



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**  
FACULTAD DE CIENCIAS E INGENIERÍA

**Caracterización de datos transcriptómicos de glutaminasas (*GLS1* y *GLS2*) en sujetos con cáncer de hígado, cérvix y cabeza y cuello, infectados con HBV/HCV y HPV y su asociación con la sobrevida global.**

Tesis para optar por el título profesional de Licenciado en Biología

Autor:

B.Sc. Livisu Pajares Rojas

Asesora:

PhD. MSc. Claudia Ines Gloria Machicado Rivero

Lima, Perú

2024

## **Agradecimientos**

*A mis padres Libertad Rojas y Ramón por su apoyo durante la carrera.*

*A la Dra. Claudia Machicado por su apoyo y enseñanzas durante el  
desarrollo de la tesis.*

## Caracterización de datos transcriptómicos de glutaminasas (GLS1 y GLS2) en sujetos con cáncer de hígado, cérvix y cabeza y cuello, infectados con HBV/HCV y HPV y su asociación con la sobrevida global

### INFORME DE ORIGINALIDAD

9%	8%	4%	2%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

### FUENTES PRIMARIAS

1	<a href="http://alicia.concytec.gob.pe">alicia.concytec.gob.pe</a> Fuente de Internet	1%
2	<a href="http://hdl.handle.net">hdl.handle.net</a> Fuente de Internet	1%
3	<a href="http://assets.researchsquare.com">assets.researchsquare.com</a> Fuente de Internet	1%
4	<a href="http://www.nature.com">www.nature.com</a> Fuente de Internet	1%
5	Livisu Pajares Rojas, Claudia Machicado Rivero. "Abstract 883: Glutaminases expression and viral infection as potential prognostic factors in cervical, head and neck and liver cancers", Cancer Research, 2023 Publicación	1%
6	<a href="http://1library.co">1library.co</a> Fuente de Internet	<1%

# Índice

<b>Resumen .....</b>	<b>1</b>
<b>Abstract .....</b>	<b>2</b>
<b>Abreviaturas .....</b>	<b>3</b>
<b>1. Introducción .....</b>	<b>8</b>
1.1. Metabolismo de la glutamina en las células cancerosas.....	8
1.2. Rol de las glutaminasas en el pronóstico del cáncer.....	9
1.3. Glutaminasas en los procesos de infección.....	9
1.4. Transformación celular promovida por infección viral. ....	10
1.5. Problema de investigación. ....	11
<b>2. Preguntas de Investigación .....</b>	<b>12</b>
<b>3. Objetivos .....</b>	<b>12</b>
3.1. Objetivo General.....	12
3.2. Objetivos Específicos .....	12
<b>4. Materiales y Métodos .....</b>	<b>13</b>
4.1. Diseño del estudio .....	13
4.2. Población y muestra de estudio .....	13
4.3. Criterios de selección de los participantes .....	15
4.3.1. Criterios de inclusión: .....	15
4.3.2. Criterios de exclusión.....	15
4.4. Variables .....	15
4.4.1 Variables Fenotípicas .....	15
4.4.1.1. Variables demográficas/descriptivas .....	15
4.4.1.2. Variables clínicas .....	16
4.4.2. Datos genómicos y transcriptómicos.....	18
4.5. Procesamiento de las bases de datos.....	19
4.5.1 Cálculo del tamaño muestral .....	22
4.5.2 Estatus de infección de la muestra .....	22
4.5.3 Segmento del número de copias .....	23
4.5.4 Expresión génica RNAseq – HTSeq-FPKM .....	24
4.6 Análisis de factores de datos mixtos (FAMD).....	26
4.7. Análisis estadísticos .....	26

4.7.1	Cálculo del Log <sub>2</sub> FC.....	28
4.7.2	Análisis de sobrevida KM.....	30
4.7.3	Cálculo de porcentajes de cada subtipo viral.....	33
4.8	Herramientas de análisis estadístico.....	33
4.10	Cuestiones éticas.....	34
<b>5.</b>	<b>Resultados.....</b>	<b>34</b>
5.1	Población de estudio.....	34
5.2	FAMD.....	35
5.2.1.	Cáncer cervical (CESC).....	35
5.2.2.	Cáncer de cabeza y cuello (HNSC).....	36
5.2.3.	Cáncer de hígado (LIHC).....	38
5.3.	Cambio de la expresión génica y la alteración del número de copias de <i>GLS1</i> y <i>GLS2</i> en sujetos con LIHC, CESC y HNSCC.....	39
5.3.1.	Cáncer cervical (CESC).....	39
5.3.2.	Cáncer de cabeza y cuello (HNSCC).....	40
5.3.3.	Cáncer de hígado (LIHC).....	41
5.4.	Relación entre la expresión de <i>GLS1/GLS2</i> y variables clínicas.....	42
5.4.1.	Cáncer cervical.....	42
5.4.2.	Cáncer de cabeza y cuello.....	45
5.4.3.	Cáncer de hígado.....	47
5.5.	Influencia de la expresión génica de <i>GLS1/GLS2</i> y del estado de la infección viral (positivo/negativo) con la sobrevida global.....	49
5.5.1.	Cáncer cervical.....	49
5.5.2.	Cáncer de cabeza y cuello.....	50
5.5.3.	Cáncer de hígado.....	51
<b>6.</b>	<b>Discusión.....</b>	<b>52</b>
<b>7.</b>	<b>Conclusiones.....</b>	<b>64</b>
<b>8.</b>	<b>Limitaciones.....</b>	<b>65</b>
<b>9.</b>	<b>Recomendaciones.....</b>	<b>65</b>
<b>10.</b>	<b>Referencias Bibliográficas.....</b>	<b>66</b>
<b>11.</b>	<b>Anexos.....</b>	<b>79</b>

## Resumen

Las infecciones crónicas por virus oncogénicos, como el virus del papiloma humano (HPV) y los virus de la hepatitis B (HBV) y C (HCV), causan distintos tipos de neoplasias incluidos el cáncer cervical (CESC), de cabeza y cuello (HNSCC), y de hígado (LIHC). La glutaminólisis es un proceso regulado por la glutaminasa 1 y 2 (GLS1/GLS2), las cuales están alteradas en HNSCC y LIHC, pero su estado es desconocido en CESC. El objetivo de esta tesis es determinar si la alteración de la expresión génica de *GLS1/GLS2* varía con el estado de infección de HPV, HBV y HCV, y si, a su vez, la expresión y el estado de infección están asociados al pronóstico de CESC, HNSCC y LIHC. Para ello analizamos datos genómicos, transcriptómicos y metadata de cohortes provenientes de *The Cancer Genome Atlas* (TCGA). La expresión de *GLS1* y *GLS2* se determinó en las cohortes de CESC, HNSCC y LIHC calculando el Log<sub>2</sub>FC y aplicando t-test de Welch y ANOVA de una vía. Finalmente, utilizamos gráficos de Kaplan-Meier (KM) para averiguar la asociación de la data genómica y clínica con la sobrevida global (OS). Encontramos que *GLS2* está subexpresado en las muestras de LIHC-HBV/HCV(+) comparadas con las muestras no tumorales con presencia de HBV/HCV(+). Así mismo, la expresión de *GLS2* estaba asociada con la OS en pacientes con CESC. En conclusión, la expresión de glutaminasas es anormal en tejidos de CESC, HNSCC y LIHC, y a su vez, depende del estado de infección viral. Como la expresión de *GLS2* e infección viral afectan la OS en los tumores analizados, esta tesis nos muestra a las glutaminasas como marcadores candidatos de pronóstico para CESC, HNSCC y LIHC, los cuales podrían ser validados en estudios futuros.

Palabras claves: sobrevida global, glutaminasas, HPV, HBV, HCV, carcinoma cervical de células escamosas, carcinoma de células escamosas de cabeza y cuello, carcinoma hepatocelular de hígado

## Abstract

Chronic infections with oncogenic viruses, such as human papillomavirus (HPV) and hepatitis B (HBV) and C (HCV) viruses, cause various types of neoplasms including cervical (CESC), head and neck (HNSCC), and liver (LIHC) cancers. Glutaminolysis is a process regulated by glutaminase 1 and 2 (GLS1/GLS2), which are altered in HNSCC and LIHC, but their status is unknown in CESC. The aim of this thesis is to determine whether altered *GLS1/GLS2* gene expression varies with HPV, HBV, and HCV infection status, and whether expression and infection status are associated with prognosis in CESC, HNSCC and LIHC. For this purpose, we analyzed genomic, transcriptomic and metadata data from The Cancer Genome Atlas (TCGA). *GLS1* and *GLS2* expression was determined in CESC, HNSCC and LIHC cohorts by calculating Log<sub>2</sub>FC and applying Welch's t-test and one-way ANOVA. Finally, we used Kaplan-Meier (KM) plots to investigate the association of genomic and clinical data with the overall survival (OS). We found that *GLS2* is under-expressed in LIHC-HBV/HCV(+) compared with non-tumorous samples infected with HBV/HCV(+). Likewise, *GLS2* expression was associated with OS in CESC patients. In conclusion, glutaminase expression is abnormal in CESC, HNSCC and LIHC tissues, and in turn, depends on viral infection status. As *GLS2* expression and viral infection affect OS in the analyzed tumors, this thesis shows glutaminases as candidate prognostic markers for CESC, HNSCC and LIHC, which could be validated in future studies.

Key words: Overall survival, glutaminases, HPV, HBC, HCV, Cervical Squamous Cell Carcinoma, Head and Neck Squamous Cell Carcinoma, Liver Hepatocellular Carcinoma

## Abreviaturas

<b>Abreviatura</b>	<b>Significado</b>
CESC:	Carcinoma cervical de células escamosas y adenocarcinoma endocervical ( <i>Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma</i> )
HNSCC:	Carcinoma de células escamosas de cabeza y cuello ( <i>Head and Neck Squamous Cell Carcinoma</i> )
LIHC:	Carcinoma hepatocelular de hígado ( <i>Liver Hepatocellular Carcinoma</i> )
HPV:	Virus del papiloma humano ( <i>Human Papiloma Virus</i> )
HBV:	Virus de la hepatitis B ( <i>Hepatitis B Virus</i> )
HCV:	Virus de la Hepatitis C ( <i>Hepatitis C Virus</i> )
HCC:	Hepatocarcinoma celular ( <i>Hepatocellular Carcinoma</i> )
TCGA GDC:	Atlas del Genoma del Cáncer Datos genómicos comunes ( <i>The Cancer Genome Atlas Genomic Data Commons</i> )
OS	Sobrevida global ( <i>Overall Survival</i> )
KM	Kaplan-Meier
IMC:	Índice de masa corporal
GLS1 – GLS2:	Glutaminasa 1 – Glutaminasa 2
pRb:	Retinoblastoma
TCA:	Vía del ácido cítrico ( <i>Tricarboxylic Acid</i> )

$\alpha$ -kG:	Alfa-cetoglutarato ( <i>Alpha-Ketoglutarate</i> )
KSHV:	Virus del sarcoma de Kaposi ( <i>Kaposi's Sarcoma-Associated Herpesvirus</i> )
MLL4:	Gen linaje mixto leucemia 4 ( <i>Myeloid/lymphoid or mixed-lineage leukemia 4</i> )
FPKM:	Fragmentos por kilobase de transcrito por millón de lecturas mapeadas ( <i>Fragments per kilobase of transcript per Million mapped reads</i> )
FC:	Proporción de cambio ( <i>Fold change</i> )
PI3K/AKT	Vía del Fosfatidilinositol-3-kinasa / Proteína Serina-Treonina kinasa ( <i>Phosphoinositide-3-Kinase/ Serine-Threonine Kinase</i> )
RT-PCR	Reacción en Cadena de la Polimerasa en Tiempo Real ( <i>Real Time Polymerase Chain Reaction</i> )
RT-qPCR	Transcripción reversa seguida de una PCR cuantitativa en tiempo real ( <i>Reverse Transcription-Real-Time Quantitative Polymerase Chain Reaction</i> )
RNA-Seq	Secuenciamiento de RNA ( <i>RNA Sequencing</i> )
GTEx	El proyecto genotipo-expresión tisular ( <i>The Genotype-Tissue Expression Project</i> )
PDK	Piruvato deshidrogenasa kinasa ( <i>Pyruvate Dehydrogenase Kinase</i> )
E6-AP	Proteína ubiquitina ligasa asociada a la proteína E6 ( <i>Ubiquitin-Protein Ligase E6-Associated Protein</i> )
p16 (p16INK4A)	Inhibidor 2A dependiente de ciclina kinasa, supresor de tumores múltiples 1 ( <i>Cyclin-Dependent Kinase Inhibitor 2A, CDKN2A, Multiple Tumor Suppressor 1</i> )

EBV	Virus del Epstein-Barr ( <i>Epstein-Barr Virus</i> )
BPTES	Bis-2-(5-fenilacetamido-1,2,4-tiadiazol-2-il)sulfuro de etilo 3 ( <i>Bis-2-(5-Phenylacetamido-1,2,4-Thiadiazol-2-yl)Ethyl Sulfide 3</i> )
GLUD1/2	Glutamato Deshidrogenasa 1 / 2 ( <i>Glutamate Dehydrogenase 1 / 2</i> )
KRAS	Ki-ras2 Homólogo del oncogén viral del sarcoma de rata de Kirsten ( <i>Ki-ras2 Kirsten rat sarcoma viral oncogene homolog</i> )
PBRM1	Proteína Polibromo 1 ( <i>Protein Polybromo 1</i> )
KDM5C	Lisina desmetilasa 5C ( <i>Lysine Demethylase 5C</i> )
BAP1	Ubiquitina Carboxil-Terminal Hidrolasa ( <i>Ubiquitin Carboxyl-Terminal Hydrolase</i> )
$\beta$ -vER	Receptores de Beta Estrógenos
GSE20017	Base de datos de eventos espaciales genómicos ( <i>Genomic Spatial Event database</i> )
SAT	Tejido Adiposo Subcutáneo ( <i>Subcutaneous Adipose Tissue</i> )
SOS Sib Pair	Sujetos obesos suecos Sib Pair ( <i>Swedish Obese Subjects Sib Pair</i> )
IL-6	Interleucina 6
G1/G2/G3	Grado Histológico 1, 2, 3
qPCR	Reacción en Cadena de la Polimerasa Cuantitativa ( <i>Quantitative Polymerase Chain Reaction</i> )

HIF-1 $\alpha$	Subunidad alfa del factor 1 inducible por hipoxia ( <i>Hypoxia Inducible Factor 1 Subunit Alpha</i> )
CD147	Cúmulo de diferenciación 147 ( <i>Cluster of Differentiation 147</i> )
SNAT1	Transportador 1 de aminoácidos neutros acoplado al sodio ( <i>Sodium-Coupled Neutral Amino Acid Transporter 1</i> )
GS	Glutamina sintetasa ( <i>Glutamine Synthetase</i> )
LAT1	Transportador 1 de aminoácidos de tipo L ( <i>L-Type Amino Acid Transporter 1</i> )
ASCT2	Transportador 2 de alanina, serina y cisteína ( <i>Alanine Serine Cysteine Transporter 2</i> )
xCT	Antiportador de cistina/glutamato ( <i>Cystine/Glutamate Antiporter</i> )
RSEM	RNA-seq por maximización de expectación ( <i>RNA-Seq by Expectation-Maximization</i> )
HT-Seq	Secuenciación masiva (High Throughput sequencing)
CNV	Variación del número de copias ( <i>Copy Number Variation</i> )
NGS	Secuenciación de Nueva Generación ( <i>Next Generation Sequencing</i> )
GRCh38/hg38	Consorcio de referencia del genoma humano constructo 38 ( <i>Genome Reference Consortium Human Build 38</i> )
CBS	Análisis de segmentación binaria circular ( <i>Circular binary segmentation analysis</i> )
STAR	Alineación de transcritos empalmados con una referencia ( <i>Spliced Transcripts Alignment to a Reference</i> )

FAMD	Análisis de factores de datos mixtos ( <i>Factor Analysis of Mixed Data</i> )
PCA	Análisis de componentes principales ( <i>Principal Component Analysis</i> )
MCA	Análisis de correspondencia múltiple ( <i>Multiple Correspondence Analysis</i> )
ANOVA	Análisis de varianza ( <i>Analysis of Variance</i> )
Q1	Cuartil Inferior
Q3	Cuartil Superior
PS	Probabilidad de Sobrevida
BPS	Bajada de la probabilidad de sobrevida
PR	Nº de pacientes en riesgo
E	Evento

## 1. Introducción

### 1.1. Metabolismo de la glutamina en las células cancerosas.

Las células cancerosas contienen una amplia variedad de alteraciones, incluyendo ajustes en su metabolismo energético (1). La modificación de las vías metabólicas descrito en las células cancerosas ocurre para satisfacer sus altas demandas metabólicas, que le permiten entre otras cosas, dividirse continuamente. El ajuste metabólico en el tejido canceroso favorece la oxidación incompleta de la glucosa a través de la vía glicolítica (2). Bajo estas condiciones, el piruvato es convertido a lactato, en lugar de entrar a la vía del ácido cítrico (TCA), incluso en presencia de oxígeno (3,4). Como consecuencia de ello, se reducen los niveles de ATP mitocondrial en células tumorales (5).

Para obtener precursores biosintéticos que ayuden a la proliferación celular, en algunos tipos de cáncer se promueve el catabolismo de glutamina (glutaminólisis), lo cual contrarresta la baja tasa de metabolismo oxidativo (6). La glutaminólisis consiste en la conversión de la glutamina a glutamato, por acción de la enzima glutaminasa (GLS1 o GLS2), para luego formar alfa-cetoglutarato ( $\alpha$ -kG). Así, el  $\alpha$ -kG alimenta la vía de TCA para poder obtener intermediarios metabólicos (7). Estas reacciones bioquímicas se logran por cambios en la expresión de glutaminasas (*GLS1* y *GLS2*) y otras enzimas reguladoras del metabolismo de glutamina (8). Con ello, la glutaminólisis junto con la glucólisis son vías que aportan energía a las células cancerosas.

Por lo expuesto antes, es de interés conocer cómo la sobreexpresión de *GLS* controla la alta demanda metabólica, aumentando la síntesis de ATP. Inclusive, en función del tipo de cáncer, existe heterogeneidad en la expresión génica de glutaminasas (9) . Por ejemplo, en cáncer pulmonar, *GLS1* subexpresado (10) mientras que en HCC, *GLS1* está sobreexpresado y *GLS2* está subexpresado, en relación a tejido sano (11).

## **1.2. Rol de las glutaminasas en el pronóstico del cáncer**

El rol regulador de glutaminasas también es diferente entre ellas. GLS1 está regulado por oncogenes y contribuye al crecimiento tumoral en gran medida (12,13). En contraste, GLS2 tiene propiedades supresoras de tumores y limita la malignidad de los tumores malignos (14,15) . Por sus efectos, se está considerando a ambas enzimas como marcadores del pronóstico de la enfermedad. Un estudio de análisis multi ómico bioinformático encontró que GLS1 y GLS2 cumplen roles tanto oncogénicos como antioncogénicos, los cuales dependen del tipo de cáncer. En dicha investigación, se determinó que *GLS1* estaba sobreexpresado en el cáncer de hígado (“LIHC” por sus siglas en inglés) y el de cabeza y cuello (“HNSCC” por sus siglas en inglés), lo cual estaba asociado a un pobre pronóstico de la enfermedad. También, *GLS2* estaba sobreexpresado en HNSCC, mientras que está subexpresado en LIHC (16). Otro estudio determinó que *GLS1* es un predictor importante para la supervivencia de los pacientes con LIHC (17).

## **1.3. Glutaminasas en los procesos de infección.**

Además de su potencial rol predictor en cáncer, ha quedado demostrado que las glutaminasas 1 y 2 son enzimas críticas para la replicación de algunos virus. Por ejemplo, la inhibición de GLS1 disminuye la replicación *in vitro* de adenovirus, herpes simplex 1 e influenza A (18). Así mismo, la infección por virus del herpes asociado al sarcoma de Kaposi (KSHV) induce la expresión de *GLS1*, sugiriendo la importancia de la glutamina en el proceso infeccioso (19). En este sentido, se ha demostrado el rol de la glutamina en la proliferación de células transformadas por KSHV, mediante la sobreexpresión de enzimas claves, como GLS2 (20). Así mismo, el virus de la hepatitis C (HCV) induce la glutaminólisis, lo que promueve la transcripción de *GLS1*, favoreciendo así la replicación viral y predisponiendo a la transformación celular a los hepatocitos infectados (21). También se encontró que en tumores hepáticos inducidos por infección crónica de virus de la hepatitis B (HBV), la actividad enzimática del GLS1 aumenta, sugiriendo la inclusión de la

glutamina como una fuente de carbono adicional para sostener el requerimiento metabólico en tumores malignos (22).

#### **1.4. Transformación celular promovida por infección viral.**

El proceso de transformación celular inducida por la infección viral crónica estimula la proliferación celular y no daña, en primera instancia, el material genético. Esta situación motiva a romper el equilibrio entre la apoptosis y la proliferación celular (23). Por ejemplo, las infecciones crónicas por HPV conllevan a la expresión de oncogenes que inactivan los genes supresores de tumor p53 y retinoblastoma (Rb), lo cual a la larga conduce a la acumulación de mutaciones en el ADN, condición conocida como inestabilidad genómica (24,25). Algunas proteínas virales de HCV, como E2, también activan oncogenes mientras que otras como NS3 inactivan genes supresores de tumores y desregulan vías de transducción de señales (26).

La inserción del material genético viral en las células huésped es un evento raro en HBV y no ocurre en HCV (26), pero es muy frecuente en HPV, afectando la expresión génica y el número de copias de determinados genes (25). Así, en HBV se fusiona el gen quimérico X (HBVx) y el gen linaje mixto leucemia 4 (MLL4). Esto lleva a una translocación del intrón 3 del MLL4 a otra región del genoma de los hepatocitos (27). En HPV, la integración del genoma viral en el genoma humano interrumpe el gen viral E2, lo cual promueve la sobreexpresión de los oncogenes E6/E7, necesarios para el mantenimiento del fenotipo celular transformado (24,25).

Desde el punto de vista metabólico, el cáncer asociado con infección viral se diferenciaría de aquel no asociado con infección. Ello fue revelado en un estudio de HNSCC, donde se investigaron las diferencias metabólicas entre neoplasias con antecedentes de infección crónica por HPV y otras sin antecedentes. En tejidos neoplásicos HPV(-) negativo, los genes asociados con la glicólisis y la oxidación fosforilativa tenían una elevada expresión, en comparación con los tejidos HPV(+) positivo (28). En particular, *GLS1* tiene una expresión significativamente elevada en

HNSCC, lo cual se asoció con un reducido tiempo libre de enfermedad después de la terapia. Sin embargo, no se indagó si la presencia de la infección de HPV afecta a la expresión de *GLS1* (29).

### **1.5. Problema de investigación.**

Hasta ahora, no se ha determinado si hay una relación entre la expresión diferencial de *GLS1/GLS2* en tejidos de cáncer cervical (“CESC” por sus siglas en inglés), HNSCC y LIHC provenientes de pacientes humanos, con el estado de infección (positivo o negativo) de los virus oncogénicos HPV, HBV y HCV. Este conocimiento podría ser obtenido a través del análisis secundario de datos genómicos de pacientes con cáncer. El aprovechamiento de dichos datos genómicos del cáncer ha sido posible gracias a proyectos internacionales como *The Cancer Genome Atlas* (TCGA), el cual logró la caracterización genómica de varios tipos de cánceres a través del secuenciamiento masivo de muestras biológicas (30–32). La información obtenida de TCGA se ha puesto a disposición de la comunidad científica desde el 2010 y ha sido aprovechada por diversos estudios interesados en indagar diversos problemas vinculados con el cáncer, lo cual demuestra que estudios dirigidos por datos (en inglés “*Data-driven research*”) son viables (33–37).

