Scientific Reports

https://doi.org/10.1038/s41598-026-43717-z

Article in Press

Spatial transcriptomic landscape of invasion patterns in human papillomavirus-associated endocervical adenocarcinoma

Received: 2 January 2026

Accepted: 5 March 2026

Published online: 12 March 2026

Cite this article as: Axelrod M.L., Zhou R. & Sun L. Spatial transcriptomic landscape of invasion patterns in human papillomavirus-associated endocervical adenocarcinoma. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-43717-z

Margaret L. Axelrod, Ruiwen Zhou & Lulu Sun

Spatial Transcriptomic Landscape of Invasion Patterns in Human Papillomavirus-Associated Endocervical Adenocarcinoma

Margaret L. Axelrod, MD, PhD[1], Ruiwen Zhou, PhD[2], and Lulu Sun, MD, PhD[1*]

[1]Department of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO, USA

[2]Center for Biostatistics and Data Science, Washington University in St. Louis, St. Louis, MO, USA

***Corresponding Author**: Lulu Sun, MD, PhD

Washington University School of Medicine, MSC 8118-04-03

660 South Euclid Avenue, St. Louis, MO 63110

Office: (314) 273-4908; Email: sunl@wustl.edu

Conflicts of Interest and Source of Funding: L.S. declares business relationships with Pairidex, Inc., and AstraZeneca, but these relationships are not relevant to the current work. M.L.A. and R.Z. do not have any conflicts of interest to declare. This work was supported by Washington University in St. Louis Department of Pathology and Immunology TRPA funding.

Abstract

Human papillomavirus (HPV)-associated endocervical adenocarcinoma is the second-most common cancer of the uterine cervix. HPV-associated endocervical adenocarcinoma can be classified into histologic Silva patterns of invasion, which are associated with clinical outcome. However, the mechanisms underlying these patterns of invasion are incompletely understood. We used whole transcriptome spatial transcriptomics to examine gene expression differences separately in the tumor epithelium and the surrounding stromal immune microenvironment (SIME). Seven cases were evaluated, focusing on cases with two distinct patterns of invasion within the same tumor, to control for inter-patient heterogeneity. The most strongly upregulated pathways in both higher-risk tumor epithelium and SIME were associated with extracellular matrix (ECM) remodeling. Transcriptomic-based inference of immune cell populations showed an increase in macrophage populations in higher-risk tumor areas, confirmed by immunohistochemistry. Finally, we derived a four-gene signature from genes upregulated in higher-risk tumor epithelium (*KRT6A, TNC, LAMC2* and *FN1*), which was associated with worse clinical outcome in an independent dataset (The Cancer Genome Atlas). Overall, this work demonstrates that ECM remodeling and macrophage presence are important in the progression to high-risk patterns of invasion in HPV-associated endocervical adenocarcinoma. In addition, we established a prognostic four-gene signature that is predictive of poor outcome.

Keywords: Endocervical adenocarcinoma, spatial transcriptomics, extracellular matrix, prognostic biomarkers

Introduction

The most common malignancies of the uterine cervix are squamous cell carcinoma and human papillomavirus (HPV)-associated endocervical adenocarcinoma. While HPV vaccination has decreased the overall incidence of cervical cancer, the proportion of adenocarcinoma relative to squamous cell carcinoma has increased (with approximately 23% of cervical cancers being adenocarcinoma)[1,2].

Historically, the same prognostic and staging factors have been used for squamous cell carcinoma and adenocarcinoma of the uterine cervix, despite differences in histopathology and clinical behavior. The three-tier Silva pattern system has more recently been used to classify HPV-associated endocervical adenocarcinomas according to the observed pattern of invasion and stromal reaction to the tumor, and is associated with clinical outcome[3]. Silva pattern A tumors have no destructive stromal invasion or single cell invasion, retain rounded glandular profiles, and lack lymphovascular space invasion. Silva pattern B tumors show limited destructive stromal invasion, while Silva pattern C tumors have diffuse destructive stromal invasion, and confluent or solid growth. Pattern A tumors have a better prognosis than pattern B and C tumors, which are independent risk factors for lymph node metastasis, higher stage, and worse survival. A two-tier system in which pattern A tumors and pattern B tumors without lymphovascular space invasion (LVSI) are considered low-risk, while pattern B tumors with LVSI and pattern C tumors are considered high-risk, also shows strong association with clinical outcome[3–5]. If two patterns of invasion are present in a tumor, the higher/more aggressive pattern is designated.

The biological mechanisms underpinning the distinct histologic patterns and clinical outcomes seen in differing Silva patterns are incompletely understood. Few previous studies have examined gene expression by Silva pattern in endocervical adenocarcinoma. To our knowledge, no prior study on endocervical adenocarcinoma has examined gene expression in a spatially resolved manner. As the classification into Silva patterns relies heavily on the stromal reaction to the tumor, we choose to separately profile gene expression in the invasive tumor epithelium compared to the surrounding stromal immune microenvironment (SIME) using the GeoMX spatial transcriptomics platform. Our study aimed to identify key genes and pathways which are upregulated in higher-risk Silva pattern tumors, and to use these findings to develop prognostic tools in the evaluation of endocervical adenocarcinoma.

Methods

Case Selection and Tissue Microarray

This study was approved by the Washington University in St. Louis Institutional Review Board (IRB #202402165) with a waiver of consent, as a retrospective study. This study was conducted in adherence to the ethical principles outlined in the Belmont Report and the Declaration of Helsinki. Archival formalin fixed paraffin embedded (FFPE) resection (cervical cone, loop electrosurgical excision procedure or hysterectomy) cases of HPV-associated, usual type, endocervical adenocarcinoma from December 2018 to December 2023 were reviewed. Silva pattern was assigned by the sign-out pathologist and reviewed by the study pathologist (all gynecologic pathology subspecialists). Cases with at least two distinct invasion patterns were included to allow for comparisons within cases. In total, seven archival FFPE cases, representing six

unique patients, were selected for downstream analysis. Within cases, distinct Silva patterns were assigned to regions of interest. Selected cases and regions were compiled into tissue microarrays (TMAs) with 2 mm cores. Areas with both tumor epithelium and stroma, at the invasive front of the tumor, were selected for malignant regions of the TMA. The number of high-quality cores on the TMAs (useable for both nanoString and immunohistochemistry) per case and pattern is shown in supplemental table 1.

