Research article

Prognostic score model-based signature genes for predicting the prognosis of metastatic skin cutaneous melanoma

Jiaping Wang*

Laboratory Medicine, Donghai County People's Hospital, Lianyungang City, Jiangsu 222300, China

* **Correspondence:** Email: 13585281988@163.com; Tel: +86051880303252.

Abstract: *Purpose:* Cutaneous melanoma (SKCM) is the most invasive malignancy of skin cancer. Metastasis to distant lymph nodes or other system is an indicator of poor prognosis in melanoma patients. The aim of this study was to identify reliable prognostic biomarkers for SKCMs. *Methods:* Four RNA-sequencing datasets associated with SKCMs were downloaded from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) database as well as corresponding clinical information. Differentially expressed genes (DEGs) were screened between primary and metastatic samples by using MetaDE tool. Weighted gene co-expression network analysis (WGCNA) was conducted to screen functional modules. A prognostic score (PS)-based predictive model and nomogram model were constructed to identify signature genes and independent clinicopathologic factors. *Results:* Based on MetaDE analysis and WGCNA, a total of 456 overlapped genes were identified as hub genes related to SKCMs progression. Functional enrichment analysis revealed these genes were mainly involved in the hippo signaling pathway, signaling pathways regulating pluripotency of stem cells, pathways in cancer. In addition, eight optimal DEGs (*RFPL1S*, *CTSV*, *EGLN3*, etc.) were identified as signature genes by using PS model. Cox regression analysis revealed that pathologic stage T, N and recurrence were independent prognostic factors. Three clinical factors and PS status were incorporated to construct a nomogram predictive model for estimating the three years and five-year survival probability of individuals. *Conclusions:* The prognosis prediction model of this study may provide a promising method for decision making in clinic and prognosis predicting of SKCM patients.

Keywords: cutaneous melanoma; WGCNA; prognosis; metastasis; nomogram model

1. Introduction

Skin cutaneous melanoma (SKCM) is an aggressive malignancy with the highest mortality in skin cancer. In general, it is characterized by high grade malignancy, and most patients are diagnosed at an advanced stage, which leads to missed therapeutic opportunities [1]. According to cancer statistics in 2020, approximately 95,710 new cases of skin melanoma are diagnosed in US [2]. Although recent checkpoint blockade immunotherapy (such as PD-1/PD-L1 antibodies) and targeted therapy have contributed to medical breakthroughs, the 5-year overall survival rate of metastatic SKCM is still less than 5% due to local recurrence and metastasis [3,4]. Therefore, it is urgent demands to develop more superior reliable biomarkers for early detection and prognosis prediction in SKCMs.

Metastasis of melanoma was associated with activation of several pathways, such as epithelial-mesenchymal transition (EMT), angiogenesis, pericytic mimicry and extravascular migratory [5,6]. However, the genetic and molecular mechanism of metastasis remains unclear. Cutaneous melanoma can metastasize haematogenously or through the lymphatic system, and three exist predominant models can explain the progression, stepwise spread, simultaneous spread and differential spread model [7,8]. Brain metastases are particularly common in metastatic melanoma patients and they can also metastasize to liver, bones, or distant lymph nodes [9]. Nowadays, the increasing application of RNA-sequencing technologies has shown effective methods for understanding tumor metastasis. With the development of gene expression profiling, researchers have identified gene expression signatures that are associated with metastasis or survival outcomes [10,11]. Recently, bioinformatic analysis of public databases has been broadly used to explore prognostic biomarkers in disease progression, and the predictive models can also be utilized to assess prognosis in melanoma patients [12,13]. Thus, the identification of reliable signature genes and clinical factors will provide a guide for decision-making in clinical.

In this study, we comprehensively analyzed the RNA-sequencing profiles of SKCMs by using weighted gene co-expression network analysis (WGCNA) and bioinformatics analysis. Based on prognostic score (PS) model and nomogram model, we screened eight signature genes and three clinical characteristics for prognosis prediction of SKCMs. Our findings might provide reliable prognostic biomarkers in SKCMs.

2. Materials and methods

2.1. Data resource

RNA-sequencing profiles of SKCMs were derived from The Cancer Genome Atlas (TCGA) (https: //gdc-portal.nci.nih.gov/) on September 9, 2020 and tested on Illumina HiSeq 2000 RNA sequencing platform. The profile included 473 SKCM tumor samples, of which 458 SKCM tumor samples with metastasis and clinical survival information were selected as training data set.

In addition, microarray data of SKCM were downloaded from the Gene Expression Omnibus (GEO) (http: //www.ncbi.nlm.nih.gov/geo/) based on following screening criteria: 1) transcriptional expression profiles; 2) skin solid tissue samples from SKCM patients; 3) human expression profiles; 4) individuals containing metastasis information; 5) samples counts should be more than fifty; 6) information with survival and prognosis. Finally, we obtained three datasets associated with SKCM,

including GSE46517 [14], GSE7553 [15], and GSE65904 [16,17]. The whole testing platform and corresponding usage of these profiles were shown in Table 1.

Table 1. Datasets and usage information included in this study.

Datasets	Platform	Component of samples	Usage
TCGA	Illumina HiSeq 2000	102 primary, 356 metastatic	Used for DEGs, WGCNA and prognostic analysis
GSE46517	GPL570 Affymetrix	31 primary, 73 metastatic	Used for DEGs, WGCNA analysis
GSE7553	GPL570 Affymetrix	14 primary, 40 metastatic	
GSE65904	GPL10558 Illumina	214 SKCM	Used for prognostic model validation

Moreover, annotation of each platform involved in Table 1 were download from Ensembl genome browser 96 databases [18] (http: //asia.ensembl.org/index.html), including RNA types, gene symbol and probe, etc. After re-annotate the detection probes in the expression profile dataset, the corresponding long non-coding RNAs (lncRNAs) and mRNA expression level were finally obtained.

2.2. Screening DEGs

MetaDE tool [19,20] in R3.6.1 software was utilized to screened significantly differentially expressed RNAs from three databases of TCGA, GSE46517 and GSE7553. Heterogeneity of gene expression value was tested with parameters of tau^2, Q-value and Qpval. The tau^2 represents the amount of heterogeneity, and Qpval represents the heterogeneity of data set. If $tau^2 = 0$ and Qpval > 0.05, it means that the gene was homogeneous and unbiased. Therefore, false discovery rate (FDR) < 0.05, $tau^2 = 0$ and Qpval > 0.05 were set as thresholds to ensure homogeneity of signature genes. Combining with |log2 fold-change (FC)| parameter, the genes with the same direction of logFC in the three datasets is screened as candidate genes for further analysis.

2.3. WGCNA

WGCNA is a method for screening diseases-related modules from thousands of genes, and then incorporating these modules to clinical characteristics [21]. In this study, we used WGCNA version 1.61 packet [22] to screen disease-related modules. Profiles of TCGA were set as training set, while GSE46517 and GSE7553 were set as validation set. All algorithms were conducted following scale-free features. The correlation of co-expression matrix was analyzed and adjacency functions were defined. By the criteria of gene counts $\geq$ 50 and cut Height = 0.995, we identified the modules associated with disease traits. The stability of modules across differential datasets was calculated based on two validation sets.

The mRNAs screened from MetaDE analysis were corresponding to WGCNA modules to identify overlapped genes significantly associated with disease traits. Hypergeometric algorithm [23] were conducted to calculated enrichment parameter and significance P value of DEGs in each module. Fold enrichment > 1 and $P < 0.05$ were set as thresholds for hub gene screening.

$$f(k, N, M, n) = C(k, M) * C(n-k, N-M)/C(n, N)$$

where "N" and "M" represent the whole gene cohorts and module-related genes screened by WGCNA. "n" and "k" represent the whole gene counts and module-related genes screened by MetaDE analysis.

2.4. Co-expression network construction

To identify crucial genes related to SKCM metastasis, we calculated the Pearson correlation coefficient (PCC) of DEGs in TCGA training set by using cor function. We constructed the co-expression network for the overlapped genes or lncRNAs screened by MetaDE analysis and WGCNA. Cytoscape 3.6.1 [24] were utilized to visualize the connection of gene pairs. GO and KEGG pathways analysis were conducted by using DAVID version 6.8 [25] tool, and p value less than 0.05 was selected as the threshold.

2.5. Identification optimal RNAs for prognostic risk prediction

Cox regression analysis were conducted to screen DEGs associated with survival time in TCGA dataset by using survival package [26]. Log-rank $p < 0.05$ was selected as a threshold.

Next, cox-proportional hazards (Cox-PH) model [27] in penalized package version 0.9-50 [28] were used to further screen the optimal RNAs associated with SKCM prognosis. The optimal parameter "lambda" in the screening model is obtained through cycling 1000 cross-validation likelihood calculation. PS model were constructed on the basis of least absolute shrinkage and selection operator (LASSO) coefficient and expression value of DEGs in training set. The PS calculation formula is as follows:

$$\text{Prognostic score (PS)} = \sum \beta_{\text{RNAs}} \times \text{Exp}_{\text{RNAs}}$$

where β_{RNAs} represents the LASSO prognostic coefficient, Exp_{RNAs} refers to expression level of RNAs in training set.

To validate the performance ability of this predictive model, we calculated PS value of each sample, and divided samples of training set and validation set into high and low risk groups by setting median PS value as criteria. The correlation between risk grouping and overall survival time was assessed by using survival package.

