Genome-wide analysis of lncRNAs, miRNAs and mRNAs forming a prognostic scoring model associated with the recurrence of osteosarcoma

Keywords
recurrence, TCGA, Osteosarcoma, prognostic scoring model

Abstract

Introduction
Purpose: Among young adults and adolescents, the most common malignant bone tumor is osteosarcoma (OS). Even patients who are cured by surgery or neoadjuvant chemotherapy still have a high possibility of recurrence. In recent years, due to the development of molecular biology research methods, many new prognostic markers based on the gene level have emerged. In addition, the mutual regulation mode among long non-coding RNA (lncRNA), miRNA and target genes are closely related to the occurrence and development of tumors. In our research, we aimed to analyze the molecular regulation mode and predict clinical outcomes by integrate three types of RNA expression.

Material and methods
Materials and Methods: We obtained the data of OS patients from The Cancer Genome Atlas (TCGA) database including expression data (RNA and miRNA expression data) and clinical data.

Results
Results: After differential gene expression analysis, Cox regression analysis and functional enrichment analysis, 1 lncRNA, 3 miRNAs and 9 mRNAs were identified as prognostic RNA. We constructed the prognostic scoring (PS) model with high predicting prognosis performance. Using PS models and clinical data, we established a nomogram to calculate patients' 3-year and 5-year survival rates.

Conclusions
Conclusions: Finally, competing endogenous RNAs (ceRNAs) network and functional enrichment analysis help us to understand molecular mechanisms associated with the recurrence of osteosarcoma.

Explanation letter

Review 1:
The manuscript is interesting, but I have some comments:
1. the summary does not reflect the content of the manuscript, especially the results and conclusions do not contain specific information.
Response: thanks for your comment. We have rewritten the summary.
2. the discussion is very superficial. The results were briefly commented, such as "Through the univariate and multivariate analyzes, we identified tumor metastasis as an independent prognostic clinical factor." It seems that this factor was previously considered as a prognostic factor. The authors should elaborate on the markers they believe have prognostic potential, including studies in other neoplasms.
Response: Thanks for your comment. We have rewritten this part. We added the impact of two clinical factor (tumor metastasis and PS model) on OS recurrence.
3. The study group is not very numerous, therefore the conclusions should be formulated more carefully, especially since the authors do not refer to basic research on the mechanisms related to the markers that

Response: Thanks for your comment. We revised our conclusion.
4. there are errors in the manuscript related to punctuation, extra spaces and typos
Response: We checked our manuscript carefully and corrected these errors.
Review 2:
The authors responded to the comments of the reviewers and improved the manuscript. English editing also improved the manuscript.
Figure legends make the images more clear, however, Figure 2 and 3 seem to be switched.
Response: Thanks for your comment and sorry for the mistake. We have corrected it.
In current Figure 3A, it is not clear what the title “Metastatic” represents. Additionally, a couple of typos remain in the manuscript.

Response: Thanks for your comment. We have removed the “Metastatic” in Figure 3A.

Response one by one .docx

Genome-wide analysis of lncRNAs, miRNAs, and mRNAs forming a prognostic scoring model associated with the recurrence of osteosarcoma

Abstract

Purpose: Among young adults and adolescents, the most common type of malignant bone tumor is osteosarcoma (OS). Even patients cured by surgery or neoadjuvant chemotherapy still have a high possibility of recurrence. Due to the development of molecular biology research methods, many new prognostic markers based on gene level have emerged. In addition, the mutual regulation mode among long noncoding RNA (lncRNA), microRNA (miRNA), and target genes is closely related to the occurrence and development of tumors. Therefore, in our research, we analyzed the molecular regulation mode and predicted clinical outcomes by integrating three types of RNA expression.

Materials and Methods: We obtained the data of patients with OS from The Cancer Genome Atlas (TCGA) database, including RNA and miRNA expression and clinical data.

Results: After performing differential gene expression, Cox regression, and functional enrichment analyses, we identified 1 lncRNA （LINC00626）, 3 miRNAs （has-miR-429，hsa-miR-526b，hsa-miR-615）, and 9 mRNAs （such as BCAS4, CA9,CPA3）as prognostic RNAs. Baed on these genes, we constructed the prognostic scoring (PS) model with a high predicting prognosis performance. Using this model and the clinical data, we established a nomogram to calculate patients’ 3- and 5-year survival rates. In addition, we also construted a competing endogenous RNA (ceRNA) network. Functional enrichment analysis shows that mRNAs in the ceRNA network were significantly related to biological processes of positive regulation of the developmental process and regulation of neurogenesis.

Conclusions: The PS model had high predicting prognosis performance. It can help us predict which patients will develop a recurrence. The ceRNA network and functional enrichment analysis can support the

understanding of the molecular mechanisms associated with OS recurrence.

Keywords: osteosarcoma, prognostic scoring model, recurrence, The Cancer Genome Atlas

Introduction

Osteosarcoma (OS), the most common malignant tumor that is prone to recurrence and metastasis (1). Itoccurs mostly occurs in young adults and adolescents and usually originates from the long bones (2). In the past, many patients died within 1 year of diagnosis due to treatment limitations. Today neoadjuvant chemotherapy has greatly improved the survival rate, but the recurrence rate remains around 35% (3). However, there are too many clinical indicators, such as age, gender, and tumor site and stage, can influence OS prognosis and cause recurrence. It is difficult to confirm which patients will develop recurrence using only these indicators. Therefore, finding effective and novel recurrent biomarkers to evaluate the prognosis of OS recurrence is important to formulating treatment strategies and predicting efficacy.

In tumorigenesis and tumor progression, protein coding genes and noncoding RNAs play important roles. MicroRNAs (miRNAs) regulate mRNA expression at a post-transcriptional level with □20 nucleotides in length. Long noncoding RNAs (lncRNAs) （>200 nucleotides）can act as competing endogenous RNAs (ceRNAs), which usually regulate mRNA expression by competing for a common pool of miRNAs. This mechanism, called “ceRNA hypothesis”(4, 5). It occurs extensively in the basic cellular processes and functions and is closely related to disease development (5, 6).

In this study, we comprehensively analyzed all OS-related data in The Cancer Genome Atlas (TCGA) database. We divided OS tissue samples into high-risk and low-risk groups, determining significant difference in overall survival through a prognostic scoring (PS) model. Then we established a nomogram to calculate patients’ 3- and 5-year survival rates using this model and clinical data. In addition, we constructed a recurrence-related ceRNA network and explored the potential molecular mechanisms of recurrence mRNAs using functional enrichment analyses. Our aim is to discover some novel clues that can effectively predict which patients will develop recurrence.

Materials and methods

Data collection

Within the TCGA database (https://gdc-portal.nci.nih.gov), we identified 169 OS samples with mRNA (which includes lncRNAs) and miRNA expression data and also recurrence prognosis information. We detected these samples with the Illumina HiSeq 2000 RNA Sequencing platform. We used these data as the training data set for this analysis.

We downloaded the validation data set from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) using the key words "osteosarcoma, homo sapiens." A total of 37 samples in GSE39058 met the following criteria: (1) solid tissue from OS tumors; (2) lncRNA, mRNA, and miRNA expression data; and (3) clinical information on recurrence and prognosis. We detected these samples using the Illumina platform.

Analysis of differentially expressed RNA in OS tissues

Using the RefSeq ID information, we compared the profiles in the training data set with the genome annotation file, which we downloaded from the HUGO Gene Nomenclature Committee (HGNC(7); http://www.genenames.org/) To further examine the differentially expressed RNAs (DERs; including lncRNA, miRNA, and mRNA) between the recurrence and non-recurrence OS samples in the training data set, we analyzed the expression data using the Limma package of the R software (https://bioconductor.org/packages/release/bioc/html/limma.html) (8). We considered $|\log2 \text{ foldchange}| > 0.5$ and FDR < 0.05 as significant. We used heatmaps and volcano plots for visualization (9).

