# Identification of Key Pathways and Genes in Nasopharyngeal Carcinoma Based on Bioinformatics Analysis

**Hao Li**
Huazhong University of Science and Technology Tongji Medical College First Clinical College: Wuhan Union Hospital

**Shimin Zong**
Huazhong University of Science and Technology Tongji Medical College First Clinical College: Wuhan Union Hospital

**Yingying Wen**
Huazhong University of Science and Technology Tongji Medical College First Clinical College: Wuhan Union Hospital

**Peiyu Du**
Huazhong University of Science and Technology Tongji Medical College First Clinical College: Wuhan Union Hospital

**Wenting Yu**
Huazhong University of Science and Technology Tongji Medical College First Clinical College: Wuhan Union Hospital

**Nan Wu**
Huazhong University of Science and Technology Tongji Medical College First Clinical College: Wuhan Union Hospital

**Gang Zong**
Huazhong University of Science and Technology Tongji Medical College First Clinical College: Wuhan Union Hospital

**Hongjun Xiao** ( xhjent_whxh@hust.edu.cn )
Department of Otorhinolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China.

---

**Research**

**Keywords:** nasopharyngeal carcinoma, bioinformatics, differentially expressed genes, signaling pathway enrichment analysis

**Posted Date:** December 31st, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-135813/v1

# Abstract

**Purpose:** The purpose of this study is to identify novel molecular markers and potential molecular targets for NPC based on bioinformatics analysis.

**Methods:** We used bioinformatics to analyze one miRNA and two mRNA expression microarray datasets from the Gene Expression Omnibus database. The study included nasopharyngeal tissue samples from 57 patients with NPC and 32 patients without NPC. Fifty-one screened differentially expressed genes (DEGs) were evaluated by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) signal pathway enrichment analyses, and a protein-protein interaction (PPI) network was constructed.

**Results:** The GO analysis results showed that the DEGs were mainly related to cell cycle checkpoints, cell division, and DNA synthesis during DNA repair. The KEGG analysis results suggested that the DEGs were mainly associated with extracellular matrix receptor interactions. In the PPI network, we identified RAD51AP1, MAD2L1, SPP1, CCNE2, CNTNAP2, and MELK as hub genes, clustered a key module, and identified eight key transcription factors: TFII-I, Pax-5, STAT4, GR-alpha, YY1, C/EBPβ, GRβ, and TFIID.

**Conclusion:** The hub genes and signaling pathways identified above may play an important role in NPC development and provide ideas for the selection of valuable prognostic markers and the development of new molecular-targeted drugs.


# Introduction

Nasopharyngeal carcinoma (NPC) is a malignant epithelial cell tumor that originates from the surface of the nasopharyngeal mucosa. NPC has a typical geographical distribution and its incidence in East and Southeast Asia is significantly higher than that in other parts of the world [1]. According to data from the International Agency for Research on Cancer, there were approximately 129 000 new cases of NPC in the world in 2018, accounting for 0.7% of all confirmed cancer cases. However, more than 70% of new cases come from East and Southeast Asia [2, 3]. Therefore, for people in East Asia and Southeast Asia, NPC still has a wider impact in tumor diseases. Radiotherapy combined with chemotherapy is currently the main treatment for NPC. This treatment has a good therapeutic effect on patients with early NPC, but the efficacy is limited in NPC patients with distant metastases [4]. In addition, due to the concealed onset of NPC and the lack of reliable early screening methods, many patients show distant metastases when symptoms appear; therefore, new and sensitive early screening or detection methods and new treatments need to be developed urgently.

With a deepened understanding of the molecular mechanism of NPC, many molecular markers related to treatment effects or prognosis have been reported, including Epstein-Barr virus (EBV) DNA, lactate dehydrogenase, and vascular endothelial growth factor VEGF [5–7]. At present, several institutions are using EBV DNA clinically for the efficacy monitoring and prognostic analysis of NPC. However, its positive predictive value for tumor screening is relatively low (11%) [8, 9]. Therefore, it is necessary to explore the molecular mechanism of NPC development in order to identify more sensitive molecular markers that can

be used for early screening. Bioinformatics is an important strategy for studying molecular mechanisms related to tumors and can provide clues and directions for discovering unknown molecular mechanisms, and new drug targets for malignant tumors.

In this study, we aimed to use bioinformatics to identify potential molecular markers for NPC. We downloaded three microarray datasets from the Gene Expression Omnibus (GEO) database and analyzed the data using a bioinformatics method to identify differentially expressed genes (DEGs) between NPC tissues and normal nasopharyngeal tissues. We further analyzed signaling pathways and hub genes related to NPC, which may be helpful for an early diagnosis, molecular targeted therapy, and prognostic analysis of NPC.

## Materials And Methods

# 2.1 Microarray data

Three microarray datasets were downloaded from the GEO database (GSE12452, GSE53819, and GSE22587). GSE12452 is an mRNA expression profile based on the GPL570 platform that contains a total of 41 samples, of which 31 are tumor tissue samples from NPC patients and 10 are nasopharyngeal tissue samples from healthy people. The detection chip used was the [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. GSE53819 is an mRNA expression profile based on the GPL6480 platform that contains 36 samples, of which 18 are tumor tissue samples from NPC patients and 18 are nasopharyngeal tissue samples from healthy people. The detection chip used was the Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F. GSE22587 is an miRNA expression profile based on the GPL8933 platform. It contains a total of twelve samples, of which eight are tumor tissue samples from NPC patients and four are nasopharyngeal tissue samples from healthy people. The detection chip used was the Illumina Human Beta-version microRNA expression BeadChip.

# 2.2 Data processing and DEG identification

The data from the three microarray datasets were analyzed by the limma and impute packages in R3.6.2 software (https://www.r-project.org/). A P-value < 0.05 and | log FC | ≥ 1 were used as criteria to determine the differential expression of mRNA in GSE53819 and GSE12452. The cluster heatmap of DEGs was drawn using gplots and RColorBrewer in R3.6.2 software. A P-value < 0.04 and | log FC | ≥ 1.5 were used as criteria to determine the differential expression of miRNA in GSE22587. In order to analyze the differentially expressed miRNA and mRNA together, we used the miRDB online tool (http://www.mirdb.org/) to predict the target mRNA of differentially expressed miRNA and built a network with Cytoscape3.7.2 software (http://cytoscape.org/). Finally, we identified the DEGs simultaneously present in all three microarray datasets by drawing a Venn diagram.

