



Dysregulation of Transcription Factor Networks Unveils Different Pathways in Human Papillomavirus 16-Positive Squamous Cell Carcinoma and Adenocarcinoma of the Uterine Cervix

Saloe Bispo^{1†}, Ticiana D. J. Farias^{1†}, Patricia Savio de Araujo-Souza^{2,3}, Ricardo Cintra⁴, Hellen Geremias dos Santos¹, Natasha Andressa Nogueira Jorge^{5,6}, Mauro Antônio Alves Castro⁷, Gabriel Wajnberg^{5,8}, Nicole de Miranda Scherer⁹, Maria Luiza Nogueira Dias Genta¹⁰, Jesus Paula Carvalho¹⁰, Luisa Lina Villa^{11,12}, Laura Sichero^{12*†} and Fabio Passetti^{1,5*†}

OPEN ACCESS

Edited by:

Shengli Li,

Shanghai Jiao Tong University, China

Reviewed by:

Yunfeng Wang,

Université Paris-Saclay, France

Madhu Khullar,

Post Graduate Institute of Medical

Education and Research

(PGIMER), India

*Correspondence:

Fabio Passetti

fabio.passetti@fiocruz.br

Laura Sichero

laura.sichero@hc.fm.usp.br

†These authors have contributed
equally to this work

Specialty section:

This article was submitted to

Cancer Genetics,

a section of the journal

Frontiers in Oncology

Received: 05 November 2020

Accepted: 17 February 2021

Published: 19 May 2021

Citation:

Bispo S, Farias TDJ, de

Araujo-Souza PS, Cintra R,

Santos HG, Jorge NAN, Castro MAA,

Wajnberg G, Scherer NM,

Genta MLND, Carvalho JP, Villa LL,

Sichero L and Passetti F (2021)

Dysregulation of Transcription Factor

Networks Unveils Different Pathways

in Human Papillomavirus 16-Positive

Squamous Cell Carcinoma and

Adenocarcinoma of the Uterine

Cervix. *Front. Oncol.* 11:626187.

doi: 10.3389/fonc.2021.626187

¹ Instituto Carlos Chagas, FIOCRUZ, Curitiba, Brazil, ² Department of Immunobiology, Biology Institute, Universidade Federal Fluminense (UFF), Niterói, Brazil, ³ Laboratory of Immunogenetics and Histocompatibility, Department of Genetics, Universidade Federal do Paraná, Curitiba, Brazil, ⁴ Department of Biochemistry, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, ⁵ Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, ⁶ Bioinformatics Group, Department of Computer Science, Interdisciplinary Center for Bioinformatics, Leipzig University, Leipzig, Germany, ⁷ Bioinformatics and Systems Biology Laboratory, Federal University of Paraná, Curitiba, Brazil, ⁸ Atlantic Cancer Research Institute, Moncton, NB, Canada, ⁹ Laboratory of Bioinformatics and Computational Biology, Division of Experimental and Translational Research, Brazilian National Cancer Institute (INCA), Rio de Janeiro, Brazil, ¹⁰ Discipline of Gynecology, Department of Obstetrics and Gynecology, Instituto do Cancer do Estado de São Paulo, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil, ¹¹ Department of Radiology and Oncology, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil, ¹² Center for Translational Research in Oncology, Instituto do Cancer do Estado de São Paulo ICESP, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo FMUSP HC, São Paulo, Brazil

Squamous cell carcinoma (SCC) and adenocarcinoma (ADC) are the most common histological types of cervical cancer (CC). The worse prognosis of ADC cases highlights the need for better molecular characterization regarding differences between these CC types. RNA-Seq analysis of seven SCC and three ADC human papillomavirus 16-positive samples and the comparison with public data from non-tumoral human papillomavirus-negative cervical tissue samples revealed pathways exclusive to each histological type, such as the epithelial maintenance in SCC and the maturity-onset diabetes of the young (MODY) pathway in ADC. The transcriptional regulatory network analysis of cervical SCC samples unveiled a set of six transcription factor (TF) genes with the potential to positively regulate long non-coding RNA genes *DSG1-AS1*, *CALML3-AS1*, *IGFL2-AS1*, and *TINCR*. Additional analysis revealed a set of MODY TFs regulated in the sequence predicted to be repressed by miR-96-5p or miR-28-3p in ADC. These microRNAs were previously described to target *LINC02381*, which was predicted to be positively regulated by two MODY TFs upregulated in cervical ADC. Therefore, we hypothesize *LINC02381* might act by decreasing the levels of miR-96-5p and miR-28-3p, promoting the MODY activation in cervical ADC. The novel TF networks here described should be explored for the development of more efficient diagnostic tools.

Keywords: cervical cancer, cervical adenocarcinoma, cervical squamous cell carcinoma, RNA-Seq, transcriptional regulatory network, lncRNA, miRNA sponge

INTRODUCTION

Cervical cancer (CC) is the fourth leading cause of death related to cancer in women worldwide (1) and the second cause of mortality in women aged 20–39 years. Higher mortality rates due to CC are observed within developing countries (2). The main risk factor for CC development is the persistent infection by high-risk human papillomavirus (HPV) (3). Squamous cell carcinoma (SCC) and adenocarcinoma (ADC) are the most common histological types of CC, comprising ~80% and 20% of the cases of all invasive CC cases, respectively (4). However, ADC cases are increasing, possibly due to differential failures in the effectiveness of screening programs for ADC detection as compared with SCC (5). Despite their epidemiological, histopathological, and prognostic differences, patients presenting cervical SCC or ADC receive identical treatments (4). Currently recommended treatments are successful for most cervical SCC lesions but frequently fail in treating cervical ADC (4, 5). This reinforces the need to identify molecular differences between cervical histological types. This knowledge is seminal for cancer biology understanding and to foster the development of novel ADC treatments.

Complex diseases, such as CC, can be evaluated by means of regulatory gene expression networks (6). This analysis identifies expression enrichment of specific transcription factors (TFs) as key regulators of various target genes, such as protein-coding, microRNAs (miRNAs), and long non-coding RNAs (lncRNAs) (6, 7). Pseudogenes are widely expressed in human cancers (8), and alterations in lncRNA gene expression levels are relevant for tumorigenesis (9). However, differences in the expression of lncRNA genes or pseudogenes between cervical SCC and ADC in the context of regulatory networks have not been described to date.

Thus, we aimed to unravel gene expression profiles that could discriminate different CC histological types in the context of transcriptional regulatory networks (TRNs) of protein-coding and non-coding RNA genes, such as lncRNAs. Here, we identified differences in gene expression profiles between cervical SCC and ADC, which were further analyzed in the context of TRNs, unveiling TFs that participate in epithelial maintenance and the maturity-onset diabetes of the young (MODY) pathways, respectively. The identification of these dysregulated networks contributes not only to the molecular characterization of CC of different histological types but may also support investigations for the identification of specific biomarkers and the development of treatments.

MATERIALS AND METHODS

Cohort Study

CC tumor biopsies were obtained from 30 untreated women with histologically proven cervical SCC ($n = 25$) or ADC ($n = 5$) referred to the Gynecological Service of the Instituto do Cancer do Estado de São Paulo (ICESP, São Paulo, Brazil) (Supplementary Table 1).

All patients signed an informed consent form before biopsies were obtained. Histological analysis was carried out using hematoxylin and eosin-stained sections at ICESP.

DNA and RNA Isolation, and Human Papillomavirus Genotyping

According to the manufacturer's instructions, DNA and RNA were isolated from each cervical tissue sample using the DNA/RNA All Prep kit (Qiagen, Germany).

Isolated DNA concentration and purity were evaluated using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA), and samples were diluted to a final concentration of 50 ng/ μ l. HPV detection and genotyping were performed using 100 ng of each DNA sample and the Inno-LiPA HPV Genotyping kit (FujiRebio, Belgium), according to manufacturer instructions.

RNA isolation was followed by DNase treatment. RNA integrity and quality were evaluated using the 2200 Tape Station System (Agilent Technologies Inc, USA), and RNA integrity number ≥ 8 was considered the quality threshold. In total, seven cervical SCCs and three ADCs with $< 20\%$ necrotic tissue were submitted to RNA sequencing (Supplementary Table 1).

RNA Sequencing

All qualified CC samples were submitted to library preparation using the TruSeq Stranded messenger RNA (mRNA) LT Sample Prep kit (Illumina Inc., USA) and paired-end (2×100 bp) sequencing using the HiSeq SBS Kit v4 and the HiSeq Flow Cell v4 with HiSeq 2500 system (Illumina Inc., USA) at the Centro de Genômica Funcional Esalq—University of São Paulo. On average, 85 million paired-end reads were sequenced per tumor sample (Supplementary Table 2).

Public Cervical Gene Expression Data Included in This Study

Transcriptome data of HPV-negative non-tumoral cervical samples (named non-CC) were downloaded from the Gene Expression Omnibus portal (accession PRJNA454568) (10).

Because gene-gene interaction analysis must be performed with a sample size bigger than the one used in this study, we used The Cancer Genome Atlas (TCGA) biolinks to extract data from the cervical tumors (TCGA-CESC). We downloaded 280 CC transcriptome data from the public TCGA repository under the tag TCGA-CESC, including 249 cervical SCC and 31 cervical ADC samples. We further compared gene expression profiles from our study samples with the TCGA dataset (Supplementary Table 3).

Transcriptome Bioinformatic Analysis

RNA sequence data output (FASTQ files) was trimmed with Trim Galore (v. 0.0.4.0) (11) and aligned onto the GRC37/hg19 version of the human genome and the HPV-16 genome (NC_001526.4, GenBank) using Hisat2 (v. 2.1.0) (12) with the following parameters: “--dta --fr -q --no-mixed --no-discordant”. The latest human gene annotation dataset (hsapiens_gene_ensembl, available at <http://grch37.ensembl.org>) was obtained using the biomaRt package (v. 2.34.2) (13) from R statistical environment. SAM files were converted, merged, sorted, and indexed using

Samtools utilities (14). The aligned reads were visualized using the IGV software (v. 2.3.82) (15). Reads counts were accessed using HTseq (v. 0.5.4p3) (16) applying the following parameters: “-m intersection-noempty -i id_gene -s reverse.”

