

## Identification of four key biomarkers and small molecule drugs in nasopharyngeal carcinoma by weighted gene co-expression network analysis

Xi Pan <sup>a</sup> and Jian-Hao Liu <sup>b</sup>

<sup>a</sup>Department of Oncology, Xiangya Third Hospital, Central South University, Changsha, China; <sup>b</sup>School of Pharmaceutical Sciences of Central South University, Changsha, 410078, China

### ABSTRACT

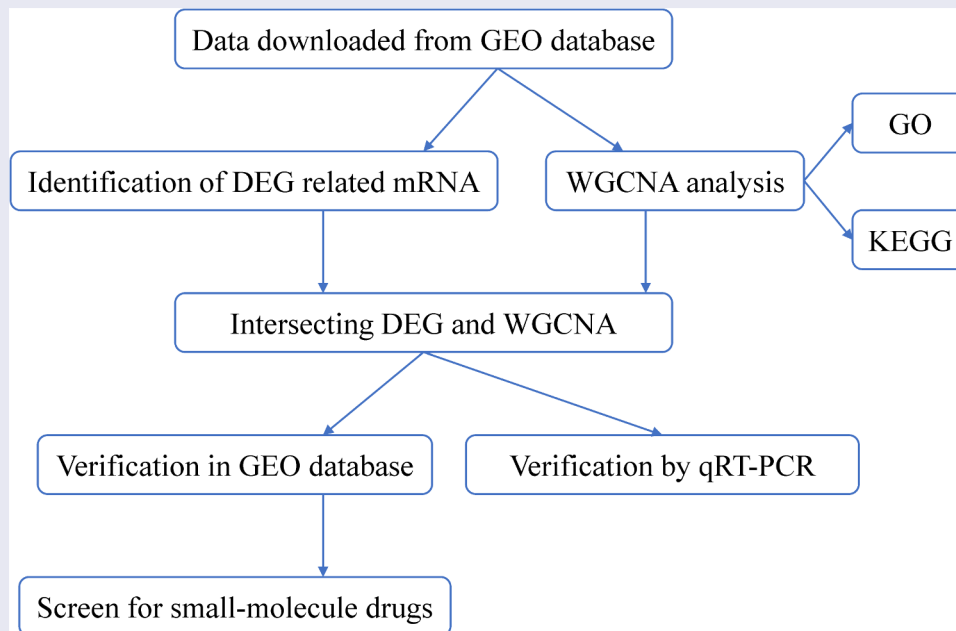
Nasopharyngeal carcinoma (NPC) is a heterogeneous carcinoma whose underlying molecular mechanisms involved in tumor initiation, progression, and migration are largely unclear. The aim of the present study was to identify key biomarkers and small-molecule drugs for screening, diagnosing, and treating NPC via gene expression profile analysis. Raw microarray data was used to identify 430 differentially expressed genes (DEGs) in the Gene Expression Omnibus (GEO) database. The key modules associated with histological grade and tumor stage were identified using weighted gene co-expression network analysis. qRT-PCR was used to verify the differential expression of hub genes. Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and the connectivity map database were used to identify potential mechanisms and screen small-molecule drugs targeting hub genes. Functional enrichment analysis showed that genes in the green module were enriched in the regulation of cell cycle, p53 signaling pathway, and cell part morphogenesis. Four DEG-related hub genes (*CRIP1*, *KITLG*, *MARK1*, and *PGAP1*) in the green module, which were considered potential diagnostic biomarkers, were taken as the final hub genes. The expression levels of these four hub genes were verified via qRT-PCR, and the results were consistent with findings from the GEO analysis. Screening was also conducted to identify small-molecule drugs with potential therapeutic effects against NPC. In conclusion, four potential prognostic biomarkers and several candidate small-molecule drugs, which may provide new insights for NPC therapy, were identified.

### ARTICLE HISTORY

Received 15 March 2021  
Revised 20 May 2021  
Accepted 21 May 2021

### KEYWORDS

Nasopharyngeal carcinoma;  
hub genes; WGCNA;  
signature; small-molecule  
drugs



## Introduction

Nasopharyngeal carcinoma (NPC) is a malignant carcinoma arising from the nasopharyngeal epithelial lining of the nasopharynx. It has a unique geographical distribution and racial prevalence. Based on GLOBOCAN cancer estimates and the International Agency for Research on Cancer, there were approximately 129,079 newly diagnosed NPC cases and 72,987 deaths in 2018, accounting for approximately 0.7% of cancer cases and deaths in that year [1,2]. The geographical distribution of global NPC incidence is extremely unbalanced; more than 80% of newly diagnosed cases are in Asia, while fewer than 0.2% of the cases occur in Oceania, resulting in an age-standardized rate of 10.0 per 100,000 persons in Southeast Asia and 0.5 per 100,000 persons in Central America [1,2]. Despite substantial improvements in NPC treatment options over the past decade, including chemotherapy, immunotherapy, and radiotherapy, the prognosis for patients with NPC remains low; approximately 20% of cases recur and the 5-year survival rate is only 50%–70% [3]. Thus, novel biomarkers and therapeutic strategies are urgently required for effective diagnosis and improved treatment for NPC patients. The remarkable geographical distribution and racial prevalence of NPC incidence has spurred studies on its risk factors, and it is thought that several etiological factors, including genetic and ethnic factors, Epstein-Barr virus (EBV) infection, and environmental factors (e.g. cigarette smoking and exposure to dust and formaldehyde) contribute to the development and progression of NPC [4–8]. EBV infection is perhaps the most extensively studied etiological factor in NPC pathogenesis. The study involves in situ hybridization of EBV-encoded RNA, which is found in all tumor cells but not in the normal nasopharyngeal epithelial lining of the nasopharynx, suggesting that EBV activation may play a vital role in the pathogenesis of NPC [9–11]. There is growing evidence that EBV is associated with multiple human cancers, including NPC [7–10]. Apart from EBV infection, genetic susceptibility is perhaps the most common cause of NPC. Studies have identified genetic susceptibility genes involved in DNA repair, and metabolic and

immune response pathways are associated with the development of NPC [12–14]. Recent linkage and association studies have described a potential association between genetic susceptibility genes and the development of NPC, including *MST1R*, *PTPRG*, *HLA*, and *MMP2* [12–17]. The rapid developments in high-throughput whole-genome sequencing and bioinformatic analyses have played a significant role in the screening for candidate biomarkers of several diseases, including NPC.

In this study, we sought to obtain relevant data from the Gene Expression Omnibus (GEO) database and undertook a weighted gene co-expression network analysis (WGCNA) to identify key modules. We then explored the underlying mechanisms of hub module genes using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Based on the connectivity map database, four potential small-molecule drugs targeting the hub genes were identified. Two NPC cohorts from the GEO database and experimental data from our study were used to verify the expression levels of these four hub genes. The aim of our study was to identify key biomarkers and validate the expression levels of four hub genes, and to further screen for small-molecule drugs of these hub genes.

## Materials and methods

### Microarray datasets

Five mRNA microarray datasets (GSE12452, GSE53819, GSE13597, GSE40290, and GSE64634) (Table 1) were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), a public functional genomics data repository that includes array- and sequence-based gene profile data [18].

**Table 1.** Primers used for the qRT-PCR.

MARK1	F: 5'- GAGCGGGACACGGAAAATCAT -3' R: 5'- TGCTACTCGACTTGGTAGGCT -3'
PGAP1	F: 5'- CTTCCGGCTTCGAGGAGAATAAG -3' R: 5'- GGGATAGCGTTTTGCCAGTTT -3'
KITLG	F: 5'- AATCCTCTCGTCAAACTGAAGG -3' R: 5'- CCATCTCGCTTATCCAACAATGA -3'
CRIP1	F: 5'- TAATGACGGCCCGGTCTTTTA -3' R: 5'- TGCAGCGTGCTGGGTTTAAT -3'
GAPDH	F: 5'- TGTGGGCATCAATGGATTGG -3' R: 5'- ACACCATGTATCCGGGTCAAT -3'

### **Data processing and identification of differentially expressed genes (DEGs)**

The robust multiarray averaging (RMA) method was used for data pre-processing, including background adjustment, quantile normalization, and log<sub>2</sub> transformation of expression matrices. The corresponding platforms were used to annotate each probe based on the Entrez ID. For genes with several probes, the medians of all probes were selected. The 'limma' package in the R statistical software (version 3.6) was used to identify differentially expressed genes associated with NPC, with  $|\log_{2}FC| > 1$  and  $P$ -value  $< 0.05$  set as the screening thresholds.

### **Construction of weighted gene co-expression networks**

The median absolute deviation of each gene expression value was calculated and genes with median absolute deviation in the top 25% were chosen for WGCNA. The 'WGCNA' package in R was used to construct weighted gene co-expression networks [19]. First, the goodSamplesGenes function was used to detect good samples and good genes. Second, the pickSoftThreshold function was used to select an appropriate soft-thresholding power. Then, the co-expression similarity was raised to achieve a scale-free topology based on the appropriate soft-threshold power. Finally, co-expression gene modules were generated using the cutreeDynamic function. Additionally, we calculated gene significance (GS), module eigengene (ME), and module membership (MM) to detect the association between clinical parameters of either the genes or modules;  $P < 0.05$  was considered statistically significant.

