



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
IZTACALA**

**IDENTIFICACIÓN DEL PERFIL DE EXPRESIÓN
GLOBAL ASOCIADO A LA INFECCIÓN POR
DISTINTOS GENOTIPOS DEL VIRUS DE
PAPILOMA HUMANO (VPH) EN CÁNCER
CERVICAL**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGO

P R E S E N T A:

ANTONIO DANIEL MARTÍNEZ GUTIÉRREZ



**DIRECTOR DE TESIS:
Dr. Carlos Guadalupe Pérez Plasencia
Los Reyes Iztacala, Tlalnepantla, Edo. De México
2016**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dedicatoria

A mis padres, Antonio y Carina, pilares fundamentales en mi vida, por motivarme a ser siempre el mejor, y por su infinito apoyo incondicional.

A Sarahi, por estar siempre conmigo y apoyarme.

Agradecimientos

Al Dr. Carlos Pérez Plasencia, por permitirme ser parte de su grupo de investigación.

Al Dr. Abraham Pedroza por el apoyo recibido durante la escritura de esta tesis.

“All we have to decide is what to do with the time that is given us”

— **J.R.R. Tolkien, *The Fellowship of the Ring***

Índice

Resumen	1
Introducción	2
Cáncer Cérvico Uterino	8
Factores de riesgo	9
Diagnóstico del cáncer cérvico uterino	10
Virus del papiloma humano	11
Características principales de los papilomavirus	12
Proteínas tempranas	13
Papilomavirus y cáncer cérvico uterino	14
Los Microarreglos	17
Los microarreglos y el virus del papiloma humano	19
Justificación	21
Objetivos	22
Objetivo general	22
Objetivos particulares	22
Hipótesis	22
Metodología	23
Obtención de los datos	23
Determinación de los genotipos más frecuentes	24
Análisis de expresión diferencial	25
Obtención de los <i>heatmaps</i>	27
Análisis de enriquecimiento de términos	29
Análisis de Supervivencia	30
Resultados	31
Discusión	55
Conclusiones	60
Anexos	61
Anexo A	61
Anexo B	63
Anexo C	64
Anexo D	68
Anexo E	71
Anexo F	89
Bibliografía	94

Índice de figuras

Figura 1.- Los hallmarks del cáncer.	3
Figura 2.- Los hallmarks del cáncer emergentes.	6
Figura 3.- Incidencia y mortalidad del cáncer en México.	9
Figura 5.- Ejemplo de la imagen obtenida de un microarreglo.	11
Figura 6.- Organización del genoma del grupo de alfa papilomavirus.	13
Figura 7. ExpressionSet obtenido de la plataforma Gene Expression Omnibus.	25
Figura 8.- Genes diferencialmente expresados entre los tumores infectados con el genotipo 16 en comparación con tumores con el genotipo 18.	33
Figura 9.- Genes diferencialmente expresados entre los tumores infectados con alguna coinfección en comparación con tumores con el genotipo 18.	34
Figura 10.- Genes diferencialmente expresados entre los tumores infectados con el genotipo 45 en comparación con tumores con el genotipo 18.	35
Figura 11. Heatmap de la expresión diferencial entre los tumores que presentaron una infección simple con el genotipo 16 y tumores que presentaron el genotipo 18.	42
Figura 12. Heatmap de la expresión diferencial entre los tumores que presentaron una infección con el genotipo 18 y los tumores que presentaron una coinfección.	43
Figura 13. Heatmap de la expresión diferencial entre los tumores que presentaron una infección con el genotipo 18 y el genotipo 45.	44
Figura 14. Número de genes diferencialmente expresados compartidos entre cada comparación de los genotipos.	45
Figura 15. Número de genes diferencialmente expresados en cada comparación.	47
Figura 16. Análisis funcional de las vías de señalización alteradas relacionadas con la progresión y desarrollo de cáncer.	48
Figura 17. Análisis funcional de los términos biológicos GO enriquecidos de los genes diferencialmente expresados.	51
Figura 18. Análisis Kaplan-Meier de supervivencia.	53

Índice de Tablas

Tabla 1.- Clasificación de los estadios del cáncer cérvico uterino	10
Tabla 2.- Características clínicas, patológicas y tipificación de las pacientes	31
Tabla 3.- Genotipos virales presentes en las coinfecciones	32
Tabla 4.- Genes subexpresados en el Genotipo 16 vs Genotipo 18	36
Tabla 5.- Genes sobreexpresados en el Genotipo 16 vs Genotipo 18	37
Tabla 6.- Genes subexpresados en pacientes con coinfección vs Genotipo 18	38
Tabla 7.- Genes sobreexpresados en pacientes con coinfección vs Genotipo 18	39
Tabla 8.- Genes subexpresados en pacientes infectados con el Genotipo 45 vs Genotipo 18	40
Tabla 9.- Genes sobreexpresados en pacientes infectados con el Genotipo 45 vs Genotipo 18	41

Resumen

El cáncer es una de las principales causas de muerte a nivel mundial, siendo el cáncer cérvico uterino la segunda causa de muerte en mujeres por cáncer en México. Uno de los principales factores que aumentan el riesgo de desarrollar cáncer cérvico uterino es la presencia de una infección por un genotipo de virus de papiloma humano (VPH) de alto riesgo.

En este trabajo se utilizaron 63 tumores de pacientes del Instituto Nacional de Cancerología (INCAN) diagnosticadas con cáncer cérvico uterino en un microarreglo de la marca NimbleGen 12x135k, posteriormente se separaron en los genotipos virales de VPH más frecuentes y se realizó un análisis de expresión diferencial utilizando el paquete Linear Models for Microarray data (Limma) del proyecto Bioconductor, basado en el lenguaje estadístico R. Posteriormente se hizo el análisis de las principales vías de señalización alteradas con el paquete de Bioconductor *gage* utilizando los genes diferencialmente expresados. Finalmente se realizó un análisis de supervivencia utilizando el estadístico Kaplan-Meier, utilizando el tiempo libre de enfermedad y la supervivencia global.

En total se encontraron 76 genes diferencialmente expresados entre las pacientes infectadas con el genotipo 16 y el genotipo 18, entre las pacientes con el genotipo 18 y las que presentaban una coinfección se encontraron 237 genes diferencialmente expresados, mientras que entre las pacientes infectadas con el genotipo 45 y el genotipo 18 se encontraron 679 genes diferencialmente expresados. Las principales vías de señalización alteradas fueron vías relacionadas con el metabolismo, proliferación celular, apoptosis, adhesión, metástasis, diferenciación celular y relacionadas con la respuesta del sistema inmune. No se encontró una relación significativa entre el genotipo de VPH y la supervivencia de las pacientes.

Se concluye que el perfil de expresión de las pacientes se ve modificado por el genotipo viral que posean las pacientes, se presume que esta diferencia está dada por las diferencias en la secuencia de nucleótidos de las proteínas de cada genotipo.

Introducción

El Cáncer

El cáncer es un problema de salud a nivel global, y es considerada una de las primeras causas de muerte en todo el mundo, tan solo en el 2012 se registraron cerca de 8.2 millones de muertes a causa de esta enfermedad (Ferlay et al., 2015). El cáncer es el resultado de un conjunto de enfermedades, producto de diversas mutaciones en las células del organismo que las llevarán a proliferar sin control, lo que tendrá como consecuencia la invasión de tejidos que normalmente están destinados a otras células, provocando alteraciones a nivel sistémico en órganos y tejidos que finalmente llevará a la muerte.

De acuerdo a la teoría celular, todas las células provienen de una célula, a través de una serie de eventos en los cuales una célula progenitora aumentará su masa celular, duplicará su genoma y se dividirá en dos, heredando una copia de su genoma a una célula hija, este conjunto de eventos es conocido como ciclo celular. Debido a su gran importancia, estos eventos se encuentran altamente regulados por diversas proteínas que evitan una proliferación sin control de todas las células. La desregulación de los mecanismos de control del ciclo celular es un evento clave en la aparición y desarrollo del cáncer.

Una vez que una célula ha perdido el control en los mecanismos que regulan la proliferación celular, esta comenzará a adquirir ciertas características que posteriormente le permitirán transformarse en un tumor. Estas características fueron descritas como los *hallmarks* del cáncer en el 2000 por Hanahan y Weinberg (Hanahan, Weinberg, & Francisco, 2000), las cuales se observan en la Figura 1 y se describen brevemente a continuación.

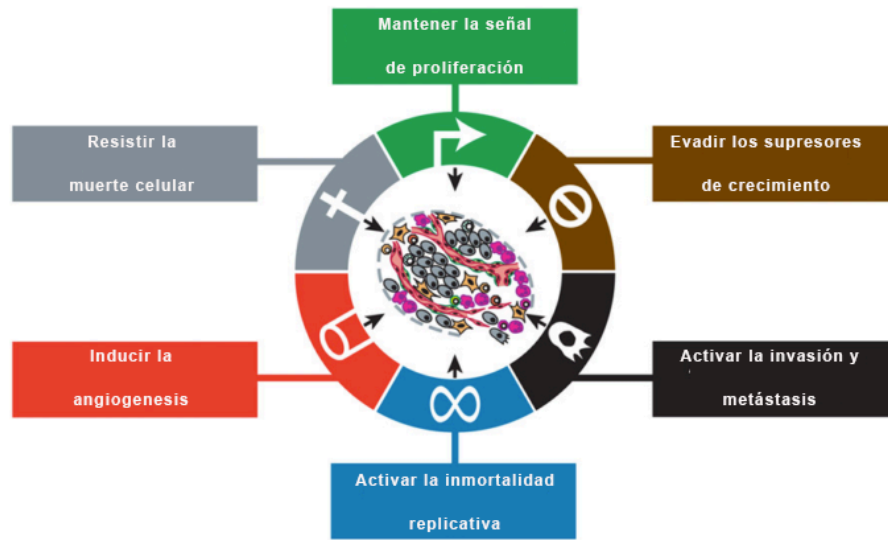


Figura 1.- Los hallmarks del cáncer.

Descritos por primera vez en el año 2000, muestran las principales características que adquieren las células normales en el transcurso de su transformación a células malignas (Editado de Hanahan & Weinberg, 2011).

Mantener la señal de proliferación: En general, el tejido sano mantiene un control muy estricto de los factores de crecimiento que le permiten regular la progresión del ciclo celular, sin embargo, las células tumorales evaden este control de diferentes maneras; ya sea mediante la producción autócrina de sus propios ligandos o mediante el envío de señales a células sanas con el fin de que estas últimas suplan a las células tumorales de los factores de crecimiento (Hanahan & Weinberg, 2011).

Evadir los supresores de crecimiento: Los genes supresores de tumores codifican proteínas que tienen como principal función controlar y regular la proliferación celular mediante la regulación de múltiples procesos que previenen la proliferación aberrante, ejemplos de genes supresores de tumores son p53 y Rb, los cuales tienen como función inhibir la proliferación celular mediante la activación de vías de señalización como la apoptosis (Fridman & Lowe, 2003). Durante el

proceso de carcinogénesis las células tumorales pierden la función de estos genes, proliferando sin control.

Evasión de la Apoptosis: La apoptosis es una forma de muerte celular programada, durante este proceso el citoesqueleto, la cromatina nuclear y la envoltura nuclear se condensan y se fragmentan. Posteriormente la célula se rompe en fragmentos llamados cuerpos apoptóticos, que de manera subsecuente serán fagocitados por células adyacentes o por macrófagos. Este proceso es activado cuando la célula ha sufrido un daño irreparable en su ADN o cuando ha sido infectada por un microorganismo, la ventaja de este proceso es que la célula es eliminada antes de que se convierta en una amenaza para el organismo (Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, 2015).

Al igual que otros procesos vitales de la célula, la apoptosis se encuentra regulada por diversas proteínas que funcionan como estimulantes o inhibidores de la apoptosis. Un ejemplo es la familia de proteínas Bcl2. Esta familia posee proteínas proapoptóticas, como *Bax* y *Bak*, y antiapoptóticas como *Bcl-2*. Se sabe que estas proteínas se unen entre si formando heterodímeros, inhibiéndose mutuamente. El equilibrio en la expresión de estas dos clases de proteínas determina si una célula activa o no la vía de señalización de la apoptosis (Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, 2015). La apoptosis también puede ser inducida por la acción de supresores de tumores como p53, el cual es activado al detectar un daño en el ADN que no puede ser reparado.

Durante el proceso de carcinogénesis las células tumorales evadirán la apoptosis mediante múltiples mecanismos; como la pérdida de genes supresores de tumores, la sobreexpresión de reguladores antiapoptóticos y supervivencia como *Bcl-2* o mediante la subexpresión de factores proapoptóticos como *Bax* (Hanahan & Weinberg, 2011).

Activar la inmortalidad replicativa: Para que las células tumorales sean capaces de formar tumores macroscópicos es necesario que evadan la principal barrera

que regula la proliferación celular; la senescencia. La senescencia es un mecanismo que regula cantidad de veces que una célula es capaz de dividirse, esto se logra mediante el acortamiento progresivo de los telómeros, los cuales son secuencias de nucleótidos repetidos ubicadas en los extremos del cromosoma, cuya función es que protegerlo del deterioro . Se cree que las telomerasas -las enzimas encargadas de replicar el ADN en los extremos de los cromosomas y que permite el alargamiento de los telómeros- se encuentran ausentes en células normales, sin embargo se encuentran sobreexpresadas de manera importante en células inmortalizadas, incluidas células tumorales, de esta manera estas últimas logran evadir la senescencia celular y adquieren la habilidad de dividirse de manera ilimitada (Cavallo, De Giovanni, Nanni, Forni, & Lollini, 2011).

Inducir la angiogénesis: La angiogénesis es el proceso mediante el cual se estimula la formación de nuevos vasos sanguíneos a partir de los vasos preexistentes, este proceso ocurre principalmente durante la embriogénesis, sin embargo se encuentra inactivado durante la etapa adulta del organismo, de esta manera se crea una barrera que dificulta el desarrollo de neoplasias (Bergers & Benjamin, 2003). Las células tumorales también requieren de los mismos nutrientes que las células normales, por lo que durante la progresión tumoral se activa el proceso de angiogénesis; lo que provoca la producción continua de vasos sanguíneos hacia las células tumorales, con la finalidad de proveer de los nutrientes necesarios para el mantenimiento y el crecimiento del tumor (Hanahan et al., 2000).

Adquirir la capacidad de hacer metástasis: La metástasis es la habilidad del tumor de escapar del tejido de origen a través del torrente sanguíneo con el objetivo de colonizar y crecer en órganos distantes (Bendas & Borsig, 2012). Se cree que uno de los procesos involucrados con la adquisición de esta habilidad es un proceso relacionado con la regulación del desarrollo durante la embriogénesis con la conocido como la transición epitelio-mesénquima (EMT), el cual permitirá al

tumor adquirir habilidades que le facilitaran invadir otros órganos (Hanahan & Weinberg, 2011).

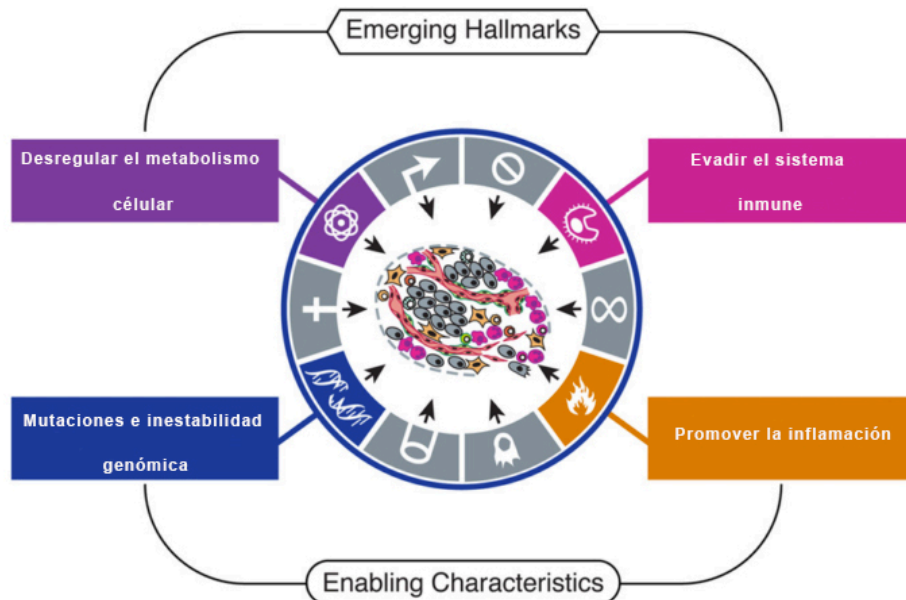


Figura 2.- Los *hallmarks* del cáncer emergentes.

En el 2011, se publicó una revisión de los *hallmarks* del cáncer, en la cual se añadirían dos nuevos *hallmarks* emergentes; desregular el metabolismo celular y evadir el sistema inmune, y dos características indispensables: mutaciones e inestabilidad genómica y promover la inflamación (Editado de Hanahan & Weinberg, 2011).

En el 2011 se publicó una revisión de los *hallmarks* del cáncer, en la cual se añadían dos *hallmarks* emergentes y dos características indispensables (ver Figura 2).

Mutaciones e inestabilidad genómica: Para que una célula tumoral adquiriera un *hallmark* es necesario que presente un conjunto de alteraciones en su genoma, por lo que se cree que la progresión tumoral es un proceso múltiple que involucra la adquisición de genotipos mutantes que le confieran a las células neoplásicas alguna ventaja que les permita sobrevivir y crecer (Hanahan & Weinberg, 2011).

Promover la inflamación: Se ha visto que la inflamación producida por las células tumorales contribuye a la adquisición de *hallmarks* mediante la producción y suministro de moléculas, como las especies reactivas de oxígeno y factores de crecimiento al microambiente tumoral (Hanahan & Weinberg, 2011). También se ha demostrado que las células tumorales se aprovechan de los procesos que ocurren durante la inflamación para invadir y provocar la metástasis (Rüegg, 2006).

Desregular el metabolismo celular: Para que una célula neoplásica que ha sido capaz de inmortalizar su replicación y proliferación pueda continuar haciéndolo requiere de modificar su metabolismo. Esto lo logra a través del fenómeno conocido como efecto Warburg, nombrado en honor de Otto Warburg, quien fue el primero en observar este proceso. En condiciones aeróbicas las células normales oxidan la glucosa para formar piruvato, el cual es posteriormente trasladado a la mitocondria donde se obtendrá ATP a través de la fosforilación oxidativa. Solamente cuando las células normales se encuentran en condiciones anaerobias producirán ATP a través de la reducción de piruvato a lactato (Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, 2015). Se sabe que las células tumorales, incluso en condiciones aeróbicas modifican su metabolismo glucolítico con el fin de aumentar considerablemente sus niveles de lactato, restringiéndose solamente a la glicolisis para la generación de ATP, este fenómeno es conocido como efecto Warburg. Como consecuencia de esta alteración, se producirá una mayor cantidad de intermediarios de la glucolisis, los cuales son usados principalmente para generar aminoácidos y nucleótidos, moléculas que son indispensables para ensamblar nuevas células (Hsieh, Walton, Altman, Stine, & Dang, 2015). El efecto Warburg también ha sido relacionado con la activación de oncogenes como RAS, MYC y con la presencia de mutaciones en genes supresores de tumores (Hanahan & Weinberg, 2011).

Evadir el sistema inmune: Los mecanismos de vigilancia en el sistema inmune reconocen e impiden el desarrollo de la mayoría de las células tumorales, sin embargo, no es totalmente eficiente, por lo que es necesario para las células tumorales desarrollar mecanismos para evadir el sistema inmune. Uno de estos mecanismos consiste en alterar las moléculas del complejo de histocompatibilidad (MHC), que impedirá el reconocimiento de la célula por parte de los linfocitos T (Cavallo et al., 2011). Además se ha visto que las células tumorales son capaces de paralizar las células Natural Killers (NK) mediante la secreción de TGF-B y otros factores supresores del sistema inmune (Hanahan & Weinberg, 2011).

Cáncer Cérvico Uterino

El cáncer cérvico uterino es el término que se le da a las neoplasias malignas que se originan a partir de células del cérvix y el útero. A pesar de los programas de detección temprana, el cáncer cérvico uterino continua siendo una de las primeras causas de muerte por neoplasias en mujeres a nivel mundial que afecta primordialmente a los países en vías de desarrollo (Bosch, F. Xavier, Silvia, 2003; Jemal et al., 2011). En nuestro país el cáncer cérvico uterino es la segunda causa de muerte por neoplasias en mujeres mayores de 25 años (ver Figura 3), anualmente se estima una ocurrencia de 20,444 casos en mujeres, con una incidencia de 35.4 casos por 100,000 mujeres (Ferlay et al., 2015).

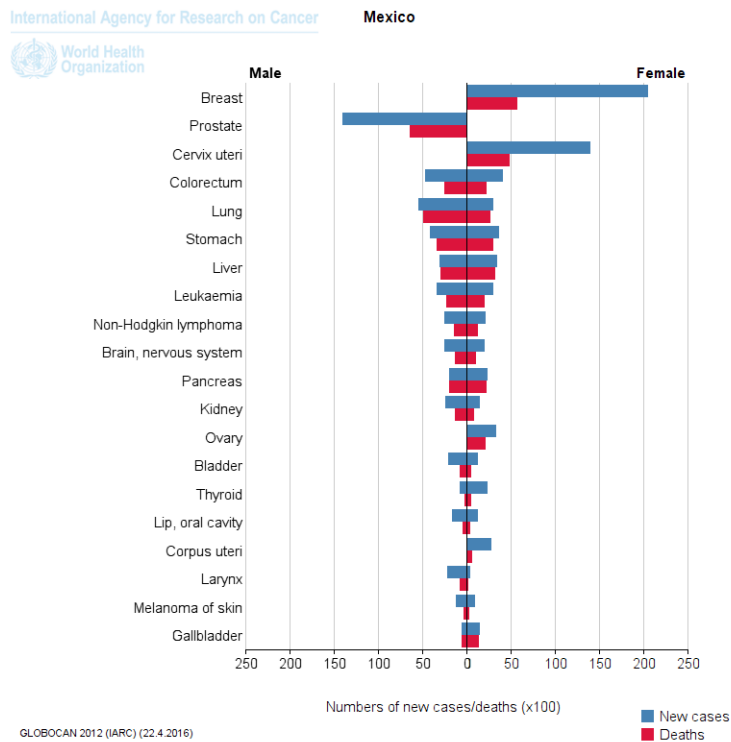


Figura 3.- Incidencia y mortalidad del cáncer en México.

Se observa que, en nuestro país, el cáncer cérvico-uterino es la segunda causa de mortalidad y de incidencia en mujeres, solamente detrás del cáncer de mama (GLOBOCAN, 2012).

Factores de riesgo

Se sabe que la infección por algún genotipo de VPH de alto riesgo es la principal causa de desarrollo de cáncer cérvico uterino (Muñoz et al., 2003). Otros factores que se han relacionados con el desarrollo de cáncer son un alto número de parejas sexuales (más de cuatro), inicio de la actividad sexual antes de los 16 años, entre otros. Un factor de riesgo independiente para desarrollar cáncer cérvico uterino es fumar tabaco, ya que se han encontrado evidencias de la presencia de carcinógenos relacionados con el tabaco en el epitelio cervical de fumadores, los cuales se sabe que pueden unirse al ADN provocando mutaciones,

umentando la probabilidad de desarrollar una neoplasia maligna (Waggoner et al., 2003).

Diagnóstico del cáncer cérvico uterino

El diagnóstico de cáncer cérvico uterino se da a través de un examen de Papanicolaou ó mediante una colposcopia. Una vez que se ha confirmado el diagnóstico de cáncer en una paciente se procede a clasificarla de acuerdo a los estadios de la International Federation of Gynecologists and Obstetricians (FIGO) (Ver Tabla 1) (Pecorelli, Zigliani, & Odicino, 2009). El estadio es determinado de manera clínica, y está basado en principalmente en el tamaño del tumor en el cérvix o su extensión en la pelvis (Waggoner et al., 2003), este paso es de gran importancia, ya que decide el tipo de tratamiento y las opciones quirúrgicas del paciente.

Tabla 1.- Clasificación de los estadios de cáncer cérvico uterino. Esta clasificación se basa primordialmente en el tamaño y la localización de las lesiones malignas.

ESTADIO	DESCRIPCIÓN
ESTADIO 0	Carcinoma in situ, carcinoma intraepitelial
ESTADIO I	Carcinoma infiltrante estrictamente confinado al cervix
IA	Carcinoma infiltrante diagnosticado mediante microscopio
IA1	Infiltracion del estroma < 3 mm en profundidad, con extension de superficie < 7 mm
IA2	Infiltracion del estroma > 3mm, pero menor de 5 mm en profundidad con extension en superficie de 7 mm
IB	Lesion visible clinicamente limitada al cervix con mayor dimension que el estadio IA2
IB1	Lesion visible clinicamente con tamaño menor a 4 cm en su diametro maximo
IB2	Lesion visible clinicamente con tamaño mayor a 4 cm en su diametro maximo
ESTADIO II	El tumor se extiende mas alla del utero sin afectar la pared pélvica o al tercio inferior de la vagina
IIA	No hay invasión de los parametrios
IIA1	Lesion visible clinicamente con tamaño menor a 4 cm en su diametro maximo
IIA2	Lesion visible clinicamente con tamaño mayor a 4 cm en su diametro maximo
IIB	Se observa invasión de los parametrios
ESTADIO III	El tumor se extiende a la pared pélvica, afectando al tercio inferior de la vagina provocando hidronefrosis o anulación de la función renal
IIIA	El tumor afecta el tercio inferior de la vagina sin extensión a la pared pélvica
IIIB	El tumor se extiende a la pared pélvica, provocando hidronefrosis o anulación de la función renal
ESTADIO IV	El tumor se extiende fuera de la pelvis o clinicamente se observa invasion de la vejiga o recto
IVA	El tumor infiltra la mucosa de la vejiga, el recto ó se extiende mas allá de la pelvis
IVB	Metástasis a distancia

La detección temprana y el apropiado tratamiento de las lesiones precancerosas juegan un papel importante en la prevención de la progresión a cáncer. Se consideran como estadios localmente avanzados a partir del estadio IB2; cuando el tumor presenta una lesión clínica visible de más de 4 cm en su diámetro máximo. Algunos autores, consideran que la supervivencia de las pacientes diagnosticadas con cáncer cérvico uterino puede ser pronosticada usando el estadio clínico. En las etapas tempranas de la enfermedad como el estadio IA se tienen un 100% de probabilidad de supervivencia a 5 años, mientras que esta probabilidad disminuye a 70-85% en el estadio IB1 y IIA. La supervivencia en los estadios localmente avanzados es muy baja, y está influenciada por factores como la edad del paciente y la presencia de otras enfermedades. En estos estadios la probabilidad de supervivencia a 5 años es de 30% al 50% para el estadio III, y de 5% a 15% para el estadio IV (Waggoner et al., 2003).

Virus del papiloma humano

Los papilomavirus son un grupo de virus de ADN circular de doble hebra con aproximadamente 7,900 pares de bases que inducen la aparición de lesiones en la piel; las cuales son denominadas verrugas o condilomas. Desde el 2002 se sabe que entre cerca del 99.9 % de los casos de cáncer cérvico uterino están relacionados con la infección por al menos un genotipo de VPH, sin embargo la infección no siempre es necesaria para que exista una progresión a cáncer

(Bosch, Lorincz, Muñoz, Meijer, & Shah, 2002; Sobota, Ramogola-Masire, Williams, & Zetola, 2014).

La mayoría de los genotipos de VPH producen lesiones benignas en la piel y no están relacionados con el desarrollo de cáncer; sin embargo se sabe que ciertos genotipos están relacionados con la progresión de las lesiones a cáncer cérvico uterino (Fields, Knipe, & Howley, 2013), estos genotipos son denominados como de alto riesgo.

Características principales de los papilomavirus

El genoma de los papilomavirus está dividido en tres regiones principales, la región larga de control (LCR), la segunda región corresponde a las proteínas tempranas E1, E2, E4, E5, E6, E7 y E8, mientras que la tercera región contiene a las proteínas tardías L1 y L2, las cuales son las proteínas de la cápside relacionadas con la entrada del virus a la célula. Las actuales vacunas profilácticas contra el VPH reconocen estas proteínas (Figura 4.) (Buck, Day, & Trus, 2013).

La región LCR contiene el inicio de la replicación del ADN viral, a través de un promotor conocido como p97 en el genotipo 16, y p105 en el genotipo 18, los cuales tienen la principal función regular la transcripción de las proteínas E6 y E7, además se sabe que esta región contienen un alto nivel de variación entre genotipos (Silva et al., 2013).

