

RESEARCH ARTICLE

Editorial Process: Submission:01/25/2021 Acceptance:06/25/2021

Analysis of Nuclear Encoded Mitochondrial Gene Networks in Cervical Cancer

Cecile Meneur^{1,2}, Sangavi Eswaran¹, Divya Adiga¹, Sriharikrishnaa S¹, Nadeem Khan G¹, Sandeep Mallya³, Sanjiban Chakrabarty¹, Shama Prasada Kabekkodu^{1*}

Abstract

Background: Cervical cancer (CC) is one of the most common female cancers in many developing and underdeveloped countries. High incidence, late presentation, and mortality suggested the need for molecular markers. Mitochondrial defects due to abnormal expression of nuclear-encoded mitochondrial genes (NEMG) have been reported during cancer progression. Nevertheless, the application of NEMG for the prognosis of CC is still elusive. Herein, we aimed to investigate the associations between NEMG and CC prognosis. **Materials and Methods:** The differentially expressed genes (DEG) in the TCGA-CESC dataset and NEMGs were retrieved from TACCO and Mitocarta2.0 databases, respectively. The impact of methylation on NEMG expression were predicted using DNMIVD and UALCAN tools. HCMDB tool was used to predict genes having metastatic potential. The prognostic models were constructed using DNMIVD, TACCO, GEPIA2, and SurvExpress. The functional enrichment analysis (FEA) was performed using clusterProfiler. The protein-protein interaction network (PPIN) was constructed to identify the hub genes (HG) using String and CytoHubba tools. Independent validation of the HG was performed using OncoPrint and Human Protein Atlas databases. The druggable genes were predicted using DGIdb. **Results:** Among the 52 differentially expressed NEMG, 15 were regulated by DNA methylation. The expression level of 16, 10, and 7 has the potential for CC staging, prediction of metastasis, and prognosis. Moreover, 1 driver gene and 16 druggable genes were also identified. The FEA identified the enrichment of cancer-related pathways, including AMPK and carbon metabolism in cancer. The combined expression of 10 HG has been shown to affect patient survival. **Conclusion:** Our findings suggest that the abnormal expression of NEMGs may play a critical role in CC development and progression. The genes identified in our study may serve as a prognostic indicator and therapeutic target in CC.

Keywords: Nuclear encoded mitochondrial genes- prognosis- survival- TCGA-CESC- cervical cancer

Asian Pac J Cancer Prev, 22 (6), 1799-1811

Introduction

Cervical cancer (CC) is a significant public health problem affecting women in countries with low resource settings. According to the Global Cancer Observatory database, there were 570,000 cases and 311,000 deaths due to CC in 2018 (Ferlay et al., 2018). Over a third of the overall global CC cases were contributed together by India and China. In 2018, the CC cases in India and China were 97,000 and 106,000, respectively, with mortality of 60,000 and 48,000, respectively (Arbyn et al., 2020; Bray et al., 2018). The Squamous cell carcinoma (SCC) is the most common CC histological type. The standard method for CC treatment includes surgery, chemotherapy, and radiotherapy. Persistent infection with high-risk HPV, high parity, multiple sexual partners, smoking cigarettes,

and long-term use of oral contraceptive pills are a few risk factors associated with CC. The introduction of Pap testing and HPV typing has significantly reduced CC incidence in many countries (Johnson et al., 2019; Wudtisan et al., 2019). Despite the availability of effective early screening and detection methods, CC is generally diagnosed at an advanced stage leading to a high mortality rate. The 5-year survival rate for stage I, stage II, stage III, and stage IV CC was 81-96%, 65-87%, 35-50%, and 15-20%, respectively. CC with advanced stage shows recurrence and therapy resistance and most patients succumb within three years (Canfell et al., 2020; Charakorn et al., 2018). This suggests the need for biomarkers to identify the patients with poor prognosis at an early stage for intensified treatment for improved patient care. In this direction, identifying the molecular markers and associated networks may be useful

¹Department of Cell and Molecular Biology, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal-576104, Karnataka, India. ²La Rochelle University, Avenue Albert Einstein, 17031, La Rochelle, France. ³Department of Bioinformatics, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal-576104, Karnataka, India. *For Correspondence: shama.prasada@manipal.edu, spbhat81@gmail.com. Cecile Meneur and Sangavi Eswaran have equal contribution in this study.

for better management of CC.

Genome-wide genetic and epigenetic studies using normal and tumor samples lead to identifying cancer-associated genes and related pathways for diagnostic and prognostic applications in cancer. Studies have shown that the genome-wide data's reanalysis may provide (i) molecular markers for diagnosis, prognosis, therapy, and (ii) a novel gene-gene regulatory networks, mechanisms, and pathways facilitating CC. Interestingly, previous studies have indicated that the integrated bioinformatics analysis of big data generated from cancer studies can develop reliable and novel biomarkers, networks, and mechanisms related to CC pathogenesis (Lin et al., 2019; Wu et al., 2019).

Mitochondria are an intracellular organelle that plays a role in cellular bioenergetics, free radical generation, and cell signaling. The human mtDNA is 16.5kb in size and codes for 37 genes linked with the OXPHOS pathway, ATP production, and apoptosis initiation (Shaughnessy et al., 2015). Defects in both structure and function of mitochondria contribute to resistance to apoptosis, abnormal cell proliferation, and therapy resistance in cancer (Indran et al., 2011). Both somatic mutation and mtDNA copy number changes are linked with cancer, including CC (Chatterjee et al., 2006; Sun et al., 2020). Although mtDNA encodes for proteins, most of the protein required for its function is synthesized by nuclear-encoded genes. As per the estimate, over 1,500 proteins are necessary for maintaining the structure and function of mitochondria. The anterograde and retrograde signaling between mitochondria and the nucleus is critical for the normal functioning of the cell (Pfanner et al., 2019). Many nuclear-encoded genes linked with mitochondrial function (NEMG) show a significant difference in their expression in normal and CC conditions (Sun et al., 2020). Therefore, profiling the expression of NEMG can be useful as a sensitive and specific biomarker for diagnostic and prognostic applications in CC.

In the present study, we have investigated the differentially expressed nuclear-encoded mitochondrial genes (DE-NEMG) and the associated network for diagnostic and prognostic applications in CC. We have analyzed The Cancer Genome Atlas, Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA-CESC) datasets for DE-NEMG in CC to evaluate its prognostic significance. Using the TCGA-CESC dataset, we have constructed the protein-protein interaction network (PPIN) and hub genes (HG) and predicted the possible drug targets. The integrated bioinformatic analysis identified interaction networks and mitochondrial targets that may be useful as a potential marker for diagnosis, treatment, and CC therapy.

Materials and Methods

Selection of NEMG

MitoCarta2.0 contains human and mouse proteins having strong mitochondrial translocation signals from fourteen different tissue samples (Calvo et al., 2016). In the present study, we downloaded 1158 NEMGs from the MitoCarta2.0 database to identify the DE-NEMGs.

TCGA-CESC datasets

The TCGA contains molecularly characterized epigenomics, transcriptomics, and proteomics data for over 20000 primary tumors and matched normal samples across 33 different cancer types. The TCGA-CESC dataset contains clinical and molecular data for 310 samples (307 tumors and 3 normal samples) (Tomczak et al., 2015). The differentially expressed genes in TCGA-CESE data sets were retrieved using the Transcriptome Alterations in Cancer Omnibus (TACCO: <http://tacco.life.nctu.edu.tw/>) online tool. The TACCO is a freely available online tool for identifying differentially expressed genes (DEG) and miRNAs from TCGA datasets. The TACCO can be used for Gene Ontology (GO), pathway enrichment, and prognostic model construction using a user-defined gene list (Chou et al., 2019).

Identification of DE-NEMG

We have downloaded a list of 1,158 NEMGs from the MitoCarta2.0 database. We downloaded the list of differentially expressed protein-coding genes (DEGs) between normal and tumor samples in the TCGA-CESC dataset via the TACCO online tool. We have identified the DEGs in the TCGA-CESC data set using the "select DEGs" function with cut-off criteria of adjusted p-value < 0.05 and expression value log 2-fold change of +2 and -2 between tumor and normal tissue samples. We have compared the DEGs in TCGA-CESC dataset with that of NEMG downloaded from the MitoCarta2.0 using Venny 2.1 (<https://bioinfogp.cnb.csic.es/tools/venny/>) online tool to identify the DE-NEMGs.