# 2.3 Functional and signal pathway enrichment analysis of DEGs

The identified DEGs were subjected to functional and pathway enrichment analysis through Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://www.david.abcc.ncifcrf.gov/). Gene Ontology (GO) terms included in the biological process (BP), cellular component (CC), and molecular function (MF) categories were selected, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database was used. Values of $P < 0.05$ were considered statistically significant.

## 2.4 Protein-protein interaction (PPI) network construction and hub gene identification

The online tool STRING (http://string-db.org/) was used to explore the relative interactions of DEGs. Interaction pairs with a combined score > 0.4 were considered significant and were reserved. Cytoscape3.7.2 software was used to construct the PPI network and create a picture. Subsequently, the Cytoscape plug-in Network Analyzer was used to further analyze the topological attributes of the PPI network and identify the hub genes. The MCODE tool in Cytoscape was used to screen the high-score modules in the PPI network (the screening criteria for each standard were a node score = 0.2, K-Core = 2, and degree = 2).

## 2.5 Transcription factor regulation network of hub genes

The UCSC Genome Browser (http://genome.ucsc.edu/) and PROMO database (http://alggen.lsi.upc.es/cgibin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) were used to predict transcription factors of the hub genes (the regions 2000 nt upstream and 100 nt downstream of the transcription start point were selected as the promoter regions, the PROMO database was used to predict the transcription factor, and the maximum fault tolerance was set to 0%). Cytoscape3.7.2 software was applied to construct a network diagram, and transcription factors that could simultaneously regulate six hub genes were shown in the form of a table.

# Results

## 3.1 DEG identification

One up-regulated and eight down-regulated miRNAs of differential expression were screened from GSE22587 according to the preset criteria: hsa-miR-18a, hsa-miR-642, hsa-miR-34c, hsa-miR-34b, hsa-miR-625, hsa-miR-576, hsa-miR-92b, hsa-miR-596, and hsa-miR-497. A network of related targets containing nine miRNAs and 3718 different targets was constructed by selecting predicted targets with scores $\geq 60$ (Fig. 1a). Seven hundred and fifty differentially expressed mRNAs were screened from GSE12452, of which 280 increased and 470 decreased in expression. Two thousand and seventy-eight differentially expressed mRNAs were screened from GSE53819, of which 864 increased and 1,214 decreased in expression. In Fig. 1b, DEGs in NPC tissues and normal tissues obtained from the two microarray datasets are displayed in the form of clustering heatmaps. From the Venn diagram (Fig. 2), 51 mRNAs in the intersection of the above three datasets were identified as DEGs. Among those 51 DEGs, 27 were up-regulated and 24 were down-regulated (Table 1).

Table 1
**Fifty-one DEGs were selected from three microarray datasets**

| DEGs | Gene symbol |
|---|---|
| Up-regulated (27) | MAD2L1 MELK ZNRF3 CCNE2 CXCL10 GBP5 KCTD3 TNFSF15 ROBO1 PUS7 CNTNAP2 LHX2 GAD1 PAPSS2 ZIC2 NUAK1 COL5A2 SPP1 BMP2 VASH2 RBBP8 FCGR3B RAD51AP1 SYNPO2 ITGAV NEDD1 MEST |
| Down-regulated (24) | ARMC3 C11orf88 CAPS2 CCDC89 CHL1 CP DNAH5 ENKUR FAM107B GABRP GDA KCNE1 LRRC34 LRRC46 MS4A1 OSBPL6 PLEKHH1 RBM24 SDR16C5 SNX31 TCTEX1D1 TTC9 UBXN10 UPK1B |

DEGs: differentially expressed genes.

# 3.2 Functional and pathway enrichment analysis of DEGs

The result of the functional and pathway enrichment analysis of DEGs is shown in Fig. 3, containing the BP, CC, and MF terms of the GO and KEGG pathways. In Fig. 3, the counts of the DEGs in each term or pathway are presented as a histogram, and the $-\log_{10}$ P-values of each term or pathway are presented as a broken line graph. Table 2 illustrates the detailed data from the GO and KEGG pathway analyses. Among these results, we found that cell adhesion and cell cycle checkpoint were the most important terms of BP, and most of the DEGs in those terms were up-regulated. Cell surface was a representative term in the CC category, and most of the DEGs in those terms were up-regulated. Retinol dehydrogenase activity was the only term that met the criteria of the MF category. Extracellular matrix (ECM)-receptor interaction was the critical signaling pathway identified in the KEGG pathway analysis. The DEGs involved in this pathway were up-regulated.

Table 2
GO and KEGG pathway enrichment analysis of DEGs

| Term | Count | P-value | Genes |
|---|---|---|---|
| **Biological process** | | | |
| GO:0007155 ~ cell adhesion | 6 | 0.005522 | ROBO1, NUAK1, ITGAV, CNTNAP2, CHL1, SPP1 |
| GO:0000075 ~ cell cycle checkpoint | 2 | 0.022292 | CCNE2, RBBP8 |
| GO:0051301 ~ cell division | 4 | 0.056741 | CCNE2, NEDD1, MAD2L1, RBBP8 |
| GO:0006954 ~ inflammatory response | 4 | 0.068641 | BMP2, GBP5, SPP1, CXCL10 |
| GO:0000731 ~ DNA synthesis involved in DNA repair | 2 | 0.084001 | RAD51AP1, RBBP8 |
| **Cellular component** | | | |
| GO:0009986 ~ cell surface | 5 | 0.035544 | BMP2, ROBO1, ITGAV, KCNE1, CNTNAP2 |
| GO:0009897 ~ external side of plasma membrane | 3 | 0.086366 | ITGAV, MS4A1, CXCL10 |
| **Molecular function** | | | |
| GO:0004745 ~ retinol dehydrogenase activity | 2 | 0.040799 | BMP2, SDR16C5 |
| GO:0050840 ~ extracellular matrix binding | 2 | 0.058407 | ITGAV, SPP1 |
| GO:0005125 ~ cytokine activity | 3 | 0.062296 | BMP2, TNFSF15, SPP1 |
| GO:0046332 ~ SMAD binding | 2 | 0.094784 | BMP2, COL5A2 |
| **KEGG pathway** | | | |
| hsa04512:ECM-receptor interaction | 3 | 0.023518 | ITGAV, COL5A2, SPP1 |
| hsa04151:PI3K-Akt signaling pathway | 4 | 0.066779 | CCNE2, ITGAV, COL5A2, SPP1 |