Screening for independent prognostic clinical factors

Univariate and multivariate Cox regression analysis was used to screen out the independent prognostic

clinical factors by R package of survival. We both selected $p_{log\text{-}rank} < 0.05$ as the threshold for significant correlation. To further investigate the relationship between independent prognostic factors and high- and low-risk groups, we conducted risk stratification analysis on these independent prognostic clinical factors.

PS model and performance evaluation

We used Cox regression analysis to assess the effects of candidate high-risk gene expression on overall survival and used univariate Cox regression analysis to test the relationship between DER expression levels and prognosis in the training data set. We used a multivariate Cox regression analysis to examine the results.

We selected $p_{log\text{-}rank} <0.05$ as the threshold for significant correlation and screened out the DERs related to independent prognosis. On the basis of DER expression, we used the R package of penalized (http://bioconductor.org/packages/penalized/) to screen out the optimal prognosis-related signature DERs using the Cox proportional-hazards model (10)

Using the signature DER expression and their prognosis coefficients in the training data set, we identified a combined signature to build a PS model. With this model, we calculated an expression-based risk score for every sample. Then the samples were classified into two groups (high-risk and low-risk) according to their median score.

Using Kaplan−Meier analysis we evaluated for the association between the samples and actual survival prognostic information (11). We also evaluated the association in the validation data set.

Nomogram construction

To further investigate the correlation among the independent prognostic clinical factors, the PS model, and survival prognosis, we established a nomogram for 3- and 5-year survival rates using the regression modeling strategies package in R(12, 13).

ceRNA network construction

Prediction of lncRNA−miRNA−mRNA interactions

We used the experimental module DIANA-LncBase Version 2 (14) (http://www.microrna.gr/LncBase) to predict the lncRNA−miRNA interactions. Then, we used T starBase Version 2.0 (15) (http://starbase.sysu.edu.cn/) to predict the interactions between the DEmiRNAs and DEmRNAs. We mapped the signature mRNAs to regulated target genes to predict the signature miRNA−mRNA interactions. Finally, we constructed an lncRNA-related ceRNA network using Cytoscape Version 3.6.1 (16). It shows which lncRNAs can affect miRNA function and how to regulate mRNA expression.

Gene function analysis

We performed the Gene Ontology (GO) biology process and the Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway enrichment analysis on the genes contained in the ceRNA network using the DAVID version 6.8(17, 18) (Database for Annotation, Visualization, and Integrated Discovery, http://david.abcc.ncifcrf.gov/). $P < 0.05$ was the threshold of enrichment significance.

Functional enrichment analysis for DERs between the high- and low-risk groups

First, we used the Limma package of R to screen DERs between the high- and low-risk groups in the training data set. We considered |log2 foldchange| >0.5 and FDR < 0.05 as significant. Then we used DAVID Version 6.8 for the functional analysis of the DERs.

Results

DE lncRNA, DEmiRNA, and DEmRNA

We found 169 OS tissue samples for combined mRNA, miRNA, and lncRNA from the TCGA database. We annotated 10,700 mRNAs, 1,029 lncRNAs, and 1,881 miRNAs. Utilizing the recurrence information, we divided the samples into recurrence and non-recurrence groups containing 28 and 141 samples, respectively.

From the analysis of the expression profile we identified 54 miRNAs, 178 mRNAs, and 47 lncRNAs differentially expressed between recurrent and non-recurrent OS. The volcano plot (Fig 1A) and heatmap (Fig 1B) for the lncRNAs, miRNAs, and mRNAs showed that the recurrence tissues clustered separately from the non-recurrence tissues.

Identifying the independent prognosis DERs and the PS model

In the TCGA dataset, 133 OS tumor samples contained recurrence prognosis information. From the univariate Cox proportional hazard regression analysis, we identified 61 significant RNAs ($P < 0.05$). We performed a multivariate Cox regression analysis for these using recurrence-free survival as an independent variable; thus, we identified 20 DERs, including 4 lncRNAs, 3 miRNAs, and 13 mRNAs, that were significantly associated with independent prognosis in OS ($P < 0.05$). Finally, from the Cox-PH model of the regularization regression algorithm, which screens the optimal combination of the signature DERs, we obtained a total of 13 DERs (Table 1).

Therefore, we proposed the following PS model for recurrence-free survival with the risk scoring method, which integrated the signature DER expression levels and their prognosis coefficients:

$$\text{Prognostic score} = \sum \beta_{DERs} \times \text{Exp}_{DERs}$$

where β_{DERs} is the prognosis coefficient of the signature DERs and Exp_{DERs} is the expression level of the signature DERs.

Screening the independent prognostic clinical factors

With the univariate and multivariate Cox regression analyses we selected the independent prognostic clinical factors. Two factors were significantly associated with independent prognosis in OS: tumor metastasis and PS model status (Table 2). As is shown in Fig 2A, samples with tumor metastasis had a worse prognosis than did samples without ($P < 0:0001$). This result is consistent with the actual situation. We then divided the samples into with− and without−tumor metastasis groups, analyzing the correlation between the prediction results of

the PS model and the actual prognosis for each group. Our results showed that the PS model remained significantly correlated with recurrence-free survival time after adjustment by tumor metastasis (Fig 2B).

Prognostic scoring model and performance evaluation

We calculated the risk scores based on the formula of the PS model and divided the samples into low-risk (n = 66) and high-risk (n = 67) groups according to median risk score. The prognostic value of these signature DERs was demonstrated in Kaplan−Meier plots (Fig 3A). For this PS model, we obtained a good area under the curve (AUC) value of 0.966 based on a 10-fold cross-validation in the training data set. Samples in the high-risk group had a worse prognosis than did those in the low-risk group (P < 0:0001). We tested the robustness of the combined prognostic signature DERs for predicting recurrence in patients with OS in the validation data set (GSE39058, n = 37) downloaded from the GEO, obtaining similar risk stratification results (Fig 3B). As with the training data set, the combined prognostic signature DERs classified 37 samples into low-risk (n = 18) and high-risk (n = 19) groups with significantly different recurrence-free survival times. The AUC of the prognostic model in the validation was 0.854.

Nomogram survival rate model with independent prognostic factors

To further analyze the correlation among the tumor metastasis and PS model status factors and survival prognosis, we conducted a nomogram survival rate model construction analysis with the TCGA samples, a practical way to predict the survival probability for OS patients (Fig 4A). As shown in Fig 4B, the 3- and 5-year C-indexes were 0.834 and 0.869, respectively, suggesting a high prediction of performance.

Construction of ceRNA regulatory network

In this research, we constructed a ceRNA network using the DERs obtained from Step 2.2. This network contained 33 miRNAs, 55 mRNAs, and 22 lncRNAs (Fig 5). We found 3 significant RNAs in the network: *CPA3*, *SERTAD4,* and *GLRB*.

By performing the GO and KEGG analyses for the mRNAs in the ceRNA network, we screened 23

significant correlations in the biological processes and 3 KEGG signal pathways. As shown in Fig 6, these mRNAs were mostly enriched in specific categories, such as positive regulation of the developmental process and regulation of neurogenesis. The results of the KEGG analysis revealed the potential biological relationships between our gene set and the endocytosis, axon guidance, and cancer pathways.

Pathway enrichment analysis of the DEGs between the high- and low-risk groups

We obtained 257 differentially expressed genes (DEGs), including 140 upregulated genes and 117 downregulated genes. Then we performed pathway enrichment analysis on these. As shown in Fig 7, the DEGs were mainly enriched in the biological processes of immune response and cytokine−cytokine receptor interaction.

