CA9 is an important molecular marker for the prognosis of osteosarcoma patients

Jiguang Chen
Changhai Hospital,Second Military Medical University
Jian Wang
Shanghai Pudong New Area People's Hospital
An Wang
Shanghai Pudong New Area People's Hospital
Jiangming Yu (18918803141@163.com)
Changzheng Hospital,Second Military Medical University https://orcid.org/0000-0001-8047-7439

Research article

Keywords: osteosarcoma, differential expression analysis, risk score system, stratification analysis, pathway enrichment analysis

Posted Date: May 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-28407/v1

Abstract

Background

As a common malignant bone tumor, osteosarcoma (OS) progresses rapidly and recurs easily. This study is aimed to build a risk score system for OS patients.

Methods

From The Cancer Genome Atlas and Gene Expression Omnibus databases, the RNA-seq data of OS (the training set) and GSE39055 (the validation sets) separately were obtained. Combined with limma package, the differentially expressed lncRNAs (DE-lncRNAs) and mRNAs (DE-mRNAs) between recurrence and non-recurrence groups were analyzed. After the RNAs correlated with independent prognosis were screened using survival package, risk score systems were constructed and the optimal one was selected. For the independent clinical prognostic factors identified by survival package, stratification analysis was conducted. Using Gene Set Enrichment Analysis, pathways were enriched for the differentially expressed genes (DEGs) between high and low risk groups.

Results

For recurrence and non-recurrence groups, 319 DE-mRNAs and 14 DE-lncRNAs were identified. Subsequently, 10 DE-mRNAs (including *ALDH1A1*, *CA9*, *GMDS*, *LCMT2*, *LRRC75A*, *METTL1*, *RAB29*, *TADA2B*, *TDRD7*, and *TIGD2*) and eight DE-lncRNAs were found to be correlated with independent prognosis. From the four risk score systems, the mRNA expression status-based risk score system was selected as the optimal one. Among the three independent clinical prognostic factors, age and recurrence were significantly related to overall survival in high risk group. Additionally, vascular smooth muscle contraction, and glycine, serine and threonine metabolism were enriched for the DEGs between high and low risk groups.

Conclusion

The mRNA expression status-based risk score system might be devoted to predict the prognosis of OS patients.

Background

As the most common malignancy of the skeletal system, osteosarcoma (OS) derives from stromal cells and is capable of producing osteoid tissues [1]. OS is more common in adolescents and occurs in the metaphysis of the long bone with abundant blood supply [2]. OS develops rapidly and is prone to recurrence and metastasis [3]. OS takes up about 5% of pediatric tumors, and most patients die within a year of diagnosis [4, 5]. Although the emergence of neoadjuvant chemotherapy greatly improves the 5-year survival rate of OS patients, the recurrence rate of OS is still high [6]. Therefore, the underlying mechanisms of OS should be deeply studied to improve its treatment and prognosis.

Along with the advances in molecular biology, some genes and long non-coding RNAs (lncRNAs) involved in OS have been revealed. For example, migration-inducing gene 7 (*MIG-7*) is of great significance for mediating vasculogenic mimicry formation in OS, which can be used as an independent adverse prognostic marker in OS patients [7]. The overall survival and the metastasisfree survival of the OS patients with low CC motif chemokine ligand 5 (*CCL5*) expression are higher than those with high *CCL5* expression, therefore, *CCL5* may be a prognostic biomarker and therapeutic target in OS [8, 9]. Increased nuclear factor-κB (*NF-κB*) and decreased phosphatase and tensin homolog deleted in chromosome 10 (*PTEN*) in OS are inversely correlated, which is important for the early diagnosis and prognosis of the tumor [10]. Overexpression of the lncRNA highly up-regulated in liver cancer (*HULC*) in OS is related to clinical stage and metastasis, suggesting that *HULC* may be applied for the prognosis prediction and targeted therapy of the disease [11]. The lncRNA carbonyl reductase 3-antisense RNA 1 (*CBR3-AS1*) exerts an oncogenic effect in OS through mediating cell proliferation, invasion, migration, and apoptosis, which functions as an independent poor prognostic factor in OS [12]. However, the prognostic mechanisms of OS have not been totally reported.

LncRNA is as important as the mRNA involved in the coding, which may be used as a biomarker to judge or predict the prognosis of human diseases [13, 14]. In this study, mRNA and lncRNA expression characteristics in patients with recurrent OS were comprehensively compared and analyzed. Besides, the mRNAs and lncRNAs closely related to the recurrence of OS were selected. Afterwards, different types of risk score systems were constructed and the optimal risk score system was screened. Moreover, stratification analysis and pathway enrichment analysis were carried out. This study might help to predict the prognosis of OS patients.

Materials And Methods

Data downloading and data preprocessing

From The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) database, the RNA-seq data of OS (downloaded in July 15, 2019; including both mRNA-seq and lncRNA-seq data; 265 OS samples; platform: Illumina HiSeq 2000 RNA Sequencing) was extracted. Among the 265 OS samples, 169 OS samples with survival prognosis information and clear information of recurrence or not were screened as the training set.

From Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database, eligible datasets were searched with “osteosarcoma”, “recurrence”, and “Homo sapiens” as the key words. The selection criteria were: (1) the OS samples had clear information of recurrence or not (recurrence/non-recurrence); (2) the OS samples had survival prognosis information. Finally, only GSE39055 (including 37 OS samples (18 recurrence samples and 19 non-recurrence

samples) with survival prognosis information; platform: Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip) was obtained as the validation set. The raw expression data of the two datasets were downloaded, and then the expression data was normalized using the unit-scale normalization algorithm [15].

Differential expression analysis

Combined with the reference sequence (RefSeq) identification numbers (IDs) of the downloaded data, the mRNAs and lncRNAs shared by the two datasets were identified based on the annotation information in HUGO Gene Nomenclature Committee (HGNC, http://www.genenames.org/) database [16].

Based on the information of recurrence or not, the OS samples in the training set were divided into recurrence and non-recurrence groups. Using the R package limma (version 3.34.7, https://bioconductor.org/packages/release/bioc/html/limma.html) [17], the differentially expressed RNAs (DERs, including both differentially expressed lncRNAs (DE-lncRNAs) and differentially expressed mRNAs (DE-mRNAs)) between recurrence and non-recurrence groups were screened. The $|\log_2$ fold change (FC)$| > 0.263$ and false discovery rate (FDR) < 0.05 were selected as the cut-off criteria. Based on the centered Pearson correlation algorithm in the R package pheatmap (version 1.0.8, https://cran.r-project.org/web/packages/pheatmap/index.html) [18], bidirectional hierarchical clustering for the expression values of the DERs were performed.

Construction of risk score systems

Using the univariate Cox regression analysis in the survival package (version 2.41-1, http://bioconductor.org/packages/survivalr/) [19] in R, the DERs significantly correlated with overall survival were identified combined with the survival prognosis information of the OS samples in the training set. Subsequently, the DERs associated with independent prognosis were further selected using the multivariate Cox regression analysis in the survival package [19]. The log-rank p-value < 0.05 was the significant threshold.