2.6. Identify independent prognostic clinicopathologic factors

Univariate and multivariate cox regression analysis were conducted to screen the clinical factors correlated to survival times of SKCMs. Log-rank p value < 0.05 was selected as a threshold. Since then, to further explore the correlation of risk grouping and variables, individuals could be divided into differential subgroups on the basis of clinicopathologic parameters, such as pathologic T, N, recurrence, etc.
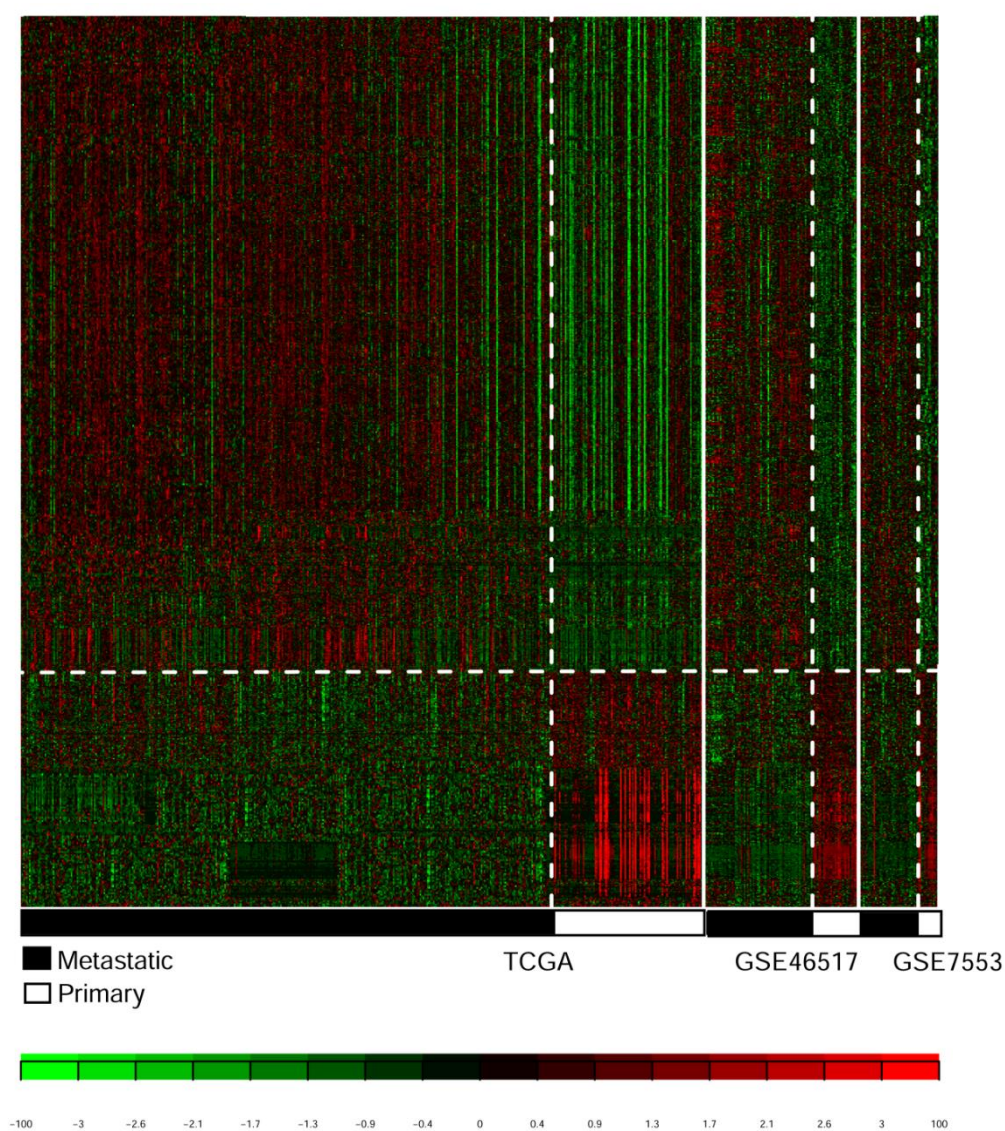
In addition, nomogram model incorporating three factors and PS status were constructed by using rms package [29] version 5.1-2 for prediction of 3- and 5-year survival probability of individuals. Nomogram is a graphic scale to visualize the contribution of each factor to survival probability prediction of an individual patient, and it has been widely applied in survival time prediction in multiple cancers [30,31], such as bladder cancer, lung cancer, renal cancer, etc.

2.7. Functional analysis of DEGs between high-risk group and low-risk group

The samples in the training set were divided into high-risk and low-risk two groups according to PS model. The DEGs between two groups were screened by using limma package version 3.34.7. FDR < 0.05 and |log2FC| > 1 were considered as screening threshold. Gene set enrichment analysis (GSEA) [32] was carried out to identify potential pathways related to risk, and $P < 0.05$ was set as criteria.

3. Results

3.1. Differentially expressed RNAs between primary and metastasis SKCMs

Figure 1. Hierarchical cluster analysis of DEGs between primary and metastasis SKCM samples. Black and white column represent primary and metastasis SKCM samples, respectively.

According to gene expression analysis, we screened numerous of differential expressed RNAs from three datasets (TCGA, GSE46517 and GSE7553). After calculated parameters (pval, FDR, tau^2, Qpval and Qval) of each gene, we finally identified 1463 differential expressed RNAs between primary and metastasis samples, including 11 lncRNAs and 1452 mRNAs. The clustering analysis results of heatmap showed these differential expressed RNAs could be clustered into the same column and rows, indicating these genes might be crucial genes related to metastasis of SKCMs (Figure 1).

3.2. Identification functional modules

While the power value was set as 14, the median connectivity of RNAs satisfied the scale-free network (Figure 2A). CutHeight was chose as 0.995, and three gene hierarchical clustering trees (dendrogram) were generated based on TCGA, GSE46517, and GSE7553 dataset, including 12 co-expression modules (Figure 2B). A heatmap were generated to visualize the associations of module and clinical traits (Figure 2C).
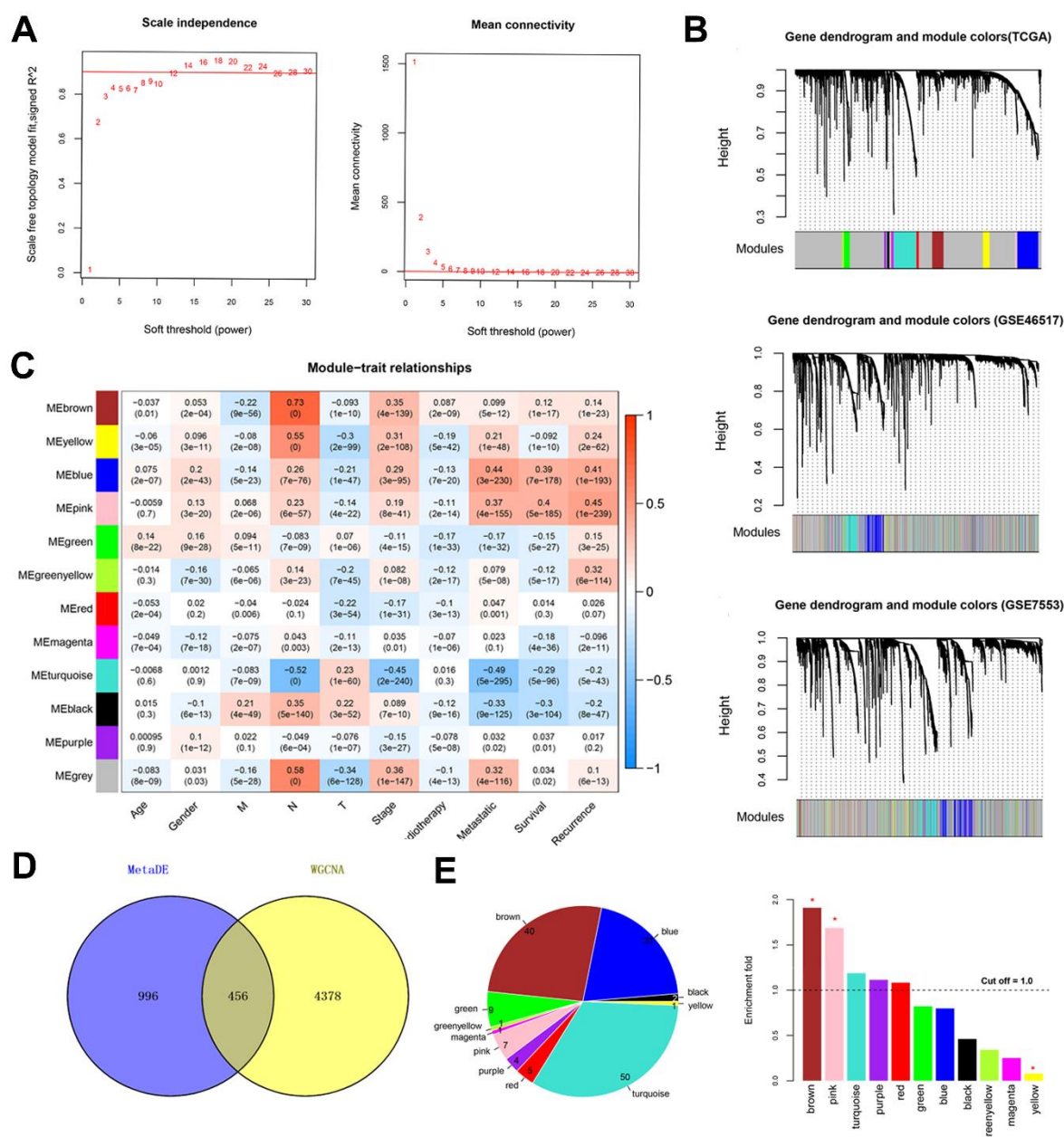
Table 2. Preservation information of TCGA, GSE46517 and GSE7553 modules.

