



**Universidad  
Europea** VALENCIA

**Grado en ODONTOLOGIA**

**Trabajo Fin de Grado**

**Curso 2022-2023**

***METAANÁLISIS DE EXPRESIÓN GÉNICA  
EN PACIENTES CON CÁNCER, FIBROSIS Y  
LEUCOPLASIAS EN LA CAVIDAD ORAL  
MEDIANTE DATOS PÚBLICOS DE RNA-SEQ.***

Presentado por: ***Agnese Formaggio***

Tutora: ***Marta Serna García***

**Campus de Valencia**  
Paseo de la Alameda, 7  
46010 Valencia  
universidadeuropea.com



**Universidad**  
**Europea** VALENCIA



<b>1. RESUMEN.....</b>	<b>5</b>
<b>2. ABSTRACT.....</b>	<b>7</b>
<b>3. PALABRAS CLAVES.....</b>	<b>9</b>
<b>4. INTRODUCCIÓN.....</b>	<b>11</b>
<b>4.1 Ciencia ómicas.....</b>	<b>11</b>
<b>4.2 Transcriptómica.....</b>	<b>12</b>
<b>4.3 RNA-seq.....</b>	<b>12</b>
<b>4.4 Tejido epitelial.....</b>	<b>14</b>
<b>4.5 Lesiones Pre-cancerígenas.....</b>	<b>15</b>
<b>4.5.1. Fibrosis.....</b>	<b>16</b>
<b>4.5.1 Leucoplasia.....</b>	<b>16</b>
<b>4.6 Cáncer Oral.....</b>	<b>18</b>
<b>5. JUSTIFICACIÓN.....</b>	<b>21</b>
<b>6. HIPOTESIS.....</b>	<b>24</b>
<b>7. OBJETIVOS.....</b>	<b>24</b>
<b>7.1 General.....</b>	<b>24</b>
<b>7.2 Especificos.....</b>	<b>24</b>
<b>8. MATERIALES Y MÉTODOS.....</b>	<b>26</b>
<b>8.1 Diseño de estudios.....</b>	<b>26</b>
<b>8.1.1 Selección de los estudios.....</b>	<b>26</b>
<b>8.1.2 Selección de datos brutos para los analisis.....</b>	<b>26</b>
<b>9. RESULTADOS.....</b>	<b>31</b>
<b>10. DISCUSIÓN.....</b>	<b>46</b>
<b>11. CONCLUSIÓN.....</b>	<b>52</b>
<b>12. BIBLIOGRAFÍA.....</b>	<b>54</b>



**Universidad**  
**Europea** VALENCIA

## 1. RESUMEN

**Introducción:** El RNA-seq es una herramienta que descubre nuevos biomarcadores de cáncer oral y lesiones pre-cancerígenas como la leucoplasia y fibrosis oral, además, estudia genes relevantes y conoce la estructura biológica pudiendo ayudar a desarrollar terapias. El objetivo fue realizar un estudio transcriptómico e identificar los procesos biológicos en la cavidad oral mediante RNA-seq para muestras obtenidas de bases de datos públicas de grupos divergentes; muestras de pacientes con lesiones pre-cancerosas: Leucoplasia (L) y Fibrosis oral (F) y Cáncer oral (C) comparada con pacientes sin patología (N).

**Materiales y métodos:** Seleccionaron tres estudios de la base de datos pública *GeneExpression Omnibus* (*GSE125866*<sup>(2019)</sup>, *GSE20116*<sup>(2010)</sup> y *GSE131568*<sup>(2020)</sup>). Todos los archivos brutos se analizaron con el mismo flujo de trabajo. La calidad de lecturas se verificó mediante el *Fastqc software* y se filtraron con *Trimmomatic software*. Se alineó con genoma *Homo Sapiens* con *Kallisto software*. *EdgeR software* y *Limma software* normalizaron las lecturas e identificaron genes expresados diferencialmente ( $FDR \leq 0.05$ ). Para determinar procesos biológicos significativos fue usado *DAVID software* ( $p \text{ value} \leq 0,05$ ). *Cytoscape software* fue utilizado para términos biológicos.

**Resultados:** Se detectaron 800 genes infraexpresados y 608 sobreexpresados (C vsN), 102 genes infraexpresado y 8 sobreexpresado (FvsN) y 2567 genes infraexpresado y 2603 sobreexpresado (LvsN). 154 genes comunes sobreexpresados (C y L), un gen (L y F), 4 genes (C y F) 148 genes comunes infraexpresado (C y L) y 104 genes (F y L) y 1 gen (C y F). EPRS fue el único gen común en las tres patologías ( $Foldchange 0.653$ ). De los 49 genes destacamos 8 involucrados con el tejido epitelial.

**Conclusiones:** Los genes ACTIN1, LIMK1, CORO1C, INF2, CFL1, FSCN1, MYO1B y SH3D21 podrían influenciar la integridad del tejido epitelial y ayudar al avance de lesiones precancerosas y cáncer. El EPRS es un posible biomarcador temprano.



**Universidad**  
**Europea** VALENCIA

## 2. ABSTRACT

**Introduction:** RNA-seq is a tool that discovers new biomarkers of oral cancer and precancerous lesions such as leukoplakia and oral fibrosis, in addition, it studies relevant genes and knows the biological structure, thus helping to develop therapies. The goal was to carry out a transcriptomic study and identify the biological processes in the oral cavity using RNA-seq for samples obtained from public databases of divergent groups; samples from patients with precancerous lesions: Leukoplakia (L) and oral fibrosis (F) and oral cancer (C) compared with patients without pathology (N).

**Materials and methods:** Three studies were selected from the public database *Gene Expression for All* (*GSE125866*<sup>(2019)</sup>, *GSE20116*<sup>(2010)</sup> y *GSE131568*<sup>(2020)</sup>). All raw files were analyzed with the same workflow. The quality of readings was verified using *Fastqc software* and filtered with *Trimmomatic software*. Aligned with genome *A wise man* with *Callisto software*. *EdgeR software* y *Glue software* normalized the reads and identified differentially expressed genes ( $FDR \leq 0.05$ ). *DAVID* was used to determine significant biological processes *software* ( $p \text{ value} \leq 0,05$ ). *Cytoscape software* was used for biological terms.

**Results:** 800 underexpressed and 608 overexpressed (C vsN) genes, 102 underexpressed and 8 overexpressed (FvsN) genes, and 2567 underexpressed and 2603 overexpressed (LvsN) genes were detected. 154 overexpressed common genes (C and L), one gene (L and F), 4 genes (C and F) 148 underexpressed common genes (C and L) and 104 genes (F and L) and 1 gene (C and F) . *EPRS* was the only common gene in the three pathologies ( $Foldchange 0.653$ ). Of the 49 genes, we highlight 8 involved with epithelial tissue.

**Conclusions:** The *ACTIN1*, *LIMK1*, *CORO1C*, *INF2*, *CFL1*, *FSCN1*, *MYO1B* and *SH3D21* genes could influence the integrity of epithelial tissue and help advance precancerous lesions and cancer. *EPRS* is a potential early biomarker.



**Universidad**  
**Europea** VALENCIA



### 3. PALABRAS CLAVES

- I. RNA-seq
- II. Cáncer oral
- III. Lesiones pre-cancerígenas
- IV. Leucoplasia
- V. Fibrosis
- VI. Biomarcadores
- VII. Genes
- VIII. Estudio transcriptómico
- IX. Tejido epitelial
- X. Procesos biológico



**Universidad**  
**Europea** VALENCIA

## 4. INTRODUCCIÓN

### 4.1 Ciencias Òmicas

Las ciencias Òmicas, son ciencia que permiten de obtener una perspectiva global del individuo a través de una visión del funcionamiento de sus células y de la influencia del ambiente que las rodea (1) además, gracias a las tecnologías Òmicas, estas ciencias, se caracterizan por ser de alto rendimiento, generando grandes cantidades de datos (denominado *big data*) en un solo experimento realizado a partir de una única muestra (2).

Tradicionalmente, la investigación del cáncer ha sido de las que más se han beneficiado con la aplicación de las ciencias Òmicas, tanto para la explicación la patología, como para la detección, el diagnóstico, el pronóstico y el tratamiento de distintos tipos de cáncer (3).

Existen diferentes áreas dentro de las ciencias Òmicas en las cuales encontramos: la Metabólica, la Transcriptómica, la Proteómica y por último la Genómica.

La Genómica es una ciencia que investiga cómo se constituye y funciona un genoma: unas de las tareas es caracterizar y cuantificar los genes (4).

El genoma es el conjunto de datos del ADN que componen un organismo (5), cada genoma humano contiene genes que codifican proteínas y genes que no codifican proteínas (6).

Una molécula de ADN está constituida por dos largas cadenas, cada cadena contiene un código genético compuesto por cuatro letras que representan las bases de nucleótidos:

- Adenina (A)
- Guanina (G)
- Timina (T)
- Citosina (C)

El orden de cada una de estas cadenas que forma parte del ADN se retuerce para formar una doble hélice (5).

El dogma central de la biología molecular describe las informaciones que se almacena en los genes como ADN, se transcribe en ARN proceso que describe la transcriptómica y finalmente se traduce en proteínas (7).

#### 4.2 Transcriptómica

En el campo de la genómica, el análisis del transcriptoma, siempre se ha considerado como un papel central para averiguar la complejidad de la regulación de la expresión génica (8).

La transcriptómica es considerada la ciencia que estudia el patrón de expresión genética de un organismo o de unas células específicas, es decir, estudia el conjunto de todos los ARN mensajeros (ARNm) y los ácidos ribonucleicos no codificantes (ARNs), tanto a nivel cuantitativo como a nivel cualitativo dando informaciones de gran interés (1,9).

Todo el patrón de expresión genética es variable a lo largo del tiempo debido a que los genes se están expresando en un momento dado, en una célula, un tejido o un organismo (1) además, puede revelar tanto la arquitectura de la expresión génica, como proporcionar detalles sobre la estructura, la modificación (10) y variación de transcriptoma individual (11). Mediante herramientas de alto rendimiento como es el RNA-seq (*RNA sequencing*) podemos estudiar la transcriptómica permitiendo obtener las secuencias de los transcritos, cuantificar el nivel de expresión génica de manera precisa y reproducible (8).

La investigación sobre el transcriptoma humano ha proporcionado soporte molecular de muchos procesos biológicos y enfermedades, entre estos destaca el cáncer (12).

#### 4.3 ARN-seq

Desde la primera aparición de la secuenciación de alto rendimiento en 2005 (13), ha sido posible comprender las actividades de la vida a nivel molecular y

realizar investigaciones detalladas para dilucidar el genoma y la transcriptoma (14).

Las muestras de especies y tejidos se pueden utilizar para la secuenciación de ARN por la identificación de los genes expresados, revelando así su función y posibles mecanismos moleculares (15), además gracias a la expresión génica diferencial se ha podido facilitar el descubrimiento de posibles biomarcadores del cáncer (16).

Esta secuenciación denominada “análisis diferencial” es el proceso donde las células son capaces de decodificar la información del material genético (los ácidos nucleicos) para la elaboración de los genes necesario de alto riesgo molecular como de riesgo funcionamiento del organismo (17); además es considerada una de las aplicaciones más comunes utilizada en la de secuenciación de ARN (18).

Como parte esencial de la secuenciación de alto rendimiento, la secuenciación de ARN (RNA-seq), pone a disposición información biológica analiza los determinantes de la heterogeneidad de la expresión infra tumoral e identifica la base molecular de la formación de muchas enfermedades oncológicas (12,14).

El establecimiento de RNA-seq es considerada como una herramienta analítica en transcriptómica de rápido desarrollo, que va reduciendo el costo de funcionamiento y ofreciendo la posibilidad de descubrir nuevos eventos relacionados con la transcripción (8).

Por tanto, la secuenciación del ARN ofrece información invaluable para la investigación y el tratamiento del cáncer (14).

El primer paso en la secuenciación del transcriptoma es el aislamiento de ARN de una muestra biológica, para garantizar un experimento de RNA-seq exitoso, el RNA debe tener la calidad suficiente para producir una biblioteca para la secuenciación (7).

El RNA-seq fue informado por primera vez por en 2009 (19) para abordar el problema de que las muestras que se podían obtener eran demasiado escasas para implementar el análisis del transcriptoma, que preferiblemente puede analizar las muestras a nivel de una sola célula (20).

En los últimos años, ha sido ampliamente adoptado por la comunidad científica y se ha transformado en el ensayo estándar para una amplitud de aplicaciones transcriptómicas (12), no es un protocolo único (21) sino más bien una familia de metodologías relacionadas (12).

La elección de la herramienta se basa en el propósito y la precisión de la aplicación (22,23), de hecho, un proceso de análisis de datos de secuenciación de ARN implica a grandes rasgos:

- a) El control de calidad de los datos sin procesar (en forma de archivo FASTQ) (14,24).
- b) La alineación con el genoma o transcriptoma de referencia (14).
- c) Ensamblaje de la transcripción de lecturas cortas (mediante la utilización de diversos softwares como “Cufflinks”, “StringTie” y/o “Trinity” (25–29).
- d) La cuantificación de la expresión (tanto a nivel de gen como a nivel de transcripción o de exón) (14).
- e) El análisis de la expresión diferencial (donde se genera una matriz de expresión usando métodos estadísticos para la identificación de los genes expresados (utilizando las “DESeq2” y “edgeR”) (14).

Es gracias a la alta resolución de RNA-seq podría detectar mutaciones tempranas, tanto de alto riesgo molecular como de riesgo más bajo, por lo que puede descubrir nuevos biomarcadores de cáncer y objetivos terapéuticos potenciales, monitorear enfermedades y guiar la terapia dirigida durante las decisiones de tratamiento temprano (14) contribuyendo también a descubrimientos de nuevos objetivos y mecanismos de cáncer (20).

El RNA-seq tiene competencia para diseccionar genes del organismo, relacionados con células cancerígenas, una población de células pequeñas con firmas genéticas específicas (19).

Diferentes estudios experimentales y clínicos sostienen que los genes tienen la capacidad de auto renovarse contribuyendo a recurrencia y metástasis del tumor (20).

#### *4.4 Tejido epitelial*

El tejido epitelial es una barrera que separa los tejidos subyacentes de su entorno (30). En la mucosa oral queratinizada, el epitelio está compuesto por cuatro capas: el estrato basal, estrato espinoso, estrato granuloso y estrato córneo mientras que, en el epitelio no queratinizado, el estrato basal es seguido por el estrato filamentososo y el estrato distendum (31).

En la mucosa oral se diferencian distintos fenotipos:

a) *Mucosa de revestimiento*

Se localiza en las estructuras móviles de la cavidad bucal como por ejemplo el velo del paladar, las mejillas, los labios y la mucosa alveolar.

b) *Mucosa masticatoria*

Cubierta protectora rígida y resistente que encontramos en la encía y del paladar duro, está estrechamente unida por el tejido conjuntivo denso al hueso subyacente.

c) *Mucosa especializada*

Se localiza en el dorso de la lengua, presenta un epitelial queratinizado en el cual incluye papilas y linguales, gustativas (30,32).

Un mal funcionamiento de las células del tejido epitelial puede dar lugar al Cáncer; el proceso tumoral provoca la aparición de cáncer asociado a los fibroblastos (CAF) promoviendo la degradación de la matriz favoreciendo su progresión (33).

#### 4.5 Lesiones Precancerígenas

En la actualidad, el término «lesiones precancerosas» o «pre-malignas» ha sido aceptada por Warnakulsuriya el cual propone denominarlas “lesiones potencialmente malignas”.

Una lesión pre-cancerígena se define como una alteración a nivel de la mucosa oral o tejido en la cual se puede desarrollar con mayor riesgo el cáncer oral (34).

La OMS la identifica como lesiones crónicas sobre las cuales es más frecuente en desarrollo de un pre-cáncer. Su morfología tiende a tener mayor predisposición en el desarrollarse en cáncer (+5%) que un tejido normal (35).

Existen diferentes lesiones potencialmente pre-malignas, entre estas encontramos la leucoplasia y la fibrosis (36).

#### *4.5.1 Fibrosis*

La fibrosis oral es una de las lesiones pre-cancerosas del carcinoma oral de células escamosas (COCE) (37), se considera estado atrófico de la mucosa bucal caracterizado por una inflamación del epitelio, una alteración fibroelástica de la lámina propia y una atrofia epitelial (36), la cual, ocasionalmente pueden extenderse hasta la faringe y el primer tercio del esófago (38).

Inicialmente se presenta con ampollas, vesículas o úlceras y solo con el avance de la lesión cambia en su forma más conocida: zona pálida y densa de la mucosa con induración y/o alteraciones eritematosas que pueden ser acompañada con sensaciones urente, trastorno en la masticación, deglución, fonación y trismo (36,38).

Histopatológicamente se trata de células inflamatorias crónicas con componente eosinofílico que se introducen en los tejidos conectivos subepiteliales.

La lesión más antigua tiene una vascularización reducida, un número reducido de células inflamatorias y haces densos y láminas de colágeno depositadas inmediatamente debajo del epitelio formando la placa blanca que la caracteriza (39).