After the prognostic coefficients of the independent prognosis-associated DE-lncRNAs and DE-mRNAs were obtained, the risk score systems were constructed. With the Monte-Carlo p-value < 0.05 as the criterion, the cutoff values of the expression levels of the independent prognosis-associated DE-lncRNAs and DE-mRNAs were selected using X-Tile Bio-Informatics Tool (https://medicine.yale.edu/lab/rimm/research/software.aspx) [20]. The status of each sample in the expression of the RNA was defined based on the cutoff value of each RNA. When the expression level of the RNA was higher than its cutoff value, status was defined as 1. When the expression level of the RNA was lower than its cutoff value, status was defined as 0. Then, the risk score system was constructed by linear combination of RNA expression status weighted by regression coefficients, and the corresponding formula for calculating risk score (RS) was:

$$\text{Status Risk Score} = \sum \beta_{RNAn} \times \text{Status}_{RNAn}$$

β and status separately represented the regression coefficient and the status variable (both the mRNA expression status-based risk score system and the lncRNA expression status-based risk score system were built here).

Besides, the expression level-based risk score systems were constructed based on the following calculation formula:

$$\text{Expression Risk score} = \sum \beta_{RNAn} \times \text{Exp}_{RNAn}$$

β and "Exp_{RNAs}" stood for the regression coefficient and the expression level of RNA. (both the mRNA expression level-based risk score system and the lncRNA expression level-based risk score system were constructed).

The RSs of the OS samples in the training set were calculated, and their median was used as the demarcation point for dividing the samples into high risk (with RS no less than the demarcation point) and low risk (with RS lower than the demarcation point) groups. Using the Kaplan-Meier (KM) method in survival package [19], the correlation between risk grouping and actual survival prognosis information was evaluated. Meanwhile, the samples in the validation set were classified into high and low risk groups according to the median of their RSs. Afterwards, C-index and Brier score calculated by survcomp package (version 1.30.0, http://www.bioconductor.org/packages/release/bioc/html/survcomp.html) [21] and log-rank p-value calculated by survival package [19] were used to assess the correlation between risk grouping and actual survival prognosis information in the validation set.

Stratification analysis

Using the univariate Cox regression analysis and multivariate Cox regression analysis in the survival package [19], the independent clinical prognostic factors were screened from the training set. The log-rank p-value < 0.05 was selected as the threshold of significant correlation.

To further investigate the relationship between the independent clinical prognostic factors in different risk groups and prognosis, stratification analysis [22] for the independent clinical prognostic factors was carried out. Risk stratification was performed for the OS samples in the training set to study the prognosis conditions of target clinical factors in different risk groups.

Pathway enrichment analysis

The RSs of the OS samples in the training set were calculated based on the optimal risk score system, and then the OS samples were divided into high and low risk groups. With the $|\log_2$ FC$| > 0.263$ and FDR < 0.05 as the thresholds, the differentially expressed genes (DEGs) between the two groups were selected using limma package [17]. In addition, the DEGs were conducted with pathway enrichment analysis using Gene Set Enrichment Analysis (GSEA, http://software.broadinstitute.org/gsea/index.jsp) [23]. The FDR < 0.05 was set as the threshold for selecting significant pathways.

Results

Differential expression analysis

Combined with HGNC database, a total of 15672 mRNAs and 993 lncRNAs shared by the two datasets were annotated. The 169 OS samples in the training set were divided into recurrence group (28 samples) and non-recurrence group (141 samples). There were 333 DERs between recurrence and non-recurrence groups, including 319 DE-mRNAs (120 up-regulated and 199 down-regulated) and 14 DE-lncRNAs (two up-regulated and 12 down-regulated). As shown in the hierarchical clustering heatmap, the expression values of the DERs in the samples could clearly classify the samples into two clusters (Fig. 1).

Construction of risk score systems

Based on the univariate Cox regression analysis, 97 DE-mRNAs and 10 DE-lncRNAs significantly correlated with overall survival were selected. Furthermore, 10 DE-mRNAs (including aldehyde dehydrogenase 1 family member A1, *ALDH1A1*; carbonic anhydrase 9, *CA9*; GDP-mannose 4,6-dehydratase, *GMDS*; leucine carboxyl methyltransferase 2, *LCMT2*; leucine rich repeat containing 75A, *LRRC75A*; methyltransferase like 1, *METTL1*; RAB29, member RAS oncogene family, *RAB29*; transcriptional adaptor 2B, *TADA2B*; tudor domain containing 7, *TDRD7*; and tigger transposable element derived 2, *TIGD2*) and eight DE-lncRNAs (including COL18A1 antisense RNA 1, *COL18A1-AS1*; long intergenic non-protein coding RNA 92, *LINC00092*; long intergenic non-protein coding RNA 298, *LINC00298*; long intergenic non-protein coding RNA 636, *LINC00636*; long intergenic non-protein coding RNA 885, *LINC00885*; long intergenic non-protein coding RNA 957, *LINC00957*; long intergenic non-protein coding RNA 1124, *LINC01124*; and ZEB1 antisense RNA 1, *ZEB1-AS1*) correlated with independent prognosis were screened using the multivariate Cox regression analysis (Table 1).

Table 1
The mRNAs and long non-coding RNAs (lncRNAs) correlated with independent prognosis.

Type	RNA	Coef	P-value	Hazard Ratio	95% CI	Cutoff
lncRNA	*COL18A1-AS1*	-1.16269	2.02E-02	0.313	0.052–0.866	0.01
	LINC00092	-0.02708	3.85E-02	0.973	0.916–0.995	0.3
	LINC00298	0.08726	4.77E-02	1.091	1.008–1.388	0.04
	LINC00636	-0.27707	1.15E-02	0.758	0.537–0.970	0.12
	LINC00885	-0.27569	4.04E-02	0.759	0.397–0.950	0.06
	LINC00957	-0.01955	3.63E-02	0.981	0.940–1.023	0.51
	LINC01124	-0.05308	1.32E-02	0.948	0.885–1.016	0.02
	ZEB1-AS1	0.03671	1.05E-02	1.037	0.992–1.084	0.21
mRNA	*ALDH1A1*	-0.043957	5.90E-03	0.957	0.928–0.987	1.7
	CA9	0.030203	3.85E-03	1.031	1.010–1.052	1.07
	GMDS	0.076649	1.39E-02	1.080	1.016–1.148	0.96
	LCMT2	-0.185329	3.72E-3	0.831	0.733–0.942	1.83
	LRRC75A	-0.068367	3.13E-02	0.934	0.878–0.994	0.92
	METTL1	0.072283	4.35E-02	1.075	1.002–1.153	0.52
	RAB29	-0.095801	1.26E-02	0.909	0.843–0.980	0.55
	TADA2B	-0.105184	2.66E-02	0.900	0.820–0.988	0.67
	TDRD7	0.075387	3.58E-02	1.078	1.005–1.157	1.93
	TIGD2	0.114616	3.53E-02	1.121	1.008–1.248	1.9

Note: CI, confidence interval.

