ABSTRACT – Objective: Cervical cancer is the fourth most common cancer in women worldwide. Therefore, it is necessary to research and propose new diagnostic, prognostic, and therapeutic biomarkers that decrease the high incidence and mortality rates. In this study, we performed bioinformatic analysis using a public dataset to investigate genes as potential diagnostic and prognostic biomarkers.

Materials and Methods: Differentially expressed genes (DEGs) were identified using GSE63678 and GSE7410 datasets in the GEO2R database. Validation of DEGs was performed in Gene Expression Profiling Interactive Analysis (GEPIA) and The Human Protein Atlas (HPA) databases. Enrichment, survival, and Receiver Operating Characteristic (ROC) analyses were performed in Enrichr, Kaplan Meier Plotter, and easyROC software, respectively. Expression according to FIGO stages and copy number were analyzed in GEPIA and cBioportal databases. Methylation was analyzed in the DiseaseMeth database.

Results: We identified 485 DEGs involved in several pathways, including central carbon metabolism. The high Solute carrier family 2, facilitated glucose transporter member 1 (SLC2A1), L-lactate dehydrogenase A chain (LDHA), and Hexokinase-2 (HK2) expression. Also, the low Fibroblast growth factor receptor 2 (FGFR2) expression correlates with poor survival. LDHA and FGFR2 methylation are associated with cervical cancer.

Conclusions: Our results suggested that SLC2A1, LDHA, HK2, and FGFR2 expression could be useful as prognostic biomarkers, while the SLC2A1 and HK2 expression could be good diagnostic biomarkers.

INTRODUCTION

The hallmarks of cancer are sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, evading immune destruction, and reprogramming cell metabolism. In cancer cells, the increased glucose uptake and fermentation of glucose to lactate is common, known as the Warburg Effect, which promotes growth, survival, proliferation, and long-term maintenance. Recent studies have shown that some metabolites or molecules could be used for biomarkers in bio-fluids and tissues, which are used to diagnose cancer stages or predict drug response in several types of cancer, including cervical cancer.

Cervical Cancer (CC) represents a significant public health problem and is the fourth most common cancer in women, with approximately 570,000 cases and 311,000 deaths reported in 2018 worldwide. In addition, CC was the most common cancer in 42 low-resource countries and the main cause of cancer-related deaths in African women. Surgery followed by chemotherapy or radiotherapy is the most common treatment for patients with CC. The main risk factors are low socioeconomic status, smoking, multiple sexual partners, multiple childbirths, and infection with high-risk human papillomavirus (HR-HPV), such as HPV16 and 18. Most Western countries' population-wide screening and early-detection programs are based on a cytological diagnostic test known as the ‘Pap test’ and HPV molecular detection. However, although CC cases have decreased worldwide, there has been reported an increase in CC cases in developing countries, and according to the Human Development Index (HDI), it has been estimated a high incidence (26.7) and mortality rate (20) in countries with low HDI. The overall survival (OS) of patients with CC remains poor, with a 5-year survival of 63.5% and 62.8% in 2001-2003 and 2004-2009, respectively. Therefore, it is necessary to research and propose new diagnostic, prognostic, and therapeutic biomarkers that decrease the high incidence and mortality rates of CC.

In the present study, we aim to investigate potential prognostic biomarkers through bioinformatic analysis using a public dataset.

MATERIALS AND METHODS

A total of 485 DEGs were identified in two datasets of patients with CC, which participate in several signaling pathways, including reprogramming of the cell metabolism. The gene expression in this pathway was validated in another two independent datasets. Survival analysis revealed that the SLC2A1, LDHA, HK2, and FGFR2 expression correlate with poor OS. The diagnostic value of the expression of these genes was determined through the ROC curve analysis, the expression of these genes was analyzed according to FIGO stages and copy numbers. Finally, the methylation analysis in their promoters was performed in cervical cancer and normal samples.