# 3.3 PPI network construction and hub gene identification

We added 51 DEGs to the online STRING database to obtain the PPI network based on those genes (Fig. 4a). Fifty-one nodes and 22 degrees were screened out by setting the bar to a combined score > 0.4. A degree ≥ 3 was selected as the screening criteria for hub genes. We obtained six hub genes, shown in Fig. 4b: *RAD51AP1* (degree = 4, Closeness centrality = 4.5, Betweenness centrality = 8), *MAD2L1* (degree = 4, Closeness centrality = 4.5, Betweenness centrality = 8), *SPP1* (degree = 4, Closeness centrality = 5, Betweenness centrality = 26), *CCNE2* (degree = 3, Closeness centrality = 4, Betweenness centrality = 0),

*CNTNAP2* (degree = 3, Closeness centrality = 4.3, Betweenness centrality = 22), and *MELK* (degree = 3, Closeness centrality = 4, Betweenness centrality = 0).

In addition, MCODE, a cystoscope tool, was used to screen for significant models (Fig. 4c). A significant model was obtained containing four genes: *MAD2L1*, *MELK*, *CCNE2*, and *RAD51AP1*. Those four genes belonged to a group of previously obtained hub genes, which might indicate that these hub genes hold an important position in the PPI network.

## 3.4 Construction of a transcriptional regulatory network for the hub genes

Transcription factors of hub genes were predicted through the UCSC Genome Browser online tool and PROMO database. *RAD51AP1*, *MAD2L1*, *SPP1*, *CCNE2*, *CNTNAP2*, and *MELK* were regulated by 20, 21, 17, 18, 19, and 24 transcription factors, respectively. Following which, a transcriptional regulatory network for hub genes was constructed using Cytoscape3.7.2 software (Fig. 5). Among the transcription factors, eight (TFII-I, Pax-5, STAT4, GR-alpha, YY1, C/EBPbeta, GR-beta, and TFIID) were capable of regulating all of the hub genes.

# Discussion

Studies have shown that DNA replication and damage repair genes may promote cancer. *RAD51AP1* regulates DNA repair processes as a critical part of RAD51-mediated homologous recombination, which suggests that *RAD51AP1* is particularly relevant in cancer. Research has shown that overexpression of *RAD51AP1* is found in many cancers, such as human non-small cell lung cancer (NSCLC), ovarian cancer, and breast cancer [10–13]. Down-regulating *RAD51AP1* not only inhibits epithelial-mesenchymal transformation and metastasis of NSCLC in vivo but also reduces the proliferation of ovarian and lung cancer cells in vitro [11, 12]. A high *RAD51AP1* expression level has been detected in the blood of patients with ovarian cancer and in human breast tumor tissue, and patients with overexpression of *RAD51AP1* have inferior overall survival [12, 13]. These studies suggest that *RAD51AP1* can promote tumor growth by regulating DNA repair processes and that it is related to an inferior overall survival. This gene may play an important role in NPC and can be used for screening, identifying new target drugs, and predicting prognosis.

As part of the spindle checkpoint, *MAD2L1* plays an important supervisory role in chromosome segregation during mitosis. Abnormal *MAD2L1* expression can promote cancer by causing aneuploids and chromosome structure variation. It has been found that *MAD2L1* is overexpressed in breast cancer cells, human liver cancer cells, and AGS and BGC-823 gastric cancer cells [14–16]. *MAD2L1* is significantly correlated with various tumor markers (including ER, P53, HER-2, and KI-67) in breast cancer [14]. Some scholars found that the proliferation of human liver cancer cells and AGS and BGC-823 gastric cancer cells could be inhibited by down-regulating *MAD2L1*. *MAD2L1* is also associated with cell cycle arrest [15, 16]. Overexpression of *MAD2L1* in patients with lung adenocarcinoma has been shown to be negatively correlated with overall survival and relapse-free survival [17]. Due to its involvement in the

development of cancer through cell cycle regulation and its association with overall survival and relapse-free survival, *MAD2L1* has the potential to be new tumor marker.

In recent years, researchers have attached more importance to the role of SPP1 (osteopontin) in the development of cancer and have found that SPP1 can regulate cell adhesion, migration, invasion, and chemotaxis through the integrin receptor and glycoprotein CD44 [18]. Research has also shown that the overexpression of *SPP1* in ovarian cancer cells promotes cell proliferation, migration, and invasion by activating the integrin β1/FAK/AKT signaling pathway. Silencing *SPP1* could inhibit the integrin β1/FAK/AKT signaling pathway and reduce cell proliferation, migration, and invasion [19]. In addition, the down-regulation of *SPP1* increases the sensitivity of cervical cancer cells to cisplatin by inhibiting the PI3K/AKT signaling pathway [20]. These studies show that *SPP1* is involved in the development of cancer and can influence many signaling pathways and the sensitivity of cancer cells to drugs. *SPP1* may become a new therapeutic target for patients with NPC.

The protein encoded by *CCNE2* belongs to a highly conserved cyclin family that has been reported to include key proteins of the G1/S transition during the cell cycle. The expression level of this gene is significantly increased in tumor cells [21]. Some scholars found that the up-regulation of Mir-3607-3p could inhibit cell proliferation and migration by inhibiting the expression of *CCNE2* [22]. The overexpression of *CCNE2* in HER2 + breast cancer cells may result in resistance to trastuzumab [23]. *CCEN2* is involved in cell cycle and can influence cell proliferation and migration. It can even lead to a drug-resistant cancer, which indicates that it may be a new target for treating drug-resistant cancer.

The contactin-associated protein encoded by *CNTNAP2* is a member of the axon protein family. It plays an important role as a cell adhesion molecule and receptor in the nervous system of vertebrates. It is associated with a variety of neurodevelopmental disorders, including autism, epilepsy, and intellectual disability [24–26]. Scientists are paying more attention to the role of *CNTNAP2* in nervous system diseases, and its role in cancer requires further studies.