ID	Color	Module size	Preservation infor		#DEGs	Enrichment infor	
			Z-score	P value		Enrichment fold [95%CI]	P$_{hyper}$
module 1	black	46	1.0948	5.30E-01	2	0.461 [0.0540-1.772]	4.33E-01
module 2	blue	412	26.1088	1.50E-32	31	0.798 [0.528-1.166]	2.86E-01
module 3	brown	222	13.0959	5.10E-07	40	1.909 [1.310-2.725]	7.61E-04
module 4	green	116	3.3211	6.80E-02	9	0.822 [0.364-1.630]	7.46E-01
module 5	green yellow	31	1.4561	2.20E-02	1	0.342 [0.00838-2.062]	5.19E-01
module 6	grey	3254	1.3379	3.10E-10	305	0.994 [0.851-1.159]	9.69E-01
module 7	magenta	42	5.5990	7.70E-01	1	0.252 [0.00623-0.493]	1.76E-01
module 8	pink	44	9.0744	1.30E-02	7	1.686 [0.637-3.802]	2.05E-01
module 9	purple	38	1.9483	6.70E-02	4	1.115 [0.288-3.119]	7.81E-01
module 10	red	49	2.2596	4.90E-01	5	1.082 [0.335-2.719]	8.06E-01
module 11	turquoise	446	31.0564	4.00E-35	50	1.188 [1.055-1.622]	2.80E-02
module 12	yellow	134	6.8541	3.50E-01	1	0.0791 [0.00198-0.450]	1.26E-04

Notes: where "Z score" represented the stability of modules. Modeled with Z score range from 5 to 10 represented the general stability, while Z score > 10 represented a high stability. The p value represented correlationships of modules.

Four modules with preservation Z score > 5 and *P* value < 0.05 were identified as stable modules associated with clinical traits (Table 2). We compared the differential expressed RNAs from MetaDE and WGCNA package, and found 456 overlapped genes with consistent expressed status (Figure 2D). These genes were mainly enriched in brown and turquoise modules according to the preset thresholds (Fold enrichment > 1, *P* value < 0.05), including 40 and 50 mRNAs respectively (Figure 2E).
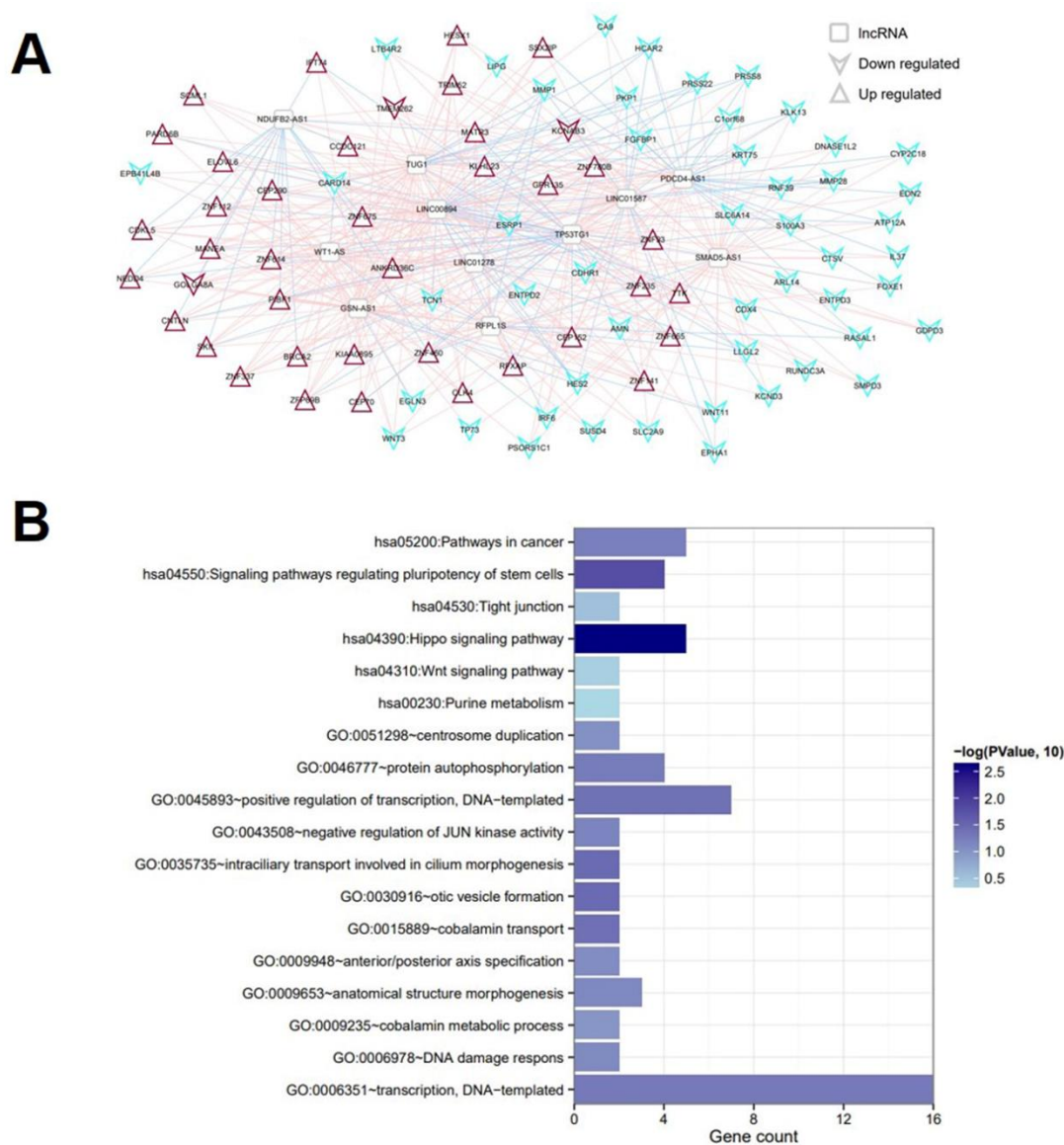
Figure 2. Weighted gene co-expression network construction. A, Scale-free network analysis. B, Division of co-expression gene modules based on TCGA, GSE46517 and GSE7553 datasets. C, Correlation matrix for each module and clinicopathological parameters. D, Venn diagram showed the overlapped genes between MetaDE analysis and WGCNA. E, Distribution of overlapped genes in each module (left) and the parameter graph of genes enriched in modules (right).

3.3. Co-expression network construction

We explore the correlation-ships of dysregulated mRNAs or lncRNAs related to SKCM prognosis. By the criteria of absolute PCC value higher than 0.4, we screened 458 pair of connections to construct network (Figure 3A). These genes were mainly related to 12 BP and 6

KEGG pathways (Table 3, Figure 3B), such as process of transcription, DNA-templated, protein autophosphorylation, DNA-templated, anatomical structure morphogenesis, pathways in cancer, hippo signaling pathway, etc.

Figure 3. Co-expression network construction A, The co-expression network of DEGs in brown and turquoise module. B, GO term and KEGG pathway enrichment analysis.

