

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Original Article

A novel three-microRNA signature for predicting survival in patients with nasopharyngeal carcinoma

Shan-Qiang Zhang ^a, Jun Liu ^a, Hai-Bin Chen ^c, Wen-Jie Dai ^a,
Li-Qing Zhou ^a, Chong-Wei Xie ^a, Ji-Cheng Li ^{a,b*}

^a Medical Research Center, Yue Bei People's Hospital, Shantou University Medical College, Shaoguan, Guangdong, China

^b Institute of Cell Biology, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

^c Department of Histology and Embryology, Shantou University Medical College, Shantou, Guangdong, China

Received 18 July 2021; Final revision received 13 August 2021

Available online 8 September 2021

KEYWORDS

Nasopharyngeal carcinoma;
microRNA;
Survival;
Prognosis;
Bioinformatics analysis

Abstract *Background/purpose:* Nasopharyngeal carcinoma (NPC) is a malignant neoplasm of the head and neck. This study aims to use integrated bioinformatics technologies to develop a predictive miRNA-signature correlated with the prognosis of NPC.

Materials and methods: Initially, the differentially expressed miRNAs (DEMs) in NPC were identified, and then DEMs related to the prognosis of NPC were further screened. Subsequently, the relatively important DEMs identified by random forest algorithm were used to construct a predictive signature by multivariate COX regression analysis. Moreover, PCA, Kaplan–Meier analysis, time-dependent ROC analysis, and univariate and multivariate COX regression analysis were performed to evaluate the ability of the signature in risk identification and prognosis prediction in NPC.

Results: Hsa-miR-29c, hsa-miR-30e and hsa-miR-93 were selected from DEMs to construct a signature, and their abnormal expression was significantly associated with poor prognosis of NPC. The average AUC values of 1- to 5-year OS, DFS and DMFS predicted by the signature were all above 0.7, and showed better clinical independence than other indexes. In addition, 295 differentially expressed mRNAs could be used as potential target genes of the 3 DEMs. Among them, 56 differentially expressed mRNAs were related to PFS. GO and KEGG enrichment analysis indicated that the poor prognosis of NPC was related to the abnormality of chromosomes, cytokines, and chemokines.

* Corresponding author. Medical Research Center, Yue Bei People's Hospital, Shantou University Medical College, No. 133, Huimin South Road, Shaoguan, 512026, China.

E-mail address: lijichen@zju.edu.cn (J.-C. Li).

<https://doi.org/10.1016/j.jds.2021.08.017>

1991-7902/© 2021 Association for Dental Sciences of the Republic of China. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Conclusion: We constructed a three-miRNA signature with good independent performance in predicting the prognosis for NPC. This study may lay the foundation for exploring new therapeutic targets and improving survival outcomes in NPC patients.

© 2021 Association for Dental Sciences of the Republic of China. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Nasopharyngeal carcinoma (NPC), a type of head and neck malignant tumor derived from the columnar epithelium of nasopharyngeal mucosa, has caused substantial mortality and a huge economic burden in Southeast Asian countries over the years due to its insidious onset and highly invasive characteristics.¹ According to the International Agency for Research on Cancer (IARC), there were an estimated 133,354 new cases of NPC and 80,008 NPC-related deaths in 2020 worldwide, showing an increasing trend compared with 2018.^{2,3} As a country with the highest incidence rate of NPC in the world, in 2018, the corresponding new cases of NPC and NPC-related deaths in China were as high as 60,558 and 31,413 respectively, and most of which were confirmed to be locoregionally advanced (LA) NPC at the time of initial diagnosis.¹ One of the predominant reasons for the high mortality of NPC is tumor relapse and distant metastases, which is mainly related to the treatment failure of patients with LA-NPC.⁴ Therefore, it is important to strengthen the prognostic evaluation of NPC in order to guide the individualized treatment for NPC patients.

Currently, the tumor-node-metastasis (TNM) staging system is the main determinant for NPC treatment, but the disadvantages of this method include difficulties to predict the prognosis of NPC patients and to identify the biological heterogeneity of the tumor.⁵ Although the application of intensity-modulated radiation therapy has been shown to improve the overall survival rate in NPC patients with early stage disease, 20–30% of the patients eventually developed tumor relapse and distant metastasis.⁶ Furthermore, the prognosis in a considerable number of LA-NPC patients has not been improved due to the induction chemotherapy or adjuvant chemotherapy, suggesting that the differences in molecular biology or gene of NPC in different individuals may be the main reason for the differences in prognosis of NPC patients.⁷ Several studies have explored the influence of Epstein–Barr virus (EBV) DNA load and antibody titer changes on the prognosis of NPC, while others also have analyzed the prognostic evaluation capacity of tissue inhibitor metalloproteinase-2 (TIMP-2), L-selectin (SELL), Matrix metalloproteinase-3 (MMP-3), interleukin-8 (IL8), as well as cytokines such as cystic fibrosis transmembrane conduction regulator (CFTR) and platelet to lymphocyte ratio (PLR) in NPC.^{8–11} However, despite the encouraging answers given by these studies, the current evidence is still not enough to start their clinical trials. Therefore, it is important to explore new molecular markers related to the survival outcome of NPC.

As a class of short (≈ 20 – 22 nucleotides), non-coding RNA molecules, micro RNAs (miRNAs) have been proved to

play important roles in the occurrence, development and outcome of a tumor.¹² In recent years, with the rapid development of high-throughput technology and bioinformatics algorithm, investigators have found that miRNAs expression in tissue and bodily fluids of NPC patients can not only be used as potential biomarkers for early diagnosis and efficacy prediction, but can also participate in prognostic evaluation of the disease.^{5,13} For instance, Zhou et al.¹⁴ found that five differentially expressed miRNAs in NPC tissues were involved in distant metastasis. Liu et al.¹⁵ identified a 5-miRNA signature in large-scale tissue samples, which were associated with the prognosis of NPC patients with a high risk of progression. Lian et al.¹⁶ found that miRNA-183 and miRNA-141 were potential risk factors for poor prognosis in NPC patients. However, although the research on miRNAs as prognostic biomarkers of NPC is developing rapidly, the overlap of the results is very limited. So, it is necessary to verify and optimize the research data to provide a more accurate basis for its feasibility of clinical transformation.

Gene Expression Omnibus (GEO) database stores a large number of published high-throughput data, which makes it more convenient for investigators to study the role of genes in tumorigenesis. In this study, we constructed a prognostic signature composed of three miRNAs such as, hsa-mir-93, hsa-mir-29c and hsa-mir-30e, based on the GSE32960 dataset from the GEO database. Comprehensive analysis showed that the three-miRNA signature has a satisfactory independent predictive performance for the overall survival (OS), disease-free survival (DFS), and distant metastasis-free survival (DMFS) in NPC patients.

Materials and methods

Data acquisition and processing

All datasets used in this study were searched and downloaded from the GEO database of national center for biotechnology information (NCBI) (<https://www.ncbi.nlm.nih.gov/geo>), including miRNA expression dataset GSE32960, mRNA expression dataset GSE12452, GSE53819 and GSE102349. Among them, the GSE32960 dataset contained two cohorts of healthy and NPC tissue samples with a total of 330 cases (18 and 312, respectively). The number of normal nasopharyngeal tissues and NPC tissues in GSE12452 and GSE53819 were 10 and 31, and 18 and 18, respectively, while GSE102349 contained 113 NPC tissue samples alone. Meanwhile, the clinical data of GSE32960 dataset and GSE 102349 dataset were also downloaded. All microarray data were processed for homogenization, and

the data of GSE102349 was processed by fragments per kilobase of exon model per million mapped reads (FPKM) and \log_2 -transformed for further analyses.

Screening of differentially expressed miRNAs and mRNAs

The limma Bioconductor R package was used to screen differentially expressed miRNAs (DEMs) between NPC tissues and normal nasopharyngeal tissues. Considering the limited expression of miRNAs in tissues and analyzing DEMs as much as possible, the P-value [false discovery rate (FDR)] < 0.05 and $|\text{fold change (FC)}| > 1.5$ were adjusted as the cut-off criteria to identify DEMs in GSE32960 dataset. Similarly, the P-value (FDR) < 0.05 and $|\text{FC}| > 2$ was adjusted as the cut-off criteria to identify differentially expressed mRNAs in GSE12452 and GSE53819 datasets.

