Observational Study

Medicine®

OPEN

Identifying key genes and small molecule compounds for nasopharyngeal carcinoma by various bioinformatic analysis

Lucheng Fang, MD[a], Licai Shi, MD[a], Wen Wang, MD[a], Qinjuan Chen, MD[b], Xingwang Rao, PhD[a,*]

Abstract

Nasopharyngeal carcinoma (NPC) is one of the most prevalent head and neck cancer in southeast Asia. It is necessary to proceed further studies on the mechanism of occurrence and development of NPC.

In this study, we employed the microarray dataset GSE12452 and GSE53819 including 28 normal samples and 49 nasopharyngeal carcinoma samples downloaded from the Gene Expression Omnibus(GEO) to analysis. R software, STRING, CMap, and various databases were used to screen differentially expressed genes (DEGs), construct the protein–protein interaction (PPI) network, and proceed small molecule compounds analysis, among others.

Totally, 424 DEGs were selected from the dataset. DEGs were mainly enriched in extracellular matrix organization, cilium organization, PI3K-Akt signaling pathway, collagen-containing extracellular matrix, and extracellular matrix-receptor interaction, among others. Top 10 upregulated and top 10 downregulated hub genes were identified as hub DEGs. Piperlongumine, apigenin, menadione, 1,4-chrysenequinone, and chrysin were identified as potential drugs to prevent and treat NPC. Besides, the effect of genes CDK1, CDC45, RSPH4A, and ZMYND10 on survival of NPC was validated in GEPIA database.

The data revealed novel aberrantly expressed genes and pathways in NPC by bioinformatics analysis, potentially providing novel insights for the molecular mechanisms governing NPC progression. Although further studies needed, the results demonstrated that the expression levels of CDK1, CDC45, RSPH4A, and ZMYND10 probably affected survival of NPC patients.

Abbreviations: BP = biological process, CC = cellular component, CMap = Connectivity Map, DEG = differentially expressed gene, GEO = Gene Expression Omnibus, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, NPC = nasopharyngeal carcinoma, PPI = Protein-protein interaction.

Keywords: gene expression, nasopharyngeal carcinoma, pathway, small molecule compound

1. Introduction

Nasopharyngeal carcinoma (NPC) is an uncommon cancer in the western countries, but it is one of the most prevalent head and neck cancer in southeast Asia, the Middle East, and North Africa.[1,2] Notably, it is in China that the annual incidence rate of NPC is the highest in the world (about 60.6 cases per 100,000 people), and the number of new cases accounted for nearly the half of the global amount.[2,3] Over the past decades, nasopharyngeal carcinoma incidence has declined gradually worldwide: substantial reductions have been observed in south and east Asia, north America, and the Nordic countries, with average annual changes of about−1% to −5%.[1,2] Undoubtedly, since the high risk population of NPC is mainly among male aged 40 to 50 years, both the social-economic loss and the medical burden are considerable.[3] World Health Organization have classified NPC into 3 pathological subtypes including keratinizing squamous, nonkeratinizing, and basaloid squamous.[1] Epstein-Barr virus infection is a generally accepted important risk factors for nonkeratinizing subtype NPC,[4] and the function of genetic factors also played an important role in tumorigenesis.[1,5] A recent study proposed a nomogram combining pre-treatment plasma Epstein-Barr virus DNA and clinicopathological variables, and found that it resulted in more accurate prognostic prediction for patients with nasopharyngeal carcinoma.[1] What's more, based on many constructive clinical trials, concurrent chemoradiotherapy is generally recommended as standard treatment for NPC.[1,6] In spite of more and more advances in diagnosis and treatment technique, it is approximately 30% of high-risk patients that still die of tumor recurrence with distant metastasis.[7] Therefore, it is necessary to proceed further studies on the mechanism of occurrence and development of NPC to improve survival for NPC patients.

Editor: Milind Chalishazar.

The authors report no conflicts of interest.

Ethical approval statement: Our study didn't involve human beings or animals, so the approval of Ethics Committee isn't necessary in our study.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

[a] First Affiliated Hospital of Wenzhou Medical University, Wenzhou City, Zhejiang Province, China, [b] Second Affiliated Hospital of Wenzhou Medical University, Wenzhou City, Zhejiang Province, China.

[*] Correspondence: Xingwang Rao, First Affiliated Hospital of Wenzhou Medical University, Wenzhou City, Zhejiang Province, China (e-mail: 11124682@cumt.edu.cn).

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Fang L, Shi L, Wang W, Chen Q, Rao X. Identifying key genes and small molecule compounds for nasopharyngeal carcinoma by various bioinformatic analysis. Medicine 2021;100:37(e27257).

Received: 30 December 2020 / Received in final form: 25 July 2021 / Accepted: 30 August 2021

http://dx.doi.org/10.1097/MD.0000000000027257

In this big data age, microarray technology with automated, integrated, and miniaturized features have been widely applied to analyze cancer-specific gene expression profiles.[8] Bioinformatic analysis can be performed more easily based on microarray technology. What's more, the establishment of multiple databases further enables us to perform gene functional analysis more efficiently and comprehensively. In this study, data from the Gene Expression Omnibus (GEO) database were employed to identify differentially expressed genes (DEGs). The Gene Ontology (GO) database, the Kyoto Encyclopedia of Genes, and Genomes (KEGG) database were used for enrichment analysis. The STRING database were applied to construct protein–protein interaction (PPI) network.[9] The GEPIA database[10] is able to perform survival analysis and the Connectivity Map (CMap) database[11] can find interaction between genes and drugs.

In summary, we believe that it is feasible to identify novel differentially expressed genes associated with NPC with the explained method, providing key insights into the NPC at the molecular level that governed the NPC development and progressionto full blown stages. What's more, we are looking forward to exploring more reliable diagnostic way and finding better personalized therapy strategy by the above analysis.

2. Methods and materials

2.1. Data resources and extraction of DEGs

Figure 1 demonstrated the brief flowchart of our study. The data of microarray datasets GSE12452 and GSE53819 downloaded from the GEO database were selected for analysis (https://www.ncbi.nlm.nih.gov/geo/). The update dates of GSE12452 and GSE53819 are March 25, 2019 and August 01, 2019. There were totally 28 normal samples and 49 NPC samples in dataset GSE12452 (Platform: GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) and dataset GSE53819 (Platform: GPL6480 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F). We identified DEGs between normal and NPC samples using the Limma software package in R. What's more, the thresholds to screen DEGs were P value <.05 and |log2 fold change (FC)| >1. Subsequently, the online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to identify overlapped DGEs between GSE12452 and GSE53819 and plotted Venn diagram.