3.4. Prognostic model for establishing optimal RNAs

The whole samples in training set and validation set were divided into low- and high-risk group after calculating PS value. We screened 41 RNAs significantly correlated to prognosis of disease by univariate analysis. Multivariate analysis showed 14 RNAs with independent prognostic values were identified as hub genes, including 2 lncRNAs and 12 mRNAs. Using cox-proportional hazards model, we final selected 8 RNAs as optimal signature genes associated with SKCM prognosis, such as *RFPL1S*, *CTSV*, *EGLN3*, *ESRP1*, *HESX1*, *MANEA*, *PKP1* and *PRSS8* (Table 4).
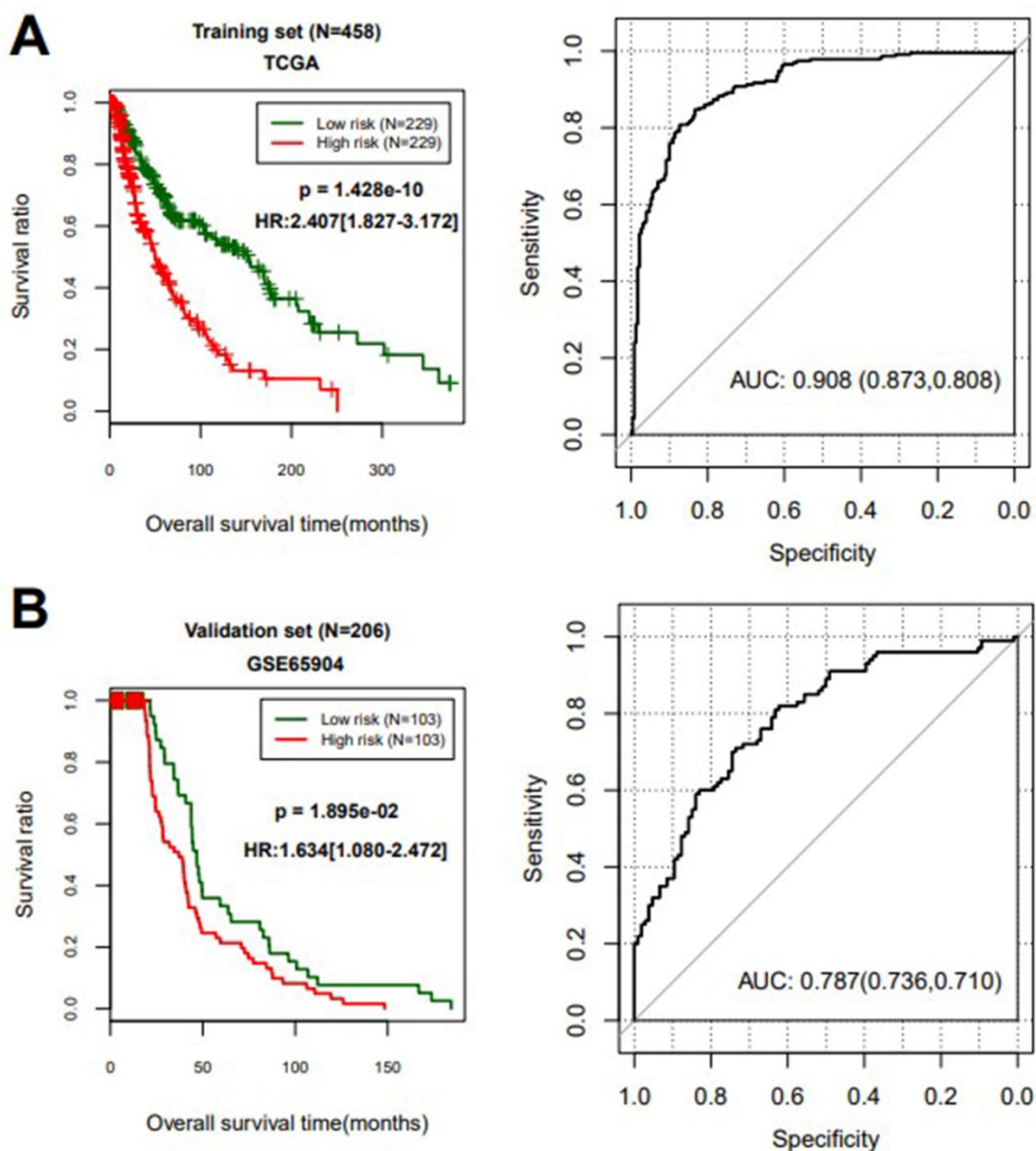
Table 3. GO terms and KEGG pathways enrichment analysis for major mRNAs in co-expression network.

Category	Term	Count	Pvalue
Biology Process	GO: 0035735~intraciliary transport involved in cilium morphogenesis	2	3.37E-02
	GO: 0030916~otic vesicle formation	2	3.37E-02
	GO: 0015889~cobalamin transport	2	3.84E-02
	GO: 0045893~positive regulation of transcription, DNA-templated	7	4.00E-02
	GO: 0006351~transcription, DNA-templated	16	4.50E-02
	GO: 0046777~protein autophosphorylation	4	4.52E-02
	GO: 0043508~negative regulation of JUN kinase activity	2	4.66E-02
	GO: 0009653~anatomical structure morphogenesis	3	4.74E-02
	GO: 0006978~DNA damage respons	2	4.75E-02
	GO: 0009948~anterior/posterior axis specification	2	4.80E-02
	GO: 0051298~centrosome duplication	2	4.89E-02
	GO: 0009235~cobalamin metabolic process	2	4.98E-02
KEGG Pathway	hsa04390: Hippo signaling pathway	5	2.29E-04
	hsa04550: Signaling pathways regulating pluripotency of stem cells	4	1.52E-03
	hsa05200: Pathways in cancer	5	5.79E-03
	hsa04530: Tight junction	2	2.82E-02
	hsa04310: Wnt signaling pathway	2	4.10E-02
	hsa00230: Purine metabolism	2	4.91E-02

Table 4. Identify the optimal RNA cohorts for prognostic prediction model.

Symbol	Type	Multi-variate Cox regression analysis			LASSO coef
		HR	95%CI	P value	
RFPL1S	lncRNA	1.424	1.018-1.990	3.89E-02	0.1846
CTSV	mRNA	1.107	1.002-1.358	3.32E-02	0.0440
EGLN3	mRNA	1.109	1.006-1.342	2.90E-02	0.0651
ESRP1	mRNA	1.122	1.033-1.268	1.64E-02	0.1092
HESX1	mRNA	0.687	0.476-0.993	4.57E-02	-0.2125
MANEA	mRNA	0.747	0.601-0.930	8.96E-03	-0.2613
PKP1	mRNA	1.241	1.080-1.426	2.36E-03	0.1473
PRSS8	mRNA	0.766	0.599-0.980	3.40E-02	-0.0959

PS value of each sample was calculated based on expression value and LASSO coefficient of eight signature genes. The solution paths and parameters of lasso regression model of 8-genes prognostic risk model were shown in Figure 4. Individuals of dataset can be divided into low- and high-risk groups. Accuracy of risk grouping was further validated by evaluating the correlation between actual survival rate and risk grouping. Patients in high risk groups exhibited a poor prognosis than those in low risk groups, indicating a significant consistency of disease grouping and actual disease prognosis (Figure 5A,B).
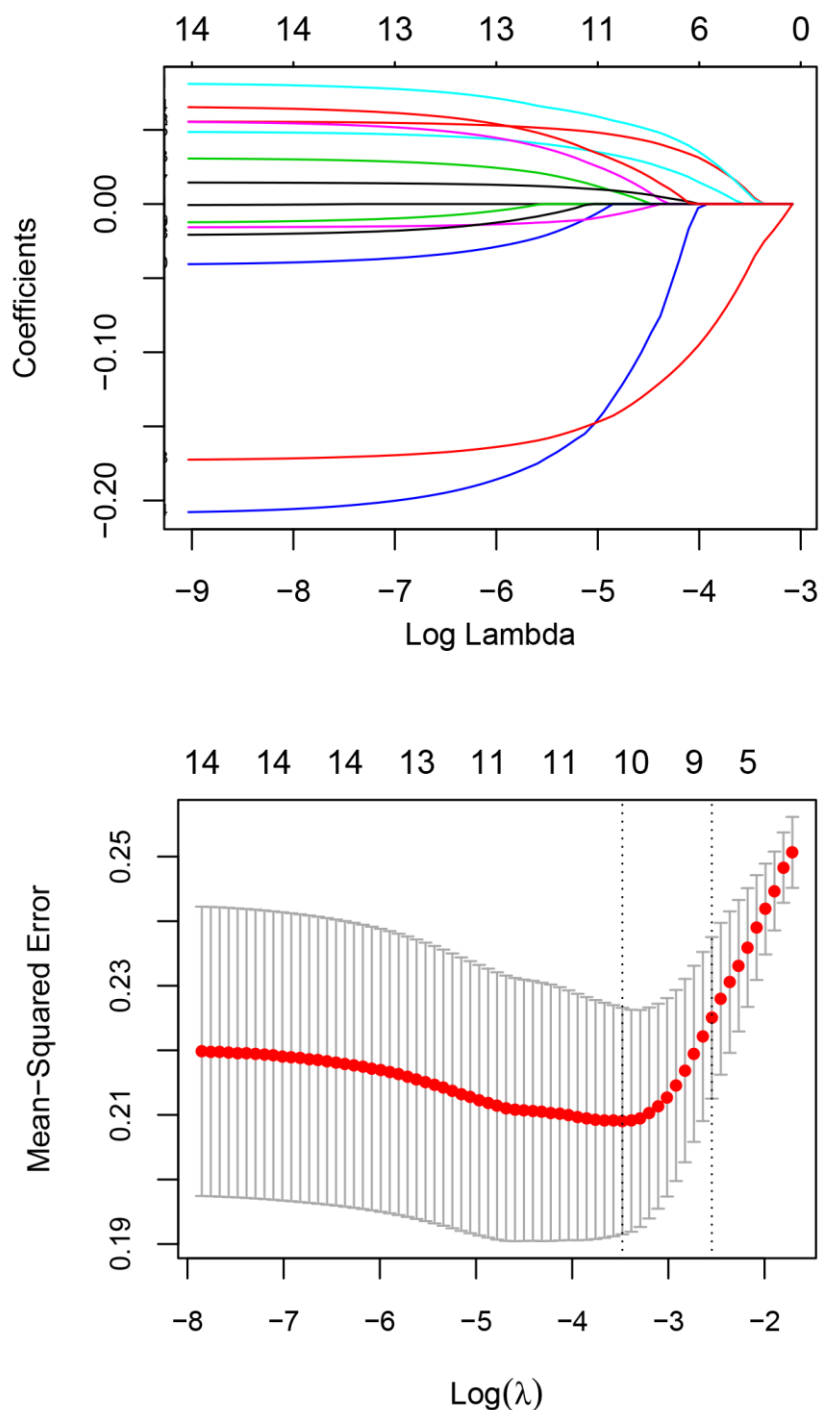
Figure 4. Kaplan-Meier curves for SKCMs in training (A) and validation set (B). Left: Overall survival time analysis based on prediction model. Right: Receiver operating characteristic curve. Data in bracket represent specificity and sensitivity of model.
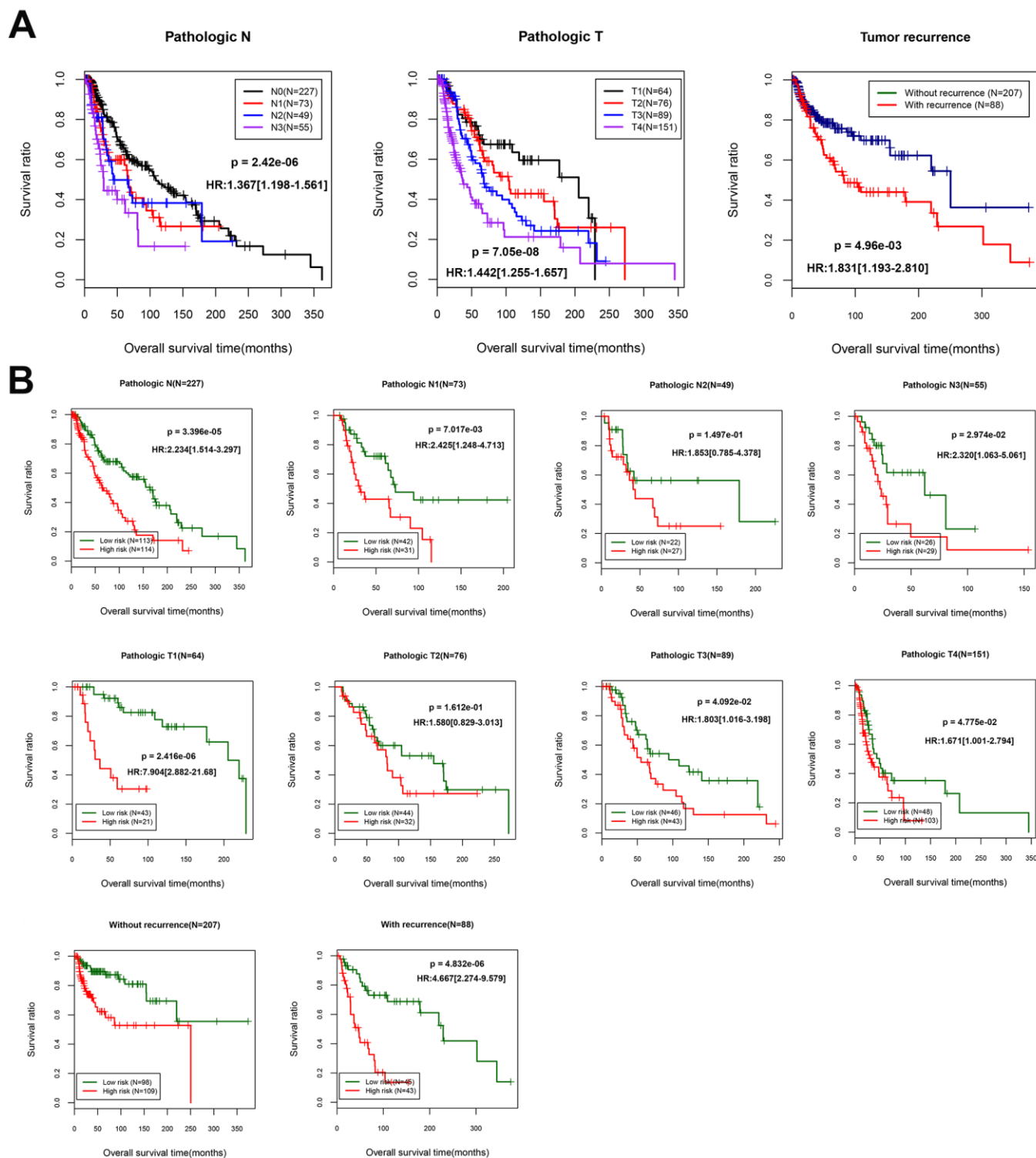
3.5. Independent prognostic clinical factors