Functional enrichment analysis

FunRich (Version 3.1.3), an online tool for functional enrichment and interaction network analysis, was used to identify the pathways involved in the DEMs and the transcription factors that regulate their expression. In addition, the “clusterProfiler”, “enrichplot” and “org.Hs.eg.db” packages were adopted to conduct gene ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the potential target genes of selected hub DEMs.

Construction and evaluation of three-miRNA signature

In order to better screen and identify the DEMs related to the prognosis of NPC, 312 patients in GSE32960 dataset were randomly divided into the training set and the testing set by R package “caret”, and the whole set was used to represent the integration of the two sets. Then, the univariate Cox regression analysis was used to identify the DEMs related to OS of patients in the training set, and the random forest algorithm was used to reduce the dimension of the survival data to complete the importance ranking of DEMs. According to relative importance > 0.1 , three hub DEMs (hsa-miR-29c, hsa-miR-30e, and hsa-miR-93) were selected for multivariate Cox regression analysis to construct the three-miRNA prognostic signature. The index of signature was calculated as follows:

Prognostic index (Pi) = $(\beta * \text{expression level of hsa-miR-29c}) + (\beta * \text{expression level of hsa-miR-30e}) + (\beta * \text{expression level of hsa-miR-93})$.

Subsequently, Kaplan–Meier survival curve and time-dependent ROC curve analysis were used to evaluate and verify the predictive capacity of the three-miRNA signature for risk identification, OS, DFS and DMFS of patients in the training set, testing set and whole set, respectively. Univariate and multivariate Cox regression analyses were used to demonstrate whether it can be used as a prognostic indicator for NPC independent of other clinical indexes.

Target gene analysis

Since miRNA can target mRNA and regulate gene expression at the post-transcriptional level, and the imbalance of miRNA-mRNA regulatory network is considered to be significantly associated with cancer treatment and prognosis, TargetScan (an online miRNA binding prediction database) was used to predict the target genes of 3 miRNAs. Furthermore, the predicted results were intersected with the differentially expressed mRNAs in GSE12452 and GSE53819 datasets to obtain the potential target genes of 3 miRNAs, and the overlapping results were displayed by drawing the Venn diagram on the bioinformatics & Evolutionary Genomics website (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Finally, the genes associated with NPC progression free survival (PFS) were screened in GSE102349 dataset by the univariate Cox regression analysis.

Statistical analysis

R software (version 3.6.3) was used for all the statistical analyses. The R statistical packages “survival” and “survminer” were used to plot the Kaplan–Meier survival curve with the median used as a cut-off value, while the package “survivalROC” was used to generate the receiver operating characteristic (ROC) curve of the three-miRNA signature.

Results

Identification and functional enrichment analysis of DEMs in NPC

The roadmap of the present study is demonstrated in Fig. 1. The clinical characteristics of 312 NPC patients in GSE32960 dataset are listed in Table 1. According to the above screening criteria ($|\text{FC}| > 1.5$ and $\text{FDR} < 0.05$), a total of 185 miRNAs in NPC tissues reached statistical significance compared with the normal tissues. Among them, 94 significantly up-regulated miRNAs and 91 significantly down-regulated miRNAs are illustrated in the volcanic map (Fig. 2a), while the heat map of these DEMs is shown in Fig. 2b. The enrichment analysis results by FunRich showed that these DEMs were mainly involved in ErbB receptor signaling network, sphingosine 1-phosphate pathway, glypican pathway, plasma Membrane Estrogen Receptor, proteoglycan syndecan-mediated signaling, vascular endothelial growth factor (VEGF) and VEGF receptor signaling network, and other biological pathways, as shown in Fig. 2c. In addition, the top 10 upstream transcription factors predicted by FunRich software, namely SP1, EGR1, SP4, POU2F1, RREB1, NFIC, SOX1, TFAP4, E2F1 and RORA, may have regulatory effects on these 185 DEMs, as shown in Fig. 2d.

Construction of three-miRNA signature and risk identification of NPC patients

The results of univariate Cox regression analysis in the training set revealed that 17 DEMs were associated with OS

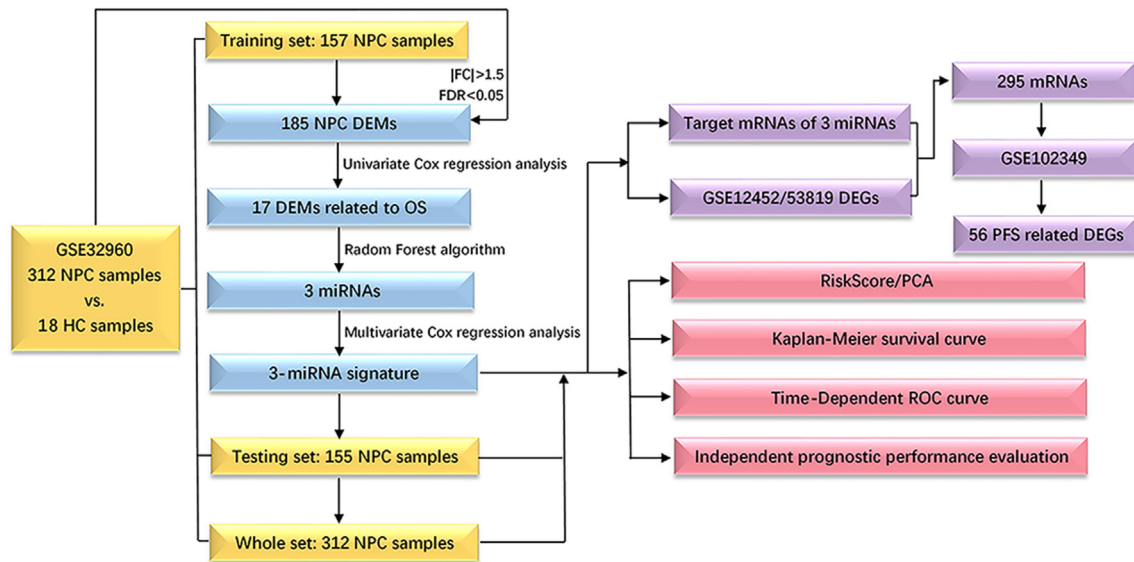


Figure 1 Roadmap of the present study.

Table 1 Baseline information of patients in GSE32960.

Clinical characteristics		Total	Percentage (%)
GSE32960 in GEO ^a		312	100
Survival status	Survival	246	78.85
	Death	66	21.15
Age	≤65 years	291	93.27
	>65 years	21	6.73
Gender	Female	79	25.32
	Male	233	74.68
Stage	I	12	3.85
	II	86	27.56
	III	91	29.17
	IV	123	39.42
RT ^b period interruptions	0 days	178	57.05
	≥1 days	134	42.95
RT boosting	Yes	149	47.76
	No	163	52.24
CCRT ^c	Yes	268	85.90
	No	44	14.10

^a GEO: Gene Expression Omnibus.

^b RT: Radiotherapy.

^c CCRT: Concurrent chemoradiotherapy.

in NPC patients (Fig. S1). By using the random forest algorithm to evaluate the importance of these DEMs, three miRNAs with relative importance >0.1, such as hsa-miR-29c, hsa-miR-30e and hsa-miR-93, were selected for future analysis. The relationship between the error rate and the number of classification trees and the out of bag importance of the 3 miRNAs is shown in Fig. 3a. Multivariate Cox regression analysis was then performed to construct a three-miRNA signature, the hazard ratios (HRs) and confidence interval of 3 miRNAs were shown in Fig. 3b. The signature risk score was calculated as follows: Risk Score = 0.843 * exp(hsa-miR-93) - 0.543 * exp(hsa-miR-29c) - 0.749 * exp(hsa-miR-30e). Next, the three-miRNA signature

was used to calculate the risk score of each patient in the training set, and the patients were divided into the high-risk group and the low-risk group according to the median value. As shown in Fig. 3c and Fig. S2a, the results showed that there were significant differences in risk score, survival time, expression of three miRNAs and principal component analysis (PCA) between the two groups, indicating that the three-miRNA signature constructed in this study has a good power to identify the prognostic risk of NPC. As expected, similar results were also observed in the testing set and the whole set, which further increased the credibility of analysis results in the present study (Fig. 3d, e and Fig. S2b, c).