Discussion

Growing research has proved that RNAs are important prognostic factors in human diseases such as OS. For example, Zhang et al.(19) has reported that LncRNA *CBR3-AS1* is an independent prognostic factor for OS. Using real-time quantitative polymerase chain reaction, Li et al. (20) has verified that *miR-1826* can be a new prognostic marker for OS. *Receptor interacting protein kinase 4 (RIPK4)*，*Matrilin-2 (MATN2),* and many other genes also can play important roles in prognosis(21-23). Therefore, our prediction model integrates multiple types of RNA, which may have better prognostic efficacy.

In our research, we proposed a novel PS model based on miRNA, lncRNA, and mRNA that had a high prediction performance. We analyzed the clinical factors（Age, Gender,Tumor multifocal, Tumor metastatic, Radiotherapy,Tumor necrosis and PS model) of the samples and screened out independent prognostic clinical factors. Through the univariate and multivariate analyses, tumor metastasis and PS model was screened out. Despite combined treatment of extensive resection and chemotherapy, 40%−50% of patients will develop lung metastases (24). Lung metastases remain an important cause of OS-related mortality (25). With the

improvement of treatment technology, the survival rates of nonmetastatic OS have been increased to 65% to 75%. (26-28). But, the survival rates of OS systemic metastasis, especially the occurrence of lung metastasis is still only 11% to 30%(28). The nomogram we constructed, which combines the PS model and tumor metastasis information, can more effectively predict the survival probability of individual patients with OS.

Noncoding RNAs can act as central players in modulating gene expression at multiple levels and can affect diverse aspects of cellular processes, including cell apoptosis, proliferation, cycle, migration, and invasion, as well as drug resistance (29). Most RNAs used in the construction of the PS model are related to malignant tumors. *miR-429* can suppress tumorigenesis in OS by affecting cell proliferation and invasion (30). *miR-526b* can regulate the initiation and progression of non-cardia gastric, esophageal squamous cell, breast, and colon cancers(31-34). *miR-615* plays an important role in renal cell carcinoma progression [(35)]. However, until now, no research has been done on *LINC00626* in cancer.

Carboxypeptidase A3 is a metalloexopeptidase that can be expressed in the subtype of mast cells (36). which can promote the development of certain malignant tumors such as stomach, prostate, or pancreatic cancers(37). Glycine receptors, including its beta receptor subunit, can inhibit neurotransmission. (38). The expression of *BCAS4* is significantly reduced in myelodysplastic syndromes patients (39). *CA9* overexpression is identified as an independent favorable prognostic marker in many cancers such as intrahepatic cholangiocarcinoma, tongue squamous cell carcinoma and (RCC) (40-42). *LOXL3* contributes to proliferation and metastasis in pancreatic ductal adenocarcinoma, gastric cancer and melanoma (43). *NRXN2* is one of neurexins genes and relates to a wide variety of neuropsychiatric disorders (44). RAMP1 plays a critical role in inflammation-related lymphangiogenesis (45). *SUSD2* expression correlates with the progression of lung adenocarcinoma, breast cancer and high grade serous ovarian cancer (46-48). *SERTAD4* is a SERTA domain-containing protein. It can interacts with I-mfa which considered to be candidate tumor suppressor gene (49).

Because post-transcriptional regulation is a complex regulatory network, we should not focus only on miRNA−mRNA silencing mechanisms. The ceRNA network is an effective tool for comprehensively

analyzing the function and regulation mechanisms. We constructed a ceRNA network in OS based on 33 miRNAs, 55 mRNAs, and 22 lncRNAs. Many factors such as heredity, inflammation, and environment can affect OS occurrence and development. OS has a complicated pathophysiological process (50). Due to mRNAs are the implementers of molecular function, we performed a GO analysis; our results revealed that mRNAs in the ceRNA network were mainly enriched in positive regulation of the developmental process and of neurogenesis. We also made a GO-enriched analysis for the DEGs between the high- and low-risk groups. The DEGs enriched the biological processes, including those for immune responses and the cytokine−cytokine receptor interaction pathways.

Conclusions

In our research, we have added the exploration of miRNAs. Therefore, our PS model was constructed using these 3 types of RNA, with high predicting prognosis performance. Combined with tumor metastasis information, this model can help us to predict survival probability through the nomogram, helping us predict the likelihood of a patient's recurrence. Through the ceRNA network and enrichment analysis, we can understand how lncRNAs can affect the function of miRNAs. These RNAs were potential biomarkers for OS diagnosis and prognosis. Our research also has some limitations. We could not explore the 5-year and 10-year survival rates using the nomogram model. In future studies, we will collect more data. In addition, we also need to validate our results through experiment.

Funding: No funding was received to assist with the preparation of this manuscript.

Declaration of Interests: No potential competing interests are reported by the authors.

Acknowledgments: None.

Data availability: The data that support the findings of this study are openly available in TCGA at https://gdc-portal.nci.nih.gov.

231

References

1. Liu Y, Teng Z, Wang Y, Gao P, Chen JJMsmimjoe, research c. Prognostic significance of survivin expression in osteosarcoma patients: a meta-analysis. 2015;21:2877.
2. Allison DC, Carney SC, Ahlmann ER, Hendifar A, Chawla S, Fedenko A, et al. A meta-analysis of osteosarcoma outcomes in the modern medical era. 2012;2012.
3. Yu W, Tang L, Lin F, Li D, Wang J, Yang Y, et al. Stereotactic radiosurgery, a potential alternative treatment for pulmonary metastases from osteosarcoma. 2014;44(4):1091-8.
4. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PPJC. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? 2011;146(3):353-8.
5. Tay Y, Rinn J, Pandolfi PPJN. The multilayered complexity of ceRNA crosstalk and competition. 2014;505(7483):344-52.
6. Karreth FA, Pandolfi PPJCd. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. 2013;3(10):1113-21.
7. Wright MWJHg. A short guide to long non-coding RNA gene nomenclature. 2014;8(1):7.
8. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. 2015;43(7):e47-e.
9. Wang L, Cao C, Ma Q, Zeng Q, Wang H, Cheng Z, et al. RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. 2014;14(1):169.
10. Tibshirani RJSim. The lasso method for variable selection in the Cox model. 1997;16(4):385-95.
11. Goeman JJJBj. L1 penalized estimation in the Cox proportional hazards model. 2010;52(1):70-84.
12. Anderson WI, Schlafer DH, Vesely KRJJowd. Thyroid follicular carcinoma with pulmonary metastases in a beaver (Castor canadensis). 1989;25(4):599-600.
13. Eng KH, Schiller E, Morrell KJO. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. 2015;6(34):36308.
14. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, et al. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. 2016;44(D1):D231-D8.
15. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-HJNar. starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. 2014;42(D1):D92-D7.
16. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. 2003;13(11):2498-504.
17. Sherman BT, Lempicki RAJNp. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. 2009;4(1):44.