Hoy en día la tasa de malignidad está en aumento, pero aún se desconoce el mecanismo de la malignidad (37).

No existe un tratamiento eficaz para la fibrosis oral y como la condición es irreversible una vez que se forma los pacientes deben tener un seguimiento estrecho y regular con su propio medico (39).

#### *4.5.2 Leucoplasia*

La leucoplasia es la lesión pre-maligna más frecuente de la mucosa oral y su prevalencia en España se sitúa en torno al 0,5% y el 1% (36).

Históricamente, el término leucoplasia se utilizaba para definir cualquier parche (o placa) blanco adherente (40), con el avance de los años se definió



como una placa blanca de mucosa elevada, bien delimitada, de 5 mm de diámetro, que no puede desprenderse al raspado y no puede clasificarse como otra lesión (36).

La leucoplasia oral puede presentarse de forma asintomática o mostrar una apariencia clínica benigna esto es una de la motivación por lo que dificulta que el médico a veces diferencie los trastornos reactivos o inflamatorios (benignos) comunes de la mucosa oral (40).

Alrededor del 5% hasta un 25% de esta lesión se desarrollan hacia neoplasia maligna (41) el riesgo anual de transformación maligna se del 2% hasta un 2% (42) por este motivo si no tenemos un descarte de nivel histológico todas las leucoplasia se consideran como una lesión pre-maligna (41).

Varios estudios han demostrado una coincidencia de leucoplasia en el momento del diagnóstico de carcinoma oral de las células escamosa (OSCC) de un 36% cuando la misma lesión presenta características displásica y un 15% sin características displásica (39).

Hay muchos factores de alarma que nos pueden indicar la transformación maligna de la lesión, como, por ejemplo, el tamaño, el subtipo clínico, la zona oral y la presencia o ausencia de displasia epitelial, la cual se evalúa con un examen histopatológico (43).

El diagnóstico se basa en la inspección visual y la palpación manual y se reconocen dos tipos de leucoplasia: homogéneas y no homogéneas (en esta última encontramos las eritroleucoplasias de formas nodulares o formas exofíticas) (36).

1. La leucoplasia homogénea es uniformemente plana y delgada, tiene una superficie lisa y pueden presentar grietas poco profundas no induradas (36,40).
2. Las leucoplasia no homogéneas suelen ser lesiones blancas o rojizas, irregulares, nodulares o exofíticas que suelen ser sintomáticas y tiene mayor tendencia de alteraciones displásicas respecto a la homogénea (36).

Las lesiones orales pre-cancerosas y cancerosas pueden dar inicio a cualquier cantidad de lesiones orales benignas, pero también pueden dejarse sin investigación y tratamiento por parte del paciente hasta que estén muy avanzadas (44).

#### *4.6 Cáncer Oral*

El cáncer es un problema de salud a nivel mundial y es una de las causas más importantes de mortalidad (45).

En las últimas décadas la detección precoz del cáncer oral está adquiriendo cada vez más importancia debido el aumento de los casos nuevos en la población (35).

Uno de los cánceres orales más frecuente es lo de cabeza y cuello; se encuentra dentro de las 10 neoplasias malignas más comunes y se observa un aumento de los casos diagnosticados en los últimos 10 años (46) y junto con los cánceres orofaríngeos, son la sexta neoplasia maligna más común en todo el mundo (47).

El cáncer oral en general tiene un pronóstico malo, con unas tasas de supervivencia global a los 5 años tan bajas como el 40%; pero si se diagnostica en las primeras etapas (I y II), las tasas de supervivencia pueden superar el 80 % (48).

La incidencia de cáncer oral aumenta con la edad, siendo más alta a partir de los 60 años, aunque los casos en personas menores de 40 años están aumentando (49).

Hasta el 50% de los cánceres se diagnostican en una etapa avanzada (III y IV), debido a que la mayoría de los pacientes no presentan síntomas en las primeras etapas y no buscan ayuda médica hasta síntomas claros como dolor, sangrado o una masa. en la boca o el cuello (44).

Cuando el diagnóstico supera un mes, el riesgo de tener un cáncer oral en etapa avanzada es significativamente mayor (50).

Como regla general, el pronóstico va a degenerar a medida que la enfermedad va avanzando y el sitio del tumor se vuelve menos accesible (47)

además se consideran lesiones potencialmente malignas cuando presentan un aspecto más blanco, rojos o moteadas (44).

Hay varios factores de riesgo conocido, pero, durante mucho tiempo el consumo de tabaco, incluido el tabaco sin humo y el consumo excesivo de alcohol, son considerados los principales factores etiológicos para el desarrollo del cáncer oral (51).

Otro factor de riesgo considerado relacionado con el cáncer es el uso y abuso de alcohol (44).

También la prevalencia de cáncer oral y orofaríngeo está asociada al VPH (principalmente VPH tipo 16) ha ido en aumento en las últimas décadas, predominantemente entre las personas más jóvenes (47).



**Universidad**  
**Europea** VALENCIA

#### **4. JUSTIFICACIÓN**

Los estudios de las enfermedades pre-cancerosa y del cáncer oral son de suma importancia debido a un aumento en el número de casos en los últimos años (52) por lo que su conocimiento, prevención y diagnóstico se va enfatizando cada día más para los odontólogos, médicos y científicos.

Es gracias a las nuevas tecnologías que el estudio del cáncer ha obtenido un avance siempre más completo, sin embargo, a pesar de la existencia de números estudios utilizando la tecnología RNA-seq para estas patologías (14,53–55) se necesita delucitar más genes que podrían ser responsables de estas patologías, así como entender mejor las funciones biológicas pudiendo en un futuro ayudar a las terapias génicas.

De esta manera, identificar genes que están siendo activados comunes a diferentes lesiones precancerosas y también con el cáncer oral comparado con personas sin patologías, podrían ser la clave para ayudar a la detección precoz de estas patologías. Existen diferentes estudios relacionados con los genes y el cáncer de páncreas, mama (56–58).

Poco son los estudios (59) que están hechos sobre el tejido epitelial de la mucosa oral y menos relacionan los genes que se pueden encontrar en las enfermedades pre-cancerígena y cancerígena con pacientes sanos utilizando la secuenciación del ARN.

Este trabajo nos permitirá en mostrar cuáles son los genes que podrían estar relacionados tanto en una leucoplasia y/o fibrosis (lesiones pre-cancerígenas) como en una lesión cancerígena (cáncer oral), además permite de evaluar cuales genes están siendo más activados cuando una persona presenta estas enfermedades respecto a una persona sana utilizando la técnica de secuenciación ARN (ARN-seq).

#### **5. HIPÓTESIS**

La hipótesis de nuestro estudio considera que reanalizando con la misma metodología bioinformática los datos brutos secuenciando el RNA mediante la técnica RNA-seq de diferentes bases de datos públicas existen genes

diferenciales comunes en muestras de pacientes con cáncer, leucoplasia y fibrosis respecto a muestras de pacientes que no presentan patologías en la cavidad oral.

*Hipótesis nula:* si se encuentran genes diferenciales comunes entre pacientes con cáncer, leucoplasia y fibrosis respecto a muestras de pacientes que no presentan patologías en la cavidad oral.

*Hipótesis alternativa:* Se rechaza la hipótesis nula: no se encuentran genes diferenciales comunes con cáncer oral, leucoplasia y fibrosis respecto a muestras de pacientes que no presentan patologías en la cavidad oral.



**Universidad**  
**Europea** VALENCIA

## 6. OBJETIVOS

### 7.1 General

Realizar un estudio transcriptómico e identificar los procesos biológicos de interés en la cavidad oral en humanos mediante la técnica de RNA-seq para diferentes muestras obtenidas de bases de datos públicas clasificadas en grupos divergentes (muestras de pacientes con lesiones pre-cancerosas comparada con muestras de pacientes sin patologías).

### 7.2 Específicos

- Detectar los genes diferenciales (sobre expresados e infra expresados) comunes entre varias bases de datos de pacientes con cáncer oral *versus* tejido sano, fibrosis *versus* tejido sano y leucoplasia *versus* tejido sano.
- Realizar un análisis de enriquecimiento para identificar procesos biológicos significativos para cáncer oral, fibrosis y leucoplasia.
- Identificar los genes diferenciales comunes más relevantes asociados a cáncer y lesiones precancerosas para en un futuro podrían utilizarse como biomarcadores.





**Universidad**  
**Europea** VALENCIA

## 7. MATERIALES Y MÉTODOS

### 8.1 Diseño del estudio

#### 8.1.1 Selección de los estudios

Fueron seleccionados diferentes estudios con la opción de descarga de datos brutos de secuenciación del RNA de muestras de pacientes a través de la técnica de RNA-seq, en las bases de datos públicas de Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), repositorio de datos de genómica funcional pública, datos basados en matrices y secuencias y BioStudies (<https://www.ebi.ac.uk/biostudies/arrayexpress>) que contiene descripciones de estudios biológicos, enlaces a datos de estos estudios en otras bases de datos en EMBL-EBI o fuera, así como datos que no caben en los archivos estructurados en EMBL-EBI.

El criterio utilizado para la búsqueda de los estudios para la base de datos de Gene Expression Omnibus fue: HNSCC [All Fields] OR ("mouth neoplasms" [MeSH Terms] OR ORAL CANCER [All Fields]) OR OSCC [All Fields] AND RNA SEQ [All Fields] NOT ("cell line" [MeSH Terms] OR CELL LINES[All Fields]) NOT scRNA-seq [All Fields] AND ("gse"[Filter] AND "Homo sapiens"[Organism]).

Lo resultante de la búsqueda fueron 14, de los cuales excluimos:

- Estudios mencionado otro cáncer que no fuera cáncer oral.
- Estudios con herpes.
- Estudios con líneas celulares o comparando líneas celulares.
- Estudios utilizando otras técnicas de secuenciación como Single-Cell

RNA.

Fueron seleccionados dos estudios con el número de acceso a la base de datos de GSE125866 y GSE20116.

El primer estudio encontramos muestras secuenciadas de pacientes con cáncer, pacientes con fibrosis y pacientes normales.

En el segundo estudio encontramos pacientes con cáncer y pacientes normales. Decidimos incluir un estudio más con el número de acceso a la base

de datos de GSE131568 con muestras de pacientes secuenciados de Leucoplasia.

El criterio de selección para la base de datos BioStudies (incluyen estudios de otras bases de datos en EMBL-EBI o fuera) [https://www.ebi.ac.uk/biostudies/arrayexpress/studies?facet.study\\_type=rna-seq%20of%20coding%20rna&facet.study\\_type=transcription%20profiling%20by%20array&facet.organism=homo%20sapiens&facet.assay\\_by\\_molecule=rna%20assay&facet.raw=true&query=OSCC&page=2](https://www.ebi.ac.uk/biostudies/arrayexpress/studies?facet.study_type=rna-seq%20of%20coding%20rna&facet.study_type=transcription%20profiling%20by%20array&facet.organism=homo%20sapiens&facet.assay_by_molecule=rna%20assay&facet.raw=true&query=OSCC&page=2)

En la búsqueda se seleccionó en la palabra: OSCC.

En el tipo de estudio se seleccionó RNA-seq con codificación del RNA y perfiles de transcripción por matriz.

En el organismo se seleccionó: *homo sapiens*.

En los ensayos se seleccionó: *ensayo de RNA y muestras sin procesar*.

Los resultados de la búsqueda fueron 31 estudios, de los cuales se excluyó:

- Estudios con líneas celulares.
- Estudios con estrategias de siRNA.
- Estudios de *non coding RNA*.

No se incluyó ningún estudio más de esta base de datos, dado que había duplicidad con la primera base de datos consultada (Gene Expression Omnibus).

### *8.1.2 Selección de datos brutos para los análisis diferencial*

El estudio **GSE125866** tenemos n=40 muestras de tejido de la cavidad oral, de las cuales n=10 son de pacientes que presentan fibrosis, n=20 muestras de pacientes que presentan cáncer y n=10 de pacientes que no muestran patología y la consideramos muestras control (normal).

El estudio **GSE20116** tenemos n=20 muestras de tejido de la cavidad oral, de las cuales n=6 muestras de pacientes que presentan cáncer y n=14 de pacientes que no muestran patología y la consideramos muestras control (normal).

El estudio **GSE131568** tenemos n=26 muestras de tejido de la cavidad oral, de las cuales n=6 muestras de pacientes que presentan leucoplasia y n=20 de pacientes que no muestran patología y la considerados muestras control (normal).

Realizamos 4 análisis diferenciales con los datos brutos de las muestras de los pacientes de los 3 estudios seleccionados:

- Para el primer análisis seleccionamos las muestras del estudio GSE125866, 10 muestras de fibrosis las comparamos con 10 muestras control.
- El segundo análisis seleccionamos las muestras del estudio GSE125866, 10 muestras de cáncer las comparamos con 10 muestras control.
- El tercer análisis seleccionamos las muestras del estudio GSE20116, 6 muestras de cáncer las comparamos con 14 muestras control.
- El cuarto análisis seleccionamos las muestras del estudio GSE131568, 6 muestras de Leucoplasia las comparamos con 20 muestras control.

### *8.1.3 Análisis de expresión diferencial*

Todos los archivos brutos correspondientes a cada muestra de los estudios se analizaron con el mismo flujo de trabajo.

La calidad de las lecturas de RNA-Seq (índices de calidad, contenido de GC, contenido de N, distribuciones de longitud, duplicación, secuencias sobrerrepresentadas y contenido de K-mer) se verificó mediante el software Fastqc (v.0.11.4) (60).

Las lecturas de baja calidad (secuencia del adaptador y lecturas que contienen poli-N) de datos sin procesar usando Trimmomatic v.0.36 (61) con los siguientes parámetros: `java -jar /usr/local/bin/trimmomatic-0.36.jar PE -phred33 <IN_R1.fastq.gz> <IN_R2.fastq.gz> <OUT_R1.fastq.gz> <OUT_R1_UN.fastq.gz> <OUT_R2.fastq.gz> <OUT_R2_UN.fastq.gz> LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50.`

Todo el análisis posterior se basó en los datos recortados con lecturas de alta calidad, se utilizó KALLISTO (62) para alinear las lecturas recortadas de

extremos emparejados con el genoma de referencia Homo Sapiens depositado en el Centro Nacional de Información Biotecnológica (NCBI) (<https://www.ncbi.nlm.nih.gov/>).

Este recuento de lecturas se llevó a cabo utilizando el script interno y Ensembl (Índice de /pub/release-105/gtf/homo sapiens/ (ensembl.org)) para las anotaciones de los genes.

Los paquetes EdgeR y Limma disponibles del proyecto Bioconductor se utilizaron para estimar lecturas de CPM normalizadas (recuentos por millón) e identificar genes expresados diferencialmente (DEG) entre los grupos.

Se utilizó EdgeR para normalizar los datos y el software Limma para los métodos estadísticos (modelado lineal y Bayes empírico), siguiendo la metodología de Law et al. (63).

Consideramos que un gen se expresa diferencialmente cuando el valor de FDR (false positive) es  $\leq 0.05$ .

#### *8.1.4 Análisis de enriquecimiento*

Se realizó análisis de enriquecimiento de los genes diferenciales encontrados se realizaron utilizando la base de datos para anotación, visualización y descubrimiento integrado DAVID v.6.8 (<https://david.ncifcrf.gov>) para mapear las rutas enriquecidas de genes usando la base de datos de la Enciclopedia de Genes y Genomas de Kyoto (KEGG), se determinó p value  $\leq 0,05$  como procesos biológicos significativos.

Por último, una vez obtenidos los genes diferenciales más relevantes, se utilizó el ClueGO, un complemento de Cytoscape software que visualiza los términos biológicos no redundantes para grandes grupos de genes en una red agrupada funcionalmente (65).



**Universidad**  
**Europea** VALENCIA

## 8. RESULTADOS

### *8.1 Genes diferenciales entre pacientes con Cáncer Oral, Leucoplasia y Fibrosis*

Se han encontrado 800 genes infraexpresado y 608 muestra sobreexpresado en el primer estudio (cáncer vs no cáncer), 102 genes infraexpresado y 8 sobreexpresado en el segundo estudio (fibrosis vs no cancer) y 2567 genes infraexpresado y 2603 muestras sobreexpresado en el tercer estudio (leucoplasia vs no cancer).