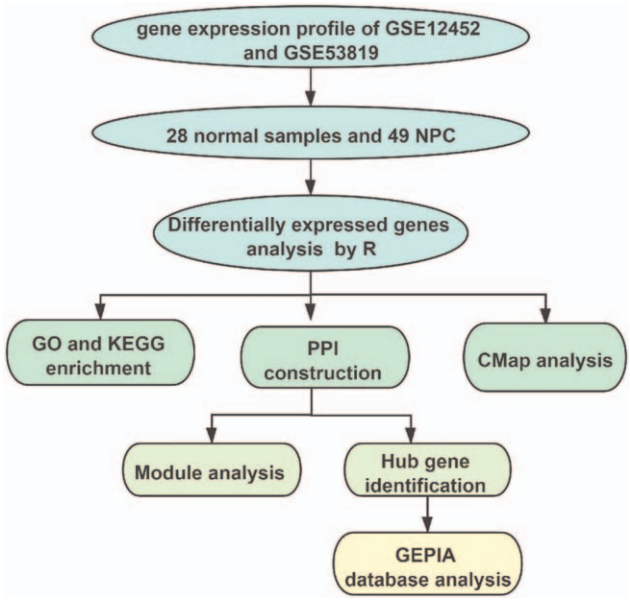

Figure 1. Flowchart of the study procedures.

2.2. GO function and KEGG pathway enrichment analysis

The GO knowledgebase is the world's largest source of information on the functions of genes. In the GO analysis, we classified genes into different biological process (BP), cellular component (CC), and molecular function. On the contrary, the KEGG database was used for identifying functional and metabolic pathways. P value <.05 was set as cutoff indicating a statistical difference. Some convenient and useful R packages, such as clusterProfiler package, topGO package, Rgraphviz package, and pathview package, were applied to postulate the possible functions and pathways of the DEGs.

2.3. PPI networks and Hub gene extraction

STRING database, an online web tool, was used to construct protein interactions network with a confidence score >0.4 set as the standard.[9] Also, visualized PPI network was built and downloaded from STRING database. Afterward, based on Cytoscape 3.8.0 software, Molecular Complex Detection (MCODE) was applied to extract hub modules with MCODE scores>10, and Cytohubba was utilized to screen top 10 hub genes ranked by Degree method.[12–14]

2.4. Validation in the GEPIA database

GEPIA is a newly opening interactive web server, developed by Peking University, for cancer and normal gene expression profiling and interactive analyses based on 9736 tumors and 8587 normal samples from the TCGA and the GTEx projects.[10] The GEPIA database allows users to perform all expression analyses such as survival analysis, correlation analysis, and similar gene detection, and so on. The impact of hub genes on head and neck squamous cell carcinoma about overall survival (OS) and disease-free survival (DFS) was estimated by taking advantage of the GEPIA database.

2.5. Screening of small-molecules compounds

We screened the compounds with molecular features that have the potential to treat NPC. To do so, we uploaded previously selected DEGs into the CMap database. Connectivity Map (CMap) is a tool utilized to study the relationship between small-molecular compounds and certain genes.[11] To upload eligible data for CMap, we converted gene symbol ID of all DEGs into probe sets. We obtained the corresponding small-molecule compounds by using probe sets as queries to search the CMap. Potential therapeutic compounds were identified based on enrichment scores and P value.

3. Results

3.1. Identification of DGEs

After standardizing the data of microarray results, we identified 2156 DEGs in GSE53919 and 1067 DEGs in GSE12452. The

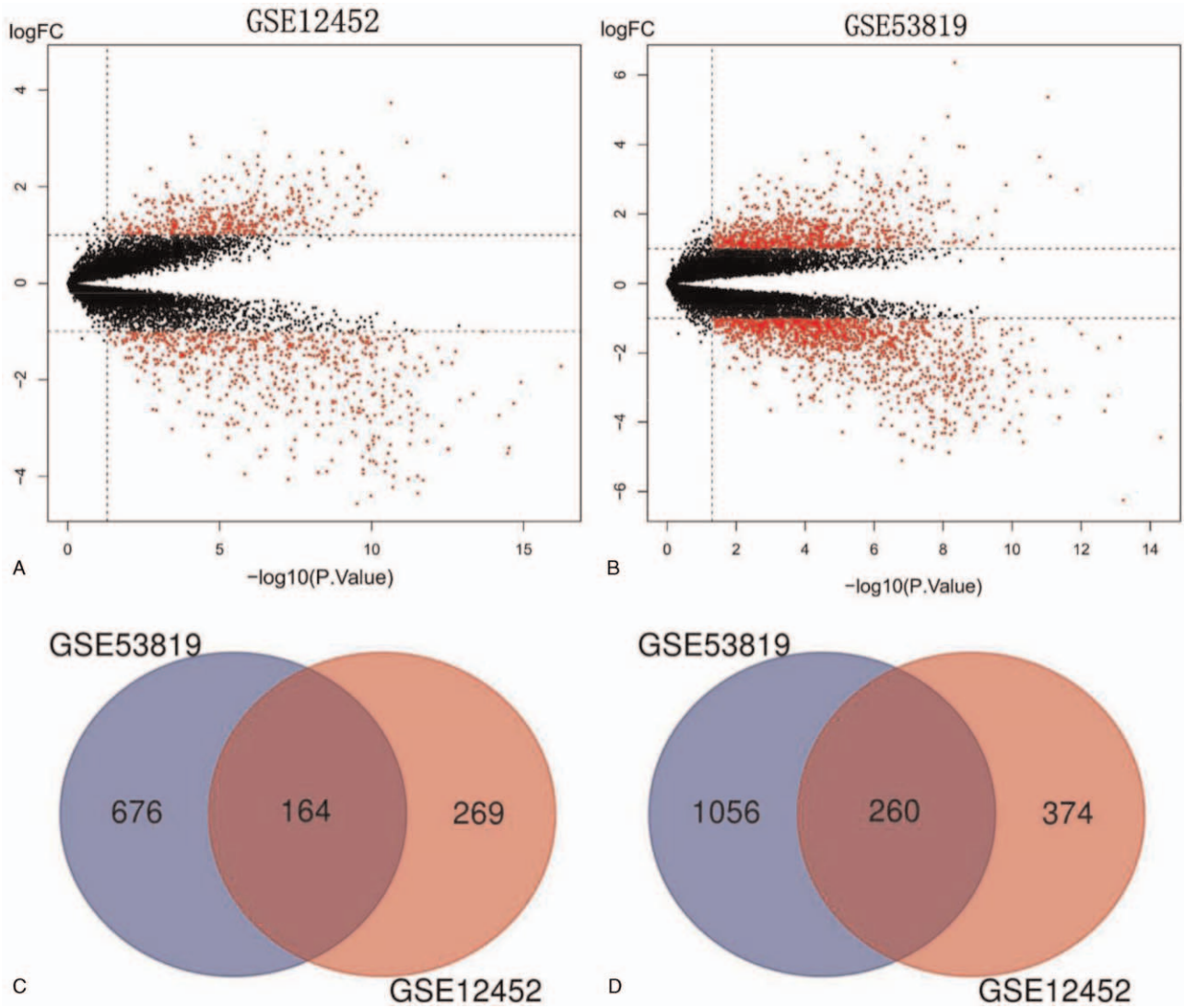

Figure 2. (A, B) Volcano plots of DEGs in GSE12452 and GSE53819. Black dots represent genes equally represented between NPC and normal samples. Red dots represent upregulated and downregulated DEGs. Volcano plots showing all DEGs. |log2FC| ≥ 1 and $P < .05$ were set as cut-off criteria. (C) High expression DEGs. (D) Low expression DEGs. DEGs = differentially expressed genes.

amount of overlapped DEGs between 2 data sets was 424. What's more, there were 164 upregulated and 260 downregulated overlapped DEGs. The volcano plots and Venn diagrams are shown in Figure 2.