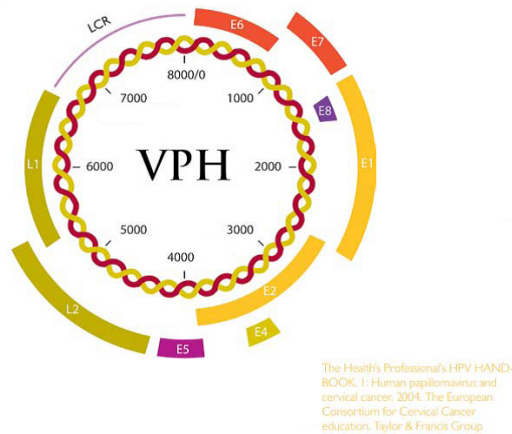


Figura 6.- Organización del genoma del grupo de alfa papilomavirus.

El genoma de los papilomavirus está compuesto por genes tempranos (E1-E6) y genes tardíos (L1-L2) (Consortium for Cervical Cancer education. Taylor & Francis Group. 2004).

Proteínas tempranas

E1: Este gen codifica a una helicasa de ADN que es necesaria para la replicación y amplificación del genoma viral. Al igual que la proteína L1, esta altamente conservada.

E2: Este gen codifica a una proteína de 45kDa, sus principales funciones son actuar como un represor de los genes E6 y E7, además de regular la transcripción y replicación del ADN viral.

E4: Codifica a una proteína de 260 pares de bases que se encuentra dentro del marco de lectura de E2, su expresión está asociada a la alteración de los filamentos intermedios de queratina, que conlleva a facilitar la liberación de los viriones de la célula.

E5: Proteína cuya principal función es inducir la transformación de las células y promover a la progresión del tumor mediante el mantenimiento de la mitosis y la

división celular en las células supra basales del epitelio (Müller, Prescott, Wasson, & Macdonald, 2015).

E6: Es una proteína pequeña de aproximadamente 150 aminoácidos, entre sus principales funciones canónicas esta ubiquitinar y degradar al supresor de tumores p53, activar la función de la telomerasa con el fin de prevenir que la célula entre en un estado de senescencia y de prevenir el daño al DNA. El resultado de su actividad es prevenir y resistir la apoptosis.

E7: Proteína que posee alrededor de 100 aminoácidos y que es esencial durante el ciclo viral y la transformación carcinogénica. Su principal función es unirse a la proteína supresora de tumores pRB con el fin de marcarla para su degradación, mientras que desregula el ciclo celular, alterando la transición de la fase G1/S mediante la unión al factor de transcripción E2F y a través de la anulación de los inhibidores de las cinasas dependientes de ciclinas (CKIs) (McLaughlin-Drubin & Münger, 2009).

E8: Es una proteína expresada solamente en los genotipo 16, 18 y 31, la cual se une al origen de la replicación (ORF) con el fin de reprimir la transcripción de E1 y E2 (Straub, Fertey, Dreer, Iftner, & Stubenrauch, 2015).

Papilomavirus y cáncer cérvico uterino

Uno de los factores que definen la capacidad de los virus de desarrollar cáncer en el hospedero es su posición filogenética. Se ha establecido la filogenia de estos virus basándose en la distancia genética entre estos, donde se considera al género como el taxón más amplio y con mayor variabilidad. Los géneros están compuestos de especies, y estas se subdividen en tipos virales. Hasta ahora solo el género alfa de los virus del papiloma humano se ha relacionado con el

desarrollo de cáncer cervical. La clasificación de los papilomavirus está basada en la similitud en la secuencia L1, se considera que son diferentes tipos cuando la similitud entre dos virus es entre 69% y 86%, mientras que se les considera subtipos si cuentan con más del 87% de similitud (De Villiers et al., 2004).

Los virus de bajo riesgo solamente son capaces de provocar una proliferación celular benigna, en esta categoría encuentran genotipos como el 6 y el 11. Los tipos virales de alto riesgo corresponden a las especies $\alpha 5$, $\alpha 6$, $\alpha 7$, $\alpha 9$ y $\alpha 11$, de éstas, los tipos virales 16 y 18 son las responsables del 70% de los casos de cáncer cérvico uterino a nivel mundial (Sobota et al., 2014). Solamente los virus pertenecientes a la categoría de alto riesgo son capaces de inmortalizar los queratinocitos primarios humanos que infectan (Li & Coffino, 1996).

Una vez que el virus infecta a una célula epitelial provoca una infección persistente, integrando su genoma en la célula hospedera, esta persistencia en una de las principales causas de progresión a cáncer. Después de su integración, el virus iniciara su actividad oncogénica, mediante la expresión de los oncogenes E6 y E7, las cuales tienen como principal blanco dos proteínas supresoras de tumores. E6 se unirá a la proteína E6-AP, provocando la ubiquitinización y degradación de p53, mientras que la unión de E7 con la proteína del retinoblastoma (Wan et al., 2008). Estas interacciones tendrán como resultado la pérdida de la regulación del ciclo celular y de la apoptosis.

Se sabe que la principal diferencia entre los virus de alto riesgo y bajo riesgo es la afinidad y el efecto de sus oncoproteínas, en específico, la oncoproteína E7 presente en el grupo de alto riesgo posee una mayor afinidad por la proteína del retinoblastoma. En el caso de E6, la variante presente en ambos grupos es capaz de interactuar con p53, sin embargo, solamente la variante presente en los virus de alto riesgo tiene la capacidad de inducir su degradación (Li & Coffino, 1996). Además, varios estudios han reportado que en cáncer cérvico uterino, la infección por genotipos de bajo riesgo está relacionada con una alta probabilidad de tener

una buena prognosis, mientras que la infección por genotipos de alto riesgo se relaciona con un pronóstico reservado (Rosty et al., 2005).

La principal vía de transmisión del VPH es por contacto sexual, lo que facilita su propagación en las personas jóvenes y se favorece la infección con diferentes genotipos del virus. Se ha reportado que la presencia de dos o más tipos de VPH de la misma especie inducen una protección natural al desarrollo de lesiones de alto grado, un efecto similar al utilizar las vacunas convencionales, ya que por las similitudes de entre los genotipos usados en estas terapias se espera que provean de una protección parcial contra otros genotipos similares mediante la neutralización cruzada de anticuerpos (Sobota et al., 2014).

La coinfección con más de un genotipo de papilomavirus es común, especialmente en mujeres jóvenes. Dado que el modo de transmisión principal del virus es la vía sexual las mujeres que poseen más de un genotipo poseen una mayor probabilidad de acarrear genotipos adicionales (Chaturvedi et al., 2011).

Varios estudios han demostrado que la coinfección con múltiples genotipos de virus de papiloma humano se asocian con una mayor probabilidad de desarrollar una neoplasia en comparación con pacientes con una sola infección (Rousseau et al., 2003). Se ha demostrado que entre el 20% y el 40% de las mujeres VPH positivas poseen al menos dos genotipos virales, los cuales pudieron ser adquiridos de manera simultánea o sucesiva. Hasta ahora se desconoce el riesgo que conlleva la infección múltiple del VPH con la persistencia del mismo y con la progresión a cáncer (Schmitt et al., 2010).

Anteriormente, Chaturvedi y colaboradores en 2011 reportaron que las mujeres con múltiples infecciones tenían un mayor riesgo de desarrollar lesiones de alto grado en comparación con aquellas que solo presentaban una infección, y en 2013 Goldman y colaboradores proponen que ciertos genotipos de VPH actúan como cofactores para facilitar la infección por un genotipo de alto riesgo con

aquellas que solo presentaban una infección, sin embargo aún no se sabe si la presencia de múltiples infecciones distintas del riesgo del virus están relacionadas con un mayor riesgo de desarrollar cáncer cérvico uterino (Goldman et al., 2013).

Los Microarreglos

La tecnología de los microarreglos fue desarrollada a principios de la década de los 90's por el Dr. Patrick Brown (Schena, Shalon, Davis, & Brown, 1995). Los arreglos de DNA son una herramienta que permite la obtención masiva de datos de expresión de proteínas, mRNA y microRNAs en un solo experimento. Esta tecnología está basada en la propiedad de la complementariedad del ADN, donde las hebras simples de ADN se unen e hibridan con su hebra complementaria, formando una doble hebra. En los microarreglos, miles de secuencias específicas de ADN son sintetizadas de manera ordenada y unidas mediante enlaces covalentes o no covalentes en una superficie 2D, donde cada punto del arreglo corresponde a una secuencia. Esta tecnología permite adherir una gran cantidad de secuencias de ácidos nucleicos a una superficie con el objetivo de medir la concentración relativa de secuencias de DNA en una solución problema mediante hibridación.

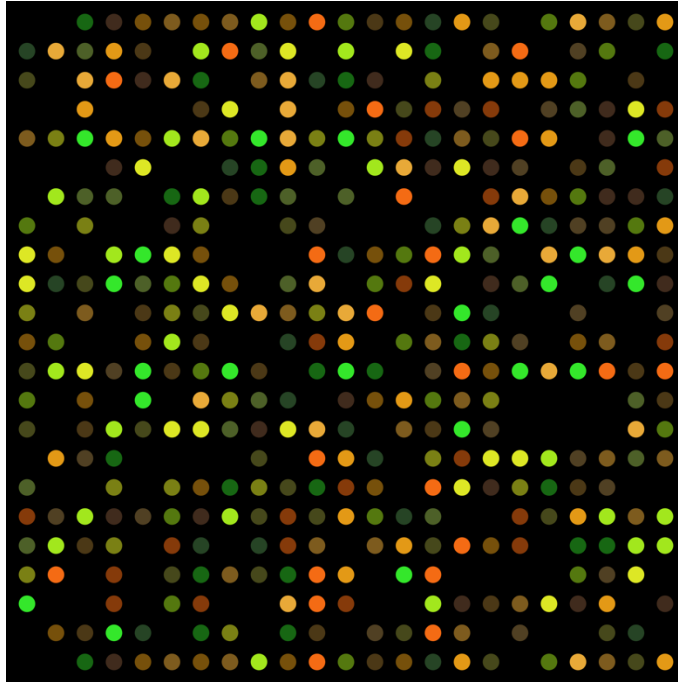


Figura 5.- Ejemplo de la imagen obtenida de un microarreglo.

Cada punto corresponde a una secuencia específica. La intensidad de la fluorescencia muestra la cantidad de cDNA que se unió a las secuencias del arreglo (Tomado de Wikimedia Commons).

La principal función de los microarreglos es medir los niveles de expresión génica, el cual se refiere al patrón de genes expresados por una célula en circunstancias específicas, a nivel de transcripción, lo que permite tener una visión global de las funciones celulares de esta en determinadas condiciones. En enfermedades multifactoriales como el cáncer el conocimiento de estos patrones cobra una especial relevancia ya que el desarrollo y progresión del tumor está dado por la mutación y expresión anormal de múltiples genes, por lo que el análisis de los patrones de expresión global pueden mejorar nuestro conocimiento acerca de los mecanismos de regulación, vías de señalización bioquímicas y funciones celulares de los tumores de una manera más amplia, con el fin de poder predecir el comportamiento biológico y las consecuencias clínicas de los diferentes perfiles de expresión presentes en las células tumorales.

El proceso de experimentación utilizando microarreglos comienza con la extracción de RNA de las células problema, a partir de este RNA se sintetiza

cDNA, el cual es marcado mediante colorantes fluorescentes como Cy3 y Cy5. Posteriormente el cDNA se hibridará con las secuencias adheridas al arreglo, acto seguido se utiliza un escáner el cual utilizará un láser que excitará los colorantes, los cuales emitirán fluorescencia que será medida. La intensidad de la fluorescencia en cada punto se considera como la unidad de expresión de cada gen (Ver Figura 5) (Bumgarner, 2013). El escáner producirá una imagen, la cual es analizada mediante un software especializado para transformar la intensidad de cada arreglo en un valor numérico.

Uno de los objetivos de los microarreglos es conocer si entre los grupos de muestras analizados existen genes que se encuentren sobreexpresados o subexpresados, lo que arrojará información acerca de las diferencias moleculares específicas que son asociadas a el efecto biológico analizado (Trevino, Falciani, & Barrera-Saldaña, 2007).

En el área de investigación del cáncer el uso de tecnologías de alto rendimiento como los microarreglos tiene una gran importancia; ya que es de gran valor entender las anomalías génicas como la alteración de las secuencias y el nivel de expresión de una gran cantidad de genes que lleven al inicio y posterior progresión del cáncer. Estas anomalías génicas, que pueden ser heredadas o adquiridas, le darán a las células tumorales la habilidad de desarrollar las características esenciales que les permitirán progresar y/o llevar a la muerte de un paciente con cierto perfil de expresión.

Los microarreglos y el virus del papiloma humano

Diversos estudios en los que se han utilizado microarreglos han aportado información sobre el efecto a nivel transcripcional que tiene la infección con el virus del papiloma humano. Lohavanichbut et al., utilizando la plataforma de Affymetrix en cáncer orofaríngeo, compararon muestras de tumores negativos y

positivos a la infección por algún genotipo de virus del papiloma humano. En este estudio reportaron 446 genes que se expresaban diferencialmente entre estos tumores, los cuales estaban implicados en vías de señalización como reparación al daño en el DNA, recombinación del DNA, replicación del DNA y el ciclo celular (Lohavanichbutr et al., 2009).

En el caso de cáncer cérvico uterino varios estudios han demostrado que cada variante de HPV16 induce un perfil de expresión global diferente, de acuerdo al polimorfismo que posean sus oncogenes. Por ejemplo, en el estudio Zacapala-Gomez *et al* se realizó un microarreglo de la plataforma Affymetrix utilizando células transfectadas con las principales variantes de VPH16. En este trabajo se reporta que cada variante induce un perfil de expresión diferencial, donde las principales vías alteradas por cada variante del virus están relacionadas con el desarrollo de cáncer cérvico uterino (Zacapala-Gómez et al., 2016).

Estos trabajos han demostrado que existe un perfil de expresión asociado a cada variante de VPH16, sin embargo, aún no se ha elucidado las diferencias entre los restantes genotipos del virus en cáncer cérvico uterino que poseen mayor prevalencia y su impacto en el perfil de expresión global.

Justificación

El cáncer cérvico uterino es una de las principales causas de muerte en mujeres de países en vías de desarrollo, donde el 99% de los tumores cérvico uterinos están infectados con al menos un genotipo de VPH. Se sabe que las variaciones en los oncogenes de los distintos genotipos de VPH tienen diferentes efectos biológicos en el organismo hospedero, diversos estudios han mostrado el impacto en el perfil de expresión que tiene la infección por algún genotipo de VPH de alto riesgo en el perfil de expresión de tumores, donde se han visto diferencias entre tumores no infectados con el virus y tumores infectados. Sin embargo, aún no se han elucidado las diferencias en los perfiles de expresión de tumores infectados con distintos genotipos virales. Ya que existe un gran número de genotipos de VPH, es de gran importancia conocer el impacto que tienen a nivel de expresión global, con el fin de obtener y desarrollar nuevas herramientas terapéuticas. Además, esta información podría mejorar nuestro conocimiento de la biología de la infección de cada genotipo de VPH y su relación con el desarrollo de cáncer cervico uterino, conociendo los genes diferencialmente expresados por cada genotipo inclusive existe la posibilidad de mejorar la elección del tratamiento suministrado a las pacientes.

Objetivos

Objetivo general

- Identificar el perfil de expresión global asociado a la infección por distintos genotipos del virus del papiloma humano

Objetivos particulares

- Obtener los genes diferencialmente expresados por cada genotipo
- Obtener las vías de señalización alteradas por cada genotipo
- Comparar la supervivencia de las pacientes de acuerdo a su genotipo viral

Hipótesis

- La afinidad y las moléculas blanco de los oncogenes del virus del papiloma humano están dadas por los polimorfismos únicos de cada genotipo viral, por lo que cada genotipo mostrara un perfil de expresión diferente en los tumores de cáncer cérvico uterino infectados por algún genotipo de VPH.

Metodología

Obtención de los datos

Se analizaron los datos de expresión global de 89 muestras de biopsias realizadas en el laboratorio de Oncogenómica del Instituto Nacional de Cancerología. Se partió de los datos generados por el microarreglo de la marca NimbleGen, utilizando la plataforma *GPL10191 NimbleGen Homo sapiens (12x135k)* y *GPL16025 NimbleGen Homo sapiens Expression Array*, los cuales están compuestos por 45000 secuencias (Fernandez-Retana et al., 2015).

Se obtuvieron los archivos *.PAIR* del microarreglo utilizando la plataforma *Gene Expression omnibus*, con el número de acceso GSE56303, donde se encuentran alojados los archivos, para lo cual se utilizó el siguiente script utilizando el lenguaje de programación R y el paquete Bioconductor.

```
library(Biobase)
```

```
library(GEOquery)
```

```
gset <- getGEO("GSE56303", GSEMatrix =TRUE, AnnotGPL=FALSE)
```

```
gset_log2 <- log2(gset)
```

Cabe destacar que los archivos .PAIR contenían los valores de las intensidades de las sondas normalizados por el método de Robust Multiarray Average, por lo que se procedió a transformar las intensidades en logaritmo base 2, siguiendo el protocolo de Quackenbush (Sturn, Quackenbush, & Trajanoski, 2002), esto con el fin de reducir la variación en las intensidades a un orden de magnitud y reducir las probabilidades de encontrar falsos positivos durante la implementación de los posteriores análisis.

El resultado es una matriz de tipo *ExpressionSet*, donde las columnas corresponden al *accession number* reportado en el NCBI de cada sonda y las columnas a cada paciente, como se muestra en la figura 7.

Determinación de los genotipos más frecuentes

Posteriormente, con los datos clínicos de las pacientes, se relacionó el número de identificación de cada microarreglo y el genotipo viral que se había detectado en cada muestra de tumor. De acuerdo a los datos clínicos de las pacientes se obtuvieron los genotipos más frecuentes y se agruparon los tumores de acuerdo a estos. Los tumores que presentaron más de un genotipo fueron clasificados como coinfecciones.

	T45_439287A03	T45_439287A09	T45_442723A09	T45_445087A03	T45_449245A11	T45_509837A02	T45_532138A03
AB000409	9.497516	8.769942	7.658502	10.891662	10.830422	7.367277	5.403211
AB000463	6.838034	7.889839	5.954606	8.938479	9.564160	8.960719	10.707050
AB000781	6.629248	4.557352	5.927786	4.501217	4.812784	6.515030	4.153934
AB001328	9.311690	6.810075	8.088714	7.744235	5.867842	11.783518	3.704500
AB002294	11.261947	10.858610	10.204058	10.974548	11.496205	4.016134	9.282234
AB002308	8.878355	11.202992	9.635301	11.343734	11.318909	11.710399	5.878468
AB002311	4.337830	8.827258	9.990198	7.297858	7.809863	10.551454	7.083620
AB002313	10.672455	11.444783	10.452367	12.242376	12.363377	4.168865	10.800827
AB002360	9.146386	8.921836	8.297345	7.678051	6.151455	7.277727	10.700799
AB002377	8.394595	8.218963	8.822148	9.938355	10.354743	8.156643	8.498644
AB002381	8.137371	8.783715	7.287175	9.518066	9.854602	7.188525	10.625443
AB002382	10.403014	11.663413	11.447914	11.778387	11.731478	8.494199	6.961683
AB002384	4.254042	6.602847	4.345146	3.829868	6.514679	5.761565	9.734218
AB003177	7.746867	10.630127	9.067262	10.794690	9.369082	6.265672	7.127754
AB003333	6.854795	11.966034	10.983869	11.486247	11.766045	6.372868	4.237941
AB006589	7.413988	7.319183	4.148270	6.964922	8.921140	5.096058	5.938515
AB006590	6.673268	8.578889	5.025729	6.396348	9.521052	10.096806	7.084766
AB006621	9.689930	8.337905	11.679258	5.633490	3.934248	4.571762	10.863407
AB006625	7.670114	7.472203	10.601485	5.578829	4.188439	6.040156	6.151633

Showing 1 to 19 of 45,033 entries

Figura 7. ExpressionSet obtenido de la plataforma Gene Expression Omnibus.

El resultado del procesado de los datos es una matriz donde las filas corresponden al *accesión number* de cada sonda de acuerdo al NCBI y las columnas a cada paciente.

Análisis de expresión diferencial

Se realizó el análisis de expresión diferencial entre cada los genotipos más frecuentes mediante el diseño de un script en el lenguaje de programación R, en el cual se utilizo un modelo lineal para cada gen utilizando el paquete de bioconductor *limma* (*Linear Models for Microarray*) (Ritchie et al., 2015).

Se asume que la fórmula de la regresión lineal es: $E[y_j] = X\alpha_j$, donde y_j contiene los datos de expresión del gen j , X es la matriz del diseño y α_j es el vector de los coeficientes.

Entonces, y_j^T es la j va fila de la matriz de expresión y contiene las intensidades de expresión en logaritmo.

Los contrastes de interés están dados por $\beta_j = \mathbf{C}^T \alpha_j$, donde \mathbf{C} es la matriz de los contrastes. Durante el análisis diferencial el componente de los coeficientes del β_j del modelo producto de la regresión lineal contiene los valores estimados de α_j , lo que permite comparar los contrastes entre ellos (G. K. Smyth, 2005). La distribución de los datos producto de la regresión lineal se ajustó usando el modelo *Empirical Bayes* propuesto por Smyth en el 2004 (G. K. Smyth, 2004), esto con el fin de resolver el hecho de que la variabilidad en los valores de expresión varía entre los genes.

Se crearon tres matrices, la matriz de expresión, de diseño y de contrastes. La primera matriz posee el *ExpressionSet* del microarreglo, la segunda matriz contiene el diseño del experimento representando y relacionando las pacientes con los tratamientos y la tercer matriz posee los contrastes del experimento, esta última matriz contiene los coeficientes definidos en la matriz de diseño combinados con los contrastes de interés. El script utilizado se muestra a continuación.

```
library(limma)
library(Biobase)
library(bioDist)
library(genefilter)
rm(list=ls(all=TRUE))

datos <- read.csv("LOG_16_VS_18.csv", header = TRUE, row.names = 1)
pData <- read.csv("pDATA_2.csv", header = TRUE, row.names = 1)
T16 <- pData$T16
T18 <- pData$T18
unidos <- cbind(T16, T18)
rownames(unidos) <- rownames(pData)
unidos <- as.data.frame(unidos)
pData <- Unidos
datosmatrix <- as.matrix(datos)
phenoData <- new("AnnotatedDataFrame", data=pData)
```

```

ExprSet<- ExpressionSet(assayData=datosmatrix,phenoData=phenoData)
design <- pData
fit <- lmFit(datos, design)
cont.matrix <- makeContrasts(T16vsT18=T18-T16, levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2)
resultados_bien <- topTable(fit2, adjust="BH")
resultados_todos_BIEN<-topTable(fit2,sort="none",n=Infadjust="BH")
write.csv(resultados_todos_BIEN,"resultados_TODOS_bien.csv")

```

Obtención de los *heatmaps*

Solo se tomaron en cuenta aquellos genes cuyo pvalue fuera menor a 0.01, estos genes fueron extraídos del *ExpressionSet* para posteriormente ser graficados en un *heatmap*

Los *heatmaps* fueron construidos utilizando la librería de Bioconductor *gplots*. Se partió del *ExpressionSet* con los datos normalizados en logaritmo base2. Con el objetivo de centrar los datos y observar los cambios en la expresión de cada gen en cada paciente se calculó el Z-Score de cada valor en el *ExpressionSet*, para lo cual se utilizó la siguiente formula:

$$z = \frac{x - \mu}{\sigma}$$

Donde:

X: Cada valor dentro del *ExpressionSet*

μ : Media de cada gen

σ : Desviación estándar de cada gen

El script en R utilizado para extraer los genes significativos y calcular el Z-Score fue el siguiente:

```
library(matrixStats)
diffexprarchivo <- read.csv("resultados_TODOS_bien.csv",header = TRUE,row.names =
1)
datosmatrix_ordenado <- datosmatrix
genes_significativos <- subset(diffexprarchivo, P.adj < 0.01)
genes_significativos_nombres<- rownames(genes_significativos)

índice<-row.names(datosmatrix_ordenado) %in% genes_significativos_nombres
matriz_selected <- datosmatrix_ordenado[índice,]
matriz_menos_ROWMEANS<-matriz_selected- rowMeans(matriz_selected)
matriz_desvest <- matriz_menos_ROWMEANS / rowSds(matriz_selected)
matriz_foldchange <- matriz_desvest
```

Los *heatmaps* se construyeron con la matriz que contenía el Z-Score de cada gen, después estos se agruparon en clústers de manera supervisada y no supervisada utilizando la distancia euclidiana, con el coeficiente de agrupamiento promedio, esto se hizo utilizando el siguiente script:

```
control_nombres_index<- which(design$T16 == 1)
caso_nombres_index <- which(design$T18 == 1)
numero_control <- length(control_nombres_index)
numero_caso <- length(caso_nombres_index)

colores<-rep(c("blue","red"), c(numero_control,numero_caso))
heatmap.2(as.matrix(matriz_foldchange),ColSideColors=colores,
col=redblue(100),trace=c("none"), margins = c(11.5, 8),
density.info=c("none"), cexRow=0.2,main="T16 VS T18",
distfun=function(x) dist(x, method="euclidian"),key = TRUE)
par(xpd=TRUE)
legend(.85,1.095,legend=c("T16","T18"),
fill=c("red","blue", border=FALSE, bty="0", y.intersp = 1, cex=0.7
```

Análisis de enriquecimiento de términos

Con el fin de observar los procesos biológicos más relevantes involucrados con el perfil de expresión asociado a cada genotipo se hizo un análisis de enriquecimiento de términos y ontología de genes utilizando el paquete de Bioconductor, *Generally Applicable Gene-set Enrichment* (GAGE) (Luo et al., 2009), este paquete permite asignar un p-value a cada vía de acuerdo al número de genes alterados, importancia y posición en cada vía. Primero se convirtieron los *accession id* de cada gen a la nomenclatura *Entrez ID*, esto se realizó utilizando la última versión del *Genome wide annotation for Human*, contenida en el paquete *org.Hs.eg.db* del proyecto Bioconductor (Carlson M, 2016). Finalmente los *Entrez ID* fueron introducidos en *GAGE* utilizando el script que se detalla a continuación, de esta plataforma se obtuvieron las principales vías de señalización alteradas de cada comparación.

```
library(pathview)
library(gage)
library(gageData)
kg.hsa <- kegg.gsets( "hsa" )
data(kegg.sets.hs)
data(sigmet.idx.hs)
kegg.sets.hs = kg.hsa$kg.sets[ kg.hsa$sigmet.idx ]
data(go.sets.hs)
gobpsets = go.sets.hs[go.subs.hs$BP]
gobpres = gage(foldchanges, gsets=gobpsets, same.dir=TRUE)
GO_ORDENADO <- lapply(gobpres, head)
GO_SOBREXPRESADOS <- GO_ORDENADO$greater
write.csv(GO_SOBREXPRESADOS,"GO_SOBREXPRESADOS.csv")
GO_SUBEXPRESADOS <- GO_ORDENADO$less
write.csv(GO_SUBEXPRESADOS,"GO_SUBEXPRESADOS.csv")
```

Análisis de Supervivencia

El análisis de supervivencia se realizó utilizando el estimador Kaplan-Meier mediante el paquete *survival* incluido en el lenguaje de programación R. El tiempo libre de enfermedad (*Disease Free Survival*) se calculó restando la fecha de diagnóstico de enfermedad recurrente de cada paciente menos la fecha de inicio de la braquiterapia, en los casos en los que no se hubiera reportado enfermedad recurrente se tomó en cuenta la fecha de último monitoreo.

El tiempo de supervivencia global (*Overall Survival*) se calculó restando la fecha de último monitoreo menos la fecha de diagnóstico por el instituto, en el caso de las pacientes fallecidas se restó la fecha de defunción menos la fecha de diagnóstico por el instituto. El script utilizado para el análisis de supervivencia se describe a continuación.

```
library(survival)

datos_survival <- read.csv("Supervivencia_VPH.csv")
supervivencia <- Surv(time = datos_survival$DFS, event =
datos_survival$STATUS.CENSORED,type = "right")
supervivencia.km <- survfit(Surv(time = DFS, event = STATUS.CENSORED, type =
"right") ~ GRUPO, data = datos_survival)
plot(supervivencia.km, lty = 1, col = c("darkred", "darkblue","green"),
      ylab = "Proporción de Supervivencia", xlab = "Días", main= "Disease Free
Survival VPH")

par(xpd=TRUE)
legend("bottomright", legend = c("Coinfeccion", "Genotipo 16","Genotipo
18","Genotipo 45"),
      lty = 1:1, col = c("darkred", "darkblue","green","black","white"))

diferencia <- survdiff(formula = Surv(time = DFS,event = STATUS.CENSORED, type =
"right") ~ GRUPO, data = datos_survival)
```

Resultados

Las muestras de los tumores analizados en el microarreglo provinieron de 89 pacientes del Instituto Nacional de Cancerología, a las cuales se les dio seguimiento de abril de 2010 a Agosto de 2012, de estas pacientes solo en 63 pacientes se detectó algún genotipo de VPH.