NanoString GeoMX Digital Spatial Profiling and Sequencing

Whole RNA transcriptome spatial gene expression was investigated using the nanoString GeoMX Digital Spatial Profiler (Bruker Spatial Biology, Bothell, WA, USA), which has a stated resolution of down to 10 microns, as previously described[6,7]. Digital spatial profiling was performed at the St. Louis University Advanced Spatial Biology and Research Histology Facility (St. Louis, MO, USA). The GeoMX platform has a maximum size of regions of interest (ROIs) for analysis (660μm x 785 μm). The ROIs were selected by a pathologist (L.S.), to include a balance of both tumor epithelium and stroma and to maximize the amount of tissue analyzed. ROIs were segregated using a nuclear counterstain and anti-cytokeratin immunofluorescence probes (clone AE1/AE3, Novus NBP2-33200) to separate RNA captured from tumor epithelium (cytokeratin-positive) from RNA in the adjacent SIME (cytokeratin negative). Aspirates from 93 ROIs were collected (25 benign cervix, 21 Silva A, 22 Silva B, and 25 Silva C). Library preparation was performed according to the manufacturer's instructions. During the library preparation process, the photocleaved oligonucleotides with tag sequence identifiers, unique molecular identifiers, and sequencing primer reads were PCR

amplified. Unique dual indexes and flow cell sequences were added. Library concentration was quantified by Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Libraries were sequenced on the Illumina NovaSeq X Plus (Illumina, Inc., San Diego, CA, USA), targeting 1.5 billion reads per library.

Bioinformatics, Data Analysis and Code

Sequencing (fastq) files were initially processed through the proprietary GeoMx NGS Pipeline. The raw reads were trimmed, paired-end reads were merged, and then these reads were aligned to the corresponding identification barcodes based on their tag sequence identifiers, which correspond to genes in the transcriptome. The number of tags were counted to determine number of transcripts, corresponding to level of RNA expression. PCR duplicates were removed by sorting the unique molecular identifiers. Data was normalized using Q3 normalization. Low quality reads (low percent aligned reads less than 90%, low percent stitched reads less than 95%, Q30 scores less than 99.0%) were excluded from analysis.

Using the initially processed data, additional data analysis was done in R version 4.4.2 using RStudio version 2024.12.0. Raw un-normalized counts were filtered for low expression and low variability genes (genes with a coefficient of variation less than one were removed) and then used as the input for differential gene expression analysis in DeSeq2 (release 3.21), per the developer's requirements[8,9]. The DeSeq2 model internally corrects for library size. Because multiple regions of interest (ROIs) were sampled from the same tumor, ROIs were not strictly independent observations. In order to address this, case was included as a variable in comparisons which included multiple cases, and the two-stage Benjamini and Hochberg (BH) procedure was used

(with an alpha of 0.05) for controlling the false discovery rate (FDR) in all differential gene expression analyses[10].

The apeglm method was used for log fold change shrinkage. To compare differentially expressed genes within and across cases, significance was defined as adjusted p-value (two-stage BH method) <0.1 and log2 fold change >0.5. The two-stage BH method was implemented with the package multtest[11]. Functional enrichment analysis was done using gprofiler2 using GO, KEGG, Reactome and WikiPathways terms, with a significance threshold of p=0.05 and multiple testing correction performed per the developer's custom algorithm[12]. Gprofiler2 was used for pathways analysis[13]. Expression of individual genes was plotted using row Z-scores on the normalized data. Kaplan-Meier curve, log-rank analysis, and Cox proportional hazards modeling were done using packages survival and survminer. Group cut-offs were determined using Youden cut point analysis in package cutpointr[14]. Immune cell populations were inferred using CIBERSORTx, using the provided LM22 matrix as a reference[15]. Only the SIME compartment was included in the cibersortx analysis. Cibersortx was run in absolute mode with 1000 permutations with quantile normalization disabled. Wilcoxon tests were used to test for differences in inferred immune cell abundance with Holm correction used for multiple comparisons adjustment. All code used in data analysis is available at https://github.com/MLAxelrod/CervicalCaDSP. The datasets generated and analyzed during the current study are available in the Gene Expression Omnibus (GEO) repository, GSE316098.

Immunohistochemistry

Immunohistochemical (IHC) staining for CD68 (clone KP-1) and cytokeratin 5/6 (CK5/6; clone D5/16B4) was performed using a clinically validated CAP/CLIA protocol on the Ventana auto-staining platform by the Anatomic and Molecular Pathology Core Lab at Washington University School of Medicine. Each stain was scored independently by a blinded pathologist (L.S.) and a blinded pathologist-in-training (M.L.A.). The scores across the two reviewers were averaged. CD68 was assigned a three-tier score (0 = rare, scattered positive cells; 1 = intermediate density; 2 = diffuse). CK5/6 was scored by both intensity (0 = absent; 1 = weak; 2= strong) and percentage of epithelial cells positive (in increments of 10%). A CK5/6 composite score was assigned as percentage times intensity.

The Cancer Genome Atlas (TCGA) data

Available transcriptome and outcome data was retrieved from the TCGA for cases of cervical adenocarcinomas via cBioportal. All cases of cervical adenocarcinoma with available data were included. Data were accessed on 12-16-2024.

Results

Description of cases and regions of interest used for analysis

The study cohort consisted of seven FFPE specimens of HPV-associated, endocervical adenocarcinoma, usual type, from six unique patients. All patients were female with an age range of 28-70 years old. All patients had Federation of Gynecology and Obstetrics (FIGO) stage IB1 or IB2 disease. All patients were alive with no evidence of disease with a minimum follow up time of 2 years (Table 1). Only excision specimens

(hysterectomy, LEEP, or cone excisions) were included, to ensure adequate evaluation of morphology and lymphovascular space invasion. Emphasis was placed on cases which included multiple distinct histologic patterns of invasion within the same case (Figure 1A). These cases were assembled into a tissue microarray and profiled using the nanoString GeoMX Digital Spatial Profiler. Immunofluorescence was used to separate RNA captured from the tumor epithelium (cytokeratin positive) from RNA in the SIME (cytokeratin negative; Figure 1B), using GeoMX platform tools. The number of ROIs from each case and risk pattern area are summarized in Table 2.