The protein encoded by *MELK* is a cyclin-dependent kinase, And an overexpression of *MELK* was reported in a variety of tumors [27–29]. Researchers found that *MELK* can specifically phosphorylate Bcl-G, a pro-apoptotic member of the Bcl-2 family that inhibits the apoptosis induced by Bcl-GL and promotes the development of breast cancer in vitro [27]. *MELK* plays an important role in gastric cancer metastasis through the FAK/Paxillin pathway [29]. MELK also regulates the cell cycle by phosphorylating FOXM1 to activate FoxM1-induced gene expression in glioma stem cells [30]. In both head and neck squamous cell carcinoma, the knockdown of *MELK* inhibits *SOX2* expression and thus inhibits the proliferation of cancer cells [31]. Therefore, it may be a potential target for novel future drugs.

In this study, ECM-receptor interaction was the critical signaling pathway identified by KEGG pathway analysis. Additionally, the hub gene *SPP1* is involved in this signaling pathway, which indicates that signaling pathways play an important role in NPC. ECM-receptor interaction is critical for cellular activities such as adhesion, migration, differentiation, proliferation, and apoptosis [32]. It has been reported that the ECM-receptor is closely associated with gastric cancer, colorectal cancer, and renal

cancer [33, 34], and that the ECM may play a role in promoting the epithelial-mesenchymal transformation of cancer cells in colorectal cancer [34]. In renal cell carcinoma cells, Twist2 promotes proliferation and invasion by regulating the expression of *ITGA6* and *CD44*, which are a part of the ECM-receptor interaction pathway [35]. This signaling pathway may influence the development of NPC by promoting epithelial-mesenchymal transformation, and regulation of this pathway may lead to a new treatment.

The significant model obtained in this study contained four genes: *MAD2L1*, *MELK*, *CCNE2*, and *RAD51AP1*. Those four genes belonged to the group of hub genes, indicating that the hub genes play a critical role in the PPI network. Furthermore, this model is associated with the cell cycle and may suggest a new therapeutic direction. The eight key transcription factors identified in this study may be clear indicators associated with hub gene regulation.

There are obvious differences between this study and that of two previous bioinformatics studies on key NPC genes [36, 37]. The most important is the original microarray datasets used for analysis are different, and therefore got different hub genes. Compared with a study published by Ye [36] which included an expression profile of an NPC cell line, the microarray datasets selected in this study were all derived from human tissue, which could thus more accurately reflect the conditions of clinical patients. Compared with a study published by Zhu [37] in which only hub genes and pathways were investigated, we not only focused on hub genes and pathways but also analyzed the transcription factors of hub genes. Our findings may provide direction for further research of hub gene regulation. Most importantly, our study identified new hub genes, *RAD51AP1*, *MAD2L1*, *SPP1*, *CCNE2*, *CNTNAP2*, and *MELK*, which may be complementary to the results of the two studies and provide the way forward for understanding and treating NPC.

In summary, we obtained 51 DEGs from NPC and normal nasopharyngeal tissue samples by using bioinformatics. In addition, we found that cell adhesion and the cell cycle checkpoint were the most important BP terms. ECM-receptor interaction was the critical signaling pathway identified in our KEGG pathway analysis and this might be playing a role in the development of NPC. The six hub genes (*RAD51AP1*, *MAD2L1*, *SPP1*, *CCNE2*, *CNTNAP2*, and *MELK*) that were screened out have the potential to be novel markers and may play an important role in the screening, treatment, and prognosis of NPC. Finally, we predicted eight transcription factors, which may serve as clear indicators associated with hub gene regulation. The results of this study may provide ideas for the selection of early screening markers for NPC and the development of new molecular-targeted drugs. However, the specific role of these hub genes in NPC need further experimental studies.

# Declarations

**Funding:** This work was supported by the National Natural Science Foundation of China [Grant numbers 81771002].

**Conflicts of interest:** All authors declare that there is no conflict of interest.

**Ethics approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

**Consent to participate:** Not applicable.

**Consent for publication:** Not applicable.

**Availability of data and material:** The datasets (GSE12452, GSE53819, and GSE22587) analysed during the current study are available in the GEO database.

**Code availability:** The download address of software used during this study is included in this article.

**Authors' contributions:**

Hao Li contributed to the conception of the study;

Shimin Zong and Hao Li contributed significantly to analysis and manuscript preparation;

Yingying Wen, Wenting Yu and Peiyu Du performed the data analyses and wrote the manuscript;

Hongjun Xiao and Gang Zhong corrected the manuscript.

**Acknowledgments**

This work was supported by a grant from the National Natural Science Foundation of China [Grant numbers 81771002].

# References

1. Lindsey A Torre, Freddie Bray, Rebecca L Siegel, et al. Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2015;65(2): p. 87-108.
2. Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6): p. 394-424.
3. Yu-Pei Chen, Anthony T C Chan, Quynh-Thu Le, et al. Nasopharyngeal carcinoma. The Lancet. 2019;394(10192): p. 64-80.
4. David Adelstein, Maura L Gillison, David G Pfister, et al. NCCN Guidelines Insights: Head and Neck Cancers, Version 2.2017. J Natl Compr Canc Netw. 2017;15(6): p.761-770.
5. Sing-fai Leung, Benny Zee, Brigette B Ma, et al. Plasma Epstein-Barr viral deoxyribonucleic acid quantitation complements tumor-node-metastasis staging prognostication in nasopharyngeal carcinoma. J Clin Oncol. 2006;24(34): p. 5414-8.
6. Selahattin Turen, Enis Ozyar, Kadri Altundag, et al. Serum lactate dehydrogenase level is a prognostic factor in patients with locoregionally advanced nasopharyngeal carcinoma treated with

chemoradiotherapy. Cancer Invest. 2007;25(5): p. 315-21.

