DNA methylation-based diagnostic and prognostic biomarkers of nasopharyngeal carcinoma patients

Zeng-hong Wu, MB[a,b], Tao Zhou, MB[a], Hai-Ying Sun, MD[a,c,*]

Abstract

Nasopharyngeal carcinoma (NPC) is the most common malignant tumor with a remarkable racial and geographical distribution including people in southern China, South East Asia, and the Middle East/North Africa. DNA methylation is an important manifestation of epigenetic modification, has been studied over several decades, and by regulating and controlling the expression of cancer-related genesis, abnormal DNA methylation can influence in a variety of human malignancy tumors.

Until now, there is no analysis focus on differentially methylated, differential expressed genes (MDEGs) study, so we make a joint analysis for both gene methylation profiling microarray and gene expression profiling microarray in NPC. Two gene expression datasets (GSE64634 and GSE12452) and gene methylation profiling data set (GSE62336) were downloaded from GEO and analyzed using the online tool GEO2R to identify MDEGs. Gene ontology (GO) functional analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the differentially methylated genes were performed. The STRING database was used to evaluate the interactions of MDEGs and to construct a protein–protein interaction (PPI) network using Cytoscape software. Hub genes were validated with the cBioPortal database.

The overlap among the 3 datasets contained 135 hypermethylation genes and 541 hypomethylation genes between NPC and non-NPC samples. A total of 4 genes (TROAP, PCOLCE2, HOXA4, and C1QB) in Hyper-LGs and 14 genes (DYNC1H1, LNX1, RAB37, ALDH3A1, SLC24A4, CP, CEP250, ANK2, DNAI2, MUC13, ACACB, GABRP, STX7, and TTC9) in Hypo-HGs were identified as hub genes.

The study of DNA methylation and gene expression provides us a strong support as well as new comprehensive information of MDEGs to the revelation of nasopharyngeal carcinoma's complex pathogenesis. However, further studies are needed to elucidate the biological function of these genes in NPC in the future.

Abbreviations: BP = biological processes, CC = cellular components, DEG = differentially expressed genes, GEO = gene expression omnibus, GO = gene ontology, KEGG = Kyoto encyclopedia of genes and genomes, MCODE = Molecular Complex Detection, MF = molecular functions, NPC = nasopharyngeal carcinoma, PPI = protein–protein interaction, STRING = search tool for the retrieval of interacting genes/proteins.

Keywords: bioinformatics, gene, methylation, nasopharyngeal carcinoma

1. Introduction

Nasopharyngeal carcinoma (NPC) is the most common malignant tumor with a remarkable racial and geographical distribution including people in southern China, South East Asia, and the Middle East/North Africa.[1,2] The main clinical manifestations of the patient are nasal congestion, blood stasis, ear blockage, hearing loss, vision, and headaches, among others. It is primarily a malignant tumor derived from nasopharyngeal epithelium located in the upper part of the nasopharyngeal cavity and on the side wall, with a strong tendency to metastasize.[3] Diagnosing the disease in the early needs a high index of clinical acumen and confirmation is only dependent on histology.[2] The potential risk

factors for NPC include Epstein–Barr virus (EBV) infection,[4] alcohol consumption, exposure to dust, formaldehyde, the function of genetic factors, and cigarette smoking.[5–8] Studies reported that ingestion of salted fish or other preserved foods was the most important part cause of NPC. Other studies found EBV infection is 90% to 100% of NPC cases in endemic regions.[9] EBV is associated with multiple types of human cancer, such as Burkitt lymphoma and Hodgkin disease, whereas in Asia it is closely association with NPC.

DNA methylation is an important manifestation of epigenetic modification, has been studied over several decades, and by regulating and controlling the expression of cancer-related genes, abnormal DNA methylation can influence in a variety of human malignant tumors. Recently epigenetic studies indicate that methylation can be of use as diagnostic biomarker and potential target for treatment.[10–14] Hypomethylation activates transcription of genes, whereas hypermethylation usually inhibits transcription of genes. CpG islands are lie in or near promoter regions of the genome, so aberrant methylation genes in CpG islands are often hypermethylated and may affect chromatic structure, upregulating or downregulating gene expression. Changes and malignant transformation of cells eventually lead to the formation of tumors.[15] About one-quarter of methylation alterations are significantly related to changes in the expression of tumor genes.[16] Zouridis et al[17] has been reported that 78% of the identified methylation expression were negative, consistent with DNA methylation used to silence local transcription. Therefore, the identification of differentially methylated, differential expressed genes (MDEGs) will be of great significance in clarifying the pathogenesis of NPC and in filtering biomarkers for diagnosis.

Many gene expression profiling analysis were introduced for differentially expressed genes (DEGs), whereas separated analysis of DEGs is limited.[18] Until now, there is no analysis focus on differentially methylated genes (DMGs) study, so we make a joint analysis for both gene methylation profiling microarray and gene expression profiling microarray in NPC. In this study, we used online bioinformatics resources to explore NPC-specific MDEGs. We identified NPC-related MDEGs, including hypomethylated, highly expressed genes (Hypo-HGs), and hypermethylated, lowly expressed genes (Hyper-LGs). Gene ontology and KEGG pathways involving the MDEGs help us understand the function of the MDEGs and we also explored the main hub nodes in protein–protein interaction (PPI) networks. Our results showed that the Hyper-HGs hub genes are *TROAP*, *PCOLCE2*, *HOXA4*, and *C1QB*; the Hypo-LGs hub gene is *DYNC1H1*, *LNX1*, *RAB37*, *ALDH3A1*, *SLC24A4*, *CP*, *CEP250*, *ANK2*, *DNAI2*, *MUC13*, *ACACB*, *GABRP*, *STX7*, and *TTC9*. The analyses provide comprehensive biological information for MDEGs, which may help to promote the understanding of the development and progression of NPC.

