J Genet Resour2023;9(1): 92-102 *Homepage: http://sc.journals.umz.ac.ir/*

RESEARCH ARTICLE *DOI: 10.22080/jgr.2023.25032.1344*

Weighted Gene Co-expression Network Analysis Unveiled Key Genes Related to Progression and Prognosis of Cervical Cancer

Mina Dehghani-Samani and Modjtaba Emadi-Baygi*

Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran

ARTICLE INFO

Article history:
Received 26 September 2022
Accepted 03 November 2022
Available online 24 January 2023

Keywords:
Cervical cancer
Differentially expressed genes
PCP4
TCGA
WGCNA

****Corresponding authors:***
✉ M Emadi-Baygi
emadi-m@sku.ac.ir

p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

Cervical cancer is one of the most common malignancies and one of the main death causes among females all over the world. The discovery of tumor-related genes is crucial for understanding tumor biology and developing preventative and therapeutic strategies. However, genes included in the tumorigenesis of cervical cancer cells are still unclear. Due to its high prevalence and mortality, understating its pathogenesis and biomarker detection is necessary. The purpose of this study was to recognize potential biomarkers related to cervical cancer and analyze their prognostic significance. The present research used the level 3 mRNA expression data and clinical data of cervical cancer from The Cancer Genome Atlas database to identify differentially expressed genes followed by gene ontology. Weighted co-expression Network Analysis was used to construct scale-free gene co-expression networks. In the co-expression network, the hub module and hub genes were identified. The significant modules associated with T, N, M, and FIGO staging in the network were subsequently screened. Next, overlapping genes between significant gene modules and DEGs were screened. *CALML5*, *TERT*, *PNPLA3*, *CHRDL1*, *C7*, *LEFTY2*, and *PCP4* were identified as hub genes. Survival analysis was performed to identify the association between these genes and survival using the GEPIA database. Survival analysis showed that *PCP4* was slightly less expressed in patients with primary solid tumors than normal, and related to poor prognosis in cervical cancer. These results show that these hub genes, especially *PCP4*, may be a potential diagnostic biomarker for cervical cancer and provide a new perspective on the pathogenesis of cervical cancer.

Please cite this paper as: Dehghani-Samani M, Emadi-Baygi. M. 2023. Weighted gene co-expression network analysis unveiled key genes related to progression and prognosis of cervical cancer. *J Genet Resour* 9(1): 92-102. doi: *10.22080/jgr.2023.25032.1344*

Introduction

Cervical cancer is the second cause of mortality in developed countries and the third most prevalent cancer among females (Petry, 2014). Annually, 500000 women are detected with cervical cancer all over the world, causing 200000 deaths (Sharma *et al*., 2020). Smoking, having multiple sexual partners, human papillomavirus, hereditary, and birth control pills/contraception are known as risk factors for cervical cancer (Wang and Chen, 2019). Currently, due to an available appropriate screening program, efficient treatment of primary tumors, and requiring prolonged time to show invasive features, cervical cancer is preventable cancer among other malignancy types (Wright *et al*., 2015). Although cervical cancer has not been completely eradicated globally, the mortality rate has decreased in the past years due to early detection (Boeker *et al*., 2016). Clinically, the pap smear is the best and foremost method for early detection of cervical cancer. Regularly, screening women between the ages of 21 and 65 with a Pap test may reduce the risk of death in patients with cervical cancer (Orlov and Baranova, 2020).

The discovery of tumor-related genes is crucial for understanding tumor biology and developing preventative and therapy strategies (Tamura, 2006). Investigating cervical cancer at the gene expression level gives crucial information for pathogenesis research, early detection, and therapy (Ibeanu, 2011). The advancements in biomarkers identified by high throughput genomic sequencing and computational analysis have garnered increasing attention during the last two decades (Chen *et al*., 2019). Computational approaches have been turned into a standard tool for detecting differences in transcriptome across various samples on a high-throughput platform that can identify novel differentially expressed genes (DEGs) from big data analysis. Several public databases such as The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and ArrayExpress are available for the analysis of microarray/NGS datasets that will provide useful information for the study of various disorders (Edgar and Barrett, 2006; Tomczak *et al*., 2015).

The phenotypic variety of complex life is frequently mediated by the interplay of numerous genes in the gene system, rather than by a single gene. Weighted Gene Co-expression Network Analysis (WGCNA) is a method that can be employed to find groups or clusters of genes that are closely correlated with each other, often known as modules, and to study the connection between modules and external sample variables. This allows for the discovery of hub genes that might be used as therapeutic targets. Hub genes in WGCNA are those that are highly related to a module or cluster of genes. These hub genes are thought to be crucial in controlling the module's function and are frequently identified as possible therapeutic targets or biomarkers for illness diagnosis. Hub genes are important in WGCNA because they are crucial actors in the molecular network, and knowing their involvement in gene regulation might give insights into disease pathophysiology and potential therapeutic approaches (Peter Langfelder, 2008). Up to this point, WGCNA has been used effectively in a variety of biological domains, including cancer, (Di *et al*., 2019; Yin *et al*., 2021; Zhai *et al*., 2017). Using WGCNA, Chen *et al*. discovered the major pathogenic gene Abnormal Spindle-like, Microcephaly-associated (ASPM) for glioblastoma, which has since become a critical gene for treatments (Chen *et al*., 2020). In the present investigation, 304 cervical cancer samples from the TCGA dataset were used to perform a WGCNA and determine module-trait correlations. Using this method, co-expression modules that were relevant and strongly correlated with tumor grade and stage were discovered. Through KEGG enrichment analysis and hierarchical clustering analysis, hub genes were also found. These hub genes may be used as cervical cancer biomarkers for diagnosis and prognosis and might be the focus of cutting-edge treatments. Figure 1 shows the pipeline used in this investigation.

