



OPEN ACCESS

EDITED BY

Sharon R. Pine,
University of Colorado Anschutz Medical
Campus, United States

REVIEWED BY

Weiren Luo,
The Second Affiliated Hospital of Southern
University of Science and Technology, China
Liang Yu,
Second Affiliated Hospital of Harbin Medical
University, China
Xiaolong Wang,
Temple University, United States

*CORRESPONDENCE

Jintao Du
✉ 9069402@qq.com

†These authors have contributed
equally to this work and share
first authorship

RECEIVED 13 December 2023

ACCEPTED 09 May 2024

PUBLISHED 24 May 2024

CITATION

Tan Y, Zhou J, Liu K, Liu R, Zhou J, Wu Z, Li L,
Zeng J, Feng X, Dong B and Du J (2024)
Novel prognostic biomarkers in
nasopharyngeal carcinoma unveiled by
mega-data bioinformatics analysis.
Front. Oncol. 14:1354940.
doi: 10.3389/fonc.2024.1354940

COPYRIGHT

© 2024 Tan, Zhou, Liu, Liu, Zhou, Wu, Li, Zeng,
Feng, Dong and Du. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Novel prognostic biomarkers in nasopharyngeal carcinoma unveiled by mega-data bioinformatics analysis

Yishuai Tan^{1,2†}, Jiao Zhou^{3†}, Kai Liu⁴, Ruowu Liu¹, Jing Zhou¹,
Zhenru Wu⁵, Linke Li¹, Jiaqi Zeng¹, Xuxian Feng¹,
Biao Dong⁶ and Jintao Du^{1*}

¹Department of Otolaryngology-Head & Neck Surgery, West China Hospital, Sichuan University, Chengdu, China, ²West China School of Pharmacy, Sichuan University, Chengdu, China, ³Department of Geriatrics, West China Hospital, Sichuan University, Chengdu, China, ⁴College of Biomedical Engineering, Sichuan University, Chengdu, China, ⁵Institute of Clinical Pathology, Key Laboratory of Transplant Engineering and Immunology, NHC, West China Hospital, Sichuan University, Chengdu, Sichuan, China, ⁶State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China

Nasopharyngeal carcinoma (NPC) is commonly diagnosed at an advanced stage with a high incidence rate in Southeast Asia and Southeast China. However, the limited availability of NPC patient survival data in public databases has resulted in less rigorous studies examining the prediction of NPC survival through construction of Kaplan-Meier curves. These studies have primarily relied on small samples of NPC patients with progression-free survival (PFS) information or data from head and neck squamous cell carcinoma (HNSCC) studies almost without NPC patients. Thus, we coanalyzed RNA expression profiles in eleven datasets (46 normal (control) vs 160 tumor (NPC)) downloaded from the Gene Expression Omnibus (GEO) database and survival data provided by Jun Ma from Sun Yat-sen University. Then, differential analysis, gene ontology (GO) enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and network analysis were performed using STRING database. After that, 2142 upregulated differentially expressed genes (DEGs) and 3857 downregulated DEGs were screened. Twenty-five of them were identified as hub genes, which were enriched in several pathways (cilium movement, extracellular matrix structural constituent, homologous recombination and cell cycle). Utilizing the comprehensive dataset we amassed from GEO database, we conducted a survival analysis of DEGs and subsequently constructed survival models. Seven DEGs (RASGRP2, MOCOS, TTC9, ARHGAP4, DPM3, CD37, and CD72) were identified and closely related to the survival prognosis of NPC. Finally, qRT-PCR, WB and IHC were performed to confirm the elevated expression of RASGRP2 and the decreased expression of TTC9, CD37, DPM3 and ARHGAP4, consistent with the DEG analysis. Conclusively, our findings provide insights into the novel prognostic biomarkers of NPC by mega-data bioinformatics analysis, which suggests that they may serve special targets in the treatment of NPC.

KEYWORDS

nasopharyngeal carcinoma, survival model, prognosis, mega-data bioinformatic analysis, scRNA-seq

Introduction

Nasopharyngeal carcinoma (NPC) is a distinct head and neck cancer, with patients mainly from East Asia, Southeast Asia, and North Africa (1). The incidence and mortality of NPC in China is high globally. An estimated 42,100 new cases and 21,320 deaths were attributed to NPC in China in 2013, accounting for 1.14% of all new cancer cases and 0.96% of all cancer-related deaths (2).

NPC is increasingly considered an ecological disease, with the tumor microenvironment (TME) playing a pivotal role in the cancer's progression (3). The NPC microenvironment is characterized by an increased abundance of suppressive regulatory T cells (Tregs) (4). Furthermore, recent research indicates that malignant epithelial cells can activate Tregs through the CD70-CD27 pathway, exacerbating immune suppression (5). Innate-like B cell has been identified as a potential biomarker for gemcitabine-plus-platinum (GP)-based treatment in NPC, which could improve patient management (6). Professor Ma Jun's team has made further refinements to the 8th edition of the AJCC/UICC TNM classification, which holds promise for enhancing patient treatment outcomes (7). Moreover, due to the lack of clinical information specific to NPC, most of studies have utilized data from head and neck squamous cell carcinoma (HNSCC) for survival analysis. However, this approach is flawed as the HNSCC database contains minimal NPC data, a tumor type distinct from HNSCC, rendering the survival results largely irrelevant (8–11). NPC can be effectively treated in early stages, yet over 50% of patients are diagnosed at advanced stages, leading to worse outcomes and prognosis (12). Consequently, there is a critical need for the identification of biomarkers that enable early diagnosis and prognostication, thus guiding high-risk individuals towards timely follow-up and intervention.

Recently, microarrays and high-throughput sequencing have become powerful and effective tools that allow the detection of genome-wide gene expression differences in healthy populations and cancer patients (13). It is possible to define the gene expression profile of the tumor and to correlate it with prognosis and metastasis (14). The Gene Expression Omnibus (GEO) repository at the National Center for Biotechnology Information (NCBI) archives and freely distributes high-throughput molecular abundance data, predominantly gene expression data generated by DNA microarray technology (15). The cBioPortal for Cancer Genomics (<http://cbioportal.org>) provides a Web resource for exploring, visualizing, and analyzing multidimensional cancer genomics data (16). However, there is a lack of multiple database combination analysis and screening of DEGs from mega data.

Hence, samples of NPC from the GEO database were collected and analyzed based on the principle of bioinformatics to screen the differentially expressed genes (DEGs). Combined with the GEO database and the clinical information with related NPC RNA profile data from Jun Ma (17), we used these screened genes to perform subsequent survival analysis. Multivariate Cox regression analysis was performed to verify the credibility of the DEGs. Finally, seven DEGs (RASGRP2, MOCOS, TTC9, ARHGAP4, DPM3, CD37, and

CD72) were identified and closely related to the survival prognosis of NPC. Subsequently, the survival models constructed by the above identified DEGs predicted patient prognosis well. Therefore, our study constructed a prognostic model and identified novel prognostic biomarkers of NPC by mega-data bioinformatics analysis, which suggests that they may serve special targets in the treatment of NPC.

Materials and methods

Data acquisition and principal component analysis

Microarray datasets (GSE12452 (18), GSE64634 (19), GSE15047, GSE15903 (20), GSE36972 (21), GSE48501 (22), GSE100193 (23), GSE79571 (24), GSE119020 (25), GSE127848, GSE128502 (26), GSE13597 (27), GSE53819 (28), GSE40290, GSE34573 (29)) were accessed from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), which is an international public repository for high-throughput microarray and next-generation sequence functional genomic datasets submitted by the research community (30). Because these datasets belong to different Geo platforms (GPL) and the number of genes varies greatly between platforms, we combined datasets that belong to the same platform (GPL570) into a group (MIX) to avoid the loss of a large number of genes. Finally, these datasets were divided into five groups for further experiments: GSE34573, GSE53819, GSE13597, GSE40290 and MIX (GSE34571: 3 normal samples and 14 tumor samples, GSE53819: 18 normal samples and 18 tumor samples, GSE13597: 3 normal samples and 25 tumor samples, GSE40290: 8 normal samples and 25 tumor samples, and MIX: 14 normal samples and 78 tumor samples). Because the MIX group's data came from different datasets, we used limma packages to remove batch effects. Meanwhile, GSE102349 and GSE132112 were obtained for the survival analysis. More detailed sample information and grouping results are presented in Table 1. PCA was conducted via R package FactoMineR and factoextra, which was further visualized by using scatterplot3d function from R package scatterplot3d.

Selection and identification of DEGs

We performed a differentially expressed gene (DEG) analysis of each of the five groups through the limma package with filter criteria: $\log_2FC > 1$ and $p \text{ value} < 0.05$. The limma package is an R/Bioconductor software package that provides an integrated solution for analyzing data from gene expression experiments (31). Then, the ComplexHeatmap function was used to display the top 50 upregulated and top 50 downregulated DEGs of each group (rank according to log fold change). Meanwhile, we continued to show the genes that met the screening criteria through the EnhancedVolcano package. The Venn diagram presents the overlap of DEGs using jvreen (<http://jvreen.toulouse.inra.fr/>) online tools.

TABLE 1 Details of the five groups.

Dataset	Platform	PMID	Sample type	group
GSE12452	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	17119049	NPC	MIX
GSE64634	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	26246469	NPC	MIX
GSE15047	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	unknown	CNE, HONE	MIX
GSE15903	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	19963132	NPC	MIX
GSE36972	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	23069661	NPC 5-8F	MIX
GSE48501	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	24498188	NPC	MIX
GSE100193	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	28906488	CNE	MIX
GSE79571	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	28423621	HONE	MIX
GSE119020	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	30683844	HONE	MIX
GSE127848	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	unknown	NPC	MIX
GSE128502	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	31321241	NPC-TW01	MIX
GSE13597	GPL96 [HG-U133A] Affymetrix Human Genome U133A Array	19142888	NPC	GSE13597
GSE53819	GPL6480 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F	24763226	NPC	GSE53819
GSE40290	GPL8380 Capitalbio 22K Human oligo array version 1.0	unknown	NPC	GSE40290
GSE34573	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	22815911	NPC	GSE34573

Enrichment analysis for nasopharyngeal carcinoma

To further acknowledge the significant role that DEGs play in biological functions, enrichment analysis was carried out. We adopted clusterProfiler together with GOplot to achieve Gene Ontology (GO) term enrichment analysis, which was divided into three types: biological processes (BP), cellular components (CC), and molecular functions (MF). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was also performed. Furthermore, we also show the significant DEGs in the key pathways through the function GOchord from the GOplot R package.

