

# Construction and Analysis of lncRNA-mediated ceRNA Network in Nasopharyngeal Carcinoma Based on Weighted Correlation Network Analysis

Zhenning Zou

Guangdong Medical University

Shuguang Liu

The Eighth Affiliated Hospital of Sun Yat-sen University

Yanping Ha

Guangdong Medical University

Bowan Huang (✉ [bowanhuang@126.com](mailto:bowanhuang@126.com))

Central People's Hospital of Zhanjiang <https://orcid.org/0000-0002-4651-2912>

---

## Primary research

**Keywords:** nasopharyngeal, carcinoma, lncRNA, ceRNA, weighted, gene, correlation, network, analysis, prognosis

**Posted Date:** September 24th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-35735/v2>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BioMed Research International on October 10th, 2020. See the published version at <https://doi.org/10.1155/2020/1468980>.

# Abstract

Increasing evidence indicated that aberrant expression of long noncoding RNAs (lncRNAs) are involved in tumorigenesis of nasopharyngeal carcinoma (NPC). The purpose of this study was to construct a lncRNA-mediated ceRNA network based on weighted correlation network analysis (WGCNA). First, modules with highly correlated genes were identified from GSE102349 via WGCNA, and the preservation of the modules was evaluated by GSE68799. Then, the differentially expressed lncRNAs and mRNAs identified from GSE12452 which belonged to the same WGCNA modules and the differentially expressed miRNAs identified from GSE32960 were used to construct a ceRNA network. The prognostic value of the network was evaluated by survival analysis. Furthermore, a risk score model for predicting progression-free survival (PFS) of NPC patients was established via lasso penalized Cox regression, and the differences in the expression of the lncRNAs between high- and low-risk groups were investigated. Finally, 14 stable modules were identified, and a ceRNA network composed of 11 lncRNAs, 15 miRNAs and 40 mRNAs was established. The lncRNAs and mRNAs in the network belonged to the turquoise and salmon modules. Survival analysis indicated that ZNF667-AS1, LDHA, LMNB2, TPI1, UNG and hsa-miR-142-3p were significantly correlated with the prognosis of NPC. Gene set enrichment analysis indicated that the up-regulation of ZNF667-AS1 was associated with some immune-related pathways. Besides, a risk score model consisting of 12 genes was constructed and showed a good performance in predicting PFS for NPC patients. Among the 11 lncRNAs in the ceRNA network, SNHG16, SNHG17, and THAP9-AS1 were up-regulated in the high-risk group of NPC, while ZNF667-AS1 was down-regulated in the high-risk group of NPC. These results will promote our understanding of the crosstalk among lncRNAs, miRNAs, and mRNAs in the tumorigenesis and progression of NPC.

## Introduction

Nasopharyngeal carcinoma (NPC) is a fatal malignancy arising from the nasopharynx epithelium. It is characterized by a distinctive geographical and ethnic distribution with a high incidence of NPC in Southern China, Southeast Asia and Middle East/North Africa [1]. Southern China accounts for 71% of new cases worldwide [2]. It was estimated that there were 129,079 new cases and 72,987 deaths around the world in 2018 [3]. With the emergence of advanced diagnostic methods and improvement of treatment methods, especially the application of intensity-modulated radiotherapy, the prognosis of NPC has been dramatically improved over the past decades[2]. Nevertheless, a portion of patients still had adverse outcomes because of locoregional recurrence and distant metastasis [4, 5]. Hence, it is necessary to explore the molecular mechanisms underlying NPC and develop more effective therapeutic approaches to reduce the occurrence of post-therapy complications and improve the prognosis of patients with NPC.

lncRNAs are referred to transcripts longer than 200 nucleotides. Although lacking protein-coding ability, lncRNAs can interact with RNA or DNA molecules via base pairing, even with double-stranded DNA, and form networks with DNA, protein complexes and RNA [6]. lncRNAs serve as competitive endogenous RNAs (ceRNAs) and implicated in the occurrence and progression of a variety of tumors, including NPC. For example, lncRNA H19 was suggested to regulate EZH2 expression by interacting with miR-630 and

promotes cell invasion in NPC [7]. LncRNA HOXC13-AS was reported to positively affect cell proliferation and invasion in NPC via modulating miR-383-3p/HMGA2 axis[8]. LncRNA FAM225A was revealed to promote NPC tumorigenesis and metastasis by sponging mir-590-3p/mir-1275 and upregulate ITGB3 [9]. LncRNA SNHG1 was shown to antagonize the effect of miR-145a-5p on the down-regulation of NUA1 in NPC[10]. The ceRNAs hypothesis supposed that any RNA transcript that harbors the miRNA recognition elements (MREs) can sequester miRNAs from other targets sharing the same MREs, thereby regulating the expression of target genes [11, 12]. However, there are still few comprehensive studies concerning lncRNA-mediated ceRNA network in NPC based on high-throughput data.

Weighted correlation network analysis (WGCNA) is a widely used and efficient system biology method [13]. This method can identify clusters (modules) of highly correlated genes based on high-dimensional data, and can further identify hub genes closely related to external phenotypic traits, which may be potential diagnostic and therapeutic markers. This process is different from many other bioinformatics methods because it does not rely on priori defined gene sets or pathways [14]. WGCNA has been applied in various biological processes, such as, cancer, genetics and brain imaging data analysis [15]. As for NPC, four miRNAs, including hsa-miR-142-3p, hsa-miR-150, hsa-miR-29b, and hsa-miR-29c, were obtained as prognostic markers by combining univariate Cox regression analysis with WGCNA [16]. Ge et al. identified several gene modules with different biological functions related to NPC through WGCNA analysis [17]. However, to our knowledge, no study has previously constructed a lncRNA-mediated ceRNA network based on WGCNA in NPC.

In the present study, we obtained RNA-Seq data of GSE102349 and performed WGCNA to enrich modules associated with NPC. Subsequently, we used mRNA profiling data (GSE12452) and miRNA profiling data (GSE32960) to identify differentially expressed genes (DEGs) and differentially expressed miRNAs (DEmiRNAs). Through annotation of DEGs and intersection with WGCNA modules, we obtained differentially expressed lncRNAs (DElncRNAs) and differentially expressed mRNAs (DEmRNAs) in the WGCNA modules. Then, those DElncRNAs, DEmiRNAs and DEmRNAs were used to construct a lncRNA-miRNA-mRNA ceRNA network based on LncBase v.2 [18] and miRTarBase [19] databases. Next, survival analysis was performed to explore the prognostic characteristics of the ceRNA network. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analysis were implemented for the WGCNA modules, lncRNAs and mRNAs in the ceRNA network. Gene set enrichment analysis (GSEA) was also carried out for the prognostic lncRNAs in the ceRNA network. Moreover, we constructed a lasso penalized Cox regression model for PFS prediction. The differences in the expression of the lncRNAs and mRNAs in the ceRNA network between high- and low-risk groups were investigated. The results of this study will promote our understanding of the crosstalk among lncRNAs, miRNAs, and mRNAs in the tumorigenesis and progression of NPC.

## Material And Methods

### Data collection

Four public datasets of NPC (GSE102349, GSE68799, GSE32960 and GSE12452) were obtained from the Gene Expression Omnibus database (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>). GSE102349, performed on the RNA-Seq platform of GPL11154 (Illumina HiSeq 2000), including 113 NPC tissues, was used to conduct WGCNA. GSE68799, performed on the same platform as GSE102349, consisting of 42 NPC and 4 non-NPC tissues, was used for module preservation analysis [20]. GSE32960, performed on the miRNA profiling platform of GPL14722 (CapitalBio, Inc., Beijing, China), containing 312 paraffin-embedded NPC and 18 paraffin-embedded normal nasopharyngeal tissues, was used to identify differentially expressed miRNAs (DEmiRNAs) between NPC and normal tissues. GSE12452 performed on the mRNA profiling platform of GPL570 [HG-U133\_Plus\_2] (Thermo Fisher Scientific, Inc.), comprising 31 NPC and 10 normal nasopharyngeal tissues, was used to identify differentially expressed genes (DEGs) between NPC and normal tissues. Since these data are downloaded from the public database, the consent of the ethics committee is not required.

### **Weighted correlation network analysis**

WGCNA was performed based on GSE102349 with “WGCNA” package in R software (version 3.5.0; <http://www.r-project.org>) for a weighted correlation network construction. The gene expression levels were measured by log<sub>2</sub>-transformed fragments per kilobase million (FPKM). Only genes with the highest 25% variance of expression values among samples were included in WGCNA. The outliers were identified and removed by hierarchical clustering analysis. Proper soft thresholding power was obtained based on the criterion of approximate scale-free topology. Next, “one-step network construction and module detection” function was used to construct a signed co-expression network and identify modules associated with NPC. The preservation of the modules was examined by GSE68799. The preservation value “medianRank” and “Zsummary.pres” (Z-score) were calculated using the *nPermutations* of 200. In general, if Z-score is lower than 2, the module will be regarded as no evidence of preservation, while Z-score > 10 will be regarded as high preservation. Furthermore, the associations between each module and the clinical traits were analyzed by Pearson’s correlation test.  $P < 0.05$  was considered significant.

