SPANDIDOS PUBLICATIONS

INTERNATIONAL JOURNAL OF ONCOLOGY 52: 1178-1188, 2018

A six-long non-coding RNA signature predicts prognosis in melanoma patients

SHUOCHENG YANG[1], JIANGUO XU[2] and XUAN ZENG[3]

[1]Department of Medical Cosmetology, Shanghai Traditional Chinese Medicine Integrated Hospital, Shanghai 200072; [2]Department of Plastic Surgery, Changhai Hospital, Shanghai 200433; [3]Department of Plastic and Reconstructive Surgery, Guangzhou General Hospital of Guangzhou Military Command, Guangzhou, Guangdong 510010, P.R. China

Received September 22, 2017; Accepted January 10, 2018

DOI: 10.3892/ijo.2018.4268

Abstract. The aim of this study was to identify long non-coding RNAs (lncRNAs) which may prove useful for risk-classifying patients with melanoma. For this purpose, based on a dataset from The Cancer Genome Atlas (TCGA), we selected and analyzed samples from melanoma stages I, II, III and IV, from which differentially expressed lncRNAs were identified. The lncRNAs were classified using two-way hierarchical clustering analysis and analysis of support vector machine (SVM), followed by Kaplan-Meier survival analysis. The prognostic capacity of the signature was verified on an independent dataset. lncRNA-mRNA networks were built using signature lncRNAs and corresponding target genes. The Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis was conducted on the target genes. A total of 48 differentially expressed lncRNAs were identified, from which 6 signature lncRNAs (AL050303 and LINC00707, LINC01324, RP11-85G21, RP4-794I6.4 and RP5-855F16) were identified. Two-way hierarchical clustering analysis revealed that the accuracy of the six-lncRNA signature in risk-stratifying samples was 84.84%, and the accuracy of the SVM classifier was 85.9%. This predictive signature performed well on the validation dataset [accuracy, 86.76; area under the ROC curve (AUROC), 0.816]. A total of 720 target genes of the 6 lncRNAs were selected for the lncRNA-mRNA networks. These genes were significantly related to mitogen-activated protein kinase (MAPK), the neurotrophin signaling pathway, focal adhesion pathways, and several immune and inflammation-related pathways. On the whole, we identified a six-lncRNA prognostic signature for risk-stratifying patients with melanoma. These lncRNAs may affect prognosis by regulating the MAPK pathway, immune and inflammation-related pathways, the neurotrophin signaling pathway and focal adhesion pathways.

Correspondence to: Dr Shuocheng Yang, Department of Medical Cosmetology, Shanghai Traditional Chinese Medicine Integrated Hospital, 230 Baoding Road, Shanghai 200072, P.R. China
E-mail: doctoryangsc@126.com

Key words: melanoma, long non-coding RNA, pathway, support vector machine classifier, survival

Introduction

Melanoma develops from pigment-containing cells known as melanocytes. It is the most aggressive type of skin cancer and caused 59,800 deaths globally in 2015 (1,2). When the disease is detected at an early stage (stages I and II), prognosis is favorable; however, the survival rates for patients with melanoma at stages III and IV are low (3). Therefore, the development of precise tests for the detection of melanoma at an early stage are required. To aid in this effort, there is an urgent need to identify novel signature molecules that can be used as prognostic biomarkers of melanoma.

Long non-coding RNAs (lncRNAs) are defined as a class of non-protein-coding RNAs which are >200 nucleotides in length. They are implicated in a variety of transcriptional and post-transcriptional gene regulatory processes, and can therefore affect cellular homeostasis (4). There is also mounting evidence to indicate that lncRNAs may play a role in the cancer paradigm (5,6). Increasing attention has been paid to the potential role of lncRNAs in the molecular mechanisms of melanoma (7). There is evidence to suggest that the lncRNA HOTAIR is linked to melanoma cell motility and invasion (8). Li *et al* reported that the lncRNA BANCR increased malignant melanoma cell proliferation, and that its expression was indicative of a higher mortality rate (9). Moreover, Chen *et al* suggested a four-lncRNA signature for predicting the prognosis of patients with cutaneous melanoma (10). Despite these advancements, the association of lncRNAs with the prognosis of patients with remains elusive.

Compared to the study by Chen *et al*, the current study not only screened for signature lncRNAs that may predict the prognosis of patients with melanoma, but also attempted to unravel the underlying mechanisms. By using a The Cancer Genome Atlas (TCGA), an mRNA dataset containing 376 melanoma samples, differentially expressed lncRNAs were identified between melanoma samples at stages I and II, and melanoma samples at stages III and IV. Out of these differentially expressed lncRNAs, optimal signature lncRNAs were identified using the random forest method and were used to construct a support vector machine (SVM) classifier. By using the SVM classifier,

all samples were then classified into an early-stage-like group and an advanced-stage-like group, and were then subjected to Kaplan-Meier survival analysis. Furthermore, the predictive capability of the lncRNA signature was verified on an independent dataset, and Cox univariate and multivariate regression analyses were employed to search for independent predictors of prognosis. In addition, lncRNA-mRNA networks were constructed using signature lncRNAs and corresponding target genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed for these target genes. The aim of this study was to provide promising prognostic candidates, and to enhance our understanding of the etiology and genetic underpinnings of melanoma.

Data collection and analysis

Data sources. An mRNA-seq expression dataset was accessed from the TCGA data portal (https://portal.gdc.cancer.gov/projects/TCGA-SKCM), which included 376 primary melanoma samples with complete clinical charateristics (Illumina HiSeq 2000 RNA Sequencing platform). The TCGA data were in the form of RNA sequencing data on an Illumina HiSeq 2000 RNA Sequencing platform.

Another mRNA expression dataset (E-MTAB-4725, A-GEOD-13369-Illumina Human Whole-Genome DASL HT platform) consisting of 204 primary melanoma samples was downloaded from EBI ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) and used as a validation set in this study. mRNA expression was assessed using the Illumina Human Whole-Genome DASL HT 12.4 whole genome array, followed by normalization using the quantile method following background correction (11). Demographic and clinical characteristics of the training set and the validation set are shown in Table I, which were compared using the Student's t-test or Chi-square test.

Screening for differentially expressed lncRNAs and hierarchical clustering analysis. The 376 samples in the training dataset were classified according to pathological stage as follows: The early-stage group (stages I and II) and the advanced-stage group (stages III and IV). Subsequently, differentially expressed lncRNAs were screened using the DEseq package (12) and edgeR package (13) in R3.1.0, with a strict cut-off set as a false discovery rate (FDR) of <0.05 and |logFC| of >0.263. The overlapping lncRNAs that were significantly differentially expressed were selected for further analysis.

Two-way hierarchical clustering analysis was performed on the expression values of the significantly overlapping lncRNAs using centered Pearson's correlation metric (14) via the pheatmap package (15) in R. The number of samples at the early or advanced stages was compared between clusters using the Chi-square test with the chisq.test function in R. Patient survival was estimated using the Kaplan-Meier method (16) in the survival package in R, and survival was compared using the log-rank test.

Determination of optimal lncRNA signatures. Random forest models are non-parametric, non-linear models characterized by less overfitting and robust performance, among other reliable features (17). To identify lncRNA signatures that discriminate between patients with the early and advanced stages of the disease in the training set, the random forest method was used via the bootstrap procedure (18) and estimated using out-of-bag (OOB) testing (18). Based on the expression values of the identified lncRNAs signature, two-way hierarchical clustering analysis was performed on the 376 samples in the training set.

Classifying samples using the SVM classifier. To determine whether the signature lncRNAs can distinguish between the two types of melanoma samples, an SVM classifier was constructed based on the expression values of the signature lncRNAs using the SVM function in e1071 package of R (19), with the Sigmoid Kernel function and a 10-fold cross-validation. By using the SVM classifier, the samples in the training set were classified into two groups as follows: the early-stage-like group and the advanced-stage-like group. The survival of the two groups was analyzed using the Kaplan-Meier method.

