Identification of Key Genes for Nasopharyngeal Carcinoma by Gene Expression and DNA Methylation Data Integration Analysis

zengyan zong
Shenzhen Luohu Hospital Group Luohu People's Hospital
https://orcid.org/0000-0003-2332-4211

dayang chen
Shenzhen Luohu Hospital Group Luohu People's Hospital

WEI WU
Shenzhen Luohu Hospital Group Luohu People's Hospital

xiaowen dou
Shenzhen Luohu Hospital Group Luohu People's Hospital

mengmeng wang
Shenzhen Luohu Hospital Group Luohu People's Hospital

xiang ji
Shenzhen Luohu Hospital Group Luohu People's Hospital

xiuming zhang (✉ zhangxiuming0760@163.com)
Medical Laboratory of the Third affiliated hospital of Shenzhen university, Shenzhen, 518001, China;

dan xiong
Shenzhen Luohu Hospital Group Luohu People's Hospital

Research

Keywords: nasopharyngeal carcinoma, gene expression, DNA methylation, GEO database

DOI: https://doi.org/10.21203/rs.3.rs-96113/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: The pathogenesis of Nasopharyngeal carcinoma (NPC) is very complicated. The present study aimed to identify some candidate genes as biomarkers for NPC diagnosis and pathogenesis.

Methods: Three Microarray datasets GSE53819, GSE64634 and GSE12452 and a methylation array (GSE52068) were re-analyzed. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the differentially expressed genes (DEGs) were applied. STRING software was used to construct a protein-protein interaction (PPI) network of DEGs and visualized by Cytoscape. Random Forest (RF) algorithm was performed to construct classifiers and identified key genes.

Results: A total of 91 DEGs were screened from the three datasets. GO term and KEGG pathway analysis suggested that the DEGs were predominantly enriched in drug metabolism-cytochrome P450 pathway, metabolism of xenobiotics by cytochrome P450 pathway, chemical carcinogenesis pathway, ciliary part, motile cilium, axoneme, microtubule and ciliary plasm. We obtained nine hub genes and one significant module. We constructed a classifier based on DEGs and found CLIC6 and CLU have the best classification ability. Finally, five hypermethylated and downregulated genes (hyper-down) were identified by integrating methylation data.

Conclusions: With gene expression and methylation data integration analysis, several key genes were identified may be potential biomarkers for NPC diagnosis and pathogenesis.

Background

Nasopharyngeal carcinoma (NPC) is the most common squamous cell carcinoma arising from nasopharynx and is characterised by its high incidence in distinct geographic and ethnic populations. The top 3 high national incidence rates are estimated to be in Malaysia, Indonesia, and Singapore. In particular, the NPC incidence rates observed in South China, including Guangdong province and Hong Kong were significantly higher than in other regions. A number of NPC patients tend to develop into an advanced stage of disease before were diagnosed with NPC. A new report reported that the 5-year overall survival of advanced NPC patients was 50.0%. Radiotherapy is still a major treatment for nonmetastatic disease and approximate 30% of cases will develop local or distant metastasis recurrence. Although, previous literatures revealed that genetic, age, Epstein-Barr virus infection and certain non-coding RNAs (circRNA and microRNA) are well associated with NPC, the molecular mechanism regulating the activation and progression of NPC is still unclear. Therefore, it is necessary to identify key genes for NPC pathogenesis.

The occurrence of cancer is the result of a series of molecular changes that occur in cells, which leads to a large number of changes in gene expression. For example, numerous studies have indicated that differentially expressed genes (DEGs) frequently associated with NPC. Abnormally expressed c9orf24,
PCDP1, and LRRC46 may serve as diagnostic and therapeutic markers for NPC\textsuperscript{10}. TSPAN8 were confirmed to highly expressed in NPC tissues and to relate with poor survival in NPC patients\textsuperscript{11}. Furthermore, the changes of dehydrogenase member 2 (DHRS2)\textsuperscript{12}, ribosomal protein genes (uS8, uS4, eS31, and uL14)\textsuperscript{12}, and homeobox protein hox-c6 (HOXC6)\textsuperscript{13} were proved to involve in the development, metastasis and prognosis of NPC.