Construction of the protein–protein interaction network and TF–gene network

Overlapping DEGs that appeared in at least four groups were used to construct the PPI network by STRING (STRING: functional protein association networks (string-db.org)). They are also displayed in Table 2. We prefer to achieve a comprehensive and objective global network, including direct (physical) as well as indirect (functional) interactions for DEGs by STRING (32). Cytoscape software was used to calculate the degree of connectivity between DEGs to identify hub genes. Concurrently, we built the TF–gene network of DEGs in NetworkAnalyst (www.networkanalyst.ca), which is a Web-based tool for creating and visualizing biological networks, and users can now perform gene expression profiling for 17 different species (33).

Manufacture of the DEG mutation spectrum diagram

To investigate the mutation standard of the overlapping DEGs, we constructed the mutation signature by cBioPortal online tools. The high mutation frequency in DEGs is shown based on NPC mutation datasets. This dataset was obtained from a study of the genomic landscape of NPC (34).

Gene screening based on clinical information

Two datasets, GSE132112 (17) and GSE 13597, were chosen for further research on DEGs on account of clinical information. Different radiotherapy and chemotherapy strategies should be applied to patients with different stages (35). Accordingly, we studied the effects of DEG quantity on stage and metastasis to assist in choosing the method of treatment. Thus, we plotted a boxplot of genes associated with tumor-node-metastasis (TNM) staging using the ggstatsplot R package. The DEGs that were strongly associated with staging were further analyzed for survival.

Survival prognostic analysis for DEGs

The Kaplan–Meier curve is a comparative analysis that depends upon the whole curve and not upon isolated points (36). The genes used to plot the curves were those identified by differential analysis. Meanwhile, some of the differentially expressed genes (DEGs) were related to the stages and grades of the patients or had a high hazard

TABLE 2 Overlapping differentially expressed genes (DEGs).

Regulation	Overlapped differentially expressed genes(DEGs)
Up-regulated DEGs (appear in five groups*)	KIF23, TOP2A, BUB1B, RBBP8, PRC1, GALNT11, VRK2, NFE2L3, TNFAIP6, LAMB1
Up-regulated DEGs (appear in four groups*)	CDK1, ASPM, RAD51AP1, LMNB2, DTL, NPL, ZWINT, PMAIP1, FJX1, GINS1, STAR, ECT2, RCN1, CCL8, COL4A1, ZIC2, IDH1, GRB10, COL5A2, USP18, TTK, PLAU, MMP1, PTGS2, CCNB2, TPX2, TMPO, GJA1, BRCA1, INSM1, GAD1, KIF14, LHX2, TFRC, PUS7, MEST, FOXM1, ITGAV, ROBO1, KREMEN2, GADD45A, ARNT2, RAI14
Down-regulated DEGs (appear in five groups*)	MSLN, LTF, SCGB1A1, CRYM, SPAG6, CYP2F1, CYP4B1, AGR2, SLPI, MUC13, MSMB
Down-regulated DEGs (appear in four groups*)	UPK1B, ALOX15, KRT4, CDH26, CAPS, ALDH3A1, DHRS9, MUC20, RRAD, TEKTI, FAM3D, S100P, DNAI1, SPA17, ROPN1L, BCAS1, ZMYND10, CH25H, CHST9, CXCL14, IQCD, C9orf116, MAL, HSPB8, CEACAM6, GCNT3, C4orf19, DNALI1, SCGB2A1, ALDH1L1, PIP, TMC5, CLDN10, ATP2C2, FOLR1, TSPAN1, PACRG, MUC16, TFF3, SCGB3A1, ZMYND12, RIBC2, SLC22A4, AK7, FUT3, VILL, MLPH, SELENBP1, ANG, NDRG2, TPPP3, NME5, LCN2, SERPINB7, ADRA2A, PROM1, FOXJ1, BLK, CR2, ATP2A3, AQP5, CLU, PIGR

*including GSE34573, GSE53819, GSE13597, GSE40290 and MIX; # Four of GSE34573, GSE53819, GSE13597, GSE40290 and MIX.

ratio. The progression-free survival information was downloaded from GSE102349 (37), which provides 113 treatment-naïve undifferentiated NPC tumor gene expression matrices and clinical information. The overall survival (OS) and failure-free survival were completed using GSE132112. All survival prognoses were analyzed for selected DEGs by the survival and survminer R packages.

Analysis of high-risk genes based on crucial cancer clinical trial endpoints

Multivariate Cox regression analysis was constructed to forecast the patient's prognosis. The above three models were arranged for overall survival (OS), progression-free survival (PFS), and failure-free survival (FFS). At the same time, patients were divided into low- and high-risk groups according to the median risk score (38, 39). The risk score was calculated as follows: risk score = $\exp(1) \cdot h(1) + \exp(2) \cdot h(2) + \exp(3) \cdot h(3) + \exp(4) \cdot h(4) + \dots + \exp(n) \cdot h(n)$, where $\exp(x)$ represents the gene expression level, and $h(x)$ represents the regression coefficient calculated by multivariate Cox proportional hazard regression). These key genes are integrated to build risk models.

Single cell sequencing data analysis for nasopharyngeal carcinoma

The public sequencing data were downloaded from the Human Cell Atlas(4 nasopharynx) (40) and GEO (GSE162025 (41):

1 nasopharynx and 5 nasopharyngeal carcinoma). All scRNA-seq data meet the quality control criteria: chondriosome gene lower than 5% and the number of total genes is between 200 and 5000. We use R package Seurat (v 4.3.0) to merge the three scRNA seq data and analyze the spatial transcriptome data. Then, R package harmony (42) was used to remove the batch effects of the three different origin datasets. Expression levels visualization was using R package ggplot2, Seurat and scCustomize.

NPC sample collection and isolation of total RNA for real-time quantitative PCR

Seven NPC samples and four control samples were collected. All patients signed informed consent forms, which were approved by the hospital ethics committee. The diagnosis of all patients was confirmed by pathology. Total RNA from specimens was isolated using the FastPure Cell/Tissue Total RNA Isolation Kit V2 (Vazyme, Nanjing, China) according to the manufacturer's instructions. mRNA to cDNA was synthesized using Bimake All-in-One cDNA Synthesis SuperMix (Bimake, Houston, USA), and qPCR of mRNAs was performed with Bimake 2x SYBR Green qPCR master mix (Bimake). Relative expression was calculated with the $2^{-\Delta\Delta CT}$ method, and levels were normalized using GAPDH for mRNAs. The sequences of primers used in this study are shown in Table 3.

Western blotting

After being thoroughly grounded in liquid nitrogen, the tissues were collected for western blotting (WB). Total protein of tissues or cells were extracted in RIPA lysis buffer with a cocktail of protease inhibitors (Bimake), and then boiling for 10 min with SDS loading buffer. Equal amounts of protein were electrophoresed on SDS-PAGE in 10% Tris-glycine gels and transferred to PVDF membranes (Millipore, MA, USA). Membranes were blocked with 5% non-fat milk in tris buffered solution containing 0.1% Tween-20 at room temperature for 1 h, followed by overnight incubation with primary antibodies against GAPDH, anti-KIF2C (Proteintech) antibody, anti-CD37 antibody (HUABIO), anti-SHP1 (PTPN6) antibody (HUABIO), CDT2 (DTL) rabbit mAb (ZENBIO), and ARHGAP4 rabbit pAb (ZENBIO). After washing thrice at room temperature, the membranes were incubated with secondary antibody (Zen Bioscience, Chengdu, China) and signals were visualized by using ECL Plus Western Blotting Reagent Pack (Bio-Rad, Hercules, USA). The band intensities were quantified by Fusion Solo Imaging System (VIBER LOURMAT, FRANCE).

Immunohistochemistry and statistical analysis

Immunohistochemistry (IHC) staining was carried out using the Biotin-Streptavidin horseradish peroxidase Detection Kit (SP-9001, ZSGB-Bio, Beijing, China) and the DAB Detection Kit (ZLI-

TABLE 3 PCR primers.

	Forward primer	Reverse primer
GABRP	TTTCTCAGGCCCAATTTTGGT	GCTGTCGGAGGTATATGGTGG
PTPN6	TGAACTGCTCCGATCCCACTA	CACGCACAAGAAACGTCCAG
ARHGAP4	GATGACGGGGATGTGCTTGAT	GGTAGAAGGTTTCGGTCTCCTG
TTC9	TGAAGCCATAGGCAAATACCAC	TGACTCGTTCATAGTTACCAGC
CD37	TCCTGAGAGGTAACGGGTCG	GGATTGTGGAGTCGTTGGTCCG
DTL	TGGTCTTCAACAATACCCTCTTCA	CTTCATTGGCAACTGCTAGTACA
RASGRP2	TCGCCTGTCAGTTGAGTGTC	CCAGTAGCCCTGCATCCTTC
DPM3	GGCCACTTTTCATGACTGCG	GAAATGGGAGGAAGGGCTGT

9061, ZSGB-Bio). For each section, three fields of view were selected, and images were captured using a microscope at 200× magnification. The tissue sections were subjected to immunostaining using the following primary antibodies: anti-KIF2C polyclonal antibody (1:200, Proteintech), anti-CD37 antibody (1:100, HUABIO), anti-SHP1 (PTPN6) antibody (1:100, HUABIO), CDT2 (DTL) rabbit mAb (1:100, ZENBIO), and ARHGAP4 rabbit pAb (1:100, ZENBIO). Immunohistochemical staining intensity was quantified by calculating the average rate of positively stained cells using ImageJ and IHC Profiler across three distinct high-power fields (200× magnification) on each slide (Supplementary Table S1). The median immunoreactivity score for each gene served as the threshold for our analyses. We assessed prognostic variables through bivariate Kaplan-Meier log-rank tests and multivariate Cox proportional hazards modeling. Statistical significance was set at a two-sided P value of less than 0.05. Survival curves were conducted using GraphPad Prism version 9.

