

TRABAJO FIN DE GRADO
CURSO ACADÉMICO 2025-2026

**Identificación de biomarcadores para predecir
lesiones HSIL en pacientes con riesgo de
infección por HPV**

AUTORA: Aida Rico Torregrosa

TUTOR: Asia Fernández Carvajal

COTUTOR: Melissa Loreto Belló Pérez

Departamento y Área: Bioquímica y Biología Molecular

Agradecimientos

A la meua família, mamà i papà, gràcies per haver dedicat tot el vostre temps i esforç a proporcionar-me els mitjans necessaris perquè jo només haguera de preocupar-me d'estudiar. Gràcies per la vostra paciència infinita i per confiar en mi des del primer minut. No hi ha paraules per agrair-vos-ho tot, us estime infinit!

A la meua germana xicoteta, Inés. Sense tu al meu costat no ho haguera aconseguit. Has sigut el meu suport incondicional en tot moment, la que sempre ha estat per a donar-me ànims, celebrar-ho o fer-me una abraçada quan tocava plorar. Gràcies per no deixar de creure en mi ni un sol segon. Tant de bo ser un gran exemple per a tu, como tú ho eres per a mi. T'estime!

I als meus iaies, que sé que des del cel m'han acompanyat en cada pas i han vist com la seua neta xicoteta ha assolit la meta universitària. Tot açò també és vostre.

I com no, al meu grup d'amigues "Las muchachas", cadascuna venim d'un poble o fins tot d'una Comunitat distinta, però una mateixa vocació ens ha unit. Gràcies per saber ser i estar, amb vosaltres confirme la mítica frase de "Elije bien con quien sentarte el primer día de la Universidad". Quina sort haver-nos trobat, ser equip, companyeres de laboratori, de bus i principalment companyeres de vida. Els nostres noms sempre quedaran gravats en els treballs e informes del laboratori. Serem unes grans biotecnòlogues. Vos estime tants quilòmetres com d'aquí a Milà, Roma i Budapest!

No puc oblidar el meu millor amic Víctor, la persona que porta sent la meua mà dreta des de que vam anar junts a preescolar. A l'institut els nostres camins es van separar, però nosaltres no ho vam fer. Gràcies per continuar al meu costat, per passar etapes on sols em podies veure cada 3 mesos a la biblioteca, gràcies per escoltar-me queixar, i gràcies per fer-me creure que sóc capaç d'aconseguir tot allò que vullga amb esforç i constància, com tu fas.

I principalment gràcies a l'Aida del passat que va elegir la carrera més bonica que hi ha, gràcies per confiar en mi mateixa, per no rendir-me quan vaig trovar el primer suspens, per seguir posant colzes i ganes quan sols voliar plorar, per la meua constància i disciplina. Gràcies, Aideta, ets molt capaç!

CONTENIDO

Resumen.....	4
Palabras clave	5
GLOSARIO:.....	6
1. INTRODUCCIÓN	8
1.1 Infecciones por el Virus del Papiloma Humano (VPH) en personas que viven con VIH (PVV)...	8
1.2 Características del VPH	8
1.3. Factores de riesgo y poblaciones vulnerables al cáncer	10
1.4. Mecanismos de oncogénesis asociados al VPH	10
1.5. Cribado y diagnóstico de lesiones anales asociadas al VPH en PVV.....	12
1.6. Biomarcadores en patología anal	13
1.7. Aproximaciones transcriptómicas para la predicción de lesiones HSIL	13
2. OBJETIVOS	16
3. MATERIALES Y MÉTODOS.....	17
3.1 Diseño del estudio	17
3.2 Extracción de ARN	17
3.3. Retrotranscripción (RT) y reacción en cadena de la polimerasa (PCR)	17
3.4 Análisis estadístico	18
3.5 Anotación y caracterización estructural de las distintas isoformas de los genes <i>S100A14</i> y <i>KRT19</i>	18
4.RESULTADOS	19
4.1 Características de los participantes	19
4.2 Evaluación de la calidad de las librerías y clasificación de lecturas	19
4.3 Análisis del uso diferencial de isoformas del gen <i>S100A14</i>	22
4.4 Análisis y caracterización estructural de las isoformas del gen <i>S100A14</i>	23
4.5 Análisis del uso diferencial de las isoformas del gen <i>KRT19</i>	25
4.6 Análisis y caracterización estructural de las isoformas del gen <i>KRT19</i>	26
5. DISCUSIÓN.....	28
6. CONCLUSIÓN Y PERSPECTIVAS FUTURAS	30
7. BIBLIOGRAFÍA.....	32

Resumen

La búsqueda de biomarcadores precisos para predecir lesiones HSIL en pacientes con VPH continúa siendo un gran desafío. Las técnicas diagnósticas actuales presentan limitaciones relacionadas con su especificidad, coste y carácter invasivo. Con el objetivo de identificar alternativas moleculares no invasivas, en este trabajo se empleó un enfoque transcriptómico mediante secuenciación de lectura larga combinado con el análisis del uso diferencial de transcritos (DTU). Los resultados revelaron alteraciones significativas en el perfil de isoformas de los genes *S100A14* y *KRT19*. En el caso de *S100A14*, destacó la isoforma *S100A14-204* por alcanzar un valor significativo respecto a los otros grupos, la elevada presencia de la isoforma *S100A14-205* en el grupo HSIL, y el intrón retenido en *S100A14-206*. Además, se determinó que todas las variantes codificantes del gen convergen en la misma secuencia proteica de 104 aminoácidos, lo que sugiere que las diferencias estructurales se encuentran en las regiones UTRs, cuya desregulación corresponde a los mecanismos de control postranscripcional. Por otra parte, el gen *KRT19* presentaba un desplazamiento simétrico, recíproco y de dirección opuesta entre sus dos isoformas principales en pacientes con HSIL, lo que apunta a un cambio en el *splicing* alternativo asociado a la progresión de la lesión. En conjunto, estos hallazgos identifican por primera vez a *S100A14* y *KRT19* como genes candidatos de interés en las lesiones HSIL y demuestran la utilidad de la secuenciación de lectura larga para el descubrimiento de biomarcadores.

Abstract

The search for accurate biomarkers to predict HSIL lesions in HPV-infected patients remains a significant challenge. Current diagnostic techniques have limitations related to their specificity, cost, and invasive nature. To identify non-invasive molecular alternatives, this study employed a transcriptomic approach using long-read sequencing combined with differential transcript use (DTU) analysis. The results revealed significant alterations in the isoform profile of the *S100A14* and *KRT19* genes. For *S100A14*, the *S100A14-204* isoform stood out due to its significantly higher value compared to other groups, the high prevalence of the *S100A14-205* isoform in the HSIL group, and the retained intron in *S100A14-206*. Furthermore, it was determined that all coding variants of the gene converge on the same 104-amino-acid protein sequence, suggesting that the structural differences lie in the UTRs, whose dysregulation corresponds to post-transcriptional control mechanisms. Moreover, the *KRT19* gene exhibited a symmetric, reciprocal, and opposite-direction shift between its two main isoforms in patients with HSIL, pointing to a change in alternative splicing associated with lesion progression. Taken together, these findings identify *S100A14* and *KRT19* for the first time as candidate genes of interest in HSIL lesions and demonstrate the utility of long-read sequencing for biomarker discovery.

Palabras clave

Virus del Papiloma Humano, HSIL (*high-grade squamous intraepithelial lesion*): lesión escamosa intraepitelial de alto grado, *S100A14*, *KRT19*, secuenciación de lectura larga, uso diferencial de transcritos

Keywords

Human Papillomavirus (HPV), high-grade squamous intraepithelial lesion (HSIL), *S100A14*, *KRT19*, long read sequencing, differential transcript use (DTU)



GLOSARIO:

aa: aminoácidos.

AAR: anoscopia de alta resolución.

ADAR1 (*adenosine deaminase acting on RNA 1*): adenosina desaminasa que actúa sobre ARN 1.

ADN: ácido desoxirribonucleico.

ADNc: ADN complementario.

AR: alto riesgo.

ARNm: ARN mensajero.

ARNr: ARN ribosómico.

ASC-H (*atypical squamous cells-cannot exclude HSIL*): células escamosas atípicas en las que no se puede descartar HSIL.

ASC-US (*atypical squamous cells of undetermined significance*): células escamosas atípicas de significado indeterminado.

Best-q: valor de corte óptimo.

BH (*Benjamini-Hochberg*): método de ajuste de p-valores de Benjamini-Hochberg.

bp (*base pairs*): pares de bases.

BR: bajo riesgo.

CCE: carcinoma de células escamosas.

CDS (*coding sequence*): secuencia codificante.

CEIM: Comité de Ética de la Investigación con Medicamentos.

CRTA (*capture and reverse transcription*): adaptador de captura o transcripción inversa.

CTRL (*control*): control.

CVRS: calidad de vida relacionada con la salud.

DTU (*differential transcript usage*): uso diferencial de transcritos.

E (*early*): región temprana.

E6AP (*E6-associated protein*): ubiquitina-ligasa intracelular asociada a E6.

EPS: examen proctológico estándar.

HGUE: Hospital General Universitario de Elche.

HO/BI: homosexual/bisexual.

HSH: hombres que tienen sexo con hombres.

hTERT (*human telomerase reverse transcriptase*): transcriptasa inversa de la telomerasa humana.

HSIL (*high-grade squamous intraepithelial lesion*): lesión escamosa intraepitelial de alto grado.

HTSX: heterosexual.

IQR (*interquartile range*): rango intercuartílico.

KRT19: citoqueratina 19.

L (*late*): región tardía.

LCR (*long control region*): región larga de control.

LSIL (*low-grade squamous intraepithelial lesion*): lesión escamosa intraepitelial de bajo grado.

M: mediana.

MANE (*matched annotation from NCBI and EMBL-EBI*): anotación concordante de NCBI y EMBL-EBI.

miRNAs (*microRNAs*): microARN.

n: número de participantes.

NIA: neoplasia intraepitelial anal.

Novel: transcrito de *novus* (no anotado en bases de datos de referencia).

ONT (*Oxford Nanopore Technologies*): tecnología de secuenciación de Oxford Nanopore Technologies.

oligo-dT: cebador compuesto por una cadena corta de desoxitimidinas.

p16INK4a / p16: inhibidor de quinasas dependiente de ciclinas.

pRB: proteína del retinoblastoma.

p-valor: p-valor ajustado con el método de Benjamini-Hochberg.

PVV: personas que viven con el VIH.

qPCR (*quantitative polymerase chain reaction*): reacción en cadena de la polimerasa cuantitativa.

RNA-seq (*RNA sequencing*): secuenciación de ARN.

RT (*reverse transcription*): retrotranscripción o transcripción inversa.

SCJ (*squamous-columnar junction*): unión escamocolumnar.

SIL (*squamous intraepithelial lesion*): lesión escamosa intraepitelial.

