JCI insight

Transcriptome and microbiome-immune changes across preinvasive and invasive anal cancer lesions

Ezequiel Lacunza, … , Juan C. Ramos, Martin C. Abba

JCI Insight. 2024. https://doi.org/10.1172/jci.insight.180907.

Research In-Press Preview

Anal squamous cell carcinoma (ASCC) is a rare gastrointestinal malignancy linked to high-risk Human papillomavirus (HPV) infection, which develops from precursor lesions like Low-Grade Squamous Intraepithelial Lesions (LGSIL) and High-Grade Squamous Intraepithelial Lesions (HGSIL). ASCC incidence varies across populations, posing increased risk for People Living with HIV (PLWH). Our investigation focused on transcriptomic and metatranscriptomic changes from Squamous Intraepithelial Lesions (SILs) to ASCC. Metatranscriptomic analysis highlighted specific bacterial species (e.g., Fusobacterium nucleatum, Bacteroides fragilis) more prevalent in ASCC than precancerous lesions. These species correlated with gene encoding enzymes (Acca, glyQ, eno, pgk, por) and oncoproteins (FadA, dnaK), presenting potential diagnostic or treatment markers. Unsupervised transcriptome analysis identified distinct sample clusters reflecting histological diagnosis, immune infiltrate, HIV/HPV status, and pathway activities, recapitulating anal cancer progression's natural history. Our study unveiled molecular mechanisms in anal cancer progression, aiding in stratifying HGSIL cases based on low- or high-risk progression to malignancy.

Transcriptome and microbiome-immune changes across preinvasive and invasive anal cancer lesions.

Ezequiel Lacunza [1,8,*], Valeria Fink [2,8], María E. Salas [1,8], Ana M. Gun [2,8], Jorge A. Basiletti [3], María A. Picconi [3], Mariano Golubicki [4], Juan Robbio [4], Mirta Kujaruk [4], Soledad Iseas [5], Sion Williams [6,8], María I. Figueroa [2,8], Omar Coso [7,8], Pedro Cahn [2,8], Juan C. Ramos [6,8], Martín C. Abba [1,8,*].

(1) Centro de Investigaciones Inmunológicas Básicas y Aplicadas (CINIBA), Facultad de Ciencias Médicas, Universidad Nacional de La Plata, La Plata, Argentina.
(2) Dirección de Investigaciones, Fundación Huésped, Buenos Aires, Argentina.
(3) Laboratorio Nacional y Regional de Referencia de Virus Papiloma Humano. Instituto Nacional de Enfermedades Infecciosas - ANLIS "Dr. Malbrán", Buenos Aires, Argentina.
(4) Unidad de Oncología, Hospital de Gastroenterología “Dr. Carlos Bonorino Udaondo”, Buenos Aires, Argentina.
(5) Medical Oncology Department, Paris-St Joseph Hospital, Paris, France
(6) University of Miami - Center for AIDS Research (UM-CFAR) / Sylvester Comprehensive Cancer Center (SCCC), University of Miami Miller School of Medicine, Miami, FL, USA.
(7) Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE-CONICET), Universidad de Buenos Aires, Buenos Aires, Argentina.
(8) University of Miami - Centre for AIDS Research/Sylvester Cancer Comprehensive Center Argentina Consortium for Research and Training in Virally Induced AIDS-Malignancies, University of Miami Miller School of Medicine, Miami, Florida, USA.
*Corresponding authors: Correspondence to E Lacunza (ez.lacunza@gmail.com) or MC Abba (mcabba@gmail.com); phone number +54 221 423 6711 (342); mailing address: Facultad de Ciencias Médicas Universidad Nacional de La Plata Calle 60 y Av. 120 B1900 La Plata, Buenos Aires, Argentina.

Abstract

Anal squamous cell carcinoma (ASCC) is a rare gastrointestinal malignancy linked to high-risk Human papillomavirus (HPV) infection, which develops from precursor lesions like Low-Grade Squamous Intraepithelial Lesions (LGSIL) and High-Grade Squamous Intraepithelial Lesions (HGSIL). ASCC incidence varies across populations, posing increased risk for People Living with HIV (PLWH). Our investigation focused on transcriptomic and metatranscriptomic changes from Squamous Intraepithelial Lesions (SILs) to ASCC. Metatranscriptomic analysis highlighted specific bacterial species (e.g., *Fusobacterium nucleatum*, *Bacteroides fragilis*) more prevalent in ASCC than precancerous lesions. These species correlated with gene encoding enzymes (Acca, glyQ, eno, pgk, por) and oncoproteins (FadA, dnaK), presenting potential diagnostic or treatment markers. Unsupervised transcriptome analysis identified distinct sample clusters reflecting histological diagnosis, immune infiltrate, HIV/HPV status, and pathway activities, recapitulating anal cancer progression's natural history. Our study unveiled molecular mechanisms in anal cancer progression, aiding in stratifying HGSIL cases based on low- or high-risk progression to malignancy.

Keywords: ASCC, Microbiome, Transcriptome, Metatranscriptome.

Introduction

Anal Squamous Cell Carcinoma (ASCC) is a rare gastrointestinal neoplasia that involves the formation of malignant tumors in the anal region. Over the past thirty years, the incidence of ASCC has been on the rise globally, particularly in mem who have sex with men (MSM) and people living with HIV (PLWH) (1).

Squamous Intraepithelial Lesions (SILs), categorized into Low-Grade (LGSIL), analogous to anal intraepithelial neoplasia I, and High-Grade (HGSIL), analogous to anal intraepithelial neoplasia II and III, often precede the progression to ASCC (2, 3) Similar to cervical cancer, ASCC development is driven by the infection with oncogenic human papillomaviruses (HPV) (4, 5). The risk of anal cancer varies significantly across different population groups, with the highest risk observed in PLWH (1). This increased susceptibility is primarily attributed to a weakened immune system, which makes it more challenging to control infections, including HPV infections (6). Beyond the potential impact of oncogenic viruses, the microbiome may also play a significant role in the development of precancerous anal lesions and ASCC, as the influence of microbes is increasingly recognized in cancer development (7,8). The microbiome can influence the balance of host cell proliferation and apoptosis, disrupt anti-tumoral immunity, and affect the metabolism of host-produced factors, ingested food components, and drugs. (9). In a recent study, we defined the microbiome composition of the anal mucosa of HIV-exposed individuals. Metagenomic sequencing enabled us to identify viral and bacterial taxa linked to the development of anal lesions. Our results confirmed the occurrence of oncogenic viromes in this population and identified *Prevotella bivia* and *Fusobacterium gonidiaformans* as two relevant bacterial species predisposing to SILs. Moreover, gene family analysis identified bacterial gene signatures associated with SILs that may have potential as prognostic and predictive biomarkers for HIV-associated malignancies (10). Other reports using 16S rRNA gene sequencing to analyze the ASCC demonstrated the role of the anal microbiota in anal cancer response to therapy and toxicity, as well as changes in taxonomic compositions among normal, dysplasia, and anal cancer samples (11,12).

The molecular biology of ASCC is complex and not completely understood (13). However, several studies have identified potential molecular targets for ASCC therapy, including regulators of apoptosis (14), agents targeting the PI3K/AKT pathway (15), antibody therapy targeting EGFR (16) or PD-L1 expression to stratify good versus poor responders to chemoradiotherapy (17). Despite advancements in understanding ASCC from various perspectives, thus far, no prognostic or predictive markers have been identified that are useful in clinical practice. Furthermore, a notable gap in existing information is the paucity of

studies employing anal cancer biopsies for gene expression profiling, particularly utilizing advanced techniques like next-generation sequencing (NGS).

Transcriptomics and metatranscriptomics profiling are powerful NGS-based tools for the functional genomics characterization of complex diseases. In this sense, bulk RNA sequencing (RNA-seq) in neoplastic disease enables the simultaneous study of the host tumor transcriptome and its microenvironment, including the tumor immune infiltrate and the associated tumor microbiome. Transcriptomic profiling provides a thorough examination of gene expression patterns, uncovering crucial insights into the molecular mechanisms driving cancer development and progression. Metatranscriptomic profiling enables researchers to analyze gene expression levels of various organisms within a microbial community, providing insights into their metabolic processes and functional activities in cancer and immune-related diseases (18). In this sense, metatranscriptomics approaches enable the analysis of the active microbiota instead of more frequent studies based on 16S rRNA sequencing, which analyzes the "total" microbiota, including active and inactive bacteria.

The aim of this study was to analyze the transcriptomic and metatranscriptomics changes that occur during the progression from LGSIL to HGSIL and ultimately to ASCC.

We collected biopsies identified as SILs and ASCC from a cohort of 70 participants, encompassing individuals both with and without HIV, all of whom provided informed consent. Biopsies were subjected to bulk RNA-seq. Our goal was to gain insights into the molecular mechanisms underlying the development and progression of anal lesions, which could potentially lead to the identification of novel biomarkers and therapeutic targets for improved diagnostic and treatment strategies in patients with ASCC.

Results

Clinical characteristics of patients and microbial community variations in SILs and ASCC cases.

Seventy patients were included in the present study. All participants underwent anal cytology and high-resolution anoscopy with biopsies. Based on cytology and histology analysis, samples were classified into LGSIL, n=23, HGSIL, n=16, and ASCC, n=23. Demographic and clinical data were collected, including age, sex at birth (male or female), gender (cisgender men (CGM), transgender women (TGW), and cisgender women (CGW)), HPV DNA status, HIV status, and antiretroviral therapy (ART). This information is summarized in Table 1.

122 We first conducted a compositional analysis of the three distinct groups—
123 LGSIL, HGSIL, and ASCC— by performing permutational multivariate ANOVA
124 (PERMANOVA) with Euclidean distance. The Principal Coordinate Analysis
125 (PCoA) defined two distinct clusters based on component I ($p<0.001$). Cluster
126 I was enriched in LGSILs, comprising 24 out of 31 samples (77%), while Cluster
127 II predominantly featured ASCC samples with 19 out of 23 (83%) (Fig. 1A).
128 HGSIL demonstrated an almost equal distribution between the two clusters,
129 with 9 out of 16 in Cluster I (56%) and 7 out of 16 in Cluster II (44%) (Fig. 1A).
130 In addition, we considered covariates such as age, gender, HIV status, and
131 high-risk HPV DNA genotyping (HR-HPV) to evaluate the factors influencing
132 cluster formation based on diagnostic groups. Employing PERMANOVA, our
133 analysis of beta diversity revealed distinctions primarily in samples positive for
134 HR-HPV types compared to samples in which these HPV types were
135 undetected (Figure 1B; Supp Figure 1).

136 The ASCC microbial community, assessed through Observed and Chao 1
137 indices based on metatranscriptome species composition, exhibited a
138 significantly higher richness compared to LGSIL (Observed, $p = 0.033$; Chao 1,
139 $p = 0.035$) and HGSIL (Observed, $p = 0.029$; Chao 1, $p = 0.034$). This trend
140 persisted when merging LGSIL and HGSIL into the group termed SILs
141 (Observed, $p = 0.012$; Chao 1, $p = 0.018$), suggesting that the ASCC
142 environment may provide a more favorable habitat for a specific range of
143 microorganisms, resulting in increased community richness (Fig. 1B;
144 Supplementary Data 1). Richness indices were also augmented in the HR-HPV
145 group compared to the negative cohort for HR-HPV types. In addition, a
146 significant association between HIV-positive status and decreased alpha
147 diversity was observed, in agreement with previous studies (10) (Fig. 1B).

148 Analysis of diversity indices (Shannon and Simpson) revealed a significant
149 increase in ASCC compared to HGSIL (Shannon, $p = 0.0082$; Simpson, $p =$
150 0.0134), while no differences were observed between LGSIL and ASCC
151 (Supplementary Data 1). These findings align with a recent study that reported
152 similar alpha diversity indices between anal dysplasia and anal cancer but
153 highlighted an elevated abundance of specific taxa in the latter (12).
154 Consistent with our prior research we further observed a negative influence
155 of aging on microbiome diversity (10) (Supplementary Data 1).

156 We analyzed bacterial abundance at the phylum level between SILs and ASCC
157 groups. Fusobacteriota, Bacteroidota, and Bacillota, among the most
158 abundant phyla, were significantly more enriched in ASCC compared to SILs
159 (Fig. 1D). Additionally, Pseudomonadota showed enrichment within the ASCC
160 group compared to precancerous lesions (Fig. 1D).

At the species level, we identified a total of 25 taxa, each exhibiting a relative abundance exceeding 20% of the overall composition in at least one of the samples (Fig. 1E). Among these taxa, *Fusobacterium nucleatum*, *Fusobacterium necrophorum*, *Bacteroides fragilis*, and *Prevotella intermedia* are well-established gut-associated bacteria with previous associations with colorectal cancer (CRC) (19). Conversely, other taxa such as *Mycoplasma hominis*, *Prevotella bivia*, *Fusobacterium gonidiaformans*, *Sneathia amnii*, *Campylobacter ureoliticus* or *Bacteroides fragilis* have been linked to HPV-related precancerous and cancerous genital lesions (10, 12, 20, 21, 22)

To identify bacterial species associated with ASCC compared to SILs, we used MaAsLin2 analysis. To account for potential confounders, we refined the model by incorporating additional covariates, including HIV status, HR-HPV DNA status, sex at birth, and age. Significant enrichment was observed for *Fusobacterium nucleatum* (p= 0.001), *Fusobacterium gonidiaformans* (p = 0.001), *Bacteroides fragilis* (p = 0.01), *Campylobacter ureolyticus* (p = 0.003), and *Criibacterium bergeronii* (p = 0.006) (Fig. 1E; Supplementary Data 2). Moreover, C. ureolyticus (p=0.002), *F. gonidiaformans* (p=0.01), and *C. bergeronii* (p=0.02) were associated with male sex (Supplementary Data 2). Additionally, *C. ureolyticus* correlated with HIV-negative cases (p=0.03) (Supplementary Data 2).

F. nucleatum and *B. fragilis* have established roles in CRC progression, highlighting their importance in ASCC development and progression (19). While knowledge about *F. gonidiaformans*, *C. ureolyticus*, and *C. bergeronii* is limited, prior associations exist between *F. gonidiaformans* and *C. ureolyticus* with HPV presence and the development of precancerous lesions in anal and cervical cancers (10, 21, 23). These findings suggest a potential contribution of specific bacteria to ASCC progression.

Exploring Viral Signatures in Anal Lesions Progression: Alpha Papillomavirus and Non-HPV Species.

