



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR)

Lidy Vannessa Mejía Guarnizo

Universidad Nacional de Colombia
Facultad de ciencias, Instituto de Biotecnología
Bogotá, Colombia

2024

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR)

Lidy Vannessa Mejía Guarnizo

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título
de:

Magister en ciencias-Microbiología

Director (a):

MSc. Josefa Antonia Rodríguez

Codirector (a):

Ph.D. Liliana López Kleine

Asesor (a):

MSc. Clara Esperanza Trujillo

Línea de Investigación:

Biología celular

Grupo de Investigación:

Instituto Nacional de Cancerología

Grupo de investigación en Biología del Cáncer

Universidad Nacional de Colombia

Facultad de ciencias, Instituto de Biotecnología

Bogotá, Colombia

2024

IV

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

A mi familia, a mi hija, a mi compañero de vida y a quienes ya no están, gracias por acompañarme y motivarme desde el inicio de este proceso. Ustedes vieron mis momentos de flaqueza y me alentaron a seguir adelante. Su apoyo incondicional, su paciencia y su amor me han demostrado que todo es posible.

Agradecimientos

Mis más sinceros agradecimientos a la Doctora Esperanza Trujillo, por su acompañamiento constante a lo largo de este proceso, desde el inicio hasta el final. Su transmisión de conocimientos y valores como investigadora me permitió construir una sólida base en mi camino de investigación. Agradezco profundamente su apoyo incondicional en los momentos difíciles, su perseverancia y su firmeza en las adversidades. Es una gran persona tanto en lo personal como en lo profesional, siempre guiándome con sabiduría y tomando decisiones con gran responsabilidad.

A la Doctora Josefa Rodríguez, mi gratitud por abrirme las puertas de su grupo y brindarme la oportunidad de trabajar con ella desde hace años. Gracias por sus valiosos consejos y por estar siempre disponible para apoyarme, por su compromiso con mi formación, por sus correcciones y por su deseo constante de ver a sus estudiantes triunfar.

A la Doctora Liliana López Kleine, agradezco sinceramente su disposición para colaborar y sus valiosos aportes durante todo el proceso. Sus sugerencias y correcciones fueron fundamentales para mi desarrollo, y su compromiso con mi formación es muy apreciado.

Mi agradecimiento al Dr. Carlos Orozco y a Sebastián Prada Padilla por su orientación y entrenamiento en el análisis de datos, así como por compartir su experiencia en bioinformática. Sus enseñanzas fueron cruciales para el manejo de los datos.

A la Doctora María Mercedes Bravo, líder del proyecto de investigación, le agradezco la oportunidad de participar en su equipo y por permitirme involucrarme en este proyecto tan significativo.

Mi gratitud al grupo de investigación en Biología del Cáncer y a todos sus miembros por su colaboración y apoyo. También agradezco al Instituto Nacional de Cancerología por su apoyo a la formación de estudiantes y por las instalaciones que facilitaron el desarrollo de la investigación. A la Universidad Distrital Francisco José de Caldas y al Profesor Duver Martínez, por confiar en nosotros y prestarnos el equipo de secuenciación, sin el cual no hubiéramos obtenido los resultados. A Viviana Quintero, por su asesoría y entrenamiento en la secuenciación.

Finalmente, quiero expresar un agradecimiento profundo a mi compañero de vida y a mi hija, quienes, a pesar del tiempo que dediqué a este proceso, siempre estuvieron incondicionales. A mis padres y hermanas, por su compañía y apoyo constante. Y a mis amigos Ana María, Mafe, Sebastián, Adriana y Felipe, por estar conmigo.

Resumen

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

El cáncer de cuello uterino (CCU) es el cuarto tipo de cáncer más común en mujeres a nivel mundial y el tercero en países en vías de desarrollo. La infección persistente por el virus del papiloma humano de alto riesgo (VPH-AR) es un factor clave en el desarrollo de neoplasia intraepitelial cervical (NIC) de grados 1 a 3. Aunque la detección de VPH-AR supera a la citología en eficacia, su valor predictivo para NIC2/3 es limitado debido a la naturaleza transitoria de muchas infecciones. Por ello, es crucial encontrar métodos complementarios para mejorar la detección y el diagnóstico. Los microRNAs (miRNAs) se presentan como biomarcadores prometedores por su alteración en diversos cánceres. Nuestro estudio tuvo como objetivo determinar los perfiles de expresión de miARNs en pacientes con lesiones de alto grado (NIC2/3) y bajo grado (SL/NIC1) positivas para VPH-AR, con el fin de identificar una firma molecular que permita diferenciar entre NIC2/3 y NIC1, y evaluar el papel de estos miARNs en la carcinogénesis cervical. Para ello, realizamos un estudio en dos fases. En la fase de descubrimiento, secuenciamos (small RNAseq) muestras de cepillados cervicales de mujeres VPH-AR+ con lesiones de bajo grado (SL/NIC1, n=31) y alto grado (NIC2/3, n=33). Construimos perfiles de expresión diferencial de miARNs y realizamos agrupamientos (clustering) por k-medias, curvas ROC y regresión lineal bivariada para identificar los miARNs con mayor potencial discriminativo. Los estudios de enriquecimiento funcional y el análisis de datos externos (GEO) respaldaron nuestros hallazgos. En la fase de validación, mediante RT-qPCR utilizamos un grupo independiente de 30 muestras para evaluar la capacidad de los miARNs con curvas ROC calculando el AUC de los miRNAs para la detección de NIC2/3. Identificamos 19 miARNs con expresión diferencial significativa. De estos, cuatro (hsa-miR-1271-5p,

hsa-miR-9-5p, hsa-miR-501-3p, y hsa-miR-1246) reunieron la mayoría de las características analizadas y se destacó su implicación en la regulación de genes clave en la infección por VPH, lo que sugiere su potencial como biomarcadores para discriminar NIC2/3. Además, se observó que el uso prolongado de anticonceptivos se asoció con un mayor riesgo de NIC2/3. En la fase de validación, tres miARNs (hsa-miR-1271, hsa-miR-501, y hsa-miR-9-5p) presentaron una expresión significativamente mayor en lesiones NIC2/3 ($p < 0.05$) y su expresión aumentó con la progresión de la enfermedad (NIC1, NIC2, NIC3). El miARN hsa-miR-501 mostró el mayor AUC (88%) para la detección de NIC2/3, y su combinación con hsa-miR-1271 y hsa-miR-9-5p mejoró la precisión del modelo a un AUC de 94%. Estos resultados, que se correlacionan entre ambas fases del estudio, subrayan el potencial de los miARNs identificados como biomarcadores prometedores para el tamizaje de lesiones cervicales preneoplásicas, con implicaciones significativas para su futura aplicación clínica en la detección de la lesión de alto grado y estratificación del riesgo.

Palabras clave: VPH-AR, neoplasia intraepitelial cervical, miRNAs, biomarcador, cepillado cervical, citología.

Abstract

Identification of a miRNA signature with potential usefulness as a biomarker for the triage of women positive for high-risk human papillomavirus (HR-HPV) infection.

Cervical cancer (CC) is the fourth most common cancer in women worldwide and the third in developing countries. Persistent infection with high-risk human papillomavirus (HPV-HR) is a key factor in the development of cervical intraepithelial neoplasia (CIN) grades 1 to 3. Although HPV-HR detection is more effective than cytology, its predictive value for CIN2/3 is limited due to the transient nature of many infections. Therefore, it is crucial to find complementary methods to improve detection and diagnosis. MicroRNAs (miRNAs) are emerging as promising biomarkers due to their alteration in various cancers. Our study aimed to determine the miRNA expression profiles in patients with high-grade (CIN2/3) and low-grade (SL/CIN1) lesions positive for HPV-HR, with the goal of identifying a molecular signature that differentiates between CIN2/3 and CIN1 and evaluating the role of these miRNAs in cervical carcinogenesis. To achieve this, we conducted a two-phase study. In the discovery phase, we sequenced (small RNAseq) cervical brush samples from HPV-HR+ women with low-grade lesions (SL/CIN1, n=31) and high-grade lesions (CIN2/3, n=33). We constructed differential miRNA expression profiles and applied k-means clustering, ROC curves, and bivariate linear regression to identify miRNAs with the highest discriminatory potential. Functional enrichment studies and external data analysis (GEO) supported our findings. In the validation phase, using RT-qPCR, we employed an independent group of 30 samples to evaluate the diagnostic capability of the miRNAs with ROC curves, calculating the AUC of the miRNAs for CIN2/3 detection. We identified 19 miRNAs with significant differential expression. Of these, four (hsa-miR-1271-5p, hsa-miR-9-5p, hsa-miR-501-3p, and hsa-miR-1246) met most of the analyzed criteria and showed involvement in regulating key genes in HPV infection, suggesting their potential as biomarkers for discriminating CIN2/3. Additionally, prolonged use of contraceptives was associated with a higher risk of CIN2/3. In the validation phase, three miRNAs (hsa-miR-1271, hsa-miR-501, and hsa-miR-9-5p) exhibited significantly higher expression in CIN2/3 lesions ($p < 0.05$), and their expression increased with disease progression (CIN1, CIN2, CIN3). miRNA hsa-miR-501 demonstrated the highest AUC (88%) for detecting CIN2/3,

and its combination with hsa-miR-1271 and hsa-miR-9-5p improved the model's accuracy to an AUC of 94%. These results, correlating across both phases of the study, underscore the potential of the identified miRNAs as promising biomarkers for cervical pre-cancerous lesion screening, with significant implications for their future clinical application in high-grade lesion detection and risk stratification.

Keywords: HPV-HR, cervical intraepithelial neoplasia, miRNAs, biomarker, cervical brush, cytology.

Contenido

1. Planteamiento del problema	5
2. Justificación.....	7
3. Objetivos	10
3.1 Objetivo General	10
3.2 Objetivos específicos	10
3.2.1 Objetivo específico 1	10
3.2.2 Objetivo específico 2	10
3.2.3 Objetivo específico 3.....	10
4. Marco de referencia.....	11
4.1 MiRNAs en Cáncer	11
4.2 MiRNAs en cáncer de cuello uterino	14
5. Marco teórico	18
5.1 Virus del papiloma humano	19
5.1.1 Estructura y genoma del VPH.....	21
5.1.2 Funciones principales de las proteínas virales	22
5.1.3 Oncoproteínas	23
5.2 VPH y carcinogénesis cervical.....	23
5.2.1 Canal cervical y anatomía del cuello uterino	23
5.2.2 Cambios patogénicos e historia natural del CCU de células escamosas	24
5.3 Cáncer de cuello uterino	25
5.4 Prevención del CCU.....	26
5.4.1 Prevención primaria.....	26
5.4.2 Prevención secundaria.....	26
5.4.3 Prevención terciaria	26
5.5 MiRNAs	27
5.5.1 Biogénesis de los miRNAs	27
6. Metodología.....	29
6.1 Diseño del estudio.....	29
6.2 Descripción de los sujetos del estudio.....	30
6.3 Consideraciones éticas.....	31
6.4 Procedimientos experimentales para el cumplimiento de los objetivos.....	32
6.4.1 Extracción del RNA	32
6.4.2 Cuantificación del RNA total.....	33
6.4.3 Construcción de bibliotecas de secuenciación	33

6.4.4	Secuenciación	34
6.4.5	Análisis Bioinformático	35
6.4.6	Control de calidad tabla de conteos.....	37
6.5	Análisis de datos para el cumplimiento de los objetivos	38
6.5.1	Caracterización Inicial de las muestras	38
6.5.2	Análisis de Clustering.....	38
6.5.3	Identificación de miRNAs candidatos	39
6.5.4	Análisis para identificar miRNAs con potencial utilidad para detectar NIC2/3	40
6.5.5	Validación experimental	42
7.	Resultados	45
7.1	Capítulo 1. Resultados de procedimientos generales.....	45
7.1.1	Elaboración de bibliotecas de secuenciación:	45
7.1.2	Calidad de los datos de secuenciación.....	48
7.1.3	Mapeo de los datos de secuenciación.....	49
7.1.4	Control de calidad de la tabla de conteos.....	50
7.2	Capítulo 2. Clustering y Análisis de expresión diferencial	54
7.2.1	Resultados objetivo 1.	54
7.2.2	Resultados objetivo 2.	60
7.3	Capítulo 3. miRNAs con alto poder discriminatorio para detectar NIC2/3.....	72
7.3.1	Resultados Objetivo 3.	72
8.	Discusión.....	88
9.	Conclusiones	95
10.	Recomendaciones y perspectivas	97
11.	BIBLIOGRAFIA.....	99

Lista de figuras

Figura 1 miRNA como oncogenes o supresores de tumor.	12
Figura 2 Árbol evolutivo basado en el alineamiento de secuencias para los genes E1, E2, L1 y L2 del VPH.	20
Figura 3 Clasificación de los VPH en alto y bajo riesgo.	21
Figura 4 Esquema del genoma del VPH-16.....	21
Figura 5 Distribución de células epiteliales escamosas normales e infectadas por el VPH en lesiones precancerosas normales (displasia leve, moderada y grave; NIC 1, NIC 2 y NIC 3, respectivamente) y CCU.	24
Figura 6 Incidencia y mortalidad del CCU en Colombia según Global Cancer Observatory: Cancer Today.	25
Figura 7 Biogénesis de los miRNAs.	28
Figura 8. Diagrama de las principales fases del diseño del estudio para la selección y obtención de las muestras.....	29
Figura 9 Esquema de las muestras seleccionadas para el estudio..	31
Figura 10 Resumen de las técnicas y equipos empleados para la obtención de las bibliotecas.	32
Figura 11 Esquema detallado de las fases de la construcción de la biblioteca.	35
Figura 12 Esquema de la presentación de los resultados para el cumplimiento de los objetivos.....	45
Figura 13 Resultados de control de calidad del RNA total obtenidos con el 2100 Bioanalyzer.	46
Figura 14 Purificación del constructo de cDNA con Gel TBE Novex™ 6%.....	47
Figura 15 Cuantificación mediante Bioanalizador del purificado de la biblioteca.	47
Figura 16 Valor Q para las lecturas de las muestras secuenciadas del pool 9.....	48
Figura 17 Cantidad y distribución de las muestras por lesiones después de realizar los filtros de control y calidad.	51
Figura 18 Variabilidad de los conteos de las lecturas para cada miRNA (n=717) en cada una de las muestras (n=64).....	53
Figura 19. Análisis de PCA con los miRNAs luego del control de calidad (n= 717 y 64)..	55
Figura 20 Análisis de clustering utilizando el algoritmo de K-medias para todos los miRNAs (n=717).	56
Figura 21 Análisis de clustering con K-medias para todos los miRNAs (n=717) usando la visualización con la función s.class.....	58
Figura 22 Mapa de calor de los miRNAs expresados diferencialmente entre SL/NIC1> vs NI2/3.....	63

Figura 23 Análisis de PCA de los miRNAs expresados diferencialmente.....	66
Figura 24 Análisis de k-medias en los dos primeros componentes principales para 2 clusters usando los niveles de expresión de los miRNAs expresados al alza	68
Figura 25 Correlación de los DEmiRNAs.....	69
Figura 26 Visualización de K-medias con la función s.class. según los miRNAs expresados diferencialmente.....	70
Figura 27 Curva ROC de los miRNA que presentaron significancia estadística.....	73
Figura 28 Análisis de Enriquecimiento para la categoría de componente celular (CC)...	79
Figura 29 Análisis de enriquecimiento para componente de procesos biológicos (BP). .	80
Figura 30 Análisis de enriquecimiento funcional para componente de función molecular (FM).....	80
Figura 31 Analisis de enriquecimiento para vias metabolicas.	81
Figura 32. Expresión relativa de los 6 miRNAs en las muestras de validación.	83
Figura 33. Expresión relativa $\text{Log}_2(2^{-\Delta\Delta C_t})$ de los miRNAs por grupo de lesión (NIC1, NIC2, NIC3) en las muestras de validación.	85
Figura 34. Curvas ROC para cada miRNA individual para la detección de NIC2/3 en los datos de validación.	86
Figura 35. Curva ROC de las combinaciones de hsa-miR-9-5p y hsa-miR-501-3p para la detección de NIC2/3.	87
Figura 36. Curva ROC de las combinaciones de hsa-miR-9-5p, hsa-miR-501-3p y hsa-miR-1271 para la detección de NIC2/3.	87

Lista de tablas

Tabla 1 MiRNAs reportados como oncogénicos o supresores de tumores en CCU.....	15
Tabla 2 Resumen de los miRNAs más relevantes encontrados en diferentes muestras provenientes de cáncer cervical.....	16
Tabla 3 Resumen de miRNAs regulados positivamente y negativamente en el contexto del cáncer cervical.	16
Tabla 4 Métricas del rendimiento de la secuenciación y el alineamiento/mapeo al transcriptoma de referencia (miRBase V.22.1).	49
Tabla 5 Número de muestras por lesión.....	49
Tabla 6. Descripción de la población de estudio.	51
Tabla 7 miRNAs expresados diferencialmente para SL vs NIC1, SL vs NIC2/3 y NIC1 vs NIC2/3.....	60
Tabla 8 miRNAs expresados diferencialmente entre las comparaciones NIC2/3 y SL/NIC1.....	62
Tabla 9 Resumen Set de datos obtenidos de GEO para la comparación del análisis de expresión diferencial.	64
Tabla 10 Tabla de contingencia con la distribución de las muestras en los clústeres 1 y 2 utilizando los DEmiRNAs compartidos entre los análisis.	65
Tabla 11. Tabla de contingencia que muestra la distribución de las muestras en los clústeres 1 y 2 utilizando los miRNAs expresados diferencialmente (al alza y a la baja) en la comparación SL/NIC1 vs NIC2/3.....	67
Tabla 12 Tabla de contingencia que muestra la distribución de las muestras en los clústeres 1 y 2 utilizando los miRNAs seleccionados: ("hsa-miR-1271-5p", "hsa-miR-9-5p", "hsa-miR-3690", "hsa-miR-6510-3p", "hsa-miR-4662a-5p", "hsa-miR-184", "hsa-miR-501-3p" y "hsa-miR-1246").	71
Tabla 13 Regresión logística bivariada.....	75
Tabla 14. Tabla de Características de miARNs para la Discriminación de NIC2/3.....	77
Tabla 15 Genes target de los miRNAs relacionados con la carcinogénesis cervical en la infección por VPH.	78
Tabla 16. Características de los miRNAs seleccionados según su desempeño en los análisis realizados para la detección de NIC2/3.	81
Tabla 17. Numero de muestras utilizadas en la cohorte de validación por RT-qPCR	82
Tabla 18. Tabla de contingencia del análisis de clustering por K-medias utilizando los miRNAs hsa-miR-1271 y hsa-miR-6510.	73

Lista de Símbolos y abreviaturas

Abreviatura	Término
CCU	Cáncer de cuello uterino
VPH	Virus del papilloma humano
VPH-AR	Virus de papiloma humano de alto riesgo
NIC	Neoplasia intraepitelial cervical
miRNA	microRNA
NIC2/3	Neoplasia intraepitelial cervical grado 2 y 3, o (lesiones de alto grado)
SL/NIC1	Sin lesión / Neoplasia intraepitelial cervical grado 1 (Lesiones de bajo grado de cérvix)
NIC 1: NIC I	Neoplasia Intraepitelial de grado 1
NIC1>	Neoplasia Intraepitelial de grado 1 o menor
NIC 2: NIC II	Neoplasia Intraepitelial de grado 2
NIC2+	Neoplasia Intraepitelial de grado 2 o peor
NIC 3: NIC III	Neoplasia Intraepitelial de grado 3
NIC3+	Neoplasia Intraepitelial de grado 3 o peor
OMS	Organización mundial de la salud
VPH-AR+	Virus de papiloma humano de alto riesgo positivo
VPH-BR	Virus de papiloma humano de bajo riesgo
RNAm	RNA mensajeros
RNA	Ribonucleic Ácid
ADN	Ácido Desoxirribonucleico
qRT-PCR:	Reacción en cadena de la polimerasa cuantitativa acoplada a transcriptasa reversa
OncomiRs	miRNAs oncogénicos
LLC	Leucemia linfocítica crónica
DEmiRNAs	miRNAs desregulados
IHC	Inmunohistoquímica
GLOBOCAN	Global Cancer Observatory: Cancer Today.
ROC	Receiver operating characteristic curve
FC	Fold Change

Introducción

El CCU es la segunda causa de muerte por cáncer en mujeres de 15 y 44 años a nivel mundial (1). De acuerdo con globocan (2022), a nivel mundial el CCU ocupa el octavo lugar en incidencia con 661.021 nuevos casos y el noveno en mortalidad con 348.189 muertes a causa de la enfermedad (2). En Colombia, ocupa el tercer lugar en incidencia y mortalidad entre los cánceres que afectan a las mujeres con 13.7 casos y 6.9 casos por cada 100.000 habitantes respectivamente (3), por lo que es considerado un problema de salud pública. Este cáncer se denomina "enfermedad de la disparidad", debido a la desproporción en las cifras de incidencia y mortalidad entre los países de ingresos bajos y medios y los países de ingresos altos (4). Las barreras socioeconómicas y de acceso a la salud contribuyen a la alta incidencia y mortalidad por este cáncer en Colombia (5,6).

El VPH se detecta en el 99.7% de los casos de CCU, por lo que es considerado como el factor etiológico de este tipo de cáncer (7), y un factor de riesgo para el desarrollo de lesiones intraepiteliales cervicales y, eventualmente, cáncer. La infección por VPH es la infección de transmisión sexual más común y aunque la mayoría de las mujeres logra eliminar la infección (7), las mujeres que tienen infecciones persistentes VPH-AR desarrollan lesiones precursoras que, de no ser detectadas y tratadas, tienen el potencial de progresar a CCU. Los VPH-AR 16 y 18 son los más comúnmente encontrados en CCU y se asocian con el 70% de todos los cánceres de cérvix (7).

La historia natural del CCU es un proceso lento, que tarda años/décadas en ocurrir (8), desde la infección inicial, pasando por las lesiones intraepiteliales de bajo, mediano y alto grado, con mayor riesgo de progresar a cáncer invasor (9), hay una ventana de tiempo suficiente para la detección y el tratamiento de estas lesiones, que podría disminuir la mortalidad por este tipo de cáncer (7).

La detección de las lesiones precursoras tradicionalmente se ha basado en la citología cervicovaginal (CCV), y confirmadas mediante una biopsia. Esta técnica ha desempeñado un papel crucial en la reducción de las tasas de mortalidad en los países que tienen amplia cobertura y calidad, sin embargo, en países en vías de desarrollo las tasas de mortalidad persisten debido a la falta de programas de detección y seguimiento eficaces (4,10). En Colombia, hubo una disminución en la incidencia desde el 2012 hasta el 2018 (de 15.1 a 13.2), pero la tasa de mortalidad se ha mantenido constante entre los años 2017 y 2021 (de 6.6 a 6.5) (11).

La citología cervicovaginal enfrenta desafíos que comprometen su eficacia como la discordancia entre observadores, que genera baja reproducibilidad de la prueba con altas tasas de falsos negativos (12), y, en especial, una baja sensibilidad de la prueba para detectar NIC2+ (13), y NIC3+ (14). La implementación de la citología en base líquida no mejora la sensibilidad ni la especificidad para la detección de lesiones intraepiteliales cervicales de alto grado (15) y su impacto en la incidencia y la mortalidad asociadas al CCU es muy limitada (12), por otra parte, su aplicación en los países en vías de desarrollo es compleja, difícil de ejecutar, ineficiente y costosa.

Dada la relación etiológica entre la infección por VPH y la carcinogénesis cervical, se ha propuesto la prueba de ADN del VPH como prueba de detección alternativa a la citología (16). La Organización Mundial de la Salud (OMS) respalda el uso de la prueba del ADN del VPH como método de tamización primaria más eficaz que la citología para reducir la incidencia de lesiones preneoplásicas y cáncer (7, 17, 18). La elección de esta prueba se justifica en su notable eficacia, ofreciendo un 60 a 70% más de protección que la tamización basada en citología, como señala un estudio de Ronco y colaboradores en 2014 (19) (20) y exhibe una sensibilidad superior al 90% para distinguir lesiones de alto grado (NIC2/3) (7, 21, 22). Esta prueba se destaca por su valor predictivo negativo cercano al 100%, es decir que un resultado negativo para ADN de VPH reduce la probabilidad de desarrollar cáncer, y su rentabilidad es mayor, comparada con la citología, porque se realiza cada 5 años en caso de que el resultado sea negativo (23).

Sin embargo, la detección primaria de VPH presenta una especificidad inferior al 50%, debido a que la mayoría de las mujeres infectadas con el VPH-AR eliminan la infección de forma espontánea y sólo un pequeño porcentaje de las mujeres infectadas con VPH-AR desarrollan infección persistente y lesiones precursoras del cáncer de cuello uterino,

por lo que esta prueba aumenta el número de resultados de detección positivos, la cantidad de remisiones a colposcopias y los sobre tratamientos, que aumentan los costos en el sistema de salud, y también se traduce en carga emocional en las mujeres. (23, 24). Por tal motivo, ante una prueba de detección primaria de VPH, es crucial desarrollar métodos que identifiquen, entre las mujeres VPH+, a aquellas con lesiones preneoplásicas de alto grado y mayor riesgo de progresión a cáncer cervical.

Se han desarrollado pruebas para clasificar el grado y riesgo de las lesiones en mujeres positivas para VPH-AR, que incluyen la genotipificación parcial de VPH 16 y 18, la inmunohistoquímica para p16/Ki-67, la metilación del ADN, la expresión de los oncogenes E6 y E7 del VPH y biomarcadores moleculares, entre otros (26). En la actualidad las pruebas conjuntas de citología y ADN VPH se han destacado en mejorar la estratificación del riesgo (21,23), sin embargo; es importante destacar que ambas son herramientas de detección, no de clasificación de lesiones. La prueba estándar es la colposcopia con biopsia para confirmar y estadificar las lesiones premalignas (26), pero esta prueba es invasiva y, en ocasiones dolorosa, por lo que, a pesar de su efectividad, es necesario realizar pruebas más precisas para identificar, entre las mujeres con VPH-AR+. Por lo tanto, se requiere una prueba que permita distinguir objetivamente las infecciones transitorias que no van más allá de causar una lesión de bajo grado de las asociadas a transformación oncogénica o de alto grado, que tienen mayor riesgo de terminar en cáncer de cérvix. Esta clasificación es fundamental para evaluar el riesgo asociado a las lesiones y mejorar la identificación de pacientes con un riesgo significativo de desarrollar cáncer cervical.

En los últimos años, el análisis de los cambios de expresión en miRNAs se ha convertido en una herramienta útil para distinguir entre células normales y tumorales y con potencial como biomarcador para el diagnóstico de diferentes tipos de cáncer (26). Los miRNAs son secuencias de RNA de 19 a 23 nucleótidos, cuya función es regular negativamente la expresión génica a nivel postranscripcional (27,28,29,30). Los miRNAs suprimen la expresión de su mRNA blanco al aparearse con secuencias complementarias en las 3'UTR, lo que desestabiliza el transcrito, reprime la traducción, o ambas (31). Dado que los miRNAs humanos se localizan con frecuencia en sitios frágiles como las regiones cromosómicas afectadas en cáncer, se postula que estas alteraciones representan un mecanismo subyacente a la expresión alterada de miRNAs en cáncer. Este mecanismo ha

sido demostrado *in vitro* en líneas celulares de melanoma, neuroblastoma, mieloma, y en cáncer de ovario y de mama (32).

La expresión aberrante de miRNAs en CCU y sus lesiones precursoras se investigó previamente (26) (33) (34), y los resultados obtenidos respaldan la idea de que los miRNAs desregulan la expresión de genes supresores de tumor vinculados a procesos carcinogénicos del VPH, como TP53, CDKN1A y CDKN1B, pRB-E2F1 (35). Sin embargo, dichos resultados no siempre fueron concluyentes en cuanto a la capacidad de los miRNA para apoyar o suprimir la carcinogénesis del cuello uterino (33) (36), probablemente debido a diferencias en los materiales o métodos utilizados para el análisis.

Los perfiles de miRNAs son fácilmente detectables con RNAseq, por lo que en el presente estudio se propuso secuenciar y determinar los perfiles de expresión de miRNAs que permitan discriminar eficazmente entre lesiones preneoplásicas de alto (NIC2/3) y bajo grado (NIC1), en mujeres VPH-AR+. La capacidad de estos perfiles para clasificar las lesiones de manera precisa no solo facilita la separación de los casos de mayor riesgo, sino que también tiene implicaciones directas para el triaje clínico. Una clasificación precisa basada en miRNAs ayudaría a priorizar el manejo de mujeres VPH-AR+ con lesiones de alto riesgo, optimizando recursos y reduciendo intervenciones innecesarias en mujeres con lesiones de bajo riesgo. Por tal razón, el objetivo de este estudio fue *“Determinar los perfiles de expresión de los miRNAs presentes en muestras de cepillados cervicales de mujeres con lesiones de alto (NIC 2/3) y bajo grado (NIC 1) o sin lesiones, positivas para infección por VPH de alto riesgo con el fin de establecer una firma de miRNAs con potencial utilidad en el triaje de mujeres positivas para VPH-AR”*, siendo este uno de los primeros estudios en el país sobre el papel de los miRNAs como biomarcadores en mujeres positivas para VPH-AR.

1. Planteamiento del problema

En 2020 la incidencia de CCU fue de 14.9 casos por cada 100.000 habitantes con 4.570 nuevos casos, lo que lo convierte en el tercer cáncer más frecuente en mujeres y en Colombia, es la tercera causa de muerte por cáncer en mujeres (2435 muertes que equivalen al 4.3% del total). A nivel mundial ocupa el cuarto lugar en mortalidad con 341.831 muertes (1,3), por lo que el CCU continúa siendo un problema de salud pública, particularmente en países en vías de desarrollo, en los cuales las estrategias de prevención y diagnóstico temprano no son tan efectivas (37,38).

La citología cervicovaginal, ha sido crucial para la reducción de la incidencia y la mortalidad por CCU en países desarrollados gracias al establecimiento de programas de tamizaje bien estructurados (38,39). Sin embargo, en países de ingresos bajos y medianos, la calidad y cobertura de la citología son inadecuadas, lo que resulta en un alto porcentaje de falsos negativos (>50%) (38) debido, en parte, a que es una prueba dependiente de la habilidad del operador, lo que dificulta la discriminación precisa entre lesiones de alto (NIC2/3) y bajo (NIC1>) grado e incrementa el riesgo de no detectar lesiones significativas. (13,20). Incluso, cuando se obtiene un resultado positivo para la lesión, se requieren remisiones adicionales a colposcopia y/o biopsia para estratificar el riesgo de la lesión. El entendimiento del papel del virus en el CCU permitió adaptar la prueba del ADN del VPH-AR como tamizaje primario para la detección del virus y la prevención de ésta enfermedad (13) Este método de detección es más sensible (>90%) y rentable que la citología para detectar lesiones premalignas, pero, su valor predictivo es bajo debido a la pequeña proporción de mujeres positivas para ADN de virus de alto riesgo que desarrolla CCU (20), lo que dificulta la estratificación del riesgo o triaje de las infecciones primarias por VPH-AR. Se estima que alrededor del 60% al 90% de NIC1 regresará a la normalidad después de 1 año, mientras que las mujeres con NIC2 y NIC3 tienen un mayor riesgo de desarrollar cáncer invasivo (15% y 10% respectivamente) (40,41,42)

El principal desafío para la prevención del CCU es el desarrollo de una clasificación precisa para las mujeres VPH-AR+. Dado que menos del 10% de las mujeres infectadas por el VPH-AR, desarrollarán una infección persistente que progresará a la aparición de lesiones preneoplásicas de alto grado. Lo anterior resalta la importancia de la identificación de biomarcadores con posible utilidad para el desarrollo de una estrategia de triaje que permita identificar el grado y riesgo de la lesión en mujeres que resulten positivas para VPH-AR en la tamización basada en detección de VPH. Aunque existe evidencia del papel de los VPH-AR como el principal factor de riesgo para desarrollar CCU, también es bien conocido que el VPH es necesario, pero no suficiente para desarrollar el CCU. En la carcinogénesis del cuello uterino, intervienen otros factores como las variaciones genéticas individuales y los factores ambientales.

Los miRNA son RNA endógenos no codificantes con una longitud de aproximadamente 22 nucleótidos, capaces de modular la expresión génica a nivel postranscripcional mediante su unión a secuencias complementarias en la región no traducida (UTR) 3' de los RNA mensajeros (RNAm). Estos miRNAs son útiles para el diagnóstico clínico gracias a su estabilidad en condiciones de temperatura y pH no controladas, a su resistencia a las ribonucleasas (43) y, a su larga duración *In vivo* (44) e *In vitro* (45), lo que facilita su análisis y la obtención de resultados automatizados (46), en diversas muestras (47). Estudios adicionales sugieren que el VPH parece desempeñar un papel crucial en la modulación de la red celular de miRNA, y que la mayoría de los miRNA vinculados a este proceso, tienden a expresarse como consecuencia de la infección por VPH (10, 48, 49).

Dado que la recolección de células cervicales constituye un método poco invasivo para realizar la citología y la detección del VPH, el cepillado cervical constituye una muestra de tejido estratégica para identificar perfiles de expresión de miRNA que pueden ser biomarcadores útiles para la detección de las lesiones preneoplásicas de alto grado, en mujeres positivas para el VPH-AR.

PREGUNTA DE INVESTIGACIÓN

¿La expresión diferencial de miRNAs en cepillados cervicales de mujeres positivas para infección por VPH-AR con lesiones de bajo (SL/NIC 1) y alto grado (NIC 2/3), puede ser utilizada como biomarcador para la clasificación y estratificación del riesgo de las lesiones pre neoplásicas del cérvix?

2. Justificación

Se ha demostrado que la prueba primaria de detección del VPH brinda una detección superior de las lesiones intraepiteliales de alto grado (>90%) y CCU en comparación con la tamización basada en citología, por lo cual las guías actuales de citología recomiendan realizar pruebas para la detección de ADN del VPH como herramienta de tamización primaria (19, 50, 51, 52). Sin embargo, la especificidad (falsos positivos) de la prueba de VPH para detectar la NIC 2/3 es baja al igual que su valor pronóstico (53), por lo que se requiere encontrar mejores métodos para seleccionar a las mujeres VPH-AR positivas que deben ser referidas a colposcopia con biopsia inmediata por su riesgo de progresión a cáncer.

Aún no se ha llegado a un consenso sobre la estrategia óptima de triaje para las mujeres con VPH positivo, lo que es crítico para desarrollar un programa de detección basado en la infección por VPH, debido a la mayor sensibilidad, pero menor especificidad de esta prueba (52, 54). La repetición de pruebas genera ansiedad e implica pérdidas en el seguimiento. Para prevenir la remisión innecesaria de mujeres con infección transitoria por VPH a colposcopia, y disminuir la ansiedad asociada a estos procedimientos, es necesario disponer de una prueba para estratificar el riesgo de lesión intraepitelial en las mujeres positivas para VPH.

Los miRNAs han emergido como biomarcadores prometedores debido a sus características deseables, como su estabilidad y capacidad para ser cuantificados

mediante técnicas avanzadas como RNAseq y RT-qPCR. Se ha demostrado que los perfiles de miRNA permiten clasificar los cánceres humanos. En muchos tipos de tumores incluido el CCU, se observan patrones característicos de expresión de miRNAs (33). En CCU se han reportado más de 246 miRNAs desregulados (55, 56, 57), y en líneas celulares, lesiones preneoplásicas, y exosomas aislados de plasma se han reportado patrones de expresión de miRNAs alterados. (58, 59, 60, 61,62, 63, 64, 65, 66, 67). Adicionalmente, existe evidencia de que los miRNAs regulan negativamente la expresión de genes supresores de tumor como TP53, CDKN1A (p21) y CDKN1B (p27), cuya función es esencial para suprimir el desarrollo tumoral. Durante la infección por VPH-AR, estos genes supresores de tumor son degradados por las oncoproteínas E6 y E7 del virus, lo que desregula el ciclo celular (68, 69,70,71).

A la fecha no se cuenta con un sistema unificado para la identificación de miRNAs celulares expresados diferencialmente que tengan potencial como biomarcador, en muestras de cepillados cervicales VPH-AR+. Aunque se han realizado intentos para comparar los miRNAs diferencialmente expresados entre las lesiones del cérvix de alto grado y las de bajo grado en mujeres VPH-AR+, y determinar su desempeño diagnóstico en términos de sensibilidad y especificidad, los resultados han sido variables y aún persiste un vacío significativo en el conocimiento. (72, 73; 66, 20, 74, 75, 67). Tian et al. informaron que la detección de miRNA en raspados cervicales positivos para VPH tiene un rendimiento superior sobre la citología para la detección de lesiones NIC de alto grado en una población de referencia, lo que indica el potencial que tiene la detección de miRNAs para las pruebas de clasificación de mujeres VPH-AR+ en programas de tamizaje para la estratificación del riesgo de progresión de las lesiones. Sin embargo, hacen falta estudios que permitan descubrir una firma de miRNAs confiable en NIC y CCU (20).

Este trabajo tiene como objetivo explorar el potencial de los miRNAs para la estadificación de mujeres con infección por VPH-AR, especialmente identificando las lesiones neoplásicas de alto grado. Al clasificar correctamente estas lesiones, se busca reconocer a las mujeres con mayor riesgo de desarrollar cáncer de cuello uterino, lo que permitirá aportar al diseño de estrategias de triaje más precisas y eficaces que contribuyan al abordaje de estrategias de prevención primaria más efectivas. La identificación de las

mujeres en riesgo permitirá reducir los procedimientos invasivos innecesarios y optimizar los recursos del sistema de salud.

HIPÓTESIS

Existen diferencias entre los perfiles de expresión de miRNAs en mujeres VPH-AR+ con NIC de alto grado (NIC2/3) y de bajo grado (NIC1>), que permiten discriminar el grado de la lesión intraepitelial, así como el riesgo de desarrollar lesiones precursoras del cáncer de cuello uterino.

3. Objetivos

3.1 Objetivo General

Determinar los perfiles de expresión de los miRNAs presentes en muestras de cepillados cervicales de mujeres con lesiones de alto (NIC 2/3) y bajo grado (NIC 1) o sin lesiones, positivas para infección por VPH de alto riesgo con el fin de establecer una firma de miRNAs con potencial utilidad en el triaje de mujeres positivas para VPH-AR.

3.2 Objetivos específicos

3.2.1 Objetivo específico 1

Realizar la caracterización y clasificación de las muestras de cepillados cervicales en: normales, con lesiones de bajo grado (NIC 1) o con lesiones de alto grado (NIC 2/3) mediante un método estadístico de clustering, con base en los datos obtenidos del análisis de expresión de miRNAs en las muestras.

3.2.2 Objetivo específico 2

Comparar los perfiles de expresión diferencial de los miRNAs obtenidos mediante la secuenciación masiva en paralelo de los miRNAs obtenidos partir de cepillados cervicales provenientes de mujeres sin lesiones del cérvix, con lesiones de bajo grado (NIC 1) o con lesiones de alto grado (NIC 2/3), positivas para VPH-AR.