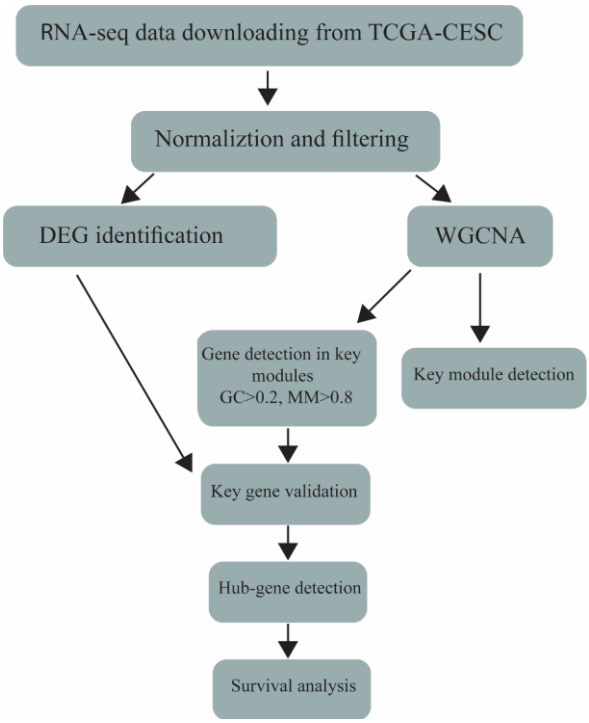

Fig. 1. Flowchart of the present study.

Materials and Methods

Data collection and preprocessing

Data related to gene expression of cervical cancer patients were obtained from the TCGA database. The data and related clinical profiles of three normal samples as well as 304 initial tumors and 2 recurrent ones were downloaded using the Genomic Data Commons (GDC) data transfer tool (https://portal.gdc.cancer.gov),

normalized, and preprocessed through TCGABiolinks pipeline (v.2.8.4) in the R software. The TCGA analysis Preprocessing function with a correlation cut-off of 0.6 (Correlation coefficient) was used to look for potential outliers in mRNA transcript data. Furthermore, TCGABiolinks provide the option to utilize normalization methods using the function TCGA analyze Normalization, to conduct between-lane normalization or within-lane normalization. Moreover, by using the function of TCGA analyze-filtering which benefits from quantile normalization, the genes were filtered by cutoff criteria of 0.25.

DEG analysis

For detecting total differential expression of genes (DEG) between primary tumor samples and recurrent samples, the edgeR package, DESeq2 package, and limma package were employed to evaluate the variations in gene expression between samples of normal cervical tissue and cancerous samples. The final DEGs for cervical cancer were chosen from the studied genes with the following inclusion criteria |log2fold change (FC)| >1 and a threshold P. value < 0.05.

Gene enrichment analysis

The cluster profiler package in R was employed for gene ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. The goals of this research were to investigate and demonstrate the biological function of the DEGs that were found. A layout of a volcano was used to illustrate the results at the end (Fig. 2).

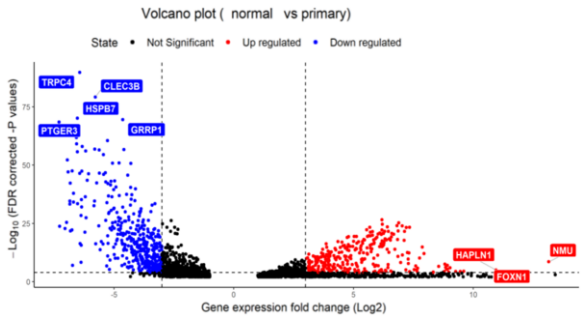

Fig. 2. Volcano plot between DEGs between primary tumor and normal samples.

WGCNA construction and GO analysis

The initial step in constructing the similarity matrix of gene expression levels was to calculate the Pearson correlation coefficient, expressed as an absolute value, between any two genes. After that step, the gene expression matrix was converted into an adjacency matrix that included a signed network type and exponent. The adjacency matrix Aij is understood to represent the contiguity of the genes I and j, with Sij standing for the Pearson's correlation that exists between the two genes. The strength of positive correlations may be increased and the strength of negative correlations can be decreased when the exponential function is applied to the Pearson correlation coefficient for each pair of genes.