### **Differential expression analysis between NPC and normal tissues**

To identify DEmiRNAs and DEGs between NPC and normal tissues, we performed differential expression analysis based on GSE32960 and GSE12452 using “limma” package [21]. Before analysis, miRNA expression values were log<sub>2</sub>-transformed and quantiled normalized across multiple arrays. Gene expression values of GSE12452 were calculated by the method described in our previous study [22]. The DEmiRNAs and DEGs with false discovery rate (FDR) < 0.05 and  $|\log \text{ fold change (FC)}| > 0.5$  were considered significant and used for subsequent analysis.

### **Identification of lncRNAs and Construction of the ceRNA network**

The lncRNAs and mRNAs in the WGCNA modules were defined and annotated based on the Gencode lncRNAs annotation file (gencode.v31.long\_noncoding\_RNAs.gtf.gz) from the GENCODE database (<https://www.gencode-genes.org>). Following, a lncRNA-mediated ceRNA network was constructed based

on the hypothesis that lncRNAs can interact with target miRNAs and invoked miRNA sponges to regulate the activity of mRNAs [11]. According to the hypothesis, both lncRNAs and mRNAs had a negative correlation with the corresponding miRNAs. We speculated that the mRNAs may be the target genes of lncRNAs that belong to the same WGCNA module, so we first obtained up- and down-regulated DElncRNAs and DEmRNAs in NPC by taking the intersection of the DEGs and the genes from the WGCNA modules. Then, we used the validated lncRNA-miRNA interaction data from LncBase v.2 experimental module [18] to predict target miRNAs for DElncRNAs. The target miRNAs were matched with DEmiRNAs to acquire up- and down-regulated miRNAs for the ceRNA network. Subsequently, we used the experimentally verified miRNA-mRNA interaction data from miRTarBase database [19] to predict target mRNAs for those up- and down-regulated miRNAs. After that, the target mRNAs were matched with the DEmRNAs in the WGCNA modules to obtain mRNAs for the ceRNA network. Thus, a ceRNA network based on the WGCNA modules was constructed and visualized by Cytoscape v3.6 software. The workflow of building the ceRNA network was provided in Figure S1.

### **Survival analysis for the lncRNAs, miRNAs and mRNAs in the ceRNA network**

There were 88 NPC samples with progression-free survival (PFS) data in GSE102349. PFS was defined as the time from the date of diagnosis to the date of objective tumor progression or death from any cause [23]. There were 312 NPC samples with prognostic information in the dataset GSE32960. The prognostic information included overall survival (OS), disease-free survival (DFS) and distant metastasis-free survival (DMFS). To explore the relationship between the expression level of lncRNAs, miRNAs and mRNAs in the ceRNA network and survival time, the samples were divided into two groups of high expression and low expression according to the median value. Then, the Kaplan-Meier survival curve combined with a Log-rank test was performed to compare the survival difference between the high- and low-expression groups. Survival analysis was implemented using “survival” and “survminer” packages [24, 25].  $P < 0.05$  was considered to be statistically significant.

### **Functional and pathway enrichment analysis for the genes in the WGCNA modules and ceRNA network**

To explore the potential biological functions for the genes in the WGCNA modules and lncRNA-mediated ceRNA network, the genes were submitted to GO and KEGG pathway enrichment analysis using “clusterProfiler” package [26]. For GO analysis, the cut-off criteria were  $P < 0.05$  and  $q < 0.05$ . For KEGG analysis, the cut-off criterion was  $P < 0.05$ .

### **GSEA for the prognostic lncRNAs in the ceRNA network**

To explore the potential pathways of the prognostic lncRNAs in the ceRNA network, the samples in the GSE102349 and GSE12452 were categorized into two groups of high-expression and low-expression according to the median value of the prognostic lncRNAs. Then, GSEA was performed by “clusterProfiler” package using the annotated gene set, “c2.cp.kegg.v7.0.symbols.gmt” from the Molecular Signature Database (MSigDB, <http://software.broadinstitute.org/gsea/msigdb/index.jsp>) as the reference gene set. False discovery rate (FDR)  $< 0.05$  was chosen as the threshold.

## Construction of risk score model

To construct a risk score model for PFS prediction, the DEGs with false discovery rate (FDR) <0.05 and  $|\log \text{fold change (FC)}| > 1$  in the high preserved WGCNA modules were subjected to lasso penalized Cox regression analysis by “glmnet” package [27, 28] in R. The function “cv.glmnet” was used to compute 10-fold cross-validation for the Cox model. “*Lambda.min*” was select as the optimal  $\lambda$  value. The prognostic genes and their coefficients were generated to calculate the risk score (RS) for each sample. The RS formula was constructed as follows: Risk score =  $\sum (\beta \times expr)$ . Here,  $\beta$  represents the coefficient of a gene, and *expr* represents the expression level of the gene. According to the median value of the RS, 88 samples in GSE102349 were classified into high- and low-risk groups. Then, the Kaplan-Meier survival curve combined with a Log-rank test was performed to evaluate the survival differences between high- and low-risk groups using “survival” and “survminer” packages. The receiver operating characteristic (ROC) curve, the area under the curve (AUC) and the C-index were applied to assess the prognostic accuracy of the risk score model using “survivalROC” package [29]. Besides, the distribution of the risk score, survival status and expression patterns of the prognostic gene signatures were explored in GSE102349. After that, the prognostic genes were subjected to univariate Cox analysis and multivariate Cox analysis.  $P < 0.05$  was considered statistically significant. Using “forestplot” package [30], a forest plot was used to illustrate the results of hazard ratio (HR), 95% CI (confidence interval) and corresponding P-value. Finally, we compared the expression level of the lncRNAs in the ceRNA network between high- and low-risk groups in GSE102349 using two-sided student’ t test.  $P < 0.05$  was considered to be statistically significant.

## Results

### Weighted correlation network analysis

A weighted correlation network of NPC based on GSE102349 was constructed using “WGCNA” package. After removing three samples (GSM2735294, GSM2735309 and GSM2735313) which were considered abnormal after sample clustering analysis, 110 samples were retained for further analysis (Fig. S2). We chose the soft-thresholding power of  $\beta$  as three, which was the lowest power for which the scale-free topology fit index (scale-free  $R^2$ ) achieved 0.9 (Fig. S3). Then, a weighted correlation network with 14 modules was constructed. The sizes of the modules ranged from 52 (cyan module) to 2045 (turquoise module) genes. A hierarchical clustering dendrogram (tree) was generated with the color assignment (Fig. 1A). The genes labeled with “gray” represented background genes that were not assigned to any co-expressive gene modules. Following, the preservation of the modules were examined using GSE68799. Results showed that the 14 modules were stable because their Z-scores were all higher than two (Fig. S4). Among them, ten modules (blue, turquoise, yellow, brown, red, cyan, magenta, black, tan and purple) were high preserved (Z-score > 10). Besides, we calculated the correlations between the 14 modules and the clinical traits (Fig. 1B). The results showed that three modules (cyan, turquoise and black) were significantly correlated with clinical stage, four modules (tan, red, turquoise and magenta) were

significantly correlated with the event, and four modules (brown, red, turquoise and black) were significantly correlated with time to event ( $P < 0.05$ ).

### **Identification of DE miRNAs and DEGs in NPC tissues**

Based on the miRNA profiling dataset GSE32960, we identified a total of 196 DE miRNAs, including 87 (44.4%) upregulated miRNAs and 109 (55.6%) downregulated miRNAs. Meanwhile, based on the mRNA profiling dataset GSE12452, we identified 2371 DEGs, including 1191 (50.2%) upregulated genes and 1180 (49.8%) downregulated genes. The DE miRNAs (Fig. 2A) and DEGs (Fig. 2B) were depicted in volcano plots, respectively.

### **Identification of lncRNAs and Construction of the ceRNA network**

Using GENCODE database, we identified 65 lncRNAs in the WGCNA modules. However, after matching with the DEGs from GSE12452, LncBase v.2 and miRTarBase databases, only 11 lncRNAs were retained to construct the lncRNA-mediated ceRNA network (Table 1). Of them, ZNF667-AS1 was down-regulated in NPC, the others (GAS5, SNHG1, SNHG12, SNHG15, SNHG16, SNHG17, SNHG6, SNHG8, THAP9-AS1 and ZFAS1) were up-regulated in NPC. As demonstrated in Figure 3, we eventually constructed a ceRNA network composed of 11 lncRNAs (Table 1), 15 miRNAs (Table S1) and 40 mRNAs (Table S2) from turquoise and salmon modules.

### **Survival analysis for the lncRNAs, miRNAs and mRNAs in the ceRNA network**

Kaplan-Meier survival analysis was performed to explore the prognostic value of the lncRNAs, miRNAs and mRNAs in the ceRNA network. Results showed that four mRNAs, including LDHA, LMNB2, TPI1 and UNG, were negatively correlated with PFS, and lncRNA ZNF667-AS1 was positively correlated with PFS. Patients in the high-expression group of ZNF667-AS1 had longer PFS than those in the low-expression group (Fig. 4A). Additionally, hsa-miR-142-3p was found to be positively correlated with OS, DFS and DMFS (Fig. 4B).

