Identification of Four Key Biomarkers and Small Molecule Drugs Innasopharyngeal Carcinoma by Weighted Gene Co-expression Network Analysis

Xi Pan (xy3panxi@csu.edu.cn)
Xiangya Hospital Central South University
Jian-Hao Liu
Xiangya Hospital Central South University

Research

Keywords: Nasopharyngeal carcinoma, Hub genes, WGCNA, Signature, Small-molecule drugs

Posted Date: July 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-49643/v1

Abstract

Background

Nasopharyngeal carcinoma (NPC) is a heterogeneous carcinoma that the underlying molecular mechanisms involved in the tumor initiation, progression, and migration are largely unclear. The purpose of the present study was to identify key biomarkers and small-molecule drugs for NPC screening, diagnosis, and therapy via gene expression profile analysis.

Methods

Raw microarray data of NPC were retrieved from the Gene Expression Omnibus (GEO) database and analyzed to screen out the potential differentially expressed genes (DEGs). The key modules associated with histology grade and tumor stage was identified by using weighted correlation network analysis (WGCNA). Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of genes in the key module were performed to identify potential mechanisms. Candidate hub genes were obtained, which based on the criteria of module membership (MM) and high connectivity. Then we used receiver operating characteristic (ROC) curve to evaluate the diagnostic value of hub genes. The Connectivity map database was further used to screen out small-molecule drugs of hub genes.

Results

A total of 430 DEGs were identified based on two GEO datasets. The green gene module was considered as key module for the tumor stage of NPC via WGCNA analysis. The results of functional enrichment analysis revealed that genes in the green module were enriched in regulation of cell cycle, p53 signaling pathway, cell part morphogenesis. Furthermore, four DEGs-related hub genes in the green module were considered as the final hub genes. Then ROC revealed that the final four hub genes presented with high areas under the curve, suggesting these hub genes may be diagnostic biomarkers for NPC. Meanwhile, we screened out several small-molecule drugs that have provided potentially therapeutic goals for NPC.

Conclusions

Our research identified four potential prognostic biomarkers and several candidate small-molecule drugs for NPC, which may contribute to the new insights for NPC therapy.

Background

Nasopharyngeal carcinoma (NPC) is a malignant carcinoma arising from the nasopharyngeal epithelial lining of the nasopharynx with a unique geographical distribution and racial prevalence. According to the GLOBOCAN cancer estimates and the International Agency for Research on Cancer (IARC), in 2018, there

were about 129079 newly diagnosed NPC cases and 72987 deaths, accounting for about 0.7% cancer cases and deaths in that year [1, 2]. Nevertheless, its geographical distribution of global incidence is extremely unbalanced; more than 80% of newly diagnosed NPC cases are in Asia and fewer than 0.2% cases are in Oceania, with an age-standardized (ASR) of 10.0 per 100000 populations in South-Eastern Asia to 0.5 per 100000 in populations that are central America [1, 2]. Despite substantial improvement in treatment options include chemotherapy and immunotherapy as well as radiotherapy for NPC during the past decade, patients with NPC prognosis remains unsatisfactory, approximately 20% NPC cases would recurrence and only offer a survival of 50–70% at five years [3]. Thus, novel effective potential diagnostic biomarkers and therapeutic strategies to combat NPC and thereby to improve treatment outcomes in patients with NPC is urgently required.

The remarkable geographical distribution and racial prevalence of NPC incidence has spurred studies on its risk factors, and it is suggested that several aetiological factors, such as genetic, ethic, Epstein-Barr virus (EBV) infection, and environmental factors (eg, cigarette smoking, exposure to dust, formaldehyde, and so on.) are contributors in the development and progression of NPC [4–8]. EBV infection is perhaps the most extensively studied etiological factor in the pathogenesis of NPC. This is achieved by probing for the EBV-encoded RNA by in-situ hybridization, the virus is exclusively found in all tumor cells but not in 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 types of human cancer, including NPC [7–10]. Besides EBV infection, genetic susceptibility is perhaps the most common causal agent of NPC. Studies have identified that genetic susceptibility genes involved in DNA repair, metabolic, and immune responses pathways 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, such as MST1R, PTPRG, HLA, MMP2, and so on [12–17].

With the rapid development of high-throughput whole-genome sequencing, including bioinformatic analysis, which play a significant role in screening candidate biomarkers for some diseases, especially NPC. Herein, the microarray data sets were applied to gain and validate key biomarkers in NPC via weighted gene co-expression network analysis (WGCNA), as well as to further identify small-molecule target drugs to treat NPC with promising perspectives.

Materials And Methods

Microarray datasets

The Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm. nih.gov/geo/), a public functional genomics data repository, which included array- and sequence-based gene profiles data [18]. Two mRNA microarray datasets, GSE12452 and GSE53819, were downloaded from the GEO database [19, 20]. GSE12452 contains 31 NPC tissues and 10 normal tissues, which sequencing platform used was GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array.

Data processing and differentially expressed genes (DEGs) identification

The Robust MultiArray Averaging (RMA) method was used for data pre-processing, including background adjustment, quantile normalization and log2 transformation of expression matrixes. The corresponding platforms were applied to annotate each probe based on Entrez ID. For genes with several probes, the median of all probes was selection. Then “limma” package of R language was applied to identify differentially expressed genes associated with NPC, with |logFC| >1 and *P*-value < 0.05 as the screening threshold.

Weighted co-expression network construction

The median absolute deviation of each gene expression value was calculated and the genes with median absolute deviation in the top 25% were chose for the WGCNA construction. The “WGCNA” package of R language was used to construct weighted gene co-expression network [21]. First, the function goodSamplesGenes was used to detect the good samples and good genes. Second, the function pickSoftThreshold was used to select an appropriate soft-thresholding power. Then the co-expression similarity was raised to achieve scale-free topology based on the appropriate soft-threshold power. Third, cox-expression gene modules were created via using the function cutreeDynamic. Fourth, we further 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-value* < 0.05 was considered as a statistically significant.

Functional enrichment analysis of hub module genes

To better understand the potential molecular mechanism of how hub module genes impact correlative clinical parameters, we have uploaded all hub module genes into Metascape (https://metascape.org/gp/index.html#/main/step1) online tool [22]. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were utilized based on the Metascape online tool. *P-value* < 0.01 was considered as a cut-off criterion.

Hub genes selection and validation

Select the gene with the highest intra-modular connectivity as candidate hub genes. Genes with the biologically significant usually have higher absolute values of GS. In the present study, candidate hub genes were screened out according to the criteria that the absolute values of GS > 0.20 and MM > 0.80. Then, the final hub genes were identified from the intersection of NPC-related DEGs and candidate hub genes. The expression level of hub genes in patients with NPC was assessed via using box plots. The diagnostic values of hub genes were evaluated via the area under the receiver operating characteristic (ROC). Meanwhile, we further validated the differential expression and diagnostic value of hub genes in patients with NPC via using separate external dataset (GSE53819).