The cutoff value of the expression level of each independent prognosis-associated DE-lncRNA and DE-mRNA was obtained, based on which the status of each sample in the expression of the RNA was defined. Then, the expression status-based risk score systems were constructed combined with the prognostic coefficients and expression status of the independent prognosis-associated DE-lncRNAs and DE-mRNAs, and the corresponding formula was:

mRNA status RS= $(-0.043957)*Status_{ALDH1A1} + (0.030203)*Status_{CA9} + (0.076649)*Status_{GMDS} + (-0.185329) *Status_{LCMT2} + (-0.068367)*Status_{LRRC75A} + (0.072283)*Status_{ETTL1} + (-0.095801)*Status_{RAB29} + (-0.105184) *Status_{TADA2B} + (0.075387)*Status_{TDRD7} + (0.114616)*Status_{TIGD2}$

lncRNA status RS $=(-1.16269)*Status_{COL18A1-AS1} + (-0.02708)*Status_{LINC00092} + (0.08726)*Status_{LINC00298} + (-0.27707)*Status_{LINC00636} + (-0.27569)*Status_{LINC00885} + (-0.01955)*Status_{LINC00957} + (-0.05308)*Status_{LINC01124} + (0.03671)*Status_{ZEB1-AS1}$

For the independent prognosis-associated DE-lncRNAs and DE-mRNAs, the expression level-based risk score systems were constructed and the formula was as follows:

mRNA exprs RS= (-0.043957)*$Exprs_{ALDH1A1}$ + (0.030203)* $Exprs_{CA9}$ + (0.076649)* $Exprs_{GMDS}$ + (-0.185329) *$Exprs_{LCMT2}$+ (-0.068367)*$Exprs_{LRRC75A}$+ (0.072283)*$Exprs_{ETTL1}$+ (-0.095801)*$Exprs_{RAB29}$+ (-0.105184) *$Exprs_{TADA2B}$ + (0.075387)* $Exprs_{TDRD7}$+ (0.114616)* $Exprs_{TIGD2}$

lncRNA exprs RS =(-1.16269)*$Exprs_{COL18A1-AS1}$+ (-0.02708)*$Exprs_{LINC00092}$ + (0.08726)* $Exprs_{LINC00298}$+ (-0.27707)*$Exprs_{LINC00636}$+ (-0.27569)*$Exprs_{LINC00885}$+ (-0.01955)*$Exprs_{LINC00957}$+ (-0.05308)*$Exprs_{LINC01124}$+ (0.03671)* $Exprs_{ZEB1-AS1}$

According to the median of the RSs, the OS samples in the training set and the validation set separately were divided into two groups (high risk group and low risk group). Based on KM method, correlation analysis for the risk grouping and actual survival prognosis information was performed. For both the training set and the validation set, the risk groups divided by the lncRNA expression status-based risk score system (training set: log-rank p-value = 2.196e-03, C-index = 0.703, Brier score = 0.0652; validation set: log-rank p-value = 3.537e-03, C-index = 0.779, Brier score = 0.0897) and mRNA expression status-based risk score system (training set: log-rank p-value = 6.264e-11, C-index = 0.878, Brier score = 0.0402; validation set: log-rank p-value = 8.905e-03, C-index = 0.822, Brier score = 0.0647) had significant correlations with the actual survival prognosis information (Fig. 2). Meanwhile, the risk groups divided by the lncRNA expression level-based risk score system (training set: log-rank p-value = 5.071e-04, C-index = 0.730, Brier score = 0.0603; validation set: log-rank p-value = 6.793e-03, C-index = 0.887, Brier score = 0.101) and mRNA expression level-based risk score system (training set: log-rank p-value = 1.927e-08, C-index = 0.869, Brier score = 0.0397; validation set: log-rank p-value = 1.132e-01, C-index = 0.733, Brier score = 0.147) were significantly correlated with the actual prognosis information (Fig. 3). Through comparing the log-rank p-values, C-indexes, and Brier scores, the mRNA expression status-based risk score system was selected as the optimal risk score system.

Stratification analysis

Combined with Cox regression analysis, three independent clinical prognostic factors (age, recurrence, and mRNA status based RS model) in the training set were identified (Table 2). KM curves showed that OS patients with lower age and non-recurrence had better prognosis (Fig. 4), which was consistent with the actual situation. Stratification analysis indicated that age and recurrence were significantly correlated with overall survival prognosis in the high risk group (Fig. 5).

Table 2
The screening of independent clinical prognostic factors.

Clinical characteristics	TCGA(N = 169)	Uni-variables cox			Multi-variables cox		
		HR	95%CI	P	HR	95%CI	P
Age(years,mean ± sd)	61.40 ± 15.24	1.019	1.001–1.037	4.24E-02	1.024	1.005–1.044	1.44E-02
Gender(Male/Female)	69/100	1.155	0.693–1.925	5.80E-01	-	-	-
Tumor multifocal(Yes/No/-)	33/127/9	1.614	0.896–2.910	1.08E-01	-	-	-
Tumor recurrence(Yes/No)	28/141	2.692	1.581–4.585	1.48E-04	2.454	1.391–4.329	1.93E-03
Tumor metastatic(Yes/No)	56/113	2.754	1.657–4.578	4.61E-05	1.709	0.958–3.050	6.96E-02
Radiotherapy(Yes/No)	61/108	0.817	0.483–1.381	4.50E-01	-	-	-
Tumor necrosis (No/Slight/Moderate/Severe/-)	59/34/59/9/8	1.193	0.927–1.535	1.69E-01	-	-	-
mRNA status based model(High/ Low)	84/85	6.454	3.437–12.12	6.26E-11	5.844	3.061–11.16	8.83E-08
Dead(Death/Alive/-)	61/108	-	-	-	-	-	-
Overall survival time(months,mean ± sd)	40.32 ± 32.59	-	-	-	-	-	-
Note: TCGA, The Cancer Genome Atlas; HR, Hazard Ratio; CI, confidence interval.							

Pathway enrichment analysis

Combined with the mRNA expression status-based risk score system, the OS samples in the training set were divided into high and low risk groups according to their RSs. There were 722 DEGs (144 up-regulated genes and 578 down-regulated genes) in high risk group in comparison to low risk group (Fig. 6). Moreover, seven pathways for the DEGs were enriched, including vascular smooth muscle contraction (FDR = 0.0000), histidine metabolism (FDR = 0.0217), and glycine, serine and threonine metabolism (FDR = 0.0436) (Table 3).

Table 3
The pathways significantly enriched for the differentially expressed genes.

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	Gene
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	9	-0.7802	-2.4985	0	0	*PPP1R14A, ITPR1,ACTG2,MYLK,ACTA2,MRVI1,MYH11,PF*
KEGG_HISTIDINE_METABOLISM	4	-0.7896	-1.8085	3.20E-03	2.17E-02	*MAOB,MAOA,ALDH3A2,ALDH1B1*
KEGG_TYROSINE_METABOLISM	3	-0.8632	-1.8135	1.60E-03	2.42E-02	*AOC3,MAOB,MAOA*
KEGG_PHENYLALANINE_METABOLISM	3	-0.9527	-1.8162	0	2.86E-02	*AOC3,MAOB,MAOA*
KEGG_TRYPTOPHAN_METABOLISM	4	-0.8705	-1.8214	0	3.40E-02	*MAOB,MAOA,ALDH3A2,ALDH1B1*
KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	3	-0.8185	-1.8611	1.50E-03	4.36E-02	*AOC3,MAOB,MAOA*
KEGG_ARGININE_AND_PROLINE_METABOLISM	4	-0.7077	-1.7245	1.08E-02	4.59E-02	*MAOB,MAOA,ALDH3A2,ALDH1B1*

Note: ES, enrichment score; NES: normalized enrichment score; NOM p-value: nominal p value, FDR, false discovery rate.