2. Materials and methods

2.1. Data resources

GEO (http://www.ncbi.nlm.nih.gov/geo)[15] is a public functional genomics data repository which included throughout gene expression data, chips and microarrays. Two gene expression datasets (GSE64634[19] and GSE12452[20]) and gene methylation profiling data set (GSE62336) were downloaded from GEO (GPL13534 Illumina HumanMethylation450 BeadChip and Affymetrix GPL570 platform, Affymetrix Human Genome

U133 Plus 2.0 Array). Observing the download of GSE64634 database included 14 NPC samples and 4 normal samples; GSE12452 contained 31 NPC samples and 10 noncancerous samples; GSE62336 dataset contained 25 NPC tissue samples and 25 noncancerous samples. Ethical approval was not necessary for this study because our study is bioinformatic analysis.

2.2. Identification of differentially expressed genes

The identification of DEGs and DMGs between NPC and noncancerous samples was performed using GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) with the criteria of $P < .05$ and $|t| > 2$. GEO2R is an online tool designed that allows users to compare different datasets in a GEO series for identify DEGs across experimental conditions. To correct the limitations of false-positives, we used Benjamini and Hochberg False Discovery Rate method.[21] Finally, we acquired hypomethylation-high expression genes (Hypo-HGs) after superimposition of upregulated and hypomethylation genes and acquired hypermethylation-low expression genes (Hyper-LGs) after superimposition of downregulated and hypermethylation genes.

2.3. KEGG and GO enrichment analyses of DMGs

The Enrichr (https://amp.pharm.mssm.edu/Enrichr/) (version 6.8)[22] which is a useful online platform database that integrates biological data and provides a comprehensive set of functional annotation information of genes as well as proteins for users to analyze the functions or signaling pathways. The Kyoto Encyclopedia of Genes and Genomes (KEGG)[23] is a database resource for understanding high-level gene functions and linking genomic information from large-scale molecular datasets. Gene ontology (GO)[24] function analysis (biological processes [BPs], cellular components [CCs], and molecular functions [MFs]) is a powerful bioinformatics tool to analyze BP and annotate genes. To analyze the function of the identified DMGs, biological analyses were performed using GO enrichment and KEGG pathway analysis via Enrichr online database. $P < .05$ as the cutoff criterion considered statistically significant.

2.4. PPI network construction and module analysis

Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org)[25] online database was used to predict the PPI network information. Analyzing the interactions and functions between DMGs may provide information about the mechanisms of generation and development of disease (PPI score >0.4). Cytoscape (version 3.7.1) is a bioinformatics platform for constructing and visualizing molecular interaction networks.[26] The plug-in Molecular Complex Detection (MCODE) of Cytoscape was applied to detect densely connected regions in PPI networks. The PPI networks were constructed using Cytoscape and the most significant module in the PPI networks was selected using MCODE. The criteria for selection were set as follows: Max depth=100, degree cut-off=2, Node score cut-off=0.2, MCODE scores >5 and K-score=2.

2.5. Hub genes selection and analysis

A network of the integrative relationships of the hub genes and their co-expression genes clinical characteristics in NPC was analyzed using cBioPortal for Cancer Genomics (http://www.
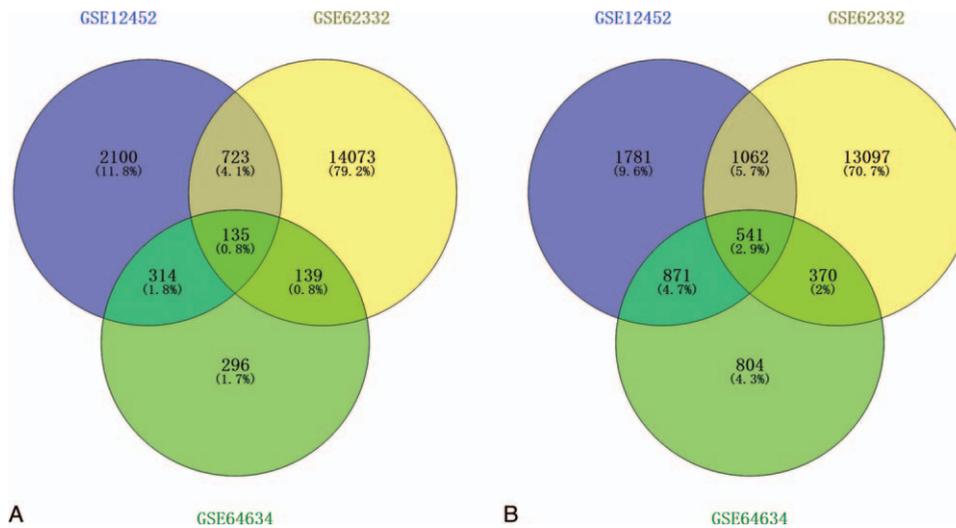
Figure 1. Abnormal methylation of expression genes identified in data sets including GSE64634 and GSE12452 for gene expression, and GSE62336 for gene methylation. (A) Highly expressed genes with low methylation. (B) Low expressed genes with high methylation.

cbioportal.org/),[27] which is an open-access resource for analyzing and exploring genetic alterations from multidimensional studies samples. The analyses of genomic mutations in the selected TCGA datasets could be analyzed in the cBioPortal online according to the instructions.

3. Results

3.1. Identification of DMGs in NPC

After standardization of the microarray results, DEGs and DMGs were identified. The overlap among the 3 datasets contained 135 hypermethylation genes and 541 hypomethylation genes between NPC and non-NPC samples the results as shown in the Venn diagram (Fig. 1).