Quality control (QC) metrics were met for all samples. All segments (epithelium or SIME areas for each ROI) had greater than 99.5% bases with Q≥30. The average percent aligned reads was 95.2% (range 92.6-95.6%). Principal component analysis after normalization demonstrated separation between the tumor and SIME compartments, supporting the expected segmentation (Supplemental Figure 1).

Five cases included sufficient ROIs from two distinct Silva patterns to allow for intra-case comparisons (Table 2). Within each case, gene expression was compared in the higher risk pattern compared to the lower risk pattern. This method allowed for minimization of inter-patient heterogeneity, and isolation of genes important for progression of invasive morphology and subsequent correlation with poor outcome.

Analysis of high-risk tumor epithelium and environment demonstrates upregulation of extracellular matrix genes

Comparing across cases, 630 genes were upregulated in higher risk Silva pattern tumor epithelium within the same case and shared in two or more cases

(Supplemental File 1: IntraCaseComparisons). Using functional enrichment analysis, upregulated pathways in higher-risk tumor epithelium include: “PI3K-Akt signaling pathway”, “Human papillomavirus infection”, “Focal adhesion”, “ECM-receptor interaction”, “Degradation of the extracellular matrix”, and “Extracellular matrix organization” (Figure 2A).

Six genes, *KRT6A, TNC, SERPINA3, LAMC2, FN1,* and *CCND1,* were upregulated in higher risk tumor epithelium and shared across four or five cases (Figure 2B). *KRT6A* encodes keratin 6, higher expression of which has been associated with poor outcomes in multiple solid tumor types, including lung cancer[16,17], bladder cancer[18], pancreatic cancer[19], and nasopharyngeal cancer[20]. *TNC* encodes tenascin-C, an extracellular matrix (ECM) glycoprotein. *SERPINA3* is a member of the SERPIN family of protease inhibitors. *LAMC2* encodes an ECM laminin. *FN1* encodes fibronectin, which has a variety of functions in the ECM. *CCND1* encodes cyclin D1, a key cell cycle protein. Interestingly, two of these genes, *SERPINA3* and *CCND1,* showed high expression in benign cervix (Figure 2B).

Applying the same analysis to the SIME, 333 genes were upregulated in higher-risk Silva pattern SIME within the same case and shared in two or more cases. Using functional enrichment analysis, upregulated pathways in higher-risk SIME include: “Complement system”, “cytokine activity”, “fibronectin binding”, “Phagosome”, “Degradation of the extracellular matrix”, “Innate Immune System”, and “Neutrophil degranulation” (Figure 3A). One gene, *CTHRC1,* was shared in 4 cases (Figure 3B). *CTHRC1* encodes collagen triplet helix repeat containing 1, which has been associated with solid tumor carcinogenesis and regulation of TGF-beta and WNT signaling

pathways[21]. Additional selected key genes driving upregulation of the indicated pathways are shown in Figure 3B. *CD68* is a key identity gene for macrophages. *SRFP2* encodes secreted frizzled related protein 2, a key component of the WNT signaling pathway. *MMP9* encodes matrix metalloprotease 9, which plays a key role in ECM remodeling. *POSTN* encodes periostin, a secreted ECM protein involved in ECM remodeling and epithelial to mesenchymal transition[22]. Notably, *FN1* was found to be upregulated in both the higher-risk SIME and also in higher risk tumor epithelium.

Similar findings of upregulated extracellular matrix organization pathways were identified when we analyzed differential gene expression from high-risk epithelium compared to low-risk epithelium, and high-risk stroma versus low-risk stroma, across all seven cases (Supplemental Figure 2, Supplemental File 2:AllHighvsLowRisk), as a between-case comparison instead of within-case comparison. This analysis showed fewer genes supporting each highlighted pathway, and fewer upregulated pathways with higher p-values compared to the within-case comparisons, supporting the utility of the within-case intratumoral approach. In addition, some pathways seen in the within-case analysis were not found in the overall between-case comparisons, including PI3K signaling and integrin cell surface interactions in the epithelium, and fibronectin binding and innate immune system categories in the SIME.

Transcriptomic-based inference of immune populations shows increased abundance of macrophage populations

The SIME analysis showed upregulation of *CD68*, a marker of macrophages. In order to investigate whether other immune cell populations, or specific macrophage subsets, might show differential abundance in tumors by risk group, we used

CIBERSORTx to infer immune cell abundance from transcriptomic data in the SIME compartment [15] (Supplemental File 3: CibersortStatistics). Macrophages, particularly M2 macrophages, showed a pattern of increasing abundance in higher risk tumors (Figure 4). In contrast, total B cells and total T cells did not show striking differences by risk group. The analysis suggested that M2 macrophages may be a key component of the SIME in higher risk tumors.

Immunohistochemistry for CD68 highlights increased macrophage presence in high-risk pattern areas

We next sought to determine whether differences in gene expression and transcriptomics-inferred immune populations could be translated to measurable protein level differences through IHC on the tissue microarrays (Figure 5A). CD68 is a marker of tissue macrophages, which were suggested to be increased by inferred immune cell abundance, and *CD68* was also upregulated in the SIME. Notably, CD68 was also available as a validated IHC marker in our laboratory. CD68 showed higher expression in the surrounding tissues of pattern C areas of tumors relative to less aggressive disease, consistent with increased macrophage presence (Figure 5B).

In the epithelium, CK5/6 was selected on the basis of higher keratin 6 gene *(KRT6A)* expression in higher risk tumor epithelium, and it being a widely used immunohistochemical stain; however, keratin 5 gene expression was not different across groups (data not shown). Interestingly, CK5/6 was more highly expressed in pattern A and B tumors relative to benign cervix, but there were mixed results with pattern C tumors (Figure 5C).