Abnormal DNA methylation can lead to differential genes expression, therefore, the alterations of DNA methylation and genes may affect the development of NPC. DNA methylation is a common epigenetic alteration that happens exclusively on cytosine nucleotides and a majority of in the background of CpG islands (specific gene regions 5′—C—phosphate—G—3′). The changes occurred in modification pattern ultimately not only influence cell predisposition and tumour phenotype, but promote the malignancy of tumor\textsuperscript{14,15}. For example, a report found that the HOP homeobox HOPX showed a highest methylation level and lead to poor clinical outcomes in NPC patients\textsuperscript{16}. The tumor suppressor gene ten-eleven translocation methyl cytosine dioxygenase1 (TET1) can facilitate the shifts of 5-methylcytosine to 5-hydroxy methyl cytosine in NPC cells\textsuperscript{17}. These findings indicated that the abnormal methylated and differentially expressed genes may serve as potential biomarkers for cancer treatment and prediction\textsuperscript{18}.

In our study, we screened 91 DEGs for further investigated. We performed PPI network construction, enrichment analysis and constructed a classifier based on gene expression data. Finally, we analyzed differentially methylated CpG islands (DMCs) based on the methylation data set and identified five hypermethylated and downregulated genes by comprehensive analysis data. Hopefully, the five hyper-down genes found in this study may act as biomarkers for NPC for NPC diagnosis and pathogenesis.

## Methods

### Data resource

Methylation profiling data from the dataset GSE52068 was downloaded from Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo, GEO) and were produced by the platform of Illumina HumanMethylation450 BeadChip (GPL13534, HumanMethylation450_15017482) (Illumina, Inc., San Diego, CA, USA). A total of 24 NPC cases and 24 controls were included in GSE52068 dataset. Three gene expression profiling datasets (GSE53819, GSE64634 and GSE12452) were obtained from GEO database. The GSE53819 dataset were produced on the platform of Agilent-014850 Whole Human Genome Microarray 4x44 K G4112F and included 18 primary NPC tumors and 18 non-cancerous nasopharyngeal tissues. The GSE64634 data set contained 4 normal nasopharyngeal tissues and 12 nasopharyngeal carcinoma tissues, the GSE12452 included 10 normal nasopharyngeal tissues and 31 nasopharyngeal carcinoma tissues. Additionally, both datasets were based on the GPL570 platforms ([HG-U133_Plus_2] Affymetrix Hu-man Genome U133 Plus 2.0 Array). The analysis process was shown in Fig. 1

### DEGs identification
The raw microarray data of the four datasets downloaded from the GEO database were processed by the online tool GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) to identify DEGs and DMRs between normal nasopharyngeal tissues and nasopharyngeal carcinoma tissues. We identified the DEGs using threshold p value < 0.05 and absolute log2FC > 2.0. The overlapping DEGs among GSE53819, GSE64634 and GSE12452 were identified.

GO term and KEGG pathway enrichment analysis

The R package “clusterProfiler” is implemented for Gene ontology (GO: http://www.geneontology.org), Kyoto Encyclopedia of Genes and Genomes (KEGG: http://www.genome.jp/kegg/pathway.html) pathway enrichment analysis. P<0.05 was considered significant.

Protein–protein interaction (PPI) network construction and hub genes identification

The PPI network of 91 overlapping DEGs was constructed with the online tool STRING (https://string-db.org/). A node represented a gene and each edge represented the interaction between genes in the PPI network. Subsequently, Cytoscape software (download from http://www.cytoscape.org/) was used to search for hub genes. Several models were obtained from the PPI network by the MCODE analysis in Cytoscape. Hub nodes were filtrated out according to the connectivity degree, betweenness centrality, and closeness centrality of a node. Degree Cutoff = 2, K-Core = 2 and Node Score Cutoff = 0.2 were regarded as cutoff parameters for module identification.

