



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

“FIRMA EPIGENÉTICA DEL CÁNCER
CERVICAL ASOCIADA A UN EXCESO
DE PESO DETERMINADA MEDIANTE
UN ANÁLISIS BIOINFORMÁTICO DE
METILACIÓN DE REGIONES
PROMOTORAS”

TESIS PARA OPTAR EL GRADO DE
MAESTRA EN INFORMÁTICA BIOMÉDICA
EN SALUD GLOBAL CON MENCIÓN EN
BIOINFORMÁTICA

ROSA MARIA OSCANO AIDA

LIMA – PERÚ

2020

ASESORA DE TESIS

Claudia Machicado Rivero, PhD, MSc

Universidad Peruana Cayetano Heredia

JURADO DE TESIS

DRA. TATIANA VIDAURRE ROJAS
PRESIDENTE

DRA. MARIA VICTORIA MARULL ESPINOZA
VOCAL

MG. CARLA MARIA GALLO LOPEZ ALIAGA
SECRETARIA

A mi familia

AGRADECIMIENTOS

A la Dra. Claudia Machicado, por su constante guía y asesoría durante toda la investigación. Sin duda, ha sido una etapa de la cual me llevo muchos aprendizajes valiosos.

A mis compañeros de laboratorio, Yomara Romero y Jorge Luis Lara, por su orientación inicial en el uso de herramientas bioinformáticas.

Al Dr. César Cárcamo, por su orientación y acertadas sugerencias en el análisis estadístico, además de su apoyo durante el programa de maestría.

A la MSc. Patricia Mallma, por su orientación en el análisis estadístico y ayuda en mis dudas correspondientes al tema.

A la Obs. Paola Pflucker y MSc. Daniel Condor por su apoyo durante el programa de maestría.

A Renzo José Carlos Calderón Anyosa, por su orientación y sugerencias en el manejo de datos y los análisis estadísticos.

Finalmente, un agradecimiento especial a mi familia: a mis padres, Rosa y Teodoro, y mi hermano, Julio. Gracias por ser mi mayor soporte y fuente de inspiración cada día.

FINANCIAMIENTO

Esta investigación fue financiada por CONCYTEC-FONDECYT.

TABLA DE CONTENIDOS

1	Introducción.....	1
1.1	Planteamiento de la investigación	1
2	Marco teórico.....	2
2.1	Epigenética: definición y mecanismos.....	2
2.2	Metilación del ADN.....	2
2.3	Epigenética en cáncer	4
2.4	Metilación de ADN aberrante en cáncer.....	4
2.4.1	Hipometilación global del ADN en cáncer	4
2.4.2	Hipermetilación en genes supresores de tumores.....	5
2.5	Cáncer cervical	5
2.6	Epigenética en cáncer cervical	7
2.7	Relación entre el exceso de peso y cáncer	8
2.8	Relación entre el exceso de peso y cáncer cervical	8
2.9	Metiloma en cáncer asociado al estado de exceso de peso	9
2.10	Estudio de la genómica del cáncer cervical mediante el proyecto <i>The Cancer Genome Atlas</i> (TCGA-CESC).....	9
2.11	Conceptos utilizados en estudios bioinformáticos	10
2.11.1	Análisis de expresión diferencial de genes	10
2.11.2	Ontología génica.....	11
2.11.3	Rutas KEGG.....	12
3	Justificación del estudio	13
4	Pregunta de investigación.....	14
5	Hipótesis	14
6	Objetivos.....	14
6.1	Objetivo principal	14
6.2	Objetivos específicos	14
7	Metodología.....	16
7.1	Diseño del estudio.....	16
7.2	Población y muestra de estudio	16
7.3	Criterios de selección.....	16
7.3.1	Criterios de inclusión	16
7.3.2	Criterios de exclusión	16
7.4	Operacionalización de variables.....	17

7.4.1	Características clínicas y demográficas	17
7.4.2	Datos genómicos y epigenómicos	21
7.5	Instrumentos para la recolección de datos	23
7.6	Procedimientos y técnicas	25
7.6.1	Construcción de BD de la población	26
7.6.2	Selección de DEGs en TCGA-CESC	26
7.6.3	Construcción de la BD de expresión de los DEGs	27
7.6.4	Redes de PPI y enriquecimiento funcional	29
7.6.5	Construcción de la BD de metilación del ADN	30
7.7	Plan de análisis	31
7.8	Consideraciones éticas	34
8	Resultados	35
8.1	Características de la población	35
8.2	Selección de DEGs en TCGA-CESC	38
8.3	Análisis de expresión génica	38
8.3.1	Análisis de expresión por estado de exceso de peso	38
8.3.2	Nivel de expresión por tipo de muestra	47
8.4	Análisis de redes de PPI y enriquecimiento funcional	47
8.4.1	Genes sub-expresados	47
8.4.2	Genes sobre-expresados	48
8.5	Análisis de metilación del ADN	64
8.5.1	Selección de transcritos	64
8.5.2	Nivel de metilación en función del estado de exceso de peso	67
8.5.3	Estado de metilación en función del estado de exceso de peso	72
8.5.4	Nivel de metilación por tipo de muestra	80
8.6	Asociación del estado de metilación con características clínicas y demográficas	80
8.7	Sobrevida global en función del estado de metilación del ADN	91
9	Discusión	96
10	Conclusiones	106
11	Limitaciones	106
12	Recomendaciones	107
13	Referencias bibliográficas	108
14	Anexos	

ÍNDICE DE TABLAS

Tabla 1 Operacionalización de variables de características clínicas y demográficas	17
Tabla 2 Operacionalización de variables de datos genómicos y epigenómicos	21
Tabla 3 Herramientas web utilizadas para la recolección de datos	23
Tabla 4 Características generales de las pacientes por grupo de estudio.....	36
Tabla 5 Resultados de expresión de los DEGs sub-expresados en CCE	41
Tabla 6 Resultados de expresión de los DEGs sobre-expresados en CCE	44
Tabla 7 Enriquecimiento funcional de términos de GO en los genes sobre-expresados	51
Tabla 8 Enriquecimiento funcional de términos KEGG en los genes sobre-expresados	53
Tabla 9 Descripción y función de los 39 DEGs seleccionados	54
Tabla 10 Detalles de los transcritos de estudio	64
Tabla 11 Nivel de metilación en función del estado de exceso de peso	68
Tabla 12 Estado de metilación de los transcritos de estudio.....	73
Tabla 13 Estado de metilación en función del estado de exceso de peso	78
Tabla 14 Asociación del estado de metilación de los ocho transcritos seleccionados con las características clínicas y demográficas de las participantes	82

ÍNDICE DE FIGURAS

Figura 1 Metilación del ADN.....	3
Figura 2 Anatomía del aparato reproductor femenino.....	6
Figura 3 Flujograma de la metodología del estudio	25
Figura 4 Flujograma de selección de DEGs.....	39
Figura 5 Red de PPI de los genes sobre-expresados seleccionados	49
Figura 6 Gráficas de KM para representar la supervivencia global en función del estado de metilación de los transcritos.....	92

ABREVIATURAS

CCE: Cáncer cervical

TCGA-CESC: Estudio de cáncer cervical del proyecto *The Cancer Genome Atlas*

DEGs: Genes diferencialmente expresados

IMC: Índice de masa corporal

CGI: Islas CpG

VPH: Virus del papiloma humano

GO: Ontología génica

KEGG: Enciclopedia de Genes y Genomas de Kioto

BD: Base de datos

RESUMEN

Introducción: El exceso de peso (sobrepeso u obesidad) está asociado al pronóstico de las pacientes con cáncer cervical (CCE), sin embargo, los mecanismos moleculares que explican esta asociación aún son desconocidos. El objetivo de este estudio fue evaluar la potencial asociación entre el exceso de peso, el estado de metilación de la región promotora de genes diferencialmente expresados (DEGs) en CCE y la sobrevida global.

Metodología: Se realizó un análisis secundario de datos genómicos y epigenómicos del proyecto de libre acceso de CCE denominado TCGA-CESC (The Cancer Genome Atlas Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma). Las pacientes fueron clasificadas de acuerdo a su Índice de Masa Corporal (IMC) como pacientes sin exceso de peso ($IMC < 25$) o con exceso de peso ($IMC \geq 25$). Se identificaron DEGs en CCE y se recolectaron datos de expresión mediante las plataformas online GEPIA y UCSC Xena, respectivamente. Se seleccionaron DEGs con expresión diferencial entre los grupos de estudio, se recolectaron datos del nivel de metilación a través de la base MethHC y se evaluó la asociación de la metilación en la región promotora y el estado de exceso de peso. Asimismo, se evaluó la asociación de características clínicas y demográficas de las participantes y el estado de metilación de la región promotora de ocho DEGs seleccionados. Finalmente, se evaluó la asociación de la sobrevida global y el estado de metilación de los DEGs en CCE seleccionados.

Resultados: Se identificaron 39 DEGs en CCE con diferencias significativas de expresión génica entre las pacientes sin exceso y con exceso de peso ($p < 0.01$). Se observó una diferencia significativa en el nivel de metilación de la región promotora

entre los grupos de estudio en siete genes ($p < 0.05$): *SPOCK1* (NM_004598), *EDIL3* (NM_005711), *SEC31B* (NM_015490), *GPM6A* (NM_005277), *C2ORF40* (NM_032411), *ETV7* (NM_016135) y *MND1* (NM_032117). Hubo una asociación significativa entre el estado de metilación de *FGF10* y el estado de exceso de peso [OR=3.81, IC al 95%; 1.36 -10.68], considerando el resto de características clínicas y demográficas de la población constantes. Finalmente, el estado de metilación del gen sub-expresado *PRICKLE2* (NM_198859) estuvo asociado con la sobrevida global ($p=0.0022$).

Conclusiones: Se identificaron diferencias en el nivel de metilación en la región promotora de siete DEGs en CCE en función del estado de exceso de peso en la cohorte TCGA-CESC. Asimismo, el estado de metilación del gen sub-expresado *PRICKLE2* (NM_198859) estuvo asociado a la sobrevida global. Nuestros resultados sugieren un rol de la metilación asociada al estado de exceso de peso en CCE y una posible implicancia en la sobrevida. Esta área de investigación amerita mayor exploración por su potencial relevancia clínica.

Palabras claves: Exceso de peso, metilación, región promotora, cáncer cervical, sobrevida, CESC-TCGA

SUMMARY

Introduction: Excess of body weight (overweight or obesity) is associated with the prognosis of cervical cancer (CCE) patients, however, the molecular mechanisms that explain this association remain unclear. The present study aimed to evaluate the potential association between excess of body weight, the methylation status of the promoter region of differentially expressed genes (DEGs) in CCE and overall survival.

Methodology: We performed a secondary analysis of genomic and epigenomic data from the CCE free-access project, TCGA-CESC (The Cancer Genome Atlas Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma). Patients were classified according to their Body Mass Index (BMI) as patients without an excess of body weight ($BMI < 25$) or patients with an excess of body weight ($BMI \geq 25$). We identified DEGs in CCE and retrieved expression level data using GEPIA and UCSC Xena, respectively. We selected DEGs with differential expression between the study groups, retrieved methylation level data through MethHC and evaluated the association between methylation in the promoter region and excess of body weight status. Additionally, we evaluated the association of clinical and demographic characteristics and the methylation status of the promoter region of eight selected DEGs. Finally, we evaluated the association between survival and methylation status of selected DEGs in CCE.

Results: We identified 39 DEGs in CCE with significant differences in gene expression levels between patients with and without excess of body weight ($p < 0.01$). A significant difference in methylation level of promoter region was observed between study groups in seven genes ($p < 0.05$): *SPOCK1* (NM_004598),

EDIL3 (NM_005711), *SEC31B* (NM_015490), *GPM6A* (NM_005277), *C2ORF40* (NM_032411), *ETV7* (NM_016135) and *MND1* (NM_032117). A significant association between *FGF10* methylation status and excess of body weight status was evident [OR=3.81, 95% CI; 1.36-10.68], considering the rest of the clinical and demographic characteristics of the population constant. Finally, the methylation status of the under-expressed gene *PRICKLE2* (NM_198859) was associated with overall survival (p=0.0022).

Conclusion: We identified differences in the methylation level of the promoter region of seven DEGs in CCE according to excess of body weight status in the CESC-TCGA cohort. What is more, the methylation status of the under-expressed gene *PRICKLE2* (NM_198859) was associated with overall survival. Our results suggest a role of methylation associated with excess of body weight in CCE and a possible implication in survival. This field of research is worth further exploration because of its potential clinical relevance.

Keywords: Excess of body weight, methylation, promoter region, cervical cancer, survival, CESC-TCGA

1 INTRODUCCIÓN

1.1 Planteamiento de la investigación

El cáncer cervical (CCE) es el cuarto tipo de cáncer más frecuente en mujeres a nivel global (1). El CCE registra una elevada mortalidad en países en vías de desarrollo, los cuales concentran el 90% de las muertes (1–3).

El desarrollo y progreso de la carcinogénesis cervical requiere de alteraciones genéticas y epigenéticas (4–6). La alteración epigenética más estudiada en CCE es la relacionada a la metilación del ADN (7). Se ha demostrado que los cambios de metilación del ADN están presentes en lesiones cervicales precursoras y malignas y se acumulan conforme progresa la enfermedad (8).

Un factor asociado al pronóstico de las pacientes con CCE es el estado del peso, definido por el Índice de Masa Corporal (IMC) (9–11), sin embargo los mecanismos moleculares que explican esta asociación aún son desconocidos. En otros tipos de neoplasias, como cáncer colorrectal, se han descrito alteraciones en el patrón global de metilación (metiloma) que distinguen a pacientes obesos de pacientes no obesos (12). Esto sugiere un vínculo entre el estado del peso del paciente con cáncer y alteraciones epigenéticas, lo cual podría tener un rol importante en la evolución de la neoplasia si dichas alteraciones ocurriesen en genes vinculados al cáncer.

En este contexto, el presente estudio planteó investigar si en CCE existe un patrón de metilación asociado al exceso de peso y si este influye en la supervivencia de las pacientes. Para ello se realizó un análisis secundario de datos genómicos y epigenómicos de una base de datos pública de CCE perteneciente al proyecto TCGA-CESC. Esta base de datos proviene de una de las cohortes de estudio de CCE más grandes desarrolladas hasta la fecha.

2 MARCO TEÓRICO

2.1 Epigenética: definición y mecanismos

La epigenética es definida como el estudio de los cambios heredables en la función de genes que no implican un cambio en la secuencia del ADN (13). Los mecanismos epigenéticos que regulan la expresión génica pueden ser agrupados en tres categorías: metilación del ADN, modificaciones post-traduccionales de histonas, remodelamiento de la cromatina y mecanismos de interferencia mediados por moléculas de ARN no codificantes (ncRNAs, por sus siglas en inglés) (14). El estado o perfil epigenético a nivel del genoma completo se denomina epigenoma (15).

2.2 Metilación del ADN

La metilación del ADN es la modificación epigenética más estudiada en humanos (15,16). Consiste en la adición covalente de un grupo metilo (-CH₃) en el carbono de la posición 5 de una citosina (C) que precede a una guanina (G), región que se denomina dinucleótido CpG o sitio CpG (15,17). Esta reacción ocurre luego de la síntesis del ADN y es realizada por una familia de enzimas denominadas ADN metiltransferasas (DNMTs por sus siglas en inglés) que catalizan la transferencia del grupo metilo de una S-adenosil-metionina (SAM) a la C (17,18) (Figura 1).

Los sitios CpG tienden a agruparse en regiones cortas o *clusters* denominadas islas CpG (CGIs, por sus siglas en inglés), definidas como regiones que contienen más de 200 bases con un contenido de G+C de al menos 50% y una razón (*ratio*) de frecuencias de CpG observadas a esperadas estadísticamente de al menos 0.6 (16). La proporción de sitios CpG en el genoma humano es más baja (~1%) de lo esperado (~4%) a partir de la abundancia total de C y G (42% de las bases del ADN)

(16,17). Las CGIs están asociadas a aproximadamente el 60% de las regiones promotoras de genes humanos (16).

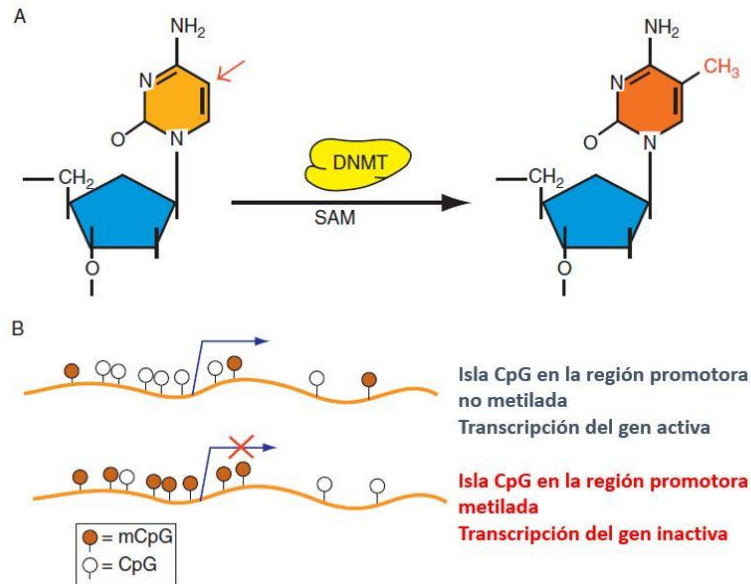


Figura 1 Metilación del ADN

(A) La adición covalente de un grupo metilo (-CH₃) en el carbono de la posición 5 de una citosina es catalizada por la enzima DNMT en presencia de SAM. (B) Los sitios CpG no metilados dentro de la región promotora de un gen no reprimen la transcripción, por lo que el gen se expresa. En contraste, cuando los sitios CpG dentro de esta región se encuentran metilados, se reprime la transcripción y el gen se silencia. Figura extraída de Kulis y Esteller, 2010 (18).

En células normales, la metilación del ADN ocurre predominantemente en regiones genómicas repetitivas, mientras que la mayoría de las CGIs se encuentran en un estado no metilado, lo que permite que los genes se expresen en presencia de activadores transcripcionales necesarios (17,19). Se estima que alrededor del 6% de CGIs se metilan de manera tejido-específica durante el desarrollo temprano o en tejidos no diferenciados (16,18).

La metilación de CGIs en la región promotora del gen está asociada a una disminución de la expresión génica, mientras que la metilación en regiones

intragénicas (cuerpo génico, 5' UTR y 3'UTR) se asocian a un incremento en la expresión (20).

2.3 Epigenética en cáncer

Las células cancerosas poseen además de mutaciones genéticas, alteraciones en el epigenoma conformados por cambios en la metilación del ADN, los patrones de modificación de histonas y los perfiles de expresión de enzimas modificadoras de la cromatina (16,21). El perfil epigenético aberrante en cáncer se ha asociado a la iniciación y progresión del cáncer (16,21).

Una de las primeras alteraciones epigenéticas en cáncer descritas fue la alteración en el patrón de metilación del ADN (15,21). En 1983, Feinberg y Vogelstein reportaron una reducción en el patrón de metilación del ADN de genes específicos en células de cáncer de colon humano en comparación a tejidos normales de colon (22). Asimismo, Gama-Sosa et al, también en 1983, describieron una reducción en el contenido de 5-metilcitosina del ADN en muestras tumorales en comparación a tejidos normales (23).

2.4 Metilación de ADN aberrante en cáncer

Cuando las células se vuelven cancerosas el patrón de metilación de ADN cambia sustancialmente debido a dos principales fenómenos: la hipometilación global del ADN y la hipermetilación en promotores de genes supresores de tumores (17).

2.4.1 Hipometilación global del ADN en cáncer

Las células cancerosas se caracterizan por una pérdida de metilación del ADN a nivel global que representa una disminución del 20-60% en el contenido total de 5-metil-citosina en comparación a un tejido normal (16,17). El grado de

hipometilación global del ADN se asoció a la progresión del cáncer a partir del estudio de Fraga et al (2004), en donde se describe que el grado de hipometilación de ADN incrementa a medida que una lesión benigna progresa a un cáncer invasivo (24).

Los mecanismos propuestos para explicar la contribución de la hipometilación global de ADN a la progresión del cáncer son tres: la generación de inestabilidad cromosomal, la reactivación de elementos transposables y la pérdida de *imprinting* (pérdida de expresión génica monoalélica específica del alelo parental) (15).

2.4.2 Hipermetilación en genes supresores de tumores

A pesar de la hipometilación global del ADN en cáncer, se ha identificado también la hipermetilación en CGIs en promotores de genes supresores de tumores, el cual es un evento clave en el origen de muchos tipos de cáncer (15). En las células cancerosas se ha observado la ganancia de metilación en CGIs que, en condiciones normales, no se encuentran metiladas, lo que conlleva al silenciamiento génico (21). Los genes que se ven afectados por la hipermetilación en CGIs están involucrados en una gama de rutas celulares importantes como el ciclo celular, la adherencia celular e invasión, apoptosis, metabolismo de carcinógenos, respuesta hormonal, señalización de Ras, microRNAs, entre otras (15–18).

Los patrones de hipermetilación son específicos para el tipo de tumor (16) y ocurren en diferentes etapas de la progresión del cáncer (18).

2.5 Cáncer cervical

El CCE es el cuarto tipo de cáncer más frecuente en mujeres (1). Se estima que en el 2018 hubo 570,000 nuevos casos a nivel mundial y 311,000 muertes por CCE (2,25). Alrededor del 90% de las muertes por CCE provienen de países de bajos y

medianos ingresos, en donde el acceso a un tamizaje efectivo de CCE es limitado (1,3). La elevada tasa de mortalidad por CCE en los países en vías de desarrollo se explica porque, con frecuencia, el CCE es detectado en estadios avanzados, en cuyo caso el pronóstico no es bueno (1). El alto número de muertes por CCE se podría reducir mediante programas de tamizaje y tratamientos efectivos (1,3).

El CCE ocurre en la parte más baja y estrecha del útero llamada cuello uterino o cérvix, el cual conecta el cuerpo del útero con la vagina (Figura 2).

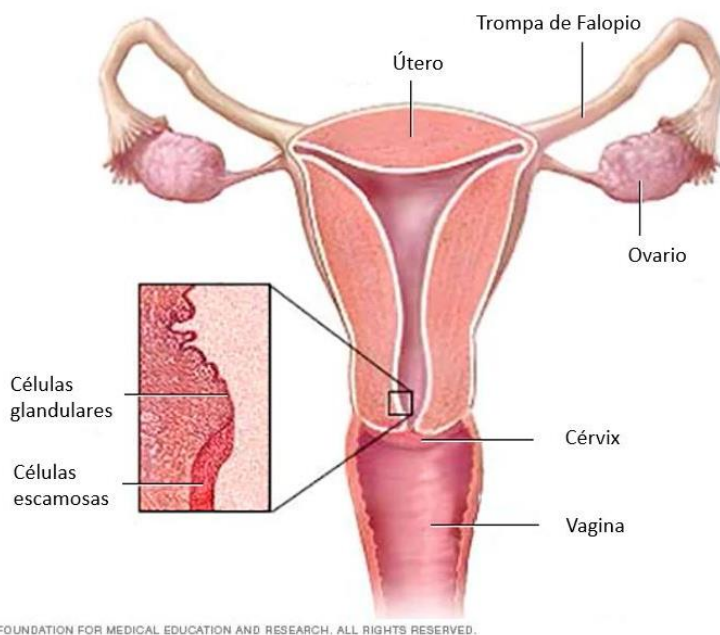


Figura 2 Anatomía del aparato reproductor femenino

El CCE ocurre en el cérvix, la parte más baja y estrecha del útero. Figura extraída de *Mayo Foundation for Medical Education and Research* (26).

Se sabe por estudios biológicos y epidemiológicos que la causa principal de CCE es el virus del papiloma humano (VPH), el cual está presente en aproximadamente el 100% de los casos de CCE (27,28). De las más de 100 cepas de VPH, solo algunas causan cambios anormales en las células del cérvix y conllevan al desarrollo de CCE. Alrededor del 70% de casos de CCE y de las lesiones precancerosas de cérvix

son producidos por las cepas de alto riesgo VPH 16 y VPH 18 (1). No obstante, una infección persistente de una cepa VPH de alto riesgo no es suficiente para immortalizar y transformar a las células epiteliales del humano, sino que se requieren de alteraciones genéticas y epigenéticas para el desarrollo de la carcinogénesis (4–6). Estos factores, en conjunto, alterarían el control del ciclo celular y conllevarían a que la célula hospedera adquiriera un fenotipo inmortal y progrese a un fenotipo maligno e invasivo (5).

2.6 Epigenética en cáncer cervical

Las alteraciones epigenéticas en CCE han sido estudiadas principalmente porque tienen el potencial de servir como biomarcadores para un diagnóstico temprano al igual que para el tratamiento de la enfermedad, ya que al no ser mutaciones genéticas estas pueden ser “revertidas” mediante la inhibición de las enzimas asociadas (29).

Se han identificado alteraciones epigenéticas en cada estadio de CCE (5). Estas alteraciones epigenéticas incluyen la hipometilación global del ADN, la hipermetilación de genes supresores de tumores, modificaciones de histonas y alteraciones en ncRNAs (5,29). La alteración epigenética más estudiada en CCE es la relacionada a la metilación del ADN (7).

En CCE se ha demostrado que los cambios de metilación de ADN se acumulan en relación al progreso de la enfermedad y han sido identificados en muestras de lesiones cervicales precursoras y malignas (8). Destaca el estudio de Kim et al (1994) en donde se reporta que el grado de hipometilación incrementa progresivamente con el avance de la neoplasia cervical (30). Además, se ha reportado que en CCE varios genes supresores de tumores presentan

hipermetilación en las regiones promotoras ricas en CGI (31). En este sentido, en CCE se reporta una variedad de perfiles de metilación alterados que conllevan a una anormal expresión de genes, activación de oncogenes y elementos transposables, pérdida de impronta y la inactivación de genes supresores de tumores (29,31).

2.7 Relación entre el exceso de peso y cáncer

El sobrepeso y la obesidad, definidas por un IMC mayor a 25 y 30, respectivamente, se asocian a una muerte prematura a través de enfermedades crónicas, incluyendo varios tipos de cáncer (32). Ambos son factores de riesgo para al menos 13 tipos de cáncer, entre los que se encuentra el cáncer de útero y de ovarios (33). La obesidad y el sobrepeso pueden causar cambios en el organismo que conllevan al cáncer a través de un incremento en los niveles de ciertas hormonas, factores de crecimiento e inflamación por parte de las células adiposas (34). Asimismo, la obesidad está asociada a un microambiente del tejido adiposo con alteraciones endocrinas que favorecerían la iniciación y progresión del tumor (35).

2.8 Relación entre el exceso de peso y cáncer cervical

Existe limitada evidencia sobre la asociación entre el sobrepeso/obesidad y el riesgo de desarrollar CCE (36). También se ha evidenciado que el peso influye en la sobrevivencia de pacientes con CCE (9–11) aunque los resultados son contradictorios. Por un lado, Clark et al (2016) encontraron que las pacientes con CCE con bajo peso (IMC<18.5) y con sobrepeso/obesidad (IMC>25) tuvieron una menor sobrevivencia que las pacientes de peso normal (IMC entre 18.9 y 24.9) (9). Mientras que, en un estudio de Xu et al (2015), las pacientes con CCE y con valores de IMC

más altos mostraron una menor mortalidad (10). Asimismo, se tiene el estudio en CCE de Grigsby et al (2018) en el que se reporta que las pacientes con obesidad severa ($IMC \geq 35$) tuvieron una mayor sobrevivencia que las pacientes de bajo peso ($IMC \leq 18.5$) y que las pacientes con un IMC entre 18.6 y 34.9 (11).

2.9 Metiloma en cáncer asociado al estado de exceso de peso

La metilación del ADN es modulada por factores ambientales (incluyendo factores nutricionales) y está influida por el estilo de vida (37). Por ejemplo, en una cohorte de individuos europeos se encontró patrones de metilación asociados al IMC (38), además, se sabe que cambios en el estilo de vida como la pérdida de peso pueden influenciar la metilación del ADN (39). Esto sugiere una asociación entre el estado de peso y el metiloma en humanos que, consecuentemente, podría tener un impacto en la expresión génica.

Por otro lado, estudios recientes han sugerido un posible rol del metiloma en la asociación entre el estado de exceso de peso y cáncer. En el 2017, Crujeiras et al describieron un metiloma asociado a obesidad y estado de menopausia en pacientes con cáncer de mama (40). Asimismo, en el 2018 se identificó un metiloma asociado a obesidad en cáncer colorrectal, el cual involucra genes que tienen un rol en procesos metabólicos e inflamatorios (12).

2.10 Estudio de la genómica del cáncer cervical mediante el proyecto *The Cancer Genome Atlas* (TCGA-CESC)

El proyecto The Cancer Genome Atlas (TCGA), es un esfuerzo colaborativo entre el National Cancer Institute (NCI) y el National Human Genome Research Institute (NHGRI) de EEUU cuyo propósito es catalogar y descubrir las principales alteraciones genómicas que desencadenan el cáncer a través de la aplicación de

herramientas de secuenciamiento de genomas en proyectos genómicos a gran escala (41). Los datos del TCGA son de libre acceso y están disponibles a través de diversas plataformas web y repositorios biológicos. Esto se debe a que la finalidad del TCGA es de que esta información genómica pueda ser de utilidad en la mejora de métodos de diagnóstico, tratamiento y prevención del cáncer (41,42).

El TCGA empezó en el año 2006 y ha recolectado información genómica en 33 tipos de cáncer, incluyendo el cáncer cervical (43). El estudio de CCE dentro del proyecto TCGA se denominó CESC, por las siglas en inglés de *Cervical squamous cell carcinoma and endocervical adenocarcinoma*. El estudio TCGA-CESC es uno de los estudios genómicos completos de cáncer cervical más grandes hasta la fecha, el cual cuenta con 308 pacientes. En el 2017, el TCGA publicó un artículo acerca de la caracterización molecular y genómica del CESC en la revista *Nature* (44).

2.11 Conceptos utilizados en estudios bioinformáticos

2.11.1 Análisis de expresión diferencial de genes

El transcriptoma es el conjunto total de moléculas de ARN (denominados también “transcritos”) en una célula o una población de células en un determinado tiempo o bajo cierta condición (45,46). Actualmente existen dos principales técnicas para el estudio de transcriptomas: los microarreglos (*microarrays*), los cuales cuantifican un conjunto de secuencias predeterminadas y, el secuenciamiento de ARN, conocido como *RNA-seq*, metodología que utiliza tecnologías de secuenciamiento masivo *Next-Generation Sequencing* (NGS, por sus siglas en inglés) (47,48).

En la última década, la técnica de RNA-seq, ha ido reemplazando a los microarreglos en el estudio de la expresión génica debido a que permite cuantificar la presencia y abundancia de ARNs en una muestra biológica con un mayor

rendimiento y cobertura (47,48). Además, a partir de los datos obtenidos con el RNA-seq, es posible realizar múltiples análisis, como el análisis de expresión diferencial de genes, el cual permite detectar genes que presentan un cambio estadísticamente significativo en sus niveles de expresión en las condiciones comparadas (por ejemplo, células con tratamiento y células sin el tratamiento o tejido tumoral y tejido sano) (48,49). La determinación de estos genes denominados “genes diferencialmente expresados” (DEGs, por sus siglas en inglés) se puede realizar con herramientas bioinformáticas de análisis de datos de RNA-seq o de manera manual con el análisis estadístico de conjuntos de datos (50).

En los análisis de identificación de DEGs, se analizan un número muy alto de genes (se estima de que en el genoma humano hay aproximadamente más de 20,000 genes), por lo cual se suele aplicar una corrección por pruebas múltiples, como la corrección de Bonferroni (50). Debido a que el número de DEGs seleccionados en los análisis tiende a seguir siendo elevado (usualmente >1000 DEGs), se emplean métodos para interpretar el significado de los cambios de expresión identificados como el análisis de enriquecimiento de rutas (por ejemplo, las rutas KEGG) o el análisis de enriquecimiento de términos de ontología génica (GO, por sus siglas en inglés) (45,50).