Evaluation of the predictive performance for OS and independent prognostic value of three-miRNA signature in NPC

To evaluate the predictive performance of three-miRNA signature for OS in NPC patients, Kaplan–Meier survival curves were generated in the training set, the testing set, and the whole set, respectively. The results revealed that the high-risk patients identified by three-miRNA signature had worse survival probability ($P < 0.001$) (Fig. 4a, b, c). Subsequently, the time-dependent ROC curve analysis was performed to further evaluate the predictive capacity of three-miRNA signature in each group. As shown in Fig. 4d, e and f, the AUC values of the signature for predicting 1- to 5-year OS of NPC patients were 0.803, 0.806, 0.793, 0.806 and 0.806 in the training set, 0.981, 0.685, 0.699, 0.685 and 0.685 in the testing set, and 0.838, 0.750, 0.750, 0.752 and 0.756 in the whole set, respectively. These results indicated that the three-miRNA signature constructed in this study is accurate and stable in predicting OS of NPC patients. Furthermore, the results of univariate and multivariate Cox regression analyses showed that whether in the training set, testing set, or whole set, three-miRNA signature was superior to any clinical index such as gender, age,

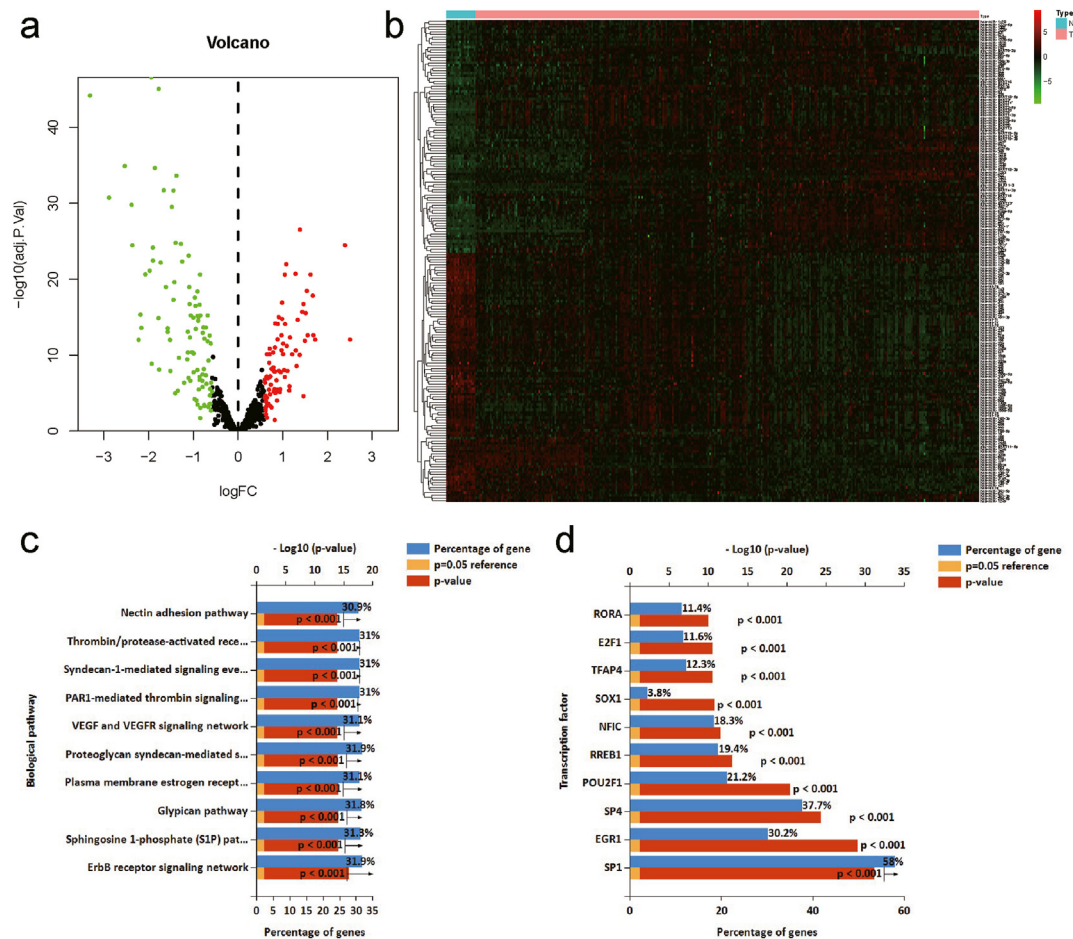


Figure 2 Identification of DEGs and enrichment analysis. (a) The volcano of DEGs. (b) The heatmap of DEGs. (c) Biological pathway analysis for DEGs using FunRich software. (d) Transcription factor analysis for DEGs using FunRich software.

and TNM stage in predicting prognosis, suggesting that it is suitable as an independent prognostic indicator for NPC patients (Fig. S3a, b and c). In addition, these results also implied that hsa-miR-93, hsa-miR-29c, and hsa-miR-30e may play important roles in NPC progression and can be used as potential targets for NPC treatment.

Evaluation of DFS and DMFS prediction performance of three-miRNA signature in NPC

In order to comprehensively evaluate the prognostic value of three-miRNA signature in NPC, the Kaplan–Meier survival curves of DFS and DMFS of NPC patients in GSE32960 dataset were further generated. The results revealed that the high-risk patients differentiated by the three-miRNA signature constructed in this study had worse survival probability ($P < 0.0001$) (Fig. 5a and b). Furthermore, time-dependent ROC curve analysis was performed to evaluate the predictive capacity of three-miRNA signature for DFS and DMFS in NPC patients. The results indicated that the AUC values of the signature for predicting 1- to 5-year DFS and DMFS of NPC patients were 0.718, 0.723, 0.702, 0.723 and 0.723, and 0.753, 0.774, 0.749, 0.774 and 0.774, respectively (Fig. 5c and d). Moreover, the results of univariate and multivariate Cox regression analyses also demonstrated

that the three-miRNA signature had better independent prognostic performance in predicting NPC DFS and DMFS than gender, age, TNM stage, radiotherapy interrupt, radiotherapy boosting and concurrent chemoradiotherapy (Fig. 6a and b). These results further corroborated the robustness of prognostic value of the three-miRNA signature for NPC patients, which is worthy of further study.

Advantages of the three-miRNA signature

Currently, gender, age and TNM stage are the main references to evaluate the prognosis of NPC patients. Therefore, in order to further confirm the advantages of the three-miRNA signature constructed in this study in predicting prognosis, time-dependent ROC curves of the comparison between three-miRNA signature, gender, age and TNM stage in predicting 1-year, 3-year and 5-year OS of NPC patients in the GSE32960 dataset were conducted. As shown in Fig. 7a, b and c, the AUC values of three-miRNA signature in predicting 1-year, 3-year, and 5-year OS of NPC patients were 0.838, 0.750, and 0.745, respectively. Compared with the clinical indexes mentioned above, the three-miRNA signature showed better predictive power. In addition, Kaplan–Meier survival curve analyses of the 3 miRNAs revealed that the patients with high has-miR-93