In other words, the correlation coefficient may be transformed using the exponential function to strengthen positive correlations and weaken negative correlations. This is due to the exponential function's ability to "boost" weak positive correlations to greater ones while "dampening" strong negative correlations to weaker ones. Afterward, an adjacency matrix was used to construct a topological matrix. A topological overlap matrix (TOM) was utilized to characterize the degree to which different genes are connected. A method known as dynamic shear was used to carry out a hierarchical clustering analysis, which was then utilized to identify further modules. It performs this by displaying, in an iterative manner, the splitting and grouping of clusters in a hierarchical tree. A dendrogram could be used to offer an initial evaluation and illustrate the differences between different subclusters. Gene significance (GS) is a quantitative metric used to evaluate the link between genes' biological importance and external data. A greater GS often denotes a stronger correlation between the gene and the outside information. GS is a critical statistic that is used in many research areas, including genetics, genomics, and bioinformatics, to identify and rank genes that are most likely to have biological significance or to be connected to diseases. In the case where GS equals zero, the gene will not be involved in the biological process.

The present study developed a scale-free network by using WGCNA as a systematic

biological technique to search the gene expression dataset for potential gene connections. Based on the target genes that were gathered in the phase before, this research generated a WGCNA and discovered co-expression modules in R by utilizing the WGCNA function (https://www.r-project.org/), which was based on the WGCNA function. Within the context of the co-expression network, the term "module eigengene" refers to the first principal component (PC) of each module's gene expression matrix.

This PC reflects the largest percentage of variance for the expression values of all module genes within a sample, and thus is referred to as the "module eigengene." A few changes were made to the cervical cancer dataset before the signed weighted network analysis was performed. In the end, GO analysis was performed on each discovered hub module by utilizing the DAVID web tool (https://david.ncifcrf.gov/).

Survival analysis of detected hub genes

In this study, the researchers downloaded the RNA-Seq datasets that contained clinical follow-up information to perform a further study on the prognosis associated with high-abundance genes in cervical cancer. After doing so, FPKM values were transformed into TPM values. The datasets had a combined 304 cervical cancer tissue samples. An online program known as GEPIA (http://gepia.cancer-pku.cn/) was used to mine the change in expression of cancer-specific genes and prognosis using data obtained from The Cancer Genome Atlas. In this regard, the overall survival curve of hub genes was investigated by employing the aforementioned online platform, and a p-value of less than 0.05 was employed to establish the importance of the data.

Results

Data preprocessing and filtering

First, RNA-seq data including two metastatic samples, three normal samples, and 304 primary cervical tumor samples were downloaded from the TCGA database. These separate files were transformed into an integrated matrix (column = samples and rows= genes) through the Summarized Experiment package.

After data preprocessing, a total of 19947 genes were identified, which decreased to 19866 and 14899 genes after data normalization and filtering, respectively.

Normal and primary tumor sample analysis

By using the TCGA Bio links package, DEGs were identified between primary tumor and normal samples. In this regard, 2557 genes were filtered, out of which 1274 and 1283 genes were upregulated and downregulated, respectively, which are shown in the volcano plot in Figure 2. Gene ontology analysis (GO) including cellular components, molecular function, and the biological process was implemented through Cluster Profiler Package.

Weighted gene co-expression network

After filtering (14899 genes), the obtained data were transformed into log-based values. The top 5000 genes were selected from 14899 genes based on the median absolute deviation criterion. This research used the Good Sample Genes function to impute missing values but did not select genes with high missing values for sampling. Clusters of genes were visualized through the Flash Clust function. In this study, four important clinical data including tumor, node, metastasis, and FIGO-staging were downloaded from TCGA. Then, samples were clustered through the average method and Flash Clust function. Furthermore, the numbers2colors function was used to give color to each chosen clinical characteristic. WGCNA's pick Soft Threshold function is calculated between 1 to 20 the power variable. A weighted co-expression network was constructed based on a signed hybrid in this study and the first β power, which had the highest fitness with scale independence topology network and was three between primary tumor and normal samples. The mean connectivity for power three was equal to 35. When the power β was three, the scale independence was equal to 0.91, which is appropriate for scale-free topology network construction. In addition, the slope of the diagram constructed with the soft connectivity function was equal to -1.61. The topological overlap matrix (TOM) and dissimilarity TOM (dissTOM) were generated using TOM similarity and dissimilarity based on the adjacency matrix.

Eventually, following hierarchical clustering analysis based on dissTOM, modules were produced using the Flash Clust function (Fig. 3). The number of genes in each module is listed in Table 1. With a deepSplit value of 2 for branch splitting (with cutree Dynamic function) and a minimum size cutoff of 30 (minimum cluster size= 30), the primary parameters were determined.

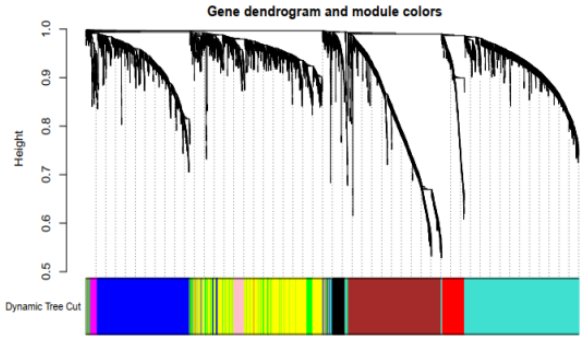

Fig. 3. Gene dendrogram and given colors to each module between primary tumor samples and normal samples.

Visualization of obtained module

By utilizing the "gplots" package and TOMplot function as well as dynamicColors and gene Tree arguments, a TOM heatmap plot was constructed in which illustrated Diagonal blocks represent obtained modules. The cluster package and cmd scale function were also used to show modules in multidimensional scales.