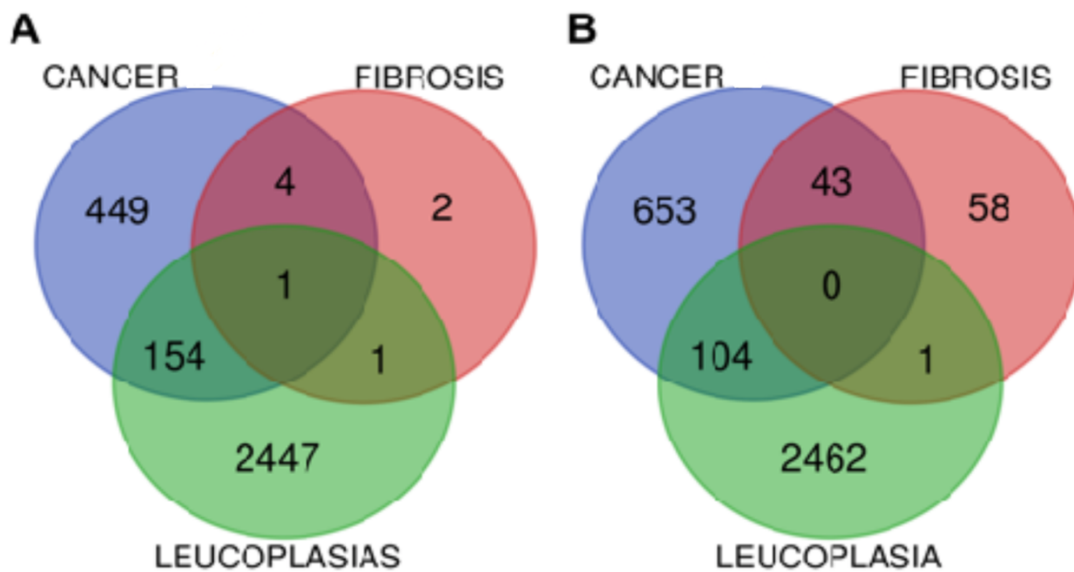
Todos los genes diferenciales obtenidos se aplicó la corrección FDR (*False Discovery Rate*) menor que 0,05.

### *9.2 Genes diferenciales comunes entre pacientes con Cáncer Oral, Leucoplasia y Fibrosis*

Se ha detectado un total de 160 genes comunes sobreexpresados las diferentes patologías. Se ha identificado un solo gen sobreexpresado común entre las tres patologías, 154 genes comunes entre el cáncer oral y la leucoplasia, un gen común entre fibrosis y leucoplasia y 4 genes comunes entre cáncer oral y fibrosis (*Figura 1A*).

Se ha detectado un total de 148 genes comunes infraexpresado comparando las diferentes patologías.

No se detectaron genes infraexpresados comunes entre las tres patologías; entre el cáncer oral y la leucoplasia encontramos 104 genes comunes, entre fibrosis y leucoplasia 1 y entre cáncer oral y fibrosis 43 (*Figura 1B*).



**Figura 1.A.** Venn Diagramm de los genes comunes entre sobreexpresados entre Leucoplasia, Cáncer Oral y Fibrosis. **1.B.** Venn Diagramm de los genes comunes infraexpresados entre Leucoplasia, Cáncer Oral, Fibrosis.

En la Tabla 1 se puede observar los genes comunes sobre expresados para Cáncer Oral (C) comparado con el grupo sano (N) y genes sobre expresados en el grupo de leucoplasia (L) comparado con pacientes sanos (N).

**Tabla 1:** Genes comunes sobreexpresado entre cáncer oral y leucoplasia.

<i>ID</i>	<b>Nombre del Gen</b>	<b>Fold change (log2)* C vs N</b>	<b>Fold change (log2)* L vs N</b>
<b>ACLY</b>	<i>ATP Citrate Lyase</i>	1.057	0.087
<b>ACTN1</b>	<i>Actinin Alpha 1</i>	1.340	0.074
<b>ADAM19</b>	<i>ADAM Metallopeptidase Domain 19</i>	1.342	0.201
<b>ADM</b>	<i>Adrenomedullin</i>	1.708	0.263
<b>AEN</b>	<i>Apoptosis Enhancing Nuclease</i>	1.371	0.339
<b>ALYREF</b>	<i>Aly/REF Export Factor</i>	1.385	0.117
<b>ATAD3A</b>	<i>ATPase Family AAA Domain Containing 3A</i>	1.434	0.159
<b>ATF5</b>	<i>Activating Transcription Factor 5</i>	1.156	0.156
<b>BOP1</b>	<i>Ribosomal Biogenesis Factor</i>	1.362	0.159



<b>C1orf216</b>	<i>Chromosome 1 Open Reading Frame 216</i>	1.091	0.481
<b>CARD10</b>	<i>Caspase Recruitment Domain Family Member 10</i>	1.377	0.174
<b>CARS</b>	<i>Nuclear Receptor Subfamily 1 Group I Member 3</i>	1.024	0.062
<b>CCT3</b>	<i>Chaperonin Containing TCP1 Subunit 3</i>	1.106	0.135
<b>CCT4</b>	<i>Chaperonin Containing TCP1 Subunit 4</i>	0.449	0.042
<b>CDV3</b>	<i>CDV3 Homolog</i>	0.37	0.158
<b>CEBPB</b>	<i>CCAAT Enhancer Binding Protein Beta</i>	1.581	0.332
<b>CFL1</b>	<i>Cofilin 1</i>	0.921	0.068
<b>CHST2</b>	<i>Carbohydrate Sulfotransferase 2</i>	3.001	0.781
<b>CHST3</b>	<i>Carbohydrate Sulfotransferase 3</i>	1.069	0.351
<b>CLIC4</b>	<i>Chloride Intracellular Channel 4</i>	0.728	0.138
<b>CLSPN</b>	<i>Claspin</i>	2.011	0.203
<b>COL27A1</b>	<i>Collagen Type XXVII Alpha 1 Chain</i>	1.357	0.114
<b>COL4A1</b>	<i>Collagen Type IV Alpha 1 Chain</i>	1.386	0.02
<b>CORO1C</b>	<i>Coronin 1C</i>	0.676	0.08
<b>CPA4</b>	<i>Carboxypeptidase A4</i>	1.993	0.329
<b>CTPS1</b>	<i>CTP Synthase 1</i>	1.327	0.077
<b>CTSL</b>	<i>Cathepsin L</i>	1.453	0.222
<b>CYP19A1</b>	<i>Cytochrome P450 Family 19 Subfamily A Member 1</i>	3.076	1.204
<b>CYP24A1</b>	<i>Cytochrome P450 Family 24 Subfamily A Member 1</i>	3.486	1.103
<b>DAP3</b>	<i>Death Associated Protein 3</i>	0.508	0.085
<b>DAPK3</b>	<i>Death Associated Protein Kinase 3</i>	1.46	0.104
<b>DDX21</b>	<i>DExD-Box Helicase 21</i>	1.452	0.116
<b>DDX23</b>	<i>DEAD-Box Helicase 23</i>	0.802	0.048
<b>DESI1</b>	<i>Desumoylating Isopeptidase 1</i>	0.838	0.147
<b>DKC1</b>	<i>Dyskerin Pseudouridine Synthase 1</i>	0.739	0.172
<b>EBNA1BP2</b>	<i>EBNA1 Binding Protein 2</i>	0.999	0.144
<b>EEF1D</b>	<i>Eukaryotic Translation Elongation Factor 1 Delta</i>	0.704	0.06
<b>EFNB1</b>	<i>Ephrin B1</i>	1.484	0.3
<b>EHD1</b>	<i>EH Domain Containing 1</i>	1.133	0.136
<b>EIF3B</b>	<i>Eukaryotic Translation Initiation Factor 3 Subunit B</i>	0.684	0.096

<b>EIF4A1</b>	<i>Eukaryotic Translation Initiation Factor 4A1</i>	0.899	0.055
<b>ELAC2</b>	<i>ElaC Ribonuclease Z 2</i>	0.685	0.061
<b>ENTPD7</b>	<i>Ectonucleoside Triphosphate Diphosphohydrolase 7</i>	1.183	0.089
<b>EPHB2</b>	<i>EPH Receptor B2</i>	1.727	0.615
<b>EPPK1</b>	<i>Epiplakin 1</i>	3.504	0.196
<b>ETV4</b>	<i>ETS Variant Transcription Factor 4</i>	1.528	0.59
<b>FBXO45</b>	<i>F-Box Protein 45</i>	2.134	0.218
<b>FMNL2</b>	<i>Formin Like 2</i>	1.37	0.084
<b>FSCN1</b>	<i>Fascin Actin-Bundling Protein 1</i>	1.996	0.224
<b>FSD1L</b>	<i>Fibronectin Type III And SPRY Domain Containing 1 Like</i>	0.831	0.328
<b>FTSJ1</b>	<i>FtsJ RNA 2'-O-Methyltransferase 1</i>	0.661	0.126
<b>GALNT18</b>	<i>Polypeptide N-Acetylgalactosaminyltransferase 18</i>	1.001	0.17
<b>GALNT6</b>	<i>Polypeptide N-Acetylgalactosaminyltransferase 6</i>	2.704	0.434
<b>GPATCH4</b>	<i>G-Patch Domain Containing 4 (Gene/Pseudogene)</i>	0.934	0.143
<b>GPR153</b>	<i>G Protein-Coupled Receptor 153</i>	1.377	0.237
<b>HARS</b>	<i>Histidyl-tRNA synthetase</i>	0.507	0.057
<b>HOMER3</b>	<i>Homer Scaffold Protein 3</i>	1.873	0.301
<b>HSP90B1</b>	<i>Heat Shock Protein 90 Beta Family Member 1</i>	0.730	0.053
<b>HSPA5</b>	<i>Heat Shock Protein Family A (Hsp70) Member 5</i>	1.242	0.068
<b>HSPH1</b>	<i>Heat Shock Protein Family H (Hsp110) Member 1</i>	1.199	0.127
<b>HYOU1</b>	<i>Hypoxia Up-Regulated 1</i>	0.682	0.106
<b>INF2</b>	<i>Inverted Formin 2</i>	0.734	0.115
<b>ITGA3</b>	<i>Integrin Subunit Alpha 3</i>	1.710	0.092
<b>KIAA0930</b>	<i>KIAA0930</i>	0.740	0.223
<b>KLF7</b>	<i>KLF Transcription Factor 7</i>	1.114	0.193
<b>KRT17</b>	<i>Keratin 17</i>	5.685	0.683
<b>LAMC2</b>	<i>Laminin Subunit Gamma 2</i>	3.083	0.101
<b>LIMK1</b>	<i>LIM Domain Kinase 1</i>	1.278	0.225
<b>LMNB2</b>	<i>Lamin B2</i>	1.505	0.111
<b>LRP8</b>	<i>LDL Receptor Related Protein 8</i>	1.381	0.338
<b>LRRC59</b>	<i>Leucine Rich Repeat Containing 59</i>	1.000	0.171

<b>LY6K</b>	<i>Lymphocyte Antigen 6 Family Member K</i>	2.447	0.696
<b>LYAR</b>	<i>Ly1 Antibody Reactive</i>	1.209	0.188
<b>MDFI</b>	<i>MyoD Family Inhibitor</i>	1.698	0.203
<b>MET</b>	<i>MET Proto-Oncogene, Receptor Tyrosine Kinase</i>	1.051	0.144
<b>MICAL2</b>	<i>Microtubule Associated Monooxygenase, Calponin And LIM Domain Containing 2</i>	1.748	0.085
<b>MRPL14</b>	<i>Mitochondrial Ribosomal Protein L14</i>	1.042	0.156
<b>MSN</b>	<i>Moesin</i>	0.726	0.228
<b>MTHFD1L</b>	<i>Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1 Like</i>	1.549	0.175
<b>MTHFD2</b>	<i>Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 2, Methenyltetrahydrofolate Cyclohydrolase</i>	1.017	0.156
<b>MYBL2</b>	<i>MYB Proto-Oncogene Like 2</i>	1.55	0.13
<b>MYO1B</b>	<i>Myosin IB</i>	1.978	0.233
<b>NACC1</b>	<i>Nucleus Accumbens Associated 1</i>	0.933	0.135
<b>NADK</b>	<i>NAD Kinase</i>	0.79	0.117
<b>NCF2</b>	<i>Neutrophil Cytosolic Factor 2</i>	1.123	0.385
<b>NKRF</b>	<i>NFKB Repressing Factor</i>	0.729	0.185
<b>NOC2L</b>	<i>NOC2 Like Nucleolar Associated Transcriptional Repressor</i>	0.907	0.052
<b>NOMO1</b>	<i>NODAL Modulator 1</i>	0.881	0.082
<b>NOP56</b>	<i>NOP56 Ribonucleoprotein</i>	0.964	0.079
<b>NPLOC4</b>	<i>NPL4 Homolog, Ubiquitin Recognition Factor</i>	0.080	0.091
<b>NRIP1</b>	<i>Nuclear Receptor Interacting Protein 2</i>	1.224	0.088
<b>NSUN2</b>	<i>NOP2/Sun RNA Methyltransferase 2</i>	0.944	0.063
<b>NUP62</b>	<i>Nucleoporin 62</i>	0.587	0.076
<b>OAS3</b>	<i>2'-5'-Oligoadenylate Synthetase 3</i>	1.877	0.094
<b>PA2G4</b>	<i>Proliferation-Associated 2G4</i>	0.869	0.111
<b>PDCD11</b>	<i>Programmed Cell Death 11</i>	1.149	0.068
<b>PFDN2</b>	<i>Prefoldin Subunit 2</i>	1.101	0.359
<b>PLAU</b>	<i>Plasminogen Activator, Urokinase</i>	2.666	0.096
<b>PLCD3</b>	<i>Phospholipase C Delta 3</i>	1.228	0.085
<b>PLEC</b>	<i>Plectin</i>	1.528	0.068
<b>PLOD3</b>	<i>Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase 3</i>	0.831	0.194

<b>PMEPA1</b>	<i>Prostate Transmembrane Protein, Androgen Induced 1</i>	1.725	0.238
<b>PNLIPRP3</b>	<i>Pancreatic Lipase Related Protein 3</i>	4.07	2.771
<b>PNP</b>	<i>Purine Nucleoside Phosphorylase</i>	1.132	0.109
<b>PPP1R14C</b>	<i>Protein Phosphatase 1 Regulatory Inhibitor Subunit 14C</i>	1.805	0.239
<b>PPP2R5B</b>	<i>Protein Phosphatase 2 Regulatory Subunit B'Beta</i>	1.053	0.154
<b>PRNP</b>	<i>Prion Protein</i>	1.164	0.253
<b>PSMB7</b>	<i>Proteasome 26S Subunit, Non-ATPase 7</i>	0.822	0.71
<b>PSMC3</b>	<i>Proteasome 26S Subunit, ATPase 3</i>	1.387	0.102
<b>PSMD2</b>	<i>Proteasome 26S Subunit Ubiquitin Receptor, Non-ATPase 2</i>	1.095	0.98
<b>PSMD3</b>	<i>Proteasome 26S Subunit, Non-ATPase 3</i>	0.607	0.103
<b>PTH1H</b>	<i>Parathyroid Hormone Like Hormone</i>	2.367	0.539
<b>PTPRZ1</b>	<i>Protein Tyrosine Phosphatase Receptor Type Z1</i>	2.561	0.062
<b>PTTG1IP</b>	<i>PTTG1 Interacting Protein</i>	0.728	0.129
<b>PUS1</b>	<i>Pseudouridine Synthase 1</i>	0.954	0.171
<b>RAD23A</b>	<i>RAD23 Homolog A, Nucleotide Excision Repair Protein</i>	0.834	0.157
<b>RGS20</b>	<i>Regulator Of G Protein Signaling 20</i>	2.397	0.654
<b>RRAS2</b>	<i>RAS Related 2</i>	0.835	0.171
<b>RRS1</b>	<i>Ribosome Biogenesis Regulator 1 Homolog</i>	1.374	0.413
<b>SDF4</b>	<i>Stromal Cell Derived Factor 4</i>	0.955	0.065
<b>SDK2</b>	<i>Sidekick Cell Adhesion Molecule 2</i>	1.792	0.564
<b>SEMA3C</b>	<i>Semaphorin 3C</i>	1.789	0.337
<b>SERPINB7</b>	<i>Serpin Family B Member 7</i>	3.093	0.737
<b>SH2D2A</b>	<i>SH2 Domain Containing 2A</i>	1.402	0.301
<b>SH3D21</b>	<i>SH3 Domain Containing 21</i>	1.216	0.112
<b>SLC15A1</b>	<i>Solute Carrier Family 15 Member 1</i>	2.131	0.258
<b>SLC16A1</b>	<i>Transmembrane Channel Like 7</i>	1.861	0.358
<b>SLC28A3</b>	<i>Solute Carrier Family 28 Member 3</i>	3.059	0.433
<b>SLC3A2</b>	<i>Solute Carrier Family 3 Member 2</i>	1.411	0.08
<b>SMOX</b>	<i>Spermine Oxidase</i>	2.494	0.339
<b>SQLE</b>	<i>Squalene Epoxidase</i>	1.701	0.138
<b>STIP1</b>	<i>Stress Induced Phosphoprotein 1</i>	1.068	0.126

<b>STRIP2</b>	<i>Striatin Interacting Protein 2</i>	2.287	0.633
<b>STX1A</b>	<i>Syntaxin 1A</i>	1.537	0.613
<b>SZRD1</b>	<i>SUZ RNA Binding Domain Containing 1</i>	0.438	0.138
<b>TGM2</b>	<i>Transglutaminase 2</i>	2.319	0.414
<b>THOC5</b>	<i>THO Complex Subunit 5</i>	0.468	0.066
<b>TIMM17A</b>	<i>Translocase Of Inner Mitochondrial Membrane 17A</i>	0.781	0.143
<b>TMC7</b>	<i>Transmembrane Channel Like 7</i>	1.244	0.602
<b>TMEM184B</b>	<i>Transmembrane Protein 184B</i>	1.027	0.17
<b>TNC</b>	<i>Tenascin C</i>	2.178	0.214
<b>TOP1MT</b>	<i>DNA Topoisomerase I Mitochondrial</i>	1.18	0.155
<b>TRMT6</b>	<i>TRNA Methyltransferase 6 Non- Catalytic Subunit</i>	0.966	0.108
<b>TRPV3</b>	<i>Transient Receptor Potential Cation Channel Subfamily V Member 3</i>	2.871	0.366
<b>TUBB3</b>	<i>Tubulin Beta 3 Class III</i>	2.060	0.614
<b>TUBG1</b>	<i>Tubulin Gamma 1</i>	1.067	0.207
<b>UHRF1</b>	<i>Ubiquitin Like With PHD And Ring Finger Domains 1</i>	1.197	0.104
<b>USB1</b>	<i>U6 SnRNA Biogenesis Phosphodiesterase 1</i>	1.561	0.175
<b>WDR5</b>	<i>WD Repeat Domain 5</i>	0.841	0.056
<b>WDR53</b>	<i>WD Repeat Domain 53</i>	1.154	0.147
<b>WDR62</b>	<i>WD Repeat Domain 62</i>	0.998	0.284
<b>WDR66</b>	<i>WD Repeat Domain 66</i>	1.971	0.665
<b>YWHAG</b>	<i>Tyrosine 3- Monoxygenase/Tryptophan 5- Monoxygenase Activation Protein Gamma</i>	0.679	0.195
<b>ZNF622</b>	<i>Zinc Finger Protein 622</i>	0.544	0.073

\*The *Fold-change* estimados (cambios de expresión relativa de los genes).