SNG: secuenciación de nueva generación.

TAR: tratamiento antirretroviral.

Undef (*undefined*): indefinida.

UTRs (*untranslated regions*): regiones no traducidas.

VIH: virus de la inmunodeficiencia humana.

VLP (*virus-like particles*): partículas similares a virus.

VPH: virus del papiloma humano.

VPH-AR: virus del papiloma humano de alto riesgo.

Δ : delta mediana (diferencia entre la mediana del grupo HSIL y la mediana del grupo control).

1. INTRODUCCIÓN

1.1 Infecciones por el Virus del Papiloma Humano (VPH) en personas que viven con VIH (PVV)

El virus del Papiloma Humano (VPH) es la infección de transmisión sexual más prevalente a nivel mundial y el principal agente etiológico del cáncer anal, entre otros tumores anogenitales. En la población general, la mayoría de las infecciones por VPH son transitorias y asintomáticas, y se resuelven espontáneamente gracias a una respuesta inmunitaria eficaz. Sin embargo, en las personas que viven con VIH (PVV), la inmunosupresión asociada a la infección por el virus compromete el aclaramiento viral, favoreciendo la persistencia de la infección por VPH y su progresión hacia lesiones escamosas intraepiteliales de alto grado (HSIL) y precursoras de cáncer (Natale et al., 2024).

La interacción entre el VIH y el VPH es bidireccional. Por un lado, la infección por VPH puede modificar el microambiente mucosal genital y alterar su integridad, favoreciendo la entrada del VIH por una baja regulación de proteínas antimicrobianas (Liebenberg et al., 2019) (Dreyer, 2017). Por otro lado, la inmunodeficiencia inducida por el VIH, caracterizada por la depleción de linfocitos T CD4+, deteriora la vigilancia inmunitaria frente al VPH, aumentando el riesgo de infección persistente y de progresión a HSIL (Man et., al 2025). Esta progresión sigue una secuencia bien establecida: desde lesiones intraepiteliales escamosas de bajo grado (LSIL) hasta HSIL y carcinoma, en la que la integración del genoma del VPH en el ADN del huésped y la expresión de las oncoproteínas E6 y E7 desempeñan un papel central (Natale et al., 2024).

Pese a los grandes avances del tratamiento, actualmente no existe una cura para el VIH, las personas con este virus mantienen tratamientos antirretrovirales (TAR) de por vida (Tarrés-Freixas et al., 2025). No obstante, gracias a la introducción de la TAR en la década de 1990, la infección por VIH ha pasado de ser una enfermedad mortal a una crónica, lo que ha permitido un aumento significativo en la esperanza de vida de las PVV. Sin embargo, estas personas presentan con frecuencia comorbilidades asociadas que pueden afectar negativamente a su calidad de vida (García-López et al., 2020).

1.2 Características del VPH

Las partículas de VPH son pequeñas (60 nm de diámetro) y comprenden 72 capsómeros pentaméricos que encierran una cadena de ADN circular y de doble hebra (Grupo I según la

clasificación de Baltimore). Pertenecen a la familia *Papillomaviridae* y subfamilia *Firstpapillomaviridae*. Su genoma se organiza en 3 regiones funcionales: la región larga de control (LCR, *Long Control Region*), la región temprana (E, *Early*) y la región tardía (L, *Late*). La LCR es una región no codificante con función reguladora de la transcripción de los genes virales, especialmente de E6 y E7. La región E codifica las proteínas tempranas y no estructurales (E1, E2, E4, E5, E6 y E7) implicadas en la replicación del genoma viral y oncogénesis. Por último, la región L codifica dos proteínas estructurales de la cápside: L1 y L2 (**Figura 1A**). L1 es la proteína mayoritaria de la cápside y se autoensambla en partículas similares a virus (*virus-like particles*, VLPs); y L2, es la proteína minoritaria que desempeña un papel en el empaquetamiento del genoma viral (**Figura 1B**) (Williamson, 2023) (Beltrán Lissabet, 2014).

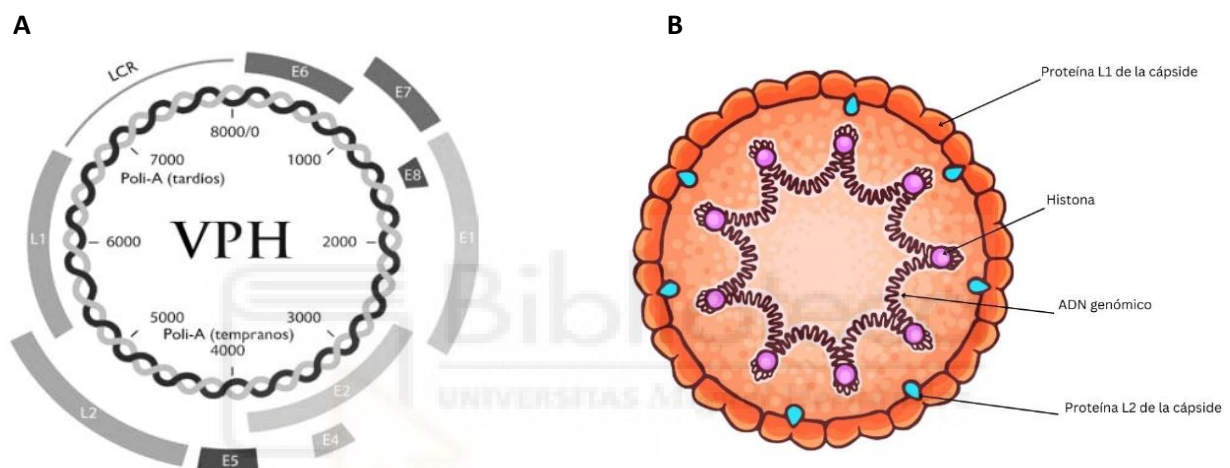


Figura 1. El panel representa la estructura y organización del VPH. El panel izquierdo (**A**) muestra la organización genómica del VPH. Se representan las tres regiones principales: la región larga de control (LCR), la región temprana (E) con sus genes E1-E8, y la región tardía (L) con las proteínas de la cápside L1 y L2. *Figura tomada de: Beltrán Lissabet, 2014.* El panel derecho (**B**) muestra la estructura de la partícula viral del VPH. Se representan los componentes principales: la cápside externa formada por las proteínas L1 y L2, y el núcleo interno constituido por el ADN genómico de doble cadena asociado a histonas. Imagen tomada de: Adaptado de IARC, 2023

Existen más de 150 genotipos de VPH, que se clasifican en dos grupos en función de su capacidad para producir lesiones cancerosas: los de alto riesgo (AR) u oncogénicos, que incluyen los genotipos 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 y 68 entre otros, y los de bajo riesgo (BR) o no oncogénicos, como los genotipos 6, 11, 40, 42, 43, 44, 54, 61, 70, 72 y 81. Entre los de AR, VPH-16 y VPH-18 son los genotipos más estudiados, ya que son responsables de aproximadamente el 70 % de los cánceres asociados a este virus (Williamson, 2023).

Específicamente, el VPH16 es el más prevalente y se asocia a carcinomas de células escamosas, con alta actividad replicativa y amplia distribución en cánceres cervicales; mientras que VPH18, tiene mayor probabilidad de integrarse en el genoma del huésped y se asocia preferentemente con adenocarcinomas (Xiao et al., 2025)

1.3. Factores de riesgo y poblaciones vulnerables al cáncer

Los grupos de mayor riesgo en los que se recomienda la detección de lesiones precancerosas anales son: 1) hombres que tienen sexo con hombres (HSH) que viven con VIH, 2) mujeres con antecedentes de HSIL o cáncer vulvar por VPH y 3) personas que se sometieron a un trasplante de órgano sólido hace más de 10 años. La tasa de progresión de HSIL a carcinoma de células escamosas (CCE) varía según la edad del paciente, la presencia de VPH16, el tamaño de las HSIL y el grado de inmunosupresión.

En PVV, la inmunodeficiencia, la inflamación crónica inducida por el virus y determinados factores como el consumo excesivo de alcohol y tabaco aumentan el riesgo de desarrollar neoplasias definitivas de SIDA, como el sarcoma de Kaposi (SK), el cáncer cervical invasivo y el linfoma no Hodgkin (LNH). Estas personas son también más susceptibles a coinfecciones con virus oncogénicos como el virus de Epstein-Barr (VEB), el herpes virus humano 8 (HHV-8), el VPH y los virus de la hepatitis B y C (VHB y VHC). A pesar de los avances con la TAR, estas neoplasias siguen diagnosticándose frecuentemente en estadios avanzados (Omar et al., 2024).

1.4. Mecanismos de oncogénesis asociados al VPH

La mayoría de las infecciones por VPH son transitorias y se resuelven espontáneamente en pocos meses sin consecuencias clínicas relevantes. Sin embargo, en determinados casos, especialmente en individuos inmunodeprimidos, la infección persistente puede favorecer la transformación celular. El desarrollo del cáncer asociado al VPH no forma parte del ciclo de replicación viral, sino que es la combinación de múltiples factores, que incluyen: la integración del ADN viral en el genoma del huésped, la alteración de la regulación génica celular, la activación de vías de señalización mediadas por proteínas virales y la inflamación crónica (Xiao et al., 2025)

Las proteínas virales E6 y E7 desempeñan un papel clave en este proceso, ya que interfieren con la función de genes supresores de tumores como p53, pRb, p21 y p16, promoviendo la proliferación celular descontrolada y la acumulación de alteraciones genéticas. Comprender los

mecanismos que actúan contra la oncogénesis ha proporcionado pistas para el desarrollo de nuevas estrategias terapéuticas: medicamentos antivirales, inhibidores de proteínas virales y vacunas preventivas, aunque el tratamiento de las neoplasias asociadas al VPH continúa siendo un desafío clínico. (Xiao et al., 2025).

La proteína E6 del VPH es uno de los principales determinantes de la capacidad oncogénica de los genotipos de alto riesgo. Su mecanismo de acción más relevante consiste en la unión a la ubiquitina-ligasa intracelular (E6AP), formando un complejo que recluta a la proteína supresora de tumores p53, promoviendo su ubiquitinación y degradación por el proteasoma. La pérdida de la función de p53 suprime la regulación de la apoptosis y favorece la acumulación de mutaciones que conducen al desarrollo de cáncer (Li et., al 2019) (**Figura 2B**). Además, E6 activa la telomerasa (hTERT), contribuyendo a la inmortalización celular (Xiao et al., 2025).

La oncoproteína E7 complementa la acción de E6 actuando sobre otra vía clave del control celular. E7 se une con alta afinidad a la proteína del retinoblastoma (pRb), la cual se encarga de mantener reprimidos los factores de transcripción E2F, frenando la progresión del ciclo celular. La unión de E7 inactiva a pRb y libera al factor de transcripción E2F, permitiendo que la célula avance directamente a la fase S donde ocurre una replicación descontrolada (Münger et al., 1989) (revista CENIC) (**Figura 2A**).