Differential gene expression analysis was performed to compare between the normalized and filtered read count among three groups: cervical SCC, ADC, and non-CC (10) using the Bioconductor package edgeR (v. 3.20.9) (17), setting the \log_2 fold change module ($|\log_2FC| \geq 2$) and false discovery rate as ≤ 0.01 . To overcome the batch effect due to combining data from independent cohorts, we used edgeR package and BatchEffect function from Limma Package, available at the Bioconductor project. Non-expressed and weakly expressed genes were defined as having ≤ 1 count (read) per million in n of samples in the smallest group for each comparison and were excluded from the differential expression analysis.

Regulatory Networks Derived From Gene Expression Profiles

The regulatory network for both cervical SSC and ADC was inferred by computing mutual information between annotated TFs and all potential target genes using the RTN package (18). In the network architecture inferred, each TF was assigned to a list of candidate targets, which could be linked to multiple TFs. As regulation can occur from both direct (TF–target) and indirect interactions (TF–TF–target), the Algorithm for the Reconstruction of Accurate Cellular Networks was used as an additional step to enrich the regulatory networks with direct TF–target interactions (19). The resulting regulatory networks were plotted with Cytoscape (v. 3.7.2) (20).

Gene Function Annotation and Pathway Enrichment Analysis

Enrichment of Gene Ontology biological processes of cervical SCC and ADC gene expression profiles were investigated using Enrichr (21). The GSEA software (version 4.0.3.4) (22) was used to identify Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways showing an overrepresentation of up- or downregulated genes in cervical ADC or SCC. Briefly, an enrichment score was calculated for each gene set (i.e., KEGG pathway) by ranking each gene according to their expression difference using Kolmogorov–Smirnov statistic, computing a cumulative sum of each gene ranked in each gene set and recording the maximum deviation from zero as the enrichment score.

Quantitative Real-Time Polymerase Chain Reaction Validation and Analysis

To validate genes with different expression levels revealed by whole RNA sequencing, we performed quantitative polymerase chain reaction (qPCR) of six selected differentially expressed lncRNA using RNA obtained from 20 cervical SCC and 5 cervical ADC samples (Supplementary Tables 1, 4). We selected for validation by qPCR lncRNAs with the highest or lowest \log_2FC values. Complementary DNA synthesis from 500 ng of total RNA isolated from CC samples was performed using the SuperScript

VILO kit (Life Technologies, USA). qPCR was carried out using the SYBR Green PCR Master Mix on an ABI 7500 real-time PCR system (Applied Biosystems, USA). lncRNAs transcript levels were normalized to the mRNA levels of the ribosomal protein lateral stalk subunit P0 (*RPLP0*) constitutive gene. The lncRNA gene annotation from LNCipedia (23), Ensembl transcript identification, qPCR pair primers pair sequences, and expected amplicon length are available in **Supplementary Table 4**. The Mann–Whitney statistical test was applied to compare cervical SCC and ADC distribution, considering $\alpha = 0.01$ as the significance level. Statistical analyses were performed using GraphPad Prism version 7 for Windows (24) and ggplot2 package (25) in the R statistical environment.

In silico Analysis of MicroRNA–Messenger RNA Target Predictions

The following databases of miRNA targeting mRNA predictions have been investigated to detect potential mRNA transcripts regulated by the miR-28-3p or miR-96-5p: mirDIP with the minimum score set to medium (26); mirMap with the default parameters (27); TargetScan with a parameter to search for conserved sites for miRNA families conserved only among mammals (28); miRDB with the default parameters (29); and DIANA-microT with a threshold set to 0.5 (30). If miR-28-3p or miR-96-5p returned any positive result for the TF genes in the MODY pathway, a bidimensional matrix was populated (Supplementary Table 5).

Development and Evaluation of a Cervical Adenocarcinoma Classifier

We developed a classifier based on TCGA data, considering the ADC histological type (possible responses: yes or no, when it was an SCC case) as our outcome of interest. According to differential expression analysis, the set of predictors was composed of those genes identified as enriched in cervical ADC gene profile followed by TRN analysis. We applied logistic regression to adjust the classifier and leave-one-out cross-validation (CV) technique to evaluate its performance in data not used for its adjustment. Briefly, on each CV iteration, the observations are randomly divided into a training and a test set, then we adjust a logistic model in the training data and estimate the risk (probability) of being classified as ADC on the test data. The predictive performance was evaluated by calculating the area under the receiver operating characteristic (AUROC) curve with 95% confidence intervals. Sensitivity (S), specificity (E), positive predictive value (PPV), and a negative predictive value (NPV) were measured for different estimated probability (p) cutoff points: one that maximizes model's sensitivity and specificity, and other aiming to improve model's specificity and, accordingly, to reduce the false-positive results (i.e., an SCC case classified as ADC). The latter approach was based on a cutoff point that distinguishes the 10% of the observations with the highest predicted risk for ADC. The adjusted model was also applied to our dataset, and S, E, PPV, and NPV were estimated according to the cutoff points described earlier. Analyses were performed using R software. Model training

was carried out using the *caret* (Classification and Regression Training) package and model evaluation through pROC and ROCR packages.

RESULTS

Primary frozen tumor biopsies obtained from 30 treatment-naive women diagnosed with CC (25 SCCs and 5 ADCs) referred to the ICESP in São Paulo, Brazil, were submitted to HPV detection and typing (**Supplementary Table 1**). HPV was detected in 23 (92.3%) of the cervical SCC samples, and HPV-16 was the most frequent type found in 13 of 23 SCC samples. Co-infections by two HPV types were found in three SCC samples. In contrast, HPV-16 was the only type identified in all ADC cases (**Supplementary Table 1**). To investigate the gene expression profiles found in the two histological cancer types without the interference of different HPV types, the transcriptomes of seven SCC and three ADC HPV-16-positive samples were evaluated by RNA-Seq (**Supplementary Table 1**). As expected, viral E6 and E7 transcripts were detected in all samples submitted to RNA-Seq (**Supplementary Table 2**).

The transcriptome data obtained from our samples were compared with four HPV-negative non-tumoral cervical tissues (hereafter nominated non-CC) downloaded from the Gene Expression Omnibus repository (PRJNA454568) (10). **Figure 1A** depicts the strategy outlined in our study. Overall, 33,915 expressed genes were identified, of which 25,998, 29,319, and 30,758 were in non-CC, SCC, and ADC samples, respectively. Although most genes (23,336/33,915; 68.81%) were found to be expressed within all sample groups, almost a quarter (7,917/33,915; 23.34%) were exclusively expressed in SCC or ADC or in both CC histological types (**Figure 1B**). In non-CC samples, 25,998 genes were expressed among the 33,915 (76.66%) genes observed among all groups. In addition, 86.45% (29,319/33,915) and 90.69% (30,758/33,915) of the genes were expressed among cervical SCC and ADC samples, respectively. Among all genes identified within each group, ADC samples harbored the highest proportion of genes exclusively expressed in these tumors (2,244/30,758; 7.30%), followed by non-CC (1,581/25,998; 6.08%), and SCC (1,266/29,319; 4.32%).

The non-CC, SCC, and ADC samples expressed distinct sets of protein-coding (mRNA), lncRNA genes, and pseudogenes in comparison with the frequency of these gene categories within the human genome (**Figure 1C**). Next, we sought to investigate the enrichment of Gene Ontology biological processes of expressed genes in each group (**Figure 1D**, **Supplementary Figure 1**). Pathways related to epithelial differentiation and lipid metabolism were identified in SCC and ADC, respectively. The distinct sets of genes reflect the different biological processes occurring in each histological tumor type.

Differential expression analysis was performed between SCC and ADC samples and SCC or ADC *vs.* non-CC samples. According to our analysis, we identified 135 upregulated and 258 downregulated genes in SCC as compared with ADC samples; 1,413 genes upregulated in SCC as compared with non-CC and

1,260 genes upregulated in ADC as compared with non-CC (**Supplementary Tables 6–8**, **Supplementary Figures 2, 3**).

To identify gene expression patterns that might discriminate cervical SCC from ADC, we selected only those genes that were differentially expressed exclusively in each group or less expressed in non-CC samples, as shown in **Figure 2A**. The selected genes were investigated in the context of TRNs. This analysis identifies expression enrichment of specific TFs as key regulators of various target genes. However, due to our samples' small sample size and the need to use more than 100 samples to perform TRN analysis, we used only the TCGA data for this analysis. It is noteworthy that the differential gene expression profiles observed between cervical SCC and ADC in our study correlated to the respective profiles derived from the TCGA dataset with a Spearman's rank correlation coefficient of 0.79 (**Supplementary Figures 3, 4**, **Supplementary Table 3**).

Overall, 774 genes exclusively upregulated in SCC samples (682 genes from SCC *vs.* non-CC, 56 genes from SCC *vs.* ADC, and 36 overlapped genes from both SCC *vs.* ADC and SCC *vs.* non-CC) and 705 genes exclusively upregulated in ADC samples (135 genes from ADC *vs.* SCC, 491 genes from ADC *vs.* non-CC, and 79 overlapped genes from both ADC *vs.* SCC and ADC *vs.* non-CC) were used for the construction of TRNs (**Figure 2A**).

The distinct TRNs for SCC (**Figure 2B**) and ADC (**Figure 2D**) reflect biological processes enriched in each CC histological type, according to the Gene Ontology annotation (**Figures 2C,E**). The cervical SCC TRN showed *FOXN1*, *POU3F1*, *SOX15*, *ZNF185*, *HES2*, and *BNC1* TF genes with the potential to positively regulate several other protein-coding genes, lncRNAs, and pseudogenes (**Figure 2B**) that are related to cornification, keratinization, and epithelial differentiation processes (**Figure 2C**). Ten TFs within the cervical ADC TRN analysis, *PPARG*, *GATA4*, *HNF4G*, *DLX6*, *INSM1*, *MYB*, *HIF3A*, *SIM2*, *MEIS2*, and *HNF4A* (**Figure 2D**), are related to carbohydrate and glucose homeostasis processes (**Figure 2E**).