Clinical significance

Nine genes signature associated with central carbon metabolism were deregulated in cervical cancer. Mainly, SLC2A1 and HK2 expression could be a good diagnosis biomarker in this disease; other genetic alterations, such as DNA methylation or copy number, are related to expression alteration of SLC2A1 and HK2 genes.

Differential expression analysis

Differential expression analysis was performed using two independent cervical cancer microarray datasets, including GSE63678 and GSE7410 in the GEO2R database, a web tool based on GEOquery and limma R packages from Bioconductor project that allows comparing two or more groups of samples to identified differentially expressed genes (DEGs). GSE63678 dataset (Platform: GPL571 [HG-U133A_2]-Affymetrix Human Genome U133A 2.0 array) includes five cervical cancer samples and five normal cervical samples and the GSE7410 dataset (Platform: GPL1708-Agilent-012391 Whole Human Genome Oligo Microarray G4112A) includes 30 CC samples and five normal cervical samples. A log2 fold change>1.0 and p-value<0.05 were considered to identify DEGs between CC and normal samples. The DEGs in common were identified through a Venn Diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/).
Pathway enrichment analysis

Enrichment analysis of DEGs was performed using the Enrichr program\textsuperscript{14,15}. Enrichr is integrative web software that identifies enriched terms using the JavaScript library. The KEGG 2019 human database includes several biological pathways, sorted by combined score ranking (\(p\)-value <0.05, adjusted \(p\)-value: <0.05, and odds ratios).

Validation analysis

DEGs were validated in two independent CC datasets, including The Cancer Genome Atlas (TCGA) dataset using the GEPIA database\textsuperscript{16} and HPA\textsuperscript{17,18}. GEPIA is a web server that helps to analyze RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from TCGA and GTEx projects. The expression data were log2 (TPM+1) transformed in each group. The significant differences were determined using a one-way ANOVA test, and a \(p\)-value<0.01 was considered significant. The HPA is a Swedish-based program that analyzes the expression of all the human proteins in cells and tissues.

Expression analysis according to FIGO stages and copy number

The expression of SLC2A1, LDHA, HK2, and FRFG2 genes was analyzed according to FIGO stages (I to IV) and copy number in CC samples from the TCGA dataset using the GEPIA database (16) and cbio Cancer Genomics portal (cbiportal) database\textsuperscript{19,20}. In the GEPIA database, the expression data were log2 (TPM+1) transformed, and a One-way ANOVA test was used to determine the differences; a \(p\)-value<0.01 was considered significant. On the other hand, cBioportal provides an open-access source to analyze cancer genomic data. The CC samples were classified into diploid, depleted, and gain groups according to copy number. The differences were determined through Student’s \(t\)-test, and a \(p\)-value<0.05 was considered significant.

Survival analysis

The prognostic value of mRNA expression of SLC2A1, LDHA, HK2, and FRFG2 genes was analyzed in the Kaplan Meier Plotter database\textsuperscript{21}. Kaplan Meier Plotter database assesses the effect of 54,000 genes on overall survival (OS) and relapse-free survival (RFS) in 21 human cancers from TCGA, GEO, and EGA datasets. Patients were divided into two groups based on the median expression of each gene (high vs. low expression). A hazard ratio (HR) with 95% CI and a log-rank \(p\)-value were calculated in each Kaplan-Meier survival plot. A log-rank \(p\)-value<0.05 was considered statistically significant.

ROC analysis

ROC curve analysis was performed from GSE7803, GSE67522, and GSE63514 datasets using easyROC software (Ver. 1.3.1)\textsuperscript{22}. GSE7803 (Platform: GPL96 [HG-U133A] Affymetrix Human Genome U133A Array) includes 21 cervical cancer samples and ten normal cervical samples\textsuperscript{23}. GSE67522 (Platform: GPL10558 Illumina HumanHT-12 V4.0 expression bead chip) includes 20 cervical cancer samples and 22 normal cervical samples\textsuperscript{24}. GSE63514 (Platform: GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) includes 28 cervical cancer samples and 24 normal cervical samples\textsuperscript{25}. A \(p\)-value<0.05 was considered significant with a 95% confidence interval (CI).