2.11.2 Ontología génica

La ontología génica (GO) es un recurso que ofrece información acerca de las funciones de los genes y de los productos génicos (51). La base de datos (BD) del GO está estructurada con una ontología formal basada en el uso de “clases” de funciones génicas (términos GO) que poseen relaciones establecidas entre ellas (51). La ontología cubre tres aspectos de la función génica: la función molecular

(MF, por sus siglas en inglés), el componente celular (CC, por sus siglas en inglés) y el proceso biológico (BP, por sus siglas en inglés):

- Función molecular (MF): comprende las actividades de un producto génico a nivel molecular. Ejemplo: la catálisis o el transporte.
- Componente celular (CC): comprende las localizaciones de estructuras celulares en donde un producto génico realiza una función. Ejemplo: la mitocondria o el ribosoma.
- Proceso biológico (BP): comprende los procesos que se llevan a cabo a partir de múltiples actividades moleculares. Ejemplo: la transducción de señales o la reparación del ADN.

2.11.3 Rutas KEGG

La Enciclopedia de Genes y Genomas de Kioto (KEGG, por sus siglas en inglés) es un recurso que integra 18 BDs con información genómica, información de sistemas, información química e información relacionada a la salud (52). El objetivo de KEGG es vincular la información de genes y genomas con información funcional a nivel molecular y de mayor orden (53). La BD principal de KEGG es KEGG PATHWAY, la cual es una colección de mapas de rutas manualmente elaboradas que representan información de redes de interacciones y relaciones moleculares. Los mapas de las rutas pueden pertenecer al metabolismo, procesamiento de información genética, procesamiento de información ambiental, procesos celulares, sistemas de órganos, enfermedades humanas o desarrollo de drogas.

3 JUSTIFICACIÓN DEL ESTUDIO

Se sabe que el exceso de peso (sobrepeso u obesidad) influye en el pronóstico de las pacientes con CCE, sin embargo, los mecanismos moleculares que explican esta asociación aún no se comprenden. La asociación podría estar mediada por una regulación epigenética en genes que tienen un rol en la progresión del CCE. Es por ello que el presente estudio tiene como objetivo evaluar la potencial asociación del exceso de peso y patrones de metilación de la región promotora en genes diferencialmente expresados (DEGs) en CCE a partir del análisis secundario de datos genómicos y epigenómicos del proyecto de libre acceso TCGA-CESC. El estudio hace uso de una extensa cantidad de datos provenientes de uno de los estudios en genómica de CCE más grandes hasta la fecha. El enfoque planteado permitirá generar información que podría ser de utilidad clínica, particularmente en la clasificación de tumores de CCE para un manejo personalizado de la enfermedad que considera el IMC de cada paciente.

4 PREGUNTA DE INVESTIGACIÓN

- ¿Existen diferencias en la metilación de la región promotora de genes diferencialmente expresados asociadas al estado de exceso de peso en pacientes con CCE del proyecto TCGA-CESC? Si existen esas diferencias, ¿ellas están asociadas a la sobrevida global?

5 HIPÓTESIS

El estado de metilación de la región promotora de los DEGs en pacientes con CCE del proyecto TCGA-CESC es distinto entre pacientes que tienen un exceso de peso y las que no tienen un exceso de peso y ello influye en la sobrevida global.

6 OBJETIVOS

6.1 Objetivo principal

- Determinar la asociación entre el exceso de peso y la metilación de la región promotora de DEGs en pacientes con CCE del proyecto TCGA-CESC a partir de un análisis secundario de datos y determinar la asociación entre el estado de metilación y la sobrevida global.

6.2 Objetivos específicos

1. Identificar DEGs que presentan una diferencia de expresión entre las pacientes sin exceso de peso y con exceso de peso y evaluar la potencial asociación de la metilación (nivel y estado de metilación) en la región promotora y el estado de exceso de peso
2. Evaluar la potencial asociación de características clínicas y demográficas y el estado de metilación de la región promotora de DEGs cuya expresión y

metilación es diferente entre las pacientes sin exceso de peso y con exceso de peso

3. Evaluar la potencial asociación de la sobrevida global y el estado de metilación de la región promotora de DEGs cuya expresión es diferente entre las pacientes sin exceso de peso y con exceso de peso

7 METODOLOGÍA

7.1 Diseño del estudio

El presente estudio es observacional, analítico y de corte transversal, basado en un análisis de datos secundarios.

7.2 Población y muestra de estudio

Se trabajó con pacientes diagnosticadas con CCE del proyecto TCGA-CESC cuyos datos clínicos y genómicos se encontraban disponibles en los repositorios de datos cBioPortal, UCSC Xena y MethHC.

7.3 Criterios de selección

7.3.1 Criterios de inclusión

- Pacientes mujeres diagnosticadas con CCE primario, mayores de 18 años, cuya información de genotipo y fenotipo estaba disponible en cBioPortal.
- Con código de identificación de la paciente, código de la muestra biológica, subtipo histológico, peso y talla

7.3.2 Criterios de exclusión

- Pacientes con diagnóstico previo de cáncer
- Pacientes con tipo de muestra metastásica

7.4 Operacionalización de variables

7.4.1 Características clínicas y demográficas

En la Tabla 1 se describe la operacionalización de las variables correspondientes a las características clínicas y demográficas de la población.

Tabla 1 Operacionalización de variables de características clínicas y demográficas

Nombre de la variable	Definición operacional	Valores	Tipo de variable según su naturaleza	Escala de medición	Fuente
Código de la paciente	Código de identificación de la paciente de CESC	Ejemplo: TCGA-EK-A2PI	Cualitativa politómica	Nominal	cBioPortal
Código de la muestra biológica	Código de identificación de la muestra biológica de CESC	Ejemplo: TCGA-EK-A2PI-01	Cualitativa politómica	Nominal	cBioPortal

Peso	Peso de la paciente (en kilogramos) al momento del diagnóstico de cáncer	Mayores a 0	Cuantitativa discreta	Razón	cBioPortal
Talla	Estatura de la paciente (en centímetros) al momento del diagnóstico de cáncer	Mayores a 0	Cuantitativa discreta	Razón	cBioPortal
IMC	Peso de la paciente (en kilogramos) dividido por la estatura (en metros) elevada al cuadrado	Mayores a 0	Cuantitativa continua	Razón	Calculado a partir del peso y la talla
Estado de exceso de peso	Categorización basada en el IMC para definir la presencia o ausencia de exceso de peso	<ul style="list-style-type: none"> • Sin exceso de peso ($IMC < 25$) • Con exceso de peso ($IMC \geq 25$) 	Cualitativa dicotómica	Ordinal	Calculado a partir del IMC
Edad	Años cumplidos al momento del diagnóstico de cáncer	18 años a más	Cuantitativa discreta	Razón	cBioPortal

Subtipo histológico	Histología del tumor	<ul style="list-style-type: none"> • Carcinoma de células escamosas • Adenocarcinoma endocervical • Carcinoma mucinoso • Otros 	Cualitativa politómica	Nominal	cBioPortal
Estadio clínico	Estadio clínico según los sistemas de clasificación FIGO y AJCC	<ul style="list-style-type: none"> • Estadio I o II (estadio temprano) • Estadio III o IV (estadio avanzado) 	Cualitativa dicotómica	Ordinal	cBioPortal
Estado de infección por VPH	Estado de infección por cepas de VPH	<ul style="list-style-type: none"> • VPH 16 o 18 • Otras cepas de VPH • Negativo 	Cualitativa politómica	Nominal	Publicación de CCE del TCGA (2017) (44)
Raza	Raza reportada por la paciente	<ul style="list-style-type: none"> • Blanca • Negra afroamericana^o • Asiática • Otras 	Cualitativa politómica	Nominal	cBioPortal

Historia de tabaquismo	Categoría que describe el estado actual y el historial de tabaquismo según lo informado por la paciente	<ul style="list-style-type: none"> • No fumadora de por vida • Actual fumadora • Actual fumadora reformada 	Cualitativa politómica	Nominal	cBioPortal El significado de la codificación se encontró en la herramienta CDE Curation Tool (54)
Historia de uso de anticonceptivos hormonales	Categoría que indica si la paciente ha usado o usa anticonceptivos hormonales	<ul style="list-style-type: none"> • Nunca ha usado • Actual usuaria • Antigua usuaria 	Cualitativa politómica	Nominal	cBioPortal
Estado de menopausia	Estado de menopausia en el momento en que la paciente fue diagnosticada con cáncer	<ul style="list-style-type: none"> • Pre-menopausia • Post-menopausia • Indeterminado 	Cualitativa politómica	Nominal	cBioPortal
Sobrevida global	Sobrevida global luego del diagnóstico de cáncer, expresado en meses	0 a más	Cuantitativa continua	Razón	cBioPortal

Estado de sobrevida global	Estado de sobrevida general	<ul style="list-style-type: none"> • Fallecido • Vivo 	Cualitativa dicotómica	Nominal	cBioPortal
----------------------------	-----------------------------	---	------------------------	---------	------------

7.4.2 Datos genómicos y epigenómicos

En la Tabla 2 se describe la operacionalización de las variables correspondientes a los datos genómicos y epigenómicos.

Tabla 2 Operacionalización de variables de datos genómicos y epigenómicos

Nombre de la variable	Definición operacional	Valores	Tipo de variable según su naturaleza	Escala de medición	Fuente
Nombre del gen	Nombre del DEG identificado	Ejemplo: PLA2G2A	Cualitativa politómica	Nominal	GEPIA

Categoría de expresión	Categoría que indica el estado de expresión del gen en el tumor primario de cérvix en comparación a la expresión en el tejido sano	<ul style="list-style-type: none"> • Sub-expresado • Sobre-expresado 	Cualitativa dicotómica	Nominal	GEPIA
Nivel de expresión	Valores de expresión expresados como el recuento normalizado de valores RSEM transformados a $\log_2(x+1)$, donde x es el valor RSEM	0 a más	Cuantitativa continua	De razón	UCSC Xena
Nivel de metilación	Valores de metilación expresados como coeficientes β	Valores entre 0 y 1	Cuantitativa continua	De razón	MethHC
Estado de metilación	Categoría que describe el estado de metilación basada en un punto de corte referencial (55,56)	<ul style="list-style-type: none"> • Hipometilado ($\beta \leq 0.1$) • Metilación normal ($0.1 < \beta < 0.3$) • Hipermetilado ($\beta \geq 0.3$) 	Cualitativa politómica	Ordinal	Basado en el nivel de metilación

7.5 Instrumentos para la recolección de datos

En la Tabla 3 se describen las herramientas web empleadas para la recolección de datos en el estudio.

Tabla 3 Herramientas web utilizadas para la recolección de datos

Nombre	Descripción	Herramienta utilizada	Utilidad en el estudio	Periodo de consulta	Sitio web	Referencia
cBioPortal	Recurso web de acceso abierto para la exploración interactiva de conjuntos de datos genómicos en cáncer (incluyendo datos del TCGA)	<i>Clinical data</i>	Obtención de datos clínicos y demográficos	Marzo - Abril 2019	https://www.cbioportal.org/	Cerami <i>et al.</i> , 2012 (57)
GEPIA	Recurso web interactivo para el análisis de datos de expresión genética, en muestras de proyectos del TCGA y GTEx	<i>Differential Expression Analysis</i>	Identificación de DEGs en TCGA-CESC	Junio 2019	http://gepia.cancer-pku.cn/	Tang <i>et al.</i> , 2017 (58)

UCSC Xena	Recurso web interactivo para la exploración de conjuntos de datos ómicos públicos (incluyendo TCGA, GDC y GTEx) y privados	<i>Xena Functional Genomics Explorer</i>	Obtención de niveles de expresión de los DEGs	Julio 2019 (TCGA-CESC), Setiembre 2019 (TCGA TARGET GTEx)	http://xena.ucsc.edu/	Goldman <i>et al.</i> , 2019 (59)
MethHC	Recurso web que permite acceder a una gran colección de datos de metilación de ADN y perfiles de expresión de mRNA/microRNA del TCGA	<i>Gene search</i>	Obtención de niveles de metilación en la región promotora	Octubre 2019	http://MethHC.mbc.nctu.edu.tw/	Huang <i>et al.</i> , 2015 (60)
STRING	BD biológica y recurso web interactivo que permite evaluar redes de PPI y enriquecimientos funcionales en listas de proteínas proporcionadas por el usuario	<i>Multiple proteins</i>	Análisis de redes de PPI y enriquecimiento funcional	Setiembre 2019	https://string-db.org/	Szklarczyk <i>et al.</i> , 2018 (61)

7.6 Procedimientos y técnicas

En el flujograma de la Figura 3 se resumen los procedimientos realizados en el estudio.

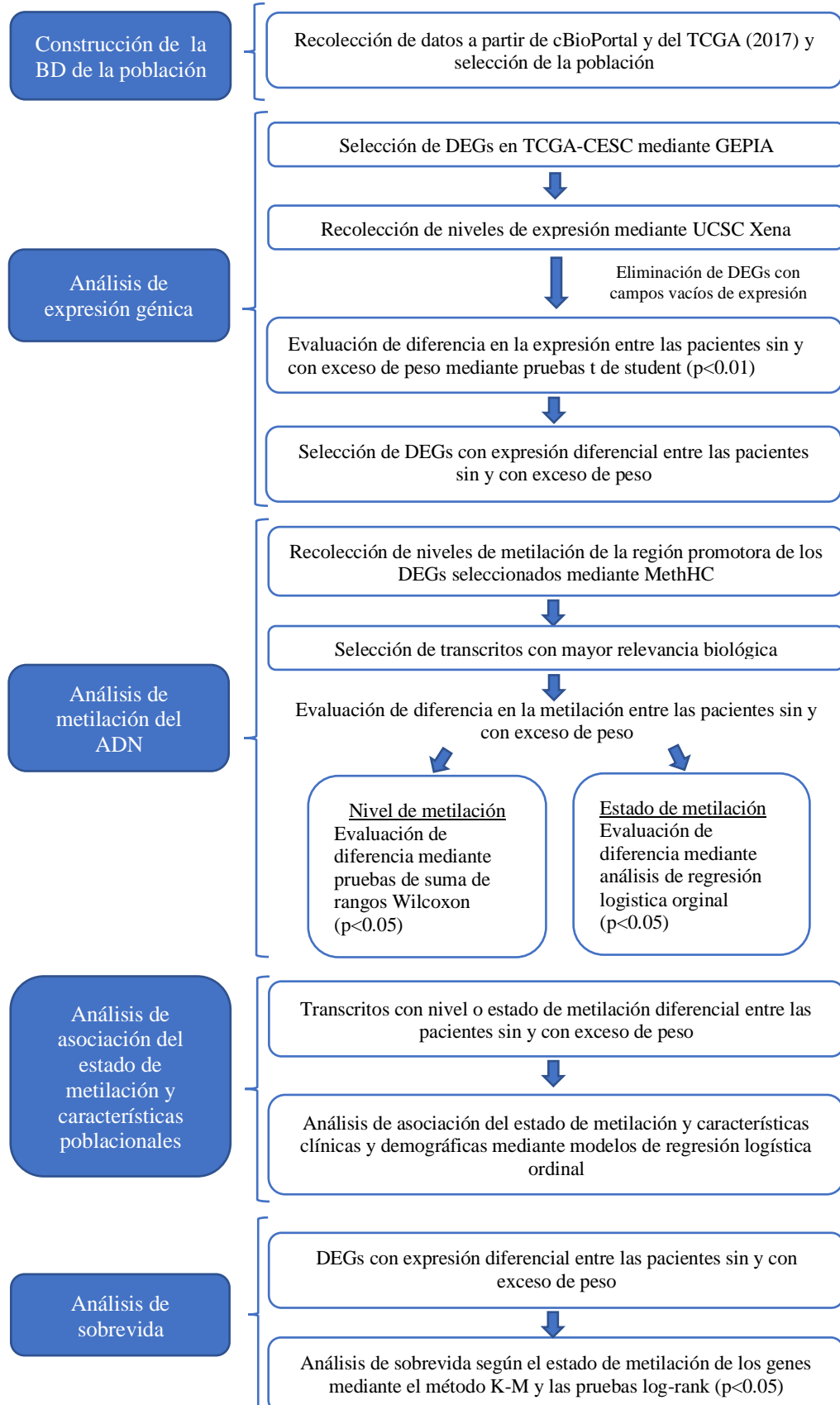


Figura 3 Flujograma de la metodología del estudio

7.6.1 Construcción de BD de la población

Los datos de las variables clínicas y demográficas de las pacientes del estudio fueron recolectados de la Base de datos provisional TCGA-CESC (*Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Firehose Legacy)*) disponible en cBioPortal. Los datos del Estado de infección por VPH fueron obtenidos del material suplementario del estudio del TCGA en CCE del 2017 (44). Los datos fueron incorporados en una BD con el programa estadístico Stata SE 15.0. El IMC fue calculado a partir de los valores de peso y talla recolectados con cBioPortal (Tabla 1) como el peso de la paciente (en kg) dividido por su estatura (en m) elevada al cuadrado. El IMC fue categorizado usando el criterio de la OMS (62): bajo peso ($IMC < 18.50$), peso normal ($18.50 \leq IMC < 25$), sobrepeso ($25 \leq IMC < 30$) y obesidad ($IMC \geq 30$). Luego, para evaluar el efecto de un exceso de peso, las pacientes fueron clasificadas en dos grupos incluyendo mujeres sin exceso de peso (bajo peso y peso normal) o con exceso de peso (sobrepeso y obesidad) (12).

7.6.2 Selección de DEGs en TCGA-CESC

La identificación de los DEGs en TCGA-CESC se realizó mediante la herramienta *Differential Expression Analysis* de GEPIA, la cual se encuentra bajo el menú de *Cancer Type Analysis*. Se seleccionó la base de datos CESC y se utilizaron los siguientes parámetros: $|\text{Log}_2\text{FC}|$ cutoff = 2, un valor de corte para el valor de $q = 0.01$ y el método de análisis diferencial de ANOVA. Los resultados obtenidos incluyeron el símbolo del gen, el código ID del gen, las medianas de expresión del tejido tumoral de cérvix y del tejido sano de cérvix, el $\text{Log}_2(\text{FC})$ y el valor de p de la prueba ANOVA con corrección por pruebas múltiples según el método de tasa

de descubrimiento falso (FDR, por sus siglas en inglés). Todos los datos fueron recolectados manualmente e ingresados a una BD en Microsoft Excel.

7.6.3 Construcción de la BD de expresión de los DEGs

7.6.3.1 BD de expresión de los DEGs en TCGA-CESC

Debido a que GEPIA no provee información de niveles de expresión génica de los DEGs por paciente, se empleó otro recurso web denominado UCSC Xena para la obtención de dicha información. Se utilizó la herramienta *Xena Functional Genomics Explorer* y se seleccionó el estudio *TCGA Cervical Cancer (CESC)* y las siguientes variables fenotípicas: Código de muestra de la paciente, Código de la muestra biológica y Tipo de muestra.

Luego, se seleccionaron únicamente las muestras de tumor primario utilizando la herramienta de filtro (*Filter actions*) en la barra de búsqueda del UCSC Xena. Esta herramienta se encuentra en la parte superior del portal y permite seleccionar muestras según un término o código. En este caso, se conservaron aquellas muestras que fueran del tipo tumor primario utilizando el término en inglés “Primary Tumor”.

Para la variable de expresión, se escogió *gene expression RNAseq - IlluminaHiSeq*. Esta variable muestra las estimaciones de transcripción a nivel del gen como el recuento normalizado de valores RSEM transformados a $\log_2(x+1)$, donde x es el valor RSEM. Los datos de expresión génica provienen de la medición experimental a través de la plataforma *Illumina HiSeq 2000 RNA Sequencing* elaborada por el Centro de Caracterización del Genoma del TCGA de la University of North Carolina.

Se ingresaron los nombres de los DEGs identificados previamente (sección 7.6.2) y se descargaron los datos de expresión como archivos .tsv. Luego, estos fueron importados y manejados en Stata en archivos de aproximadamente 200 DEGs cada uno, debido a la gran cantidad de datos.

Finalmente, se procedió a eliminar a los DEGs cuyos valores de expresión estaban conformados por campos vacíos utilizando el paquete *missings* en el programa Stata. Los valores de expresión por paciente fueron enlazados con los valores de IMC recolectados en la BD de la población elaborada en la Sección (7.6.1).

7.6.3.2 BD de expresión de DEGs seleccionados en TCGA TARGET GTEx

Se analizó exploratoriamente la expresión en los DEGs seleccionados por su asociación con el estado de obesidad (Sección 7.7.1.2.1) según el tipo de muestra (tumoral y tejido normal). Este análisis tuvo la finalidad de corroborar el estado de expresión de los genes en tejido tumoral de cérvix en comparación a un tejido sano obtenido en GEPIA. Debido a que el proyecto TCGA-CESC solo cuenta con 3 muestras de tejido sano de cérvix, se empleó el estudio *TCGA TARGET GTEx* en UCSC Xena para recolectar los datos de expresión de DEGs seleccionados en muestras de tejido sano y compararlas con la expresión en tejido tumoral. Los datos de este estudio provienen del *UCSC RNA-seq Compendium* y comprende muestras tanto del TCGA como del GTEx (repositorio de expresión de tejidos sanos). Las muestras del estudio *TCGA TARGET GTEx* han sido reanalizadas siguiendo una misma metodología de análisis de RNA-seq, lo que posibilita la comparación de la expresión de las muestras tumorales de cérvix (del TCGA-CESC) y las muestras de tejido sano (del GTEx) dentro del mismo estudio.

El procedimiento para la construcción de la BD de expresión fue similar al utilizado para la BD de expresión en TCGA-CESC en la sección anterior (7.6.3.1). Se seleccionó, en Xena, la base de datos *TCGA TARGET GTEX* y las variables: Estudio, Tipo de muestra, Sitio primario, Código de muestra de la paciente o del sujeto sano, Código de la muestra biológica y Tipo de muestra.

A continuación, se seleccionaron únicamente las muestras de cérvix con la herramienta de filtro como se describió en la sección anterior (7.6.3.1), esta vez usando como término de selección de muestras la palabra “cervix”.

Luego, se seleccionó la variable de expresión *gene expression RNAseq - RSEM norm_count*, cuya unidad de medida es igual a la descrita en la sección anterior (7.6.3.1). Finalmente, se ingresaron los nombres de los DEGs seleccionados y se descargaron los datos en un archivo .tsv. Los archivos fueron importados y manejados en Stata.

7.6.4 Redes de PPI y enriquecimiento funcional

Se determinaron las interacciones entre las proteínas codificadas por los genes con diferencia de expresión entre los grupos de estudio, consultando la base de datos STRING. El análisis se ejecutó con la herramienta *Multiple Proteins by Names/Identifiers*, para lo cual se ingresaron los nombres de los genes y se seleccionó el organismo *Homo sapiens*. Se empleó un puntaje mínimo de interacción requerida de 0.400 (confianza media) y los bordes de la red se configuraron por colores que indican el tipo de interacción evidenciada. Se construyeron redes de PPI cuya significancia estadística estuvo dada por un valor de p. Asimismo, se evaluaron enriquecimientos funcionales según los sistemas de

clasificación de GO y KEGG. La significancia estadística de cada enriquecimiento funcional estuvo dada por un valor de p corregido por el método FDR.

7.6.5 Construcción de la BD de metilación del ADN

Para la recolección de los valores de metilación de la región promotora de los genes seleccionados se utilizó la herramienta *Gene search* de MethHC. Se seleccionaron las siguientes opciones en el portal: estudio *CESC*, región génica *Promoter* y método de evaluación de nivel de metilación *Median*.

En el caso de que el portal tuviera disponible los datos de más de un transcrito por gen, se escogieron los transcritos que codifican a la isoforma canónica o predominante según lo reportado por las bases de datos NCBI, Ensembl y UniProt.

Los valores de metilación están expresados como coeficientes β y fueron obtenidos mediante un *Beadchip Illumina Infinium HumanMethylation 450 Dell TCGA* (44).

Los valores de β van de 0 (no metilado) a 1 (completamente metilado) y son calculados mediante la siguiente fórmula:

$$\beta = \frac{M}{(U + M + 100)}$$

donde:

M = Intensidad de la sonda metilada

U = Intensidad de la sonda no metilada

El valor de metilación en cada muestra estuvo dada por la mediana del coeficiente β de todas las sondas que mapean la región promotora del transcrito (63). Los valores para cada transcrito fueron descargados como archivos .txt, los cuales luego fueron convertidos a formato .dta en Stata y finalmente fueron ingresados en una sola BD de metilación.

Adicionalmente, se crearon variables para el estado de metilación (variable categórica ordinal) para los transcritos seleccionados a partir de los respectivos

niveles de metilación (variable cuantitativa continua). La clasificación del estado de metilación de cada gen en CCE se basó en puntos de cortes del coeficiente β reportados y empleados en otros estudios de epigenómica del cáncer (55,56). Se consideró como hipometilado si $\beta \leq 0.1$, metilación normal si $0.1 < \beta < 0.3$ e hipermetilado si $\beta \geq 0.3$. Esta clasificación fue empleada en lugar de una clasificación basada en controles de tejido sano de cérvix debido a que el estudio TCGA-CESC solo cuenta con tres muestras sanas con información epigenómica.

7.7 Plan de análisis

7.7.1.1 Análisis descriptivo de las características clínicas y demográficas

Se compararon las características clínicas y demográficas entre los grupos de estudio. Las variables cualitativas se expresaron como n (%) y se evaluó su asociación con el exceso de peso mediante la prueba exacta de Fisher o Chi². Las variables cuantitativas se expresaron como media (desviación estándar) y fueron evaluadas mediante la prueba t de student si presentaron distribución normal, de lo contrario se expresaron como mediana [rango intercuartil] y fueron evaluadas con la prueba de suma de rangos Wilcoxon. La tabla de resumen de los datos y los análisis estadísticos fueron realizados con el paquete *compareGroups* (64) en el programa R. El nivel de significancia estuvo dado por un valor de $p < 0.05$.

7.7.1.2 Análisis de expresión génica

7.7.1.2.1 Análisis de expresión por estado de exceso de peso

Se comparó el nivel de expresión entre las muestras de cérvix tumorales de las pacientes con exceso de peso y sin exceso de peso para los DEGs identificados en la Sección 7.6.3.1. Para ello se aplicaron pruebas t de student. El nivel de significancia estuvo dado por un valor de $p < 0.01$. Se elaboró un código (*script*) en

Stata para el análisis estadístico de los 1695 DEGs según el estado de exceso de peso. El código consistió en generar un bucle que permita ejecutar de manera automatizada las pruebas t de student y guardar los principales resultados obtenidos, como el valor de p, la media y la desviación estándar para cada grupo de estudio. El código se basó en el uso de los comandos *post* en Stata y se encuentra detallado en el Anexo 1.

7.7.1.2.2 Nivel de expresión por tipo de muestra

Para comparar el nivel de expresión de muestras tumorales de cérvix y muestras de tejido sano, se evaluó la diferencia de expresión basada en las medias mediante pruebas t de student. El nivel de significancia estuvo dado por un valor de $p < 0.05$. Se empleó un código de análisis en Stata similar al previamente descrito (Anexo 2)

7.7.1.3 Análisis de metilación del ADN

7.7.1.3.1 Nivel y estado de metilación en función del estado de exceso de peso

Se comparó el nivel de metilación de la región promotora de los transcritos seleccionados, entre las muestras de cérvix tumorales de las pacientes con exceso de peso y sin exceso de peso. Para ello se aplicaron pruebas de suma de rangos Wilcoxon. El nivel de significancia estuvo dado por un valor de $p < 0.05$. Los análisis fueron realizados en el programa R.

Asimismo, se evaluó la asociación del estado de metilación de los transcritos con el estado de exceso de peso mediante modelos de regresión logística ordinal. El nivel de significancia estuvo dado por un valor de $p < 0.05$ y los intervalos de confianza (IC) fueron reportados al 95%. El análisis estadístico se realizó con el programa Stata y se crearon gráficos utilizando el programa Microsoft Excel.

7.7.1.3.2 Nivel de metilación por tipo de muestra

Se comparó el nivel de metilación de la región promotora de los transcritos seleccionados, entre muestras tumorales y muestras sanas, mediante pruebas de los rangos con signo de Wilcoxon (prueba no paramétrica para muestras pareadas). Los niveles de metilación fueron representados en diagramas de caja según el tipo de muestra. El nivel de significancia estuvo dado por un valor de $p < 0.05$. Los análisis fueron realizados en el programa R y los gráficos creados con el paquete *ggplot2* (65)

7.7.1.4 Asociación del estado de metilación con características clínicas y demográficas

Se seleccionaron los transcritos que presentaron una asociación entre el nivel de metilación o estado de metilación con el estado de exceso de peso (descrito en la Sección 7.7.1.3.1) y se evaluó en ellos la asociación del estado de metilación con características clínicas y demográficas mediante modelos de regresión logística ordinal. La construcción de los modelos se realizó de manera independiente para cada transcrito, donde la variable dependiente fue el estado de metilación y las variables independientes correspondieron a características clínicas y demográficas. Las variables independientes con más de dos categorías fueron dicotomizadas debido al bajo número de casos en algunas categorías. Las variables independientes para los análisis fueron edad, tipo de CCE (Carcinoma de células escamosas, Otros), estadio clínico del CCE (Estadio I o II/Estadio III o IV), estado de infección por VPH (Infección por VPH 16 ó 18/Infección por otras cepas), raza (Blanca, Otras), historia de tabaquismo (no fumadora de por vida/fumadora antigua o actual),

historia de uso de anticonceptivos hormonales (nunca ha usado/usuario antigua o actual) y estado de menopausia (pre-menopausia/post-menopausia).

Primero, se construyeron modelos bivariados para cada factor. Luego, se construyeron modelos multivariados considerando los factores asociados significativamente con el estado de metilación en el modelo anterior ($p < 0.05$). El nivel de significancia fue dado por un valor de $p < 0.05$ y los intervalos de confianza (IC) fueron reportados al 95%. Todo el análisis estadístico se realizó utilizando el programa Stata.

7.7.1.5 Sobrevida global en función del estado de metilación

Se utilizó el método de Kaplan-Meier (K-M) para crear las gráficas de sobrevida global según el estado de metilación de los transcritos de aquellos DEGs que mostraron diferencia de expresión asociada al estado de exceso de peso (Sección 7.7.1.2.1) y la prueba de log-rank para evaluar diferencias entre las curvas de sobrevida. El nivel de significancia fue dado por un valor de $p < 0.05$. Se excluyeron del análisis aquellos transcritos con más del 97% de datos para una sola categoría en el estado de metilación. El análisis estadístico se realizó utilizando el programa Stata.

7.8 Consideraciones éticas

El proyecto fue exento de aprobación por el Comité Institucional de Ética de la Universidad Peruana Cayetano Heredia (SIDISI 103802), con fecha 06 de marzo del 2019. La confidencialidad de las participantes se mantuvo durante todo el estudio ya que las pacientes están identificadas exclusivamente por códigos elaborados por el Proyecto TCGA-CESC.

8 RESULTADOS

8.1 Características de la población

El proyecto TCGA-CESC en el repositorio cBioPortal comprende 308 pacientes con CCE. Se evaluó la elegibilidad de las 308 pacientes del estudio, de las cuales fueron excluidas 54 incluyendo a dos que presentaban tipos de muestras metastásicas, ocho que tuvieron diagnóstico previo de cáncer, una que no presentaba dato de diagnóstico previo de cáncer, 29 que no tuvieron registro de peso y 14 que no tuvieron registro de talla. En consecuencia, se recolectaron datos de las características clínicas y demográficas de 254 pacientes.

En la Tabla 4 se presentan las características generales de la muestra según el grupo de estudio (sin exceso de peso y con exceso de peso) y a nivel total. El número de de pacientes con CCE que presentaban exceso de peso (n=157, 61.8%) fue mayor que el de las pacientes sin exceso de peso (n=97, 38.2%). Hubo una diferencia significativa en el IMC entre los grupos, lo cual concuerda con lo esperado ya que la clasificación se basó en un punto de corte de IMC igual a 25.