In terms of viral composition analysis, among the 40 species identified at the transcript level in all samples, eight were the most prevalent, with abundances greater than 30% of the total abundance in any sample and detected more than three times. Notably, seven of these species belonged to the *Alpha Papillomavirus* (Alpha-PV) genus, along with the *Human endogenous retrovirus K* (HERV-K), with evident variations in their relative abundances across distinct diagnostic groups (Fig. 2A). MaAsLin2 analysis revealed a higher abundance of Alpha-PV-10, which includes low-risk genotypes like HPV6 and HPV11, in both LGSIL and HGSIL compared to ASCC. (Fig. 2A, B, Supplementary Data 2). Conversely, Alpha-PV-9 (HPV16, 31, 33, 52, 58) and Alpha-PV-7 (HPV18, 39, 59, 68, 45, 70) were significantly associated with HGSIL and ASCC

201 (Fig. 2A, B, Supplementary Data 2). This trend persisted when considering the
202 number of positive cases for these species independent of their relative
203 abundance (Fig. 2C). Although the significance was not established for Alpha-
204 PV-10, it remained significant for Alpha-PV-7 and Alpha-PV-9 (Fig. 2C).

205 The HPV DNA genotyping data highlighted a robust association between
206 HPV16 and both HGSIL and ASCC, correlating with the pattern observed with
207 Alpha-PV-9 (Fig. 2C). However, HPV18 was detected in only one case of ASCC,
208 contrasting with Alpha-PV-7 detected at the RNA level in over 20% of
209 participants (Fig. 2A). This discrepancy could be due to Alpha-PV-7 containing
210 other HPV genotypes (24). HPV6 and HPV11 were predominantly linked to
211 LGSIL (Fig. 2D). Analyzing positive and negative cases for all low-risk (LR) HPV
212 types and high-risk (HR) HPV types identified within the cohort revealed
213 negative ($p<0.001$) and positive associations ($p<0.05$), respectively, with the
214 diagnostic groups (Fig. 2E, F). These results confirm the prominence of HR and
215 LR HPV types, particularly HPV6 and HPV16, in delineating the diagnostic
216 groups (25).

217 Among the non-HPV species, it is noteworthy to highlight a significant increase
218 in the relative abundance of the endogenous HERV-K in ASCC compared with
219 HGSIL ($p<0.01$; Supplementary Data 2). HERV-K overexpression is widely
220 associated with malignant phenotypes and is upregulated in various cancers
221 such as breast lymphoma, germ-line tumors, and melanoma (26). Additionally,
222 Human betaherpesvirus 5 (HCMV), although with low relative abundance,
223 demonstrated significant enrichment in ASCC compared with SILs ($p<0.05$;
224 Supplementary Data 2). HCMV is linked to several cancer types, including
225 lymphoma, cervical cancer, Kaposi's sarcoma, CRC, prostate cancer, skin
226 cancer, and glioblastomas (27). However, it remains unclear whether HCMV
227 actively contributes to malignant tumor progression or is reactivated under
228 conditions leading to chronic inflammation or immunosuppression (27).

229 Overall, these findings confirm the significance of specific viral Alpha
230 Papillomavirus species and their association with SILs toward ASCC
231 progression. Furthermore, our data reveals a potential involvement of HERV-
232 K and HCMV in ASCC tumorigenesis. Additionally, the use of
233 metatranscriptomics demonstrates remarkable reliability, sensitivity, and
234 specificity in detecting the presence of HPV types, even in cases where DNA
235 genotyping results were negative.

236 Metabolic Pathways in ASCC Progression

237 To understand the functional implications of microbial community changes
238 between SILs and ASCC, we conducted metatranscriptomics analysis, revealing
239 20 MetaCyc modules as significantly enriched pathways in ASCC compared to
240 SILs (Table 2). These modules encompassed Nucleotide, Amino Acid, and Lipid

241 Biosynthesis pathways. This finding aligns with our prior observations, where
242 pathways related to amino acid and de novo nucleotide biosynthesis were
243 enriched in HIV individuals with anal precancerous lesions (10). These
244 pathways are vital for cell growth and proliferation, as cells require energy and
245 nutrients from their environment to support these processes. Similarly, cancer
246 cells exhibit metabolic adaptations essential for their growth (28). Hence, our
247 data suggest that certain bacteria within the evolving microenvironment
248 during malignancy may exploit these pathways to thrive and proliferate, like
249 cancer cells.

250 Microbial Contributions to Anal Lesions: Enriched Proteins and Taxonomic 251 Associations.

252 To go further, we next explored the gene proteins contributed by the microbial
253 organisms in the comparison of SILs versus ASCC. MaAsLin2 analysis yielded a
254 total of 2523 UniRef90 sequence proteins differentially expressed
255 (Supplementary Data 3). We further employed the KEGG database to annotate
256 387 proteins of which 349 were significantly enriched in ASCC and 37 in SILs
257 (Supplementary Data 3). Functional annotation using KEGG Mapper revealed
258 metabolic pathways such as glycolysis, lipid, amino acid, and nucleotide
259 biosynthesis, contributed by 60 bacterial proteins enriched in ASCC (Fig. 3A).
260 Proteins like Acca (*acetyl-CoA carboxylase carboxyl transferase subunit alpha*),
261 glyA (glycine hydroxymethyltransferase), glyQ (*glycyl-tRNA synthetase alpha*
262 *chain*), eno (enolase), pgk (*phosphoglycerate kinase)* and por (pyruvate-
263 ferredoxin/flavodoxin oxidoreductase), previously identified in anal samples
264 from individuals with precancerous anal lesions (10), underline their potential
265 roles as metabolic markers in anal cancer progression. In addition, among
266 these 60 proteins, we identified the enrichment of the oncogenic FadA
267 adhesion protein from *F. nucleatum* in ASCC, a factor widely associated with
268 CRC; and dnaK, a protein kinase with a known involvement in carcinogenesis
269 and cancer progression (29,30). These findings align with the taxonomic
270 abundance analysis, highlighting the significant role of bacteria like B. fragilis,
271 *F. nucleatum*, and *C. ureolyticus*, alongside other relevant and distinct gut
272 microbiota taxa, in orchestrating these processes (Fig. 3A). Furthermore, four
273 proteins linked to the oncogene E6 from Human Papillomavirus 16 were
274 enriched in ASCC (Fig. 3B). E6 oncoprotein promotes p53 degradation,
275 contributing to keratinocyte immortalization. In SILs, 37 enriched proteins
276 were detected, all predominantly associated with genes from the LR Human
277 Papillomavirus genomes HPV6 and HPV11, underscoring their potential role as
278 drivers or sustainers of precancerous anal lesions (31) (Fig. 3B).

Transcriptomic profiling and functional insights across anal lesion progression.

We then explored the host transcriptome of LGSIL, HGSIL, and ASCC. Like the metatranscriptomes analysis, the unsupervised clustering of samples revealed two primary clusters (Fig. 4A). Cluster I is predominantly composed of LGSILs, with the inclusion of some HGSILs. In contrast, Cluster II comprises most anal cancer samples, alongside a subgroup of SILs. One plausible interpretation for this distribution is that precancerous lesions may be at varying stages of progression, with some nearing malignant transformation and others in a regressive or early stage (32).

Next, we applied supervised comparative analysis between LGSIL and HGSIL as well as HGSIL and ASCC. The analysis revealed a higher number of differentially expressed genes (DEG; FC>2, FDR < 0.05) in the transition from HGSIL to ASCC (544 DEG) than in the comparison among the two SIL groups (121 DEG) (Fig. 4B, C; Supplementary Data 4). Among the most significant genes, a decrease in keratins in HGSIL compared to LGSIL stands out (Fig. 4B) as well as the overexpression of members of the MAGE gene family of cancer/testis antigens in ASCC compared with HGSIL, like *MAGEA4*, *MAGEA3*, and *MAGEA1* (Fig. 4C). The MAGE family has gained attention as a potential cancer biomarker and immunotherapy (33). Notably, a phase I trial for autologous T-cell therapy targeting *MAGEA4*-positive solid cancers is currently underway (34).

To comprehend the functional significance of DEG, we employed Gene Set Enrichment Analysis (GSEA) on Gene Ontology (GO), Cancer Hallmarks, and Disease Ontology (DO) terms. GSEA revealed activated processes such as nuclear division, chromatin modification, and cell proliferation, along with suppressed pathways like keratinocyte differentiation and leukocyte-mediated immunity in HGSIL compared to LGSIL (Fig. 4D; Supplementary Data 5). These processes align with the histopathological features of HGSIL, including a higher nuclear-to-cytoplasmic ratio, decreased organization of cell layers, a greater degree of nuclear pleomorphism, and increased mitotic index (35). Furthermore, analysis of Cancer Hallmarks indicated the activation of pathway terms associated with sustaining proliferative signaling, such as MYC targets, E2F targets, G2M checkpoint, or mitotic spindle (Fig. 4E). Notably, there was a decrease in genes related to IFN-alpha and IFN-gamma levels, potentially compromising the ability of the immune system to mount an effective defense against viral infections and favoring persistent infection and progression to HGSIL (36) (Fig. 4E). The activation of DNA repair genes may be a response to potential damage caused by viral oncoproteins E6 or E7, which

317 aim to integrate the host genome through DNA double breakpoints (35) (Fig.
318 4E).

319 The network representation resulting from GSEA with GO comparing LGSIL to
320 HGSIL provided valuable insights into the molecular landscape
321 (Supplementary Fig. 2A). Three distinct clusters emerged, each revealing
322 specific functional themes: a DNA and chromosome organization cluster,
323 characterized by a dense interconnection of genes primarily related to
324 histones and chromatin modifiers, suggesting a potential role in the epigenetic
325 regulation and structural integrity of the genome; a chromosome segregation
326 cluster with genes predominantly linked to processes such as the mitotic
327 spindle and cell division; and a skin development cluster, offering insights into
328 the gene network governing epidermal differentiation (Supplementary Fig.
329 2A). These findings suggest a complex interplay of molecular events involving
330 DNA organization, chromosome segregation, and skin differentiation in the
331 transition from LGSIL to HGSIL. Some of these events may be attributed to HPV
332 E6 oncoprotein. The expression of viral E6 enhances cell cycle progression and
333 induces mitotic defects leading to centrosome amplification observed in
334 keratinocytes, contributing to chromosomal instability through aberrant
335 chromosome segregation (37).

336 Moreover, Disease Ontology (DO) revealed additional clusters of genes related
337 to gut inflammatory processes, HIV disease, and B cell immunodeficiency
338 (Supplementary Fig. 2B). Together, these data unveil the impact on the anal
339 transcriptome caused during the transition from LGSIL to HGSIL, defining
340 distinct driver processes, including several genes that can be new avenues for
341 further research.

342 Conversely, in comparing HGSIL and ASCC, GO analysis revealed a
343 predominant activation of immune response in ASCC but a decrease in
344 epidermal differentiation-related genes (Fig. 4F). Hallmarks analysis
345 demonstrated activation of IFN pathways emphasizing immune activation.
346 Remarkably, suppression of the p53 pathway may be linked to the
347 overexpression of HPV16 E6 protein (Fig. 4G). The network representation of
348 GO revealed clusters of genes mainly representing immune activation,
349 leukocyte migration, cytokine and immunoglobulin production but also
350 epidermal cell differentiation (Supplementary Fig. 2C). Additionally, DO yield
351 terms related to inflammatory processes of colon, HIV, and skin disease
352 (Supplementary Fig. 2D).

353 Therefore, unlike the comparison between LGSIL and HGSIL, the data suggest
354 that the transition from HGSIL to ASCC is characterized by a predominance of
355 immune response activation over processes related to cell proliferation or
356 DNA modifications (38).

Host transcriptome reveals two intrinsic signatures with varied features and prognoses.

GSEA highlighted deregulated processes across anal lesion stages, emphasizing central roles for the cell cycle, immune response, viral infection, and epidermal differentiation. We focused on significant gene signatures obtained by GSEA related to these processes to visualize gene expression patterns including epidermal differentiation (30 genes – Fig. 5A, Supplementary Data 6), immune response (72 genes – Fig. 5B, Supplementary Data 6), and cell cycle (86 genes – Fig. 5C, Supplementary Data 6) Heatmaps revealed at least two subtypes within each diagnosis group, one with high expression of the gene signature and the other with low expression. To categorize samples, we introduced "high" and "low" scores based on the average expression of each gene signature, divided by the median value (Fig. 5 A-C).

Next, we incorporated these signatures along with LR and HR HPV and HIV status into the unsupervised clustering of samples. This allowed us to discern two primary clusters with distinct characteristics (Fig. 6A). Cluster I primarily comprised SILs ($p< 0.01$; 24 out of 26 in Cluster I) with a low immune signature ($p<0.001$), high epidermal differentiation ($p<0.001$), a low cell cycle signature ($p<0.05$), and a smaller number of samples infected with HR HPV types detected at both RNA ($p<0.05$) and DNA ($p<0.05$) levels compared to Cluster II. In contrast, Cluster II encompasses 91% of anal cancer cases ($p< 0.01$; 21 out of 23 ASCC) and 62 % of HGSIL (10 out of 16 HSGIL) It exhibits a higher immune signature score ($p<0.001$), low epidermal differentiation ($p<0.001$), a greater number of samples with a high cell cycle signature ($p<0.05$), and a higher prevalence of HR HPV infections ($p<0.05$; Fig. 6A). Of note, Cluster II included most of the subjects without HIV (92%; 11 out of 12 HIV-negative cases) compared with Cluster I ($p<0.05$) which was mainly integrated with PLWH (25 out of 26 cases in Cluster I).

Immune infiltration and cell composition analysis.

We utilized EPIC and ESTIMATE algorithms for predicting immune infiltration and cell fraction composition (Fig. 6B). Cluster II exhibited a higher level of immune infiltration, as determined by EPIC ($p<0.001$). The analysis of cell composition revealed a significant increase in B cells ($p<0.001$), CD4 T cells ($p<0.001$), CD8 T cells ($p<0.05$), and macrophages ($p<0.001$), aligning with the high immune signature assigned to this cluster (Supplementary Data 7). A possible explanation for these findings could be the higher prevalence of HIV-negative cases in Cluster II, suggesting a potentially less compromised immune system compared to individuals in Cluster I.

To explore this further, we conducted a comparison of the immune profile between HIV-positive and HIV-negative individuals, irrespective of their cluster assignment. Results revealed a significant reduction in B cells ($p<0.01$) and CD4+ T cells ($p<0.001$) among PLWH in our cohort (Supplementary Data 7). This aligns with the asymptomatic phase of HIV infection, characterized by ongoing viral replication leading to a gradual depletion of CD4+ T cells, which can be partially restored with ART. While the impact of HIV on B-cell numbers is less clear, studies indicate a reduction in B-cell counts in HIV-infected individuals (39). Dysregulation of B cells during HIV infection is also influenced by ART therapy. Of note, a significant portion of individuals in our HIV-infected cohort were on ART during recruitment, contributing to observed variations in B cell composition.