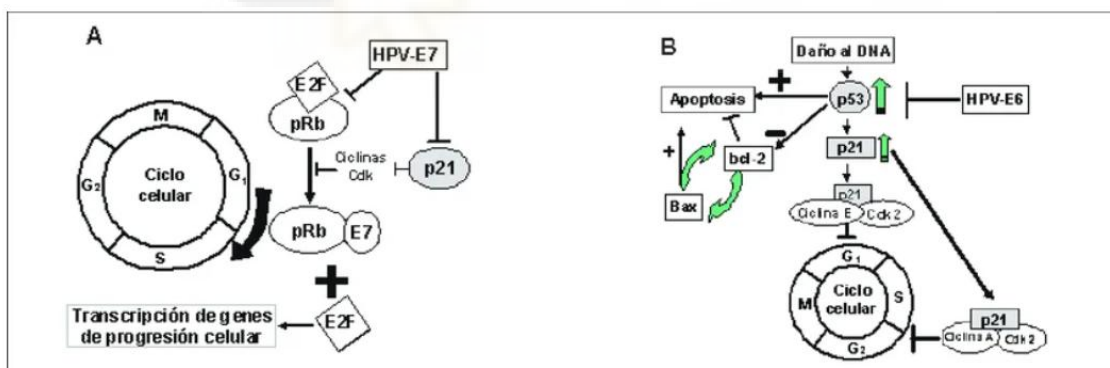


Figura 2. Mecanismos oncogénicos de las proteínas E6 y E7 de HPV. A) Efectos de la oncoproteína E7 sobre la regulación del ciclo celular, se muestra la liberación de E2F ocasionado por la unión de E7 con pRb. La interacción entre E7 y p21 libera la represión que ésta última ejerce sobre Cdk2 y ciclinas E y A. B) Efectos de la proteína E6 en la regulación del ciclo celular. La proteína E6 de HPV bloquea la apoptosis y permite la progresión celular al causar la degradación de p53 (revista CENIC)

La vacunación frente al VPH es un método de prevención primaria eficaz y costo-efectivo, con gran potencial de gran impacto para reducir la incidencia de las enfermedades inducidas por el VPH (Liebenberg et al., 2019).

1.5. Cribado y diagnóstico de lesiones anales asociadas al VPH en PVV

El cribado de lesiones anales en PVV ha cobrado gran importancia debido a su mayor riesgo de infección persistente por VPH y progresión a lesiones HSIL. En este contexto, el estudio ANCHOR (Palefsky, 2022) demostró que el tratamiento de las lesiones HSIL reduce en un 57% el riesgo de progresión a cáncer anal en esta población, lo que refuerza la importancia del diagnóstico precoz y el seguimiento sistemático. En PVV con síntomas anorrectales o como parte del seguimiento clínico, puede realizarse un examen proctológico estándar (EPS), que incluye inspección del margen anal, tacto anorrectal digital y anoscopia simple.

Aunque la EPS es generalmente bien tolerada, su rendimiento para detectar HSIL es deficiente al carecer de magnificación óptica y tinción con ácido acético, características propias de la anoscopia de alta resolución (AAR). Se comprobó que de 446 hombres que viven con VIH, no logró detectar ninguno de los 146 casos de HSIL (Spinder et al., 2023).

La citología anal constituye una de las principales herramientas de cribado en PVV, debido a su capacidad para detectar alteraciones celulares asociadas a infecciones por VPH. En esta población, se utiliza como prueba inicial de cribado. Sin embargo, presenta limitaciones importantes, como una especificidad moderada y una correlación variable con los hallazgos histológicos. Por ello, su mayor utilidad se observa cuando se combina con la detección de genotipos de AR de VPH y con la AAR. La AAR constituye la técnica de referencia para el diagnóstico confirmatorio de HSIL en PVV. Permite la visualización directa del epitelio anal y la toma de biopsias dirigidas, lo que mejora significativamente la precisión diagnóstica. Presenta una sensibilidad entre el 59% - 100% y una especificidad entre 66% - 74% (Sprinder et al., 2024) (Natale et al., 2024).

La detección de VPH-AR se basa en la detección del genoma viral por PCR en células extraídas de la zona de transición anal. La prueba para VPH-AR es extremadamente sensible para la detección de HSIL, con una sensibilidad del 90%. Por el contrario, su especificidad es baja (40-47% para mujeres y solo del 35% para HSH VIH+) (Sprinder et al., 2024) (Natale et al., 2024)

1.6. Biomarcadores en patología anal

Debido a las limitaciones de la citología y la imposibilidad de realizar de forma masiva la AAR, existe una necesidad urgente de biomarcadores objetivos que sirvan como herramientas de triaje en el cribado de lesiones anales asociadas al VPH (Cohen & Clarke, 2023) (Gaisa et al., 2021). En este contexto, en los últimos años se han evaluado varios biomarcadores relacionados con la infección por VPH, entre los que destacan el ARNm de las oncoproteínas E6/E7 y la proteína p16INK4a (p16) (**Tabla 1**), ambos asociados a procesos de transformación celular inducidos por el virus (Rollo et al., 2021).

A diferencia de la detección de ADN viral, que únicamente confirma la presencia del VPH, la detección de ARNm permite identificar infecciones activas pues refleja la expresión de las oncoproteínas E6 y E7 implicadas en la carcinogénicos por VPH.

Estudios comparativos han mostrado que las pruebas basadas en ARNm presentan una especificidad superior a la detección de ADN viral, especialmente si se limita a los cinco genotipos altamente oncogénicos (16, 18, 31, 33 y 45). Esta mayor especificidad permite mejorar la identificación de HSIL. Por el contrario, las técnicas basadas en ADN presentan mayor sensibilidad al incluir un mayor número de genotipos, aunque a costa de una menor especificidad y valor predictivo positivo. Por ello, se plantean estrategias combinadas que podrían optimizar la selección de pacientes candidatos a AAR y mejorar el diagnóstico precoz de lesiones HSIL (Castle et al., 2011).

La proteína p16INK4a es un inhibidor de las quinasas dependientes de ciclina cuya sobreexpresión se produce como consecuencia de la actividad de la oncoproteína viral E7 del VPH. Por este motivo, se considera un biomarcador indirecto de transformación celular asociada a infección persistente por VPH-AR. Su expresión se ha correlacionado con la presencia de HSIL, lo que confiere utilidad como herramienta complementaria en el diagnóstico de estas lesiones (Rollo et al., 2021).

1.7. Aproximaciones transcriptómicas para la predicción de lesiones HSIL

Las limitaciones de las técnicas diagnósticas convencionales, como la citología anal y la AAR, han impulsado el desarrollo de nuevas estrategias moleculares para la estratificación del riesgo de progresión a HSIL. En este contexto, las tecnologías ómicas, y en particular la transcriptómica, surge como una aproximación prometedora.

A diferencia de los tests basados en ADN, la detección de ARN mensajero (ARNm) permite evaluar el estado funcional de la infección. En concreto, se han propuesto ensayos dirigidos a los transcritos virales E6 y E7, cuya expresión persistente constituye un factor clave en la progresión hacia HSIL. La integración del genoma viral en el genoma del huésped, suelen interrumpir los genes virales E1 y E2, lo que provoca la sobreexpresión desregulada de E6 y E7. Estas oncoproteínas están implicadas en múltiples procesos característicos del cáncer, incluyendo la degradación de p53, la inactivación de pRb y la evasión de los puntos de control del ciclo celular. (Faillace Thiesen et al., 2025) (Ayass et al., 2026). Se ha observado que los niveles de expresión de E6/E7 son significativamente mayores en HSIL (80%) en comparación con LSIL (45%) y lesiones negativas, mientras que el marcador de infección productiva L1 (**Tabla 1**) presenta el patrón inverso, siendo más abundante en lesiones negativas y disminuyendo en LSIL y HSIL (Ayass et al., 2026).

Entre los biomarcadores emergentes destaca la Adenosina Desaminasa que actúa sobre el ARN 1 (ADAR1), una enzima que se encuentra sobreexpresada en las células cancerosas (**Tabla 1**). Su isoforma p150 es inducida por un interferón que facilita la progresión de las lesiones asociadas al VPH al modular la respuesta inmunitaria del huésped y suprimir la actividad del interferón. La cuantificación del ARNm de ADAR1 representa un método prometedor para mejorar la detección de estas lesiones y reducir procedimientos invasivos innecesarios (Bello-Pérez et al., 2025)

Tabla 1: Biomarcadores de patología anal y sus características

Biomarcador	Tipo	Mecanismo biológico	Relación con HSIL
ARNm de E6/E7	Biomarcador transcriptómico viral.	Se expresa durante la degradación de p53 e inactivación de pRb.	Mayor especificidad que el ADN. permite una mejor identificación de HSIL activas.
p16INK4a (p16)	Biomarcador proteico celular.	Se sobreexpresa cuando la oncoproteína E7 altera la vía pRb.	Indicador indirecto de transformación celular e infección persistente.
ARNm L1	Biomarcador transcriptómico viral.	Expresión de proteínas de la cápside cuando hay infección por VPH.	Su expresión disminuye progresivamente cuando la lesión llega a HSIL.
ADAR1 (ARNm)	Biomarcador transcriptómico celular.	La isoforma p150 (inducida por interferón) modula la respuesta inmunitaria y favorece la progresión tumoral.	Sobreexpresada en células cancerosas y asociada a lesiones HSIL.

Abreviaturas: ADAR1: Adenosina Desaminasa que actúa sobre ARN1; ADN: Ácido Desoxirribonucleico; ARNm: ácido ribonucleico mensajero; E6 y E7: proteínas tempranas 6 y 7 (*Early 6* y *Early 7*); HSIL: lesión Intraepitelial de Alto Grado; L1: proteína tardía 1; p150: isoforma 150 de la proteína ADAR1; p16INK4a: Inhibidor de quinasas dependiente de ciclinas; p53: proteína supresora de tumores; pRb: proteína del Retinoblastoma; VPH: Virus del Papiloma Humano.

Por otro lado, la secuenciación de ARN de célula única (scRNA-seq) permite analizar la heterogeneidad tumoral y el microambiente inmunológico, habiéndose observado diferencias significativas en las poblaciones linfocitarias entre tumores VPH positivos y negativos. (Su et al., 2025). No obstante, estas técnicas presentan limitaciones importantes para su traslación clínica, incluyendo el elevado coste, el reducido tamaño muestral de los estudios actuales y la complejidad técnica para su implementación en el contexto asistencia (Xia et al., 2025).

En este contexto, existe una necesidad urgente de identificar biomarcadores transcriptómicos accesibles y reproducibles que permitan predecir la presencia de HSIL a partir de muestras citológicas de rutina, evitando procedimientos invasivos y optimizando la selección de pacientes candidatos a AAR.