Verification using an independent set. The signature lncRNAs were further verified on the test set (EBI set). Two-way hierarchical clustering analysis, SVM classifier analysis and Kaplan-Meier survival analysis were conducted sequentially on all samples in the EBI set, based on the lncRNA signature.

Association of clinical factors with prognosis. In the training set, Cox univariate and multivariate regression analyses were performed to determine the association of survival with the following clinical variables: Age, sex, pathologic_M, pathologic_N, pathologic_T, new tumors, radiation therapy and SVM prediction. The melanoma samples were stratified by each clinical variable, and further classified into the early-stage-like group and advanced-stage-like group using the SVM classifier. Subsequently, the survival of the two groups was analyzed using Kaplan-Meier survival analysis.

Construction of lncRNA-mRNA networks and KEGG pathway enrichment analysis. In the training set, correlations between each signature lncRNA with corresponding target genes were computed using the COR function of R. Genes that showed correlations with one or more lncRNA were retained, and then numbered according to the absolute value of correlation co-efficient (R), in descending order. The top 1% target genes were selected for the construction of lncRNA-mRNA networks using the STRING database (http://string-db.org) (20), with the cut-off set at a string score of >0.8. Using The Database for Annotation, Visualization and Integrated Discovery (DAVID) software (21), KEGG pathway enrichment analysis was performed for the genes positively or negatively related to the signature lncRNAs, respectively. Pathways with a P-value <0.05 were selected as significant pathways.

Results

Selection of differentially expressed lncRNAs. The training set included 191 early-stage samples and 185 advanced-stage samples. A total of 107 differentially expressed lncRNAs were selected between the early-stage samples and advanced-stage samples using the edge R package, while 55 differentially expressed lncRNAs were selected using the DEseq package. The 48 overlapping, differentially expressed lncRNAs were selected for further analysis.

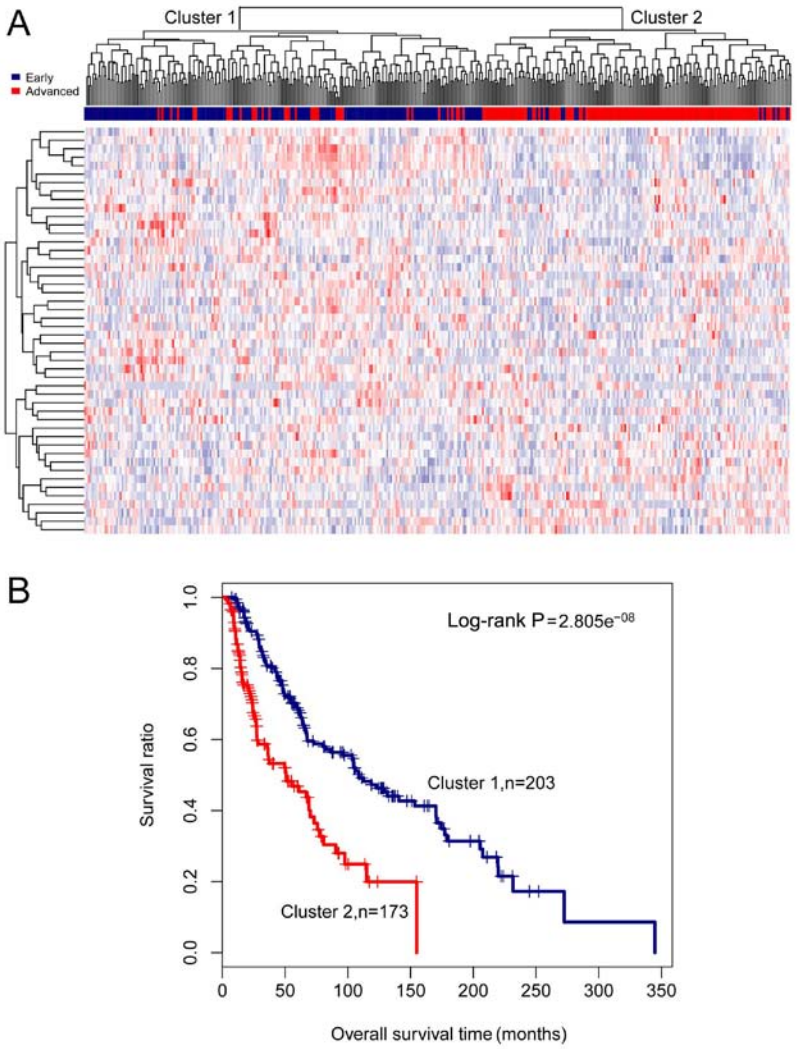

Figure 1. Results of two-way hierarchical clustering analysis based on 48 differentially expressed long non-coding RNAs (lncRNAs). (A) A heatmap for cluster analysis on the training set. All samples are classified into cluster 1 and cluster 2. (B) Kaplan-Meier survival curves of cluster 1 (blue) and cluster 2 (red) obtained from the two-way hierarchical clustering analysis. Survival time is compared between cluster 1 and cluster 2 using the log-rank test.

Hierarchical clustering analysis of differentially expressed lncRNAs. Based on the expression values of the 48 lncRNAs, the samples in the training set were subjected to two-way hierarchical clustering analysis. Two clusters were identified, and these are presented in Fig. 1A. Cluster 1 consisted of 175 early-stage samples and 28 advanced-stage samples, and cluster 2 contained 16 early-stage samples and 157 advanced-stage samples. As the 28 advanced-stage samples in cluster 1, and the 16 early-stage samples in cluster 2 were incorrectly clustered, the accuracy was 88.3% (332/376). A number of early- and advanced-stage samples were differed markedly between the two clusters (χ^2=218.2596, P-value =2.2e^{-16}). Kaplan-Meier survival analysis revealed that survival in cluster 1 was significantly greater compared to that in cluster 2 (log-rank P-value =2.805e^{-08}). Similarly, the mean survival time in cluster 1 was significantly longer compared to that in cluster 2 (79.88±64.70 months vs. 33.31±30.09 months, P-value =1.025e^{-17}) (Fig. 1B).

Identification of optimal signature lncRNAs using the random forest method. Using the random forest method, six lncRNAs with the smallest OOB error (0.162) were identified as an optimal set of lncRNAs and a potential signature for use in patient classification (Fig. 2). The 6 signature lncRNAs are shown in Table II. Among the six signature lncRNAs, the expression of AL050303 and LINC00707 was significantly elevated in the early-stage group compared with the advanced-stage group, while LINC01324, RP11-85G21, RP4-794I6.4 and RP5-855F16 expression was significantly lower in the early stage-group compared with the advanced-stage group (P-value <0.05) (Fig. 3).

Based on expression values of the 6 lncRNAs, two-way hierarchical clustering analysis was performed on the training set. As shown in Fig. 4A, all samples were classified into cluster 1 and cluster 2. Specifically, 172 out of the 210 samples in cluster 1 were early-stage samples, and 147 out of the 166 samples in cluster 2 were advanced-stage samples. The accuracy was 84.84% (319/376), similar to the accuracy of the clustering analysis based on the 48 differentially expressed lncRNAs (88.3%). Moreover, cluster 1 had a significantly better survival (log-rank P-value =8.451e^{-04}) and a markedly longer survival time in comparison with cluster 2 (76.08±63.45 months vs. 35.86±35.61 months, P-value =9.509e^{-14}) (Fig. 4B). These results imply that the 6 signature lncRNAs may represent the 48 differentially expressed lncRNAs.

Table I. Clinical characteristics of patients in the TCGA and E-MTAB-4725 datasets.