Candidate small-molecule drugs identification

Connectivity map (CMAP) is an online database (https://portals.broadinstitute.org/cmap), which contains more than 7000 gene-expression profiles representing 1309 compounds, to predict small-molecule drugs for a specific disease[23]. To further identify small-molecule target drugs to treat NPC, the hub genes were uploaded to the CMAP in the 'query' page to identify the potential small-molecule target drugs to treat NPC. Correlation between the small-molecule drugs and hub genes were evaluated by enrichment scores, ranging from − 1 to 1. Chemical molecular structures of the small-molecule drugs were downloaded from PubChem Compound (https://pubchem.ncbi.nlm.nih.gov/).

Statistical analysis

To identify differentially expressed genes associated with NPC, the two-tailed Student's t-test was applied and genes with |logFC| >1 and *P-value* < 0.05 were screened out as the significant DEGs. All statistical analyses were conducted using R software (version 4.0.0) or MedCalc software (version 19.1).

Results

DEGs identification

Two microarray datasets of GSE12452 and GSE53819 were obtained from GEO database for DEGs identifying between NPC samples and normal samples. Based on the threshold of |logFC| >1 and *P-value* < 0.05, all DEGs were identified. The volcano plots for DEGs of two datasets separately were shown in Fig. 1A, 1B. 430 common genes were identified from the intersection of DEGs of two datasets for further WGCNA analysis. Among 430 genes, 165 genes were significantly up-regulated in NPC, while 265 genes were significantly down-regulated in NPC (Fig. 1C, 1D).

Weighted co-expression network construction and hub modules identification

The genes (3949 genes) with median absolute deviation in the top 25% with 31 NPC samples and 10 normal samples in GSE12452 were selected. Samples were clustered to remove outlier samples. WGCNA was then conducted to identify genes that were highly co-expressed across 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 in the present study, as shown in Fig. 2. Then the similar modules were merged based on the height of ME in the clustering equal to 0.25 (Fig. 3A). Finally, 12 co-expressed gene modules were identified. The "grey" module, which contained 583 genes not attributed to any modules, was removed in the subsequent analysis (Fig. 3B). The relationships of the 11 modules were analyzed with clinical parameters (including histology 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 (Fig. 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 based on the Metascape online tool. Terms with a *P-value* less than 0.01, minimum overlap genes equal to three, and an enrichment factor more than 1.50 were set as the cut-off criteria. GO analysis showed that green module genes were mainly enriched in plasma membrane bounded cell projection assembly, smoothened signaling pathway, cellular response to nutrient levels, regulation of cell cycle process, neuro maturation, protein kinase complex, and so on (Fig. 5). KEGG pathway enrichment was also presented in Fig. 5. Two significantly enriched pathways, p53 signaling pathway and proteoglycans in cancer, were identified association with genes in the green module.

Hub genes identification and validation

According to the criteria that the absolute values of GS > 0.2 and MM > 0.8, a total of 14 genes were screened as candidate hub genes, which have greatest connectivity in the green module (Fig. 6). Then, the final four hub genes were identified from the intersection of NPC-related DEGs and candidate hub genes, including cysteine-rich intestinal protein 1 (CRIP1), KIT ligand (KITLG), microtubule affinity regulating kinase 1 (MARK1), and post-GPI attachment to proteins 1 (PGAP1). Both in the datasets GSE12452 and GSE53819, compared with normal tissues, three of the final four hub genes showed significantly higher expression in NPC tissues, including PGAP1, MARK1, KITLG, while CRIP1 was significantly lower expression in NPC tissues (Fig. 7). In order to verify the diagnostic values of the final four hub genes, a ROC analysis was performed in the datasets GSE12452 and GSE53819. The results of ROC analysis demonstrated these four genes have excellent diagnostic efficiency in NPC (Fig. 8).

The potential small-molecule drugs for NPC treatment

The hub genes were uploaded to the CMAP to identify potential small-molecule drugs to treat NPC. According to the criterion that the number of instances (n > 5) and *P*-value (< 0.05), the top ten most significant potential small-molecule drugs were vorinostat, pioglitazone, fludrocortisone, thalidomide, bendroflumethiazide, chlorcyclizine, etiocholanolone, hyoscyamine, sulfadiazine, and biperiden, which might provide potentially therapeutic goals for NPC (Table 1 and Fig. 9).

Table 1
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
chlorcyclizine	-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

Discussion

According to estimates from the GLOBOCAN, an estimated 129079 newly diagnosed cases of NPC and 72987 patients die of this neoplasm were reported for 2018 worldwide [1]. Etiologic factors for NPC include genetic, ethic, EBV infection, and environmental factors [4–8]. Despite substantial improvement in treatment for NPC during the past decade, patients with NPC diagnosis and prognosis remains largely unsatisfactory due to lack of obvious clinical signs in its early stages and most of them were late stages at the time of clinical discovery [3]. Therefore, to reveal the molecular mechanism of NPC is great significance.

In our study, WGCNA was applied to identify potential molecular biomarkers and small-molecule drugs associated with the progression of NPC patients. We totally found 430 DEGs and 11 co-expressed gene modules through WGCNA analysis. The green module had the greatest correlation with NPC tumor stage. Four genes were identified as the final hub genes, which had high functional significance in the clinically significant module. Compared with normal tissues, PGAP1, MARK1, and KITLG were significantly higher expression in patients with NPC, while CRIP1 was significantly lower expression in patients with NPC in both test and validation datasets. Meanwhile, the results of ROC analysis showed that these four hub genes have excellent diagnostic efficiency for 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 proteins 1, which is closely associated with cell cycle process, cell proliferation, cell differentiation, and multiple signaling transduction pathways. PGAP1

can directly and/or indirectly induce carcinogenesis by tumor cell proliferation, migration, and multiple drug resistance [24–27]. MARK1, also known as partitioning defective gene 1 (Par-1), is a member of the family of serine-threonine kinases, which plays a vital role in the regulations of cell migration, cell proliferation, establishing and maintaining cell polarity, and microtubule [28–30]. Several research results revealed that MARK1 plays an important role in tumorigenesis, Natalia *et al*. found that MARK1 was a novel functional target for miR-125a-5p with implications on the regulation stimulated cell migration of cervical tumor cell lines [31]. Tang and his colleagues identified that MARK1 overexpression was correlated with cell migration and proliferation in colorectal cancer via regulated miR-23a expression [32]. KITLG, the gene encoding the ligand of the tyrosine-kinase receptor encoded by the KIT locus, which plays an essential role in the regulation of cell survival and proliferation, stem cell maintenance, mast cell development, migration. KITLG/SCF binding can active several signal transduction pathways [33–35]. KITLG mRNA expression levels were evaluated in testicular germ cell cancer, breast cancer, and gastrointestinal stromal tumors [36–39]. Overexpression of KITLG protein levels were correlated with increased tumor cell migration, proliferation, suggesting the tumorigenic role of KITLG [36, 38]. KITLG was significantly higher expressed in normal mammary gland epithelium tissue compared to breast tumor tissue, indicated that KITLG might be involved in the homeostasis of normal mammary and its disruption would confer to breast carcinogenesis [40]. CRIP1 is a member of the LIM/double zinc finger protein family, which is highly expressed in intestine and immune cells [41]. In gastric cancer, CRIP1 was over-expressed in primary tumor tissues, and confirmed as an independent prognostic factor. Patients with gastric carcinoma over-expression of CRIP1 had a shorter overall survival [42]. Compared to adjacent normal tissues, CRIP1 was significantly higher expressed in cervical cancer, and it was significantly associated with FIGO stage [43].