Furthermore, we explored whether there was an association between these immune profiling differences and HPV16 infection. Results indicated a significantly higher immune profile of macrophages in HPV16-infected cases ($p<0.01$; Supplementary Data 7). Previous studies have reported that M2-like macrophages infiltrate HPV16-associated tumors, suppressing antitumor T-cell response and facilitating tumor growth (40).

Overall, Cluster II is represented by ASCC tumors and precancerous lesions with a high immune infiltration. The significance of tumor-infiltrating lymphocytes (TILs) in influencing favorable outcomes across various tumor types, including ASCC, has been reported in the literature (41, 42, 43, 44). Our recent study demonstrated the crucial role of PD-L1 expression in influencing complete response rates and survival outcomes in non-metastatic ASCC patients undergoing standard definitive chemoradiotherapy (17). Motivated by the importance of immune factors in ASCC, we employed the T cell dysfunction and exclusion score (TIDE) in our current study to predict cancer immunotherapy response.

The results yielded a compelling connection between immune-related characteristics and treatment response. Cluster II, characterized by a higher immune signature and immune cell infiltration, exhibited a significantly higher number of responders ($p<0.05$; Fig. 6C). The TIDE analysis highlighted specific immune cell changes associated with responders, including an increase in CD4+ TILs ($p<0.05$) and macrophages ($p<0.05$), and a concurrent decrease in cancer-associated fibroblasts (CAFs, $p < 0.01$) and endothelial cells ($p<0.01$) (Supplementary Data 7). These findings underscore the potential predictive value of immune-related parameters in discerning responders and non-responders to cancer immunotherapy in the context of anal cancer progression.

435 Furthermore, we compared the gene expression profiles of two surrogate
436 markers for HPV-related malignancy, Ki67 and p16. Results showed that both
437 markers were higher in Cluster II (Fig. 6D). Additionally, Cluster I was linked to
438 younger subjects and MSM, while Cluster II was associated with older patients,
439 enriched in TGW and cis-gender women (Fig. 6E). In coincidence with the
440 latter, high p16 expression has been shown to correlate with the female sex
441 and with better outcomes following chemo-radiotherapy (45,46,47).

442 These findings might help to better understand the molecular landscape
443 within and between different stages of anal lesions and reveal potential
444 biomarkers and therapeutic pathways for further research.

445 Immune profiling of p16, CD3 / CD8 cells and PD-L1 expression among 446 ASCC

447 The immunohistochemical (IHC) analysis of p16, CD3, CD8, and PD-L1 in anal
448 cancer not only provides valuable insights into the tumor microenvironment,
449 but also serves as a guide for treatment decisions and aids in predicting patient
450 outcomes (17).

451 In our study, we explored these markers in 10 (for p16) and 14 (for CD3, CD8
452 and PD-L1) out of the 23 ASCC samples using IHC. Ninety percent of ASCC (9
453 out of 10) showed a diffusely positive pattern of p16 (Fig. 7A). The density of
454 CD3 and CD8 TILs was moderate to high in 47% (6 out of 14) of ASCC samples
455 (Fig. 7B). Of note, all these samples exhibited a high immune signature,
456 correlating with increased immune infiltration as assessed by EPIC (Fig. 7C). In
457 this context, tumors with moderate to high CD3 and CD8 expression were
458 associated with lower tumor purity scores ($p<0.01$) and higher cell fractions of
459 cancer-associated fibroblasts (CAFs) ($p<0.05$), macrophages ($p<0.05$), and CD4
460 T cells ($p<0.05$) as revealed by EPIC analysis (Fig. 7D). The PD-L1 expression
461 status was assessed in the 14 ASCC cases using the Combined Positive Score
462 (CPS). Notably, 57% of positive cases (8 out of 14) exhibited moderate to high
463 PD-L1 expression levels (CPS > 5%), while the remaining samples showed low
464 PD-L1 expression levels (CPS < 5%;6 out of 14) (Fig. 7B). This analysis indicates
465 a complex relationship between TILs and tumor microenvironment factors,
466 shaping the immune profile of ASCC tumors and potentially influencing
467 treatment approaches.

468 Comparative transcriptome analysis of HPV-related squamous cell 469 carcinomas.

470 We analyzed relevant HPV-associated cancer studies to compare the gene
471 expression signatures identified in ASCC with head and neck squamous cell
472 carcinomas (HNSCC) and cervical squamous cell carcinomas (CSCC) cases. In a

473 previous study, Zhang et al. conducted RNA-seq on 36 HNSCC (18 HPV+ and 18
474 HPV-), identifying two HPV+ subtypes. One subtype was enriched in "immune
475 response" related genes, while the other was enriched in "keratinocyte
476 differentiation" related genes (48), which is consistent with our ASCC findings.
477 We applied the gene signature distinguishing these subtypes in HNSCC across
478 our sample cohort, sorted by immune score (Figure 8A; Supplementary Data
479 8). Additionally, we employed our gene signature, derived from the most
480 significantly deregulated genes in the HGSIL vs. ASCC comparison, on HNSCC
481 samples, grouped by the subtypes defined by the authors (Figure 8B;
482 Supplementary Data 8). Results indicate similar gene expression patterns
483 between locations, with variations in gene composition, yet lined with similar
484 biological processes. For CSCC, we utilized den Boon et al.'s study, despite
485 being microarray-based, due to its comprehensive analysis of premalignant
486 (CIN1, CIN2, and CIN3) and CSCC specimens (49). Like our approach, we
487 established a gene signature by comparing CIN2/CIN3 (comparable to HGSIL)
488 versus CSCC and visualized the gene expression profile in our sample cohort
489 (Figure 8C; Supplementary Data 8). This analysis and the application of our
490 signature to cervical lesion samples, sorted by immune score (Figure 8D;
491 Supplementary Data 8), showed an almost mutually exclusive relationship
492 between immune and epidermal differentiation processes. This suggests a
493 significant decrease in keratinocyte differentiation as the disease progresses,
494 alongside a significant increase in immune response genes.

495 Mutational profiling of cancer driver genes among ASCC and other 496 squamous cell carcinomas

497 We conducted mutational profiling on ASCC biopsies from 23 patients based
498 on RNA-seq data, revealing 51 somatic missense mutations in cancer driver
499 genes among 87% of ASCC cases (20 out of 23). We identified mutations in
500 *KMT2C* (also known as *MLL3*, 30%), *PIK3CA* (20%), *EP300* (20%), *NOTCH1*
501 (15%), *IDH1* (15%), *PRDM1* (15%), *FGFR2* (15%), *SETD2* (15%), *FGFR3* (10%),
502 *MAP3K1* (10%), and *MET* (10%). Single cases of mutations were found affecting
503 *TP53, TET2, ATM, TSC1, EZH2, CASP8, ARID1B, APC, NCOR1, SF3B1, STK11,*
504 *BRCA1, KDM6A*, and *STAG2* (Fig. 9A). Several of these mutated genes are
505 commonly found in HPV-driven squamous cancers like cervix, head and neck,
506 vulva, and anus, including *KMT2C, EP300, PIK3CA, NOTCH1, FGFR2, ATM,*
507 *TP53*, and *BRCA1* (50, 51, 17).

508 Consistent with our results, comparable frequencies of *KMT2C, PIK3CA* and
509 the chromatin remodeler *EP300*, have been reported at the genomic level

510 through NGS or targeted sequencing among the most mutated genes in ASCC
511 (51,52,53,17).

512 Our data revealed *KMT2C* mutations at comparable rates in the early stages of
513 anal lesions, reaching 30% in HGSIL and 42% in LGSIL (Fig. 9A), suggesting a
514 potential pivotal role for *KMT2C* as a driver gene in anal carcinogenesis
515 progression. Additionally, increased mutation frequencies for *EP300* (21% in
516 ASCC, 4% in HGSIL, and 13% in LGSIL) and *PI3KCA* (17% in ASCC, 8% in HGSIL,
517 4% in LGSIL) were observed compared to earlier stages of anal lesions (Fig. 9A),
518 indicating potential shifts in the molecular landscape during disease
519 progression.

520 A higher mutation rate of 3.5 (21 mutations in 6 samples) was observed in
521 ASCC with a low immune signature compared to the high immune signature
522 group ($p < 0.01$), which had a mutation rate of 1.76 (30 mutations in 17
523 samples). This implies distinct tumor subpopulations with mutations in cancer
524 driver genes (Fig. 9A).

525 Furthermore, all mutations in *KMT2C* (7 mutations in 6 cases), *PRDM1* (3
526 mutations in 3 cases) and *FGFR2* (3 mutations in 2 cases) occurred in HPV16-
527 infected cases, comprising 25% of total mutations (Fig. 9A). *PRDM1* is a master
528 regulator of lymphoid cell differentiation and a tumor suppressor gene in
529 lymphoma (54). It has been identified as a master regulator for HPV16 E6/E7
530 proteins (55) Aberrant *FGFR* signaling and HPV16 E5 expression have been
531 shown to be correlated with cervical cancer progression (56). Furthermore,
532 the interaction between HPV16 E5 and *FGFR2* alters keratinocyte
533 differentiation and inhibits tumor-suppressive genes, suggesting a role in the
534 early stages of HPV infection and transformation (56).

535 Consistent with our findings, previous studies have recognized *KMT2C* and
536 *EP300* as the most frequently mutated genes in metastatic ASCC (51). KMT2C
537 mutations are associated with abnormal H3K4 methylation, linked to
538 oncogenic transformation in preclinical models (57). K*MT2C* plays a crucial role
539 in activating *TP53* gene expression, demonstrated by targeted inactivation
540 studies in mice (58).

541 Regarding *EP300*, the oncoprotein HPV/E6 mediates *TP53* degradation by
542 binding to the histone acetyltransferase *EP300*, inhibiting *EP300*-mediated
543 *TP53* acetylation, and promoting *TP53* degradation (59,60). Consequently,
544 dysregulated histone/chromatin modulation within the context of impaired
545 DNA repair mechanisms emerges as a driver of malignancy. We categorized
546 mutated genes into cancer hallmarks and observed that Genome Instability
547 predominated (Supplementary Data 9). Genes like *KMT2C*, *EP300, IDH1,*

548 *SETD2, TET2, BRCA1, TP53, APC, ATM, KDM6A, NCOR1, SF3B1, and STAG2*
549 defined a gene network critical for ASCC, regardless of HPV infection, aligning
550 with *TP53* association with HPV-HR negativity in our study, consistent with
551 prior research (17,51,52).

552 To perform a comparative analysis of the mutational profile identified in ASCC
553 with other squamous cell carcinomas, we analyzed two combined cervical
554 cancer datasets (MSK-CESC and TCGA-CESC) and a head and neck cancer
555 dataset (TCGA-HNSC) retrieved from cBioPortal online resource
556 (http://www.cbioportal.org/). Only drivers and putative drivers' somatic
557 missense or truncating mutations were considered for frequency estimations
558 among cohorts. The comparative analysis showed that one third of the most
559 frequents cancer driver mutations identified in ASCC (8 out of 25 genes) were
560 also frequently mutated (>5% of cases) in CSCC and HNSCC (*KMT2C, EP300,*
561 *PIK3CA, NOTCH1, TP53, CASP8, STK11* and *KDM6A*) (Fig. 9B).

562 Our mutational profiling of ASCC biopsies from 23 patients offered valuable
563 insights into the somatic mutation landscape of cancer driver genes,
564 particularly given their derivation from transcriptomic data. However, we
565 recognize the significance of the limited sample size when drawing definitive
566 conclusions.

567 Discussion

568 ASCC represents only 2% of all gastrointestinal tumors but is characterized by
569 high morbidity and mortality. Unfortunately, treatment options for ASCC have
570 not evolved in the past 20 years; concurrent chemoradiotherapy continues to
571 be the standard care strategy for non-metastatic cases. For patients with
572 metastasis at diagnosis or those who develop metastatic recurrences after
573 chemoradiation therapy, the 5-year survival rate is below 20% (61). To date,
574 platinum-based chemotherapy doublets are the most commonly used
575 anticancer drugs for palliative chemotherapy, and no targeted agents have
576 been approved. In clinical practice, prognostic factors of survival in ASCC are
577 the T and N stage, sex, differentiation, tumor location, high-risk HPV infection,
578 and occurrence of a complete response after CRT (17). These clinical
579 parameters related to survival cannot be used to personalize therapy or
580 predict treatment response in individual patients. Less is known regarding
581 early-stage prognostic biomarkers of ASCC.

582 Comprehensive characterization of anal squamous precancerous and
583 cancerous lesions at metatranscriptome and transcriptome levels allowed us
584 to identify the most relevant changes that occur at the cell host and their
585 associated microenvironment – the immune infiltrate and the microbiome –

586 during the progression from preinvasive to the invasive stages. Unsupervised
587 analyses allowed us to identify two patient clusters (Cluster I and Cluster II)
588 based on their histological diagnosis, microbial composition, cell cycle,
589 immune infiltrate, immune response, viral infection (HIV and HPV), epidermal
590 differentiation and activity of specific metabolic and signaling pathways.
591 Cluster I was mainly composed by LGSIL and HGSIL differentiated and low
592 proliferative cases with low immune infiltrate and almost infected by low-risk
593 HPV types. Meanwhile Cluster II was significantly enriched in ASCC and HGSIL
594 cases with higher immune signature score, low epidermal differentiation, a
595 greater number of samples with a high cell cycle signature, and a higher
596 prevalence of high-risk HPV. In this sense, Cluster II was associated with higher
597 expression of Ki67 and p16, older patients, TGW, and females. These findings
598 align with previous studies that have implicated specific viral infections,
599 immune responses, and molecular pathways in the progression of anal lesions
600 (4,10,17). The observed distinctions between Cluster I and Cluster II provide
601 valuable insights into potential prognostic and therapeutic considerations in
602 the management of anal squamous lesions (62).