Clinical characteristics	TCGA (n=376)	E-MTAB-4725 (n=204)	P-value
Age (years; means ± SD)	57.64±5.44	55.73±12.97	0.1157[a]
Sex (male/female)	235/141	100/104	0.0023[b]
Pathologic_M (M0/M1/-)	351/19/6	202/2	0.0211[b]
Pathologic_N (N0/N1/N2/N3/-)	182/71/43/56/24	182/6/13/3	2.2E^{-16}[b]
Pathologic_T (T1/T2/T3/T4/-)	62/73/80/128/33	6/66/72/59	1.72E^{-08}[b]
Pathologic_stage (I/II/III/IV/-)	80/111/166/19	58/123/21/2	2.2E^{-16}[b]
Radiation therapy (yes/no)	40/336	-	-
New tumor (yes/no)	220/153/3	-	-
Deceased (deceased/alive)	179/197	120/102	0.6192[b]
Overall survival (months; means ± SD)	58.33±56.59	73.57±40.16	0.0001[a]

Clinical characteristics between TCGA and E-MTAB-4725 were compared using the [a]Student's t-test or [b]Chi-square test. SD, standard deviation. The hyphen (-) indicates that data were unavailable. TCGA, The Cancer Genome Atlas.

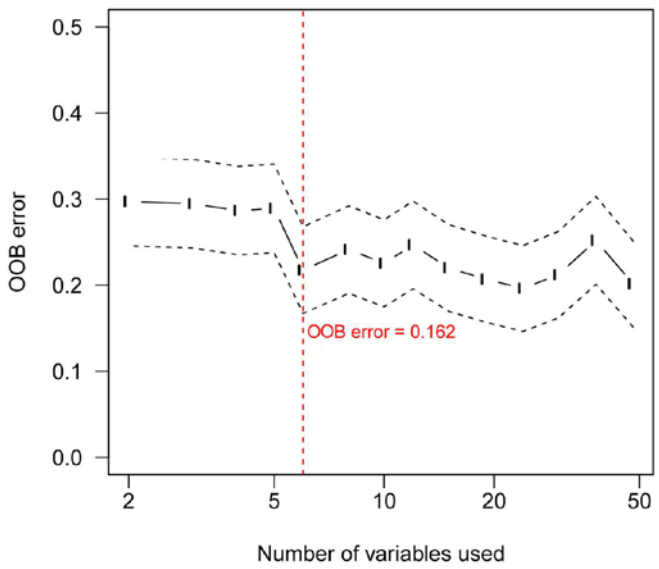

Figure 2. Out-of-bag (OOB) error. The selected OOB error is marked by the red line.

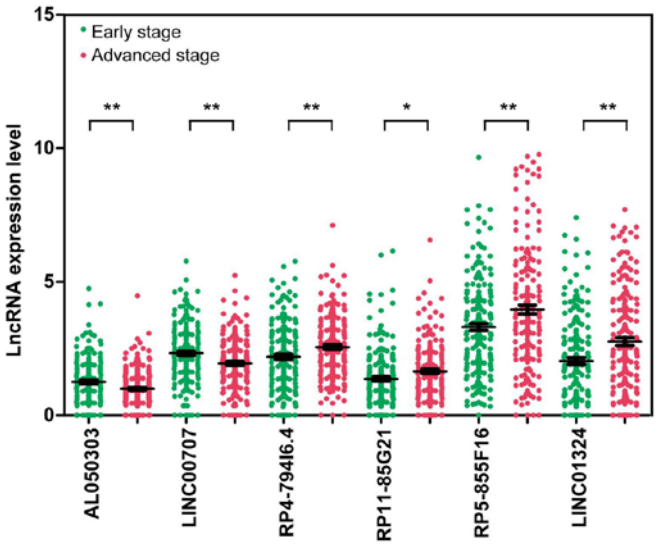

Figure 3. Expression of six signature long non-coding RNAs (lncRNAs) in the early-stage group and the advanced-stage group. Expression in the early-stage group is shown in green; expression in the advanced-stage group is shown in red. *P<0.05 and **P<0.005.

Sample classification using an SVM classifier. Based on the expression values of the six signature lncRNAs, an SVM classifier was built and used to classify the samples in the training set into early-stage-like samples and advanced-stage-like samples. As a result, 23 early-stage samples and 30 advanced-stage samples were incorrectly classified. The accuracy was 85.9% with a sensitivity of 87.29%, a specificity of 84.62%, a positive predictive value (PPV) of 84.04%, a negative predictive value (NPV) of 87.77% and an area under the receiver operating characteristic curve (AUROC) of 0.962 (Fig. 5A). Similarly, as shown in Fig. 5B, the early-stage-like samples had a more favorable survival (log-rank P-value =1.619e^{-03}) and a longer mean survival time compared to the advanced-stage-like samples (67.71±61.76 vs. 48.95±49.29 months, P-value =0.0012).

Validation using an EBI set. The predictive power of the six signature lncRNAs identified using the training set was tested on an EBI set (E-MTAB-4725). The results of two-way hierarchical clustering analysis revealed that the samples in the validation dataset were classified into cluster 1 and cluster 2 (Fig. 6A). Specifically, 1 advanced-stage sample was incorrectly clustered into cluster 1, and 47 early-stage samples were incorrectly clustered into cluster 2. The accuracy was 71.57%. Fig. 6B shows that cluster 1 exhibited a better survival compared to cluster 2 (log-rank P-value =2.716e^{-03}; mean survival time, 78.84±39.43 vs. 65.23±40.14 months, P-value =0.0187).

The performance of an SVM classifier based on the six-lncRNA signature was tested on the EBI set. The results revealed that 1 advanced-stage sample and 26 early-stage samples were incorrectly classified by the SVM classifier with an accuracy of 86.76% and an AUROC of 0.816 (sensitivity, 95.65%; specificity, 85.64%; PPV, 75.83%; NPV, 87.08%) (Fig. 7A). Likewise, the survival of early-stage-like patients (n=156) was much improved in comparison with the advanced-stage-like patients (n=48) (log-rank P-value =1.397e^{-03}; mean survival time,

Table II. Six signature lncRNAs.

		edgeR test			Deseq		
Gene name	Chromosome location	logFC	P-value	FDR	logFC	P-value	FDR
AL050303	Chromosome 21: 13,769,932-13,771,740(+)	-0.4022	0.0002	0.0051	-0.4617	$3.01E^{-05}$	0.0024
LINC00707	Chromosome 10: 6,779,598-6,842,906(+)	-0.3735	$7.63E^{-05}$	0.0018	-0.3905	$3.95E^{-06}$	0.0003
LINC01324	Chromosome 3: 164,714,095-164,831,480(-)	0.5550	$4.98E^{-08}$	$1.17E^{-06}$	0.5727	$1.72E^{-06}$	0.0001
RP11-85G21	Chromosome 1: 157,232,231-157,237,136(-)	0.3565	0.0002	0.0058	0.4001	0.0002	0.0148
RP4-794I6.4	Chromosome 20: 3,239,705-3,245,382(+)	0.3301	0.0001	0.0037	0.3499	$6.13E^{-05}$	0.0050
RP5-855F16	Chromosome 7: 10,940,423-10,940,735(+)	0.4914	$1.31E^{-08}$	$3.08E^{-07}$	0.4639	$5.01E^{-06}$	0.0004

lncRNAs, long non-coding RNAs; FDR, false discovery rate; FC, fold change.