3.2. GO enrichment and KEGG analyses of DMGs

To further investigate the biological functions and mechanisms of the DMGs, functional and pathway enrichment analyses were performed using Enrichr tool. Hyper-LGs GO analysis results showed that changes in BPs of DMGs were significantly enriched in sensory organ morphogenesis, polyol biosynthetic process, oxygen homeostasis, among others. Changes in MF were mainly enriched in nucleoside kinase activity, DNA *N*-glycosylase activity, and so on. Changes in CC of DMGs were mainly enriched in the axonal growth cone, microtubule plus-end, ribonucleoprotein granule, and so on. KEGG pathway analysis revealed that the DMGs were mainly enriched in protein digestion and absorption, ECM-receptor interaction, PI3K-Akt signaling pathway, and so on (Fig. 2).

Figure 2. Gene ontology (GO) functional analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the Hyper-LGs. (A) Biological processes (BP). (B) Molecular functions (MF). (C) Cellular components (CC). (D) KEGG pathway.
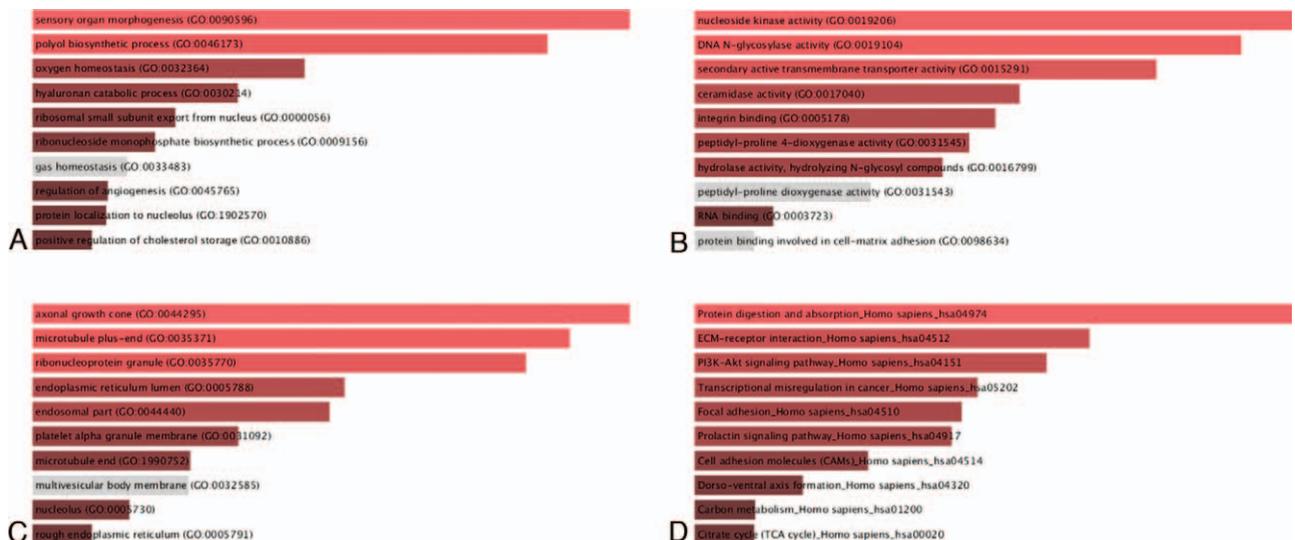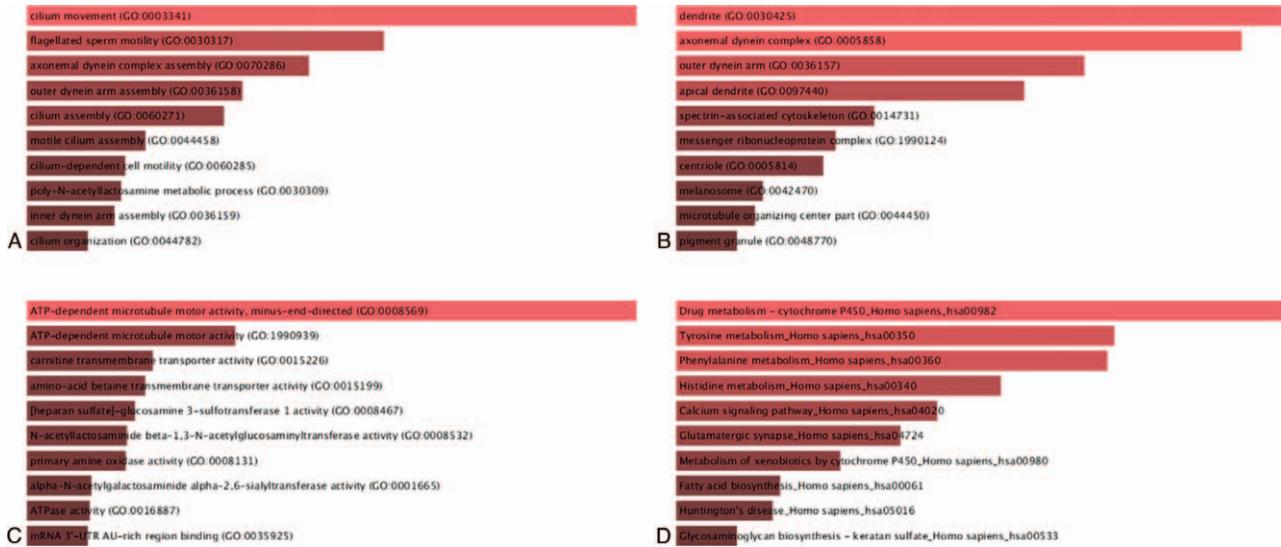
Figure 3. Gene ontology (GO) functional analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the Hypo-HGs. (A) Biological processes (BP). (B) Molecular functions (MF). (C) Cellular components (CC). (D) KEGG pathway.

