

RESEARCH

Open Access



# Identification of a five-miRNA signature as a novel potential prognostic biomarker in patients with nasopharyngeal carcinoma

Bo Tu<sup>1†</sup>, Ling Ye<sup>2†</sup>, Qingsong Cao<sup>1</sup>, Sisi Gong<sup>1</sup>, Miaohua Jiang<sup>1</sup> and Hui Li<sup>1\*</sup>

## Abstract

**Background:** MicroRNAs (miRNAs) are involved in the prognosis of nasopharyngeal carcinoma (NPC). This study used clinical data and expression data of miRNAs to develop a prognostic survival signature for NPC patients to detect high-risk subject.

**Results:** We identified 160 differentially expressed miRNAs using RNA-Seq data from the GEO database. Cox regression model consisting of hsa-miR-26a, hsa-let-7e, hsa-miR-647, hsa-miR-30e, and hsa-miR-93 was constructed by the least absolute contraction and selection operator (LASSO) in the training set. All the patients were classified into high-risk or low-risk groups by the optimal cutoff value of the 5-miRNA signature risk score, and the two risk groups demonstrated significant different survival. The 5-miRNA signature showed high predictive and prognostic accuracies. The results were further confirmed in validation and external validation set. Results from multivariate Cox regression analysis validated 5-miRNA signature as an independent prognostic factor. A total of 13 target genes were predicted to be the target genes of miRNA target genes. Both PPI analysis and KEGG analysis networks were closely related to tumor signaling pathways. The prognostic model of mRNAs constructed using data from the dataset GSE102349 had higher AUCs of the target genes and higher immune infiltration scores of the low-risk groups. The mRNA prognostic model also performed well on the independent immunotherapy dataset Imvigor210.

**Conclusions:** This study constructed a novel 5-miRNA signature for prognostic prediction of the survival of NPC patients and may be useful for individualized treatment of NPC patients.

**Keywords:** Nasopharyngeal carcinoma, GEO, KEGG, miRNA, Prognostic

## Background

Nasopharyngeal carcinoma (NPC) occurs in the roof and lateral wall of the nasopharyngeal cavity. The common symptoms of NPC include nasal congestion, blood in the mucus, stuffiness in the ears, hearing loss, facial numbness, diplopia, and headache. It is associated with EBV infection, environment, genetics, smoking, and some other factors [1, 2]. NPC is particularly common in the Southern part of China and Southeast Asia, where it ranks first among head and neck malignant tumors in terms of morbidity and mortality [3, 4].

\*Correspondence: huilidoc@163.com

<sup>†</sup>Bo Tu and Ling Ye equally contributed to this work.

<sup>1</sup> Department of Otorhinolaryngology and Head Neck Surgery, The first affiliated hospital of Jinan University, Guangzhou 510630, Guangdong, China

Full list of author information is available at the end of the article

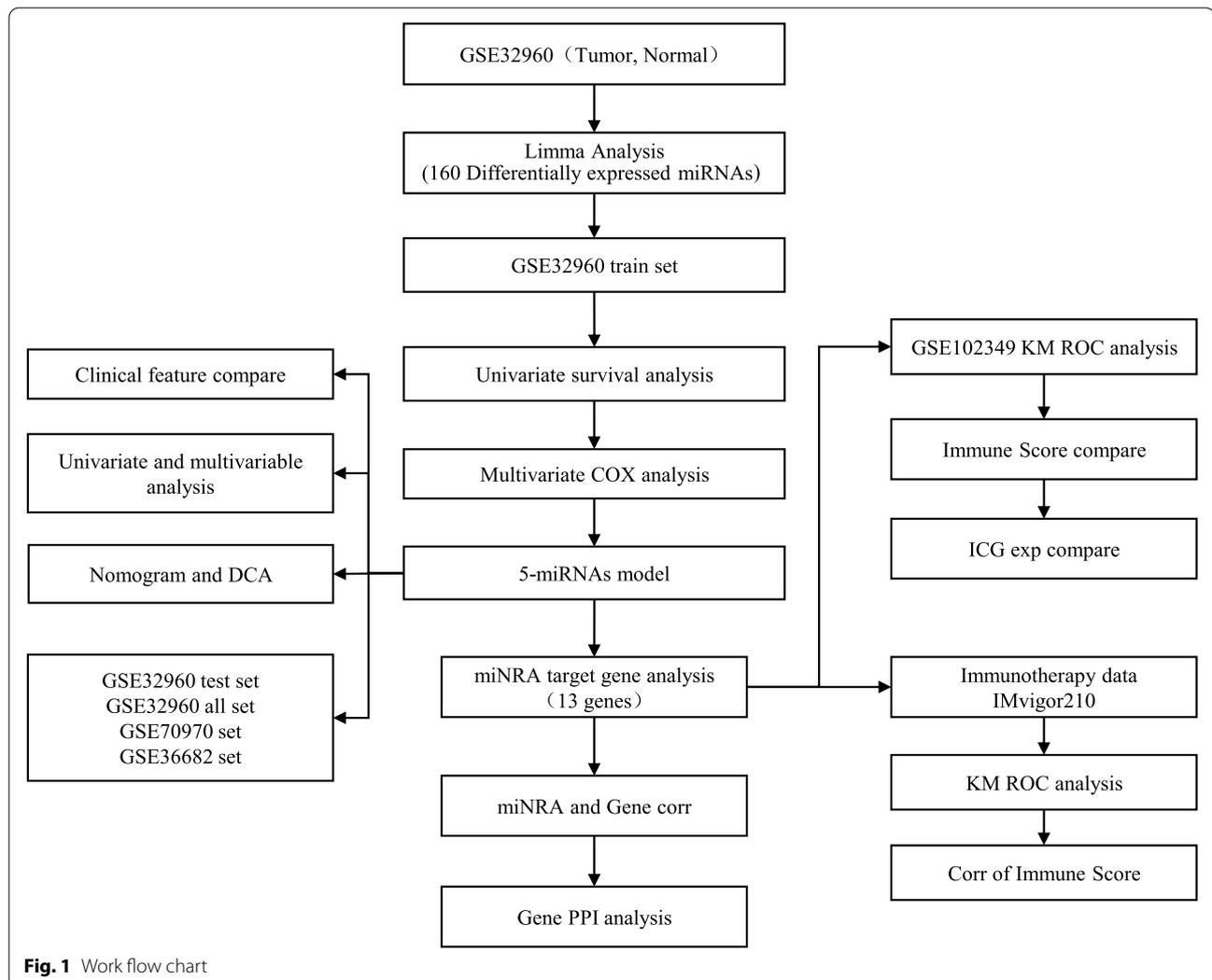


Despite sensitivity to radiotherapy and chemotherapy, the failure rate of nasopharyngeal cancer treatment remains high, with the main causes of failure being local recurrence and metastasis. Approximately 30–40% of patients with locally advanced NPC eventually develop distant metastases after radical treatment [5]. The median survival of patients with distant metastasis of NPC was only 19~21 months [6]. High risk of metastasis and recurrence of NPC will increase the complexity of and difficulty of treatment, and currently there is an urgent need to improve the treatment effect with effective biomarkers and treatment strategies. To achieve such a goal, it is very highly important to explore the biological and molecular mechanism of NPC and identify biomarkers related to the stratification of prognosis risk.

MicroRNAs (miRNAs) are a class of short, endogenous primed non-coding RNAs (18–25 nucleotides)

that regulate gene expression through pairing with 3'-untranslated region (3'-UTR) bases to mediate translation inhibition or degradation of homologous mRNAs [7]. Extensive evidence suggested that miRNAs may act as tumor suppressor or oncogenes and play critical roles in the proliferation, invasion, apoptosis, differentiation and metabolism of tumor progression [8, 9]. In addition, about 50% of miRNAs were found to be located in “fragile loci” in the genome, which are most frequently amplified or lost in cancer.

It has been found that some miRNAs, including let-7a [10], miR-34 [11] and miR-93 [12], are associated with the progression of NPCs through regulating cell metastasis, growth, and apoptosis. 8-miRNA and 16-miRNA markers identified by Wen et al. can be used in diagnosing NPC, and these two are the first diagnostic markers of NPC screened from the whole blood sample so far. In



particular, this study found that 16 miRNAs could effectively differentiate NPC from head and neck tumors [13]. Based on plasma EBVDNA and clinicopathological variables, a recent study established a nomogram as a more accurate predictor of NPC prognosis [14]. Ma and co-workers developed five miRNA signatures associated with NPC survival [15]. A recent study explored gene expression differences in patients with and without metastatic locally advanced NPC after treatment, and found a distant metastatic gene signature of 13 genes in locally advanced NPC, which has been considered as a reliable prognostic tool for recognizing distant metastasis of NPC [16]. Peter Shaw et al. also indicated prognostic significance of miRNAs as biomarkers in NPC patients [17]. These studies suggested that miRNAs could be expected to be effective prognostic markers and biological targets for NPC. The advantage of miRNA-based therapies is that miRNAs can simultaneously target multiple effector molecules in tumor cell differentiation and proliferation pathways. At present, these exploratory biomarker analyses are not sufficiently effective, thus, it is necessary to comprehensively explore the miRNA and regulatory mechanism related to NPC.

This study applied bioinformatics to systematically analyze the microarray data of multiple gene expression profiles to screen the key genes, signal pathways and regulatory mechanisms in the occurrence and development of NPC (Fig. 1).