The Connectivity map database was further used to screen out small-molecule drugs with promising capacity as therapeutic goals for NPC. To screen out potential small-molecule drugs to treat NPC, 66 candidate small-molecule drugs were obtained from the prediction of CMAP database based on the final hub genes of NPC. Among the top ten most significant potential small-molecule drugs, vorinostat and thalidomide were particularly interesting and have anti-tumor effects in NPC treatment. Vorinostat is a histone deacetylase inhibitors (HDACi), which has shown strong anti-tumor effects in various types of solid cancers, when combined with other traditional chemotherapeutic drugs like camptothecin and gemcitabine[44–49]. Thalidomide, is a glutamic acid derivative, which was a classic drug with immune-modulatory properties and tumor necrosis factor alpha inhibits [50–52]. Subsequent studies found that thalidomide has anti-angiogenic, anti-inflammatory, anti-fibrotic and immune-modulatory effects, and the clinical effectiveness of thalidomide has been confirmed in the treatment of certain disease such as refractory Crohn's disease, lung metastasis, multiple myeloma, hepatocellular carcinoma, and breast carcinoma[53–60]. Our results based on the CMAP database may provide several potential small-molecule drugs as to future therapy for NPC; nevertheless, studies in vivo and in vitro experiments are necessary.

In summary, through the construction of weighted gene co-expression network with data from the GEO database, our study identified four hub genes correlated with prognostic in NPC tumor, which may

contribute to the new insights for NPC therapy, and several potential small-molecule drugs for NPC treatment. Meanwhile, further studies, including in vivo and in vitro experiments, are necessary to elucidate potential molecular mechanism 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.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

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

Funding

Not applicable.

Availability of data and materials

The publicly published GEO datasets are available online.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All the authors have consented for the publication.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, 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):1355–67.
3. Bensouda Y, et al. Treatment for metastatic nasopharyngeal carcinoma. Eur Ann Otorhinolaryngol Head Neck Dis. 2011;128(2):79–85.
4. Henderson BE, et al. Risk factors associated with nasopharyngeal carcinoma. N Engl J Med. 1976;295(20):1101–6.
5. Nakanishi Y, et al. Progression of understanding for the role of Epstein-Barr virus and management of nasopharyngeal carcinoma. Cancer Metastasis Rev. 2017;36(3):435–47.
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):619–22.
7. Tsang CM, Tsao SW. The role of Epstein-Barr virus infection in the pathogenesis of nasopharyngeal carcinoma. Virol Sin. 2015;30(2):107–21.
8. Tsao SW, et al. The role of Epstein-Barr virus in epithelial malignancies. J Pathol. 2015;235(2):323–33.
9. Yang Z, et al. Epstein-Barr Virus-Encoded Products Promote Circulating Tumor Cell Generation: A Novel Mechanism of Nasopharyngeal Carcinoma Metastasis. Onco Targets Ther. 2019;12:11793–804.
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).
11. Chan AS, et al. Frequent chromosome 9p losses in histologically normal nasopharyngeal epithelia from southern Chinese. Int J Cancer. 2002;102(3):300–3.
12. Bei JX, 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):107–16.
15. Dai W, 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):3317–22.
16. Yu G, 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 AK, et al. PTPRG suppresses tumor growth and invasion via inhibition of Akt signaling in nasopharyngeal carcinoma. Oncotarget. 2015;6(15):13434–47.
18. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207–10.
19. Sengupta S, et al. Genome-wide expression profiling reveals EBV-associated inhibition of MHC class I expression in nasopharyngeal carcinoma. Cancer research. 2006;66(16):7999–8006.
20. Bao Y-N, et al. Urokinase-type plasminogen activator receptor signaling is critical in nasopharyngeal carcinoma cell growth and metastasis. Cell Cycle. 2014;13(12):1958–69.
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.
22. Zhou Y, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature communications. 2019;10(1):1523.
23. Lamb J, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929–35.
24. Li C, et al. Binding of pro-prion to filamin A disrupts cytoskeleton and correlates with poor prognosis in pancreatic cancer. J Clin Invest. 2009;119(9):2725–36.
25. Ho JC, 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):1330–7.
26. Jiang WW, et al. Alterations of GPI transamidase subunits in head and neck squamous carcinoma. Mol Cancer. 2007;6:74.
27. Amit M, et al., Post-translational Regulation of Radioactive Iodine Therapy Response in Papillary Thyroid Carcinoma. J Natl Cancer Inst, 2017. 109(12).
28. Hurov J, Piwnica-Worms H. The Par-1/MARK family of protein kinases: from polarity to metabolism. Cell Cycle. 2007;6(16):1966–9.
29. Matenia D, Mandelkow EM. The tau of MARK: a polarized view of the cytoskeleton. Trends Biochem Sci. 2009;34(7):332–42.
30. McDonald JA. Canonical and noncanonical roles of Par-1/MARK kinases in cell migration. Int Rev Cell Mol Biol. 2014;312:169–99.
31. Natalia MA, et al. MARK1 is a Novel Target for miR-125a-5p: Implications for Cell Migration in Cervical Tumor Cells. Microrna. 2018;7(1):54–61.
32. Tang X, et al. MicroRNA-23a promotes colorectal cancer cell migration and proliferation by targeting at MARK1. Acta Biochim Biophys Sin (Shanghai). 2019;51(7):661–8.
33. Zazo Seco C, 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):647–60.
34. Wang ZQ, et al. Gain-of-function mutation of KIT ligand on melanin synthesis causes familial progressive hyperpigmentation. Am J Hum Genet. 2009;84(5):672–7.