18. Huang DW, Sherman BT, Lempicki RAJNar. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. 2009;37(1):1-13.
19. 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.
20. Li P, Wei L, Zhu W. Downregulation of miR-1826 Indicates a Poor Prognosis for Osteosarcoma Patients and Regulates Tumor Cell Proliferation, Migration, and Invasion. International journal of genomics. 2020;2020:7968407.
21. Jiang H, Guo W, Yuan S, Song L. Matrilin-2 is a novel prognostic marker in osteosarcoma. International journal of clinical and experimental pathology. 2019;12(10):3752-60.
22. Liu S, Liu J, Yu X, Shen T, Fu QJFiO. Identification of a two-gene (PML-EPB41) signature with independent prognostic value in osteosarcoma. 2019;9.
23. Yi Z, Pu Y, Gou R, Chen Y, Ren X, Liu W, et al. Silencing of RIPK4 inhibits epithelial -mesenchymal transition by inactivating the Wnt/β-catenin signaling pathway in osteosarcoma. 2020;21(3):1154-62.
24. Kager L, Zoubek A, Po¨tschger U, Kastner U, Flege S, Kempf-Bielack B, et al. Primary metastatic osteosarcoma: presentation and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study Group protocols. 2003;21(10):2011-8.
25. Fan TM, Roberts RD, Lizardo MMJFiO. Understanding and Modeling Metastasis Biology to Improve Therapeutic Strategies for Combating Osteosarcoma Progression. 2020;10.
26. Eilber F, Giuliano A, Eckardt J, Patterson K, Moseley S, Goodnight J. Adjuvant chemotherapy for osteosarcoma: a randomized prospective trial. 1987;5(1):21-6.
27. Allison DC, Carney SC, Ahlmann ER, Hendifar A, Chawla S, Fedenko A, et al. A meta-analysis of osteosarcoma outcomes in the modern medical era. Sarcoma. 2012;2012:704872.
28. Meyers PA, Heller G, Healey JH, Huvos A, Applewhite A, Sun M, et al. Osteogenic sarcoma with clinically detectable metastasis at initial presentation. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1993;11(3):449-53.
29. Wang JY, Yang Y, Ma Y, Wang F, Zhang QAJB, Biomedecine p, et al. Potential regulatory role of lncRNA-miRNA-mRNA axis in osteosarcoma. 2020;121:109627.
30. Li X, Jiang H, Xiao L, Wang S, Zheng JJMsmimjoe, research c. miR-200bc/429 inhibits osteosarcoma cell proliferation and invasion by targeting PMP22. 2017;23:1001.
31. Fan Q-H, Yu R, Huang W-X, Cui X-X, Luo B-H, Zhang L-Y. The has-miR-526b binding-site rs8506G>a polymorphism in the lincRNA-NR_024015 exon identified by GWASs predispose to non-cardia gastric cancer risk. PLoS One. 2014;9(3):e90008-e.
32. Han L, Liu S, Liang J, Guo Y, Shen S, Guo X, et al. A genetic polymorphism at miR‐526b binding‐site in the lincRNA‐NR_024015 exon confers risk of esophageal squamous cell carcinoma in a population of North China. 2017;56(3):960-71.

33. Majumder M, Landman E, Liu L, Hess D, Lala PKJMCR. COX-2 elevates oncogenic miR-526b in breast cancer by EP4 activation. 2015;13(6):1022-33.
34. Zhang R, Zhao J, Xu J, Wang J, Jia JJAjotr. miR-526b-3p functions as a tumor suppressor in colon cancer by regulating HIF-1α. 2016;8(6):2783.
35. Wang Q, Wu G, Zhang Z, Tang Q, Zheng W, Chen X, et al. Long non-coding RNA HOTTIP promotes renal cell carcinoma progression through the regulation of the miR-615/IGF-2 pathway. International journal of oncology. 2018;53(5):2278-88.
36. Balzar S, Fajt ML, Comhair SA, Erzurum SC, Bleecker E, Busse WW, et al. Mast cell phenotype, location, and activation in severe asthma: data from the severe asthma research program. 2011;183(3):299-309.
37. Varricchi G, Galdiero MR, Loffredo S, Marone G, Iannone R, Marone G, et al. Are mast cells MASTers in cancer? 2017;8:424.
38. Lueken U, Kuhn M, Yang Y, Straube B, Kircher T, Wittchen H, et al. Modulation of defensive reactivity by GLRB allelic variation: converging evidence from an intermediate phenotype approach. 2017;7(9):e1227-e.
39. Shiseki M, Ishii M, Okada M, Ohwashi M, Wang YH, Osanai S, et al. Expression analysis of genes located within the common deleted region of del(20q) in patients with myelodysplastic syndromes. Leukemia research. 2019;84:106175.
40. Gu M. CA9 overexpression is an independent favorable prognostic marker in intrahepatic cholangiocarcinoma. International journal of clinical and experimental pathology. 2015;8(1):862-6.
41. Guan C, Ouyang D, Qiao Y, Li K, Zheng G, Lao X, et al. CA9 transcriptional expression determines prognosis and tumour grade in tongue squamous cell carcinoma patients. Journal of cellular and molecular medicine. 2020;24(10):5832-41.
42. Li G, Bilal I, Gentil-Perret A, Feng G, Zhao A, Peoc'h M, et al. CA9 as a molecular marker for differential diagnosis of cystic renal tumors. Urologic oncology. 2012;30(4):463-8.
43. Laurentino TS, Soares RDS, Marie SKN, Oba-Shinjo SM. LOXL3 Function Beyond Amino Oxidase and Role in Pathologies, Including Cancer. International journal of molecular sciences. 2019;20(14).
44. Kasem E, Kurihara T, Tabuchi K. Neurexins and neuropsychiatric disorders. Neuroscience research. 2018;127:53-60.
45. Tsuru S, Ito Y, Matsuda H, Hosono K, Inoue T, Nakamoto S, et al. RAMP1 signaling in immune cells regulates inflammation-associated lymphangiogenesis. Laboratory investigation; a journal of technical methods and pathology. 2020;100(5):738-50.
46. Guo W, Shao F, Sun S, Song P, Guo L, Xue X, et al. Loss of SUSD2 expression correlates with poor prognosis in patients with surgically resected lung adenocarcinoma. Journal of Cancer. 2020;11(7):1648-56.
47. Hultgren EM, Patrick ME, Evans RL, Stoos CT, Egland KA. SUSD2 promotes tumor-associated macrophage recruitment by increasing levels of MCP-1 in breast cancer. PLoS One.

2017;12(5):e0177089.
48. Xu Y, Miao C, Jin C, Qiu C, Li Y, Sun X, et al. SUSD2 promotes cancer metastasis and confers cisplatin resistance in high grade serous ovarian cancer. Experimental cell research. 2018;363(2):160-70.
49. Kusano S, Shiimura Y, Eizuru Y. I-mfa domain proteins specifically interact with SERTA domain proteins and repress their transactivating functions. Biochimie. 2011;93(9):1555-64.
50. Zhang G, Li Y, Xu J, Xiong ZJOM. Advances in the role of miRNAs in the occurrence and development of osteosarcoma. 2020;15(1):1003-11.

Figure captions

Fig 1. The DER results. **A.** Volcano plot. The horizontal dashed line indicates FDR < 0.05; the two vertical dashed lines indicate | log2FC | >0.5. The point size represents the absolute value of logFC; the larger the value, the larger the point. **B.** Heatmap.

Fig 2. A. Kaplan–Meier curve of tumor metastasis related to prognosis in the training data set. Blue and red curves indicate no-recurrence and recurrence groups for the OS tumor samples, respectively. **B.** The OS sample group with and without tumor metastasis is based on the PS prediction model and the prognosis-related Kaplan–Meier curve line graphs. Blue: low-risk samples; red: high-risk samples.

Fig 3. Training and validation data sets used in the study. **A.** Training set. **B.** Validation set. Left: Kaplan–Meier survival plots of low-grade and high-grade samples. Right: receiver operating characteristic (ROC) curve for the PS model. The area under the curve (AUC) values of 0.966 and 0.854 showed a good performance of the risk prediction.

Fig 4. Nonogram and concordance plots for the study. A. Nomogram of prognosis. **B.** Concordance plots of predictions of 3- and 5-year recurrence survival probability with actual recurrence survival probability. Red: 3 years; blue: 5 years.

Fig 5. ceRNA network. Squares: lncRNA; triangles: miRNA; circles: mRNA. The change in color from blue to red indicates the change in logFC expression from down-regulation to up-regulation, respectively, and the nodes with larger numbers represent signature RNAs.

372 **Fig 6.** GO analysis of the mRNA in the ceRNA network. Horizontal axis: number of genes; vertical axis: GO
373 entry name. Dot size: number of genes involved; dot color: −log10 (FDR). The closer the color is to red, the
374 higher the significance.