Teniendo en cuenta todo lo expuesto, nos preguntamos si el patrón de expresión génica de *GLS1* y *GLS2* cambia con la presencia de infección viral por HPV, HBV y HCV en sujetos diagnosticados con cáncer de hígado, cáncer cervical y cabeza y cuello. También nos preguntamos si la expresión génica de *GLS1/GLS2* y la presencia de infección por HPV, HBV y HCV están asociados con la sobrevida global de las neoplasias anteriormente mencionadas.

Esta investigación es relevante ya que el HBV y HPV son agentes etiológicos frecuentes del cáncer de hígado y cáncer cervical, respectivamente, en el Perú. De existir una correlación entre la infección viral y la sobrevida de los pacientes, ello podría ser útil para mejorar las políticas públicas de prevención de HBV/HCV y HPV

que busquen disminuir la incidencia de dichas infecciones en el Perú y el mundo. Además, la expresión de glutaminasas podría ser evaluada como un marcador de pronóstico en futuros estudios.

## **2. Preguntas de Investigación**

- ¿Cuál es el patrón de expresión génica de glutaminasas (*GLS1* y *GLS2*) en pacientes con cáncer de hígado, cérvix y cabeza y cuello que presentan infección viral por HPV, HBV y HCV al momento del diagnóstico del cáncer?
- ¿Se asocia el patrón de expresión génica de *GLS1/GLS2* y la presencia de infección viral por HPV, HBV y HCV con la sobrevida global?

## **3. Objetivos**

### **3.1. Objetivo General**

Determinar la influencia de la presencia de infección viral por HPV, HBV y HCV sobre la expresión génica de *GLS1/GLS2* y su asociación con la sobrevida global mediante un análisis de datos transcriptómicos derivados de las cohortes de TCGA de cáncer de hígado, cérvix y cabeza y cuello.

### **3.2. Objetivos Específicos**

- Conocer la expresión génica y la alteración del número de copias de *GLS1* y *GLS2* en sujetos con cáncer de hígado, cérvix y cabeza y cuello que poseen resultados de pruebas moleculares de infección viral (HPV, HCV, HBV).
- Determinar el patrón de la expresión génica o la asociación con la alteración del número de copias de *GLS1/GLS2* con la presencia de infección viral por HPV, HBV y HCV en cáncer de hígado, cérvix y cabeza y cuello.
- Indagar si la expresión génica de *GLS1/GLS2*, así como la presencia de infección viral por HPV, HBV y HCV están asociadas con la sobrevida global en cáncer hígado, cérvix, de cabeza y cuello.

## 4. Materiales y Métodos

### 4.1. Diseño del estudio

El presente estudio es correlacional, retrospectivo, sin hipótesis, analítico de cohortes, basado en un análisis secundario de datos públicos.

### 4.2. Población y muestra de estudio

Se consideraron pacientes diagnosticados con CESC, LIHC y HNSCC, participantes de los proyectos TCGA (30–32), cuyos datos clínicos y genómicos se encontraron disponibles en los repositorios de datos UCSC Xena (37). En la Tabla 1 se muestran los enlaces originales de las bases de datos descargados del UCSC Xena Browser. Los datos considerados para los análisis realizados en esta tesis se encuentra en el Anexo 1.

*Tabla 1. Enlaces de los datos descargados en UCSC Xena Browser.*

Tipo de dato	Cohorte	Enlace
Segmento del Número de Copias	CEC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-CEC.cnv.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-CEC.cnv.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443</a>
Expresión Génica RNAseq HTSeq - FPKM	CEC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-CEC.htseq_fpkm.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-CEC.htseq_fpkm.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443</a>
Fenotipo	CEC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-CEC.GDC_phenotype.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-CEC.GDC_phenotype.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443</a>
Datos de Sobrevida	CEC	<a href="https://xenabrowser.net/datap">https://xenabrowser.net/datap</a>

		<a href="https://xenabrowser.net/datapages/?dataset=TCGA-CESC.survival.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">ages/?dataset=TCGA-CESC.survival.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>
Segmento del Número de Copias	HNSC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.cnv.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.cnv.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>
Expresión Génica RNAseq HTSeq - FPKM	HNSC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.htseq_fpkm.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.htseq_fpkm.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>
Fenotipo	HNSC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.GDC_phenotype.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.GDC_phenotype.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>
Datos de Sobrevida	HNSC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.survival.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-HNSC.survival.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>
Segmento del Número de Copias	LIHC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-LIHC.cnv.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-LIHC.cnv.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>
Expresión Génica RNAseq HTSeq - FPKM	LIHC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-LIHC.htseq_fpkm.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-LIHC.htseq_fpkm.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>
Fenotipo	LIHC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-LIHC.GDC_phenotype.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-LIHC.GDC_phenotype.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fxenana.treehouse.gi.ucsc.edu%3A443</a>

Datos de Sobrevida	LIHC	<a href="https://xenabrowser.net/datapages/?dataset=TCGA-LIHC_survival.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fena.treehouse.gi.ucsc.edu%3A443">https://xenabrowser.net/datapages/?dataset=TCGA-LIHC_survival.tsv&amp;host=https%3A%2F%2Fgdc.xenahubs.net&amp;removeHub=https%3A%2F%2Fena.treehouse.gi.ucsc.edu%3A443</a>
--------------------	------	---

### 4.3. Criterios de selección de los participantes

#### 4.3.1. Criterios de inclusión:

- Sujetos varones o mujeres diagnosticados con cáncer primario de cualquiera de los 3 tipos a estudiar (cáncer cervical, cabeza y cuello e hígado), mayores de 18 años, cuya información de genotipo y fenotipo (variables clínicas y descriptivas) esté disponible en la base de datos de TCGA GDC (30–32).
- Participantes que cuenten con el estado de infección viral documentado.

#### 4.3.2. Criterios de exclusión

- Muestras con estados de coinfección por dos o más subtipos de HPV o HBV/HCV.

### 4.4. Variables

#### 4.4.1 Variables Fenotípicas

Las variables fenotípicas incluyen variables demográficas y variables clínicas.

##### 4.4.1.1. Variables demográficas/descriptivas

En la Tabla 2 se describe la operacionalización de las variables de estudio, correspondientes a las características demográficas de la población. Estas variables solo sirven para describir las poblaciones de las tres cohortes (CESC, HNSC y LIHC)

*Tabla 2. Operacionalización de variables demográficas/descriptivas*

Nombre de la variable	Definición operacional	Base de datos a la que	Valores	Tipo de variable	Escala de medición	Fuente
-----------------------	------------------------	------------------------	---------	------------------	--------------------	--------

		<b>pertenece</b>				
Edad	Años cumplidos al momento del diagnóstico de cáncer	LIHC, CESC y HNSCC	18 años a más	Cuantitativa discreta	Razón	UCSC Xena
Género	Género reportado por los pacientes	LIHC, CESC y HNSCC	Hombre, Mujer	Cualitativa dicotómica	Nominal	UCSC Xena
Etnicidad	Etnicidad reportada por el paciente	LIHC, CESC y HNSCC	- Latino o hispano - No hispano o latino	Cualitativa politómica	Nominal	UCSC Xena
Raza	Raza reportada por el paciente	LIHC, CESC y HNSCC	- Asiático u otro - Negro - Blanco	Cualitativa politómica	Nominal	UCSC Xena
Índice de masa corporal (IMC)	Peso de la paciente (en kilogramos) dividido por la estatura (en metros) elevada al cuadrado	LIHC, CESC	Mayores a 0	Cuantitativa continua	Razón	UCSC Xena
Categoría de IMC	Determinar si el paciente tiene un peso saludable	LIHC, CESC	- Bajo de peso < 18.5 - Peso de normal entre 18.5 y 24.9 - Subido de peso entre 25 y 29.9 - Obesidad > 30	Cualitativa politómica	Ordinal	UCSC Xena

#### 4.4.1.2. Variables clínicas

En la Tabla 3 se describe la operacionalización de las variables clínicas.

*Tabla 3. Operacionalización de variables de características clínicas.*

Nombre de la variable	Definición operacional	Base de datos a la que pertenece	Valores	Tipo de variable	Escala de medición	Fuente
Tipo de	Tumor	CESC, HNSCC	Tumor	Cualitativa	Nominal	UCSC

muestra	Primario o Tejido Sólido (Control)	y LIHC	Control	dicotómica		Xena
Estatus de HPV	Presencia viral de HPV	CESC y HNSCC	Negativo Positivo	Cualitativa dicotómica	Nominal	Refs. (30) y (31)
Hepatitis estatus	Presencia viral de HBV o HCV	LIHC	Negativo Positivo	Cualitativa dicotómica	Nominal	Ref. (32)
Tipo de HPV	Cepas de HPV	CESC y HNSCC	HPV18, 16, 42, 52, 58, 35, 45, 68, 31, 39, 30, 56, 59, 73, 70, 60, 33.	Cualitativa politómica	Nominal	Refs. (30) y (31)
Tipo viral de hepatitis	Tipo de virus detectado en la biopsia del tejido (RNAseq)	LIHC	HBV HCV	Cualitativa dicotómica	Nominal	Ref. (32)
Estatus de infección en la muestra	Si hay presencia del virus en la muestra tumoral primaria o en el tejido normal sólido.	CESC, HNSCC y LIHC	Tumor HPV(+) (CESC, HNSCC) Tumor HPV(-) (CESC, HNSCC) Control HPV(+) (HNSCC) Control HPV(-) (CESC, HNSCC) Tumor positivo (LIHC) Tumor negativo (LIHC) Control positivo (LIHC) Control negativo (LIHC)	Cualitativa politómica	Nominal	UCSC Xena
Grado histológico	Describe cuántas células cancerosas se parecen a las saludables	LIHC, CESC y HNSCC	G1, G2, G3	Cuantitativa politómica	ordinal	UCSC Xena
Estadío clínico o del tumor	Estadío de la enfermedad. Cuánto se ha extendido el cáncer (tamaño y difusión)	LIHC, CESC y HNSCC	I, II, III, IV	Cualitativa politómica	ordinal	UCSC Xena

Tiempo de supervivencia global (OS)	OS luego del diagnóstico de cáncer, expresado en días	LIHC, CESC y HNSCC	0 a más	Cuantitativa discreta	Razón	UCSC Xena
Consumo de Tabaco	Categoría que describe el estado actual y el historial de tabaquismo.	CESC y HNSCC	- No fumador - Fumador reformado > a 15 años - Fumador reformador < = a 15 años	Cualitativa politómica	Nominal	UCSC Xena
Consumo de Alcohol	Describe consumo alcohol.	HNSCC	Sí No	Cualitativa dicotómica	Nominal	UCSC Xena

#### 4.4.2. Datos genómicos y transcriptómicos

En la Tabla 4 se describe la operacionalización de las variables correspondientes a los datos genómicos y transcriptómicos de cada participante.

*Tabla 4. Operacionalización de variables de datos genómicos y transcriptómicos.*

Nombre de la variable	Definición operacional	Valores	Tipo de variable	Escala de medición	Fuente
Segmento del número de copias <i>GLS1/GLS2</i>	Valores de variación del número de copias ( <i>copy number</i> ) transformadas en Segmento de medias = $\log_2(\text{copy number}/2)$	Desde -1 a 1	Cuantitativa continua	De razón	UCSC Xena
Estatus de la alteración del número de copias <i>GLS1/GLS2</i>	Indica si ocurrió un cambio en el número de copias del gen. Si es que existe un cambio, indica si es delección o amplificación	< -0.3 = delección > 0.3 = amplificación Entre -0.3 y 0.3 = sin cambio (38)	Cualitativa politómica	Nominal	UCSC Xena
Expresión	Valores de	0 a más	Cuantitativa continua	De razón	UCSC

génica <i>GLS1/GLS2</i> HTSeq-FPKM	expresión expresados como el recuento normalizado de valores RSEM transformados a $\log_2(x+1)$ , donde x es el valor RSEM. Determinados por RNAseq.				Xena
LOG <sub>2</sub> FC de <i>GLS1/GLS2</i>	Es el logaritmo de base 2 de la mediana de tumor en fpkm +1 menos el logaritmo de la mediana del tejido saludable en fpkm+1	Puede tener un rango menor a -2 y llegar a tener valores mayores a 2.	Cuantitativa continua	De Razón	GEPIA
Estatus de expresión de <i>GLS1/GLS2</i>	Esta variable sirve para saber si una de las muestras tumorales se sobreexpresa o subexpresa en contraste con la mediana de los tejidos normales o controles.	LOG <sub>2</sub> FC $\geq 2$ : Sobreexpresión LOG <sub>2</sub> FC $\leq -2$ : Subexpresión Si LOG <sub>2</sub> FC está entre 2 y -2: no hay cambio en la expresión (39).	Cualitativa politómica	Nominal	GEPIA

#### 4.5. Procesamiento de las bases de datos

En este estudio se consultaron bases de datos genómicos de acceso abierto, provenientes de UCSC Xena (37). Las cohortes analizadas fueron: GDC TCGA Cervical Cancer (CESC) (30), GDC TCGA Head and Neck Cancer (HNSC) (31), GDC TCGA Liver Cancer (LIHC) (32).

El número total de sujetos disponibles por cada cohorte es el siguiente:

- GDC TCGA Cervical Cancer (CESC): 317
- GDC TCGA Head and Neck Cancer (HNSC): 612
- GDC TCGA Liver Cancer (LIHC): 469

Se descargaron los datos para las variables de interés (Tabla 2, Tabla 3 y Tabla 4) por cada cohorte y se vaciaron en hojas de cálculo de Excel (Figura 1). Algunas variables clínicas (Tabla 3) como estatus de HPV, tipo de HPV (excepto en HNSCC), estatus de hepatitis y tipo viral de hepatitis fueron obtenidas directamente de los materiales suplementarios provenientes de los artículos científicos de cada cohorte (30–32).

En este estudio contamos con controles que se describen a continuación:

- **Controles del grupo cáncer:** muestras no cancerosas y con prueba positiva para infección viral.
- **Controles completos:** muestras no cancerosas y con prueba negativa para infección viral

Estas muestras controles son provenientes de tejido no cancerígeno o tejido sólido adyacente al tumor primario del paciente, es decir, el tejido tumoral.

En la Figura 2 se muestra un flujograma del diseño experimental, el cual será descrito en las siguientes subsecciones.

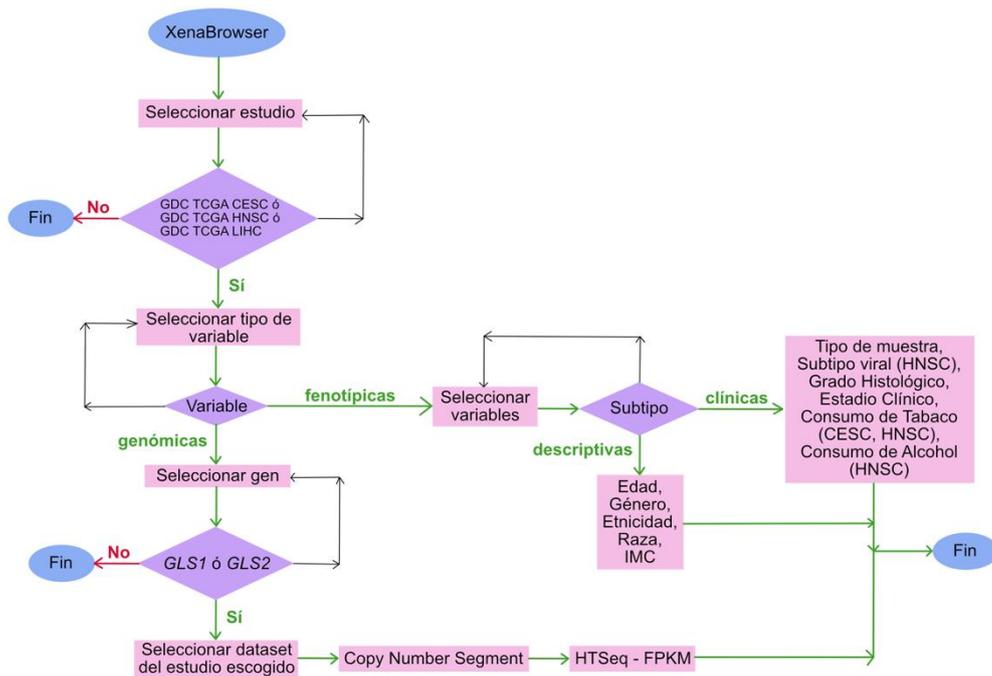


Figura 1. Flujoograma del proceso de descarga de la base de datos.

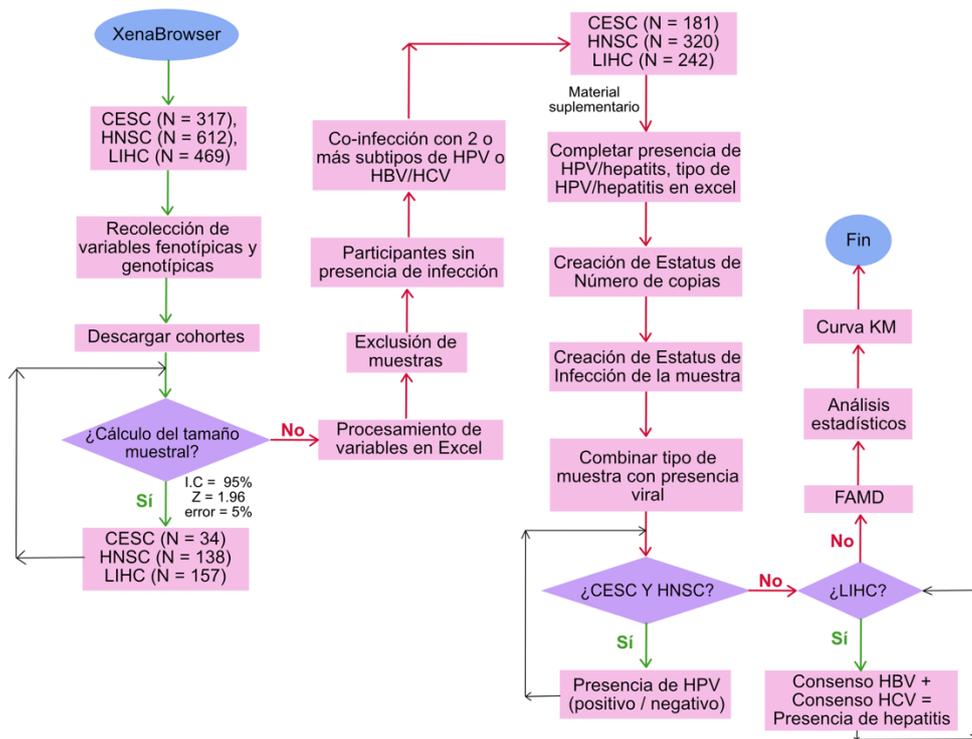


Figura 2. Flujoograma del diseño experimental completo.

#### 4.5.1 Cálculo del tamaño muestral

El cálculo del tamaño muestral mínimo esperado se realizó en Excel, en donde primero se calcularon las proporciones del tejido no canceroso y canceroso en cada cohorte. Luego se usó la Ecuación 1:

*Ecuación 1. Cálculo del tamaño muestral.*

$$n = \frac{N \times Z_{\alpha}^2 \times p \times (1 - p)}{d^2 \times (N - 1) + Z_{\alpha}^2 \times p \times (1 - p)}$$

Donde,  $N$ : tamaño inicial de la población,  $Z$ : coeficiente de confiabilidad 1.96 con un nivel de confianza de 95%,  $p$ : proporción del tejido no canceroso y  $d$ : error máximo admisible del 5%.

Los resultados del cálculo fueron los siguientes:

- GDC TCGA Cervical Cancer (CESC): 34
- GDC TCGA Head and Neck Cancer (HNSC): 138
- GDC TCGA Liver Cancer (LIHC): 157

En este estudio, se contaron con más muestras de las necesarias de acuerdo al cálculo de la Ecuación 1, inclusive después de aplicar los criterios de elegibilidad en los participantes.

#### 4.5.2 Estatus de infección de la muestra

Para la construcción de la variable estatus de infección de la muestra, se combinó la variable tipo de muestra con estatus de HPV en CESC y HNSCC, mientras que en la cohorte LIHC, se combinaron el tipo de muestra con el estatus de hepatitis. Se obtuvieron los siguientes valores:

Para CESC y HNSCC:

- Tumor HPV(+)
- Tumor HPV(-)
- Control HPV(+) = controles del grupo de cáncer
- Control HPV(-) = controles completos

Para LIHC:

- Tumor HBV/HCV(+)
- Tumor HBV/HCV(-)
- Control HBV/HCV(+) = controles del grupo de cáncer
- Control HBV/HCV(-) = controles completos

#### **4.5.3 Segmento del número de copias**

La “Variación del Número de Copias” (CNV) fue obtenido por el TCGA mediante *Next Generation Sequencing* (NGS) o Affymetrix SNP 6.0 (SNP6) microarrays. Luego, se identificaron las regiones genómicas que se repiten. A partir de ahí, se infirió el “Número de Copias” (*Copy Number*) de esas repeticiones (40).

Para procesar los datos, se produjo un set de sondas moleculares tipo SNP6. Se mapearon las secuencias de dichas sondas al genoma de referencia GRCh38 o hg38 (41).

Los archivos “*tangent.copynumber*” de una alícuota del TCGA nivel 2 SNP6 fueron la base para crear el tercer set de *pipelines* CNV. Para generar estos archivos, se normalizó los valores de intensidad de los microarrays, estimando así el número de copias sin procesar. A partir de estos resultados, se realizó una normalización tangencial (40). Esta asume que el ruido, transformado en unidades logarítmicas, de los datos de arrays de SNPs se distribuyen de igual forma en muestras cancerígenas

vs no cancerígenas. Por lo tanto, para minimizar el ruido, se restaron los perfiles de ruido estimados, los cuales se calcularon individualmente para cada tumor usando data de todas las muestras no cancerígenas (42). Esto tiene como propósito reducir la variación en el set de las muestras no cancerígenas.

Luego se realizó un análisis de segmentación binaria circular (CBS). Este convierte la intensidad de las medidas ruidosas en regiones cromosómicas con número de copias equivalentes (43). Lo que resultó de este análisis, se segmentó en regiones genómicas con el número de copias estimado para cada región (40).

Las regiones genómicas corresponden al “Número de Copias” (*Copy Number*). La GDC transforma estos valores en la variable descargada por el XenaBrowser “Segmento del Número de Copias” (Tabla 4, Ecuación 2). Esta variable asocia regiones cromosómicas contiguas con las medias del segmento genómico (40).

*Ecuación 2. Transformación del Segmento del número de copias por la GDC.*

$$\text{Segmento del Número de copias} = \log_2 \left( \frac{\text{Número de Copias}}{2} \right)$$

Con la variable “Segmento del Número de Copias”, en Excel creamos una nueva variable categórica llamada “Estatus de la alteración del número de copias”, donde para cada muestra (38):

- Segmento del Número de Copias < -0.3 = delección,
- Segmento del Número de Copias > 0.3 = amplificación
- $-0.3 \leq \text{Segmento del Número de Copias} \leq 0.3$  = No hay cambio del número de copias.

#### **4.5.4 Expresión génica RNAseq – HTSeq-FPKM**

Para realizar el secuenciamiento RNA-seq, el TCGA usó la plataforma IlluminaHiSeq. El GDC utilizó el programa “alineación de transcritos empalmados con una referencia” (STAR) para medir el nivel de expresión génica mediante el análisis de

cuantificación de Ácido ribonucleico mensajero. STAR alinea cada grupo de lectura usando el genoma de referencia GRCh38 (41) de forma separada y luego junta todos los alineamientos en uno (44). Para cuantificar las lecturas alineadas o mapeadas, el GDC usó el algoritmo RNA-Seq por maximización de expectación (RSEM). Para estimar las cantidades de lecturas alineadas por STAR, RSEM usa el alineamiento de las lecturas del RNA-seq a los transcritos de referencia. Finalmente, estos recuentos sin procesar se transforman a unidades FPKM (44).