603 Microbiome changes in preinvasive and invasive stages of anal cancer

604 A comparison of the microbiota composition at phylum and species levels
605 reveals expected differences between SILs and ASCC regarding the prevalence
606 of HR HPV subtypes but also identifies several viruses and bacteria species
607 significantly associated with anal cancer not previously reported. In this sense,
608 *Fusobacterium nucleatum*, *Fusobacterium gonidiaformans* and *Bacteroides*
609 *fragilis,* previously associated with CRC progression at early stages (17), were
610 significantly enriched in ASCC compared with premalignant lesions. More
611 importantly, these taxa together with HPV16 contributed with gene encoding
612 enzymes (e.g.: Acca, glyQ, eno, pgk and por) and oncoproteins (FadA and dnaK)
613 and a distinctive ASCC metabolic profile characterized by the enrichment of
614 pathways related to oxidative, energetics or biosynthetic processes, including
615 glycolysis, lipid, amino acid, and nucleotide biosynthesis that could facilitate
616 and promote the survival and proliferation of cancerous cells (10, 29, 30).
617 Among these enzymes and proteins, Acca, glyA, glyQ, eno, pgk, and por were
618 identified in our previous study as associated with precancerous anal lesions,
619 highlighting their roles as metabolic markers in cancer progression (10). In line
620 with our results, Serrano and Villar also found pgk and eno overexpressed in
621 the microbiome of HGSIL subjects, while they propose succinyl-CoA and
622 cobalamin as markers associated with HGSIL (8). This reinforces the idea that
623 HPV-infected cells can modify metabolism by regulating genes involved in
624 cellular growth and metabolism, which is crucial to oncogenesis (63).
625 Considering and validating these microbial proteins as markers could offer
626 alternative tools in cancer prevention.

Cell signaling pathways affected in preinvasive and invasive stages of anal cancer

Our integrative analysis of the host transcriptome provided valuable insights into the molecular landscape underlying anal cancer development. The transition from HGSIL to ASCC was characterized by a statistically significant number of DEG, with notable alterations in keratin expression and overexpression of members of the MAGE gene family in ASCC. Functional analysis revealed key biological processes and pathways associated with each stage. In HGSIL compared to LGSIL, activated processes included nuclear division, chromatin modification, and cell proliferation, aligning with histopathological features indicative of high-grade lesions (35). Conversely, the transition from HGSIL to ASCC revealed immune response activation, marked by upregulation of IFN pathways, highlighting the role of the immune system in the progression to anal squamous cell carcinoma (38). Indeed, patients of Cluster II were characterized by a higher immune signature and immune cell infiltration, as assessed by gene expression profiling of immune cell fractions, IHC of CD3 and CD8 TILs as well as PDL-1 expression.

Therefore, in comparing HGSIL and ASCC, the data underscored the predominance of immune response activation in ASCC, contrasting with the cell proliferation and DNA modification processes observed in the transition from LGSIL to HGSIL. Noteworthy findings included the suppression of the p53 pathway potentially linked to the overexpression of HPV16 E6 protein, highlighting the intricate interplay between viral oncoproteins and host cellular processes in the progression to ASCC (64).

Shared and unique immune and molecular changes across squamous cell carcinomas

Through the integration of transcriptomic studies on HNSCC and CSCC with our ASCC transcriptome, we found shared gene expression patterns across tumor sites indicating a shift towards immune response genes and a decrease in keratinocyte differentiation genes during disease progression from preinvasive to invasive stages. These patterns align with the known biology of HPV carcinogenesis, where HPV E6 oncoprotein downregulates keratinocyte differentiation genes and upregulates mesenchymal lineage genes (65). Regarding the immune response, the elevated immune score and the high frequency of TILs cell fraction observed in cluster II samples from our analysis are also in line with a higher prevalence of HR-HPV. HPV-positive tumors may have increased numbers of TILs, myeloid dendritic cells, and proinflammatory chemokines, which are thought to improve treatment response in patients with head and neck and cervical cancers (66,67). Our results showed that a strong immune response is associated with better treatment outcomes, as

667 indicated by TIDE score analysis. Studies have shown that TILs may improve
668 treatment responses or outcomes in CSCC patients undergoing chemotherapy
669 or radiotherapy (68,69). TIL-based immunotherapy has shown promise as an
670 alternative treatment for advanced cervical cancer, with positive results (70).
671 In advanced ASCC, immunotherapy trials primarily focus on targeting
672 PD1/PDL1 and E6/E7 proteins (71). Therefore, combining TIL therapy with
673 checkpoint blockade and HPV E6/E7 vaccination offers a potent anti-tumor
674 therapy with the potential to eradicate malignancy in ASCC completely.

675 Furthermore, ASCC exhibited somatic missense mutations in cancer driver
676 genes, with *KMT2C*, *PIK3CA* and *EP300*, being the most mutated genes in
677 agreement with previous reports. The prevalence of these mutations varied at
678 different stages of anal precancerous lesions, suggesting their involvement in
679 early stages of anal cancer development. Additionally, there were distinct
680 tumor subpopulations with different mutation rates and immune signatures.
681 Mutations in *KMT2C*, *PRDM1,* and *FGFR2* were predominantly found in HPV16-
682 infected cases, indicating their association with HPV-related carcinogenesis
683 (51,52). The comparative analysis of mutational profiles across different
684 squamous cell carcinomas, including ASCC, HNSCC, and CSCC, revealed
685 significant overlaps in the mutation landscape. By examining datasets from
686 cBioPortal, encompassing a substantial number of cases, we identified
687 common mutations in several cancer driver genes among these carcinomas.
688 Approximately one third of the frequently mutated genes in ASCC were also
689 prevalent in HNSCC and CSCC, suggesting potential shared molecular
690 mechanisms underlying these cancers. Key drivers such as *KMT2C*, *EP300*,
691 *PIK3CA*, *NOTCH1*, *TP53*, *CASP8*, *STK11*, and *KDM6A* emerged as recurrently
692 altered across these cohorts. However, the remaining two thirds of the
693 mutated genes appear to be specific to ASCC, indicating distinct genetic
694 alterations driving the development and progression of anal cancer.

695 Our study has a number of limitations given its cross-sectional nature and the
696 low sample size utilized for data collection due the rarity of anal cancer. The
697 small sample size might not fully represent the biological diversity and
698 variability within the population under investigation, potentially limiting the
699 generalization of the findings. Furthermore, the high risk of false discovery
700 poses a considerable concern, especially in exploratory analyses or when
701 multiple comparisons are conducted. Due to the cross-sectional design
702 adopted in this study, establishing causal associations becomes challenging.
703 However, this is the first cross-sectional study that identifies
704 metatranscriptomics and transcriptomics changes among premalignant and
705 the malignant stages of anal cancer. Furthermore, these findings provide
706 valuable insights into novel prognostic biomarkers that may help to stratify
707 patients with precancerous lesions in low- vs. high-risk groups of progression

to the malignant stage. Future research employing larger sample sizes and longitudinal designs would be needed to address these limitations and corroborate our findings.

Methods

Sex as a biological variable

Sex at birth (male or female) and gender identity (cisgender men (CGM), transgender women (TGW), and cisgender women (CGW)), were incorporated into our study design as biological variables.

Sample Collection and RNA Sequencing

We collected 70 anal biopsies from patients with different stages of anal lesions: 31 LGSIL, 16 HGSIL, and 23 ASCC, stored in RNAlater (Thermo Fisher Scientific, USA). Clinical data including age, HPV status, ART treatment, and HIV status were recorded at enrollment. RNA was extracted using miRNeasy Tissue/Cells Advanced Kits (Qiagen), and its quality was assessed on an Agilent 2100 Bioanalyzer. Samples with RNA integrity number (RIN) >7 were chosen for RNA-seq. Directional RNA-seq libraries were prepared using Illumina Total RNA Prep with Ribo-Zero Plus kit. Sequencing was performed on an Illumina Novaseq 6000 platform, yielding approximately 80 million clusters per sample with >92% >Q30 quality scores.

DNA purification and HPV detection and genotyping

Samples were collected using Qiagen specimen collection devices (Qiagen, USA) by qualified staff at Fundación Huésped and Hospital Udaondo. DNA purification utilized QIAMP DNA kits (Qiagen, USA). DNA integrity and concentration were assessed by Nanodrop spectrophotometry. HPV detection was performed at Institute Malbrán via PCR using biotinylated Broad-Spectrum General Primers BSGP5+/GP6, designed to amplify a 140 bp fragment of the HPV-L1 gene. Genotyping was conducted using reverse line blot hybridization (RLB) for 36 HPV genotypes (validated by Global HPV LabNet) (72). Biotinylated amplicons were denatured and hybridized with genotype-specific oligonucleotide probes immobilized as parallel lines on membrane strips.

Metatranscriptomic Data Analysis.

For metatranscriptomics analysis, the obtained RNA-seq data were processed using the Biobakery suite of tools: KneadData was used to separate the human and the non-human reads; taxonomic profiling was performed using MetaPhlAn to identify and quantify microbial taxa at species level present in the anal samples (73).

745 Species richness and diversity were calculated using the R function
746 estimate_richness from R package phyloseq (74). We considered the observed
747 species and Chao1 indices for richness, and the Shannon and Simpson indices
748 for diversity. Beta diversity was measured by Bray–Curtis, weighted UniFrac,
749 and unweighted UniFrac. For Principal Coordinate Analysis, the Aitchison
750 distance was used as the distance metric to analyze the compositional data.
751 To test whether the samples cluster beyond that expected by sampling
752 variability we applied permutational multivariate analysis of variance
753 (PERMANOVA) Differences in richness and diversity indices between groups
754 were determined using the Wilcoxon rank sum test with a significance level of
755 0.05. For relative abundance analysis and visualization, we used R phyloseq
756 packages.

757 Differential abundance analysis

758 For determining the relative differential abundance and the multivariable
759 association between subjects' metadata and microbial features, we used the
760 MaAsLin2 package from the bioBakery suite in R/Bioconductor (75). We used
761 default parameters for normalization (TSS method), transformation (Log),
762 analysis method (LM), correction method (BH), and significance threshold (q-
763 value < 0.25). The minimum abundance for each feature was set to 0.001
764 (0.1%) while the minimum percent of samples for which a feature was
765 detected (prevalence, Pr) at minimum abundance was used as follows: 0.05
766 (5%) for viruses, 0.1 (10%) for bacteria and pathways and 0.2 (20%) for gene
767 families.

768 Pathways and gene family analysis

769 Metatranscriptomics pathway analysis was conducted using the HMP Unified
770 Metabolic Analysis Network 3 (HUMAnN3) pipeline to investigate potential
771 variations in metabolic pathways. HUMAnN3 employs a multifaceted
772 approach, extracting species profiles from KneadData output, aligning reads
773 to pan-genomes, executing translated searches on unclassified reads, and
774 quantifying gene families and pathways. By default, gene families are
775 UniRef90 annotated and metabolic pathways are annotated using MetaCyc
776 database (76,77). The UniRef90 gene family abundance from HUMAnN3 was
777 then regrouped to Kyoto Encyclopedia of Genes and Genomes (KEGG)
778 orthology (KO) (78) We used the KEGGREST package in R/Bioconductor for KO
779 identifiers and KEGG Mapper reconstruct tool for KEGG pathway maps (79,80).

780 Data visualization

781 We used the R package "phyloseq" to create a heatmap representation of taxa
782 abundances. For the unsupervised ordination of samples, we applied the
783 NMDS method and Bray distance in the plot_heatmap function. Heatmap

visualization of differentially represented gene proteins was done with R/Bioconductor and the MultiExperiment Viewer software (MeV v4.9).

Transcriptomic, functional enrichment and immune infiltrate analysis.

The raw short-read sequences were preprocessed using Rfastp from the R/Bioconductor package Rsubread. Quality checks, adapter removal, and trimming of low-quality bases were conducted with Rfastp. Reads were aligned to the human genome reference GRCh38 using the Subread aligner algorithm from Rsubread. Gene expression abundance at the whole-genome level was calculated using featureCounts from Rsubread. Differential gene expression analysis between anal lesion stages (LGSIL vs. HGSIL and HGSIL vs. ASCC) utilized edgeR, with fold changes and adjusted p-values computed based on normalized log2 count per million values. Genes with a log-fold change >1 and adjusted p-value <0.05 were considered differentially expressed.

Functional enrichment analysis of differentially expressed genes employed the clusterProfiler package for Gene Set Enrichment Analysis (GSEA) (81). Functional enrichment results were visualized using enrichplot for Gene Ontology, Hallmark of Cancers, and Disease Ontology terms. Heatmaps were generated using MultiExperiment Viewer (MeV) 4.9.0.

Tumor purity, immune cell infiltration, and T cell dysfunction/exclusion scores were estimated using ESTIMATE, EPIC, and TIDE algorithms, respectively, on normalized count matrices.

For comparative transcriptomics analysis, the GSE74927 dataset for HNSCC and GSE63514 for CSCC were utilized. Raw data were imported into R using GEOquery to obtain normalized matrices for each study. Differential gene expression analysis employed DESeq2 for GSE74927 and limma for GSE63514. To visualize gene expression patterns, we defined the following gene signatures: for ASCC, we used the gene signature obtained from our comparison of HGSIL vs. ASCC; for HNSCC, we used the gene signature provided in the study by Zhang et al., derived from the differential expression analysis between two HPV+ subgroups of HNSCC (48); for CSCC, we obtained a gene signature from the comparison between CIN2/CIN3 samples (comparable to HGSIL) and CSCC. Gene expression profiles across ASCC, HNSCC, and CSCC matrices were visualized after filtering out genes with <50% variance within each signature. Functional enrichment analysis of resulting genes used the ClusterProfiler package. Heatmaps were generated in MeV 4.9.0 based on immune scores.

Mutational analysis based on RNA-seq data.

822 The preprocessed reads previously used for the transcriptomic analysis were
823 aligned and mapped to the human genome reference GRCh38 using the
824 Subjunc aligner algorithm provided by Rsubread R/Bioconductor package.
825 Subjunc aligner was developed for aligning RNA-seq reads and for the
826 detection of exon-exon junctions at the same time. The Subjunc mapping
827 results (BAM files) were used for genomic variants detection using the
828 exactSNP variant caller algorithm provided by Rsubread package. The
829 VariantAnnotation R/Bioconductor package was subsequently used for SNPs
830 and InDels filtering of the obtained VCF files based on quality (QUAL > 20) and
831 coverage (DP>10) metrics. Identified variants were annotated, filtered and
832 interpreted using OpenCRAVAT and their aggregated variant databases and
833 resources (GnomAD, Cancer Genome Interpreter, Cancer Hotspots, CIVIC,
834 Cosmic, SIFT, PolyPhen2) for the prediction of somatic mutations in cancer
835 driver genes.

836 In addition, to perform a comparative analysis of the mutational profile
837 identified in ASCC with other squamous cell carcinomas, we analyzed HNSCC
838 and CSCC datasets obtained from cBioPortal online resource
839 (http://www.cbioportal.org/). Briefly, the mutational profiles of the 25 cancer
840 driver genes mutated in ASCC were retrieved from two combined cervical
841 cancer datasets (MSK-CESC and TCGA-CESC, n=468) and a head and neck
842 cancer dataset (TCGA-HNSC, n=510). Only drivers and putative drivers'
843 somatic missense or truncating mutations were considered for frequency
844 estimations among cohorts.