Results

Identification of differentially expressed genes between nasopharyngeal carcinoma and normal tissues after data standardization

Principal component analysis (PCA) was performed to determine whether cancer and normal samples could be separated in each group of data. The results showed an apparent distinction between NPC and control tissue (Figures 1A–F). Then, the expression value of each group was designed as a boxplot with the sample median and expression value being similar in each group, which means that the quality of each group of data was acceptable (Figures 1G–K). Jointly, these results demonstrated that our grouping was reasonable and met the conditions for further identification of DEGs.

Heatmaps of the top 50 upregulated DEGs and top 50 downregulated DEGs in each group are shown (Figure 1A). Additionally, the volcano plots showed that DEGs were

differentially expressed between NPC and control tissue samples with the following filter criteria: $|\log_2(\text{fold change})| > 1$ and $p\text{-value} < 0.05$ (Figure 2B). All DEGs from the five groups were plotted as Venn diagrams (Figures 2C, D). The overlapping DEGs that appeared in at least four groups and the other overlapping DEGs are listed in the supplement (Supplementary Data Sheets 1, 2). Eventually, we identified 2142 upregulated genes and 3157 downregulated genes (filter criteria: $\log_2(\text{fold change}) > 1$ and $p\text{-value} < 0.05$). A more stringent filter of $|\log_2(\text{fold change})| > 1.5$ with the same p-value criterion revealed 1,316 upregulated and 2,718 downregulated genes (Supplementary Data Sheet 3).

Functional enrichment analysis of overlapping DEGs

To further study the function of DEGs (appearing in at least four groups), we carried out enrichment analysis by GOplot and the clusterProfiler R package. The GO analysis showed that the following BPs (biological processes) were notably enriched among the DEGs: axoneme assembly, cilium movement and microtubule-based movement. The CCs (cellular components) were mainly enriched in motile cilia, microtubules, axonemes and mitotic spindles, and MFs (molecular functions), such as extracellular matrix structural constituent and microtubule motor activity, were highly associated with the DEGs (Figures 3A–F). The KEGG pathway analysis revealed that DEGs in NPC were mainly enriched in the cell cycle, platinum drug resistance, p53 signaling pathway, metabolism of xenobiotics by cytochrome P450, amoebiasis and viral protein interaction with cytokine receptor (Figures 3G, H). Among all three GO terms and KEGG pathway analyses, we classified primary pathway-relevant DEGs through a chordal graph. These pathways included BP (BUB1, CXCL10, SLPI, GNLY, PRC1, TPPP3) (Figure 3D); CC(CDK1, GJA1, COL22A1, LAMB1) (Figure 3E); MF (DYNLRB2, DEFB1, KIF23) (Figure 3F); and KEGG(CDK1, CXCL17, CYP2F1) (Figure 3I), which contributed to critical directions for our future studies. These enrichment pathways indicated that the DEGs were involved in tumor migration, tumor microenvironment and pathogenesis of NPC.

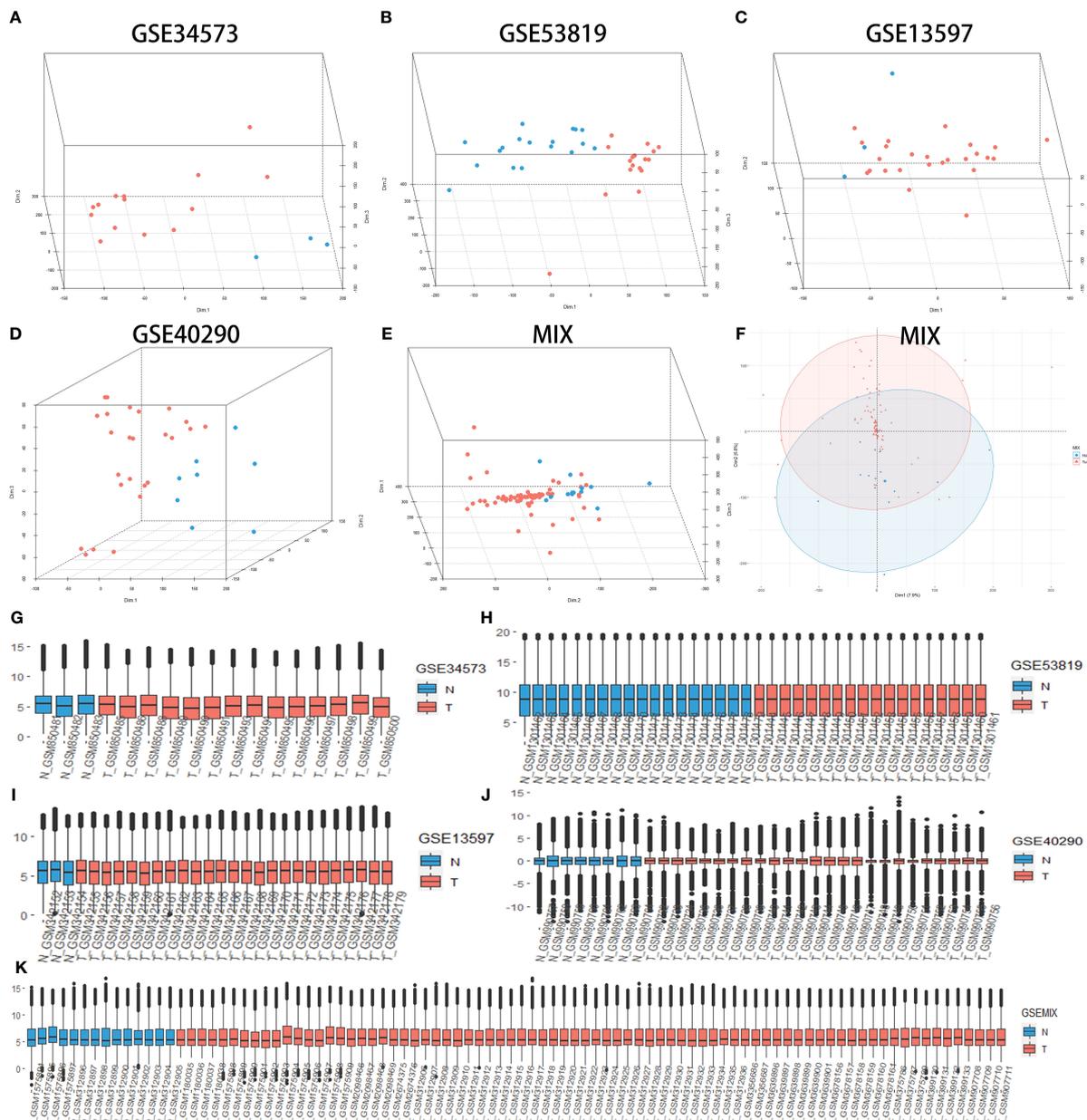
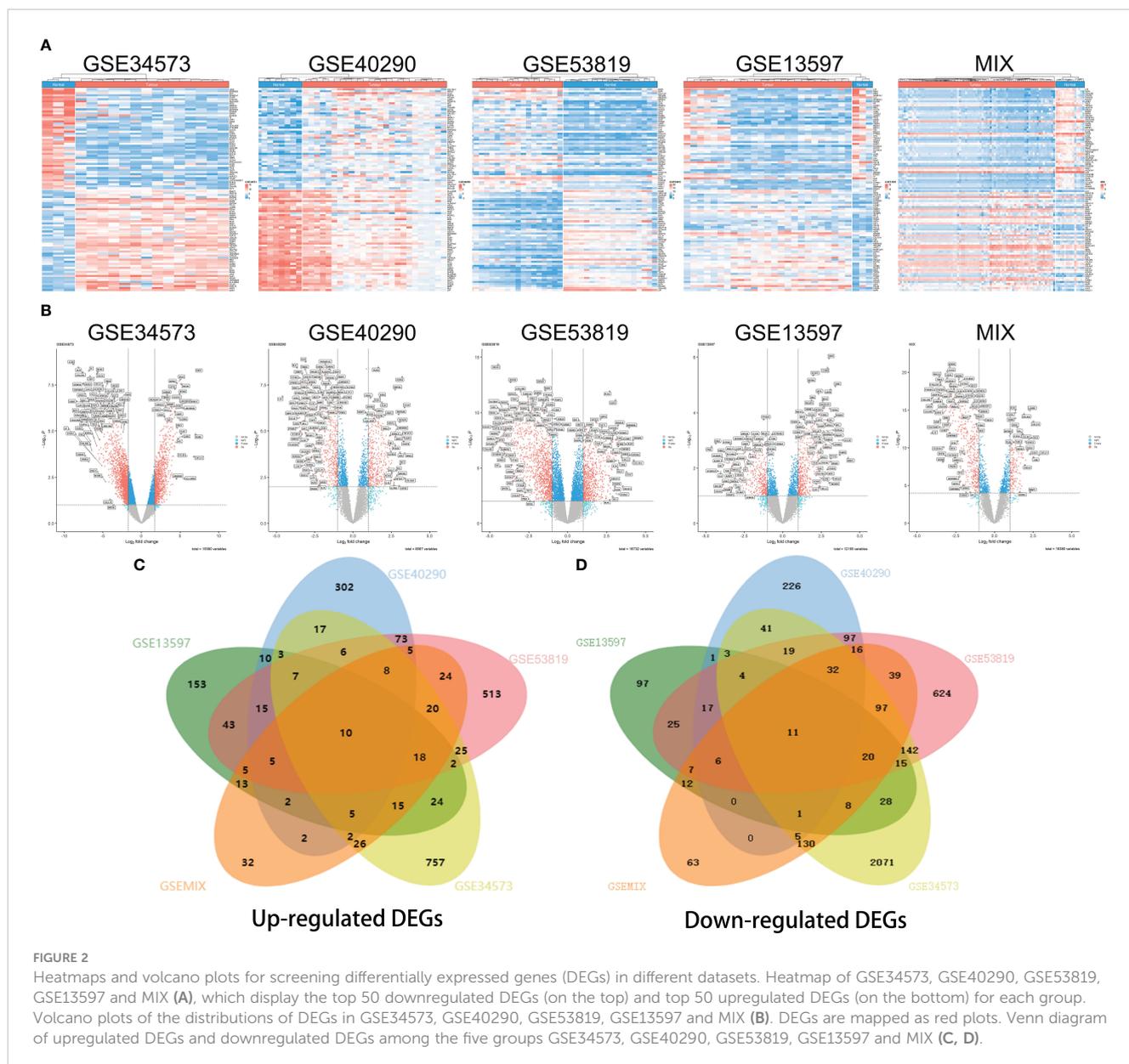


FIGURE 1 Principal component analysis (PCA) shows the clustering of different samples in the RNA expression matrix in different groups, quality control (QC) for all groups. PCA between nasopharyngeal carcinoma (NPC) and normal tissue samples (control) in GSE34573 (A), GSE53819 (B), GSE13597 (C), GSE40290 (D) and MIX (E, F). Blue plots represent normal tissue, and red plots represent NPC. QC for GSE34573, GSE53819, GSE13597, GSE40290 and MIX (G-K). The red box represents normal tissue, and the cyan box represents NPC.