7. Xing Lv 1, Yan-Qun Xiang, Su-Mei Cao, et al. Prospective validation of the prognostic value of elevated serum vascular endothelial growth factor in patients with nasopharyngeal carcinoma: more distant metastases and shorter overall survival after treatment. Head Neck.2011;33(6): p. 780-5.
8. Kelly Y Kim, Quynh-Thu Le, Sue S Yom, et al. Clinical Utility of Epstein-Barr Virus DNA Testing in the Treatment of Nasopharyngeal Carcinoma Patients. International Journal of Radiation Oncology Biology Physics. 2017;98(5): p. 996-1001.
9. K C Allen Chan, John K S Woo, Ann King, et al. Analysis of Plasma Epstein–Barr Virus DNA to Screen for Nasopharyngeal Cancer. New England Journal of Medicine. 2017;377(6): p. 513-522.
10. Pires E, Sung P, Wiese C. Role of RAD51AP1 in homologous recombination DNA repair and carcinogenesis. DNA Repair (Amst).2017;59: p. 76-81.
11. Yuanyuan Wu, Huifeng Wang, Lijiao Qiao, et al. Silencing of RAD51AP1 suppresses epithelial-mesenchymal transition and metastasis in non-small cell lung cancer. Thorac Cancer. 2019;10(9): p. 1748-1763.
12. Dimple Chudasama, Valeria Bo, Marcia Hall, et al. Identification of cancer biomarkers of prognostic value using specific gene regulatory networks (GRN): a novel role of RAD51AP1 for ovarian and lung cancers. Carcinogenesis. 2018;39(3): p. 407-417.
13. Rajneesh Pathania, Sabarish Ramachandran, Gurusamy Mariappan, et al. Combined Inhibition of DNMT and HDAC Blocks the Tumorigenicity of Cancer Stem-like Cells and Attenuates Mammary Tumor Growth. Cancer Res. 2016;76(11): p. 3224-35.
14. Xiao-Fei Zhu, Min Yi, Juan He, et al. Pathological significance of MAD2L1 in breast cancer: an immunohistochemical study and meta-analysis. Int J Clin Exp Pathol. 2017;10(9):9190-9201.
15. Yuanhang Li, Weijun Bai, Jianjun Zhang. MiR-200c-5p suppresses proliferation and metastasis of human hepatocellular carcinoma (HCC) via suppressing MAD2L1. Biomed Pharmacother. 2017;92: p. 1038-1044.
16. Yu Wang, Fenghui Wang, Jing He, et al. miR-30a-3p Targets MAD2L1 and Regulates Proliferation of Gastric Cancer Cells. Onco Targets Ther. 2019;12: p. 11313-11324.
17. Yuan-Xiang Shi, Tao Zhu, Ting Zou, et al. Prognostic and predictive values of CDK1 and MAD2L1 in lung adenocarcinoma. Oncotarget. 2016;7(51):85235-85243.
18. Pieter H Anborgh, Jennifer C Mutrie, Alan B Tuck, et al. Role of the metastasis-promoting protein osteopontin in the tumour microenvironment. J Cell Mol Med. 2010;14(8): p. 2037-44.
19. Biao Zeng, Min Zhou, Huan Wu, et al. SPP1 promotes ovarian cancer progression via Integrin beta1/FAK/AKT signaling pathway. Onco Targets Ther. 2018;11: p. 1333-1343.
20. Xing Chen, Dongsheng Xiong, Liya Ye, et al. SPP1 inhibition improves the cisplatin chemo-sensitivity of cervical cancer cell lines. Cancer Chemother Pharmacol.2019;83(4): p. 603-613.
21. J M Gudas, M Payton, S Thukral, et al. Cyclin E2, a novel G1 cyclin that binds Cdk2 and is aberrantly expressed in human cancers. Mol Cell Biol.1999;19(1):612-622.

22. Peng Gao, Huan Wang, Jiarui Yu, et al. miR-3607-3p suppresses non-small cell lung cancer (NSCLC) by targeting TGFBR1 and CCNE2. PLoS Genet.2018;14(12): p. e1007790.
23. Eduardo Tormo, Anna Adam-Artigues, Sandra Ballester, et al. The role of miR-26a and miR-30b in HER2+ breast cancer trastuzumab resistance and regulation of the CCNE2 gene. Sci Rep. 2017; 7: p. 41309.
24. Emanuel Lauber, Federica Filice, Beat Schwaller. Dysregulation of Parvalbumin Expression in the Cntnap2-/- Mouse Model of Autism Spectrum Disorder. Front Mol Neurosci. 2018;11: p. 262.
25. Emanuela Leonardi, Emanuela Dazzo, Maria Cristina Aspromonte, et al. CNTNAP2 mutations and autosomal dominant epilepsy with auditory features. Epilepsy Res.2018;139: p. 51-53.
26. Christiane Zweier, Eiko K de Jong, Markus Zweier, et al. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. Am J Hum Genet. 2009;85(5): p. 655-66.
27. Meng-Lay Lin, Jae-Hyun Park, Toshihiko Nishidate, et al. Involvement of maternal embryonic leucine zipper kinase (MELK) in mammary carcinogenesis through interaction with Bcl-G, a pro-apoptotic member of the Bcl-2 family. Breast Cancer Res. 2007; 9(1): p. R17.
28. Mark R Pickard, Andrew R Green, Ian O Ellis, et al. Dysregulated expression of Fau and MELK is associated with poor prognosis in breast cancer. Breast Cancer Res. 2009;11(4): p. R60.
29. Tao Du, Ying Qu, Jianfang Li, Hao Li, et al. Maternal embryonic leucine zipper kinase enhances gastric cancer progression via the FAK/Paxillin pathway. Mol Cancer.2014;13:100.
30. Kaushal Joshi, Yeshavanth Banasavadi-Siddegowda, Xiaokui Mo, et al. MELK-dependent FOXM1 phosphorylation is essential for proliferation of glioma stem cells. Stem Cells. 2013;31(6): p. 1051-63.
31. Lili Ren, Boya Deng, Vassiliki Saloura, et al. MELK inhibition targets cancer stem cells through downregulation of SOX2 expression in head and neck cancer cells. Oncol Rep. 2019; 41(4): p. 2540-2548.
32. A van der Flier, A Sonnenberg, Function and interactions of integrins. Cell Tissue Res. 2001;305(3): p. 285-98.
33. Ping Yan, Yingchun He, Kexin Xie, et al. In silico analyses for potential key genes associated with gastric cancer. PeerJ. 2018;6: p. e6092.
34. Nuh N Rahbari, Dmitriy Kedrin, Joao Incio, et al. Anti-VEGF therapy induces ECM remodeling and mechanical barriers to therapy in colorectal cancer liver metastases. Sci Transl Med. 2016;8(360): p. 360ra135.
35. Hao-Jie Zhang, Jing Tao, Lu Sheng, et al. Twist2 promotes kidney cancer cell proliferation and invasion by regulating ITGA6 and CD44 expression in the ECM-receptor interaction pathway. Onco Targets Ther.2016;9: p. 1801-12.
36. Zhimin Ye, Fangzheng Wang, Fengqin Yan, et al. Identification of candidate genes of nasopharyngeal carcinoma by bioinformatical analysis. Arch Oral Biol. 2019;106: p. 104478.