En la Tabla 2 se puede observar los genes sobre expresados entre paciente que presenta Cáncer Oral (C) comparado con paciente sano (N) y paciente que presenta Fibrosis (F) comparado con paciente sano (N).

**Tabla 2:** Genes comunes sobreexpresado entre pacientes con cáncer y Fibrosis

<b>ID</b>	<b>Nombre gen</b>	<b>Fold change (log2)* C vs N</b>	<b>Fold change (log2)*</b>
-----------	-------------------	---	------------------------------------

			<b>F vs N</b>
<b>RP11-23P13.6</b>	<i>Long-non-coding</i>	1.291	1.366
<b>STARD 4</b>	<i>StAR Related Lipid Transfer Domain Containing 4</i>	1.292	1.150
<b>RP11-420K8.1</b>	<i>Long-non-coding RNA</i>	1.293	1.538
<b>NUS1</b>	<i>NUS1 Dehydrodolichyl Diphosphate Synthase Subunit</i>	1.294	0.614

El único gen sobre expresado común encontrado entre paciente que presentan cáncer oral, fibrosis o leucoplasia, comparados con pacientes sanos fue el EPRS (glutamyl-prolyl-tRNA synthetase 1) con un Fold change de 0.653 en el estudio entre cáncer (C) y paciente sano (N), un Fold change de 0.492 en el estudio entre fibrosis (F) y paciente sano (N) y un Fold change de 0.091 en el último estudio entre leucoplasia (L) y paciente sano (N).

## 9.2 Análisis de enriquecimiento de los genes

Se realizó un análisis de enriquecimiento de los 160 genes (Tabla 1) sobreexpresado comunes entre las tres patologías mediante el software DAVID (v6.8) siendo posible explorar las funciones biológicas (GO) de los genes diferenciales detectados (Tabla 3).

Se realizó utilizando las bases de datos secundarias GOTERM\_BP\_DIRECT, GOTERM\_CC\_DIRECT y GOTERM\_MF\_DIRECT. Destacamos Procesos biológicos (BP), procesos moleculares, (MF) y componentes celulares (CC).

**Tabla 3:** Procesos biológicos significativos involucrados 160 genes sobreexpresados comunes entre los 3 estudios.

Category	Term	Count	Pvalue	Genes
<b>GOTERM_MF_DIRECT</b>	GO:0051015~actin	7	0.00	MYO1B, FMNL2, ACTN1, CFL1, FSCN1, CORO1C, PLEC

	filament binding			
<b>UP_KW_MOLECULAR_FUNCTION</b>	KW-0009~Actin-binding	8	0.01	MYO1B, INF2, ACTN1, CFL1, FSCN1, MICAL2, CORO1C, PLEC
<b>GOTERM_CC_DIRECT</b>	GO:0005903~brush border	4	0.01	SLC15A1, MYO1B, ACTN1, PLEC
<b>GOTERM_BP_DIRECT</b>	GO:0030036~actin cytoskeleton organization	6	0.01	INF2, ACTN1, CFL1, LIMK1, FSCN1, CORO1C
<b>GOTERM_MF_DIRECT</b>	GO:0003779~actin binding	8	0.01	MYO1B, INF2, ACTN1, CFL1, FSCN1, MICAL2, MSN, PLEC
<b>GOTERM_BP_DIRECT</b>	GO:0007015~actin filament organization	5	0.02	MYO1B, SH3D21, ACTN1, FSCN1, CORO1C
<b>GOTERM_BP_DIRECT</b>	GO:0051017~actin filament bundle assembly	3	0.02	MYO1B, ACTN1, FSCN1
<b>GOTERM_CC_DIRECT</b>	GO:0030027~lamellipodium	5	0.04	TUBB3, CFL1, LIMK1, FSCN1, CORO1C
<b>UP_SEQ_FEATURE</b>	DOMAIN:Fibronectin type-III	5	0.04	FSD1L, PTPRZ1, TNC, EPHB2, SDK2
<b>UP_SEQ_FEATURE</b>	REGION:Leucine-zipper	4	0.05	CEBPB, EEF1D, DAPK3, ATF5

Destacamos 49 genes involucrados en los procesos biológicos significativos en la funciones moleculares y componentes moleculares (pvalor menor a 0,05) (Tabla 4).

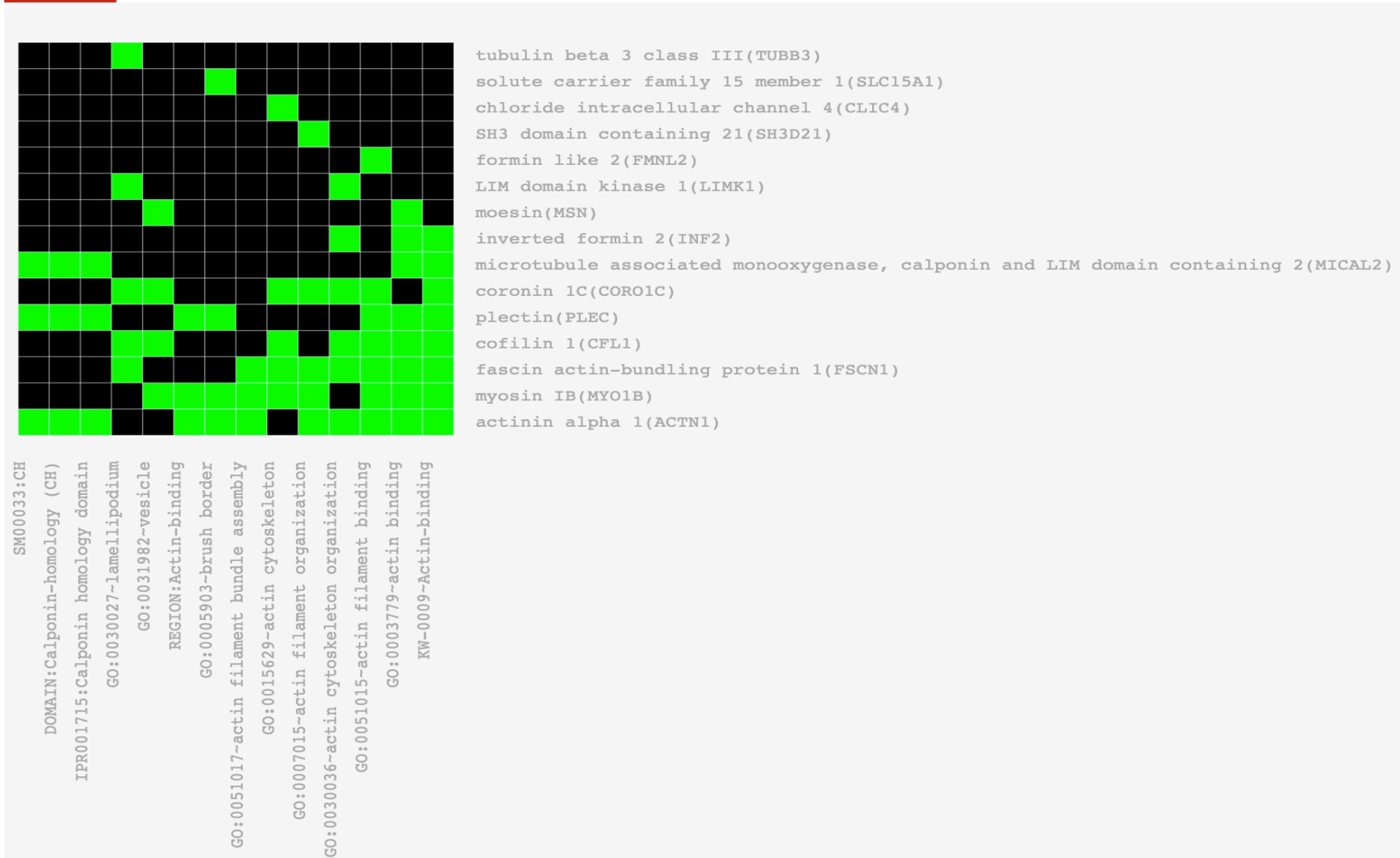
**Tabla 4.** Genes involucrados con Función Biológica, Función Molecular y Componente Molecular significativas.

<b>ID</b>	<b>Nombre gen</b>
<b>DDX21</b>	<i>DExD-Box Helicase 21</i>
<b>DDX23</b>	<i>DExD-Box Helicase 23</i>
<b>EEF1D</b>	<i>Eukaryotic Translation Elongation Factor 1 Delta</i>
<b>UHRF1</b>	<i>Ubiquitin Like With PHD And Ring Finger Domains 1</i>
<b>ACTN1</b>	<i>Actinin Alpha 1</i>
<b>ATF5</b>	<i>Activating Transcription Factor 5</i>
<b>CCT3</b>	<i>Chaperonin Containing TCP1 Subunit 3</i>
<b>CEBPB</b>	<i>CCAAT Enhancer Binding Protein Beta</i>
<b>CFL1</b>	<i>Cofilin 1</i>
<b>CLIC4</b>	<i>Chloride Intracellular Channel 4</i>
<b>CORO1C</b>	<i>Coronin 1C</i>
<b>DAPK3</b>	<i>Death Associated Protein Kinase 3</i>
<b>EPHB2</b>	<i>EPH Receptor B2</i>
<b>EPPK1</b>	<i>Epiplakin 1</i>
<b>ETV4</b>	<i>ETS Variant Transcription Factor 4</i>
<b>FMNL2</b>	<i>Formin Like 2</i>
<b>FSCN1</b>	<i>Fascin Actin-Bundling Protein 1</i>
<b>FSD1L</b>	<i>Fibronectin Type III And SPRY Domain Containing 1 Like</i>
<b>INF2</b>	<i>Inverted Formin 2</i>
<b>KLF7</b>	<i>KLF Transcription Factor 7</i>
<b>KRT17</b>	<i>Keratin 17</i>
<b>LIMK1</b>	<i>LIM Domain Kinase 1</i>
<b>LMNB2</b>	<i>Lamin B2</i>

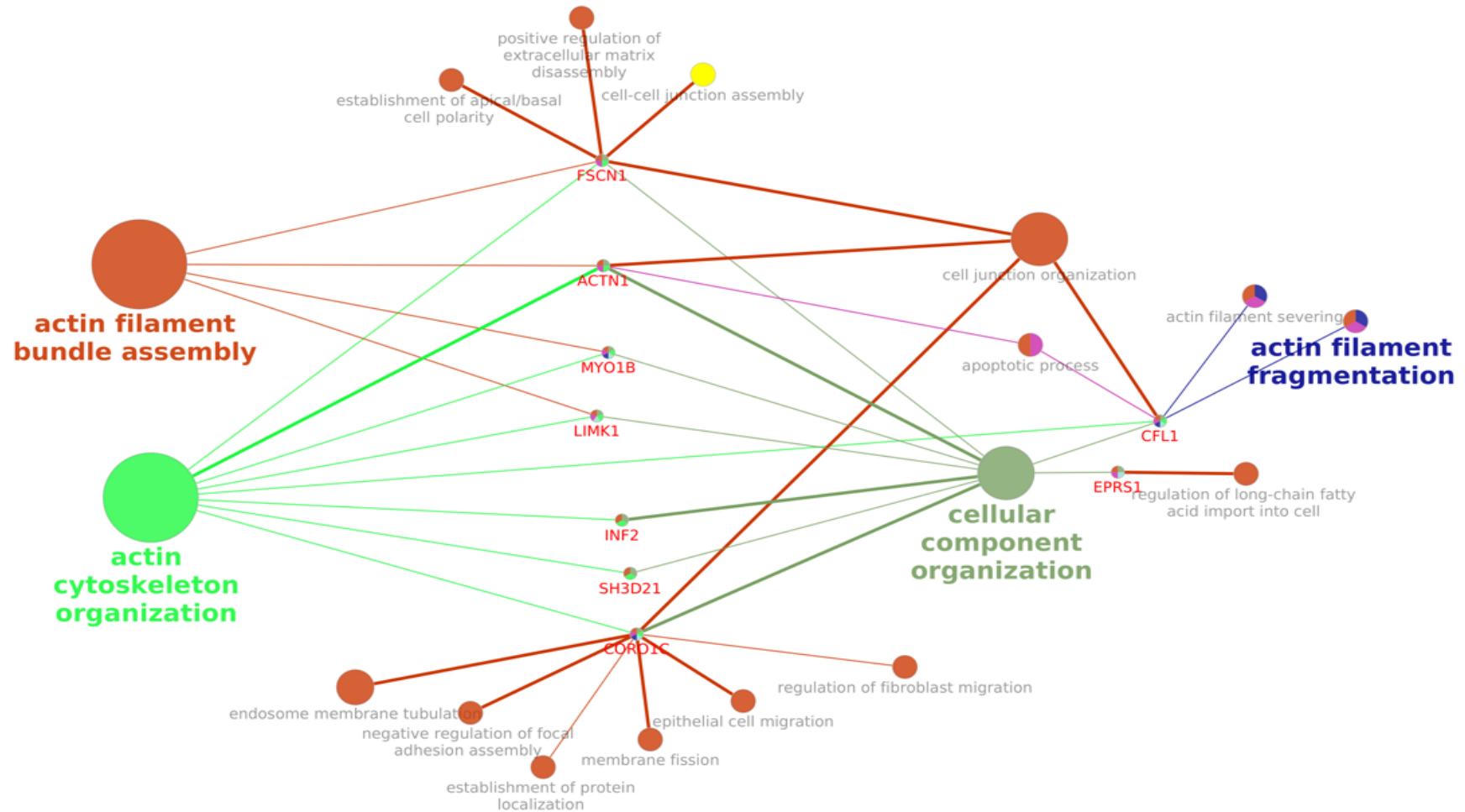


<b>LYAR</b>	<i>Ly1 Antibody Reactive</i>
<b>MDFI</b>	<i>MyoD Family Inhibitor</i>
<b>MET</b>	<i>MET Proto-Oncogene, Receptor Tyrosine Kinase</i>
<b>MICAL2</b>	<i>Microtubule Associated Monooxygenase, Calponin And LIM Domain Containing 2</i>
<b>MSN</b>	<i>Moesin</i>
<b>MYBL2</b>	<i>MYB Proto-Oncogene Like 2</i>
<b>MYO1B</b>	<i>Myosin IB</i>
<b>NACC1</b>	<i>Nucleus Accumbens Associated 1</i>
<b>NKRF</b>	<i>NFKB Repressing Factor</i>
<b>NOC2L</b>	<i>NOC2 Like Nucleolar Associated Transcriptional Repressor</i>
<b>NOMO1</b>	<i>NODAL Modulator 1</i>
<b>NRIP1</b>	<i>Nuclear Receptor Interacting Protein 1</i>
<b>PA2G4</b>	<i>Proliferation-Associated 2G4</i>
<b>PLEC</b>	<i>Plectin</i>
<b>PPP2R5B</b>	<i>Protein Phosphatase 2 Regulatory Subunit B'Beta</i>
<b>PSMC3</b>	<i>Proteasome 26S Subunit, ATPase 3</i>
<b>PTPRZ1</b>	<i>Protein Tyrosine Phosphatase Receptor Type Z1</i>
<b>SDK2</b>	<i>Sidekick Cell Adhesion Molecule 2</i>
<b>SEMA3C</b>	<i>Semaphorin 3C</i>
<b>SH3D21</b>	<i>SH3 Domain Containing 21</i>
<b>SLC15A1</b>	<i>Solute Carrier Family 15 Member 1</i>
<b>TGM2</b>	<i>Transglutaminase 2</i>
<b>TNC</b>	<i>Tenascin C</i>
<b>TUBB3</b>	<i>Tubulin Beta 3 Class III</i>
<b>UHRF1</b>	<i>Ubiquitin Like With PHD And Ring Finger Domains 1</i>
<b>WDR5</b>	<i>WD Repeat Domain 5</i>