Multiple network analysis of DEGs and identification of DEG mutation spectrum

To further explore the relationship among these overlapping DEGs at the protein level, we constructed a protein–protein interaction (PPI) network. Based on the network analysis of Cytoscape, 16 upregulated DEGs, CDK1, DTL, FOXM1, TPX2, CCNB2, BUB1B, KIF23, RAD51AP1, ASPM, TTK, TOP2A, ECT2, PRC1, ZWINT, BRCA1 and IDH1, were identified as hub genes with degrees>15 (Figure 4A). Nine downregulated DEGs, SCGB1A1, ROPN1L, FOXJ1, LTF, DNAL1, TFF3, SPAG6, AGR2, and

CEACAM6, were confirmed as hub genes with degrees>5 (Figure 4C). Moreover, the TF-gene interaction network showed that ELF1, SMAD5, TFDP1 and MAZ were highly related to DEGs, which played key roles in DEG transcription (Figures 4B, D). Furthermore, POLR1B and PTPN6 play a key role as nodes in the network diagram drawn by DEGs (Figures 4E, F). When all the DEGs were analyzed together, the results showed that the upregulated and downregulated hub DEGs were related to each other (CDK1-BLK, MMP1-LCN2, KIF14-VILL) (Figure 4G). In summary, these results indicated that interactions between DEGs could contribute to the occurrence and development of NPC.



Moreover, we used the cBioProtal Web tools online to carry out genetic mutation detection based on NPC datasets. Accordingly, 32 DEGs were found in the top mutation, including 12 upregulated DEGs (Supplementary Figure S1A) and 20 downregulated DEGs (Supplementary Figure S1B). Among them, TTN, FAT2, KLHDC7A and TNXB were highly associated with survival prognosis (Supplementary Figure S2). This result indicated that a considerable number of valuable DEGs were mutated in NPC and were associated with the survival of NPC.

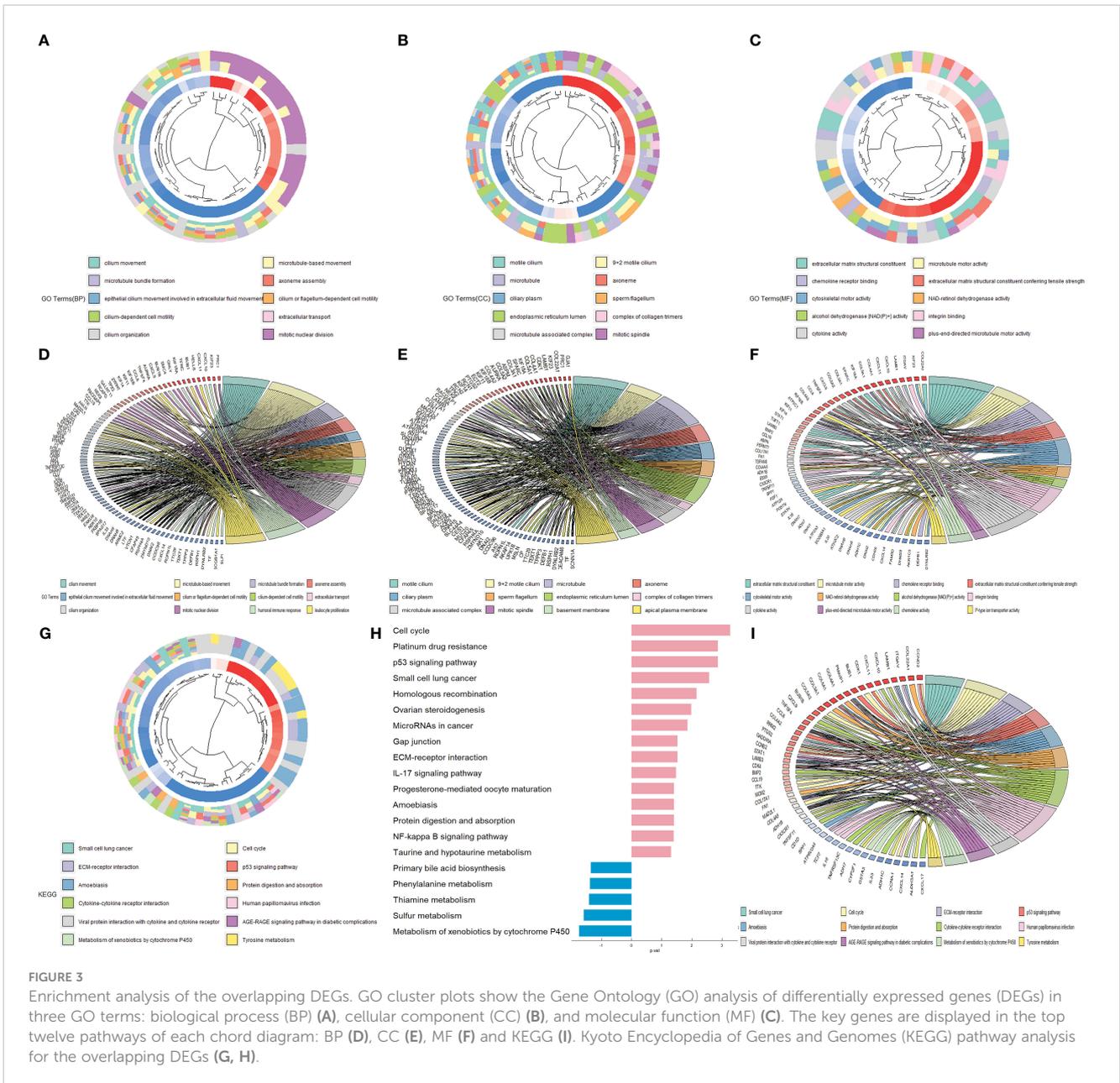
Survival analysis of the overlapping DEGs

Kaplan–Meier survival curves were drawn for NPC patients based on their overall survival, progression-free survival and failure-free survival information. Eight overlapping DEGs were screened out from more than 1000 overlapping DEGs, which have a strong

relationship with patient survival (Figure 5). Among them, two upregulated DEGs (KIF2C and DTL) contributed to poor prognosis with lower survival rates (Figures 5A, B). Meanwhile, six downregulated DEGs (CD37, DPM3, GABRP, PTPN6, TTC9, ARHGAP4) were also screened out with a strong association with decreased OS, FFS and PFS (Figures 5C–H). These findings validated that these identified DEGs were closely related to the prognosis of the patients.

Multivariate Cox regression analysis of DEGs and construction of the survival prediction model

After selected genes were subjected to multiple Cox analysis, the DEG risk forest diagrams were constructed. According to the calculation of the risk score, patients were divided into high-risk



and low-risk groups (Supplementary Data Sheet 4). Remarkably, some of the selected genes were present in the overlapping DEGs (identified from the five groups) (Supplementary Data Sheets 1, 2). In-depth analysis revealed that seven DEGs (MOCOS, RASGRP2, CD37, DPM3, CD72, TTC9, and ARHGAP4) play a critical role in the survival of patients with NPC (Figures 6A–C). Notably, the DEGs we selected can jointly construct three survival prediction models (OS: p value=0.00044; FFS: p value<0.0001; PFS: p value<0.0001). The survival prognosis of patients classified into the high-risk group in these three models was significantly worse than that of patients in the low-risk group (Figures 6D–F). The value of the ROC curve is close to one, which proves that the reliability of the three models is very high (Figures 6G–I). Combined with the above PPI network analysis, several DEGs as

the nodes in the network diagram were the key genes in the survival model, including POLR1B and PTPN6. Together, these models could be applied to predict the prognosis of patients, which indicated that the selected DEGs were positive factors affecting the prognosis of the patients.

The significance of clinical information for genetic screening

Correlation analysis was performed between DEGs and clinical information (TNM staging system), and six overlapping DEGs were closely related to clinical information. The upregulated DEGs (TOP2A, INSM1) increased with the progression of T and N