Classifier construction

Random Forest (RF) algorithm was applied to construct one RF classifier based on gene expression data to distinguish NPC patients from normal controls. We used 93 samples from the three datasets and the 91 DEGs expression data to train the classifier. The leave-one-out cross validation (LOOCV) was used to assess classification ability. Then, we assessed the average performance of each DEG for classifier using the “importance” function of “Random Forest” package, and the DEGs were ordered in Mean Decrease Accuracy based on their importance.

Identification of differentially methylated CpG sites (DMCs)

Differential methylation sites were screened with P <0.05 and absolute log2FC > 0.2 as thresholds between NPC tissues samples and normal tissues by GEO2R. Then, DMCs were mapped onto gene sub-regions (5'UTR, TSS200, TSS150, 1stExon, body and 3'UTR) and methylated genomic regions (N_Shore, N_Shelf, CpG island, S_Shelf and S_Shore). Subsequently, we selected downregulated genes with hypermethylated sites (hyper-down) and upregulated genes with hypomethylated sites (hypo-up) by comprehensive analysis of DMCs and DEGs.

Results

DEGs identification
To identify the DEGs between normal tissues and NPC tissues, we analyze the data of GSE12452, GSE53819 and GSE64634 datasets. Then, the volcano plots of DEGs were shown in Fig.2. A total of 276 (52 upregulated and 224 downregulated), 638 (188 upregulated and 450 downregulated) and 441 (67 upregulated and 374 downregulated) genes were identified, respectively. With the online tool Venny for the integrated analysis, 91 genes were overlapped among three datasets, including 5 upregulated and 86 downregulated genes (Fig.3). These 91 overlapping genes were confirmed as candidate DEGs and employed for further analysis.

**GO term and KEGG pathway analysis**

The top enriched GO terms and KEGG pathways were listed in Fig. 4. As Fig.4 shows, the 91 overlapping genes were highly associated with pathways including drug metabolism-cytochrome P450 pathway, metabolism of xenobiotics by cytochrome P450 pathway and chemical carcinogenesis pathway. With the analysis of GO terms, these genes were largely enriched in cell component (CC) including ciliary part, motile cilium, axoneme, microtubule and ciliary plasm. The enriched biological process (BP) notations included cilium organization, axoneme assembly and cilium movement.

**Protein–protein interaction (PPI) network construction**

The DEG expression PPI network was showed in Fig.5. The PPI network contained 91 nodes and 9 DEGs were recognized as hub genes (see Table 1) with node degree≥10. In addition, in this PPI network, we observed three significant models after MCODE analysis in Cytoscape and we choose the most significant module with MCODE score = 7.7 and number of nodes =9. The module revealed that seven of the nine genes (DNALI1, RSPH1, SPAG6, ARMC4, DNAI2, DNAH9, and RSPH4A) were belong to the hub genes. With regards to the functions, these DEGs were closely involved the cilium movement, ciliary part and cilium organization, in which DNALI1, TEKT1, ARMC4, DNAI2, DNAH9, RSPH4A, SPAG6 and RSPH1 were highly enriched\textsuperscript{19-27}. ARMC4 is a top tumor suppressor gene and mutation occurs in breast cancer\textsuperscript{28}. SNTN is a kind of apical structure protein and its abnormal expression was verified to be related to pathological and cancerous phenotypes\textsuperscript{29}. These findings indicated that these tightly interactional genes associated molecular pathways and might activate pathogenesis of NPC.