Calculation of representative module

By using the module Eigengenes function and complementing the Principal Components Analysis (PCA) input expression file of WGCNA, the first PCA was determined as a representative module (module eigengenes) for each module. Next, module eigengenes were clustered by the Flash Clust function. Moreover, the interactions of each module with each other were visualized. Furthermore, because the merging threshold of green and yellow modules was less than 0.25, they were merged with the merge Close Modules function and clustered by the same aforementioned function, and module eigengenes were illustrated in plotDendroAndColors were used to illustrate merged module eigengenes in the dendrogram. Then, the module Trait Cor function helped to determine the correlation of considered characteristics (Tumor, Node, Metastasis, and FIGO-staging) with module eigengenes (through Pearson's correlation) and the calculation of the p-value was based on the "Trait P value" function module. Ultimately, module membership was calculated with the Pearson test.

Table 1. The number of genes in the module in primary tumor samples and normal samples.

Modules	N. Genes*	Modules	N. Genes
Black	132	Pink	122
Blue	984	Red	238
Brown	966	Turquoise	1277
Green	252	Yellow	958
Gray	6	Dark Red	65

*= Number of genes

Bioinformatic analysis

Gene significance (coefficient of gene and clinical traits) between datExpr and tumor trait was calculated through gene Trait Significance function (through Pearson test) and finally, GSP value was assessed. Subsequently, the key module was detected based on the data revealed from the gene significance score. In this regard, for the tumor trait, a turquoise-colored module was identified with a p-value = 1.9 × 10-31. Gene enrichment analysis through the DAVID database was implemented. Finally, dendrogram and heatmap diagrams were illustrated through the plot of Eigengene Networks. A scatter plot for the turquoise module (with a correlation of 0.37 and a p-value of 1 × 10-42) was also illustrated between gene significance and module membership. By applying cut-off criteria of gene Module Membership greater than 0.8 and gene Trait Significance less than 0.2, a total of 33 hub genes were identified within the turquoise module, which consists of 1277 genes. Eventually, the CALML5 (Calmodulin-like 5) gene was found as a common gene between differentially expressed genes and identified hub genes for tumor clinical traits.

By implementing the same above-mentioned approaches, the red module was detected as a key module for the node clinical trait with a p-value of 3.3 ×10-33. Dendrogram and heatmap plots were visualized. Similarly, a scatter plot for the red module with a correlation of 0.37 and a p-value of 1× 10- 42 was displayed between gene significance and module membership.

According to the analysis conducted using a specific criterion, it has been observed that out of the 238 genes present in the red module, a subset of 21 genes has been identified as hub genes. As shown in the Venn Diagram, two common genes were detected between differentially expressed genes and discovered hub genes (TERT [telomerase reverse transcriptase] and PNPLA3 [Patatin-like phospholipase domain containing 3]).

For the metastasis clinical trait, the blue module was identified as a key module with a p-value of 3.1 ×10-75. Enrichment analysis was performed through the DAVID database for cellular components, molecular functions, and biological process items. Heatmap and dendrogram plots were represented through the plot Eigengene Networks function. A scatter plot displaying gene significance and module membership in the blue module was illustrated with a correlation of 0.49 and a p-value of 1.5 ×10-60. In the blue module, 30 genes were elucidated as hub genes by the previous criterion of 984 genes. By depicting Venn Diagram, it was clear that four common genes were the same between differentially expressed genes and unveiled hub genes (CHRDL1 [chordin like 1], C7 [complement C7], LEFTY2 [left-right determination factor 2], and PCP4 [Purkinje cell protein 4]).

The Pink module was detected as a key module for the FIGO-staging clinical trait with a p-value of 6.3 ×10-11. Enrichment analysis for GO terms was carried out through the DAVID database. The plot Eigengene Networks function was used to draw a heat map, a dendrogram, and the modules of the FIGO-staging clinical trait. Thereafter, a scatter plot is shown, with a correlation of 0.32 and a p-value of 0.00033 between gene importance and membership in the pink module.

In the pink module, 122 genes existed and one of them was elucidated as hub genes. It is worth mentioning that these genes were not common within identified differentially expressed genes as can be shown in the Venn Diagram.

Survival analysis

The outcomes related to gene expression were then compared using survival analysis. CALML5, TERT, PNPLA3, CHRDL1, C7, LEFTY2, and PCP4 were among the six genes that were shown to be predictive of cervical cancer prognosis and were carried out through GEPIA database and Kaplan-Meier curves and are shown in Figure 4.

Discussion

One of the most prevalent cancers among women worldwide is cervical cancer. By the end of 2020, over 600,000 additional cases had been identified. Because of inadequate medical conditions, death rates are greater in low- and middle-income nations (Denny, 2015; Vu *et al*., 2018). Targeted therapy is now a popular topic in the treatment of cervical cancer (Kumar *et al*., 2018). In past years, many biological markers for cervical cancer have been discovered. In this case, Heng Zou *et al*. identified a novel regulatory pathway involving non-coding circular RNAs termed 0018289, miR-183-5p, and TMED5, which provides new insights into the molecular mechanisms underlying cervical cancer (Zou *et al*., 2021). MiR-186-3p may control the emergence of cervical cancer, according to studies. By modulating the Vascular endothelial growth factor (VEGF) and phosphoinositide-3-kinase–protein kinase B/Akt (PI3K-PKB/Akt) signaling pathways, miR-125 has been demonstrated to prevent the development and progression of cervical cancer. These findings offer valuable insights into the development of effective therapies for cervical cancer (Lu *et al*., 2021). However, the biomarkers of targeted treatment for cervical cancer are complicated and varied and must be thoroughly investigated (Wang *et al*., 2016).