375 **Fig 7.** Comparisons between the high- and low-risk groups. (A) GO results for the DEmRNAs between the
376 high- and low-risk groups. (B) 8 enrichment of KEGG pathway analysis of DEmRNAs between the high- and
377 low-risk groups. Horizontal axis: number of genes; vertical axis: GO entry name. Dot size: number of genes
378 involved; dot color: −log10 (FDR). The closer the color is to red, the higher the significance.

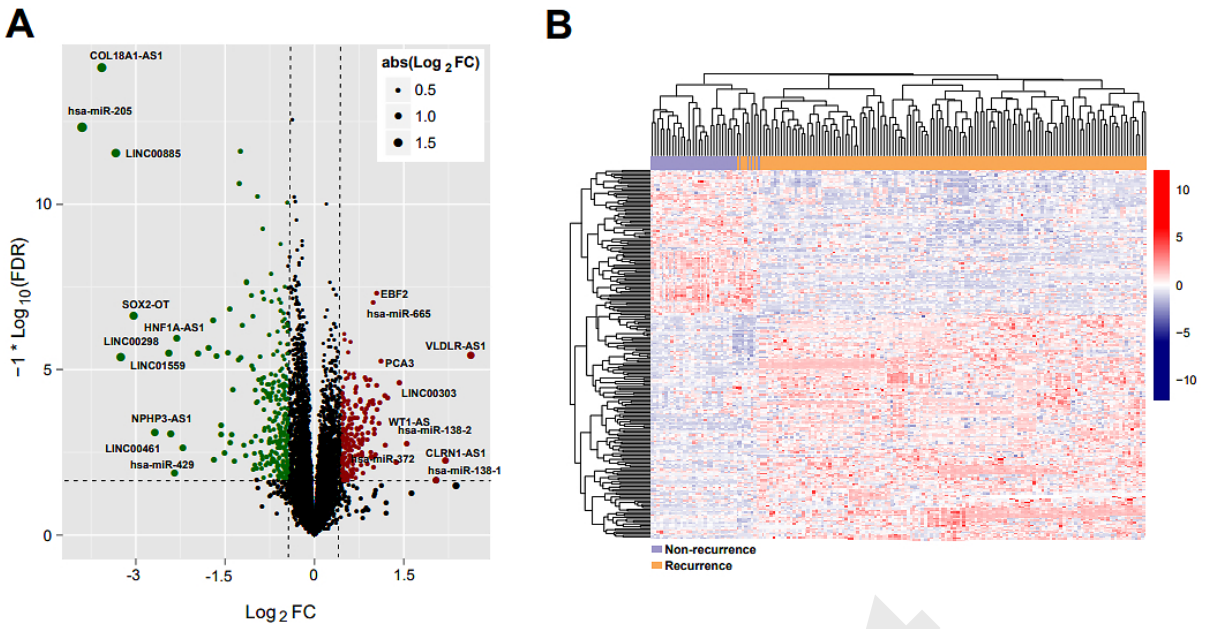
1 **Table 1. The optimal combination of signature DERs.**

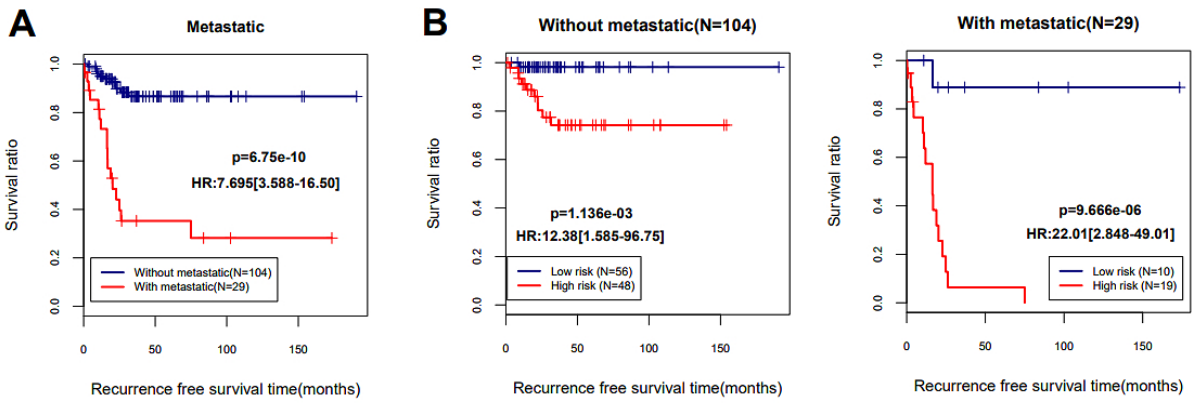
2

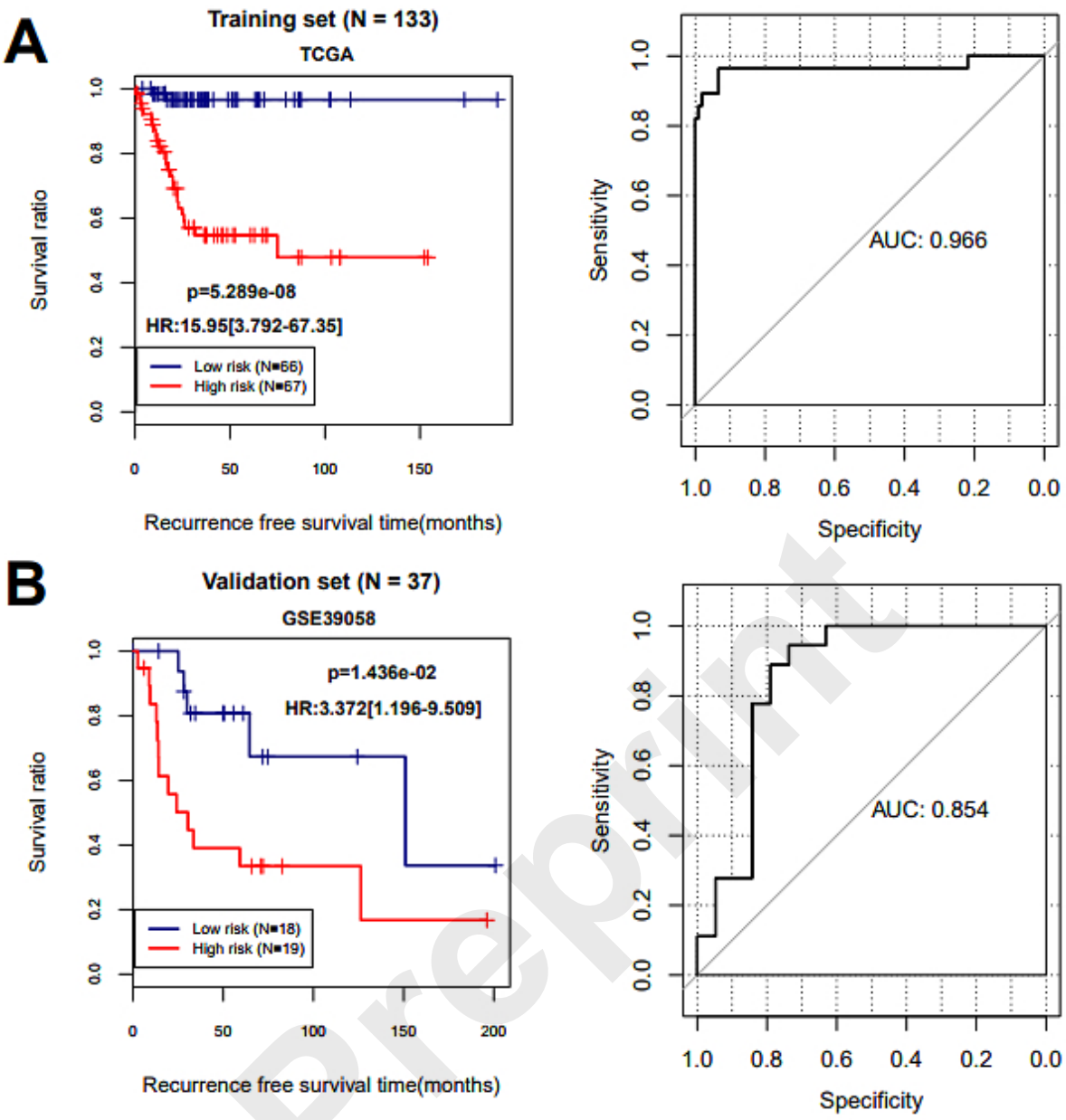
Symbol	Type	Multi-variate Cox regression analysis			LASSO coefficient
		HR	95%CI	P value	
LINC00626	lncRNA	1.0056	1.0011-1.0101	3.840E-03	0.02461
has-miR-429	miRNA	1.0166	1.0092-1.042	4.858E-02	0.33082
hsa-miR-526b	miRNA	1.0211	1.0099-1.043	1.565E-02	0.24799
hsa-miR-615	miRNA	0.9967	0.9934-0.9999	1.155E-02	-0.14581
BCAS4	mRNA	0.9663	0.9419-0.9914	8.742E-03	-0.67382
CA9	mRNA	1.0111	1.0031-1.0191	6.424E-03	0.17397
CPA3	mRNA	1.0125	1.0013-1.0238	2.826E-02	0.13737
GLRB	mRNA	0.9613	0.9375-0.9856	1.965E-03	-0.18084
LOXL3	mRNA	0.9776	0.9644-0.9911	1.158E-03	-0.21229
NRXN2	mRNA	0.9154	0.9035-0.9275	1.108E-02	-0.13537
RAMP1	mRNA	0.9911	0.9826-0.9997	4.173E-02	-0.03370
SERTAD4	mRNA	0.9864	0.973-0.9999	4.833E-02	-0.50416
SUSD2	mRNA	1.0123	1.0034-1.0212	6.659E-03	0.29587

