



A potential prognostic prediction model for metastatic osteosarcoma based on bioinformatics analysis

Yan WANG, Guangfu MING, Bohua GAO

Department of Orthopedics, Hainan General Hospital (Hainan Affiliated Hospital of Hainan Medical University), Xiuying district, Haikou, China.

Correspondence at: Bohua Gao, Phone:+86-137075772257, Fax: +86-0898-68642204, E-mail: abzyxymgf@163.com

Osteosarcoma (OS) is a malignant primary bone tumor with a high incidence. This study aims to construct a prognostic prediction model by screening the prognostic mRNA of metastatic OS. Data on four eligible expression profiles from the National Center for Biotechnology Information Gene Expression Omnibus repository were obtained based on inclusion criteria and defined as the training set or the validation set. The differentially expressed genes (DEGs) between metastatic and non-metastatic OS samples in the training set were first identified, and DEGs related to prognosis were screened by univariate Cox regression analysis. In total, 107 DEGs related to the prognosis of metastatic OS were identified. Then, 46 DEGs were isolated as the optimized prognostic gene signature, and a metastatic-OS discriminating classifier was constructed, which had a high accuracy in distinguishing metastatic from non-metastatic OS samples. Furthermore, four optimized prognostic gene signatures (ALOX5AP, COL21A1, HLA-DQB1, and LDHB) were further screened, and the prognostic prediction model for metastatic OS was constructed. This model possesses a relatively satisfying prediction ability both in the training set and validation set. The prognostic prediction model that was constructed based on the four prognostic mRNA signatures has a high predictive ability for the prognosis of metastatic OS.

Keywords: osteosarcoma, tumor metastasis, differentially expressed genes, prognosis prediction model.

INTRODUCTION

Osteosarcoma (OS) is a malignant primary bone tumor, with an annual incidence of 3.4/1,000,000 worldwide¹. Currently, the conventional treatment method for OS is a combination of surgery and adjuvant chemotherapy, which has improved long-term outcomes for patients with a localized OS². The five-year survival rate of patients with non-metastatic OS has reached 65% to 70%, whereas no significant improvements have been found in patients with metastatic OS, who have a poor prognosis and an overall survival rate of 20% to 30%³. Despite the significant difference in the survival rate between patients with metastatic and non-metastatic OS in the last three decades, effective treatment for metastatic OS is still lacking^{1,4}. An understanding of the mechanisms of OS metastasis is urgently needed¹. Therefore, it is crucial to identify novel diagnostic markers and therapeutic methods, as well as underlying prognostic factors for patients with metastatic OS.

Encouragingly, several underlying mechanisms of OS have been identified based on a wide range of analytic approaches, including conventional experimental techniques, association analysis, microarrays,

and high-throughput sequencing^{5,6}. Fuja *et al.* have emphasized that transglutaminase-2 (TGM2) is under-expressed in metastatic OS lesions compared with primary bone tumors, and TGM2 knockdown inhibits the proliferation and migration of metastatic OS cells in vitro, as well as suppressing metastatic potential and improving survival outcome in vivo⁷. Another research group focused on Notch signaling abnormalities in OS, and they found that Notch3 is a prognostic factor for OS. They then further demonstrated that Notch3 silencing reduced the OS-associated pulmonary metastasis through the effector matrix metalloproteinase 7 and its downstream target gene hairy and enhancer of split-1⁸. Additionally, it has been reported that a cluster of differentiation 44 (CD44) is also associated with the high potential for metastasis and poor survival in patients with OS; in vitro experiments have revealed that CD44 silencing impairs the migratory and invasive functions in highly metastatic OS cell lines⁹. However, few studies have investigated the prognostic prediction model of metastatic OS.

In the current study, the prognostic genes of metastatic OS were screened based on expression profile data from the National Center for Biotechnology

Information (NCBI) Gene Expression Omnibus (GEO) repository. Then, these prognostic genes were used to construct a metastatic-OS discriminating classifier as well as a prognostic prediction model of metastatic OS using bioinformatics methods. We aimed to provide some useful insights for improving the prognosis of patients with metastatic OS.

MATERIALS AND METHODS

The data for the expression profiles were preliminarily screened according to the search words “osteosarcoma and Homo sapiens” from the NCBI GEO repository (<http://www.ncbi.nlm.nih.gov/geo/>). Then, the eligible data were included in this study as the following inclusion criteria: (1) the selected datasets were gene expression profiles data; (2) the samples in the data were solid tissues of OS; (3) the samples were divided into metastatic and non-metastatic OS tissues; (4) the total number of samples was not less than 30, and (5) the datasets contained survival and prognosis information of the samples. A dataset that met the (1), (2), (3), and (4) search condition was used as a validation set of the support vector machine (SVM) classifier. The dataset that met the (1), (2), (4), and (5) search condition was used as a validation set for the prognostic model. According to the above search strategy, the eligible datasets were included as follows.

