

# Analysis of the Differentially Expressed Genes and microRNAs and Prediction of miRNA-mRNA Negative Regulatory Network in Nasopharyngeal Carcinoma

Ping Ouyang<sup>1,†</sup>, Wenyan Wu<sup>1,†</sup>, Rang Li<sup>2</sup>, Xuefeng Zhou<sup>1</sup>, Tao Li<sup>1,\*</sup>

<sup>1</sup>Guangdong Provincial Key Laboratory of Medical Molecular Diagnostics, Guangdong Medical University, Dongguan, P. R. China

<sup>2</sup>Institute of Medical Systems Biology, School of Public Health, Guangdong Medical University, Dongguan, P. R. China

## Email address:

litao@gdmu.edu.cn (Tao Li)

\*Corresponding author

† Ping Ouyang and Wenyan Wu are co-first authors.

## To cite this article:

Ping Ouyang, Wenyan Wu, Rang Li, Xuefeng Zhou, Tao Li. Analysis of the Differentially Expressed Genes and microRNAs and Prediction of miRNA-mRNA Negative Regulatory Network in Nasopharyngeal Carcinoma. *American Journal of Internal Medicine*.

Vol. 9, No. 1, 2021, pp. 36-48. doi: 10.11648/j.ajim.20210901.16

Received: January 5, 2021; Accepted: January 22, 2021; Published: January 28, 2021

**Abstract:** Introduction: Nasopharyngeal carcinoma (NPC) is an endemic cancer in southern China, particularly in Guangdong population, but the prognosis of NPC is poor. Recently microRNA (miR) has been shown to have function in aiding the treatment of cancer. Thus, in this study, miRNAs and genes associated with NPC were analyzed. Methods: mRNA-sequencing and miR-sequencing data were obtained from the Gene Expression Omnibus. The differentially expressed genes (DEGs) and miRNAs (DEMs) were filter out. Then, the gene function annotations about the DEGs were predicted using Gene Ontology (GO) and KEGG pathway. Subsequently, the protein-protein interaction (PPI) network was established based on the STRING database, and function modules were identified using Cytoscape. Finally, DEGs targeted by DEMs were predicted by using the miRDB, miRTarBase, TargetScan and DIANA databases, and the DEM-DEG negative interaction network was built. Results: In all, 704 DEGs (about 49.9% upregulated) were enriched in 234 GO terms and 53 KEGG pathways. Seven hub genes (APP, GNG2, VAV1, RAC2, YES1, EGFR and GNB5) in 6 function modules were found for the PPI network. In addition, 86 DEMs were identified containing 56 upregulated and 30 downregulated miRNAs. There were 538 DEM-DEG pairs, of which miR-93-5p/TGFB2, miR-455-3p/STK17B and miR-766-5p/ITGAV had functions in other cancers, moreover, these pairs may potentially contributed to NPC pathogenesis. Conclusion: The constructed miRNA-mRNA negative regulatory network will give help in elucidating the molecular mechanisms of NPC. The important DEGs, DEMs and DEM-DEG pairs associated with NPC may contribute to the diagnosis and treatment of NPC in the future.

**Keywords:** Nasopharyngeal Carcinoma (NPC), Differentially Expressed Genes (DEGs), Differentially Expressed microRNAs (DEMs), miRNA-mRNA Regulatory Network

## 1. Introduction

Nasopharyngeal carcinoma (NPC) is an endemic cancer in southern China, particularly in Guangdong population [1, 2]. There are about 129,000 new NPC cases and 73,000 associated deaths in 2018, moreover, it showed an uptrend every year [3]. Multiple factors contribute to the accumulation of the NPC development, including predisposing genetic factors, environmental carcinogens, and Epstein-Barr virus infection [2, 4]. So far, the NPC

pathogenesis has not been well known, which limits the diagnosis and therapy of NPC [5]. Thus, exploring the molecules taken part in the NPC progression and their relationships is essential.

RNA sequencing (RNA-Seq) is a revolutionary tool that has been widely used in cancer research [6]. MicroRNAs (miRNAs) are useful diagnostic and prognostic indicators for human cancer [7, 8]. By comparing the RNA sequencing results of cancer tissue and normal tissue samples can provide insights into the molecular

mechanism of tumorigenesis [9]. The subsequent combination of bioinformatics analysis plays an important role in screening tumor biomarkers [10]. Ge et al [11] performed bioinformatic analysis of the GSE12452 dataset which included 31 cancer tissues of NPC and 10 control samples, and identified some important genes that may contribute in NPC development. Zhu et al [12] analyzed three mRNA expression datasets (GSE12452, GSE34573 and GSE64634) comprising of 59 NPC cases and 17 controls and also identified some important pathways and significant expressed genes in NPC development by bioinformatic tools. Liu et al [13] collected 3 mRNA expression profiles (GSE53819, GSE13597, GSE12452) and 5 miRNA expression profiles (GSE32906, GSE46172, GSE22587, GSE32960 and GSE36682) and found some miRNAs and genes might play essential functions in the NPC development process. Thus, screening the differentially expressed miRNAs or genes in case and control helps in elucidating the molecular mechanisms of NPC. Nevertheless, the studies focused on the bioinformatics analysis of the regulation between miRNAs and mRNAs in NPC are still limited, the combination analysis of the miRNAs and mRNAs may be more helpful in uncovering the mechanisms of the occurrence and development of NPC.

Based on the previous research, here, three expression profile datasets were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Among which, GSE68799 and GSE118719 were mRNA data and GSE118720 was miRNA data. The differentially expressed genes (DEGs) and miRNAs (DEMs) were screened out by comparing the data of the NPC cancer tissues and normal nasopharyngeal mucosal controls. Then, the DEGs were carried out a series of bioinformatics analysis, including the functional enrichment analysis, the protein-protein interactions, the functional modules and the hub genes analysis. After that, for the DEMs, the potential downstream target DEGs were predicted, and the miRNAs-mRNAs negative regulated network in NPC progression were constructed. The results will be helpful in further uncovering the NPC pathogenesis.

## 2. Methods

### 2.1. The mRNA and miRNA Sequencing Data

The mRNA expression profiles (GSE68799 and GSE118719) and miRNA expression profile (GSE118720) were obtained from GEO. Data GSE68799 included 42 Chinese NPC specimens and 4 control specimens, while data GSE118719 and GSE118720 were from the same group containing 7 NPC tissues and 4 normal nasopharyngeal mucosal controls. In addition, GSE68799 was achieved using the platform of GPL11154 Illumina HiSeq 2000, while GSE118719 and GSE118720 were obtained using GPL20301 Illumina HiSeq 4000 (Illumina, San Diego, CA, USA).

### 2.2. Screening out NPC-associated mRNAs and microRNAs

After three expressing profiles GSE68799, GSE118719 and GSE118720 were downloaded, the data were processed using R/Bioconductor package limma linear models for sequencing data to analyze the DEGs and DEMs between the NPC tissues and non-NPC tissues. The false discovery rates (FDR) < 0.05 and the fold-change (FC)  $\geq 2$  were used as a cut-off.

### 2.3. Functional Annotation and Pathway Enrichment Analysis of the DEGs

DAVID (<https://david.ncicrf.gov>) is a well-known database which can supply functional interpretation of gene clusters [14]. The functional enrichment analysis includes Gene Ontology (GO) functional enrichment analysis comprised of biology process (BP), molecular function (MF) and cellular component (CC). In this study, DEGs and the predicted miRNA target genes were analyzed using DAVID, including enrichment analysis of GO functional terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.  $P < 0.05$  was regarded as significant.