Identification of differentially methylated NEMG

We have used UALCAN (<http://ualcan.path.uab.edu/index.html>) and DNMIIVD (<http://119.3.41.228/dnmivd/index/>) online resources for identification of methylation regulated NEMGs. UALCAN online tool can identify differentially expressed, methylation-regulated genes, along with clinical attributes such as survival, age, histology, tumor grade, and nodal metastasis. A beta values of 0 and 1 are considered as unmethylated and completely methylated respectively. Beta value: 0.7 - 0.5 and 0.3 - 0.25 and a P<0.05 considered significantly hypermethylated and hypomethylated (Chandrashekar et al., 2017). DNMIIVD is a user-friendly interactive visualization of the DNA methylation profile of genes in TCGA datasets. The tool has a module for gene expression analysis, methylation and expression correlation, survival analysis, diagnostic and prognostic model generation. The samples were divided into high (H) and low (L) based on median DNA methylation beta values as a threshold. DNA methylation beta values of 0.3 and 0.7 were used as a cutoff for sample categorization (Ding et al., 2020).

Evaluation of metastatic potential of DE-NEMG

The Human Cancer Metastasis Database (HCMDB <https://hcmdb.i-sanger.com/>) is a freely available online tool for assessing gene metastatic potential. The interactive web tool contains metastatic data from 38 metastasis sites and 29 cancer types (Zheng et al., 2018). In the present study, the HCMDB database was used to test the metastatic

potential of DE-NEMG in TCGA-CESC data.

Construction of PPIN and identification HG

The PPIN and HG play a crucial role in governing the biological process and signaling pathways. They are commonly used to predict cellular function, understand disease mechanisms, and design drug targets. Using the Search Tool for the Retrieval of Interacting Genes (STRING: <https://string-db.org/>) version 11, the PPIN of the DE-NEMG was constructed with the highest confidence score of 0.9 and minimum interaction number =2 (Szkłarczyk et al., 2017). The cytoHubba V 0.1 plugin of Cytoscape was used to identify the HG of the PPIN. The 10 HG were selected based on the highest degree of connectivity on the PPIN. The two-dimensional visualization of the network was performed through Cytoscape 3.7.2 (<https://cytoscape.org/>) (Chin et al., 2014; Shannon et al., 1971).

Functional enrichment analysis (FEA) of DE-NEMG

The FEA included prediction of gene ontology (GO) [biological process (BP), cellular component (CComp) and molecular function (MF)], and pathway enrichment analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG: <http://www.genome.jp>) (Kanehisa, 2000). The analysis and visualization of functional and pathway enrichment for 52 DE-NEMG were performed using the clusterProfiler tool in the Bioconductor package in the R statistical environment (Yu et al., 2012).

Survival analysis of DE-NEMG

In the present study, prognostic prediction and pathological classification models were generated for CC using the 52 differentially expressed NEMG via the Random forest algorithm using TACCO online tool. Using GEPIA2 (<http://gepia2.cancer-pku.cn/#index>), we have predicted the association between DE-NEMGs with disease-free survival (DFS) and or overall survival (OS) at the individual gene level (Tang et al., 2019).

Identification of driver gene and enrichment analysis

The driver gene analysis was carried out using DriverDBv3 (<http://driverdb.tms.cmu.edu.tw/>) online tool. The DriverDBv3 uses genetic and epigenetic changes in OMICS databases to compare against clinical data to identify cancer driver genes (Liu et al., 2020). We have used the CCDB database (<https://webs.iitd.edu.in/raghava/ccdb/index.php>) (Agarwal et al., 2011) to identify the DE-NEMG, which are already reported in CC. Also, the disease enrichment analysis of 52 DE-NEMG was performed using Comparative Toxicogenomics Database (<http://ctdbase.org>) (Davis et al., 2019).

Analysis of the HG

We evaluated the relative expression 10 HG via the GEPIA2 tool. The GEPIA2 is an online tool for in silico differential expression analysis of genes using TCGA transcriptome and GTEx expression data. It allows the expression analysis between normal and tumor samples, multiple gene comparisons, and survival analysis. We analyzed the survival data of 10 HGs using TCGA-CESC

data. Next, we analyzed the expression of 10 HGs at the protein level using The Human Protein Atlas (HPA: <http://www.proteinatlas.org/>) (Thul and Lindskog, 2018).

Drug and DE-NEMG interaction analysis

The drug-gene interaction database (DGIdb, <https://www.dgldb.org/>) was used to screen the drugs that interact with DE-NEMGs. The DGIdb is a user-friendly online resource for screening a druggable genome. The tool has browse and search options to identify DGI and enrichment. Besides, the DGIdb has options for searching approved, antineoplastic, and immunotherapies options to filter the drugs (Cotto et al., 2018).

Results

Data Retrieval and differential gene expression analysis

Figure 1 depicts the workflow of the study. We first retrieved the DEGs in the TCGA-CESC dataset using the TACCO database. Our analysis identified 2020 (Figure 2A) DEGs with a fold change of 2 and $P < 0.05$. Among those 2020 DEGs, 802 and 1218 genes were upregulated and downregulated, respectively. The MitoCarta2.0 database contained 1158 NEMG. Figure 2B shows the expression of 1158 NEMG in the TCGA-CESC dataset. The comparison between 2020 DEGs with that of 1158 NEMG lead to the identification of 52 (27 upregulated and 25 downregulated) DE- NEMG (Table 1, Figure 2C, and 2D). The list of all DEGs in the TCGA-CESC dataset and NEMG retrieved from Mitocarta2.0 are shown in Supplementary Tables 1, 2 and 3.

DNA methylation regulated NEMG

Next, we analyzed the impact of promoter DNA methylation on the expression of 52 DE-NEMG using DNMIIVD tools by Pearson's correlation analysis. We considered genes that are commonly predicted as differentially methylated by the UALCAN and DNMIIVD tools for all subsequent analysis. Among the 52 differentially expressed genes, 19 genes showed significant aberrant promoter methylation. Among the 19 differentially methylated genes, 8 and 11 genes were hypomethylated and hypermethylated, respectively (Table 1). Furthermore, 27 genes showed an inverse correlation between methylation and expression by Pearson correlation analysis (Table 1, Supplementary Figure 1). Out of 27 genes, 19 of them showed significant differential methylation between normal and tumor samples. Interestingly, among the 19 significantly differentially methylated genes, only 15 revealed the inverse correlation between methylation and expression (Supplementary Figure 2).

Identification of DE-NEMG for CC staging application

By using UALCAN, we analyzed the expression of 52 DE-NEMG in various stages of CC. Among the DEGs, 16 of them showed the potential to differentiate CC stages (Table 1). Expression analysis of 6 genes (PIF1, DHCR24, CPT1B, FASN, DMGDH, and PDK4) can distinguish stage 1 from stage 2 CC. Genes, namely, HK2, NCEH1, NIPSNAP3B, ABCD2, ACACB, DEPP1,

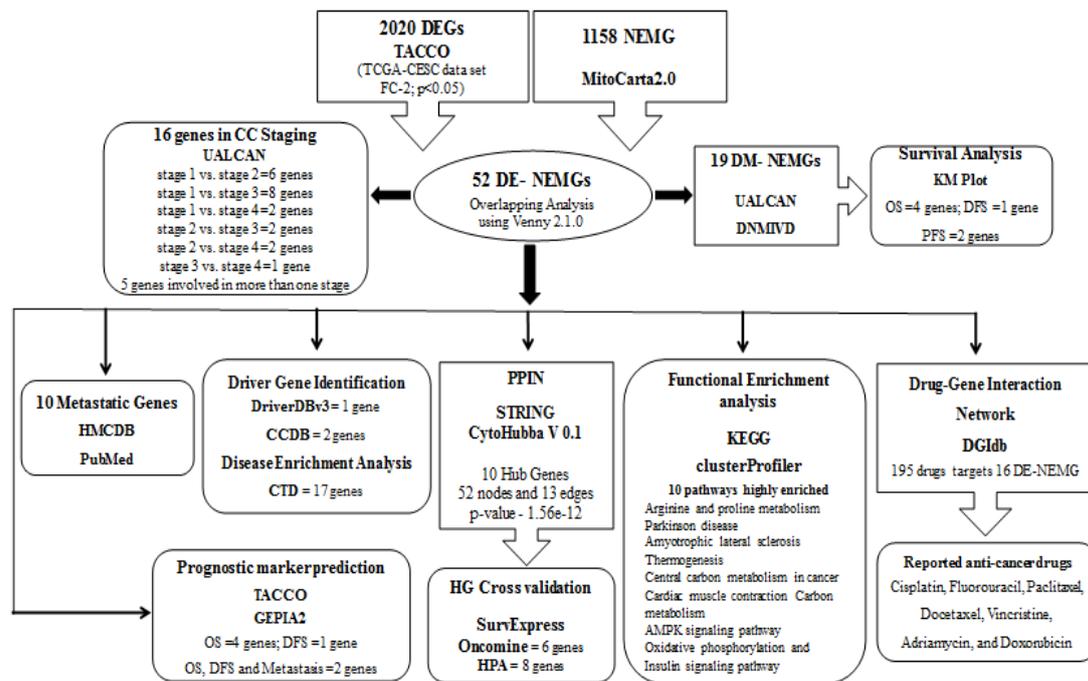


Figure 1. The Flow Diagram of Bioinformatic Analysis NEMG and Associated Pathways in CC.

