

**Grado en ODONTOLOGÍA**

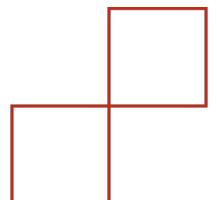
**Trabajo Fin de Grado**

**Curso 2023-24**

**Impacto del virus del papiloma humano en el carcinoma de células escamosas de cabeza y cuello. Metaanálisis de expresión génica.**

**Presentado por: Anna Isabel Gil Whitfield**

**Tutoras: Marta Serna García  
Nicla Flacco**



## **Agradecimientos**

En primer y más importante lugar quiero agradecer a mis padres, gente que en todo mi camino me ha apoyado y ayudado a ser mi mejor versión y me ha inspirado en todo momento. Gracias por vuestras enseñanzas, todas las experiencias juntas y vuestra confianza en mis capacidades y mi persona. Agradezco mucho tener a padres cómo vosotros en mi vida. Sin todo lo que habéis invertido no sé dónde estaría.

En segundo lugar, quiero agradecer a mi pareja de la vida, Alejandro. Eres un apoyo demasiado fuerte para mí y aprecio mucho todo lo que me has enseñado y cómo has creído en mí siempre. Espero que podamos llegar los dos a dónde queremos estar y espero que podamos hacerlo juntos para seguir inspirando uno al otro y apoyándonos en nuestros caminos.

Tercero, quiero agradecer a mi increíble círculo de amigos. Gente de todas edades y todos ámbitos todos con una cosa en común, el fuego dentro de querer crear una comunidad y crear proyectos. Cómo me encanta haber encontrado a un círculo de personas con los cuales puedo contar para cualquier cosa. Gracias por entender además que muchas veces los estudios han sido prioridad y que no he podido estar en todas, aunque lo intente.

Cuarto, me gustaría agradecer a mis tutoras de este TFG. Habéis sido fenomenales guiándome en este trabajo y aprecio cada tutoría que me habéis dado y vuestra confianza en mí. Ha sido un placer realizar este trabajo con vuestra ayuda.

Quinto, quiero agradecer a todas aquellas personas me han inspirado, ayudado, apoyado, me han hecho feliz, me han hecho reír y me han mostrado alguna lección de la vida; Mis abuelos y mi familia, Andrea & Gerhardt, Maggi & Wilfried, y muchos más. Que agradecida estoy de tener en mi vida a una tan gran cantidad de gente con buenos corazones.

Por último, quiero agradecer a la Universidad Europea por su apoyo y su confianza. Ha sido una experiencia lindísima cursar mis estudios con Ustedes.

## ÍNDICE

<b>1.</b>	<b>RESUMEN.....</b>	<b>4</b>
<b>2.</b>	<b>ABSTRACT.....</b>	<b>5</b>
<b>3.</b>	<b>PALABRAS CLAVES Y ABREVIATURAS.....</b>	<b>6 – 7</b>
<b>4.</b>	<b>INTRODUCCIÓN.....</b>	<b>8 – 21</b>
<b>5.</b>	<b>JUSTIFICACIÓN E HIPÓTESIS.....</b>	<b>22 – 23</b>
<b>6.</b>	<b>OBJETIVOS.....</b>	<b>24</b>
<b>7.</b>	<b>MATERIALES Y MÉTODOS.....</b>	<b>25 – 28</b>
<b>8.</b>	<b>RESULTADOS.....</b>	<b>29 – 44</b>
<b>9.</b>	<b>DISCUSIÓN.....</b>	<b>45 – 77</b>
<b>10.</b>	<b>CONCLUSIÓN.....</b>	<b>77 – 78</b>
<b>11.</b>	<b>BIBLIOGRAFÍA.....</b>	<b>79 – 94</b>
<b>12.</b>	<b>ANEXOS.....</b>	<b>95 – 135</b>
	<b>12.1.ANEXO 1.....</b>	<b>95 – 119</b>
	<b>12.2.ANEXO 2.....</b>	<b>120 – 135</b>

## 1. RESUMEN

### **Introducción:**

El carcinoma de células escamosas de cabeza y cuello (HNSCC) se caracteriza por diagnósticos tardíos, altas tasas de mortalidad y una incidencia en aumento. La asociación del virus del papiloma humano (HPV) con estos cánceres ha impulsado la necesidad de un análisis comparativo para comprender mejor los genes involucrados y proponer biomarcadores de pronóstico.

### **Materiales y Métodos:**

Este estudio se basó en 112 muestras de pacientes con HNSCC obtenidas de la base de datos TCGA, de las cuales 20 fueron HPV+ y 92 HPV- (para el HPV de tipo 16). Se efectuó un estudio transcriptómico, incluyendo un análisis diferencial de expresión génica, un análisis de enriquecimiento utilizando los software GSEA y DAVID, un análisis de curvas ROC y un análisis de supervivencia.

### **Resultados:**

El análisis diferencial reveló una sobreexpresión de 5.203 genes y una infraexpresión de 752 genes (FDR <0,01) en pacientes HNSCC-HPV+. De estos, 350 genes sobreexpresados y 11 genes infraexpresados mostraron un Area Under Curve (AUC) >0,7. El análisis funcional indicó que estos genes están relacionados con procesos celulares clave, como el ciclo celular, la organización de la cromatina y la queratinización, entre otros. Destacamos los genes EZH2, HDAC10, CDT1, CDK20, TET1, SPC24, SETMAR, BIRC6, KMT5C, KDM5A y SPRR2G, debido a su capacidad discriminativa como biomarcadores, su relación con los procesos de interés y su influencia positiva sobre la supervivencia en el grupo HNSCC-HPV+. El gen BIRC6 influyó de manera significativa (pvalor = 0,02).

### **Conclusión:**

Los genes EZH2, BIRC6, SPC24 y CDK20 demuestran un potencial valor como biomarcadores de pronóstico del HNSCC en pacientes HPV+. La identificación de estos biomarcadores podría permitir determinar qué pacientes HPV+ tienen un mayor riesgo de desarrollar cáncer, mejorando así el seguimiento y permitiendo un diagnóstico precoz del cáncer que podría mejorar su pronóstico.

## **2. ABSTRACT**

### **Introduction:**

Head and neck cancer, characterized by late diagnosis and high mortality rates, often involves treatments with side effects. The growing association of human papillomavirus (HPV) with these cancers has prompted a comparative analysis of samples from the TCGA database to better understand the involved genes and propose enhanced prognostic biomarkers.

### **Materials and Methods:**

For this study, information was obtained from 530 samples of head and neck squamous cell carcinomas (HNSCC) from the TCGA database. Of the 530 HNSCC samples, 112 included data on HPV type 16 infection; 20 were HPV+, and 92 were HPV-. Transcriptomic analysis was conducted, including differential analysis, enrichment analysis using GSEA and DAVID software, ROC curve analysis, and survival analysis.

### **Results:**

Differential analysis revealed overexpression of 5.203 genes and underexpression of 752 genes (FDR <0.01) in HNSCC-HPV+ patients compared to HNSCC-HPV-. Among the overexpressed genes, 350, and among the underexpressed genes, 11 showed an AUC>0.7. Functional analysis indicated their involvement in cellular processes, chromatin related processes, and keratinization, among others. Genes EZH2, HDAC10, CDT1, CDK20, TET1, SPC24, SETMAR, BIRC6, KMT5C, KDM5A, and SPRR2G were highlighted, and all had a positive effect on survival in HNSCC-HPV+ group in accordance with the expression found in our study. BIRC6 influenced survival significantly (p-value = 0.02).

### **Conclusion:**

Genes EZH2, BIRC6, SPC24, and CDK20 demonstrate potential value as biomarkers in HPV+ patients. Using these, it could be determined which HPV+ patients are at higher risk of developing cancer, allowing for focused monitoring and early cancer diagnosis to improve prognosis.

### 3. PALABRAS CLAVES Y ABREVIATURAS

- I. **Cáncer de cabeza y cuello**
- II. **Virus del papiloma humano**
- III. **Estudio de transcriptómica**
- IV. **Análisis diferencial**
- V. **Análisis de enriquecimiento**
- VI. **DAVID**
- VII. **Análisis de supervivencia**
- VIII. **Biomarcadores de pronóstico**
- IX. **Genes**

#### ABREVIATURAS

AUC:	Área debajo de la curva, del inglés <i>"Area under Curve"</i>
CDK:	Quinasa dependiente de ciclina, del inglés <i>"Cyclin dependant kinase"</i>
DAVID:	<i>Database for Annotation, Visualization and Integrated Discovery</i>
DEG:	Genes diferencialmente expresados, del inglés <i>"Differentially expressed genes"</i>
DNA:	Ácido desoxirribonucleico, del inglés <i>"deoxyribonucleic acid"</i>
FDR:	<i>False Discovery Rate</i>
FC:	<i>Fold Change</i>
GO:	<i>Gene Ontology</i>
GSEA:	<i>Gene set enrichment analysis</i>
HNSCC:	Carcinoma de cabeza y cuello de células escamosas, del inglés <i>"Head and Neck Squamous Cell Carcinoma"</i>
HPV:	Virus del papiloma humano, del inglés <i>"Human papilloma virus"</i>
HSPG:	Proteoglicanos de heparán sulfato, del inglés <i>"Heparan Sulfate Proteo Glycan"</i>
ICD:	Clasificación internacional de enfermedades, del inglés <i>"International Classification of Diseases"</i>
NCI:	Instituto Nacional de Cáncer, del inglés <i>"National Cancer Institute"</i>

NGS:	<i>Next Generation Sequencing</i>
NK:	<i>Natural Killer</i>
ODS:	Objetivo de Desarrollo Sostenible
OPSCC:	Carcinoma orofaríngeo de células escamosas, del inglés " <i>Oropharyngeal squamous cell carcinoma</i> "
ORF:	Secuencias codificantes de proteínas, del inglés " <i>open reading frame</i> "
OSCC:	Carcinoma oral de células escamosas, del inglés " <i>Oral squamous cell carcinoma</i> "
PCR:	Reacción en cadena de la polimerasa, del inglés " <i>Polymerase Chain Reaction</i> "
RNA:	Ácido ribonucleico, del inglés " <i>ribonucleic acid</i> "
ROC:	<i>Receiver Operating Characteristics</i>
ROS:	Especies reactivas de oxígeno, del inglés " <i>High reactive oxygen species</i> "
SCC:	Carcinoma de células escamosas, del inglés " <i>squamous cell carcinoma</i> "
SNP:	Polimorfismo de nucleótido único, del inglés " <i>Single Nucleotide Polymorphisms</i> "
TCGA:	<i>The Cancer Genome Atlas Project</i>
TME:	Microambiente tumoral, del inglés " <i>Tumor microenvironment</i> "
WHO:	Organización Mundial de la Salud, del inglés " <i>World Health Organization</i> "

## 4. INTRODUCCIÓN

### 4.1. Cáncer de cabeza y cuello

Los cánceres de cabeza y cuello representan un gran porcentaje de todos los cánceres mundiales. Es el sexto cáncer más común con una incidencia anual de entre 630.000 y 660.000 pacientes diagnosticados, y entre 325.000 y 350.000 fallecimientos anuales (1).

La gran mayoría (90%) de los cánceres de cabeza y cuello son carcinomas de células escamosas, los cuales se presentan en la mucosa oral (OSCC), la orofaringe (OPSCC) y la laringe (1-5). Los carcinomas orales de células escamosas (OSCC) más frecuentemente se desarrollan en el suelo de boca, la lengua, los labios, las encías, y la mucosa yugal (2). Las zonas de desarrollo del cáncer dan lugar a la clasificación ICD-10 (*International Classification for Diseases*) que fue desarrollada por la WHO (*World Health Organization*), la cual incluye 19 zonas desde los labios y la mucosa oral hasta la laringe, los senos y la faringe, entre otros. Aparte de tener una incidencia alta, también está en tendencia de crecimiento; se habla de una estimación de crecimiento de 30% de 2022 hasta 2030 (4).

Clásicamente estos cánceres se ven relacionados con el tabaquismo y el alcoholismo, los cuales contribuyen al desarrollo de 80% de los casos mundiales (3). Según Gormley y cols (2022), la combinación del alcohol y del tabaco tiene una mayor asociación a la aparición de cánceres de la laringe (89%) que a los cánceres orales (64%) (4). Aparte de estos factores de riesgo, también está relacionado con la masticación de la nuez de areca, la cual se consume mayormente en países de desarrollo por su efecto eufórico, psicoactivo, y de supresión de apetito; y con infecciones virales como el HPV (Virus del papiloma humano) o el Virus de Epstein-Barr (6).

Otros factores de riesgo incluyen la inflamación crónica, cómo por ejemplo por la periodontitis, una exposición crónica a carcinógenos en el ámbito laboral o ambiental, una nutrición e higiene oral inadecuada, una preposición genética y la presencia de estados inmunodeficientes (6). Lesiones traumáticas de la mucosa, una microbiota oral alterada, además de patologías relacionadas con una disminución de la reparación de DNA cómo pueden ser la ataxia-telangiectasia, xeroderma pigmentosum y la anemia de



Fanconi también cuentan como factores de riesgo para el desarrollo de carcinoma de cabeza y cuello de células escamosas (HNSCC) (3).

La mayoría de los factores de riesgo del cáncer oral son factores modificables, por lo cual juega un rol importante en la prevención la motivación al cambio de los estilos de vida, y la vacunación contra el HPV (3, 6).

El OSCC representa el cáncer más común en hombres y el tercero más común en mujeres en países de riesgo alto como son Bangladesh, India, Sri Lanka y Pakistán. Dos tercios de los nuevos casos de cánceres orales se diagnostican en estos países de alto riesgo (3). En Europa, el país más afectado por el OSCC es Francia (1). La incidencia además está relacionada con la edad, siendo mayor en personas con una edad mayor a 60 años (3).

Aunque se ha visto un descenso de la incidencia del carcinoma de células escamosas oral y de la laringe relacionados a factores de riesgo clásicos como el tabaco tanto en los Estados Unidos como también por ejemplo en Hong Kong y Canadá, se ha visto un incremento de aquellos cánceres de cabeza y cuello relacionados al HPV (HNSCC-HPV+) (4).

El OPSCC ha visto un incremento, lo cual se atribuye a una asociación con el HPV. Este SCC (Carcinoma de células escamosas) en específico relacionado con el HPV afecta a la base del  $\frac{1}{3}$  posterior de la lengua, las amígdalas y las paredes laterales de la orofaringe, además del anillo de Waldeyer incluyendo la nasofaringe (1). El tipo de HPV que se considera de mayor riesgo para el desarrollo del cáncer de la orofaringe es el tipo 16, y se piensa que la transmisión mayormente se da por vía oro-genital (4, 7). Se ha visto que un 90% de las infecciones con HPV se dan por actividad sexual (7).

El HPV se relaciona hoy en día con más de 75% de los OPSCC, siendo el factor responsable del incremento mayor a 25% de los HNSCC en los Estados Unidos, principalmente en hombres de edad media (1). Se ve una mayor relación entre el HPV y el HNSCC en países desarrollados (72%) que en países en desarrollo (13%) (6). La incidencia de HNSCC-HPV+ en hombres fue de 6,2/100.000 casos y en mujeres de 1,4/100.000 casos en los Estados Unidos entre 2004-2008 (8). A nivel global un 4,5% de los cánceres diagnosticados se pueden asociar al HPV, es decir unos 630.000 casos al año, de los cuales la mayoría son cánceres cervicales. Pero en ciertos países como los Estados Unidos la incidencia de HPV asociado a cánceres orofaríngeos sobrepasa la de

HPV asociado a cánceres cervicales. Se estima que la incidencia en adultos entre 70 y 83 años de OPSCC-HPV+ seguirá creciendo hasta 2045 de 16,8 a 29,0 por 100.000 casos (9). Según Vigneswaran y cols (2014) los HNSCC que afectan a los  $\frac{2}{3}$  anteriores de la lengua, el suelo de la boca, el paladar, la mucosa bucal, el sulco y en la encía no se consideran relacionados al HPV (1).

Los cánceres relacionados al HPV tienen un mejor pronóstico (4, 7, 10). Según Gormley, se piensa que podría estar relacionado a la edad menor de las personas con infección por HPV, una inmunidad anti-tumoral mayor y una comorbilidad menor. Aparte, se ha llegado a postular que los tumores HPV-positivos podrían tener una mayor sensibilidad al tratamiento radiológico (4). Según Barsouk y cols (2023) otros factores que mejoran el pronóstico de los cánceres relacionados al HPV son que típicamente presentan un menor número de mutaciones, tienen una mayor infiltración de linfocitos T en el microambiente tumoral (TME) y tienen una respuesta apoptótica adecuada (6). Aquellos pacientes HNSCC-HPV+ tienen una supervivencia general a los 5 años 30% mejor que aquellos pacientes HNSCC-HPV- (10).

Los tratamientos más apoyados en la investigación y más utilizados hoy en día para tratar los cánceres, incluyendo a los cánceres de cabeza y cuello son la cirugía, la radioterapia y la quimioterapia. Los tres métodos son métodos con desventajas y efectos adversos que afectan de manera impactante a la calidad de vida de los pacientes (2).

Cuando hablamos de un cáncer, debemos conocer el término *tumor microenvironment* (TME). Este ambiente próximo al foco principal del cáncer juega un rol importante en factores del cáncer cómo pueden ser la recurrencia y la metástasis (2, 11). Se compone de fibroblastos, células inmunes y componentes de la matriz extracelular, aparte de vasos sanguíneos, linfocitos, células inflamatorias y moléculas señalizadoras (12). Controla el desarrollo del tumor por medio de vías de señalización. Un TME anormal puede aumentar las tasas de infiltración y de metástasis, siendo caracterizado por una matriz extracelular con hipoxia, un pH disminuido, la presencia de ROS (*high reactive oxygen species*), una vascularización alterada y una presión osmótica aumentada. Es común ver alteraciones a nivel de la matriz extracelular en los TMEs; se ha visto que los tumores provocan la formación de matriz extracelular y que en cambio la matriz extracelular puede regular la proliferación de las células tumorales (2).

Las funciones generales de la matriz extracelular son la regulación del crecimiento, de la supervivencia, la diferenciación y de la movilidad celular. Estas funciones se llevan a cabo por medio de uniones a diferentes receptores como por ejemplo las integrinas. Además, la matriz extracelular se encarga de la correcta homeostasis celular y está adaptada según el órgano, para poder asegurar su función de una manera óptima (13, 14).

Los componentes más importantes de la matriz extracelular son el colágeno, la fibronectina, la elastina, la laminina, proteoglicanos y el ácido hialurónico (2, 14, 15).

El componente más abundante de la matriz extracelular es el colágeno, y existen 28 tipos de este (16). El colágeno es un componente que aporta fortaleza a los tejidos, además de formar redes, interactuar con otros componentes como por ejemplo los factores de crecimiento y tener funciones relacionadas a la adhesión celular (15).

La fibronectina es un componente que actúa uniendo proteínas estructurales de la matriz extracelular, además de interactuando con proteínas para desempeñar funciones regulatorias (2). Es un componente de unión entre lo que es la célula y lo que es la matriz extracelular (16). Una de sus funciones principales es la unión de integrina a la membrana celular, lo cual juega un rol importante en la transducción de señales intracelulares. Aparte, puede interactuar directamente con una variedad de factores de crecimiento, lo cual le da un importante rol en la transformación maligna de tumores (2).

Otro componente importante de las matrices extracelulares es la elastina, la cual se halla en tejidos con cierta necesidad elástica, cómo por ejemplo los pulmones, cartílago y los vasos sanguíneos (15). La elastina es una proteína que aporta elasticidad tisular, mientras que la laminina está involucrada en la maduración de la vascularización y conforma parte de la membrana basal junto al colágeno (2). Aparte de estas funciones, la laminina juega un rol en la diferenciación, adhesión y migración celular, además de la resistencia contra la apoptosis y el mantenimiento del fenotipo celular (16).

Los proteoglicanos son responsables de funciones biológicas y estructurales, aportando resistencia mecánica y siendo responsables de atrapar factores de crecimiento. Además de estas funciones, juegan un rol en la hidratación celular (14, 16).

El ácido hialurónico cumple funciones relacionadas con la permeabilidad vascular, la difusión de materiales y la promoción de curación de heridas, además del

equilibrio hídrico y osmótico de la matriz extracelular (2, 14). Las células tumorales son capaces de identificar al ácido hialurónico por medio del receptor de membranas CD44, lo cual les ayuda a migrar e infiltrarse (2).

Dado que los cánceres de cabeza y cuello, y más específicamente el cáncer oral, presentan una patología con una tasa de supervivencia de meramente 40-50% a los 5 años es de alta importancia el enfoque en el diagnóstico temprano de estas patologías (5, 17). La tasa de supervivencia sube hasta el 80% si se diagnostica el cáncer en un estadio temprano (I o II) (17). Se ha visto que una exploración intraoral puede detectar hasta un 99% de los cánceres orales (3).

Warnakulasuriya y cols (2021) señalan que la asintomatología en estadios tempranos es un factor clave que impide el diagnóstico temprano de las lesiones cancerosas de la cavidad oral, y hace hincapié en la necesidad de efectuar *screenings* para detectar esta patología de manera precoz (18). Lamentablemente un 50% de los cánceres orales se diagnostican ya estando en un estadio avanzado (III o IV), por esta falta de sintomatología en los estadios tempranos (3). Para efectuar un *screening* se realiza una exploración intraoral visual con ayuda de la palpación tanto de zonas orales como también de los ganglios linfáticos de la zona para determinar si el paciente presenta cáncer o alguna lesión precancerosa que podría aumentar su riesgo de desarrollar cáncer en un futuro (18).

En cuanto a las lesiones precancerosas más comunes destacan la leucoplasia, la eritroplasia, el liquen plano y la fibrosis submucosa (3). Según Woo y cols (2019), de estas, la leucoplasia, la eritroplasia y la fibrosis submucosa son lesiones con potencial de malignidad, mientras que el potencial maligno del liquen plano está considerado controversial. Estas lesiones se relacionan fuertemente con la displasia del epitelio oral (19). La tasa media estimada de transformación maligna de una leucoplasia oral es 9,8% (5).

Una displasia se considera aquella alteración que afecta a la estructura característica de la mucosa, y está relacionada con cambios en genes relacionados con la señalización celular, la angiogénesis, el crecimiento celular, la supervivencia celular y el control del ciclo celular (3). La presencia de displasia está asociada a una progresión hacia el cáncer (20).

Para diagnosticar una displasia se busca evidencia de una disminución de la polaridad de células basales, una estratificación irregular del epitelio, un aumento de las tasas de mitosis, mitosis superficiales y disqueratosis, entre otros (3).

Para establecer el diagnóstico definitivo de una lesión cancerosa se utiliza la biopsia con estudio histopatológico, pero se están desarrollando nuevos métodos de diagnóstico. Entre ellos se incluye el diagnóstico con biomarcadores de la saliva (3). Según Warnakulasuriya y cols (2021), este método diagnóstico está todavía en fase de desarrollo y la principal desventaja de él, es la falta de profesionales cualificados para su uso en los países en desarrollo (18).

Este método diagnóstico estudia niveles de citocinas, niveles de receptores como el CD44, y marcadores de DNA (Ácido desoxirribonucleico) y RNA (Ácido ribonucleico) que están sobreexpresados en procesos carcinógenos (3). Es un método que en un futuro podría aportar una manera fácil y accesible de diagnosticar cánceres orales antes de que estos estén en estadios más avanzados, pero falta una investigación más exhaustiva del uso de estos para afirmar su efectividad (18).

Aparte del diagnóstico temprano por medio de biopsia y biomarcadores, se pueden efectuar exploraciones clínicas con enfoque en las zonas más afectadas por estos cánceres, y pruebas con azul de toluidina o citologías exfoliativas. Cabe decir que el coste de efectuar estas exploraciones en toda la población sería considerablemente alto, por lo cual resulta eficaz enfocarse en los grupos de alto riesgo cómo pueden ser aquellos mayores de 40 años con hábitos de tabaquismo y alcoholismo (21).

#### **4.2. Virus del papiloma humano**

El HPV es un virus de doble cadena de tipo circular. Estos virus son capaces de infectar células epiteliales (9, 22). Hay una gran variedad de tipos de HPV, que se diferencian en un mínimo de 10% en sus secuencias de nucleótidos del gen L1. Se clasifican además según su afiliación a ubicaciones epiteliales o cutáneas (22). Además, se pueden clasificar en HPV de bajo riesgo o llamados no-oncogénicos, asociados al desarrollo de verrugas cutáneas y papilomas en el tracto respiratorio; y de alto riesgo, los cuales se asocian a neoplasias cervicales, anogenitales, y orofaríngeas (9). Se estima

que el tiempo de latencia entre una infección con HPV y el desarrollo de un HNSCC es de entre 10 y 20 años (23).

En el año 2018 se estimó que 80 millones de estadounidenses fueron infectados por este virus, con una estimada prevalencia de 7% de HPV oral en adultos entre 18 y 69 años, el cuál es más frecuentemente causado por sexo oral. La prevalencia en los Estados Unidos es mayor en hombres que en mujeres (11,5% vs 3,3% respectivamente). La prevalencia mayor se halla en países de la Oceanía y África, seguido de Europa, las Américas y Asia (9).

El gen L1 es el que más se conserva entre todos los tipos de HPV y es aquel en el cual se basó la clasificación de estos virus en una entidad separada de virus denominada Papillomaviridae. Los HPV de las mucosas comúnmente pertenecen al genus alpha, mientras que los HPV cutáneos se incluyen en el genus beta. Adicionalmente, existen HPV asociados a papilomas y verrugas de la piel, los cuales se incluyen en los genes gamma, mu y nu (22).

El HPV tipo 16, el cual es el de mayor riesgo para el desarrollo del cáncer oral, pertenece al genus alpha (24).

Estos virus, en general, tienen aproximadamente 8.000 pares de bases (bp) y sus secuencias codificantes de proteínas (ORFs) se restringen a una sola cadena (22).

Un ORF es lo suficientemente grande para codificar para una proteína, y las ORFs se clasifican en *early* (E) y *late* (L), según los genes que contiene. Los HPV comúnmente contienen los ORFs E1, E2, E4, E5, E6 y E7, además de L1 y L2, que codifican las proteínas L1 y L2 (22, 24). E1 y E2 son conocidos por regular la replicación del virus, mientras que E4-E7 están involucrados en la oncogénesis, teniendo el E5 influencia sobre la diferenciación de queratinocitos y E6 y E7 influencia sobre *checkpoints* celulares (24). Algunos HPV específicos contienen también los ORFs E3 y E8, pero se desconoce su función (22). En general se observa que los oncogenes no tienen funciones enzimáticas, sino que interactúan con proteínas celulares y por este medio causan desregulaciones en ciertas vías biológicas en las células del huésped. Esto causa alteraciones en el crecimiento y en la evasión inmune, entre otros (25).

El virión se compone de 360 copias de la proteína L1 (*major capsid protein*) y hasta 72 moléculas de la proteína L2 (*minor capsid protein*) (22). La región no codificante está involucrada en la adhesión de factores de transcripción y replicación, permitiendo

la transcripción por medio de la RNA-polimerasa II (24). Esta región se halla entre L1 y E6 (22).

En los cánceres este virus se encuentra tanto en forma episomal, integrada cómo mixta (22, 24). La integración al genoma del huésped normalmente se da en la región E1 o E2 del virus (22). Estas regiones regulan la replicación del virus, primero de forma lenta en la capa basal de células indiferenciadas, y luego de manera más rápida cuando ya queda integrado en la capa suprabasal (24). Según Rautawa y cols (2012) este virus infecta a células basales indiferenciadas del epitelio después de que esté sufre trauma o erosión. Después se da una replicación del DNA vegetativa y se amplifica el genoma a un número de más de 1.000 copias por célula. Esto induce la expresión de los genes L virales y la producción de partículas virales maduras que son liberadas de las células en las capas más superficiales del epitelio diferenciado (22). Demuestran tropismo específico hacia el epitelio escamoso (26).

Se ha visto que la integración del HPV a una célula aumenta la tasa de crecimiento de la célula afectada (27). Al integrarse en el genoma del huésped se inhibe la zona E2 del virus, la cual se encarga teóricamente de la represión de E6 y E7. Es decir, que, al integrarse, estas dos zonas (E6 y E7) aumentan su actividad, lo cual aumenta la tasa de replicación del virus, por su capacidad de aumentar la tasa de proliferación celular, inhibir la apoptosis, e inducir la angiogénesis, infiltración y metástasis (24).

La proteína E7 del HPV es capaz de inducir la proliferación celular por medio de varias vías por su capacidad de inhibir la actividad de los inhibidores p21 y p27 de la quinasa dependiente de ciclina (CDK), de activar CDKs y de desestabilizar la proteína Rb supresora de tumores (22).

E6 y E7 pueden además destruir la proteína supresora de tumores p53 (27). Según Pal (2020), el E6 tiene como función oncogénica más importante la capacidad de degradación del p53, lo cual consigue a través de la unión a la proteína E6AP, mientras que el E7 tiene como capacidad oncogénica más importante su inhibición de la proteína pRb, la cual está normalmente regulando el proceso de la apoptosis por medio de su unión a los factores de transcripción de la familia E2F. Esta unión permanece siempre y cuando las células no están preparadas para pasar de la fase G1 a la fase S del ciclo celular. E7 interactúa con pRb y causa la liberación de los factores de transcripción E2F, lo cual induce que las células pasen a la fase S del ciclo celular, aunque no estén

preparadas para ello, además de inducir la liberación del gen p16INK4A, que nos indica que hay infección de HPV de alto riesgo (24, 28).

Es por estas dos proteínas (E6 y E7) que se relacionan los HPV con el desarrollo de los cánceres, dado que la alteración que causan estas proteínas sobre las proteínas supresoras de tumores tiene como consecuencia alteraciones en la proliferación, estabilidad y apoptosis (26).

Como se ha comentado anteriormente, se ha visto en los últimos años una asociación creciente de cánceres de cabeza y cuello y el HPV, más específicamente el HPV 16 (22).

Los HPV de alto riesgo se han relacionado frecuentemente con los cánceres anogenitales y orofaríngeos, pero sigue siendo una pregunta de investigación la posible relación con el cáncer oral. Se consideran como HPV de mayor riesgo al subtipo 16 (4, 7). Aparte de este, el tipo 18, 31 y el 33 también han sido relacionados a un mayor riesgo del desarrollo de cánceres (7).

Sin embargo, es importante destacar que estudios que han relacionado los HPV con el OSCC han podido tener sesgos a la hora de la toma de la muestra, a la hora de definir la zona anatómica, de juntar cánceres de varias zonas de la cabeza y cuello en un solo grupo, y una ausencia de evidencia que justifica que el HPV esté activo oncogénicamente, lo cual se debe valorar según la expresión de los oncogenes E6 y E7 (26).

Existe una vacuna para el HPV, la cual ha demostrado eficiencia en relación con la reducción de la incidencia de cáncer cervical, pero este efecto no se ha podido demostrar efectivamente en los cánceres orofaríngeos (23). Solamente un 51% de los adolescentes entre 13 y 15 años han sido vacunados con las tres dosis contra el HPV en los Estados Unidos, un valor que no llega al valor objetivo de 80%. Se ha visto aparte que la polémica del COVID-19 ha causado una reducción en las tasas de vacunación infantiles afectando también a las vacunaciones del HPV (7, 9).

En un metaanálisis se vio que de 12.263 muestras de HNSCC un 45,8% de cánceres orofaríngeos, un 24,4% de cánceres orales y un 22,1% de cánceres laríngeos se atribuyeron al HPV. Estos valores se basaron en la detección de DNA del HPV por medio de PCR (Reacción en cadena de la polimerasa) en las muestras de los tumores. Se evaluó en algunos casos de OPSCC también la expresión de p16 o de mRNA de la proteína E6 o



E7 del HPV. Esta evaluación dual no se efectuó en la mayoría de las muestras de cánceres orales o cánceres laríngeos, aunque se considera una técnica de mayor valor para el establecimiento de la causalidad (10, 23).

El mecanismo de entrada del HPV a las células del huésped se basa en la interacción de este virus con los proteoglicanos de heparán sulfato (HSPG). La eliminación de estos proteoglicanos resulta en una reducción del 80-90% de la unión del virus a las células del huésped (29). Se especula además que la  $\alpha 6$ -integrina y la 5-laminina están involucradas en la unión y en la internalización del virus hacia la parte interior de las células (28, 29). En el caso de los cánceres cervicales, el HPV infecta exclusivamente a las células basales, en cuyo interior puede mantenerse latente. En la cavidad oral hay una zona en la cual hay una exposición de células basales con el medio ambiente exterior; la bolsa periodontal. Cuando esta zona está inflamada se miden niveles de carga viral mayores y hay un riesgo mayor de la transmisión del virus. Se especula que la inflamación crónica de esta zona y la consecuente proliferación continua podría favorecer la replicación del HPV (30).

#### **4.3. Transcriptómica**

La transcriptómica es una rama de la investigación que pertenece al área de la genómica. La genómica por su parte es una ciencia que se concentra en la investigación de las secuencias de genes a nivel del genoma (*DNA*), transcriptoma (*mRNA*) y proteoma (proteínas), además del estudio de las interacciones entre estos componentes y en el entendimiento de vías metabólicas de los organismos. Se dedica a documentar las secuencias génicas de los organismos (31). Es un área interdisciplinar que se ocupa de estudiar el genoma completo (32).

La historia de la genómica empezó con Fred Sanger, dado que este científico fue el primero en secuenciar un genoma completo de un bacteriófago en el año 1977 y gracias a los avances tecnológicos de las últimas cuatro décadas, se ha podido desarrollar de manera sustancial. Hoy en día ayuda en el desarrollo de métodos preventivos y de fármacos, pudiendo por ejemplo la farmacogenómica ayudar en valorar si un fármaco será eficiente en una persona específica (31, 32).

La genómica funcional es un área de la genómica, la cual se divide en genómica estructural, funcional y comparativa, que se ocupa de estudiar la expresión génica de organismos y que caracteriza al complejo de genes existentes en el genoma. A esta rama de la genómica se le conoce con el término transcriptómica (31).

El transcriptoma se refiere al conjunto de moléculas de *RNA*, es decir *mRNA*, *rRNA*, *tRNA* y subtipos de *RNA* no codificantes (31, 33). Cuando se estudia el transcriptoma de tejidos, organismos y células, la expresión génica, además de las interacciones y dinámicas de ellos en un momento dado, se habla de la transcriptómica. Hoy en día hay un método de secuenciación avanzado que ayuda en el estudio del transcriptoma - el RNA-seq (31, 34).

El análisis del transcriptoma nos ayuda a entender procesos moleculares involucrados en ciertos procesos biológicos y procesos patológicos (33, 35).

#### **4.4. RNA-Seq**

El *RNA-Seq* es una herramienta que forma parte de las tecnologías *NGS* (*Next Generation Sequencing*), o también llamadas *high-throughput* o *deep sequencing*, que revolucionaron el ámbito de la biología molecular (36, 37). El *RNA-Seq* nos permite determinar niveles de expresión, el *splicing* diferencial y la expresión específica de alelos con una mayor precisión y a un coste más eficiente que las técnicas antiguas. Permite estudiar y definir el transcriptoma completo de una célula individual. El *RNA-Seq*, además puede ayudar en identificar isoformas de *splicing* nuevas, además de caracterizarlas y cuantificarlas, por lo cual es un método de alto beneficio para la rama de la transcriptómica, dado que los métodos *microarray* no podían detectar transcriptomas nuevos (37). Hoy en día es posible aplicar esta técnica a células únicas, lo cual puede permitir estudiar las diferencias de expresión génica entre grupos de células (36).

Además, el *RNA-Seq* puede identificar transcritos no caracterizados, fusiones génicas y *novel splicing* (36). Asimismo,, el *RNA-Seq* ayuda a definir los SNPs (*Single Nucleotide Polymorphisms*) y a determinar la anotación génica correcta (37).

Los pasos de los estudios de transcriptómica comúnmente incluyen el análisis cuantitativo, en el cual se determina la cantidad de genes expresados en una muestra

de *RNA-seq*, el análisis de la expresión diferencial para comparar la expresión génica entre muestras y células diferentes, y la anotación, en la cual se determina la función biológica de cada transcriptoma (36).

El objetivo de un análisis diferencial es averiguar diferencias en los transcriptomas de tejidos y células, además de la sobre- e infraexpresión de los genes de ellos, es decir se centra en evaluar la diferencia de expresión génica entre muestras. Es una técnica, que nos ayuda a entender de mejor manera los procesos biológicos, la diferenciación celular y la respuesta inmunológica a patógenos, entre otros (36).

Al hacer un análisis diferencial los resultados se clasifican por su *Fold Change* (FC), lo cual es un valor de diferencia relativa que se utiliza para estudiar la expresión diferencial génica (38). Es decir, un *Fold Change* alto, quiere decir que un gen está altamente sobreexpresado. Un *Fold Change* negativo, por ejemplo, quiere decir que un gen está infraexpresado.

*RNA-Seq* nos permite profundizar nuestro conocimiento del contenido del transcriptoma génico y nos provee con un método que nos aporta perfiles de expresión génica con desviaciones técnicas menores (36, 38).

Se puede decir que los métodos *NGS*, entre ellos incluido el *RNA-seq*, han revolucionado el ámbito de la biología molecular (37).

#### **4.5. Biomarcadores de pronóstico**

Los biomarcadores de pronóstico son secuencias moleculares que ayudan a establecer un pronóstico en pacientes con cierta patología. Los factores que pueden pronosticar son la recurrencia y la mortalidad. Con estos biomarcadores será posible intervenir de manera temprana en pacientes identificados de alto riesgo, pudiendo por ejemplo establecer en cuáles pacientes se debería intervenir de manera más agresiva (39).

Para la búsqueda de información sobre posibles biomarcadores relacionados a cierta patología, se puede realizar un análisis de curvas ROC (*Receiver operating characteristics*).

En sí, las curvas ROC sirven para evaluar la relación entre verdaderos positivos (sensibilidad) y verdaderos negativos (especificidad) de un clasificador binario (38).

El valor que se utiliza para determinar la exactitud de una prueba se valora con el AUC (*area under ROC curve*). En el eje y se encuentra la fracción de verdaderos positivos y en el eje x la fracción de verdaderos negativos. Si estos dos solapan completamente en cualquier umbral, nos encontraríamos con una curva ROC diagonal con un valor de AUC de 0,5, es decir 50% del área total. Esto nos indica que una prueba no tiene valor, o que su habilidad discriminativa iguala a la casualidad, es decir que no la tiene. En comparación, una curva ROC con valor máximo, sería aquella que tiene un AUC de 1,0, es decir un 100% del área total. En este caso se habría registrado el punto óptimo de operación que se iguala a un valor de fracción de verdadero positivo de 1, y un valor de fracción de verdadero negativo de 0; y no habría solapamiento (40). Aquel valor sería el de habilidad discriminativa perfecta (41).

En nuestro estudio este análisis sirve para establecer cuáles genes nos servirán de biomarcadores de pronóstico con alta sensibilidad (identificación de positivos verdaderos) y especificidad (identificación de negativos verdaderos) (42).

#### **4.6. Análisis de enriquecimiento**

El análisis de enriquecimiento sirve para interpretar datos de expresión génica, de tal manera que podamos entender de manera más profunda los mecanismos biológicos en los cuales los genes expresados están involucrados (43).

Este método intenta averiguar si un conjunto de genes está asociado a una característica. Para ello calcula un *enrichment score* para el conjunto de genes y se define la significancia de este valor respecto sus mecanismos biológicos (44). Muchos métodos de análisis de enriquecimiento utilizan estadísticas de solapamiento para determinar si los genes diferenciales están enriquecidos para un término ontológico o una vía biológica (45).

En este trabajo utilizamos dos softwares para efectuar este análisis sobre nuestro conjunto de genes; el GSEA (*Gene set enrichment análisis*) y el DAVID (*Database for Annotation, Visualization and Integrated Discovery*).

