Molecular Biology Reports (2019) 46:2829–2840
https://doi.org/10.1007/s11033-019-04729-3

ORIGINAL ARTICLE

The identification of key genes in nasopharyngeal carcinoma by bioinformatics analysis of high-throughput data

Yanshan Ge[1,2,3,4] · Zhengxi He[1,2,3,4] · Yanqi Xiang[5] · Dawei Wang[6] · Yuping Yang[7] · Jian Qiu[8] · Yanhong Zhou[1,2,3,4]

Received: 16 November 2018 / Accepted: 28 February 2019 / Published online: 4 March 2019

Abstract

Nasopharyngeal carcinoma (NPC) is a common pattern of regional malignancy in the south of China, especially in Guangdong province. The development of computerized tomography (CT) technology and the improvement of radiotherapy scheme can improve the survival rate of NPC patients. However, the prevalence and recurrence rate of NPC are increasing every year. It is urgent for us to uncover the molecular mechanism of NPC. In this study, we used scientific information retrieval from the GEO (gene expression omnibus) database to download the GSE12452, which contained 41 samples, including 31 nasopharyngeal carcinoma samples and 10 control samples. With the help of GO (gene ontology) analysis, KEGG (kyoto encyclopedia of genes and genomes) analysis, PPI (protein–protein interaction) network model construction, and WGCNA (weighted gene co-expression network analysis), we found 6896 differentially expressed genes, which affected the biological processes included cell cycle process, DNA metabolic process, DNA repairing, immune response, cell activation, regulation of immune system process, inflammatory response. The 20 hub genes present in front of us are SYK, PIK3CG, FYN, ACACB, LRRK2, RIPK4, RAC2, PIK3CD, PTPRC, LCR, RAD51, MAD2L1, CDK1, PCNA, GMPS, CCNB1, GAPDH, CCNA2, RFC4, TOP2A. In the future, these are the areas where we need to focus on the molecular mechanism of NPC.

Keywords Nasopharyngeal carcinoma · Bioinformatics analysis · Differentially expressed gene

✉ Yanhong Zhou
zhouyanhong@csu.edu.cn

1 The Key Laboratory of Carcinogenesis of the Chinese Ministry of Health, Xiangya Hospital, Central South University, Changsha 410078, Hunan, China

2 Basic School of Medicine, Central South University, Changsha 410078, Hunan, China

3 Cancer Research Institute, Central South University, Changsha 410078, Hunan, China

4 Key Laboratory of Carcinogenesis of Ministry of Health and Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Cancer Research Institute, Central South University, Changsha 410078, Hunan, China

5 Department of Nursing, the Second Xiangya Hospital, Central South University, 410078 Changsha, Hunan, China

6 Department of Gastrointestinal Surgery, Tengzhou City Center People's Hospital, Zaozhuang 277599, Shandong, China

7 Department of Emergency, Tengzhou City Center People's Hospital, Zaozhuang 277599, Shandong, China

8 Department of Emergency, Tancheng City Center People's Hospital, Linyi 276100, Shandong, China

Abbreviations

NPC	Nasopharyngeal carcinoma
EBV	Epstein–Barr virus
DEGs	Differentially expressed genes
GEO	Gene expression omnibus
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
PPI	Protein–protein interaction
RLE	Relative log expression

Introduction

Nasopharyngeal carcinoma (NPC) is the most common malignant tumor in southern China and South East Asia. It is primarily a highly tumor derived from the nasopharyngeal epithelium, with a strong tendency to metastasize [1]. The risk factors for NPC include Epstein–Barr virus (EBV) infection [2], the function of genetic factors, intake of salted food, and cigarette smoking [3–9]. Female patients account for about 30%, suggesting that NPC is more frequent in male than in female due to male smoking. Nowdays, radiation therapy is the preferred method currently found in the

treatment of NPC, because most of the NPC is poorly differentiated carcinoma. To increase the survival rate of NPC patients, surgical treatment, chemotherapy and radiotherapy are often used [10]. Compared with conventional chemotherapy, induction chrono-chemotherapy seemed to reduce chemotherapy-related toxicity and improve average local relapse time in patients treated with combined chemoradiotherapy for NPC [11].

Microarray gene expression is a frontier biotechnology. In this larger data age, it combines high throughput and simultaneously detects thousands of genes. Microarray gene expression dazzling features are automated, integrated, miniaturized. We used microarrays to identify down-regulated or up-regulated genes in NPC compared with non-malignant controls. In the powerful function of gene chip technology, our analysis found that many important key roles of differentially expressed genes, which played an important role in the initiation and development of NPC. Differentially expressed genes (DEGs) are the primary choice for the study of molecular targets and diagnostic markers. For example, EBV is associated with various forms of cancer, including malignant EBV immunoblastic lymphoma, harmful lung cancer, Hodgkin's lymphoma, Burkitt lymphoma, gastric cancer and NPC [12, 13]. Epstein–Barr virus latent protein expression was limited to EBNA1 nuclear antigen and latent membrane protein (LMP1, LMP2A, LMP2B) [14, 15]. Zhang et al. report a novel mechanism driving NPC metastasis through the EBV-encoded LMP1-mediated metabolic reprogramming, via activation of IGF1-mTORC2 signaling and nuclear acetylation of the Snail promoter by the PDHE1α, an enzyme involved in glucose metabolism [16]. Bose et al. demonstrated that anti-cancer gene ATM in NPC is down, Down-regulation of DEGs and tissue activity mediated by EBV infection in NPC cells [17]. In many cancers, DNA methylation is the most famous epigenetic marker that can serve as a biomarker of prognosis. Jiang et al. [18] proved that 6-hypermethylated gene panel was associated with poor survival in patients with NPC. In this study, we identified key DEGs related to NPC in tumors and normal samples, then we screened these differential genes according to statistical methods. We took advantage of the current very powerful analysis software and statistical methods including GO terminology, enrichment analysis, KEGG path analysis, PPI network, Gene co-expression network analysis. Finally, we found that NPC was closely related to the key genes and pathways. The 20 hub genes present in front of us were SYK, PIK3CG, FYN, ACACB, LRRK2, RIPK4, RAC2, PIK3CD, PTPRC, LCR, RAD51, MAD2L1, CDK1, PCNA, GMPS, CCNB1, GAPDH, CCNA2, RFC4, TOP2A.

Methods and materials

Gained microarray data

The gene expression database is a common functional genome database that supported MIAME-compliant data submission. The database provided a gene expression that helped the user query and download the experiment. First of all, we cautiously selected NPC-associated dataset GSE12452 from NCBI. Observing the download of this database included 31 NPC samples and 10 normal samples [19].

Identification of DEGs

We used a powerful statistical software R and some packages of Bioconductor to prominence analysis of DEGs between NPC patient and normal person. The affyPLM package could fit the original data of the chip, get the weights and residuals diagram, the relative log expression (RLE), the relative standard deviation (NUSE, Normalized unscaled standard errors) box diagram. Before we analyzed the data, we would have carefully quality testing of microarray data. In this process, we made use of some powerful and accurate R packets, such as packages of affyPLM, packages of affy, packages of RColorBrewer. Removing unqualified samples, after that, we got a reasonable and useful sample. According to ''limma'' package of Bioconductor, we picked out key DEGs. $P < 0.05$ was considered to indicate a statistically significant difference.

GO term and KEGG pathway enrichment analysis

The statistically significant difference in gene expression of the gene ontology (GO) classification was commonly used to interpret transcriptomics data. Classification of genes into different functional groups through gene ontology (GO) was a widely method for comprehension functional genomics and apprehend molecular pathways. GO analysis was extensively used in a short list of genes with statistically significant differences in expression. GO divided into three areas, respectively biological process, cellular component and molecular function. Determine the function of the protein domain was obvious. The KEGG website was an online database of genomes, enzymatic pathways and biochemicals. It provided free gene database, access database. Different input data methods were based on different analyzes. Compared with other databases, KEGG was a significant feature of a powerful graphics function, used graphics rather than harassment of

the text to introduce a large number of metabolic pathways and the relationship between the various ways ($P < 0.05$).