Three clinical factors (pathologic N, T, recurrence) and PS status were identified as independent prognosis-related factors (Table 5, Table S1) according to regression analysis. Patients could be divided into differential groups according to three clinical factors. Kaplan-Meier curve demonstrated patients in groups of high-grade pathologic T, N and recurrence statue exhibited a poor prognosis. The results were consistent with actual disease progression (Figure 6A). Furthermore, individuals could be divided into differential subgroups

based on pathologic stage and recurrence status. The patients in differential subgroups were further divided into high and low risk groups according to PS value. Patients in low risk groups exhibited a better prognosis than those in high risk groups (Figure 6B).

Figure 5. LASSO profiles of the prognostic genes in SKCM. (A) LASSO coefficient profiles of the prognostic genes in SKCM. (B) Lasso deviance profiles of the prognostic genes in SKCM.

Figure 6. The results of identify prognostic clinical factors in SKCMs. A, Kaplan-Meier curves for clinical factors pathologic N, T and recurrence. B, Kaplan-Meier curves showed the correlation of risk grouping and prognosis.

Table 5. Screening independent clinical factors for survival analysis.

Clinical characteristics	TCGA (N = 458)	Uni-variables cox			Multi-variables cox		
		HR	95%CI	P	HR	95%CI	P
Age (years, mean ± sd)	58.09 ± 15.74	1.024	1.015-1.034	1.71E-07	1.008	0.992-1.025	3.19E-01
Gender (Male/Female)	284/174	1.148	0.866-1.522	3.32E-01	-	-	-
Pathologic M (M0/M1/-)	408/24/26	1.877	1.019-3.459	4.34E-02	1.478	0.439-4.971	5.28E-01
Pathologic N (N0/N1/N2/N3/-)	227/73/49/55/54	1.367	1.198-1.561	2.42E-06	1.405	1.201-1.975	4.98E-02
Pathologic T (T1/T2/T3/T4/-)	64/76/89/151/78	1.442	1.255-1.657	7.05E-08	1.412	1.100-1.811	6.75E-03
Pathologic stage (I/II/III/IV/-)	82/138/169/23/46	1.419	1.203-1.674	2.73E-05	1.395	0.775-2.513	2.67E-01
Radiotherapy (Yes/No/-)	49/408/1	0.522	0.317-0.857	8.93E-03	1.207	0.538-2.709	6.48E-01
Tumor recurrence (Yes/No/-)	88/207/163	1.831	1.193-2.810	4.96E-03	1.915	1.110-3.302	1.95E-02
Prognostic model (High/Low)	229/229	2.407	1.827-3.172	1.43E-10	3.23	1.854-5.628	3.47E-05
Death (Yes/No)	221/237	-	-	-	-	-	-
Overall survival time (months mean ± sd)	61.45 ± 64.69	-	-	-	-	-	-

3.6. Nomogram model to predict survival probability of individuals

Nomogram model incorporating four variables (pathologic N, T, recurrence and PS status) were constructed to predict 3- and 5-year survival probability of individuals (Figure 7). The total score of each factor corresponded to an estimated survival rate in nomograms. "Total Points" axis in the first row incorporates the four variables to predict prognosis of patients.

3.7. Screening of crucial genes DEGs between high-risk group and low-risk group
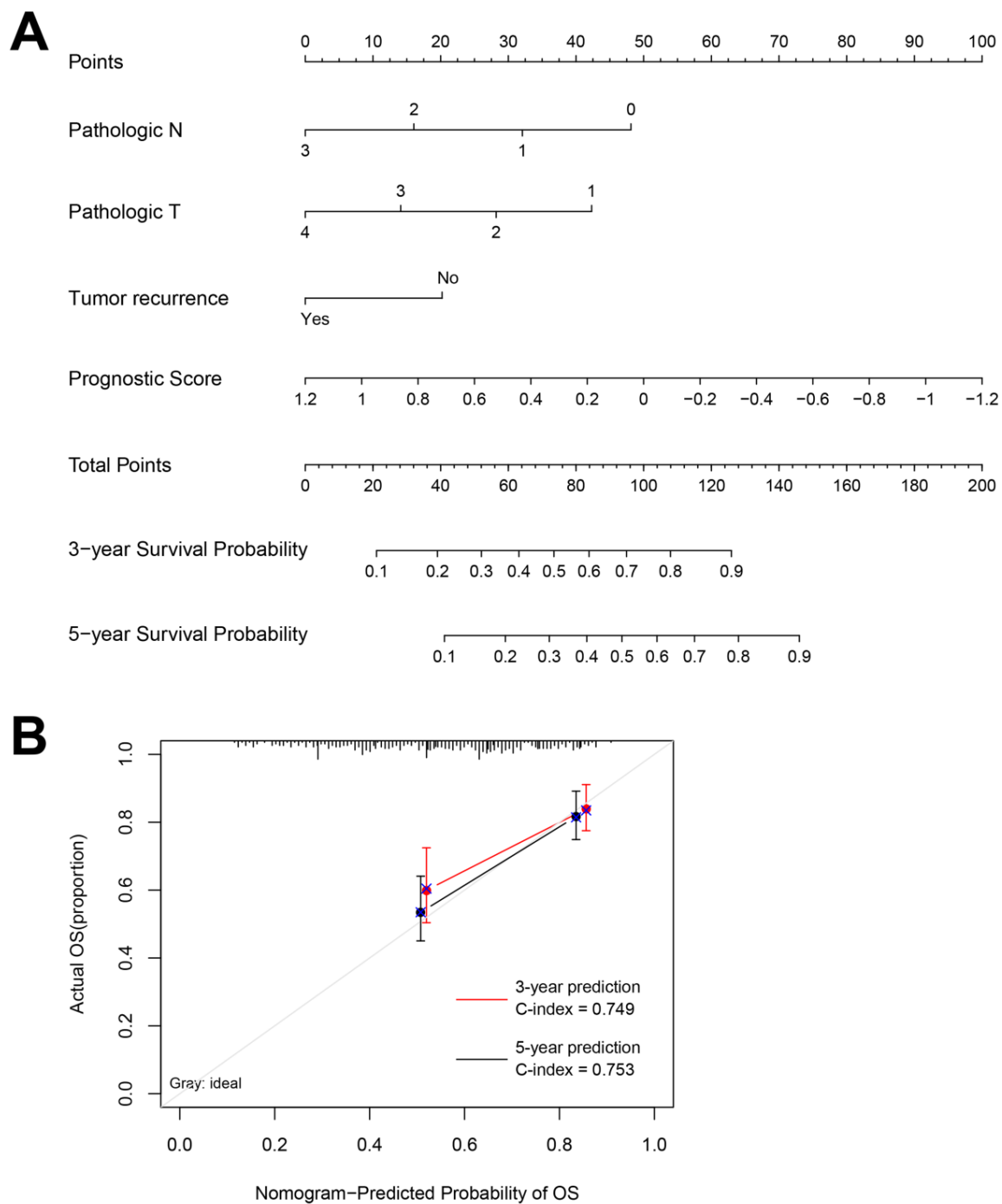
Table 6. KEGG pathways analysis for critical differential expressed genes related to risk grouping.