To confirm the observed differences in gene expression levels obtained from the differential expression analysis of our RNA-seq data, we selected six lncRNA genes differently expressed between cervical SCC and ADC for validation in a larger number of samples (**Supplementary Tables 4, 6**). qPCR results revealed that *DSG1-AS1*, *CALML3-AS1*, *IGFL2-AS1*, and *TINCR* genes were upregulated in cervical SCC as compared with ADC. Interestingly, these lncRNA genes are regulated by FOXN1 and HES2 TFs. *LINC02381* and *LINC01833* lncRNA genes were upregulated in cervical ADC as compared with SCC, corroborating our transcriptome results (**Figure 3**, **Supplementary Table 6**).

In addition to the biological processes observed in SCC and ADC samples (**Figure 2**), we investigated enriched biological pathways using the KEGG database based on genes exclusively expressed in each CC histological type (**Supplementary Table 9**). Several pathways were enriched in SCC, such as the estrogen signaling pathway (KEGG ID map04915; **Supplementary Figure 5**), whereas the MODY pathway was enriched in ADC (**Figure 4A**, **Supplementary Figure 6**). At last, we filtered all TFs from the MODY pathway (**Supplementary Table 10**) together with

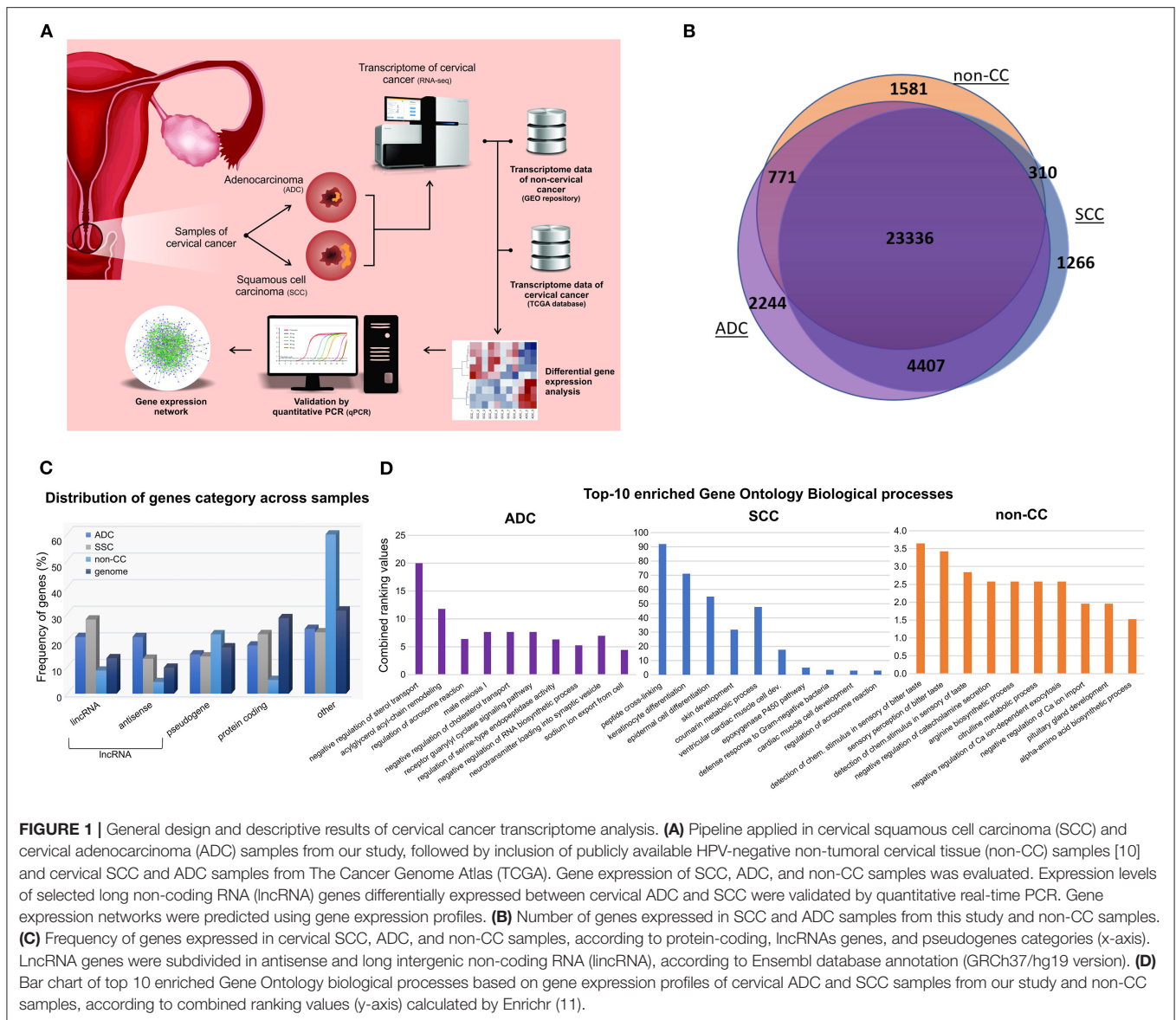


FIGURE 1 | General design and descriptive results of cervical cancer transcriptome analysis. **(A)** Pipeline applied in cervical squamous cell carcinoma (SCC) and cervical adenocarcinoma (ADC) samples from our study, followed by inclusion of publicly available HPV-negative non-tumoral cervical tissue (non-CC) samples [10] and cervical SCC and ADC samples from The Cancer Genome Atlas (TCGA). Gene expression of SCC, ADC, and non-CC samples was evaluated. Expression levels of selected long non-coding RNA (lncRNA) genes differentially expressed between cervical ADC and SCC were validated by quantitative real-time PCR. Gene expression networks were predicted using gene expression profiles. **(B)** Number of genes expressed in SCC and ADC samples from this study and non-CC samples. **(C)** Frequency of genes expressed in cervical SCC, ADC, and non-CC samples, according to protein-coding, lncRNAs genes, and pseudogenes categories (x-axis). lncRNA genes were subdivided in antisense and long intergenic non-coding RNA (lincRNA), according to Ensembl database annotation (GRCh37/hg19 version). **(D)** Bar chart of top 10 enriched Gene Ontology biological processes based on gene expression profiles of cervical ADC and SCC samples from our study and non-CC samples, according to combined ranking values (y-axis) calculated by Enrichr (11).

upregulated genes detected in the TRN of ADC to obtain a MODY-specific TRN in ADC. This analysis identified *HNF4A*, *HNF4G*, *FOXA2*, and *PAX6* TF genes in cervical ADC TRN, including a set of protein-coding and lncRNA genes, such as the *LINC02381* gene (highlighted in red in **Figure 4B**).

To evaluate the potential of genes encoding MODY pathway TFs in discriminating between ADC and SCC CC types, we adjusted a classifier considering the ADC histological type (possible responses: yes or no, when it was an SCC case) as our outcome of interest and the set of genes *HNF4A*, *HNF4G*, *FOXA3*, *PAX6*, *HNF1A*, *PDX1*, as well as the lncRNA *LINC02381*, as our predictors. For each iteration of the leave-one-out CV process, the adjusted classifier was evaluated in the test data. At the end of the CV process, the estimated risk from iterative test data

was aggregated to estimate the AUROC curve, which results indicate a good discrimination performance for our classifier (AUROC curve: 0.8978; 95%; confidence interval: 0.808–0.987) (**Figure 5**).

Two scenarios were defined according to the cutoff points we adopted for the estimated probability. For the cutoff point that maximizes model's sensitivity and specificity ($p \geq 0.077$), it is possible to successfully predict 87.10% (27 of 31 cases) of ADC cases. Nonetheless, it is important to note that, although the negative predictive value is close to 1 (that is, when the classifier indicates an SCC case, it is almost always correct) because ADC is a rare event, the positive predictive value is low (if the classifier indicates an ADC case, then further investigation is necessary because, in half of the times, it is a false result). So, this would be a good classifier to rule out the possibility of SCC, requiring as a next step some approach to confirm