ACSS3, and PABPC5, were differentially expressed between stage 1 vs. stage 3 CC. NIPSNAP3B and DMGDH expression were significantly different in stage 1 vs. stage 4 CC. Gene, namely, NCEH1 and ABCD2 expression, was significantly different in stage 2 vs. stage 3 CC. SLC25A10 and NIPSNAP3B expression showed significant differential expression between stage 2 vs. stage 4 CC. ACACB expression was significantly different between stage 3 vs. stage 4 CC. Interestingly, genes such as NCEH1, NIPSNAP3B, ABCD2, DMGDH, and ACACB showed a significant difference in expression

in more than one CC stage (Supplementary Table 4).

Identification of NEMG for prognostic application

We investigated the differentially methylated and DEGs for their predictive utility in CC using DNMIIVD, UALCAN, and TACCO tools with default parameters. The multivariate proportional hazard regression model was applied to divide the patients into high-risk and low-risk groups to generate the Kaplan Meier (KM) plot. The prognostic model generated using 19 differentially methylated NEMG suggested a significant prognostic

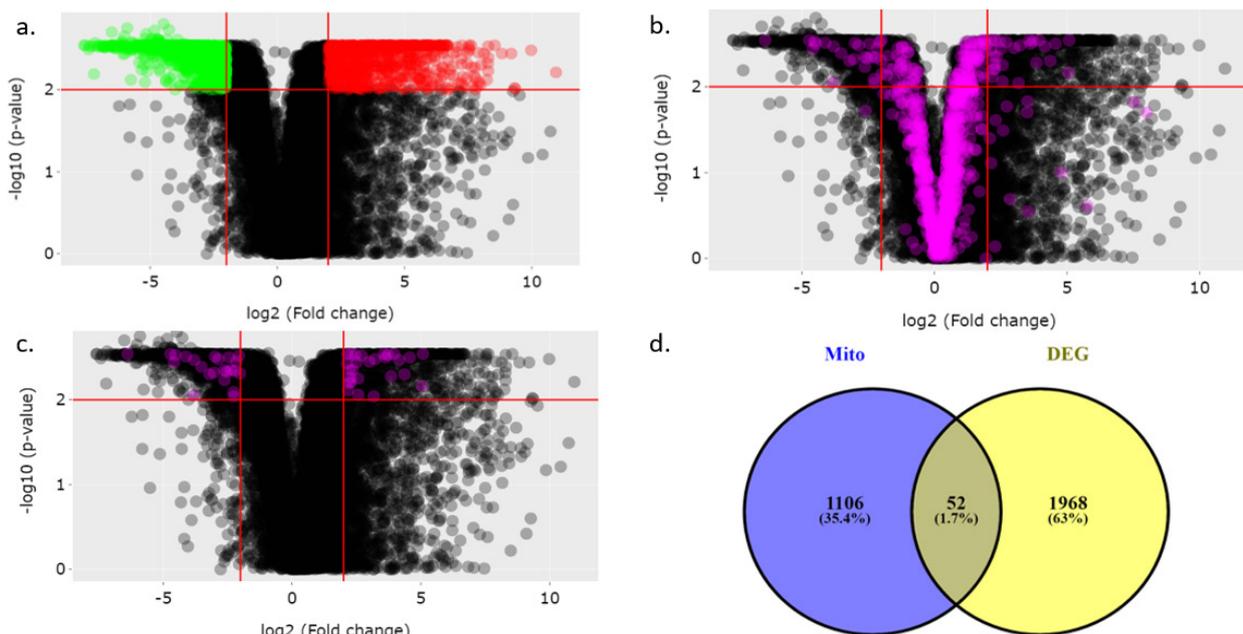


Figure 2. Nuclear Encoded Mitochondrial Genes (NEMG) are Differentially Expressed between Normal and Cervical Cancer Samples. A) Represents the expression profile of genes in TCGA-CESC dataset. B) Represents the expression profile of 1158 NEMG in TCGA-CESC dataset. C) Represents the expression profile of 52 DE-NEMGs in TCGA-CESC dataset. D) Venn diagram showing 52 DE-NEMGs.

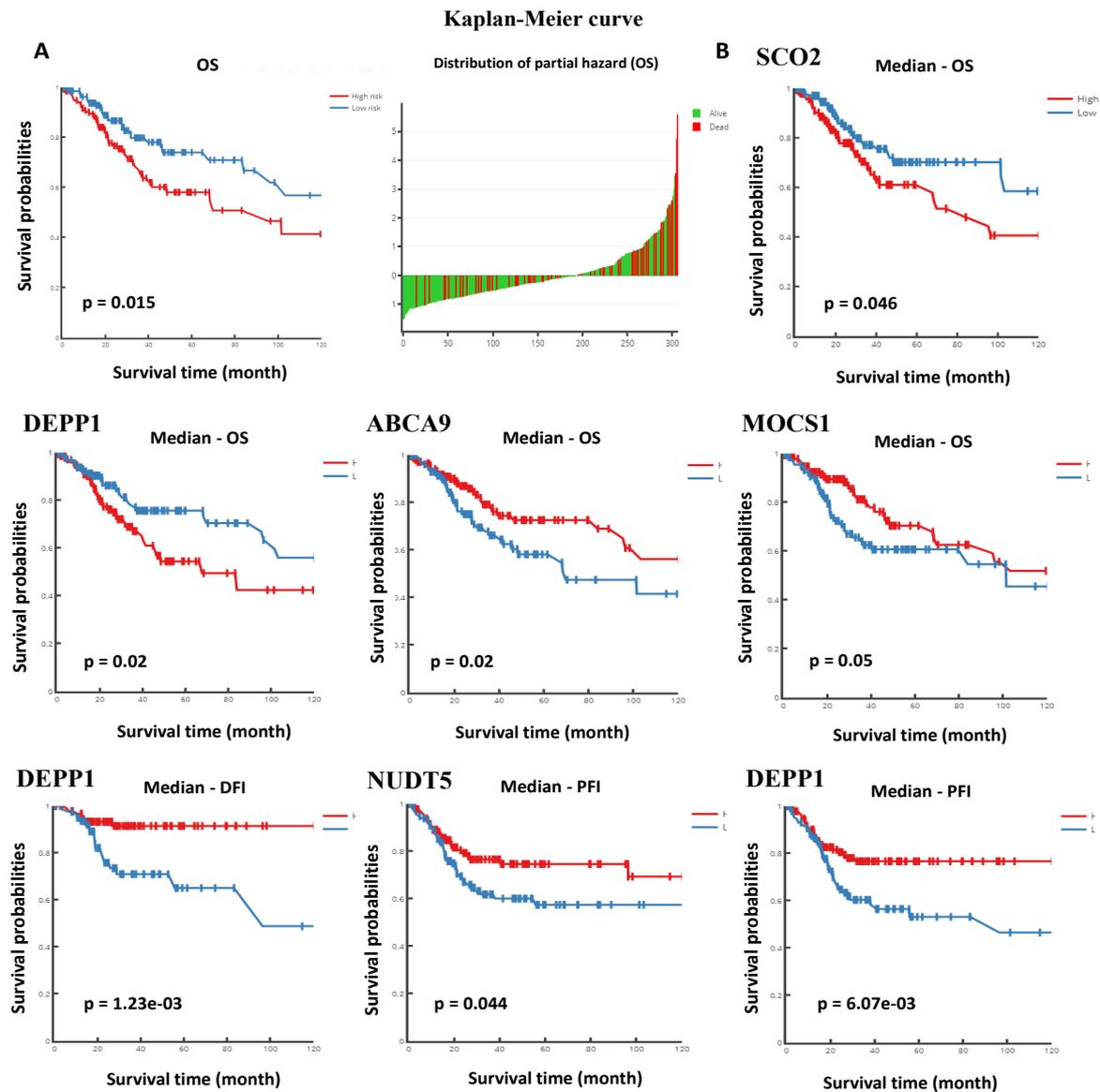


Figure 3. Prognostic Model Based on DNA Methylation. A) Represents the KM plot generated using multivariate proportional hazard regression model based on the 19 differentially methylated genes. B) Represents the KM plot for differentially methylated individual genes linked with OS, DFS and PFS.

value in CESC (Figure 3A). The high-risk category patients showed lower OS as opposed to low-risk categories. When analyzed individually, the differential methylation of gene promoters had substantial implications in CC prognosis (Figure 3B). Methylation status of SCO2, MOCS1, DEPP1, and ABCA9 was linked with OS, and DEPP1 was linked with DFS, and NUDT5, DEPP1 with PFS (Figure 3B) (Table 1).