El GSEA se enfoca en analizar conjuntos de datos que comparten funciones biológicas, ubicaciones cromosómicas y regulación. Este análisis puede servir para

identificar vías metabólicas alteradas e identificar procesos involucrados en la aparición de diversas patologías (43).

El método GSEA incluye 3 elementos;

1. La calculación del *Enrichment Score* (ES).
2. La estimación del nivel de significancia del ES.
3. El ajuste para pruebas de hipótesis múltiples.

El GSEA fue creado con el fin de disminuir problemas asociados a métodos previos que se enfocaban únicamente en algunos genes de ambos extremos de la lista. Estos problemas incluían que muchos genes no llegaban a ser estadísticamente significativos después del tercer paso; se obtenía una lista muy larga de genes estadísticamente significativos, pero sin correlación biológica; no se podía estudiar adecuadamente los efectos sobre los *pathways*; había una gran probabilidad de que dos grupos que estudiaban el mismo sistema biológico obtuvieran listas diferentes de genes estadísticamente significativos. Dado que el GSEA analiza todos los genes, y no solamente los cuales con cierto FDR y pvalor, y preserva la correlación entre genes, provee un método mejor en comparación con otros análisis de enriquecimiento (43).

El DAVID es una herramienta que ayuda a entender el significado biológico de una lista de genes diferenciales. Sirve para la anotación funcional y para efectuar un análisis de enriquecimiento de un conjunto de genes diferenciales, para poder entender más profundamente su significancia biológica. Las herramientas de este software pueden entre otros identificar funciones biológicas enriquecidas, o términos GO, descubrir grupos de genes relacionados en función y enriquecidos y agrupar términos de anotación redundantes (45, 46).

## 5. JUSTIFICACIÓN

El carcinoma de células escamosas de cabeza y cuello representa uno de los cánceres más comunes en el mundo, con una tasa de supervivencia a los 5 años de 50% y una incidencia creciente (47).

Dentro de los factores de riesgo, el consumo de tabaco y alcohol ha sido clásicamente reconocido como el determinante principal. Sin embargo, en los últimos años se ha visto que la infección por HPV es un factor etiológico crucial, especialmente en los cánceres de la orofaringe, aumentando su incidencia particularmente en los países desarrollados (4).

El diagnóstico y tratamiento temprano del carcinoma de células escamosas de cabeza y cuello son esenciales para mejorar las tasas de supervivencia, las cuales pueden aumentar significativamente si se detecta el cáncer en etapas iniciales. No obstante, la falta de síntomas en fases tempranas y la alta tasa de diagnósticos en etapas avanzadas complican este objetivo. La detección mediante biomarcadores es una metodología aún en desarrollo.

En este contexto, el estudio de la expresión génica en cánceres asociados al HPV representa una oportunidad para desarrollar estrategias diagnósticas y terapéuticas más efectivas. Un metaanálisis centrado en la expresión génica de los cánceres de cabeza y cuello relacionados con el HPV permitirá identificar patrones moleculares específicos que pueden diferenciar estos tumores de aquellos no relacionados con el virus.

Esta diferenciación podría facilitar el desarrollo de biomarcadores diagnósticos y tratamientos personalizados que puedan mejorar los resultados clínicos y reducir la morbilidad y mortalidad asociadas a estos cánceres.

Además, dado que los cánceres relacionados con el HPV tienden a tener un mejor pronóstico y respuesta al tratamiento radiológico, comprender los mecanismos moleculares subyacentes a esta diferencia puede ofrecer información valiosos para mejorar el manejo clínico de los pacientes con carcinoma de células escamosas de cabeza y cuello.

Este trabajo de fin de Grado se enmarca en el Objetivo de Desarrollo Sostenible (ODS) número 3, "Salud y Bienestar". Al profundizar en la investigación del carcinoma

oral, se busca contribuir al objetivo global de promover una vida saludable y bienestar para todos. Los resultados obtenidos podrían tener aplicaciones clínicas concretas, mejorando la detección temprana y el desarrollo de terapias personalizadas, alineándose así con la promoción de la salud y la prevención de las enfermedades.

## **6. HIPÓTESIS**

Existen diferencias significativas en la expresión génica entre pacientes con carcinoma de células escamosas de cabeza y cuello que son HPV positivos en comparación con aquellos que son HPV negativos. Se espera que los genes diferencialmente expresados tengan una influencia positiva sobre el pronóstico.

## **7. OBJETIVOS**

### **7.1. General**

Identificar y caracterizar los genes diferencialmente expresados en el cáncer de células escamosas de cabeza y cuello en pacientes HPV positivos en comparación con pacientes HPV negativo, para mejorar la comprensión de las bases moleculares de la enfermedad y desarrollar potenciales biomarcadores diagnósticos y terapéuticos.

### **7.2. Específicos**

1. Detectar genes diferenciales en muestras de carcinoma de células escamosas de cabeza y cuello, comparando pacientes HPV positivos (HNSCC-HPV+) con pacientes HPV negativos (HNSCC-HPV-).
2. Investigar las funciones biológicas asociadas a los genes diferenciales.
3. Evaluar el impacto de los genes diferenciales en la supervivencia de los pacientes con carcinoma de células escamosas de cabeza y cuello.
4. Identificar posibles biomarcadores con potencial predictivo para el carcinoma de células escamosas de cabeza y cuello.



## **8. MATERIALES Y MÉTODOS**

Realizamos un análisis bioinformático de expresión diferencial en muestras de pacientes con carcinoma de células escamosas de cabeza y cuello, utilizando datos de la base de datos TCGA (*The Cancer Genome Atlas Project*), con el propósito de identificar genes diferencialmente expresados en pacientes HPV positivos en comparación con pacientes HPV negativo.

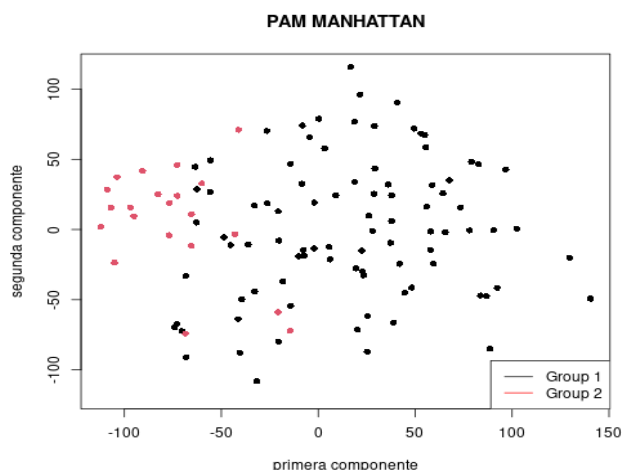
### **8.1. Diseño del estudio**

#### **8.1.1. Selección de los datos**

Para realizar este estudio se utilizaron datos de la base de datos TCGA (*The Cancer Genome Atlas Project*), <http://tcga.cancer.gov> Este portal permite el análisis y la descarga de una serie extensa de datos 'ómicos' relacionados con el cáncer. Creado por el NCI (*National Cancer Institute*) y el Instituto Nacional de Investigación del Genoma, ha generado más de 2,5 petabytes de datos genómicos, epigenómicos, transcriptómicos y proteómicos. Estos datos ya han conducido a mejoras en la capacidad para diagnosticar, tratar y prevenir el cáncer.

Dentro del portal TCGA se filtraron las muestras según el cáncer de enfoque para nuestro estudio: cáncer de cabeza y cuello. Se obtuvieron 530 muestras de carcinoma de células escamosas del estudio *Firehose Legacy data atlas* de las cuales utilizamos 112 que contenían datos sobre la presencia del HPV p16 o ausencia de HPV en las muestras.

Una vez realizada la selección de las muestras se realizó un Análisis de Componentes Principales (ACP) de las 112 muestras: 20 muestras presentaban una infección con el HPV 16 (Grupo 2) y 92 muestras presentaban un HPV negativo (Grupo 1). Se demostró una separación de ambos grupos, la cual se ve reflejada en la Figura 1.



**Figura 1.** Análisis de Componentes Principales de 112 muestras HPV positivo (20 muestras: Grupo 2) y HPV negativo (92 muestras: Grupo 1)

## 8.2. Análisis de expresión diferencial

Para el análisis diferencial, se compararon las 20 muestras pertenecientes a pacientes con cáncer oral y HPV *versus* las 92 pertenecientes a pacientes con cáncer oral sin infección por HPV.

Para llevar a cabo el análisis se usaron los softwares EdgeR y Limma (Bioconductor), donde se estimaron lecturas de TPMs normalizadas y se identificaron genes diferencialmente expresados (DEGs) a través de métodos estadísticos (i.e., *linear modeled and empirical Bayes*) según la metodología de Law y cols. Se consideraron aquellos genes con un FDR (*false discovery rate*)  $\leq 0,01$  como expresados diferencialmente (48).

## 8.3. Análisis de curvas ROC

Para obtener posibles biomarcadores se generaron curvas ROC. Utilizamos el paquete *genefilters* del proyecto Bioconductor con la función de “*rowpAUCs*” aplicada en los valores normalizados (TPMs) en ambos grupos genéticos (independientes).

Un área debajo de la curva ROC de 0,5 representa una prueba sin habilidad discriminativa (no mejor que casualidad), mientras que un AUC de 1 representa una prueba de habilidad discriminativa perfecta (41).

Los valores de AUC que consideramos significativos y con habilidad discriminativa fueron aquellos  $\geq 0,7$ .

#### **8.4. Análisis de enriquecimiento**

Para el estudio de procesos biológicos de los genes mediante análisis de enriquecimiento se usaron dos programas el GSEA y el DAVID.

Se usó el software GSEA con la base de datos MSigDB v7.5 para el estudio de procesos biológicos asociados a todos los genes que habían tenido un cambio de expresión independiente si se detectaron como significativos o no, teniendo una visión más amplia de todo lo que está puede estar cambiando en el transcriptoma del grupo HPV positivo. Aquellos procesos biológicos con un FDR  $\leq 0,01$  se consideraron cómo estadísticamente significativos (42, 49).

Se utilizó el software DAVID para observar los procesos biológicos asociados a los genes diferencialmente expresados (DEGs) (44, 45).

Aquellos procesos biológicos con un FDR  $\leq 0,01$  se consideraron cómo estadísticamente significativos en el DAVID realizado con los genes sobreexpresados, mientras que aquellos procesos biológicos con un pvalor  $\leq 0,01$  se consideraron estadísticamente significativos en el DAVID realizado con los genes infraexpresados.

Además, se utilizó ClueGO un complemento Cytoscape para la interpretación gráfica de las funciones biológicas no redundantes de nuestros genes de interés. ClueGO integra términos de *Gene Ontology* (GO), así como vías KEGG/BioCarta, y crea una red de términos GO/vías funcionalmente organizadas (50).

#### **8.5. Análisis de supervivencia**

Utilizamos el paquete R “Survival” (51) para analizar la información clínica de las 112 muestras de carcinoma de células escamosas de cabeza y cuellos de TCGA del estudio *Firehose Legacy data atlas* con información relevante a la infección con HPV. Se analizó los genes que se hallaron cómo relevantes según nuestro estudio y como ellos pueden estar influenciando a la supervivencia general de la enfermedad. Este análisis se centra clásicamente en estudiar el tiempo traspasado hasta la ocurrencia de un evento

preestablecido, en este caso el fallecimiento. Se ha utilizado una prueba de Fisher para ver si existe significancia entre la expresión de genes de interés y la supervivencia en pacientes con HNSCC-HPV+.

## 9. Resultados

### 9.1. Análisis de expresión diferencial

Se han encontrado un total de 5.955 genes diferenciales con un FDR (*False Discovery Rate*) de menos de 0,01 en 20 muestras de tejido oral en pacientes HNSCC-HPV+ frente a las 92 muestras de tejido oral en pacientes HPV negativo. De ellos 5.203 genes fueron sobreexpresados (*Fold Change* positivo) y 752 genes infraexpresados (*Fold Change* negativo) en los pacientes HSNCC-HPV+. En la tabla 1 se observa los 20 primeros genes diferenciales con mayor sobreexpresión y en la tabla 2 se observan los 20 primeros genes diferenciales con menor infraexpresión. En el anexo se pueden apreciar los 5.955 DEGS con FDR<0,01 (Tabla 1 y 2 anexo).

**Tabla 1.** Los 20 genes diferenciales con FDR <0,01 con mayor sobreexpresión (*Fold Change* positivo)

ID	Nombre completo	<i>fold change (FC)* (log2)</i>	pvalor	padj (FDR)
COLCA1	Colorectal cancer associated 1	2,995	5,652E-09	2,455E-06
SMC1B	Structural Maintenance Of Chromosomes 1B	2,663	9,705E-07	4,585E-05
NOS2	Nitric Oxide Synthase 2	2,528	5,668E-08	8,289E-06
ZFR2	Zinc Finger RNA Binding Protein 2	2,465	1,254E-04	1,511E-03
SYCP2	Synaptonemal Complex Protein 2	2,427	1,504E-07	1,444E-05
CLGN	Calmequin	2,382	2,617E-07	1,997E-05
CD8B2	CD8B Family Member 2	2,367	3,847E-05	6,178E-04
KRT19	Keratin 19	2,355	3,230E-04	3,097E-03
MUC4	Mucin 4, Cell Surface Associated	2,334	1,278E-04	1,531E-03
YBX2	Y-Box Binding Protein 2	2,260	3,462E-06	1,099E-04
BTNL9	Butyrophilin Like 9	2,241	4,790E-07	2,885E-05
MAP7D2	MAP7 Domain Containing 2	2,238	5,408E-05	7,951E-04
ABCA17P	ATP Binding Cassette Subfamily A Member 17, Pseudogene	2,218	1,410E-05	2,916E-04

<b>TDRD5</b>	<b>Tudor Domain Containing 5</b>	2,172	1,804E-05	3,483E-04
<b>IRX6</b>	<b>Iroquois Homeobox 6</b>	2,162	3,567E-05	5,837E-04
<b>ADARB2</b>	<b>Adenosine Deaminase RNA Specific B2 (Inactive)</b>	2,143	5,056E-07	2,992E-05
<b>IL17REL</b>	<b>Interleukin 17 Receptor E Like</b>	2,134	1,731E-05	3,378E-04
<b>DMRTA2</b>	<b>DMRT Like Family A2</b>	2,134	3,685E-04	3,436E-03
<b>WNK2</b>	<b>WNK Lysine Deficient Protein Kinase 2</b>	2,130	3,489E-06	1,101E-04
<b>FOXA1</b>	<b>Forkhead box A1</b>	2,119	2,327E-06	8,293E-05

\*Fold Change se refiere al grupo HPV+ en comparación con el HPV-

**Tabla 2.** Los 20 genes diferenciales con FDR <0,01 con menor infraexpresión (*Fold Change* negativo)

<b>ID</b>	<b>Nombre completo</b>	<b><i>Fold Change</i> (FC)* (log2)</b>	<b>pvalor</b>	<b>Padj (FDR)</b>
<b>SPRR2G</b>	<b>Small Proline Rich Protein 2G</b>	-3,091	1,208E-06	5,308E-05
<b>KRTDAP</b>	<b>Keratinocyte Differentiation Associated Protein</b>	-2,905	1,900E-05	3,605E-04
<b>KLK5</b>	<b>Kallikrein Related Peptidase 5</b>	-2,810	8,350E-06	1,980E-04
<b>SBSN</b>	<b>Suprabasin</b>	-2,601	9,770E-07	4,598E-05
<b>WFDC12</b>	<b>WAP Four-Disulfide Core Domain 12</b>	-2,452	7,558E-06	1,844E-04
<b>CASP14</b>	<b>Caspase 14</b>	-2,367	1,080E-03	7,938E-03
<b>REG</b>	<b>Epiregulin</b>	-2,357	1,398E-05	2,897E-04
<b>KLK7</b>	<b>Kallikrein related peptidase 7</b>	-2,342	9,056E-05	1,186E-03
<b>PRR9</b>	<b>Proline Rich 9</b>	-2,244	1,907E-04	2,071E-03
<b>KRT75</b>	<b>Keratin 75</b>	-2,228	1,658E-05	3,278E-04
<b>CPA4</b>	<b>Carboxypeptidase A4</b>	-2,212	7,010E-06	1,754E-04
<b>KRT14</b>	<b>Keratin 14</b>	-2,193	3,133E-07	2,221E-05
<b>LINC01404</b>	<b>long intergenic non-protein coding RNA 1404</b>	-2,192	2,120E-08	5,042E-06
<b>FOLR3</b>	<b>Folate Receptor Gamma</b>	-2,191	8,405E-07	4,185E-05
<b>DSG1</b>	<b>Desmoglein 1</b>	-2,175	7,577E-04	6,021E-03
<b>LCE3D</b>	<b>Late Cornified Envelope 3D</b>	-2,159	2,019E-04	2,162E-03

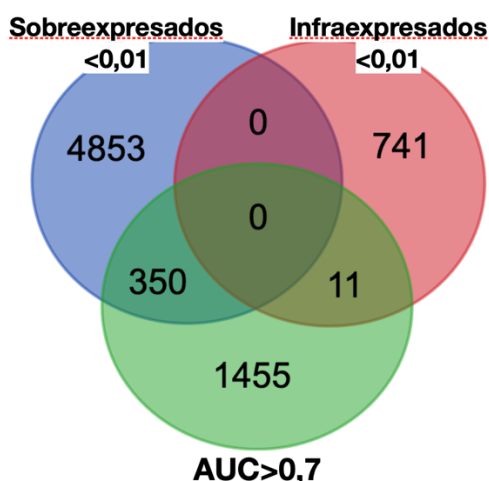
<b>MMP3</b>	<b>Matrix metallopeptidase 3</b>	-2,131	6,223E-05	8,873E-04
<b>KRT16P6</b>	<b>Keratin 16 Pseudogene 6</b>	-2,083	6,677E-05	9,377E-04
<b>CRCT1</b>	<b>Cysteine Rich C-Terminal 1</b>	-2,056	3,821E-04	3,541E-03
<b>PAEP</b>	<b>Progestagen Associated Endometrial Protein</b>	-2,054	9,951E-07	4,628E-05

\*Fold Change se refiere al grupo HPV+ en comparación con el HPV-

## 9.2. Análisis de curvas ROC

Se realizó un análisis de curvas ROC de todos los genes detectados obteniendo un total de 1.816 genes con AUC >0,7 considerándose buenos marcadores biológicos. En el anexo se observan los 1.816 genes con AUC >0,7 (Tabla 3 anexo).

En la figura 2 se puede apreciar los genes diferenciales (FDR <0,01) sobre e infraexpresados comunes con AUC >0,7. De los 1.816 genes con AUC >0,7 fueron 350 genes sobreexpresados comunes (Tabla 4 anexo) destacando los 20 genes con mayor AUC (Tabla 3) y 11 genes infraexpresados comunes (Tabla 4).



**Figura 2.** Diagrama de Venn de AUC >0,7, Genes sobreexpresados y Genes infraexpresados

**Tabla 3.** Los 20 genes diferenciales sobreexpresados (FDR <0,01) con mayor AUC (AUC >0,7).

ID	Nombre completo	AUC
THUMPD3-AS1	THUMPD3 Antisense RNA 1	0,817
TRAPPC2	Trafficking Protein Particle Complex Subunit 2	0,810
RNF32-AS1	RNF32 Antisense RNA 1	0,802
ALOX12-AS1	ALOX12 Antisense RNA 1	0,792

HLCS-AS1	HLCS Antisense RNA 1	0,792
FZD9	Frizzled Class Receptor 9	0,791
CDKL3	Cyclin Dependent Kinase Like 3	0,789
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit	0,788
LEAP2	Liver Enriched Antimicrobial Peptide 2	0,786
E2F1	E2F Transcription Factor 1	0,784
FAM98C	Family With Sequence Similarity 98 Member C	0,778
UBXN11	UBX Domain Protein 11	0,777
IFT27	Intraflagellar Transport 27	0,777
HDAC10	Histone Deacetylase 10	0,777
KDM5A	Lysine Demethylase 5A	0,776
POLR3F	RNA Polymerase III Subunit F	0,776
NPRL2	NPR2 Like, GATOR1 Complex Subunit	0,773
CDK5RAP1	CDK5 Regulatory Subunit Associated Protein 1	0,772
RNF32-DT	RNF32 Divergent Transcript	0,772
ATAT1	Alpha Tubulin Acetyltransferase 1	0,770

**Tabla 4.** Los 11 genes diferenciales infraexpresados con mayor AUC (AUC >0,7)

ID	Nombre completo	AUC
EHD2	EH Domain Containing 2	0,771
GDPD2	Glycerophosphodiester Phosphodiesterase Domain Containing 2	0,754
PPP2R2C	Protein Phosphatase 2 Regulatory Subunit Bgamma	0,734
HTRA1	HtrA Serine Peptidase 1	0,731
CNGB1	Cyclic Nucleotide Gated Channel Subunit Beta 1	0,723
MMP3	Matrix Metallopeptidase 3	0,719
SPRR2G	Small Proline Rich Protein 2G	0,707
GSDMA	Gasdermin A	0,706
CLMP	CXADR Like Membrane Protein	0,702
WFDC12	WAP Four-Disulfide Core Domain 12	0,702
TNC	Tenascin C	0,701

### 9.3. Análisis de enriquecimiento con GSEA software

Se realizó un análisis con el software GSEA para visualizar las funciones biológicas de todos los genes con cambio de expresión. Se observan 1.731 procesos biológicos con FDR <0,01 activándose (LOR positivo) y 26 inhibiéndose (LOR negativo) en las muestras HPV positivas (tabla 5 anexo).

Se observaron los 20 procesos biológicos activados con menor FDR (Tabla 5) y los 26 procesos biológicos inactivados con menor FDR (Tabla 6). Destacamos en color naranja en la tabla 5 los procesos biológicos activados en pacientes HSNCC-HPV+ de



interés para nuestro estudio relacionados con la replicación de DNA, el *RNA splicing*, la segregación de cromosomas y la recombinación de DNA y destacamos en la tabla 6 procesos biológicos inhibidos en muestras con HSNCC-HPV+ relacionados con la queratinización, el desarrollo de la epidermis y la organización de la matriz extracelular. Destacamos estos procesos por su vínculo a procesos que podrían estar involucrados en la oncogénesis.

**Tabla 5.** Los 20 procesos biológicos activados con menor FDR según GSEA

GO TERM	Count	LOR	pvalor	FDR	Función
GO:0006397	498	1,016	1,565E-105	9,310E-101	mRNA processing
GO:0008380	428	1,002	6,383E-97	1,898E-92	RNA splicing
GO:0000375	342	1,030	1,971E-85	3,908E-81	RNA splicing, via transesterification reactions
GO:0000377	339	1,028	3,024E-84	2,997E-80	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
GO:0000398	339	1,028	3,024E-84	2,997E-80	mRNA splicing, via spliceosome
GO:0016570	448	1,013	2,959E-84	2,997E-80	histone modification
GO:0016569	468	1,000	7,685E-84	6,530E-80	covalent chromatin modification
GO:0007059	308	1,079	5,954E-74	4,426E-70	chromosome segregation
GO:0018205	390	1,018	2,326E-73	1,537E-69	peptidyl-lysine modification
GO:0044782	363	0,975	4,680E-70	2,784E-66	cilium organization
GO:0006261	153	1,414	4,934E-68	2,667E-64	DNA-dependent DNA replication
GO:0006260	276	1,280	1,138E-66	5,641E-63	DNA replication
GO:0060271	345	0,977	1,432E-66	6,552E-63	cilium assembly
GO:0033044	328	0,972	1,073E-62	4,560E-59	regulation of chromosome organization
GO:0098813	251	1,093	1,169E-62	4,636E-59	nuclear chromosome segregation
GO:0006302	230	1,171	2,221E-62	8,254E-59	double-strand break repair
GO:0006310	271	1,101	4,480E-61	1,567E-57	DNA recombination
GO:0034660	439	0,759	5,606E-60	1,852E-56	ncRNA metabolic process
GO:0000819	181	1,181	4,082E-58	1,278E-54	sister chromatid segregation
GO:0022613	418	0,744	6,547E-56	1,947E-52	ribonucleoprotein complex biogenesis

**Tabla 6.** Los procesos biológicos inhibidos con FDR <0,01 según el GSEA

GO TERM	COUNT	LOR	pvalor	FDR	Función
---------	-------	-----	--------	-----	---------

<b>GO:0050911</b>	384	-0,667	6,166E-42	7,640E-39	detection of chemical stimulus involved in sensory perception of smell
<b>GO:0050907</b>	433	-0,615	2,739E-39	2,585E-36	detection of chemical stimulus involved in sensory perception
<b>GO:0009593</b>	469	-0,571	3,701E-36	3,057E-33	detection of chemical stimulus
<b>GO:0007608</b>	415	-0,592	1,037E-34	7,344E-32	sensory perception of smell
<b>GO:0007606</b>	485	-0,533	1,896E-32	1,116E-29	sensory perception of chemical stimulus
<b>GO:0031424</b>	224	-0,776	4,337E-31	2,324E-28	keratinization
<b>GO:0050906</b>	490	-0,522	4,610E-31	2,448E-28	detection of stimulus involved in sensory perception
<b>GO:0030216</b>	303	-0,546	1,431E-20	3,325E-18	keratinocyte differentiation
<b>GO:0070268</b>	111	-0,835	5,786E-17	1,064E-14	cornification
<b>GO:0018149</b>	59	-0,985	3,174E-15	4,866E-13	peptide cross-linking
<b>GO:0009913</b>	358	-0,424	3,196E-15	4,886E-13	epidermal cell differentiation
<b>GO:0043588</b>	417	-0,333	2,064E-11	2,080E-09	skin development
<b>GO:0008544</b>	466	-0,298	2,241E-10	1,968E-08	epidermis development
<b>GO:0030198</b>	371	-0,297	1,242E-08	8,817E-07	extracellular matrix organization
<b>GO:0030049</b>	39	-0,838	1,631E-08	1,124E-06	muscle filament sliding
<b>GO:0033275</b>	39	-0,838	1,631E-08	1,124E-06	actin-myosin filament sliding
<b>GO:0043062</b>	372	-0,294	1,670E-08	1,149E-06	extracellular structure organization
<b>GO:0022617</b>	81	-0,588	1,525E-07	9,226E-06	extracellular matrix disassembly
<b>GO:0031581</b>	12	-1,188	8,961E-06	4,106E-04	hemidesmosome assembly
<b>GO:0061844</b>	66	-0,537	9,044E-06	4,133E-04	antimicrobial humoral immune response mediated by antimicrobial peptide
<b>GO:0019730</b>	117	-0,397	1,612E-05	7,058E-04	antimicrobial humoral response
<b>GO:0060371</b>	10	-1,162	3,267E-05	1,352E-03	regulation of atrial cardiac muscle cell membrane depolarization
<b>GO:0032963</b>	114	-0,373	7,018E-05	2,710E-03	collagen metabolic process
<b>GO:0030574</b>	46	-0,589	1,250E-04	4,608E-03	collagen catabolic process
<b>GO:0007210</b>	36	-0,609	1,424E-04	5,202E-03	serotonin receptor signaling pathway
<b>GO:0098664</b>	31	-0,639	1,732E-04	6,206E-03	G protein-coupled serotonin receptor signaling pathway

#### 9.4. Análisis de enriquecimiento con DAVID software

Se realizó un análisis de enriquecimiento con el software DAVID insertando los 5.203 genes diferenciales sobreexpresados con FDR <0,01 y los 752 genes diferenciales infraexpresados con FDR <0,01.

Se obtuvieron 98 procesos biológicos enriquecidos con FDR <0,01 relacionados a los DEGs sobreexpresados (Tabla 6 anexo) y 11 procesos biológicos relacionados a los DEGs infraexpresados con pvalor <0,01 (Tabla 8).

En la siguiente tabla se visualizan los 20 procesos biológicos influenciados por los DEGs sobreexpresados (Tabla 7).

Destacamos en naranja en las tablas 7 y 8 procesos biológicos similares a aquellos hallados en el GSEA como por ejemplo *chromatin remodeling and organization*, *cell division*, *cellular response to DNA damage stimulus*, *DNA repair*, *mRNA processing*, *cell cycle*, *mRNA splicing via spliceosome*, *RNA splicing*, *chromosome segregation*, *epidermis development*, *keratinization* y *keratinocyte differentiation*, los cuales son de interés para nuestro estudio por su influencia conocida sobre los procesos oncogénicos.

**Tabla 7.** Los 20 procesos biológicos con menor FDR enriquecidos según el análisis DAVID.

Term	Proceso biológico	Count	pvalor	FDR
GO:0006357	regulation of transcription from RNA polymerase II promoter	562	7,985E-37	5,360E-33
GO:0006281	DNA repair	158	1,236E-35	4,148E-32
GO:0006355	regulation of transcription, DNA-templated	363	6,685E-33	1,496E-29
GO:0006338	chromatin remodeling	160	9,386E-30	1,575E-26
GO:0006325	chromatin organization	152	8,522E-28	1,015E-24
GO:0006260	DNA replication	81	9,070E-28	1,015E-24
GO:0051301	cell division	171	1,147E-27	1,100E-24
GO:0006974	cellular response to DNA damage stimulus	138	2,551E-24	2,141E-21

GO:0006397	mRNA processing	108	1,923E-18	1,434E-15
GO:0007049	cell cycle	144	1,940E-17	1,302E-14
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	367	3,052E-17	1,862E-14
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	310	7,795E-16	4,360E-13
GO:0000398	mRNA splicing, via spliceosome	91	1,194E-15	6,166E-13
GO:0045893	positive regulation of transcription, DNA-templated	235	4,194E-15	2,011E-12
GO:0000724	double-strand break repair via homologous recombination	57	4,530E-15	2,027E-12
GO:0060271	cilium assembly	100	9,216E-15	3,866E-12
GO:0008380	RNA splicing	91	6,223E-14	2,457E-11
GO:0007059	chromosome segregation	53	2,623E-13	9,779E-11
GO:0000070	mitotic sister chromatid segregation	29	1,608E-12	5,679E-10
GO:0032508	DNA duplex unwinding	37	1,503E-11	5,044E-09

**Tabla 8.** Los 11 procesos biológicos con pvalor <0,01 inhibidos según el análisis DAVID.

Term	Proceso biológico	Count	pvalor	FDR
GO:0008544	Epidermis development	15	3,735E-12	4,837E-09
GO:0031424	keratinization	14	1,232E-10	7,979E-08
GO:0030216	keratinocyte differentiation	11	1,149E-08	4,958E-06

GO:0007186	G-protein coupled receptor signaling pathway	28	9,923E-05	3,212E-02
GO:0070268	cornification	5	2,173E-04	4,721E-02
GO:0045109	intermediate filament organization	7	2,187E-04	4,721E-02
GO:0022617	extracellular matrix disassembly	6	2,788E-04	5,158E-02
GO:0061436	establishment of skin barrier	5	9,602E-04	1,554E-01
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	15	1,273E-03	1,832E-01
GO:0043616	keratinocyte proliferation	4	4,670E-03	6,047E-01
GO:0071467	cellular response to pH	3	5,408E-03	6,366E-01

### 9.5. Análisis de coincidencias entre los genes involucrados en los procesos biológicos enriquecidos de interés y los genes con AUC >0,7

Determinamos a los siguientes procesos biológicos significativos por el programa DAVID cómo de interés primario para nuestro estudio, dado que son procesos fuertemente asociados a la oncogénesis, además de ser procesos biológicos significativos que tienen relación en los dos análisis de enriquecimiento: *cell division*, *cell cycle*, *chromatin organization*, *chromatin remodelling*, además de los procesos biológicos inhibidos de *keratinization*, *epidermis development* y, *keratinocyte differentiation*. Determinamos los genes; BIRC6, CDK20, CDT1, EZH2, HDAC10, KDM5A, KMT5C, SETMAR, SPC24, SPRR2G y TET1 que están relacionados con estos procesos biológicos de interés y además que tenían un AUC >0,7 en nuestro análisis ROC (Tabla 9

y 10).

**Tabla 9.** Posibles genes de interés organizados según el proceso enriquecido vinculado.

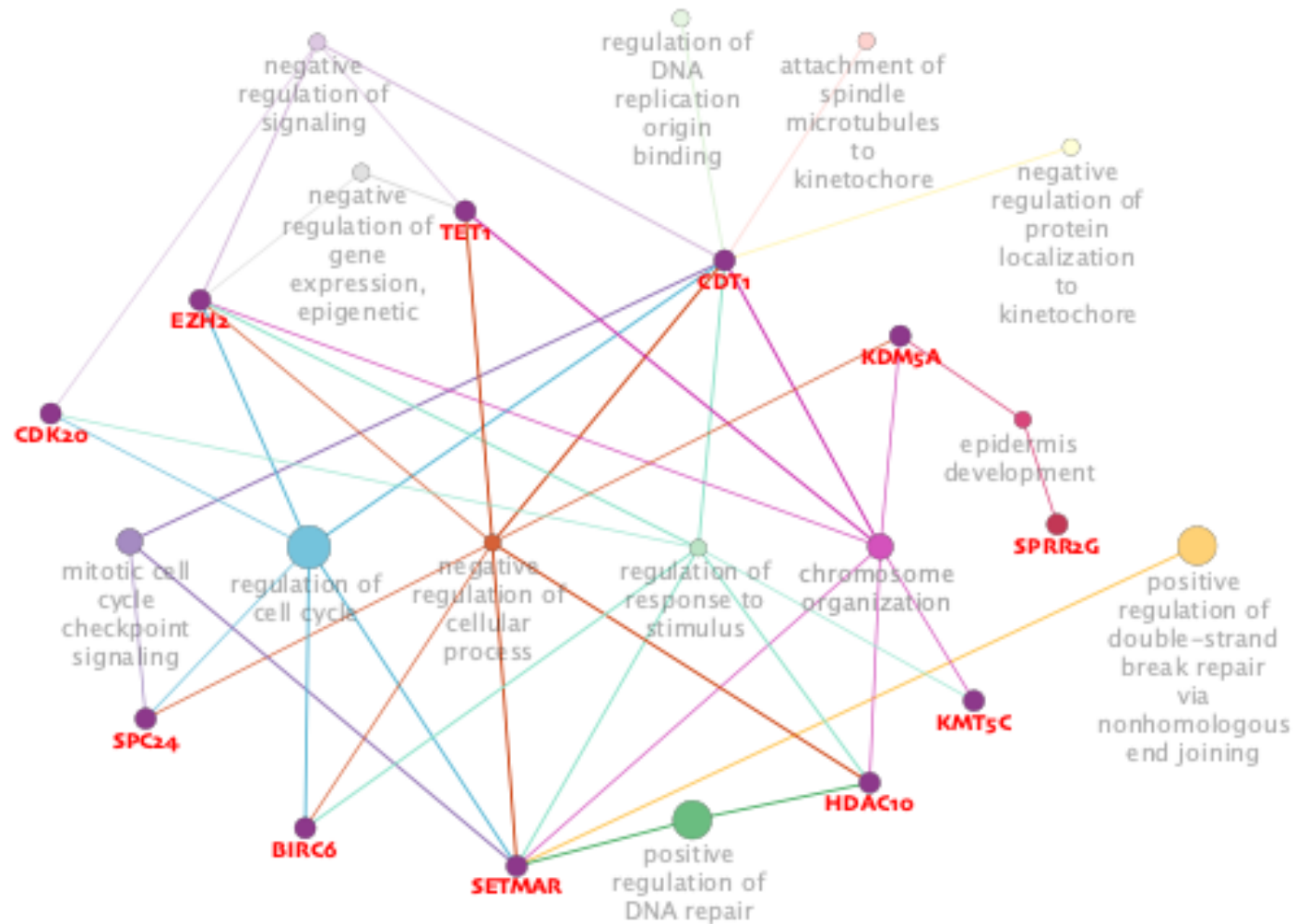
Proceso biológico	Genes sobreexpresado	Gen infraexpresado
<i>Chromatin remodelling</i>	KMT5C, TET1, HDAC10, SETMAR, KDM5A	
<i>Chromatin organization</i>	KMT5C, TET1, HDAC10, SETMAR, EZH2	
<i>Cell division</i>	CDK20, CDT1, BIRC6, SPC24	
<i>Cell cycle</i>	CDK20, CDT1, BIRC6, SPC24	
<i>Keratinization</i>		SPRR2G
<i>Keratinocyte differentiation</i>		SPRR2G
<i>Epidermis development</i>		SPRR2G

En la Tabla 10 se puede observar todos los genes de interés en orden descendente del AUC pudiéndose apreciar el FDR, el *Fold Change*, y el nombre completo de los genes. Además, en la Figura 3 podemos observar los genes con sus funciones biológicas no redundantes.

**Tabla 10.** Los genes de posible interés para nuestro estudio

	Nombre completo	Fold Change	FDR	AUC
<b>EZH2</b>	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit	0,715	6,128E-04	0,788
<b>HDAC10</b>	Histone Deacetylase 10	0,785	7,843E-03	0,777
<b>KDM5A</b>	Lysine Demethylase 5A	0,468	3,921E-03	0,776
<b>BIRC6</b>	Baculoviral IAP Repeat Containing 6	0,510	4,445E-03	0,748
<b>SETMAR</b>	SET Domain And Mariner Transposase Fusion Gene	0,611	1,724E-03	0,732
<b>SPC24</b>	SPC24 Component Of NDC80 Kinetochore Complex	0,610	6,774E-03	0,729
<b>CDK20</b>	Cyclin Dependent Kinase 20	0,912	1,605E-03	0,724
<b>KMT5C</b>	Lysine Methyltransferase 5C	0,505	4,986E-03	0,720
<b>CDT1</b>	Chromatin Licensing And DNA Replication Factor 1	0,702	2,317E-03	0,715

<b>TET1</b>	Tet Methylcytosine Dioxygenase 1	0,848	6,457E-03	0,712
<b>SPRR2G</b>	Small Proline Rich Protein 2G	-3,091	5,308E-05	0,707

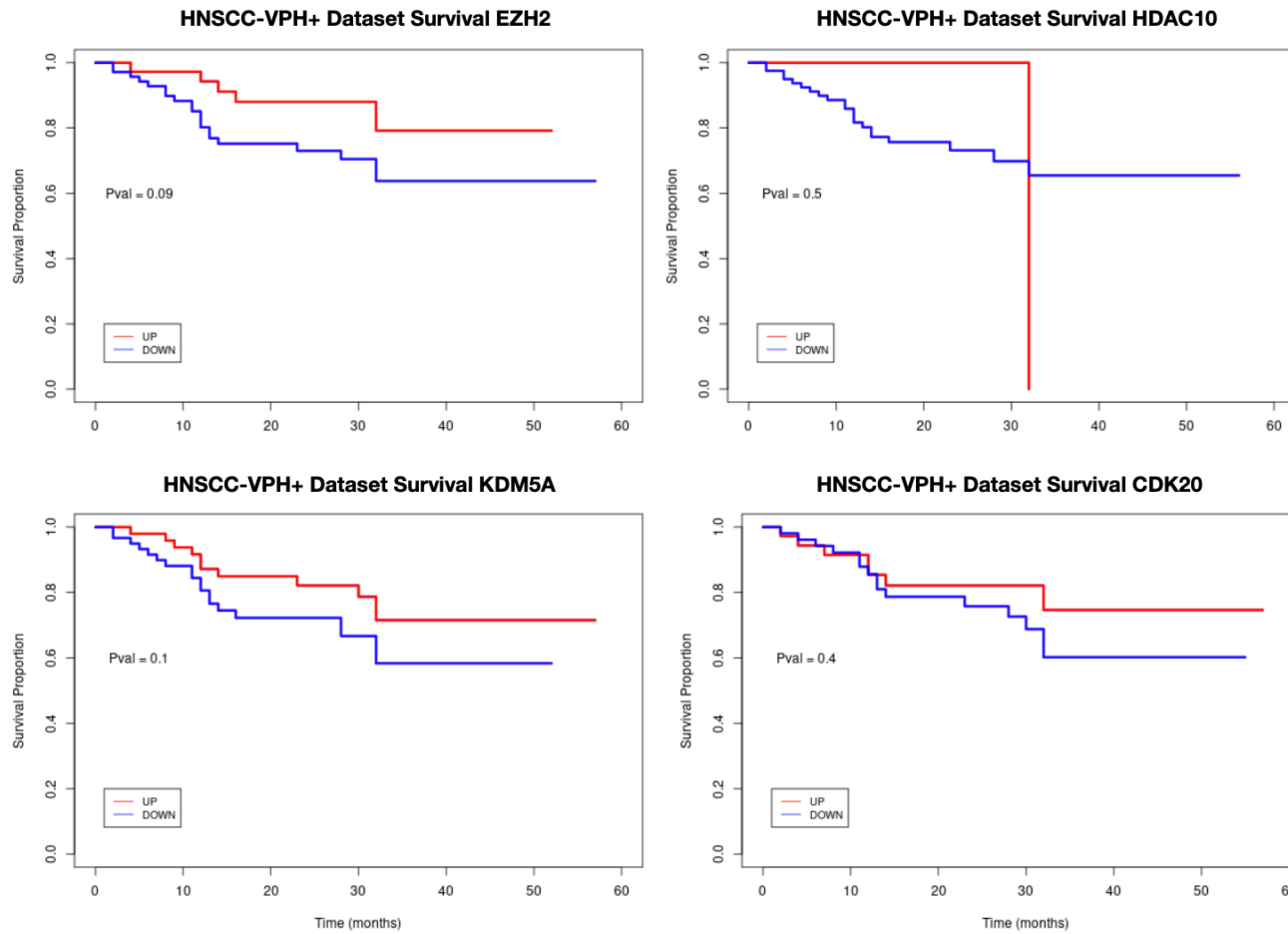


**Figura 3.** Genes de interés con sus funciones biológicas relacionadas realizado mediante el programa Cytoscape, plug-in ClueGo.