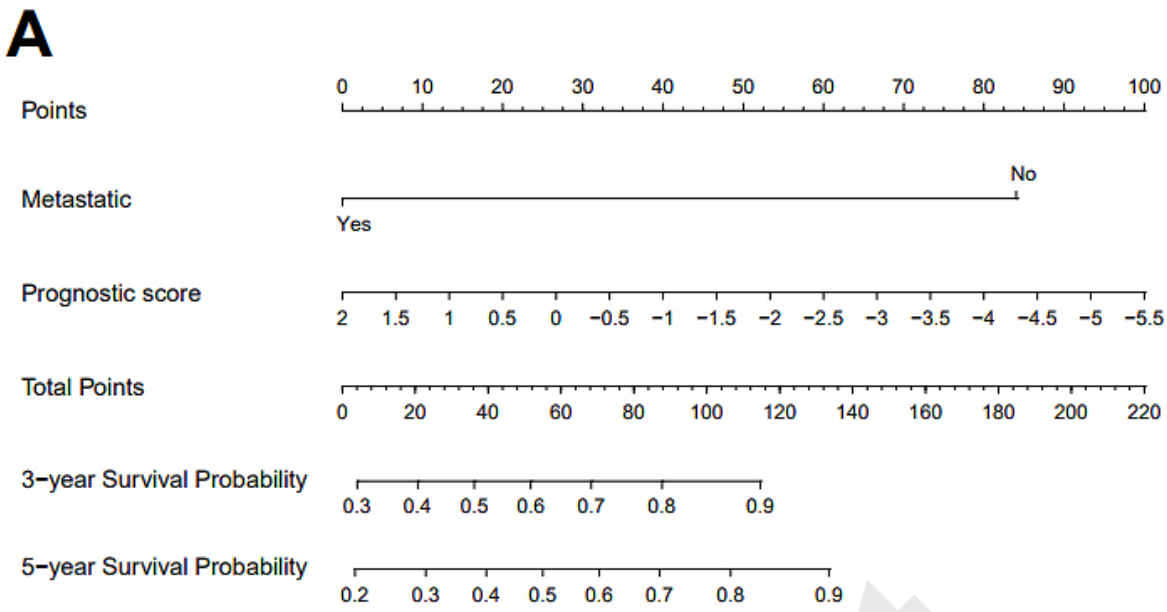
Table 2. Univariate and multivariate Cox regression analysis for clinical factors.

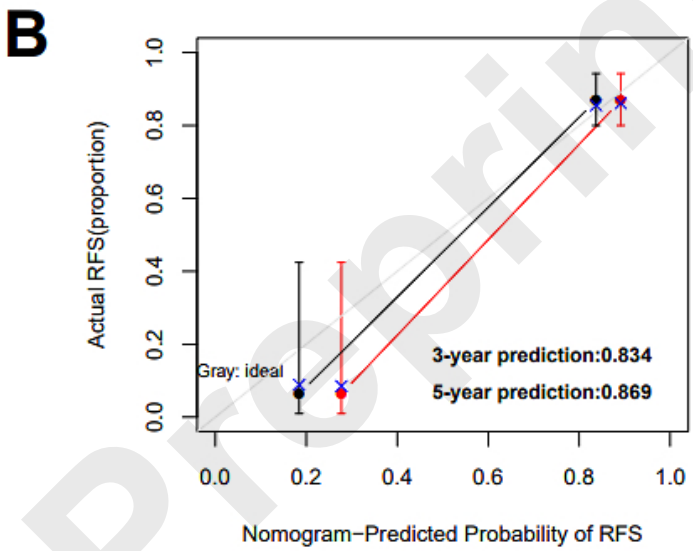
Clinical characteristics	TCGA(N=133)	Uni-variables cox			Multi-variables cox		
		HR	95%CI	P	HR	95%CI	P
Age(years,mean±sd)	38.81±35.63	1.001	0.977-1.025	9.59E-01	-	-	-
Gender(Male/Female)	52/81	1.935	0.920-4.070	8.24E-02	-	-	-
Tumor multifocal(Yes/No/-)	27/98/8	2.803	1.290-6.092	6.56E-03	1.688	0.753-3.784	2.04E-01
Tumor metastatic(Yes/No)	29/104	7.695	3.588-16.50	6.75E-10	7.094	3.179-15.83	1.70E-06
Radiotherapy(Yes/No)	46/87	1.262	0.597-2.670	5.42E-01	-	-	-
Tumor necrosis(No/Slight/Moderate/Severe/-)	49/26/44/6/8	1.257	0.842-1.876	2.60E-01	-	-	-
PS model(High/ Low)	68/69	15.98	3.792-67.35	5.29E-08	14.01	3.309-59.35	3.38E-04
Tumor recurrence(Yes/No)	28/105	-	-	-	-	-	-
Recurrence free survival time(months,mean±sd)	60.40±15.51	-	-	-	-	-	-