Tabla 2.- Características clínicas de las pacientes. Se muestran las características patológicas y tipificación de las pacientes, la muestra fue de 63 pacientes infectadas con algún genotipo de VPH

Características		Pacientes		Genotipificación VPH (Frecuencia)			
Edad		N		Infección simple		Pacientes	Porcentaje
	Media	46		Tipo 16	29	46.0%	
	Rango	29-65		Tipo 18	5	7.9%	
Tipo histológico			Porcentaje	Tipo 45	9	14.2%	
	Carcinoma epidermoide	57	90.4%	Otro	7	11.1%	
	Adenocarcinoma	6	9.6%	Pacientes con coinfección	13	20.6%	
Tamaño del tumor				Pacientes sin coinfección	50	79.3%	
	<=4 cm	20	31.7%	Coinfección	Tipo 16	6	46.2%
	>4 cm	40	63.4%		Tipo 18	4	30.8%
	Sin datos	3	4.7%		Tipo 51	1	7.7%
			Tipo 33		1	7.7%	
Estadio Clínico				Tipo 52	4	30.8%	
	IB2	9	14.2%	Tipo 35	1	7.7%	
	IIA	1	1.6%	Tipo 58	3	23.1%	
	IIB	39	61.9%	Tipo 49	1	7.7%	
	IIIA	1	1.6%	Tipo 59	1	7.7%	
	IIIB	13	20.6%	Respuesta al tratamiento	Respuesta completa	36	57.1%
			Sin respuesta		26	41.2%	
			Sin datos		1	1.6%	

La infección simple fue el tipo de infección más común, donde 50 pacientes presentaron solo un genotipo viral, mientras que 13 pacientes dieron positivo a múltiples genotipos virales. El genotipo viral con mayor frecuencia en las pacientes fue el tipo 16, el cual estuvo presente en 35 pacientes, de las cuales 29 tuvieron infección simple (46%) y 6 como parte de una coinfección (Ver Tabla 2).

En las infecciones simples, el segundo genotipo más frecuente fue el tipo 45, el cual fue encontrado en 9 pacientes (14.2%), seguido del tipo 18, presente en 5 pacientes (7.9%)

Tabla 3.- Genotipos virales presentes en las coinfecciones. Las pacientes que presentaban más de una infección se denominaron coinfecciones, en total encontramos 13 casos de coinfecciones.

Coinfección	Número de pacientes	Porcentaje
16 Y 18	1	8%
16, 31 Y 59	1	8%
16 Y 32	1	8%
16, 33 Y 52	1	8%
16, 35 Y 52	1	8%
16 Y 49	1	8%
18 Y 16	1	8%
45 Y 51	1	8%
58 Y 52	2	15%
58, 52 Y 45	1	8%
6 Y 18	2	15%

En las pacientes que presentaron coinfecciones, se encontraron 9 genotipos diferentes, de los cuales el genotipo 18 y 52 fueron los segundos tipos más frecuentes, al presentarse en 4 pacientes cada uno (30.8%), mientras que el tipo 58 fue el tercero más frecuente, al estar infectando a 3 pacientes (23.1%). Las coinfecciones con una mayor frecuencia fueron la infección con los genotipos 58 con 52 y los genotipos 6 con 18. El único genotipo de bajo riesgo encontrado tanto en infecciones simples como en las coinfecciones fue el tipo 6 (Ver Tabla 3).

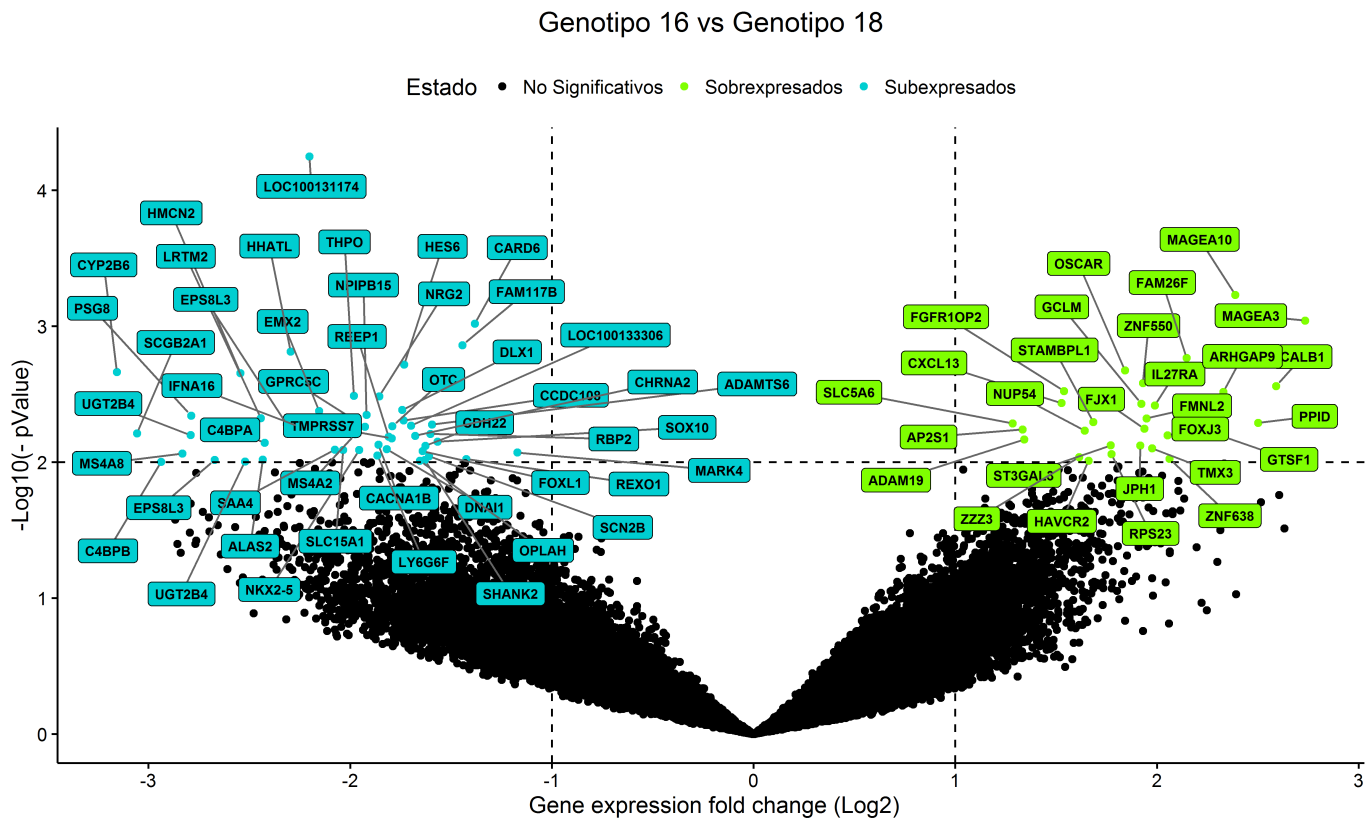


Figura 8.- Genes diferencialmente expresados entre los tumores infectados con el genotipo 16 en comparación con tumores con el genotipo 18.

La grafica de volcán muestra en color verde los genes que se encontraban sobreexpresados y en azul los subexpresados, además de tener un pvalue < 0.01.

Posteriormente se separaron las pacientes de acuerdo a su genotipo viral en 4 grupos; Genotipo 18, Genotipo 16, Genotipo 45 y Coinfección. Estos grupos se compararon durante el análisis de expresión diferencial utilizando el paquete de bioconductor *limma*, utilizando la siguiente fórmula para los contrastes: $\text{Contraste1vsContraste2} = \text{Contraste2} - \text{Contraste1}$ (Smyth, G. K, 2004). Para filtrar a los genes diferencialmente expresados se estableció un límite basado en un fold change mínimo de 1 ó -1 y un p value menor a 0.01, donde el Fold change se refiere a la absoluta diferencia en los ratios de expresión (Figura 8, 9 y 10).

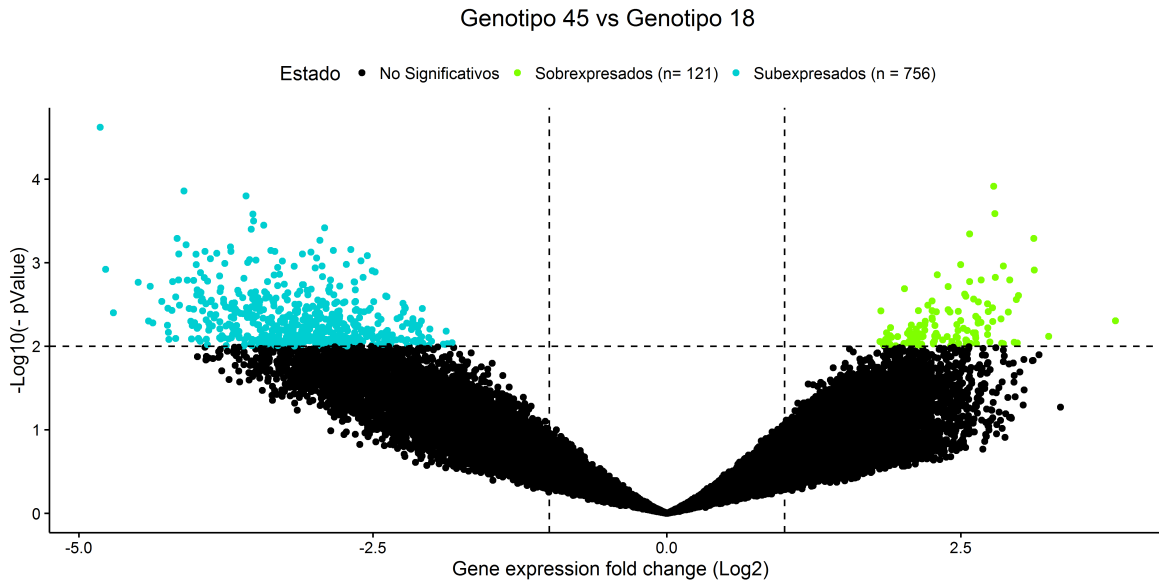


Figura 10.- Genes diferencialmente expresados entre los tumores infectados con el genotipo 45 en comparación con tumores con el genotipo 18.

La grafica de volcán muestra en color verde los genes que se encontraban sobrexpresados y en azul los sobrexpresados, además de tener un pvalue < 0.01

El análisis de expresión diferencial entre los tumores con genotipo 16 y el genotipo 18 mostro que 76 genes se encontraban diferencialmente expresados entre estos grupos, de los cuales 48 genes se encontraban subexpresados, con fold changes que iban de -3.15 a -1.17 (Ver Tabla 4), mientras que 28 genes se encontraban sobrexpresados, con fold changes de 1.28 a 2.73 (Ver Tabla 5).

Con el fin de facilitar la lectura de este manuscrito solo se presentan los primeros 20 genes diferencialmente expresados en cada análisis, la lista completa de genes puede encontrarse en la sección de anexos.

Tabla 4.- Genes subexpresados en la comparación entre el genotipo 16 y el genotipo 18. En total se encontraron 48 genes que se subexpresaban en el perfil de expresión entre estos tumores, se tomaron en cuenta aquellos genes cuyo pvalue < 0.01. Se muestran los primeros 20 genes subexpresados, ordenados de acuerdo al logFC. La lista completa de genes puede encontrarse en el Anexo A.

SEQ_ID	Símbolo del Gen	Nombre del gen	logFC
BC067430	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	-3.156410655
NM_002407	SCGB2A1	secretoglobin, family 2A, member 1	-3.056968509
NM_000716	C4BPB	complement component 4 binding protein, beta	-2.937943288
BC022895	MS4A8	membrane-spanning 4-domains, subfamily A, member 8	-2.831534973
AY529122	UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	-2.790705225
CR749812	PSG8	pregnancy specific beta-1-glycoprotein 8	-2.788361544
NM_133181	EPS8L3	EPS8-like 3	-2.672363275
AK128680	HMCN2	hemicentin 2	-2.544149595
NM_021139	UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	-2.522004017
NM_001039029	LRTM2	leucine-rich repeats and transmembrane domains 2	-2.442761206
NM_000032	ALAS2	5'-aminolevulinate synthase 2	-2.434168695
NM_000715	C4BPA	complement component 4 binding protein, alpha	-2.423784254
NM_024526	EPS8L3	EPS8-like 3	-2.315194387
NM_020707	HHATL	hedgehog acyltransferase-like	-2.295836097
NM_002173	IFNA16	interferon, alpha 16	-2.253146359
AK130404	LOC100131174	uncharacterized LOC100131174	-2.202769364
NM_004098	EMX2	empty spiracles homeobox 2	-2.154788573
NM_006512	SAA4	serum amyloid A4, constitutive	-2.076821893
NM_005073	SLC15A1	solute carrier family 15 (oligopeptide transporter), member 1	-2.035558419
NM_199228	THPO	thrombopoietin	-1.98184416

Tabla 5.- Genes sobreexpresados en la comparación entre el genotipo 16 y el genotipo 18. En total se encontraron 28 genes que se sobreexpresaban en el perfil de expresión entre estos tumores, se tomaron en cuenta aquellos genes cuyo pvalue < 0.01. Se muestran los primeros 20 genes sobreexpresados, ordenados de acuerdo al logFC. La lista completa de genes puede encontrarse en el Anexo B.

SEQ_ID	Simbolo del Gen	Nombre del gen	logFC
NM_005362	MAGEA3	melanoma antigen family A3	2.734401484
NM_004929	CALB1	calbindin 1, 28kDa	2.590339336
NM_005038	PPID	peptidylprolyl isomerase D	2.501957528
NM_001011543	MAGEA10	melanoma antigen family A10	2.388234965
NM_144594	GTSF1	gametocyte specific factor 1	2.337821902
NM_032496	ARHGAP9	Rho GTPase activating protein 9	2.328701732
NM_001010919	FAM26F	family with sequence similarity 26, member F	2.146629999
AF273049	ZNF638	zinc finger protein 638	2.062361562
NM_014947	FOXJ3	forkhead box J3	2.053159197
NM_004843	IL27RA	interleukin 27 receptor, alpha	1.988933437
BC093792	TMX3	thioredoxin-related transmembrane protein 3	1.975284148
NM_001004417	FMNL2	formin-like 2	1.948763145
BC071898	FJX1	four jointed box 1	1.938081027
BC034810	ZNF550	zinc finger protein 550	1.930424657
NM_002061	GCLM	glutamate-cysteine ligase, modifier subunit	1.922287974
NM_020647	JPH1	junctophilin 1	1.91687764
BC035023	OSCAR	osteoclast associated, immunoglobulin-like receptor	1.841238364
NM_001025	RPS23	ribosomal protein S23	1.774331437
AY167994	ST3GAL3	ST3 beta-galactoside alpha-2,3-sialyltransferase 3	1.770324578
NM_020799	STAMBPL1	STAM binding protein-like 1	1.684275625

Cuando se compararon los tumores de pacientes que tenían una coinfección y las que poseían el genotipo viral 18 se obtuvieron 237 genes diferencialmente expresados que fueron significativos, de los cuales 96 genes se encontraban subexpresados, con fold changes que iban de -3.57 a -1.51 (Ver Tabla 6) y 141 genes sobreexpresados, con un rango de fold changes de 3.35 a 1.58 (Ver Tabla 7).

Tabla 6.- Genes subexpresados en la comparación entre pacientes que presentaban una coinfección y entre pacientes que estaban infectados con el genotipo 18.

En total se encontraron 96 genes que se subexpresaban en el perfil de expresión entre estos tumores, se tomaron en cuenta aquellos genes cuyo pvalue < 0.01. Se muestran los primeros 20 genes subexpresados, ordenados de acuerdo al logFC. La lista completa de genes puede encontrarse en el Anexo C.

SEQ_ID	Símbolo del gen	Nombre del gen	logFC
BC020864	CALB1	calbindin 1, 28kDa	-3.574370935
NM_014229	SLC6A11	solute carrier family 6 (neurotransmitter transporter), member 11	-3.08086244
BC009568	SYNGR3	synaptogyrin 3	-3.057607468
AB209790	AAMP	angio-associated, migratory cell protein	-3.056432036
NM_001039569	AP1S3	adaptor-related protein complex 1, sigma 3 subunit	-2.794810043
NM_001008660	PICALM	phosphatidylinositol binding clathrin assembly protein	-2.784595504
BC033637	SSX2IP	synovial sarcoma, X breakpoint 2 interacting protein	-2.778113542
NM_007026	DUSP14	dual specificity phosphatase 14	-2.767832347
NM_005038	PPID	peptidylprolyl isomerase D	-2.76456386
NM_001011543	MAGEA10	melanoma antigen family A10	-2.679500451
BC109372	MRPS35	mitochondrial ribosomal protein S35	-2.671048529
X82460	HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	-2.641096722
AK074748	ELOVL5	ELOVL fatty acid elongase 5	-2.557841702
BC110820	PHLDA1	pleckstrin homology-like domain, family A, member 1	-2.551086162
NM_001740	CALB2	calbindin 2	-2.516769728
CR933633	STK4	serine/threonine kinase 4	-2.509801105
BC093792	TMX3	thioredoxin-related transmembrane protein 3	-2.49797601
NM_004165	RRAD	Ras-related associated with diabetes	-2.470471282
BC062779	ABCA3	ATP-binding cassette, sub-family A (ABC1), member 3	-2.458916156
NM_003639	IKBKG	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma	-2.444407748

Tabla 7.- Genes sobreexpresados en la comparación entre pacientes que presentaban una coinfección y entre pacientes que estaban infectados con el genotipo 18. En total se encontraron 141 genes que se sobreexpresaban en el perfil de expresión entre estos tumores, se tomaron en cuenta aquellos genes cuyo pvalue < 0.01. Se muestran los primeros 20 genes sobreexpresados, ordenados de acuerdo al logFC. La lista completa de genes puede encontrarse en el Anexo D.

SEQ_ID	Símbolo del gen	Nombre del gen	logFC
AY341951	FAM138D	family with sequence similarity 138, member D	4.178833605
BC110891	GSTA1	glutathione S-transferase alpha 1	3.871957744
BC008608	SMOC1	SPARC related modular calcium binding 1	3.860941096
AF346307	FAM138B	family with sequence similarity 138, member B	3.727853165
BC008405	PSG4	pregnancy specific beta-1-glycoprotein 4	3.714423903
NM_174962	SSX9	synovial sarcoma, X breakpoint 9	3.692214485
NM_032785	AGBL4	ATP/GTP binding protein-like 4	3.686885706
NM_001005484	OR4F5	olfactory receptor, family 4, subfamily F, member 5	3.662567527
NM_001005283	OR9Q2	olfactory receptor, family 9, subfamily Q, member 2	3.612348726
AK094547	SLC7A14	solute carrier family 7, member 14	3.580650867
NM_001004688	OR2M2	olfactory receptor, family 2, subfamily M, member 2	3.553429991
CR749812	PSG8	pregnancy specific beta-1-glycoprotein 8	3.522400136
NM_001004695	OR2T33	olfactory receptor, family 2, subfamily T, member 33	3.514389063
NM_000716	C4BPB	complement component 4 binding protein, beta	3.505570859
AY792621	OR4F13P	olfactory receptor, family 4, subfamily F, member 13 pseudogene	3.483231037
NM_175067	TAAR6	trace amine associated receptor 6	3.455847011
NM_001034832	SSX4B	synovial sarcoma, X breakpoint 4B	3.447173774
BC110494	PIGR	polymeric immunoglobulin receptor	3.419482112
NM_001005323	OR5AK2	olfactory receptor, family 5, subfamily AK, member 2	3.407403368
NM_021016	PSG3	pregnancy specific beta-1-glycoprotein 3	3.393350186

Tabla 8.- Genes subexpresados en pacientes infectados con el genotipo 45 y pacientes infectados con el genotipo 18. En total se encontraron 117 genes que se subexpresaban en el perfil de expresión entre estos tumores, se tomaron en cuenta aquellos genes cuyo pvalue < 0.01. Se muestran los primeros 20 genes subexpresados, ordenados de acuerdo al logFC. La lista completa de genes puede encontrarse en el Anexo E.

SEQ_ID	Símbolo del gen	Nombre del Gen	logFC
BC008405	PSG4	pregnancy specific beta-1-glycoprotein 4	-4.819313911
NM_001005501	OR4K2	olfactory receptor, family 4, subfamily K, member 2	-4.773906534
NM_172194	OR4Q3	olfactory receptor, family 4, subfamily Q, member 3	-4.708373645
NM_001332	CTNND2	catenin (cadherin-associated protein), delta 2	-4.4963236
BC001003	SSX1	synovial sarcoma, X breakpoint 1	-4.408074137
AY341951	FAM138D	family with sequence similarity 138, member D	-4.394538959
NM_001011719	ARSH	arylsulfatase family, member H	-4.372404
NM_001005270	OR4C12	olfactory receptor, family 4, subfamily C, member 12	-4.296423777
NM_001004715	OR4K17	olfactory receptor, family 4, subfamily K, member 17	-4.247616526
BC027895	REG1B	regenerating islet-derived 1 beta	-4.244033948
NM_001005280	OR10A7	olfactory receptor, family 10, subfamily A, member 7	-4.240636455
NM_001005281	OR6B1	olfactory receptor, family 6, subfamily B, member 1	-4.236678702
NM_001004473	OR10K1	olfactory receptor, family 10, subfamily K, member 1	-4.208789764
AK094547	SLC7A14	solute carrier family 7, member 14	-4.20369805
NM_020178	CA10	carbonic anhydrase X	-4.178550715
NM_003996	GPX5	glutathione peroxidase 5	-4.177121857
NM_153444	OR5P2	olfactory receptor, family 5, subfamily P, member 2	-4.165421479
NM_178466	BPIFA3	BPI fold containing family A, member 3	-4.154304537
NM_001004688	OR2M2	olfactory receptor, family 2, subfamily M, member 2	-4.151217622
NM_000936	PNLIP	pancreatic lipase	-4.145549092

Finalmente, en el análisis entre tumores infectados con el genotipo 45 y el genotipo 18 encontramos el mayor número de genes diferencialmente expresados entre todas las comparaciones, con 679 genes significativos, de los cuales 117 genes se encontraban subexpresados (Ver Tabla 8) y 562 genes sobreexpresados (Ver Tabla 9), con fold changes que iban en un rango de -4.81 a 3.81.

Tabla 9.- Genes sobreexpresados en la comparación entre pacientes infectados con el genotipo 45 y pacientes infectados con el genotipo 18. En total se encontraron 562 genes que se sobreexpresaban en el perfil de expresión entre estos tumores, se tomaron en cuenta aquellos genes cuyo pvalue < 0.01. Se muestran los primeros 20 genes sobreexpresados, ordenados de acuerdo al logFC. La lista completa de genes puede encontrarse en el Anexo F.

SEQ_ID	Símbolo del gen	Nombre del Gen	logFC
NM_000269	NME1	NME/NM23 nucleoside diphosphate kinase 1	3.814219077
BC011396	PUS7	pseudouridylate synthase 7 (putative)	3.246815566
NM_005038	PPID	peptidylprolyl isomerase D	3.124050817
NM_003385	VSNL1	visinin-like 1	3.11986836
AK000591	C1orf159	chromosome 1 open reading frame 159	2.989473217
BC063787	KREMEN1	kringle containing transmembrane protein 1	2.980064313
BC062427	NSMCE4A	NSE4 homolog A, SMC5-SMC6 complex component	2.970632781
BC010901	CKAP2	cytoskeleton associated protein 2	2.952725396
BC034580	SLC44A5	solute carrier family 44, member 5	2.915816956
NM_024698	SLC25A22	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	2.902916752
BC004202	CHEK1	checkpoint kinase 1	2.861238565
X73568	SYK	spleen tyrosine kinase	2.860175507
NM_001017973	P4HA2	prolyl 4-hydroxylase, alpha polypeptide II	2.8472834
NM_173547	TRIM65	tripartite motif containing 65	2.841614199
NM_017422	CALML5	calmodulin-like 5	2.792744413
NM_005252	FOS	FBJ murine osteosarcoma viral oncogene homolog	2.789291476
NM_144693	ZNF558	zinc finger protein 558	2.778064451
NM_004431	EPHA2	EPH receptor A2	2.771499601
NM_030816	ANKRD13C	ankyrin repeat domain 13C	2.768976687
BC093031	TSPAN4	tetraspanin 4	2.747098654

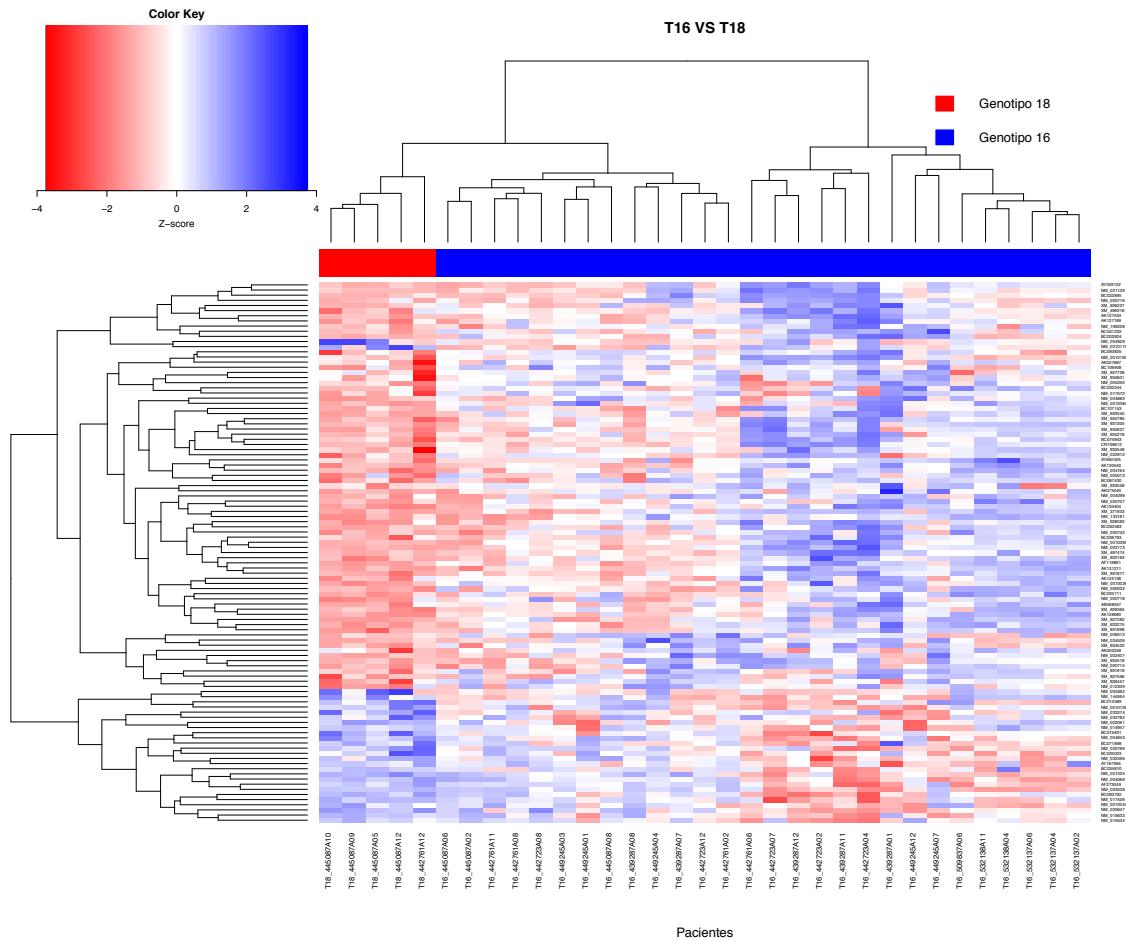


Figura 11. Heatmap de la expresión diferencial entre los tumores que presentaron una infección simple con el genotipo 16 y tumores que presentaron el genotipo 18.

Se muestra el agrupamiento en clúster de acuerdo a la expresión de los genes diferencialmente expresados. La expresión de los genes se muestra en un gradiente de color donde el rojo denota una subexpresión y el color azul una sobrexpresión.

Una vez obtenidos los genes diferencialmente expresados se procedió a realizar un análisis de agrupamiento, esto con el fin de averiguar si la similitud entre la expresión de estos genes permitía agrupar en clústers a las pacientes de acuerdo al genotipo viral con el que estuvieran infectadas. En las figuras 11, 12 y 13 se muestran las gráficas de calor de los perfiles de expresión de las pacientes, donde

los dendogramas indican la similitud que tienen tanto las pacientes como los genes.

En la gráfica de calor entre las pacientes infectadas con el genotipo 18 vs las pacientes con una coinfección se pueden observar claramente dos grupos, los cuales se encuentran bien definidos de acuerdo al tipo de infección.