**Classifier construction**

We built the RF classifier using the function in “RandomForest” package for R and added these DEGs to the classifier one by one in order of importance (supplementary table 1). For a variable, the higher decrease in Gini and the higher MeanDecreaseAccuracy measure, the greater is its importance in the classification process. The top 30 DEGs obtained from both rankings showed in Figure 6C. As Fig. 6A shows, the top 2, including CLIC6 and CLU, are the best predictors of the classifier, and have a good prediction power—AUC:0.985. The chloride intracellular channel (CLIC) family contains six homologs in human (CLIC1-6) which plays a critical role in the processes of human cancer. A previous study suggested that the expression level of CLICS (CLIC1,4,5 and 6) were significantly correlated with histological tumor grade in breast cancer\textsuperscript{30}. Clusterin (CLU) was regulated by N,N'-Dinitrosopiperazine
(DNP) and promotes NPC metastasis. In the case of unknown clinical information, it is also very good at identifying tumors and healthy tissues based on two-dimensional representation of unsupervised classification analysis (Fig. 6B).

**Identification of differentially methylated CpG sites (DMCs)**

To our knowledge, the distribution of DMCs and their influence on key functional genomic elements including CpG islands in NPC are poorly defined. We studied the global methylation patterns of 91 DEGs across gene sub-regions and methylated genomic regions. CpG island hypermethylation is a landmark epigenetic modification of cancer. In our study, 3264 DMCs were identified from NPC tumor genomes, among of them, 2599 were hypermethylated CpG sites (hyper-CpGs) and 665 were hypomethylated CpG sites (hypo-CpGs). The number of hyper-CpGs was greater than the hypo-CpGs in NPC tissues. As the Fig. 7 shows that 1132 hyper-DMCs were enriched the body, and 1959 of hyper-CpGs were distinctly enriched with CpG islands (79.80%), but very few hyper-CpGs located in the shelves. Our analyses also suggested that the regions close to TSS (TSS1500, TSS200, and 5’UTR) and the CpG islands (shores), in general, have more DMCs in NPC tumor tissues. In addition, the distribution of few hypo-CpGs were comparatively even among shelves, s-shore and CpG islands, except for higher in N-shore. Expectedly, we observed that the CpG island and its neighborhood location (shores) were predominantly hypermethylated which is similar with an early study. The promoter regions (TSS, 1stExon and 5’UTR) had nearly twice hyper-CpGs (63.41% vs 36.59%) compared to outside of promoter regions (body and 3’UTR), similar observations were also reported by Nitish Kumar Mishra et al. Hyper-CpGs in these promoter regions lead directly to the genetic inactivation of tumour-suppressor genes and silence transcriptional initiation. This finding indicated that hypermethylation related genes may were highly associated with the cell division, cell cycle and development of NPC. Finally, by comprehensive analysis of DMCs and DEGs, we observed 5 differential methylation genes (DMGs), including CLDN10, CLIC6, LRRC10B, SORBS2 and UBXN10, which both hypermethylated and downregulated. The correlation analysis between the 5 DMGs and DMCs was generalized in Table 2.

**Discussion**

In this study, we analyzed DEGs and DMGs comprehensively by bioinformatics analysis. 91 DEGs were identified in NPC, including 5 up-regulated genes and 86 down-regulated genes. CLIC6 and CLU selected from 91 DEGs have the best classification ability by RF classifier. Five genes (CLDN10, CLIC6, SORBS2, UBXN10 and LRRC10B) which were both hypermethylated and downregulated in NPC were identified as well. There were no one that both hypomethylated and upregulated genes (hypo-up), which reveal that the hypermethylated genes may largely related to epigenetic modification in NPC and might lead to downregulation of these genes.

Claudin-10 (CLDN10) is the member of claudins family. Claudins are a key element of tight junction (TJ) protein and its abnormal expression can lead to the invasion and distant metastasis of tumor cells. The previous report indicated that CLDN10 can enhance the metastatic phenotype of human osteosarcoma.
CLDN10 also shows high expression level in the patients who developed papillary thyroid cancer and had a worse survival\textsuperscript{38}. A previous study reported that its homologous protein \textit{CLDN11} was an anti-tumor gene, abnormal methylated and expressed of \textit{CLDN11} regulated the development of the NPC by inhibiting cell migration and invasion\textsuperscript{39}(Li et al. 2018)(39)(39)\textsuperscript{39}. As described earlier, \textit{CLIC6} had best predictive ability and may play important role in NPC. Although, their relationship with NPC are not well known, these findings suggested \textit{CLDN10} and \textit{CLIC6} as biomarkers for NPC are feasible. Especially, \textit{CLIC6} has the best predictive power and has three hyper-CpGs (1stExon, TSS200 and 1stExon), which indicates that it has great potential as a biomarker for NPC.