35. Yuzawa S, et al. Structural basis for activation of the receptor tyrosine kinase KIT by stem cell factor. Cell. 2007;130(2):323–34.
36. Hirano 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.
37. Ulivi P, et al. c-kit and SCF expression in normal and tumor breast tissue. Breast Cancer Res Treat. 2004;83(1):33–42.
38. Lammie A, et al. Expression of c-kit and kit ligand proteins in normal human tissues. J Histochem Cytochem. 1994;42(11):1417–25.
39. Chanock S. High marks for GWAS. Nat Genet. 2009;41(7):765–6.
40. Chen W, et al. A functional p53 responsive polymorphism in KITLG, rs4590952, does not affect the risk of breast cancer. Sci Rep. 2014;4:6371.
41. 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.
42. Balluff B, et al. MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer. Am J Pathol. 2011;179(6):2720–9.
43. Zhang LZ, et al. CRIP1 promotes cell migration, invasion and epithelial-mesenchymal transition of cervical cancer by activating the Wnt/β–catenin signaling pathway. Life Sci. 2018;207:420–7.
44. Liu S, et al. Co-Prodrugs of 7-Ethyl-10-hydroxycamptothecin and Vorinostat with in Vitro Hydrolysis and Anticancer Effects. ACS Omega. 2020;5(1):350–7.
45. Duvic M, Vu J. Vorinostat: a new oral histone deacetylase inhibitor approved for cutaneous T-cell lymphoma. Expert Opin Investig Drugs. 2007;16(7):1111–20.
46. 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.
47. Han L, et al. A facile route to form self-carried redox-responsive vorinostat nanodrug for effective solid tumor therapy. Int J Nanomedicine. 2016;11:6003–22.
48. Bruzzese F, 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):3075–87.
49. Hui KF, 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):1930–40.
50. Franks ME, Macpherson GR, Figg WD. Thalidomide Lancet. 2004;363(9423):1802–11.
51. Sampaio EP, et al. Thalidomide selectively inhibits tumor necrosis factor alpha production by stimulated human monocytes. J Exp Med. 1991;173(3):699–703.
52. Corral LG, et al. Selection of novel analogs of thalidomide with enhanced tumor necrosis factor alpha inhibitory activity. Mol Med. 1996;2(4):506–15.

53. He Y, 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.
54. Hu H, Wang X, Liu S. Thalidomide induces mucosal healing in postoperative Crohn disease endoscopic recurrence: Case report and literature review. Medicine. 2016;95(36):e4799.
55. Bauditz J, Wedel S, Lochs H. Thalidomide reduces tumour necrosis factor alpha and interleukin 12 production in patients with chronic active Crohn's disease. Gut. 2002;50(2):196–200.
56. Miyazato K, Tahara H, Hayakawa Y. Anti-metastatic effects of thalidomide by inducing the functional maturation of peripheral NK cells. Cancer Sci, 2020.
57. Wang A, 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):1779–84.
58. Aguiar PM, 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.
59. Cao DD, et al. Thalidomide combined with transcatheter artierial chemoembolzation for primary hepatocellular carcinoma: a systematic review and meta-analysis. Oncotarget. 2017;8(27):44976–93.
60. Wang X, 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.

Figures

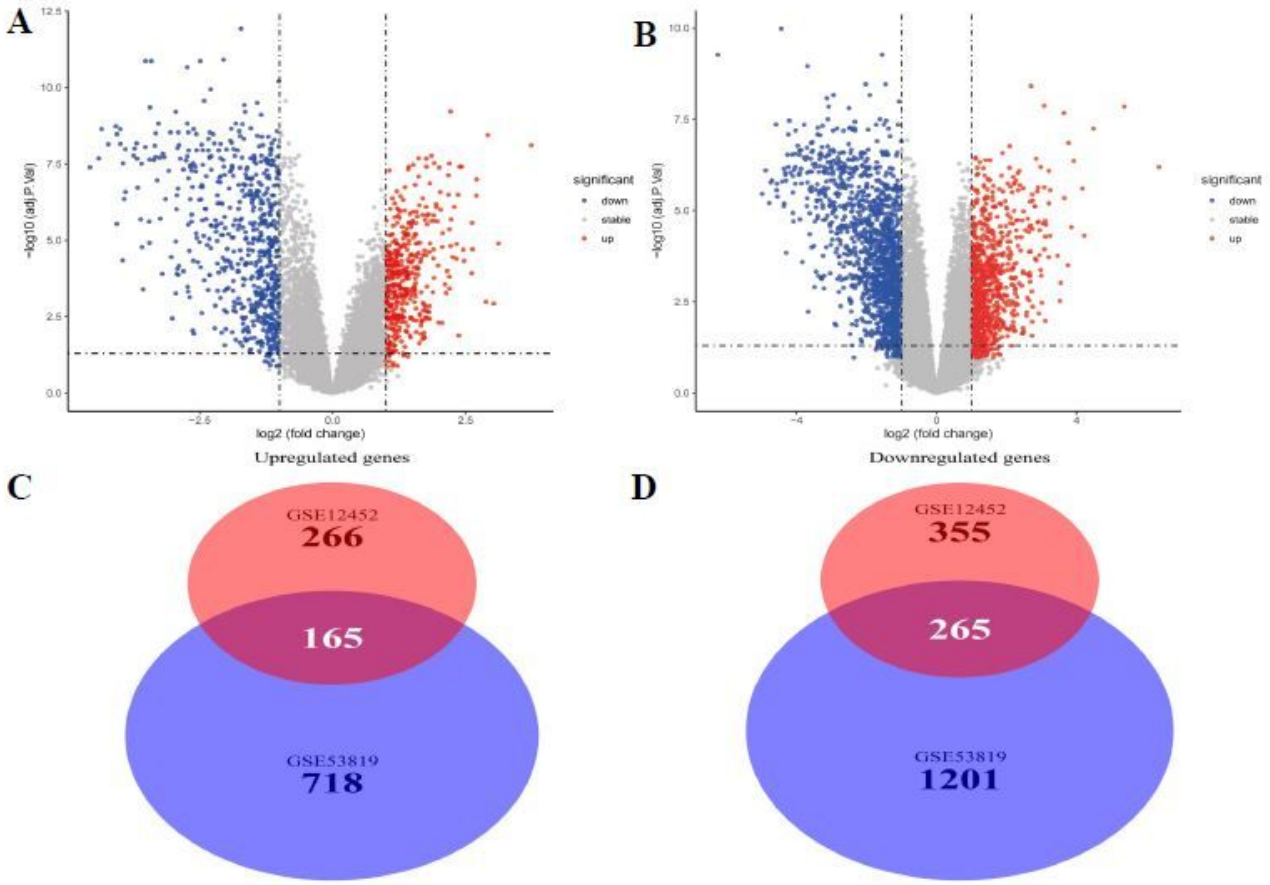

Figure 1

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.