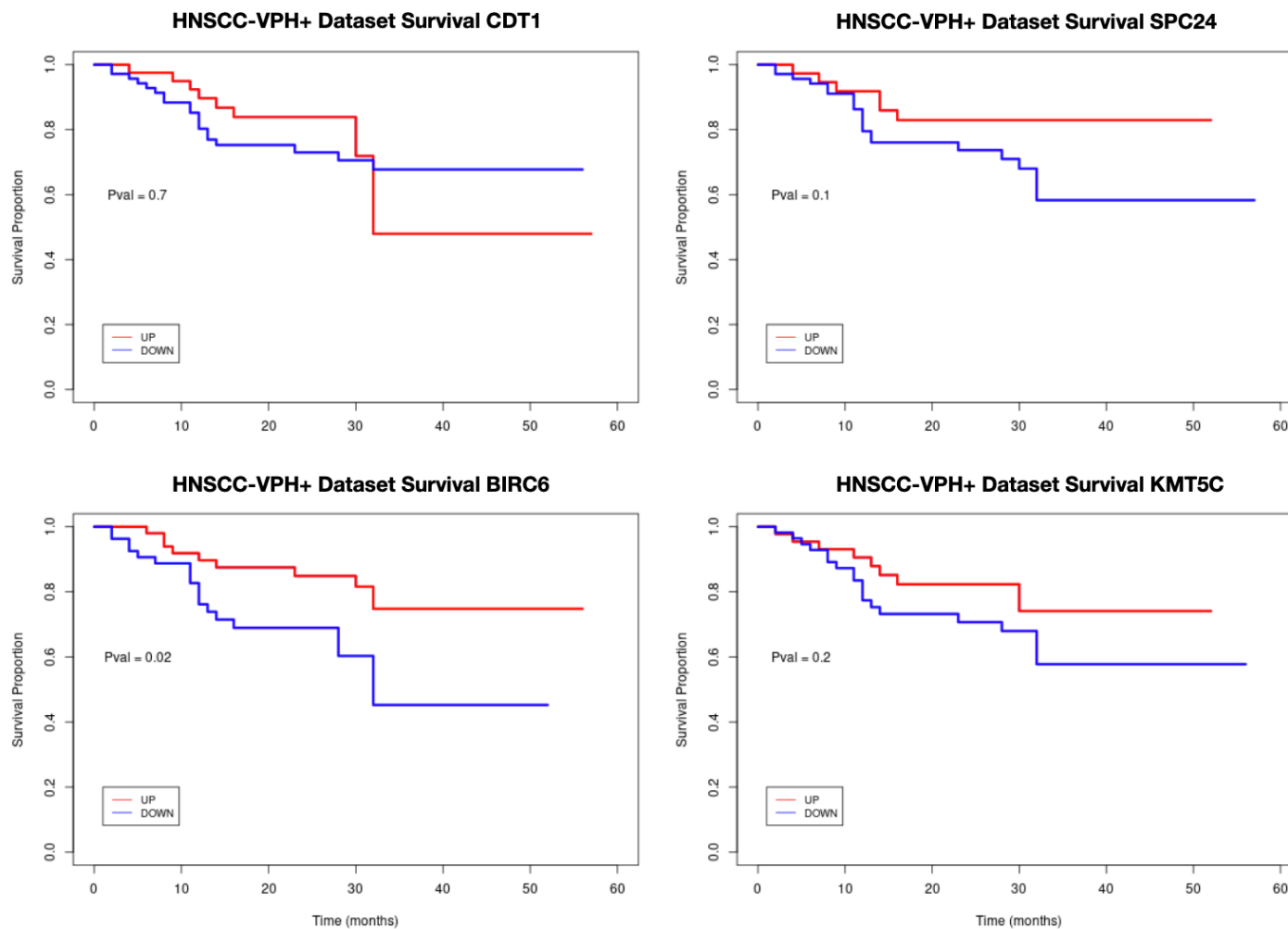


## 9.6. Análisis de supervivencia

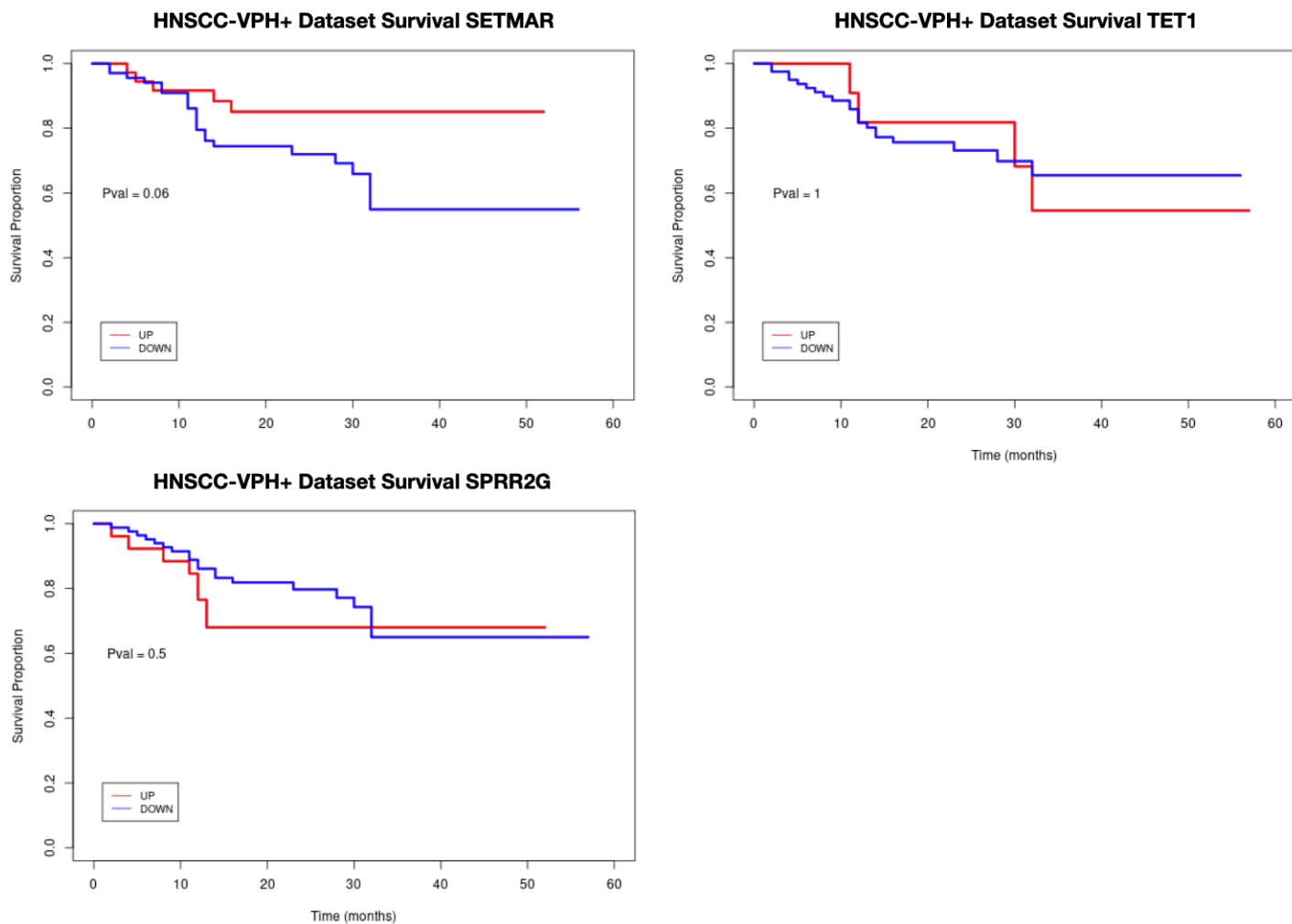
Se realizó un análisis de supervivencia para los genes de interés con las 112 muestras con HNSCC con información perteneciente a la infección o ausencia de infección por el HPV tipo 16 y se les aplicó una prueba estadística de Fischer para validar la significancia. Acorde a su expresión diferencial hallada en nuestro análisis diferencial, todos los genes discutidos demostraron tener una tendencia de un efecto positivo sobre la supervivencia (Figuras 4-6). El único gen que demostró esta tendencia de manera significativa (pvalor = 0,02) fue BIRC6.



**Figura 4.** Análisis de supervivencia de los genes EZH2, HDAC10, KDM5A, CDK20. La línea azul representa la infra expresión del gen y la línea roja la sobreexpresión de los genes referentes a las muestras HNSCC-HPV.



**Figura 5.** Análisis de supervivencia de los genes CDT1, SPC24, BIRC6, KMT5C. La línea azul representa la infra expresión del gen y la línea roja la sobreexpresión de los genes referentes a las muestras HNSCC-HPV.



**Figura 6.** Análisis de supervivencia de los genes SETMAR, TET1, SPRR2G. La línea azul representa la infra expresión del gen y la línea roja la sobreexpresión de los genes referentes a las muestras HNSCC-HPV.

## 10. Discusión

En nuestro estudio los genes que han resultado de interés han sido EZH2, KDM5A, KMT5C, TET1, HDAC10, SETMAR, CDK20, CDT1, BIRC6, SPC24 y SPRR2G, de los cuales SPRR2G es el único gen infraexpresado.

Estos genes han resultado de interés por su expresión diferencial significativa (FDR <0,01), su AUC >0,7 y su influencia sobre procesos biológicos enriquecidos de posible interés para el desarrollo tumoral como por ejemplo el ciclo celular, la división celular, la organización y la remodelación de la cromatina en caso de los genes sobreexpresados, además de la queratinización, la diferenciación de queratinocitos y el desarrollo de la epidermis en caso de nuestro gen infraexpresado SPRR2G en el grupo HNSCC-HPV+.

### EZH2

EZH2 podría tener influencia sobre la oncogénesis por su rol en la regulación de la expresión génica, por la trimetilación de Lys 27 de la histona 3, la inactivación de genes supresores de tumores, la activación de factores de transcripción STAT3 y NFκB, además del silenciamiento de H3K9me3 (5, 52-56). El EZH2 es un gen con influencia sobre el ciclo celular influyendo en procesos de apoptosis, la promoción de la reparación de DNA y la inhibición de la senescencia celular (5, 53, 54, 57). En nuestro estudio se ha visto activando la organización de la cromatina en el grupo con papiloma.

Bugide y cols (2021) demostraron que EZH2 reprime CXCL10, una quimiocina que regula la migración de células inmunes NK (*Natural Killer*). Esta represión tiene como efecto una inhibición de la migración de estas células NK, es decir, una menor respuesta antitumoral. El reclutamiento de EZH2 a CXCL10 es dependiente de otro gen de interés y sobreexpresado en nuestro estudio: HDAC10 (58). Además de esta acción sobre el sistema inmunológico también juega un rol en la modulación del sistema inmune por medio de regularización de la diferenciación y polarización de linfocitos T además de la liberación de citocinas, la expresión de *check-points* inmunológicos y la represión de la capacidad de migración de células Langerhans (53, 54).

En nuestro estudio este gen ha demostrado una sobreexpresión en pacientes HNSCC-HPV+ como de manera similar el estudio de Hyland y cols (2011) sobre tejidos HPV positivos en lesiones cervicales displásicas y cancerosas (59). Según Holland y cols (2008) la proteína E7 del HPV tipo 16, el mismo tipo de HPV que en nuestro estudio, activa al EZH2 por medio de liberación de E2F. Además, parece ser necesaria la expresión continuada de EZH2 para la proliferación de células cervicales cancerosas HPV positivas (60). El mecanismo de activación de EZH2 por la liberación de E2F gracias a E7 se da por la unión de E7 a pRb lo que lo inhibe causando la activación de los factores de transcripción E2F. Esto a su vez causa la progresión de la fase G1 a la fase S en el ciclo celular (61).

En la tabla 1 del anexo podemos ver que 7 de los genes de la familia E2F se hallaron diferencialmente expresados en los pacientes HNSCC-HPV+. Más específicamente E2F1, E2F2, E2F5, E2F6, E2F7, E2F8 se hallaron sobreexpresados.

Por el mecanismo de acción que tiene sobre STAT3 y NF- $\kappa$ B hay una involucración con vías biológicas de la inflamación, la proliferación celular, el sistema inmune, la evasión inmune y la transcripción de DNA, teniendo influencia sobre la oncogénesis por la desregulación vinculada de la vía JAK/STAT (62-66). Además, la vía JAK/STAT parece influir sobre la liberación de E2F (70). Esto crearía un círculo vicioso en el cual el HPV E7 activaría a EZH2 por medio de liberación de E2F, EZH2 activaría a la vía JAK/STAT, lo cual liberaría más E2F provocando en teoría un nuevo aumento de EZH2.

EZH2 ha sido también relacionado al citomegalovirus, viéndose correlacionado con éste en muestras de carcinoma sérico ovárico de alto grado (67).

Otros cánceres que han demostrado una sobreexpresión de EZH2, sin estar estudiada la relación al HPV, son el cáncer de próstata, de mama, de hígado, de vejiga y cánceres orales y orofaríngeos (52, 57, 68). Más específicamente Nienstedt y cols (2018) vieron una mayor asociación de la expresión de EZH2 en los cánceres orales y de la orofaringe que en los cánceres laríngeos (68).

En cuanto al pronóstico se ha podido establecer una relación de la sobreexpresión con un pronóstico empeorado en cánceres de pulmón, de próstata, de riñones, de mama, de hígado y gástricos (52, 53, 55, 69, 70). Ganesh y cols (2023) también han podido establecer que la sobreexpresión está relacionada a una tasa de supervivencia más baja y que hubo una asociación a un riesgo 13 veces mayor de tejido

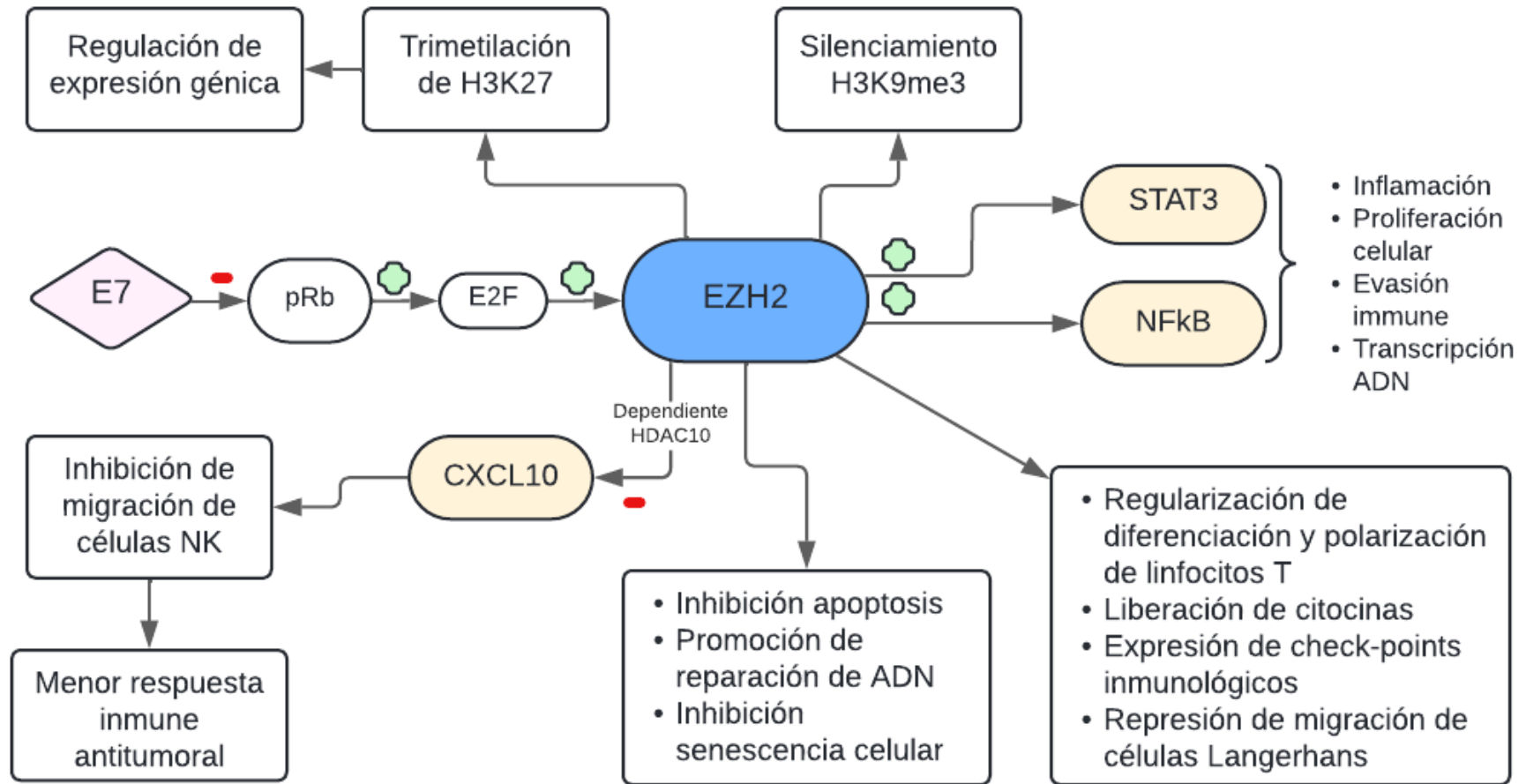
precanceroso oral de transformarse en tejido canceroso escamoso en presencia de la sobreexpresión de EZH2 (5).

Aparte de esto, EZH2 parece influir positivamente sobre la metástasis en ganglios linfáticos y sobre la resistencia a la apoptosis (59, 68). En casos de inhibición de EZH2 se ha podido incrementar la tasa de apoptosis y se piensa que esto va relacionado a la regulación de los factores Puma y Bad, el incremento de p21 y la disminución del número de células en la fase S y G2/M (71).

En cuanto al tratamiento del cáncer con enfoque en este gen, se ha visto que EPZ-6438 y UNC1999 podrían ser posibles inhibidores de él (56). Según Khattri y cols (2023) es importante no solamente enfocarse en inhibir las funciones independientes de la metiltransferasa sino también las funciones relacionadas a ella (61). La inhibición de EZH2 ha demostrado una disminución de la tasa proliferativa en el carcinoma sérico ovárico de alto grado (67).

A través de nuestro análisis de supervivencia, hemos podido observar que la sobreexpresión de EZH2 demuestra un impacto positivo sobre la supervivencia de pacientes HNSCC-HPV+. Esto se puede apreciar en la figura 4 (Figura 4).

Para una mejor visualización de las vías y otros componentes biológicos influenciados por EZH2 hemos elaborado un *Flow chart* (Figura 7).



**Figura 7.** Flow Chart del gen EZH2. Vemos la influencia de EZH2 sobre STAT3, NFκB, ULBP1 y CXCL10, aparte de un vínculo del HPV a un componente con influencia sobre la expresión de EZH2 (E2F). Elaboración propia



## HDAC10

Otro gen de interés que ha demostrado tener un efecto positivo sobre la expresión de STAT3 es HDAC10. Su deficiencia se ha relacionado con un efecto negativo sobre la expresión de STAT3, lo cual implicaría que la sobreexpresión de HDAC10 jugaría un rol promotor sobre procesos cancerígenos y la inflamación (64). Según nuestro estudio este gen está sobreexpresado en pacientes HNSCC-HPV+ y está involucrado en funciones de remodelación y organización de cromatina.

En mayor detalle este gen influye sobre la desacetilación de histonas. El efecto de este proceso conlleva una represión de procesos transcripcionales, dado que la cromatina es convertida en una forma no funcional (72). Además, el HDAC10 juega un rol en procesos de reparación de DNA (73). Liu y cols (2020) también comentan de su rol en la diferenciación celular, la proliferación celular, la regulación del ciclo celular y procesos de autofagia, además de su influencia en vías biológicas involucradas en procesos oncogénicos y metastásicos (74).

En el estudio de Bugide y cols (2021) en el cual se estableció que el HDAC10 es necesario para el reclutamiento de EZH2 a CXCL10 para que lo reprima, también se realizó un *knockdown* de HDAC10 lo cual promovió la migración de células NK (*Natural Killer*) (58). Es decir que la sobreexpresión de HDAC10 podría estar vinculada a una menor función inmunológica antitumoral, dado que reprime la migración de células NK.

Concordando con nuestro resultado, Yang y cols (2016) establecieron una sobreexpresión de este gen en el cáncer de pulmón y vieron que el *knockdown* de éste causó una detención del ciclo celular. Además, en un modelo en ratones podían ver que la sobreexpresión incrementó las tasas de proliferación (75). Liu y cols (2020) podían establecer que la sobreexpresión estuvo relacionada a un peor pronóstico en carcinomas pulmonares (74). Guo y cols (2023) también demostraron una sobreexpresión en cáncer pulmonar. El *knockdown* resultó en una inhibición de la proliferación y una promoción de la ferroptosis por la influencia de HDAC10 sobre los niveles de ROS (*High reactive oxygen species*) (76).

En contraste, Gu y cols (2022) tuvieron resultados contrarios. La sobreexpresión parecía tener una influencia negativa sobre la capacidad de invasión y migración de las células cancerosas e inhibir el proceso canceroso cervical. Vieron una infraexpresión en los tejidos cancerosos cervicales y una asociación entre la infraexpresión con un

pronóstico desfavorable (72). Zhu y cols 2021 también vieron una infraexpresión en tejido canceroso, específicamente cáncer cervical y postularon que HDAC10 podría reducir la capacidad proliferativa. La sobreexpresión del HDAC10 en un modelo de ratones resultó en tasas de crecimiento tumoral reducidas. La expresión alta se asoció a una mejor tasa de supervivencia (77).

Juega también un posible rol en la vía PI3K/AKT, lo cual podría tener efectos oncogénicos (75). Esto se debería estudiar más profundamente.

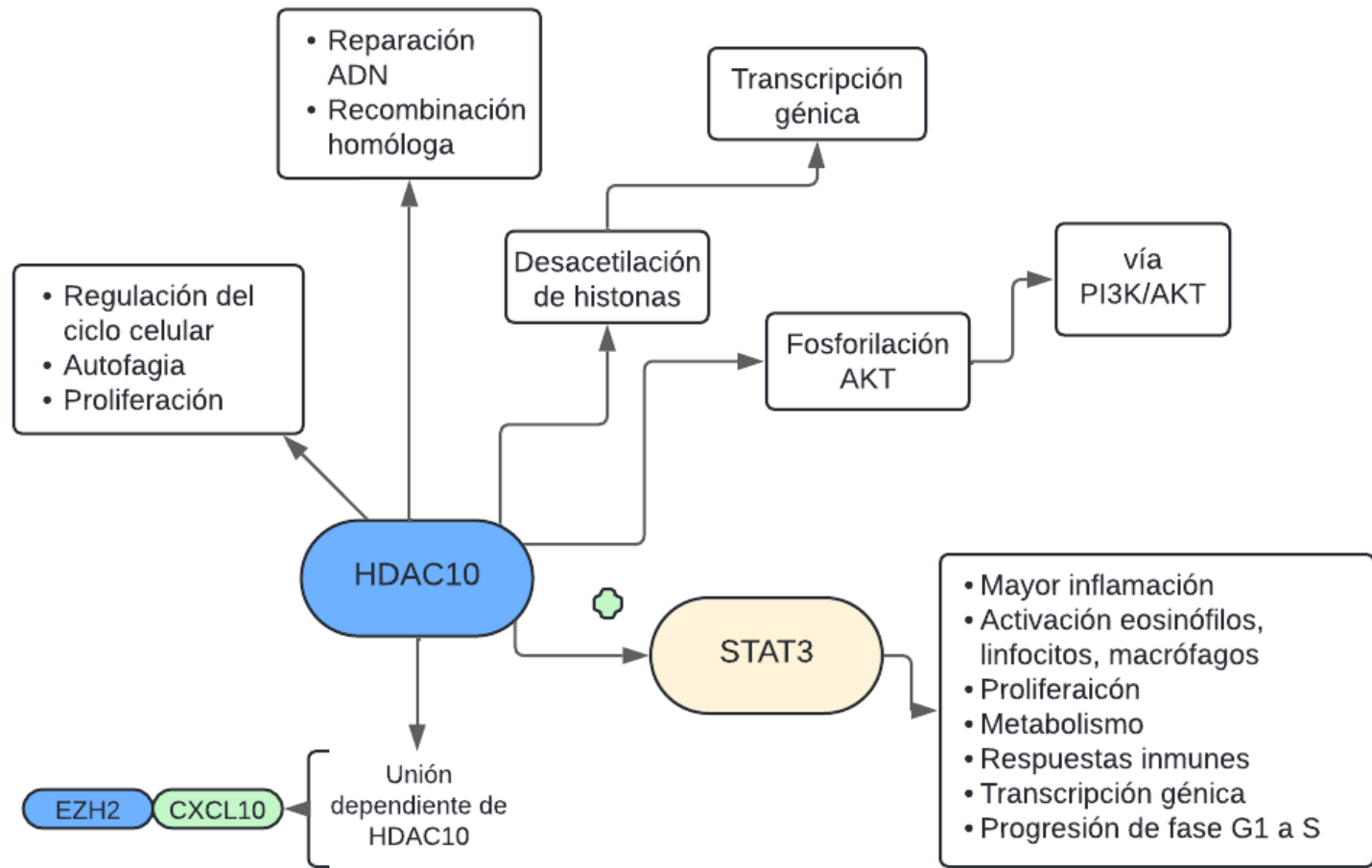
Se ha visto que los niveles disminuidos de HDAC10 tienen un efecto positivo sobre la sensibilidad a los tratamientos quimioterapéuticos basados en platino (73). Esto concuerda con lo postulado por Ling y cols (2023).

En general, hay una mayor variabilidad en los resultados de los estudios sobre este gen en comparación con los estudios sobre EZH2. Nuestro estudio refuerza la noción de que este gen podría jugar un rol oncogénico y vincula este gen al HPV. Es posible que el HDAC10 sea oncogénico, sin embargo, hay evidencia de una asociación negativa, por lo cual es importante establecer de manera mejor el rol oncogénico o supresor de HDAC10.

Sería también de interés estudiar los posibles mecanismos vinculantes entre el HPV y el HDAC10. Sería de mucho interés evaluar si efectivamente tiene una función activadora sobre el STAT3, dado que esto reforzaría que este gen esté involucrado en procesos patológicos, por la influencia de STAT3 sobre procesos oncogénicos.

Para interpretar correctamente el efecto de la expresión de este gen sobre la supervivencia de pacientes HNSCC-HPV+/ HPV- se necesitan de más muestras, dado que en nuestro análisis se observa una falta de muestras con sobreexpresión de este gen como para interpretar la gráfica. De igual modo se puede observar el resultado en la figura 4.

En la siguiente figura se puede observar un *Flow chart* de los procesos asociados a este gen y en qué manera influye sobre otros componentes biológicos (Figura 8).



**Figura 8.** Flow Chart del gen HDAC10. Se pueden apreciar en cuales componentes biológicos influye este gen y con cuales procesos está vinculado. Elaboración propia

## KDM5A

Otro de nuestros genes de interés con funciones relacionadas a la remodelación de cromatina es el KDM5A. El mecanismo de acción de éste se basa en la eliminación de di- y trimetil *moieties* de la cuarta lisina de la histona 3 (H3K4me2/3) lo cual conlleva un efecto sobre la transcripción. Otras funciones vinculadas son la represión de la diferenciación celular, la promoción de la angiogénesis, la movilidad celular, la transición epitelial-mesenquimal, la proliferación y la invasión celulares (78).

Este gen tiene también un rol en la presentación de antígenos, silenciando genes involucrados para este proceso. Según Liu y cols (2022) inhibe los procesos antitumorales y la infiltración de linfocitos T CD8+ (79).

Este gen se ha visto sobreexpresado en la leucemia mieloide aguda y otros tipos de cánceres y parece tener un efecto negativo sobre la supervivencia de los pacientes (78). Liu y cols (2022) también vieron que la inhibición de este gen tuvo un efecto positivo sobre los procesos antitumorales en un estudio sobre ratones, lo cual refuerza la noción de que tenga un rol oncogénico. La inhibición de KDM5A restauró la expresión de los genes involucrados en las vías de procesamiento y presentación de antígenos (80). Reforzando este estudio Oser y cols (2019) vieron que la inhibición de KDM5A resultó en una menor tumorigénesis en ratones (80).

Kataria y cols (2023) también concuerdan con el rol oncogénico viendo una sobreexpresión en cánceres gástricos, pulmonares y de mama, además de una inhibición de inhibidores de CDK inducida por la sobreexpresión de KDM5A, lo cual incrementó la capacidad invasiva y proliferativa de las células cancerosas. En cánceres ováricos aumentó la transición epitelial-mesenquimal y la proliferación (81).

Estos resultados concuerdan con el nuestro; la sobreexpresión de KDM5A está asociada al cáncer, en nuestro caso de cabeza y cuello en pacientes HPV+.

En cuanto a las vías en las cuales podría influir se ha estudiado una influencia negativa sobre la vía Notch en el carcinoma pulmonar de células pequeñas (80). La vía Notch tiene mecanismos tumorigénicos por su involucración en procesos relacionados a la señalización celular, la regulación del TME y un metabolismo alterado (82).

Kataria y cols (2023) también demostraron su rol supresor sobre p53, lo cual tendría un efecto negativo sobre la apoptosis, lo cual significaría un mayor potencial oncogénico (81).

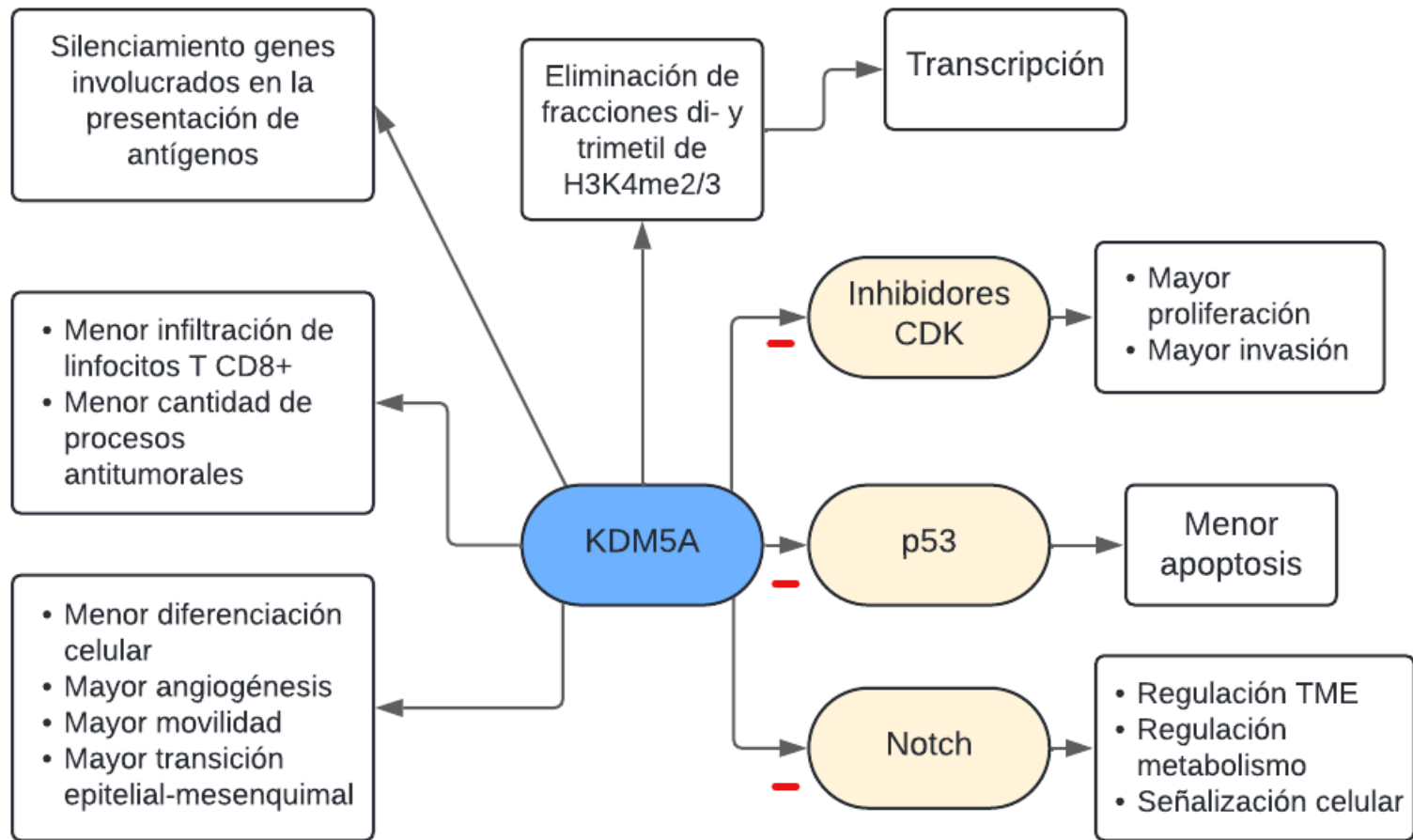
La inhibición de KDM5A no solo tiene efectos sobre el cáncer, sino que también tiene efectos protectores sobre la ototoxicidad provocada por uso del compuesto cisplatino (81).

En cuanto al pronóstico no se ha estudiado de manera profunda la asociación de la sobreexpresión de KDM5A con el pronóstico. Sería interesante vincular estas dos cosas en futuros estudios para determinar si KDM5A sería un buen biomarcador de pronóstico. Además, la posible relación entre el HPV y el KDM5A no ha sido estudiado. Se debería hacer hincapié en esta relación para entender de mejor manera los posibles mecanismos por los cuales el HPV podría aumentar la expresión de este gen.

El HPV parece influir sobre la vía Notch dado que E6 se une a NFX1-123 cuya sobreexpresión en queratinocitos está vinculada a una mayor tasa de crecimiento, una menor tasa de apoptosis, menor expresión de genes de la respuesta inmune, mayor actividad de telomerasa y una regulación positiva sobre Notch1 (83). En teoría entonces el HPV activaría la vía Notch mientras que KDM5A la inactivaría. La postulación del efecto negativo de KDM5A sobre la vía Notch se debería estudiar más profundamente para poder establecer los posibles mecanismos por los cuales influye en la oncogénesis.

Según nuestro análisis de supervivencia la sobreexpresión de KDM5A en pacientes HNSCC-HPV+ demostró tener un efecto positivo sobre la supervivencia (Figura 4).

En el *Flow Chart* de la página 50 se pueden apreciar aquellas funciones relacionadas a este gen y aquellos componentes que se vean afectados por él (Figura 9).



**Figura 9.** Flow chart del KDM5A. Se visualiza la influencia de KDM5A sobre inhibidores CDK, p53 y Notch y funciones relacionadas con este gen. Elaboración propia

## **CDK20**

El E6, además del E7, ambos componentes del HPV asociados fuertemente con la oncogénesis regulan la vía Wnt/ $\beta$ -catenina porque causan una menor degradación de la  $\beta$ -catenina (84, 85).

El gen CDK20 ha sido asociado a la activación de esta vía (86). Esto implicaría un efecto sobre el desarrollo del cáncer, dado que los productos finales de esta vía influyen sobre procesos relacionados a la evasión inmune y la proliferación (87).

En nuestro estudio CDK20 resultó sobreexpresado en pacientes HNSCC-HPV+ relacionado con la división y el ciclo celular. Más específicamente la familia de CDKs juega un rol en la progresión de la célula a través del ciclo celular, modulando la transcripción génica y la división celular por medio de la integración de señales intra- y extracelulares. El CDK20 específicamente estimula el ciclo celular por su efecto sobre la vía  $\beta$ -catenina/TCF y se ha visto que la inhibición de CDK20 resulta en la inhibición del ciclo celular (86).

Otros estudios han podido relacionar este gen con la hepatocarcinogénesis, cánceres pulmonares, hepáticos, ováricos y colorrectales (88, 89). Zhou y cols (2018) también comentan de su rol conector entre varios reguladores transcripcionales como el EZH2 (88).

Feng y cols (2015) refuerzan esta interacción con EZH2. En su estudio vieron que la expresión de CDK20 en tejidos hepáticos causó una sobreexpresión de EZH2 y que esta sobreexpresión aumentó la expresión y actividad de los receptores de andrógenos lo cual aumentó la expresión de CDK20 (90). Es decir, CDK20 y EZH2 parecen estimular sus respectivas expresiones en un tipo de círculo vicioso. En nuestro estudio ambos genes se encuentran sobreexpresados, lo cual podría reforzar esta asociación, aunque se tendría que estudiar en mayor detalle y en los tejidos orales.