2. OBJETIVOS

El **objetivo principal** de este TFG es evaluar el perfil transcriptómico en exudados anales de pacientes con VIH para identificar biomarcadores asociados a la presencia de lesiones HSIL anales.

Para alcanzar este objetivo principal se establecieron los siguientes **objetivos secundarios**:

1. Caracterizar los perfiles de expresión génica en exudados anales en pacientes con y sin lesiones HSIL.
2. Identificar firmas transcriptómicas diferencialmente expresadas asociadas a HSIL.
3. Evaluar el potencial predictivo de los biomarcadores transcriptómicos identificados.
4. Analizar las vías moleculares y los procesos biológicos diferencialmente regulados entre pacientes con HSIL y sin HSIL.



3. MATERIALES Y MÉTODOS

3.1 Diseño del estudio

Este estudio prospectivo se llevó a cabo en el Centro de Salud Sexual del Hospital General Universitario de Elche (HGUE), en España. Los participantes que se incluyeron eran personas con VIH con resultados de citología anal alterada—incluyendo células escamosas atípicas de significado indeterminado (ASC-US), ASC-H (células escamosas atípicas en las que no se puede descartar HSIL), lesiones intraepiteliales de bajo grado (LSIL) y lesiones de alto grado (HSIL). Todos los participantes se sometieron a una AAR para la toma de un exudado anal y una biopsia. La biopsia fue analizada por el servicio de anatomía patológica del HGUE, para el diagnóstico de HSIL. Los criterios de exclusión para la AAR comprendían antecedentes de cirugía ano rectal reciente, presencia de proctitis activa, diagnóstico reciente de cáncer anal y embarazo.

Este estudio fue aprobado por el Comité de Ética de la Investigación con Medicamentos (2025/119). Todos los participantes firmaron el consentimiento informado.

3.2 Extracción de ARN

El ARN total de los exudados anales de cada participante se aisló utilizando el kit comercial RNeasy Mini Kit (QIAGEN, Alemania), siguiendo las instrucciones del fabricante. Brevemente, las muestras fueron lisadas y homogeneizadas. Posteriormente, se adicionó etanol en la misma proporción que de lisado y se cargó la muestra en la columna con membranas de sílice, aquí el ARN se unió a matriz. A continuación, el ARNm se purificó del ARN total con el Magnetic mRNA Isolation Kit (New England Biolabs, USA) para eliminar el ARN ribosómico. El kit emplea unas bolas magnéticas conjugadas con oligo-dT que hibridan con las colas poly-A del ARNm. El ARNm se cuantificó mediante el sistema fluorimétrico QuantiFluor® RNA System (Promega, Estados Unidos), que permite una cuantificación específica del ARN mediante la unión de un fluorocromo intercalante, siguiendo el protocolo *Low Standard* (LS) del fabricante.

3.3. Retrotranscripción (RT) y reacción en cadena de la polimerasa (PCR)

Con el objetivo de analizar el perfil transcriptómico de las muestras, se llevó a cabo la síntesis de ADNc mediante retrotranscripción a partir de ARNm purificado, siguiendo el protocolo del kit cDNA-PCR Sequencing V14-Barcoding (SQK-PCB114.24) (Oxford Nanopore Technologies (ONT), Reino Unido). Este enfoque permite la secuenciación de transcritos de longitud completa, lo que facilita la identificación y caracterización de isoformas de *splicing* alternativo.

Brevemente, se llevó a cabo el anillamiento del adaptador CRTA (*Capture and Reverse Transcription Adapter*) al ARNm molde. Este adaptador bicatenario presenta un saliente de poli(T) que hibrida específicamente con la cola poli(A) del extremo 3' del ARNm, garantizando la retrotranscripción de la longitud completa. A continuación, se sintetizó la primera cadena de ADNc usando la enzima *Maxima H Minus Reverse Transcriptase*.

Posteriormente, el ADNc obtenido se amplificó mediante PCR con la polimerasa KOD One PCR Master Mix (Sigma-Aldrich, Estados Unidos), utilizando cebadores específicos por muestra que incorporan códigos únicos en ambos extremos del amplicón. Este diseño permite el multiplexado simultáneo de hasta 24 librerías en un único ciclo de secuenciación Nanopore. Tras la amplificación, se añadió exonucleasa para degradar los cebadores no incorporados. Finalmente, se realizó una purificación con bolas magnéticas (AMPure XP), obteniendo la librería de ADNc final, lista para empezar el ciclo.

3.4 Análisis estadístico

Para el análisis estadístico del uso diferencial de transcritos (*Differential Transcript Usage*, DTU), las variables cuantitativas se expresaron mediante la mediana y el rango intercuartílico (IQR), mientras que las variables cualitativas se expresaron como frecuencias absolutas y porcentajes. La comparación entre el grupo control (sin HSIL) y el grupo HSIL se realizó mediante la prueba U de Mann-Whitney. Los p-valores se ajustaron por comparaciones múltiples mediante el método de Benjamini-Hochberg (BH), aplicado al conjunto de transcritos de cada gen. La delta mediana (Δ) se calculó como la diferencia entre la mediana del grupo HSIL y la mediana del grupo control, utilizando Microsoft Excel.

3.5 Anotación y caracterización estructural de las distintas isoformas de los genes *S100A14* y *KRT19*.

Con el fin de contextualizar funcionalmente los transcritos identificados, se llevó a cabo la anotación y caracterización estructural de las distintas isoformas de los genes *S100A14* y *KRT19*. Para ello, se consultaron las bases de datos de Uniprot, Ensembl Genome Browser y AlphaFold. A partir de dichas herramientas bioinformáticas se obtuvo información relativa a la secuencia proteica, la localización celular, la secuencia codificante (CDS), la longitud del transcrito y la estructura tridimensional predicha de la proteína funcional de cada isoforma.

4.RESULTADOS

4.1 Características de los participantes

Entre septiembre y noviembre de 2025, se recogieron exudados anales y biopsias de 7 pacientes con diagnóstico de infección por VPH. De acuerdo con los resultados de anatomía patológica, 4 presentaron HSL y 3 no presentaron HSIL. Todos los participantes fueron hombres, cuya mediana de edad era 55 años. La mayor parte de los sujetos procedían de España (85%) y todos ellos estaban vacunados frente al VPH. Las características basales de los participantes se encuentran en la **Tabla 2**.

Tabla 2. Características basales de los participantes

Características	Participantes con HSIL (n=4)	Participantes sin HSIL (n=3)	Total pacientes (n=7)	Valor p (n=7)
Edad, mediana (IQR)	47 (37,50 – 55,00)	57 (55,50- 60,00)	55 (42,00- 63,00)	0,114
Sexo, n (%)				No aplicable
Hombres	4 (100)	3(100)	7 (100)	
Mujeres	0	0	0	
Lugar de nacimiento				1,00
España	3 (75,00)	3(100)	6 (85,72)	
Otros	1 (25,00)	0	1 (14,28)	
Ruta de transmisión VIH, n (%)				1,00
HO/BI	3 (75,00)	3 (100)	6 (85,71)	
HTSX	1 (25,00)	0	1 (14,28)	
CD4 NADIR, células/μL, mediana (IQR)	350 (182- 358)	311 (310- 385,)	56,00 (310- 366)	0,857
Vacunación HPV (%)	100	33,30	71,42	0,143

Abreviaturas: %: porcentaje de expresión; HO/BI: homosexual/bisexual; HPV: Virus del Papiloma Humano; HTSX: heterosexual; IQR: rango intercuartílico; n: número de participantes.

4.2 Evaluación de la calidad de las librerías y clasificación de lecturas

La calidad de las librerías de secuenciación se evaluó mediante Pychopper, herramienta que utiliza la detección de cebadores como medida indirecta para identificar, orientar y validar la

integridad de los transcritos. Este análisis nos permite distinguir entre lecturas con cebadores detectados, lecturas rescatadas y lecturas no clasificables.

En la librería control, el porcentaje de lecturas con cebadores detectados fue del 92.6% (76.582 lecturas) mientras que en la librería HSIL fue del 93% (88.622 lecturas), en ambos casos el porcentaje de lecturas con cebadores supera el 75%, umbral a partir del cual la calidad de la librería se considera aceptable para análisis posteriores. En cambio, el porcentaje de lecturas rescatadas y de no clasificables fue inferior al 10% en ambas librerías (**Figura 3A y 3B**).

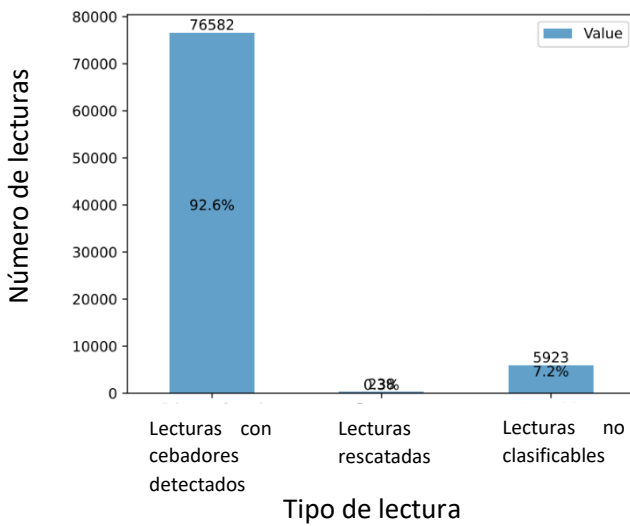
El análisis de la orientación de las lecturas mostró distribución equilibrada entre ambas hebras en las dos librerías. En la librería control se observó un 53,7 % de lecturas en orientación positiva (+) y un 46,3 % en orientación negativa (-); en la librería HSIL la distribución fue de 53,6 % y 46,4 %, respectivamente. La ausencia de sesgo indica que ninguna de las dos hebras fue amplificada preferentemente, reflejando una correcta construcción de la librería (**Figura 3C y 3D**).

Dentro de las lecturas no clasificables, la mayoría son no interpretables, es decir, sin señal de cebadores detectable. Concretamente, se registraron aproximadamente 120.000 lecturas con 0 hits (sin señal de cebadores) y menos de 10.000 con 2 hits (2 señales parecidas a cebadores). Esto quiere decir que el sistema discrimina bien entre señal y ruido de fondo (**Figura 3E y 3F**).