### **Functional enrichment analysis of hub module genes**

To better understand the potential molecular mechanisms by which hub module genes impact correlative clinical parameters, we uploaded all hub module genes to the Metascape online tool (<https://metascape.org/gp/index.html#/main/step1>) [20] where Gene ontology (GO) and Kyoto

Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed.  $P < 0.01$  was considered statistically significant.

### **Selection and validation of hub genes**

Genes with the highest intra-modular connectivity were selected as candidate hub genes. Genes with biological significance usually have higher absolute GS values. In this study, candidate hub genes were screened based on the criteria: absolute GS value  $> 0.20$  and MM  $> 0.80$ . The final hub genes were identified from the intersection of NPC-related DEGs with candidate hub genes. Expression levels of hub genes in patients with NPC were assessed using box plots. The diagnostic values of hub genes were evaluated using the area under the receiver operating characteristic (ROC) curve. Differential expression and diagnostic values of hub genes were validated in NPC patients using a separate external dataset (GSE53819). Three microarray datasets (GSE13597, GSE40290, and GSE64634) were then combined to further verify the expression levels of hub genes.

### **Identification of candidate small-molecule drugs**

The connectivity map (CMAP) is an online database (<https://portals.broadinstitute.org/cmap>) with more than 7000 gene expression profiles of 1309 compounds that is used to predict small-molecule drugs for specific diseases [21]. To identify small-molecule target drugs for treating NPC, the hub genes were uploaded to the CMAP database's 'query' page. Correlation between the small-molecule drugs and hub genes was evaluated using enrichment scores ranging from  $-1$  to  $1$ . The chemical and molecular structures of the small-molecule drugs were downloaded from the PubChem Compound database (<https://pubchem.ncbi.nlm.nih.gov/>).

### **Tissue collection**

NPC specimens and adjacent nasopharyngeal epithelial tissues were collected from 20 patients, immediately transferred into liquid nitrogen and then preserved at  $-80^{\circ}\text{C}$ . This study was approved by the Ethics Committee of the Xiangya Third

Hospital, Central South University and was conducted in accordance with the Declaration of Helsinki.

### Cell culture

The human immortalized nasopharyngeal cell line, NP69, was purchased from the American Tissue Culture Collection (ATCC; Manassas, VA, USA) and human nasopharyngeal carcinoma cell lines (CNE1, HNE3, 5–8 F, and HK-1) were purchased from the cell bank of the Chinese Academy of Sciences in Shanghai (Shanghai, China). The NP69 cell line was cultured in keratinocyte serum-free medium (KFSM; Gibco, Grand Island, NY, USA) containing 10%–15% fetal bovine serum (FBS; Gibco), 100 U/ml penicillin, and 100 µg/mL streptomycin and incubated in a saturated humidity incubator set at 37°C and 5% CO<sup>2</sup>. CNE1, HNE3, 5–8 F, and HK-1 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Gibco, Grand Island, NY, USA) containing 10%–15% FBS, 100 U/ml penicillin, and 100 µg/mL streptomycin and incubated in a saturated humidity incubator set at 37°C and 5% CO<sup>2</sup>.

### RNA extraction and quantitative reverse transcription PCR (qRT-PCR)

RNA was extracted from the cells using TRIzol reagent and reverse transcribed into cDNA using the following reaction mixture: 2 µL cDNA, 10 µL SYBR® Premix Ex Taq<sup>TM</sup> 11 (Takara Bio, Tokyo, Japan), 0.4 µL forward and reverse primers, and 6.4 µL ddH<sub>2</sub>O. Each sample was run in triplicate and GAPDH was used as the internal control. The reaction conditions used were: pre-denaturation at 95°C for 10 min followed by 40 cycles of denaturation at 95°C for 15 s, annealing at 60°C for 30 s, and extension at 70°C for 30 s. The data was

analyzed using the  $\Delta\Delta$ CT method and the average of the three replicate experiments for each sample was taken. The primers used are shown in Table 2.

### Statistical analysis

Two-tailed Student's t-test was used to identify differentially expressed genes associated with NPC. Statistical analyses were conducted in R (version 4.0.0), MedCalc (version 19.1) or SPSS (version 24.0) software. Genes with  $|\log_{2}FC| > 1$  and  $P < 0.05$  were considered statistically significant DEGs.

### Results

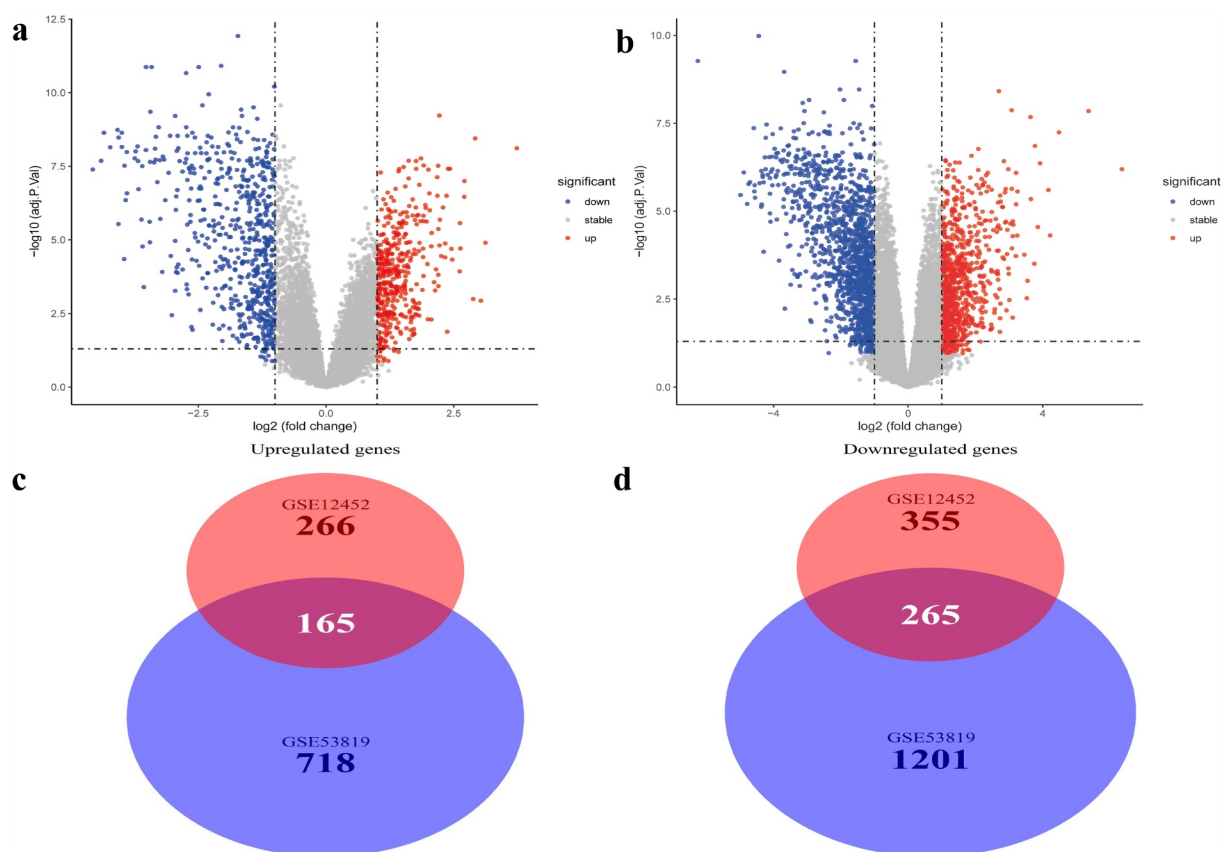
This study used WGCNA to identify key gene modules associated with NPC histological grade and tumor stage. Genes with the highest intra-modular connectivity were selected as candidates. These hub genes were identified from the intersection of NPC-related DEGs with candidate hub genes. GO and KEGG analyses were performed to explore the mechanisms underlying the hub module genes. The connectivity map database was used to screen for potential small-molecule drug targets of the hub genes. Finally, two NPC cohorts from the GEO database were used to verify the expression levels of the four hub genes.

