genetic heterogeneity during the development and progression of melanoma. Another important clinical indicator of melanoma is the clinical stage. The third critical factor for prognosis is the treatment methods. Among the aspects that influence the clinical outcome of melanoma, treatment method is controllable, and clinical stage could be estimated using the clinical indicators, while methods to evaluate the genetic heterogeneity of melanoma is immature.

Although clinicopathological indicators provide information regarding cancer development levels but poorly reflect biological heterogeneity. Thus, they are insufficient to predict the clinical outcome of melanoma. Although a single biomarker based on genomic alterations and expression can provide insight into carcinogenesis and development, the heterogeneity of melanoma weakens the utilization of a single indicator. One of the clues is that none of the genes was significantly associated with survival in all datasets assayed in this study. Multiple genes-based model predicts the survival of melanoma using genetic information and overcomes the genetic heterogeneity to some extent. Therefore, a multiple genes-based model is currently desired, and the multiple genes-based models have been developed in other cancer types [14,23,24]. By integrating clinical information and gene-based model predicted the prognosis more precisely [19], and some models were assayed to be effective for guiding adjuvant therapy [17]. In this study, we developed a melanoma score model based on the expression of four genes, and the model proved to be robust across datasets. Notably, the model reflected multiple biological conditions of melanoma.

Among the candidate genes, PRKCD was shown to alter the metastasis and progression of ovarian cancer [25]. In addition, expression of PRKCD was associated with radiation resistance in cervical cancer [26]. Genome-wide association studies revealed that ANKRD30A is a susceptibility gene for triple-negative breast cancer, ovarian cancer and sarcoma for both carcinogenesis and development[27, 28]. GPR64, a G-protein-coupled receptor, was reported to promote both cancer metastasis and invasion [29, 30]. The clinical significance of the candidate genes suggests that the genes used for our model development are biologically functional. The biological and clinical functions of these candidate genes in melanoma and other cancer types indicate that these genes selected were rational. Thus, the model successfully predicted the clinical outcome of melanoma. It is noticeable that although the model performs well in prognosis, none of the genes was significantly associated with survival in all the cohorts assayed, which is consistent with the fact that the multiple genes based model excelled the single gene model.

One of the most important limitations of this study is that detailed information is missing, including time to recurrence/metastasis and treatment method. Thus, the performance of the melanoma score could not be evaluated in subgroups. Another limitation is that the melanoma score was calculated with z-score-transformed gene expression data. Thus pooled data are needed.

In conclusion, the model we developed using the expression of four genes is robust in predicting the survival of melanoma across cohorts. It could be used as an independent clinical indicator for melanoma prognosis.

Acknowledgement

National Key R&D Program of China (2018YFB1307700) to Liping Sun.

Conflict of interests

The authors declare no conflict of interests.