Using the Random Forest model, we assessed the association between DE-NEMG with CC prognosis using the TACCO tool (Figure 4A). The DE-NEMGs displayed a sensitivity and specificity of 0.91 and 0.89, respectively, to distinguish high risk from low-risk categories (Figure 4B). The prognostic ability of the individual genes was tested by the GEPIA2 tool using Cox Proportional-Hazards Model. The genes with significant OS (499 genes) and DFS (500 genes) in the TCGA-CESC dataset were downloaded from GEPIA2 and compared against 52 DE-NEMG using Venny 2.1.0 tool. We found that genes,

namely, HK2, MSRB3, FASN, and BDH1, showed a significant prognostic value towards OS; while CKMT2 towards DFS (Figure 4C) (Table 1). Besides, PDK4 and BCL2 are additional genes significantly associated with (i) OS and DFS and (ii) metastasis by literature analysis. Next, these genes (HK2, MSRB3, FASN, BDH1, CKMT2, PDK4, and BCL2) were used for risk score construction using the SurvExpress tool (<http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp>) (Aguirre-Gamboa et al., 2013). CC patients were categorized into high-risk and low-risk groups as per the median risk score for survival outcomes. The KM plot suggested that the high-risk patients showed a significantly lower survival rate than the low risk-group (Figure 4D).

Identification of DE-NEMG associated with metastasis

The DE-NEMG with metastatic potential in CC was identified using HCMDB online tool and by literature search using PubMed. Among the 52 DEGs, 10 genes

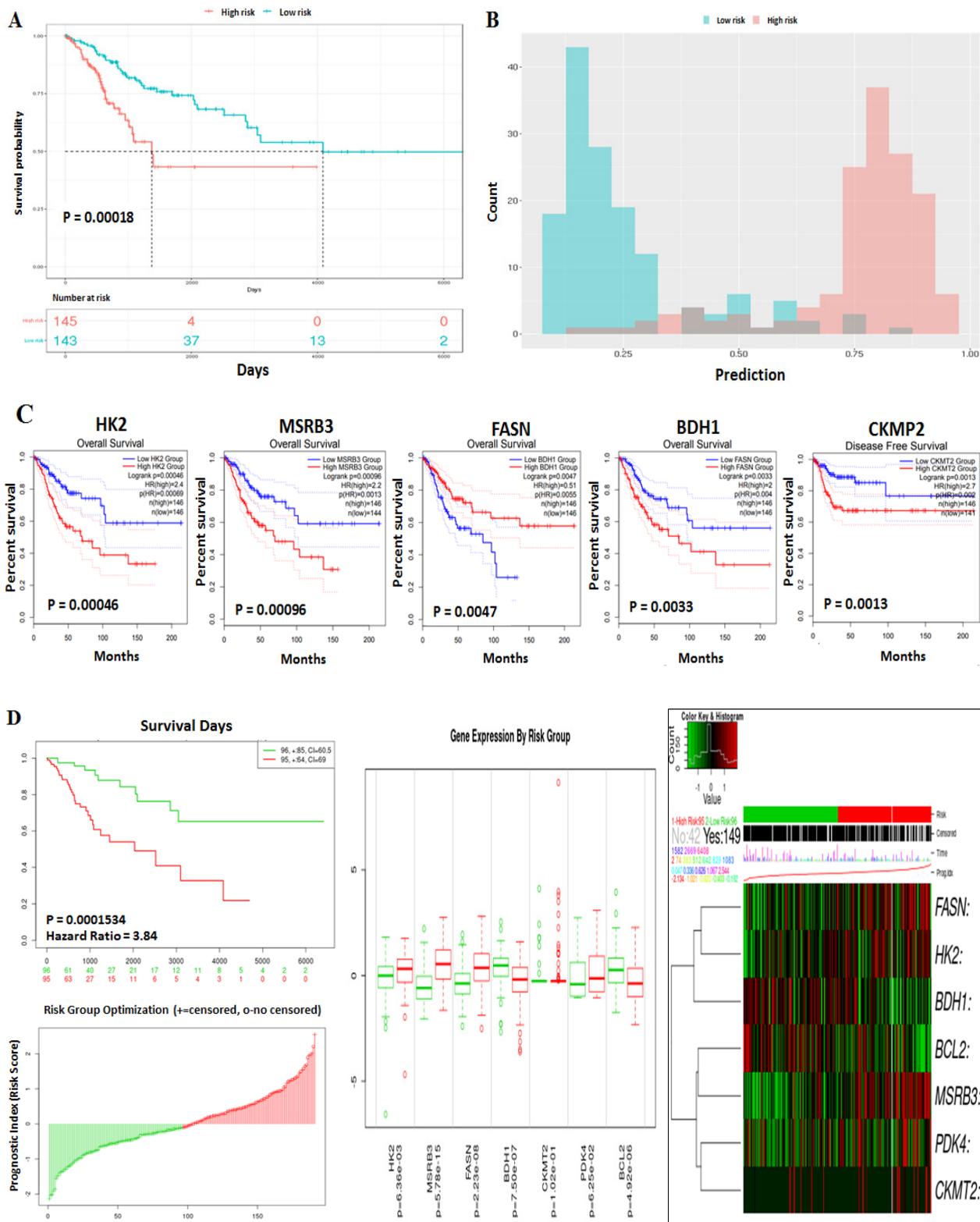


Figure 4. Survival Analysis of DE-NEMGs. A) The association between DE-NEMG with CC prognosis generated using random forest model. B) Represents the sensitivity and specificity of DE-NEMG. C) Represents the KM plot of DE-NEMG associated with OS and DFS. D) Represents the KM plot and risk score construction using 52 DE-NEMGs.

(BDH1, CPT1B, CKMT1A, MSRB3, MGARP, PMAIP1, TDRKH, CKMT1B, PDK4, and BCL2) were linked with head and neck, lung, or lymph node metastasis (Table 1).

PPIN and HG analysis

The PPINs of 52 DEGs were generated using the

STRING database. The PPIN of DE-NEMG displayed 52 nodes and 13 edges with a PPIN enrichment p-value of 1.56e-12 (Figure 5A). The top 10 connection proteins or HG were predicted using the CytoHubba tool and visualized via the Cytoscape tool. The CytoHubba analysis identified FASN, HK2, ACACB, PIF1, COX7A1,

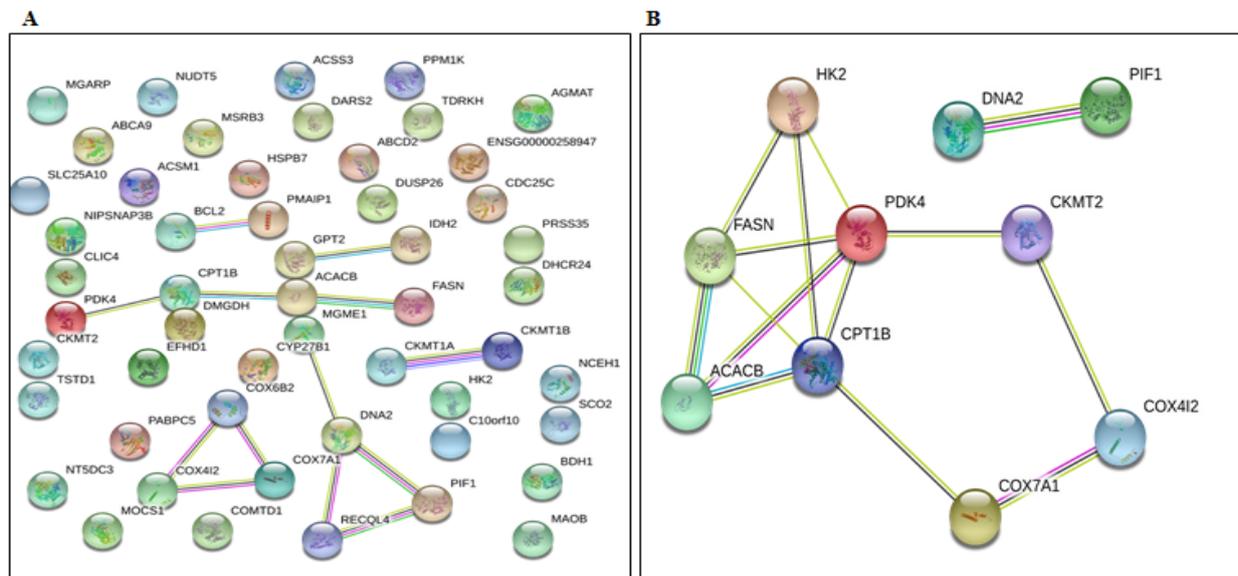


Figure 5. PPIN of DE-NEMG. A) Represents the PPIN of 52 DE-NEMGs. B) Represents the PPIN of 19 hub genes.

CKMT2, CPT1B, PDK4, DNA2, and COX4I2 as top 10 Hub genes, and these 10 proteins may play a key role in CC development or progression (Figure 5B).