Hypo-HGs GO analysis results showed that changes in BPs of DMGs were significantly enriched in cilium movement, flagellated sperm motility, axonemal dynein complex assembly, and so on. Changes in MF were mainly enriched in ATP-dependent microtubule motor activity, minus-end-directed, carnitine transmembrane transporter activity, and so on,. Changes in cell component (CC) of DMGs were mainly enriched in the dendrite, axonemal dynein complex, outer dynein arm, and so on. KEGG pathway analysis revealed that the DMGs were mainly enriched in drug metabolism-cytochrome P450, tyrosine metabolism, phenylalanine metabolism, and so on (Fig. 3).

3.3. PPI network construction and module analysis

To further explore the connection between Hyper-LGs and Hypo-HGs at the protein level, the PPI networks were constructed based on the interactions of DMGs (Fig. 4) and the most significant module was obtained using Cytoscape (Fig. 5). A total of 157 interactions and 95 nodes in Hyper-LGs and a total of 964 interactions and 434 nodes in Hypo-HGs were screened to establish the PPI network and the biological functional analyses of genes involved in this most significant module were analyzed using Enrichr tool. The most significant module KEGG results and GO analysis results showed in Tables 1 and 2.
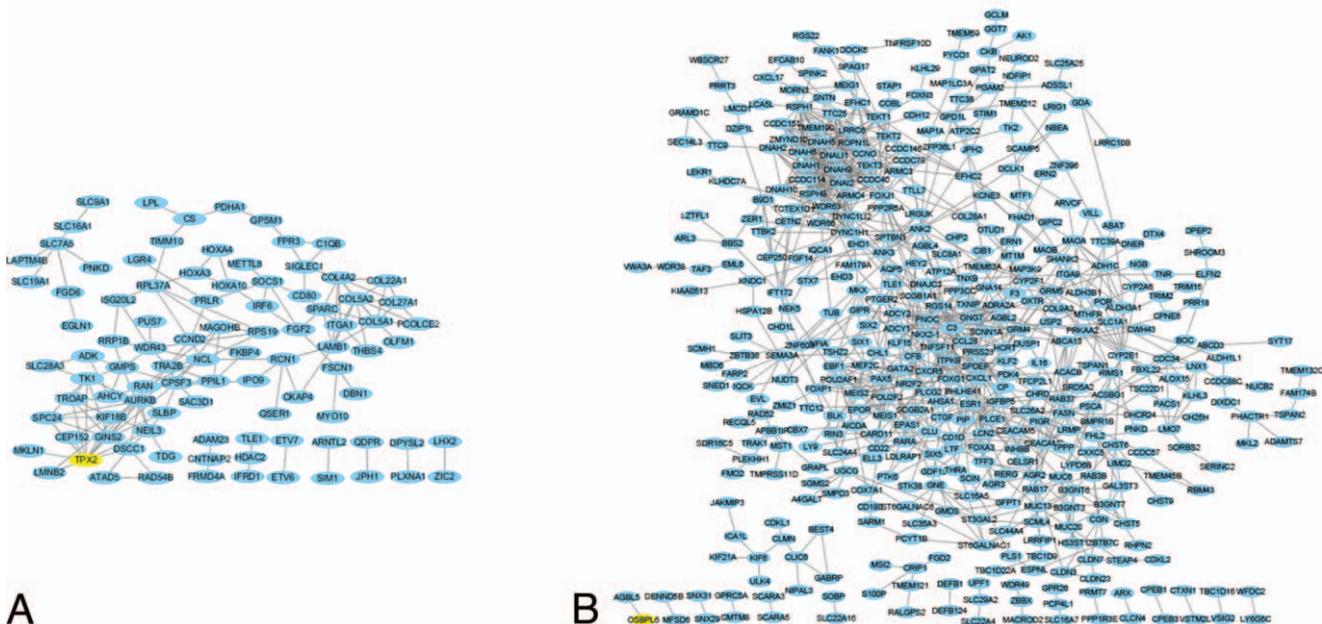

Figure 4. The protein–protein interaction (PPI) network of differentially methylated genes was constructed using Cytoscape. (A) hypermethylated, lowly expressed genes (Hyper-LGs). (B) hypomethylated, highly expressed genes (Hypo-HGs).
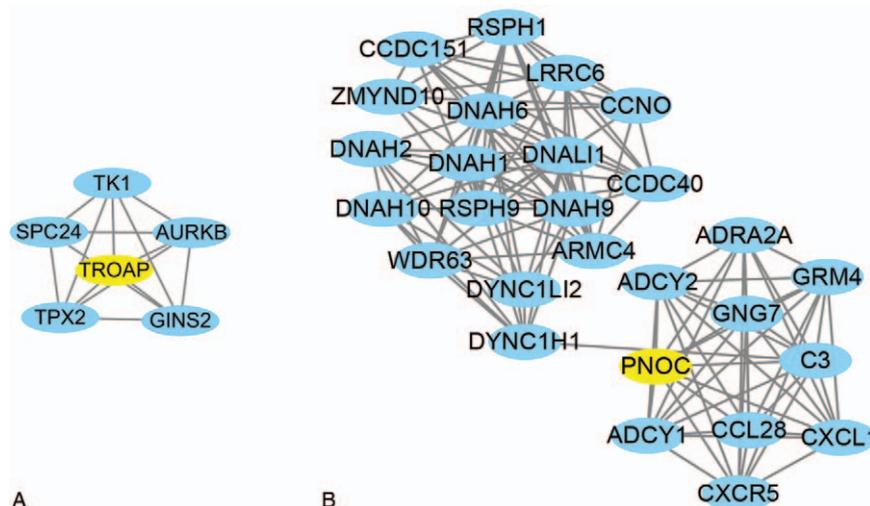
Figure 5. The protein–protein interaction (PPI) network of the most significant module was obtained using Cytoscape. (A) hypermethylated, lowly expressed genes (Hyper-LGs). (B) hypomethylated, highly expressed genes (Hypo-HGs).