### Identification of DEGs

Two microarray datasets, GSE12452 and GSE53819, from the GEO database, were used to identify DEGs between normal samples and NPC samples. DEGs were identified based on the thresholds  $|\log_{2}FC| > 1$  and  $P < 0.05$ . Volcano plots of DEGs in the two separate datasets are shown in Figure 1a and Figure 1b. A total of 430 common genes were identified at the intersection

**Table 2.** Characteristics of datasets included in the integrated analysis.

Year	Source	First author	Country	Assay type	Platform	PMID
2008	GSE12452	Ahlquist P	USA	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	22,880,099
2014	GSE53819	Qian CN	China	Agilent-014850 Whole Human Genome Microarray 4x44K	GPL6480	24,763,226
2009	GSE13597	Yap L	United Kingdom	Affymetrix Human Genome U133A Array	GPL96	19,142,888
2016	GSE40290	Huang ZX	China	Capitalbio 22 K Human oligo array version 1.0	GPL8380	NA
2017	GSE64634	Xiong W	China	Affymetrix Human Genome U133 Plus 2.0 Array	GPL570	26,246,469



**Figure 1.** Differentially expressed genes and common differentially expressed genes in two datasets. (a and b) The volcano plots visualize the differentially expressed genes in GSE12452 and GSE53819. (c and d) Common differentially expressed genes in two datasets.

of DEGs in the two datasets for further WGCNA. Of the 430 genes, 165 were significantly upregulated while 265 were significantly downregulated in NPC patients (Figure 1c and Figure 1d).

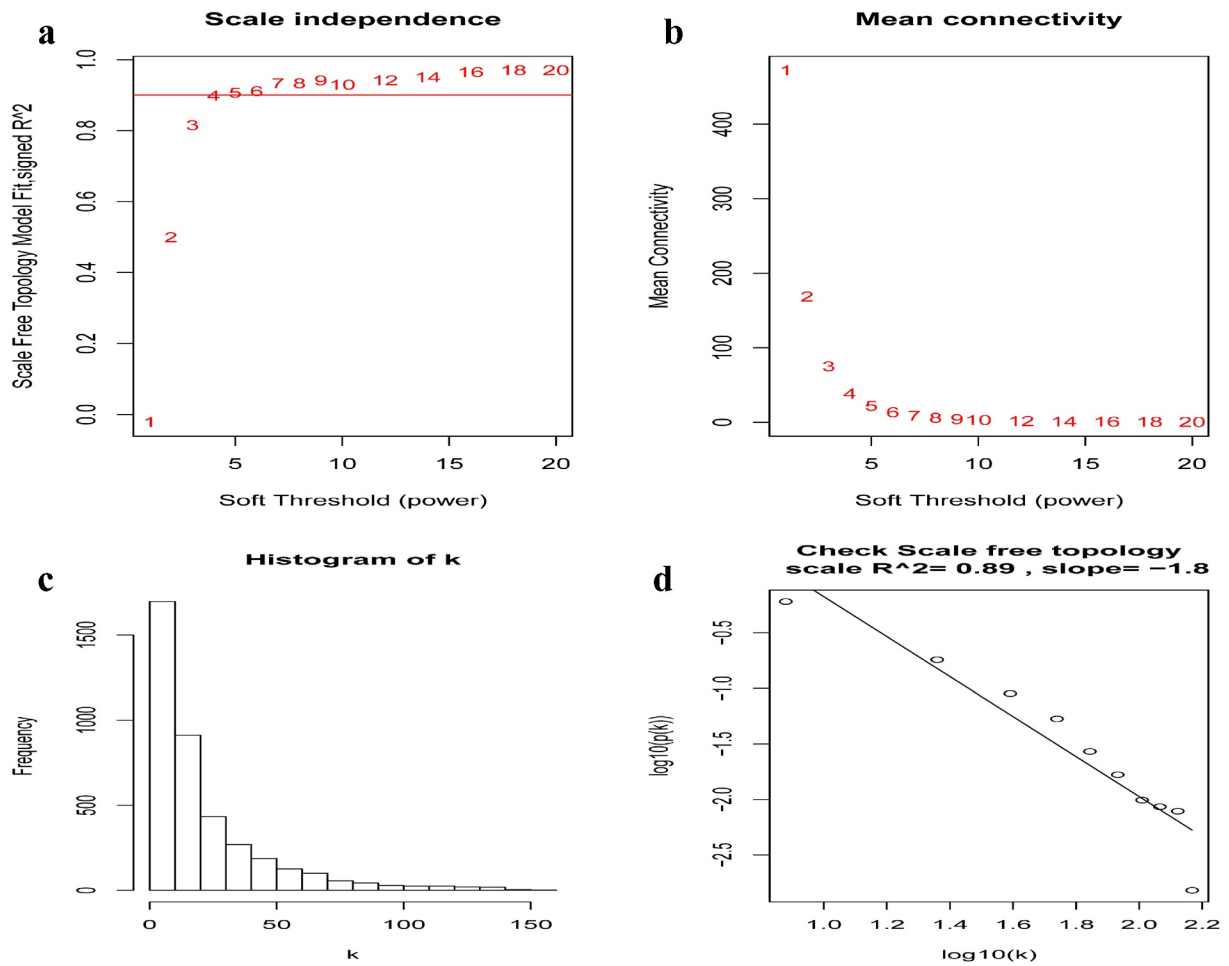
### **Construction of weighted gene co-expression network and identification of hub modules**

A total of 3949 genes with median absolute deviations in the top 25% in 31 NPC samples and 10 normal samples in GSE12452 dataset were selected. The samples were clustered to remove outlier samples. WGCNA was then used to identify genes that were highly co-expressed across the groups. To ensure a scale-free network, the power of  $\beta = 5$  (scale-free  $R^2 = 0.89$ , slope =  $-1.80$ ) was selected as the soft-thresholding, as shown in Figure 2. Then, similar modules were merged based on the height of ME in the clustering equal to 0.25 (Figure 3a). 12 co-expressed gene modules were identified.

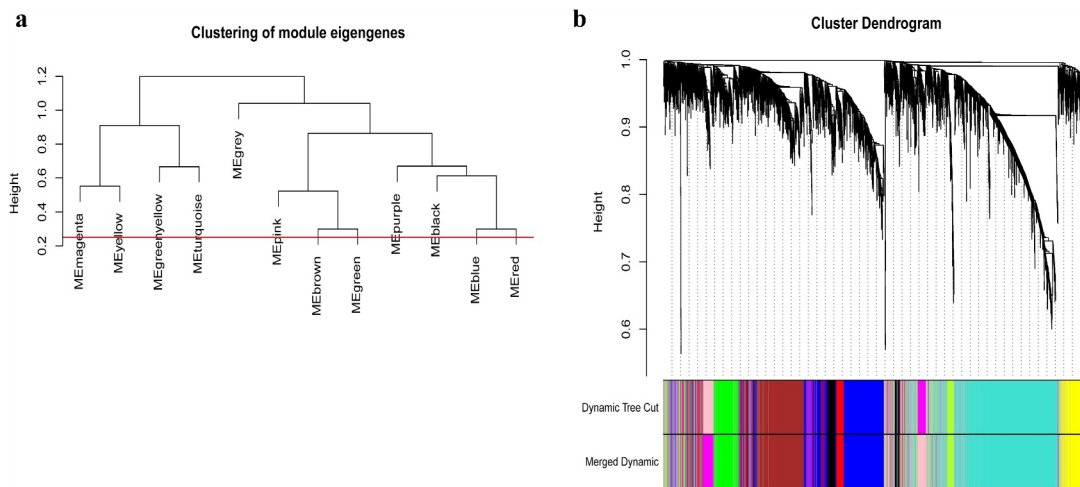
The ‘grey’ module, which contained 583 genes not attributed to any modules (Figure 3b), was removed in the subsequent analysis. The relationships among the 11 modules were analyzed using clinical parameters (including histological grade and tumor stage). The green module was significantly associated with NPC tumor stage (N stage) and was chosen as the clinically significant module for subsequent analyses (Figure 4).

### **Functional enrichment analysis of hub module genes**

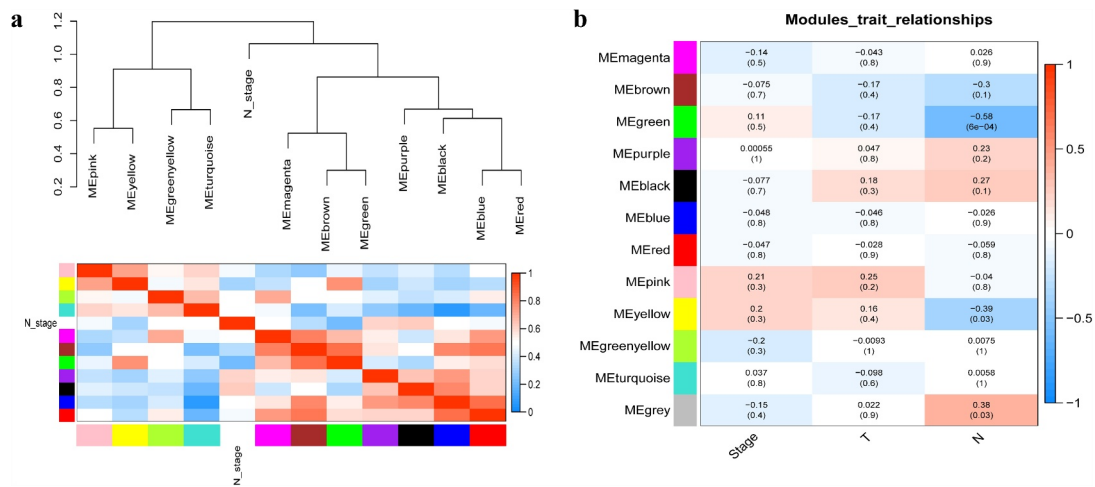
GO and KEGG pathway enrichment analysis of the above green module was carried out to mine potential biological functions associated with NPC using the Metascape online tool.  $P < 0.01$ , minimum overlap genes = 3, and an enrichment factor  $> 1.50$  were set as the cutoff criteria. GO analysis showed that genes in the green module were mainly enriched in plasma membrane-bound cell



**Figure 2.** Determination of soft-thresholding power in the weighted gene co-expression network analysis (WGCNA). (a) Analysis of the scale-free fit index for various soft-thresholding powers ( $\beta$ ). The red line indicates where the correlation coefficient is 0.9, and the corresponding soft-thresholding power is 5. (b) Analysis of the mean connectivity for various soft-thresholding powers. The red line indicates where the correlation coefficient is 0.9, and the corresponding soft-thresholding power is 5. (c) Histogram of connectivity distribution when  $\beta = 5$ . (d) Checking the scale-free topology when  $\beta = 5$ .