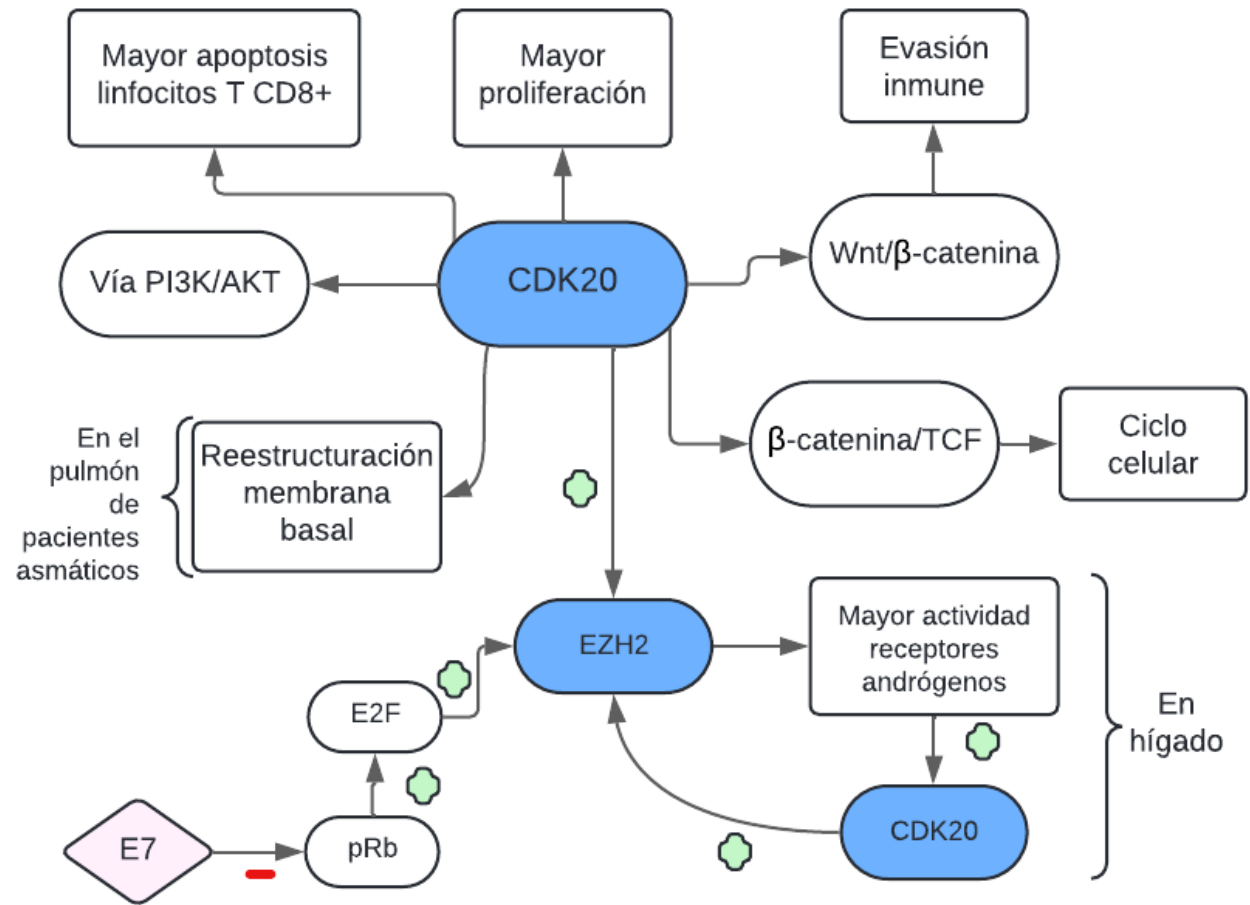
Otro resultado interesante fue del estudio de Bazan-Socha y cols (2021) sobre tejidos pulmonares en pacientes asmáticos. En este estudio pudieron establecer una asociación entre la sobreexpresión de CDK20 y la anchura de la membrana basal indicando un rol en la reestructuración de este (91). Posiblemente juega un rol similar en los tejidos epiteliales de la cabeza y cuello, que refuerza el potencial oncogénico. Se debería estudiar esto.

Se ha demostrado una inhibición efectiva de CDK20 y de la proliferación en hepatomas con sobreexpresión de CDK20 (89).

Nuestro análisis de supervivencia demostró aparte una tendencia mejorada en el pronóstico de pacientes HPV+ en presencia de la sobreexpresión de CDK20 (Figura 4).

En la siguiente figura se puede observar de manera visual la influencia que tiene este gen sobre otros componentes y vías biológicas, además de ver las funciones con las que está relacionado (Figura 10).





**Figura 10.** Flow Chart de CDK20. Se ven las vías y moléculas influenciadas por este gen. Aparte se ve un posible vínculo con el HPV a través de la molécula E2F y su influencia sobre EZH2. Elaboración propia

## **CDT1**

Otro gen que hemos encontrado sobreexpresado en pacientes HNSCC-HPV+ es el CDT1. Tiene funciones sobre la división y el ciclo celular.

Esta enzima forma parte del complejo de pre-replicación, por lo cual juega un rol en la replicación de DNA en el ciclo celular (92). Según Karantzelis y cols (2022) inhibe la unión temprana de MCM a la cromatina durante la fase S y G2 del ciclo celular y su desregulación puede causar inestabilidad genómica además de carcinogénesis (93).

La sobreexpresión relacionada al cáncer se ha visto en varios estudios. Ogawa y cols (2023) lo relacionaron con tejidos hepáticos cancerosos, Jiang y cols (2023) lo relacionaron al adenocarcinoma pulmonar y Sarafidis y cols (2022) lo relacionaron al cáncer urinario de vejiga, siendo asociado a un peor pronóstico en los tejidos hepáticos cancerosos y en el adenocarcinoma (94-96). Además, se relacionó con un pronóstico desfavorable en el cáncer de mama (92).

En el estudio de Ogawa y cols (2023) la sobreexpresión se relacionó a un mayor infiltrado inflamatorio en los lóbulos hepáticos, un nivel de transaminasa sérica elevada y un mayor grado de fibrosis en el tejido hepático, lo cual le da valor pronóstico al CDT1 para averiguar el potencial maligno del cáncer (94).

Li y cols (2021) pudieron demostrar un mayor potencial de invasión y de proliferación en aquellos cánceres con el CDT1 sobreexpresado (97). Wang y cols (2022) también indican que en el cáncer de próstata facilitó la metástasis (98).

CDT1 necesita ser regulado para poder realizar sus funciones de manera correcta. Sus niveles se encuentran en su punto más bajo en la fase S/G2 y en su punto más alto en las fases M y G1. Es posible que la interacción de estas enzimas tenga repercusiones sobre la replicación de DNA, pero se tiene que estudiar de mejor manera (99).

En un estudio de Karakaidos y cols (2004) la expresión de CDT1 se vinculó con la expresión E2F1. En este estudio se vio que E2F1 sobrerregula un promotor de CDT1 (100). Como indicado en el apartado de EZH2, efectivamente es un gen con sobreexpresión en los pacientes HNSCC-HPV+ de nuestro estudio. Es posible entonces que por esta sobreexpresión se dé la sobreexpresión de CDT1.

La sobreexpresión de tanto CDT1 promovió la rereplicación en aquellos casos en los cuales el p53 no estuvo activo, pero no en casos en los cuales el p53 estuvo activo

(101). Esto sería interesante en relación con el HPV, dado que el HPV degrada el p53, sin embargo, sería necesario confirmar los resultados de Vaziri y cols (2003) con otros estudios.

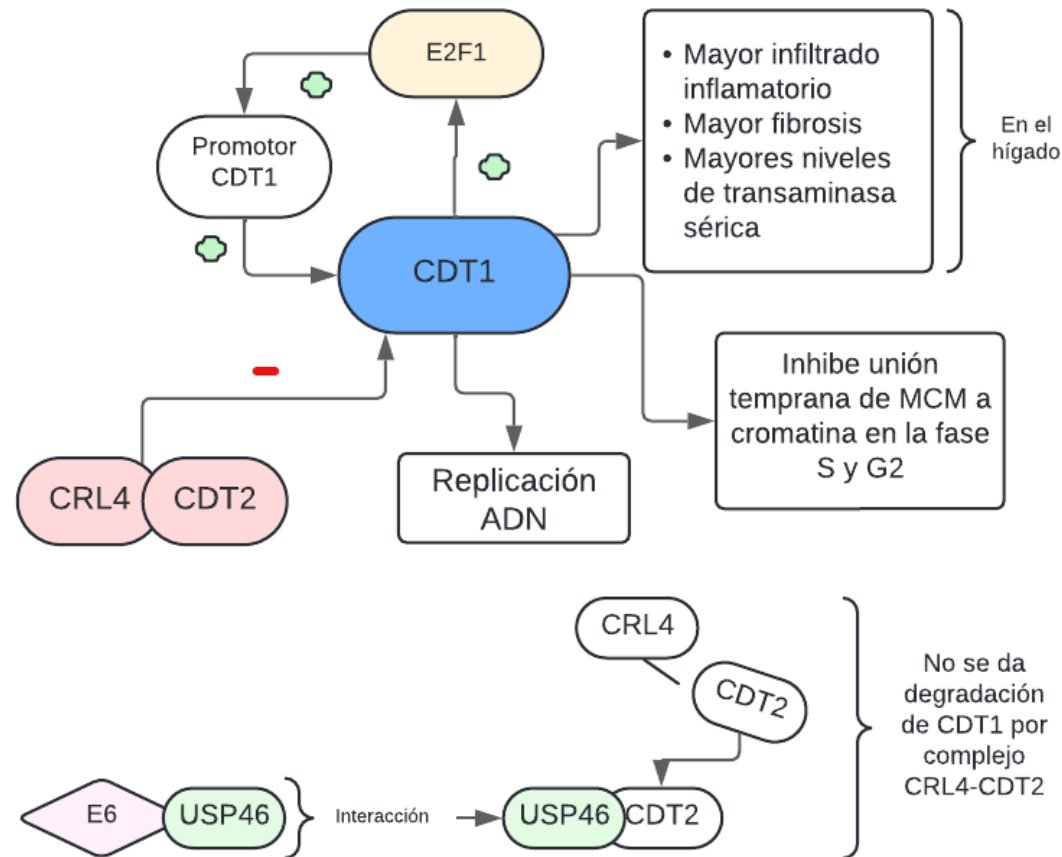
En el estudio de Kiran y cols (2018) indican que CDT1 se degrada por el complejo CRL4-CDT2. El E6 del HPV interactúa a su vez con USP46 y promueve una interacción entre USP46 y CDT2, causando una estabilización de CDT2 y por lo tanto el complejo CRL4-CDT2 no degrada sus varias enzimas dianas entre ellas CDT1. Esto permite la proliferación de células HPV+ porque CDT1 se mantiene con niveles altos (102).

La expresión elevada de CDT1 se asoció al grado tumoral y a la metástasis mayor y a una supervivencia general reducida. En el carcinoma hepatocelular resultó ser un biomarcador pronóstico de valor (103).

Se ha visto que la inhibición de la expresión de CDT1 provocó una disminución de la tasa de replicación de DNA (93).

El análisis de supervivencia de CDT1 demuestra una tendencia a un mejor pronóstico en pacientes con sobreexpresión de CDT1, aunque a los 30 meses hay una inversión de esta tendencia y parece que la sobreexpresión afecta negativamente al pronóstico (Figura 5). Se debería estudiar el efecto de este gen sobre el pronóstico en más estudios.

Hemos elaborado un *Flow chart* que nos ayuda en visualizar todos los procesos, funciones y moléculas que se vean afectadas por CDT1 en relación con el cáncer. Se puede ver en la figura de abajo (Figura 11).



**Figura 11.** Flow Chart de influencia del gen CDT1 sobre procesos y componentes biológicos. Además, se visualiza una posible explicación de la sobreexpresión en casos de HPV, por la menor degradación de CDT1 mediante el complejo CRL4-CDT2, dado que el E6 interactúa con USP46 lo cual provoca una unión de USP46 y CDT2. Elaboración propia

## TET1

El gen TET1 demuestra resultados variados en cuanto a la relación entre su sobreexpresión y la promoción de la oncogénesis. En nuestro estudio ha sido hallado como un gen sobreexpresado en pacientes HNSCC-HPV+.

Según nuestro análisis este gen enriquece funciones biológicas relacionadas a la remodelación y organización de la cromatina. Aparte, Alzahayga y cols (2022) y Stolz y cols (2022) comentan de su función regulador sobre la expresión génica por su acción de convertir 5-metilcitosina en 5-hidroximetilcitosina, 5-formilcitosina y 5-carboxilcitosina. Este mecanismo tiene un efecto sobre la homeostasis tisular (104, 105).

Este gen ha sido hallado con sobreexpresión en carcinomas esofágicos, carcinoma de cabeza y cuello, adenocarcinomas pulmonares, carcinomas pulmonares de células escamosas y sarcomas. Tiene una función promotora sobre la oncogénesis cuando esté bajo condiciones de hipoxia en este caso será indicador de un peor pronóstico general. Aunque se hayan observado su sobreexpresión en los nombrados cánceres, también se ha visto infraexpresado en otros cánceres como el carcinoma de mama invasivo, cromófobos renales, carcinomas renales de células papilares y en carcinomas tiroideos (106).

El estudio de Alzahayga y cols (2022) comenta que la sobreexpresión suprime fenotipos oncogénicos en el cáncer de mama, y que la infraexpresión da un pronóstico desfavorable, aparte de que la sobreexpresión ha sido asociada a la inhibición de la metástasis e invasión tumoral, aunque es cierto que comenta que este gen ha sido relacionado tanto con la oncogénesis como con la inhibición de ella (104). Se ha visto desregulado en el cáncer cervical, gástrico, hepático, del colon, de mama, pulmonar, tiroideo y esofágico (104, 107).

Chen y cols (2019) indican que aquellos estudios que lo relacionan a la supresión tumoral podrían contener sesgos de selección y posibles errores o afirmaciones sin respaldo. En su estudio detectan un pronóstico desfavorable en el carcinoma ovárico con sobreexpresión de TET1 (108).

TET1 parece promover la transcripción de CK2B, incrementar la fosforilación de AKT, GSK3 $\beta$  y aumentar los niveles de  $\beta$ -catenina, pudiendo por estas vías influir sobre

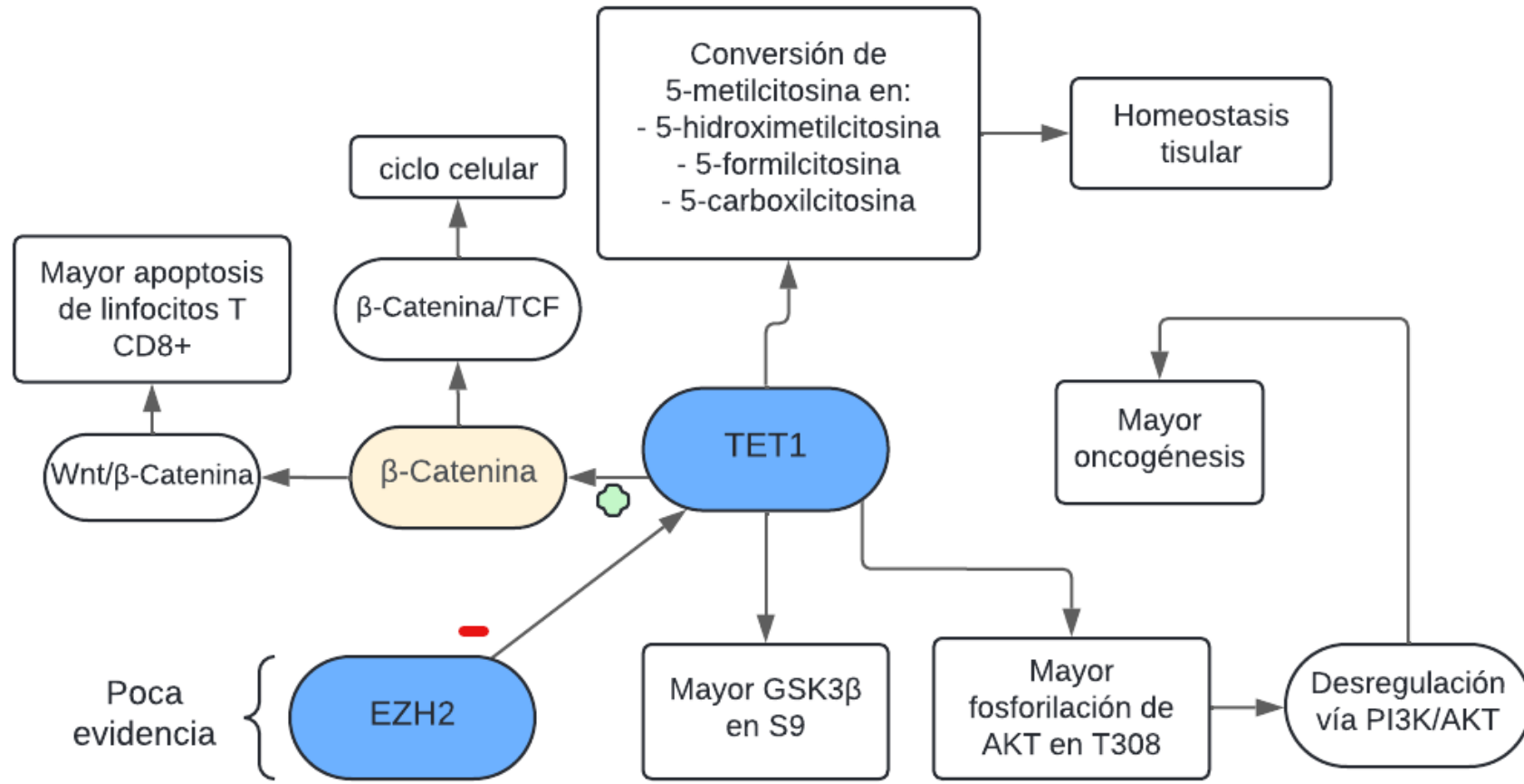
la oncogénesis (106). El uso de inhibidores de Ck2a ha demostrado ayudar en el tratamiento (106, 108).

En cuanto al HPV no se ha estudiado la posible relación entre este gen y la progresión del HPV. Sería interesante estudiar esta área con mayor hincapié en una posible relación entre ellos y mecanismos que podrían estar involucrados.

En cuanto a nuestros otros genes de interés, según Alzahayga y cols (2022) el EZH2 regula negativamente la expresión de TET1 (104). En nuestro estudio se ha visto una sobreexpresión de ambos genes. Se tendría que estudiar esta posible relación entre TET1 y EZH2 de mejor manera para concluir los posibles efectos que uno podría tener sobre el otro.

En nuestro análisis de supervivencia la sobreexpresión de TET1 demostró tener efectos positivos sobre la supervivencia de pacientes HNSCC-HPV+, aunque a partir de los 30 meses parece haber una inversión de esta tendencia (Figura 5).

En la siguiente figura se puede observar un *Flow chart* con los procesos y componentes biológicos asociados a este gen en relación con procesos oncogénicos (Figura 12).



**Figura 12.** Flow Chart de los procesos y componentes influenciados por TET1 con relación a la oncogénesis. Se aprecia su influencia sobre la β-catenina y la vía PI3K/AKT. Elaboración propia

## **BIRC6**

Otro gen de nuestro estudio que tiene influencia sobre la vía  $\beta$ -catenina es el BIRC6, el cual activa esta vía por su efecto degradador sobre la axina y la consiguiente acumulación de  $\beta$ -catenina. Esto promueve la transición epitelial-mesenquimal y da una mayor resistencia a la quimioterapia aparte de una mayor proliferación celular (109).

Según nuestro análisis este gen está implicado en procesos biológicos vinculados a la división y el ciclo celular. Además, inhibe la apoptosis, interactúa con enzimas, y participa en la mitosis por su acción sobre la formación y el movimiento de los *spindles* desde los microtúbulos intermediarios hasta el anillo intermedio. En cánceres de hígado, aparte, se ha visto que promueve la degradación de p53 y por ello promueve el avance del proceso oncogénico (109).

En nuestro estudio este gen ha resultado sobreexpresado en pacientes HNSCC-HPV+.

Se ha visto también una sobreexpresión de BIRC6 en cánceres epiteliales ováricos, cánceres de la vejiga, cánceres renales, colorrectales, pulmonares de células no pequeñas, de próstata y de mama triple negativos (110-117). Aparte, ha sido relacionada al OSCC (118).

La sobreexpresión en los tejidos colorrectales se asoció a una supervivencia general peor y un menor tiempo de supervivencia sin enfermedad (113). Aparte, el *knockdown* ha sido asociado a una menor proliferación (111, 113, 116). Hu y cols (2015) también pudieron observar una mayor sensibilidad al tratamiento de quimioterapia, una parada del ciclo celular en la fase S1 en aquellas células en las cuales se efectuó un *knockdown* (113). Añadiendo a esto, el *knockdown* resultó en una menor transición epitelial-mesenquimal, menor invasión y migración (111). En los cánceres pulmonares con sobreexpresión de BIRC6 se observaron tasas de metástasis mayores, una mayor involucración ganglionar, un grado tumoral peor y una peor supervivencia a los 3 años con mayores tasas de recurrencias (115). Tang y cols (2015) refuerzan la esta tendencia viendo un mal pronóstico y altas recurrencias en pacientes con carcinoma hepatocelular con la expresión elevada de BIRC6. El *knockdown* causó una menor proliferación, una mayor sensibilidad al tratamiento con sorafenib y una detención del ciclo celular en G1/S (119).

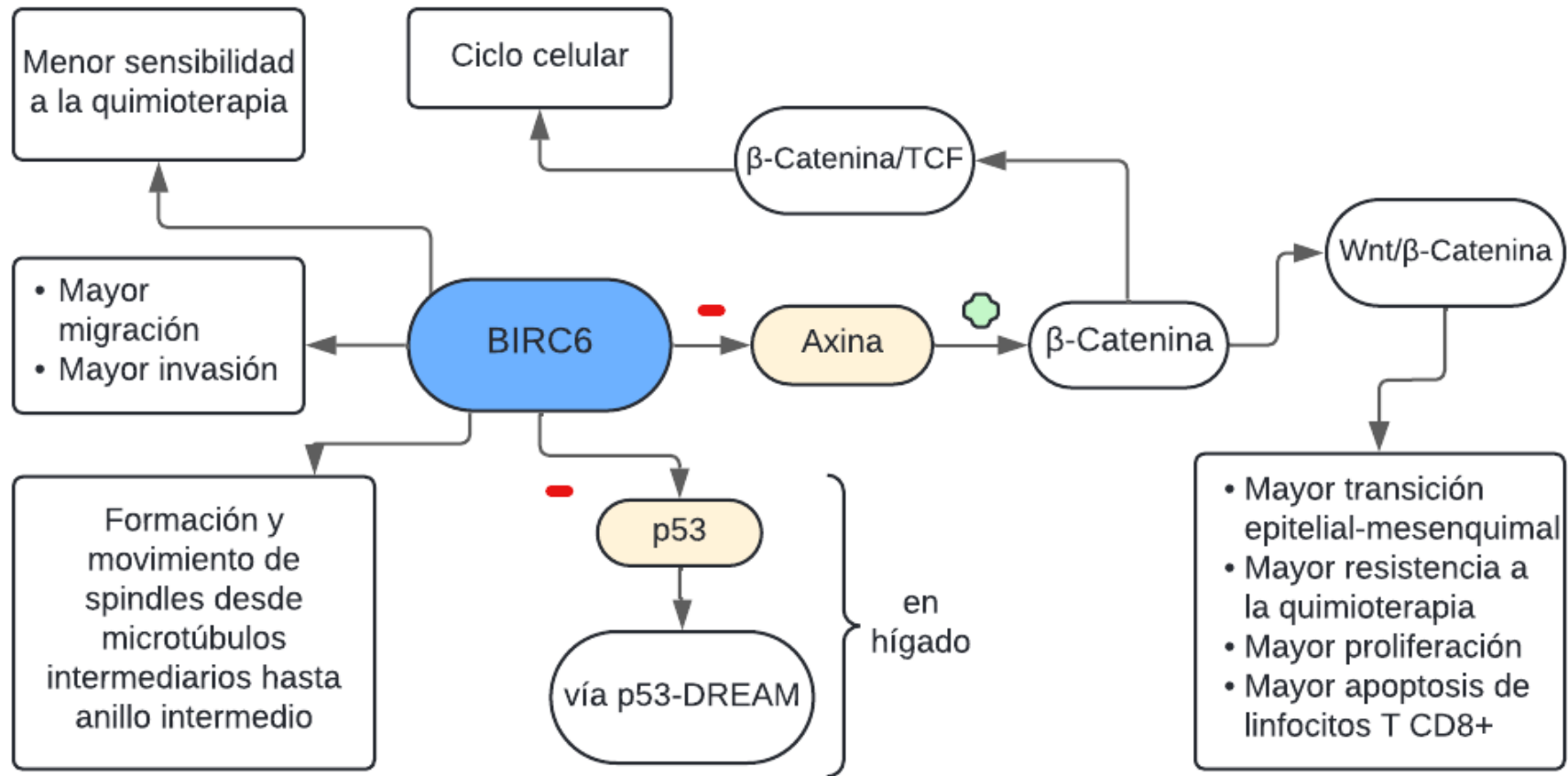


Según Zhao y cols (2023) y Tang y cols (2015) BIRC6 degrada p53, lo cual tendría efectos antiapoptóticos (119, 120). Esto se refuerza con el hecho de que Wang y cols (2019) lograron un aumento en los niveles de expresión de p53 cuando disminuyeron BIRC6 mediante miR-204 (110).

En general, este gen demuestra una fuerte asociación a los cánceres y con el mal pronóstico de estos. Podría entonces ser un buen biomarcador de pronóstico para determinar de antemano cuales pacientes podrían estar a mayor riesgo de desarrollar un cáncer más agresivo. Además, como la sobreexpresión está asociada a una resistencia mayor a la quimioterapia, según como comentan Dong y cols (2013) y Van Houdt y cols (2011), podría ser efectivo enfocar el tratamiento en la inhibición de este gen (114, 115).

En el análisis de supervivencia efectuado en nuestro estudio se ve una tendencia fuerte de un mejor pronóstico en aquellos pacientes HPV+ con una sobreexpresión de este gen. Es el único gen con influencia significativa ( $p$ valor = 0,02) sobre la supervivencia del grupo HNSCC-HPV+ (Figura 5). Esto es contrario a la postulación de que la sobreexpresión aumente la resistencia a la quimioterapia.

En la próxima página se puede observar el *Flow chart* elaborado para la mejor visualización de los efectos *downstream* de BIRC6 (Figura 13).



**Figura 13.** Flow Chart de BIRC6 en el cual se aprecian los componentes y procesos biológicos asociados. Elaboración propia

## SPC24

Otro gen que parece tener relación a la vía PI3K/AKT es el SPC24, el cual ha demostrado sobreexpresarse en caso de activar a la vía PI3K/AKT (121). Según Zhou y cols (2018) SPC24 regula a la vía PI3K/AKT (122).

Este gen salió sobreexpresado en pacientes HNSCC-HPV+ en nuestro análisis.

En los cánceres se ha observado una sobreexpresión el cáncer de próstata, cáncer pulmonar, cáncer laríngeo de células escamosas, cánceres de mama, cánceres tiroideos, carcinoma renal (122-126).

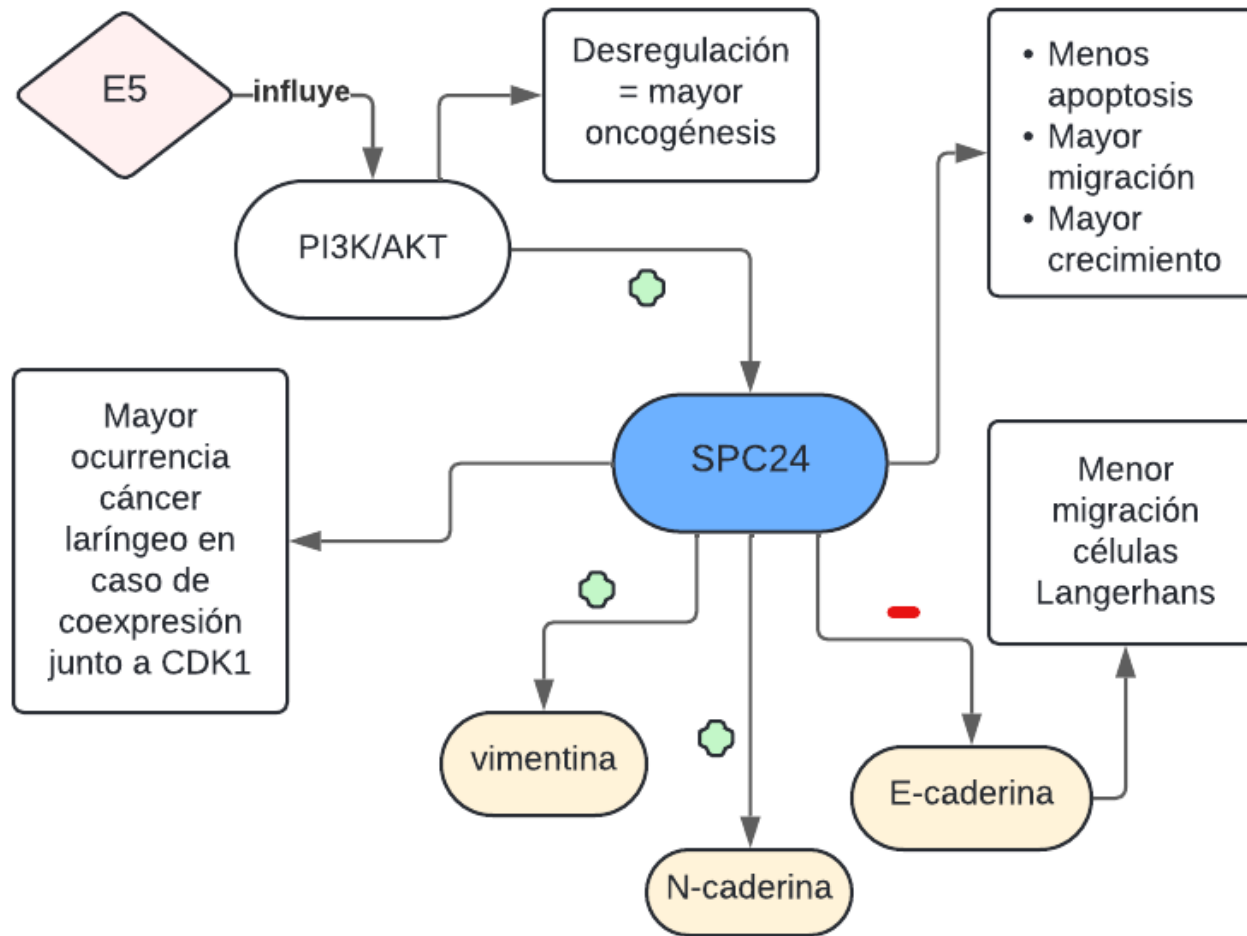
La sobreexpresión de este gen en los carcinomas renales ha sido asociada a un estadio tumoral más grave y un peor pronóstico (125). Esto concuerda con el estudio de Chen y cols (2021) en el cual se observó una asociación entre la sobreexpresión y el estadio tumoral, una mayor tasa de metástasis y un pronóstico peor (123).

El *knockdown* de SPC24 se asoció a una menor capacidad de migración, un crecimiento lento, apoptosis incrementada, mayores niveles de E-caderina, y menores niveles de vimentina y n-caderina (122, 124). Yin y cols (2017) también observaron un crecimiento menor, una menor capacidad invasiva, mayores tasas de apoptosis y un aumento de los niveles de e-caderina en células en los cuales se efectuó un *knockdown* de SPC24. La e-caderina es una enzima que está involucrada en la transición epitelial-mesenquimal, más específicamente induce a este proceso cuándo sus niveles están disminuidos, y por lo tanto aumenta la capacidad invasiva de las células cancerosas (127).

En general SPC24 parece tener gran vínculo a la oncogénesis, lo cual podría ser a través de su influencia sobre la vía PI3K/AKT, además de su influencia sobre los niveles de E-caderina, vimentina y N-caderina, pero se debería estudiar en mayor profundidad el mecanismo detrás de esta promoción oncogénica y el posible vínculo con el HPV. Sí que el HPV parece influir sobre la vía PI3K/AKT (84, 122, 128). Quizás por este mecanismo se da la sobreexpresión de SPC24 y de ahí la exacerbación del proceso tumoral.

En nuestro análisis de supervivencia aquellos pacientes HPV+ con una sobreexpresión de SPC24 demostraron tener una supervivencia general mejorada en comparación con los HNSCC-HPV- (Figura 5). Esto es contrario a lo hallado en el estudio de Chen y cols y Fu y cols. Se tendría que estudiar esta tendencia en futuros estudios.

En el *Flow Chart* de la siguiente página se pueden visualizar los procesos y componentes influenciados por SPC24 (Figura 14).



**Figura 14.** Flow Chart de los componentes y procesos biológicos influenciados por SPC24 en relación con el cáncer. Elaboración propia

## **SETMAR**

El SETMAR es un gen que ha salido sobreexpresado en nuestro análisis y que tiene involucración en los procesos de remodelación y organización de la cromatina. Según Tellier y cols (2021) participa en procesos de reparación de DNA, regularización génica, integración ilegítima de DNA, la decatenación de DNA y las mitosis celulares (129, 130)

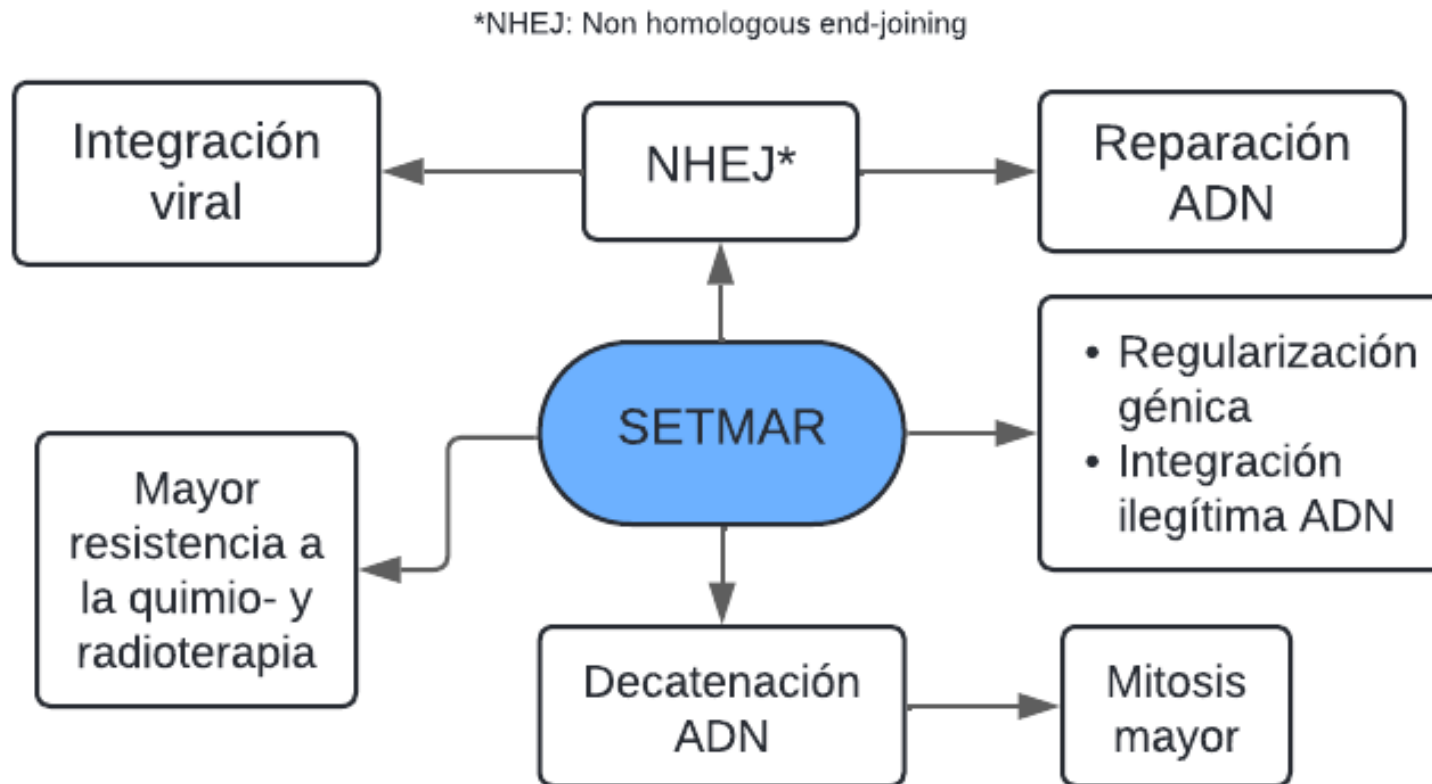
Aparte, podría facilitar la integración de virus al genoma del huésped por su rol modificador sobre la cromatina (131).

En general hay pocos estudios que relacionan este gen al cáncer, pero Tellier y cols (2021) observaron un incremento en la tasa de crecimiento en caso de sobreexpresión de SETMAR (129). Además, se vio una sobreexpresión en el glioblastoma, pero los niveles altos en los tejidos perilesionales dieron un mejor pronóstico (132). Fnu y cols (2011) observaron una resistencia mayor a la quimioterapia y radioterapia en casos de sobreexpresión (133).

Williamson y cols (2008) hicieron un estudio sobre el SETMAR en relación con los lentivirus y observaron una correlación entre la integración del virus y los niveles altos de SETMAR (131). Podría entonces estar jugando un rol similar en el HPV, pero se tendría que estudiar esto.

En nuestro análisis de supervivencia SETMAR demostró tener un efecto positivo sobre la supervivencia general de los pacientes HNSCC-HPV+ con sobreexpresión de SETMAR (Figura 6). Esto es contrario a lo postulado por Fnu y cols (2011), dado que observaron una resistencia mayor a la quimio- y radioterapia en casos de sobreexpresión. Se debería estudiar entonces el efecto de la expresión de SETMAR sobre el pronóstico en futuros estudios para establecer su rol de mejor manera.

Para una mejor visualización de la influencia de SETMAR sobre los procesos biológicos relacionados al cáncer, hemos desarrollado el siguiente Flow chart (Figura 15).



**Figura 15.** Flow Chart de la influencia de SETMAR sobre procesos biológicos relacionados con el cáncer y el HPV. Elaboración propia

## **KMT5C**

KMT5C es un gen con influencia sobre procesos de la remodelación y organización de la cromatina. En nuestro análisis diferencial este gen salió sobreexpresado.

El rol de este gen es el de metilizar a H4K20me1 a H4K20me2 y posteriormente a H4K20me3. En tejidos adiposos está involucrada en procesos de expresión génica (134).

Hay pocos estudios que relacionen este gen con el cáncer y no hay consenso entre ellos. Zeng y cols (2024) comentan que ha sido hallado sobreexpresado en el carcinoma hepático y en el carcinoma renal de células claras, pero infraexpresado en el cáncer de mama con altas tasas de metástasis (135).

La depleción de KMT5C ha sido asociada a la promoción de la elongación de los telómeros y por ello a la tumorigénesis (136). Zhao y cols (2020) pudieron observar un aumento de los niveles de p53 en tejidos adiposos con la depleción de KMT5C (134).

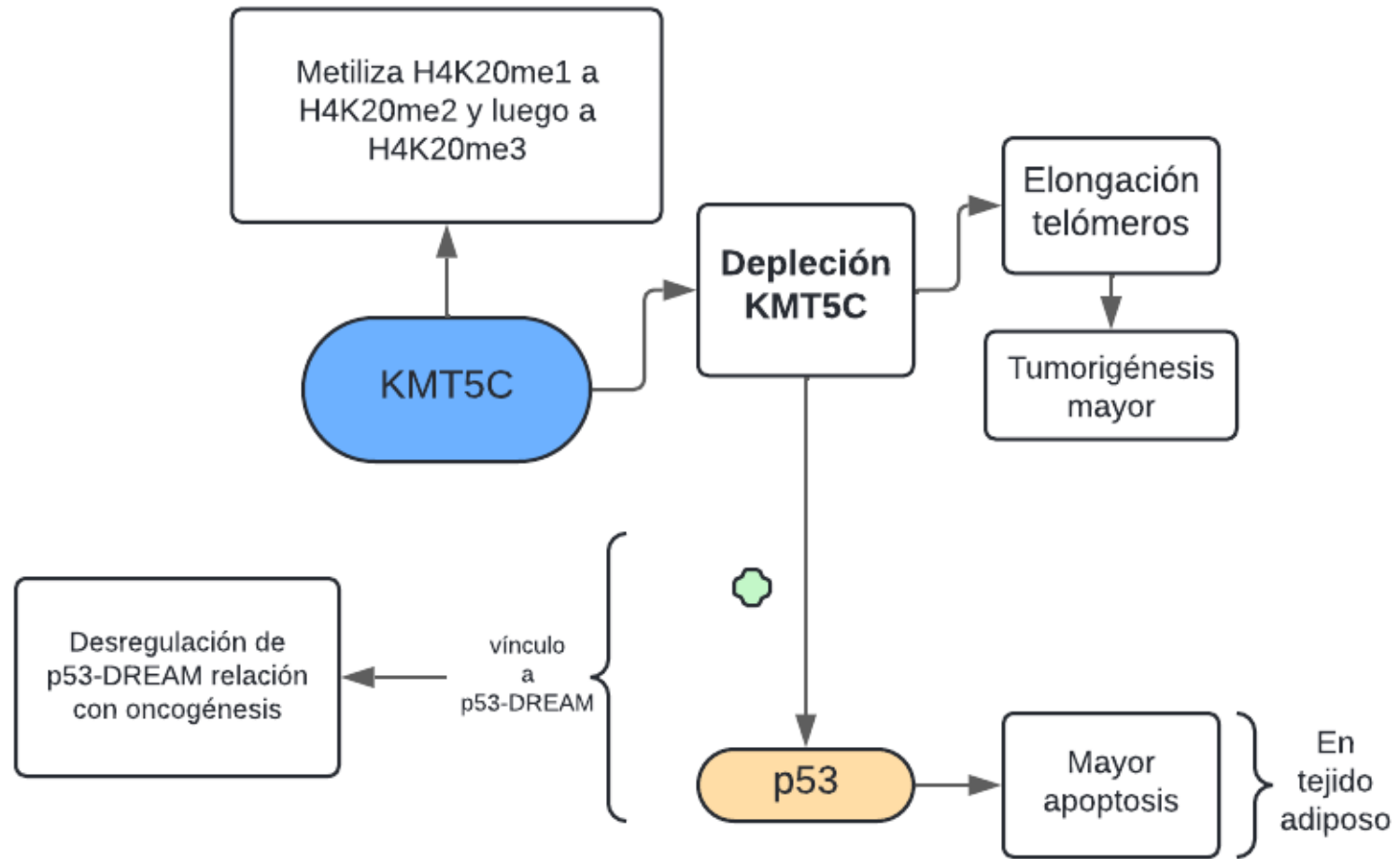
En el cáncer de mama la pérdida de la trimetilación de H4K20 provocó un peor pronóstico (137).