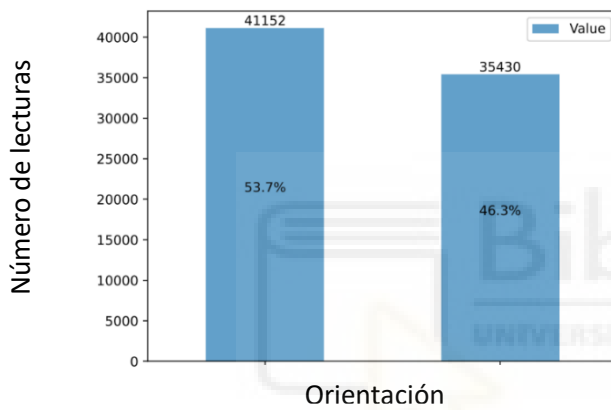
Por último, la métrica más relevante para la evaluación de la calidad es el valor de corte óptimo (*Best-q*), que determina el umbral a partir del cual el software maximiza la cantidad de bases útiles sin incorporar lecturas de baja calidad. En ambas librerías, el valor de *Best-q* fue de 0,5172, dentro del rango típico (~0,4–0,7) descrito para conjuntos de datos de ADNc de ONT (**Figura 3G y 3H**). Este resultado confirma que el filtrado de calidad aplicado es apropiado y que las lecturas seleccionadas para análisis son representativas y adecuadas para los análisis subsiguientes.

Librería Control

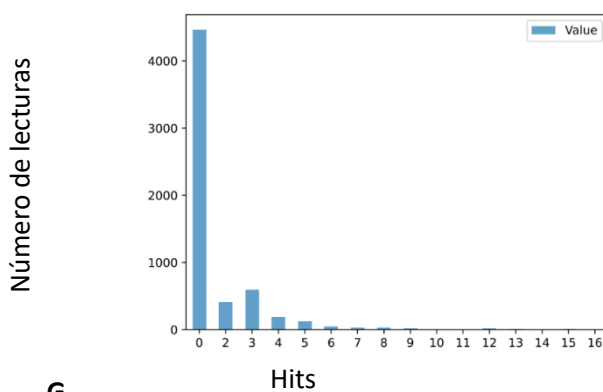
A Clasificación de las lecturas de salida



C Orientación de las lecturas

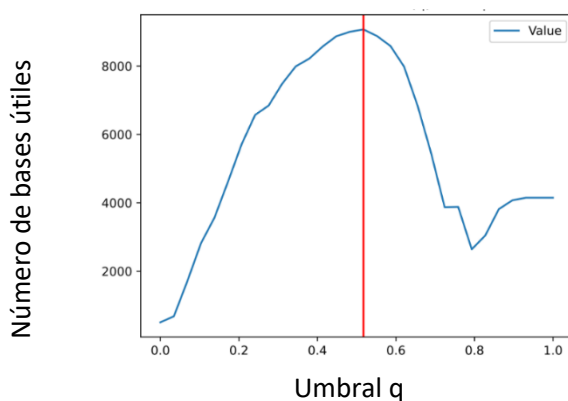


E Número de hits en lecturas no clasificables



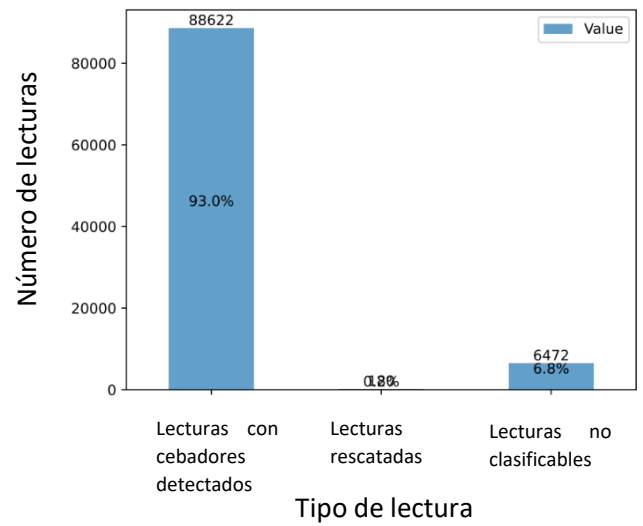
G

Bases útiles en función del umbral (q). Mejor 1=0.5172

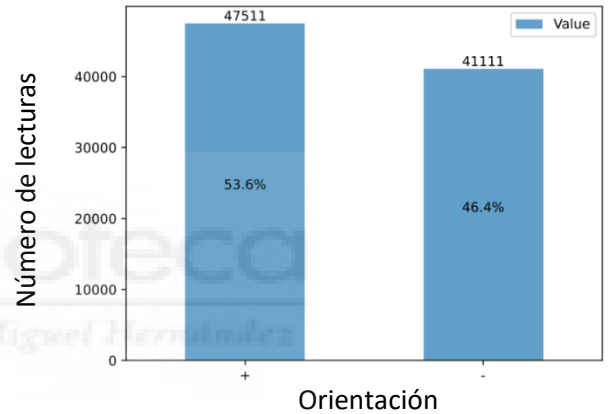


Librería HSIL

B Clasificación de las lecturas de salida

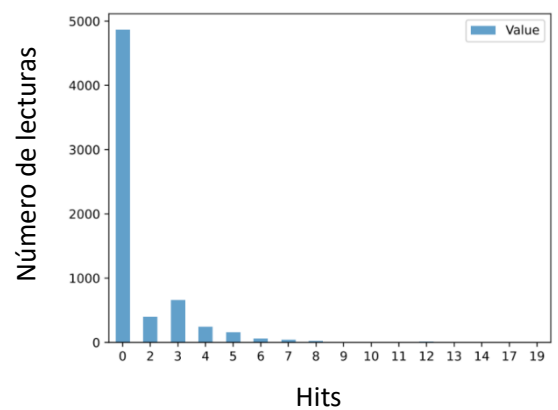


D Orientación de las lecturas



F

Número de hits en lecturas no clasificables



H

Bases útiles en función del umbral (q). Mejor q=0.5172

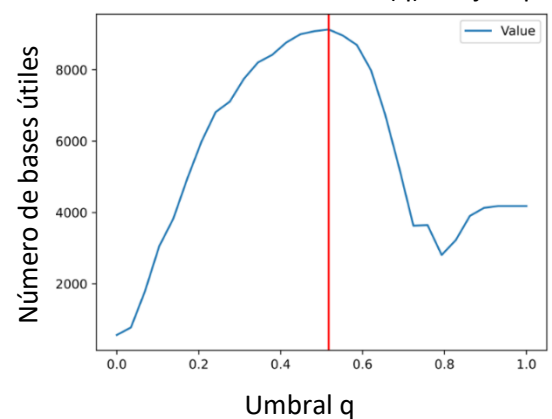


Figura 3. Control de calidad de las lecturas de ADNc procesadas mediante Pychopper. Los paneles de la columna izquierda (A, C, E y G) corresponden a las muestras control (sin HSIL) y los de la columna derecha (B, D, F, H) a las muestras HSIL. **A-B:** Distribución de las lecturas según su clasificación: lecturas con cebadores detectados, lecturas rescatadas y lecturas no clasificables. **C-D:** Orientación de las lecturas respecto a la secuencia original del ADNc; (+) indica orientación positiva y (-) orientación negativa. **E-F:** Número de hits de cebadores detectados en las lecturas no clasificables. **G-H:** Número de bases útiles en función del valor de corte óptimo (*Best-q*); la línea vertical roja indica el valor óptimo seleccionado.

4.3 Análisis del uso diferencial de isoformas del gen *S100A14*

El análisis DTU (*Differential Transcript Usage*) del gen *S100A14* no reveló diferencias estadísticamente significativas en la proporción relativa de sus isoformas entre pacientes con y sin HSIL, pues la significancia estadística tras la corrección de Benjamini-Hochberg estaba dentro del rango no significativo ($p\text{-adj} < 0.05$). A excepción del transcrito procesado *S100A14-204*, cuyo valor estadístico ($p = 0.0436$) indica una distribución diferencial significativa entre ambos grupos (**Tabla 3**).

De los seis transcritos analizados, cuatro de ellos corresponden a isoformas con biotipo codificante (*S100A14-201*, *-202*, *-203*, *-205*), uno es un transcrito procesado (*S100A14-204*) y otro presenta un intrón retenido (*S100A14-206*). En los pacientes control, las isoformas *201*, *202*, *203*, *204* presentaron medianas de uso relativo de entre 24,49 % y 24,63 %, con una variabilidad intragrupal muy baja ($IQR \leq 0,26$). En el grupo HSIL, sus medianas fueron inferiores, oscilando entre 22,22 % y 23,20 % ($IQR \leq 0,92$), lo que supone una disminución de entre 1,42 y 2,27 puntos porcentuales respecto al control (Δ mediana = -1,42 a -2,27).

Es destacable el comportamiento de la isoforma *S100A14-205*, un transcrito codificante de expresión minoritaria en condiciones basales. En el grupo control, su mediana de uso relativo fue de 1,36 % ($IQR = 0,76$), mientras que en el grupo HSIL ascendió a 5,95 % ($IQR = 1,49$), lo que supone un incremento de 4,59 puntos porcentuales (Δ mediana = +4,59). Aunque este cambio no alcanzó significación estadística, la magnitud del desplazamiento podría ser de interés biológico en el contexto de la transformación neoplásica.

Por último, el transcrito *S100A14-206*, clasificado como intrón retenido, mostró un aumento de 1,99 puntos en el grupo HSIL (mediana = 0,18 % en control frente a 2,17 % en HSIL; Δ mediana = +1,99) sin alcanzar significación estadística ($p = 0,2845$).

Tabla 3. Análisis del uso diferencial de las isoformas del gen *S100A14* entre pacientes sin y con HSIL, expresado en proporción a la expresión total del gen.

Transcrito	Biotipo/ Nombre	M. CTRL (%)	IQR CTRL	M. HSIL (%)	IQR HSIL	Δ Mediana	p-valor
ENST00000368702	S100A14-201 (codificante)	24.63	0.14	23.20	0.76	-1.42	0.0571
ENST00000344616	S100A14-202 (codificante)	24.63	0.14	23.20	0.76	-1.42	0.0571
ENST00000368701	S100A14-203 (codificante)	24.49	0.26	22.51	0.92	-1.98	0.0571
ENST00000476873	S100A14-204 (transcrito procesado)	24.49	0.26	22.22	0.48	-2.27	0.0436
ENST00000368700	S100A14-205 (codificante)	1.36	0.76	5.95	1.49	+4.59	0.0571
ENST00000469571	S100A14-206 (intrón retenido)	0.18	0.34	2.17	2.89	+1.99	0.2845

Abreviaturas: %, porcentaje sobre la expresión total del gen; Δ mediana, diferencia entre la mediana del grupo HSIL y la mediana del grupo control; CTRL, grupo control; IQR, rango intercuartílico; M., mediana; p-valor, p-valor ajustado mediante el método de Benjamini-Hochberg.

4.4 Análisis y caracterización estructural de las isoformas del gen *S100A14*

Para determinar la diversidad funcional del gen *S100A14*, se realizó una caracterización *in silico* de sus distintas isoformas. El análisis reveló la existencia de seis transcritos principales en *Homo sapiens*, cuyas características se resumen en la **Tabla 4**.