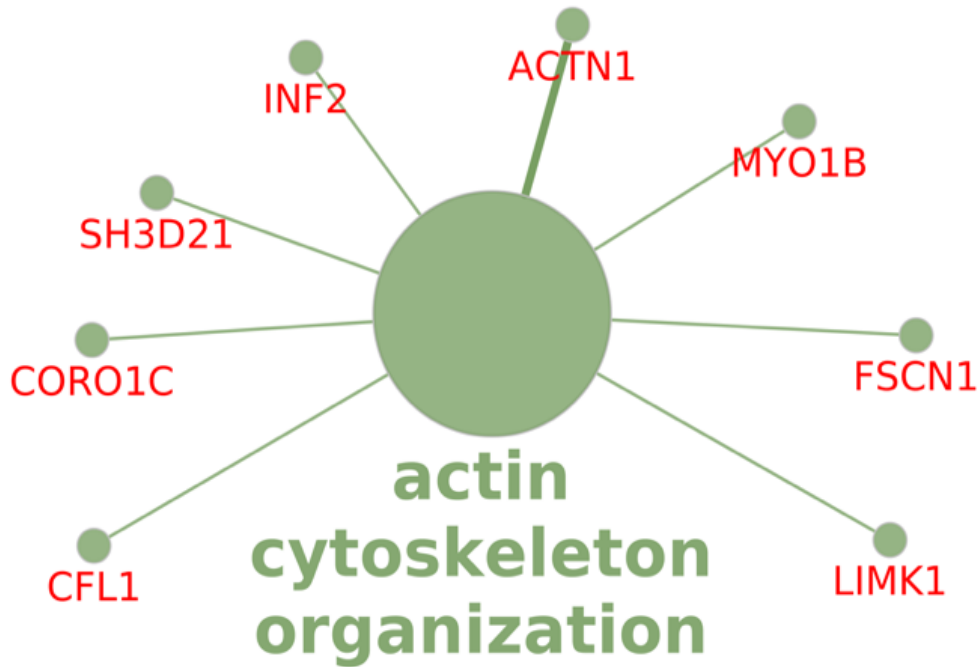
De los 49 genes encontrados destacamos los genes: **INF2**, **ACTN1**, **CFL1**, **LIMK1**, **FSCN1**, **MYO1B**, **SH3D21**, **CORO1C** siendo involucrados directamente en la formación del tejido epitelial (Figura 2, 3 y 4).



**Figura 2.** Imagen de la correlación positiva entre de los 8 genes destacados con procesos que pueden estar involucrados en la formación del tejido epitelial.



**Figura 3.** Construcción de redes de interacción entre los genes ACNT1 INF2, ACTN1, CFL1, LIMK1, FSCN1, MYO1B, SH3D21, CORO1C y sus procesos biológicos mediante el software cytoscape.



**Figura 4.** Genes comunes sobrespresados entre cáncer oral, leucoplasia y fibrosis involucrados con la organización del citoesquelto de actina.



**Universidad**  
**Europea** VALENCIA

## 9. DISCUSSION

Las patologías pre-cancerígenas y cancerígenas de los tejidos orales son enfermedades complejas y multifactoriales que pueden ser causadas por una combinación de factores genéticos y ambientales.

Diversos genes están implicados en el desarrollo del cáncer oral, leucoplasia y fibrosis oral pero aún poco se sabe sobre la implicación de estos genes en el tejido epitelial siendo un tipo de tejido que recupera las superficies externas e internas del cuerpo y cumple funciones importantes en la protección, absorción y secreción en diversos órganos y sistemas.

Los genes más relevantes en nuestro estudio fueron ACTIN1, LIMK1, CORO1C, INF2, CFL1, FSCN1, MYO1B SH3D21 sobreexpresado en las muestras de tejido epiteliales de pacientes que presentan Cáncer oral, leucoplasia y fibrosis los cuales están involucrados con procesos biológicos relacionados con la regulación y función del tejido epitelial.

El tejido epitelial es de gran importancia a la hora que hablamos de cáncer; está constituido de múltiples células y cada una de ellas contiene material genético. En general, los cánceres son consecuencia de desregulaciones que pueden ser debidas a errores en el sistema de regulación que controla la expresión de los genes produciendo alguna patogenicidad de algunos cánceres.

El tejido epitelial está estrechamente relacionado con el cáncer, por esto es importante saber la importancia que juega. El tejido epitelial está compuesto de una multitud de células las cuales cada una está constituida por un ADN; el cáncer se forma empezado de una mutación de la división celular que pueden ser debidas tanto por errores en la división de las células o por exposición a sustancias carcinógenas, las cuales dañan el ADN de las células que forman (66).

De los 160 genes totales sobreexpresados el EPRS1 es el único que encontramos común en las tres patologías.

El gen EPRS1 (Glutamyl-prolyl-tRNA synthetase 1), es un gen codificador de proteínas, considerado el más relevante por encontrarse sobreexpresado en muestras asociadas al cáncer oral, fibrosis oral y leucoplasia oral.

En el estudio de Krishnendu Khan y cols. el gen EPRS1 presenta una mayor resistencia al tamoxifeno, que son moduladores selectivos de receptores de estrógeno utilizado como medicamentos contra el cáncer, por eso la presencia de este gen sobreexpresado es considerado mal pronóstico en las salas de emergencia (67).

Se ha observado que el EPRS1 estaba sobreexpresado en el tejido tumoral de mama en comparación con el tejido normal y se asoció con una mayor destrucción celular y una peor supervivencia de los pacientes (68) también en cáncer de pulmón promueve la extinción y la invasión de las células cancerígenas, y que su inhibición reduzca significativamente el crecimiento tumoral y la capacidad metastásica (69).

Tannous et al. 2015 investigaron los perfiles de expresión génica en muestras de leucoplasia oral y cáncer oral donde encontraron el gen EPRS más elevado (70). Además, este gen podría tener relación con el desarrollo y mantenimiento del tejido epitelial. Se ha observado en células epiteliales de la piel y cabello la pérdida de EPRS1 condujo a alteraciones en el desarrollo y función de la piel (71). Este gen también se vio relacionado en la integridad de la barrera intestinal al regular la expresión de genes relacionados con la función de barrera y la respuesta inflamatoria (72).

Sin embargo, se necesitan más estudios para comprender completamente los mecanismos y la implicación precisa de EPRS1 en el tejido epitelial.

Otro gen relevante es el ACTN1, denominada también “actinas no musculares”, son las principales proteínas del citoesqueleto y realizan importantes funciones como la regulación de la citocinesis, la adhesión celular, la migración y median la función del sarcómero (73).

En nuestro estudio se mostró que los niveles de proteína Actina 1 del tejido enfermo estaban mayores respecto a lo del tejido sano; lo que indica que la proteína ACTN1 están mayormente expresada en paciente que padece de cáncer oral; De acuerdo con el estudio de Guo-feng Xie y cols. los cuales consideran el ACTIN1 una de los principales marcadores de los tejidos que presentaban el cáncer carcinoma oral de la célula escamosa (OSCC),

además, se vieron los niveles aumentados estaban asociado a metástasis de los ganglios con un mal pronóstico (74).

De manera similar el ACTIN 1 está relacionado también a otras tipologías de cáncer, como por ejemplo el cáncer de mama o el carcinoma hepatocelular (75,76).

También está respaldado por parte de Xu y cols. Los cuales con su estudio encontraron también los niveles de ACTIN1 en el tejido con OSCC que en el tejido normal.

Otro gen sobreexpresado que se ha detectado relevante en nuestro estudio es el LIM Domain Kinase 1 (LMNK1), es una serina quinasa que cataliza la fosforilación del grupo hidróxido de la serina, la cuales regulan la polimerización de la actina; Este gen tiene la función de regular los movimientos celulares y el ciclo celular (77).

De acuerdo con nuestro estudio también en el estudio de Bingxia Shi y cols. se ha detectado sobreexpresado el gen en el cáncer oral, sobre todo en el cáncer OSCC, el cual se afirma que en la tumorigenesis y metástasis del cáncer está afectada cuando LIMK1, además en este estudio se ha detectado también una mayor activación del gen LIM Domain Kinase 2 (LIMK2). En ambos genes el papel se desenvuelve (78).

El tercer gen involucrado en nuestro estudio es Coronin 1C (COROC1), es un gen que ejecuta como una plataforma estable para la interacción con otras proteínas.

En nuestro estudio si encontramos una sobreexpresión del este gen en pacientes que presentan lesiones orales cancerosas, conforme con el estudio de Xiaonfan Dai y cols. los cuales afirman que, debido a su función de motilidad celular, el gen CORO1C juega un papel fundamental, debido a que la emigración es importante para la metástasis de tumores malignos demostrando la función de gen en el cáncer (79).

No obstante, el estudio se analizó sobre todos tejido de cáncer de próstata podemos también en el tejido del cáncer oral se ha encontrado el gen sobreexpresado. El gen INF2 (Inverted Formin 2) es el cuarto gen sobreexpresado que encontramos en nuestro estudio; Es INF2 es una



formina que tiene la capacidad de acelerar tanto la polimerización como la despolimerización de actina y además tiene la capacidad de unir también los filamentos de actinas en fibras de tensión (80).

De acuerdo con nuestro estudio, Vanina D. Heuser et cols. Encontró actividad del INF2 sobreelevada en el cáncer de mama de tipo basal, forma de cáncer más agresiva, además en este estudio se notó que la eliminación de INF2 iba a reducir la proliferación de las células tumorales (81).

El gen CFL1 hace de función como receptor juega un el importante en la proliferación y migración celular en el cuales encontramos funciones en la aparición, desarrollo, infiltración y metástasis de lesiones tumorales.

En nuestro estudio encontramos niveles del gen sobreexpresado en el tejido epitelial de la cavidad oral en pacientes que presentan lesiones malignas.

Dritan Turhani y cols. Encontraron en sus estudios CFL1 altamente expresado en el cáncer carcinoma oral de la célula escamosa (82) además, Min- li y cols. encontraron relaciones en los cánceres del páncreas, mama, vesicular biliar y demostraron que la inhibición de la actividad de CFL1 en un tumor maligno puede ser un objetivo potencial para inhibir la progresión tumoral y la metástasis (83).

El gen FSCN1 está centrada en el papel de micropuntas, filopodios, seudópodos lamelares y otras protuberancias basadas en actina debajo de la membrana plasmática, que son importantes para procesos que incluyen la migración celular y la adhesión a la matriz celular, Zhongxun Li y cols ha indicado una expresión altamente significativa en los tejido de cáncer de la células escamosas de la laringe, además se descubrió que el FSCN1 unido a otra proteína estaba involucrado en la progresión maligna de dicho cáncer; Además de estar involucrado en cáncer de laringe también lo encontramos sobreexpresado en cáncer respiratorios, del aparato digestivo, de sistema urinario y del sistema Genital (84).

Lii C y cols. considera la inhibición del FSCN1 con fármacos contra el cáncer tiene un potencial significativo para el tratamiento y para aplicaciones clínicas (85).

MYO1B es un gen considerado importante para la migración de la células y la motilidad celulares; en nuestro estudio encontramos nivel alto del gen de acuerdo con el estudio de Lang Xie y cols. Los cuales después de un estudio profundizó llego a la conclusión que el MYO1B es considerado un biomarcador fundamental para la metástasis del cáncer colonrrectal y de mal pronóstico, pero además de esto lo considera como un futuro objetivo terapéutico (86).

Por último, tenemos el gen S3D21 con una elevada actividad en nuestro estudio.

Masoundi M y cols. considera que, aunque no sea un gen muy estudiado y que no se ha encontrado mucho informe para el estudio, el S3D21 puede ser relacionados con la activación y sensibilización de las células del Cáncer, denominándose, así, gen de supervivencia para el cáncer de los pulmones (87).

Estos genes sobreexpresado pueden ser considerado importante porque podrían estar desrgularizando la formación y la integridad del tejido epitelial tnato oral como por otros tipos de cánceres, promovendo la producción de células cancerígena, sin embargo, destacamos la poca información existente en la bibliografía sobre el funcionamiento estos genes en el tejido epitelial oral siendo sumamente importante en la evolución de patologías.

Es necesario estudios adicionales para comprender completamente los mecanismos biológicos subyacentes para el cáncer oral, leucoplasia oral y la fibrosis oral.



**Universidad**  
**Europea** VALENCIA

## 10. CONCLUSION

- Se detectaron 800 genes diferenciales infraexpresados y 608 sobreexpresados (C vs N), 102 genes diferenciales infraexpresado y 8 sobreexpresado (F vs N) y 2567 genes diferenciales infraexpresado y 2603 sobreexpresado (L vs N).

- Un total de 160 genes comunes fueron sobreexpresados en paciente que presenta dos de las tres patologías; 154 genes fueron comunes sobreexpresados (en C y L), un gen (en L y F), 4 genes (en C y F.) Un total de 148 fueron genes comunes infraexpresado dos de las tres patologías; 104 genes (en C y L) y 1 gene (en F y L) y 43 genes (en C y F).

-Encontramos procesos biológicos significativos involucrados 49 genes de los 160 genes comunes sobreexpresados presentes en dos de las tres patologías (F, C y L) destacando los genes ACTIN1, LIMK1, CORO1C, INF2, CFL1, FSCN1, MYO1B y SH3D21 teniendo relación con procesos biológicos que podrían estar involucrados en el desequilibrio de la integridad del tejido epitelial de la cavidad oral.

-El único gen encontrado sobreexpresado común en las tres patologías fue EPRS1, que podría en un futuro considerarse como un biomarcador temprano de Cáncer Oral.



**Universidad**  
**Europea** VALENCIA

## 11. BIBLIOGRAFIA

1. Orfao A, Benitez J, Corrales F, Martin- Subero I, Maria Ordovas J. *Ciencia Omicas.C.O*;2019;15:1-32.
2. Barh Debmalya, Blum Kenneth, Madigan MA. *Omics: biomedical perspectives and applications*. 2012;591.
3. Developing pathways for community-led research with big data: a content analysis of stakeholder interviews.*Health and Medicine*.2020;76.
4. Serna M, Directora G, María D, Santacreu A, Directora J, Rosa D, et al. *Escuela tecnica superior de ingenieria agronomica y medio natural Grado en Ingeniería Agroalimentaria y del Medio Rural estudio del tejido ovarico para mejorar la produccion a traves del tamaño de camada en conejos*. 2016.
5. Bruce A, Johnson A, Lewis J, Raff M, Keith R, Walter P. *The Structure and Function of DNA*.NCBI.2002;4(3).
6. Roth SC. What is genomic medicine? *Journal of the Medical Library Association*.2019;107(3):442–8.
7. Kukurba KR, Montgomery SB. *RNA sequencing and analysis*. Cold Spring Harb Protoc.2015;(11):951–69.
8. Nookaew I, Papini M, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlén M, et al. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res*.2012;40(20):10084–97.
9. Hasin Y, Seldin M, Lusic A. Multi-omics approaches to disease. *Genome Biology*.2017;18(1):1–15.
10. Frye M, Jaffrey SR, Pan T, Rechavi G, Suzuki T. RNA modifications: What have we learned and where are we headed? *Nature Reviews Genetics*. Nature Publishing Group; 2016;17:365–72.

11. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, et al. Genome-Wide Survey of Human Alternative Pre-mRNA Splicing with Exon Junction Microarrays. *2003;302(5653):2141-4.*
12. Cieřlik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. *Nature Reviews Genetics. Nature Publishing Group; 2018;19:93–109.*
13. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.2005;437(7057):376–80.*
14. Hong M, Tao S, Zhang L, Diao LT, Huang X, Huang S, et al. RNA sequencing: new technologies and applications in cancer research. *Journal of Hematology and Oncology. BioMed Central Ltd; 2020;13.*
15. Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nat Biotechnol.2015;33(3):306–12.*
16. Govindarajan M, Wohlmuth C, Waas M, Bernardini MQ, Kislinger T. High-throughput approaches for precision medicine in high-grade serous ovarian cancer. *J Hematol Oncol-2020;13.*
17. Giménez Llorente Tutora D, Rodríguez Luján Ponente I, María González Marcos A. Analisis de expression diferencial para datos de next generation sequencing (NGS) con multiples condiciones experimentales. *2017.*
18. Zubovic L, Piazza S, Tebaldi T, Cozzuto L, Palazzo G, Sidarovich V, et al. The altered transcriptome of pediatric myelodysplastic syndrome revealed by RNA sequencing. *J Hematol Oncol. 2020;13(1).*
19. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods. 2009;6(5):377–82.*
20. Wu J, Ding Y, Wang J, Lyu F, Tang Q, Song J, et al. Single-cell RNA sequencing in oral science: Current awareness and perspectives. *Cell Prolif.2022;55(10):13287.*

21. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008 Jul;5(7):621–8.
22. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biology*.2016;17(1):1–19.
23. Soverini S, Abruzzese E, Bocchia M, Bonifacio M, Galimberti S, Gozzini A, et al. Next-generation sequencing for BCR-ABL1 kinase domain mutation testing in patients with chronic myeloid leukemia: A position paper. *J Hematol Oncol*. 2019;12(1).
24. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*.2010;38(6):1767–71.
25. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and analysis of RNA-seq data. *Nat Methods*. 2010;7(11):909–12.
26. Hurgobin B. Short Read Alignment Using SOAP2. *Methods Mol Biol*.2016;1374:241–52.
27. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
28. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5.
29. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
30. Groeger S, Meyle J. Oral Mucosal Epithelial Cells. *Front Immunol*. 2019;10:208.
31. L.P Gartner.Oral anatomy and tissue types.*Semin Dermatol*.1994;13(2):68-73.