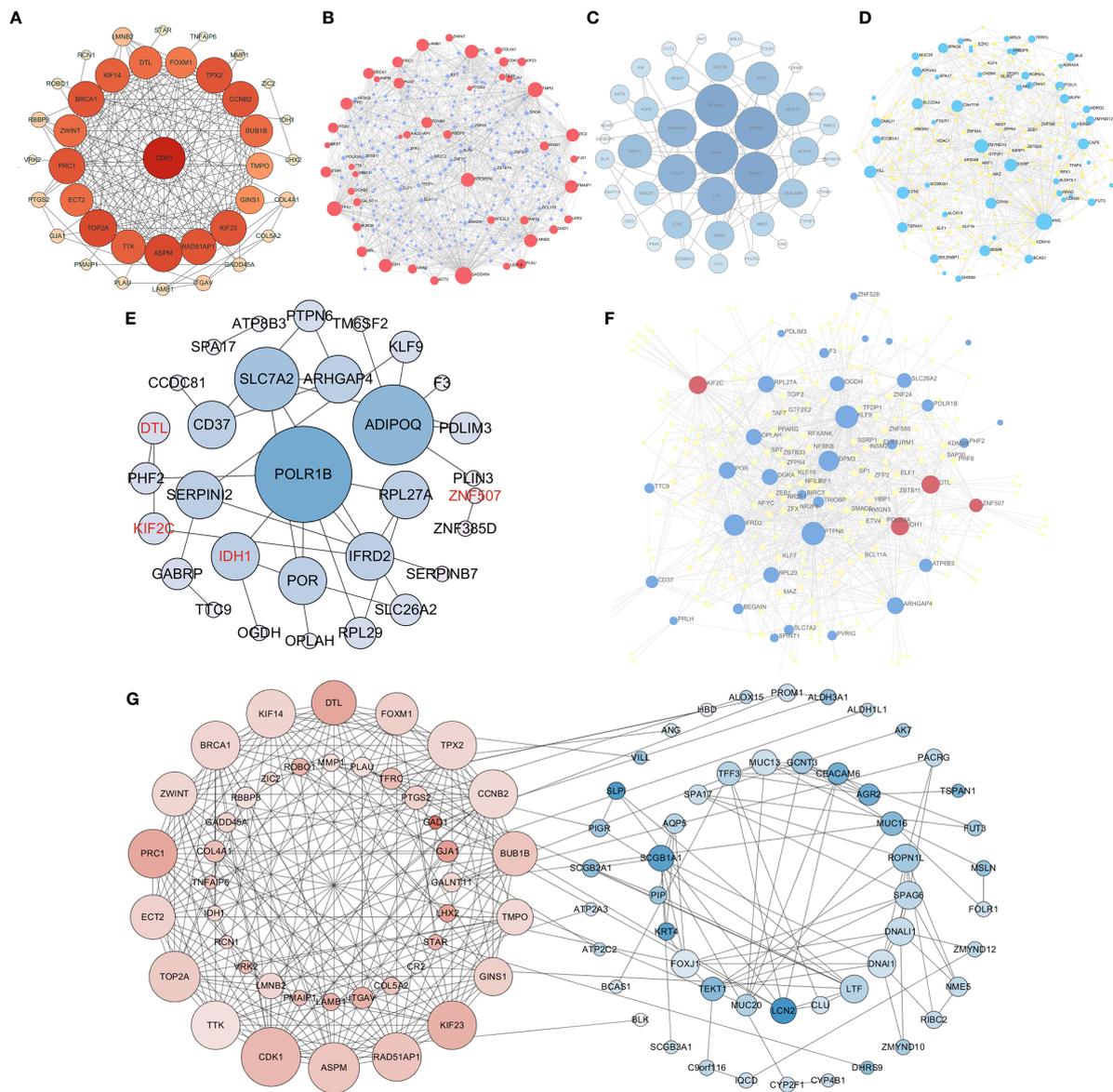
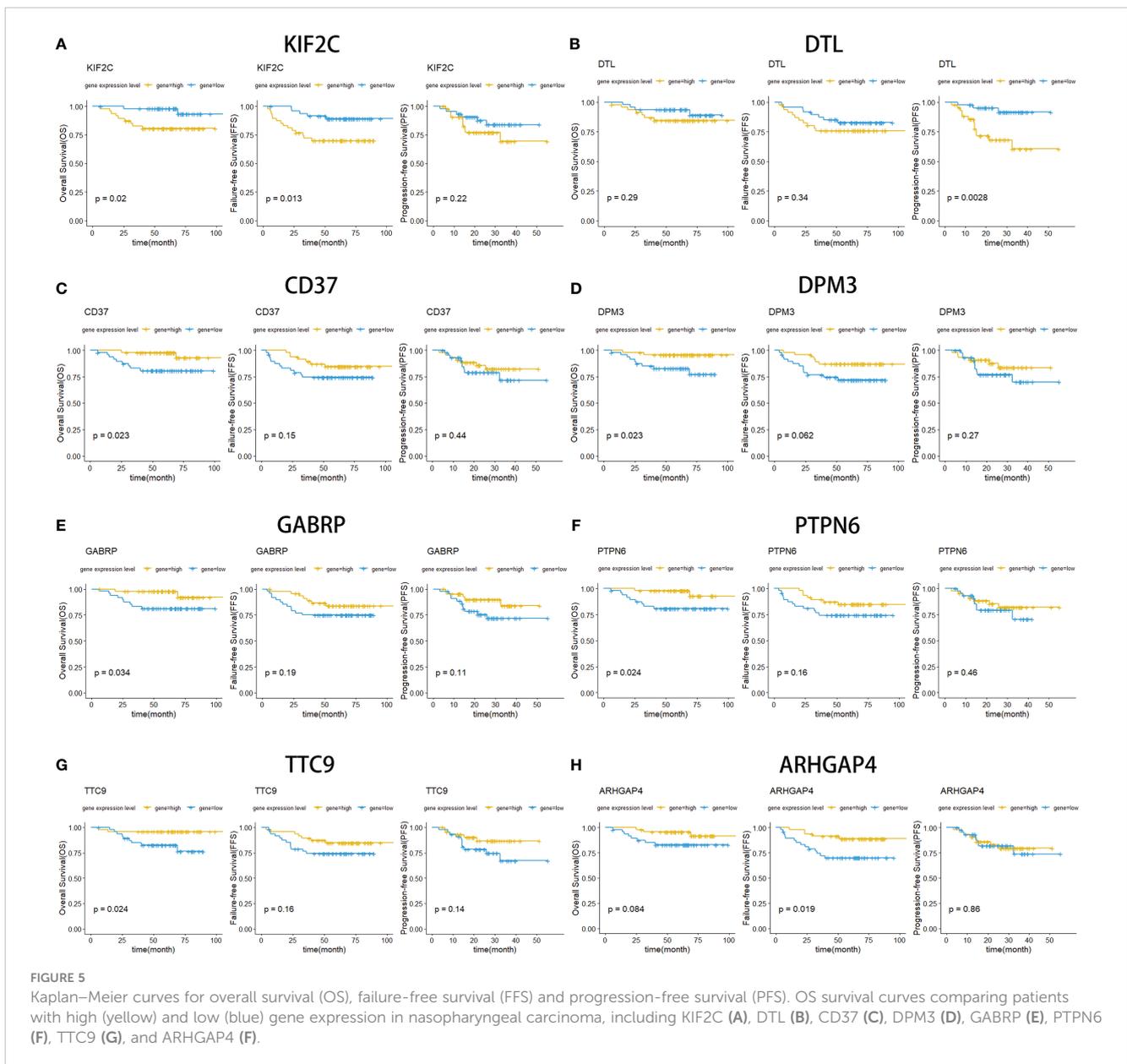


FIGURE 4
 Protein-protein interaction (PPI) network for differentially expressed genes and construction of a transcription factor (TF)-gene network. PPI network for upregulated DEGs (A). TF-gene interaction network of upregulated DEGs (B). PPI network for downregulated DEGs (C). TF-gene interaction network of downregulated DEGs (D). PPI network and TF-gene interaction network of DEGs from the survival model (E, F). The interaction between upregulated DEGs and downregulated DEGs (G). Red nodes are upregulated DEGs, blue nodes are downregulated DEGs, and purple/yellow nodes represent transcription factors (TFs). The red font represents upregulated DEGs from survival models.

staging (Supplementary Figures S3A, B). The downregulated DEGs (GABRP, DPM3, CD37 and ARHGAP4) decreased with the progression of T and N staging (Supplementary Figures S3C–F). This result suggested that the more malignant the cancer was, the higher the upregulated genes and the lower the downregulated genes were expressed. Simultaneously, we assessed the clinical information about staging and found that the upregulated DEG TOP2A expression in stages III/IV was higher than that in stages I/II (Supplementary Figure S3A), which might be critical evidence of tumor aggressiveness (43). Additionally, there was obviously

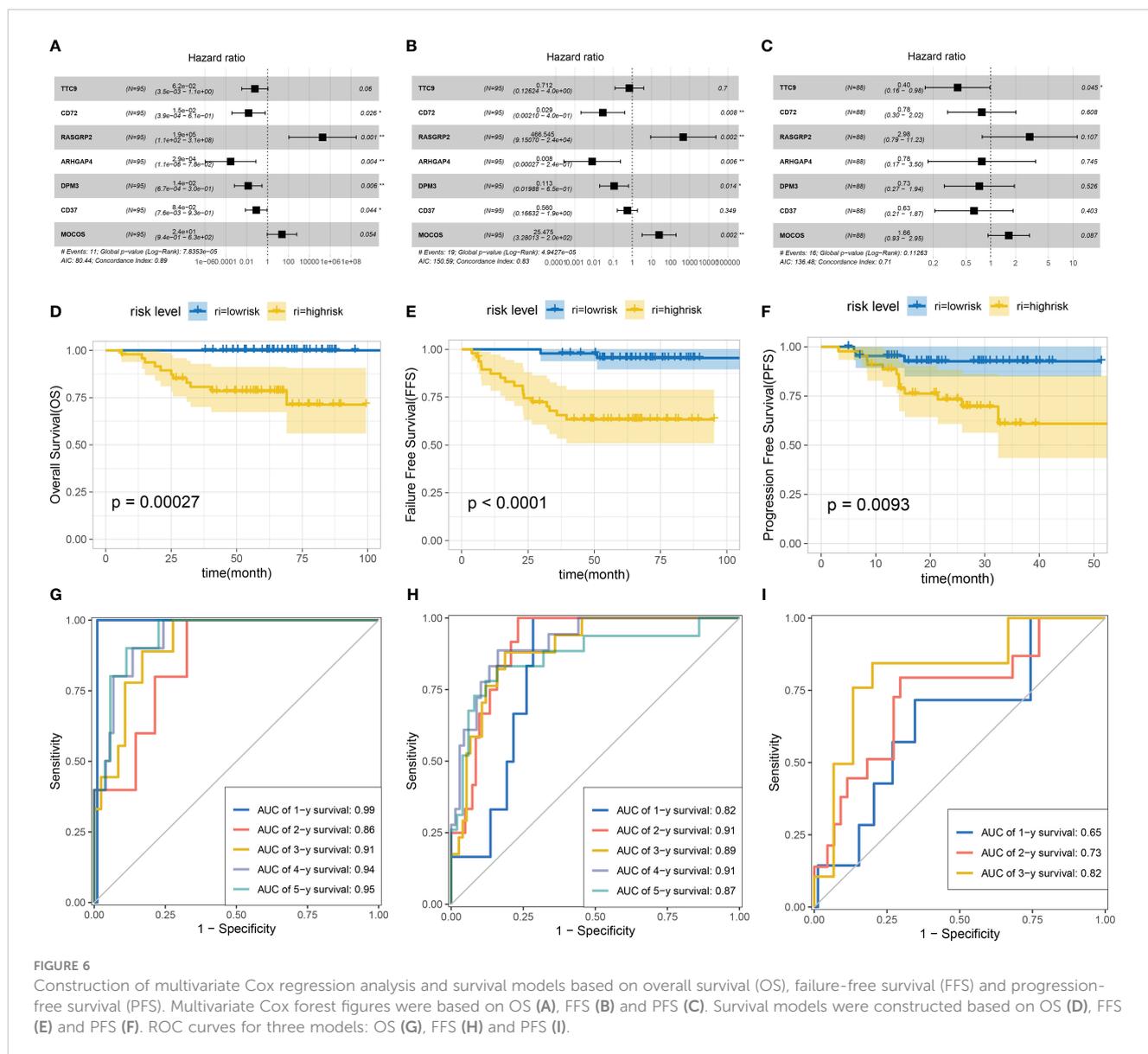
decreased CD37 expression with increasing T stage (Supplementary Figure S3E). Moreover, we detected a significant reduction in LCK in the N3 group, which might mean that low LCK expression is related to regional lymph node metastasis (Supplementary Figure S4). It is also worth mentioning that S100P, CEACAM6, MSLN and KRT23 also showed a high correlation with the progress of T and N staging (Supplementary Figure S4). Together, these results indicated that the expression levels of the screened genes were strongly associated with tumor progression.