### 2.4. Protein-protein Interaction (PPI) Network Construction and Module Analysis

The interactions between different proteins encoded by the DEGs were performed using STRING database (<https://string-db.org/>) [15], and the combined score of > 0.9 was thought as a cut-off. The PPI network was visualized using Cytoscape software v3.7.2 (<http://www.cytoscape.org>). The functional modules within the PPI network were established using the Molecular Complex Detection (MCODE) in Cytoscape, and the significance score of the module was set to > 5.0.

### 2.5. miRNA-mRNA Interaction Network Construction

The mRNAs targeted by different DEMs in this study were predicted by four databases: miRDB (<http://mirdb.org/>), miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/>), TargetScan ([http://www.targetscan.org/vert\\_72/](http://www.targetscan.org/vert_72/)) and DIANA ([http://carolina.imis.athena-innovation.gr/diana\\_tools/web/index.php](http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php)). It should be noted that the DIANA database was used for predicting the mRNAs targeted by EBV BART miRNAs specially. All the predicted mRNA targets were taken intersected with the DEGs, and the overlapping genes were defined as the mRNA target dataset. Then the miRNA-mRNA (DEMs-DEGs) interaction network with negative interaction relationships was constructed and visualized using Cytoscape software v3.7.2.

## 3. Results

### 3.1. Expression Profiling Datasets Retrieved from GEO Datasets of NCBI

The sequencing samples contained in the present study

were NPC tissue samples and non-NPC tissue samples. In this study, mRNA expression profiles GSE68799 included a collection of 42 Chinese Nasopharyngeal Carcinoma patients tissues and 4 non-NPC nasopharyngeal tissues. mRNA expression profiles GSE118719 and miRNA expression profiles GSE118720 included a collection of 7 NPC samples and 4 normal nasopharyngeal mucosal samples. Information of expression profiling sequencing datasets was present in Table 1.

### 3.2. DEGs and DEMs Analysis

Following analysis of GSE68799, GSE118719, GSE118720 datasets, there were 2339 DEGs (1352 up and 987 down), 2732 DEGs (945 up and 1787 down) and 86 DEMs (56 up and 30 down) were screened out, respectively. The cluster analysis of DEGs showed great differences in the normal nasopharyngeal controls and NPC specimens (Figure 1, Figure 2 and Figure 3). Furthermore, using Venn diagram analysis, GSE68799 and GSE118719 shared 704 identical DEGs, among which, 351 were upregulated (Figure 4).

### 3.3. Functional Annotation and Pathway Enrichment Analysis of DEGs

The above mentioned 704 DEGs were then performed functional enrichment analysis, and there were 129, 54 and 51 significant functional terms for BP, CC and MF, respectively. Positive regulation of GTPase activity ( $p=2.01E-12$ ), signal transduction ( $p=3.15E-12$ ) and B cell receptor signaling pathway ( $p=4.68E-09$ ) were the three most significant function terms for the biological process (Figure 5A). Plasma membrane ( $p=2.45E-20$ ), integral component of plasma membrane ( $p=9.95E-11$ ) and cell-cell junction ( $p=1.05E-08$ ) were the three most significant function terms for cellular component (Figure 5B). GTPase activator activity ( $p=1.11E-11$ ), guanyl-nucleotide exchange factor activity ( $p=4.11E-09$ ), SH3 domain binding ( $p=2.76E-08$ ) were the most three enriched terms for molecular function (Figure 5C).

The pathway enrichment analysis was performed and there were 53 KEGG significant pathways. The top ten significant pathways were shown in Figure 5D, among which, the three primarily associated pathways were as follows: B cell receptor signaling pathway ( $p=2.32E-14$ ), Pathways in cancer ( $p=1.30E-08$ ), Focal adhesion ( $p=3.95E-07$ ) (Figure 5D).

### 3.4. PPI Network Construction

The PPI network based on 704 proteins was constructed using STRING, and it included 287 proteins and 903 interaction pairs (Figure 6). Fifty-seven proteins had a degree of more than 10, and 7 proteins had a degree of more than

25, which were regarded as hub genes as follows: amyloid beta precursor protein (APP, Degree=34), G protein subunit gamma 2 (GNG2, Degree=29), vav guanine nucleotide exchange factor 1 (VAV1, Degree=28), Rac family small GTPase 2 (RAC2, Degree=28), yes proto-oncogene 1, src family tyrosine kinase (YES1, Degree=28), epidermal growth factor receptor (EGFR, Degree=27), and G protein subunit beta 5 (GNB5, Degree=27).

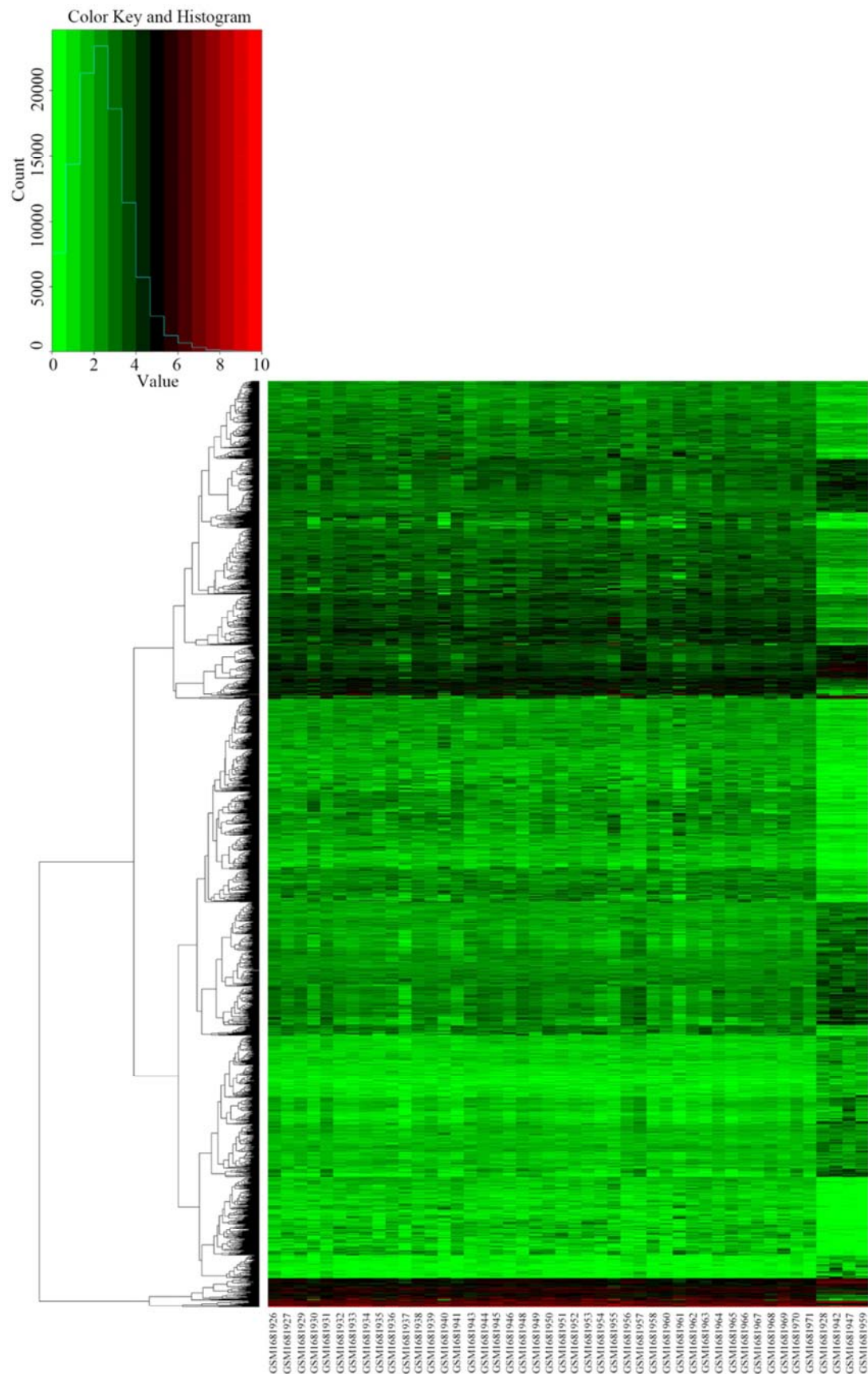
### 3.5. Module Analysis

Module analysis was conducted using the MCODE in Cytoscape to predict protein complexes. The significant score of the modules were set to  $>5.000$ . As a result, there were 6 modules were identified (Figure 7). The detailed information for module 1 (Nodes=17, Edges=136, Score=17.000), module 2 (Nodes=10, Edges=45, Score=10.000), module 3 (Nodes=21, Edges=73, Score=7.300), module 4 (Nodes=7, Edges=21, Score=7.000), module 5 (Nodes=6, Edges=15, Score=6.000), and module 6 (Nodes=6, Edges=13, Score=5.200).