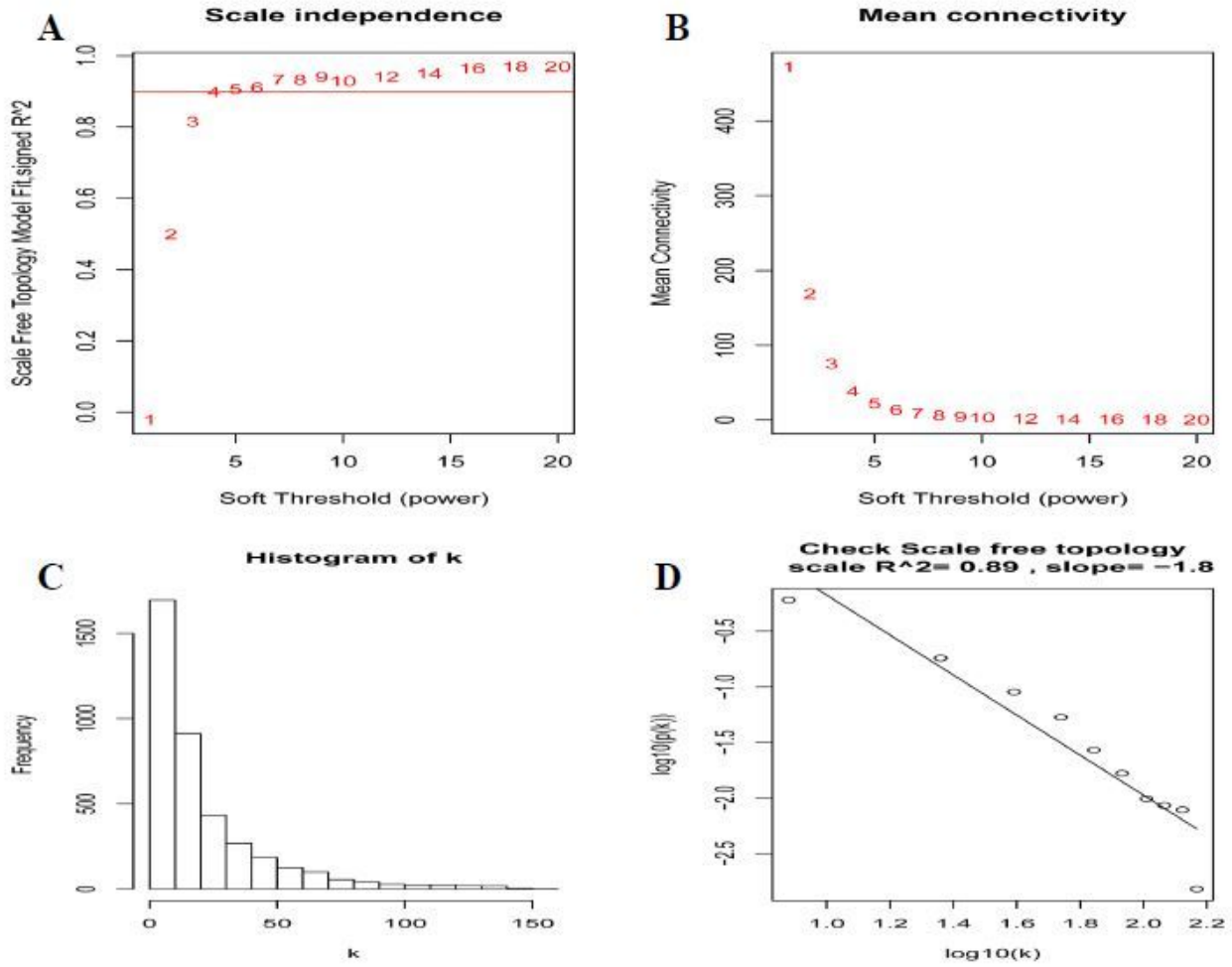

Figure 2

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 (β). 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 β = 5. (D) Checking the scale-free topology when β = 5.

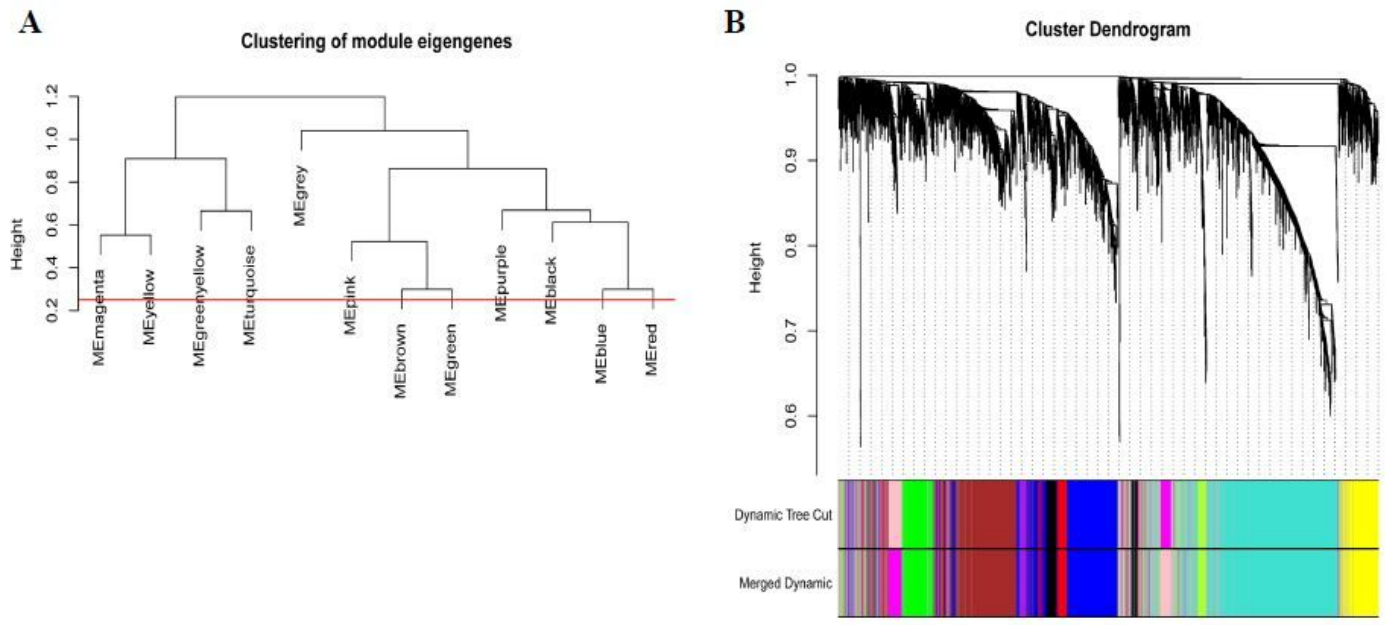

Figure 3

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.