Sorbin and SH3 domain-containing protein2 (\textit{SORBS2}) is a scaffolding protein and acts to refrain the ovarian cancer invasion\textsuperscript{40}. \textit{SORBS2} suppressed the proliferation, diffusion and migration of hepatocellular carcinoma (HCC) cell both in vivo and in vitro studies\textsuperscript{41}. Similarly, two other studies showed that \textit{SORBS2} may as a tumor suppressor gene not only is associated with cervical carcinogenesis but also negatively regulated the pancreatic cancer\textsuperscript{42,43}. Hence, we can speculate that the hyper-down \textit{SORBS2} may work in the same way as in ovarian and HCC. \textit{UBXN10} and \textit{LRRC10B} have not been reported in NPC. A study reported that \textit{LRRC10B} might be a target gene regulated by RPLP0P2. \textit{RPLP0P2} was involved into lower prognosis, proliferation and adhesion ability of lung tumor cells\textsuperscript{44}. Malavika Raman et al demonstrated \textit{UBXN10} locates in VCP-dependent manner cilia\textsuperscript{45}. VCP and \textit{UBXN10} are essential for promoting cilia formation. As mentioned above, these genes are abundant in ciliary part, motile cilium, cilium organization and cilium movement and therefore may perform certain functions in NPC progress. However, this hypothesis requires to be verified in further experiments.

Pathways in cancer, we found that these abnormal genes were primarily enriched metabolism-cytochrome P450 pathway (CYP450), metabolism of xenobiotics by cytochrome P450 pathway and chemical carcinogenesis pathway. CYP450 was a group of enzymes that appears highly polymorphic and their variants play an important role in cancer. Currently, CYP450 was used in the 80% of drugs, including some anticancer drugs\textsuperscript{46}. Similarly, a previous report indicated that one of the expression patterns of cytochrome P450 (CYP2A6) genes had an increased risk for NPC and affected the susceptibility of NPC\textsuperscript{47}. Hence, the DEGs were involved in metabolism of xenobiotics by CYP450 pathway may ultimately affected the treatment of NPC. The chemical carcinogenesis pathway has been proven to connected with high incidence of NPC\textsuperscript{48,49}. Active cilia can remove inhaled harmful particles and bacteria. Therefore, without normal cilia, mucociliary clearance is severely impaired, and bacterial mucus stays in the nose and sinuses causing inflammation and recurrent infections. Therefore, we can acquire an insight to understand the mechanism of cancer by analysis of these related signaling pathways.

\section*{Conclusions}

In our study, we verified some signature genes that were directly or indirectly involved with the pathological process of NPC. Some of them were identified in NPC for the first time and may act as potential biomarkers.
Abbreviations

NPC: Nasopharyngeal carcinoma; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: Differentially expressed genes; PPI: protein-protein interaction; RF: Random Forest; CpG islands: 5′—C—phosphate—G—3′; DMCs: Differentially methylated CpG sites.

Declarations

Availability of data and materials

Data are available at Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo,GEO) as GSE53819, GSE64634 and GSE12452.

Authors' Contributions

Zengyan Zong and Dayang Chen contributed equally to this work in the design and write original draft should be considered as co-first authors; Wei Wu and Mengmeng Wang collected and analyzed the data; Xiaowen Dou and Xiang Ji performed the literature search and were responsible for data visualization; Xiuming Zhang and Dan Xiong revised the paper carefully and rigorously; All authors have read and approved the final manuscript.

Acknowledgements

Not applicable.