A complicated network exists inside the biological system. WGCNA offers several advantages over other methods, and its conclusions are more biologically relevant and trustworthy since it emphasizes the connection between co-expression modules and clinical traits. In the present WGCNA-bases investigation, the present study aimed to elucidate hub genes related to tumorigenesis and progression of cervical cancer in primary tumors in comparison with normal and metastatic samples, considering four main clinical traits including tumor, node, metastasis, and FIGO-staging. To analyze gene co-expression, the distribution of gene node degrees was examined

by incorporating the data into a human PPI interaction network and a co-expression interaction network. Since adjacent proteins are typically involved in the same disease pathways or biological processes, there was a strong positive association between the expressions of the core proteins and those of the close proteins. As a result, these investigations enable the identification of hub genes and modules that are important for biology. They may act as therapeutic targets or tumor-specific biomarkers. This study's analysis identified hub genes and co-expression modules that were significantly associated with cervical cancer grade and metastasis. These findings have potential clinical implications as TCGA provides a valuable resource for identifying novel therapeutic targets and biomarkers for oncogenic factors (15 and 16). After that, survival analysis was utilized to compare the results related to gene expression. CALML5, TERT, PNPLA3, CHRDL1, C7, PCP4, and LEFTY2 were seven genes that were shown to be hub genes and possible prognostic factors for cervical cancer (Table 2).

Table 2. Key genes and their expression in different modules in the normal samples and primary tumor.

MM.GS & DEGs in turquoise for T/log FC	
CALML5	10.34
MM.GS & DEGs in red for N/log FC	
TERT	9.52
PNPLA3	4.77
MM.GS & DEGs in blue for M/log FC	
CHRDL1	-7.04
C7	-6.37
LEFTY2	-6.42
PCP4	-5.19

While most (non-cycling) normal cells lack telomerase activity, their transcription and activity are upregulated in 80-90% of malignant tumors. TERT promoter mutations, TERT mRNA overexpression, and clinicopathological features are all linked with a variety of cancers. Tumor suppressor protein p27/kip1 may limit TERT mRNA expression and telomerase activity in human cervical cancer cell lines by downregulating telomerase-involved RNA helicase 1 (IRF1) and RNA helicase 2. This suggests that p27 functions as a tumor suppressor by reducing TERT expression. Increasing the expression and transcriptional activity of the cell cycle regulator cyclin D1 in human prostate epithelial cell lines was also demonstrated to boost TERT expression. This data provides support for the hypothesis that TERT regulates cyclin D1 expression. Independent of telomerase activity and telomere preservation, TERT may stimulate vascular endothelial growth factor (VEGF) transcription in WI-38 and HeLa cells (Pestana *et al*., 2017). Calmodulin-like protein 5 (CALML5), also known as calmodulin-like skin protein (CLSP), is found on 10p15.1. This area is commonly changed in HPV-immortalized cells and cervical carcinomas. CALML5 encodes a new calcium-binding protein that is produced in the epidermis and is linked to the calmodulin family of calcium-binding proteins. CALML5 is associated with keratinocyte final differentiation (Kitazawa *et al*., 2021). CALML5 was discovered to be particularly increased in the proportion of nuclear structural proteins of human breast cancer in comparison to normal and benign control samples (Debald *et al*., 2013). To the best of our knowledge, the precise role and function of CALML5 have not been investigated yet and it is suggested that this gene is a potential gene for further cancer research investigations. PNPLA3 is a member of the patatin-like phospholipase domain-containing (PNPLA) family of lipid-metabolizing enzymes, which was initially discovered in potato tubers (Sookoian and Pirola, 2014). The other members of the PNPLA family are soluble proteins with nonspecific lipid acyl hydrolase activity, but PNPLA3 is intimately associated with membranes and lipid droplets (BasuRay *et al*., 2019). In 2001, PNPLA3 was identified as a nutritionally controlled gene that was substantially upregulated during the murine cell line 3T3-L1 differentiation into adipose cells (Chang *et al*., 2013). Similar to CALML5, the connection between PNPLA3 and cervical cancer progression has not been studied yet and can be the topic of the next in vitro and in vivo studies. Otherwise, this research's data showed that LEFTY2, PCP4, C7, and CHRDL1 are down-regulated in primary tumors in comparison to normal samples.