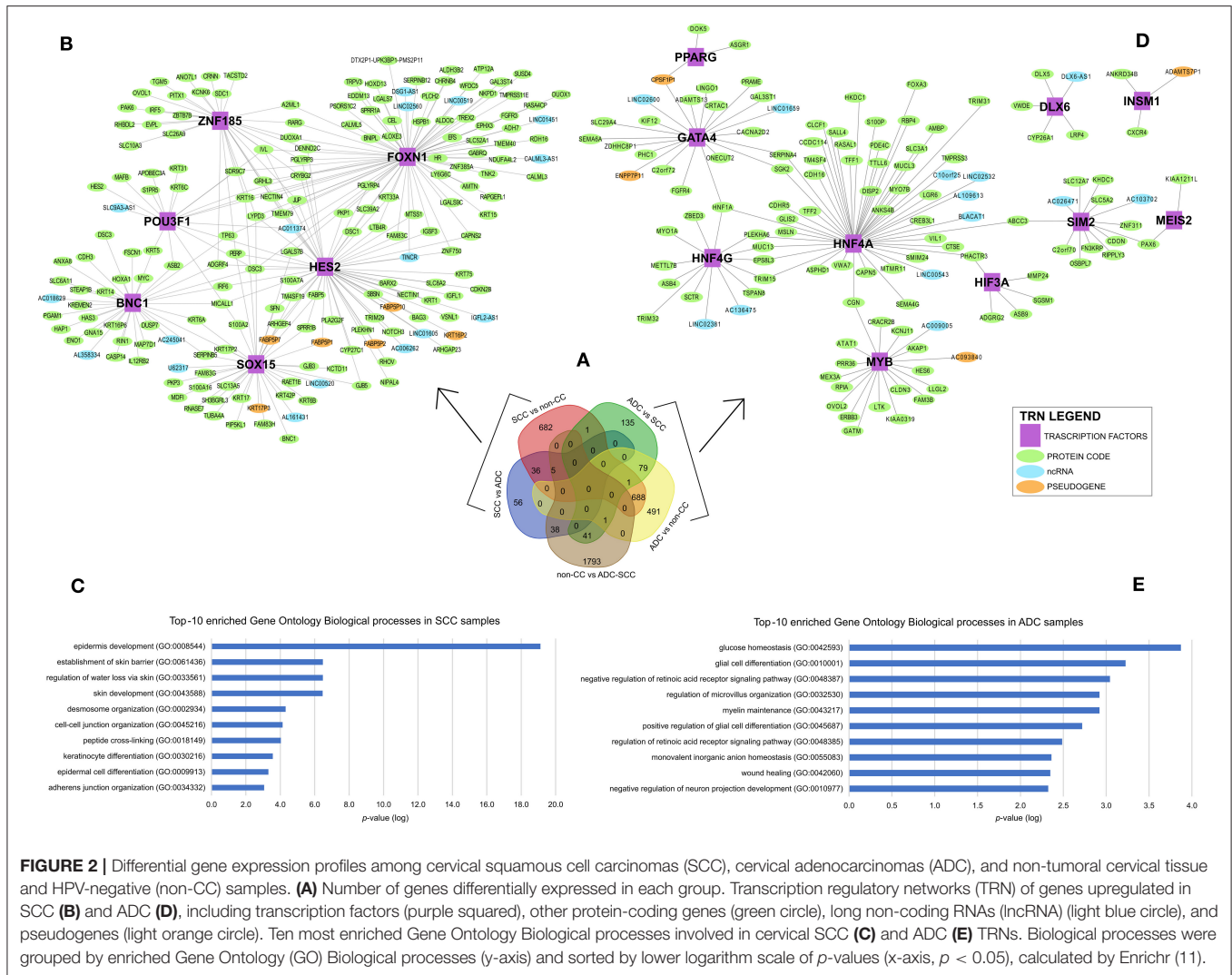


FIGURE 2 | Differential gene expression profiles among cervical squamous cell carcinomas (SCC), cervical adenocarcinomas (ADC), and non-tumoral cervical tissue and HPV-negative (non-CC) samples. **(A)** Number of genes differentially expressed in each group. Transcription regulatory networks (TRN) of genes upregulated in SCC **(B)** and ADC **(D)**, including transcription factors (purple squared), other protein-coding genes (green circle), long non-coding RNAs (lncRNA) (light blue circle), and pseudogenes (light orange circle). Ten most enriched Gene Ontology Biological processes involved in cervical SCC **(C)** and ADC **(E)** TRNs. Biological processes were grouped by enriched Gene Ontology (GO) Biological processes (y-axis) and sorted by lower logarithm scale of p-values (x-axis, $p < 0.05$), calculated by Enrichr (11).

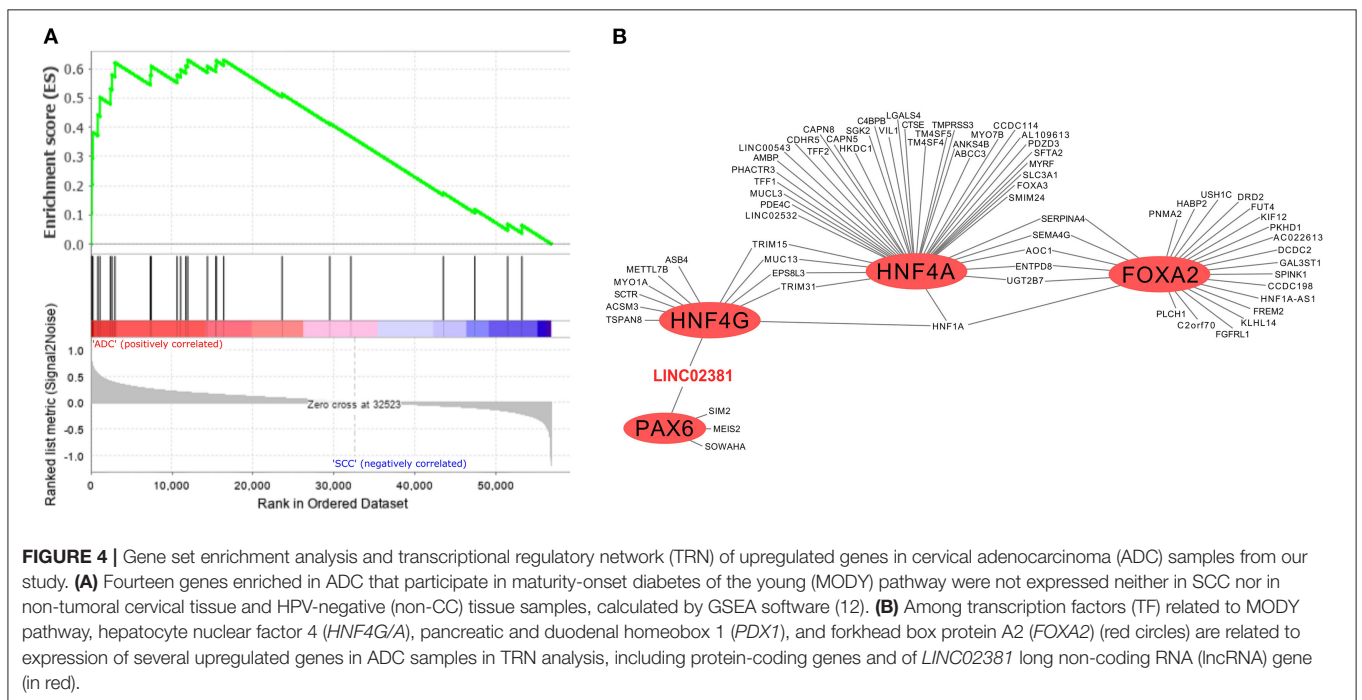
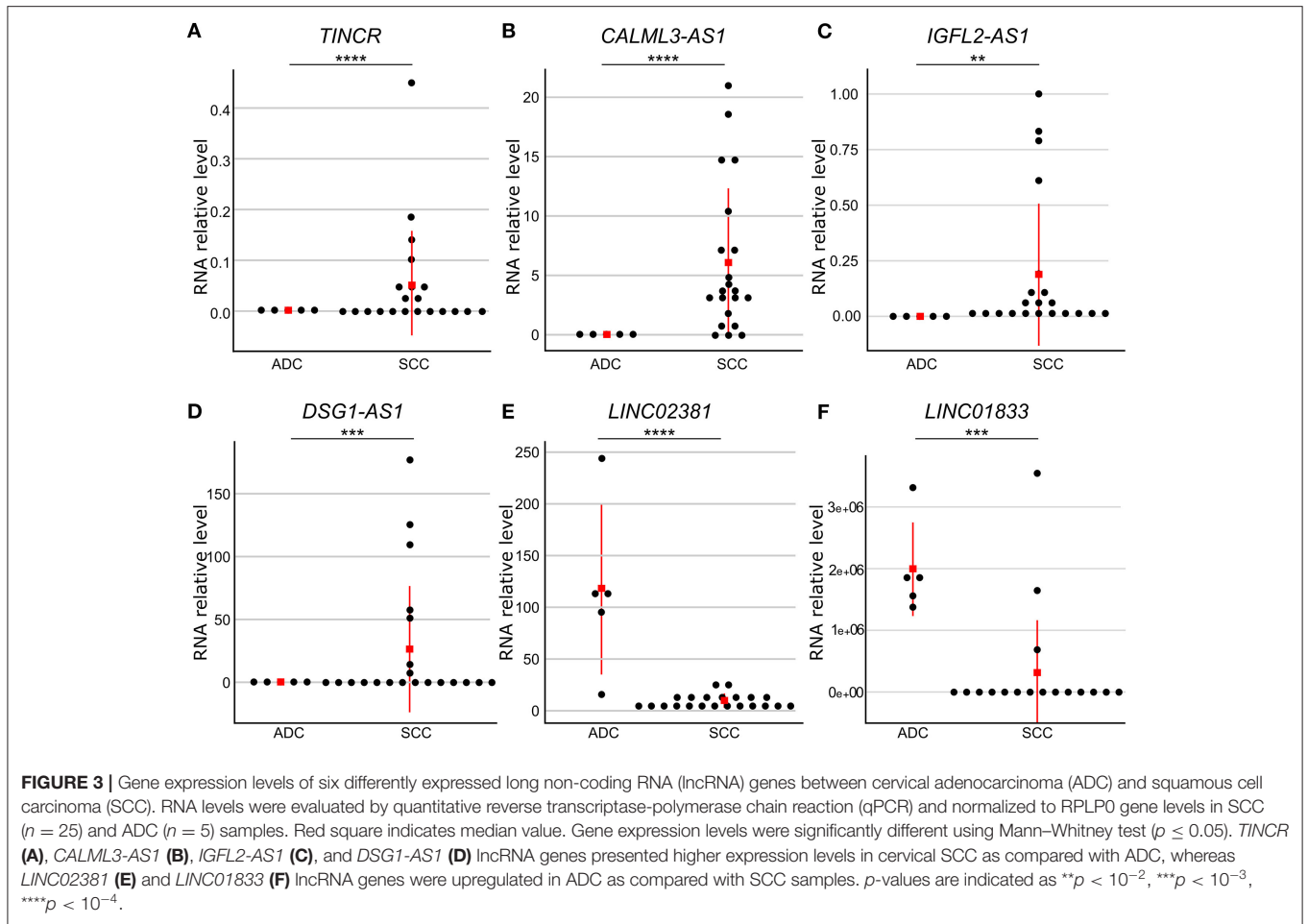
ADC. Additionally, aiming to improve the model's specificity and, accordingly, to reduce false-positive results, we chose a cutoff point that might discriminate those patients in the 10% highest risk strata ($p \geq 0.2082$). In this scenario, it is possible to successfully predict 67.74% (21 of 31 cases) of ADC cases. Additionally, because the false-positive rate has reduced from 8 to 3% (in comparison with the first scenario we described), now the positive predictive value is 75.00%. This classifier would be a good tool when we want to be more confident about the classifier result because, for example, it is necessary to prioritize a group (higher risk ADC group) for some intervention (**Supplementary Table 11**).