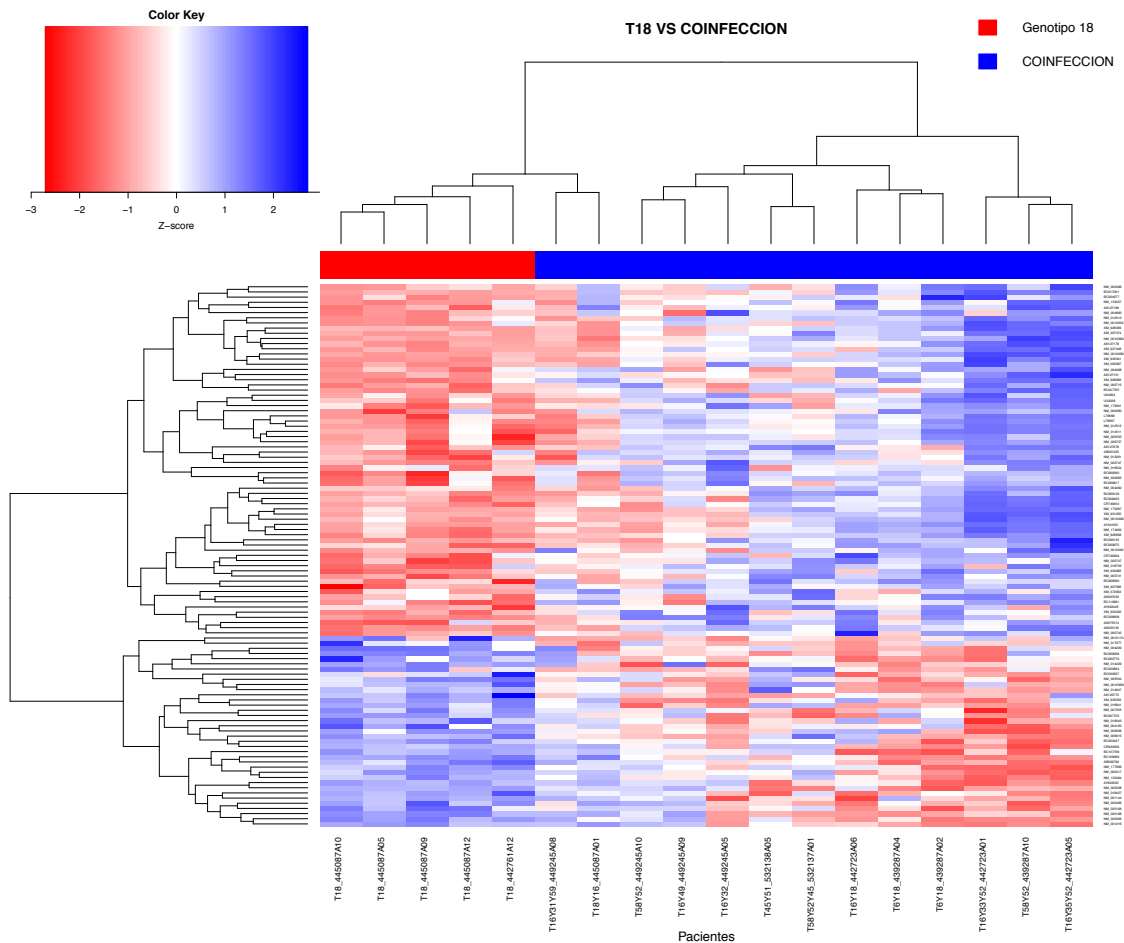


Figura 12. Heatmap de la expresión diferencial entre los tumores que presentaron una infección con el genotipo 18 y los tumores que presentaron una coinfección.

Se muestra el agrupamiento en clúster de acuerdo a la expresión de los genes diferencialmente expresados. La expresión de los genes se muestra en un gradiente de color donde el rojo denota una subexpresión y el color azul una sobreexpresión.

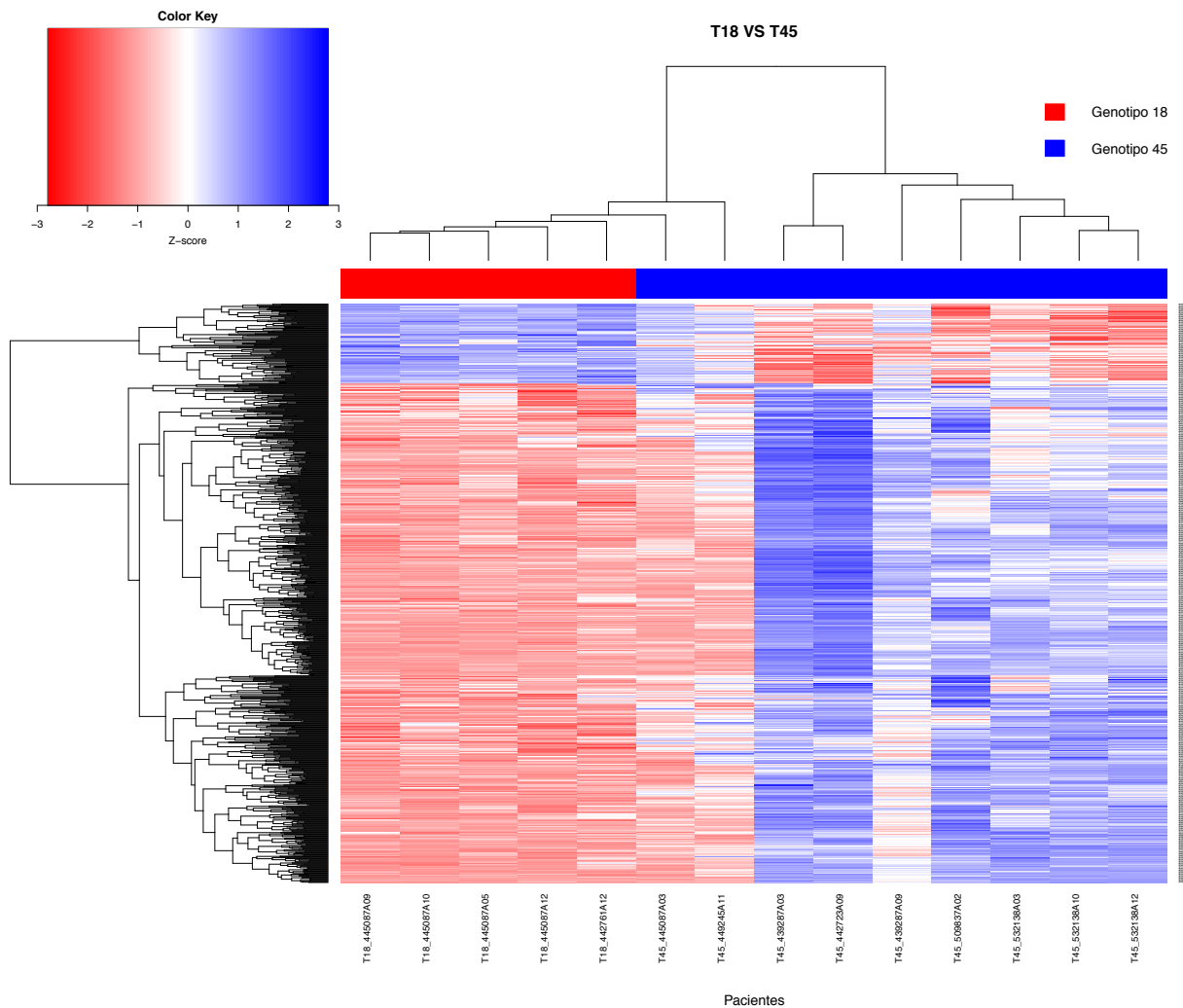


Figura 13. Heatmap de la expresión diferencial entre los tumores que presentaron una infección con el genotipo 18 y el genotipo 45.

Se muestra el agrupamiento en clúster de acuerdo a la expresión de los genes diferencialmente expresados. La expresión de los genes se muestra en un gradiente de color donde el rojo denota una subexpresión y el color azul una sobreexpresión.

En estas tres figuras se puede apreciar que todas las pacientes se agruparon de acuerdo al genotipo viral con el que estuvieron infectadas, lo que sugiere que el

perfil de expresión de los tumores se ve afectado por el genotipo viral presente en las pacientes.

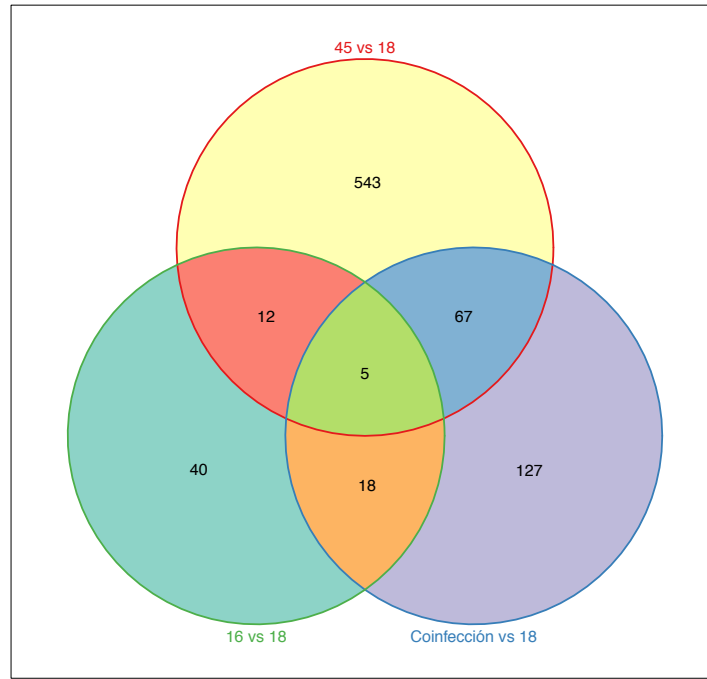


Figura 14. Número de genes diferencialmente expresados compartidos entre **cada comparación de los genotipos**.

Cada círculo corresponde a cada comparación entre los genotipos más frecuentes. Se observa que en total se comparten 5 genes diferencialmente expresados por los genotipos analizados.

Posteriormente comparamos las listas de genes diferencialmente expresados entre las comparaciones, con el fin de observar que genes se compartían entre estos. En total se encontraron 5 genes que se expresaban diferencialmente en las tres comparaciones, los cuales son: CHRNA2, MS4A8, PPID, PSG8 y UGT2B4 (Figura 14).

CHRNA2 codifica para la subunidad alfa del receptor nicotínico de acetilcolina neuronal, el cual está relacionado con la transmisión sináptica rápida. El gen MS4A8 codifica a la proteína del mismo nombre perteneciente a la familia de

proteínas *membrane spanning 4-domains*, caracterizadas por poseer cuatro regiones transmembranales, aunque aún no se conoce la función exacta de estas proteínas se sabe que están relacionadas con la señalización intracelular y el proceso de diferenciación de las células hematopoyéticas y epiteliales, además de que recientemente se ha reportado que en cáncer de colon se pierde su expresión, mientras que su sobreexpresión inhibe la capacidad de las células de proliferar y migrar (Michel et al., 2013).

El gen PPID codifica a una proteína perteneciente a la familia *peptidil-prolil cis-trans isomerasa*, una de sus funciones principales es facilitar el plegamiento de proteínas (Jandova, Janda, & Sligh, 2013). En varios reportes se ha relacionado la alteración de este gen en células tumorales, donde se ha encontrado sub-expresado en líneas celulares de próstata y cáncer de mama (Periyasamy, Warriar, Tillekeratne, Shou, & Sanchez, 2007; Ward, Mark, Ingram, Minchin, & Ratajczak, 1999). PSG8 pertenece a una familia de proteínas conocidas como human *pregnancy-specific glycoproteins*, las cuales son producidas durante el embarazo por los sincitiotrofoblastos placentarios. Además, comprenden un subgrupo de la familia de los antígenos carcinoembrionicos (CEA), los cuales pertenecen a la súper familia de las inmunoglobulinas. No hay reportes en la literatura que relacionen a esta proteína con cáncer.

El gen UGT2B4 codifica para una importante enzima perteneciente a la familia de UDP-glucuronosiltransferasas, expresándose principalmente en el hígado, corazón, próstata, glándulas mamarias y en los riñones. Esta enzima está involucrada principalmente en el metabolismo de varios xenobioticos y substratos

endógenos, especialmente hormonas esteroides y ácidos bilícos. Se sabe que mutaciones en este gen provocan alteraciones graves en el nivel de hormonas esteroides en el cuerpo, y ya que se expresa en las glándulas mamarias se presume que es un factor de riesgo para desarrollar cáncer de mama (Sun et al., 2011).

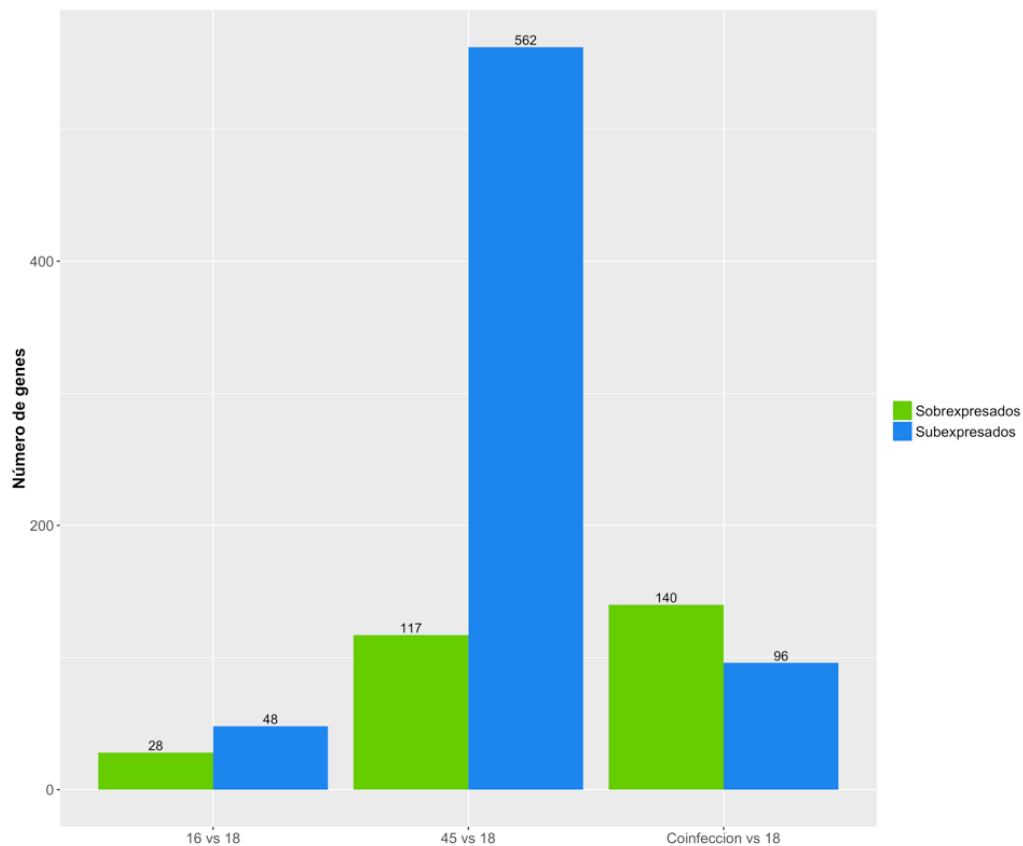


Figura 15. Número de genes diferencialmente expresados en cada comparación.

En color verde se muestran los genes sobrexpresados y en azul los genes sobrexpresados. El mayor número de genes diferencialmente expresados se encontró en la comparación entre tumores infectados con el genotipo 45 y tumores infectados con el genotipo 18.

La comparación de los perfiles de expresión entre las comparaciones muestra que el cambio más dramático en el número de genes diferencialmente expresados se da entre el genotipo 45 y el genotipo 18, donde se ven alterados 679 genes, de los cuales 562 genes se subexpresaron y 112 se subexpresaron. La comparación entre los tumores que presentan infección con el genotipo 18 y los tumores con

coinfecciones muestran un cambio en la expresión de 236 genes, subexpresando 96 genes y sobreexpresando 140, mientras que el menor número de genes diferencialmente expresados se encontró en el perfil de expresión entre la comparación de los tumores con el genotipo 16 y los tumores con el genotipo 18, donde solo se obtuvieron 48 genes subexpresados y 28 sobreexpresados (Ver Figura 15).

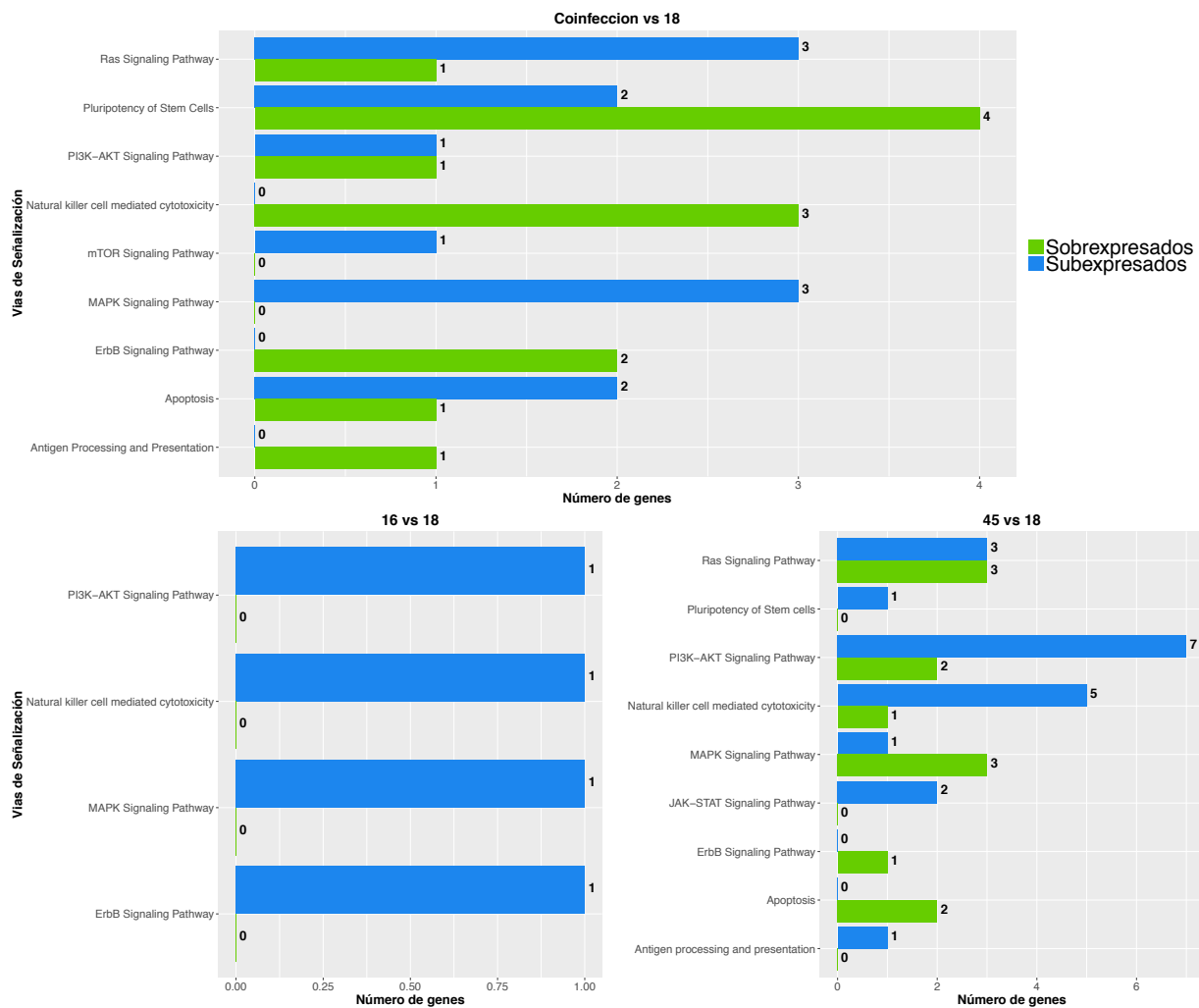


Figura 16. Análisis funcional de las vías de señalización alteradas relacionadas con la progresión y desarrollo de cáncer.

Se observa que cada genotipo afecta diferentes vías de señalización, siendo la vía de pluripotencialidad de células madre la que presenta un mayor número de genes diferencialmente expresados por la infección del genotipo 18 y las coinfecciones.

Posteriormente, se analizó los genes diferencialmente expresados entre cada comparación con la paquetería *gage* de bioconductor. Entre las principales vías alteradas se encontraron vías de señalización involucradas con el metabolismo, sin embargo el objetivo del presente trabajo es reportar las vías relacionadas con la carcinogénesis, por lo que solo se tomó en cuentas aquellas vías canónicas involucradas en la proliferación celular, apoptosis, adhesión, metástasis, diferenciación celular y relacionadas con la respuesta del sistema inmune.

La vía de señalización que mostro un mayor número de genes diferencialmente expresados por la coinfección y el genotipo 18 fue en las vías que regulan la pluripotencialidad de células madre, donde se encontró que los genes *Nanog* y *Pax6* estaban sobreexpresados , mientras que el gen *SMAD`s* se encontraban subexpresandose. En la vía de RAS encontramos que *Mst1* e *IKK* se encontraban sobreexpresados, en esta vía el único gen subexpresado fue *PAK*. En la vía de la citotoxicidad mediada por células natural killer solo encontramos genes sobreexpresados, los cuales fueron: *KIR3DL1*, *KIR3DL2* , *KIR2DS* y *KIR2DL*. En la vía de señalización de las MAP cinasas se encontró solamente 3 genes subexpresados: *G12*, *IKK* y *MST1/2*. En la vía de la apoptosis, los genes *IKK* e *AIP* se encontraron subexpresados, mientras que *HRK* se encontró sobreexpresado. La vía de señalización de ERBB se encontraba alterada por la sobreexpresión de *NRG2* y *PAK*. En la vía PI3K-AKT se encontró a *GPCR* sobreexpresado, mientras que *IKK* estaba sobreexpresado. La vía de mTOR solo poseía un único gen alterado; *GATOR2*, el cual se encontraba subexpresado. La vía de presentación y procesamiento de antígenos presentaba solamente un gen alterado; el cual fue *KIR* , que se encontraba sobreexpresado (Ver Figura 16).

En el perfil de expresión del genotipo 45 vs el genotipo 18 la vía de señalización con el mayor número de genes diferencialmente expresados fue la vía de PI3K-AKT, encontrándose 7 genes subexpresados: *GF*, *CITOKINE*, *GPCR*, *PP2A*, *SGK*, *PEPCK* y *CREB*, mientras que los dos genes sobreexpresados fueron *RTK* y *SYK*. La segunda vía con el mayor número de genes alterados fue la vía de la

citotoxicidad mediada por células natural killers, de manera interesante se encontró que esta vía poseía los mismos genes alterados que la comparación entre el genotipo 45 vs 18; sin embargo la expresión de estos genes se encontraba invertida, en este caso los genes KIR3DL1, KIR3DL2, KIR2DL y KIR2DS se encontraban subexpresados, en esta vía el único gen que presentó una sobreexpresión fue SYK. La vía de Ras presentó 3 genes subexpresados; GBY, GF y PLA, mientras que los genes sobreexpresados fueron RTK, CAM Y JNK. La vía de las MAP cinasas presentaba alteración en los genes JNK, c-fos y G12, los cuales se encuentran sobreexpresados, mientras que FGF mostró una subexpresión. La vía de la apoptosis solo presentaba dos genes alterados, los cuales se encontraban sobreexpresados, estos genes son: AP1 y JNK. La vía de JAK-STAT solo presentó dos citocinas subexpresadas, mientras que la vía de ERBB solo poseía a JNK sobreexpresado. Al igual que en la comparación entre la coinfección y el genotipo 18 se encontró que la vía de presentación y procesamiento de antígenos presentaba al gen KIR subexpresado. Solo un gen presentaba alteración en la vía de la regulación de la pluripotencialidad de las células madre; el cual fue Lefty2, que presentaba una subexpresión (Ver Figura 16).

La comparación entre el genotipo 16 y el genotipo 18 fue la que presentó la menor cantidad de vías de señalización alteradas, lo cual está relacionado con el pequeño número de genes diferencialmente expresados comparados con las restantes comparaciones. En esta comparación solo se obtuvo información de 4 vías relacionadas con la carcinogénesis, de las cuales cada vía solo tuvo un gen alterado e interesantemente solo se encontraron genes subexpresados. Estos genes fueron CACN, NRG2, CITOCINAS, e IFN, los cuales estaban alterados en las vías MAPK, ERBB, PI3K-AKT y citotoxicidad mediada por células natural killers respectivamente (Ver Figura 16).

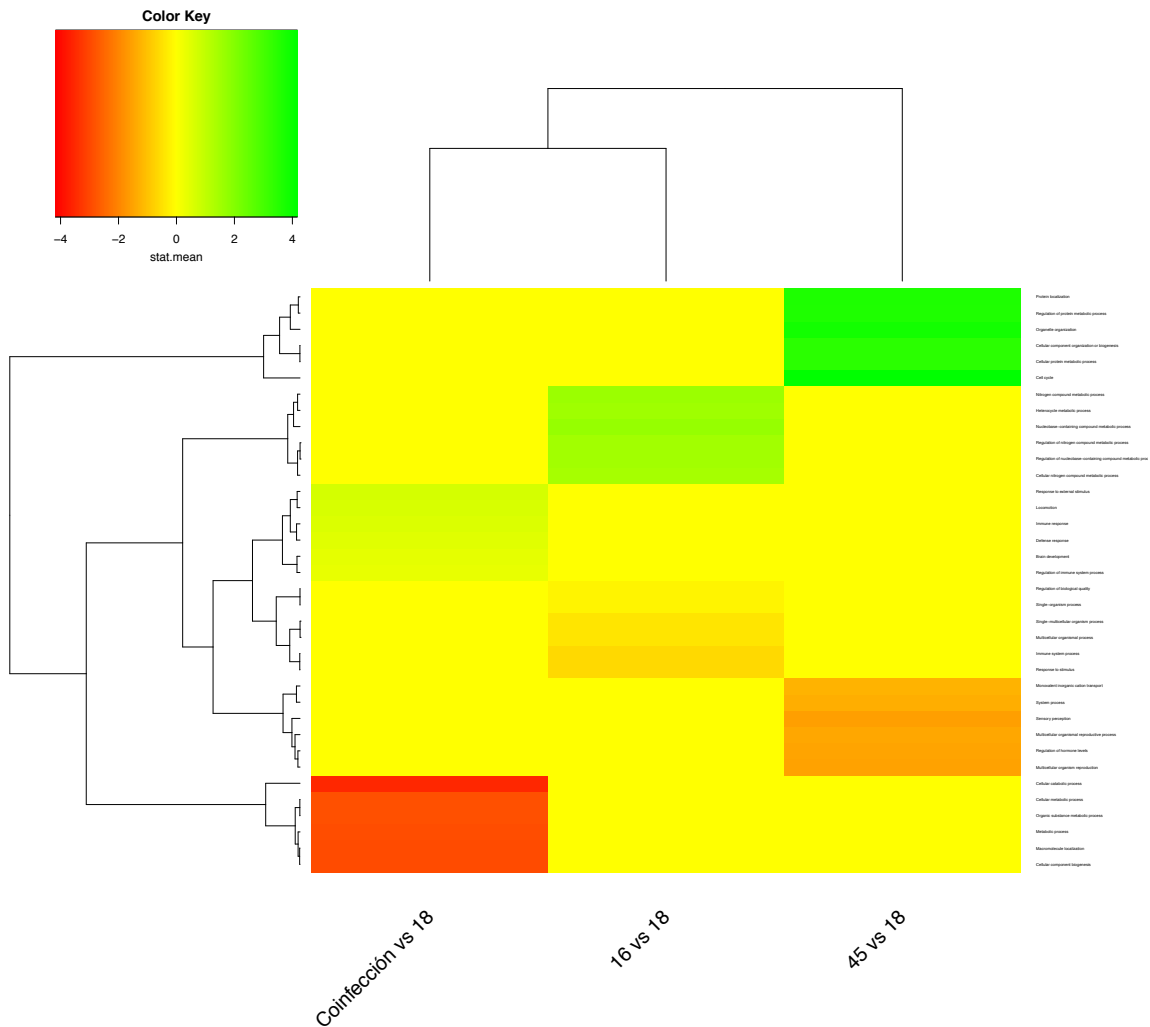


Figura 17. Análisis funcional de los términos biológicos GO enriquecidos de los genes diferencialmente expresados.

Se observa un *heatmap* que denota magnitud de los cambios en la expresión de los genes en los procesos biológicos GO. En color rojo se muestran los términos que presentan una subexpresión y en verde los términos que se encuentran sobreexpresados.