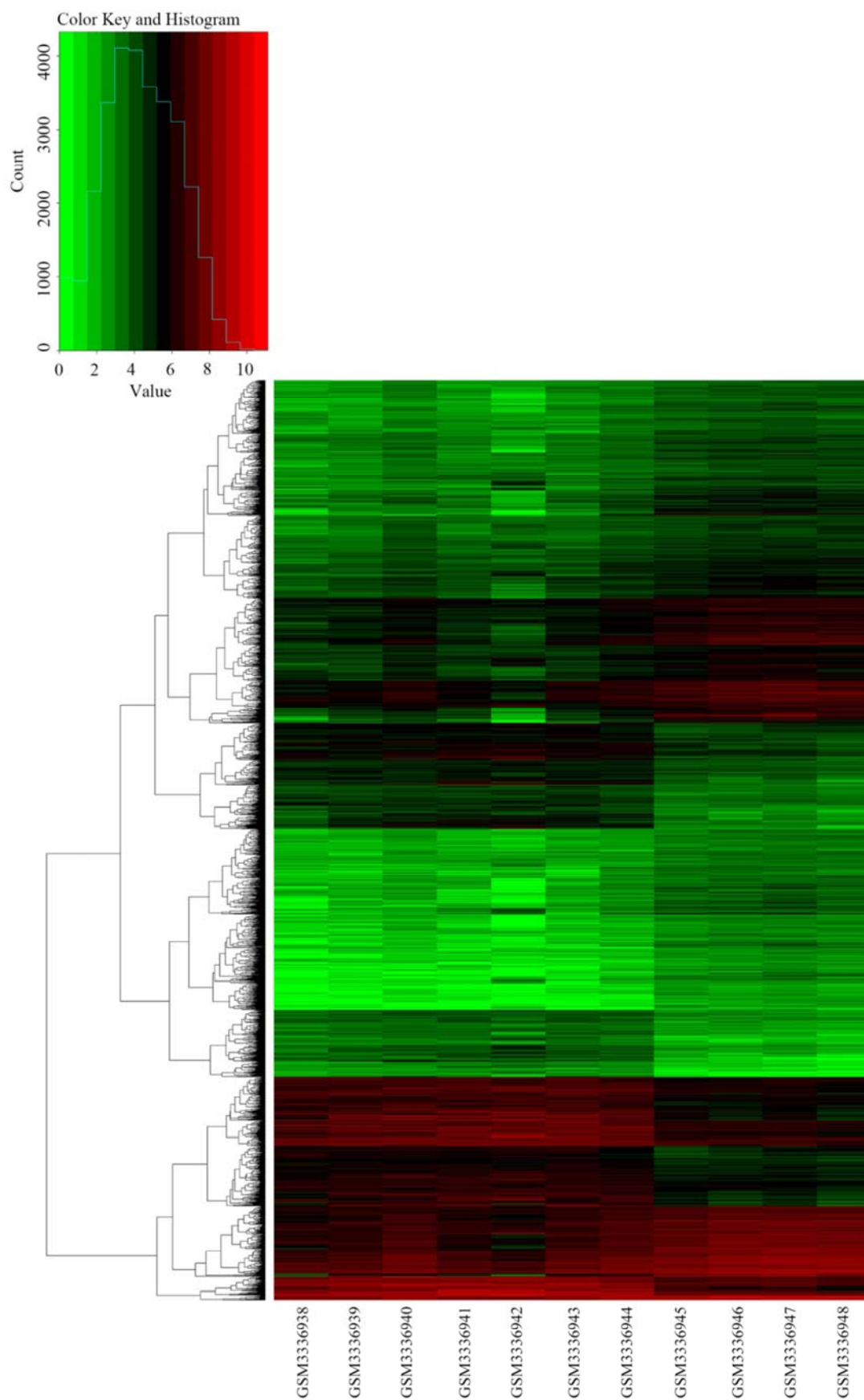
Moreover, the DEGs included in different modules were carried out pathway enrichment analysis. For DEGs in module 1, the most significant enriched KEGG pathway was Chemokine signaling pathway. The DEGs in module 2 were all down-regulated, and the most significant pathway was B cell receptor signaling pathway. In module 3, the top enriched pathway was ECM-receptor interaction. The DEGs involved in module 4 and module 5 were upregulated, the top enriched pathways were Protein digestion and absorption and Arrhythmogenic right ventricular cardiomyopathy, respectively. In module 6, the top enriched KEGG pathway was Regulation of actin cytoskeleton (Table 2).

### 3.6. miRNA-mRNA Interaction Network Construction

In this study, miRDB, miRTarBase, TargetScan and DIANA databases were applied to predict the target genes of miRNA. Since there were so many interaction pairs and the network was huge, we chose to focus the negatively regulated miRNA-mRNA interactions. The miRNA-mRNA negative regulatory network contained 43 DEMs, 284 DEGs and 538 miRNA-mRNA pairs (Figure 8). Some miRNAs had many targets, among which, the top 5 upregulated miRNAs were hsa-miR-548p, hsa-miR-93-5p, hsa-miR-548ah-3p, hsa-miR-329-3p and hsa-miR-455-3p; while the top five downregulated miRNAs were hsa-miR-1273h-5p, hsa-miR-29a-3p, hsa-miR-766-5p, hsa-miR-29b-3p and hsa-miR-29c-3p. Many miRNA-mRNA pairs were reported in different cancer pathogenesis, such as miR-93-TGFBR2, miR-455-3p-STK17B and miR-29-VEGFA (Table 3).