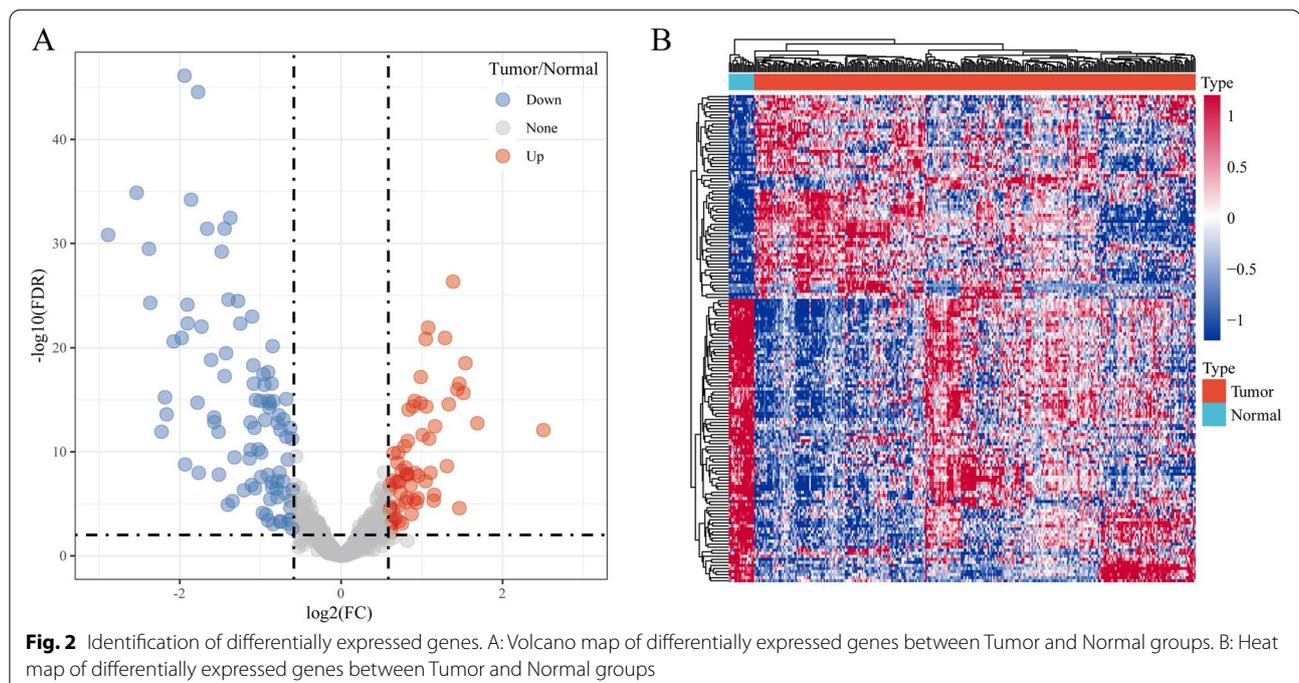
## Results

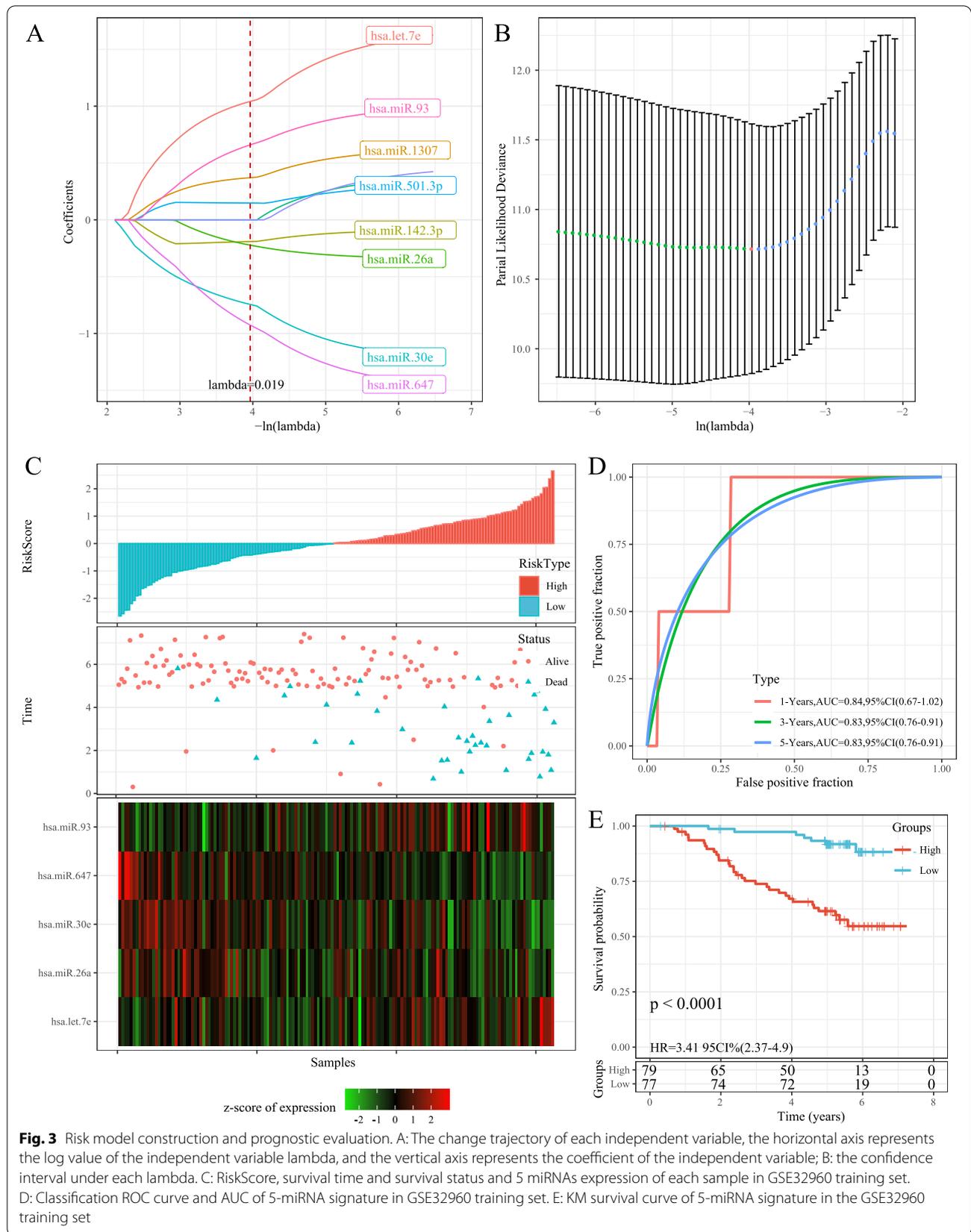
### Identification of differentially expressed miRNAs

Between tumor and normal in the GSE32960 dataset, the limma package was used to calculate the differentially expressed miRNAs. A total of 160 differentially expressed miRNAs obtained included 67 up-regulated miRNAs and 93 down-regulated miRNAs (Fig. 2A, B). The univariate Cox proportional risk regression model was further conducted on the differentially expressed miRNAs using the R package survival coxph function, and we obtained a total of 11 prognostically related miRNAs here.

### Risk model construction and prognostic evaluation

For the 11 prognostic-related miRNAs, the R software package glmnet was used to perform lasso cox regression analysis. From the change trajectory of the independent variable, as the lambda gradually increased, the number of independent variable coefficients close to 0 also increased gradually (Fig. 3A). The model was constructed applying 10-fold cross-validation to determine the confidence interval of each lambda. The model was the optimal when  $\lambda = 0.019$  (Fig. 3B). Hence, 8 miRNAs with  $\lambda = 0.019$  were chosen as the target genes. The stepAIC method in the MASS package further reduced 8 miRNAs to 5 miRNAs (hsa-miR-26a, hsa-let-7e, hsa-miR-647, hsa-miR-30e, and hsa-miR-93). The 5-miRNA signature formula was as follows:





$$\begin{aligned} \text{RiskScore} = & 1.302 * (\text{hsa} - \text{let} - 7e) - 0.468 * (\text{hsa} - \text{miR} - 26a) \\ & - 1.108 * (\text{hsa} - \text{miR} - 30e) - 1.453 * (\text{hsa} - \text{miR} - 647) \\ & + 0.88 * (\text{hsa} - \text{miR} - 93) \end{aligned}$$

Risk score of each sample in the GSE32960 training set was calculated based on the expression level of genes. We divided the 406 patients into high- and low-risk groups with median risk scores as the cutoff value. OS of the low-risk group was obviously longer than the high-risk group. Heatmap of miRNA showed that miR-26a, hsa-miR-30e and hsa-miR-647 were protective factors, while hsa-let-7e and hsa-miR-93 were risk factors (Fig. 3C). The 1-year, 2-year and 5-year AUC for ROC curve of training data set was 0.84, 0.83 and 0.83, respectively (Fig. 3D). KM survival curve showed that the OS time in high-risk group had a shorter time than that in low-risk group ( $p < 0.0001$ , Fig. 3E).

#### Robustness of risk models

To analyze the model robustness, we used the same model and coefficients in the validation set of GSE32960 and the full data set of GSE32960 as training set to calculate the risk score of each sample based on sample expression level, and the samples were classified into high-risk and low-risk groups based on the median value. Similarly, the OS of the low-risk group was obviously longer than that of the high-risk group. Heatmap of miRNA showed that miR-26a, hsa-miR-30e and hsa-miR-647 were protective factors, while hsa-let-7e and hsa-miR-93 were risk factors (Fig. 4A, D). The AUC for ROC curve of test dataset and entire GSE32960 dataset were all higher (Fig. 4B, E). KM survival curve also demonstrated that the OS time in high-risk group had a shorter time than that in low-risk group (Fig. 4C, F). The external datasets GSE70970 and GSE36682 were further used to verify the prognostic prediction ability of the gene signature. The OS of the low-risk group was obviously longer than that of the high-risk group in GSE70970 and GSE36682 dataset. Heatmap of miRNA showed that miR-26a, hsa-miR-30e and hsa-miR-647 were protective factors, while hsa-let-7e and hsa-miR-93 were risk factors (Fig. 5A, D). The AUC for ROC curve of GSE70970 and GSE36682 dataset were all higher (Fig. 5B, E). KM survival curve also showed that the

OS time in high-risk group had a shorter time than that in low-risk group (Fig. 5C, F).