PPI network analysis

PPI network significant role was from the protein interaction database, such as HPRD/DIP, etc., extract all or interested in the interaction of protein pairs, and then used the network visualization tools, such as Cytoscape, drawing network diagram. PPI network was an experimental network, which belonged to biological network visualization. Gene co-expression network was trying to reconstruct the biological network through the expression of gene expression, and hoped to discover new rules. In the PPI network, nodes and edges (lines) represented proteins and their interactions ($P < 0.05$).

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was based on large sample transcriptome data, according to the similarity of expression patterns, the gene was divided into different gene modules, which describe as a frequently systematic biology method. The clustering criteria of WGCNA were biologically significant, and the consequence of this method has a higher degree of credibility. We conducted these results to do a lot of further work, such as association traits, metabolic pathway modeling, establishment of gene interaction networks. In a co-expression network, the expression of each gene at a particular time or space was treated as a node. WGCNA uses a soft threshold based on the determination method. Commonly used weighting functions included sigmoid functions and power functions. The similarity and proximity of gene co-expression were calculated using soft threshold power. Analysis of network topology confirmed the final structure of soft thresholding power.

Ethical statement

The study was approved by the Ethics Committee of Xiangya Hospital, Central South University and conducted in accordance with the Declaration of Helsinki. Prior to the start of the study, all participants gave written informed consent.

Patient samples

Five NPC specimens and 5 non-tumor nasopharyngeal epithelial tissues were collected from the Xiangya Hospital, Central South University (Changsha, Hunan, China). Consent forms were obtained from individual patients, and experimental protocols were approved by the Institutional Review Board of Xiangya Hospital, Central South University.

RNA extraction and reverse transcription-PCR analysis

Total RNA was extracted from the biopsy samples with RNeasy® Kit (Qiagen, Carlsbad, CA, USA) according to the manufacture's recommendations. The total RNA samples (1 ug) were used to generate cDNA. Reverse transcription was carried out as the manufacture's recommendations. After the reverse transcription reaction, the PCR reaction was preceded by 94 °C for 5 min, then 30 cycles for genes of 94 °C for 45 s, 55 °C for 45 s, and 72 °C for 1 min followed by 72 °C for 7 min. All RT-PCR reactions were repeated at least three times at different numbers of the extension cycle to avoid false results of the PCR. GAPDH was used as an endogenous control for normalization. The sequences of the primers used for RT-PCR were as follows: SYK forward, 5′-gaagccatatcgagggatga-3′, reverse, 5′-ccacatcgtatgtccagcac-3′; FYN forward, 5′-tggggaagtatggatgggta-3′, reverse, 5′-caccactgcatagagctgga-3′; LRRK2 forward, 5′-tgcctctgttgatcgtcttg-3′, reverse, 5′-cccaatcatttccaacatcc-3′, RIPK4 forward, 5′-cttctgaaaccgaggacctg-3′, reverse, 5′-tcggagaggctgtagtcgtt-3′; ACACB forward, 5′-ctgaccacaggtgaagctga-3′, reverse, 5′-gacaaactcagcgggagaag-3′; The expression of mRNA was assessed by evaluated threshold cycle (CT) values. The CT values were normalized with the expression levels of GAPDH and the relative amount of mRNA specific to each of the target genes was calculated using the $2^{-\Delta\Delta CT}$ method.

Statistical analysis

All results were presented as the mean ± standard error (SE) of three independent experiments. A one-way ANOVA test or unpaired t-test was used for statistical analysis. All statistical tests were performed with GraphPad Prism 5.0 (GraphPad Sofware Inc., La Jolla, CA, USA). All statistical tests were two tailed, and $P < 0.05$ was considered significant.

Results

Identification of differentially expressed genes

GSE12452 expression profile shows the tumor and normal healthy epithelium clearly distinguish. Isolation and identification of differentially expressed genes not only helped to discover the function of genes, but also helped to reveal the pathogenesis of the disease. Gene chip raw data performs detection quality, 4 low-quality samples were removed from 10 normal nasopharyngeal epithelial tissues samples, and

14 low-quality samples were removed from 31 NPC samples. The remaining samples needed to be integrated and processed. Preprocessing with statistical analysis software R was very necessary and important. Then, the microarray data of gene differential expression was analyzed. The data from the statistical results included the 6896 differentially expressed genes. We selected 3306 differentially expressed genes (fold change > 2) for Fig. 1, and then we chose a very differentially and obvious expressed gene (fold change > 2) for Fig. 2 ($P < 0.05$).

GO term enrichment analysis of DEGs

These results showed that our acquired clean reads were of a sufficiently high quality for subsequent analysis. The up-regulated differentially expressed genes were primarily enriched in cell cycle process, biosynthetic process, DNA metabolic process, DNA repairing, gene expression and regulation of signal transduction by p53 class mediator. Down-regulation of the differentially expressed genes was mainly enriched by the response to the immune response, cell activation, regulation of immune system process, inflammatory response, calcium ion transmembrane import into cytosol, B cell receptor signaling pathway. Simultaneously, It has a closely relation with regulation of GTPas, activity signal transduction and regulation of hemopoiesis. The key to the differential gene GO analysis was statistically enriched by gene enrichment. For cellular component, to analyze the biological functions, biological processes and subcellular localization of these differentially expressed genes, the up-regulated DEGs basically enriched in protein complex, membrane-enclosed lumen, mitochondrion, transferase complex, ribosome and intracellular ribonucleoprotein complex. After the scientific analysis, we found that down-regulated genes were mainly found in plasma membrane, immunological synapse, dynein complex, and supramolecular complex. For biological cell molecular function, the DEGs which were up regulated significantly enriched in RNA binding, protein binding, DNA-dependent ATPase activity, ATPase activity, ATP binding and ribonucleotide binding. The DEGs of NPC which were down regulated significantly enriched in molecular function regulator, endopeptidase inhibitor activity, hospholipid binding and G-protein coupled purinergic nucleotide receptor activity. Figures 3 and 4 show more comprehensive and detailed GO enrichment analysis results.

KEGG pathway analysis of NPC DEGs

KEGG is a database of systematic analysis of gene function, genome information, which helps to analyze genes

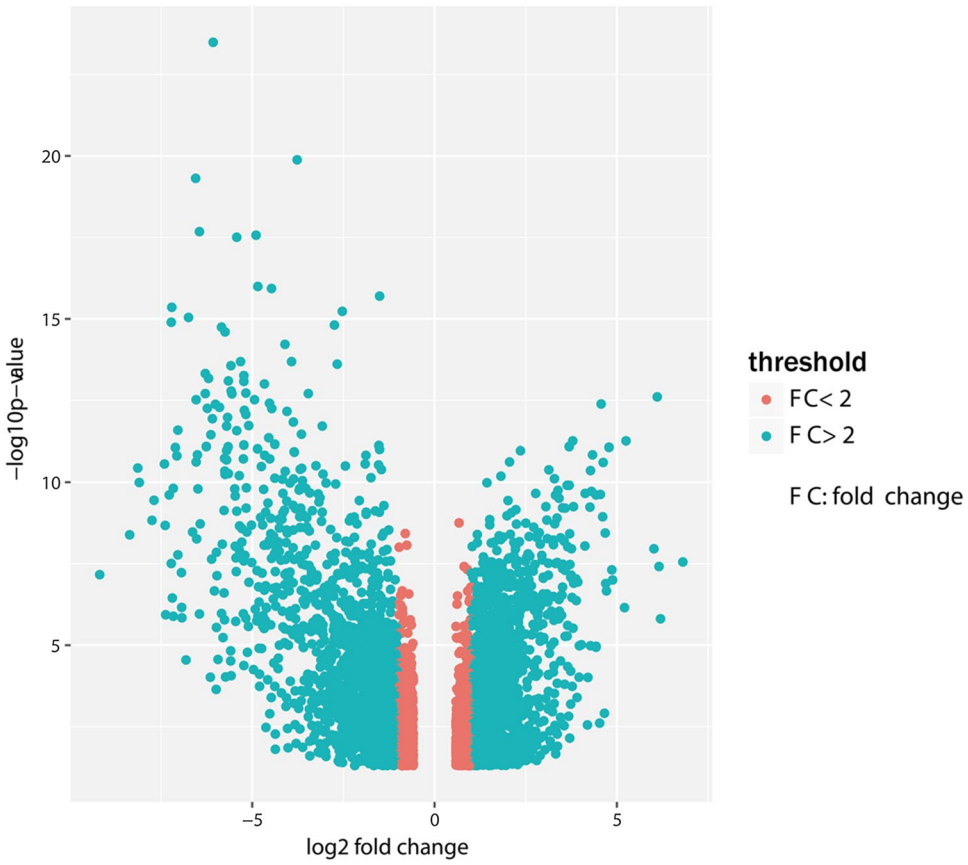

Fig. 1 Volcano plot of 3306 DEGs. Turquoise represents the fold change value of DEGs > 2; red represents the fold change value of DEGs < 2. $P < 0.05$ was considered statistically significant. (Color figure online)