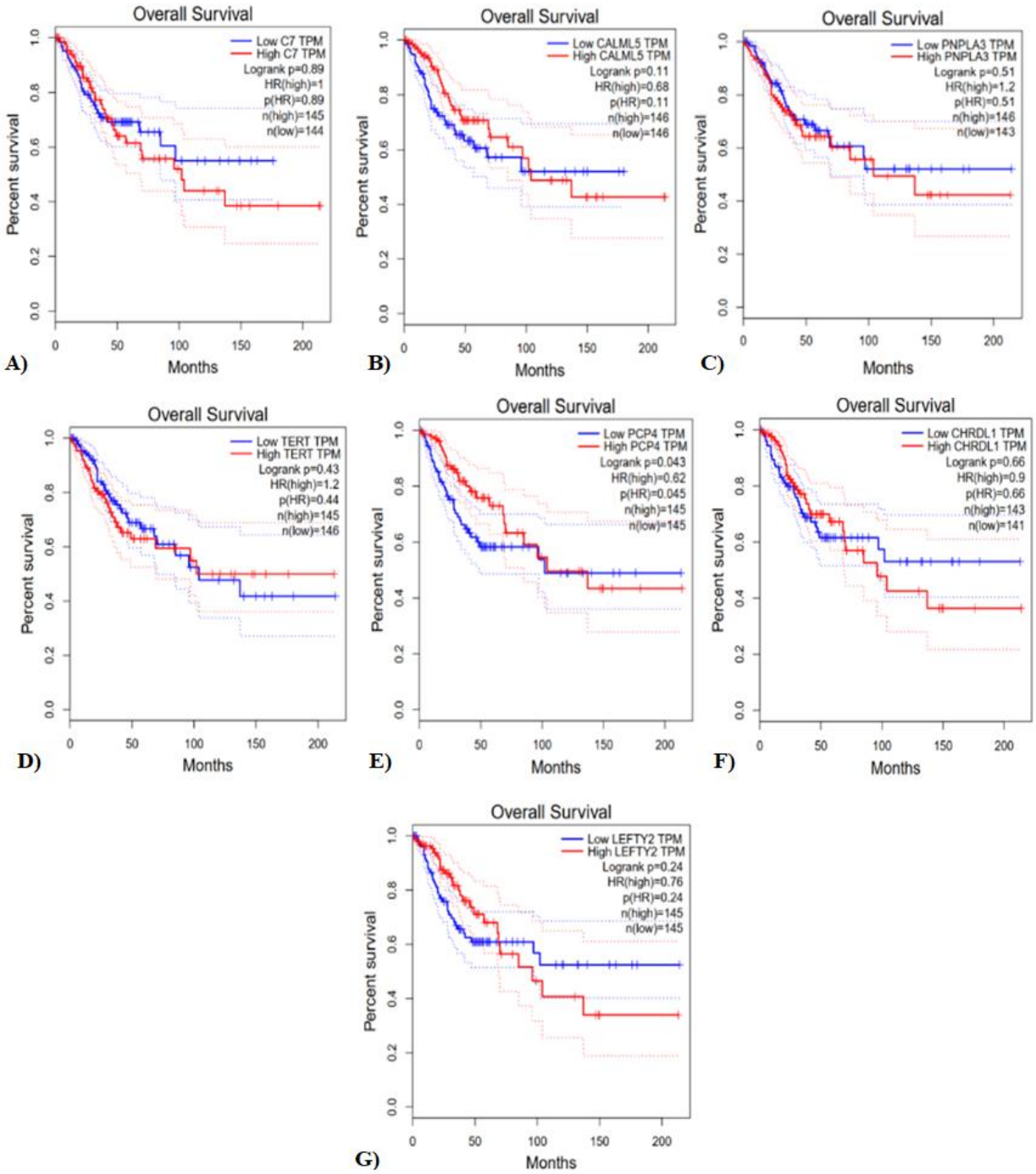

Fig. 4. Survival plots analysis: Survival analysis for A) C7, B) CALML5, C) PNPLA3, D) TERT, E) PCP4, F) CHRDL1, and G) LEFTY2 genes.

In this regard, Ventroptin, a bone morphogenic protein antagonist, is encoded by CHRDL1. It is thought to have a role in the determination of topographic retinotectal projection (Webb *et al*., 2012). Mamoor *et al*. showed that CHRDL1, despite its involvement in cervical cancer not being explicitly detected, is differently expressed at the mRNA level in cancer and dysplasia of the human vulvar, a gynecologic malignancy. CHRDL1 is a potential gene that has been studied in the context of vulvar cancer transcriptional biology (Mamoor, 2021). The C7, along with complement elements of C5b, C6, C8, and C9, may generate a membrane attack

complex. This complex functioned as part of the innate immune system's final complement pathway. The findings indicate that the presence of C7 on the cell surface acts as a modulator for excessive proinflammatory responses (Barroso *et al*., 2006). Furthermore, Li *et al*. found that decreased C7 expression in non-small cell lung cancer (NSCLC) may operate as a tumor suppressor, which is linked to tumor growth and prognosis (Ying *et al*., 2016). Dickinson *et al*. also showed that in individuals with oropharyngeal squamous cell carcinoma, serum C7 levels distinguish between HPV-positive and HPV-negative patients (Dickinson *et al*., 2020). LEFTY2, a negative regulator of cancer cell reprogramming, is an inhibitor of cell proliferation, tumor development, stemness, and embryonic differentiation. In this situation, Alowayed *et al*. discovered that LEFTY2 inhibits the marker of proliferation Ki-67 (MKi67) expression and focal adhesion kinase (FAK) function while increasing miR-200a and E-cadherin expression, making it a potent inhibitor of endometrial cell proliferation and metastasis (Alowayed *et al*., 2016). Purkinje cell protein (PCP) 4 was discovered in the rat cerebellum as a 7.6 kDa polypeptide with similarity to the S100 protein's calcium-binding-chain. It interacts with calmodulin (CaM) and regulates CaM-dependent signaling by modifying calcium/CaM-dependent protein kinase (CaMK) activity to impact a range of events in neurons, including apoptosis (Hamada *et al*., 2014). Yoshimura *et al*. hypothesized that PCP4 may boost cancer cell adhesion, migration, and invasion and that PCP4 might be a novel therapeutic target for drugs that block epithelial-mesenchymal transition and increase cell apoptosis in breast cancer patients (Yoshimura *et al*., 2016). The results of survival analysis also provided the fact that PCP4 downregulation represents poor prognosis due to increased metastasis rate in these patients.