Las pacientes tuvieron una edad promedio de 48.1 ± 13.3 años al momento del diagnóstico del CCE. La mediana de la sobrevida global fue de 23.6 [13.7;40.9] meses y el 79.9% de pacientes estaban vivas. La mayoría presentó el subtipo histológico de carcinoma de células escamosas (n=209, 82.3%) y un estadio clínico temprano del CCE (estadio I o II) (n=194, 76.4%). Asimismo, el 43.7% presentó infección por VPH 16 ó 18 (n=111). El 39% de las pacientes no presentó información para esta característica.

La mayoría de las pacientes fueron de raza blanca (n=183, 72%). Hubo una diferencia significativa en la raza entre los grupos ($p < 0.001$), probablemente debido a que la fracción de pacientes de raza asiática es mucho mayor en el grupo sin exceso de peso (n=16, 16.5%) que en el grupo con exceso de peso (n=3, 1.9%).

Más de la mitad reportó ser no fumadora de por vida (n=131, 51.6%). No se contaron con datos acerca de la historia del uso de anticonceptivos hormonales en una fracción importante de la población (41.2%) y, cuando sí se conoció dicha información, la categoría más frecuente fue la de nunca haber usado anticonceptivos hormonales (n=84, 33.1%). Asimismo, la mayoría se encontraba en la pre-menopausia (n=132, 52%).

A excepción de la raza, no hubo diferencias en las variables demográficas o clínicas entre ambos grupos ($p > 0.05$).

Tabla 4 Características generales de las pacientes por grupo de estudio

	Total (n=254)	IMC<25 (n=97)	IMC≥25 (n=157)	P
IMC (kg/m²)	28.0 (7.62)	21.3 (2.45)	32.0 (6.82)	<0.001
Edad (años)	48.1 (13.3)	49.2 (14.2)	47.5 (12.8)	0.345
Sobrevida global (meses)	23.6 [13.7;40.9]	21.4 [13.6;36.2]	25.0 [14.0;44.6]	0.533
Estado de sobrevida				0.195
Vivo	203 (79.9%)	73 (75.3%)	130 (82.8%)	
Fallecido	51 (20.1%)	24 (24.7%)	27 (17.2%)	
Subtipo histológico				0.54
Carcinoma de células escamosas	209 (82.3%)	81 (83.5%)	128 (81.5%)	
Adenocarcinoma endocervical	25 (9.84%)	7 (7.22%)	18 (11.5%)	
Carcinoma mucinoso	15 (5.91%)	6 (6.19%)	9 (5.73%)	
Otros	5 (1.97%)	3 (3.09%)	2 (1.27%)	
Estadio clínico				0.071
Estadio I o II	194 (76.4%)	67 (69.1%)	127 (80.9%)	
Estadio III o IV	55 (21.7%)	27 (27.8%)	28 (17.8%)	

No documentado	5 (1.97%)	3 (3.09%)	2 (1.27%)	
Estado de infección por VPH				0.955
VPH 16 ó 18	111 (43.7%)	38 (39.2%)	73 (46.5%)	
Otras cepas	36 (14.2%)	13 (13.4%)	23 (14.6%)	
Negativo	8 (3.15%)	3 (3.09%)	5 (3.18%)	
No documentado	99 (39.0%)	43 (44.3%)	56 (35.7%)	
Raza				<0.001
Blanca	183 (72.0%)	61 (62.9%)	122 (77.7%)	
Negra o afroamericana	21 (8.27%)	8 (8.25%)	13 (8.28%)	
Asiática	19 (7.48%)	16 (16.5%)	3 (1.91%)	
Otras	8 (3.15%)	2 (2.06%)	6 (3.82%)	
No documentado	23 (9.06%)	10 (10.3%)	13 (8.28%)	
Historia de tabaquismo				0.084
No fumadora de por vida	131 (51.6%)	53 (54.6%)	78 (49.7%)	
Actual fumadora	48 (18.9%)	20 (20.6%)	28 (17.8%)	
Actual fumadora reformada	44 (17.3%)	10 (10.3%)	34 (21.7%)	
No documentado	31 (12.2%)	14 (14.4%)	17 (10.8%)	
Historia de uso de anticonceptivos hormonales				0.286
Nunca ha usado	84 (33.1%)	34 (35.1%)	50 (31.8%)	
Antigua usuaria	49 (19.3%)	22 (22.7%)	27 (17.2%)	
Actual usuaria	14 (5.51%)	3 (3.09%)	11 (7.01%)	
No documentado	107 (42.1%)	38 (39.2%)	69 (43.9%)	
Estado de menopausia				0.787
Pre-menopausia	132 (52.0%)	50 (51.5%)	82 (52.2%)	
Post-menopausia	72 (28.3%)	27 (27.8%)	45 (28.7%)	
Indeterminado	2 (0.79%)	0 (0.00%)	2 (1.27%)	
No documentado	48 (18.9%)	20 (20.6%)	28 (17.8%)	

8.2 Selección de DEGs en TCGA-CESC

A través de la herramienta web GEPIA, se identificaron inicialmente 2042 DEGs en CCE, de los cuales 1264 fueron DEGs sub-expresados y 778 DEGs sobre-expresados. Luego de la descarga de los datos de expresión a través de la herramienta web UCSC Xena, se eliminaron 347 DEGs cuyos valores de expresión contenían campos vacíos. El número final de DEGs a analizar fue de 1695.

8.3 Análisis de expresión génica

8.3.1 Análisis de expresión por estado de exceso de peso

Se identificaron 39 DEGs que mostraron diferencias significativas ($p < 0.01$) de expresión génica entre las pacientes sin exceso y con exceso de peso, lo cual representa el 2.3% del total de DEGs. En la Figura 4 se presenta un flujograma del procesamiento realizado. De los 39 genes seleccionados, 19 fueron sub-expresados y 20 fueron sobre-expresados.

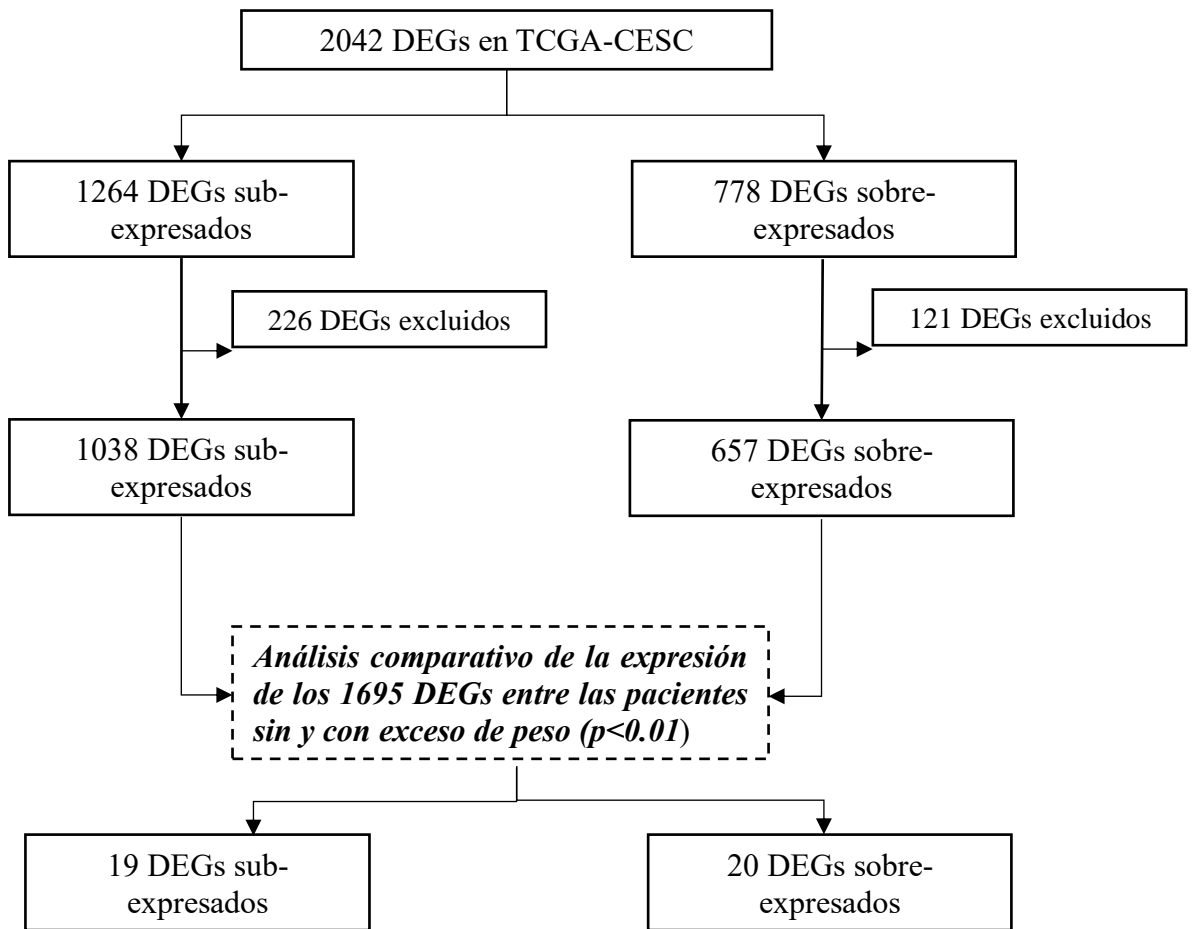


Figura 4 Flujo de selección de DEGs

Los resultados obtenidos para los DEGs sub-expresados y sobre-expresados se resumen en la Tabla 5 y Tabla 6, respectivamente.

De los 19 genes sub-expresados seleccionados, se evidencia que el gen más sub-expresado en CCE es *C2ORF40*, el cual se expresa aproximadamente 6 veces menos que en tejido cervical sano (Tabla 5). Otros genes sub-expresados en CCE fueron *SRPX* (Log2FC=-4.72), *CHRDL2* (Log2FC=-4.26) y *TUB* (Log2FC=-4.08) (Tabla 5).

Con respecto a los 20 genes sobre-expresados seleccionados, se evidencia que los dos genes más sobre-expresados en CCE son *ASF1B* y *PTTG1*, los cuales se expresan 5 veces más que en tejido cervical sano (Tabla 6). Otros genes con mayor sobre-expresión en CCE son *KIFCI* (Log2FC=4.49) y *LMNB1* (Log2FC=3.9) (Tabla 6).

En 18 de los 19 DEGs sub-expresados en CCE, la media de expresión del grupo con exceso de peso fue menor que la media del grupo sin exceso de peso (Tabla 5).

Por otro lado, en 19 de los 20 DEGs sobre-expresados en CCE, la media de expresión del grupo con exceso de peso fue mayor que la media del grupo sin exceso de peso (Tabla 6).

Tabla 5 Resultados de expresión de los DEGs sub-expresados en CCE

Gen	Proteína codificada	Expresión (UCSC Xena ^a)			Expresión (UCSC Xena ^b)			Expresión (GEPIA)			
		IMC<25 (n=96)	IMC≥25 (n=154)	p	Tumor (n=304)	Sano (n=10)	p	Tumor	Sano	Log2(FC)	FDR
<i>PLA2G2A</i>	Fosfolipasa A2, asociada a membrana	3.52 (2.98)	2.24 (2.38)	0.0002	3.56 (2.96)	7.10 (3.01)	2.37E-04	0.38	4.63	-2.03	0.000293
<i>SPOCK1</i>	Testican-1	7.52 (2.12)	6.63 (1.95)	0.0009	7.73 (2.04)	11.13 (1.28)	2.99E-07	1.58	25.03	-3.34	1.13E-07
<i>CRMP1</i>	Proteína mediadora de respuesta de colapsina 1	6.95 (2.08)	6.18 (1.71)	0.0014	7.44 (1.78)	9.94 (0.75)	1.36E-05	1.85	17.78	-2.72	5.74E-06
<i>NOS3</i>	Sintasa de óxido nítrico 3	6.82 (1.11)	6.39 (1.05)	0.0024	7.39 (1.05)	9.67 (0.90)	6.46E-11	2.19	13.31	-2.16	2.56E-14
<i>EML1</i>	Proteína 1 similar a proteína asociada a microtúbulos de equinodermo	6.95 (1.52)	6.42 (1.25)	0.0031	7.50 (1.35)	10.34 (0.50)	1.23E-10	1.63	12.39	-2.35	5.95E-14
<i>SHANK3</i>	Proteína 3 SH3 con múltiples dominios	8.15 (0.82)	7.85 (0.77)	0.0031	8.72 (0.77)	11.89 (0.96)	3.09E-30	2.05	15.67	-2.45	2.97E-30

	repetidos de anquirina										
<i>C15ORF59</i>	Factor sináptico inhibitorio 1	3.83 (2.01)	3.11 (1.80)	0.0034	4.64 (2.04)	8.47 (1.07)	9.31E-09	0.57	6.94	-2.34	1.32E-07
<i>EDIL3</i>	Proteína 3 que contiene repeticiones tipo EGF y dominio tipo discoidina I	4.94 (2.26)	4.08 (2.28)	0.0041	7.77 (1.80)	9.70 (1.59)	8.93E-04	1.36	10.87	-2.33	4.44E-05
<i>TUB</i>	Homóloga de la proteína Tubby	5.92 (1.87)	5.23 (1.81)	0.0041	6.32 (1.94)	11.54 (1.30)	1.14E-15	0.49	24.13	-4.08	2E-35
<i>SEC31B</i>	Proteína de transporte de proteínas Sec31B	6.30 (1.15)	6.73 (1.16)	0.0042	7.64 (1.11)	10.38 (0.48)	1.16E-13	2.52	13.18	-2.01	2.48E-14
<i>GPM6A</i>	Glicoproteína M6A de membrana neuronal	1.78 (1.57)	1.27 (1.36)	0.0066	1.89 (1.75)	7.50 (1.44)	1.46E-20	0.04	4.25	-2.34	0
<i>C2ORF40</i>	Augurina	2.12 (2.08)	1.49 (1.55)	0.0071	2.34 (1.96)	9.54 (1.75)	1.04E-25	0.19	60.23	-5.68	0
<i>MAP1B</i>	Proteína 1B asociada a microtúbulo	7.96 (1.77)	7.34 (1.79)	0.0072	8.35 (1.84)	11.93 (0.73)	2.60E-09	0.75	12.15	-2.91	6.57E-11

<i>FGF10</i>	Factor de crecimiento de fibroblasto 10	0.37 (0.67)	0.18 (0.45)	0.0073	1.86 (1.66)	9.46 (1.31)	5.59E-36	0.04	13.02	-3.75	0
<i>TTYH2</i>	Homólogo 2 de proteína tweety	5.71 (1.57)	5.19 (1.39)	0.0076	6.33 (1.44)	9.10 (0.74)	4.48E-09	0.73	7.48	-2.29	9.71E-15
<i>CHRD2</i>	Proteína 2 tipo similar a cordina	3.30 (2.32)	2.57 (1.98)	0.0086	4.01 (2.17)	8.76 (4.18)	2.02E-10	0.57	28.97	-4.26	1.16E-20
<i>SRPX</i>	Proteína SRPX que contiene repetición Sushi	5.43 (1.95)	4.80 (1.77)	0.0088	5.96 (1.82)	10.77 (1.30)	3.18E-15	1.26	58.66	-4.72	1.02E-21
<i>PRICKLE2</i>	Proteína 2 similar a prickle	7.33 (1.22)	6.91 (1.26)	0.0090	7.82 (1.26)	11.37 (0.87)	7.62E-17	0.93	14.34	-2.99	5.48E-23
<i>TSPAN18</i>	Tetraspanina 18	6.59 (1.80)	6.05 (1.45)	0.0096	7.07 (1.66)	9.79 (1.15)	4.92E-07	1.10	10.54	-2.46	3.61E-10

Los valores de expresión de UCSC Xena se expresan como media (desviación estándar) del nivel de expresión. El nivel de expresión se representa como el recuento normalizado de valores RSEM transformados a $\log_2(\text{RSEM}+1)$. Los valores de expresión de GEPIA se expresan como la mediana del nivel de expresión. El nivel de expresión se representa como transcritos por millón (TPM) transformados al $\log_2(\text{TPM}+1)$. El Log2FC es definido como la mediana (Tumor) – mediana (Sano).

^a: datos del estudio “TCGA Cervical Cancer (CESC)”

^b: datos del estudio “TCGA TARGET GTEX”

p: valor de p de la prueba t de student

FDR: valor de p de la prueba ANOVA con corrección por pruebas múltiples según el método FDR (recolectado de GEPIA)

Tabla 6 Resultados de expresión de los DEGs sobre-expresados en CCE

Gen	Proteína	Expresión (UCSC Xena ^a)			Expresión (UCSC Xena ^b)			Expresión (GEPIA)			
		IMC<25 (n=96)	IMC≥25 (n=154)	P	Tumor (n=304)	Sano (n=10)	p	Tumor	Sano	Log2(FC)	FDR
<i>CDC25C</i>	Fosfatasa 3 inductora de fase M	7.75 (0.70)	8.06 (0.65)	0.0003	8.67 (0.66)	2.07 (2.11)	0.00E+00	7.26	0.09	2.92	2.78E-32
<i>LMNB1</i>	Lamin B1	11.16 (0.71)	11.44 (0.62)	0.0011	12.06 (0.68)	7.40 (1.34)	0.00E+00	60.18	3.10	3.90	6.59E-30
<i>IL2RG</i>	Subunidad gama común de receptor de citoquina	8.76 (1.49)	9.39 (1.63)	0.0022	9.96 (1.60)	6.86 (0.84)	3.55E-09	25.17	3.97	2.40	1.54E-05
<i>PSMB9</i>	Subunidad del tipo beta 9 del proteosoma	10.27 (1.49)	10.82 (1.35)	0.0032	11.54 (1.42)	8.94 (0.52)	1.75E-08	165.78	32.55	2.31	1.06E-05
<i>GNLV</i>	Granulisina	7.24 (1.96)	8.03 (2.11)	0.0035	8.74 (2.02)	6.09 (0.52)	4.59E-05	20.72	2.77	2.53	0.00316
<i>CIORF210</i>	Proteína TEMP de membrana de endosoma tipo III	7.81 (1.03)	8.18 (0.94)	0.0041	8.68 (1.04)	3.76 (4.10)	6.44E-29	10.45	0.20	3.26	8.62E-10
<i>CD52</i>	Antígeno CAMPATH-1	7.76 (1.42)	8.30 (1.44)	0.0042	8.76 (1.46)	5.58 (0.91)	4.51E-11	53.63	8.61	2.51	6.92E-06

<i>CCL5</i>	Quimiocina 5 con motivo C-C	9.38 (1.59)	10.04 (1.88)	0.0046	10.45 (1.76)	7.32 (1.07)	5.29E-08	42.22	7.23	2.39	0.000173
<i>ETV7</i>	Factor de transcripción ETV7	8.12 (1.45)	8.63 (1.34)	0.0047	9.17 (1.38)	5.63 (1.52)	2.63E-14	14.56	1.39	2.70	2.48E-09
<i>HIST1H2BJ</i>	Histona H2B tipo 1-J	6.59 (1.34)	7.10 (1.41)	0.0054	7.12 (1.41)	4.97 (2.02)	4.15E-06	12.00	1.58	2.33	0.000242
<i>MND1</i>	Homólogo de la proteína 1 de división nuclear meiótica	7.12 (0.82)	7.41 (0.81)	0.0066	8.08 (0.78)	2.97 (1.44)	0.00E+00	11.37	0.45	3.09	8.53E-34
<i>POC1A</i>	Homólogo A de proteína centriolar POC1	8.53 (0.79)	8.80 (0.72)	0.0068	9.41 (0.75)	5.80 (1.00)	4.72E-38	13.00	0.96	2.84	1.94E-24
<i>KIFC1</i>	Proteína KIFC1 similar a kinesina	10.52 (0.59)	10.73 (0.61)	0.0071	11.34 (0.61)	5.68 (1.71)	0.00E+00	38.61	0.76	4.49	0
<i>FAM110C</i>	Proteína FAM110C	6.12 (1.30)	5.59 (1.63)	0.0074	10.24 (1.47)	6.66 (3.72)	1.18E-11	15.22	1.70	2.59	2.77E-05
<i>ASF1B</i>	Chaperona de histona ASF1B	10.52 (0.73)	10.78 (0.73)	0.0076	11.37 (0.72)	4.93 (1.94)	0.00E+00	48.38	0.45	5.09	0
<i>CD3D</i>	Cadena delta CD3 de glucoproteína	6.49 (1.74)	7.10 (1.75)	0.0077	7.67 (1.69)	4.61 (1.14)	3.41E-08	12.49	1.86	2.24	3.66E-05

	de superficie de células T										
<i>PTTG1</i>	Securina	10.57 (0.83)	10.85 (0.79)	0.0084	11.47 (0.79)	6.25 (1.66)	0.00E+00	157.74	3.80	5.05	0
<i>GINS2</i>	Proteína PSF2 del complejo de replicación del ADN GINS	9.57 (0.69)	9.81 (0.72)	0.0089	10.50 (0.69)	5.44 (1.25)	0.00E+00	19.35	0.76	3.53	1.4E-36
<i>ERBB3</i>	Tirosina proteína quinasa receptora erbB-3	11.21 (0.86)	11.52 (0.89)	0.0090	12.28 (0.93)	10.28 (1.57)	2.89E-10	44.62	9.68	2.10	1.57E-11
<i>PCNA</i>	Antígeno nuclear de célula proliferante	12.44 (0.69)	12.66 (0.59)	0.0092	13.25 (0.63)	10.08 (0.82)	0.00E+00	281.66	37.48	2.88	3.19E-28

Los valores de expresión de UCSC Xena se expresan como media (desviación estándar) del nivel de expresión. El nivel de expresión se representa como el recuento normalizado de valores RSEM transformados a $\log_2(\text{RSEM}+1)$. Los valores de expresión de GEPIA se expresan como la mediana del nivel de expresión. El nivel de expresión se representa como transcritos por millón (TPM) transformados al $\log_2(\text{TPM}+1)$. El Log2FC es definido como la mediana (Tumor) – mediana (Sano).

^a: datos del estudio TCGA Cervical Cancer (CESC)

^b: datos del estudio TCGA TARGET GTEx

p: valor de p de la prueba t de student

FDR: valor de p de la prueba ANOVA con corrección por pruebas múltiples según el método FDR (recolectado de GEPIA)

8.3.2 Nivel de expresión por tipo de muestra

Debido a que GEPIA sólo brindó la información cualitativa de expresión génica, se recurrió a UCSC Xena (BD seleccionada: TCGA TARGET GTEx) para conocer la información cuantitativa de expresión génica de los 39 genes de interés. Así, se comparó la expresión de los genes entre tejido tumoral de cérvix y tejido sano de cérvix. Como era de esperarse, los genes sub-expresados, las medias de expresión del tejido de tumor primario fueron menores que las medias de expresión de tejido sano ($p < 0.001$) (Tabla 5). En contraste, para los genes sobre-expresados, las medias de expresión del tejido de tumor primario fueron mayores que las medias de expresión de tejido sano ($p < 0.001$) (Tabla 6). Estos resultados coinciden con los esperado en base a los resultados de GEPIA y confirman el estado de expresión de los DEGs codificantes seleccionados en la población de estudio.

8.4 Análisis de redes de PPI y enriquecimiento funcional

A través del programa STRING se construyeron redes de PPI y se evaluaron enriquecimientos funcionales utilizando los sistemas de clasificación de GO y KEGG.

8.4.1 Genes sub-expresados

En el caso de los genes sub-expresados, se analizaron las proteínas codificadas por 18 de los 19 genes seleccionados (PLA2G2A, SPOCK1, CRMP1, NOS3, EML1, C15ORF59, EDIL3, TUB, SEC31B, GPM6A, C2ORF40, MAP1B, FGF10, TTYH2, CHRDL2, SRPX, PRICKLE2, TSPAN18). La proteína SHANK3 no se incluyó en el análisis porque no fue identificada en el programa STRING. No se evidenció una red estadísticamente significativa conformada por las 18 proteínas ($p\text{-valor}=1$) y tampoco hubo algún enriquecimiento funcional.

8.4.2 Genes sobre-expresados

En cuanto a los genes sobre-expresados, se analizaron las proteínas codificadas por los 20 genes seleccionados: CDC25C, LMNB1, IL2RG, PSMB9, GNLY, C1ORF210, CD52, CCL5, ETV7, HIST1H2BJ, MND1, POC1A, KIFC1, FAM110C, ASF1B, CD3D, PTTG1, GINS2, ERBB3, PCNA.

La red de enriquecimiento de PPI fue significativa ($p=5.86e-10$). En la Figura 5 se ilustra la red PPI conformada por dos sub-redes (una de 10 proteínas y otra de 3 proteínas) y 23 bordes en total. En la figura de la red se excluyeron aquellas proteínas que no presentaron interacciones, las cuales fueron CCL5, C1ORF210, FAM110C, GNLY, ERBB3, POC1A, ETV7.

La red de 10 proteínas estuvo conformada por la proteína CDC25C como nodo central y las proteínas PSMB9, PTTG1, KIFC1, GINS2, ASF1B, PCNA, MND1, HIST1H2BJ y LMNB1. En ella, se evidenciaron 19 bordes, de los cuales 14 fueron de asociaciones por co-expresión (líneas negras), 2 asociaciones por evidencia experimental (líneas moradas) y 3 asociaciones por evidencia en BD curadas (líneas celestes). Por otro lado, la red de tres proteínas estuvo conformada por las proteínas CD52, IL2RG y CD3D. Se evidenciaron cuatro bordes conformados por tres asociaciones por co-expresión (líneas negras) y una asociación por evidencia en BD curadas (líneas celestes).

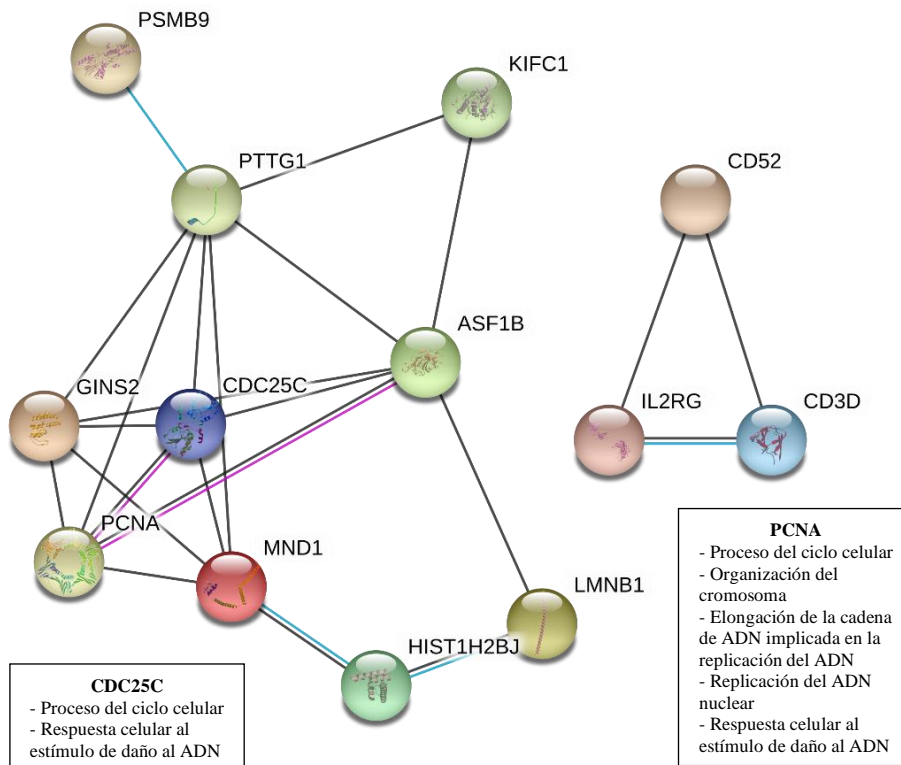


Figura 5 Red de PPI de los genes sobre-expresados seleccionados

La red PPI resume la red de asociaciones predichas para el grupo de genes sobre-expresados seleccionados. Los nodos de la red son proteínas. Los bordes representan las asociaciones predichas y están coloreados en base al tipo de evidencia utilizada para la predicción. Línea morada: evidencia experimental, línea celeste: evidencia de BDs curadas línea negra: evidencia de co-expresión. Se excluyeron de la red las proteínas que no presentaron interacciones (en total 7). Se incluye en recuadros, los principales BP en los que las proteínas CDC25C y PCNA participan.

Asimismo, se evidenció enriquecimiento funcional para 12 procesos biológicos (BP), y 8 rutas KEGG. Los detalles del enriquecimiento se resumen en las Tabla 7 y Tabla 8. Entre los términos BP enriquecidos se encontraron términos relacionados al ciclo celular (“Proceso del ciclo celular”, “Proceso del ciclo celular mitótico”, “Organización del cromosoma”), a la replicación del ADN (“Elongación de la cadena de ADN implicada en la replicación del ADN”, “Replicación del ADN nuclear”), a procesos que involucran otros organismos (“Proceso de múltiples organismos”, “Interacción entre especies entre organismos”) y a procesos

relacionados a la reproducción (“Espermatogénesis”, “Proceso reproductivo”, “Proceso reproductivo de organismos multicelulares”). No hubo enriquecimiento en términos GO para MF ni CC.

Por último, los términos KEGG enriquecidos incluyeron rutas relacionadas al ciclo celular (“ciclo celular”, “meiosis de oocitos”), infecciones (“infección HTLV-1”, “enfermedad de Chagas”) y al sistema inmune (“inmunodeficiencia primaria, diferenciación de células Th1 y Th2, diferenciación de células Th17).

Las funciones de los 39 genes de estudio han sido resumidas en la Tabla 9 para lo cual se consultaron las BDs de Uniprot y el NCBI.

