

**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 CÉRVIX**

Martha Isabel González Ramírez

Trabajo presentado como requisito para optar al título de
Magíster en Ciencias Básicas Biomédicas
Énfasis Virología

Tutora

Gloria Inés Sánchez Vásquez, MSc, PhD

Comité Asesor

Olga María Bermúdez Muñoz, MSc, PhD
Luis Fernando Barrera Robledo, MSc, PhD
Silvio Urcuqui Inchimá, MSc, PhD

Corporación Ciencias Básicas Biomédicas
Grupo Infección y Cáncer
Universidad de Antioquia
Medellín
2020

Tabla de contenido

Lista de Figuras	4
Lista de Tablas.....	4
Material Suplementario	4
Abreviaturas.....	5
NOTA AL LECTOR	5
Summary.....	6
Resumen	7
Agradecimientos	8
1. Introducción	9
2. Antecedentes	12
<i>Papel de los miRNAs en Procesos Oncológicos.....</i>	12
<i>Aproximación in-silico: interacción miRNAs y RNAm dianas</i>	14
<i>Relación miRNAs y Cáncer Cervical.....</i>	16
<i>miRNAs como biomarcadores</i>	17
3. Marco Conceptual.....	20
<i>Virus del Papiloma Humano VPH.....</i>	21
<i>Oconproteínas E6 y E7 del VPH-AR.....</i>	22
<i>microRNAs.....</i>	25
<i>Biogénesis de los miRNAs</i>	26
<i>Regulación de la expresión génica por miRNAs</i>	27
4. Planteamiento del problema.....	28
5. Justificación	29
6. Pregunta de Investigación e Hipótesis	30
7. Objetivos	31
<i>Objetivo general</i>	31
<i>Objetivos específicos</i>	31
8. Materiales y Métodos.....	32
<i>Predicción in-silico.....</i>	32
<i>Descripción de la Población</i>	34

<i>Obtención de RNA a partir de TFFIP</i>	38
<i>Preparación de las muestras TFFIP</i>	39
<i>Extracción de RNA total</i>	39
<i>Secuenciación de miRNAs – smallRNAseq</i>	40
<i>Obtención del cDNA y qRT-PCR</i>	42
<i>Manejo de datos de secuenciación</i>	44
<i>Análisis de expresión diferencial</i>	44
<i>Análisis de curvas ROC</i>	47
9. Resultados	48
<i>Identificación de miRNAs por la estrategia de búsqueda in-silico</i>	48
<i>Alineamiento de miRNAs y pre-procesamiento de los datos de secuenciación</i>	50
<i>Análisis de expresión diferencial de los datos de secuenciación</i>	53
<i>Selección de miRNAs para validación</i>	60
<i>Niveles de expresión relativa de miR-1, miR-2, miR-3, miR-4 y miR-5 por qRT-PCR</i>	60
<i>Exploración de la utilidad clínica de miRNAs</i>	64
10. Discusiones	68
<i>Normalización de datos de qRT-PCR</i>	68
<i>Variabilidad de la expresión de miRNAs y correlación entre técnicas</i>	69
<i>Identificación de miRNAs in-silico</i>	70
<i>Validación en tejidos de miRNAs encontrados in-silico y por secuenciación</i>	71
11. Conclusiones	77
12. Recomendaciones.....	79
13. Perspectivas.....	80
Bibliografía.....	82

Lista de Figuras

- Figura 1. Papel de los miRNAs en Tumorigénesis
 Figura 2. Potenciales biomarcadores para cáncer cervical
 Figura 3. Diagrama del genoma del VPH tipo 16
 Figura 4. Esquema comparativo de la regulación del ciclo celular
 Figura 5. Biogénesis de los miRNAs
 Figura 6. Esquema de la metodología para el cumplimiento de los objetivos
 Figura 7. Preparación de la biblioteca de secuenciación de miRNA
 Figura 8. Representación de las secuencias de las librerías a ser secuenciadas
 Figura 9. El sistema miRCURY LNA miRNA PCR
 Figura 10. El Nucleótido LNA
 Figura 11. Estrategia *in-silico* para la selección de miRNAs
 Figura 12. Número y porcentaje de los miRNA diferencialmente expresados por secuenciación
 Figura 13. Número y porcentaje de los miRNAs diferencialmente expresados por secuenciación EN COMÚN con los encontrados *in-silico*
 Figura 14. Variabilidad entre los conteos de lecturas normalizados para cada miRNA (n=25)
 Figura 15. Expresión relativa Log2FC (FC=2^{-ΔΔCT}) de los cinco miRNAs en cada grupo de lesión.
 Figura 16. Expresión relativa Log2FC (FC=2^{-ΔΔCT}) de los cinco miRNAs, en cada estadio de la enfermedad cervical NEG, NIC1 y NIC3+, sin incluir las muestras del grupo NIC2
 Figura 17. Expresión relativa Log2FC (FC=2^{-ΔΔCT}) de los cinco miRNAs entre los grupos de lesiones de alto grado vs. de bajo grado
 Figura 18. Expresión relativa Log2FC (FC=2^{-ΔΔCT}) de los cinco miRNAs entre los grupos de lesiones de alto grado vs. de bajo grado. Solo con las muestras del grupo de validación
 Figura 19. Curvas ROC para cada miRNA y combinaciones para la detección de NIC2+
 Figura 20. Curvas ROC para cada miRNA y combinaciones para la detección de NIC3+
 Figura 21. Red de interacciones predichas en miRWalk2.0 de miRNAs encontrados diferencialmente expresados por secuenciación y sus respectivos RNA mensajeros correspondientes a dianas E6/E7 VPH-AR
 Figura 22. Cuantificación de miRNAs extraídos de exfoliados cervicales

Lista de Tablas

- Tabla 1. Proteínas blanco de acción de las E6 y E7 del VPH-AR
 Tabla 2. Descripción de la población de estudio
 Tabla 3. Número de muestras en cada conjunto de experimentos
 Tabla 4. Resumen del alineamiento de las lecturas secuenciadas
 Tabla 5. Lista de miRNAs diferencialmente expresados encontrados en común entre la aproximación *in-silico* y el análisis de los datos de secuenciación en los programas R y DAC
 Tabla 6. Top 25 de los miRNA diferencialmente expresados por secuenciación
 Tabla 7. Selección de cinco miRNAs para validación por qRT-PCR
 Tabla 8. Valores de AUC y valor p de cada miRNAs y sus combinaciones propuestas para la detección de lesiones de alto grado de cérvix (NIC2+ y NIC3+) en mujeres VPH-AR+
 Tabla 9. Porcentaje de lesión y valores del CT del gen SNORD44 en las muestras del grupo de descubrimiento

Material Suplementario

- Suplemento 1. Red de interacciones entre los 454 miRNAs candidatos
 Suplemento 2. Variabilidad entre los conteos de lecturas normalizados para cada miRNA (n=2186)
 Suplemento 3. Análisis de componentes principales
 Suplemento 4. Variabilidad entre los conteos de lecturas normalizados para cada miRNA (n=162)
 Suplemento 5. Lista de 162 miRNA diferencialmente expresados en el programa DAC
 Suplemento 6. Correlación de los valores de FC y Log2FC de los cinco miRNAs entre la secuenciación y qRT-PCR

Abreviaturas

AUC: Area under the curve: Área bajo la curva

CEC: Carcinoma Escamocelular

DNA: Deoxyribonucleic Acid (ADN: Ácido Desoxirribonucleico)

DNAC: DNA complementario/copia

FFPE: Formalin-Fixed Paraffin-Embedded

FC: Fold Change (Veces de Cambio)

hrHPV: high-risk HPV

HPV: Human papillomavirus

miRNA: microRNAs

Nt: nucleótidos

NEG: Negativo para Lesión

NIC 1: NIC I: Neoplasia Intraepitelial de grado 1

NIC 2: NIC II: Neoplasia Intraepitelial de grado 2

NIC 3: NIC III: Neoplasia Intraepitelial de grado 3

NIC 2+: NIC II+: Neoplasia Intraepitelial de grado 2 o peor (NIC3- CEC)

Pb: Pares de bases

qRT-PCR: Reacción en cadena de la polimerasa cuantitativa acoplada a transcriptasa reversa

RNA: Ribonucleic acid (ARN: Ácido Ribonucleico)

RNAm: RNA mensajero

ROC: Receiver operating characteristic curve

TFFIP: Tejido Fijado en formalina e incluidos en parafina

VPH-AR: Virus del Papiloma Humano de Alto Riesgo

VPH-AR+: Positivo para infección con el Virus del Papiloma Humano de Alto Riesgo

NOTA AL LECTOR

El presente trabajo de investigación está en proceso de Protección de Propiedad Intelectual con la Unidad de Transferencia de Conocimiento, Dirección de Innovación de la Universidad de Antioquia. Por tanto, la nomenclatura de los miRNAs encontrados en esta investigación se encuentra modificada, se usaron números del 1 al 162 (miR-1, miR-2,, miR-162) para referirse a los miRNAs productos de protección. Los demás miRNAs y moléculas de trabajos citados en este manuscrito no fueron modificados.

Summary

Cervical cancer ranks first in cancer mortality among women of low-middle income countries where 80% of the 570,000 cases and 311,000 worldwide deaths estimated for 2018 occurred. Persistent infections with high-risk HPV (hrHPV) genotypes can lead to cervical high-grade lesions (Cervical Intraepithelial Neoplasia Grade 2 or more severe disease (CIN II+), that if left untreated progress to cancer. hrHPV test has high sensitivity but because many women infected with hrHPV genotypes will clear the infection spontaneously, it has low specificity to detect CIN II+. MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression and show differential profiles in cervical intraepithelial lesions.

Aim. Identify miRNAs that regulated target proteins of E6/E7 of hrHPV, differentially expressed in high- versus low-grade lesions from hrHPV infected women.

Methods. miRNAs were searched *in-silico* using the strategy based on hrHPV- E6/E7 target proteins. We compared the miRNAs expression pattern between FFPE tissues from hrHPV infected women with low-grade lesions (n=10) and with high-grade lesions (n=10), using smallRNAseq, and five miRNAs were further validated by qRT-PCR (miR-1, 2, 3, 4 and 5) in an independent set of tissues FFPE (n=50). ROC analyses were used to calculate the AUC of the five miRNA and their combinations for NIC3 detection.

Results. We identified 454 miRNAs regulated by some of the 19 target proteins of E6/E7. The smallRNAseq analysis identified 162 miRNAs differential expressed in high- versus low-grade lesions. Of which 47 were predicted *in-silico*. The relative expression by qRT-PCR of five miRNAs was higher in high-grade versus low-grade ($p < 0,05$), and the levels of miR-2, 3 and 4 increased with disease progression (from healthy, CIN I, CIN II, CIN III+). The highest FC were obtained by miR-1, 2 and 3, of 19.3, 10 y 9.4 respectively, in the sequencing, and of 16.8, 6.7 y 6.7 respectively, by qRT-PCR, in high- versus low-grade lesions ($p < 005$). The Pearson's correlation coefficient, R, between NGS analysis and qRT-PCR results was 0.86. ROC analysis exploration showed that miR-3 and its combination with miR-1 exhibit the highest AUC of 79 % (95% CI: 65-93). In conclusion, we identified 162 miRNAs by sequencing, of which 47 regulate some of the 19 target proteins of E6/E7, and we validated five by qRT-PCR. These results show that miRNAs that regulate target proteins of E6/E7; are expressed differently in high- versus low-grade lesions. Further validation on a larger cohort of samples is needed to confirm the potential role of these miRNAs to triage hrHPV infected women.

Resumen

El cáncer cervical ocupa el primer lugar en mortalidad por cáncer en mujeres de los países de bajos y medianos ingresos, donde se produjo el 80% de los 570,000 casos y 311,000 muertes en el mundo, estimadas para 2018. La infección persistente con genotipos de alto riesgo del virus del papiloma humano (VPH-AR) puede conducir a lesiones de alto grado (NIC-2+), que si no se tratan progresan a cáncer. La prueba de VPH-AR tiene alta sensibilidad, pero dado que muchas mujeres infectadas eliminarán la infección espontáneamente, tiene baja especificidad para detectar NIC-2+. Los miRNA regulan la expresión génica a nivel postranscripcional y han mostrado perfiles diferenciales en el espectro de la enfermedad. Objetivo. Identificar miRNAs reguladores de proteínas diana de E6/E7 del VPH-AR, expresados diferencialmente entre lesiones de alto grado y lesiones de bajo grado del cérvix, en mujeres VPH-AR+. Métodos. Se buscaron miRNAs *in-silico* con la estrategia basada en proteínas dianas de E6/E7 del VPH-AR. Se comparó el patrón de expresión de miRNAs entre tejidos fijado en formalina e incluido en parafina (TFFIP) VPH-AR+, con lesiones de bajo grado (n = 10) y lesiones de alto grado (n = 10) con datos de smallRNAseq y validamos cinco miRNAs por qRT-PCR en un grupo independiente de 50 TFFIP. Análisis exploratorios de curvas ROC se usaron para calcular el AUC de los miRNAs para la detección de NIC-3.

Resultados. Se identificaron 454 miRNAs reguladores de alguno de los 19 genes que codifican para proteínas dianas de E6/E7. Por secuenciación se obtuvieron 162 miRNA diferencialmente expresados entre lesiones de alto grado y bajo grado, de los cuales 47 fueron predichos *in-silico*. La expresión relativa de cinco miRNAs por qRT-PCR, fue mayor entre alto grado y bajo grado ($p < 0,05$) y los niveles de miR-2, 3 y 4 aumentaron con la progresión de la enfermedad (sano, NIC 1, NIC 2, NIC 3+). Los mayores FC fueron obtenidos por el miR-1, 2 y 3, de 19.3, 10 y 9.4 respectivamente, en secuenciación, y de 16.8, 6.7 y 6.7 respectivamente, por qRT-PCR, entre alto y bajo grado ($p < 0,05$). El coeficiente de correlación de Pearson entre smallRNAseq y qRT-PCR fue 0.86. El miR-3 y su combinación con miR-1 obtuvo el mayor AUC de 79,09 % para la detección de NIC 3+. En conclusión, se identificaron 162 miRNA por secuenciación, de los cuales 47 regulan transcritos que codifican para proteínas diana de E6/E7, a su vez cinco de estos miRNAs (miR-1, 2, 3, 4 y 5) fueron validados por qRT-PCR mostrando que miRNAs reguladores de transcritos, cuyos genes codifican para proteínas diana de E6/E7, están alterados entre lesiones de alto grado y bajo grado. Se necesita una validación adicional en una cohorte de muestras más grande para confirmar el potencial de estos miRNA para clasificar mujeres VPH-AR+.

Agradecimientos

Mis más especiales agradecimientos para mi tutora la Dra. Gloria Inés Sánchez por transmitirme excelentes valores como profesional e investigadora, también porque desde el pregrado me abrió las puertas del grupo de investigación permitiéndome consolidar bases inquebrantables en el camino de la investigación; y porque me ha apoyado y brindado grandes oportunidades que han cambiado mi forma de ver el mundo.

A mi comité tutorial la Dra. Olga María Bermúdez, el Dr. Luis Fernando Barrera y el Dr. Silvio Urcuqui, de quienes recibí excelentes sugerencias y siempre estuvieron como un libro abierto para ilustrarme, corregirme y asesorarme, gracias por su compromiso con mi formación.

Quiero expresar mi gratitud especialmente al Dr. Jovanny Zabaleta, por haber estado siempre presto a ayudarme y porque gracias a su colaboración fue posible llevar a cabo los experimentos del presente trabajo, también a su equipo de trabajo del laboratorio Translational Genomics Core de Louisiana Cancer Research Center, Louisiana State University Health Sciences Center U.S.A, las Dras. Jone Garai y Li Li por su amabilidad y colaboración durante mi pasantía en New Orleans.

Al Dr. Armando Baena, la Dra. Tatiana Ramírez, la Médico y candidata a Doctora María Cecilia Agudelo y a Melisa Castañeda por ser miembros esenciales en la ejecución de los estudios ASCUS y Biomarcadores. También a todo los asesores nacionales, internacionales, profesionales de la salud, auxiliares de enfermería y miembros del Grupo Infección y Cáncer quienes reclutaron las participantes de los estudios, procesaron las muestras, y velan por el mantenimiento de la información de los estudios, y a las mujeres que aceptaron participar en los estudios, sin la participación de todos ellos no hubiese sido posible realizar este trabajo.

Al Dr. Carlos Orozco quien me asesoró, me aconsejó y brindó mejoras a mi trabajo con mucho entusiasmo y compromiso. También quiero agradecer a Diego Álvarez y al profesor Rodrigo Ochoa por su asesoría, entrenamiento bioinformático y compartirme su experiencia en el manejo de programas especializados.

También quiero agradecer profundamente a mis padres César y Martha que, pese a la distancia, siempre estuvieron incondicionales y a mi hermana Jessica por su compañía y apoyo en este camino. A mis grandes amigos Juan Daniel, Juliana, Cindy y Juan David que siempre me dieron ánimos y me recuerdan esos motivos por lo que hemos decidido hacer Ciencia.

A COLCIENCIAS y el CODI por el financiamiento de los estudios, al concurso Mujeres Jóvenes Talento de la Ciudad de Medellín por la beca de Maestría, al programa posgrados nacionales de SAPIENCIA, al programa Jóvenes Investigadores de COLCIENCIAS. Que hicieron posible la realización de esta Maestría.

1. Introducción

El cáncer cervical es un problema grave de salud pública en el mundo, es la primera causa de muerte por cáncer en mujeres de países con bajos y medianos ingresos, en donde ocurrieron el 95% de 570,000 nuevos casos reportados y de 311,000 muertes estimadas para el 2018 (Ferlay et al. 2018). Además, en los países en desarrollo se presentan barreras al acceso a los servicios de salud, en donde las tasas de incidencia y mortalidad se han mantenido constantes, mientras que en los países desarrollados han disminuido drásticamente en los últimos 50 años (Ferlay et al. 2019). Se espera un incremento en el número de casos y muertes debido al aumento en el tamaño y en la esperanza de vida de la población y porque las vacunas profilácticas tendrán efecto 30-50 años después de lograr alta cobertura en su implementación (Cuzick, Castañón, y Sasieni 2010). Teniendo en cuenta la información disponible sobre las tendencias reportadas cada 4 años por los registros poblacionales, en Colombia, se ha observado una reducción en la tasa de mortalidad por cada 100.000 mujeres, desde un 18.5 en el periodo 1984-88, a una tasa de 7 por cada 100.000 mujeres en el periodo 2009-2011, gracias a la implementación de campañas de tamización y cambios en la exposición a factores de riesgo, como disminución en las tasas de natalidad (Muñoz y Bravo 2014). GLOBOCAN, que alberga los datos globales de cada país, reporta actualmente que en el 2018 la tasa de mortalidad estimada, en Colombia fue del 5,7 (Ferlay et al. 2018).

La incidencia y mortalidad del cáncer de cuello uterino ha disminuido en países que han implementado y sostenido programas masivos de detección basados en citología, sin embargo, estos programas requieren repeticiones de la prueba de citología, múltiples visitas al ginecólogo para procesos de diagnóstico e identificación de las mujeres que están en alto riesgo y que requieren tratamiento. Estos programas también requieren de infraestructura de salud bien constituida, con laboratorios para el procesamiento e interpretación de frotis cervical, mecanismos para el control de calidad, remisión a colposcopia, tratamiento de lesiones precursoras y seguimiento para detectar fallas en el tratamiento. Si bien este enfoque ha tenido éxito en la prevención del cáncer cervical cuando se implementa correctamente, ha resultado excesivamente complejo, difícil de implementar, ineficiente y costoso en países en desarrollo (Denny 2012). Desde otro punto de vista, la citología tiene una sensibilidad limitada, perdiendo aproximadamente el 50% de probabilidad de detectar la lesión de alto grado presente en el momento de la recolección de la muestra; además, múltiples factores como el personal de la toma de muestra, el correcto extendido y coloración, el

citotecnólogo o patólogo que lea la placa, afectan la reproducibilidad de esta prueba. En este contexto, con muy pocas excepciones, incluso en áreas donde la citología está disponible, el impacto es muy limitado sobre la incidencia y la mortalidad del cáncer de cuello uterino (Kitchener, Castle, y Cox 2006).

El virus del Papiloma Humano (VPH), es el agente causal necesario de todos los casos de cáncer cervical y de una fracción de otros cánceres anogenitales, y orofaríngeos. Trece genotipos de VPH son considerados oncogénicos o de alto riesgo (VPH-AR), ellos son, el 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 y 68. Los genotipos 16 y 18 son los más comúnmente encontrados en cáncer cervical y se asocian con el 70% de todos los cánceres (Muñoz et al. 2004; de Sanjose et al. 2010).

La prueba de ADN del VPH (Prueba de VPH) ofrece ventajas sobre la citología a nivel diagnóstico pues ha mostrado alta sensibilidad (superior al 90%), para la diferenciar neoplasias intraepiteliales cervicales grado 2-3 (NIC 2-3) de neoplasias intraepiteliales de bajo grado (NIC 1). Esta prueba además tiene un alto valor predictivo negativo, es decir, un resultado VPH negativo confiere baja probabilidad de desarrollar una neoplasia; por tal razón ha sido introducida como prueba de tamizaje primario para la detección de cáncer de cérvix en mujeres mayores de 30 años (Meijer et al. 2009; Murphy et al. 2012). Sin embargo, ésta presenta una especificidad menor del 50%, dado que las mujeres infectadas con el VPH-AR (VPH-AR+) en su mayoría eliminan la infección viral espontáneamente reduciendo significativamente el riesgo de cáncer (Rijkaart et al. 2012). Esta infección transitoria se presenta principalmente en mujeres menores de 30 años, quienes resuelven la infección sin desarrollar lesiones neoplásicas y cáncer de cuello uterino como resultado de la activación de la respuesta inmune (Doorbar 2006). Lo que quiere decir que solo una baja proporción de las mujeres infectadas con el VPH-AR (VPH+) desarrollan una infección persistente y lesiones precursoras del cáncer cervical (Castle et al. 2009). Dado lo anterior, la Prueba de VPH aumenta la remisión de mujeres VPH+ que no van a desarrollar la enfermedad a procedimientos de diagnóstico adicionales, como colposcopia y tratamientos innecesarios que aumentan los costos en el sistema de salud, pero también se traduce en carga emocional produciendo ansiedad en las mujeres (Castle et al. 2011).

Avances en el entendimiento de los mecanismos moleculares del VPH durante la transformación de la célula hospedera, han mostrado que la expresión de los oncogenes virales E6 y E7 del VPH bloquean la acción supresora tumoral de la retinoblastoma (pRb) y p53, respectivamente. Esto ha

llevado a la evaluación de marcadores que permitan predecir el potencial oncogénico de estas infecciones, mediante la detección directa de las oncoproteínas E6/E7, y de proteínas implicadas en el ciclo celular, tales como p16, la proteína indicadora de división celular Ki67 que usualmente no se expresa en células en reposo como los queratinocitos (Brown et al. 2012). A pesar de los resultados tan promisorios obtenidos con estas pruebas, el hecho de ser proteínas implica que las muestras deben ser preservadas adecuadamente para su procesamiento y en muchos casos requiere inclusive de la lectura e intervención de personal entrenado para su uso. Por esta razón, se necesita desarrollar pruebas que permitan reconocer de las mujeres que son VPH+, aquellas que tienen verdadero riesgo de desarrollar lesiones preneoplásicas de alto grado o cáncer, que a su vez puedan representar una solución para mitigar las limitaciones de los biomarcadores basados en análisis de proteínas ya mencionados.

Los miRNAs son RNAs pequeños no codificantes, de 16 a 24 nucleótidos cuya función es regular negativamente la expresión génica a nivel de la traducción del RNA a proteína. El entendimiento del papel que desempeñan los miRNA en los seres vivos, ha llevado a un gran número de investigaciones sobre su utilidad como indicadores de enfermedades. La expresión diferencial de miRNAs se ha reportado en diversas enfermedades humanas infecciosas, cardiovasculares, autoinmunes, neurológicas y en diversos tipos de cáncer, incluido el cáncer cervical y sus lesiones preneoplásicas en pacientes infectadas con VPH (He et al. 2016). Existe evidencia de que los miRNAs modulan negativamente la expresión de genes supresores de tumores relacionados con procesos carcinogénicos del VPH como *TP53*, *CDKN1A* (p21) y *CDKN1B* (p27) (Beaudenon y Huibregtse, 2008; Milde-Langosch et al., 2001). Estas proteínas supresoras de tumores, en el panorama de la infección por el VPH, son degradadas por las oncoproteínas virales E6 y E7, siendo un paso clave para la desregulación del ciclo celular (Beaudenon y Huibregtse 2008; Milde-Langosch et al. 2001). Los altos niveles de expresión de los miRNAs que modulan la expresión de genes supresores de tumores, se correlaciona positivamente con los niveles de expresión de E6/E7 *in-vitro*, lo que sugiere que la desregulación de miRNAs por la acción de E6/E7 también media la carcinogénesis cervical (Chiantore et al. 2016; Harden et al. 2017; Harden y Munger 2017)

Los perfiles de miRNAs son fácilmente detectables con RNAseq, sin embargo, estos datos deben ser validados por tecnologías más sensibles como la PCR, por lo que el presente estudio se propone hacer una búsqueda *in-silico* y por secuenciación de miRNAs reguladores de transcritos, cuyos

genes codifican para proteínas diana de E6/E7 del VPH-AR, y una posterior validación de cinco miRNAs por PCR en un grupo de muestras independiente.

A pesar de estos avances, no se han llevado a cabo estudios que conduzcan a la identificación racional de miRNAs expresados diferencialmente en lesiones de alto grado y cáncer, respecto a lesiones de bajo grado, como potenciales biomarcadores para la detección de lesiones de alto grado de cérvix. Por tal razón, el objetivo de este trabajo fue *Identificar miRNAs celulares reguladores de las proteínas blanco de acción de las oncoproteínas E6 y E7 del VPH, expresados de manera diferencial en lesiones intraepiteliales de alto grado versus lesiones intraepiteliales de bajo grado, en mujeres infectadas con el VPH-AR.*

El presente proyecto es el primer acercamiento en el país, sobre el papel de los miRNAs como biomarcadores estudiados en biopsias de lesiones preneoplásicas del cérvix de mujeres infectadas con el VPH-AR.

2. Antecedentes

Papel de los miRNAs en Procesos Oncológicos

El desarrollo del cáncer comprende la aparición de seis capacidades biológicas, señalización proliferativa sostenida, evasión de supresores de crecimiento, resistencia a muerte celular, inmortalidad replicativa, inducción de angiogénesis y activación de procesos de invasión y metástasis (Hanahan y Weinberg 2011). El desarrollo y la aparición de un cáncer es causado principalmente por mutaciones en genes supresores tumorales o alteraciones en las proteínas encargadas de controlar la proliferación y el ciclo celular, tales como MYC, proteínas de la familia RAS, supresoras de tumores como p53, pRb, entre otras. Los niveles de estas mismas proteínas en una célula son a su vez controlados, en parte, mediante represión postranscripcional, en gran medida por los miRNAs y se estima que más del 30% de la traducción de proteínas está regulada por miRNAs (Rajewsky 2006), de tal forma que alteraciones en la secuencia de un miRNA o en su biogénesis conducirán a modificaciones en su expresión, aumentando el riesgo de pérdida del control del ciclo celular. Si el RNA mensajero objeto de regulación por un miRNA, codifica para una proteína clave en el control del ciclo celular, fallas en el proceso de control por parte del

miRNA, puede ser la causa desencadenante de alteraciones cancerígenas. En esencia los miRNAs poseen dos mecanismos por los que pueden desencadenar un proceso tumoral (Figura 1).

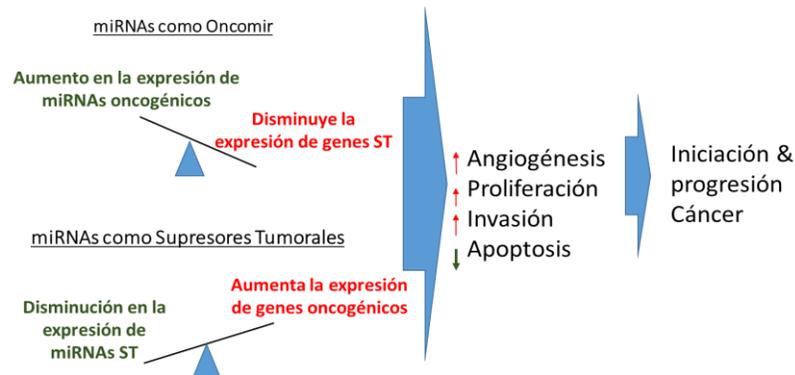


Figura 1. Papel de los miRNAs en Tumorigénesis. El aumento de la expresión de miRNAs de tipo Oncomir, aumenta la regulación negativa de genes que codifican para proteínas ST, y la disminución de la expresión de miRNAs de tipo ST desencadena un aumento en la expresión de genes proto oncoogénicos, que no logran ser regulados negativamente por estos miRNAs. Estos dos fenómenos dirigen a la aparición de “hallmarks” del cáncer como la angiogénesis, pérdida del control celular, invasión y reducción de la apoptosis. ST: Supresores tumorales. Elaboración propia

- I. A través de la reducción en la cantidad en que se produce, si se trata de un “miRNA supresor de tumores”, es decir, cuya función es reprimir la traducción de un RNAm que puede codificar para una proteína implicada en la regulación de la expresión de proteínas oncoogénicas. La reducción de su expresión o disponibilidad dirigirá la función celular hacia la formación de tumor, lo cual puede deberse a un defecto genético en cualquier etapa de su biogénesis que provoque una expresión defectuosa o reducida del miRNA.
- II. Aumento en los niveles de expresión de un miRNA oncoogénico “Oncomir”, es decir, cuya función normal es reprimir la traducción de un RNAm que puede codificar para una proteína implicada en la regulación de la expresión de proteínas supresoras de tumores, y al estar aumentado el Oncomir, bloquea el paso de RNAm a proteínas supresoras, aportando al desarrollo del tumor. Además de estos mecanismos, cualquier alteración genética en el extremo 5’ de la región semilla de un miRNA puede provocar que éste cambie su diana y altere cascadas de señalización que provoquen por su parte el desarrollo tumoral (Esquela-Kerscher y Slack 2006)

Aproximación in-silico: interacción miRNAs y RNAm dianas

Entre las tecnologías para determinar la abundancia relativa de los miRNAs está la secuenciación de RNAs pequeños, qRT-PCR e hibridación (microarreglos) (Mestdagh et al. 2014). El RNAseq se basa en la secuenciación de DNA copia (DNAc), comúnmente usada para la búsqueda masiva de perfiles transcriptómicos, actualmente dependiendo de lo que se quiera secuenciar se acopla con kit comerciales de captura de RNA total, para enriquecer RNAs pequeños, por ejemplo, luego se convierte el RNA en DNAc usando un enzima transcriptasa reversa. Se ligan adaptadores en sus extremos 3' y 5' del DNAc y se amplifica para obtener la librería que será secuenciada en una plataforma NGS (Illumina, SOLiD y Roche las más usadas). El RNAseq permite explorar masivamente, sin necesidad de conocer previamente el transcriptoma o el genoma y no requiere hibridación de sondas. El RNAseq posee alta sensibilidad analítica, requiere entre 5 y 10 µg, con una concentración alrededor de 500 ng/µl de RNA, ofrece cobertura completa de los transcritos y alta correlación con los resultados que se obtienen por PCR (Carolina et al. 2012).