Discussion

There were 319 DE-mRNAs (120 up-regulated and 199 down-regulated) and 14 DE-lncRNAs (two up-regulated and 12 down-regulated) between recurrence and non-recurrence groups. Through performing Cox regression analysis, 10 DE-mRNAs (including *ALDH1A1*, *CA9*, *GMDS*, *LCMT2*, *LRRC75A*, *METTL1*, *RAB29*, *TADA2B*, *TDRD7*, and *TIGD2*) and eight DE-lncRNAs correlated with independent prognosis were identified. After four risk score systems were built, the mRNA expression status-based risk score system was selected as the optimal one. Three independent clinical prognostic factors (age, recurrence, and mRNA status based RS model) were selected, among which age and recurrence were significantly correlated with overall survival prognosis in the high risk group. Previous study has revealed that *CA9* is overexpressed in OS compared with bone marrow stromal cells (BMSCs) and CA9 inhibitor 3 can reduce tumor growth via causing significant necrosis [24]. *CA9* expression is detected in OS patients, and its inhibitors is able to inhibit cell proliferation, cell migration, and chemoresistance in OS under hypoxic conditions [25]. Therefore, *CA9* might act in the prognosis of OS patients.

Although *ALDH1A1*, *GMDS*, *LCMT2*, *LRRC75A*, *METTL1*, *TADA2B*, and *TDRD7* have not been reported to be directly correlated with OS, they have found to be able to affect other tumors. Through upregulating the cancer stem cell marker *ALDH1A1*, CCAAT-Enhancer-Binding Protein β1 (*C/EBPβ1*) accelerates transformation and results in chemoresistance in Ewing sarcoma [26]. *GMDS* deficiency promotes the resistance of colon cancer cells to the apoptosis caused by tumor necrosis factor-related apoptosis-inducing ligand (*TRAIL*), leading to tumor progression and metastasis through evading tumor immune surveillance [27, 28]. Some cancer-related genes (including *LCMT2*) possess frameshift mutations and mutational intratumoral heterogeneity in colorectal cancer (CRC) with high microsatellite instability, and their frameshift mutations may be involved in the tumorigenesis of the disease [29]. Decreased LRRC75A antisense RNA 1 (*LRRC75A-AS1*) plays an anti-tumor role in the tumorigenesis and progression of CRC, which may serve as a diagnostic marker in CRC [30]. Upregulated *METTL1* is related to unfavorable outcomes in hepatocellular carcinoma, and *METTL1* contributes to cell proliferation and migration in the tumor by inhibiting *PTEN* signaling [31]. Several highest-ranking genes (including *TADA2B*) are selected by RNAi Gene Enrichment Ranking algorithm, and their loss contributes to vemurafenib resistance in melanoma [32]. *TDRD7*, heat shock protein 90B (*HSP90B*), *RAB1A*, and vimentin are found to be pseudopodia-localizing proteins, which play roles in cell migration and affect tumor malignancy [33]. These suggested that *ALDH1A1*, *GMDS*, *LCMT2*, *LRRC75A*, *METTL1*, *TADA2B*, and *TDRD7* might also be implicated in the prognosis of OS patients.

Additionally, 722 DEGs (144 up-regulated genes and 578 down-regulated genes) between high risk and low risk groups were screened, for which seven pathways (including vascular smooth muscle contraction, and glycine, serine and threonine metabolism) were enriched. Structural damage and contractile dysfunction of vascular smooth muscle and anti-cancer effects may be induced by the ginsenoside Rg3 via destroying the actin [34]. Smooth muscle contraction, differentiation, and migration, as well as cancers and cardiovascular diseases are affected by the actin-associated protein palladin [35]. The mTOR complex 1 (mTORC1)/serine/glycine metabolic axis contributes to antioxidant ability, cell proliferation, and cell survival in OS, indicating that mTORC1-mediated serine/glycine metabolism exerts a protective effect on OS cells [36, 37]. Thus, vascular smooth muscle contraction, and glycine, serine and threonine metabolism might be related to the prognosis of OS patients.

After differential expression analysis for recurrence and non-recurrence groups was conducted, the DE-mRNAs and DE-lncRNAs correlated with independent prognosis were selected. Moreover, risk score systems were constructed based on the prognosis-associated RNAs, followed by the independent clinical prognostic factors correlated with prognosis were revealed by stratification analysis. Nevertheless, these findings were obtained on the basis of bioinformatics analyses and needed to be confirmed by adequate experiments.

Conclusion

In conclusion, 319 DE-mRNAs and 14 DE-lncRNAs were identified from recurrence and non-recurrence groups. Besides, the mRNA expression status-based risk score system (involving *ALDH1A1*, *CA9*, *GMDS*, *LCMT2*, *LRRC75A*, *METTL1*, *RAB29*, *TADA2B*, *TDRD7*, and *TIGD2*) might be applied for predicting the prognosis

of OS patients. Furthermore, vascular smooth muscle contraction, and glycine, serine and threonine metabolism might have correlations with the prognosis of OS.

Abbreviations

osteosarcoma (OS)

long non-coding RNAs (lncRNAs)

migration-inducing gene 7 (MIG-7)

C C motif chemokine ligand 5 (CCL5)

nuclear factor-κB (NF-κB)

Gene Expression Omnibus (GEO)

reference sequence (RefSeq)

differentially expressed mRNAs (DE-mRNAs))

fold change (FC)|

false discovery rate (FDR)

Kaplan-Meier (KM)

CCAAT-Enhancer-Binding Protein β1 (C/EBPβ1)

colorectal cancer (CRC)

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:
Not applicable

Availability of data and material

The raw data were collected and analyzed by the Authors, and are not ready to share their data because the data have not been published.

Competing interests:
The authors declare no conflict of interest.

Funding:

none

Authors' contributions

JGC and JW participated in the design of this study, and they both performed the statistical analysis. AW carried out the study and collected important background information. JMY drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