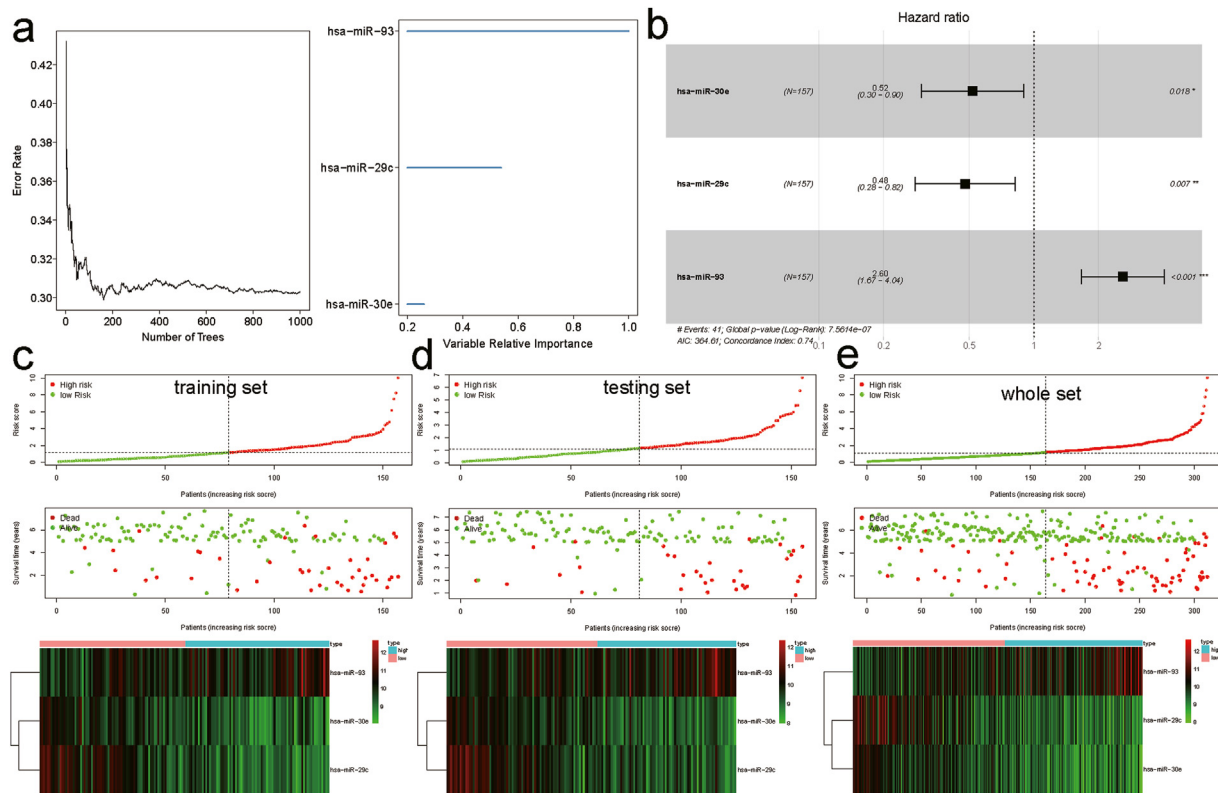


Figure 3 Construction and verification of the signature. (a) Decision tree constructed using random forest algorithm and screening of relative importance genes. (b) Three-miRNA signature was built by step multivariate Cox regression. Distribution of Risk score, survival time and miRNAs expression between the high-risk and low-risk groups in training set (c), testing set (d), and whole set (e).

expression had poor prognosis (OS, $P < 0.001$), and low expressions of has-miR-30e and has-miR-29c were associated with poor prognosis (OS, $P < 0.001$ and $P = 0.002$) (Fig. 7d, e, f). These results demonstrated that the three-miRNA signature is a better potential prognostic biomarker for NPC compared with the commonly used clinical indexes, and can be used to guide the individualized treatment of NPC.

Analysis of three-miRNA target genes

Using TargetScan, a total of 9642 target genes for the 3 miRNAs were predicted. Meanwhile, 1322 and 2439 differentially expressed mRNAs were screened in GSE12452 and GSE53819 datasets, respectively. The intersection of overlapping genes between the three sets of data is shown in the Venn diagram, which indicated that 295 differentially expressed mRNAs could be the potential target genes of the 3 miRNAs (Fig. 8a). The results of GO enrichment analysis revealed that the 295 differentially expressed mRNAs were mainly involved in biological processes such as, nuclear chromosome segregation, and endodermal cell differentiation. In terms of cellular components, these potential target genes were mainly enriched the apical part of the cell, endoplasmic reticulum lumen, and apical plasma membrane, while their molecular function were mainly associated with the extracellular matrix structural

constituent, cytokine activity, chemokine receptor binding, and extracellular matrix structural constituent conferring tensile strength (Fig. 8b). Furthermore, KEGG analysis indicated that the potential target genes of 3 miRNAs were significantly enriched in the cytokine–cytokine receptor interaction, chemokine signaling pathway, PI3K-Akt signaling pathway, and other pathways associated with tumorigenesis (Fig. 8c). Finally, the univariate Cox regression analysis revealed that there were 56 differentially expressed genes in GSE102349 dataset, which were related to progression-free survival (PFS), and the top 40 genes were illustrated in Fig. 8d.

Discussion

Currently, miRNAs have been demonstrated to play an important role in the radioresistance and chemoresistance of NPC, and can affect the survival outcome of patients.¹⁴ Several studies have explored the potential role of miRNAs as prognostic biomarkers for NPC based on different human biological samples.^{15–17} In spite of the emerging biological importance of miRNAs, their characteristics such as small molecular weight, low expression and high sequence homology, as well as the expensive, time-consuming and complex operation of microarray and quantitative real time polymerase chain reaction (qRT-PCR) can restrict the further clinical transformation research of

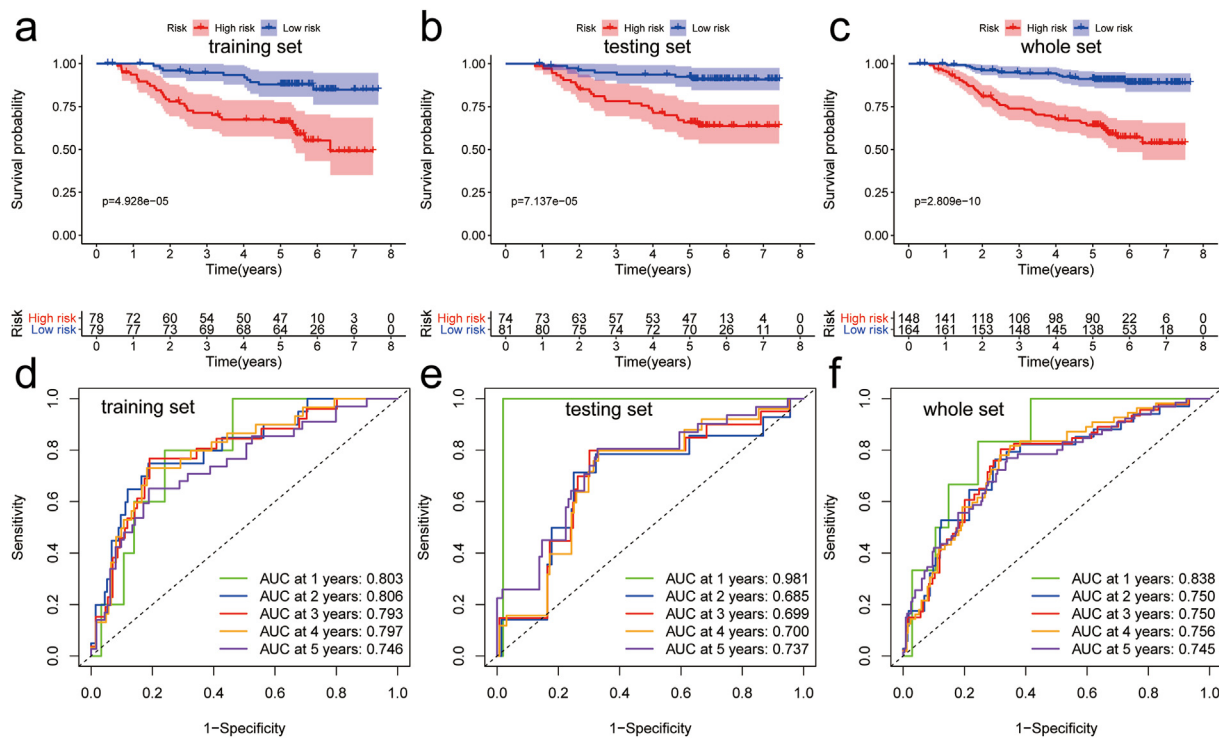


Figure 4 Kaplan–Meier plotter and ROC analysis and Independent prognostic factors determined. Overall survival analysis of the three-miRNA signature in training set (a), testing set (b), and whole set (c). The ROC analysis of the three-miRNA signature for 1-, 2-, 3-, 4-, and 5-years in training set (d), testing set (e), and whole set (f).