Los recientes avances tecnológicos, especialmente la secuenciación de alto rendimiento, han conducido a un crecimiento exponencial en la generación de datos relacionados con miRNAs. Por lo anterior, se han desarrollado herramientas bioinformáticas para gestionar y manejar el flujo creciente de datos relacionados con miRNAs, como herramientas de predicción de interacciones, y de búsqueda de interacciones miRNAs-RNAm validadas. Las herramientas de predicción de miRNAs o de RNAm que son regulados por miRNAs, se basan en esencia en características intrínsecas de la interacción miRNAs-RNAm. Siendo un reto para los algoritmos computacionales superar las particularidades de la naturaleza de los miRNAs, tal como su promiscuidad en el reconocimiento del RNAm, que le otorga la capacidad, a un solo miRNA, de interactuar y regular cientos de RNAm. Al tiempo un mismo RNAm puede ser blanco de diferentes miRNAs, al coincidir parcial o totalmente su secuencia en la región 3' UTR con la secuencia semilla (6-8 nt) de diferentes miRNAs, lo cual incrementa la posibilidad de que se produzca de manera efectiva su silenciamiento (Peter 2010).

En general, todos los algoritmos bioinformáticos para predecir interacciones miRNA-RNAm tienen en consideración cinco aspectos:

- ✓ Patrón del apareamiento de bases en la secuencia semilla
- ✓ Estabilidad termodinámica del dúplex miRNA/RNAm, considerado por la mayoría de los algoritmos de predicción, como la mínima energía libre de unión.
- ✓ Conservación de los sitios blanco: Regiones 3'-UTR ortólogas en RNAm
- ✓ Presencia de múltiples sitios blanco.
- ✓ Estructura secundaria de la región 3' UTR del transcrito.

Estas y otras características permiten asignar un puntaje para establecer si un RNAm es o no considerado blanco de uno o varios miRNAs. La base de datos miRWalk.2.0 (última actualización 15.02.2016) (<http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/>) contiene información producida con TarPmiR, una herramienta de predicción de sitios de unión de miRNA-mRNA que utiliza datos de experimentos de unión de miRNA-RNAm como CLASH (*crosslinking ligation and sequencing of hybrids*) (Dweep y Gretz 2015; Dweep, Sticht, y Gretz 2013). TarPmiR aplica el algoritmo Random Forest para integrar las características convencionales para predecir los sitios objetivo de miRNA (Ding, Li, y Hu 2016).

El algoritmo TarPmiR comienza con un heptámero que inicia en el nucleótido 1 o 2 del extremo 5' del miRNA y recorre la secuencia completa del gen que esté analizando hasta el extremo 3'. Tan pronto como identifica un apareamiento perfecto del heptámero, extiende la longitud del alineamiento hasta que aparece un “*mismatch*” (desajuste). En consecuencia, el algoritmo devuelve todos los posibles “hits” con 7 o más nucleótidos con apareamiento perfecto. El algoritmo asigna la predicción de los resultados en cinco partes: región promotora, 5'-UTR (“untranslated region”), CDS (“coding sequence”), 3'-UTR, y genes mitocondriales. Adicionalmente, este algoritmo permite, dentro de su configuración, la combinación de 12 bases de datos diferentes (DIANA-microTv4.0, DIANA-microT-CDS, miRanda, mirBridge, miRDB4.0, miRmap, miRNAMap, PicTar2, PITA, RNA22v2, RNAhybrid2.1 y TargetsCan6.2), de manera que los resultados de la predicción obtenidos tienen menor probabilidad de falsos positivos (Dweep y Gretz 2015).

Relación miRNAs y Cáncer Cervical

La expresión aberrante de miRNAs ha sido reconocida como un mecanismo molecular importante que conduce a la carcinogénesis, dada la capacidad que tienen los miRNAs de regular la expresión de varios genes implicados en procesos celulares importantes, e influir en la regulación de las vías de señalización relacionadas con el cáncer (Calin y Croce 2006).

Gran número de revisiones resumen los resultados de estudios que muestran la relación entre miRNAs y la carcinogénesis cervical mediante la comparación de los perfiles de expresión de miRNAs, entre tejidos cancerosos y tejidos cervicales normales, sugiriendo que la expresión aberrante de miRNAs es un factor importante en el desarrollo de este cáncer (González-Quintana et al., 2016; Guerrero y Guerrero, 2016; Rao et al, 2012). Adicionalmente, se han recopilado evidencias sobre el papel de los miRNAs en las etapas de progresión y desarrollo del cáncer, lo que ha motivado a estudiar la mecánica molecular de los miRNAs en el cáncer cervical y sus lesiones precursoras (Gocze et al. 2013; Li et al. 2011).

A medida que avanzan las investigaciones sobre miRNAs asociados a cáncer cervical, el número de miRNAs implicados en su desarrollo incrementa considerablemente. Algunos de los principales miRNAs encontrados con altos niveles de expresión en cáncer cervical versus pacientes sanas son: miR-10b, miR-15a, miR-16, miR-17, miR-20b, miR-21, miR-93, miR-106a, miR-106b, miR-1143, miR-146b-5p, miR-155, miR-185, miR-195, miR-339-5p, miR-625, miR-941 y miR-1224-5p; Los miRNAs encontrados con bajos niveles de expresión son: miR-99a, miR-100, miR-125, miR-139-5p, miR-139-3p, miR-199a, miR-199b-5p, miR-149, miR-328, miR-375, miR-379, miR-381, miR-497, miR-574-3p, y miR-617 (Hu et al. 2010; Huang et al. 2012; Yoon y De Micheli 2005), y cada vez son más los estudios que describen perfiles de miRNAs expresados diferencialmente en lesiones avanzadas y cáncer versus lesiones leves y/o pacientes sanas.

Se han reportado alteraciones en patrones de expresión de algunos miRNAs durante la progresión de tejidos cervicales normales a lesiones intraepiteliales de alto grado (NIC2 y NIC3) y/o tejidos con cáncer invasivo. Tal es el caso de los análisis de microarreglos de Liang y colaboradores (2014) que mostraron que los niveles de miR-182, miR-183 y miR-210 están más altos y miR-128, miR-143, miR-145 y miR-195 están más bajos en cáncer cervical con respecto a tejido normal. Un conjunto particular de estos miRNAs se asoció con la progresión del epitelio cervical normal a

NIC2 y de NIC3 a Carcinoma Escamocelular Invasivo. Estas observaciones sugieren una asociación entre la desregulación de miRNAs en cáncer cervical con la progresión de la enfermedad (Liang et al. 2014).

miRNAs como biomarcadores

El VPH desempeña un papel fundamental en la etiología de distintos carcinomas, incluidos cáncer cervical y cáncer de cabeza y cuello. El cáncer cervical sigue siendo el cuarto cáncer más común en todo el mundo y la principal causa de mortalidad por cáncer en países en desarrollo. Sin embargo, independientemente de los esfuerzos sustanciales en el campo para desarrollar métodos de detección y prevención, no hay una solución efectiva disponible para combatir estas enfermedades letales. La progresión de la enfermedad mediada por la infección del VPH se asocia con diversos eventos como cambios en los mecanismos moleculares que podrían discriminar la infección con VPH, de la lesión precursora de alto grado, y eventualmente cáncer. Estos eventos representan biomarcadores potenciales y comprenden principalmente cambios epigenéticos como la metilación viral y del hospedero, y la expresión desregulada de miRNAs celulares (Figura 2). Estos biomarcadores podrían mejorar la fiabilidad de la detección y prevención del cáncer (Sahasrabudde, Luhn, y Wentzensen 2011; Schiffman y Wentzensen 2013).

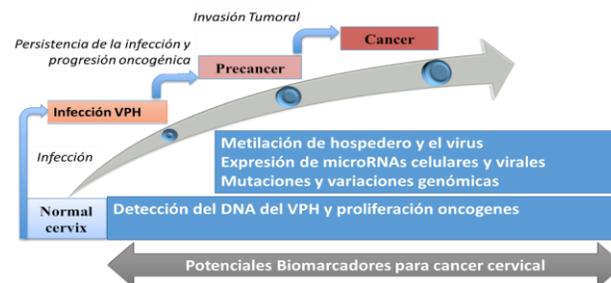


Figura 2. Potenciales biomarcadores para cáncer cervical. Tomado y adaptado de (Kumar Gupta y Kumar 2015).

Perfiles o firmas de expresión de miRNAs se han observado en patologías que van desde el cáncer hasta enfermedades parasitarias (Cheng, 2015), sugiriendo que los miRNAs tienen patrones específicos al tipo de tejido y estado fisiológico. Los miRNAs poseen propiedades deseables en comparación con otros biomarcadores, son útiles para el diagnóstico clínico dada su gran estabilidad a condiciones no controladas de temperatura y pH, son altamente estables frente a la acción de las ribonucleasas (Turchinovich et al. 2011), poseen larga duración (vida media) con respecto a los RNAm tanto *in-vivo* (Lim et al. 2005; Xi et al. 2007) como *in-vitro* (Tang et al. 2006), permitiendo el análisis en TFFIP, sangre, suero y orina para aplicaciones rutinarias de

diagnóstico (Glinge et al. 2017; Jung et al. 2010; Kakimoto et al. 2016). Por estas razones, gran número de investigaciones han dirigido sus esfuerzos para estudiar el papel de los miRNAs y se han postulado como biomarcadores de lesiones preneoplásicas de alto grado y en cáncer cervical (He et al. 2016).

Los últimos estudios en el tema se han centrado en identificar los miRNAs expresados diferencialmente en lesiones de alto grado versus lesiones de bajo grado del cérvix de mujeres VPH-AR+. Estudios más recientes aún, se enfocan en determinar el desempeño diagnóstico en términos de sensibilidad y especificidad de miRNAs candidatos (Babion et al. 2019; Babion et al. 2018; Ivanov et al. 2018; Snoek et al. 2019; Tian et al. 2014; Wang et al. 2014; Wilting et al. 2013a).

Algunos de estos se describen a continuación:

Babion y colaboradores (2019) evaluaron el potencial marcador de 10 miRNA candidatos en 209 exfoliados cervicales de mujeres sin lesión, o con lesión NIC 1-2-3 y cáncer, todas VPH-AR+, e identificaron un clasificador compuesto por miR-149, miR-20a y miR-93 con un AUC de 83% para la detección de NIC 3 (Babion et al. 2019). También evaluaron el valor clínico de 8 miRNAs genética o epigenéticamente desregulados, como marcadores para el triaje en 58 biopsias y 225 exfoliados cervicales de mujeres sin lesión, o con lesión NIC 1-2-3 y cáncer, todas VPH-AR+, y logran identificar un clasificador compuesto por miR-15b-5p y miR-375 con un AUC de 62%, sensibilidad y especificidad de 55 y 70 % respectivamente, para la detección de lesiones de alto grado y cáncer (Babion et al., 2018)

Snoek y colaboradores (2019) usaron un análisis de Genome-Wide para identificar miRNAs, el análisis por curvas ROC de datos de secuenciación mostró un marcador compuesto por 9 miRNAs (let-7b, miR-9, miR-15b, miR-20a, miR-31, miR-93, miR-183, miR-184, miR-222) con un AUC de 89% para la detección de NIC3. Además, validaron 6 miRNAs por PCR en un grupo de 191 muestras independientes (101 NEG, 48 NIC3, 41 SCC) VPH+ (de autotoma). Cinco de los miRNAs (let-7b, miR-15b, miR-20a, miR-93 y miR-222) presentaron un valor de AUC combinada de 78% para la detección de NIC3+ (Snoek et al., 2019).

Otros autores, Tian y colaboradores (2014) proponen un panel de miRNAs compuesto por miR-424/375/218 con valores de sensibilidad y especificidad más altos que la citología para detectar

lesiones NIC2+ (Tian et al. 2014). Ivanov y colaboradores (2018) evaluaron 25 miRNAs (20a-5p, -21-5p, -99a-5p, -146b-5p, -200b-3p, -23a-3p, -106b-5p, -155-5p, -203a-3p, -31-5p, -125b-5p, -181b-5p, -375, -34a-5p, -126-3p, -191-5p, -1246, -96-5p, -145-5p, -192-5p, let-7d, -143-3p, -196b-5p, -146a-5p, -197-3p) por qRT-PCR a partir de exfoliados cervicales de 174 mujeres (40 NEG, 34 NIC1, 57 NIC2-3 y 43 cáncer invasivo), de las cuales 144 eran HPV-AR+; y determinaron un clasificador compuesto por 8 miRNAs (miR-20a/375/21/34a/96/145/196b/1246), con un AUC de 82,4% para la detección de NIC2+ (Ivanov et al. 2018).

Es también importante resaltar, que la expresión de los miRNAs, así como la expresión de los genes que codifican proteínas, está controlada por mecanismos genéticos y epigenéticos. Y es la interrupción de estos procesos de control, la que conduce a expresión anormal de miRNAs en cáncer (Moutinho y Esteller 2017). Dado que los miRNAs pueden regularse epigenéticamente mediante metilación del DNA y/o modificaciones específicas de histonas, varios estudios han explorado estos mecanismos en búsqueda de biomarcadores de lesiones de alto grado de cérvix.

Algunos de estos estudios se describen a continuación:

Jiménez-Wences y colaboradores (2016) encontraron que la expresión de miR-124 y miR-218 fue significativamente menor, y la de miR-193b significativamente mayor, en cáncer cervical comparado con lesiones de bajo grado (10 NEG/VPH-, 7 NEG/VPH+, 16 NIC1 y 16 cáncer). Además, las diferencias de expresión se correlacionaron con niveles de metilación, de los promotores de estos miRNAs: niveles de metilación significativamente alto en los promotores de miR-124-2, miR-218-1/2 y miR-34b/c fueron significativamente mayores en cáncer cervical comparado con el nivel de metilación en lesiones de bajo grado. Por el contrario, el nivel de metilación del promotor miR-193b fue significativamente más bajo en cáncer cervical que en lesiones de bajo grado. La selección de los miRNAs se basó a reportes previos sobre el papel de éstos en otros tipos de cáncer (Jiménez-Wences et al. 2016).

Wilting y colaboradores (2013b) llevaron a cabo un análisis de metilación específica por PCR (MSP) en un panel de líneas celulares que representan los diferentes estados de la transformación inducida por la infección de VPH, encontrando un incremento en la metilación de miR-149, -203 y -375 en líneas de carcinoma (Wilting et al., 2013b). Estos investigadores, también evaluaron la función supresora tumoral y el silenciamiento mediado por la metilación del miR-124-1/2 en 139

biopsias mostrando diferencias en la frecuencia metilación en biopsias de tejido normal (0%) hasta 93% en cáncer cervical. Los niveles de metilación de miR-124-1/2 fueron significativamente correlacionados con la reducción de la expresión de miR-124 en cáncer cervical (Wilting et al., 2010)

Botezatu y colaboradores (2011) llevaron a cabo un análisis de metilación específica por PCR (MSP), evaluaron la metilación de las islas CpG alrededor de miR-124a, miR-34b y miR-203 en DNA extraído de 29 lesiones precursoras del cáncer, 31 cáncer y 30 sanos, todos VPH+, y encontraron hipermetilación de estos miRNAs en lesiones precursoras y cáncer comparado con muestras de individuos sanos (Botezatu et al. 2011).

Los anteriores avances confirman la participación de alteraciones epigenéticas en la oncogénesis cervical y sugieren que los miRNAs representan un biomarcador candidato para uso clínico en el triaje de mujeres VPH-AR+ (es decir, para clasificar mujeres VPH-AR+ con mayor riesgo de desarrollar una lesión de alto grado o cáncer cervical, de mujeres VPH+ con menor riesgo).

3. Marco Conceptual

La infección persistente por genotipos del VPH-AR es la causa necesaria mas no suficiente de cáncer cervical (Meijer et al. 2009). La infección por el VPH se adquiere con el inicio de las relaciones sexuales; en la gran mayoría de los casos (95%, aproximadamente) se presenta una infección autocontrolada que se resuelve en un plazo de 18 meses (Rodríguez et al. 2008). Ocasionalmente, dicha infección puede inducir la aparición de una Neoplasia Intraepitelial Cervical (NIC), que en su mayoría son benignas o displasias de bajo grado o leves (NIC 1); sin embargo, estas lesiones regresan espontáneamente en la mayoría de los casos. Cuando el virus no se elimina y la infección con VPH-AR persiste, la lesión precursora se mantiene y entre un 5 y un 10% de estas lesiones pueden progresar hasta una lesión intraepitelial de alto grado, o mayor (neoplasia intraepitelial cervical de grado 2 , grado 3 (NIC 2 y NIC 3) la cual, si no es tratada a tiempo, progresa a carcinoma escamo celular (CEC) (Lizano, Carrillo, y Contreras s. f.; Rodríguez et al. 2008; Schiffman et al. 2007)

La progresión desde la infección persistente con VPH hasta el desarrollo de cáncer invasivo puede tomar en promedio 15-20 años (Castle et al. 2009). Los cofactores principales que modulan el riesgo de cáncer cervical en mujeres VPH-AR positivas son ambientales (uso de anticonceptivos hormonales y el tabaquismo), virales (infección por tipos específicos del VPH, las variantes del VPH y la carga viral), y relacionados con el hospedero (genética, las hormonas y las respuestas inmunes) (Aguilar-Fabré y Ríos-Hernández, 2008).

Virus del Papiloma Humano VPH

El VPH tiene un genoma de DNA circular de doble cadena de aproximadamente 8000 pares de bases (pb) y una cápside de simetría icosaédrica sin envoltura. El VPH tiene tropismo por células del epitelio escamoso, cutáneo o mucosas. Existen aproximadamente 150 tipos de VPH, que se diferencian principalmente por su tropismo epitelial (epitelio mucoso o cutáneo). Generalmente, la infección por VPH causa condilomas, o es asintomática; sin embargo, existen genotipos de alto riesgo potencialmente oncogénicos, tal como el 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 y 68 (de Sanjosé et al. 2010). Del genoma de VPH se diferencian tres regiones (Figura 3), i) la región no codificante, conocida como región larga de control (“Long control regulator LCR”), o región reguladora principal, además de contener el origen de la replicación y los sitios de unión para factores de transcripción viral y celular, también se encarga de la regulación de la expresión de los genes virales E6 y E7, ii) la región codificadora de proteínas no estructurales conocida también por estar compuesta por los genes de expresión temprana o E (“early”), E1, E2, E4, E5, E6 y E7 que codifican para proteínas involucradas en la transcripción, replicación y oncogénesis viral y cuya función es garantizar la replicación completa del virus. Las oncoproteínas E6 y E7 participan principalmente en el proceso oncogénico desregulando el ciclo celular inhibiendo p53 y pRb, y activando factores de transcripción necesarios para la replicación celular; esta región se localiza en posición adyacente y corriente abajo del LCR. iii) la región codificadora de proteínas estructurales de la cápside, compuesta por genes de expresión tardía o (“late”), L1 y L2 facilitadores del empaquetamiento y la maduración del DNA viral (Doorbar 2006).

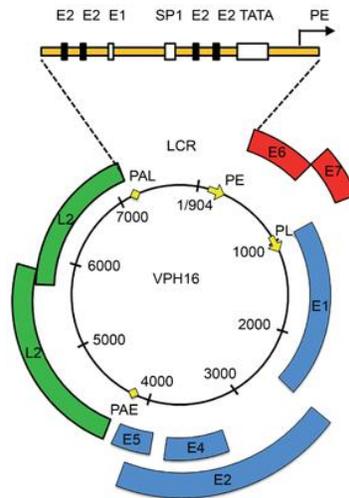


Figura 3. Diagrama del genoma del VPH tipo 16. Genoma de doble cadena circularizado de 7904 pb. Se señalan los promotores PE y PL, para la expresión de los genes tempranos y tardíos, respectivamente, así como de los elementos PAE y PAL, que son señales de poliadenilación para los genes tempranos y tardíos, respectivamente. En la parte superior se observa una amplificación de la región larga de control (LCR) con diferentes elementos de respuesta a factores de transcripción virales y celulares. Figura tomada y modificada de Doorbar y colaboradores (Doorbar et al. 2012)

Oncoproteínas E6 y E7 del VPH-AR

Las propiedades oncogénicas del VPH AR se han asociado principalmente con la actividad de sus proteínas E6 y E7. E6 inmortaliza células humanas con baja eficiencia, pero con la cooperación de E7, transforma queratinocitos humanos (Harden y Munger 2017). Algunos estudios han evidenciado que E6 y E7 contribuyen de diferentes maneras a la carcinogénesis en ratones transgénicos. E7 promueve el comienzo de la formación del tumor, mientras que E6 acelera la progresión de los tumores a la malignidad; de esta forma, E6 y E7 cooperan juntas a la inducción del tumor (Munger et al. 2004).

En particular, E6 es una proteína de 158 aminoácidos altamente conservados, pues contiene dos motivos o dedos de Zinc Cys-X-X-Cys, cuya integridad es esencial para su interacción con proteínas de la célula hospedera (Vande Pol y Klingelutz 2013). E6 está presente tanto en núcleo como citoplasma de la célula hospedera. En el extremo C-terminal la proteína E6 tiene el motivo X-(T/S)-X-(V/I/L), que se conserva entre los VPH-AR y que se une a la familia de las proteínas PDZ. La proteína E6 es esencial para el ciclo productivo del virus y el mantenimiento del DNA viral en los queratinocitos y presenta actividades oncogénicas claramente descritas; además, se une a un sinnúmero de blancos celulares, entre los que se resalta la proteína supresora de tumores p53.

E6 se une y degrada a p53 a través de la proteína ubiquitina ligasa celular E3, llamada proteína asociada a E6 (E6-AP) (Beaudenon y Huibregtse 2008) por lo cual resulta en la ubiquitinación y degradación de p53, lo que le permite bloquear la apoptosis, interferir con la diferenciación celular e inducir la transformación celular en las células infectadas con el VPH (Vande Pol y Klingelutz 2013).

E6 también se une al coactivador de p53, p300/CBP, inhibiendo la actividad transcripcional intrínseca y reduciendo la capacidad de p300 de coactivar elementos promotores de p53 (Patel et al. 1999). E6 puede interrumpir la apoptosis mediante mecanismos independientes de la p53, uniéndose a Bak (una proteína proapoptótica de la familia Bcl-2) e induciendo su degradación (Mantovani y Banks 2001). E6 induce, además, la degradación de la proteína humana PDZ hDlg, la cual está involucrada en la formación de uniones celulares, la regulación de la adhesión celular y la proliferación de los tejidos epiteliales. Esta alteración induce una desregulación en el control del crecimiento invasivo de las células epiteliales, favoreciendo la formación e invasión del tumor maligno (Pim y Banks 2010). E6, además de unirse a la telomerasa y activar la transcripción de la subunidad catalítica de esta enzima, lo cual ayuda al mantenimiento de las estructuras teloméricas contenidas al final de los cromosomas, permite la proliferación celular sostenida (Vande Pol y Klingelutz 2013), e interfiere con la respuesta inmunológica ante la infección viral. La proteína E6 tiene una función inmunosupresora mediante su unión al transactivador de transcripción IRF-3, afectando su función de activador transcripcional, lo que directamente interrumpe la expresión de los genes inducidos por el interferón. Además, se une a la proteína tirosina cinasa 2 (tyrosine-protein kinase 2, Tyk2), una proteína transductora de señales de la vía Jak-STAT y, por lo tanto, interfiere en la activación de la vía de señalización del IFN- α (Vande Pol y Klingelutz 2013)

Por otro lado, la oncoproteína E7, es una fosfoproteína con cerca de 100 aminoácidos, posee dos dedos de zinc en su extremo carboxilo terminal, que usa para su dimerización (Roman y Munger 2013). E7 puede auto ensamblarse dentro de oligómeros esféricos y es fosforilada por la caseína quinasa II, acomplejando pRb, p107, y p130, resultando en la fosforilación y la liberación de E2F, y de esta forma promueve la progresión del ciclo celular de G1 a la fase S (Felsani, Mileo, y Paggi 2006). E7 también se une a las histonas deacetilasas (HDACs), a la subunidad 4 (S4) del proteosoma, a Mi2B, AP-1, MPP2, y proteínas de unión a la región TATA (White, Pagano, y Khalili 2014). E7 contiene dos regiones homólogas con la proteína E1A del adenovirus en la región

amino terminal la; estos dominios son importantes para la actividad transformadora, pues permiten la interacción con diferentes proteínas celulares, incluidas la unión y la degradación de la proteína supresora de tumores del retinoblastoma (pRb), lo que contribuye a la progresión maligna y la inducción de la inestabilidad genómica en las células infectadas (Roman y Munger 2013).

E7 se acompleja con la forma hipofosforilada de pRb e induce su degradación, induciendo a la vez la activación constitutiva del factor de transcripción E2F, promoviendo la activación de genes involucrados en la síntesis DNA durante la fase S del ciclo celular (Roman y Munger 2013). E7 se une a miembros de la familia de factores de transcripción AP-1, tales como c-Jun, JunD y c-Fos, importantes en las vías de señalización de la mitosis y de diferenciación celular. E7 se une a c-Jun y activa su promotor por un mecanismo de transactivación, potenciando la activación de genes involucrados en la progresión del ciclo celular y la mitosis (Antinore et al. 1996).

La figura 4 representa los principales mecanismos moleculares que son alteradas por E6 y E7 del VPH-AR. E7 se dirige a los miembros de la familia de la proteína del Retinoblastoma (pRb) para la degradación, liberando E2F, que permite que las células basales y parabasales entren en fase S. La liberación de E2F depende de factores de crecimiento externos, que estimulan la actividad de ciclina dependientes de quinasas (CDKs). La expresión de proteínas celulares involucradas en la progresión del ciclo celular está regulada por la proteína p16, y esta a su vez en un ciclo de retroalimentación negativa mediante la supresión de la actividad de la ciclina D/CDK4/6. Por otro lado, la oncoproteína viral E6 ubiquitina a p53 induciendo degradación a través de la ruta del proteosoma, y en las células no infectadas (Figura 4A), los niveles de p53 se mantienen en un nivel bajo (Doorbar et al. 2015).

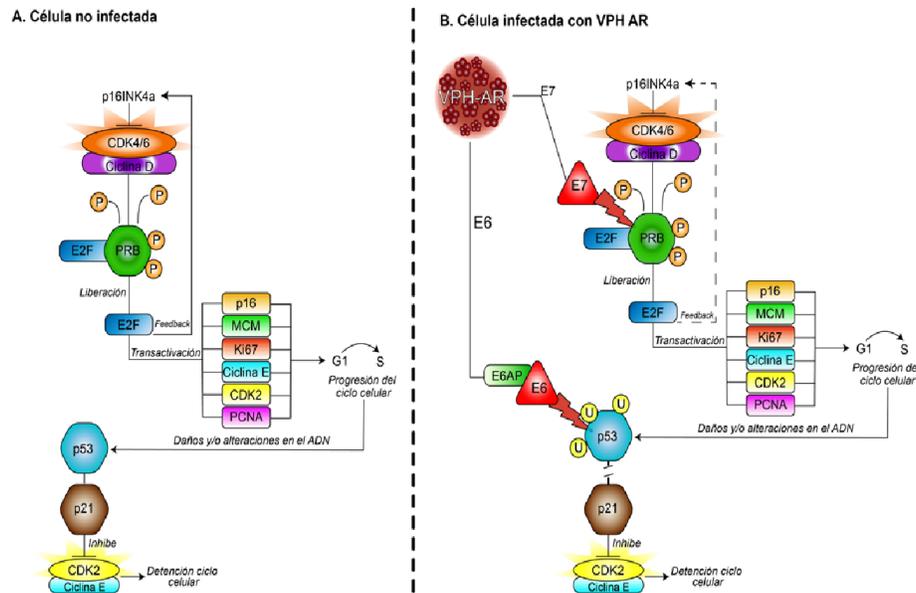


Figura 4. Esquema comparativo de la regulación del ciclo celular. En una célula normal no infectada (A) versus una célula infectada con VPH-AR (B). En una célula no infectada las proteínas supresoras de tumores pRb y p53 regulan la proliferación celular, mientras que en las células con infección transformadora por la acción del VPH-AR, las oncoproteínas E6 y E7 desregulan los puntos de control de la proliferación celular, inhibiendo el mecanismo de control de p53 y pRb, respectivamente. Tomada de Ramírez-Pineda 2019 (Ramírez-Pineda et al., 2019).

microRNAs

Los miRNAs son un grupo de pequeños RNAs de 19 a 25 nucleótidos de longitud, no codificantes, conservados evolutivamente, altamente estables, y participan en la regulación génica a nivel postranscripcional. Por tal razón, los miRNAs intervienen en procesos celulares importantes, como diferenciación celular, proliferación celular, apoptosis y en el desarrollo del cáncer, mediante la promoción o inhibición del crecimiento del tumor, la progresión, y metástasis (Turchinovich et al. 2011). Un miRNA por lo general bloquea o inhibe la traducción de múltiples RNAs mensajeros (hasta 200), lo que les permite regular muchos genes en una ruta o proceso fisiológico al mismo tiempo (Krek et al. 2005). Diversos estudios muestran que la expresión aberrante de miRNAs está asociada al desarrollo de cáncer cervical, indicando que pueden llegar a ser valiosos biomarcadores (Ma et al., 2014; Park et al., 2014). De acuerdo con la última versión de la base de datos de secuencias miRBase (publicada en octubre 2018, versión 22.1), más de 38.589 miRNAs han sido registrados en 271 especies. En el genoma humano se han identificado alrededor de 2000 miRNAs precursores y 2654 miRNAs maduros (Griffiths-Jones 2006).

cola Poli A en el extremo 3'. Los pri-miRNA son clivados por la RNasa III (enzima endoribonucleasa), llamada también DROSHA, dando lugar a intermediarios (pre-miRNA) con aproximadamente 70 nucleótidos. Los pre-miRNA son exportados al citoplasma a través de la proteína transmembranal nuclear Exportina 5, donde luego son clivados por otra RNasa III llamada DICER, dando lugar a un RNA de cadena doble de 16 a 24 nucleótidos que se disocia por acción helicasa y en la que cada cadena es un miRNA. Sin embargo, uno es degradado y el otro, que proviene de la cadena guía, es funcional y es incorporado en el complejo de silenciamiento inducido por el RNA (RISC), compuesto, a su vez, por proteínas argonautas (1- 6). El complejo RISC-miRNA se dirige a sus ARNm dianas y se une a los sitios complementarios: si la complementariedad de bases entre miRNA y ARNm es completa, se degrada el transcrito; si la complementariedad es parcial, hay represión de la traducción. La expresión de las oncoproteínas E6 y E7 (rombos) en una célula infectada con VPH-AR, desregula los niveles de proteínas supresoras tumorales (círculos) importantes en el controlar el ciclo celular (la desregulación se representa con rayas negras), y a su vez se induce la sobreexpresión de los miRNA que reprimen la expresión de las mismas proteínas (la represión se representa con rayas azules). Elaboración propia.