Tabla 7 Enriquecimiento funcional de términos de GO en los genes sobre-expresados

Tipo de término GO	ID del término GO	Descripción del término GO	Recuento de genes	Recuento de genes de fondo	FDR	Proteínas de la red
BP	GO:0022402	Proceso del ciclo celular	7	890	0.0175	CDC25C, GINS2, KIFC1, MND1, PCNA, POC1A, PTTG1
BP	GO:0051704	Proceso de múltiples organismos	10	2222	0.0175	ASF1B, CCL5, CDC25C, GNLY, HIST1H2BJ, IL2RG, KIFC1, POC1A, PSMB9, PTTG1
BP	GO:0044419	Interacción entre especies entre organismos	6	724	0.0201	CCL5, CDC25C, GNLY, HIST1H2BJ, IL2RG, PSMB9
BP	GO:0006271	Elongación de la cadena de ADN implicada en la replicación del ADN	2	18	0.0205	GINS2, PCNA
BP	GO:0007283	Espermatogénesis	5	490	0.0205	ASF1B, CDC25C, KIFC1, POC1A, PTTG1
BP	GO:0048609	Proceso reproductivo de organismos multicelulares	6	766	0.0205	ASF1B, CDC25C, KIFC1, PCNA, POC1A, PTTG1
BP	GO:1903047	Proceso del ciclo celular mitótico	5	564	0.0205	CDC25C, GINS2, KIFC1, PCNA, POC1A
BP	GO:0022414	Proceso reproductivo	7	1350	0.0221	ASF1B, CDC25C, KIFC1, MND1, PCNA, POC1A, PTTG1
BP	GO:0051276	Organización del cromosoma	6	999	0.0238	ASF1B, GINS2, HIST1H2BJ, KIFC1, PCNA, PTTG1
BP	GO:0033260	Replicación del ADN nuclear	2	35	0.0352	GINS2, PCNA

BP	GO:0006974	Respuesta celular al estímulo de daño al ADN	5	749	0.0389	CDC25C, GINS2, MND1, PCNA, PTTG1
BP	GO:0035821	Modificación de la morfología o fisiología de otro organismo	3	182	0.0389	CCL5, GNLY, HIST1H2BJ

GO: Ontología génica, BP: Proceso biológico, FDR: valor de p para el enriquecimiento funcional corregido por el método FDR (obtenido de STRING)

Tabla 8 Enriquecimiento funcional de términos KEGG en los genes sobre-expresados

ID del término KEGG	Descripción de la ruta KEGG	Recuento de genes	Recuento de genes de fondo	FDR	Proteínas de la red
hsa05166	Infección HTLV-1	4	250	0.0051	CD3D, IL2RG, PCNA, PTTG1
hsa04110	Ciclo celular	3	123	0.0062	CDC25C, PCNA, PTTG1
hsa05340	Inmunodeficiencia primaria	2	37	0.0108	CD3D, IL2RG
hsa04658	Diferenciación de células Th1 y Th2	2	88	0.0423	CD3D, IL2RG
hsa04114	Meiosis de oocitos	2	116	0.044	CDC25C, PTTG1
hsa04659	Diferenciación de células Th17	2	102	0.044	CD3D, IL2RG
hsa05142	Enfermedad de Chagas (trypanosomiasis americana)	2	101	0.044	CCL5, CD3D
hsa05162	Sarampión	2	133	0.0465	CD3D, IL2RG

FDR: valor de p para el enriquecimiento funcional corregido por el método FDR (obtenido de STRING)

Tabla 9 Descripción y función de los 39 DEGs seleccionados

Gen	Código del gen	Proteína codificada	Código Uniprot	Descripción y función de la proteína	Tamaño
Genes sub-expresados					
<i>PLA2G2A</i>	ENSG00000188257	Fosfolipasa A2, asociada a membrana	P14555	Es un miembro de la familia de fosfolipasas A2 (PLA2). Podría tener un rol en la regulación del metabolismo de los fosfolípidos en las biomembranas, incluyendo la biosíntesis de eicosanoides. Actúa como ligando para integrinas.	144 aa
<i>SPOCK1</i>	ENSG00000152377	Testican-1	Q08629	Podría tener un rol en las interacciones célula-célula y célula-matriz. Puede contribuir en varios mecanismos neuronales en el sistema nervioso central.	439 aa
<i>CRMP1</i>	ENSG00000072832	Proteína mediadora de respuesta de colapsina 1	Q14194	Es necesaria para la señalización por semaforinas de clase 3 y la posterior remodelación del citoesqueleto. Tiene un rol en la orientación del axón. Participa en el crecimiento invasivo y la migración celular. También tiene un rol en la citoquinesis.	686 aa

<i>NOS3</i>	ENSG00000164867	Sintasa de óxido nítrico 3	P29474	Proteína que produce óxido nítrico (NO), el cual está implicado en la relajación del músculo liso vascular a través de una vía de transducción de señales mediada por cGMP. El NO interviene en la angiogénesis inducida por el factor de crecimiento endotelial vascular (VEGF) en los vasos coronarios y promueve la coagulación de la sangre a través de la activación de las plaquetas.	1203 aa
<i>EML1</i>	ENSG00000066629	Proteína 1 similar a proteína asociada a microtúbulos de equinodermo	O00423	Modula el ensamblaje y la organización del citoesqueleto de microtúbulos. Probablemente tiene un rol en la regulación de la orientación del huso mitótico y la orientación del plano de división celular. Es necesaria para la proliferación normal de células progenitoras neuronales en el cerebro en desarrollo y para el desarrollo normal del cerebro.	815 aa
<i>SHANK3</i>	ENSG00000251322	Proteína 3 SH3 con múltiples dominios repetidos de anquirina	Q9BYB0	Es un miembro de la familia de proteínas Shank, las cuales son proteínas de andamio multidominio de la densidad postsináptica que conectan receptores de neurotransmisores, canales iónicos y otras proteínas de membrana al citoesqueleto de actina y a rutas de señalización acopladas a proteínas G. Estas proteínas tienen un rol en la formación de la sinapsis y en la maduración de la columna dendrítica.	1731 aa

<i>C15ORF59</i>	ENSG00000205363	Factor sináptico inhibitorio 1	Q2T9L4	Es un componente de la maquinaria de proteínas en las sinapsis inhibitorias, en donde probablemente actúa como un andamio. Las sinapsis inhibitorias amortiguan la actividad neuronal a través de la hiperpolarización postsináptica y son fundamentales para el funcionamiento del sistema nervioso central.	293 aa
<i>EDIL3</i>	ENSG00000164176	Proteína 3 que contiene repeticiones tipo EGF y dominio tipo discoidina I	O43854	Es un ligando de integrina que inhibe la formación de estructuras vasculares. Podría estar involucrada en la regulación de la morfogénesis vascular de la remodelación en el desarrollo embrionario.	480 aa
<i>TUB</i>	ENSG00000166402	Homóloga de la proteína Tubby	P50607	Es un miembro de la familia Tubby de factores de transcripción bipartitos. Tiene un rol en la transducción de señales de receptores heterotriméricos acoplados a proteína G. Puede contribuir a la regulación de la transcripción en el núcleo y podría estar involucrada en la regulación hipotalámica del peso corporal. Puede contribuir a la estimulación de la fagocitosis de las células y los macrófagos del epitelio pigmentario retiniano apoptótico (RPE).	506 aa
<i>SEC31B</i>	ENSG00000075826	Proteína de transporte de proteínas Sec31B	Q9NQW1	Es un componente del complejo de proteínas de cubierta II (COPII). Puede estar involucrada en la formación de vesículas y en la exportación de carga desde el retículo endoplásmico.	1179 aa

<i>GPM6A</i>	ENSG00000150625	Glicoproteína M6A de membrana neuronal	P51674	Tiene un rol en la diferenciación neuronal, incluyendo la diferenciación y migración de células madres neuronales. También contribuye en la plasticidad neuronal y participa en el crecimiento de neuritas y filopodias, la motilidad de folipodios y probablemente en la formación de sinapsis.	278 aa
<i>C2ORF40</i>	ENSG00000119147	Augurina	Q9H1Z8	Probablemente es una hormona que atenúa la proliferación celular e induce la senescencia de oligodendrocitos y células precursoras neurales en el sistema nervioso central.	148 aa
<i>MAP1B</i>	ENSG00000131711	Proteína 1B asociada a microtúbulo	P46821	Es un miembro de la familia de proteínas asociadas a microtúbulos. Se piensa que esta familia tiene un rol en el ensamblaje de microtúbulos, función que es esencial en la neurogénesis.	2468 aa
<i>FGF10</i>	ENSG00000070193	Factor de crecimiento de fibroblasto 10	O15520	Tiene un rol importante en la regulación del desarrollo embrionario, la proliferación celular y la diferenciación celular. Es requerido para la morfogénesis de ramificación normal y podría cumplir un rol en la curación de heridas.	208 aa
<i>TTYH2</i>	ENSG00000141540	Homólogo 2 de proteína tweety	Q9BSA4	Miembro de la familia de proteínas Tweety, las cuales funcionan como canales de aniones de cloruro. TTYH2 funciona como un canal de cloruro activado por calcio (2+) y podría estar involucrada en la proliferación y agregación celular.	534 aa

<i>CHRDL2</i>	ENSG00000054938	Proteína 2 tipo similar a cordina	Q6WN34	Es un miembro de la familia de proteínas cordina. Tiene un efecto regulador negativo en la formación/regeneración de cartílago de células mesenquimales inmaduras. Está implicada en la angiogénesis tumoral. Podría tener un rol durante la diferenciación y maduración de mioblastos y osteoblastos.	451 aa
<i>SRPX</i>	ENSG00000101955	Proteína SRPX que contiene repetición Sushi	P78539	Proteína que podría tener un rol en la fagocitosis durante el desprendimiento del disco, la adhesión celular a células distintas del epitelio pigmentario o la transducción de señales.	464 aa
<i>PRICKLE2</i>	ENSG00000163637	Proteína 2 similar a prickle	Q7Z3G6	Es una proteína homóloga a la proteína Prickle de <i>Drosophila</i> . Se desconoce su función exacta, sin embargo, a partir de estudios en ratones se hipotetiza que podría estar involucrada en la prevención de convulsiones.	844 aa
<i>TSPAN18</i>	ENSG00000157570	Tetraspanina 18	Q96SJ8	No se tiene información de su función	248 aa
Genes sobre-expresados					

<i>CDC25C</i>	ENSG00000158402	Fosfatasa 3 inductora de fase M	P30307	Tiene un rol clave en la regulación de la división celular. CDC25C dirige la defosforilación de CD52 unido a la ciclina B y desencadena la entrada a la mitosis. Asimismo, suprime el arresto de crecimiento inducido por p53.	473 aa
<i>LMNB1</i>	ENSG00000113368	Lamin B1	P20700	Es un componente de la lámina nuclear, la cual es una capa fibrosa de la membrana nuclear interna.	586 aa
<i>IL2RG</i>	ENSG00000147168	Subunidad gama común de receptor de citoquina	P31785	Es una subunidad común de receptores de interleuquinas.	369 aa
<i>PSMB9</i>	ENSG00000240065	Subunidad del tipo beta 9 del proteosoma	P28065	Es un componente del proteosoma. El proteosoma es un complejo proteinasa multicatalítico que se caracteriza por su capacidad de cortar péptidos con grupos Arginina (Arg), Fenilalanina (Phe), Tirosina (Tyr), Leucina (Leu) y Ácido glutámico (Glu) adyacentes al grupo saliente a un pH neutro o ligeramente básico. Esta subunidad está involucrada en el procesamiento de antígenos para generar péptidos de unión a la clase I.	219 aa

<i>GNLY</i>	ENSG00000115523	Granulisina	P22749	Es una proteína antimicrobiana que mata patógenos intracelulares. Es activo contra una amplia gama de microorganismos, incluyendo bacterias, hongos y parásitos.	145 aa
<i>CIORF210</i>	ENSG00000253313	Proteína TEMP de membrana de endosoma tipo III	Q8IVY1	Podría tener un rol en el tráfico de membrana entre los endosomas y la membrana plasmática.	113 aa
<i>CD52</i>	ENSG00000169442	Antígeno CAMPATH-1	P31358	Podría tener un rol en el transporte y la orientación de los carbohidratos.	61 aa
<i>CCL5</i>	ENSG00000271503	Quimiocina 5 con motivo C-C	P13501	Es un tipo de quimiocina, las cuales son una superfamilia de proteínas secretadas involucradas en procesos inmunoregulatorios e inflamatorios. La quimiocina CCL5 funciona como un quimioatrayente de monocitos de la sangre, células T cooperadoras de memoria y eosinófilos. Ocasiona la liberación de histamina de los basófilos y activa a los eosinófilos y es uno de los principales factores supresores de VIH producidos por las células T CD8+.	91 aa

<i>ETV7</i>	ENSG00000010030	Factor de transcripción ETV7	Q9Y603	Es un represor transcripcional que se une a la secuencia de ADN 5'-CCGGAAGT-3. Pertenece a la familia de factores de transcripción ETS, un grupo grande de reguladores transcripcionales evolutivamente conservados que tienen un rol importante en procesos celulares durante el desarrollo y diferenciación. También están relacionados a la oncogénesis.	341 aa
<i>HIST1H2BJ</i>	ENSG00000124635	Histona H2B tipo 1-J	P06899	Es un componente central de los nucleosomas, las cuales son estructuras que envuelven y compactan el ADN en cromatina. Posee actividad antibacteriana.	126 aa
<i>MND1</i>	ENSG00000121211	Homólogo de la proteína 1 de división nuclear meiótica	Q9BWT6	Es una proteína requerida para el adecuado emparejamiento de cromosomas homólogos y una eficiente recombinación cruzada e intragénica durante la meiosis. Se asocia con HOP2 para formar un complejo heterodimérico estable que se une al ADN y estimula la actividad recombinasa de RAD51 y DMC1.	205 aa
<i>POCIA</i>	ENSG00000164087	Homólogo A de proteína centriolar POC1	Q8NBT0	Forma parte de los centrosomas. Tiene un rol importante en el ensamblaje y estabilidad del centriolo, al igual que en la formación de cilios.	407 aa

<i>KIFC1</i>	ENSG00000237649	Proteína KIFC1 similar a kinesina	Q9BW19	Es una proteína requerida en la formación del huso bipolar. Puede contribuir al movimiento de vesículas endocíticas tempranas y regula la formación y estructura del cilio.	673 aa
<i>FAM110C</i>	ENSG00000184731	Proteína FAM110C	Q1W6H9	Puede tener un rol en la organización de microtúbulos, la propagación celular y la migración de células epiteliales. Su función puede involucrar la vía de señalización AKT1.	321 aa
<i>ASF1B</i>	ENSG00000105011	Chaperona de histona ASF1B	Q9NVP2	Es una chaperona de histona que facilita el depósito, intercambio y eliminación de histonas durante el ensamblaje y desmontaje del nucleosoma. Actúa en conjunto con el factor de ensamblaje de cromatina 1 (CAF-1). Es requerida para la espermatogénesis.	202 aa
<i>CD3D</i>	ENSG00000167286	Cadena delta CD3 de glucoproteína de superficie de células T	P04234	Es una proteína que forma parte del complejo receptor de célula T/CD3 (complejo TCR-CD3) presente en la superficie celular de linfocitos T. Está involucrada en el desarrollo de células T y en la transducción de señales. Esta proteína es la subunidad delta del complejo CD3.	171 aa

<i>PTTG1</i>	ENSG00000164611	Securina	O95997	Es una proteína regulatoria que tiene un rol central en la estabilidad de los cromosomas, en la ruta p53/TP53 y en la reparación del ADN. Se hipotetiza que actúa bloqueando la acción de proteínas claves. Durante la mitosis, bloquea la función Separasa/ESPL1. Al inicio de la anafase está ubiquitinada, lo que conduce a su destrucción y a la liberación de ESPL1. Regula negativamente la actividad transcripcional y la actividad de apoptosis relacionada a TP53.	202 aa
<i>GINS2</i>	ENSG00000131153	Proteína PSF2 del complejo de replicación del ADN GINS	Q9Y248	El complejo GINS tiene un rol esencial en la iniciación de la replicación del ADN y en la progresión de hebras de replicación del ADN.	185 aa
<i>ERBB3</i>	ENSG00000065361	Tirosina proteína quinasa receptora erbB-3	P21860	Es una tirosina proteína quinasa que tiene un rol esencial como receptor de la superficie celular para neuregulinas. La unión del ligando incrementa la fosforilación en residuos de tirosina y promueve su asociación con la subunidad p85 de la quinasa fosfatidil inositol 3. Está involucrada en la diferenciación de células mieloides.	1342 aa
<i>PCNA</i>	ENSG00000132646	Antígeno nuclear de célula proliferante	P12004	Es un cofactor de la ADN polimerasa delta. Participa en el control de la replicación del ADN eucariota al aumentar la capacidad de procesamiento de la polimerasa durante el alargamiento de la cadena líder.	261 aa

8.5 Análisis de metilación del ADN

8.5.1 Selección de transcritos

Para el análisis de metilación de los 39 genes seleccionados se escogieron los transcritos que codifican a la isoforma canónica o predominante. Los genes *PSMB9* y *KIFC1* no se encontraron en la base de datos de MethHC, por lo que no pudieron ser analizados. En total se seleccionaron 40 transcritos que tuvieron un rango de entre 467 y 11790 pb, cuyos detalles se resumen en la Tabla 10.

Tabla 10 Detalles de los transcritos de estudio

Gen	Transcrito	Tamaño del transcrito
Genes sub-expresados		
<i>PLA2G2A</i>	NM_000300	991 pb
<i>SPOCK1</i>	NM_004598	4824 pb
<i>CRMP1</i>	NM_001014809	3174 pb
	NM_001313	2948 pb
<i>NOS3</i>	NM_000603	4366 pb
<i>EML1</i>	NM_004434	4460 pb
<i>SHANK3</i>	NM_033517	7345 pb
<i>C15ORF59</i>	NM_001039614	5730 pb
<i>EDIL3</i>	NM_005711	4814 pb
<i>TUB</i>	NM_177972	6175 pb
<i>SEC31B</i>	NM_015490	4634 pb

<i>GPM6A</i>	NM_005277	3215 pb
	NM_201591	3697 pb
<i>C2ORF40</i>	NM_032411	753 pb
<i>MAP1B</i>	NM_005909	11790 pb
<i>FGF10</i>	NM_004465	6113 pb
<i>TTYH2</i>	NM_032646	3433 pb
<i>CHRD2</i>	NM_015424	1963 pb
<i>SRPX</i>	NM_006307	1846 pb
<i>PRICKLE2</i>	NM_198859	9947 pb
<i>TSPAN18</i>	NM_130783	4557 pb
Genes sobre-expresados		
<i>CDC25C</i>	NM_001790	2093 pb
<i>LMNB1</i>	NM_005573	2890 pb
<i>IL2RG</i>	NM_000206	1560 pb
<i>GNLY</i>	NM_006433	777 pb
<i>CIORF210</i>	NM_182517	1486 pb
<i>CD52</i>	NM_001803	467 pb
<i>CCL5</i>	NM_002985	1217 pb
<i>ETV7</i>	NM_016135	1605 pb
<i>HIST1H2BJ</i>	NM_021058	481 pb
<i>MND1</i>	NM_032117	929 pb

<i>POCIA</i>	NM_015426	1936 pb
<i>FAM110C</i>	NM_001077710	3880 pb
<i>ASF1B</i>	NM_018154	1689 pb
<i>CD3D</i>	NM_000732	771 pb
<i>PTTG1</i>	NM_004219	715 pb
<i>GINS2</i>	NM_016095	2628 pb
<i>ERBB3</i>	NM_001982	5765 pb
<i>PCNA</i>	NM_002592	1355 pb
	NM_182649	1276 pb

pb: pares de bases. Se detallan los códigos de los transcritos que se emplearon en los análisis para cada uno de los DEGs seleccionados, al igual que el tipo de variante y el tamaño (en pb).

8.5.2 Nivel de metilación en función del estado de exceso de peso

Con el recurso MethHC, se obtuvieron los datos de metilación de 181 pacientes, de las cuales 67 pertenecieron al grupo sin exceso de peso y 114 pertenecieron al grupo con exceso de peso. En la Tabla 11 se detalla el nivel de metilación de los 40 transcritos de estudio y los resultados de las pruebas de suma de rangos Wilcoxon. Se observó una diferencia significativa ($p < 0.05$) en el nivel de metilación entre los grupos de estudio para siete transcritos pertenecientes a cinco genes sub-expresados *SPOCK1* (NM_004598), *EDIL3* (NM_005711), *SEC31B* (NM_015490), *GPM6A* (NM_005277), *C2ORF40* (NM_032411) y a dos genes sobre-expresados *ETV7* (NM_016135) y *MND1* (NM_032117) (Tabla 11).

En los cinco genes sub-expresados mencionados, se evidencia que los niveles de metilación son mayores en el grupo con exceso de peso. Por otro lado, para el gen sobre-expresado *ETV7* (NM_016135) se observa que los niveles de metilación son mayores en el grupo sin exceso de peso (0.07 [0.06; 0.08]) en comparación al grupo con exceso de peso (0.06 [0.06;0.07]). En el caso del segundo gen sobre-expresado seleccionado, *MND1* (NM_032117), se evidenció una diferencia entre los grupos a pesar de que se registraron valores iguales de mediana y RI (0.03 [0.03;0.04]). Al graficar la distribución de los datos por grupo en histogramas (Anexo 5), se observó que, en el grupo sin exceso de peso, existen valores aberrantes cercanos a 0.1, mientras que, en el grupo con exceso de peso, el 100% de sus valores están por debajo de 0.06. Probablemente la presencia de estos valores aberrantes influya en que ambos grupos cuenten con igual mediana y RI, aunque la distribución de sus respectivos datos sea diferente.

Tabla 11 Nivel de metilación en función del estado de exceso de peso

Gen (transcrito)	Nivel de metilación			p
	Total de participantes (n=181)	IMC<25 (n=67)	IMC≥25 (n=114)	
Genes sub-expresados				
<i>PLA2G2A</i> (NM_000300)	0.64 [0.48;0.76]	0.63 [0.51;0.78]	0.65 [0.47;0.75]	0.927
<i>SPOCK1</i> (NM_004598)	0.20 [0.07;0.52]	0.12 [0.05;0.50]	0.23 [0.09;0.52]	0.045*
<i>CRMP1</i> (NM_001014809)	0.52 [0.37;0.64]	0.49 [0.38;0.61]	0.53 [0.36;0.65]	0.348
<i>CRMP1</i> (NM_001313)	0.06 [0.04;0.16]	0.06 [0.04;0.10]	0.06 [0.04;0.18]	0.36
<i>NOS3</i> (NM_000603)	0.84 [0.75;0.87]	0.84 [0.75;0.87]	0.84 [0.76;0.87]	0.808
<i>EML1</i> (NM_004434)	0.33 [0.07;0.52]	0.26 [0.07;0.48]	0.37 [0.08;0.54]	0.152
<i>SHANK3</i> (NM_033517)	0.15 [0.08;0.31]	0.14 [0.08;0.28]	0.16 [0.09;0.37]	0.317
<i>C15ORF59</i> (NM_001039614)	0.07 [0.05;0.20]	0.07 [0.05;0.16]	0.07 [0.05;0.22]	0.603
<i>EDIL3</i> (NM_005711)	0.16 [0.06;0.53]	0.08 [0.05;0.44]	0.22 [0.06;0.55]	0.007*
<i>TUB</i> (NM_177972)	0.09 [0.05;0.17]	0.08 [0.06;0.11]	0.10 [0.05;0.21]	0.241

<i>SEC31B</i> (NM_015490)	0.66 [0.55;0.73]	0.63 [0.50;0.71]	0.67 [0.58;0.73]	0.016*
<i>GPM6A</i> (NM_005277)	0.47 [0.35;0.53]	0.40 [0.32;0.49]	0.48 [0.39;0.54]	0.026*
<i>GPM6A</i> (NM_201591)	0.38 [0.29;0.51]	0.39 [0.30;0.49]	0.38 [0.29;0.54]	0.891
<i>C2ORF40</i> (NM_032411)	0.53 [0.36;0.62]	0.49 [0.33;0.58]	0.55 [0.40;0.64]	0.043*
<i>MAP1B</i> (NM_005909)	0.05 [0.04;0.09]	0.05 [0.03;0.07]	0.05 [0.04;0.11]	0.147
<i>FGF10</i> (NM_004465)	0.53 [0.43;0.64]	0.50 [0.41;0.61]	0.55 [0.45;0.65]	0.178
<i>TTYH2</i> (NM_032646)	0.03 [0.03;0.03]	0.03 [0.03;0.04]	0.03 [0.03;0.03]	0.463
<i>CHRD2</i> (NM_015424)	0.48 [0.30;0.60]	0.48 [0.29;0.60]	0.48 [0.32;0.62]	0.684
<i>SRPX</i> (NM_006307)	0.47 [0.41;0.55]	0.46 [0.39;0.54]	0.49 [0.43;0.55]	0.221
<i>PRICKLE2</i> (NM_198859)	0.49 [0.34;0.64]	0.50 [0.38;0.64]	0.48 [0.31;0.63]	0.334
<i>TSPAN18</i> (NM_130783)	0.66 [0.50;0.82]	0.69 [0.55;0.84]	0.64 [0.46;0.81]	0.197
Genes sobre-expresados				
<i>CDC25C</i> (NM_001790)	0.05 [0.04;0.05]	0.05 [0.04;0.05]	0.05 [0.04;0.05]	0.563

<i>LMNB1</i> (NM_005573)	0.03 [0.03;0.04]	0.03 [0.03;0.04]	0.03 [0.03;0.04]	0.761
<i>IL2RG</i> (NM_000206)	0.40 [0.31;0.48]	0.41 [0.30;0.47]	0.39 [0.32;0.48]	0.824
<i>GPLY</i> (NM_006433)	0.75 [0.64;0.81]	0.77 [0.66;0.82]	0.75 [0.64;0.81]	0.272
<i>CIORF210</i> (NM_182517)	0.33 [0.27;0.42]	0.34 [0.25;0.42]	0.32 [0.28;0.42]	0.646
<i>CD52</i> (NM_001803)	0.68 [0.59;0.74]	0.68 [0.59;0.74]	0.68 [0.59;0.73]	0.926
<i>CCL5</i> (NM_002985)	0.55 [0.39;0.69]	0.57 [0.44;0.70]	0.54 [0.39;0.68]	0.645
<i>ETV7</i> (NM_016135)	0.07 [0.06;0.08]	0.07 [0.06;0.08]	0.06 [0.06;0.07]	0.007*
<i>HIST1H2BJ</i> (NM_021058)	0.03 [0.03;0.04]	0.03 [0.03;0.04]	0.03 [0.03;0.04]	0.379
<i>MND1</i> (NM_032117)	0.03 [0.03;0.04]	0.03 [0.03;0.04]	0.03 [0.03;0.04]	0.004*
<i>POCIA</i> (NM_015426)	0.04 [0.03;0.04]	0.04 [0.03;0.04]	0.04 [0.03;0.04]	0.793
<i>FAM110C</i> (NM_001077710)	0.08 [0.06;0.11]	0.08 [0.05;0.11]	0.08 [0.06;0.12]	0.667
<i>ASF1B</i> (NM_018154)	0.02 [0.02;0.03]	0.02 [0.02;0.03]	0.02 [0.02;0.03]	0.595

<i>CD3D</i> (NM_000732)	0.73 [0.64;0.80]	0.73 [0.66;0.79]	0.73 [0.63;0.80]	0.356
<i>PTTG1</i> (NM_004219)	0.06 [0.06;0.07]	0.07 [0.06;0.07]	0.06 [0.06;0.07]	0.359
<i>GINS2</i> (NM_016095)	0.04 [0.03;0.04]	0.04 [0.04;0.04]	0.04 [0.03;0.04]	0.669
<i>ERBB3</i> (NM_001982)	0.04 [0.03;0.04]	0.04 [0.03;0.04]	0.04 [0.03;0.04]	0.641
<i>PCNA</i> (NM_002592)	0.02 [0.02;0.02]	0.02 [0.02;0.02]	0.02 [0.02;0.02]	0.236
<i>PCNA</i> (NM_182649)	0.03 [0.02;0.03]	0.03 [0.02;0.03]	0.03 [0.02;0.03]	0.697

Los valores de metilación de los transcritos se expresan como mediana [rango intercuartílico]. El nivel de metilación corresponde a los coeficientes β , cuyos valores están entre 0 y 1. Los datos fueron obtenidos de MethHC.p: valor de p obtenido en la prueba de suma de rangos Wilcoxon. Se marcó con un asterisco (“*”) y se resaltó en negrita aquellos transcritos que tuvieron un valor de $p < 0.05$ en la prueba estadística.

8.5.3 Estado de metilación en función del estado de exceso de peso

El estado de metilación de los 40 transcritos de estudio, por grupo de estudio y a nivel total, se resumen en la Tabla 12.

De los 21 transcritos pertenecientes a genes sub-expresados, se observa que, en 13 de ellos, la mayoría de pacientes presentan hipermetilación (entre 53% y 99.4%). Por otro lado, en cinco transcritos la categoría predominante es la hipometilación (entre 56.9% y 100%). En el caso de los transcritos *SPOCK1* (NM_004598), *SHANK3* (NM_033517) y *EDIL3* (NM_005711), ninguna categoría de estado de metilación contiene más del 50% de los datos. En *SPOCK1* (NM_004598), el 44.2% presenta hipermetilación, el 34.3% hipometilación y el 21.5% metilación normal. En *SHANK3* (NM_033517), el 43.1% presenta metilación normal, el 30.9% hipometilación y el 26% hipermetilación. En *EDIL3* (NM_005711), 43.1% presenta hipometilación, el 42% hipermetilación y un 14.9% metilación normal.

En el caso de los transcritos que pertenecen a los genes sobre-expresados, se observa que, en 13 de los 19 transcritos, la mayoría presenta hipometilación (entre 68% y 100%). Por otro lado, en los seis transcritos restantes, predominó la hipermetilación (entre 60.8% y 98.3%).