845 Immunohistochemistry analysis of ASCC.

846 Immunostaining utilized a Roche Benchmark XT system with anti-CD3 (Clone
847 2GV6, Ventana - Roche), anti-CD8 (Clone SP57, Ventana - Roche), anti-PD-L1
848 (Clone SP263, Ventana - Roche), and anti-p16 (Clone 6H12, Leica Biosystems)
849 antibodies. Evaluation involved two independent pathologists, with
850 discrepancies resolved by a senior pathologist in four cases. CD3 and CD8
851 expression levels were averaged across intra- and peritumoral areas and
852 categorized as low (0–34%), moderate (35–64%), or high (65–100%) based on
853 total tumor-related lymphocyte staining. PD-L1 expression was assessed using
854 the Combined Positive Score (CPS) for gastric/gastroesophageal junction
855 adenocarcinoma.

856 Statistics

857 We used R/Bioconductor for different statistical comparisons outside of
858 MaAsLin's analysis. To analyze continuous variables, we utilized either two
859 tailed t-tests or Wilcoxon tests as appropriate. For categorical data, we
860 employed Chi-squared and Fisher tests. Box-plots were created in R using the
861 ggplot package.

Study approval.

This study was approved by the institutional review boards of Fundación Huésped and Hospital de Gastroenterología "Dr. Carlos Bonorino Udaondo," both in Buenos Aires, Argentina. All participants included in this study signed informed consent before being involved in the project.

Data availability

The raw data have been submitted to NCBI GEO database with accession number GSE253560. Supporting data values of figures and table 1 are available as supplementary files. The rest of the data are available from the corresponding author upon reasonable request. All codes and scripts used for data preprocessing and analysis are available at the following Github repository: https://github.com/mabba777/ASCC-transcriptomics.

Author contributions

EL: Investigation, formal analysis, writing the article. VF, MIF, PC: resources; logistics of obtaining samples and clinical data of participants. MES, AMG, JAB, MAP, MG, JR, MK, SI, SW: methodology, research assistance, clinical data of participants. OC, JCR: resources. MCA: Conceived the study, supervision, formal analysis, writing the article.

Acknowledgements

The authors want to acknowledge the invaluable contributions from all study participants and from all the research team at Fundación Huésped, where participants were recruited. This work was supported by the NIH grant CA221208.

Conflict of interest

The authors have declared that no conflict of interest exists.

References

1. Clifford, GM. et al. A meta-analysis of anal cancer incidence by risk group: toward a unified anal cancer risk scale. Int. J. Cancer. 2020; 148, 38–40

2. Darragh TM, et al. The Lower Anogenital Squamous Terminology Standardization project for HPV-associated lesions: background and consensus recommendations from the College of American Pathologists and

894 the American Society for Colposcopy and Cervical Pathology. Arch Pathol Lab
895 Med. 2012; 136: 1266–97.

896 3. Berry JM, et al. Progression of anal high-grade squamous intraepithelial
897 lesions to invasive anal cancer among HIV-infected men who have sex with
898 men. Int J Cancer. 2014; 134: 1147-55.

899 4. Hoots BE, et al. Human papillomavirus type distribution in anal cancer and
900 anal intraepithelial lesions. Int J Cancer. 2009;124: 2375-83.

901 5. de Martel C, et al. Worldwide burden of cancer attributable to HPV by site,
902 country and HPV type. Int J Cancer. 2017; 141: 664–70.

903 6. Kelly H, et al. Association of antiretroviral therapy with anal high-risk human
904 papillomavirus, anal intraepithelial neoplasia, and anal cancer in people living
905 with HIV: a systematic review and meta-analysis. Lancet HIV. 2020 ;7(4): e262-
906 e278.

907 7. Kadosh E, et al. The gut microbiome switches mutant p53 from tumour-
908 suppressive to oncogenic. Nature. 2020; 586(7827):133-138.

909 8. Serrano-Villar, et al. Microbiome-derived cobalamin and succinyl-CoA as
910 biomarkers for improved screening of anal cancer. Nat Med. 2023; 29, 1738–
911 1749.

912 9. Herrera, S., et al. HIV, cancer, and the microbiota: common pathways
913 influencing different diseases. Front Immunol. 2019; 10, 1466.

914 10. Lacunza E, et al. Oral and anal microbiome from HIV-exposed individuals:
915 role of host-associated factors in taxa composition and metabolic pathways.
916 NPJ Biofilms Microbiomes. 2023;9(1):48.

917 11. Lin D, et al. Microbiome dynamics during chemoradiation therapy for anal
918 cancer. Int J Radiat Oncol Biol Phys. 2022; 113:974–84.

919 12. Elnaggar JH, et al. HPV-related anal cancer is associated with changes in
920 the anorectal microbiome during cancer development. Front Immunol. 2023;
921 14:1051431.

922 13. Bernardi MP, et al. Molecular biology of anal squamous cell carcinoma:
923 implications for future research and clinical intervention. Lancet Oncol. 2015;
924 16(16):e611-21.

925 14. Fraunholz I, et al. High survivin expression as a risk factor in patients with
926 anal carcinoma treated with concurrent chemoradiotherapy. Radiat Oncol.
927 2012; 7: 88.

928 15. Sheppard K, et al. Targeting PI3 kinase/AKT/mTOR signaling in cancer. Crit
929 Rev Oncog. 2012; 17: 69–95.

930 16. Gilbert DC, et al. p16INK4A, p53, EGFR expression and KRAS mutation
931 status in squamous cell cancers of the anus: correlation with outcomes
932 following chemo radiotherapy. Radiother Oncol. 2013; 109: 146–51.

933 17. Iseas S, et al. A clinical and molecular portrait of non-metastatic anal
934 squamous cell carcinoma. Transl Oncol. 2021; 14(6):101084.

935 18. Zhang Y, et al.. Metatranscriptomics for the Human Microbiome and
936 Microbial Community Functional Profiling. Annu Rev Biomed Data Sci. 2021
937 ;4:279-311.

938 19. Bullman S, et al. Analysis of Fusobacterium persistence and antibiotic
939 response in colorectal cancer. Science. 2017; 358(6369):1443-1448.

940 20. Klein C, et al. Mycoplasma Co-Infection Is Associated with Cervical Cancer
941 Risk. Cancers (Basel). 2020;12(5):1093.

942 21. Zhang C, et al. The direct and indirect association of cervical microbiota
943 with the risk of cervical intraepithelial neoplasia. Cancer Med. 2018; 7(5):2172-
944 2179.

945 22. Nguyen HDT, et al. Relationship between Human Papillomavirus Status and
946 the Cervicovaginal Microbiome in Cervical Cancer. Microorganisms. 2023;
947 11(6):1417.

948 23. Liu CJ, et al. Genital Microbiota of Women From Six Ethnic Groups With
949 and Without Human Papillomavirus Infection in Shangri-La, China. Front Cell
950 Infect Microbiol. 2022; 12:935068.

951 24. Chen Z, et al. Evolution and taxonomic classification of alphapapillomavirus
952 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70. PLoS
953 One. 2013;8(8):e72565.

954 25. Lin C, et al. Human papillomavirus types from infection to cancer in the
955 anus, according to sex and HIV status: a systematic review and meta-analysis.
956 Lancet Infect Dis. 2018; 18(2):198-206.

957 26. Rivas SR, et al. The Role of HERV-K in Cancer Stemness. Viruses. 2022
958 ;14(9):2019.

959 27. Alibek K, et al. Implication of human herpesviruses in oncogenesis through
960 immune evasion and supression. Infect Agent Cancer. 2014; 9(1):3.

961 28. Zhu J, Thompson CB. Metabolic regulation of cell growth and proliferation.
962 Nat Rev Mol Cell Biol. 2019; 20(7):436-450.

963 29. Guo P, et al. FadA promotes DNA damage and progression of
964 Fusobacterium nucleatum-induced colorectal cancer through up-regulation of
965 chk2. J Exp Clin Cancer Res. 2020; 39(1):202.

966 30. Benedetti F, et al. Exogenous bacterial DnaK increases protein kinases
967 activity in human cancer cell lines. J Transl Med. 2021; 19(1):60.

968 31. Scheffner M, et al. The E6 oncoprotein encoded by human papillomavirus
969 types 16 and 18 promotes the degradation of p53. Cell. 1990; 63(6):1129-36.

970 32. Svidler López L, La Rosa L. Human Papilloma Virus Infection and Anal
971 Squamous Intraepithelial Lesions. Clin Colon Rectal Surg. 2019; 32(5):347-357.

972 33. Caballero OL, Chen YT. Cancer/testis (CT) antigens: potential targets for
973 immunotherapy. Cancer Sci. 2009; 100(11):2014-21.

974 34. Hong DS, et al. Autologous T cell therapy for MAGE-A4+ solid cancers in
975 HLA-A*02+ patients: a phase 1 trial. Nat Med. 2023;29(1):104-114.

976 35. Khieu M, Butler SL. High-Grade Squamous Intraepithelial Lesion of the
977 Cervix. 2023. In: StatPearls [Internet]. Treasure Island (FL): StatPearls
978 Publishing.

979 36. Nees M, et al. Papillomavirus type 16 oncogenes downregulate expression
980 of interferon-responsive genes and upregulate proliferation-associated and
981 NF-kappaB-responsive genes in cervical keratinocytes. J Virol. 2001;
982 75(9):4283-96.

983 37. Cosper PF, et al. HPV16 E6 induces chromosomal instability due to polar
984 chromosomes caused by E6AP-dependent degradation of the mitotic kinesin
985 CENP-E. Proc Natl Acad Sci U S A. 2023; 120(14):e2216700120.

986 38. Martin D, et al. The immune microenvironment and HPV in anal cancer:
987 Rationale to complement chemoradiation with immunotherapy. Biochim
988 Biophys Acta Rev Cancer. 2017; 1868(1):221-230.

989 39. Moir S, Fauci AS. B cells in HIV infection and disease. Nat Rev Immunol.
990 2009 ;9(4):235-45.

991 40. Lepique AP, et al. HPV16 tumor associated macrophages suppress
992 antitumor T cell responses. Clin Cancer Res. 2009;15(13):4391-400.

993 41. Grabenbauer GG, et al. Tumor-infiltrating cytotoxic T cells but not
994 regulatory T cells predict outcome in anal squamous cell carcinoma. Clin
995 Cancer Res. 2006; 12: 3355–3360.

996 42. Mei Z, Liu Y, et al. Tumour-infiltrating inflammation and prognosis in
997 colorectal cancer: systematic review and meta-analysis. Br J Cancer. 2014;
998 110(6): 1595–1605.

999 43. Matsumoto H, et al. Role of inflammatory infiltrates in triple negative
1000 breast cancer. J Clin Pathol. 2015; 68(7): 506–510.

1001 44. Gilbert DC, et al. Tumour-infiltrating lymphocyte scores effectively stratify
1002 outcomes over and above p16 post chemo-radiotherapy in anal cancer. Br J
1003 Cancer. 2016; 114(2):134-7.

1004 45. Gilbert DC, et al. p16(INK4A), p53, EGFR expression and KRAS mutation
1005 status in squamous cell cancers of the anus: Correlation with outcomes
1006 following chemo-radiotherapy. Radiother Oncol. 2013; 109: 146–151.

1007 46. Koerber SA, et al. Influence of human papillomavirus and p16(INK4a) on
1008 treatment outcome of patients with anal cancer. Radiother Oncol. 2014; 113:
1009 331–336.

1010 47. Rodel F, et al. Human papillomavirus DNA load and p16INK4a expression
1011 predict for local control in patients with anal squamous cell carcinoma treated
1012 with chemoradiotherapy. Int J Cancer. 2014; 136: 278–288.

1013 48. Zhang Y, et al. Subtypes of HPV-Positive Head and Neck Cancers Are
1014 Associated with HPV Characteristics, Copy Number Alterations, PIK3CA
1015 Mutation, and Pathway Signatures. Clin Cancer Res. 2016; 15;22(18):4735-45.

1016 49. den Boon JA, et al. Molecular transitions from papillomavirus infection to
1017 cervical precancer and cancer: Role of stromal estrogen receptor signaling.
1018 Proc Natl Acad Sci U S A. 2015;112(25): E3255-64.

1019 50. Litwin TR, et al. Somatic Host Cell Alterations in HPV Carcinogenesis.
1020 Viruses. 2017;9(8):206.

1021 51. Morris, V et al. Comprehensive Genomic Profiling of Metastatic Squamous
1022 Cell Carcinoma of the Anal Canal. Mol. Cancer Res. 2017; 15, 1542–1550.

1023 52. Chung, et al. Comprehensive genomic profiling of anal squamous cell
1024 carcinoma reveals distinct genomically defined classes. Ann Oncol. 2016; 27,
1025 1336–1341.

1026 53. Smaglo, et al. Comprehensive multiplatform biomarker analysis of 199 anal
1027 squamous cell carcinomas. Oncotarget. 2015; 6(41):43594-604.

1028 54. Mandelbaum J. et al. BLIMP1 is a tumor suppressor gene frequently
1029 disrupted in activated B cell-like diffuse large B cell lymphoma. Cancer Cell.
1030 2010; 18(6):568-79.

1031 55. Smith SP et al. Identification of host transcriptional networks showing
1032 concentration-dependent regulation by HPV16 E6 and E7 proteins in basal
1033 cervical squamous epithelial cells. Sci Rep. 2016; 6:29832.

1034 56. Mahmood HA, et al. FGF signalling facilitates cervical cancer progression.
1035 FEBS J. 2022; 289(12):3440-3456.

1036 57. Hu D, et al. The MLL3/MLL4 branches of the COMPASS family function as
1037 major histone H3K4 monomethylases at enhancers. Mol Cell Biol. 2013;
1038 33:4745–54.

1039 58. Lee J, et al. A tumor suppressive coactivator complex of p53 containing
1040 ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4.
1041 Proc Natl Acad Sci U S A. 2009; 106(21):8513–8518.

1042 59. Zimmermann H, et al. The human papillomavirus type 16 E6 oncoprotein
1043 can down-regulate p53 activity by targeting the transcriptional coactivator
1044 CBP/p300. J Virol. 1999;73(8):6209–6219.

1045 60. Gillison ML, et al. Human papillomavirus and the landscape of secondary
1046 genetic alterations in oral cancers. Genome Res. 2019;29(1):1–17.

1047 61. Nilsson MP, et al. Patterns of recurrence in anal cancer: a detailed analysis.
1048 Radiat Oncol. 2020;15(1):125.

1049 62. Jones CM, et al. Biomarkers in anal cancer: from biological understanding
1050 to stratified treatment. Br J Cancer. 2017; 116(2):156-162.