3.2. GO and KEGG pathway enrichment analyses

The GO annotation and KEGG pathway enrichment analyses of overlapped DEGs were implemented using R packages. As for BP, highly expressed DEGs were mainly enriched in "extracellular matrix organization," "extracellular structure organization," "chromosome segregation," "nuclear division and organelle fission," and so on (Fig. 3A), whereas DEGs with low expression were mainly associated with "cilium organization," "cilium assembly, microtubule-based movement," "microtubule bundle formation," and "cilium movement," and so on (Fig. 3B). In the analysis of the CC enrichment, highly expressed DEGs correlated with "collagen-containing extracellular matrix, endoplasmic reticulum lumen," "spindle," "chromosomal region and chromosome," "centromeric region," and so on (Fig. 3C). For another, lowly expressed DEGs were predominant in "motile cilium," "plasma membrane bounded cell projection cytoplasm," "ciliary plasm," and "9+2 motile cilium," and so on (Fig. 3D). About molecular function, highly expressed DEGs were primarily enriched in "extracellular matrix structural constituent," "cell adhesion molecule binding," "receptor ligand activity," "signaling receptor activator activity," and "cytokine activity," and so on (Fig. 3E), whereas DEGs with low expression were mostly enriched in "'oxidoreductase activity acting on the NAD or NADP as acceptor," "oxidoreductase activity acting on CH-OH group of donors," "motor activity," "microtubule motor activity," and "retinol dehydrogenase activity," and so on (Fig. 3F). However, for KEGG pathway enrichment, highly expressed DEGs mainly involved in "PI3K-Akt signaling pathway," "Cytokine-cytokine receptor interaction," and "ECM-receptor interaction," and so on (Fig. 3G), whereas lowly expressed DEGs mainly involved in "hematopoietic cell lineage," "B cell receptor signaling pathway," and "complement

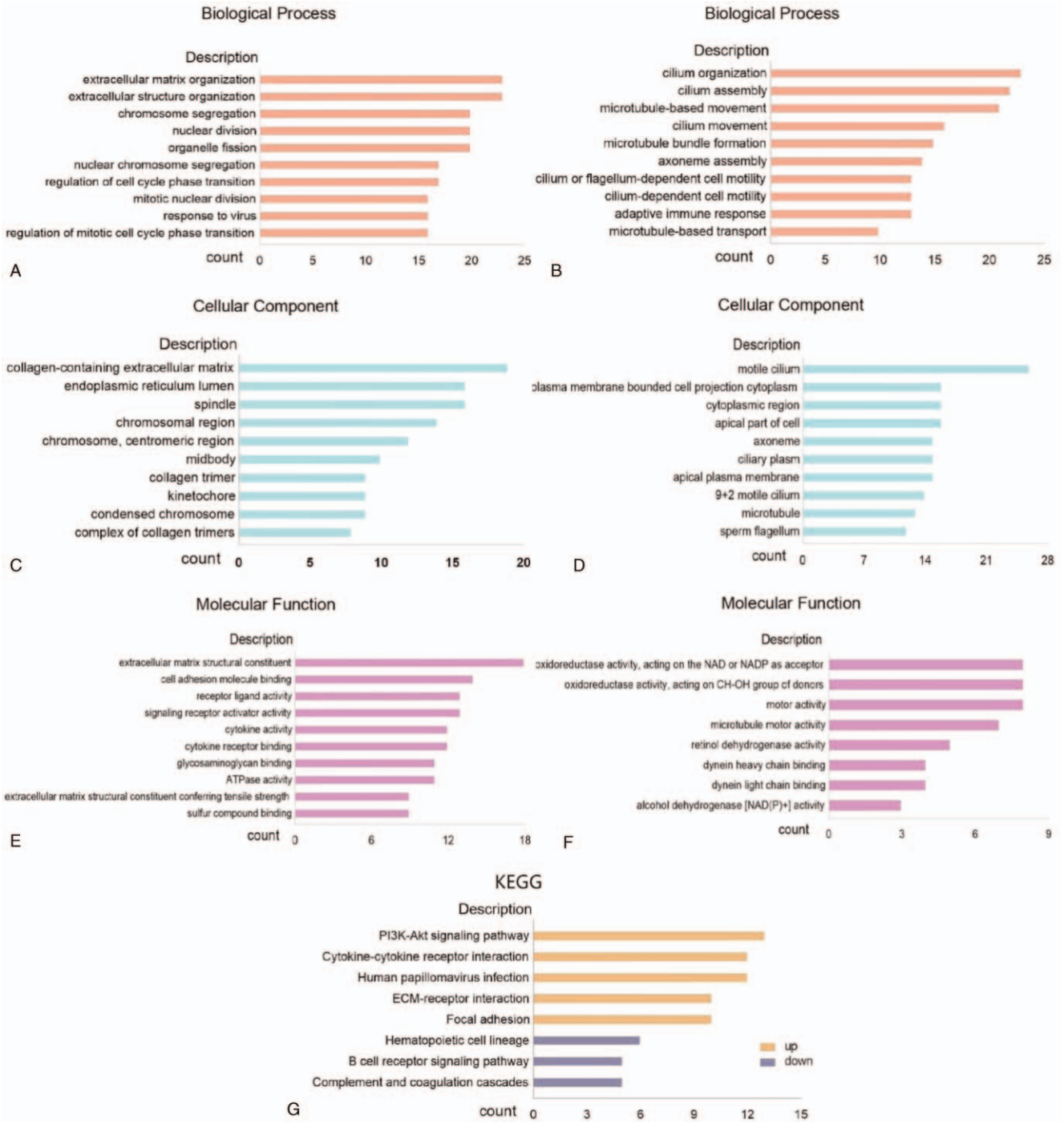

Figure 3. The gene ontology annotation and pathway enrichment analysis. The high count means that the number of genes were found in the particular ontology or pathway. (A, C, E) GO of upregulated DEGs. (B, D, F) GO of downregulated DEGs. (G) KEGG of DEGs, orange represented upregulated DEGs and blue represented downregulated DEGs. DEG = differentially expressed genes, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes.

and coagulation cascades" (Fig. 3G). The detailed information of GO and KEGG are shown in Table 1.