3.2.3 Objetivo específico 3

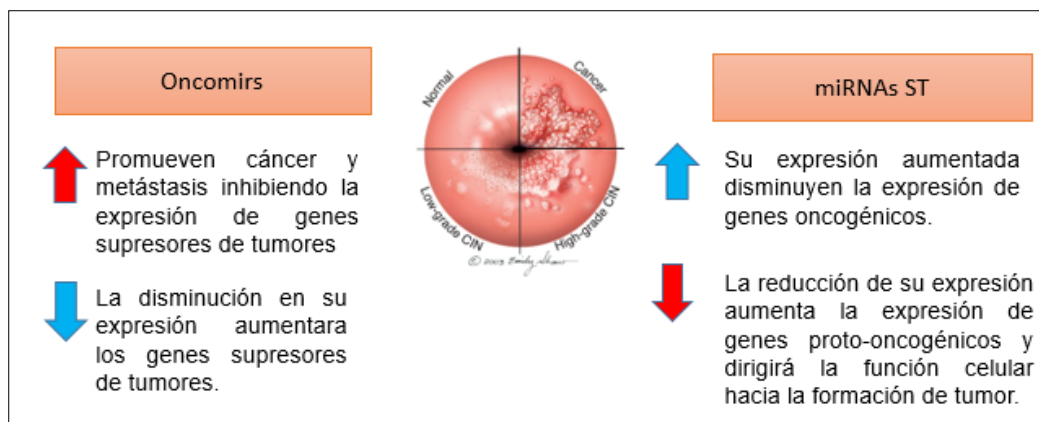
Identificar mediante el análisis bioinformático de las secuencias obtenidas, los miRNAs con potencial utilidad para detectar NIC2/3 con un alto poder discriminatorio e identificar los genes blanco de éstos miRNAs y sus implicaciones en los procesos relacionados con la carcinogénesis.

4. Marco de referencia

4.1 MiRNAs en Cáncer

Los biomarcadores han evolucionado desde enfoques únicos hasta adaptaciones necesarias para abordar procesos complejos como la carcinogénesis. Para ser considerado un biomarcador de pronóstico, el gen o la proteína debe estar expresada diferencialmente entre células normales y tumorales, debe cumplir ciertos criterios y ser validados mediante estrategias que van desde la evaluación molecular hasta su implementación en protocolos clínicos (76). Mediante la genómica, la proteómica y la patología molecular se han identificado múltiples biomarcadores con potencial aplicación en la clínica, para mejorar la precisión diagnóstica y la atención integral del paciente (76). Los biomarcadores de RNA, los RNAm y los miRNA, son los objetivos terapéuticos clave debido a su papel en la regulación de la expresión génica. Los miRNAs realizan una regulación postraduccional asociada con el silenciamiento de genes, por degradación del RNA mensajero (RNAm) o por interferencia con su traducción. Más del 30% de la traducción de proteínas se regula de esta manera (77). Alteraciones en la secuenciación o biogénesis de los miRNAs pueden provocar cambios en su expresión, lo que incrementa el riesgo de pérdida de control del ciclo celular y contribuye al desarrollo tumoral. Existe evidencia de que la función de los miRNA en la carcinogénesis está mediada por la regulación positiva de los miRNA supresores de tumores (ST) o la regulación negativa de los miRNA oncogénicos (oncomiRs) que conducen al inicio y la progresión del cáncer (78, 79). (Figura 1).

Figura 1 miRNA como oncogenes o supresores de tumor. Los miRNAs sobreexpresados son conocidos como oncomiRs, mientras que aquellos con disminución en su expresión se denominan miRNAs ST. Dependiendo del tipo de cáncer, los miRNAs pueden tener una función dual. Los oncomiRs actúan como oncogenes al promover la proliferación, angiogénesis e invasión celular, estimulando la formación de tumores a, mientras que los miRNAs supresores de tumor tienen el efecto contrario al inhibir la carcinogénesis (78,79,80).



La ubicación específica de los miRNA en el genoma es crucial en la forma como se lleva a cabo esta regulación transcripcional. Pueden dividirse en intragénicos e intergénicos según su posición genómica (82), y se cree que los miRNA intragénicos, localizados dentro de intrones o exones en la misma cadena del gen, se regulan conjuntamente con sus genes hospedadores mediante la RNA polimerasa II. En contraste, los miRNA intergénicos, situados entre genes, se transcriben desde sus propios promotores por la RNA polimerasa II o III. No obstante, investigaciones recientes han demostrado que los miRNA intragénicos, incluidos los intrónicos, no siempre se transcriben simultáneamente con los genes hospedadores (79). Alrededor del 50% de los miRNA humanos anotados se agrupan en regiones del genoma, consideradas sitios frágiles que infieren el papel crucial de los miRNA en el desarrollo del cáncer (83)

Las alteraciones en los cromosomas que albergan clusters de miRNA, la expresión anormal de factores de transcripción y los cambios en la vía de procesamiento de los miRNA pueden desregular su expresión en cáncer. Se han descrito una gran cantidad de

miRNA sobreexpresados y reprimidos en diferentes tumores sólidos mediante técnicas como microarreglos, PCR cuantitativa, secuenciación, Northern blot y NanoString, en modelos animales, líneas celulares y muestras de tejido humanas. Por ejemplo, en 2002, el primer informe sobre la función de los miRNA en el cáncer mostró que en la mayoría de las personas con leucemia linfocítica crónica (LLC), los miRNA miR-15 y miR-16 están eliminados (84). Otros estudios revelaron que miR-15 y miR-16 funcionan como supresores de tumores al atacar el oncogén BCL2, que codifica una proteína de supervivencia celular (83). El miR-21 es otro miRNA, que se sobreexpresa en la mayoría de los tumores e incrementa la proliferación y disminuye la apoptosis, al atacar varios genes supresores de tumores, incluido PTEN (85). Las alteraciones en genes de la maquinaria de biogénesis de los miRNAs, como por ejemplo Drosha y Dicer, también se han asociado con la generación de varios tipos de cáncer, como el de ovario, CCU, esófago, y gástrico, entre otros (85). Un gran número de estudios han reportado la expresión alterada de miRNA en diversos tipos de cáncer: let-7, regulado negativamente en cáncer de mama, colon y pulmón, actúa como supresor tumoral al reprimir RAS o MYC. miR-34a, regulado por p53, y miR-331-3p, expresados a la baja en carcinoma nasofaríngeo, también tienen funciones supresoras de tumor. En contraste, la sobreexpresión de miR-21, miR-155, el grupo miR-17~19b y miR-106b-5p tiene función oncogénica en diversos tumores sólidos y neoplasias hematológicas (83, 85); miR221/222, miR-663, y miR-545 interactúan con blancos importantes en la proliferación celular; miR 17-92, miR-34, miR-122 están relacionados con la apoptosis y, miR-155, miR-200, miR-203, miR-9, y miR-212, con la invasión y metástasis celular (67).

El interés por la investigación de los miRNAs en oncología ha aumentado significativamente en los últimos años debido a su papel en la patogénesis del cáncer y su potencial como biomarcadores con potencial para el diagnóstico, tratamiento y pronóstico de la enfermedad. El hecho de que los miRNAs sean cortos y estables, les confiere una ventaja práctica sobre el RNA mensajero para la investigación en cáncer, y resultan muy convenientes como biomarcadores para el diagnóstico y pronóstico del cáncer.

4.2 MiRNAs en cáncer de cuello uterino

Un número considerable de estudios ha demostrado que los miRNAs regulan la expresión génica modulando la expresión de genes regulatorios importantes, lo que favorece la carcinogénesis del cérvix (86,87). Mediante la comparación de los perfiles de expresión de miRNAs, entre tejidos tumorales y normales del cuello uterino, se ha establecido una asociación entre la expresión aberrante de los miRNAs y el cáncer de cuello uterino (61). Adicionalmente, existe evidencia del papel de los miRNAs en las lesiones precursoras del cérvix, lo que ha llevado a estudiar el mecanismo molecular de los miRNAs en este tipo de cáncer.

La mayoría de estudios han identificado miRNAs desregulados en muestras de tejido del cérvix mediante análisis de microarreglos. Muchos de estos estudios se han enfocado en el uso diagnóstico y pronóstico de los miRNAs empleando PCR cuantitativa en biopsias de cérvix o investigando la contribución funcional de los miRNAs individuales al desarrollo del cáncer de cérvix (71,72). Sharma et al. realizaron una revisión comprensiva de los miRNAs desregulados en CCU (DEmiRNAs), y encontraron 246 DEmiRNAs reportados en 51 estudios, de los cuales sólo 41 han sido reportados en 5 o más estudios, y 116 han sido reportados en un solo estudio (56). He y colaboradores realizaron un metaanálisis en el que, luego de excluir del análisis los miRNAs reportados en un único estudio, encuentran 42 miRNAs sobre regulados entre diferentes estadios del cáncer de cérvix, 21 regulados a la baja y 35 con resultados variables (46). Pardini et al, realizaron una revisión sistemática en la que evaluaron los miRNAs como marcadores de progresión en CCU, incluyeron 24 estudios, 18 que emplearon PCR cuantitativa en tiempo real basada en miRNAs seleccionados de otros estudios, cinco basados en microarrays y uno por RT-PCR cuantitativa. Se identificaron 17 miRNAs sobre regulados asociados con la progresión a cáncer de cérvix y 13 regulados a la baja. Sin embargo, la coincidencia entre los DEmiRNAs descritos en diferentes estudios, es muy pequeña (61). Recientemente se han reportado varios estudios de secuenciación masiva en paralelo para determinar los perfiles de miRNAs en la carcinogénesis del cuello uterino, usando diferentes tipos de muestra: exosomas en plasma (60, 63), líneas celulares derivadas de CCU (57), biopsias (62,66) suero (62), cultivos organotípicos (74), y material de autotoma de citología (65,89). Al igual

que en los estudios basados en microarrays, la coincidencia en cuanto a los perfiles de miRNAs descritos es poca, y esto puede explicarse en parte por los diferentes tipos de muestra empleados, puesto que es bien sabido que los perfiles de miRNAs difieren según la muestra empleada para su análisis (90). A la fecha no encontramos en bases de datos como *Pubmed* reportes de perfiles de expresión de miRNAs en cepillados del cuello uterino de mujeres con lesiones preneoplásicas mediante secuenciación, adicionalmente, existen pocos estudios acerca de la aplicabilidad clínica del análisis de expresión de miRNAs en cepillados cervicales, y los que hay, se han basado principalmente en el descubrimiento de firmas de miRNAs obtenidas a partir de microrarrays en biopsias de tumor y de lesiones preneoplásicas, con resultados variables (20, 64, 71, 87).

A medida que las investigaciones sobre miRNAs en CCU avanzan, la cantidad de ellos que están involucrados en su desarrollo aumenta notablemente. La tabla 1 resume los miRNAs que se han reportado como oncogénicos y supresores de tumores en CCU.

Tabla 1 MiRNAs reportados como oncogénicos o supresores de tumores en CCU. Información resumida de (91).

miRNAs oncogénicos	miRNAs supresores de tumores	
miR-21	miR-23b	miR-107
miR-20a y miR-106a	miR-329-3p	miR-145
miR-221-3p	miR-200c	miR-22
miR-1246	miR-132 y miR-183	miR-124
miR-150	miR-454-3p	miR-494
miR-31,	miR-486-3p	miR-let-7a
miR-222	miR-375	miR-101
miR-221	miR-296	miR-143/miR-107
miR-590-5p	miR-337	miR-139-3p
miR-10a	miR-424-5p	miR-138
miR-501	miR-140-5p	miR-379
miR-92a		miR-485
miR-181a-5p		
miR-19a/b		

Se han realizado numerosos estudios de análisis de expresión diferencial comparando tejido normal y CCU, NIC de bajo y alto grado, y NIC en diferentes grados, utilizando diferentes tipos de muestras, en los que se han encontrado algunos miRNAs

sobre expresados o sub expresados. Las Tablas 2 y 3 resumen los resultados de algunas revisiones sistemáticas que han reportado miRNAs desregulados al alza, o a la baja.

Tabla 2 Resumen de los miRNAs más relevantes encontrados en diferentes muestras provenientes de cáncer cervical, especificando su regulación (positiva o negativa) y los tipos de muestras donde se han identificado. Información resumida de (92).

MicroRNA	Regulación	Tipos de Muestras
miR-20a-5p	Positiva	Biopsias de tejido, células exfoliadas, suero
miR-92a-3p	Positiva	Biopsias de tejido, suero
miR-101	Positiva/Neg.	Biopsias de tejido, suero
miR-141-3p	Positiva	Suero, moco cervical
miR-144-3p	Positiva	Moco cervical
miR-155	Positiva	Suero, moco cervical
miR-205	Positiva	Suero, moco cervical
miR-375	Positiva/Neg.	Biopsias de tejido, células exfoliadas
miR-424	Positiva	Biopsias de tejido, células exfoliadas
miR-494	Positiva	Suero, células cervicales exfoliadas
miR-142-3p	Negativa	Suero
miR-218-5p	Negativa	Suero

Tabla 3 Resumen de miRNAs regulados positivamente y negativamente en el contexto del cáncer cervical (91, 93, 94).

Regulados positivamente	Regulados negativamente
miR-10a, miR-501, miR-1246, miR-221-3p,	miR-20a, miR-106a, miR-150, miR-31,
miR-221, miR-222, miR-92a, miR-181a-5p,	miR-23b, miR-329-3p, miR-132, miR-7,
miR-19a/b, miR-29a, miR-21, miR-25	miR-144, miR-375, miR-337, miR-296,
miR-106a, miR-185, miR-20b,	miR-424-5p, miR-140-5p, miR-139-3p,
miR-31, miR-155, miR-92b, miR-93,	miR-138, miR-107, miR-143, miR-101,
miR-146a, miR-200a, miR-27a	miR-206, miR-145, miR-214, miR-124,
	miR-22, miR-362, miR-494, miR-379,
	miR-485, miR-let-7a, miR-99a/b
	miR-375, miR-203, miR-34a, miR-23b, miR-100, miR-193b, miR-497

Las alteraciones en la expresión de los miRNAs son un mecanismo molecular crucial en el desarrollo del cáncer. Estos pequeños RNAs no codificantes pueden controlar la expresión de múltiples genes y regular las vías de señalización asociadas con la oncogénesis. Los miRNAs tienen potencial como biomarcadores en CCU, capaces de reflejar la evolución de la enfermedad, caracterizarla e incluso predecir la respuesta a los tratamientos farmacológicos (95).

Los virus participan en la regulación de la expresión génica de la célula hospedera, en parte debido a cambios en la expresión de miRNAs que, en la carcinogénesis del cuello uterino, parecen expresarse como resultado de la infección por el VPH, que contribuye con la carcinogénesis, por lo cual la detección de estas alteraciones en la expresión de miRNAs podría ser útil para diferenciar entre la infección por VPH, las lesiones precursoras de alto grado y, eventualmente, el cáncer (48,88,96). Adicionalmente, ciertos miRNA originados a partir del genoma del VPH parecen regular genes del hospedero esenciales para la infección viral y el mantenimiento del virus, que están relacionados con la progresión tumoral, la migración y la respuesta inmune, pero se requieren más estudios para entender estos mecanismos (9). Por el contrario, las oncoproteínas E6 y E7 del VPH han sido ampliamente estudiadas y su impacto en el genoma humano, regulando la expresión de varios genes, incluidos los miRNAs y sus dianas, está bien descrito. Todos estos eventos moleculares pueden convertirse en biomarcadores potenciales (9).

El análisis de los niveles de expresión de miRNAs en células del cuello uterino exfoliadas durante la toma de la citología cervicouterina abre nuevas oportunidades de investigación en biomarcadores moleculares para programas de prevención primaria, puesto que los miRNAs pueden regularse epigenéticamente mediante metilación del DNA y/o modificaciones específicas de histonas, y la infección por VPH, el genotipo viral y otros factores de riesgo carcinogénico introducen cambios en el estado de metilación del ADN celular (97). Dado que los miRNAs también están asociados con resistencia al tratamiento en CCU, nuevas combinaciones terapéuticas podrían incluir quimioterapia y radioterapia con inhibidores de miRNAs, para ofrecer a los pacientes un tratamiento personalizado, un diagnóstico temprano y mejores resultados clínicos (98,99).

Estos mecanismos genéticos y epigenéticos que interrumpen los procesos de control y conducen a la expresión anormal de miRNAs en el cáncer, junto con los estudios mencionados anteriormente, demuestran que los miRNAs pueden considerarse biomarcadores potenciales y sus perfiles de expresión podrían utilizarse como prueba de triaje en mujeres con VPH-AR+ para determinar el riesgo de desarrollar lesiones de alto grado.

5. Marco teórico

En los últimos 60 años, se han realizado avances significativos en la comprensión del vínculo entre infección y cáncer, lo que ha mejorado las posibilidades de prevención de la enfermedad. Muchas de estas infecciones son factores de riesgo que pueden modificarse, y ya existen herramientas de prevención efectivas. Entre los 11 patógenos clasificados como carcinógenos del grupo 1 por la Agencia Internacional para la Investigación del Cáncer (IARC), los más importantes son *Helicobacter pylori*, el VPH-AR, el virus de la hepatitis B (VHB) y el virus de la hepatitis C (VHC) (7). Juntos, estos cuatro patógenos representan más del 90% de los cánceres relacionados con infecciones en todo el mundo (100).

Después del *Helicobacter pylori*, el VPH es la segunda causa infecciosa de cáncer. Entre los 690.000 casos de cáncer atribuibles al VPH, 570.000 (83%) son de CCU, y de estos, 500.000 (72%) pueden atribuirse a los tipos de alto riesgo 16 y 18 del VPH (VPH-AR) y 120.000 (17%) a los tipos de VPH 31, 33, 45, 52 y 58 (101). Por lo que, en mayo de 2018, la OMS anunció un llamado mundial a la acción para eliminar el CCU mediante la ampliación de la vacunación, la detección y el tratamiento eficaces (7).

La infección por el VPH es la infección de transmisión sexual (ITS) más común en hombres y mujeres sexualmente activos. Esta infección tiene un impacto negativo en la vida social de las personas infectadas (102). Clasificado dentro de la familia Papillomaviridae, el VPH

es un virus de ADN bicatenario que tiene tropismo por células del epitelio escamoso, cutáneo o mucoso y se divide en dos categorías: de bajo riesgo (VPH-BR), responsable de la aparición de verrugas anogenitales y cutáneas, y de alto riesgo (VPH-AR), implicado en el desarrollo de varios tipos de cáncer, incluyendo el CCU, anal, vulvar, vaginal y peniano. La infección por VPH-AR es una causa necesaria pero no suficiente para el desarrollo del cáncer cervical, ya que la mayoría de las infecciones son controladas y se resuelven en aproximadamente 2 años, desde la adquisición de la infección, y persisten solamente en el 10% de las mujeres, aunque aún es materia de discusión si el virus se elimina completamente, o si más bien se queda latente en las células basales, lo que haría posible su potencial reactivación en algunos casos (7, 103). Cuando la infección persiste, puede inducir el desarrollo de lesiones precursoras, también llamadas neoplasias intraepiteliales cervicales (NIC) de bajo (NIC1) y alto grado (NIC2/3), que pueden progresar a carcinoma escamocelular si no se tratan.

Se estima que el lapso entre la aparición de la lesión precursora y el desarrollo de cáncer es de 5 a 15 años, y el pico de detección de la lesión es alrededor de los 30 años de edad. Aunque el VPH-AR es esencial para el desarrollo del CCU, existen otros cofactores que pueden contribuir, tales como, la iniciación de la vida sexual temprana, el tabaquismo, el número de embarazos, el uso prolongado de anticonceptivos hormonales, otras infecciones de transmisión sexual, y factores relacionados con el hospedero (7).

Los estudios epidemiológicos han demostrado que casi todos los casos de CCU contienen el genoma del VPH, lo que resalta el impacto de la infección por este virus en la etiología de esta enfermedad. La persistencia de la infección por VPH-AR y su potencial para inducir cambios epigenéticos, como la metilación del ADN, o la alteración en clusters de miRNAs juegan un papel crucial en la progresión del CCU, resaltando su potencial para el descubrimiento de potenciales biomarcadores para la identificación de lesiones de alto grado.

5.1 Virus del papiloma humano

Los virus de papiloma humanos (VPHs) son virus que infectan el epitelio escamoso y causan lesiones proliferativas en piel y mucosas (104, 105). Se han identificado más de

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

200 genotipos de VPH (Figura 3), los cuales se han clasificado con base en su potencial para progresar a cáncer como de alto, intermedio y bajo riesgo. (Figura 2) (101). Aunque la mayoría de las infecciones son asintomáticas, los VPH de alto riesgo están vinculados a la progresión de lesiones preneoplásicas a cáncer. Los VPH 16 y 18 son los tipos virales asociados con el desarrollo del CCU más predominantes y están presentes en más del 50% de los casos de CCU diagnosticados anualmente (9).

Figura 2 Árbol evolutivo basado en el alineamiento de secuencias para los genes E1, E2, L1 y L2 del VPH. En gris se observan cutáneos de bajo riesgo, mucosos de bajo riesgo en naranja, de alto riesgo en rosado, verde de alto y bajo riesgo, y azul lesiones cutáneas de bajo riesgo. Imagen tomada de (106).

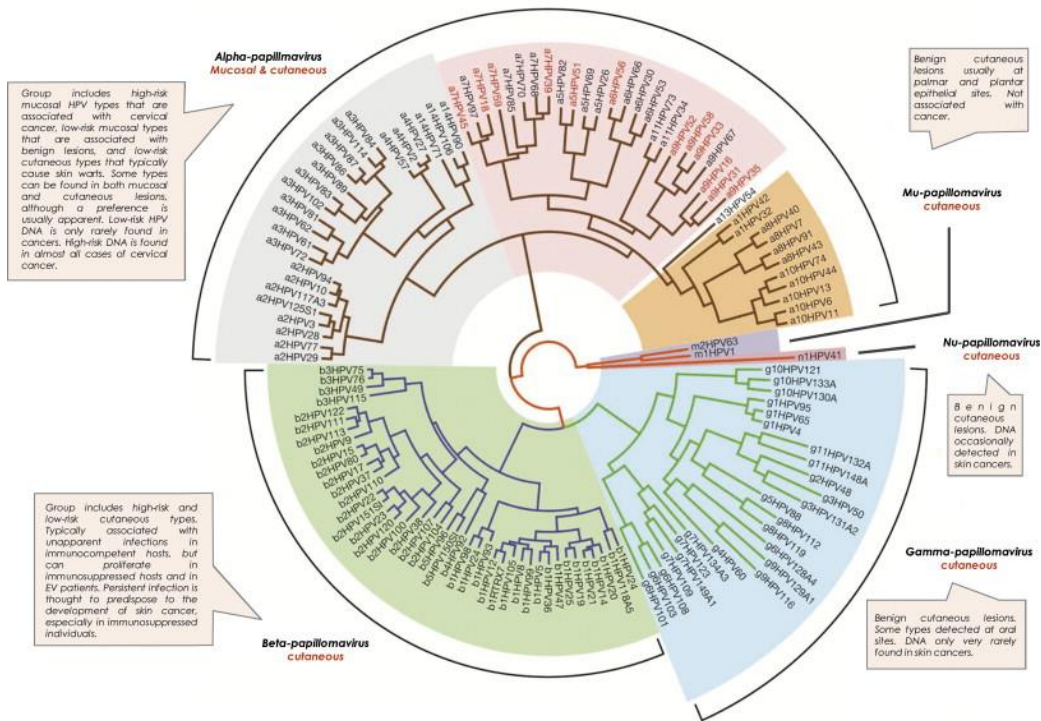
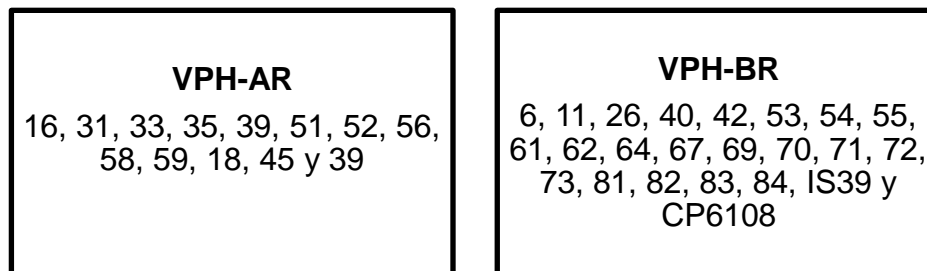


Figura 3 Clasificación de los VPH en alto y bajo riesgo.

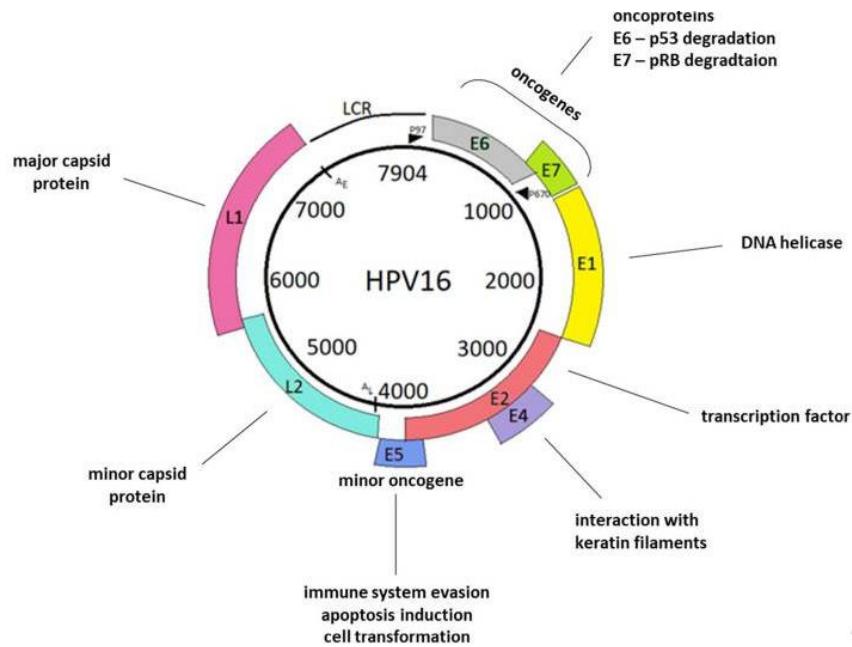


5.1.1 Estructura y genoma del VPH

Los VPH son virus icosaédricos de ADN bicatenario de la familia Papillomaviridae, desnudos, de 50 a 60 nm de diámetro, y su genoma contiene aproximadamente 8000 pares de bases que componen 3 regiones: una región reguladora (LCR) (107), una región temprana (E) que codifica las oncoproteínas (E6, E7, E5) y las proteínas reguladoras (E1, E2, E4), y una región tardía (L) que codifica proteínas estructurales (L1, L2) (Figura 4) (7).

Figura 4 Esquema del genoma del VPH-16. El genoma viral del VPH incluye los genes L1 y L2, que codifican la proteína principal de la cápside L1 y la proteína secundaria de la cápside L2; la región reguladora larga (LCR), que es la menos conservada del genoma y contiene el promotor p97 (sitio de inicio de la transcripción de genes tempranos virales) y varias secuencias que actúan como potenciadores y silenciadores de la transcripción viral. Los genes tempranos del VPH comprenden E1, E2, E4, E5, E6 y E7. En el genoma viral, también se observan sitios de poliadenilación tardía, AL; sitios de poliadenilación temprana, AE; y un promotor tardío, P670 (sitio de inicio de la transcripción de genes tardíos virales) (101). Figura tomada de (101)

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).



5.1.2 Funciones principales de las proteínas virales

Las proteínas E1 y E2, son esenciales para la replicación del genoma viral. E6 y E7 son las principales oncoproteínas del virus y están implicadas en la degradación de p53 y la inactivación de pRb promoviendo la transformación celular y la oncogénesis. La E5, aunque no esencial para la oncogénesis, contribuye a la invasión tumoral y la evasión inmune. Las proteínas L1 y L2, forman la cápside icosaédrica del virión. (101, 108).

Las proteínas E1 y E2, regulan la replicación del genoma viral y la transcripción de proteínas de expresión temprana, que mantienen bajos los niveles de las oncoproteínas E6 y E7, por acción de la proteína E2 que reprime su transcripción al unirse a sitios específicos en las secuencias promotoras del ADN. La pérdida de función de E2 puede conducir a la expresión desregulada de E6 y E7, que inician el proceso de transformación celular cuando el genoma del VPH se integra en las células del hospedero (101, 108).

5.1.3 Oncoproteínas

E6 y E7 juegan un papel fundamental en la carcinogénesis mediante la desregulación de los mecanismos de control del ciclo celular. E6 mediante su interacción con la proteína asociada a E6 (E6AP), dirige la degradación de p53 por la vía ubiquitina-proteosoma, afectando el control del ciclo celular, lo que contribuye a la tumorigénesis (109, 110, 111). E7 se une a pRb, e impide la represión de E2F, permitiendo la transición de la célula a la fase S del ciclo celular y la división de las células infectadas con VPH. E7 adicionalmente interactúa con la ADN metiltransferasa DNMT1, para inducir la hipermetilación y el silenciamiento de genes, lo que contribuye a la tumorigénesis (107).

Por su parte, la E5 es una oncoproteína accesoria que juega el papel de soporte, mediante la modulación de la señalización celular y la evasión inmune, sin que sea necesaria para la oncogénesis. (101). Esta proteína tiene la capacidad de incrementar la expresión del receptor para el factor de crecimiento epidermoide (EGFR) y el c-MET y altera la síntesis y función de las proteínas del complejo mayor de histocompatibilidad (MHC) de clase I y II promoviendo la evasión inmune al impedir el reconocimiento de las células infectadas por parte de los linfocitos T citotóxicos. In vitro, las proteínas E5 y E6, participan en la formación de coilocitos, marcadores morfológicos de la infección por VPH. Además, E5 puede contribuir a la transformación celular mediante la regulación de otras proteínas virales y la inducción de la invasión celular, aunque su papel exacto en la transformación neoplásica aun no es claro (101).

5.2 VPH y carcinogénesis cervical

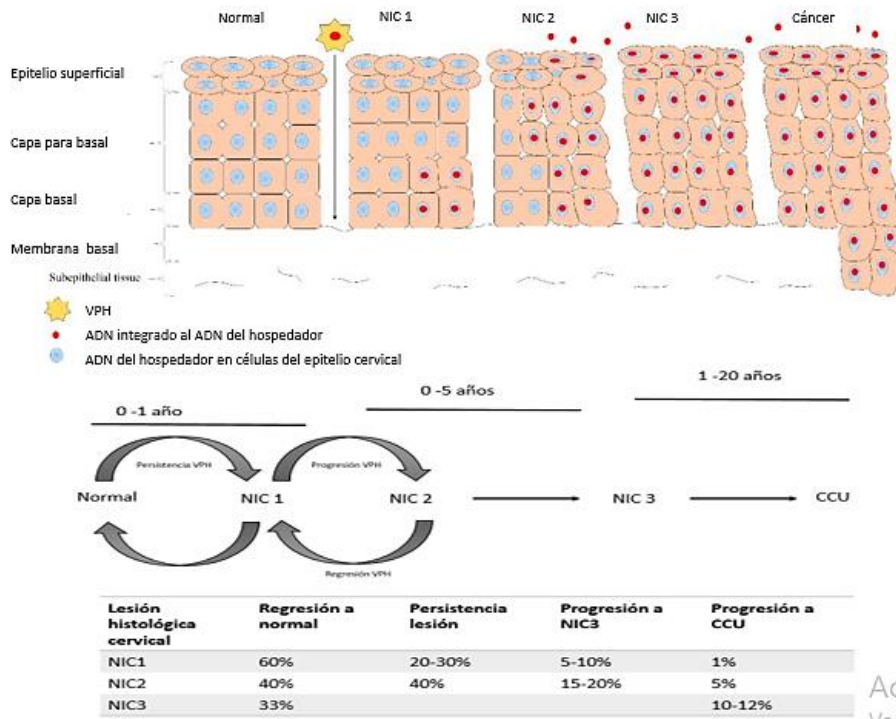
5.2.1 Canal cervical y anatomía del cuello uterino

El cuello uterino, conecta el útero con la vagina y su zona de transformación (112), donde se encuentran el epitelio escamoso y el columnar, es la zona más susceptible a la infección por VPH. (107). Dado que esta zona se renueva continuamente, es susceptible al desarrollo de displasias que pueden ser precursoras del cáncer de cuello uterino (106), particularmente del carcinoma de células escamosas (80-90% de los casos) (113). Factores como el envejecimiento y el uso prolongado de anticonceptivos, pueden aumentar el riesgo de infección persistente por VPH y el desarrollo de CCU (112).

5.2.2 Cambios patogénicos e historia natural del CCU.

La mayoría de las infecciones por VPH son transitorias y desaparecen en 12 a 24 meses en individuos inmunocompetentes (114), pero las persistentes, especialmente con VPH-AR, pueden conducir al desarrollo de lesiones preneoplásicas que pueden progresar a CCU. Las lesiones de bajo grado (NIC-I) tienden a regresar espontáneamente sin necesidad de tratamiento (solo el 5-10% progresan a HSIL) (115) (116), mientras que las de alto grado (NIC 2 /3) tienen mayor riesgo de progresión a CCU invasivo. La persistencia de la infección por VPH16, por ejemplo, aumenta significativamente el riesgo de progresión a cáncer (7) (117) (Figura 5).

Figura 5 Distribución de células epiteliales escamosas normales e infectadas por el VPH en lesiones precancerosas normales (displasia leve, moderada y grave; NIC 1, NIC 2 y NIC 3, respectivamente) y CCU. Figura adaptada de (116,119).

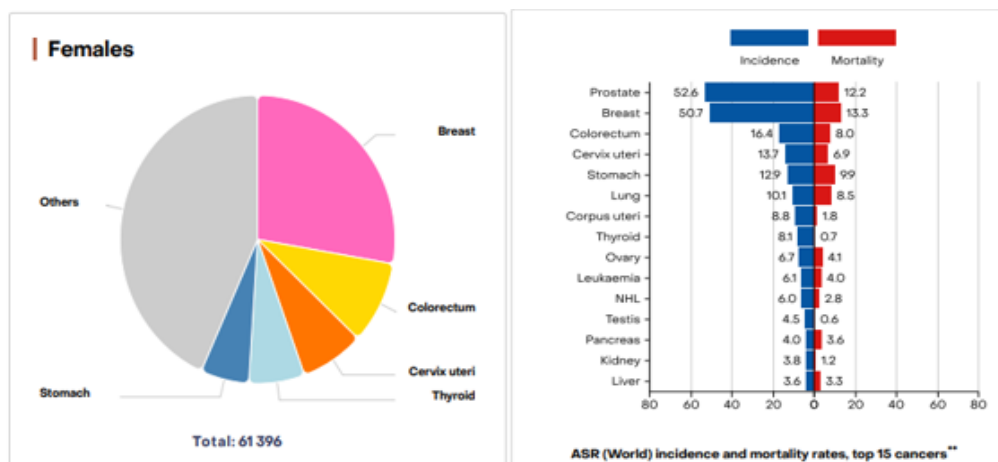


La detección temprana mediante citología y la intervención oportuna son clave para reducir la incidencia de CCU (7), y la vacunación contra el VPH ha demostrado ser efectiva en la prevención de nuevas infecciones. (7).

5.3 Cáncer de cuello uterino

El CCU es el cuarto cáncer más frecuente en mujeres a nivel mundial (3), con 604,000 diagnósticos y 342,000 muertes en 2020 (103). La incidencia y mortalidad son más altas (Figura 6) en países de ingresos bajos y medios, donde el acceso a medidas de prevención y tratamiento es limitado. (118). El 90% de los casos de CCU corresponden a carcinoma escamo celular el 10 % a adenocarcinomas (10%) (103), aunque se viene observando un incremento en la presentación de este último, particularmente entre las mujeres jóvenes en países desarrollados (119).

Figura 6 Incidencia y mortalidad del CCU en Colombia según Global Cancer Observatory: Cancer Today.



Dado que la detección temprana y el tratamiento oportuno aumentan significativamente las posibilidades de curación, la OMS busca eliminar el CCU utilizando tres estrategias fundamentales: La vacunación del 90% de las niñas antes de los quince años; el tamizaje

del 70% de mujeres a los 35 y a los 45 años y, el tratamiento adecuado del 90% de las mujeres con lesiones precancerosas o CCU) (118, 120).

5.4 Prevención del CCU

5.4.1 Prevención primaria

El desarrollo de vacunas profilácticas contra los tipos virales de VPH más cancerígenos ha sido crucial para la prevención del CCU y otras neoplasias asociadas al VPH (103). Inicialmente, se recomendaban tres dosis de la vacuna, pero se ha demostrado que esquemas de dos, o incluso una dosis proporcionan una protección similar a la del esquema original (7), demostrando que la eficacia de la vacunación es alta para la reducción del riesgo de desarrollar lesiones precursoras y carcinoma in situ (121). La vacuna nonavalente, aprobada en 2014, ofrece protección adicional contra otros tipos de VPH. La OMS recomienda la vacunación en una sola dosis antes del inicio de la actividad sexual con cobertura ampliada hasta los 45 años (7).

5.4.2 Prevención secundaria

El tamizaje basado en la detección molecular de VPH-AR es la estrategia más efectiva y tiene una mayor sensibilidad que la citología cervicouterina o la inspección visual con ácido acético. El tamizaje de la infección por VPH-AR reduce significativamente la incidencia del CCU e incluye el uso de otras estrategias de triaje, como la tinción dual de Ki-67 y p16, para identificar a mujeres que tengan lesiones con mayor riesgo de progresión. Una de las estrategias más efectivas y rentables para el cribado del VPH es la genotipificación parcial de HPV16 y 18, derivando a colposcopia a pacientes positivas y utilizando citología reflejo o triaje para otros genotipos. Otros biomarcadores, como FAM19A4 y miR-124, son alternativas efectivas para el triaje en regiones con acceso limitado a pruebas convencionales (7).

5.4.3 Prevención terciaria

El tratamiento del CCU en etapas tempranas es quirúrgico, mientras que, para los estadios avanzados, la combinación de radioterapia y quimioterapia sigue siendo el estándar.

Nuevas terapias como el bevacizumab, pembrolizumab y las inmunoterapias han mejorado la supervivencia, aunque el acceso a estos tratamientos sigue siendo un reto global (7).