37. Hong-Ming Zhu, Qian Fei, Lu-Xi Qian, et al. Identification of key pathways and genes in nasopharyngeal carcinoma using bioinformatics analysis. Oncol Lett. 2019;17(5): p. 4683-4694.

# Figures

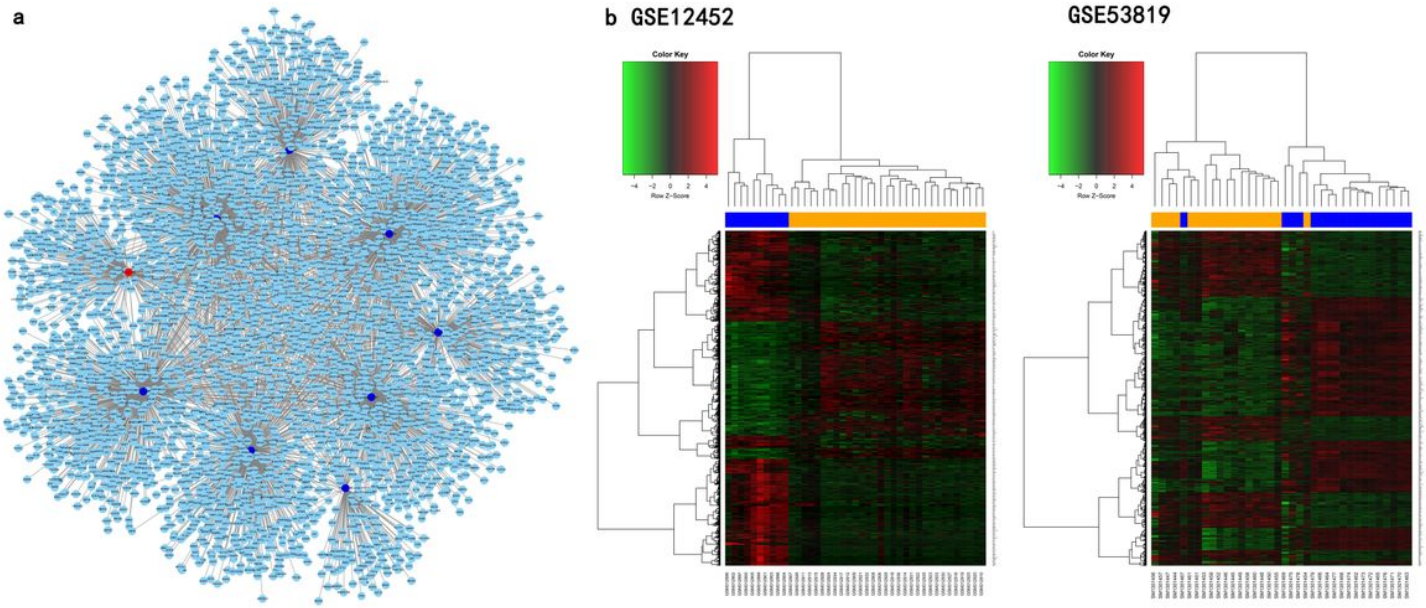



**Figure 1**

miRNA target network of GSE22587 (a) and heatmap of DEGs in GSE12452 (a) and GSE53819 (b) a. The red node represents up-regulated miRNA, the mazarine node represents down-regulated miRNA, and the blue node represents the target gene. b. The abscissa represents different samples; orange represents the cancer sample, blue represents the control sample, and the vertical axis represents clusters of DEGs. Red represents up-regulation, and green represents down-regulation.