### **Functional and pathway enrichment analysis for the genes in the WGCNA modules**

The potential biological function of the genes in the WGCNA modules was investigated by GO and KEGG pathway enrichment analysis. Results indicated that the genes in the modules were associated with different functional categories and pathways (Fig. S5A-B). For example, "Wnt signaling pathway" was enriched in the black module; "Small cell lung cancer" was enriched in the brown module; "p53 signaling pathway" was enriched in the brown and tan modules; "PI3K-Akt signaling pathway" was enriched in the brown, turquoise and yellow modules; "Epstein-Barr virus infection" was enriched in the cyan and turquoise modules; "IL-17 signaling pathway" was enriched in the green modules; "Amoebiasis" was enriched in the purple, turquoise and yellow modules; "Cell cycle", "Oocyte meiosis", "DNA replication" were enriched in the tan module; "Protein digestion and absorption" was enriched in the yellow module.

### **Functional and pathway enrichment analysis for the genes in the ceRNA network**

Through GO and KEGG pathway enrichment analysis, we also studied the potential biological functions of the genes in the ceRNA network. Results showed that the genes were significantly enriched in 76 GO terms for biological processes (BP), two GO terms for molecular function (MF) and one GO terms for cellular components (CC) (Table S3). For BP, enriched terms included “nucleocytoplasmic transport”, “nuclear transport”, “regulation of organ morphogenesis”, “regulation of morphogenesis of an epithelium” and “sulfur amino acid metabolic process”. For CC, the enriched term was “nuclear periphery”. For MF, the enriched terms included “ribosomal small subunit binding” and “histone kinase activity”. Additionally, KEGG pathway enrichment analysis indicated that “Wnt signaling pathway” and “HIF-1 signaling pathway” were significantly enriched. We displayed the results as 4 networks that depict the representative GO terms and KEGG pathways (Fig. 5).

### **GSEA for ZNF667-AS1**

As shown in the above analysis, of the 11 lncRNAs, only ZNF667-AS1 was correlated with PFS significantly. To better understand the role of ZNF667-AS1 in NPC, GSEA was performed on GSE102349 and GSE12452. The results showed that 58 pathways were enriched in GSE102349 (Table S4), and 54 pathways were enriched in GSE12452 (Table S5). Further analysis revealed that these two data sets shared 44 pathways. Figure 6 showed the top five activated pathways and five suppressed pathways when ZNF667-AS1 was up-regulated and down-regulated. The up-regulation of ZNF667-AS1 was associated with “intestinal immune network for IgA production”, “hematopoietic\_cell\_lineage”, “B cell receptor signaling pathway”, “T cell receptor signaling pathway”, “leukocyte transendothelial migration” and other pathways. The down-regulation of ZNF667-AS1 was related to “DNA replication”, “cell cycle”, “P53 signaling pathway”, “spliceosome” and other pathways.

### **Construction of risk score model**

Lasso penalized Cox regression analysis along with 10-fold cross-validation was conducted to build risk score model for PFS prediction (Fig. 7A-B), which resulted in the identification of 12 prognostic genes: RRM2, VILL, MANSC1, CYP4B1, LXN, MLF1, CRIP1, WDR54, MNS1, CNN3, CTHRC1 and NFE2L3. Risk score for each sample was calculated based on the following formula:  $RS = 0.07785695 \times \text{expr}_{RRM2} - 0.05729219 \times \text{expr}_{VILL} + 0.15392701 \times \text{expr}_{MANSC1} - 0.02185837 \times \text{expr}_{CYP4B1} - 0.05553768 \times \text{expr}_{LXN} + 0.19470422 \times \text{expr}_{MLF1} - 0.44565573 \times \text{expr}_{CRIP1} + 0.22242843 \times \text{expr}_{WDR54} + 0.39891041 \times \text{expr}_{MNS1} - 0.27436890 \times \text{expr}_{CNN3} - 0.03554977 \times \text{expr}_{CTHRC1} + 0.08434817 \times \text{expr}_{NFE2L3}$ . According to the median value of RSs, 88 samples in the dataset GSE102349 were separated into high- and low-risk groups. Kaplan-Meier analysis showed that patients in the high-risk group had poorer PFS than those in the low-risk group (Fig. 7C). Time-dependent ROC analysis at varying follow-up times was used to explore the prognostic capacity of the risk score model for PFS. Results showed that the AUC received 0.932, 0.943 and 0.967 at 12-, 24- and 36-months (Fig. 7D). The C-index for the model was 0.891. Therefore, this model has good PFS prediction ability in GSE102349. The distribution of the risk score indicated patients in the high-risk group had a worse outcome of PFS than those in the low-risk group (Fig. 7E-F), and the former tended to have overexpression of RRM2, MANSC1, MLF1, WDR54, MNS1 and NFE2L3, whereas

the latter tended to have overexpression of VILL, CYP4B1, LXN, CRIP1, CNN3 and CTHRC1 (Fig. 7G). Besides, multivariate Cox regression analysis indicated that MANSC1, CYP4B1, MLF1, CRIP1, MNS1, CNN3 and CTHRC1 were independent prognostic factors associated with the DFS in NPC (Fig. 7H). Among the 11 lncRNAs in the ceRNA network, SNHG16, SNHG17 and THAP9-AS1 were up-regulated in the high-risk group of NPC, while ZNF667-AS1 was down-regulated in the high-risk group of NPC (Fig. 8).

## Discussion

lncRNAs, function as ceRNAs by binding miRNAs, have been reported to be involved in the physiological and pathological processes of various diseases. One of the most effective methods to predict the function of ceRNAs was to construct a ceRNA network based on the high-throughput data together with bioinformatic tools and computational approaches [31]. However, the comprehensive analysis of lncRNA-mediated ceRNA regulatory network in NPC remains scarce.

In the present study, we conducted WGCNA using RNA-Seq data of GSE102349 to enrich modules associated with NPC. Finally, we built a weighted correlation network composed of 14 modules. Through enrichment analysis, we found that the modules were related to some terms and pathways that had been previously reported by us or other researchers, such as “Wnt signaling pathway”, “Small cell lung cancer”, “PI3K–Akt signaling pathway”, “Epstein–Barr virus infection”, “Cell cycle”, and “DNA replication” [17, 22, 32].

To better understand the potential roles of lncRNAs in NPC, we identified lncRNAs from the WGCNA modules. Combining with mRNA expression profile data, we identified the up- and down-regulated lncRNAs and mRNAs in the same module. The mRNAs in the same module may be the regulatory targets for lncRNAs, so we used the lncRNAs and mRNAs in the same module to construct a ceRNA network. To improve the reliability of the ceRNA network, we identified the DE miRNAs from GSE32960. Following, LncBase v.2 database was utilized to predict target miRNAs for lncRNAs, and miRTarBase was utilized to predict target mRNAs for miRNAs. Finally, we developed a ceRNA network composed of 11 lncRNAs, 15 miRNAs and 40 mRNAs from turquoise and salmon modules. It is noteworthy that the genes in the ceRNA network were significantly enriched in the “Wnt signaling pathway” and “HIF-1 signaling pathway” which had been shown to be related to the development of NPC. For instance, Wang et al. [33] revealed that Wnt/ $\beta$ -catenin signaling (including  $\beta$ -catenin, cyclin D1, c-Myc, and MMP-7) and p-eIF4E expression were elevated in NPC compared with non-cancerous nasopharyngeal epithelial tissues and associated with clinical characteristics of NPC patients. Sung et al. [34] showed that HIF-1 alpha, HIF-2 alpha, CA IX, and VEGF were frequently coexpressed in NPC biopsies, and associated with poor outcomes after radiotherapy.

Among the 40 mRNAs in the ceRNAs, four mRNAs were found to have a significant impact on the prognosis of NPC, including LDHA, LMNB2, TPI1 and UNG. High expression of LDHA, LMNB2, TPI1 and UNG indicated unfavorable outcomes in patients with NPC. LDHA is up-regulated in NPC tissues and cells, and it was reported to be an independent adverse prognostic factor of NPC [35, 36]. Through

inhibition of LDHA, miR-34b-3 and miR-449a suppressed NPC progression and metastasis [37]. Therefore, the ceRNA network was reliable because the regulatory relationship between LDHA and miR-449a was indicated in our ceRNA network. LMNB2, a B type nuclear lamin, binds to the C-terminus of MCM7 and competes with the binding of the tumor suppressor RB protein, thus regulates human non-small cell lung cancer progression [38]. TPI1, an enzyme that catalyzes the interconversion of DHAP and G3P in glycolysis and gluconeogenesis, might be a novel prognostic factor to evaluate gastric cancer patients' survival [39, 40]. UNG is a critical mediator of pemetrexed sensitivity in lung cancer [41]. The role and mechanism of LMNB2, TPI1 and UNG in NPC remain elusive.

Among the 15 miRNAs in the ceRNAs, hsa-miR-142-3p was demonstrated to be down-regulated in NPC, and a high level of hsa-miR-142-3p indicated favorable prognosis, which was consistent with the previous studies [16, 42]. Li Y et al. revealed that miR-142-3p was epigenetically silenced by EZH2-recruited DNMT1 and suppress NPC cell metastasis and EMT through targeting ZEB2 [43]. However, Qi et al. reported that miR-142-3p inhibits the expression of SOCS6 and promotes cell proliferation in NPC [44]. Therefore, the role of miR-142-3p in the progression of NPC deserves further studied.