5.5 MiRNAs

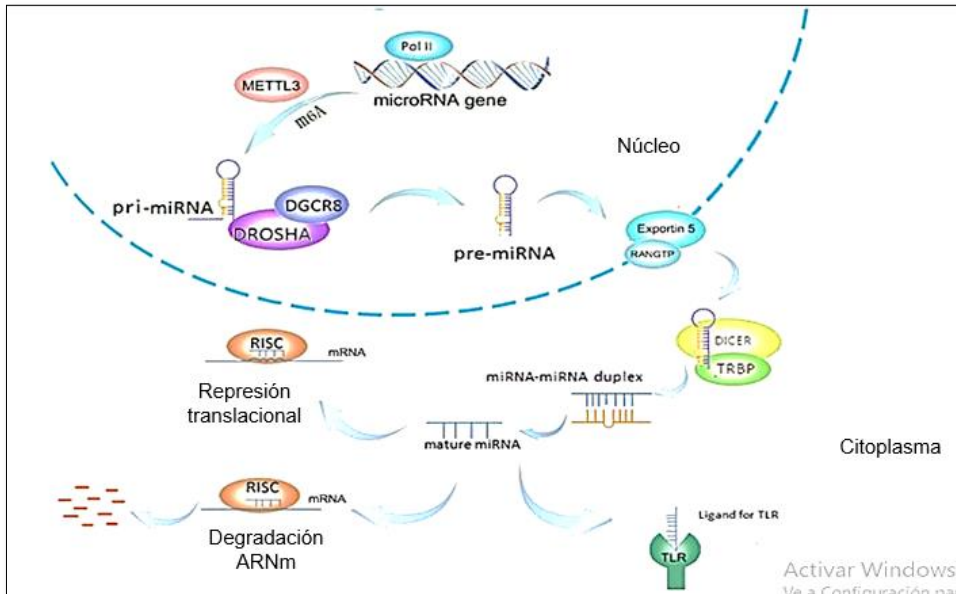
Los miRNA son moléculas pequeñas de RNA no codificantes, de 17 a 25 nucleótidos (122, 82) cuya función es la regulación postranscripcional de la expresión génica (82) al suprimir la traducción del RNA mensajero mediante su interacción con secuencias específicas en la región 3'UTR, lo que resulta en la regulación de alrededor del 30% de los genes humanos (122). Los miRNA desempeñan roles esenciales en procesos biológicos como el desarrollo, la proliferación y diferenciación celular, la apoptosis y el metabolismo (122, 82) y actúan como reguladores en enfermedades como el cáncer, donde pueden actuar como supresores tumorales o promotores al modular la proliferación celular, la transición epitelio-mesenquimal y la invasión tumoral (81). Las alteraciones en la expresión de los miRNA se asocian con el desarrollo de tumores y otras neoplasias malignas (84).

5.5.1 Biogénesis de los miRNAs:

La biogénesis de los miRNA (Figura 7) comienza con la transcripción del gen correspondiente por la RNA polimerasa II, generando un pri-miRNA que luego es procesado por Drosha, y exportado al citoplasma, donde Dicer lo convierte en un miRNA maduro. El miRNA maduro se une al complejo RISC, que dirige la represión o la degradación del RNAm objetivo (68, 83, 85).

Inicialmente se pensaba que los miRNA regulaban negativamente la expresión génica en el citoplasma mediante degradación específica de RNAm o represión traduccional en la región 3' UTR, pero se han descubierto funciones adicionales de los miRNA en la regulación de la transcripción génica en el núcleo (79). En cáncer los miRNAs están involucrados en la progresión del tumor por alteraciones en su biogénesis y por su capacidad para modular la apoptosis y promover la proliferación celular y la angiogénesis (84).

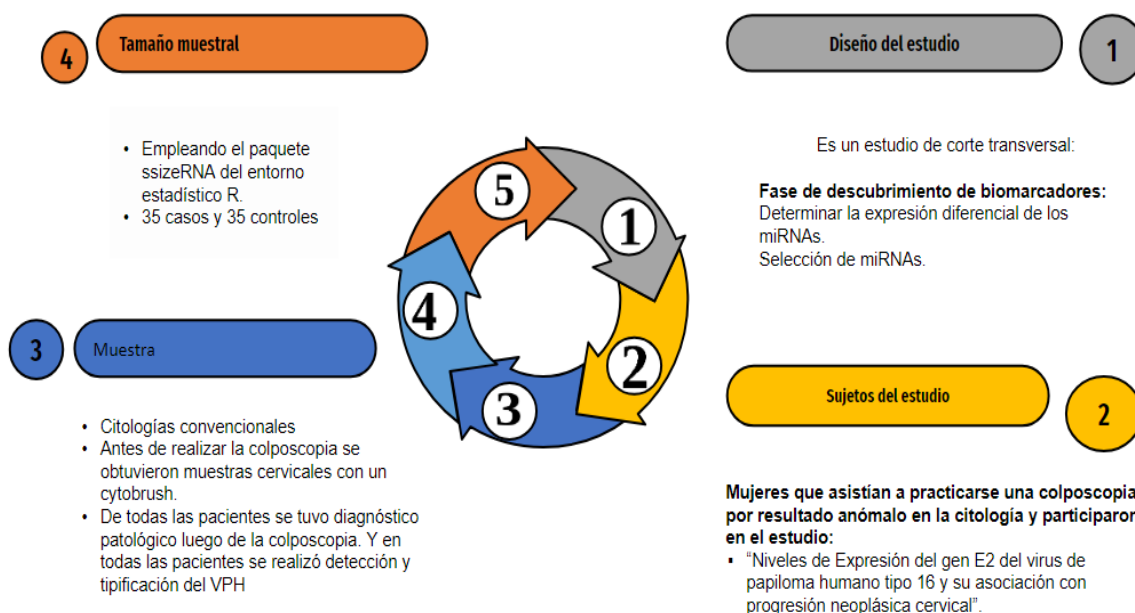
Figura 7 Biogénesis de los miRNAs (67).



6. Metodología

La metodología se muestra en tres fases distintas que abarcan el diseño del estudio, los procedimientos experimentales y los procedimientos de análisis de datos para cada objetivo. La Figura 8 presenta de manera resumida el diseño del estudio.

Figura 8. Diagrama de las principales fases del diseño del estudio para la selección y obtención de las muestras.



6.1 Diseño del estudio

Estudio de corte transversal, que contempla dos fases importantes: la fase 1 de descubrimiento de biomarcadores y la fase 2 de validación. El presente proyecto contempla únicamente la fase 1, que consistió en la determinación del perfil de expresión de miRNAs en mujeres con lesiones de bajo grado (NIC1 o menor) y alto grado (NIC 2/3) positivas para VPH-AR mediante secuencia masiva en paralelo del RNA (RNAseq), en esta fase se seleccionaron los miRNAs que presentaron el mayor poder discriminatorio para las lesiones de alto grado (NIC2/3).

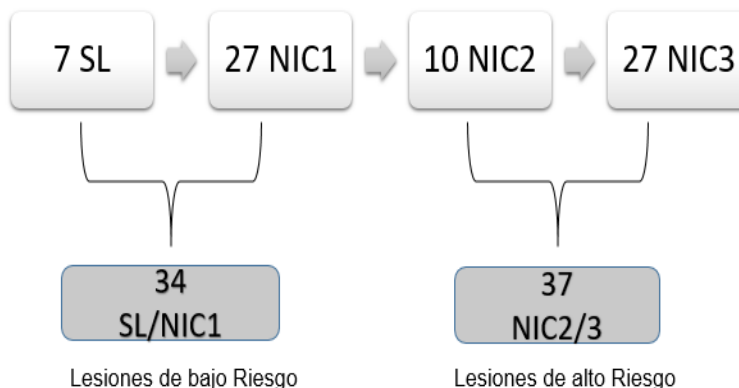
6.2 Descripción de los sujetos del estudio

Las muestras analizadas corresponden a cepillados cervicales obtenidos de participantes del estudio “Niveles de Expresión del gen E2 del virus de papiloma humano tipo 16 y su asociación con progresión neoplásica cervical” durante los años 2010 y 2011 en el Instituto Nacional de Cancerología (INC). Fueron procesadas y almacenadas en alícuotas a -80°C hasta su uso y se recolectó la información del diagnóstico reportado en la citología convencional y la biopsia guiada por colposcopia. Se realizó la detección y tipificación del VPH.

Se empleó el paquete **ssizeRNA** del entorno estadístico R para el cálculo del tamaño de muestra y el poder estadístico, utilizando 35 controles y 35 casos. El diseño experimental incluyó la secuenciación masiva de RNA en paralelo para identificar genes con expresión diferencial entre los dos grupos. Se consideraron un mínimo de 5 lecturas por gen, una dispersión máxima de 0,28, un factor de normalización de 1 y un *Fold Change* (FC) de 2. Con estas condiciones, se alcanzó un poder del 80% para rechazar la hipótesis nula de igualdad de medias poblacionales, con un valor de FDR de 0,05.

Se incluyeron 71 muestras en total para la etapa de descubrimiento, seleccionadas conforme a la clasificación de la biopsia: 7 muestras correspondientes a SL, 27 a NIC1, 10 a NIC2 y 27 a NIC3. Al agruparlas en lesiones de bajo grado, se obtuvieron 34 muestras SL/NIC1 y 37 muestras NIC2/3 para las lesiones de alto grado (Figura 9). Los criterios de inclusión de las muestras fueron: Ser positivas para VPH-AR, tener datos completos de citología y patología, tener las variables sociodemográficas mínimas para análisis. La diferencia en el número de muestra por condición estuvo sujeta a la disponibilidad de la muestra y a los controles de calidad posteriores a la extracción de RNA.

Figura 9 Esquema de las muestras seleccionadas para el estudio. Se muestra la cantidad de muestras para cada condición y el agrupamiento que se realiza en dos condiciones para el análisis de expresión diferencial.



Las características sociodemográficas de las mujeres cuyas muestras componen el grupo de descubrimiento fueron evaluadas considerando variables como edad, ocupación, escolaridad, estrato socioeconómico, número de partos, número de parejas sexuales, uso de anticonceptivos y hábito de fumar. La comparación de estas características en la población de estudio se realizó mediante la prueba de Chi-Cuadrado considerando como significativo un p-value menor a 0.05. El análisis descriptivo se llevó a cabo utilizando el programa R, versión 4.3.2. utilizando la función `chisq.test` (123).

6.3 Consideraciones éticas

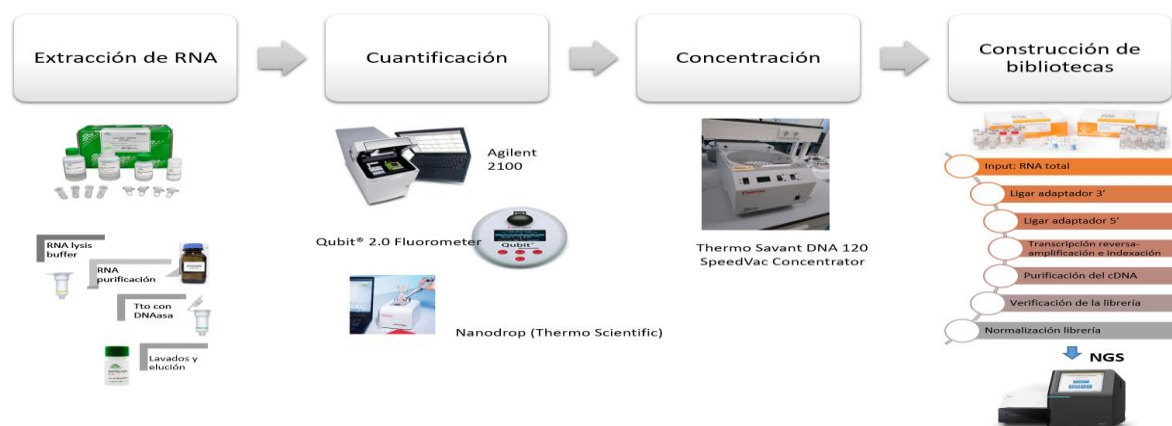
El proyecto se ajusta a la normatividad internacional (particularmente a la declaración de Helsinki y a las pautas éticas para la investigación biomédica preparadas por el Consejo de Organizaciones Internacionales de las Ciencias Médicas –CIOMS-) y a los parámetros establecidos en el ámbito nacional por la “Resolución 8430 de 1993” y “Resolución 2378 de 2008” (expedida por el Ministerio de la Protección Social). Este es un estudio sin riesgo dado que, la participación en el estudio no reviste riesgo para los sujetos de investigación, porque las muestras fueron recolectadas en un estudio previo y cuentan con consentimiento informado para usos futuros.

No se empleará ningún procedimiento que implique riesgo ecológico o ambiental. Los sujetos de estudio serán anonimizados, los datos del estudio estarán bajo custodia del investigador principal y del coordinador operativo.

6.4 Procedimientos experimentales para el cumplimiento de los objetivos

Para alcanzar los objetivos del estudio se llevó a cabo la fase experimental, la cual se centró en la aplicación de la técnica de RNAseq, como se resume en la (Figura 10). Esta metodología permitió obtener los datos de expresión de los miRNAs necesarios para el análisis y la consecución de los objetivos planteados.

Figura 10 Resumen de las técnicas y equipos empleados para la obtención de las bibliotecas.



6.4.1 Extracción del RNA

La extracción del RNA se realizó a partir de 500uL de muestra de los cepillados cervicales utilizando el kit comercial **Quick-RNA™ Miniprep Kit Catalog No. R1054 & R1055** de ZYMO Research, siguiendo las instrucciones del fabricante con algunas modificaciones y debido a que las muestras utilizadas han estado almacenadas a -80°C durante mucho tiempo, se mantuvo un estricto control de temperatura para optimizar la obtención del RNA requerido para la secuenciación.

6.4.2 Cuantificación del RNA total

Para determinar la concentración del RNA, se realizaron análisis con Agilent 2100 usando el kit RNA 6000 nano, el Qubit® 2.0 Fluorometer y el Nanodrop (Thermo Scientific).

Después de determinar la concentración del RNA total extraído para cada muestra, en algunas muestras concentró el RNA obtenido utilizando el Thermo Savant DNA 120 SpeedVac, para poder adicionar 1 µg de RNA total en un volumen de reacción de 5 µl para la construcción de la biblioteca.

6.4.3 Construcción de bibliotecas de secuenciación

Se utilizó el kit **TruSeq Small RNA Sample Preparation (Illumina)** siguiendo las instrucciones del fabricante que se resumen en la Figura 11. Para la generación de la biblioteca, se utilizó el RNA total extraído que contiene la fracción de RNAs pequeños. Para garantizar una concentración uniforme de RNA por muestra y tener un input inicial equivalente y normalizado, la biblioteca se construyó en grupos de cuatro muestras con concentraciones similares para mejorar la profundidad de la secuenciación y lograr que las muestras tuvieran un número de lecturas (reads) similares.

El RNA total se ligó con adaptadores RA3 (RNA3' Adapter) y RA5 (RNA5' Adapter) usando la T4 RNA Ligasa 2 Deletion Mutant y T4 RNA Ligase respectivamente. Se realizó la transcripción reversa usando la enzima SuperScript II Reverse Transcriptase de Invitrogen y se incubó la reacción a 50°C por 1 hora. Para comprobar la integridad del proceso se procedió a correr cada librería en el Bioanalyzer con el kit High Sensitivity DNA.

Posteriormente, se llevó a cabo una selección por tamaño mediante geles de poliacrilamida en condiciones no denaturantes al 6% (6% Novex TBE PAGE Gel, Invitrogen). Dado que las bibliotecas con índices únicos pueden ser agrupadas y purificadas conjuntamente en el gel, se procedió a realizar el pool de las muestras, y se tomaron 20 µl del cDNA amplificado de cada una de las cuatro muestras iniciales, en un tubo Eppendorf de 1.5 ml, junto con 20 µl del buffer de corrida Novex TBE (5x), para obtener un volumen final de 100µl. Se adicionaron 25 µl de la mezcla en cada uno de los 4 pozos del gel, para incrementar la concentración de las bandas que se cortaron. El gel se corrió a 145V por

aproximadamente 60 minutos o hasta que el frente de corrida abandonó el gel. Los fragmentos de cDNA con tamaños entre 145 y 160 pb, (correspondientes a miRNAs), fueron recortados del gel, transferidas a un tubo rompe-gel de 0.5 mL, centrifugadas a 17000 xg por 3 min., y eluidas en 200 µl de agua ultrapura. Posteriormente, el eluido y los residuos de gel se transfirieron a la parte superior de un filtro de 5 µm y se centrifugaron durante 2 minutos a 600 x g. El producto purificado se cuantificó utilizando el sistema Qubit y un Bioanalizador Agilent 2000 con el kit de alto rendimiento para ADN.

▪ **Normalización y desnaturalización del pool de la librería:**

Una vez cuantificada la concentración del pool de cDNA purificado a través del gel de poliacrilamida nativa al 6%, se realizó la conversión de unidades de nanogramos a nanomoles, utilizando la siguiente fórmula:

$$ng \text{ de la librería} \times \frac{1.000.000}{\text{longitud del fragmento} \times 660}$$

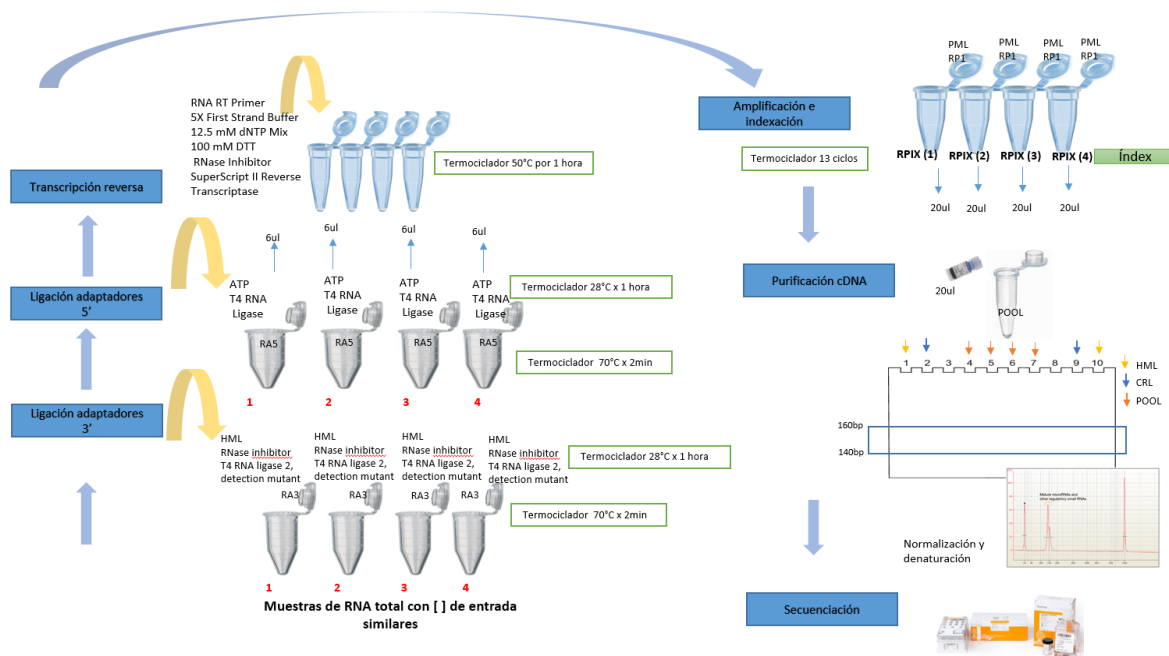
La concentración obtenida fue ajustada a 4nM en un volumen de 10ul, y se procedió a desnaturalizarlas. Se tomaron 5µl de las muestras de bibliotecas agrupadas ajustadas a una concentración de 4nM y se adicionaron 5µl de NaOH 0.2 N y se incubaron por 5 min. Se adicionaron 990 ul de HT1 a los 10ul de la librería denaturada para obtener la concentración final de 20pM. Se tomaron 600 µl de las bibliotecas y fueron sembrados en el cartucho del estuche comercial MiSeq Reagent Kit v2 (50-cycles). Finalmente, se realiza el montaje de las muestras en el cartucho de secuencia y este se instala en el equipo MiSeq Illumina® para realizar la secuenciación.

6.4.4 Secuenciación

Los templates identificados con sus index fueron secuenciados, usando el kit MiSeq Reagent Kit v2 (50-cycles) de Illumina en un equipo MiSeq con lecturas single end. Posterior al montaje, el equipo empieza a mostrar las métricas de calidad a partir del ciclo 25: El porcentaje de clusters que pasan el filtro (%PF: es una indicación de la pureza de la

señal de cada cluster) y el Q30, lo que da un primer acercamiento a la calidad de la secuencia.

Figura 11 Esquema detallado de las fases de la construcción de la biblioteca.



6.4.5 Análisis Bioinformático

Se siguieron algunas de las recomendaciones de la página RNA-seqlopedia (<https://rnaseq.uoregon.edu/#rna-prep>) para el manejo de los datos. Todo el procesamiento bioinformático se realizó en el sistema operativo Linux ejecutado en un sistema operativo Windows 11 y, utilizando el programa FASTQC versión 0.12.0., se realizó la verificación inicial de la calidad de la secuenciación y de las lecturas (124).

Con base en los ejemplos proporcionados para small RNA en la página de bioinformática del instituto barbraham (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), se revisaron las siguientes métricas: estadísticas básicas, calidad por secuencia de bases, puntuaciones de calidad por secuencia, contenido de N por base, distribución de la longitud de la secuencia).

Usando la herramienta Cutadapt (125), se procedió a cortar el adaptador 3' (TGGAATTCTCGGGTGCCAAGG), utilizado en el protocolo de secuenciación que se

aplicó, para asegurar la precisión en la identificación de secuencias genuinas de miRNA. Se seleccionaron secuencias con una longitud entre <15 nucleótidos y >27 nucleótidos para abarcar la longitud de los miRNAs y descartar transcritos no relevantes para el análisis. Se eliminaron caracteres comodines N, se filtraron las lecturas con calidad menor a Q30 y se descartaron las que no presentaban adaptador. Se evaluó la calidad de las lecturas mediante el análisis de FASTQC para verificar la calidad, longitud y correcta eliminación de adaptadores en las secuencias obtenidas.

Una vez depuradas las lecturas, se realizó la anotación y detección de miRNAs conocidos utilizando el software miRDeep2 que puede: (1) identificar miRNAs conocidos y nuevos en todos los clados principales de animales; (2) distinguir miRNAs de otros RNAs pequeños unidos a argonata; (3) informar miRNAs que pueden resistir la validación de alto rendimiento, y además de ser fácil de usar, es eficiente en consumo de memoria y tiempo (126).

Se procedió al mapeo de las lecturas contra el genoma de referencia GRCh38 versión 14. Esta tarea fundamental permitió obtener conteos detallados de lecturas asignados a cada miRNA en cada muestra, un paso esencial para el análisis detallado y preciso de la expresión de miRNAs. Para realizar esta tarea se empleó la herramienta mirdeep2, aprovechando sus distintas funcionalidades: los módulos como mapper.pl utilizan herramientas como Bowtie para alinear las secuencias de lecturas de RNA pequeño contra el genoma de referencia., quantifier.pl para cuantificar el perfil de expresión de forma precisa, y mirdeep2.pl para identificar con exactitud los miRNAs. Es importante destacar que todos los mapeos se realizaron considerando la información contenida en la base de datos miRBase versión 4.3.2, una fuente confiable y actualizada de secuencias de miRNAs. Los datos resultantes del mapeo y cuantificación se recopilan en una tabla de los conteos crudos.

Las lecturas no se alinearon con otras bases de datos debido a que no eran propósito del presente estudio. No obstante, se llevó a cabo una exploración adicional mediante el uso de herramientas en línea como sRNAtoolbox (<https://RNA.ugr.es/srnatoolbox/>) y miRmaster (<https://ccb-compute.cs.uni-saarland.de/mirmaster2>). Estas herramientas

ofrecieron información acerca de los reads similares a lo obtenido en miRbase, así como con otras bases de datos de RNAs pequeños y contaminantes.

6.4.6 Control de calidad tabla de conteos

Después de obtener la tabla de conteos, se llevó a cabo un control de calidad de los datos (n=71) con el propósito de analizar su comportamiento y distribución, identificar artefactos y detectar outliers que afecten la calidad de los datos. Utilizando el paquete ggplot2 del programa estadístico Rstudio versión 4.3.2 (127), se realizaron diversas técnicas para la exploración de los datos, tales como boxplots, correlogramas, gráficos de dispersión y análisis de componentes principales (PCA) (128).

Para reducir el ruido en los datos, primero se llevó a cabo un paso de filtración, que consistió en descartar los genes que tuvieran muy pocas lecturas o que no se alineó ninguna lectura. El criterio inicial fue eliminar las filas (los miRNAs) de la matriz que tuviesen menos de 70 conteos en la suma de todas las muestras y muestras con baja correlación.

Los datos de expresión de miRNAs fueron transformados a Counts Per Million (CPM), dividiendo el recuento de lecturas de un miRNA por el total de lecturas en la muestra y multiplicando por un millón, para normalizar los datos, ajustando las diferencias en la profundidad de secuenciación entre muestras. La transformación a CPM facilita la comparación directa entre muestras al expresar los conteos en una escala común. Posteriormente, los datos CPM se transformaron a $\log_2(\text{CPM} + 1)$ (logCPM) para pre-procesar los datos de expresión génica antes de realizar análisis, como clustering, análisis de expresión diferencial y visualización, para garantizar que los resultados del análisis reflejen las diferencias biológicas reales entre las muestras y no las técnicas.

Se aplicó el escalado Z-score para estandarizar los datos de modo que cada miRNA tuviera una media de cero y una desviación estándar de uno para la identificación de outliers. Adicionalmente, se utilizó la transformación regularizada logarítmica (rlog) disponible en herramientas como DESeq2. La rlog estabiliza la variabilidad de los datos de conteo y es eficaz para identificar artefactos técnicos. Estas transformaciones y normalizaciones combinadas permitieron realizar una evaluación de la calidad de los datos, e identificar y

corregir posibles artefactos y variaciones técnicas, asegurando la fiabilidad de los análisis subsiguientes.

6.5 Análisis de datos para el cumplimiento de los objetivos

6.5.1 Caracterización Inicial de las muestras

Se empleó el análisis de componentes principales (PCA) para la caracterización inicial de los datos que pasaron el control de calidad para comparar los resultados y obtener una visión clara de la distribución y el agrupamiento potencial de las muestras según los perfiles de expresión de miRNAs, facilitando la interpretación de los análisis. El PCA se realizó en Rstudio V 4.3.2 utilizando la función `prcomp`.

6.5.2 Análisis de Clustering

Para el agrupamiento y clasificación de las muestras de acuerdo con los niveles de expresión de los miRNAs, se utilizó el método de clustering k-means (129), empleando los datos de expresión transformados y filtrados como se indicó previamente. La visualización se hizo a través de la función `plot()` o `fviz_cluster()` en un diagrama de dispersión, donde los puntos se colorean según los clusters asignados por el algoritmo de k-medias. Esta aplicación permite asignar las muestras a grupos basados en las similitudes de sus perfiles de expresión génica, dividiendo el conjunto de datos en clusters distintos. Para determinar el número óptimo de clusters, se utilizó la función `fviz_nbclust` del paquete `factoextra` (130) en RStudio.

También se utilizó la función `s.class` de RStudio para graficar las muestras en el PCA y, para graficar el resultado obtenido por k-medias. Esta función proporciona una visualización clara de la distribución de las muestras en las condiciones experimentales que estamos investigando. Con esta función se representó gráficamente la distribución de las muestras en los clusters formados y se comparó con las condiciones experimentales conocidas. Además, se construyeron tablas de contingencia para detallar la cantidad de muestras de cada condición experimental asignadas a cada cluster. Se calculó el porcentaje de error del agrupamiento de los miRNAs seleccionados, dividiendo el número

de muestras mal clasificadas (errores) por el total de muestras clasificadas y multiplicando el resultado por 100 para obtener el porcentaje. La sensibilidad se refiere a la capacidad del modelo de clustering para identificar correctamente las muestras de NIC2/3, y se calculó como el número de muestras de NIC2/3 correctamente clasificadas entre el total de muestras de NIC2/3. Por otro lado, la especificidad mide la capacidad del modelo para identificar correctamente las muestras de SL/NIC1, y se calculó como el número de muestras de SL/NIC1 correctamente clasificadas entre el total de muestras de SL/NIC1.

Para guiar la identificación de los miRNAs con mayor poder discriminatorio en la detección de NIC2/3, se empleó el análisis de clustering con k-medias como herramienta clave en diversas etapas de la fase de descubrimiento con el fin de explorar cómo los miRNAs influían en la separación de las muestras en grupos distintivos (SL/NIC1 y NIC2/3), basándose en sus perfiles de expresión, sin necesidad de una variable de clase predefinida, lo que refinó la clasificación. El clustering se aplicó a diferentes conjuntos de miRNAs, incluyendo aquellos expresados diferencialmente, los regulados al alza, y los seleccionados como candidatos a partir de los demás análisis de datos mencionados en la sección 6.5.3, con el objetivo de identificar los miRNAs más efectivos para identificar patrones en los datos y verificar si los miRNAs seleccionados podían separar naturalmente las muestras en grupos que correspondieran a las categorías clínicas histopatológicas (como NIC2/3 y SL/NIC1), convirtiéndose en un excelente complemento para explorar el comportamiento de los miRNAs seleccionados durante el proceso.

6.5.3 Identificación de miRNAs candidatos

- **Análisis de expresión diferencial**

El análisis de expresión diferencial se realizó siguiendo el procedimiento establecido por el paquete DESeq2 de Bioconductor (131), dentro del entorno del programa R versión 4.3.2. La prueba estadística empleada por DESeq2 para llevar a cabo este análisis de expresión diferencial es la prueba de LRT, diseñada para comparaciones múltiples. Se consideraron miRNAs como diferencialmente expresados entre ambos grupos de comparación aquellos que presentaran valores de p ajustados menores de 0.05,

calculados mediante el método de Benjamini-Hochberg FDR, y con un Log2FoldChange mayor a 1.

▪ **Análisis de expresión diferencial con datos públicos:**

Se empleó la base de datos GEO Expression Omnibus del NCBI con el propósito de llevar a cabo un análisis exploratorio de otros conjuntos de datos que presentaran diseños experimentales afines al nuestro, permitiendo así comparaciones de los resultados obtenidos. Para la búsqueda de estos datos se utilizaron palabras clave como: miRNA, cáncer cervical, lesión cervical intraepitelial neoplásica, neoplasia intraepitelial cervical, VPH+; y se excluyeron datos provenientes de líneas celulares. Los conjuntos de datos identificados fueron sometidos a análisis de control y calidad, y aquellos que tenían diseños de estudio similares fueron seleccionados para los análisis. Utilizando el mismo flujo de trabajo de DESeq2, se reanalizaron dos conjuntos de datos públicos del NCBI-GEO con el fin de llevar a cabo análisis exploratorios y comparativos de los datos.

6.5.4 Análisis para identificar miRNAs con potencial utilidad para detectar NIC2/3

▪ **Análisis de correlación de miRNAs**

Se realizó un análisis de correlaciones para determinar qué miRNAs estaban más relacionados entre sí y cómo influían las coordenadas de cada componente principal. Un mayor grado de correlación entre los miRNAs separaría las mismas muestras, por lo que el objetivo era seleccionar miRNAs con correlaciones distintas. Se utilizó un "Círculo de correlaciones" mediante la función **fviz_pca_var()** que permite visualizar las cargas de los miRNAs en cada componente principal, identificando aquellos que contribuyen significativamente a la variabilidad observada en los datos. Además, se empleó la función **cor()** para visualizar la matriz de correlación de estos miRNAs, proporcionando una representación clara de las relaciones entre ellos.

- **Curvas ROC (Receiver operating characteristic curve)**

Se realizaron curvas ROC para evaluar el poder discriminatorio de cada miRNA candidato de la fase de descubrimiento. Se utilizó el software SPSS Statistics versión 25. Como datos de entrada se utilizaron los conteos de expresión de miRNAs normalizados. Los grupos de interés se definieron como SL/NIC1 versus NIC2/3. Usando la mediana de la expresión de cada miRNA se determinó su puntuación. Se asignó 1 si la expresión estaba por encima, y 0 si estaba por debajo de la mediana.

- **Regresión logística bivariada:**

Para identificar miRNAs asociados con lesiones preneoplásicas de alto grado de cérvix (NIC2/3), se realizó una regresión logística bivariada. Esta técnica estadística permite examinar la relación entre múltiples variables independientes (los niveles de miRNAs y variables socioeconómicas) y una variable dependiente dicotómica (la presencia de lesiones de alto grado= 1, frente a la ausencia, o presencia de lesiones de bajo grado= 0). Cada miRNA candidato de la fase de descubrimiento se evaluó individualmente para determinar su capacidad de predecir el riesgo de lesiones. El criterio para seleccionar los miRNAs candidatos fue tener un valor de $P \leq 0.05$, lo que indica una asociación estadísticamente significativa. Además, se evaluaron los Odds Ratios (OR) y sus intervalos de confianza al 95% (IC 95%). Un OR mayor a 1 indica un aumento en la probabilidad de la condición de interés, mientras que un OR menor a 1 indica una disminución. Para que exista una asociación estadísticamente significativa entre el miRNA y las lesiones preneoplásicas, el intervalo de confianza del OR debe excluir el valor 1. Si el intervalo incluye el valor 1, no se puede concluir que haya una asociación significativa entre la expresión del miRNA y el riesgo de desarrollar lesiones preneoplásicas. Este análisis se realizó en se utilizó el software SPSS Statistics versión 25.

▪ **Análisis de enriquecimiento funcional**

Para describir los genes diana de los miRNA, se utilizó desde el programa estadístico Rstudio V 4.3.2 el paquete `get_multimir` (132), una colección completa de interacciones miRNA-objetivo predichas y validadas y sus asociaciones con enfermedades y medicamentos. Tiene tres bases de datos principales para encontrar los genes blanco (mirecords, mirtarbase, tarbase), sin embargo, recopila información de más de 14 de bases de datos. Estas bases de datos proporcionan los genes blanco de los miRNAs candidatos de la fase de descubrimiento.

Con los genes blanco, se usaron las funciones de Clusterprofiler (133) y Enrichgo (134). El análisis de enriquecimiento se realizó utilizando las categorías de ontología genética: componentes celulares (CC), función molecular (MF), proceso biológico (BP) y vías metabólicas en KEGG. El método de Benjamini-Hochberg se utilizó para ajustar los p-valores y controlar la tasa de descubrimientos falsos.

6.5.5 Validación experimental

Extracción del RNA:

Se utilizó el kit Quick-RNA™ Miniprep Kit Catalog No. R1054 & R1055 de ZYMO Research, para el aislamiento del RNA de cepillados cervicales, según instrucciones del fabricante. La concentración de los RNAs pequeños se determinó con el kit Qubit RNA assay kit en el fluorómetro Qubit.

Síntesis de cDNA:

La transcripción reversa de todos los miRNAs seleccionados se realizó en un formato multiplex, tal como lo describe Babion et al. (71). Para esto, se utilizó un grupo de cebadores específicos de los ensayos TaqMan MicroRNA y reactivos del TaqMan MicroRNA Reverse Transcription Kit. Se preparó un 'pool' de cebadores, combinando los primers específicos de transcripción reversa para cada uno de los miRNAs seleccionados, lo que permitió realizar la transcripción de múltiples miRNAs en una sola reacción.

Reacción de PCR cuantitativa:

La reacción de qPCR se realizó en el equipo QuantStudio 3 en formato de 96 pozos con los reactivos para Taqman para un volumen total de 10ul. Se utilizó como gen de referencia (control endógeno) el hsa-miR-423-3p. Todas las muestras se analizaron por duplicado y aquellas muestras con una media geométrica Ct >38 se consideraron negativas.

EL CT para cada muestra se obtuvo como el promedio entre los CT de la misma muestra. Los valores de umbral de ciclo (Ct) obtenidos se normalizaron contra el control endógeno para calcular los valores ΔCt ($\Delta Ct = Ct_{miRNA} - Ct_{genref}$, el ΔCt es la diferencia entre CT del gen de interés en la muestra, y el CT del gen de referencia). Posteriormente, se calcularon los valores $\Delta\Delta Ct$ comparando los ΔCt de las muestras de NIC2+ con los de las muestras de NIC1 ($\Delta\Delta Ct = \Delta Ct_{casos} - \Delta Ct_{controles}$, diferencia entre el ΔCt del gen de interés en la muestra, y el promedio del ΔCt del gen de interés en los controles para cada muestra, relación 2:1). Finalmente se halló el $2^{-\Delta\Delta Ct}$, y el Log2 de este valor para mejorar la visualización. Este método permitió evaluar las diferencias relativas en la expresión de los miARNs entre los dos grupos de lesiones.

Análisis estadístico

Se realizaron pruebas de normalidad sobre las variables numéricas continuas utilizando la prueba de Shapiro-Wilk, se usó la prueba de U de Man Whitney para la comparación de medianas entre los grupos NIC1 y NIC2/3, y Kruskal Wallis seguido de la prueba de Dunn para múltiples comparaciones (NIC1, NIC2, NIC3). Los análisis estadísticos y la elaboración de gráficos se realizaron con el software GraphPad Prism. Los miARNs se consideraron diferencialmente expresados cuando $p < 0.05$. Se presentaron los resultados en gráficos de dispersión con barras de error que muestran la media y el error estándar de la media (SEM). Los gráficos indican los niveles de expresión de cada microARN analizado en las muestras de NIC1 y NIC2/3, así como los valores p correspondientes a las comparaciones entre los grupos.

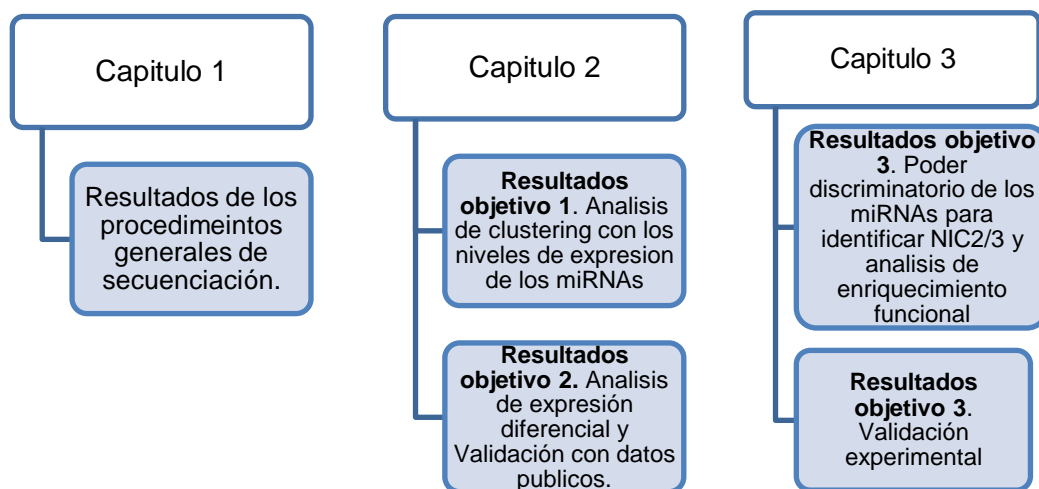
Se exploró la posible utilidad diagnóstica de cada miRNA individualmente para detectar NIC2/3 mediante la construcción de curvas ROC utilizando los valores de $\log_2(2^{-\Delta\Delta Ct})$. Los valores p fueron determinados por una prueba de distribución estándar con el software GraphPad Prism, los valores de $p < 0.05$ fueron considerados significativos para descartar

la hipótesis nula que hace referencia a una curva con un AUC de 50%. Se realizó un análisis de curvas ROC en Rstudio utilizando una firma de miARNs seleccionados por su buen desempeño individual en la separación de NIC2/3. Se recopilaron los valores de expresión de estos miARNs en $\log_2(2^{-\Delta\Delta Ct})$ para cada muestra y se ajustó un modelo de regresión logística multivariada que generó un score combinado. Este score estimó la probabilidad de que una muestra pertenezca al grupo NIC2/3 en función de la expresión de los miARNs combinados. El área bajo la curva (AUC) se calculó para medir la precisión del marcador, con su intervalo de confianza del 95% para evaluar su eficacia discriminativa. Finalmente, se determinaron los puntos de corte óptimos para maximizar la sensibilidad y especificidad del marcador, utilizando los métodos de Youden y MaxSp, que proporcionan umbrales clínicamente relevantes para la interpretación de los resultados.