Tabla 12 Estado de metilación de los transcritos de estudio

Gen (transcrito)	Estado de metilación	Total de participantes (n=181)	IMC<25 (n=67)	IMC≥25 (n=114)
Genes sub-expresados				
<i>PLA2G2A</i> (NM_000300)	Hipometilado	1 (0.55%)	0 (0.00%)	1 (0.88%)
	Normal	11 (6.08%)	4 (5.97%)	7 (6.14%)
	Hipermetilado	169 (93.4%)	63 (94.0%)	106 (93.0%)
<i>SPOCK1</i> (NM_004598)	Hipometilado	62 (34.3%)	31 (46.3%)	31 (27.2%)
	Normal	39 (21.5%)	9 (13.4%)	30 (26.3%)
	Hipermetilado	80 (44.2%)	27 (40.3%)	53 (46.5%)
<i>CRMP1</i> (NM_001014809)	Hipometilado	10 (5.52%)	5 (7.46%)	5 (4.39%)
	Normal	28 (15.5%)	8 (11.9%)	20 (17.5%)
	Hipermetilado	143 (79.0%)	54 (80.6%)	89 (78.1%)
<i>CRMP1</i> (NM_001313)	Hipometilado	120 (66.3%)	48 (71.6%)	72 (63.2%)
	Normal	30 (16.6%)	11 (16.4%)	19 (16.7%)
	Hipermetilado	31 (17.1%)	8 (11.9%)	23 (20.2%)
<i>NOS3</i> (NM_000603)	Normal	1 (0.55%)	0 (0.00%)	1 (0.88%)
	Hipermetilado	180 (99.4%)	67 (100%)	113 (99.1%)
<i>EML1</i> (NM_004434)	Hipometilado	58 (32.0%)	26 (38.8%)	32 (28.1%)
	Normal	27 (14.9%)	10 (14.9%)	17 (14.9%)
	Hipermetilado	96 (53.0%)	31 (46.3%)	65 (57.0%)
<i>SHANK3</i> (NM_033517)	Hipometilado	56 (30.9%)	22 (32.8%)	34 (29.8%)
	Normal	78 (43.1%)	30 (44.8%)	48 (42.1%)
	Hipermetilado	47 (26.0%)	15 (22.4%)	32 (28.1%)

<i>C15ORF59</i> (NM_001039614)	Hipometilado	117 (64.6%)	44 (65.7%)	73 (64.0%)
	Normal	28 (15.5%)	12 (17.9%)	16 (14.0%)
	Hipermetilado	36 (19.9%)	11 (16.4%)	25 (21.9%)
<i>EDIL3</i> (NM_005711)	Hipometilado	78 (43.1%)	38 (56.7%)	40 (35.1%)
	Normal	27 (14.9%)	6 (8.96%)	21 (18.4%)
	Hipermetilado	76 (42.0%)	23 (34.3%)	53 (46.5%)
<i>TUB</i> (NM_177972)	Hipometilado	103 (56.9%)	45 (67.2%)	58 (50.9%)
	Normal	44 (24.3%)	12 (17.9%)	32 (28.1%)
	Hipermetilado	34 (18.8%)	10 (14.9%)	24 (21.1%)
<i>SEC31B</i> (NM_015490)	Normal	11 (6.08%)	4 (5.97%)	7 (6.14%)
	Hipermetilado	170 (93.9%)	63 (94.0%)	107 (93.9%)
<i>GPM6A</i> (NM_005277)	Hipometilado	23 (12.7%)	10 (14.9%)	13 (11.4%)
	Normal	12 (6.63%)	4 (5.97%)	8 (7.02%)
	Hipermetilado	146 (80.7%)	53 (79.1%)	93 (81.6%)
<i>GPM6A</i> (NM_201591)	Normal	48 (26.5%)	16 (23.9%)	32 (28.1%)
	Hipermetilado	133 (73.5%)	51 (76.1%)	82 (71.9%)
<i>C2ORF40</i> (NM_032411)	Hipometilado	2 (1.10%)	1 (1.49%)	1 (0.88%)
	Normal	32 (17.7%)	15 (22.4%)	17 (14.9%)
	Hipermetilado	147 (81.2%)	51 (76.1%)	96 (84.2%)
<i>MAP1B</i> (NM_005909)	Hipometilado	141 (77.9%)	57 (85.1%)	84 (73.7%)
	Normal	9 (4.97%)	0 (0.00%)	9 (7.89%)
	Hipermetilado	31 (17.1%)	10 (14.9%)	21 (18.4%)
<i>FGF10</i> (NM_004465)	Hipometilado	6 (3.31%)	3 (4.48%)	3 (2.63%)
	Normal	12 (6.63%)	9 (13.4%)	3 (2.63%)
	Hipermetilado	163 (90.1%)	55 (82.1%)	108 (94.7%)

<i>TTYH2</i> (NM_032646)	Hipometilado	181 (100%)	67 (100%)	114 (100%)
<i>CHRD2</i> (NM_015424)	Hipometilado	4 (2.21%)	2 (2.99%)	2 (1.75%)
	Normal	42 (23.2%)	17 (25.4%)	25 (21.9%)
	Hipermetilado	135 (74.6%)	48 (71.6%)	87 (76.3%)
<i>SRPX</i> (NM_006307)	Hipometilado	1 (0.55%)	0 (0.00%)	1 (0.88%)
	Normal	12 (6.63%)	5 (7.46%)	7 (6.14%)
	Hipermetilado	168 (92.8%)	62 (92.5%)	106 (93.0%)
<i>PRICKLE2</i> (NM_198859)	Normal	35 (19.3%)	9 (13.4%)	26 (22.8%)
	Hipermetilado	146 (80.7%)	58 (86.6%)	88 (77.2%)
<i>TSPAN18</i> (NM_130783)	Normal	6 (3.31%)	1 (1.49%)	5 (4.39%)
	Hipermetilado	175 (96.7%)	66 (98.5%)	109 (95.6%)
Genes sobre-expresados				
<i>CDC25C</i> (NM_001790)	Hipometilado	180 (99.4%)	67 (100%)	113 (99.1%)
	Normal	1 (0.55%)	0 (0.00%)	1 (0.88%)
<i>LMNB1</i> (NM_005573)	Hipometilado	181 (100%)	67 (100%)	114 (100%)
<i>IL2RG</i> (NM_000206)	Hipometilado	1 (0.55%)	0 (0.00%)	1 (0.88%)
	Normal	39 (21.5%)	15 (22.4%)	24 (21.1%)
	Hipermetilado	141 (77.9%)	52 (77.6%)	89 (78.1%)
<i>GNLY</i> (NM_006433)	Normal	5 (2.76%)	0 (0.00%)	5 (4.39%)
	Hipermetilado	176 (97.2%)	67 (100%)	109 (95.6%)
<i>CIORF210</i> (NM_182517)	Hipometilado	1 (0.55%)	0 (0.00%)	1 (0.88%)
	Normal	70 (38.7%)	26 (38.8%)	44 (38.6%)
	Hipermetilado	110 (60.8%)	41 (61.2%)	69 (60.5%)
<i>CD52</i> (NM_001803)	Normal	3 (1.66%)	1 (1.49%)	2 (1.75%)
	Hipermetilado	178 (98.3%)	66 (98.5%)	112 (98.2%)

<i>CCL5</i> (NM_002985)	Normal	18 (9.94%)	6 (8.96%)	12 (10.5%)
	Hipermetilado	163 (90.1%)	61 (91.0%)	102 (89.5%)
<i>ETV7</i> (NM_016135)	Hipometilado	165 (91.2%)	58 (86.6%)	107 (93.9%)
	Normal	15 (8.29%)	9 (13.4%)	6 (5.26%)
<i>HIST1H2BJ</i> (NM_021058)	Hipermetilado	1 (0.55%)	0 (0.00%)	1 (0.88%)
	Hipometilado	180 (99.4%)	67 (100%)	113 (99.1%)
<i>MND1</i> (NM_032117)	Normal	1 (0.55%)	0 (0.00%)	1 (0.88%)
<i>POCIA</i> (NM_015426)	Hipometilado	181 (100%)	67 (100%)	114 (100%)
<i>FAM110C</i> (NM_001077710)	Hipometilado	181 (100%)	67 (100%)	114 (100%)
	Hipometilado	123 (68.0%)	47 (70.1%)	76 (66.7%)
	Normal	56 (30.9%)	19 (28.4%)	37 (32.5%)
<i>ASF1B</i> (NM_018154)	Hipermetilado	2 (1.10%)	1 (1.49%)	1 (0.88%)
	Hipometilado	181 (100%)	67 (100%)	114 (100%)
<i>CD3D</i> (NM_000732)	Hipometilado	181 (100%)	67 (100%)	114 (100%)
	Normal	3 (1.66%)	1 (1.49%)	2 (1.75%)
<i>PTTG1</i> (NM_004219)	Hipermetilado	178 (98.3%)	66 (98.5%)	112 (98.2%)
<i>GINS2</i> (NM_016095)	Hipometilado	181 (100%)	67 (100%)	114 (100%)
<i>ERBB3</i> (NM_001982)	Hipometilado	181 (100%)	67 (100%)	114 (100%)
	Hipometilado	180 (99.4%)	67 (100%)	113 (99.1%)
<i>PCNA</i> (NM_002592)	Normal	1 (0.55%)	0 (0.00%)	1 (0.88%)
	Hipometilado	181 (100%)	67 (100%)	114 (100%)
<i>PCNA</i> (NM_182649)	Hipometilado	181 (100%)	67 (100%)	114 (100%)

Para cada transcrito, se indica el número de casos y el porcentaje (%) correspondiente para cada una de las categorías del estado de metilación. Estado de metilación: Hipometilado ($\beta \leq 0.1$), Metilación normal ($0.1 < \beta < 0.3$), Hipermetilado ($\beta \geq 0.3$).

Los resultados también se representaron gráficamente según el grupo en el Anexo 6 y 7. En el Anexo 6, correspondiente al estado de metilación de los transcritos de genes sub-expresados, se distinguen patrones de metilación diferentes en cinco transcritos: *SPOCK1* (NM_004598), *EDIL3* (NM_005711), *TUB* (NM_177972), *MAP1B* (NM_005909) y *FGF10* (NM_004465). En los 16 transcritos restantes, los patrones de metilación son similares entre los grupos. En el Anexo 7 se observa que los patrones del estado de metilación de los genes sobre-expresados son similares entre los grupos.

A partir del análisis de regresión logística ordinal para evaluar la asociación del estado de metilación y el estado de exceso de peso, se identificaron tres genes cuyo estado de metilación estuvo asociado al exceso de peso ($p < 0.05$) (Tabla 13). Los genes fueron sub-expresados y, en los tres casos, se muestra una tendencia hacia una mayor metilación en presencia de exceso de peso: *EDIL3* (NM_005711) [OR=2.05, IC al 95%, 1.14 – 3.69], *TUB* (NM_177972) [OR=1.87, IC al 95%, 1.01-3.45] y *FGF10* (NM_004465) [OR=3.81, IC al 95%, 1.35 – 10.68].

Tabla 13 Estado de metilación en función del estado de exceso de peso

Gen (transcrito)	OR	p	[IC 95%]	
Genes sub-expresados				
<i>PLA2G2A</i> (NM_000300)	0.83	0.775	0.24	2.88
<i>SPOCK1</i> (NM_004598)	1.70	0.071	0.96	3.01
<i>CRMP1</i> (NM_001014809)	0.90	0.781	0.43	1.90
<i>CRMP1</i> (NM_001313)	1.54	0.19	0.81	2.92
<i>NOS3</i> (NM_000603)	-	-	-	-
<i>EML1</i> (NM_004434)	1.57	0.124	0.88	2.81
<i>SHANK3</i> (NM_033517)	1.23	0.46	0.71	2.16
<i>C15ORF59</i> (NM_001039614)	1.15	0.667	0.62	2.12
<i>EDIL3</i> (NM_005711)	2.05	0.016	1.14	3.69
<i>TUB</i> (NM_177972)	1.87	0.045	1.01	3.45
<i>SEC31B</i> (NM_015490)	0.97	0.963	0.27	3.45
<i>GPM6A</i> (NM_005277)	1.20	0.641	0.56	2.54
<i>GPM6A</i> (NM_201591)	0.80	0.538	0.40	1.61
<i>C2ORF40</i> (NM_032411)	1.67	0.18	0.79	3.55
<i>MAP1B</i> (NM_005909)	1.89	0.112	0.86	4.17
<i>FGF10</i> (NM_004465)	3.81	0.011	1.36	10.68
<i>TTYH2</i> (NM_032646)	-	-	-	-
<i>CHRD2</i> (NM_015424)	1.29	0.468	0.65	2.55
<i>SRPX</i> (NM_006307)	1.06	0.924	0.33	3.38

<i>PRICKLE2</i> (NM_198859)	0.53	0.127	0.23	1.20
<i>TSPAN18</i> (NM_130783)	0.33	0.317	0.04	2.89
Genes sobre-expresados				
<i>CDC25C</i> (NM_001790)	-	-	-	-
<i>LMNB1</i> (NM_005573)	-	-	-	-
<i>IL2RG</i> (NM_000206)	1.02	0.967	0.49	2.10
<i>GNLY</i> (NM_006433)	-	-	-	-
<i>CIORF210</i> (NM_182517)	0.96	0.894	0.52	1.78
<i>CD52</i> (NM_001803)	0.85	0.894	0.08	9.54
<i>CCL5</i> (NM_002985)	0.84	0.733	0.30	2.34
<i>ETV7</i> (NM_016135)	0.43	0.109	0.15	1.21
<i>HIST1H2BJ</i> (NM_021058)	-	-	-	-
<i>MND1</i> (NM_032117)	-	-	-	-
<i>POCIA</i> (NM_015426)	-	-	-	-
<i>FAM110C</i> (NM_001077710)	1.16	0.652	0.61	2.23
<i>ASF1B</i> (NM_018154)	-	-	-	-
<i>CD3D</i> (NM_000732)	0.85	0.894	0.08	9.54
<i>PTTG1</i> (NM_004219)	-	-	-	-
<i>GINS2</i> (NM_016095)	-	-	-	-
<i>ERBB3</i> (NM_001982)	-	-	-	-
<i>PCNA</i> (NM_002592)	-	-	-	-
<i>PCNA</i> (NM_182649)	-	-	-	-

Resultados del análisis de regresión logística ordinal (RLO) para evaluar la asociación del estado de metilación con el estado de exceso de peso de las participantes. OR: Odds Ratio, [IC 95%]: intervalo de confianza al 95%, p: valor de p del modelo de RLO. Los OR de tener un mayor grado de metilación se calcularon teniendo como categoría de referencia al grupo que no tiene exceso de peso (IMC<25). Se marcó con un asterisco (“*”) y se resaltó en negrita aquellos transcritos que resultaron significativos en los análisis de regresión logística ordinal. Se denota con un guión (“-“) aquellos transcritos que no pudieron ser analizados debido a que sus datos del estado de metilación correspondían a una sola categoría o tenían categorías con un bajo número de casos.

8.5.4 Nivel de metilación por tipo de muestra

Se evaluó el nivel de metilación de los genes por tipo de muestra de manera exploratoria. No se evidenció una diferencia en los niveles de metilación de la región promotora entre las muestras tumorales y las muestras sanas de cérvix en ninguno de los transcritos evaluados ($p>0.05$) (Anexo 3 y 4). Debemos remarcar que este análisis se realizó solo con tres muestras de tejido sano.

8.6 Asociación del estado de metilación con características clínicas y demográficas

Los resultados de la asociación del estado de metilación de los transcritos seleccionados en las secciones 8.5.2 y 8.5.3 con características clínicas y demográficas se resumen en la Tabla 14. Los transcritos analizados fueron ocho: *SPOCK1* (NM_004598), *EDIL3* (NM_005711), *TUB* (NM_177972), *SEC31B* (NM_015490), *GPM6A* (NM_005277), *C2ORF40* (NM_032411), *FGF10* (NM_004465) y *ETV7* (NM_016135). Se excluyó el transcrito *MND1* (NM_032117) ya que el 100% de los datos fueron hipometilados.

Los resultados muestran una tendencia hacia una mayor metilación de *FGF10* (NM_004465) en presencia de exceso de peso [OR= 3.81, IC al 95%, 1.36-10.68].

Por otro lado, se observa una tendencia a una mayor metilación asociada a la presencia de un subtipo histológico de CCE diferente al Carcinoma de células

escamosas en *SPOCK1* (NM_004598) [OR=9.29, IC al 95%, 2.60 - 33.18], *EDIL3* (NM_005711) [OR=4.87, IC al 95%, 1.63-14.53], *TUB* (NM_177972) [OR=11.72, IC al 95%, 4.78-28.72] y *ETV7* (NM_016135) [OR=6.00, IC al 95%, 2.06-17.54]. La tendencia es inversa en la metilación de *SEC31B* (NM_015490) [OR=0.10, IC al 95%, 0.03-0.36].

En cuanto a la asociación del estado de infección por VPH y el estado de metilación, se observa una tendencia hacia una menor metilación ante la infección por cepas diferentes a la VPH 16 ó 18 en *SPOCK1* (NM_004598) [OR=0.31, IC al 95%, 0.14-0.67] y *EDIL3* (NM_005711) [OR=0.38, IC al 95%, 0.16-0.91].

Se observó una asociación significativa entre el estado de menopausia y el estado de metilación de *EDIL3* (NM_005711), en donde las pacientes post-menopausicas, en comparación a las pre-menopausicas presentan una tendencia a una mayor metilación [OR=2.87, IC al 95%, 1.34-6.18].

Asimismo, se observó una asociación significativa entre la historia de uso de anticonceptivos hormonales y el estado de metilación de *C2ORF40* (NM_032411), en donde las pacientes que son usuarias antiguas o actuales en comparación a las que nunca han sido usuarias, presentan una tendencia a una mayor metilación [OR=3.85, IC al 95%, 1.08-13.70].

No se evidenció una asociación significativa entre el estado de metilación de la región promotora de los ocho transcritos seleccionados (Sección 7.7.1.3.1) y la edad al diagnóstico, el estado clínico de CCE, la raza y la historia de tabaquismo.

Tabla 14 Asociación del estado de metilación de los ocho transcritos seleccionados con las características clínicas y demográficas de las participantes

	Modelo bivariado				Modelo multivariado			
	OR crudo	p	[IC 95%]		OR ajustado	p	[IC 95%]	
SPOCK1 (NM_004598)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	-	-	-	-	-	-	-	-
Con exceso de peso (IMC≥25)	1.70	0.071	0.96	3.01	-	-	-	-
Edad al diagnóstico (n=181)	0.98	0.133	0.96	1.01	-	-	-	-
Subtipo histológico (n=181)								
Carcinoma de células escamosas (ref.)	1.00	-	-	-	1.00	-	-	-
Otros	5.47	0.000	2.24	13.39	9.29	0.001	2.60	33.18
Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-
Estadio III o IV	1.39	0.359	0.69	2.80	-	-	-	-
Estado de infección por VPH (n=147)								
VPH 16 o 18 (ref.)	1.00	-	-	-	1.00	-	-	-
Otras cepas	0.30	0.001	0.15	0.61	0.31	0.003	0.14	0.67
Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	1.00	-	-	-
Otras	0.50	0.036	0.26	0.96	0.62	0.212	0.29	1.32
Historia de tabaquismo (n=151)								
No fumadora de por vida (ref.)	1.00	-	-	-	-	-	-	-
Fumadora antigua o actual	0.66	0.171	0.36	1.20	-	-	-	-

Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	-	-	-	-
Usuaria antigua o actual	1.56	0.236	0.75	3.26	-	-	-	-
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	-	-	-	-
Post-menopausia	0.76	0.393	0.40	1.43	-	-	-	-
EDIL3 (NM_005711)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	1.00	-	-	-	1.00	-	-	-
Con exceso de peso (IMC≥25)	2.05	0.016	1.14	3.69	1.64	0.210	0.76	3.57
Edad al diagnóstico (n=181)	1.01	0.356	0.99	1.03	-	-	-	-
Subtipo histológico (n=181)								
Carcinoma de células escamosas (ref.)	1.00	-	-	-	1.00	-	-	-
Otros	3.69	0.001	1.65	8.22	4.87	0.005	1.63	14.53
Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-
Estadio III o IV	0.56	0.106	0.27	1.13	-	-	-	-
Estado de infección por VPH (n=147)								
VPH 16 o 18 (ref.)	1.00	-	-	-	1.00	-	-	-
Otras cepas	0.45	0.030	0.22	0.92	0.38	0.029	0.16	0.91
Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	-	-	-	-
Otras	1.12	0.738	0.58	2.17	-	-	-	-
Historia de tabaquismo (n=151)								

No fumadora de por vida (ref.)	1.00	-	-	-	-	-	-	-
Fumadora antigua o actual	0.55	0.060	0.30	1.02	-	-	-	-
Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	-	-	-	-
Usuaría antigua o actual	1.13	0.750	0.54	2.37	-	-	-	-
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	1.00	-	-	-
Post-menopausia	1.96	0.040	1.03	3.74	2.87	0.007	1.34	6.18
TUB (NM_177972)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	1.00	-	-	-	1.00	-	-	-
Con exceso de peso (IMC≥25)	1.87	0.045	1.01	3.45	1.77	0.121	0.86	3.63
Edad al diagnóstico (n=181)	1.00	0.721	0.97	1.02				
Subtipo histológico (n=181)								
Carcinoma de células escamosas (ref.)	1.00	-	-	-	1.00	-	-	-
Otros	11.38	0.000	4.96	26.13	11.72	0.000	4.78	28.72
Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-
Estadio III o IV	1.00	0.990	0.49	2.01	-	-	-	-
Estado de infección por VPH (n=147)								
VPH 16 o 18 (ref.)	1.00	-	-	-	-	-	-	-
Otras cepas	0.65	0.248	0.31	1.35	-	-	-	-
Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	-	-	-	-

Otras	0.91	0.784	0.46	1.80	-	-	-	-
Historia de tabaquismo (n=151)								
No fumadora de por vida (ref.)	1.00	-	-	-	1.00	-	-	-
Fumadora antigua o actual	0.47	0.022	0.24	0.90	0.50	0.050	0.25	1.00
Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	-	-	-	-
Usuaría antigua o actual	1.28	0.522	0.60	2.76	-	-	-	-
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	-	-	-	-
Post-menopausia	0.79	0.484	0.40	1.54	-	-	-	-
SEC31B (NM_015490)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	1.00	-	-	-	-	-	-	-
Con exceso de peso (IMC≥25)	0.97	0.963	0.27	3.45	-	-	-	-
Edad al diagnóstico (n=181)	1.00	0.986	0.95	1.05	-	-	-	-
Subtipo histológico (n=181)								
Carcinoma de células escamosas (ref.)	1.00	-	-	-	-	-	-	-
Otros	0.10	0.000	0.03	0.36	-	-	-	-
Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-
Estadio III o IV	2.67	0.357	0.33	21.58	-	-	-	-
Estado de infección por VPH (n=147)								
VPH 16 o 18 (ref.)	1.00	-	-	-	-	-	-	-
Otras cepas	1.65	0.652	0.19	14.62	-	-	-	-

Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	-	-	-	-
Otras	1.44	0.651	0.30	6.94	-	-	-	-
Historia de tabaquismo (n=151)								
No fumadora de por vida (ref.)	1.00	-	-	-	-	-	-	-
Fumadora antigua o actual	1.28	0.745	0.29	5.54	-	-	-	-
Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	-	-	-	-
Usaria antigua o actual	0.67	0.665	0.11	4.17	-	-	-	-
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	-	-	-	-
Post-menopausia	1.10	0.895	0.26	4.60	-	-	-	-
GPM6A (NM_005277)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	1.00	-	-	-	-	-	-	-
Con exceso de peso (IMC≥25)	1.20	0.641	0.56	2.54	-	-	-	-
Edad al diagnóstico (n=181)	1.01	0.341	0.98	1.05	-	-	-	-
Subtipo histológico (n=181)								
Carcinoma de células escamosas (ref.)	1.00	-	-	-	-	-	-	-
Otros	0.61	0.285	0.25	1.51	-	-	-	-
Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-
Estadio III o IV	0.80	0.630	0.33	1.95	-	-	-	-
Estado de infección por VPH (n=147)								

VPH 16 o 18 (ref.)	1.00	-	-	-	-	-	-	-
Otras cepas	0.73	0.498	0.29	1.83	-	-	-	-
Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	-	-	-	-
Otras	1.08	0.862	0.45	2.59	-	-	-	-
Historia de tabaquismo (n=151)								
No fumadora de por vida (ref.)	1.00	-	-	-	-	-	-	-
Fumadora antigua o actual	1.62	0.256	0.70	3.74	-	-	-	-
Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	-	-	-	-
Usuaria antigua o actual	1.79	0.249	0.66	4.82	-	-	-	-
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	-	-	-	-
Post-menopausia	1.11	0.809	0.49	2.50	-	-	-	-
C2ORF40 (NM_032411)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	1.00	-	-	-	-	-	-	-
Con exceso de peso (IMC≥25)	1.67	0.180	0.79	3.55	-	-	-	-
Edad al diagnóstico (n=181)	1.00	0.776	0.97	1.02	-	-	-	-
Subtipo histológico (n=181)								
Carcinoma de células escamosas (ref.)	1.00	-	-	-	-	-	-	-
Otros	2.45	0.162	0.70	8.58	-	-	-	-
Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-

Estadio III o IV	1.26	0.643	0.48	3.31	-	-	-	-
Estado de infección por VPH (n=147)								
VPH 16 o 18 (ref.)	1.00	-	-	-	1.00	-	-	-
Otras cepas	0.37	0.029	0.15	0.90	0.36	0.117	0.10	1.29
Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	-	-	-	-
Otras	0.68	0.364	0.29	1.57	-	-	-	-
Historia de tabaquismo (n=151)								
No fumadora de por vida (ref.)	1.00	-	-	-	-	-	-	-
Fumadora antigua o actual	1.50	0.367	0.62	3.62	-	-	-	-
Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	1.00	-	-	-
Usaria antigua o actual	3.25	0.027	1.15	9.24	3.85	0.038	1.08	13.70
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	-	-	-	-
Post-menopausia	1.01	0.986	0.44	2.30	-	-	-	-
FGF10 (NM_004465)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	1.00	-	-	-	-	-	-	-
Con exceso de peso (IMC≥25)	3.81	0.011	1.36	10.68	-	-	-	-
Edad al diagnóstico (n=181)	1.00	0.827	0.96	1.03	-	-	-	-
Subtipo histológico (n=181)								
Carcinoma de células escamosas (ref.)	1.00	-	-	-	-	-	-	-
Otros	0.99	0.991	0.27	3.66	-	-	-	-

Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-
Estadio III o IV	2.23	0.300	0.49	10.16	-	-	-	-
Estado de infección por VPH (n=147)								
VPH 16 o 18 (ref.)	1.00	-	-	-	-	-	-	-
Otras cepas	0.47	0.207	0.14	1.53	-	-	-	-
Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	-	-	-	-
Otras	1.57	0.495	0.43	5.75	-	-	-	-
Historia de tabaquismo (n=151)								
No fumadora de por vida (ref.)	1.00	-	-	-	-	-	-	-
Fumadora antigua o actual	1.55	0.447	0.50	4.76	-	-	-	-
Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	-	-	-	-
Usaria antigua o actual	1.32	0.689	0.33	5.24	-	-	-	-
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	-	-	-	-
Post-menopausia	0.56	0.270	0.20	1.56	-	-	-	-
ETV7 (NM_016135)								
Estado de exceso de peso (n=181)								
Sin exceso de peso (IMC<25) (ref.)	1.00	-	-	-	-	-	-	-
Con exceso de peso (IMC≥25)	0.43	0.109	0.15	1.21	-	-	-	-
Edad al diagnóstico (n=181)	1.01	0.625	0.97	1.05	-	-	-	-
Subtipo histológico (n=181)								

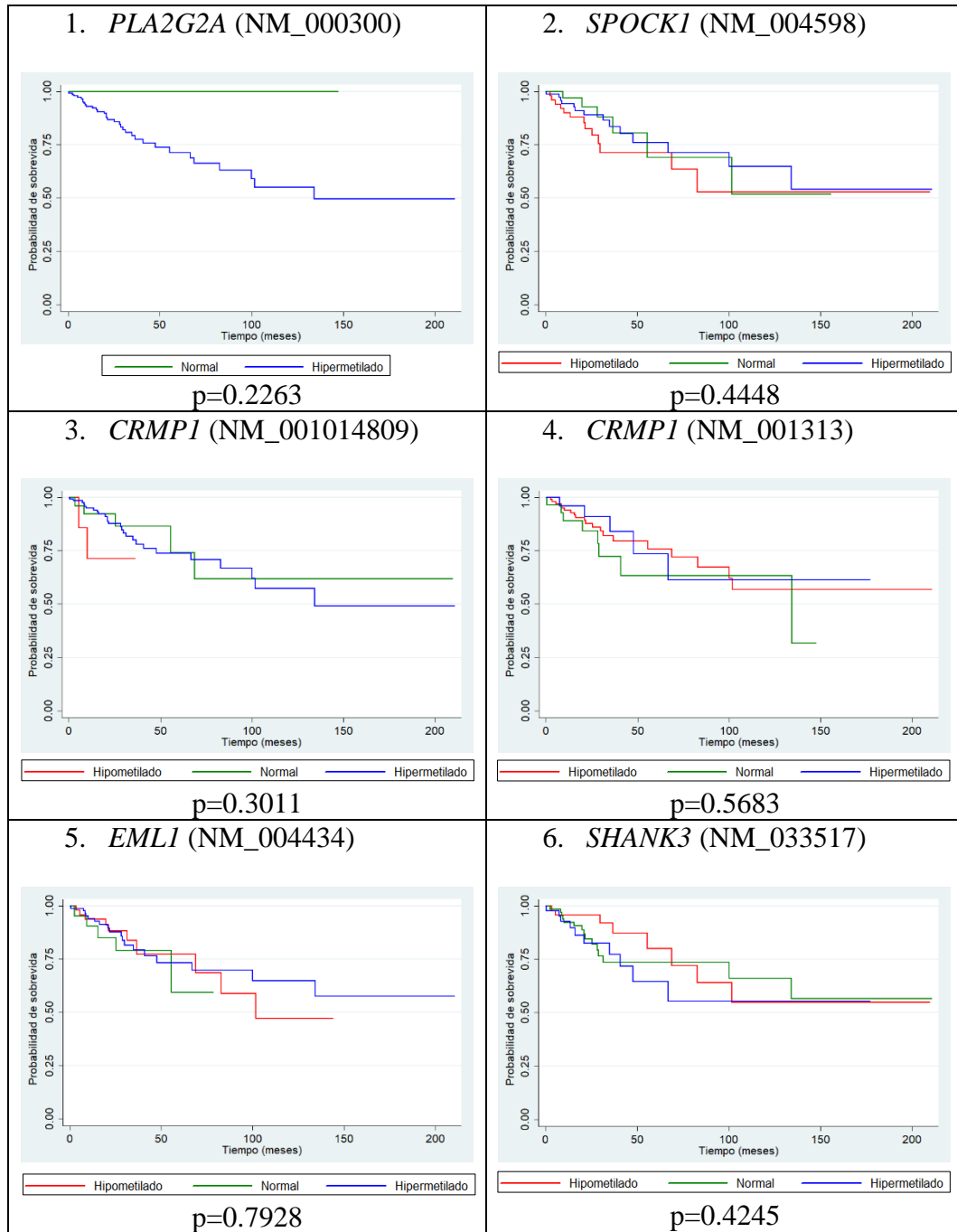
Carcinoma de células escamosas (ref.)	1.00	-	-	-	-	-	-	-
Otros	6.00	0.001	2.06	17.54	-	-	-	-
Estadio del tumor (n=177)								
Estadio I o II (ref.)	1.00	-	-	-	-	-	-	-
Estadio III o IV	0.89	0.860	0.24	3.30	-	-	-	-
Estado de infección por VPH (n=147)								
VPH 16 o 18 (ref.)	1.00	-	-	-	-	-	-	-
Otras cepas	0.42	0.430	0.05	3.57	-	-	-	-
Raza (n=172)								
Blanca (ref.)	1.00	-	-	-	-	-	-	-
Otras	0.78	0.708	0.21	2.90	-	-	-	-
Historia de tabaquismo (n=151)								
No fumadora de por vida (ref.)	1.00	-	-	-	-	-	-	-
Fumadora antigua o actual	0.80	0.713	0.25	2.58	-	-	-	-
Historia de uso de anticonceptivos hormonales (n=101)								
Nunca ha usado (ref.)	1.00	-	-	-	-	-	-	-
Usuaría antigua o actual	1.63	0.423	0.49	5.36	-	-	-	-
Estado de menopausia (n=147)								
Pre-menopausia (ref.)	1.00	-	-	-	-	-	-	-
Post-menopausia	0.48	0.277	0.13	1.80	-	-	-	-

Resultados del análisis de regresión logística ordinal (RLO) para evaluar la asociación del estado de metilación con características clínicas y demográficas de las participantes. OR: Odds Ratio, [IC 95%]: intervalo de confianza al 95%, p: valor de p del modelo de RLO. Los OR de tener un mayor grado de metilación según cada característica clínica o demográfica se calcularon en base a una categoría de referencia, indicada por el término "(ref.)". Se denota con un guión ("-") aquellos análisis que no fueron ejecutados ya sea porque se trata de la categoría de referencia utilizada en el modelo de RLO o porque la variable independiente no fue significativa en el modelo bivariado y por lo tanto fue excluida del análisis en un modelo multivariado.

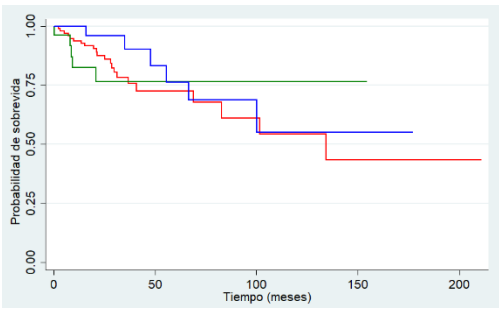
8.7 Sobrevida global en función del estado de metilación del ADN

Se evaluó la asociación del estado de metilación de la región promotora de los 40 transcritos de estudio en la sobrevida global en la población. Se excluyeron aquellos transcritos que presentaron más del 97% de datos para una sola categoría en el estado de metilación, por lo que, se analizaron 24 transcritos en total. Los resultados se detallan en la Figura 6. Solo se encontró una asociación significativa para el estado de metilación del gen sub-expresado *PRICKLE2* (NM_198859) y la sobrevida global, en donde la curva K-M sugiere una mayor sobrevida en las pacientes que presentan hipermetilación en el promotor de *PRICKLE2* en comparación a las que presentan una metilación normal ($p=0.0022$) (Figura 6, imagen 18).