Dataset A: The gene expression profile data (GSE21257)¹⁰ contained 34 metastatic samples and 19 non-metastatic OS samples, and clinical survival and prognosis information were also obtained. All samples were sequenced on the Illumina human-6 v2.0 expression BeadChip. The dataset that met all the search conditions was used as the training set.

Dataset B: The gene expression profile data (GSE87624)¹¹ included three normal samples, 8 metastatic samples, 25 non-metastatic samples, and 16 OS samples with an unknown nature of the metastatic. All the samples were sequenced on an Illumina HiSeq 2000 (Homo sapiens) and used as the validation set for the SVM classifier.

Dataset C: The gene expression profile data (GSE39055)¹² included 37 OS samples with clinical survival and prognosis information. All the samples were sequenced on an Illumina HumanHT-12 WG-DASL V4.0 R2 expression BeadChip and used as the validation set for the prognostic model.

Dataset D: The gene expression profile data (GSE1609)¹³ included 34 OS samples with clinical survival and prognosis information. All the samples were sequenced on an Affymetrix Human Genome

U133A Array and used as the validation set for the prognostic model.

Data preprocessing and screening of differentially expressed mRNAs

Limma package (version 3.34.0, <https://bioconductor.org/packages/release/bioc/html/limma.html>)¹⁴ in R 3.4.1 was used for data transformation of raw datasets A and C of TXT format, and data standardization was performed using the median method. Meanwhile, for dataset D, raw data were pre-processed using oligo (version 1.41.1, <http://www.bioconductor.org/packages/release/bioc/html/oligo.html>)¹⁵ in R 3.4.1, including the missing value supplement using the median method, background correction by the MAS method, and normalization using the quantiles method. Also, data in dataset B were normalized using the quantiles method based on preprocessCore (Version 1.44.0, <https://www.bioconductor.org/packages/release/bioc/html/preprocessCore.html>)¹⁶ in R 3.4.1. Based on the normalized data in the training set (dataset A), limma package (version 3.34.7)¹⁴ in R 3.4.1 was used to screen DEGs between 34 metastatic and 19 non-metastatic OS samples. The thresholds were set as $|\log \text{ fold change (FC)}| > 0.5$ and false discovery rate (FDR) < 0.05 . Then, bidirectional hierarchical clustering based on a centered Pearson correlation algorithm was performed by pheatmap (version 1.0.8, <https://cran.r-project.org/web/packages/pheatmap/index.html>)¹⁷ according to the expression values of the DEGs in the training set. Furthermore, based on the Database for Annotation, Visualization and Integrated Discovery (DAVID) program (version 6.8, <https://david.ncifcrf.gov/>)^{18,19}, gene ontology (GO) functional annotation associated with biological process analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed with P-value < 0.05 as the screening threshold.

Identification of differentially expressed mRNAs related to prognosis

DEGs related to the overall survival were screened by the univariate Cox regression analysis survival package (version 2.41.3, <https://cran.r-project.org/web/packages/survival/index.html>)²⁰ in R 3.4.1 with the threshold of log-rank P-value < 0.05 .

The establishment of the SVM classifier

First, based on the DEGs related to prognosis, the optimized DEG signatures were screened by a recursive

feature elimination (RFE) algorithm using the caret package (version 6.0-76, <https://cran.r-project.org/web/packages/caret>)^{21,22} in R 3.4.3. Then, to evaluate the sensitivity and specificity of the feature gene in the discriminative prediction, the SVM-based classifier was established using the e1071 package (<https://cran.r-project.org/web/packages/e1071>)²³ in R 3.4.3. The model was built based on the training set using default parameters (Sigmoid Kernel and 100-fold cross-validation). Finally, receiver operating characteristic (ROC) curve analysis was used to evaluate the predictive power of this SVM classifier in the training set and validation set (dataset B, excluding the 16 samples with unknown nature regarding the metastatic status). Also, the sensitivity, specificity, positive prediction value (PPV), and negative prediction value (NPV) were calculated using pROC (version 1.12.1, <https://cran.r-project.org/web/packages/pROC/index.html>)²⁴ in R 3.4.1.

Construction of a prognostic prediction model

Combined with the clinical prognostic information in the training set and the feature genes based on the SVM classifier, the optimized set of prognostic gene signatures was identified using the Cox proportional-hazards (Cox-PH) model²⁵, which is based on the L1-penalized regularization regression algorithm of the penalized package (version 0.9.50, <http://bioconductor.org/packages/penalized/>)²⁶ in R 3.4.1. The optimized parameter, lambda, in this model was obtained by 1,000 cycles of calculation of a cross-validation likelihood (cvl) algorithm. Subsequently, the prognostic prediction model was constructed based on the prognostic regression coefficient of the DEGs, and the risk score (RS) for each sample was calculated as $RS = \sum \text{coef}_{\text{DEGs}} \times \text{Exp}_{\text{DEGs}}$. The $\text{Coef}_{\text{gene}}$ represents the prognostic regression coefficient, and Exp_{DEGs} is defined as the expression value of the corresponding gene. According to the median value of RS, all samples in the training set were divided into high-risk and low-risk groups. The Kaplan-Meier (K-M) survival curve analysis based on the survival package (version 2.41.1) in R 3.4.1 was used to estimate this model. This model was further verified in the validation set of the prognostic model (datasets C and D). The workflow for this study is shown in Figure 1.