1051 63. Arizmendi-Izazaga, A. et al. Metabolic reprogramming in cancer: role of
1052 HPV 16 variants. Pathogens. 2021; 10(3):347.

1053 64. Zhu X, et al. Molecular and immunophenotypic characterization of anal
1054 squamous cell carcinoma reveals distinct clinicopathologic groups associated
1055 with HPV and TP53 mutation status. Mod Pathol. 2021; 34(5):1017-1030.

1056 65. Ganti K, et al. The human papillomavirus E6 PDZ binding motif: from life
1057 cycle to malignancy. Viruses. 2015; 7:3530–3551.

1058 66. Partlová S, et al. Distinct patterns of intratumoral immune cell infiltrates in
1059 patients with HPV-associated compared to non-virally induced head and neck
1060 squamous cell carcinoma. Oncoimmunology. 2015; 4(1):e965570.

1061 67. Tang Y, et al. Prognostic and therapeutic TILs of cervical cancer-Current
1062 advances and future perspectives. Mol Ther Oncolytics. 2021; 22:410-430.

1063 68. Martins PR, et al. Cervical cancer patients that respond to chemoradiation
1064 therapy display an intense tumor infiltrating immune profile before treatment.
1065 Exp Mol Pathol. 2019; 111:104314.

1066 69. Heeren AM, et al. Neoadjuvant cisplatin and paclitaxel modulate tumor-
1067 infiltrating T cells in patients with cervical cancer. Cancer Immunol
1068 Immunother. 2019; 68(11):1759-1767.

1069 70. Zhu Y, et al. Adoptive tumor infiltrating lymphocytes cell therapy for
1070 cervical cancer. Hum Vaccin Immunother. 2022; 18(5):2060019.

1071 71. Dhawan N, et al. Immunotherapy in Anal Cancer. Curr Oncol. 2023;
1072 30(5):4538-4550.

1073 72. Eklund C, et al. The 2019 HPV Labnet international proficiency study: Need
1074 of global Human Papillomavirus Proficiency Testing. J Clin Virol. 2021;
1075 141:104902.

1076 73. McIver LJ, et al. bioBakery: a meta'omic analysis environment.
1077 Bioinformatics. 2018; 34(7):1235-1237.

1078 74. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible
1079 interactive analysis and graphics of microbiome census data. PLoS ONE. 2013;
1080 8(4):e61217.

1081 75. Mallick, H. et al. Multivariable association discovery in population-scale
1082 metaomics studies. PLoS Comput Biol. 2021; 17(11):e1009442.

1083 76. Suzek, BE, et al. UniProt Consortium. UniRef clusters: a comprehensive and
1084 scalable alternative for improving sequence similarity searches.
1085 Bioinformatics. 2015; 31, 926–932.

1086 77. Caspi, R et al. The MetaCyc database of metabolic pathways and enzymes
1087 and the BioCyc collection of pathway/genome databases. Nucleic Acids Res.
1088 2016; 44(D1), D471–D480.

1089 78. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes.
1090 Nucleic Acids Res. 2000; 28, 27–30.

1091 79. Tenenbaum D, Maintainer B. KEGGREST: Client-side REST access to the
1092 Kyoto Encyclopedia of Genes and Genomes (KEGG). R package version 1.44.1.
1093 Bioconductor. 2024.

1094 80. Kanehisa, M, Sato Y. KEGG Mapper for inferring cellular functions from
1095 protein sequences. Protein Sci. 2020; 29, 28–35.

1096 81. Wu T, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting
1097 omics data. Innovation (Camb). 2021;2(3):100141.

1098

1099

1100

1101

1102

1103

1104

1105 **Table 1:** Clinical and demographics data of patients.

Baseline characteristics	LGSIL n=31	HGSIL n=16	ASCC n=23	P-value *
Age (years)	33.22 ± 12.88	41.31 ± 11.14	52.23 ± 13.30	< 0.001
Sex at birth				0.011
Male: 61 (CGM+TGW)	31 (100%)	16 (100%)	14 (61%)	
Female: 9	0	0	9 (39%)	
Gender				0.027
CGM: 53 (all MSM)	28 (90%)	11 (69%)	14 (61%)	
TGW: 8	3 (10%)	5 (31%)	0	
CGW: 9	0	0	9 (39%)	
HPV DNA status				< 0.001
Low Risk	21(67%)	5 (32%)	2 (9%)	
High Risk	3 (10%)	8 (50%)	16 (70%)	
Undetected	7 (23%)	3 (18%)	5 (21%)	
HIV status				0.004
Positive	28 (90%)	15 (94%)	9 (52%)	
Negative	3 (10%)	1 (6%)	8 (48%)	
NA			6	
ART (HIV positive cases)				0.321
Treated	27 (96%)	15 (100%)	5 (83)	
Untreated	1 (4%)	0	1 (17%)	
NA	0	0	3	

1106 * **Fisher Exact Test**

1107

1108

1109

1110

1111

1112

1113

1114

1115

1116

1117

1118

1119

1120

1121 **Table 2: Metabolic pathways enriched in ASCC compared with SILs.**

PWY_ID	Name	Description	coef	pval	qval
PWY-7219	adenosine_ribonucleotides_de_novo_biosynthesis	*Nucleotides synthesis*	2.44	0.0015	0.0103
PWY-7221	guanosine_ribonucleotides_de_novo_biosynthesis	*Nucleotides synthesis*	1.98	0.0035	0.0113
SER_GLYSYN	PWY-_superpathway_of_L_serine_and_glycine_biosynthesis_I	*Amino acids synthesis*	1.93	0.0141	0.0206
PWY-6122	5_aminoimidazole_ribonucleotide_biosynthesis_II	*Nucleotides synthesis*	1.76	0.0010	0.0103
PWY-6277	superpathway_of_5_aminoimidazole_ribonucleotide_biosynthesis	*Nucleotides synthesis*	1.76	0.0010	0.0103
PWY-6121	5_aminoimidazole_ribonucleotide_biosynthesis_I	*Nucleotides synthesis*	1.72	0.0012	0.0103
VALSYN-PWY	L_valine_biosynthesis	*Amino acids synthesis*	1.65	0.0022	0.0113
PWY-7208	superpathway_of_pyrimidine_nucleobases_salvage	*Nucleotides synthesis*	1.55	0.0034	0.0113
PWY-7228	superpathway_of_guanosine_nucleotides_de_novo_biosynthesis_I	*Nucleotides synthesis*	1.54	0.0018	0.0103
COA-PWY	coenzyme_A_biosynthesis_I	*Coenzymes synthesis*	1.48	0.0046	0.0113
PWY-7220	adenosine_deoxyribonucleotides_de_novo_biosynthesis_II	*Nucleotides synthesis*	1.47	0.0015	0.0103
PWY-7222	guanosine_deoxyribonucleotides_de_novo_biosynthesis_II	*Nucleotides synthesis*	1.47	0.0015	0.0103
PWY-7663	gondoate_biosynthesis	*Fatty acid synthesis*	1.46	0.0133	0.0206
PWY-5973	cis_vaccenate_biosynthesis	*Fatty acid synthesis*	1.41	0.0126	0.0206
PWY-6151	S_adenosyl_L_methionine_salvage_I	*Amino acids synthesis*	1.38	0.0051	0.0116
FASYN-INITIAL-PWY	superpathway_of_fatty_acid_biosynthesis_initiation	*Fatty acid synthesis*	1.31	0.0038	0.0113
UDPNAGSYN-PWY	UDP_N acetyl_D_glucosamine_biosynthesis_I	*Nucleotides sugar synthesis*	1.25	0.0047	0.0113
PWY-6125	superpathway_of_guanosine_nucleotides_de_novo_biosynthesis_II	*Nucleotides synthesis*	1.22	0.0040	0.0113
PWY-2942	L_lysine_biosynthesis_III	*Amino acids synthesis*	1.16	0.0148	0.0206
PWY-6124	inosine_5._phosphate_biosynthesis_II	*Amino acids synthesis*	1.15	0.0040	0.0113

1122

1123

1124

1125

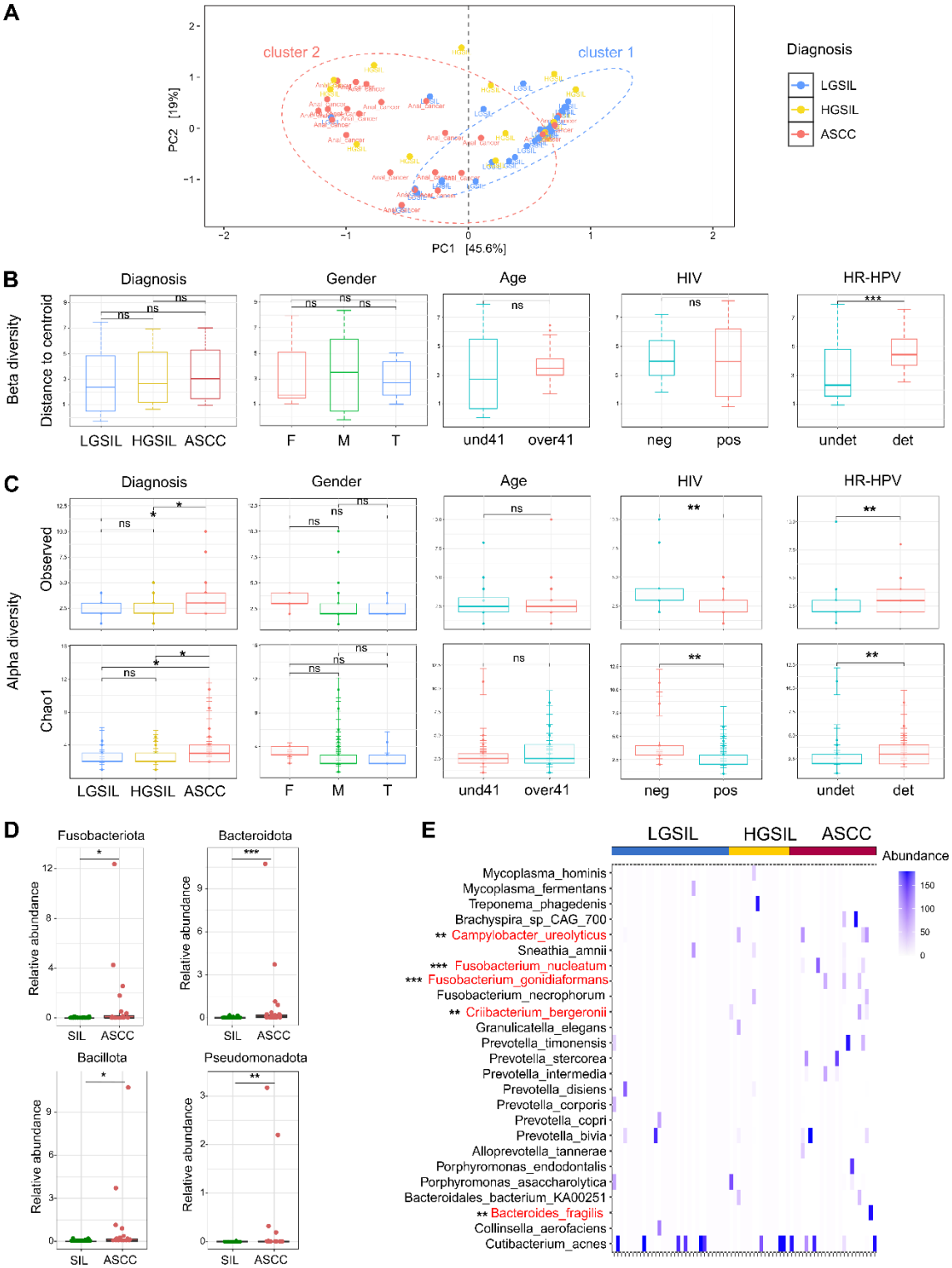


Figure 1: Richness, diversity, and microbial profile of LGSIL, HGSIL and ASCC. A. Principal Coordinate Analysis depicting the unsupervised distribution of samples, assessed at the species level based on microbiota composition, and evaluated through Euclidean distance. **B.** Beta diversity comparison between diagnosis groups and covariates. **C.** Observed and Chao1 richness indices obtained at species level by metatranscriptome analysis. **D.** Significantly altered phyla Fusobacteriota, Bacteroidota, Bacillota and Pseudomonadota, were related to ASCC. Statistical significance was calculated with the Wilcoxon signed-rank test. **E.** Heatmap representation of the relative

abundances of the most abundant bacterial species identified across all samples. Highlighted in red are the taxa significantly enriched in ASCC compared with SIL obtained by MaAsLin2 analysis. * p <0.05; ** p<0.01; *** p <0.001.