Driver Gene Identification

The DriverDBv3 database has predicted 337 genes as a potential driver gene in the TCGA-CESC dataset. The comparison of 52 DE-NEMG with that of 337 identified IDH2 as a potential driver gene. In addition, the CCDB database compared 538 genes reported in CC with 52 DE-NEMG identified 2 common genes (BCL2, DUSP26). The disease enrichment analysis of 52 DE-NEMG using Comparative Toxicogenomics Database identified 17 genes (ABCA9, ABCD2, ACSM1, BCL2, CDC25C, CLIC4, CPT1B, DEPP1, EFHD1, FASN, HK2, IDH2,

MAOB, PDK4, PMAIP1, SLC25A10, and TUBB3) associated with various cancers.

Functional enrichment analysis of DE-NEMG

The GO and KEGG pathway analysis were carried out for 52 DE-NEMG using clusterProfiler to understand their function in CC (Figure 6A). The pathway enrichment using KEGG annotations identified genes related to pathways such as arginine and proline metabolism, Parkinson disease, Amyotrophic lateral sclerosis, Thermogenesis, Central carbon metabolism in cancer, Cardiac muscle contraction, Carbon metabolism, AMPK signaling pathway, Oxidative phosphorylation, and Insulin signaling pathway as highly enriched. The 52 differentially expressed NEMG linked with biological process, cellular

Table 1. NEMG and TCGA-CESC

Expression	Gene Name
Upregulated (27)	BDH1; SLC25A10; IDH2; DARS2; SCO2; COX6B2; GTP2; CPT1B; HK2; CKMT1A; TUBB3; FASN; TSTD1; RECQL4; DHCR24; NUDT5; NCEH1; PIF1; AGMAT; COMTD1; PMAIP1; TDRKH; CYP27B1; MGME1; CKMT1B; CDC25C; DNA2
Downregulated (25)	PDK4; COX7A1; MOCS1; ACSM1; CKMT2; COX4I2; DMGDH; ABCD2; ACACB; ACSS3; HSPB7; ABCA9; NT5DC3; MAOB; PPM1K; MSRB3; NIPSNAP3B; MGARP; DUSP26; EFHD1; BCL2; PABPC5; PRSS35; C10orf10; CLIC4
Methylation	
Hypermethylated (11)	CPT1B; DMGDH; CKMT2; ACSS3; COX7A1; MAOB; PDK4; PABPC5; MSRB3; MGARP; HSPB7
Hypomethylated genes (8)	CKMT1A; CKMT1B; TSTD1; AGMAT; COMTD1; SCO2; NCEH1; BCL2
Inverse correlation (Methylation vs Expression)	COX6B2; CYP27B1; CKMT1A; RECQL4; CKMT1B; PIF1; PMAIP1; TSTD1; AGMAT; GPT2; COMTD1; TDRKH; CPT1B; NT5DC3; MOCS1; DMGDH; COX4I2; CKMT2; PPM1K; ACSM1; ACSS3; COX7A1; PDK4; PABPC5; MSRB3; MGARP; HSPB7
Staging (16)	HK2; PIF1; DHCR24; SLC25A10; CPT1B; FASN; NCEH1; NIPSNAP3B; ABCD2; DMGDH; ACACB; DEPP1; ACSS3; MAOB; PDK4; PABPC5
Metastasis (10)	BDH1; CPT1B; CKMT1A; MSRB3; MGARP; PMAIP1; TDRKH; CKMT1B; PDK4; BCL2
Prognosis	
Methylation	OS: SCO2; MOCS1; DEPP1; ABCA9 DFS: DEPP1 PFS: NUDT5; DEPP1
Expression	OS: HK2; MSRB3; FASN; BDH1 DFS: CKMT2

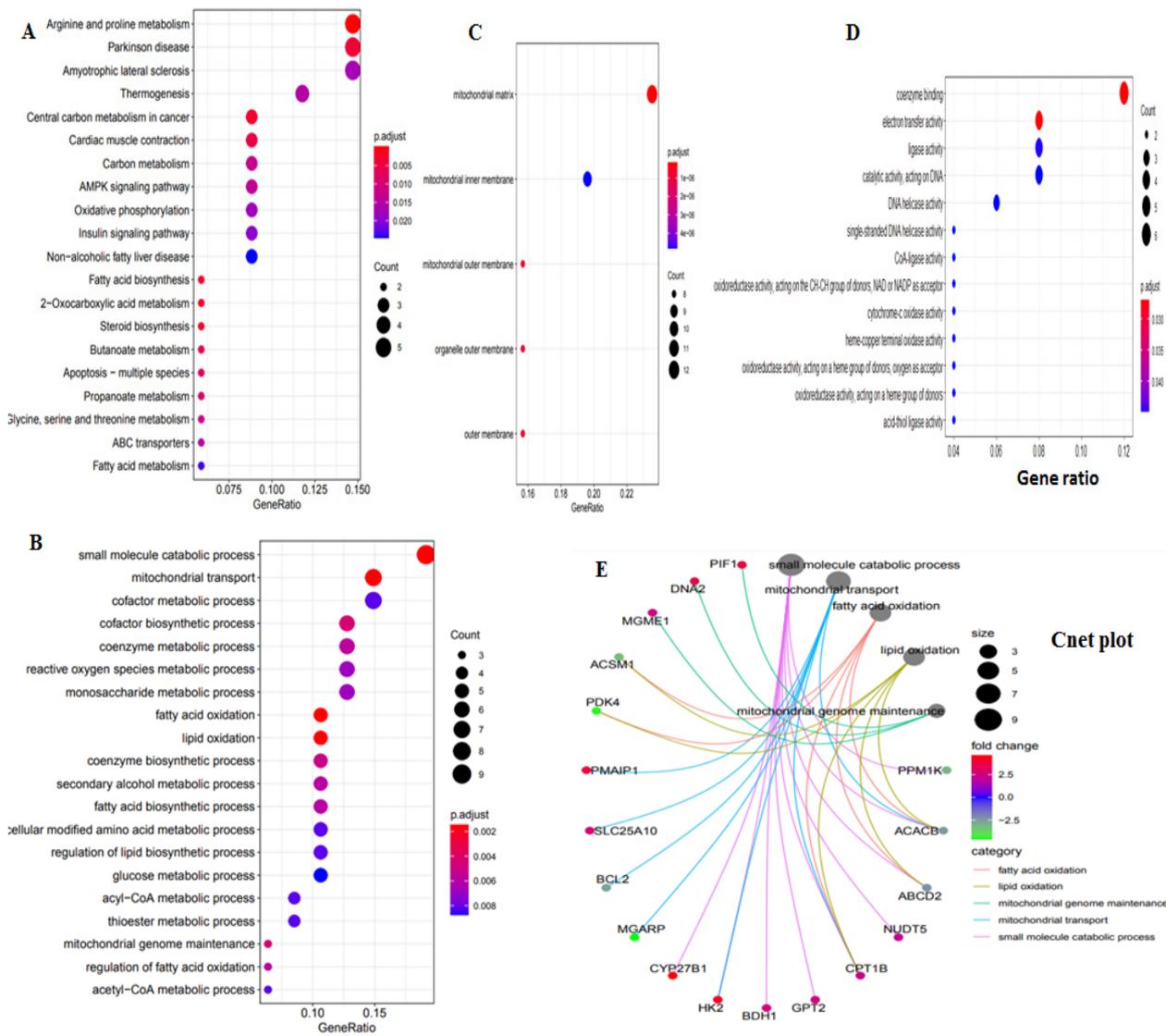


Figure 6. Functional and Pathway Enrichment Analysis of DE-NEMG. A) Signaling pathways analysis (B) biological process (C) cellular components (D) molecular components. E) Cent plot represent connection between genes and enriched ontology terms.

components, and molecular functions are shown in Figure 6B-6E.

HG cross validation in independent datasets

The HG in the PPIN were predicted using CytoHubba and visualized using Cytoscape. The top 10 HG identified by CytoHubba includes FASN, HK2, ACACB, PIF1, COX7A1, CKMT2, CPT1B, PDK4, DNA2, and COX4I2. The HG analysis using SurvExpress showed that the low-risk group patients showed significantly higher survival than high-risk group patients (Risk group hazard ratio: 3.64, p=0.0002. Independent validation of HG expression was performed using Oncomine Research Edition (<https://www.oncomine.org/resource/login.html>) online tool (Rhodes et al., 2007). Oncomine analysis identified genes, namely, HK2, PIF1, COX7A1, CPT1B, DNA2, and COX4I2, were significantly differentially expressed in more than two independent datasets. Besides, 10 hub genes' expression was also independently validated using the human protein atlas (HPA) database (<https://www.proteinatlas.org/>). Among the 10 genes,

immunohistochemical data were available for only 8 genes (FASN, HK2, ACACB, PIF1, CKMT2, CPT1B, PDK4, and DNA2). The expression of the 8 genes was in concordance with TCGA-CESC expression data (Figure 7, Supplementary Table 5).