RESULTS

A total of 400 DEGs were identified between the metastatic and non-metastatic OS samples based on

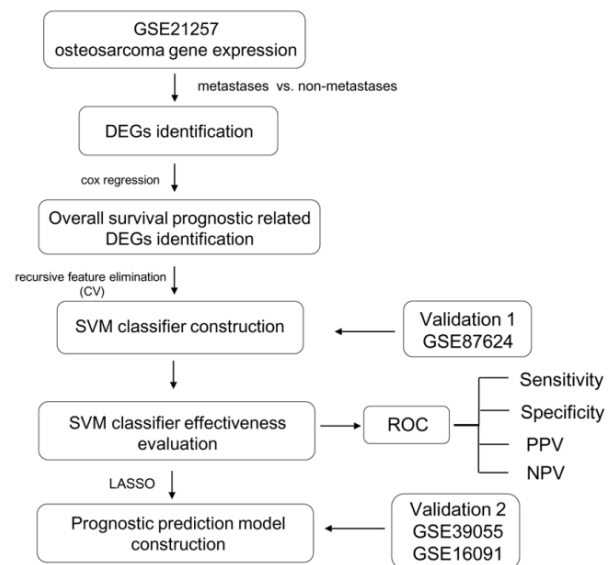


Figure 1 — The workflow of the construction of the prognostic prediction model for metastatic osteosarcoma.

the selective criteria, including 194 downregulated and 206 upregulated DEGs (Figure 2A). Then, bidirectional hierarchical clustering was conducted for these 400 DEGs, indicating that these DEGs could significantly distinguish between metastatic and non-metastatic OS samples (Figure 2B). To further identify the functional characteristics of DEGs, the functional enrichment analysis of genes was conducted using DAVID. Consequently, the GO analysis of the DEGs revealed that the significantly enriched terms primarily concentrated on immune response, inflammatory response, the interferon-gamma-mediated signaling pathway, antigen processing and presentation, and innate immune response (Figure 2C). Also, the KEGG pathway analysis implied that these DEGs were responsible for antigen processing and presentation, cell adhesion molecules, allograft rejection, complement and coagulation cascades, and osteoclast differentiation (Figure 2C).

The univariate Cox regression analyses in the training set revealed that a total of 107 DEGs are significantly associated with the patient's prognosis. The RFE algorithm showed that the optimized DEGs' signatures contain 46 genes when accuracy is the highest value (Figure 3A). Then, a metastatic-OS discriminating SVM classifier was established based on the optimized 46 DEG signatures. The sample distribution based on this classifier in the training set and validation set is shown in Figure 3B and 3C, respectively. The ROC curve analysis revealed that the area under the curve (AUC), Sen, Spe, PPV, and NPV of this SVM classifier are 0.986, 0.917, 0.941, 0.971, and 0.842 in the training