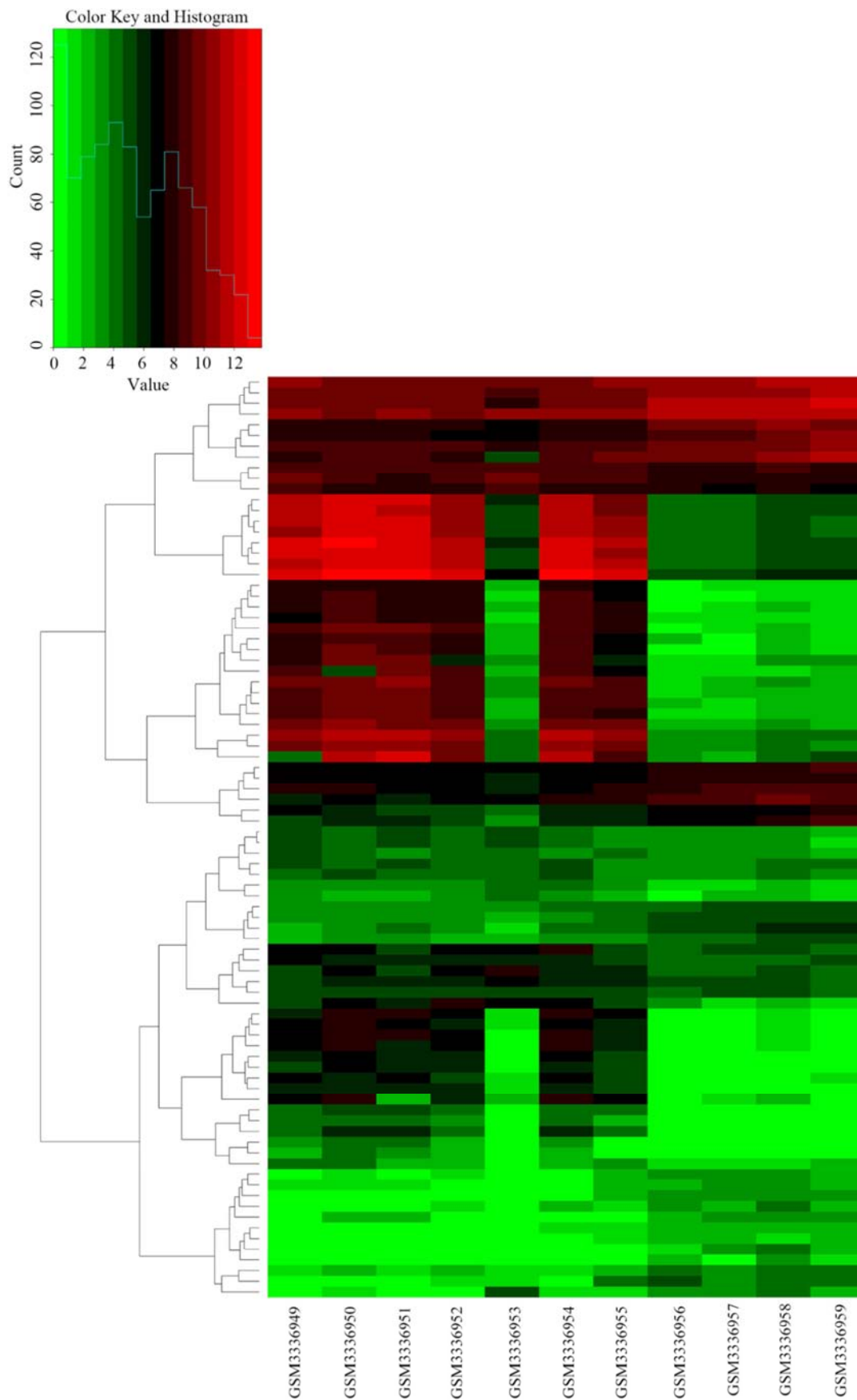


**Figure 1.** Cluster analysis of the mRNAs in GSE68799. Row and column represented mRNAs and tissue samples, respectively. The color scale represented the expression levels.



**Figure 2.** Cluster analysis of the mRNAs in GSE118719. Row and column represented mRNAs and tissue samples, respectively. The color scale represented the expression levels.

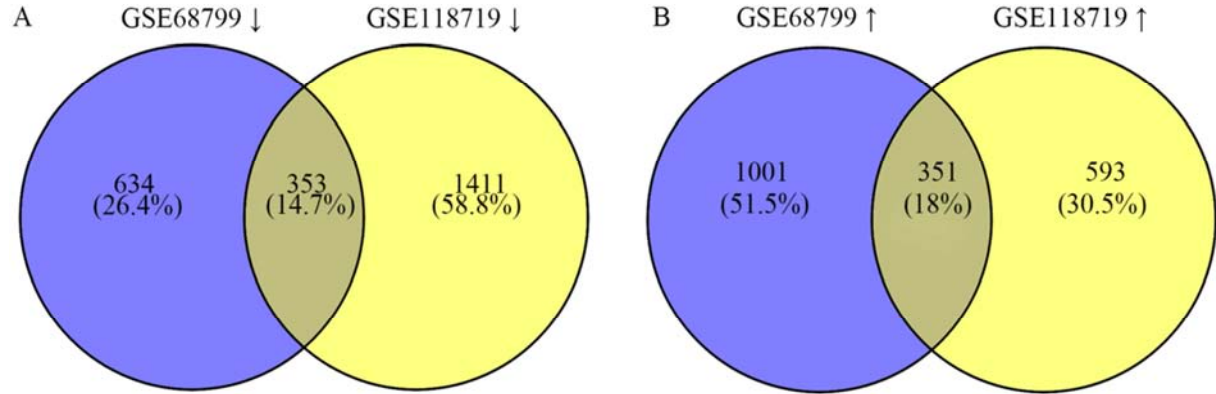
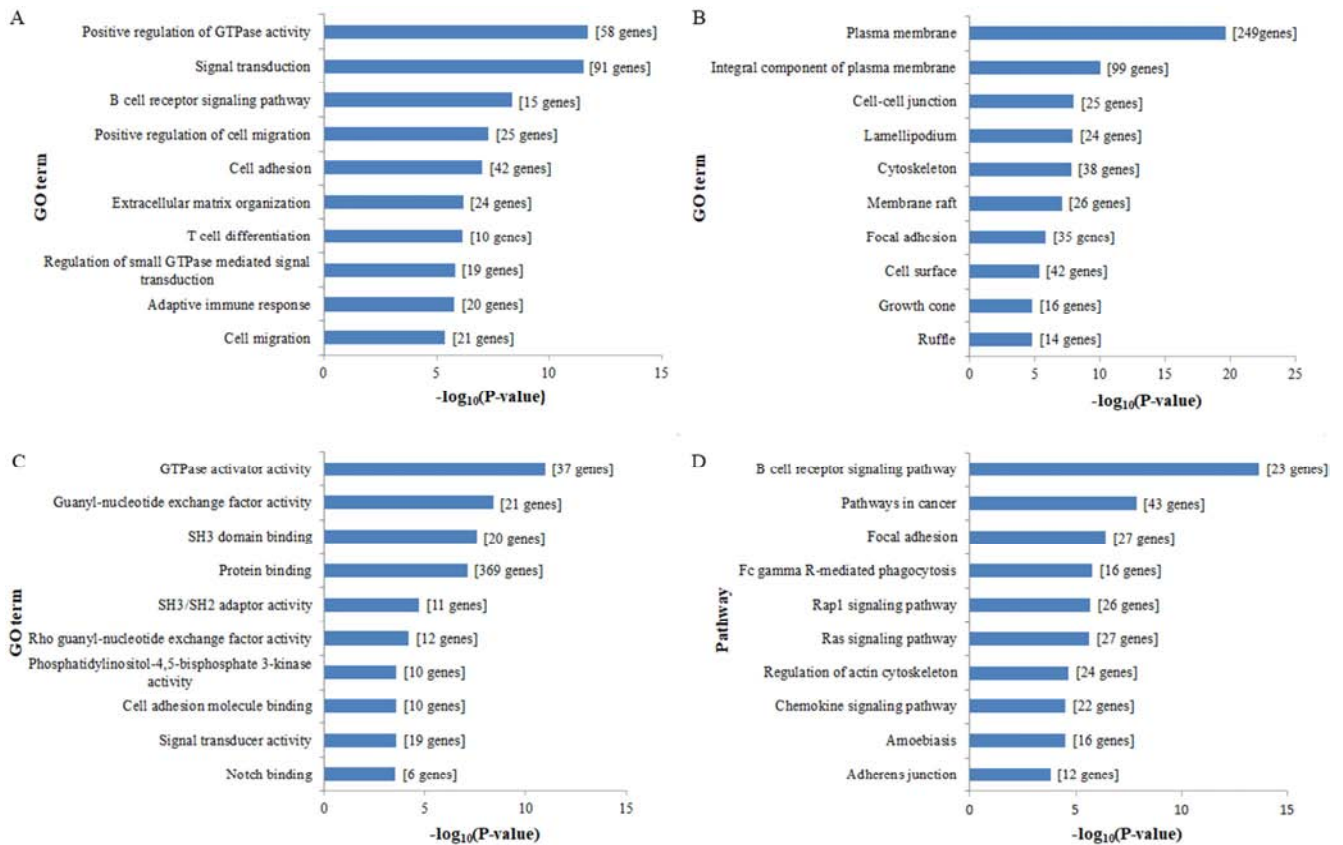