3.4. Hub gene selection and analysis

A total of 4 genes (*TROAP*, *PCOLCE2*, *HOXA4*, and *C1QB*) in Hyper-LGs and 14 genes (*DYNC1H1*, *LNX1*, *RAB37*, *ALDH3A1*, *SLC24A4*, *CP*, *CEP250*, *ANK2*, *DNAI2*, *MUC13*, *ACACB*, *GABRP*, *STX7*, and *TTC9*) in Hypo-HGs were identified as hub genes with degrees ≥10. The full names, abbreviations, also known as, and functions for these hub genes are shown in Tables 3 and 4. A network of the hub genes and their co-expression genes were performed via cBioPortal online platform (Fig. 6).

4. Discussion

It is known that NPC located at the head of the head and neck cancer poses a great threat to human health, but there has been

limited research on the pathogenesis of NPC. Tumor is a product of epigenetic, cumulative genetic, somatic, and endocrine aberrations.[28] DNA methylation can alter the expression of genes and provided a novel idea to understand the pathogenesis of malignant tumor.[29] DNA methylation aberrations are more common contrast to genomic aberrations in the cancer genome. The rapid development of high-throughput sequencing technologies and microarray has allowed us observed thousands of genes methylation levels simultaneously in the human genome, so we can investigate the key genes affected by methylation now. In our study, we identified 135 Hyper-LGs as well as 541 Hypo-HGs that may be involved in molecular regulation with the development of NPC.

The DMGs GO analysis results showed that the main MFs of the Hyper-LGs were nucleoside kinase activity and DNA

Table 1

GO and KEGG pathway enrichment analysis of Hyper-LGs in the most significant module.

Term	Description	Count in gene set	*P*
GO: MF (top 5)			
GO:0004797	Thymidine kinase activity	1	6.430E-4
GO: 0061676	Importin-alpha family protein binding	1	9.643E-4
GO: 0019136	Deoxynucleoside kinase activity	1	1.286E-3
GO: 0035174	Histone serine Kinase activity	1	2.570E-3
GO: 0043138	3'-5' DNA helicase activity	1	4.173E-3
GO: CC (top 5)			
GO:0005876	Spindle microtubule	2	1.450E-4
GO: 0000779	condensed Chromosome, centromeric region	2	4.444E-4
GO:0000776	Kinetochore	2	6.191E-4
GO: 0000922	Spindle pole	2	7.117E-4
GO:0000811	GINS complex	1	9.441E-4
GO: BP (top 5)			
GO:1903047	Mitotic cell cycle process	4	8.585E-5
GO: 0000278	Mitotic cell cycle	4	1.209E-4
GO:1902850	Microtubule cytoskeleton organization involved in mitosis	2	1.419E-4
GO: 0090307	Mitotic spindle assembly	2	1.419E-4
GO:1902292	Cell cycle DNA replication initiation	1	3.223E-4
KEGG		*z* score	
hsa00240	Pyrimidine metabolism	−1.76	<.05
hsa00983	Drug metabolism—other enzymes	−1.89	<.05

BP = biological process, CC = cellular component, GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular function, NPC = nasopharyngeal carcinoma.

Table 2

GO and KEGG pathway enrichment analysis of Hypo-HGs in the most significant module.

Term	Description	Count in gene set	*P*
GO: MF (top 5)			
GO:0003777	Microtubule motor activity	8	1.205E-13
GO: 0003774	Motor activity	8	1.003E-11
GO: 0017111	Nucleoside-triphosphatase activity	8	1.030E-5
GO: 0016462	Pyrophosphatase activity	8	1.506E-5
GO: 0016818	Hydrolase activity, acting on acid anhydrides	8	1.532E-5
GO: CC (top 5)			
GO:0097014	Ciliary plasm	10	7.024E-17
GO: 0005930	Axoneme	10	7.024E-17
GO:0030286	Dynein complex	8	1.073E-15
GO: 0044441	Ciliary part	11	7.530E-13
GO:0005929	Cilium	12	2.619E-12
GO: BP (top 5)			
GO:0003341	Cilium movement	10	1.723E-19
GO: 0007018	Microtubule-based movement	13	2.528E-18
GO:0007017	Microtubule-based process	15	1.674E-15
GO: 0035082	Axoneme assembly	8	4.480E-15
GO:0001578	Microtubule bundle formation	8	1.516E-13
KEGG (Top 5)		*z* score	
hsa04062	Chemokine signaling pathway	−1.93	1.550e-7
hsa05016	Huntington disease	−1.88	1.868e-7
hsa04724	Glutamatergic synapse	−1.88	<.05
hsa04727	GABAergic synapse	−1.74	<.05
hsa04713	Circadian entrainment	−1.77	<.05

BP = biological Process, CC = cellular Component, GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular function, NPC = nasopharyngeal carcinoma.

Table 3

Functional roles of Hyper-HGs hub genes with degree ≥10.

Gene symbol	Full name	Also known as	Function
TROAP	Trophinin-associated protein	TASTIN	Cell adhesion molecule; enhances malignancy but promotes tumor development
PCOLCE2	Procollagen C-endopeptidase enhancer 2	PCPE2	Adult heart and strong expression primarily in nonossified cartilage in developing tissues
HOXA4	Homeobox A4	HOX1; HOX1D	Regulate gene expression, morphogenesis, and differentiation
C1QB	Complement C1q B chain	NA	Glomerulonephritis and lupus erythematosus.