Después de obtener las vías de señalización alteradas en cada comparación se procedió a realizar un análisis de enriquecimiento de términos de ontología de genes (GO). Una anotación de ontología de genes es una base de datos en la que se encuentran asociados u genes con sus respectivos términos en las ontologías, estos términos están basados en tres características principales de la biología

celular; como lo son la función molecular, los procesos biológicos y los componentes celulares, por lo general estas relaciones están generadas por modelos predictivos o de forma manual por un curador (Yon Rhee, Wood, Dolinski, & Draghici, 2008). En estas ontologías los genes se asocian con todos los términos posibles que reflejen toda la información que se conoce acerca de los genes, entre las ventajas se encuentran la capacidad de inferir que tan relacionados se encuentran los términos de una base de datos mediante la similitud que estos tienen con los términos que son asociados a los genes. Para hacer el análisis de enriquecimiento de términos GO se utilizaron los genes diferencialmente expresados y significativos de cada comparación, los cuales fueron introducidos en la paquetería *gage*. En la figura 17 se muestran los resultados del análisis, en la gráfica de calor se representa el estadístico *stat.mean* de los términos biológicos de cada comparación, el cual es un valor absoluto que representa la magnitud de los cambios en los genes introducidos, además de que muestra la dirección de los cambios, siendo una sobreexpresión cuando el valor es positivo y una subexpresión cuando el valor es negativo. Los resultados de este análisis muestran que los genes de cada comparación tienen procesos biológicos alterados distintos, donde ninguno de ellos es compartido.

En el caso de la comparación entre las coinfecciones y el genotipo 18 se observó una subexpresión en los términos biológicos biogénesis de componentes celulares, localización de macromoléculas, procesos metabólicos, procesamiento de sustancias metabólicas orgánicas, procesos metabólicos celulares y procesos celulares catabólicos; mientras que existe una ligera sobreexpresión en términos como la regulación del sistema inmune, desarrollo cerebral, respuesta inmune, locomoción y respuesta a estímulos externos (Ver Figura 17).

Los términos biológicos alterados entre las pacientes con infección por el genotipo 16 en comparación con el genotipo 18 fue una subexpresión en la regulación de la calidad biológica, procesos de organismos multicelulares, procesos de respuesta inmune y respuesta a estímulos; mientras que los términos biológicos

sobreexpresados fueron: procesamiento metabólico del nitrógeno, procesos metabólicos heterocíclicos y metabolismo de compuestos con nucleobases (Ver Figura 17).

En la comparación entre el genotipo 45 y el genotipo 18 se encontraron términos biológicos sobreexpresados relacionados con localización de proteínas, regulación del metabolismo de proteínas, organización de los organelos, biogénesis de componentes celulares, metabolismo de proteínas celulares y ciclo celular; mientras que se halló una subexpresión en el transporte monovalente de cationes inorgánicos, procesamiento de sistemas, percepción sensorial, procesos reproductivos de organismos multicelulares, regulación de niveles hormonales y reproducción de organismos multicelulares (Ver Figura 17).

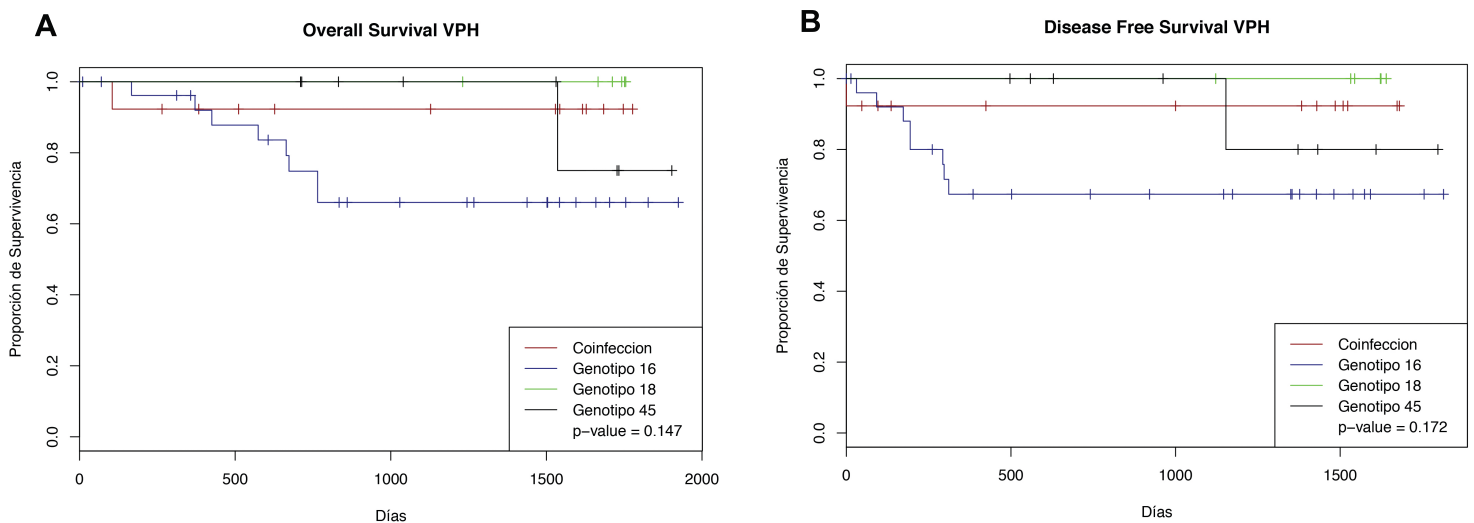


Figura 18. Análisis Kaplan-Meier de supervivencia.

Supervivencia media de las pacientes de acuerdo al genotipo viral que poseen (A) y supervivencia libre de enfermedad (B). En ambas graficas la prueba *log-rank* muestra que no existe una diferencia significativa en la supervivencia de las pacientes.

Finalmente se realizó un análisis de supervivencia de las pacientes de acuerdo a su genotipo viral mediante la estimación de Kaplan-Meier utilizando la supervivencia general de las pacientes, la cual es la diferencia en días entre la fecha de la fecha de defunción de la paciente o último monitoreo y la fecha de diagnóstico por el instituto, además también se utilizó la supervivencia libre enfermedad, obtenida mediante la resta entre la fecha de diagnóstico de recurrencia y la fecha de término de la braquiterapia. Para ambos análisis se compararon las curvas de supervivencia de cada genotipo mediante la prueba de *log-rank*. Los resultados obtenidos muestran que no hay una diferencia estadísticamente significativa en la supervivencia de las pacientes de acuerdo a su genotipo viral (Ver Figura 18).

Discusión

La identificación de perfiles de expresión usando microarreglos ha demostrado ser una herramienta poderosa para la búsqueda de genes que se encuentren diferencialmente expresados en diferentes enfermedades, las ventajas de esta tecnología se aprecian principalmente en el estudio de cáncer, donde es de gran ayuda la identificación de diferencias génicas entre tumores. En este trabajo analizamos los perfiles de expresión de 63 tumores de cáncer cérvico uterino localmente avanzado, los cuales fueron clasificados de acuerdo al genotipo viral que poseían, con el objetivo de identificar los genes diferencialmente expresados que estuvieran relacionados al genotipo viral. Este es el primer trabajo en cáncer cérvico uterino donde se utilizan microarreglos para ver la diferencia y el impacto que tiene la infección entre más de dos diferentes genotipos virales.

Una explicación de las diferencias en los perfiles de expresión asociados al genotipo viral es la diferencia en sus proteínas virales. Anteriormente ya se ha demostrado que alrededor del 4% de la variación en la secuencia entre genotipos de VPH tipo 16 pertenecientes a la misma especie se da entre los 8 genes del virus, además se ha reportado que el 9.9% de las posiciones de los aminoácidos son variables; donde la mayor proporción de variaciones sinónimas y no sinónimas se da en los genes E2 y E7. Variaciones similares en los oncogenes E6 y E7 también han sido reportadas entre los genotipos 18 y 45 (Chen et al., 2014; Garbuglia et al., 2007). La mayoría de estudios que reportan variaciones en las secuencias del VPH se centran solamente en el genotipo 16, esto se debe a la prevalencia y el riesgo que conlleva la infección por este genotipo. Estudios epidemiológicos en países como Japón, Tailandia y el Reino Unido muestran que la variante molecular del genotipo 16 se encuentra relacionada con el aumento o disminución en el riesgo de desarrollar neoplasias malignas, donde las diferencias en entre estas variantes moleculares pueden ser tan simples como una mutación

puntual en el gen E6 que llevará a la sustitución de un aminoácido por otro (Ishimi et al., 2003; Matsumoto et al., 2011). En el caso del genotipo 18 se ha reportado que la variabilidad en el gen E2 promueve la existencia de un subtipo molecular con un potencial oncogénico drásticamente disminuido (Hecht, Kadish, Jiang, & Burk, 1995). En el caso de la proteína viral E6, esta posee un dominio de zinc unido mediante una región compuesta de 35 aminoácidos. Además, contiene un dominio de unión PDZ que facilita el acople e interacción con proteínas que presenten un dominio PDZ en su estructura. Ambos dedos de zinc están compuestos por un par de motivos CXXC, mientras que el motivo PDZ contiene un submotivo XTXV/L (Wallace & Galloway, 2015). La principal función de estos motivos es permitir la unión de la proteína E6 a una gran cantidad de proteínas en el hospedero, por lo que las variaciones en esta proteína en los diferentes genotipos explicarían los diferentes genes alterados y perfiles de expresión obtenidos.

El genotipo 18 y el genotipo 16 son los tipos virales que más alto riesgo tienen de promover y facilitar el desarrollo de cáncer, además de ser los más frecuentes. Los pocos genes diferencialmente expresados en la comparación entre el genotipo 16 contra el genotipo 18 puede ser explicada por la similitud en la afinidad de los oncogenes de ambas variantes, ya que existen reportes mostrando que la región LCR de ambos genotipos posee una mayor afinidad y actividad transcripcional que otros genotipos (Sichero, Franco, & Villa, 2005). Una de las implicaciones biológicas de alteraciones en esta región del virus es un cambio en la afinidad de los factores de transcripción virales, por lo que se presume que los oncogenes de estos genotipos tienen una mayor similitud en las proteínas diana sobre las que actúan.

Se sabe que aproximadamente entre el 20 y el 40% de las mujeres que son positivas a VPH poseen al menos dos genotipos virales (Schmitt et al., 2010), aunque hasta el momento no se ha podido relacionar la presencia de coinfecciones con un mayor riesgo de desarrollar cáncer. En el presente trabajo demostramos que el perfil de expresión de las pacientes se ve modificado en comparación con la infección por los demás genotipos, aunque cabe destacar que

en las coinfecciones, salvo en una paciente, había presencia de por lo menos un genotipo viral de alto riesgo, principalmente el genotipo 16 y el 18. Anteriormente se ha reportado que durante las coinfecciones con los genotipos 16, 31, 33, 35, 52, 58 y 57 la carga viral del genotipo 16 se ve drásticamente disminuida en comparación con los demás genotipos, efecto que sucede de igual manera durante las coinfecciones con los genotipos 18, 39, 45, 59, 68 y 70, donde hay un menor número de viriones, y por ende una menor replicación del genotipo 18 (Mejlhede, Pedersen, Frisch, & Fomsgaard, 2010); este fenómeno podría explicar la diferencia en el perfil de expresión que observamos en las pacientes con coinfecciones cuando se les comparaba con las pacientes con infección por el genotipo 18, por lo que el perfil de expresión se vería afectado principalmente por la expresión de las proteínas virales presentes en los genotipos que no son tan comunes.

El mayor número de genes diferencialmente expresados se obtuvo en la comparación entre los tumores con infección por el genotipo 45 y los tumores infectados con el genotipo 18, este resultado es de gran importancia ya que el genotipo 45 posee solamente un 25% de diferencia en la homología con el genotipo 18 (De Villiers, Fauquet, Broker, Bernard, & Zur Hausen, 2004); lo que hace a los dos genotipos filogenéticamente cercanos, por lo que las diferencias en la expresión de los 679 genes diferencialmente expresados están dadas por la mínima diferencia en la secuencia entre estos dos genotipos, una de estas pequeñas diferencias pudiera ser los cambios en los aminoácidos en los ORF's de las proteínas E6 y E7 del genotipo 45, ya que se ha reportado que han reportado frecuencias de variaciones de aminoácidos de hasta el 20% en estas proteínas del genotipo 45 (Chen et al., 2014), que si bien no se encuentran en las regiones críticas de E6 y E7, si podrían explicar las diferencias biológicas en los perfiles de expresión.

En general se observó que el perfil de expresión inducido por el genotipo 18 siempre presentaba diferencias cuando se le comparaba con los perfiles de

expresión inducidos por los restantes genotipos analizados en este estudio, lo que indica la infección el genotipo 18 induce un perfil de expresión con una mayor cantidad de variaciones únicas que no se presentan en la infección por los restantes genotipos. Estas variaciones podrían explicar el por qué estudios anteriores predicen una menor supervivencia, prognosis y una rápida progresión a un carcinoma progresivo durante la infección por este genotipo (Nakagawa et al., 1996). En nuestros resultados se observó que las pacientes infectadas por el genotipo 16 presentaron un mayor número de fallecimientos ($n=6$) en comparación con el genotipo 18 ($n=1$), sin embargo, los análisis de supervivencia no mostraron una diferencia estadísticamente significativa en la supervivencia de las pacientes, estos resultados pueden ser explicados por el pequeño número de pacientes fallecidas durante el tiempo de estudio del experimento.

Estudios anteriores de VPH en cáncer cérvico uterino han demostrado que los principales procesos biológicos alterados por la infección de VPH están involucradas con la progresión del ciclo celular, reparación del DNA, apoptosis y diferenciación celular (Buitrago-Pérez, Garaulet, Vázquez-Carballo, Paramio, & García-Escudero, 2009). En este estudio se encontraron genes relacionados con estos procesos alterados, sin embargo, también se observó una gran cantidad de genes diferencialmente expresados que se encuentran relacionados con la respuesta inmune a la infección.

En nuestros resultados, una de las vías de señalización con más alteraciones fue la inducción de citotoxicidad por células natural killers, donde se observó una mayor desregulación de genes durante la infección por el genotipo 18, en estos casos había una sobreexpresión de KIR2DL3, KIR2DS1, KIR2DS2, IFNA13, KIR3DL2, KIR2DS4 e IFNA21. En 2014, Rizzo y colaboradores reportaron una relación en la expresión de KIR2DL2 y KIR2DL3 en la infección de pacientes con tipos de VPH de alto riesgo en 33 pacientes con cáncer, relacionando esta

expresión con la transformación y desarrollo de lesiones neoplásicas (Rizzo et al., 2014). Nuestros datos confirmarían los hallazgos de Rizzo e indicarían que esta sobreexpresión promueve la inhibición de las células natural killers, lo que facilitaría que las células infectadas evadan la detección del sistema inmune. Estos resultados se reafirman con el análisis de enriquecimiento, el cual muestra la presencia de una sobreexpresión en los procesos biológicos relacionados con el sistema inmune. Se presume que si bien, todos los genotipos virales poseen y activan mecanismos para evadir el sistema inmune, no todos son tan eficaces ni lo alteran de igual forma. En este caso, es posible que las proteínas virales del genotipo 18 posean una mayor capacidad inmunoevasora.

En total se encontraron cinco genes diferencialmente expresados compartidos entre los diferentes genotipos analizados. De estos cinco genes compartidos, cuatro han sido relacionados con al menos un tipo de cáncer, mientras que PPID ya se ha reportado como diferencialmente expresado en pacientes con cáncer de cabeza y cuello positivos a VPH (Schlecht et al., 2007), en este estudio solo se reportan casos positivos a los principales virus pertenecientes al grupo de alto riesgo, el genotipo 16 y el genotipo 18, por lo que nuestros datos sugieren que este gen es uno de los principales blancos canónicos de al menos varios genotipos virales que pertenecen a este grupo filogénico.

Conclusiones

- El genotipo viral 18 mostro una mayor alteración de genes relacionados con la evasión del sistema inmune.
- Las diferencias en la secuencia de nucleótidos entre el genotipo 45 y el 18 promueve el mayor número de genes diferencialmente expresados en el perfil de expresión.
- El genotipo viral no está relacionado con la supervivencia de las pacientes ni con el tiempo libre de enfermedad.
- Existen diferencias en el perfil de expresión en las pacientes de cáncer cérvico uterino dependiendo del genotipo viral de VPH que posean, las cuales se presume pueden deberse a las variaciones en la secuencia de nucleótidos de los oncogenes virales.

Anexos

Anexo A

GENES SUBEXPRESADOS EN EL GENOTIPO 16 y GENOTIPO 18

SEQ_ID	Símbolo del Gen	Nombre del gen	logFC
BC067430	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	-3.156410655
NM_002407	SCGB2A1	secretoglobin, family 2A, member 1	-3.056968509
NM_000716	C4BPB	complement component 4 binding protein, beta	-2.937943288
BC022895	MS4A8	membrane-spanning 4-domains, subfamily A, member 8	-2.831534973
AY529122	UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	-2.790705225
CR749812	PSG8	pregnancy specific beta-1-glycoprotein 8	-2.788361544
NM_133181	EPS8L3	EPS8-like 3	-2.672363275
AK128680	HMCN2	hemicentin 2	-2.544149595
NM_021139	UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	-2.522004017
NM_001039029	LRTM2	leucine-rich repeats and transmembrane domains 2	-2.442761206
NM_000032	ALAS2	5'-aminolevulinate synthase 2	-2.434168695
NM_000715	C4BPA	complement component 4 binding protein, alpha	-2.423784254
NM_024526	EPS8L3	EPS8-like 3	-2.315194387
NM_020707	HHATL	hedgehog acyltransferase-like	-2.295836097
NM_002173	IFNA16	interferon, alpha 16	-2.253146359
AK130404	LOC100131174	uncharacterized LOC100131174	-2.202769364
NM_004098	EMX2	empty spiracles homeobox 2	-2.154788573
NM_006512	SAA4	serum amyloid A4, constitutive	-2.076821893
NM_005073	SLC15A1	solute carrier family 15 (oligopeptide transporter), member 1	-2.035558419
NM_199228	THPO	thrombopoietin	-1.98184416
BC074843	MS4A2	membrane-spanning 4-domains, subfamily A, member 2	-1.975125697
BC025711	NKX2-5	NK2 homeobox 5	-1.955291633
AK000249	GPRC5C	G protein-coupled receptor, class C, group 5, member C	-1.927183791
NM_001018059	NPIP15	nuclear pore complex interacting protein family, member B15	-1.919203608
NM_001003693	LY6G6F	lymphocyte antigen 6 complex, locus G6F	-1.865009957
NM_000718	CACNA1B	calcium channel, voltage-dependent, N type, alpha 1B subunit	-1.860544429
NM_004883	NRG2	neuregulin 2	-1.855881765
BC030583	DNAI1	dynein, axonemal, intermediate chain 1	-1.81923175
NM_022912	REEP1	receptor accessory protein 1	-1.809071144

AK131211	TMPRSS7	transmembrane protease, serine 7	-1.793857601
BC107153	OTC	ornithine carbamoyltransferase	-1.791426152
NM_001038493	DLX1	distal-less homeobox 1	-1.741619096
AK127189	CCDC108	coiled-coil domain containing 108	-1.734645805
AK075040	HES6	hes family bHLH transcription factor 6	-1.732805025
AK125136	LOC100133306	uncharacterized LOC100133306	-1.698759044
NM_000742	CHRNA2	cholinergic receptor, nicotinic, alpha 2 (neuronal)	-1.677490693
NM_017570	OPLAH	5-oxoprolinase (ATP-hydrolysing)	-1.653962659
NM_005250	FOXL1	forkhead box L1	-1.640995063
NM_012309	SHANK2	SH3 and multiple ankyrin repeat domains 2	-1.630077252
BC031232	CDH22	cadherin 22, type 2	-1.627389441
BC036793	SCN2B	sodium channel, voltage gated, type II beta subunit	-1.608910577
NM_004164	RBP2	retinol binding protein 2, cellular	-1.60270346
AY692425	ADAMTS6	ADAM metalloproteinase with thrombospondin type 1 motif, 6	-1.594778161
BC002824	SOX10	SRY (sex determining region Y)-box 10	-1.566445279
BC106906	FAM117B	family with sequence similarity 117, member B	-1.442984409
BC032244	REXO1	REX1, RNA exonuclease 1 homolog	-1.424316207
BC093825	CARD6	caspase recruitment domain family, member 6	-1.382294106
AB088047	MARK4	MAP/microtubule affinity-regulating kinase 4	-1.170816958

Anexo B

GENES SOBREPRESADOS EN EL GENOTIPO 16 y GENOTIPO 18

SEQ_ID	Símbolo del Gen	Nombre del gen	logFC
BC015631	SLC5A6	solute carrier family 5 (sodium/multivitamin and iodide cotransporter), member 6	1.284496282
NM_004069	AP2S1	adaptor-related protein complex 2, sigma 1 subunit	1.332994939
NM_033274	ADAM19	ADAM metallopeptidase domain 19	1.342024014
BC012589	CXCL13	chemokine (C-X-C motif) ligand 13	1.526374887
NM_015633	FGFR1OP2	FGFR1 oncogene partner 2	1.539087625
NM_015534	ZZZ3	zinc finger, ZZ-type containing 3	1.614671259
NM_017426	NUP54	nucleoporin 54kDa	1.641968781
NM_032782	HAVCR2	hepatitis A virus cellular receptor 2	1.660802499
NM_020799	STAMBPL1	STAM binding protein-like 1	1.684275625
AY167994	ST3GAL3	ST3 beta-galactoside alpha-2,3-sialyltransferase 3	1.770324578
NM_001025	RPS23	ribosomal protein S23	1.774331437
BC035023	OSCAR	osteoclast associated, immunoglobulin-like receptor	1.841238364
NM_020647	JPH1	junctionophilin 1	1.91687764
NM_002061	GCLM	glutamate-cysteine ligase, modifier subunit	1.922287974
BC034810	ZNF550	zinc finger protein 550	1.930424657
BC071898	FJX1	four jointed box 1	1.938081027
NM_001004417	FMNL2	formin-like 2	1.948763145
BC093792	TMX3	thioredoxin-related transmembrane protein 3	1.975284148
NM_004843	IL27RA	interleukin 27 receptor, alpha	1.988933437
NM_014947	FOXJ3	forkhead box J3	2.053159197
AF273049	ZNF638	zinc finger protein 638	2.062361562
NM_001010919	FAM26F	family with sequence similarity 26, member F	2.146629999
NM_032496	ARHGAP9	Rho GTPase activating protein 9	2.328701732
NM_144594	GTSF1	gametocyte specific factor 1	2.337821902
NM_001011543	MAGEA10	melanoma antigen family A10	2.388234965
NM_005038	PPID	peptidylprolyl isomerase D	2.501957528
NM_004929	CALB1	calbindin 1, 28kDa	2.590339336
NM_005362	MAGEA3	melanoma antigen family A3	2.734401484

Anexo C

GENES SOBREPRESADOS EN LAS COINFECCIONES Y GENOTIPO 18

SEQ_ID	Símbolo del gen	Nombre del gen	logFC
AY341951	FAM138D	family with sequence similarity 138, member D	4.178833605
BC110891	GSTA1	glutathione S-transferase alpha 1	3.871957744
BC008608	SMOC1	SPARC related modular calcium binding 1	3.860941096
AF346307	FAM138B	family with sequence similarity 138, member B	3.727853165
BC008405	PSG4	pregnancy specific beta-1-glycoprotein 4	3.714423903
NM_174962	SSX9	synovial sarcoma, X breakpoint 9	3.692214485
NM_032785	AGBL4	ATP/GTP binding protein-like 4	3.686885706
NM_001005484	OR4F5	olfactory receptor, family 4, subfamily F, member 5	3.662567527
NM_001005283	OR9Q2	olfactory receptor, family 9, subfamily Q, member 2	3.612348726
AK094547	SLC7A14	solute carrier family 7, member 14	3.580650867
NM_001004688	OR2M2	olfactory receptor, family 2, subfamily M, member 2	3.553429991
CR749812	PSG8	pregnancy specific beta-1-glycoprotein 8	3.522400136
NM_001004695	OR2T33	olfactory receptor, family 2, subfamily T, member 33	3.514389063
NM_000716	C4BPB	complement component 4 binding protein, beta	3.505570859
AY792621	OR4F13P	olfactory receptor, family 4, subfamily F, member 13 pseudogene	3.483231037
NM_175067	TAAR6	trace amine associated receptor 6	3.455847011
NM_001034832	SSX4B	synovial sarcoma, X breakpoint 4B	3.447173774
BC110494	PIGR	polymeric immunoglobulin receptor	3.419482112
NM_001005323	OR5AK2	olfactory receptor, family 5, subfamily AK, member 2	3.407403368
NM_021016	PSG3	pregnancy specific beta-1-glycoprotein 3	3.393350186
NM_001037732	DEFB128	defensin, beta 128	3.392475062
NM_006737	KIR3DL2	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2	3.390489465
NM_153444	OR5P2	olfactory receptor, family 5, subfamily P, member 2	3.362342786
NM_001004703	OR4C46	olfactory receptor, family 4, subfamily C, member 46	3.356330007
BC017021	MEOX2	mesenchyme homeobox 2	3.350361327
NM_020178	CA10	carbonic anhydrase X	3.335858081
NM_153445	OR5P3	olfactory receptor, family 5, subfamily P, member 3	3.324202304
BC069129	XAGE5	X antigen family, member 5	3.315865427
NM_001004724	OR4N5	olfactory receptor, family 4, subfamily N, member 5	3.294174382
NM_001005492	OR5J2	olfactory receptor, family 5, subfamily J, member 2	3.285034046
AB061325	PPARGC1A	peroxisome proliferator-activated receptor gamma, coactivator 1 alpha	3.234931933

NM_001003954	ANXA13	annexin A13	3.223488386
BC069996	LPAR4	lysophosphatidic acid receptor 4	3.184058224
NM_020535	KIR2DL5A	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 5A	3.171557837
NM_001005286	OR6F1	olfactory receptor, family 6, subfamily F, member 1	3.167854907
NM_016945	TAS2R16	taste receptor, type 2, member 16	3.149914454
NM_001038706	FER1L6-AS1	FER1L6 antisense RNA 1	3.148107805
NM_014511	KIR2DL3	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3	3.139517366
NM_001017364	C4BPB	complement component 4 binding protein, beta	3.130972913
BC110058	LRRTM4	leucine rich repeat transmembrane neuronal 4	3.085025893
NM_145740	GSTA1	glutathione S-transferase alpha 1	3.075042217
NM_000846	GSTA2	glutathione S-transferase alpha 2	3.069434938
BC101977	KIR2DS4	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 4	3.053769965
BC022895	MS4A8	membrane-spanning 4-domains, subfamily A, member 8	3.043072048
NM_012313	KIR2DS3	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 3	3.041057215
NM_006615	CAPN9	calpain 9	3.028159607
AY529122	UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	3.003981374
NM_000847	GSTA3	glutathione S-transferase alpha 3	3.003539006
NM_002747	MAPK4	mitogen-activated protein kinase 4	2.994023509
NM_012312	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	2.982678463
L76668	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	2.974990574
BC095538	LPAR4	lysophosphatidic acid receptor 4	2.952636052
U33328	KIR3DL1	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 1	2.933790312
NM_021139	UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	2.926708064
NM_002407	SCGB2A1	secretoglobin, family 2A, member 1	2.919962481
AJ002104	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	2.886969395
NM_001034852	SMOC1	SPARC related modular calcium binding 1	2.870842245
AJ002102	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	2.854905353
BC069360	XCL2	chemokine (C motif) ligand 2	2.85479969
NM_203451	SERTM1	serine-rich and transmembrane domain containing 1	2.853218484
NM_173643	ZNF663P	zinc finger protein 663, pseudogene	2.850851462
NM_001005471	OR2T6	olfactory receptor, family 2, subfamily T, member 6	2.844606814
NM_001039029	LRTM2	leucine-rich repeats and transmembrane domains 2	2.821378905