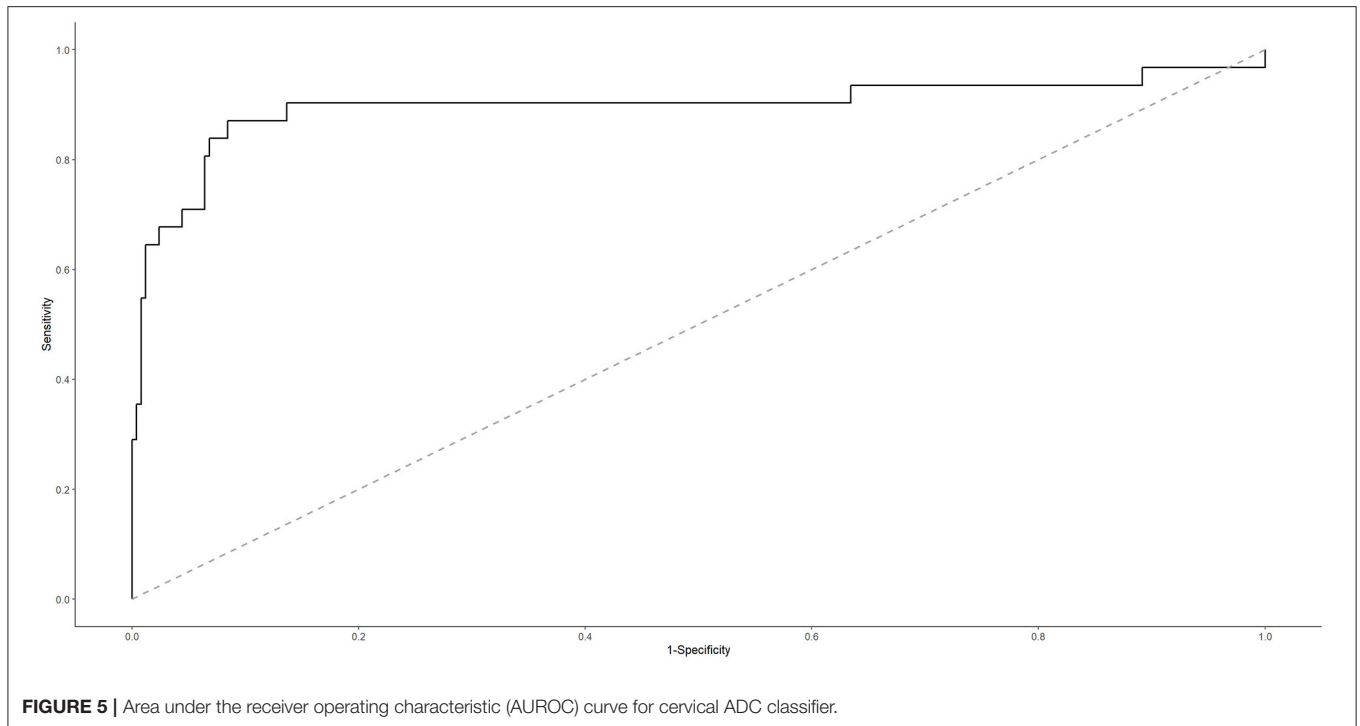
Considering the former or the latter cutoff point for classifying our study data led to the same results: S, E, PPV, and NPV were, respectively, 0.667, 1.00, 1.00, and 0.875. Of the three ADC cases, two presented an estimated probability higher than that of the SCC cases; however, one of the ADC samples had an estimated probability of 0.016 (below the two cutoff points described earlier), and it was labeled as SCC (a false-negative result).

DISCUSSION

Differences in cervical SCC and ADC transcriptomes have been previously investigated, focusing on protein-coding genes. Additionally, functional network analysis (31) and drug repurposing for cervical ADC or cervical SCC have been proposed (32, 33). However, no set of TFs and lncRNAs coordinately regulated with the potential to explain cervical ADC carcinogenesis have been described to date. In this study, we confirmed the upregulation of protein-coding genes previously identified in ADC, such as *IGFBP2* (34), and in SCC, we identified *CDKN2A*, which encodes p16 (35), and *E2F1* (36).

Furthermore, we describe TRNs that discriminate between HPV-16-positive cervical SCC and ADC and between both histological types and non-CC tissues. Genes found upregulated among SCC samples are involved in biological processes related to epithelium structure and function, including maintenance of skin barrier due to epithelium attachment, in addition to cornification and keratinization processes. We observed





the enrichment of keratin coding genes highly expressed in cervical SCC, similar to previous data of the TCGA consortium (34). On the other hand, the highly expressed genes among cervical ADC samples are associated with carbohydrate and glucose homeostasis.

It is important to note, however, that one limitation of this study is that we used previously published RNA-Seq data from normal cervical tissues (non-CC), which may have a distinct genetic background compared with the CC samples sequenced here. Furthermore, the restricted sample number unable us to discriminate whether the genes are differentially expressed between tumor and non-tumor samples results from HPV-16 infection. Nevertheless, despite the few cervical ADC samples included in this study, we believe that our data enabled the identification of biological signals to computationally discriminate cervical SCC from ADC. Such promising findings must be further confirmed and investigated in depth by additional studies.

Epithelium Structure and Function Pathways Are Enriched in Cervical Squamous Cell Carcinoma Gene Profile

The upregulation of genes involved in epithelium maintenance among cervical SCC samples was further endorsed by the TRN analysis, which identified regulons whose lncRNA genes are highly expressed in SCC in contrast to ADC or non-CC samples. Epithelial differentiation and homeostasis are governed by several TFs (37–39). Among our findings, POU Class 3 Homeobox 1 (*POU3F1*) (40), Basonuclin 1 (*BNC1*) (41), Human Forkhead-box N1 gene (*FOXN1*) (42), hes family bHLH

transcription factor 2 (*HES2*) (43), and zinc finger protein 185 with LIM domain (*ZNF185*) (44) were previously described to be involved in keratinocyte differentiation, or skin development, homeostasis, and wound healing.

One important regulator of the expression of several genes involved in keratinocyte proliferation and differentiation is p63 (45), which is encoded by the tumor protein p63 (*TP63*) gene. We found *TP63* upregulated in cervical SCC as compared with ADC and non-CC (**Supplementary Table 6, Figure 2B**). Indeed, p63 was previously detected by immunohistochemistry in 97% of cervical SCC, whereas it was absent in ADC (46). In this study, *TP63* integrates the SCC TRN as a gene potentially regulated by *BNC1* and *FOXN1*. Expression of *BNC1* and *ZNF185* have been previously described to be positively regulated by p63 (47, 48). Higher expression levels of *BNC1* and SRY-Box Transcription Factor 15 (*SOX15*) genes have also been previously detected in SCC samples from other anatomical sites compared with their related non-tumoral tissues (49–51). In contrast, *ZNF185* displayed a discordant expression pattern in cervical SCC compared with our findings (48). However, because the expression of the protein encoded by this gene was shown to vary widely among skin layers (44), we suggest that our findings must be investigated in-depth in the future. Two of the six TFs within the cervical SCC TRN were also previously reported to be positively regulated by p63. We thus conclude that the cervical SCC TRN proposed in the present study encompasses a set of TFs with functions associated with epithelial differentiation and homeostasis, supporting its role in cervical SCC.

In addition to the putative role of different TFs upon cervical SCC TRN, several lncRNAs genes regulated by these proteins might also affect integrity maintenance and differentiation of

the epithelium. Using qPCR, we validated the higher expression of *TINCR* ubiquitin domain-containing (*TINCR*), *CALML3* antisense RNA 1 (*CALML3-AS1*), *DSG1* antisense RNA 1 gene (*DSG1-AS1*), and IGF-like family member 2 antisense RNA 1 (*IGFL2-AS1*) lncRNA genes in cervical SCC in comparison with cervical ADC samples. Nevertheless, it needs to be highlighted that, to date, the functions of most of the cataloged human lncRNA are not clarified due to the lack of specific literature.

Increased *TINCR* lncRNA levels have been found within differentiated superficial epidermis layers and were shown to positively regulate the expression of the MAF:MAFB TF dimer (52). Corroborating our findings, higher expression of *TINCR* in cervical SCC was previously observed (53), although these data contrast with data obtained from TCGA (54). All other validated lncRNA genes in our study were classified as antisense transcripts (55). We observed higher expression levels of the *IGFL2-AS1* in cervical SCC compared with cervical ADC and non-CC. Higher levels of this antisense gene have been described in renal cell carcinoma (56) and gastric cancer (57); in contrast, a lower expression of *IGFL2-AS1* was observed in breast ADCs (58). The *CALML3-AS1* lncRNA was shown to increase proliferation, migration, and invasion of CC cells and to decrease the apoptosis of these cells. Indeed, high *CALML3-AS1* levels were previously observed in CC samples of TCGA (59). In our study, lower *CALML3-AS1* levels were detected in cervical ADC and non-CC compared with cervical SCC, indicating this molecule may be further explored as a molecular target for specific diagnosis of this histological tumor type.

In conclusion, although the function of most of the lncRNAs remains unclear, a selected set of such transcripts were successfully validated as highly expressed in cervical SCC as compared with ADC. Among them are *TINCR* and *CALML3-AS1*, whose role in CC is under investigation. Taken together, our data reinforce the molecular differences between cervical SCC and ADC, whose significance should be explored in the prognostic and therapeutic management of these tumors.

Genes Encoding Maturity-Onset Diabetes of the Young Pathway Transcription Factors Are Enriched in Cervical Adenocarcinoma Gene Profile

ADC arises from the cervical glandular tissue and leads to abnormal epithelium structure and production of secreted molecules. In this study, we describe a set of 10 TF genes upregulated in ADC as compared with SCC or non-CC. All TFs within the ADC TRN have been previously associated with the genesis of ADCs, in particular pancreatic ductal adenocarcinoma (PDA), with the exception of the insulinoma-associated protein 1 (*INSM1*). Although no literature regarding *INSM1* in PDA, it was described to have a key role in pancreatic development (60), similar to the GATA binding protein 4 (*GATA4*) (61).