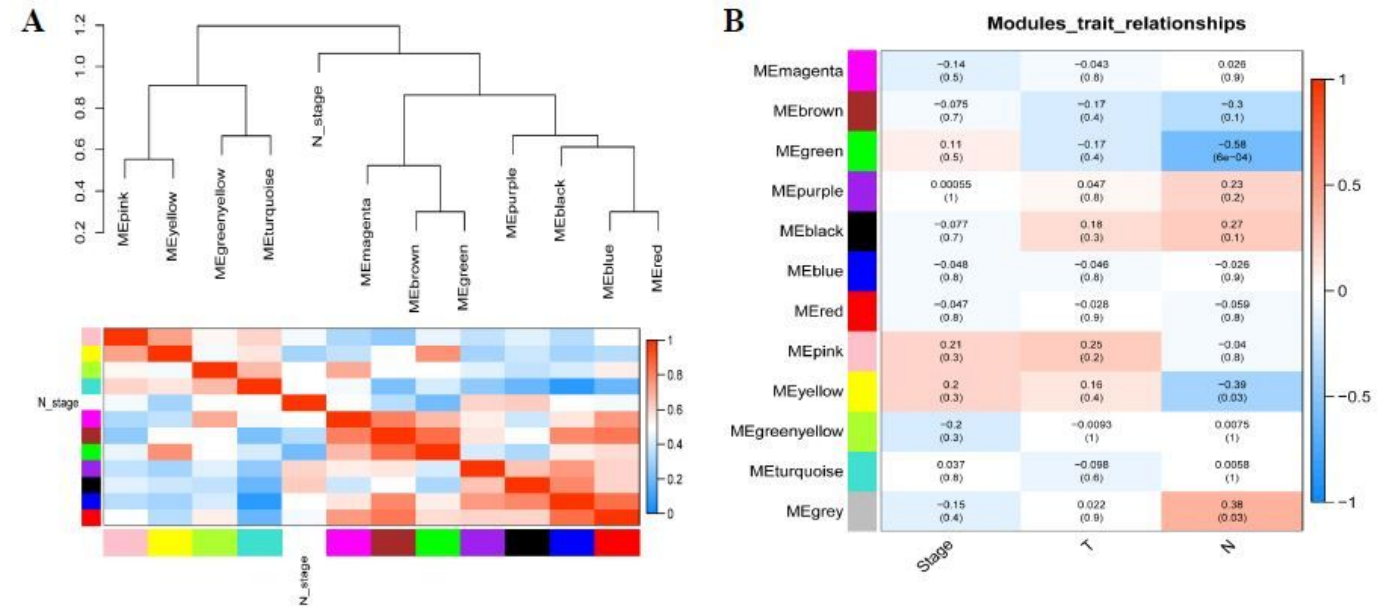

Figure 4

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 Nstage.

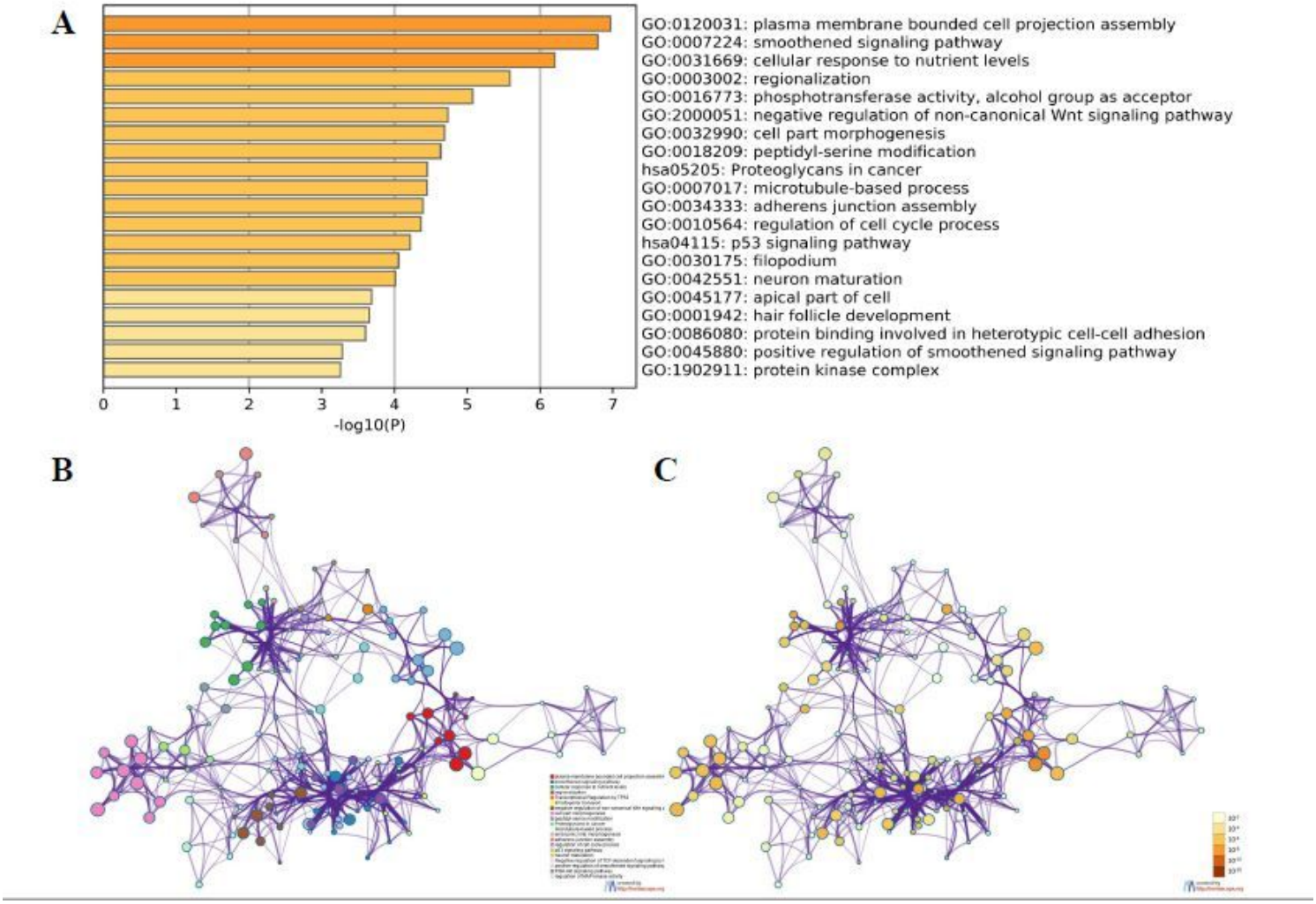

Figure 5

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 -log10(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.