Among the 11 lncRNAs in the ceRNAs, ZNF667-AS1 was found to be related to the PFS of patients with NPC. ZNF667-AS1, also known as MORT, is located in 19q13.43. Dysregulated expression of ZNF667-AS1 has been reported in many tumors, including breast cancer, cervical cancer, laryngeal squamous cell carcinoma and esophageal squamous cell carcinoma [45-48]. Li et al. found that ZNF667-AS1 reduces tumor invasion and metastasis in cervical cancer by counteracting microRNA-93-3p-dependent PEG3 downregulation [46]. Dong et al. demonstrated that aberrant hypermethylation-mediated downregulation of ZNF667-AS1 and ZNF667 correlate with progression and prognosis of esophageal squamous cell carcinoma [47]. Further, ZNF667-AS1 was revealed to be silenced by aberrant DNA methylation in 22 of 33 of TCGA cancer types [49]. In this study, ZNF667-AS1 was shown to be down-regulated in NPC compared with normal tissues, and it was down-regulated in the high-risk patients with poor prognosis. The expression level of ZNF667-AS1 was positively correlated with PFS of NPC. Combined with the above literature, we speculated that the down-regulation of ZNF667-AS1 might also be caused by methylation in NPC, but it needs further molecular experiments to prove this hypothesis. Besides, we found that ZNF667-AS1 may regulate the expression of PRKCB and PAX5 through competitively binding to hsa-miR-574-5p. What's more, GSEA analysis showed that ZNF667-AS1 was associated with some important pathways related to tumorigenesis. Therefore, ZNF667-AS1 might act as a tumor suppressor and participate in the occurrence and development of NPC. To the best of our knowledge, the present study was the first to demonstrate the relation between ZNF667-AS1 and NPC using bioinformatics analysis.

In the present study, we constructed a risk score model for predicting PFS of NPC patients. This model consists of 12 genes RRM2, VILL, MANSC1, CYP4B1, LXN, MLF1, CRIP1, WDR54, MNS1, CNN3, CTHRC1 and NFE2L3. Among them, RRM2, MANSC1, MLF1, WDR54, MNS1 and NFE2L3 were associated with high-risk of poor prognosis, while VILL, CYP4B1, LXN, CRIP1, CNN3 and CTHRC1 were associated with low-risk of poor prognosis. Multivariate Cox regression analysis indicated that MANSC1, CYP4B1, MLF1, CRIP1, MNS1, CNN3 and CTHRC1 were independent prognostic factors associated with the DFS in NPC.

Kaplan-Meier survival analysis, ROC curve and the C-index demonstrated that the predictive capability of the risk model was successfully in GSE102349. Among the prognostic genes in the model, RRM2 encodes a subunit of ribonucleotide reductase, which catalyzes the conversion of ribonucleotides into deoxyribonucleotides. Overexpression of RRM2 predicts an unfavorable prognosis for patients with NPC [50]. CYP4B1 encodes a member of the cytochrome P450 superfamily of enzymes, which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. Although CYP4B1 was detected in NPC tissues[51], the role of CYP4B1 is little known. In addition to RRM2 and CYP4B1, the relationship between NPC and other prognostic genes in the risk model has not been reported in the literature.

What's more, we analyzed the expression differences of lncRNAs in the ceRNA network between two risk groups. Results showed that SNHG16, SNHG17 and THAP9-AS1 were up-regulated in the high-risk group of NPC, while ZNF667-AS1 was down-regulated in the high-risk group of NPC. SNHG16 was regarded as an oncogene and associated with neuroblastoma, bladder cancer, colorectal cancer, esophageal squamous cell carcinoma, hepatocellular carcinoma [52-54]. Zhang et al. suggested that SNHG16 promotes tumor progression through acting as an endogenous "sponge" by competing with miR-140-5p, thereby regulating target ZEB1 [53]. SNHG17 was reported to promote gastric cancer progression by epigenetically silencing of p15 and p57, and it was an unfavorable prognostic factor in colorectal cancer and gastric cancer [55, 56]. Extensive overexpression of THAP9-AS1 was observed in pancreatic ductal adenocarcinoma which is associated with poor clinical outcomes [57]. Function as ceRNA, THAP9-AS1 facilitated YAP expression by sequestering miR-484, and it can bind YAP to inhibit phosphorylation-mediated inactivation by LATS1 [57]. The function and mechanism of those identified lncRNAs in NPC warrant further study.

Altogether, based on the WGCNA method and ceRNAs hypothesis, combined with bioinformatics database, this study constructed a lncRNA-mediated ceRNA network, which is of great significance for us to identify lncRNAs with important biological significance in NPC, especially the lncRNAs relevance to prognosis and risk, and to understand the regulatory mechanism and function of lncRNAs. The lncRNAs, miRNAs, and mRNAs identified in this study may serve as candidate diagnostic and therapeutic targets for NPC, and provide clues for further in-depth research in the future.

It should be acknowledged that there were certain limitations to the present study. Firstly, only 11 differential expressed lncRNAs were obtained from the integrated analysis of RNA-seq data and mRNA profiling data. Some lncRNAs may be ignored, for only a portion of the possible lncRNAs was included in HG-U133 Plus 2.0 platform. Secondly, the sample size of GSE102349 was relatively limited. There were only 113 NPC samples included in GSE102349, and only 88 samples had survival information, thus research with large sample sizes, high-quality high-throughput data are required to verify our findings in the future. Thirdly, due to the lack of important clinical information such as gender and age, we can't confirm the independence of the risk model. Moreover, due to the absence of similar NPC public datasets with survival information, external validation was not conducted. Finally, although we have constructed a ceRNA network using multiple datasets and experimental-based databases, the regulatory relationship

among lncRNAs, miRNAs and mRNAs still require to verify using molecular experiments in vivo and in vitro.

## Conclusions

Collectively, we constructed a ceRNA network based on WGCNA. The network can reflect the mutual regulated relationships among lncRNAs, miRNAs and mRNAs, and it can reflect the changes in their expression level between NPC and normal tissues. This ceRNA network provided novel insights into the understanding of the crosstalk among lncRNAs, miRNAs and mRNAs in the tumorigenesis and progression of NPC. Moreover, we developed a LASSO penalized Cox regression model for NPC that may contribute to predict PFS of NPC patients. Furthermore, we identified several potential prognostic biomarkers and high-risk biomarkers from the ceRNA network. Further evaluation of those biomarkers is needed by well-designed scientific experiments in the future.

## Abbreviations

NPC: nasopharyngeal carcinoma; lncRNA: long noncoding RNA; miRNA: microRNA; ceRNA: competing endogenous RNA; WGCNA: Weighted correlation network analysis; PFS: progression-free survival; MREs: miRNA recognition elements; DEGs: differentially expressed genes; DE miRNAs: differentially expressed miRNAs; DE lncRNAs: differentially expressed lncRNAs;

DE mRNAs: differentially expressed mRNAs; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: Gene set enrichment analysis; GEO: Gene Expression Omnibus;

FPKM: fragments per kilobase million; FDR: false discovery rate; FC: fold change; RS: risk score;

ROC: receiver operating characteristic; AUC: area under the curve; CI: confidence interval; HR: hazard ratio; OS: overall survival; DFS: disease-free survival; DMFS: distant metastasis-free survival.

## Declarations

## Data Availability

The data used for analysis in this study are available from the Gene Expression Omnibus database.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Funding Statement

This study was supported by Scientific and Technological Project of Zhanjiang (grant no. 2010C3104008), Medical Science and Technology Research Project of Guangdong (grant no. A2019527) and Doctoral Research Start-up Fund of Guangdong Medical University (grant no. B2019025).

## Acknowledgments

We thank Research Square Preprint Platform for sharing our work at <https://www.researchsquare.com/article/rs-35735/v1>.

## References

- [1] Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2006;15:1765-77.
- [2] Chua MLK, Wee JTS, Hui EP, Chan ATC. Nasopharyngeal carcinoma. *The Lancet* 2016;387:1012-24.
- [3] 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.
- [4] Cheng SH, Jer-Min Jian J, Tsai SYC, Yen KL, Chu N-M, Chan K-Y, et al. Long-term survival of nasopharyngeal carcinoma following concomitant radiotherapy and chemotherapy. *International Journal of Radiation Oncology, Biology, Physics* 2000;48:1323-30.
- [5] Nakanishi Y, Wakisaka N, Kondo S, Endo K, Sugimoto H, Hatano M, et al. Progression of understanding for the role of Epstein-Barr virus and management of nasopharyngeal carcinoma. *Cancer and Metastasis Reviews* 2017;36:435-47.
- [6] Schmitz SU, Grote P, Herrmann BG. Mechanisms of long noncoding RNA function in development and disease. *Cell Mol Life Sci* 2016;73:2491-509.
- [7] Xudong L, Yan L, Xi Y, Xiaoguang W, Xiaoguang H. Long noncoding RNA H19 regulates EZH2 expression by interacting with miR-630 and promotes cell invasion in nasopharyngeal carcinoma. *Biochem Biophys Res Commun* 2016.
- [8] Gao C, Lu W, Lou W, Wang L, Xu Q. Long noncoding RNA HOXC13-AS positively affects cell proliferation and invasion in nasopharyngeal carcinoma via modulating miR-383-3p/HMGA2 axis. *J Cell Physiol* 2019;234:12809-20.
- [9] Zheng Z-Q, Li Z-X, Zhou G-Q, Lin L, Zhang L-L, Lv J-W, et al. Long Noncoding RNA FAM225A Promotes Nasopharyngeal Carcinoma Tumorigenesis and Metastasis by Acting as ceRNA to Sponge miR-590-3p/miR-1275 and Upregulate ITGB3. *Cancer Res* 2019;79:4612-26.