GATA4 gene expression levels were shown to vary in ADCs arising from distinct tissues (62, 63). In PDA, *GATA4* protein levels were detected with higher intensity in women than in men (64). In addition to *GATA4* gene expression, the peroxisome proliferator-activated receptor gamma gene (*PPARG*) (65), the

hepatocyte nuclear factor-4-alpha (*HNF4A*) (66), the hypoxia-inducible factor 3 subunit alpha (*HIF3A*) (67), the MYB proto-oncogene (*MYB*) (68), and the single-minded 2 (*SIM2*) (69) genes were also previously described to be highly expressed in PDA. To our knowledge, this is the first report to identify the surprising similarity of TFs overexpressed in cervical ADC and PDA. Such similarity encourages additional investigations. Interestingly, although highly expressed in PDA, high levels of *SIM2* in cervical SCC have been associated with better overall survival compared with tumors expressing low levels of this gene (70). The Meis homeobox 2 (*MEIS2*) gene encodes for a TF that was identified as a key signal transduction pathway regulator in breast ADC (71). Recently, *MEIS2* was identified as highly expressed in cervical SCC and showed prognostic value (54). Therefore, we believe that additional studies must be carried out to clarify the possible roles of *SIM2* and *MEIS2* in the distinct histological types of CC.

The pathway enrichment analysis in cervical ADC unveiled the MODY pathway. MODY TFs have also been extensively investigated in the context of pancreas and liver development. In addition to *HNF4A* expression in PDA, *HNF4A* and *HNF4G* genes were reported to be positively regulated by HNF1 homeobox A (*HNF1A*) protein in the pancreatic cell (72). Corroborating our findings, *HNF1A* was shown to have higher expression levels in CC compared with the normal cervix (73), despite the fact that the histological type was not indicated in their study. Nevertheless, we may speculate that *HNF1A* was identified as differentially expressed in their study due to an overrepresentation of cervical ADC among samples. Another MODY pathway member, the gene encoding the TF pancreatic and duodenal homeobox 1 (*PDX1*), had the literature reviewed in the context of pancreas embryogenesis (74), including PAD (75). In an extensive analysis regarding pancreatic cancer, *PDX1*, *MNX1*, *HNF4G*, *HNF4A*, *HNF1B*, *HNF1A*, *FOXA2*, *FOXA3*, and *HES1* were proposed as the key TFs associated with the pancreatic progenitor subtype (76). Except for *FOXA2* and *HES1*, all of these TFs were identified to be overexpressed in cervical ADC ($\log_2FC \geq 2$ and false discovery rate < 0.05 ; **Supplementary Table 8**), providing stronger support to the hypothesis that cervical ADC and PDA share molecular similarities. It is important to stress, however, that such relationships regarding cervical ADC regulation are predictions and must be further investigated and validated.

The construction of a TRN including all MODY TFs highly expressed in cervical ADC revealed *PAX6*, *HNF4A*, *HNF4G*, and *FOXA2* genes with the potential to positively regulate other protein-coding or lncRNA genes (**Figure 4B**). At the moment, the role of these TFs within the context of cervical ADC is unknown. However, the MODY pathway was previously associated with other ADCs, such as EAC and PAD. Our TRN analysis showed that *PAX6* and *HNF4G* TF genes might positively regulate the expression of the lncRNA *LINC02381* gene (also known as *LOC400043*). Although there is limited literature regarding *LINC02381*'s biological role, it was described to have a suppressive effect upon the tumorigenesis of human colorectal ADC (77). These data indicate that MODY TF genes may be associated with other molecular processes in addition to those

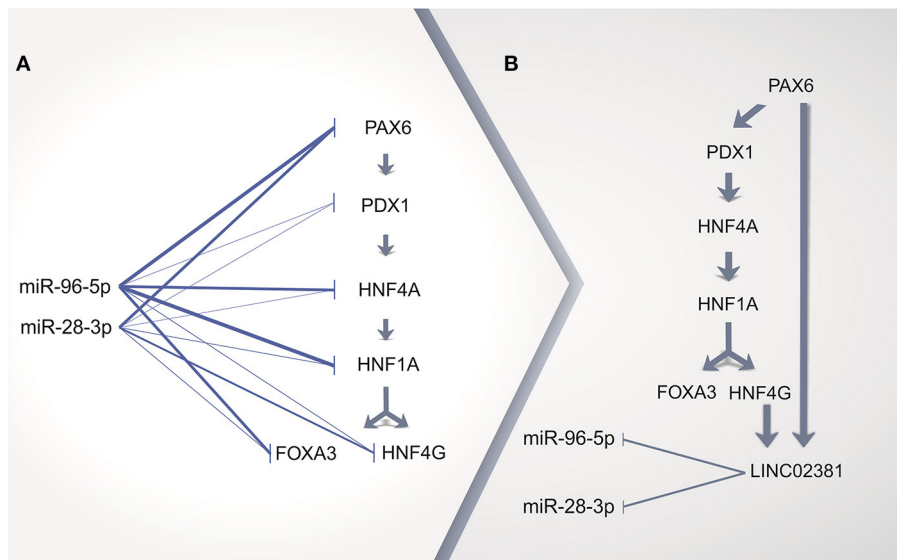


FIGURE 6 | Proposed scheme of cervical adenocarcinoma (ADC) gene regulation by long non-coding RNA (lncRNA) acting as a sponge of microRNA (miRNA). **(A)** *in silico* prediction analysis of miRNA repressing all transcription factor (TF) genes upregulated in cervical ADC as compared with squamous cell carcinoma (SCC) in maturity-onset diabetes of the young (MODY) KEGG pathway. **(B)** TF genes HNF4G and PAX6 might upregulate LINC02381 lncRNA, which could act as a miRNA sponge, as positive feedback for all TF genes pathway expression. Thicker lines on **(A)** represent more positive predictions of miRNA targeting to a given mRNA (**Supplementary Table 5**).

previously described in onset diabetes. Additionally, we were able to validate that *LINC02381* and *LINC01833* had higher expression levels in cervical ADC than in cervical SCC by qPCR.

Although the role of lncRNAs in CC development is not completely elucidated, it is well-established that they can compete for binding to mRNAs with other regulatory molecules, such as miRNAs (78, 79). In the context of HPV-associated tumors, several studies indicate that E6/E7 viral oncoproteins target lncRNAs, subverting cellular processes seminal to carcinogenesis (80). Previous *in silico* miRNA to mRNA targeting predictions analyses, followed by experimental validation, identified that miR-96-5p and miR-28-3p could directly bind to *LINC02381* (79).

The inspection of the entire MODY pathway (KEGG ID hsa04950) regarding the highly expressed TFs in cervical ADC allowed for the identification of the following set of MODY TF genes regulated in sequence: *PAX6*, *PDX1*, *HNF4A*, *HNF1A*, *HNF4G*, and *FOXA3* (**Supplementary Table 10**, **Supplementary Figure 6**). Surprisingly, according to at least one of five prediction tools, this set of TF genes within the MODY pathway is a potential target of either miRNAs miR-96-5p or miR-28-3p (**Figure 6A**, **Supplementary Table 5**). Interestingly, miR-96-5p has been associated with CC tumorigenesis progression by regulating the PTEN pathway (81).

Integration of MODY TRN and MODY TFs regulated in sequence in cervical ADC revealed that *HNF4G* and *PAX6* might lead to increased expression of *LINC02381*. We speculate that the enhanced expression of *LINC02381* in cervical ADC could dysregulate the repression role of miR-96-5p and miR-28-3p, acting as a sponge of such miRNAs (79). Therefore, the repressed

MODY pathway previously silenced in the cervix would become active in cervical ADC (**Figure 6B**).

Taken together, our TRN analysis revealed a set of MODY TFs upregulated in cervical ADC. Among these, two TFs could positively regulate the lncRNA *LINC02381*. According to the literature, *LINC02381* can directly bind to two miRNAs acting as a sponge, reducing the availability of miR-96-5p and miR-28-3p. Computational predictions showed that miR-96-5p and miR-28-3p might negatively regulate this set of TFs. Therefore, we speculate *LINC02381* may act by decreasing the levels of miR-96-5p and miR-28-3p and promoting the upregulation of the MODY pathway in cervical ADC. Such findings, when considered as predictors in a classifier to differentiate ADC and SCC cases, resulted in encouraging results. Nonetheless, it is important to note that external validation, based on a larger dataset, is required to corroborate our results because, due to the small sample size (especially for the positive ADC type), we chose to use leave-one-out CV process and, therefore, effectively all data were used to estimate the predictive performance of the classifier, as this process implies a rotation of roles (training and testing) for each sample at each iteration performed.

CONCLUSION

Cervical SCC and ADC patients frequently receive identical treatments. Therefore, it is urgent to identify novel therapeutic targets to treat cervical ADC. In the present study, we report distinct TRNs for cervical SCC and ADC. Further analysis uncovered the TF encoding genes *PAX6*, *PDX1*, *HNF4A*, *HNF1A*,

HNF4G, and *FOXA3* with higher expression levels in cervical ADC as compared with cervical SCC or HPV-negative non-tumoral cervical tissues. This set of six TF genes is regulated in sequence in the *MODY* pathway. Based on bioinformatics predictions, we propose a hypothetical activation molecular mechanism of this set of six TFs, whose importance has been reported previously in pancreatic cancer. Our findings might provide novel directions to the development of more specific diagnostic tools for cervical ADC.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/>, PRJNA521318.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo Av. Dr. Arnaldo, 251 - 21^o andar - sala 36 Cerqueira César - São Paulo - SP - Brazil, CEP: 01246-000, +55 11 3893 4401, cep.fm@usp.br. This study was performed according to the Brazilian federal laws and approved by the Ethics Committee of the Faculdade de Medicina, Universidade de São Paulo (National Committee in Research Ethics, CONEP, number 53156815.5.0000.0065). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PSA-S, LLV, LS, and FP: conceptualization. SB, TDJF, PSA-S, NANJ, GW, HGS, RC, LS, and FP: data curation. SB, TDJF, PSA-S, NANJ, HGS, RC, LS, and FP: formal analysis. LLV: funding acquisition. SB, TDJF, PSA-S, NANJ, HGS, MAAC, RC, LLV, LS, and FP: investigation. SB, TDJF, PSA-S, NANJ, HGS, MAAC, NMS, GW, RC, LLV, LS, and FP: methodology. MLNDG, JPC, LLV, and LS: sample acquisition. PSA-S, LLV, LS, and FP: supervision. SB, TDJF, PSA-S, and FP: writing—original draft. SB, TDJF, PSA-S, NANJ, LLV, LS, and FP: writing—review and

editing. All the authors contributed to the manuscript revision, read and approved the submitted version.