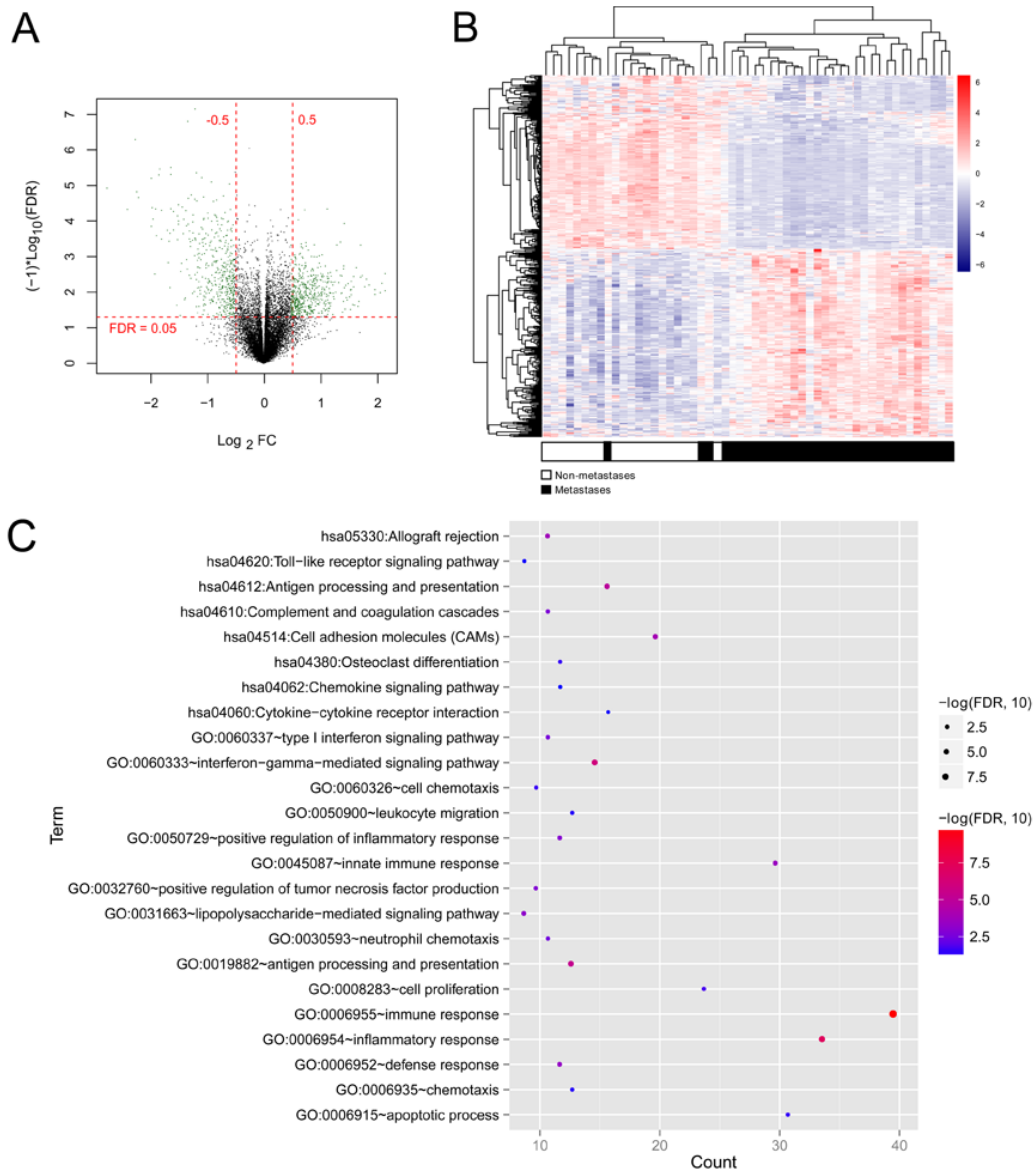


Figure 2 — Identification and functional characteristics of differentially expressed genes. A. Volcano map. Green dots indicate DEGs, and the red horizontal dotted line represents the false discovery rate (FDR) <0.05; two red vertical dotted lines represent $|\log \text{fold change (FC)}| > 0.5$. B. A bidirectional hierarchical clustering map based on 400 DEGs. White and black sample bars represent non-metastatic and metastatic OS samples, respectively. C. GO and KEGG enrichment analyses of DEGs. The horizontal axis and vertical axis represent the gene number and term, respectively; the color and size of the dots indicate the significant P-value, and the closer the color is to red, the higher the significance.

set, respectively. Similarly, these evaluation indexes are 0.847, 0.875, 0.960, 0.889, and 0.917 in the validation set, respectively (Figure 3B and 3C). These results indicate that this SMV classifier has high accuracy when distinguishing metastatic and non-metastatic OS samples.

Based on these 46 DEGs, the optimized set of prognostic gene signatures was screened by Cox P-H. After 1000 cycles of calculation of the cvl algorithm, lambda was confirmed as 7.8565, and the maximum

value of the cvl was -101.8146 (Figure 4A). As a result, four optimized prognostic gene signatures were identified, including ALOX5AP, COL21A1, HLA-DQB1, and LDHB (Figure 4B, Table 1). According to the prognostic regression coefficient of these DEGs, the prognostic prediction model was constructed, and RS was calculated as $RS = (-0.1453) \times \text{Exp}_{\text{ALOX5AP}} + (0.0491) \times \text{Exp}_{\text{COL21A1}} + (-0.0188) \times \text{Exp}_{\text{HLA-DQB1}} + (0.1567) \times \text{Exp}_{\text{LDHB}}$. The estimation of the K-M survival analysis showed that the overall survival time of the