Identification of CpG islands

The SLC2A1, LDHA, HK2, and FRFG2 gene promoters, specifically of -2000 pb to +2000 pb relative to transcription start site (TSS), were downloaded from the ExPASy Bioinformatics Resource Portal\textsuperscript{26}. ExPASy is the SIB Bioinformatics Resource Portal that provides access to scientific databases and software
tools, including genomic data. The prediction of CpG islands was performed considering a window >200 bp with a CG content ≥ of 50% and a ratio of observed CpG/expected CpG > 0.6 in the Methprimer program. Methprimer serves to identify CpG islands and designed primers for bisulfite genomic sequencing and Methylation-specific PCR (MSP) assays.

Correlation analysis

The correlation between methylation and expression of SLC2A1, LDHA, HK2, and FRFG2 genes was performed from the TCGA dataset using the cBioportal database. The correlation was evaluated through Spearman, Pearson, and R correlation coefficients, and a p-value<0.05 was defined as statistically significant.

Methylation analysis

Methylation analysis was performed in SLC2A1, LDHA, HK2, and FRFG2 gene promoters from the TCGA dataset using the DiseaseMeth version 2 database. DiseaseMeth is a human disease methylation database, especially for various human cancers. Methylation was analyzed in the whole promoter (2.0 kb upstream of TSS to 0.5 kb downstream) to each gene from 450k microarray (Illumina Infinium Human-Methylation450 BeadChip). Student’s t-test was performed to compare two groups, and a p-value<0.01 was considered statistically significant.

RESULTS

Central carbon metabolism-associated genes are deregulated in cervical cancer

We analyzed the gene expression in CC samples to identify DEGs and compared them with normal cervical samples from GSE63678 and GSE7410 datasets using the GEO2R database. We found 485 overlapping DEGs, including 260 genes overexpression and 225 genes downregulation (Figure 1A). We selected the top ten DEGs (five upregulated and five downregulated) to validate their expression in CC samples, compared with normal cervical samples from the TCGA dataset using the GEPIA database, and we found consistent results (Figure 1B). Finally, we performed a pathways enrichment analysis with the 485 DEGs using Enrichr software, and we found several pathways well known in CC, such as DNA replication, cell cycle, and p53 signaling pathway. Interestingly, the central carbon metabolism pathway in cancer was observed (Figure 1C), and taking into account its role in cancer development, we focused our attention on this pathway.

To validate the expression of DEGs involved in metabolism in CC (SLC2A1, LDHA, HK2, TIGAR, SCO2, PIK3R1, FGFR1, PDGFRB, and FGFR2), we analyzed their expression at mRNA and protein level in CC samples. We compared with normal cervical samples from TCGA and HPA datasets using the GEPIA database. We found consistent results: the expression of SLC2A1, LDHA, HK2, TIGAR, and SCO2 genes is significantly increased (Figure 2A and C). In contrast, PIK3R1, FGFR1, and PDGFRB gene expression were significantly downregulated (Figure 2B and D) in CC samples compared with normal cervical samples. These data suggest that SLC2A1, LDHA, HK2, TIGAR, SCO2, PIK3R1, FGFR1, and PDGFRB genes could be involved in CC.