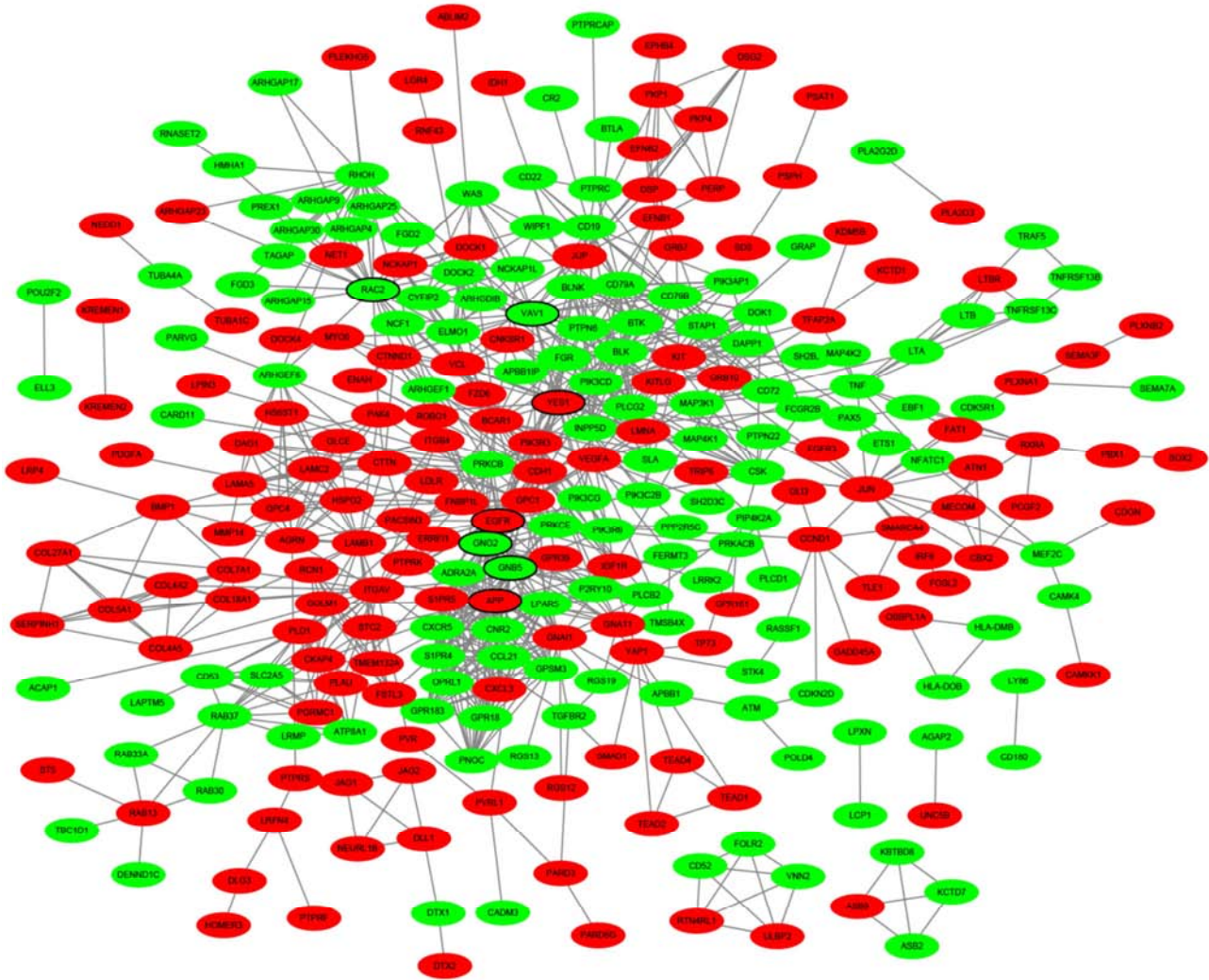


**Figure 3.** Cluster analysis of the microRNAs in GSE118720. Row and column represented microRNAs and tissue samples, respectively. The color scale represented the expression levels.

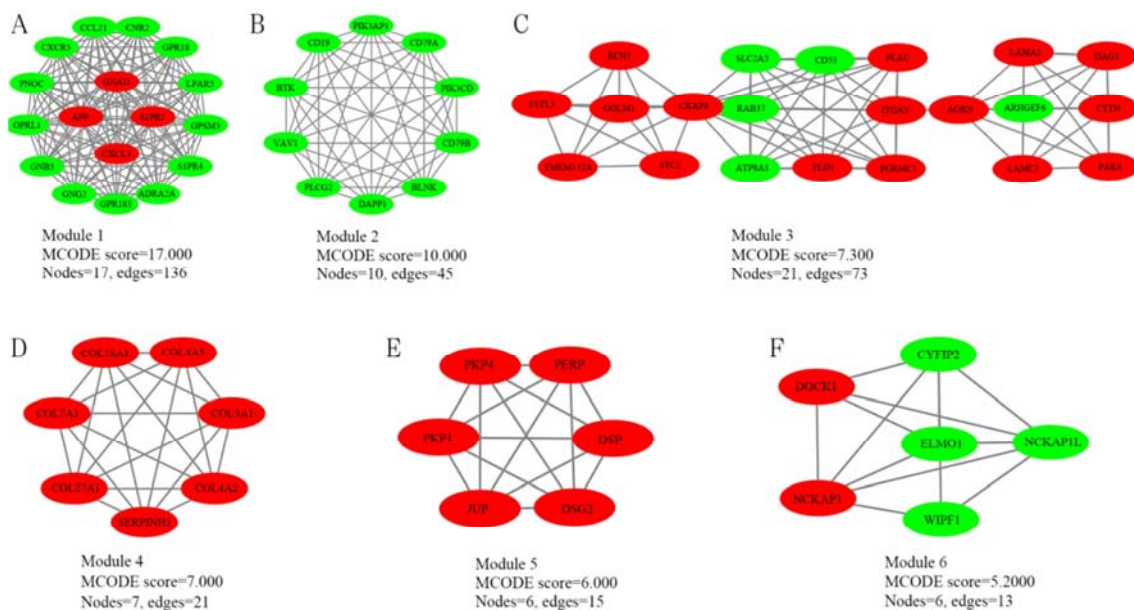
**Table 1.** Information of sequencing datasets.

Accession number	Platform	Organism	Experiment type	Information of samples
GSE68799	GPL11154	Homo sapiens	mRNA sequencing	42 NPC and 4 non-NPC
GSE118719	GPL20301	Homo sapiens	mRNA sequencing	7 NPC and 4 non-NPC
GSE118720	GPL20301	Homo sapiens	microRNA sequencing	7 NPC and 4 non-NPC

**Figure 4.** Venn diagram analysis of the DEGs in different datasets. The blue circle represented the GSE68799 dataset and the yellow circle represented the GSE118719 dataset. The intersection of the two circles represented the identical DEGs between the two datasets. (A) The upregulated DEGs from both datasets; and (B) The downregulated DEGs from both datasets.**Figure 5.** The top 10 significantly enriched GO terms and KEGG pathways of the DEGs. The X-axis represented  $-\log_{10}(P\text{-value})$  and the Y-axis showed GO terms or KEGG pathways. (A) Biological Process; (B) Cellular Component; (C) Molecular Function; and (D) KEGG pathways.