32. Van Rensburg BG Jansen. Oral biology. 995;524.
33. Díaz-Peña R, Castro-Santos P. Histological Changes Implicated in Metastasis. *International Journal of Morphology*.2014;32(3):935–41.
34. Eccles K, Carey B, Cook R, Escudier M, Diniz-Freitas M, Limeres-Posse J, et al. Artículo educativo Trastornos orales potencialmente malignos: recomendaciones sobre el abordaje en la atención primaria. *J Oral Med Oral Surg*.2022;38:36.
35. Carter LM, Ogden GR. Oral cancer awareness of undergraduate medical and dental students. *BMC Med Educ*.2007;7(1):1–8.
36. Donado M, Martinez JM. Donado cirugía bucal Patología y técnica.4th.Barcelona:Elsevier;2014.
37. Zhou S, Zhu Y, Li Z, Zhu Y, He Z, Zhang C. Exosome-derived long non-coding RNA ADAMTS9-AS2 suppresses progression of oral submucous fibrosis via AKT signalling pathway. *J Cell Mol Med*. 2021;25(4):2262–73.
38. Iparraguirre MF, Fajardo X, Carneiro E, Couto PH, Iparraguirre MF, Fajardo X, et al. Desordenes orales potencialmente malignos-Lo que el odontólogo debe conocer. *Revista Estomatológica Herediana*.2020;30(3):216–23.
39. Mithani SK, Mydlarz WK, Grumbine FL, Smith IM, Califano JA. Molecular genetics of premalignant oral lesions. *Oral Dis*. 2007;13(2):126–33.
40. Warnakulasuriya S. Clinical features and presentation of oral potentially malignant disorders.*Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*. Mosby.2018;125:582–90.
41. Kumar V, Abbas AK, Aster JC. Patología estructural y funcional.10th.Barcelona:Elsevier;2021.
42. Carrard VC, van der Waal I. A clinical diagnosis of oral leukoplakia; A guide for dentists. *Med Oral Patol Oral Cir Bucal*. 2018;23(1):59–64.
43. Lodi G, Franchini R, Warnakulasuriya S, Varoni EM, Sardella A, Kerr AR, et al. Interventions for treating oral leukoplakia to prevent oral cancer. *Cochrane Database Syst Rev*. 2016;2016(7):92–3.
44. McCullough MJ, Prasad G, Farah CS. Oral mucosal malignancy and potentially malignant lesions: an update on the epidemiology, risk factors, diagnosis and management. *Australian dental journal*. 2010;55(1):61–5.

45. Antón MSM, Pérez SM. Revisión de la literatura. *Av Odontoestomatol.* 2015;31(4):247–59.
46. Izaguirre Perez PM, Benevente Lipa LA. Nivel de conocimiento sobre el cancer oral de estudiante del ultimo ano de carrera profesional de odontologia de la universidades de Lima. 2012.
47. Abati S, Bramati C, Bondi S, Lissoni A, Trimarchi M. Oral cancer and precancer: A narrative review on the relevance of early diagnosis. *MDPI AG* 2020;17(24):9160.
48. Silverman S, Kerr AR, Epstein JB. Oral and Pharyngeal Cancer Control and Early Detection. *Journal of Cancer Education.* 2010;25(3):279.
49. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncology.*2009;45:309–16.
50. Gómez I, Seoane J, Varela-Centelles P, Diz P, Takkouche B. Is diagnostic delay related to advanced-stage oral cancer? A meta-analysis. 2009;117(5):541-6.
51. Dhanuthai K, Rojanawatsirivej S, Thosaporn W, Kintarak S, Subarnbhesaj A, Darling M, et al. Oral cancer: A multicenter study. *Med Oral Patol Oral Cir Bucal.* 2018;23(1):23–9.
52. Andrés U, Chile B. Epidemiological characterization of oral cancer. Literature review.2023;12(1).
53. Smith CC, Bixby LM, Miller KL, Selitsky SR, Bortone DS, Hoadley KA, et al. Using RNA Sequencing to Characterize the Tumor Microenvironment. *Methods Mol Biol.*2020;2055:245–72.
54. Wu J, Ding Y, Wang J, Lyu F, Tang Q, Song J, et al. Single-cell RNA sequencing in oral science: Current awareness and perspectives. *Cell Prolif.* 2022;55(10):13287.
55. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the Impact of Smoking and Lung Cancer on the Airway Transcriptome Using RNA-Seq. *Cancer Prevention Research.* 2011;4(6):803–17.

56. Dietlein F, Wang AB, Fagre C, Tang A, Besselink NJM, Cuppen E, et al. Genome-wide analysis of somatic noncoding mutation patterns in cancer. *Science*. 2022;376(6589).
57. Wang SW, Gao C, Zheng YM, Yi L, Lu JC, Huang XY, et al. Current applications and future perspective of CRISPR/Cas9 gene editing in cancer. *Mol Cancer*. 2022;21(1).
58. Timar J, Kashofer K. Molecular epidemiology and diagnostics of KRAS mutations in human cancer. *Cancer Metastasis Rev*. 2020;39(4):1029–38.
59. Yao L, Guo B, Wang J, Wu J. Analysis of transcriptome expression profiling data in oral leukoplakia and early and late-stage oral squamous cell carcinoma. *Oncol Lett*. 2023;25(4).
60. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res*. 2018;7:1338.
61. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
62. Eduardo A, Cubillos R, Laura B, Jiménez P, Adriana J, Bernal G. analizando datos de Rna-seq en procariotas: una revision para no experto: A Review for Non-experts. *Acta Biolo Colomb*. 2014;19(2):131–42.
63. Smyth GK, Ritchie ME, Law CW, Alhamdoosh M, Su S, Dong X, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res*. 2018;5.
64. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
65. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
66. Marzo-Castillejo M, Vela-Vallespín C, Bellas-Beceiro B, Bartolomé-Moreno C, Ginés-Díaz Y, Melús-Palazón E. Grupos de Expertos de Cáncer del PAPPS. Recomendaciones de Prevención del Cáncer. Actualización PAPPS 2020. *Aten Primaria*. 2020;52(2):44.

67. Khan K, Gogonea V, Fox PL. Aminoacyl-tRNA synthetases of the multi-tRNA synthetase complex and their role in tumorigenesis. Vol. 19, Translational Oncology. Neoplasia Press, Inc.; 2022.
68. Berral AV, Murua CM, de la Vega FA, García IH, Rosino AA, de Liaño ÁD, et al. Situación actual en el tratamiento del cáncer gástrico. *Revista Española de Enfermedades Digestivas*. 2012;104(3):134–41.
69. Castañeda CA, Casavilca S, Orrego E, García-Corrochano P, Deza P, Heinike H, et al. Glioblastoma: Análisis molecular y sus implicancias clínicas. *Rev Peru Med Exp Salud Publica*. 2015;32(2):316–25.
70. Martínez-Sahuquillo Marquez A, Gallardo Castillo I, et al. La leucoplasia oral: Su implicación como lesión precancerosa. *Av Odontoestomatol*. 2008;24(1).
71. Fuchs-Telem D, Stewart H, Rapaport D, Nousbeck J, Gat A, Gini M, et al. CEDNIK syndrome results from loss-of-function mutations in SNAP29. *Br J Dermatol*. 2011;164(3):610–6.
72. Kim S, You S, Hwang D. Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nature Reviews Cancer*. 2011;11(10):708–18.
73. Foley KS, Young PW. The non-muscle functions of actinins: An update. *Biochemical Journal*. Portland Press Ltd. 2014;459:1–13.
74. Xie GF, Zhao LD, Chen Q, Tang DX, Chen QY, Lu HF, et al. High ACTN1 Is Associated with Poor Prognosis, and ACTN1 Silencing Suppresses Cell Proliferation and Metastasis in Oral Squamous Cell Carcinoma. *Drug Des Devel Ther*. 2020;14:1717–27.
75. Chen Q, Zhou XW, Zhang AJ, He K. ACTN1 supports tumor growth by inhibiting Hippo signaling in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2021;40(1).
76. Zhang S, Wang J, Chen T, Wang J, Wang Y, Yu Z, et al.  $\alpha$ -Actinin1 promotes tumorigenesis and epithelial-mesenchymal transition of gastric cancer via the AKT/GSK3 $\beta$ / $\beta$ -Catenin pathway. *Bioengineered*. 2021;12(1):5688–704.

77. Su J, Zhou Y, Pan Z, Shi L, Yang J, Liao A, et al. Downregulation of LIMK1–ADF/cofilin by DADS inhibits the migration and invasion of colon cancer. *Scientific Reports* 2017 7:1. 2017;7(1):1–12.
78. Xu J, Lv H, Zhang B, Xu F, Zhu H, Chen B, et al. miR-30b-5p acts as a tumor suppressor microRNA in esophageal squamous cell carcinoma. *J Thorac Dis.* 2019;11(7):3015–29.
79. Dai X, Liang Z, Liu L, Guo K, Xu S, Wang H. Silencing of MALAT1 inhibits migration and invasion by sponging miR-1-3p in prostate cancer cells. *Mol Med Rep.*2019;20(4):3499–508.
80. Ramabhadran V, Korobova F, Rahme GJ, Higgs HN. Splice variant–specific cellular function of the formin INF2 in maintenance of Golgi architecture. *Mol Biol Cell.* 2011;22(24):4822.
81. Heuser VD, Mansuri N, Mogg J, Kurki S, Repo H, Kronqvist P, et al. Formin Proteins FHOD1 and INF2 in Triple-Negative Breast Cancer: Association With Basal Markers and Functional Activities. *Breast Cancer (Auckl).* 2018];12:117.
82. Turhani D, Krapfenbauer K, Thurnher D, Langen H, Fountoulakis M. Identification of differentially expressed, tumor-associated proteins in oral squamous cell carcinoma by proteomic analysis. *Electrophoresis.* 2006;27(7):1417–23.
83. Wang L, Xiong L, Wu Z, Miao X, Liu Z, Li D, et al. Expression of UGP2 and CFL1 expression levels in benign and malignant pancreatic lesions and their clinicopathological significance. *World J Surg Oncol.* 2018];16(1):11.
84. Li Z, Shi J, Zhang N, Zheng X, Jin Y, Wen S, et al. FSCN1 acts as a promising therapeutic target in the blockade of tumor cell motility: a review of its function, mechanism, and clinical significance. *J Cancer [Internet].* 2022;13(8):2528.
85. Lii CK, Chang JW, Chen JJ, Chen HW, Liu KL, Yeh SL, et al. Docosahexaenoic acid inhibits 12-O-tetradecanoylphorbol-13- acetate-induced fascin-1-dependent breast cancer cell migration by suppressing the PKC $\delta$ - and Wnt-1/ $\beta$ -catenin-mediated pathways. *Oncotarget.* 2016;7(18):25162–79.

86. Xie L, Huang H, Zheng Z, Yang Q, Wang S, Chen Y, et al. MYO1B enhances colorectal cancer metastasis by promoting the F-actin rearrangement and focal adhesion assembly via RhoA/ROCK/FAK signaling. *Ann Transl Med.* 2021;9(20):1543–1543.
87. Masoudi M, Seki M, Yazdanparast R, Yachie N, Aburatani H. A genome-scale CRISPR/Cas9 knockout screening reveals SH3D21 as a sensitizer for gemcitabine. *Sci Rep.* 2019;9(1):19188.



**Universidad**  
**Europea** VALENCIA

**META-ANALYSIS OF GENE EXPRESSION IN PATIENTS WITH CANCER,  
FIBROSIS AND LEUKOPLASIA IN THE ORAL CAVITY USING PUBLIC  
RNA-SEQ DATA**

**Short title: Meta-analysis of gene expression in patients with oral cancer  
and precancerous diseases with public RNA-seq data**

**Authors:**

**Agnese Formaggio <sup>1</sup>, Marta Serna-Garcia <sup>2</sup>**

*<sup>1</sup> 5th year student of the Dentistry degree at European University of Valencia,  
Valencia, Spain.*

*<sup>2</sup> Department of Dentistry, Faculty of Health Sciences, European University of  
Valencia, Valencia, Spain.*

**Correspondence**

Marta Serna Garcia  
Paseo de Alameda 7, Valencia  
46010, Valencia  
marta.serna@universidadeuropes.es



## **Abstract**

**Introduction:** RNA-seq is a tool that discovers new biomarkers of oral cancer and precancerous lesions such as leukoplakia and oral fibrosis, in addition, it studies relevant genes and knows the biological structure, thus helping to develop therapies. The goal was to carry out a transcriptomic study and identify the biological processes in the oral cavity using RNA-seq for samples obtained from public databases of divergent groups; samples from patients with precancerous lesions: Leukoplakia (L) and oral fibrosis (F) and oral cancer (C) compared with patients without pathology (N).

**Materials and methods:** Three studies were selected from the public database *GeneExpression for All* (*GSE125866*<sup>(2019)</sup>, *GSE20116*<sup>(2010)</sup> y *GSE131568*<sup>(2020)</sup>). All raw files were analyzed with the same workflow. The quality of readings was verified using *Fastqcsoftware* and filtered with *Trimmomaticsoftware*. Aligned with genome *A wise man* with *Callistosoftware*. *EdgeR software* y *Gluesoftware* normalized the reads and identified differentially expressed genes ( $FDR \leq 0.05$ ). *DAVID* was used to determine significant biological processes *software* ( $p \text{ value} \leq 0,05$ ). *Cytoscape software* was used for biological terms.

**Results:** 800 underexpressed and 608 overexpressed (C vsN) genes, 102 underexpressed and 8 overexpressed (FvsN) genes, and 2567 underexpressed and 2603 overexpressed (LvsN) genes were detected. 154 overexpressed common genes (C and L), one gene (L and F), 4 genes (C and F) 148 underexpressed common genes (C and L) and 104 genes (F and L) and 1 gene (C and F) . *EPRS* was the only common gene in the three pathologies ( $Foldchange 0.653$ ). Of the 49 genes, we highlight 8 involved with epithelial tissue.

**Conclusions:** The *ACTIN1*, *LIMK1*, *CORO1C*, *INF2*, *CFL1*, *FSCN1*, *MYO1B* and *SH3D21* genes could influence the integrity of epithelial tissue and help advance precancerous lesions and cancer. *EPRS* is a potential early biomarker.

**Keywords:** *RNA-seq, oral cancer, fibrosis, leukoplakia, epithelial tissue, transcriptomic analysis*

## Introduction

Cancer is a global health problem and is one of the leading causes of mortality (1). One of the most frequent oral cancers is that of the head and neck; It is among the 10 most common malignant neoplasms and there has been an increase in cases diagnosed in the last 10 years (2).

Studies of oral pre-cancerous and cancerous diseases are of paramount importance due to an increase in the number of cases in recent years.(3) Therefore, its knowledge, prevention and diagnosis are being emphasized more and more for dentists, doctors and scientists; Thanks to RNA seq technology, there are numerous studies for these pathologies (4–7) but more genes that could be responsible for these pathologies still need to be elucidated, as well as a better understanding of their biological functions, which may help gene therapies in the future. In this way, identifying genes that are being activated common to different precancerous lesions and also with oral cancer compared to people without pathologies, could be the key to help early detection of these pathologies.

This work will allow us to analyze the expression of genes and identify the biological processes of interest in the human oral cavity using the RNA-seq technique for different samples (from patients with pre-cancerous lesions compared to samples from patients without pathologies). HEdetect differential genes (over and under expressed) common in patients with oral cancer *versus* healthy tissue, fibrosis *versus* healthy tissue and Leukoplakia *versus* healthy tissue, an enrichment analysis will be carried out to identify significant biological processes for oral cancer, fibrosis and Leukoplakia and the most relevant common differential genes associated with cancer and precancerous lesions will be identified so that in the future they could be used as biomarkers.

## Materials and methods

### - Study design

Different studies of raw RNA sequencing data from patient samples were selected using the RNA-seq technique, in the public databases of Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), public functional genomics data repository, sequence and array-based data, and BioStudies (<https://www.ebi.ac.uk/biostudies/arrayexpress>). Three studies with the accession number to the database of GSE125866, GSE20116 and GSE131568 were selected. The study *GSE125866* we have n=40 tissue samples from the oral cavity, of which n=10 are from patients with fibrosis, n=20 samples from patients with cancer and n=10 from patients who do not show pathology and we consider it control samples (normal). The study *GSE20116* we have n=20 tissue samples from the oral cavity, of which n=6 samples from patients presenting with cancer and n=14 from patients that do not show pathology and we consider it control samples (normal). The study *GSE131568* have n=26 tissue samples from the oral cavity, of which n=6 samples from patients presenting Leukoplakia and n=20 from patients who do not show pathology and that considered control samples (normal).