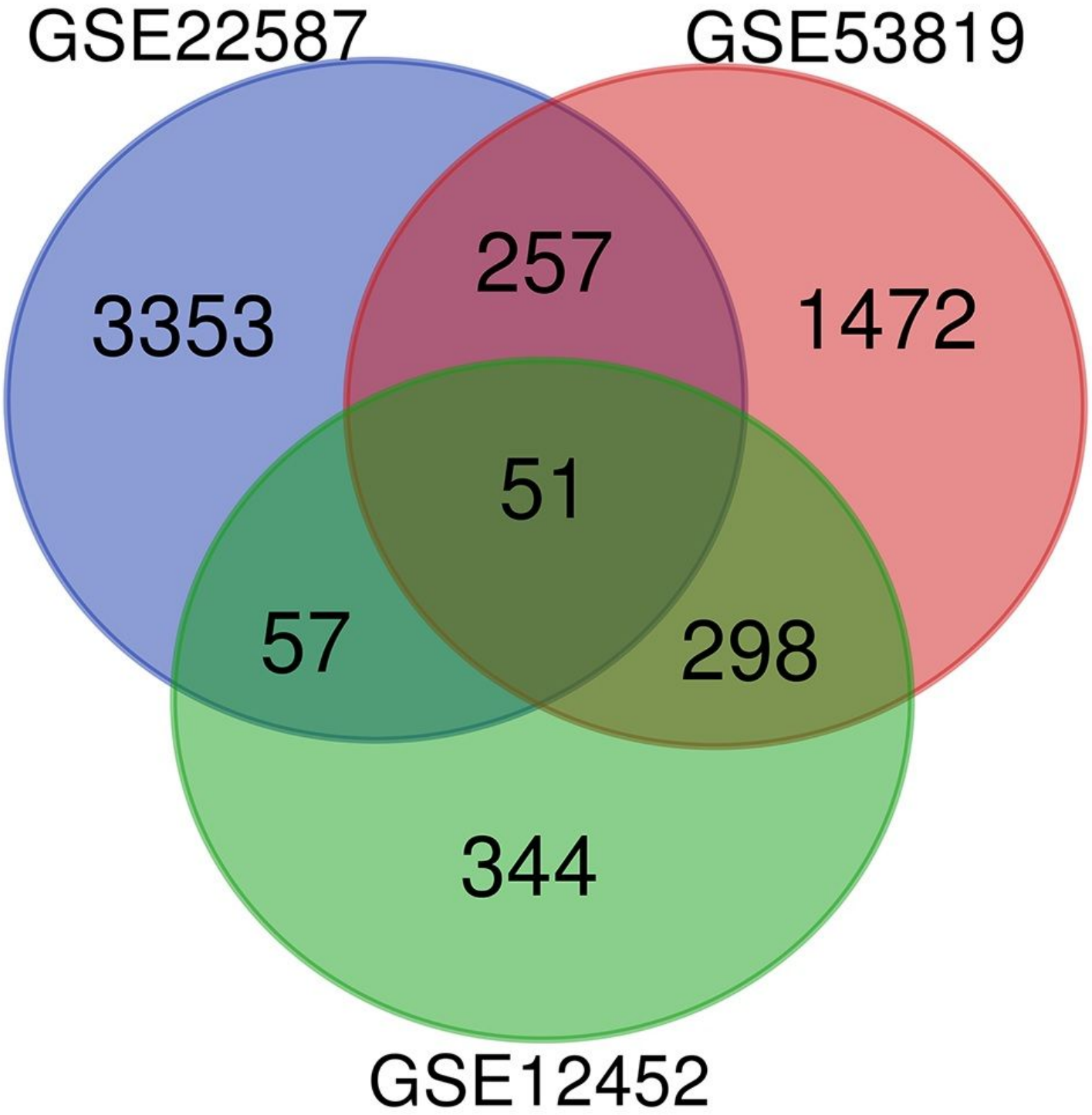
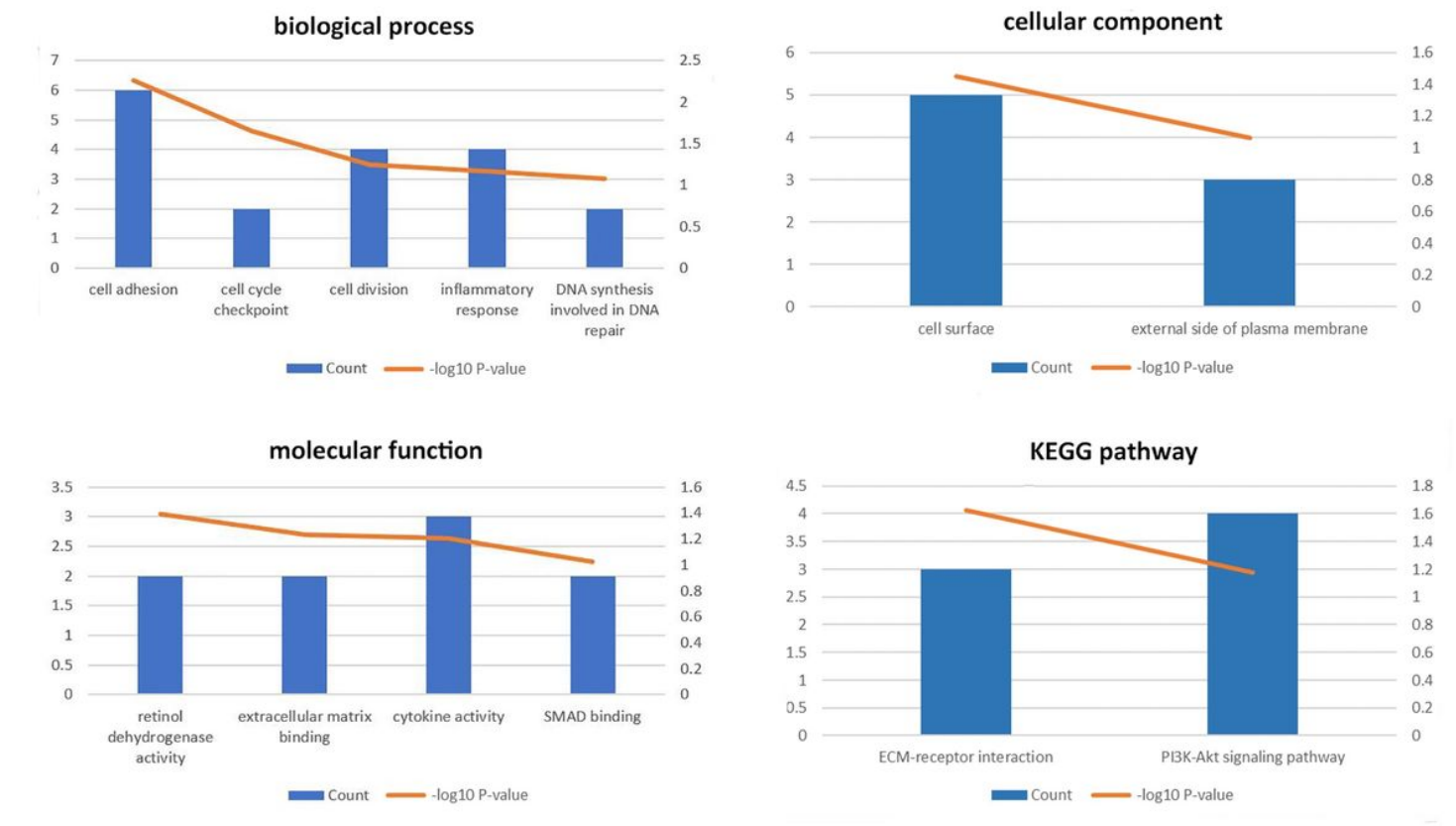