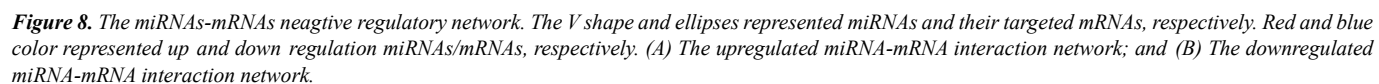


**Figure 6.** Protein-protein interaction (PPI) networks of the DEGs. The red and green ellipses represented proteins encoded by the up and downregulated DEGs. Ellipses with black border were the 7 hub DEGs.



**Figure 7.** The six MCODE modules screened out from the PPI network. The red and green ellipses represented proteins encoded by the up and down regulated DEGs. (A) Module 1; (B) Module 2; (C) Module 3; (D) Module 4; (E) Module 5; and (F) Module 6.





**Table 2.** The most significant pathways for the DEGs in different modules of the PPI network.

Module	Term	Description	Count	Genes	P-Value
1	hsa04062	Chemokine signaling pathway	6	CXCR5, GNAI1, CCL21, CXCL3, GNB5, GNG2	9.30E-06
2	hsa04662	B cell receptor signaling pathway	10	CD19, DAPP1, PIK3CD, PLCG2, PIK3AP1, CD79B, CD79A, PIK3R3, BLNK, BTK	5.19E-19
3	hsa04512	ECM-receptor interaction	5	LAMA5, ITGAV, DAG1, LAMC2, AGRN	7.36E-06
4	hsa04974	Protein digestion and absorption	6	COL18A1, COL4A2, COL7A1, COL27A1, COL5A1, COL4A5	3.06E-10
5	hsa05412	Arrhythmogenic right ventricular cardiomyopathy	3	JUP, DSG2, DSP	2.79E-04
6	hsa04810	Regulation of actin cytoskeleton	4	DOCK1, CYFIP2, NCKAP1L, NCKAP1	2.68E-04

**Table 3.** The top 5 up- and down-regulated miRNAs and their targets in the miRNA-mRNA network.

miRNA	Expression level	Target Number	Prediction targets
hsa-miR-548p	up	23	AP1S2, ATP2A3, ATP8A1, BANK1, DOK3, ELMO1, ETS1, HHEX, LBH, MEF2C, OSBP10, PLA2G2D, PRKACB, PXX, RALGPS2, SLC38A1, STAMBPL1, STK4, STX7, TBC1D1, TOX, WIPF1, ZFP82
hsa-miR-93-5p	up	20	BCAS4, CD69, DENND5B, GNB5, KBTBD8, KIAA0922, MAP3K8, NIN, ORAI2, PIP4K2A, PREX1, PRKACB, PRKCB, PXX, RAB30, SEMA7A, SLC46A3, STK17B, TGFB2, TTC9
hsa-miR-548ah-3p	up	19	AFF3, ATP8A1, CBA2T3, CD1D, EMB, ETS1, GNG2, IKZF1, MAP3K1, MEF2C, NIN, P2RY10, PLCL2, PRKCB, RAB33A, SGPP1, SLAIN1, ST6GAL1, TOX
hsa-miR-329-3p	up	13	ANKRD44, ARHGEF6, CD180, CD19, CDKN2D, ETS1, KCTD7, MEF2C, PIK3CG, PLEKHA2, RAB37, TNFRSF13C, TRAF5
hsa-miR-455-3p	up	12	BCAS4, GNB5, KIF21B, MAPRE2, PSD4, RASGRP3, STK17B, STK4, STX7, TMEM154, TNFRSF13C, TTC9
hsa-miR-1273h-5p	down	32	AFAP1, CDC42BPA, CDH1, CDK2AP1, COL5A1, CTNND1, CXCR5, DAG1, FAM83B, FAM83H, FNDC3B, FOSL2, FUT1, GPC4, KIAA1671, LRP4, MEX3A, MYO10, NAV1, NFIC, OSBP1A, PLEKHG5, PVR, RNF207, SCD5, SOX13, ST5, TSKU, UNC5B, VCL, ZNF286A, ZNF503
hsa-miR-29a-3p	down	28	ANKRD13B, ATP1B1, CBX2, CD276, CDC42BPA, CLDN1, COL4A2, COL4A5, COL5A1, COL7A1, CUEDC1, DAG1, FRAS1, KDM5B, KREMEN2, LDOC1L, NAV1, PLXNA1, PTPRK, RNF165, ROBO1, SERPINH1, SHROOM2, TMEM132A, TMTC3, ULBP2, VEGFA, ZNF286A
hsa-miR-766-5p	down	27	CBS, CD276, CDK2AP1, CUEDC1, DAG1, EHBPI, FAM160A1, FOSL2, FRAS1, ITGAV, KREMEN1, LAMC2, LZTS1, MAL2, MEX3A, MTSS1L, MYO6, PBX1, PCSK6, PTPRF, RBFOX2, SHROOM2, SOX4, STC2, TBC1D16, TMEM63B, TSPAN9
hsa-miR-29b-3p	down	26	ANKRD13B, ATP1B1, BMP1, CBX2, CD276, CDC42BPA, CLDN1, COL4A2, COL4A5, COL5A1, COL7A1, CUEDC1, FRAS1, KDM5B, LDOC1L, NAV1, PLXNA1, PTPRK, RNF165, ROBO1, SERPINH1, SHROOM2, TMEM132A, TMTC3, ULBP2, VEGFA

## 4. Discussion

In the present research, bioinformatics methods were applied to analyze three public NPC sequencing data, and found some potential important genes and miRNAs related with the pathogenesis of NPC, the miRNA-mRNA correlation pairs may supply some clues for the molecular mechanisms of NPC pathogenesis.

Through protein-protein interaction network analysis, 7 hub genes (APP, GNG2, VAV1, RAC2, YES1, EGFR and GNB5) were found, which suggested their potential important roles in the pathogenesis of NPC. Among these genes, EGFR is a well known cancer-related gene [13]. It were reported to be overexpressed in most head-and-neck cancers [14]. Aberrant expression of EGFR presented potential unfavorable prognosis for patients with NPC [16, 17], and activation of EGFR signal transduction can promote the invasion and metastasis of NPC cells [15]. Interesting, another hub gene APP was reported to be a downstream protein of EGFR in NPC. Knockdown of APP can suppress the progression of NPC through inhibiting cell viability, migration and invasion [18], and upregulation of secretory

APP by EGFR may relate to the pathogenesis of NPC [19]. Our studies were consistent with these previous research that EGFR/APP took part in NPC progression. RAC2 can regulate many biological processes in cells, such as secretion, and phagocytosis [20]. RAC2 was also reported to play roles in cancer, and it may be a prognostic biomarker in clear cell renal cell carcinoma [21]. Moreover, RAC2 had also been regarded as a hub gene in the NPC high-throughput data bioinformatics analysis [11], but the molecular mechanism of RAC2 in NPC was not clear and deserved researching in the next experiment. The research of other 4 hub genes (GNG2, VAV1, GNB5 and YES1) in NPC was limited. However, VAV1, GNB5 and YES1 were reported to affect the progression of other cancer types. For example, YES1 can drive lung cancer growth and progression and predict sensitivity to Dasatinib [22]. What's more, YES1 has an important role in antitumor immunity, and phosphorylated Yes (p-Yes) level combined with the pIgR level can work as a prognostic biomarker in specific hepatocellular carcinoma patients [23]. Among the 7 hub genes, GNG2 was the least reported in tumor. It was reported the GNG2 expression level was dropped in malignant melanoma, which indicated GNG2 may be a cancer suppress gene in NPC and this needed

further confirmation [24].