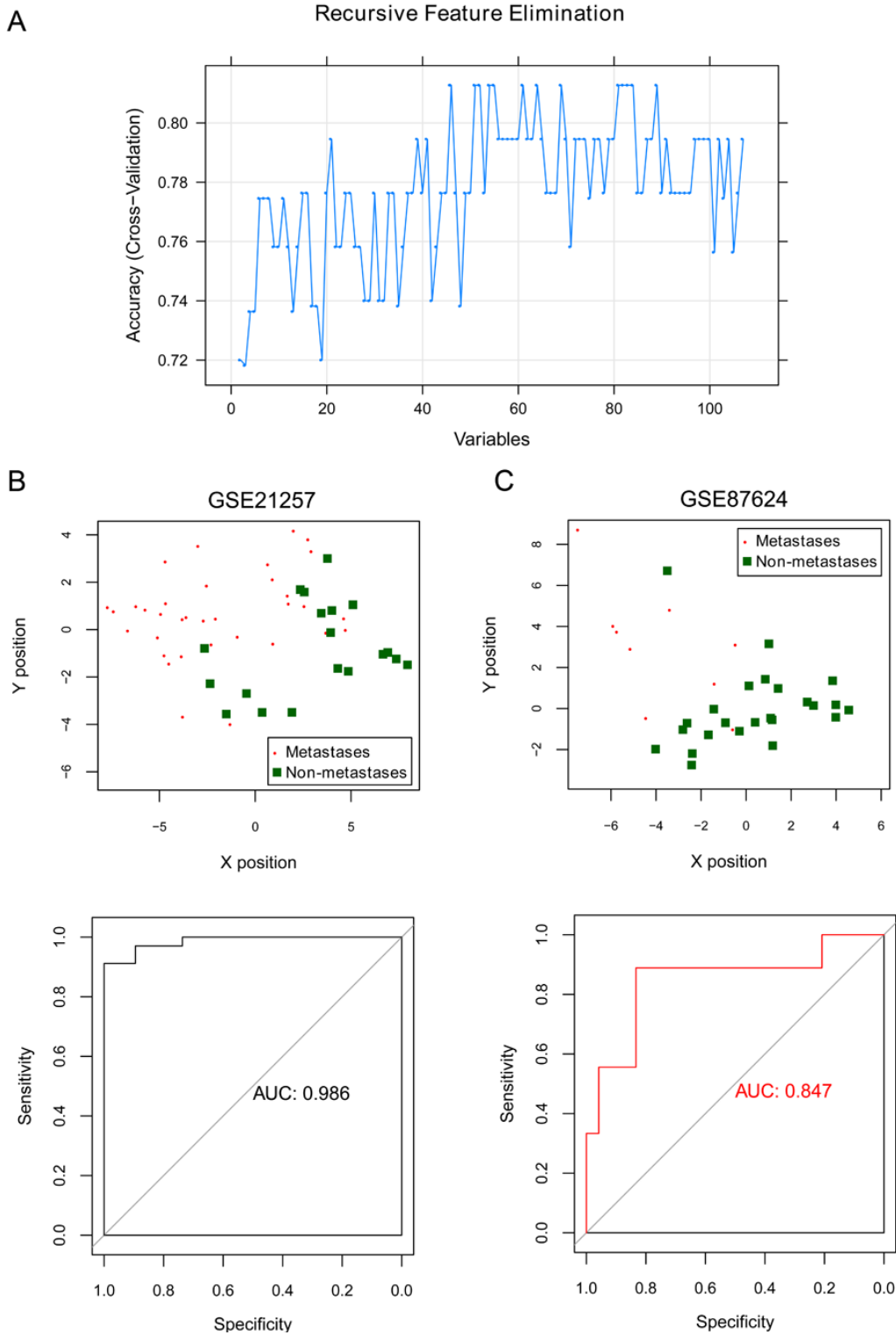


Figure 3 — Construction of a metastatic osteosarcoma discriminating classifier. A. Accuracy curve of the optimized gene signature using the RFE algorithm. B. Scatter plot (above) and ROC curve (below) based on the SVM classifier in the training set. C. Scatter plot (above) and ROC curve (below) based on the SVM classifier in the validation set.

patients was significantly longer in the low-risk group compared with the high-risk group ($P = 1.720 \times 10^{-3}$) in the training set. Meanwhile, the K-M survival analysis

also showed that the low- and high-risk sample prediction is closely associated with the patients' overall survival time in validation set 1 (GSE16091, $P = 1.393$