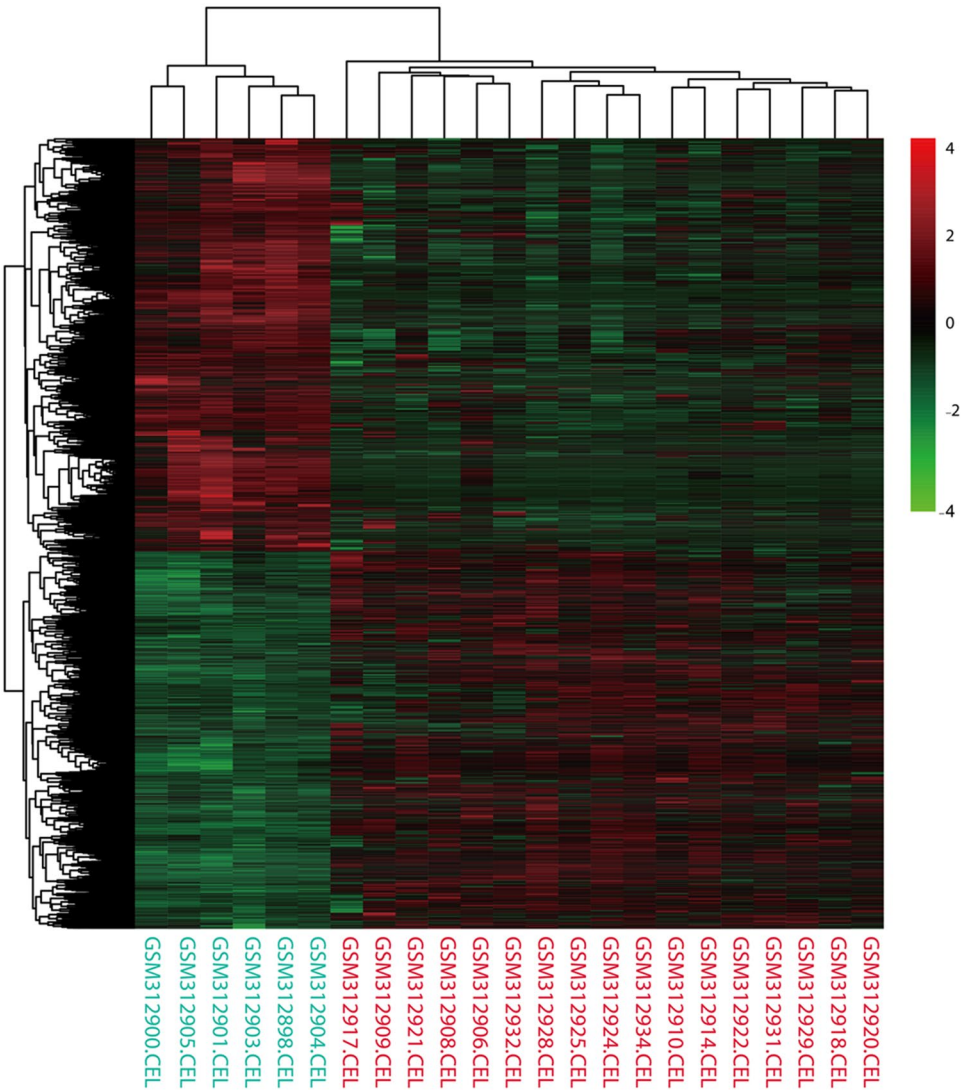

Fig. 2 The heat map for nasopharyngeal carcinoma DEGs. Red represents up-regulation; green represents down-regulated DEGs. $P < 0.05$ was considered statistically significant. (Color figure online)

as a whole network. We found five significantly enriched pathways by KEGG pathway analysis (Fig. 5). First, enrichment aspect of amoebiasis included fifteen DEGS which were down regulated and four down-regulated DEGs. Second, enrichment aspect of DNA replication included thirty up-regulated DEGs and nine down-regulated DEGs. Third, enrichment aspect of one carbon pool by folate included twenty up-regulated DEGs and two down-regulated DEGs. Fourth, enrichment aspect of B cell receptor signaling pathway included eighteen up-regulated DEGs and three down-regulated DEGs. Fifth, enrichment aspect of in cell cycle included five up-regulated DEGs and Thirty-two down-regulated DEGs. The difference in the results represents a complex molecular mechanism of NPC (Fig. 6).

Analysis of DEGs by protein–protein interaction network

Gene co-expression and protein–protein interaction (PPI) network have a guiding role in the study of NPC target gene and target protein, and have a prominent function in the future study of nasopharyngeal. Based on STRING database, we chose the top 10 interacting proteins, the protein–protein interactional (PPI) network of DEGs included SYK, PIK3CG, FYN, ACACB, LRRK2, RIPK4, RAC2, PIK3CD, PTPRC, LCR, RAD51, MAD2L1, CDK1, PCNA, GMPS, CCNB1, GAPDH, CCNA2, RFC4, TOP2A (Figs. 7, 8). The correlation of various protein molecules was marked out, helping us to observe and understand the molecular mechanism of NPC.

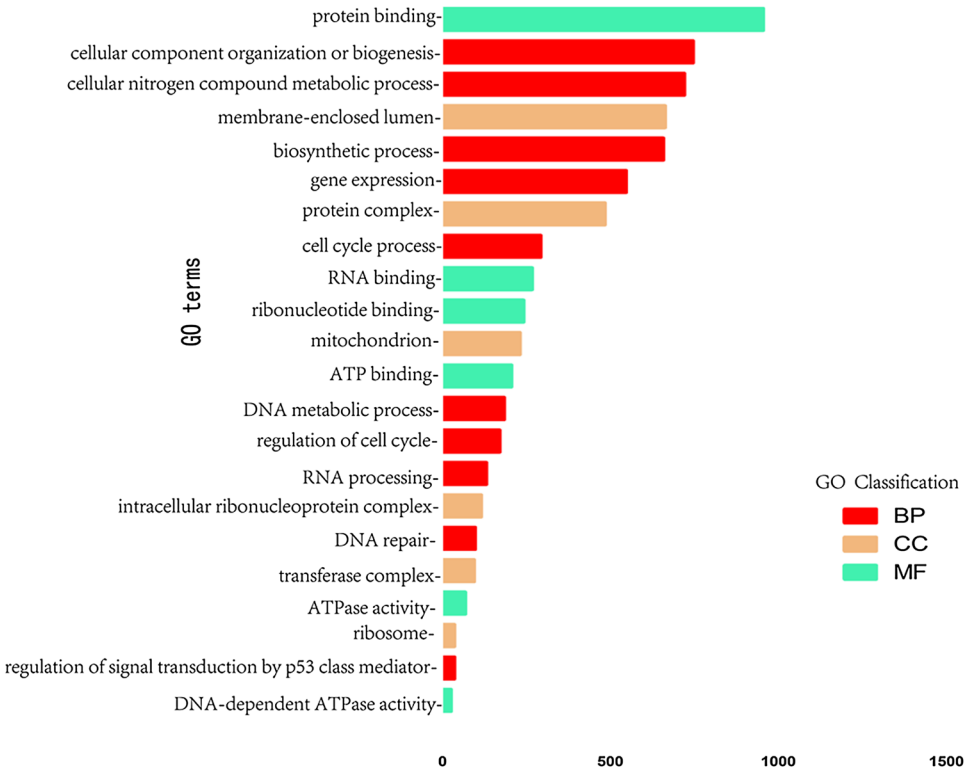

Fig. 3 GO terms: up-regulation; red represents the biological process (BP) of DEGs; coral represents the cellular component (CC) of DEGs; aqua represents the molecular function (MF) of DEGs (FC > 2). $P < 0.05$ was considered statistically significant. (Color figure online)