none

References

1. Li S, Sun W, Wang H, Zuo D, Hua Y, Cai Z. Research progress on the multidrug resistance mechanisms of osteosarcoma chemotherapy and reversal. Tumor Biology. 2015;36(3):1329–38.
2. Duffaud F, Digue L, Baciuchka-Palmaro M, Volot F, Perles-Daniel C, Garbe L, Favre R. Osteosarcomas of flat bones in adolescents and adults. Cancer. 2015;88(2):324–32.
3. Kansara M, Teng MW, Smyth MJ, Thomas DM. Translational biology of osteosarcoma. Nat Rev Cancer. 2014;14(11):722–35.
4. Biazzo A, De Paolis M. Multidisciplinary approach to osteosarcoma. Acta Orthop Belg. 2016;82(4):690–8.
5. Taran SJ, Taran R, Malipatil NB. Pediatric Osteosarcoma: An Updated Review. Indian Journal of Medical Paediatric Oncology Official Journal of Indian Society of Medical Paediatric Oncology. 2017;38(1):33–43.
6. Wan Y, Xu L, Zhuo N, Ge J, Chen G, Lu XB. The clinical significance of neoadjuvant chemotherapy in improving the drug resistance of osteosarcoma. Minerva Med. 2017;108(5):479.
7. Ren K, Zhang J, Gu X, Wu S, Shi X, Ni Y, Chen Y, Lu J, Gao Z, Wang C. Migration-inducing gene-7 independently predicts poor prognosis of human osteosarcoma and is associated with vasculogenic mimicry. Exp Cell Res. 2018;369(1):80–9.
8. Sun K, Gong C, Peng H, Fang H, Zhou J, Li J, Chen S, Zheng H. High CCL5 expression is associated with osteosarcoma metastasis and poor prognosis of patients with osteosarcoma. Mol Med Rep. 2017;16(5):6953–7.
9. Wang SW, Liu SC, Sun HL, Huang TY, Chan CH, Yang CY, Yeh HI, Huang YL, Chou WY, Lin YM. CCL5/CCR5 axis induces vascular endothelial growth factor-mediated tumor angiogenesis in human osteosarcoma microenvironment. Carcinogenesis. 2015;36(1):104–14.
10. Teng G, Su X, Xia Q, Wang J, Kan S. Expression of NF-κB and PTEN in osteosarcoma and its clinical significance. Oncology Letters. 2017;14(6):6744–8.
11. Xiao-Hui S, Li-Bin Y, Xiao-Lin G, Ran W, Zhi-Chang Z. Increased expression of lncRNA HULC indicates a poor prognosis and promotes cell metastasis in osteosarcoma. International Journal of Clinical Experimental Pathology. 2015;8(3):2994–3000.
12. Zhang Y, Meng W, Cui H. LncRNA CBR3-AS1 predicts unfavorable prognosis and promotes tumorigenesis in osteosarcoma. Biomedicine pharmacotherapy = Biomedecine pharmacotherapie. 2018;102:169–74.
13. Heery R, Finn SP, Cuffe S, Gray SG. Long Non-Coding RNAs: Key Regulators of Epithelial-Mesenchymal Transition, Tumour Drug Resistance and Cancer Stem Cells. Cancers. 2017;9(4):38.
14. Gupta SC, Tripathi YN. Potential of long non-coding RNAs in cancer patients: From biomarkers to therapeutic targets. Int J Cancer. 2017;140(9):1955–67.
15. Chaudhary K, Poirion OB, Lu L, Garmire LX. (2017) Deep Learning based multi-omics integration robustly predicts survival in liver cancer. Clinical Cancer Research 24 (6):clincanres.0853.2017.
16. Povey S,., Lovering R, Bruford, Wright E, Lush M, Wain M H. The HUGO Gene Nomenclature Committee (HGNC). Hum Genet. 2001;109(6):678.
17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
18. Kolde R. (2015) pheatmap: Pretty Heatmaps.R package version 102 ed 2015. Available: https://cran.r-project.org/web/packages/pheatmap/index.html.
19. Austin PC. A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications. Int Stat Rev. 2017;85(2):185.
20. Camp RL, Dolledfilhart M, Rimm DL. X-Tile A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization. Clin Cancer Res. 2004;10(21):7252–9.
21. Schröder MS, Culhane AC, Quackenbush J, Haibekains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27(22):3206–8.
22. Burkhardt JK, Lasker GF, Winkler EA, Kim H, Lawton MT. Microsurgical resection of brain arteriovenous malformations in the elderly: outcomes analysis and risk stratification. J Neurosurg. 2017;129(5):1.
23. Li L, Wang X, Xiao G, Gazdar A. Integrative gene set enrichment analysis utilizing isoform-specific expression. Genet Epidemiol. 2017;41(6):498.
24. Perut F, Carta F, Bonuccelli G, Grisendi G, Pompo GD, Avnet S, Sbrana FV, Hosogi S, Dominici M, Kusuzaki K. Carbonic anhydrase IX inhibition is an effective strategy for osteosarcoma treatment. Expert Opin Ther Targets. 2015;19(12):1593–605.
25. Okuno K, Matsubara T, Nakamura T, Iino T, Kakimoto T, Asanuma K, Matsumine A, Sudo A. Carbonic anhydrase IX enhances tumor cell proliferation and tumor progression in osteosarcoma. Onco Targets Ther. 2018;11:6879–86.
26. Gardiner JD, Abegglen LM, Huang X, Carter BE, Schackmann EA, Stucki M, Paxton CN, Randall RL, Amatruda JF, Putnam AR. C/EBPβ-1 promotes transformation and chemoresistance in Ewing sarcoma cells. Oncotarget. 2017;8(16):26013–26.
27. Kenta M, Shinichiro S, Eiji M. GDP-mannose-4,6-dehydratase (GMDS) deficiency renders colon cancer cells resistant to tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) receptor- and CD95-mediated apoptosis by inhibiting complex II formation. J Biol Chem. 2011;286(50):43123.
28. Wei X, Zhang K, Qin H, Zhu J, Qin Q, Yu Y, Wang H. GMDS knockdown impairs cell proliferation and survival in human lung adenocarcinoma. Bmc Cancer. 2018;18(1):600.
29. Su YY, Yun SJ, Choi EJ, Min SK, Yoo NJ, Lee SH. Frameshift Mutations in Repeat Sequences of ANK3, HACD4, TCP10L, TP53BP1, MFN1, LCMT2, RNMT, TRMT6, METTL8 and METTL16 Genes in Colon Cancers. Pathology Oncology Research. 2017;24(3):1–6.
30. Chen J, Lan J, Ye Z, Duan S, Hu Y, Zou Y, Zhou J. (2019) Long noncoding RNA LRRC75A-AS1 inhibits cell proliferation and migration in colorectal carcinoma. Exp Biol Med (Maywood):1535370219874339. doi:10.1177/1535370219874339.
31. Tian QH, Zhang MF, Zeng JS, Luo RG, Wen Y, Chen J, Gan LG, Xiong JP. (2019) METTL1 overexpression is correlated with poor prognosis and promotes hepatocellular carcinoma via PTEN. J Mol Med (Berl):1–11. doi:10.1007/s00109-019-01830-9.

32. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science. 2014;343(6166):84.
33. Ito A, Mimae T, Yamamoto YS, Hagiyama M, Nakanishi J, Ito M, Hosokawa Y, Okada M, Murakami Y, Kondo T. Novel application for pseudopodia proteomics using excimer laser ablation and two-dimensional difference gel electrophoresis. Lab Invest. 2012;92(9):1374.
34. Jung Y, Kim K, Bian Y, Ngo T, Bae ON, Lim KM, Chung JH. Ginsenoside Rg3 disrupts actin-cytoskeletal integrity leading to contractile dysfunction and apoptotic cell death in vascular smooth muscle. Food Chem Toxicol. 2018;118:645–52.
35. Jin L. The actin associated protein palladin in smooth muscle and in the development of diseases of the cardiovasculature and in cancer. J Muscle Res Cell Motil. 2011;32(1):7–17.
36. Wang DW, Wu L, Cao Y, Yang L, Liu W, Ji EXQ, Bi G ZG. A novel mechanism of mTORC1-mediated serine/glycine metabolism in osteosarcoma development. Cell Signal. 2017;29:107–14.
37. Gorska-Ponikowska M, Perricone U, Kuban-Jankowska A, Lo BG, Barone G. 2-Methoxyestradiol Impacts on Amino Acids-mediated Metabolic Reprogramming in Osteosarcoma Cells by Interaction with NMDA Receptor. J Cell Physiol. 2017;232(11):3030–49.