Derived four-gene signature is associated with worse outcomes in an independent dataset

Finally, we aimed to test association of key genes of interest with outcome in an independent data set. We considered the four genes that were shared across higher-risk tumor epithelium and with higher expression in tumor than benign cervix to create a four-gene risk score (sum of z-scores for *KRT6A, TNC, LAMC2* and *FN1;* Figure 2B). As expected, the four genes combined had higher expression in higher-risk Silva pattern tumor epithelium (Figure 6A). Using The Cancer Genome Atlas (TCGA)[23–25] entire data set of cervical adenocarcinomas with available transcriptome and outcome data (n=26), we modeled this four-gene score as a continuous predictor of outcome using a univariate Cox proportional hazards model, which showed a trend towards higher four-gene expression being associated with worse outcome (hazard ratio 2.17, 95% confidence interval 0.57-8.31, p-value = 0.258). With the practical aim of using the four-gene signature as a potential prognostic biomarker, we applied Youden cut point analysis to determine the threshold for dichotomization of into high and low score (z-score cut point = -0.3076). Applying this threshold to the same TCGA dataset, we found that high four-gene risk score was associated with worse overall survival (Figure 6B).

The majority of the TCGA cases were T1 or T2 (57.7% T1 and 23.0% T2, with an additional 3.8% T3 and 15.4% stage unknown). Interestingly, the survival curve of the TCGA cohort stratified by stage showed that T1 tumors had trended toward worse overall survival compared to T2, although this was not statistically significant (Figure 6C). Table 3 shows the distribution of stage by low or high four-gene risk score. Although the sample size is small, the presence of multiple low stage patients with high

risk scores and poor outcome may suggest that the risk score could be of additional prognostic value compared to stage alone. Including stage as a covariate in the Cox proportional hazards model showed a similar result to the univariate analysis (hazard ratio 1.5, 95% confidence interval 0.25 - 8.95, p=0.65). Silva pattern was unknown for the TCGA cohort and was not incorporated into the analysis.

Similar analyses did not show a significant association of key genes upregulated in higher risk SIME (*CD68, CTHRC1, SFRP2, MMP9, FN1,* or *POSTN*) with survival outcomes in TCGA (data not shown).

Discussion

To our knowledge, this is the first study to examine gene expression in distinct Silva patterns of endocervical adenocarcinoma in a spatially resolved manner while controlling for inter-patient variation. Pathways associated with ECM remodeling were upregulated in both higher-risk tumor epithelium and SIME. This finding is concordant with the histologic features of enhanced desmoplastic stroma in higher-risk tumors. Upregulation of ECM pathways and genes, such as *FN1, TNC, LAMC2, MMP9*, and *POSTN,* may be key drivers of worse outcomes for higher-risk Silva pattern tumors. ECM dysregulation in solid tumors has been associated with tumor aggressiveness, angiogenesis, immune-dysregulation and impaired drug delivery of systemic therapies. *POSTN* showed particularly striking upregulation in the SIME of pattern C tumors. *POSTN* is a key marker expressed by immunosuppressive cancer associated fibroblast

in several cancer types and has been associated with epithelial to mesenchymal transition[22,26,27].

The strong association of ECM pathways and higher risk disease is also intriguing as many facets of ECM biology are possible drug targets [28,29]. Efforts to target the ECM have included drugs which regulate MMPs, collagen synthesis, hyaluronic acid synthesis, and TGF-beta signaling, among others. Other strategies have included using fibronectin as a target for more specific drug delivery. Future work should investigate the potential use of ECM-targeting therapies in high-risk Silva pattern endocervical adenocarcinoma.

Additionally, PI3K-Akt signaling was upregulated in higher-risk tumor epithelium. PI3K inhibition is also an intriguing drug target, with FDA-approved agents in breast cancer and some hematologic malignancies. Combinatorial strategies using PI3K inhibitors with immune-targeting agents are actively being investigated[30].

In the SIME, upregulation of innate immune system pathways was associated with higher risk disease. This was validated by increased macrophages highlighted by CD68 IHC staining in higher Silva pattern disease. This finding may be causally linked to upregulation of ECM-related pathways, as ECM-immune interactions can recruit immunosuppressive myeloid cells including inducing polarization of macrophages toward a more pro-tumor “M2” phenotype[28,29]. Indeed, we found via gene expression-based imputation of immune populations that M2 macrophages were at estimated higher abundance in high-risk pattern SIME. Notably, *FN1,* which encodes fibronectin 1, was upregulated in both the high-risk tumor epithelium and SIME. Overexpression of fibronectin 1 has been shown to induce macrophage M2 polarization and tumor

progression in other solid tumors[31]. These data suggest that interactions between the tumor epithelium and the SIME may promote polarization toward a pro-tumor phenotype, and may play a role in progression of endocervical adenocarcinoma.

We aimed to develop useful prognostic indicators by translating some of the gene expression differences that we discovered into IHC biomarkers. In particular, we chose CD68 and CK5/6 for validation efforts as they are clinically validated in our laboratory, commonly used in surgical pathology, and widely available. Our aim was to select markers which could be easily adapted into clinical practice. We found higher CD68 density in higher-risk Silva pattern tumors. Interestingly, CK5/6 showed more ambivalent results. This antibody recognizes both keratin 5 and keratin 6, although only the gene encoding keratin 6 showed differential expression in our dataset. CK5/6 is more commonly used as a marker of squamous differentiation, although expression, especially of cytokeratin 6, can be seen in adenocarcinomas[31,32]. Our data show higher expression of CK5/6 in pattern A and B tumors compared to benign cervix, but mixed expression in pattern C tumors. Further exploration in a larger dataset is needed. Alternatively, IHC for only CK6 could be evaluated.

Finally, we developed a four-gene risk score, derived from genes upregulated in higher-risk tumor epithelium (*KRT6A, TNC, LAMC2* and *FN1*), which was associated with worse overall survival in the TCGA dataset. The four-gene risk score was better able to stratify overall survival compared to stage alone in this limited cohort, which showed a trend of worse survival in the stage T1 group compared to stage T2. Gene expression profiling is commonly used in other solid tumor types to guide therapeutic management decisions and discussions of prognosis. For example, the multi-gene

expression assay OncotypeDX is widely used in breast cancer and impacts treatment strategies for patients[33].