#### Analysis of risk scores on clinical characteristics

Comparison of the distribution of RiskScore among clinical feature groups in the GSE32960 data set (Fig. 6) showed significant differences between T Stage, Stage, and the risk of developing metastasis ( $p < 0.05$ ). More advanced stages such as T Stage and Stage were positively related to a higher risk score. For samples with metastasis risk, the risk score of the sample with metastasis was significantly higher than those without metastasis.

#### Independence analysis of 5-miRNA signature

The relationship between prognostic indicators and clinicopathological characteristics was analyzed, including Age, Gender, T Stage, N Stage, Stage, Metastasis, Relapse, Radiotherapy interrupt, Radiotherapy boosting, and Cocurrent chemotherapy. Univariate analyses demonstrated that RiskType, Stage, Metsastasisi, Replase were significant risk factors for poor outcome (Fig. 7A). Multivariate analysis indicated that a high riskscore and high Metsastasisi, Replase were independently associated with worse OS (Fig. 7B).

#### Establishment of the nomogram and assessment of OS prediction

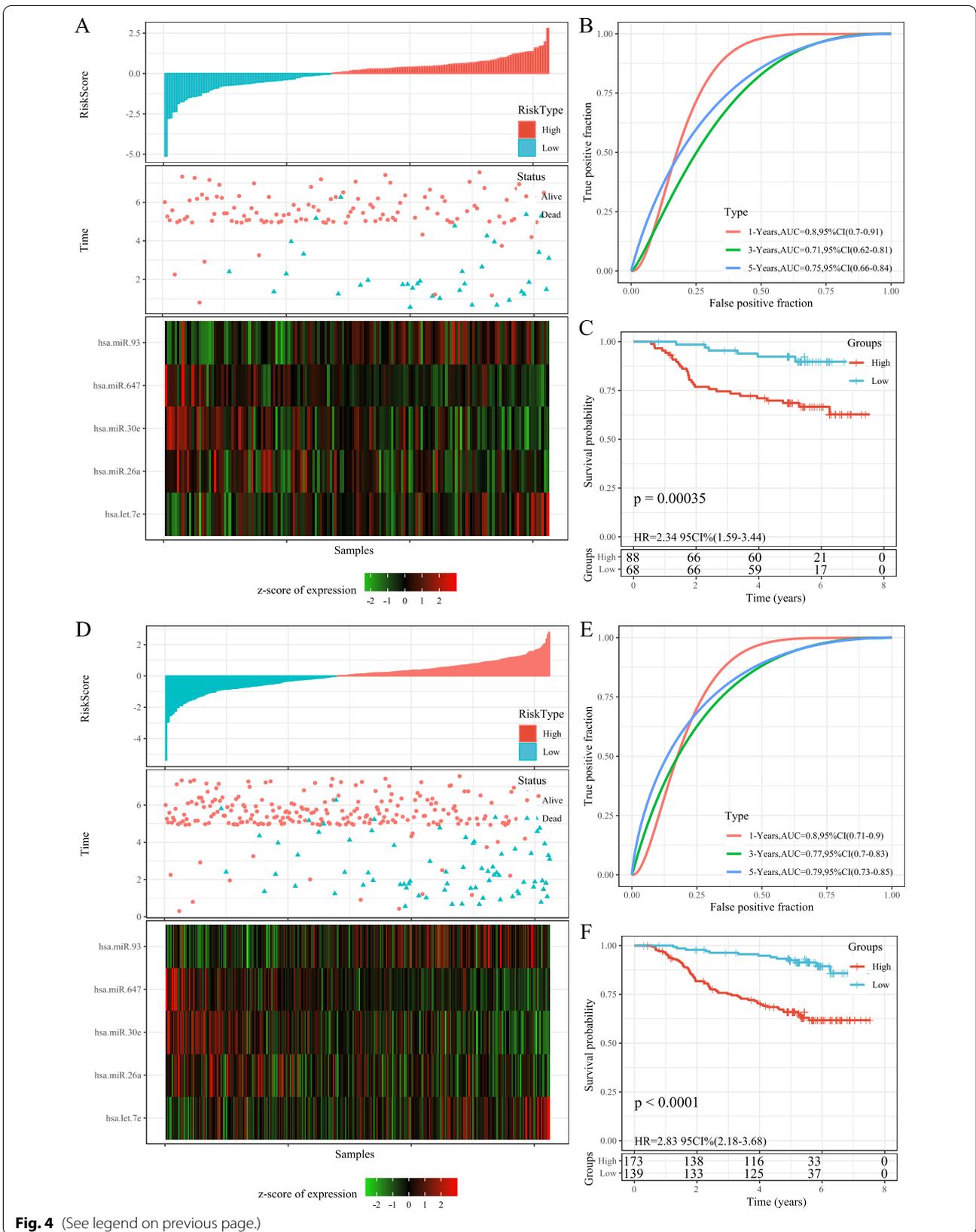
According to the results of univariate and multivariate Cox analysis, the 5-lncRNA signature was a reliable method to predict OS of patients with NPC. Then, according to the results from the multivariate Cox regression, we established a nomogram consisting of Age, N Stage, Metastasis, Relapse and RiskScore (Fig. 8A). The calibration plots verified the satisfactory predictive value between predictive values and observation values (Fig. 8B). The DCA curve also showed that RiskScore and Nomogram had a strong predictive effects (Fig. 8C).

#### PPI analysis of miRNA target genes

A total of 13 genes were predicted to be miRNA target genes, among them, 6 were the target genes of hsa-miR-26a, 4 were the target genes of hsa-miR-30e and 3

(See figure on next page.)

**Fig. 4** Robustness of risk models. A: RiskScore, survival time and survival status and 5 miRNAs expression of each sample in GSE32960 test dataset. B: classification ROC curve and AUC of 5-miRNA signature in GSE32960 test dataset. C: KM survival curve of 5-miRNA signature in the GSE32960 test dataset. D: RiskScore, survival time and survival status and 5 miRNAs expression of each sample in entire GSE32960 dataset. E: classification ROC curve and AUC of 5-miRNA signature in entire GSE32960 dataset. F: KM survival curve of 5-miRNA signature in the entire GSE32960 dataset



**Fig. 4** (See legend on previous page.)

were the target genes of hsa-miR-93 (Fig. 9A-C). In the GSE118721 data set, Pearson correlation coefficients between miRNAs and corresponding target genes were calculated, and the results showed that most miRNAs were negatively correlated with target genes (Fig. 9D-F). PPI analysis was performed on the 13 target genes identified by miRNA analysis using NetworkAnalyst 3.0, and PPI correlation network analysis was performed using STRING. There were two networks, and KEGG pathway enrichment analysis was conducted using the related genes in the network. Subnet 1 annotated tumor-related pathways such as Wnt signaling pathway, Cell cycle, Prostate cancer, Pathways in cancer, Hepatitis C, Hepatitis B (Fig. 9G). Subnet 2 was annotated to Wnt signaling pathway, mTOR signaling pathway, Pathways in cancer, Proteoglycans in cancer and other tumor-related pathways (Fig. 9H).

#### Prognostic effect and immune infiltration analysis of target gene model

Multivariate Cox analysis was performed on 13 target genes in the data set GSE102349 to obtain the risk coefficient of each gene. The samples were divided into high- and low-risk groups, and survival curve analysis and ROC curve analysis were performed. The results showed that the high- and low-risk groups of the data set GSE102349 had significant survival differences (Fig. 10A). The ROC curve of 1, 2, and 3 years had a high AUC value (above 0.88, Fig. 10B). The immune scores of the three software all demonstrated that the immune scores of the high-risk group were lower than those of the low-risk group (Fig. 10C, D, E). Furthermore, we compared the expression of PDCD1 (PD-L1), CTLA4 and IFNG (IFN- $\gamma$ ) genes in the high- and low-risk groups, and found that the expression of these three genes in the high-risk group was significantly lower than low-risk group (Fig. 10F). At the same time, correlation calculations between PDCD1 (PD-L1), CTLA4, and IFNG (IFN- $\gamma$ ) with T cells and CD8 T cells all showed a positive correlation (Fig. 10G). The above results indicated that low-risk grouped samples may have a better response to immunotherapy.

#### Prediction of immunotherapy by risk modeling of target genes

Currently, effective predictive markers for immunotherapy are limited. Identification of novel predictive