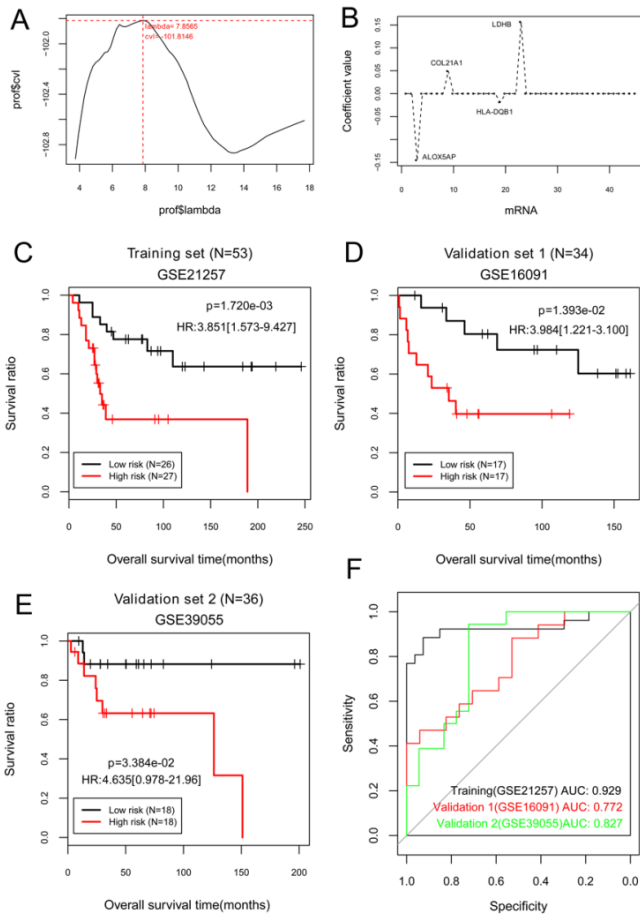


Figure 4 — Construction and verification of the prognostic prediction model. A. The lambda parametric curves by the cross-validation likelihood (cvl) algorithm. The horizontal axis and vertical axis represent lambda and cvl, respectively; the intersection of the red dotted line represents the maximum value of cvl at -101.8146 when lambda was 7.8565 . B. Coefficient distribution diagram of the four optimized DEGs related to prognosis based on the L1-penalized Cox-PH regression model. C. K-M survival analysis based on the prognostic prediction model in the training set (GSE21257); HR represents the hazard ratio. D. K-M survival analysis based on the prognostic prediction model in validation set 1 (GSE16091); HR represents the hazard ratio. E. K-M survival analysis based on the prognostic prediction model in validation set 2 (GSE39055); HR represents the hazard ratio. F. ROC curve analysis of the prognostic prediction model in the training set (GSE21257), validation set 1 (GSE16091), and validation set 2 (GSE39055).

Table I. — The four optimized differentially expressed mRNAs related to prognosis in metastatic osteosarcoma

Gene	Coef	P-value	Hazard Ratio	95%CI
ALOX5AP	-0.1453	0.03236	0.8023	0.518–0.942
COL21A1	0.0491	0.00846	1.4918	1.047–2.350
HLA-DQB1	-0.0188	0.03289	0.8972	0.578–0.933
LDHB	0.1567	0.01301	1.3655	1.192–2.044

$\times 10^{-2}$) and validation set 2 (GSE39055, $P = 3.384 \times 10^{-2}$) (Figure 4C–4E). Also, the ROC curve analysis revealed that the AUC of this model is 0.929, 0.772, and 0.827 in the training set, validation set 1, and validation set 2, respectively, indicating that this model has a satisfactory prediction ability both in the training set and validation set (Figure 4F).

DISCUSSION

In this study, 107 DEGs related to the prognosis of metastatic OS were identified based on the univariate Cox regression analysis of expression profile data from the NCBI GEO. Then, 46 DEGs were further

screened, and a metastatic-OS discriminating SMV classifier was constructed, which has high accuracy when distinguishing metastatic and non-metastatic OS samples. Furthermore, four optimized prognostic gene signatures (ALOX5AP, COL21A1, HLA-DQB1, and LDHB) were screened, based on these 46 DEGs, by Cox-PH analysis, and the prognostic prediction model for metastatic OS was constructed based on the prognostic regression coefficient of these four DEGs and possesses a relatively satisfactory prediction ability both in the training set and validation set.

The mining of a large amount of genetic data in various diseases has been enhanced by the rapid technological advances in high-throughput sequencing and bioinformatics²⁷. This study integrates the eligible expression profile data of OS from the NCBI GEO and screened 107 DEGs of metastatic and non-metastatic OS samples concerning prognosis. Then, the optimized 46 DEG signatures were used to establish a discriminating SVM classifier that could distinguish between metastatic and non-metastatic OS. SVM, as a powerful classification tool, is widely applied in cancer genomic subtyping or classification²⁸. Due to the classification feature of SVM and based on the

vast amounts of genomic data, SVM has been used to discover novel biomarkers or drug targets²⁸, as well as distinguish between disease subtypes²⁹.

Similarly, this study also constructed a metastatic-OS discriminating classifier using SMV with a high differentiation capability for metastatic and non-metastatic OS. Furthermore, the prognostic prediction model for metastatic OS was constructed using the Cox-PH analysis, a semi-parametric, proportional-hazards model. The availability of this model in survival analysis has been confirmed in recent studies^{30,31}. The prognostic prediction model in this study also showed a high predictive ability in the prognosis of metastatic OS.

In this study, the prognostic prediction model was constructed based on four optimized prognostic gene signatures, including ALOX5AP, COL21A1, HLA-DQB1, and LDHB. Arachidonate 5-lipoxygenase activating protein (ALOX5AP) is an important enzyme in leukotriene biosynthesis, which is involved in inflammatory responses^{32,33}. Currently, studies of ALOX5AP are focused on ischemic stroke and myocardial infarction, and ALOX5AP genetic polymorphisms have been demonstrated to increase the risk of myocardial infarction, ischemic stroke, and cardiovascular diseases³⁴⁻³⁶. Also, overexpression of ALOX5AP is found in patients with insulin resistance and obesity and can contribute to the inflammatory response³⁷. Furthermore, ALOX5AP is reported to be significantly related to the survival time of patients with esophageal squamous cell carcinoma³⁸. Collagen type XXI alpha 1 chain (COL21A1), as a component of the extracellular matrix, is a member of the fibril-associated collagens with interrupted helices (FACIT) collagen family and is expressed in many tissues, such as the placenta, skeletal muscle, kidney, stomach, and heart³⁹. Collagen XII, another member of the FACIT collagen family, has been demonstrated to have a potential role in bone metabolism⁴⁰. Major histocompatibility complex, class II, DQ beta 1 (HLA-DQB1) belongs to human leukocyte antigen (HLA) class II, which consists of a heterodimer consisting of an alpha (DQA) and a beta chain (DQB)⁴¹. HLA-DQB1 can peptides secreted from extracellular proteins, and exerts a central role in the immune system⁴¹. A previous study has shown that HLA II DQB1 polymorphisms are associated with the risk of cervical cancer⁴². These studies indicate that ALOX5AP, COL21A1, and HLA-DQB1 may be implicated in the inflammation and immune response; however, the influence of these genes on metastatic OS is still unclear. Lactate dehydrogenase B (LDHB) is considered as a key component in the oxidative pathway