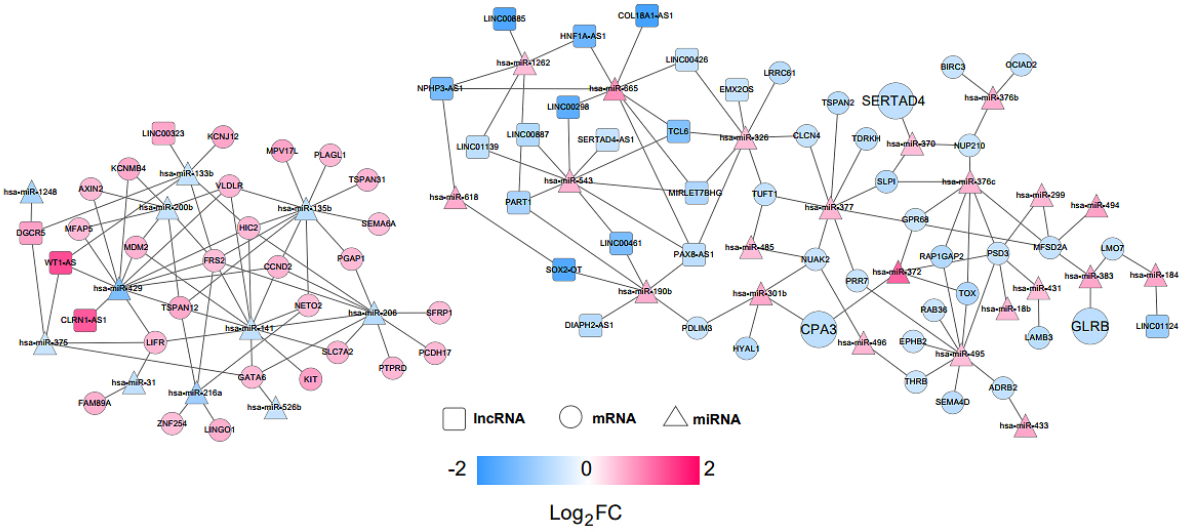

A
B
COL18A1-AS1
hsa-miR-205
LINC00885
SOX2-OT
HNF1A-AS1
LINC00298
LINC01559
NPHP3-AS1
LINC00461
hsa-miR-429
EBF2
hsa-miR-665
VLDLR-AS1
PCA3
LINC00303
WT1-AS
hsa-miR-138-2
hsa-miR-372
CLRN1-AS1
hsa-miR-138-1
abs(Log 2 FC)
0.5
1.0
1.5
−1 * Log 10(FDR)
Log 2 FC
Non-recurrence
Recurrence

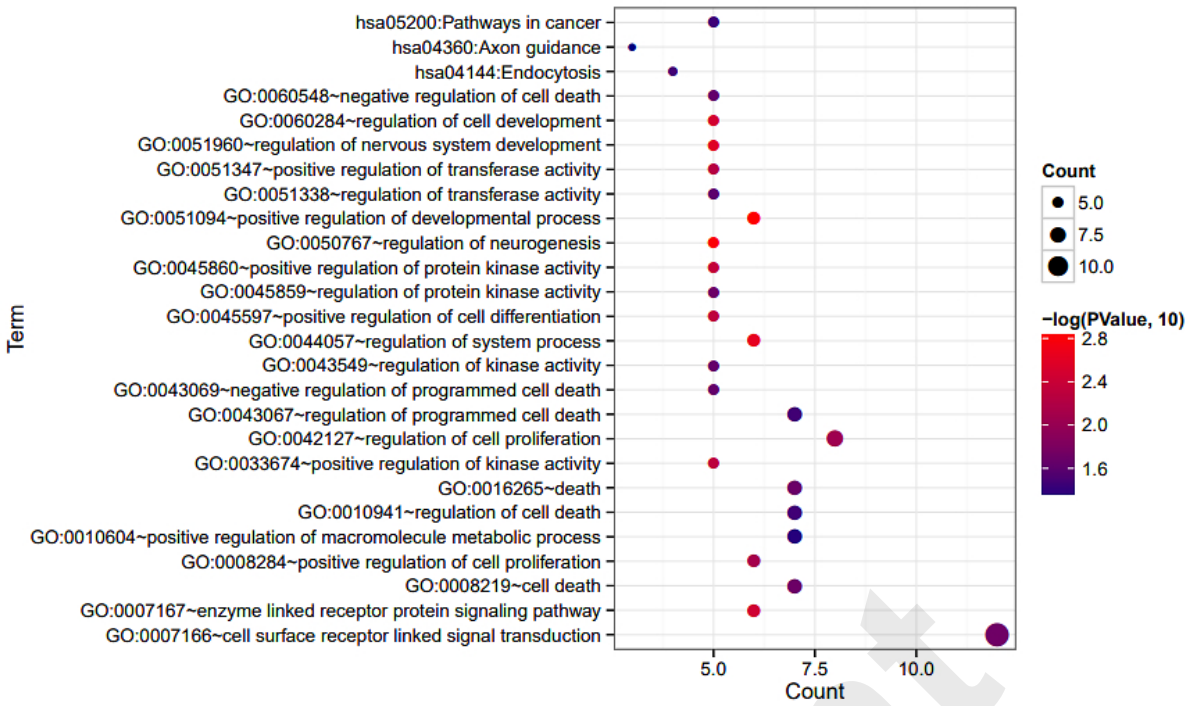

A
Metastatic
Survival ratio
p=6.75e-10
HR:7.695[3.588-16.50]
Without metastatic(N=104)
With metastatic(N=29)
Recurrence free survival time(months)
B
Without metastatic(N=104)
Survival ratio
p=1.136e-03
HR:12.38[1.585-96.75]
Low risk (N=56)
High risk (N=48)
Recurrence free survival time(months)
With metastatic(N=29)
Survival ratio
p=9.666e-06
HR:22.01[2.848-49.01]
Low risk (N=10)
High risk (N=19)
Recurrence free survival time(months)