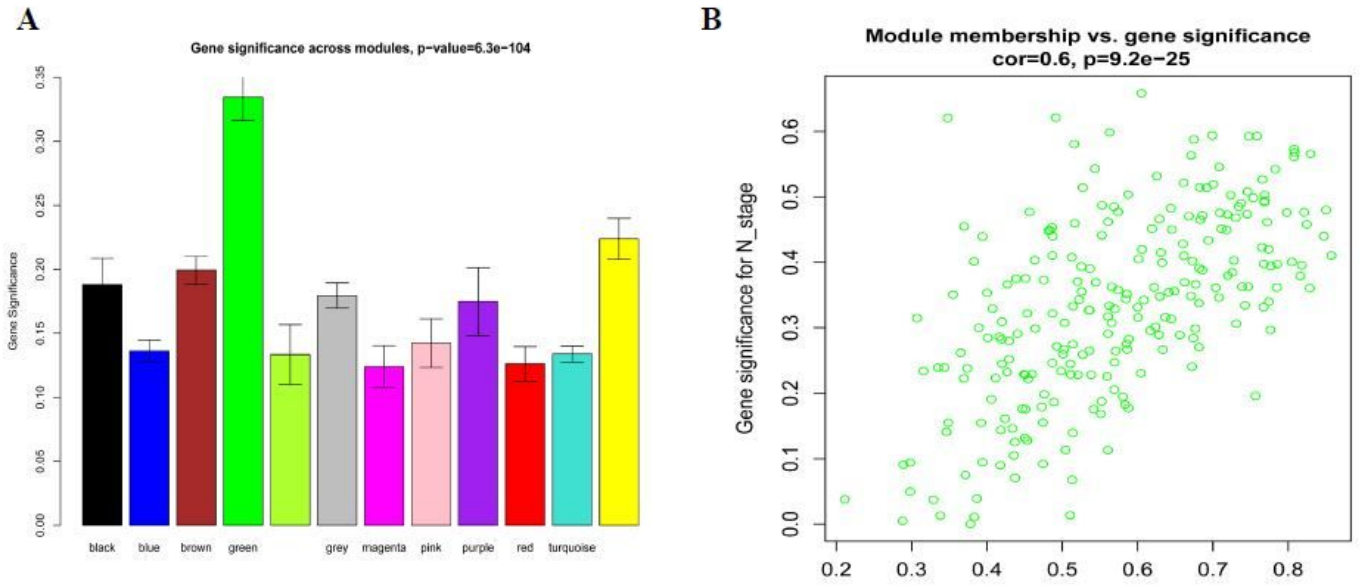

Figure 6

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.

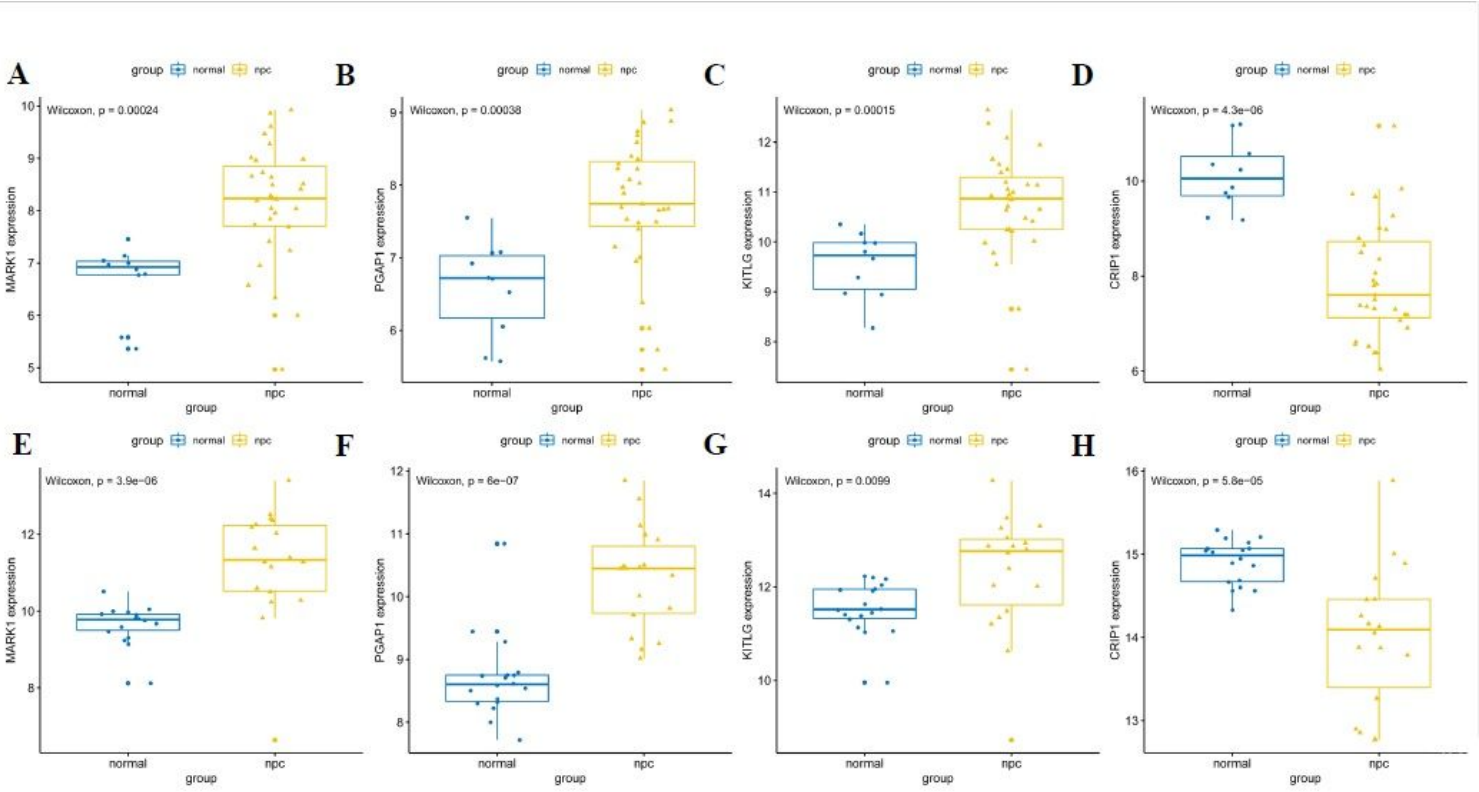

Figure 7

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.