and plays an important role in lysosomal activity and autophagy in cancer cells⁴³. Notably, the involvement of LDHB in metastasis has been demonstrated in tumor metastasis, including hepatocellular carcinoma⁴⁴ and prostate cancer⁴⁵. However, the function of LDHB in metastatic OS should be further investigated.

CONCLUSIONS

In conclusion, this study discusses the constructed a metastatic-OS discriminating classifier and prognostic prediction model for metastatic OS, which both have high accuracy. Four optimized prognostic gene signatures (ALOX5AP, COL21A1, HLA-DQB1, and LDHB) may have clinical implications in the prognosis of metastatic OS. However, the prognostic significance and mechanism of these four genes should be further confirmed in metastatic OS.

Consent statement: Not applicable.

Ethics approval: Not applicable.

REFERENCES

- Misaghi A, Goldin A, Awad M, Kulidjian AA. Osteosarcoma: a comprehensive review. *SICOT-J*. 2018;4:12.
- Isakoff MS, Bielack SS, Meltzer P, Gorlick R. Osteosarcoma: current treatment and a collaborative pathway to success. *J Clin Oncol*. 2015;33(27):3029.
- Meazza C, Scanagatta P. Metastatic osteosarcoma: a challenging multidisciplinary treatment. *Expert Rev Anticancer Ther*. 2016;16(5):543-56.
- Harrison DJ, Geller DS, Gill JD, Lewis VO, Gorlick R. Current and future therapeutic approaches for osteosarcoma. *Expert Rev Anticancer Ther*. 2018;18(1):39-50.
- Bao Q, Gong L, Wang J, Wen J, Shen Y, Zhang W. Extracellular vesicle RNA sequencing reveals dramatic transcriptomic alterations between metastatic and primary osteosarcoma in a liquid biopsy approach. *Ann Surg Oncol*. 2018;25(9):2642-51.
- Sayles LC, Breese MR, Koehne AL, Leung SG, Lee AG, Liu H-Y, et al. Genome-informed targeted therapy for osteosarcoma. *Cancer Discov*. 2019;9(1):46-63.
- Fuja DG, Rainusso NC, Shuck RL, Kurenbekova L, Donehower LA, Yustein JT. Transglutaminase-2 promotes metastatic and stem-like phenotypes in osteosarcoma. *Am J Cancer Res*. 2018;8(9):1752-63.
- Tang X-F, Cao Y, Peng D-B, Zhao G-S, Zeng Y, Gao Z-R, et al. Overexpression of Notch3 is associated with metastasis and poor prognosis in osteosarcoma patients. *Cancer Manag Res*. 2019;11:547-59.
- Liu T, Yan Z, Liu Y, Choy E, Hornicek FJ, Mankin H, et al. CRISPR-Cas9-Mediated Silencing of CD44 in Human Highly Metastatic Osteosarcoma Cells. *Cell Physiol Biochem*. 2018;20(1):1218-30.
- Buddingh EP, Kuijper ML, Duim RA, Bürger H, Agelopoulos K, Myklebost O, et al. Tumor-infiltrating macrophages are associated with metastasis suppression in high-grade osteosarcoma: a rationale for treatment with macrophage activating agents. *Clin Cancer Res*. 2011;17(8):2110-9.
- Scott MC, Temiz NA, Sarver AE, LaRue RS, Rathe SK, Varshney J, et al. Comparative transcriptome analysis quantifies