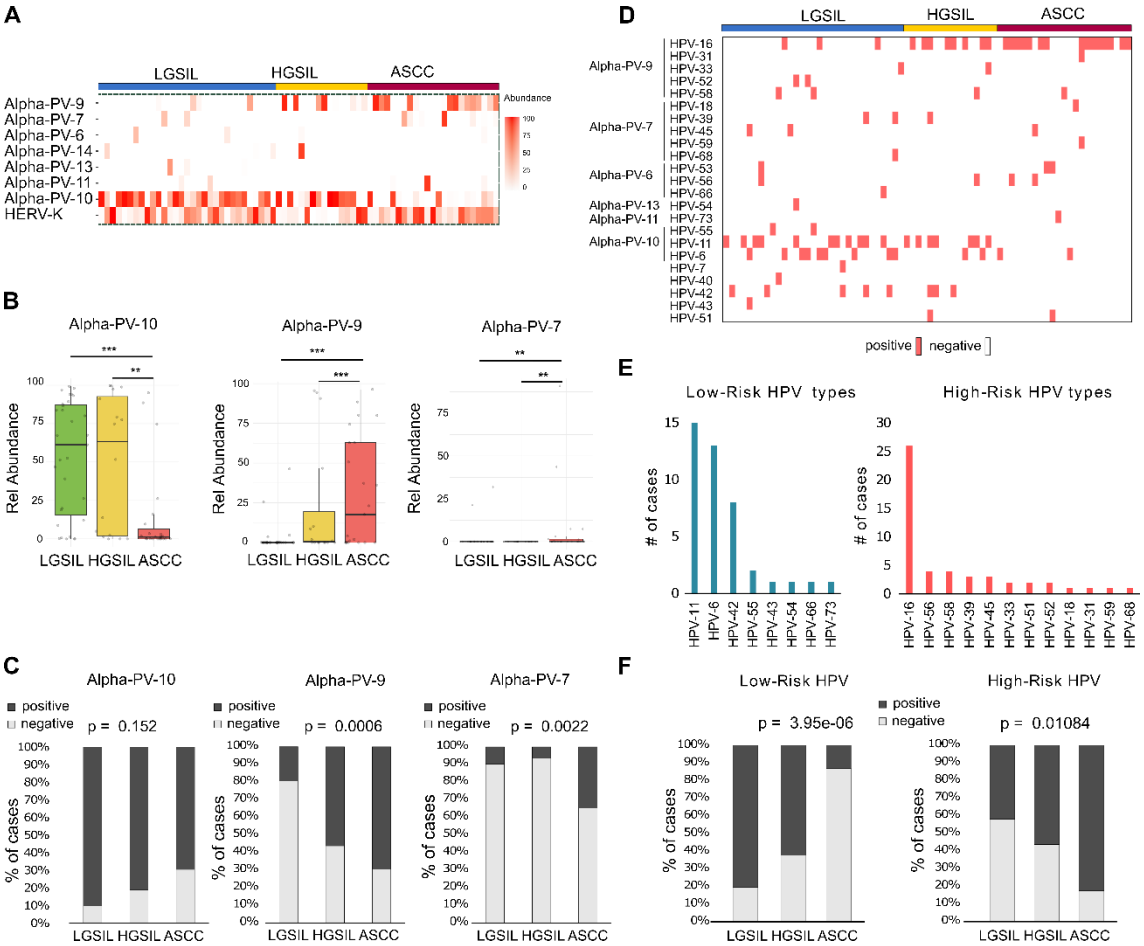


Figure 2: Viral composition of LGSIL, HGSIL and ASCC. A. Relative abundance heatmap showing the most prevalent viral species identified in the diagnosis groups using meta-transcriptome analysis (RNA Level). **B.** Alpha PV-10 was found to be linked to the SIL group, whereas Alpha PV-9 and Alpha PV-7 were associated with ASCC. Statistical significance was derived from MaAsLin2 analysis **C.** Percentage of patients with detectable viruses of the species Alpha PV 10, 9, and 7 assessed by meta-transcriptome. Statistical significance was determined through the application of the Fisher exact test**.** **D.** Tile plot visualizing the HPV types identified through DNA genotyping across the different diagnosis groups. **E.** Percentage distribution of HPV types, assessed by DNA genotyping and classified into low-risk and high-risk categories. **F.** Percentage of patients in each diagnostic group with detectable low-risk and high-risk HPV types identified through DNA genotyping. Statistical significance was determined through the application of the Fisher exact test**.**

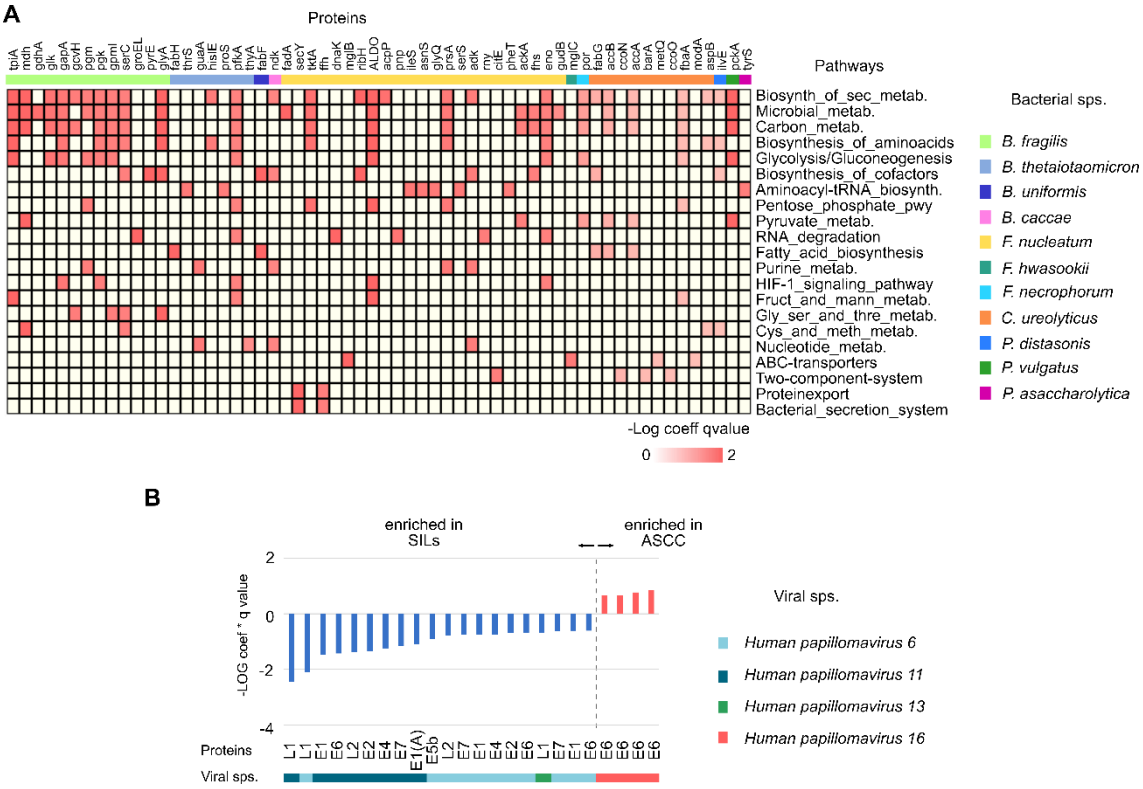


Figure 3: Functional and taxonomic enrichment of microbial gene proteins associated with anal lesions. A. Heatmap representation of metabolic pathways enriched in ASCC compared with SILs represented by 60 gene proteins contributed by relevant gut microbiota taxa of which *F nucleatum*, *B fragilis*, and *C ureolyticus* are predominant. **B**. Viral proteins identified as differentially abundant in ASCC relative to SILs contributed by high risk and low risk HPV.

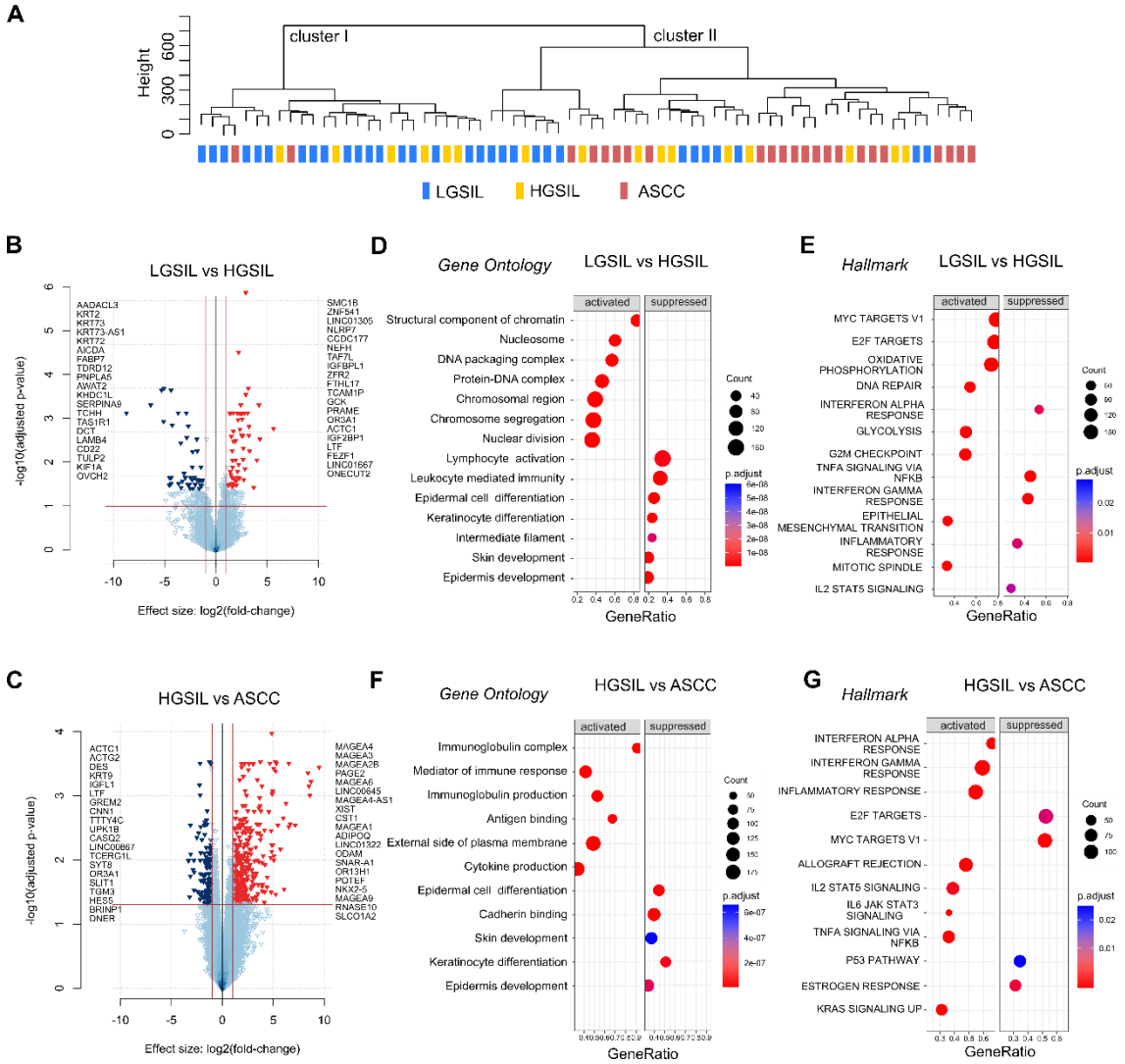
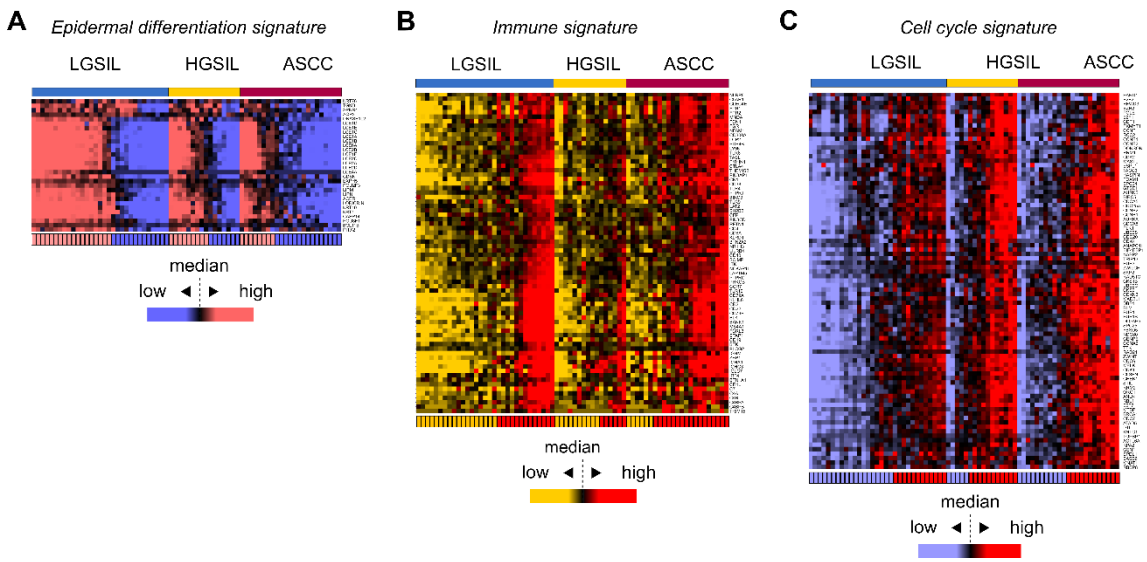


Figure 4: Differential gene expression analysis and functional enrichment of transcriptomic data. A. Unsupervised hierarchical clustering of samples classified according to diagnosis groups **B-C.** Volcano plots representing significant differentially expressed genes (LogFC > 1, adj p-value < 0.05) from the comparisons between LGSIL and HGSIL (**B**) and between HGSIL and ASCC (**C**). Upregulated genes are indicated by red arrowheads, while downregulated genes are indicated by blue arrowheads. The top 20 significant genes are shown. **D-G**. Dot plots of Gene Set Enrichment Analysis obtained from the comparisons between LGSIL and HGSIL (D-E) and between HGSIL and ASCC (F-G). **D.** Dot plot of significantly activated and suppressed Gene Ontology pathways in HGSIL compared with LGSIL. **E.** Dot plot of significantly activated and suppressed Hallmarks of Cancer in HGSIL compared with LGSIL. **F.** Dot plot of significantly activated and suppressed Gene Ontology pathways in ASCC compared with HGSIL. **G.** Dot plot of significantly activated and suppressed Hallmarks of Cancer in ASCC compared with HGSIL.



1181

1182 **Figure 5: Heatmaps illustrating the expression profiles of gene signatures across**
1183 **diagnostic groups—LGSIL, HGSIL, and ASCC. A.** Epidermal differentiation signature. **B.**
1184 Immune signature. **C.** Cell cycle signature. The color coding bar at the bottom of each
1185 heatmap indicates the score (high or low) assigned to each sample based on the average
1186 expression of the gene signature divided by the median value.