This study was limited by a small cohort size (n=7 cases from six unique patients), all of whom had similar FIGO stage disease. Our study was also based primarily on gene expression, with additional IHC validation. Future work could include analysis of additional molecule types, including glycosaminoglycans of the ECM. In addition, it is possible that since this study focused on cases with mixed morphology, low-risk areas could harbor abnormalities and phenotypic properties that promote high-risk behavior. Future studies that specifically look at gene expression in cases with only low-risk or high-risk patterns could identify additional genes and pathways for risk stratification. In addition, spatial transcriptomic platforms may have segmentation errors with overlap between cell boundaries, although this is more problematic for single-cell and subcellular resolution technologies as opposed to the region of interest approach with the GeoMX platform used here[34,35]. This known technical limitation could result in mis-assignment of transcripts to tumor versus SIME categories. The development of more accurate segmentation and data analysis methods may be able to address these issues. Finally, as there were limited external TCGA data for cervical adenocarcinomas, we were not able to use separate datasets for cut point definition and evaluation, or to stratify survival analysis by stage, Silva pattern, or other histopathologic features. Using the same dataset for cut point definition and evaluation is likely to overestimate the prognostic value in independent datasets. The utility of a gene expression-based prognostic biomarker in endocervical adenocarcinoma would need to be validated in further, large scale studies.

In summary, our study used spatially resolved whole transcriptome gene expression profiling to identify ECM remodeling as a key factor upregulated in higher risk tumor epithelium and SIME in HPV-associated endocervical adenocarcinoma. Furthermore, innate immune pathways and CD68 expression by immunohistochemistry were associated with higher risk SIME. We also identified a four-gene risk score which was associated with worse overall survival in an independent dataset. Overall, these data suggest that therapeutic strategies targeting the ECM and innate immune system warrant further investigation in HPV-associated endocervical adenocarcinoma. Use of prognostic gene-expression based biomarkers may be useful in stratifying patients and further informing personalized treatment decisions.

List of abbreviations: Human papillomavirus (HPV), stromal immune microenvironment (SIME), extracellular matrix (ECM), lymphovascular space invasion (LVSI), formalin fixed paraffin embedded (FFPE), tissue microarrays (TMAs), regions of interest (ROIs), false discovery rate (FDR), immunohistochemical (IHC), The Cancer Genome Atlas (TCGA), Federation of Gynecology and Obstetrics (FIGO), loop electrosurgical excision procedure (LEEP), Adenocarcinoma in situ (Ais), no evidence of disease (NED), quality control (QC)

References

1. Smith HO, Tiffany MF, Qualls CR, Key CR. The Rising Incidence of Adenocarcinoma Relative to Squamous Cell Carcinoma of the Uterine Cervix in the United States—A 24-Year Population-Based Study. *Gynecol Oncol*. 2000;78(2):97-105. doi:10.1006/GYNO.2000.5826

2. Smith AJB, Beavis AL, Rositch AF, Levinson K. Disparities in Diagnosis and Treatment of Cervical Adenocarcinoma compared to Squamous Cell Carcinoma: An Analysis of the National Cancer Database, 2004–2017. *J Low Genit Tract Dis*. 2022;27(1):29. doi:10.1097/LGT.0000000000000702

3. De Vivar AD, Roma AA, Park KJ, Alvarado-Cabrero I, Rasty G, Chanona-Vilchis JG, et al. Invasive endocervical adenocarcinoma: Proposal for a new pattern-based classification system with significant clinical implications: A multi-institutional study. *Int J Gynecol Pathol*. 2013;32(6):592-601. doi:10.1097/PGP.0B013E31829952C6

4. Powell A, Hodgson A, Cohen PA, Rabban TJ, Park KJ, McCluggage WG, et al. Improved Risk Prediction in Human Papillomavirus–Associated Endocervical Adenocarcinoma Through Assessment of Binary Silva Pattern-based Classification: An International Multicenter Retrospective Observational Study Led by the International Society of Gynecological Pathologists (ISGyP). *Int J Gynecol Pathol*. 2024;43(5):436-446. doi:10.1097/PGP.0000000000001033

5. Alvarado-Cabrero I, Parra-Herran C, Stolnicu S, Roma A, Oliva E, Malpica A. The Silva Pattern-based Classification for HPV-associated Invasive Endocervical Adenocarcinoma and the Distinction Between In Situ and Invasive Adenocarcinoma: Relevant Issues and Recommendations From the International Society of Gynecological Pathologists. *Int J Gynecol Pathol*. 2021;40(suppl 1):S48-S65. doi:10.1097/PGP.0000000000000735

6. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol*. 2020;38(5):586-599. doi:10.1038/s41587-020-0472-9

7. Zimmerman SM, Fropf R, Kulasekara BR, Griswold M, Appelbe O, Bahrami A, et al. Spatially resolved whole transcriptome profiling in human and mouse tissue using Digital Spatial Profiling. *Genome Res*. 2022;32(10):1892-1905. doi:10.1101/GR.276206.121/-/DC1

8. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-21. doi:10.1186/S13059-014-0550-8/FIGURES/9

9. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*. 2019;35(12):2084-2092. https://dx.doi.org/10.1093/bioinformatics/bty895. Accessed March 30, 2021.

10. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that

control the false discovery rate. *Biometrika*. 2006;93(3):491-507. doi:10.1093/biomet/93.3.491

11. Pollard KS, Dudoit S, van der Laan MJ. Multiple Testing Procedures: the multtest Package and Applications to Genomics. 2005:249-271. doi:10.1007/0-387-29362-0_15

12. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

13. Peterson H, Kolberg L, Raudvere U, Kuzmin I, Vilo J. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research 2020 9709*. 2020;9:709. doi:10.12688/f1000research.24956.2

14. Thiele C, Hirschfeld G. cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R. *J Stat Softw*. 2021;98:1-27. doi:10.18637/JSS.V098.I11

15. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773-782. doi:10.1038/s41587-019-0114-2

16. Che D, Wang M, Sun J, Li B, Xu T, Lu Y, et al. KRT6A Promotes Lung Cancer Cell Growth and Invasion Through MYC-Regulated Pentose Phosphate Pathway. *Front Cell Dev Biol*. 2021;9:694071. doi:10.3389/FCELL.2021.694071/BIBTEX