El FPKM son los fragmentos por kilobase del exón por cada millón de lecturas mapeadas. El objetivo de usar FPKM como normalización es controlar la longitud de los transcritos y la cantidad total de secuencias (44).

El GDC calculó el FPKM como se muestra en la Ecuación 3:

*Ecuación 3. Cálculo del FPKM hecho por la GDC*

$$FPKM = \frac{C_g \times 10^9}{(\sum_{i=1}^N C_i) L_g}$$

Donde,  $C_g$ : número de lecturas mapeadas al gen "g",  $N$ : número de genes codantes,  $\sum_{i=1}^N C_i$ : el número total de lecturas mapeadas a secuencias codantes,  $L_g$ : longitud del gen en pares de bases.

Los valores de FPKM se anotaron con el símbolo del gen y el biotipo del gen (44). Finalmente, XenaBrowser realizó una transformación logarítmica de estos resultados con la Ecuación 4 (45).

*Ecuación 4. Normalización realizada por el XenaBrowser.*

$$y = \log_2(x + \theta)$$

$$y = \log_2(x + 1)$$

$$HTSeq - FPKM = \log_2(FPKM + 1)$$

Donde  $x$  corresponde al tipo de transformación de los resultados de RSEM, en este caso, siendo el FPKM.  $\theta$  es un número pequeño, en este caso 1. Esto se añade debido a que no se puede sacar el log de cero (46).

#### **4.6 Análisis de factores de datos mixtos (FAMD)**

A continuación, se realizó un análisis de factores de datos mixtos (FAMD). El FAMD es una técnica que mezcla el análisis de componentes principales (PCA), el cual se usa para datos cuantitativos continuos y el análisis de correspondencia múltiple (MCA), el cual se usa para variables cualitativas (47,48).

El propósito de realizar FAMD fue ver qué variables contribuyen mejor a la variabilidad del set y así reducir el número de variables para el análisis.

Para realizar este análisis se usó R con ayuda de los paquetes tidyverse (49), naniar (50), missMDA (51), factoextra (52), FactoMineR (53) y ggpubr (54). Los scripts de los análisis se encuentran en el Anexo 2.

No se tomaron en cuenta las variables demográficas, ya que estas sirven solo para describir a las cohortes (Tabla 2). Se agregaron como datos suplementarios a las variables genómicas y transcriptómicas (Tabla 4), además de la OS (Tabla 3)Tabla 3. Esto sirve para que sean tomadas en cuenta en la creación del modelo, pero no salgan en el porcentaje de contribución a los componentes principales, ya que estas variables son importantes para nuestros objetivos.

Finalmente, las variables clínicas (Tabla 3) que contribuyeron menos a los componentes principales de CESC, HNSC y LIHC fueron excluidas de los análisis posteriores.

#### **4.7. Análisis estadísticos**

Los libros de cálculo de Excel, conteniendo todos los datos de interés en la población de estudio, se subieron a la plataforma del XenaBrowser en “View my

*data*". Ahí, se realizaron las pruebas estadísticas como el chi-cuadrado, t-welch, ANOVA de una vía y la curva de KM.

Para alcanzar el primer objetivo específico, primero se quiso saber si había asociación entre el tipo de muestra, es decir tejidos tumorales vs no tumorales, y el estatus de la alteración del número de copias de *GLS1/GLS2*. Para ello, se realizó un análisis de chi-cuadrado ( $p < 0.05$ ). Luego se comparó el nivel de expresión génica de *GLS1/GLS2* en las muestras tumorales vs no tumorales. Se evaluó la diferencia de expresión basada en las medias mediante pruebas t de Welch ( $p < 0.05$ ). Luego, las cohortes (CESC, HNSC, LIHC) que tuvieron diferencias de expresión significativa ya sea en *GLS1* o *GLS2*, se estimó en ellas el  $\text{Log}_2\text{FC}$  en Excel para ver si existía una sobreexpresión o subexpresión (véase sección 4.6.1).

Para el segundo objetivo, se aplicó la prueba de chi cuadrado ( $p < 0.05$ ) para ver si existe una asociación entre el estatus de alteración de copias *GLS1/GLS2* y las variables clínicas que quedaron después de realizar el FAMD (sección 4.4.2). Luego, para comparar las medianas de expresión génica de *GLS1/GLS2* y las variables clínicas que quedaron, se usó ANOVA de una vía ( $p < 0.05$ ). Finalmente, las cohortes que tuvieron diferencias significativas de expresión ya sea en *GLS1* o *GLS2*, se estimó en ellas el  $\text{Log}_2\text{FC}$  (véase sección 4.6.1). En la Figura 3 se observa el flujograma del procedimiento de los análisis estadísticos realizados.

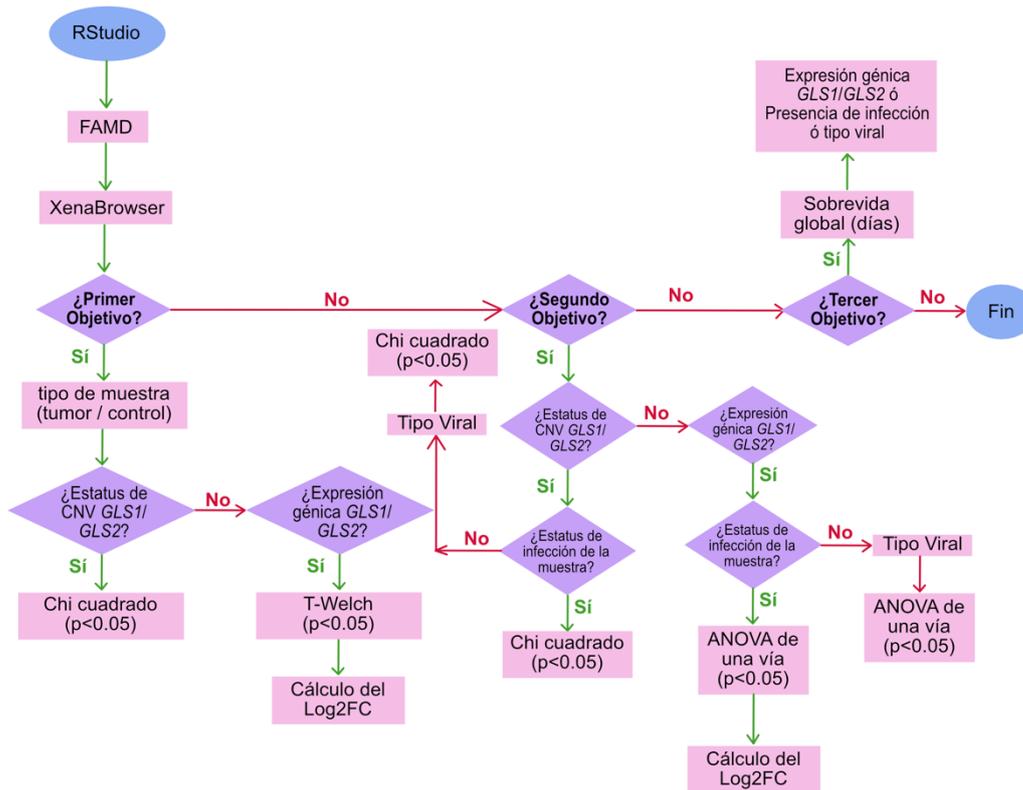


Figura 3. Flujograma del proceso de los análisis estadísticos.

#### 4.7.1 Cálculo del Log<sub>2</sub>FC

Como se mencionó en la sección 4.5.3, XenaBrowser realizó una transformación logarítmica de la data de la expresión génica en FPKM de ambas glutaminasas.

Para saber la tasa de cambio o diferencia que hay entre los tejidos cancerosos vs los no cancerosos en cada glutaminasa, se usó la proporción de cambio (FC). Esta usa la regla del cociente de los logaritmos, la cual establece lo siguiente en la Ecuación 5 (46):

Ecuación 5. Regla del cociente de logaritmos aplicado a la proporción de cambios.

$$\log_2 FC = \log_2 \left( \frac{\text{Mediana tumorales} + 1}{\text{Mediana no tumorales} + 1} \right)$$

$$\log_2 FC = \log_2(\text{Mediana tumorales} + 1) - \log_2(\text{Mediana no tumorales} + 1)$$

Esto nos da como resultado el  $\text{Log}_2\text{FC}$ . El cálculo se realizó en Excel, sin embargo, como la data estaba transformada, solo se restó lo que se ve en la Ecuación 6:

*Ecuación 6. Cálculo del  $\text{Log}_2\text{FC}$  en Excel*

$$\log_2FC = \text{mediana}(\text{tumoraes}) - \text{mediana}(\text{no tumoraes})$$

Primero se calculó el  $\text{Log}_2\text{FC}$  de cada glutaminasa de forma general, es decir, solo basado en el tipo de muestra. Para ello se agrupó los datos de expresión génica de cada glutaminasa en muestras tumorales y en muestras no tumorales. Se le calculó la mediana a cada grupo y se restaron ambos valores como indica la Ecuación 6.

Después se calculó el  $\text{Log}_2\text{FC}$  de cada glutaminasas en base la variable estatus de infección de la muestra, la cual indica si la muestra, además de ser cancerígena o no cancerígena, presenta infección viral, ya sea por HPV, HBV o HCV. En este caso, se agrupó los datos de expresión génica de cada glutaminasa en muestras tumorales con estado de infección positivo/negativo y muestras no tumorales (controles) con estado de infección positivo/negativo. Se calculó la mediana de los 4 grupos y se restaron los valores de la manera como indica la Ecuación 7:

*Ecuación 7. Cálculo del  $\text{Log}_2\text{FC}$  según la presencia o ausencia de infección en Excel.*

$$\log_2FC = \text{mediana}(\text{tumoraes}_x) - \text{mediana}(\text{controles}_x)$$

$$\log_2FC = \text{mediana}(\text{tumoraes}_y) - \text{mediana}(\text{controles}_y)$$

Donde  $x$  es la presencia de infección e  $y$  es la ausencia de infección por HPV, HCV o HBV. Luego, para determinar si los genes estaban sub o sobreexpresados, los resultados se interpretan de la siguiente manera:

- $\log_2FC \geq 2 =$  El gen *GLS1* o *GLS2* está **sobreexpresado** en muestras tumorales (con/sin infección por HPV, HBV/HCV) comparadas con muestras no tumorales en la cohorte CESC, HNSC o LIHC.

- $\log_2 FC \leq -2$  = El gen *GLS1* o *GLS2* está **subexpresado** en muestras tumorales (con/sin infección por HPV, HBV/HCV) comparadas con muestras no tumorales en la cohorte CESC, HNSC o LIHC.
- $-2 \leq \log_2 FC \leq 2$  = No hay ni sobreexpresión ni subexpresión.

#### 4.7.2 Análisis de sobrevida KM

Finalmente, para estimar la influencia de la expresión génica de *GLS1/GLS2* y del estado de infección (positivo o negativo) con la sobrevida global, graficamos un KM en XenaBrowser. Se dividieron las variables expresión génica en dos grupos basados en el cuartil inferior (Q1) y el cuartil superior (Q3).

Se usaron múltiples curvas en cada gráfico de KM. En ese caso, para saber si las curvas eran significativamente diferentes, se usó la prueba de rango logarítmico ( $p < 0.05$ ). Esta prueba se basa en el estadístico de chi-cuadrado y el valor p está dado por la distribución de chi cuadrado (55).

El **tiempo hasta el evento (días o meses)** se define como la duración para cada paciente en el estudio. Se inicia cuando el paciente es enrolado en el estudio y, en caso de la **sobrevida global (OS)**, al tiempo de muerte o cuando se censura al paciente del estudio. Censurar un paciente significa que el tiempo total de sobrevida para el paciente no se puede determinar por cualquier motivo, por ejemplo, cuando el paciente abandona el estudio, se perdió durante el seguimiento o no murió durante todo el intervalo de estudio (56).

El **evento** indica si el paciente fue censurado o no. Esta data se codifica de manera binaria, en donde 1 significa que el paciente murió (evento) y 0 que el paciente fue censurado (no ocurrió el evento). Los pacientes censurados se muestran mediante una marca vertical en cada escalón de la función KM (56).

Para graficar el análisis de KM, se crea una tabla con 3 columnas: el ID del paciente, **el tiempo hasta el evento**, en este caso corresponde al **OS en días**, y el **evento** en sí.

En el eje  $x$ , se coloca el tiempo  $t_n$ , en este caso, el OS en días. En el eje  $y$ , se grafica la probabilidad de supervivencia ( $PS_{t_n}$ ). En general, la probabilidad de supervivencia baja en el eje  $y$ , en unidades dadas por la Ecuación 8. Esto hace que la función sea en forma de una escalera (57).

*Ecuación 8. Bajada de la probabilidad de supervivencia*

$$BPS_{t_n} = \frac{E_n}{PR_n}$$

En donde,  $BPS$  es la bajada de probabilidad de supervivencia,  $PR$  es el número de pacientes en riesgo,  $E$  es el tipo de evento (1 o 0),  $n$  es el nº o ID del paciente (57).

Para el primer paciente  $n = 1$ , el nº de pacientes en riesgo es el número de participantes totales en la cohorte ( $N$ ) (Ecuación 9) (57).

*Ecuación 9. Nº de pacientes en riesgo cuando  $n = 1$*

$$PR_n = PR_1 = N$$

Cuando  $n \geq 2$ , el nº de pacientes en riesgo se determina según la Ecuación 10 y el nº de pacientes en riesgo va disminuyendo en cada paciente participante (57):

*Ecuación 10. Nº de pacientes en riesgo cuando  $n \geq 2$*

$$PR_n = PR_{n-1} - 1$$

Por lo que, para  $n \geq 2$ , la bajada de probabilidad de supervivencia se denota por la Ecuación 11. Esta se ha obtenido combinando la Ecuación 8 con la Ecuación 10 (57):

*Ecuación 11. Bajada de probabilidad de supervivencia cuando  $n \geq 2$*

$$BPS_{t_n} = \frac{E_n}{PR_{n-1} - 1}$$

En general, la probabilidad de supervivencia para  $n \geq 1$  se da por la Ecuación 12.

*Ecuación 12. Probabilidad de supervivencia para  $n \geq 1$*

$$PS_{t_n} = PS_{t_{n-1}} - BPS_{t_n}$$

Finalmente, la probabilidad de supervivencia se obtiene reemplazando la Ecuación 8 en la Ecuación 12 (57).

*Ecuación 13. Probabilidad de supervivencia para  $n \geq 1$*

$$PS_{t_n} = PS_{t_{n-1}} - \frac{E_n}{PR_n}$$

Ahora, la Ecuación 13 se modifica dependiendo de  $n$ . Para el primer paciente,  $n = 1$ , El número de pacientes en riesgo está dado por la Ecuación 9, la cual se reemplaza en la Ecuación 13 (57).

*Ecuación 14. Probabilidad de supervivencia cuando  $n = 1$*

$$PS_{t_1} = PS_{t_0} - \frac{E_1}{N}$$

En donde  $PS_{t_0} = 1$ , ya que antes que se dé el primer evento, es decir la primera muerte, la probabilidad de supervivencia es del 100%. Al alcanzar el evento del primer paciente en un tiempo  $t_1$ , ocurre la bajada de probabilidad de la supervivencia (57).

Cuando  $n \geq 2$ , el número de pacientes en riesgo se da por la Ecuación 10, la cual se reemplaza en la Ecuación 13, por lo que la probabilidad de supervivencia es (57):

*Ecuación 15. Probabilidad de supervivencia cuando  $n \geq 2$*

$$PS_{t_n} = PS_{t_{n-1}} - \frac{E_n}{PR_{n-1} - 1}$$

Para continuar con la gráfica, una vez que lleguemos al primer paciente con evento censurado ( $E_n = E_{n_0} = 0$ ), se coloca una marca vertical pequeña y no hay bajada en la probabilidad de supervivencia, sin embargo, el número de pacientes en riesgo sigue disminuyendo como se muestra en la Ecuación 16 (57).

*Ecuación 16. Bajada de la probabilidad de supervivencia cuando  $E = 0$*

$$BPS_{t_{n_0}} = \frac{E_{n_0}}{PR_{n-1} - 1} = \frac{0}{PR_{n-1} - 1} = \frac{0}{PR_{n_0}} = 0$$

$$BPS_{t_{n_0+1}} = \frac{E_{n_0+1}}{PR_{n_0} - 1}$$

En donde  $n_0$  es el primer paciente con  $E = 0$  y  $n_0 + 1$  es el paciente que le sigue al paciente  $n_0$  (57).

Si  $E_{n_0+1} = 1$ , entonces el  $BPS_{t_{n_0+1}}$  sigue bajando en el eje  $y$ , de lo contrario, si  $E_{n_0+1} = 0$ , el  $BPS_{t_{n_0+1}}$  no se mueve, quedando una línea recta horizontal con una marca vertical. Sin embargo, el  $PR$  seguirá disminuyendo (57).

#### 4.7.3 Cálculo de porcentajes de cada subtipo viral

En Excel, se calculó la proporción de cada subtipo de HPV. Los subtipos HPV menos comunes se agruparon en la categoría “otros”. También se calculó la proporción de muestras con infección por HBV y con HCV. Luego, se realizó el gráfico de tartas en R.

#### 4.8 Herramientas de análisis estadístico

Para realizar los análisis estadísticos, incluyendo los gráficos de caja y bigotes, se empleó la plataforma UCSC Xena (37). Para las gráficas de tarta se emplearon utilizando R (versión 4.3.2), RStudio (versión 2023.12.1+402). Para el FAMD se utilizó R (versión 4.3.2) (48), RStudio (versión 2023.12.1+402), con los paquetes tidyverse (49), naniar (50), missMDA (51), factoextra (52), FactoMineR (53) y ggpubr (54). 4.9. Herramientas bioinformáticas

En la Tabla 5 se describen los recursos bioinformáticos que se usaron en el presente estudio.

*Tabla 5. Recursos bioinformáticos empleados en el estudio*

Nombre	Descripción	Utilidad en el	Sitio Web	Referencia
--------	-------------	----------------	-----------	------------

		<b>estudio</b>		
UCSC Xena	Recurso web interactivo para la exploración de conjuntos de datos ómicos públicos (incluyendo TCGA, GDC y GTEx) y privados	Análisis de Kaplan Meier para OS. Acceso a los datos ómicos y clínicas. Análisis estadísticos.	<a href="https://xena.ucsc.edu/">https://xena.ucsc.edu/</a>	Goldman MJ <i>et al.</i> , 2020 (37)
GEPIA	Recurso web interactivo para el análisis de datos de expresión genética, en muestras de proyectos del TCGA y GTEx	Consulta de fórmulas para el Log <sub>2</sub> FC y cómo calcularlas.	<a href="http://gepia.cancer-pku.cn/">http://gepia.cancer-pku.cn/</a>	Tang <i>et al.</i> , 2017 (60)
KEGG	Base de datos que integra información genómica, química y funcional sistémica.	Consulta de rutas metabólicas.	<a href="https://www.genome.jp/kegg/kegg1.html">https://www.genome.jp/kegg/kegg1.html</a>	Kanehisa <i>et al.</i> , 2019 (61)

#### 4.10. Cuestiones éticas

El presente estudio gestionó de forma secundaria los datos provenientes de pacientes estadounidenses, en su mayoría de raza blanca. El estudio no puso en riesgo a la salud ni tuvo acceso a datos confidenciales de los participantes. Este estudio fue registrado con código de SIDISI N° 203509 y fue aprobado por el comité de ética de la UPCH el 2 de noviembre del 2020.

## 5. Resultados

### 5.1 Población de estudio

Se aplicaron todos los criterios de selección a los participantes (véase la [sección 4.3](#)), lo cual redujo el número de sujetos disponible para cada cohorte:

- GDC TCGA Cervical Cancer (CESC): 181
- GDC TCGA Head and Neck Cancer (HNSC): 320

Cabe resaltar que el número de muestras por cohorte es mayor a las calculadas en la sección 4.5.1.

## 5.2 FAMD

### 5.2.1. Cáncer cervical (CESC)

Para la cohorte de CESC, los dos primeras componentes principales o dimensiones (PC) contribuyen al 31.5% de la variabilidad del set. La PC1 contribuye al 19.9% de la variabilidad mientras que el PC2 contribuye al 11.6% (Figura 4).

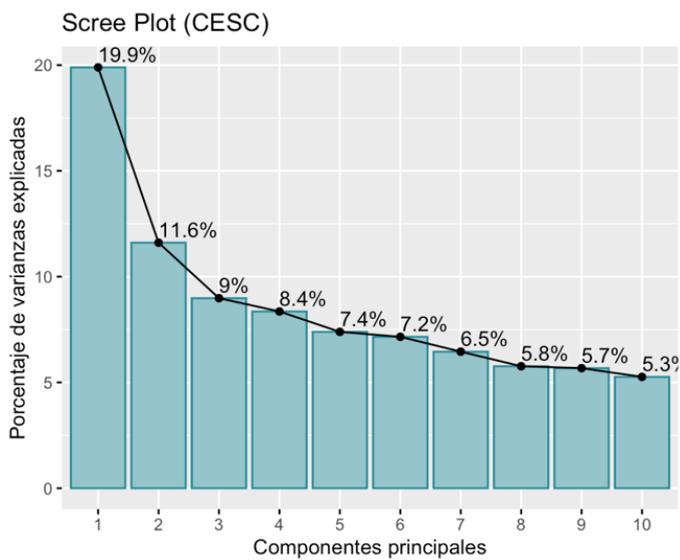
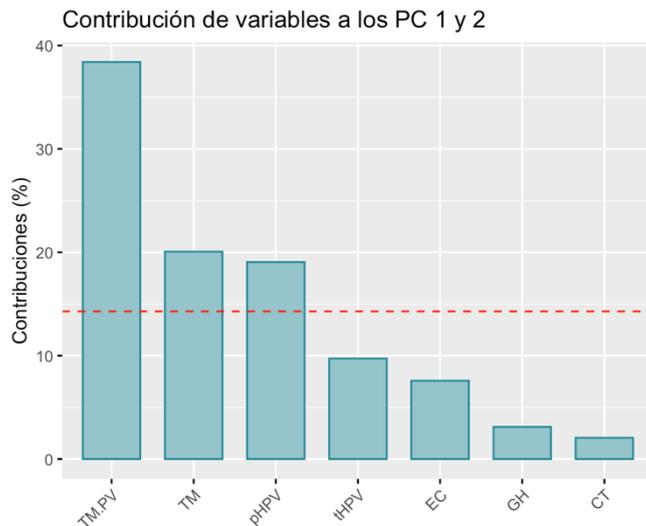


Figura 4. Screen plot del porcentaje de varianzas explicadas por cada PC en CESC.

Las variables que tienen mayor contribución a los dos primeros componentes principales son estatus de infección de la muestra, el tipo de muestra y la presencia de HPV (Figura 5).



*Figura 5. Porcentaje de contribución de las variables a los primeros dos PC (CESC). Línea roja: porcentaje de contribución mínima. TM.PV: Estatus de infección de la muestra, TM: Tipo de muestra, pHPV: Presencia de HPV, tHPV: Tipo de HPV, EC: Estadio clínico, GH: Grado histológico, CT: Consumo de Tabaco.*

Como podemos ver en la Figura 5, el tipo de HPV, estadio clínico, grado histológico y consumo de tabaco no contribuyen lo suficiente en estos componentes. Sin embargo, mantuvimos el tipo de HPV porque es una variable de interés primario para este estudio.

### **5.2.2. Cáncer de cabeza y cuello (HNSC)**

Para la cohorte de HNSC, las dos primeras componentes principales o dimensiones (PC) contribuyen al 41.9% de la variabilidad del set. La PC1 contribuye al 17.3% de la variabilidad, el PC2 contribuye al 14.5% y el PC3 10.1% (Figura 6).