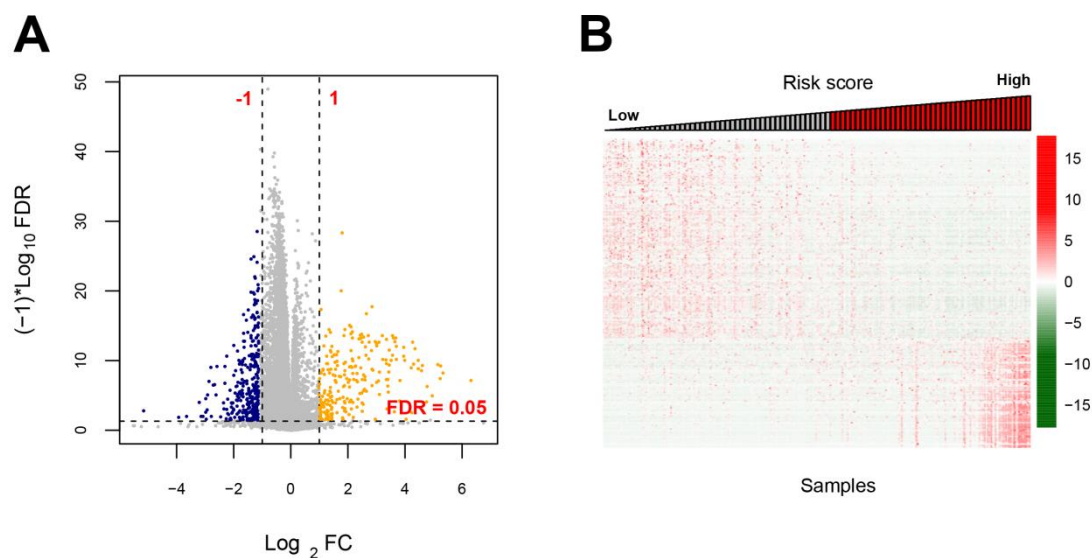| NAME	Gene count	ES	NES	NOM p-val
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	20	-0.5980	-1.8426	6.623E-03
KEGG_FOCAL_ADHESION	5	-0.7394	-1.5914	1.843E-02
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	3	-0.7673	-1.5308	3.419E-02
KEGG_JAK_STAT_SIGNALING_PATHWAY	7	-0.5553	-1.5085	3.601E-02
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	25	-0.4286	-1.4432	3.915E-02
KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	6	-0.7082	-1.4858	4.270E-02
KEGG_MELANOGENESIS	3	0.8835	1.6179	6.012E-03
KEGG_TYROSINE_METABOLISM	2	0.8191	1.4877	3.184E-02

Samples of TCGA set were divided into two groups after calculating PS values. We finally obtained 662 DEGs (including 427 downregulated and 235 upregulated genes) between the high-risk and low-risk groups, and the volcanic diagram was shown in Figure 8A. The heatmap incorporate the gene expression level and Risk Score to visualize the distribution of crucial genes (Figure 8B).

GSEA results revealed these DEGs were mainly related to eight pathways, such as melanogenesis (Table 6, Figure 9), JAK-STAT signaling pathway, cytokine receptor interaction, neuroactive ligand receptor interaction, etc.
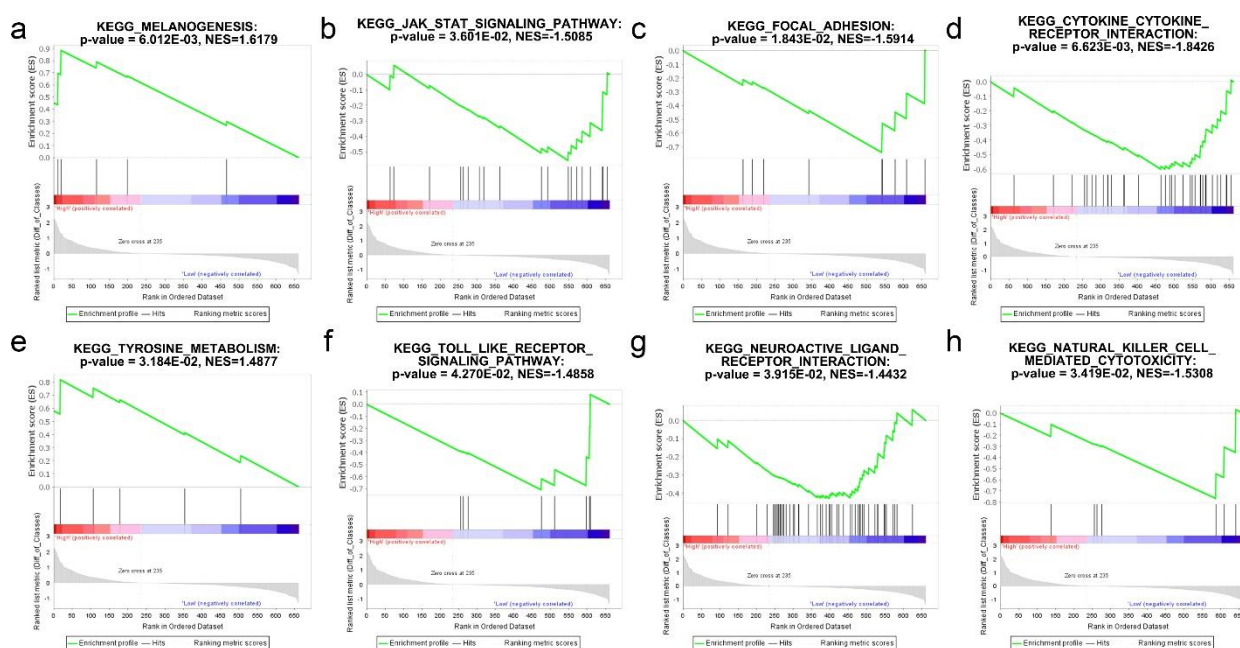
Figure 7. Nomogram for predicting survival probability in the TCGA data set. A, A nomogram model of independent prognostic factor for survival probability prediction. B, Calibration curve for the nomogram predicting overall survival.

Figure 8. Screen DEGs between high-risk group and low-risk group based on TCGA training set. A, Volcanic diagram showed the DEGs of high and low risk groups. Green and orange dots represent DEGs, while gray represents general genes. B, Heatmap showed the different group of DEGs changing from low to high risk score.

Figure 9. Enrichment map of KEGG pathways significantly related to risk grouping.

4. Discussion

In this study, a total of 1463 DEGs were screened between primary tumor and metastatic samples based on three datasets, TCGA, GSE46517 and GSE7553. We conducted WGCNA and identified two stable functional modules. Functional enrichment analysis revealed module-related

DEGs were associated with pathways in cancer, hippo signaling pathway, regulating pluripotency of stem cells. After PS model and nomogram model construction, we identified 8 optimal signature RNAs (such as lncRNA-*RFPL1S*, *CTSV*, *ESRP1*, etc.) and three independent clinical factors (Pathologic N, T and Recurrence) for prognosis prediction of individuals.

In this study, a lncRNA-mRNA network was constructed by differentially expressing lncRNAs and DEGs. In order to elucidate the underlying mechanisms, pathway enrichment analysis was performed on these lncRNAs and DERs. The results of the study showed that these DEGs were significantly associated with the Hippo signaling pathway, Signaling pathways regulating pluripotency of stem cells and Pathways in cancer. The Hippo pathway is a critical regulator of organ growth and cell fate that is dysregulated in many cancers [33]. Cancer malignancy has been linked to distinct subsets of stem-like cells, known as cancer stem cells, which persist during treatment and appear to lead to drug-resistant recurrence [34]. Besides, melanomas develop from melanoma-competent melanocyte stem cells in response to UVB stimulation, which causes melanocyte stem cell activation and translocation through an inflammation-dependent process [35]. These findings indicate that the Hippo signaling pathway, Signaling pathways regulating pluripotency of stem cells and Pathways in cancer are important for SKCM, and that disease-related RNAs RNAs are involved in the regulation of these pathways in SKCM.

We explored the potential biological function of eight signature genes by searching the published papers. lncRNAs are types of RNAs molecules without protein encoding function, and have been reported involved in the tumorgenesis and development. *LncRNAs-RFPL1S* is antisense to *RFPL1* gene and may function in the post-transcriptional regulation of this gene [36]. *RFPL* transcripts encode proteins with tripartite structure and it has been reported regulating cell-cycle progression through cyclin B1/Cdc2 degradation [37]. A recent study showed *RFPL1S* were down-regulated in brain tumors samples and expression patterns of this lncRNA were correlated with malignancy grade in gliomas [38].