SLC2A1, LDHA, HK2, and FGFR2 expression correlates with poor overall survival in cervical cancer

To analyze the prognostic value of the genes analyzed above, we analyzed the OS according to the expression of these genes in patients with CC from the TCGA dataset using the KM-plotter database. The results revealed that high SLC2A1, LDHA, and HK2 expression. Also, low FGFR2 expression correlates with poor OS in patients with CC (Figure 3A-D). Together, these results suggest that the expression of SLC2A1, LDHA, HK2, and FGFR2 genes could be useful as a prognostic biomarker in CC.
Diagnostic value of SLC2A1 and HK2 expression for cervical cancer

To determine whether the expression of SLC2A1, LDHA, HK2, and FGFR2 genes could be useful as a diagnostic biomarker, we performed a ROC analysis from GSE7803, GSE67522, and GSE63514 datasets using easyROC software. In GSE7803, the results showed an Area Under Curve (AUC) of 0.8619 (95% CI: 0.73-0.98, sensitivity: 0.71, specificity: 1.0 and \( p < 0.05 \)) and 0.88095 (95% CI: 0.76-0.99, sensitivity: 0.66, specificity: 1.0 and \( p < 0.05 \)) to SLC2A1 and HK2 genes, respectively (Figure 4A). Similar results were obtained from GSE67522 dataset with an AUC of 0.83182 (95% CI: 0.69-0.96, sensitivity: 0.75, specificity: 0.86 and \( p < 0.05 \)) and 0.72273 (95% CI: 0.56-0.88, sensitivity: 0.70, specificity: 0.72 and \( p < 0.05 \)) (Figure 4B) and GSE63514 dataset with an AUC of 0.76786 (95% CI: 0.63-0.90, sensitivity: 0.92, specificity: 0.62 and \( p < 0.05 \)) and 0.64732(95% CI: 0.49-0.80, sensitivity: 0.64, specificity: 0.75 and \( p < 0.05 \)) (Figure 4C) to SLC2A1 and HK2 genes, respectively. These results suggest that expression of SLC2A1 and HK2 are good candidates to discriminate tumor tissues from normal tissues.
Figure 2. Validation of DEGs involved in metabolism in CC. The expression of 9 DEGs involved in metabolism was validated at the mRNA level from the TCGA dataset using the GEPIA database (A and B) and at the protein level from The HPA dataset (C and D). The significant differences were determined through a one-way ANOVA test. A p-value<0.01 was considered statistically significant.
SLC2A1 and HK2 expression increases in advanced FIGO stages in cervical cancer

To determine SLC2A1, LDHA, HK2, and FGFR2 gene expression in stages of CC, we analyzed their expression according to FIGO stages in CC samples from the TCGA dataset using the GEPIA database. We found that the expression of SLC2A1 and HK2 (Figure 5A and C) genes increases, the expression of the LDHA gene slightly increases (Figure 5B), and the expression of FGFR2 gene slightly decreases (Figure 5D) in advanced FIGO stages in CC. These results suggest that SLC2A1 and HK2 expression could be involved in the progression of CC.

SLC2A1 and LDHA expression correlates with alterations in copy number in cervical cancer

To explore whether genetic alterations, such as variations in copy numbers, are involved in the deregulation of expression of SLC2A1, LDHA, HK2, and FGFR2 genes in CC, we analyzed their expression in CC samples with depletion and gain copy, compared with CC samples with diploid from TCGA dataset using cBioportal database. We found that depletion and gain genic could be involved in the expression of SLC2A1 and LDHA in CC samples (Figure 6A and B). However, we do not find changes in the expression of HK2 and FGFR2 genes (Figure 6C and D). All these data suggest that the depletion and gain in copy number could be involved in deregulating the expression of SLC2A1 and LDHA genes in CC.
Figure 4. Diagnostic value of expression of SLC2A1, LDHA, HK2, and FGFR2 genes for CC. ROC curve analysis of expression of SLC2A1, LDHA, HK2, and FGFR2 genes in CC was performed from GSE7803 (A), GSE67522 (B), and GSE63514 (C) datasets using easyROC software.
Methylation in promoters of SLC2A1, LDHA, HK2, and FGFR2 genes negatively correlates with their expression in cervical cancer