3.3. PPI networks and hub genes

Based on STRING database, the most significant module of upregulated genes was obtained using Cytoscape (Fig. 4A), and it was mainly linked to extracellular matrix organization, extracellular structure organization, chromosome segregation, and so on (Table 2). Top 10 hub genes with highly expression were BUB1B, CDK1, CDC6, KIF23, ZWINT, CDC45, IRC5, UBE2C, TTK, and TOP2A (Fig. 5A). The most significant module of downregulated genes is shown in Figure 4B, and it was mainly linked to cilium organization, cilium assembly, motile cilium, and so on

Table 1

GO and KEGG pathway enrichment analysis of DEGs.

Term	Description	Count in gene set	*P*
Upregulated DEGs			
GO:BP (Top 5)			
GO:0030198	Extracellular matrix organization	23	1.24E-13
GO:0043062	Extracellular structure organization	23	1.31E-13
GO:0007059	Chromosome segregation	20	5.84E-12
GO:0000280	Nuclear division	20	4.07E-10
GO:0048285	Organelle fission	20	2.23E-09
GO:CC (Top 5)			
GO:0062023	Collagen-containing extracellular matrix	19	9.57E-10
GO:0005788	Endoplasmic reticulum lumen	16	5.24E-09
GO:0005819	Spindle	16	2.69E-08
GO:0098687	Chromosomal region	14	1.03E-06
GO:0000775	Chromosome, centromeric region	12	6.41E-08
GO:MF (Top 5)			
GO:0005201	Extracellular matrix structural constituent	18	4.44E-15
GO:0050839	Cell adhesion molecule binding	14	.00013
GO:0048018	receptor ligand activity	13	.00033
GO:0030546	Signaling receptor activator activity	13	.00036
GO:0005125	Cytokine activity	12	5.50E-07
KEGG			
hsa04151	PI3K-Akt signaling pathway	13	.00024
hsa04060	Cytokine-cytokine receptor interaction	12	.00016
hsa05165	Human papillomavirus infection	12	.00047
hsa04512	ECM-receptor interaction	10	6.65E-08
hsa04510	Focal adhesion	10	.00012
Downregulated DEGs			
GO:BP (Top 5)			
GO:0044782	Cilium organization	23	7.79E-11
GO:0060271	Cilium assembly	22	1.93E-10
GO:0007018	Microtubule-based movement	21	1.16E-11
GO:0003341	Cilium movement	16	5.78E-17
GO:0001578	Microtubule bundle formation	15	9.78E-14
GO:CC (Top 5)			
GO:0031514	Motile cilium	26	1.92E-21
GO:0032838	Plasma membrane bounded cell projection cytoplasm	16	2.13E-09
GO:0099568	Cytoplasmic region	16	3.11E-08
GO:0045177	Apical part of cell	16	8.79E-06
GO:0005930	Axoneme	15	1.01E-11
GO:MF (Top 5)			
GO:0016616	Oxidoreductase activity, acting on the NAD or NADP as acceptor	8	6.90E-05
GO:0016614	Oxidoreductase activity, acting on CH-OH group of donors	8	.00012
GO:0003774	Motor activity	8	.00018
GO:0003777	Microtubule motor activity	7	5.06E-05
GO:0004745	Retinol dehydrogenase activity	5	2.55E-06
KEGG			
hsa04640	Hematopoietic cell lineage	6	.00017
hsa04662	B cell receptor signaling pathway	5	.00060
hsa04610	Complement and coagulation cascades	5	.00071

BP = biological process, CC = cellular component, DEG = differentially expressed genes, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular function.

(Table 2). Top 10 hub genes with lowly expression were DNALI1, DNAH5, DNAI2, RSPH1, RSPH4A, CCDC65, ZMYND10, DNAAF1, LRRC6, and WDR63 (Fig. 5B).

3.4. Validation of the hub genes

It is necessary to explore the relationship between hub genes and OS as well as DFS in NPC. Therefore, thanks to GEPIA database, we evaluated the prognostic value of 20 hub genes. Patients with high expression of RSPH4A and ZMYND10 were associated with longer OS (Fig. 6A and B). Furthermore, because of high expression of CDK1 and CDC45, patients were presented with shorter DFS (Fig. 6C and D). Except the above genes, we did not observe that other hub genes were significantly associated with OS or DFS for NPC patients.

3.5. CMap analysis of DEGs

Small-molecule compounds with the potential to treat NPC patients were screened by us from the CMap database based on the uploaded DEGs. Furthermore, the potential small molecular compounds ($P < .05$) were ranked according to *P* values. Among these molecules, piperlongumine, apigenin, menadione, 1,4-chrysenequinone, and chrysin were screened as top 5 negative

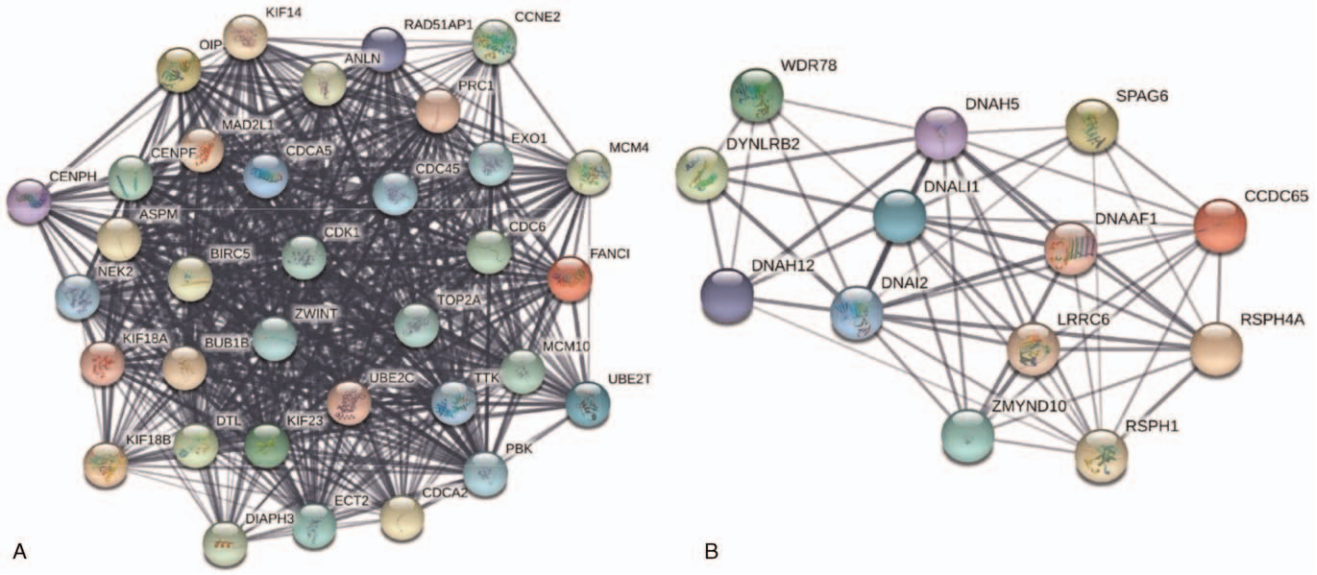

Figure 4. (A) The most significant module of high-expression genes. (B) The most significant module of low-expression genes.

correlation small molecular compounds in terms of connectivity scores. The detailed features are shown in Table 3 and Figure 7.