7.Resultados

Los resultados se dividen en tres secciones o capítulos representadas en la Figura 12:

Figura 12 Esquema de la presentación de los resultados para el cumplimiento de los objetivos.



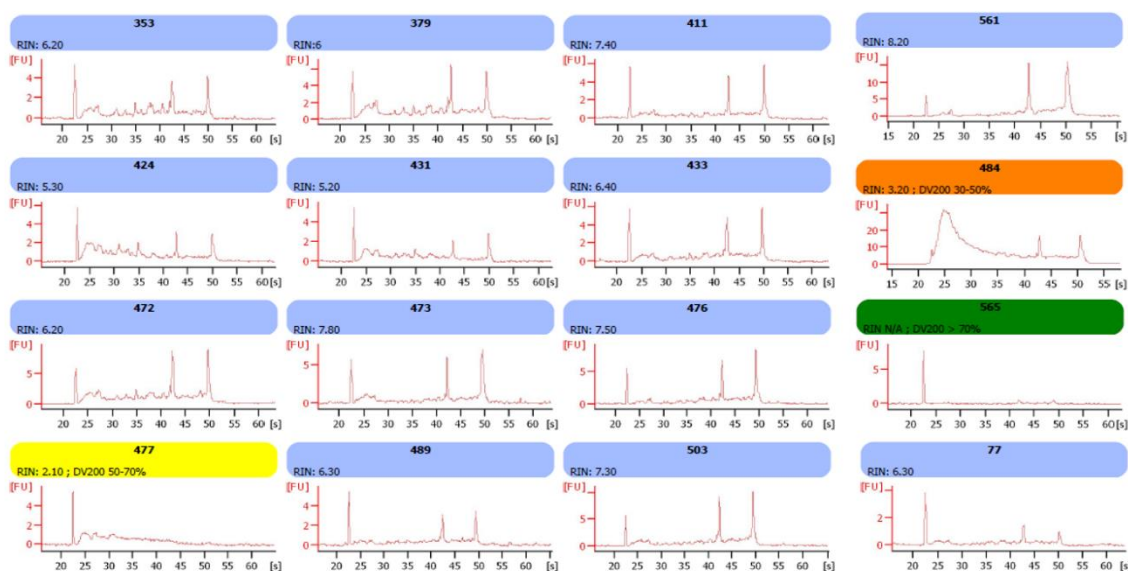
7.1 Capítulo 1. Resultados de procedimientos generales

7.1.1 Elaboración de bibliotecas de secuenciación:

Se seleccionaron un total de 103 muestras positivas para VPH-AR, abarcando lesiones de alto (NIC2-NIC3) y de bajo grado (SL-NIC1), para realizar la extracción de RNA total. La concentración inicial de RNA fue un factor crítico para la construcción de la biblioteca, porque no todas las muestras alcanzaban la concentración mínima de 1 µg. De este conjunto inicial, se escogieron 71 muestras con concentraciones (>600ug), junto con valores de RIN superiores a 6 y DV200 por encima del 30%. Esta selección se respaldó mediante un proceso estandarizado y ensayos previos que incluyeron muestras con concentraciones variables y valores de RIN bajos. Además, se consultó la bibliografía pertinente, que demostró la calidad de la secuenciación incluso con concentraciones bajas de RNA (135). Este enfoque permitió la inclusión de muestras con bajas concentraciones

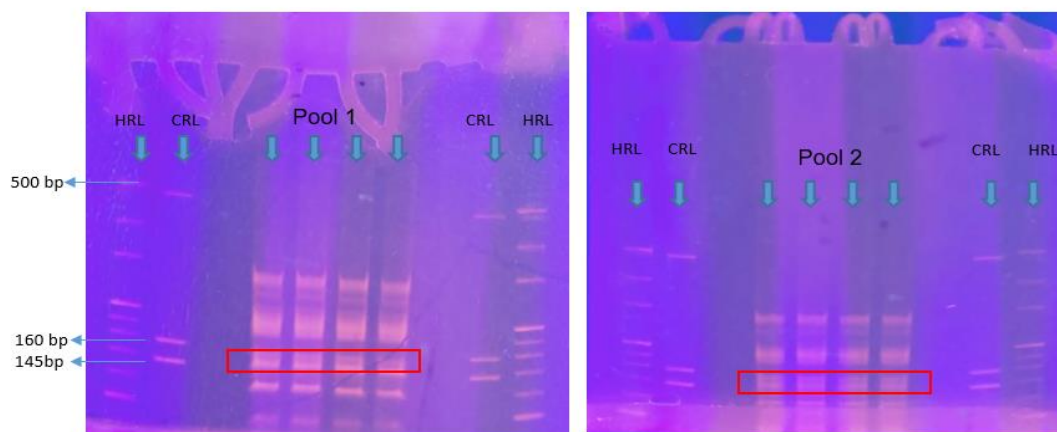
de RNA, se monitoreó rigurosamente la calidad (Figura 13). Debido a que los miRNAs constituyen una pequeña proporción en relación con el RNA total, es necesario trabajar con una buena concentración de material inicial. A pesar de las bajas concentraciones de RNA en algunas muestras, el proceso de construcción de bibliotecas se llevó a cabo de manera óptima, garantizando resultados confiables y de alta calidad.

Figura 13 Resultados de control de calidad del RNA total obtenidos con el 2100 Bioanalyzer. Se muestran algunas de las muestras y sus respectivas mediciones de calidad, evaluadas a través de las bandas 18S y 28S.



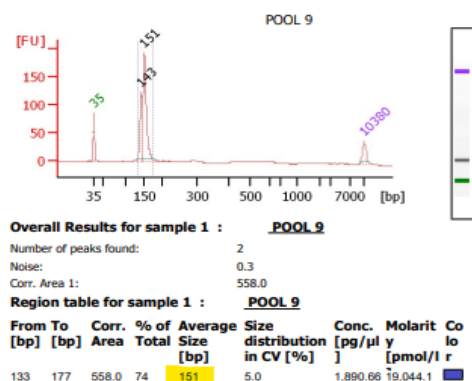
Se verificó la correcta elaboración de la biblioteca mediante la realización de un gel de poliácridamida nativa al 6% (Geles TBE Novex™, 6%), donde se observó un producto de purificación en un rango de 160 bp y 145 bp (Figura 14). La banda de 147 nt contiene principalmente miRNA maduro, generado a partir de fragmentos de RNA pequeño de aproximadamente 22 nts. La banda de 157 nt incluye RNA de interacción con Piwi, algunos microRNA y otras moléculas reguladoras de RNA pequeño generadas a partir de fragmentos de RNA de alrededor de 30 nt. Los demás productos inespecíficos observados corresponden a dímeros de primers. El producto amplificado se ubicó entre 160 y 145 pb, del marcador de peso molecular CRL Illumina®, confirmando la longitud esperada de las bibliotecas de small RNA y corroborando su adecuada elaboración.

Figura 14 Purificación del constructo de cDNA con Gel TBE Novex™ 6%. Se destacan las bandas obtenidas (señaladas en rojo), que corresponden a los microRNAs y RNA pequeños, conforme a las indicaciones de la guía de referencia del kit TruSeq Small RNA y al marcador de peso CRL de Illumina.



Luego de purificar el cDNA, se extrajeron las bandas de interés y se comprobó la longitud del pool de las bibliotecas mediante el Bioanalyzer 2100 empleando el DNA High Sensitivity Chip de Agilent (Figura 15). Los reportes en los que se observa un pico aproximado de 150bp corresponden a la longitud esperada para la librería y en donde se encuentran los miRNAs.

Figura 15 Cuantificación mediante Bioanalizador del purificado de la biblioteca.



7.1.2 Calidad de los datos de secuenciación

Las métricas de calidad de los datos obtenidos, presentaron una calidad igual o superior a Q30, considerado como de alta calidad (Figura 16). Los 18 pools realizados, lograron un porcentaje promedio del 90%, indicando una buena calidad de las secuencias.

En el proceso de control y calidad de los archivos FASTQ, se llevó a cabo la remoción del adaptador 3' (ATCTCGTATGCCGTCTTCTGCTT) de todas las lecturas y los filtros que se mencionan en la sección de materiales y métodos. En la tabla 4, se presentan de manera detallada las diferencias en el número total de lecturas antes y después del filtrado. Aunque se estimó una concentración teórica común inicial para cada pool, no se observó un porcentaje uniforme de lecturas para cada una de las bibliotecas realizadas. En promedio, se obtuvieron aproximadamente 3,9 millones de lecturas por muestra durante cada reacción y aproximadamente 1,394,429 millones de lecturas mapearon a secuencias de miRNAs maduros.

Figura 16 Valor Q para las lecturas de las muestras secuenciadas del pool 9. Se observa en %PF los grupos que pasaron el filtro Q30, el número de ciclos, y el rendimiento.

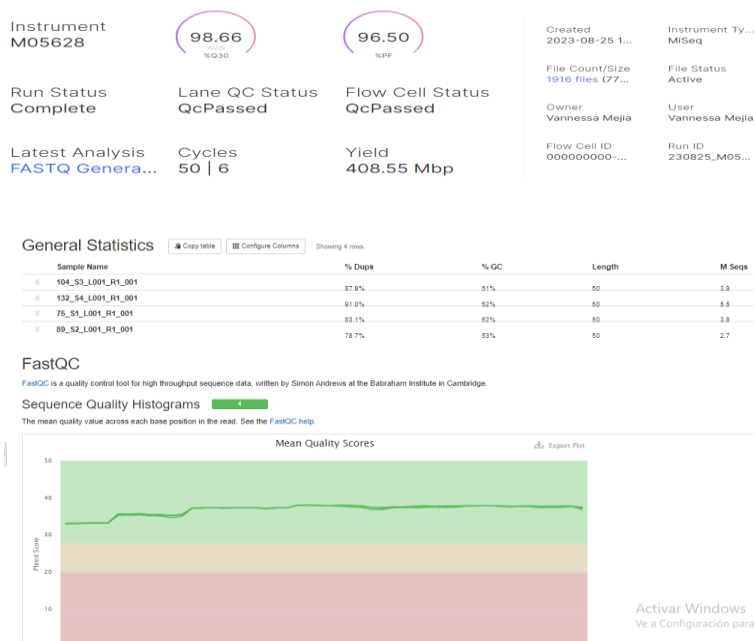


Tabla 4 Métricas del rendimiento de la secuenciación y el alineamiento/mapeo al transcriptoma de referencia (miRBase V.22.1).

Lecturas	Promedio
Totales	3.929.794
Total de secuencias filtradas y limpias	3.358.343
Lecturas eliminadas	571.472
Mapeados a miRbase	1,394,429
miRNAs con conteos por muestra	1807
Mapeados a miRbase	35,7%

7.1.3 Mapeo de los datos de secuenciación

La tabla 5 representa las muestras secuenciadas por tipo de lesión y tipo de VPH-AR.

Tabla 5 Número de muestras por lesión. Se llevó a cabo la secuenciación de RNAs pequeños de 71 muestras de cepillados cervicales todas VPH-AR+.

Lesión	Riesgo por lesión	por N° de muestras	Riesgo por VPH
SL	BR	7	AR
NIC 1	BR	27	AR
NIC 2	AR	10	AR
NIC 3	AR	27	AR
Total=		71	AR

A partir de los archivos FASTQ ya procesados, se procedió al mapeo de estos contra el genoma de referencia GRCh38 versión 14, para obtener conteos de lecturas asignados a

cada miRNA en cada muestra. Se empleó la herramienta mirdeep2, aprovechando sus distintas funcionalidades: los módulos mapper.pl para procesar los archivos FASTQ, y quantifier.pl para cuantificar el perfil de expresión de forma precisa. Todos los mapeos se realizaron considerando la información contenida en la base de datos miRBase V.22.1. Las lecturas mapeadas correspondieron a 2656 miRNAs maduros que contiene la base de datos miRBase V.22.1., y de estos 1807 contenían conteos (Tabla 4). Finalmente, se generó una tabla de conteos con 2656 microRNA en las filas y 71 muestras en las columnas. La tabla de conteos representa la cantidad de lecturas asignadas a cada gen en cada muestra.

7.1.4 Control de calidad de la tabla de conteos

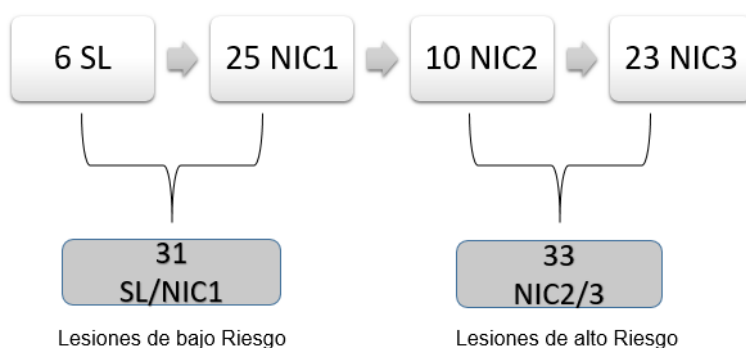
Se realizó un análisis exploratorio de calidad con el objetivo de entender el comportamiento y la distribución de las muestras de acuerdo con los niveles de expresión. A través de gráficos como boxplot, correlogramas, diagramas de dispersión y gráfico de análisis de componentes principales (PCA), se identificaron patrones que podrían indicar variabilidad extrínseca, y basados en estos se aplicaron filtros específicos para mitigar esta variabilidad y asegurar la robustez de los datos para análisis subsiguientes.

Se realizó un filtrado de los transcritos eliminando aquellos miRNAs cuyos recuentos fueran menores a 40 en todas las muestras, considerando que al menos hubiese un recuento por muestra; y también se filtró por microRNAs que no presentaran varianza. Estos umbrales son más bien arbitrarios, y se exploraron diferentes límites para filtro de bajos recuentos (10, 20, 30, 40, 50, 60, 70), observando resultados comparables en el análisis de expresión diferencial.

Para corregir las diferencias en los datos crudos a causa de variaciones en la profundidad de secuenciación, se aplicaron métodos de transformación y normalización antes del análisis principal. Los datos se transformaron a CPM (conteos por millón), y se transformaron tomando el logaritmo en base 2 para estabilizar la varianza. Luego, se normalizaron a Zscore, y por último se aplicó la transformación regularizada logarítmica (rlog), disponible en herramientas como DESeq2. Con cada transformación, se realizaron los gráficos de control de calidad para eliminar aquellos que generaban ruido.

Se eliminaron las muestras: SL_380, NIC1_564, NIC1_113, NIC3_477, NIC3_431, NIC3_411 y NIC3_277 debido a su baja correlación en diferentes etapas del proceso de control de calidad, baja profundidad, baja calidad y datos atípicos. Después de aplicar los filtros mencionados anteriormente, obtuvimos una matriz de conteos de 717 miRNAs y 64 muestras distribuidas según la Figura 17 y, se realizaron los siguientes análisis.

Figura 17 Cantidad y distribución de las muestras por lesiones después de realizar los filtros de control y calidad.



Las características sociodemográficas de las mujeres a las que pertenecen las muestras de cepillado cervical del grupo de descubrimiento se resumen en la Tabla 6. Se evidencia que las categorías aplicadas a cada una de las variables no influyen en el diagnóstico histopatológico, pues en la prueba de chi-cuadrado no se obtuvo ningún valor p significativo ($p < 0.05$), lo que elimina posibles factores de confusión en los resultados, para que los biomarcadores (miRNAs) sean verdaderos predictores del diagnóstico.

Tabla 6. Descripción de la población de estudio. *Se aplicó la prueba de chi-cuadrado.

Variable socio-demográfica	<i>SL/NIC1</i>		<i>NIC2/3</i>		<i>Todos</i>		<i>Valor p*</i>
	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	
Edad (años)							
Menor 30	16	61,54	10	38,46	26	36,11	0,13
Mayor 30	15	39,47	23	60,53	38	52,78	
Ocupación							
ama de casa	7	36,84	12	63,16	19	26,39	0,52
técnico	1	33,33	2	66,67	3	4,17	

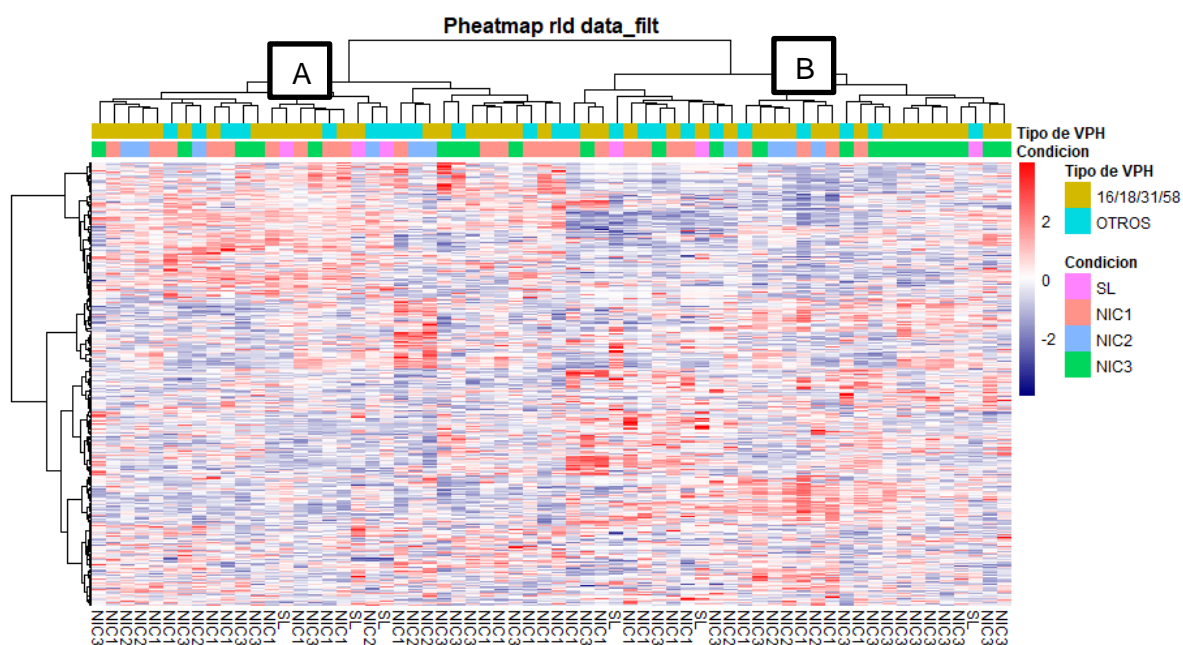
Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

<i>profesional</i>	9	50	9	50	18	25	
<i>otras</i>	14	58,33	10	41,67	24	33,33	
Educación							
<i>primaria</i>	3	33,33	6	66,67	9	12,5	0,66
<i>secundaria</i>	7	50	7	50	14	19,44	
<i>técnica</i>	5	41,67	7	58,33	12	16,67	
<i>universitaria</i>	16	55,17	13	44,83	29	40,28	
Estrato							
<i>Bajo</i>	3	33,33	6	66,67	9	12,5	0,13
<i>Medio</i>	25	48,08	27	51,92	52	72,22	
<i>Alto</i>	3	100	0	0	3	4,17	
Partos							
<i>0</i>	11	57,89	8	42,11	19	26,39	0,23
<i>1-2</i>	13	54,17	11	45,83	24	33,33	
<i>3+</i>	7	33,33	14	66,67	21	29,17	
Compañeros Sexuales							
<i>1-2</i>	9	40,91	13	59,09	22	30,56	0,54
<i>3+</i>	22	52,38	20	47,62	42	58,33	
Uso de anticonceptivos							
<i>Ninguno</i>	12	70,59	5	29,41	17	23,61	0,29
<i>Diu</i>	3	33,33	6	66,67	9	12,5	
<i>Hormonales</i>	9	45	11	55	20	27,78	
<i>Pomeroi</i>	4	40	6	60	10	13,89	
<i>Preserv</i>	3	37,5	5	62,5	8	11,11	
Tiempo uso anticonceptivos							
<i>1-48 meses</i>	26	63,41	15	36,59	41	56,94	0,16
<i>60 a 230 meses</i>	5	21,74	18	78,26	23	31,94	
Consumo de tabaco							
<i>Nunca</i>	24	50	24	50	48	66,67	0,73
<i>Fuma</i>	2	33,33	4	66,67	6	8,33	
<i>Dejo</i>	5	50	5	50	10	13,89	

La variabilidad de los conteos después de aplicar los filtros de control y calidad se representa en un mapa de calor en la Figura 18. El dendrograma de las columnas revela una clara división en dos grupos distintos. En el grupo A, se observa una expresión más elevada de los miRNAs, reflejada en colores cálidos predominantes en la parte superior del mapa. En contraste, el grupo B exhibe colores más fríos (menor expresión). Además, se observa una distribución leve de las muestras según su condición: el grupo A contiene

principalmente muestras de NIC1, mientras que el grupo B está compuesto en su mayoría por muestras NIC3.

Figura 18 Variabilidad de los conteos de las lecturas para cada miRNA (n=717) en cada una de las muestras (n=64). Realizado con la función pheatmap y el método de clustering Ward.D.



7.2 Capítulo 2. Clustering y Análisis de expresión diferencial

7.2.1 Resultados objetivo 1. *Realizar la caracterización y clasificación de las muestras de cepillados cervicales en: normales, lesiones de bajo grado (NIC 1) o lesiones de alto grado (NIC 2/3) mediante un método estadístico de clustering, con base en los datos obtenidos del análisis de expresión de miRNAs en las muestras.*

- **Caracterización y Clasificación de muestras de Cepillados Cervicales con base en los datos obtenidos de expresión de miRNAs:**

Para caracterizar las muestras de cepillado cervical, se utilizaron todos los miRNAs obtenidos después del filtrado de control de calidad (n=717). Se aplicó un análisis no supervisado y se evaluó la distribución de los conteos, entre las muestras, utilizando un PCA que reduce la dimensionalidad de los datos y visualiza las relaciones y las diferencias entre muestras. Esto proporcionó una visión clara de la distribución y el agrupamiento de las muestras de acuerdo con la expresión de los miRNAs.

Para la visualización de las muestras en el gráfico de PCA se dividieron por las condiciones de acuerdo con lo planteado en el objetivo 1, SL(Azul) NIC1(rosa) y NIC2/3(verde), se observó que las muestras no presentan una agrupación clara en las componentes principales de acuerdo con su condición conocida (Figura 18-B-C). Las muestras de lesiones cervicales de bajo grado (NIC1) se agrupan junto con las de lesiones de alto grado (NIC2/3), y las muestras sin lesión (SL), se mezclan con ambas categorías, tanto para los datos transformados y normalizados en Zscore, como para la transformación logarítmica regularizada (rlog) y logCPM. Este fenómeno puede atribuirse a la baja variabilidad biológica que capturan las 2 primeras componentes (~32%) (Figura 19-A) y a la complejidad de las condiciones cervicales analizadas que muestran una alta variabilidad entre lesiones.

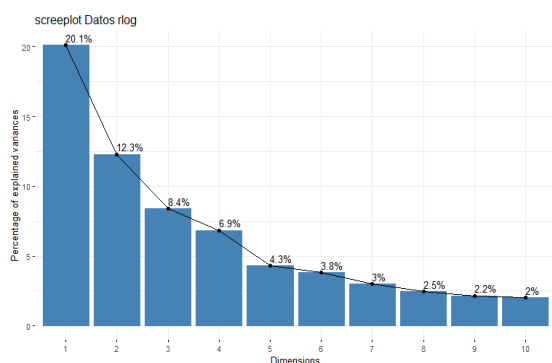
Al utilizar la transformación rlog, la variabilidad explicada en el scree plot era ligeramente mayor (Figura 19-A), y la visualización de los posibles grupos en el PCA era más clara en

comparación con las otras transformaciones utilizadas. Por esta razón, se decidió emplear la transformación rlog para los análisis subsecuentes (figura 19).

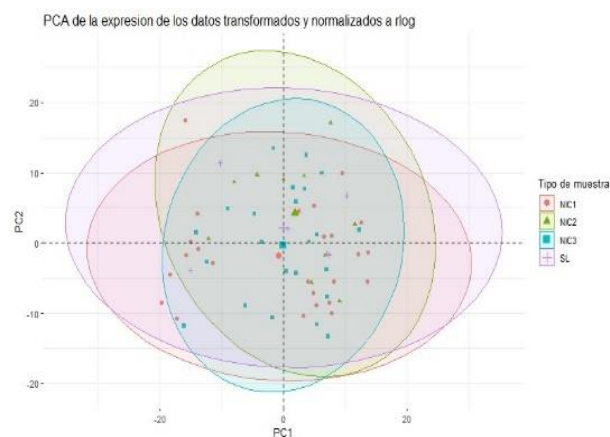
La transformación (rlog) separó ligeramente las muestras de alto grado NIC2/3 hacia la parte superior del plano, aunque no fue suficiente para resaltar las diferencias entre las condiciones (Figura 19-C). En la Figura 18-B, se observa la orientación de las muestras según su condición individual (SL, NIC1, NIC2, NIC3), lo que permite identificar sus patrones de distribución, lo que sugiere que simplificar el enfoque facilitaría la interpretación de los datos (Figura 19-C). Las diferencias en cuanto a miRNAs entre estas condiciones pueden no ser lo suficientemente distintivas para ser captadas por las dos primeras componentes del PCA y sugiere la presencia de otras características no capturadas por estas dos primeras componentes que influyen en la estructura de los datos.

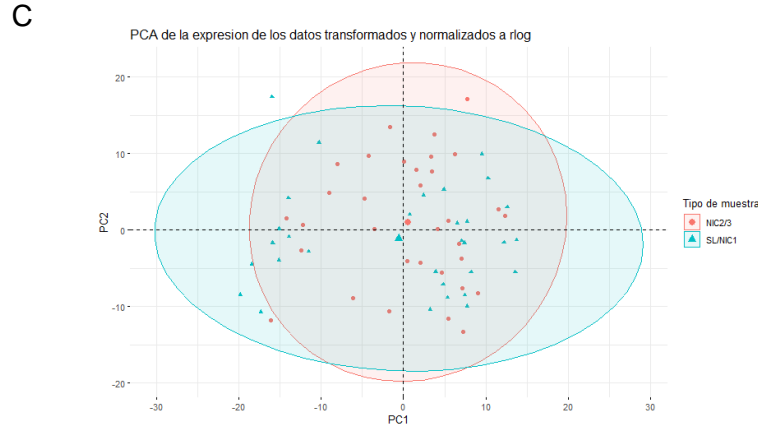
Figura 19. Análisis de PCA con los miRNAs luego del control de calidad (n= 717 y 64). (A) Variabilidad explicada para cada componente principal. Los componentes 1 y 2 explican la variabilidad en un 20% y 12%, respectivamente. fviz_eig se utilizó en Rstudio para visualizar los valores propios. (B) Datos transformados a rlog con las muestras separadas por condición individual. (C) Datos transformados a rlog con las muestras separadas en lesión de alto y bajo grado. La función prcomp de Rstudio se utilizó para realizar los PCA.

A



B

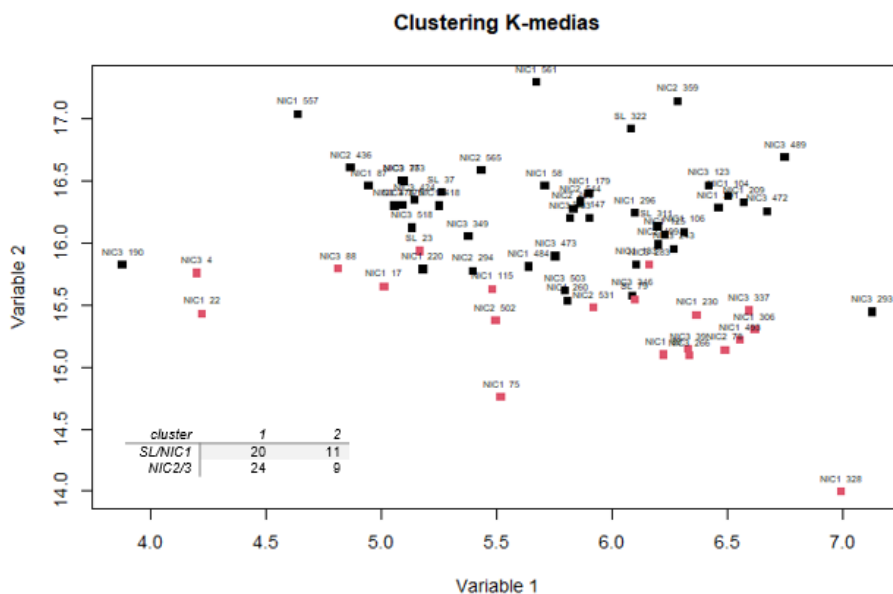




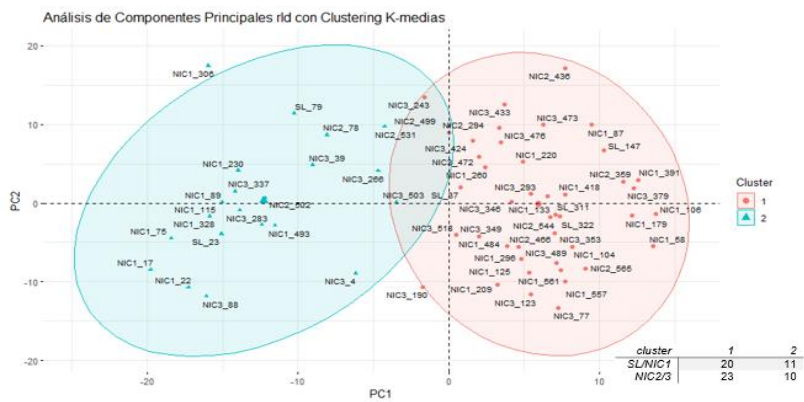
Debido a los resultados del PCA, se realizó el análisis de agrupamiento con k-medias utilizando la expresión de los microRNAs (N=717) sin reducir dimensiones, con el fin de capturar la variabilidad presente entre las condiciones del estudio. Exploramos diferentes configuraciones de clusters, incluyendo 2 y 3, de acuerdo con el cálculo de la inercia, para determinar su número óptimo. Los puntos fueron asignados a los clusters correspondientes y coloreados según su pertenencia, con la etiqueta de la muestra respectiva (Figura 20-A). Sin embargo, el análisis reveló que los clusters formados de las muestras individuales (SL, NIC1, NIC2, NIC3), no exhibían una tendencia clara de agrupamiento entre muestras de la misma condición o entre lesiones de bajo y alto grado. Se quiso explorar el resultado reduciendo las dimensiones con PCA y posteriormente aplicando el k-medias y el resultado del clustering fue igual (Figura 20-B), lo que significa que la reducción de dimensionalidad proporcionada por el PCA no estaría afectando significativamente el resultado del clustering. Esto confirma la variabilidad en la expresión entre las muestras, la cual no se ajusta completamente a las categorías de lesiones cervicales conocidas o tipos de VPH (Figura 20-C).

Figura 20 Análisis de clustering utilizando el algoritmo de K-medias para todos los miRNAs (n=717) (A) K-medias con dos clusters de datos transformados en rlog sin reducción de PCA. (B) K-medias con datos reducidos por PCA. (C) K-medias con 3 clusters y tipos principales de VPH-AR, datos transformados a Log2CPM. La función prcomp de Rstudio se utilizó para realizar los PCA.

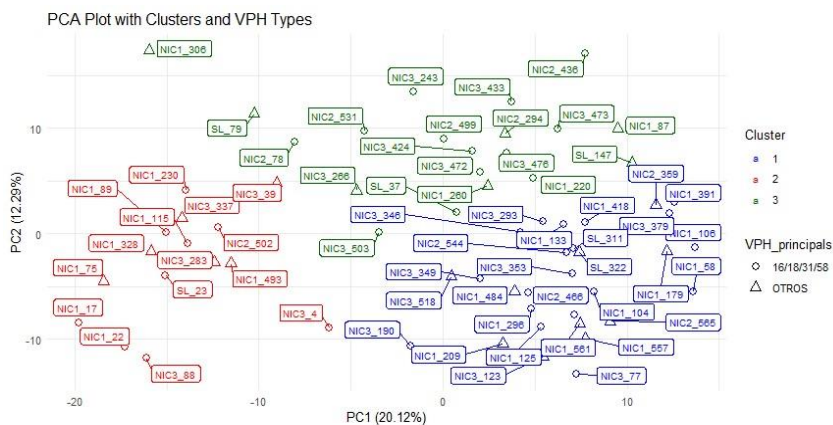
A



B



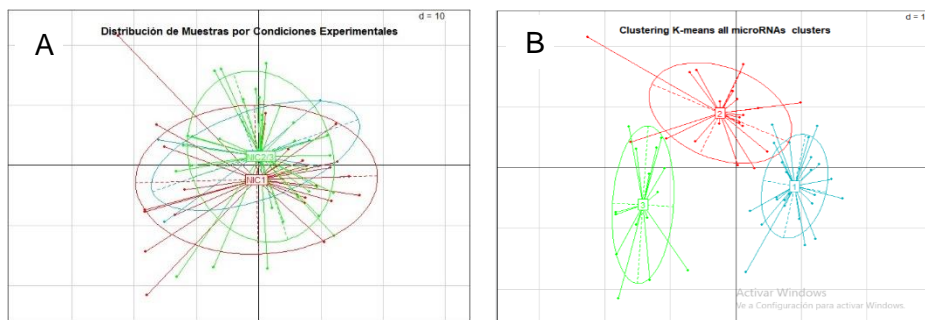
C



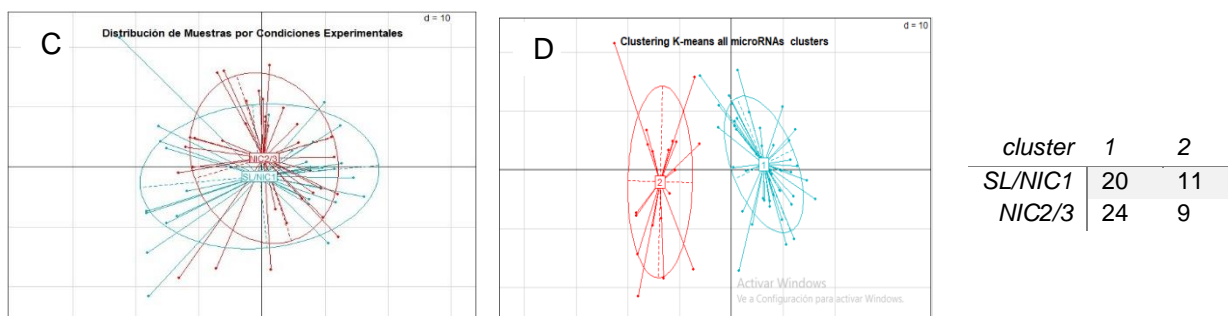
Se utilizó la función `s.class` para visualizar la distribución de las muestras en el espacio de componentes principales (Figura 21-A y C), y se confirmó que las muestras no se agrupan distintivamente según las condiciones experimentales conocidas. En la gráfica, se observa que las coordenadas de las muestras correspondientes a las condiciones "SL", "NIC1" y "NIC2/3" (Figura 21-A) están superpuestas al igual que en la Figura 21-C para dos condiciones, indicando que no hay una separación entre los grupos de acuerdo con estas condiciones.

Al realizar el k medias de las muestras, tanto en el escenario de tres clusters (Figura 21-B) como en el de dos (Figura 21-D), las elipses que rodean a las muestras indican la dispersión y variabilidad dentro de ese cluster. Adicionalmente, el k-medias no refleja patrones definidos de agrupamiento y en las tablas de contingencia realizadas, se observa que las muestras de las distintas condiciones están dispersas en los diferentes clusters, sin una asignación predominante a uno específico.

Figura 21 Análisis de clustering con K-medias para todos los miRNAs (n=717) usando la visualización con la función `s.class`. (A) Distribución de las condiciones (SL, NIC1, NIC2/3) en las coordenadas del PCA. (B) Distribución de las muestras por condición en 3 clusters. (C) Distribución de las condiciones (SL/NIC1, NIC2/3) en las coordenadas del PCA. (D) Distribución de las muestras por condición en 2 clusters.



<i>cluster</i>	1	2	3
SL	2	3	1
NIC1	13	14	8
NIC2/3	14	13	6



La variabilidad entre replicas observada en nuestro estudio refleja la complejidad inherente de los datos. Probamos la reducción de dimensiones con el algoritmo UMAP y el análisis de clustering general con DBSCAN (datos no presentados ya que son exploratorios), así como el K-medias sin reducción de dimensiones por PCA. Con todos estos enfoques, no logramos una agrupación clara de las muestras según su condición, utilizando los perfiles generales de expresión de miRNAs. Esto sugiere que existen factores adicionales contribuyendo a la variabilidad biológica en los perfiles de expresión. En nuestro caso, el tipo de VPH-AR, es un factor crítico porque muchas muestras contienen coinfecciones por VPH-AR.

Se realizó el clustering para caracterizar los datos y evaluar la separación entre las condiciones utilizando los perfiles de expresión diferencial de miRNAs. De esta manera, se puede mejorar la capacidad de clasificar las muestras en su categoría correspondiente de lesiones de bajo y alto grado. Dado que considerar todos los miRNAs de manera indiscriminada puede no ser informativo, hasta ahora el clustering ha servido principalmente para caracterizar los datos. La identificación de miRNAs diferencialmente expresados es crucial para refinar el análisis y mejorar la capacidad de clasificar las muestras en las categorías de lesiones correspondientes.

7.2.2 Resultados objetivo 2. Comparar los perfiles de expresión diferencial de los miRNAs obtenidos mediante la secuenciación masiva en paralelo a partir de cepillados cervicales provenientes de mujeres sin lesiones, con lesiones de bajo grado (NIC 1) y con lesiones de alto grado (NIC 2/3), positivas para VPH-AR.

▪ **Análisis de expresión diferencial de los datos:**

Se realizó el análisis de expresión diferencial entre los grupos de muestras: sin lesiones (SL) vs NIC1, y SL vs NIC2/3, tomando como grupo de referencia SL; también se comparó NIC1 vs NIC2/3 tomando como referencia a NIC1. En la tabla 7 se muestran los miRNAs expresados diferencialmente para cada una de las comparaciones:

Tabla 7 miRNAs expresados diferencialmente para SL vs NIC1, SL vs NIC2/3 y NIC1 vs NIC2/3. Los miRNAs resaltados en color azul son aquellos expresados a la baja y los de color rojo son los expresados al alza. (SL=6, NIC1=24, NIC2=10, NIC3=23).