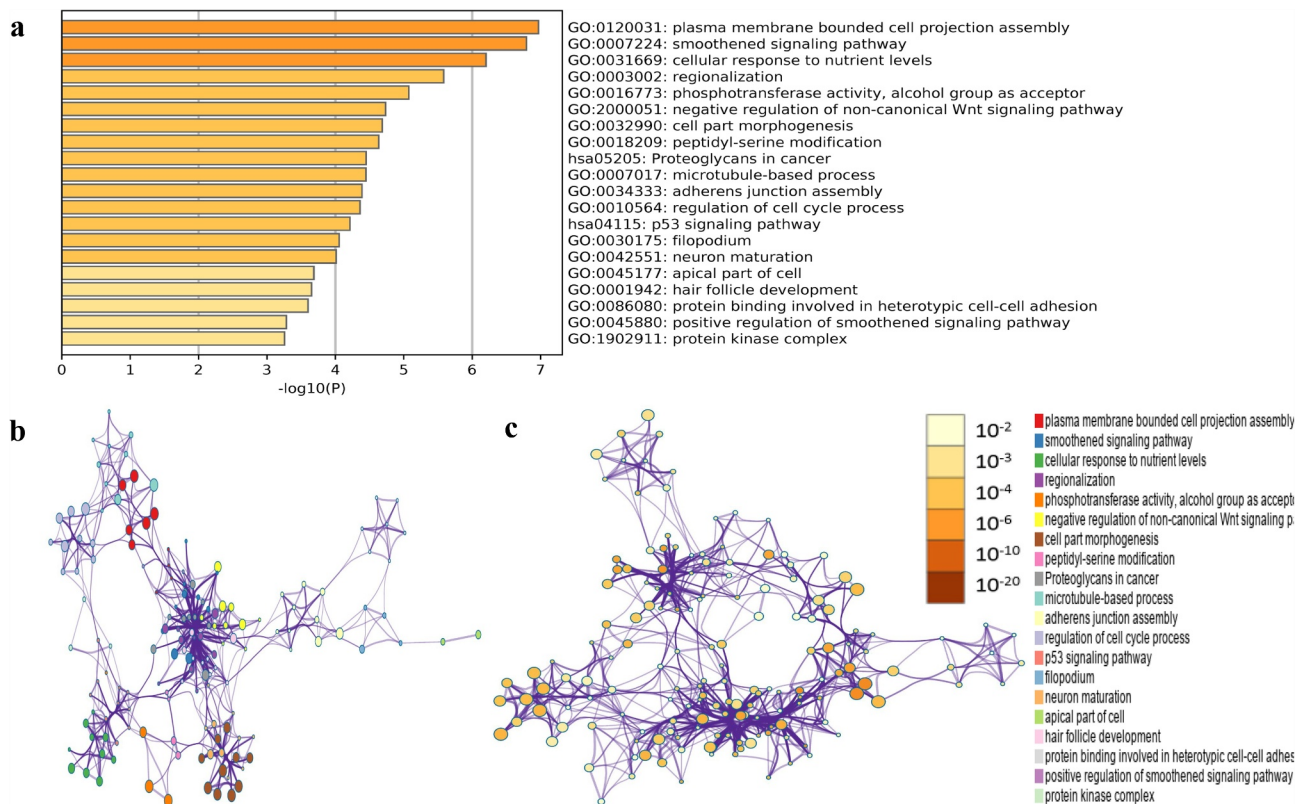
**Figure 3.** Construction of WGCNA modules. (a) The cluster dendrogram of module eigengenes. (b) The cluster dendrogram of the genes with median absolute deviation in the top 25% in the GSE12452. Each branch in the figure represents one gene, and every color below represents one co-expression module.



**Figure 4.** Relationship between modules and clinical traits. (a) Module eigengene dendrogram and eigengene network heatmap summarize the modules yielded in the clustering analysis. (b) Heatmap of the module-trait relationships. The green module was significantly associated with N stage.

projection assembly, smoothed signaling pathway, cellular response to nutrient levels, regulation of cell cycle process, neuro maturation, and protein kinase complex (Figure 5). Two significantly

enriched pathways, the p53 signaling pathway and proteoglycans in cancer, were associated with genes in the green module.

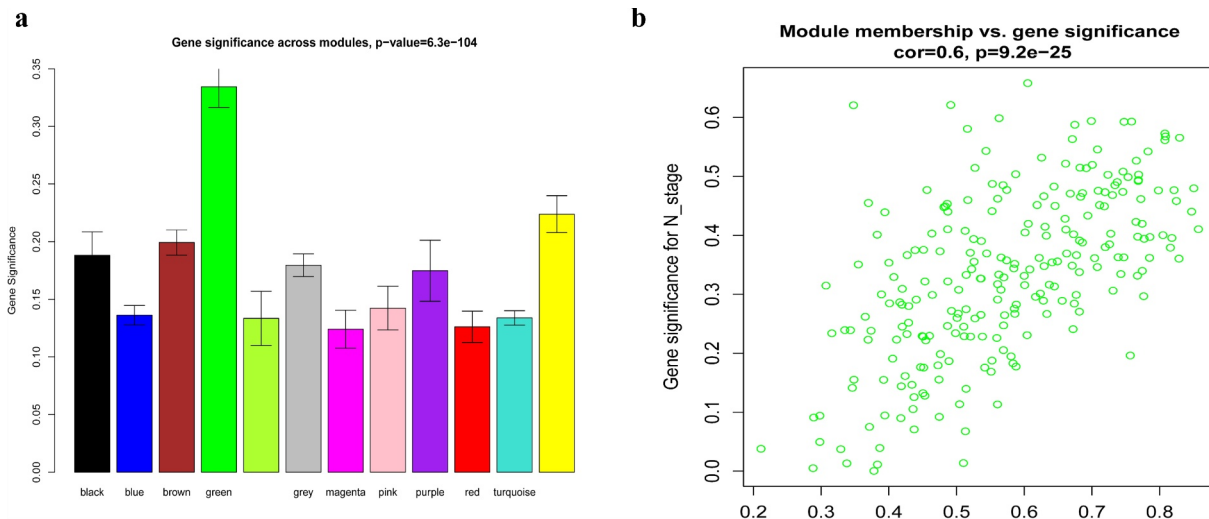


**Figure 5.** Functional enrichment analysis of hub module genes. (a) GO terms and KEGG pathway were presented, and each band represents one enriched term or pathway colored according to the  $-\log_{10}(p)$ . (b) Network of the enriched terms and pathways. Nodes represent enriched terms or pathway with node size indicating the number of hub module genes involved in. Nodes sharing the same cluster are typically close to each other, and the thicker the edge displayed.