- [10] Lan X, Liu X. LncRNA SNHG1 functions as a ceRNA to antagonize the effect of miR-145a-5p on the down-regulation of NUA1 in nasopharyngeal carcinoma cell. *J Cell Mol Med* 2019;23:2351-61.
- [11] Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;146:353-8.
- [12] Karreth FA, Pandolfi PP. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov* 2013;3:1113-21.
- [13] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 2008;9:559.
- [14] Liu R, Zhang W, Liu ZQ, Zhou HH. Associating transcriptional modules with colon cancer survival through weighted gene co-expression network analysis. *BMC genomics* 2017;18:361.
- [15] Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Experimental eye research* 2018;166:13-20.
- [16] Zhang S, Yue W, Xie Y, Liu L, Li S, Dang W, et al. The four-microRNA signature identified by bioinformatics analysis predicts the prognosis of nasopharyngeal carcinoma patients. *Oncol Rep* 2019;42:1767-80.
- [17] Ge Y, He Z, Xiang Y, Wang D, Yang Y, Qiu J, et al. The identification of key genes in nasopharyngeal carcinoma by bioinformatics analysis of high-throughput data. *Mol Biol Rep* 2019;46:2829-40.
- [18] Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, et al. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res* 2016;44:D231-8.
- [19] Chou C-H, Shrestha S, Yang C-D, Chang N-W, Lin Y-L, Liao K-W, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res* 2017;46:D296-D302.
- [20] Langfelder P, Luo R, Oldham MC, Horvath S. Is My Network Module Preserved and Reproducible? *PLoS computational biology* 2011;7.
- [21] 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 research* 2015;43:e47.
- [22] Zou Z, Gan S, Liu S, Li R, Huang J. Investigation of differentially expressed genes in nasopharyngeal carcinoma by integrated bioinformatics analysis. *Oncol Lett* 2019;18:916-26.
- [23] Zhang L, Maclsaac KD, Zhou T, Huang PY, Xin C, Dobson JR, et al. Genomic Analysis of Nasopharyngeal Carcinoma Reveals TME-Based Subtypes. *Molecular cancer research : MCR* 2017;15:1722-32.

- [24] Therneau T. A Package for Survival Analysis in S. version 2.38. 2015.
- [25] Kassambara A, Kosinski M. survminer: Drawing survival curves using “ggplot2,” 2018. URL <https://CRAN.R-project.org/package=survminer> R package version 04 2018;3.
- [26] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* 2012;16:284-7.
- [27] TIBSHIRANI R. THE LASSO METHOD FOR VARIABLE SELECTION IN THE COX MODEL. *Statistics in Medicine* 1997;16:385-95.
- [28] Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33:1-22.
- [29] Saha-Chaudhuri P, Hapb P. survivalROC: Time-dependent ROC curve estimation from censored survival data. R Package Version 2013;1.
- [30] Gordon M, Lumley T. Forestplot: advanced forest plot using ‘grid’graphics. 2017. URL <https://CRAN.R-project.org/package=forestplot> R package version 2017;1:7.
- [31] Guo L-L, Song C-H, Wang P, Dai L-P, Zhang J-Y, Wang K-J. Competing endogenous RNA networks and gastric cancer. *World J Gastroenterol* 2015;21:11680-7.
- [32] Chen J, Yang R, Zhang W, Wang Y. Candidate pathways and genes for nasopharyngeal carcinoma based on bioinformatics study. *International Journal of Clinical and Experimental Pathology* 2015;8:2026-32.
- [33] Wang W, Wen Q, Luo J, Chu S, Chen L, Xu L, et al. Suppression Of  $\beta$ -catenin Nuclear Translocation By CGP57380 Decelerates Poor Progression And Potentiates Radiation-Induced Apoptosis in Nasopharyngeal Carcinoma. *Theranostics* 2017;7:2134-49.
- [34] Sung FL, Hui EP, Tao Q, Li H, Tsui NB, Lo YM, et al. Genome-wide expression analysis using microarray identified complex signaling pathways modulated by hypoxia in nasopharyngeal carcinoma. *Cancer Lett* 2007;253:74-88.
- [35] Li AC, Xiao WW, Wang L, Shen GZ, Xu AA, Cao YQ, et al. Risk factors and prediction-score model for distant metastasis in nasopharyngeal carcinoma treated with intensity-modulated radiotherapy. *Tumour Biol* 2015;36:8349-57.
- [36] Su Y, Yu QH, Wang XY, Yu LP, Wang ZF, Cao YC, et al. JMJD2A promotes the Warburg effect and nasopharyngeal carcinoma progression by transactivating LDHA expression. *BMC Cancer* 2017;17:477.
- [37] Li H, Li X, Ge X, Jia L, Zhang Z, Fang R, et al. MiR-34b-3 and miR-449a inhibit malignant progression of nasopharyngeal carcinoma by targeting lactate dehydrogenase A. *Oncotarget* 2016.

- [38] Ma Y, Fei L, Zhang M, Zhang W, Liu X, Wang C, et al. Lamin B2 binding to minichromosome maintenance complex component 7 promotes non-small cell lung carcinogenesis. *Oncotarget* 2017;8:104813-30.
- [39] Maquat LE, Chilcote R, Ryan PM. Human triosephosphate isomerase cDNA and protein structure. Studies of triosephosphate isomerase deficiency in man. *J Biol Chem* 1985;260:3748-53.
- [40] Chen T, Huang Z, Tian Y, Lin B, He R, Wang H, et al. Clinical significance and prognostic value of Triosephosphate isomerase expression in gastric cancer. *Medicine (Baltimore)* 2017;96:e6865.
- [41] Weeks LD, Fu P, Gerson SL. Uracil-DNA glycosylase expression determines human lung cancer cell sensitivity to pemetrexed. *Molecular cancer therapeutics* 2013;12:2248-60.
- [42] 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 Oncology* 2012;13:633-41.
- [43] Li Y, He Q, Wen X, Hong X, Yang X, Tang X, et al. EZH2-DNMT1-mediated epigenetic silencing of miR-142-3p promotes metastasis through targeting ZEB2 in nasopharyngeal carcinoma. *Cell death and differentiation* 2019;26:1089-106.
- [44] Qi X, Li J, Zhou C, Lv C, Tian M. MiR-142-3p Suppresses SOCS6 Expression and Promotes Cell Proliferation in Nasopharyngeal Carcinoma. *Cell Physiol Biochem* 2015;36:1743-52.
- [45] Vrba L, Futscher BW. Epigenetic Silencing of MORT Is an Early Event in Cancer and Is Associated with Luminal, Receptor Positive Breast Tumor Subtypes. *J Breast Cancer* 2017;20:198-202.
- [46] Li Y-J, Yang Z, Wang Y-Y, Wang Y. Long noncoding RNA ZNF667-AS1 reduces tumor invasion and metastasis in cervical cancer by counteracting microRNA-93-3p-dependent PEG3 downregulation. *Molecular oncology* 2019;13:2375-92.
- [47] Dong Z, Li S, Wu X, Niu Y, Liang X, Yang L, et al. Aberrant hypermethylation-mediated downregulation of antisense lncRNA ZNF667-AS1 and its sense gene ZNF667 correlate with progression and prognosis of esophageal squamous cell carcinoma. *Cell Death Dis* 2019. p. 930.
- [48] Meng W, Cui W, Zhao L, Chi W, Cao H, Wang B. Aberrant methylation and downregulation of ZNF667-AS1 and ZNF667 promote the malignant progression of laryngeal squamous cell carcinoma. *J Biomed Sci* 2019. p. 13.
- [49] Vrba L, Futscher BW. Epigenetic silencing of lncRNA MORT in 16 TCGA cancer types. *F1000Research* 2018. p. 211.
- [50] Han P, Lin Z-R, Xu L-H, Zhong Q, Zhu X-F, Liang F-Y, et al. Ribonucleotide reductase M2 subunit expression and prognostic value in nasopharyngeal carcinoma. *Mol Med Rep* 2015;12:401-9.