Verification of the expression level of the above screened DEGs through qRT–PCR, WB, IHC and single cell sequencing data (scRNA-seq)

To further validate the differential expression of the key genes above, we investigated the expression levels of RASGRP2, ARHGAP4, CD37, DPM3, GAPRB, TTC9, PTPN6 and DTL in NPC specimens and control specimens. Each of these genes is associated with a poor prognosis in patients with NPC, and they are all overlapping DEGs (Supplementary Data Sheets 1, 2). At the same time, CD37, DTL, and ARHGAP4 are also hub genes. Two genes (RASGRP2 and DTL) (Figures 7A, B, Supplementary Figure S5A) were upregulated, and five genes (ARHGAP4, TTC9, CD37, PTPN6, GABRP and DPM3) (Figures 7C–H, Supplementary Figure S5A) were downregulated in NPC, which is consistent with our

previous results. Furthermore, we were able to identify a few genes that displayed significant differences in expression based on the WB results, including KIF2C, DTL, CD37, ARHGAP4, and PTPN6. DTL, KIF2C, ARHGAP4, PTPN6, and CD37 were differentially expressed, aligning with our computational and experimental findings (Supplementary Figures S6A–G). Additionally, we found that 10 up-regulated DEGs and 11 down-regulated DEGs (Table 2), grouped into five categories each, corresponded with expression levels in scRNA-seq data; LTF was the exception, showing no significant difference (Supplementary Figures S6H–K). To validate these findings, immunohistochemical staining was performed, and the results showed that KIF2C and DTL exhibited positive staining in tumor samples, which aligns with our above obtained results. Conversely, PTPN6, ARHGAP4, and CD37 demonstrated weaker positive staining in tumor samples, consistent with our previous findings (Figure 7I). The IHC results also illustrated that up-



regulated KIF2C was associated with patients' poor prognosis (Supplementary Figures S5B–D). Together, these results implied that the expression levels of key genes screened from mega-data bioinformatics analysis basically reflected the real situation in patients with NPC, which could predict the prognosis of NPC patients.

Discussion

In our study, an extensive dataset sourced from reputable databases was meticulously analyzed through bioinformatics methods to screen for differentially expressed genes (DEGs) in nasopharyngeal carcinoma (NPC). This approach allowed us to construct a unique survival model for NPC patients, opening up exciting new avenues for research. A flow chart summarizing the overall results was shown in Supplementary Figure S7. Currently, there is no overall survival information for NPC in public databases,

such as GEO and TCGA, so it has not been reported previously that using mega-data to identify the DEGs, carrying out in-depth bioinformatics analysis and then combining clinical information to analyze the survival of NPC patients. First, we selected 15 datasets from GEO and merged them into five groups, including 46 controls and 160 NPC patients, with increasing accuracy and credibility. On this basis, further bioinformatics analysis identified 2142 upregulated DEGs and 3857 downregulated DEGs. Significantly, DPM3 was differentially expressed in European populations, while CD37, ARHGAP4, and CD72 showed differential expression in Asian cohorts. Furthermore, survival analysis was performed, and survival models were constructed based on these identified DEGs, with the minority of them being reported in a previous study of NPC (44). Finally, the consistent differential expression of these DEGs in NPC samples was confirmed by qRT-PCR, WB and IHC, which confirmed the accuracy of mega-data bioinformatics analysis. All of the above findings from mega-data bioinformatics analysis provide insights into the novel prognostic biomarkers of NPC,

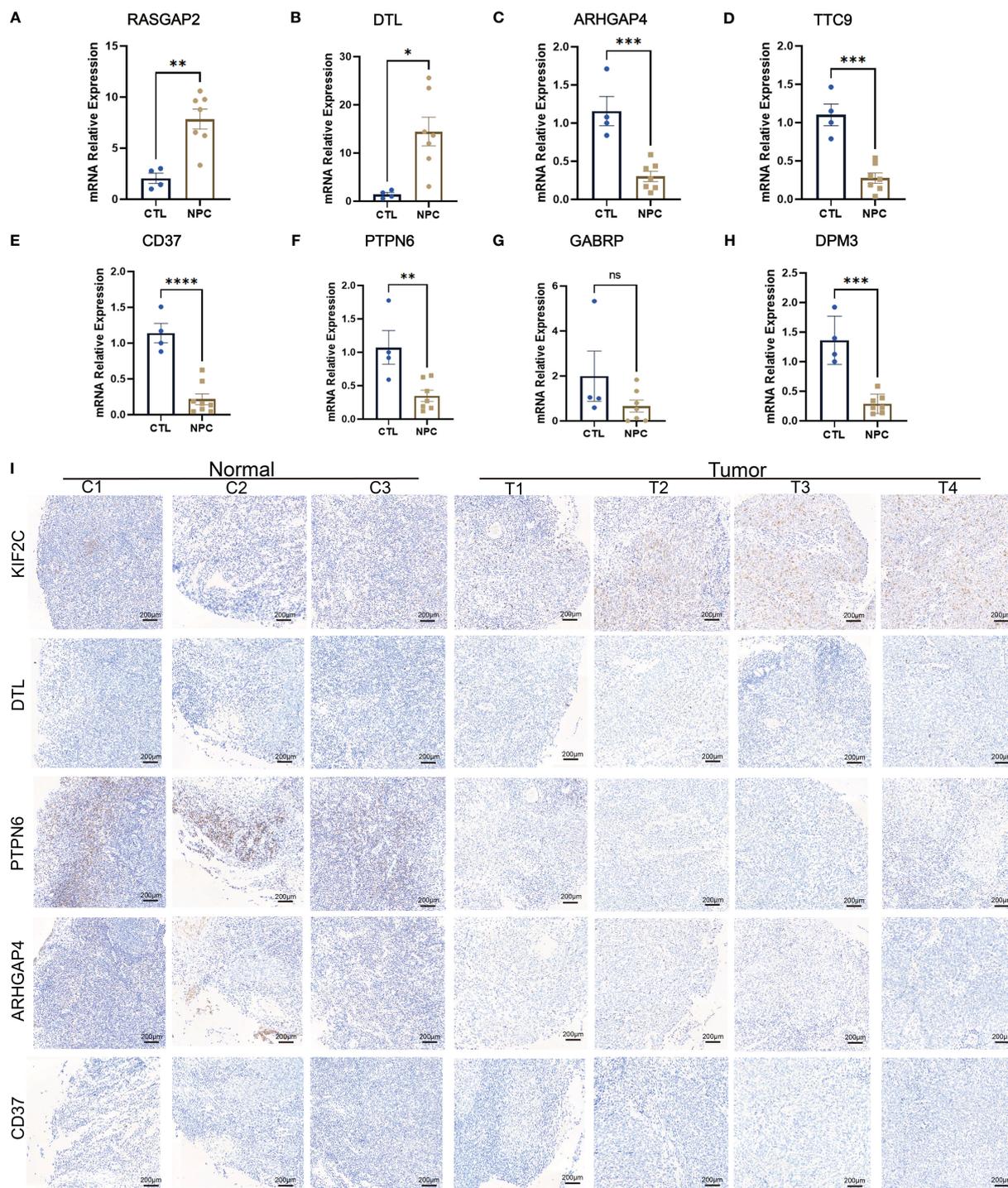


FIGURE 7
 PCR results of several identified differentially expressed genes (DEGs), including KIF2C (A), DTL (B), ARHGAP4 (C), TTC9 (D), CD37 (E), PTPN6 (F), GABRP (G), and DPM3 (H). Immunohistochemistry results of KIF2C, DTL, PTPN6, ARHGAP4 and CD37 (I). ns: $P > 0.05$; *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$.

which suggests that they may serve special targets in the treatment of NPC.

Our results are derived from the analysis of mega-data, endowing them with enhanced reliability and accuracy in contrast to studies conducted with limited sample sizes (45, 46). First, several DEGs from our screening were consistent with previous reports (11, 47). Hub genes identified by previous experiments, such as CD37

and PTPN6, were also discovered in our study (48, 49). In addition, hub genes that were used to construct models also appeared in previous studies. ARHGAP4 plays a pivotal role in governing the intricate processes of cell migration and invasion in pancreatic cancer through modulation of the HDAC2/ β -catenin signaling pathway (50). CD37 emerges as an autonomous prognostic determinant, exhibiting immunosuppressive properties within the

domain of diffuse gliomas (51). Furthermore, DPM3/prostin-1, a novel gene regulated by phospholipase C-gamma, displays a correlation with prostate tumor invasion and represents a noteworthy target for inhibiting invasive behaviors (52). Moreover, the consistent differential expression of these DEGs between the profile screen and verification of tissues was confirmed by qRT-PCR, WB and IHC. These results indicate the accuracy of the mega-data bioinformatics analysis.

In our research, more sample clinical information was collected, especially survival information, which is distinct from other studies (53, 54). Previous studies often performed survival analysis by using GSE102349 reported in 2017, which only contained progression-free survival (PFS) information for NPC patients (55–57). The survival analysis results they obtained were limited to a small number of samples, leading to significant biases in the outcomes. Moreover, owing to the absence of clinical information on NPC, some studies resorted to using data from head and neck squamous cell carcinoma (HNSCC) for survival analysis. However, since the HNSCC database hardly includes any information on NPC, which is a tumor type vastly distinct from HNSCC, the derived survival results are virtually devoid of any meaningful reference value (8–11). To overcome these limitations, we obtained NPC RNA expression profiles from GSE132112 and corresponding clinical information from Professor Jun Ma at Sun Yat-sen University. This allowed us to establish a comprehensive survival analysis by exploring the relationship between DEGs and overall survival (OS), failure-free survival (FFS), and progression-free survival (PFS), uniquely constructing survival models using NPC-specific DEGs. This allows for our subsequent analysis of clinical information and survival analysis to better reflect the real situation. Therefore, with the support of these crucial clinical data, our analysis is more capable of reflecting the reality accurately.