4. Discussion

Although NPC is a relatively rare cancer in the western countries, it is particularly prevalent in east and southeast Asia.[1,2] The annual incidence rate of NPC in China, especially the southern regions, is the highest in the world.[3] Therefore, it is necessary to explore the potential mechanisms of NPC in order to benefit the diagnosis, treatment, and prognosis assessment. In this present study, we downloaded the 2 gene expression datasets from GEO, and a total of 447 DEGs (164 DEGs with high-expression and 260 DEGs with low-expression) between NPC and normal tissue were identified by using multiple bioinformatics tools.

Table 2

XXX.

Cluster	Term	Description	Count in gene set	*P*
High expression Module 1				
BP	GO:0007059	Chromosome segregation	19	1.85E-22
	GO:0000280	Nuclear division	19	8.83E-21
	GO:0048285	Organelle fission	19	3.83E-20
CC	GO:0005819	Spindle	15	4.44E-16
	GO:0098687	Chromosomal region	13	3.76E-13
	GO:0000775	Chromosome, centromeric region	11	4.23E-13
MF	GO:0016887	ATPase activity	7	1.25E-05
	GO:0008017	Microtubule binding	6	5.49E-06
	GO:0015631	tubulin binding	6	3.22E-05
KEGG	hsa04110	Cell cycle	8	5.33E-11
	hsa04114	Oocyte meiosis	3	.00230
	hsa05166	Human T-cell leukemia virus 1 infection	3	.01010
Low expression Module 1				
BP	GO:0060271	Cilium assembly	10	2.01E-15
	GO:0044782	Cilium organization	10	3.27E-15
	GO:0035082	axoneme assembly	9	1.89E-20
CC	GO:0005930	Axoneme	8	2.28E-15
	GO:0097014	Ciliary plasm	8	2.44E-15
	GO:0031514	Motile cilium	8	4.88E-14
MF	GO:0003774	Motor activity	6	1.64E-10
	GO:0003777	Microtubule motor activity	5	1.65E-09
	GO:0045504	Dynein heavy chain binding	3	8.64E-08
KEGG	hsa05016	Huntington disease	4	9.81E-06
	hsa05014	Amyotrophic lateral sclerosis	4	1.96E-05
	hsa05022	Pathways of neurodegeneration—multiple diseases	4	5.64E-05

BP = biological process, CC = cellular component, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular function.

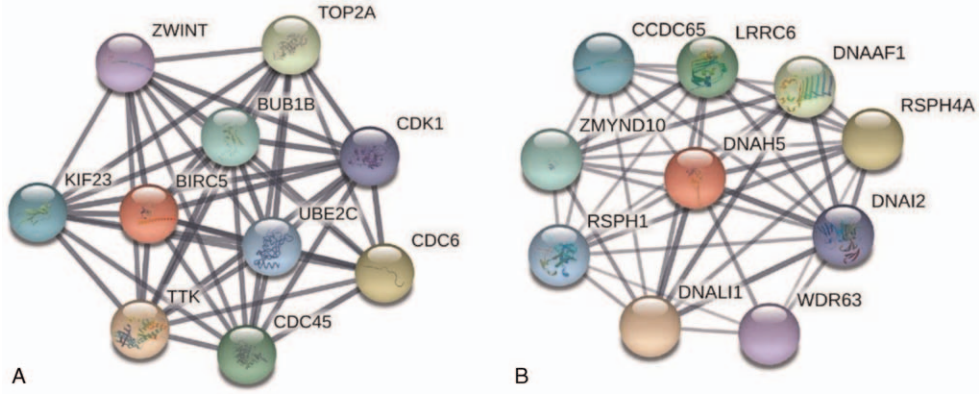

Figure 5. (A) The top 10 upregulation hub genes. (B) The top 10 downregulation hub genes.