Funding

This study was supported by the National Natural Science Foundation of China (Grant NO.81772921, 81502344). A grant from the Science and Technology Planning Project of Shenzhen City of China(No. JCYJ20180306172209668). The discipline construction ability promotion project of Shenzhen health and population family planning commission (NO. SZXJ2017018, SZXJ2018031). Sanming Project of Medicine in Shenzhen (NO. SZSM201601062).

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors state that there are no financial, personal, or professional conflicts of interests that may hinder this work.
References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. _CA Cancer J Clin_. 2015;65(2):87-108.

2. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. _Lancet_. 2019;394(10192):64-80.

3. Wong EHC, Liew YT, Loong SP, Prepageran N. Five-year Survival Data on the Role of Endoscopic Endonasal Nasopharyngectomy in Advanced Recurrent rT3 and rT4 Nasopharyngeal Carcinoma. _Ann Otol Rhinol Laryngol_. 2019:3489419887410.

4. Bensouda Y, Kaikani W, Ahbeddou N, et al. Treatment for metastatic nasopharyngeal carcinoma. _Eur Ann Otorhinolaryngol Head Neck Dis_. 2011;128(2):79-85.

5. Zhang L, Chen QY, Liu H, Tang LQ, Mai HQ. Emerging treatment options for nasopharyngeal carcinoma. _Drug Des Devel Ther_. 2013;7:37-52.

6. Zhu L, Liu Y, Yang Y, Mao XM, Yin ZD. CircRNA ZNF609 promotes growth and metastasis of nasopharyngeal carcinoma by competing with microRNA-150-5p. _Eur Rev Med Pharmacol Sci_. 2019;23(7):2817-2826.

7. Bruce JP, Liu FF. MicroRNAs in nasopharyngeal carcinoma. _Chin J Cancer_. 2014;33(11):539-544.

8. Lu J, He ML, Wang L, et al. MiR-26a inhibits cell growth and tumorigenesis of nasopharyngeal carcinoma through repression of EZH2. _Cancer Res_. 2011;71(1):225-233.

9. Dai W, Zheng H, Cheung AK, 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.

10. Xue K, Cao J, Wang Y, et al. Identification of Potential Therapeutic Gene Markers in Nasopharyngeal Carcinoma Based on Bioinformatics Analysis. _Clin Transl Sci_. 2019.

11. Lin X, Bi Z, Hu Q, et al. TSPAN8 serves as a prognostic marker involving Akt/MAPK pathway in nasopharyngeal carcinoma. _Ann Transl Med_. 2019;7(18):470.

12. Luo X, Li N, Zhao X, et al. DHRS2 mediates cell growth inhibition induced by Trichothecin in nasopharyngeal carcinoma. _J Exp Clin Cancer Res_. 2019;38(1):300.

13. Chang SL, Chan TC, Chen TJ, Lee SW, Lin LC, Win KT. HOXC6 Overexpression Is Associated With Ki-67 Expression and Poor Survival in NPC Patients. _J Cancer_. 2017;8(9):1647-1654.

14. Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. _Lancet_. 2018;392(10149):777-786.

15. Bruce JP, Yip K, Bratman SV, Ito E, Liu FF. Nasopharyngeal Cancer: Molecular Landscape. _J Clin Oncol_. 2015;33(29):3346-3355.

16. Ren X, Yang X, Cheng B, et al. HOPX hypermethylation promotes metastasis via activating SNAIL transcription in nasopharyngeal carcinoma. _Nat Commun_. 2017;8:14053.
17. Fan J, Zhang Y, Mu J, et al. TET1 exerts its anti-tumor functions via demethylating DACT2 and SFRP2 to antagonize Wnt/beta-catenin signaling pathway in nasopharyngeal carcinoma cells. *Clin Epigenetics*. 2018;10(1):103.

18. Laird PW. The power and the promise of DNA methylation markers. *Nat Rev Cancer*. 2003;3(4):253-266.

19. Viswanadha R, Sale WS, Porter ME. Ciliary Motility: Regulation of Axonemal Dynein Motors. *Cold Spring Harb Perspect Biol*. 2017;9(8).