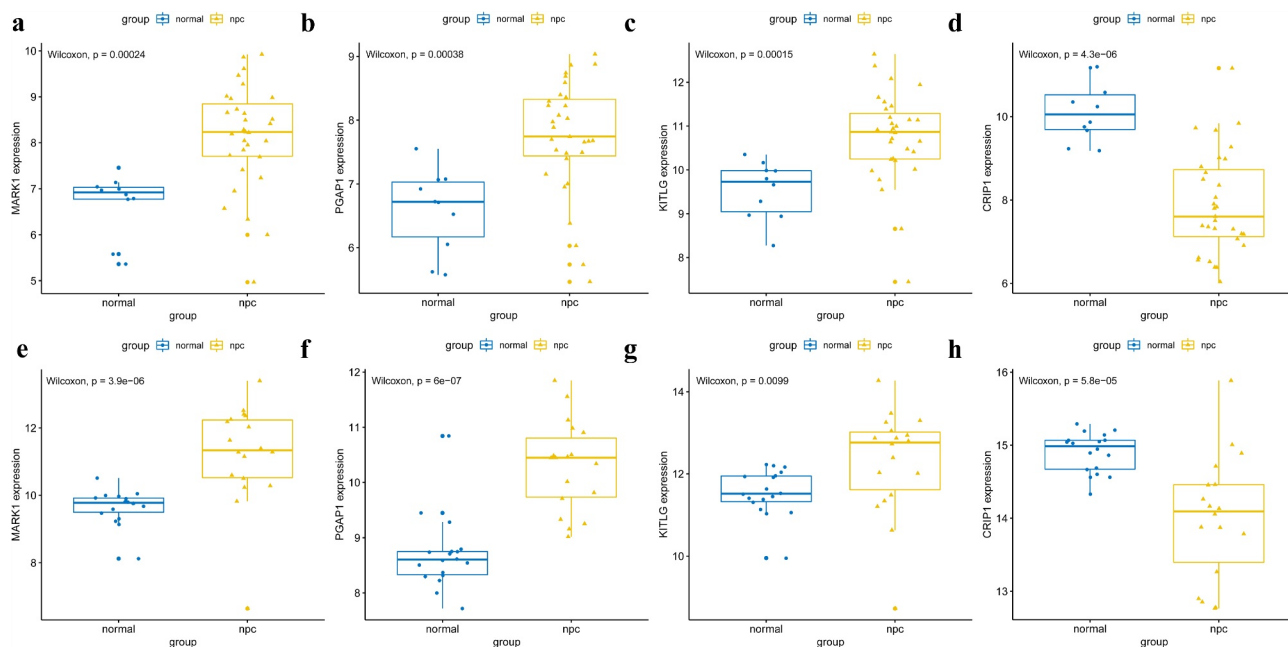
### Identification and validation of hub genes

A total of 14 genes with the greatest connectivity in the green module were identified as candidate hub genes based on  $GS > 0.2$  and  $MM > 0.8$  (Figure 6). The final four hub genes: cysteine-rich intestinal protein 1 (*CRIP1*), KIT ligand (*KITLG*), microtubule affinity regulating kinase 1 (*MARK1*), and post-GPI attachment to protein 1 (*PGAP1*), were identified from the intersection of NPC-related DEGs with candidate hub

genes. Data in the GSE12452 and GSE53819 datasets were compared with that of normal tissues, and three of the final four hub genes: *PGAP1*, *MARK1*, and *KITLG* showed significantly higher expression in NPC tissues, while the expression of *CRIP1* was significantly downregulated (Figure 7). To verify the diagnostic potential of these four hub genes, ROC analysis was performed on the GSE12452 and GSE53819 datasets. The results of the ROC analysis



**Figure 6.** Identification of hub module and candidate hub gene. (a) Distribution of average gene significance and errors in the modules associated with the NPC. (b) Scatter plot for correlation between gene module membership in the green module and gene significance.



**Figure 7.** Validation of hub genes in the gene expression level. (a-d) Validation of hub genes in the GSE12452. *PGAP1*, *MARK1*, and *KITLG* were significantly higher expression in NPC compared with normal tissues, while *CRIP1* was significantly lower expression in NPC compared with normal tissues. (e-h) Validation of hub genes in the GSE53819 and the results were the same as earlier.



demonstrated that these four genes had excellent diagnostic efficiencies in NPC patients (Figure 8). The expression levels of the four hub genes were further validated in the combined microarray datasets (Figure 9).

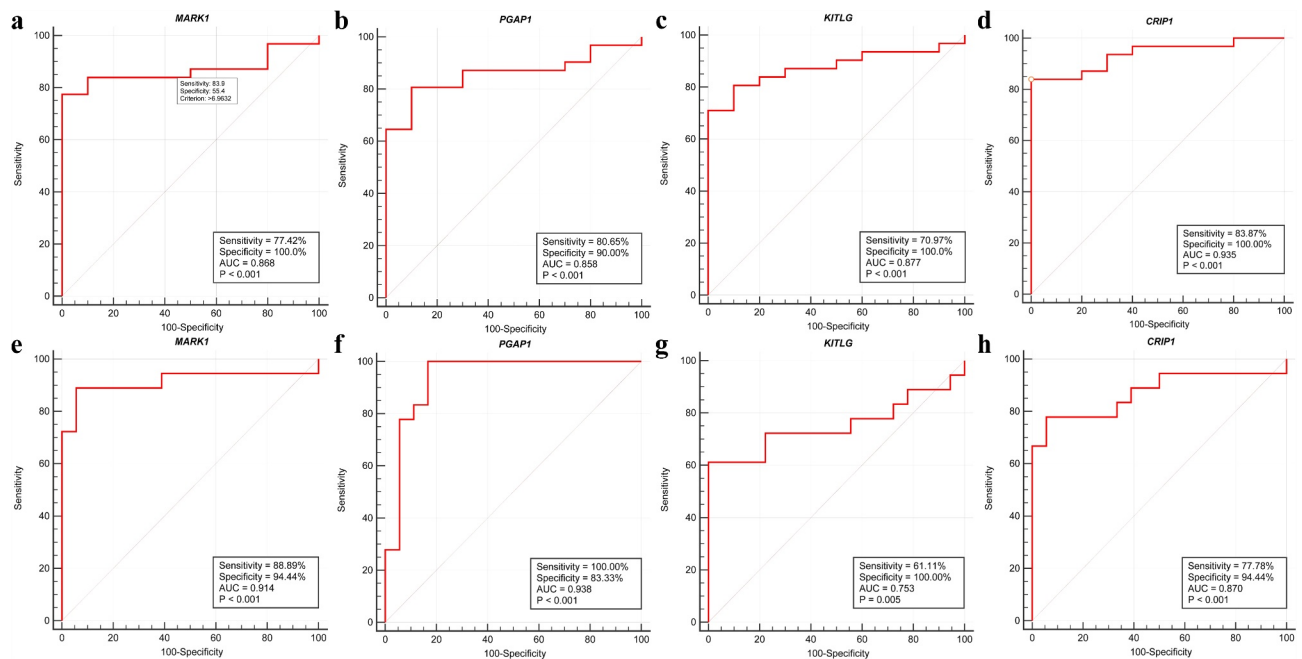
### Identification of potential small-molecule drugs for NPC treatment

The hub genes were uploaded to the CMAP database to identify potential small-molecule drugs for NPC treatment. Based on the criteria: number of instances  $n > 5$  and  $P < 0.05$ , the top ten most

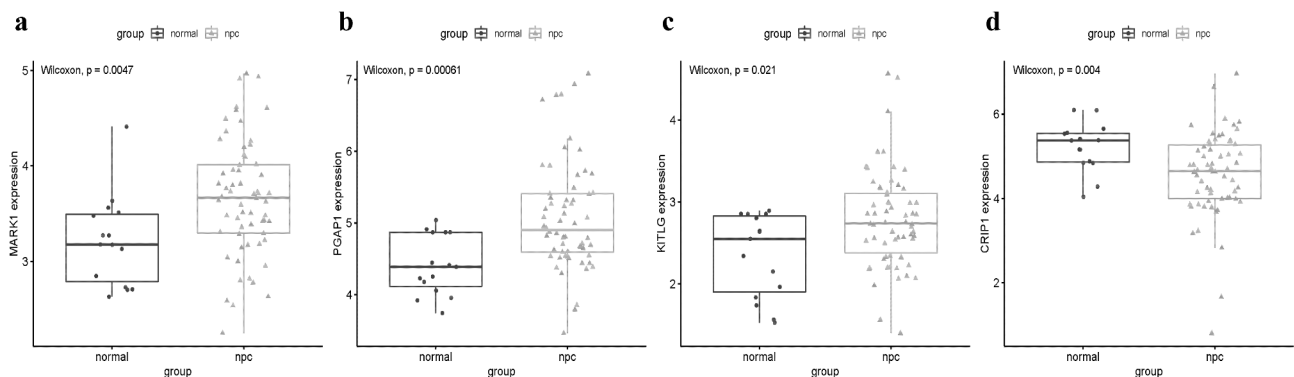
significant small-molecule drugs, which might be potential therapeutic targets for NPC treatment, were identified to be vorinostat, pioglitazone, fludrocortisone, thalidomide, bendroflumethiazide, chlorcyclizine, etiocholanolone, hyoscyamine, sulfadiazine, and biperiden (Table 3 and Figure 10).

### Validation of the expression levels of four hub genes in NPC tissues and cell lines

The mRNA expression levels of *MARK1*, *PGAP1*, *KITLG*, and *CRIP1* genes in 20 pairs of NPC and adjacent tissues were quantified via qRT-PCR.



**Figure 8.** Validation of hub genes in the diagnostic value. (a-d) Validation of hub genes in the GSE12452. Receiver operating characteristic (ROC) curves and area under the curve (AUC) statistics are used to evaluate the capacity to discriminate NPC from normal controls with excellent sensitivity and specificity. (e-h) Validation of hub genes in the GSE53819 and the results were the similar as earlier. These findings indicated these four hub genes have excellent diagnostic efficiency in NPC.