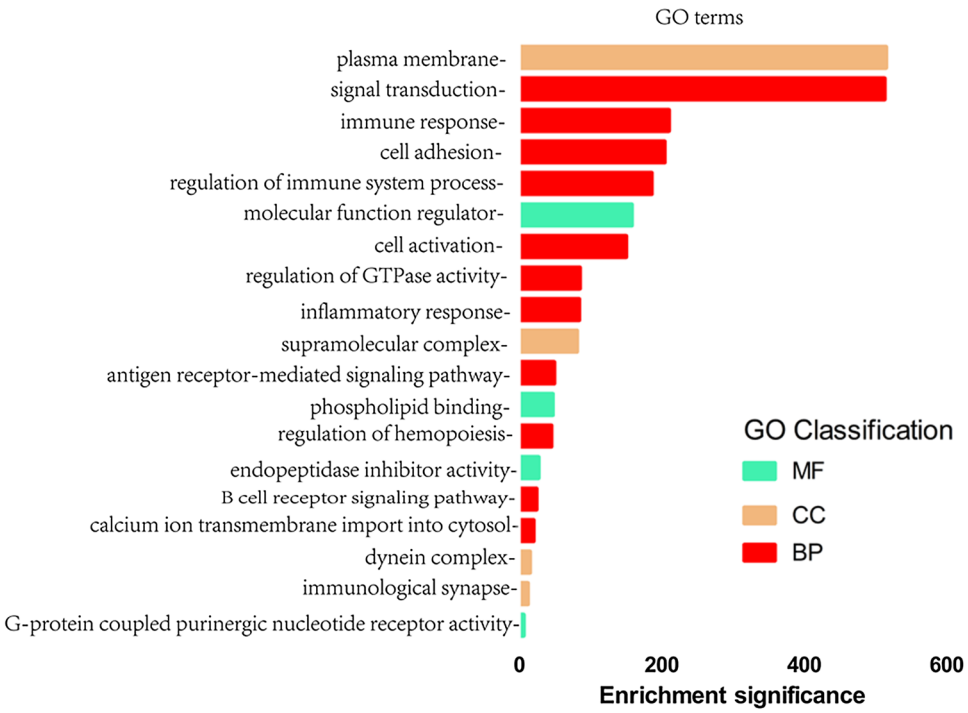

Fig. 4 GO terms: down-regulation; red represents the biological process (BP) of DEGs; coral represents the cellular component (CC) of DEGs; aqua represents the molecular function (MF) of DEGs (FC > 2). $P < 0.05$ was considered statistically significant. (Color figure online)

Weighted correlation network analysis of DEGs

Weighted gene co-expression network analysis (WGCNA), as an efficient and complex biology method, is widely used in biological information analysis. Through correlation analysis, the gene modules related to specific sample traits were quickly screened from the data. More results were shown in Fig. 8. From the figure we saw the regulatory or interrelated genes tended to be show the same or similar expression pattern. Thus, we constructed a gene

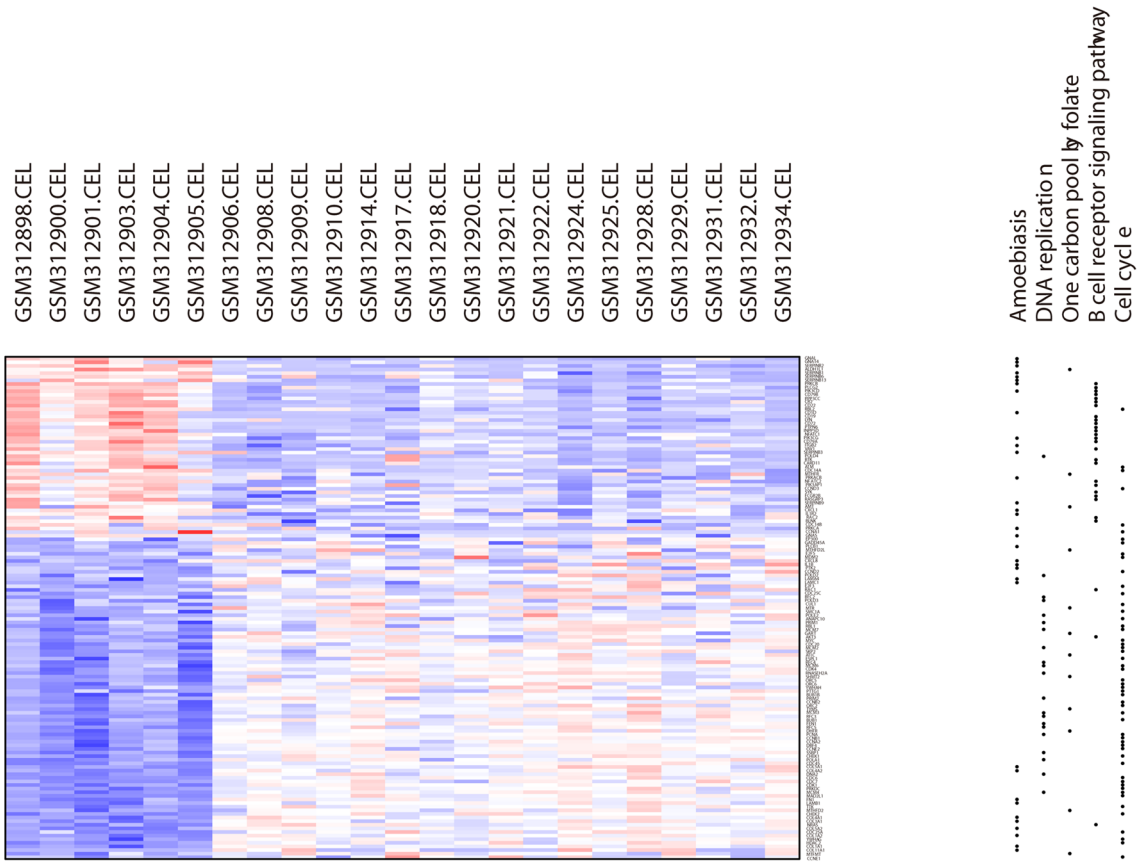

Fig. 5 The heat map for nasopharyngeal carcinoma DEGs in KEGG pathways. Red represents up-regulation; blue represents down-regulation. $P < 0.05$ was considered statistically significant. (Color figure online)

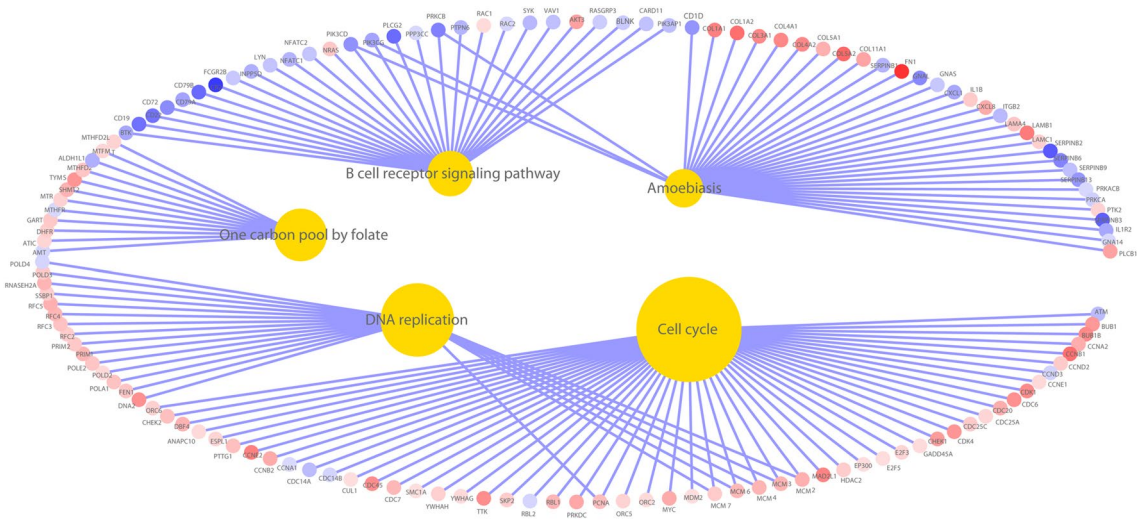

Fig. 6 Yellow represents the main relationship of DEGs. Red represents up-regulation; blue represents down-regulation (FC > 2). $P < 0.05$ was considered statistically significant. (Color figure online)