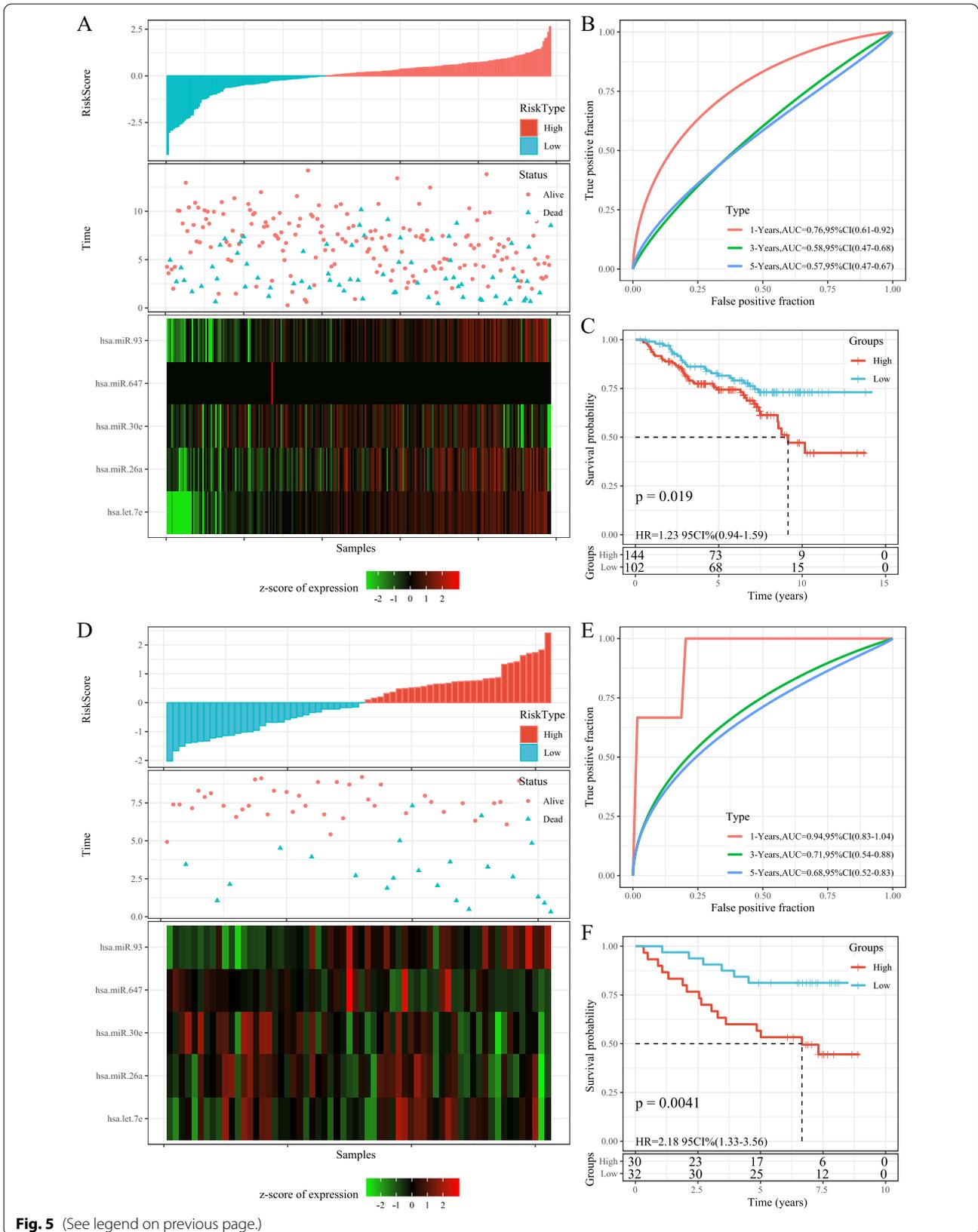
markers is critical to further develop precision immunotherapy. We searched an immunotherapy dataset (Imvigor210), which stored expression data in human mUC samples with patients' response to anti-PD-L1 immunotherapy. From Kaplan-Meier curves, it could be found that in mUC patients receiving immunotherapy, those with higher RiskScore values had poorer survival (Fig. 11A). The ROC curve showed that RiskScore had a higher AUC (Fig. 11B). Significant differences between immunotherapy responders and non-responders in high- and low-risk subgroups were detected, with a smaller proportion of patients in the high-risk subgroup showing response to immunotherapy (Fig. 11C). The immune cell scores of Imvigor210 samples were calculated using MCPcounter, and the correlation between RiskScore and TMB, NEO and immune cell scores was calculated. We found that RiskScore showed a negative correlation with TMB and NEO (Fig. 11D). At the same time, we compared the differences between RiskScore in different subgroups, and the data showed that RiskScore had significant differences in the response of effectiveness subgroup to immunotherapy (Fig. 11E), but RiskScore did not show significant differences between immune cell subgroup, tumor cells or immune phenotype (Fig. 11F-H).

#### Discussion

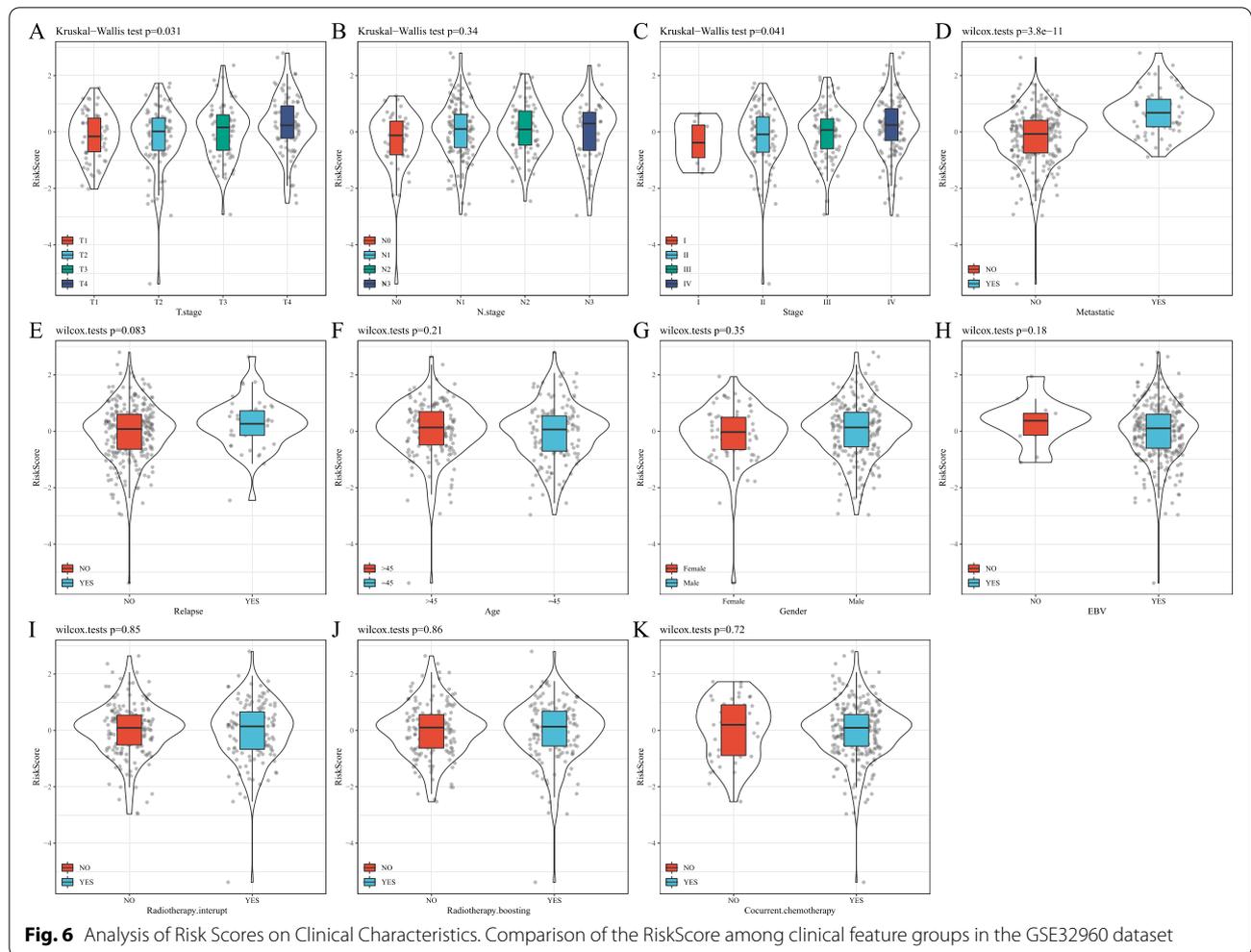
Some of the key microRNAs identified in this study were consistent with other previous findings. Study showed miR-26a is downregulated in NPC and targets mRNAs, and that an increase in protein-encoding mRNAs may promote NPC invasion and metastasis [18, 19]. Therefore, downregulating miR-26a expression in NPC cells could contribute to NPC cell invasion and metastasis, leading to a poor overall survival. miR-93 was significantly upregulated in NPC cell lines and clinical specimens, and depletion of miR-93 inhibited NPC growth, invasion and migration in vitro and in vivo [12]. Li Jiang et al. reported that 3 small extracellular vesicles-derived miRNAs (miR-134-5p, miR-205-5p, and miR-409-3p) could potentially act as an alternative or complementary approach for diagnosing NPC [20]. Ma and co-workers identified five miRNAs (miR-142-3p, miR-29c, miR-26a, and miR-30e and miR-93) significantly associated with DFS as independent prognostic factors for NPC [15]. In contrast, this

(See figure on next page.)

**Fig. 5** Robustness of risk models. A: RiskScore, survival time and survival status and 5 miRNAs expression of each sample in GSE70970 dataset. B: classification ROC curve and AUC of 5-miRNA signature in GSE70970 dataset. C: KM survival curve of 5-miRNA signature in the GSE70970 dataset. D: RiskScore, survival time and survival status and 5 miRNAs expression of each sample in GSE70970 dataset. E: classification ROC curve and AUC of 5-miRNA signature in GSE70970 dataset. F: KM survival curve of 5-miRNA signature in the GSE70970 dataset