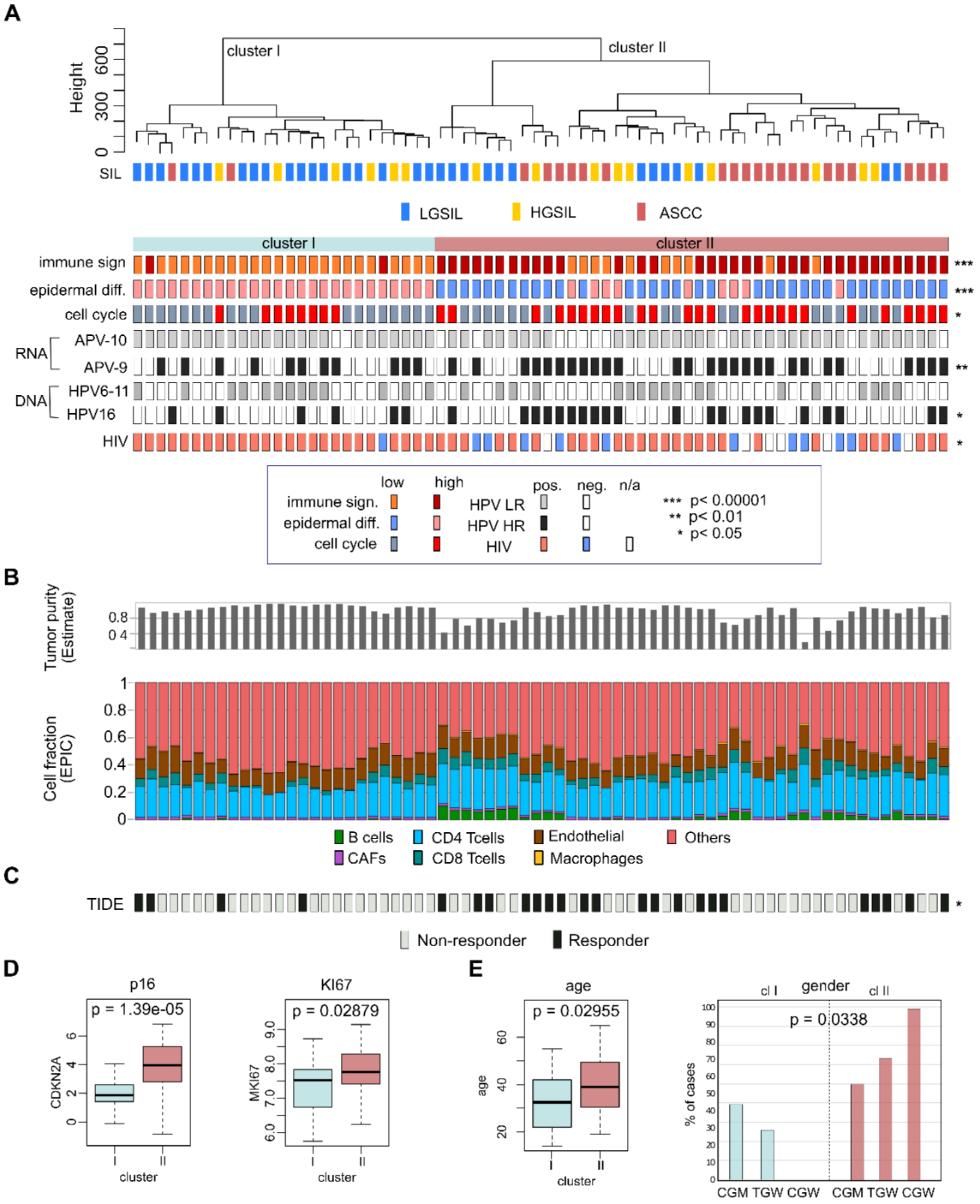


Figure 6: Integrative analysis of host transcriptome of LGSIL, HGSIL, and ASSC. A. Tile plot illustrating signatures scores, HPV status, and HIV status of samples distributed according to the unsupervised clustering analysis. Statistical significance was determined through the application of the Fisher exact test**. B.** Immune profiling and cell fraction composition for each sample using Estimate and Epic, respectively. **C.** T cell dysfunction and exclusion score for each sample. Statistical significance was determined through the application of the Fisher exact test **D.** Relative mRNA abundance of CDKN2A (p16) and MKI67 (Ki67) across samples in Cluster I versus Cluster II. **E.** Comparative analysis of clusters for age and gender. Statistical significance was determined through the application of a t-test for age and Chi-square test for gender. * p <0.05; ** p<0.01; *** p <0.001.

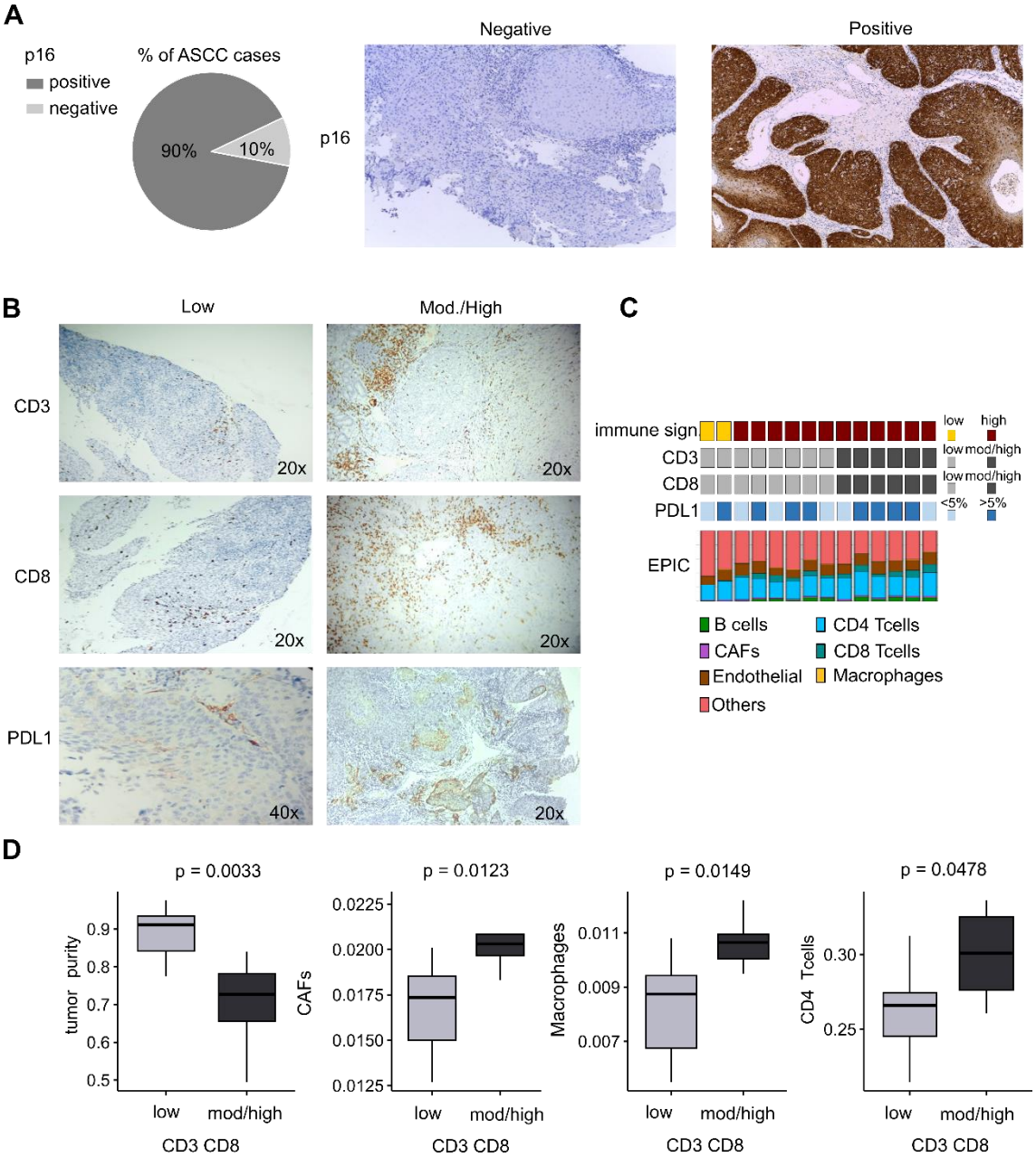


Figure 7: Comprehensive analysis of p16, CD3, CD8 TILs density, and PD-L1 expression in the tumor microenvironment of ASCC. A. Immunohistochemistry (IHC) results of p16 in 10 ASCC cases. Microphotographs represent negative and diffusely positive p16 staining on ASCC (10X) **Be.** Representative IHC results depicting high and low expression levels of CD3, CD8, and PD-L1. **C.** Tile plot illustrating ASCC samples analyzed by IHC, showcasing scores for immune signature, CD3, CD8, and PD-L1 IHC results, along with EPIC cell fractions. **D.** Box-Plots comparing tumor purity, CAFs, and macrophage levels, as obtained by EPIC, between tumors with low (n=8) and high (n=6) CD3/CD8 TILs. Statistical significance was calculated with the Wilcoxon signed-rank test.

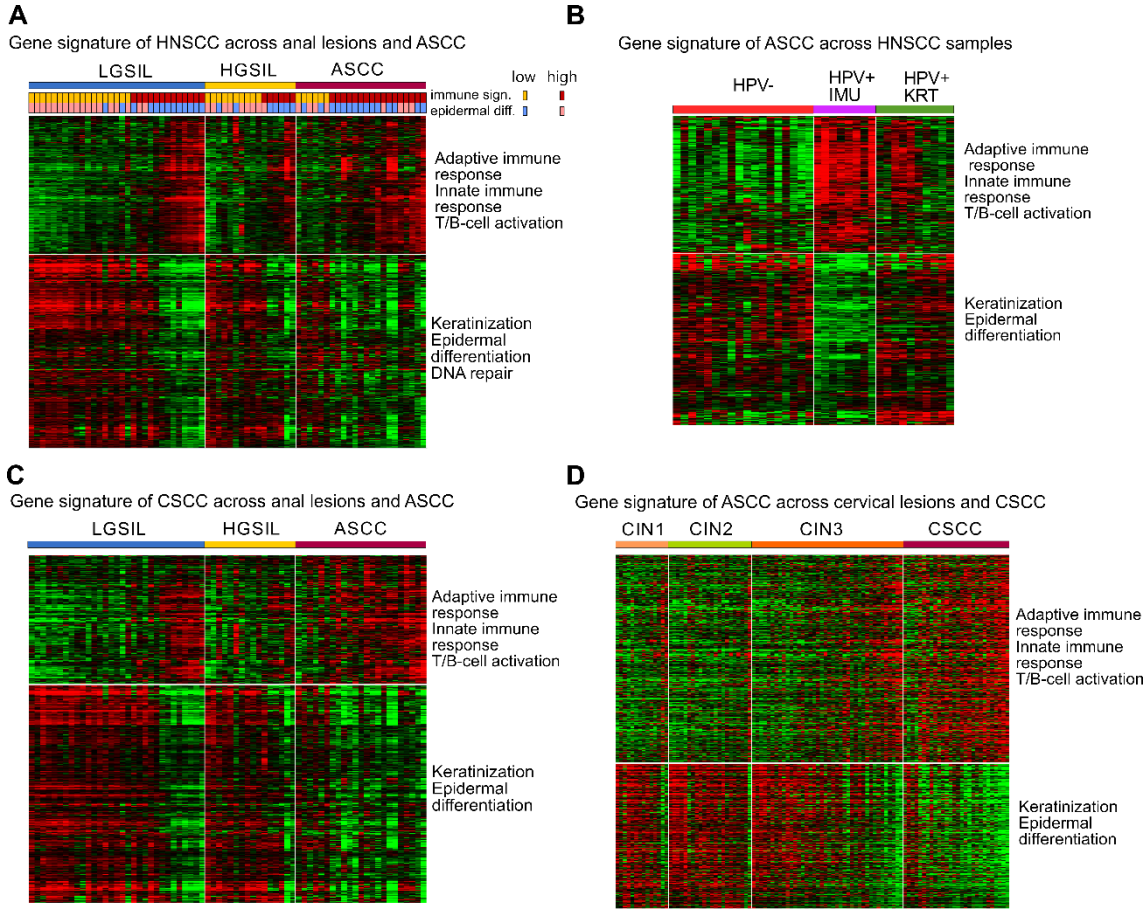


Figure 8: Comparative analysis of gene signature expression patterns and enriched pathways in HNSSC, cervical lesions and anal lesions. A. Heatmap visualization of HNSCC gene signature across our sample cohort, grouped by immune score within each diagnosis category. Additionally, the epidermal differentiation score is displayed. **B.** Heatmap visualization of the ASCC gene signature expression profile in HNSCC samples organized by subtype classification according to Zhang et al 2016 **C.** Heatmap visualization of CSCC gene signature across our sample cohort grouped by immune score within each diagnosis category. **D.** Heatmap visualization of the ASCC gene signature across cervical lesions, arranged in ascending order based on the immune gene profile within each diagnosis category.

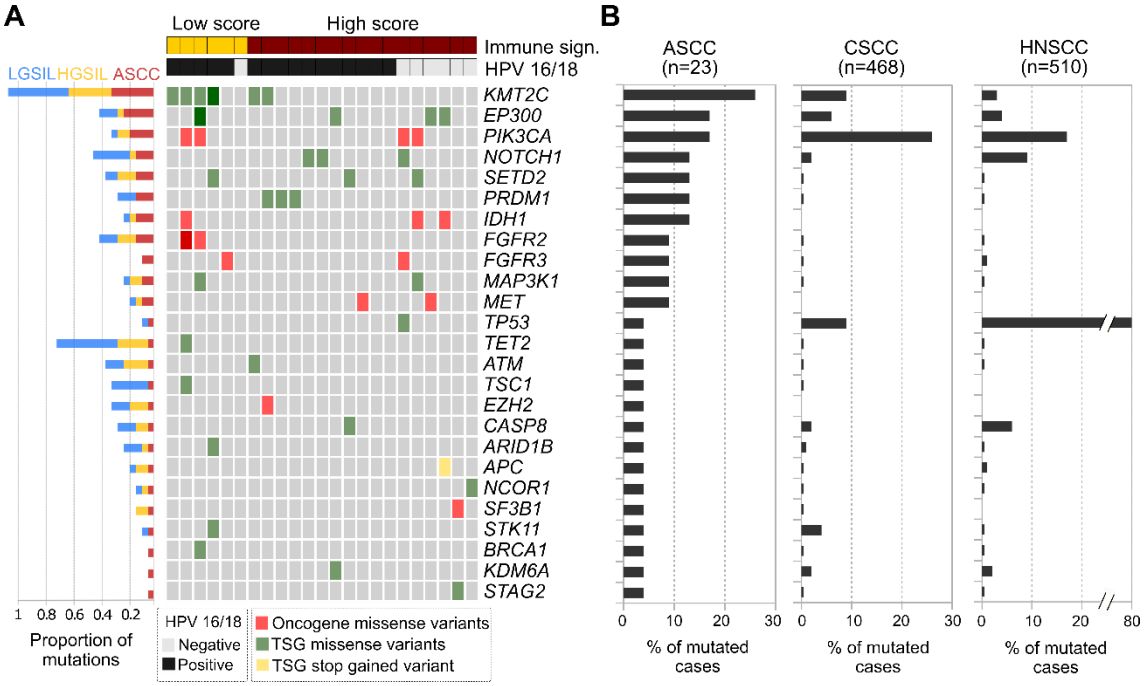


Figure 9: Mutational profiles among squamous cell carcinomas. A. Tile plot of the most prevalent somatic cancer driver mutations identified in 23 ASCC cases through transcriptome-based sequencing. The upper color-coded bars provide an indication of the immune signature score and HR-HPV status for each respective sample. On the left barplot, the proportions of somatic mutations within each group are presented, relative to the total number of cases in that specific group. TSG: Tumor Suppressor Gene. **B.** Comparative frequency of the mutations identified in the ASCC cohort with respect to CSCC and HNSCC retrieved from the TCGA cohorts.