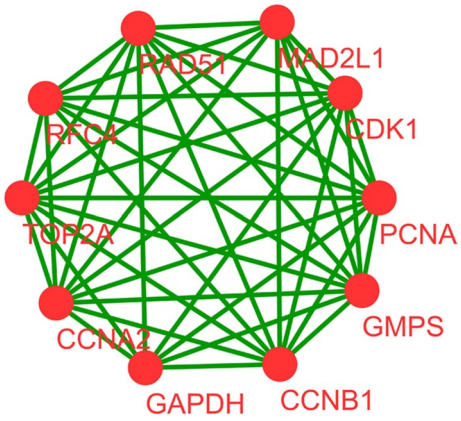

Fig. 7 The up-regulation DEGs with highest connectivity degree of PPI network. Red represents up-regulation. $P < 0.05$ was considered statistically significant. (Color figure online)

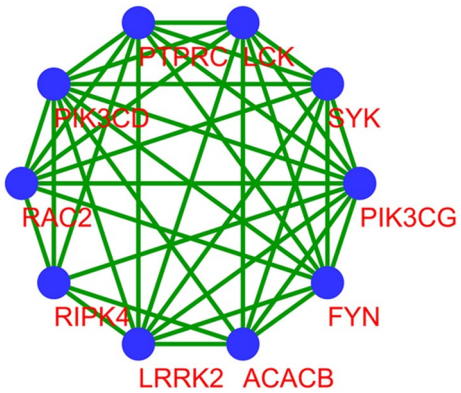

Fig. 8 The down-regulation DEGs with highest connectivity degree of PPI network. Blue represents down-regulation. $P < 0.05$ was considered statistically significant. (Color figure online)

co-expressing the network's screen with a similar expression profile gene module. Identify key modules and biological processes that are closely related to NPC. The DEGs which were in blue module significantly enriched in cell communication, immune system process and signal transduction. The DEGs of brown module significantly enriched in leukocyte activation, antigen receptor-mediated signaling pathway. The DEGs of green module significantly enriched in epidermal cell differentiation, ribonucleoprotein complex binding and skin development. The DEGs of in red module significantly enriched in transmembrane signaling receptor activity and mature B cell differentiation, immune response. The DEGs which were in turquoise module significantly enriched in mitotic cell cycle process and cell cycle process and chromosome segregation. The DEGs which were in yellow module significantly enriched in cell morphogenesis involved in differentiation and protein targeting to lysosome and regulation of cellular process (Fig. 9).

The validation of differential genes by q-PCR

We sought to determine whether the differentially expressed genes identified in our microarray analysis were upregulated or downregulated in a sample of clinical NPC patients. We take normal tissue and cancer tissue from the patient, and validated with qPCR (Fig. 10). Our results showed that the gene FYN, ACACB, RIPK4, LRRK2 expression down-regulated significantly. It is the same as the result of our microarray analysis. The results demonstrated that microarray analysis may provide reference for the identification of molecular markers and therapeutic targets for NPC.

Discussion

Epidemiological trends during the past decade suggest that although incidence of NPC is gradually declining, even in endemic regions, mortality from the disease has fallen substantially [20]. Simultaneously, the sooner found the better curative effect of NPC. NPC is usually (> 70%) diagnosed at a locoregionally advanced stage [21–23]. A comprehensive understanding of the molecular mechanism of NPC was essential to the diagnosis and treatment of disease. Since the microarray achieves rapid detection of gene information by detecting the corresponding position hybridization probes. In this study, we used bioinformatics analysis to obtain differentially expressed genes from GSE12452 in NPC samples and normal samples. The function of these differential genes was mainly concentrated in the cell cycle, biosynthetic process, DNA metabolic process, DNA repairing, and immune response. A comprehensive analysis of the PPI, we found that these genes of DEGs had a special effect on the NPC molecular mechanism. At last, 3306 DEGs were identified in NPC, including 1579 up-regulated genes and 1727 down-regulated genes.

From the top 20 hub genes, we found that the protein encoded by the RAC2 gene regulated secretion, phagocytosis and cellular polarization. The activity of this protein is also involved in the production of reactive oxygen species. The mutation of this gene is associated with neutrophil immunodeficiency syndrome [24]. The proteins encoded by SYK, FYN, RIPK4 genes are associated with proliferation, differentiation and angiogenesis [25, 26]. The protein encoded by PTPRC gene is a member of the protein tyrosine phosphatase (PTP) family. It is a signaling molecule that regulates a variety of cellular processes including cell growth, differentiation, mitosis, and oncogenic transformation [27]. The proteins encoded by PIK3CG and PIK3CD genes are importantly related to immunity [28]. The protein encoded by the ACACB and LRRK2 genes are involved in energy metabolism [29]. Simultaneously, LRRK2 is a complex signaling protein which is PD-related biomarker

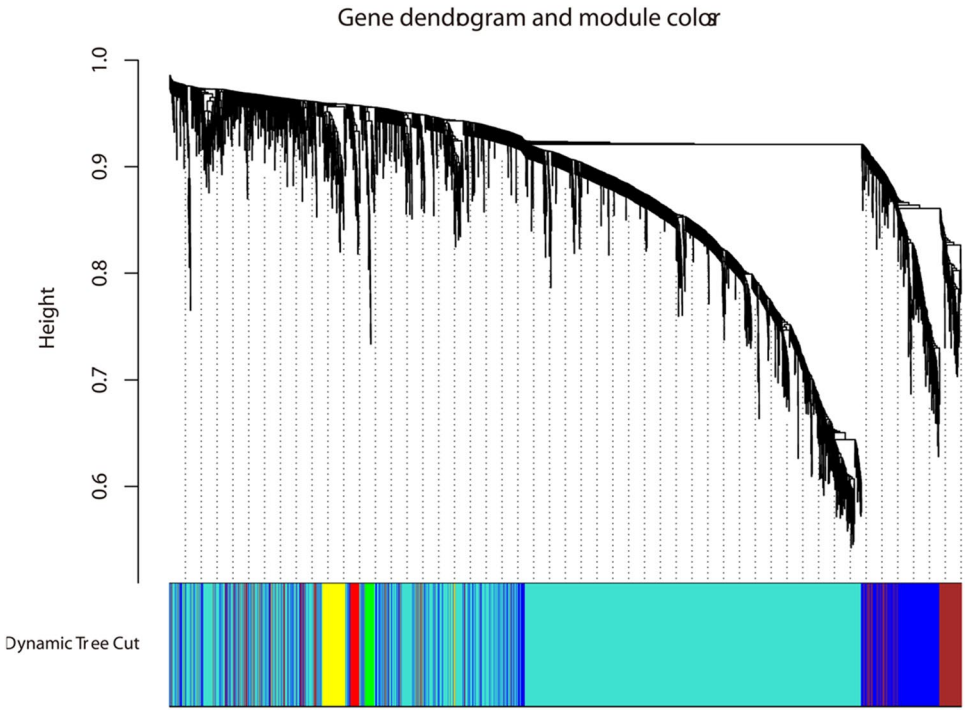

Fig. 9 DEGs in NPC clustering and module screening based on gene expression pattern. Different modules have different colors. $P < 0.05$ was considered statistically significant. (Color figure online)