Cuatro de estos transcritos (S100A14-201, -203, -204 y -206) poseen capacidad codificante, dando lugar a una proteína de 104 aminoácidos, a pesar de las diferencias en la longitud de las pares de bases (bp) de los ARNm debido a las variaciones en las regiones no traducidas (UTRs). Aunque estas isoformas producen la misma proteína, las diferencias en las UTRs pueden tener consecuencias en la regulación postranscripcional. Los transcritos S100A14-202 y S100A14-205 se clasifican como biotipos no codificantes y no generan producto proteico definido; su presencia

podría tener un papel regulador a nivel postranscripcional. Todas las isoformas presentan localización citoplasmática.

Tabla 4. Análisis y caracterización estructural de las isoformas del gen *S100A14*.

Transcrito ID	Nombre	Bp	Biotipo	Proteína	Localización	Organismo
ENST00000368702	S100A14-201	1054	Proteína codificante	104aa	Citoplasma	<i>Homo sapiens</i>
ENST00000344616	S100A14-202	494	No proteína	—	Citoplasma	<i>Homo sapiens</i>
ENST00000368701	S100A14-203	104	Proteína codificante	104aa	Citoplasma	<i>Homo sapiens</i>
ENST00000476873	S100A14-204	1218	Proteína codificante	104aa	Citoplasma	<i>Homo sapiens</i>
ENST00000368700	S100A14-205	435	No proteína	—	Citoplasma	<i>Homo sapiens</i>
ENST00000469571	S100A14-206	1025	Proteína codificante	104aa	Citoplasma	<i>Homo sapiens</i>

Abreviaturas: aa, aminoácidos; bp: pares de bases.

La estructura tridimensional de la proteína fue obtenida mediante el servidor AlphaFold (**Figura 6**). El modelo revela una proteína compacta formada por cuatro hélices alfa principales organizadas en dos motivos *EF-hand*, estructura característica y conservada de la familia S100. El dominio *EF-hand* es el responsable de la unión a calcio (Ca^{2+}), que actúa como señal reguladora de la actividad de la proteína. A nivel funcional, S100A14 actúa como un modulador de p53 participando en la regulación de la proliferación, diferenciación y supervivencia celular.



Figura 4. Modelado estructural de la proteína S100A14 obtenido mediante AlphaFold. Se representan las hélices alfa características del dominio EF-hand. El código de colores indica la confianza de la predicción: azul oscuro (>90), muy alta; azul claro, alta (70-90); amarillo, baja (50-70); naranja, muy baja (<50). Las regiones de mayor confianza coinciden con el núcleo hidrofóbico formado por las hélices mientras que los extremos N y C-terminal muestran menor confianza, coherente con su naturaleza flexible.

4.5 Análisis del uso diferencial de las isoformas del gen *KRT19*

El análisis DTU del gen *KRT19* se realizó sobre dos transcritos identificados en los datos de secuenciación: la isoforma *KRT19-201*, designada como *MANE Select* que es codificante, y un transcrito *novel MSTRG.2898.1*, con secuencia codificante (CDS) no definida. Ninguno de los dos transcritos alcanzó significación estadística ($p = 0,0571$).

No obstante, cabe destacar la magnitud del cambio recíproco observado entre ambas isoformas. En el grupo control, la isoforma canónica *KRT19-201* presentó una mediana de uso relativo del 53,33 % (IQR= 1,40), mientras que en el grupo HSIL esta proporción ascendió al 75,86% (IQR= 1,71), lo que supone un incremento de 22,53 puntos porcentuales (Δ mediana= +22,53). De forma complementaria, el transcrito *MSTRG.2898.1* redujo su mediana del 46,67 % en el grupo control al 24,14 % en el grupo HSIL (Δ mediana= -22,53). Sus alteraciones son simétricas y de sentido opuesto, lo que confirma que se trata de un desplazamiento entre las dos isoformas dentro del mismo gen, sin cambios en la expresión génica total (**Tabla 5**).

Tabla 5. Análisis del uso diferencial de las isoformas del gen *KRT19* entre pacientes sin y con HSIL, expresado en proporción a la expresión total del gen.

Transcrito	Biotipo/ Nombre	M. CTRL (%)	IQR CTRL	M. HSIL (%)	IQR HSIL	Δ Mediana	p-valor
<i>ENST00000361566</i>	KRT19-201 MANE Select (coding)	53.33	1.40	75.86	1.71	+22.53	0.0571
<i>MSTRG.2898.1</i>	KRT19- novel (CDS undef.)	46.67	1.40	24.14	1.71	-22.53	0.0571

Abreviaturas: %, porcentaje sobre la expresión total del gen; Δ mediana, diferencia entre la mediana del grupo HSIL y la mediana del grupo control; CDS, *Coding Sequence* (secuencia codificante); *coding*, codificante; CTRL, grupo control; IQR, rango intercuartílico; M., mediana; MANE, *Matched Annotation from NCBI and EMBL-EBI* (anotación concordante de NCBI y EMBL-EBI); novel, transcrito de nuevo, no anotado en bases de datos de referencia; p-valor, p-valor ajustado mediante el método de Benjamini-Hochberg; undef, indefinida;

4.6 Análisis y caracterización estructural de las isoformas del gen *KRT19*

La caracterización *in silico* del gen *KRT19* se llevó a cabo mediante un análisis bioinformático de sus variantes de transcripción. Los datos obtenidos están en la **Tabla 6** y muestran la existencia de dos transcritos principales.

El transcrito ENST00000361566, designado como *MANE Select*, presenta un ARNm de 1576 bp que codifica una proteína de 403 aminoácidos correspondiente a la queratina de tipo I citoplasmática. Por otro lado, el transcrito *novel* MSTRG.2898.1, de 546 pb, no tiene CDS definida en las bases de datos de referencia, por lo que su potencial codificante y función biológica permanecen por determinar.

Tabla 6. Análisis y caracterización estructural de las isoformas del gen *KRT19*

Transcrito ID	Nombre	Bp	Biotipo	Proteína	Localización	Organismo
ENST00000361566	KRT19-201 <i>MANE Select</i>	1576	Proteína codificante	403aa	Citoplasma	<i>Homo Sapiens</i>
MSTRG.2898.1	KRT19-novel	546	No proteína	—	Citoplasma	<i>Homo Sapiens</i>

Abreviaturas: aa, aminoácido; bp, pares de bases; novel, de nuevo.

La estructura tridimensional de la proteína KRT19 fue predicha mediante AlphaFold (**Figura 5**). El modelo revela una arquitectura filamentosa organizada en torno a un dominio central de varillas formado por dos hélices alfa en disposición *coiled-coil*, estructura característica y conservada de todas las queratinas y filamentos intermedios del citoesqueleto. Este dominio es el responsable de la polimerización de la proteína en filamentos y presenta la mayor confianza de predicción. De especial relevancia es la ausencia del dominio de cola C-terminal, rasgo distintivo de KRT19 respecto al resto de queratinas de tipo I, que podría condicionar sus interacciones proteína-proteína y papel en la dinámica del citoesqueleto epitelial terminal. Su localización es citoplasmática y se integra en la red de filamentos intermedios del citoesqueleto.

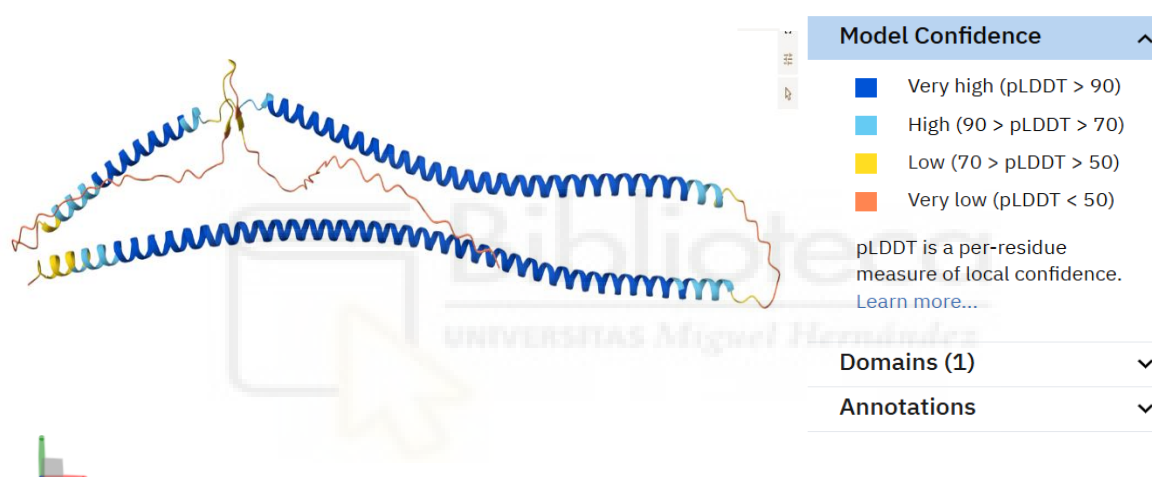


Figura 5. Modelado estructural de la proteína KRT19 obtenido mediante AlphaFold. Se representa una arquitectura filamentos organizada en torno al dominio central de varilla, formado por hélices alfa dispuestas en *coiled-coil*, estructura muy conservada en filamentos e imprescindible para la polimerización y el andamiaje citoplasmático. Ausencia del dominio de cola C-terminal. El código de colores indica la confianza de la predicción por residuos: azul oscuro, muy alta (>90); azul claro, alta (70-90); amarillo, baja (50-70); naranja, muy baja (<50), localizándose la máxima fiabilidad en el núcleo helicoidal de la proteína y en las regiones terminales una menor fiabilidad.

5. DISCUSIÓN

Hasta el momento las técnicas empleadas para la detección de HSIL en pacientes con VPH tienen escasa especificidad, exceso de procedimientos invasivos, elevado coste y malestar para el paciente (Chen et al., 2013). En este contexto, el presente estudio exploró el potencial de la secuenciación de lectura larga (ONT) combinada con análisis de uso diferencial de transcritos (DTU), con objetivo de identificar alteraciones transcriptómicas asociadas a la presencia de lesiones HSIL. Como resultado, se identificaron cambios en el uso relativo de isoformas de los genes *S100A14* y *KRT19*, lo que señala a estos genes como candidatos de interés para futuros estudios dirigidos a evaluar el valor biológico y diagnóstico de los eventos de *splicing* asociados a HSIL.

Cada vez hay más evidencia que sugiere la importancia de las proteínas de la familia S100 en la tumorigénesis y la progresión tumoral, siendo reguladoras clave en la progresión del ciclo celular, el crecimiento, la invasión y la migración celular (Wang et al., 2015). La proteína S100A14 pertenece a esta familia de proteínas de bajo peso molecular y contiene dominios y motivos estructurales especializados en la unión a calcio que desempeñan un papel fundamental en la regulación de múltiples procesos celulares. Su expresión es altamente heterogénea entre los distintos tejidos humanos, lo que sugiere que no presenta una función universal, sino que su función depende del tejido y contexto biológico específico (Basnet et al., 2019).