Regulación de la expresión génica por miRNAs

El complejo RISC-miRNA reconoce a los RNAm dianas específicos al unirse a secuencias complementarias y modular negativamente la síntesis de las proteínas. Esta regulación se da de acuerdo con el grado de complementariedad entre la secuencia semilla del miRNA (7-9 nt) y su respectiva diana, conllevando a una de las dos formas siguientes (figura 5 paso 6):

- I. Si la unión por complementariedad de bases entre el miRNAs y su RNAm diana es imperfecta o parcial, la traducción del RNAm diana es bloqueada, al inhibir la unión de la subunidad pequeña del ribosoma al transcrito. Generalmente los sitios de unión están ubicados en regiones no traducidas del extremo 3' (UTR 3') de RNAm diana.
- II. Por otro lado, los miRNAs que se unen perfectamente o casi totalmente a su RNAm diana, inducen la escisión del RNAm diana. Estos sitios de unión se encuentran generalmente en la secuencia codificante o marco de lectura abierto (ORF) del RNAm de la diana (Esquela-Kerscher y Slack 2006).

4. Planteamiento del problema

Como se mencionaba en la introducción, el VPH-AR es el agente etiológico del cáncer cervical. Sin embargo, dado que se requieren otros factores de riesgo que llevan a la aparición de lesiones preneoplásicas y el cáncer, la infección persistente por genotipos de VPH-AR es la causa necesaria, **MAS NO SUFICIENTE**, para la aparición de este cáncer. Dichos factores pueden reflejarse en la capacidad de la mujer para eliminar la infección o en los mecanismos que se afectan a nivel celular. Se sabe por ejemplo, que durante el desarrollo del cáncer cervical ocurre una cascada de eventos anormales que altera el control del ciclo celular, induce la perturbación de la respuesta inmunitaria antitumoral y la alteración de la expresión génica, y alteración de procesos epigenéticos como la metilación y la desregulación de los miRNA (Kori y Arga 2018).

El vacío en el conocimiento de la prevención del cáncer cervical es resolver cómo clasificar correctamente las mujeres VPH-AR+, dado que, según la historia natural de la enfermedad del cáncer cervical, la gran mayoría de mujeres VPH+ eliminarán la infección y solo en el 10% de mujeres infectadas por el VPH, la infección será persistente, llevando a la aparición de la neoplasia intraepitelial de grado 2 o más severas lesiones preneoplásicas (NIC 2, NIC3 y carcinoma in situ) o cáncer). Se requiere desarrollar pruebas que permitan reconocer a las mujeres con alto riesgo de desarrollar cáncer entre las mujeres infectadas con el VPH de alto riesgo, es decir, una prueba que con mayor especificidad y de esta forma se logre hacer triaje inmediato y acertado a las mujeres infectadas con VPH a procedimientos de colposcopia para la confirmación del diagnóstico por un análisis histopatológico (prueba estándar de oro), eliminando visitas de seguimiento innecesarias en mujeres VPH-AR positivas, pero sin este riesgo.

Se ha demostrado que la transformación maligna por la infección VPH-AR, se induce por la acción de las oncoproteínas virales E6 y E7, lo cual ha llevado al descubrimiento y validación de biomarcadores, que predicen el potencial oncogénico de estas infecciones, tales como la detección directa de las oncoproteínas E6/E7 y la detección de proteínas del ciclo celular como p16/Ki67. A pesar de los resultados tan promisorios con estas pruebas, el hecho de que son pruebas que detectan proteínas, implica que las muestras deben ser preservadas adecuadamente para su procesamiento y en muchos casos inclusive requiere de la lectura e intervención de personal entrenado para la lectura e interpretación del resultado. Por esta razón se necesita ofrecer pruebas cuyos resultados automatizados eviten la necesidad de intervención con personal altamente entrenado, más estables

y simples de usar especialmente en regiones remotas de países y regiones de bajos ingresos e índices de desarrollo.

5. Justificación

Los miRNAs son RNAs pequeños no codificantes que poseen características deseables como biomarcadores; son moléculas estables y fácilmente identificables con tecnologías de secuenciación de alto rendimiento (RNAseq), pueden ser cuantificados en tiempo real mediante Reacción en Cadena de la Polimerasa acoplada a la transcripción reversa (qRT-PCR), ensayos de fluorescencias (FISH) y de hibridación directa en pequeños formatos (Condorelli y Dimmeler 2008). La expresión aberrante de miRNAs se ha reportado en diversas enfermedades humanas infecciosas, cardiovasculares (Zhou et al. 2018), autoinmunes (Luck, Muljo, y Collins 2015), neurológicas (Wang, Kwon, y Tsai, 2012) y en diversos tipos de cáncer (Peng y Croce 2016), incluido el cáncer cervical y sus respectivos estadios preneoplásicos en pacientes infectadas con VPH (He et al. 2016).

Existe evidencia de que los miRNAs modulan la expresión de genes supresores de tumores relacionados con procesos carcinogénicos del VPH-AR como *TP53*, *CDKN1A* (p21) y *CDKN1B* (p27) (Beaudenon y Huibregtse 2008; Milde-Langosch et al. 2001), esas proteínas supresoras de tumores, en el panorama de infección por el VPH-AR, son degradadas por las oncoproteínas virales E6 y E7 del VPH-AR, siendo un paso clave para la desregulación del ciclo celular. La alteración en la expresión de miRNAs que modulan la expresión de genes supresores de tumores del ciclo celular se correlaciona con la expresión de E6/E7 *in-vitro*, sugiere que la desregulación de miRNAs por la acción de E6/E7 también podría mediar la carcinogénesis cervical (Harden et al. 2017; Harden y Munger 2017).

A pesar de estos avances, no se han identificado miRNAs celulares expresados diferencialmente con potencial biomarcador en muestras de lesiones preneoplásicas del cérvix VPH+ en términos de indicadores del desempeño diagnóstico, que evalúen la sensibilidad y especificidad para la detección de lesiones de alto grado de cérvix (Gocze et al. 2013; Li et al. 2011).

Los miRNAs representan un biomarcador candidato para uso clínico en el triaje de mujeres VPH-AR+ (es decir, para separar mujeres VPH-AR+ con mayor riesgo de desarrollar una lesión de alto grado o cáncer cervical, de mujeres VPH+ con menor riesgo). Aunque este campo apenas se está estudiando, varios esfuerzos se han centrado en identificar los miRNAs expresados diferencialmente en lesiones de alto grado versus lesiones de bajo grado del cérvix de mujeres VPH-AR+. Estudios más recientes aún, se han interesado en determinar el desempeño diagnóstico en términos de sensibilidad y especificidad de miRNAs candidatos (Babion et al. 2018, Babion et al. 2019; Ivanov et al. 2018; Snoek et al. 2019; Tian et al. 2014; Wang et al. 2014; Wilting et al. 2013a)

La principal perspectiva de llegar a implementar una prueba de triaje, en este caso, de miRNAs, es lograr disminuir el número de mujeres que son remitidas a colposcopia y, disminuir el número de visitas de mujeres al centro de salud para seguimiento. Lo que se traduce en la reducción de costos para el sistema de salud y un mejor trato al paciente, ya que, una colposcopia, es comúnmente incómoda y dolorosa para la mujer.

El presente trabajo pretende aportar entendimiento al vacío en el conocimiento de la prevención del cáncer cervical: **cómo identificar mujeres infectadas con VPH-AR que tienen mayor probabilidad de desarrollar lesiones neoplásicas de alto grado y cáncer cervical.**

6. Pregunta de Investigación e Hipótesis

Pregunta

¿Los miRNAs celulares que regulan las proteínas blanco de acción de las oncoproteínas E6 y E7 del VPH se expresan de manera diferente en lesiones intraepiteliales de alto grado versus lesiones intraepiteliales de bajo grado, en mujeres infectadas con el VPH-AR?

Hipótesis

Los miRNAs celulares que regulan las proteínas blanco de acción de las oncoproteínas E6 y E7 del VPH están expresados de manera diferente en lesiones intraepiteliales de alto grado versus lesiones intraepiteliales de bajo grado, en mujeres infectadas con el VPH-AR.

7. Objetivos

Objetivo general

Identificar miRNAs celulares reguladores de las proteínas blanco de acción de las oncoproteínas E6 y E7 del VPH, expresados de manera diferencial en lesiones intraepiteliales de alto grado versus lesiones intraepiteliales de bajo grado, en mujeres infectadas con el VPH-AR.

Objetivos específicos

- ✓ Los miRNAs celulares sobreexpresados en lesiones intraepiteliales de alto grado versus lesiones intraepiteliales de bajo grado del cérvix en mujeres infectadas con el VPH-AR, reportados en la literatura, ¿son posibles reguladores de proteínas blanco de acción de E6 y E7 del VPH-AR?

Objetivo específico.1: Predecir *in-sílico* los miRNAs reguladores de genes que codifican para proteínas blanco de acción de E6 y E7 del VPH-AR.

- ✓ ¿Cuáles miRNAs celulares tienen niveles de expresión altos en lesiones de alto grado versus lesiones de bajo grado del cérvix, en mujeres infectadas con el VPH-AR?

Objetivo específico. 2: Cuantificar los niveles de expresión de miRNAs celulares, en lesiones intraepiteliales de alto grado versus lesiones intraepiteliales de bajo grado del cérvix, en mujeres infectadas con el VPH-AR

- ✓ ¿Cuál es el potencial de los miRNAs con altos niveles de expresión en lesiones de alto grado versus bajo grado del cérvix, en la identificación de mujeres con lesiones de alto grado de cérvix, entre las mujeres infectadas con VPH-AR?

Objetivo específico. 3: Explorar el potencial de los miRNAs con altos niveles de expresión en lesiones de alto grado versus bajo grado del cérvix, para identificar lesiones de alto grado de cérvix, en mujeres infectadas con VPH-AR.

8. Materiales y Métodos

La metodología llevada a cabo para el cumplimiento del objetivo general del presente trabajo se describe de acuerdo con cada objetivo específico, en el diagrama presentado en la Figura 6.

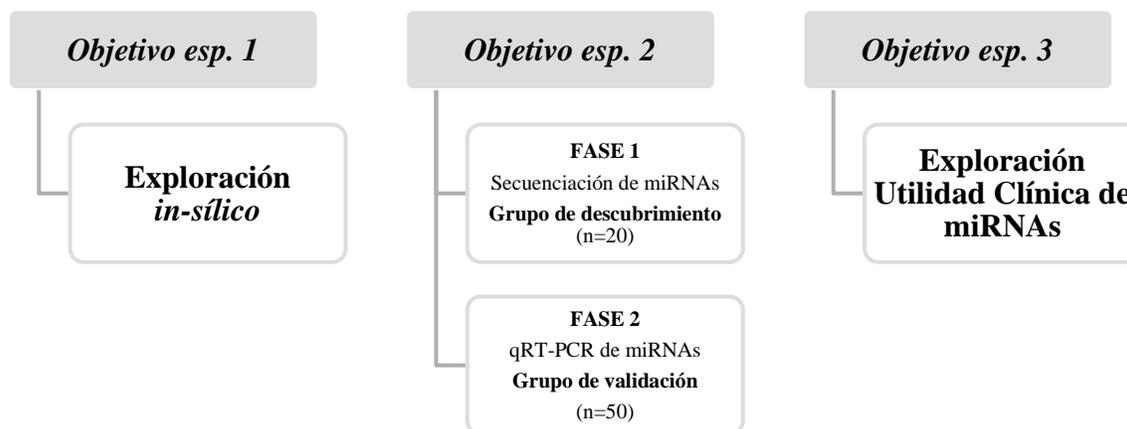


Figura 6. Esquema de la metodología para el cumplimiento de los objetivos. Para el cumplimiento del objetivo esp. 1 se hizo una exploración *in-silico* de miRNAs usando miRWalk2.0 y revisión de la literatura; para el objetivo esp. 2, se llevaron a cabo dos fases de estudio, la fase 1 consistió en la secuenciación de 20 muestras que forman el grupo de descubrimiento de miRNAs diferencialmente expresados en lesiones de alto grado versus lesiones de bajo grado del cérvix. La fase 2 consistió en la validación de cinco miRNAs en 50 muestras que forman el grupo de validación (independientes al grupo de descubrimiento) por qRT-PCR. Para el cumplimiento del objetivo esp. 3, la exploración de la utilidad clínica se llevó a cabo usando los valores CT normalizado de la fase 2 del objetivo esp. 2, a través del cálculo de la AUC a partir de curvas ROC.

Objetivo 1. Predecir *in-silico* los miRNAs reguladores de genes que codifican para proteínas blanco de acción de E6 y E7 del VPH-AR.

Predicción in-silico

Con el objetivo de explorar miRNAs como candidatos a ser utilizados como biomarcadores para la detección de lesiones de alto grado de cérvix en mujeres infectadas con VPH-AR, se hizo la predicción *in-silico* de los miRNAs reguladores de proteínas blanco de la acción de E6 y E7 del VPH usando el programa miRWalk2.0 (Dweep, Sticht, y Gretz 2013).

Como estrategia de predicción se emplearon los códigos EntrezID (National Center for Biotechnology Information, NCBI) de 19 proteínas diana de la acción de E6 y E7 de VPH-AR (Tabla 1); para hacer la predicción de miRNAs en el programa miRWalk2.0, el algoritmo TarPmiR Random Forest de miRWalk2.0 (Ding, Li, y Hu 2016; Dweep et al. 2011) se usó para reunir los resultados de la predicción de interacciones. Seguido de la predicción, se seleccionaron las

interacciones miRNA-mRNA reportadas previamente en la literatura, en el módulo de validación de miRWalk2.0.

Los parámetros de predicción de miRNAs fueron

- ✓ Mínima longitud de la región semilla: 7 nucleótidos
- ✓ Región de regulación en el transcrito de RNAm: 3' UTR
- ✓ Valor $p < 0,05$ de la predicción
- ✓ Posición de inicio de la región semilla en el miRNA: primera posición.

Tabla 1. Proteínas blanco de acción de las E6 y E7 del VPH-AR (White, Pagano, y Khalili 2014)

Oncoproteína viral	Gen	Entrez ID	Efecto biológico de la oncoproteína sobre la proteína	Referencia
E6	<i>DLG1</i>	1739	Induce pérdida de DLG1, causando desorden en la polaridad celular y descontrol en la proliferación.	(Bugnon Valdano et al. 2016)
	<i>E6AP = UBE3A</i>	7337	Degradación de p53	(Beaudenon y Huibregtse 2008)
	<i>EP300-P300</i>	2033	Inhibe la transcripción	(Patel et al. 1999)
	<i>IRF-3</i>	3661	Bloqueo de la transactivación de interferones, previniendo la inducción de IFN-B	(Ronco et al. 1998)
	<i>MAML1</i>	9794	Inhibición de la ruta de señalización Notch	(Meyers, Spangle, y Munger 2013)
	<i>PXN</i>	5829	Disrupción de la unión de actina al citoesqueleto celular	(Tong y Howley 1997)
	<i>SCRIB/HDIG</i>	23513	Disrupción de la polaridad celular	(Nakagawa y Huibregtse 2000)
	<i>SIPAL1=E6TP1</i>	26037	Degradación de E6TP1, provoca alteración en la proliferación.	(Gao et al. 1999)
	<i>TERT</i>	7015	Se activa la telomerasa	(Veldman, Horikawa, y Barrett 2001)
	<i>TP53</i>	7157	Es degradada p53 por acción directa sobre E6AP quien la ubiquitina en presencia de E6	(Beaudenon y Huibregtse 2008)
E7	<i>IGFBP-3</i>	3486	Inhibición en la señalización de factores de crecimiento tipo Insulina	(Mannhardt et al. 2000)
	<i>CDKN1B=p27</i>	1027	Se elimina la inhibición del crecimiento celular	(Milde-Langosch et al. 2001)
	<i>CDKN1A=p21</i>	1026	Se elimina la inhibición del crecimiento celular	(Milde-Langosch et al. 2001)
	<i>CUL2</i>	8453	Degradación de pRb y activación de E2F	(Huh et al. 2007)
	<i>E2F1</i>	1869	E2F, es liberado, promoviendo la proliferación celular.	(Zhang et al. 2015)
	<i>GAA</i>	2548	Activación alostérica de alfa-glucosidasa provocando agotamiento de glucógeno intracelular	(Zwerschke et al. 2000)
	<i>HDAC9</i>	9734	Estimula la transcripción	(Munger et al. 2004) (Longworth y Laimins 2004)
	<i>IRF-9= ISGF3</i>	10379	Perdida de transducción de señales mediada por INF-a, la infección evita el sistema inmune.	(Zimmermann et al. 1999)
<i>UBR4=p600</i>	23352	Al unirse con E7 contribuye al crecimiento independiente de anclaje y la transformación celular.	(Huh K. et al. 2005)	

Objetivo 2. Cuantificar los niveles de expresión de miRNAs celulares, en lesiones intraepiteliales de alto grado versus lesiones intraepiteliales de bajo grado del cérvix, en mujeres infectadas con el VPH-AR

Descripción de la Población

Las muestras usadas fueron biopsias/tejidos del cérvix fijados con formalina e incluidos en parafina (TFFIP) de participantes de los estudios *i-ii* del Grupo de Investigación Infección y Cáncer UdeA, y recuperadas directamente de laboratorios de patología. Estas biopsias fueron colectadas por ginecólogos colposcopistas en consultorios de colposcopia de la ciudad de Medellín-Colombia. Los dos estudios se describen a continuación:

- i. *Evaluación de estrategias para el manejo clínico óptimo de mujeres con células escamosas atípicas de significado indeterminado (ASCUS)*, financiado por COLCIENCIAS, es un estudio clínico pragmático aleatorizado de tres brazos sin cegamiento, de grupos paralelos. Las mujeres de 20 a 69 años (n = 2661) con una citología por primera vez ASCUS (células escamosas atípicas de significado indeterminado) fueron reclutadas en los servicios de tamización de rutina y asignadas al azar para recibir colposcopia inmediata (brazo CI; n = 882), repetir la citología a los 6 y 12 meses (brazo RC; n = 890) o una prueba de VPH (brazo de VPH; n = 889). El reclutamiento de mujeres se hizo entre los años 2011 y 2014 y las mujeres fueron seguidas con pruebas de VPH y citología 24 meses más tarde después del ingreso. El estudio finalizó el seguimiento en enero de 2016. Al menos 1410 mujeres tuvieron un diagnóstico histológico emitido por patólogos de la ciudad de Medellín y 756 fueron VPH positivas. Estos diagnósticos histológicos fueron releídos independientemente por dos patólogos expertos y externos al estudio. Entre las mujeres VPH positivas se identificaron 185 con lesiones de alto grado de cérvix (NIC2 o más).
- ii. *Evaluación de la utilidad clínica de proteínas reguladoras del ciclo celular y del patrón de metilación del genoma del VPH en la detección de lesiones preneoplásicas del cérvix*, financiado por El Comité para el Desarrollo de la Investigación CODI, es un estudio de corte transversal en el que se incluyeron mujeres entre 25 y 69 años (n = 1325), que asistieron a centros de colposcopia para un diagnóstico histopatológico, por presentar citología anormal (\geq ASCUS) o infección por **VPH-AR**, entre los años 2014

y 2016. El estatus de la infección se determinó por la prueba de Cobas® 4800 para la detección del VPH-AR la cual, detecta 14 genotipos de alto riesgo incluyendo VPH 16 y VPH 18. Las participantes firmaron consentimiento informado para estudios posteriores.

Ambos estudios fueron aprobados por el comité de bioética de la Sede de Investigación Universitaria (SIU) - Universidad de Antioquia (UdeA). Todas las mujeres firmaron un consentimiento informado el cual fue aprobado por el mismo comité de Bioética. El consentimiento informado contiene un resumen de los antecedentes de la propuesta, procedimientos del estudio, riesgos y beneficios, mantenimiento de la confidencialidad, uso de muestras e información de la persona de contacto y se resguardan bajo las respectivas normativas en las instalaciones del Grupo Infección y Cáncer. Ambos estudios se ajustan a las "Normas científicas, técnicas y administrativas para la investigación en Salud" establecidas en la resolución No 008430 de 1993 del Ministerio de Salud. El presente trabajo de investigación cuenta con el aval del comité de bioética SIU-UdeA. No. Acta 19-40-841

Para el cumplimiento del objetivo específico 2, se llevaron a cabo dos fases que se describen a continuación:

Fase 1: Secuenciación y análisis de expresión diferencial en 20 TFFIP VPH-AR+, este grupo se denotó como el “grupo de descubrimiento”, por ser las muestras usadas para la búsqueda ciega y masiva de miRNAs diferencialmente expresados entre lesiones de alto grado versus lesiones de bajo grado del cérvix. Este grupo estuvo compuesto por 10 muestras que pertenecen al proyecto i y 10 del proyecto ii. De estas 20, 10 tenían diagnósticos de lesiones intraepiteliales de alto grado (NIC2=7 & NIC3=3) y 10 eran de lesiones intraepiteliales de bajo grado o sin lesión (NIC1=9 & Negativo para lesión=1). Se identificaron miRNAs a través del análisis de expresión diferencial de datos obtenidos por secuenciación.

Fase 2: Validación mediante amplificación de miRNAs específicos (qRT-PCR), en un “grupo de validación” compuesto por 50 muestras VPH-AR+, **independientes al grupo de descubrimiento**. Las 50 muestras pertenecían al proyecto i, contaban con la relectura del diagnóstico patológico por un panel de expertos en patología cervical, 25 muestras tenían diagnósticos de lesiones intraepiteliales de alto grado (NIC3=23 & CEC=2) y 25 eran de lesiones de bajo grado o sin lesión

(NIC1=7 & Negativo para lesión=18). Se validaron 5 miRNAs diferencialmente expresados por PCR usando detección por SYBER™ Green y cebadores LNA (Qiagen®).

Las características sociodemográficas de las mujeres a las que pertenecen las muestras TFFIP del grupo de descubrimiento y del grupo de validación, son: edad, EPS, estado civil, nivel educativo, edad de la primera relación sexual, número de compañeros sexuales, número de hijos, año de uso de anticonceptivos y frecuencia de la citología. La prueba de Chi-Cuadrado se usó para comparar las características de la población de estudio, así mismo las Medianas y Rangos Intercuartílicos (RIC) para las variables continuas (edad, edad de la primera relación sexual, número de compañeros sexuales, número de hijos y años de uso de anticonceptivos). La Tabla 2 muestra las características sociodemográficas del total de la población incluida en el estudio (n=70).

Tabla 2. Descripción de la población de estudio

Característica	≤NIC1		NIC2+		Todos		valor p*
	n	%	n	%	n	%	
	35	(50)	35	(50)	70	(100)	
Edad (años)							
Mediana [RIC]	31	[25 - 43]	31	[27 - 37]	31	[26 - 38]	
≤30	17	(53,13)	15	(46,88)	32	(45,71)	0.8104
>30	18	(47,37)	20	(52,63)	38	(54,29)	
EPS							
COMFENALCO	3	(27,27)	8	(72,73)	11	(15,71)	0.3917
COMFAMA	15	(57,69)	11	(42,31)	26	(37,14)	
SURA	12	(50)	12	(50)	24	(34,29)	
SAVIA SALUD	5	(55,56)	4	(44,44)	9	(12,86)	
Estado Civil							
Casada/Unión Libre	17	(44,74)	21	(55,26)	38	(54,29)	0.6088
Divorciada/separada/viuda	2	(50)	2	(50)	4	(5,71)	
Soltera	16	(57,14)	12	(42,86)	28	(40)	
Nivel Educativo							
Técnica/tecnología/Profesional	9	(42,86)	12	(57,14)	21	(30)	0.7279
Secundaria completa	13	(52)	12	(48)	25	(35,71)	
Sin educación/P. I/P.C/S. I	13	(54,17)	11	(45,83)	24	(34,29)	
Edad de la primera relación sexual (años)							
Mediana [RIC]	16	[15 - 18]	16	[15 -18]	16	[15 - 18]	
≥ 18	12	(54,55)	10	(45,45)	22	(31,43)	0.7547
16 - 17	12	(44,44)	15	(55,56)	27	(38,57)	
≤ 15	11	(52,38)	10	(47,62)	21	(30)	
Número de parejas sexuales							
Mediana [RIC]	3	[2 - 5]	4	[3 - 7]	3	[2 - 6]	
1 - 3	22	(59,46)	15	(40,54)	37	(52,86)	0.1186
≥ 4	13	(40,63)	19	(59,38)	32	(45,71)	
No responde	0	(0)	1	(100)	1	(1,43)	
Número de hijos							
Mediana [RIC]	2	[1 - 2]	1	[1 - 2]	1	[1 - 2]	
Ninguno	8	(53,33)	7	(46,67)	15	(21,43)	0,7513
1 - 2	19	(46,34)	22	(53,66)	41	(58,57)	
≥ 3	8	(57,14)	6	(42,86)	14	(20)	
Consumo de Anticonceptivos (años)							
Mediana [RIC]	2,5	[0,45 - 8,75]	3,3	[0,5 - 7,1]	3	[0,4 - 7,3]	
Nunca	8	(17,78)	8	(17,78)	45	(64,29)	1
< 5	13	(16,05)	13	(16,05)	81	(115,71)	
≥5	14	(16,67)	14	(16,67)	84	(120)	
Frecuencia de la citología							
Una o más veces en un año	20	(54,05)	17	(45,95)	37	(52,86)	0.7656
Una vez cada 2-3 años	9	(45)	11	(55)	20	(28,57)	
Una vez cada 4-5 años	5	(45,45)	6	(54,55)	11	(15,71)	
No sabe/ No responde	1	(50)	1	(50)	2	(2,86)	

*Test Chi-cuadrado de Pearson, ≤NIC1: Neoplasia Intraepitelial Cervical grado 1 o bajo grado; NIC2+: Neoplasia Intraepitelial Cervical grado 2 o más o alto grado; RIC: Rango Intercuartílico, EPS: Empresas de Salud Pública. P.I: Primaria Incompleta, P.C: Primaria Completa, S.I: Secundaria Incompleta.

Tabla 3. Número de muestras en cada conjunto de experimentos

Diagnóstico Histopatológico	Grupo Descubrimiento		Grupo Validación	
	n	%	n	%
	20	100	50	100
NEG	1	5	18	36
NIC1	9	45	7	14
NIC2	7	35	0	0
NIC3	3	15	23	46
CEC	0	0	2	4

la Tabla 3 muestra el número y distribución de muestras de acuerdo con el grupo de estudio (descubrimiento/ validación) y al diagnóstico histopatológico. Todas las muestras fueron seleccionadas por muestreo aleatorizado simple del total muestras VPH+ disponibles, y procesadas en ciego hasta el análisis de los datos. El análisis descriptivo se llevó a cabo en el programa R. v3.6.1. Las categorías para cada una de las variables NO INFLUYE en el diagnóstico histopatológico de las muestras seleccionadas para este estudio, evidenciado por un valor $p > 0,05$ en la Tabla 2.

Obtención de RNA a partir de TFFIP

Para la extracción del RNA a partir de tejidos fijados con formalina (formaldehído al 37%) e incluidos en parafina, se debe considerar que los largos periodos de tiempo en que permanecen los tejidos pueden provocar la transformación de Formaldehído en Ácido Fórmico, el cual produce entrecruzamiento (“cross-linking”) entre ácidos nucleicos y proteínas (RNA-RNA y RNA-proteína) (Feldman 1973); hidroliza puentes fosfodiéster del DNA (Douglas y Rogers 1998) y modifica covalentemente el RNA por adición de grupos metilos en las bases nitrogenadas (Feldman 1973; Masuda et al. 1999). Todo lo anterior, genera problemas en la obtención de material genético óptimo para realizar estudios moleculares posteriores. No obstante, se ha demostrado que el tratamiento con proteinasas mejora la recuperación del DNA/RNA, ya que permite liberar los ácidos nucleicos atrapados en los complejos proteicos. Además, se ha mostrado que el entrecruzamiento induce la inactivación de proteínas celulares (ribonucleasas), promoviendo la conservación, estabilidad e integridad de los miRNAs, garantizando la extracción de estas moléculas (Díaz-Cano y Brady 1997). Las muestras de TFFIP usadas en este trabajo oscilan entre 2 a 7 años de almacenamiento a temperatura ambiente.

Preparación de las muestras TFFIP

Los TFFIP fueron recuperadas directamente de los laboratorios de patología de la ciudad de Medellín donde fueron atendidas las participantes de los proyectos i-ii. Un total de 70 biopsias (20 del grupo de descubrimiento y 50 del grupo de validación), fueron incluidas en casetes, se eliminó el exceso de parafina y se pulieron los bordes cuidando de no resquebrajar el bloque ni cortar la muestra. Con un micrótopo y usando una cuchilla nueva se colectaron 4 secciones de 10µm de grosor por cada bloque en tubos libres de RNasas de 1,5 mL y almacenadas a -20°C hasta su procesamiento.