NM_207474	BCRP2	breakpoint cluster region pseudogene 2	2.812402092
NM_012314	KIR2DS4	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 4	2.811388628
L76667	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	2.795504749
BC069428	CIB3	calcium and integrin binding family member 3	2.763756297
NM_000280	PAX6	paired box 6	2.740940498
BC074752	KCNJ1	potassium channel, inwardly rectifying subfamily J, member 1	2.740291054
NM_002995	XCL1	chemokine (C motif) ligand 1	2.73762885
NM_005172	ATOH1	atonal bHLH transcription factor 1	2.706029279
BC069817	XCL1	chemokine (C motif) ligand 1	2.704186203
NM_000715	C4BPA	complement component 4 binding protein, alpha	2.69649759
NM_018702	ADARB2	adenosine deaminase, RNA-specific, B2 (non-functional)	2.683613443
BC098267	LRRTM4	leucine rich repeat transmembrane neuronal 4	2.666962886
BC047765	ABCA8	ATP-binding cassette, sub-family A (ABC1), member 8	2.646124321
NM_153443	KIR3DL3	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 3	2.645252279
NM_020769	RGAG1	retrotransposon gag domain containing 1	2.598181702
AK123428	PLXNA4	plexin A4	2.587931513
BC035626	GFI1B	growth factor independent 1B transcription repressor	2.584114981
BC112166	UTS2B	urotensin 2B	2.56705301
AY366243	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	2.560558225
AK127178	LOC101929106	uncharacterized LOC101929106	2.559407099
BC024179	PAK7	p21 protein (Cdc42/Rac)-activated kinase 7	2.549354802
NM_020747	ZNF608	zinc finger protein 608	2.521776621
NM_001007470	TRPM3	transient receptor potential cation channel, subfamily M, member 3	2.509219383
BC096135	TNP2	transition protein 2 (during histone to protamine replacement)	2.501944509
NM_012309	SHANK2	SH3 and multiple ankyrin repeat domains 2	2.484396623
CR749624	ZNF608	zinc finger protein 608	2.462980822
AK097517	GOLGA6L1	golgin A6 family-like 1	2.457244526
NM_145234	CHRDL1	chordin-like 1	2.4562289
AK127151	CYP2G1P	cytochrome P450, family 2, subfamily G, polypeptide 1 pseudogene	2.454768186
BC034977	UBQLNL	ubiquilin-like	2.441358581
AK026597	CFAP43	cilia and flagella associated protein 43	2.431217097
NM_015868	KIR2DL3	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3	2.415228842
NM_152338	ZG16	zymogen granule protein 16	2.394797518
NM_013261	PPARGC1A	peroxisome proliferator-activated receptor gamma, coactivator 1 alpha	2.388963549
NM_001337	CX3CR1	chemokine (C-X3-C motif) receptor 1	2.369430609
NM_012360	OR1F1	olfactory receptor, family 1, subfamily F, member 1	2.354940738

BC069609	SPRY3	sprouty RTK signaling antagonist 3	2.33905539
NM_000036	AMPD1	adenosine monophosphate deaminase 1	2.28938691
BC029488	SPACA1	sperm acrosome associated 1	2.28896923
BC023152	GYG2	glycogenin 2	2.260821174
BC058010	CPS1	carbamoyl-phosphate synthase 1, mitochondrial	2.257167845
NM_207390	CLEC17A	C-type lectin domain family 17, member A	2.250301554
NM_004482	GALNT3	polypeptide N-acetylgalactosaminyltransferase 3	2.243929601
NM_016524	SYT17	synaptotagmin XVII	2.243396858
NM_004098	EMX2	empty spiracles homeobox 2	2.241820976
AK127189	CCDC108	coiled-coil domain containing 108	2.228836911
NM_001024613	FEZF1	FEZ family zinc finger 1	2.221082476
NM_024082	PRRG3	proline rich Gla (G-carboxyglutamic acid) 3 (transmembrane)	2.211880244
NM_173493	PASD1	PAS domain containing 1	2.202003675
AK127678	LOC100129322	uncharacterized LOC100129322	2.172678507
NM_000486	AQP2	aquaporin 2 (collecting duct)	2.155252626
NM_004883	NRG2	neuregulin 2	2.14335029
NM_138703	MAGEE2	melanoma antigen family E2	2.141018066
NM_000742	CHRNA2	cholinergic receptor, nicotinic, alpha 2 (neuronal)	2.126203955
AK075513	PLEKHG4B	pleckstrin homology domain containing, family G (with RhoGef domain)	2.125429382
AF202640	GPRC5B	G protein-coupled receptor, class C, group 5, member B	2.122299065
BC009975	WSCD1	WSC domain containing 1	2.111578775
NM_133457	COL26A1	collagen, type XXVI, alpha 1	2.105064324
AL832601	RIC3	RIC3 acetylcholine receptor chaperone	2.075289954
NM_003558	PIP5K1B	phosphatidylinositol-4-phosphate 5-kinase, type I, beta	2.068213834
NM_024847	TMC7	transmembrane channel-like 7	2.067906493
NM_003741	CHRD	chordin	2.03987796
NM_003806	HRK	harakiri, BCL2 interacting protein	2.005848033
BC111529	CMYA5	cardiomyopathy associated 5	1.981722919
NM_001038493	DLX1	distal-less homeobox 1	1.953391836
BC069807	NANOG	Nanog homeobox	1.939378457
NM_018073	TRIM68	tripartite motif containing 68	1.918473026
BC041961	C8orf34	chromosome 8 open reading frame 34	1.879957336
BC031232	CDH22	cadherin 22, type 2	1.861271712
BC065369	LOC100129534	small nuclear ribonucleoprotein polypeptide N pseudogene	1.838585802
AK055135	TOX2	TOX high mobility group box family member 2	1.795642191
NM_175901	FBXL19-AS1	FBXL19 antisense RNA 1 (head to head)	1.778699539
NM_001037	SCN1B	sodium channel, voltage gated, type I beta subunit	1.772461595
AY692425	ADAMTS6	ADAM metalloproteinase with thrombospondin type 1 motif, 6	1.771282601
NM_016157	TRO	trophinin	1.705816805
NM_173479	WDR88	WD repeat domain 88	1.592452218

Anexo D**GENES SUBEXPRESADOS EN LAS COINFECCIONES Y GENOTIPO 18**

SEQ_ID	Simbolo del gen	Nombre del gen	logFC
BC020864	CALB1	calbindin 1, 28kDa	-3.574370935
NM_014229	SLC6A11	solute carrier family 6 (neurotransmitter transporter), member 11	-3.08086244
BC009568	SYNGR3	synaptogyrin 3	-3.057607468
AB209790	AAMP	angio-associated, migratory cell protein	-3.056432036
NM_001039569	AP1S3	adaptor-related protein complex 1, sigma 3 subunit	-2.794810043
NM_001008660	PICALM	phosphatidylinositol binding clathrin assembly protein	-2.784595504
BC033637	SSX2IP	synovial sarcoma, X breakpoint 2 interacting protein	-2.778113542
NM_007026	DUSP14	dual specificity phosphatase 14	-2.767832347
NM_005038	PPID	peptidylprolyl isomerase D	-2.76456386
NM_001011543	MAGEA10	melanoma antigen family A10	-2.679500451
BC109372	MRPS35	mitochondrial ribosomal protein S35	-2.671048529
X82460	HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	-2.641096722
AK074748	ELOVL5	ELOVL fatty acid elongase 5	-2.557841702
BC110820	PHLDA1	pleckstrin homology-like domain, family A, member 1	-2.551086162
NM_001740	CALB2	calbindin 2	-2.516769728
CR933633	STK4	serine/threonine kinase 4	-2.509801105
BC093792	TMX3	thioredoxin-related transmembrane protein 3	-2.49797601
NM_004165	RRAD	Ras-related associated with diabetes	-2.470471282
BC062779	ABCA3	ATP-binding cassette, sub-family A (ABC1), member 3	-2.458916156
NM_003639	IKBKG	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma	-2.444407748
NM_025168	LRRC1	leucine rich repeat containing 1	-2.441440605
NM_021095	SLC5A6	solute carrier family 5 (sodium/multivitamin and iodide cotransporter), member 6	-2.436667842
NM_007353	GNA12	guanine nucleotide binding protein (G protein) alpha 12	-2.410346296
NM_014947	FOXJ3	forkhead box J3	-2.408431154
BC032544	IPP	intracisternal A particle-promoted polypeptide	-2.401757443
NM_022458	LMBR1	limb development membrane protein 1	-2.380232914

NM_004209	SYNGR3	synaptogyrin 3	-2.36096862
NM_000788	DCK	deoxycytidine kinase	-2.356455697
AL832139	A2ML1	alpha-2-macroglobulin-like 1	-2.354105042
NM_004760	STK17A	serine/threonine kinase 17a	-2.353547286
NM_005504	BCAT1	branched chain amino-acid transaminase 1, cytosolic	-2.335742235
BC067822	HIBCH	3-hydroxyisobutyryl-CoA hydrolase	-2.33443713
NM_005891	ACAT2	acetyl-CoA acetyltransferase 2	-2.330048012
BC039498	SLC39A6	solute carrier family 39 (zinc transporter), member 6	-2.315869977
NM_003385	VSNL1	visinin-like 1	-2.300361298
NM_018941	CLN8	ceroid-lipofuscinosis, neuronal 8	-2.279409273
AY823523	POT1	protection of telomeres 1	-2.278679668
NM_020198	CCDC47	coiled-coil domain containing 47	-2.26858056
AK131501	GLG1	golgi glycoprotein 1	-2.268379665
NM_001144	AMFR	autocrine motility factor receptor, E3 ubiquitin protein ligase	-2.264247924
BC051810	RELT	RELT tumor necrosis factor receptor	-2.261897683
NM_002417	MKI67	marker of proliferation Ki-67	-2.24912644
NM_133494	NEK7	NIMA-related kinase 7	-2.241789344
BC107769	LOC646358	DnaJ (Hsp40) homolog, subfamily B, member 14 pseudogene	-2.212670211
BC109089	ATG14	autophagy related 14	-2.208141548
BC032858	TTK	TTK protein kinase	-2.205944333
NM_032466	ASPH	aspartate beta-hydroxylase	-2.184098578
NM_018043	ANO1	anoctamin 1, calcium activated chloride channel	-2.166011111
NM_032552	DAB2IP	DAB2 interacting protein	-2.145747689
BC059782	LGALS2	lectin, galactoside-binding, soluble, 2	-2.144333648
NM_024345	DCAF10	DDB1 and CUL4 associated factor 10	-2.144165446
NM_001001419	SMAD5	SMAD family member 5	-2.138116514
NM_153371	LNX2	ligand of numb-protein X 2	-2.135345752
NM_031216	SEH1L	SEH1-like nucleoporin	-2.130601927
BC005338	CAPZA2	capping protein (actin filament) muscle Z-line, alpha 2	-2.129111632
NM_001006635	MTX2	metaxin 2	-2.127704307
NM_017958	PLEKHB2	pleckstrin homology domain containing, family B (evectins) member 2	-2.125544353
NM_033505	EPT1	ethanolaminephosphotransferase 1	-2.115192538
NM_006470	TRIM16	tripartite motif containing 16	-2.110392168

NM_018427	RRN3	RRN3 homolog, RNA polymerase I transcription factor	-2.10624743
BC007725	CLN8	ceroid-lipofuscinosis, neuronal 8	-2.093714267
NM_173177	C1D	C1D nuclear receptor corepressor	-2.090657942
AK090677	MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1	-2.084237729
NM_177966	PDE12	phosphodiesterase 12	-2.077677693
NM_015360	SKIV2L2	superkiller viralicidic activity 2-like 2 (S. cerevisiae)	-2.069221364
BC010942	ACAT1	acetyl-CoA acetyltransferase 1	-2.051737901
NM_002061	GCLM	glutamate-cysteine ligase, modifier subunit	-2.050839864
NM_005920	MEF2D	myocyte enhancer factor 2D	-2.049108031
AF093825	MOB4	MOB family member 4, phocein	-2.03617267
NM_004308	ARHGAP1	Rho GTPase activating protein 1	-2.003958106
BX647267	HMGB1	high mobility group box 1	-1.987796417
NM_014043	CHMP2B	charged multivesicular body protein 2B	-1.986604434
NM_018416	FOXJ2	forkhead box J2	-1.969016629
AF118070	SLC43A3	solute carrier family 43, member 3	-1.951781979
BC023247	TOR1AIP1	torsin A interacting protein 1	-1.945991003
NM_002069	GNAI1	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	-1.940473374
BC051695	FRMD8	FERM domain containing 8	-1.91167928
NM_006915	RP2	retinitis pigmentosa 2 (X-linked recessive)	-1.865691364
NM_152902	TIPRL	TOR signaling pathway regulator	-1.856171382
NM_024713	KATNBL1	katanin p80 subunit B-like 1	-1.843894283
NM_018000	MREG	melanoregulin	-1.83626023
BC010618	DTD2	D-tyrosyl-tRNA deacylase 2 (putative)	-1.825216509
BC034468	CMTR2	cap methyltransferase 2	-1.821855138
BC050695	PLEKHB2	pleckstrin homology domain containing, family B (evectins) member 2	-1.809315094
NM_017677	MTMR8	myotubularin related protein 8	-1.802656703
NM_152903	KBTBD6	kelch repeat and BTB (POZ) domain containing 6	-1.775470801
NM_003921	BCL10	B-cell CLL/lymphoma 10	-1.771934939
NM_138635	H2AFV	H2A histone family, member V	-1.75413828
BC015631	SLC5A6	solute carrier family 5 (sodium/multivitamin and iodide cotransporter), member 6	-1.708759266
BC036773	OPN3	opsin 3	-1.702490692
NM_014263	YME1L1	YME1-like 1 ATPase	-1.693964198
AF054589	MDFIC	MyoD family inhibitor domain containing	-1.645973004
NM_001001740	RFWD2	ring finger and WD repeat domain 2, E3 ubiquitin protein ligase	-1.645698619
NM_018343	RIOK2	RIO kinase 2	-1.619531482
BC017220	NGLY1	N-glycanase 1	-1.570659308
NM_020727	ZBTB21	zinc finger and BTB domain containing 21	-1.51597157

Anexo E

GENES SUBEXPRESADOS EN EL GENOTIPO 45 Y GENOTIPO 18

SEQ_ID	Simbolo del gen	Nombre del Gen	logFC
BC008405	PSG4	pregnancy specific beta-1-glycoprotein 4	-4.819313911
NM_001005501	OR4K2	olfactory receptor, family 4, subfamily K, member 2	-4.773906534
NM_172194	OR4Q3	olfactory receptor, family 4, subfamily Q, member 3	-4.708373645
NM_001332	CTNND2	catenin (cadherin-associated protein), delta 2	-4.4963236
BC001003	SSX1	synovial sarcoma, X breakpoint 1	-4.408074137
AY341951	FAM138D	family with sequence similarity 138, member D	-4.394538959
NM_001011719	ARSH	arylsulfatase family, member H	-4.372404
NM_001005270	OR4C12	olfactory receptor, family 4, subfamily C, member 12	-4.296423777
NM_001004715	OR4K17	olfactory receptor, family 4, subfamily K, member 17	-4.247616526
BC027895	REG1B	regenerating islet-derived 1 beta	-4.244033948
NM_001005280	OR10A7	olfactory receptor, family 10, subfamily A, member 7	-4.240636455
NM_001005281	OR6B1	olfactory receptor, family 6, subfamily B, member 1	-4.236678702
NM_001004473	OR10K1	olfactory receptor, family 10, subfamily K, member 1	-4.208789764
AK094547	SLC7A14	solute carrier family 7, member 14	-4.20369805
NM_020178	CA10	carbonic anhydrase X	-4.178550715
NM_003996	GPX5	glutathione peroxidase 5	-4.177121857
NM_153444	OR5P2	olfactory receptor, family 5, subfamily P, member 2	-4.165421479
NM_178466	BPIFA3	BPI fold containing family A, member 3	-4.154304537
NM_001004688	OR2M2	olfactory receptor, family 2, subfamily M, member 2	-4.151217622
NM_000936	PNLIP	pancreatic lipase	-4.145549092
NM_004316	ASCL1	achaete-scute family bHLH transcription factor 1	-4.108181388
BC069312	LGALS13	lectin, galactoside-binding, soluble, 13	-4.089904868

NM_001011717	NMS	neuromedin S	-4.079292647
BC096288	GCM1	glial cells missing homolog 1 (Drosophila)	-4.062828336
NM_199131	VAX1	ventral anterior homeobox 1	-4.049630325
NM_001005496	OR5D16	olfactory receptor, family 5, subfamily D, member 16	-4.042924538
BC069454	CELA1	chymotrypsin-like elastase family, member 1	-4.036255512
NM_024492	LPAL2	lipoprotein, Lp(a)-like 2, pseudogene	-4.024916587
NM_001004734	OR14I1	olfactory receptor, family 14, subfamily I, member 1	-4.024817444
NM_001005211	OR9I1	olfactory receptor, family 9, subfamily I, member 1	-4.017413915
BC029057	DAO	D-amino-acid oxidase	-4.004339011
NM_001034832	SSX4B	synovial sarcoma, X breakpoint 4B	-4.002844299
BC093992	TAS2R7	taste receptor, type 2, member 7	-3.99959765
NM_032785	AGBL4	ATP/GTP binding protein-like 4	-3.995871914
BC106932	AMBN	ameloblastin (enamel matrix protein)	-3.994029958
NM_001037668	DEFB107A	defensin, beta 107A	-3.984268486
NM_001001912	OR4E2	olfactory receptor, family 4, subfamily E, member 2	-3.983800525
NM_001004195	OR4F4	olfactory receptor, family 4, subfamily F, member 4	-3.981619717
AJ783421	DSCR8	Down syndrome critical region 8	-3.969228622
NM_013937	OR11A1	olfactory receptor, family 11, subfamily A, member 1	-3.968973643
BC069295	WFDC9	WAP four-disulfide core domain 9	-3.966370644
NM_080831	DEFB129	defensin, beta 129	-3.963424816
BC069128	BPIFA4P	BPI fold containing family A, member 4, pseudogene	-3.959950808
NM_001005512	OR4A47	olfactory receptor, family 4, subfamily A, member 47	-3.95250622
NM_017545	HAO1	hydroxyacid oxidase (glycolate oxidase) 1	-3.948156284
NM_005912	MC4R	melanocortin 4 receptor	-3.946437378
NM_001005214	LRRC52	leucine rich repeat containing 52	-3.934993433
BC009567	BAAT	bile acid CoA:amino acid N-acyltransferase	-3.928532285
NM_002909	REG1A	regenerating islet-derived 1 alpha	-3.921693412
NM_001004458	OR1S1	olfactory receptor, family 1, subfamily S, member 1 (gene/pseudogene)	-3.918651472
NM_001032412	ACTR3BP2	ACTR3B pseudogene 2	-3.915873054

NM_152291	MUC7	mucin 7, secreted	-3.909262176
NM_174962	SSX9	synovial sarcoma, X breakpoint 9	-3.907414769
NM_030657	LIM2	lens intrinsic membrane protein 2, 19kDa	-3.89653247
NM_001004328	ZNF705A	zinc finger protein 705A	-3.896055732
NM_006507	REG1B	regenerating islet-derived 1 beta	-3.888986287
NM_147198	WFDC9	WAP four-disulfide core domain 9	-3.883285808
BC110494	PIGR	polymeric immunoglobulin receptor	-3.882733558
NM_003007	SEMG1	semenogelin I	-3.881110103
NM_000192	TBX5	T-box 5	-3.873604617
NM_001005484	OR4F5	olfactory receptor, family 4, subfamily F, member 5	-3.866811851
NM_001004703	OR4C46	olfactory receptor, family 4, subfamily C, member 46	-3.850085284
NM_001031839	C8orf49	chromosome 8 open reading frame 49	-3.843631261
AF336106	KLK2	kallikrein-related peptidase 2	-3.836236462
NM_016945	TAS2R16	taste receptor, type 2, member 16	-3.828536417
BC039387	TRY2P	trypsinogen-like pseudogene	-3.804230666
NM_000957	PTGER3	prostaglandin E receptor 3 (subtype EP3)	-3.804095577
NM_033106	GALP	galanin-like peptide	-3.802351584
NM_203303	ZCCHC13	zinc finger, CCHC domain containing 13	-3.802136758
AL136545	TRPM3	transient receptor potential cation channel, subfamily M, member 3	-3.800337729
NM_001001709	C9orf170	chromosome 9 open reading frame 170	-3.794758635
NM_001001674	OR4F15	olfactory receptor, family 4, subfamily F, member 15	-3.793516927
NM_001039702	OLAH	oleoyl-ACP hydrolase	-3.792182904
NM_182511	CBLN2	cerebellin 2 precursor	-3.784337599
NM_174932	BPIFC	BPI fold containing family C	-3.770776813
BC067430	CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	-3.770110854
AF479699	DEFB119	defensin, beta 119	-3.769125792
NM_004088	DNTT	DNA nucleotidylexotransferase	-3.760617177
AK128835	LOC100287704	uncharacterized LOC100287704	-3.754728997
NM_199286	DPPA3	developmental pluripotency associated 3	-3.749733141
NM_173486	C2orf73	chromosome 2 open reading frame 73	-3.747453359
NM_001011878	DEFB121	defensin, beta 121	-3.745848054
NM_001010889	PRAMEF6	PRAME family member 6	-3.728486553
NM_001004755	OR51L1	olfactory receptor, family 51, subfamily L, member 1	-3.72626437

NM_001005172	OR52K2	olfactory receptor, family 52, subfamily K, member 2	-3.712434819
BC017021	MEOX2	mesenchyme homeobox 2	-3.712010178
NM_001005164	OR52E2	olfactory receptor, family 52, subfamily E, member 2	-3.709208314
NM_030589	CYP2A7	cytochrome P450, family 2, subfamily A, polypeptide 7	-3.707886988
NM_032588	TRIM63	tripartite motif containing 63, E3 ubiquitin protein ligase	-3.705633659
AK124766	MOBP	myelin-associated oligodendrocyte basic protein	-3.70465137
NM_182560	C14orf177	chromosome 14 open reading frame 177	-3.697313486
BC029353	DSCR8	Down syndrome critical region 8	-3.694801984
NM_178556	TRIML1	tripartite motif family-like 1	-3.694097576
NM_001443	FABP1	fatty acid binding protein 1, liver	-3.684113897
NM_198139	SEMG1	semenogelin I	-3.683729913
NM_001005492	OR5J2	olfactory receptor, family 5, subfamily J, member 2	-3.677186638
NM_002063	GLRA2	glycine receptor, alpha 2	-3.674577389
NM_030901	OR7A17	olfactory receptor, family 7, subfamily A, member 17	-3.670627071
NM_001003745	OR10A3	olfactory receptor, family 10, subfamily A, member 3	-3.664273693
NM_001004724	OR4N5	olfactory receptor, family 4, subfamily N, member 5	-3.66075081
BC110891	GSTA1	glutathione S-transferase alpha 1	-3.649414386
BC112209	TAAR5	trace amine associated receptor 5	-3.647117894
AK097656	TXNDC2	thioredoxin domain containing 2 (spermatozoa)	-3.646656043
AF159054	LOC51145	uncharacterized LOC51145	-3.640746017
NM_001002917	OR8D1	olfactory receptor, family 8, subfamily D, member 1	-3.640435243
BC069171	TAAR5	trace amine associated receptor 5	-3.639751123
BC112008	C5orf64	chromosome 5 open reading frame 64	-3.638982816
NM_003967	TAAR5	trace amine associated receptor 5	-3.638381219
NM_001005237	OR51G1	olfactory receptor, family 51, subfamily G, member 1 (gene/pseudogene)	-3.625222368
BC005279	CPA1	carboxypeptidase A1 (pancreatic)	-3.608574258
NM_020129	LGALS14	lectin, galactoside-binding, soluble, 14	-3.600705769
BC045626	PKNOX2	PBX/knotted 1 homeobox 2	-3.599704409

NM_145651	SCGB1C1	secretoglobin, family 1C, member 1	-3.597298218
NM_001039769	UG0898H09	uncharacterized LOC643763	-3.591077334
NM_138294	PATE1	prostate and testis expressed 1	-3.590888402
BC105047	SLC5A5	solute carrier family 5 (sodium/iodide cotransporter), member 5	-3.579515515
BC040321	LHX8	LIM homeobox 8	-3.574630507
BC022257	LGALS14	lectin, galactoside-binding, soluble, 14	-3.56990287
NM_001001914	OR2G3	olfactory receptor, family 2, subfamily G, member 3	-3.56964789
BC027938	RSPO2	R-spondin 2	-3.567857773
NM_001001919	OR13C4	olfactory receptor, family 13, subfamily C, member 4	-3.567298718
AK095460	NLRP12	NLR family, pyrin domain containing 12	-3.566130434
NM_002042	GABRR1	gamma-aminobutyric acid (GABA) A receptor, rho 1	-3.563649594
NM_020638	FGF23	fibroblast growth factor 23	-3.556340869
BC069428	CIB3	calcium and integrin binding family member 3	-3.550002671
BC028719	TPTE	transmembrane phosphatase with tensin homology	-3.541882777
NM_000047	ARSE	arylsulfatase E (chondrodysplasia punctata 1)	-3.538669231
NM_020973	GBA3	glucosidase, beta, acid 3 (gene/pseudogene)	-3.5365068
NM_000032	ALAS2	5'-aminolevulinatase synthase 2	-3.535696382
NM_005924	MEOX2	mesenchyme homeobox 2	-3.534978388
AK126199	LOC100506688	uncharacterized LOC100506688	-3.534486459
BC101977	KIR2DS4	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 4	-3.532156014
NM_001001955	OR4C13	olfactory receptor, family 4, subfamily C, member 13	-3.528687297
NM_001005180	OR56B1	olfactory receptor, family 56, subfamily B, member 1	-3.52652085
BC074920	MAGEB6	melanoma antigen family B6	-3.526495115
NM_031457	MS4A8	membrane-spanning 4-domains, subfamily A, member 8	-3.523570452
NM_001038706	FER1L6-AS1	FER1L6 antisense RNA 1	-3.522706528
BC109208	SLC17A4	solute carrier family 17, member 4	-3.522672872
AK127532	LOC100130372	uncharacterized LOC100130372	-3.520386455

NM_001003750	OR8I2	olfactory receptor, family 8, subfamily I, member 2	-3.518974447
NM_006192	PAX1	paired box 1	-3.518406754
BC035071	GABBR2	gamma-aminobutyric acid (GABA) B receptor, 2	-3.515264101
NM_001004726	OR4X1	olfactory receptor, family 4, subfamily X, member 1 (gene/pseudogene)	-3.508988226
X98858	KIR2DS1	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 1	-3.507131179
NM_001004704	OR4C6	olfactory receptor, family 4, subfamily C, member 6	-3.502944546
BC069692	FGF9	fibroblast growth factor 9	-3.499692229
NM_000334	SCN4A	sodium channel, voltage gated, type IV alpha subunit	-3.496763178
NM_145234	CHRDL1	chordin-like 1	-3.495615656
NM_033056	PCDH15	protocadherin-related 15	-3.490947632
NM_001033660	LINC01555	long intergenic non-protein coding RNA 1555	-3.490165078
NM_030946	OR14J1	olfactory receptor, family 14, subfamily J, member 1	-3.485848675
NM_016247	IMPG2	interphotoreceptor matrix proteoglycan 2	-3.484550136
BC069996	LPAR4	lysophosphatidic acid receptor 4	-3.48122456
NM_006737	KIR3DL2	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2	-3.47821357
NM_001004714	OR4K13	olfactory receptor, family 4, subfamily K, member 13	-3.477774838
NM_001005468	OR8B2	olfactory receptor, family 8, subfamily B, member 2	-3.47367256
BX537502	SAG	S-antigen; retina and pineal gland (arrestin)	-3.471390268
NM_003886	AKAP4	A kinase (PRKA) anchor protein 4	-3.470043255
NM_001004745	OR5T1	olfactory receptor, family 5, subfamily T, member 1	-3.467873531
NM_002644	PIGR	polymeric immunoglobulin receptor	-3.461611616
NM_003117	SPAM1	sperm adhesion molecule 1 (PH-20 hyaluronidase, zona pellucida binding)	-3.459658989
BC058286	TRPM1	transient receptor potential cation channel, subfamily M, member 1	-3.458795258
BC101001	SEC14L3	SEC14-like lipid binding 3	-3.45695666
NM_001004462	OR10G4	olfactory receptor, family 10, subfamily G, member 4	-3.454940798
NM_032607	CREB3L3	cAMP responsive element binding protein 3-like 3	-3.449339881
AK125824	LOC100130698	uncharacterized LOC100130698	-3.445243642
CR749812	PSG8	pregnancy specific beta-1-glycoprotein 8	-3.439489338
AK097310	MYOC	myocilin, trabecular meshwork inducible glucocorticoid response	-3.438368133
NM_032884	C1orf94	chromosome 1 open reading frame 94	-3.434960341
BC111775	SLC13A1	solute carrier family 13 (sodium/sulfate symporter), member 1	-3.430655556
AY429509	KLK2	kallikrein-related peptidase 2	-3.428059386