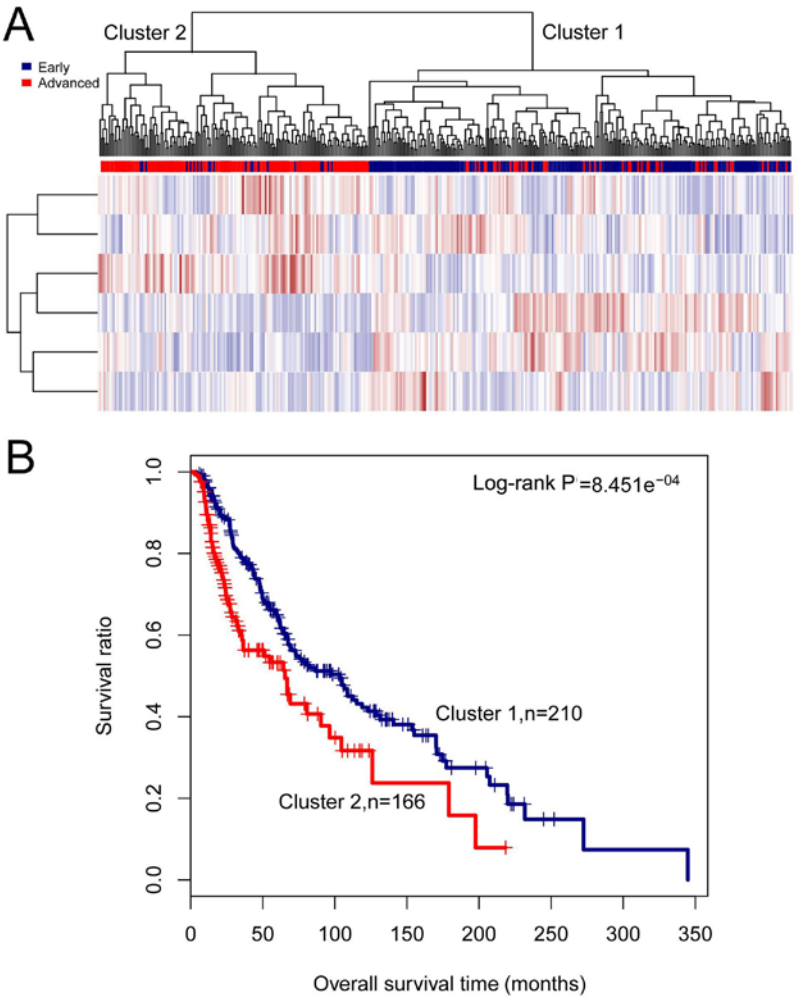

Figure 4. Results of the two-way hierarchical clustering analysis based on six signature long non-coding RNAs (lncRNAs) in the training set. (A) A heatmap for clustering analysis. All samples in the training set are stratified into cluster 1 and cluster 2. (B) Kaplan-Meier survival curves of cluster 1 (blue) and cluster 2 (red) obtained from the two-way hierarchical clustering analysis.

76.96±37.31 vs. 62.54±47.05 months, P-value <0.050 (Fig. 7B). These results confirmed the reliability of the six signature lncRNAs in distinguishing different stages of melanoma samples.

Correlation of clinical characteristics with survival. Using Cox univariate and multivariate regression analyses, we found that based on the six-lncRNA signature SVM prediction, Pathologic_N, Pathologic_T, and new tumors were independent predictors of prognosis of melanoma in the training set (Table III and Fig. 8).

Furthermore, the samples were stratified by clinical characteristics and classified using the six-lncRNA signature-based SVM classifier. As shown in Table IV, the SVM classifier was also effective in distinguishing the early-stage samples from

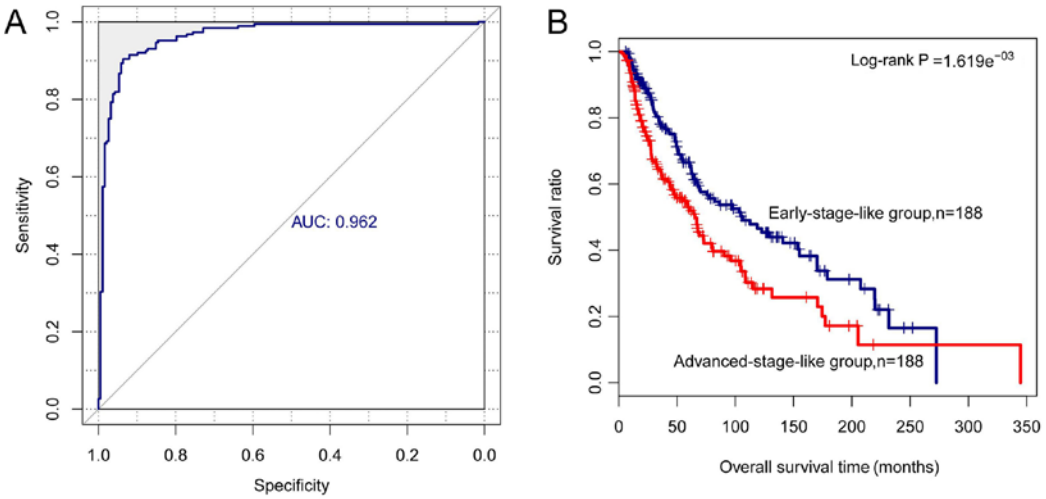

Figure 5. Performance of the support vector machine (SVM) classifier based on the six-long non-coding RNA (lncRNA) signature in the training set. (A) ROC analysis of the SVM classifier. All samples in the training set are classified into early-stage-like group and advanced-stage-like group via the SVM classifer. (B) Kaplan-Meier survival curves for early-stage-like samples (blue) and advanced-stage-like samples (red).

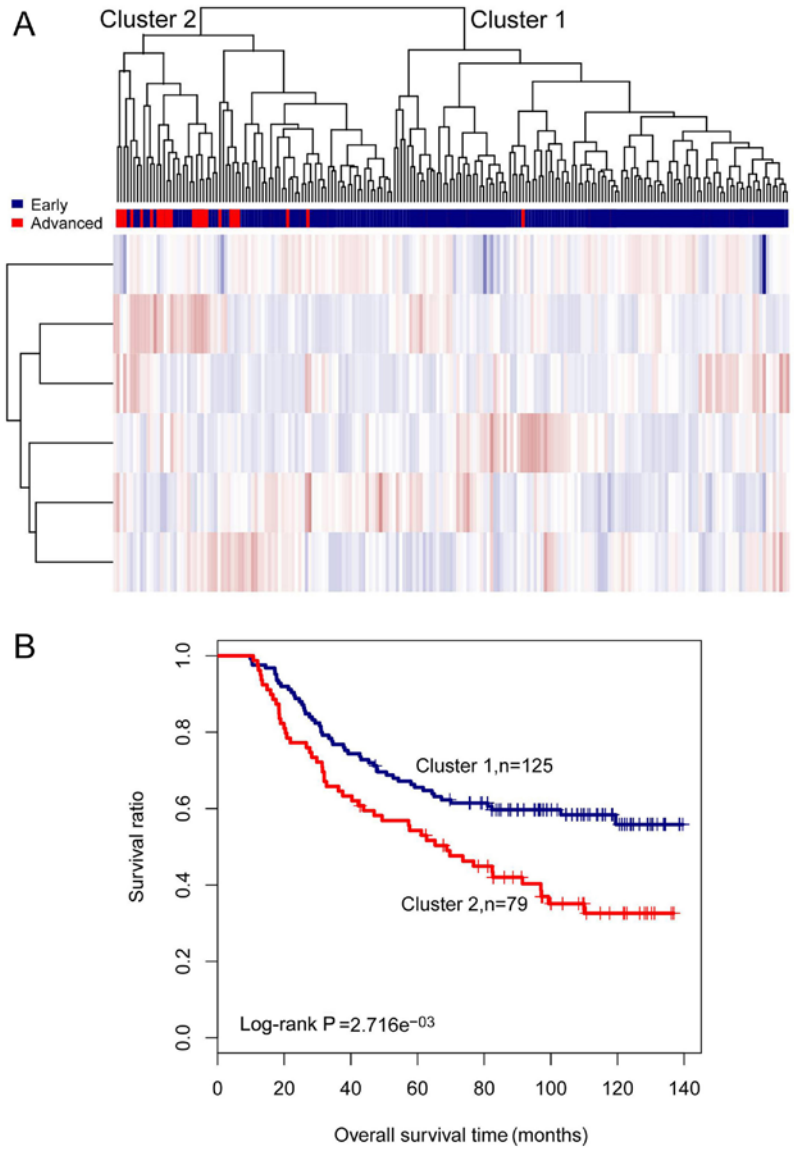

Figure 6. Results of the two-way hierarchical clustering analysis based on 6 signature long non-coding RNAs (lncRNAs) in the validation set. (A) A heatmap of clustering analysis. All samples are clustered into cluster 1 and cluster 2. (B) Kaplan-Meier survival curves for cluster 1 (blue) and cluster 2 (red). Survival time is compared between cluster 1 and cluster 2 using the log-rank test.