- [51] Jiang JH, Jia WH, Qin HD, Liang H, Pan ZG, Zeng YX. Expression of cytochrome P450 enzymes in human nasopharyngeal carcinoma and non-cancerous nasopharynx tissue. *Ai Zheng* 2004;23:672-7.
- [52] Christensen LL, True K, Hamilton MP, Nielsen MM, Damas ND, Damgaard CK, et al. SNHG16 is regulated by the Wnt pathway in colorectal cancer and affects genes involved in lipid metabolism. *Molecular oncology* 2016;10:1266-82.
- [53] Zhang K, Chen J, Song H, Chen L-B. SNHG16/miR-140-5p axis promotes esophagus cancer cell proliferation, migration and EMT formation through regulating ZEB1. *Oncotarget* 2017;9:1028-40.
- [54] Chen H, Li M, Huang P. LncRNA SNHG16 Promotes Hepatocellular Carcinoma Proliferation, Migration and Invasion by Regulating miR-186 Expression. *J Cancer* 2019;10:3571-81.
- [55] Ma Z, Gu S, Song M, Yan C, Hui B, Ji H, et al. Long non-coding RNA SNHG17 is an unfavourable prognostic factor and promotes cell proliferation by epigenetically silencing P57 in colorectal cancer. *Molecular bioSystems* 2017;13:2350-61.
- [56] Chen LL, He J, Qiu XT, Yu J, Wang ZM. The prognostic roles of long non-coding RNA SNHG17 in the patients with gastric cancer. *European review for medical and pharmacological sciences* 2019;23:1063-8.
- [57] Li N, Yang G, Luo L, Ling L, Wang X, Shi L, et al. LncRNA THAP9-AS1 promotes pancreatic ductal adenocarcinoma growth and leads to a poor clinical outcome via sponging miR-484 and interacting with YAP. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2019:clincanres.0674.2019.

## Tables

Table 1: The differential expressed lncRNAs in the ceRNA network

Module Color	Symbol	logFC	adj.P.Val	Group	Target miRNA
salmon	ZFAS1	0.804750464	2.10E-04	up	hsa-miR-34c-5p; hsa-miR-421; hsa-miR-449a
salmon	SNHG12	0.710805371	0.007409834	up	hsa-miR-34c-5p; hsa-miR-449a
salmon	GAS5	0.902974269	1.73E-05	up	hsa-miR-34c-5p; hsa-miR-449a
salmon	SNHG6	0.615722821	0.003071714	up	hsa-miR-421
salmon	SNHG1	1.067332781	2.98E-07	up	hsa-miR-125a-5p; hsa-miR-34c-5p; hsa-miR-421; hsa-miR-449a
salmon	SNHG8	1.065333272	1.10E-05	up	hsa-miR-125a-5p
turquoise	SNHG16	0.722248011	9.24E-04	up	hsa-miR-140-3p; hsa-miR-142-3p; hsa-miR-342-3p; hsa-miR-361-3p; hsa-miR-449a; hsa-miR-564
turquoise	SNHG17	0.627392375	0.001876364	up	hsa-miR-125a-5p; hsa-miR-140-5p; hsa-miR-28-5p; hsa-miR-342-3p
turquoise	SNHG15	0.668944147	1.22E-04	up	hsa-miR-125a-5p; hsa-miR-28-5p; hsa-miR-449a
turquoise	THAP9-AS1	0.759104166	0.003243755	up	hsa-miR-142-5p; hsa-miR-146b-5p; hsa-miR-28-5p; hsa-miR-361-5p
turquoise	ZNF667-AS1	-1.117518496	5.97E-07	down	hsa-miR-574-5p

**Abbreviation:** logFC, log<sub>2</sub>-fold change; adj.P.Val, p-value adjusted for multiple testing.

## Figures

Figure 1

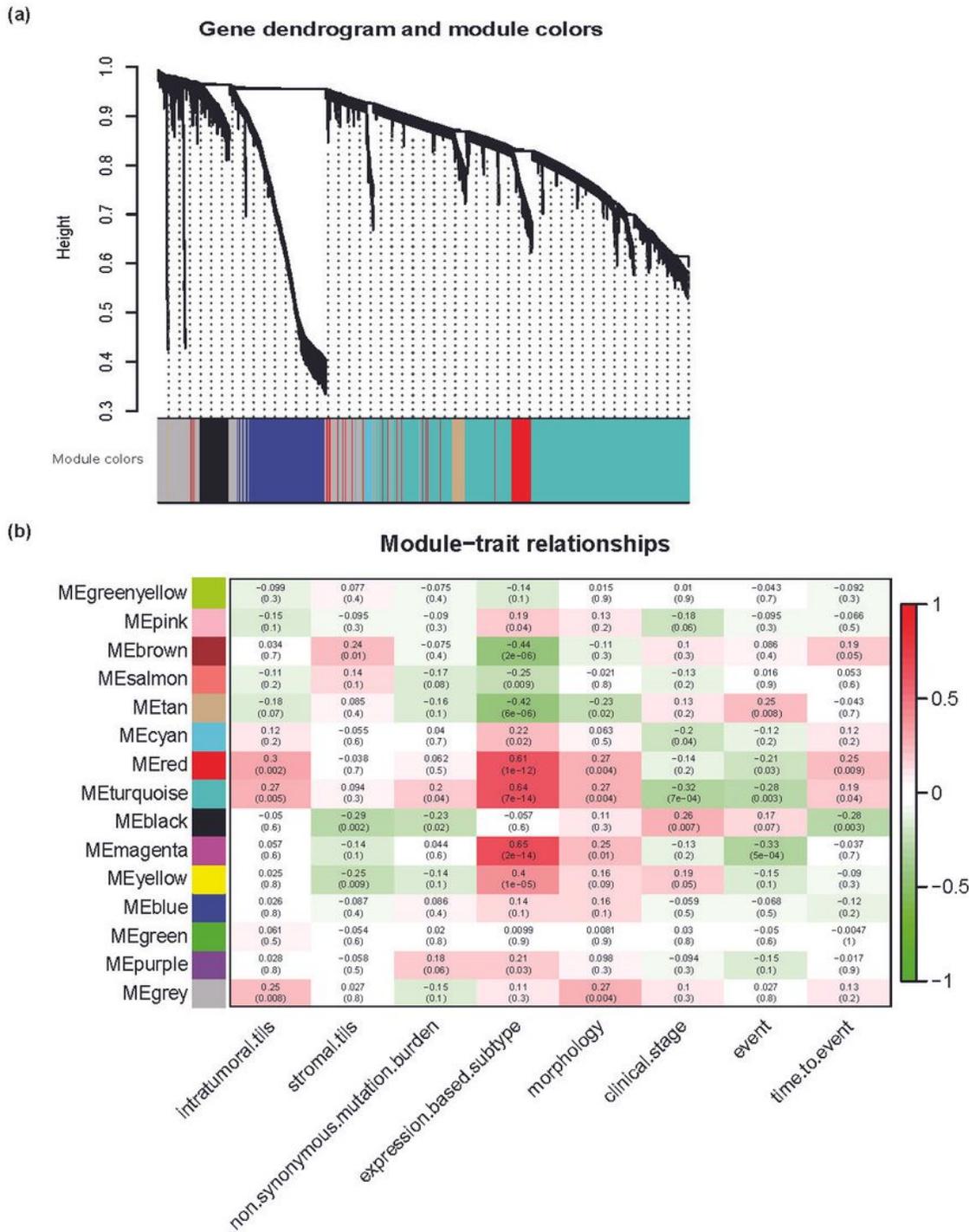


Figure 1

Weighted gene co-expression network analysis. (A) Hierarchical clustering dendrogram of the 14 modules. The branches of the tree represent the clusters of genes. Colors below the tree display the gene modules that correspond to the clusters. (B) Heatmap of the correlation between module eigengenes and clinical traits of NPC. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and p-value. NPC: nasopharyngeal carcinoma