Extracción de RNA total

Se usó el kit de Qiagen miRNeasy para TFFIP, el cual proporciona condiciones optimizadas de lisis e incubación para revertir el efecto de la formalina, y purifica el RNA total, incluyendo miRNAs a partir de 18 nucleótidos. Los pasos para la obtención del RNA total fueron:

- I. Desparafinización: La parafina se eliminó de las secciones de TFFIP al agregar Xileno a cada muestra (4 secciones de 10 µm con un área superficial de máx. 250 mm² cada una) seguido de lavado con etanol para extraer el exceso de Xileno.
- II. Lisis: Se incubó con tampón de lisis de la compañía optimizado para la digestión con Proteinasa K, permitiendo la liberación de los ácidos nucleicos de las secciones de tejido evitando la degradación del RNA.
- III. Revertir el entrecruzamiento (“cross-linking”): Dos pasos de desnaturalización/incubación a 56 °C y 80 °C por 15 min cada uno, se llevaron a cabo para revertir el entrecruzamiento de ácidos nucleicos y proteínas producto de la fijación con la formalina, mejorando así el rendimiento y la calidad del RNA.
- IV. Tratamiento con DNasas por 15 min optimizado para eliminar el DNA genómico, incluidos fragmentos de DNA pequeños que a menudo se presentan en biopsias expuestas a fijación prolongada con formalina y/o largos tiempos de almacenamiento.
- V. Unión del RNA a la columna: Se adicionó un tampón de sal de la compañía que contiene Clorhidrato de Guanidina, para acondicionar la unión de ácidos nucleicos a la columna de separación RNeasy MinElute, donde el RNA total se une a la membrana de sílica en presencia de la sal de Guanidina, y los demás componentes celulares se eliminan en el

sobrenadante. Las columnas RNeasy MinElute tienen un 80% de eficiencia de retención (Kemp et al. 2015).

- VI. Lavados: Se adicionó dos veces etanol a las columnas RNeasy MinElute que contienen el RNA total, seguido de un tampón de lavado de la compañía que elimina los rastros de sales de la columna.
- VII. Elución: Se adicionó 20µL de H₂O libre de RNasas. Los RNAs fueron almacenados a -80°C hasta su procesamiento.

Un microlitro (1µL) de la muestra se usó para medir la concentración del RNA total extraído usando el kit Qubit™ RNA HS Assay (Invitrogen™) en el Fluorómetro Qubit 3.0 (Thermo). Este kit incluye dos estándares (0 ng/µl–500 ng/µl) y un reactivo fluorescente (confidencial de la compañía) que reconoce el ácido nucleico a cuantificar.

Secuenciación de miRNAs – smallRNAseq

Preparación de librerías de miRNA

Se obtuvieron las librerías de miRNAs de 20 muestras usando el QIAseq miRNA Library kit de Qiagen (Figura 7). Los adaptadores se ligan secuencialmente a los extremos 3' y 5' de los miRNAs. Posteriormente, se realiza la síntesis universal de DNAc con la asignación de un índice molecular único (UMI)* (Figura 8). Seguido de un paso de limpieza de DNAc, amplificación del DNAc para la obtención de las librerías y finalmente la limpieza de librerías. El flujo de trabajo se explica a continuación:

- ✓ Ligadura al extremo 3': Un adaptador de DNA preadenilado se liga al extremo 3' de todos los miRNAs. Mediante la ligasa QIAseq miRNA NGS3 (3'AACTGTAGGCACCATCAAT).
- ✓ Ligadura del extremo 5': Un adaptador de RNA está ligado al extremo 5' de miRNAs maduros (5' GUUCAGAGUUCUACAGUCCGACGAUC).
- ✓ Síntesis de DNAc: El cebador de transcripción inversa (RT) contiene una UMI integrada y se une a una región del adaptador 3' y facilita la conversión de los miRNA ligados 3'-5' en DNAc, al tiempo se asigna una UMI (Índice Molecular Único) a cada molécula de miRNA. Durante la transcripción inversa, también se agrega una secuencia universal (barra verde

figura 7) que es reconocida por los cebadores de indexación de la muestra durante la amplificación de la librería.

- ✓ Limpieza DNAC: Después de la transcripción inversa, la limpieza del DNAC se hace utilizando perlas magnéticas.
- ✓ Amplificación de la librería: Se logra utilizando un cebador directo universal unido a un índice personalizado único, que es asignado diferencialmente a cada una de las muestras.
- ✓ Limpieza de la librería: después de la amplificación de la librería, la limpieza se hace utilizando perlas magnéticas.

*Los UMI (Índice Molecular Único) son secuencias aleatorias de 15 nt que permiten cuantificar con precisión las moléculas miRNAs después de un protocolo de secuenciación. La cuantificación basada en UMI refleja la cantidad original de moléculas de miRNAs, puesto que al unirse las UMIs a los miRNAs maduros, las moléculas de miRNA son contadas individualmente, y de esta forma se evita la secuenciación de la misma molécula de miRNA una y otra vez resultando en la sobreestimación de la expresión de miRNA (Obtenido de Qiagen).

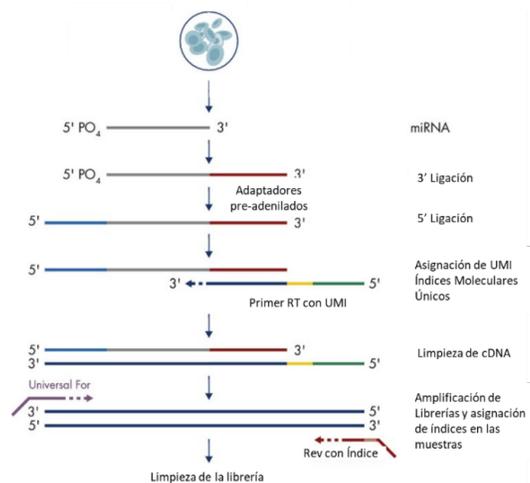


Figura 7. Preparación de la biblioteca de secuenciación de miRNA. Utilizando el QIAseq miRNA Library kit de Qiagen. Los adaptadores 3' y 5' están diseñados para ligarse a miRNA maduros. Los miRNA ligados se transcriben inversamente a DNAC utilizando un cebador de transcripción inversa (RT) con una UMI. Después de la limpieza de DNAC, la amplificación de las librerías se logra con un cebador “Forward” universal y cebadores “Reverse” de indexación. Después de una limpieza final, las librerías están listas para el control de calidad y NGS posteriores. A partir de RNA total aislado, se llevó a cabo el flujo de trabajo del kit de biblioteca QIAseq miRNA. Tomado y adaptado de Qiagen.

P5 capture	Adaptador 5' 19 nt	miRNA maduro ≈14 nt	Adaptador 3' 26 nt	UMI 12 nt	Índice de muestra 4 nt	P7 capture
------------	-----------------------	------------------------	-----------------------	--------------	---------------------------	------------

Figura 8. Representación de las secuencias de las librerías a ser secuenciadas. P5 y P7 capture, son dos oligo-secuencias de Illumina que permiten la unión a las celdas de flujo. Los adaptadores 3' y 5' tienen un tamaño de 19 nt y 26 nt respectivamente, los UMI de 12 nt, los índices de muestras de 4 nt. Por lo cual las librerías derivadas del kit QIAseq miRNA Library requieren secuenciar lecturas de una longitud de 75pb

Control de calidad de librerías pre-secuenciación

Para determinar el tamaño y calidad de las librerías de miRNAs se usó 1 μ L de cada librería, en un chip High Sensitivity DNA en el equipo Agilent 2100 Bioanalyzer (Agilent Technologies). El tamaño de una librería de miRNA se evidenció como un pico con una intensidad de fluorescencia alrededor de los 300 FU y un tamaño aproximado de 180 pb en el electroferograma. Dos (2 μ L) de cada librería fueron usados para determinar la concentración usando el kit Qubit™ dsDNA HS Assay (Invitrogen™), en el Fluorómetro Qubit 3.0 (Thermo).

Secuenciación

El QIAseq miRNA NGS Kit de Qiagen se usó para la secuenciación de las librerías diluidas a 4 nM. Se hizo 2 mezclas de 10 librerías cada uno, dependiendo de los UMIs, en cantidades equimolares. Las dos mezclas se prepararon y se secuenciaron en un MiSeq (Illumina) a 1 x 75 bp en un solo extremo (single-end) siguiendo las indicaciones del inserto *MiSeq System Denature and Dilute Libraries Guide*.

Obtención del cDNA y qRT-PCR

Síntesis de cDNA de miRNAs

Se usó el Kit miRCURY LNA RT de Qiagen, cuyo principio se basa en agregar una cola de PoliA en el extremo 3' de las secuencias de los miRNAs maduros (Figura 9, Paso1-A). Luego, el DNAC se sintetiza con un primer de PoliT que tiene un OH-3' libre para ser reconocido por la polimerasa y además tiene unida una etiqueta universal en el extremo 5' la cual es una secuencia específica de nucleótidos para hacer más extensa la secuencia permitiendo así que la transcriptasa reversa sintetice la hebra de DNAC (Figura 9, Paso1-B). El volumen de la reacción fue de 10 μ L, contenía: 2 μ L de 5X miRCURY RT Reaction Buffer (el primer RT y los dNTPs), 1 μ L de 10X miRCURY

RT Enzyme mix (PolyA polimerasa y ReversoTranscriptasa), 5 μ L de H₂O libre de RNAsas y 2 μ L de RNA total extraído a 5ng/ μ L. Las condiciones de la reacción fueron 60 min a 42°C (reacción RT), 5 min a 95°C (inactivación de la reacción) y 4°C. El DNAc se almacenó a 20°C hasta su uso.

Detección de miRNAs

Para la detección de miR-1, miR-2, miR-3, miR-4 y miR-5 se usó un set de cebadores (forward/reverse) de miRCURY LNA miRNA PCR Assay de Qiagen (ver Figura 9-Paso 2-A-B y Figura 10)

Para la amplificación se utilizó el miRCURY LNA SYBR® Green PCR Kit de Qiagen y como gen normalizador se usó el *SNORD44*. Para cada muestra el marcador fue medido por duplicado. El volumen final de la reacción fue de 10 μ L, así: 5 μ L de 2X miRCURY SYBR® Green Master Mix, 1 μ L de PCR Primer Mix, 1 μ L H₂O libre de RNAsas y 3 μ L de una dilución 1:60 del DNAc. Las condiciones de la reacción fueron 2 min a 95°C, seguido 40 ciclos de 10 s a 95°C (desnaturalización) y 60 s a 56°C (alineación/extensión). Los resultados con CT>39 fueron considerados negativos.

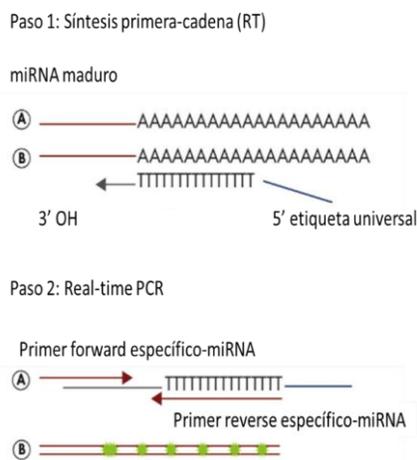


Figura 9. El sistema miRCURY LNA miRNA PCR. En el paso 1 utiliza una única reacción UNIVERSAL de síntesis de DNAc para todas las amplificaciones, reduciendo el pipeteo, ahorrando tiempo y muestras. En el Paso 2. Dos cebadores PCR específicos de miRNA mejorados con LNA permiten una amplificación altamente específica y sensible, incluso para la discriminación a nivel de un solo nucleótido. Tomado y adaptado de Qiagen

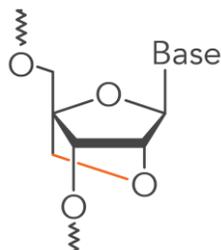


Figura 10. El Nucleótido LNA tiene el anillo de ribosa conectado por un puente de metileno entre los átomos de 2'-O y 4'-C, "bloqueando" el anillo de ribosa en la conformación ideal para la unión de Watson-Crick. Esta modificación ofrece una afinidad mayor por su cadena complementaria, en comparación con los oligonucleótidos de DNA o RNA tradicionales. Esto otorga mayor sensibilidad y especificidad siendo ideales para la detección diferencial de miRNAs, incluso si varían en un solo nucleótido.

Manejo de datos de secuenciación

Para el manejo adecuado de datos de secuenciación de RNA, se consideraron las recomendaciones del sitio web de la Universidad de Oregon (<https://rnaseq.uoregon.edu>). Se usaron 10 réplicas biológicas (10 muestras en cada grupo de comparación), sin embargo, no se hicieron réplicas técnicas. Para disminuir falsos positivos se llevó a cabo primero un paso de filtración, que consistió en descartar aquellos genes para los que se encontraron muy pocas lecturas o no se alineó ninguna lectura, el criterio fue eliminar las filas de la matriz de conteos (los miRNAs) que tuvieran cero en 3 o más muestras. También se hizo la normalización de los datos con el objetivo de controlar los efectos de la variación técnica y convertir las librerías de cada grupo de comparación a una misma escala, ya que una librería con mayor profundidad de secuenciación tiene más probabilidad de tener genes diferencialmente expresados respecto a otra, sin deberse estas diferencias a la condición bajo estudio.

Análisis de expresión diferencial

Datos de secuenciación

Para obtener los perfiles de expresión de miRNAs, se llevó a cabo secuenciación de RNAs pequeños y los datos crudos en formato .fastaq fueron procesados en el programa Data Analysis Center de Qiagen – DAC (<https://geneglobe.qiagen.com/us/analyze/>), siguiendo los protocolos QIAseq miRNA Library kit Primary Quantification y QIAseq miRNA Library kit –Secondary

Analysis (<https://www.qiagen.com/us/resources/download.aspx?id=bea2dcfa-0a5c-47c5-afd8-8b0fe90a471a&lang=en>) y (https://dataanalysis.qiagen.com/QIASeqmiRNA/documents/QIAseq_miRNA_Library_Kit_Secondary_Data_Analysis_Handbook_v1.1.pdf).

El protocolo QIAseq miRNA Library kit Primary Quantification consistió en el despojamiento de los adaptadores (Trim adapter) 3' AACTGTAGGCACCATCAAT (19nt) y 5' GTTCAGAGTTCTACAGTCCGACGATC (26nt) usando cutadapt (cutadapt.readthedocs.io/en/stable/guide.html), seguido de la identificación de secuencias insertadas y de las UMI, en donde las lecturas con menos de 16 pb o secuencias UMI de menos de 10pb fueron descartadas. Finalmente se llevó a cabo una estrategia de alineamiento/mapeo secuencial a distintas bases de datos usando el alineador Bowtie (bowtie-bio.sourceforge.net/index.shtml), si el ajuste era perfecto se mapeó a la base de datos de miRNAs maduros, si no, a la base de datos miRBase hairpin, seguido de RNA no codificantes (piRNA, rRNA, tRNA y otrosRNA) y luego RNAm y nuevamente a miRNA maduros tolerando como mínimo dos desajustes (“mismatches”). Para cada paso, solamente las secuencias no asignadas pasaron al siguiente paso. La base de datos de miRBase V.21 se usó para miRNA maduros, la piRNABank se usó para piRNA y los otros RNA no codificantes y RNAm se mapearon al Genome Reference Consortium GRCm38. Finalmente se obtuvo la tabla de conteos de lecturas de cada miRNAs para cada muestra.

El protocolo QIAseq miRNA Library kit –Secondary Analysis, se llevó a cabo para normalizar los conteos crudos por el algoritmo DESeq2, es decir, usando el método LogGeometricMean (LGM), que da cuenta de la composición de miRNA en cada muestra y utiliza un factor de escala para convertir a la misma escala los recuentos de UMI en todas las muestras. El factor de escala de cada muestra se calculó como la mediana de las razones de conteos observados con la media geométrica de cada miRNA correspondiente en todas las muestras (Love, Huber, y Anders 2014).

El análisis de expresión diferencial entre lesiones de alto grado (NIC2=7, NIC3=3) y lesiones de bajo grado (NIC1=9, NEG=1), se hizo desde dos programas diferentes, pero usando el mismo algoritmo DESeq2 de Bioconductor. Uno de los programas fue R V3.6.1 corriendo directamente el script de DESeq2 y el otro fue el Data Analysis Center (DAC) de Qiagen (<https://geneglobe.qiagen.com/us/analyze/>). En el programa R se consideraron miRNA como

diferencialmente expresados aquellos con valor p ajustado $<0,05$ (calculado a partir de la prueba de Wald y ajustado por el método de Benjamini-Hochberg (FDR)) y con DAC se consideraron miRNA diferencialmente expresados aquellos con valor $p <0,05$ (calculado a partir de la prueba de Wald).

Dado el gran número de comparaciones (una comparación por cada gen) necesarias para llevar a cabo un análisis de expresión diferencial de datos de RNAseq, se debe aplicar reiteradamente una prueba de comparación de medias simple, tal como lo es la prueba de Wald (Love, Huber, y Anders 2014) utilizado aquí, para un nivel de significancia de 0,05, conduciendo a un número grande de rechazos de la hipótesis nula, aunque no existan diferencias reales. Por tal razón, en la determinación de la diferencia significativa, siempre se verificó que las diferencias halladas no se debían a la variabilidad intrínseca de los datos, mediante el uso de herramientas estadísticas que permiten controlar la tasa de falsos descubrimientos cuando se procesan datos genómicos tal como la corrección por FDR (Benjamini, Yoav ; Hochberg 1995).

Datos de qRT-PCR

EL CT para cada muestra se obtuvo como el promedio entre los CT de las reacciones por duplicado del gen evaluado en la misma muestra. El Fold Change (FC) se determinó a partir del cálculo del Leevak $2^{-\Delta\Delta CT}$ (Livak y Schmittgen 2001). Para esto se calculó el $\Delta CT = CT_{\text{gen}_x_y} - CT_{\text{norm}_y}$ (El ΔCT es la diferencia entre CT del gen de interés x en la muestra y , y el CT del gen normalizador en la muestra y), luego se calculó el $\Delta\Delta CT = \Delta CT_{\text{casos}} - \Delta CT_{\text{controles}}$ (Es la diferencia entre el ΔCT del gen de interés x en la muestra, y el promedio del ΔCT del gen de interés x en todos los controles, para cada muestra). El promedio ΔCT del gen de interés x en todos los controles varía de acuerdo con el grupo que se seleccione como control para hacer la comparación deseada.

Los valores de FC fueron transformados a la función $\text{Log}_2(\text{FC})$, dada la alta variabilidad de los datos, con el objetivo de representar gráficamente mejor y aplicar las pruebas estadísticas. Se probó normalidad y homocedasticidad con las pruebas Shapiro Wilk y Brown-Forsythe, respectivamente. Se usó la prueba de U de Man Whitney para la comparación de medianas entre dos grupos, Kruskal Wallis para más de dos grupos seguido de la post-prueba de Dunn para múltiples comparaciones.

Se consideraron diferencias significativamente estadísticas para valores de $p < 0.05$. Para los análisis se usó el programa estadístico GraphPadPrism 7 (v.7.02).

Objetivo 3. Explorar el potencial de los miRNAs con altos niveles de expresión en lesiones de alto grado versus bajo grado del cérvix, para identificar lesiones de alto grado de cérvix, en mujeres infectadas con VPH-AR.

Análisis de curvas ROC

Se hizo una aproximación y se exploró la posible utilidad diagnóstica de los cinco miRNAs y cuatro combinaciones para detectar NIC2+ y NIC3, mediante la construcción de curvas ROC utilizando los valores de ΔCT ($\Delta CT_y = CT_{miRNAx, y} - CT_{SNORD44y}$). Los valores p fueron determinados por una prueba de distribución estándar, 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%.

La prueba estándar de oro para las lesiones cervicales es la biopsia dirigida por colposcopia, seguida de un análisis histopatológico. La colposcopia es un examen del cuello uterino que utiliza un lente de aumento e iluminación brillante para identificar pistas sugestivas de tejido anormal. Las biopsias pueden ser tomado de áreas que parecen anormales, o si las anomalías no son evidentes, una técnica de muestreo puede ser usada (ACCP 2011)

Para la definición del diagnóstico, el índice y la distribución de células en proliferación a través del epitelio estratificado del cérvix, y la experticia del patólogo juegan un papel clave para determinar correctamente el grado de lesión cervical mediante la lectura de placas con cortes del tejido teñidos con Eosina y Hematoxilina (Stoler et al. 2015). Para las muestras del grupo de validación el diagnóstico histopatológico se obtuvo por la relectura de las placas por un panel de patólogos expertos en cérvix y para las muestras del grupo de descubrimiento, el diagnóstico de 10 de ellas fue dado por el mismo panel de expertos, y las otras 10 por patólogos de los centros de colposcopia de la ciudad de Medellín.

Se construyeron curvas para cada miRNA individualmente: miR-1, miR-2, miR-3, miR-4 y miR-5 y cuatro combinaciones propuestas, que se explican a continuación:

1. miR-1, miR-2, miR-3: Fueron los tres miRNAs que presentaron mayor FC por smallRNAseq y en la validación por qRT-PCR
2. miR-4 y miR-5: Fueron los dos miRNAs que presentaron menor variabilidad en su expresión por smallRNAseq y en la validación por qRT-PCR
3. miR-1 y miR-3: Fueron los dos miRNAs que presentaron mayor AUC independientemente.
4. miR-1, miR-2, miR-3, miR-4 y miR-5: La combinación de los cinco miRNAs

Las curvas ROC y los AUC fueron obtenidos usando el programa estadístico GraphPadPrism 7 (v.7.02).

9. Resultados

Identificación de miRNAs por la estrategia de búsqueda in-silico

La aproximación *in-silico* consistió en dos etapas, esquematizadas en la Figura 11. La primera etapa (región verde) de predicción *in-silico*, utilizo los códigos ID de los genes que codifican para 19 proteínas dianas de E6 y E7 (ver tabla 1) en el módulo de predicción del programa miRWalk2.0, e identificó 2515 miRNAs con potencial de unirse a transcritos de genes que codifican para proteínas dianas de E6 (*DLG1*, *E6AP*, *EP300*, *IRF-3*, *MAML1*, *PXN*, *SCRIB/HDIG*, *SIPAL1*, *TERT* y *TP53(p53)*) (White, Pagano, y Khalili 2014) (ver tabla 1), y también identificó 2468 con potencial de unirse a transcritos de genes que codifican para proteínas dianas de E7 (*IGFBP-3*, *CDKN1B(p27)*, *CDKN1A(p21)*, *CUL2*, *E2F1*, *GAA*, *HDAC9*, *IRF-9* y *UBR4(p600)*) (White, Pagano, y Khalili 2014) (ver tabla 1). Una segunda etapa (región naranja) consistió en usar el módulo de validaciones del programa miRWalk2.0, en donde es posible conocer si la interacción predicha ha sido reportada previamente en la literatura. Se encontraron 190 y 319 interacciones validadas de miRNAs que hacen blanco en los 19 transcritos cuyos genes codifican para proteínas dianas de E6 y E7 respectivamente (ver tabla 1). En total se encontraron 454 miRNAs por la estrategia de búsqueda *in-silico*, de los cuales 55 miRNAs regulan tanto dianas de E6 como dianas

de E7, tal como lo indica el diagrama de venn en la región morada de la Figura 11. Los 454 miRNAs fueron seleccionados como potenciales candidatos a interactuar con las proteínas E6 y E7 del VPH.

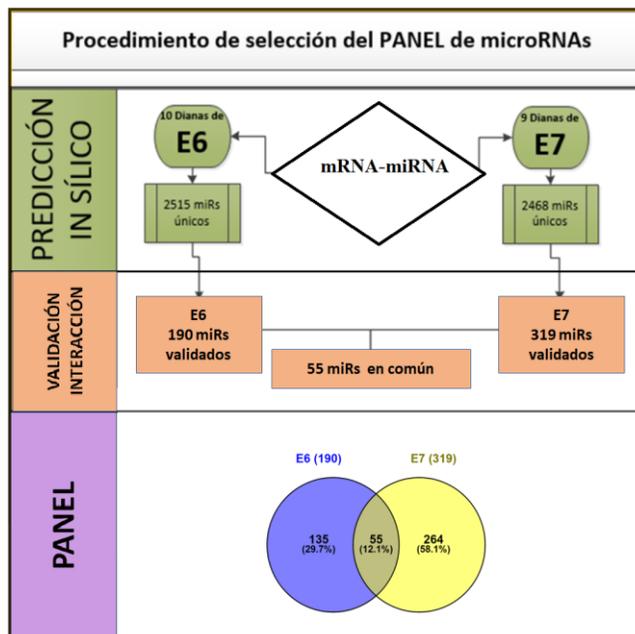
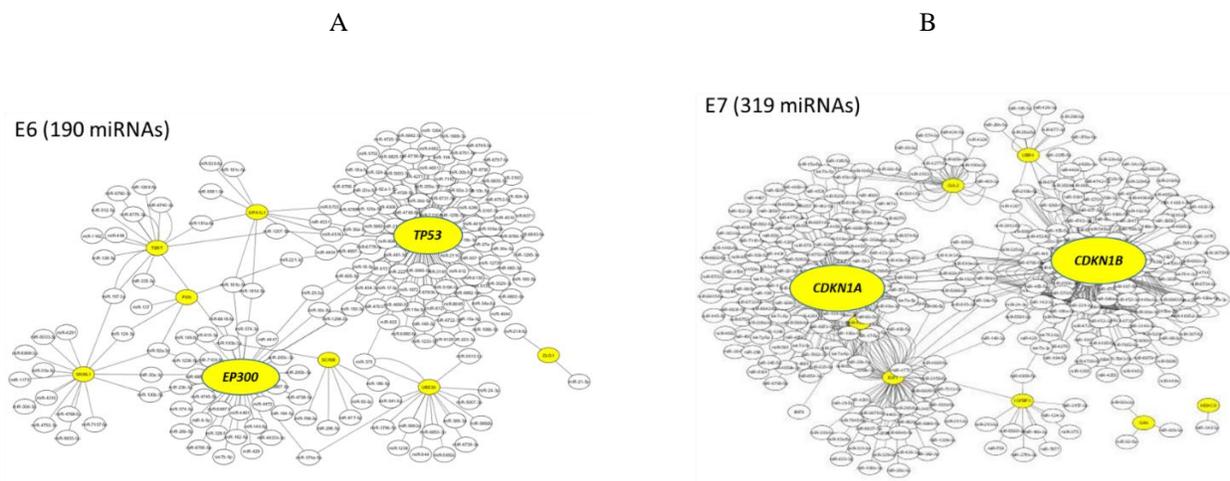


Figura 11. Estrategia *in-silico* para la selección de miRNAs

La red de interacciones entre los 454 miRNAs candidatos (encontrados por la estrategia de búsqueda *in-silico* explicada en la figura 11) con sus correspondientes RNA mensajeros diana, que a su vez son blanco de acción de las oncoproteínas E6 y E7, se presentan en el Suplemento 1A y 1B. Estas figuras muestran las interacciones miRNA-RNA halladas en la estrategia de predicción del objetivo específico número 1, los miRNAs en óvalos blancos y los nodos de regulación o proteínas dianas de E6 y E7, en óvalos amarillo. Los nodos de concentración más regulados fueron las dianas *TP53*(p53) y *EP300*(p300) de E6 y *CDKN1A*(p27) y *CDKN1B*(p21) de E7, Inhibidores de los Iniciadores del ciclo celular las quinasas dependientes de ciclinas (CDKs).



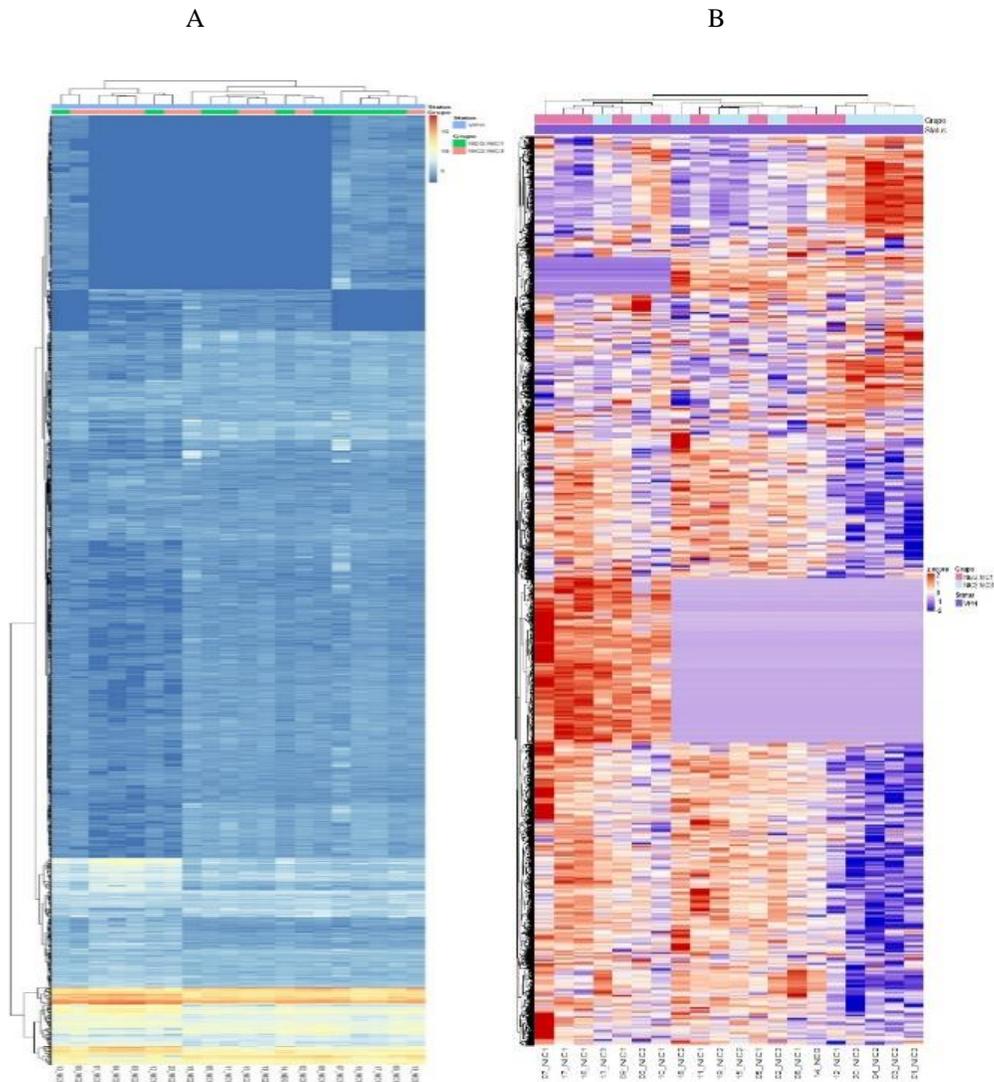
Suplemento 1. Red de interacciones entre los 454 miRNAs candidatos (propuestos por la estrategia de búsqueda *in-silico* explicada en la Figura 10) y sus correspondientes blancos, que son a su vez blanco de acción de las oncoproteínas E6 (A (190 miRNAs) y E7 (B (319 miRNAs)). Los óvalos en amarillo representan los RNA mensajeros dianas de los miRNA en óvalos blancos. Elaboración propia en Cytoscape V.3.7.1.