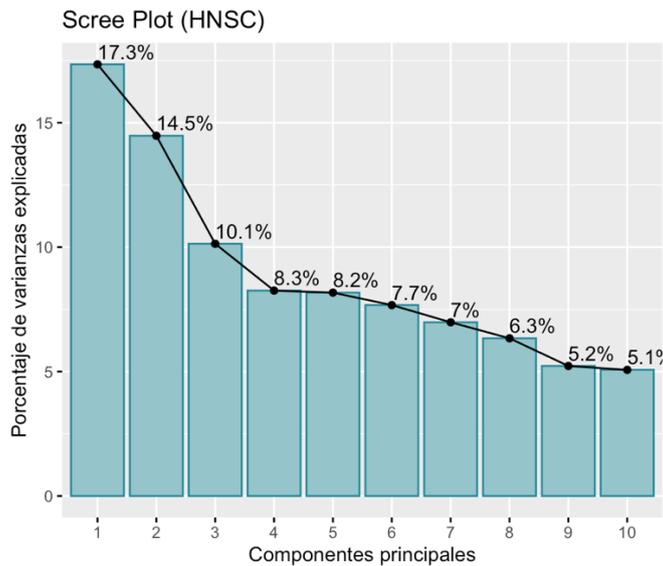


Figura 6. Screen plot del porcentaje de varianzas explicadas por cada PC en HNSC.

Las variables que tienen mayor contribución a los tres primeros componentes principales son estatus de infección, el tipo de muestra y la presencia de HPV (Figura 7).

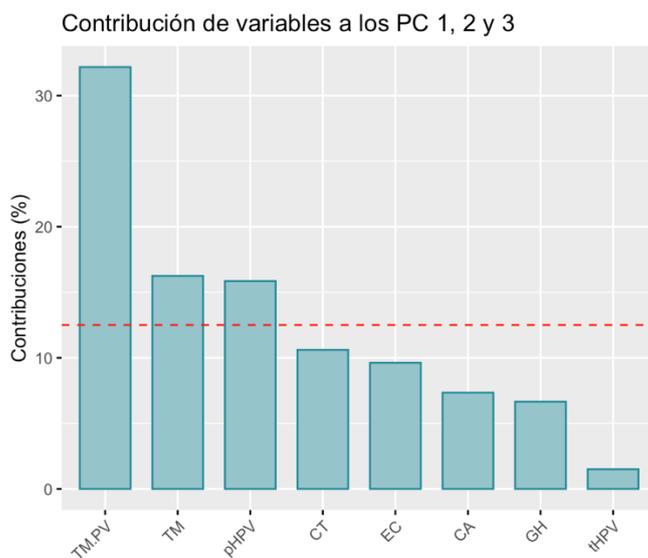


Figura 7. Porcentaje de contribución de las variables a los primeros dos PC (HNSC). Línea roja: porcentaje de contribución mínima. TM.PV: Estatus de infección de la muestra, TM: Tipo de muestra, pHPV: Presencia de HPV, tHPV: Tipo de HPV, EC: Estadio clínico, GH: Grado histológico, CT: Consumo de Tabaco, CA: Consumo de Alcohol.

Como podemos ver en la Figura 7, el tipo de HPV, estadio clínico, grado histológico, consumo de tabaco y de alcohol no contribuyen lo suficiente en estos componentes. Sin embargo, mantuvimos el tipo de HPV porque es una variable de interés primario para este estudio.

### 5.2.3. Cáncer de hígado (LIHC)

Para la cohorte de LIHC, los dos primeras componentes principales o dimensiones (PC) contribuyen al 31.8% de la variabilidad del set. La PC1 contribuye al 17.3% de la variabilidad mientras que el PC2 contribuye al 14.5% (Figura 8).

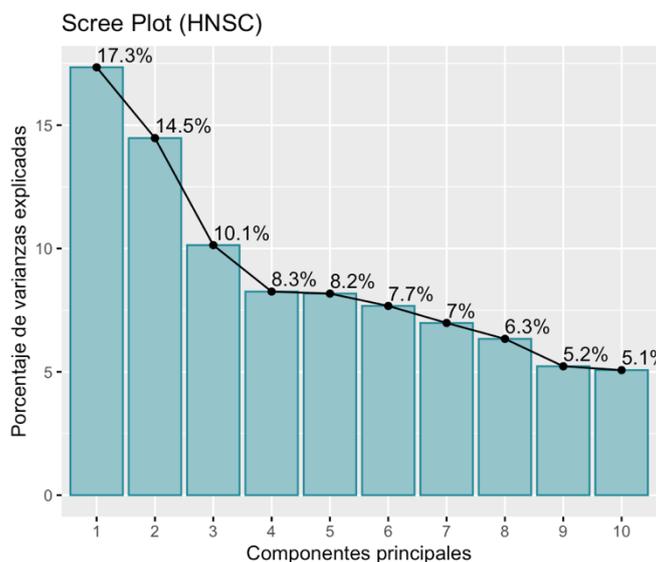


Figura 8. Scree plot del porcentaje de varianzas explicadas por cada PC en LIHC.

Las variables que tienen mayor contribución a los dos primeros componentes principales son estatus de infección de la muestra, el tipo de muestra y la presencia de HBV o HCV (Figura 9).

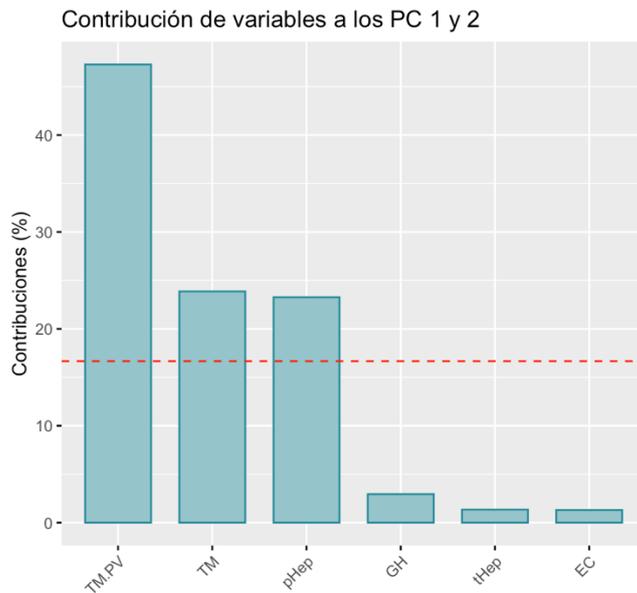


Figura 9. Porcentaje de contribución de las variables a los primeros dos PC (LIHC). Línea roja: porcentaje de contribución mínima. TM.PV: Estatus de infección de la muestra, TM: Tipo de muestra, pHep: Presencia de Hepatitis, tHep: Tipo de Hepatitis, EC: Estadio clínico, GH: Grado Histológico

Como podemos ver en la Figura 9, el tipo de Hepatitis (HBV o HCV), estadio clínico y grado histológico no contribuyen lo suficiente en estos componentes. Sin embargo, mantuvimos el tipo de hepatitis porque es una variable de interés primario para este estudio

### 5.3. Cambio de la expresión génica y la alteración del número de copias de *GLS1* y *GLS2* en sujetos con LIHC, CESC y HNSCC.

#### 5.3.1. Cáncer cervical (CESC)

La expresión de *GLS1* y *GLS2* fue equivalente entre el tejido cervical neoplásico y el tejido no canceroso para la variable del tipo de muestra ( $p > 0.05$ ; Anexo 3).

No apreciamos alguna alteración en el número de copias de *GLS1/GLS2* en las muestras cancerosas cuando se les comparó con tejido no canceroso, en la cohorte de CESC (Anexo 4).

### 5.3.2. Cáncer de cabeza y cuello (HNSCC)

Los pacientes con HNSCC tienen una mayor expresión de *GLS1* en comparación con los que no tienen cáncer ( $p < 0.0001$ ) (Tabla 6, Figura 10). Sin embargo, el valor de  $\text{Log}_2\text{FC}$  no es suficiente para ser considerado una sobreexpresión con respecto al control. Para *GLS2*, no hubo diferencias entre la expresión de pacientes con HNSCC y el grupo control ( $p > 0.05$ , Anexo 7).

Tabla 6. Expresión génica de *GLS1/GLS2* (expresados en mediana de  $\log_2(\text{fpkm}+1)$ ) y el  $\text{Log}_2\text{FC}$  por el tipo de muestra en cáncer de cabeza y cuello.

Gen	Variable		Test estadístico	$\text{Log}_2\text{FC}^*$	Estatus de expresión
	Tipo de muestra				
	Tumor	Control	Test T de Welch		
<i>GLS1</i>	3 (n=274)	2.23 (n=44)	$p = 1.634e-22$ ( $t = -12.90$ )	0.77	Indeterminado

\* $\text{Log}_2\text{FC} \leq -2 = \text{subexpresado}$ ,  $\text{Log}_2\text{FC} \geq 2 = \text{sobreexpresión}$ ,  $-2 < \text{Log}_2\text{FC} < 2 = \text{indeterminado}$

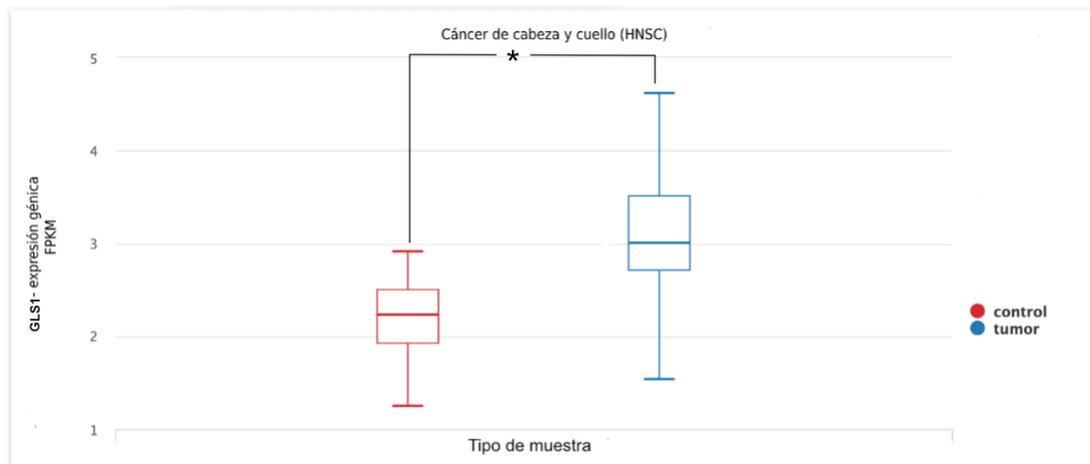


Figura 10. Gráfica de caja y bigotes de la expresión génica de *GLS1* según el tipo de muestra de tejido de cabeza y cuello (cancerosa o normal).

No hubo alteración en el número de copias al comparar las muestras cancerosas con las normales, en HNSCC (Anexo 8).

### 5.3.3. Cáncer de hígado (LIHC)

Los pacientes con LIHC tienen una mayor expresión de *GLS1* y una menor expresión de *GLS2*, comparados con el grupo control ( $p < 0.0001$ ) (Tabla 7, Figura 11, Figura 12) Considerando el  $\log_2FC$ , se determinó que los tejidos con cáncer de hígado tienen el gen *GLS2* subexpresado en comparación con los tejidos no cancerosos (Tabla 7). Sin embargo, para el gen *GLS1*, el estado de expresión es indeterminado.

Tabla 7. Expresión génica (expresados en mediana de  $\log_2(fpkm+1)$ ) de *GLS1/GLS2* y el  $\text{Log}_2FC$  por el tipo de muestra en cáncer de hígado.

Gen	Variable		Test estadístico	Log2FC	Estatus de expresión
	Tipo de muestra		Test T de Welch		
	Tumor	Control			
<i>GLS1</i>	2.07 (n=189)	1.07 (n=50)	$p = 4.263e-23$ ( $t = -11.25$ )	1.01	Indeterminado
<i>GLS2</i>	0.368 (n=189)	2.36 (n=50)	$p = 5.258e-18$ ( $t = -10.18$ )	-2	Subexpresión

\*  $\text{Log}_2FC \leq -2 = \text{subexpresado}$ ,  $\text{Log}_2FC \geq 2 = \text{sobreexpresión}$ ,  $-2 < \text{Log}_2FC < 2 = \text{indeterminado}$

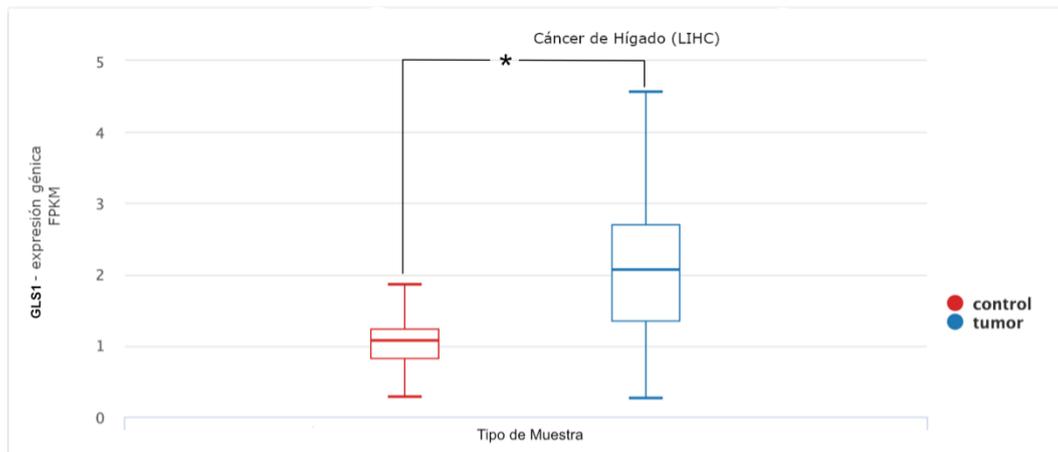


Figura 11. Gráfica de caja y bigotes de la expresión génica de *GLS1* según el tipo de muestra de tejido hepático (cancerosa o normal).

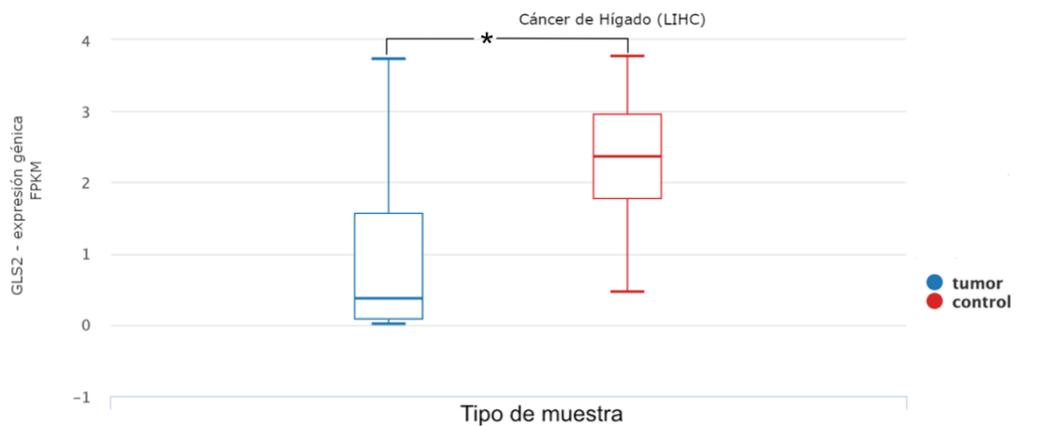


Figura 12. Gráfica de caja y bigotes de la expresión génica de *GLS2* según el tipo de muestra de tejido hepático (cancerosa o normal).

No hubo alteración en el número de copias entre muestras cancerosas y no tumorales en LIHC (Anexo 11).

#### 5.4. Relación entre la expresión de *GLS1/GLS2* y variables clínicas.

##### 5.4.1. Cáncer cervical

Encontramos que la presencia en conjunto de neoplasia cervical y HPV se asocian con una mayor expresión de *GLS2* ( $p < 0.05$ ) (Tabla 8, Figura 13). Sin embargo, no hubo diferencias en la expresión génica de *GLS1* cuando se agrupó por el estatus de infección de la muestra ( $p > 0.05$ ) (Anexo 5).

Tabla 8. Expresión génica de *GLS1* y *GLS2* agrupadas por variables clínicas con diferencias significativas y el  $\text{Log}_2\text{FC}$  en la cohorte de CESC.

Gen	Expresión génica (mediana de $\log_2(\text{fpkm}+1)$ )			Prueba estadística
	Variable: Estatus de infección de la muestra			
	Tumor con infección HPV(+)	Tumor sin infección HPV(-)	Control sin infección HPV(-) (tejido no tumoral sin infección)	ANOVA de una vía
<i>GLS2</i>	0.303 (n=169)	0.0844 (n=9)	0.0755 (n=3)	p = 0.03516 (f)

	Log2FC: 0.3; indeterminado	Log2FC: 0.009; indeterminado			= 3.412)
Gen	Expresión génica (mediana de log2(fpkm+1))				Prueba estadístico
	Variable: Subtipo viral				
	HPV18	HPV16	Otros		ANOVA de una vía
GLS2	0.157 (n=27)	0.374 (n=103)	0.182 (n=39)		p = 0.001247 (f = 6.964)

\*  $Log2FC \leq -2$  = subexpresado,  $Log2FC \geq 2$  = sobreexpresión,  $-2 < Log2FC < 2$  = indeterminado

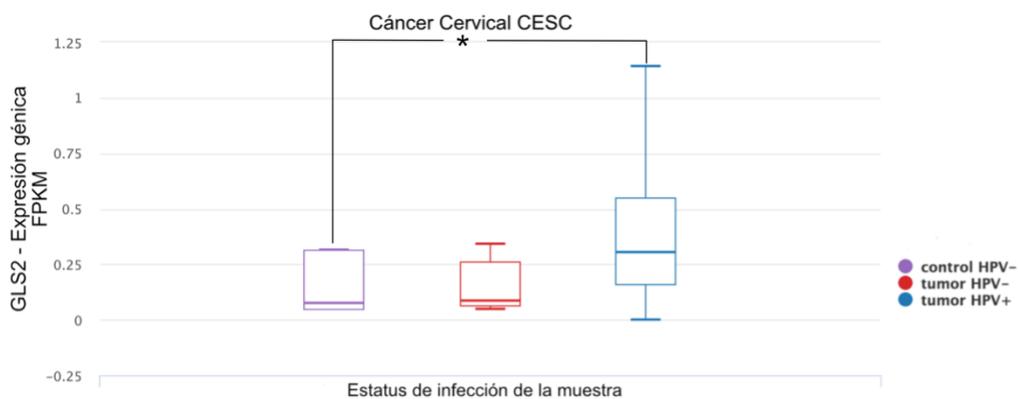


Figura 13. Gráfica de caja y bigotes de la expresión génica de GLS2 según el estado de infección por HPV (positivo o negativo) en muestras de tejido cervical (canceroso o normal).

El 93% (n = 169) de pacientes en la cohorte CESC (n = 181) presentaron infección por HPV. De ellas, la mayoría presentó infección por HPV16 (61%, n = 103), existiendo también casos de infección por HPV18 (16%, n = 27) y otros subtipos (23%, n = 39) (Figura 14).

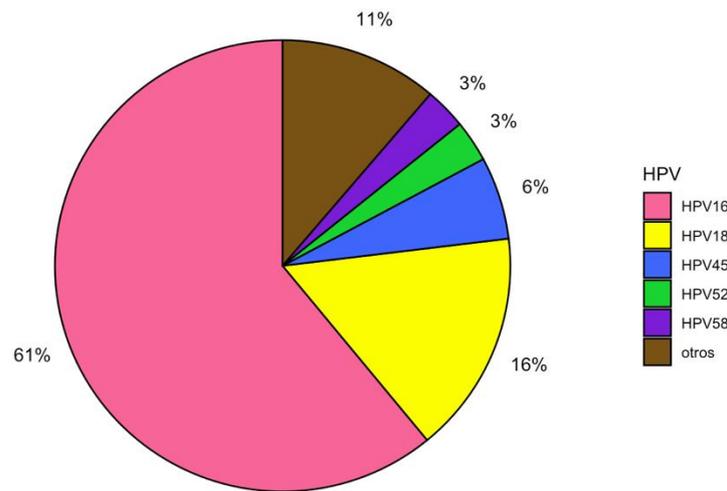


Figura 14. Distribución de distintos subtipos de HPV en la cohorte de cáncer cervical.

En pacientes con cáncer cervical, la infección por HPV16 se asocia con una mayor expresión de *GLS2* en comparación con las infecciones por otros subtipos virales ( $p < 0.01$ ) (Tabla 8, Figura 15). Al contrario, la expresión de *GLS1* no varió en función del subtipo viral ( $p > 0.05$ , Anexo 5).

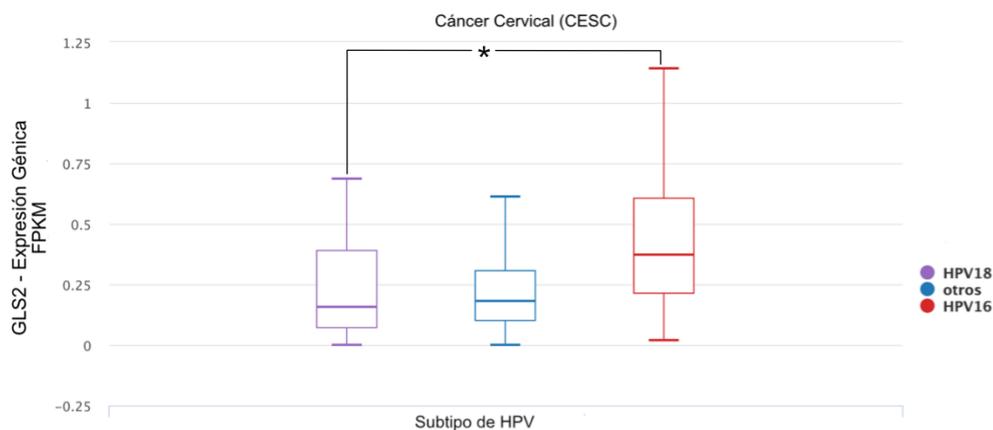


Figura 15. Gráfica de caja y bigotes de la expresión génica de *GLS2* según el subtipo viral de HPV en cáncer cervical.

El número de copias de *GLS1/GLS2* no se vio afectado por ninguna de las variables clínicas investigadas, incluyendo el estatus de infección de la muestra ni el subtipo de HPV (Anexo 6).

#### 5.4.2. Cáncer de cabeza y cuello

En los sujetos con HNSCC, la ausencia de la infección por HPV se asocia con mayor expresión de *GLS1* ( $p < 0.0001$ ), mientras que la infección por HPV en pacientes con HNSCC se asocia con una mayor expresión del gen *GLS2* ( $p < 0.0001$ ) (Tabla 9, Figura 16, Figura 17). El estatus de expresión de glutaminasas fue “indeterminado” a considerar por los valores de  $\log_2FC$ .

Tabla 9. Expresión génica de *GLS1* y *GLS2* agrupadas por variables clínicas y  $\log_2FC$  en la cohorte de HNSCC.