Figures

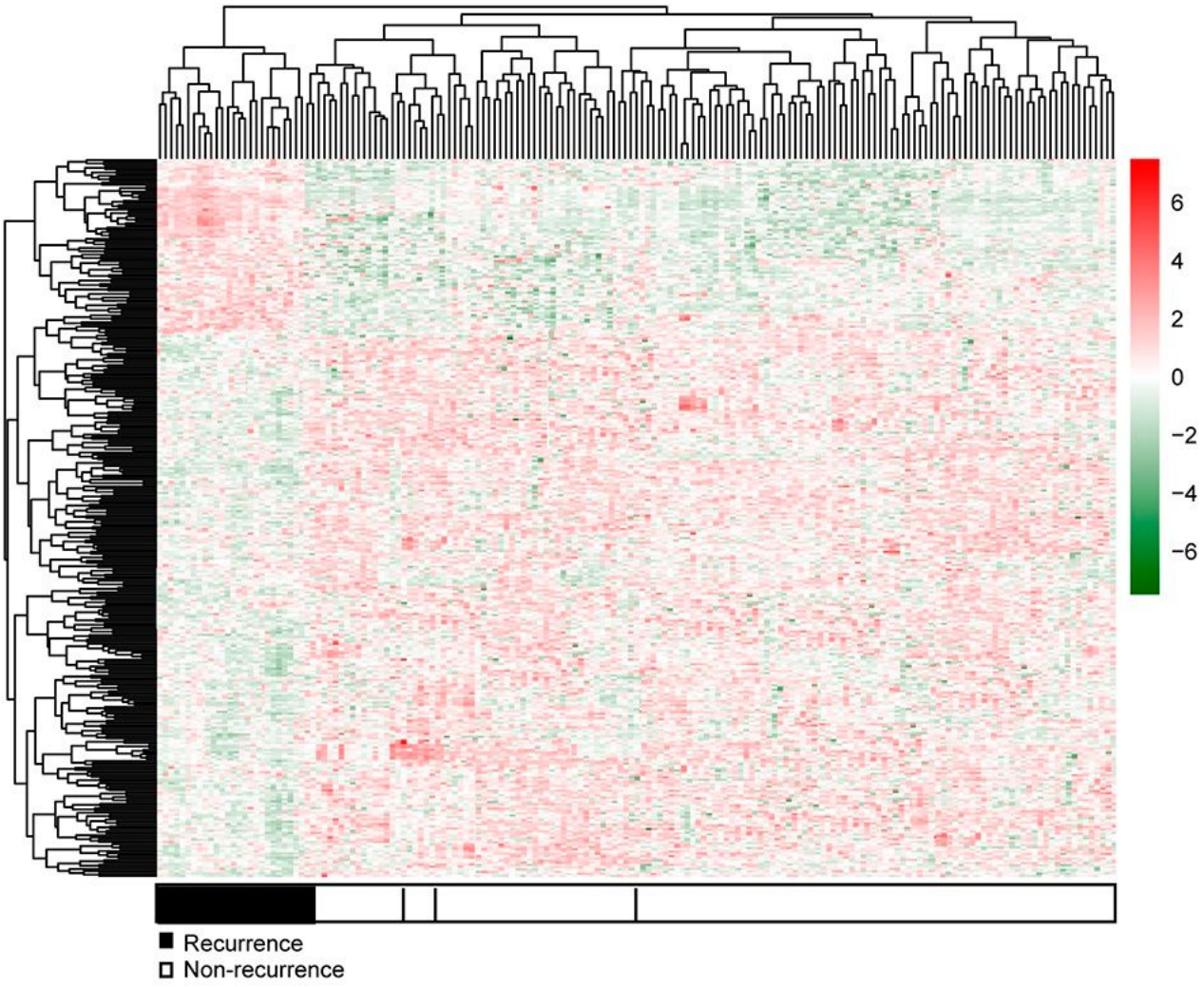

Figure 1

The hierarchical clustering heatmap based on the expression values of the differentially expressed RNAs (DERs). Black and white sample strips represent the samples in recurrence and non-recurrence groups, respectively.

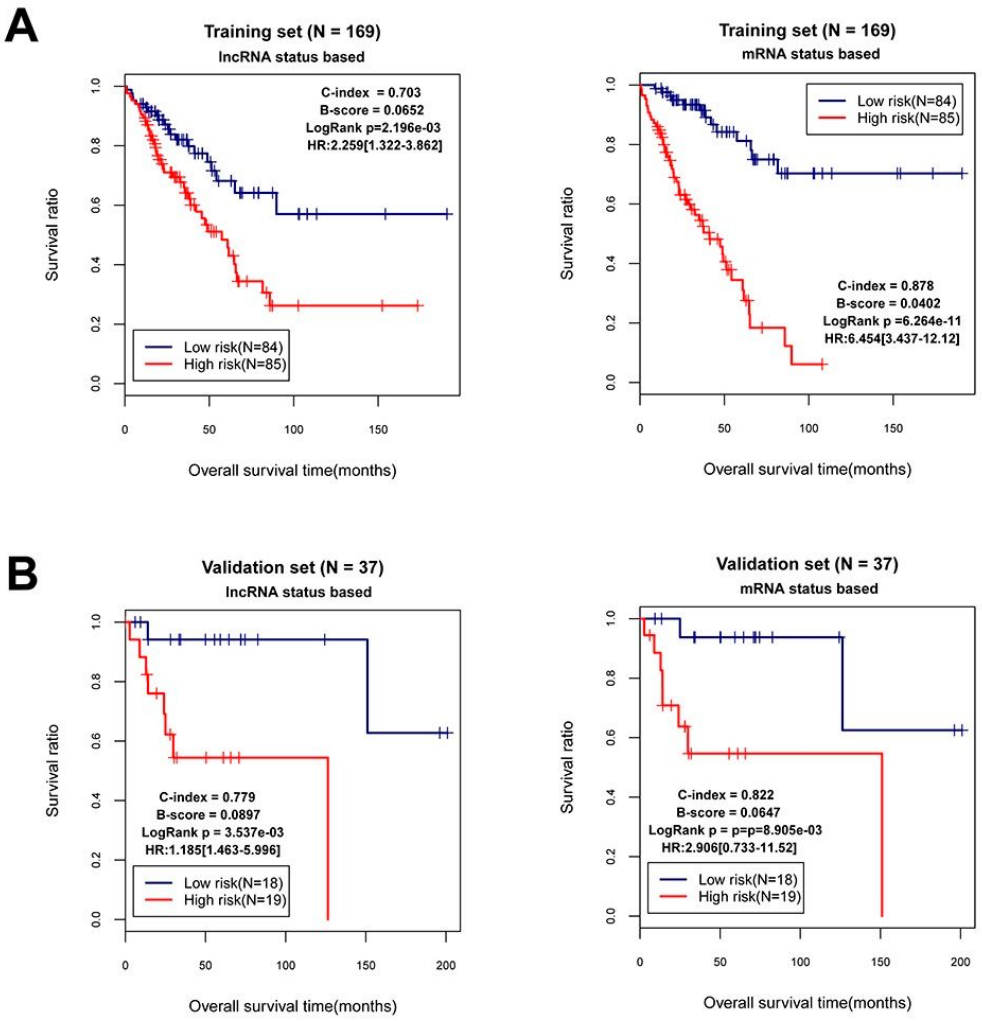

Figure 2

Correlation analysis for the risk groups divided by the expression status-based risk score system and actual survival prognosis information. (A) The Kaplan-Meier (KM) curves for the training set (left: lncRNA expression status-based risk score system; right: mRNA expression status-based risk score system); (B) The KM curves for the validation set (left: lncRNA expression status-based risk score system; right: mRNA expression status-based risk score system). In KM curves, blue and red separately represent low risk and high risk groups.