To explore if DNA methylation is involved in the deregulation of expression of SLC2A1, LDHA, HK2, and FGFR2 genes in CC, first, we searched the presence of CpG islands in their promoters, and we found a CpG island located in -2000 pb to +2000 pb region in each gene (Figure 7A-D; upper). Then, we performed a correlation analysis between methylation and expression of SLC2A1, LDHA, HK2, and FGFR2 genes in CC samples from the TCGA dataset using the cBioportal database, and we found a negative correlation between methylation and expression to each gene (Figure 7A-D; middle). Finally, we analyzed the methylation level in the promoters of these genes in CC samples. We compared them with normal cervical samples from the TCGA dataset using the DiseaseMeth database, and we found that methylation in the promoter of SLC2A1 genes lightly decreases (Figure 7A; bottom) in CC samples. To confirm this result, we analyzed the methylation level in the promoter of the SLC2A1 gene in CC samples. We compared it with normal cervical samples from the GSE30760 dataset using the GEO2R database, and

Figure 5. Expression of SLC2A1, LDHA, HK2, and FGFR2 genes in FIGO stages of CC. Expression of SLC2A1 (A), LDHA (B), HK2 (C), and FGFR2 (D) genes were analyzed in FIGO stages (I to IV) of CC from the TCGA dataset using the GEPIA database. The differences were determined using a one-way ANOVA test, and a p-value<0.01 was considered significant.
we found that methylation level in the promoter SLC2A gene decreases in CC samples (Data not shown). The methylation level in the promoters of LDHA (Figure 7B; base) and HK2 (Figure 7C; base) genes decreases, while the methylation level in the promoter FGFR2 gene increases (Figure 7D; base) in CC samples. Overall, these results suggest that the abnormal methylation in the promoters of SLC2A1, LDHA, HK2, and FGFR2 genes could be involved in deregulating their expression in CC.

**DISCUSSION**

Currently, the mortality and morbidity rates for CC are a serious public health problem because this type of cancer is the fourth most common cancer among women worldwide⁴. The lack of screening methods for the early stage of CC, with high sensitivity and specificity, is the main reason for the failure to eradicate this disease⁵. Therefore, it is necessary to identify new biomarkers for the early screening of CC.
Figure 7. Methyl-ation analysis in the promoters of SLC2A1, LDHA, HK2, and FGFR2 genes in CC. Upper: CpG islands in the promoters of SLC2A1 (A), LDHA (B), HK2 (C), and FGFR2 (D) genes. Arrow: Transcription Start Site. White box: Exon 1. Vertical line: CpG. Middle: Correlation analysis between the methylation and expression of each gene. Spearman, Pearson, and R correlation coefficients were determined, and a p-value < 0.05 was defined as statistically significant. The methylation level in the SLC2A1, LDHA, HK2, and FGFR2 gene promoters was analyzed in CC samples and compared with normal cervical cancer from the TCGA dataset using the DiseaseMeth database. The Student’s t-test determined the differences, and a p-value < 0.01 was considered statistically significant.
In this study, we found 485 DEGs through a differential expression analysis, and the pathway analysis revealed key pathways in carcinogenesis, such as DNA replication\(^9,30\), cell cycle\(^31\), and p53 signaling pathway\(^30\). Interestingly, we found the pathway of central carbon metabolism in cancer, which is one of the hallmarks in cancer, known as reprogramming of the cell metabolism\(^1\), where SLC2A1, HK2, LDHA, and other molecules play an important role in glucose consumption and lactate production for cell proliferation; in fact, the most important function of SLC2A1 is the glucose entry to the cell\(^32\), while HK2 is a key factor for the conversion of the substrate glucose into glucose-6-phosphate\(^33\). Finally, LDHA is the main mediator for L-lactate production\(^34\) (Figure 8). Several studies have reported that SLC2A1 is associated with invasiveness through forming stromal protrusions in CC\(^35\)–\(^37\), while LDHA is over-expressed in CC with pelvic lymph node metastasis\(^38\). On the other hand, HK2\(^39,40\) and FGR2 expression are usually associated with cell growth and progression of cervical dysplasia\(^41,42\). Interestingly, we found eight DEGs deregulates involved in the pathway of central carbon metabolism in CC; they were validated at mRNA and protein levels using two other independent datasets. Similar results were obtained, supporting our findings. However, only the high expression of SLC2A1, LDHA, and HK2 genes and low FGFR2 correlate with poor OS in patients with CC. Similarly, previous studies have shown that the high expression of SLC2A1\(^43\), HK2\(^44\), and the negative FGFR2 expression\(^45\) correlates with poor OS in patients with CC. Similarly, the survival analysis shows that the high expression of SLC2A1, LDHA, HK2, and the low expression of FGFR2 confers a poor prognostic at 50 months in the patients with CC (Figure 4). To explore the molecular mechanisms involved in SLC2A1, LDHA, and HK2, and low FGFR2 expression in CC, we analyzed the genetic alterations, such as variations in copy number\(^46\) and epigenetic factors, such as methylation in their promoter, due to this, the methylation is highly related to Warburg effect in cancer cells, particularly in CC deregulation in DNA methylation is associated with metabolic alterations\(^47\). Only SLC2A1 and FGFR2 methylation had been associated with lung adenocarcinoma and gastric cancer prognostic, respectively. Interestingly, we reported for the first time that LDHA and