We performed 4 differential analyzes with the raw data of the patient samples from the 3 selected studies. For the first analysis we selected the samples from study *GSE125866*, 10 fibrosis samples were compared with 10 control samples. In the second analysis we selected the samples from the *GSE125866* study, 10 cancer samples were compared with 10 control samples. In the third analysis we selected the samples from the *GSE20116* study, 6 cancer samples were compared with 14 control samples. In the fourth analysis we selected the samples from the *GSE131568* study, 6 Leukoplakia samples were compared with 20 control samples.

- **Differential expression analysis**

All raw files corresponding to each study sample were analyzed with the same workflow.

The quality of the RNA-Seq reads (quality indices, GC content, N content, length distributions, duplication, overrepresented sequences, and K-mer content) was verified using Fastqc software (v.0.11.4) (8).

Low quality reads (adapter sequence and poly-N containing reads) from raw data using Trimmomatic v.0.36 (61) with the following parameters: `java -jar /usr/local/bin/trimmomatic-0.36.jar PE -phred33 <IN_R1.fastq.gz> <IN_R2.fastq.gz> <OUT_R1.fastq.gz> <OUT_R1_UN.fastq.gz> <OUT_R2.fastq.gz> <OUT_R2_UN.fastq.gz> LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50`.

All subsequent analysis was based on data clipped with high quality readings, KALLISTO was used (9) Bray NL, Pimentel H., Melsted P., & Pachter L., 2016. Near-optimal RNA-seq probabilistic quantification. *Nat. Biotechnology*.34 : 525–527. 10.1038/nbt.3519 (OTHER) THIS IS THE CITE INCLUDE IT to align the paired-end trimmed reads with the Homo Sapiens reference genome deposited at the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>).

This read count was performed using the internal script and Ensembl (Index to `/pub/release-105/gtf/homo_sapiens/` ([ensembl.org](http://ensembl.org))) for gene annotations.

The EdgeR and Limma packages available from the Bioconductor project were used to estimate normalized CPM reads (counts per million) and identify differentially expressed genes (DEGs) between groups.

EdgeR was used to normalize the data and the Limma software for the statistical methods (linear modeling and empirical Bayes), following the Law et al. methodology (10).

We consider that a gene is differentially expressed when the value of FDR (false positive) is  $\leq 0.05$

#### - **Enrichment analysis**

Differential gene enrichment analysis was performed using the DAVID v.6.8 software (<https://david.ncifcrf.gov>) using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database ( $p$  value  $\leq 0.05$ ). ClueGO was used, a Cytoscape software plugin that visualizes non-redundant biological terms for large groups of genes in a functionally clustered network (11).

### **Results**

In the first study (cáncer vs non cáncer) have been found 800 underexpressed genes and 608 overexpressed, 102 underexpressed genes and 8 overexpressed in the second study (fibrosis vs non cáncer) and 2567 underexpressed genes and 2603 overexpressed samples in the third study (leukoplakia vs non cancer).

All the differential genes obtained were applied the FDR correction (*False Discovery Rate*) less than 0.05. A total of 160 common genes overexpressed in the different pathologies have been detected. A single common overexpressed gene has been identified among the three pathologies, 154 common genes between oral cancer and leukoplakia, one common gene between fibrosis and leukoplakia, and 4 common genes between oral cancer and fibrosis.

A total of 148 underexpressed common genes have been detected comparing the different pathologies. No common underexpressed genes have been detected among the three pathologies; We found 104 common genes between oral cancer and leukoplakia, between fibrosis and leukoplakia 1, and between oral cancer and fibrosis 43. The only common overexpressed gene found among patients with oral cancer, fibrosis, or leukoplakia, compared with healthy patients, was EPRS1. (glutamyl-prolyl-tRNA synthetase 1) with a fold change of 0.653 in the study between cancer (C) and healthy patient (N), a fold change of 0.492 in the study between fibrosis (F) and healthy patient (N) and a fold change of 0.091 in the last study between leukoplakia (L) and healthy patient (N). We highlight 49 genes involved in biological processes, in molecular function and in the molecular component of interest in the study ( $p$  value less than 0.05). Of the 49 genes found, we highlight the genes: *INF2*, *ACTN1*, *CFL1*, *LIMK1*, *FSCN1*, *MYO1B*, *SH3D21*, *CORO1C*) being directly involved in the formation of epithelial tissue.

## Discussion

The EPRS1 gene (Glutamyl-prolyl-tRNA synthetase 1) is considered the most relevant because it is overexpressed in samples associated with oral cancer, oral fibrosis and oral leukoplakia. This gene is a protein-coding gene. In Krishnendu Khan's study et al. the EPRS1 gene presents greater resistance to tamoxifen, which are selective modulators of estrogen receptors used as anticancer drugs, and in turn a poor prognosis in emergency rooms (12).

In another study, EPRS1 was observed to be overexpressed in breast tumor tissue compared to normal tissue and was associated with greater cell destruction and poorer patient survival (13). It also promotes extinction and invasion in lung cancer. cancer cells, and that its inhibition significantly reduces tumor growth and metastatic capacity (14). Tannous et al. in 2015 they investigated the gene expression profiles in samples of oral leukoplakia and oral cancer where they found the highest EPIRS gene (15). In a study (16) with epithelial cells of the skin and hair, it was seen that the loss of EPRS1 in epithelial cells led to alterations in the development and function of the skin.

This gene was also related to the integrity of the intestinal barrier by regulating the expression of genes related to barrier function and the inflammatory response (17). However, further studies are needed to fully understand the mechanisms and the precise involvement of EPRS1 in epithelial tissue.

Another relevant gene is Actin (ACTN) -1, also called "non-muscle actin", are the main proteins of the cytoskeleton and perform important functions such as regulation of cytokinesis, cell adhesion, migration and mediate sarcomere function. (18). In our study it was shown that the levels of Actin 1 protein in diseased tissue were higher than in healthy tissue; which indicates that the ACTN1 protein is more expressed in patients suffering from oral cancer; According to the study by Guo-feng Xie et al. who consider ACTIN 1 one of the main markers of tissues presenting oral squamous cell carcinoma (OSCC) cancer, in addition, increased levels were seen to be associated with lymph node metastasis with a poor prognosis (19). Similarly, ACTIN 1 is also related to other types of cancer, such as breast cancer or hepatocellular carcinoma (20); It is also

supported by part of Xu et al. Which with their study also found ACTIN1 levels in tissue with OSCC than in normal tissue. Another overexpressed gene that has been detected to be relevant in our study is LIM Domain Kinase 1 (LMNK1), it is a serine kinase that catalyzes the phosphorylation of the serine hydroxide group, which regulates actin polymerization; This gene has the function of regulating cell movements and the cell cycle (21).

In agreement with our study, also in the study by Bingxia Shi et al. Overexpression of the gene has been detected in oral cancer, especially in OSCC cancer, which is stated to be affected in cancer tumorigenesis and metastasis when LIMK1 is present. Furthermore, in this study a greater activation of the LIM Domain gene has also been detected. Kinase 2 (LIMK2): The role it plays in both genes is unknown (22).

The third overexpressed gene is Coronin 1C (COROC1), a gene that functions as a stable platform for interaction with other proteins, according to the study by Xiaonfan Dai et al. who affirm that, due to its cell motility function, the CORO1C gene plays a fundamental role for the metastasis of malignant tumors, demonstrating the gene function in cancer (23).

The INF2 gene (Inverted Formin 2) is the fourth overexpressed gene that we found in our study; Consistent with our study, Vanina D Heuser et al. He found elevated INF2 activity in basal breast cancer, a more aggressive form of cancer. Furthermore, in this study it was noted that the elimination of INF2 would reduce the proliferation of tumor cells (24).

The CFL1 gene acts as a receptor plays an important role in cell proliferation and migration in which we find functions in the appearance, development, infiltration and metastasis of tumor lesions. In our study we found overexpressed levels of the gene in the epithelial tissue of the oral cavity in patients with malignant lesions. Dritan Turhani et al. They found CFL1 highly expressed in oral squamous cell carcinoma cancer (25) besides, Min- Li et al. found relationships in cancers of the pancreas, breast, gallbladder and demonstrated that inhibition of CFL1 activity in a malignant tumor may be a potential target to inhibit tumor progression and metastasis (26).

The FSCN1 gene is focused on the role of microspikes, filopodia, lamellar pseudopods, and other actin-based protrusions below the plasma membrane, which are important for processes including cell migration and cell matrix adhesion, Zhongxun Li et al. highly significant in squamous cell cancer tissue of the Larynx, it was also discovered that FSCN1 bound to another protein was involved in the malignant progression of said cancer; In addition to being involved in laryngeal cancer, we also found it overexpressed in respiratory, digestive, urinary, and genital system cancers.(27). Lii C et al. considers that inhibition of FSCN1 with anticancer drugs has significant potential for treatment and clinical applications (28).

MYO1B is considered important for cell migration and cell motility; in our study we found high levels in accordance with Lang Xie et al. considers it a fundamental biomarker for cancer metastasis (29).

Lastly, we have the S3D21 gene with high activity in our study. Masoundi M et al. It is considered that, although it is not a highly studied gene and that not much report has been found for the study, S3D21 may be related to the activation and sensitization of cancer cells, being called a survival gene for lung cancer (30).

In conclusion, a total of 160 overexpressed common genes were found in a patient presenting two of the three pathologies; 154 genes were common overexpressed (in C and L), one gene (in L and F), 4 genes (in C and F.) A total of 148 were common underexpressed two of the three pathologies; 104 genes (in C and L) and 1 gene (in F and L) and 43 genes (in C and F). We found significant biological processes involved in 49 genes out of the 160 overexpressed common genes present in two of the three pathologies (F, C and L), highlighting the genes ACTIN1, LIMK1, CORO1C, INF2, CFL1, FSCN1, MYO1B and SH3D21 having a relationship with biological processes that could be involved in the imbalance of the integrity of the epithelial tissue of the oral cavity. The only gene found to be overexpressed common in the three pathologies was EPRS1, which could in the future be considered as an early biomarker of Oral Cancer.



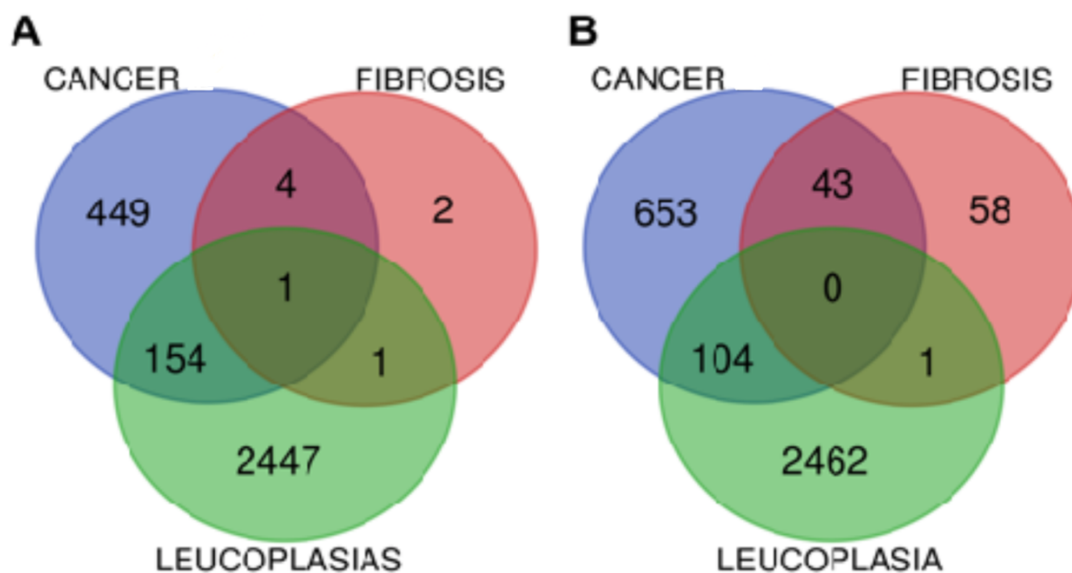
## **Bibliografía**

1. Odontoestomatología AE, Antón MSM, Pérez SM. Revisión de la literatura. *Av Odontoestomatol.* 2015;31(4):247–59.
2. Izaguirre Perez PM, Benevente Lipa LA. Nivel de conocimiento sobre el cancer oral de estudiante del ultimo ano de carrera professional de odontologia de la universidades de Lima. 2012.
3. Andrés U, Chile B. Epidemiological characterization of oral cancer. Literature review. 2023;12(1).
4. Nookaew I, Papini M, Pornputtpong N, Scalcinati G, Fagerberg L, Uhlén M, et al. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2012;40(20):10084–97.
5. Smith CC, Bixby LM, Miller KL, Selitsky SR, Bortone DS, Hoadley KA, et al. Using RNA Sequencing to Characterize the Tumor Microenvironment. *Methods Mol Biol.* 2020;20:245–72.
6. Wu J, Ding Y, Wang J, Lyu F, Tang Q, Song J, et al. Single-cell RNA sequencing in oral science: Current awareness and perspectives. *Cell Prolif.* 2022;55(10):13287.
7. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the Impact of Smoking and Lung Cancer on the Airway Transcriptome Using RNA-Seq. *Cancer Prevention Research.* 2011;4(6):803–17.
8. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.* 2018;7:1338.
9. Eduardo A, Cubillos R, Laura B, Jiménez P, Adriana J, Bernal G. Analizando datos de RNA-seq en procariotas: una revisión para no expert. *Acta Biolo Colomb.* 2014 [cited 2023 Feb 15];19(2):131–42.
10. Smyth GK, Ritchie ME, Law CW, Alhamdoosh M, Su S, Dong X, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res.* 2016;5.
11. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504
12. Khan K, Gogonea V, Fox PL. Aminoacyl-tRNA synthetases of the multi-tRNA synthetase complex and their role in tumorigenesis. *Translational Oncology.* Neoplasia Press, Inc. 2022;19
13. Berral AV, Murua CM, de la Vega FA, García IH, Rosino AA, de Liaño AD, et al. Situación actual en el tratamiento del cáncer gástrico. *Revista Española de Enfermedades Digestivas [Internet].* 2012 [cited 2023 Jun 3];104(3):134–41.

14. Castañeda CA, Casavilca S, Orrego E, García-Corrochano P, Deza P, Heinike H, et al. Glioblastoma: Análisis molecular y sus implicancias clínicas. *Rev Peru Med Exp Salud Publica*.2015;32(2):316–25.
15. Martínez-Sahuquillo Marquez A, Gallardo Castillo I, et al. La leucoplasia oral: Su implicación como lesión precancerosa. *Av Odontoestomatol*-2008;24(1)
16. Fuchs-Telem D, Stewart H, Rapaport D, Nousbeck J, Gat A, Gini M, et al. CEDNIK syndrome results from loss-of-function mutations in SNAP29. *Br J Dermatol*.2011;164(3):610–6.
17. Kim S, You S, Hwang D. Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nature Reviews Cancer*.2011;11(10):708–18.
18. Foley KS, Young PW. The non-muscle functions of actinins: An update. *Biochemical Journal*.2014;459:1–13.
19. Xie GF, Zhao LD, Chen Q, Tang DX, Chen QY, Lu HF, et al. High ACTN1 Is Associated with Poor Prognosis, and ACTN1 Silencing Suppresses Cell Proliferation and Metastasis in Oral Squamous Cell Carcinoma. *Drug Des Devel Ther*.2020;14:1717–27.
20. Chen Q, Zhou XW, Zhang AJ, He K. ACTN1 supports tumor growth by inhibiting Hippo signaling in hepatocellular carcinoma. *J Exp Clin Cancer Res*.2021;40(1).
21. Su J, Zhou Y, Pan Z, Shi L, Yang J, Liao A, et al. Downregulation of LIMK1–ADF/cofilin by DADS inhibits the migration and invasion of colon cancer. *Scientific Reports*.2017;7(1):1–12.
22. Xu J, Lv H, Zhang B, Xu F, Zhu H, Chen B, et al. miR-30b-5p acts as a tumor suppressor microRNA in esophageal squamous cell carcinoma. *J Thorac Dis*.2019;11(7):3015–29.
23. Dai X, Liang Z, Liu L, Guo K, Xu S, Wang H. Silencing of MALAT1 inhibits migration and invasion by sponging miR-1-3p in prostate cancer cells. *Mol Med Rep*.2019;20(4):3499–508.
24. Heuser VD, Mansuri N, Mogg J, Kurki S, Repo H, Kronqvist P, et al. Formin Proteins FHOD1 and INF2 in Triple-Negative Breast Cancer: Association With Basal Markers and Functional Activities. *Breast Cancer*.2018;12.
25. Turhani D, Krapfenbauer K, Thurnher D, Langen H, Fountoulakis M. Identification of differentially expressed, tumor-associated proteins in oral squamous cell carcinoma by proteomic analysis. *Electrophoresis*. 2006;27(7):1417–23.
26. Wang L, Xiong L, Wu Z, Miao X, Liu Z, Li D, et al. Expression of UGP2 and CFL1 expression levels in benign and malignant pancreatic lesions and their clinicopathological significance. *World J Surg Oncol*.2018;16(1).
27. Li Z, Shi J, Zhang N, Zheng X, Jin Y, Wen S, et al. FSCN1 acts as a promising therapeutic target in the blockade of tumor cell motility: a review

- of its function, mechanism, and clinical significance. *J Cancer*.2022;13(8):2528.
28. Lii CK, Chang JW, Chen JJ, Chen HW, Liu KL, Yeh SL, et al. Docosahexaenoic acid inhibits 12-O-tetradecanoylphorbol-13- acetate-induced fascin-1-dependent breast cancer cell migration by suppressing the PKC $\delta$ - and Wnt-1/ $\beta$ -catenin-mediated pathways. *Oncotarget*. 2016;7(18):25162–79.
  29. Xie L, Huang H, Zheng Z, Yang Q, Wang S, Chen Y, et al. MYO1B enhances colorectal cancer metastasis by promoting the F-actin rearrangement and focal adhesion assembly via RhoA/ROCK/FAK signaling. *Ann Transl Med*.2021;9(20):1543–1543.
  30. Masoudi M, Seki M, Yazdanparast R, Yachie N, Aburatani H. A genome-scale CRISPR/Cas9 knockout screening reveals SH3D21 as a sensitizer for gemcitabine. *Sci Rep*.2019;9(1).