Identification of Drugs targeting NEMG

The potential druggable candidates in the 52 DE-NEMGs were predicted using the drug-gene interaction database (DGIdb; <https://www.dgiddb.org/>). The DGI was evaluated for 52 DE-NEMG against approved, anti-neoplastic, and immunotherapeutic agents. Our analysis identified 195 drugs targeting 16 DE-NEMG (Supplementary Table 6). Drugs such as Cisplatin, Fluorouracil, Paclitaxel, Docetaxel, Vincristine, Adriamycin, and Doxorubicin are already being used to treat various cancers. Besides, our analysis also identified several novel drugs that can be repurposed to treat CC.

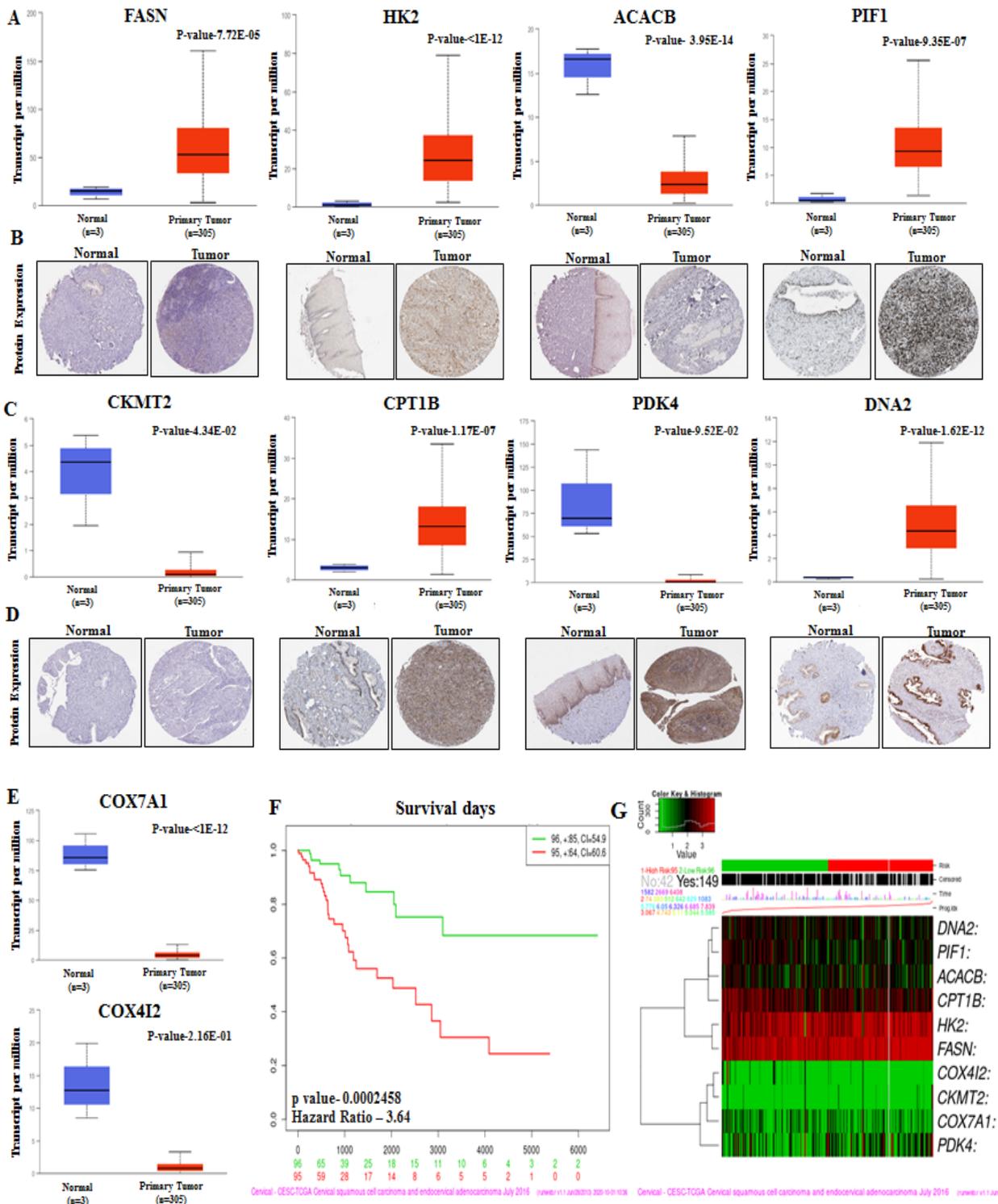


Figure 7. Validation of the HG Expression at Protein Level. A) The bar graph and IHC images represents the expression of HG at RNA and protein level. B) Survival analysis of hub genes performed by using SurvExpress. C) Heatmap showing the hub gene expression.

Discussion

CC is one of the common gynecological problems in many developing and underdeveloped countries (Arbyn et al., 2020). Various risk factors such as infection with high-risk HPV, smoking, alcohol consumption, multiple sexual partners, parity, and oral contraceptives are well recognized to influence CC’s development and progression (Johnson et al., 2019; Wudtisan et al., 2019). Despite the

availability of early diagnostic procedures and advances in CC treatment, there is little improvement in the disease prognosis. The detection of the disease at an advanced stage contributes to poor clinical outcomes and high mortality (Canfell et al., 2020). Besides, only limited prognostic markers are available for clinical use in CC. Hence, it is indispensable to explore new prognostic markers and additional target genes for CC treatment.

Mitochondria play a critical role in bioenergetics

and cellular metabolism. Besides, mitochondria also participate in various signaling, programmed cell death, heme, and steroid synthesis (Shaughnessy et al., 2015). In humans, dysregulation of mitochondria's structure and function is linked with all stages of cancer progression. Additionally, studies have reported the association between pathogenic mtDNA variations with CC (Sun et al., 2020). Moreover, mitochondrial defects can promote tumor growth and metastasis by providing energy to growing cancer cells, promoting cell survival via activation of anti-apoptotic signaling pathways. Both anterograde and retrograde signaling is critical for normal homeostasis of the cells. Defective NEMG signaling contributes and controls several aspects of carcinogenesis, such as growth, proliferation, migration, and metastasis. The abnormal expression of NEMG has been reported as a potential diagnostic and prognostic marker and a novel target for therapy in various cancers (Chatterjee et al., 2006; Indran et al., 2011). However, such studies are limited in CC. Hence, a comprehensive investigation of NEMG may significantly improve CC's management by providing markers for diagnosis, prognosis, and treatment. Accordingly, in the present study, we identified the NEMG with potential prognostic applications in CC.

Cancer staging is vital for designing the treatment and predicting the prognosis. Towards this, using the UALCAN database, we predicted the potential of 52 DE-NEMG for CC staging. The 52 genes were significantly differentially expressed between normal and tumor samples, suggesting their potential in CC diagnosis. However, out of 52 DEGs, the differential expression of 16 genes could differentiate specific stages of CC. Thus, our study has identified a panel of genes whose expression profiling could be useful for staging CC and required to be evaluated further in clinical samples.

Metastasis is a critical factor for therapeutic resistance and poor patient survival. Thus, identifying the metastasis-associated genes is proposed to improve patient care. Towards this, we investigated the potential of 52 DE-NEMG and their contribution towards the induction of metastasis in CC. Among the 52 DE-NEMG, 10 genes have been linked to head and neck, lung, or lymph node metastasis in CC. Among the various metastatic genes identified, elevated expression of CPT1B is reported in chemoresistant metastatic breast tumors (Wang et al., 2018). An axis comprising of lncRNA n335586/miR-924/CKMT1A is reported in metastasis of hepatocellular carcinoma cells (Fan et al., 2018). Overexpression of MSRB3 is reported in peritoneal metastasis and poor prognosis in gastric cancer (Zhang et al., 2020). The role of PDK4 in colon cancer metastasis is also reported. The PDK4 mediated metabolic reprogramming contributes significantly to the metastasis cascade (Leclerc et al., 2017). BCL2 promotes and accelerates metastasis in breast and colorectal cancer (Um, 2016). In-silico analysis using HCMDDB identified 10 proteins associated with CC metastasis based on clinical correlation analysis. Many of these genes are either directly or indirectly participating in metabolic reprogramming and/or adaptation to promote cancer cells' growth and survival. However, more functional studies are required to identify the precise role

of these genes in metastasis and metabolic reprogramming in CC.