In our study, we must consider the inherent biases associated with bioinformatics analyses, including the potential for skewed data representation and the complexity of merging datasets from diverse sources, which may be inaccurate. Moreover, the generalizability of our prognostic model across different ethnicities and regions remains uncertain, given the variable prevalence and genetic landscape of NPC globally. Lastly, while qRT-PCR and IHC have validated the expression of certain DEGs, the absence of functional assays in our study leaves the direct impact of these genes on NPC progression and patient outcomes unverified.

To advance our understanding of the functional roles that hub genes and DEGs play in NPC, future research should prioritize targeted *in vitro* and *in vivo* studies, such as gene knockdown and overexpression experiments, to examine their effects on tumor biology and treatment response. Prospective clinical trials are warranted to validate our prognostic model, ensuring its applicability and integration into clinical workflows for improved patient management. Furthermore, concerted efforts to develop novel targeted therapies based on our identified DEGs, particularly those unexplored in NPC context, and to investigate the tumor-immune interface, will be crucial for pioneering new avenues in NPC treatment and immunotherapeutic strategies.

In summary, our study confirmed the DEGs between NPC and normal tissue that are crucial for the prognosis of NPC through bioinformatics methods from mega-data. Furthermore, our results imply that mega-data analysis would provide more accurate predictions and hints in NPC prognosis. Overall, our study not only integrated mega-data analysis and clinical information but also constructed a promising survival model specific to NPC patients by combining several key genes as novel prognostic biomarkers, which suggests that they may serve special targets in the treatment of NPC.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/>, GSE12452, GSE64634, GSE15047, GSE15903, GSE36972, GSE48501, GSE100193, GSE79571, GSE119020, GSE127848, GSE128502, GSE13597, GSE53819, GSE40290, GSE34573, GSE102349, and GSE132112 and <https://explore.data.humancellatlas.org/projects/111d272b-c25a-49ac-9b25-e062b70d66e0>.

Ethics statement

The studies involving humans were approved by West China Hospital of Sichuan University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YT: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft. JiaoZ: Formal analysis, Funding acquisition, Investigation, Supervision, Validation, Writing – review & editing. KL: Formal analysis, Investigation, Methodology, Writing – original draft. RL: Formal analysis, Methodology, Validation, Writing – original draft. JingZ: Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft. ZW: Methodology, Supervision, Validation, Writing – original draft. LL: Data curation, Resources, Supervision, Writing – original draft. JZ: Writing – original draft. XF: Writing – original draft. BD: Writing – original draft. JD: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Sichuan Agency of Science and Technology

(2022YFS0246) by JD, the China Postdoctoral Science Foundation (2021M702367) by JiaoZ, the National Natural Science Fund of China (82101658) by JiaoZ and the University Student Innovation and Entrepreneurship Fund (C2023122946) by YT.

Acknowledgments

The authors would like to thank Professor Jun Ma (Sun Yat-sen University Cancer Center, the State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University, Guangzhou, China) for kindly providing us with clinical information and related NPC RNA profile data of NPC patients. We thank Dr. Jianming Zeng (University of Macau) and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2024.1354940/full#supplementary-material>

References

- Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet (London England)*. (2019) 394:64–80. doi: 10.1016/S0140-6736(19)30956-0
- Wei KR, Zheng RS, Zhang SW, Liang ZH, Li ZM, Chen WQ. Nasopharyngeal carcinoma incidence and mortality in China, 2013. *Chin J Cancer*. (2017) 36:90. doi: 10.1186/s40880-017-0257-9
- Luo W. Nasopharyngeal carcinoma ecology theory: cancer as multidimensional spatiotemporal "unity of ecology and evolution" pathological ecosystem. *Theranostics*. (2023) 13:1607–31. doi: 10.7150/thno.82690
- Gong L, Kwong DL, Dai W, Wu P, Li S, Yan Q, et al. Comprehensive single-cell sequencing reveals the stromal dynamics and tumor-specific characteristics in the microenvironment of nasopharyngeal carcinoma. *Nat Commun*. (2021) 12:1540. doi: 10.1038/s41467-021-21795-z
- Gong L, Luo J, Zhang Y, Yang Y, Li S, Fang X, et al. Nasopharyngeal carcinoma cells promote regulatory T cell development and suppressive activity via CD70-CD27 interaction. *Nat Commun*. (2023) 14:1912. doi: 10.1038/s41467-023-37614-6
- Lv J, Wei Y, Yin JH, Chen YP, Zhou GQ, Wei C, et al. The tumor immune microenvironment of nasopharyngeal carcinoma after gemcitabine plus cisplatin treatment. *Nat Med*. (2023) 29:1424–36. doi: 10.1038/s41591-023-02369-6
- Du XJ, Wang GY, Zhu XD, Han YQ, Lei F, Shen LF, et al. Refining the 8th edition TNM classification for EBV related nasopharyngeal carcinoma. *Cancer Cell*. (2024) 42:464–473.e463. doi: 10.1016/j.ccell.2023.12.020
- Guo W, Zheng X, Hua L, Zheng X, Zhang Y, Sun B, et al. Screening and bioinformatical analysis of differentially expressed genes in nasopharyngeal carcinoma. *J Cancer*. (2021) 12:1867–83. doi: 10.7150/jca.48979
- Fang L, Shi L, Wang W, Chen Q, Rao X. Identifying key genes and small molecule compounds for nasopharyngeal carcinoma by various bioinformatical analysis. *Medicine*. (2021) 100:e27257. doi: 10.1097/MD.00000000000027257
- Liu G, Zeng X, Wu B, Zhao J, Pan Y. RNA-Seq analysis of peripheral blood mononuclear cells reveals unique transcriptional signatures associated with radiotherapy response of nasopharyngeal carcinoma and prognosis of head and neck cancer. *Cancer Biol Ther*. (2020) 21:139–46. doi: 10.1080/15384047.2019.1670521

SUPPLEMENTARY FIGURE 1

Mutation plots of differentially expressed genes (DEGs). The top 12 most mutated upregulated DEGs (A) and top 20 most mutated downregulated DEGs (B) in NPC.

SUPPLEMENTARY FIGURE 2

Kaplan–Meier curves for overall survival (OS), failure-free survival (FFS) and progression-free survival (PFS). OS survival curves comparing patients with high (yellow) and low (blue) gene expression in nasopharyngeal carcinoma, including TTN (A), FAT2 (B), KLHDC7A (C), TNXB (D), MOCOS (E), and CD72 (F).

SUPPLEMENTARY FIGURE 3

Relationship between gene expression and different clinical information. Boxplot of expression differences of DEGs grouped by staging of cancer and tumor-node-metastasis (TNM) staging system, including TOP2A (A), INSM1 (B), GABRP (C), DPM3 (D), CD37 (E), and ARHGAP4 (F).

SUPPLEMENTARY FIGURE 4

Relationship between gene expression and different clinical information. Boxplot of expression differences of DEGs grouped by staging of cancer and tumor-node-metastasis (TNM) staging system, including LCK (A), MSLN (B), S100P (C), KRT23 (D) and CEACAM6 (E).

SUPPLEMENTARY FIGURE 5

Western blotting (WB) and survival analysis based on IHC results. (A) WB results of KIF2C, DTL, ARHGAP4, CD37 and PTPN6. (B–D) Survival analysis based on IHC results for KIF2C, PTPN6 and ARHGAP4.

SUPPLEMENTARY FIGURE 6

Validation differentially expressed genes using scRNA-seq data. (A) Umap plot for nasopharyngeal carcinoma samples and control samples. Blue: nasopharyngeal carcinoma. Red: Control. (B–F) Umap plots for the expression levels of DTL, KIF2C, ARHGAP4, PTPN6 and CD37.

SUPPLEMENTARY FIGURE 7

A flow chart of the overall results.

SUPPLEMENTARY DATA SHEET 1

Upregulated differentially expressed genes (DEGs).

SUPPLEMENTARY DATA SHEET 2

Downregulated differentially expressed genes (DEGs).

SUPPLEMENTARY DATA SHEET 3

Upregulated and Downregulated differentially expressed genes with $|\log_2FC| > 1.5$ and $p < 0.05$ (DEGs).

SUPPLEMENTARY DATA SHEET 4

Cox regression analysis of differentially expressed genes (DEGs).