Alineamiento de miRNAs y preprocesamiento de los datos de secuenciación

Con el fin de identificar miRNAs expresados diferencialmente en lesiones de alto grado (NIC2 = 7; NIC3=3) versus lesiones de bajo grado (NIC1=9 y NEG=1) de cervix, se llevó a cabo la secuenciación de RNAs pequeños en 20 muestras de TFFIP todas VPH-AR+, llamado también el grupo de descubrimiento. El rendimiento de la secuenciación y el alineamiento/mapeo al miRnoma de referencia (miRBase V.21) y a las bases de datos descritas en Materiales y Métodos (MyM) se resume en la Tabla 4. En promedio se obtuvieron aproximadamente 2,8 millones de lecturas en cada muestra, más de 1,2 millones de lecturas en promedio (el 40% del total de lecturas) mapearon a secuencias de miRNAs maduras, las lecturas mapeadas correspondieron a 2186 miRNAs maduros. El resto de las lecturas mapearon a otras bases de datos sin interés para el presente trabajo, no obstante, se muestran en la Tabla 4.

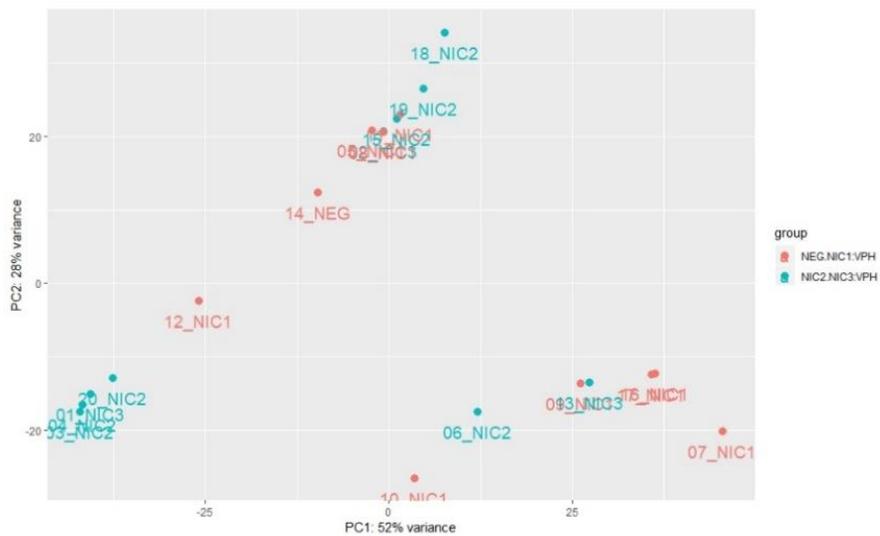
Tabla 4. Resumen del alineamiento de las lecturas secuenciadas.	
Lecturas	Promedio
Total	2.756.137,80
Adaptadores	49.696,40
Cortas <16bt	776.326,90
UMIs defectuosos <10pb	15.074,40
miRNAs maduros	1.278.746,10
miRNAs hairpin	1.481,80
piRNAs	30.443,40
rRNA	157.494,70
tRNA	99.604,10
mRNA	12.547,80
otros RNA	52.749,60
Mapeado no caracterizado	88.189,50
No mapeado no caracterizado	193.783,30
hairpin: estructura en bucle de los miRNAs no maduros. piRNAs: RNAs piwis, rRNA: RNA ribosomal, tRNA: RNA de transcripción, mRNA: RNA mensajero	

En total se mapearon 2186 miRNAs, la variabilidad de los conteos de lecturas normalizados para cada miRNA mapeado en las 20 muestras, se presenta en un mapa de calor en el Suplemento 2. En la parte inferior del mapa de calor, se observa en colores cálidos un conjunto de miRNAs con alta expresión en algunas de las muestras, pero al mismo tiempo con colores fríos en otras muestras, lo que indica que posiblemente estos son los miRNAs diferencialmente expresados.



Suplemento 2. Variabilidad entre los conteos de lecturas normalizados para cada miRNA (n=2186) en cada una de las 20 muestras. A. La agrupación color verde hace referencia a los NIC1 y NEG y la agrupación color rosado a los NIC2 y NIC3, los valores van entre 5 y 15 lecturas normalizadas. **B.** Transformación a la puntuación Z (z-score) de los conteos de lecturas normalizadas, la agrupación color rosado hace referencia a los NIC1 y NEG y la agrupación color azul a los NIC2 y NIC3, los valores de z van entre -2 y 2, con una escala creciente desde colores azul oscuro hasta color rojo oscuro. Gráficos realizados en R. V.3.6.2. (A y B fueron obtenidos con diferentes paquetes de R)

Para dilucidar mejor los resultados de la secuenciación es pertinente hacer un análisis no supervisado y evaluar la distribución de los conteos de las lecturas normalizadas, entre las muestras, para esto el Suplemento 3, presenta el Análisis de Componentes Principales que incluye todos los miRNAs mapeados (2186).



Suplemento 3. Análisis de componentes principales. PC1 y PC2 componentes principales 1 y 2 explican la variabilidad en un 52% y 29%, respectivamente, de la expresión de todos los miRNAs mapeados, en cada muestra. Los puntos azules hacen referencia a las muestras NIC2 y NIC3 y los puntos color salmón a las muestras NEG y NIC1.

Aunque la agrupación por lesiones de alto o bajo grado no está claramente reflejada en los componentes principales, es posible identificar pequeños agrupamientos dispersos en el plano, entre muestras de lesiones de alto grado y entre lesiones de bajo grado.

Análisis de expresión diferencial de los datos de secuenciación

El análisis de expresión diferencial entre los grupos de muestras NIC2+ vs. \leq NIC1, se llevó a cabo siguiendo el flujo de trabajo del paquete DESeq2 de Bioconductor, este se realizó en el programa gratuito Data Analysis Center (DAC) de Qiagen y los resultados se confirmaron independientemente en el programa R (v3.6.3), llevando a cabo el mismo flujo de trabajo DESeq2. La prueba estadística utilizada por DESeq2 para el análisis de expresión diferencial es la prueba de Wald para comparaciones múltiples, se consideraron miRNAs diferencialmente expresados entre ambos grupos de comparación aquellos que se presentaron valores p ajustados menores de 0.05 (Benjamini-Hochberg FDR).

El programa R identificó 25 miRNAs expresados diferencialmente, los cuales corresponden al Top 25 de miRNAs (de un total de 162) encontrados diferencialmente expresados en el programa DAC (ver figura 12). Estos resultados sugieren una alta reproducibilidad del flujo de trabajo DESeq2 de Bioconductor entre ambos programas.

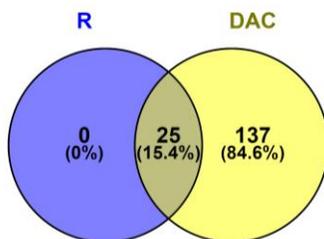


Figura 12. Número y porcentaje de los miRNA diferencialmente expresados por secuenciación entre lesiones de alto y bajo grado del cérvix utilizando los programas R (círculo azul) y DAC (círculo amarillo).

Contrastando los miRNA encontrados en el análisis de expresión diferencial a partir de los datos de secuenciación con los miRNAs encontrados en la estrategia de búsqueda *in-silico*, se obtuvieron en común 6 miRNAs identificados por el programa R, contenidos en los 47 miRNAs en común identificados por el programa DAC; es decir, estos 47 miRNAs regulan RNAs mensajeros que codifican para proteínas dianas de E6 y E7 (ver figura 13 y tabla 5).

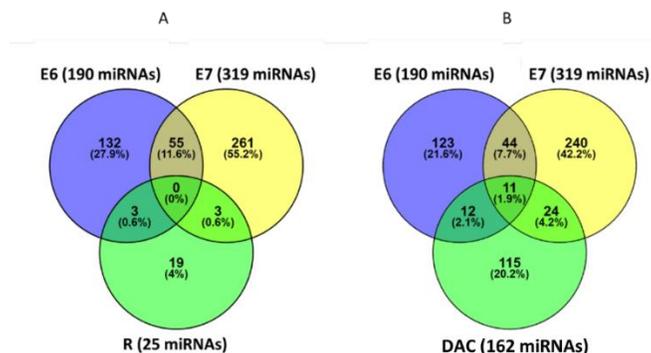


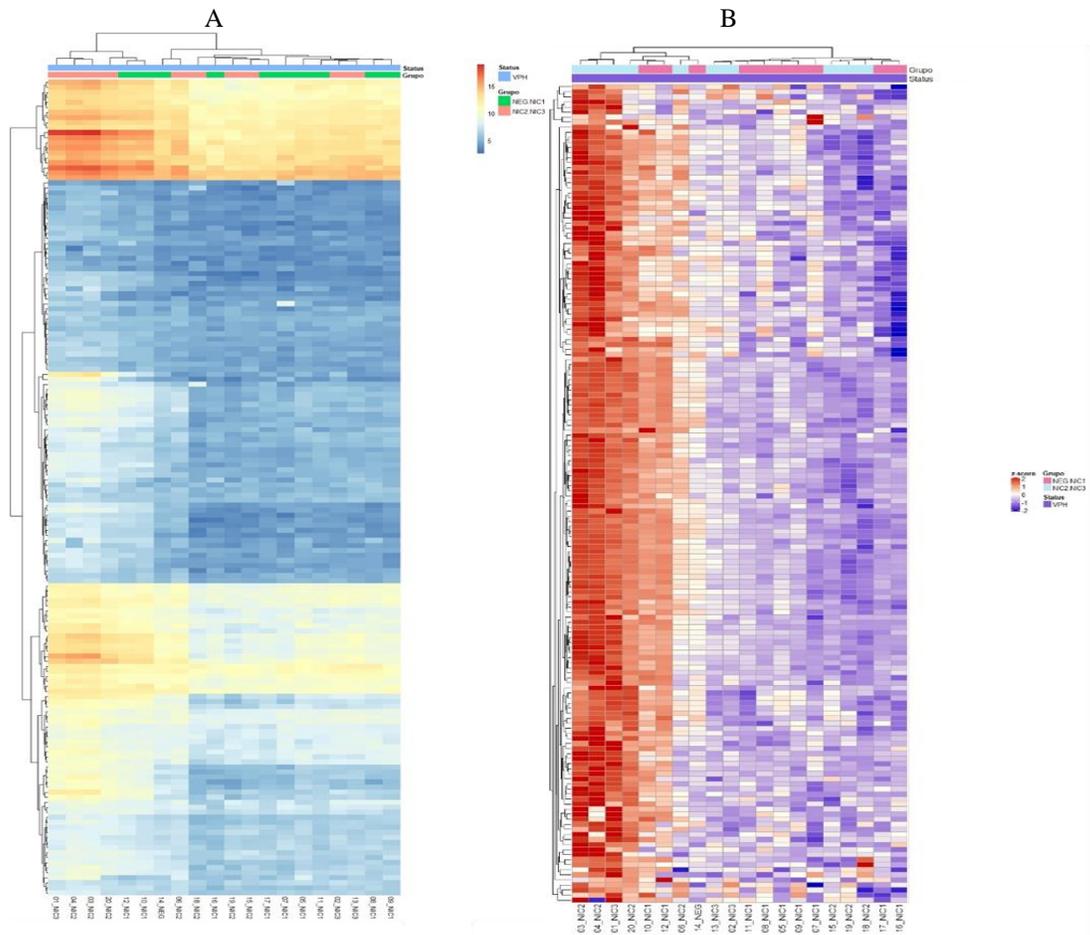
Figura 13. Número y porcentaje de los miRNAs diferencialmente expresados por secuenciación EN COMÚN con los encontrados *in-silico*. Si el análisis se hizo en el programa R (A) o el DAC (B) en círculos verdes. Y si la interacción es con un RNA mensajero que codifica para una proteína diana de E6 (círculo azul) o de E7 (círculo amarillo).

Tabla 5. Lista de miRNAs diferencialmente expresados encontrados en común entre la aproximación *in-silico* y el análisis de los datos de secuenciación en los programas R y DAC

Programa R		Programa DAC		
E6 n=3	E7 n=3	E6yE7 n=11	E6 n=12	E7 n=24
miR-5	miR-1	miR-92	miR-61	miR-1
miR-23	miR-6	miR-90	miR-81	miR-12
miR-9	miR-12	miR-36	miR-114	miR-6
		miR-143	miR-148	miR-63
		miR-74	miR-23	miR-55
		miR-144	miR-71	miR-65
		miR-104	miR-155	miR-58
		miR-80	miR-5	miR-28
		miR-97	miR-91	miR-38
		miR-34	miR-139	miR-149
		miR-75	miR-9	miR-86
			miR-54	miR-64
				miR-84
				miR-78
				miR-27
				miR-102
				miR-94
				miR-145
				miR-96
				miR-26
				miR-32
				miR-138
				miR-131
				miR-35

DAC: Programa Data Analysis Center - Qiagen; R: Programa R; n: número de miRNAs predichos de acuerdo con si interactúa con un gen blanco de acción de E6 o E7, o ambos. En letra de color rojo están señalados los miRNAs en común entre R y DAC.

El suplemento 4, muestra la variabilidad de los conteos de lecturas normalizados de los 162 miRNA encontrados diferencialmente expresados en el programa DAC para las 20 muestras.



Suplemento 4. Variabilidad entre los conteos de lecturas normalizados para cada miRNA (n=162) en cada una de las 20 muestras. A. La agrupación color verde hace referencia a los NIC1 y NEG y la agrupación color rosado a los NIC2 y NIC3, los valores van entre 5 y 15 lecturas normalizadas. **B.** Transformación a la puntuación Z (z-score) de los conteos de lecturas normalizadas, la agrupación color rosado hace referencia a los NIC1 y NEG y la agrupación color azul a los NIC2 y NIC3, los valores de z van entre -2 y 2, con una escala creciente desde colores azul oscuro hasta color rojo oscuro. Gráficos realizados en R. V.3.6.2. (A y B fueron obtenidos con diferentes paquetes de R).

La lista de los 162 miRNAs diferencialmente expresados entre lesiones de alto grado versus lesiones de bajo, encontrados en el programa DAC, se listan en la tabla del Suplemento 5.

Suplemento 5. Lista de 162 miRNA diferencialmente expresados en el programa DAC			Cont...			Cont....		
miRNA	programa DAC FC	valor p	miRNA	programa DAC FC	valor p	miRNA	programa DAC FC	valor p
miR-8	2,81	0,001	miR-86	2,42	0,007	miR-90	1,52	0,019
miR-6	2,98	0,001	miR-64	2,39	0,007	miR-117	2,42	0,02
miR-20	2,4	0,002	miR-30	1,74	0,007	miR-80	1,51	0,02
miR-4	2,58	0,002	miR-45	2,28	0,008	miR-111	2,4	0,02
miR-17	2,47	0,002	miR-83	2,38	0,008	miR-150	2,15	0,021
miR-9	2,91	0,002	miR-93	2,1	0,008	miR-75	1,9	0,021
miR-14	3,11	0,002	miR-39	2,03	0,008	miR-74	2	0,022
miR-23	3,21	0,002	miR-77	1,85	0,008	miR-99	1,8	0,022
miR-16	3,06	0,002	miR-140	1,51	0,008	miR-106	1,67	0,022
miR-21	2,08	0,002	miR-89	1,98	0,009	miR-69	1,64	0,023
miR-19	3,2	0,002	miR-51	2,76	0,009	miR-47	1,69	0,024
miR-13	2,87	0,002	miR-67	2,65	0,009	miR-79	1,51	0,025
miR-5	2,1	0,002	miR-50	2,6	0,01	miR-100	1,44	0,026
miR-25	2,81	0,003	miR-125	2,72	0,01	miR-130	2,17	0,026
miR-15	2,88	0,003	miR-136	2,72	0,01	miR-145	1,58	0,027
miR-7	2,41	0,003	miR-137	2,01	0,011	miR-159	1,63	0,028
miR-1	6,92	<0,001	miR-108	2,1	0,011	miR-96	1,7	0,03
miR-10	6,42	<0,001	miR-85	2,35	0,011	miR-26	1,69	0,03
miR-2	5,09	<0,001	miR-121	1,72	0,011	miR-32	1,59	0,03
miR-3	4,99	<0,001	miR-84	2,08	0,012	miR-82	1,46	0,03
miR-12	4,79	<0,001	miR-78	1,78	0,012	miR-105	1,53	0,032
miR-11	4,48	<0,001	miR-58	2,01	0,012	miR-156	1,58	0,033
miR-18	3,77	<0,001	miR-139	2,28	0,012	miR-91	1,43	0,033
miR-24	3,37	<0,001	miR-62	2,38	0,012	miR-72	1,76	0,033
miR-22	3,3	<0,001	miR-92	1,66	0,012	miR-61	1,41	0,033
miR-41	2,34	0,002	miR-36	2,09	0,013	miR-107	1,67	0,035
miR-31	2,7	0,003	miR-43	2,43	0,013	miR-29	1,8	0,035
miR-52	2,6	0,003	miR-116	2,66	0,013	miR-138	1,88	0,036
miR-46	2,28	0,004	miR-88	1,81	0,014	miR-76	1,63	0,036
miR-65	2,77	0,004	miR-155	2,06	0,014	miR-132	1,62	0,037
miR-115	3,19	0,004	miR-141	2,51	0,014	miR-60	1,81	0,037
miR-49	2,73	0,004	miR-27	1,95	0,014	miR-151	1,71	0,039
miR-135	2,98	0,004	miR-55	1,81	0,014	miR-147	1,32	0,039
miR-34	2,71	0,004	miR-37	1,98	0,014	miR-126	2,03	0,039
miR-95	2,48	0,004	miR-102	1,71	0,015	miR-53	1,49	0,041
miR-63	2,87	0,005	miR-57	1,79	0,015	miR-158	1,59	0,041
miR-119	2,92	0,005	miR-123	2,41	0,015	miR-112	1,63	0,041
miR-157	2,09	0,005	miR-40	1,73	0,015	miR-148	1,61	0,041
miR-142	2,27	0,005	miR-70	1,78	0,015	miR-131	2,07	0,042
miR-127	2,41	0,005	miR-143	1,95	0,015	miR-110	1,87	0,043
miR-161	1,91	0,005	miR-97	1,91	0,015	miR-101	1,73	0,043
miR-120	2,68	0,006	miR-56	1,69	0,016	miR-144	1,63	0,044
miR-68	2,81	0,006	miR-114	1,5	0,017	miR-153	1,59	0,045
miR-134	2,49	0,006	miR-109	2,03	0,017	miR-35	1,42	0,046
miR-54	2,02	0,006	miR-146	1,57	0,017	miR-104	1,67	0,046
miR-28	1,96	0,006	miR-42	2,46	0,017	miR-113	1,58	0,048
miR-38	2,13	0,006	miR-160	1,91	0,018	miR-152	1,7	0,049
miR-66	2,74	0,006	miR-129	2,41	0,018	miR-154	1,64	0,049
miR-81	2,2	0,006	miR-44	2,01	0,018	miR-33	1,62	0,05
miR-59	2,76	0,006	miR-73	1,76	0,018	miR-128	1,71	0,05
miR-149	2,7	0,006	miR-162	1,85	0,018	miR-87	1,51	0,05
miR-98	2,72	0,006	miR-94	1,63	0,019	miR-48	3,02	<0,001
miR-124	2,22	0,007	miR-122	1,85	0,019			
miR-103	1,89	0,007	miR-133	2,39	0,019			
miR-71	2,39	0,007	miR-118	2,47	0,019			

Valor p: Valores p, Prueba de Wald - algoritmo DESeq2. FC: Fold Change (casos/controles) Resaltado en fondo rosado los miRNAs diferencialmente expresados encontrados en común con el programa R.

La variabilidad entre los conteos de lecturas normalizados para el Top 25 de un total de 162 miRNA encontrados diferencialmente expresados en el programa DAC, (El Top 25 fue confirmado y encontrado en el programa R), se presenta en el mapa de calor de la figura 14. Estos 25 miRNAs fueron encontrados diferencialmente expresados significativos valor p ajustado menor de 0,05 (FDR<0.05) (prueba de Wald para comparar las lecturas de miRNAs mapeadas en los NIC2 y NIC3 con las lecturas de miRNAs mapeadas en los NEG y NIC).

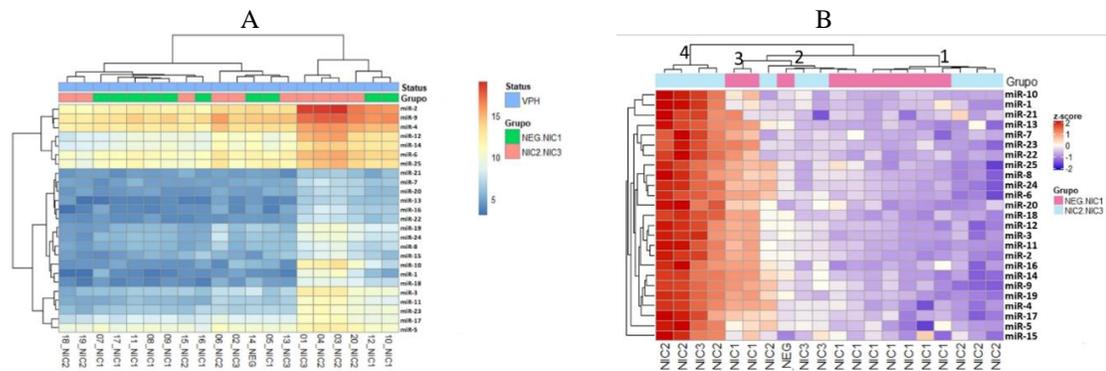


Figura 14. Variabilidad entre los conteos de lecturas normalizados para cada miRNA (n=25) en cada una de las 20 muestras. A. La agrupación color verde hace referencia a los NIC1 y NEG y la agrupación color rosado a los NIC2 y NIC3, los valores van entre 5 y 15 lecturas normalizadas. **B.** Transformación a la puntuación Z (z-score) de los conteos de lecturas normalizadas, la agrupación color rosado hace referencia a los NIC1 y NEG y la agrupación color azul a los NIC2 y NIC3, los valores de z van entre -2 y 2, con una escala creciente desde colores azul oscuro hasta color rojo oscuro. Gráficos realizados en R. V.3.6.2. (A y B fueron obtenidos con diferentes paquetes de R heatmap y mheatmap)

En el mapa de calor Figura 14.B se observa una alta variabilidad en los conteos de lecturas de los miRNAs entre las muestras secuenciadas, pues se evidencia solapamiento entre las lesiones de alto y bajo grado, el subgrupo uno (1) incluye muestras tanto de NIC1 como NIC2, subgrupo dos (2) NIC1 y NIC3, pero el subgrupo tres (3) está formado solo por muestras NIC1 y el subgrupo cuatro (4) por lesiones de alto grado NIC2 y NIC3.

En la tabla 6 se lista el Top 25 de los miRNAs encontrados en el programa DAC y confirmados en el programa R, con su respectivo valor de Veces de Cambio (Fold Change FC) y el valor p ajustado (FDR), obtenido por cada uno de los programas (R y DAC) y los valores del coeficiente de variabilidad (%CV) entre los conteos de lecturas normalizados (obtenidas al mapear las lecturas de la secuenciación) para cada miRNA en todas las muestras. Los valores de FC y valores p de los 162 miRNAs que se encontraron diferencialmente expresados en el análisis con el programa DAC, pero no en el programa R se presentan en la tabla del Suplemento 5.

Tabla 6. Top 25 de los miRNA diferencialmente expresados por secuenciación.						
miRNA		programa R		programa DAC		%CV
		FC	Valor p	FC	Valor p	
1	miR-5	2,3	0,05	2,1	0,002	9%
2	miR-4	2,9	0,05	2,58	0,002	10%
3	miR-9	3,6	0,05	2,91	0,002	12%
4	miR-17	2,8	0,049	2,47	0,002	13%
5	miR-6	3,7	0,049	2,98	0,001	13%
6	miR-25	3,5	0,05	2,81	0,003	14%
7	miR-21	2,2	0,049	2,08	0,002	15%
8	miR-14	4,1	0,05	3,11	0,002	18%
9	miR-20	2,7	0,049	2,4	0,002	18%
10	miR-2	10	0,012	5,09	<0,001	19%
11	miR-7	2,7	0,05	2,41	0,003	19%
12	miR-8	3,4	0,049	2,81	0,001	21%
13	miR-12	8,6	0,012	4,79	<0,001	22%
14	miR-15	3,7	0,05	2,88	0,003	25%
15	miR-23	4,5	0,05	3,21	0,002	25%
16	miR-24	4,5	0,04	3,37	<0,001	25%
17	miR-22	4,2	0,026	3,3	<0,001	26%
18	miR-13	3,6	0,05	2,87	0,002	27%
19	miR-19	4,4	0,05	3,2	0,002	28%
20	miR-16	4	0,05	3,06	0,002	29%
21	miR-3	9,4	0,012	4,99	<0,001	30%
22	miR-11	7,5	0,017	4,48	<0,001	30%
23	miR-18	5,9	0,049	3,77	<0,001	36%
24	miR-1	19,3	0,001	6,92	<0,001	44%
25	miR-10	32,8	0,001	6,42	<0,001	48%

Valor p ajustado: FDR Valores p ajustados por Benjamini y Hochberg FC: Fold Change (NIC2-NIC3/NEG-NIC1). Valor p, prueba de Wald - algoritmo DESeq2 en R y DCA. %CV: Porcentaje de coeficiente de variabilidad total (n=20 muestras) de los conteos de lecturas normalizados para cada miRNA

El análisis de expresión diferencial de los datos de secuenciación en el programa R, muestra que 25 miRNAs están expresados diferencialmente. De forma particular, estos 25 miRNA están **sobreexpresados/sobrerregulados** en lesiones de alto grado (NIC2=7-NIC3=3) respecto a las lesiones bajo grado o sin lesión (NIC1=9-NEG=1), es decir, los valores de FC son mayores a 1 (FC>1) (ver tabla 6). Así mismo los 162 miRNAs encontrados con el programa DAC (ver tabla del Suplemento 5), ninguno se encontró con valores de FC<1, es decir, ningún miRNAs diferencialmente expresado se encontró subexpresados en lesiones de alto grado versus lesiones de bajo grado.

Selección de miRNAs para validación

Cinco de los 25 miRNAs encontrados diferencialmente expresados, fueron seleccionados para validar su expresión en un grupo de muestras independientes (n=50). Se validaron los tres miRNAs con el más alto valor del FC y los dos miRNAs con el más bajo coeficiente de variabilidad (ver Tabla 7).

	R		DCA		Variabilidad
	FC	Valor p	FC	Valor p	CV
miR-1	19,3	0,001	6,9	<0,001	0,44
miR-2	10,0	0,012	5,1	<0,001	0,19
miR-3	9,4	0,012	5,0	<0,001	0,30
miR-4	2,9	0,050	2,6	0,002	0,10
miR-5	2,3	0,050	2,1	0,002	0,09

FC: Fold Change (NIC2+ / ≤ NIC1); prueba de Wald de DESeq2. programas R y DAC; CV: Coeficiente de variabilidad en todas las muestras (n = 20).

Los miRNAs (miR-1, miR-2, miR-3) fueron los que reportaron mayores valores de FC en lesiones de alto grado respecto a lesiones de bajo grado (valores resaltados en color amarillo). Sin embargo, miR-4 y miR-5 si bien tienen valores de FC>2, son los que presentan menor coeficiente de variabilidad de los conteos de lecturas normalizados para cada miRNA en todas las muestras (valores resaltados en color amarillo). Entre los cinco miRNAs seleccionados con estos criterios, dos de ellos, el miR-1 y miR-5 fueron identificados previamente en la estrategia de búsqueda *in-silico*, es decir que interactúan con los transcritos de los genes de proteínas blanco de acción de las oncoproteínas E6 y E7 del VPH (ver tabla 5).

Niveles de expresión relativa de miR-1, miR-2, miR-3, miR-4 y miR-5 por qRT-PCR

La expresión diferencial de estos 5 miRNA fue evaluada mediante la técnica de PCR en tiempo real. No se obtuvieron valores de CT para 6 muestras del grupo de descubrimiento (n=20) y para 6 muestras del grupo de validación (n=50). De acuerdo con lo anterior el número de muestras incluidas en los análisis de validación fueron:

- Grupo de descubrimiento (n=14): 1 NEG - 4 NIC1 - 6 NIC2 - 3 NIC3- 0 CEC
- Grupo de validación (n=44): 14 NEG - 7 NIC1 - 0 NIC2 - 21 NIC3- 2 CEC
- Total: 15 NEG - 11 NIC1 - 6 NIC2 - 24 NIC3- 2 CEC

Dada la representatividad de las tres lesiones preneoplásicas del cérvix en todo el conjunto de muestras (grupo de descubrimiento y validación), es de interés ver la expresión relativa de los cinco miRNAs en cada una de estas (ver figura 15).