SL vs NIC1	SL vs NIC2/3	NIC1 vs NIC2/3	
hsa-miR-4488	hsa-miR-221-5p	hsa-miR-1271-5p	hsa-miR-422a
		hsa-miR-1246	hsa-miR-193b-5p
		hsa-miR-9-5p	hsa-miR-210-3p
		hsa-miR-501-3p	hsa-miR-365a-5p
		hsa-miR-221-5p	hsa-miR-324-3p
		hsa-miR-99a-5p	hsa-miR-29b-2-5p
		hsa-miR-96-5p	hsa-miR-203a-5p
		hsa-miR-4662a-5p	hsa-miR-320b
		hsa-miR-196a-5p	hsa-miR-378g

Dado el desequilibrio en el número de muestras entre los estadios analizados, con un considerablemente menor número de muestras en la categoría "sin lesión" (SL) En comparación con las que tenían lesiones intraepiteliales de bajo (NIC1) y alto grado (NIC2/3). Es probable que esta disparidad haya influido en los resultados del análisis de expresión diferencial, ya que observamos una detección limitada de microRNAs diferencialmente expresados entre SL y lesiones NIC1 y NIC2/3, mientras que al realizar

la comparación entre NIC1 y NIC2/3, con un número de muestras comparable, se identifican varios microRNAs expresados diferencialmente tanto al alza como a la baja.

Para abordar el desequilibrio en el número de muestras y mejorar la robustez de los resultados, agrupamos las muestras en dos condiciones generales (bajo y alto grado): SL/NIC1 versus NIC2/3, para simplificar el análisis y resaltar las diferencias más relevantes. De esta manera, se facilitó la comparación entre lesiones de bajo y alto grado. Desde una perspectiva biológica, agrupar NIC1 y muestras sin lesión (SL) está justificado, por el potencial de regresión espontánea de NIC1 (8,16,65), mientras que las lesiones NIC2/3, al compartir características moleculares y morfológicas dadas por su potencial de progresión a cáncer, representan el grupo objetivo más adecuado para la evaluación de biomarcadores.

▪ **Análisis de expresión diferencial lesiones de las lesiones de bajo grado (SL/NIC1) vs alto grado (NIC2/3)**

Se realizó el análisis de expresión diferencial entre los grupos de muestras SL/NIC1 vs. NIC2/3, siguiendo el procedimiento establecido por el paquete DESeq2 de Bioconductor, dentro del entorno del programa R versión 4.3.2. La prueba estadística empleada por DESeq2 para llevar a cabo este análisis de expresión diferencial es la prueba de Wald que compara dos condiciones. Se detectaron 19 miRNAs con expresión diferencial entre las dos condiciones. 12 presentan una regulación al alza en NIC2/3 en comparación con el grupo de control (SL/NIC1), mientras que 8 exhiben una regulación a la baja en NIC2/3 con respecto al control, adicionalmente, se identificaron miRNAs compartidos con el análisis de expresión diferencial realizado previamente entre NIC1 vs NIC2/3 (Tabla 7). La presencia de estos miRNAs comunes sugiere que pueden estar implicados en la transición de lesiones de bajo a alto grado, y resalta la similitud entre las lesiones SL y NIC1, o bien, indica que las lesiones NIC1 pueden estar influyendo en la expresión observada. Estos miRNAs podrían servir como marcadores de riesgo para lesiones de alto grado, dado que muestran los mayores valores de razón de cambio (FC) y valores ajustados de significancia (padj). En la Tabla 8 se presentan los miRNAs expresados diferencialmente entre SL/NIC1 y NIC2/3.

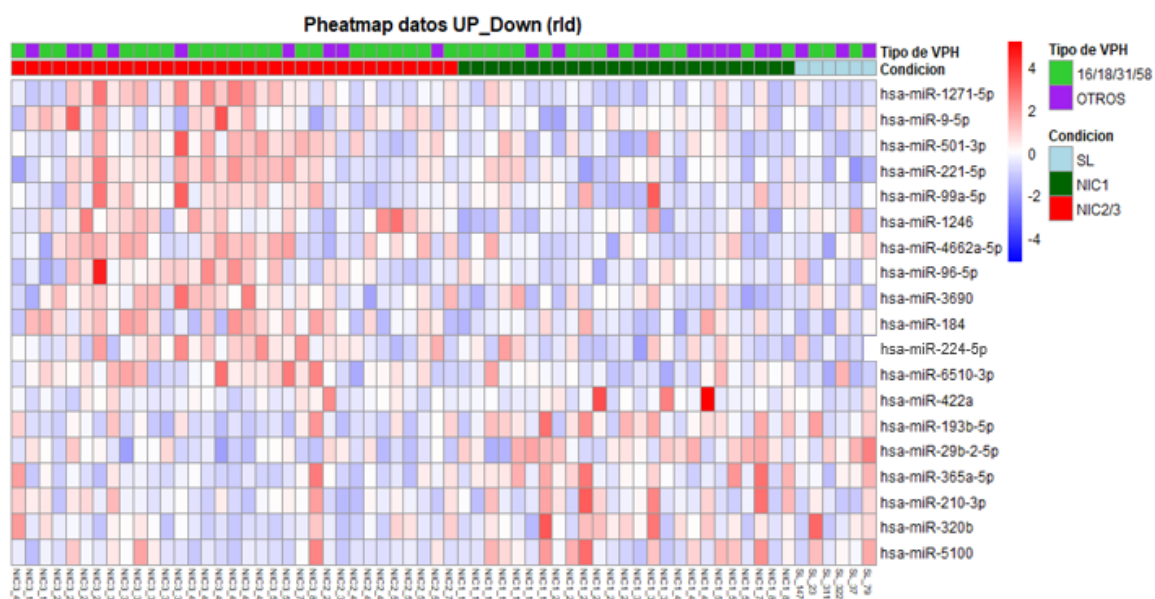
Tabla 8 miRNAs expresados diferencialmente entre las comparaciones NIC2/3 y SL/NIC1. Los que se resaltan en negrita son aquellos que se comparten con las comparaciones de la tabla 7.

	<i>baseMean</i>	<i>FC</i>	<i>log2FC</i>	<i>pvalue</i>	<i>padj</i>	<i>expressed</i>
<i>hsa-miR-1271-5p</i>	8,94	7,87	2,97	1,05E-07	5,50E-05	UP
<i>hsa-miR-9-5p</i>	60,34	3,71	1,89	4,92E-06	0,0012	UP
<i>hsa-miR-501-3p</i>	167,42	3,63	1,85	1,29E-05	0,0016	UP
<i>hsa-miR-221-5p</i>	50,62	2,35	1,23	3,26E-05	0,0034	UP
<i>hsa-miR-99a-5p</i>	491,60	2,93	1,55	0,00010	0,0060	UP
<i>hsa-miR-1246</i>	11,06	3,37	1,75	0,00027	0,011	UP
<i>hsa-miR-4662a-5p</i>	3,67	2,60	1,38	0,00029	0,011	UP
<i>hsa-miR-96-5p</i>	38,60	2,39	1,25	0,00059	0,018	UP
<i>hsa-miR-3690</i>	16,53	2,83	1,50	0,00080	0,023	UP
<i>hsa-miR-184</i>	96,24	2,95	1,56	0,00090	0,024	UP
<i>hsa-miR-224-5p</i>	3125,91	2,28	1,19	0,0019	0,045	UP
<i>hsa-miR-6510-3p</i>	9,42	2,67	1,41	0,0022	0,048	UP
<i>hsa-miR-422a</i>	5,31	0,15	-2,73	9,23E-06	0,0016	DOWN
<i>hsa-miR-193b-5p</i>	13,87	0,28	-1,79	5,59E-05	0,0048	DOWN
<i>hsa-miR-29b-2-5p</i>	3,55	0,45	-1,13	8,30E-05	0,0058	DOWN
<i>hsa-miR-365a-5p</i>	4,60	0,25	-1,99	0,00022	0,011	DOWN
<i>hsa-miR-210-3p</i>	1189,88	0,36	-1,44	0,00033	0,012	DOWN
<i>hsa-miR-320b</i>	430,58	0,46	-1,11	0,00037	0,012	DOWN
<i>hsa-miR-5100</i>	9,0067	0,34	-1,52	0,0023	0,049	DOWN

La Figura 22 muestra el patrón de expresión y variabilidad de las muestras en relación con los miRNAs diferencialmente expresados. Se observa que los miRNAs con expresión aumentada son más prominentes en las muestras de lesiones de alto grado, siendo más intensos en NIC3 y seguidos por NIC2, mientras que en las muestras de lesiones de bajo grado (SL y NIC1) se observa una coloración más tenue, lo que indica una menor expresión. Por otro lado, los miRNAs con expresión disminuida presentan el patrón opuesto: se expresan de manera más intensa en las muestras de lesiones de bajo grado y muestran una reducción gradual en las muestras de NIC2 y NIC3. Aunque predominan estos patrones, también se evidencian algunos solapamientos en la intensidad de color, lo

que sugiere variabilidad en la expresión de ciertos miRNAs entre las diferentes condiciones clínicas.

Figura 22 Mapa de calor de los miRNAs expresados diferencialmente entre SL/NIC1> vs NI2/3, datos transformados a escala logarítmica regularizada. La agrupación de color rojo hace referencia a los NIC2/3, la verde a los NIC1 y la azul a los SL. Se muestra la agrupación realizada con los tipos de VPH-AR.



Además de llevar a cabo el análisis de expresión diferencial con nuestros propios datos, integramos conjuntos de datos públicos para enriquecer nuestros hallazgos y respaldar la selección de miRNAs con mayor potencial para agrupar nuestras muestras según el estadio de las lesiones intraepiteliales cervicales, para distinguir de manera más precisa entre las lesiones de alto y bajo grado.

▪ Análisis de datos públicos:

En total se encontraron 5 set de datos de secuenciación de próxima generación (NGS), y 6 set de datos de microarreglos. Se seleccionaron dos conjuntos de datos, GSE104758 y GSE167858, correspondientes a estudios cuyo diseño era similar al nuestro. Se encontraron expresados diferencialmente y 4 de ellos, **hsa-miR-1246** y **hsa-miR-9-5p**, **hsa-miR-184** y **hsa-miR-99a-5p**, se comparten con nuestros datos. La Tabla 9 muestra los miRNAs expresados diferencialmente.

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

Tabla 9 Resumen Set de datos obtenidos de GEO para el análisis de expresión diferencial. En la última columna los miRNAs diferencialmente expresados. Los marcados en rojo coinciden con nuestro conjunto de datos de descubrimiento y con los artículos publicados para cada set de datos.

# de acceso del estudio	Año del estudio	Título	plataforma	Descripción del estudio	Tipos de muestras	DEGs
GSE104758	2018	Genome-wide miRNA analysis of HPV-positive self-samples yields novel triage markers for early detection of cervical cancer.	HiSeq 2500 (Illumina)	Establecer perfiles de miRNA de todo el genoma en automuestras positivas para VPH para identificar miRNA que pueden predecir la presencia de CIN3 y CCU.	36 mujeres control con \leq CIN1 y 38 mujeres (casos) que fueron diagnosticados histológicamente con una lesión CIN3.	UP: hsa-miR-1246 hsa-miR-20a-3p hsa-miR-9-5p hsa-miR-93-5p hsa-miR-184 Down: hsa-miR-22-3p
GSE167858	2023	miRNAs signature as potential biomarkers for cervical precancerous lesions in Human Papillomavirus Positive women	Illumina MiSeq (Homo sapiens)	Secuenciación de RNA pequeño (miRNA-seq) para comparar los patrones de expresión de miRNA en 20 tejidos fijados con formalina e incluidos en parafina (FFPE) de mujeres positivas para VPHar de Medellín, Colombia.	Muestras de casos =10 muestras de lesión cervical de alto grado Muestras de control = 10 muestras de lesión cervical de bajo grado o sin lesión	hsa-miR-1-3p hsa-miR-133a-3p hsa-miR-133b hsa-miR-143-3p hsa-miR-143-5p hsa-miR-145-3p hsa-miR-145-5p hsa-miR-7976 hsa-miR-204-5p hsa-miR-30b-5p hsa-miR-99a-5p hsa-miR-9-5p Down: ninguno

▪ **Clustering de las muestras de cepillados cervicales usando los miRNAs expresados diferencialmente**

Para determinar si los DEmiRNAs logran discriminar entre las lesiones SL/NIC1 y NIC2/3 del CCU, se utilizó el análisis de clustering por K-medias, como se hizo previamente. Se seleccionaron los miRNAs expresados diferencialmente (al alza y a la baja) compartidos entre las diferentes comparaciones (Tabla 7 y 8) y datos de GEO analizados (Tabla 9). Como se muestra en la Tabla 10, en el clúster 2, las muestras de NIC2/3 tienden a agruparse, y aunque no todas las muestras de NIC2/3 se encuentran en este clúster, son más predominantes en comparación con las muestras de SL/NIC1.

Tabla 10 Tabla de contingencia con la distribución de las muestras en los clústeres 1 y 2 utilizando los DEmiRNAs compartidos entre los análisis. En rojo los que son regulados al alza, en azul los regulados a la baja y en negrita los que se encontraron en los datos analizados de GEO ("**hsa-miR-1271-5p**", "**hsa-miR-9-5p**", "**hsa-miR-99a-5p**", "**hsa-miR-1246**", "**hsa-miR-221-5p**", "**hsa-miR-501-3p**", "**hsa-miR-4662a-5p**", "**hsa-miR-96-5p**", "**hsa-miR-184**", "**hsa-miR-422a**", "**hsa-miR-193b-5p**", "**hsa-miR-29b-2-5p**", "**hsa-miR-365a-5p**", "**hsa-miR-210-3p**", "**hsa-miR-320b**").

	1	2
<i>SL/NIC1</i>	30	1
<i>NIC2/3</i>	21	12

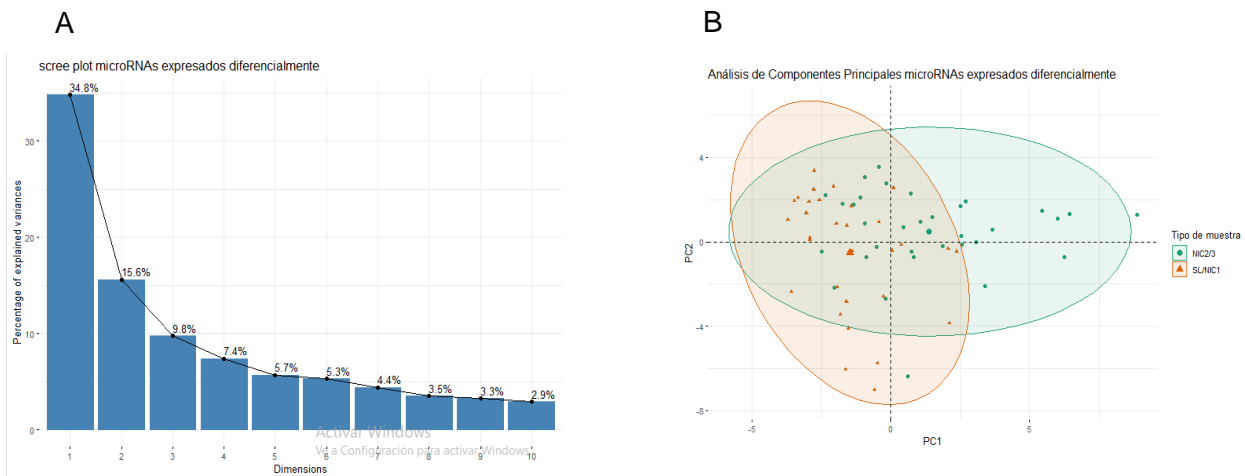
Aunque en la separación inicial se observó una tendencia a agrupar las muestras de NIC2/3 en el Cluster 2, un número considerable de ellas compartía características del cluster 1, lo que indica una separación subóptima. Dado que existía una tendencia a discriminar las lesiones de alto grado, el desafío residía en identificar los miRNAs más adecuados para mejorar esta clasificación y lograr una mayor separación entre las muestras de los diferentes grupos. Por esta razón, se decidió emplear el análisis de clustering por K-medias como un enfoque adicional para apoyar la selección de los miRNAs con mayor poder discriminatorio. Esta técnica permite agrupar las muestras de manera no supervisada, basándose en los patrones de expresión de los miRNAs, lo que facilita la identificación de aquellos que mejor contribuyen a la separación de las condiciones.

Se incluyeron todos los miRNAs diferencialmente expresados en la comparación SL/NIC1 vs NIC2/3 (Tabla 8), debido que los resultados de dicha comparación incluyen los miRNAs previamente identificados en el análisis de expresión diferencial (Tabla 7), los datos obtenidos del análisis de GEO, y por un análisis adicional realizado con el software gratuito ExpressAnalyst, en el que también se compararon las condiciones SL/NIC1 vs NIC2/3 (datos no mostrados).

Se realizó PCA para observar la varianza explicada y la variabilidad de los datos por cada componente principal (Figura 23-A), destacando que las dos primeras componentes principales capturan el 50.4% de esta. En la figura 23-B se observa la distribución de las

muestras en los dos primeros componentes principales con respecto a los miRNAs expresados diferencialmente, en contraste con el análisis de PCA realizado con todos los miRNAs (Figura 19) se puede notar que los NIC2/3 exhiben una tendencia a ubicarse hacia el lado derecho del gráfico y se separan más claramente de las lesiones de bajo grado.

Figura 23 Análisis de PCA de los miRNAs expresados diferencialmente. (A) Variabilidad retenida en los componentes principales de los miRNAs expresados diferencialmente entre las condiciones de SL/NIC1 vs NIC2/3. (B) PCA de las muestras.



Se realizó un análisis de K-medias sobre los resultados de los dos componentes principales y también directamente sobre los datos de expresión sin reducción de componentes. Ambas estrategias mostraron una tendencia a agrupar las muestras de alto grado en el clúster 2, como se observa en la Tabla 11. Sin embargo, en este análisis, el número de muestras SL/NIC1 en el clúster 2 aumentó de uno a tres en comparación con el análisis de la Tabla 10. Adicionalmente, el incremento de las muestras NIC2/3 en el clúster 2 no fue significativo. Al comparar con el análisis de K-medias realizado previamente para todos los miRNAs (Figuras 20 y 21), se evidencia que, usando los miRNAs expresados diferencialmente, se logra una separación más clara entre las lesiones de alto y bajo grado en los clusters, y aunque esta separación no abarca un gran número de muestras según su tipo de lesión, sí muestra una tendencia más marcada a distinguir entre las diferentes lesiones, por lo que es necesario seguir explorando, entre los miRNAs expresados diferencialmente, aquellos que proporcionen un mejor resultado de clustering.

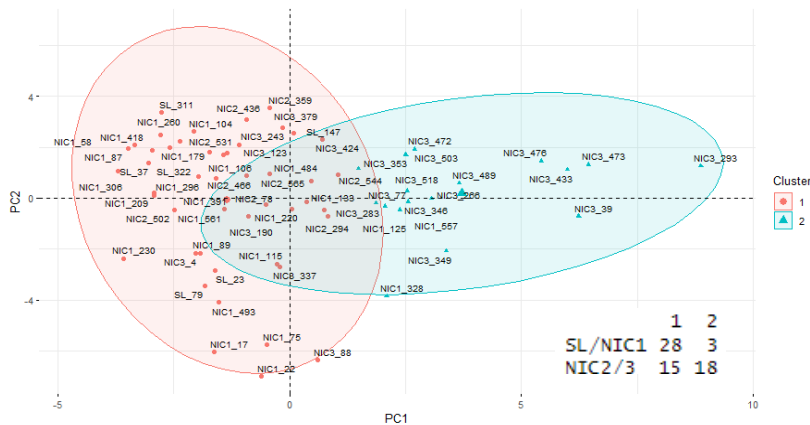
Tabla 11. Tabla de contingencia que muestra la distribución de las muestras en los clústeres 1 y 2 utilizando los miRNAs expresados diferencialmente (al alza y a la baja) en la comparación SL/NIC1 vs NIC2/3.

	1	2
SL/NIC1	28	3
NIC2/3	19	14

Aunque se han usado los miRNAs expresados diferencialmente al alza y a la baja para el análisis de K-medias, se observó que los miRNAs expresados al alza generaban una mejor separación, aumentando significativamente la cantidad de muestras NIC2/3 en el cluster 2 (Figura 24), por lo cual, se decidió analizar únicamente los miRNAs expresados al alza, por su mayor poder discriminatorio entre las condiciones estudiadas y en la búsqueda de biomarcadores, es preferible utilizar los miRNAs al alza porque representan genes cuya expresión se incrementa en una condición patológica determinada; que podría a menudo indicar una mayor relevancia biológica y funcional en el proceso de la enfermedad. Además, los genes expresados al alza son más fácilmente detectables y cuantificables, lo que mejora la robustez y sensibilidad de los biomarcadores más efectivos y clínicamente relevantes para la detección de NIC2/3.

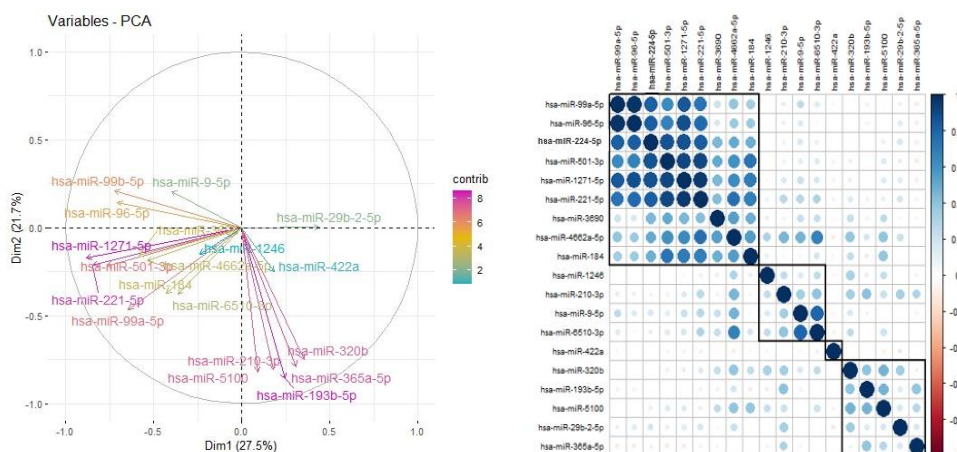
La Figura 23, muestra la tabla de contingencia con la distribución de las muestras en los clústeres 1 y 2 utilizando los miRNAs expresados diferencialmente al alza. Se observó que el clúster 2 aumentó el número de muestras NIC2/3, por lo que continuar explorando con estos miRNAs podría reducir el número total de miRNAs necesarios para discriminar entre este tipo de lesiones, facilitando su validación y la selección de los mejores.

Figura 24 Análisis de k-medias en los dos primeros componentes principales para 2 clusters usando los niveles de expresión de los miRNAs expresados al alza, y una tabla de contingencia que detalla la distribución de las muestras en los clusters para las condiciones SL/NIC1 y NIC2/3.



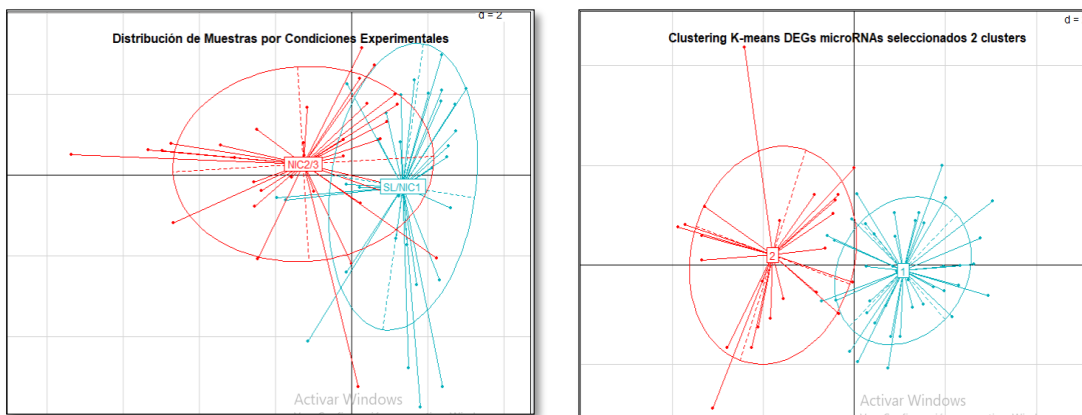
La selección de los miRNAs con mayor capacidad discriminadora entre SL/NIC1 y NIC2/3 mediante clustering, se optimizó con la elaboración de gráficos de correlación de los DE miRNAs, para ayudar a la selección dentro de los miRNAs sobreexpresados (Figura 25 A-B), e identificar los miRNAs con mayor correlación y variabilidad entre sí, para capturar la diversidad en los patrones de expresión de los miRNAs, y mejorar el poder discriminador del análisis. Con base en estas características, los miRNAs seleccionados fueron: 'hsa-miR-1271-5p', 'hsa-miR-9-5p', 'hsa-miR-3690', 'hsa-miR-6510-3p', 'hsa-miR-4662a-5p', 'hsa-miR-184', 'hsa-miR-501-3p' y 'hsa-miR-1246'.

Figura 25 Correlación de los DEmiRNAs. (A) Visualización de las variables (microRNAs) en el espacio definido por los dos primeros componentes principales obtenidos del análisis de PCA. Cada variable se representa como un punto en el plano definido por los dos primeros componentes principales, y el color de cada punto se determina según la contribución de esa variable a los componentes principales. (B) Grafico de correlaciones entre miRNAs, los menos correlacionados tienden a agrupar diferentes conjuntos de muestras.



Se realizó el agrupamiento mediante K-medias con los miRNAs seleccionados para validar visualmente su capacidad discriminadora. Los miRNAs seleccionados agrupaban las muestras en grupos bien definidos, confirmando su potencial para discriminar entre las condiciones experimentales. Al agrupar las condiciones experimentales en dos clusters, el dos agrupó un número mayor de lesiones NIC2/3 con respecto a los resultados anteriores, y la separación entre condiciones es mucho más evidente. En la figura 26-A, los puntos se colorean según condiciones experimentales conocidas, mientras que, en el segundo gráfico (Figura 26-B), los puntos se colorean según los clusters asignados por k-medias, lo que facilita la comparación entre los clusters formados y las condiciones experimentales.

Figura 26 Visualización de K-medias con la función s.class. (A) distribución de las muestras por condición experimental (SL/NIC1 y NIC2/3) según los miRNAs expresados diferencialmente. (B) Coordenadas por color de las muestras que pertenecen a los clusters 1 y 2, utilizando los miRNAs seleccionados por su mayor potencial discriminativo ("hsa-miR-1271-5p", "hsa-miR-9-5p", "hsa-miR-3690", "hsa-miR-6510-3p", "hsa-miR-4662a-5p", "hsa-miR-184", "hsa-miR-501-3p" y "hsa-miR-1246").



Se elaboró una tabla de contingencia para detallar cómo se distribuyen las muestras de las diferentes condiciones en los clusters formados (Tabla 12). En esta tabla se observa que hay una mayor cantidad de muestras de la condición NIC2/3 que se separan de las SL/NIC1, evidenciando la capacidad discriminativa del análisis de clustering para esta condición. Se exploraron varias combinaciones de las condiciones de estudio para determinar la efectividad de los miRNAs seleccionados en la separación de las muestras, observando que la combinación SL/NIC1 vs NIC2/3 ofrece una mejor interpretación de la información. Sin embargo, se reconoce que las otras combinaciones de condiciones también aportan datos relevantes para el análisis:

Tabla 12 Tablas de contingencia que muestran la distribución de las muestras en los clústeres 1 y 2 utilizando los miRNAs seleccionados: ("hsa-miR-1271-5p", "hsa-miR-9-5p", "hsa-miR-3690", "hsa-miR-6510-3p", "hsa-miR-4662a-5p", "hsa-miR-184", "hsa-miR-501-3p" y "hsa-miR-1246"). (A) Distribución basada en dos clusters y dos condiciones clínicas (SL/NIC1 vs. NIC2/3). (B) Distribución de las muestras en los clústeres 1 y 2, esta vez separando NIC2 de NIC3, pero manteniendo la clasificación de bajo grado (SL/NIC1). (C) Distribución desglosada según tres condiciones clínicas (SL, NIC1, NIC2/3).

A

	Cluster 1	Cluster 2
SL/NIC1	27	4
NIC2/3	12	21

B

	Cluster 1	Cluster 2
SL/NIC1	27	4
NIC2	8	2
NIC3	4	19

C

	Cluster 1	Cluster 2
SL	6	0
NIC1	21	4
NIC2/3	12	21

Para cuantificar la precisión del clustering, se calculó el porcentaje de error con base en la cantidad de muestras mal clasificadas. El porcentaje de error resultante fue del 26,56%. La sensibilidad del clustering fue del 62% y la especificidad del 87.1%. Con estos miRNAs seleccionados aumentó considerablemente la capacidad discriminatoria para identificar las lesiones NIC2/3 de las de SL/NIC1. Sin embargo, dado que el análisis no supervisado agrupa las muestras en función de patrones inherentes y no etiquetas predefinidas, es importante interpretar estos resultados con cautela. El hecho de que se estén evaluando las mismas muestras utilizadas para identificar los miRNAs seleccionados podría introducir sesgos en la evaluación de la precisión. Para mejorar la validez de estos resultados, sería recomendable validar estos miRNAs en un conjunto de datos independiente o mediante validación cruzada, reduciendo así el riesgo de sobreajuste y asegurando su capacidad discriminatoria en otras cohortes.

7.3 Capítulo 3. miRNAs con alto poder discriminatorio para detectar NIC2/3

7.3.1 Resultados Objetivo 3. *Identificar mediante el análisis bioinformático de las secuencias obtenidas, los miRNAs con potencial utilidad para detectar NIC2/3 con un alto poder discriminatorio e identificar los genes blanco de éstos miRNAs y sus implicaciones en los procesos relacionados con la carcinogénesis.*

- **Curvas ROC para evaluar el potencial de los miRNAs de forma individual para discriminar lesiones de alto grado NIC2/3.**

Este análisis permitió validar el desempeño individual de cada miRNA regulado al alza en la identificación de NIC2/3, ayudando a seleccionar los miRNAs más efectivos como biomarcadores para estas lesiones. Además, proporcionó una base cuantitativa adicional que refuerza las conclusiones sobre la utilidad de los miRNAs en la detección de NIC2/3 con base en los datos de secuenciación. **hsa-miR-1271 y hsa-miR-6510** fueron los miRNAs que mostraron significancia estadística, sensibilidad y especificidad media (Figura 26). Los resultados obtenidos para el miR-1271 (Figura 27-A), muestran un área bajo la curva (AUC) de 0,710 con una desviación estándar de 0,066, lo que indica una capacidad de discriminación moderada. El miR6510 (Figura 27-B) mostró un área bajo la curva (AUC) de 0,660 con una desviación estándar de 0,068, que también indica una capacidad moderada de discriminación.

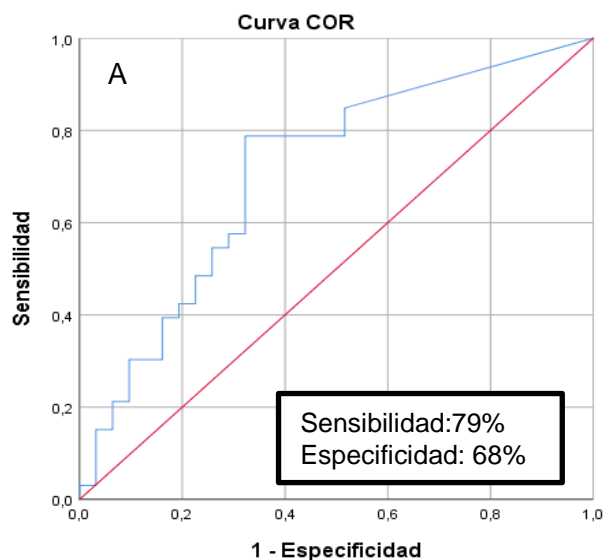
Se realizó un análisis de clustering exploratorio utilizando únicamente los niveles de expresión de estos dos miRNAs y se encontró que estos miRNAs son los principales responsables de la separación observada entre los grupos (Tabla 13):

Tabla 13. Tabla de contingencia del análisis de clustering por K-medias utilizando los miRNAs hsa-miR-1271 y hsa-miR-6510.

	1	2
<i>SL/NIC1</i>	28	3
<i>NIC2/3</i>	16	17

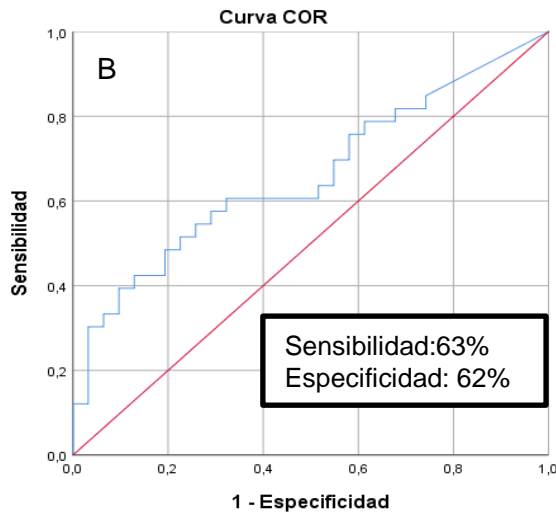
A pesar de que los otros miRNAs evaluados no alcanzaron significancia estadística (valor p), presentaron valores de AUC superiores a 0.50 (miR-3690, miR-184, miR-9-5p). Esto sugiere que, aunque la asociación no sea estadísticamente significativa, estos miRNAs podrían tener potencial como marcadores diagnósticos. La falta de significancia podría deberse al tamaño reducido de la muestra o a la variabilidad intrínseca de los datos, y no necesariamente descarta su relevancia biológica.

Figura 27 Curva ROC de los miRNA que presentaron significancia estadística. (A) miR1271 (B) miR6510.



Variables de resultado **miR1271** de prueba:

Área	Desv. Error ^a	Significación asintótica ^b	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
0,710	0,066	0,004	0,581	0,839



Variables de resultado de prueba: **miR6510**

Área	Desv. Error ^a	Significación asintótica ^b	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
0,660	0,068	0,028	0,526	0,794

▪ Regresión logística

Con base en los resultados obtenidos (Tabla 14), algunas variables mostraron una asociación estadísticamente significativa con lesiones cervicales de alto grado. El uso de anticonceptivos hormonales y de dispositivos intrauterinos (DIU) aumenta el riesgo de tener lesiones cervicales de alto grado en más de 5.2 veces en comparación con mujeres que no utilizan ningún método anticonceptivo. Además, se observó que el uso de cualquier tipo de anticonceptivo durante un período de 1 a 48 meses está asociado con un aumento del riesgo de lesiones cervicales de alto grado en 4.5 veces en comparación con mujeres que nunca han utilizado anticonceptivos.

El análisis de regresión logística bivariada reveló que dos miRNAs, **hsa-miR-1271-5p** y **hsa-miR-4662a-5p**, mostraron una asociación estadísticamente significativa con las lesiones preneoplásicas de alto grado en cérvix. El miRNA hsa-miR-1271-5p presentó un valor de P de 0.005, un OR de 1.252 y un IC 95% de 1.171 a 1.464, lo que sugiere que un incremento en la expresión de éste miRNA se asocia con un aumento del 1.25%—en la probabilidad de encontrar lesiones preneoplásicas de alto grado en comparación con lesiones de bajo grado.

Por su parte, hsa-miR-4662a-5p mostró un valor de P de 0.002, un OR de 1.298 y un IC 95% de 1.102 a 1.529, lo que indica que el incremento en la expresión de éste miRNA se asocia con un aumento del 1.29% en la probabilidad de presentar lesiones preneoplásicas de alto grado, en comparación con lesiones de bajo grado. Estos hallazgos sugieren que estos miARNs son más indicativos de la presencia de lesiones de alto grado en lugar de predecir la progresión de lesiones de bajo grado a lesiones de alto grado. Por tanto, su utilidad se centra más en el diagnóstico que en la predicción del riesgo de progresión.

Tabla 14 Regresión logística bivariada. En la tabla se muestran los resultados de una regresión logística bivariada que utilizó como variable de desenlace tener una lesión de bajo = 0 o alto grado =1. (Ref= grupo de referencia)

Característica clínica	Valor de P	OR	95% C.I._ menor	95% C.I._ mayor
Ocupación				
Ama de casa (Ref)				
Técnico o Profesional	0.433	0.615	0.183	2.072
Otras	0.661	0.769	0.238	2.488
Escolaridad				
Primaria y secundaria (Ref)				
Técnica y/o Universitaria	0.942	0.963	0.346	2.674
Estrato Socioeconómico				
1+2 (Ref)				
3	0.565	0.640	0.140	2.921
4+5+6	0.772	0.800	0.177	3.618
Partos				
1 o 2 (Ref)	0.358	1.806	0.512	6.363
Mas de 2	0.283	1.920	0.583	6.324
Numero de Compañeros sexuales				
Hasta 2 (Ref)				
Mas de 2	0.384	0.629	0.222	1.787
Uso de Anticonceptivos				
Ninguno (Ref)				

Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

Hormonales + DIU	0.023	5.200	1.253	21.572
Otros	0.311	1.878	0.555	6.358
Tiempo de uso de Anticonceptivos				
No usa (Ref)				
De 1 a 48 meses	0.029	4.500	1.166	17.373
De 60 a 230 meses	0.441	1.587	0.490	5.138
Consumo de Tabaco				
Nunca (Ref)				
Fuma actualmente o fumo en el pasado	0.472	1.519	0.486	4.748
hsa-miR-1246	0.056	1.051	0.999	1.106
hsa-miR-1271-5p	0.005	1.252	1.171	1.464
hsa-miR-184	0.033	1.005	1.000	1.010
hsa-miR-221-5p	0.003	1.033	1.011	1.056
hsa-miR-3690	0.024	1.053	1.007	1.101
hsa-miR-4662a-5p	0.002	1.298	1.102	1.529
hsa-miR-501-3p	0.023	1.005	1.001	1.010
hsa-miR-6510-3p	0.076	1.058	0.994	1.127
hsa-miR-9-5p	0.051	1.019	1.000	1.039
hsa-miR-96-5p	0.038	1.024	1.001	1.047
hsa-miR-99a-5p	0.345	1.000	1.000	1.001
hsa-miR-224-5p	0.038	1.000	1.000	1.000

En negrita variables con valores p significativos

La tabla 15 reúne las características de cada miARN según su desempeño en los análisis realizados, para identificar aquellos miARNs que cumplieran con el mayor número de características relevantes, ya que podrían ser los candidatos más prometedores para discriminar entre NIC2 y NIC3. Se asignó un punto a cada miARN que cumplía con una característica específica, y se sumaron los puntos al final para obtener una puntuación total. Este método, aunque empírico, permitió una evaluación preliminar para determinar cuáles miARNs podrían ser los más adecuados para una validación más detallada. Los miARNs **hsa-miR-1271-5p**, **hsa-miR-9-5p**, **hsa-miR-501-3p**, **hsa-miR-1246**, y **hsa-miR-4662a-5p** son los que reúnen la mayor cantidad de características, lo que sugiere que podrían tener una capacidad superior para discriminar entre lesiones NIC2 y NIC3. Por otro lado, los miARNs **hsa-miR-3690**, **hsa-miR-6510-3p**, y **hsa-miR-184** presentan una puntuación media, indicando un desempeño intermedio en la discriminación de estas lesiones.