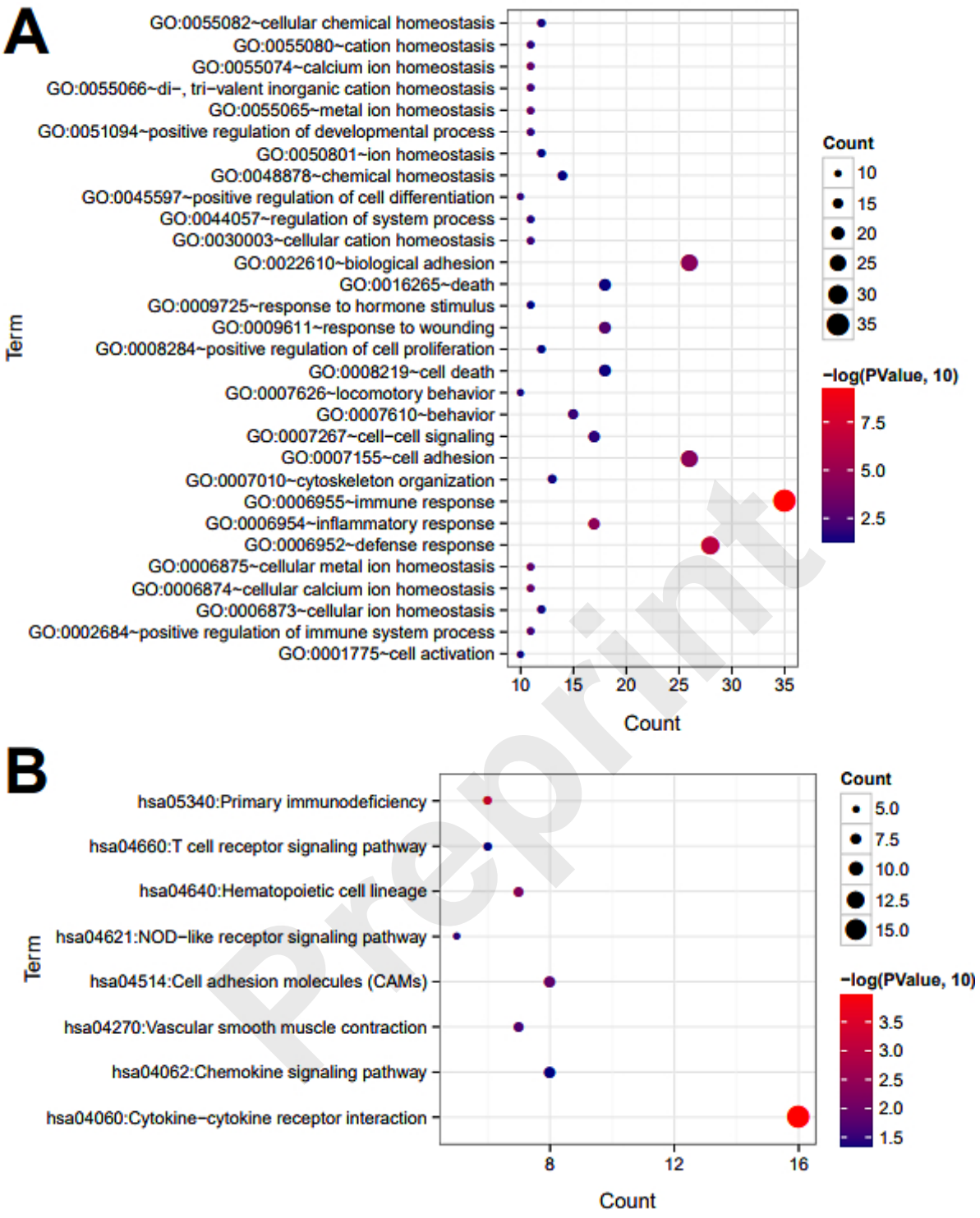

A
Training set (N = 133)
TCGA
Survival ratio
p=5.289e-08
HR:15.95[3.792-67.35]
Low risk (N=66)
High risk (N=67)
Recurrence free survival time(months)
Sensitivity
AUC: 0.966
Specificity
B
Validation set (N = 37)
GSE39058
Survival ratio
p=1.436e-02
HR:3.372[1.196-9.509]
Low risk (N=18)
High risk (N=19)
Recurrence free survival time(months)
Sensitivity
AUC: 0.854
Specificity


A
Points
0 10 20 30 40 50 60 70 80 90 100
Metastatic
No
Yes
Prognostic score
2 1.5 1 0.5 0 −0.5 −1 −1.5 −2 −2.5 −3 −3.5 −4 −4.5 −5 −5.5
Total Points
0 20 40 60 80 100 120 140 160 180 200 220
3−year Survival Probability
0.3 0.4 0.5 0.6 0.7 0.8 0.9
5−year Survival Probability
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9


B
Actual RFS(proportion)
0.0 0.2 0.4 0.6 0.8 1.0
Gray: ideal
3-year prediction:0.834
5-year prediction:0.869
0.0 0.2 0.4 0.6 0.8 1.0
Nomogram−Predicted Probability of RFS


LINC00685
COL18A1-AS1
HNF1A-AS1
hsa-miR-1262
LINC00426
LRRC61
NPHP3-AS1
hsa-miR-365
EMX2OS
LINC00298
TSPAN2
SERTAD4
BIRC3
OCIAD2
hsa-miR-376b
LINC01139
LINC00887
SERTAD4-AS1
TCL6
hsa-miR-326
CLCN4
TDRKH
hsa-miR-370
NUP210
hsa-miR-543
SLPI
hsa-miR-376c
hsa-miR-618
PART1
MIRLET7BHG
TUFT1
hsa-miR-377
GPR68
hsa-miR-299
hsa-miR-494
LINC00461
PAX8-AS1
hsa-miR-485
NUAK2
RAP1GAP2
PSD3
MFSD2A
LMO7
SOX2-OT
hsa-miR-372
PRR7
hsa-miR-383
hsa-miR-184
hsa-miR-190b
TOX
hsa-miR-431
hsa-miR-301b
RAB36
hsa-miR-18b
DIAPH2-AS1
PDLIM5
CPA3
GLRB
LINC01124
HYAL1
hsa-miR-496
EPHB2
LAMB3
hsa-miR-495
THRB
ADRB2
SEMA4D
hsa-miR-433
LINC00323
KCNJ12
MPV17L
PLAGL1
KCNMB4
hsa-miR-133b
TSPAN31
hsa-miR-1248
AXIN2
VLDLR
hsa-miR-200b
hsa-miR-135b
SEMA6A
DGCR5
MFAP5
HIC2
MDM2
FRS2
CCND2
PGAP1
WT1-AS
hsa-miR-429
TSPAN12
NETO2
hsa-miR-206
SFRP1
CLRN1-AS1
hsa-miR-375
LIFR
hsa-miR-141
SLC7A2
PCDH17
PTPRD
GATA6
KIT
hsa-miR-31
hsa-miR-216a
FAM89A
hsa-miR-526b
ZNF254
LINGO1
lncRNA
mRNA
miRNA
-2
0
2
Log2FC


hsa05200:Pathways in cancer
hsa04360:Axon guidance
hsa04144:Endocytosis
GO:0060548~negative regulation of cell death
GO:0060284~regulation of cell development
GO:0051960~regulation of nervous system development
GO:0051347~positive regulation of transferase activity
GO:0051338~regulation of transferase activity
GO:0051094~positive regulation of developmental process
GO:0050767~regulation of neurogenesis
GO:0045860~positive regulation of protein kinase activity
GO:0045859~regulation of protein kinase activity
GO:0045597~positive regulation of cell differentiation
GO:0044057~regulation of system process
GO:0043549~regulation of kinase activity
GO:0043069~negative regulation of programmed cell death
GO:0043067~regulation of programmed cell death
GO:0042127~regulation of cell proliferation
GO:0033674~positive regulation of kinase activity
GO:0016265~death
GO:0010941~regulation of cell death
GO:0010604~positive regulation of macromolecule metabolic process
GO:0008284~positive regulation of cell proliferation
GO:0008219~cell death
GO:0007167~enzyme linked receptor protein signaling pathway
GO:0007166~cell surface receptor linked signal transduction
Term
5.0
7.5
10.0
Count
Count
5.0
7.5
10.0
−log(PValue, 10)
2.8
2.4
2.0
1.6


A
GO:0055082~cellular chemical homeostasis
GO:0055080~cation homeostasis
GO:0055074~calcium ion homeostasis
GO:0055066~di−, tri−valent inorganic cation homeostasis
GO:0055065~metal ion homeostasis
GO:0051094~positive regulation of developmental process
GO:0050801~ion homeostasis
GO:0048878~chemical homeostasis
GO:0045597~positive regulation of cell differentiation
GO:0044057~regulation of system process
GO:0030003~cellular cation homeostasis
GO:0022610~biological adhesion
GO:0016265~death
GO:0009725~response to hormone stimulus
GO:0009611~response to wounding
GO:0008284~positive regulation of cell proliferation
GO:0008219~cell death
GO:0007626~locomotory behavior
GO:0007610~behavior
GO:0007267~cell−cell signaling
GO:0007155~cell adhesion
GO:0007010~cytoskeleton organization
GO:0006955~immune response
GO:0006954~inflammatory response
GO:0006952~defense response
GO:0006875~cellular metal ion homeostasis
GO:0006874~cellular calcium ion homeostasis
GO:0006873~cellular ion homeostasis
GO:0002684~positive regulation of immune system process
GO:0001775~cell activation
Term
10
15
20
25
30
35
Count
Count
10
15
20
25
30
35
−log(PValue, 10)
7.5
5.0
2.5
B
hsa05340:Primary immunodeficiency
hsa04660:T cell receptor signaling pathway
hsa04640:Hematopoietic cell lineage
hsa04621:NOD−like receptor signaling pathway
hsa04514:Cell adhesion molecules (CAMs)
hsa04270:Vascular smooth muscle contraction
hsa04062:Chemokine signaling pathway
hsa04060:Cytokine−cytokine receptor interaction
Term
8
12
16
Count
Count
5.0
7.5
10.0
12.5
15.0
−log(PValue, 10)
3.5
3.0
2.5
2.0
1.5