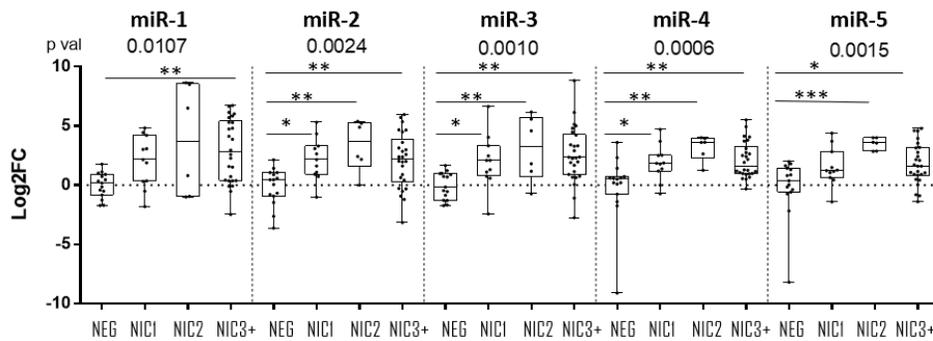


Figura 15. Expresión relativa Log2FC ($FC=2^{-\Delta\Delta CT}$) de los cinco miRNAs en cada grupo de lesión. En muestras del grupo de descubrimiento y validación. Las barras representan la mediana y los extremos los valores máx. y mín. El valor p se determinó por la prueba de Kruskal-Wallis. Para múltiples comparaciones la prueba de Dunn ($p < 0,05^*$; $p < 0,01^{**}$; $p < 0,001^{***}$) NEG: Negativo para lesión NIC: Neoplasia Intraepitelial Cervical grado 1,2 o 3. CEC: Carcinoma Escamocelular Cervical. Los tamaños de muestra fueron: NEG=15, NIC1=11, NIC2=6, NIC3=24, CEC=2 para todos los grupos.

Los niveles de expresión relativa para los cinco miRNAs fueron diferencialmente significativos (valor $p < 0,05$ Kruskal-Wallis) entre al menos dos grupos de comparación (NEG, NIC1, NIC2 o NIC3+). El grupo CEC se suma al grupo NIC3 dado el poco tamaño de muestra del grupo CEC ($n = 2$) y la similitud de NIC3 considerado cáncer *in-situ* con el CEC. La expresión relativa para el miR-2, 3 y 4 fue significativamente mayor en los grupos de lesión NIC3+, NIC2 y NIC1 con respecto al grupo NEG, los niveles de expresión relativa del miR-5 fueron significativamente mayores en los grupos NIC3+ y NIC2 con respecto al grupo NEG y los niveles de expresión relativa del miR-1 fueron significativamente mayores solo en el grupo NIC3+ con respecto al grupo NEG (estas diferencias entre grupos se evidencian por valores de $p < 0,05$ por la prueba de Dunn).

La figura 15, también muestra una alta variabilidad en la expresión relativa del miR-1-2 y 3 en las muestras del grupo NIC2 evidenciado por un Rango Intercuartílico alto, o cajas alargadas, pero no para los miR-4 y 5, que, en concordancia con los datos de secuenciación, fueron los miRNAs con menor coeficiente de variabilidad, y se puede observar que su variabilidad es pequeña en cada grupo de lesión.

La alta variabilidad en las muestras NIC2 puede deberse en primera instancia al pequeño tamaño de muestras, $n=6$, siendo necesario más muestras que representen mejor este grupo, y en segunda instancia se puede explicar también desde la naturaleza de la lesión, dado que la definición del diagnóstico histopatológico NIC2 tiene baja reproducibilidad, porque molecular y morfológicamente las células se encuentran en una transición entre los grados NIC1 y NIC3 (Stoler et al. 2015)

El diagnóstico histopatológico NIC2 se caracteriza por ser difícil de determinar y por tanto tiene muy baja reproducibilidad; sin embargo, la detección de biomarcadores como p16/Ki67 pueden aumentar la especificidad de detectar verdaderos NIC2. La sobreexpresión de la proteína p16 indica transformación por VPH, y la presencia de ki67 indica proliferación celular (Tsoumpou et al. 2009). De las 6 muestras NIC2, una (1) fue positivo para el biomarcador p16/Ki67 y cuatro (4) correspondían a diagnósticos releídos por un panel de patólogos expertos, pero no fueron probados para p16/ki67. Por las anteriores razones en el análisis de la figura 16 se excluyen las muestras del grupo NIC2.

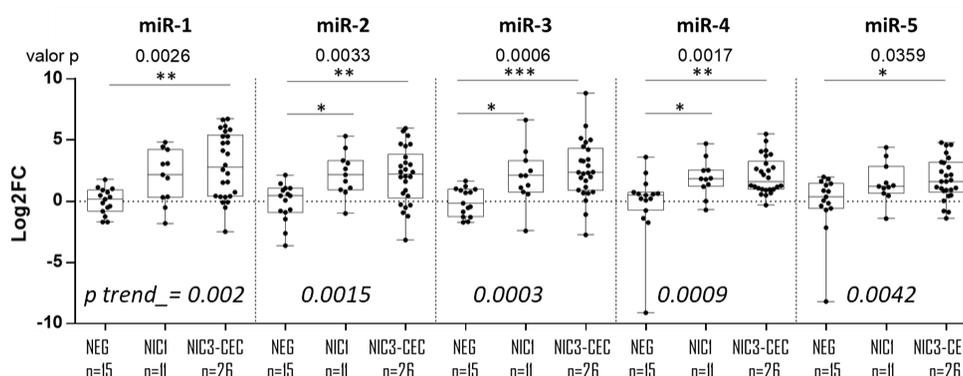


Figura 16. Expresión relativa Log2FC ($FC=2^{-\Delta\Delta CT}$) de los cinco miRNAs, en cada estadio de la enfermedad cervical NEG, NIC1 y NIC3+, sin incluir las muestras del grupo NIC2. En muestras del grupo de descubrimiento y validación. Las barras representan la mediana y los extremos los valores máx. y mín. El valor p se determinó por la prueba de Kruskal-Wallis Para múltiples comparaciones la prueba de Dunn ($p < 0,05^*$; $p < 0,01^{**}$; $p < 0,001^{***}$), p tren: prueba de tendencia lineal. NEG: Negativo para lesión. NIC: Neoplasia Intraepitelial Cervical grado 1 o 3. CEC: Carcinoma Escamocelular Cervical.

En la figura 16, se presenta la expresión relativa para cada grupo de lesión, excluyendo las muestras del grupo NIC2, se mantiene la observación que la expresión relativa de los cinco miRNAs es diferente entre al menos dos grupos de lesión ($p < 0,05$ Kruskal-Wallis), y también se mantiene que la expresión relativa del miR-2, 3 y 4 es significativamente mayor en el grupo NIC3+ y NIC1 con respecto al grupo NEG ($p < 0,05$ Dunn). Y los niveles de expresión del miR-1 y miR-5 sólo

fueron significativamente mayores entre el grupo NIC3+ respecto al grupo NEG ($p < 0,05$ Dunn). Aunque no se encontraron diferencias significativas en la expresión relativa entre cada grupo de lesión y el grupo inmediatamente antecesor, si es posible observar un aumento en la expresión relativa de los cinco miRNAs, con el avance o progresión de la enfermedad, evidenciado esto, por valores de p significativos para la prueba de tendencia lineal ($p < 0,05$).

Al comparar los niveles de expresión relativa de los cinco miRNAs entre lesiones de bajo grado (NEG-NIC1) y lesiones de alto grado (NIC2-NIC3-CEC) en mujeres VPH-AR+, en la figura 17 se muestra que los niveles para los cinco miRNAs son significativamente mayores en las lesiones de alto grado de cérvix con respecto a las lesiones de bajo grado del cérvix ($p < 0,05$ U-Mann-Whitney), indicando que estos miRNAs se encuentran en mayores niveles o sobreexpresados en mujeres VPH-AR+ que presentan lesiones preneoplásicas mayores o iguales a NIC2. De notar que las diferencias de los niveles de expresión diferencial de miR-1,3,4 y 5 entre lesiones de alto grado vs. de bajo grado fueron **muy** significativas ($p < 0,01$).

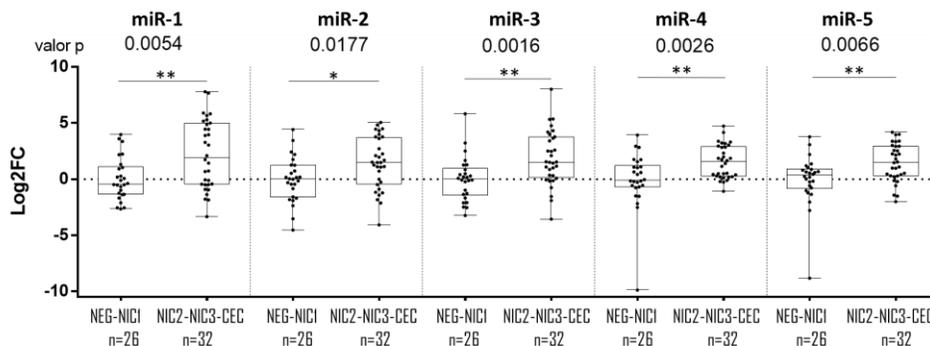


Figura 17. Expresión relativa Log2FC ($FC=2^{-\Delta\Delta CT}$) de los cinco miRNAs entre los grupos de lesiones de alto grado vs. de bajo grado. En muestras del grupo de descubrimiento y validación. Comparación de medianas entre dos grupos U Mann-Whitney $p < 0,05^*$; $p < 0,01^{**}$; $p < 0,001^{***}$. NEG=15, NIC1=11, NIC2=6, NIC3=24, CEC=2 para todos los grupos.

La figura 18, replica el análisis de la figura 17, **pero sólo con las muestras analizadas del grupo de validación (n=44)**, un grupo de muestras independientes a las usadas en la secuenciación, **notar que el grupo de muestras de validación no incluye muestras del grupo NIC2**, la figura 18 muestra que los niveles de expresión relativa para cuatro miRNAs (miR-1,2,3 y 4) fueron significativamente mayores en las lesiones de alto grado de cérvix (NIC3+) con respecto a las lesiones de bajo grado del cérvix (\leq NIC1) ($p < 0,05$ U-Mann-Whitney), indicando que estos miRNAs se encuentran en mayores niveles o sobreexpresados en mujeres VPH-AR+ que presentan

lesiones preneoplásicas mayores o iguales a NIC3. De notar que las diferencias de los niveles de expresión diferencial de miR-1 y miR-3 entre lesiones de alto grado vs. de bajo grado fueron **muy** significativas ($p < 0,01$). Aunque no hubo diferencia significativa entre los niveles de expresión del miR-5 entre lesiones de alto grado vs. bajo grado, si es posible notar un leve incremento en la expresión en las lesiones de alto grado vs. bajo grado con un valor p marginal ($p=0,0541$).

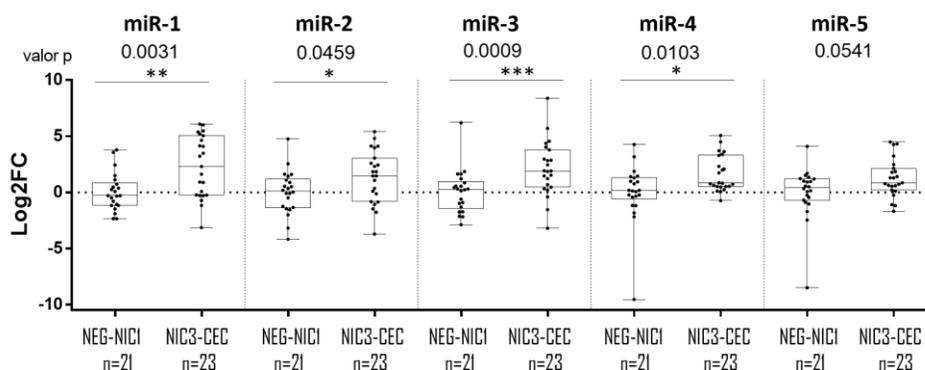


Figura 18. Expresión relativa Log2FC ($FC=2^{-\Delta\Delta CT}$) de los cinco miRNAs entre los grupos de lesiones de alto grado vs. de bajo grado. Solo con las muestras del grupo de validación. NEG=14, NIC1=7, NIC3=21, CEC=2 para todos los grupos.

Exploración de la utilidad clínica de miRNAs

Se obtuvieron curvas ROC para cada uno de los cinco miRNAs y también se hicieron curvas ROC promediando los valores de los ΔCT s de las combinaciones de todos los cinco miRNAs, los dos miRNAs que obtuvieron mejor AUC individual, los tres miRNAs con mayor FC y los dos miRNAs con menor %CV.

Se construyeron las curvas ROC para dos desenlaces:

Desenlace 1. Para la detección de NIC2+, usando las muestras de ambos grupos (descubrimiento y validación) ($n=58$), fueron considerados como controles las muestras NIC1-NEG ($n=26$) y el grupo desenlace fue NIC2-NIC3-CEC ($n=32$) (figura 19)

Desenlace 2. Para la detección de NIC3+, usando solo las muestras del grupo de validación ($n=44$), fueron considerados como controles las muestras NIC1-NEG ($n=21$) y el grupo desenlace fue NIC3-CEC ($n=23$) (figura 20)

La figura 19 muestra los valores de AUC y los valores p, para cada miRNA independiente (A-E) y las cuatro combinaciones propuestas: miR-1, 2 y 3 por tener los valores más altos de FC (F), miR-4 y miR-5 por tener los valores más bajos de %CV (G), miR-3 y miR-4 que reportaron los mayores valores de AUC independientemente (H) y la combinación de los cinco miRNAs (I).

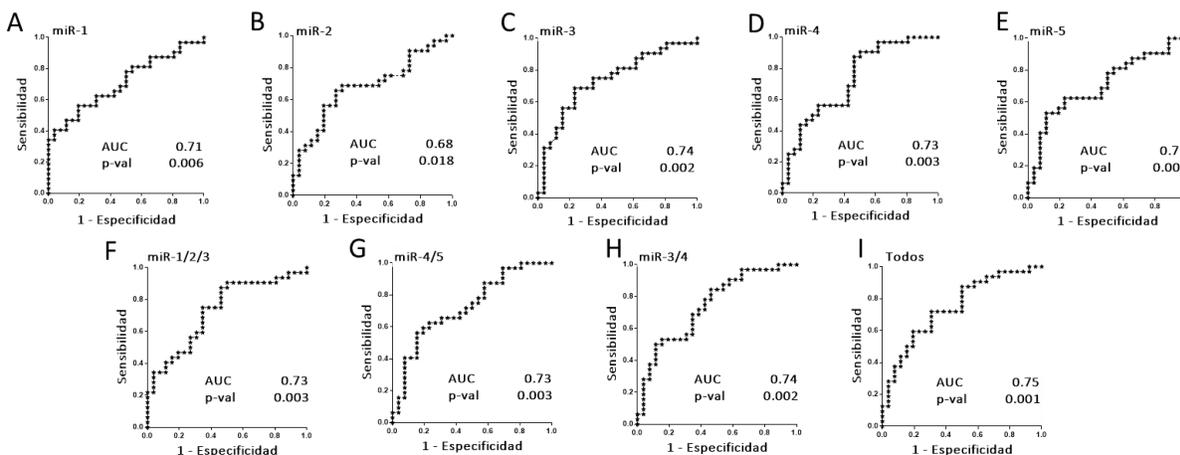


Figura 19. Curvas ROC para cada miRNA y combinaciones para la detección de NIC2+. Se usaron los valores del Δ CT (CT miRNA – CTgenRef). El diagnóstico histopatológico de biopsia dirigida por colposcopia se consideró la prueba estándar de oro. Valor $p < 0.05$, distribución normal al comparar con una curva ROC de AUC de 0.5. NIC2+ (NIC2=6, NIC3=24, CEC=2), \leq NIC1 (NEG=15, NIC1=11).

Las curvas ROC de la figura 19, se construyeron con las muestras del grupo de descubrimiento y validación. Los valores obtenidos indican que los cinco miRNAs individualmente y las cuatro combinaciones distinguen NIC2+ de \leq NIC1, esto evidenciado por valores p significativos ($p < 0.05$) al comparar con una curva con un área de 0,5 (hipótesis nula h_0 : el marcador no distingue entre NIC2+ y \leq NIC1). La combinación de los cinco miRNAs, seguido de la combinación del miR-3 y miR-4, fueron los clasificadores con más altos valores de AUC de 74,52% (IC 95%: 61,85-87,18) y 74,16% (IC 95%: 61,35-86,97) respectivamente.

La figura 20 muestra los valores de AUC y los valores p, para cada miRNA independiente (A-E) y las cuatro combinaciones propuestas: miR-1, 2 y 3 por tener los valores más altos de FC (F), miR-4 y miR-5 por tener los valores más bajos de %CV (G), miR-1 y miR-3 que reportaron los mayores valores de AUC independientemente (H) y la combinación de los cinco miRNAs (I).

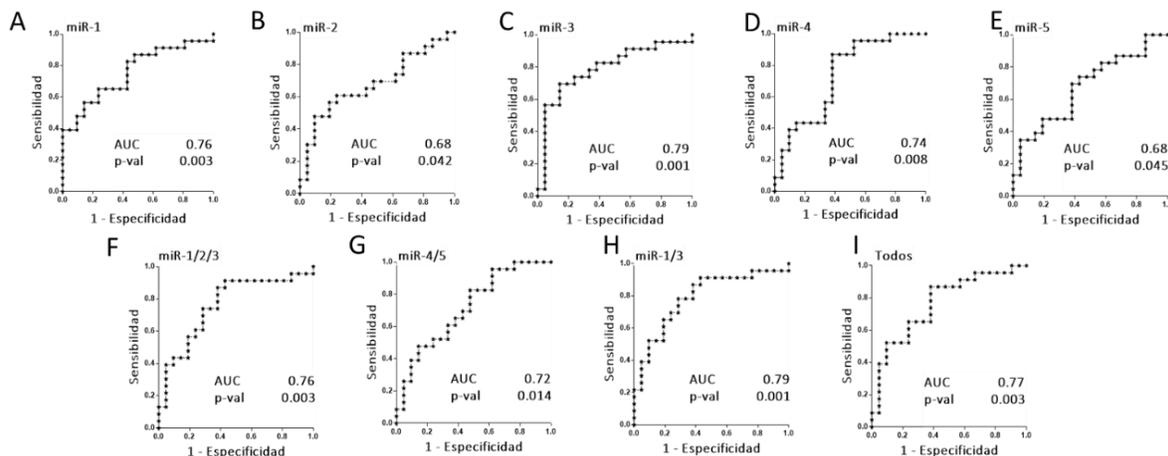


Figura 20. Curvas ROC para cada miRNA y combinaciones para la detección de NIC3+. Se usaron los valores del Δ CT (CT miRNA – CTgenRef). El diagnóstico histopatológico de biopsia dirigida por colposcopia se consideró la prueba estándar de oro. Valor $p < 0.05$, distribución normal al comparar con una curva ROC de AUC de 0.5. NIC3+ (NIC3=21, CEC=2) vs. \leq NIC1 (NEG=14, NIC=7).

Las curvas ROC de la figura 20, se construyeron sólo con las muestras del grupo de validación. Los valores obtenidos indican que los cinco miRNAs individualmente y las cuatro combinaciones distinguen NIC3+ de \leq NIC1, esto evidenciado por valores p significativos ($p < 0,05$) al comparar con una curva con un área de 0,5 (hipótesis nula: h_0 : h_0 : el marcador no distingue entre NIC3+ y \leq NIC1). El miR-3 exhibe el mayor valor de AUC de 79,09% (IC 95%: 65,07-93,11) junto a su combinación con miR-1 que exhibe un valor de AUC de 79,09 % (IC 95%: 65,42-92,76), seguido de la combinación de los cinco miRNAs con un AUC de 76,60% (IC 95%: 62,40-90,81) (tabla 8).

Tabla 8. Valores de AUC y valor p de cada miRNAs y sus combinaciones propuestas para la detección de lesiones de alto grado de cérvix (NIC2+ y NIC3+) en mujeres VPH-AR+

miRNAs individuales y combinaciones	NIC2+		NIC3+	
	%AUC (95%IC)	Valor p*	%AUC (95%IC)	Valor p*
miR-1	71,15 (58,00-84,31)	0,006	75,78 (61,48-90,07)	0,003
miR-2	68,15 (54,32-81,98)	0,018	67,91 (51,86-83,96)	0,042
miR-3	73,74 (60,73-86,75)	0,002	79,09 (65,07-93,11)	0,001
miR-4	72,84 (59,63-86,04)	0,003	73,50 (58,35-88,65)	0,008
miR-5	70,67 (57,22-84,13)	0,007	67,70 (51,79-83,61)	0,045
miR-1/2/3	73,08 (60,05-86,11)	0,003	76,19 (61,62-90,77)	0,003
miR-4/5	72,60 (59,46-85,73)	0,003	71,64 (56,49-86,79)	0,014
miR-3/4	74,16 (61,35-86,97)	0,002		
miR-1/3			79,09 (65,42-92,76)	0,001
miR-1/2/3/4/5	74,52 (61,85-87,18)	0,001	76,60 (62,40-90,81)	0,003

AUC calculado a partir de curvas ROC construidas con los valores de la diferencia entre el CT de cada miRNAs y el CT del SNORD44 en cada muestra. Desenlace NIC2+ (NIC2=6, NIC3 = 24, CEC = 2) vs (NEG = 15, NIC1=11) y Desenlace NIC3+ (NIC3 = 21, CEC = 2) vs (NEG = 14, NIC1=7). IC: Intervalo de confianza al 95%. *Valores p significativos $< 0,05$ determinado por distribución normal (dos-colas) asumiendo que un área del 50 % (hipótesis nula). El diagnóstico histopatológico a partir de biopsia tomada por colposcopia fue considerado la prueba estándar de oro.

En la tabla 8 se resume los resultados obtenidos del análisis de curvas ROC para la detección de NIC2+ y NIC3+. En amarillo se resaltan los valores de AUC obtenidos que mejor distinguen lesiones de alto grado NIC2+ o NIC3+ de lesiones de bajo grado (\leq NIC1) en mujeres VPH-AR+.

El mejor clasificador de lesiones NIC2+ es el marcador formado por la combinación de todos los miRNAs (figura 19), sin embargo, al incluir las muestras del grupo de descubrimiento en este análisis los resultados podrían estar sesgados, por lo que una mejor conclusión es de acuerdo a la figura 20, cuyo análisis fue exclusivamente con los datos del grupo de muestras independientes de validación, en este caso el mejor clasificador para distinguir lesiones NIC3+ de lesiones NIC1 y NEG en mujeres VPH-AR+, es el miR-3 con un AUC de 79,09% (IC 95%: 65,07-93,11) y aunque su combinación con miR-1 exhibe un valor de AUC de 79,09 % (IC 95%: 65,42-92,76) muy similar, es evidente que la fuerza para distinguir la lesión NIC3+ es otorgada principalmente por miR-3 por lo que no es necesario usar la combinación, y por tanto con la evaluación de sólo el miR-3 es suficiente para alcanzar la misma AUC.

10. Discusiones

Normalización de datos de qRT-PCR

Es conocido que la expresión de los genes es altamente variable entre individuos, tipos de muestras o métodos utilizados, y por ello en el método de qRT-PCR se utilizan los valores de detección por PCR del gen normalizador *SNORD44*, (propuesto como un adecuado gen normalizador para miRNAs (Masè et al. 2017)). Los datos presentados en la tabla 9 muestran que independiente del grado de la lesión o el porcentaje de la lesión en cada biopsia, los CTs para este gen no presentaron mayor variabilidad, indicando que hubo una adecuada normalización que controla los efectos de las distribuciones tan variables de la expresión de los genes normalmente esperada entre las diferentes muestras.

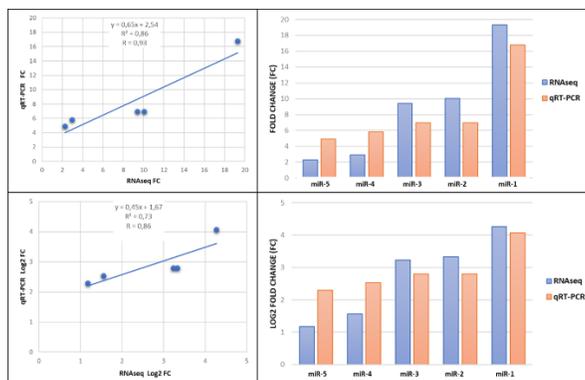
Tabla 9. Porcentaje de lesión y valores del CT del gen SNORD44 en las muestras del grupo de descubrimiento.

Muestra	% Lesión	CT SNORD44
4_NIC2	45	26,47
15_NIC2	20	26,9
12_NIC1	18	26,64
1_NIC3	15	25,35
18_NIC2	12	26,98
16_NIC1	10	TL
10_NIC1	5	25,68
20_NIC2	5	26,51
5_NIC1	3	TL
14_NEG	0	25,31
3_NIC2	2	25,84
7_NIC1	2	FR
17_NIC1	1	TL
2_NIC3	0	26,9
6_NIC3	0	26,9
8_NIC1	0	TL
9_NIC1	0	TL
11_NIC1	0	26,29
13_NIC3	0	25,79
19_NIC2	0	FR
Promedio	7,05	26,27
Desviación estándar	11	0,61
Coefficiente Variabilidad	1,82	0,02

El % Lesión fue leído a partir de cortes posteriores a los cortes usados para extracción de RNA. Por lo cual en algunos fue 0. FR: Fuera de Rango (CT>30 valor atípico extremos). TL: Too low, concentración muy baja de RNA insuficiente para la síntesis de DNAc.

Variabilidad de la expresión de miRNAs y correlación entre técnicas

Se presentó alta variabilidad en la expresión de los miRNAs entre las muestras secuenciadas (ver figura 14 y Suplemento 3 (PCA)), evidenciado en el solapamiento en los niveles de expresión entre lesiones de diferente grado, ya que los subgrupos uno y dos incluyen muestras tanto NIC1, como NIC2 y NIC3 en el mapa de calor presentado en la figura 14). Pese a esto, se logró validar cinco miRNAs por qRT-PCR que fueron diferencialmente expresados en la secuenciación. El coeficiente de correlación de Pearson fue de $R=0,93$ y $R=0,86$, al correlacionar los valores de FC y Log2FC respectivamente, entre las técnicas de RNAseq y qRT-PCR. Dado que estos análisis se llevaron a cabo en dos grupos de muestras completamente independientes esta correlación sugiere consistencia de los hallazgos por dos diferentes técnicas (ver Suplemento 6). Chen y colaboradores (2013), desarrollaron una estrategia para la identificación de miRNAs basándose en la secuenciación seguido de una validación de más de 100 miRNAs por qRT-PCR, pese a la alta variabilidad en la expresión de los miRNAs (evidenciada como el solapamiento entre los grupos de comparación), pudieron confirmar una alta correlación entre los perfiles de miRNAs de la secuenciación y los resultados por qRT-PCR, el coeficiente de correlación de Pearson fue de $R = 0.84$ (Cheng et al. 2013).



Suplemento 6. Correlación de los valores de FC y Log2FC de los cinco miRNAs entre la secuenciación y qRT-PCR. La expresión diferencial de cinco miRNAs fue confirmada por qRT-PCR. LA correlación entre los niveles de expresión obtenidos por ambas técnicas, secuenciación y qRT-PCR para cada uno de 5 miRNAs validados. El coeficiente de correlación de Pearson fue de $R=0,93$ y $R=0,86$ para el FC y el log2FC respectivamente. Valores altos sugieren consistencia entre los perfiles de miRNA y los de qRT-PCR.

Identificación de miRNAs in-silico

Inicialmente se hizo búsqueda *in-silico* de miRNAs que tuvieran capacidad de interactuar con los transcritos de 19 genes que codifican para proteínas blanco de acción de E6 y E7. De los 454 miRNAs (ver Figura 11) encontrados en la estrategia *in-silico*, seis presentaron expresión diferencial en el método de secuenciación cuando se utilizó el programa R, y 47 cuando se utilizó el programa DAC (ver tabla 5). A pesar de que se necesitan hacer ensayos posteriores en grupos de muestras independientes, estos hallazgos están en línea con nuestra hipótesis de que estos son biomarcadores con potencial biológico, ya que se ha mostrado que la desregulación de los miRNAs en el panorama de la infección con VPH es inducida por la acción de las oncoproteínas virales E6 y E7, lo que puede constituir pasos importantes en la carcinogénesis cervical (Harden et al. 2017; Harden y Munger 2017).

De manera interesante, 13 de los miRNA encontrados en la estrategia *in-silico* hacen parte de una lista de 51 miRNA asociados a la sobreexpresión *in-vitro* de E6 y E7 en cultivos primarios de queratinocitos de prepucio humano (Harden et al. 2017; Harden y Munger 2017). Estos 13 miRNAs son reguladores de genes que codifican para siete proteínas supresoras tumorales, tales como la *TP53* (p53), *CDKN1A* (p21), *CDKN1B* (p27) (Beaudenon y Huibregtse 2008; Milde-Langosch et al. 2001), *pRB-E2F1* (Zhang et al. 2015), *CUL2* (Huh et al. 2007), *EP300* (p300) (Patel et al. 1999) y *UBR4*(p600) (Huh K et al. 2005), ver interacciones miRNA-RNA en la figura 5.

Los transcritos *TP53* de los genes que codifican para la proteína p53, son regulados por miR-106b, miR-16, miR-18 y miR-25; *CDKN1A* (p21) por miR-106b, miR-15b, miR-16, miR-20b; *CDKN1B* (p27) por miR-155; *E2F1* por miR-106b, miR-181b, miR-205, miR-20b y miR-10a; *UBR4* (p300) por miR-378a, miR-155, *CUL2* por miR-15b y miR-16, y de *EP300* por miR-181b, miR-25. Las interacciones de estos 13 miRNAs se muestran en la Figura 5.

Interesantemente, se exploró la predicción de interacciones del top 25 de los miRNAs con expresión diferencial encontrados a partir de la secuenciación, usando el programa miRWalk2.0 (sin considerar el módulo de validación), la Figura 21 muestra la red de posibles blancos de acción de 19 de los 25 miRNA más significativos. Las interacciones encontradas reportaron probabilidades de unión significativas, con puntuaciones cercanas a 1 en una escala de cero a uno, lo que indica alta probabilidad de unión.