**Fig. 5** (See legend on previous page.)



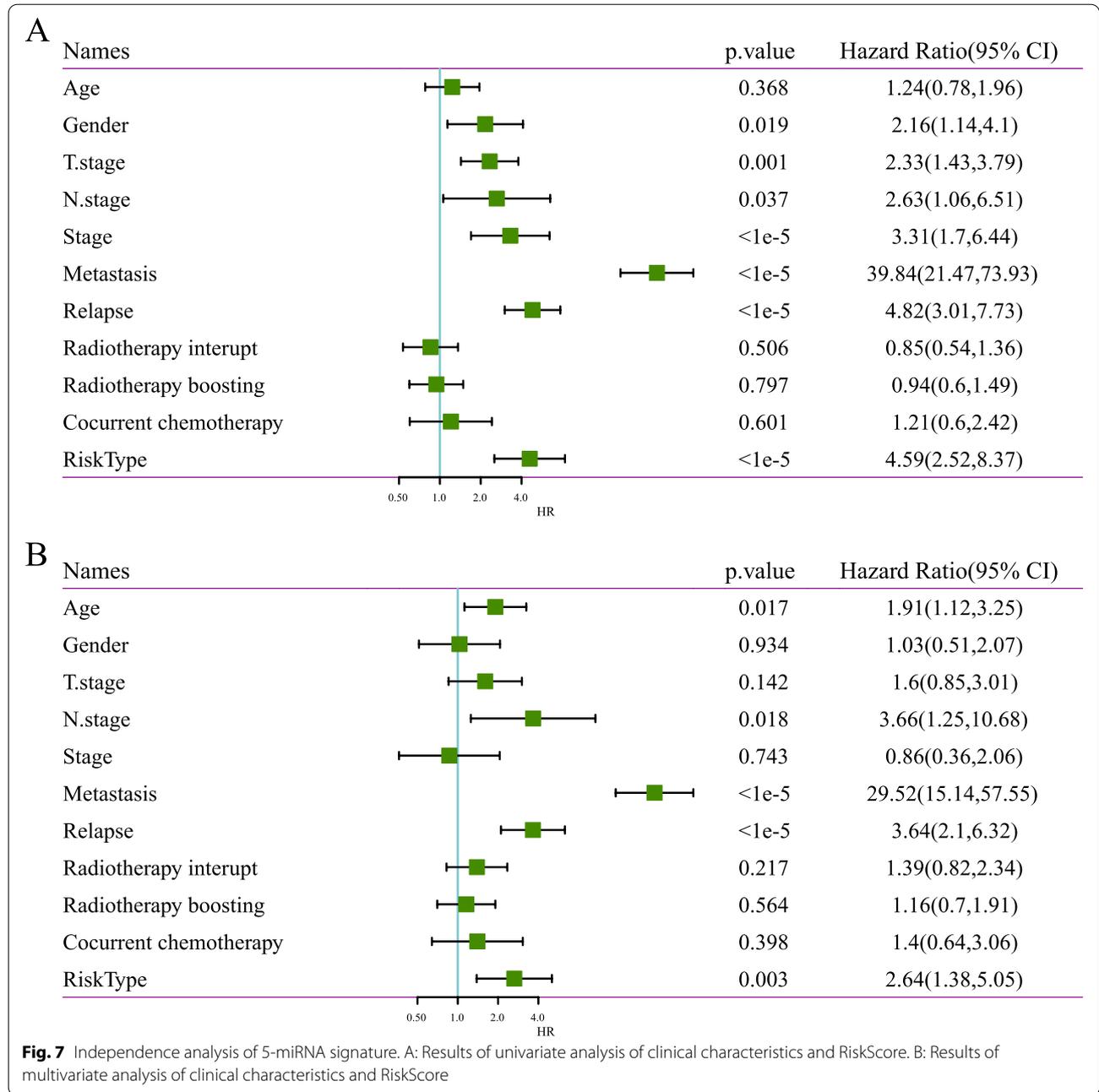
study again verified that miR-26a, miR-30e, and miR-93 could predict the prognosis of patients with NPC from multiple data sets, providing evidence for clinical diagnosis and treatment to find potential molecular markers and therapeutic targets. There is nearly no available research on the expression of hsa-miR-30e, hsa-miR-647 and hsa-let-7e in NPC, but they were the key microRNAs discovered in this study.

In this study, multiple databases were used to predict DEM target genes to improve the accuracy of predicting NPC prognosis. The intersection of DEM and differentially expressed genes of NPC were taken, a total of 13 intersection genes were found to be involved in the miRNA-mediated regulation of NPC. To understand the potential functional role of miRNA, target genes of miRNA were analyzed by PPI and KEGG functional annotation, and the results showed that DEG was mainly involved in Wnt signaling pathway,

Cell cycle, mTOR signaling pathway, Pathways in cancer. Recently, Wnt/ $\beta$ -catenin as the key pathway for carcinogenesis has been reported to play a critical role in the induction and maintenance of EMT and LNM [21]. Interestingly, dysregulated signaling of the Wnt/ $\beta$ -catenin signaling pathway enhances the malignancy of many human cancers, including NPC [22]. The mTOR signaling pathway could promote cell proliferation and inhibiting apoptosis of NPC [23, 24]. The analysis of the signaling pathways involved in NPC-related intersection genes may provide new insights into the mechanisms of NPC cells.

### Conclusions

In summary, we successfully identified five miRNAs that may play a key role in the development of NPC metastasis using a comprehensive bioinformatics analysis. Hsa-miR-30e, HSA-miR-647 and HSA-let-7E, which have

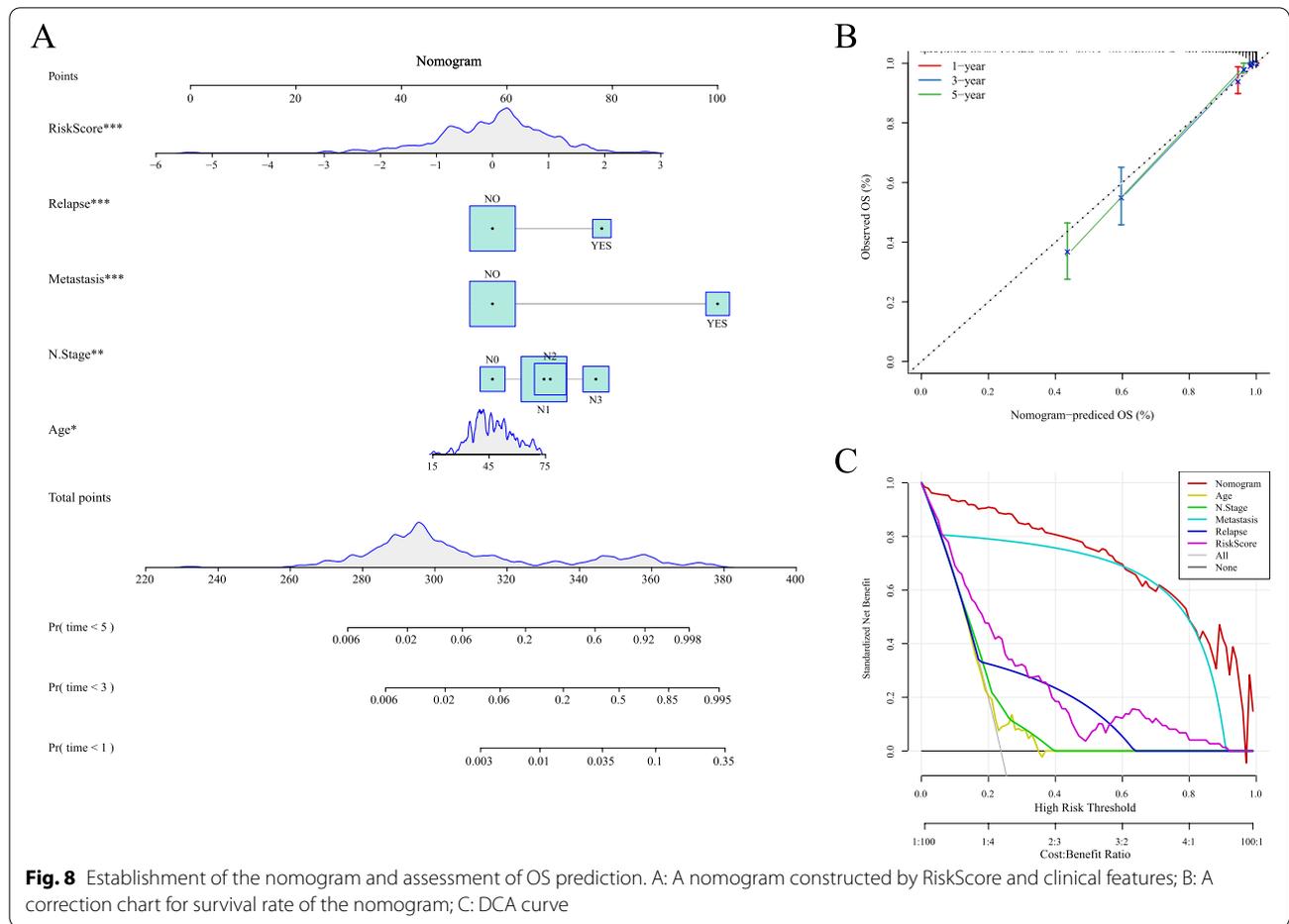


not been reported to be related to NPC, can be used as new potential markers of NPC. The regulatory role of relevant miRNAs in the mechanism of NPC occurrence and development and the detailed regulatory mechanism should be further studied. This study contributed to the promotion of individualized treatment of NPC in the future.

**Methods**

**Data acquisition and processing**

Clinical data and gene expression data of NPC were downloaded from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) database [25] including the GSE32960 [26], GSE70970 [27], GSE36682 [28], GSE118721, and GSE102349 [29]