Table III. Results of Cox univariate and multivariate regression analyses.

	Univariate analysis			Multivariate analysis		
Variables	HR	95% CI	P-value	HR	95% CI	P-value
SVM prediction						
Early/advanced stage	1.61	1.194-2.17	0.0016	1.618	1.139-2.299	0.0073
Age (years)						
≤60/>60	1.528	1.131-2.064	0.0055	1.238	0.878-1.745	0.224
Sex						
Male/female	1.098	0.801-1.505	0.561	1.298	0.9089-1.856	0.152
Pathologic_M stage						
M0/M1	2.278	1.195-4.342	0.0101	1.714	0.723-4.066	0.221
Pathologic_N stage						
N0-N1/N2-N3	1.581	1.111-2.251	0.0103	1.806	1.167-2.794	0.0080
Pathologic_T stage						
T0-T2/T3-T4	1.938	1.405-2.673	$4.05E^{-05}$	1.91	1.343-2.717	0.0003
New tumor						
Yes/no	2.687	1.831-3.944	$1.48E^{-07}$	3.125	1.972-4.955	$1.25E^{-06}$
Radiation therapy						
Yes/no	0.4771	0.271-0.841	0.0088	0.866	0.440-1.703	0.677

HR, hazard ratio; CI, confidence interval; SVM, support vector machine.

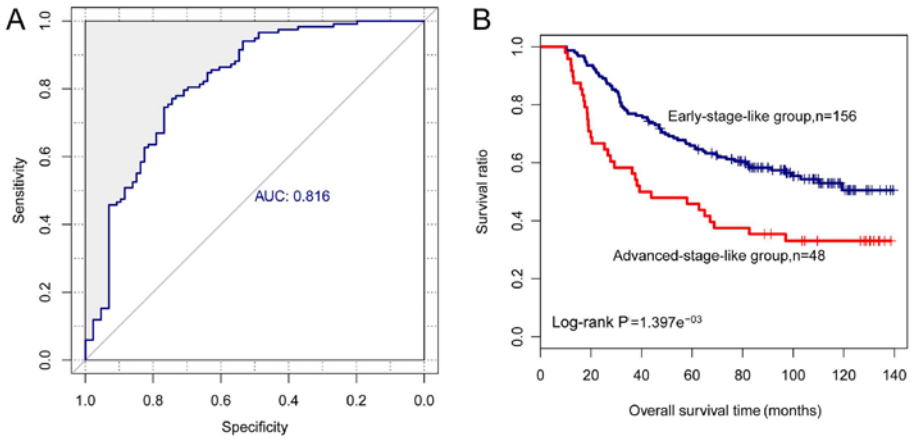

Figure 7. Performance of the six-long non-coding RNA (lncRNA) signature-based support vector machine (SVM) classifier in the validation set. (A) ROC analysis of the SVM classifier. All samples in the validation set are divided into early-stage-like group and advanced-stage-like group via the SVM classifer. (B) Kaplan-Meier survival curves for early-stage-like samples (blue) oradvanced-stage-like samples (red).

the advanced-stage samples for patients of any age, male patients, patients with pathologic_M0 or pathologic_N2-N3 or pathologic_T3-T4, patients with new tumors, and patients who did not receive radiation therapy (P-value <0.05) (Fig. 9). It should be noted that some information for several samples was not available in the dataset.

Pathway enrichment analysis of the six-lncRNA signature. Functional analysis was employed to determine the possible role of the six-lncRNA signature in the pathogenesis of melanoma. In the training set, the association of each signature lncRNA with its target genes was analyzed. A total of 720 genes that were associated with the signature lncRNAs were obtained, 637 of which were positively related to the signature lncRNAs and 83 of which were negatively related to the signature lncRNAs. Additionally, lncRNA-mRNA networks were constructed using the lncRNA-mRNA pairs (score >0.8) (Fig. 10).

As shown in Fig. 11, the negatively associated genes were significantly clustered in 6 pathways, including the mitogen-activated protein kinase (MAPK) signaling pathway, pathway in cancer, neurotrophin signaling pathway, long-term potentiation, and the natural killer cell mediated cytotoxicity pathway. The positively related genes were significantly enriched in 8 pathways, including the intestinal immune

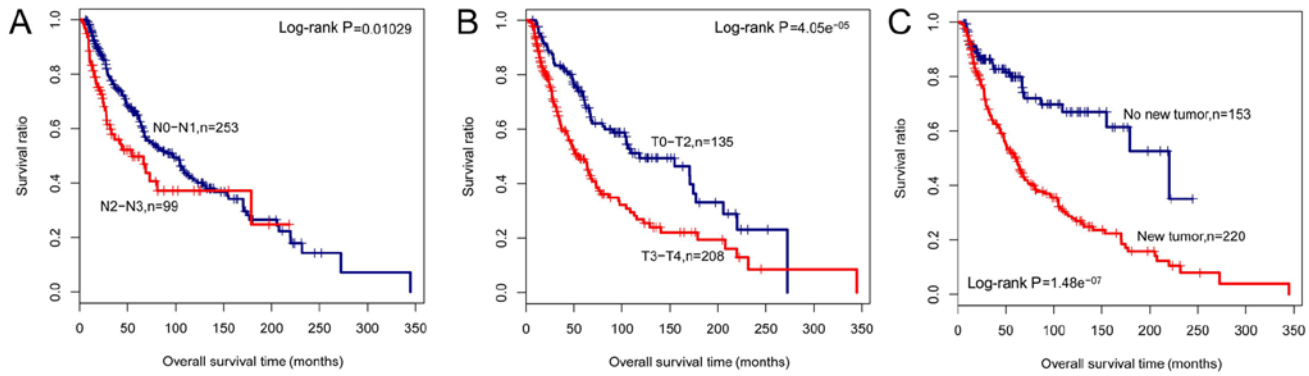

Figure 8. Kaplan-Meier survival analysis for the determined independent prognostic factors in melanoma. (A) Kaplan-Meier survival curves for patients with pathologic_N0-N1 or N2-N3 stage; (B) Kaplan-Meier survival curves for patients with pathologic_T0-T2 or T3-T3 stage; (C) Kaplan-Meier survival curves for patients with or without new tumors.

Table IV. Associations of clinical features with the prognostic capability of the SVM classifier.