Figura 6 Gráficas de KM para representar la supervivencia global en función del estado de metilación de los transcritos



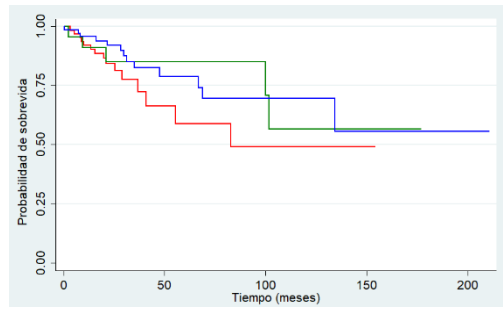
7. *C15ORF59* (NM_001039614)



Hipometilado Normal Hipermetilado

p=0.6311

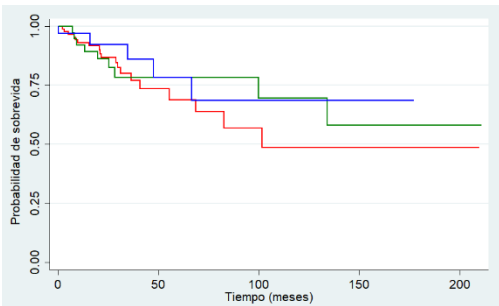
8. *EDIL3* (NM_005711)



Hipometilado Normal Hipermetilado

p=0.3396

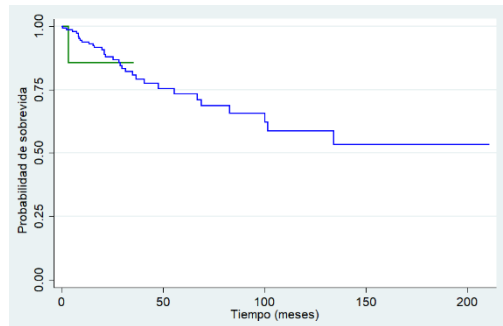
9. *TUB* (NM_177972)



Hipometilado Normal Hipermetilado

p=0.6840

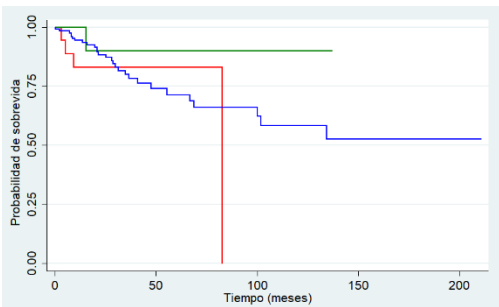
10. *SEC31B* (NM_015490)



Normal Hipermetilado

p=0.7846

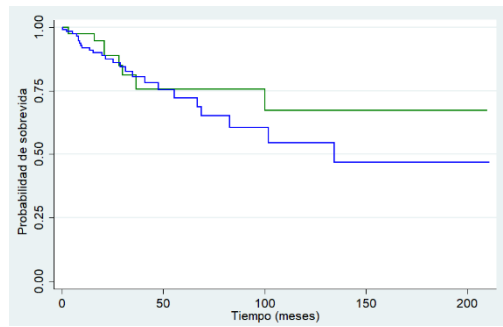
11. *GPM6A* (NM_005277)



Hipometilado Normal Hipermetilado

p=0.2217

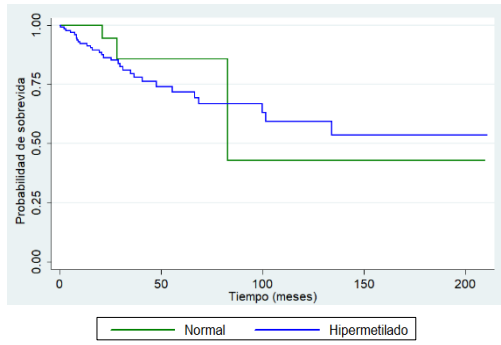
12. *GPM6A* (NM_201591)



Normal Hipermetilado

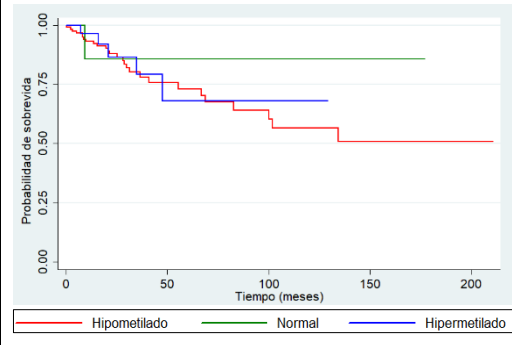
p=0.4115

13. *C2ORF40* (NM_032411)



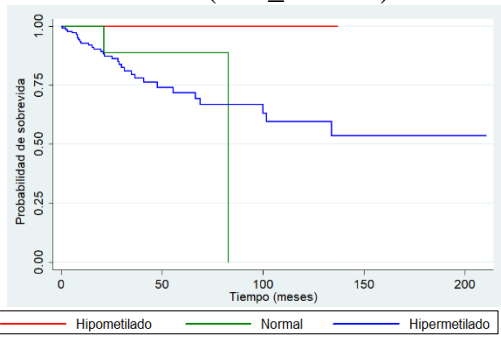
p=0.7377

14. *MAP1B* (NM_005909)



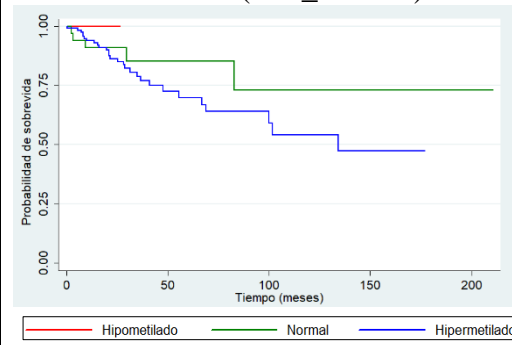
p=0.7837

15. *FGF10* (NM_004465)



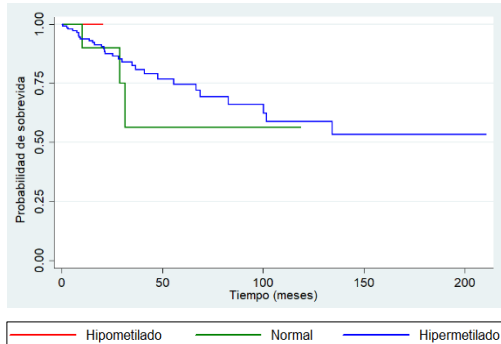
p=0.6045

16. *CHRD2* (NM_015424)



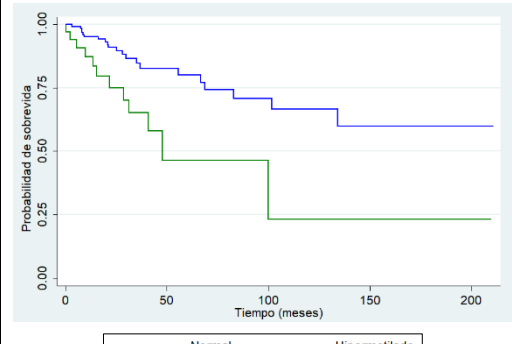
p=0.4534

17. *SRPX* (NM_006307)



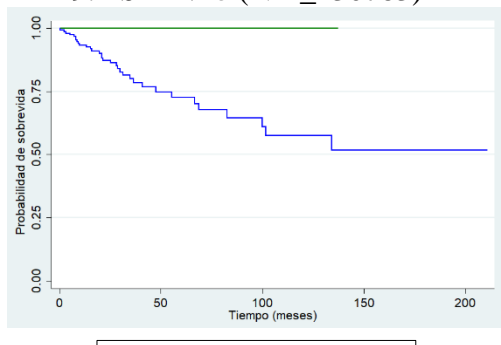
p=0.7525

18. *PRICKLE2* (NM_198859)



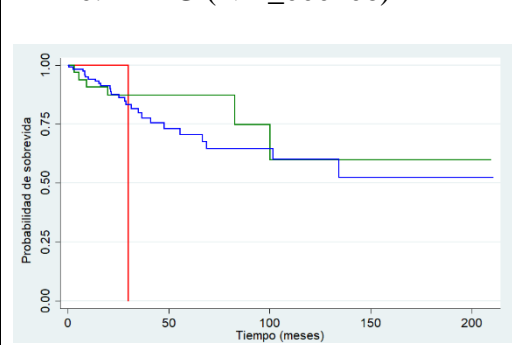
p=0.0022*

19. *TSPAN18* (NM_130783)

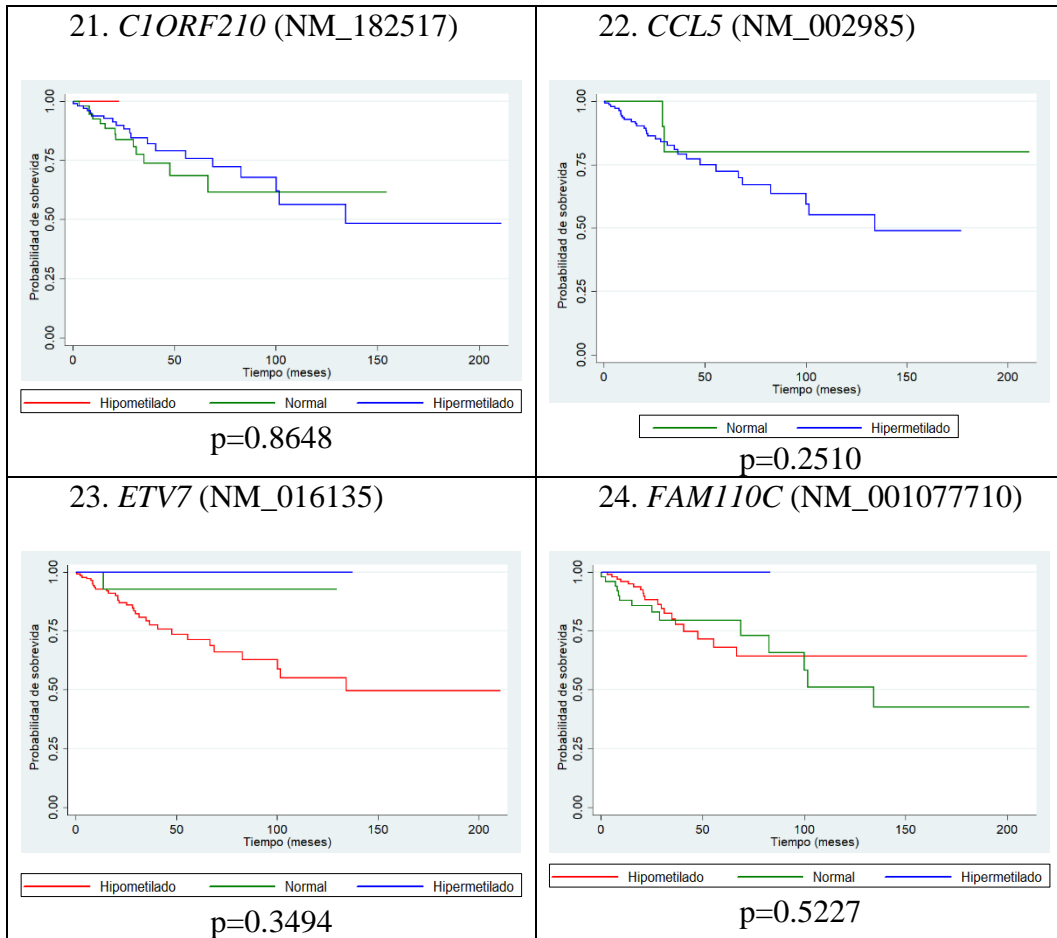


p=0.2605

20. *IL2RG* (NM_000206)



p=0.1296



Curvas de supervida global K-M según el estado de metilación de cada Gen (transcrito). Rojo: Hipometilado ($\beta \leq 0.1$), Verde: Metilación normal ($0.1 < \beta < 0.3$), Azul: Hipermetilado ($\beta \geq 0.3$). Se indican los valores de p obtenidos en las pruebas log-rank para evaluar las diferencias entre las curvas de supervida. Se marcó con un asterisco (“*”) y se resaltó en **negrita** aquellos transcritos que tuvieron un valor de $p < 0.05$ en la prueba estadística.

9 DISCUSIÓN

Mediante un análisis secundario de datos genómicos y epigenómicos del proyecto TCGA-CESC, el presente estudio buscaba evaluar la existencia de una firma epigenética de metilación en regiones promotoras de DEGs en CCE asociada al estado de exceso de peso y si esta a su vez estaba asociada a la sobrevida.

Se identificaron siete DEGs en CCE que presentaron diferencias en el nivel metilación en la región promotora en función del estado de exceso de peso. Dichos genes comprometieron cinco genes sub-expresados (*SPOCK1* [NM_004598], *EDIL3* [NM_005711], *SEC31B* [NM_015490], *GPM6A* [NM_005277], *C2ORF40* [NM_032411]) y dos genes sobre-expresados (*ETV7* [NM_016135] y *MND1* [NM_032117]). Hasta donde tenemos conocimiento, este es el primer reporte que describe la asociación de la metilación de DEGs en CCE en asociación con el estado de exceso de peso.

La metilación en la región promotora de genes es un mecanismo epigenético importante en la regulación de la expresión génica en cáncer. Si dicha región génica se encuentra hipermetilada, la transcripción del gen tiende a estar inactiva, en contraste, si se encuentra hipometilada, la transcripción tiende a estar activa (18).

Nuestros resultados evidencian que, en los cinco DEGs sub-expresados en CCE, el grupo con exceso de peso presenta un mayor nivel de metilación que el grupo sin exceso de peso, con lo cual se esperaría que los niveles de expresión en el primer grupo fuesen menores. Los resultados de expresión indican que esto se cumple para *SPOCK1*, *EDIL3*, *GPM6A* y *C2ORF40*, mas no para *SEC31B*, en el que se observó que el grupo con exceso de peso tenía una mayor expresión que el grupo sin exceso de peso. Este hallazgo podría indicar que para *SPOCK1*, *EDIL3*, *GPM6A* y

C2ORF40 la metilación en la región promotora asociada al estado de exceso de peso juega un rol en la regulación de su expresión en CCE, mientras que, para el gen *SEC31B* otros mecanismos epigenéticos distintos de la metilación al ADN podrían estar involucrados. Estos resultados coinciden con lo reportado en otros estudios de metilación génica en cáncer asociada al estado de peso, como en cáncer colorrectal y cáncer de mama, en donde pacientes obesos presentaban mayores niveles de metilación que pacientes no obesos (12,66). Dicha tendencia se evidenció en sitios CpG correspondientes a los genes *TMEM106A*, *ZNF543*, *ZNF3970S*, *ADCYA*, *DLX4*, *GSX1* y *SNCA* en el estudio de cáncer colorrectal (12) y, en el caso de cáncer de mama, en sitios CpG correspondientes a los genes *SH3BP2* y *XIST* (66).

En contraste, en los dos DEGs sobre-expresados en CCE (*ETV7* y *MND1*) el grupo con exceso de peso presentó menores niveles de metilación, con lo cual se esperaría que los niveles de expresión en este grupo fuesen mayores en comparación al grupo sin exceso de peso. Los resultados de expresión coinciden con lo esperado y se evidencia que, en las pacientes con exceso de peso, *ETV7* y *MND1* se expresan más que en las pacientes sin exceso de peso. Estos resultados sugieren que, también en estos genes, la metilación en la región promotora en función del estado de exceso de peso podría influenciar su expresión en CCE.

La metilación en DEGs en CCE también fue evaluada considerando el estado de metilación en función del estado de exceso de peso. Se evidenció que, en presencia de exceso de peso hay una tendencia hacia una mayor metilación en los DEGs sub-expresados *EDIL3* (NM_005711), *TUB* (NM_177972) y *FGF10* (NM_004465). Al realizar el análisis de asociación del estado de metilación y el estado de exceso de

peso considerando también características clínicas y demográficas de la población (edad, subtipo histológico de CCE, estadio del tumor, estado de infección por VPH, raza, historia de tabaquismo, historia de uso de anticonceptivos hormonales y estado de menopausia), solo se evidenció una asociación significativa entre el estado de metilación del gen *FGF10* y el estado de exceso de peso [OR=3.81, IC al 95%; 1.36-10.68]. Debemos resaltar, sin embargo, que el IC de esta asociación es amplio, lo que sugiere cierto grado de imprecisión en el cálculo del OR, posiblemente relacionado a una baja muestra en las categorías del estado de metilación del gen. La relación entre la metilación de *FGF10* y el estado de peso o CCE no se conoce a detalle. Se ha reportado que el gen *FGF10* presenta niveles de expresión elevados en tejido adiposo de individuos con obesidad extrema (67). En nuestro estudio, las pacientes con exceso de peso tuvieron una menor expresión de *FGF10* en comparación a las que no tienen exceso de peso ($p=0.0073$), lo que sugiere que la metilación en la región promotora podría tener un rol en su regulación génica. Es importante recalcar que la clasificación del estado de metilación de cada gen en CCE no se basó en controles de tejido sano de cérvix debido a que el estudio TCGA-CESC solo cuenta con tres muestras sanas con información epigenómica. Ante la limitación en el número de controles, se empleó una estrategia alternativa para clasificar el estado de metilación en base a puntos de cortes del coeficiente β reportados y empleados en otros estudios de epigenómica del cáncer (55,56). En consecuencia, es probable que el estado de metilación asignado para cada gen (transcrito) en cada paciente no represente el verdadero estado de metilación en CCE (hiper- o hipo- metilado con respecto a un estado sano), información que

podría brindar un mejor entendimiento de la asociación entre IMC, CCE y metilación génica.

De los 39 DEGs en CCE que se expresan distinto en pacientes con exceso de peso y sin exceso de peso, solo en dos de ellos (*CCL5* y *PCNA*) se ha reportado una asociación entre su expresión, cáncer y adiposidad. D'Esposito et al (2016) reportaron que las células adiposas crean un microambiente que contribuye a que células de cáncer de mama triple negativo adquieran un carácter invasivo y diseminativo, en parte, debido a la producción de la quimiocina *CCL5* (68). Por otro lado, *PCNA* es una proteína indicadora de crecimiento celular y en un estudio *in vitro* de Nair et al (2017) se reportó que los niveles de esta proteína incrementaban cuando se co-cultivaban células de adipocitos humanos (células SW872) con células epiteliales glandulares endometriales, en comparación a cuando se cultivaban dichas células solas (69). La adición de las células de adipocitos favorecía la proliferación celular de las células endometriales, en parte, debido a la secreción del factor de necrosis tumoral alfa ($TNF-\alpha$).

Dentro de los hallazgos más relevantes de nuestro estudio se encuentra que los 20 DEGs sobre-expresados en CCE seleccionados enriquecen rutas relacionadas al ciclo celular, replicación del ADN, procesos de infección y/o del sistema inmune. Este hallazgo coincide con lo reportado por Hair et al (2015), en donde se encontró que la obesidad estaba asociada a la metilación de genes involucrados en la respuesta inmune, crecimiento celular y reparación del ADN en tumores de cáncer de mama (66). Se debe precisar que, en dicho estudio, se evaluó la asociación del IMC de las pacientes y la metilación en un panel definido de genes seleccionados

en un estudio previo por su relevancia en cáncer (70), el cual consistió en 935 sitios CpG.

De manera similar, en nuestro estudio se analizó la potencial asociación entre el estado de exceso de peso y la metilación en la región promotora en 39 DEGs en CCE que se expresan distinto en pacientes con exceso de peso y sin exceso de peso. Este enfoque buscaba identificar DEGs en CCE cuya expresión pudiese estar desregulada debido a la metilación en la región promotora y que ello estuviera asociada al estado de exceso de peso. Una limitación de nuestra aproximación metodológica es que, debido a que el panel de genes analizado fue delimitado a DEGs en CCE (en total 1695 genes), solo se consideró un subconjunto de genes cuya metilación podría estar diferenciada en función del estado de exceso de peso. Por lo tanto, es probable que nuestros resultados estén sesgados hacia genes involucrados en cáncer, posibilidad que también es discutida en el estudio de Hair et al (2015) debido al enfoque metodológico empleado (66). Esto podría explicar por qué nuestros hallazgos difieren de lo reportado por Crujeiras et al (2018) en donde se analizaron alrededor de 450 000 sitios CpG y se identificó un metiloma de cáncer colorrectal asociado a obesidad con rutas enriquecidas relacionadas a procesos metabólicos e inflamatorios (12). En este contexto, el estudio del metiloma completo en pacientes con CCE brindaría un panorama más completo de los potenciales sitios de metilación diferenciados en función del estado de exceso de peso, por lo que resulta un área que amerita ser explorada.

Adicionalmente, se debe remarcar que no se aplicó un método de corrección por pruebas múltiples en los análisis estadísticos de expresión diferencial y metilación en función del estado de exceso de peso. Se decidió emplear los valores de p crudos

con la finalidad de analizar un subconjunto de genes en términos de expresión, metilación y enriquecimiento de redes PPI y rutas biológicas. Bajo este enfoque, fue posible seleccionar siete genes cuyos niveles de metilación en la región promotora estarían asociados al estado de exceso de peso en CCE, los cuales deberían de ser estudiados a mayor profundidad para determinar su rol en CCE. Los estudios de asociación entre metilación, obesidad y cáncer no son muy abundantes, y se han realizado principalmente en cáncer colorrectal (12) y cáncer de mama (40,66,71). Por otro lado, a pesar de que existe limitada evidencia del sobrepeso u obesidad como factor de riesgo de CCE (36), es interesante notar que en nuestra muestra las pacientes con exceso de peso conformaron aproximadamente el 62% de la muestra.

Si bien no hubo una red PPI significativa ni términos GO o KEGG enriquecidos por los 19 DEGs sub-expresados en CCE analizados, resulta interesante que ocho de ellos tengan un rol en el sistema nervioso central (*SPOCK*, *CRMP1*, *EML1*, *SHANK3*, *C15ORF59*, *GPM6A*, *C2ORF40*, *MAP1B*). Estos hallazgos guardan relación con lo encontrado en un estudio reciente de Xu et al (2019) en donde se realizó un análisis integrativo de metilación y expresión génica en TCGA-CESC, encontrándose que en genes hipermetilados sub-expresados había un enriquecimiento rutas de la transmisión sináptica química y el desarrollo del sistema nervioso, entre otros términos de GO (72).

El estudio de la asociación entre la metilación génica, cáncer y exceso de peso es complejo ya que la relación entre estos tres componentes no es unidireccional, es decir, la metilación podría ser la causa o la consecuencia de la presencia de exceso de peso (73). En este estudio, clasificamos a las pacientes con CCE según el estado

de exceso de peso definido por un valor de corte de IMC, al momento del diagnóstico, mayor o igual a 25. Nuestra hipótesis fue que el estado de exceso de peso en CCE está asociado a una firma epigenética de metilación y que esto tendría una implicancia en la sobrevida global. Sin embargo, en la relación de estos elementos debemos considerar también que, de manera independiente, existen firmas de metilación del ADN descritas en obesidad (74) al igual que en CCE (7,8,29,31,75).

Debemos remarcar que ninguno de los 39 DEGs identificados en nuestro estudio está asociado con obesidad en adultos según nuestra búsqueda en literatura y en bases de datos como la herramienta de expresión génica “*Expression Atlas*” (76) y el catálogo “*GWAS Catalog*” (77) que contiene información de estudios de asociación del genoma completo (GWAS, por sus siglas en inglés). De cierta forma, esto nos indicaría que su expresión/metilación no se debería a la obesidad. Resulta necesario entonces, tomar en consideración los distintos factores mencionados al interpretar nuestros resultados, no obstante, nuestro estudio logra revelar información innovadora acerca de genes con diferencias en la metilación en la región promotora en función del estado de exceso de peso a partir de un subconjunto de DEGs en CCE. Solo en el caso del gen *MND1*, un estudio reportó su hipermetilación en niños varones obesos en comparación a niños no obesos, sin una diferencia de expresión entre los grupos de estudio (78). No obstante, debemos señalar que en este estudio se analizaron células sanguíneas y solo se evaluaron 12 individuos, todos de ancestría asiática.

Con respecto a la asociación del estado de metilación de los genes y características clínicas y demográficas de las pacientes, se evidenció una asociación entre el estado

de metilación de los genes *SPOCK1* (NM_004598), *EDIL3* (NM_005711), *TUB* (NM_177972), *ETV7* (NM_016135) y *SEC31B* (NM_015490) y el subtipo histológico de CCE. No existen reportes de una asociación entre alguno de estos genes y el subtipo histológico de CCE, sin embargo, en estudios previos con datos de la cohorte TCGA-CESC, se han descrito diferencias en patrones de metilación en función del subtipo histológico (44,72).

Por otro lado, se observa una tendencia hacia una menor metilación en la región promotora de *SPOCK1* (NM_004598) y *EDIL3* (NM_005711) ante la infección por cepas diferentes a la VPH 16 ó 18. Estos resultados sugieren un potencial rol de la infección del VPH en alteraciones epigenéticas en CCE en donde una infección con VPH 16 ó 18, cepas de alto riesgo, estaría asociada a una mayor metilación de *SPOCK1* y *EDIL3*, lo que podría conllevar a una menor expresión génica. Una menor expresión de *SPOCK1* disminuiría las interacciones entre las células y la matriz extracelular, mientras que, en *EDIL3* favorecería la formación de estructuras vasculares, por lo que se trataría de regulaciones que podrían tener un impacto en la progresión de la neoplasia. Existe evidencia de que la infección por cepas de VPH causa alteraciones de metilación en ciertos genes del hospedero humano (75). Uno de los mecanismos sería a través de la oncoproteína viral E7 de la cepa VPH 16, la cual puede estimular la actividad metiltransferasa de la enzima DNMT1 *in vitro* (79). Asimismo, se ha evidenciado que la proteína E7 de VPH 16 aumenta la metilación en las regiones promotoras de los genes *CXCL14* (80) y *CCNA1* (81) y ello conllevaría a una disminución en sus niveles de expresión.

Se observó también que en las pacientes post-menopáusicas hubo una tendencia a una mayor metilación de *EDIL3* (NM_005711) en comparación a las pre-

menopáusicas, lo que podría sugerir que en este grupo hay una menor expresión de *EDIL3*. El estado de menopausia ha sido reportado como un factor asociado tanto a la metilación génica como al IMC en otros tipos de cáncer, como el cáncer de mama (40,66).

Por último, se encontró una asociación entre el estado de metilación de la región promotora de *PRICKLE2* y la sobrevida global en CCE, en donde la curva K-M sugiere una mayor sobrevida en las pacientes que presentan hipermetilación en el promotor de *PRICKLE2* en comparación a las que presentan una metilación normal ($p=0.0022$). El rol de *PRICKLE2*, ya sea a nivel de expresión o metilación, y la sobrevida en cáncer no ha sido descrita anteriormente.

En nuestro estudio, *PRICKLE2* estuvo sub-expresado en CCE. La sub-expresión de este gen también ha sido reportada en carcinoma renal de células claras (82), aunque se desconocen los mecanismos de regulación implicados y su rol concreto en la carcinogénesis. *PRICKLE2* es una proteína homóloga a la proteína Prickle de *Drosophila* (83). Se sabe que participa en la vía de señalización de Wnt, una ruta implicada en varios tipos de cáncer (84) y en donde se modulan elementos del citoesqueleto y se controla la polaridad de la célula. La asociación entre *PRICKLE2* y CCE solo ha sido descrita en un estudio de Senchenko et al (2013) en donde se postula a este gen como potencial marcador para discriminar entre carcinoma de células escamosas y adenocarcinomas (85).

Solo hemos encontrado un estudio en donde se ha evaluado el potencial rol de la metilación del ADN, la obesidad y la sobrevida en cáncer de mama realizado por McCullough et al (2016) (71). En este, se evidenció que la mortalidad específica

por cáncer de mama fue más alta en pacientes obesas con metilación en la región promotora de los genes *APC* y *TWIST1* en tejidos tumorales.

En este contexto, sería conveniente una mayor investigación en el potencial rol del estado de metilación de *PRICKLE2* en CCE, pues si bien no hubo diferencias en su metilación en función al estado de peso en nuestra cohorte estudiada, la expresión fue mayor en el grupo con exceso de peso en comparación al grupo sin exceso de peso. Se debe notar que, en muchos casos no fue posible evaluar la sobrevida en función del estado de metilación de los genes debido a que el número de pacientes en algunas categorías del estado de metilación era muy bajo. Ello impidió tanto la ejecución del análisis como la detección de diferencias significativa entre las categorías comparadas.

Los resultados de nuestro estudio sugieren una potencial asociación entre el estado de exceso de peso y la metilación en la región promotora de algunos DEGs en CCE. Es importante reconocer que estos resultados se basan en un análisis secundario de datos, por lo que sería necesario corroborar nuestros hallazgos en posteriores estudios con un mayor número de participantes y con ensayos experimentales.

10 CONCLUSIONES

- Se identificaron 39 DEGs en CCE con diferencias de expresión génica entre las pacientes sin exceso de peso y con exceso de peso en la cohorte TCGA-CESC.
- Se reconocieron diferencias en el nivel metilación en la región promotora de siete DEGs en CCE en función del estado de exceso de peso en la cohorte TCGA-CESC, los cuales incluyeron a los cinco genes sub-expresados *SPOCK1* (NM_004598), *EDIL3* (NM_005711), *SEC31B* (NM_015490), *GPM6A* (NM_005277), *C2ORF40* (NM_032411) y a los dos genes sobre-expresados *ETV7* (NM_016135) y *MND1* (NM_032117).
- Hubo una asociación significativa entre el estado de metilación del gen sub-expresado *PRICKLE2* (NM_198859) y la sobrevida global en la cohorte TCGA-CESC, en donde la curva K-M sugiere una mayor sobrevida en las pacientes que presentan hipermetilación en comparación a las que presentan una metilación normal.

11 LIMITACIONES

- Una limitación de nuestro estudio es que el análisis de genes estuvo delimitado a DEGs en CCE en la cohorte TCGA-CESC, por lo que solo se consideró un subconjunto de genes cuya metilación podría estar diferenciada en función del estado de exceso de peso.
- Por otro lado, no se analizó el estado de metilación en base a controles de tejido sano debido al limitado número de muestras sanas con información epigenómica, lo cual podría brindar más información acerca de la asociación entre IMC, CCE y metilación.

12 RECOMENDACIONES

- Se recomienda evaluar la asociación entre exceso de peso en CCE y metilación en regiones promotoras de DEGs en estudios con un mayor número de participantes, pues ayudarían a corroborar lo reportado en el presente estudio.
- Para posteriores estudios en el tema, se recomienda tomar en consideración la potencial influencia de los eventos de pérdida y ganancia de peso en pacientes con CCE en el metiloma.
- Para posteriores estudios sería interesante analizar la metilación de otras regiones génicas (no promotores) de DEGs en CCE, pues también podrían estar asociadas al estado de exceso de peso y a la expresión génica.

13 REFERENCIAS BIBLIOGRÁFICAS

1. World Health Organization. Cervical cancer, Early diagnosis and screening [Internet]. 2018 [cited 2019 Jan 31]. Available from: <https://www.who.int/cancer/prevention/diagnosis-screening/cervical-cancer/en/>
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* [Internet]. 2018 Nov;68(6):394–424. Available from: <http://doi.wiley.com/10.3322/caac.21492>
3. World Health Organization. Human papillomavirus (HPV) and cervical cancer [Internet]. 2019 [cited 2019 Jan 31]. Available from: [https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-\(hpv\)-and-cervical-cancer](https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer)
4. Saavedra KP, Brebi PM, Roa JCS. Epigenetic alterations in preneoplastic and neoplastic lesions of the cervix. *Clin Epigenetics* [Internet]. 2012;4(1):13. Available from: <http://clinicaepigeneticsjournal.biomedcentral.com/articles/10.1186/1868-7083-4-13>
5. Szalmás A, Kónya J. Epigenetic alterations in cervical carcinogenesis. *Semin Cancer Biol*. 2009;19(3):144–52.
6. Fang J, Zhang H, Jin S. Epigenetics and cervical cancer: From pathogenesis to therapy. *Tumor Biol*. 2014;35(6):5083–93.
7. Steenbergen RDM, Snijders PJF, Heideman DAM, Meijer CJLM. Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. *Nat Rev Cancer* [Internet]. 2014;14(6):395–405. Available from: <http://dx.doi.org/10.1038/nrc3728>
8. Bhat S, Kabekkodu SP, Noronha A, Satyamoorthy K. Biological implications and therapeutic significance of DNA methylation regulated genes in cervical cancer. *Biochimie* [Internet]. 2016;121:298–311. Available from: <http://dx.doi.org/10.1016/j.biochi.2015.12.018>
9. Clark LH, Jackson AL, Soo AE, Orrey DC, Gehrig PA, Kim KH. Extremes in body mass index affect overall survival in women with cervical cancer. *Gynecol Oncol* [Internet]. 2016;141(3):497–500. Available from: <http://dx.doi.org/10.1016/j.ygyno.2016.03.035>
10. Xu X, Zhou L, Miao R, Chen W, Zhou Y, Pang Q, et al. Association of cancer mortality with postdiagnosis overweight and obesity using body mass index. *Oncotarget*. 2015;7(4):5023–9.
11. Grigsby P, Elhammali A, Ruiz F, Markovina S, McLellan MD, Miller CA, et al. Clinical outcomes and differential effects of PI3K pathway mutation in obese versus non-obese patients with cervical cancer. *Oncotarget* [Internet]. 2018;9(3):4061–73. Available from: <http://www.oncotarget.com/fulltext/23664%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/29423104%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5790521>
12. Crujeiras AB, Morcillo S, Diaz-Lagares A, Sandoval J, Castellano-Castillo D, Torres E, et al. Identification of an episinature of human colorectal cancer associated with obesity by genome-wide DNA methylation analysis.