Figure 2

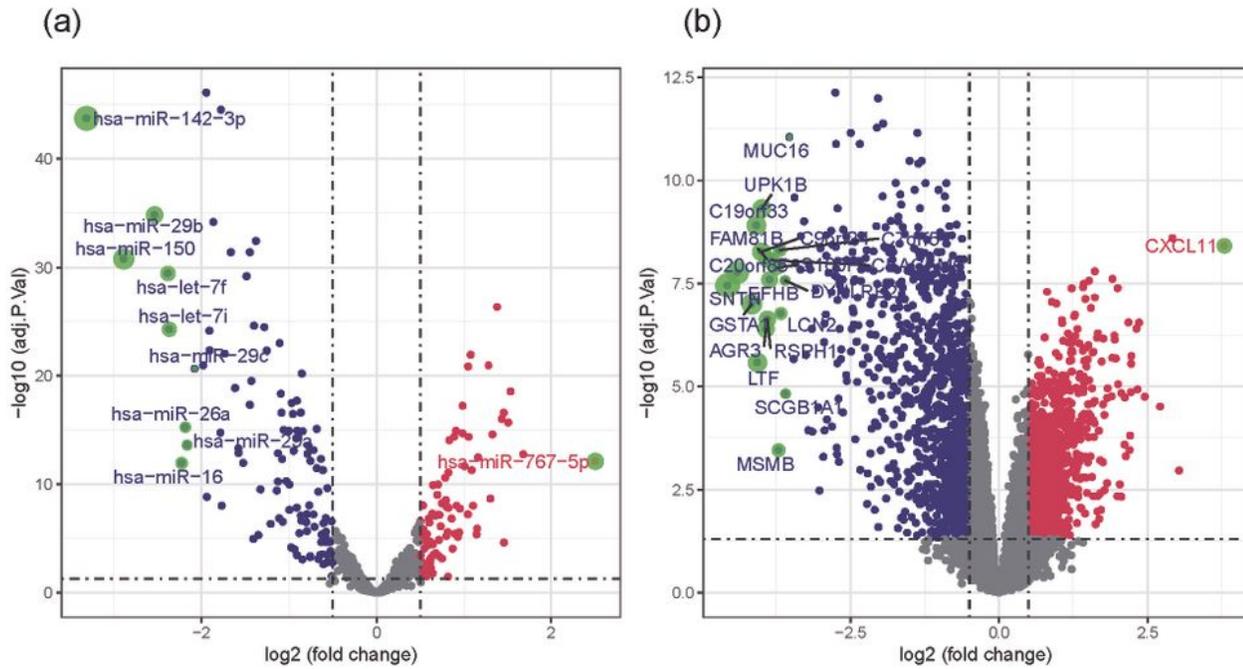


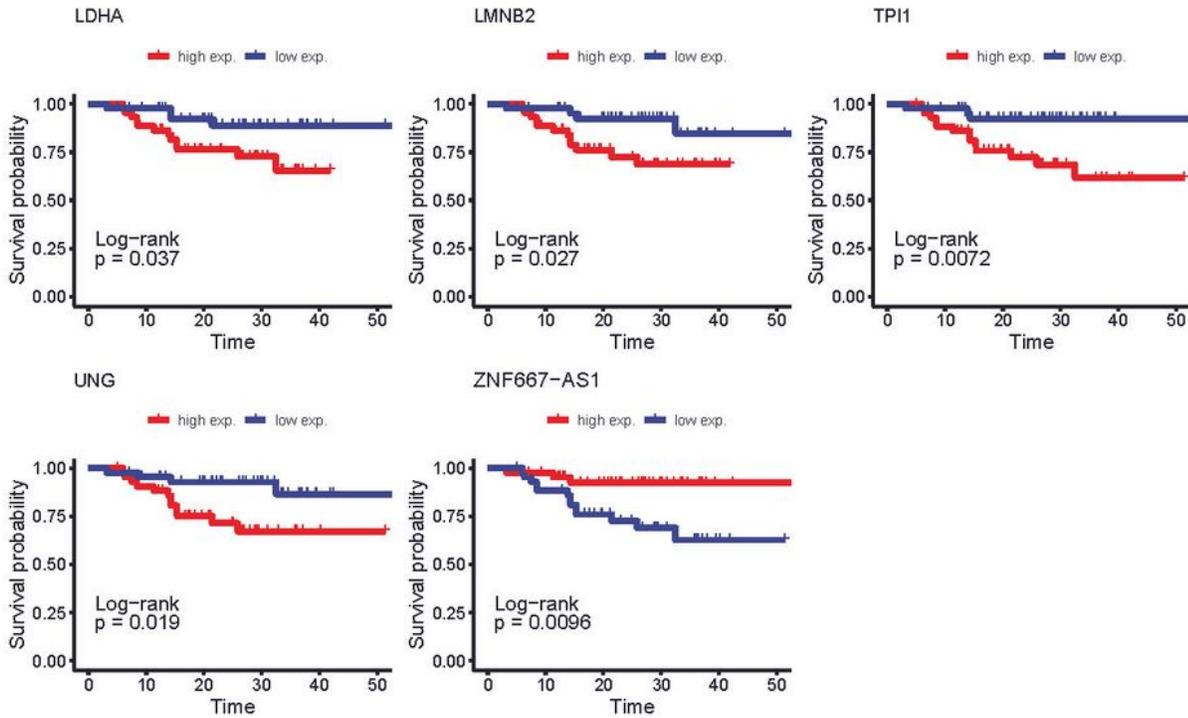
Figure 2

Volcano plots of DE miRNAs (A) and DEGs (B). Spots on the right represent up-regulated genes, and spots on the left represent down-regulated genes. False discovery rate (FDR)  $< 0.05$  and  $|\log$  fold change (FC)|  $> 0.5$  were considered significant. DE miRNAs: differentially expressed miRNAs; DEGs: differentially expressed genes

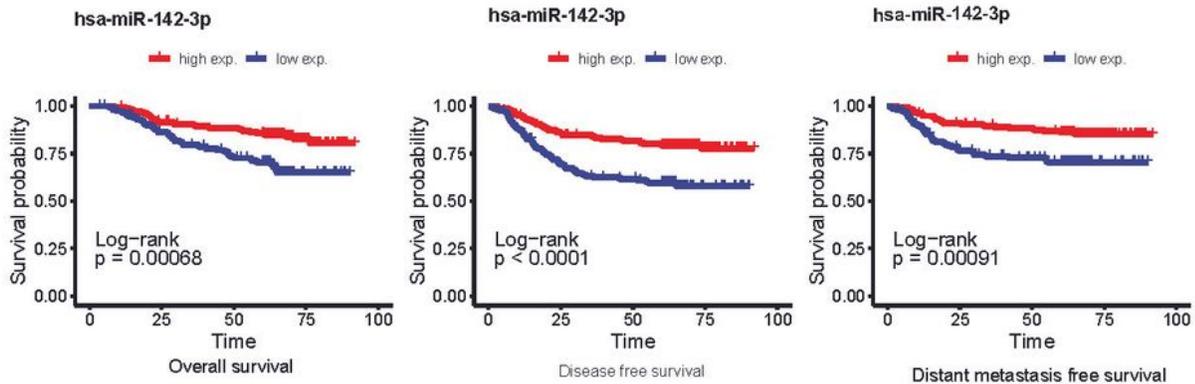


**Figure 4**

(a)



(b)



**Figure 4**

Survival analysis for the lncRNAs, miRNAs and mRNAs in the ceRNA network. LDHA, LMNB2, TPI1 and UNG and ZNF667-AS1 were significantly correlated with PFS of patients with NPC (A). Hsa-miR-142-3p were significantly correlated with OS, DFS and DMFS of patients with NPC (B). NPC: nasopharyngeal carcinoma; PFS: progression-free survival; OS: overall survival; DFS: disease-free survival; DMFS: distant metastasis-free survival

Figure 5

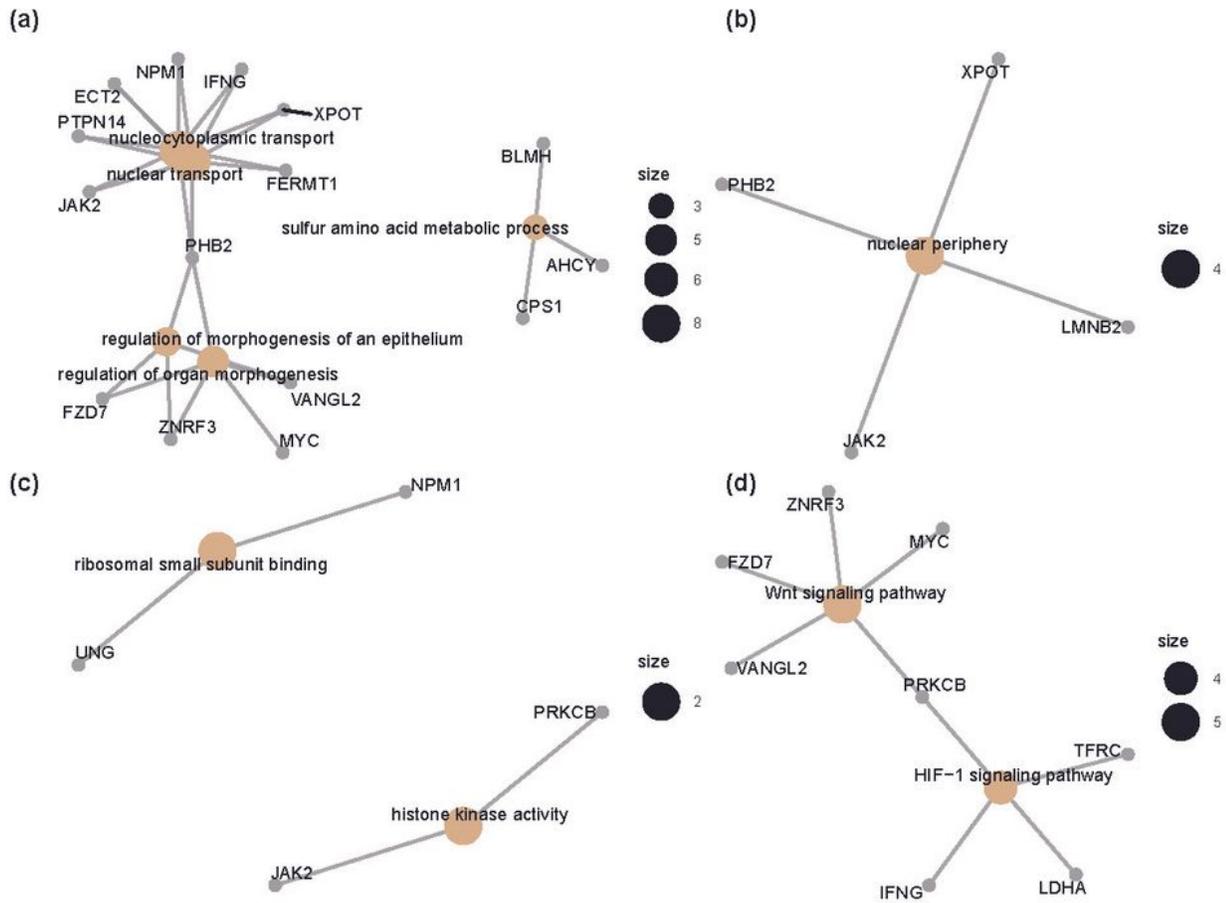


Figure 5

Network of representative GO terms and KEGG pathways. GO term enrichment analysis for biological process (A), molecular function (B), cellular component (C). KEGG pathway analysis (D). GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