CTSV or cathepsin L2 is a lysosomal cysteine proteinase that associated with extracellular-matrix degradation and tumor progress. High expression of *CTSV* is reported to involved in metastasis and poor prognosis in breast ductal cancer, and it is a potential biomarker for prognosis prediction breast cancer [39]. Moreover, two gene *CTSV* and *CTSC* were involved in invasion of RCC, and a recent study showed praeruptorin B could promote the metastatic ability of RCC cells through targeting *CTSC* and *CTSV* expression [40]. In addition, ELGN3 or PHD3 is a member of PHDs, which induced in hypoxia. The biological role of PHD3 was related to HIF-1α hydroxylation, suppression of tumor angiogenesis under hypoxic conditions [41,42]. Further study demonstrated the tumor inhibition role of EGLN3, and it could regulates p53 stability by hydroxylating proline 359 and resulted to cell cycle arrest and apoptosis [43]. In murine melanoma model, treatment of PHD3 inhibitor could decreases tumor growth and angiogenesis through increasing sVEGFR-1 generation from tumor-associated macrophages, indicating a critical role PHD3 in tumor progression [44]. Whereas, PHD3 might function as the tumor promotion factor in some tumors. By integrative -omics and HLA-ligand omics analysis, Anna identified that EGLN3-derived peptides were associated with higher infiltration of tumor by CD8+ T cells; functional analyses revealed EGLN3 might play the role of pro-proliferative and anti-apoptotic in several RCC cell lines [45]. Moreover, EGLN3 were also elaborated that related to chemotherapy resistance of cancer, such as prostate cancer and pheochromocytoma [46,47].

Furthermore, protein of ESRP1 belong to RNA-binding proteins family and it involved in tumor

progression through regulating of gene posttranscriptional process [48]. Studies have shown that downregulation of ESRPs was associated with tumor infiltration in various type of cancers, including prostate cancer, breast cancer and pancreatic cancers [49,50]. Recently, ESRP1 was identified as potential prognostic biomarker in cutaneous malignant melanoma [51], and patients with higher ESRP1 level had a poor overall survival rate than those with lower ESRP1 level; the further analysis showed ESRP1 was negatively associated with infiltration of DC and Treg cells. It is well known that DC can promote tumor metastasis by up-regulating Treg cells and down-regulating the cytotoxicity of CD8+ T cells [52]. In fact, data analysis in our study also identified dysregulated ESRP1 from metastatic SKCM sample, which consistent with previous studies.

In addition, PKP1/2/3 played an potential role in tumor invasion and metastasis in various malignancy and *PKP1* mutation could result in skin fragility syndrome [53,54]. Lee et al reported that phosphorylation of Pkp1 by RIPK4 (receptor-interacting serine-threonine kinase 4) regulated epidermal differentiation and skin tumorigenesis [55]. Moreover, the serine protease PRSS8 could suppress colorectal carcinogenesis and metastasis [56]. Hypermethylated PRSS8 in ESCC tissues was linked to the downregulation of PRSS8; The reduction of PRSS8 was well correlated with shorter survival time in ESCC patients [57]. High levels of PRSS8 and prostasin has been identified as potential clinical biomarkers for ovarian cancer early detection [58]. Taken together, these findings indicated that protein of CTSV, ESRP1, ELGN3 might be critical genes associated with tumor metastasis, and the eight genes may be reliable biomarkers for predictive of SKCM prognosis.

Univariate and multivariate cox regression analysis revealed three clinicopathological variables and PS status were correlated to prognosis of SKCMs. Nomogram model incorporating these factors to predict 3- and 5-year survival probability of individuals. Previous studies showed this model has been widely used to provide prognosis prediction for cancer patients. A recent study showed six clinical factors were significantly correlated with sentinel node status, which is major factor of melanoma prognosis; six variables-based nomogram model exhibited a good predictive performance and suggested to be a decision aid for T1 melanoma being considered sentinel node biopsy [59]. Using nomogram model, another paper identified five immune-related genes and several clinical parameters for prognosis prediction of SKCMs [60], such as pathological T, N, AJCC stage etc.. In our study, we selected common clinical factors of SKCMs, and these clinicopathologic parameters were relative easily available and comprehensive in clinical work. Moreover, our nomogram exhibited a better discrimination ability for predicting prognosis.

The potential limitation of this study should be enumerated. Firstly, individual number was small and the whole data analysis were conducted on the basis of public databases (TCGA and GEO), thus more external samples from multiple medical centers should be concerned validate the performance of prediction models. Secondly, the function of optimal RNAs also needed to be further investigated in SKCM not only by data analysis but also by molecular and cellular experiments.

5. Conclusions

In conclusion, we screened cohorts of overlapping genes related to SKCM metastasis by using DEGs analysis and WGCNA. On the basis of PS model, we identified eight signature genes and pathologic N, T and recurrence as prognostic factors for disease prediction. Furthermore, eight signature genes based predictive model and nomogram might be useful methods for prognostic prediction of SKCM patients.

Conflict of interest

The authors report no conflicts of interest.

References

1. D. Burns, J. George, D. Aucoin, J. Bower, N. Bower, The pathogenesis and clinical management of cutaneous melanoma: an evidence-based review, *J. Med. Imaging Radiat. Sci.*, **50** (2019), 460–469.

2. R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, *CA. Cancer J. Clin.*, **70** (2020), 7–30.

3. T. Crosby, R. Fish, B. Coles, M. Mason, Systemic treatments for metastatic cutaneous melanoma, *Cochrane Database Syst. Rev.*, **2** (2018), CD001215.

4. L. C. van Kempen, M. Redpath, C. Robert, A. Spatz, Molecular pathology of cutaneous melanoma, *Melanoma Manag.*, **1** (2014), 151–164.

5. C. Lugassy, S. Zadran, L. A. Bentolila, M. Wadehra, R. Prakash, S. T. Carmichael, et al., Angiotropism, pericytic mimicry and extravascular migratory metastasis in melanoma: an alternative to intravascular cancer dissemination, *Cancer Microenviron.*, **7** (2014), 139–152.

6. S. L. V. Es, M. Colman, J. F. Thompson, S. W. McCarthy, R. A. Scolyer, Angiotropism is an independent predictor of local recurrence and in-transit metastasis in primary cutaneous melanoma, *Am. J. Surg. Pathol.*, **32** (2008), 1396–1403.

7. L. Mervic, Time course and pattern of metastasis of cutaneous melanoma differ between men and women, *PLoS One.*, **7** (2012), e32955.

8. N. R. Adler, A. Haydon, C. A. McLean, J. W. Kelly, V. J. Mar, Metastatic pathways in patients with cutaneous melanoma, *Pigment Cell Melanoma Res.*, **30** (2017), 13–27.

9. I. J. Fiddler, Melanoma metastasis, *Cancer Control*, **2** (1995), 398–404.

10. C. Haqq, M. Nosrati, D. Sudilovsky, J. Crothers, D. Khodabakhsh, B. L. Pulliam, et al., The gene expression signatures of melanoma progression, *Proc. Natl. Acad. Sci. U. S. A.*, **102** (2005), 6092–6097.

11. S. Mandruzzato, A. Callegaro, G. Turcatel, S. Francescato, M. C. Montesco, V. Chiarion-Sileni, et al., A gene expression signature associated with survival in metastatic melanoma, *J. Transl. Med.*, **4** (2006), 1479–5876.

12. B. Huang, W. Han, Z. F. Sheng, G. L. Shen, Identification of immune-related biomarkers associated with tumorigenesis and prognosis in cutaneous melanoma patients, *Cancer Cell Int.*, **20** (2020), 020–01271.

13. M. Liao, F. Zeng, Y. Li, Q. Gao, M. Yin, G. Deng, et al., A novel predictive model incorporating immune-related gene signatures for overall survival in melanoma patients, *Sci. Rep.*, **10** (2020), 12462.

14. O. Kabbarah, C. Nogueira, B. Feng, R. M. Nazarian, M. Bosenberg, M. Wu, et al., Integrative genome comparison of primary and metastatic melanomas, *PLoS One*, **5** (2010), 0010770.

15. A. I. Riker, S. A. Enkemann, O. Fodstad, S. Liu, S. Ren, C. Morris, et al., The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis, *BMC Med. Genomics*, **1** (2008), 1755–8794.

16. H. Cirenajwis, H. Ekedahl, M. Lauss, K. Harbst, A. Carneiro, Molecular stratification of metastatic melanoma using gene expression profiling : Prediction of survival outcome and benefit from molecular targeted therapy, *Oncotarget*, **6** (2015), 12297–12309.

17. R. Cabrita, M. Lauss, A. Sanna, M. Donia, G. Jönsson, Tertiary lymphoid structures improve immunotherapy and survival in melanoma, *Nature*, **577** (2020), 561–565.

18. V. Nicolaidou, C. Papaneophytou, C. Koufaris, Detection and characterisation of novel alternative splicing variants of the mitochondrial folate enzyme MTHFD2, *Mol. Biol. Rep.*, **47** (2020), 1–8.

19. C. Qi, L. Hong, Z. Cheng, Q. Yin, Identification of metastasis-associated genes in colorectal cancer using metaDE and survival analysis, *Oncol. Lett.*, **11** (2015), 568–574.