Enrichment analysis for biological processes revealed that most of the genes are enriched in the cell cycle such as DNA replication, nucleus division, cell cycle and its regulation, and chromosome separation. Cellular component enrichment analysis unveiled that the genes are mostly enriched in different cell locations including kinetochore, cell-cell interactions, centromere, and mitotic spindles. Regarding molecular function, the results showed that the genes are enriched in functions correlated with the cell cycle like DNA helicase, actin-binding site, integrin and tubulin, and ATPase activity related to DNA. Moreover, KEGG pathway analysis also revealed the fact that the identified genes are enriched in pathways related to cancer progression involving cell cycle, apoptosis, DNA replication, p53 signaling pathway, prostate cancer, calcium signaling pathway, and DNA repair. Based on the weighted gene co-expression network analysis, this research identified nine modules; among them, the turquoise module, red module, blue module, and pink module were related to the tumor, node, metastasis, and FIGO-staging clinical traits. The main part of the bioinformatics-based study was related to hub gene detection. In this case, seven hub genes termed CALML5, TERT, PNPLA3, CHRDL1, C7, PCP4, and LEFTY2, were elucidated. Among these hub genes TERT, CALML5, and PNPLA3 were shown to be up-regulated and their tumorigenesis effects were validated through the COREMINE database.

This research, however, has several limitations; first, the over-expression or down-regulation of elucidated genes should be validated through in vitro and in vivo investigations. The molecular mechanisms through which identified hub genes promote cervical cancer progression are unknown, although they might be investigated in the future. Furthermore, bigger sample size investigations are required to confirm the clinical significance of obtained hub genes in cervical cancer prognosis.

Acknowledgments

The data analyzed here are in whole or part based upon data generated by the TCGA: http://www.cancer.gov/tcga. Furthermore, the researchers would like to express their sincere appreciation to Fatemeh Akhoundi, Dr. Parvaneh Nikpour, and the Fast-computing center of Shahrekord University. Besides, this work was supported in part by Shahrekord University.

Conflict of Interest

The authors declare no conflict of interest.

References

Alowayed N, Salker MS, Zeng N, Singh Y, Lang F. 2016. LEFTY2 controls migration of human endometrial cancer cells via focal adhesion kinase activity (FAK) and miRNA-200a. *Cell Physiol Biochem* 39(3): 815-826.

Barroso S, Rieubland C, JoséÁlvarez A, López-Trascasa M, Bart PA, Núñez-Roldán A, …, Sánchez B. 2006. Molecular defects of the C7 gene in two patients with complement C7 deficiency. *Immunology* 118(2): 257-260.

BasuRay S, Wang Y, Smagris E, Cohen JC, Hobbs HH. 2019. Accumulation of PNPLA3 on lipid droplets is the basis of associated hepatic steatosis. *Proc Natl Acad Sci* 116(19): 9521-9526.

Boeker M, França F, Bronsert P, Schulz S. 2016. TNM-O: ontology support for staging of malignant tumours. *J Biomed Semant* 7: 1-11.

Chang PA, Sun YJ, Huang FF, Qin WZ, Chen YY, Zeng X, …, Wu YJ. 2013. Identification of human patatin-like phospholipase domain-containing protein 1 and a mutant in human cervical cancer HeLa cells. *Mol Biol Rep* 40: 5597-5605.

Chen L, Lu D, Sun K, Xu Y, Hu P, Li X, …, Xu F. 2019. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* 692: 119-125.

Chen X, Huang L, Yang Y, Chen S, Sun J, Ma C, …, Yang J. 2020. ASPM promotes glioblastoma growth by regulating G1 restriction point progression and Wnt-β-catenin signaling. *Aging (Albany NY)* 12(1): 224-241.

Debald M, Schildberg FA, Linke A, Walgenbach K, Kuhn W, Hartmann G, …, Walgenbach-Bruenagel G. 2013. Specific expression of k63-linked ubiquitination of calmodulin-like protein 5 in breast cancer of premenopausal patients. *J Cancer Res Clin Oncol* 139: 2125-2132.

Denny L. 2015. Control of cancer of the cervix in low-and middle-income countries. *Ann Surg Oncol* 22: 728-733.

Di Y, Chen D, Yu W, Yan L. 2019. Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis. *Hereditas* 156(1): 1-11.