Asimismo, S100A14 se ha identificado como modulador de la diferenciación de queratinocitos en epitelio escamoso, un proceso esencial para el mantenimiento de la homeostasis epitelial y cuya alteración se ha asociado al desarrollo y progresión de diversas neoplasias escamosas (Chen et al., 2013). Sin embargo, la función biológica de S100A14 parece depender en gran medida del contexto tisular. Estudios previos han descrito comportamientos opuestos de esta proteína en tumores de origen epiteliales y mesenquimal, actuando en algunos casos como supresor tumoral y en otros favoreciendo la progresión tumoral. Aunque los mecanismos responsables de estas diferencias no se conocen completamente, se ha propuesto que podría estar relacionado con las proteínas efectoras con las que interactúa y las vías de señalización activas en cada tejido (Basnet et al., 2019).

En el presente estudio, el gen *S100A14* no reveló diferencias globales significativas en el uso de sus isoformas entre participantes con y sin HSIL, con la excepción del transcrito *S100A14-204* ($p=0,0436$), que sí presentó una diferencia significativa entre grupos. Además, se observó un aumento de más del cuádruple de la isoforma codificante *S100A14-205* en el grupo HSIL. Aunque este último cambio no fue estadísticamente significativo debido al número de muestras, nos indica una posible tendencia a un mayor uso de esta isoforma en el contexto de HSIL. Por último, la

isoforma *S100A14-206*, que presenta un intrón retenido, aumentó casi el doble en el grupo HSIL. Este cambio podría sugerir una alteración en el proceso de *splicing*, ya que la retención de intrones puede dar lugar a transcritos menos estables o potencialmente no funcionales. Sin embargo, ninguno de estos cambios alcanzó significación estadística, probablemente debido al tamaño muestral reducido. Un hallazgo de especial relevancia es que todas las variantes codificantes convergen en la misma secuencia de 104 aa. Por tanto, las diferencias observadas se localizan en las regiones no traducidas (UTRs). Esto sugiere que podrían existir cambios en los mecanismos de regulación postranscripcional, como la unión de micro ARNs o proteínas que interactúan con el ARN, alterando la estabilidad del ARNm o su eficiencia de traducción, sin necesidad de modificar los niveles de transcripción del gen.

KRT19 (citoqueratina 19) es una proteína de los filamentos intermedios, expresada principalmente en células epiteliales, donde desempeña un papel estructural como citoesqueleto y participa en proceso de diferenciación epitelial y señalización celular (Saha et al., 2018). Su estructura destaca por la ausencia del dominio de cola C-terminal, lo que le confiere alta flexibilidad en el ensamblaje de los filamentos intermedios (Moll et al., 2008).

Al igual que *S100A14*, la función de *KRT19* parece depender del tipo de tejido y del contexto tumoral. Se ha descrito que participa en procesos relacionados con la progresión tumoral en distintos tipos de cáncer, aunque sus efectos pueden variar según el órgano de origen. Por ello, su papel biológico todavía no se comprende completamente y parece estar condicionado por las características específicas de cada tejido (Saha et al., 2018).

En el contexto de las lesiones intraepiteliales escamosas asociadas a VPH, diversos estudios sugieren una posible implicación de *KRT19* en la carcinogénesis escamosa. Se ha observado su expresión en la unión escamocolumnar (SCJ) del cérvix, una región especialmente susceptible a la infección por VPH de AR. Su expresión fue positiva en la totalidad de los casos de CIN3 y carcinoma escamoso cervical analizados, lo que apunta a su posible implicación en la carcinogénesis cervical mediada por VPH de AR (Santoro et al., 2025).

También se ha descrito que las oncoproteínas virales E6 y E7 de los VPH de alto riesgo inducen la inmortalización de queratinocitos y alteran el patrón de expresión de citoqueratinas en epitelio escamoso estratificado, incluyendo la desregulación de *KRT19* favoreciendo así el proceso escalonado que conduce al carcinoma escamoso (Serrano-Villar et al., 2017).

Sin embargo, hasta donde alcanza nuestro conocimiento, no existían estudios que hubieran evaluado específicamente los cambios de isoformas de KRT19 en las lesiones anales HSIL mediante el uso diferencial de transcritos.

En nuestro estudio, el análisis DTU del gen *KRT19* identificó dos isoformas principales: la canónica KRT19-201 y el transcrito de *nov*, MSTRG.2898.1, con secuencia codificante no definida. Los resultados revelaron un cambio en el uso relativo de ambas isoformas entre los grupos estudiados. En los pacientes HSIL, la proporción de KRT19-201 aumentó su uso relativo del 53,33% al 75,86%, mientras que la de MSTRG.2898.1 disminuyó del 46,67% al 24,14% ($p = 0,0571$).

Aunque esta diferencia no alcanzó significación estadística, los resultados sugieren una tendencia hacia un mayor uso de la isoforma canónica en las lesiones HSIL. Este patrón podría reflejar cambios en los mecanismos de *splicing* asociados a la progresión de la lesión, más que alteraciones en la expresión global del gen.

La relevancia funcional de este hallazgo todavía no está clara. Sin embargo, dado que KRT19 participa en la organización estructural del epitelio y que su expresión se encuentra alterada en distintos procesos neoplásicos asociados al VPH, este cambio de isoformas podría relacionarse con alteraciones celulares que acompañan al desarrollo de HSIL.

En conjunto, estos hallazgos implican por primera vez a *KRT19* como un gen de interés para futuros estudios sobre *splicing* alternativo en lesiones anales asociadas al VPH. La interpretación de estos resultados debe realizarse teniendo en cuenta las características de la cohorte estudiada. Aunque el tamaño fue reducido ($n=7$), los grupos presentaron niveles de CD4 nadir comparables, lo que podría haber reducido el efecto de la inmunosupresión como factor de confusión sobre los patrones de *splicing* observados. La validación en cohortes de mayor tamaño será necesaria para determinar la reproducibilidad de este hallazgo y su posible utilidad como biomarcador.

6. CONCLUSIÓN Y PERSPECTIVAS FUTURAS

En conclusión, los principales hallazgos de este trabajo son:

1. La combinación de la secuenciación de lectura larga (ONT) y el análisis de uso diferencial de transcritos (DTU) demostró ser una aproximación metodológica viable y eficaz para identificar alteraciones transcriptómicas asociadas a la presencia de lesiones HSIL en pacientes con VPH.
2. El análisis molecular identificó, por primera vez en el contexto de lesiones anales HSIL, alteraciones en el perfil de isoformas de los genes *S100A14* y *KRT19*, consolidándolos como potenciales candidatos de interés biológico y diagnóstico.

3. En *S100A14* se observaron cambios compatibles con mecanismos de regulación postranscripcional.
4. En *KRT19* se detectó un cambio coordinado en el uso relativo de sus principales isoformas, sugiriendo una posible alteración en el splicing alternativo asociada a la progresión de la lesión.

El primer paso para consolidar estos resultados sería su validación en cohortes de mayor tamaño, así como en muestras de tejido obtenidas mediante biopsia. Esta aproximación permitirá determinar si las alteraciones en el uso de isoformas observadas en los exudados relejan los cambios transcriptómicos presentes en el tejido lesionado, reforzando el potencial de los exudados como fuente de biomarcadores no invasivos para la detección de HSIL.

Una vez confirmada esta asociación, estudios futuros podrán centrarse en entender mejor los mecanismos de regulación postranscripcional propuestos para *S100A14*. Sería de gran interés explorar *in silico* las regiones UTR de sus transcritos, mediante herramientas bioinformáticas de predicción. Se podrían identificar sitios de unión para microARNs (miRNAs) o proteínas de unión a ARN (RBPs) específicos que modulen la estabilidad y la eficiencia de traducción del ARNm en el microambiente de las lesiones HSIL. Una vez identificadas estas dianas, se confirmaría experimentalmente su interacción funcional mediante ensayos de gen reportero de luciferasa.

Por otro lado, los cambios observados en *KRT19* justifican la evaluación del posible papel de las oncoproteínas virales E6 y E7 de los genotipos de VPH-AR sobre la maquinaria de *splicing* celular y comprender si la transición de la isoforma *MSTRG.2898.1* a la canónica *KRT19-201* es una consecuencia de la expresión de estas oncoproteínas. Para ello, podemos emplear modelos celulares de queratinocitos anales inmortalizados donde se silencien o sobreexpresen selectivamente dichos factores virales.

En conjunto, estas líneas de investigación futuras permitirán profundizar en el conocimiento sobre los eventos de *splicing* alternativo asociados a HSIL, mejorando nuestra comprensión de la patogénesis molecular de las lesiones anales inducidas por VPH y contribuyendo al desarrollo de nuevos biomarcadores de diagnóstico no invasivos y de alta especificidad.

7. BIBLIOGRAFÍA

1. Grünberger, F., Ferreira-Cerca, S., & Grohmann, D. (2022). Nanopore sequencing of RNA and cDNA molecules in *Escherichia coli*. *RNA*, 28(3), 400–417. <https://doi.org/10.1261/rna.078937.121>
2. Saha, S. K., Kim, K., Yang, G. M., Choi, H. Y., & Cho, S. G. (2018). Cytokeratin 19 (KRT19) has a dual role as a tumor suppressor and an oncogene in various cancers. *International Journal of Molecular Sciences*, 19(5), Artículo 1423. <https://doi.org/10.3390/ijms19051423>
3. Moll, R., Divo, M., & Langbein, L. (2008). The human keratins: Biology and pathology. *Histochemistry and Cell Biology*, 129(6), 705–733. <https://doi.org/10.1007/s00418-008-0435-6>
4. Chen, H., Ma, J., Sunkel, B., Luo, A., Ding, F., Li, Y., He, H., Zhang, S., Xu, C., Jin, Q., Wang, Q., & Liu, Z. (2013). S100A14 is a novel modulator of terminal differentiation of esophageal squamous cell carcinoma. *Molecular Cancer Research*, 11(12), 1542–1553. <https://doi.org/10.1158/1541-7786.MCR-13-0317>
5. Wang, X., Yang, J., Qian, J., Liu, Z., Chen, H., & Cui, Z. (2015). S100A14, a mediator of epithelial-mesenchymal transition, regulates proliferation, migration and invasion of human cervical cancer cells. *American Journal of Cancer Research*, 5(4), 1484–1495.
6. Santoro, A., Pannone, G., Ninivaggi, R., Petrucci, M., Santarelli, A., Russo, G. M., Lepore, S., Pietrafesa, M., Laurenzana, I., Leonardi, R., Bucci, P., Natalicchio, M. I., Lucchese, A., Papagerakis, S., & Bufo, P. (2015). Relationship between CK19 expression, deregulation of normal keratinocyte differentiation pattern and high-risk human papillomavirus infection in oral and oropharyngeal squamous cell carcinoma. *Infectious Agents and Cancer*, 10, 46. <https://doi.org/10.1186/s13027-015-0041-x>
7. Saha, S. K., Kim, K., Yang, G.-M., Choi, H. Y., & Cho, S.-G. (2018). Cytokeratin 19 (KRT19) has a role in the reprogramming of cancer stem cell-like cells to less aggressive and more drug-sensitive cells. *International Journal of Molecular Sciences*, 19(5), 1423. <https://doi.org/10.3390/ijms19051423>
8. Serrano-Villar, S., Hernández-Novoa, B., de Benito, A., del Romero, J., Ocampo, A., Blanco, J. R., Masiá, M., Sendagorta, E., Sanz, G., Moreno, S., & Pérez-Molina, J. A. (2017). Screening for precancerous anal lesions with P16/Ki67 immunostaining in HIV-infected MSM. *PLOS ONE*, 12(11), e0188851. <https://doi.org/10.1371/journal.pone.0188851>
9. Wang, X., Yang, J., Qian, J., Liu, Z., Chen, H., & Cui, Z. (2015). S100A14, a mediator of epithelial–mesenchymal transition, regulates proliferation, migration and invasion of human cervical cancer cells. *American Journal of Cancer Research*, 5(4), 1484–1495.