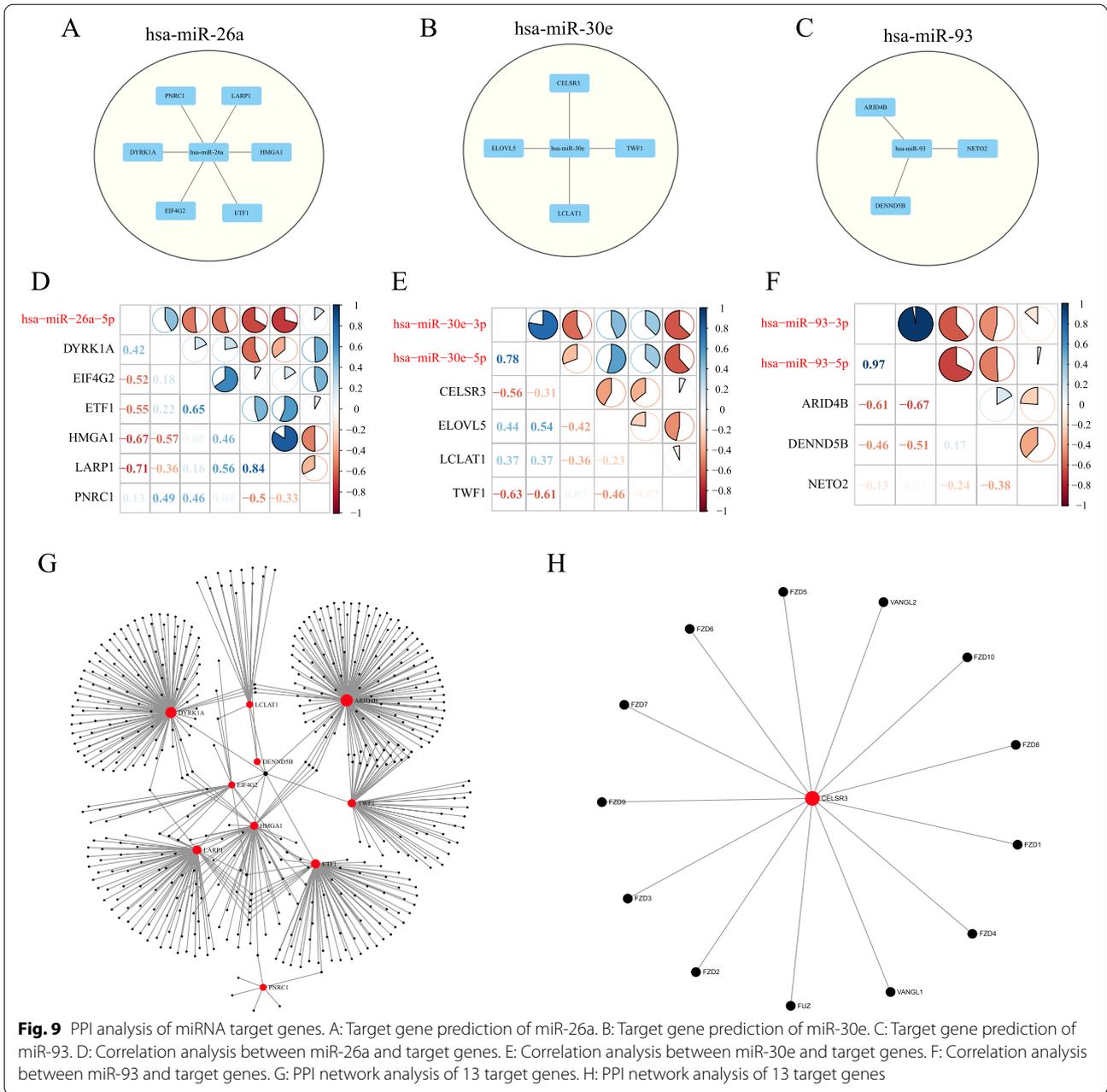


datasets. The GSE32960, GSE70970, and GSE36682 datasets were miRNA data containing clinical survival information; the GSE118721 dataset contained miRNA and mRNA data but without clinical survival information; and the GSE102349 dataset was mRNA data containing clinical survival information. The immunotherapy dataset used IMvigor210 [30] with expression data and clinical data.

For the GEO dataset, samples without clinical follow-up, survival time, or survival status were removed. Probes were converted to gene symbols. When a probe corresponded to multiple genes, the probe was removed, and the median expression value was taken from multiple gene symbols. For miRNA data, only human-associated miRNA expression data were retained. The sample information after data preprocessing is shown in Table 1. See Fig. 1 for the workflow diagram.

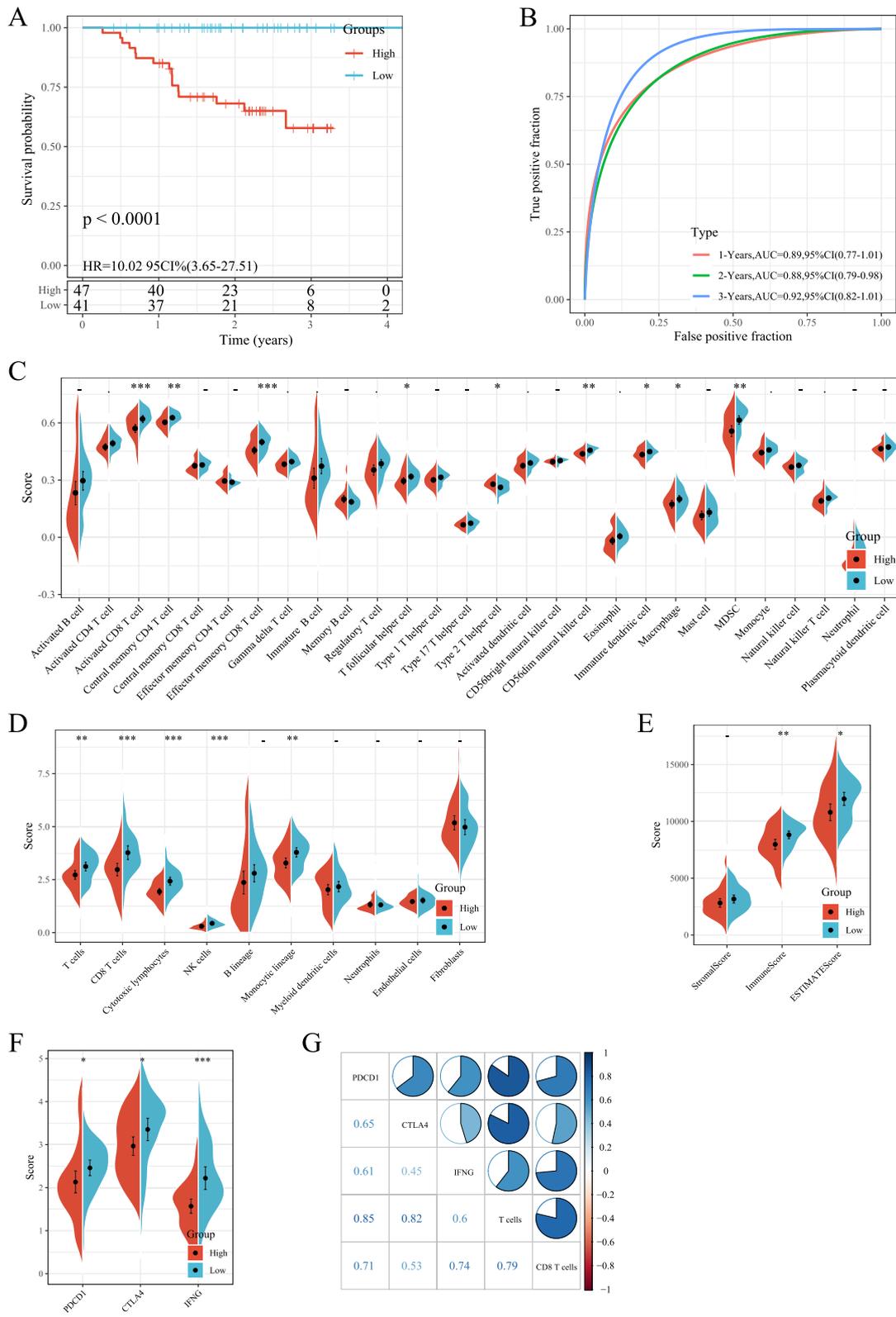
**Sample grouping**

Firstly, 312 tumor samples from the GSE32960 data set were assigned into training set and validation set. All the samples were put back into random groups for 100 times to avoid random allocation bias that may affect the stability of subsequent modeling. Here, grouping sampling was conducted in accordance at the ratio of training set: validation set = 1:1. Here, follow-up time, sex, age distribution, and proportion of deceased patients were similar in the two groups. The two randomly assigned data sets were clustered and the number of dichotomous samples was similar to select the most appropriate training set and verification set. There are 156 samples in the final GSE32960 training set and 156 samples in the GSE32960 test set (Table 2). The training set and test set samples were subjected to Chi-square test, and no significant difference between groups was shown ( $p > 0.05$ ).



(See figure on next page.)

**Fig. 10** Prognostic effect and immune infiltration analysis of target gene model. A: KM survival curves for high and low risk groups in the target gene prognostic model in dataset GSE102349. B: ROC curves of the target gene prognostic model in dataset GSE102349. C: Comparison of ssGSEA immunization scores for high- and low-risk subgroups in the GSE102349 dataset. D: Comparison of MCPcounter immunization scores for high- and low-risk subgroups in the GSE102349 dataset. E: Comparison of ESTIMATE immunization scores for high- and low-risk subgroups in the GSE102349 dataset. F: Comparison of immunotherapy gene expression between high and low risk groups in GSE102349 data set. G: Correlation between immune checkpoint and immune score in GSE102349 dataset



**Fig. 10** (See legend on previous page.)

### Identification of differentially expressed genes

Limma package [31] were employed to calculate the differentially expressed miRNA between Tumor and Normal in the GSE32960 dataset under the condition of  $FDR < 0.05$  and  $|\log_2FC| > 1.5$ .

### Cox risk analysis for univariate survival

Univariate Cox proportional risk regression models were constructed using the R package survival coxph function [32] for each differential miRNA as well as survival data in the GSE32960 training set.  $p < 0.05$  was the threshold value.

### Model construction

To simplify the risk model, the genes obtained from univariate Cox analysis were further filtered by Lasso cox regression using the R package glmnet. In stepwise regression, AIC red pool information criterion, which considers the statistical fit of the model and parameters number, was used. The stepAIC method in the MASS package started with the most complex model and sequentially removed a variable to reduce the AIC, with a smaller value indicating a better model. Combining the expression of each prognosis-related gene, we developed an independent prognosis model. The *RiskScore* was calculated using the following formula:

$$\begin{aligned} \text{RiskScore} = & 1.302 * (\text{hsa} - \text{let} - 7e) - 0.468 * (\text{hsa} - \text{miR} - 26a) \\ & - 1.108 * (\text{hsa} - \text{miR} - 30e) - 1.453 * (\text{hsa} - \text{miR} - 647) \\ & + 0.88 * (\text{hsa} - \text{miR} - 93) \end{aligned}$$

### Evaluation of the riskcore in GEO dataset

Each patient in the GSE32960 cohort, the GSE70970 dataset and GSE36682 dataset was assigned with a risk score using prognostic model. The median risk score as a cutoff was applied to classify NPC patient subjects into low-risk or high-risk group. We plotted survival curves using Kaplan-Meier (KM), and log-rank tests were conducted to evaluate survival differences in the two groups. Receiver operating characteristic curve (ROC) was established using “timeROC” software package [33], and we calculated area under the curve (AUC) value to analyze model sensitivity and specificity. In addition, a prognostic nomogram based on the

Cox proportional hazards regression model was performed to visualize the correlation between individual predictors and survival in patients with NPC by using “rms” package of R software [34]. The performance of the prognostic line graph was assessed by C index and calibration curves.