En nuestro análisis de supervivencia la sobreexpresión de KMT5C demostró tener un efecto positivo sobre la supervivencia de pacientes HSNCC-HPV+ (Figura 6).

En la figura de la próxima página se visualizan los vínculos de este gen a procesos y componentes biológicos relacionados al cáncer, aunque cabe decir que este gen debe ser estudiado en mayor profundidad para establecer mejor los posibles vínculos con la carcinogénesis (Figura 16).

En general este gen se tiene que estudiar en mayor medida, dado que hay muy poca evidencia científica que lo vincula al cáncer, y ninguna que lo relaciona con el HPV. Además, no se entiende muy bien en cuales vías biológicas podría estar influyendo de cual manera.





**Figura 16.** Flow Chart del gen KMT5C en el cual se puede observar un posible vínculo a la vía p53-DREAM. Elaboración propia.

## **SPRR2G**

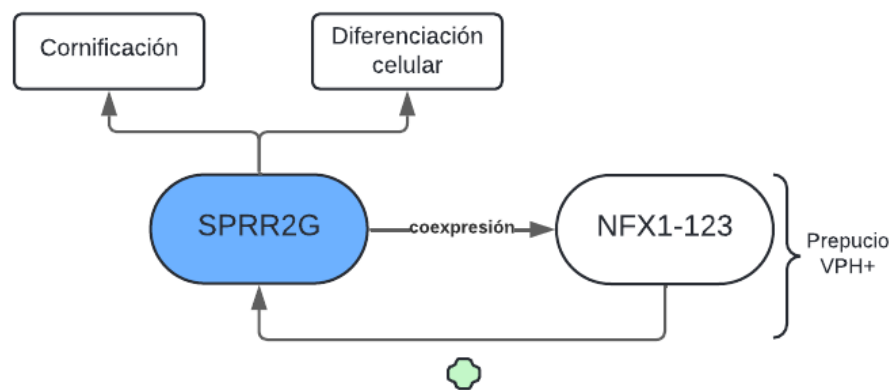
El gen SPRR2G es el único de los genes destacados que se halló con infraexpresión en los pacientes HNSCC-HPV+. Influye en procesos de queratinización, diferenciación de queratinocitos y en el desarrollo de la epidermis según nuestro estudio. Rakha y cols (2020) especifican que este gen codifica para una proteína de queratinocitos y que luego interactúa con la transglutimasa y así se da la unión a proteínas de la membrana. Además, influye sobre la cornificación y la diferenciación de células epiteliales (138).

Es un gen poco estudiado y no hay mucha información sobre su vínculo al cáncer ni al HPV, pero se ha hallado sobreexpresado en otro estudio sobre el HNSCC de Chintala y cols (2022), y la sobreexpresión ha sido asociado al grado tumoral según indica Rakha y cols (2022) (138, 139).

Según Chintala y cols no tiene relación con el HPV (139). Vliet-Gregg y cols (2013) indican que la sobreexpresión estuvo asociada a una expresión elevada de NFX1-123 en queratinocitos del prepucio humano HPV positivos (140). Según Chintala y cols (2022) el NFX1-123 sobrerregula SPRR2G (139). Esto iría en contra de nuestro hallazgo, pero dado que hay muy poca evidencia hace falta estudiar este gen con mayor profundidad y con mayor hincapié en la relación al proceso cancerosos y al HPV.

En el análisis de supervivencia la infraexpresión de SPRR2G en pacientes HPV+ dio un mejor pronóstico (Figura 6).

En la siguiente figura se observan los vínculos de este gen a procesos relacionados a la oncogénesis, aunque igual como pasa con el KMT5C, hay una falta de estudios para crear una red más completa de las asociaciones de SPRR2G al cáncer (Figura 17).



**Figura 17.** *Flow Chart* del gen SPRR2G con su vínculo a los procesos de la cornificación y diferenciación celular, aparte de su vínculo a NFX1-123 en el prepucio HPV+. Elaboración propia

### Relación entre el HPV y el cáncer

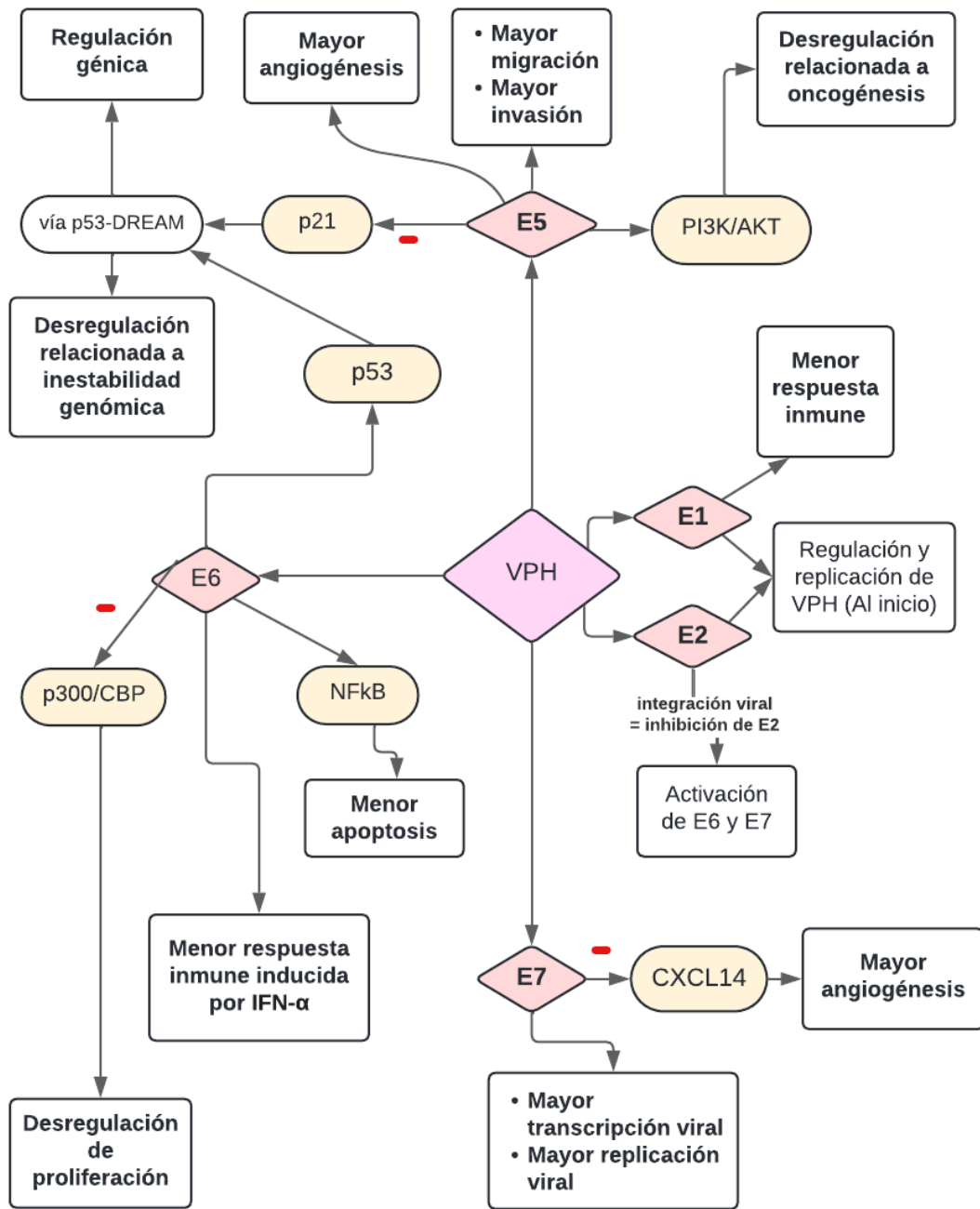
La relación entre el HPV y el cáncer se ha establecido principalmente en los cánceres cervicales. Sin embargo, el 25% de los casos de HNSCC están asociados a HPV, lo que ha llevado a un aumento en el estudio de la relación entre estas dos patologías (141).

El HPV actúa por una gran cantidad de vías para facilitar su replicación y progreso, y algunas de estas vías como la PI3K/AKT, la p53-DREAM, la NF- $\kappa$ B están asociados a la oncogénesis (142). Hay entonces un fuerte vínculo entre el HPV y la oncogénesis.

En cuanto al pronóstico, los cánceres asociados a HPV suelen tener un pronóstico mejor en comparación con los casos HPV negativos. De hecho, la mayoría de los estudios indican un mejor pronóstico en muestras tumorales HPV+ (143 – 150), aunque algunos no encuentran diferencias significativas (151, 152).

Para una visualización completa de los efectos del HPV sobre los componentes celulares y los procesos relacionados con la oncogénesis hemos elaborado un *Flow Chart*, que muestra como las proteínas virales E1, E2, E5, E6 y E7 interactúan con componentes y vías biológicas clave (Figura 18).

Es importante entonces detectar aquellos pacientes HPV+ con un riesgo elevado de desarrollar un proceso canceroso, y para ello sería interesante utilizar biomarcadores como los propuestos en nuestro estudio.



**Figura 18.** Flow Chart del impacto del HPV sobre procesos celulares vinculados a la oncogénesis. Se aprecian los vínculos de E1, E2, E5, E6 y E7 a componentes y vías biológicas involucradas en la oncogénesis. Elaboración propia.

## **Limitaciones**

Nuestro estudio presenta algunas limitaciones, como la escasez de investigaciones que vinculen el HPV con la mayoría de los genes discutidos, por lo cual podrían existir relaciones que no se han investigado todavía y, por lo tanto, que no hemos podido comentar. Además, muchos estudios se centran en cánceres distintos al de cabeza y cuello o no especifican la zona específica del HNSCC. Podría existir también un sesgo de publicación, dado que aquellos estudios con resultados no significativos son menos propensos a ser publicados. Esto podría influir en la visión global de los hechos. Además, la influencia de los factores personales, como hábitos tabáquicos, de consumo de otras sustancias o la dieta, puede no estar ajustada en algunos estudios, impactando en los resultados de pronóstico.

## **11. Conclusión**

El presente trabajo aporta información valiosa acerca de las diferencias genéticas en pacientes HNSCC-HPV+ frente a aquellos HPV-, reforzando la idea de que el HPV mejora el pronóstico en pacientes con procesos cancerígenos. Se ha obtenido una expresión diferencial de 5.955 genes en las muestras HNSCC-HPV+ en comparación con las muestras HNSCC-HPV-.

Los análisis GSEA y DAVID han identificado funciones biológicas asociadas a estos genes diferenciales, destacando procesos clave como el ciclo celular, la división celular, la segregación de cromosomas, la reparación y replicación de ADN, el RNA *splicing*, la recombinación de ADN, la remodelación y organización de cromatina, la queratinización, la diferenciación de queratinocitos, el desarrollo de la epidermis y la cornificación, entre otros.

De los genes diferenciales, hemos destacado a EZH2, HDAC10, KDM5A, CDK20, CDT1, TET1, SPC24, BIRC6, SETMAR, KMT5C y SPRR2G, debido a su AUC>0,7 y su relación a procesos de interés como al ciclo celular, la división celular, la organización y la remodelación de la cromatina, la queratinización, el desarrollo de la epidermis y la diferenciación de queratinocitos. Estos genes demostraron una tendencia a impactar positivamente la supervivencia de pacientes HNSCC-HPV+.

Destacamos los genes EZH2, BIRC6 y CDT1 como los genes con mayor consenso y evidencia sobre el hecho de que su sobreexpresión influye negativamente sobre la supervivencia de los pacientes. Sin embargo, nuestros análisis de supervivencia mostraron un efecto positivo sobre el pronóstico en el caso de la sobreexpresión de estos genes en el grupo HNSCC-HPV+, lo cual contrasta con la literatura existente.

Este estudio ayuda a identificar genes que podrían aumentar la frecuencia de desarrollo de cáncer en pacientes HPV+, permitiendo identificar a estos pacientes y diferenciarlos de pacientes HPV+ con menor riesgo de desarrollar cáncer.

En general nuestro estudio es de gran valor por la gran cantidad de genes diferenciales hallados en los pacientes de HNSCC-HPV+ y la evidencia de que su expresión influye positivamente en la supervivencia.

Existe una diferencia significativa en la expresión génica de pacientes HNSCC-HPV+ frente a HNSCC-HPV- y los DEGs tienen una relación a procesos biológicos estrechamente relacionados con el desarrollo de procesos oncogénicos. El HPV parece activar varias vías biológicas que contribuyen al desarrollo tumoral.

En particular, los genes EZH2, BIRC6 y CDT1 podrían ser buenos biomarcadores de pronóstico para pacientes HPV+, ayudando a establecer su riesgo de desarrollar un cáncer en un futuro. Sin embargo, son necesarios más estudios in vitro y ex vivo para validar estos hallazgos.

## 12. BIBLIOGRAFÍA

1. Vigneswaran N, Williams MD. Changing epidemiology of head and neck cancer. *Oral Maxillofac Surg Clin North Am.* 2014;26(2):123–41.
2. Liu C, Wang M, Zhang H, Li C, Zhang T, Liu H, et al. Tumor microenvironment and immunotherapy of oral cancer. *Eur J Med Res.* 2022;27(1):1–19. Available from: <https://doi.org/10.1186/s40001-022-00835-4>
3. Abati S, Bramati C, Bondi S, Lissoni A, Trimarchi M. Oral cancer and precancer: A narrative review on the relevance of early diagnosis. *Int J Environ Res Public Health.* 2020;17(24):1–14.
4. Gormley M, Creaney G, Schache A, Ingarfield K, Conway DI. Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. *Br Dent J.* 2022;233(9):780–6.
5. Ganesh D, Dafar A, Niklasson J, Sandberg I, Braz-Silva P, Sapkota D, et al. EZH2 Expression Correlates With T-Cell Infiltration in Oral Leukoplakia and Predicts Cancer Transformation. *Anticancer Res.* 2023;43(4):1533–42.
6. Barsouk A, Aluru JS, Rawla P, Saginala K, Barsouk A. Epidemiology, Risk Factors, and Prevention of Head and Neck Squamous Cell Carcinoma. *Med Sci.* 2023;11(2):42.
7. Nogues JC, Fassas S, Mulcahy C, Zapanta PE. Human papillomavirus-associated head and neck cancer. *J Am Board Fam Med.* 2021;34(4):832.
8. Centers for Disease Control and Prevention (CDC). Human papillomavirus-associated cancers - United States, 2004-2008. *MMWR Morb Mortal Wkly Rep.* 2012 Apr 20;61:258-61. PMID: 22513527
9. Roman BR, Aragonés A. Epidemiology and incidence of HPV-related cancers of the head and neck. 2022;124(6):920–2.
10. Wittekindt C, Wagner S, Sharma SJ, Würdemann N, Knuth J, Reder H, et al. HPV - A different view on Head and Neck Cancer. *Laryngorhinootologie.* 2018;97(S 01):S48–113. Available from: <https://pubmed.ncbi.nlm.nih.gov/29905354/>
11. Jarosz-Biej M, Smolarczyk R, Cichoń T, Kułach N. Tumor microenvironment as a “game changer” in cancer radiotherapy. *Int J Mol Sci.* 2019;20(13):18–20.
12. Arneth B. Tumor microenvironment. *Med.* 2020;56(1).

13. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep.* 2014 Dec;15(12):1243-53. doi: 10.15252/embr.201439246. PMID: 25381661; PMCID: PMC4264927.
14. Karamanos NK, Theocharis AD, Piperigkou Z, Manou D, Passi A, Skandalis SS, et al. A guide to the composition and functions of the extracellular matrix. *FEBS J.* 2021;288(24):6850–912.
15. Theocharis AD, Manou D, Karamanos NK. The extracellular matrix as a multitasking player in disease. *FEBS J.* 2019;286(15):2830–69.
16. Walker C, Mojares E, Del Río Hernández A. Role of extracellular matrix in development and cancer progression. *Int J Mol Sci.* 2018;19(10):3028.
17. Poudel P, Srii R, Marla V. Oral cancer awareness among undergraduate dental students and dental surgeons: A descriptive cross-sectional study. *J Nepal Med Assoc.* 2020;58(222):102–7.
18. Warnakulasuriya S, Kerr AR. Oral Cancer Screening: Past, Present, and Future. *J Dent Res.* 2021 Nov;100(12):1313-1320. PMID: 34036828; PMCID: PMC8529297.
19. Woo SB. Oral epithelial dysplasia and premalignancy. *Head Neck Pathol.* 2019;13(3):423–39. Available from: <http://dx.doi.org/10.1007/s12105-019-01020-6>
20. Li B, Gu ZY, Yan KX, Wen ZN, Zhao ZH, Li LJ, et al. Evaluating oral epithelial dysplasia classification system by near-infrared Raman spectroscopy. *Oncotarget.* 2017 Jul 18;8(44):76257-76265. PMID: 29100309; PMCID: PMC5652703.
21. Joseph BK. Oral Cancer: Prevention and Detection. *Med Princ Pract.* 2002 Jul 1;11(Suppl. 1):32–35. <https://doi.org/10.1159/000057776>
22. Rautava J, Syrjänen S. Biology of Human Papillomavirus Infections in Head and Neck Carcinogenesis. *Head Neck Pathol.* 2012;6(Suppl. 1):3–15.
23. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of human papillomavirus-positive head and neck squamous cell carcinoma. *J Clin Oncol.* 2015;33(29):3235–42.
24. Pal A, Kundu R. Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy. *Front Microbiol.* 2020;10:.
25. Chintala S, Quist KM, Gonzalez-DeWhitt PA, Katzenellenbogen RA. High expression of NFX1-123 in HPV positive head and neck squamous cell carcinomas. *Head Neck.* 2022;44(1):177–88.



26. Melo BA de C, Vilar LG, Oliveira NR de, Lima PO de, Pinheiro M de B, Domingueti CP, et al. Human papillomavirus infection and oral squamous cell carcinoma - a systematic review. *Braz J Otorhinolaryngol*. 2021;87(3):346–52.
27. Ruutu M, Peitsaro P, Johansson B, Syrjänen S. Transcriptional profiling of a human papillomavirus 33-positive squamous epithelial cell line which acquired a selective growth advantage after viral integration. *Int J Cancer*. 2002;100(3):318–26.
28. Senapati R, Senapati NN, Dwibedi B. Molecular mechanisms of HPV mediated neoplastic progression. *Infect Agent Cancer*. 2016;11.
29. Mukherjee A, Ye Y, Wiener HW, Kuniholm MH, Minkoff H, Michel K, et al. Variations in Genes Encoding Human Papillomavirus Binding Receptors and Susceptibility to Cervical Precancer. *Cancer Epidemiol Biomarkers Prev*. 2023;32(9):1190–7.
30. Hübbers CU, Akgül B. HPV and cancer of the oral cavity. *Virulence*. 2015;6(3):244–8.
31. Solanke A, Tribhuvan K, Kanika. Genomics: An Integrative Approach for Molecular Biology. *Biotechnol Prospect*. 2015 Sep:234–70.
32. Khodadadian A, Darzi S, Haghi-Daredeh S, Eshaghi FS, Babakhanzadeh E, Mirabutalebi SH, et al. Genomics and transcriptomics: The powerful technologies in precision medicine. *Int J Gen Med*. 2020;13:627–40.
33. Dong Z, Chen Y. Transcriptomics: advances and approaches. *Sci China Life Sci*. 2013 Oct;56(10):960-7.
34. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol*. 2017;13(5).
35. Qi YX, Liu YB, Rong WH. [RNA-Seq and its applications: a new technology for transcriptomics]. *Yi Chuan*. 2011 Nov;33(11):1191-202.
36. Passos GA. Transcriptomics in Health and Disease, Second Edition. 2022:1–475.
37. Costa V, Angelini C, De Feis I, Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol*. 2010;2010.
38. Feng J, Meyer CA, Wang Q, Liu JS, Liu XS, Zhang Y. GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*. 2012;28(21):2782–8.
39. Ogunwobi OO, Mahmood F, Akingboye A. Biomarkers in colorectal cancer: Current research and future prospects. *Int J Mol Sci*. 2020;21(15).
40. Flach P. ROC Analysis. SpringerReference. 2012:1–12.

41. van Erkel AR, Pattynama PMT. ROC. *Eur J Radiol.* 1998;27:88–94.
42. Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J.* 2017;34:357–359.
43. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
44. Lin R, Dai S, Irwin RD, Heinloth AN, Boorman GA, Li L. Gene set enrichment analysis for non-monotone association and multiple experimental categories. *BMC Bioinformatics.* 2008;9.
45. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research.* 23 March 2022. doi:10.1093/nar/gkac194.
46. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* 2009;4(1):44-57.
47. Chen Y, Feng Y, Yan F, Zhao Y, Zhao H, Guo Y. A Novel Immune-Related Gene Signature to Identify the Tumor Microenvironment and Prognose Disease Among Patients With Oral Squamous Cell Carcinoma Patients Using ssGSEA: A Bioinformatics and Biological Validation Study. *Front Immunol.* 2022;13.
48. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research.* 2016;5:1408.
49. Mootha V, Lindgren C, Eriksson KF, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34:267–273.
50. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003 Nov;13(11):2498-504.
51. Schober P, Vetter TR. Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. *Anesth Analg.* 2018 Sep;127(3):792-798.
52. Yu W, Liu N, Song X, Chen L, Wang M, Xiao G, et al. EZH2: An Accomplice of Gastric Cancer. *Cancers (Basel).* 2023;15(2).
53. Wang B, Liu Y, Liao Z, Wu H, Zhang B, Zhang L. EZH2 in hepatocellular carcinoma: progression, immunity, and potential targeting therapies. *Exp Hematol Oncol.* 2023;12(1).

54. Duan R, Du W, Guo W. EZH2: A novel target for cancer treatment. *J Hematol Oncol.* 2020;13(1).
55. Entezari M, Taheriazam A, Paskeh MDA, Sabouni E, Zandieh MA, Aboutalebi M, et al. The pharmacological and biological importance of EZH2 signaling in lung cancer. *Biomed Pharmacother.* 2023;160:114313.
56. Kim KH, Roberts CW. Targeting EZH2 in cancer. *Nat Med.* 2016 Feb;22(2):128-34.
57. Huang J, Gou H, Yao J, Yi K, Jin Z, Matsuoka M, et al. The noncanonical role of EZH2 in cancer. *Cancer Sci.* 2021;112(4):1376–82.
58. Bugide S, Gupta R, Green MR, Wajapeyee N. EZH2 inhibits NK cell–mediated antitumor immunity by suppressing CXCL10 expression in an HDAC10-dependent manner. *Proc Natl Acad Sci U S A.* 2021;118(30).
59. Hyland PL, McDade SS, McCloskey R, Dickson GJ, Arthur K, McCance DJ, et al. Evidence for Alteration of EZH2, BMI1, and KDM6A and Epigenetic Reprogramming in Human Papillomavirus Type 16 E6/E7-Expressing Keratinocytes. *J Virol.* 2011;85(21):10999–1006.
60. Holland D, Hoppe-Seyler K, Schuller B, Lohrey C, Maroldt J, Dürst M, et al. Activation of the enhancer of zeste homologue 2 gene by the human papillomavirus E7 oncoprotein. *Cancer Res.* 2008;68(23):9964–72.
61. Khattri M, Amako Y, Gibbs JR, Collura JL, Arora R, Harold A, et al. Methyltransferase-independent function of enhancer of zeste homologue 2 maintains tumorigenicity induced by human oncogenic papillomavirus and polyomavirus. *Tumour Virus Res [Internet].* 2023;16(March):200264. Available from: <https://doi.org/10.1016/j.tvr.2023.200264>
62. Dolcet X, Llobet D, Pallares J, Matias-Guiu X. NF-κB in development and progression of human cancer. *Virchows Arch.* 2005 May;446(5):475-82. doi: 10.1007/s00428-005-1264-9. Epub 2005 Apr 27. PMID: 15856292.
63. Bedoya FA. El factor de transcripción nuclear NF-κB en cáncer NF-κB transcription factor in cancer. *Horiz Médico.* 2023;23(1).
64. Zhong Y, Huang T, Huang J, et al. The HDAC10 instructs macrophage M2 program via deacetylation of STAT3 and promotes allergic airway inflammation. *Theranostics.* 2023;13(11):3568–81.

65. Xue C, Yao Q, Gu X, et al. Evolving cognition of the JAK-STAT signaling pathway: autoimmune disorders and cancer. *Signal Transduct Target Ther*. 2023;8(1).
66. Kassab A, Gupta I, Moustafa AE Al. Role of E2F transcription factor in oral cancer: Recent insight and advancements. *Semin Cancer Biol* [Internet]. 2023;92(February):28–41. Available from: <https://doi.org/10.1016/j.semcancer.2023.03.004>
67. El Baba R, Haidar Ahmad S, Monnien F, Mansar R, Bibeau F, Herbein G. Polyploidy, EZH2 upregulation, and transformation in cytomegalovirus-infected human ovarian epithelial cells. *Oncogene*. 2023;42(41):3047–61.
68. Nienstedt JC, Schroeder C, Clauditz T, et al. EZH2 overexpression in head and neck cancer is related to lymph node metastasis. *J Oral Pathol Med*. 2018; 47: 240–245. <https://doi.org/10.1111/jop.12673>
69. Varambally S, Dhanasekaran S, Zhou M, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002;419:624–9.
70. Kleer CG, Cao Q, Varambally S, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A*. 2003;100(20):11606–11.
71. Hubaux R, Thu KL, Coe BP, et al. EZH2 promotes E2F-driven SCLC tumorigenesis through modulation of apoptosis and cell-cycle regulation. *J Thorac Oncol*. 2013 Aug;8(8):1102-6. doi: 10.1097/JTO.0b013e318298762f. PMID: 23857401; PMCID: PMC3713495.
72. Gu YY, Zhou GN, Li Y, et al. HDAC10 Inhibits Cervical Cancer Progression through Downregulating the HDAC10-microRNA-223-EPB41L3 Axis. *J Oncol*. 2022;2022.
73. Islam MM, Banerjee T, Packard CZ, et al. HDAC10 as a potential therapeutic target in ovarian cancer. *Gynecol Oncol*. 2017 Mar;144(3):613-620. doi: 10.1016/j.ygyno.2017.01.009. Epub 2017 Jan 7. PMID: 28073598; PMCID: PMC6020686.
74. Liu X, Wang Y, Zhang R, et al. HDAC10 Is Positively Associated With PD-L1 Expression and Poor Prognosis in Patients With NSCLC. *Front Oncol*. 2020;10(April):1–7.
75. Yang Y, Huang Y, Wang Z, et al. HDAC10 promotes lung cancer proliferation via AKT phosphorylation. *Oncotarget*. 2016;7(37):59388–401.

- 76.** Guo H, Ren H, Han K, et al. Knockdown of HDAC10 inhibits POLE2-mediated DNA damage repair in NSCLC cells by increasing SP1 acetylation levels. *Pulm Pharmacol Ther.* 2023 Dec;83:102250. doi: 10.1016/j.pupt.2023.102250. Epub 2023 Aug 30. PMID: 37657752.
- 77.** Zhu J, Han S. Histone deacetylase 10 exerts antitumor effects on cervical cancer via a novel microRNA-223/TXNIP/Wnt/ $\beta$ -catenin pathway. *IUBMB Life.* 2021;73(4):690–704.
- 78.** Yang GJ, Zhu MH, Lu XJ, et al. The emerging role of KDM5A in human cancer. *J Hematol Oncol* [Internet]. 2021;14(1):1–18. Available from: <https://doi.org/10.1186/s13045-021-01041-1>
- 79.** Liu H, Lin J, Zhou W, et al. KDM5A Inhibits Antitumor Immune Responses Through Downregulation of the Antigen-Presentation Pathway in Ovarian Cancer. *Cancer Immunol Res.* 2022 Aug 3;10(8):1028-1038. doi: 10.1158/2326-6066.CIR-22-0088. PMID: 35726891; PMCID: PMC9357105.
- 80.** Oser MG, Sabet AH, Gao W, et al. The KDM5A/RBP2 histone demethylase represses NOTCH signaling to sustain neuroendocrine differentiation and promote small cell lung cancer tumorigenesis. *Genes Dev.* 2019;33(23–24):1718–38.
- 81.** Kataria A, Tyagi S. Domain architecture and protein–protein interactions regulate KDM5A recruitment to the chromatin. *Epigenetics* [Internet]. 2023;18(1):1–19. Available from: <https://doi.org/10.1080/15592294.2023.2268813>
- 82.** Porcheri C, Mitsiadis TA. Notch in Head and Neck Cancer. *Adv Exp Med Biol.* 2021;1287:81-103. doi: 10.1007/978-3-030-55031-8\_7.
- 83.** Chintala S, Quist KM, Gonzalez-DeWhitt PA, Katzenellenbogen RA. High expression of NFX1-123 in HPV positive head and neck squamous cell carcinomas. *Head Neck.* 2022;44(1):177–88.
- 84.** Stoiber S, Brkic FF, Maier T, et al.  $\beta$ -CATENIN is a positive prognostic marker for HPV-positive head and neck squamous cell carcinoma. *J Cancer Res Clin Oncol.* 2023 Aug;149(10):7743-7750. doi: 10.1007/s00432-023-04712-3. Epub 2023 Apr 3. PMID: 37010585; PMCID: PMC10374714.
- 85.** Bello JO, Nieva LO, Paredes AC, et al. Regulation of the Wnt/ $\beta$ -Catenin Signaling Pathway by Human Papillomavirus E6 and E7 Oncoproteins. *Viruses.* 2015 Aug 19;7(8):4734-55. doi: 10.3390/v7082842. PMID: 26295406; PMCID: PMC4576203.

86. Malumbres M. Cyclin-dependent kinases. *Genome Biol.* 2014;15(6):1–10.
87. Aghbash PS, Hemmat N, Baradaran B, et al. The effect of Wnt/ $\beta$ -catenin signaling on PD-1/PDL-1 axis in HPV-related cervical cancer. *Oncol Res.* 2022;30(3):99–116.
88. Zhou J, Liu M, Sun H, et al. Hepatoma-intrinsic CCRK inhibition diminishes myeloid-derived suppressor cell immunosuppression and enhances immune-checkpoint blockade efficacy. *Gut.* 2018 May;67(5):931-944. doi: 10.1136/gutjnl-2017-314032. Epub 2017 Sep 22. PMID: 28939663; PMCID: PMC5961939.
89. Ren F, Ding X, Zheng M, et al. AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. *Chem Sci.* 2023;14(6):1443–52.
90. Feng H, Yu Z, Tian Y, et al. A CCRK-EZH2 epigenetic circuitry drives hepatocarcinogenesis and associates with tumor recurrence and poor survival of patients. *J Hepatol* [Internet]. 2015;62(5):1100–11. Available from: <http://dx.doi.org/10.1016/j.jhep.2014.11.040>
91. Bazan-Socha S, Buregwa-Czuma S, Jakiela B, Zareba L, Zawlik I, Myszka A, Soja J, Okon K, Zarychta J, Kozlik P, Dziedzina S, Padjas A, Wojcik K, Kepski M, Bazan JG. Reticular Basement Membrane Thickness Is Associated with Growth- and Fibrosis-Promoting Airway Transcriptome Profile-Study in Asthma Patients. *Int J Mol Sci.* 2021 Jan 20;22(3):998. doi: 10.3390/ijms22030998. PMID: 33498209; PMCID: PMC7863966.
92. Mahadevappa R, Neves H, Yuen SM, Bai Y, McCrudden CM, Yuen HF, et al. The prognostic significance of Cdc6 and Cdt1 in breast cancer /692/4028/67/1347 /692/4028/67/1857 /13 /13/109 /38 /13/106 /38/39 /82/80 article. *Sci Rep.* 2017;7(1):1–11.
93. Karantzelis N, Petropoulos M, De Marco V, Egan DA, Fish A, Christodoulou E, et al. Small Molecule Inhibitor Targeting CDT1/Geminin Protein Complex Promotes DNA Damage and Cell Death in Cancer Cells. *Front Pharmacol.* 2022;13(April):1–16.
94. Ogawa M, Moriyama M, Midorikawa Y, Nakamura H, Shibata T, Kuroda K, et al. The significance of CDT1 expression in noncancerous and cancerous liver in cases with hepatocellular carcinoma. *J Clin Biochem Nutr.* 2023;73(3):234–48.
95. Jiang J, Zhang Y, Wang J, Yang X, Ren X, Huang H, et al. Identification of CDT1 as a prognostic marker in human lung adenocarcinoma using bioinformatics approaches. *PeerJ.* 2023;11:1–22.

96. Sarafidis M, Lambrou GI, Zoumpourlis V, Koutsouris D. An Integrated Bioinformatics Analysis towards the Identification of Diagnostic, Prognostic, and Predictive Key Biomarkers for Urinary Bladder Cancer. *Cancers (Basel)*. 2022 Jul 10;14(14):3358. doi: 10.3390/cancers14143358. PMID: 35884419; PMCID: PMC9319344.
97. Li W, Huang K, Wen F, Cui G, Guo H, He Z, et al. Intermittent hypoxia-induced downregulation of microRNA-320b promotes lung cancer tumorigenesis by increasing CDT1 via USP37. *Mol Ther Nucleic Acids [Internet]*. 2021;24(1):528–41. Available from: <https://doi.org/10.1016/j.omtn.2020.12.023>
98. Wang C, Che J, Jiang Y, Chen P, Bao G, Li C. CDT1 facilitates metastasis in prostate cancer and correlates with cell cycle regulation. *Cancer Biomark*. 2022;34(3):459-469. doi: 10.3233/CBM-210389. PMID: 35253732.
99. Hernández-Pérez S, Cabrera E, Amoedo H, Rodríguez-Acebes S, Koundrioukoff S, Debatisse M, et al. USP37 deubiquitinates Cdt1 and contributes to regulate DNA replication. *Mol Oncol*. 2016;10(8):1196–206.
100. Karakaidos P, Taraviras S, Vassiliou LV, Zacharatos P, Kastrinakis NG, Kougiou D, Kouloukoussa M, Nishitani H, Papavassiliou AG, Lygerou Z, Gorgoulis VG. Overexpression of the replication licensing regulators hCdt1 and hCdc6 characterizes a subset of non-small-cell lung carcinomas: synergistic effect with mutant p53 on tumor growth and chromosomal instability--evidence of E2F-1 transcriptional control over hCdt1. *Am J Pathol*. 2004 Oct;165(4):1351-65. doi: 10.1016/S0002-9440(10)63393-7. PMID: 15466399; PMCID: PMC1618634.
101. Vaziri C, Saxena S, Jeon Y, Lee C, Murata K, Machida Y, Wagle N, Hwang DS, Dutta A. A p53-dependent checkpoint pathway prevents rereplication. *Mol Cell*. 2003 Apr;11(4):997-1008. doi: 10.1016/s1097-2765(03)00099-6. Erratum in: *Mol Cell*. 2003 May;11(5):1415. PMID: 12718885.
102. Kiran S, Dar A, Singh SK, Lee KY, Dutta A. The Deubiquitinase USP46 Is Essential for Proliferation and Tumor Growth of HPV-Transformed Cancers. *Mol Cell [Internet]*. 2018;72(5):823-835.e5. Available from: <https://doi.org/10.1016/j.molcel.2018.09.019>
103. Karavias D, Maroulis I, Papadaki H, Gogos C, Kakkos S, Karavias D, Bravou V. Overexpression of CDT1 Is a Predictor of Poor Survival in Patients with

- Hepatocellular Carcinoma. *J Gastrointest Surg.* 2016 Mar;20(3):568-79. doi: 10.1007/s11605-015-2960-7. Epub 2015 Sep 25. PMID: 26408331.
- 104.** Alzahayqa M, Jamous A, Khatib AAH, Salah Z. TET1 Isoforms Have Distinct Expression Pattern, Localization and Regulation in Breast Cancer. *Front Oncol.* 2022;12(May):1–10.
- 105.** Stolz P, Mantero AS, Tvardovskiy A, Ugur E, Wange LE, Mulholland CB, et al. TET1 regulates gene expression and repression of endogenous retroviruses independent of DNA demethylation. *Nucleic Acids Res.* 2022;50(15):8491–511.
- 106.** Yang Q, Dang H, Liu J, Wang X, Wang J, Lan X, et al. Hypoxia switches TET1 from being tumor-suppressive to oncogenic. *Oncogene.* 2023;42(20):1634–48.
- 107.** Schagdarsurengin U, Luo C, Slanina H, Sheridan D, Füssel S, Böğürçü-Seidel N, Gattenloehner S, Baretton GB, Hofbauer LC, Wagenlehner F, Dansranjav T. Tracing TET1 expression in prostate cancer: discovery of malignant cells with a distinct oncogenic signature. *Clin Epigenetics.* 2021 Nov 29;13(1):211. doi: 10.1186/s13148-021-01201-7. PMID: 34844636; PMCID: PMC8630881.
- 108.** Chen LY, Huang RL, Chan MW, Yan PS, Huang TS, Wu RC, Suryo Rahmanto Y, Su PH, Weng YC, Chou JL, Chao TK, Wang YC, Shih IM, Lai HC. TET1 reprograms the epithelial ovarian cancer epigenome and reveals casein kinase 2 $\alpha$  as a therapeutic target. *J Pathol.* 2019 Jul;248(3):363-376. doi: 10.1002/path.5266. Epub 2019 Apr 23. PMID: 30883733; PMCID: PMC6579655.
- 109.** Zhong K, Wang X, Zhang H, Chen N, Mai Y, Dai S, Yang L, Chen D, Zhong W. BIRC6 Modulates the Protein Stability of Axin to Regulate the Growth, Stemness, and Resistance of Renal Cancer Cells via the  $\beta$ -Catenin Pathway. *ACS Omega.* 2024 Feb 7;9(7):7782-7792. doi: 10.1021/acsomega.3c07265. PMID: 38405482; PMCID: PMC10882609.
- 110.** Wang L, Chen YJ, Hou J, Wang YY, Tang WQ, Shen XZ, Tu RQ. Expression and clinical significance of BIRC6 in human epithelial ovarian cancer. *Tumour Biol.* 2014 May;35(5):4891-6. doi: 10.1007/s13277-014-1641-6. Epub 2014 Jan 23. PMID: 24453032.
- 111.** Zhou L, Wang B, Zhang Y, Yao K, Liu B. Silencing circ BIRC6 inhibits the proliferation, invasion, migration and epithelial mesenchymal transition of bladder cancer cells by targeting the miR 495 3p/XBP1 signaling axis. *Mol Med Rep.* 2021