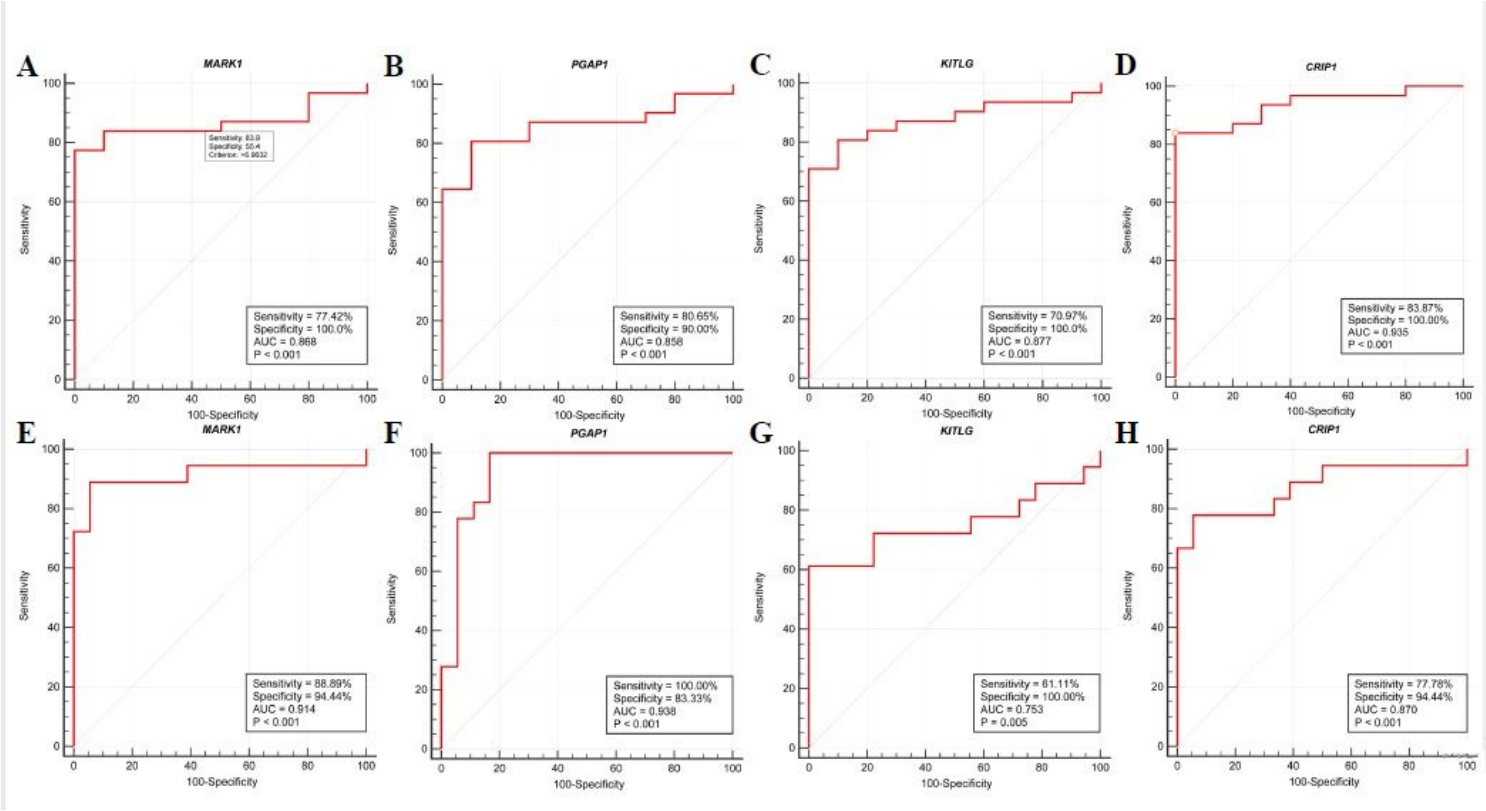

Figure 8

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.

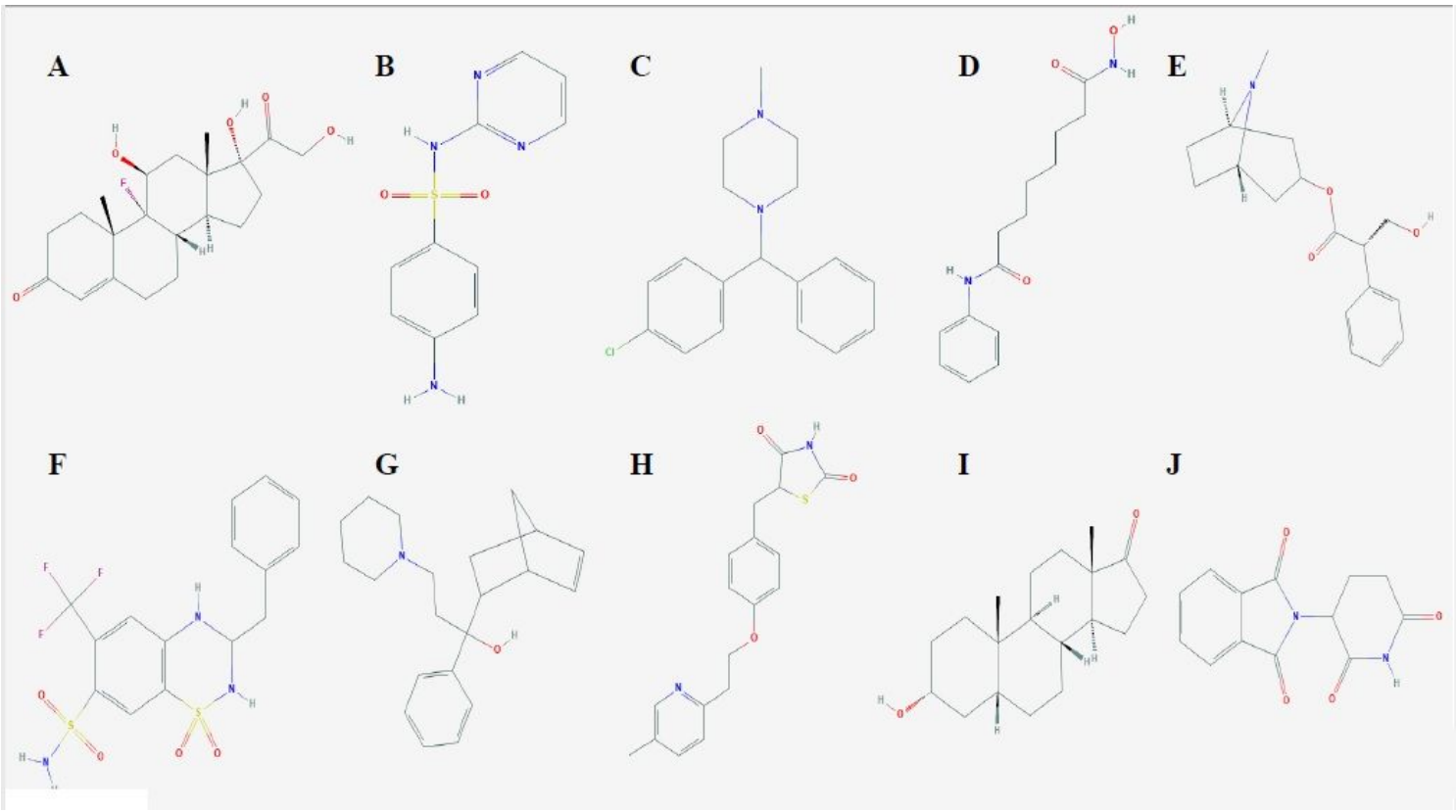

Figure 9

The 2D molecular structures of the top ten most significantsmall-molecule drugs. (A) Fludrocortisone. (B) Sulfadiazine.(C) Chlorcyclizine. (D) Vorinostat. (E) Hyoscyamine. (F)Bendroflumethiazide. (G) Biperiden. (H) Pioglitazone. (I) Etiocholanolone.(J) Thalidomide.