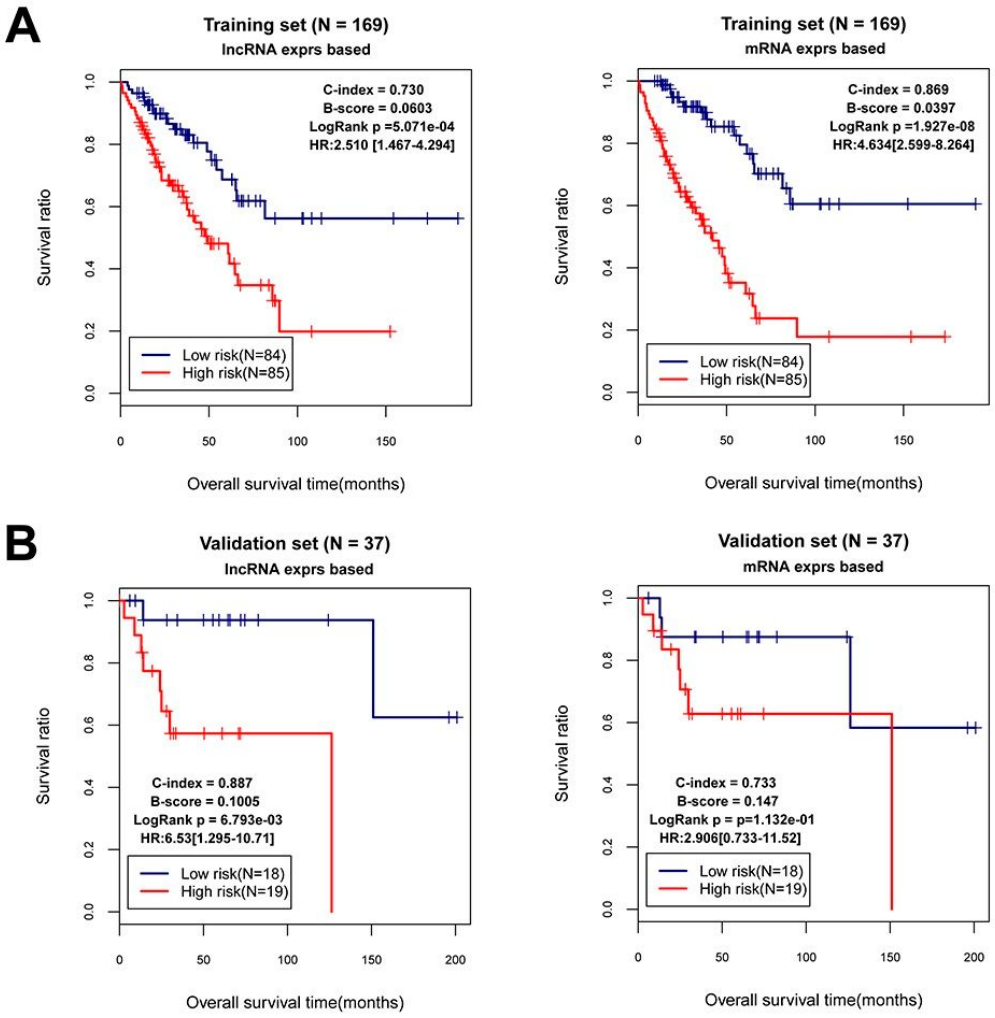

Figure 3

Correlation analysis for the risk groups divided by the expression level-based risk score system and actual survival prognosis. (A) The Kaplan-Meier (KM) curves for the training set (left: lncRNA expression level-based risk score system; right: mRNA expression level-based risk score system); (B) The KM curves for the validation set (left: lncRNA expression level-based risk score system; right: mRNA expression level-based risk score system). Blue and red curves represent low risk and high risk groups, respectively.

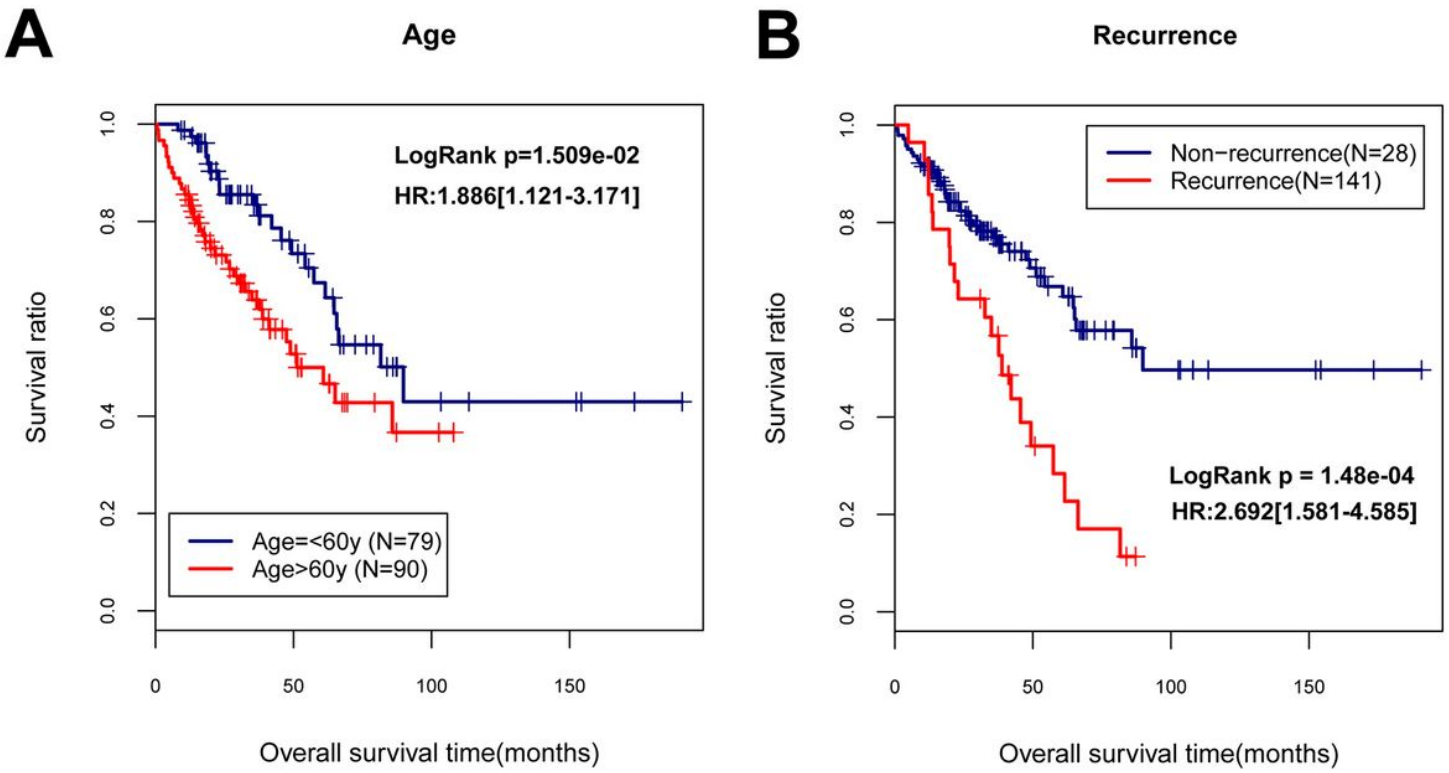

Figure 4

Correlation analysis for age/recurrence and survival prognosis. (A). The Kaplan-Meier (KM) curves for age (blue and red curves separately represent the samples no older than 60 years and the samples aged over 60 years); (B) The KM curves for recurrence (blue and red curves represent the samples in non-recurrence and recurrence groups, respectively).