- Nov;24(5):811. doi: 10.3892/mmr.2021.12451. Epub 2021 Sep 20. PMID: 34542161; PMCID: PMC8477182.
- 112.** Zhong K, Chen D, Wu Z, Wang X, Pan B, Chen N, Zhong W. [Effect of small interfering RNA-mediated BIRC6 silencing on apoptosis and autophagy of renal cancer 786-O cells]. *Nan Fang Yi Ke Da Xue Xue Bao*. 2020 Nov 30;40(11):1651-1655. Chinese. doi: 10.12122/j.issn.1673-4254.2020.11.18. PMID: 33243730; PMCID: PMC7704388.
- 113.** Hu T, Weng S, Tang W, Xue R, Chen S, Cai G, Cai Y, Shen X, Zhang S, Dong L. Overexpression of BIRC6 Is a Predictor of Prognosis for Colorectal Cancer. *PLoS One*. 2015 May 1;10(5):e0125281. doi: 10.1371/journal.pone.0125281. PMID: 25933218; PMCID: PMC4416929.
- 114.** Van Houdt WJ, Emmink BL, Pham TV, Piersma SR, Verheem A, Vries RG, Fratantoni SA, Pronk A, Clevers H, Borel Rinkes IH, Jimenez CR, Kranenburg O. Comparative proteomics of colon cancer stem cells and differentiated tumor cells identifies BIRC6 as a potential therapeutic target. *Mol Cell Proteomics*. 2011 Dec;10(12):M111.011353. doi: 10.1074/mcp.M111.011353. Epub 2011 Jul 25. PMID: 21788403; PMCID: PMC3237084.
- 115.** Dong X, Lin D, Low C, Vucic EA, English JC, Yee J, Murray N, Lam WL, Ling V, Lam S, Gout PW, Wang Y. Elevated expression of BIRC6 protein in non-small-cell lung cancers is associated with cancer recurrence and chemoresistance. *J Thorac Oncol*. 2013 Feb;8(2):161-70. doi: 10.1097/JTO.0b013e31827d5237. PMID: 23287853.
- 116.** Zhuang W, Zhang C, Hao F, Sun X. Baculoviral IAP Repeat Containing 6 (BIRC6) Is a Predictor of Prognosis in Prostate Cancer. *Med Sci Monit*. 2018 Feb 10;24:839-845. doi: 10.12659/MSM.904052. PMID: 29429983; PMCID: PMC5816567.
- 117.** Li Y, Tan Y, Wen L, Xing Z, Wang C, Zhang L, Wu K, Sun H, Li Y, Lei Q, Wu S. Overexpression of BIRC6 driven by EGF-JNK-HECTD1 signaling is a potential therapeutic target for triple-negative breast cancer. *Mol Ther Nucleic Acids*. 2021 Sep 28;26:798-812. doi: 10.1016/j.omtn.2021.09.011. PMID: 34729249; PMCID: PMC8526501.
- 118.** Eskandari F, Razavian A, Zare R, Ejlali S, Razmahang A, Zanjani M, Aghili SS, Mahdiyar MA, Mofidi H, Abbasi K, Badkoobeh A, Shamloo N, Hakim LK, Hussain A, Tebyaniyan H. Evaluation of BIRC6 Expression in Oral Squamous Cell Carcinoma,

- Epithelial Dysplasia, Lichen Planus with and without Dysplasia, and Hyperkeratosis. *Diagnostics*. 2023 Nov 29;13(23):3560. doi: 10.3390/diagnostics13233560.
- 119.** Tang W, Xue R, Weng S, Wu J, Fang Y, Wang Y, Ji L, Hu T, Liu T, Huang X, Chen S, Shen X, Zhang S, Dong L. BIRC6 promotes hepatocellular carcinogenesis: interaction of BIRC6 with p53 facilitating p53 degradation. *Int J Cancer*. 2015 Mar 15;136(6):E475-87. doi: 10.1002/ijc.29194. Epub 2014 Sep 18. PMID: 25196217.
- 120.** Zhao C, Ma B, Yang ZY, Li O, Liu SL, Pan LJ, Gong W, Dong P, Shu YJ. Inhibition of XPO1 impairs cholangiocarcinoma cell proliferation by triggering p53 intranuclear accumulation. *Cancer Med*. 2023 Mar;12(5):5751-5763. doi: 10.1002/cam4.5322. Epub 2022 Oct 5. PMID: 36200270; PMCID: PMC10028126.
- 121.** Yue H, Wu K, Liu K, Gou L, Huang A, Tang H. LINC02154 promotes the proliferation and metastasis of hepatocellular carcinoma by enhancing SPC24 promoter activity and activating the PI3K-AKT signaling pathway. *Cell Oncol (Dordr)*. 2022 Jun;45(3):447-462. doi: 10.1007/s13402-022-00676-7. Epub 2022 May 11. PMID: 35543858.
- 122.** Zhou J, Pei Y, Chen G, Cao C, Liu J, Ding C, Wang D, Sun L, Xu P, Niu G. SPC24 regulates breast cancer progression by PI3K/AKT signaling. *Gene*. 2018 Oct 30;675:272-277. doi: 10.1016/j.gene.2018.07.017. Epub 2018 Jul 6. PMID: 30180968.
- 123.** Chen S, Wang X, Zheng S, Li H, Qin S, Liu J, Jia W, Shao M, Tan Y, Liang H, Song W, Lu S, Liu C, Yang X. Increased SPC24 in prostatic diseases and diagnostic value of SPC24 and its interacting partners in prostate cancer. *Exp Ther Med*. 2021 Sep;22(3):923. doi: 10.3892/etm.2021.10355. Epub 2021 Jun 30. PMID: 34306192; PMCID: PMC8281004.
- 124.** Zhou J, Yu Y, Pei Y, Cao C, Ding C, Wang D, Sun L, Niu G. A potential prognostic biomarker SPC24 promotes tumorigenesis and metastasis in lung cancer. *Oncotarget*. 2017 Jul 4;8(39):65469-65480. doi: 10.18632/oncotarget.18971. PMID: 29029446; PMCID: PMC5630346.
- 125.** Chen J, Luo J, He J, Jiang X, Jiang N, Yang C, Zhong S. Cell cycle-related gene SPC24: A novel potential diagnostic and prognostic biomarker for laryngeal squamous cell cancer. *Biomed Res Int*. 2023 Jan 21;2023:1733100. doi: 10.1155/2023/1733100. PMID: 36718148; PMCID: PMC9884166.

- 126.** Fu S, Gong B, Ding Y, Zhuang C, Chen Q, Wang S, Li Z, Ma M, Liu Y, Zhang Z, Sun T. An In Silico Investigation of SPC24 as a Putative Biomarker of Kidney Renal Clear Cell Carcinoma and Kidney Renal Papillary Cell Carcinoma for Predicting Prognosis and/or Immune Infiltration. *Comb Chem High Throughput Screen.* 2022;25(13):2278-2294. doi: 10.2174/1386207325666220315105054. PMID: 35293292.
- 127.** Yin H, Meng T, Zhou L, Chen H, Song D. SPC24 is critical for anaplastic thyroid cancer progression. *Oncotarget.* 2017 Mar 28;8(13):21884-21891. doi: 10.18632/oncotarget.15670. PMID: 28423533; PMCID: PMC5400631.
- 128.** Zhang L, Wu J, Ling MT, Zhao L, Zhao KN. The role of the PI3K/Akt/mTOR signalling pathway in human cancers induced by infection with human papillomaviruses. *Mol Cancer.* 2015 Apr 17;14:87. doi: 10.1186/s12943-015-0361-x. PMID: 26022660; PMCID: PMC4498560.
- 129.** Tellier M. Structure, Activity, and Function of SETMAR Protein Lysine Methyltransferase. *Life (Basel).* 2021 Dec 4;11(12):1342. doi: 10.3390/life11121342. PMID: 34947873; PMCID: PMC8704517.
- 130.** Wray J, Williamson EA, Sheema S, Lee SH, Libby E, Willman CL, Nickoloff JA, Hromas R. Metnase mediates chromosome decatenation in acute leukemia cells. *Blood.* 2009 Aug 27;114(9):1852-8. doi: 10.1182/blood-2008-08-175760. Epub 2009 May 20. PMID: 19458360; PMCID: PMC2738570.
- 131.** Williamson EA, Farrington J, Martinez L, Ness S, O'Rourke J, Lee SH, Nickoloff J, Hromas R. Expression levels of the human DNA repair protein metnase influence lentiviral genomic integration. *Biochimie.* 2008 Sep;90(9):1422-6. doi: 10.1016/j.biochi.2008.05.010. Epub 2008 May 23. PMID: 18549821; PMCID: PMC2556226.
- 132.** Lié O, Virolle T, Gabut M, Pasquier C, Zemmoura I, Augé-Gouillou C. SETMAR Shorter Isoform: A New Prognostic Factor in Glioblastoma. *Front Oncol.* 2022 Jan 3;11:638397. doi: 10.3389/fonc.2021.638397. PMID: 35047379; PMCID: PMC8761672.
- 133.** Fnu S, Williamson EA, De Haro LP, Brenneman M, Wray J, Shaheen M, Radhakrishnan K, Lee SH, Nickoloff JA, Hromas R. Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. *Proc Natl Acad Sci U S A.*

- 2011 Jan 11;108(2):540-5. doi: 10.1073/pnas.1013571108. Epub 2010 Dec 27. PMID: 21187428; PMCID: PMC3021059.
- 134.** Geris JM, Amirian ES, Marquez-Do DA, Guillaud M, Dillon LM, Follen M, Scheurer ME. Polymorphisms in the Nonhomologous End-joining DNA Repair Pathway are Associated with HPV Integration in Cervical Dysplasia. *Cancer Prev Res (Phila)*. 2023 Aug 1;16(8):461-469. doi: 10.1158/1940-6207.CAPR-23-0051. PMID: 37217238; PMCID: PMC10524768.
- 135.** Zeng B, Wan R, Chang K, Li J, Zhang X, Shi G, Ye D, Xu F. Lysine methyltransferase 5C increases the proliferation and metastatic abilities of clear cell renal cell carcinoma via aerobic glycolysis. *Int J Oncol*. 2024 Apr;64(4):45. doi: 10.3892/ijo.2024.5633. Epub 2024 Mar 1. PMID: 38426605; PMCID: PMC10919755.
- 136.** Nakshatri H, Appaiah HN, Anjanappa M, Gilley D, Tanaka H, Badve S, Crooks PA, Mathews W, Sweeney C, Bhat-Nakshatri P. NF- $\kappa$ B-dependent and -independent epigenetic modulation using the novel anti-cancer agent DMAPT. *Cell Death Dis*. 2015 Jan 22;6(1):e1608. doi: 10.1038/cddis.2014.569. PMID: 25611383; PMCID: PMC4669767.
- 137.** Yokoyama Y, Matsumoto A, Hieda M, Shinchi Y, Ogihara E, Hamada M, Nishioka Y, Kimura H, Yoshidome K, Tsujimoto M, Matsuura N. Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. *Breast Cancer Res*. 2014 Jun 22;16(3):R66. doi: 10.1186/bcr3681. PMID: 24953066; PMCID: PMC4229880.
- 138.** Rakha EA, Alsaleem M, ElSharawy KA, Toss MS, Raafat S, Mihai R, Minhas FA, Green AR, Rajpoot NM, Dalton LW, Mongan NP. Visual histological assessment of morphological features reflects the underlying molecular profile in invasive breast cancer: a morphomolecular study. *Histopathology*. 2020 Oct;77(4):631-645. doi: 10.1111/his.14199. Epub 2020 Sep 1. PMID: 32618014.
- 139.** Chintala S, Quist KM, Gonzalez-DeWhitt PA, Katzenellenbogen RA. High expression of NFX1-123 in HPV positive head and neck squamous cell carcinomas. *Head Neck*. 2022 Jan;44(1):177-188. doi: 10.1002/hed.26906. Epub 2021 Oct 25. PMID: 34693597; PMCID: PMC8688290.
- 140.** Vliet-Gregg PA, Hamilton JR, Katzenellenbogen RA. NFX1-123 and human papillomavirus 16E6 increase Notch expression in keratinocytes. *J Virol*. 2013

- Dec;87(24):13741-50. doi: 10.1128/JVI.02582-13. Epub 2013 Oct 9. PMID: 24109236; PMCID: PMC3838236.
- 141.** Yeo-Teh NSL, Ito Y, Jha S. High-Risk Human Papillomaviral Oncogenes E6 and E7 Target Key Cellular Pathways to Achieve Oncogenesis. *Int J Mol Sci.* 2018 Jun 8;19(6):1706. doi: 10.3390/ijms19061706. PMID: 29890655; PMCID: PMC6032416.
- 142.** He Y, Sun MM, Zhang GG, Yang J, Chen KS, Xu WW, Li B. Targeting PI3K/Akt signal transduction for cancer therapy. *Signal Transduct Target Ther.* 2021 Dec 16;6(1):425. doi: 10.1038/s41392-021-00828-5. PMID: 34916492; PMCID: PMC8677728.
- 143.** Thorpe LM, Yuzugullu H, Zhao JJ. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat Rev Cancer.* 2015 Jan;15(1):7-24. doi: 10.1038/nrc3860. PMID: 25533673; PMCID: PMC4384662.
- 144.** Vasudevan KM, Barbie DA, Davies MA, Rabinovsky R, McNear CJ, Kim JJ, Hennessy BT, Tseng H, Pochanard P, Kim SY, Dunn IF, Schinzel AC, Sandy P, Hoersch S, Sheng Q, Gupta PB, Boehm JS, Reiling JH, Silver S, Lu Y, Stemke-Hale K, Dutta B, Joy C, Sahin AA, Gonzalez-Angulo AM, Lluch A, Rameh LE, Jacks T, Root DE, Lander ES, Mills GB, Hahn WC, Sellers WR, Garraway LA. AKT-independent signaling downstream of oncogenic PIK3CA mutations in human cancer. *Cancer Cell.* 2009 Jul 7;16(1):21-32. doi: 10.1016/j.ccr.2009.04.012. PMID: 19573809; PMCID: PMC2752826.
- 145.** Göttgens EL, Ostheimer C, Span PN, Bussink J, Hammond EM. HPV, hypoxia and radiation response in head and neck cancer. *Br J Radiol.* 2019 Jan;92(1093):20180047. doi: 10.1259/bjr.20180047. Epub 2018 Mar 14. PMID: 29493265; PMCID: PMC6435089.
- 146.** Nicolás I, Saco A, Barnadas E, Marimon L, Rakislova N, Fusté P, Rovirosa A, Gaba L, Buñesch L, Gil-Ibañez B, Pahisa J, Díaz-Feijoo B, Torne A, Ordi J, Del Pino M. Prognostic implications of genotyping and p16 immunostaining in HPV-positive tumors of the uterine cervix. *Mod Pathol.* 2020 Jan;33(1):128-137. doi: 10.1038/s41379-019-0360-3. Epub 2019 Sep 6. PMID: 31492932.
- 147.** Burbure N, Handorf E, Ridge JA, Bauman J, Liu JC, Giri A, Galloway TJ. Prognostic significance of human papillomavirus status and treatment modality in hypopharyngeal cancer. *Head Neck.* 2021 Oct;43(10):3042-3052. doi: 10.1002/hed.26793. Epub 2021 Jun 24. PMID: 34165223; PMCID: PMC9631108.

- 148.** Tong F, Geng J, Yan B, Lou H, Chen X, Duan C, He J, Zhang S, Xie H, Li H, Yuan D, Zhang F, Meng H, Wei L. Prevalence and Prognostic Significance of HPV in Laryngeal Squamous Cell Carcinoma in Northeast China. *Cell Physiol Biochem*. 2018;49(1):206-216. doi: 10.1159/000492858. Epub 2018 Aug 22. PMID: 30134232.
- 149.** Axelsson, L., Nyman, J., Haugen-Cange, H. et al. Prognostic factors for head and neck cancer of unknown primary including the impact of human papilloma virus infection. *J of Otolaryngol - Head & Neck Surg* 46, 45 (2017). <https://doi.org/10.1186/s40463-017-0223-1>
- 150.** Benson E, Li R, Eisele D, Fakhry C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. *Oral Oncol*. 2014 Jun;50(6):565-74. doi: 10.1016/j.oraloncology.2013.09.008. Epub 2013 Oct 14. PMID: 24134947; PMCID: PMC4391706.
- 151.** Šimić I, Božinović K, Milutin Gašperov N, Kordić M, Pešut E, Manojlović L, Grce M, Dediol E, Sabol I. Head and Neck Cancer Patients' Survival According to HPV Status, miRNA Profiling, and Tumour Features-A Cohort Study. *Int J Mol Sci*. 2023 Feb 7;24(4):3344. doi: 10.3390/ijms24043344. PMID: 36834756; PMCID: PMC9959828.
- 152.** Xu X, Feng T, Li D, Lou H, Lan H. Prevalent distribution and survival outcome of HPV infection in patients with early-stage cervical cancer in Hangzhou, China. *BMC Infect Dis*. 2022 Dec 15;22(1):941. doi: 10.1186/s12879-022-07888-0. PMID: 36522614; PMCID: PMC9753312.

## 13. ANEXO

### 13.1. ANEXO 1

**Tabla 1 Anexo.** Los 200 genes diferenciales sobreexpresados con menor FDR.

\*Fold Change se refiere al grupo HNSCC-HPV+

ID	fold change (FC)*	pvalor	padj (FDR)
PIAS2	0,882	1,78E-11	3,64E-07
DNAJC27	0,748	1,11E-11	3,64E-07
KIF27	1,012	2,82E-11	3,84E-07
ZNF169	0,866	7,21E-11	7,36E-07
MECOM	1,739	1,46E-10	7,95E-07
ANKRD36B	1,131	1,56E-10	7,95E-07
PRDM15	0,894	1,41E-10	7,95E-07
MSL2	0,781	1,50E-10	7,95E-07
KATNAL2	1,609	2,30E-10	8,90E-07
TBCCD1	0,852	2,62E-10	8,90E-07
SEPSECS	0,737	2,61E-10	8,90E-07
SUZ12	0,709	2,00E-10	8,90E-07
ZNF510	0,707	2,85E-10	8,96E-07
RPS10P7	1,232	3,74E-10	9,55E-07
FIGNL1	0,980	3,36E-10	9,55E-07
LIN54	0,771	3,67E-10	9,55E-07
LIN9	0,894	4,25E-10	1,01E-06
MTMR1	0,801	4,45E-10	1,01E-06
PAXBP1-AS1	1,051	4,81E-10	1,03E-06
TEX21P	1,203	5,98E-10	1,09E-06
ZNF37BP	1,200	6,54E-10	1,09E-06
KLHL24	1,120	6,95E-10	1,09E-06
ATAD5	1,014	6,30E-10	1,09E-06
ACAP2	0,981	6,61E-10	1,09E-06
ATAD2B	0,729	6,87E-10	1,09E-06
SENP1	0,704	6,82E-10	1,09E-06
MYNN	0,671	7,20E-10	1,09E-06
FAM161A	1,226	9,09E-10	1,12E-06
BRCA1P1	1,020	8,68E-10	1,12E-06
C9orf72	1,006	9,01E-10	1,12E-06
PHC3	0,906	8,61E-10	1,12E-06
CEP78	0,804	8,20E-10	1,12E-06
GPATCH1	0,665	8,27E-10	1,12E-06
JPX	1,185	1,08E-09	1,13E-06
MGA	0,887	1,01E-09	1,13E-06
MORC4	0,873	1,04E-09	1,13E-06
CCDC121	0,816	1,02E-09	1,13E-06
ZNF782	0,787	9,95E-10	1,13E-06
ZNF507	0,766	1,07E-09	1,13E-06
MSH6	0,842	1,13E-09	1,15E-06

KNTC1	0,945	1,29E-09	1,22E-06
ZNF33A	0,716	1,25E-09	1,22E-06
SCYL3	0,686	1,26E-09	1,22E-06
CYCSP10	1,217	1,52E-09	1,41E-06
WDR19	0,845	1,65E-09	1,43E-06
TRMT2B	0,835	1,63E-09	1,43E-06
SMC5	0,665	1,73E-09	1,43E-06
ANGEL2	0,652	1,79E-09	1,43E-06
KDM2B	0,593	1,77E-09	1,43E-06
RUNDC1	0,526	1,69E-09	1,43E-06
ZNF2	0,630	1,89E-09	1,48E-06
FAM228B	0,815	2,12E-09	1,63E-06
TXNDC16	1,072	2,30E-09	1,71E-06
ZFP62	0,801	2,29E-09	1,71E-06
CABLES2	0,876	2,56E-09	1,83E-06
MIR5581	0,867	2,60E-09	1,83E-06
WASHC2A	0,579	2,56E-09	1,83E-06
EFCAB13	1,315	2,80E-09	1,93E-06
HYKK	1,060	2,86E-09	1,93E-06
CHD6	0,901	2,88E-09	1,93E-06
RMI1	0,854	2,97E-09	1,96E-06
POU2AF3	2,041	3,19E-09	1,98E-06
CFAP70	1,242	3,34E-09	1,98E-06
ZNF337-AS1	1,102	3,27E-09	1,98E-06
ANKRD26	0,873	3,35E-09	1,98E-06
TTC41P	0,831	3,25E-09	1,98E-06
POGZ	0,697	3,10E-09	1,98E-06
DENND4A	0,691	3,26E-09	1,98E-06
CEP152	0,930	3,44E-09	2,01E-06
CBFA2T2	0,851	3,77E-09	2,14E-06
ZNF641	0,803	3,74E-09	2,14E-06
BRCA2	1,085	3,97E-09	2,19E-06
KDM3A	0,740	3,96E-09	2,19E-06
EMC3-AS1	1,423	4,09E-09	2,23E-06
TADA1	0,668	4,36E-09	2,31E-06
USP30	0,516	4,32E-09	2,31E-06
ENTPD1-AS1	0,582	4,44E-09	2,32E-06
IFT172	0,840	4,53E-09	2,34E-06
EHHADH	0,838	4,58E-09	2,34E-06
DIS3L	0,854	4,73E-09	2,35E-06
PAXBP1	0,756	4,77E-09	2,35E-06
LRRC37B	0,619	4,68E-09	2,35E-06
C18orf25	0,603	4,91E-09	2,39E-06
SCAI	0,812	4,99E-09	2,40E-06
CEP19	1,137	5,10E-09	2,41E-06
DNA2	0,926	5,14E-09	2,41E-06
LNCOC1	1,594	5,53E-09	2,43E-06



HCN3	1,083	5,41E-09	2,43E-06
CENPJ	1,003	5,45E-09	2,43E-06
CARF	0,953	5,48E-09	2,43E-06
ZNF397	0,781	5,39E-09	2,43E-06
KANSL3	0,581	5,25E-09	2,43E-06
COLCA1	2,995	5,65E-09	2,45E-06
TARBP1	0,808	6,06E-09	2,59E-06
CEP250	0,747	6,12E-09	2,59E-06
ago-04	0,701	6,16E-09	2,59E-06
CEP135	0,888	6,25E-09	2,61E-06
PASK	0,950	6,52E-09	2,66E-06
ALDH6A1	0,934	6,47E-09	2,66E-06
ENTPD4	0,689	6,57E-09	2,66E-06
ARHGAP26	1,252	6,83E-09	2,73E-06
ACP6	1,134	7,00E-09	2,75E-06
EIF4A2	0,893	6,97E-09	2,75E-06
TRIM52	0,865	7,39E-09	2,83E-06
ACTMAP	0,666	7,29E-09	2,83E-06
POU2F1	0,647	7,41E-09	2,83E-06
NEMP1	0,960	7,54E-09	2,85E-06
PABIR1	0,427	8,02E-09	3,00E-06
FANCI	0,901	8,21E-09	3,05E-06
RN7SKP80	1,195	8,68E-09	3,16E-06
THAP9	0,718	8,63E-09	3,16E-06
C2orf68	0,621	8,86E-09	3,20E-06
PHF8	0,942	9,10E-09	3,26E-06
STRADA	0,984	9,23E-09	3,28E-06
TEFM	0,578	9,35E-09	3,29E-06
TICRR	1,057	9,72E-09	3,31E-06
MPHOSPH9	0,883	9,56E-09	3,31E-06
TIMELESS	0,851	9,59E-09	3,31E-06
PARP16	0,757	9,72E-09	3,31E-06
KIF15	1,039	9,98E-09	3,34E-06
SLF2	0,685	9,99E-09	3,34E-06
C5	1,349	1,04E-08	3,43E-06
PUM2	0,566	1,04E-08	3,43E-06
SENP5	0,658	1,07E-08	3,50E-06
GLUD1P3	0,641	1,08E-08	3,51E-06
RPL21P118	1,160	1,15E-08	3,59E-06
SRRM2-AS1	1,071	1,14E-08	3,59E-06
ZNF41	0,871	1,17E-08	3,59E-06
C1orf112	0,864	1,13E-08	3,59E-06
ZNF639	0,704	1,17E-08	3,59E-06
ZNF674	0,691	1,13E-08	3,59E-06
PPM1D	0,683	1,16E-08	3,59E-06
ZNF620	1,339	1,20E-08	3,59E-06
PXYLP1	1,310	1,19E-08	3,59E-06

TTF2	0,881	1,20E-08	3,59E-06
OGT	0,723	1,21E-08	3,59E-06
BRIP1	1,081	1,26E-08	3,67E-06
ESCO2	1,017	1,25E-08	3,67E-06
DDX17	0,519	1,24E-08	3,67E-06
ESCO1	0,617	1,28E-08	3,71E-06
ZNF37A	0,822	1,30E-08	3,72E-06
HELLS	1,065	1,33E-08	3,73E-06
RBBP4	0,798	1,32E-08	3,73E-06
MDM4	0,761	1,33E-08	3,73E-06
KDM4C	0,654	1,32E-08	3,73E-06
TMPO	0,941	1,37E-08	3,74E-06
LAMTOR5-AS1	0,939	1,35E-08	3,74E-06
FBXO48	0,599	1,36E-08	3,74E-06
WRN	0,725	1,44E-08	3,91E-06
RAD51D	0,725	1,46E-08	3,96E-06
RFFL	0,844	1,52E-08	4,07E-06
ZNF33B	0,782	1,53E-08	4,07E-06
KANSL2	0,578	1,53E-08	4,07E-06
SMARCA2	0,919	1,55E-08	4,08E-06
SLC25A36	0,781	1,56E-08	4,08E-06
PPP1R3D	0,818	1,59E-08	4,14E-06
BARD1	0,917	1,66E-08	4,30E-06
SPDYE18	1,337	1,82E-08	4,59E-06
FOXRED2	1,225	1,81E-08	4,59E-06
RNF138	0,738	1,80E-08	4,59E-06
PHF12	0,656	1,81E-08	4,59E-06
MAN2A2	1,043	1,87E-08	4,60E-06
CIT	0,969	1,84E-08	4,60E-06
ZNF654	0,656	1,87E-08	4,60E-06
TBL1XR1	0,645	1,86E-08	4,60E-06
KIAA1328	0,738	1,93E-08	4,71E-06
ACTL6A	0,800	1,95E-08	4,73E-06
LONRF1	0,886	2,10E-08	5,04E-06
ALMS1	0,859	2,10E-08	5,04E-06
ZNF24	0,572	2,12E-08	5,04E-06
RIC8B	0,589	2,14E-08	5,06E-06
KCTD13-DT	1,234	2,24E-08	5,12E-06
PAN2	0,952	2,26E-08	5,12E-06
MTUS1	0,871	2,24E-08	5,12E-06
GMCL1	0,724	2,23E-08	5,12E-06
ZWILCH	0,664	2,24E-08	5,12E-06
RBBP9	0,657	2,18E-08	5,12E-06
JMY	0,620	2,24E-08	5,12E-06
PARP1	0,635	2,29E-08	5,16E-06
STX18-AS1	0,794	2,32E-08	5,20E-06
CCDC171	0,955	2,35E-08	5,25E-06

TAF1	0,845	2,40E-08	5,33E-06
CCDC163	0,859	2,44E-08	5,38E-06
FECH	0,653	2,61E-08	5,71E-06
CGGBP1	0,583	2,61E-08	5,71E-06
EPM2A-DT	0,925	2,68E-08	5,76E-06
F8	0,923	2,67E-08	5,76E-06
PRIM1	0,910	2,70E-08	5,76E-06
ZNF189	0,610	2,69E-08	5,76E-06
EPC1	0,514	2,71E-08	5,77E-06
CCP110	0,723	2,76E-08	5,84E-06
TSSK4	1,096	2,81E-08	5,91E-06
ACAP2-IT1	1,304	3,07E-08	5,97E-06
SRRM5	1,296	2,91E-08	5,97E-06
CHPT1	1,238	2,99E-08	5,97E-06
DZANK1	0,960	3,09E-08	5,97E-06
MIR3153	0,935	2,96E-08	5,97E-06
GABPB2	0,850	3,04E-08	5,97E-06
SLF1	0,815	3,08E-08	5,97E-06

Para una visualización de los 5203 genes diferenciales sobreexpresados (FDR<0,01) acceden a este código QR:



**Tabla 2 Anexo.** Los 200 genes diferenciales infraexpresados con menor FDR.

\*Fold Change se refiere al grupo HNSCC-HPV+

ID	Fold Change (FC)*	pvalor	Padj (FDR)
MTCO1P21	-1,005	1,666E-09	1,433E-06
LINC01404	-2,192	2,120E-08	5,042E-06
MEP1AP1	-0,888	6,688E-08	9,194E-06
GPC3-AS1	-0,669	8,943E-08	1,084E-05
ATP6V0E1P3	-0,412	1,063E-07	1,179E-05
PANX3	-1,618	1,183E-07	1,278E-05
MTND4P11	-1,319	1,218E-07	1,292E-05
MRPL57P7	-0,468	1,255E-07	1,308E-05
LINC01654	-1,097	1,909E-07	1,657E-05
PLCE1-AS2	-1,897	2,395E-07	1,898E-05
RNA5SP103	-0,293	2,442E-07	1,925E-05
MTATP6P17	-0,591	3,113E-07	2,210E-05
KRT14	-2,193	3,133E-07	2,221E-05
RN7SL63P	-0,533	3,977E-07	2,560E-05
ELOCP33	-1,204	4,333E-07	2,709E-05
RNY3P12	-0,192	6,409E-07	3,522E-05
RNY1P15	-0,253	7,423E-07	3,851E-05
MT2A	-1,172	7,772E-07	3,972E-05

FOLR3	-2,191	8,405E-07	4,185E-05
KRT16P4	-1,706	8,434E-07	4,194E-05
MTCO2P27	-0,879	8,862E-07	4,323E-05
SBSN	-2,601	9,770E-07	4,598E-05
PAEP	-2,054	9,951E-07	4,628E-05
RPL7P39	-0,970	1,015E-06	4,668E-05
CRABP2	-1,384	1,087E-06	4,914E-05
RPS7P7	-1,158	1,181E-06	5,226E-05
SPRR2G	-3,091	1,208E-06	5,308E-05
FSTL3	-1,275	1,409E-06	5,918E-05
FLRT2-AS1	-1,549	1,415E-06	5,929E-05
LINC02454	-1,648	1,444E-06	6,004E-05
MIR5089	-0,156	1,479E-06	6,064E-05
RN7SKP85	-0,495	1,557E-06	6,251E-05
OR8S1	-1,224	1,652E-06	6,506E-05
WWC3-AS1	-0,989	1,653E-06	6,506E-05
NME2P2	-0,619	1,701E-06	6,632E-05
AADA4L4	-1,236	1,718E-06	6,666E-05
KRT16	-1,853	1,817E-06	6,920E-05
OOEP-AS1	-0,606	1,864E-06	7,071E-05
GAPDHP49	-1,259	2,218E-06	7,994E-05
TUSC2P2	-0,884	2,390E-06	8,426E-05
MYOM3	-1,987	2,755E-06	9,396E-05
RN7SL411P	-0,696	2,816E-06	9,518E-05
MIR1273H	-0,273	2,900E-06	9,701E-05
MTCO2P21	-0,876	2,915E-06	9,730E-05
RPS13P5	-0,499	3,145E-06	1,026E-04
IGLVIVOR22-1	-0,768	3,251E-06	1,048E-04
TUBB2A	-0,978	3,546E-06	1,114E-04
CFAP97D2	-1,641	3,561E-06	1,115E-04
LINC01615	-1,603	3,561E-06	1,115E-04
CDA	-1,854	3,781E-06	1,167E-04
RNU6-188P	-0,215	3,861E-06	1,185E-04
USP2	-1,486	3,940E-06	1,199E-04
SLC25A5P4	-0,941	4,066E-06	1,224E-04
FAM99A	-1,071	4,125E-06	1,236E-04
PHF5AP5	-0,666	4,438E-06	1,295E-04
MIR760	-0,177	4,679E-06	1,335E-04
SMTN	-0,739	4,883E-06	1,378E-04
PCDHB1-AS1	-1,307	4,897E-06	1,380E-04
TRBV8-1	-0,386	5,132E-06	1,427E-04
FGF14-AS1	-0,578	5,199E-06	1,441E-04
APOBEC4	-1,842	5,214E-06	1,442E-04
CXCR1	-1,631	5,328E-06	1,459E-04
ZFPM2-AS1	-1,985	5,402E-06	1,474E-04
PDLIM7	-0,644	5,439E-06	1,479E-04
MED15P1	-0,807	5,519E-06	1,499E-04

RN7SL259P	-0,423	5,543E-06	1,503E-04
KCTD10P1	-0,509	6,021E-06	1,584E-04
TUBA4A	-0,850	6,383E-06	1,646E-04
RN7SL667P	-0,339	6,827E-06	1,717E-04
CPA4	-2,212	7,010E-06	1,754E-04
AREG	-1,642	7,118E-06	1,770E-04
MMP1	-1,971	7,186E-06	1,781E-04
C1orf167-AS1	-0,664	7,260E-06	1,793E-04
WFDC12	-2,452	7,558E-06	1,844E-04
S100A10	-0,634	7,677E-06	1,861E-04
LGALS1	-0,831	8,180E-06	1,947E-04
KRT16P1	-1,533	8,283E-06	1,968E-04
DPRXP1	-0,751	8,286E-06	1,968E-04
KLK5	-2,810	8,350E-06	1,980E-04
FAM242C	-1,321	8,571E-06	2,018E-04
TRAV28	-0,444	8,735E-06	2,051E-04
POU5F1P2	-0,685	9,009E-06	2,093E-04
VN1R71P	-0,523	9,027E-06	2,096E-04
CNTNAP2-AS1	-1,027	9,131E-06	2,115E-04
CHCHD2P5	-0,586	9,465E-06	2,171E-04
TFAMP1	-0,955	9,594E-06	2,194E-04
BIRC8	-1,008	1,004E-05	2,270E-04
C16orf92	-1,320	1,017E-05	2,292E-04
RGS20	-1,155	1,029E-05	2,305E-04
GCNT1P1	-1,340	1,063E-05	2,361E-04
CLTB	-0,538	1,069E-05	2,370E-04
INHBA-AS1	-1,482	1,112E-05	2,443E-04
RNU6-650P	-0,171	1,274E-05	2,710E-04
KRT89P	-1,307	1,289E-05	2,728E-04
VN1R53P	-0,913	1,330E-05	2,786E-04
MIR3663	-0,139	1,393E-05	2,890E-04
EREG	-2,357	1,398E-05	2,897E-04
RPL36AP48	-0,733	1,415E-05	2,921E-04
OOSP1P2	-0,666	1,470E-05	3,011E-04
MTCO3P1	-0,677	1,498E-05	3,057E-04
GOLGA6A	-1,730	1,508E-05	3,071E-04
NIFKP6	-1,090	1,509E-05	3,072E-04
SH3BGRL3	-0,357	1,517E-05	3,082E-04
KRT16P3	-1,411	1,566E-05	3,149E-04
MIPEPP2	-0,684	1,634E-05	3,246E-04
ACSL6-AS1	-0,726	1,636E-05	3,249E-04
KRT75	-2,228	1,658E-05	3,278E-04
CT69	-1,634	1,712E-05	3,350E-04
EEF1A1P39	-0,523	1,711E-05	3,350E-04
RNA5SP212	-0,141	1,725E-05	3,369E-04
RNA5SP307	-0,116	1,744E-05	3,391E-04
MIR193BHG	-1,010	1,747E-05	3,393E-04

SLIT3-AS1	-1,773	1,818E-05	3,502E-04
TNFRSF12A	-0,898	1,861E-05	3,552E-04
KRTDAP	-2,905	1,900E-05	3,605E-04
MTND6P21	-0,871	1,900E-05	3,605E-04
KCNK7	-1,172	1,902E-05	3,607E-04
ENPP7P10	-1,176	1,928E-05	3,644E-04
RN7SL404P	-0,464	1,984E-05	3,727E-04
OR51F4P	-0,614	2,033E-05	3,796E-04
RCC2P2	-0,504	2,089E-05	3,874E-04
DMKN	-1,675	2,150E-05	3,962E-04
YPEL5P1	-0,369	2,154E-05	3,965E-04
HTRA1	-0,889	2,237E-05	4,082E-04
RN7SKP189	-0,363	2,357E-05	4,244E-04
KRT17P6	-1,161	2,370E-05	4,261E-04
OR6C73P	-0,373	2,476E-05	4,412E-04
SNORA70G	-0,376	2,510E-05	4,459E-04
OR2Z1	-1,028	2,723E-05	4,765E-04
RN7SL692P	-0,337	2,775E-05	4,832E-04
GUSBP4	-1,012	2,831E-05	4,901E-04
CST6	-1,483	2,873E-05	4,960E-04
NF1P1	-0,742	2,904E-05	5,000E-04
AKIRIN1P1	-0,372	2,934E-05	5,035E-04
LINC00682	-1,646	2,945E-05	5,049E-04
LINC00221	-1,869	2,974E-05	5,086E-04
KCNQ1DN	-1,171	2,975E-05	5,087E-04
RNU6-1284P	-0,122	2,987E-05	5,099E-04
UPP1	-0,875	3,014E-05	5,130E-04
RBP1	-0,954	3,026E-05	5,144E-04
MEP1AP4	-0,642	3,039E-05	5,154E-04
BDKRB2	-0,857	3,098E-05	5,245E-04
IGFBP6	-1,583	3,124E-05	5,270E-04
CYP2A7P2	-0,800	3,304E-05	5,503E-04
METRNL	-0,649	3,423E-05	5,652E-04
TAF11L14	-0,511	3,544E-05	5,804E-04
GPR101	-2,024	3,595E-05	5,875E-04
KRT17	-1,334	3,597E-05	5,875E-04
RN7SL810P	-0,338	3,643E-05	5,941E-04
KCNK18	-0,679	3,656E-05	5,953E-04
PODNL1	-0,931	3,696E-05	6,006E-04
RBMV2FP	-0,737	3,730E-05	6,051E-04
CD151	-0,486	3,824E-05	6,154E-04
RHOD	-0,957	3,836E-05	6,167E-04
MT2P1	-1,148	3,853E-05	6,184E-04
FYTTD1P1	-0,930	3,995E-05	6,356E-04
RN7SL492P	-0,292	4,342E-05	6,790E-04
LINC02690	-1,475	4,368E-05	6,817E-04
GOLGA7B	-1,519	4,558E-05	7,050E-04