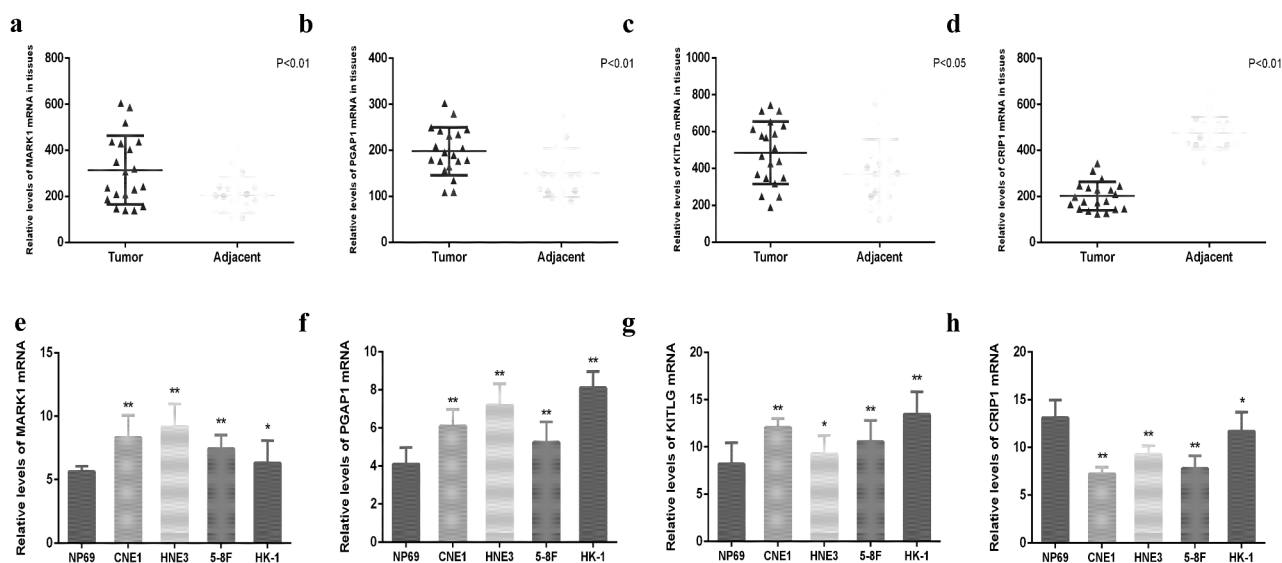


**Figure 9.** Validation the expression levels of four hub genes in the combined microarray datasets. (a-d) PGAP1, MARK1, and KITLG were significantly higher expression in NPC compared with normal tissues, while CRIP1 was significantly lower expression in NPC compared with normal tissues.

**Table 3.** Top 10 most significant small molecule drugs based on CMAP.

CMAP name	Mean	Number	Enrichment	P-value	Specificity	Percent non-null
vorinostat	0.369	12	0.423	0.0171	0.6181	58
pioglitazone	0.234	11	0.412	0.03326	0.2313	54
fludrocortisone	-0.265	8	-0.469	0.03975	0.3732	50
thalidomide	0.497	7	0.607	0.00501	0	71
bendroflumethiazide	0.33	6	0.598	0.01363	0.0145	50
chlorthalidone	-0.486	6	-0.592	0.01619	0.0089	66
etiocholanolone	-0.481	6	-0.581	0.01903	0.2208	66
hyoscyamine	0.62	5	0.717	0.00411	0	100
sulfadiazine	0.667	5	0.711	0.00473	0	100
biperiden	-0.354	5	-0.674	0.00863	0.1905	60

**Abbreviations:** CMAP, Connectivity map.



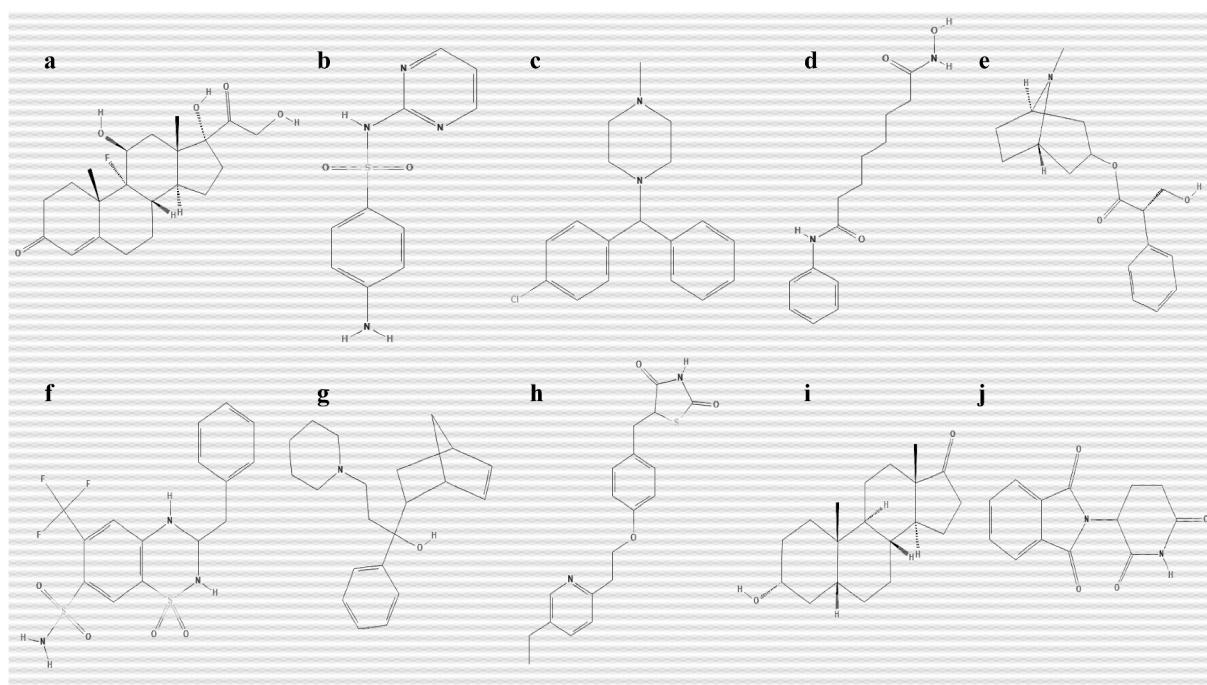
**Figure 10.** Relative expression of *MARK1*, *PGAP1*, *KITLG* and *CRIP1* mRNA in clinical nasopharyngeal carcinoma tissues and cell lines. (a-d) The relative expression of *MARK1*, *PGAP1*, *KITLG* and *CRIP1* mRNA in 20 paired nasopharyngeal carcinoma tissues and adjacent tissues were detected by qRT-PCR. (e-h) The relative expression of *MARK1*, *PGAP1*, *KITLG* and *CRIP1* mRNA in 5 cell lines (NP69, CNE1, HNE3, 5-8 F and HK-1) were detected by q-RT-PCR. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

The results showed that the expression of *MARK1*, *PGAP1*, and *KITLG* was upregulated in NPC samples relative to adjacent tissues (Figure 11(a-c)), while the expression of *CRIP1* was downregulated in NPC samples relative to adjacent tissues (Figure 11d). The expression of *MARK1*, *PGAP1*, *KITLG*, and *CRIP1* genes in four NPC cell lines (CNE1, HNE3, 5-8 F, and HK-1) and normal human immortalized nasopharynx cell line (NP69) were measured via qRT-PCR. The results showed that *MARK1*, *PGAP1*, and *KITLG* expression was upregulated in NPC cell lines relative to the normal human nasopharynx cell line (Figure 11(e-g)), while *CRIP1* expression was downregulated in NPC cell lines relative to normal human nasopharynx cell lines (Figure 11h).

## Discussion

Based on estimates from GLOBOCAN, there were 129,079 newly diagnosed NPC cases and 72,987 deaths reported globally in 2018 [1]. Etiologic factors associated with NPC include genetic, ethnic, EBV infection, and environmental factors [4-8]. Despite substantial improvements in the treatment of NPC over the past decade, the prognosis for patients with NPC diagnosis remain largely unsatisfactory due to a lack of obvious clinical signs in its early stages, with most initial diagnoses being made when the disease is at an advanced stage [3]. Therefore, elucidating the molecular mechanisms underlying NPC is of great significance.

WGCNA was used to identify potential molecular biomarkers and small-molecule drugs



**Figure 11.** The 2D molecular structures of the top ten most significant small-molecule drugs. (a) Fludrocortisone. (b) Sulfadiazine. (c) Chlorcyclizine. (d) Vorinostat. (e) Hyoscyamine. (f) Bendroflumethiazide. (g) Biperiden. (h) Pioglitazone. (i) Etiocholanolone. (j) Thalidomide.

associated with NPC progression. We identified 430 DEGs and 11 co-expressed gene modules. The green module had the highest correlation with NPC tumor stage. Four genes, which had high functional significance in the clinically significant module, were identified as the final hub genes. Compared with normal tissues, *PGAP1*, *MARK1*, and *KITLG* showed significantly higher expression in patients with NPC, while *CRIP1* expression was significantly lower in patients with NPC in both the test and validation datasets. These results were confirmed via qRT-PCR. The results of the ROC analysis showed that these four hub genes had excellent diagnostic efficiency in tumor and normal tissues. Furthermore, functional enrichment analysis showed that green module genes were mainly enriched in cellular response to nutrient levels, regulation of cell cycle process, and p53 signaling pathway.