FUNDING

This work was supported by Carlos Chagas Institute, Fiocruz Paraná (post-doctoral scholarship to SB), Inova Fiocruz/Fundação Oswaldo Cruz (post-doctoral scholarship to TDJF and SB), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001 (Ph.D. scholarship to NANJ and GW), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) scholarship (Ph.D. scholarship 154933/2016-3 to RC) and grant (573799/2008-3 and 306326/2015-9 to LLV; 303431/2018-0 to LS; 308697/2019-7 to FP), and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Grant number 08/57889-1 to LLV). The authors declare that any of the above grants have funds to pay for open access publication fees.

ACKNOWLEDGMENTS

We would like to thank all patients that donated samples to our study, as well the hospital staff. We would like to thank Miuyki Uno and Maria José Ferreira Alves from the Tumor Biobank at ICESP for helping with sample management and DNA and RNA extraction and Barbara Mello for HPV DNA detection and genotyping. We would like to thank the digital designer Wagner Nagib from Carlos Chagas Institute, Fiocruz Paraná, for helping us with the design of **Figures 1, 6**. We thank Lucas Antunes Wendhausen for helping us with **Figures 2, 4** in high resolution. The authors acknowledge the Fiocruz Bioinformatics Facility RPT04A for computational power to perform the analyses of this study. We thank the English language editing service supplied by Dr. Matthew Thomas Ferreira. The Illumina HiSeq 2500 image in **Figure 1A** is a courtesy of Illumina, Inc.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.626187/full#supplementary-material>

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *Cancer J Clin.* (2019) 69:7–34. doi: 10.3322/caac.21551
- Schiffman M, Doorbar J, Wentzensen N, De Sanjosé S, Fakhry C, Monk BJ, et al. Carcinogenic human papillomavirus infection. *Nat Rev Dis Prim.* (2016) 2:16086. doi: 10.1038/nrdp.2016.86
- Gien LT, Beauchemin MC, Thomas G. Adenocarcinoma: a unique cervical cancer. *Gynecol Oncol.* (2010) 116:140–6. doi: 10.1016/j.ygyno.2009.09.040
- Castanon A, Landy R, Sasieni PD. Is cervical screening preventing adenocarcinoma and adenosquamous carcinoma of the cervix. *Int J Cancer.* (2016) 139:1040–5. doi: 10.1002/ijc.30152
- De Jong JJ, Liu Y, Robertson AG, Seiler R, Groeneveld CS, Van Der Heijden MS, et al. Long non-coding RNAs identify a subset of luminal muscle-invasive bladder cancer patients with favorable prognosis. *Genome Med.* (2019) 11:60. doi: 10.1186/s13073-019-0669-z
- Campbell TM, Castro MAA, de Oliveira KG, Ponder BAJ, Meyer KB. Era binding by transcription factors NFIB and YBX1 enables FGFR2 signaling to modulate estrogen responsiveness in breast cancer. *Cancer Res.* (2018) 78:410–21. doi: 10.1158/0008-5472.CAN-17-1153

8. Han L, Yuan Y, Zheng S, Yang Y, Li J, Edgerton ME, et al. The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. *Nat Commun.* (2014) 5:3963. doi: 10.1038/ncomms4963
9. Carlevaro-Fita J, Lanzas A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, et al. Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun Biol.* (2020) 3:56. doi: 10.1038/s42003-019-0741-7
10. Xu J, Liu H, Yang Y, Wang X, Liu P, Li Y, et al. Genome-wide profiling of cervical rna-binding proteins identifies human papillomavirus regulation of rnaaseh2a expression by viral e7 and e2f1. *mBio.* (2019) 10:e02687-18. doi: 10.1128/mBio.02687-18
11. Felix Krueger. *Trim Galore*. Available online at: <https://github.com/FelixKrueger/TrimGalore>
12. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* (2015) 12:357-60. doi: 10.1038/nmeth.3317
13. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc.* (2009) 4:1184-91. doi: 10.1038/nprot.2009.97
14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* (2009) 25:2078-9. doi: 10.1093/bioinformatics/btp352
15. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* (2011) 29:24-6. doi: 10.1038/nbt.1754
16. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics.* (2015) 31:166-9. doi: 10.1101/002824
17. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2009) 26:139-40. doi: 10.1093/bioinformatics/btp616
18. Castro MAA, De Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat Genet.* (2015) 48:12-21. doi: 10.1038/ng.3458
19. Meyer PE, Lafitte F, Bontempi G. Minet: a r/bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinform.* (2008) 9:461. doi: 10.1186/1471-2105-9-461
20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498-504. doi: 10.1101/gr.1239303
21. Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* (2016) 44:W90-7. doi: 10.1093/nar/gkw377
22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert B, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide. *Proc Natl Acad Sci USA.* (2005) 102:15545-50. doi: 10.1073/pnas.0506580102
23. Volders PJ, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, et al. Lncpedia 5: towards a reference set of human long non-coding rnas. *Nucleic Acids Res.* (2019) 47:D135-9. doi: 10.1093/nar/gky1031
24. GraphPad. *GraphPad Software*. San Diego, CA. Available online at: www.graphpad.com
25. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer (2010). doi: 10.1007/978-0-387-98141-3
26. Tokar T, Pastrello C, Rossos AEM, Abovsky M, Hauschild AC, Tsay M, et al. MirDIP 4.1 - Integrative database of human microRNA target predictions. *Nucleic Acids Res.* (2018) 46:D360-70. doi: 10.1093/nar/gkx1144
27. Vejnar CE, Zdobnov EM. MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.* (2012) 40:11673-83. doi: 10.1093/nar/gks901
28. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife.* (2015) 4:e05005. doi: 10.7554/eLife.05005.028
29. Chen Y, Wang X. MiRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* (2020) 48:D127-31. doi: 10.1093/nar/gkz757
30. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* (2013) 41:W169-73. doi: 10.1093/nar/gkt393
31. Kim YW, Bae SM, Kim YW, Park DC, Lee KH, Liu H-B, et al. Target-based molecular signature characteristics of cervical adenocarcinoma and squamous cell carcinoma. *Int J Oncol.* (2013) 43:539-47. doi: 10.3892/ijo.2013.1961
32. Yang PM, Chou CJ, Tseng SH, Hung CF. Bioinformatics and *in vitro* experimental analyses identify the selective therapeutic potential of interferon gamma and apigenin against cervical squamous cell carcinoma and adenocarcinoma. *Oncotarget.* (2017) 8:46145-62. doi: 10.18632/oncotarget.17574
33. van Dam PA, van Dam PJHH, Rolfo C, Giallombardo M, van Berckelaer C, Trinh XB, et al. *In silico* pathway analysis in cervical carcinoma reveals potential new targets for treatment. *Oncotarget.* (2016) 7:2780-95. doi: 10.18632/oncotarget.6667
34. The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature.* (2017) 543:378-84. doi: 10.1038/nature21386
35. Park JS, Dong SM, Kim HS, Lee JY, Um SJ, Park IS, et al. Detection of p16 gene alteration in cervical cancer using tissue microdissection and LOH study. *Cancer Lett.* (1999) 136:101-8. doi: 10.1016/S0304-3835(98)00366-8
36. Santin AD, Zhan F, Bignotti E, Siegel ER, Cané S, Bellone S, et al. Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology.* (2005) 331:269-91. doi: 10.1016/j.virol.2004.09.045
37. Bukowska J, Kopcewicz M, Walendzik K, Gawronska-Kozak B. Foxn1 in skin development, homeostasis and wound healing. *IJMS.* (2018) 19:1956. doi: 10.3390/ijms19071956
38. Siebel C, Lendahl U. Notch signaling in development, tissue homeostasis, and disease. *Physiol Rev.* (2017) 97:1235-94. doi: 10.1152/physrev.00005.2017
39. Tremblay M, Sanchez-Ferraz O, Bouchard M. GATA transcription factors in development and disease. *Development.* (2018) 145:dev164384. doi: 10.1242/dev.164384
40. Faus I, Hsu HJ, Fuchs E. Oct-6: a regulator of keratinocyte gene expression in stratified squamous epithelia. *Mol Cell Biol.* (1994) 14:3263-75. doi: 10.1128/MCB.14.5.3263
41. Tseng H, Green H. Basonuclin: a keratinocyte protein with multiple paired zinc fingers. *Proc Natl Acad Sci USA.* (1992) 89:10311-5. doi: 10.1073/pnas.89.21.10311
42. Janes SM. Transient activation of FOXN1 in keratinocytes induces a transcriptional programme that promotes terminal differentiation: contrasting roles of FOXN1 and Akt. *J Cell Sci.* (2004) 117:4157-68. doi: 10.1242/jcs.01302
43. Ishibashi M, Sasai Y, Nakanishi S, Kageyama R. Molecular characterization of HES-2, a mammalian helix-loop-helix factor structurally related to Drosophila hairy and Enhancer of split. *Eur J Biochem.* (1993) 215:645-52. doi: 10.1111/j.1432-1033.1993.tb18075.x
44. Smirnov A, Cappello A, Lena AM, Anemona L, Mauriello A, Di Daniele N, et al. ZNF185 is a p53 target gene following DNA damage. *Aging.* (2018) 10:3308-26. doi: 10.18632/aging.101639
45. Kouwenhoven EN, Oti M, Niehues H, van Heeringen SJ, Schalkwijk J, Stunnenberg HG, et al. Transcription factor p63 bookmarks and regulates dynamic enhancers during epidermal differentiation. *EMBO Rep.* (2015) 16:863-78. doi: 10.15252/embr.201439941
46. Wang T, Chen B, Yang Y, Chen H, Wang Y, Cviko A, et al. Histologic and immunophenotypic classification of cervical carcinomas by expression of the p53 homologue p63: a study of 250 cases. *Hum Pathol.* (2001) 32:479-86. doi: 10.1053/hupa.2001.24324
47. Boldrup L, Coates PJ, Laurell G, Nylander K. p63 transcriptionally regulates BNC1, a Pol I and Pol II transcription factor that regulates ribosomal biogenesis and epithelial differentiation. *Eur J Cancer.* (2012) 48:1401-6. doi: 10.1016/j.ejca.2011.06.032
48. Smirnov A, Lena AM, Cappello A, Panatta E, Anemona L, Bischetti S, et al. ZNF185 is a p63 target gene critical for epidermal differentiation and squamous cell carcinoma development. *Oncogene.* (2019) 38:1625-38. doi: 10.1038/s41388-018-0509-4