The role of abnormal DNA methylation in carcinogenesis is well established. Both promoter hypermethylation resulting in tumor suppressor gene silencing and hypomethylation activating oncogenes are reported to promote CC. Previous studies have indicated that measuring DNA methylation has the potential diagnostic and prognostic value in CC (Kabekkodu et al., 2014). Herein, we investigated the association between methylation and expression via correlation analysis. Our analysis identified 19 gene promoters were significantly differentially methylated between normal and tumor samples. Interestingly, 15 out of 19 NEMG showed an inverse correlation between methylation and expression. Moreover, our analysis recommends the use of SCO2, MOCS1, DEPP1, and ABCA9 for OS, and NUDT5, DEPP1 for PFS in CC after further validation.

A PPIN was constructed to better understand the interaction and functions of 52 DE-NEMG. PPIN identified 52 nodes and 13 edges, suggesting that these proteins' interactions may have a critical role in CC development and progression. CytoHubba analysis identified FASN, HK2, ACACB, PIF1, COX7A1, CKMT2, CPT1B, PDK4, DNA2, and COX4I2 as critical members of the network or HG. Our study's 10 HG were linked to fatty acid biosynthesis, AMPK signaling, insulin signaling, oxidative phosphorylation, and metabolic process. Abnormal activation or suppression of these pathways has already been linked to various cancers, including CC. The hub genes' cross-validation using the OncoPrint tool in 5 independent data sets identified HK2, PIF1, COX7A1, CPT1B, DNA2, and COX4I2 as commonly altered in CC. Thus, abnormal expression of these genes might be critical for CC pathogenesis. The details of these hub genes commonly identified in multiple datasets are described below.

The conversion of glucose to glucose-6-phosphate is mediated by HK2 (Hexokinase 2) enzyme. Cancer cells activate glycolytic enzymes to promote metabolic changes to support the growth and proliferation of cancer cells. HK2 is often upregulated in numerous cancers, including CC. HK2 overexpression promotes proliferation, migration, and inhibition of apoptosis in SiHa cells via phosphorylation of AKT. High HK2 expression correlates with the size of the tumor, pathological grade, and prognosis (Liu et al., 2019). Inhibition of HK2 sensitized the CC cells to radiation and induced apoptosis via caspase-3 and PARP cleavage activation. Thus, targeting HK2 can be used to sensitize cancer cells to radiotherapy and control tumor cells' growth and proliferation (Liu et al., 2017). PIF1 (5'-To-3' DNA Helicase) encodes for a protein with 5' to 3' DNA helicase function required to maintain nuclear and mtDNA genome stability. PIF1 acts as a tumor promoter in CC via suppressing the TERT. CC cells show a high expression of PIF1 and promote proliferation by inhibiting apoptosis by targeting BAX and Caspase-3 (Wang et al., 2020).

COX7A1 (Cytochrome C Oxidase Subunit 7A1) takes part in a biochemical reaction involving electron transfer activities during the OXPHOS reaction. COX7A1

acts as a tumor suppressor by inhibiting non-small cell lung cancer's growth and proliferation by induction of apoptosis via blocking autophagic flux (Zhao et al., 2019). Another study demonstrated the downregulation of COX7A1 by promoter hypermethylation in breast cancer (He et al., 2019). Besides, COX7A1 is implicated in tumor metabolism and therapy. However, their precise role in CC is yet to be established. CPT1B (Carnitine Palmitoyltransferase 1B) is a rate-controlling enzyme that participates in the transfer of long-chain fatty acyl-CoAs from the cytoplasm to mitochondria. Yeh and colleagues in 2006 demonstrated the overexpression of this gene in colorectal cancer (Yeh et al., 2006). In muscle-invasive bladder cancer, CPT1B deregulation correlated with a higher rate of mortality (Kim et al., 2016). In our study, we found overexpression of CPT1B in CC. This suggests that the overexpression of CPT1B leading to metabolic alterations may be a prerequisite for cancer cells' growth and survival. In gastrointestinal cancer, higher expression of CPT1B displayed a superior response to Carnitine palmitoyltransferase inhibitors (Wang et al., 2020). Furthermore, targeting CPT1B in cancers with abnormal lipid metabolism and fatty acid oxidation (castration-resistant prostate cancer) is proposed. CPT1B overexpression is correlated with poor prognosis in prostate cancer (Abudurexiti et al., 2020). However, the role of CPT1B in CC is yet to be uncovered.

DNA2 (DNA Replication Helicase/Nuclease 2) is a protein required to maintain mitochondrial and nuclear DNA stability by participating in the replication and repair process. Similar to our findings, Li et al. 2018 reported the upregulation of DNA2 in CC (Li et al., 2018). However, the molecular mechanism and pathways regulated by DNA2 in CC are still elusive. Overexpression of DNA2 is reported in multiple human cancers. By counteracting the replication stress, DNA2 can act as a cancer promoter (Zheng et al., 2020). Targeting DNA2 is proposed as a therapeutic target to control tumor growth in pancreatic cancer (Kumar et al., 2017). Thus, DNA2 may act as a potential driver of carcinogenesis. COX4I2 (Cytochrome C Oxidase Subunit 4I2) is an enzyme that drives oxidative phosphorylation and is downregulated in our analysis. Hypoxic condition induces COX4I2 expression as it harbors hypoxia-responsive elements. In HeLa cells, hypoxia enhances the promoter activity of COX4I2 (Fukuda et al., 2007).

Therapy resistance is one of the major problems in cancer treatment. To overcome the drug resistance, researchers are looking for new targets and repurposing of the existing drugs. Herein, we have tested the potential druggable DE-NEMG using the DGIdb database. Our study leads to the identification of 16 druggable genes and 195 drugs. Based on our analysis, we propose the use of a combination of drugs for treating CC. The various known drugs identified in our study, which are already in use for treating cancer in general and CC in particular, include Cisplatin, Fluorouracil, Paclitaxel, Docetaxel, Vincristine, Adriamycin, and Doxorubicin.

The present study has some limitations. CC has several histological types. The data used in the study included all histological types without performing any subgroup

analysis. Besides, our survival analysis also contained data for all histological types of CC without performing subgroup analysis. Lack of experimental validation is another limitation of our study.

Nevertheless, our study is the first comprehensive study investigating the role of DE-NEMG as potential markers for prognostic application in CC. Our *in silico* investigation identified new insights into the interaction of NEMG and associated pathways regulated during CC. Although we have analyzed many tumor samples, the normal samples included are still inadequate, suggesting the need for further experimental validation before further conclusions are drawn. Besides, several genes reported as significantly differentially expressed in TCGA-CESC data require further validation in patient cohort and functional studies using *in vitro* and *in vivo* models.

In the present study, we assessed transcriptomic and corresponding clinical data from the TCGA-CESC dataset and identified 52 significantly DE-NEMGs. The prognostic utility, functional and pathway enrichment, and PPIN were performed using *in silico* methods. The PPIN constructed suggested the novel PPIN in CC. The findings of our study have potential prognostic implications in CC. Our DGI analysis identified novel drugs that can be repurposed to treat CC. The druggable genes identified in our study may facilitate the development of specific and more effective treatments against CC.

Author Contribution Statement

SPK designed the study; CM and SE acquired, analyzed, and interpreted the data; DA, NKG and SS drafted the initial manuscript. SM, SC and SPK helped in the critical revisions of the manuscript. All the authors have read and approved the final draft of the manuscript.

Acknowledgments

We thank Department of Science and Technology (DST), Government of India (Grant No: EMR/2016/002314), Science and Engineering Research Board (SERB), DST-Ph.D. fellowship, KSTePS, DST, Government of Karnataka (Reference ID- DST/KSTePS/Ph.D. Fellowship/LIF-11: 2019-20), ICMR-Senior Research fellowship, Government of India (Reference ID- 2019/4115/CMB/BMS and 2020/8704/CMB/BMS), Directorate of Minorities Fellowship, Government of Karnataka (DOM/FELLOWSHIP/CR-10/2019-20), and Dr. TMA Pai Structured PhD fellowship, MAHE, for financial support. We acknowledge Manipal Academy of Higher Education (MAHE) for infrastructure support.

Funding statement

This work is supported by Department of Biotechnology, Government of India under pilot project on cancer (Sanction number: 6242-P8/RGCB/PMD/DBT/SPDK/2015).

The study includes secondary data analysis, therefore doesn't require approval from scientific body. The present study was a part of student thesis (Ms. Cecile Maneur, UG Student, La Rochelle University, France).

Ethical approval was not required for the present study. The study data was retrieved and synthesized from the already published studies (secondary data analysis).

Conflict of interest

All the authors declare no conflict of interest.