Tabla 15. Tabla de Características de miARNs para la Discriminación de NIC2/3.

	1	2	3	4	5	6	7	8	suma
<i>hsa-miR-1271-5p</i>	1	1	1	1	1	1	0	1	7
<i>hsa-miR-9-5p</i>	1	1	1	1	0	0	1	1	6
<i>hsa-miR-501-3p</i>	1	1	1	1	0	0	0	1	5
<i>hsa-miR-1246</i>	1	1	1	1	0	0	1	1	6
<i>hsa-miR-4662-5p</i>	1	1	1	0	0	1	0	1	5
<i>hsa-miR-3690</i>	1	1	1	0	0	0	0	1	4
<i>hsa-miR-184</i>	1	0	1	0	0	0	1	1	4
<i>hsa-miR-6510-3p</i>	1	0	1	0	1	0	0	1	4
<i>hsa-miR-221-5p</i>	1	1	1	0	0	0	0	0	3
<i>hsa-miR-99a-5p</i>	1	1	0	0	0	0	1	0	3
<i>hsa-miR-96-5p</i>	1	1	1	0	0	0	0	0	3
<i>hsa-miR-224-5p</i>	1	0	1	0	0	0	0	0	2

*1=Expresados diferencialmente SL/NIC1 vs NIC2/3, *2= Expresados diferencialmente NIC1 vs NIC2/3, *3= Expresados diferencialmente ExpressAnalyst, *4= FoldChange mayor a 3, *5= Curva ROC, *6= Regresión logística, *7= GEO, *8= Seleccionados en el clustering

▪ Análisis de enriquecimiento funcional y genes blanco

Se utilizó el paquete `get_multimir` de RStudio para identificar los genes blanco de los miRNA que obtuvieron las mayores puntuaciones y las puntuaciones medias en la Tabla 14: ("hsa-miR-1271-5p", "hsa-miR-9-5p", "hsa-miR-3690", "hsa-miR-6510-3p", "hsa-miR-4662a-5p", "hsa-miR-184", "hsa-miR-501-3p" y "hsa-miR-1246"), para realizar el enriquecimiento. Se identificaron un total de 2805 genes objetivos de estos miARNs pero como no se encontraron genes comunes entre los blancos de todos los miARNs, se decidió realizar el análisis de enriquecimiento utilizando el conjunto completo de 2805 genes recopilados.

Además, para investigar la relación entre los genes objetivo de los miRNAs seleccionados y el CCU asociado con la infección por VPH, se realizó una búsqueda en la literatura (49, 152, 153, 154) centrada en genes implicados en la carcinogénesis cervical inducida por VPH. Se identificaron 93 genes relevantes (no mostrados), que se analizaron para determinar si podían ser objetivos de los miRNAs seleccionados. Los resultados (Tabla 16) revelaron que 5 de los 8 miRNAs utilizados en el enriquecimiento regulan genes clave asociados con las oncoproteínas E6, E7 y E5, indicando una significativa relevancia

biológica de estos miRNAs en el contexto del CCU relacionado con VPH, pudiendo ser claves para un proceso de validación.

Tabla 16 Genes objetivo de los miRNAs relacionados con la carcinogénesis cervical en la infección por VPH.

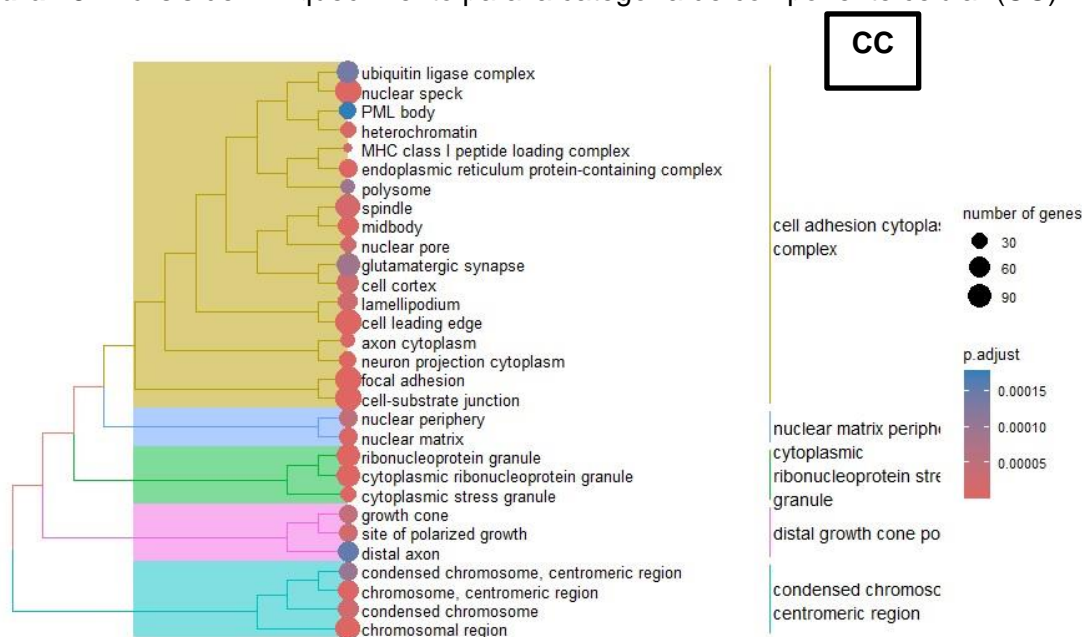
hsa-miR-1271	hsa-miR9-5p	hsa-miR-501	hsa-miR-184	hsa-miR-1246
CDKN1B	EP300	SCRIB	PRKCB	TP53
CCND2	MAML1	CDKN1B	MYC	PIK3CB
FOXO1	TP53	E2F1		TP63
IFNAR1	CDKN1A	CCND2		
IGF2R	UBR4	FOXO1		
CCND1	RABEP1	E2F		
	PHGDH			
	TGFBR2			
	PRKCB			
	KDELRL1			
	TJP2			
	FOXO1			
	FNDC3B			
	SDC1			
	ERBIN			
	CXCL8			
	KDM6B			
	E2F			

En el análisis de enriquecimiento, el análisis GO en la categoría de componente celular (CC) (Figura 28) incluye principalmente términos como "Complejo citoplasmático de Adhesión Celular", "Matriz Nuclear Periférica", "Gránulos de estrés de ribonucleoproteínas", "crecimiento del cónico distal" y "región cromosómica". En la categoría de proceso biológico (BP) (Figura 29) los términos más relevantes son "fosforilación de peptidil-serina", lo que sugiere alteraciones en las vías de señalización celular debido a modificaciones postraduccionales de proteínas, y "respuesta a la radiación", que indica una mayor actividad en la reparación del ADN y mecanismos de defensa celular frente al daño genético, posiblemente causado por el VPH. Además, el término "respuesta a la hipoxia" sugiere que las células se están adaptando a condiciones de hipoxia, comunes en tejidos preneoplásicos. Los términos "desfosforilación de

proteínas" y "actividad fosfatasa de fosfoproteínas" resaltan la importancia de la regulación postraduccional de proteínas en la progresión de estas lesiones. Finalmente, en la categoría de función molecular (MF) (Figura 30) los términos más relevantes incluyen "unión a cadherinas", que sugiere alteraciones en la adhesión y comunicación celular, "unión a enzimas", "actividad de la quinasa de ADN", y "unión a pequeñas GTPasas", resalta el papel de las proteínas G pequeñas en la señalización intracelular y la regulación del citoesqueleto.

Como resultado del análisis de KEGG (Figura 31), se asociaron varias vías que incluyen las alteraciones en el ciclo celular, vías de señalización críticas como P53, que está involucrada en la regulación de la apoptosis y la respuesta al daño del ADN; PI3K-Akt, que juega un papel esencial en la supervivencia celular y la resistencia a la apoptosis; FoxO, que está implicada en la regulación del estrés oxidativo y la longevidad de las células cancerosas vías asociadas con la infección viral y el VPH.

Figura 28 Análisis de Enriquecimiento para la categoría de componente celular (CC).



Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

Figura 29 Análisis de enriquecimiento para componente de procesos biológicos (BP).

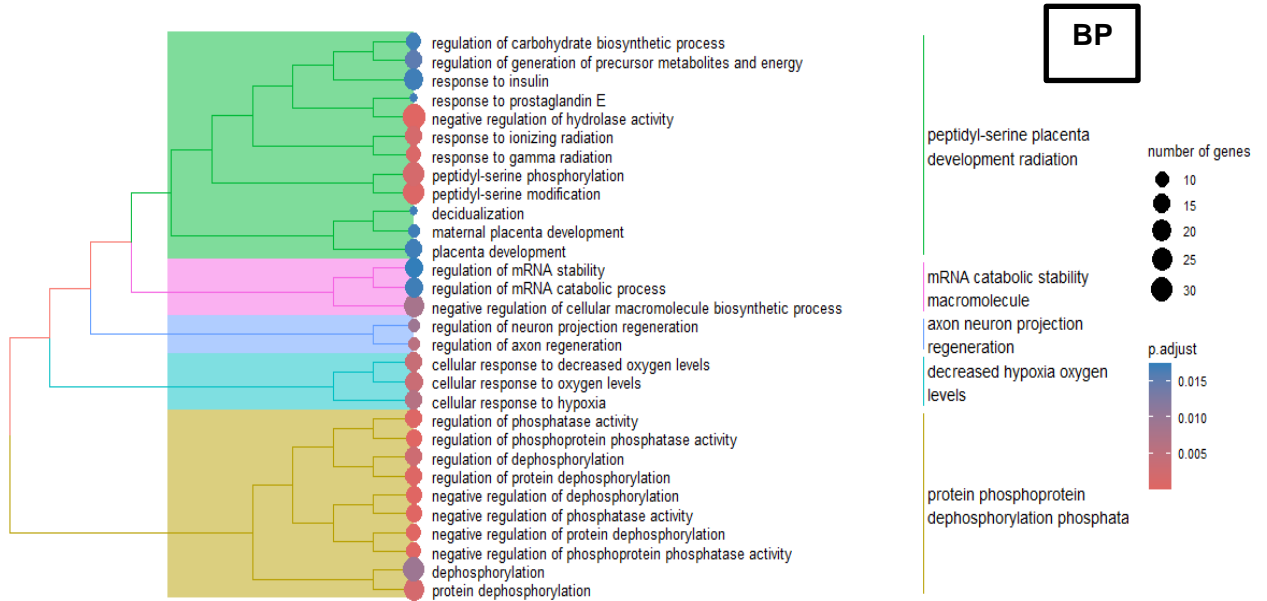


Figura 30 Análisis de enriquecimiento funcional para componente de función molecular.

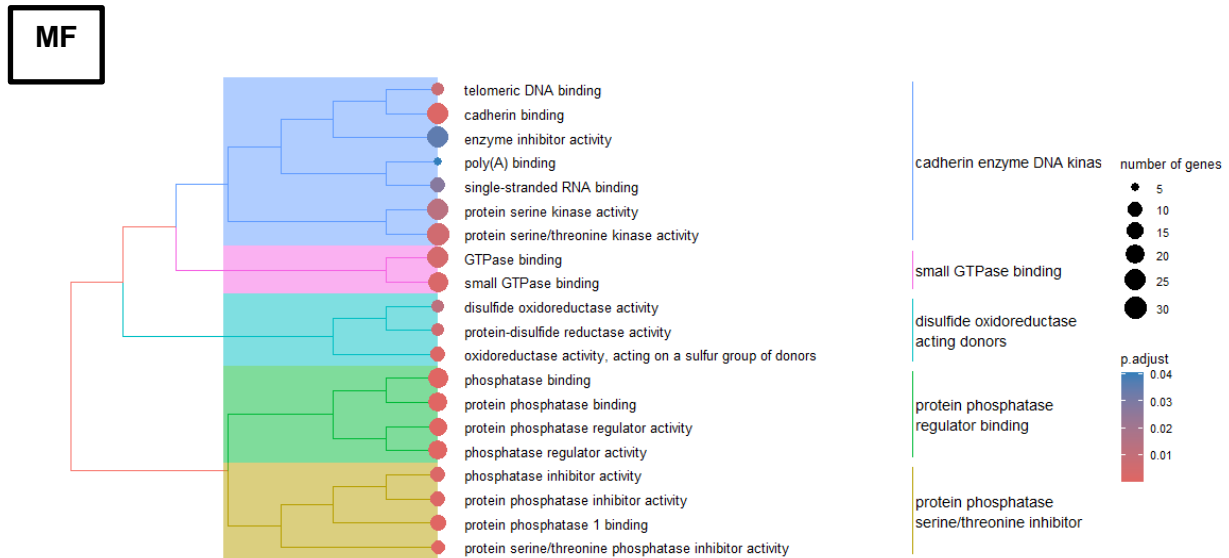
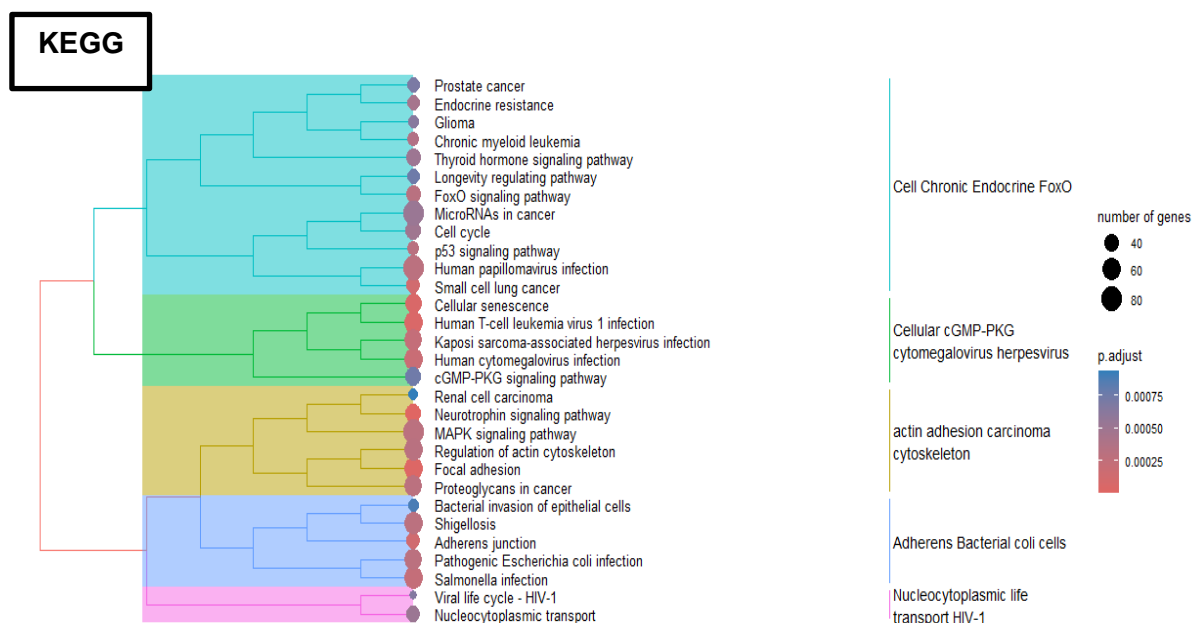


Figura 31 Analisis de enriquecimiento para vias metabolicas.



Al asociar la relevancia biológica observada en los miRNAs con los resultados obtenidos en los distintos análisis, los mejores candidatos para una validación experimental de acuerdo a los análisis del grupo de descubrimiento serían: **hsa-miR-1271-5p**, **hsa-miR-9-5p**, **hsa-miR-501-3p** y **hsa-miR-1246** (Tabla 17).

Tabla 17. Características de los miRNAs seleccionados según su desempeño en los análisis realizados para la detección de NIC2/3.

	1	2	3	4	5	6	7	8	9	suma
hsa-miR-1271	1	1	1	1	1	1	0	1	1	8
hsa-miR-9-5P	1	1	1	1	0	0	1	1	1	7
hsa-miR-501	1	1	1	1	0	0	0	1	1	6
hsa-miR-1246	1	1	1	1	0	0	1	1	1	7
<i>hsa-miR-4662</i>	1	1	1	0	0	1	0	1	0	5
<i>hsa-miR-3690</i>	1	1	1	0	0	0	0	1	0	4
<i>hsa-miR-184</i>	1	0	1	0	0	0	1	1	1	5
<i>hsa-miR-6510</i>	1	0	1	0	1	0	0	1	0	4

*1=Expresados diferencialmente SL/NIC1 vs NIC2/3, *2= Expresados diferencialmente NIC1 vs NIC2/3, *3= Expresados diferencialmente ExpressAnalyst, *4= FoldChange mayor a 3, *5= Curva ROC, *6= Regresión logística, *7= GEO, *8= Seleccionados en el clustering, *9= genes blancos implicados en la carcinogénesis cervical inducida por VPH

▪ **Validación experimental**

Aunque la validación experimental de los resultados obtenidos en la fase de descubrimiento no estaba inicialmente prevista, la validación con datos externos es una parte fundamental en los estudios de descubrimiento de biomarcadores. El objetivo 3 del proyecto contemplaba la identificación de miRNAs con potencial utilidad para detectar NIC2/3 y en línea con este objetivo, se logró realizar una validación experimental exploratoria para evaluar algunos de los miRNAs seleccionados como candidatos en la fase de descubrimiento (hsa-miR-1271-5p, hsa-miR-9-5p, hsa-miR-501-3p, hsa-miR-3690, hsa-miR-6510-3p, hsa-miR-4662a-5p, hsa-miR-184 y hsa-miR-1246). Esta validación, fue considerada para reforzar y respaldar los hallazgos del análisis bioinformático. Se validó experimentalmente la expresión de 4 (**hsa-miR-1271-5p, hsa-miR-9-5p, hsa-miR-501-3p, hsa-miR-4662a-5p**) de los miARNs candidatos mediante PCR en tiempo real (qPCR), junto con 2 miARNs adicionales que no estaban dentro de los candidatos (**hsa-miR-224 y hsa-miR-99a-5p**) (ver Tabla 15).

Dado que 4 de los 8 miARNs candidatos mostraban un alto potencial como biomarcadores para la detección de lesiones de alto grado (NIC2/3), y no era viable validarlos todos, se eligieron aquellos con mayor puntaje (**hsa-miR-1271-5p, hsa-miR-501-3p, hsa-miR-9-5p y hsa-miR-4662a-5p**). Además, se seleccionaron dos miARNs adicionales (**hsa-miR-99a-5p, hsa-miR-224-5p**) exclusivamente por su nivel de expresión diferencial, a pesar de no haber estado dentro de los candidatos (ver Tabla 15).

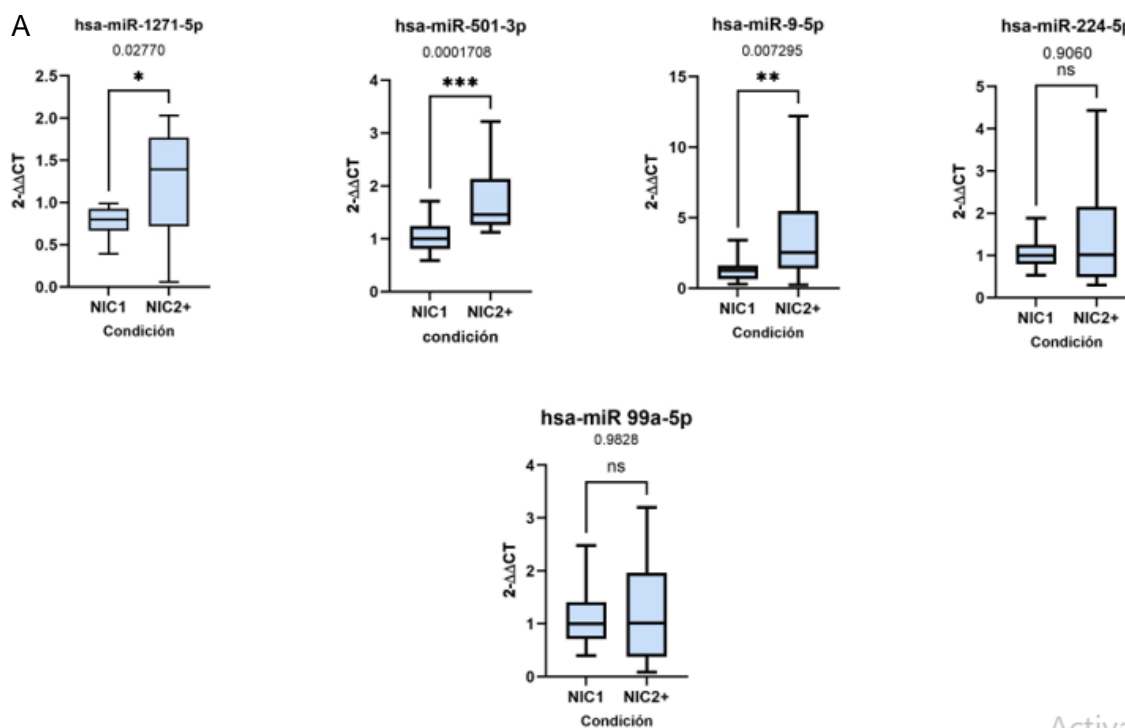
La expresión diferencial de estos 6 miRNAs fue evaluada mediante la técnica de PCR en tiempo real y cuantificación relativa a través de la comparación de CTs. No se obtuvieron resultados de amplificación para 14 de las 30 muestras del miRNA hsa-miR-4662a-5p, por lo que se decidió excluirlo de los análisis por falta de datos. Las muestras que se utilizaron se resumen en la tabla 18:

Tabla 18.Numero de muestras utilizadas en la cohorte de validación por RT-qPCR

<i>Tipo Lesión</i>	<i>n</i>	<i>Total %</i>	<i>Tipo VPH</i>
NIC1	20	100%	VPH-AR
NIC2	5		VPH-AR
NIC3	5	100%	VPH-AR

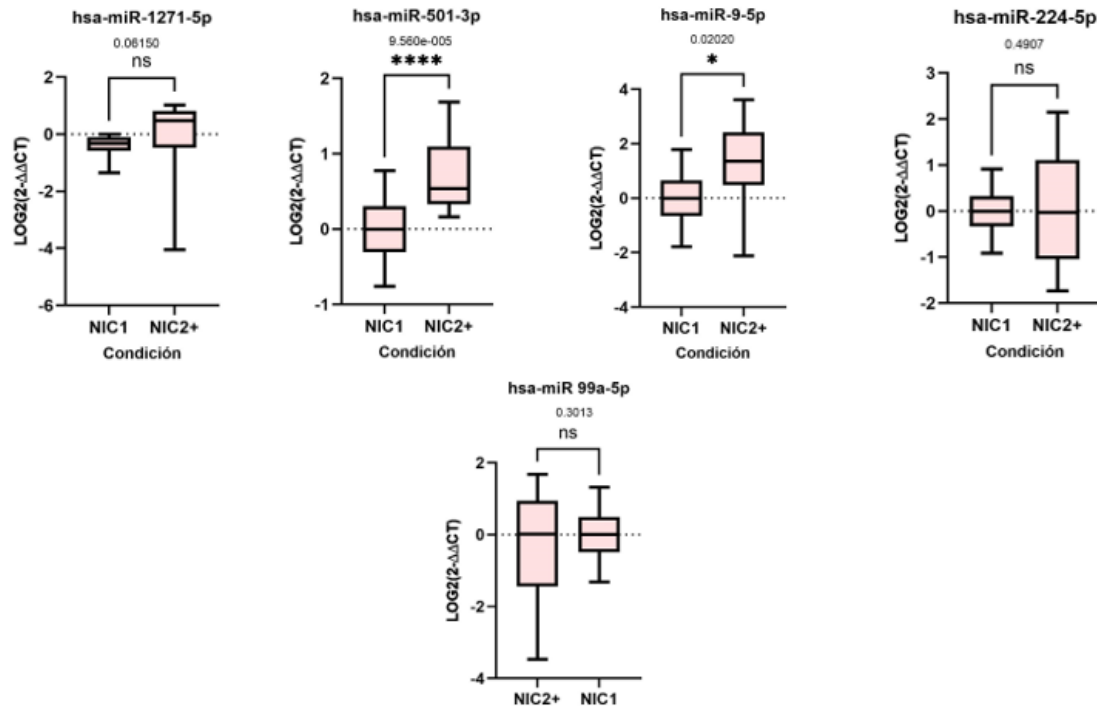
Los niveles de expresión relativa fueron significativamente diferentes ($p < 0.05$, prueba de Mann-Whitney) en la comparación entre los grupos NIC1 y NIC2/3 para los miRNAs hsa-miR-501 y hsa-miR-9-5p (Figura 32-A, B), indicando que estos miRNAs tienen una expresión significativamente mayor en las lesiones de alto grado con respecto a las lesiones de bajo grado de cérvix positivas para VPH-AR. En el caso del hsa-miR-1271-5p, aunque no resultó significativo al utilizar el logaritmo base 2 de $2^{-\Delta\Delta Ct}$ (Figura 31-B), se ve el leve incremento en la expresión de lesiones NIC2/3, y al usar el valor de la expresión relativa $2^{-\Delta\Delta Ct}$, sí mostró significancia (Figura 31-A). El resto de miRNAs evaluados, no mostraron significancia estadística.

Figura 32. Expresión relativa de los 6 miRNAs en las muestras de validación. (A) Expresión relativa en escala $2^{-\Delta\Delta Ct}$ (B) Expresión relativa en escala $\text{Log}_2(2^{-\Delta\Delta Ct})$. (NIC2+=NIC2/3)



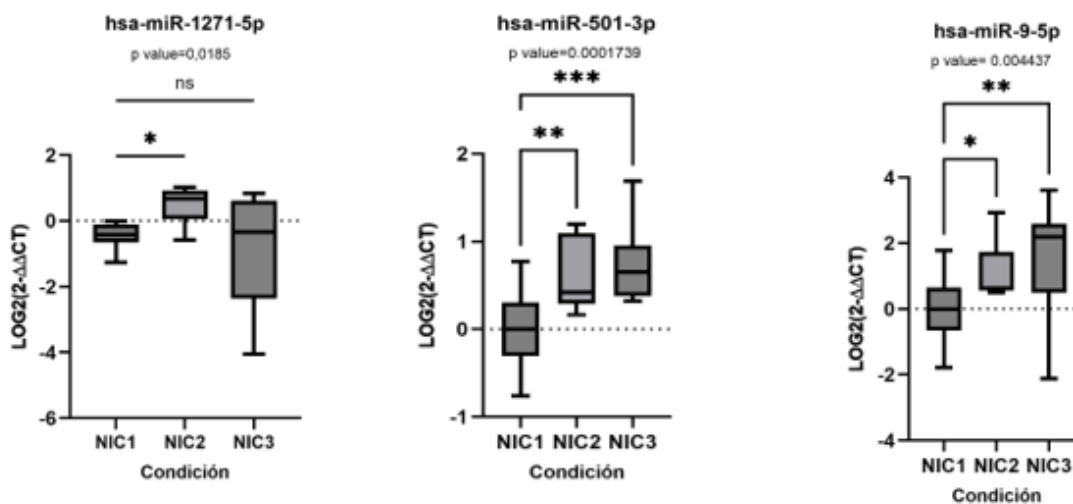
Identificación de una firma de miRNAs con potencial utilidad como biomarcador para el triaje de mujeres positivas para la infección por el virus del papiloma humano de alto riesgo (VPH-AR).

B



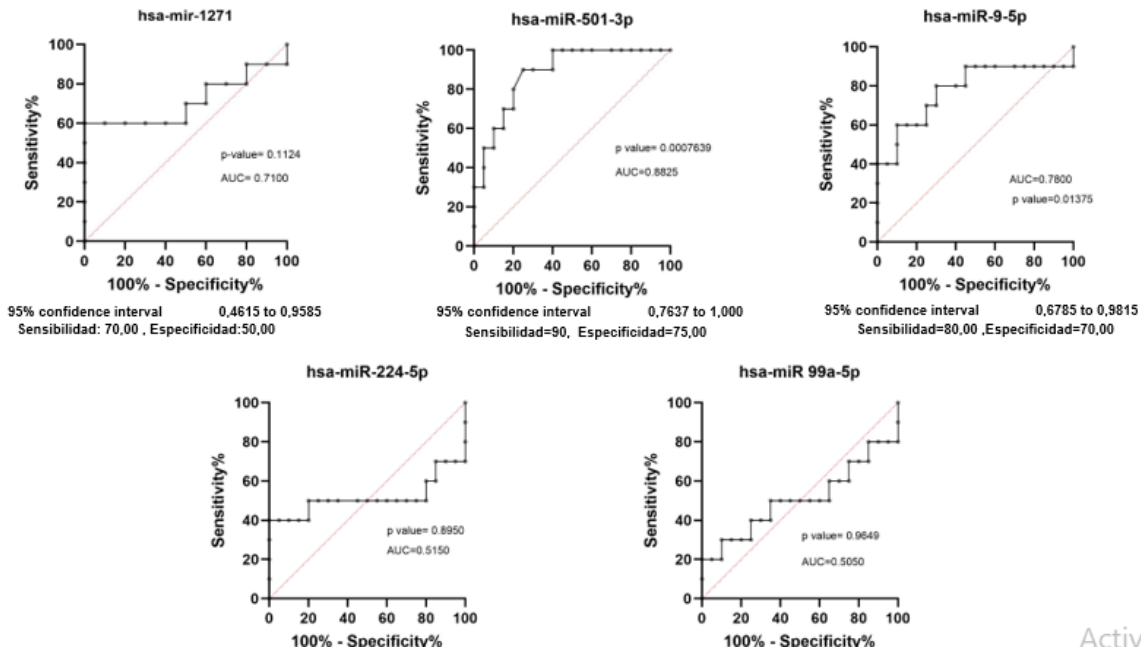
Se exploraron otros grupos de comparación considerando la distribución de las tres lesiones pre-neoplásicas del cérvix en el conjunto de muestras (20 NIC1, 5 NIC2 y 5 NIC3). La expresión relativa de los miRNAs mostró diferencias significativas ($p < 0.05$, Kruskal-Wallis) entre al menos uno de los grupos de comparación. En particular, los miRNAs hsa-miR-501 y hsa-miR-9-5p presentaron una mayor significancia entre los grupos NIC1 y NIC3, que entre NIC1 y NIC2 que también fueron significativos, lo que sugiere su potencial rol en la identificación de las lesiones consideradas precursoras del CCU. Por otro lado, el hsa-miR-1271 mostró significancia estadística entre los grupos NIC1 y NIC2, lo que puede indicar su papel en la identificación temprana de las lesiones de alto grado. Es posible que diversos mecanismos moleculares influyan en que su expresión sea más baja en NIC3. Además, es importante señalar que el número limitado de muestras (5 NIC2 y 5 NIC3) podría no ser suficiente para detectar diferencias significativas adicionales entre los grupos NIC1 y NIC3. Por lo tanto, se requieren estudios adicionales con un tamaño de muestra mayor para validar estos hallazgos y comprender mejor el papel de hsa-miR-1271 en la progresión de las lesiones pre-neoplásicas del cérvix (figura 33).

Figura 33. Expresión relativa $\text{Log}_2(2^{-\Delta\Delta\text{Ct}})$ de los miRNAs por grupo de lesión (NIC1, NIC2, NIC3) en las muestras de validación.



Para evaluar la utilidad clínica de los miRNAs en el grupo de validación, se generaron curvas ROC para el desenlace NIC2/3, tanto para cada miRNA individual como para combinaciones de miRNAs que mostraron significancia individual. Los análisis se realizaron utilizando los valores de expresión relativa $\text{Log}_2(2^{-\Delta\Delta\text{Ct}})$. En la figura 34 se presentan los valores de p, AUC y los puntos de corte correspondientes a la sensibilidad y especificidad para cada miRNA por separado. Destaca el hsa-miR-501-3p, que mostró una excelente capacidad discriminativa para distinguir entre NIC2/3, con un valor p significativo ($p < 0.00076$) y un AUC del 88% (IC 95%: 0.76-1.0). El miRNA hsa-miR-9a-5p también demostró un buen rendimiento, con un valor p de 0.001 y un AUC del 76% (IC 95%: 0.67-0.98).

Figura 34. Curvas ROC para cada miRNA individual para la detección de NIC2/3 en los datos de validación.



Activ

Basándose en los resultados de las curvas ROC para los miRNAs individuales, se decidió generar una curva ROC combinando los miRNAs que demostraron una alta significancia en la separación de NIC2/3. Específicamente, se combinó hsa-miR-501-3p con hsa-miR-9-5p (Figura 35) para evaluar el efecto de una posible firma compuesta por estos dos miRNAs. Esta combinación mostró un AUC del 90%, con una sensibilidad del 90%, especificidad del 75%, y un valor p de 0.002. Dado que hsa-miR-1271 presentó valores muy cercanos a la significancia estadística y se destacó en varios análisis a lo largo del estudio, se decidió incluirlo en una segunda curva ROC junto con los otros dos miRNAs (Figura 36). Esta combinación resultó en una curva muy buena, con un AUC del 94% y una sensibilidad y especificidad del 90%.

Figura 35. Curva ROC de las combinaciones de hsa-miR-9-5p y hsa-miR-501-3p para la detección de NIC2/3.

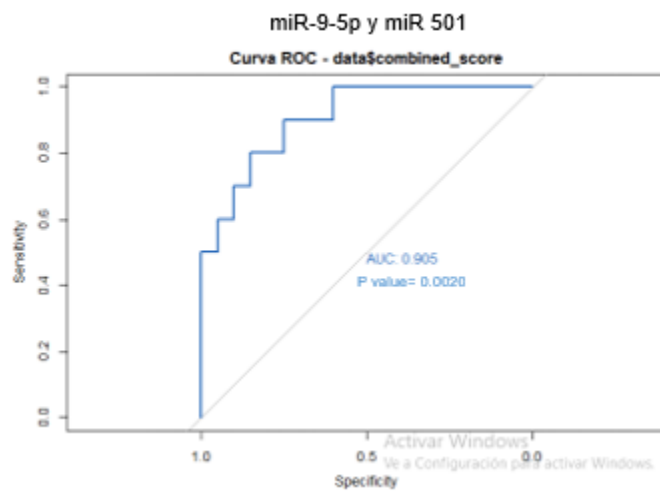
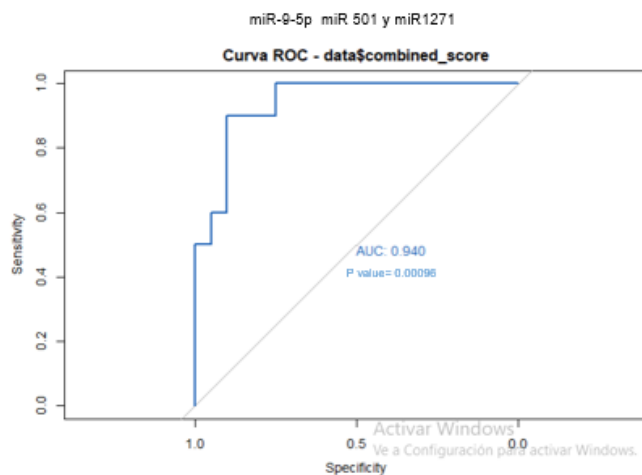


Figura 36. Curva ROC de las combinaciones de hsa-miR-9-5p, hsa-miR-501-3p y hsa-miR-1271 para la detección de NIC2/3.



8. Discusión

La prueba de tamizaje para VPH-AR ha demostrado ser altamente sensible en la detección de neoplasias cervicales, superando a la citología. Sin embargo, su valor predictivo positivo es limitado debido a la alta prevalencia de infecciones transitorias por VPH, muchas de las cuales no progresan a neoplasias (135, 72). Esto subraya la necesidad de un triaje más preciso para identificar a las mujeres con riesgo de progresión a cáncer de cervix y que requieren remisión a biopsia. En este contexto, la desregulación de miRNAs ha emergido como un potencial biomarcador para el diagnóstico, pronóstico y tratamiento del cáncer cervical (58,72,136, 142,143).

Nuestro estudio utilizó secuenciación de ARN pequeño en cepillados cervicales para identificar miRNAs desregulados. A pesar de las dificultades en la recolección de muestras, obtuvimos secuencias de alta calidad, validando la robustez de la metodología. Investigaciones previas han mostrado que la secuenciación de miRNAs es reproducible incluso en muestras con bajos niveles de ARN, utilizando plataformas como NEB Next, NEXTFlex y TruSeq, las cuales proporcionan resultados confiables (137, 141), lo que nos indica que las variaciones en la calidad del ARN no afectaron significativamente nuestros hallazgos (144).

Aunque la variabilidad observada en los datos de expresión de miRNAs entre replicas, dificultó la separación clara entre las lesiones de bajo y alto grado, sugiriendo que las diferencias en la expresión de miRNAs entre estos grupos son sutiles como para ser detectadas por los primeros componentes principales. Este hallazgo es coherente con estudios previos que señalan las limitaciones del PCA en la identificación de diferencias genéticas cuando las variaciones son mínimas (145). El uso de técnicas de agrupamiento no supervisado como el clustering jerárquico y K-means tampoco permitió una separación clara de las muestras según las condiciones histológicas conocidas. Este solapamiento entre los grupos coincide con trabajos que han señalado la alta variabilidad de los miRNAs entre réplicas de lesiones intraepiteliales de diferentes grados, resaltando la importancia de interpretar con cautela el uso de los miRNAs como biomarcadores en lesiones preneoplásicas (66, 146). Más que los problemas técnicos, la heterogeneidad de las

muestras, la carga viral, los diferentes tipos de VPH y las diferencias interindividuales en la expresión de miRNAs son los factores que podrían estar influyendo en esta falta de separación clara como demostró Pereira et al (146).

Además, la variabilidad en el número de muestras entre las categorías estudiadas, particularmente la baja cantidad de muestras en la categoría "sin lesión" (SL), pudo haber limitado nuestra capacidad para detectar diferencias significativas en la expresión de miRNAs, puesto que la literatura sugiere que se necesitan al menos 20 muestras por grupo para garantizar la robustez de los análisis de expresión diferencial (147). Por otra parte, la baja cantidad de miRNAs desregulados en la comparación entre SL y NIC1 sugiere que las lesiones de bajo grado (SL y NIC1) comparten características moleculares similares, lo que concuerda con el estudio de Wittenborn et al., quien reportó que las diferencias entre muestras sanas y NIC1 no fueron significativas, por lo que excluyó los NIC1 de sus análisis (185). La agrupación de las muestras en categorías más amplias (SL/NIC1 y NIC2/3) refleja un enfoque más clínico y práctico, alineado con la necesidad de simplificar los análisis sin comprometer la relevancia clínica de los hallazgos (8, 16, 65).