miRNAs.^{18,19} Therefore, integrative bioinformatics analysis leveraging a number of public datasets has become one of the effective methods to excavate miRNAs with better clinical application potential.²⁰

In the present study, miRNA expression data from GSE32960 dataset were adopted for differential expression analysis, and 185 DEMs were identified. The miRNA-enrichment analysis showed that these DEMs were involved in various pathophysiological pathways such as tumor proliferation, invasion, metastasis and angiogenesis, indicating that they can play an indispensable role in the prognosis of NPC. Considering that transcription factors and miRNAs are crucial trans-regulatory factors that can play important roles in controlling gene regulation at the transcriptional and post-transcriptional level, the transcription factors that regulate these DEMs were also predicted. As a sequence-specific DNA binding protein, transcription factor specificity protein-1 (SP1) has been proved to be involved in regulating the expression of genes that play a key role in the proliferation and metastasis of a variety of tumor cells, and the high expression level of SP1 protein is often considered as a negative prognostic factor in tumor patients.²¹ A recent study showed that exogenous SP1 could inhibit the expression of miR7-5p and could reverse the growth of NPC cells inhibited by solamargine.²² Ho et al.²³ found that melatonin can regulate the expression of matrix metalloprotein 9 (MMP-9) gene induced by tumor promoter 12-o-tetradecanoylphosphate-3-acetate (TPA) by inhibiting the binding ability of SP1 to DNA, thus inhibiting the metastasis activity of NPC. Our prediction showed that SP1 may regulate nearly 58% of DEMs, which is helpful to

further clarify the invasion and metastasis mechanism of NPC, and more experiments are needed to verify it in the future.

For clinicians, the current difficulty in NPC treatment is that some patients often do not respond to radiotherapy or chemotherapy, resulting in tumor progression or distant metastasis due to not being treated in time, leading to poor prognosis.^{7,24} Therefore, we constructed a three-miRNA signature to evaluate the prognosis of patients with NPC. Different from other studies, we employed random forest algorithm to select hsa-miR-93, hsa-miR-29c and hsa-miR-30e in the training set to form the prognostic signature, and confirmed the accuracy and stability of the signature in the testing set and the whole set. It is known that hsa-miR-93 is overexpressed in many types of cancer, and can target transforming growth factor beta-receptor type II (TGFβR2) and disabled-2 (DAB2) through a variety of signaling pathways to promote the growth and metastasis of NPC cells.^{25,26} Recent studies have shown that upregulated miR-30e-5p can target metastasis-associated protein 1 (MTA1) and ubiquitin-specific protease 22 (USP22), respectively, to reduce the ability of 5–8F cells migration, invasion and metastasis.^{27,28} In addition, hsa-miR-29c can directly target T lymphoma invasion and metastasis inducing factor 1 (TIAM1), inhibit the migration and invasion of NPC cells, and downregulate Integrin Beta 1 (ITGB1) to enhance the sensitivity of NPC cells to paclitaxel.^{29,30} In this study, we found high expression of hsa-miR-93 and low expression of has-miR-30e and hsa-miR-29c in NPC tissues through data analysis, and Kaplan–Meier curves indicated that patients with this expression pattern tend to have shorter OS. These

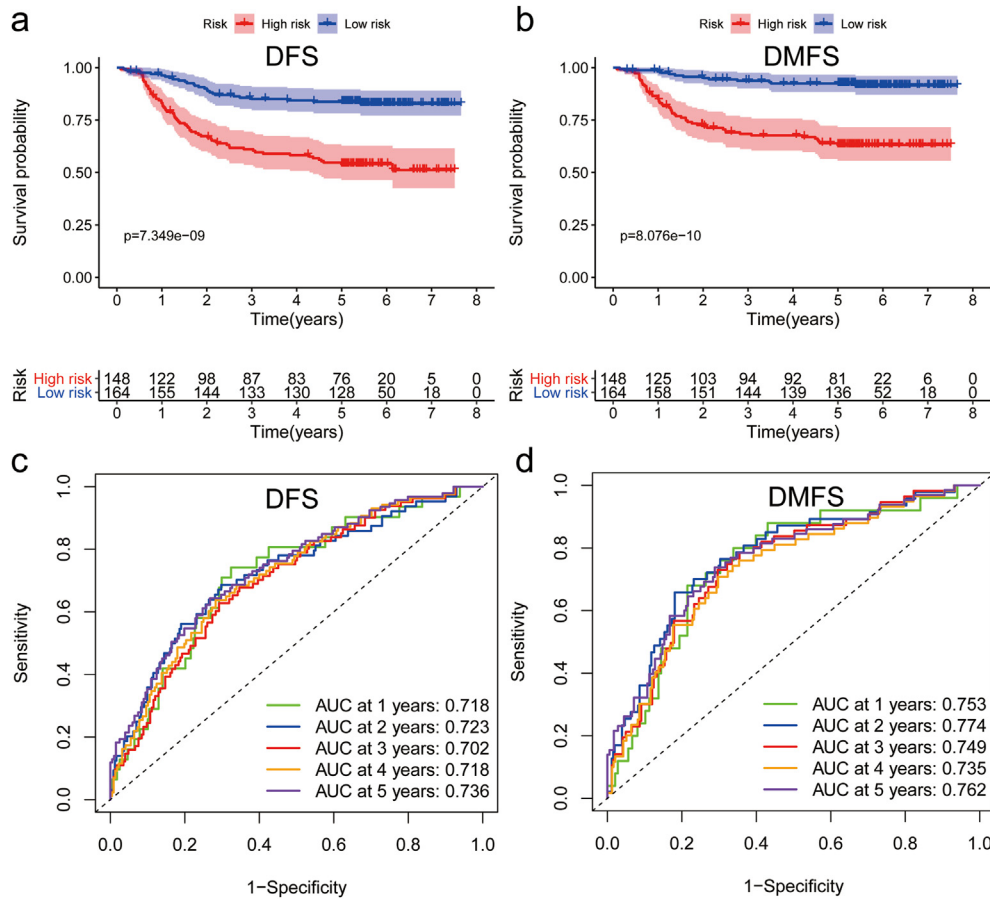


Figure 5 Kaplan–Meier plotter and ROC analysis. Disease-free survival (DFS) (a) and distant-metastasis-free survival (DMFS) (b) analysis of the three-miRNA signature in whole set. ROC curve for the prediction of DFS (c) and DMFS (d) in patients with NPC.

results are consistent with the findings of other studies mentioned above, which support the feasibility of three-miRNA signature as the prognostic risk factor for patients with NPC. In addition to the three miRNAs used to construct signature, we also found that hsa-miR-29a, hsa-miR-29b and hsa-miR-30d were significantly correlated with the survival of NPC patients. It was found that the down-regulation of hsa-miR-29a and hsa-miR-29b was related to the pathogenesis of NPC, and could enhance the migration and invasiveness of NPC cells by regulating the expression of secreted protein acidic and rich in cysteine (SPARC) and Collagen type III alpha 1 chain (COL3A1).³¹ In another study, Gao et al.³² found that up-regulating hsa-miR-29a may also reduce the resistance of nasopharyngeal carcinoma CNE-1 cells to paclitaxel by inhibiting the expression of signal transducer and activator of transcription 3 (STAT3) and B-cell leukemia/lymphoma 2 (Bcl-2). One team reported the power of signature containing hsa-miR-29b to predict the prognosis of NPC patients, but the AUC values of the risk score in DFS and DMFS were only 0.69.³³ In addition, hsa-miR-30d was found to be low expressed in a variety of tumor tissues, and its expression level is helpful to distinguish different stages of non-small cell lung cancer and predict the results of sorafenib in the treatment of hepatocellular carcinoma.^{34,35} However, there are few reports on hsa-miR-30d in NPC-related studies, so its prognostic significance for NPC needs to be clarified in future studies.