	Univariate analysis		
Variables	HR	95% CI	P-value
Age (years)			
≤60 (n=209)	1.755	1.161-2.655	0.0069
>60 (n=167)	1.642	1.048-2.573	0.0289
Sex			
Male (n=235)	1.746	1.21-2.521	0.0025
Female (n=141)	1.468	0.869-2.482	0.1491
Pathologic_M stage			
M0 (n=351)	1.547	1.138-2.103	0.0050
M1 (n=19)	2.990	0.374-3.239	0.2790
Pathologic_N stage			
N0-N1 (n=253)	1.401	0.979-2.007	0.0642
N2-N3 (n=99)	3.765	1.34-10.58	0.0070
Pathologic_T stage			
T0-T2 (n=135)	1.642	0.978-2.756	0.0582
T3-T4 (n=208)	1.510	1.019-2.239	0.0387
New tumor			
Yes (n=220)	1.642	1.179-2.288	0.0031
No (n=153)	1.894	0.901-3.983	0.0873
Radiation therapy			
Yes (n=40)	1.979	0.614-6.378	0.2444
No (n=336)	1.549	1.137-2.111	0.0052

The patients are stratified by different clinical characteristics and further classified into early-stage-like samples and advanced-stage-like samples by a six lncRNAs-based SVM classifer. The survival of the early-stage-like samples and advanced-stage-like samples was compared using a log-rank test. SVM, support vector machine; HR, hazard ratio; CI, confidence interval.

network for IgA production, leukocyte transendothelial migration, complement and coagulation cascades, cell adhesion molecules (CAMs), chemokine signaling pathway, cytokine-cytokine receptor interaction, the MAPK pathway, and keratan sulfate biosynthesis. Notably, the MAPK pathway was significantly enriched with 16 positively associated genes and 11 negatively associated genes, such as mitogen-activated protein kinase kinase kinase kinase 1 (*MAP4K1*), RAS guanyl releasing protein 2 (*RASGRP2*), mitogen-activated protein kinase 8 interacting protein 3 (*MAPK8IP3*), mitogen-activated protein kinase kinase 5 (*MAP2K5*) and the B-Raf proto-oncogene, serine/threonine kinase (*BRAF*).

Discussion

Melanoma is an aggressive skin cancer, and the importance of lncRNAs in the biology of melanoma has been increasingly acknowledged in recent years. To the best of our knowledge, the functions of ~13 lncRNAs in melanoma have been determined (7). Nevertheless, there are limited studies discussing the association of lncRNAs with patient prognosis. Based on a TCGA dataset that included 376 samples, this study identified a potential prognostic six-signature lncRNA. This signature included AL050303, LINC00707, LINC01324, RP11-85G21, RP4-794I6.4 and RP5-855F16. Of these lncRNAs, AL050303 and LINC00707 were upregulated, while RP11-85G21, RP4-794I6.4 and RP5-855F16 were downregulated in the early-stage samples compared to the advanced-stage samples.

The classification capability of the signature lncRNAs was verified on an independent dataset that included 204 samples. Two-way hierarchical clustering analysis, SVM classifier analysis and Kaplan-Meier analysis achieved consistent results that support the conclusion that this six-lncRNA signature exhibited reliable predictive accuracy. Furthermore, Cox univariate and multivariate regression analyses revealed that the six-lncRNA signature-based SVM prediction was an independent predictor of prognosis. To the best of our knowledge, the prognostic value of this multi-marker signature in melanoma has not been previously reported. Therefore, the current study provides new insight into the improved risk-stratification and prediction of survival in patients with melanoma.

A growing number of studies have demonstrated a key role for MAPK dysregulation in melanoma, which largely results from mutations in the *B-RAF* and *RAS* genes (22,23). Moreover, BRAF and MEK inhibitors have been developed and have achieved unprecedented treatment outcomes in clinic practice (24). In the

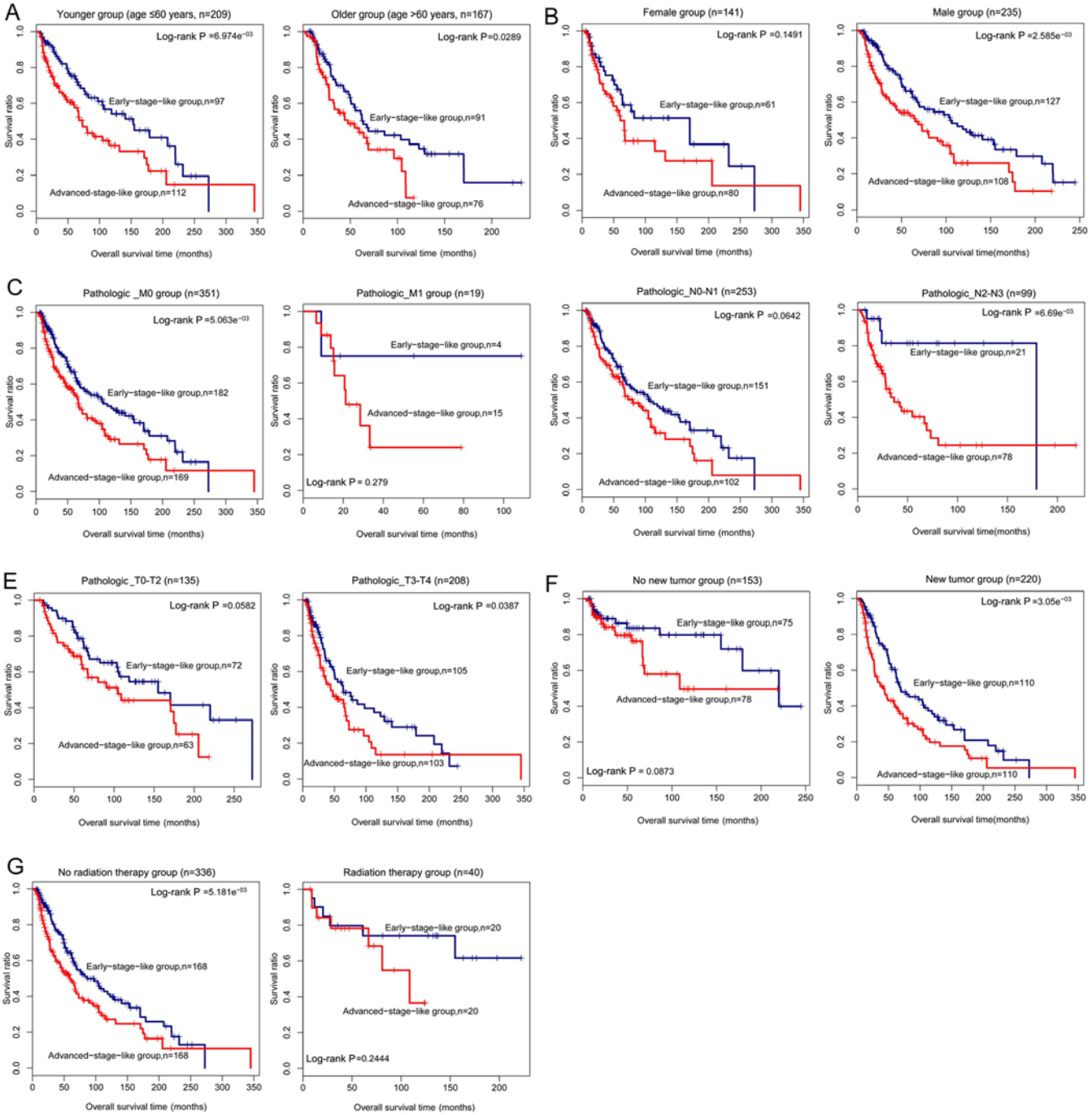

Figure 9. Kaplan-Meier survival analysis for patients stratified by different clinical factors. (A) Kaplan-Meier survival curves for patients ≤60 years (left panel) or >60 years (right panel) of age. (B) Kaplan-Meier survival curves for female (left panel) or male (right panel) patients. (C) Kaplan-Meier survival curves for patients with pathologic_M0 (left panel) or pathologic_M1 stage (right panel). (D) Kaplan-Meier survival curves for patients with pathologic_N0-N1 (left panel) or pathologic_N2-N3 stage (right panel). (E) Kaplan-Meier survival curves for patients with pathologic_T0-T2 (left panel) or pathologic_T3-T4 stage (right panel). (F) Kaplan-Meier survival curves for patients with (right panel) or without new tumors (left panel). (G) Kaplan-Meier survival curves for patients receiving radiation therapy (right panel) or not (left panel). Patients stratified by different clinical factors are further classified into early-stage-like group and advanced-stage-like group using the support vector machine (SVM) classifer. Survival curves for early-stage-like group and advanced-stage-like group are labeled in blue and red, respectively.

present study, *MAP4K1*, *RASGRP2*, *MAPK8IP3*, *MAP2K5* and *BRAF* were identified as target genes of the six-lncRNA signature, which was significantly enriched in MAPK pathway genes. MAP4K1, and MAP2K5 are members of the MAP kinase family. MAPK8IP3 has been found to interact with various members of the MAP kinase family as well as C-Raf (25). The protein encoded by RASGRP2 can activate RAS and RAP1/RAS3. These findings suggest that the six signature lncRNAs may affect prognosis in melanoma by modulating the MAPK pathway.