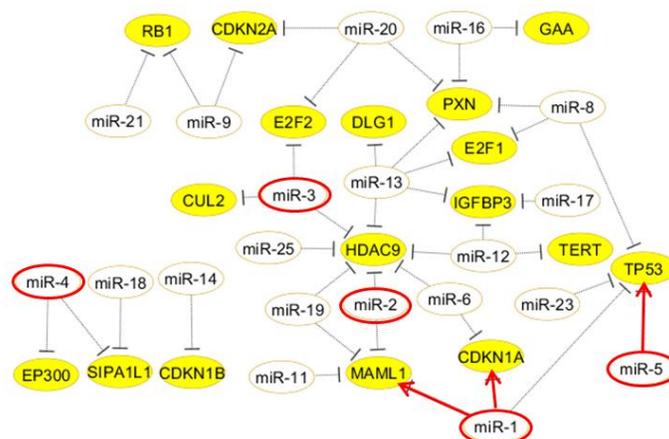


Figura 21. Red de interacciones predichas en miRWalk2.0 de miRNAs encontrados diferencialmente expresados por secuenciación y sus respectivos RNA mensajeros correspondientes a dianas E6/E7 VPH-AR. Elaborado en Cytoscape v.3.7.1. La predicción de dianas del Top 25 de miRNAs alterados (óvalos blancos), identificó diferentes RNA mensajeros (óvalos amarillos) correspondientes a dianas de E6 y E7. Las interacciones encontradas reportaron probabilidades de unión significativas, con puntuaciones cercanas a un 1 (en una escala de 0 a 1).

Los miRNAs en óvalos rojos fueron los seleccionados para validación por qRT-PCR, notar que tres interacciones (flechas de color rojo) entre dos de los miRNAs, el miR-1 y el miR-5, han sido reportadas/descritas previamente en la literatura (coincidieron con los resultados de la estrategia *in-silico*) (Ivanovska et al. 2008; Shi et al. 2019; Wu et al. 2014) , por lo cual estos dos miRNAs junto a los otros tres miRNAs seleccionados (miR-2, 3 y 4) para validación, representan ser buenos candidatos de biomarcadores, porque regulan RNAs mensajeros de dianas de E6/E7, por lo cual se justifica validar estos miRNAs en el contexto de la infección por VPH en lesiones cervicales.

Validación en tejidos de miRNAs encontrados in-silico y por secuenciación

De los 454 miRNAs encontrados en la estrategia *in-silico*, 47 coinciden con los encontrados en la estrategia de secuenciación, de estos 47 miRNAs se seleccionaron dos para validación, el miR-1 por tener el mayor FC y el miR-5 por tener el menor CV, por aparte se seleccionaron otros dos miRNAs, el miR-2 y el miR-3, por tener los más altos valores de FC después del miR-1, y se seleccionó también el miR-4 por tener el menor valor de CV después del miR-5, aunque miR-2, 3 y 4 no coincidieron con los resultados de la estrategia *in-silico*, la predicción de dianas muestra que la posibilidad de regulación de estos miRNAs con dianas de E6 y E7 es alta (figura 21). El miR-1 regula *MAML1*, *CDKN1A* y *TP53*; miR-2 regula *MAML1* y *HDAC9*; miR-3 regula *CUL2*, *E2F2* y *HDAC9*; miR-4 regula *EP300* y *SIPA1L1*, y el miR-5 regula *TP53*. Es importante aclarar que,

aunque estos hallazgos fortalecen la hipótesis del papel de los miRNAs en la oncogénesis inducida por el VPH, son necesarios ensayos controlados *in-vitro* para llegar a una conclusión directa de estas sugerencias.

Interferencia de microRNAs virales

El papel de los miRNAs codificados por el VPH ha sido estudiado; Virtanen y colaboradores (2016), estudiaron los niveles de expresión de miRNAs codificados por el genoma de VPH-16 en 20 TFFIP, 16 exfoliados cervicales recolectados por citología líquida, 16 exfoliados cervicales de mujeres que atendieron a colposcopia por anomalías cervicales, y en 4 líneas celulares VPH-16 positivo. Estos autores encontraron que los niveles de 5 miRNAs codificados por el genoma de VPH-16 fueron regulados negativamente en todas las muestras analizadas. Entre los virus de DNA, como los poliomavirus BK, JC y Merkel codifican miRNA, que regulan negativamente la expresión temprana de genes virales, su papel es ayudar al escape del ataque inmune del huésped y facilitan la replicación viral (Bauman et al. 2011; Lee et al. 2011). La detección de miRNAs virales en estas muestras sugiere que pueden regular la replicación viral y la expresión de genes celulares y, por lo tanto, tener un papel en la carcinogénesis (Virtanen et al. 2016). Análisis *in-silico* por Qian y colaboradores (2013), muestran que hay 459 posibles transcritos blancos de HPV16-miR-H6, sin embargo, hasta el momento no se ha probado la regulación de RNAm del hospedero por miRNAs de VPH-16 (Qian et al. 2013). Un análisis *in-silico* de predicción de interacción entre la secuencia de los cinco miRNAs de VPH-16 y los 19 RNA mensajeros de las proteínas dianas de E6 y E7, no mostró puntuaciones significativas para la posibilidad de interacción *in-silico*.

Validación de cinco miRNAs por qRT-PCR

El análisis de expresión diferencial de los datos de secuenciación encontró que los niveles de 162 miRNAs fueron más altos en lesiones de alto grado versus bajo grado del cérvix (ver tabla del Suplemento 5), y cinco de estos miRNAs, claramente no solo se comportan como potenciales biomarcadores para distinguir lesiones de alto grado de lesiones bajo grado en mujeres VPH-AR+ (miR-1, miR-2, miR-3, miR-4 y miR-5 con valores de FC 19.3, 10.0, 9.4, 2.9 y 2.3 respectivamente entre NIC2+ vs. \leq NIC1, ver tabla 7, también con valores p de 0.0054, 0,0177, 0.0016, 0.0026, y 0.0066 respectivamente entre NIC2+ vs. \leq NIC1, ver figura 17, y con valores p de 0,0031, 0,0459,

0.0009, 0.0103, 0.0541 respectivamente entre NIC3+ vs. \leq NIC1, ver figura 18, y con valores $p < 0.05$ para cada miRNA y combinaciones para la detección de NIC2+ y NIC3+, ver figura 19 y 20 respectivamente), sino que también se observa que la expresión relativa de los cinco miRNAs aumenta con la progresión de la enfermedad con una tendencia lineal, desde NEG a NIC1 y luego a NIC3+, p tendencia $< 0,05$ (ver figura 16) y aunque no se encontraron diferencias significativas entre cada lesión y su antecesora, esta observación le otorga relevancia biológica a estas moléculas como biomarcadores, con un importante valor clínico en términos de la transformación maligna y desarrollo de la enfermedad. Sin embargo, sólo la expresión relativa de miR-2, miR-3 y miR-4 fue significativamente mayor ($p < 0,05$) en los NIC3+, NIC2 y NIC1 respecto a los NEG (ver figura 15 y 16), lo cual sugiere que estos tres miRNAs podrían ser biomarcadores candidatos de transformación, desde un temprano grado de lesión preneoplásica del cérvix.

El análisis de datos de secuenciación mostró que los valores más altos de FC fueron obtenidos por miR-1, miR-2 y miR-3, en contraste el miR-4 y el miR-5 obtuvieron valores de FC no muy altos (2,9 y 2,3 respectivamente) (ver tabla 7). Entre estos cinco miRNAs diferencialmente expresados (valor p ajustado $< 0,05$), el hecho de que el miR-4 y 5 tienen menor valor de FC no significa necesariamente que no puedan ser tan buenos biomarcadores como los otros tres que tienen los más altos valores de FC, dado que mínimos cambios en moléculas que en estado normal no varían entre individuos, pueden llegar a representar un cambio significativo cuando se trata de moléculas que son altamente controladas a nivel celular (Fisher 2015), de hecho el miR-4 y 5 son los que presentaron menor variabilidad (ver tabla 7).

Los datos de la figura 15, se obtuvieron con muestras de ambos grupos de experimentos (descubrimiento y validación), y se realizó con el objetivo de analizar la expresión relativa de los miRNAs en todo el espectro de transformación de la enfermedad cervical, por lo cual se tienen muestras para cada grupo de lesión: NEG para lesión, NIC1, NIC2, NIC3 y CEC, dado el pequeño tamaño de muestras en el grupo CEC ($n=2$), estas se incluyen al grupo NIC3+, no obstante las diferencias que se obtienen con el grupo NIC3 no se ven alteradas al incluir las dos muestras de CEC (datos no mostrados). La figura 16, replica el análisis de la figura 15, excluyendo al grupo de muestras NIC2, en primera instancia por el pequeño tamaño de muestra ($n=6$), y en segunda instancia por la baja reproducibilidad de este diagnóstico histopatológico, ya que molecular y morfológicamente está en transición entre un NIC1 y NIC3. De hecho, de las lesiones NIC2 un

40% regresan y un 20 % avanzan a NIC3, por lo que en algunos casos la biopsia no logra concluir el grado de lesión, y se requiere confirmación por inmunohistoquímica de las proteínas p16/Ki67, la cual se evidencia positiva para p16 como un patrón difuso en células basales y parabasales indicando transformación por VPH, y la presencia de Ki67 indicador de proliferación celular (Tsoumpou et al. 2009). Para contrarrestar la proliferación y la transformación maligna inducida por la infección con VPH-AR, en la célula se induce la expresión de la proteína p16INK4a para bloquear la quinasa dependiente de ciclina 4/6 (CDK4/6), cuya función es restablecer la actividad anti-proliferativa de la pRb, pero su efecto se anula en presencia del VPH, ya que E7 al unirse y degradar a pRb induce la liberación no controlada de E2F independiente de la señalización por CDK4/6 (ver figura 4).

La figura 17 compara lesiones de alto grado con lesiones de bajo grado, con muestras de ambos grupos de estudio (descubrimiento y validación), y la figura 18, replica el análisis de la figura 17, pero solo con las muestras del grupo de validación. El razonamiento para la selección de las muestras del grupo de validación consistió en seleccionar solamente lesiones NIC3 y CEC, y no lesiones NIC2, por las razones expuestas en los párrafos anteriores, de tal forma que el grupo de muestras independientes para validación contará con rigurosa robustez en términos del diagnóstico histopatológico, para una correcta validación de los hallazgos que se obtuvieran en el grupo de descubrimiento el cual si contiene muestras de lesiones NIC2. Esto permite inferir que los miRNAs evaluados en el grupo de validación, es decir, los miRNAs sobreexpresados en lesiones NIC3+, posiblemente también estén sobreexpresados en NIC2+ (aunque esta lesión no se incluyó en el grupo de muestras de validación), los resultados de la secuenciación y de qRT-PCR en la figura 17 muestran que los cinco miRNAs están sobreexpresados en lesiones NIC2+.

El grupo de validación no incluyó muestras del grupo NIC2, dado que se deseaba validar un biomarcador de lesiones de alto grado, el mejor grupo de muestras que representa molecularmente las lesiones de alto grado son las lesiones NIC3 porque son las que anteceden molecular y morfológicamente al cáncer cervical, dada la claridad en la definición del diagnóstico NIC3 (alta reproducibilidad), dadas sus características como distribución de células en proliferación en los 3/3 del epitelio estratificado, se considera un cáncer *in-situ* y por tanto la lesión más cercana/similar a cáncer cervical, en la que se busca que el biomarcador esté presente. Por lo cual, NIC3 al ser una lesión altamente patognomónica, es decir que la presencia de esta lesión indica presencia de la

enfermedad (cáncer *in-situ*) y ser la lesión más cercana a cáncer cervical, se convierte en el grupo objetivo para la evaluación de biomarcadores que distingan lesiones de alto grado de lesiones de bajo grado o sin lesión.

Análisis de curvas ROC

En un ejercicio exploratorio y teniendo en cuenta que se requiere un tamaño de muestra más grande y una mejor representatividad de todo el espectro de la enfermedad, se calcularon los valores del Área bajo la curva (AUC) de los cinco miRNA validados por qRT-PCR y cuatro combinaciones de ellos, para la detección de NIC2+ usando las muestras de ambos grupos (descubrimiento y validación n=58) y para la detección de NIC3+ usando sólo las muestras del grupo de validación n=44). Para la detección de NIC2+, los cinco miRNAs individualmente y las cuatro combinaciones distinguen NIC2+ de \leq NIC1, esto evidenciado por valores p significativos ($p < 0,05$) al comparar con una curva de área de 0,5 (hipótesis nula h_0 : el marcador no distingue entre NIC2+ y \leq NIC1). La combinación de los cinco miRNAs, seguido de la combinación del miR-3 y miR-4, fueron los clasificadores con más altos valores de AUC de 74,52% (IC 95%: 61,85-87,18) y 74,16% (IC 95%: 61,35-86,97) respectivamente. Para la detección de NIC3+, los cinco miRNAs individualmente y las cuatro combinaciones distinguen NIC3+ de \leq NIC1, esto evidenciado por valores p significativos ($p < 0,05$) al comparar con una curva con un área de 0,5 (hipótesis nula: h_0 : h_0 : el marcador no distingue entre NIC3+ y \leq NIC1). El miR-3 presentó el mayor valor de AUC de 79,09% (IC 95%: 65,07-93,11) junto a su combinación con el miR-1 que presentó un valor de AUC de 79,09 % (IC 95%: 65,42-92,76), seguido de la combinación de los cinco miRNAs con un AUC de 76,60% (IC 95%: 62,40-90,81) (ver tabla 8).

Nuevamente se debe señalar que estos hallazgos deben ser comprobados en análisis posteriores, puesto que el tamaño de muestra de este estudio representa una importante limitación, pues con una potencia estadística del 80% para encontrar diferencias significativas entre un valor AUC deseado de 80% versus un AUC del 50%, con una confianza del 95%, se requieren al menos 30 muestras por grupo de comparación y en este estudio, las curvas para la detección de NIC2+ se construyeron con \leq NIC1=**26** y NIC2+=**32** y las curvas para detección de NIC3+ se construyeron con \leq NIC1=**21** y NIC3+=**23**. Dada esta limitante, los resultados de las curvas ROC de este trabajo deben ser interpretados únicamente como resultados exploratorios y no se consideran aptos para contrastar con la literatura.

No obstante a continuación se expone la literatura reportada en el tema: Babion y colaboradores (2019a), evaluaron el potencial marcador de miRNA en 132 exfoliados cervicales (NEG=77, NIC=55) VPH-AR+ e identificaron un clasificador compuesto por **miR-149, miR-20a y miR-93** con un AUC de 83% para la detección de NIC3 (Babion et al. 2019). Los mismos investigadores evaluaron en otro grupo de 187 exfoliados cervicales (NEG=66, NIC3=121) VPH-AR+, e identificaron un clasificador compuesto por **miR-15b-5p y miR-375** con un AUC de 62% para la detección de NIC3 (Babion et al. 2018); Snoek y colaboradores (2019), evaluaron el valor clínico de miRNA en 190 exfoliados cervicales de autotoma (NEG=101, NIC3=48 y CEC:41) VPH-AR+, e identificaron un marcador compuesto por cinco miRNAs (**let-7b, miR-15b, -20a, -93, y 222**) con un AUC de 78% para la detección de NIC3+ (Snoek et al., 2019). Ivanov y colaboradores (2018b), identificaron un clasificador compuesto por dos miRNAs (**miR-21 y -375**) a partir de 174 exfoliados cervicales (NEG=40, NIC1=34, NIC2-3=57 CEC=43), HPV-AR+, con AUC de 91% para la detección de NIC2+ (Ivanov et al. 2018). Cabe mencionar que 3 miRNAs (**miR-104, miR-12 y miR-65**) encontrados diferencialmente expresados por secuenciación en este trabajo, han sido evaluados en los trabajos anteriormente descritos, pero ninguno de estos se obtuvo dentro los mejores clasificadores listados en este párrafo.

Poniendo en contraste los resultados obtenidos de la aproximación de la evaluación clínica de los miRNAs con la literatura actual, se puede inferir que es suficiente un clasificador formado por dos a cinco miRNAs para distinguir lesiones de alto grado de las de bajo grado en mujeres VPH-AR+, los datos preliminares de este trabajo muestran que incluso con un sólo miRNA (miR-3) se logra distinguir lesiones NIC3+. No obstante al combinar biomarcadores de tipo miRNAs con marcadores como: metilación de genes supresores tumorales (Babion et al. 2019), genotificación de VPH 16/18 (Babion et al. 2018), carga viral de VPH (Ivanov et al. 2018) es posible aumentar el área bajo la curva a valores por encima de 90%.

Cada paso antes, durante y después del RNAseq puede influir en la reproducibilidad de los resultados, por lo cual se deben adoptar recomendaciones, como el manejo adecuado de las muestras, en el caso del presente trabajo, la asignación correcta del diagnóstico, el uso de replicas biológicas, es decir más de una unidad experimental por grupo de estudio, si es posible el uso de replicas técnicas, y usar kits de captura del RNA estándares, como captura de RNA a partir de la cola poli(A) (Carolina et al. 2012), , en el caso del presente trabajo, las secuencias maduras de los

miRNAs no tienen cola PoliA, por lo que el kit utiliza una polimerasa poli(A) seguido de cebadores poli(T).

11. Conclusiones

Las evidencias obtenidas en el presente trabajo que aportan a la aprobación de la hipótesis “*los miRNAs reguladores de las proteínas blanco de acción de las oncoproteínas E6 y E7, están sobreexpresados en lesiones de alto grado versus lesiones de bajo grado, y que por tanto tienen potencial como marcadores para la estratificación de las mujeres con mayor riesgo de presentar lesiones cervicales de alto grado*” son:

- ✓ Se identificaron 454 miRNAs en la estrategia *in-sílico*, 162 miRNAs en la secuenciación por el programa DAC y el top 25 de estos, concordó con los encontrados en el programa R. De los 162 miRNAs, 47 concuerdan con los resultados de la estrategia *in-sílico* (Tabla 5), sugiriendo que la ruta basada en dianas de E6 y E7 es acertada para dirigir la búsqueda de marcadores.
- ✓ De los cinco miRNAs validados por qRT-PCR, dos fueron encontrados en la estrategia *in-sílico* con tres interacciones validadas, **miR-1** regulando la expresión de *CDKN1A* y *MAML1* y **miR-5** regulando a *TP53*, genes que codifican para las proteínas blanco de la acción de E6 y E7 (Tabla 1). La estrategia *in-sílico* consistió en una etapa de predicción, seguido de una etapa de validación en la que se garantiza que la interacción ya ha sido encontrada y reportada previamente *in-vitro* (Figura 11). Esto indica que miR-1 y miR-5 se relacionan con procesos de transformación inducidos por el VPH-AR.
- ✓ Por otro lado, para los otros tres miRNAs validados por qRT-PCR (**miR-2, 3 y 4**), la predicción de dianas mostró que **miR-2** regula a *MAML1* y *HDAC9*, **miR-3** a *CUL2*, *E2F2* y *HDAC9* y **miR-4** a *EP300* y *SIPA1L1*, que codifican para proteínas blanco de acción de E6 y E7 (Tabla 1). Esta predicción arrojó probabilidades de interacción altas con puntuaciones cercanas a uno (ver Figura 21). Estos hallazgos fortalecen la hipótesis del papel de los miRNAs en la oncogénesis inducida por el VPH-AR, es necesario probar estas sugerencias *in-sílico*.
- ✓ Los niveles de expresión de **miR-1-2-3-4 y 5 fueron** mayores en NIC2+ vs. <NIC1, y en NIC3+ vs. <NIC1 excepto **miR-5**; los cinco mostraron diferencias en al menos dos grupos de lesión; y los niveles de **miR-2, 3 y 4** fueron mayores en NIC3+, NIC2 y NIC1 vs. NEG, sugiriendo un papel en la progresión de la enfermedad (p tendencia<0.05). Aunque los

niveles de expresión relativa del miR-5 no fueron significativamente mayores en lesiones de alto grado vs de bajo grado ($p=0,054$) (figura 16), sí fue encontrado en la estrategia *in-silico*, es decir, regula a *TP53* que codifica para la diana p53 de E6, esta evidencia es suficientemente robusta para sugerirlo como un potencial biomarcador de lesiones de alto grado en mujeres VPH-AR+ porque existe un vínculo que lo relaciona con los procesos de transformación inducido por el virus.

- ✓ La alta correlación entre RNAseq y qRT-PCR ($R=0.93$), indica consistencia entre ambos métodos. Este resultado respalda la necesidad de validar los demás miRNAs encontrados en común entre la estrategia *in-silico* y el RNAseq, porque reflejan eventos moleculares del desarrollo de la enfermedad.
- ✓ Los cinco miRNA y combinaciones, distinguen NIC3+ y NIC2+ de <NIC1 (AUC $p<0.05$). Es necesario contextualizar los valores de AUC, sensibilidad y especificidad para la detección de NIC3+, respecto a otros marcadores como VPH 16/18 y citología. Y de acuerdo con esto definir puntos de corte que optimicen el desempeño de miRNAs como marcadores de NIC2+ en mujeres VPH AR+.

Retomando, se identificaron 454 miRNAs reguladores de alguno de los 19 genes que codifican para proteínas blanco de acción de E6 y E7 del VPH-AR. De estos datos se tiene la certeza que la interacción es 100% real, dado que fueron encontrados en la sección de validación de miR-Walk.2.0. Se encontraron 25 miRNAs diferencialmente expresados por secuenciación en el programa R, contenidos en 162 miRNAs encontrados en el programa DAC, de estos 162, 47 miRNAs fueron predichos en la exploración *in-silico*, sugiriendo que la estrategia basada en proteínas diana de E6 y E7 es acertada para dirigir la búsqueda de marcadores moleculares. Se logró validar cinco de los miRNAs identificados por secuenciación (miR-1, miR-2, miR-3, miR-4 y miR-5) en un grupo independiente de 50 TFFIP, por qRT-PCR, dos de ellos, el miR-1 y miR-5 fueron predichos por la estrategia *in-silico*, los cuales bloquean los transcritos de los genes *CDKN1A*, *TP53* y *MAML1* que codifican para proteínas dianas de E6 y E7. Adicionalmente, miR-3 y su combinación con miR-1 otorgan el mayor AUC de 79,09 % (IC 95%: 65,07-93,11) y 79,09 % (IC 95%: 65,42-92,76) respectivamente, seguido de la combinación de los cinco miRNAs con un AUC de 76,60% (IC 95%: 62,40-90,81) para la detección de NIC3+. El mayor cambio en la expresión relativa lo confiere el miR-1, miR-2 y el miR-3 con valores de FC de 19.3, 10 y 9.4

respectivamente en el secuenciación, y de 16.8, 6.7 y 6.7 respectivamente en la validación por qRT-PCR, entre lesiones de alto grado versus lesiones de bajo grado (FC=NIC2+/NEG-NIC1), el coeficiente de correlación de Pearson, para los valores de FC de los cinco miRNAs entre ambas técnicas fue de $R=0,93$.

Finalmente, en conjunto los hallazgos descritos anteriormente muestran que los niveles de miRNAs reguladores o potencialmente reguladores de las proteínas blanco de acción de las oncoproteínas E6 y E7, están alterados (más altos) en lesiones de alto grado versus lesiones de bajo grado del cérvix, y podrían tener potencial biomarcador para distinguir lesiones de alto grado de cérvix de lesiones de bajo grado o sin lesión en mujeres VPH-AR+.

12. Recomendaciones

- ✓ Si bien esta claramente descrito la relación de miRNAs alterados (que regulan dianas de E6 y E7) con la sobreexpresión de E6 y E7 (Harden y Munger 2017), aún no se ha descrito la relación de miRNAs alterados (que regulan dianas de E6 y E7) con la reducción de la expresión de proteínas supresoras tumorales del ciclo celular (que son también blancos de E6 y E7). Por tanto, es necesario realizar más estudios que prueben la asociación directa de la desregulación molecular inducida por las oncoproteínas del VPH y el perfil de expresión de miRNAs en lesiones de alto grado versus bajo grado del cérvix.
- ✓ El smallRNAseq al ser una técnica de alta sensibilidad y cobertura, al mapear las lecturas, nos permite inferir que el gran número de lecturas no mapeadas representan una fuente invaluable de información y moléculas no anotadas; es muy probable que miRNAs no reportados, desempeñen un papel en lesiones de alto grado y cáncer cervical. Por lo cual, estudios más detallados de estas secuencias son necesarios para identificar y anotar nuevos miRNAs, así mismo, estudios funcionales para probar la relación de los miRNAs con los mecanismos de carcinogénesis cervical.
- ✓ Aumentar el tamaño de muestra del análisis de las curvas ROC, para alcanzar mayor robustez y mejor potencia estadística, y así lograr explorar correctamente la utilidad clínica y el desempeño diagnóstico de estos miRNAs para detectar lesiones de alto grado, en términos de sensibilidad y especificidad.

- ✓ Validar el resto de miRNAs diferencialmente expresados encontrados por secuenciación, en un grupo independiente de TFFIP de mayor tamaño. Puesto que cualquiera de los 162 miRNAs encontrados en la secuenciación, son candidatos y podrían ser buenos biomarcadores del cambio de lesión de bajo grado a lesión de alto grado de cérvix.
- ✓ Dada la facilidad de degradación de los ácidos nucleicos, siempre será mejor trabajar con muestras frescas, suspendidas en soluciones que desactiven la acción de RNAsas y conserven la estabilidad de los ácidos nucleicos, como el RNAlater.

13. Perspectivas

- ✓ Validar los hallazgos en muestras de exfoliado cervical; dado que estudios que evalúen el valor clínico de un biomarcador deberían realizarse en muestras clínicamente asequeables que puedan ser procesadas ojalá inmediatamente la muestra es tomada. En este caso, el exfoliado cervical que se toma para realizar la prueba de VPH podría servir para la evaluación de la expresión relativa de los miRNAs, en este sentido sería posible acoplar la detección de los miRNAs y el VPH en el mismo tubo con una misma técnica de PCR y de esta forma mejorar el diagnóstico de las mujeres VPH-AR+ con riesgo a desarrollar una lesión preneoplásica de alto grado.
- ✓ Ensayos pilotos por el Grupo Infección y Cáncer han mostrado que es posible cuantificar y detectar miRNAs extraídos de muestras de exfoliados cervicales, de mujeres con lesiones de bajo y alto grado de cérvix, todas VPH-AR+, suspendidas en diferentes medios de almacenamiento: STM (Specimen Transport Medium), STM+DNR (reactivo desnaturizante), PBS y RNAlater. Usando kits comerciales de extracción y los reactivos de Qiagen para la detección, los resultados muestran que el medio de STM+DNR conserva mejor los miRNAs, es una solución diluida de NaOH, que disminuye la acidez de la muestra, aumentando el pH y causando desnaturización de proteínas, por tanto, parece que este medio bloquea la acción de nucleasas y de esta forma se preservan mejor los miRNAs. Aunque la concentración de miRNAs es baja (promedio de 2,9 ng/uL de miRNAs extraídos a partir de exfoliados almacenados en STM+DNR), esta cantidad es suficiente para la detección por qRT-PCR, sin embargo, métodos de extracción que optimicen la concentración de miRNAs extraídos deben ser evaluados (Gonzalez y Sanchez 2019). Babion y colaboradores (2019) han trabajado en la detección de miRNAs extraídos de muestras de exfoliado cervical almacenadas

en medio PreservCyt, que es una solución de alcoholes presente en los viales de toma de muestra para la prueba de VPH, y han usado protocolos de extracción de RNA total basados en el reactivo TRIZol (Babion et al. 2019). Al lograr escalar el análisis de expresión diferencial relativa de miRNAs en muestras de exfoliado cervical, se podría pensar en la implementación de una prueba que detecte miRNAs, cuando la mujer asista a su cita de tamizaje por la prueba de VPH.

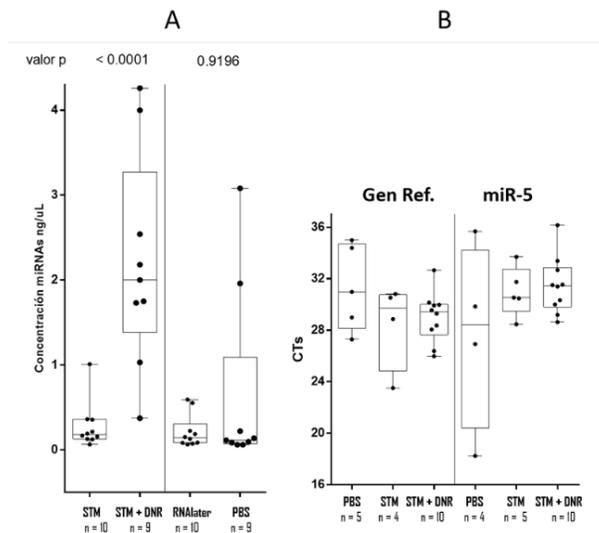


Figura 22. Cuantificación de miRNAs extraídos de exfoliados cervicales. A. concentración de miRNAs en muestras de exfoliados cervicales almacenados en 4 diferentes medios, (STM, STM+DNR, RNAlater y PBS) n es el tamaño de muestra en cada grupo. Las cajas representan el rango intercuartílico, los bigotes el valor mínimo y máximo, las líneas la mediana. Comparación de medianas U Mann Whitney. **B.** Valores de CTs para un gen de referencia (SNORD44 y el miR-5) de muestras de exfoliados cervicales almacenadas en diferentes medios de transporte.

- ✓ Los esfuerzos en la búsqueda de biomarcadores de lesiones de alto grado de cérvix en mujeres VPH-AR+, se centran en la evaluación de moléculas que regulen la expresión génica, si bien los miRNAs controlan los niveles de proteínas mediante represión postranscripcional, y se estima que más del 30% de la traducción de proteínas está regulada por miRNAs (Rajewsky 2006). Recientemente la regulación epigenética ha sido estudiada, la metilación de genes supresores tumorales involucrados en la desregulación inducida por la infección del VPH-AR, la metilación de las proteínas L1 y L2 del VPH-AR y la metilación de reguladores de la expresión génica tal como los miRNAs (Botezatu et al. 2011; Jiménez-Wences et al. 2016; Wilting et al. 2010; Wilting et al., 2013). El significado de la metilación en islas CpG alrededor de genes de miRNAs, radica en que los miRNAs con actividad supresora tumoral (regulan negativamente transcritos cuyos genes codifican para oncoproteínas), estarían hipermetilados

en cáncer por lo que se encontrarían con expresión reducida; y los miRNAs con actividad oncomir (regulan negativamente transcritos cuyos genes codifican para proteínas supresoras tumorales) estarían hipometilados en cáncer por lo que se encontrarían altamente expresados (Wang y Chim 2015).