By analyzing the risk score, survival time, the expression of three miRNAs, and other data of patients, we found that the signature composed of hsa-miR-93, hsa-miR-29c and hsa-miR-30e could classify patients into the high-risk group and the low-risk group in GSE32960 dataset, and the Kaplan–Meier analysis showed that these identified high-risk patients have worse survival probability, which indicated that the three-miRNA signature constructed in this study has a better prognosis risk identification ability for NPC patients. Time-dependent ROC analysis showed that AUC values of three-miRNA signature in predicting 5-year OS of patients within the training set, testing set and whole set were 0.806, 0.685 and 0.756, respectively, which were better than the three-miRNA signature constructed by Wang et al.,³⁶ 4-miRNA signature constructed by Zhang et al.,³³ 4-miRNA signature constructed by Zhao et al.³⁷ and 4-miRNA signature constructed by Liu et al.³⁸ Furthermore, compared with the published signatures mentioned above, the AUC values of the three-miRNA signature constructed in this study for predicting 1- to 5-year DFS and DMFS of NPC patients were all above 0.7, and showed better power in predicting OS than gender, age and TNM stage, indicating that the three-miRNA signature constructed in this study has the potential to be an independent prognostic factor for NPC patients.

In the following analysis, the results of target gene prediction showed that 295 differentially expressed mRNAs

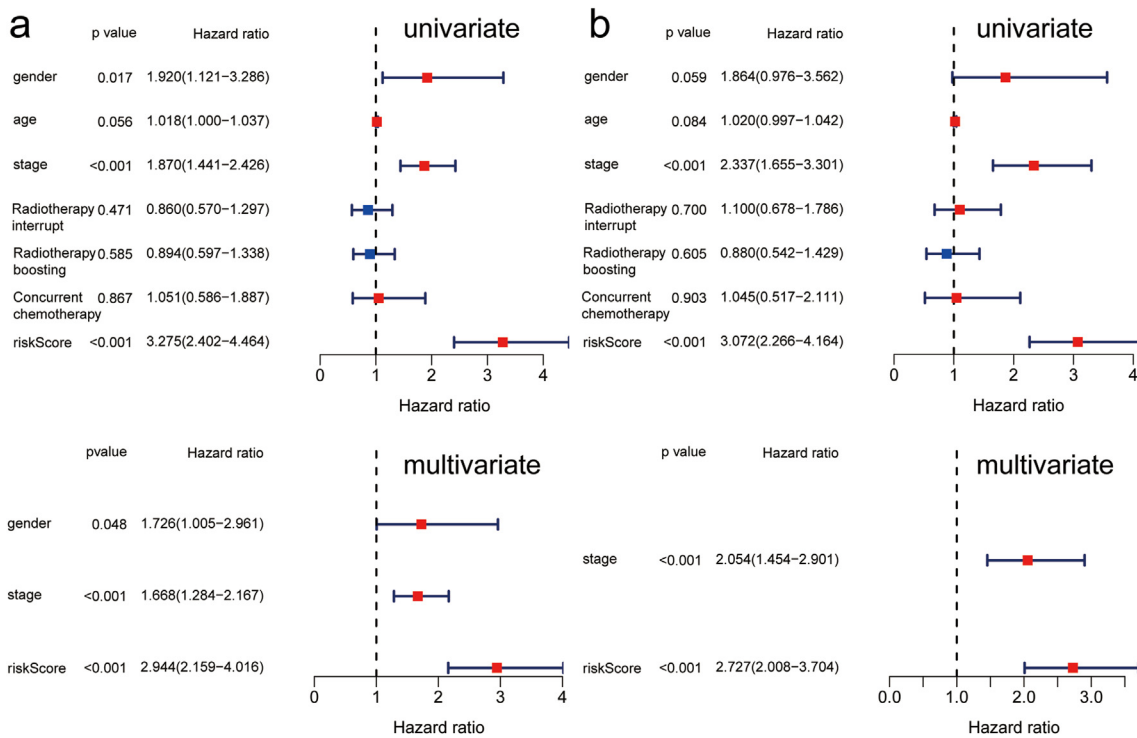


Figure 6 Independent prognostic prediction analysis. The three-miRNA signature was an independent prognostic factor for Disease-free survival (a) and distant-metastasis-free survival (b) by Univariate and multivariate Cox regression analyses.

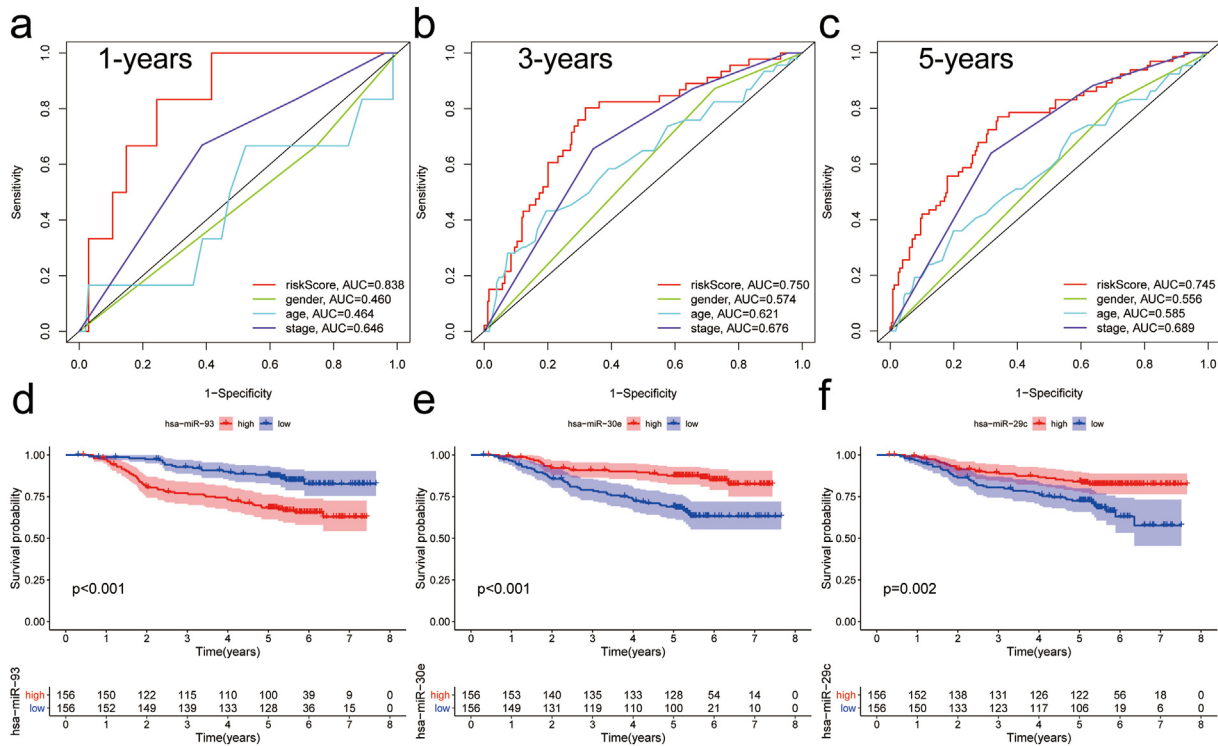


Figure 7 ROC analysis of clinical feature and risk score, and Kaplan–Meier survival curve for three miRNAs. ROC analysis of risk score, gender, age, and TNM stage for 1- (a), 3- (b), and 5-years (c) overall survival in the whole set. The overall survival analysis of has-miR-93 (d), has-miR-30e (e), and has-miR-29c (f) in the whole set.