Whether the model could be used as an independent prognostic factor was examined through including sex, age, T, M, and N, stage as independent variables. We also performed univariate cox regression analyses and multivariate cox regression analyses on changes of survival outcomes and survival time.

### Prediction of miRNA target genes

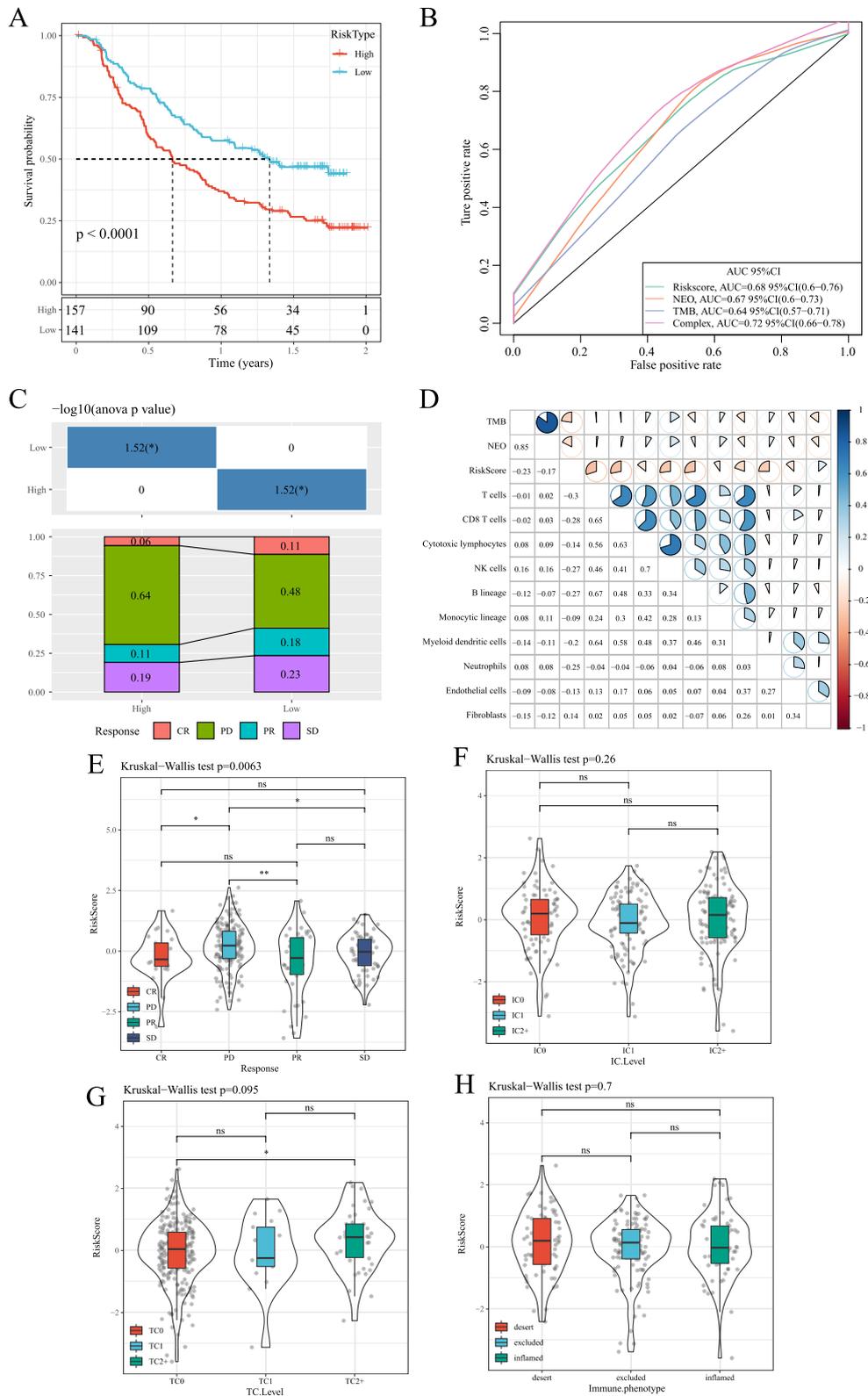
Target genes of the five prognostic-related miRNA genes were predicted using microT [35], miRanda [36], mircode [37], miRDB [38], miRmap [39], miRtarbase [40], PicTar [41], PITA, TargetMiner [42], and TargetScan [43]. Target genes predicted in all of the 10 databases were retained.

### Protein-protein interaction analysis

NetworkAnalyst 3.0 (<http://www.networkanalyst.ca/>) [44] is a web-based visual analysis platform for comprehensive interpretation of gene expression data at the level of system. We used NetworkAnalyst 3.0 to generate protein-protein interaction (PPI) network diagrams using experimentally validated data from STRING (version v11.0, <https://string-db.org/>) [45]. STRING (<https://string-db.org/>) is a web-based visualization platform for comprehensive analysis and interpretation of system-level gene expression data. The database (.org/) is a searchable database of known interactions between proteins and predicted interactions between proteins. In PPI networks, biological systems are described based on proteins (e.g., nodes) and their relationships (physical/functional interactions) (e.g., edges), here each node represents a gene or protein and each edge represents an interaction between a pair of genes or proteins. Tools such as STRING databases and Cytoscape software are used for developing protein interaction networks for genes. Differentially expressed genes were subjected to KEGG pathway functional enrichment analysis in R software package WebGestaltR (V0.4.2) [46].

(See figure on next page.)

**Fig. 11** Prediction of immunotherapy by risk modeling of target genes. A: KM curve of target gene prognostic model in the Imvigor210 dataset. B: ROC curves of the target gene prognostic model in dataset Imvigor210. C: Comparison of immune cell scores between high-risk and low-risk groups in the Imvigor210 dataset. D: Comparison of immune cell scores between high-risk and low-risk groups in the Imvigor210 dataset. E: Comparison of StromalScore, ImmuneScore and ESTIMATEScore between high-risk and low-risk groups in the Imvigor210 dataset. F: Comparison of PDCD1, CTLA4 and IFNG expression between high-risk and low-risk groups in the Imvigor210 dataset. G: Correlation analysis between immune genes and immune cells in the Imvigor210 dataset



**Fig. 11** (See legend on previous page.)

**Table 1** Clinical information of the samples

Clinical Features	GSE32960(miRNA)	GSE36682(miRNA)	GSE70970(miRNA)	GSE102349(mRNA)
<b>Type</b>				
Normal	18			
Tumor	312	62	246	88
<b>OS</b>				
0	238	40	176	72
1	74	22	70	16
<b>Metastatic</b>				
NO	246			
YES	66			
<b>Relapse</b>				
NO	268			
YES	44			
<b>T Stage</b>				
T1	66			
T2	89			
T3	71			
T4	86			
<b>N Stage</b>				
N0	44			
N1	148			
N2	72			
N3	48			
<b>Stage</b>				
I	12			
II	86			
III	91			
IV	123			
<b>Gender</b>				
Male	233			
Female	79			
<b>Age</b>				
≤ 45	148			
>45	164			
<b>EBV</b>				
YES	299			
NO	13			
<b>Radiotherapy interrupt</b>				
NO	178			
YES	134			
<b>Radiotherapy boosting</b>				
NO	163			
YES	149			
<b>Cocurrent chemotherapy</b>				
NO	44			
YES	268			

**Table 2** Sample Informations of GSE32960 Training Set and Validation Set

Clinical Features	GSE32960(train)	GSE36682(test)	P
<b>OS</b>			
0	117	121	0.6897
1	39	35	
<b>Metastatic</b>			
NO	122	124	0.8897
YES	34	32	
<b>Relapse</b>			
NO	132	136	0.6256
YES	24	20	
<b>T Stage</b>			
T1	30	36	0.7632
T2	48	41	
T3	36	35	
T4	42	44	
<b>N Stage</b>			
N0	23	21	0.7131
N1	73	75	
N2	39	33	
N3	21	27	
<b>Stage</b>			
I	5	7	0.8443
II	45	41	
III	47	44	
IV	59	64	
<b>Gender</b>			
Male	119	114	0.6025
Female	37	42	
<b>Age</b>			
≤ 45	78	70	0.2474
>45	78	86	
<b>EBV</b>			
YES	150	149	1
NO	6	7	
<b>Radiotherapy interrupt</b>			
NO	90	88	0.9089
YES	66	68	
<b>Radiotherapy boosting</b>			
NO	85	78	0.4965
YES	71	78	
<b>Cocurrent chemotherapy</b>			
NO	26	18	0.2549
YES	130	138	

**Immunoassay**

StromalScore, ImmuneScore, and ESTIMATEScore immune scores were examined using the R software package ESTIMATE [47]. 10 immune cell scores were analyzed by MCPcounter [48] and 28 immune cell scores were assessed using the ssGSEA method in the GSVA package [49].

**Statistical analysis**

Benjamini & Hochberg was used for multiple inspection correction in the current study.

**Abbreviations**

miRNAs: MicroRNAs; LASSO: Least absolute contraction and selection operator; NPC: Nasopharyngeal carcinoma; 3'-UTR: 3'-untranslated region..

**Authors' contributions**

Hui Li conceived and designed the research; Bo Tu and Ling Ye drafted the manuscript and got agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved; Miaohua Jiang contributed to date acquisition; Qingsong Cao analyzed data; Sisi Gong interpreted data; Hui Li revised the manuscript for important intellectual content; Bo Tu and Ling Ye contributed equally to this work. All authors approved the final version to be published.

**Funding**

This study was supported by Guangdong Medical Science and Technology Research Foundation [B2020025], Fundamental Research Funds for the Central Universities [21618305], and Project of Administration of Traditional Chinese Medicine of Guangdong Province of China [20201108].