17. Sui P, Liu X, Zhong C, Sha Z. Integrated single-cell and bulk RNA-Seq analysis enhances prognostic accuracy of PD-1/PD-L1 immunotherapy response in lung adenocarcinoma through necroptotic anoikis gene signatures. *Sci Reports 2024 141*. 2024;14(1):1-17. doi:10.1038/s41598-024-61629-8

18. Chen Y, Ji S, Ying J, Sun Y, Liu J, Yin G. KRT6A expedites bladder cancer progression, regulated by miR-31-5p. *Cell Cycle*. 2022;21(14):1479. doi:10.1080/15384101.2022.2054095

19. Shinomiya Y, Kouchi Y, Harada-Kagitani S, Ishige T, Takano S, Ohtsuka M, et al. ECM1 and KRT6A are involved in tumor progression and chemoresistance in the effect of dexamethasone on pancreatic cancer. *Cancer Sci*. 2024;115(6):1948-1963. doi:10.1111/CAS.16175

20. Chen C, Shan H. Keratin 6A gene silencing suppresses cell invasion and metastasis of nasopharyngeal carcinoma via the β-catenin cascade. *Mol Med Rep*. 2019;49(5):3477-3484. doi:10.3892/mmr.2019.10055

21. Liu YJ, Du J, Li J, Tan XP, Zhang Q. CTHRC1, a novel gene with multiple functions in physiology, disease and solid tumors (Review). *Oncol Lett*. 2023;25(6):266. doi:10.3892/OL.2023.13852

22. Dorafshan S, Razmi M, Safaei S, Gentilin E, Madjd Z, Ghods R. Periostin: biology and function in cancer. *Cancer Cell Int 2022 221*. 2022;22(1):1-28. doi:10.1186/S12935-022-02714-8

23. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. 2012;2(5):401-404. doi:10.1158/2159-8290.CD-12-0095

24. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. 2013;6(269):1-1. /doi/pdf/10.1126/scisignal.2004088?download=true. Accessed May 10, 2022.

25. de Bruijn I, Kundra R, Mastrogiacomo B, Tran TN, Sikina L, Mazor T, et al. Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res*. 2023;83(23):3861-3867. doi:10.1158/0008-5472.CAN-23-0816/728880/AM/ANALYSIS-AND-VISUALIZATION-OF-LONGITUDINAL-GENOMIC

26. Chen C, Guo Q, Liu Y, Hou Q, Liao M, Guo Y, et al. Single-cell and spatial transcriptomics reveal POSTN + cancer-associated fibroblasts correlated with immune suppression and tumour progression in non-small cell lung cancer. *Clin Transl Med*. 2023;13(12):e1515. doi:10.1002/CTM2.1515

27. Wang H, Liang Y, Liu Z, Zhang R, Chao J, Wang M, et al. POSTN+ cancer-associated fibroblasts determine the efficacy of immunotherapy in hepatocellular carcinoma. *J Immunother Cancer*. 2024;12(7):8721. doi:10.1136/JITC-2023-008721

28. Prakash J, Shaked Y. The Interplay between Extracellular Matrix Remodeling and Cancer Therapeutics. *Cancer Discov*. 2024;14(8):1375-1388. doi:10.1158/2159-8290.CD-24-0002

29. Huang J, Zhang L, Wan D, Zhou L, Zheng S, Lin S, et al. Extracellular matrix and its therapeutic potential for cancer treatment. *Signal Transduct Target Ther 2021 61*. 2021;6(1):1-24. doi:10.1038/s41392-021-00544-0

30. Vanhaesebroeck B, Perry MWD, Brown JR, André F, Okkenhaug K. PI3K inhibitors are finally coming of age. *Nat Rev Drug Discov 2021 2010*. 2021;20(10):741-769. doi:10.1038/s41573-021-00209-1

31. Völkel C, De Wispelaere N, Weidemann S, Gorbokon N, Lennartz M, Luebke AM, et al. Cytokeratin 5 and cytokeratin 6 expressions are unconnected in normal and cancerous tissues and have separate diagnostic implications. *Virchows Arch*. 2022;480(2):433-447. doi:10.1007/S00428-021-03204-4/TABLES/2

32. Chu PG, Weiss LM. Expression of Cytokeratin 5/6 in Epithelial Neoplasms: An Immunohistochemical Study of 509 Cases. *Mod Pathol*. 2002;15(1):6-10. doi:10.1038/MODPATHOL.3880483

33. Schaafsma E, Zhang B, Schaafsma M, Tong CY, Zhang L, Cheng C. Impact of Oncotype DX testing on ER+ breast cancer treatment and survival in the first decade of use. *Breast Cancer Res*. 2021;23(1):1-11. doi:https://doi.org/10.1186/s13058-021-01453-4

34. Mitchel J, Gao T, Petukhov V, Cole E, Kharchenko P V. Impact and correction of segmentation errors in spatial transcriptomics. *Nat Genet*. 2026;58(2). doi:10.1038/s41588-025-02497-4

35. Jones DC, Elz AE, Hadadianpour A, Ryu H, Glass DR, Newell EW. Cell simulation as cell segmentation. *Nat Methods 2025 226*. 2025;22(6):1331-1342. doi:10.1038/s41592-025-02697-0

Acknowledgements: We thank the St. Louis University Advanced Spatial Biology and Research Histology Facility (Caroline Murphy and Michelle Brennan, PhD) and the Anatomic and Molecular Pathology Core Lab, Washington University School of Medicine. We also thank the Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH.

Author Contributions: L.S. and M.L.A. designed the study, collected and analyzed the data, and wrote and reviewed the manuscript. R.Z. provided biostatistics consultation. All authors read and approved the final paper.

Data Availability Statement: All code used in data analysis is available at https://github.com/MLAxelrod/CervicalCaDSP . The datasets generated and analyzed during the current study are available in the Gene Expression Omnibus (GEO) repository, GSE316098.

Competing Interests: L.S. declares business relationships with Pairidex, Inc., and AstraZeneca, but these relationships are not relevant to the current work. M.L.A. and R.Z. do not have any conflicts of interest to declare.

Funding: This work was supported by Washington University in St. Louis Department of Pathology and Immunology TRPA funding.