Dickinson A, Saraswat M, Syrjänen S, Tohmola T, Silén R, Randén-Brady R, …, Mattila P. 2020. Comparing serum protein levels can aid in differentiating HPV-negative and-positive oropharyngeal squamous cell carcinoma patients. *PLoS One* 15(6): e0233974.

Edgar R, Barrett T. 2006. NCBI GEO standards and services for microarray data. *Nat Biotechnol* 24(12): 1471-1472.

Hamada T, Souda M, Yoshimura T, Sasaguri S, Hatanaka K, Tasaki T, …, Tsutsui M. 2014. Anti-apoptotic effects of PCP4/PEP19 in human breast cancer cell lines: a novel oncotarget. *Oncotarget* 5(15): 6076-6086

Ibeanu OA. 2011. Molecular pathogenesis of cervical cancer. *Cancer Biol Ther* 11(3): 295-306.

Kitazawa S, Takaoka Y, Ueda Y, Kitazawa R. 2021. Identification of calmodulin-like protein 5 as tumor-suppressor gene silenced during early stage of carcinogenesis in squamous cell carcinoma of uterine cervix. *Int J Cancer* 149(6): 1358-1368.

Kumar L, Harish P, Malik PS, Khurana S. 2018. Chemotherapy and targeted therapy in the management of cervical cancer. *Curr Probl Cancer* 42(2): 120-128.

Langfelder P, Horvath S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform* 9(1): 1-13.

Lu X, Song X, Hao X, Liu X, Zhang X, Yuan N, …, Zhang Z. 2021. MiR-186-3p attenuates tumorigenesis of cervical cancer by targeting IGF1. *World J Surg Oncol* 19(1): 1-10.

Mamoor S. 2021. Differential expression of CHRDL1 in cancer of the vulva. doi:10.31219/osf.io/6mnkr

Orlov YL, Baranova AV. 2020. Bioinformatics of genome regulation and systems biology. *Front Genet* 11:625. doi.org/10.3389/fgene.2020.00625

Pestana A, Vinagre J, Sobrinho-Simões M, Soares P. 2017. TERT biology and function in cancer: beyond immortalisation. *J Mol Endocrinol* 58(2): 129-146.

Petry KU. 2014. HPV and cervical cancer. *Scand. J Clin Lab* 74(244): 59-62.

Sharma S, Deep A, Sharma AK. 2020. Current treatment for cervical cancer: an update. *Anti-Cancer Agents Med Chem*. 20(15): 1768-1779.

Sookoian S, Pirola CJ. 2014. PNPLA3, the history of an orphan gene of the potato tuber protein family that found an organ: the liver. *Hepatol* 59(6): 2068-2071.

Tamura G. 2006. Alterations of tumor suppressor and tumor-related genes in the development and progression of gastric cancer. *World J Gastroenterol* 12(2): 192.

Tomczak K, Czerwińska P, Wiznerowicz M. 2015. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol* 2015(1): 68-77.

Vu M, Yu J, Awolude OA, Chuang L. 2018. Cervical cancer worldwide. *Curr Probl Cancer* 42(5): 457-465.

Wang F, Nazarali AJ, Ji S. 2016. Circular RNAs as potential biomarkers for cancer diagnosis and therapy. *Am J Cancer Res* 6(6): 1167-1176.

Wang Jy, Chen Lj. 2019. The role of miRNAs in the invasion and metastasis of cervical cancer. *Biosci Rep* 39(3). doi: 10.1042/BSR20181377

Webb TR, Matarin M, Gardner JC, Kelberman D, Hassan H, Ang W, …, Hardcastle AJ. 2012. X-linked megalocornea caused by mutations in CHRDL1 identifies an essential role for ventroptin in anterior segment development. *Am J Hum Genet* 90(2): 247-259.

Wright JD, Chen L, Tergas AI, Burke WM, Hou JY, Neugut AI, …, Hershman DL. 2015. Population-level trends in relative survival for cervical cancer. *Am J Obstet Gynecol* 213(5): 670. e1-670. e7.

Yin X, Wang P, Yang T, Li G, Teng X, Huang W, …, Yu H. 2021. Identification of key modules and genes associated with breast cancer prognosis using WGCNA and ceRNA network analysis. *Aging (Albany NY)* 13(2): 2519.

Ying L, Zhang F, Pan X, Chen K, Zhang N, Jin J, …, Jin H. 2016. Complement component 7 (C7), a potential tumor suppressor, is correlated with tumor progression and prognosis. *Oncotarget* 7(52): 86536-86546.

Yoshimura T, Hamada T, Hijioka H, Souda M, Hatanaka K, Yoshioka T, …, Tanimoto A. 2016. PCP4/PEP19 promotes migration, invasion and adhesion in human breast cancer MCF-7 and T47D cells. *Oncotarget* 7(31): 49065-49074.

Zhai X, Xue Q, Liu Q, Guo Y, Chen Z. 2017. Colon cancer recurrence-associated genes revealed by WGCNA co-expression network analysis. *Mol Med Rep* 16(5): 6499-6505.

Zou H, Chen H, Liu S, Gan X. 2021. Identification of a novel circ-0018289/miR-183-5p/TMED5 regulatory network in cervical cancer development. *World J Surg Oncol* 19(1): 1-13.