20. X. Wang, D. D. Kang, K. Shen, C. Song, S. Lu, L. C. Chang, et al., An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection, *Bioinformatics*, **28** (2012), 2534–2536.

21. X. Zhai, Q. Xue, Q. Liu, Y. Guo, Z. Chen, Colon cancer recurrenceassociated genes revealed by WGCNA coexpression network analysis, *Mol. Med. Rep.*, **16** (2017), 6499–6505.

22. P. Langfelder and S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.*, **9** (2008), 1471–2105.

23. J. Cao, S. Zhang, A Bayesian extension of the hypergeometric test for functional enrichment analysis, *Biometrics.*, **70** (2014), 84–94.

24. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13** (2003), 2498–2504.

25. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc.*, **4** (2009), 44–57.

26. P. Wang, Y. Wang, B. Hang, X. Zou, J. H. Mao, A novel gene expression-based prognostic scoring system to predict survival in gastric cancer, *Oncotarget*, **7** (2016), 55343–55351.

27. R. Tibshirani, The lasso method for variable selection in the Cox model, *Stat. Med.*, **16** (1997), 385–395.

28. J. J. Goeman, L1 penalized estimation in the Cox proportional hazards model, *Biom. J.*, **52** (2010), 70–84.

29. K. H. Eng, E. Schiller, K. Morrel, On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve, *Oncotarget*, **6** (2015), 36308–36318.

30. W. Liang, L. Zhang, G. Jiang, Q. Wang, J. He, Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer, *J. Clin. Oncol.*, **33** (2015), 861–869.

31. C. Zhang, F. Wang, F. Guo, C. Ye, B. Yang, A 13-gene risk score system and a nomogram survival model for predicting the prognosis of clear cell renal cell carcinoma, *Urol. Oncol.*, **38** (2020), 74.e1–74.e11.

32. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.*, **102** (2005), 15545–15550.

33. X. Zhang, L. Yang, P. Szeto, G. K. Abali, Y. Zhang, A. Kulkarni, et al., The Hippo pathway oncoprotein YAP promotes melanoma cell invasion and spontaneous metastasis, *Oncogene*, **39** (2020), 5267–5281.

34. Z. Kozovska, V. Gabrisova and L. Kucerova, Malignant melanoma: diagnosis, treatment and cancer stem cells, *Neoplasma*, **63** (2016), 510–517.

35. H. Moon, L. R. Donahue, E. Choi, P. O. Scumpia, W. E. Lowry, J. K. Grenier, et al., Melanocyte Stem Cell Activation and Translocation Initiate Cutaneous Melanoma in Response to UV Exposure, *Cell Stem. Cell*, **21** (2017), 665–678.

36. E. Seroussi, D. Kedra, H. Q. Pan, M. Peyrard, C. Schwartz, P. Scambler, et al., Duplications on human chromosome 22 reveal a novel Ret Finger Protein-like gene family with sense and endogenous antisense transcripts, *Genome Res.*, **9** (1999), 803–814.

37. J. Bonnefont, T. Laforge, O. Plastre, B. Beck, S. Sorce, C. Dehay, et al., Primate-specific RFPL1 gene controls cell-cycle progression through cyclin B1/Cdc2 degradation, *Cell Death Differ.*, **18** (2011), 293–303.

38. X. Zhang, S. Sun, J. K. Pu, A. C. Tsang, D. Lee, V. O. Man, et al., Long non-coding RNA expression profiles predict clinical phenotypes in glioma, *Neurobiol. Dis.*, **48** (2012), 1–8.

39. M. Toss, I. Miligy, K. Gorringe, K. Mittal, R. Aneja, I. Ellis, et al., Prognostic significance of cathepsin V (CTSV/CTSL2) in breast ductal carcinoma in situ, *J. Clin. Pathol.*, **73** (2020), 76–82.

40. C. -L. Lin, T. -W. Hung, T. -H. Ying, C. -J. Lin, Y. -H. Hsieh, C. -M. Chen, Praeruptorin B mitigates the metastatic ability of human renal carcinoma cells through targeting CTSC and CTSV expression, *Int. J. Mol. Sci.*, **21** (2020), 2919.

41. Q. L. Liu, Q. L. Liang, Z. Y. Li, Y. Zhou, W. T. Ou, Z. G. Huang, Function and expression of prolyl hydroxylase 3 in cancers, *Arch Med. Sci.*, **9** (2013), 589–593.

42. N. Pescador, Y. Cuevas, S. Naranjo, M. Alcaide, D. Villar, M. O. Landázuri, et al., Identification of a functional hypoxia-responsive element that regulates the expression of the egl nine homologue 3 (egln3/phd3) gene, *Biochem. J.*, **390** (2005), 189–197.

43. J. Rodriguez, A. Herrero, S. Li, N. Rauch, A. Quintanilla, K. Wynne, et al., PHD3 regulates p53 protein stability by hydroxylating proline 359, *Cell Rep.*, **24** (2018), 1316–1329.

44. J. M. Roda, Y. Wang, L. A. Sumner, G. S. Phillips, C. B. Marsh, T. D. Eubank, Stabilization of HIF-2α induces sVEGFR-1 production from tumor-associated macrophages and decreases tumor growth in a murine melanoma model, *J. Immunol.*, **189** (2012), 3168–3177.

45. A. Reustle, M. Di Marco, C. Meyerhoff, A. Nelde, J. S. Walz, S. Winter, et al., Integrative -omics and HLA-ligandomics analysis to identify novel drug targets for ccRCC immunotherapy, *Genome Med.*, **12** (2020), 32–32.

46. Y. Wang, X. Li, W. Liu, B. Li, D. Chen, F. Hu, et al., MicroRNA-1205, encoded on chromosome 8q24, targets EGLN3 to induce cell growth and contributes to risk of castration-resistant prostate cancer, *Oncogene*, **38** (2019), 4820–4834.

47. S. Li, J. Rodriguez, W. Li, P. Bullova, S. M. Fell, O. Surova, et al., EglN3 hydroxylase stabilizes BIM-EL linking VHL type 2C mutations to pheochromocytoma pathogenesis and chemotherapy resistance, *Proc. Natl. Acad. Sci. U. S. A.*, **116** (2019), 16997–17006.

48. T. W. Bebee, J. W. Park, K. I. Sheridan, C. C. Warzecha, B. W. Cieply, A. M. Rohacek, et al., The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development, *Elife*, **15** (2015), 08954.

49. K. Horiguchi, K. Sakamoto, D. Koinuma, K. Semba, A. Inoue, S. Inoue, et al., TGF-β drives epithelial-mesenchymal transition through δEF1-mediated downregulation of ESRP, *Oncogene*, **31** (2012), 3190–3201.

50. J. Ueda, Y. Matsuda, K. Yamahatsu, E. Uchida, Z. Naito, M. Korc, et al., Epithelial splicing regulatory protein 1 is a favorable prognostic factor in pancreatic cancer that attenuates pancreatic metastases, *Oncogene*, **33** (2014), 4485–4495.

51. B. Wang, Y. Li, C. Kou, J. Sun, X. Xu, Mining database for the clinical significance and prognostic value of ESRP1 in cutaneous malignant melanoma, *Biomed. Res. Int.*, **5** (2020), 4985014.

52. A. Sawant, J. A. Hensel, D. Chanda, B. A. Harris, G. P. Siegal, A. Maheshwari, et al., Depletion of plasmacytoid dendritic cells inhibits tumor growth and prevents bone metastasis of breast cancer cells, *J. Immunol.*, **189** (2012), 4258–4265.

53. A. E. Boyce, J. A. McGrath, T. Techanukul, D. F. Murrell, C. W. Chow, L. McGregor, et al., Ectodermal dysplasia-skin fragility syndrome due to a new homozygous internal deletion mutation in the PKP1 gene, *Australas. J. Dermatol.*, **53** (2012), 61–65.

54. I. Hofmann, Plakophilins and their roles in diseased states, *Cell Tissue Res.*, **379** (2020), 5–12.

55. P. Lee, S. Jiang, Y. Li, J. Yue, X. Gou, S. Y. Chen, et al., Phosphorylation of Pkp1 by RIPK4 regulates epidermal differentiation and skin tumorigenesis, *Embo. J.*, **36** (2017), 1963–1980.

56. Y. Bao, Y. Guo, Y. Yang, X. Wei, S. Zhang, Y. Zhang, et al., PRSS8 suppresses colorectal carcinogenesis and metastasis, *Oncogene*, **38** (2019), 497–517.

57. Y. Bao, Q. Wang, Y. Guo, Z. Chen, K. Li, Y. Yang, et al., PRSS8 methylation and its significance in esophageal squamous cell carcinoma, *Oncotarget*, **7** (2016), 28540–28555.

58. A. Tamir, A. Gangadharan, S. Balwani, T. Tanaka, U. Patel, A. Hassan, et al., The serine protease prostasin (PRSS8) is a potential biomarker for early detection of ovarian cancer, *J. Ovarian Res.*, **9** (2016), 016–0228.

59. A. Maurichi, R. Miceli, H. Eriksson, J. Newton-Bishop, J. Nsengimana, M. Chan, et al., Factors affecting sentinel node metastasis in thin (T1) cutaneous melanomas: development and external validation of a predictive nomogram, *J. Clin. Oncol.*, **38** (2020), 1591–1601.

60. B. Hu, Q. Wei, C. Zhou, M. Ju, L. Wang, L. Chen, et al., Analysis of immune subtypes based on immunogenomic profiling identifies prognostic signature for cutaneous melanoma, *Int. Immunopharmacol.*, **6** (2020), 107162.