- immune cell transcript levels, metastatic progression, and survival in osteosarcoma. *Cancer Res.* 2018;78(2):326-37.
12. Kelly AD, Haibe-Kains B, Janeway KA, Hill KE, Howe E, Goldsmith J, et al. MicroRNA paraffin-based studies in osteosarcoma reveal reproducible independent prognostic profiles at 14q32. *Genome Med.* 2013;5(1):2.
 13. Paoloni M, Davis S, Lana S, Withrow S, Sangiorgi L, Picci P, et al. Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. *BMC Genomics.* 2009;10(1):625.
 14. Parrish RS, Spencer III HJ. Effect of normalization on significance testing for oligonucleotide microarrays. *J Biopharm Stat.* 2004;14(3):575-89.
 15. 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. *Nucleic Acids Res.* 2015;43(7):e47.
 16. Heng L, Jia Z, Bai J, Zhang K, Zhu Y, Ma J, et al. Molecular characterization of metastatic osteosarcoma: Differentially expressed genes, transcription factors and microRNAs. *Mol Med Rep.* 2017;15(5):2829-36.
 17. 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. *BMC Plant Biol.* 2014;14(1):169.
 18. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-57.
 19. Da WH, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37(1):1-13.
 20. Wang P, Wang Y, Hang B, Zou X, Mao JH. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. *Oncotarget.* 2016;7(34):55343-51.
 21. Deist TM, Dankers F, Valdes G, Wijsman R, Hsu IC, Oberije C, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers. *Med Phys.* 2018;45(7):3449-59.
 22. Lu X, Yang Y, Wu F, Gao M, Xu Y, Zhang Y, et al. Discriminative analysis of schizophrenia using support vector machine and recursive feature elimination on structural MRI images. *Medicine.* 2016;95(30):e3973.
 23. Wang Q, Liu X. Screening of feature genes in distinguishing different types of breast cancer using support vector machine. *Onco Targets Ther.* 2015;8:2311-7.
 24. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;17:77.
 25. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* 1997;16(4):385-95.
 26. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J.* 2010;52(1):70-84.
 27. Rodriguez-Ezpeleta N, Hackenberg M, Aransay AM. *Bioinformatics for high throughput sequencing*: Springer Science & Business Media; 2011.
 28. Huang S, Cai N, Pacheco PP, NARRANDES S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics.* 2018;15(1):41-51.
 29. Feres M, Louzoun Y, Haber S, Faveri M, Figueiredo LC, Levin L. Support vector machine-based differentiation between aggressive and chronic periodontitis using microbial profiles. *Int Dent J.* 2018;68(1):39-46.
 30. Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput Biol.* 2018;14(4):e1006076.
 31. Liang R, Wang M, Zheng G, Zhu H, Zhi Y, Sun Z. A comprehensive analysis of prognosis prediction models based on pathwaylevel, genelevel and clinical information for glioblastoma. *Int J Mol Med.* 2018;42(4):1837-46.
 32. White RK, Montgomery S. Leukotrienes: Inflammatory mediators—A review. *Or Surg Or Med Or Pa.* 1986;61(5):514-8.
 33. Helgadóttir A, Manolescu A, Thorleifsson G, Gretarsdóttir S, Jonsdóttir H, Thorsteinsdóttir U, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet.* 2004;36(3):233-9.
 34. Ye H, Zhang X, Chen Z, Li X, Zhang T, Yang C, et al. Association between the polymorphism (rs17222919,-1316T/G) of 5-lipoxygenase-activating protein gene (ALOX5AP) and the risk of stroke: A meta analysis. *Medicine.* 2018;97(41):e12682.
 35. Joob B, Wiwanitkit V. ALOX5AP genetic variants and risk of atherothrombotic stroke; a molecular structure analysis. *Journal of Ischemia and Tissue Repair.* 2019;3(1):e01.
 36. Linsel-Nitschke P, Götz A, Medack A, König IR, Bruse P, Lieb W, et al. Genetic variation in the arachidonate 5-lipoxygenase-activating protein (ALOX5AP) is associated with myocardial infarction in the German population. *Clin Sci (Lond).* 2008;115(10):309-15.
 37. Kaaman M, Rydén M, Axelsson T, Nordström E, Sicard A, Bouloumie A, et al. ALOX5AP expression, but not gene haplotypes, is associated with obesity and insulin resistance. *Int J Obes (Lond).* 2006;30(3):447-52.
 38. Wu B, Bai C, Du Z, Zou H, Wu J, Xie W, et al. The arachidonic acid metabolism protein-protein interaction network and its expression pattern in esophageal diseases. *Am J Transl Res.* 2018;10(3):907-24.
 39. Fitzgerald J, Bateman JF. A new FACIT of the collagen family: COL21A1. *FEBS Lett.* 2001;505(2):275-80.
 40. TANAKA Y, KAMADA A, IKEO T. Response of Collagen XII Splice Variants in Human Osteoblast-like Cells to IL-1 β . *Journal of Oral Tissue Engineering.* 2007;5(1):15-22.
 41. Megiorni F, Pizzuti A. HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. *J Biomed Sci.* 2012;19(1):88.
 42. Madeleine MM, Johnson LG, Smith AG, Hansen JA, Nisperos BB, Li S, et al. Comprehensive analysis of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 loci and squamous cell cervical cancer risk. *Cancer Res.* 2008;68(9):3532-9.
 43. Brisson L, Bański P, Sboarina M, Dethier C, Danhier P, Fontenille M-J, et al. Lactate dehydrogenase B controls lysosome activity and autophagy in cancer. *Cancer Cell.* 2016;30(3):418-31.
 44. Ding SJ, Li Y, Shao XX, Zhou H, Zeng R, Tang ZY, et al. Proteome analysis of hepatocellular carcinoma cell strains, MHCC97-H and MHCC97-L, with different metastasis potentials. *Proteomics.* 2004;4(4):982-94.
 45. Leiblich A, Cross S, Catto J, Phillips J, Leung H, Hamdy F, et al. Lactate dehydrogenase-B is silenced by promoter hypermethylation in human prostate cancer. *Oncogene.* 2006;25(20):2953-60