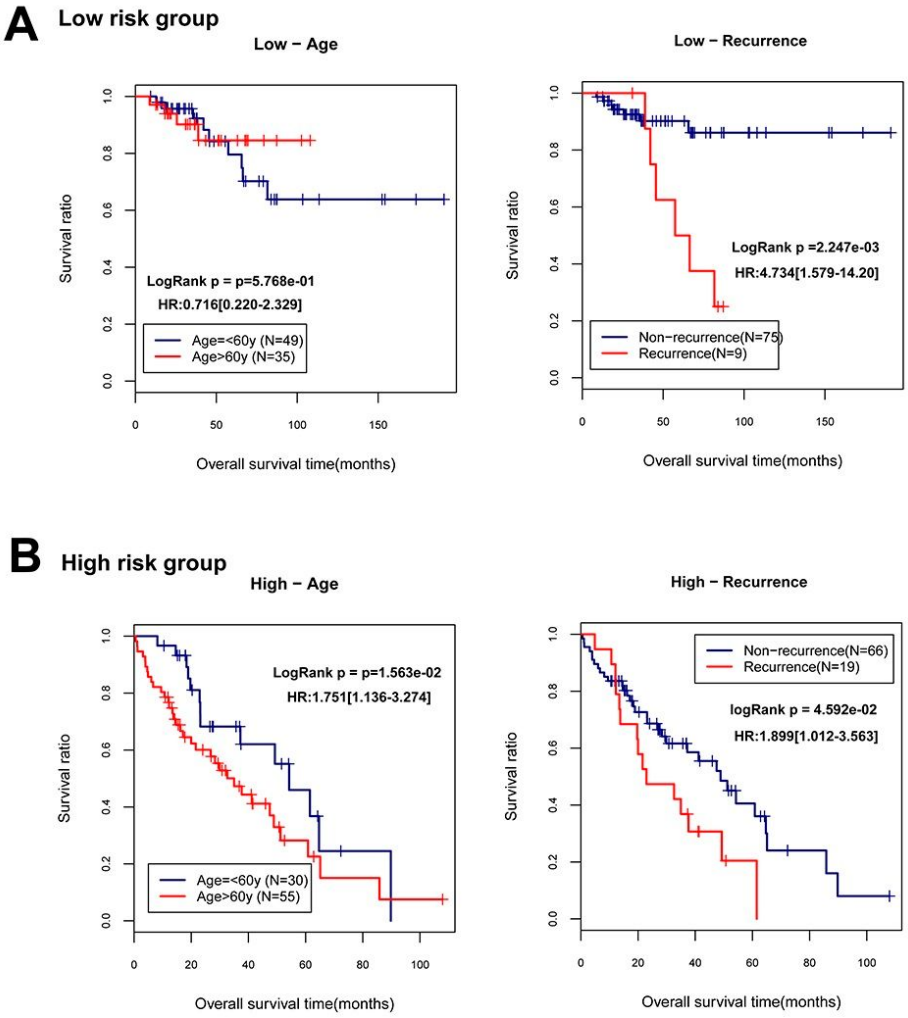

Figure 5

Correlation analysis for age/recurrence and survival prognosis in low risk group and high risk group. (A) The Kaplan-Meier (KM) curves for low risk group (left: age; right: recurrence); (B) The KM curves for high risk group (left: age; right: recurrence). In the KM curves for age, blue and red curves separately represent the samples no older than 60 years and the samples aged over 60 years. In the KM curves for recurrence, blue and red curves represent the samples in non-recurrence and recurrence groups, respectively.

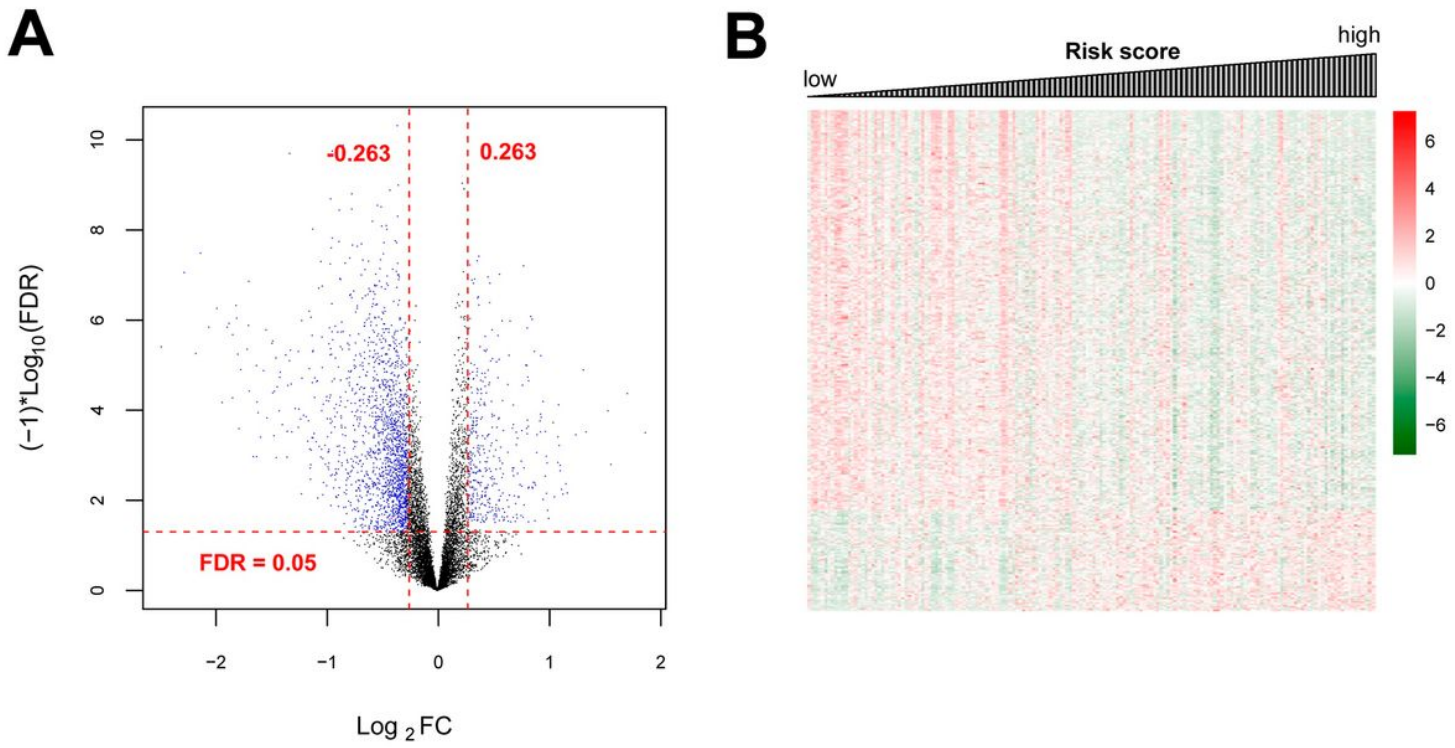

Figure 6

The volcano plot and expression heatmap of the differentially expressed genes (DEGs). (A) The volcano plot (blue and black dots separately represent DEGs and non-DEGs); (B) The heatmap based on the expression changes of the DEGs with risk scores from low to high.