Table 4

Functional roles of 14 Hypo-HGs hub genes with degree ≥10.

| Gene symbol	Full name	Also known as	Function
DYNC1H1	Dynein cytoplasmic 1 heavy chain 1	p22; DHC1; DNCL; DYHC; HL-3	Motor neuron axon degenerative diseases
LNX1	Ligand of numb-protein X 1	LNX; MPDZ; PDZRN2	Junction adhesion molecule; protein interactions and signal transduction
RAB37	RAB37, member RAS oncogene family	NA	GTPases that are critical regulators of vesicle trafficking
ALDH3A1	Aldehyde dehydrogenase 3 family member A1	ALDH3; ALDHIII	Protect cells from injury
SLC24A4	Solute carrier family 24 member 4	AI2A5; NCKX4; SHEP6; SLC24A2	Risk of late-onset Alzheimerdisease; enamel maturation
CP	Ceruloplasmin	CP-2	Aceruloplasminemia; diabetes; neurologic abnormalities
CEP250	Centrosomal protein 250	CEP2; CNAP1; C-NAP1; CRDHL2	Encodes a core centrosomal protein required for centriole-centriole cohesion
ANK2	Ankyrin 2	LQT4; ANK-2; brank-2	Na/Ca exchanger 1 in Cardiomyocytes
DNAI2	Dynein axonemal intermediate chain 2	DIC2; CILD9	Associated with primary ciliary dyskinesia type 9
MUC13	Mucin 13, cell surface associated	DRCC1; MUC-13	Secreted and cell surface glycoproteins expressed by ductal and glandular epithelial tissues
ACACB	Acetyl-CoA carboxylase beta	ACC2; ACCB; HACC275	Rate-limiting step in fatty acid uptake and oxidation by mitochondria
GABRP	Gamma-aminobutyric acid type A receptor pi subunit	NA	Expressed in several non-neuronal tissues including the uterus and ovaries
STX7	Syntaxin 7	NA	A syntaxin family membrane receptor involved in vesicle transport
TTC9	Tetratricopeptide repeat domain 9	TTC9A	Hormonally regulated in breast cancer cells and may play a role in cancer cell invasion and metastasis
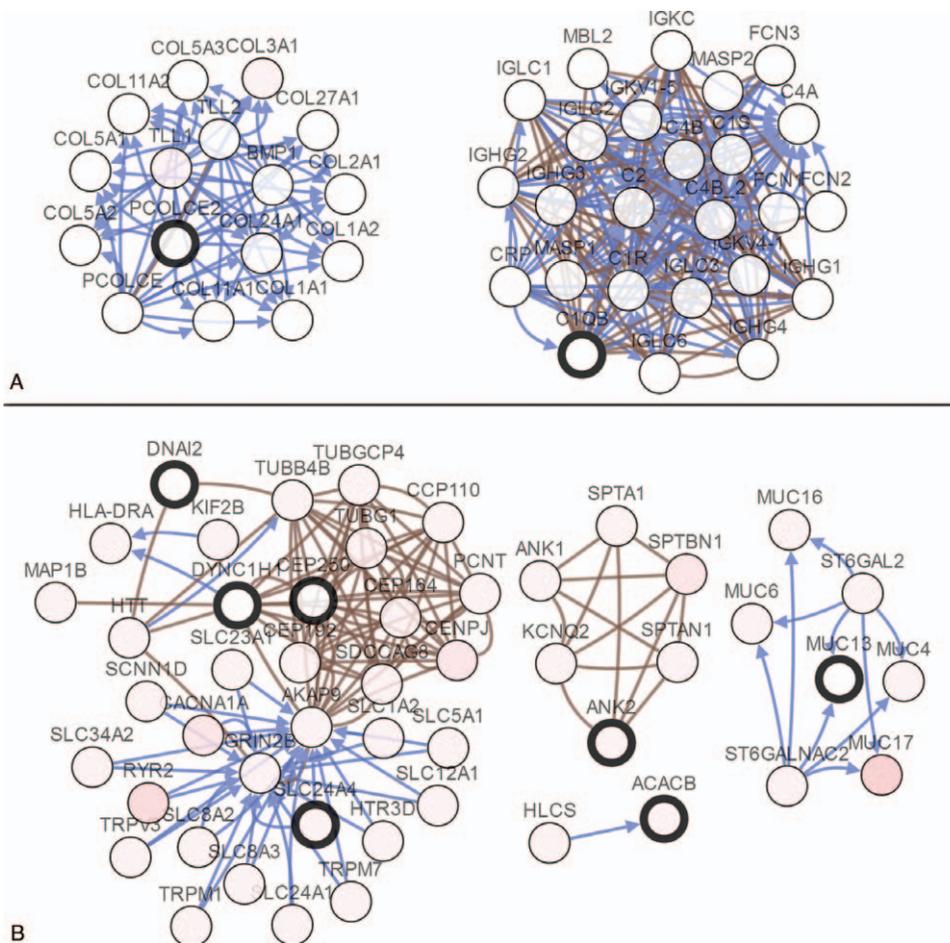
Figure 6. Interaction network and biological process analysis of the hub genes. Hub genes and their co-expression genes were analyzed using cBioPortal. Nodes with bold black outline represent hub genes. Nodes with thin black outline represent the co-expression genes. (A) hypermethylated, lowly expressed genes (Hyper-LGs). B hypomethylated, highly expressed genes (Hypo-HGs).

N-glycosylase activity indicating that hypermethylation mainly affected enzyme activity. In addition, the functions of the Hypo-HGs BPs of DMGs were significantly enriched in cilium movement, flagellated sperm motility, and axonemal dynein complex assembly suggesting that nasal mucociliary dysfunction may be an important reason for NPC development. Hyper-LGs KEGG pathway analysis revealed that the DMGs were mainly enriched in protein digestion and absorption, ECM-receptor interaction, PI3K-Akt signaling pathway suggesting that the disorder of regulation of various cellular functions such as cell proliferation, differentiation, apoptosis, and glucose transport could promote the evolution of tumors. However, Hypo-HGs KEGG pathway analysis revealed that the DMGs were mainly enriched in drug metabolism-cytochrome P450, tyrosine metabolism, phenylalanine metabolism prompting that hypomethylated more involvement in the regulation of metabolic pathways.