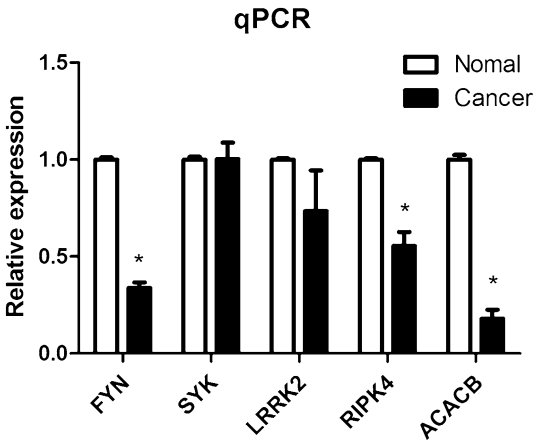

Fig. 10 The gene FYN, ACACB, RIPK4, LRRK2, SYK validated with qPCR. $P < 0.05$ was considered statistically significant

[30]. The proteins encoded by MAD2L1, CDK1, CCNB1, CCNA2 genes are essential for their G1/S and G2/M transitions for eukaryotic cell cycle [31, 32]. The proteins encoded by PCNA and GAPDH genes are involved in DNA damage repair [33, 34]. In addition, the protein encoded by the GAPDH gene was identified as having uracil DNA glycosylase activity in the nucleus. The proteins encoded by GMPS and RFC4 genes have a significant connection with metabolic enzymes. The protein encoded by the RAD51 gene is a member of the RAD51 protein family, which involves in the homologous recombination and repair of DNA. Inactivation of RAD51 protein may be a key event that leads to genomic instability and tumorigenesis [35]. In other words, these differentially expressed genes play an integral role in the progression of NPC.

By observing the results of our bioinformatics analysis, we found that cell cycle process-related genes and pathways. There were many special genes in NPC that regulated the cell cycle [36–39]. In the analysis of the results of KEGG, ATM and ATM protein expression had a significant effect on the proliferation, apoptosis and radiosensitivity of NPC cells [40, 41]. ATM was down-regulated in the cell cycle. In contrast, cyclin dependent kinase 4 was up-regulated in the cell cycle. Simultaneously, it regulates the proliferation and differentiation of neural stem cells [42]. CDK4 regulates cell cycle initiation and progression, inhibiting the growth of most tumors [43–46]. The protein encoded by CDK1 is a member of the Ser/Thr protein kinase family. This protein is a catalytic subunit of the highly conserved protein kinase complex known as M-phase promoting factor (MPF), which is essential for G1/S and G2/M phase transitions of eukaryotic cell cycle [47]. Mitotic cyclins stably associate with this protein and function as regulatory subunits. This allows cancer cells to enter the S phase in time from G1 phase, and maintain a strong ability to divide [48]. Studies had found that CDK1 promoted cell proliferation and survival via phosphorylation and FoxO1 transcription factor inhibition [49]. Furthermore, microRNAs are confirmed to be involved in the regulation of CDK1 gene. Therefore, we initially concluded that many genes were involved in the cell cycle, including chromosome segregation and organelle fission,

mitotic nuclear division, and DNA replication to influence tumor occurrence and progression. Both ATM and Smad pathways contribute to the cell cycle arrest and cell apoptosis during NPC cells treated with radiation [50]. NPC risk factors included EBV infection, which activates DNA-damaging intermediate, involved in DNA repair [51–53].

From our bioinformatics analysis results, we found that metabolic process-related genes and pathways are involved. Some key genes in NPC regulated metabolic process, such as LDHA and argininosuccinate synthetase 1 (ASS1) [54, 55]. LDHA is essential for the proliferation and migration of vascular smooth muscle cells [56]. It provides metabolic advantages to cancer cells growth. Meanwhile, LDHA plays an important role in the regulation of glycolysis in cancer cells. Metabolic reprogramming in tumors has a necessarily close connection with LDHA [57]. We had identified the ASS1 gene from DEGs by our bioinformatics analysis. The down-regulation of ASS1 which have tumor suppressor function, is expressed in many types of human malignancies. P53 [58, 59] from DEGS promotes cell proliferation, migration and invasion [60].

In additional, we found that immune response-related genes and pathways are involved too. GO term enrichment analysis showed that 24 down-regulated DEGs regulate immune response. From the results of KEGG, we clearly saw that 28 DEGS regulated B cell receptor signaling pathway, including 3 up-regulated genes and 25 down-regulated gene, such as PTPN6 and PIK3CD [61]. PTPN6 associates with autoimmune diseases and hematological malignancies in humans [62]. Neutrophils drive the cutaneous inflammation in PTPN6-deficient mice, suggests that PTPN6 plays an important role in the immune response [63]. PI3KCD which is up-regulated drives normal cells to transform into tumor cells, suggests that promote the tumor process obviously [64, 65]. Furthermore, we hypothesized that immune response originating in the local microenvironment could influence NPC initiation and progression from various aspects [66].

In conclusion, we conduct the differential expression genes of GSE12452 by systematic bioinformatics analysis followed by q-PCR validation for some key genes. The advantage of this study is to find some biomarkers, indicates that DEGs have important effects for the development and progress of NPC. However, more biology experiments are needed to confirm the function of these identified genes in NPC.

Acknowledgements This work was supported by Science and technology foundation survey project of Ministry of science and technology of China (2018FY100900, 2018FY10090004), the National Natural Sciences Foundation of China (81672685, 81571151), Hunan Provincial Natural Science Foundation of China (2016JJ2164, 2018JJ3782), Changsha Science and Technology Bureau Foundation (Grant No. kq1701085), the New Xiangya Talent Project of the Third Xiangya Hosipital of Central South University (20160308).

Authors contributions GY and ZY conceived and supervised the study; GY and HZ designed experiments; GY and XY performed experiments; ZY provided new tools and reagents; YY and WD analysed data; GY and QJ wrote the manuscript; ZY and XY made manuscript revisions. All authors reviewed the manuscript.

Compliance with ethical standards

Conflict of interests The authors declare that they have no conflict of interests.

References

1. Chua MLK, Wee JTS, Hui EP, Chan ATC (2016) Nasopharyngeal carcinoma 387:1012–1024
2. Scott RS (2017) Epstein–Barr virus: a master epigenetic manipulator. Curr Opin Virol 26:74–80
3. Zhou Y, Zeng Z, Zhang W, Xiong W, Li X, Zhang B, Yi W, Xiao L, Wu M, Shen S, Li X, Cao L, Tang K, Li G (2008) Identification of candidate molecular markers of nasopharyngeal carcinoma by microarray analysis of subtracted cDNA libraries constructed by suppression subtractive hybridization. Eur J Cancer Prev 17:561–571
4. Nakanishi Y, Wakisaka N, Kondo S, Endo K, Sugimoto H, Hatano M, Ueno T, Ishikawa K, Yoshizaki T. Progression of understanding for the role of Epstein–Barr virus and management of nasopharyngeal carcinoma. Cancer Metastasis Rev. 2017;36:435–47
5. Zeng Z, Zhou Y, Xiong W, Luo X, Zhang W, Li X, Fan S, Cao L, Tang K, Wu M, Li G (2007) Analysis of gene expression identifies candidate molecular markers in nasopharyngeal carcinoma using microdissection and cDNA microarray. J Cancer Res Clin Oncol 133:71–81
6. Zeng Z, Zhou Y, Zhang W, Li X, Xiong W, Liu H, Fan S, Qian J, Wang L, Li Z, Shen S, Li G (2006) Family-based association analysis validates chromosome 3p2as a putative nasopharyngeal carcinoma susceptibility locus. Genet Med 8:156–160
7. Xiong W, Zeng ZY, Xia JH, Xia K, Shen SR, Li XL, Hu DX, Tan C, Xiang JJ, Zhou J, Deng H, Fan SQ, Li WF, Wang R, Zhou M, Zhu SG et al (2004) A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. Cancer Res 64:1972–1974
8. Fang Y, Hu W, Lee MK, King MC, Wang H, Wu S, Zhu J, Xia J, Mo H, Deng M, Huang H, Liu W, Zhao Y, Feng B, Huang L, Zeng Y et al. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. Nat Genet 2002; 31:395–399
9. Bei JX, Li Y, Jia WH, Feng BJ, Zhou G, Chen LZ, Feng QS, Low HQ, Zhang H, He F, Tai ES, Kang T, Liu ET, Liu J, Zeng YX (2010) A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat Genet 42:599–603
10. Kwong DL, Sham JS, Au GK, Chua DT, Kwong PW, Cheng AC, Wu PM, Law MW, Kwok CC, Yau CC, Wan KY, Chan RT, Choy DD (2004) Concurrent and adjuvant chemotherapy for nasopharyngeal carcinoma: a factorial study. J Clin Oncol 22:2643–2653
11. Gou XX, Jin F, Wu WL, Long JH, Li YY, Gong XY, Chen GY, Chen XX, Liu LN (2018) Induction chronomodulated chemotherapy plus radiotherapy for nasopharyngeal carcinoma: a phase II prospective randomized study. J Cancer Res Ther 14(7):1613–1619
12. MacMahon EM, Glass JD, Hayward SD, Mann RB, Becker PS, Charache P, McArthur JC, Ambinder RF (1991) Epstein–Barr