49. Cui C, Elsam T, Tian Q, Seykora JT, Grachtchouk M, Dlugosz A, et al. Gli proteins up-regulate the expression of basonuclin in basal cell carcinoma. *Cancer Res.* (2004) 64:5651–8. doi: 10.1158/0008-5472.CAN-04-0801
50. Zhang M, Wang J, Gao T, Chen X, Xu Y, Yu X, et al. Inhibition of SOX15 sensitizes esophageal squamous carcinoma cells to paclitaxel. *CMM.* (2019) 19:349–56. doi: 10.2174/1566524019666190405121139
51. Thu KL, Becker-Santos DD, Radulovich N, Pikor LA, Lam WL, Tsao MS. SOX15 and other SOX family members are important mediators of tumorigenesis in multiple cancer types. *Oncoscience.* (2014) 1:326–35. doi: 10.18632/oncoscience.46
52. Lopez-Pajares V, Qu K, Zhang J, Webster DE, Barajas BC, Saprashvili Z, et al. A lncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation. *Dev Cell.* (2015) 32:693–706. doi: 10.1016/j.devcel.2015.01.028
53. Hou A, Zhang Y, Zheng Y, Fan Y, Liu H, Zhou X. LncRNA terminal differentiation-induced ncRNA (TINCR) sponges miR-302 to upregulate cyclin D1 in cervical squamous cell carcinoma (CSCC). *Hum Cell.* (2019) 32:515–21. doi: 10.1007/s13577-019-00268-y
54. Hazawa M, Lin DC, Handral H, Xu L, Chen Y, Jiang YY, et al. ZNF750 is a lineage-specific tumour suppressor in squamous cell carcinoma. *Oncogene.* (2017) 36:2243–54. doi: 10.1038/onc.2016.377
55. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. *Nat Rev Genet.* (2013) 14:880–93. doi: 10.1038/nrg3594
56. Cheng G, Liu D, Liang H, Yang H, Chen K, Zhang X. A cluster of long non-coding RNAs exhibit diagnostic and prognostic values in renal cell carcinoma. *Aging.* (2019) 11:9597–615. doi: 10.18632/aging.102407
57. Ma Y, Liu Y, Pu Y, Cui M, Mao Z, Li Z, et al. LncRNA IGFL2-AS1 functions as a ceRNA in regulating ARPP19 through competitive binding to miR-802 in gastric cancer. *Mol Carcinog.* (2020) 59:311–22. doi: 10.1002/mc.23155
58. Tracy KM, Tye CE, Page NA, Fritz AJ, Stein JL, Lian JB, et al. Selective expression of long non-coding RNAs in a breast cancer cell progression model. *J Cell Physiol.* (2018) 233:1291–9. doi: 10.1002/jcp.25997
59. Liu CN, Zhang HY, Liu CL, Wang CC. Upregulation of lncRNA CALML3-AS1 promotes cell proliferation and metastasis in cervical cancer via activation of the Wnt/ β -catenin pathway. *Eur Rev Med Pharmacol Sci.* (2019) 23:5611–20. doi: 10.26355/eurrev_201907_18295
60. Gierl MS. The Zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. *Genes Dev.* (2006) 20:2465–78. doi: 10.1101/gad.381806
61. Watt AJ, Zhao R, Li J, Duncan SA. Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Dev Biol.* (2007) 7:37. doi: 10.1186/1471-213X-7-37
62. Wakana K, Akiyama Y, Aso T, Yuasa Y. Involvement of GATA-4/-5 transcription factors in ovarian carcinogenesis. *Cancer Lett.* (2006) 241:281–8. doi: 10.1016/j.canlet.2005.10.039
63. the Oesophageal Cancer Clinical Molecular Stratification (OCCAMS) Consortium, Frankell AM, Jammula S, Li X, Contino G, Killcoyne S, et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet.* (2019) 51:506–16. doi: 10.1038/s41588-018-0331-5
64. Karafin MS, Cummings CT, Fu B, Iacobuzio-Donahue CA. The developmental transcription factor Gata4 is overexpressed in pancreatic ductal adenocarcinoma. *Int J Clin Exp Pathol.* (2009) 3:47–55.
65. Zhou J, Hui X, Mao Y, Fan L. Identification of novel genes associated with a poor prognosis in pancreatic ductal adenocarcinoma via a bioinformatics analysis. *Biosci Rep.* (2019) 39:BSR20190625. doi: 10.1042/BSR20190625
66. Noll EM, Eisen C, Stenzinger A, Espinet E, Muckenhuber A, Klein C, et al. CYP3A5 mediates basal and acquired therapy resistance in different subtypes of pancreatic ductal adenocarcinoma. *Nat Med.* (2016) 22:278–87. doi: 10.1038/nm.4038
67. Zhou X, Guo X, Chen M, Xie C, Jiang J. HIF-3 α promotes metastatic phenotypes in pancreatic cancer by transcriptional regulation of the RhoC-ROCK1 signaling pathway. *Mol Cancer Res.* (2018) 16:124–34. doi: 10.1158/1541-7786.MCR-17-0256
68. Srivastava SK, Bhardwaj A, Arora S, Singh S, Azim S, Tyagi N, et al. MYB is a novel regulator of pancreatic tumour growth and metastasis. *Br J Cancer.* (2015) 113:1694–703. doi: 10.1038/bjc.2015.400
69. DeYoung MP, Tress M, Narayanan R. Down's syndrome-associated single minded 2 gene as a pancreatic cancer drug therapy target. *Cancer Lett.* (2003) 200:25–31. doi: 10.1016/S0304-3835(03)00409-9
70. Nakamura K, Komatsu M, Chiwaki F, Takeda T, Kobayashi Y, Banno K, et al. SIM2 attenuates resistance to hypoxia and tumor growth by transcriptional suppression of HIF1A in uterine cervical squamous cell carcinoma. *Sci Rep.* (2017) 7:14574. doi: 10.1038/s41598-017-15261-4
71. Tapia-Carrillo D, Tovar H, Velazquez-Caldelas TE, Hernandez-Lemus E. Master regulators of signaling pathways: an application to the analysis of gene regulation in breast cancer. *Front Genet.* (2019) 10:1180. doi: 10.3389/fgene.2019.01180
72. Boj SF, Parrizas M, Maestro MA, Ferrer J. A transcription factor regulatory circuit in differentiated pancreatic cells. *Proc Natl Acad Sci USA.* (2001) 98:14481–6. doi: 10.1073/pnas.241349398
73. Hu Y, Wu F, Liu Y, Zhao Q, Tang H. DNMT1 recruited by EZH2-mediated silencing of miR-484 contributes to the malignancy of cervical cancer cells through MMP14 and HNF1A. *Clin Epigenet.* (2019) 11:186. doi: 10.1186/s13148-019-0786-y
74. Hui H, Perfetti R. Pancreas duodenum homeobox-1 regulates pancreas development during embryogenesis and islet cell function in adulthood. *Eur J Endocrinol.* (2002) 146:129–41. doi: 10.1530/eje.0.1460129
75. Naqvi AAT, Hasan GM, Hassan MI. Investigating the role of transcription factors of pancreas development in pancreatic cancer. *Pancreatol.* (2018) 18:184–90. doi: 10.1016/j.pan.2017.12.013
76. Australian Pancreatic Cancer Genome Initiative, Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature.* (2016) 531:47–52. doi: 10.1038/nature16965
77. Jafarzadeh M, Soltani BM, Soleimani M, Hosseinkhani S. Epigenetically silenced LINC02381 functions as a tumor suppressor by regulating PI3K-Akt signaling pathway. *Biochimie.* (2020) 171–2:63–71. doi: 10.1016/j.biochi.2020.02.009
78. Tornesello ML, Faraonio R, Buonaguro L, Annunziata C, Starita N, Cerasuolo A, et al. The role of microRNAs, long non-coding RNAs, and circular RNAs in cervical cancer. *Front Oncol.* (2020) 10:150. doi: 10.3389/fonc.2020.00150
79. Militello G, Weirick T, John D, Döring C, Dimmeler S, Uchida S. Screening and validation of lncRNAs and circRNAs as miRNA sponges. *Brief Bioinform.* (2016) 18:780–88. doi: 10.1093/bib/bbw053
80. Sharma S, Mungar K. The role of long noncoding RNAs in human papillomavirus-associated pathogenesis. *Pathogens.* (2020) 9:289. doi: 10.3390/pathogens9040289
81. Shao S, Wang C, Wang S, Zhang H, Zhang Y. LncRNA STXBP5-AS1 suppressed cervical cancer progression via targeting miR-96-5p/PTEN axis. *Biomed Pharmacother.* (2019) 117:109082. doi: 10.1016/j.biopha.2019.109082

Conflict of Interest: LLV is speaker and consultant of Merck Sharp & Dohme for HPV prophylactic vaccines.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Bispo, Farias, de Araujo-Souza, Cintra, Santos, Jorge, Castro, Wajnberg, Scherer, Genta, Carvalho, Villa, Sichero and Passetti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.