In this study, a total of 86 DEMs were found in NPC. We focused on the DEMs which had more target genes within the DEGs. Among the up-regulated DEMs, hsa-miR-548p, hsa-miR-93-5p, hsa-miR-548ah-3p, hsa-miR-329-3p and hsa-miR-455-3p had the top 5 target mRNAs (Table 2). Among which, many of the miRNAs or the related miRNA-gene pairs have been reported in NPC or other cancer types. For example, hsa-miR-93-5p was probably to be a candidate biomarker in NPC [25]. The hsa-miR-93/TGFBR2 (Transforming Growth Factor Beta Receptor 2, TGFBR2) pair promoted the proliferation and invasion in prostate tumor [13, 14]. hsa-miR-455-3p inhibited the proliferation and invasion of esophageal squamous cell carcinoma, and could serve as a prognostic marker [9]. The hsa-miR-455-3p/STK17B (serine/threonine kinase 17B, STK17B) pair can stimulate hepatocellular carcinoma cell proliferation and metastasis [14]. Interesting, in this study, we also predicted hsa-miR-93/TGFBR2 and hsa-miR-455-3p/STK17B took part in NPC progression, and this will need further experimental verification. It should be noted that hsa-miR-548p was reported to be a tumor suppressors in breast cancer and hepatitis B virus X protein associated hepatocellular carcinoma [26, 27], however, it was upregulated in this study, which suggested it may be a tumor promoter in NPC. Now the research about hsa-miR-548p is not enough, and whether hsa-miR-548p plays different roles in various cancer types will need further experiments.

For the down-regulated DEMs, hsa-miR-1273h-5p, hsa-miR-29a-3p, hsa-miR-766-5p, hsa-miR-29b-3p and hsa-miR-29c-3p had the top 5 target mRNAs. All the 5 miRNAs were cancer-related. hsa-miR-766-5p was downregulated in lung cancer tissues and cells [28] and inhibited the cancer progression in colorectal cancer [29]. hsa-miR-1273h-5p can inhibit cell proliferation, migration and invasion by targeting chemokine ligand (CXCL12) in gastric cancer [30]. The hsa-miR-29 family members (miR-29a, -29b, and -29c) were cancer suppressors in different cancer types. In nasopharyngeal carcinoma cells, hsa-miR-29-3p can improve the radiotherapy sensitivity by targeting COL1A1 3'-UTR [31]. In this study, we predicted the other 4 collagen genes (COL4A2, COL4A5, COL5A1, COL7A1) may be potential targets of hsa-miR-29a-3p (Table 2). MiR-29b-3p was sharply reduced in NPC and this result was in consistence with a meta analysis of NPC [32]. Notably, most predicted targeted mRNAs of miR-29 family were cancer associated, such as vascular endothelial growth factor A (VEGFA) [16], serpin family h member 1 (SERPINH1) [15], collagen type IV alpha 2 (COL4A2) [11]. Of these targets, VEGFA was a key target for the hsa-miR-29 family. For example, miR-29a can suppress the growth and invasion of gastric cancer cells by targeting VEGFA [23], hsa-miR-29b can inhibit angiogenesis by targeting VEGFA in endometrial carcinoma [23] and hsa-miR-29c can function as a tumor suppressor through targeting VEGFA in lung cancer [33]. SERPINH1 was also an important target for the

hsa-miR-29a, -29b, and -29c, it had been reported that hsa-miR-29a, -29b, and -29c can directly regulate SERPINH1 and participates in the pathogenesis of renal cell carcinoma [24]. Therefore, hsa-miR-29 was a potential important regulator in the NPC progression. There are still some limitation in this study, the research was performed by a series of bioinformatics analysis and lack of experiment verification. In the future, more biological experiments will be needed to study the functions of the hub genes and the predicted miRNA-mRNA pairs in NPC.

## 5. Conclusions

Taken together, the present study have found many DEGs and DEMs associated with NPC, and constructed a miRNA-mRNA negative regulatory network. The important DEGs, DEMs and DEM-DEG pairs associated with NPC may contribute to the diagnosis and treatment of NPC in the future. The constructed miRNA-mRNA negative regulatory network will give help in elucidating the molecular mechanisms of NPC.

## Authors Contribution

Ping Ouyang: Methodology, Validation, Writing- Original draft preparation, Funding acquisition. Wenyan Wu: Software, Data curation, Visualization. Rang Li: Visualization, Investigation. Xuefeng Zhou: Validation. Tao Li: Conceptualization, Supervision, Writing- Reviewing and Editing.

## Financial Support

The present study was funded by The Medical Research Fund of Guangdong Province (grant no. B2020064) and The General Project of Guangdong University (grant no. GDMUM201820).

## Declaration of Competing Interests

The authors declare that they have no competing interest.

## References

- [1] M. L. K. Chua, J. T. S. Wee, E. P. Hui, A. T. C. Chan, Nasopharyngeal carcinoma, *Lancet* 387 (10022) (2016) 1012-1024.
- [2] Y. P. Chen, A. T. C. Chan, Q. T. Le, P. Blanchard, Y. Sun, J. Ma, Nasopharyngeal carcinoma, *Lancet* 394 (10192) (2019) 64-80.
- [3] H. M. Lee, K. S. Okuda, F. E. Gonzalez, V. Patel, Current Perspectives on Nasopharyngeal Carcinoma, *Adv Exp Med Biol* 1164 (2019) 11-34.
- [4] W. Dai, H. Zheng, A. K. Cheung, M. L. Lung, Genetic and epigenetic landscape of nasopharyngeal carcinoma, *Chin Clin Oncol* 5 (2) (2016) 16.