KRT16P5	-1,706	4,818E-05	7,308E-04
MORC1	-1,676	4,824E-05	7,308E-04
SPDYE10	-1,284	4,857E-05	7,347E-04
MIR31HG	-1,505	4,876E-05	7,357E-04
MARK2P18	-1,268	4,877E-05	7,357E-04
THAP5P2	-1,226	4,870E-05	7,357E-04
PPP1R14B-AS1	-0,962	5,001E-05	7,504E-04
FAM90A2P	-1,342	5,091E-05	7,607E-04
KDM5C-IT1	-1,038	5,119E-05	7,633E-04
MIPEPP1	-1,488	5,173E-05	7,691E-04
KRT6B	-1,658	5,206E-05	7,729E-04
RN7SKP224	-0,447	5,211E-05	7,733E-04
GARIN6	-1,082	5,306E-05	7,838E-04
PICSAR	-1,963	5,339E-05	7,877E-04
FAM83A-AS1	-1,460	5,441E-05	7,989E-04
PIWIL3	-1,707	5,463E-05	8,008E-04
KRT17P1	-1,220	5,545E-05	8,103E-04
BDKRB1	-1,095	5,769E-05	8,355E-04
MYB-AS1	-0,621	5,899E-05	8,516E-04
PRR23C	-1,440	5,939E-05	8,553E-04
RPL34P17	-0,453	6,095E-05	8,728E-04
TEX28	-0,667	6,209E-05	8,864E-04
MMP3	-2,131	6,223E-05	8,873E-04
MTND2P31	-0,713	6,250E-05	8,899E-04
CCNB1IP1P1	-0,732	6,303E-05	8,938E-04
ARHGAP22-IT1	-0,470	6,391E-05	9,056E-04
CAV1	-1,089	6,411E-05	9,079E-04
IL9RP4	-0,773	6,431E-05	9,104E-04
LINC02132	-0,653	6,485E-05	9,161E-04
KRT16P6	-2,083	6,677E-05	9,377E-04
SHMT1P1	-1,175	6,816E-05	9,530E-04
KRT18P46	-0,728	6,859E-05	9,581E-04
PAQR5	-1,201	6,886E-05	9,611E-04
ANAPC15P2	-0,532	6,903E-05	9,626E-04
TAS2R41	-0,836	6,949E-05	9,670E-04
CEACAMP2	-0,592	6,969E-05	9,688E-04
GNG14	-0,715	7,017E-05	9,744E-04
PABPC1P9	-0,848	7,051E-05	9,778E-04
HEPHL1	-1,947	7,131E-05	9,856E-04
RAB38	-1,049	7,240E-05	9,952E-04
ATOSB	-0,500	7,252E-05	9,959E-04

Para una visualización de los 752 genes diferenciales infraexpresados (FDR<0,01) acceden a este código QR:



**Tabla 3 Anexo.** Los 200 genes diferenciales con mayor AUC

ID	AUC
SYCE1L	0,848
TOMM5	0,843
ZNF232-AS1	0,842
NDUFA11	0,839
PICK1	0,836
CFAP119	0,827
MRPL20	0,825
PRKACB-DT	0,825
SNHG25	0,824
SCAT2	0,823
FLJ40194	0,821
NDUFA2	0,821
TMA7	0,821
LIN7B	0,820
NOL12	0,818
WDR45	0,818
THUMPD3-AS1	0,817
RPP21	0,817
CDKN2AIPNL	0,814
SNORD104	0,812
RNVU1-30	0,812
RPS10	0,810
TRAPPC2	0,810
FANCG	0,809
LCDR	0,809
DICER1-AS1	0,809
SNORA13	0,808
CENPS	0,807
NPEPL1	0,806
TNKS2-DT	0,805
PXMP2	0,804
IQCC	0,804
SNHG12	0,804
SRA1	0,803
PET100	0,803
ERC1	0,802
RNF32-AS1	0,802
C15orf61	0,801
AGER	0,801
DMAP1	0,801
MIIP	0,799
RNASEH2C	0,798
SS18L2	0,797
SWAP70	0,797
MESP1	0,796
MIRLET7A1HG	0,796
CAMTA1-DT	0,795
CCNL2	0,795
DUSP16	0,794
HNRNPD-DT	0,793
CDK5	0,793
ATP5IF1	0,793
CCDC107	0,793
TGFBR1	0,793
SPON1	0,793
ALOX12-AS1	0,792
HLCS-AS1	0,792
LINC01023	0,792



TRPC6	0,792
CALML6	0,792
COL4A5	0,792
MIR25	0,792
RANBP3-DT	0,792
UEVLD	0,791
FZD9	0,791
DPH7	0,791
HEMK1	0,790
SNORD99	0,790
BLOC1S1	0,790
CPEB1	0,789
CDKL3	0,789
GLI4	0,789
UQCRQ	0,789
FAM229A	0,788
SLC20A1-DT	0,788
CAHM	0,788
CCDC12	0,788
EZH2	0,788
LIN37	0,788
SLC18A2	0,787
ARL6IP4	0,786
RBIS	0,786
LEAP2	0,786
LSM3	0,785
MRPL2	0,785
SNX29	0,785
ZNHIT1	0,785
E2F1	0,784
IFRD1	0,784
MRPL20-DT	0,784
SNRNP70	0,784
CLASRP	0,783
RB1	0,783
SPPL2B	0,782
MAD2L2	0,782
POP4	0,782
UBE2J2	0,782
NDUFS6	0,782
QTRT1	0,782
AURKAIP1	0,781
ENTPD3-AS1	0,781
MIR17HG	0,781
SPATA1	0,781
NDUFAF6	0,780
SHISAL1	0,780
BABAM1	0,780
ABHD17A	0,780
ELP6	0,780
TUSC1	0,779
COPS9	0,779
CLDND2	0,779
MORN2	0,779
RPL13AP7	0,779
HAUS7	0,778
RPS19BP1	0,778
ATG16L2	0,778
METTL15P1	0,778
SNHG11	0,778
BRK1	0,778
FAM98C	0,778
LINC01311	0,778
SNHG1	0,778
UBXN11	0,777
FMC1	0,777
IFT27	0,777
LINC01089	0,777

HDAC10	0,777
REX1BD	0,777
SKP1	0,777
CRNDE	0,776
GMNN	0,776
KDM5A	0,776
POLR3F	0,776
UQCC3	0,776
ZRSR2	0,776
ZSCAN16-AS1	0,776
BGLAP	0,775
OTULIN-DT	0,775
UBE2M	0,775
FASTK	0,774
COL12A1	0,774
NAV1	0,774
NSUN5P1	0,774
BCL2L12	0,774
CEMIP	0,774
ENDOG	0,774
PSMD9	0,774
BORCS8	0,773
ILKAP	0,773
NPRL2	0,773
RGS4	0,773
NPIPA1	0,773
SNHG10	0,773
ZNF205	0,773
PFKFB3	0,773
TTC9B	0,772
CYFIP1	0,772
CDK5RAP1	0,772
MXD3	0,772
POC1A	0,772
RNF32-DT	0,772
WNK1	0,772
NDUFB11	0,772
POP5	0,772
PPIH	0,772
ATAD3B	0,771
ATP5PO	0,771
FAAP20	0,771
PLEKHJ1	0,771
SNHG32	0,771
SNORA33	0,771
CRYZP1	0,771
EHD2	0,771
GPS2	0,771
IL6R	0,771
RBMS1P1	0,771
SP2-DT	0,771
ARHGAP11B	0,770
ATAT1	0,770
C1orf159	0,770
JTB	0,770
MBLAC1	0,770
METTL6	0,770
ZNF594-DT	0,770
CCDC167	0,770
XRCC3	0,770
CTTNBP2NL	0,770
RNU6-8	0,769
CFAP298	0,769
PAXX	0,769
RABAC1	0,769
SLC25A33	0,769
ZNF783	0,768
CAP1P1	0,768

IFI27L1	0,768
SCIRT	0,768
FBLL1	0,768
SMIM45	0,768
CCHCR1	0,767
CDK10	0,767

Para la visualización de los 1.816 genes con AUC>0,7 acceden al siguiente código QR:



**Tabla 4 Anexo.** Los DEGs sobreexpresados con FDR<0,01 y AUC>0,7 en orden alfabético

ID	
AAK1	0,757
ABCA17P	0,751
ABHD1	0,729
ACBD4	0,726
ACP6	0,719
ADCY10P1	0,738
ADD3	0,705
AGBL3	0,755
AKAP1-DT	0,722
ALG1L2	0,703
ALG6	0,707
ALOX12-AS1	0,792
ANKHD1	0,741
ANKHD1-DT	0,741
AP4B1	0,734
APOO	0,742
APRG1	0,717
ARHGAP11B	0,770
ARHGAP12	0,717
ARHGAP33	0,764
ARRHGEF33	0,701
ARID5B	0,711
ARIH2OS	0,761
ASB16-AS1	0,739
ASXL2	0,705
ATAT1	0,770
ATE1	0,701
ATP23	0,730
ATP6V0E2-AS1	0,718
BIRC6	0,748
BLTP2	0,708
BMS1P4	0,715
BOLA1	0,743

BRME1	0,728
BSG-AS1	0,732
C21orf58	0,738
C3orf33	0,733
C5orf34	0,745
C8orf44	0,757
CCDC125	0,754
CCDC150	0,761
CCDC177	0,711
CCDC183	0,716
CCNY-AS1	0,758
CDC25C	0,734
CDH24	0,724
CDK12	0,701
CDK20	0,724
CDK5RAP1	0,772
CDK5RAP3	0,736
CDKL3	0,789
CDKN2B-AS1	0,704
CDT1	0,715
CENPM	0,729
CHAF1A	0,709
CHAF1B	0,751
CHASERR	0,701
CHTOP	0,717
CIDECP1	0,729
CIRBP-AS1	0,714
CNTD1	0,737
CRBN	0,714
CROCCP2	0,728
CSAD	0,716
CYP2D6	0,731
CYP2D7	0,710
DDX11-AS1	0,717
DET1	0,711
DLEU2	0,700
DNAJC9-AS1	0,703
DOC2GP	0,745
DONSON	0,723
DSN1	0,743
DUSP12	0,732
DUT	0,735
E2F1	0,784
ECHDC2	0,719
EFCAB15P	0,730
EIF4EBP3	0,721
ELAC1	0,737

EMC1-AS1	0,709
EOLA1-DT	0,737
ERO1B	0,707
EXOSC9	0,723
EZH2	0,788
FAHD2A	0,706
FAHD2CP	0,752
FAM241B	0,741
FAM98C	0,778
FBH1	0,755
FBXO43	0,740
FEM1A	0,706
FZD9	0,791
GCDH	0,701
GKAP1	0,707
GNRH1	0,757
GOLGA2P5	0,707
GOLGA6L9	0,711
GPR89B	0,723
GRIN2C	0,767
GUSBP18	0,713
H1-10-AS1	0,713
HACL1	0,741
HAUS5	0,728
HCG27	0,718
HDAC10	0,777
HESX1	0,730
HIVEP1	0,714
HLCS-AS1	0,792
HMG20A	0,707
HMG2P15	0,732
HNRNPA1P16	0,704
HPS5	0,723
HULC	0,700
IER3IP1	0,728
IFT22	0,746
IFT27	0,777
IL17RB	0,708
INO80B	0,740
INO80E	0,703
IRX6	0,734
KANSL1L-AS1	0,714
KDM5A	0,776
KDSR-DT	0,755
KIAA1217	0,745
KLHDC9	0,713
KLHL11	0,713

KMT5C	0,720
KRBOX4	0,720
KRTCAP2	0,725
LAMB2P1	0,704
LARP4B	0,707
LEAP2	0,786
LIFR-AS1	0,705
LINC01138	0,720
LINC01275	0,714
LINC01355	0,721
LINC01409	0,707
LINC02609	0,720
LINC02615	0,701
LIPE	0,717
LNCTAM34A	0,718
LPCAT3	0,746
LRGUK	0,712
LRRC56	0,744
LSM8	0,732
LUC7L2	0,713
LYRM9	0,719
MAMDC4	0,735
MATN1-AS1	0,718
MCM8	0,752
MELTF-AS1	0,752
METTL17	0,727
METTL18	0,719
METTL27	0,729
METTL2A	0,708
MIR3936HG	0,715
MORN1	0,703
MPC2	0,726
MRPL46	0,713
MST1	0,753
MST1P2	0,752
MTERF2	0,736
MTHFS	0,728
MYL5	0,739
MYO15A	0,723
NADK2	0,715
NASP	0,704
NCOA3	0,703
NCOA4	0,717
NDUFA6-DT	0,703
NEIL1	0,745
NF1	0,734
NFATC3	0,744

NIPSNAP3B	0,715
NPRL2	0,773
NT5M	0,754
NUDT6	0,710
NUFIP2	0,716
NUP98	0,710
OBSCN-AS1	0,700
OFD1	0,758
ORMDL1	0,720
OSBPL7	0,713
OSGEP	0,722
OXER1	0,727
OXNAD1	0,742
PABPC1L	0,721
PABPN1	0,724
PAM16	0,724
PARTICL	0,742
PCBD2	0,734
PCBP2P2	0,735
PCNA	0,738
PCP2	0,713
PDCL3P4	0,702
PDHA1	0,714
PIDD1	0,743
PIF1	0,730
PIK3C2A	0,710
PIK3IP1-DT	0,711
PJVK	0,731
PNMA3	0,720
POC1B-AS1	0,718
PODXL2	0,722
POLR2H	0,718
POLR3F	0,776
PPIEL	0,713
PPM1L-DT	0,712
PPOX	0,732
PPP1R3F	0,718
PRKAG2-AS1	0,724
PROCA1	0,701
PRORS1P	0,706
PRRT3	0,715
PSEN2	0,709
PSMA3-AS1	0,705
PSTK	0,747
PTBP2	0,728
PTOV1-AS2	0,736
PTPRG-AS1	0,717

RAD1	0,714
RAD18	0,706
RAD51-AS1	0,759
RAD54B	0,711
RAD9A	0,712
RAPGEF5	0,707
RBM6	0,731
RBM8A	0,708
RBP7	0,752
REL	0,740
REM2	0,761
RESF1	0,764
RFC4	0,750
RFESD	0,727
RHEBL1	0,733
RHPN1	0,721
RIBC2	0,737
RIPPLY3	0,721
RNA5SP383	0,702
RNF111	0,721
RNF32	0,755
RNF32-AS1	0,802
RNF32-DT	0,772
RPA2	0,737
RPL36A	0,715
RRP7BP	0,714
RSAD1	0,741
SBF2	0,725
SCAMP1-AS1	0,707
SDHAP3	0,721
SELENOO-AS1	0,746
SEPTIN7P13	0,718
SEPTIN7P14	0,730
SEPTIN7P2	0,724
SETD4	0,705
SETMAR	0,732
SETX	0,701
SF11	0,702
SH3BP5-AS1	0,707
SHOC1	0,711
SIGIRR	0,704
SINHCAF	0,710
SIX2	0,724
SLC16A4-AS1	0,741
SLC16A8	0,731
SLC25A14	0,765
SLC25A21-AS1	0,742



SLC4A5	0,720
SLC5A3	0,742
SMC1B	0,742
SMG1P7	0,744
SMG9	0,705
SMIM19	0,717
SMIM45	0,768
SMPD4BP	0,709
SNX32	0,717
SPAG8	0,752
SPC24	0,729
SPICE1	0,737
SRRM2-AS1	0,712
SRSF3	0,704
STAG3	0,728
SUGP1	0,701
SUV39H1	0,730
SUZ12P1	0,750
SYCE2	0,736
SYCP2	0,708
TAF6L	0,716
TAMM41	0,701
TASOR2	0,715
TATDN2	0,715
TCAM1P	0,745
TESMIN	0,702
TET1	0,712
TEX22	0,741
TEX45	0,747
THUMPD3-AS1	0,817
TMED2-DT	0,739
TMEM161B-DT	0,720
TMEM18	0,702
TMPO-AS1	0,727
TONSL	0,740
TOR3A	0,707
TRAIP	0,749
TRAPPC2	0,810
TRIM45	0,722
TRIM74	0,711
TSNAXIP1	0,718
TTC21A	0,715
TTC23L	0,729
TTC28-AS1	0,703
TTC32	0,753
TUBGCP6	0,700
UBE2Q2P13	0,741

UBXN11	0,777
USP20	0,717
USP34	0,715
VHL	0,729
WAPL-DT	0,736
WASF2	0,707
WASH2P	0,728
WBP1	0,719
WBP2NL	0,710
WDR62	0,734
WDR93	0,722
WDR97	0,714
WIPF2	0,718
WRAP73	0,745
YBX2	0,725
YJU2B	0,722
YPEL1	0,704
ZBTB38	0,753
ZKSCAN2-DT	0,713
ZMPSTE24-DT	0,739
ZMYND10	0,708
ZMYND11	0,732
ZNF213-AS1	0,707
ZNF217	0,707
ZNF22	0,706
ZNF32	0,716
ZNF407-AS1	0,702
ZNF460	0,729
ZNF567-DT	0,707
ZNF576	0,726
ZNF594-DT	0,770
ZNF767P	0,755
ZNF783	0,768
ZNF815P	0,707
ZRANB2	0,701

**Tabla 5 anexo:** Los 150 procesos activados con menor FDR según el GSEA.

GO TERM	Count	LOR	pvalor	FDR	Función
GO:0006397	498	1,016	1,565E-105	9,310E-101	mRNA processing
GO:0008380	428	1,002	6,383E-97	1,898E-92	RNA splicing
GO:0000375	342	1,030	1,971E-85	3,908E-81	RNA splicing, via transesterification reactions
GO:0016570	448	1,013	2,959E-84	2,997E-80	histone modification
GO:0000377	339	1,028	3,024E-84	2,997E-80	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
GO:0000398	339	1,028	3,024E-84	2,997E-80	mRNA splicing, via spliceosome
GO:0016569	468	1,000	7,685E-84	6,530E-80	covalent chromatin modification
GO:0007059	308	1,079	5,954E-74	4,426E-70	chromosome segregation
GO:0018205	390	1,018	2,326E-73	1,537E-69	peptidyl-lysine modification
GO:0044782	363	0,975	4,680E-70	2,784E-66	cilium organization
GO:0006261	153	1,414	4,934E-68	2,667E-64	DNA-dependent DNA replication

GO:0006260	276	1,280	1,138E-66	5,641E-63	DNA replication
GO:0060271	345	0,977	1,432E-66	6,552E-63	cilium assembly
GO:0033044	328	0,972	1,073E-62	4,560E-59	regulation of chromosome organization
GO:0098813	251	1,093	1,169E-62	4,636E-59	nuclear chromosome segregation
GO:0006302	230	1,171	2,221E-62	8,254E-59	double-strand break repair
GO:0006310	271	1,101	4,480E-61	1,567E-57	DNA recombination
GO:0034660	439	0,759	5,606E-60	1,852E-56	ncRNA metabolic process
GO:0000819	181	1,181	4,082E-58	1,278E-54	sister chromatid segregation
GO:0022613	418	0,744	6,547E-56	1,947E-52	ribonucleoprotein complex biogenesis
GO:1901987	471	0,759	1,018E-55	2,884E-52	regulation of cell cycle phase transition
GO:0071103	267	1,048	1,868E-55	5,049E-52	DNA conformation change
GO:0048285	450	0,794	1,222E-54	3,160E-51	organelle fission
GO:1901990	437	0,775	3,968E-54	9,833E-51	regulation of mitotic cell cycle phase transition
GO:0043414	295	0,957	5,819E-54	1,384E-50	macromolecule methylation
GO:0032259	352	0,882	4,663E-53	1,067E-49	methylation
GO:0034470	365	0,770	2,162E-52	4,761E-49	ncRNA processing
GO:0044786	71	1,593	6,241E-52	1,326E-48	cell cycle DNA replication
GO:0000280	410	0,815	1,934E-51	3,967E-48	nuclear division
GO:0033260	60	1,557	2,974E-50	5,896E-47	nuclear DNA replication
GO:0090305	291	0,886	2,340E-49	4,489E-46	nucleic acid phosphodiester bond hydrolysis
GO:0000070	149	1,160	2,927E-49	5,441E-46	mitotic sister chromatid segregation
GO:0051169	346	0,819	1,158E-48	2,087E-45	nuclear transport
GO:0006913	343	0,819	3,446E-48	6,027E-45	nucleocytoplasmic transport
GO:0010498	461	0,670	1,046E-47	1,777E-44	proteasomal protein catabolic process
GO:0051052	337	0,857	1,213E-47	2,004E-44	regulation of DNA metabolic process
GO:0000725	132	1,298	1,532E-47	2,463E-44	recombinational repair
GO:0071824	216	1,002	2,272E-47	3,556E-44	protein-DNA complex subunit organization
GO:0000724	131	1,290	7,040E-47	1,074E-43	double-strand break repair via homologous recombination
GO:0006403	227	0,951	2,065E-46	3,071E-43	RNA localization
GO:0031503	274	0,886	4,199E-46	6,090E-43	protein-containing complex localization
GO:0043161	409	0,690	1,321E-45	1,871E-42	proteasome-mediated ubiquitin-dependent protein catabolic process
GO:0140014	272	0,927	1,370E-44	1,894E-41	mitotic nuclear division
GO:0065004	178	1,038	1,232E-43	1,665E-40	protein-DNA complex assembly
GO:0010948	357	0,749	1,111E-42	1,468E-39	negative regulation of cell cycle process
GO:0015931	238	0,907	4,609E-42	5,959E-39	nucleobase-containing compound transport
GO:0032200	146	1,107	6,069E-42	7,640E-39	telomere organization
GO:0000075	215	1,041	6,754E-42	8,198E-39	cell cycle checkpoint
GO:1903311	328	0,756	1,557E-41	1,852E-38	regulation of mRNA metabolic process
GO:0072331	265	0,935	1,695E-41	1,938E-38	signal transduction by p53 class mediator
GO:0000723	145	1,105	1,679E-41	1,938E-38	telomere maintenance
GO:0044839	269	0,875	2,750E-41	3,085E-38	cell cycle G2/M phase transition
GO:0022616	24	1,693	3,400E-41	3,745E-38	DNA strand elongation
GO:0051236	193	0,987	6,432E-41	6,955E-38	establishment of RNA localization
GO:0006338	189	1,021	7,981E-41	8,476E-38	chromatin remodeling
GO:2001020	211	0,947	1,352E-40	1,411E-37	regulation of response to DNA damage stimulus
GO:0070646	305	0,765	1,481E-40	1,518E-37	protein modification by small protein removal
GO:0006405	133	1,113	5,909E-40	5,956E-37	RNA export from nucleus
GO:0051168	192	1,012	6,134E-40	6,080E-37	nuclear export
GO:0050657	190	0,975	1,619E-39	1,553E-36	nucleic acid transport
GO:0050658	190	0,975	1,619E-39	1,553E-36	RNA transport
GO:0006399	163	0,931	6,373E-39	5,923E-36	tRNA metabolic process
GO:0043543	246	0,885	1,609E-38	1,472E-35	protein acylation
GO:0000209	302	0,719	6,142E-38	5,535E-35	protein polyubiquitination
GO:0006479	172	1,035	1,186E-36	1,038E-33	protein methylation
GO:0008213	172	1,035	1,186E-36	1,038E-33	protein alkylation
GO:0032392	94	1,192	2,675E-36	2,306E-33	DNA geometric change
GO:0016579	288	0,740	3,010E-36	2,557E-33	protein deubiquitination
GO:0000086	252	0,838	3,238E-36	2,713E-33	G2/M transition of mitotic cell cycle
GO:0071426	125	1,089	4,070E-36	3,316E-33	ribonucleoprotein complex export from nucleus

GO:0071166	126	1,082	4,992E-36	4,012E-33	ribonucleoprotein complex localization
GO:0034508	41	1,531	5,517E-36	4,375E-33	centromere complex assembly
GO:0031124	97	1,155	8,872E-36	6,942E-33	mRNA 3'-end processing
GO:0009314	441	0,622	2,380E-35	1,838E-32	response to radiation
GO:0006611	177	0,982	2,565E-35	1,956E-32	protein export from nucleus
GO:0050684	140	1,039	5,099E-35	3,839E-32	regulation of mRNA processing
GO:1902749	214	0,897	5,855E-35	4,353E-32	regulation of cell cycle G2/M phase transition
GO:0031570	155	1,143	6,443E-35	4,731E-32	DNA integrity checkpoint
GO:0043967	65	1,343	8,552E-35	6,202E-32	histone H4 acetylation
GO:0007051	169	0,955	8,666E-35	6,210E-32	spindle organization
GO:0006473	203	0,929	1,427E-34	9,982E-32	protein acetylation
GO:2001252	170	0,984	1,757E-34	1,215E-31	positive regulation of chromosome organization
GO:0006271	17	1,804	2,843E-34	1,944E-31	DNA strand elongation involved in DNA replication
GO:0051028	149	1,039	4,171E-34	2,819E-31	mRNA transport
GO:0007018	338	0,677	5,558E-34	3,714E-31	microtubule-based movement
GO:0010639	379	0,641	6,434E-34	4,252E-31	negative regulation of organelle organization
GO:0045930	333	0,686	6,521E-34	4,262E-31	negative regulation of mitotic cell cycle
GO:0034502	80	1,331	1,668E-33	1,079E-30	protein localization to chromosome
GO:0006914	491	0,545	2,045E-33	1,294E-30	autophagy
GO:0061919	491	0,545	2,045E-33	1,294E-30	process utilizing autophagic mechanism
GO:0006406	109	1,148	3,009E-33	1,864E-30	mRNA export from nucleus
GO:0071427	109	1,148	3,009E-33	1,864E-30	mRNA-containing ribonucleoprotein complex export from nucleus
GO:0044843	290	0,731	3,178E-33	1,949E-30	cell cycle G1/S phase transition
GO:0051656	397	0,606	5,241E-33	3,180E-30	establishment of organelle localization
GO:0045814	57	1,355	6,650E-33	3,995E-30	negative regulation of gene expression, epigenetic
GO:0016311	470	0,553	1,556E-32	9,251E-30	dephosphorylation
GO:1901796	179	0,992	4,109E-32	2,396E-29	regulation of signal transduction by p53 class mediator
GO:0000082	272	0,740	5,354E-32	3,091E-29	G1/S transition of mitotic cell cycle
GO:0031023	133	1,211	7,065E-32	4,040E-29	microtubule organizing center organization
GO:0006282	120	1,083	7,655E-32	4,336E-29	regulation of DNA repair
GO:0032508	85	1,185	8,181E-32	4,590E-29	DNA duplex unwinding
GO:0018394	169	0,996	1,006E-31	5,590E-29	peptidyl-lysine acetylation
GO:0043981	16	1,815	1,309E-31	7,144E-29	histone H4-K5 acetylation
GO:0043982	16	1,815	1,309E-31	7,144E-29	histone H4-K8 acetylation
GO:0051054	189	0,878	2,350E-31	1,270E-28	positive regulation of DNA metabolic process
GO:0045787	379	0,627	9,884E-31	5,202E-28	positive regulation of cell cycle
GO:0031123	148	0,945	1,133E-30	5,913E-28	RNA 3'-end processing
GO:0071826	195	0,790	1,303E-30	6,736E-28	ribonucleoprotein complex subunit organization
GO:0000077	143	1,078	1,322E-30	6,779E-28	DNA damage checkpoint
GO:0006323	173	0,961	1,672E-30	8,499E-28	DNA packaging
GO:0010389	198	0,863	2,293E-30	1,156E-27	regulation of G2/M transition of mitotic cell cycle
GO:0006401	394	0,573	2,491E-30	1,245E-27	RNA catabolic process
GO:0022618	188	0,799	3,114E-30	1,544E-27	ribonucleoprotein complex assembly
GO:0006475	166	0,976	3,455E-30	1,698E-27	internal protein amino acid acetylation
GO:0006333	152	1,019	5,797E-30	2,826E-27	chromatin assembly or disassembly
GO:0018022	123	1,173	6,051E-30	2,926E-27	peptidyl-lysine methylation
GO:0034724	39	1,401	7,422E-30	3,560E-27	DNA replication-independent nucleosome organization
GO:0031055	32	1,516	9,376E-30	4,461E-27	chromatin remodeling at centromere
GO:0009451	156	0,875	1,192E-29	5,628E-27	RNA modification
GO:0018393	161	0,988	1,402E-29	6,565E-27	internal peptidyl-lysine acetylation
GO:0043044	73	1,229	1,849E-29	8,593E-27	ATP-dependent chromatin remodeling
GO:0007093	166	0,984	1,922E-29	8,861E-27	mitotic cell cycle checkpoint
GO:0007098	124	1,217	3,076E-29	1,407E-26	centrosome cycle
GO:0032875	10	2,001	3,739E-29	1,697E-26	regulation of DNA endoreduplication
GO:0016573	157	1,003	3,985E-29	1,795E-26	histone acetylation
GO:0090329	56	1,511	5,469E-29	2,445E-26	regulation of DNA-dependent DNA replication
GO:0002429	468	0,507	5,877E-29	2,589E-26	immune response-activating cell surface receptor signaling pathway
GO:0002757	468	0,507	5,877E-29	2,589E-26	immune response-activating signal transduction

GO:0006336	38	1,402	6,051E-29	2,646E-26	DNA replication-independent nucleosome assembly
GO:0070507	185	0,840	6,876E-29	2,985E-26	regulation of microtubule cytoskeleton organization
GO:0016571	134	1,116	7,502E-29	3,233E-26	histone methylation
GO:0008033	124	0,917	8,700E-29	3,723E-26	tRNA processing
GO:0040029	133	1,064	1,191E-28	5,061E-26	regulation of gene expression, epigenetic
GO:1902850	130	0,961	1,516E-28	6,393E-26	microtubule cytoskeleton organization involved in mitosis
GO:0051983	99	1,104	1,877E-28	7,863E-26	regulation of chromosome segregation
GO:0034248	470	0,514	2,916E-28	1,213E-25	regulation of cellular amide metabolic process
GO:1902275	173	0,902	5,533E-28	2,285E-25	regulation of chromatin organization
GO:0099111	181	0,837	7,301E-28	2,995E-25	microtubule-based transport
GO:0043484	137	0,939	1,147E-27	4,673E-25	regulation of RNA splicing
GO:0043087	475	0,503	1,187E-27	4,801E-25	regulation of GTPase activity
GO:0032886	233	0,738	1,752E-27	7,042E-25	regulation of microtubule-based process
GO:2000104	26	1,696	2,367E-27	9,448E-25	negative regulation of DNA-dependent DNA replication
GO:0007062	56	1,447	2,383E-27	9,449E-25	sister chromatid cohesion
GO:0006417	418	0,533	3,678E-27	1,449E-24	regulation of translation
GO:0042254	285	0,619	4,677E-27	1,830E-24	ribosome biogenesis
GO:0034080	30	1,489	4,869E-27	1,881E-24	CENP-A containing nucleosome assembly
GO:0061641	30	1,489	4,869E-27	1,881E-24	CENP-A containing chromatin organization
GO:2001251	132	1,002	6,181E-27	2,372E-24	negative regulation of chromosome organization
GO:0006402	361	0,561	6,656E-27	2,538E-24	mRNA catabolic process
GO:0016925	79	1,108	1,178E-26	4,462E-24	protein sumoylation

Para la visualización de los 1.731 procesos biológicos activados con FDR <0,01 acceden al siguiente código QR:



**Tabla 6 anexo:** Los 98 procesos biológicos enriquecidos con FDR <0,01 según análisis DAVID

GO TERM	Count	Pvalor	FDR	Function
GO:0006357	562	7,985E-37	5,360E-33	regulation of transcription from RNA polymerase II promoter
GO:0006281	158	1,236E-35	4,148E-32	DNA repair
GO:0006355	363	6,685E-33	1,496E-29	regulation of transcription, DNA-templated
GO:0006338	160	9,386E-30	1,575E-26	chromatin remodeling
GO:0006325	152	8,522E-28	1,015E-24	chromatin organization
GO:0006260	81	9,070E-28	1,015E-24	DNA replication
GO:0051301	171	1,147E-27	1,100E-24	cell division
GO:0006974	138	2,551E-24	2,141E-21	cellular response to DNA damage stimulus
GO:0006397	108	1,923E-18	1,434E-15	mRNA processing
GO:0007049	144	1,940E-17	1,302E-14	cell cycle
GO:0045944	367	3,052E-17	1,862E-14	positive regulation of transcription from RNA polymerase II promoter
GO:0000122	310	7,795E-16	4,360E-13	negative regulation of transcription from RNA polymerase II promoter
GO:0000398	91	1,194E-15	6,166E-13	mRNA splicing, via spliceosome
GO:0045893	235	4,194E-15	2,011E-12	positive regulation of transcription, DNA-templated
GO:0000724	57	4,530E-15	2,027E-12	double-strand break repair via homologous recombination

GO:0060271	100	9,216E-15	3,866E-12	cilium assembly
GO:0008380	91	6,223E-14	2,457E-11	RNA splicing
GO:0007059	53	2,623E-13	9,779E-11	chromosome segregation
GO:0000070	29	1,608E-12	5,679E-10	mitotic sister chromatid segregation
GO:0032508	37	1,503E-11	5,044E-09	DNA duplex unwinding
GO:0045892	187	2,301E-11	7,354E-09	negative regulation of transcription, DNA-templated
GO:0007099	20	2,140E-10	6,530E-08	centriole replication
GO:0000278	66	3,135E-10	9,149E-08	mitotic cell cycle
GO:1902459	30	5,595E-10	1,565E-07	positive regulation of stem cell population maintenance
GO:0031297	26	1,226E-09	3,292E-07	replication fork processing
GO:0010212	27	1,869E-09	4,824E-07	response to ionizing radiation
GO:0000723	34	2,454E-09	6,100E-07	telomere maintenance
GO:0043484	41	5,600E-09	1,343E-06	regulation of RNA splicing
GO:0006302	38	8,851E-09	2,049E-06	double-strand break repair
GO:0006468	140	1,133E-08	2,535E-06	protein phosphorylation
GO:0000381	35	1,329E-08	2,793E-06	regulation of alternative mRNA splicing, via spliceosome
GO:0006606	48	1,331E-08	2,793E-06	protein import into nucleus
GO:0016567	153	1,794E-08	3,649E-06	protein ubiquitination
GO:0036297	25	1,870E-08	3,693E-06	interstrand cross-link repair
GO:0006275	28	1,974E-08	3,723E-06	regulation of DNA replication
GO:0006406	33	1,997E-08	3,723E-06	mRNA export from nucleus
GO:0042073	23	4,237E-08	7,687E-06	intraciliary transport
GO:0006270	20	5,833E-08	1,030E-05	DNA replication initiation
GO:0016310	98	6,929E-08	1,192E-05	phosphorylation
GO:0051321	37	1,265E-07	2,123E-05	meiotic cell cycle
GO:0051028	42	1,337E-07	2,189E-05	mRNA transport
GO:0051726	91	1,388E-07	2,219E-05	regulation of cell cycle
GO:0051984	13	2,608E-07	4,072E-05	positive regulation of chromosome segregation
GO:0043966	27	5,292E-07	8,072E-05	histone H3 acetylation
GO:0006913	28	5,497E-07	8,180E-05	nucleocytoplasmic transport
GO:0006282	30	5,606E-07	8,180E-05	regulation of DNA repair
GO:0007062	18	6,656E-07	9,505E-05	sister chromatid cohesion
GO:0007098	26	8,401E-07	1,175E-04	centrosome cycle
GO:0018105	66	1,012E-06	1,386E-04	peptidyl-serine phosphorylation
GO:0006298	20	1,269E-06	1,703E-04	mismatch repair
GO:0006268	16	1,389E-06	1,793E-04	DNA unwinding involved in DNA replication
GO:0071539	16	1,389E-06	1,793E-04	protein localization to centrosome
GO:0090307	23	1,982E-06	2,510E-04	mitotic spindle assembly
GO:0000375	17	2,188E-06	2,719E-04	RNA splicing, via transesterification reactions
GO:0007095	22	3,144E-06	3,836E-04	mitotic G2 DNA damage checkpoint
GO:1905515	29	3,247E-06	3,892E-04	non-motile cilium assembly
GO:2000781	23	3,372E-06	3,971E-04	positive regulation of double-strand break repair
GO:0031124	13	3,469E-06	4,014E-04	mRNA 3'-end processing
GO:0030174	12	4,260E-06	4,765E-04	regulation of DNA-dependent DNA replication initiation

GO:0031053	12	4,260E-06	4,765E-04	primary miRNA processing
GO:0080182	15	4,758E-06	5,235E-04	histone H3-K4 trimethylation
GO:0030030	28	5,136E-06	5,510E-04	cell projection organization
GO:0032259	47	5,172E-06	5,510E-04	methylation
GO:0016573	21	8,542E-06	8,959E-04	histone acetylation
GO:2000001	12	1,305E-05	1,348E-03	regulation of DNA damage checkpoint
GO:0006303	20	1,361E-05	1,384E-03	double-strand break repair via nonhomologous end joining
GO:1905168	21	1,423E-05	1,425E-03	positive regulation of double-strand break repair via homologous recombination
GO:1900264	11	1,677E-05	1,655E-03	positive regulation of DNA-directed DNA polymerase activity
GO:0048478	9	1,966E-05	1,891E-03	replication fork protection
GO:1903706	10	1,973E-05	1,891E-03	regulation of hemopoiesis
GO:0009411	23	2,235E-05	2,113E-03	response to UV
GO:0000731	13	2,356E-05	2,196E-03	DNA synthesis involved in DNA repair
GO:0006310	35	2,697E-05	2,480E-03	DNA recombination
GO:0045739	23	3,408E-05	2,976E-03	positive regulation of DNA repair
GO:0043982	12	3,413E-05	2,976E-03	histone H4-K8 acetylation
GO:0043981	12	3,413E-05	2,976E-03	histone H4-K5 acetylation
GO:1902455	14	3,414E-05	2,976E-03	negative regulation of stem cell population maintenance
GO:0007094	18	3,465E-05	2,981E-03	mitotic spindle assembly checkpoint
GO:2000045	25	4,395E-05	3,734E-03	regulation of G1/S transition of mitotic cell cycle
GO:0051310	11	4,805E-05	4,032E-03	metaphase plate congression
GO:2000819	16	5,049E-05	4,171E-03	regulation of nucleotide-excision repair
GO:0034644	27	5,183E-05	4,171E-03	cellular response to UV
GO:0044818	13	5,220E-05	4,171E-03	mitotic G2/M transition checkpoint
GO:0007064	13	5,220E-05	4,171E-03	mitotic sister chromatid cohesion
GO:0043967	21	5,668E-05	4,476E-03	histone H4 acetylation
GO:0000281	29	5,739E-05	4,479E-03	mitotic cytokinesis
GO:0015031	124	6,277E-05	4,826E-03	protein transport
GO:0000727	10	6,471E-05	4,826E-03	double-strand break repair via break-induced replication
GO:1902275	10	6,471E-05	4,826E-03	regulation of chromatin organization
GO:0034501	10	6,471E-05	4,826E-03	protein localization to kinetochore
GO:0036092	14	6,806E-05	5,020E-03	phosphatidylinositol-3-phosphate biosynthetic process
GO:0000082	29	7,820E-05	5,705E-03	G1/S transition of mitotic cell cycle
GO:0031573	12	7,906E-05	5,706E-03	intra-S DNA damage checkpoint
GO:0051571	15	8,029E-05	5,710E-03	positive regulation of histone H3-K4 methylation
GO:0035825	9	8,082E-05	5,710E-03	reciprocal DNA recombination
GO:0006511	86	9,407E-05	6,577E-03	ubiquitin-dependent protein catabolic process
GO:0042789	23	1,088E-04	7,528E-03	mRNA transcription from RNA polymerase II promoter
GO:0048025	14	1,277E-04	8,748E-03	negative regulation of mRNA splicing, via spliceosome

13.2. ANEXO 2

**"Impact of Human Papillomavirus on Head and Neck Squamous Cell Carcinoma: Gene Expression Meta-Analysis"**

**Anna Isabel Gil Whitfield<sup>1</sup>, Marta Serna García<sup>1</sup>, Nicla Flacco<sup>1</sup>**

<sup>1</sup>Universidad Europea de Valencia, Faculty of Health Sciences, Department of Dentistry, Valencia, Spain

*Correspondence:*  
*Universidad Europea de Valencia*  
*Paseo Alameda, 7*  
*46010 – Valencia, Spain*  
*Annagil98@gmail.com*



**Abstract:**

**Background:** This study delves into the genetic landscape of head and neck squamous cell carcinomas (HNSCC), particularly focusing on human papillomavirus (HPV) type 16-associated cases. The aim is to identify differentially expressed genes in HPV-positive (HPV+) HNSCC cases and explore potential prognostic biomarkers, given the increasing number of cancer cases associated with HPV.

**Materials and Methods:** Leveraging data from the TCGA database, transcriptomic analysis was performed on 112 samples with available information on HPV type 16 infection. This included ROC-analysis, functional analysis, gene enrichment analysis via GSEA and DAVID software, and survival analysis using the corrected Fischer test.