References

1. Chen W, Zheng R, Baade PD et al. Cancer statistics in China, 2015. CA: Cancer J Clin 2016;66:115-32.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA: Cancer J Clin 2015;65:5-29.
3. Shields CL, Kaliki S, Furuta M, Mashayekhi A, Shields JA. Clinical spectrum and prognosis of uveal melanoma based on age at presentation in 8,033 cases. Retina (Philadelphia, Pa.) 2012;32: 1363-72.
4. Shields CL, Kaliki S, Furuta M, Fulco E, Alarcon C, Shields JA. American Joint Committee on Cancer Classification of Uveal Melanoma (Anatomic Stage) Predicts Prognosis in 7,731 Patients: The 2013 Zimmerman Lecture. Ophthalmology 2015;122:1180-6.
5. Balch CM, Soong SJ, Gershenwald JE et al. Age as a prognostic factor in patients with localized melanoma and regional metastases. Ann Surg Oncol 2013;20:3961-8.
6. Morton DL, Davtyan DG, Wanek LA, Foshag LJ, Cochran AJ. Multivariate analysis of the relationship between survival and the microstage of primary melanoma by Clark level and Breslow thickness. Cancer 1993:71:3737-43.
7. Guo H, Cheng Y, Martinka M, Mcelwee K. High LIFr expression stimulates melanoma cell migration and is associated with unfavorable prognosis in melanoma. Oncotarget 2015;6:25484-98.
8. Ratzinger G, Mitteregger S, Wolf B et al. Association of TNFRSF10D DNA-methylation with the survival of melanoma patients. Int J Molec Sci 2014;15:11984-95.
9. Linsley PS, Speake C, Whalen E, Chaussabel D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. PLoS One 2014;9: e109760.
10. Wang K, Zhang ZW. Expression of miR-203 is decreased and associated with the prognosis of melanoma patients. Int J Clin Exper Pathol 2015;8:13249-54.
11. Nikolin B, Djan I, Trifunovic J et al. MIA, S100 and LDH as important predictors of overall survival of patients with stage IIb and IIc melanoma. JBUON 2016;21:691-7.
12. Yang Z, Qi Y, Lu C, Zhang J, Luo R, Kang S. Small interfering RNA (siRNA)-mediated knockdown of Notch1 suppresses tumor growth and enhances the effect of IL-2 immunotherapy in malignant melanoma. JBUON 2015;20:1553-64.
13. Wu X, Weng L, Li X et al. Identification of a 4-microRNA signature for clear cell renal cell carcinoma metastasis and prognosis. PLoS One 2012;7:e35661.
14. Bou Samra E, Klein B, Commes T, Moreaux J. Identification of a 20-gene expression-based risk score as a predictor of clinical outcome in chronic lymphocytic leukemia patients. BioMed Res Int 2014;423174.
15. Liu Q, Diao R, Feng G, Mu X, Li A. Risk score based on three mRNA expression predicts the survival of bladder cancer. Oncotarget 2017;8:61583-91.
16. Zhang ZL, Zhao LJ, Chai L et al. Seven LncRNA-mRNA based risk score predicts the survival of head and neck squamous cell carcinoma. Sci Rep 2017;7:309.
17. Cardoso F, Van't Veer LJ, Bogaerts J et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. New Engl J Med 2016;375:717-29.
18. Toole MJ, Kidwell KM, Van Poznak C. Oncotype dx results in multiple primary breast cancers. Breast Cancer Basic Clin Res 2014;8:1-6.
19. You YN, Rustin RB, Sullivan JD. Oncotype DX((R)) colon cancer assay for prediction of recurrence risk in patients with stage II and III colon cancer: A review of the evidence. Surg Oncol 2015;24:61-6.
20. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 2010;11:94.
21. Robin X, Turck N, Hainard A et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.
22. 22. Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Academy of Sciences USA 2005;102:15545-50.
23. Massari F, Bria E, Ciccarese C et al. Prognostic Value of Beta-Tubulin-3 and c-Myc in Muscle Invasive Urothelial Carcinoma of the Bladder. PLoS One 2015;10:e0127908.
24. Buttner F, Winter S, Rausch S et al. Survival Prediction of Clear Cell Renal Cell Carcinoma Based on Gene Expression Similarity to the Proximal Tubule of the Nephron. Eur Urol 2015;68:1016-20.
25. Yao L, Wang L, Li F, Gao X, Wei X, Liu Z. MiR181c inhibits ovarian cancer metastasis and progression by targeting PRKCD expression. Int J Clin Exper Med 2015;8:15198-205.
26. Ke G, Liang L, Yang JM et al. MiR-181a confers resistance of cervical cancer to radiation therapy through targeting the pro-apoptotic PRKCD gene. Oncogene 2013;32:3019-27.
27. Chen C, Li Z, Yang Y, Xiang T, Song W, Liu S. Microarray expression profiling of dysregulated long non-coding RNAs in triple-negative breast cancer. Cancer Biol Ther 2015;16:856-65.
28. Agelopoulos K, Richter GH, Schmidt E et al. Deep Sequencing in Conjunction with Expression and Functional Analyses Reveals Activation of FGFR1 in Ewing Sarcoma. Clin Cancer Res 2015;21:4935-46.
29. Peeters MC, Fokkelman M, Boogaard B et al. The adhesion G protein-coupled receptor G2 (ADGRG2/GPR64) constitutively activates SRE and NFkappaB and is involved in cell adhesion and migration. Cell Signalling 2015;27:2579-88.
30. Richter GH, Fasan A, Hauer K et al. G-Protein coupled receptor 64 promotes invasiveness and metastasis in Ewing sarcomas through PGF and MMP1. J Pathol 2013;230:70-81.