- [5] P. Paul, H. Deka, A. K. Malakar, B. Halder, S. Chakraborty, Nasopharyngeal carcinoma: understanding its molecular biology at a fine scale, *Eur J Cancer Prev* 27 (1) (2018) 33-41.
- [6] H. Nakagawa, M. Fujita, Whole genome sequencing analysis for cancer genomics and precision medicine, *Cancer science* 109 (3) (2018) 513-522.
- [7] R. Rupaimoole, F. J. Slack, MicroRNA therapeutics: towards a new era for the management of cancer and other diseases, *Nature reviews. Drug discovery* 16 (3) (2017) 203-222.
- [8] Y. S. Lee, A. Dutta, MicroRNAs in cancer, *Annual review of pathology* 4 (2009) 199-227.
- [9] H. Yang, Y. N. Wei, J. Zhou, T. T. Hao, X. L. Liu, MiR-455-3p acts as a prognostic marker and inhibits the proliferation and invasion of esophageal squamous cell carcinoma by targeting FAM83F, *Eur Rev Med Pharmacol Sci* 21 (14) (2017) 3200-3206.
- [10] G. Chen, T. Shi, L. Shi, Characterizing and annotating the genome using RNA-seq data, *Sci China Life Sci* 60 (2) (2017) 116-125.
- [11] Y. Ge, Z. He, Y. Xiang, D. Wang, Y. Yang, J. Qiu, Y. Zhou, The identification of key genes in nasopharyngeal carcinoma by bioinformatics analysis of high-throughput data, *Mol Biol Rep* 46 (3) (2019) 2829-2840.
- [12] H. M. Zhu, Q. Fei, L. X. Qian, B. L. Liu, X. He, L. Yin, Identification of key pathways and genes in nasopharyngeal carcinoma using bioinformatics analysis, *Oncol Lett* 17 (5) (2019) 4683-4694.
- [13] M. Liu, K. Zhu, X. Qian, W. Li, Identification of miRNA/mRNA-Negative Regulation Pairs in Nasopharyngeal Carcinoma, *Medical science monitor: international medical journal of experimental and clinical research* 22 (2016) 2215-34.
- [14] G. Dennis, Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, R. A. Lempicki, DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome biology* 4 (5) (2003) P3.
- [15] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, *Nucleic acids research* 45 (D1) (2017) D362-D368.
- [16] P. Zhang, S. K. Wu, Y. Wang, Z. X. Fan, C. R. Li, M. Feng, P. Xu, W. D. Wang, J. Y. Lang, p53, MDM2, eIF4E and EGFR expression in nasopharyngeal carcinoma and their correlation with clinicopathological characteristics and prognosis: A retrospective study, *Oncol Lett* 9 (1) (2015) 113-118.
- [17] X. J. Cao, J. F. Hao, X. H. Yang, P. Xie, L. P. Liu, C. P. Yao, J. Xu, Prognostic value of expression of EGFR and nm23 for locoregionally advanced nasopharyngeal carcinoma, *Med Oncol* 29 (1) (2012) 263-71.
- [18] J. Xu, Y. Ying, G. Xiong, L. Lai, Q. Wang, Y. Yang, Amyloid beta precursor protein silencing attenuates epithelial-mesenchymal transition of nasopharyngeal carcinoma cells via inhibition of the MAPK pathway, *Mol Med Rep* 20 (1) (2019) 409-416.
- [19] C. E. Tang, Y. J. Guan, B. Yi, X. H. Li, K. Liang, H. Y. Zou, H. Yi, M. Y. Li, P. F. Zhang, C. Li, F. Peng, Z. C. Chen, K. T. Yao, Z. Q. Xiao, Identification of the amyloid beta-protein precursor and cystatin C as novel epidermal growth factor receptor regulated secretory proteins in nasopharyngeal carcinoma by proteomics, *J Proteome Res* 9 (12) (2010) 6101-11.
- [20] A. P. Hsu, A. Donkó, M. E. Arrington, M. Swamydas, D. Fink, A. Das, O. Escobedo, V. Bonagura, P. Szabolcs, H. N. Steinberg, J. Bergerson, A. Skoskiewicz, M. Makhija, J. Davis, L. Foruraghi, C. Palmer, R. L. Fuleihan, J. A. Church, A. Bhandoola, M. S. Lionakis, S. Campbell, T. L. Leto, D. B. Kuhns, S. M. Holland, Dominant activating RAC2 mutation with lymphopenia, immunodeficiency, and cytoskeletal defects, *Blood* 133 (18) (2019) 1977-1988.
- [21] Y. Liu, G. Cheng, Z. Song, T. Xu, H. Ruan, Q. Cao, K. Wang, L. Bao, J. Liu, L. Zhou, D. Liu, H. Yang, K. Chen, X. Zhang, RAC2 acts as a prognostic biomarker and promotes the progression of clear cell renal cell carcinoma, *International journal of oncology* 55 (3) (2019) 645-656.
- [22] I. Garmendia, M. J. Pajares, F. Hermida-Prado, D. Ajona, C. Bértolo, C. Sainz, A. Lavín, A. B. Remírez, K. Valencia, H. Moreno, I. Ferrer, C. Behrens, M. Cuadrado, L. Paz-Ares, X. R. Bustelo, I. Gil-Bazo, D. Alameda, F. Lecanda, A. Calvo, E. Felip, M. Sánchez-Céspedes, Wistuba, II, R. Granda-Díaz, J. P. Rodrigo, J. M. García-Pedrero, R. Pio, L. M. Montuenga, J. Agorreta, YES1 Drives Lung Cancer Growth and Progression and Predicts Sensitivity to Dasatinib, *American journal of respiratory and critical care medicine* 200 (7) (2019) 888-899.
- [23] X. Yue, J. Ai, Y. Xu, Y. Chen, M. Huang, X. Yang, B. Hu, H. Zhang, C. He, X. Yang, W. Tang, X. Peng, L. Dong, H. Wang, J. Fan, J. Ding, M. Geng, Polymeric immunoglobulin receptor promotes tumor growth in hepatocellular carcinoma, *Hepatology (Baltimore, Md.)* 65 (6) (2017) 1948-1962.
- [24] I. Yajima, M. Y. Kumasaka, Y. Naito, T. Yoshikawa, H. Takahashi, Y. Funasaka, T. Suzuki, M. Kato, Reduced GNG2 expression levels in mouse malignant melanomas and human melanoma cell lines, *Am J Cancer Res* 2 (3) (2012) 322-9.
- [25] J. F. Tang, Z. H. Yu, T. Liu, Z. Y. Lin, Y. H. Wang, L. W. Yang, H. J. He, J. Cao, H. L. Huang, G. Liu, Five miRNAs as novel diagnostic biomarker candidates for primary nasopharyngeal carcinoma, *Asian Pac J Cancer Prev* 15 (18) (2014) 7575-81.
- [26] X. M. Hu, X. H. Yan, Y. W. Hu, J. L. Huang, S. W. Cao, T. Y. Ren, Y. T. Tang, L. Lin, L. Zheng, Q. Wang, miRNA-548p suppresses hepatitis B virus X protein associated hepatocellular carcinoma by downregulating oncoprotein hepatitis B x-interacting protein, *Hepatol Res* 46 (8) (2016) 804-15.
- [27] Y. Liang, X. Song, Y. Li, P. Su, D. Han, T. Ma, R. Guo, B. Chen, W. Zhao, Y. Sang, N. Zhang, X. Li, H. Zhang, Y. Liu, Y. Duan, L. Wang, Q. Yang, circKDM4C suppresses tumor progression and attenuates doxorubicin resistance by regulating miR-548p/PBLD axis in breast cancer, *Oncogene* 38 (42) (2019) 6850-6866.
- [28] Y. Bai, G. Zhang, R. Cheng, R. Yang, H. Chu, CASC15 contributes to proliferation and invasion through regulating miR-766-5p/ KLK12 axis in lung cancer, *Cell cycle (Georgetown, Tex.)* 18 (18) (2019) 2323-2331.
- [29] B. Jia, L. Xia, F. Cao, The role of miR-766-5p in cell migration and invasion in colorectal cancer, *Exp Ther Med* 15 (3) (2018) 2569-2574.



- [30] X. Z. Dong, Z. R. Zhao, Y. Hu, Y. P. Lu, P. Liu, L. Zhang, LncRNA COL1A1-014 is involved in the progression of gastric cancer via regulating CXCL12-CXCR4 axis, *Gastric Cancer* 23 (2) (2020) 260-272.
- [31] Y. Guo, J. Zhai, J. Zhang, C. Ni, H. Zhou, Improved Radiotherapy Sensitivity of Nasopharyngeal Carcinoma Cells by miR-29-3p Targeting COL1A1 3'-UTR, *Medical science monitor: international medical journal of experimental and clinical research* 25 (2019) 3161-3169.
- [32] J. Luan, J. Wang, Q. Su, X. Chen, G. Jiang, X. Xu, Meta-analysis of the differentially expressed microRNA profiles in nasopharyngeal carcinoma, *Oncotarget* 7 (9) (2016) 10513-21.
- [33] L. Liu, N. Bi, L. Wu, X. Ding, Y. Men, W. Zhou, L. Li, W. Zhang, S. Shi, Y. Song, L. Wang, MicroRNA-29c functions as a tumor suppressor by targeting VEGFA in lung adenocarcinoma, *Mol Cancer* 16 (1) (2017) 50.