11. Yin DP, Zheng YF, Sun P, Yao MY, Xie LX, Dou XW, et al. The pro-tumorigenic activity of p38 γ overexpression in nasopharyngeal carcinoma. *Cell Death Dis.* (2022) 13:210. doi: 10.1038/s41419-022-04637-8
12. Wong KCW, Hui EP, Lo KW, Lam WKJ, Johnson D, Li L, et al. Nasopharyngeal carcinoma: an evolving paradigm. *Nat Rev Clin Oncol.* (2021) 18:679–95. doi: 10.1038/s41571-021-00524-x
13. Hoffmann J, Wilhelm J, Olschewski A, Kwapiszewska G. Microarray analysis in pulmonary hypertension. *Eur Respir J.* (2016) 48:229–41. doi: 10.1183/13993003.02030-2015
14. D'Angelo G, Di Rienzo T, Ojetti V. Microarray analysis in gastric cancer: a review. *World J Gastroenterol.* (2014) 20:11972–6. doi: 10.3748/wjg.v20.i34.11972
15. Barrett T, Edgar R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods enzymology.* (2006) 411:352–69. doi: 10.1016/S0076-6879(06)11019-8
16. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signaling.* (2013) 6:p11. doi: 10.1126/scisignal.2004088
17. Lei Y, Li YQ, Jiang W, Hong XH, Ge WX, Zhang Y, et al. A gene-expression predictor for efficacy of induction chemotherapy in locoregionally advanced nasopharyngeal carcinoma. *J Natl Cancer Institute.* (2021) 113:471–80. doi: 10.1093/jnci/djaa100
18. Dodd LE, Sengupta S, Chen IH, den Boon JA, Cheng YJ, Westra W, et al. Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma. *Cancer epidemiology Biomarkers prevention: Publ Am Assoc Cancer Research cosponsored by Am Soc Prev Oncol.* (2006) 15:2216–25. doi: 10.1158/1055-9965.EPI-06-0455
19. Bo H, Gong Z, Zhang W, Li X, Zeng Y, Liao Q, et al. Upregulated long non-coding RNA AFAP1-AS1 expression is associated with progression and poor prognosis of nasopharyngeal carcinoma. *Oncotarget.* (2015) 6:20404–18. doi: 10.18632/oncotarget.v6i24
20. Li J, Fan Y, Chen J, Yao KT, Huang ZX. Microarray analysis of differentially expressed genes between nasopharyngeal carcinoma cell lines 5–8F and 6–10B. *Cancer Genet cytogenetics.* (2010) 196:23–30. doi: 10.1016/j.cancergencyto.2009.08.004
21. Deng M, Zhang W, Tang H, Ye Q, Liao Q, Zhou Y, et al. Lactotransferrin acts as a tumor suppressor in nasopharyngeal carcinoma by repressing AKT through multiple mechanisms. *Oncogene.* (2013) 32:4273–83. doi: 10.1038/ncr.2012.434
22. Li XH, Qu JQ, Yi H, Zhang PF, Yi HM, Wan XX, et al. Integrated analysis of differential miRNA and mRNA expression profiles in human radioresistant and radiosensitive nasopharyngeal carcinoma cells. *PLoS One.* (2014) 9:e87767. doi: 10.1371/journal.pone.0087767
23. Han P, Chen RH, Wang F, Zeng JY, Yu ST, Xu LH, et al. Novel chimeric transcript RRM2-c2orf48 promotes metastasis in nasopharyngeal carcinoma. *Cell Death Dis.* (2017) 8:e3047. doi: 10.1038/cddis.2017.402
24. Gao W, Li ZH, Chen S, Chan JY, Yin M, Zhang MJ, et al. Epstein-Barr virus encoded microRNA BART7 regulates radiation sensitivity of nasopharyngeal carcinoma. *Oncotarget.* (2017) 8:20297–308. doi: 10.18632/oncotarget.v8i12
25. Yuan J, Chen L, Xiao J, Qi XK, Zhang J, Li X, et al. SHROOM2 inhibits tumor metastasis through RhoA-ROCK pathway-dependent and -independent mechanisms in nasopharyngeal carcinoma. *Cell Death Dis.* (2019) 10:58. doi: 10.1038/s41419-019-1325-7
26. Hsu YB, Huang CF, Lin KT, Kuo YL, Lan MC, Lan MY. Podoplanin, a potential therapeutic target for nasopharyngeal carcinoma. *BioMed Res Int.* (2019) 2019:7457013. doi: 10.1155/2019/7457013
27. Bose S, Yap LF, Fung M, Starczynski J, Saleh A, Morgan S, et al. The ATM tumour suppressor gene is down-regulated in EBV-associated nasopharyngeal carcinoma. *J Pathol.* (2009) 217:345–52. doi: 10.1002/path.2487
28. Bao YN, Cao X, Luo DH, Sun R, Peng LX, Wang L, et al. Urokinase-type plasminogen activator receptor signaling is critical in nasopharyngeal carcinoma cell growth and metastasis. *Cell Cycle (Georgetown Tex).* (2014) 13:1958–69. doi: 10.4161/cc.28921
29. Hu C, Wei W, Chen X, Woodman CB, Yao Y, Nicholls JM, et al. A global view of the oncogenic landscape in nasopharyngeal carcinoma: an integrated analysis at the genetic and expression levels. *PLoS One.* (2012) 7:e41055. doi: 10.1371/journal.pone.0041055
30. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* (2013) 41:D991–995. doi: 10.1093/nar/gks1193
31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
32. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, 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–d613. doi: 10.1093/nar/gky1131
33. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* (2019) 47:W234–w241. doi: 10.1093/nar/gkz240
34. Lin DC, Meng X, Hazawa M, Nagata Y, Varela AM, Xu L, et al. The genomic landscape of nasopharyngeal carcinoma. *Nat Genet.* (2014) 46:866–71. doi: 10.1038/ng.3006
35. Chen YP, Ismaila N, Chua MLK, Colevas AD, Haddad R, Huang SH, et al. Chemotherapy in combination with radiotherapy for definitive-intent treatment of stage II-IVA nasopharyngeal carcinoma: CSCO and ASCO guideline. *J Clin Oncol.* (2021) 39:840–59. doi: 10.1200/JCO.20.03237
36. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngology–head Neck Surg.* (2010) 143:331–6. doi: 10.1016/j.otohns.2010.05.007
37. Zhang L, MacIsaac KD, Zhou T, Huang PY, Xin C, Dobson JR, et al. Genomic analysis of nasopharyngeal carcinoma reveals TME-Based subtypes. *Mol Cancer research: MCR.* (2017) 15:1722–32. doi: 10.1158/1541-7786.MCR-17-0134
38. Lin DY. MULCOX: a computer program for the Cox regression analysis of multiple failure time variables. *Comput Methods programs biomedicine.* (1990) 32:125–35. doi: 10.1016/0169-2607(90)90092-N
39. Kajiyama Y, Tsurumaru M, Udagawa H, Tsutsumi K, Kinoshita Y, Ueno M, et al. Prognostic factors in adenocarcinoma of the gastric cardia: pathologic stage analysis and multivariate regression analysis. *J Clin Oncol.* (1997) 15:2015–21. doi: 10.1200/JCO.1997.15.5.2015
40. Ziegler CGK, Miao VN, Owings AH, Navia AW, Tang Y, Bromley JD, et al. Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19. *Cell.* (2021) 184:4713–33.e4722. doi: 10.1016/j.cell.2021.07.023
41. Liu Y, He S, Wang XL, Peng W, Chen QY, Chi DM, et al. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat Commun.* (2021) 12:741. doi: 10.1038/s41467-021-21043-4
42. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* (2019) 16:1289–96. doi: 10.1038/s41592-019-0619-0
43. Sang J, Li X, Lv L, Zhang C, Zhang X, Li G. Circ-TOP2A acts as a ceRNA for miR-346 and contributes to glioma progression via the modulation of sushi domain-containing 2. *Mol Med Rep.* (2021) 23(4):255. doi: 10.3892/mmr
44. Lu X, Chen X, Wang X, Qing J, Li J, Pan Y. Construction of lncRNA and mRNA co-expression network associated with nasopharyngeal carcinoma progression. *Front Oncol.* (2022) 12:965088. doi: 10.3389/fonc.2022.965088
45. Ye Z, Wang F, Yan F, Wang L, Li B, Liu T, et al. Bioinformatic identification of candidate biomarkers and related transcription factors in nasopharyngeal carcinoma. *World J Surg Oncol.* (2019) 17:60. doi: 10.1186/s12957-019-1605-9
46. Zhu HM, Fei Q, Qian LX, Liu BL, He X, Yin L. Identification of key pathways and genes in nasopharyngeal carcinoma using bioinformatics analysis. *Oncol Lett.* (2019) 17:4683–94. doi: 10.3892/ol
47. Chen C, Wu B, Wang M, Chen J, Huang Z, Shi JS. GABRP promotes CD44s-mediated gemcitabine resistance in pancreatic cancer. *PeerJ.* (2022) 10:e12728. doi: 10.7717/peerj.12728
48. Liu K, Kang M, Zhou Z, Qin W, Wang R. Bioinformatics analysis identifies hub genes and pathways in nasopharyngeal carcinoma. *Oncol Lett.* (2019) 18:3637–45. doi: 10.3892/ol
49. Wang S, Jin F, Fan W, Liu F, Zou Y, Hu X, et al. Gene expression meta-analysis in diffuse low-grade glioma and the corresponding histological subtypes. *Sci Rep.* (2017) 7:11741. doi: 10.1038/s41598-017-12087-y
50. Shen Y, Xu L, Ning Z, Liu L, Lin J, Chen H, et al. ARHGAP4 regulates the cell migration and invasion of pancreatic cancer by the HDAC2/ β -catenin signaling pathway. *Carcinogenesis.* (2019) 40:1405–14. doi: 10.1093/carcin/bgz067
51. Yan X, Zhou Q, Zhu H, Liu W, Xu H, Yin W, et al. The clinical features, prognostic significance, and immune heterogeneity of CD37 in diffuse gliomas. *iScience.* (2021) 24:103249. doi: 10.1016/j.isci.2021.103249
52. Manos EJ, Kim ML, Kassir J, Chang PY, Wells A, Jones DA. Dolichol-phosphate-mannose-3 (DPM3)/prost-in-1 is a novel phospholipase C-gamma regulated gene negatively associated with prostate tumor invasion. *Oncogene.* (2001) 20:2781–90. doi: 10.1038/sj.onc.1204379
53. Hu R, Xu X, Mo L, Chen M, Liu Y. Bioinformatics analysis identifies potential biomarkers involved in the metastasis of locoregionally advanced nasopharyngeal carcinoma. *Medicine.* (2022) 101:e30126. doi: 10.1097/MD.00000000000030126
54. Ge Y, He Z, Xiang Y, Wang D, Yang Y, Qiu J, et al. The identification of key genes in nasopharyngeal carcinoma by bioinformatics analysis of high-throughput data. *Mol Biol Rep.* (2019) 46:2829–40. doi: 10.1007/s11033-019-04729-3
55. Dai Y, Chen W, Huang J, Xie L, Lin J, Chen Q, et al. Identification of key pathways and genes in nasopharyngeal carcinoma based on WGCNA. *Auris nasus larynx.* (2023) 50:126–33. doi: 10.1016/j.anl.2022.05.013
56. Huang D, Liu Q, Zhang W, Huang C, Zheng R, Xie G, et al. Identified IGSF9 association with prognosis and hypoxia in nasopharyngeal carcinoma by bioinformatics analysis. *Cancer Cell Int.* (2020) 20:498. doi: 10.1186/s12935-020-01587-z
57. Chen X, Ding Q, Lin T, Sun Y, Huang Z, Li Y, et al. An immune-related prognostic model predicts neoplasm-immunity interactions for metastatic nasopharyngeal carcinoma. *Front Immunol.* (2023) 14:1109503. doi: 10.3389/fimmu.2023.1109503