The PPI network showed the functional connectivity of the hypermethylation-low expression genes, among which the hub genes were *TROAP*, *PCOLCE2*, *HOXA4*, and *C1QB*. *TROAP* is involved with trophinin and bystin in cell adhesion molecule, and the trophinin–cell-adhesion molecule complex mediates an initial attachment of the blastocyst to uterine epithelial cells.[30,31] Some studies reported that TROAP expression not only enhances malignancy but also promotes tumor development in colorectal

cancer, ovarian adenocarcinomas, breast cancer, bladder urothelial carcinoma, hepatocellular carcinoma, and other cancers.[32,33] Thus, TROAP may be a potential biomarker may be used to predict cancer prognosis and sensitivity to cancer treatment. Procollagen C-proteinase enhancer 2 (PCOLCE2) protein confirmed as a differentially expressed epithelial transcript.[34] PCOLCE2 at high levels in the adult heart and strong expression primarily in nonossified cartilage in developing tissues.[35] HOXA4 is part of the A cluster on chromosome 7 and encodes a DNA-binding transcription factor, which may regulate gene expression, morphogenesis, and differentiation. Accumulated evidence has indicated the abnormal expression HOXA4 contributions to carcinogenesis.[36] HOXA4 is reportedly overexpressed in epithelial ovarian cancer and colorectal cancer. Study further reported that HOXA4 suppresses migration via β1 integrin in ovarian cancer cell lines.[37–39] C1QB deficiency is associated with glomerulonephritis and lupus erythematosus.

The PPI network showed the functional connectivity of the selected hypomethylation-high expression genes, among which the first 5 hub genes were *DYNC1H1*, *LNX1*, *RAB37*, *ALDH3A1*, and *SLC24A4*. DYNC1H1 encodes a heavy chain of cytoplasmic dynein for axonal transport[40] and involvement in development of motor neuron axon degenerative diseases.[41] Further studies proved the association of CpG island methylation

level in *DYNC1H1* and spinal muscular atrophy severity was substantiated. *LNX1* encodes a membrane-bound protein that is involved in protein interactions and signal transduction and may play an indispensable role in tumorogenesis. LNX1 not only interacts with and ubiquitinates c-Src kinase but also facilitates the endocytosis of junction adhesion molecule.[42,43]*RAB37* belong to member RAS oncogene family that are critical regulators of exocytosis of secreted glycoproteins.[44] Patients with preserved RAB37 protein expression were related to suppress cancer metastasis and better prognosis. *ALDH3A1* are most abundant in stem cells as a cytoprotective aldehyde dehydrogenases family members, which protect cells from injury.[45] SLC24A4 encodes a member of the potassium-dependent sodium/calcium exchanger protein family. SLC24A4 was shown to be closely related to the risk of late-onset Alzheimer disease[46] and reveals that enamel maturation is dependent upon SLC24A4 function.[47]

The study of DNA methylation and gene expression provides us a strong support as well as new comprehensive information of MDEGs to the revelation of NPC's complex pathogenesis. As for abnormal methylation regions often contributed to changes in gene expression, this study indirectly reflects the indispensable role of abnormal methylation in the pathogenesis of NPC. A total of 4 genes (*TROAP*, *PCOLCE2*, *HOXA4*, and *C1QB*) in Hyper-LGs and 14 genes (*DYNC1H1*, *LNX1*, *RAB37*, *ALDH3A1*, *SLC24A4*, *CP*, *CEP250*, *ANK2*, *DNAI2*, *MUC13*, *ACACB*, *GABRP*, *STX7*, and *TTC9*) in Hypo-HGs were selected as effective biomarkers by a series of the advanced data analysis. Moreover, a set of hub genes in network modules related to MDEGs were identified. These genes may have potential value as methylation-based biomarkers for the diagnosis and treatment of NPC. However, further studies are needed to elucidate the biological function of these genes in this type of cancer in the future.

Author contributions

W.Z.H. Z.T. designed and analyzed the research study; W.Z.H. and S.H.Y. wrote and revised the manuscript.
Formal analysis: Zeng-hong Wu.
Investigation: Zeng-hong Wu.
Methodology: Zeng-hong Wu.
Writing – original draft: Zeng-hong Wu.
Writing – review & editing: Zeng-hong Wu.

References

[1] Chua MLK, Wee JTS, Hui EP, et al. Nasopharyngeal carcinoma. Lancet 2016;387:1012–24.

[2] Yu MC. Nasopharyngeal carcinoma: epidemiology and dietary factors. IARC Sci Publ 1991;105:39–47.

[3] Kamran SC, Riaz N, Lee N. Nasopharyngeal carcinoma[J]. Surg Oncol Clin North Am 2015;24:547–61.

[4] Yu MC. Nasopharyngeal carcinoma: epidemiology and dietary factors. 1991;IARC Sci Publ, 39-47.

[5] Tsao SW, Tsang CM, To KF, et al. The role of Epstein-Barr virus in epithelial malignancies[J]. J Pathol 2015;235:323–33.

[6] 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;36:435–47.

[7] Henderson BE, Louie E, SooHoo Jing J, et al. Risk factors associated with nasopharyngeal carcinoma. N Engl J Med 1976;295:1101–6.

[8] 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;84:619–22.

[9] Ho Y, Tsao SW, Zeng M, et al. STAT3 as a therapeutic target for epstein-barr virus (EBV): associated nasopharyngeal carcinoma. Cancer Lett 2013;330:141–9.

[10] Jones PA. DNA methylation and cancer. Cancer Res 1986;46:461–6.

[11] Kulis M, Esteller M. DNA methylation and cancer. Adv Genet 2010;70:27–56.

[12] Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 2003;3:253–66.

[13] Rakyan VK, Down TA, Balding DJ, et al. Epigenome-wide association studies for common human diseases. Nat Rev Genet 2011;12:529.

[14] Lo KW, Huang DP. Genetic and epigenetic changes in nasopharyngeal carcinoma. Semin Cancer Biol 2002;12:451–62.