Gen	Expresión génica (mediana de $\log_2(fpkm+1)$ )				Prueba estadística
	Variable: Estatus de infección de la muestra				
	Tumor con infección HPV(+)	Tumor sin infección HPV(-)	Control sin infección HPV(-) (tejido no tumoral sin infección)	Control HPV(+) (tejido no tumoral con infección)	ANOVA de una vía
<i>GLS1</i>	2.87 (n=55)	3.04 (n=219)	2.31 (n=38)	2.23 (n=6)	p = 3.331e-16 (f = 28.17)
	Log <sub>2</sub> FC: 0.73; indeterminado	Log <sub>2</sub> FC: 0.77; indeterminado			
<i>GLS2</i>	0.389 (n=55)	0.0767 (n=219)	0.224 (n=38)	0.308 (n=6)	p = 7.262e-12 (f = 19.96)
	Log <sub>2</sub> FC: 0.14; indeterminado	Log <sub>2</sub> FC: -0.15; indeterminado			

\*\*  $\log_2FC \leq -2$  = subexpresado,  $\log_2FC \geq 2$  = sobreexpresión,  $-2 < \log_2FC < 2$  = indeterminado

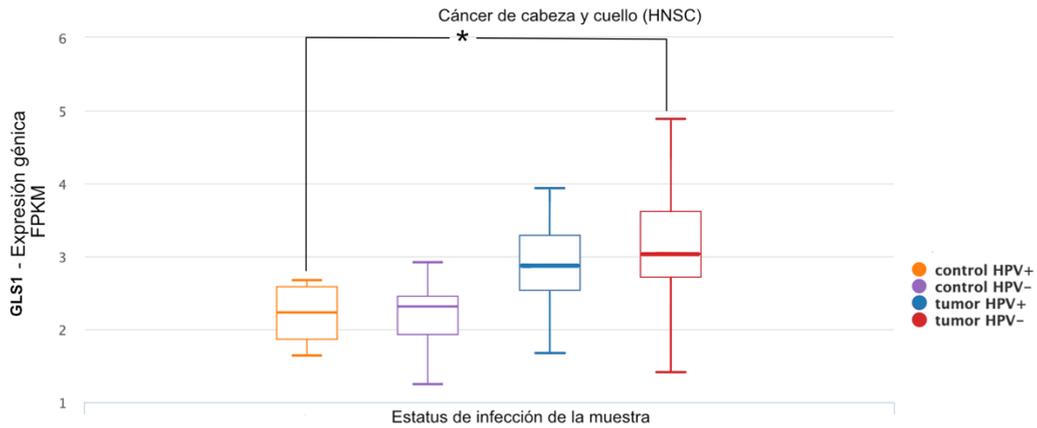


Figura 16. Gráfica de caja y bigotes de expresión génica de GLS1 según el estado de la infección (HPV+)/HPV(-)) y el tipo de muestra de tejido de cabeza y cuello (cancerosa o normal).

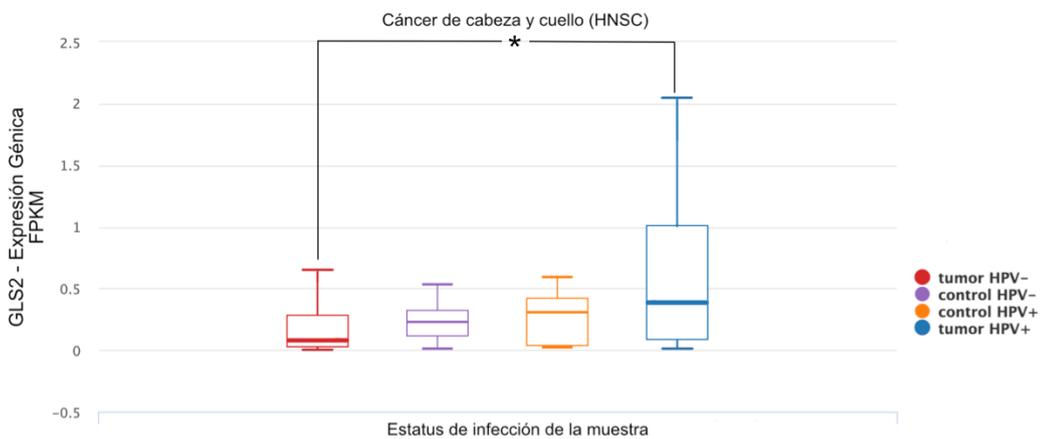


Figura 17. Gráfica de caja y bigotes de expresión génica de GLS2 según el estado de la infección (HPV+)/HPV(-)) y el tipo de muestra de tejido de cabeza y cuello (cancerosa o normal).

El 19% (n = 61) de pacientes en la cohorte HNSCC (n = 320) presentaron infección por HPV. De ellos, los subtipos virales más frecuentes fueron HPV 16 (69%, n = 42) y HPV 33 (25%, n = 15) siendo los subtipos 18 y 35 menos frecuentes (7%, n = 4) (Figura 18).

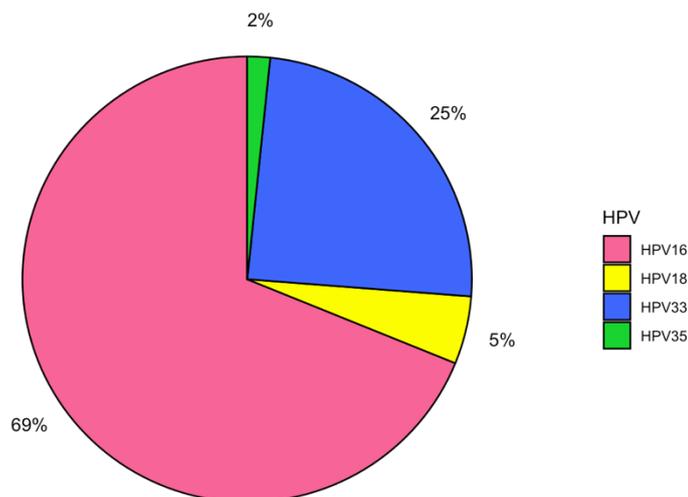


Figura 18. Distribución de los subtipos de HPV en la cohorte de cáncer de cabeza y cuello.

No hubo diferencias en la expresión de *GLS1* y *GLS2* en los diferentes subtipos de HPV dentro la cohorte de HNSCC ( $p > 0.05$ ) (Anexo 9).

En HNSCC no observamos cambio en el número de copias de glutaminasas con relación a las variables clínicas incluidas estatus de infección de la muestra y subtipo de HPV (Anexo 10).

### 5.4.3. Cáncer de hígado

Los sujetos con LIHC y con infección por el HBV o HCV tuvieron una expresión de *GLS1* similar a los pacientes con LIHC sin infección, la cual es significativamente mayor en comparación con los grupos control (no cancerosos con y sin infección). Al contrario, los sujetos con LIHC, con infección o sin ella, presentaron niveles significativamente menores de expresión de *GLS2*, comparados con los controles (no cancerosos con y sin infección) ( $p < 0.0001$ ). (Tabla 10, Figura 19, Figura 20). Al calcular el  $\text{Log}_2\text{FC}$ , se pudo determinar que *GLS2* estaba subexpresado en las muestras cancerosas con infección en relación con los tejidos sin cáncer (controles) con infección.

Tabla 10. Expresión génica, alteración del número de copias de *GLS1* y *GLS2* agrupados por las variables clínicas con diferencias significativas y el Log2FC en la cohorte de LIHC.

Gen	Expresión génica (mediana de log2(fpkm+1))				Prueba estadística
	Variable: Estatus de infección de la muestra				
	Tumor HBV/HCV (+)	Tumor HBV/HCV (-)	Control del tumor HBV/HCV (-)	Control del tumor HBV/HCV (+)	ANOVA de una vía
<i>GLS1</i>	2.09 (n=71)	2.07 (n=118)	1.06 (n=36)	1.07 (n=14)	p = 4.303e-10 (f = 17.13)
	Log2FC: 1.02; indeterminado	Log2FC: 1.007; indeterminado			
<i>GLS2</i>	0.368 (n=71)	0.399 (n=118)	2.32 (n=36)	2.96 (n=14)	p = 1.203e-12 (f = 22.11)
	Log2FC: -2.45; Subexpresión	Log2FC: -1.89; indeterminado			

\*\*  $\text{Log2FC} \leq -2$  = subexpresado,  $\text{Log2FC} \geq 2$  = sobreexpresión,  $-2 < \text{Log2FC} < 2$  = indeterminado

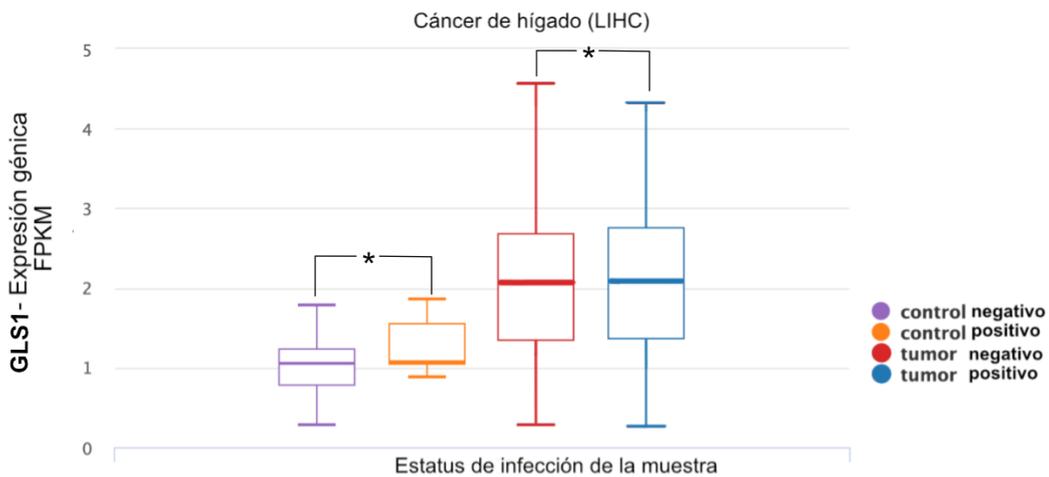


Figura 19. Gráfica de caja y bigotes para la expresión génica de *GLS1* del estado de la muestra y el estado de la infección en cáncer de hígado.

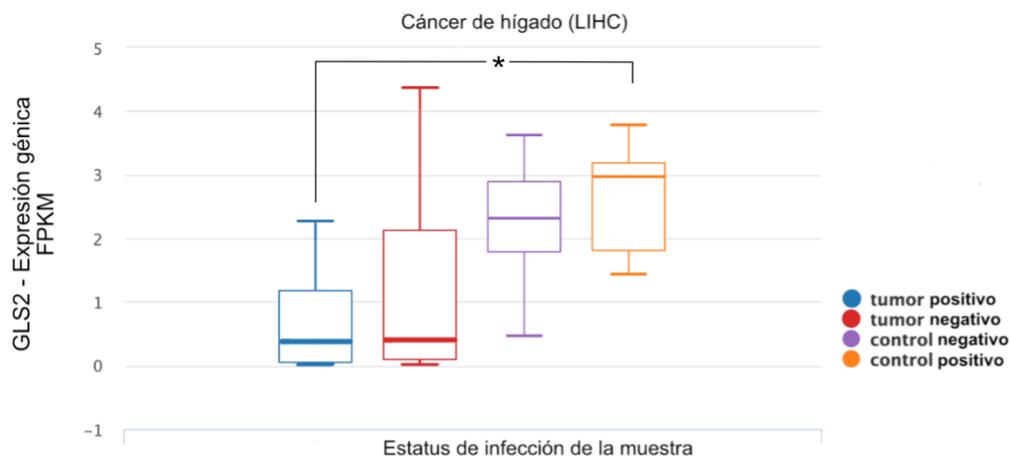


Figura 20. Gráfica de caja y bigotes para la expresión génica de *GLS2* del estado de la muestra y el estado de la infección en cáncer de hígado.

El 35% (n = 84) de los pacientes en la cohorte LIHC (n = 242) presentaron infección por HBV o HCV. De ellos, el 57% (n = 48) está infectado con HBV, mientras que el 43% está infectado con HCV (n = 36). No encontramos diferencias en expresión de *GLS1* y *GLS2* en función del agente viral (HBV o HCV) ( $p > 0.05$ ) (Anexo 12).

No se observó la ocurrencia de alteración del número de copias de *GLS1/GLS2* con las variables de interés incluyendo el estatus de infección de la muestra y el subtipo viral (HBV/HCV) (Anexo 13).

### 5.5. Influencia de la expresión génica de *GLS1/GLS2* y del estado de la infección viral (positivo/negativo) con la sobrevida global.

#### 5.5.1. Cáncer cervical

Al graficar la curva de KM resultó que las pacientes con CESC, cuya mediana de expresión de *GLS2* en tejido tumoral fue mayor de 0.59 (valor de  $\log_2(\text{fpkm}+1)$ ), tuvieron un mayor OS que las pacientes con valores de mediana de expresión menores a 0.15 ( $p < 0.05$ ) (Figura 21). No se encontró alguna asociación entre la expresión de *GLS1* y el OS en la cohorte de CESC.

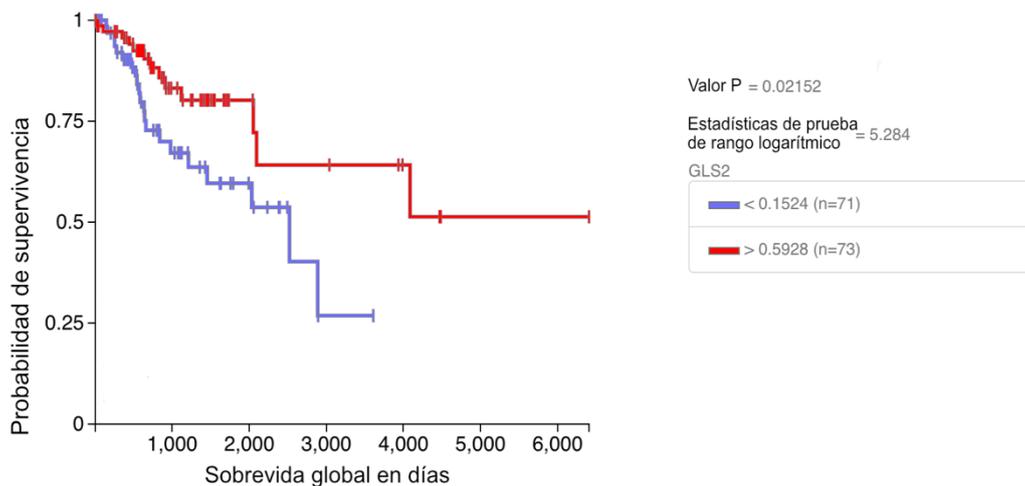


Figura 21. Curva de KM representando OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con CESC agrupadas por los niveles de expresión génica de GLS2. Valores de expresión agrupados en cuartiles.

Al evaluarse la influencia del estado de infección con OS en CESC, se apreció que la infección por HPV no afecta el pronóstico de la enfermedad ( $p > 0.05$ ), tampoco el subtipo de HPV ( $p > 0.05$ ).

### 5.5.2. Cáncer de cabeza y cuello

La gráfica de KM en la cohorte de HNSCC mostró que ni la expresión de *GLS1/GLS2* ni el estado de infección influyeron en OS ( $p > 0.05$ ).

Los pacientes con HNSCC con infección viral por HPV16 y HPV33 tuvieron menor OS que los pacientes con infección viral por otros tipos de HPV ( $p < 0.01$ ) (Figura 22).

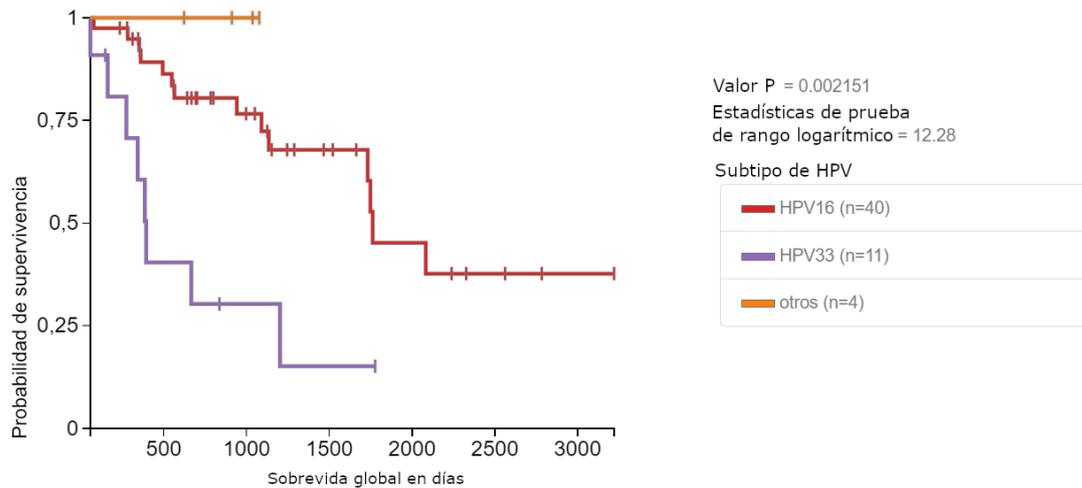


Figura 22. Curva de KM representando OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con HNSCC agrupados por tipo de HPV.

### 5.5.3. Cáncer de hígado

Para pacientes con LIHC, ni la expresión de *GLS1/GLS2* ni el estado de infección influyeron en OS ( $p > 0.05$ ). Sin embargo, los pacientes con LIHC y con infección por HCV tuvieron un mayor OS que los pacientes con infección por HBV ( $p < 0.05$ ) (Figura 23).

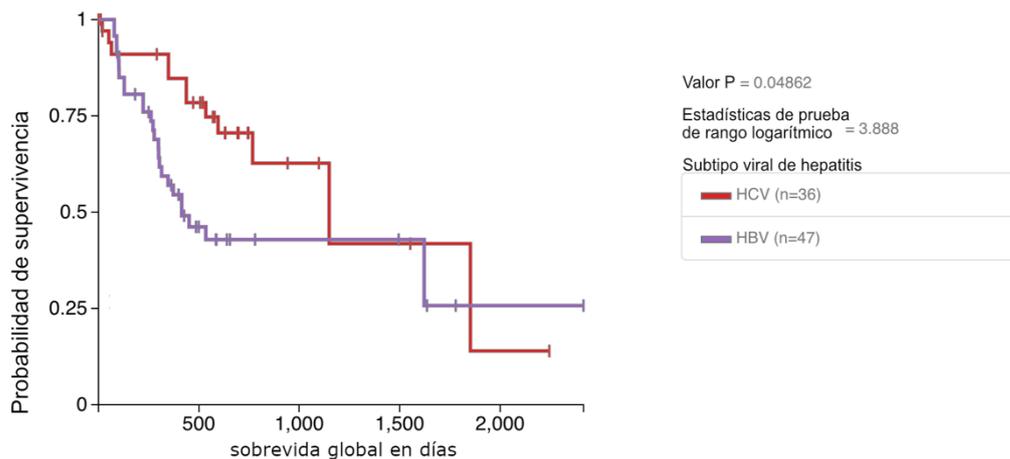


Figura 23. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con LIHC agrupados por tipo de virus de la hepatitis HCV o HBV.

## 6. Discusión

Las glutaminasas GLS1 y GLS2 son enzimas que participan de la glutaminólisis, proceso que permite a las células cancerosas complementar, junto con la glucólisis, el abastecimiento de energía necesario para la proliferación celular. La glutaminólisis es un proceso que genera energía y que se inicia con el ingreso de la glutamina a la mitocondria y su posterior conversión a glutamato, para finalmente integrarse al ciclo de Krebs (TCA) como  $\alpha$ -kG (7).

La glutamina, a su vez, es sustrato para la síntesis de aminoácidos no esenciales como alanina y aspartato, a partir de los cuales se sintetizan *de novo* nucleótidos. De hecho, en varios tipos de neoplasias, como LIHC, la conversión de glutamina en glutamato por acción de las glutaminasas está promovida, lo cual contribuye a mantener un pool de moléculas y un nivel energético necesario para la alta proliferación celular (7).

GLS1 y GLS2 juegan roles diferentes en el desarrollo de la carcinogénesis y en los resultados clínicos (*outcome clínico*). Por un lado, GLS1 puede ser usado como un blanco terapéutico y como promotor tumoral (12,16,62–64), mientras que GLS2

sería útil como un marcador de pronóstico y supresor de tumores dependiendo del tipo de cáncer (16,65,66).

Numerosas células cancerosas exhiben un notable apetito por la glucosa, prefiriendo producir ácido láctico en lugar de catabolizar la glucosa mediante el ciclo del TCA. Este último proceso desempeña un papel fundamental en la generación de ATP en células normales no expuestas a hipoxia. Este hecho se puede evidenciar al realizar una tomografía por emisión de positrones utilizando fluorodesoxiglucosa, en donde los tumores son los que brillan más con esta técnica, evidenciando su consumo ávido de glucosa comparado con los tejidos normales (4). Hace más de ocho décadas, Otto Warburg notó que se registraban elevadas tasas de absorción de glucosa y generación de lactato al examinar secciones delgadas de tumores humanos y animales en condiciones *ex vivo* (3). Es por esta razón que la transición hacia la producción de lactato en las neoplasias malignas, aunque estén en condiciones de oxigenación adecuada se le conoce como Efecto Warburg o glucólisis aeróbica (4).

En esta tesis, buscamos determinar si la expresión de los genes de glutaminasas *GLS1* y *GLS2* variaba en función del estado de infección viral en pacientes con CESC, HNSCC y LIHC. También nos propusimos determinar si la alteración de la expresión génica de *GLS1/GLS2* y el estado de infección estaban asociados al sobrevida global.

En lo que concierne a cáncer cervical, no se encontraron diferencias en la expresión de *GLS1* y *GLS2*, entre tejidos tumorales y tejidos normales. Anteriormente ya se había investigado el cambio de expresión de ambas enzimas en CESC, encontrándose un menor nivel de expresión de *GLS1* y una mayor expresión de *GLS2* en tejidos con cáncer que en tejidos sin cáncer (16). Es posible que la discrepancia entre nuestros resultados y los de Saha et al. (2019) se deban a que ellos tuvieron en cuenta, no solo la base de datos del TCGA, sino también de GTEx (*The Genotype-Tissue Expression Project*).

Así mismo, en CESC, observamos que las infecciones por HPV aumentan la expresión de *GLS2*, mas no la de *GLS1*. Este hallazgo dista del reportado por Pyeon et al. (2007), quienes encontraron que el estado viral HPV(+) no conlleva a un aumento de la expresión de *GLS2*, comparado con el cáncer cervical HPV(-) (67). La discrepancia entre estos hallazgos puede deberse, por un lado, al tamaño de la muestra, que en nuestro estudio fue de 131 pacientes de la cohorte CESC TCGA, siendo mayor que la del estudio de D. Pyeon et al. (2007). Además, en dicha investigación se analizaron datos de expresión obtenidos por microarreglos de RNA, mientras que nosotros analizamos datos de RNAseq (30).

Por otro lado, la prevalencia del HPV en la cohorte CESC (n = 181) estudiada fue del 93% (n = 169), lo cual es un hecho bien reconocido ya que el HPV es un virus oncogénico, responsable de la mayoría de los casos de transformación oncogénica en CESC (24,25,68,69). La distribución de subtipos virales en la cohorte CESC estudiada también está en concordancia con estudios previos, siendo más común la infección por HPV16 que por el resto de subtipos (70,71).

Por otro lado, hasta donde sabemos, es la primera vez que se demuestra la asociación entre la infección por HPV16 y la sobreexpresión de *GLS2*, en comparación con el resto de los subtipos de HPV en CESC. El subtipo HPV16 secreta la oncoproteína E7 (HPV16-E7), la cual promueve el flujo glucolítico a través de una estimulación directa a la enzima PKM2 (72) y activa la glutaminólisis, promoviendo la vía anaplerótica, tal cual lo haría un tumor (73). Por lo tanto, cabe la posibilidad que la activación de la glutaminólisis por acción del subtipo HPV16 pueda ocurrir a través de la sobreexpresión de *GLS2*, conduciendo a contar con una fuente de energía extra en las células infectadas a través de la degradación de glutamina.