A rich body of evidence has demonstrated that the immune system and inflammation are closely associated with cancer progression, including melanoma (26,27). In this study, target genes of the multi-marker signature were identified in several immune and inflammation-related pathways including the following: Complement and coagulation cascades, leukocyte transendothelial migration, the chemokine signaling pathway, intestinal immune network for IgA production, and natural killer cell-mediated cytotoxicity pathways. Melanocytes express

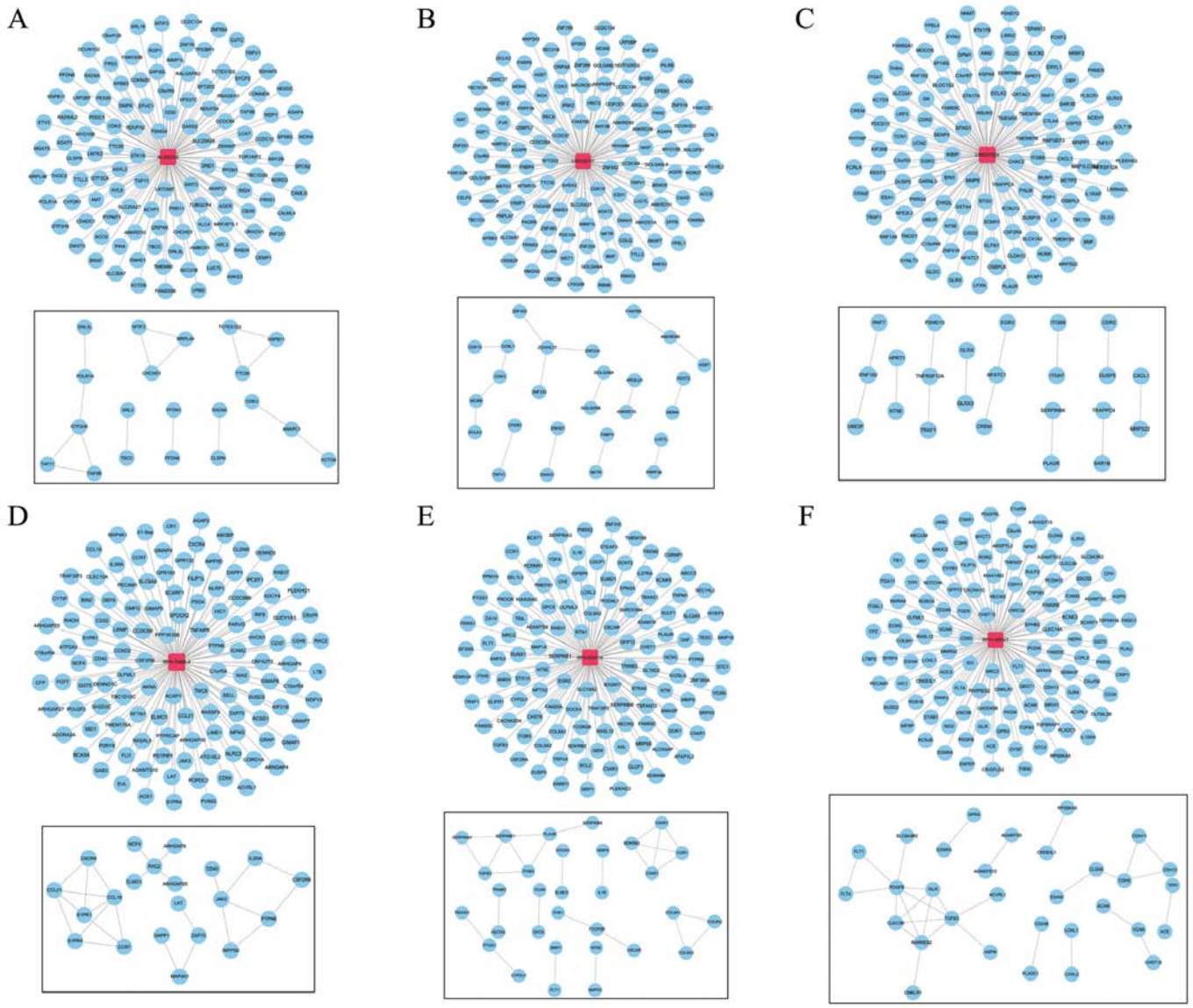

Figure 10. Long non-coding RNA (lncRNA)-mRNA networks. Red square nodes indicate lncRNAs, and blue round nodes indicate target genes of lncRNAs. (A-F) lncRNA-mRNA networks for AL050303, LINC00707, LINC01324, RP4-794I6.4, RP5-855F16 and RP11-85G21, separately.

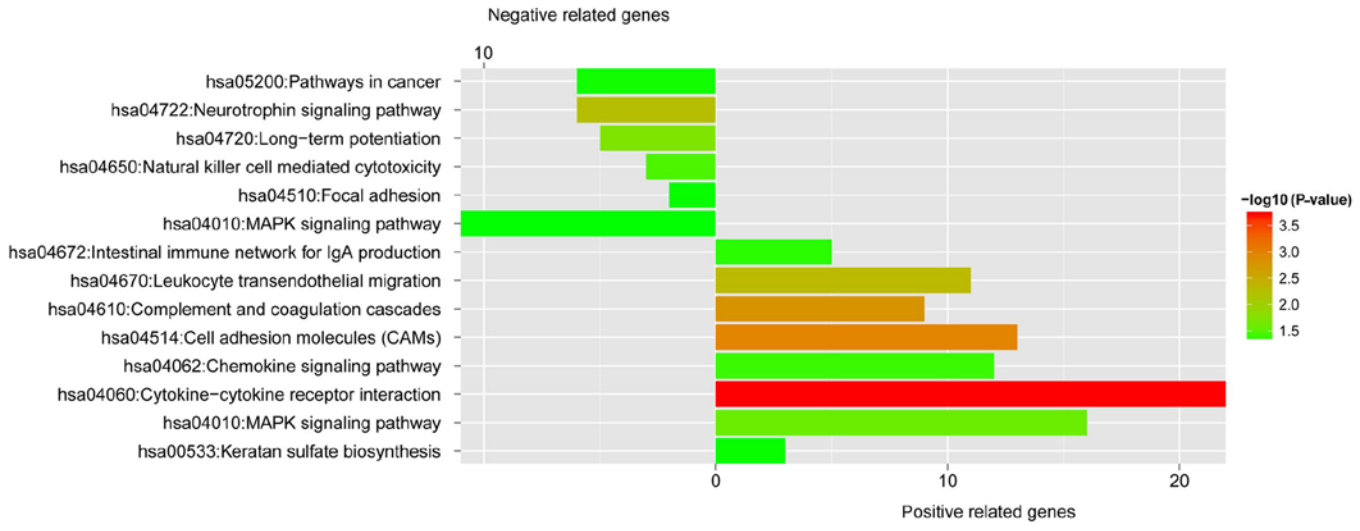

Figure 11. Significant KEGG pathways enriched with positively or negatively related genes. Vertical axis, number of genes enriched in each pathway; horizontal axis, and significant KEGG pathways.

neurotrophins and their receptors, which play an important part in modulating melanoma cell proliferation and migration (28). Focal adhesion kinases are implicated in regulating melanoma cell motility and migration (29,30). The present study found that the neurotrophin signaling pathway and the focal adhesion pathway were significantly linked to the target genes of the six-lncRNA signature. These results imply that the six-lncRNA signature may be involved in regulating immune and inflammation-related pathways, the neurotrophin signaling pathway, and the focal adhesion pathway, thereby influencing the survival of patients.