**Availability of data and materials**

The datasets generated and/or analyzed during the current study are available in the [GSE32960] repository [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32960>], in [GSE70970] repository [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>], in [GSE36682] repository [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>], in [GSE118721] repository [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118721>], and in [GSE102349] repository [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>].

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Department of Otorhinolaryngology and Head Neck Surgery, The first affiliated hospital of Jinan University, Guangzhou 510630, Guangdong, China.

<sup>2</sup>Department of oncology, The first affiliated hospital of Jinan University, Guangzhou 510630, Guangdong, China.

Received: 6 August 2021 Accepted: 24 November 2021

Published online: 08 January 2022

## References

- Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet* (London, England). 2019;394(10192):64–80.
- Lee HM, Okuda KS, González FE, Patel V. Current perspectives on nasopharyngeal carcinoma. *Adv Exp Med Biol*. 2019;1164:11–34.
- Xia C, Yu XQ, Zheng R, Zhang S, Zeng H, Wang J, et al. Spatial and temporal patterns of nasopharyngeal carcinoma mortality in China, 1973–2005. *Cancer Lett*. 2017;401:33–8.
- Tang LL, Chen WQ, Xue WQ, He YQ, Zheng RS, Zeng YX, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. *Cancer Lett*. 2016;374(1):22–30.
- Hui EP, Leung SF, Au JS, Zee B, Tung S, Chua D, et al. Lung metastasis alone in nasopharyngeal carcinoma: a relatively favorable prognostic group. A study by the Hong Kong nasopharyngeal carcinoma study group. *Cancer*. 2004;101(2):300–6.
- Jin Y, Shi YX, Cai XY, Xia XY, Cai YC, Cao Y, et al. Comparison of five cisplatin-based regimens frequently used as the first-line protocols in metastatic nasopharyngeal carcinoma. *J Cancer Res Clin Oncol*. 2012;138(10):1717–25.
- Mohr AM, Mott JL. Overview of microRNA biology. *Semin Liver Dis*. 2015;35(1):3–11.
- Chan SH, Wang LH. Regulation of cancer metastasis by microRNAs. *J Biomed Sci*. 2015;22(1):9.
- Notarte KI, Senanayake S, Macaranas I, Albano PM, Mundo L, Fennell E, et al. MicroRNA and other non-coding RNAs in Epstein-Barr virus-associated cancers. *Cancers*. 2021;13(15).
- Wu A, Wu K, Li J, Mo Y, Lin Y, Wang Y, et al. Let-7a inhibits migration, invasion and epithelial-mesenchymal transition by targeting HMGA2 in nasopharyngeal carcinoma. *J Transl Med*. 2015;13:105.
- Li YQ, Ren XY, He QM, Xu YF, Tang XR, Sun Y, et al. MiR-34c suppresses tumor growth and metastasis in nasopharyngeal carcinoma by targeting MET. *Cell Death Dis*. 2015;6(1):e1618.
- Xu YF, Mao YP, Li YQ, Ren XY, He QM, Tang XR, et al. MicroRNA-93 promotes cell growth and invasion in nasopharyngeal carcinoma by targeting disabled homolog-2. *Cancer Lett*. 2015;363(2):146–55.
- Wang Y, Zhao Q, Lan N, Wang S. Identification of methylated genes and miRNA signatures in nasopharyngeal carcinoma by bioinformatics analysis. *Mol Med Rep*. 2018;17(4):4909–16.
- Tang LQ, Li CF, Li J, Chen WH, Chen QY, Yuan LX, et al. Establishment and validation of prognostic nomograms for endemic nasopharyngeal carcinoma. *J Natl Cancer Inst*. 2016;108(1).
- Liu N, Chen NY, Cui RX, Li WF, Li Y, Wei RR, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. *Lancet Oncol*. 2012;13(6):633–41.
- Tang XR, Li YQ, Liang SB, Jiang W, Liu F, Ge WX, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. *Lancet Oncol*. 2018;19(3):382–93.
- Shaw P, Senthilnathan R, Krishnan S, Suresh D, Shetty S, Muthukalianan GK, et al. A clinical update on the prognostic effect of microRNA biomarkers for survival outcome in nasopharyngeal carcinoma. *Syst Rev Meta-Anal Cancers*. 2021;13(17).
- Yu L, Lu J, Zhang B, Liu X, Wang L, Li SY, et al. miR-26a inhibits invasion and metastasis of nasopharyngeal cancer by targeting EZH2. *Oncol Lett*. 2013;5(4):1223–8.
- Lu J, He ML, Wang L, Chen Y, Liu X, Dong Q, et al. MiR-26a inhibits cell growth and tumorigenesis of nasopharyngeal carcinoma through repression of EZH2. *Cancer Res*. 2011;71(1):225–33.
- Jiang L, Zhang Y, Li B, Kang M, Yang Z, Lin C, et al. miRNAs derived from circulating small extracellular vesicles as diagnostic biomarkers for nasopharyngeal carcinoma. *Cancer Sci*. 2021;112(6):2393–404.
- Han X, Fang X, Lou X, Hua D, Ding W, Foltz G, et al. Silencing SOX2 induced mesenchymal-epithelial transition and its expression predicts liver and lymph node metastasis of CRC patients. *PLoS One*. 2012;7(8):e41335.
- Liang TS, Zheng YJ, Wang J, Zhao JY, Yang DK, Liu ZS. MicroRNA-506 inhibits tumor growth and metastasis in nasopharyngeal carcinoma through the inactivation of the Wnt/ $\beta$ -catenin signaling pathway by down-regulating LHX2. *J Exp Clin Cancer Res: CR*. 2019;38(1):97.
- Dai W, Dong P, Liu J, Gao Y, Hu Y, Lin H, et al. Euscaphic acid inhibits proliferation and promotes apoptosis of nasopharyngeal carcinoma cells by silencing the PI3K/AKT/mTOR signaling pathway. *Am J Transl Res*. 2019;11(4):2090–8.
- Lin YT, Wang HC, Chuang HC, Hsu YC, Yang MY, Chien CY. Pre-treatment with angiotensin-(1-7) inhibits tumor growth via autophagy by downregulating PI3K/Akt/mTOR signaling in human nasopharyngeal carcinoma xenografts. *J Mol Med (Berlin, Germany)*. 2018;96(12):1407–18.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(Database issue):D991–5.
- 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(1):16.
- Xu Y, Huang X, Ye W, Zhang Y, Li C, Bai P, et al. Comprehensive analysis of key genes associated with ceRNA networks in nasopharyngeal carcinoma based on bioinformatics analysis. *Cancer Cell Int*. 2020;20:408.
- Lu ZW, Du MY, Qian LX, Zhang N, Gu JJ, Ding K, et al. MiR-152 functioning as a tumor suppressor that interacts with DNMT1 in nasopharyngeal carcinoma. *Oncotargets Ther*. 2018;11:1733–41.
- Huang D, Liu Q, Zhang W, Huang C, Zheng R, Xie G, et al. Identified IGSF9 association with prognosis and hypoxia in nasopharyngeal carcinoma by bioinformatics analysis. *Cancer Cell Int*. 2020;20:498.
- Mariathanas S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGF $\beta$  attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018;554(7693):544–8.
- Xu Z, Wang Z, Sun H, Xin H. Evaluation of Exosomal miRNA in blood as a potential diagnostic biomarker for human non-small cell lung cancer. *Med Sci Monitor: Intl Med J Exp Clin Res*. 2020;26:e924721.
- Bhattacharjee A, Rajendra J, Dikshit R, Dutt S. HER2 borderline is a negative prognostic factor for primary malignant breast cancer. *Breast Cancer Res Treat*. 2020;181(1):225–31.
- Zhang C, Zheng Y, Li X, Hu X, Qi F, Luo J. Genome-wide mutation profiling and related risk signature for prognosis of papillary renal cell carcinoma. *Annals Trans Med*. 2019;7(18):427.
- McCool R, Wilson K, Arber M, Fleetwood K, Toupin S, Thom H, et al. Systematic review and network meta-analysis comparing ocrelizumab with other treatments for relapsing multiple sclerosis. *Multiple Sclerosis and Related Disorders*. 2019;29:55–61.
- Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic acids research*. 2013;41(Web Server issue):W169–73.
- Erickson SL, Salekin KL, Johnson LN, Doran SC. The predictive power of intelligence: Miranda abilities of individuals with intellectual disability. *Law Hum Behav*. 2020;44(1):60–70.
- Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics (Oxford, England)*. 2012;28(15):2062–3.
- Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res*. 2015;43(Database issue):D146–52.
- Vejnar CE, Blum M, Zdobnov EM. miRmap web: Comprehensive microRNA target prediction online. *Nucleic acids research*. 2013;41(Web Server issue):W165–8.
- Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020;48(D1):D148–d54.
- Lu T, Chen S, Qu L, Wang Y, Chen HD, He C. Identification of a five-miRNA signature predicting survival in cutaneous melanoma cancer patients. *PeerJ*. 2019;7:e7831.
- Bandyopadhyay S, Mitra R. TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples. *Bioinformatics (Oxford, England)*. 2009;25(20):2625–31.
- Zhao X, Ren Y, Cui N, Wang X, Cui Y. Identification of key microRNAs and their targets in exosomes of pancreatic cancer using bioinformatics analysis. *Medicine*. 2018;97(39):e12632.
- Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res*. 2019;47(W1):W234–w41.

45. Zhao C, Sahni S. String correction using the Damerau-Levenshtein distance. *BMC Bioinformatics*. 2019;20(Suppl 11):277.
46. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*. 2019;47(W1):W199–w205.
47. Luo Q, Vögeli TA. A methylation-based reclassification of bladder Cancer based on immune cell genes. *Cancers*. 2020;12(10).
48. Shi J, Jiang D, Yang S, Zhang X, Wang J, Liu Y, et al. LPAR1, correlated with immune infiltrates, is a potential prognostic biomarker in prostate Cancer. *Front Oncol*. 2020;10:846.
49. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics*. 2013;14:7.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