virus in AIDS-related primary central nervous system lymphoma. Lancet 338:969–973
13. Zhang L, Zhang J, Lambert Q, Der CJ, Del VL, Miklossy J, Khalili K, Zhou Y, Pagano JS (2004) Interferon regulatory factor 7 is associated with Epstein–Barr virus-transformed central nervous system lymphoma and has oncogenic properties. J Virol 78:12987–12995
14. Pathmanathan R, Prasad U, Sadler R, Flynn K, Raab-Traub N (1995) Clonal proliferations of cells infected with Epstein–Barr virus in preinvasive lesions related to nasopharyngeal carcinoma. N Engl J Med 333:693–698
15. Coghill AE, Wang CP, Verkuijilen S, Yu KJ, Hsu WL, Middeldorp JM, Hildesheim A. Evaluation of Nasal and Nasopharyngeal Swab Collection for the Detection of Epstein–Barr Virus in Nasopharyngeal Carcinoma. J Med Virol. 2017;90:191–195
16. Zhang J, Jia L, Liu T, Yip YL, Tang WC, Lin W, Deng W, Lo KW, You C, Lung ML, Lung HL, Cheung AL, Tsao SW, Tsang CM. mTORC2-mediated PDHE1alpha nuclear translocation links EBV-LMP1 reprogrammed glucose metabolism to cancer metastasis in nasopharyngeal carcinoma. Oncogene. 2019. https://doi.org/10.1038/s41388-019-0749-y
17. Bose S, Yap LF, Fung M, Starzcynski J, Saleh A, Morgan S, Dawson C, Chukwuma MB, Maina E, Buettner M, Wei W, Arrand J, Lim PV, Young LS, Teo SH, Stankovic T et al (2009) The ATM tumour suppressor gene is down-regulated in EBV-associated nasopharyngeal carcinoma. J Pathol 217:345–352
18. Jiang W, Liu N, Chen XZ, Sun Y, Li B, Ren XY, Qin WF, Jiang N, Xu YF, Li YQ, Ren J, Cho WC, Yun JP, Zeng J, Liu LZ, Li L et al (2015) Genome-Wide Identification of a Methylation Gene Panel as a Prognostic Biomarker in Nasopharyngeal Carcinoma. Mol Cancer Ther 14:2864–2873
19. Dodd LE, Sengupta S, Chen IH, den Boon JA, Cheng YJ, Westra W, Newton MA, Mittl BF, McShane L, Chen CJ, Ahlquist P, Hildesheim A (2006) Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma. Cancer Epidemiol Biomark Prev 15:2216–2225
20. Nieborowska-Skorska M, Kopinski PK, Ray R, Hoser G, Ngaba D (2016) KAI1 overexpression promotes apoptosis and inhibits proliferation, cell cycle, migration, and invasion in nasopharyngeal carcinoma cells. Am J Otolaryngol 14:34–51
21. Mohr S, Doebele C, Comoglio F, Berg T, Beck J, Bohnenberger H, Alexe G, Corso J, Strobel P, Wachter A, Beissbarth T, Schnutgen F, Cremer A, Haetscher N, Gollner S, Rouhi A et al (2017) Hoxa9 and Meis1 Cooperatively Induce Addiction to Syk Signaling by Suppressing miR-146a in Acute Myeloid Leukemia. Cancer Cell 31:549–562
22. Chua M, Wee J, Hui EP, Chan A (2016) Nasopharyngeal carcinoma. Lancet 387(10022):1012–1024
23. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J (2016) Cancer statistics in China, 2015. CA Cancer J Clin 66(2):115–132
24. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A (2015) Global cancer statistics, 2012. CA Cancer J Clin 65(2):87–108
25. Zhang LF, Li YH, Xie SH, Ling W, Chen SH, Liu Q, Huang QH, Cao SM (2015) Incidence trend of nasopharyngeal carcinoma from 1987 to 2011 in Sihui County, Guangdong Province, South China: an age-period-cohort analysis. Chin J Cancer 34(8):350–357
26. Xie YG, Yu Y, Hou LK, Wang X, Zhang B, Cao XC (2016) FYN promotes breast cancer progression through epithelial-mesenchymal transition. Oncol Rep 36:1000–1006
27. Lee YH, Bae S Associations between PTPRC rs10919563 A/G and FCGR2A R131H polymorphisms and responsiveness to TNF blockers in rheumatoid arthritis: a meta-analysis. Rheumatol Int. 2016; 36:837–844
28. Li KC, Yu SH, Zhuge BZ (2017) PIK3CG single nucleotide polymorphisms are associated with poor responsiveness to clopidogrel and increased risk of ischemia in patients with coronary heart disease. Medicine (Baltimore) 96:e7566
29. An L, Jiang H, Tang RN (2015) The ACACB gene rs2268388 polymorphism is associated with nephropathy in Caucasian patients with diabetes: a meta-analysis. Ren Fail 37:925–928
30. Taymans JM, Mutez E, Drouyer M, Sibran W, Chartier-Harlin MC (2017) LRRK2 detection in human biofluids: potential use as a Parkinson's disease biomarker? Biochem Soc Trans 45:207–212
31. Faesen AC, Thanasoula M, Maffini S, Breit C, Muller F, van Gerwen S, Bange T, Musacchio A (2017) Basis of catalytic assembly of the mitotic checkpoint complex. Nature 542:498–502
32. Zuryn A, Krajewski A, Szulc D, Litwiniec A, Grzanka A (2016) Activity of cyclin B1 in HL-60 cells treated with etoposide. Acta Histochem 118:537–543
33. Beaver JM, Lai Y, Rolle SJ, Liu Y (2016) Proliferating cell nuclear antigen prevents trinucleotide repeat expansions by promoting repeat deletion and hairpin removal. DNA Repair 48:17–29
34. Kosova AA, Khodyreva SN, Lavrik OI (2017) Role of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in DNA repair. Biochemistry 82:643–654
35. Xu J, Zhao L, Xu Y, Zhao W, Sung P, Wang HW (2017) Cryo-EM structures of human RAD51 recombinase filaments during catalysis of DNA-strand exchange. Nat Struct Mol Biol 24:40–46
36. Guo Z, Wang Y, Yang J, Zhong J, Liu X, Xu M (2016) KAI1 overexpression promotes apoptosis and inhibits proliferation, cell cycle, migration,and invasion in nasopharyngeal carcinoma cells. Am J Otolaryngol 38:511–517
37. Chen X, Yue B, Zhang C, Qi M, Qiu J, Wang Y, Chen J (2017) MiR-130a-3p inhibits the viability, proliferation, invasion, and cell cycle, and promotes apoptosis of nasopharyngeal carcinoma cells by suppressing BACH2 expression. Biosci Rep 37:BSR20160576
38. Li X, Liu F, Lin B, Luo H, Liu M, Wu J, Li C, Li R, Zhang X, Zhou K, Ren D (2017) miR150 inhibits proliferation and tumorigenicity via retarding G1/S phase transition in nasopharyngeal carcinoma. Int J Oncol 50:1097–1108
39. Jiang X, Feng L, Dai B, Li L, Lu W. Identification of key genes involved in nasopharyngeal carcinoma. Braz J Otorhinolar. 2016;83:670–6
40. Wang M, Liu G, Shan GP, Wang BB (2017) In vivo and in vitro effects of ATM/ATR signaling pathway on proliferation, apoptosis, and radiosensitivity of nasopharyngeal carcinoma cells. Cancer Biother Radiopharm 32:193–203
41. Royba E, Miyamoto T, Natsuko AS, Hosoba K, Tauchi H, Kudo Y, Tashiro S, Yamamoto T, Matsuura S (2017) Evaluation of ATM heterozygous mutations underlying individual differences in radiosensitivity using genome editing in human cultured cells. Sci Rep 7:5996
42. Goel S, DeCristo MJ, Watt AC, BrinJones H, Sceneay J, Li BB, Khan N, Ubellacker JM, Xie S, Metzger-Filho O, Hoog J, Ellis MJ, Ma CX, Ramm S, Krop IE, Winer EP et al (2017) CDK4/6 inhibition triggers anti-tumour immunity. Nature 548:471–475
43. Yu Q, Sicinska E, Geng Y, Ahnstrom M, Zagozdzon A, Kong Y, Gardner H, Kiyokawa H, Harris LN, Stal O, Sicinski P (2006) Requirement for CDK4 kinase function in breast cancer. Cancer Cell 9:23–32
44. Malumbres M (2012) Cell cycle-based therapies move forward. Cancer Cell 22:419–420
45. Patnaik A, Rosen LS, Tolaney SM, Tolcher AW, Goldman JW, Gandhi L, Papadopoulos KP, Beeram M, Rasco DW, Hilton JF, Nasir A, Beckmann RP, Schade AE, Fulford AD, Nguyen TS, Martinez R et al (2016) Efficacy and safety of abemaciclib, an inhibitor of CDK4 and CDK6, for patients with breast cancer,