Figure 6

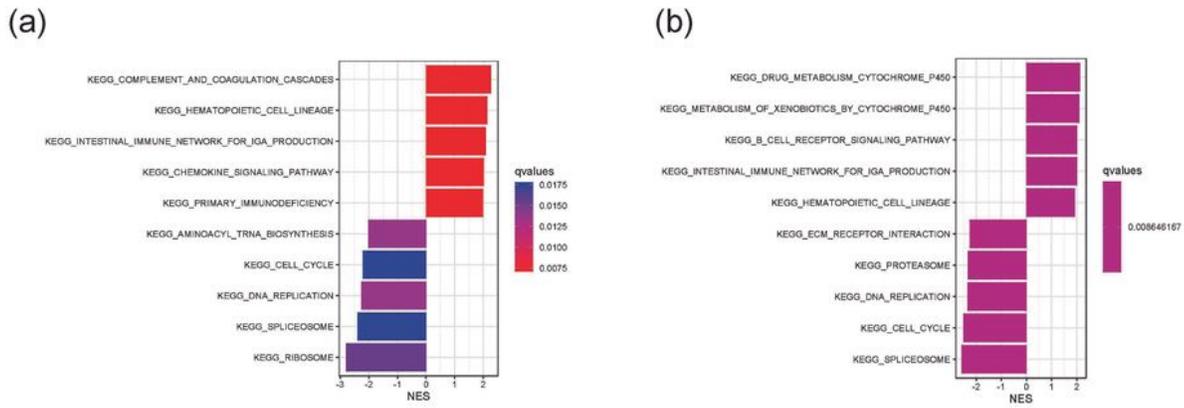
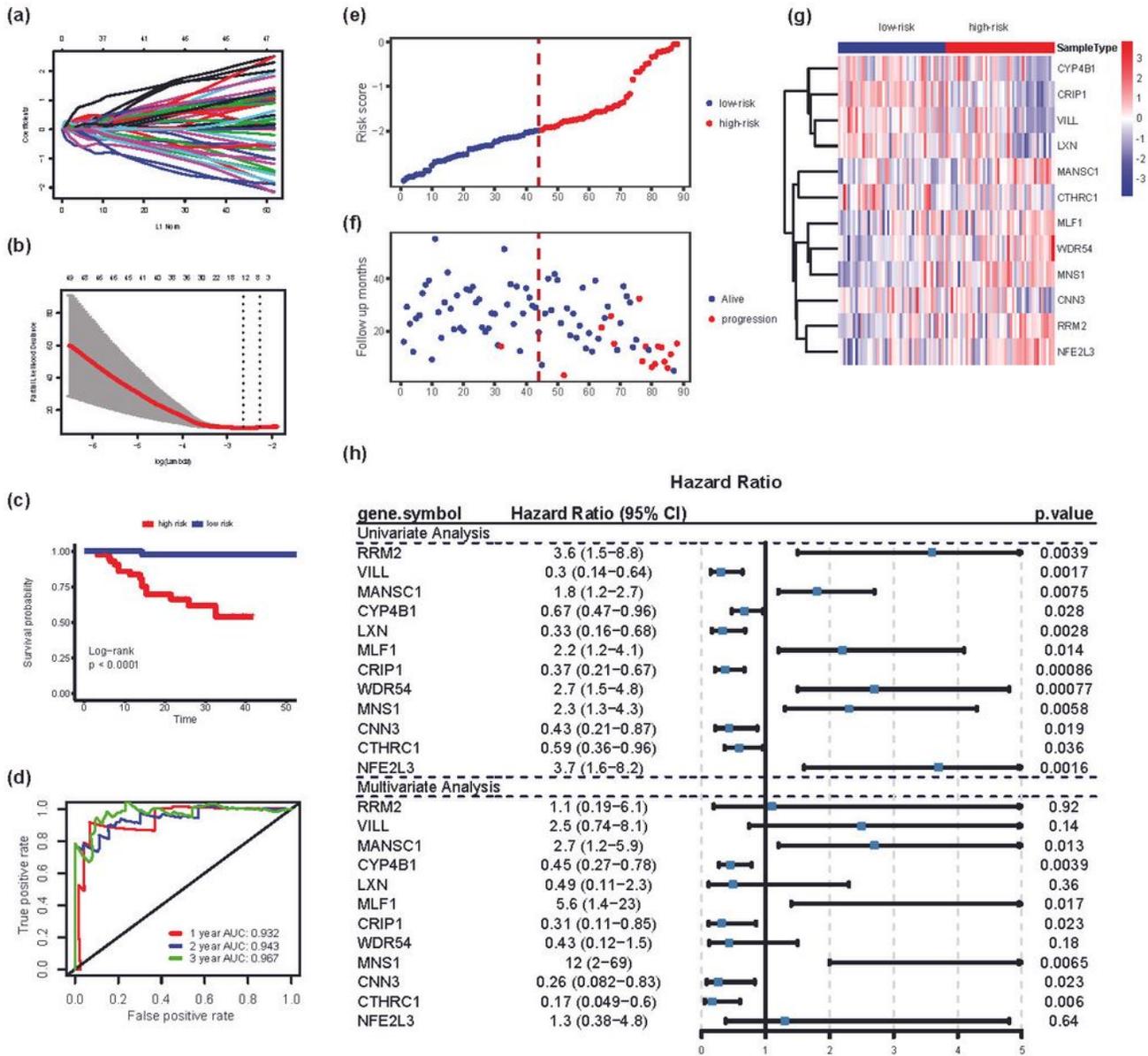


Figure 6

Gene set enrichment analysis (GSEA) for ZNF667-AS1 using GSE102349 (A) and GSE12452 (B). The pathway with normalized enrichment score (NES) >0 may be activated when ZNF667-AS1 was up-regulated, otherwise it may be suppressed

**Figure 7**

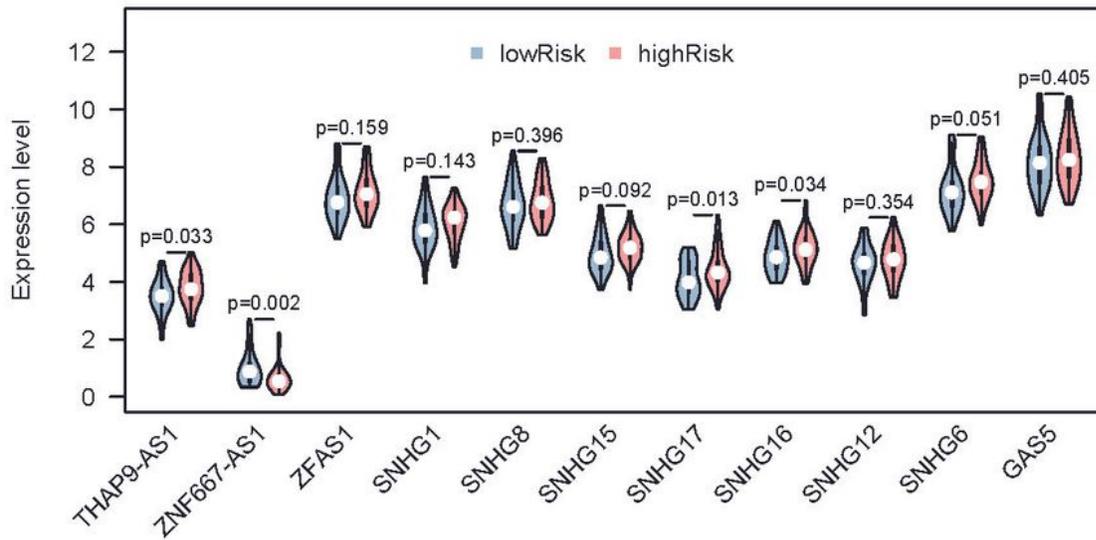


**Figure 7**

Construction of risk score model based on LASSO penalized Cox regression analysis. LASSO coefficients at varying levels of penalty were shown in (A). Plot of the cross-validation error rates was shown in (B). The left vertical line in the plot shows the CV-error curve hits its minimum. The right vertical line shows the most regularized model with CV-error within 1 standard deviation of the minimum. Base on the lasso penalized Cox regression of PFS, the RS for each sample was calculated. Then, according to the median

value of RS, the samples were separated into high- and low-risk groups. Kaplan–Meier survival analysis for the risk score model was shown in (C). Time-dependent ROC curve analysis for the risk score model was shown in (D). Risk score distribution of samples was shown in (E). Survival profile of samples was shown in (F). A heat map was generated to show the expression of 12 prognostic genes in NPC (G). A forrest plot of the univariate and multivariate Cox regression analysis for the prognostic genes associated with PFS was shown in (H). NPC: nasopharyngeal carcinoma; CI: confidence interval; PFS: progression-free survival

**Figure 8**



## Figure 8

Comparison of the expression level of the lncRNAs in the ceRNA network between high- and low-risk groups in NPC. NPC: nasopharyngeal carcinoma

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementalmaterial.pdf](#)