It should be noted that the results of this study may have been influenced by sample heterogeneity and/or differing sample collection or RNA extraction methods (31). Additionally, the sample size of this study was limited. Further studies with a larger cohort of patients and timely follow-up are warranted in order to confirm the predictive capacity of this signature in melanoma.

In conclusion, in this study, we identified a six-lncRNA signature as a useful prognostic biomarker for risk-classifying patients with melanoma. The lncRNAs may affect prognosis partly by modulating MAPK, immune and inflammation-related pathways, the neurotrophin signaling pathway, and the focal adhesion pathway. These findings provide novel insight into the correlation of lncRNAs with prognosis, and help lay a foundation for improving the survival of patients with melanoma. Further studies are warranted to validate this prognostic signature.

Competing interests

The authors declare that they have no competing interests.

References

1. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, Casey DC, Charlson FJ, Chen AZ, Coates MM, *et al*; GBD 2015 Mortality and Causes of Death Collaborators: Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: A systematic analysis for the Global Burden of Disease Study 2015. Lancet 388: 1459-1544, 2016.
2. MacKie RM, Hauschild A and Eggermont AMM: Epidemiology of invasive cutaneous melanoma. Ann Oncol 20 (Suppl 6): vi1-vi7, 2009.
3. Piris A, Lobo AC and Duncan LM: Melanoma staging: Where are we now? Dermatol Clin 30: 581-592, 2012.
4. Mercer TR, Dinger ME and Mattick JS: Long non-coding RNAs: Insights into functions. Nat Rev Genet 10: 155-159, 2009.
5. Gibb EA, Brown CJ and Lam WL: The functional role of long non-coding RNA in human carcinomas. Mol Cancer 10: 38, 2011.
6. Huarte M: The emerging role of lncRNAs in cancer. Nat Med 21: 1253-1261, 2015.
7. Hulstaert E, Brochez L, Volders PJ, Vandesompele J and Mestdagh P: Long non-coding RNAs in cutaneous melanoma: Clinical perspectives. Oncotarget 8: 43470-43480, 2017.
8. Tang L, Zhang W, Su B and Yu B: Long noncoding RNA HOTAIR is associated with motility, invasion, and metastatic potential of metastatic melanoma. Biomed Res Int 2013: 251098, 2013.
9. Li R, Zhang L, Jia L, Duan Y, Li Y, Bao L and Sha N: Long non-coding RNA BANCR promotes proliferation in malignant melanoma by regulating MAPK pathway activation. PLoS One 9: e100893, 2014.
10. Chen X, Guo W, Xu XJ, Su F, Wang Y, Zhang Y, Wang Q and Zhu L: Melanoma long non-coding RNA signature predicts prognostic survival and directs clinical risk-specific treatments. J Dermatol Sci 85: 226-234, 2017.
11. Nsengimana J, Laye J, Filia A, Walker C, Jewell R, Van den Oord JJ, Wolter P, Patel P, Sucker A, Schadendorf D, *et al*: Independent replication of a melanoma subtype gene signature and evaluation of its prognostic value and biological correlates in a population cohort. Oncotarget 6: 11683-11693, 2015.
12. Anders S and Huber W: Differential expression analysis for sequence count data. Genome Biol 11: R106, 2010.
13. Robinson MD, McCarthy DJ and Smyth GK: edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140, 2010.
14. Eisen MB, Spellman PT, Brown PO and Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863-14868, 1998.
15. Wang L, Cao C, Ma Q, Zeng Q, Wang H, Cheng Z, Zhu G, Qi J, Ma H, Nian H, *et al*: RNA-seq analyses of multiple meristems of soybean: Novel and alternative transcripts, evolutionary and functional implications. BMC Plant Biol 14: 169, 2014.
16. Wang P, Wang Y, Hang B, Zou X and Mao JH: A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget 7: 55343-55351, 2016.
17. Tolosi L and Lengauer T: Classification with correlated features: Unreliability of feature ranking and solutions. Bioinformatics 27: 1986-1994, 2011.
18. Zapf A, Brunner E and Konietschke F: A wild bootstrap approach for the selection of biomarkers in early diagnostic trials. BMC Med Res Methodol 15: 43, 2015.
19. Wang Q and Liu X: Screening of feature genes in distinguishing different types of breast cancer using support vector machine. Onco Targets Ther 8: 2311-2317, 2015.
20. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, *et al*: The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 39: D561-D568, 2011.
21. Huang W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57, 2009.
22. Inamdar GS, Madhunapantula SV and Robertson GP: Targeting the MAPK pathway in melanoma: Why some approaches succeed and other fail. Biochem Pharmacol 80: 624-637, 2010.
23. Grimaldi AM, Simeone E, Festino L, Vanella V, Palla M and Ascierto PA: Novel mechanisms and therapeutic approaches in melanoma: Targeting the MAPK pathway. Discov Med 19: 455-461, 2015.
24. Wellbrock C and Arozarena I: The Complexity of the ERK/MAP-kinase pathway and the treatment of melanoma skin cancer. Front Cell Dev Biol 4: 33, 2016.
25. Ito M, Yoshioka K, Akechi M, Yamashita S, Takamatsu N, Sugiyama K, Hibi M, Nakabeppu Y, Shiba T and Yamamoto KI: JSAP1, a novel jun N-terminal protein kinase (JNK)-binding protein that functions as a Scaffold factor in the JNK signaling pathway. Mol Cell Biol 19: 7539-7548, 1999.
26. Meyer C, Sevko A, Ramacher M, Bazhin AV, Falk CS, Osen W, Borrello I, Kato M, Schadendorf D, Baniyash M, *et al*: Chronic inflammation promotes myeloid-derived suppressor cell activation blocking antitumor immunity in transgenic mouse melanoma model. Proc Natl Acad Sci USA 108: 17111-17116, 2011.
27. Keibel A, Singh V and Sharma MC: Inflammation, microenvironment, and the immune system in cancer progression. Curr Pharm Des 15: 1949-1955, 2009.
28. Truzzi F, Marconi A, Lotti R, Dallaglio K, French LE, Hempstead BL and Pincelli C: Neurotrophins and their receptors stimulate melanoma cell proliferation and migration. J Invest Dermatol 128: 2031-2040, 2008.
29. Wang R, Bi J, Ampah KK, Ba X, Liu W and Zeng X: Lipid rafts control human melanoma cell migration by regulating focal adhesion disassembly. Biochim Biophys Acta 1833: 3195-3205, 2013.
30. Hess AR, Postovit LM, Margaryan NV, Seftor EA, Schneider GB, Seftor RE, Nickoloff BJ and Hendrix MJ: Focal adhesion kinase promotes the aggressive melanoma phenotype. Cancer Res 65: 9851-9860, 2005.
31. Tímár J, Gyorffy B and Rásó E: Gene signature of the metastatic potential of cutaneous melanoma: Too much for too little? Clin Exp Metastasis 27: 371-387, 2010.