Figure Legends

Figure 1. Digital Spatial Profiling on distinct patterns of invasion within the same case. A) Hematoxylin and eosin-stained sections showing distinct patterns of invasion (left = A; right = B/C) within the same case. B) Immunofluorescence for cytokeratin (green) and DAPI (blue) showing region of interest selection and segmentation into tumor epithelium (cytokeratin positive) and stromal immune microenvironment (cytokeratin negative).

Figure 2. Extracellular matrix remodeling is upregulated in higher risk tumor epithelium. A) Pathway analysis showing upregulated pathways in the 630 genes that are upregulated in higher risk compared to lower risk tumor epithelium within the same case and shared by at least two cases. Dot size represents the number of genes intersecting between the genes of interest and the indicated pathway. Color indicates the term source. GO:MF = Gene Ontology: Molecular Function; KEGG = Kyoto Encyclopedia of Genes and Genomes; REAC = Reactome Pathway Database. B) Expression (Z-score) of key genes shown across all regions of interest, by Silva pattern. Box plots show the median, first and third quartiles. P values represent Holm-corrected Wilcoxon tests. Color indicates case.

Figure 3. Innate immune pathways and extracellular matrix remodeling are upregulated in higher risk SIME. A) Pathway analysis showing upregulated pathways in the 333 genes that are upregulated in higher risk compared to lower risk SIME within the same case and shared by at least two cases. Dot size represents the number of genes intersecting between the genes of interest and the indicated pathway. Color indicates the term source. GO:MF = Gene Ontology: Molecular Function; KEGG =

Kyoto Encyclopedia of Genes and Genomes; REAC = Reactome Pathway Database. B) Expression (Z-score) of key genes shown across all regions of interest, by Silva pattern. Box plots show the median, first and third quartiles. P values represent Holm-corrected Wilcoxon tests. Color indicates case.

Figure 4. Transcriptomic-inferred immune population analysis identifies higher abundance of M2 macrophages in higher risk tumor SIME. CIBERSORTx inferred immune cell abundance is shown for key immune cell populations. Box plots show the median, first and third quartiles. P values represent Holm-corrected Wilcoxon tests. Color indicates case.

Figure 5. Immunohistochemistry shows higher CD68 expression in higher risk Silva patterns. A) Example immunohistochemical staining for CD68 and CK5/6 with representative scoring. B-C) Quantification of immunohistochemical scoring (averaged across two reviewers) for CD68 (scored by abundance) and CK5/6 (scored by percent of tumor cells positive times intensity). Box plots show the median, first and third quartiles. P values represent Holm-corrected Wilcoxon tests. Color indicates case.

Figure 6. Four gene risk score predicts survival in an independent dataset. A) Four gene risk score (sum of Z-scores of KRT6A, TNC, LAMC2 and FN1) across all regions of interest, by Silva pattern. Box plots show the median, first and third quartiles. P values represent Holm-corrected Wilcoxon tests. Color indicates case. B) Kaplan-Meier plot of overall survival in TCGA dataset of 26 endocervical adenocarcinomas, stratified by expression of the four-gene risk score. C) Kaplan-Meier plot of overall survival in TCGA dataset, stratified by AJCC stage.

Supplemental Figure 1. Principal component analysis of spatial transcriptomic tumor epithelium and SIME segments. Color indicates compartment (epithelium or SIME) and shape indicates Silva pattern risk group.

Supplemental Figure 2. Extracellular matrix remodeling is upregulated in higher risk tumor epithelium and SIME, across all cases. Pathway analysis showing upregulated pathways in higher risk tumor epithelium (top) and SIME (bottom) when including all cases. Dot size represents the number of genes intersecting between the genes of interest and the indicated pathway. Color indicates the term source. GO:MF = Gene Ontology: Molecular Function; KEGG = Kyoto Encyclopedia of Genes and Genomes; REAC = Reactome Pathway Database.

Tables

ID	Sex	Age	Specimen type	Silva Patterns	LVSI	FIGO Stage	Treatment	Status	Follow up time (years)
WU-01	F	28	Cone	B, A	Negative	IB1	Hysterectomy	Alive, NED	3.1
WU-02 *same patient as WU-01	F	28	Hysterectomy	B, A	Negative	IB1	Hysterectomy	Alive, NED	3.1
WU-03	F	28	Cone	C, A	Suspicious	IB2	Hysterectomy	Alive, NED	3.1
WU-04	F	34	Cone	B, A	Positive	IB1	Hysterectomy	Alive, NED	5.2
WU-05	F	70	Hysterectomy	C, B	Positive	IB2	Hysterectomy, radiation, carboplatin	Alive, NED	2.1
WU-06	F	33	LEEP	C, Ais	Negative	IB1	LEEP	Alive, NED	2.1
WU-07	F	32	LEEP	B, A, Ais	Positive	IB1	Hysterectomy, radiation	Alive, NED	2.8

Table 1: Clinicopathologic characteristics. FIGO = Federation of Gynecology and Obstetrics; LEEP = loop electrosurgical excision procedure; Ais = Adenocarcinoma in situ; LVSI = Lymphovascular space invasion; NED = No evidence of disease.

	WU-01	*WU-02*	*WU-03*	*WU-04*	*WU-05*	WU-06	WU-07
Benign	*8*	*0*	*4*	*4*	*2*	4	3
A (low risk)	*5*	*3*	*8*	*5*	*0*	0	0
B (low risk)	*4*	*1*	*0*	*0*	*0*	0	0
B (high risk)	*0*	*0*	*0*	*6*	*7*	0	4
C (high risk)	*0*	*0*	*10*	*0*	*8*	7	0

Table 2. Number of regions of interest per case and risk pattern area used in nanoString GeoMX. Italicized cases were included in intra-case risk pattern analyses.

Stage	Low 4-gene risk score	High 4-gene risk score
T1	4	11

T2	4	2
T3	0	1
Unknown	0	4

Table 3. Number of patients with each AJCC stage cervical adenocarcinoma by four-gene risk score group in the TCGA dataset.