References

- Abudurexiti M, Zhu W, Wang Y, et al (2020). Targeting CPT1B as a potential therapeutic strategy in castration-resistant and enzalutamide-resistant prostate cancer. *Prostate*, **80**, 950–61.
- Agarwa SM, Raghav D, Singh H, Raghava GPS (2011). CCDB: A curated database of genes involved in cervix cancer. *Nucleic Acids Res*, **39**, 975–9.
- Aguirre-Gamboa R, Gomez-Rueda H, Martínez-Ledesma E, et al (2013). SurvExpress: An online biomarker validation tool and database for cancer gene expression data using survival analysis. *PLoS One*, **8**, e74250.
- Arbyn M, Weiderpass E, Bruni L, et al (2020). Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob Heal*, **8**, 191–203.
- Bray F, Ferlay J, Soerjomataram I (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, **68**, 394–424.
- Calvo SE, Clauser KR, Mootha VK (2016). MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res*, **44**, 1251–7.
- Canfell K, Kim JJ, Brisson M, et al (2020). Mortality impact of achieving WHO cervical cancer elimination targets: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *Lancet*, **395**, 591–603.
- Chandrashekar DS, Bachel B, Balasubramanya SAH (2017). UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*, **19**, 649–658.
- Charakorn C, Thadanipon K, Chaijindaratana S, et al (2018). The association between serum squamous cell carcinoma antigen and recurrence and survival of patients with cervical squamous cell carcinoma: A systematic review and meta-analysis. *Gynecol Oncol*, **150**, 190–200.
- Chatterjee A, Mambo E, Sidransky D, (2006). Mitochondrial DNA mutations in human cancer. *Oncogene*, **25**, 4663–74.
- Chin CH, Chen SH, Wu HH, et al (2014). cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*, **8**, 1–7.
- Chou PH, Liao WC, Tsai KW, et al (2019). TACCO, a database connecting transcriptome alterations, pathway alterations and clinical outcomes in cancers. *Sci Rep*, **9**, 1–10.
- Cotto KC, Wagner AH, Feng YY, (2018). DGIdb 3.0: A redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res*, **46**, 1068–73.
- Davis AP, Grondin CJ, Johnson RJ, et al (2019). The comparative toxicogenomics database: Update 2019. *Nucleic Acids Res*, **47**, 948–54.
- Ding W, Chen J, Feng G, et al (2020). DNMIIVD: DNA methylation interactive visualization database. *Nucleic Acids Res*, **48**, 856–62.
- Fan H, Lv P, Mu T, et al (2018). LncRNA n335586/miR-924/CKMT1A axis contributes to cell migration and invasion in hepatocellular carcinoma cells. *Cancer Lett*, **429**, 89–99.
- Ferlay J, Ervik J, Lam F, et al (2018). Global Cancer Observatory: Cancer Today. Lyon, Fr. Int. Agency Res. Cancer.
- Fukuda R, Zhang H, Kim JW, et al (2007). HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells. *Cell*, **129**, 111–22.
- He Z, Wang F, Zhang W, et al (2019). Comprehensive and integrative analysis identifies COX7A1 as a critical methylation-driven gene in breast invasive carcinoma. *Ann Transl Med*, **7**.
- Indran IR, Tufo G, Pervaiz S, Brenner C (2011). Recent advances in apoptosis, mitochondria and drug resistance in cancer cells. *Biochim Biophys Acta Bioenerg*, **1807**, 735–45.
- Johnson CA, James D, Marzan A, Armaos M, (2019). Cervical Cancer: An Overview of Pathophysiology and Management. *Semin Oncol Nurs*, **35**, No. 2, pp 166–74.
- Kabekkodu SP, Bhat S, Radhakrishnan R, et al (2014). DNA promoter methylation-dependent transcription of the double C2-like domain β (DOC2B) gene regulates tumor growth in human cervical cancer. *J Biol Chem*, **289**, 10637–49.
- Kanehisa M (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, **28**, 27–30.
- Kim WT, Yun SJ, Yan C, et al (2016). Metabolic pathway signatures associated with urinary metabolite biomarkers differentiate bladder cancer patients from healthy controls. *Yonsei Med J*, **57**, 865.
- Kumar S, Peng X, Daley J, et al (2017). Inhibition of DNA2 nuclease as a therapeutic strategy targeting replication stress in cancer cells. *Oncogenesis*, **6**, 319.
- Leclerc D, Pham DNT, Lévesque N, et al (2017). Oncogenic role of PDK4 in human colon cancer cells. *Br J Cancer*, **116**, 930–6.
- Li X, Tian R, Gao H, et al (2018). Identification of significant gene signatures and prognostic biomarkers for patients with cervical cancer by integrated bioinformatic methods. *Technol Cancer Res Treat*, **17**, 1533033818767455.
- Lin M, Ye M, Zhou J, Wang ZP, Zhu X (2019). Recent advances on the molecular mechanism of cervical carcinogenesis based on systems biology technologies. *Comput Struct Biotechnol J*, **17**, 241–50.
- Liu C, Wang X, Zhang Y (2019). The roles of HK2 on tumorigenesis of cervical cancer. *Technol Cancer Res Treat*, **18**, 1533033819871306.
- Liu SH, Shen PC, Chen CY, et al (2020). DriverDBv3: A multi-omics database for cancer driver gene research. *Nucleic Acids Res*, **48**, 863–70.
- Liu Y, Murray-Stewart T, Casero RA, et al (2017). Targeting hexokinase 2 inhibition promotes radiosensitization in HPV16 E7-induced cervical cancer and suppresses tumor growth. *Int J Oncol*, **50**, 2011–23.
- Shannon P, Markiel A, Ozier O, et al (1971). Cytoscape: A Software Environment for Integrated Models. *Genome Res*, **13**, 426.
- Pfanner N, Warscheid B, Wiedemann N (2019). Mitochondrial proteins: from biogenesis to functional networks. *Nat Rev Mol Cell Biol*, **20**, 267–84.
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al (2007). OncoPrint 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*, **9**, 166–80.
- Shaughnessy DT, McAllister K, Worth L, et al (2015). Mitochondria, energetics, epigenetics, and cellular responses to stress. *Environ Health Perspect*, **122**, 1271–8.
- Sun W, Qin X, Zhou J, et al (2020). Mitochondrial DNA copy number in cervical exfoliated cells and risk of cervical cancer among HPV-positive women. *BMC Womens Health*, **20**, 1–7.
- Szklarczyk D, Morris JH, Cook H, et al (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*, **45**, 362–8.
- Tang Z, Kang B, Li C, Chen T, Zhang Z (2019). GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*, **47**, 556–60.

- Thul PJ, Lindskog C (2018). The human protein atlas: A spatial map of the human proteome. *Protein Sci*, **27**, 233–44.
- Tomczak K, Czerwińska P, Wiznerowicz M (2015). The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Wspolczesna Onkol*, **1**, 68–77.
- Um HD (2016). Bcl-2 family proteins as regulators of cancer cell invasion and metastasis: A review focusing on mitochondrial respiration and reactive oxygen species. *Oncotarget*, **7**, 5193.
- Wang J, Zhu X, Ying P, Zhu Y (2020). PIF1 affects the proliferation and apoptosis of cervical cancer cells by influencing tert. *Cancer Manag Res*, **12**, 7827.
- Wang T, Fahrman JF, Lee H, et al (2018). JAK/STAT3-regulated fatty acid β -Oxidation is critical for breast cancer stem cell self-renewal and chemoresistance. *Cell Metab*, **27**, 136-50.
- Wang Y, Lu JH, Wang F, et al (2020). Inhibition of fatty acid catabolism augments the efficacy of oxaliplatin-based chemotherapy in gastrointestinal cancers. *Cancer Lett*, **473**, 74-89.
- Wu X, Peng L, Zhang Y, et al (2019). Identification of key genes and pathways in cervical cancer by bioinformatics analysis. *Int J Med Sci*, **16**, 800–812.
- Wudtisan J, Tantipalakorn C, Charoenkwan K, Sreshtaputra RA, Srisomboon J (2019). Factors associated with development of high-grade squamous intraepithelial lesions of the uterine cervix in women younger than 30 years. *Asian Pacific J Cancer Prev*, **20**, 1031.
- Yeh CS, Wang JY, Cheng TL, et al (2006). Fatty acid metabolism pathway play an important role in carcinogenesis of human colorectal cancers by Microarray-Bioinformatics analysis. *Cancer Lett*, **233**, 297-308.
- Yu G, Wang LG, Han Y, He QY (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi A J Integr Biol*, **16**, 284-7.
- Zhang S, Zang D, Cheng Y, et al (2020). Identification of key gene and pathways for the prediction of peritoneal metastasis of gastric cancer by co-expression analysis. *J Cancer*, **11**, 3041.
- Zhao L, Chen X, Feng Y, et al (2019). COX7A1 suppresses the viability of human non-small cell lung cancer cells via regulating autophagy. *Cancer Med*, **8**, 7762-73.
- Zheng G, Ma Y, Zou Y, et al (2018). HCMDB: The human cancer metastasis database. *Nucleic Acids Res*, **46**, 950–5.
- Zheng L, Meng Y, Campbell JL, Shen B (2020). Multiple roles of DNA2 nuclease/helicase in DNA metabolism, genome stability and human diseases. *Nucleic Acids Res*, **48**, 16–35.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.