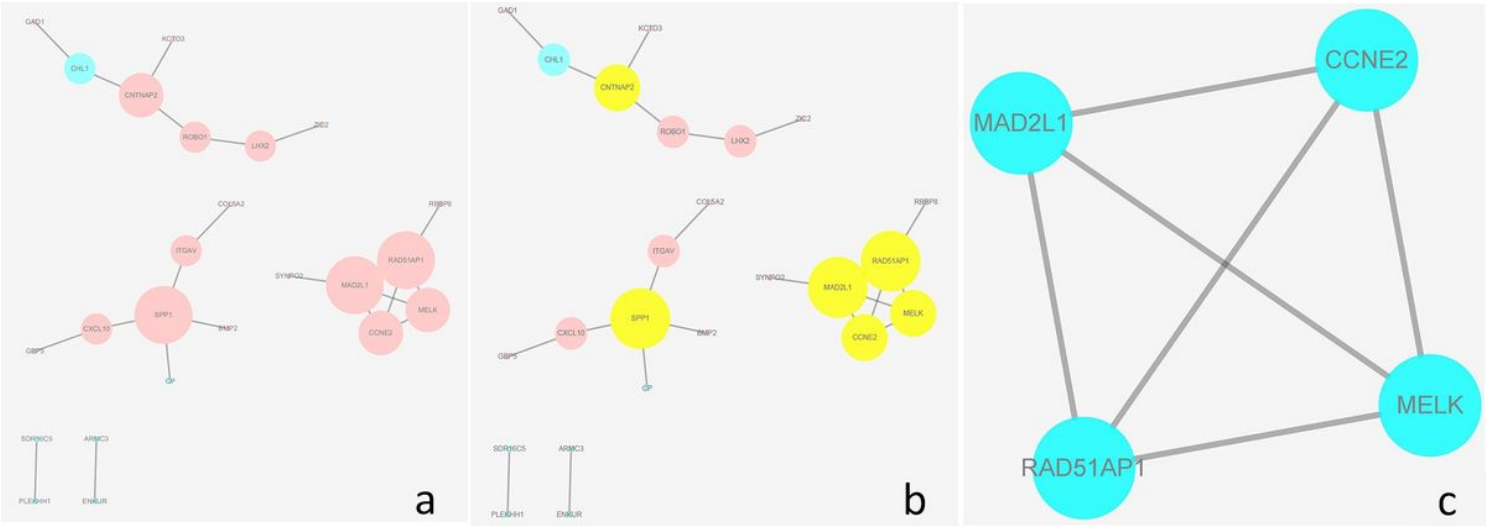

**Figure 2**

Venn diagram of three microarray datasets




## Figure 3

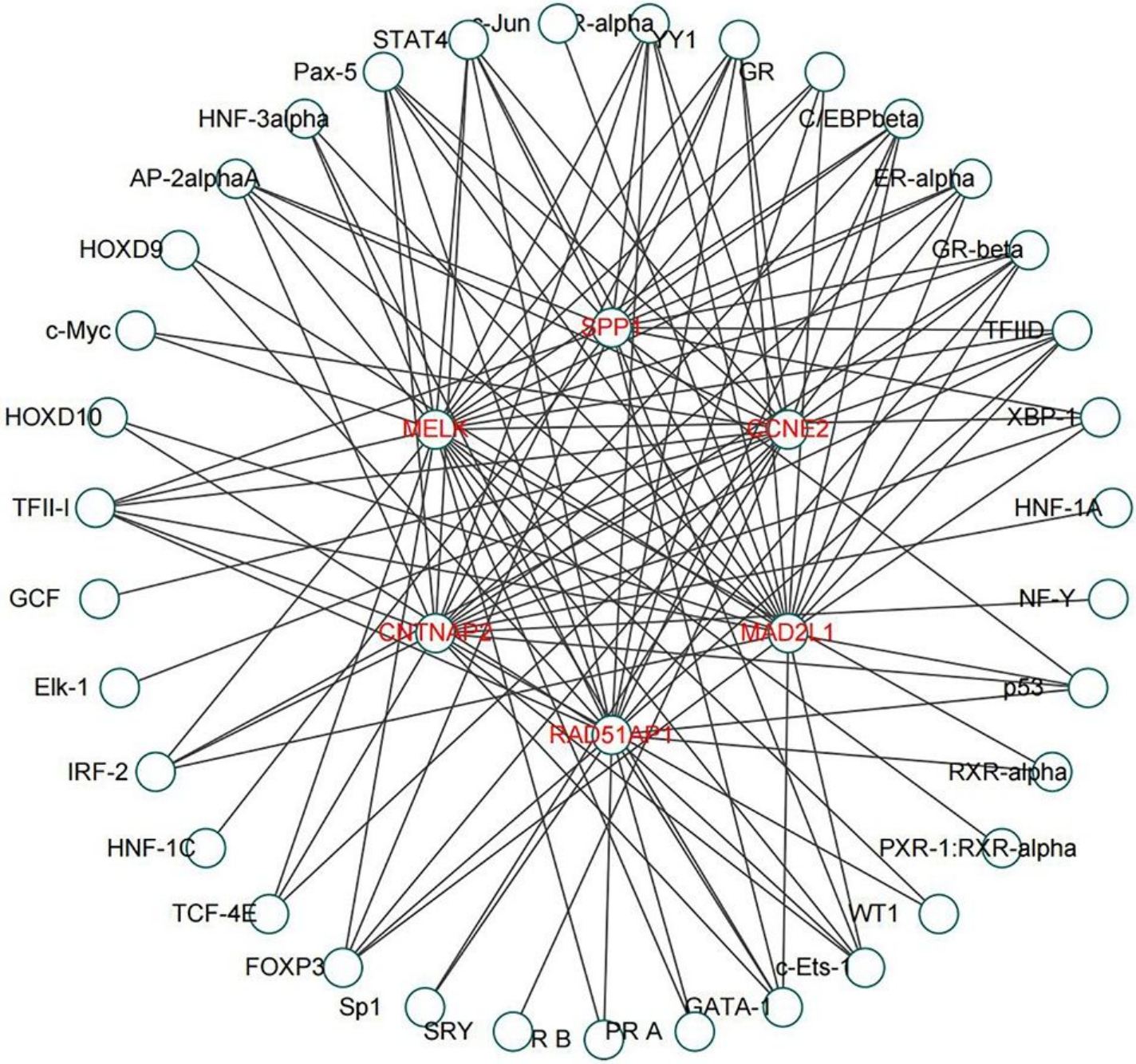
Functional and pathway enrichment analysis of DEGs The left-hand abscissa represents the count of DEGs in each term or pathway; the right-hand abscissa represents the -log10 P-value of the term or pathway.




## Figure 4

PPI network of DEGs a: The PPI network of 51 DEGs. Red nodes represent up-regulated genes; blue nodes represent down-regulated genes. Larger the degree, larger the diameter of the node. b: The PPI network

showing the hub genes. The yellow nodes represent hub genes. c: The significant model.




## Figure 5

The hub gene-transcriptional factor regulatory network Hub gene names are in red. Transcription factor names are in black.