Identificamos 19 miRNAs diferencialmente expresados entre SL/NIC1 y NIC2/3, de los cuales 14 fueron comunes en ambas comparaciones y cuatro miRNAs (hsa-miR-1246, hsa-miR-9-5p, hsa-miR-184, y hsa-miR-99a-5p) fueron validados mediante comparación con conjuntos de datos de GEO, fortaleciendo la consistencia y validez de nuestros resultados. Este tipo de integración de datos es clave para mejorar la reproducibilidad en estudios de biomarcadores y subraya la necesidad de estandarizar técnicas en la investigación de miRNAs en lesiones pre neoplásicas y cáncer cervical, lo que puede resolver las inconsistencias en los perfiles de expresión de miRNAs evidenciados en la literatura y mejorar la identificación de biomarcadores precisos (148).

El análisis de clustering no supervisado con k-medias como estrategia inicial para identificar los DEmiARNs que mejor revelaran patrones de agrupamiento en los datos, sin depender de etiquetas predefinidas, permitió descubrir patrones significativos que podrían no ser evidentes (150, 151, 152), y reveló que 8 miARNs (hsa-miR-1271-5p, hsa-miR-9-5p, hsa-miR-3690, hsa-miR-6510-3p, hsa-miR-4662a-5p, hsa-miR-184, hsa-miR-501-3p y hsa-miR-1246) segregaron las muestras en dos clusters, y que en el Cluster 2

predominaron las muestras NIC2/3 (21 muestras) y solo 4 SL/NIC1. Aunque la separación no fue perfecta, los miARNs explicaron más del 50% de la variabilidad. Las coinfecciones con distintos tipos de VPH-AR podrían haber influido en estos resultados (146,153, 154).

Los resultados de nuestro estudio resaltan la importancia de tener en cuenta el tipo de virus presente en estas infecciones al buscar biomarcadores o evaluar la expresión diferencial de miRNAs, especialmente para los tipos de VPH-AR más prevalentes y oncogénicos, como el VPH16 y VPH18, responsables de más del 50% de los cánceres cervicales (9, 174). Aún así, es importante resaltar que al incluir varios genotipos de VPH-AR, se refleja la realidad clínica, donde las infecciones por múltiples tipos son comunes y aumentan el riesgo de progresión a lesiones de alto grado. (155,7, 156).

La precisión en la clasificación de muestras es clave para el análisis de clustering. En este estudio, las etiquetas de las lesiones cervicales se basaron en diagnósticos por biopsias guiadas por colposcopia, considerado el estándar clínico (157). No obstante, la experiencia del colposcopista y el número de biopsias tomadas pueden influir en la precisión diagnóstica, especialmente en el caso de las NIC2, lo que puede afectar tanto la interpretación de los datos como el clustering (158, 8, 20). Etiquetas inexactas pueden alterar la capacidad del análisis no supervisado para agrupar correctamente las muestras. En estudios previos, como el de Snoek et al., se excluyeron las NIC2 debido a su ambigüedad diagnóstica (65), mientras que Herrington et al. sugieren una clasificación de dos niveles (SIL bajo y alto grado) en lugar del sistema de tres grados NIC (159). En nuestro caso, aunque se observó una separación entre clusters, esta no fue perfecta, lo que coincide con investigaciones previas donde el clustering no supervisado no logró diferenciar completamente entre lesiones como NIC2 y NIC3 (160).

El análisis de la capacidad discriminativa de los miRNAs mediante curvas ROC mostró que miRNAs como hsa-miR-1278 y hsa-miR-6510-3p tuvieron una capacidad moderada para discriminar lesiones de alto grado. Estos resultados son consistentes con el análisis de clustering, donde estos miRNAs también fueron identificados como relevantes para separar las muestras en dos clusters principales. La inclusión de una firma combinada de varios miRNAs podría mejorar la capacidad discriminativa y aumentar la precisión del triaje, como han sugerido otros estudios (149, 150, 151, 152).

Finalmente, los análisis de regresión logística bivariada indicaron que hsa-miR-1271-5p y hsa-miR-4662a-5p están asociados significativamente con el riesgo de presentar lesiones de alto grado (NIC2/3). Estos hallazgos concuerdan con estudios previos que sugieren que el uso de combinaciones de miRNAs y variables clínicas, como el uso prolongado de anticonceptivos hormonales, puede mejorar la predicción del riesgo de progresión a cáncer de cervix (161). La combinación de análisis ROC, clustering y regresión logística nos permitió identificar un conjunto de miRNAs candidatos para realizar la fase de validación que nos permita, mejorar la estratificación del riesgo en mujeres con infección por VPH-AR y avanzar hacia una firma basada en miRNAs.

El análisis de enriquecimiento funcional enfatiza la importancia de los miRNAs para la progresión de las lesiones cervicales asociadas al VPH. Los genes regulados por miRNAs como hsa-miR-1271, hsa-miR-9-5p, hsa-miR-501 y hsa-miR-1246 están involucrados en procesos clave como la adhesión celular, remodelación del citoesqueleto, regulación nuclear, y el ciclo celular, lo que sugiere que los miRNAs modulan eventos esenciales en la tumorigénesis, como la pérdida de adhesión y la inestabilidad genómica (162,163,164). Estos mecanismos también se ven reflejados en componentes citoesqueléticos y gránulos de estrés, los cuales indican que las células preneoplásicas experimentan estrés celular inducido por el virus, facilitando la progresión hacia estados malignos. En términos de Procesos Biológicos y Función Molecular, se destacan vías como la fosforilación de peptidil-serina y la respuesta a la hipoxia, fundamentales para enfrentar el daño celular causado por el VPH y evitar la acumulación de alteraciones genéticas (165, 166, 167,168). El análisis de las vías KEGG subraya la implicación de estos miRNAs en rutas virales, especialmente "Human papillomavirus infection", lo que enfatiza su potencial como biomarcadores para el triaje en mujeres con lesiones preneoplásicas. También se observaron vías relacionadas con otros virus oncogénicos, como el virus de Epstein-Barr y el HTLV-1, que, al igual que el VPH, se asocian con la carcinogénesis cervical (162,163,164, 165). Esto resalta la complejidad de la interacción entre los miRNAs y las células infectadas por VPH en la progresión del cáncer cervical.

En este estudio, el miRNA hsa-miR-423 se utilizó como gen normalizador en la qPCR para minimizar la variabilidad técnica entre las muestras cervicales, siguiendo la recomendación de Babion et al., que lo identificó como un gen de referencia adecuado para raspados

cervicales (169). Esto permitió la validación de seis miRNAs previamente identificados en la fase de descubrimiento: se seleccionaron, cuatro (hsa-miR-1271-5p , hsa-miR-9-5p, hsa-miR-501-3p, hsa-miR-4662a-5p) que destacaron en los análisis de la fase de descubrimiento y estaban dentro de los 4 mejores miRNAs con gran potencial para discriminar entre NIC2/3 vs lesiones de bajo grado, mientras que dos (hsa-miR-224 y hsa-miR-99a-5p) se seleccionaron únicamente por su expresión diferencial entre SL/NIC1 vs NIC2/3. No se logró validar el hsa-miR-4662a-5p debido a sus bajos niveles de expresión, esta discrepancia puede deberse a diferencias entre los métodos utilizados (qPCR versus sRNA-Seq), como han señalado estudios anteriores, donde Snoek et al. tampoco pudieron validar tres miRNAs de la fase de descubrimiento. Aunque varios estudios han mostrado correlaciones entre qPCR y RNA-Seq, también se han informado diferencias notables en la cuantificación de miRNAs. La presencia de isomiRs, variantes de longitud de miRNAs, podría haber afectado la precisión de la detección mediante qPCR, lo que coincide con el hecho de que el mapeo en sRNA-Seq no permitió analizar estos isomiRs (65). A pesar de estas limitaciones, la validación general fue satisfactoria, ya que el uso de muestras independientes demuestra consistencia en los hallazgos obtenidos en las dos fases del estudio.

Tres de los cinco miRNAs (hsa-miR-1271-5p, hsa-miR-9-5p, hsa-miR-501-3p) mostraron diferencias significativas en la expresión relativa entre NIC1 y NIC2/3 ($p < 0.05$). No obstante, al transformar los datos a escala logarítmica, hsa-miR-1271-5p perdió significancia estadística, aunque seguía mostrando una tendencia hacia el aumento en NIC2/3. Esto es sorprendente, ya que, en la fase de descubrimiento, este miRNA mostró el mayor potencial discriminativo y se esperaba un resultado consistente. Este cambio puede deberse a que la transformación logarítmica, utilizada para estabilizar la varianza, también puede reducir las diferencias entre grupos y la significancia estadística; o las diferencias metodológicas entre RNA-Seq y qPCR, como se mencionó previamente. Por otro lado, hsa-miR-224 y hsa-miR-99a-5p, aunque expresados diferencialmente en la fase inicial, no lograron un poder discriminativo suficiente en la validación. Esto subraya la importancia de no basar la selección de biomarcadores únicamente en la expresión diferencial inicial, sino considerar múltiples análisis y contextos experimentales para obtener resultados robustos. Finalmente, se observó un incremento lineal en la expresión

de los tres miRNAs a medida que progresa la enfermedad, con una tendencia significativa ($p < 0.05$). Este patrón apoya su asociación con la progresión de las lesiones cervicales. No obstante, hsa-miR-1271-5p mostró un aumento significativo entre NIC1 y NIC2, pero no entre NIC1 y NIC3, lo que sugiere su posible relevancia en las etapas iniciales de la progresión, y se debería explorar en estudios futuros.

La capacidad discriminativa de los miRNAs en la validación para detectar lesiones de alto grado (NIC2/3) se evaluó mediante curvas ROC. Aunque el tamaño muestral no es ideal, los resultados obtenidos ofrecen una perspectiva preliminar sobre el valor de estos miRNAs como biomarcadores destacando que estos resultados deben ser interpretados únicamente como resultados exploratorios. Entre los miRNAs evaluados, hsa-miR-501-3p destacó con un AUC de 88.25% (IC 95%: 76.37-100.0), lo que indica una alta capacidad discriminatoria para NIC2/3. Este hallazgo está en línea con estudios recientes que han identificado a hsa-miR-501-5p como un oncomir crucial en la tumorigénesis del CCU. Se ha demostrado que miR-501-5p favorece la proliferación celular e inhibe la apoptosis a través de la regulación negativa de CYLD, un gen que codifica una proteína involucrada en la diferenciación y división celular, así como en la apoptosis. Además, su asociación con el grado histológico y la supervivencia en pacientes con CCU refuerza su potencial tanto como biomarcador diagnóstico como pronóstico, tal como lo confirman múltiples estudios que lo han incluido en firmas moleculares para el diagnóstico de este tipo de cáncer (87, 170, 171, 172).

Por su parte, hsa-miR-9-5p mostró un desempeño notable, con un AUC de 78.0% (IC: 67.85-98.15). La literatura científica ha documentado su implicación en la transformación neoplásica del cérvix, siendo sobreexpresado en tejidos NIC. Su relación con la regulación negativa de FOXO3, un factor clave en la reparación de daños en el ADN (DDR), sugiere que miR-9-5p juega un papel fundamental en la progresión de las lesiones preneoplásicas hacia el cáncer invasivo. Además, análisis bioinformáticos indican que miR-9 podría inhibir la diferenciación celular y promover la proliferación celular durante las primeras etapas de la transformación cervical inducida por el VPH (10, 172, 173, 174, 175). Cabe destacar que este miRNA también se identificó diferencialmente expresado en los análisis de datos de GEO realizados en este estudio.

En contraste, hsa-miR-1271 no mostró significancia estadística en el análisis ROC para la detección de NIC2/3, a pesar de haberse identificado como un marcador prometedor en la fase de descubrimiento. Esta falta de significancia podría deberse a artefactos técnicos o a la presencia de outliers en las muestras de NIC3, donde no se observó una diferencia estadística en comparación con NIC1. A diferencia, en las muestras de NIC2 se observó una expresión significativa de miR-1271 en relación con NIC1, lo que sugiere que se necesita un mayor tamaño muestral para confirmar su potencial diagnóstico u otros análisis para determinar el papel de este miRNA en el desarrollo del CCU. Lo anterior basado en la literatura, donde hsa-miR-1271 ha sido descrito como un supresor tumoral en diversos tipos de cáncer con expresión a la baja por lo general, y como un posible oncogén en el carcinoma hepatocelular, lo que sugiere una función dual dependiendo del contexto biológico (176). Sin embargo, su rol específico en el cáncer de cuello uterino (CCU) y lesiones preneoplásicas no ha sido descrito previamente, siendo este estudio una primera aproximación a su implicación en esta patología. Esta dualidad funcional podría explicar la inconsistencia observada en nuestros resultados, donde, a pesar de no alcanzar significancia en el análisis ROC, se evidenció una tendencia hacia una mayor expresión en lesiones de alto grado, específicamente en NIC2 en comparación con NIC1. Es fundamental investigar más a fondo el papel de este miRNA en el CCU, ya que podría representar un biomarcador clave en las primeras etapas de la progresión neoplásica o incluso ser útil en la identificación de subgrupos de pacientes con diferente pronóstico, lo que contribuiría a mejorar las estrategias de detección temprana y manejo personalizado de las lesiones cervicales.

La combinación de miR-9-5p, miR-501-3p, y miR-1271-5p resultó ser la combinación más eficaz para la detección de NIC2/3, con un AUC de 94.0%, superando a los miRNAs individuales. Este hallazgo refuerza la idea de que una firma combinada de miRNAs puede ser más robusta que la evaluación de marcadores individuales, tal como se ha observado en otros estudios donde la combinación de 2-5 miRNAs logró clasificaciones más precisas de lesiones cervicales de alto grado (71, 65, 66). En este contexto, es relevante destacar que incluso un solo miRNA, como miR-501-3p, puede alcanzar un AUC adecuado para la identificación de NIC2/3. No obstante, se requieren estudios adicionales que incluyan muestras más grandes y la correlación con biomarcadores clínicos establecidos, como la

genotipificación del VPH y la tinción dual p16/Ki67, la metilación (66) y factores clínicos o sociodemográficos para una mayor comprensión del papel biológico de los miRNAs en la progresión de las lesiones cervicales y, al mismo tiempo, reforzaría su potencial utilidad como herramientas de triaje clínico.

9. Conclusiones

- Se identificaron 19 miRNAs con expresión diferencial entre lesiones de bajo (SL/NIC1) y alto grado (NIC2/3). De estos, 14 coincidieron con resultados de las comparaciones entre los diferentes tipos de lesiones (SL, NIC1, NIC2/3), 14 coincidieron con los resultados obtenidos a través de la herramienta ExpressAnalyst, y 4 miRNAs se encontraron en común con datos externos. Estos hallazgos sugieren una sólida correlación entre los miRNAs identificados y refuerzan su relevancia en la carcinogénesis cervical, subrayando su potencial como biomarcadores en diversos contextos.
- Es fundamental reconocer que el clustering no supervisado es una herramienta exploratoria para identificar patrones en los datos y en este estudio guió la selección de miRNAs informativos mediante el agrupamiento de las muestras según su grado de lesión. Aunque debe interpretarse con cautela y complementarse con métodos supervisados y validaciones con datos externos para mejorar la precisión y reproducibilidad en la clasificación en distintos contextos.
- El descubrimiento de biomarcadores requiere un enfoque cauteloso y el uso de métodos complementarios. En nuestro estudio, la evaluación de diversos parámetros, como se detalla en las tablas 15 y 17, identificó a **hsa-miR-1271**, **hsa-miR-9-5p**, **hsa-miR-501** y **hsa-miR-1246** como los principales candidatos, ya que cumplen con la mayoría de los criterios evaluados. Entre ellos, **hsa-miR-1271** destaca como el biomarcador que reúne la

mayoría de características analizadas para la discriminación de lesiones cervicales de alto grado (NIC2/3), considerándose como los principales miRNAs para una validación futura.

- La regresión logística bivariada identificó el uso prolongado de anticonceptivos como un factor de riesgo significativo para el desarrollo de lesiones cervicales de alto grado. Este hallazgo subraya la necesidad de integrar tanto los marcadores biológicos como los factores de riesgo clínicos en la detección temprana de estas lesiones y evaluar su impacto a través de modelos multivariados.

- De los seis miRNAs validados mediante qPCR, tres mostraron cambios significativos en su expresión relativa, en lesiones NIC2/3: **hsa-miR-1271, hsa-miR-9-5p, y hsa-miR-501**. Estos miRNAs no solo exhibieron una expresión diferencial destacada, sino que también fueron seleccionados como candidatos durante la fase de descubrimiento debido a sus altos puntajes. Además, tienen una relevancia biológica considerable al regular genes asociados con el cáncer de cuello uterino y la infección por VPH, como se detalla en la Tabla 15.

- Los miRNAs validados en este estudio, **hsa-miR-1271, hsa-miR-9-5p, y hsa-miR-501**, desempeñan un papel crucial en la progresión de las lesiones intraepiteliales cervicales. Como se muestra en la Figura 32, los niveles de hsa-miR-1271, hsa-miR-9-5p, y hsa-miR-501 fueron significativamente mayores en lesiones NIC2 y NIC3 en comparación con NIC1 (p tendencia <0.05). Sin embargo, los niveles de expresión relativa de hsa-miR-1271 no fueron significativamente diferentes entre lesiones NIC3 y NIC1.

- Los miRNAs hsa-miR-9-5p, hsa-miR-501 y sus combinaciones (hsa-miR-9-5p + hsa-miR-501 y hsa-miR-9-5p + hsa-miR-501 + hsa-miR-1271) mostraron una capacidad discriminativa destacada para diferenciar NIC2/3 de NIC1, con AUC de 78%, 88%, 90%, 94%, respectivamente. Aunque hsa-miR-501 sobresale individualmente, las combinaciones de miRNAs ofrecen mayor robustez diagnóstica.

- Los resultados obtenidos respaldan la importancia de validar los demás miRNAs identificados como candidatos en la fase de descubrimiento, ya que muestran un papel prometedor en la detección de lesiones de alto grado y en la evaluación del riesgo de cáncer de cuello uterino (CCU), reflejando eventos moleculares clave en el desarrollo de la enfermedad. Es crucial utilizar un número óptimo de muestras para garantizar que los

resultados sean comparables y estadísticamente significativos, lo que fortalecerá la fiabilidad de los hallazgos y su relevancia en la progresión de la enfermedad.

- Los hallazgos de este estudio, realizados en muestras de exfoliado cervical, refuerzan la utilidad clínica de este tipo de muestra para la detección de miRNAs. Dado que el exfoliado cervical es fácilmente accesible y se utiliza comúnmente en la prueba de VPH, su empleo para evaluar la expresión de miRNAs permite integrar la detección de ambos biomarcadores en una única técnica de PCR. Esto podría mejorar el diagnóstico y la estratificación del riesgo en mujeres VPH-AR+ para identificar lesiones preneoplásicas de alto grado.

10. Recomendaciones y perspectivas

- Aumentar el tamaño de la muestra en el grupo de validación para evaluar con mayor precisión la utilidad clínica y el desempeño diagnóstico de los miRNAs en la detección de lesiones de alto grado.

- Para futuros estudios se recomienda el uso de muestras frescas para minimizar la degradación del RNA y optimizar la calidad y fiabilidad de los resultados obtenidos.

- Desarrollar ensayos funcionales que confirmen la participación de los miRNAs en el proceso carcinogénico del cérvix, y comprender mejor su papel en la progresión de la enfermedad.

- Considerar la selección de muestras y el tipo específico de VPH-AR en los análisis para determinar si existen diferencias significativas entre tipos de VPH-AR y evaluar si el tipo de VPH influye en la identificación de biomarcadores en mujeres positivas para VPH.

- Realizar una evaluación de los miRNAs en las lesiones intraepiteliales cervicales (SL, NIC1, NIC2, NIC3, CÁNCER) utilizando un número suficiente de muestras para esclarecer

el papel de estos miRNAs en la progresión de la enfermedad y determinar su potencial como biomarcadores y su relevancia clínica para la estratificación del riesgo. Además, investigar el papel de los miRNAs que se regulan a la baja, para lograr una comprensión más profunda del proceso carcinogénico.

- Realizar estudios exhaustivos a las secuencias para identificar miRNAs no reportados que desempeñen un papel clave en el desarrollo de lesiones de alto grado y cáncer cervical. El uso de smallRNAseq, por su alta sensibilidad y capacidad de cobertura, revela un gran número de lecturas no mapeadas que constituyen una fuente valiosa de información sobre moléculas no anotadas.

- Los miRNAs que, en la fase de descubrimiento, fueron identificados como prometedores pero que no fueron validados, requieren ser evaluados en futuros estudios para confirmar su potencial como biomarcadores.

- Considerar la evaluación de factores adicionales que puedan influir en la regulación de los miRNAs, y las características clínicas de las pacientes para determinar su potencial utilidad en la estratificación del riesgo.

11. BIBLIOGRAFIA

1. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, Bray F. Cancer statistics for the year 2020: An overview. *Int J Cancer*. 2021 Apr 5. doi: 10.1002/ijc.33588. Epub ahead of print. PMID: 33818764.
2. Unión Internacional contra el Cáncer. (2020, diciembre 16). GLOBOCAN 2020: Global cancer data. UICC. Recuperado de <https://www.uicc.org/news/globocan-2020-global-cancer-data>
3. International Agency for Research on Cancer. (2022). Colombia fact sheet. GLOBOCAN 2020. Recuperado de <https://gco.iarc.who.int/media/globocan/factsheets/populations/170-colombia-fact-sheet.pdf>
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021; 71: 209-249. <https://doi.org/10.3322/caac.21660>
5. Chayo I, Uribe Perez CJ, De Vries E, Pinheiro PS. The impact of health insurance affiliation and socioeconomic status on cervical cancer survival in Bucaramanga, Colombia. *Cancer Epidemiol*. 2023 Aug;85:102375. doi: 10.1016/j.canep.2023.102375. Epub 2023 May 5. PMID: 37150101.
6. Hernández Vargas JA, Ramírez Barbosa PX, Valbuena-García AM, Acuña L, González-Díaz JA. Factors associated with delays in time to treatment initiation in Colombian women with cervical cancer: A cross-sectional analysis. *Gynecol Oncol Rep*. 2021 Jan 7;35:100697. doi: 10.1016/j.gore.2021.100697. PMID: 33532535; PMCID: PMC7822946.
7. Piña-Sánchez P. Human Papillomavirus: Challenges and Opportunities for the Control of Cervical Cancer. *Arch Med Res*. 2022 Dec;53(8):753-769. doi: 10.1016/j.arcmed.2022.11.009. Epub 2022 Nov 30. PMID: 36462952.
8. Ye J, Cheng XD, Cheng B, Cheng YF, Chen XJ, Lu WG. MiRNA detection in cervical exfoliated cells for missed high-grade lesions in women with LSIL/CIN1 diagnosis after colposcopy-guided biopsy. *BMC Cancer*. 2019 Jan 30;19(1):112. doi: 10.1186/s12885-019-5311-3. PMID: 30700264; PMCID: PMC6354336.
9. Berti FCB, Salviano-Silva A, Beckert HC, de Oliveira KB, Cipolla GA, Malheiros D. From squamous intraepithelial lesions to cervical cancer: Circulating microRNAs as potential biomarkers in cervical carcinogenesis. *Biochim Biophys Acta Rev Cancer*. 2019 Dec;1872(2):188306. doi: 10.1016/j.bbcan.2019.08.001. Epub 2019 Aug 6. PMID: 31398380.
10. Choi PW, Liu TL, Wong CW, Liu SK, Lum YL, Ming WK. The Dysregulation of MicroRNAs in the Development of Cervical Pre-Cancer-An Update. *Int J Mol Sci*. 2022 Jun 27;23(13):7126. doi: 10.3390/ijms23137126. PMID: 35806128; PMCID: PMC9266862.
11. InfoCancer. (n.d.). Portal de InfoCancer. Recuperado de <https://www.infocancer.co/portal/#/home>
12. Yunes-Díaz E, Ruiz PA, Lazcano-Ponce E. Assessment of the Validity and Reproducibility of the Pap Smear in Mexico: Necessity of a Paradigm Shift. *Arch Med Res*. 2015 May;46(4):310-6. doi: 10.1016/j.arcmed.2015.05.013. Epub 2015 Jun 6. PMID: 26054968.
13. Wu Q, Wang L, Zhao X, Tian Q, Wang F, Sima N, Qiu L, Lu W, Xie X, Wang X, Cheng X. The Value of MicroRNA-375 Detection for Triaging Primary Human Papillomavirus Positive Women: A Cross-Sectional Study in a General Population. *Front Oncol*. 2021 Oct 28;11:771053. doi: 10.3389/fonc.2021.771053. PMID: 34778090; PMCID: PMC8581639.
14. Whitlock EP, Vesco KK, Eder M, Lin JS, Senger CA, Burda BU. Liquid-based cytology and human papillomavirus testing to screen for cervical cancer: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2011 Nov 15;155(10):687-97. W214-5. doi: 10.7326/0003-4819-155-10-2011111150-00376. Epub 2011 Oct 17. Erratum in: *Ann Intern Med*. 2012 Jan 3;156(1 Pt 1):71-2. PMID: 22006930.
15. Siebers AG, Klinkhamer PJ, Grefte JM, Massuger LF, Vedder JE, Beijers-Broos A, Bulten J, Arbyn M. Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: a randomized controlled trial. *JAMA*. 2009 Oct 28;302(16):1757-64. doi: 10.1001/jama.2009.1569. Erratum in: *JAMA*. 2009 Dec 2;302(21):2322. PMID: 19861667.
16. Koliopoulos G, Nyaga VN, Santesso N, Bryant A, Martin-Hirsch PP, Mustafa RA, Schünemann H, Paraskevaidis E, Arbyn M. Cytology versus HPV testing for cervical cancer screening in the general population. *Cochrane Database Syst Rev*. 2017 Aug 10;8(8):CD008587. doi: 10.1002/14651858.CD008587.pub2. PMID: 28796882; PMCID: PMC6483676.
17. Huh WK, Ault KA, Chelmow D, Davey DD, Goulart RA, Garcia FA, Kinney WK, Massad LS, Mayeaux EJ, Saslow D, Schiffman M, Wentzensen N, Lawton HW, Einstein MH. Use of primary high-risk human papillomavirus testing for cervical cancer screening: interim clinical guidance. *J Low Genit Tract Dis*. 2015 Apr;19(2):91-6. doi: 10.1097/LGT.000000000000103. PMID: 25574659.
18. Perkins RB, Guido RS, Castle PE, Chelmow D, Einstein MH, Garcia F, Huh WK, Kim JJ, Moscicki AB, Nayar R, Saraiya M, Sawaya GF, Wentzensen N, Schiffman M; 2019 ASCCP Risk-Based Management Consensus Guidelines Committee. 2019 ASCCP Risk-Based Management Consensus Guidelines for Abnormal Cervical Cancer Screening Tests and Cancer Precursors. *J Low Genit Tract Dis*. 2020 Apr;24(2):102-131. doi: 10.1097/LGT.0000000000000525. Erratum in: *J Low Genit Tract Dis*. 2020 Oct;24(4):427. doi: 10.1097/LGT.0000000000000563. PMID: 32243307; PMCID: PMC7147428.
19. Ronco G, Dillner J, Elfström KM, Tunesi S, Snijders PJ, Arbyn M, Kitchener H, Segnan N, Gilham C, Giorgi-Rossi P, Berkhof J, Peto J, Meijer CJ; International HPV screening working group. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet*. 2014 Feb

- 8;383(9916):524-32. doi: 10.1016/S0140-6736(13)62218-7. Epub 2013 Nov 3. Erratum in: *Lancet*. 2015 Oct 10;386(10002):1446. doi: 10.1016/S0140-6736(15)00411-0. PMID: 24192252.
20. Tian Q, Li Y, Wang F, Li Y, Xu J, Shen Y, Ye F, Wang X, Cheng X, Chen Y, Wan X, Lu W, Xie X. MicroRNA detection in cervical exfoliated cells as a triage for human papillomavirus-positive women. *J Natl Cancer Inst*. 2014 Sep 4;106(9):dju241. doi: 10.1093/jnci/dju241. PMID: 25190727; PMCID: PMC4188123.
 21. Garzón-Olivares CD, Grillo-Ardila CF, Amaya-Guio J, Vallejo-Ortega MT, Diaz-Cruz LA, Fernández-Motta C, et al. Tamización primaria con prueba ADN-VPH en mujeres menores de 30 años: evaluación de tecnología sanitaria. *Rev. colomb. obstet. ginecol.* [Internet]. 30 de junio de 2022 [citado 21 de junio de 2024];73(2):203-22. Disponible en: <https://revista.fecolsog.org/index.php/rcog/article/view/3866>
 22. Chrysostomou AC, Stylianou DC, Constantinidou A, Kostrikis LG. Cervical Cancer Screening Programs in Europe: The Transition Towards HPV Vaccination and Population-Based HPV Testing. *Viruses*. 2018 Dec 19;10(12):729. doi: 10.3390/v10120729. PMID: 30572620; PMCID: PMC6315375.
 23. Hashim D, Engesæter B, Baadstrand Skare G, Castle PE, Bjørge T, Tropé A, Nygård M. Real-world data on cervical cancer risk stratification by cytology and HPV genotype to inform the management of HPV-positive women in routine cervical screening. *Br J Cancer*. 2020 May;122(11):1715-1723. doi: 10.1038/s41416-020-0790-1. Epub 2020 Apr 3. PMID: 32242098; PMCID: PMC7250848.
 24. Castle PE, Stoler MH, Wright TC Jr, Sharma A, Wright TL, Behrens CM. 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. *Lancet Oncol*. 2011 Sep;12(9):880-90. doi: 10.1016/S1470-2045(11)70188-7. Epub 2011 Aug 22. PMID: 21865084.
 25. Franceschi S, Denny L, Irwin KL, Jeronimo J, Lopalco PL, Monsonego J, Peto J, Ronco G, Sasieni P, Wheeler CM. Eurogin 2010 roadmap on cervical cancer prevention. *Int J Cancer*. 2011 Jun 15;128(12):2765-74. doi: 10.1002/ijc.25915. Epub 2011 Mar 25. PMID: 21207409.
 26. Wittenborn J, Weikert L, Hangarter B, Stickeler E, Maurer J. The use of micro RNA in the early detection of cervical intraepithelial neoplasia. *Carcinogenesis*. 2020 Dec 31;41(12):1781-1789. doi: 10.1093/carcin/bgaa046. PMID: 32417880.
 27. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature*. 2008 Sep 4;455(7209):64-71. doi: 10.1038/nature07242. Epub 2008 Jul 30. PMID: 18668037; PMCID: PMC2745094.
 28. Bushati N, Cohen SM. microRNA functions. *Annu Rev Cell Dev Biol*. 2007;23:175-205. doi: 10.1146/annurev.cellbio.23.090506.123406. PMID: 17506695.
 29. Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005 Feb 17;433(7027):769-73. doi: 10.1038/nature03315. Epub 2005 Jan 30. PMID: 15685193.
 30. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. *Nature*. 2008 Sep 4;455(7209):58-63. doi: 10.1038/nature07228. Epub 2008 Jul 30. PMID: 18668040.
 31. Filipowicz, W., Bhattacharyya, S. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?. *Nat Rev Genet* 9, 102–114 (2008). <https://doi.org/10.1038/nrg2290>
 32. Witting SM, Snijders PJ, Verlaet W, Jaspers A, van de Wiel MA, van Wieringen WN, Meijer GA, Kenter GG, Yi Y, le Sage C, Agami R, Meijer CJ, Steenbergen RD. Altered microRNA expression associated with chromosomal changes contributes to cervical carcinogenesis. *Oncogene*. 2013 Jan 3;32(1):106-16. doi: 10.1038/onc.2012.20. Epub 2012 Feb 13. PMID: 22330141.
 33. He Y, Lin J, Ding Y, Liu G, Luo Y, Huang M, Xu C, Kim TK, Etheridge A, Lin M, Kong D, Wang K. A systematic study on dysregulated microRNAs in cervical cancer development. *Int J Cancer*. 2016 Mar 15;138(6):1312-27. doi: 10.1002/ijc.29618. Epub 2015 Jun 12. PMID: 26032913.
 34. Kawai S, Fujii T, Kukimoto I, Yamada H, Yamamoto N, Kuroda M, Otani S, Ichikawa R, Nishio E, Torii Y, Iwata A. Identification of miRNAs in cervical mucus as a novel diagnostic marker for cervical neoplasia. *Sci Rep*. 2018 May 4;8(1):7070. doi: 10.1038/s41598-018-25310-1. PMID: 29728572; PMCID: PMC5935744.
 35. Beaudenon, S., Huibregtse, J.M. HPV E6, E6AP and cervical cancer. *BMC Biochem* 9 (Suppl 1), S4 (2008). <https://doi.org/10.1186/1471-2091-9-S1-S4>
 36. Granados López AJ, López JA. Multistep model of cervical cancer: participation of miRNAs and coding genes. *Int J Mol Sci*. 2014 Sep 4;15(9):15700-33. doi: 10.3390/ijms150915700. PMID: 25192291; PMCID: PMC4200848.
 37. Chocontá-Piraquive LA, Alvis-Guzman N, De la Hoz-Restrepo F. How protective is cervical cancer screening against cervical cancer mortality in developing countries? The Colombian case. *BMC Health Serv Res*. 2010 Sep 16;10:270. doi: 10.1186/1472-6963-10-270. PMID: 20846446; PMCID: PMC2949854.
 38. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021 May;71(3):209-249. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.
 39. Attademo L, Tuninetti V, Pisano C, Cecere SC, Di Napoli M, Tambaro R, Valabrega G, Musacchio L, Setola SV, Piccirillo P, Califano D, Spina A, Losito S, Greggi S, Pignata S. Immunotherapy in cervix cancer. *Cancer Treat Rev*. 2020 Nov;90:102088. doi: 10.1016/j.ctrv.2020.102088. Epub 2020 Aug 7. PMID: 32827839.
 40. Baena-Acevedo, J. D., Adriana, D., García-Robayo, Á. C.-A., Aristizábal-Gutiérrez, F., & Castañeda-Peláez, D. A. (2021). SFRP1, Posible biomarcador en la progresión o regresión de lesiones de cérvix asociado al Virus del Papiloma Humano. *Infectio*, 25(4), 270. <https://doi.org/10.22354/in.v25i4.959>

41. Castle PE, Schiffman M, Wheeler CM, Solomon D. Evidence for frequent regression of cervical intraepithelial neoplasia-grade 2. *Obstet Gynecol.* 2009 Jan;113(1):18-25. doi: 10.1097/AOG.0b013e31818f5008. PMID: 19104355; PMCID: PMC2694845.
42. Ostör AG. Natural history of cervical intraepithelial neoplasia: a critical review. *Int J Gynecol Pathol.* 1993 Apr;12(2):186-92. PMID: 8463044.
43. Turchinovich A, Weiz L, Langheinz A, Burwinkel B. Characterization of extracellular circulating microRNA. *Nucleic Acids Res.* 2011 Sep 1;39(16):7223-33. doi: 10.1093/nar/gkr254. Epub 2011 May 24. PMID: 21609964; PMCID: PMC3167594.
44. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature.* 2005 Feb 17;433(7027):769-73. doi: 10.1038/nature03315. Epub 2005 Jan 30. PMID: 15685193.
45. Tang F, Hajkova P, Barton SC, Lao K, Surani MA. MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Res.* 2006 Jan 24;34(2):e9. doi: 10.1093/nar/gnj009. PMID: 16434699; PMCID: PMC1351374.
46. He Y, Lin J, Ding Y, Liu G, Luo Y, Huang M, Xu C, Kim TK, Etheridge A, Lin M, Kong D, Wang K. A systematic study on dysregulated microRNAs in cervical cancer development. *Int J Cancer.* 2016 Mar 15;138(6):1312-27. doi: 10.1002/ijc.29618. Epub 2015 Jun 12. PMID: 26032913.
47. Jung M, Schaefer A, Steiner I, Kempkensteffen C, Stephan C, Erbersdobler A, Jung K. Robust microRNA stability in degraded RNA preparations from human tissue and cell samples. *Clin Chem.* 2010 Jun;56(6):998-1006. doi: 10.1373/clinchem.2009.141580. Epub 2010 Apr 8. PMID: 20378769.
48. Bañuelos-Villegas EG, Pérez-yPérez MF, Alvarez-Salas LM. Cervical Cancer, Papillomavirus, and miRNA Dysfunction. *Front Mol Biosci.* 2021 Dec 10;8:758337. doi: 10.3389/fmolb.2021.758337. PMID: 34957212; PMCID: PMC8703027.
49. Zheng ZM, Wang X. Regulation of cellular miRNA expression by human papillomaviruses. *Biochim Biophys Acta.* 2011 Nov-Dec;1809(11-12):668-77. doi: 10.1016/j.bbtagrm.2011.05.005. Epub 2011 May 17. PMID: 21616186; PMCID: PMC3175324.
50. Luttmmer R, De Strooper LM, Steenbergen RD, Berkhof J, Snijders PJ, Heideman DA, Meijer CJ. Management of high-risk HPV-positive women for detection of cervical (pre)cancer. *Expert Rev Mol Diagn.* 2016 Sep;16(9):961-74. doi: 10.1080/14737159.2016.1217157. Epub 2016 Aug 5. PMID: 27459506.
51. Wentzensen N, Schiffman M, Palmer T, Arbyn M. Triage of HPV positive women in cervical cancer screening. *J Clin Virol.* 2016 Mar;76 Suppl 1(Suppl 1):S49-S55. doi: 10.1016/j.jcv.2015.11.015. Epub 2015 Nov 28. PMID: 26643050; PMCID: PMC4789103.
52. Hanley SJB, Fujita H, Aoyama-Kikawa S, Kasamo M, Torigoe T, Matsuno Y, Noriaki S; COMPACT study group. Evaluation of partial genotyping with HPV16/18 for triage of HPV positive, cytology negative women in the COMPACT study. *J Gynecol Oncol.* 2021 Nov;32(6):e86. doi: 10.3802/jgo.2021.32.e86. PMID: 34708593; PMCID: PMC8550923.
53. Cuzick J, Clavel C, Petry KU, Meijer CJ, Hoyer H, Ratnam S, Szarewski A, Birembaut P, Kulasingam S, Sasieni P, Iftner T. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. *Int J Cancer.* 2006 Sep 1;119(5):1095-101. doi: 10.1002/ijc.21955. PMID: 16586444.
54. Dick S, Vink FJ, Heideman DAM, Lissenberg-Witte BI, Meijer CJLM, Berkhof J. Risk-stratification of HPV-positive women with low-grade cytology by FAM19A4/miR124-2 methylation and HPV genotyping. *Br J Cancer.* 2022 Feb;126(2):259-264. doi: 10.1038/s41416-021-01614-4. Epub 2021 Nov 6. PMID: 34743198; PMCID: PMC8770638.
55. Cuschieri K, Ronco G, Lorincz A, Smith L, Ogilvie G, Mirabello L, Carozzi F, Cubie H, Wentzensen N, Snijders P, Arbyn M, Monsonogo J, Franceschi S. Eurogin roadmap 2017: Triage strategies for the management of HPV-positive women in cervical screening programs. *Int J Cancer.* 2018 Aug 15;143(4):735-745. doi: 10.1002/ijc.31261. Epub 2018 Feb 8. PMID: 29341110.
56. Sharma, Garima & Dua, Pradeep & Agarwal, Subhash. (2014). A Comprehensive Review of Dysregulated miRNAs Involved in Cervical Cancer. *Current Genomics.* 15. 10.2174/1389202915666140528003249
57. Nair VB, Manasa VG, Sinto MS, Jayasree K, James FV, Kannan S. Differential Expression of MicroRNAs in Uterine Cervical Cancer and Its Implications in Carcinogenesis; An Integrative Approach. *Int J Gynecol Cancer.* 2018 Mar;28(3):553-562. doi: 10.1097/IGC.0000000000001203. PMID: 29466255.
58. Lin W, Feng M, Chen G, Zhou Z, Li J, Ye Y. Characterization of the microRNA profile in early-stage cervical squamous cell carcinoma by next-generation sequencing. *Oncol Rep.* 2017 Mar;37(3):1477-1486. doi: 10.3892/or.2017.5372. Epub 2017 Jan 16. PMID: 28098890.
59. Li Y, Wang F, Xu J, Ye F, Shen Y, Zhou J, Lu W, Wan X, Ma D, Xie X. Progressive miRNA expression profiles in cervical carcinogenesis and identification of HPV-related target genes for miR-29. *J Pathol.* 2011 Aug;224(4):484-95. doi: 10.1002/path.2873. Epub 2011 Apr 18. PMID: 21503900.
60. Zheng M, Hou L, Ma Y, Zhou L, Wang F, Cheng B, Wang W, Lu B, Liu P, Lu W, Lu Y. Exosomal let-7d-3p and miR-30d-5p as diagnostic biomarkers for non-invasive screening of cervical cancer and its precursors. *Mol Cancer.* 2019 Apr 2;18(1):76. doi: 10.1186/s12943-019-0999-x. PMID: 30940131; PMCID: PMC6446401.
61. Pardini B, De Maria D, Francavilla A, Di Gaetano C, Ronco G, Naccarati A. MicroRNAs as markers of progression in cervical cancer: a systematic review. *BMC Cancer.* 2018 Jun 27;18(1):696. doi: 10.1186/s12885-018-4590-4. PMID: 29945565; PMCID: PMC6020348.
62. Shukla V, Varghese VK, Kabekkodu SP, Mallya S, Chakrabarty S, Jayaram P, Pandey D, Banerjee S, Sharan K, Satyamoorthy K. Enumeration of deregulated miRNAs in liquid and tissue biopsies of cervical cancer. *Gynecol Oncol.* 2019 Oct;155(1):135-143. doi: 10.1016/j.ygyno.2019.08.012. Epub 2019 Aug 18. PMID: 31434614.