non-small cell lung cancer, and other solid tumors. Cancer Discov 6:740–753
46. Corona SP, Ravelli A, Cretella D, Cappelletti MR, Zanotti L, Dester M, Gobbi A, Petronini PG, Generali D (2017) CDK4/6 inhibitors in HER2-positive breast cancer. Crit Rev Oncol Hematol 112:208–214
47. Bednarek K, Kiwerska K, Szaumkessel M, Bodnar M, Kostrzewska-Poczekaj M, Marszalek A, Janiszewska J, Bartochowska A, Jackowska J, Wierzbicka M, Grenman R, Szyfter K, Giefing M, Jarmuz-Szymczak M (2016) Recurrent CDK1 overexpression in laryngeal squamous cell carcinoma. Tumor Biol 37:11115–11126
48. Landry BD, Mapa CE, Arsenault HE, Poti KE, Benanti JA (2014) Regulation of a transcription factor network by Cdk1 coordinates late cell cycle gene expression. EMBO J 33:1044–1060
49. Liu P, Kao TP, Huang H (2008) CDK1 promotes cell proliferation and survival via phosphorylation and inhibition of FOXO1 transcription factor. Oncogene 27:4733–4744
50. Li MY, Liu JQ, Chen DP, Li ZY, Qi B, He L, Yu Y, Yin WJ, Wang MY, Lin L (2017) Radiotherapy induces cell cycle arrest and cell apoptosis in nasopharyngeal carcinoma via the ATM and Smad pathways. CANCER BIOL THER 18:681–693
51. Ward MH, Pan WH, Cheng YJ, Li FH, Brinton LA, Chen CJ, Hsu MM, Chen IH, Levine PH, Yang CS, Hildesheim A (2000) Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. Int J Cancer 86:603–609
52. Hildesheim A, Anderson LM, Chen CJ, Cheng YJ, Brinton LA, Daly AK, Reed CD, Chen IH, Caporaso NE, Hsu MM, Chen JY, Idle JR, Hoover RN, Yang CS, Chhabra SK (1997) CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst 89:1207–1212
53. Kongruttanachok N, Sukdikul S, Setavarin S, Kerekhjanarong V, Supiyaphun P, Voravud N, Poovorawan Y, Mutirangura A (2001) Cytochrome P450 2E1 polymorphism and nasopharyngeal carcinoma development in Thailand: a correlative study. BMC Cancer 1:4
54. Henriet E, Hammoud AA, Dupuy JW, Dartigues B, Ezzoukry Z, Dugot-Senant N, Leste-Lasserre T, Pallares-Lupon N, Nikolski M, Le Bail B, Blanc JF, Balabaud C, Bioulac-Sage P, Raymond AA, Saltel F (2017) Argininosuccinate synthase 1 (ASS1): A marker of unclassified hepatocellular adenoma and high bleeding risk. Hepatology 66:2016–2028
55. Kremer JC, Van Tine BA (2017) Therapeutic arginine starvation in ASS1-deficient cancers inhibits the Warburg effect. Mol Cell Oncol 4:e1295131
56. Kim JH, Bae KH, Byun JK, Lee S, Kim JG, Lee IK, Jung GS, Lee YM, Park KG (2017) Lactate dehydrogenase-A is indispensable for vascular smooth muscle cell proliferation and migration. Biochem Biophys Res Commun 492:41–47
57. Boudreau A, Purkey HE, Hitz A, Robarge K, Peterson D, Labadie S, Kwong M, Hong R, Gao M, Del NC, Pusapati R, Ma S, Salphati L, Pang J, Zhou A, Lai T et al (2016) Metabolic plasticity underpins innate and acquired resistance to LDHA inhibition. Nat Chem Biol 12:779–786
58. Ahn JH, Kim TJ, Lee JH, Choi JH (2017) Mutant p53 stimulates cell invasion through an interaction with Rad21 in human ovarian cancer cells. Sci Rep 7:9076
59. Lagares MH, Silva K, Barbosa AM, Rodrigues DA, Costa IR, Martins J, Morais MP, Campedelli FL, Moura K. Analysis of p53 gene polymorphism (codon 72) in symptomatic patients with atherosclerosis. Genet Mol Res. 2017; 16
60. Zhou R, Xu A, Gingold J, Strong LC, Zhao R, Lee D (2013) Li–Fraumeni syndrome disease model: a platform to develop precision cancer therapy targeting oncogenic p53. Trends Pharmacol Sci 26:316–333
61. Bu Q, You F, Pan G, Yuan Q, Cui T, Hao L, Zhang J (2017) MiR-125b inhibits anaplastic thyroid cancer cell migration and invasion by targeting PIK3CD. Biomed Pharmacother 88:443–448
62. Zhang X, Yang L, Liu X, Nie Z, Wang X, Pan Y, Luo J (2017) Research on the epigenetic regulation mechanism of the PTPN6 gene in advanced chronic myeloid leukaemia. Br J Haematol 178:728–738
63. Raymond AA, Javary J, Breig O, Neaud V, Rosenbaum J (2017) Reptin regulates insulin-stimulated Akt phosphorylation in hepatocellular carcinoma via the regulation of SHP-1/PTPN6. Cell Biochem Funct 35:289–295
64. Rae W, Gao Y, Ward D, Mattocks CJ, Eren E, Williams AP (2017) A novel germline gain-of-function variant in PIK3CD. Clin Immunol 181:29–31
65. Yuan T, Yang Y, Chen J, Li W, Li W, Zhang Q, Mi Y, Goswami RS, You JQ, Lin D, Qian MD, Calin S, Liang Y, Miranda RN, Calin GA, Zhou X et al (2017) Regulation of PI3K signaling in T-cell acute lymphoblastic leukemia: a novel PTEN/Ikaros/miR-26b mechanism reveals a critical targetable role for PIK3CD. Leukemia 31:2355
66. Ai P, Li Z, Jiang Y, Song C, Zhang L, Hu H, Wang T (2017) Tumor microenvironment contributes to Epstein–Barr virus antinuclear antigen-1 antibody production in nasopharyngeal carcinoma. Oncol Lett 14:2458–2462