20. Yu X, Ng CP, Habacher H, Roy S. Foxj1 transcription factors are master regulators of the motile ciliogenic program. *Nat Genet*. 2008;40(12):1445-1453.

21. Ryan R, Failer M, Reilly ML, et al. Functional characterization of tektin-1 in motile cilia and evidence for TEKT1 as a new candidate gene for motile ciliopathies. *Hum Mol Genet*. 2018;27(2):266-282.

22. Loges NT, Antony D, Mauer A, et al. Recessive DNAH9 Loss-of-Function Mutations Cause Laterality Defects and Subtle Respiratory Ciliary-Beating Defects. *Am J Hum Genet*. 2018;103(6):995-1008.

23. Lucas JS, Barbato A, Collins SA, et al. European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. *Eur Respir J*. 2017;49(1).

24. Shapiro A, Davis S, Manion M, Briones K. Primary Ciliary Dyskinesia (PCD). *Am J Respir Crit Care Med*. 2018;198(2):P3-P4.

25. Kott E, Legendre M, Copin B, et al. Loss-of-function mutations in RSPH1 cause primary ciliary dyskinesia with central-complex and radial-spoke defects. *Am J Hum Genet*. 2013;93(3):561-570.

26. Loges NT, Olbrich H, Fenske L, et al. DNAI2 mutations cause primary ciliary dyskinesia with defects in the outer dynein arm. *Am J Hum Genet*. 2008;83(5):547-558.

27. Raidt J, Wallmeier J, Hjeij R, et al. Ciliary beat pattern and frequency in genetic variants of primary ciliary dyskinesia. *Eur Respir J*. 2014;44(6):1579-1588.

28. Pongor L, Kormos M, Hatzis C, Pusztai L, Szabo A, Gyorffy B. A genome-wide approach to link genotype to clinical outcome by utilizing next generation sequencing and gene chip data of 6,697 breast cancer patients. *Genome Med*. 2015;7:104.

29. Park JS, Choi SB, Kim HJ, et al. Intraoperative Diagnosis Support Tool for Serous Ovarian Tumors Based on Microarray Data Using Multicategory Machine Learning. *Int J Gynecol Cancer*. 2016;26(1):104-113.

30. Ko JH, Ko EA, Gu W, Lim I, Bang H, Zhou T. Expression profiling of ion channel genes predicts clinical outcome in breast cancer. *Mol Cancer*. 2013;12(1):106.

31. Li Y, Lu J, Zhou S, et al. Clusterin induced by N,N'-Dinitrosopiperazine is involved in nasopharyngeal carcinoma metastasis. *Oncotarget*. 2016;7(5):5548-5563.

32. Dai W, Cheung AK, Ko JM, et al. Comparative methylome analysis in solid tumors reveals aberrant methylation at chromosome 6p in nasopharyngeal carcinoma. *Cancer Med*. 2015;4(7):1079-1090.

33. Mishra NK, Southekal S, Guda C. Survival Analysis of Multi-Omics Data Identifies Potential Prognostic Markers of Pancreatic Ductal Adenocarcinoma. *Front Genet*. 2019;10:624.
34. Mishra NK, Guda C. Genome-wide DNA methylation analysis reveals molecular subtypes of pancreatic cancer. *Oncotarget.* 2017;8(17):28990-29012.

35. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet.* 2002;3(6):415-428.

36. Tabaries S, Siegel PM. The role of claudins in cancer metastasis. *Oncogene.* 2017;36(9):1176-1190.

37. Zhang X, Wang X, Wang A, Li Q, Zhou M, Li T. CLDN10 promotes a malignant phenotype of osteosarcoma cells via JAK1/Stat1 signaling. *J Cell Commun Signal.* 2019;13(3):395-405.

38. Zhou Y, Xiang J, Bhandari A, et al. CLDN10 is Associated with Papillary Thyroid Cancer Progression. *J Cancer.* 2018;9(24):4712-4717.