**Results:** Comparative analysis revealed significant gene expression differences between HPV+ and HPV-negative (HPV-) HNSCC patients. HPV+ patients exhibited overexpression of 5203 genes and underexpression of 752 genes. Gene enrichment analysis linked these genes to crucial biological processes such as the cell cycle, chromatin organization, DNA replication, keratinization, and DNA repair. Noteworthy genes EZH2, KDM5A, HDAC10, BIRC6, and SPRR2G were highlighted for their significant differential expression (FDR <0,01), association with biological processes of interest, and potential as biomarkers, determined by an AUC >0,7 (Area under Curve) in ROC-analysis (Receiver Operating Characteristics). Further, these genes showed a tendency towards superior overall survival in accordance with the expression found in this study in the survival analysis.

**Conclusions:** Notably, genes EZH2 and BIRC6 emerged as promising biomarkers in HPV+ patients and showed a strong link to tumorigenesis. KDM5A, HDAC10 and SPRR2G need to be further studied in future investigations. All 5 genes showed an improvement of overall survival in accordance with their expression in our study in our HNSCC-HPV+ patient subset. These findings suggest the potential for personalized monitoring and early detection strategies in HPV+ patients to enhance patient outcomes related to tumorigenesis. Additionally, it underscores the generally better prognosis of HPV+ tumors compared to HPV- ones.

**Keywords:** Head and neck squamous cell carcinomas, Human papillomavirus, Functional analysis, Gene enrichment analysis, Survival analysis, EZH2, KDM5A, HDAC10, BIRC6, SPRR2G

**Introduction:**

Head and neck cancers account for a significant proportion of global cancer cases. They rank as the sixth most common cancer, with an annual incidence of between 630,000 and 660,000 diagnosed patients, and between 325,000 and 350,000 annual deaths [1]. Approximately 90% of head and neck cancers are squamous cell carcinomas, which manifest in the oral mucosa (OSCC), oropharynx (OPSCC), and larynx (ICD-10) [1-5]. OSCCs most frequently develop in the floor of the mouth, tongue, lips, gums, and buccal mucosa [2]. Their sites of origin correspond to the ICD-10 classification, comprising 19 regions from the lips and oral mucosa to the larynx, sinuses, and pharynx, among others. Aside from their high incidence, these cancers are also on a rising trend, estimated to increase by 30% from 2022 to 2030 [4].

Traditionally, these cancers have been associated with smoking and alcohol consumption, contributing to 80% of global cases [3]. In addition to these risk factors, chewing areca nut, primarily consumed in developing countries for its euphoric, psychoactive, and appetite-suppressing effects, is implicated, as are viral infections such as HPV or Epstein-Barr virus [6]. Other risk factors include chronic inflammation, such as periodontitis, chronic exposure to carcinogens in the workplace or environment, inadequate oral hygiene and nutrition, genetic predisposition, and immunodeficiency states, traumatic mucosal lesions, altered oral microbiota, and pathologies associated with decreased DNA repair [3, 6].

Most oral cancer risk factors are modifiable, underscoring the importance of lifestyle changes and HPV vaccination in prevention efforts [3, 6]. Oral squamous cell carcinoma (OSCC) is the most common cancer in men and the third most common in women in high-risk countries such as Bangladesh, India, Sri Lanka, and Pakistan. Two-thirds of new oral cancer cases are diagnosed in these high-risk countries [3]. In Europe, France is the most affected by OSCC [1]. Incidence is also age-related, being higher in individuals over 60 years old [3].

Although there has been a decline in oral and laryngeal squamous cell carcinoma incidence related to classical risk factors such as tobacco in the United States and other regions like Hong Kong and Canada, there has been an increase in head and neck cancers related to human papillomavirus [4].

This study aims to detect differential genes in samples of squamous cell carcinoma of the head and neck, comparing HPV positive patients (HNSCC-HPV+) with HPV negative patients (HNSCC-HPV-), investigate the biological functions associated with these genes, evaluate their impact on patient survival, and identify potential predictive biomarkers for the disease.

**Materials and Methods:**

To conduct this study, data from The Cancer Genome Atlas Project (TCGA) were utilized. TCGA, available at <http://tcga.cancer.gov>, offers extensive cancer-related 'omic' data for analysis and download. This resource, created by the National Cancer Institute and the National Human Genome Research Institute, has generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data, leading to improvements in cancer diagnosis, treatment, and prevention. Within the portal, samples were filtered

based on the focus cancer for our study: Head and Neck Cancer. From the TCGA Firehose Legacy data atlas, 530 samples of head and neck squamous cell carcinoma were observed, from which 112 were utilized, as they provided data on the presence or absence of HPV infection.

Following sample selection, Principal Component Analysis (PCA) was conducted comparing the 112 samples with a history of human papillomavirus (HPV) infection (Group 2) and those without (Group 1). A clear separation between the two groups was demonstrated.

For the differential analysis, these 112 samples were compared. EdgeR and Limma software from Bioconductor were employed, estimating normalized TPM readings and identifying differentially expressed genes (DEGs) through statistical methods (i.e., linear modeling and empirical Bayes) following Law et al.'s methodology. Genes with a false discovery rate (FDR)  $\leq 0.01$  were considered differentially expressed [7].

For ROC curve analysis, predictors or biomarkers were obtained using Receiver Operating Characteristic (ROC) curves. The genefilters package from the Bioconductor project was utilized, applying the "rowpAUCs" function to normalized values (TPMs) in both genetic groups (independent). An area under the ROC curve of 0.5 represents a test with no discriminative ability (no better than chance), while an AUC of 1 represents a test with perfect discriminative ability [8]. AUC values  $\geq 0.7$  were deemed significant and indicative of discriminative ability.

Enrichment analysis of biological processes using GSEA and DAVID software was conducted. GSEA utilized the MSigDB v7.5 database to study biological processes associated with all genes that exhibited independent expression changes, irrespective of significance, providing a broader view of potential transcriptomic changes in the HPV-positive group. Biological processes with an FDR  $\leq 0.01$  were considered statistically significant [9, 10]. DAVID software was used to observe biological processes associated with differentially expressed genes (DEGs) [11, 12]. Processes with an FDR  $\leq 0.01$  were considered statistically significant for overexpressed genes, while those with a p-value  $\leq 0.01$  were considered significant for underexpressed genes. Additionally, ClueGO, a Cytoscape plugin, was employed for graphical interpretation of non-redundant biological functions of our genes of interest, integrating Gene Ontology (GO) terms as well as KEGG/BioCarta pathways [13].

Survival analysis was performed using the R package "Survival" to analyze clinical information from the 530 head and neck squamous cell carcinoma samples from TCGA's Firehose Legacy data atlas. This analysis focused on genes identified as relevant in our study and their potential influence on overall disease survival. The analysis traditionally examines the time elapsed until the occurrence of a predefined event, in this case, death, using a 95% confidence interval [14].

## **Results:**

A total of 5,955 differential genes were detected with a False Discovery Rate (FDR) of less than 0.01 in 20 samples of oral tissue from HPV-positive HNSCC patients compared to 92 samples of oral tissue from HPV-negative patients. Among them, 5,203 genes were overexpressed (positive Fold Change) and 752 genes were underexpressed (negative Fold

Change) in HPV-positive patients compared to HPV-negative ones. The following 2 tables display the top 20 genes with the highest overexpression and the lowest underexpression among genes with FDR <0.01 (Tables 1 and 2).

**Table 1.** The 20 differentially expressed genes with FDR <0,01 with the highest fold change.

<b>ID</b>	<b>Nombre completo</b>	<b>Fold change (FC)*</b>	<b>p-value</b>	<b>padj (FDR)</b>
<b>COLCA1</b>	<b>Colorectal cancer associated 1</b>	2,995	5,652E-09	2,455E-06
<b>SMC1B</b>	<b>Structural Maintenance Of Chromosomes 1B</b>	2,663	9,705E-07	4,585E-05
<b>NOS2</b>	<b>Nitric Oxide Synthase 2</b>	2,528	5,668E-08	8,289E-06
<b>ZFR2</b>	<b>Zinc Finger RNA Binding Protein 2</b>	2,465	1,254E-04	1,511E-03
<b>SYCP2</b>	<b>Synaptonemal Complex Protein 2</b>	2,427	1,504E-07	1,444E-05
<b>CLGN</b>	<b>Calmegin</b>	2,382	2,617E-07	1,997E-05
<b>CD8B2</b>	<b>CD8B Family Member 2</b>	2,367	3,847E-05	6,178E-04
<b>KRT19</b>	<b>Keratin 19</b>	2,355	3,230E-04	3,097E-03
<b>MUC4</b>	<b>Mucin 4, Cell Surface Associated</b>	2,334	1,278E-04	1,531E-03
<b>YBX2</b>	<b>Y-Box Binding Protein 2</b>	2,260	3,462E-06	1,099E-04
<b>BTNL9</b>	<b>Butyrophilin Like 9</b>	2,241	4,790E-07	2,885E-05
<b>MAP7D2</b>	<b>MAP7 Domain Containing 2</b>	2,238	5,408E-05	7,951E-04
<b>ABCA17P</b>	<b>ATP Binding Cassette Subfamily A Member 17, Pseudogene</b>	2,218	1,410E-05	2,916E-04
<b>TDRD5</b>	<b>Tudor Domain Containing 5</b>	2,172	1,804E-05	3,483E-04
<b>IRX6</b>	<b>Iroquois Homeobox 6</b>	2,162	3,567E-05	5,837E-04
<b>ADARB2</b>	<b>Adenosine Deaminase RNA Specific B2 (Inactive)</b>	2,143	5,056E-07	2,992E-05
<b>IL17REL</b>	<b>Interleukin 17 Receptor E Like</b>	2,134	1,731E-05	3,378E-04
<b>DMRTA2</b>	<b>DMRT Like Family A2</b>	2,134	3,685E-04	3,436E-03
<b>WNK2</b>	<b>WNK Lysine Deficient Protein Kinase 2</b>	2,130	3,489E-06	1,101E-04
<b>FOXA1</b>	<b>Forkhead box A1</b>	2,119	2,327E-06	8,293E-05

\*Fold Change estimates refer to the HPV+ group

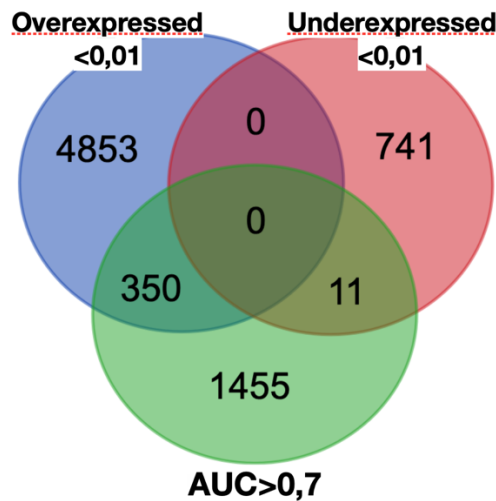
**Table 2.** The 20 differentially expressed genes with FDR <0,01 with the lowest fold change.

<b>ID</b>	<b>Nombre completo</b>	<b>Fold change (FC)*</b>	<b>p-value</b>	<b>Padj (FDR)</b>
<b>SPRR2G</b>	<b>Small Proline Rich Protein 2G</b>	-3,091	1,208E-06	5,308E-05
<b>KRTDAP</b>	<b>Keratinocyte Differentiation Associated Protein</b>	-2,905	1,900E-05	3,605E-04
<b>KLK5</b>	<b>Kallikrein Related Peptidase 5</b>	-2,810	8,350E-06	1,980E-04
<b>SBSN</b>	<b>Suprabasin</b>	-2,601	9,770E-07	4,598E-05
<b>WFDC12</b>	<b>WAP Four-Disulfide Core Domain 12</b>	-2,452	7,558E-06	1,844E-04
<b>CASP14</b>	<b>Caspase 14</b>	-2,367	1,080E-03	7,938E-03
<b>EREG</b>	<b>Epiregulin</b>	-2,357	1,398E-05	2,897E-04
<b>KLK7</b>	<b>Kallikrein related peptidase 7</b>	-2,342	9,056E-05	1,186E-03
<b>PRR9</b>	<b>Proline Rich 9</b>	-2,244	1,907E-04	2,071E-03
<b>KRT75</b>	<b>Keratin 75</b>	-2,228	1,658E-05	3,278E-04
<b>CPA4</b>	<b>Carboxypeptidase A4</b>	-2,212	7,010E-06	1,754E-04
<b>KRT14</b>	<b>Keratin 14</b>	-2,193	3,133E-07	2,221E-05
<b>LINC01404</b>	<b>long intergenic non-protein coding RNA 1404</b>	-2,192	2,120E-08	5,042E-06
<b>FOLR3</b>	<b>Folate Receptor Gamma</b>	-2,191	8,405E-07	4,185E-05
<b>DSG1</b>	<b>Desmoglein 1</b>	-2,175	7,577E-04	6,021E-03
<b>LCE3D</b>	<b>Late Cornified Envelope 3D</b>	-2,159	2,019E-04	2,162E-03
<b>MMP3</b>	<b>Matrix metallopeptidase 3</b>	-2,131	6,223E-05	8,873E-04
<b>KRT16P6</b>	<b>Keratin 16 Pseudogene 6</b>	-2,083	6,677E-05	9,377E-04
<b>CRCT1</b>	<b>Cysteine Rich C-Terminal 1</b>	-2,056	3,821E-04	3,541E-03
<b>PAEP</b>	<b>Progesterone Associated Endometrial Protein</b>	-2,054	9,951E-07	4,628E-05

\*Fold Change estimates refer to the HPV+ group

A ROC curve analysis was performed, resulting in a total of 1,816 genes with an AUC>0.7, considered as good biological markers.

Out of the 1,816 genes with AUC >0.7, 361 were differential genes with FDR <0.01, with 350 overexpressed genes and 11 underexpressed genes. The distribution of over- and underexpressed genes with AUC >0.7 can be observed in the figure below (Figure 1).



**Figure 1.** Venn Diagram in which one can observe the distribution of the DEGs with AUC >0,7.

An analysis was conducted using GSEA software to visualize the biological functions of all genes with changes in expression. There are 1,731 biological processes with FDR <0.01 that are activated (positive LOR) and 26 that are inhibited (negative LOR) in HPV positive samples (Table 5, annex). Highlighted processes relate to epithelial tissue integrity and the extracellular matrix, among others: keratinocyte differentiation, keratinization, cornification, skin development, epidermal cell differentiation, extracellular matrix organization and disassembly, hemidesmosome assembly, collagen metabolic and catabolic process, extracellular structure organization, epidermis development, DNA recombination, DNA replication, DNA-dependent DNA replication, chromosome segregation, covalent chromatin modification, histone modification, various RNA splicing processes, as well as mRNA processing.

An enrichment analysis was performed using DAVID software by inserting the 5,203 overexpressed genes with FDR <0.01 and the 752 underexpressed genes. 98 enriched biological processes with FDR <0.01 and 11 inhibited biological processes were obtained with a p-value <0.01 for the underexpressed genes. We determined the following biological processes to be of interest to our study, Cell cycle, cell division, chromatin remodeling, chromatin organization, as well as the inhibited processes of keratinization, keratinocyte differentiation and epidermis development.

We analyzed which genes involved in these processes had an AUC >0.7 in our ROC analysis to determine the genes to focus our discussion on. Among the genes involved in these processes, EZH2, HDAC10, and KDM5A demonstrated the highest AUC values; 0.788, 0.777, and 0.776 respectively. All of these are involved in either chromatin organization and/or chromatin remodeling. We also highlight BIRC6 as the gene with both involvement in the cell cycle as well as cell division with the highest AUC. The underexpressed DEG SPRR2G was shown to be involved in all three of our inhibited processes of interest, for which it was also discussed in further depth in our discussion.

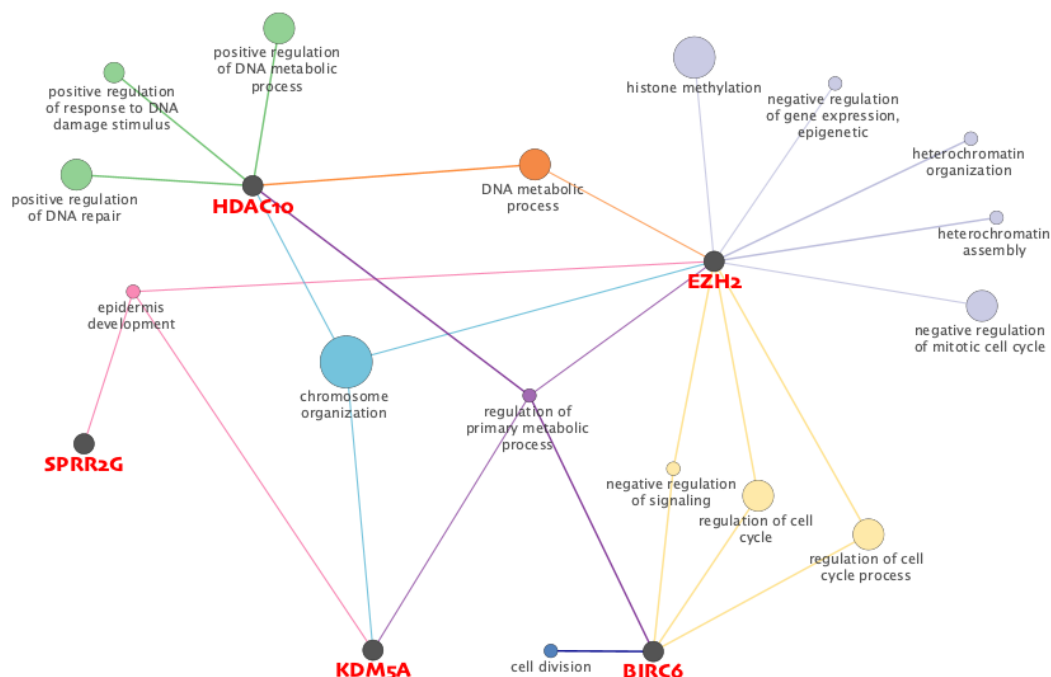
In the following table we have summarized these 5 genes for a better overview of their Fold change, FDR, and AUC values (Table 3).

**Table 3.** The 5 DEGS (FDR<0,01) of interest, due to their involvement in biological processes of interest, such as chromatin remodeling and organization, cell cycle and division, as well as keratinization, epidermis development and keratinocyte differentiation.

ID	Nombre completo	Fold Change (log2)*	FDR	AUC
<b>EZH2</b>	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit	0,715	6,128E-04	0,788
<b>HDAC10</b>	Histone Deacetylase 10	0,785	7,843E-03	0,777
<b>KDM5A</b>	Lysine Demethylase 5A	0,468	3,921E-03	0,776
<b>BIRC6</b>	Baculoviral IAP Repeat Containing 6	0,510	4,445E-03	0,748
<b>SPRR2G</b>	Small Proline Rich Protein 2G	-3,091	5,308E-05	0,707

\*Fold change estimates refer to the HNSCC-HPV+ group.

For a graphical visualization of the biological functions of the chosen genes, we used the Cytoscape program along with the ClueGO plug-in. This figure provides an overall visualization of the selected genes and aids in understanding their functions. It can be visualized below (Figure 2).

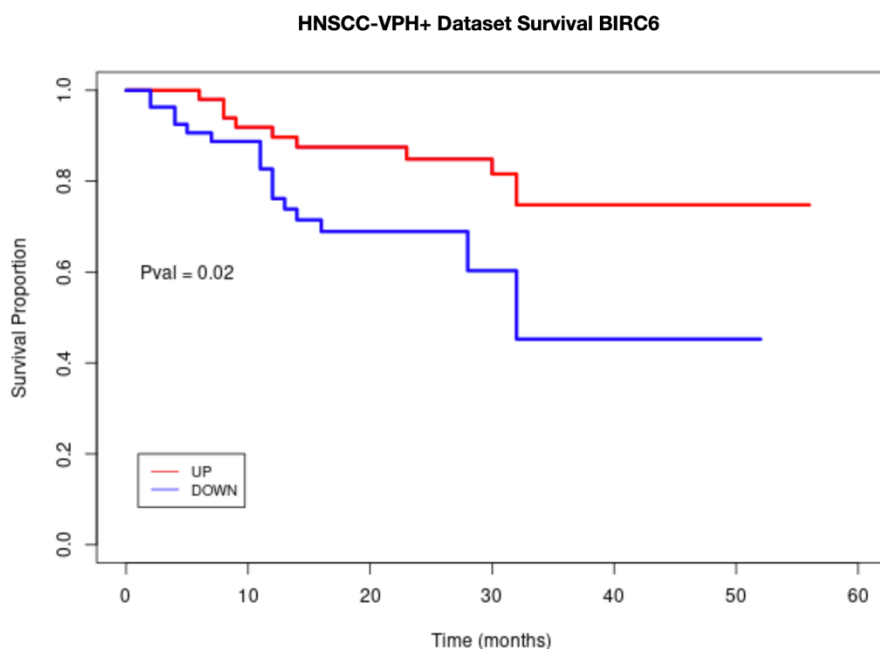


**Figure 2.** Cytoscape visualization of the genes identified to be of interest for our study with biological processes related to them.

We conducted a survival analysis for all the genes discussed, both globally across the 544 HNSCC samples and more specifically in the HPV-positive and HPV-negative samples and applied a Fisher statistical test to validate significance. In accordance with the

differential expression analyzed for each gene in our differential analysis, all the discussed genes demonstrated a trend of a positive effect on survival in the HNSCC-HPV+ group.

Below you can appreciate the survival analysis results of BIRC6, the only gene discussed with a significant impact on survival (p-value = 0,02) (Figure 3).



**Figure 3.** Survival analysis results of the gene BIRC6 in the HNSCC-HPV+ group. The blue line represents the infraexpression, whereas the red line represents the overexpression.

## Discussion:

### EZH2

EZH2 is a gene with influence over processes related to gene expression, immune system modulation and immunological check-point expression, due to its involvement in the trimethylation of Lys 27 of histone 3 (H3K27me3) and the interaction and activation of transcription factors STAT3 and NFκB [15, 16]. The overexpression is linked to cell proliferation, an increased invasive and migration capacity as well as a decrease in apoptosis [16].

In relationship to HPV, it has been found that the protein E7 of the papilloma virus has causes the liberation of E2F, which in turn activates EZH2 [17].

In our study EZH2 showed an overexpression in HNSCC-HPV+ patients. This is in accordance with the results of the study of Hyland et al. in which EZH2 was overexpressed in HPV+ dysplastic and cancerous cervical lesions [17].

Other cancers in which EZH2 has shown an overexpression are cancers of the prostate, stomach, esophagus, breast, thyroid and the nasopharynx [15].

It has also been shown that the overexpression is linked to a negative outcome in oral squamous cell carcinoma [19]. This is contrary to our survival analysis results in which



the overexpression in our HNSCC-HPV+ subset influenced the general survival positively.

Two inhibitors of EZH2 that have shown potential have been EPZ-6438 and UNC1999 [17].

Although the link of EZH2 to tumorigenesis is well established, and the link to HPV+ could be given through the interplay between E2F liberation caused by E7 of HPV, it is necessary to further study the impact this gene's overexpression has on survival of patients, to identify HPV+ patients at risk of developing a more lethal tumor.

### **HDAC10**

HDAC10 is a gene which resulted overexpressed in our study of HNSCC-HPV+ patients. This gene's main function involves the deacetylation of histones which has effects on transcriptional processes [20]. It is also involved in DNA repair, cellular differentiation, and proliferation, as well as the regulation of the cell cycle and autophagic processes [21]. Although the overexpression of HDAC10 has been linked to several cancers, and has shown to increment the proliferation rates, its underexpression has also demonstrated a link to cancers and an increased migration and invasion capacity [20, 22]. The prognosis likewise has shown both a negative and positive outcome in cases of overexpression [20, 21].

In our study, the overexpression showed a tendency to a better overall outcome in HNSCC-HPV+ patients, although it is necessary to say, that there was a lack of data. For this reason, it would be important to further study the influence of this gene's expression on the outcomes of HNSCC-HPV+ patients.

There is a higher grade of variability in the studies encompassing the relationship between the expression of HDAC10 and cancerous processes, for which reason it would be important to further study this gene and establish its definitive role in the development of cancer.

HDAC10 influences signaling pathways such as PI3K/AKT, this increments its oncogenic potential [22]. This influence and that over other biological pathways such as STAT/JAK should be studied in more depth.

### **KDM5A**

KDM5A also showed an overexpression in HNSCC-HPV+ patients in our analysis. This gene is involved in functions related to transcriptional processes, cellular differentiation, angiogenesis, cellular mobility, proliferation, invasion, and epithelial-mesenchymal transition [23]. It has also been shown to inhibit antitumoral processes and the infiltration of T lymphocytes [24].

This gene's overexpression has been linked to cancerous processes, for example in gastric cancers, lung cancers and breast cancers. It has also been shown to suppress p53, leading to a decrease in apoptosis [25].

The effect of the overexpression of KDM5A has not been studied profoundly. In our survival analysis it was shown that HNSCC-HPV+ patients had better outcomes when KDM5A was overexpressed. It would be interesting to further study the link between

KDM5A and the prognosis of cancers, as well as to understand possible links between the overexpression and the infection with papilloma virus.

### **BIRC6**

Another one of the genes showing overexpression in HNSCC-HPV+ patients in our study is BIRC6. This gene has been shown to inhibit apoptosis, participate in mitosis and interact with various other enzymes promoting the epithelial-mesenchymal transition and giving a higher resistance to chemotherapy [26].

The overexpression has been demonstrated in colorectal cancers, epithelial ovarian cancers, as well as hepatocellular cancers, among others [27]. The overexpression was linked to a worsened survival, and the knockdown proved to diminish the proliferation rates, heighten sensibility to chemotherapy, decrease the epithelial-mesenchymal transition as well as invasive and migration capacity, and detain the cellular cycle [28-30].

The overexpression of BIRC6 has been linked strongly to cancerous processes. The prognosis in our set of HNSCC-HPV+ patients was shown to be significantly improved (p-value = 0,02) in cases of BIRC6 overexpression. This is contrary to the findings of other authors, who for example found a higher resistance to chemotherapy in lung cancer patients with overexpression of BIRC6 [30].

It would be of interest to study the link between chemotherapy resistance and BIRC6 in further studies.

There is a lack of investigation on the link between HPV and BIRC6 overexpression.

### **SPRR2G**

SPRR2G is the only gene of our discussion that was found to be underexpressed in HNSCC-HPV+ patients. This gene is involved in keratinization, keratinocyte differentiation and the development of the epidermis, as well as cornification and the differentiation of epithelial cells [31].

The relationship between this gene and cancer and HPV has not been studied in many investigations. It has been found to be overexpressed in HNSCC and this overexpression seems to be associated with the tumoral grade [31, 32]. Nevertheless, it is necessary to investigate this further.

In our survival analysis the patients with HPV+ HNSCC and an infraexpression of SPRR2G tended to have better survival outcomes than those of which with an overexpression of SPRR2G.

The only study we found that related SPRR2G expression to HPV was the study of Vliet-Gregg et al. in which the overexpression was related to a heightened expression of NFX1-123 in keratinocytes of HPV+ human foreskin [33]. This would be contrary to our results, in which SPRR2G was underexpressed in HPV+ patients.

There is a definite lack of evidence to support notions of the impact of this gene's expression on the development and prognosis of cancer and its link to HPV. To further establish its role future studies would need to investigate it further.

### **HPV and its link to cancer**

In general, HPV has been associated to 25% of HNSCC cases [34]. This makes it important to investigate the link between HPV and cancer development.

The HPV through its different proteins has effects on various molecules pertaining to several biological pathways that have been shown to have oncogenic effects when dysfunctional, including PI3K/AKT, Notch and Wnt [35].

The prognosis of patients with HPV+ tumors of the head and neck region has been shown to be improved in most studies [36-40]. This is in accordance with our survival analysis results in which all the discussed differential genes impacted the prognosis of the HPV+ patients positively at the expression found in our study.

### **Conclusion:**

With our study we were able to determine that HNSCC-HPV+ patients tend to have a large quantity of differential genes when compared to HNSCC-HPV- patients and these genes were shown to influence processes related to the cell cycle, cell division and keratinization, among others. We identified the following genes as genes of interest, due to their potential as biomarkers ( $AUC > 0,7$ ) and their involvement in processes which have oncogenic effects when dysregulated such as the cell cycle, cell division, chromatin remodeling, chromatin organization, as well as keratinization, keratinocyte differentiation and epidermis development; EZH2, HDAC10, KDM5A, BIRC6 and SPRR2G.

In accordance with the expression levels found for these genes in our study, all of them influenced the survival of the HNSCC-HPV+ patients positively, although HDAC10 had a lack of data. BIRC6 did so significantly ( $p\text{-value} = 0,02$ ). This supports the notion that HPV+ tumors have an improved prognosis when compared to HPV- tumors.

Of our 5 discussed genes EZH2 and BIRC6 have the most available evidence supporting their role in the development of cancer, so these could be promising biomarkers to help identify those HPV+ patients at heightened risk of developing cancer. EZH2 expression also seems to have a strong link to the papilloma virus, due to the interplay between E6, E2F and EZH2.

## Bibliography

1. Vigneswaran N, Williams MD. Changing epidemiology of head and neck cancer. *Oral Maxillofac Surg Clin North Am.* 2014;26(2):123–41.
2. Liu C, Wang M, Zhang H, Li C, Zhang T, Liu H, et al. Tumor microenvironment and immunotherapy of oral cancer. *Eur J Med Res.* 2022;27(1):1–19. Available from: <https://doi.org/10.1186/s40001-022-00835-4>
3. Abati S, Bramati C, Bondi S, Lissoni A, Trimarchi M. Oral cancer and precancer: A narrative review on the relevance of early diagnosis. *Int J Environ Res Public Health.* 2020;17(24):1–14.
4. Gormley M, Creaney G, Schache A, Ingarfield K, Conway DI. Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. *Br Dent J.* 2022;233(9):780–6.
5. Ganesh D, Dafar A, Niklasson J, Sandberg I, Braz-Silva P, Sapkota D, et al. EZH2 Expression Correlates With T-Cell Infiltration in Oral Leukoplakia and Predicts Cancer Transformation. *Anticancer Res.* 2023;43(4):1533–42.
6. Barsouk A, Aluru JS, Rawla P, Saginala K, Barsouk A. Epidemiology, Risk Factors, and Prevention of Head and Neck Squamous Cell Carcinoma. *Med Sci.* 2023;11(2):42.
7. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research.* 2016;5:1408.
8. van Erkel AR, Pattynama PMT. ROC. *Eur J Radiol.* 1998;27:88–94.
9. Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J.* 2017;34:357–359.
10. Mootha V, Lindgren C, Eriksson KF, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34:267–273.
11. Lin R, Dai S, Irwin RD, Heinloth AN, Boorman GA, Li L. Gene set enrichment analysis for non-monotone association and multiple experimental categories. *BMC Bioinformatics.* 2008;9.
12. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research.* 2022 Mar. doi:10.1093/nar/gkac194.
13. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003 Nov;13(11):2498-504.
14. Schober P, Vetter TR. Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. *Anesth Analg.* 2018 Sep;127(3):792-798.
15. Duan R, Du W, Guo W. EZH2: A novel target for cancer treatment. *J Hematol Oncol.* 2020;13(1).
16. Kim KH, Roberts CW. Targeting EZH2 in cancer. *Nat Med.* 2016 Feb;22(2):128-34.
17. Khattri M, Amako Y, Gibbs JR, Collura JL, Arora R, Harold A, et al. Methyltransferase-independent function of enhancer of zeste homologue 2 maintains tumorigenicity induced by human oncogenic papillomavirus and polyomavirus.

- Tumour Virus Res [Internet]. 2023 Mar;16. Available from: <https://doi.org/10.1016/j.tvr.2023.200264>
18. Hyland PL, McDade SS, McCloskey R, Dickson GJ, Arthur K, McCance DJ, et al. Evidence for Alteration of EZH2, BMI1, and KDM6A and Epigenetic Reprogramming in Human Papillomavirus Type 16 E6/E7-Expressing Keratinocytes. *J Virol*. 2011;85(21):10999–1006.
  19. Zheng M, Cao MX, Luo XJ, Li L, Wang K, Wang SS, Wang HF, Tang YJ, Tang YL, Liang XH. EZH2 promotes invasion and tumour glycolysis by regulating STAT3 and FoxO1 signalling in human OSCC cells. *J Cell Mol Med*. 2019 Oct;23(10):6942-6954. doi: 10.1111/jcmm.14579. Epub 2019 Jul 31. PMID: 31368152; PMCID: PMC6787444.
  20. Gu YY, Zhou GN, Li Y, et al. HDAC10 Inhibits Cervical Cancer Progression through Downregulating the HDAC10-microRNA-223-EPB41L3 Axis. *J Oncol*. 2022;2022.
  21. Liu X, Wang Y, Zhang R, et al. HDAC10 Is Positively Associated With PD-L1 Expression and Poor Prognosis in Patients With NSCLC. *Front Oncol*. 2020;10(April):1–7.
  22. Yang Y, Huang Y, Wang Z, et al. HDAC10 promotes lung cancer proliferation via AKT phosphorylation. *Oncotarget*. 2016;7(37):59388–401.
  23. Yang GJ, Zhu MH, Lu XJ, et al. The emerging role of KDM5A in human cancer. *J Hematol Oncol* [Internet]. 2021;14(1):1–18. Available from: <https://doi.org/10.1186/s13045-021-01041-1>
  24. Liu H, Lin J, Zhou W, et al. KDM5A Inhibits Antitumor Immune Responses Through Downregulation of the Antigen-Presentation Pathway in Ovarian Cancer. *Cancer Immunol Res*. 2022 Aug 3;10(8):1028-1038. doi: 10.1158/2326-6066.CIR-22-0088. PMID: 35726891; PMCID: PMC9357105.
  25. Kataria A, Tyagi S. Domain architecture and protein–protein interactions regulate KDM5A recruitment to the chromatin. *Epigenetics* [Internet]. 2023;18(1):1–19. Available from: <https://doi.org/10.1080/15592294.2023.2268813>
  26. Zhong K, Wang X, Zhang H, Chen N, Mai Y, Dai S, Yang L, Chen D, Zhong W. BIRC6 Modulates the Protein Stability of Axin to Regulate the Growth, Stemness, and Resistance of Renal Cancer Cells via the  $\beta$ -Catenin Pathway. *ACS Omega*. 2024 Feb 7;9(7):7782-7792. doi: 10.1021/acsomega.3c07265. PMID: 38405482; PMCID: PMC10882609.
  27. Zhuang W, Zhang C, Hao F, Sun X. Baculoviral IAP Repeat Containing 6 (BIRC6) Is a Predictor of Prognosis in Prostate Cancer. *Med Sci Monit*. 2018 Feb 10;24:839-845. doi: 10.12659/MSM.904052. PMID: 29429983; PMCID: PMC5816567.
  28. Zhou L, Wang B, Zhang Y, Yao K, Liu B. Silencing circ BIRC6 inhibits the proliferation, invasion, migration and epithelial mesenchymal transition of bladder cancer cells by targeting the miR 495 3p/XBP1 signaling axis. *Mol Med Rep*. 2021 Nov;24(5):811. doi: 10.3892/mmr.2021.12451. Epub 2021 Sep 20. PMID: 34542161; PMCID: PMC8477182.
  29. Hu T, Weng S, Tang W, Xue R, Chen S, Cai G, Cai Y, Shen X, Zhang S, Dong L. Overexpression of BIRC6 Is a Predictor of Prognosis for Colorectal Cancer. *PLoS*

- One. 2015 May 1;10(5):e0125281. doi: 10.1371/journal.pone.0125281. PMID: 25933218; PMCID: PMC4416929.
30. Dong X, Lin D, Low C, Vucic EA, English JC, Yee J, Murray N, Lam WL, Ling V, Lam S, Gout PW, Wang Y. Elevated expression of BIRC6 protein in non-small-cell lung cancers is associated with cancer recurrence and chemoresistance. *J Thorac Oncol.* 2013 Feb;8(2):161-70. doi: 10.1097/JTO.0b013e31827d5237. PMID: 23287853.
  31. Rakha EA, Alsaleem M, ElSharawy KA, Toss MS, Raafat S, Mihai R, Minhas FA, Green AR, Rajpoot NM, Dalton LW, Mongan NP. Visual histological assessment of morphological features reflects the underlying molecular profile in invasive breast cancer: a morphomolecular study. *Histopathology.* 2020 Oct;77(4):631-645. doi: 10.1111/his.14199. Epub 2020 Sep 1. PMID: 32618014.
  32. Chintala S, Quist KM, Gonzalez-DeWhitt PA, Katzenellenbogen RA. High expression of NFX1-123 in HPV positive head and neck squamous cell carcinomas. *Head Neck.* 2022 Jan;44(1):177-188. doi: 10.1002/hed.26906. Epub 2021 Oct 25. PMID: 34693597; PMCID: PMC8688290.
  33. Vliet-Gregg PA, Hamilton JR, Katzenellenbogen RA. NFX1-123 and human papillomavirus 16E6 increase Notch expression in keratinocytes. *J Virol.* 2013 Dec;87(24):13741-50. doi: 10.1128/JVI.02582-13. Epub 2013 Oct 9. PMID: 24109236; PMCID: PMC3838236.
  34. Yeo-Teh NSL, Ito Y, Jha S. High-Risk Human Papillomaviral Oncogenes E6 and E7 Target Key Cellular Pathways to Achieve Oncogenesis. *Int J Mol Sci.* 2018 Jun 8;19(6):1706. doi: 10.3390/ijms19061706. PMID: 29890655; PMCID: PMC6032416.
  35. Bhattacharjee R, Das SS, Biswal SS, Nath A, Das D, Basu A, Malik S, Kumar L, Kar S, Singh SK, Upadhye VJ, Iqbal D, Almojam S, Roychoudhury S, Ojha S, Ruokolainen J, Jha NK, Kesari KK. Mechanistic role of HPV-associated early proteins in cervical cancer: Molecular pathways and targeted therapeutic strategies. *Crit Rev Oncol Hematol.* 2022 Jun;174:103675. doi: 10.1016/j.critrevonc.2022.103675. Epub 2022 Apr 4. PMID: 35381343.
  36. Göttgens EL, Ostheimer C, Span PN, Bussink J, Hammond EM. HPV, hypoxia and radiation response in head and neck cancer. *Br J Radiol.* 2019 Jan;92(1093):20180047. doi: 10.1259/bjr.20180047. Epub 2018 Mar 14. PMID: 29493265; PMCID: PMC6435089.
  37. Burbure N, Handorf E, Ridge JA, Bauman J, Liu JC, Giri A, Galloway TJ. Prognostic significance of human papillomavirus status and treatment modality in hypopharyngeal cancer. *Head Neck.* 2021 Oct;43(10):3042-3052. doi: 10.1002/hed.26793. Epub 2021 Jun 24. PMID: 34165223; PMCID: PMC9631108.
  38. Tong F, Geng J, Yan B, Lou H, Chen X, Duan C, He J, Zhang S, Xie H, Li H, Yuan D, Zhang F, Meng H, Wei L. Prevalence and Prognostic Significance of HPV in Laryngeal Squamous Cell Carcinoma in Northeast China. *Cell Physiol Biochem.* 2018;49(1):206-216. doi: 10.1159/000492858. Epub 2018 Aug 22. PMID: 30134232.
  39. Axelsson, L., Nyman, J., Haugen-Cange, H. et al. Prognostic factors for head and neck cancer of unknown primary including the impact of human papilloma virus

infection. J of Otolaryngol - Head & Neck Surg 46, 45 (2017).  
<https://doi.org/10.1186/s40463-017-0223-1>

40. Benson E, Li R, Eisele D, Fakhry C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. Oral Oncol. 2014 Jun;50(6):565-74. doi: 10.1016/j.oraloncology.2013.09.008. Epub 2013 Oct 14. PMID: 24134947; PMCID: PMC4391706.