- Int J Obes [Internet]. 2018;1–13. Available from: <http://dx.doi.org/10.1038/s41366-018-0065-6>
13. Wu C, Morris JR. Genes, Genetics, and Epigenetics: A Correspondence. *Science* (80-). 2001;293(5532):1103–5.
 14. Lopomo A, Coppedè F. Epigenetic Signatures in the Diagnosis and Prognosis of Cancer [Internet]. *Epigenetic Mechanisms in Cancer*. Elsevier Inc.; 2018. 313–343 p. Available from: <https://doi.org/10.1016/B978-0-12-809552-2.00012-7>
 15. Esteller M. Epigenetics in Cancer. *N Engl J Med*. 2008;358(11):1148–59.
 16. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol* [Internet]. 2010;28(10):1057–68. Available from: <http://dx.doi.org/10.1038/nbt.1685>
 17. Ropero S, Esteller M. Epigenetics and Cancer: DNA Methylation. In: Esteller M, editor. *Epigenetics in Biology and Medicine*. 1st ed. CRC Press; 2009. p. 3–16.
 18. Kulis M, Esteller M. DNA Methylation and Cancer. *Adv Genet*. 2010;70:27–56.
 19. Robertson KD. DNA methylation and human disease. *Nat Rev Genet*. 2005;6(8):597–610.
 20. Jones PA. Functions of DNA methylation : islands , start sites , gene bodies and beyond. *Nat Rev Genet* [Internet]. 2012;13(7):484–92. Available from: <http://dx.doi.org/10.1038/nrg3230>
 21. Sandoval J, Esteller M. Cancer epigenomics : beyond genomics. *Curr Opin Genet Dev* [Internet]. 2012;22(1):50–5. Available from: <http://dx.doi.org/10.1016/j.gde.2012.02.008>
 22. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* [Internet]. 1983;301(5895):89–92. Available from: <https://doi.org/10.1038/301089a0>
 23. Gama-sosal MA, Slagell VA, Trewyn RW, Oxenhandler R, Kuo C, Gehrke CW, et al. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res*. 1983;11(19):6883–94.
 24. Fraga MF, Herranz M, Espada J, Ballestar E, Paz MF, Ropero S, et al. A Mouse Skin Multistage Carcinogenesis Model Reflects the Aberrant DNA Methylation Patterns of Human Tumors. *Cancer Res*. 2004;64(16):5527–34.
 25. World Health Organization. The Global Cancer Observatory, Global burden of cervical cancer [Internet]. 2018. 2018 [cited 2019 Jan 31]. Available from: <https://gco.iarc.fr/today/data/factsheets/cancers/23-Cervix-uteri-fact-sheet.pdf>
 26. MFMER. Cervical cancer: Symptoms & causes [Internet]. [cited 2020 Mar 15]. Available from: <https://www.mayoclinic.org/diseases-conditions/cervical-cancer/symptoms-causes/syc-20352501>
 27. Walboomers JMM, Jacobs M V., Manos MM, Bosch FX, Kummer JA, Shah K V., et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* [Internet]. 1999 Sep;189(1):12–9. Available from: <http://doi.wiley.com/10.1002/%28SICI%291096-9896%28199909%29189%3A1%3C12%3A%3AAID-PATH431%3E3.0.CO%3B2-F>

28. Woodman CBJ, Collins SI, Young LS. The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer* [Internet]. 2007 Jan 1;7(1):11–22. Available from: <http://www.nature.com/articles/nrc2050>
29. Soto D, Song C, McLaughlin-Drubin ME. Epigenetic alterations in human papillomavirus- associated cancers. *Viruses*. 2017;9(9).
30. Kim YI, Giuliano A, Hatch KD, Schneider A, Nour MA, Dallal GE, et al. Global DNA hypomethylation increases progressively in cervical dysplasia and carcinoma. *Cancer* [Internet]. 1994 Aug 1;74(3):893–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8039116>
31. Dueñas-González A, Lizano M, Candelaria M, Cetina L, Arce C, Cervera E. Epigenetics of cervical cancer. An overview and therapeutic perspectives. *Mol Cancer*. 2005;4:1–24.
32. World Health Organization. Obesity and overweight [Internet]. 2018 [cited 2019 Feb 12]. Available from: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>
33. Center for Disease Control and Prevention. Cancer and obesity. Overweight and obesity are associated with cancer. [Internet]. 2017 [cited 2019 Jan 31]. Available from: <https://www.cdc.gov/vitalsigns/obesity-cancer/index.html>
34. WCRF/AICR. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective [Internet]. 2018. Available from: <http://www.dietandcancerreport.org>
35. Park J, Morley TS, Kim M, Clegg DJ, Scherer PE. Obesity and cancer - Mechanisms underlying tumour progression and recurrence. *Nat Rev Endocrinol* [Internet]. 2014;10(8):455–65. Available from: <http://dx.doi.org/10.1038/nrendo.2014.94>
36. Norat, Teresa., Abar, Leil, Vieira Ana Rita, Chan, Doris, Vingeliene, Snieguole, Polemiti E. World Cancer Research Fund International Systematic Literature Review The Associations between Food, Nutrition and Physical Activity and the Risk of Cervical Cancer [Internet]. 2018. Available from: <https://www.wcrf.org/sites/default/files/Cervical-cancer-slr.pdf>
37. Milagro FI, Mansego ML, De Miguel C, Martínez JA. Dietary factors, epigenetic modifications and obesity outcomes: Progresses and perspectives. *Mol Aspects Med* [Internet]. 2013;34(4):782–812. Available from: <http://dx.doi.org/10.1016/j.mam.2012.06.010>
38. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, et al. DNA methylation and body-mass index: A genome-wide analysis. *Lancet*. 2014;383(9933):1990–8.
39. Milagro FI, Gómez-Abellán P, Campián J, Martínez JA, Ordovás JM, Garaulet M. CLOCK, PER2 and BMAL1 DNA methylation: Association with obesity and metabolic syndrome characteristics and monounsaturated fat intake. *Chronobiol Int*. 2012;29(9):1180–94.
40. Crujeiras AB, Diaz-Lagares A, Stefansson OA, Macias-Gonzalez M, Sandoval J, Cueva J, et al. Obesity and menopause modify the epigenomic profile of breast cancer. *Endocr Relat Cancer*. 2017;24(7):351–63.
41. TCGA. The Cancer Genome Atlas Homepage. About TCGA. 2018 [Internet]. [cited 2019 Jan 13]; Available from: <https://cancergenome.nih.gov/abouttcga>

42. Tomczak K, Czerwińska P, Wiznerowicz M. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkol* [Internet]. 2015;1A:68–77. Available from: <http://www.termedia.pl/doi/10.5114/wo.2014.47136>
43. NCI. TCGA Cancers Selected for Study [Internet]. [cited 2020 Feb 17]. Available from: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/studied-cancers>
44. Burk RD, Chen Z, Saller C, Tarvin K, Carvalho AL, Scapulatempo-Neto C, et al. Integrated genomic and molecular characterization of cervical cancer. *Nature* [Internet]. 2017 Mar 16;543(7645):378–84. Available from: <http://www.nature.com/articles/nature21386>
45. Srivastava A, George J, Karuturi RKM. Transcriptome Analysis. In: *Encyclopedia of Bioinformatics and Computational Biology* [Internet]. Elsevier; 2019. p. 792–805. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780128096338201611>
46. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* [Internet]. 2009 Jan;10(1):57–63. Available from: <http://www.nature.com/articles/nrg2484>
47. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLOS Comput Biol*. 2017;13(5):e1005457.
48. Finotello F, Di Camillo B. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Brief Funct Genomics* [Internet]. 2015 Mar 1;14(2):130–42. Available from: <https://academic.oup.com/bfg/article-lookup/doi/10.1093/bfgp/elu035>
49. McDermid A, Monier B, Zhao J, Liu B, Ma Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief Bioinform* [Internet]. 2019 Nov 27;20(6):2044–54. Available from: <https://academic.oup.com/bib/article/20/6/2044/5066173>
50. Chen L, Wong G. Transcriptome Informatics. In: *Encyclopedia of Bioinformatics and Computational Biology* [Internet]. Elsevier; 2019. p. 324–40. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B9780128096338202045>
51. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330–8.
52. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. 2019;47(October 2018):590–5.
53. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG : new perspectives on genomes , pathways , diseases and drugs. 2016;1–9.
54. NIH. CDE Curation Tool [Internet]. [cited 2019 Jun 10]. Available from: <https://cdecurate.nci.nih.gov/cdecurate/NCICurationServlet?reqType=view&idseq=D4A6A07C-55F4-25A1-E034-0003BA12F5E7>
55. Hinoue T, Weisenberger DJ, Lange CPE, Shen H, Byun H-M, Van Den Berg D, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res* [Internet]. 2012 Feb 1;22(2):271–82. Available from: <http://genome.cshlp.org/cgi/doi/10.1101/gr.117523.110>
56. Akbani R, Akdemir KC, Aksoy BA, Albert M, Ally A, Amin SB, et al.

- Genomic Classification of Cutaneous Melanoma. *Cell* [Internet]. 2015 Jun;161(7):1681–96. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867415006340>
57. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Arman B, et al. In Focus The cBio Cancer Genomics Portal : An Open Platform for Exploring Multidimensional Cancer Genomics Data. 2012;(May).
 58. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA : a web server for cancer and normal gene expression profiling and interactive analyses. 2017;45(April):98–102.
 59. Goldman M, Craft B, Hastie M, Repe K, Mcdade F, Kamath A. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. :1–16.
 60. Huang WY, Hsu S Da, Huang HY, Sun YM, Chou CH, Weng SL, et al. MethHC: A database of DNA methylation and gene expression in human cancer. *Nucleic Acids Res*. 2015;43(D1):D856–61.
 61. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* [Internet]. 2019 Jan 8;47(D1):D607–13. Available from: <https://academic.oup.com/nar/article/47/D1/D607/5198476>
 62. Akram DS, Astrup A V., Atinmo T, Boissin JL, Bray GA, Carroll KK, et al. Obesity: Preventing and managing the global epidemic. World Health Organization - Technical Report Series. 2000.
 63. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell*. 2018;33(4):690-705.e9.
 64. Subirana I, Sanz H, Vila J. Building Bivariate Tables: The compareGroups Package for R. *J Stat Softw* [Internet]. 2014;57(12):1–16. Available from: <http://www.jstatsoft.org/v57/i12/>.
 65. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.
 66. Hair BY, Troester MA, Edmiston SN, Parrish EA, Robinson WR, Wu MC, et al. Body Mass Index Is Associated with Gene Methylation in Estrogen Receptor-Positive Breast Tumors. *Cancer Epidemiol Biomarkers Prev* [Internet]. 2015 Mar 1;24(3):580–6. Available from: <http://cebp.aacrjournals.org/cgi/doi/10.1158/1055-9965.EPI-14-1017>
 67. Gerhard GS, Styer AM, Strodel WE, Roesch SL, Yavorek A, Carey DJ, et al. Gene expression profiling in subcutaneous, visceral and epigastric adipose tissues of patients with extreme obesity. *Int J Obes* [Internet]. 2014 Mar 16;38(3):371–8. Available from: <http://www.nature.com/articles/ijo2013152>
 68. D’Esposito V, Liguoro D, Ambrosio MR, Collina F, Cantile M, Spinelli R, et al. Adipose microenvironment promotes triple negative breast cancer cell invasiveness and dissemination by producing CCL5. *Oncotarget* [Internet]. 2016 Apr 26;7(17):24495–509. Available from: <http://www.oncotarget.com/fulltext/8336>
 69. Nair S, Nguyen H, Salama S, Al-Hendy A. Obesity and the Endometrium: Adipocyte-Secreted Proinflammatory TNF α Cytokine Enhances the

- Proliferation of Human Endometrial Glandular Cells. *Obstet Gynecol Int* [Internet]. 2013;2013:1–7. Available from: <http://www.hindawi.com/journals/ogi/2013/368543/>
70. Conway K, Edmiston SN, May R, Kuan PF, Chu H, Bryant C, et al. DNA methylation profiling in the Carolina Breast Cancer Study defines cancer subclasses differing in clinicopathologic characteristics and survival. *Breast Cancer Res* [Internet]. 2014 Oct 7;16(5):450. Available from: <http://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-014-0450-6>
 71. McCullough LE, Chen J, Cho YH, Khankari NK, Bradshaw PT, White AJ, et al. DNA methylation modifies the association between obesity and survival after breast cancer diagnosis. *Breast Cancer Res Treat* [Internet]. 2016 Feb 5;156(1):183–94. Available from: <http://link.springer.com/10.1007/s10549-016-3724-0>
 72. Xu W, Xu M, Wang L, Zhou W, Xiang R, Shi Y, et al. Integrative analysis of DNA methylation and gene expression identified cervical cancer-specific diagnostic biomarkers. *Signal Transduct Target Ther* [Internet]. 2019 Dec 13;4(1):55. Available from: <http://dx.doi.org/10.1038/s41392-019-0081-6>
 73. Crujeiras AB, Casanueva FF. Obesity and the reproductive system disorders: epigenetics as a potential bridge. *Hum Reprod Update* [Internet]. 2015 Mar 1;21(2):249–61. Available from: <http://academic.oup.com/humupd/article/21/2/249/784796/Obesity-and-the-reproductive-system-disorders>
 74. van Dijk SJ, Molloy PL, Varinli H, Morrison JL, Muhlhausler BS. Epigenetics and human obesity. *Int J Obes* [Internet]. 2015 Jan 25;39(1):85–97. Available from: <http://dx.doi.org/10.1038/ijo.2014.34>
 75. Feng C, Dong J, Chang W, Cui M, Xu T. The Progress of Methylation Regulation in Gene Expression of Cervical Cancer. *Int J Genomics* [Internet]. 2018;2018:1–11. Available from: <https://www.hindawi.com/journals/ijg/2018/8260652/>
 76. Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, et al. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res* [Internet]. 2016 Jan 4;44(D1):D746–52. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1045>
 77. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* [Internet]. 2019 Jan 8;47(D1):D1005–12. Available from: <https://academic.oup.com/nar/article/47/D1/D1005/5184712>
 78. Rhee J-K, Lee J-H, Yang HK, Kim T-M, Yoon K-H. DNA Methylation Profiles of Blood Cells Are Distinct between Early-Onset Obese and Control Individuals. *Genomics Inform* [Internet]. 2017;15(1):28. Available from: <http://genominfo.org/journal/view.php?doi=10.5808/GI.2017.15.1.28>
 79. Burgers WA, Blanchon L, Pradhan S, Launoit Y de, Kouzarides T, Fuks F. Viral oncoproteins target the DNA methyltransferases. *Oncogene* [Internet]. 2007 Mar 18;26(11):1650–5. Available from:

- <http://www.nature.com/articles/1209950>
80. Cicchini L, Westrich JA, Xu T, Vermeer DW, Berger JN, Clambey ET, et al. Suppression of Antitumor Immune Responses by Human Papillomavirus through Epigenetic Downregulation of CXCL14. *MBio* [Internet]. 2016 Jul 6;7(3):1–13. Available from: <http://mbio.asm.org/lookup/doi/10.1128/mBio.00270-16>
 81. Chalertpet K, Pakdeechaidan W, Patel V, Mutirangura A, Yanatatsaneejit P. Human papillomavirus type 16 E7 oncoprotein mediates CCNA1 promoter methylation. *Cancer Sci* [Internet]. 2015 Oct;106(10):1333–40. Available from: <http://doi.wiley.com/10.1111/cas.12761>
 82. Rudenko EE, Gerashchenko G V., Lapska Y V., Vozianov SO, Zgonnyk YM, Kashuba VI. PPM1M and PRICKLE2 are potential tumor suppressor genes in human clear-cell renal cell carcinoma. *Biopolym Cell* [Internet]. 2014 May 20;30(3):229–33. Available from: <http://biopolymers.org.ua/doi/10.7124/bc.00089A>
 83. Katoh M, Katoh M. Identification and characterization of human PRICKLE1 and PRICKLE2 genes as well as mouse Prickle1 and Prickle2 genes homologous to *Drosophila* tissue polarity gene prickle. *Int J Mol Med* [Internet]. 2003 Feb 1;11(2):249–56. Available from: <http://www.spandidos-publications.com/10.3892/ijmm.11.2.249>
 84. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene* [Internet]. 2017 Mar 12;36(11):1461–73. Available from: <http://www.nature.com/articles/onc2016304>
 85. Senchenko VN, Kisseljova NP, Ivanova TA, Dmitriev AA, Krasnov GS, Kudryavtseva A V., et al. Novel tumor suppressor candidates on chromosome 3 revealed by NotI-microarrays in cervical cancer. *Epigenetics* [Internet]. 2013 Apr 27;8(4):409–20. Available from: <http://www.tandfonline.com/doi/abs/10.4161/epi.24233>

14 ANEXOS

Anexo 1

```
//Script para ejecutar pruebas t-student y guardar variables de resultados//
*****
tempname memhold
tempfile results
postfile `memhold' name pval t statistic df t mean 1 sd 1 mean 2 sd 2 n1 n2 p 1 p u
combined sd est error using `results'
foreach var of varlist krt17-prkcz {
  ttest `var',by(estado_obesidad)
  post `memhold' (`var') (`r(p)') (`r(t)') (`r(df_t)') (`r(mu_1)') (`r(sd_1)') (`r(mu_2)')
  ) (`r(sd_2)') (`r(N_1)') (`r(N_2)') (`r(p_1)') (`r(p_u)') (`r(sd)') (`r(se)')
}
postclose `memhold'
use `results', clear
list
*****
```

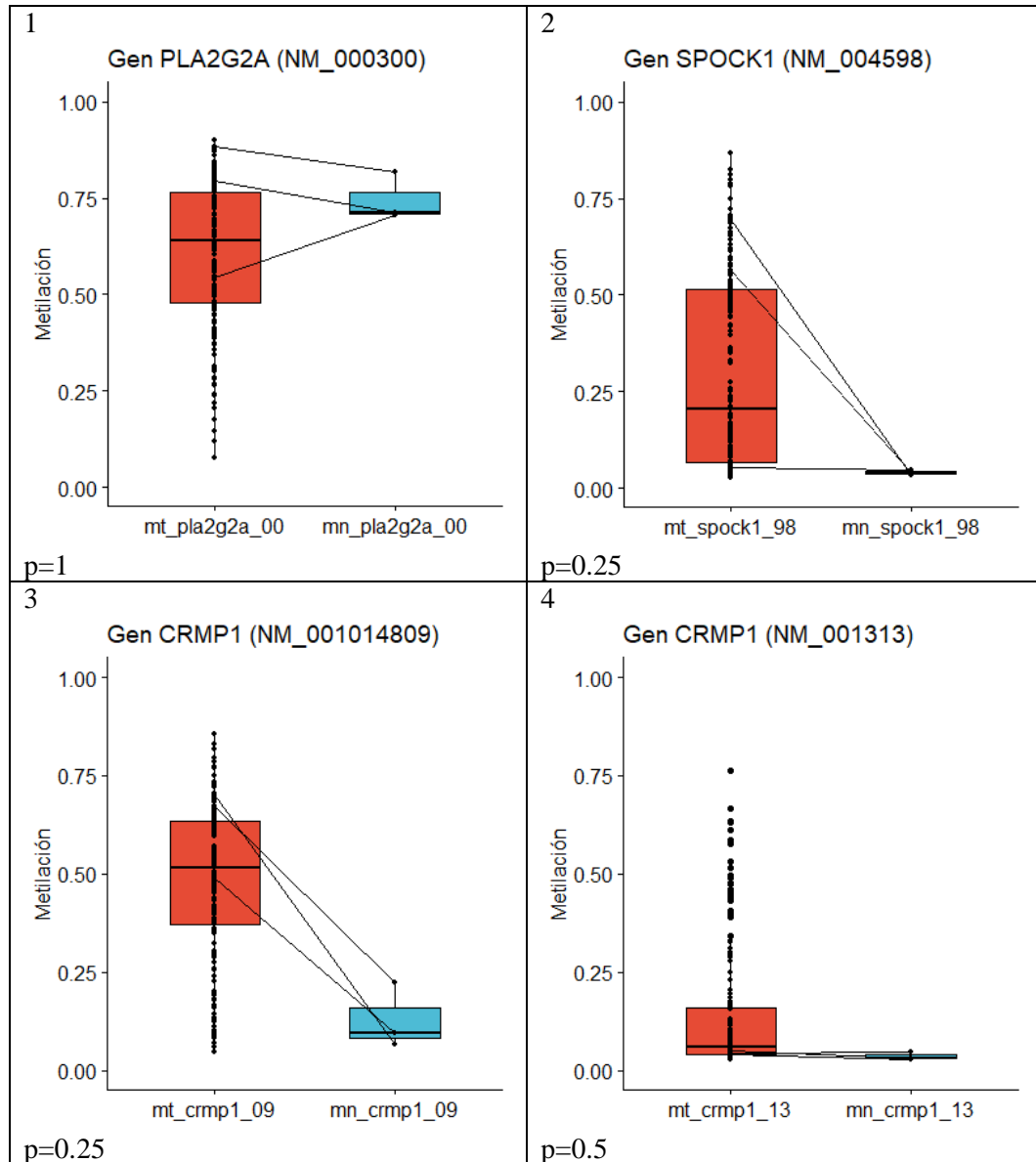
Anexo 2

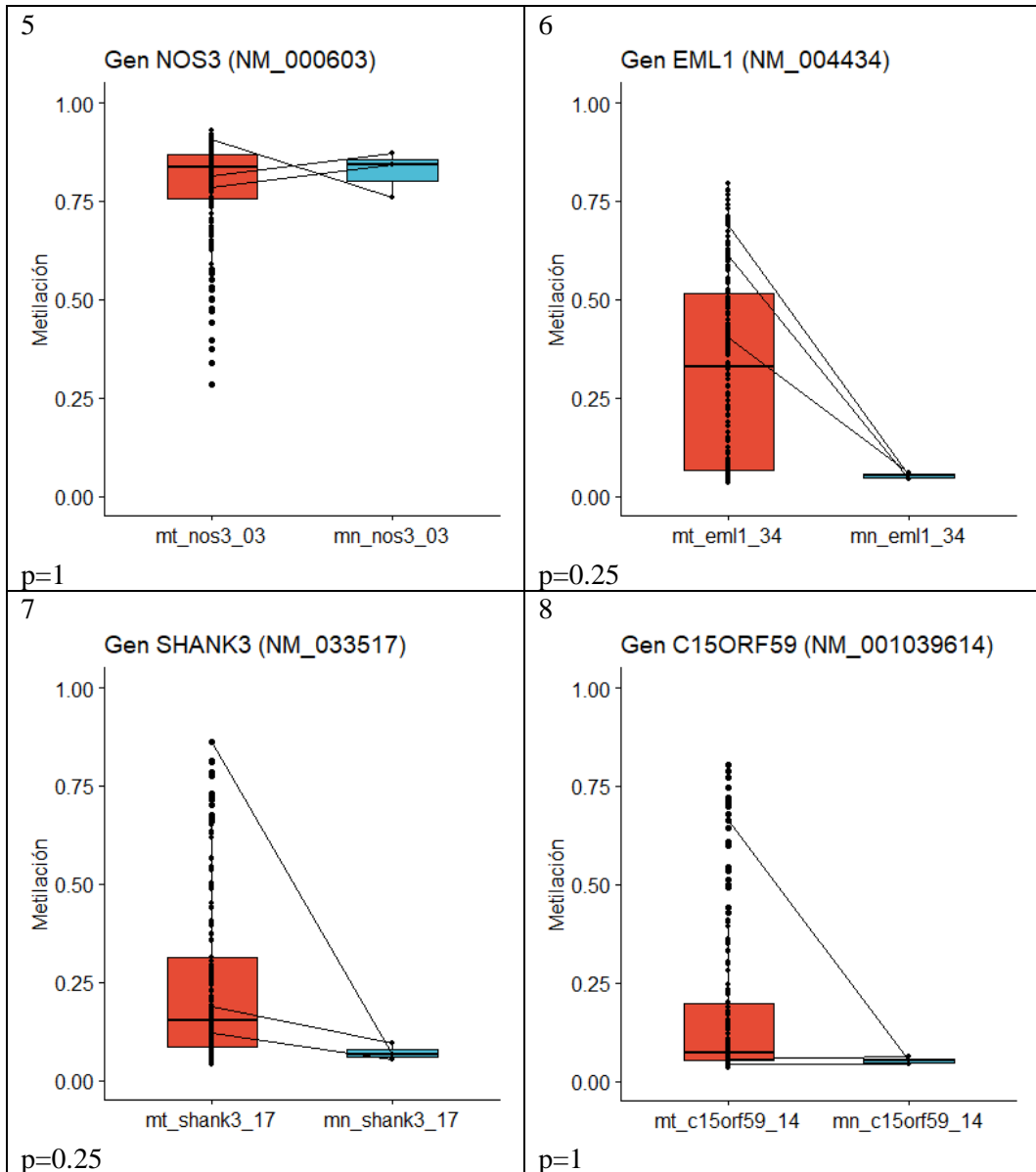
```
//Script para ejecutar pruebas t-student y guardar variables de resultados//
*****
tempname memhold
tempfile results
postfile `memhold' name pval t_statistic df_t mean_1 sd_1 mean_2 sd_2 n1 n2 p_1 p_u
combined_sd est_error using `results'
foreach var of varlist pla2g2a-scn7a {
  ttest `var',by(group_sample)
  post `memhold' (`var') (`r(p)') (`r(t)') (`r(df_t)') (`r(mu_1)') (`r(sd_1)') (`r(mu_2)')
  ) (`r(sd_2)') (`r(N_1)') (`r(N_2)') (`r(p_1)') (`r(p_u)') (`r(sd)') (`r(se)')
}
postclose `memhold'
use `results', clear
list
*****
```

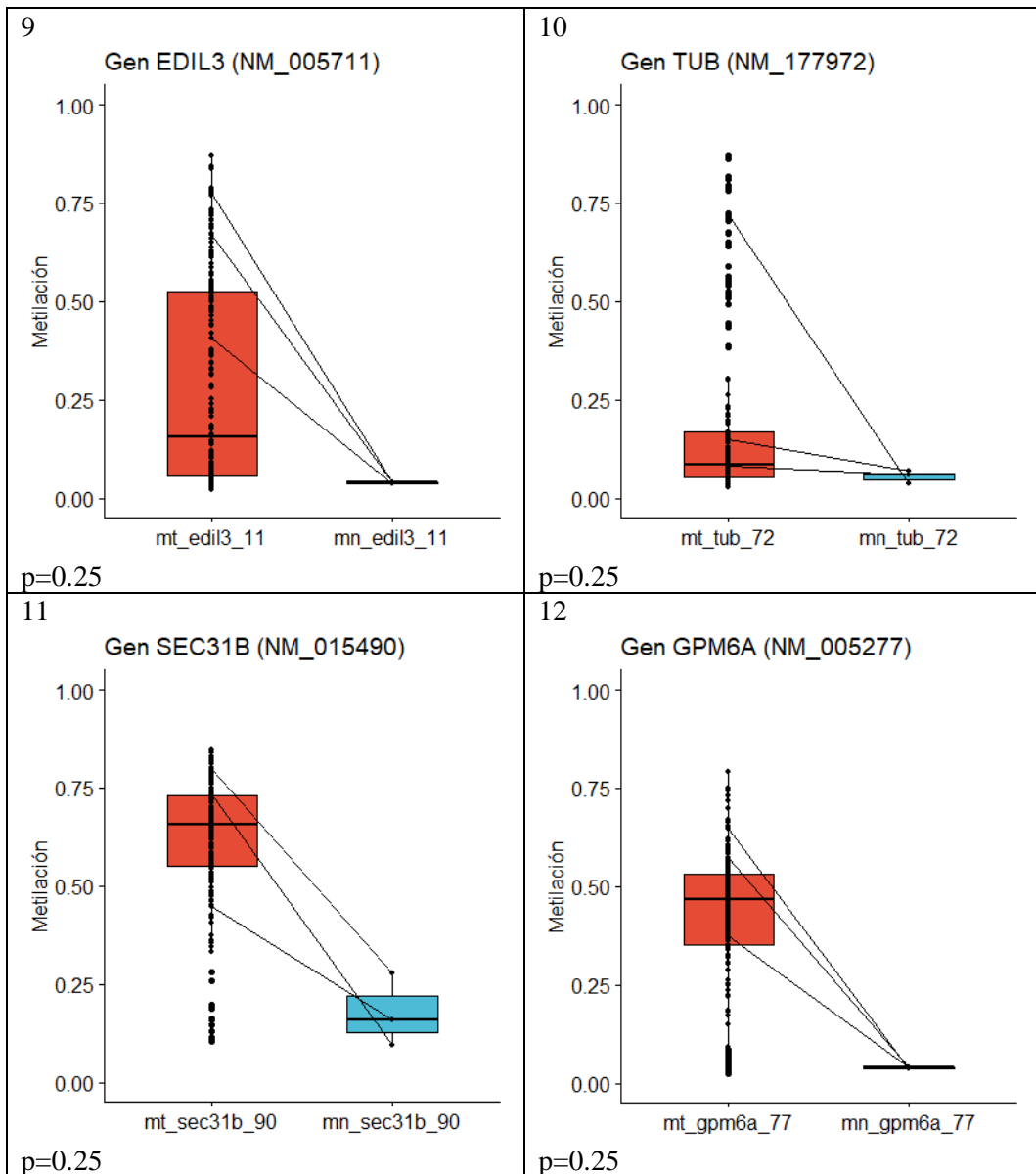
Anexo 3

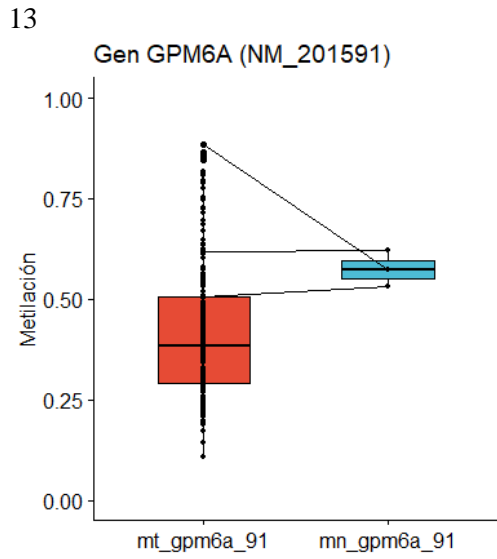
Transcritos de genes sub-expresados (21 transcritos)

Gráficos de cajas del nivel de metilación metilación en las muestras tumorales y muestras sanas de cérvix para los 21 transcritos seleccionados correspondientes a los genes sub-expresados y valores de p de las pruebas de los rangos con signo de Wilcoxon. Se indica el nombre del gen, los dos últimos números de su respectivo código de identificación y un prefijo que indica si la muestra es tumoral o sana (mt: muestra tumoral, mn:muestra).