*PGAP1*, the gene encoding the post-GPI attachment to protein 1, is closely associated with the cell cycle process, cell proliferation, cell differentiation, and multiple signaling transduction pathways. *PGAP1* can directly and/or indirectly induce carcinogenesis through tumor cell proliferation, migration, and resistance to multiple drugs [22–25].

*MARK1*, also known as partitioning defective gene 1 (*Par-1*), is a member of the serine-threonine kinase family, which plays a vital role in the regulation of cell migration, cell proliferation, and establishment and maintenance of cell polarity and microtubules [26–28]. Several studies have revealed that *MARK1* plays an important role in tumorigenesis, and Natalia *et al* [29] found that *MARK1* is a novel functional target for miR-125a-5p, with implications in the regulation of stimulated cell migration of cervical tumor cell lines. Tang *et al* [30] showed that *MARK1* overexpression was correlated with cell migration and proliferation in colorectal cancer through its regulation of miR-23a expression. *KITLG*, encoding the ligand for the tyrosine kinase receptor encoded by the *KIT* locus, plays an essential role in the regulation of cell survival and proliferation, stem cell maintenance, mast cell development, and migration. *KITLG*/SCF binding can activate several signal transduction pathways [31–33]. *KITLG* gene expression levels were evaluated in testicular germ cell cancer, breast cancer, and gastrointestinal stromal tumors [34–37]. Overexpression of *KITLG* was correlated with increased tumor cell migration and proliferation, suggesting a tumorigenic role for *KITLG* [34,36]. *KITLG* was

expressed at significantly higher levels in normal mammary gland epithelial tissue than in breast tumor tissue, indicating that it might be involved in the homeostasis of normal mammary tissue and its disruption would confer breast carcinogenesis [38]. CRIP1 is a member of the LIM/double zinc finger protein family, which is highly expressed in the intestine and immune cells [39]. In gastric cancer, *CRIP1* was overexpressed in primary tumor tissues and confirmed to be an independent prognostic factor. Patients with gastric carcinoma overexpressing *CRIP1* had shorter overall survival [40]. Compared with adjacent normal tissues, *CRIP1* was significantly more highly expressed in cervical cancer and was associated with FIGO stage [41].

The connectivity map database was used to screen small-molecule drugs with promising therapeutic capabilities against NPC. To screen out potential small-molecule drugs for treating NPC, 66 candidate small-molecule drugs were obtained from the CMAP database prediction based on the final NPC hub genes. Among the top ten most significant potential small-molecule drugs, vorinostat and thalidomide are particularly interesting and have anti-tumor effects in NPC treatment. Vorinostat is a histone deacetylase inhibitor (HDACi) that has shown strong antitumor effects in various types of solid cancers when combined with other traditional chemotherapeutic drugs such as camptothecin and gemcitabine [42–47]. Thalidomide is a glutamic acid derivative and is a classic drug with immunomodulatory properties that inhibits tumor necrosis factor alpha [48–50]. Subsequent studies found that thalidomide has anti-angiogenic, anti-inflammatory, anti-fibrotic, and immunomodulatory effects, and the clinical effectiveness of thalidomide in the treatment of diseases such as refractory Crohn's disease, lung metastasis, multiple myeloma, hepatocellular carcinoma, and breast carcinoma has been confirmed [51–58]. The CMAP database results provide several potential small-molecule drugs for future NPC therapy. However, *in vivo* and *in vitro* studies are necessary to validate their therapeutic efficacies.

## Conclusion

In summary, by constructing a weighted gene co-expression network with data from the GEO database, our study identified four hub genes

correlated with prognosis in NPC tumors and several potential small-molecule drugs for NPC treatment, which may contribute to new insights for NPC therapy. Further studies, including *in vivo* and *in vitro* experiments, are necessary to elucidate the molecular mechanisms of hub genes for future clinical applications.

## Abbreviations

NPC: Nasopharyngeal carcinoma; IARC: International Agency for Research on Cancer; ASR: age-standardized; EBV: Epstein-Barr virus; WGCNA: weighted gene co-expression network analysis; GEO: Gene Expression Omnibus; RMA: Robust MultiArray Averaging; GS: gene significance; ME: module eigengene; MM: module membership; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; ROC: receiver operating characteristic; CMAP: Connectivity map; CRIP1: cysteine-rich intestinal protein 1; KITLG: KIT ligand; MARPK1: microtubule affinity regulating kinase 1; PGAP1: post-GPI attachment to proteins 1; HDACi: histone deacetylase inhibitors.

## Highlights

- (1) CRIP1, KITLG, MARK1 and PGAP1 have excellent diagnostic efficiency in NPC.
- (2) Four prognostic biomarkers and several candidate small-molecule drugs for NPC.
- (3) Ten small-molecule drugs might provide potentially therapeutic goals for NPC.

## Acknowledgements

Not applicable.

## Availability of data and materials

The publicly published GEO datasets are available online.

## Disclosure statement

The authors declare that they have no competing interests.

## Author contributions

X.P. made substantial contributions to conception and design, acquisition of data, and analysis and interpretation of data; J.H.L. contributed to data analyses and the interpretation and completion of the figures. All authors read and approved the final article.

**ORCID**Xi Pan  <http://orcid.org/0000-0002-4344-1370>Jian-Hao Liu  <http://orcid.org/0000-0002-4177-012X>**References**

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* **2018**;68(6):394–424.
- [2] Pathmanathan R, et al. Undifferentiated, nonkeratinizing, and squamous cell carcinoma of the nasopharynx. variants of epstein-barr virus-infected neoplasia. *Am J Pathol.* **1995**;146(6):1355–1367.
- [3] Bensouda Y, Kaikani W, Ahbeddou N, et al. Treatment for metastatic nasopharyngeal carcinoma. *Eur Ann Otorhinolaryngol Head Neck Dis.* **2011**;128(2):79–85.
- [4] Henderson BE, Louie E, Jing SH, et al. Risk factors associated with nasopharyngeal carcinoma. *N Engl J Med.* **1976**;295(20):1101–1106.
- [5] Nakanishi Y, Wakisaka N, Kondo S, et al. Progression of understanding for the role of epstein-barr virus and management of nasopharyngeal carcinoma. *Cancer Metastasis Rev.* **2017**;36(3):435–447.
- [6] Nam JM, McLaughlin JK, Blot WJ. Cigarette smoking, alcohol, and nasopharyngeal carcinoma: a case-control study among U.S whites. *J Natl Cancer Inst.* **1992**;84(8):619–622.
- [7] Tsang CM, Tsao SW. The role of Epstein-Barr virus infection in the pathogenesis of nasopharyngeal carcinoma. *Virol Sin.* **2015**;30(2):107–121.
- [8] Tsao SW, Tsang CM, To K-F, et al. The role of Epstein-Barr virus in epithelial malignancies. *J Pathol.* **2015**;235(2):323–333.
- [9] Yang Z, Wang J, Zhang Z, et al. Epstein-barr virus-encoded products promote circulating tumor cell generation: a novel mechanism of nasopharyngeal carcinoma metastasis. *Onco Targets Ther.* **2019**;12:11793–11804.
- [10] Tsao SW, Tsang CM, Lo KW. Epstein-Barr virus infection and nasopharyngeal carcinoma. *Philos Trans R Soc Lond B Biol Sci.* **2017**;372(1732):20160270.
- [11] Chan AS, To KF, Lo KW, et al. Frequent chromosome 9p losses in histologically normal nasopharyngeal epithelia from southern Chinese. *Int J Cancer.* **2002**;102(3):300–303.
- [12] Bei J-XJ-X, Zuo X-Y, Liu W-S, et al. Genetic susceptibility to the endemic form of NPC. *Chin Clin Oncol.* **2016**;5(2):15.
- [13] Bei JX, Jia WH, Zeng YX. Familial and large-scale case-control studies identify genes associated with nasopharyngeal carcinoma. *Semin Cancer Biol.* **2012**;22(2):96–106.
- [14] Hildesheim A, Wang CP. Genetic predisposition factors and nasopharyngeal carcinoma risk: a review of epidemiological association studies, 2000-2011: rosetta stone for NPC: genetics, viral infection, and other environmental factors. *Semin Cancer Biol.* **2012**;22(2):107–116.
- [15] Dai W, Zheng H, Cheung AKL et al. Whole-exome sequencing identifies MST1R as a genetic susceptibility gene in nasopharyngeal carcinoma. *Proc Natl Acad Sci U S A.* **2016**;113(12):3317–3322.
- [16] Yu G, Hsu W-L, Coghill AE, et al. Whole-exome sequencing of nasopharyngeal carcinoma families reveals novel variants potentially involved in nasopharyngeal carcinoma. *Sci Rep.* **2019**;9(1):9916.
- [17] Cheung AKLAK, Ip JCY, Chu ACH et al. PTPRG suppresses tumor growth and invasion via inhibition of Akt signaling in nasopharyngeal carcinoma. *Oncotarget.* **2015**;6(15):13434–13447.
- [18] Edgar R. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**;30(1):207–210.
- [19] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* **2008**;9(1):559.
- [20] Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* **2019**;10(1):1523.
- [21] Lamb J, Crawford ED, Peck D, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science.* **2006**;313(5795):1929–1935.
- [22] Li C, Yu S, Nakamura F, et al. Binding of pro-priorin to filamin A disrupts cytoskeleton and correlates with poor prognosis in pancreatic cancer. *J Clin Invest.* **2009**;119(9):2725–2736.
- [23] Ho JC, Cheung ST, Patil M, et al. Increased expression of glycosyl-phosphatidylinositol anchor attachment protein 1 (GPAA1) is associated with gene amplification in hepatocellular carcinoma. *Int J Cancer.* **2006**;119(6):1330–1337.
- [24] Jiang WW, Zahurak M, Zhou Z-T, et al. Alterations of GPI transamidase subunits in head and neck squamous carcinoma. *Mol Cancer.* **2007**;6(1):74.
- [25] Amit M, Na'ara S, Francis D, et al. Post-translational regulation of radioactive iodine therapy response in papillary thyroid carcinoma. *J Natl Cancer Inst.* **2017**;109(12). DOI:10.1093/jnci/djx092.
- [26] Hurov J, Piwnicka-Worms H. The par-1/MARK family of protein kinases: from polarity to metabolism. *Cell Cycle.* **2007**;6(16):1966–1969.
- [27] Matenia D, Mandelkow EM. The tau of MARK: a polarized view of the cytoskeleton. *Trends Biochem Sci.* **2009**;34(7):332–342.
- [28] McDonald JA. Canonical and noncanonical roles of Par-1/MARK kinases in cell migration. *Int Rev Cell Mol Biol.* **2014**;312:169–199.