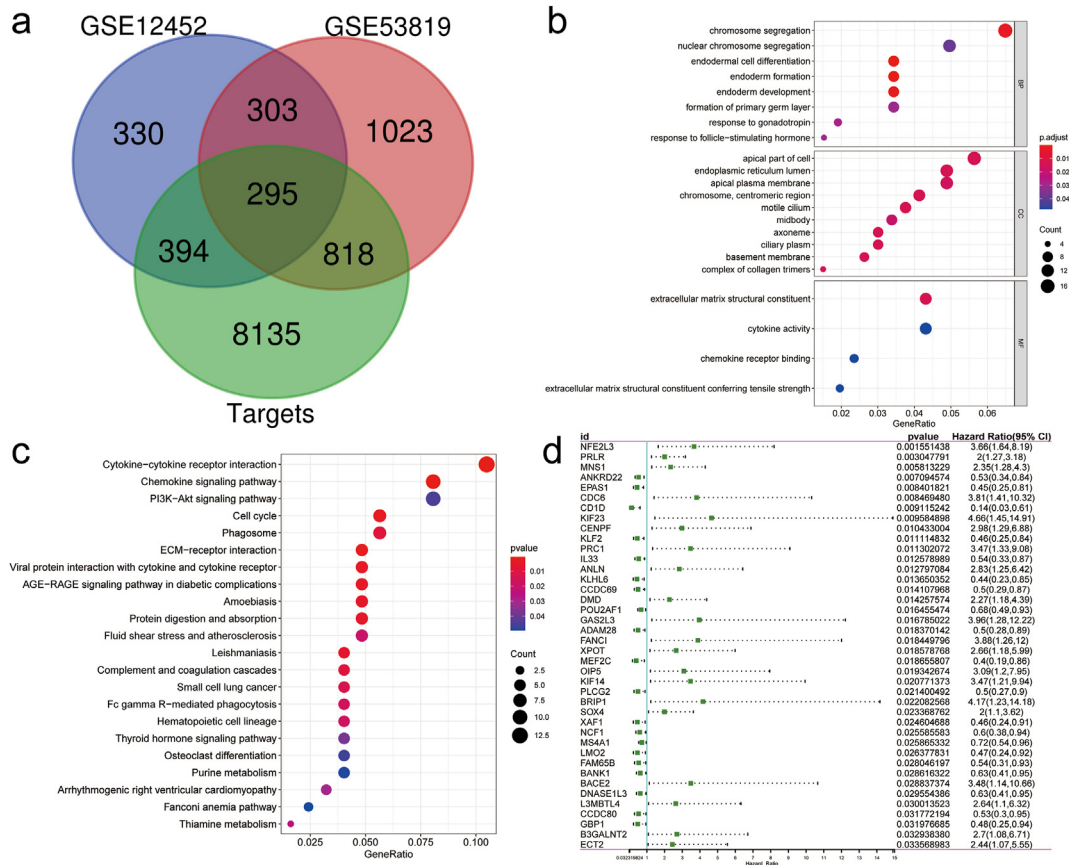


Figure 8 Three miRNA targets prediction and enrichment analysis. (a) Venn diagram of differential genes and target genes. (b) GO enrichment analysis. (c) KEGG enrichment analysis. (d) Univariate Cox regression analysis on progression-free survival of differential target genes in GSE102349 cohort.

may be the potential target genes for the three miRNAs, indicating that the dysregulation of these miRNA-mRNA pairs may predict the poor survival outcome of NPC patients. However, since mRNA can also be targeted by a variety of miRNAs, their complete interaction network remains to be further revealed.³⁹ In the subsequent GO enrichment analysis, the results showed that the biological process and cellular components of these target genes were related to the nuclear chromosome segregation and centromeric region of the chromosome, respectively, suggesting that chromosomal abnormalities are important factors for poor prognosis of NPC. At present, studies have shown that chromosomal aberration, loss of heterozygosity, chromosome rearrangements and chromosomal imbalances can play important role in the occurrence and development of NPC.⁴⁰ We speculated that during NPC resistance to radiotherapy or chemotherapy, these post-transcriptional target genes were regulated by the 3 miRNAs, resulting in chromosomal rearrangements and causing distant metastasis or recurrence of the tumor. KEGG analysis showed that the pathways involved in these target genes, such as PI3K-Akt signaling pathway and viral protein interaction with cytokine and cytokine receptors, were related to virus infection. This may be that EBV activated these signaling pathways, which leads to the dysregulation of the hub DEMs, promoted the proliferation of tumor cells, and finally worsens the disease.⁴¹ In addition, some other signal

pathways related to tumor invasion and metastasis, such as small cell lung cancer and osteoclast differentiation, further confirmed our hypothesis.^{25,42} In further analysis, the molecular function annotation and KEGG analysis showed that these target genes were mainly enriched in cytokine activity, chemokine receptor binding, and cytokine-cytokine receptor interaction and cytokine signaling pathway, which may be due to the reason that these potential target genes promoted the secretion of cytokines and chemokines by NPC cells under the regulation of the 3 miRNAs, or interacted with stromal cells by releasing tumor exosomes, resulting in immune escape and enhanced tumor metastasis.⁴³ On the other hand, the three-miRNA signature constructed in this study may be adopted as a target for NPC immunotherapy. Of course, further research is needed to clarify the mechanism of their interaction with EBV and tumor microenvironment.⁴⁴

Although the above analysis has fully demonstrated that our constructed three-miRNA signature may be a good predictive tool for the survival outcome of patients with NPC, this study still has the following limitations. Initially, because it is difficult to obtain large-scale tissue microarray datasets with complete clinical information, we only adopted the GSE32960 dataset to screen and identify DEMs, and more external validations of the predictive performance of the developed signature is needed as the next step. In addition, because most of the histopathological

subtypes of the tissue samples from datasets are undifferentiated non-keratinized carcinoma, the prognostic value of the signature constructed in this study in other subtypes remains to be explored. Moreover, the specific mechanism of the 3 miRNAs in NPC needs to be further studied by molecular experiments, and the related lncRNA-miRNA-mRNA competitive endogenous RNA (CeRNA) networks need to be established. Eventually, the performance of three-miRNA signature constructed in this study needs to be further modified and optimized in combination with clinical research. Although this signature is expected to become an auxiliary tool for post-treatment follow-up monitoring and prognosis evaluation of NPC patients to better improve the prognosis. However, because tissue biopsy is invasive and it is difficult to obtain samples after treatment, it is worth further research to detect the expression of these miRNAs based on serological detection or explore more promising non-invasive methods, such as nasopharyngeal swab collection.^{45–47}

Taken together, in this study, three DEMs (hsa-miR-93, hsa-miR-29c and hsa-miR-30e) were screened from NPC tissue microarray dataset by using a novel random forest algorithm. The results of transcription factor prediction, prognosis evaluation and target gene prediction demonstrated that the signature, which is composed of the three DEMs, has the potential to be used as an independent prognostic factor for NPC. The results of this study may lay the foundation for further elucidating the underlying mechanism of miRNAs influencing the prognosis of NPC and exploring new therapeutic targets of NPC to improve the survival outcome of patients.

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgements

We acknowledge GEO database for providing their platforms and contributors for uploading their meaningful datasets. This research was supported by the Shaoguan Science and Technology Plan Projects in 2020 (200812094530421).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jds.2021.08.017>.