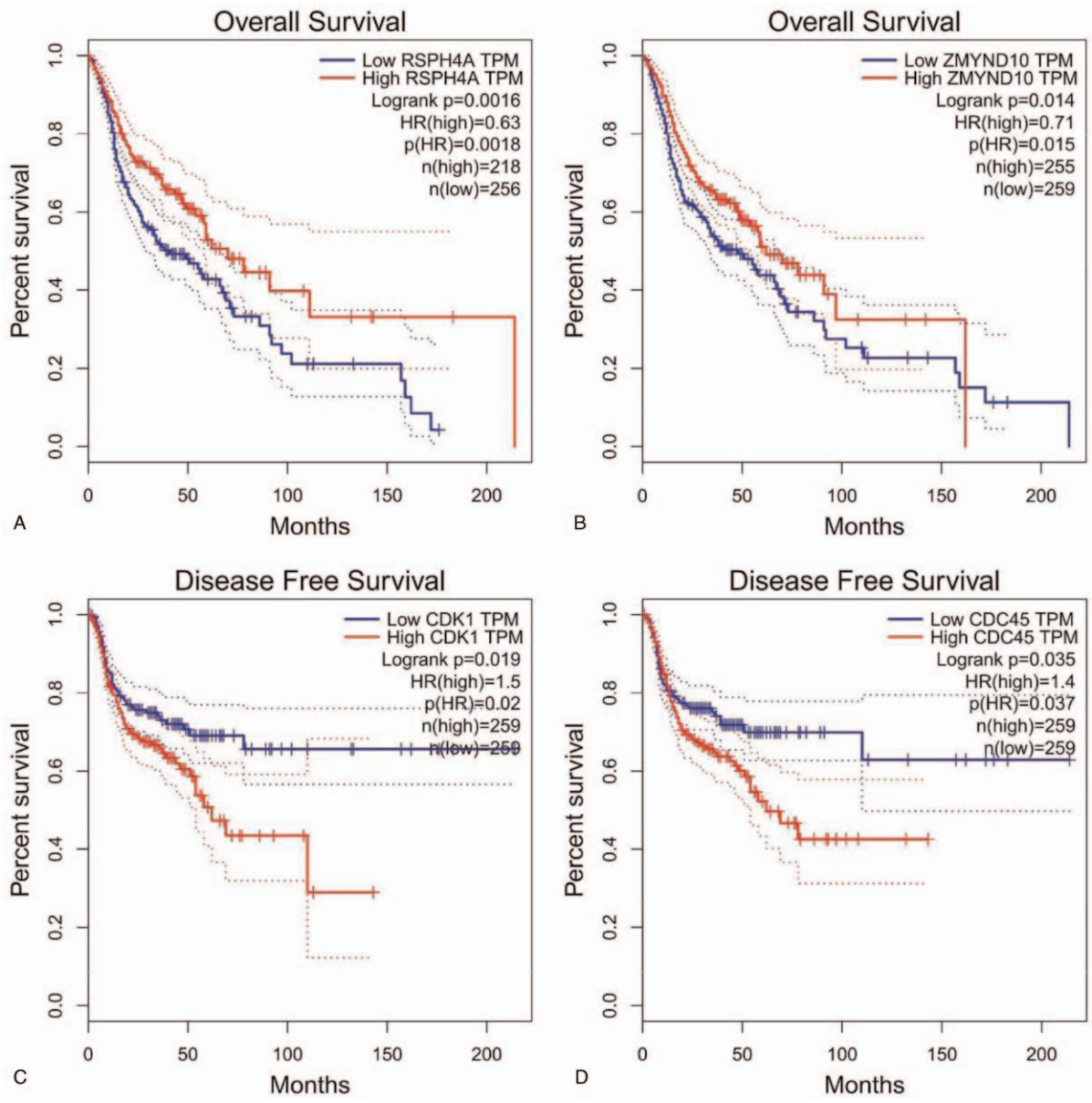

Figure 6. Prognostic value of genes detected in GEPIA database. The survival curve comparing the patients with high (red) and low (blue) expression in NPC. (A) RSPH4A, (B) ZMYD10, (C) CDK1, (D) CDC45.

Table 3

The top 10 most significant small-molecule compounds that could reverse the tumoral metastasis status of NPC.

Rank	Cmap name	Mean	Count	Enrichment score	Molecular formula	*P*
1	Piperlongumine	−0.775	2	−0.959	$C_{17}H_{19}NO_5$	.00364
2	Apigenin	−0.758	4	−0.913	$C_{15}H_{10}O_5$	.00010
3	Menadione	−0.755	2	−0.957	$C_{11}H_8O_2$	.00388
4	1,4-Chrysenequinone	−0.734	2	−0.901	$C_{18}H_{10}O_2$	.01976
5	Chrysin	−0.707	3	−0.883	$C_{15}H_{10}O_4$	.00316

CMap = connectivity map, NPC = nasopharyngeal carcinoma.

About the result in GO-BP and GO-CC enrichment analysis, the majority of upregulated DEGs were mainly enriched in chromosome segregation, nuclear division, nuclear chromosome segregation, spindle, and so on. Obviously, the initiation and progression of NPC are probably associated with the process and component of mitosis. Previous research has established that the prolonged mitosis, in nontransformed cells, is able to trigger a tumor suppressor protein p53-dependent cell cycle arrest.[15] What's more, circular acentric chromosomes, deriving from the shattering and reconstructed chromosome fragments, can be present at very high numbers and often harbor oncogenes that drive tumor development.[16,17] Notably, nearly half of tumors have observed circular acentric chromosomes which could help tumors more adaptable to changing environmental conditions.[18] As for downregulated DEGs, the majority of genes were concerned with cilia tissues, such as cilium assembly, cilium organization, ciliary plasm, and motile cilium. Reviewing previous literature, we speculate that abnormally expressed genes about cilia organization are probably associated with ciliary loss and ciliary dysmorphism in infundibulum mucosa of NPC patients.[19–21] Notably, both GO and KEGG pathway enrichment analyses were associated with extracellular matrix (ECM) for highly expressed genes, such as extracellular matrix organization, extracellular structure organization, extracellular matrix structural constituent, and ECM-receptor interaction. It is not without significance that ECM is important in regulating cell behavior. Tumor cellular transformation and metastasis arose on the basis that ECM structural constituent became disorganized during cancer progression.[22] Additionally, ECM structural constituent could, to some extent, be affected by cancer cells leading to the change in the properties of tumor microenvironment.[23] Notably, Bao et al[24] sequenced normal samples and tumor samples, and then through a comparative analysis, it was found that ECM-receptor interaction pathway is probably an important functional pathway of breast cancer. Therefore, the study of ECM may provide a new direction for future research of NPC.

In the top 10 upregulated hub genes, high-level cyclin-dependent kinase 1 (CDK1) and cell division cycle 45 (CDC45) played important roles in influencing DFS according to survival analysis in the GEPIA database. CDK1 is the only cyclin-dependent kinase that is indispensable for cell cycle progression and CDK1 kinase activity is required for mitotic entry and several mitotic events.[25,26] Furthermore, it has been described in previous studies that CDK1 is essential for tumor formation and progression, such as hepatic tumor and colorectal tumor.[27,28] CDK1 is overexpressed in cancers, undoubtedly, the possibility to inhibit CDK1 for specifically killing tumor cells without little side effects really intrigue researchers. Goga et al[29] proposed that CDK1 inhibition is probably useful with overexpressed MYC in human malignancies. Notably, Luo et al found that miR-96-5p Suppressed the progression of NPC by targeting CDK1.[30] Admittedly, both proliferation and apoptosis of neoplastic cells are related, to some extent, to the defect in the mitotic cell cycle.[31] The protein encoded by CDC45 plays a crucial role in the progression of DNA replication.[32] What's more, Cdc45, as a key target of a Chk1-mediated DNA S-phase checkpoint, also participated in a DNA damage-dependent signal transduction pathway.[33] In non-small cell lung cancer, Huang et al demonstrated that both in vitro and in vivo, CDC45 with low expression could inhibit non-small cell lung cancer cell proliferation by arresting the cells in the G2/M phase of the cell cycle.[34]

On the contrary, in the top 10 downregulated hub genes, radial spoke head component 4A (RSPH4A) and Zinc finger MYND domain-containing protein 10 (ZMYND10) were likely to be associated with OS. The similar study also find that DNALI1, RSPH4A, and DNAI2 are hub genes.[35] RSPH4A, affecting motile cilia, is related to primary ciliary dyskinesia (PCD) of which chronic rhinosinusitis is the representative symptom.[36] What's more, among other hub genes, DNAH5 and DNALI1, belonging to axonemal dynein family, also could cause PCD because of gene mutation.[37] Reviewing literature, some previous studies have reported that epithelium and cilia desquamated after radiotherapy for NPC patients.[19,20] According to Zhou et al's study, the long-term defects of nasal epithelium barrier functions in NPC patients after chemoradiotherapy is related to abnormal expression levels of DNAH5 and DNAI1 and RSPH4A.[21] The protein encoded by ZMYND10 has a zinc finger MYND (myeloid Nervy deformed epidermal auto-regulatory factor-1) with 440 amino acid residues.[38] Maimoona et al used a high-throughput technique to resequence ZMYND10 exons in primary ciliary dyskinesia patients, which demonstrated that ZMYND10 is necessary for motile ciliary function.[39] Furthermore, ZMYND10 also is one of tumor suppressors. According to a previous report, ZMYND10 was able to inhibit growth of nasopharyngeal carcinoma cells by regulation of the JNK-cyclin D1 axis to exert tumor suppression.[40] Therefore, gene CDK1, CDC45, RSPH4A, and ZMYND10 are likely to provide new potential biomarkers for clinical practice or treatment of NPC with further research.