Se ha encontrado que las oncoproteínas E6 y E7 aumentan la proliferación celular mediante la glutaminólisis en HPV16. Ambas oncoproteínas aumentan los niveles del transportador 1 de aminoácidos neutros acoplado al sodio (SNAT1), *GLS2* y Glutamina Sintetasa (GS). E6 incrementa los niveles del transportador de

aminoácidos tipo L 1 (LAT1), mientras que E7 aumenta transportador 2 de alanina, serina y cisteína (ASCT2) y Transportador de cisteína/glutamato (xCT). Estos transportadores son importantes para la entrada de la glutamina dentro de la célula, mientras que la glutamina también puede obtenerse por la actividad de la GS (74).

Ortiz-Pedraza et al. (2013) también encontraron que el transportador SNAT1 disminuye en células Ca Ski cuando la expresión de E6 y E7 fue noqueada (74). Las células Ca Ski es una línea celular infectada con HPV16 que sirve como modelo para el estudio en CESC avanzado (75). Finalmente, encontraron que la expresión de *SNAT1* era significativamente más alta en tejidos con CESC comparados con tejidos normales, además de estar asociada a un pobre OS en CESC (74).

Un hallazgo remarcable de nuestra investigación fue la influencia de la expresión de *GLS2* sobre la OS en las pacientes con CESC, en donde las mujeres tienen mejor pronóstico si los valores de expresión en mediana de  $\log_2(\text{fpkm}+1)$  de *GLS2* son mayores a 0.59. Este hallazgo es novedoso en CESC, aunque ha sido descrito en otros tipos de cáncer que un aumento de la expresión de *GLS2* se asocia con un pobre OS en varios tipos de cáncer, incluyendo ovario, leucemias, timoma (carcinoma de timo) y colon. Esta diferencia con nuestros resultados puede deberse al tipo de bases de datos consultada en otros estudios como PROGeneV2database (16).

Por otro lado, los pacientes con HNSCC tienen mayor expresión de *GLS1* en comparación con los que no tienen cáncer. Nuestros resultados concuerdan con el estudio multiómico *in silico* de Saha et al. (2019) donde se observó que la expresión relativa de mRNA en *GLS1* era mayor en HNSCC que en tejidos normales (16). El incremento de la expresión de *GLS1* podría estar asociado con un peor control tumoral, tal como se demostró en un estudio donde la inhibición de *GLS1* con telaglenastat, en líneas celulares y modelos de ratones con cáncer de HNSCC, reduce la proliferación celular (76).

La expresión de *GLS1* en HNSCC ha sido investigada previamente en algunos estudios cuyos resultados también concuerdan con los nuestros. Entre ellos está el estudio de Cetindis et al. (2016), los cuales usaron PCR en tiempo real en especímenes tumorales (77). También el de Zhang et al. (2020) y el de Basheer et al. (2022), los cuales utilizaron inmunohistoquímica en especímenes tumorales (78,79). Este último también encontró que *GLS2* tiene una expresión más alta en HNSCC que en tejidos normales, lo cual difirió con nuestros resultados. La razón por la cual nuestros resultados difirieron es debido a que analizaron la expresión de glutaminasas a nivel proteínico y no a nivel genómico. Ellos usaron 80 muestras que variaban desde de normales a metastáticas en tejidos de microarrays preparadas por la empresa US Biomax Inc (79).

Similar a nuestros resultados, Yang et al. (2019) encontraron que los niveles de expresión de *GLS1* son significativamente altos en HNSCC en líneas celulares de faringe mediante western blot. Ellos hipotetizaron que el inhibidor selectivo de *GLS1*, BPTES, el cual recorta el consumo de glutamina de la célula, inhibiría el crecimiento de las células con HNSCC. Entonces, al inhibir *GLS1* usando BPTES, se incrementaron los niveles de especies oxidativas y también los niveles de mRNA de varios piruvatos deshidrogenasas (*PK2/3/4*). Esto es causado por un aumento de la expresión de *p21*, un inhibidor del ciclo celular (29). Tomando en cuenta los estudios de los inhibidores de glutaminasas, se deduce que la suplementación de glutamina promueve el crecimiento celular en HNSCC (29,76).

*c-MYC* es un oncogén que acelera la glutaminólisis facilitando la expresión de las glutaminasas, especialmente en *GLS1*, donde *c-MYC* está sobreexpresado durante la transcripción y post-transcripción. La sobreexpresión de *c-MYC* hace que las células sean sensibles a la deprivación de glutamina. Por ejemplo, altos niveles de anabolismo de glutamina promueven el crecimiento tumoral y la resistencia a la radioterapia (80).

Uno de los factores que causan la carcinogénesis de HNSCC es la infección por HPV (81–83). En esta cohorte, 19% de los pacientes con HNSCC (n = 320) presentaron infección por HPV, siendo menor de lo esperado por tratarse de participantes principalmente caucásicos solo de Estados Unidos. De hecho, se estima que los países occidentales tienen una prevalencia del 40–60% de HPV en casos de HNSCC, en comparación con países como China e India, donde la prevalencia fluctúa entre 13% y 24% (84).

Así mismo, los subtipos virales más comunes en la cohorte HNSCC estudiada fueron HPV16 (69%) y HPV33 (25%). Esto fue documentado previamente en otro estudio, donde los cánceres de cabeza y cuello ubicados en la cavidad oro faríngea, provenientes de Europa y Norte América, mostraron que el HPV16 y el HPV33 son los subtipos virales más comunes (85). La prevalencia de HPV16 en HNSCC en China (71%, n = 80 HPV(+)) es similar al observado en nuestro estudio, donde representa el subtipo viral más común, seguido del subtipo HPV18 con 5% (86). Esto indica que, al igual que en CESC, el subtipo de HPV más común es el 16 (87). El HPV16 y los subtipos oncogénicos de alto riesgo, entre ellos el HPV18 y el 33, poseen oncoproteínas virales como E6 y E7. La E6 se une por un lado a una ubiquitina/proteína ligasa celular (E6-AP) y, simultáneamente, también se asocia a p53, una proteína supresora de tumores. Esto resulta en la ubiquitinación de p53 y su degradación proteolítica, con la consecuente disrupción de los mecanismos reguladores y reparadores de daño en el DNA (87). Por su parte, E7 se une y desestabiliza la proteína supresora de tumores retinoblastoma (pRb). Eso previene que pRb se una al factor de transcripción E2F y, en consecuencia, que se promueva la progresión del ciclo celular. La inactivación de pRb resulta en la sobreexpresión de la proteína supresora de tumores de p16 llamada p16INK4A (87).

El hallazgo de la expresión reducida de *GLS1* en los individuos que tienen infección por HPV, en comparación con los pacientes sin infección, es novedosa en HNSCC. Krishna et al. (2020) encontraron lo opuesto en otro virus oncogénico. Las células infectadas por el virus del Epstein-Barr (EBV) tienen una mayor expresión de *GLS1*,

la cual está asociada a la infección de EBV. Esto claramente difiere de nuestros resultados debido a que Krishna et al. (2020) usaron líneas celulares con linfoma de Burkitt' EBV(+) y de carcinoma de epitelio nasal los cuales no son lo mismo al HPV con excepción de ser oncogénico. Además, la expresión fue cuantificada por RT-qPCR y midieron los niveles de glutamina y  $\alpha$ -kG (88).

Krishna et al. (2020) encontraron un aumento de los niveles intracelulares de glutamato en líneas celulares EBV y el tratamiento con BPTES, un inhibidor de la glutaminólisis, redujo estos niveles. En consecuencia, también se incrementó los niveles de  $\alpha$ -kG, ya que el glutamato, resultante de la catálisis de glutamina por GLS1, es desaminado por las enzimas glutamato deshidrogenasa 1 y 2 (GLUD1/2) en la mitocondria para formar  $\alpha$ -kG, un compuesto importante para la generación de energía por el TCA. Sus resultados tienen sentido, ya que el aumento de expresión de *GLS1* juega un rol en iniciar y mantener glutaminólisis, la cual impulsa la generación de energía mediante el ciclo de TCA en células infectadas por EBV (88).

Al mismo tiempo, la expresión de *GLS2* es mayor en pacientes con HNSCC HPV(+) que en los pacientes con HNSCC HPV(-). Este hallazgo también es novedoso. Al analizar la expresión de *GLS2* en la línea celular Huh7.5 y en hepatocitos primarios infectados (HPIs) con HCV, Lévy et al. (2016) no encontró diferencias significativas en su expresión. Solo se encontró que HCV indujo un aumento en los niveles de transcritos de *GLS1* en Huh7.5 y en HPIs (21).

En los estudios sobre cáncer resulta una prioridad encontrar factores que influyen en el pronóstico de la enfermedad, lo cual fue de interés en nuestro estudio. Por un lado, no encontramos relación entre la infección y el pronóstico. Sin embargo, Alsbeih et al. (2019) encontraron una mejor OS para HNSCC con HPV(+) que el HPV(-), aunque los resultados no fueron significativos ( $p = 0.12$ ). Los autores concluyen que es posible que la baja cantidad de muestras con HPV(+) ( $n = 10$ ) pudo haber influido en este resultado (89).

Otros estudios como el de Kian et al. (2010) (HPV(+) = 206, HPV(-) = 117) también han encontrado asociación entre el estatus de infección y la OS, en donde, los pacientes con HNSCC orofaríngeo HPV(+) tienen un mejor pronóstico que los pacientes con HNSCC HPV(-) (90). La diferencia con el estudio de Alsbeith et al. (2019) es en la cantidad de muestras, lo cual hace este hallazgo significativo. El estudio de Fakhry et al. (2017) también encontró este mismo resultado, sin embargo no encontraron asociación entre estatus de HPV y el OS en HNSCC no orofaríngeos (91).

Estos estudios difieren con nuestros resultados, probablemente porque las muestras provienen de pacientes de diferentes hospitales de Arabia Saudita, Reino Unido y los Estados Unidos (89–91), mientras que el nuestro proviene de la base de datos de TCGA. Nosotros tampoco estudiamos la asociación entre el OS y estatus de HPV por subtipos de HNSCC (91).

Al contrario, nosotros encontramos que el subtipo de HPV se asocia con la OS, donde los pacientes con HNSCC con infección por HPV16 tuvieron una mayor supervivencia que los que tenían una infección con HPV33. Esto concuerda con hallazgos previos en donde la infección con HPV33 estuvo asociada con un pobre pronóstico en HNSCC (92). Según Chatfield-Reed et al. (2020), es posible que varios factores influyan en reducir la OS cuando hay infección por HPV33 en comparación con HPV16 (92). Uno de esos factores sería la firma mutacional asociada a las enzimas APOBEC, que está más enriquecida en tumores HPV16(+) a comparación de tumores HPV33(+). Dicha firma mutacional está asociada al daño en el ADN en presencia de la infección viral. En relación a ello, otro factor que influye en el peor pronóstico sería las pérdidas del brazo cromosómico 3p a una tasa más alta en tumores HPV33(+) que en tumores HPV16(+), lo cual sumado a la mutación de p53 representan marcadores de peor pronóstico (92).

En nuestro estudio no encontramos asociación entre la expresión de *GLS1* o *GLS2* y el OS de los pacientes con HNSCC. Sin embargo, otras investigaciones sí han

documentado que la reducción de expresión de *GLS1* se asocia con un mejor pronóstico en pacientes con HNSCC (16,78). Esa discrepancia podría deberse, por un lado, a las diferentes unidades de abundancia de mRNA empleadas, siendo  $\log_2(\text{FPKM}+1)$  en nuestro estudio y la expresión relativa de mRNA en un estudio previo (16) y por otro lado el análisis de cohortes distintas a TCGA (78).

Por otro lado, en esta tesis también analizamos la expresión de glutaminasas en cáncer de hígado, apreciando que la expresión de *GLS1* es mayor en tejido canceroso que en el tejido normal. Algo similar se describió anteriormente en un modelo de ratones infectados con HBV crónica y hepatocarcinoma celular, en quienes la actividad enzimática de *GLS1* era mucho mayor en tumores mayores a 3 mm (22). También Suzuki et al. (2022) llegaron a los mismos resultados al examinar la base de datos del TCGA, en donde los niveles de mRNA de *GLS1* son un poco más altos en pacientes con HCC (66). Al contrario, nosotros determinamos que el gen *GLS2* está subexpresado en LIHC en comparación con el grupo control. Estos hallazgos son concordantes con otros cuatro estudios incluido el de Saha et al. (2019) y Suzuki et al. (2022) quienes analizaron la expresión de mRNA relativa en la misma cohorte de TCGA (LIHC), el de Chen et al. (2016) con qPCR y Nwosu et al. (2017) con data de microarrays (11,16,17,66).

Por su parte, *GLS2* también ha sido vinculado con la ferroptosis, un tipo de muerte celular programada, dependiente del hierro y los peróxidos de lípidos, que es distinto de la apoptosis (93). Usando hepatocitos primarios de ratones *knockout* de *Gls2* y de líneas celulares humanas (SKHep1 WT, HepG2) derivadas de HCC, Suzuki et al, 2022 encontraron que la ferroptosis es necesaria para la supresión de tumores derivados de células que expresan *GLS2* en ratones. También encontraron que los tumores de pacientes humanos y de ratones con HCC muestran bajos niveles de *GLS2*. Todo esto sugiere que *GLS2* tiene un rol modulador en la ferroptosis y como consecuencia, en la supresión de tumores (66).

Esto tiene sentido con nuestros resultados, ya que, al estar *GLS2* subexpresado en LIHC, cabe la posibilidad que haya una supresión de la ferroptosis y, en consecuencia, la proliferación de células cancerosas con HCC. Se ha postulado que la ferroptosis influye en el desarrollo de HCC en ratones *knockout* de *Gls2* (94–96). También la data de Suzuki et al. (2022) indica que la ferroptosis está envuelta en prevenir la aparición de HCC y en su progresión (66).

Por otra parte, se sabe que una de las causas de la carcinogénesis de LIHC es la infección por HBV o HCV (26,97,98). En esta cohorte, más de la mitad de los pacientes (65%) no tenían infección por virus de hepatitis, mientras que quienes sí presentaban la infección, esta era por HBV (57%) o HCV (43%). Ello concuerda con la mayor prevalencia de infección por HBV que HCV descrita en población caucásica, como la que compone la cohorte TCGA analizada.

En este trabajo encontramos que los pacientes con LIHC tuvieron una expresión más alta de *GLS1* que los controles, al mismo tiempo que una menor expresión de *GLS2* comparado con los controles. *GLS2* es una isoforma que se expresa en el hígado bajo condiciones normales (21). Nuestros resultados sugieren que el desarrollo de la neoplasia hepática disminuye la actividad de *GLS2* en hígado, lo cual podría reducir la tasa de glutaminólisis, aunque dicho efecto podría ser contrarrestado en vista del incremento de niveles de *GLS1* en el tejido neoplásico.

Por otro lado, se observó que *GLS2* estaba subexpresado en muestras cancerosas con infección de HBV/HCV en comparación con controles infectados. Sin embargo, no se observaron diferencias en la expresión de *GLS2* en muestras cancerosas sin infección en comparación con controles no infectados. Sin embargo, un estudio *in vitro* demostró que el HCV inducía un aumento en los niveles de transcritos de mRNA de *GLS1*, aunque no de *GLS2* (21). Otro estudio de Suzuki et al. (2023) examinó la base de datos *Gene Expression Omnibus* (GEO) el cual reveló que los niveles de mRNA de *GLS2* no eran significativamente diferentes en tejidos de hígado infectados con HBV/HCV comparados con tejido de hígado sano. Sin embargo, ellos

no distinguen si los tejidos de hígado infectados provienen de HCC o de tejido sin cáncer (99).

En esta tesis, no apreciamos alguna alteración de la expresión génica de *GLS1* o *GLS2* en función del estado de infección en LIHC. Sin embargo, sí se observó cambios de glutaminasas en CESC y HNSCC dependientes de la infección por HPV, por lo que es posible que este virus se valga de un ajuste metabólico mediado por *GLS2* en CESC y HNSCC.

En nuestro estudio, reportamos que, en los pacientes con LIHC existe una asociación entre el tipo viral y la OS de los pacientes. Así, los pacientes con LIHC y con infección por HCV tuvieron una mayor sobrevida que los pacientes con LIHC y con infección por HBV. Esto coincide con el estudio de Lee et al. (2021) en donde la OS de pacientes con LIHC HCV(+) era superior al de los pacientes con LIHC HBV(+), cuando los pacientes no recibían terapia antiviral. El escenario se invertía al usar terapia antiviral, donde la OS de LIHC infectados con HBV(+) aumentaba con respecto a los pacientes infectados con HCV (100). En nuestro trabajo, no contamos con información del tratamiento antiviral por lo que esta variable no fue investigada. Estudios previos se han interesado por averiguar la influencia de la infección viral y el pronóstico del cáncer de hígado. Se sabe, por ejemplo, que la resección hepática es un factor que influye en la OS, siendo menor en pacientes HCC-HBV(+) que en HCC-HCV(+) que tuvieron dicho procedimiento (101).

Así mismo, Cantarini et al. (2006) reveló que los pacientes con HCC-HCV(+) tuvieron mejor OS que los HCC-HBV(+), tal como lo demostramos en esta tesis. Sin embargo, las diferencias reportadas anteriormente ocurren cuando el HCC estaba en estado avanzado (tumor >5 cm) (102). Para poder explicar por qué la diferencia de sobrevida entre HCC-HCV(+) y HCC-HBV(+) diverge cuando se llega a un estadio avanzado, los investigadores proponen que esto es parcialmente explicado por una mutación en los receptores de beta-estrógenos ( $\beta$ -vER), la cual es más común en

HCC-HBV(+). Esta mutación ocurre en estados tardíos de la progresión tumoral, lo cual llevaría a tumores más agresivos en HCC-HBV(+) (102,103).

En tanto, no encontramos que la expresión de *GLS1/GLS2* influya en la OS en pacientes con LIHC. Sin embargo, otros estudios discrepan de nuestros resultados. Así, Saha et al. (2019) demostraron que una baja expresión de *GLS1* está positivamente correlacionados con un pobre pronóstico en LIHC (16). Es posible que las diferencias entre ambos resultados se deban a que, usaron datos de la base de datos del TCGA-LIHC y GSE20017 (104). Además, al procesar la data, no eliminaron a pacientes sin presencia de infección por HBV/HCV y con co-infección de HBV/HCV, lo cual sí se realizó en el presente trabajo. Así mismo, en el Kaplan-Meier Plotter (105), la plataforma usada por Saha et al. (2019), solo toma en cuenta pacientes con historia clínica disponible. Todos estos factores llevaron a un número superior de muestras analizadas por Saha et al. (2019) (n = 363 vs n = 242) (16).

Yu et al. (2015) identificaron que, en pacientes diagnosticados con HCC, aquellos que presentaban niveles elevados de expresión de *GLS1* experimentaban una baja sobrevida, mientras que aquellos con una alta expresión de *GLS2* en los tejidos tumorales mostraban una prolongada sobrevida. Nosotros no encontramos dicha asociación por lo que esta discrepancia podría deberse a que en dicho estudio midieron la expresión con un ensayo de inmunoreactividad contra *GLS1* y *GLS2* en el citoplasma de células con HCC mientras que en nuestro caso trabajamos con datos transcriptómicos de la expresión génica de *GLS1* y *GLS2* (106).

Suzuki et al. (2022) también realizaron un análisis de sobrevida usando Kaplan-Meier Plotter donde dividieron la cohorte en baja expresión de *GLS2* bajo la mediana y en alta expresión de *GLS2* sobre la mediana. Por estas razones, sus resultados demostraron que los niveles de baja expresión de *GLS2* están significativamente correlacionados con un pobre OS en pacientes con HCC (16,66).

En otro estudio, también se demostró que los pacientes con cáncer de hígado y con alta expresión de *GLS2* tienden a tener mejor pronóstico que aquellos con baja expresión de *GLS2*. En dicho trabajo se analizó una cohorte en Shanghái, no los datos de TCGA como en nuestro trabajo (17). Dichas diferencias en los procedimientos de estudio pueden explicar la incongruencia entre nuestros hallazgos y lo documentado con anterioridad. En este sentido, las cohortes de TCGA han sido ampliamente analizadas y contienen datos confiables y validados, por lo que nuestros resultados están respaldados por la solidez de los datos.

Por último, nuestros resultados sugieren que el número de copias de *GLS1/GLS2* no varía en CESC, HNSCC ni en LIHC, es decir, no parece ser una estrategia de alteración genética en ninguno de estos tres tipos de cáncer. En cambio, la expresión génica de *GLS1/GLS2* sí varía en pacientes que sufren de estos cánceres, en comparación con individuos sanos. Esto quiere decir que, en CESC, HNSCC y LIHC el número de copias de *GLS1/GLS2* se mantienen intactos, pudiendo ser suficiente los cambios en la expresión génica para efectos de la modulación del metabolismo de la glutamina.

## **7. Conclusiones**

Las neoplasias de cérvix, de cabeza y cuello y de hígado presentan cambios en la expresión génica de *GLS1* o *GLS2*, pero mantienen intacto el número de copias de ambos genes.

*GLS2* está subexpresado en tejidos cancerosos de LIHC en comparación con tejidos no cancerosos. La presencia de infección por HCV/HBV está asociada con la subexpresión génica de *GLS2* en LIHC. Esto refuerza la posibilidad que HBV o HCV modulen la expresión génica y el metabolismo de la glutamina al infectar crónicamente a las células eucariotas, conduciendo así a la transformación celular.

En la cohorte de CESC, la expresión de *GLS2* fue proporcional a la OS, estando esta última también relacionada con el subtipo viral. Así mismo, los pacientes de la cohorte de HNSCC infectados con HPV16 mostraron una mayor OS que los

pacientes con HNSCC infectados con HPV33. Similarmente, los pacientes de LIHC infectados con HCV tienen una OS mayor que los pacientes con LIHC infectados con HBV.

La principal contribución de este estudio es haber identificado los patrones de expresión génica de *GLS1/GLS2* con algunos factores clínicos de interés como la presencia infección viral y subtipo viral. También el detectar si existe una asociación entre la OS y otros factores, como la expresión de glutaminasas y el subtipo viral. En un futuro, se podría indagar si la expresión génica de la *GLS2* sería un posible candidato a marcador de supervivencia global en CESC.

## **8. Limitaciones**

- Las muestras controles son tejidos de pacientes con cáncer, que fue denominado “normal” al análisis histopatológico. Sin embargo, estos tejidos no son totalmente sanos y pueden tener cierto grado de inflamación lo cual puede llevar a tener ciertos sesgos en los resultados.
- Existe asimetría entre el tamaño muestral de varios grupos de estudio. Por ejemplo, en el grupo de CESC con 178 muestras neoplásicas vs 3 muestras no tumorales. En HNSCC con 6 muestras no tumorales con presencia viral de HPV, lo cual fue insuficiente para la variable estatus de infección de la muestra.
- No contamos con la variable de carga viral, la cual podría ser un factor influyente en la expresión de glutaminasas en muestras con presencia de infección viral.
- Tampoco se contó con secuencias virales en los genomas y transcritos del TCGA como dato complementario al subgenotipo viral.

## **9. Recomendaciones**

- Explorar otras cohortes de pacientes con cáncer distintas de TCGA, para corroborar los hallazgos de la presente tesis.
- Para un futuro estudio sería recomendable contar con la carga viral y su relación con la expresión de glutaminasas.

- Se sugiere realizar estudios con genómica funcional para encontrar otros indicios de interacción. Por ejemplo, realizar un *circus plot* o una red de interacciones proteicas (*protein network*) para identificar los genes target de las glutaminasas.
- Buscar secuencias virales en los genomas y transcritos del TCGA como dato complementario al subgenotipo viral. Esto también ayudaría a identificar diferencias génicas incluso en virus con el mismo subgenotipo.