FGFR2 methylation are associated with CC (Figure 7). At the same time, the SLC2A and LDHA depletion is related to a low expression at the mRNA level (Figure 6). Further studies on other molecular mechanisms, such as miRNAs and IncRNAs, are required to regulate SLC2A1, LDHA, HK2 and also the FGFR2 expression.

To our knowledge, this is the first integral and comprehensive bioinformatic study focused on the pathway of central carbon metabolism in CC. However, this study has some limitations. First, an internal validation cohort is lacking to confirm all our results despite validation analysis in independent datasets. Second, a small number of samples could affect the diagnostic value of the LDHA and FGFR2 genes.

CONCLUSIONS

Our results show a nine-gene signature associated with central carbon metabolism in cancer (SLC2A1, LDHA, HK2, TIGAR, SCO2, PIK3R1, FGFR1, PDGFRB, and FGFR2) are deregulated in CC. Interestingly, SLC2A1, LDHA, HK2, and FGFR2 expression could be useful as prognostic biomarkers. In addition, the expression of SLC2A1 and HK2 genes together with the detection of high-risk HPV can improve diagnostic efficiency in cervical cancer. Finally, DNA methylation and copy number are involved in the altered expression of these genes.

FUNDING:
No funding was received.

AUTHOR CONTRIBUTIONS:
Eric G. Salmerón-Bárcenas, Pedro A. Ávila-López and Francisco I. Torres-Rojas: takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation, drafting the article. Eric G. Salmerón-Bárcenas, Miguel Á. Mendoza-Catalán and Ana E. Zacapala-Gómez: takes responsibility for statistical analyses and interpretation of data. Berenice Illades-Aguiar and Francisco I. Torres-Rojas: takes responsibility for complete text evaluation and guidance final approval of the version to be submitted. All authors read and approved the final manuscript.

ORCID ID:
Eric G. Salmerón-Bárcenas: https://orcid.org/0000-0002-1830-06724
Ana E. Zacapala-Gómez: https://orcid.org/0000-0003-4935-6094
Pedro A. Ávila-López: https://orcid.org/0000-0002-0104-0771
Miguel Á. Mendoza-Catalán: https://orcid.org/0000-0002-6247-8590
Berenice Illades-Aguiar: https://orcid.org/0000-0003-3937-335X
Francisco I. Torres-Rojas: https://orcid.org/0000-0003-4303-8677

CONFLICT OF INTEREST:
The authors declare that they have no conflict of interest to disclose.

ETHICS COMMITTEE APPROVAL AND INFORMED CONSENT:
Not required for this study

DATA AVAILABILITY STATEMENT:
Data sharing does not apply to this article as no datasets were generated or analyzed during the current study.

REFERENCES