Undoubtedly, chemotherapy exerted a great influence over the treatment for NPC patients. Some small molecules that possess therapeutic efficacy for anti-tumor were identified using cMap tool. Previous research has reported that these small agents, including piperlongumine, apigenin, menadione, and chrysin probably displayed anti-tumor activity. Piperlongumine, a molecule promoting reactive oxygen species production, is able to eliminate malignant cells in an reactive oxygen species-dependent style.[41] Apigenin, extensively existing in plants, is a

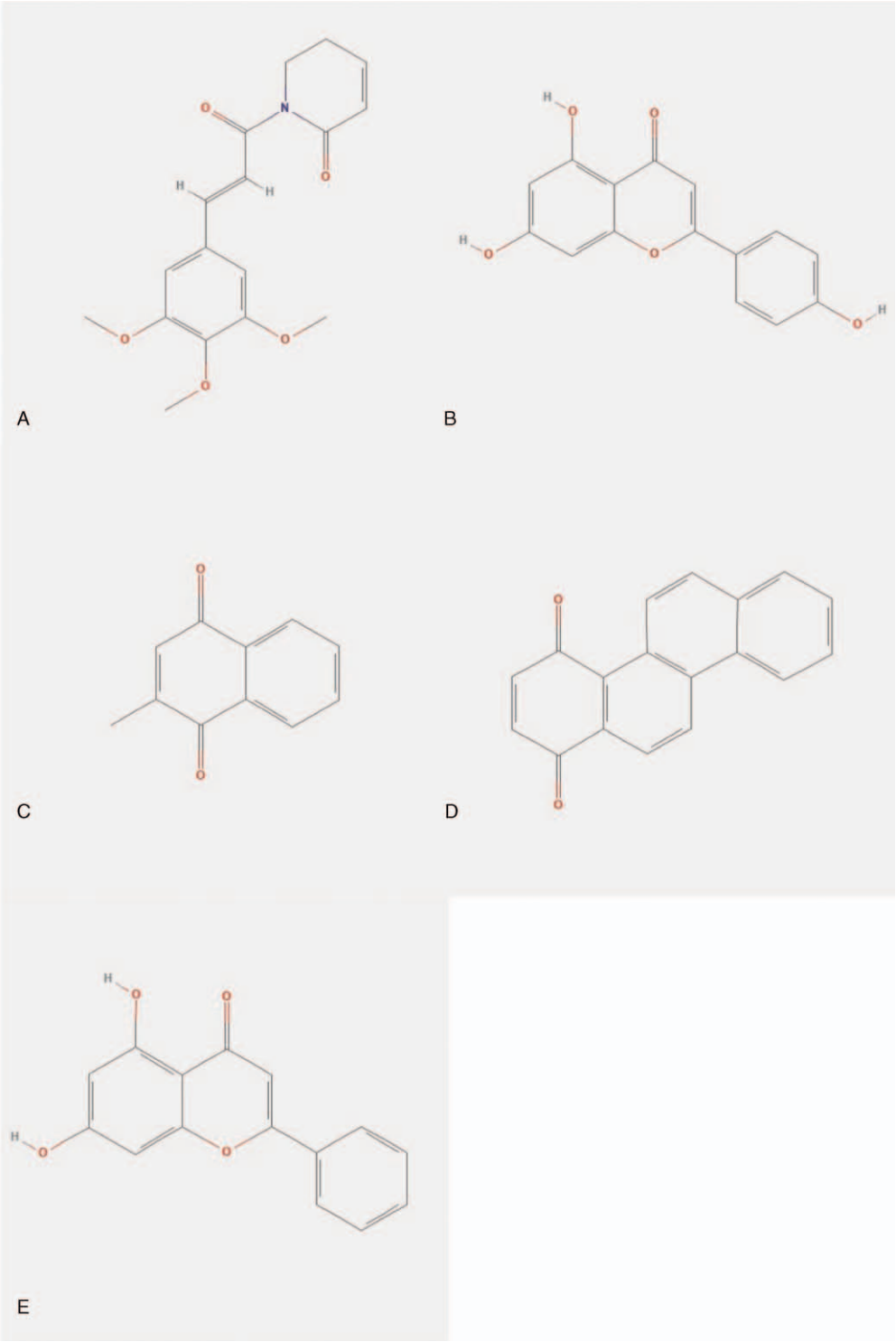

Figure 7. The 2D structures of the 5 compounds. (A) piperlongumine, (B) apigenin, (C) menadione, (D) 1,4-chrysenequinone, (E) chrysin.

natural flavonoid with various biological properties, such as anti-inflammatory, anti-thrombotic, anti-cancer, and so on .[42] Similarly, chrysin is also a natural flavone with anti-cancer effects. Ryu et al have reported that chrysin can reduce expression of proliferating cell nuclear antigen in the prostate cancer and induce apoptosis of cancer cells.[43] As for menadione (vitamin K3), it can trigger cancer cell death combining ascorbate, through inducing replicative stress and activating anti-oxidant systems in colorectal cancer cells.[44] In summary, further research about these above small molecules is needed before they can be brought to clinical trials.

There is no denying that some limitations still existed in this study. Inevitably, a limitation of this analysis is that we only used data from GEO database, and the sample size was relatively

inadequate. It meant that further vivo experiments in the future to verify the outcomes are necessary. What's more, the information acquired from all databases was limited, so with the improvement of the databases, the outcomes probably varied from researchers to researchers.

5. Conclusions

Finally, we identified a total of 424 DEGs and 20 hub genes. Various databases were used to perform bioinformatics analysis. GO function and KEGG pathway enrichment analysis provide more detailed molecular mechanisms for understanding tumorigenesis. Notably, the expression levels of CDK1, CDC45, RSPH4A, and ZMYND10 probably affected survival of NPC patients according to GEPIA database. Also, some possible small molecules were identified, including piperlongumine, apigenin, menadione, and chrysin. These findings may provide a better understanding of the prognostic value of genes and related mechanisms. Besides, our study identified small-molecule compounds which may be efficacious in the treatment of NPC. We hope the above findings had the potential to provide reference for further studies.

Author contributions

Conceptualization: Lucheng Fang.
Data curation: Lucheng Fang.
Formal analysis: Lucheng Fang.
Funding acquisition: Xingwang Rao.
Methodology: Lucheng Fang, Xingwang Rao.
Project administration: Xingwang Rao.
Software: Lucheng Fang, Qinjuan Chen.
Supervision: Qinjuan Chen.
Validation: Wen Wang, Licai Shi.
Writing – original draft: Lucheng Fang, Wen Wang, Licai Shi.
Writing – review & editing: Lucheng Fang, Licai Shi.