## 10. Referencias Bibliográficas

1. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646–74.
2. Semenza GL, Artemov D, Bedi A, Bhujwalla Z, Chiles K, Feldser D, et al. “The metabolism of tumours”: 70 years later. *Novartis Found Symp*. el 1 de enero de 2001;240:251–60.
3. Warburg O. On the origin of cancer cells. *Science*. el 24 de febrero de 1956;123(3191):309–14.
4. Kim JW, Dang CV. Cancer’s molecular sweet tooth and the warburg effect. *Cancer Res*. el 15 de septiembre de 2006;66(18):8927–30.
5. Frezza C, Gottlieb E. Mitochondria in cancer: Not just innocent bystanders. *Semin Cancer Biol*. febrero de 2009;19(1):4–11.
6. DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, et al. Beyond aerobic glycolysis: Transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc Natl Acad Sci*. el 4 de diciembre de 2007;104(49):19345–50.

7. Mates JM, Segura JA, Martin-Rufian M, Campos-Sandoval JA, Alonso FJ, Marquez J. Glutaminase Isoenzymes as Key Regulators in Metabolic and Oxidative Stress Against Cancer. *Curr Mol Med.* el 1 de abril de 2013;13(4):514–34.
8. Xiang Y, Stine ZE, Xia J, Lu Y, O'Connor RS, Altman BJ, et al. Targeted inhibition of tumor-specific glutaminase diminishes cell-autonomous tumorigenesis. *J Clin Invest.* el 1 de junio de 2015;125(6):2293–306.
9. Cluntun AA, Lukey MJ, Cerione RA, Locasale JW. Glutamine Metabolism in Cancer: Understanding the Heterogeneity. *Trends Cancer.* el 1 de marzo de 2017;3(3):169–80.
10. Kodama M, Oshikawa K, Shimizu H, Yoshioka S, Takahashi M, Izumi Y, et al. A shift in glutamine nitrogen metabolism contributes to the malignant progression of cancer. *Nat Commun.* 2020;11(1):1–13.
11. Nwosu ZC, Megger DA, Hammad S, Sitek B, Roessler S, Ebert MP, et al. Identification of the Consistently Altered Metabolic Targets in Human Hepatocellular Carcinoma. *Cmgh.* 2017;4(2):303-323.e1.
12. Wang JB, Erickson JW, Fuji R, Ramachandran S, Gao P, Dinavahi R, et al. Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. *Cancer Cell.* 2010;18(3):207–19.
13. Gao P, Tchernyshyov I, Chang TC, Lee YS, Kita K, Ochi T, et al. C-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature.* 2009;458(7239):762–5.
14. Suzuki S, Tanaka T, Poyurovsky MV, Nagano H, Mayama T, Ohkubo S, et al. Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. *Proc Natl Acad Sci U S A.* 2010;107(16):7461–6.

15. Hu W, Zhang C, Wu R, Sun Y, Levine A, Feng Z. Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. *Proc Natl Acad Sci U S A*. 2010;107(16):7455–60.
16. Saha S, Islam SM, Abdullah-AL-Wadud M, Islam S, Ali F, Park K. Multiomics Analysis Reveals that GLS and GLS2 Differentially Modulate the Clinical Outcomes of Cancer. *J Clin Med*. 2019;8(3):355.
17. Chen S, Zhu M, Ling Q, Li N, Zheng J, Yu K, et al. Glutaminase 2- a valuable predictor for hepatocellular carcinoma patients' survival. *Int J Clin Exp Med*. 2016;9(9):17322–33.
18. Thai M, Thaker SK, Feng J, Du Y, Hu H, Ting Wu T, et al. MYC-induced reprogramming of glutamine catabolism supports optimal virus replication. *Nat Commun*. el 12 de noviembre de 2015;6.
19. Valiya Veetil M, Dutta D, Bottero V, Bandyopadhyay C, Gjyshi O, Sharma-Walia N, et al. Glutamate Secretion and Metabotropic Glutamate Receptor 1 Expression during Kaposi's Sarcoma-Associated Herpesvirus Infection Promotes Cell Proliferation. *PLoS Pathog* [Internet]. el 1 de octubre de 2014 [citado el 16 de julio de 2020];10(10). Disponible en: </pmc/articles/PMC4192595/?report=abstract>
20. Zhu Y, Li T, Da Silva SR, Lee JJ, Lu C, Eoh H, et al. A critical role of glutamine and asparagine  $\gamma$ -Nitrogen in nucleotide biosynthesis in cancer cells hijacked by an oncogenic virus. *mBio*. el 1 de julio de 2017;8(4):1179–96.
21. Lévy PL, Duponchel S, Eischeid H, Molle J, Michelet M, Diserens G, et al. Hepatitis C virus infection triggers a tumor-like glutamine metabolism. *Hepatology*. el 1 de marzo de 2017;65(3):789–803.
22. Bard-Chapeau EA, Nguyen AT, Rust AG, Sayadi A, Lee P, Chua BQ, et al. Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model. *Nat Genet*. 2014;46(1):24–32.

23. Müller-Coan BG, Caetano BFR, Pagano JS, Elgui de Oliveira D. Cancer Progression Goes Viral: The Role of Oncoviruses in Aggressiveness of Malignancies. *Trends Cancer*. 2018;4(7):485–98.
24. Moody CA, Laimins LA. Human papillomavirus oncoproteins: Pathways to transformation. *Nat Rev Cancer*. 2010;10(8):550–60.
25. Martínez-Ramírez I, Carrillo-García A, Contreras-Paredes A, Ortiz-Sánchez E, Cruz-Gregorio A, Lizano M. Regulation of cellular metabolism by high-risk human papillomaviruses. *Int J Mol Sci* [Internet]. el 1 de julio de 2018 [citado el 18 de julio de 2020];19(7). Disponible en: [/pmc/articles/PMC6073392/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/30000000/)
26. Lin MV, King LY, Chung RT. Hepatitis C virus-associated cancer. *Annu Rev Pathol Mech Dis*. 2015;10(October 2014):345–70.
27. Saigo K, Yoshida K, Ikeda R, Sakamoto Y, Murakami Y, Urashima T, et al. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. *Hum Mutat*. 2008;29(5):703–8.
28. Fleming JC, Woo J, Moutasim K, Mellone M, Frampton SJ, Mead A, et al. HPV, tumour metabolism and novel target identification in head and neck squamous cell carcinoma. *Br J Cancer*. el 5 de febrero de 2019;120(3):356–67.
29. Yang J, Guo Y, Seo W, Zhang R, Lu C, Wang Y, et al. Targeting cellular metabolism to reduce head and neck cancer growth. *Sci Rep*. el 1 de diciembre de 2019;9(1):1–10.
30. Burk RD, Chen Z, Saller C, Tarvin K, Carvalho AL, Scapulatempo-Neto C, et al. Integrated genomic and molecular characterization of cervical cancer. *Nature*. 2017;543(7645):378–84.

31. Lawrence MS, Sougnez C, Lichtenstein L, Cibulskis K, Lander E, Gabriel SB, et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517(7536):576–82.
32. Ally A, Balasundaram M, Carlsen R, Chuah E, Clarke A, Dhalla N, et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell*. 2017;169(7):1327-1341.e23.
33. Zhu S, Jin J, Gokhale S, Lu AM, Shan H, Feng J, et al. Genetic alterations of TRAF proteins in human cancers. *Front Immunol* [Internet]. el 20 de septiembre de 2018 [citado el 2 de septiembre de 2020];9. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/30294322/>
34. Nagy Á, Lánckzy A, Menyhárt O, Gyorffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep* [Internet]. el 1 de diciembre de 2018 [citado el 2 de septiembre de 2020];8(1). Disponible en: <https://pubmed.ncbi.nlm.nih.gov/29907753/>
35. Heineman TE, Widman A, Kuan EC, St John M. The genetic landscape of programmed death ligand-1 (PD-L1) alterations in head and neck cancer. *Laryngoscope Investig Otolaryngol*. 2017;2(3):99–103.
36. Foy JP, Bazire L, Ortiz-Cuaran S, Deneuve S, Kielbassa J, Thomas E, et al. A 13-gene expression-based radioresistance score highlights the heterogeneity in the response to radiation therapy across HPV-negative HNSCC molecular subtypes. *BMC Med*. 2017;15(1):1–13.
37. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. el 1 de junio de 2020;38(6):675–8.
38. National Cancer Institute. Copy Number Variation Analysis Pipeline Archived [Internet]. USA: National Cancer Institute; [citado el 17 de octubre de 2020]. (GDC

Documentation). Disponible en:  
[http://web.archive.org/web/20201017230302/https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/CNV\\_Pipeline/#copy-number-estimation](http://web.archive.org/web/20201017230302/https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/#copy-number-estimation)

39. Warden CD, Yuan YC, Wu X. Optimal calculation of RNA-Seq fold-change values. En 2013. Disponible en: <https://api.semanticscholar.org/CorpusID:15499199>

40. National Cancer Institute. Copy Number Variation Analysis Pipeline [Internet]. USA: National Cancer Institute; [citado el 14 de febrero de 2024]. (GDC Documentation). Disponible en:  
[https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/CNV\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/)

41. National Cancer Institute. GDC Reference Files: GRCh38.d1.vd1 Reference Sequence [Internet]. Genomic Data Commons; [citado el 14 de febrero de 2024]. Disponible en: <https://gdc.cancer.gov/about-data/gdc-data-processing/gdc-reference-files>

42. Gao GF, Oh C, Saksena G, Deng D, Westlake LC, Hill BA, et al. Tangent normalization for somatic copy-number inference in cancer genome analysis. Schwartz R, editor. *Bioinformatics*. el 14 de octubre de 2022;38(20):4677–86.

43. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. el 1 de octubre de 2004;5(4):557–72.

44. National Cancer Institute. mRNA Analysis Pipeline [Internet]. USA: National Cancer Institute; (GDC Documentation). Disponible en:  
[https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/Expression\\_mRNA\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/)

45. XenaBrowser. Data and datasets: TCGA Gene expression RNAseq (IlluminaHiSeq) [Internet]. USA: University of California Santa Cruz; 2022. (XenaBrowser User Help Pages). Disponible en: <https://ucsc->

xena.gitbook.io/project/faq/advanced-data-and-datasets#tcga-gene-expression-rnaseq-illumina-hiseq

46. XenaBrowser. Data and datasets: How do I calculate fold change (FC)? [Internet]. USA: University of California Santa Cruz; 2022. (XenaBrowser User Help Pages). Disponible en: <https://ucsc-xena.gitbook.io/project/faq/advanced-data-and-datasets#how-do-i-calculate-fold-change-fc>

47. Pagès J. Factorial Analysis of Mixed Data. En: Multiple Factor Analysis by Example Using R. 1a ed. USA: Chapman & Hall; 2014. p. 67–72. (The R Series).

48. KASSAMBARA A. Factor Analysis of Mixed Data. En: Practical Guide to Principal Component Methods in R. 1a ed. STHDA; 2017. p. 108–19. (Multivariate Analysis II).

49. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. el 21 de noviembre de 2019;4(43):1686.

50. Tierney N, Cook D. Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations. J Stat Softw [Internet]. 2023 [citado el 4 de marzo de 2024];105(7). Disponible en: <https://www.jstatsoft.org/v105/i07/>

51. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. J Stat Softw [Internet]. 2016 [citado el 4 de marzo de 2024];70(1). Disponible en: <http://www.jstatsoft.org/v70/i01/>

52. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses [Internet]. 2020 [citado el 4 de marzo de 2024]. Disponible en: <https://cran.r-project.org/web/packages/factoextra/index.html>

53. Lê S, Josse J, Husson F. **FactoMineR** : An R Package for Multivariate Analysis. J Stat Softw [Internet]. 2008 [citado el 4 de marzo de 2024];25(1). Disponible en: <http://www.jstatsoft.org/v25/i01/>
54. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots [Internet]. 2023 [citado el 4 de marzo de 2024]. Disponible en: <https://cran.r-project.org/web/packages/ggpubr/index.html>
55. XenaBrowser. Kaplan Meier Plots [Internet]. USA: University of California Santa Cruz; [citado el 16 de marzo de 2024]. Disponible en: <https://ucsc-xena.gitbook.io/project/overview-of-features/kaplan-meier-plots>
56. Rich JT, Neely JG, Paniello RC, Voelker CCJ, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. Otolaryngol Neck Surg. septiembre de 2010;143(3):331–6.
57. Van Paemel R. Kaplan Meier curves: an introduction [Internet]. 2019 mar [citado el 15 de marzo de 2024]. Disponible en: [https://rmvpaeme.github.io/KaplanMeier\\_intro/](https://rmvpaeme.github.io/KaplanMeier_intro/)
58. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Disponible en: <https://www.r-project.org/>
59. Wickham H, Chan W, Henry L, Lin T, Kohske T, Claus W, et al. ggplot2: Elegant Graphics for Data Analysis [Internet]. New York: Springer-Verlag New York; 2016. Disponible en: <https://ggplot2.tidyverse.org>
60. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. el 3 de julio de 2017;45(W1):W98–102.

61. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* el 8 de enero de 2019;47(D1):D590–5.
62. Kahlert UD, Cheng M, Koch K, Marchionni L, Fan X, Raabe EH, et al. Alterations in cellular metabolome after pharmacological inhibition of Notch in glioblastoma cells: Notch Metabolome. *Int J Cancer.* el 1 de marzo de 2016;138(5):1246–55.
63. Le A, Lane AN, Hamaker M, Bose S, Gouw A, Barbi J, et al. Glucose-Independent Glutamine Metabolism via TCA Cycling for Proliferation and Survival in B Cells. *Cell Metab.* enero de 2012;15(1):110–21.
64. Lobo C, Ruiz-Bellido MA, Aledo JC, Márquez J, Núñez De Castro I, Alonso FJ. Inhibition of glutaminase expression by antisense mRNA decreases growth and tumourigenicity of tumour cells. *Biochem J.* el 1 de junio de 2000;348 Pt 2(Pt 2):257–61.
65. Zhang C, Liu J, Zhao Y, Yue X, Zhu Y, Wang X, et al. Glutaminase 2 is a novel negative regulator of small GTPase Rac1 and mediates p53 function in suppressing metastasis. *eLife.* el 11 de enero de 2016;5:e10727.
66. Suzuki S, Venkatesh D, Kanda H, Nakayama A, Hosokawa H, Lee E, et al. GLS2 Is a Tumor Suppressor and a Regulator of Ferroptosis in Hepatocellular Carcinoma. *Cancer Res.* el 16 de septiembre de 2022;82(18):3209–22.
67. Pyeon D, Newton MA, Lambert PF, Den Boon JA, Sengupta S, Marsit CJ, et al. Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res.* 2007;67(10):4605–19.
68. Burk RD, Chen Z, Van Doorslaer K. Human papillomaviruses: Genetic basis of carcinogenicity. *Public Health Genomics.* 2009;12(5–6):281–90.

69. French D, Lorenzon L. HPV Infections: Basis of Neoplastic Transformation and Related Molecular Tests. *Curr Pharm Des.* 2013;19(8):1371–8.
70. Clifford GM, Smith JS, Aguado T, Franceschi S. Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: A meta-analysis. *Br J Cancer.* 2003;89(1):101–5.
71. Clifford G, Franceschi S, Diaz M, Muñoz N, Villa LL. Chapter 3: HPV type-distribution in women with and without cervical neoplastic diseases. *Vaccine.* 2006;24(SUPPL. 3):26–34.
72. Lévy P, Bartosch B. Metabolic reprogramming: a hallmark of viral oncogenesis. *Oncogene.* 2016;35(32):4155–64.
73. Mazurek S, Zwerschke W, Jansen-Dürr P, Eigenbrodt E. Effects of the human papilloma virus HPV-16 E7 oncoprotein on glycolysis and glutaminolysis: role of pyruvate kinase type M2 and the glycolytic-enzyme complex. *Biochem J.* el 15 de mayo de 2001;356(Pt 1):247–56.
74. Ortiz-Pedraza Y, Muñoz-Bello JO, Ramos-Chávez LA, Martínez-Ramírez I, Olmedo-Nieva L, Manzo-Merino J, et al. HPV16 E6 and E7 Oncoproteins Stimulate the Glutamine Pathway Maintaining Cell Proliferation in a SNAT1-Dependent Fashion. *Viruses.* el 24 de enero de 2023;15(2):324.
75. Choo CK, Rorke EA, Eckert RL. Differentiation-independent Constitutive Expression of the Human Papillomavirus Type 16 E6 and E7 Oncogenes in the CaSki Cervical Tumour Cell Line. *J Gen Virol.* el 1 de mayo de 1994;75(5):1139–47.
76. Wicker CA, Hunt BG, Krishnan S, Aziz K, Parajuli S, Palackdharry S, et al. Glutaminase inhibition with telaglenastat (CB-839) improves treatment response in combination with ionizing radiation in head and neck squamous cell carcinoma models. *Cancer Lett.* abril de 2021;502:180–8.

77. Cetindis M, Biegner T, Munz A, Teriete P, Reinert S, Grimm M. Glutaminolysis and carcinogenesis of oral squamous cell carcinoma. *Eur Arch Otorhinolaryngol.* 2016;273(2):495–503.
78. Zhang Z, Liu R, Shuai Y, Huang Y, Jin R, Wang X, et al. ASCT2 (SLC1A5)-dependent glutamine uptake is involved in the progression of head and neck squamous cell carcinoma. *Br J Cancer.* 2020;122(1):82–93.
79. Basheer HA, Elsalem L, Salem A, Tailor A, Hunter K, Afarinkia K. The Expression of Glutaminases and their Association with Clinicopathological Parameters in the Head and Neck Cancers. *Curr Cancer Drug Targets.* febrero de 2022;22(2):169–79.
80. Liu N, Shi F, Yang L, Liao W, Cao Y. Oncogenic viral infection and amino acid metabolism in cancer progression: Molecular insights and clinical implications. *Biochim Biophys Acta BBA - Rev Cancer.* mayo de 2022;1877(3):188724.
81. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primer.* 2020;6(1):92.
82. Chin D, Boyle GM, Theile DR, Parsons PG, Coman WB. Molecular introduction to head and neck cancer (HNSCC) carcinogenesis. *Br J Plast Surg.* 2004;57(7):595–602.
83. Canning M, Guo G, Yu M, Myint C, Groves MW, Byrd JK, et al. Heterogeneity of the head and neck squamous cell carcinoma immune landscape and its impact on immunotherapy. *Front Cell Dev Biol.* 2019;7(APR):1–19.
84. Economopoulou P, Kotsantis I, Psyrri A. Special issue about head and neck cancers: HPV positive cancers. *Int J Mol Sci.* 2020;21(9).
85. Ndiaye C, Mena M, Alemany L, Arbyn M, Castellsagué X, Laporte L, et al. HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: A systematic review and meta-analysis. *Lancet Oncol.* 2014;15(12):1319–31.

86. Ni G, Huang K, Luan Y, Cao Z, Chen S, Ma B, et al. Human papillomavirus infection among head and neck squamous cell carcinomas in southern China. *PLoS ONE*. 2019;14(9):1–13.
87. Lajer CB, Buchwald CV. The role of human papillomavirus in head and neck cancer. *Apmis*. 2010;118(6–7):510–9.
88. Krishna G, Soman Pillai V, Valiya Veettil M. Upregulation of GLS1 Isoforms KGA and GAC Facilitates Mitochondrial Metabolism and Cell Proliferation in Epstein–Barr Virus Infected Cells. *Viruses*. el 27 de julio de 2020;12(8):811.
89. Alsbeih G, Al-Harbi N, Judia SB, Al-Qahtani W, Khoja H, El-Sebaie M, et al. Prevalence of human papillomavirus (HPV) infection and the association with survival in saudi patients with head and neck squamous cell carcinoma. *Cancers*. 2019;11(6).
90. K. Kian A, Jonathan H, Richard W, Randal W, David I. R, Phuc Felix NT, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;36(5):750–1.
91. Fakhry C, Westra WH, Wang SJ, van Zante A, Zhang Y, Rettig E, et al. The prognostic role of sex, race, and human papillomavirus in oropharyngeal and nonoropharyngeal head and neck squamous cell cancer. *Cancer*. 2017;123(9):1566–75.
92. Chatfield-Reed K, Gui S, O’Neill WQ, Teknos TN, Pan Q. HPV33+ HNSCC is associated with poor prognosis and has unique genomic and immunologic landscapes. *Oral Oncol*. enero de 2020;100:104488.
93. Gao M, Monian P, Quadri N, Ramasamy R, Jiang X. Glutaminolysis and Transferrin Regulate Ferroptosis. *Mol Cell*. julio de 2015;59(2):298–308.

94. Nie J, Lin B, Zhou M, Wu L, Zheng T. Role of ferroptosis in hepatocellular carcinoma. *J Cancer Res Clin Oncol*. diciembre de 2018;144(12):2329–37.
95. Xia X, Fan X, Zhao M, Zhu P. The Relationship between Ferroptosis and Tumors: A Novel Landscape for Therapeutic Approach. *Curr Gene Ther*. el 20 de agosto de 2019;19(2):117–24.
96. Yu H, Guo P, Xie X, Wang Y, Chen G. Ferroptosis, a new form of cell death, and its relationships with tumourous diseases. *J Cell Mol Med*. abril de 2017;21(4):648–57.
97. Torresi J, Tran BM, Christiansen D, Earnest-Silveira L, Schwab RHM, Vincan E. HBV-related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. *BMC Cancer*. 2019;19(1):707.
98. Balogh J, Iii DV, Gordon S, Li X, Ghobrial RM, Jr HPM. Hepatocellular carcinoma: a review. *J Hepatocell Carcinoma*. 2016;Volume 3:41–53.
99. Suzuki S, Venkatesh D, Tanaka T, Prives C. GLS2 shapes ferroptosis in hepatocellular carcinoma. *Oncotarget*. el 12 de enero de 2023;14(1):900–3.
100. Lee YC, Wang JH, Chen CH, Hung CH, Lo KC, Yen YH, et al. Sorafenib use in hepatitis B virus- or hepatitis C virus-related hepatocellular carcinoma: A propensity score matching study. *Kaohsiung J Med Sci*. el 24 de junio de 2021;1–9.
101. Wu CC, Ho WL, Chen JT, Tang JS, Yeh DC, P'eng FK. Hepatitis viral status in patients undergoing liver resection for hepatocellular carcinoma. *Br J Surg*. el 6 de julio de 1999;86(11):1391–6.
102. Cantarini MC, Trevisani F, Morselli-Labate AM, Rapaccini G, Farinati F, Poggio PD, et al. Effect of the Etiology of Viral Cirrhosis on the Survival of Patients with Hepatocellular Carcinoma. *Am J Gastroenterol*. enero de 2006;101(1):91–8.

103. Ng J, Wu J. Hepatitis B- and Hepatitis C-Related Hepatocellular Carcinomas in the United States: Similarities and Differences. *Hepat Mon* [Internet]. el 25 de octubre de 2012 [citado el 13 de octubre de 2022];12(10 HCC). Disponible en: <https://brief.land/hepatmon/articles/70615.html>
104. Menyhárt O, Nagy Á, Gyórfy B. Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. *R Soc Open Sci.* diciembre de 2018;5(12):181006.
105. Lániczky A, Gyórfy B. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. *J Med Internet Res.* el 26 de julio de 2021;23(7):e27633.
106. Yu D, Shi X, Meng G, Chen J, Yan C, Jiang Y, et al. Kidney-type glutaminase (GLS1) is a biomarker for pathologic diagnosis and prognosis of hepatocellular carcinoma. *Oncotarget.* el 10 de abril de 2015;6(10):7619–31.

## 11. Anexos

*Anexo 1. Enlace de la base de datos usada para los análisis realizados.*

<https://github.com/livisupajares/undergrad-thesis/tree/master/database/fixed>

*Anexo 2. Repositorio con los scripts realizados en R.*

<https://github.com/livisupajares/undergrad-thesis/tree/master/scripts>

*Anexo 3. Expresión génica (expresados en mediana de  $\log_2(\text{fpkm}+1)$ ) de GLS1 y GLS2 por el tipo de muestra en cáncer cervical.*

Gen	Tumor	Control	Welch's t-test
GLS1	2.63 (n=178)	3.23 (n=3)	p = 0.1634 (t = -1.975)
GLS2	0.288 (n=178)	0.0755 (n=3)	p = 0.07329 (t = -3.051)

Anexo 4. Alteración del número de copias (ANC) de GLS1 y GLS2 por el tipo de muestra en cáncer cervical.