Figure 1

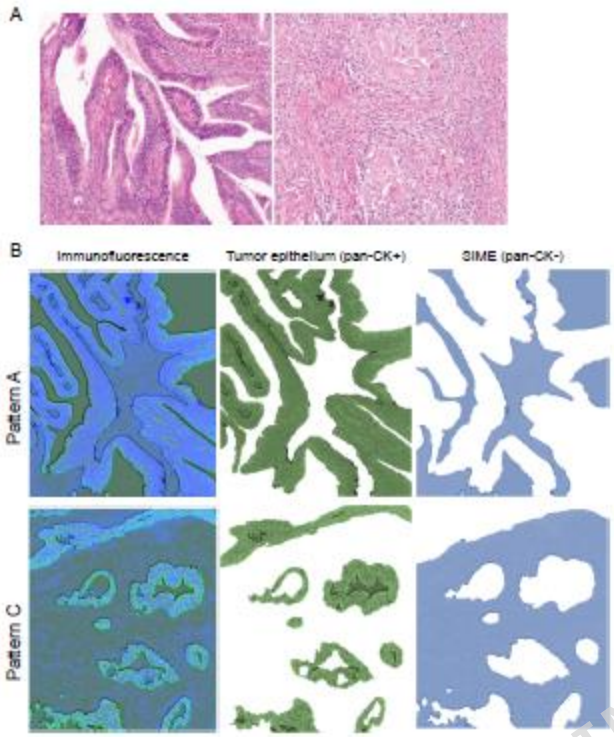

Figure 2

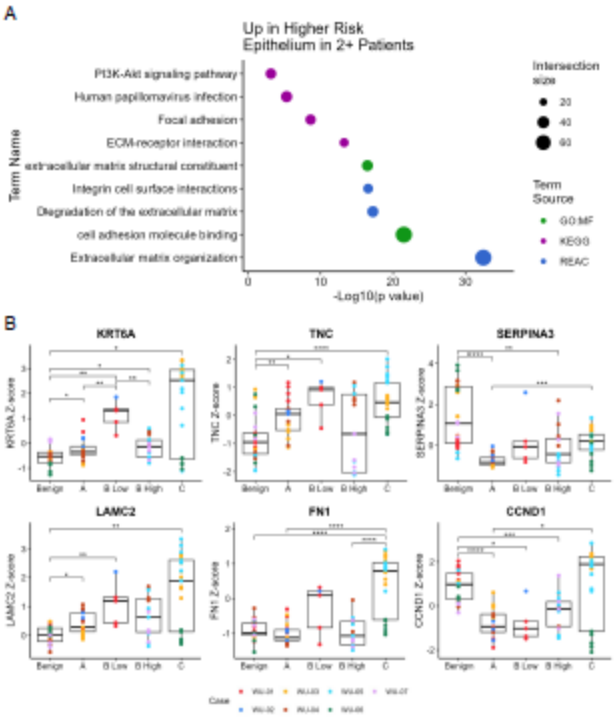

Figure 3

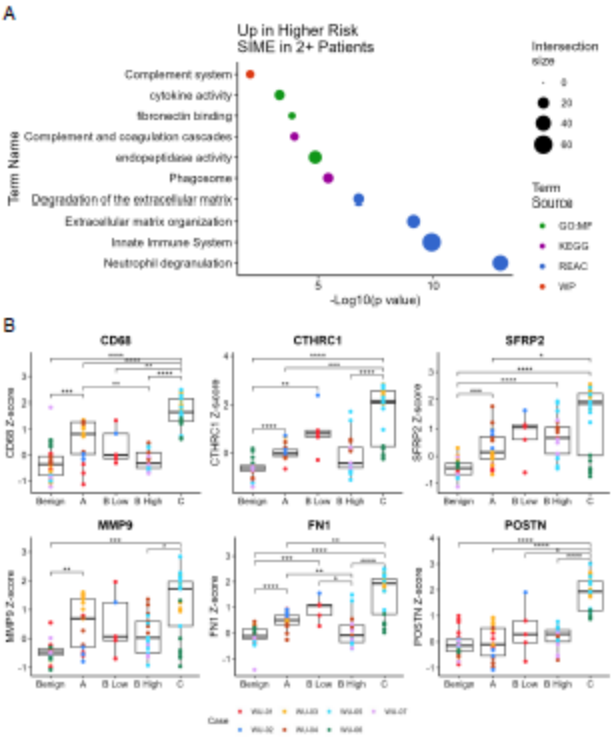

Figure 4

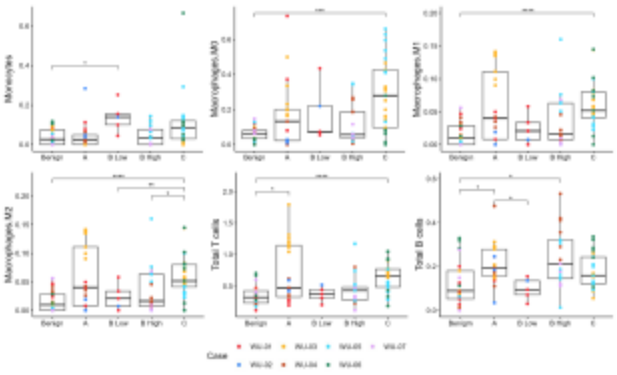

Figure 5

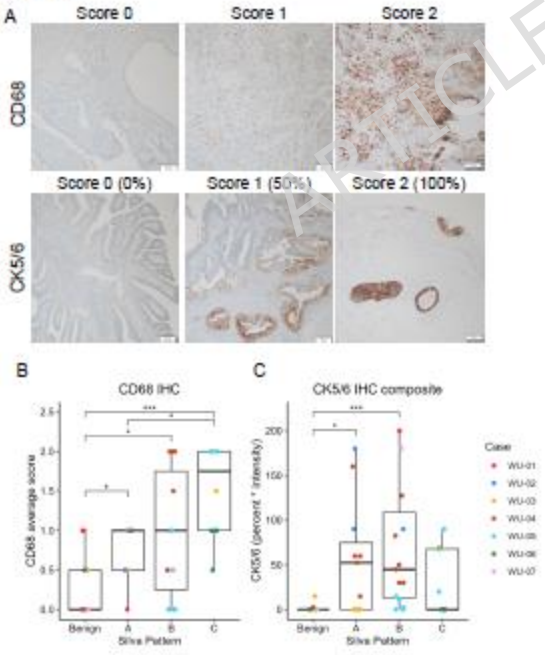

Figure 6