References

- Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet* 2019;394:64–80.
- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA - Cancer J Clin* 2021;71:209–49.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA - Cancer J Clin* 2018;68:394–424.
- Zhang Y, Chen L, Hu GQ, et al. Gemcitabine and cisplatin induction chemotherapy in nasopharyngeal carcinoma. *N Engl J Med* 2019;381:1124–35.
- Wen W, Mai SJ, Lin HX, et al. Identification of two microRNA signatures in whole blood as novel biomarkers for diagnosis of nasopharyngeal carcinoma. *J Transl Med* 2019;17:186.
- Jin YB, Zhang GY, Lin KR, et al. Changes of plasma cytokines and chemokines expression level in nasopharyngeal carcinoma patients after treatment with definitive intensity-modulated radiotherapy (IMRT). *PLoS One* 2017;2:e0172264.
- Liang Y, Li J, Li Q, et al. Plasma protein-based signature predicts distant metastasis and induction chemotherapy benefit in Nasopharyngeal Carcinoma. *Theranostics* 2020;10:9767–78.
- Lee HM, Okuda KS, González FE, Patel V. Current perspectives on nasopharyngeal carcinoma. *Adv Exp Med Biol* 2019;1164:11–34.
- Gong D, Li Z, Ding R, et al. Extensive serum biomarker analysis in patients with nasopharyngeal carcinoma. *Cytokine* 2019;118:107–14.
- Tu Z, Chen Q, Zhang JT, Jiang X, Xia Y, Chan HC. CFTR is a potential marker for nasopharyngeal carcinoma prognosis and metastasis. *Oncotarget* 2016;7:76955–65.
- Zhang J, Feng W, Ye Z, Wei Y, Li L, Yang Y. Prognostic significance of platelet-to-lymphocyte ratio in patients with nasopharyngeal carcinoma: a meta-analysis. *Future Oncol* 2020;16:117–27.
- Spence T, Bruce J, Yip KW, Liu FF. MicroRNAs in nasopharyngeal carcinoma. *Chin Clin Oncol* 2016;5:17.
- Wang S, Claret FX, Wu W. MicroRNAs as therapeutic targets in nasopharyngeal carcinoma. *Front Oncol* 2019;9:756.
- Zhou W, Chang A, Zhao H, Ye H, Li D, Zhuo X. Identification of a novel microRNA profile including miR-106b, miR-17, miR-20b, miR-18a and miR-93 in the metastasis of nasopharyngeal carcinoma. *Canc Biomarkers* 2020;27:533–9.
- Liu N, Chen NY, Cui RX, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. *Lancet Oncol* 2012;13:633–41.
- Lian J, Li Y, Yu M. MicroRNA-183 and microRNA-141 are potential risk factors for poor prognosis in patients with nasopharyngeal carcinoma. *Oncol Lett* 2019;17:1172–6.
- Zhang Z, Huang J, Wang G, Jin F, Zheng J, Xiao H. Serum miRNAs, a potential prognosis marker of loco-regionally advanced nasopharyngeal carcinoma patients treated with CCRT. *BMC Canc* 2020;20:183.
- Simonson B, Das S. MicroRNA therapeutics: the next magic bullet? *Mini Rev Med Chem* 2015;15:467–74.
- Ganju A, Khan S, Hafeez BB, et al. miRNA nanotherapeutics for cancer. *Drug Discov Today* 2017;22:424–32.
- Chen L, Heikkinen L, Wang C, Yang Y, Sun H, Wong G. Trends in the development of miRNA bioinformatics tools. *Briefings Bioinf* 2019;20:1836–52.
- Vizcaino C, Mansilla S, Portugal J. Sp1 transcription factor: a long-standing target in cancer chemotherapy. *Pharmacol Ther* 2015;152:111–24.
- Wu J, Tang X, Ma C, Shi Y, Wu W, Hann SS. The regulation and interaction of colon cancer-associated transcript-1 and miR7-5p contribute to the inhibition of SP1 expression by solamargine in human nasopharyngeal carcinoma cells. *Phytother Res* 2020;34:201–13.
- Ho HY, Lin CW, Chien MH, et al. Melatonin suppresses TPA-induced metastasis by downregulating matrix metalloproteinase-9 expression through JNK/SP-1 signaling in nasopharyngeal carcinoma. *J Pineal Res* 2016;61:479–92.
- Zhang SQ, Pan SM, Liang SX, Han YS, Chen HB, Li JC. Research status and prospects of biomarkers for nasopharyngeal carcinoma in the era of high-throughput omics. *Int J Oncol* 2021;58:1.

25. Lyu X, Fang W, Cai L, et al. TGF β R2 is a major target of miR-93 in nasopharyngeal carcinoma aggressiveness. *Mol Canc* 2014; 13:51.
26. Xu YF, Mao YP, Li YQ, et al. MicroRNA-93 promotes cell growth and invasion in nasopharyngeal carcinoma by targeting disabled homolog-2. *Canc Lett* 2015;363:146–55.
27. Hu W, Yao W, Li H, Chen L. MiR-30e-5p inhibits the migration and invasion of nasopharyngeal carcinoma via regulating the expression of MTA1. *Biosci Rep* 2020;40. BSR20194309.
28. Ma YX, Zhang H, Li XH, Liu YH. MiR-30e-5p inhibits proliferation and metastasis of nasopharyngeal carcinoma cells by targeting USP22. *Eur Rev Med Pharmacol Sci* 2018;22:6342–9.
29. Liu N, Tang LL, Sun Y, et al. MiR-29c suppresses invasion and metastasis by targeting TIAM1 in nasopharyngeal carcinoma. *Canc Lett* 2013;329:181–8.
30. Huang L, Hu C, Chao H, et al. miR-29c regulates resistance to paclitaxel in nasopharyngeal cancer by targeting ITGB1. *Exp Cell Res* 2019;378:1–10.
31. Qiu F, Sun R, Deng N, et al. miR-29a/b enhances cell migration and invasion in nasopharyngeal carcinoma progression by regulating SPARC and COL3A1 gene expression. *PLoS One* 2015; 10:e0120969.
32. Gao J, Shao Z, Yan M, Fu T, Zhang L, Yan Y. Targeted regulation of STAT3 by miR-29a in mediating Taxol resistance of nasopharyngeal carcinoma cell line CNE-1. *Canc Biomarkers* 2018; 22(4):641–8.
33. Zhang S, Yue W, Xie Y, et al. The four-microRNA signature identified by bioinformatics analysis predicts the prognosis of nasopharyngeal carcinoma patients. *Oncol Rep* 2019;42: 1767–80.
34. Hosseini SM, Soltani BM, Tavallaei M, et al. Clinically significant dysregulation of hsa-miR-30d-5p and hsa-let-7b expression in patients with surgically resected non-small cell lung cancer. *Avicenna J Med Biotechnol (AJMB)* 2018;10:98–104.
35. Kohno T, Morishita A, Iwama H, et al. Comprehensive analysis of circulating microRNAs as predictive biomarkers for sorafenib therapy outcome in hepatocellular carcinoma. *Oncol Lett* 2020;20:1727–33.
36. Wang T, Wu J, Wu Y, Cheng Y, Deng Y, Liao J. A novel microRNA-based signature predicts prognosis among nasopharyngeal cancer patients. *Exp Biol Med* 2021;246:72–83.
37. Zhao L, Fong AHW, Liu N, Cho WCS. Molecular subtyping of nasopharyngeal carcinoma (NPC) and a microRNA-based prognostic model for distant metastasis. *J Biomed Sci* 2018;25:16.
38. Liu N, Cui RX, Sun Y, et al. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. *Int J Canc* 2014;134: 1359–68.
39. Zhao H, Chang A, Ling J, Zhou W, Ye H, Zhuo X. Construction and analysis of miRNA-mRNA regulatory networks in the radioresistance of nasopharyngeal carcinoma. *J Biotech* 2020;10: 511.
40. Xu Y, Huang X, Ye W, et al. Comprehensive analysis of key genes associated with ceRNA networks in nasopharyngeal carcinoma based on bioinformatics analysis. *Canc Cell Int* 2020; 20:408.
41. Lung RW, Hau PM, Yu KH, et al. EBV-encoded miRNAs target ATM-mediated response in nasopharyngeal. *J Pathol* 2018;244: 394–407.
42. Jiang X, Feng L, Dai B, Li L, Lu W. Identification of key genes involved in nasopharyngeal carcinoma. *Braz J Otorhinolaryngol* 2017;83:670–6.
43. Huang SCM, Tsao SW, Tsang CM. Interplay of viral infection, host cell factors and tumor microenvironment in the pathogenesis of nasopharyngeal carcinoma. *Cancers (Basel)* 2018;10: 106.
44. Cortez MA, Anfossi S, Ramapriyan R, et al. Role of miRNAs in immune responses and immunotherapy in cancer. *Genes Chromosomes Cancer* 2019;58:244–53.
45. Coghill AE, Wang CP, Verkuijlen SAWM, et al. Evaluation of nasal and nasopharyngeal swab collection for the detection of Epstein-Barr virus in nasopharyngeal carcinoma. *J Med Virol* 2018;90:191–5.
46. Hao SP, Tsang NM, Chang KP, Ueng SH. Molecular diagnosis of nasopharyngeal carcinoma: detecting LMP-1 and EBNA by nasopharyngeal swab. *Otolaryngol Head Neck Surg* 2004;131: 651–4.
47. Chang HW, Chan A, Kwong DL, Wei WI, Sham JS, Yuen AP. Detection of hypermethylated RIZ1 gene in primary tumor, mouth, and throat rinsing fluid, nasopharyngeal swab, and peripheral blood of nasopharyngeal carcinoma patient. *Clin Canc Res* 2003;9:1033–8.