63. Zheng M, Hou L, Ma Y, Zhou L, Wang F, Cheng B, Wang W, Lu B, Liu P, Lu W, Lu Y. Exosomal let-7d-3p and miR-30d-5p as diagnostic biomarkers for non-invasive screening of cervical cancer and its precursors. *Mol Cancer*. 2019 Apr 2;18(1):76. doi: 10.1186/s12943-019-0999-x. PMID: 30940131; PMCID: PMC6446401.
64. Liu SS, Chan KKL, Chu DKH, Wei TN, Lau LSK, Ngu SF, Chu MMY, Tse KY, Ip PPC, Ng EKO, Cheung ANY, Ngan HYS. Oncogenic microRNA signature for early diagnosis of cervical intraepithelial neoplasia and cancer. *Mol Oncol*. 2018 Dec;12(12):2009-2022. doi: 10.1002/1878-0261.12383. Epub 2018 Sep 27. PMID: 30221475; PMCID: PMC6275249.
65. Snoek BC, Verlaet W, Babion I, et al. Genome-wide microRNA analysis of HPV-positive self-samples yields novel triage markers for early detection of cervical cancer. *International Journal of Cancer*. 2019 Jan;144(2):372-379. DOI: 10.1002/ijc.31855. PMID: 30192375; PMCID: PMC6518875.
66. González-Ramírez MI, Cardona YT, Agudelo MC, López C, Florez-Acosta JJ, Agudelo-Gamboa S, Garai J, Li L, Orozco-Castaño CA, Zabaleta J, Sánchez GI. miRNAs signature as potential biomarkers for cervical precancerous lesions in human papillomavirus positive women. *Sci Rep*. 2023 Jun 17;13(1):9822. doi: 10.1038/s41598-023-36421-9. PMID: 37330541; PMCID: PMC10276834.
67. Peng Y, Croce CM. The role of MicroRNAs in human cancer. *Signal Transduct Target Ther*. 2016 Jan 28;1:15004. doi: 10.1038/sigtrans.2015.4. PMID: 29263891; PMCID: PMC5661652.
68. Basukala O, Banks L. The Not-So-Good, the Bad and the Ugly: HPV E5, E6 and E7 Oncoproteins in the Orchestration of Carcinogenesis. *Viruses*. 2021 Sep 22;13(10):1892. doi: 10.3390/v13101892. PMID: 34696321; PMCID: PMC8541208.
69. Beaudenon, Sylvie, y Jon M Huijbregtse. 2008. «HPV E6, E6AP and cervical cancer.» *BMC biochemistry* 9 Suppl 1(Suppl 1): S4. <http://www.biomedcentral.com/1471-2091/9/S1/S4>
70. Harden ME, Prasad N, Griffiths A, Munger K. Modulation of microRNA-mRNA Target Pairs by Human Papillomavirus 16 Oncoproteins. *mBio*. 2017 Jan 3;8(1):e02170-16. doi: 10.1128/mBio.02170-16. PMID: 28049151; PMCID: PMC5210503.
71. Babion, I., Snoek, BC, Novianti, PW et al. Triage de mujeres positivas para VPH de alto riesgo en detección poblacional mediante análisis de expresión de miARN en raspados cervicales; un estudio de viabilidad. *Clin Epigenet* 10, 76 (2018). <https://doi.org/10.1186/s13148-018-0509-9>
72. Ivanov MK, Titov SE, Glushkov SA, Dzyubenko VV, Malek AV, Arkhangelskaya PA, Samsonov RB, Mikhethko AA, Bakhidze EV, Berlev IV, Kolesnikov NN. Detection of high-grade neoplasia in air-dried cervical PAP smears by a microRNA-based classifier. *Oncol Rep*. 2018 Mar;39(3):1099-1111. doi: 10.3892/or.2018.6214. Epub 2018 Jan 12. PMID: 29328473; PMCID: PMC5802032.
73. Wang X, Wang HK, Li Y, Hafner M, Banerjee NS, Tang S, Briskin D, Meyers C, Chow LT, Xie X, Tuschl T, Zheng ZM. microRNAs are biomarkers of oncogenic human papillomavirus infections. *Proc Natl Acad Sci U S A*. 2014 Mar 18;111(11):4262-7. doi: 10.1073/pnas.1401430111. Epub 2014 Mar 3. PMID: 24591631; PMCID: PMC3964092.
74. Witting SM, Snijders PJ, Verlaet W, Jaspers A, van de Wiel MA, van Wieringen WN, Meijer GA, Kenter GG, Yi Y, le Sage C, Agami R, Meijer CJ, Steenbergen RD. Altered microRNA expression associated with chromosomal changes contributes to cervical carcinogenesis. *Oncogene*. 2013 Jan 3;32(1):106-16. doi: 10.1038/onc.2012.20. Epub 2012 Feb 13. PMID: 22330141.
75. Cohen SM, Ellwein LB. Cell proliferation in carcinogenesis. *Science*. 1990 Aug 31;249(4972):1007-11. doi: 10.1126/science.2204108. PMID: 2204108.
76. Endale HT, Mariye YF, Negash HK, Hassen FS, Asrat WB, Mengstie TA, et al. MiRNA in cervical cancer: Diagnosis to therapy: Systematic review. *Heliyon* [Internet]. 2024;10(3):e24398. Available from: <https://www.sciencedirect.com/science/article/pii/S2405844024004298>
77. Lizano M, Carrillo-García A, De La Cruz-Hernández E, Castro-Muñoz LJ, Contreras-Paredes A. Promising predictive molecular biomarkers for cervical cancer (Review). *Int J Mol Med*. 2024 Jun;53(6):50. doi: 10.3892/ijmm.2024.5374. Epub 2024 Apr 12. PMID: 38606495; PMCID: PMC11090266
78. Oliveto S, Mancino M, Manfrini N, Biffo S. Role of microRNAs in translation regulation and cancer. *World J Biol Chem*. 2017 Feb 26;8(1):45-56. doi: 10.4331/wjbc.v8.i1.45. PMID: 28289518; PMCID: PMC5329714.
79. Liu B, Shyr Y, Cai J, Liu Q. Interplay between miRNAs and host genes and their role in cancer. *Brief Funct Genomics*. 2018 Jul 22;18(4):255-266. doi: 10.1093/bfpg/elz002. PMID: 30785618; PMCID: PMC6609535.
80. Doghish AS, Ali MA, Elyan SS, Elrebehy MA, Mohamed HH, Mansour RM, et al. miRNAs role in cervical cancer pathogenesis and targeted therapy: Signaling pathways interplay. *Pathol Res Pract* [Internet]. 2023;244:154386. Available from: <https://www.sciencedirect.com/science/article/pii/S0344033823000869>
81. <https://www.sciencedirect.com/science/article/pii/S0344033823000869>
82. Jie M, Feng T, Huang W, Zhang M, Feng Y, Jiang H, Wen Z. Subcellular Localization of miRNAs and Implications in Cellular Homeostasis. *Genes (Basel)*. 2021 Jun 2;12(6):856. doi: 10.3390/genes12060856. PMID: 34199614; PMCID: PMC8226975.
83. Ali Syeda Z, Langden SSS, Munkhzul C, Lee M, Song SJ. Regulatory Mechanism of MicroRNA Expression in Cancer. *Int J Mol Sci*. 2020 Mar 3;21(5):1723. doi: 10.3390/ijms21051723. PMID: 32138313; PMCID: PMC7084905.
84. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002 Nov 26;99(24):15524-9. doi: 10.1073/pnas.242606799. Epub 2002 Nov 14. PMID: 12434020; PMCID: PMC137750.

85. Budakoti M, Panwar AS, Molpa D, Singh RK, Büsselberg D, Mishra AP, Coutinho HDM, Nigam M. Micro-RNA: The darkhorse of cancer. *Cell Signal*. 2021 Jul;83:109995. doi: 10.1016/j.cellsig.2021.109995. Epub 2021 Mar 27. PMID: 33785398.
86. Rao Q, Shen Q, Zhou H, et al. Aberrant microRNA expression in human cervical carcinomas. *Medical Oncology* (Northwood, London, England). 2012 Jun;29(2):1242-1248. DOI: 10.1007/s12032-011-9830-2. PMID: 21264530.
87. Ribeiro, J. and Sousa, H. (2014) MicroRNAs as Biomarkers of Cervical Cancer Development: A Literature Review on miR-125b and miR-34a. *Molecular Biology Reports*, 41, 1525-1531. <https://doi.org/10.1007/s11033-013-2998-0>
88. Wang JY, Chen LJ. The role of miRNAs in the invasion and metastasis of cervical cancer. *Biosci Rep*. 2019 Mar 15;39(3):BSR20181377. doi: 10.1042/BSR20181377. PMID: 30833362; PMCID: PMC6418402.
89. Snoek BC, Splunter APV, Bleeker MCG, Ruiten MCV, Heideman DAM, Rurup WF, Verlaet W, Schotman H, Gent MV, Trommel NEV, Steenbergen RDM. Cervical cancer detection by DNA methylation analysis in urine. *Sci Rep*. 2019 Feb 28;9(1):3088. doi: 10.1038/s41598-019-39275-2. PMID: 30816167; PMCID: PMC6395822.
90. Hesselink AT, Heideman DA, Steenbergen RD, Gök M, van Kemenade FJ, Wilting SM, Berkhof J, Meijer CJ, Snijders PJ. Methylation marker analysis of self-sampled cervico-vaginal lavage specimens to triage high-risk HPV-positive women for colposcopy. *Int J Cancer*. 2014 Aug 15;135(4):880-6. doi: 10.1002/ijc.28723. Epub 2014 Jan 28. PMID: 24474183.
91. Doghish AS, Ali MA, Elyan SS, Elrebehy MA, Mohamed HH, Mansour RM, et al. miRNAs role in cervical cancer pathogenesis and targeted therapy: Signaling pathways interplay. *Pathol Res Pract* [Internet]. 2023;244:154386. Available from: <https://www.sciencedirect.com/science/article/pii/S0344033823000869>
92. Tornesello ML, Faraonio R, Buonaguro L, Annunziata C, Starita N, Cerasuolo A, Pezzuto F, Tornesello AL, Buonaguro FM. The Role of microRNAs, Long Non-coding RNAs, and Circular RNAs in Cervical Cancer. *Front Oncol*. 2020 Feb 20;10:150. doi: 10.3389/fonc.2020.00150. PMID: 32154165; PMCID: PMC7044410.
93. Nagandla K, Lin KH, Chitra E, Jamli MFBM. Role of microRNAs as biomarkers of cervical carcinogenesis: a systematic review. *Obstet Gynecol Sci*. 2021 Sep;64(5):419-436. doi: 10.5468/ogs.21123. Epub 2021 Aug 13. PMID: 34384196; PMCID: PMC8458608.
94. Shuna Shen, Shengbin Zhang, Ping Liu, Junfei Wang, Haiyan Du, Potential role of microRNAs in the treatment and diagnosis of cervical cancer, *Cancer Genetics*, Volumes 248–249, 2020, Pages 25-30, ISSN 2210-7762, <https://doi.org/10.1016/j.cancergen.2020.09.003>.
95. Ovando Ricárdez JA, Domínguez Méndez VB, López García MI, Sacramento Barranco ML, González Castro TB, Hernández Díaz Y, et al. MicroARN's (miARN's) como reguladores de la expresión génica y su importancia clínica. *Kuxulkab'* [Internet]. 2022 Jan 17;28(60):05–13. Available from: <https://revistas.ujat.mx/index.php/kuxulkab/article/view/4657>
96. Liu J, Li Y, Chen X, Xu X, Zhao H, Wang S, et al. Upregulation of miR-205 induces CHN1 expression, which is associated with the aggressive behaviour of cervical cancer cells and correlated with lymph node metastasis. *BMC Cancer* [Internet]. 2020;20(1):1029. Available from: <https://doi.org/10.1186/s12885-020-07478-w>
97. Moutinho C, Esteller M. MicroRNAs and Epigenetics. *Adv Cancer Res*. 2017;135:189-220. doi: 10.1016/bs.acr.2017.06.003. Epub 2017 Aug 4. PMID: 28882223.
98. Mitra T, Elangovan S. Cervical cancer development, chemoresistance, and therapy: a snapshot of involvement of microRNA. *Mol Cell Biochem* [Internet]. 2021;476(12):4363–85. Available from: <https://doi.org/10.1007/s11010-021-04249-4>
99. De Martel C, Georges D, Bray F, Ferlay J, Clifford GM, Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis, *Lancet Glob Health*, Published online 17 December 2019; [https://doi.org/10.1016/S2214-109X\(19\)30488-7](https://doi.org/10.1016/S2214-109X(19)30488-7)
100. De Martel C, Georges D, Bray F, Ferlay J, Clifford GM, Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis, *Lancet Glob Health*, Published online 17 December 2019; [https://doi.org/10.1016/S2214-109X\(19\)30488-7](https://doi.org/10.1016/S2214-109X(19)30488-7)
101. Hareża DA, Wilczyński JR, Paradowska E. Human Papillomaviruses as Infectious Agents in Gynecological Cancers. *Oncogenic Properties of Viral Proteins*. *Int J Mol Sci*. 2022 Feb 5;23(3):1818. doi: 10.3390/ijms23031818. PMID: 35163748; PMCID: PMC8836588.
102. Kombe Kombe AJ, Li B, Zahid A, Mengist HM, Bounda G-A, Zhou Y and Jin T (2021) Epidemiology and Burden of Human Papillomavirus and Related Diseases, *Molecular Pathogenesis, and Vaccine Evaluation*. *Front. Public Health* 8:552028. doi: 10.3389/fpubh.2020.552028
103. Bhatla N, Aoki D, Sharma DN, Sankaranarayanan R. Cancer of the cervix uteri: 2021 update. *International Journal of Gynecology & Obstetrics* [Internet]. 2021 Oct 1;155(S1):28–44. Available from: <https://doi.org/10.1002/ijgo.13865>
104. McBride AA. Human papillomaviruses: diversity, infection and host interactions. *Nat Rev Microbiol* [Internet]. 2022;20(2):95–108. Available from: <https://doi.org/10.1038/s41579-021-00617-5>
105. Van Doorslaer K. Evolution of the papillomaviridae. *Virology*. 2013 Oct;445(1-2):11-20. doi: 10.1016/j.virol.2013.05.012. Epub 2013 Jun 14. PMID: 23769415.
106. Doorbar John, Wim Quint, Lawrence Banks, Ignacio G. Bravo, Mark Stoler, Tom R. Broker, Margaret A. Stanley, *The Biology and Life-Cycle of Human Papillomaviruses*, *Vaccine*, Volume 30, Supplement 5, 2012, Pages F55-F70, ISSN 0264-410X, <https://doi.org/10.1016/j.vaccine.2012.06.083>. (<https://www.sciencedirect.com/science/article/pii/S0264410X12009735>)
107. Wang X, Huang X, Zhang Y. Involvement of Human Papillomaviruses in Cervical Cancer. *Front Microbiol*. 2018 Nov 28;9:2896. doi: 10.3389/fmicb.2018.02896. PMID: 30546351; PMCID: PMC6279876.
108. Nelson CW, Mirabello L. Human papillomavirus genomics: Understanding carcinogenicity. *Tumour Virus Res*. 2023 Jun;15:200258. doi: 10.1016/j.tvr.2023.200258. Epub 2023 Feb 20. PMID: 36812987; PMCID: PMC10063409.

109. Frederique J. Vink, Stéfanie Dick, Daniëlle A. M. Heideman, Lise M. A. De Strooper, Renske D. M. Steenbergen, Birgit I. Lissenberg-Witte, DNTP Group, Arno Floore et al. (2021). Classification of high-grade cervical intraepithelial neoplasia by p16ink4a, Ki-67, HPV E4 and FAM19A4/miR124-2 methylation status demonstrates considerable heterogeneity with potential consequences for management. *International Journal of Cancer*. <https://doi.org/10.1002/ijc.33566>
110. Narisawa-Saito M, Kiyono T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins. *Cancer Sci*. 2007 Oct;98(10):1505-11. doi: 10.1111/j.1349-7006.2007.00546.x. Epub 2007 Jul 23. PMID: 17645777; PMCID: PMC11158331.
111. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, Cousido-Siah A, Masson M, Vande Pol S, Podjarny A, Travé G, Zanier K. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature*. 2016 Jan 28;529(7587):541-5. doi: 10.1038/nature16481. Epub 2016 Jan 20. PMID: 26789255; PMCID: PMC4853763.
112. Organización Mundial de la Salud. (2014). *Comprehensive cervical cancer control: A guide to essential practice* (2nd ed.). World Health Organization. https://iris.who.int/bitstream/handle/10665/144785/9789241548953_eng.pdf?sequence=1
113. Brianti P, De Flammineis E, Mercuri SR. Review of HPV-related diseases and cancers. *New Microbiol*. 2017 Apr;40(2):80-85. Epub 2017 Apr 3. PMID: 28368072.
114. Jaisamrarn U, Castellsagué X, Garland SM, Naud P, Palmroth J, Del Rosario-Raymundo MR, et al. (2013) Correction: Natural History of Progression of HPV Infection to Cervical Lesion or Clearance: Analysis of the Control Arm of the Large, Randomised PATRICIA Study. *PLoS ONE* 8(12): 10.1371/annotation/cea59317-929c-464a-b3f7-e095248f229a. <https://doi.org/10.1371/annotation/cea59317-929c-464a-b3f7-e095248f229a>
115. Basu P, Taghavi K, Hu SY, Mogri S, Joshi S. Management of cervical premalignant lesions. *Curr Probl Cancer*. 2018 Mar-Apr;42(2):129-136. doi: 10.1016/j.cuprob.cancer.2018.01.010. Epub 2018 Jan 11. PMID: 29428790.
116. Sudenga SL, Shrestha S. Key considerations and current perspectives of epidemiological studies on human papillomavirus persistence, the intermediate phenotype to cervical cancer. *Int J Infect Dis*. 2013 Apr;17(4):e216-20. doi: 10.1016/j.ijid.2012.12.027. Epub 2013 Feb 27. PMID: 23453716; PMCID: PMC3602330.
117. Wood, M. A., Kerrigan, K. L., Burns, M. K., Glenn, T. L., Ludwin, A., Christianson, M. S., Bhagavath, B., & Lindheim, S. R. (2018). Overcoming the Challenging Cervix: Identification and Techniques to Access the Uterine Cavity. *Obstetrical and Gynecological Survey*, 73(11), 641–649. <https://doi.org/10.1097/OGX.0000000000000614>
118. Guida F, Kidman R, Ferlay J, Schüz J, Soerjomataram I, Kithaka B, et al. Global and regional estimates of orphans attributed to maternal cancer mortality in 2020. *Nat Med [Internet]*. 2022;28(12):2563–72. Available from: <https://doi.org/10.1038/s41591-022-02109-2>
119. Stolnicu S, Hoang L, Soslow RA. Recent advances in invasive adenocarcinoma of the cervix. *Virchows Arch*. 2019 Nov;475(5):537-549. doi: 10.1007/s00428-019-02601-0. Epub 2019 Jun 17. PMID: 31209635; PMCID: PMC6864265.
120. Stelzle D, Tanaka LF, Lee KK, Ibrahim Khalil A, Baussano I, Shah AS V, et al. Estimates of the global burden of cervical cancer associated with HIV. *Lancet Glob Health [Internet]*. 2021 Feb 1;9(2):e161–9. Available from: [https://doi.org/10.1016/S2214-109X\(20\)30459-9](https://doi.org/10.1016/S2214-109X(20)30459-9)
121. Orumaa M, Kjaer SK, Dehlendorff C, Munk C, Olsen AO, Hansen BT, Campbell S, Nygård M. The impact of HPV multi-cohort vaccination: Real-world evidence of faster control of HPV-related morbidity. *Vaccine*. 2020 Feb 5;38(6):1345-1351. doi: 10.1016/j.vaccine.2019.12.016. Epub 2020 Jan 6. PMID: 31917039.
122. Wang Z, Yao H, Lin S, Zhu X, Shen Z, Lu G, Poon WS, Xie D, Lin MC, Kung HF. Transcriptional and epigenetic regulation of human microRNAs. *Cancer Lett*. 2013 Apr 30;331(1):1-10. doi: 10.1016/j.canlet.2012.12.006. Epub 2012 Dec 11. PMID: 23246373.
123. Test estadísticos para variables cualitativas: test exacto de Fisher, chi-cuadrado de Pearson, McNemar y Q-Cochran by Joaquín Amat Rodrigo, available under a Attribution 4.0 International (CC BY 4.0) at https://www.cienciadedatos.net/documentos/22.2_test_exacto_de_fisher_chi-cuadrado_de_pearson_mcnemar_qcochran.html
124. Andrews S. (2010). FastQC: una herramienta de control de calidad para datos de secuencia de alto rendimiento. Disponible en línea en: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
125. MARTIN, Marcel. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, [S.l.], v. 17, n. 1, p. pp. 10-12, may 2011. ISSN 2226-6089. Available at: <<https://journal.embnet.org/index.php/embnetjournal/article/view/200>>. Date accessed: 08 july 2024. doi:<https://doi.org/10.14806/ej.17.1.200>.
126. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012 Jan;40(1):37-52. doi: 10.1093/nar/gkr688. Epub 2011 Sep 12. PMID: 21911355; PMCID: PMC3245920.
127. R Core Team (2023). *_R: A Language and Environment for Statistical Computing_*. R. Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>
128. Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>
129. Wu J. *Advances in K-means clustering: a data mining thinking*. Springer Science & Business Media; 2012.
130. Kassambara, A. and Mundt, F. (2020) *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R Package Version 1.0.7. <https://CRAN.R-project.org/package=factoextra>

131. Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 *Genome Biology* 15(12):550 (2014)
132. Yuanbin Ru, Matt Mulvahill, Spencer Mahaffey and Katerina Kechris (2017). multiMiR: Integration of multiple microRNA-target databases with their disease and drug associations. R package version 0.98.0.2. <https://github.com/KechrisLab/multiMiR>
133. T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo, and G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*. 2021, 2(3):100141
134. Ludwig Geistlinger, Gergely Csaba and Ralf Zimmer (2016) Bioconductor's. EnrichmentBrowser: seamless navigation through combined results of set- & network-based enrichment analysis. *BMC Bioinformatics*, 17:45. doi: 10.1186/s12859-016-0884-1
135. Hernández-Aguado JJ, de La Fuente-Valero J, Ramírez Mena M, Ortega-Medina L, Vidart Aragón, JA, Galán JC. Comparative pilot study about HPV test with partial genotyping in primary screening versus other strategies for cervical cancer population screening, CRYGEN 16/18 study. *EnfermInfecc Microbiol Clin (Engl Ed)*. 2023 May;41(5):262-268. doi: 10.1016/j.eimce.2022.08.001. Epub 2022 Aug 10. PMID: 35963773.
136. Shukla, Vaibhav & Varghese, Vinay & Kabekkodu, Shama & Mallya, Sandeep & Chakrabarty, Sanjiban & Jayaram, Pradyumna & Pandey, Deeksha & Banerjee, Sourjya & Sharan, Krishna & Satyamoorthy, Kapaettu. (2019). Enumeration of deregulated miRNAs in liquid and tissue biopsies of cervical cancer. *Gynecologic Oncology*. 155. 10.1016/j.ygyno.2019.08.012.
137. Yeri A, Courtright A, Danielson K, Hutchins E, Alsop E, Carlson E, Hsieh M, Ziegler O, Das A, Shah RV, Rozowsky J, Das S, Van Keuren-Jensen K. Evaluation of commercially available small RNASeq library preparation kits using low input RNA. *BMC Genomics*. 2018 May 5;19(1):331. doi: 10.1186/s12864-018-4726-6. PMID: 29728066; PMCID: PMC5936030.
138. Wang, Jingjing MSA,b; Li, Menghuan MSA,c; Zhao, Lixian MDa; Zhou, Bingjie MSA,d; Chen, Huaqiu MSA,e; Duan, Fuhui MDa; Wang, Guangming PhDa,* . Whole genome sequencing in high-grade cervical intraepithelial neoplasia patients from different ethnic groups in China. *Medicine* 102(45):p e35953, November 10, 2023. | DOI: 10.1097/MD.00000000000035953
139. Chen, M., Wang, H., Liang, Y. et al. Establishment of multifactor predictive models for the occurrence and progression of cervical intraepithelial neoplasia. *BMC Cancer* 20, 926 (2020). <https://doi.org/10.1186/s12885-020-07265-7>
140. Zhang L, Yu H, Deng T, Ling L, Wen J, Lv M, Ou R, Wang Q, Xu Y. FNDC3B and BPGM Are Involved in Human Papillomavirus-Mediated Carcinogenesis of Cervical Cancer. *Front Oncol*. 2021 Dec 16;11:783868. doi: 10.3389/fonc.2021.783868. PMID: 34976823; PMCID: PMC8716600.
141. Lopez JP, Diallo A, Cruceanu C, Fiori LM, Laboissiere S, Guillet I, Fontaine J, Ragoussis J, Benes V, Turecki G, Ernst C. Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med Genomics*. 2015 Jul 1;8:35. doi: 10.1186/s12920-015-0109-x. PMID: 26130076; PMCID: PMC4487992.
142. Babion I, Snoek BC, Novianti PW, Jaspers A, van Trommel N, Heideman DAM, Meijer CJLM, Snijders PJF, Steenbergen RDM, Wilting SM. Triage of high-risk HPV-positive women in population-based screening by miRNA expression analysis in cervical scrapes; a feasibility study. *Clin Epigenetics*. 2018 Jun 7;10:76. doi: 10.1186/s13148-018-0509-9. PMID: 29930741; PMCID: PMC5992707.
143. Calin GA, Croce CM. Firmas de microARN en cánceres humanos. *Nat Rev Cancer*. 2006; 6 :857–866.
144. Li S, Liu J, Zhao M, Su Y, Cong B, Wang Z. RNA quality score evaluation: A preliminary study of RNA integrity number (RIN) and RNA integrity and quality number (RNA IQ). *Forensic Sci Int*. 2024 Apr;357:111976. doi: 10.1016/j.forsciint.2024.111976. Epub 2024 Mar 2. PMID: 38447345
145. Ramirez I. MicroRNAs celulares reguladores de proteínas blanco de la acción de e6 y e7 del vph como potenciales biomarcadores de lesiones preneoplásicas de alto grado de cervix [Tesis de maestría]. Universidad de Antioquia; 2020.
146. Pereira PM, Marques JP, Soares AR, Carreto L, Santos MA. MicroRNA expression variability in human cervical tissues. *PLoS One*. 2010 Jul 26;5(7):e11780. doi: 10.1371/journal.pone.0011780. PMID: 20668671; PMCID: PMC2909898.
147. Maleki, F., Ovens, K., McQuillan, I. et al. El tamaño importa: cómo el tamaño de la muestra afecta la reproducibilidad y especificidad del análisis de conjuntos de genes. *Hum Genomics* 13 (Supl. 1), 42 (2019). <https://doi.org/10.1186/s40246-019-0226-2>
148. Leal-Calvo T, Moraes MO. Reanalysis and integration of public microarray datasets reveals novel host genes modulated in leprosy. *Mol Genet Genomics*. 2020 Nov;295(6):1355-1368. doi: 10.1007/s00438-020-01705-6. Epub 2020 Jul 13. PMID: 32661593.
149. Wu FX. Genetic weighted k-means algorithm for clustering large-scale gene expression data. *BMC Bioinformatics*. 2008 May 28;9 Suppl 6(Suppl 6):S12. doi: 10.1186/1471-2105-9-S6-S12. PMID: 18541047; PMCID: PMC2423435.
150. Wilkin GA, Huang X. A practical comparison of two K-Means clustering algorithms. *BMC Bioinformatics*. 2008 May 28;9 Suppl 6(Suppl 6):S19. doi: 10.1186/1471-2105-9-S6-S19. PMID: 18541054; PMCID: PMC2423442.
151. Lu Y, Lu S, Fotouhi F, Deng Y, Brown SJ. Incremental genetic K-means algorithm and its application in gene expression data analysis. *BMC Bioinformatics*. 2004 Oct 28;5:172. doi: 10.1186/1471-2105-5-172. PMID: 15511294; PMCID: PMC543472.
152. Dubey, A. K., Gupta, U., & Jain, S. (2016). Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. *International Journal of Computer Assisted Radiology and Surgery*, 11(11), 2033–2047. doi:10.1007/s11548-016-1437-9

153. Martinez I, Gardiner AS, Board KF, Monzon FA, Edwards RP, Khan SA. Human papillomavirus type 16 reduces the expression of microRNA-218 in cervical carcinoma cells. *Oncogene*. 2008 Apr 17;27(18):2575-82. doi: 10.1038/sj.onc.1210919. Epub 2007 Nov 12. PMID: 17998940; PMCID: PMC2447163.
154. Han MS, Lee JM, Kim SN, Kim JH, Kim HS. Human Papillomavirus 16 Oncoproteins Downregulate the Expression of miR-148a-3p, miR-190a-5p, and miR-199b-5p in Cervical Cancer. *Biomed Res Int*. 2018 Nov 29;2018:1942867. doi: 10.1155/2018/1942867. PMID: 30627542; PMCID: PMC6304571.
155. De Brot L, Pellegrini B, Moretti ST, Carraro DM, Soares FA, Rocha RM, Baiocchi G, da Cunha IW, de Andrade VP. Infections with multiple high-risk HPV types are associated with high-grade and persistent low-grade intraepithelial lesions of the cervix. *Cancer Cytopathol*. 2017 Feb;125(2):138-143. doi: 10.1002/cncy.21789. Epub 2016 Nov 21. PMID: 27870295
156. De Brot L, Pellegrini B, Moretti ST, Carraro DM, Soares FA, Rocha RM, Baiocchi G, da Cunha IW, de Andrade VP. Infections with multiple high-risk HPV types are associated with high-grade and persistent low-grade intraepithelial lesions of the cervix. *Cancer Cytopathol*. 2017 Feb;125(2):138-143. doi: 10.1002/cncy.21789. Epub 2016 Nov 21. PMID: 27870295
157. ACCP. 2011. Recent Evidence on Cervical Cancer Screening in Low-Resource Settings Cervical cancer incidence and mortality update. http://screening.iarc.fr/doc/ACCP_cxca_screening_2011.pdf
158. Zuchna C, Hager M, Tringler B, Georgouloupoulos A, Ciresa-Koenig A, Volgger B, Widschwendter A, Staudach A. Precisión diagnóstica de las biopsias cervicales guiadas: un estudio multicéntrico prospectivo que compara la histopatología de la biopsia simultánea y la muestra de cono. *Am J Obstet Gynecol*. 2010; 203(4): 321.e1-6.
159. Herrington CS. The terminology of pre-invasive cervical lesions in the UK cervical screening programme. *Cytopathology*. 2015 Dec;26(6):346-50. doi: 10.1111/cyt.12307. PMID: 26767601.
160. Cheung, T., Man, K. M., Yu, M., Yim, S., Siu, N. S. S., Lo, K. W. K., ... Wong, Y. (2012). Dysregulated microRNAs in the pathogenesis and progression of cervical neoplasm. *Cell Cycle*, 11(15), 2876-2884. doi:10.4161/cc.21278 10.4161/cc.21278
161. Li XY, Li G, Gong TT, Lv JL, Gao C, Liu FH, Zhao YH, Wu QJ. Non-Genetic Factors and Risk of Cervical Cancer: An Umbrella Review of Systematic Reviews and Meta-Analyses of Observational Studies. *Int J Public Health*. 2023 Mar 31;68:1605198. doi: 10.3389/ijph.2023.1605198. PMID: 37065642; PMCID: PMC10103589
162. Wu X, Peng L, Zhang Y, Chen S, Lei Q, Li G, Zhang C. Identification of Key Genes and Pathways in Cervical Cancer by Bioinformatics Analysis. *Int J Med Sci*. 2019 Jun 2;16(6):800-812. doi: 10.7150/ijms.34172. PMID: 31337953; PMCID: PMC6643108.
163. Wu B, Xi S. Bioinformatics analysis of differentially expressed genes and pathways in the development of cervical cancer. *BMC Cancer*. 2021 Jun 26;21(1):733. doi: 10.1186/s12885-021-08412-4. PMID: 34174849; PMCID: PMC8236200.
164. Wu K, Yi Y, Liu F, Wu W, Chen Y, Zhang W. Identification of key pathways and genes in the progression of cervical cancer using bioinformatics analysis. *Oncol Lett*. 2018 Jul;16(1):1003-1009. doi: 10.3892/ol.2018.8768. Epub 2018 May 22. PMID: 29963176; PMCID: PMC6019941.
165. Nguyen TB, Do DN, Nguyen-Thi ML, Hoang-The H, Tran TT, Nguyen-Thanh T. Identification of potential crucial genes and key pathways shared in Inflammatory Bowel Disease and cervical cancer by machine learning and integrated bioinformatics. *Comput Biol Med*. 2022 Oct;149:105996. doi: 10.1016/j.combiomed.2022.105996. Epub 2022 Aug 27. PMID: 36049413.
166. Hoppe-Seyler K, Mändl J, Adrian S, Kuhn BJ, Hoppe-Seyler F. Virus/Host Cell Crosstalk in Hypoxic HPV-Positive Cancer Cells. *Viruses*. 2017 Jul 5;9(7):174. doi: 10.3390/v9070174. PMID: 28678198; PMCID: PMC5537666.
167. Bossler F, Hoppe-Seyler K, Hoppe-Seyler F. PI3K/AKT/mTOR Signaling Regulates the Virus/Host Cell Crosstalk in HPV-Positive Cervical Cancer Cells. *Int J Mol Sci*. 2019 May 3;20(9):2188. doi: 10.3390/ijms20092188. PMID: 31058807; PMCID: PMC6539191.
168. Cruz-Gregorio A, Aranda-Rivera AK, Pedraza-Chaverri J. Human Papillomavirus-related Cancers and Mitochondria. *Virus Res*. 2020 Sep;286:198016. doi: 10.1016/j.virusres.2020.198016. Epub 2020 May 20. PMID: 32445871.
169. Babion I, Snoek BC, van de Wiel MA, Wiltng SM, Steenberg RDM. A Strategy to Find Suitable Reference Genes for miRNA Quantitative PCR Analysis and Its Application to Cervical Specimens. *J Mol Diagn*. 2017 Sep;19(5):625-637. doi: 10.1016/j.jmoldx.2017.04.010. PMID: 28826607.
170. Sanches JGP, Xu Y, Yabasin IB, Li M, Lu Y, Xiu X, Wang L, Mao L, Shen J, Wang B, Hou L, Ju J, Zhao J, Song B. miR-501 is upregulated in cervical cancer and promotes cell proliferation, migration and invasion by targeting CYLD. *Chem Biol Interact*. 2018 Apr 1;285:85-95. doi: 10.1016/j.cbi.2018.02.024. Epub 2018 Feb 23. PMID: 29477382
171. Wang JY, Chen LJ. The role of miRNAs in the invasion and metastasis of cervical cancer. *Biosci Rep*. 2019 Mar 15;39(3):BSR20181377. doi: 10.1042/BSR20181377. PMID: 30833362; PMCID: PMC6418402.
172. Wang S, Zhang S. Systematic analyses of a novel circRNA-related miRNAs prognostic signature for Cervical Cancer. *Genet Mol Biol*. 2022 Jun 24;45(2):e20210405. doi: 10.1590/1678-4685-GMB-2021-0405. PMID: 35766420; PMCID: PMC9241030.
173. Babion I, Jaspers A, van Splunter AP, van der Hoorn IAE, Wiltng SM, Steenberg RDM. miR-9-5p Exerts a Dual Role in Cervical Cancer and Targets Transcription Factor TWIST1. *Cells*. 2019 Dec 26;9(1):65. doi: 10.3390/cells9010065. PMID: 31888045; PMCID: PMC7017350.
174. Zhang H, Zhang Z, Wang S, Zhang S, Bi J. The mechanisms involved in miR-9 regulated apoptosis in cervical cancer by targeting FOXO3. *Biomed Pharmacother*. 2018 Jun;102:626-632. doi: 10.1016/j.biopha.2018.03.019. Epub 2018 Apr 5. PMID: 29602130.

175. Liu W, Gao G, Hu X, Wang Y, Schwarz JK, Chen JJ, Grigsby PW, Wang X. Activation of miR-9 by human papillomavirus in cervical cancer. *Oncotarget*. 2014 Nov 30;5(22):11620-30. doi: 10.18632/oncotarget.2599. PMID: 25344913; PMCID: PMC4294330.
176. Zhou K, Cai C, Zou M, He Y, Duan S. Molecular Mechanisms of miR-1271 Dysregulation in Human Cancer. *DNA Cell Biol*. 2021 Jun;40(6):740-747. doi: 10.1089/dna.2021.0100. Epub 2021 May 20. PMID: 34015233