References

[1] Chen Y-P, Chan ATC, Le Q-T, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet 2019;394:64–80.
[2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68:394–424.
[3] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115–32.
[4] Scott R. Epstein-Barr virus: a master epigenetic manipulator. Curr Opin Virol 2017;26:74–80.
[5] Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. Nat Rev Cancer 2004;4:177–83.
[6] Peng L, Liu J-Q, Chen Y-P, Ma J. The next decade of clinical trials in locoregionally advanced nasopharyngeal carcinoma. Br J Radiol 2019;92.
[7] Li W-F, Chen N-Y, Zhang N, et al. Concurrent chemoradiotherapy with/without induction chemotherapy in locoregionally advanced nasopharyngeal carcinoma: Long-term results of phase 3 randomized controlled trial. Int J Cancer 2019;145:295–305.
[8] Tao Z, Shi A, Li R, Wang Y, Wang X, Zhao J. Microarray bioinformatics in cancer- a review. J BUON 2017;22:838–43.
[9] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607–13.
[10] Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017;45.
[11] Lamb J. The Connectivity Map: a new tool for biomedical research. Nat Rev Cancer 2007;7:54–60.
[12] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[13] Chin C, Chen S, Wu H, Ho C, Ko M. Lin C. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;(8 suppl):S11.
[14] Bader G, Hogue C. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform 2003;4:2.
[15] Uetake Y, Sluder G. Prolonged prometaphase blocks daughter cell proliferation despite normal completion of mitosis. Curr Biol 2010; 20:1666–71.
[16] Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011;144:27–40.
[17] Zhang C, Spektor A, Cornils H, et al. Chromothripsis from DNA damage in micronuclei. Nature 2015;522:179–84.
[18] Turner K, Deshpande V, Beyter D, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature 2017;543:122–5.
[19] Kamel R, Al-Badawy S, Khairy A, Kandil T, Sabry A. Nasal and paranasal sinus changes after radiotherapy for nasopharyngeal carcinoma. Acta Otolaryngol 2004;124:532–5.
[20] Yin G-D, Xiong G-X, Zhao C, Chen Y-Y. Damage of nasal mucociliary movement after intensity-modulated radiation therapy of nasopharyngeal carcinoma. Chin J Cancer 2010;29:824–9.
[21] Zhou S, Huang H, Chen Q, et al. Long-term defects of nasal epithelium barrier functions in patients with nasopharyngeal carcinoma post chemoradiotherapy. Radiother Oncol 2020;148:116–25.
[22] Crotti S, Piccoli M, Rizzolio F, Giordano A, Nitti D, Agostini M. Extracellular matrix and colorectal cancer: how surrounding microenvironment affects cancer cell behavior? J Cell Physiol 2017;232: 967–75.
[23] Malandrino A, Mak M, Kamm RD, Moeendarbary E. Complex mechanics of the heterogeneous extracellular matrix in cancer. Extreme Mech Lett 2018;21:25–34.
[24] Bao Y, Wang L, Shi L, et al. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. Cell Mol Biol Lett 2019;24:38.
[25] Otto T, Sicinski P. Cell cycle proteins as promising targets in cancer therapy. Nat Rev Cancer 2017;17.
[26] Santamaría D, Barrière C, Cerqueira A, et al. Cdk1 is sufficient to drive the mammalian cell cycle. Nature 2007;448:811–5.
[27] Diril MK, Ratnacaram CK, Padmakumar VC, et al. Cyclin-dependent kinase 1 (Cdk1) is essential for cell division and suppression of DNA re-replication but not for liver regeneration. Proc Natl Acad Sci U S A 2012;109:3826–31.
[28] Costa-Cabral S, Brough R, Konde A, et al. CDK1 is a synthetic lethal target for KRAS mutant tumours. PLoS One 2016;11: e0149099.
[29] Goga A, Yang D, Tward AD, Morgan DO, Bishop JM. Inhibition of CDK1 as a potential therapy for tumors over-expressing MYC. Nat Med 2007;13:820–7.
[30] Luo X, He X, Liu X, Zhong L, Hu W. miR-96-5p suppresses the progression of nasopharyngeal carcinoma by targeting CDK1. Onco Targets Ther 2020;13:7467–77.
[31] Williams GH, Stoeber K. The cell cycle and cancer. J Pathol 2012; 226:352–64.
[32] Moyer SE, Lewis PW, Botchan MR. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc Natl Acad Sci U S A 2006;103:10236–41.
[33] Liu P, Barkley LR, Day T, et al. The Chk1-mediated S-phase checkpoint targets initiation factor Cdc45 via a Cdc25A/Cdk2-independent mechanism. J Biol Chem 2006;281:30631–44.
[34] Huang J, Li Y, Lu Z, et al. Analysis of functional hub genes identifies CDC45 as an oncogene in non-small cell lung cancer - a short report. Cell Oncol (Dordr) 2019;42:571–8.
[35] Ye Z, Wang F, Yan F, et al. Bioinformatic identification of candidate biomarkers and related transcription factors in nasopharyngeal carcinoma. World J Surg Oncol 2019;17:60.
[36] Yoke H, Ueno H, Narita A, et al. Rsph4a is essential for the triplet radial spoke head assembly of the mouse motile cilia. PLoS Genet 2020;16: e1008664.

[37] Lucas JS, Barbato A, Collins SA, et al. European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. Eur Respir J 2017;49.
[38] Hesson LB, Cooper WN, Latif F. Evaluation of the 3p21.3 tumour-suppressor gene cluster. Oncogene 2007;26:7283–301.
[39] Zariwala MA, Gee HY, Kurkowiak M, et al. ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6. Am J Hum Genet 2013;93:336–45.
[40] Zhang X, Liu H, Li B, Huang P, Shao J, He Z. Tumor suppressor BLU inhibits proliferation of nasopharyngeal carcinoma cells by regulation of cell cycle, c-Jun N-terminal kinase and the cyclin D1 promoter. BMC Cancer 2012;12:267.
[41] Chen W, Lian W, Yuan Y, Li M. The synergistic effects of oxaliplatin and piperlongumine on colorectal cancer are mediated by oxidative stress. Cell Death Dis 2019;10:600.
[42] Tang D, Chen K, Huang L, Li J. Pharmacokinetic properties and drug interactions of apigenin, a natural flavone. Expert Opin Drug Metab Toxicol 2017;13:323–30.
[43] Ryu S, Lim W, Bazer F, Song G. Chrysin induces death of prostate cancer cells by inducing ROS and ER stress. J Cell Physiol 2017;232:3786–97.
[44] Kishore C, Sundaram S, Karunagaran D. Vitamin K3 (menadione) suppresses epithelial-mesenchymal-transition and Wnt signaling pathway in human colorectal cancer cells. Chem Biol Interact 2019; 309:108725.