10. Chen, H., Ma, J., Sunkel, B., Luo, A., Ding, F., Li, Y., He, H., Zhang, S., Xu, C., Jin, Q., Wang, Q., & Liu, Z. (2013). S100A14: A novel modulator of terminal differentiation in esophageal squamous cell carcinoma. *Molecular Cancer Research*, 11(12), 1542–1553. <https://doi.org/10.1158/1541-7786.MCR-13-0317>
11. Basnet, S., Sharma, S., Costea, D. E., & Sapkota, D. (2019). S100A14 expression profile and function in human cancer. *Oncotarget*, 10(31), 2996–3012. <https://doi.org/10.18632/oncotarget.26861>
12. Natale, A., Brunetti, T., Orioni, G., & Gaspari, V. (2024). Screening of anal HPV precancerous lesions: A review after last recommendations. *Journal of Clinical Medicine*, 13, 5246. <https://doi.org/10.3390/jcm13175246>
13. Dreyer, G. (2017). Clinical implications of the interaction between HPV and HIV infections. *Best Practice & Research Clinical Obstetrics & Gynaecology*. <https://doi.org/10.1016/j.bpobgyn.2017.08.011>
14. Beltrán Lissabet, J. F., (2014). Aspectos generales sobre la estructura y función de las proteínas codificadas por el virus del Papiloma Humano. *Revista CENIC. Ciencias Biológicas*, 45(2), 108-118.
15. Man, S., Jansen, J., Kroeze, S., Geijtenbeek, T. B. H., & Kootstra, N. A. (2025). Transcriptomic HIV-1 reservoir profiling reveals a role for mitochondrial functionality in HIV-1 latency. *PLOS Pathogens*, 21(1), e1012822. <https://doi.org/10.1371/journal.ppat.1012822> Williamson, A.-L. (2023). Recent developments in human papillomavirus (HPV) vaccinology. *Viruses*, 15, 1440. <https://doi.org/10.3390/v15071440>
16. Tarrés-Freixas, F., Trinité, B., Carrillo, J., & Blanco, J. (2025). Under the radar: Immune evasion, pathogenesis and control of HIV infection. *International Journal of Molecular Sciences*, 26, 11381. <https://doi.org/10.3390/ijms262311381>
17. García-López, Y. L., Bernal-Soriano, M. C., Torrús-Tendero, D., Delgado de los Reyes, J. A., & Castejón-Bolea, R. (2020). Factores relacionados con la calidad de vida en personas que viven con el VIH en Alicante, España. *Enfermedades Infecciosas y Microbiología Clínica*. <https://doi.org/10.1016/j.eimc.2020.03.005>
18. Münger, K., Phelps, W. C., Bubb, V., Howley, P. M., & Schlegel, R. (1989). The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *Journal of Virology*, 63(10), 4417–4421. <https://doi.org/10.1128/JVI.63.10.4417-4421.1989>
19. Liebenberg, L. J. P., McKinnon, L. R., Yende-Zuma, N., Garrett, N., Baxter, C., Mansoor, L. E., Kharsany, A. B. M., Archary, D., Rositch, A., Samsunder, N., Passmore, J.-A. S., Abdool Karim, S. S., & Abdool Karim, Q. (2019). HPV infection and the genital cytokine milieu in women at

- high risk of HIV acquisition. *Nature Communications*, 10, 5227. <https://doi.org/10.1038/s41467-019-13089-2>
- Omar, A., Marques, N., & Crawford, N. (2024). Cancer and HIV: The molecular mechanisms of the deadly duo. *Cancers*, 16, 546. <https://doi.org/10.3390/cancers16030546>
20. Williamson, A.-L. (2023). Recent developments in human papillomavirus (HPV) vaccinology. *Viruses*, 15, 1440. <https://doi.org/10.3390/v15071440>
21. Xiao, Q., Liu, Y., Li, T., Wang, C., He, S., Zhai, L., Yang, Z., Zhang, X., Wu, Y., & Liu, Y. (2025). Viral oncogenesis in cancer: From mechanisms to therapeutics. *Signal Transduction and Targeted Therapy*, 10, 151. <https://doi.org/10.1038/s41392-025-02197-9>
22. Natale, A., Brunetti, T., Orioni, G., & Gaspari, V. (2024). Screening of anal HPV precancerous lesions: A review after last recommendations. *Journal of Clinical Medicine*, 13, 5246. <https://doi.org/10.3390/jcm13175246>
23. Spindler, L., Etienney, I., Abramowitz, L., de Parades, V., Pigot, F., Siproudhis, L., Adam, J., Balzano, V., Bouchard, D., Bouta, N., Bucau, M., Carlo, A., Chanal, J., Charpentier, C., Clifford, G., Draullette, M., Fathallah, N., Ferré, V., Fléjou, J.-F., Fouéré, S., Higuero, T., Kassouri, L., Kurt, S., Laurain, A., Leclerc, E., Lepiller, Q., Lesage, A.-C., Mège, D., Ménard, A., Merle, P., Mortreux, P., Noël, C., Péré, H., Prétet, J.-L., Roland, D., Staumont, G., Tracanelli, L., Vuitton, L., Wylomanski, S., Zaegel-Faucher, O., & Société Nationale Française de Colo-Proctologie. (2024). Screening for precancerous anal lesions linked to human papillomaviruses: French recommendations for clinical practice. *Techniques in Coloproctology*, 28, 23. <https://doi.org/10.1007/s10151-023-02899-8>
24. Rollo, F., Latini, A., Giuliani, M., Giglio, A., Donà, M. G., & Benevolo, M. (2021). Evaluation of HPV-related biomarkers in anal cytological samples from HIV-uninfected and HIV-infected MSM. *Pathogens*, 10(7), 888. <https://doi.org/10.3390/pathogens10070888>
25. Gaisa, M. M., Sigel, K. M., Deshmukh, A. A., Lenskaya, V., Chan, C. A., Silvera, R., Winters, J., & Liu, Y. (2021). Comparing anal cancer screening algorithms using cytology and human papillomavirus DNA testing in 3 high-risk populations. *The Journal of Infectious Diseases*, 224(5), 881–888. <https://doi.org/10.1093/infdis/jiaa801>
26. Cohen, C. M., & Clarke, M. A. (2023). Anal cancer and anal cancer screening. *Clinical Obstetrics and Gynecology*, 66(3), 516–533. <https://doi.org/10.1097/GRF.0000000000000789>
27. Castle, P. E., Stoler, M. H., Wright, T. C., Sharma, A., Wright, T. L., & Behrens, C. M. (2011). Performance of carcinogenic human papillomavirus (HPV) testing and HPV16 or HPV18 genotyping for cervical cancer screening of women aged 25 years and older: A subanalysis

- of the ATHENA study. *Gynecologic Oncology*, 123(3), 880–890. <https://doi.org/10.1016/j.ygyno.2011.08.005>
28. Faillace Thiesen, J., Jacquemet, E., Campagne, P., Chatelain, D., Brochot, E., Herpe, Y.-E., Dheilly, N. M., Bouilloux, F., Rognon, B., Douablin, A., Leboucher, G., Percher, F., Eloit, M., & Pérot, P. (2025). Molecular detection of hrHPV-induced high-grade squamous intraepithelial lesions of the cervix through a targeted RNA next generation sequencing assay. *Molecular Medicine*. <https://doi.org/10.1186/s10020-025-01238-x>
 29. Ayass, M. A., Zaman, N., Griko, N., Pashkov, V., Zhu, K., Abbas, G., Ghelan, M., Ramankutty Nair, R., Okyay, T., Zhang, J., & Abi-Mosleh, L. (2026). Non-invasive host transcriptome and HPV oncogene expression map the molecular landscape of HPV-driven cervical lesions. *PLOS ONE*, 21(3), e0339693. <https://doi.org/10.1371/journal.pone.0339693>
 30. Xia, P., Zhou, J., Shen, R., & Wan, D. (2025). Deciphering the cellular and molecular landscape of cervical cancer progression through single-cell and spatial transcriptomics. *npj Precision Oncology*. <https://doi.org/10.1038/s41698-025-00948-z>
 31. Bello-Perez, M., Mascarell, P., Fernández-González, M., García, J. A., Gutiérrez-Ortiz de la Tabla, A., Ledesma, C., De la Rica, A., Galiana, A., López, L., Ocete, M. D., Salvador, C., Gimeno, C., García-Abellán, J., Padilla, S., Gutiérrez, F., & Masiá, M. (2025). ADAR1 mRNA quantification for predicting HSIL in persons with HIV and abnormal anal cytology. *International Journal of Cancer*. <https://doi.org/10.1002/ijc.70050>
 32. Su, Q., Tian, X., Li, F., Yu, X., Gong, W., Chen, Y., Wang, J., Yang, S., Zhang, S., Zhang, Q., & Yang, S. (2025). Integrated multi-omics analysis of single-cell and spatial transcriptomics reveals distinct HPV-associated immune microenvironment features and prognostic signatures in cervical cancer. *Frontiers in Immunology*, 16, 1612623. <https://doi.org/10.3389/fimmu.2025.1612623>
 33. Palefsky, J. M., Lee, J. Y., Jay, N., Goldstone, S. E., Darragh, T. M., Dunlevy, H. A., Rosa-Cunha, I., Arons, A., Pugliese, J. C., Vena, D., Sparano, J. A., Wilkin, T. J., Bucher, G., Stier, E. A., Tirado Gomez, M., Flowers, L., Barroso, L. F., Mitsuyasu, R. T., Lensing, S. Y., ... ANCHOR Investigators Group. (2022). *Treatment of anal high-grade squamous intraepithelial lesions to prevent anal cancer*. The New England Journal of Medicine, 386(24), 2273–2282. <https://doi.org/10.1056/NEJMoa2201048>