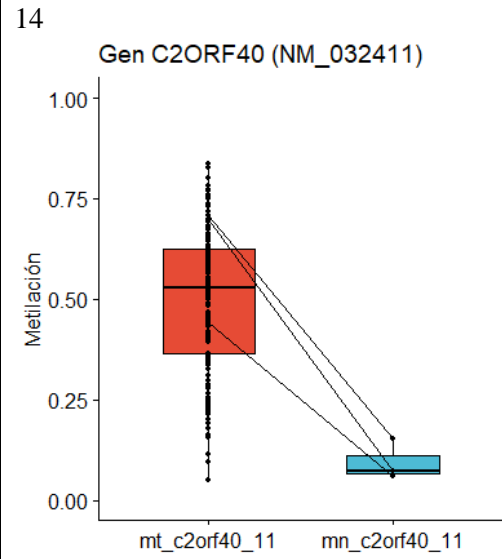




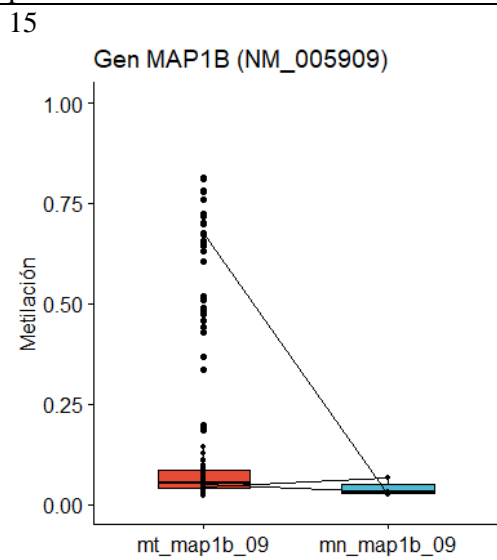




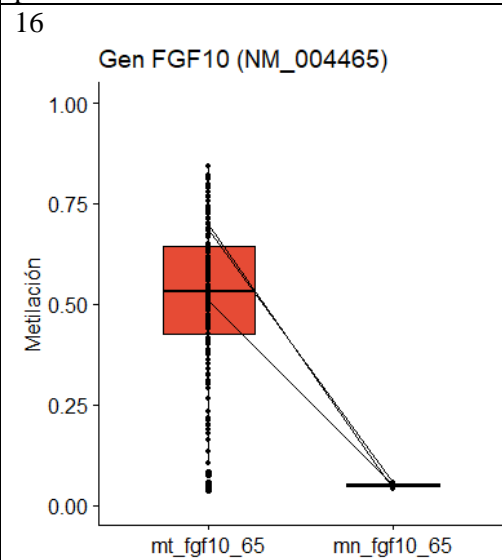
p=1



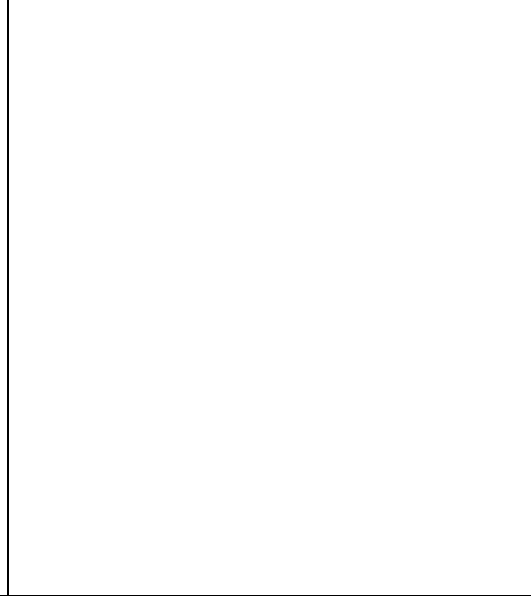
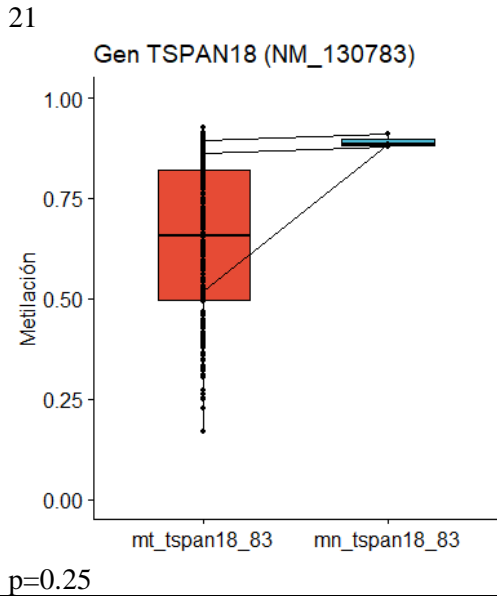
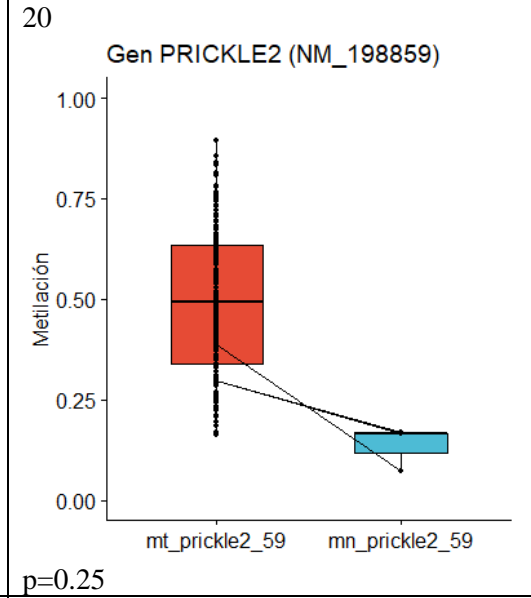
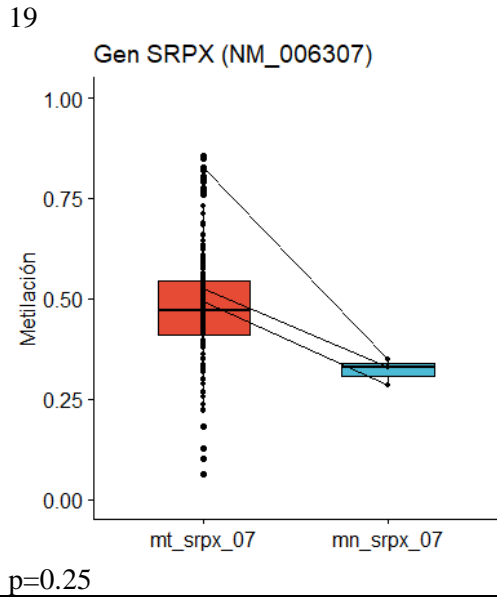
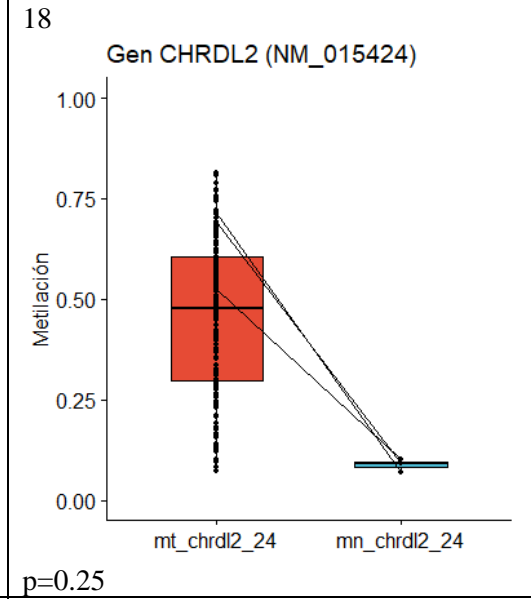
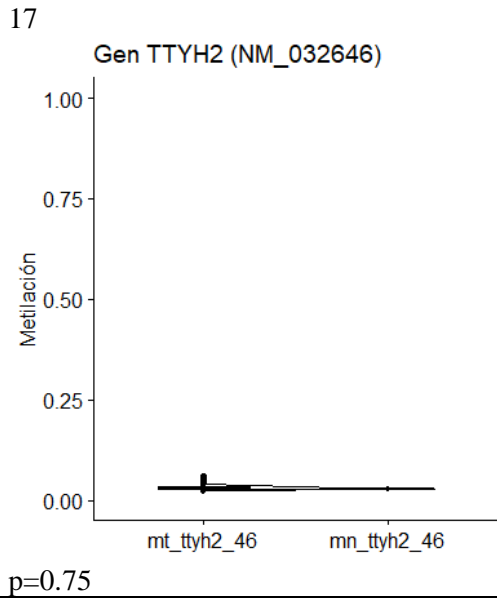
p=0.25



p=0.75



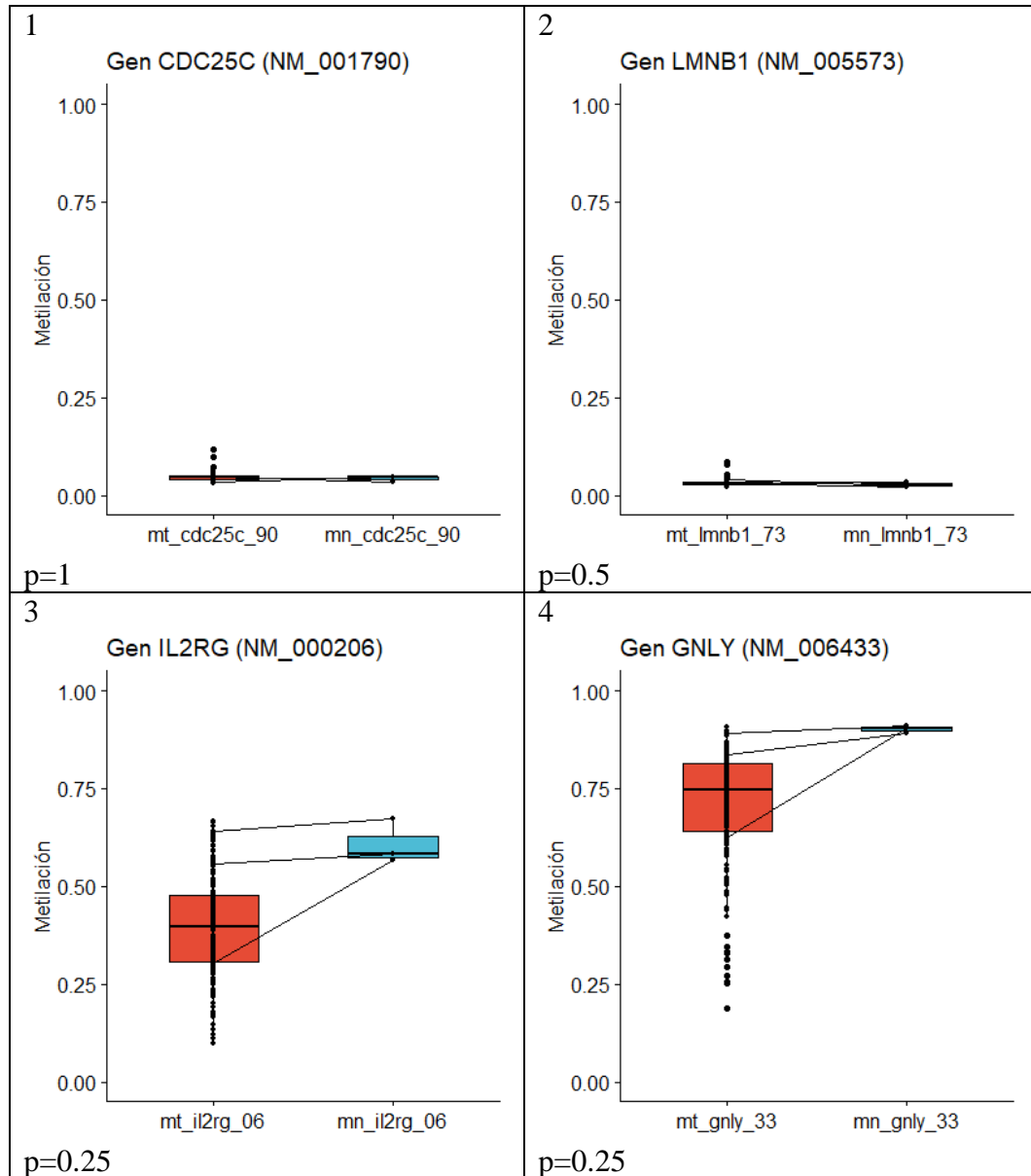
p=0.25

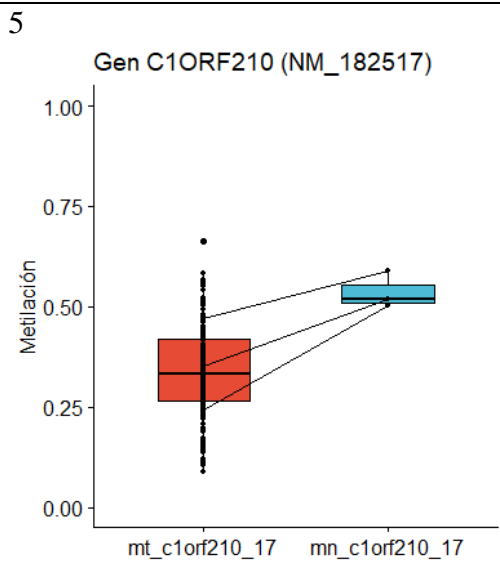


Anexo 4

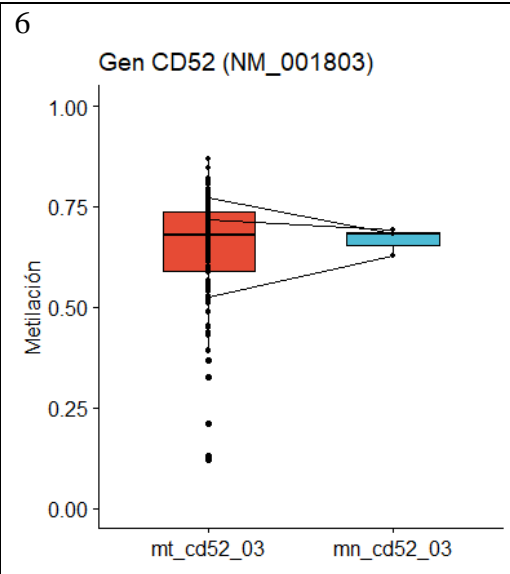
Transcritos de los genes sobre-expresados (19 transcritos)

Gráficos de cajas del nivel de metilación en las muestras tumorales y muestras sanas de cérvix para los 19 transcritos seleccionados correspondientes a los genes sob-reexpresados y valores de p de las pruebas de los rangos con signo de Wilcoxon. Se indica el nombre del gen, los dos últimos números de su respectivo código de identificación y un prefijo que indica si la muestra es tumoral o sana (mt: muestra tumoral, mn:muestra).

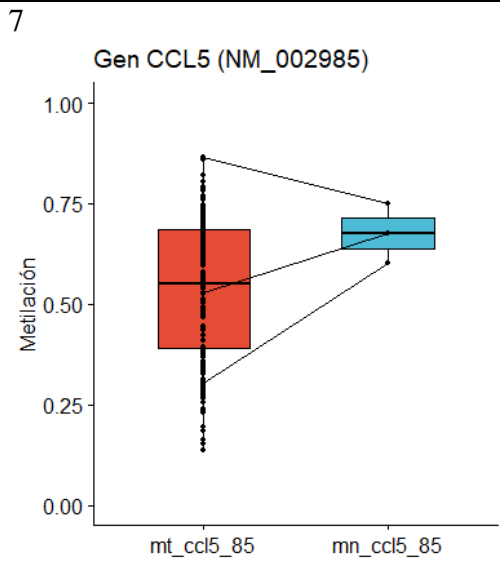




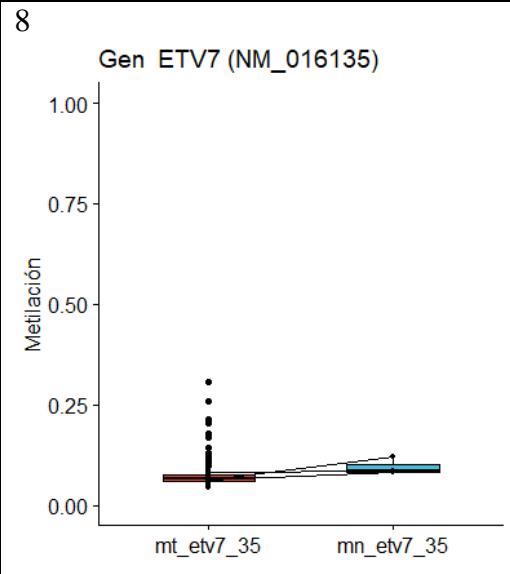
p=0.25



p=1



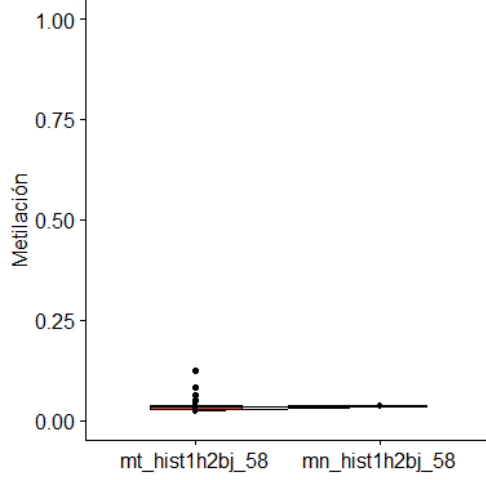
p=0.5



p=0.25

9

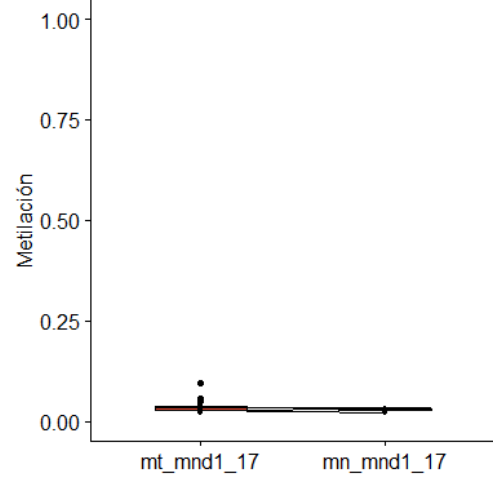
Gen HIST1H2BJ (NM_021058)



p=0.5

10

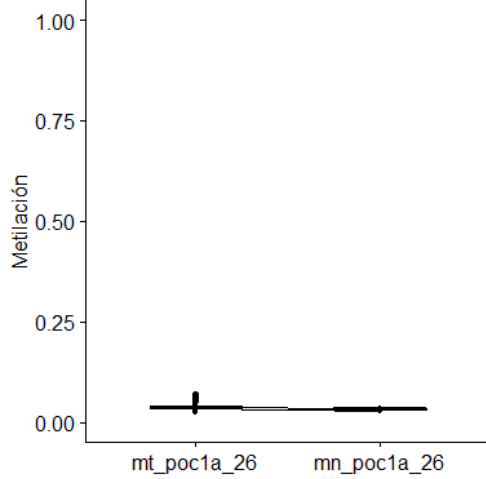
Gen MND1 (NM_032117)



p=0.5

11

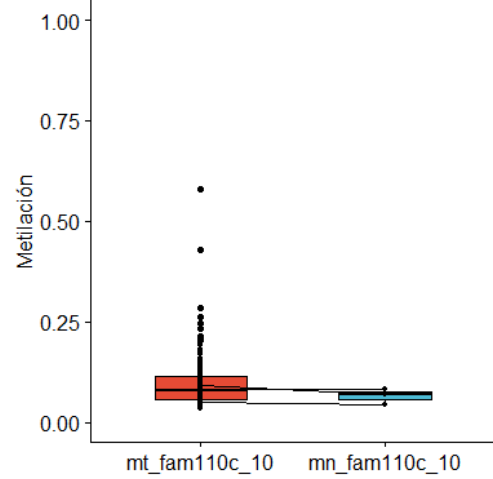
Gen POC1A (NM_015426)



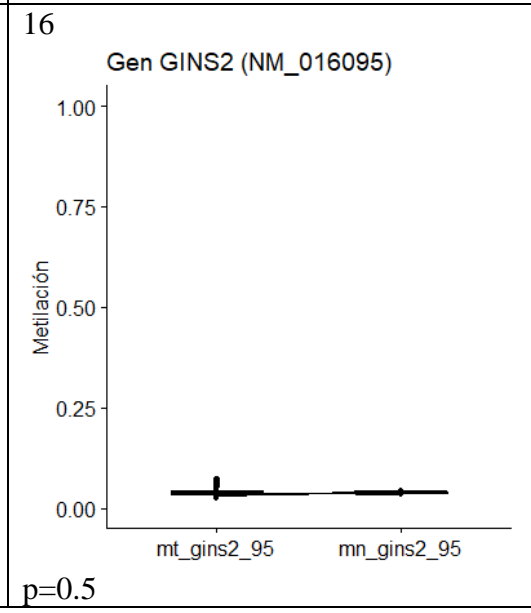
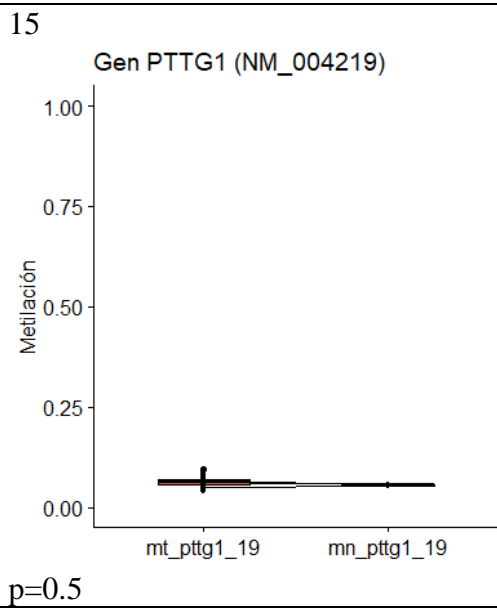
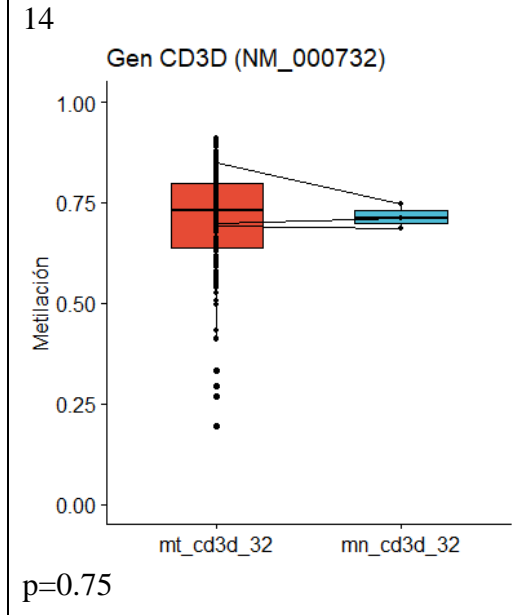
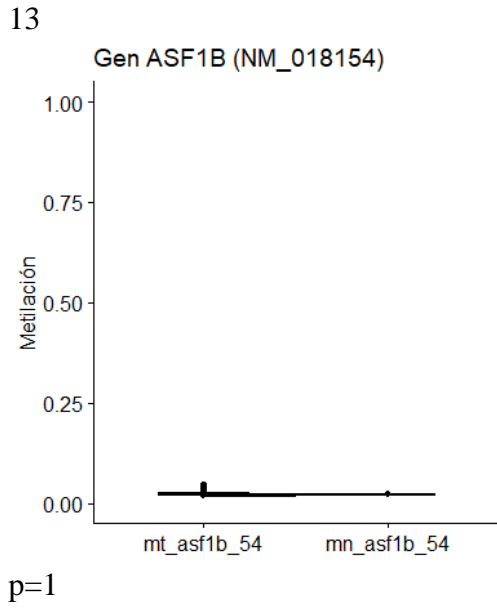
p=0.25

12

Gen FAM110C (NM_001077710)

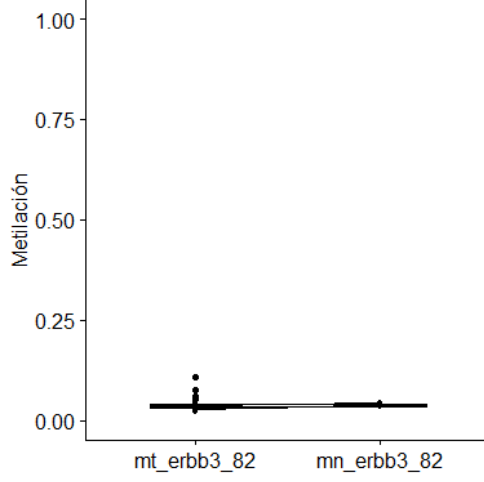


p=0.5



17

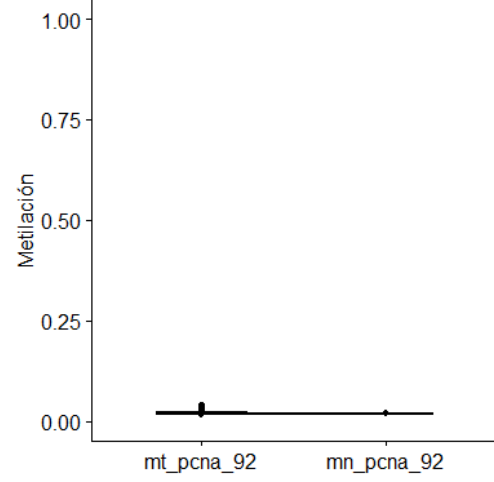
Gen ERBB3 (NM_001982)



p=0.5

18

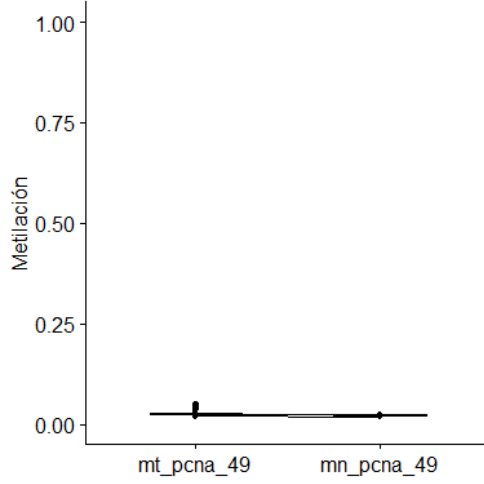
Gen PCNA (NM_002592)



p=0.75

19

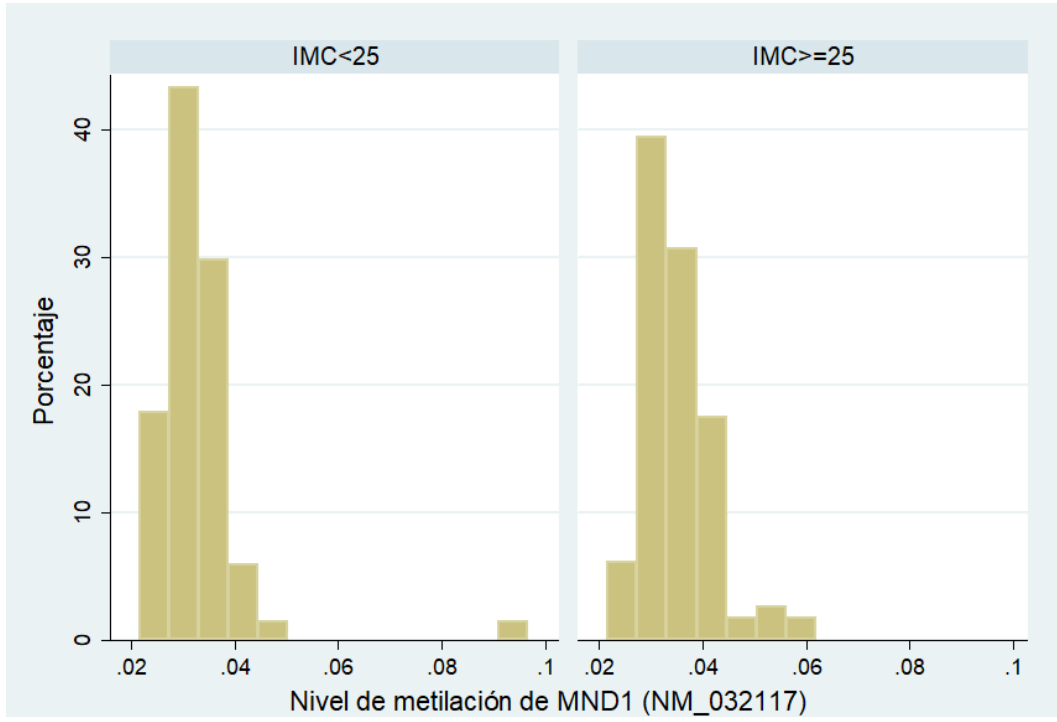
Gen PCNA (NM_182649)



p=0.5

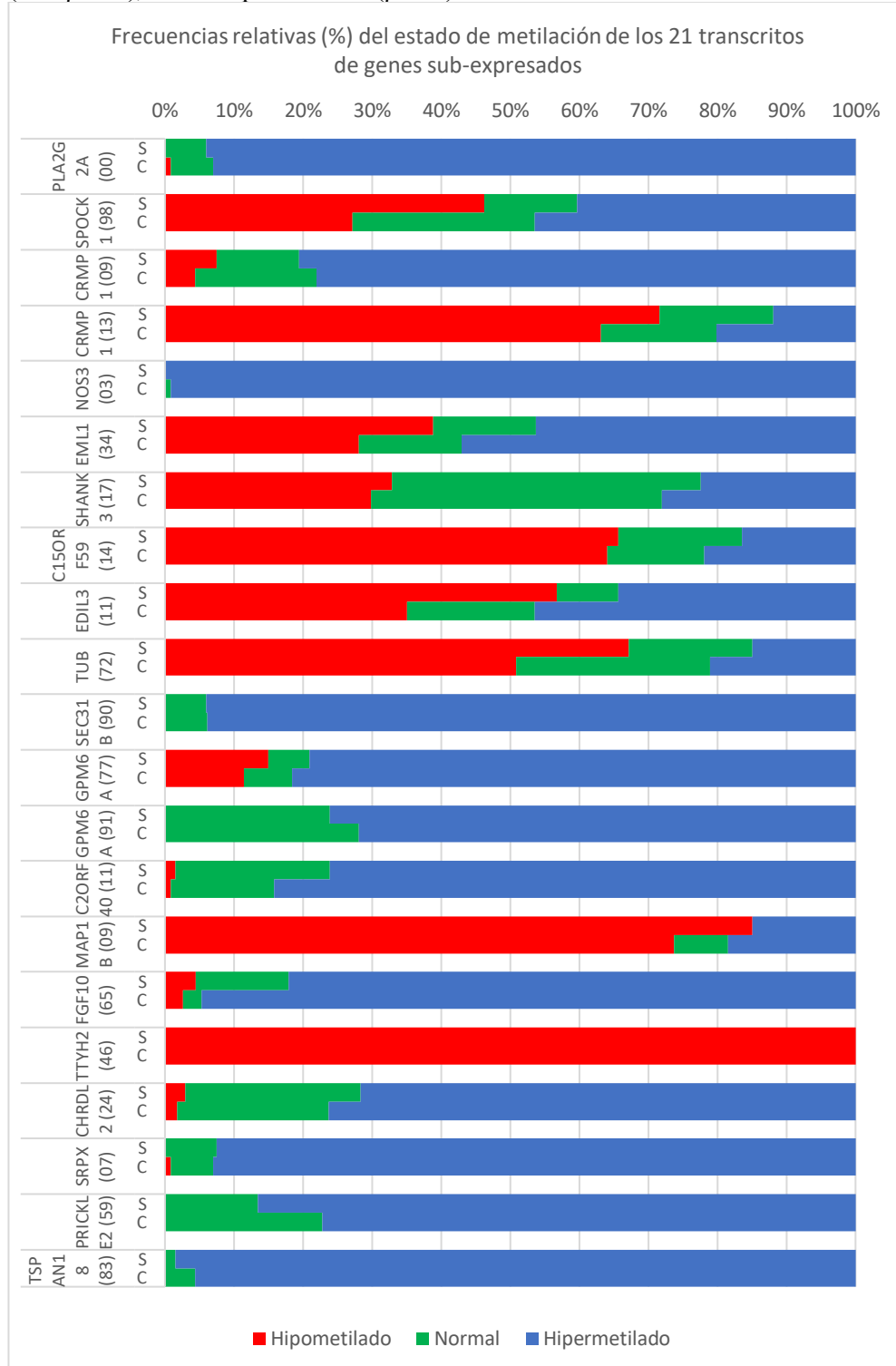
Anexo 5

Histograma del nivel de metilación del gen *MND1* (NM_032117) en función del estado de exceso de peso. IMC<25: Grupo sin exceso de peso, IMC≥25: Grupo con exceso de peso.



Anexo 6

Gráficos de barras del estado de metilación de los transcritos de los genes sub-expresados. En cada barra, se indica el nombre del gen y los dos últimos números de su respectivo código de identificación. Se indica el % de casos para las categorías del estado de metilación según el grupo de estudio. S: grupo sin exceso de peso ($IMC < 25$), C: grupo con exceso de peso ($IMC \geq 25$). Rojo: Hipometilado ($\beta \leq 0.1$), Verde: Metilación normal ($0.1 < \beta < 0.3$), Azul: Hipermetilado ($\beta \geq 0.3$).



Anexo 7

Gráficos de barras del estado de metilación de los transcritos de los genes sobre-expresados. En cada barra, se indica el nombre del gen y los dos últimos números de su respectivo código de identificación. Se indica el % de casos para las categorías del estado de metilación según el grupo de estudio. S: grupo sin exceso de peso ($IMC < 25$), C: grupo con exceso de peso ($IMC \geq 25$). Rojo: Hipometilado ($\beta \leq 0.1$), Verde: Metilación normal ($0.1 < \beta < 0.3$), Azul: Hipermetilado ($\beta \geq 0.3$).