- [29] Natalia MA, Alejandro G-T, Virginia T-VJ et al. MARK1 is a novel target for miR-125a-5p: implications for cell migration in cervical tumor cells. *Microna*. 2018;7(1):54–61.
- [30] Tang X, Yang M, Wang Z, et al. MicroRNA-23a promotes colorectal cancer cell migration and proliferation by targeting at MARK1. *Acta Biochim Biophys Sin (Shanghai)*. 2019;51(7):661–668.
- [31] Zazo Seco C, Serrão de Castro L, van Nierop J, et al. Allelic mutations of KITLG, encoding KIT ligand, cause asymmetric and unilateral hearing loss and waardenburg syndrome type 2. *Am J Hum Genet*. 2015;97(5):647–660.
- [32] Wang ZQ, Si LZ, Tang Q, et al. Gain-of-function mutation of KIT ligand on melanin synthesis causes familial progressive hyperpigmentation. *Am J Hum Genet*. 2009;84(5):54–61.
- [33] Yuzawa S, Opatowsky Y, Zhang Z, et al. Structural basis for activation of the receptor tyrosine kinase KIT by stem cell factor. *Cell*. 2007;130(2):323–334.
- [34] Hirano K, Kitazawa A, Kojima K, et al. Expression of stem cell factor (SCF), a KIT ligand, in gastrointestinal stromal tumors (GISTs): a potential marker for tumor proliferation. *Pathol Res Pract*. 2008;204(11):799–807.
- [35] Ulivi P, Zoli W, Medri L, et al. c-kit and SCF expression in normal and tumor breast tissue. *Breast Cancer Res Treat*. 2004;83(1):33–42.
- [36] Lammie A, Drobnjak M, Gerald W, et al. Expression of c-kit and kit ligand proteins in normal human tissues. *J Histochem Cytochem*. 1994;42(11):1417–1425.
- [37] Chanock S. High marks for GWAS. *Nat Genet*. 2009;41(7):765–766.
- [38] Chen W, Li J, Liu C, et al. A functional p53 responsive polymorphism in KITLG, rs4590952, does not affect the risk of breast cancer. *Sci Rep*. 2014;4(1):6371.
- [39] Cousins RJ, Lanningham-Foster L. Regulation of cysteine-rich intestinal protein, a zinc finger protein, by mediators of the immune response. *J Infect Dis*. 2000;182(Suppl 1):S81–4.
- [40] Balluff B, Rauser S, Meding S, et al. MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer. *Am J Pathol*. 2011;179(6):2720–2729.
- [41] Zhang LZ, Huang L-Y, Huang A-L, et al. CRIP1 promotes cell migration, invasion and epithelial-mesenchymal transition of cervical cancer by activating the Wnt/ $\beta$ -catenin signaling pathway. *Life Sci*. 2018;207:420–427.
- [42] Liu S, Hu Z, Zhang Q, et al. Co-Prodrugs of 7-Ethyl-10-hydroxycamptothecin and vorinostat with in vitro hydrolysis and anticancer effects. *ACS Omega*. 2020;5(1):350–357.
- [43] Duvic M, Vu J. Vorinostat: a new oral histone deacetylase inhibitor approved for cutaneous T-cell lymphoma. *Expert Opin Investig Drugs*. 2007;16(7):1111–1120.
- [44] Carew JS, Giles FJ, Nawrocki ST. Histone deacetylase inhibitors: mechanisms of cell death and promise in combination cancer therapy. *Cancer Lett*. 2008;269(1):7–17.
- [45] Han L, Wang T, Wu J, et al. A facile route to form self-carried redox-responsive vorinostat nanodrug for effective solid tumor therapy. *Int J Nanomedicine*. 2016;11:6003–6022.
- [46] Bruzzese F, Rocco M, Castelli S, et al. Synergistic antitumor effect between vorinostat and topotecan in small cell lung cancer cells is mediated by generation of reactive oxygen species and DNA damage-induced apoptosis. *Mol Cancer Ther*. 2009;8(11):3075–3087.
- [47] Hui KF, Ho DN, Tsang CM, et al. Activation of lytic cycle of Epstein-Barr virus by suberoylanilide hydroxamic acid leads to apoptosis and tumor growth suppression of nasopharyngeal carcinoma. *Int J Cancer*. 2012;131(8):1930–1940.
- [48] Franks ME, Macpherson GR, Figg WD. Thalidomide. *Lancet*. 2004;363(9423):1802–1811.
- [49] Sampaio EP, Sarno EN, Galilly R, et al. Thalidomide selectively inhibits tumor necrosis factor alpha production by stimulated human monocytes. *J Exp Med*. 1991;173(3):699–703.
- [50] Corral LG, Muller GW, Moreira AL, et al. Selection of novel analogs of thalidomide with enhanced tumor necrosis factor alpha inhibitory activity. *Mol Med*. 1996;2(4):506–515.
- [51] He Y, Mao R, Chen F, et al. Thalidomide induces clinical remission and mucosal healing in adults with active Crohn's disease: a prospective open-label study. *Therap Adv Gastroenterol*. 2017;10(5):397–406.
- [52] Hu H, Wang X, Liu S. Thalidomide induces mucosal healing in postoperative crohn disease endoscopic recurrence: case report and literature review. *Medicine (Baltimore)*. 2016;95(36):e4799.
- [53] Bauditz J. Thalidomide reduces tumour necrosis factor alpha and interleukin 12 production in patients with chronic active crohn's disease. *Gut*. 2002;50(2):196–200.
- [54] Miyazato K, Tahara H, Hayakawa Y. Anti-metastatic effects of thalidomide by inducing the functional maturation of peripheral NK cells. *Cancer Sci*. 2020;111(8):2770–2778.
- [55] Wang A, Duan QH, Liu X, et al. (Bortezomib plus lenalidomide/thalidomide)- vs. (bortezomib or lenalidomide/thalidomide)-containing regimens as induction therapy in newly diagnosed multiple myeloma: a meta-analysis of randomized controlled trials. *Ann Hematol*. 2012;91(11):1779–1784.
- [56] Aguiar PM, de # T, Colleoni GWB et al. Efficacy and safety of bortezomib, thalidomide, and lenalidomide in multiple myeloma: an overview of systematic reviews with meta-analyses. *Crit Rev Oncol Hematol*. 2017;113:195–212.

- [57] Cao DD, Xu H-L, Liu L, et al. Thalidomide combined with transcatheter arterial chemoembolization for primary hepatocellular carcinoma: a systematic review and meta-analysis. *Oncotarget*. 2017;8(27):44976–44993.
- [58] Wang X, Shen Y, MengLv L, et al. Thalidomide suppresses breast cancer tumor growth by inhibiting tumor-associated macrophage accumulation in breast tumor-bearing mice. *Eur J Pharm Sci*. 2020;151:105302.