[15] Wang Qian, Wang Gang, Liu Chaoxu, et al. Prognostic value of CpGnisland methylator phenotype among hepatocellular carcinoma patients: a systematic review and meta-analysis. Int J Surg 2018;54:92–9.

[16] Li H, Liu JW, Liu S, et al. Bioinformatics-based identification of methylated-differentially expressed genes and related pathways in gastric cancer[J]. Dig Dis Sci 2017;62:3029–39.

[17] Zouridis H, Deng N, Ivanova T, et al. Methylation subtypes and large-scale epigenetic alterations in gastric cancer. Sci Transl Med 2012;4:156ra140.

[18] Ge Y, He Z, Xiang Y, et al. The identification of key genes in nasopharyngeal carcinoma by bioinformatics analysis of high-through-put data[J]. Mol Biol Rep 2019;46:2829–40.

[19] Bo H, Gong Z, Zhang W, 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.

[20] Dodd LE, Sengupta S, Chen IH, et al. Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev 2006;15:2216–25.

[21] Benjamini Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 1995;57:289–300.

[22] Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013;128:

[23] Altermann E, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. BMC Genomics 2005;6:60.

[24] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.

[25] Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: proteinprotein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447–52.

[26] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.

[27] Jianjiong G, Bülent Arman A, Ugur D, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6:pl1.

[28] Xu Z, Wang Z, Jia X, et al. MMGZ01, an anti-DLL4 monoclonal antibody, promotes nonfunctional vessels and inhibits breast tumor growth. Cancer Lett 2016;372:118–27.

[29] Gyorffy B, Bottai G, Fleischer T, et al. Aberrant DNA methylation impacts gene expression and prognosis in breast cancer subtypes. Int J Cancer 2016;138:87–97.

[30] Nadano D, Nakayama J, Matsuzawa S, et al. Human tastin, a proline-rich cytoplasmic protein, associates with the microtubular cytoskeleton. Biochem J 2002;364:669–77.

[31] Suzuki N, Zara J, Sato T, et al. A cytoplasmic protein, bystin, interacts with trophinin, tastin, and cytokeratin and may be involved in trophinin-mediated cell adhesion between trophoblast and endometrial epithelial cells. Proc Natl Acad Sci U S A 1998;95:5027–32.

[32] Partheen K, Levan K, Osterberg L, et al. Four potential biomarkers as prognostic factors in stage III serous ovarian adenocarcinomas. Int J Cancer 2008;123:2130–7.

[33] Godoy H, Mhawech-Fauceglia P, Beck A, et al. Developmentally restricted differentiation antigens are targets for immunotherapy in epithelial ovarian carcinoma. Int J Gynecol Pathol 2013;32:536–40.

[34] Steiglitz BM, Keene DR, Greenspan DS. PCOLCE2 encodes a functional procollagen C-proteinase enhancer (PCPE2) that is a collagenbinding protein differing in distribution of expression and post-translational modification from the previously described PCPE1. J Biol Chem 2002;277:49820–30.

[35] Steiglitz BM. PCOLCE2 encodes a functional procollagen C-proteinase enhancer (PCPE2) that is a collagen-binding protein differing in distribution of expression and post-translational modification from the previously described PCPE1[J]. J Biol Chem 2002;277:49820–30.

[36] Cheng S, Qian F, Huang Q, et al. HOXA4, down-regulated in lung cancer, inhibits the growth, motility and invasion of lung cancer cells[J]. Cell Death Dis 2018;9:465.

[37] Bhatlekar S, Addya S, Salunek M, et al. Identification of a developmental gene expression signature, including HOX genes, for the normal human colonic crypt stem cell niche: overexpression of the signature parallels stem cell overpopulation during colon tumorigenesis. Stem Cells Dev 2014;23:167–79.

[38] Yamashita T, Tazawa S, Yawei Z, et al. Suppression of invasive characteristics by antisense introduction of overexpressed HOX genes in ovarian cancer cells. Int J Oncol 2006;28:931–8.

[39] Klausen C, Leung PC, Auersperg N. Cell motility and spreading are suppressed by HOXA4 in ovarian cancer cells: possible involvement of beta1 integrin. Mol Cancer Res 2009;7:1425–37.

[40] Hafezparast M, Klocke R, Ruhrberg C, et al. Mutations in dynein link motor neuron degeneration to defects in retrograde transport. Science 2003;300:808–12.

[41] Scoto M, Rossor AM, Harms MB, et al. Novel mutations expand the clinical spectrum of DYNC1H1-associated spinal muscular atrophy. Neurology 2015;84:668–79.

[42] Weiss A, Baumgartner M, Radziwill G, et al. c-Src is a PDZ interaction partner and substrate of the E3 ubiquitin ligase Ligand-of-Numb protein X1. FEBS Lett 2007;581:5131–6.

[43] Kansaku A, Hirabayashi S, Mori H, et al. Ligand-of numb protein X is an endocytic scaffold for junctional adhesion molecule 4. Oncogene 2006;25:5071–84.

[44] Tzeng HT, Li TH, Tang YA, et al. Phosphorylation of Rab37 by protein kinase C alpha inhibits the exocytosis function and metastasis suppression activity of Rab37[J]. Oncotarget 2017;8: 108556–70.

[45] Ma I, Allan AL. The role of human aldehyde dehydrogenase in normal and cancer stem cells. Stem Cell Rev 2011;7:292–306.

[46] Lu H, Zhu XC, Wang HF, et al. Lack of association between SLC24A4 polymorphism and late-onset Alzheimer's disease in Han Chinese[J]. Curr Neurovasc Res 2016;13:

[47] Wang S, Choi M, Richardson AS, et al. STIM1 and SLC24A4 Are Critical for Enamel Maturation. J Dent Res 2014;93(7 Suppl): 94S–100S.