39. Li HP, Peng CC, Wu CC, et al. Inactivation of the tight junction gene CLDN11 by aberrant hypermethylation modulates tubulins polymerization and promotes cell migration in nasopharyngeal carcinoma. *J Exp Clin Cancer Res.* 2018;37(1):102.

40. Zhao L, Wang W, Huang S, et al. The RNA binding protein SORBS2 suppresses metastatic colonization of ovarian cancer by stabilizing tumor-suppressive immunomodulatory transcripts. *Genome Biol.* 2018;19(1):35.

41. Han L, Huang C, Zhang S. The RNA-binding protein SORBS2 suppresses hepatocellular carcinoma tumourigenesis and metastasis by stabilizing RORA mRNA. *Liver Int.* 2019;39(11):2190-2203.

42. Taieb D, Roignot J, Andre F, et al. ArgBP2-dependent signaling regulates pancreatic cell migration, adhesion, and tumorigenicity. *Cancer Res.* 2008;68(12):4588-4596.

43. Backsch C, Rudolph B, Steinbach D, et al. An integrative functional genomic and gene expression approach revealed SORBS2 as a putative tumour suppressor gene involved in cervical carcinogenesis. *Carcinogenesis.* 2011;32(7):1100-1106.

44. Chen J, Hu L, Chen J, et al. Low expression IncRNA RPLP0P2 is associated with poor prognosis and decreased cell proliferation and adhesion ability in lung adenocarcinoma. *Oncol Rep.* 2016;36(3):1665-1671.

45. Raman M, Sergeev M, Garnaas M, et al. Systematic proteomics of the VCP-UBXD adaptor network identifies a role for UBXN10 in regulating ciliogenesis. *Nat Cell Biol.* 2015;17(10):1356-1369.

46. Mittal B, Tulsyan S, Kumar S, Mittal RD, Agarwal G. Cytochrome P450 in Cancer Susceptibility and Treatment. *Adv Clin Chem.* 2015;71:77-139.

47. Tiwawech D, Srivatanakul P, Karalak A, Ishida T. Cytochrome P450 2A6 polymorphism in nasopharyngeal carcinoma. *Cancer Lett.* 2006;241(1):135-141.

48. Fang CY, Huang SY, Wu CC, et al. The synergistic effect of chemical carcinogens enhances Epstein-Barr virus reactivation and tumor progression of nasopharyngeal carcinoma cells. *PLoS One.* 2012;7(9):e44810.

49. Costa S, Garcia-Leston J, Coelho M, et al. Cytogenetic and immunological effects associated with occupational formaldehyde exposure. *J Toxicol Environ Health A.* 2013;76(4-5):217-229.
## Tables

**Table 1.** The 9 hub genes of PPI network

| Gene symbol | Degree | Betweenness Centrality | Closeness Centrality | Gene Name |
|-------------|--------|-------------------------|-----------------------|-----------|
| DNALI1      | 13     | 0.21653                 | 0.3981                | dynein axonemal light intermediate chain 1 |
| RSPH1       | 13     | 0.08551                 | 0.3874                | Radial spoke head 1 homolog |
| ARMC4       | 12     | 0.03963                 | 0.3739                | Armadillo repeat containing 4 |
| SPAG6       | 12     | 0.14531                 | 0.3707                | Sperm-associated antigen 6 |
| DNAI2       | 11     | 0.02607                 | 0.3613                | Dynein, axonemal, intermed. polypeptide 2 |
| DNAH9       | 11     | 0.05441                 | 0.3772                | Dynein heavy chain 9 |
| SNTN        | 11     | 0.20846                 | 0.3772                | Sentan, cilia apical structure protein |
| RSPH4A      | 10     | 0.02664                 | 0.3675                | Radial spoke head component 4A |
| TEKT1       | 10     | 0.08609                 | 0.3839                | Tektin 1 |

Due to technical limitations, table 2 is only available as a download in the Supplemental Files section.