NM_199234	GDNF	glial cell derived neurotrophic factor	-3.425405743
NM_001002918	OR8D2	olfactory receptor, family 8, subfamily D, member 2 (gene/pseudogene)	-3.42304055
NM_001013355	OR2G6	olfactory receptor, family 2, subfamily G, member 6	-3.421629639
NM_001005279	OR6K2	olfactory receptor, family 6, subfamily K, member 2	-3.419231034
NM_000685	AGTR1	angiotensin II receptor, type 1	-3.412798968
BC100991	USP17L9P	ubiquitin specific peptidase 17-like family member 9, pseudogene	-3.410940663
BC101541	MMP26	matrix metalloproteinase 26	-3.410610266
NM_001004758	OR51S1	olfactory receptor, family 51, subfamily S, member 1	-3.405145113
NM_001005160	OR52A5	olfactory receptor, family 52, subfamily A, member 5	-3.401426612
NM_153323	DEFB119	defensin, beta 119	-3.398848215
NM_001004461	OR10A6	olfactory receptor, family 10, subfamily A, member 6 (gene/pseudogene)	-3.396237871
NM_133181	EPS8L3	EPS8-like 3	-3.394762527
NM_024114	TRIM48	tripartite motif containing 48	-3.393132399
NM_001005286	OR6F1	olfactory receptor, family 6, subfamily F, member 1	-3.391617438
AF282269	GRK7	G protein-coupled receptor kinase 7	-3.389146952
NM_173610	EWSAT1	Ewing sarcoma associated transcript 1	-3.3819815
NM_201565	FAM230B	family with sequence similarity 230, member B (non-protein coding)	-3.38129067
BC022895	MS4A8	membrane-spanning 4-domains, subfamily A, member 8	-3.371640503
NM_001033576	UNC45B	unc-45 myosin chaperone B	-3.368168313
NM_033226	ABCC12	ATP-binding cassette, sub-family C (CFTR/MRP), member 12	-3.366397882
NM_001001915	OR2G2	olfactory receptor, family 2, subfamily G, member 2	-3.365377655
NM_134444	NLRP4	NLR family, pyrin domain containing 4	-3.364139389
NM_139136	KCNC2	potassium channel, voltage gated Shaw related subfamily C, member 2	-3.363153498
AK093369	LUZP4	leucine zipper protein 4	-3.362747172
NM_033303	ADRA1A	adrenoceptor alpha 1A	-3.360916375
NM_003008	SEMG2	semenogelin II	-3.358140597
AK127250	GSTM5	glutathione S-transferase mu 5	-3.35728273
NM_001029888	FAM24A	family with sequence similarity 24, member A	-3.353330377
NM_001005238	OR51G2	olfactory receptor, family 51, subfamily G, member 2	-3.351919881
NM_004063	CDH17	cadherin 17, LI cadherin (liver-intestine)	-3.350759945
AJ437318	HTR3D	5-hydroxytryptamine (serotonin) receptor 3D, ionotropic	-3.344644811
NM_001017929	FRMPD2	FERM and PDZ domain containing 2	-3.342874928

BC106070	MME	membrane metallo-endopeptidase	-3.340418285
AK058151	RANBP3L	RAN binding protein 3-like	-3.339329898
NM_014513	KIR2DS5	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 5	-3.337426795
NM_018189	DPPA4	developmental pluripotency associated 4	-3.335309124
NM_001005243	OR9K2	olfactory receptor, family 9, subfamily K, member 2	-3.327980755
NM_000412	HRG	histidine-rich glycoprotein	-3.326953172
NM_006900	IFNA13	interferon, alpha 13	-3.324812915
BC061519	VIT	vitrin	-3.323069111
AK091315	SCN8A	sodium channel, voltage gated, type VIII alpha subunit	-3.317621428
AY792621	OR4F13P	olfactory receptor, family 4, subfamily F, member 13 pseudogene	-3.31597157
NM_001039372	HEPACAM2	HEPACAM family member 2	-3.310951584
BC095538	LPAR4	lysophosphatidic acid receptor 4	-3.310046097
M64930	PPP2R2B	protein phosphatase 2, regulatory subunit B, beta	-3.309659773
NM_033053	DMRTC1	DMRT-like family C1	-3.308823333
BC007028	CELA3A	chymotrypsin-like elastase family, member 3A	-3.308743471
NM_032518	COL25A1	collagen, type XXV, alpha 1	-3.303164377
NM_023013	PRAMEF1	PRAME family member 1	-3.29996693
NM_052996	PRDM7	PR domain containing 7	-3.293553104
NM_020866	KLHL1	kelch-like family member 1	-3.290434381
BC032842	FAM71F1	family with sequence similarity 71, member F1	-3.286392676
NM_001004727	OR4X2	olfactory receptor, family 4, subfamily X, member 2 (gene/pseudogene)	-3.285555932
NM_022436	ABCG5	ATP-binding cassette, sub-family G (WHITE), member 5	-3.280160865
BC046111	CABS1	calcium-binding protein, spermatid-specific 1	-3.279441645
NM_181608	KRTAP19-2	keratin associated protein 19-2	-3.278835305
NM_001005191	OR7D4	olfactory receptor, family 7, subfamily D, member 4	-3.278108252
AK126941	TF	transferrin	-3.270148783
NM_138283	CSTL1	cystatin-like 1	-3.26960393
NM_003880	WISP3	WNT1 inducible signaling pathway protein 3	-3.266759555
NM_032298	SYT3	synaptotagmin III	-3.263526213
AY358338	VIT	vitrin	-3.256793241
AK127178	LOC101929106	uncharacterized LOC101929106	-3.254577502
AK131386	FRMPD2	FERM and PDZ domain containing 2	-3.252901458
BC096698	EDDM3A	epididymal protein 3A	-3.251579356
NM_003053	SLC18A1	solute carrier family 18 (vesicular monoamine transporter), member 1	-3.247554899

AK057306	SLC22A16	solute carrier family 22 (organic cation/carnitine transporter), member 16	-3.236773127
NM_001005213	OR9G1	olfactory receptor, family 9, subfamily G, member 1	-3.232313049
NM_182935	MOBP	myelin-associated oligodendrocyte basic protein	-3.231725378
NM_022437	ABCG8	ATP-binding cassette, sub-family G (WHITE), member 8	-3.231724952
BC108917	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	-3.228430404
NM_053276	VIT	vitrin	-3.226806464
BC020707	TAT	tyrosine aminotransferase	-3.225405636
AK127092	SLC24A2	solute carrier family 24 (sodium/potassium/calcium exchanger), member 2	-3.225159983
AY529122	UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	-3.224962307
NM_001013356	OR8U8	olfactory receptor, family 8, subfamily U, member 8	-3.213480247
BC054496	ADIPOQ	adiponectin, C1Q and collagen domain containing	-3.210538648
NM_005635	SSX1	synovial sarcoma, X breakpoint 1	-3.207147911
NM_052917	GALNT13	polypeptide N-acetylgalactosaminyltransferase 13	-3.203676337
XM_929916	FAM188B2	family with sequence similarity 188, member B2	-3.203265319
NM_001003954	ANXA13	annexin A13	-3.202220055
BC009387	SLC18A1	solute carrier family 18 (vesicular monoamine transporter), member 1	-3.199407444
NM_000846	GSTA2	glutathione S-transferase alpha 2	-3.199002424
NM_001004729	OR5AN1	olfactory receptor, family 5, subfamily AN, member 1	-3.196805532
AK093225	SPATA9	spermatogenesis associated 9	-3.195668195
BC096699	IFNA21	interferon, alpha 21	-3.194478952
AJ002102	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	-3.188739741
BC020619	GSTA3	glutathione S-transferase alpha 3	-3.188267519
BC040656	LRRC3B	leucine rich repeat containing 3B	-3.187009489
NM_017504	OR2M4	olfactory receptor, family 2, subfamily M, member 4	-3.185406637
NM_031282	FCRL4	Fc receptor-like 4	-3.178427535
NM_003700	OR2D2	olfactory receptor, family 2, subfamily D, member 2	-3.177016001
NM_173801	OOSP2	oocyte secreted protein 2	-3.174812323
NM_001039910	NXF2	nuclear RNA export factor 2	-3.17291869
BC101489	OR6A2	olfactory receptor, family 6, subfamily A, member 2	-3.1720099
NM_002747	MAPK4	mitogen-activated protein kinase 4	-3.170916129
NM_130852	BPIFA1	BPI fold containing family A, member 1	-3.168025289
NM_153217	TMEM174	transmembrane protein 174	-3.163858282
NM_001012264	RNASE13	ribonuclease, RNase A family, 13 (non-active)	-3.163676332

BC111734	GRIA1	glutamate receptor, ionotropic, AMPA 1	-3.16275414
NM_013377	PDZRN4	PDZ domain containing ring finger 4	-3.161000736
NM_004276	CABP1	calcium binding protein 1	-3.159841063
NM_207186	OR10A4	olfactory receptor, family 10, subfamily A, member 4	-3.157343989
BX161483	LOC100129345	uncharacterized LOC100129345	-3.154427932
U92992	ADD3-AS1	ADD3 antisense RNA 1	-3.153728906
BC101640	IFNA21	interferon, alpha 21	-3.150597262
AB095365	RFX4	regulatory factor X, 4 (influences HLA class II expression)	-3.149799994
BC030780	CCDC54	coiled-coil domain containing 54	-3.149375079
NM_198185	OVCH2	ovochymase 2 (gene/pseudogene)	-3.142648918
AB180041	KRTAP19-2	keratin associated protein 19-2	-3.141806452
AK128800	LOC100133857	uncharacterized LOC100133857	-3.141720042
NM_000307	POU3F4	POU class 3 homeobox 4	-3.141172053
BC069129	XAGE5	X antigen family, member 5	-3.139132502
NM_173357	SSX6	synovial sarcoma, X breakpoint 6 (pseudogene)	-3.135533795
AK098841	VWA3B	von Willebrand factor A domain containing 3B	-3.131694977
NM_174901	FAM9C	family with sequence similarity 9, member C	-3.130182444
AY358390	DPEP3	dipeptidase 3	-3.129907102
NM_015868	KIR2DL3	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3	-3.121628061
BC039730	GPC5	glypican 5	-3.121453316
BC104456	ODF1	outer dense fiber of sperm tails 1	-3.114891182
NM_001005514	OR5H14	olfactory receptor, family 5, subfamily H, member 14	-3.110121304
NM_001031666	MS4A3	membrane-spanning 4-domains, subfamily A, member 3 (hematopoietic cell-specific)	-3.10682383
BC105113	CST8	cystatin 8 (cystatin-related epididymal specific)	-3.106027276
BC101526	PHACTR1	phosphatase and actin regulator 1	-3.09964898
NM_138932	A1CF	APOBEC1 complementation factor	-3.099560078
NM_001001963	OR2L8	olfactory receptor, family 2, subfamily L, member 8 (gene/pseudogene)	-3.098512301
NM_178127	ANGPTL5	angiotensin-like 5	-3.096673041
NM_001004052	OR52B2	olfactory receptor, family 52, subfamily B, member 2	-3.094877137
BC022316	PSG2	pregnancy specific beta-1-glycoprotein 2	-3.094842891
NM_000223	KRT12	keratin 12, type I	-3.091301577
BC093635	KCNC2	potassium channel, voltage gated Shaw related subfamily C, member 2	-3.090188109
NM_001005283	OR9Q2	olfactory receptor, family 9, subfamily Q, member 2	-3.089559301
NM_199451	ZNF365	zinc finger protein 365	-3.088322307
NM_001005167	OR52E6	olfactory receptor, family 52, subfamily E, member 6	-3.077799164

NM_007163	SLC14A2	solute carrier family 14 (urea transporter), member 2	-3.074996224
NM_170745	HIST1H2AA	histone cluster 1, H2aa	-3.073870421
NM_000111	SLC26A3	solute carrier family 26 (anion exchanger), member 3	-3.073096227
NM_001013675	LRRRC74B	leucine rich repeat containing 74B	-3.072475758
NM_002782	PSG6	pregnancy specific beta-1-glycoprotein 6	-3.07057713
NM_001005480	OR2A2	olfactory receptor, family 2, subfamily A, member 2	-3.063979647
NM_214675	CLEC4M	C-type lectin domain family 4, member M	-3.063816732
NM_001001917	OR56A1	olfactory receptor, family 56, subfamily A, member 1	-3.060257662
BC112126	GLP1R	glucagon-like peptide 1 receptor	-3.059662016
NM_002591	PCK1	phosphoenolpyruvate carboxykinase 1 (soluble)	-3.059325869
NM_000354	SERPINA7	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7	-3.055326874
NM_001001731	HHCM	Mahlavu hepatocellular carcinoma	-3.053970454
NM_000198	HSD3B2	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2	-3.053739734
NM_001005469	OR5B3	olfactory receptor, family 5, subfamily B, member 3	-3.052838792
NM_001023569	DUX4L9	double homeobox 4 like 9	-3.041474369
NM_000847	GSTA3	glutathione S-transferase alpha 3	-3.041401321
BC106753	FBXL21	F-box and leucine-rich repeat protein 21 (gene/pseudogene)	-3.039613491
NM_001001954	OR5A2	olfactory receptor, family 5, subfamily A, member 2	-3.03939439
NM_001004490	OR2AG2	olfactory receptor, family 2, subfamily AG, member 2	-3.037654311
NM_001836	CMA1	chymase 1, mast cell	-3.037305181
NM_001013732	PTCHD4	patched domain containing 4	-3.028129593
BC034977	UBQLNL	ubiquilin-like	-3.026889302
NM_032599	FAM71F1	family with sequence similarity 71, member F1	-3.026316635
NM_001014342	FLG2	filaggrin family member 2	-3.022181744
BC031616	NXF3	nuclear RNA export factor 3	-3.014398471
L76668	KIR2DS2	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 2	-3.012414449
NM_000453	SLC5A5	solute carrier family 5 (sodium/iodide cotransporter), member 5	-3.010917652
NM_054032	MRGPRX4	MAS-related GPR, member X4	-3.003750637
BC033839	C10orf126	chromosome 10 open reading frame 126	-2.997625233
NM_006905	PSG1	pregnancy specific beta-1-glycoprotein 1	-2.993888085
BC093773	CCL26	chemokine (C-C motif) ligand 26	-2.991579043
NM_032411	C2orf40	chromosome 2 open reading frame 40	-2.990458365
NM_001004692	OR2T12	olfactory receptor, family 2, subfamily T, member 12	-2.989581957
BC065723	SUMO1P1	SUMO1 pseudogene 1	-2.987379793
NM_080718	TBX5	T-box 5	-2.98423934

NM_006422	AKAP3	A kinase (PRKA) anchor protein 3	-2.982683184
NM_016116	ASB4	ankyrin repeat and SOCS box containing 4	-2.981359916
NM_032738	FCRLA	Fc receptor-like A	-2.981347486
BC040047	IQCF2	IQ motif containing F2	-2.980389965
NM_152751	BEND7	BEN domain containing 7	-2.979165004
NM_080390	TCEAL2	transcription elongation factor A (SII)-like 2	-2.978407769
BC074752	KCNJ1	potassium channel, inwardly rectifying subfamily J, member 1	-2.978177586
NM_002233	KCNA4	potassium channel, voltage gated shaker related subfamily A, member 4	-2.971751533
NM_006798	UGT2A1	UDP glucuronosyltransferase 2 family, polypeptide A1, complex locus	-2.969983703
BC101303	FBXW12	F-box and WD repeat domain containing 12	-2.967712178
NM_152250	DEFB105A	defensin, beta 105A	-2.965173362
NM_001005162	OR52B6	olfactory receptor, family 52, subfamily B, member 6	-2.956231564
AK021788	TRPM3	transient receptor potential cation channel, subfamily M, member 3	-2.953655306
BC070175	DLGAP1	discs, large (Drosophila) homolog-associated protein 1	-2.950728066
NM_003469	SCG2	secretogranin II	-2.948416225
AK096981	LOC729173	uncharacterized LOC729173	-2.947842837
NM_005071	SLC1A6	solute carrier family 1 (high affinity aspartate/glutamate transporter), member 6	-2.94365395
NM_181654	CPLX4	complexin 4	-2.941660675
NM_001073	UGT2B11	UDP glucuronosyltransferase 2 family, polypeptide B11	-2.936110669
NM_005458	GABBR2	gamma-aminobutyric acid (GABA) B receptor, 2	-2.934088702
NM_002045	GAP43	growth associated protein 43	-2.933576299
NM_207283	NPSR1-AS1	NPSR1 antisense RNA 1	-2.931703562
NM_080912	ASGR2	asialoglycoprotein receptor 2	-2.929386741
NM_152460	C17orf77	chromosome 17 open reading frame 77	-2.929277456
NM_000959	PTGFR	prostaglandin F receptor (FP)	-2.927210832
NM_153453	VGLL2	vestigial-like family member 2	-2.918309514
NM_000083	CLCN1	chloride channel, voltage-sensitive 1	-2.911838332
NM_152598	MARCH10	membrane-associated ring finger (C3HC4) 10, E3 ubiquitin protein ligase	-2.911182129
BC107951	KIR3DL2	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 2	-2.907178171
BC110798	FRMPD2B	FERM and PDZ domain containing 2B, pseudogene	-2.906454811

BC069441	UGT2B11	UDP glucuronosyltransferase 2 family, polypeptide B11	-2.903201382
NM_181461	PAX3	paired box 3	-2.901784167
AY803021	DPF3	D4, zinc and double PHD fingers, family 3	-2.901113318
NM_033050	SUCNR1	succinate receptor 1	-2.898706214
NM_022788	P2RY12	purinergic receptor P2Y, G-protein coupled, 12	-2.897375887
NM_000460	THPO	thrombopoietin	-2.897300999
AK092825	LOC100131894	uncharacterized LOC100131894	-2.89726417
NM_001005284	OR9G4	olfactory receptor, family 9, subfamily G, member 4	-2.896004206
AY509882	GYPB	glycophorin B (MNS blood group)	-2.894026557
AK129590	B3GALT5-AS1	B3GALT5 antisense RNA 1	-2.87758756
NM_022052	NXF3	nuclear RNA export factor 3	-2.873403249
NM_001001786	BLID	BH3-like motif containing, cell death inducer	-2.871198392
NM_001013630	AADACL4	arylacetamide deacetylase-like 4	-2.870412009
NM_004122	GHSR	growth hormone secretagogue receptor	-2.870392229
NM_005577	LPA	lipoprotein, Lp(a)	-2.867380034
AK092698	LMO7DN	LMO7 downstream neighbor	-2.866347743
NM_001004450	OR1B1	olfactory receptor, family 1, subfamily B, member 1 (gene/pseudogene)	-2.864386088
NM_004485	GNG4	guanine nucleotide binding protein (G protein), gamma 4	-2.864086913
NM_015687	FILIP1	filamin A interacting protein 1	-2.860905819
NM_153443	KIR3DL3	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 3	-2.860764425
BC012108	LOC440700	carbonic anhydrase XIV (CA14) pseudogene	-2.859907634
BC042481	KIF26B	kinesin family member 26B	-2.854583303
AY429510	KLK2	kallikrein-related peptidase 2	-2.853026342
NM_000036	AMPD1	adenosine monophosphate deaminase 1	-2.846768933
BC101540	KIR3DL1	killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 1	-2.842867993
NM_012314	KIR2DS4	killer cell immunoglobulin-like receptor, two domains, short cytoplasmic tail, 4	-2.840774695
NM_172314	IL25	interleukin 25	-2.840221211
NM_052956	ACSM1	acyl-CoA synthetase medium-chain family member 1	-2.837154889
BC042113	TPD52L3	tumor protein D52-like 3	-2.834000445
NM_175725	IL5RA	interleukin 5 receptor, alpha	-2.832823639
NM_054110	GALNT15	polypeptide N-acetylgalactosaminyltransferase 15	-2.831519878
NM_001024822	RNASE12	ribonuclease, RNase A family, 12 (non-active)	-2.827944533
AF259970	KLK4	kallikrein-related peptidase 4	-2.827834889
BC021742	C2orf40	chromosome 2 open reading frame 40	-2.824312719

NM_001033018	DEFB136	defensin, beta 136	-2.813727705
AK026597	CFAP43	cilia and flagella associated protein 43	-2.809991191
NM_031994	RNF17	ring finger protein 17	-2.80950331
NM_003240	LEFTY2	left-right determination factor 2	-2.807869413
BC093647	FLJ34503	uncharacterized FLJ34503	-2.806839505
NM_001040054	LINC00488	long intergenic non-protein coding RNA 488	-2.806008399
NM_153191	SLC22A2	solute carrier family 22 (organic cation transporter), member 2	-2.801359606
NM_001029866	CCT8L1P	chaperonin containing TCP1, subunit 8 (theta)-like 1, pseudogene	-2.799912663
AK093161	FAM71F1	family with sequence similarity 71, member F1	-2.799891971
NM_001001952	OR5D18	olfactory receptor, family 5, subfamily D, member 18	-2.799819093
NM_145016	GLYATL2	glycine-N-acyltransferase-like 2	-2.791235137
NM_001005609	EDA	ectodysplasin A	-2.788697596
NM_006552	SCGB1D1	secretoglobin, family 1D, member 1	-2.787933779
BC074770	DAO	D-amino-acid oxidase	-2.787254768
M64752	GRIA1	glutamate receptor, ionotropic, AMPA 1	-2.785707442
NM_000615	NCAM1	neural cell adhesion molecule 1	-2.782677322
NM_139054	ADAMTS18	ADAM metallopeptidase with thrombospondin type 1 motif, 18	-2.782095605
NM_203281	BMX	BMX non-receptor tyrosine kinase	-2.781222883
NM_138450	ARL11	ADP-ribosylation factor-like 11	-2.780846865
AK128431	MORN1	MORN repeat containing 1	-2.776172156
AY358150	SCARA5	scavenger receptor class A, member 5	-2.772536267
NM_006308	HSPB3	heat shock 27kDa protein 3	-2.754492344
NM_001005182	OR6C1	olfactory receptor, family 6, subfamily C, member 1	-2.752409847
NM_130794	CST11	cystatin 11	-2.751596923
AK124170	LOC100128170	uncharacterized LOC100128170	-2.751588319
NM_021631	TPT1P8	tumor protein, translationally-controlled 1 pseudogene 8	-2.747851129
NM_173489	MROH2B	maestro heat-like repeat family member 2B	-2.746592842
NM_001004702	OR4C3	olfactory receptor, family 4, subfamily C, member 3	-2.745166081
BC069603	GCM2	glial cells missing homolog 2 (Drosophila)	-2.740037989
NM_024082	PRRG3	proline rich Gla (G-carboxyglutamic acid) 3 (transmembrane)	-2.738083114
NM_005383	NEU2	sialidase 2 (cytosolic sialidase)	-2.726399869
BC078658	TMEM35	transmembrane protein 35	-2.723808907
NM_138703	MAGEE2	melanoma antigen family E2	-2.720794969
AK055470	LOC100130507	uncharacterized LOC100130507	-2.715758042
NM_005588	MEP1A	meprin A, alpha (PABA peptide hydrolase)	-2.714371413

BC060041	SLC2A2	solute carrier family 2 (facilitated glucose transporter), member 2	-2.710963557
NM_198448	REG3G	regenerating islet-derived 3 gamma	-2.691887096
AK127678	LOC100129322	uncharacterized LOC100129322	-2.689347335
AK124722	SV2B	synaptic vesicle glycoprotein 2B	-2.688674546
NM_199356	THPO	thrombopoietin	-2.687591591
NM_002424	MMP8	matrix metalloproteinase 8	-2.682824403
BC052271	ATP1A2	ATPase, Na+/K+ transporting, alpha 2 polypeptide	-2.678090653
AY033996	ELAVL4	ELAV like neuron-specific RNA binding protein 4	-2.673433167
NM_144699	ATP1A4	ATPase, Na+/K+ transporting, alpha 4 polypeptide	-2.665251522
BC025711	NKX2-5	NK2 homeobox 5	-2.664314196
NM_144675	GSG1L	GSG1-like	-2.661009752
U94363	GYG2	glycogenin 2	-2.660561337
BC069819	SLC38A4	solute carrier family 38, member 4	-2.657105807
NM_004590	CCL16	chemokine (C-C motif) ligand 16	-2.656431197
BC096135	TNP2	transition protein 2 (during histone to protamine replacement)	-2.653837899
NM_005012	ROR1	receptor tyrosine kinase-like orphan receptor 1	-2.653329699
NM_020632	ATP6V0A4	ATPase, H+ transporting, lysosomal V0 subunit a4	-2.652123427
NM_007072	HHLA2	HERV-H LTR-associating 2	-2.650633297
NM_001869	CPA2	carboxypeptidase A2 (pancreatic)	-2.65058445
BC075052	NPY2R	neuropeptide Y receptor Y2	-2.647502257
NM_031429	RTBDN	retbindin	-2.641028532
NM_054113	CIB3	calcium and integrin binding family member 3	-2.640869548
BC074733	GRPR	gastrin-releasing peptide receptor	-2.639200453
NM_025031	C7orf69	chromosome 7 open reading frame 69	-2.628596328
NM_002652	PIP	prolactin-induced protein	-2.625700442
NM_002241	KCNJ10	potassium channel, inwardly rectifying subfamily J, member 10	-2.625115425
NM_001001670	SPATA31D1	SPATA31 subfamily D, member 1	-2.623816096
AB093548	SCN1A	sodium channel, voltage gated, type I alpha subunit	-2.614506042
NM_020212	WDR93	WD repeat domain 93	-2.609348135
BC017898	P2RY12	purinergic receptor P2Y, G-protein coupled, 12	-2.605813127
NM_030906	STK33	serine/threonine kinase 33	-2.602971373
NM_033302	ADRA1A	adrenoceptor alpha 1A	-2.600358583
NM_152869	RGN	regucalcin	-2.597462247
NM_001007534	C3orf56	chromosome 3 open reading frame 56	-2.592917689
NM_024693	ECHDC3	enoyl CoA hydratase domain containing 3	-2.586402472
NM_207646	LGALS17A	Charcot-Leyden crystal protein pseudogene	-2.585233042
BC022189	C17orf47	chromosome 17 open reading frame 47	-2.584386662

BC069348	GABRG2	gamma-aminobutyric acid (GABA) A receptor, gamma 2	-2.580497378
NM_021232	PRODH2	proline dehydrogenase (oxidase) 2	-2.576675707
NM_172005	WFDC13	WAP four-disulfide core domain 13	-2.559061024
AY359037	GFRA3	GDNF family receptor alpha 3	-2.551280206
NM_001638	APOF	apolipoprotein F	-2.548759687
AK130379	LL22NC01-81G9.3	uncharacterized protein FLJ39582-like	-2.548192608
BC009975	WSCD1	WSC domain containing 1	-2.538035642
BC025311	HHIP	hedgehog interacting protein	-2.522446024
NM_017716	MS4A12	membrane-spanning 4-domains, subfamily A, member 12	-2.521117723
NM_017928	LINC00483	long intergenic non-protein coding RNA 483	-2.508165861
BC014144	ELAVL3	ELAV like neuron-specific RNA binding protein 3	-2.507914874
NM_133371	MYOZ3	myozenin 3	-2.507074772
NM_032257	ZMYND12	zinc finger, MYND-type containing 12	-2.506418221
BC050640	ANGPTL1	angiotensin-like 1	-2.50559183
NM_001004329	DBX2	developing brain homeobox 2	-2.502482819
NM_000353	TAT	tyrosine aminotransferase	-2.497160048
NM_022912	REEP1	receptor accessory protein 1	-2.492727201
NM_174880	CLRN1	clarin 1	-2.481785085
AK090929	LINC00269	long intergenic non-protein coding RNA 269	-2.478041608
NM_207335	KBTBD12	kelch repeat and BTB (POZ) domain containing 12	-2.462328092
BC104214	ADGRF1	adhesion G protein-coupled receptor F1	-2.457909703
AK093069	IGSF10	immunoglobulin superfamily, member 10	-2.457859973
AK125708	LOC100128059	uncharacterized LOC100128059	-2.452655715
BC069609	SPRY3	sprouty RTK signaling antagonist 3	-2.445245135
NM_000742	CHRNA2	cholinergic receptor, nicotinic, alpha 2 (neuronal)	-2.43099665
BC021719	RHOXF2	Rhox homeobox family, member 2	-2.427851757
BC100878	CRYGN	crystallin, gamma N	-2.423177632
AK128370	CD160	CD160 molecule	-2.418219163
NM_207411	XKR5	XK, Kell blood group complex subunit-related family, member 5	-2.412866404
NM_032843	FIBCD1	fibrinogen C domain containing 1	-2.406663087
NM_001004342	TRIM67	tripartite motif containing 67	-2.400377598
AK126399	UNQ6494	uncharacterized LOC100129066	-2.390864443
NM_003984	SLC13A2	solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 2	-2.387879473
BC069677	RGS8	regulator of G-protein signaling 8	-2.386508225
NM_024847	TMC7	transmembrane channel-like 7	-2.380816425