Gen	Estatus de la alteración del número de copias (ANC)				Test estadístico
	Variable: Tipo de tejido				
	Tumor (n=178)		Control (n=3)		
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Chi cuadrado de Pearson
GLS1	12.4% (n=22)	87.6% (n=156)	0%	100% (n=3)	p = 0.5159 ( $\chi^2 = 0.4221$ )
GLS2	6.74% (n=12)	93.3% (n=166)	0%	100% (n=3)	p = 0.6416 ( $\chi^2 = 0.2166$ )

Anexo 5. Expresión génica de GLS1 y GLS2 agrupadas por variables clínicas con diferencias no significativas en la cohorte de CESC.

Gen	Expresión génica (mediana de $\log_2(\text{fpkm}+1)$ )				Prueba estadístico
	Variable: Estatus de infección de la muestra				
	Tumor con infección HPV(+)	Tumor sin infección HPV(-)	Control sin infección HPV(-) (tejido no tumoral sin infección)		
GLS1	2.62 (n=169)	2.74 (n=9)	3.23 (n=3)		ANOVA de una vía  p = 0.5039 (f = 0.6881)
Gen	Expresión génica (mediana de $\log_2(\text{fpkm}+1)$ )				Prueba estadística
	Variable: Subtipo viral				
	HPV18	HPV16	Otros		
GLS1	2.48 (n=27)	2.69 (n=103)	2.53 (n=39)		ANOVA de una vía  p = 0.2751 (f = 1.301)

Anexo 6. Alteración del número de copias (ANC) de GLS1 y GLS2 agrupadas por variables clínicas sin diferencias significativas en la cohorte de CESC.

Gen	Estatus de la alteración del número de copias (ANC)	Test
-----	---	------

	Variable: Estatus de infección de la muestra						estadístico
	Tumor HPV(+) (n=169)		Tumor HPV(-) (n=9)		Control HPV(-) (n=3)		
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Chi cuadrado de Pearson
GLS1	13% (n=22)	87% (n=147)	0%	100% (n=9)	0%	100% (n=3)	p=0.4110 ( $\chi^2=1.778$ )
GLS2	6.51% (n=11)	93.5% (n=158)	11.1% (n=1)	88.9% (n=8)	0%	100% (n=3)	p = 0.7753 ( $\chi^2 = 0.5090$ )
Gen	Estatus de la alteración del número de copias (ANC)						Test estadístico
	Variable: Subtipo viral						
	HPV18 (n=27)		HPV16 (n=106)		Otros (n=39)		
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Chi cuadrado de Pearson
GLS1	7.41% (n=2)	92.6% (n=25)	13.6% (n=14)	86.4% (n=89)	15.4% (n=6)	84.6% (n=33)	p = 0.5812 ( $\chi^2 = 1.085$ )
GLS2	11.1% (n=3)	88.9% (n=24)	6.8% (n=7)	93.2% (n=96)	2.56% (n=1)	97.4% (n=38)	p = 0.3862 ( $\chi^2 = 1.903$ )
GLS2	5.08% (n=3)	94.9% (n=56)	6.52% (n=6)	93.5% (n=86)	11.1% (n=3)	88.9% (n=24)	p = 0.5814 ( $\chi^2 = 1.085$ )

Anexo 7. Expresión génica de GLS2 (expresados en mediana de  $\log_2(\text{fpm}+1)$ ) en cáncer de cabeza y cuello.

Gen	Variable		Test estadístico
	Tipo de muestra		
	Tumor	Control	Test T de Welch
GLS2	0.106 (n=274)	0.224 (n=44)	p = 0.3289 (t = 0.9801)

Anexo 8. Alteración del número de copias (ANC) de GLS1 y GLS2 no significativas por el tipo de muestra en cáncer de cabeza y cuello.

Gen	Estatus de la alteración del número de copias (ANC)				Test estadístico
	Variable: Tipo de tejido				
	tumor (n=276)		control (n=44)		
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Chi cuadrado de Pearson
GLS1	10.9% (n=30)	89.1% (n=246)	0%	100% (n=44)	p = 0.02160 ( $\chi^2 = 5.277$ )
GLS2	4.71% (n=13)	95.3% (n=263)	0%	100% (n=44)	p = 0.1416 ( $\chi^2 = 2.160$ )

Anexo 9. Expresión génica de GLS1 y GLS2 no significativas agrupadas por variables clínicas en la cohorte de HNSCC.

Gen	Expresión génica (mediana de log2(fpkm+1))			Prueba estadística
	Variable: Subtipo viral			
	HPV16	HPV33	otros	
GLS1	2.75 (n=42)	2.82 (n=15)	3.15 (n=4)	ANOVA de una vía p = 0.1633 (f = 1.870)
GLS2	0.542 (n=42)	0.308 (n=15)	0.134 (n=4)	p = 0.5159 (f = 0.6695)

Anexo 10. Alteración del número de copias (ANC) de GLS1 y GLS2 agrupadas por variables clínicas sin diferencias significativas en la cohorte de HNSC.

Gen	Estatus de la alteración del número de copias (ANC)						Test estadístico
	Variable: Estatus de infección de la muestra						
	Tumor HPV(+) (n=55)		Tumor HPV(-) (n=221)		Control HPV(-) (n=38)		
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Chi cuadrado de Pearson

GLS1	10.9% (n=6)	89.1% (n=49)	10.9% (n=24)	89.1% (n=197)	0%	100% (n=38)	p = 0.1526 ( $\chi^2 = 5.277$ )	
GLS2	3.64% (n=2)	96.4% (n=53)	4.98% (n=11)	95% (n=210)	0%	100% (n=38)	p = 0.5005 ( $\chi^2 = 2.363$ )	
Gen	Control HPV(+) (n=6)							
	cambio nº copias	Sin cambio nº de copias						
GLS1	0%	100% (n=6)						
GLS2	0%	100% (n=6)						
Gen	Estatus de la alteración del número de copias (ANC)						Test estadístico	
	Variable: Subtipo viral							
	HPV33 (n=11)		HPV16 (n=40)		Otros (n=4)			
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Chi cuadrado de Pearson	
GLS1	9.09% (n=1)	90.9% (n=10)	10% (n=4)	90% (n=36)	25% (n=1)	75% (n=3)	p = 0.6403 ( $\chi^2 = 0.8915$ )	
GLS2	0%	100% (n=11)	5% (n=2)	95% (n=38)	0%	100% (n=4)	p = 0.6877 ( $\chi^2 = 0.7489$ )	

*Anexo 11. Alteración del número de copias (ANC) de GLS1 y GLS2 no significativas por el tipo de muestra en cáncer de hígado.*

Gen	Estatus de la alteración del número de copias (ANC)				Test estadístico
	Variable: Tipo de tejido				
	Tumor (n=192)		Control (n=50)		
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Chi cuadrado de Pearson
GLS1	6.77% (n=13)	93.2% (n=179)	0%	100% (n=50)	p = 0.05856 ( $\chi^2 =$

					3.578)
GLS2	5.21%(n=10)	94.8% (n=182)	0%	100% (n=50)	p = 0.09932 ( $\chi^2 =$ 2.716)

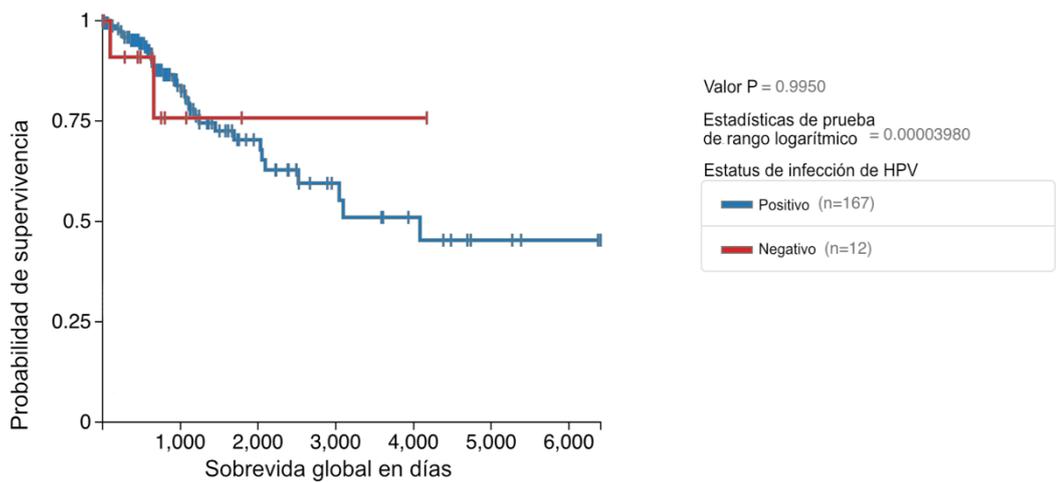
Anexo 12. Expresión génica, alteración del número de copias de GLS1 y GLS2 agrupados por las variables clínicas con diferencias no significativas en la cohorte de LIHC.

Gen	Expresión génica (mediana de log <sub>2</sub> (f <sub>pk</sub> m+1))			Prueba estadística
	Variable: Subtipo viral			
	HBV (n=36)	HCV (n=48)		Test T de Welch
GLS1	2.08 (n=48)	1.84 (n=36)		p = 0.3413 (t = - 0.9577)
GLS2	0.384 (n=48)	0.666 (n=36)		p = 0.6337 (t = - 0.4784)

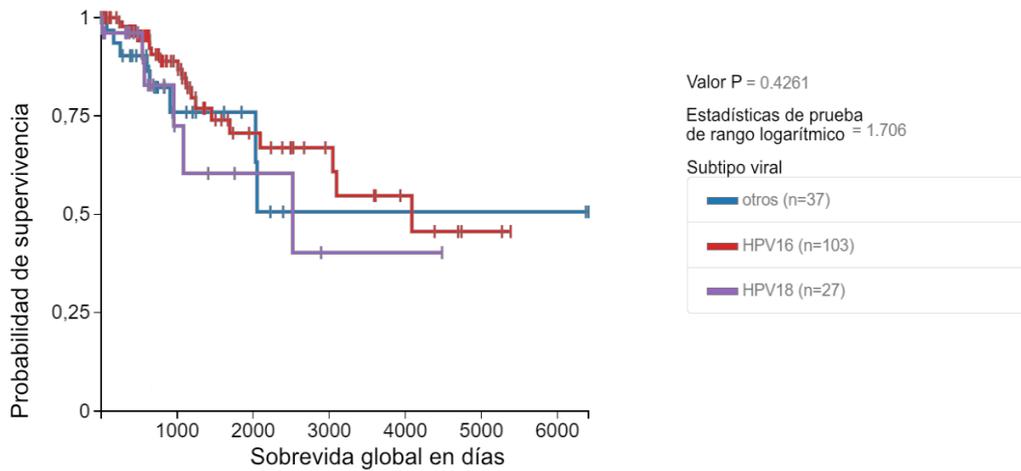
Anexo 13. Alteración del número de copias (ANC) de GLS1 y GLS2 agrupadas por variables clínicas sin diferencias significativas en la cohorte de LIHC.

Gen	Estatus de la alteración del número de copias (ANC)						Test estadístico
	Variable: Estatus de infección de la muestra						
	Control negativo (n=36)		Tumor negativo (n=121)		Control positivo (n=14)		Chi cuadrado de Pearson
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	
GLS1	0%	100% (n=36)	4.96% (n=6)	95% (n=115)	0%	100% (n=14)	p = 0.1276 ( $\chi^2 =$ 5.691)
GLS2	0%	100% (n=36)	5.79% (n=7)	94.2% (n=114)	0%	100% (n=14)	p = 0.3930 ( $\chi^2 =$ 2.991)

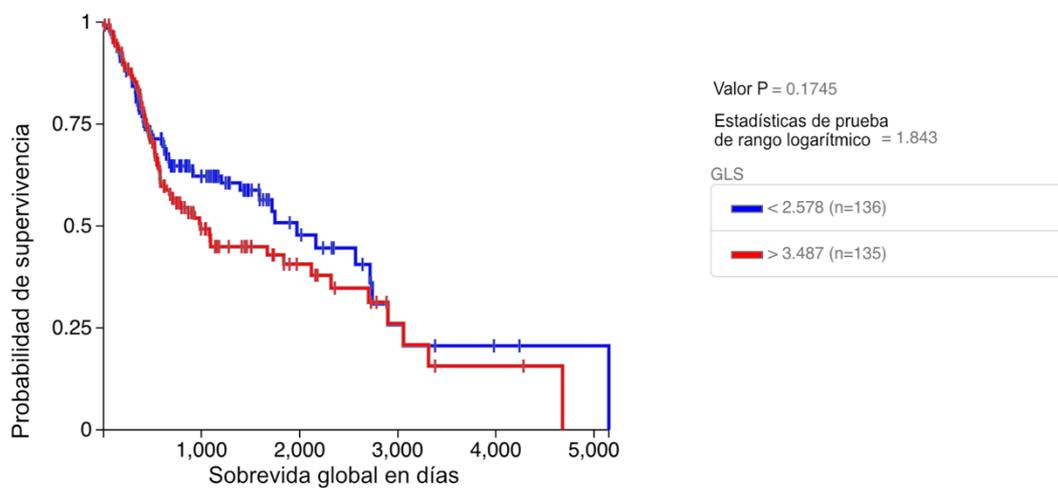
Gen	Tumor positivo (n=71)				Test estadístico
	Cambio nº copias	Sin cambio nº de copias			
GLS1	9.86% (n=7)	90.1% (n=64)			Chi cuadrado de Pearson
GLS2	4.23% (n=3)	95.8% (n=68)			
Gen	Estatus de la alteración del número de copias (ANC)				Test estadístico
	Variable: Subtipo viral				
Gen	HBV (n=40)		HCV (n=31)		Chi cuadrado de Pearson
	Cambio nº copias	Sin cambio nº de copias	Cambio nº copias	Sin cambio nº de copias	
GLS1	12.5% (n=5)	87.5% (n=35)	6.45% (n=2)	93.5% (n=29)	p = 0.1488 (χ <sup>2</sup> = 2.085)
GLS2	5% (n=2)	95% (n=38)	3.23% (n=1)	96.8% (n=30)	p = 0.6169 (χ <sup>2</sup> = 0.2503)



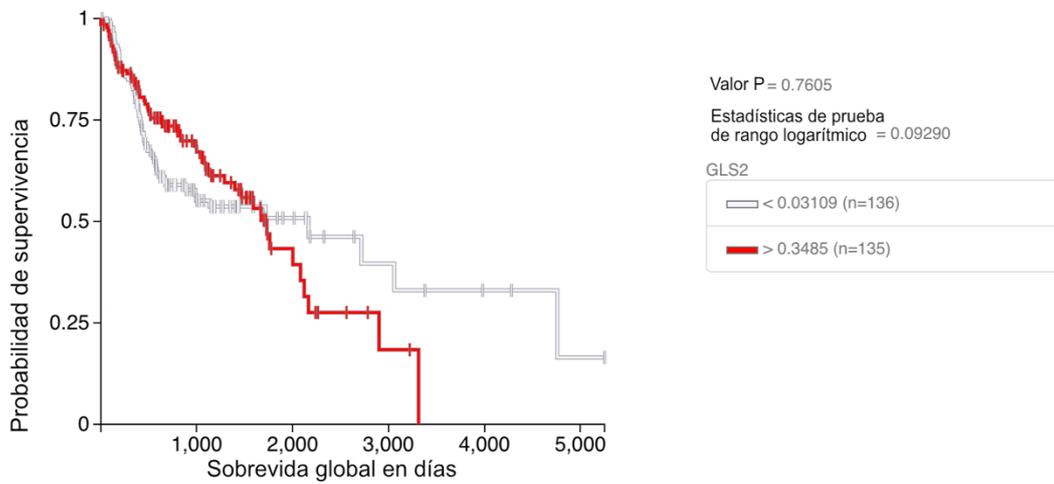
Anexo 14. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con cáncer cervical agrupados por estatus de infección por HPV.



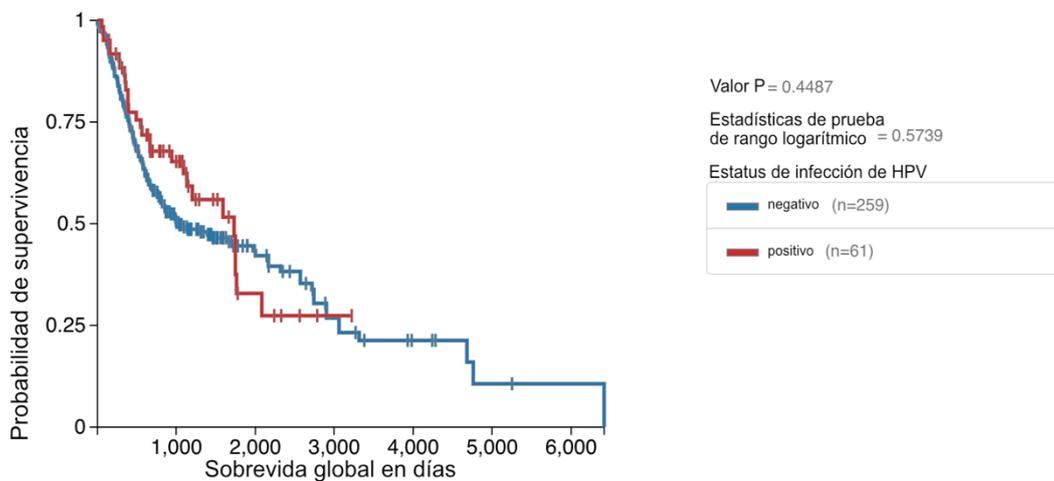
Anexo 15. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervida (eje Y) en pacientes con cáncer cervical agrupados por el tipo de HPV.



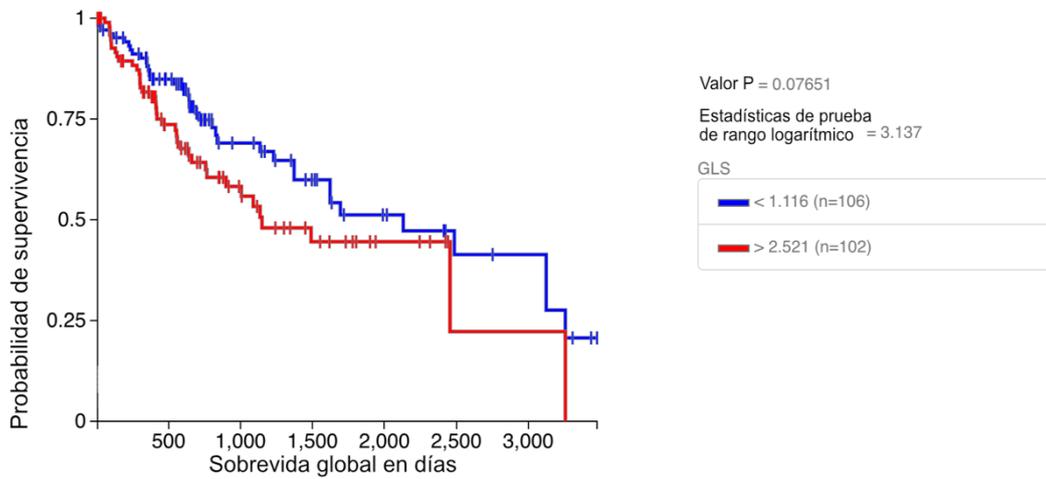
Anexo 16. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervida (eje Y) en pacientes con cáncer de cabeza y cuello agrupados por los niveles de expresión génica de GLS1. Los datos de expresión están agrupados en cuartiles.



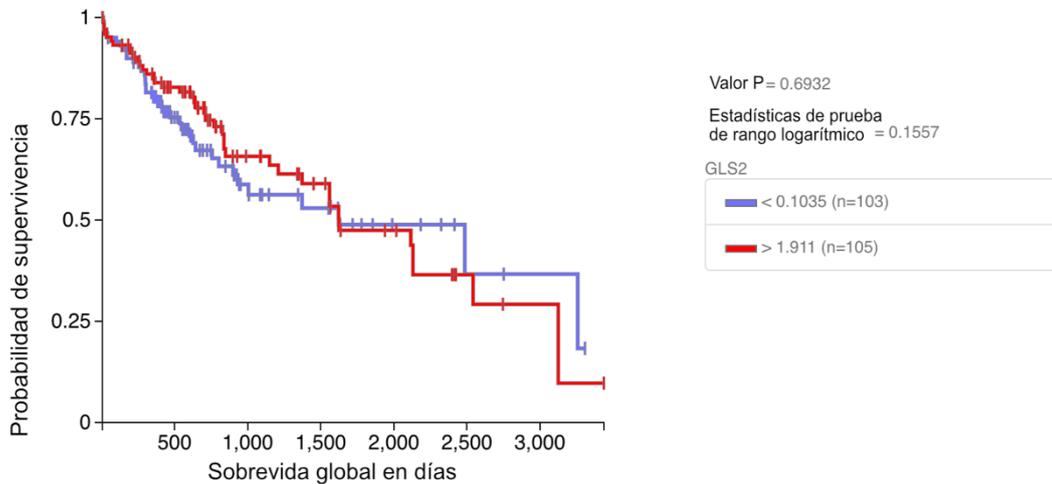
Anexo 17. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con cáncer de cabeza y cuello agrupados por los niveles de expresión génica de GLS2. Los datos de expresión están agrupados en cuartiles.



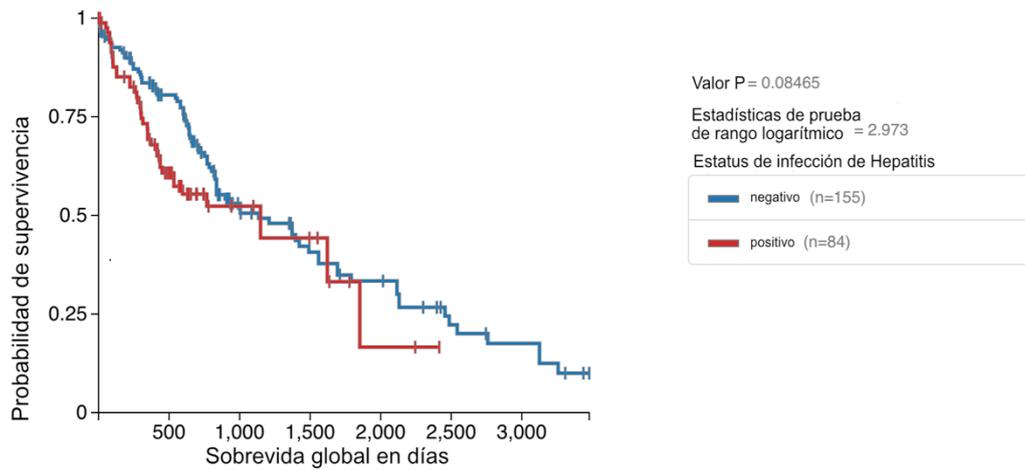
Anexo 18. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con cáncer de cabeza y cuello agrupados por estatus de infección por HPV.



Anexo 19. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con cáncer de hígado agrupados por los niveles de expresión génica de GLS1. Los datos de expresión están agrupados en cuartiles.



Anexo 20. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con cáncer de hígado agrupados por los niveles de expresión génica de GLS2. Los datos de expresión están agrupados en cuartiles.



*Anexo 21. Curva de KM representando el OS (sobrevida global) en días (eje x) y la probabilidad de supervivencia (eje Y) en pacientes con cáncer de hígado agrupados por estatus de infección de HBV/HCV.*