Bibliografía

1. ACCP. 2011. *Recent Evidence on Cervical Cancer Screening in Low-Resource Settings Cervical cancer incidence and mortality update*. http://screening.iarc.fr/doc/ACCP_exca_screening_2011.pdf (18 de junio de 2020).
2. Aguilar-Fabré, Kenia, y María de los Ángeles Ríos-Hernández. 2008. «Human papilloma virus and cervical cancer». *Revista Cubana de Obstetricia y Ginecología* 34(1): 0-0. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-600X2008000100003 (15 de mayo de 2019).
3. Antinore, M. J. et al. 1996. «The human papillomavirus type 16 E7 gene product interacts with and trans-activates the AP1 family of transcription factors.» *The EMBO Journal* 15(8): 1950-60.
4. Babion, Iris et al. 2018. «Triage of high-risk HPV-positive women in population-based screening by miRNA expression analysis in cervical scrapes; a feasibility study». *Clinical Epigenetics* 10(1): 1-10.
5. ———. 2019. «Complementarity between miRNA expression analysis and DNA methylation analysis in hrHPV-positive cervical scrapes for the detection of cervical disease». *Epigenetics* 14(6): 558-67. <https://doi.org/10.1080/15592294.2019.1600390>.
6. Bartel, David P. 2004. «MicroRNAs: Genomics, Biogenesis, Mechanism, and Function». *Cell* 116(2): 281-97. <https://www.sciencedirect.com/science/article/pii/S0092867404000455?via%3Dihub> (15 de mayo de 2019).
7. Bauman, Yoav et al. 2011. «An identical miRNA of the human JC and BK polyoma viruses targets the stress-induced ligand ULBP3 to escape immune elimination». *Cell Host and Microbe* 9(2): 93-102. <http://www.ncbi.nlm.nih.gov/pubmed/21320692> (9 de febrero de 2020).
8. Beaudenon, Sylvie, y Jon M Huibregtse. 2008. «HPV E6, E6AP and cervical cancer.» *BMC biochemistry* 9 Suppl 1(Suppl 1): S4. <http://www.biomedcentral.com/1471-2091/9/S1/S4>.
9. Benjamini, Yoav ; Hochberg, Yocef. 1995. «Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995.pdf». *Journal of the Royal Statistical Society Series B (Methodological)* 57(1): 289-300. <http://www.jstor.org/stable/2346101> (28 de octubre de 2019).
10. Botezatu, Anca et al. 2011. «Quantitative analysis of the relationship between microRNA-124a, -34b and -203 gene methylation and cervical oncogenesis». *Molecular Medicine Reports* 4(1): 121-28.
11. Brown, Charlotte A. et al. 2012. «Role of protein biomarkers in the detection of high-grade disease in cervical cancer screening programs». *Journal of Oncology*.
12. Bugnon Valdano, M. et al. 2016. «Disc large 1 expression is altered by human papillomavirus E6/E7 proteins in organotypic cultures of human keratinocytes». *Journal of General Virology* 97(2): 453-62.
13. Calin, George A., y Carlo M. Croce. 2006. «MicroRNA signatures in human cancers». *Nature Reviews Cancer* 6(11): 857-66. <http://www.nature.com/articles/nrc1997> (15 de mayo de 2019).
14. Carolina, Johana, Soto Sedano, Y Camilo, y Ernesto López Carrascal. 2012. 16 Fitosanidad *RNA-seq: herramienta transcriptómica útil para el estudio de interacciones planta-patógeno*.
15. Castle, Philip E et al. 2009. «Short term persistence of human papillomavirus and risk of cervical precancer and cancer: population based cohort study.» *BMJ (Clinical research ed.)* 339: b2569. <http://www.ncbi.nlm.nih.gov/pubmed/19638649> (15 de mayo de 2019).
16. ———. 2011. «Performance of carcinogenic human papillomavirus (HPV) testing and HPV16 or HPV18 genotyping for cervical cancer screening of women aged 25 years and older: a subanalysis of the ATHENA study». *The Lancet Oncology* 12(9): 880-90. <http://www.ncbi.nlm.nih.gov/pubmed/21865084> (15 de mayo de 2019).
17. Cheng, Guofeng. 2015. «Circulating miRNAs: Roles in cancer diagnosis, prognosis and therapy». *Advanced Drug Delivery Reviews* 81: 75-93. <http://www.ncbi.nlm.nih.gov/pubmed/25220354> (15 de mayo de 2019).
18. Cheng, Wei Chung et al. 2013. «YM500: A small RNA sequencing (smRNA-seq) database for microRNA research». *Nucleic Acids Research* 41(D1).
19. Chiantore, Maria Vincenza et al. 2016. «Human papillomavirus E6 and E7 oncoproteins affect the expression of cancer-related microRNAs: additional evidence in HPV-induced tumorigenesis». *Journal of Cancer Research and Clinical Oncology* 142(8): 1751-63.
20. Condorelli, G., y S. Dimmeler. 2008. «MicroRNAs: components of an integrated system controlling cardiac development, physiology, and disease pathogenesis». *Cardiovascular Research* 79(4): 551-52. <http://www.ncbi.nlm.nih.gov/pubmed/18614548> (15 de mayo de 2019).
21. Cuzick, J, A Castañón, y P Sasieni. 2010. «Predicted impact of vaccination against human papillomavirus 16/18 on cancer incidence and cervical abnormalities in women aged 20-29 in the UK.» *British journal of cancer* 102(5): 933-39. <http://www.ncbi.nlm.nih.gov/pubmed/20104226> (15 de mayo de 2019).
22. Denny, Lynette. 2012. «Cytological screening for cervical cancer prevention». *Best Practice & Research Clinical Obstetrics & Gynaecology* 26(2): 189-96. <https://www.sciencedirect.com/science/article/pii/S1521693411001477?via%3Dihub> (15 de mayo de 2019).
23. Díaz-Cano, S J, y S P Brady. 1997. «DNA extraction from formalin-fixed, paraffin-embedded tissues: protein digestion as a limiting step for retrieval of high-quality DNA.» *Diagnostic molecular pathology: the American journal of surgical pathology, part B* 6(6): 342-46. <http://www.ncbi.nlm.nih.gov/pubmed/9559294> (15 de mayo de 2019).

24. Ding, Jun, Xiaoman Li, y Haiyan Hu. 2016. «TarPmiR: a new approach for microRNA target site prediction.» *Bioinformatics (Oxford, England)* 32(18): 2768-75. <http://www.ncbi.nlm.nih.gov/pubmed/27207945> (28 de octubre de 2019).
25. Doorbar, John. 2006. «Molecular biology of human papillomavirus infection and cervical cancer». *Clinical Science* 110(5): 525-41. <http://www.ncbi.nlm.nih.gov/pubmed/16597322> (15 de mayo de 2019).
26. ———. 2012. «The biology and life-cycle of human papillomaviruses». *Vaccine* 30(SUPPL.5): F55-70. <http://dx.doi.org/10.1016/j.vaccine.2012.06.083>.
27. ———. 2015. «Human papillomavirus molecular biology and disease association». *Reviews in Medical Virology* 25(S1): 2-23.
28. Douglas, Michael P., y Scott O. Rogers. 1998. «DNA damage caused by common cytological fixatives». *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 401(1-2): 77-88.
29. Dweep, Harsh, y Norbert Gretz. 2015. «miRWalk2.0: a comprehensive atlas of microRNA-target interactions». *Nature Methods* 12(8): 697-697. <http://www.nature.com/articles/nmeth.3485> (22 de mayo de 2019).
30. Dweep, Harsh, Carsten Sticht, y Norbert Gretz. 2013. «In-Silico Algorithms for the Screening of Possible microRNA Binding Sites and Their Interactions.» *Current genomics* 14(2): 127-36. <http://www.ncbi.nlm.nih.gov/pubmed/24082822> (15 de mayo de 2019).
31. Dweep, Harsh, Carsten Sticht, Priyanka Pandey, y Norbert Gretz. 2011. «MiRWalk - Database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes». *Journal of Biomedical Informatics* 44(5): 839-47. <http://dx.doi.org/10.1016/j.jbi.2011.05.002>.
32. Esquela-Kerscher, Aurora, y Frank J Slack. 2006. «Oncomirs - microRNAs with a role in cancer.» *Nature reviews. Cancer* 6(4): 259-69. <http://www.ncbi.nlm.nih.gov/pubmed/16557279>.
33. Feldman, M. Ya. 1973. «Reactions of Nucleic Acids and NucleoDroteins with Formaldehyde». *Progress in Nucleic Acid Research and Molecular Biology* 13(C): 1-49.
34. Felsani, A, A M Mileo, y M G Paggi. 2006. «Retinoblastoma family proteins as key targets of the small DNA virus oncoproteins». *Oncogene* 25(38): 5277-85. <http://www.nature.com/articles/1209621> (15 de mayo de 2019).
35. Ferlay, J et al. 2018. «Global Cancer Observatory: Cancer Today. 2018». <https://gco.iarc.fr/today/help?tab=6> (1 de mayo de 2019).
36. ———. 2019. «Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods.» *International journal of cancer* 144(8): 1941-53. <http://www.ncbi.nlm.nih.gov/pubmed/30350310>.
37. Fisher, Mike. 2015. «Biomarkers for diagnosis: Looking for change». *Biomarkers in Medicine* 9(12): 1275-77. <https://www.futuremedicine.com/doi/abs/10.2217/bmm.15.109> (20 de junio de 2020).
38. Gao, Q et al. 1999. «The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation.» *Molecular and cellular biology* 19(1): 733-44. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=83930&tool=pmcentrez&rendertype=abstract>.
39. Glinge, Charlotte et al. 2017. «Stability of circulating blood-based microRNAs-Pre-Analytic methodological considerations». *PLoS ONE* 12(2).
40. Gocze, Katalin et al. 2013. «Unique microRNA expression profiles in cervical cancer.» *Anticancer research* 33(6): 2561-67. <http://www.ncbi.nlm.nih.gov/pubmed/23749909> (15 de mayo de 2019).
41. González-Quintana, Víctor et al. 2016. «MicroRNAs are involved in cervical cancer development, progression, clinical outcome and improvement treatment response (Review)». *Oncology Reports* 35(1): 3-12. <https://www.spandidos-publications.com/10.3892/or.2015.4369> (15 de mayo de 2019).
42. Gonzalez, Martha I, y Gloria I Sanchez. 2019. «MicroRNAs reguladores de proteínas blanco de la acción de E6 y E7 de VPH como biomarcadores de Lesiones Preneoplásicas de alto grado de cérvix (NIC 2+) | González | Iatreia». <http://aprendeenlinea.udea.edu.co/revistas/index.php/iatreia/article/view/337741> (21 de noviembre de 2019).
43. Griffiths-Jones, Sam. 2006. «miRBase: The MicroRNA Sequence Database». En *MicroRNA Protocols*, New Jersey: Humana Press, 129-38. <http://link.springer.com/10.1385/1-59745-123-1:129> (15 de mayo de 2019).
44. Guerrero, Alexandra, y Milena Guerrero. 2016. «MicroRNAs asociados al Cáncer de Cuello Uterino y sus lesiones precursoras: Una revisión sistemática MicroRNAs associated with Cervical Cancer and its precursor lesions: A systematic Review». *Revista Universidad y Salud*. 28(2): 1-26.
45. Hanahan, Douglas, y Robert A. Weinberg. 2011. «Hallmarks of cancer: The next generation». *Cell* 144(5): 646-74. <http://dx.doi.org/10.1016/j.cell.2011.02.013>.
46. Harden, Mallory E., y Karl Munger. 2017. «Human papillomavirus 16 E6 and E7 oncoprotein expression alters microRNA expression in extracellular vesicles». *Virology* 508: 63-69. <https://www.sciencedirect.com/science/article/pii/S0042682217301538?via%3Dihub> (15 de mayo de 2019).
47. Harden, Mallory E, Nripesh Prasad, Anthony Griffiths, y Karl Munger. 2017. «Modulation of microRNA-mRNA Target Pairs by Human Papillomavirus 16 Oncoproteins.» *mBio* 8(1). <http://www.ncbi.nlm.nih.gov/pubmed/28049151> (15 de mayo de 2019).
48. He, Yuqing et al. 2016. «A systematic study on dysregulated microRNAs in cervical cancer development». *International Journal of Cancer* 138(6): 1312-27.
49. Hu, Xiaoxia et al. 2010. «A microRNA expression signature for cervical cancer prognosis.» *Cancer research* 70(4): 1441-48. <http://www.ncbi.nlm.nih.gov/pubmed/20124485> (15 de mayo de 2019).
50. Huang, Long et al. 2012. «Downregulation of six microRNAs is associated with advanced stage, lymph node metastasis and poor prognosis in small cell carcinoma of the cervix.» *PloS one* 7(3): e33762. <http://www.ncbi.nlm.nih.gov/pubmed/22438992> (15 de mayo de 2019).
51. Huh K. W, DeMasi J, Ogawa H, Nakatani Y, Howley P. M, Munger K. 2005. «Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600.» *Proceedings of the National Academy of Sciences of the United States of America* 102(32): 11492-97.
52. Huh, KyungWon et al. 2007. «Human papillomavirus type 16 E7 oncoprotein associates with the cullin 2 ubiquitin ligase complex, which contributes to degradation of the retinoblastoma tumor suppressor.» *Journal of virology* 81(18): 9737-47. <http://www.ncbi.nlm.nih.gov/pubmed/17609271> (31 de enero de 2020).
53. Ivanov, Mikhail K. et al. 2018. «Detection of high-grade neoplasia in air-dried cervical PAP smears by a microRNA-based classifier.» *Oncology Reports* 39(3): 1099-1111.
54. Ivanovska, Irena et al. 2008. «MicroRNAs in the miR-106b Family Regulate p21/CDKN1A and Promote Cell Cycle Progression». *Molecular and Cellular Biology* 28(7): 2167-74.
55. Jiménez-Wences, Hilda et al. 2016. «Methylation and expression of miRNAs in precancerous lesions and cervical cancer with HPV16

- infection». *Oncology Reports* 35(4): 2297-2305.
56. Jung, Monika et al. 2010. «Robust MicroRNA stability in degraded RNA preparations from human tissue and cell samples». *Clinical Chemistry* 56(6): 998-1006.
 57. Kakimoto, Yu et al. 2016. «MicroRNA Stability in FFPE Tissue Samples: Dependence on GC Content» ed. Andre van Wijnen. *PLOS ONE* 11(9): e0163125. <http://dx.plos.org/10.1371/journal.pone.0163125> (8 de noviembre de 2019).
 58. Kemp, Brian M., Misa Winters, Cara Monroe, y Jodi Lynn Barta. 2015. «How much DNA is lost? Measuring DNA loss of short-tandem-repeat length fragments targeted by the powerplex 16® system using the Qiagen Minelute purification kit». *Human Biology* 86(4): 313-29.
 59. Kitchener, Henry C., Philip E. Castle, y J. Thomas Cox. 2006. «Chapter 7: Achievements and limitations of cervical cytology screening». *Vaccine* 24: S63-70. <https://www.sciencedirect.com/science/article/pii/S0264410X0600658X?via%3Dihub> (15 de mayo de 2019).
 60. Kori, Medi, y Kazim Yalcin Arga. 2018. «Potential biomarkers and therapeutic targets in cervical cancer: Insights from the meta-analysis of transcriptomics data within network biomedicine perspective». *PLoS ONE* 13(7).
 61. Krek, Azra et al. 2005. «Combinatorial microRNA target predictions». *Nature Genetics* 37(5): 495-500. <http://www.nature.com/articles/ng1536> (15 de mayo de 2019).
 62. Kumar Gupta, Amit, y Manoj Kumar. 2015. «HPVbase - A knowledgebase of viral integrations, methylation patterns and microRNAs aberrant expression: As potential biomarkers for Human papillomaviruses mediated carcinomas». *Scientific Reports* 5.
 63. Lee, Sherry et al. 2011. «Identification and validation of a novel mature microRNA encoded by the Merkel cell polyomavirus in human Merkel cell carcinomas». *Journal of Clinical Virology* 52(3): 272-75.
 64. Li, Yang et al. 2011. «Progressive miRNA expression profiles in cervical carcinogenesis and identification of HPV-related target genes for miR-29». *The Journal of Pathology* 224(4): 484-95. <http://doi.wiley.com/10.1002/path.2873> (15 de mayo de 2019).
 65. Liang, Shuang et al. 2014. «Microarray analysis revealed markedly differential miRNA expression profiles in cervical intraepithelial neoplasias and invasive squamous cell carcinoma». *Future Oncology* 10(13): 2023-32. <https://www.futuremedicine.com/doi/10.2217/fo.14.38> (15 de mayo de 2019).
 66. Livak, Kenneth J, y Thomas D Schmittgen. 2001. «Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2 C T Method». *METHODS* 25: 402-8.
 67. Lizano, Marcela, Adela Carrillo, y Adriana Contreras. *Infección por virus del Papiloma Humano: Epidemiología, Historia Natural y Carcinogénesis*. <http://incan-mexico.org/revistainvestiga/elementos/documentosPortada/1272302572.pdf> (15 de mayo de 2019).
 68. Longworth, Michelle S., y Laimonis A. Laimins. 2004. «The Binding of Histone Deacetylases and the Integrity of Zinc Finger-Like Motifs of the E7 Protein Are Essential for the Life Cycle of Human Papillomavirus Type 31». *Journal of Virology* 78(7): 3533-41.
 69. Love, Michael I., Wolfgang Huber, y Simon Anders. 2014. «Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2». *Genome Biology* 15(12).
 70. Luck, Marisa E., Stefan A. Muljo, y Colm B. Collins. 2015. «Prospects for Therapeutic Targeting of MicroRNAs in Human Immunological Diseases». *The Journal of Immunology* 194(11): 5047-52.
 71. Ma, Qianhui et al. 2014. «Serum microRNA-205 as a novel biomarker for cervical cancer patients». *Cancer Cell International* 14(1): 81. <http://cancerbiomedcentral.com/articles/10.1186/s12935-014-0081-0> (15 de mayo de 2019).
 72. Mannhardt, B et al. 2000. «Human papillomavirus type 16 E7 oncoprotein binds and inactivates growth-inhibitory insulin-like growth factor binding protein 3.» *Molecular and cellular biology* 20(17): 6483-95.
 73. Mantovani, F, y L Banks. 2001. «The human papillomavirus E6 protein and its contribution to malignant progression.» *Oncogene* 20(54): 7874-87. <http://www.ncbi.nlm.nih.gov/pubmed/11753670>.
 74. Masè, Michela et al. 2017. «Selection of reference genes is critical for miRNA expression analysis in human cardiac tissue. A focus on atrial fibrillation». *Scientific Reports* 7(1): 41127. <http://www.nature.com/articles/srep41127> (6 de septiembre de 2019).
 75. Masuda, Norikazu et al. 1999. «Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples». *Nucleic Acids Research* 27(22): 4436-43.
 76. Meijer, Chris J L M et al. 2009. «Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older.» *International journal of cancer* 124(3): 516-20. <http://www.ncbi.nlm.nih.gov/pubmed/18973271> (15 de mayo de 2019).
 77. Mestdagh, Pieter et al. 2014. «Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study». *Nature Methods* 11(8): 809-15. <http://www.nature.com/articles/nmeth.3014> (15 de mayo de 2019).
 78. Meyers, J. M., J. M. Spangle, y K. Munger. 2013. «The Human Papillomavirus Type 8 E6 Protein Interferes with NOTCH Activation during Keratinocyte Differentiation». *Journal of Virology* 87(8): 4762-67.
 79. Milde-Langosch, Karin et al. 2001. «Expression of cyclin-dependent kinase inhibitors p16MTS1, p21WAF1, and p27KIP1 in HPV-positive and HPV-negative cervical adenocarcinomas». *Virchows Archiv* 439(1): 55-61.
 80. Moutinho, Catia, y Manel Esteller. 2017. «MicroRNAs and Epigenetics». En *Advances in Cancer Research*, Academic Press Inc., 189-220.
 81. Munger, K. et al. 2004. «Mechanisms of Human Papillomavirus-Induced Oncogenesis». *Journal of Virology* 78(21): 11451-60.
 82. Muñoz, Nubia et al. 2004. «Against which human papillomavirus types shall we vaccinate and screen? the international perspective». *International Journal of Cancer* 111(2): 278-85. <http://doi.wiley.com/10.1002/ijc.20244> (15 de mayo de 2019).
 83. Muñoz, Nubia, y Luis Eduardo Bravo. 2014. «Epidemiology of cervical cancer in Colombia.» *Salud publica de Mexico* 56(5): 431-39. <http://www.ncbi.nlm.nih.gov/pubmed/25604289> (15 de mayo de 2019).
 84. Murphy, Joan et al. 2012. «HPV Testing in Primary Cervical Screening: A Systematic Review and Meta-Analysis». *Journal of Obstetrics and Gynaecology Canada* 34(5): 443-52. <https://www.sciencedirect.com/science/article/pii/S1701216316352410?via%3Dihub> (15 de mayo de 2019).
 85. Nakagawa, S, y J M Huibregtse. 2000. «Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase.» *Molecular and cellular biology* 20(21): 8244-53.
 86. Park, Hyun et al. 2014. «Dysregulated microRNA expression in adenocarcinoma of the uterine cervix: Clinical impact of miR-363-3p». *Gynecologic Oncology* 135(3): 565-72. <https://www.sciencedirect.com/science/article/pii/S0090825814013134?via%3Dihub> (15 de mayo de 2019).
 87. Patel, D, S M Huang, L A Baglia, y D J McCance. 1999. «The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300». *Embo J.* 18(0261-4189 SB-M): 5061-72.
 88. Peng, Yong, y Carlo M. Croce. 2016. «The role of MicroRNAs in human cancer». *Signal Transduction and Targeted Therapy*

- 1(November 2015): 15004. <http://www.nature.com/articles/sigtrans20154>.
89. Peter, M E. 2010. «Targeting of mRNAs by multiple miRNAs: the next step». *Oncogene* 29(15): 2161-64. <http://www.nature.com/articles/onc201059> (15 de mayo de 2019).
 90. Pim, David, y Lawrence Banks. 2010. «Interaction of viral oncoproteins with cellular target molecules: Infection with high-risk vs low-risk human papillomaviruses». *Apmis* 118(6-7): 471-93.
 91. Vande Pol, Scott B, y Aloysius J Klingelutz. 2013. «Papillomavirus E6 oncoproteins». <http://pave.niaid.nih.gov/#home> (31 de enero de 2020).
 92. Qian, Kui et al. 2013. «Identification and Validation of Human Papillomavirus Encoded microRNAs» ed. Paulo Lee Ho. *PLoS ONE* 8(7): e70202. <http://dx.plos.org/10.1371/journal.pone.0070202> (9 de febrero de 2020).
 93. Rajewsky, Nikolaus. 2006. «microRNA target predictions in animals». *Nature Genetics* 38(S6): S8-13. <http://www.nature.com/articles/ng1798> (15 de mayo de 2019).
 94. Ramirez-Pineda, Arianis Tatiana et al. 2019. «Filogenia y oncogénesis del virus del papiloma humano: una aproximación translacional al descubrimiento de biomarcadores para la detección de lesiones precancerosas de cérvix». *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* 43(168): 351-65.
 95. Rao, Qunxian et al. 2012. «Aberrant microRNA expression in human cervical carcinomas». *Medical Oncology* 29(2): 1242-48. <http://link.springer.com/10.1007/s12032-011-9830-2> (15 de mayo de 2019).
 96. Rijkaart, Dorien C. et al. 2012. «Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: Final results of the POBASCAM randomised controlled trial». *The Lancet Oncology* 13(1): 78-88.
 97. Rodríguez, Ana Cecilia et al. 2008. «Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections». *Journal of the National Cancer Institute* 100(7): 513-17.
 98. Roman, Ann, y Karl Munger. 2013. «The papillomavirus E7 proteins».
 99. Ronco, Lucienne V., Alla Y. Karpova, Marc Vidal, y Peter M. Howley. 1998. «Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity». *Genes and Development* 12(13): 2061-72.
 100. Sahasrabudhe, Vikrant V., Patricia Luhn, y Nicolas Wentzensen. 2011. «Human papillomavirus and cervical cancer: Biomarkers for improved prevention efforts». *Future Microbiology* 6(9): 1083-98.
 101. de Sanjose, Silvia et al. 2010. «Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study». *The Lancet Oncology* 11(11): 1048-56. <http://www.ncbi.nlm.nih.gov/pubmed/20952254> (15 de mayo de 2019).
 102. Schiffman, Mark et al. 2007. «Human papillomavirus and cervical cancer». *The Lancet* 370(9590): 890-907. <http://www.ncbi.nlm.nih.gov/pubmed/17826171> (23 de mayo de 2019).
 103. Schiffman, Mark, y Nicolas Wentzensen. 2013. «Human papillomavirus infection and the multistage carcinogenesis of cervical cancer». *Cancer Epidemiology Biomarkers and Prevention* 22(4): 553-60.
 104. Shi, Wanyue et al. 2019. «Methylation-mediated silencing of miR-133a-3p promotes breast cancer cell migration and stemness via miR-133a-3p/MAML1/DNMT3A positive feedback loop». *Journal of Experimental & Clinical Cancer Research* 38(1): 429. <https://jccr.biomedcentral.com/articles/10.1186/s13046-019-1400-z> (19 de junio de 2020).
 105. Snoek, Barbara C. et al. 2019. «Genome-wide microRNA analysis of HPV-positive self-samples yields novel triage markers for early detection of cervical cancer». *International Journal of Cancer* 144(2): 372-79.
 106. Stoler, Mark H. et al. 2015. «The interpretive variability of cervical biopsies and its relationship to HPV status». *American Journal of Surgical Pathology* 39(6): 729-36.
 107. Tang, Fuchou et al. 2006. «MicroRNA expression profiling of single whole embryonic stem cells.» *Nucleic acids research* 34(2): e9. <http://www.ncbi.nlm.nih.gov/pubmed/16434699> (15 de mayo de 2019).
 108. Tian, Qifang et al. 2014. «MicroRNA detection in cervical exfoliated cells as a triage for human papillomavirus-positive women». *Journal of the National Cancer Institute* 106(9).
 109. Tong, Xiao, y Peter M. Howley. 1997. «The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton». *Proceedings of the National Academy of Sciences of the United States of America* 94(9): 4412-17.
 110. Tsoumpou, I. et al. 2009. «p16INK4a immunostaining in cytological and histological specimens from the uterine cervix: A systematic review and meta-analysis». *Cancer Treatment Reviews* 35(3): 210-20. <https://www.ncbi.nlm.nih.gov/books/NBK76839/> (20 de junio de 2020).
 111. Turchinovich, Andrey, Ludmila Weiz, Anne Langheinz, y Barbara Burwinkel. 2011. «Characterization of extracellular circulating microRNA.» *Nucleic acids research* 39(16): 7223-33. <http://www.ncbi.nlm.nih.gov/pubmed/21609964> (15 de mayo de 2019).
 112. Veldman, T I M, Izumi Horikawa, y J Carl Barrett. 2001. «Transcriptional Activation of the Telomerase hTERT Gene by Human Papillomavirus Type 16 E6 Oncoprotein». *Journal of Virology* 75(9): 4467-72.
 113. Virtanen, Elina et al. 2016. «Low expression levels of putative HPV encoded microRNAs in cervical samples». *SpringerPlus* 5(1): 1856. <http://springerplus.springeropen.com/articles/10.1186/s40064-016-3524-3>.
 114. Wang, Lu Qian, y Chor Sang Chim. 2015. «DNA methylation of tumor-suppressor miRNA genes in chronic lymphocytic leukemia». *Epigenomics* 7(3): 461-73.
 115. Wang, Wenyuan, Ester J. Kwon, y Li Huei Tsai. 2012. «MicroRNAs in learning, memory, and neurological diseases». *Learning and Memory* 19(9): 359-68.
 116. Wang, Xiaohong et al. 2014. «microRNAs are biomarkers of oncogenic human papillomavirus infections.» *Proceedings of the National Academy of Sciences of the United States of America* 111(11): 4262-67. <http://www.ncbi.nlm.nih.gov/pubmed/24591631> (15 de mayo de 2019).
 117. White, Martyn K, Joseph S Pagano, y Kamel Khalili. 2014. «Viruses and human cancers: a long road of discovery of molecular paradigms.» *Clinical microbiology reviews* 27(3): 463-81. <http://www.ncbi.nlm.nih.gov/pubmed/24982317> (15 de mayo de 2019).
 118. Wilting, S. M. et al. 2013a. «Altered microRNA expression associated with chromosomal changes contributes to cervical carcinogenesis». *Oncogene* 32(1): 106-16.
 119. ———. 2013b. «Methylation-mediated transcriptional repression of microRNAs during cervical carcinogenesis». *Epigenetics* 8(2): 220-28.
 120. Wilting, Saskia M. et al. 2010. «Methylation-mediated silencing and tumour suppressive function of hsa-miR-124 in cervical cancer». *Molecular Cancer* 9: 1-14.
 121. Wu, Junnan et al. 2014. «Downregulation of MicroRNA-30 facilitates podocyte injury and is prevented by glucocorticoids». *Journal of the American Society of Nephrology* 25(1): 92-104.

122. Yoon, S., y G. De Micheli. 2005. «Prediction of regulatory modules comprising microRNAs and target genes». *Bioinformatics* 21(Suppl 2): ii93-100. <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti1116> (15 de mayo de 2019).
123. Zhang, Weifang et al. 2015. «Cancerous inhibitor of protein phosphatase 2A contributes to human papillomavirus oncoprotein E7-induced cell proliferation via E2F1». *Oncotarget* 6(7): 5253-62. <http://www.impactjournals.com/oncotarget/index.php?journal=oncotarget&page=article&op=view&path%5B%5D=2867&path%5B%5D=6087>.
124. Zhou, Shan Shan et al. 2018. «MiRNAS in cardiovascular diseases: Potential biomarkers, therapeutic targets and challenges review-article». *Acta Pharmacologica Sinica* 39(7): 1073-84.
125. Zimmermann, H, R Degenkolbe, H U Bernard, y M J O'Connor. 1999. «The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300.» *Journal of virology* 73(8): 6209-19. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=112697&tool=pmcentrez&rendertype=abstract>.
126. Zwerschke, Werner et al. 2000. «Allosteric activation of acid α -glucosidase by the human papillomavirus E7 protein». *Journal of Biological Chemistry* 275(13): 9534-41.