AK075513	PLEKHG4B	pleckstrin homology domain containing, family G (with RhoGef domain) member 4B	-2.379687678
NM_152738	LINC00518	long intergenic non-protein coding RNA 518	-2.376715039
NM_001008748	MGAM2	maltase-glucoamylase 2 (putative)	-2.375513536
AY098593	KGFLP1	fibroblast growth factor 7 pseudogene	-2.364590292
NM_001012714	PRR26	proline rich 26	-2.347941736
AK058167	AXDND1	axonemal dynein light chain domain containing 1	-2.335718272
NM_199352	SLC22A25	solute carrier family 22, member 25	-2.319398676
BC026251	PAH	phenylalanine hydroxylase	-2.303299065
NM_002381	MATN3	matrilin 3	-2.294849261
BC029488	SPACA1	sperm acrosome associated 1	-2.269459135
AK130385	LOC100130157	uncharacterized LOC100130157	-2.261758343
NM_005511	MLANA	melan-A	-2.260699758
BC111524	SLC5A7	solute carrier family 5 (sodium/choline cotransporter), member 7	-2.248536328
NM_182589	HTR3E	5-hydroxytryptamine (serotonin) receptor 3E, ionotropic	-2.244244923
NM_198481	VSTM1	V-set and transmembrane domain containing 1	-2.240779687
NM_031900	AGXT2	alanine--glyoxylate aminotransferase 2	-2.230685726
NM_004164	RBP2	retinol binding protein 2, cellular	-2.227882265
NM_017584	MIOX	myo-inositol oxygenase	-2.225039062
NM_198720	PTGER3	prostaglandin E receptor 3 (subtype EP3)	-2.224049761
NM_012228	MSRB2	methionine sulfoxide reductase B2	-2.213674203
NM_004959	NR5A1	nuclear receptor subfamily 5, group A, member 1	-2.208140804
BC103496	CLSTN2	calsyntenin 2	-2.207040385
BC006523	SGK2	serum/glucocorticoid regulated kinase 2	-2.191552605
NM_004621	TRPC6	transient receptor potential cation channel, subfamily C, member 6	-2.178452696
BC074742	OPCML	opioid binding protein/cell adhesion molecule-like	-2.173240895
NM_019605	SERTAD4	SERTA domain containing 4	-2.1610728
NM_020683	TMIGD3	transmembrane and immunoglobulin domain containing 3	-2.160399711
NM_004129	GUCY1B2	guanylate cyclase 1, soluble, beta 2 (pseudogene)	-2.156489084
BC101090	HTR3D	5-hydroxytryptamine (serotonin) receptor 3D, ionotropic	-2.143411023
NM_001739	CA5A	carbonic anhydrase VA, mitochondrial	-2.142372118
NM_001011705	LECT1	leukocyte cell derived chemotaxin 1	-2.141233167
AF349540	PLA2G12B	phospholipase A2, group XIIB	-2.128014303
BC067107	IGDCC3	immunoglobulin superfamily, DCC subclass, member 3	-2.119531625

NM_001646	APOC4	apolipoprotein C-IV	-2.111431996
BC067753	KLHL10	kelch-like family member 10	-2.106393462
BC093658	TRPC6	transient receptor potential cation channel, subfamily C, member 6	-2.103582881
NM_001014336	IL31	interleukin 31	-2.098687627
AY326436	RIC3	RIC3 acetylcholine receptor chaperone	-2.08910414
CR749624	ZNF608	zinc finger protein 608	-2.08530182
NM_000415	IAPP	islet amyloid polypeptide	-2.079886086
BC032037	FTCD	formimidoyltransferase cyclodeaminase	-2.076132839
BC016665	TOX	thymocyte selection-associated high mobility group box	-2.028969909
NM_144684	ZNF480	zinc finger protein 480	-2.025783133
NM_006512	SAA4	serum amyloid A4, constitutive	-2.016263253
BC011804	BEND5	BEN domain containing 5	-2.002069165
BC100879	CRYGN	crystallin, gamma N	-1.901752337
AK098761	TMLHE-AS1	TMLHE antisense RNA 1	-1.884816015
BC069440	CDYL2	chromodomain protein, Y-like 2	-1.878507354
NM_003918	GYG2	glycogenin 2	-1.867686142
AK123991	PSMD5-AS1	PSMD5 antisense RNA 1 (head to head)	-1.834489184
NM_206893	MS4A10	membrane-spanning 4-domains, subfamily A, member 10	-1.826470241

Anexo F

GENES SOBREPRESADOS EN EL GENOTIPO 45 Y GENOTIPO 18

SEQ_ID	Simbolo del gen	Nombre del Gen	logFC
NM_000269	NME1	NME/NM23 nucleoside diphosphate kinase 1	3.814219077
BC011396	PUS7	pseudouridylate synthase 7 (putative)	3.246815566
NM_005038	PPID	peptidylprolyl isomerase D	3.124050817
NM_003385	VSNL1	visinin-like 1	3.11986836
AK000591	C1orf159	chromosome 1 open reading frame 159	2.989473217
BC063787	KREMEN1	kringle containing transmembrane protein 1	2.980064313
BC062427	NSMCE4A	NSE4 homolog A, SMC5-SMC6 complex component	2.970632781
BC010901	CKAP2	cytoskeleton associated protein 2	2.952725396
BC034580	SLC44A5	solute carrier family 44, member 5	2.915816956
NM_024698	SLC25A22	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	2.902916752
BC004202	CHEK1	checkpoint kinase 1	2.861238565
X73568	SYK	spleen tyrosine kinase	2.860175507
NM_001017973	P4HA2	prolyl 4-hydroxylase, alpha polypeptide II	2.8472834
NM_173547	TRIM65	tripartite motif containing 65	2.841614199
NM_017422	CALML5	calmodulin-like 5	2.792744413
NM_005252	FOS	FBJ murine osteosarcoma viral oncogene homolog	2.789291476
NM_144693	ZNF558	zinc finger protein 558	2.778064451
NM_004431	EPHA2	EPH receptor A2	2.771499601
NM_030816	ANKRD13C	ankyrin repeat domain 13C	2.768976687
BC093031	TSPAN4	tetraspanin 4	2.747098654
BC036582	RPL3	ribosomal protein L3	2.733333064
NM_032970	SEC22C	SEC22 homolog C, vesicle trafficking protein	2.72933961
BC040849	HBS1L	HBS1-like translational GTPase	2.727328521

NM_019600	FAM214A	family with sequence similarity 214, member A	2.727282845
NM_002885	RAP1GAP	RAP1 GTPase activating protein	2.673130588
NM_018397	CHDH	choline dehydrogenase	2.670920071
NM_020198	CCDC47	coiled-coil domain containing 47	2.647456989
NM_022780	RMND5A	required for meiotic nuclear division 5 homolog A	2.64632392
BC015231	PLEKHG5	pleckstrin homology domain containing, family G member 5	2.645046017
NM_052925	LENG8	leukocyte receptor cluster (LRC) member 8	2.632542925
NM_005029	PITX3	paired-like homeodomain 3	2.626983399
NM_017852	NLRP2	NLR family, pyrin domain containing 2	2.626768699
NM_181471	RFC2	replication factor C (activator 1) 2, 40kDa	2.62372048
BC016333	CCNL2	cyclin L2	2.622542514
AK093170	SLC44A5	solute carrier family 44, member 5	2.60746028
NM_016069	PAM16	presequence translocase-associated motor 16 homolog	2.60544039
NM_013396	USP25	ubiquitin specific peptidase 25	2.583191657
NM_145300	ALMS1P	Alstrom syndrome 1 pseudogene	2.573957894
AK092491	NOC2L	NOC2-like nucleolar associated transcriptional repressor	2.573198023
NM_173177	C1D	C1D nuclear receptor corepressor	2.553025905
NM_005891	ACAT2	acetyl-CoA acetyltransferase 2	2.546979818
BC016140	MVK	mevalonate kinase	2.54303573
BC111997	UBXN2A	UBX domain protein 2A	2.542459299
NM_033505	EPT1	ethanolaminephosphotransferase 1	2.54103763
BC005944	EIF3E	eukaryotic translation initiation factor 3, subunit E	2.532086521
NM_001089	ABCA3	ATP-binding cassette, sub-family A (ABC1), member 3	2.521102109
NM_198705	HSD11B1L	hydroxysteroid (11-beta) dehydrogenase 1-like	2.497656878
AB209790	AAMP	angio-associated, migratory cell protein	2.496846279
BC037785	FGFR1OP	FGFR1 oncogene partner	2.480406671
NM_152988	SPPL2B	signal peptide peptidase like 2B	2.477934208
NM_004843	IL27RA	interleukin 27 receptor, alpha	2.475700471

NM_024830	LPCAT1	lysophosphatidylcholine acyltransferase 1	2.468929752
AL137686	GOLGA8A	golgin A8 family, member A	2.465089655
NM_002750	MAPK8	mitogen-activated protein kinase 8	2.459730853
BC019255	PAICS	phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase	2.423847091
AB209633	MLLT6	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 6	2.412041994
BC005248	EIF1AY	eukaryotic translation initiation factor 1A, Y-linked	2.406610708
NM_021078	KAT2A	K(lysine) acetyltransferase 2A	2.390822839
NM_012325	MAPRE1	microtubule-associated protein, RP/EB family, member 1	2.390509882
NM_015658	NOC2L	NOC2-like nucleolar associated transcriptional repressor	2.382763836
BC000422	ARIH2	ariadne RBR E3 ubiquitin protein ligase 2	2.344976769
NM_001032998	KYNU	kynureninase	2.301173719
NM_018000	MREG	melanoregulin	2.299841567
NM_014229	SLC6A11	solute carrier family 6 (neurotransmitter transporter), member 11	2.290984984
NM_206920	MAMDC4	MAM domain containing 4	2.290105149
NM_014916	LMTK2	lemur tyrosine kinase 2	2.290000819
NM_001012762	CTU2	cytosolic thiouridylase subunit 2 homolog (S. pombe)	2.279842349
BC040008	TPT1	tumor protein, translationally-controlled 1	2.276555141
BC018426	URGCP	upregulator of cell proliferation	2.274735899
NM_014263	YME1L1	YME1-like 1 ATPase	2.267039838
NM_020647	JPH1	junctionophilin 1	2.264900708
NM_012318	LETM1	leucine zipper-EF-hand containing transmembrane protein 1	2.256405541
BC069009	MRS2	MRS2 magnesium transporter	2.254835942
BC008499	RSL24D1	ribosomal L24 domain containing 1	2.254562782
NM_138635	H2AFV	H2A histone family, member V	2.254128626
NM_032514	MAP1LC3A	microtubule-associated protein 1 light chain 3 alpha	2.218204574

NM_079421	CDKN2D	cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4)	2.195957383
NM_001025	RPS23	ribosomal protein S23	2.193239527
NM_016339	RAPGEFL1	Rap guanine nucleotide exchange factor (GEF)-like 1	2.191886532
NM_015305	ANGEL1	angel homolog 1 (Drosophila)	2.177673489
AY258037	TAZ	tafazzin	2.171415086
BC114480	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	2.169222082
NM_024678	NARS2	asparaginyl-tRNA synthetase 2, mitochondrial (putative)	2.167478282
NM_006031	PCNT	pericentrin	2.16022076
NM_001006635	MTX2	metaxin 2	2.154727279
NM_017426	NUP54	nucleoporin 54kDa	2.149675919
BC006568	PPP6R2	protein phosphatase 6, regulatory subunit 2	2.143584214
BC065198	LUC7L	LUC7-like	2.139328568
NM_003350	UBE2V2	ubiquitin-conjugating enzyme E2 variant 2	2.135119763
NM_007353	GNA12	guanine nucleotide binding protein (G protein) alpha 12	2.130055571
BC032747	MREG	melanoregulin	2.128742803
NM_181781	ZNF326	zinc finger protein 326	2.120356159
NM_005049	PWP2	PWP2 periodic tryptophan protein homolog (yeast)	2.108399975
NM_020232	PSMG2	proteasome (prosome, macropain) assembly chaperone 2	2.106243285
NM_002090	CXCL3	chemokine (C-X-C motif) ligand 3	2.085553968
BC039871	CCPG1	cell cycle progression 1	2.077606797
BC040950	MED20	mediator complex subunit 20	2.068217421
NM_002069	GNAI1	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	2.061707342
NM_017768	LRRC40	leucine rich repeat containing 40	2.060761814
BC012390	F12	coagulation factor XII (Hageman factor)	2.060254019
NM_031207	HY1	hydroxypyruvate isomerase (putative)	2.047206087
NM_012321	LSM4	LSM4 homolog, U6 small nuclear RNA and mRNA degradation associated	2.045883737
BC003590	MRPS25	mitochondrial ribosomal protein S25	2.019573418
BC020502	SEPT10	septin 10	2.016654292

BC074818	SMAD7	SMAD family member 7	1.962745688
AK092816	PDDC1	Parkinson disease 7 domain containing 1	1.946457552
NM_005671	UBXN8	UBX domain protein 8	1.908872959
NM_003249	THOP1	thimet oligopeptidase 1	1.902217506
NM_020142	NDUFA4L2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2	1.898555291
BC013420	AZIN1	antizyme inhibitor 1	1.897649603
AK128061	TAF1D	TATA box binding protein (TBP)-associated factor, RNA polymerase I, D, 41kDa	1.894005249
BC028685	LANCL1	LanC lantibiotic synthetase component C-like 1 (bacterial)	1.878302222
NM_153454	LINC00205	long intergenic non-protein coding RNA 205	1.867562187
NM_181558	RFC3	replication factor C (activator 1) 3, 38kDa	1.859515914
NM_001268	RCBTB2	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 2	1.838873602
BC009006	TMEM30A	transmembrane protein 30A	1.819937891
NM_183050	BCKDHB	branched chain keto acid dehydrogenase E1, beta polypeptide	1.811338792

Bibliografía

- Bendas, G., & Borsig, L. (2012). Cancer cell adhesion and metastasis: selectins, integrins, and the inhibitory potential of heparins. *International Journal of Cell Biology*, 2012, 676731. <https://doi.org/10.1155/2012/676731>
- Bergers, G., & Benjamin, L. E. (2003). Angiogenesis: Tumorigenesis and the angiogenic switch. *Nature Reviews Cancer*, 3(6), 401–410. <https://doi.org/10.1038/nrc1093>
- Bosch, F. Xavier, Silvia, de S. (2003). Human Papillomavirus and Cervical Cancer—Burden and Assessment of Causality. Retrieved August 18, 2015, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.333.9676&rep=rep1&type=pdf>
- Bosch, F. X., Lorincz, A., Muñoz, N., Meijer, C. J. L. M., & Shah, K. V. (2002). The causal relation between human papillomavirus and cervical cancer. *Journal of Clinical Pathology*, 55(4), 244–65. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11919208>
- Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, P. W. (2015). *Molecular Biology of the Cell*. (G. Science, Ed.) (6th ed.).
- Buck, C. B., Day, P. M., & Trus, B. L. (2013). The papillomavirus major capsid protein L1. *Virology*, 445(1), 169–174. <https://doi.org/10.1016/j.virol.2013.05.038>
- Buitrago-Pérez, A., Garaulet, G., Vázquez-Carballo, A., Paramio, J. M., & García-Escudero, R. (2009). Molecular Signature of HPV-Induced Carcinogenesis: pRb, p53 and Gene Expression Profiling. *Current Genomics*, 10(1), 26–34. <https://doi.org/10.2174/138920209787581235>
- Bumgarner, R. (2013). Overview of DNA microarrays: types, applications, and their future. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, Chapter 22, Unit 22.1. <https://doi.org/10.1002/0471142727.mb2201s101>
- Carlson M. (2016). org.Hs.eg.db: Genome wide annotation for Human.
- Cavallo, F., De Giovanni, C., Nanni, P., Forni, G., & Lollini, P. L. (2011). 2011: The immune hallmarks of cancer. In *Cancer Immunology, Immunotherapy* (Vol. 60, pp. 319–326).
- Chaturvedi, A. K., Katki, H. A., Hildesheim, A., Rodríguez, A. C., Quint, W., Schiffman, M., ... Herrero, R. (2011). Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease. *The Journal of Infectious Diseases*, 203(7), 910–20. <https://doi.org/10.1093/infdis/jiq139>
- Chen, A. A., Heideman, D. A. M., Boon, D., Gheit, T., Snijders, P. J. F., Tommasino, M., ... Clifford, G. M. (2014). Human Papillomavirus 45 Genetic Variation and Cervical Cancer Risk Worldwide. *Journal of Virology*, 88(8), 4514–4521. <https://doi.org/10.1128/JVI.03534-13>
- De Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., & Zur Hausen, H. (2004). Classification of papillomaviruses. *Virology*, 324(1), 17–27. <https://doi.org/10.1016/j.virol.2004.03.033>
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., ... Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods

- and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5), E359-86. <https://doi.org/10.1002/ijc.29210>
- Fernandez-Retana, J., Lasa-Gonsebatt, F., Lopez-Urrutia, E., Coronel-Martínez, J., Cantu De Leon, D., Jacobo-Herrera, N., ... Perez-Plasencia, C. (2015). Transcript Profiling Distinguishes Complete Treatment Responders With Locally Advanced Cervical Cancer. *Translational Oncology*, 8(2), 77-84. <https://doi.org/10.1016/j.tranon.2015.01.003>
- Fields, B. N., Knipe, D. M. (David M., & Howley, P. M. (2013). *Fields virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Fridman, J. S., & Lowe, S. W. (2003). Control of apoptosis by p53. *Oncogene*, 22(56), 9030-9040. <https://doi.org/10.1038/sj.onc.1207116>
- Garbuglia, A. R., Carletti, F., Minosse, C., Piselli, P., Zaniratti, M. S., Serraino, D., & Capobianchi, M. R. (2007). Genetic variability in E6 and E7 genes of human papillomavirus -16, -18, -31 and -33 from HIV-1-positive women in Italy. *The New Microbiologica*, 30(4), 377-82. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18080672>
- Goldman, B., Rebolj, M., Rygaard, C., Preisler, S., Ejegod, D. M., Lynge, E., & Bonde, J. (2013). Patterns of cervical coinfection with multiple human papilloma virus types in a screening population in Denmark. *Vaccine*, 31(12), 1604-9. <https://doi.org/10.1016/j.vaccine.2012.12.084>
- Hanahan, D., & Weinberg, R. a. (2011). Hallmarks of cancer: The next generation. *Cell*, 144(5), 646-674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Hanahan, D., Weinberg, R. a, & Francisco, S. (2000). The Hallmarks of Cancer Review University of California at San Francisco, 100, 57-70.
- Hecht, J. L., Kadish, A. S., Jiang, G., & Burk, R. D. (1995). Genetic characterization of the human papillomavirus (HPV) 18 E2 gene in clinical specimens suggests the presence of a subtype with decreased oncogenic potential. *International Journal of Cancer*, 60(3), 369-376. <https://doi.org/10.1002/ijc.2910600317>
- Hsieh, A. L., Walton, Z. E., Altman, B. J., Stine, Z. E., & Dang, C. V. (2015). MYC and Metabolism on the Path to Cancer. *Seminars in Cell & Developmental Biology*. <https://doi.org/10.1016/j.semcd.2015.08.003>
- Ishimi, Y., Okayasu, I., Kato, C., Kwon, H.-J., Kimura, H., Yamada, K., & Song, S.-Y. (2003). Enhanced expression of Mcm proteins in cancer cells derived from uterine cervix. *European Journal of Biochemistry*, 270(6), 1089-1101. <https://doi.org/10.1046/j.1432-1033.2003.03440.x>
- Jandova, J., Janda, J., & Slich, J. E. (2013). Cyclophilin 40 alters UVA-induced apoptosis and mitochondrial ROS generation in keratinocytes. *Experimental Cell Research*, 319(5), 750-60. <https://doi.org/10.1016/j.yexcr.2012.11.016>
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (n.d.). Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61(2), 69-90. <https://doi.org/10.3322/caac.20107>
- Li, X., & Coffino, P. (1996). High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. *Journal of Virology*, 70(7), 4509-16. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8676476>
- Lohavanichbutr, P., Houck, J., Fan, W., Yueh, B., Mendez, E., Futran, N., ... Chen,

- C. (2009). Genomewide gene expression profiles of HPV-positive and HPV-negative oropharyngeal cancer: potential implications for treatment choices. *Archives of Otolaryngology--Head & Neck Surgery*, *135*(2), 180–8. <https://doi.org/10.1001/archoto.2008.540>
- Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., Woolf, P. J., Luo, W., ... Akil, H. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*, *10*(1), 161. <https://doi.org/10.1186/1471-2105-10-161>
- Matsumoto, K., Oki, A., Furuta, R., Maeda, H., Yasugi, T., Takatsuka, N., ... Yoshikawa, H. (2011). Predicting the progression of cervical precursor lesions by human papillomavirus genotyping: a prospective cohort study. *International Journal of Cancer*, *128*(12), 2898–910. <https://doi.org/10.1002/ijc.25630>
- McLaughlin-Drubin, M. E., & Münger, K. (2009). The human papillomavirus E7 oncoprotein. *Virology*, *384*(2), 335–44. <https://doi.org/10.1016/j.virol.2008.10.006>
- Mejlhede, N., Pedersen, B. V., Frisch, M., & Fomsgaard, A. (2010). Multiple human papilloma virus types in cervical infections: competition or synergy? *APMIS*, *118*(5), 346–352. <https://doi.org/10.1111/j.1600-0463.2010.2602.x>
- Michel, J., Schönhaar, K., Schledzewski, K., Gkaniatsou, C., Sticht, C., Kellert, B., ... Schmieder, A. (2013). Identification of the novel differentiation marker MS4A8B and its murine homolog MS4A8A in colonic epithelial cells lost during neoplastic transformation in human colon. *Cell Death & Disease*, *4*(1), e469. <https://doi.org/10.1038/cddis.2012.215>
- Müller, M., Prescott, E. L., Wasson, C. W., & Macdonald, A. (2015). Human papillomavirus E5 oncoprotein: function and potential target for antiviral therapeutics. *Future Virology*, *10*(1), 27–39. <https://doi.org/10.2217/fvl.14.99>
- Muñoz, N., Bosch, F. X., de Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K. V., ... Meijer, C. J. L. M. (2003). Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer. *New England Journal of Medicine*, *348*(6), 518–527. <https://doi.org/10.1056/NEJMoa021641>
- Nakagawa, S., Yoshikawa, H., Onda, T., Kawana, T., Iwamoto, A., & Taketani, Y. (1996). Type of human papillomavirus is related to clinical features of cervical carcinoma. *Cancer*, *78*(9), 1935–41. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8909314>
- Pecorelli, S., Zigliani, L., & Odicino, F. (2009). Revised FIGO staging for carcinoma of the cervix. *International Journal of Gynecology & Obstetrics*, *105*(2), 107–108. <https://doi.org/10.1016/j.ijgo.2009.02.009>
- Periyasamy, S., Warriar, M., Tillekeratne, M. P. M., Shou, W., & Sanchez, E. R. (2007). The immunophilin ligands cyclosporin A and FK506 suppress prostate cancer cell growth by androgen receptor-dependent and -independent mechanisms. *Endocrinology*, *148*(10), 4716–26. <https://doi.org/10.1210/en.2007-0145>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Rizzo, R., Gentili, V., Rotola, A., Bortolotti, D., Cassai, E., & Di Luca, D. (2014).

- Implication of HLA-C and KIR alleles in human papillomavirus infection and associated cervical lesions. *Viral Immunology*, 27(9), 468–70.
<https://doi.org/10.1089/vim.2014.0017>
- Rosty, C., Sheffer, M., Tsafrir, D., Stransky, N., Tsafrir, I., Peter, M., ... Sastre-Garau, X. (2005). Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma. *Oncogene*, 24(47), 7094–7104.
<https://doi.org/10.1038/sj.onc.1208854>
- Rousseau, M.-C., Abrahamowicz, M., Villa, L. L., Costa, M. C., Rohan, T. E., & Franco, E. L. (2003). Predictors of cervical coinfection with multiple human papillomavirus types. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 12(10), 1029–37. Retrieved from <http://cebp.aacrjournals.org/content/12/10/1029.abstract>
- Rüegg, C. (2006). Leukocytes, inflammation, and angiogenesis in cancer: fatal attractions. *Journal of Leukocyte Biology*, 80(4), 682–4.
<https://doi.org/10.1189/jlb.0606394>
- Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (New York, N.Y.)*, 270(5235), 467–70. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7569999>
- Schlecht, N. F., Burk, R. D., Adrien, L., Dunne, A., Kawachi, N., Sarta, C., ... Belbin, T. J. (2007). Gene expression profiles in HPV-infected head and neck cancer. *The Journal of Pathology*, 213(3), 283–93.
<https://doi.org/10.1002/path.2227>
- Schmitt, M., Dondog, B., Waterboer, T., Pawlita, M., Tommasino, M., & Gheit, T. (2010). Abundance of multiple high-risk human papillomavirus (HPV) infections found in cervical cells analyzed by use of an ultrasensitive HPV genotyping assay. *Journal of Clinical Microbiology*, 48(1), 143–9.
<https://doi.org/10.1128/JCM.00991-09>
- Sichero, L., Franco, E. L., & Villa, L. L. (2005). Different P105 promoter activities among natural variants of human papillomavirus type 18. *The Journal of Infectious Diseases*, 191(5), 739–42. <https://doi.org/10.1086/427825>
- Silva, R., León, D., Brebi, P., Ili, C., Roa, J. C., & Sánchez, R. (2013). Diagnóstico de la infección por virus papiloma humano en el hombre. *Revista Chilena de Infectología*, 30(2), 186–192. <https://doi.org/10.4067/S0716-10182013000200009>
- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3, Article3. <https://doi.org/10.2202/1544-6115.1027>
- Smyth, G. K. (2005). limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (pp. 397–420). New York: Springer-Verlag. https://doi.org/10.1007/0-387-29362-0_23
- Sobota, R. S., Ramogola-Masire, D., Williams, S. M., & Zetola, N. M. (2014). Co-infection with HPV types from the same species provides natural cross-protection from progression to cervical cancer. *Infectious Agents and Cancer*,

- 9(1), 26. <https://doi.org/10.1186/1750-9378-9-26>
- Straub, E., Fertey, J., Dreer, M., Iftner, T., & Stubenrauch, F. (2015). Characterization of the Human Papillomavirus 16 E8 Promoter. *Journal of Virology*, 89(14), 7304–13. <https://doi.org/10.1128/JVI.00616-15>
- Sturn, A., Quackenbush, J., & Trajanoski, Z. (2002). Genesis: cluster analysis of microarray data. *Bioinformatics (Oxford, England)*, 18(1), 207–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11836235>
- Sun, C., Huo, D., Southard, C., Nemesure, B., Hennis, A., Cristina Leske, M., ... Di Rienzo, A. (2011). A signature of balancing selection in the region upstream to the human UGT2B4 gene and implications for breast cancer risk. *Human Genetics*, 130(6), 767–75. <https://doi.org/10.1007/s00439-011-1025-6>
- Trevino, V., Falciani, F., & Barrera-Saldaña, H. A. (2007). DNA microarrays: a powerful genomic tool for biomedical and clinical research. *Molecular Medicine (Cambridge, Mass.)*, 13(9–10), 527–41. <https://doi.org/10.2119/2006-00107.Trevino>
- Waggoner, S. E., Parkin, D., Whelan, S., Ferlay, J., Al, E., Jemal, A., ... Pongratz, E. (2003). Cervical cancer. *The Lancet*, 361(9376), 2217–2225. [https://doi.org/10.1016/S0140-6736\(03\)13778-6](https://doi.org/10.1016/S0140-6736(03)13778-6)
- Wallace, N. A., & Galloway, D. A. (2015). Novel Functions of the Human Papillomavirus E6 Oncoproteins. *Annual Review of Virology*, 2(1), 403–423. <https://doi.org/10.1146/annurev-virology-100114-055021>
- Wan, F., Miao, X., Quraishi, I., Kennedy, V., Creek, K. E., & Pirisi, L. (2008). Gene expression changes during HPV-mediated carcinogenesis: a comparison between an in vitro cell model and cervical cancer. *International Journal of Cancer*, 123(1), 32–40. <https://doi.org/10.1002/ijc.23463>
- Ward, B. K., Mark, P. J., Ingram, D. M., Minchin, R. F., & Ratajczak, T. (1999). Expression of the estrogen receptor-associated immunophilins, cyclophilin 40 and FKBP52, in breast cancer. *Breast Cancer Research and Treatment*, 58(3), 267–80. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10718488>
- Yon Rhee, S., Wood, V., Dolinski, K., & Draghici, S. (2008). Use and misuse of the gene ontology annotations. *Nature Reviews Genetics*, 9(7), 509–515. <https://doi.org/10.1038/nrg2363>
- Zacapala-Gómez, A. E., Del Moral-Hernández, O., Villegas-Sepúlveda, N., Hidalgo-Miranda, A., Romero-Córdoba, S. L., Beltrán-Anaya, F. O., ... Illades-Aguiar, B. (2016). Changes in global gene expression profiles induced by HPV 16 E6 oncoprotein variants in cervical carcinoma C33-A cells. *Virology*, 488, 187–95. <https://doi.org/10.1016/j.virol.2015.11.017>