**Fig. 1.A** Venn Diagramm de los genes comunes entre sobreexpresados entre Leucoplasia, Cáncer Oral y Fibrosis. **1.B.** Venn Diagramm de los genes comunes infraexpresados entre Leucoplasia, Cáncer Oral, Fibrosis



**Tabla 1:** Genes comunes sobreexpresado entre cáncer oral y leucoplasia.

<i>ID</i>	Nombre del Gen	<i>Fold change (log2)* C vs N</i>	<i>Fold change (log2)* L vs N</i>
<b>ACLY</b>	<i>ATP Citrate Lyase</i>	1.057	0.087
<b>ACTN1</b>	<i>Actinin Alpha 1</i>	1.340	0.074
<b>ADAM19</b>	<i>ADAM Metallopeptidase Domain 19</i>	1.342	0.201
<b>ADM</b>	<i>Adrenomedullin</i>	1.708	0.263
<b>AEN</b>	<i>Apoptosis Enhancing Nuclease</i>	1.371	0.339
<b>ALYREF</b>	<i>Aly/REF Export Factor</i>	1.385	0.117
<b>ATAD3A</b>	<i>ATPase Family AAA Domain Containing 3A</i>	1.434	0.159
<b>ATF5</b>	<i>Activating Transcription Factor 5</i>	1.156	0.156
<b>BOP1</b>	<i>Ribosomal Biogenesis Factor</i>	1.362	0.159
<b>C1orf216</b>	<i>Chromosome 1 Open Reading Frame 216</i>	1.091	0.481
<b>CARD10</b>	<i>Caspase Recruitment Domain Family Member 10</i>	1.377	0.174
<b>CARS</b>	<i>Nuclear Receptor Subfamily 1 Group 1 Member 3</i>	1.024	0.062
<b>CCT3</b>	<i>Chaperonin Containing TCP1 Subunit 3</i>	1.106	0.135

<b>CCT4</b>	<i>Chaperonin Containing TCP1 Subunit 4</i>	0.449	0.042
<b>CDV3</b>	<i>CDV3 Homolog</i>	0.37	0.158
<b>CEBPB</b>	<i>CCAAT Enhancer Binding Protein Beta</i>	1.581	0.332
<b>CFL1</b>	<i>Cofilin 1</i>	0.921	0.068
<b>CHST2</b>	<i>Carbohydrate Sulfotransferase 2</i>	3.001	0.781
<b>CHST3</b>	<i>Carbohydrate Sulfotransferase 3</i>	1.069	0.351
<b>CLIC4</b>	<i>Chloride Intracellular Channel 4</i>	0.728	0.138
<b>CLSPN</b>	<i>Claspin</i>	2.011	0.203
<b>COL27A1</b>	<i>Collagen Type XXVII Alpha 1 Chain</i>	1.357	0.114
<b>COL4A1</b>	<i>Collagen Type IV Alpha 1 Chain</i>	1.386	0.02
<b>CORO1C</b>	<i>Coronin 1C</i>	0.676	0.08
<b>CPA4</b>	<i>Carboxypeptidase A4</i>	1.993	0.329
<b>CTPS1</b>	<i>CTP Synthase 1</i>	1.327	0.077
<b>CTSL</b>	<i>Cathepsin L</i>	1.453	0.222
<b>CYP19A1</b>	<i>Cytochrome P450 Family 19 Subfamily A Member 1</i>	3.076	1.204
<b>CYP24A1</b>	<i>Cytochrome P450 Family 24 Subfamily A Member 1</i>	3.486	1.103
<b>DAP3</b>	<i>Death Associated Protein 3</i>	0.508	0.085
<b>DAPK3</b>	<i>Death Associated Protein Kinase 3</i>	1.46	0.104
<b>DDX21</b>	<i>DEAD-Box Helicase 21</i>	1.452	0.116
<b>DDX23</b>	<i>DEAD-Box Helicase 23</i>	0.802	0.048
<b>DES1</b>	<i>Desumoylating Isopeptidase 1</i>	0.838	0.147
<b>DKC1</b>	<i>Dyskerin Pseudouridine Synthase 1</i>	0.739	0.172
<b>EBNA1BP2</b>	<i>EBNA1 Binding Protein 2</i>	0.999	0.144
<b>EEF1D</b>	<i>Eukaryotic Translation Elongation Factor 1 Delta</i>	0.704	0.06
<b>EFNB1</b>	<i>Ephrin B1</i>	1.484	0.3
<b>EHD1</b>	<i>EH Domain Containing 1</i>	1.133	0.136
<b>EIF3B</b>	<i>Eukaryotic Translation Initiation Factor 3 Subunit B</i>	0.684	0.096
<b>EIF4A1</b>	<i>Eukaryotic Translation Initiation Factor 4A1</i>	0.899	0.055
<b>ELAC2</b>	<i>ElaC Ribonuclease Z 2</i>	0.685	0.061
<b>ENTPD7</b>	<i>Ectonucleoside Triphosphate Diphosphohydrolase 7</i>	1.183	0.089
<b>EPHB2</b>	<i>EPH Receptor B2</i>	1.727	0.615
<b>EPPK1</b>	<i>Epiplakin 1</i>	3.504	0.196
<b>ETV4</b>	<i>ETS Variant Transcription Factor 4</i>	1.528	0.59
<b>FBXO45</b>	<i>F-Box Protein 45</i>	2.134	0.218
<b>FMNL2</b>	<i>Formin Like 2</i>	1.37	0.084
<b>FSCN1</b>	<i>Fascin Actin-Bundling Protein 1</i>	1.996	0.224
<b>FSD1L</b>	<i>Fibronectin Type III And SPRY Domain Containing 1 Like</i>	0.831	0.328

<b>FTSJ1</b>	<i>FtsJ RNA 2'-O-Methyltransferase 1</i>	0.661	0.126
<b>GALNT18</b>	<i>Polypeptide N-Acetylgalactosaminyltransferase 18</i>	1.001	0.17
<b>GALNT6</b>	<i>Polypeptide N-Acetylgalactosaminyltransferase 6</i>	2.704	0.434
<b>GPATCH4</b>	<i>G-Patch Domain Containing 4 (Gene/Pseudogene)</i>	0.934	0.143
<b>GPR153</b>	<i>G Protein-Coupled Receptor 153</i>	1.377	0.237
<b>HARS</b>	<i>Histidyl-tRNA synthetase</i>	0.507	0.057
<b>HOMER3</b>	<i>Homer Scaffold Protein 3</i>	1.873	0.301
<b>HSP90B1</b>	<i>Heat Shock Protein 90 Beta Family Member 1</i>	0.730	0.053
<b>HSPA5</b>	<i>Heat Shock Protein Family A (Hsp70) Member 5</i>	1.242	0.068
<b>HSPH1</b>	<i>Heat Shock Protein Family H (Hsp110) Member 1</i>	1.199	0.127
<b>HYOU1</b>	<i>Hypoxia Up-Regulated 1</i>	0.682	0.106
<b>INF2</b>	<i>Inverted Formin 2</i>	0.734	0.115
<b>ITGA3</b>	<i>Integrin Subunit Alpha 3</i>	1.710	0.092
<b>KIAA0930</b>	<i>KIAA0930</i>	0.740	0.223
<b>KLF7</b>	<i>KLF Transcription Factor 7</i>	1.114	0.193
<b>KRT17</b>	<i>Keratin 17</i>	5.685	0.683
<b>LAMC2</b>	<i>Laminin Subunit Gamma 2</i>	3.083	0.101
<b>LIMK1</b>	<i>LIM Domain Kinase 1</i>	1.278	0.225
<b>LMNB2</b>	<i>Lamin B2</i>	1.505	0.111
<b>LRP8</b>	<i>LDL Receptor Related Protein 8</i>	1.381	0.338
<b>LRRC59</b>	<i>Leucine Rich Repeat Containing 59</i>	1.000	0.171
<b>LY6K</b>	<i>Lymphocyte Antigen 6 Family Member K</i>	2.447	0.696
<b>LYAR</b>	<i>Ly1 Antibody Reactive</i>	1.209	0.188
<b>MDFI</b>	<i>MyoD Family Inhibitor</i>	1.698	0.203
<b>MET</b>	<i>MET Proto-Oncogene, Receptor Tyrosine Kinase</i>	1.051	0.144
<b>MICAL2</b>	<i>Microtubule Associated Monooxygenase, Calponin And LIM Domain Containing 2</i>	1.748	0.085
<b>MRPL14</b>	<i>Mitochondrial Ribosomal Protein L14</i>	1.042	0.156
<b>MSN</b>	<i>Moesin</i>	0.726	0.228
<b>MTHFD1L</b>	<i>Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1 Like</i>	1.549	0.175
<b>MTHFD2</b>	<i>Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 2, Methenyltetrahydrofolate Cyclohydrolase</i>	1.017	0.156
<b>MYBL2</b>	<i>MYB Proto-Oncogene Like 2</i>	1.55	0.13
<b>MYO1B</b>	<i>Myosin IB</i>	1.978	0.233
<b>NACC1</b>	<i>Nucleus Accumbens Associated 1</i>	0.933	0.135
<b>NADK</b>	<i>NAD Kinase</i>	0.79	0.117

<b>NCF2</b>	<i>Neutrophil Cytosolic Factor 2</i>	1.123	0.385
<b>NKRF</b>	<i>NFKB Repressing Factor</i>	0.729	0.185
<b>NOC2L</b>	<i>NOC2 Like Nucleolar Associated Transcriptional Repressor</i>	0.907	0.052
<b>NOMO1</b>	<i>NODAL Modulator 1</i>	0.881	0.082
<b>NOP56</b>	<i>NOP56 Ribonucleoprotein</i>	0.964	0.079
<b>NPLOC4</b>	<i>NPL4 Homolog, Ubiquitin Recognition Factor</i>	0.080	0.091
<b>NRIP1</b>	<i>Nuclear Receptor Interacting Protein 2</i>	1.224	0.088
<b>NSUN2</b>	<i>NOP2/Sun RNA Methyltransferase 2</i>	0.944	0.063
<b>NUP62</b>	<i>Nucleoporin 62</i>	0.587	0.076
<b>OAS3</b>	<i>2'-5'-Oligoadenylate Synthetase 3</i>	1.877	0.094
<b>PA2G4</b>	<i>Proliferation-Associated 2G4</i>	0.869	0.111
<b>PDCD11</b>	<i>Programmed Cell Death 11</i>	1.149	0.068
<b>PFDN2</b>	<i>Prefoldin Subunit 2</i>	1.101	0.359
<b>PLAU</b>	<i>Plasminogen Activator, Urokinase</i>	2.666	0.096
<b>PLCD3</b>	<i>Phospholipase C Delta 3</i>	1.228	0.085
<b>PLEC</b>	<i>Plectin</i>	1.528	0.068
<b>PLOD3</b>	<i>Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 3</i>	0.831	0.194
<b>PMEPA1</b>	<i>Prostate Transmembrane Protein, Androgen Induced 1</i>	1.725	0.238
<b>PNLIPRP3</b>	<i>Pancreatic Lipase Related Protein 3</i>	4.07	2.771
<b>PNP</b>	<i>Purine Nucleoside Phosphorylase</i>	1.132	0.109
<b>PPP1R14C</b>	<i>Protein Phosphatase 1 Regulatory Inhibitor Subunit 14C</i>	1.805	0.239
<b>PPP2R5B</b>	<i>Protein Phosphatase 2 Regulatory Subunit B'Beta</i>	1.053	0.154
<b>PRNP</b>	<i>Prion Protein</i>	1.164	0.253
<b>PSMB7</b>	<i>Proteasome 26S Subunit, Non-ATPase 7</i>	0.822	0.71
<b>PSMC3</b>	<i>Proteasome 26S Subunit, ATPase 3</i>	1.387	0.102
<b>PSMD2</b>	<i>Proteasome 26S Subunit Ubiquitin Receptor, Non-ATPase 2</i>	1.095	0.98
<b>PSMD3</b>	<i>Proteasome 26S Subunit, Non-ATPase 3</i>	0.607	0.103
<b>PTHLH</b>	<i>Parathyroid Hormone Like Hormone</i>	2.367	0.539
<b>PTPRZ1</b>	<i>Protein Tyrosine Phosphatase Receptor Type Z1</i>	2.561	0.062
<b>PTTG1IP</b>	<i>PTTG1 Interacting Protein</i>	0.728	0.129
<b>PUS1</b>	<i>Pseudouridine Synthase 1</i>	0.954	0.171
<b>RAD23A</b>	<i>RAD23 Homolog A, Nucleotide Excision Repair Protein</i>	0.834	0.157
<b>RGS20</b>	<i>Regulator Of G Protein Signaling 20</i>	2.397	0.654
<b>RRAS2</b>	<i>RAS Related 2</i>	0.835	0.171
<b>RRS1</b>	<i>Ribosome Biogenesis Regulator 1 Homolog</i>	1.374	0.413
<b>SDF4</b>	<i>Stromal Cell Derived Factor 4</i>	0.955	0.065

<b>SDK2</b>	<i>Sidekick Cell Adhesion Molecule 2</i>	1.792	0.564
<b>SEMA3C</b>	<i>Semaphorin 3C</i>	1.789	0.337
<b>SERPINB7</b>	<i>Serpin Family B Member 7</i>	3.093	0.737
<b>SH2D2A</b>	<i>SH2 Domain Containing 2A</i>	1.402	0.301
<b>SH3D21</b>	<i>SH3 Domain Containing 21</i>	1.216	0.112
<b>SLC15A1</b>	<i>Solute Carrier Family 15 Member 1</i>	2.131	0.258
<b>SLC16A1</b>	<i>Transmembrane Channel Like 7</i>	1.861	0.358
<b>SLC28A3</b>	<i>Solute Carrier Family 28 Member 3</i>	3.059	0.433
<b>SLC3A2</b>	<i>Solute Carrier Family 3 Member 2</i>	1.411	0.08
<b>SMOX</b>	<i>Spermine Oxidase</i>	2.494	0.339
<b>SQLE</b>	<i>Squalene Epoxidase</i>	1.701	0.138
<b>STIP1</b>	<i>Stress Induced Phosphoprotein 1</i>	1.068	0.126
<b>STRIP2</b>	<i>Striatin Interacting Protein 2</i>	2.287	0.633
<b>STX1A</b>	<i>Syntaxin 1A</i>	1.537	0.613
<b>SZRD1</b>	<i>SUZ RNA Binding Domain Containing 1</i>	0.438	0.138
<b>TGM2</b>	<i>Transglutaminase 2</i>	2.319	0.414
<b>THOC5</b>	<i>THO Complex Subunit 5</i>	0.468	0.066
<b>TIMM17A</b>	<i>Translocase Of Inner Mitochondrial Membrane 17A</i>	0.781	0.143
<b>TMC7</b>	<i>Transmembrane Channel Like 7</i>	1.244	0.602
<b>TMEM184B</b>	<i>Transmembrane Protein 184B</i>	1.027	0.17
<b>TNC</b>	<i>Tenascin C</i>	2.178	0.214
<b>TOP1MT</b>	<i>DNA Topoisomerase I Mitochondrial</i>	1.18	0.155
<b>TRMT6</b>	<i>TRNA Methyltransferase 6 Non-Catalytic Subunit</i>	0.966	0.108
<b>TRPV3</b>	<i>Transient Receptor Potential Cation Channel Subfamily V Member 3</i>	2.871	0.366
<b>TUBB3</b>	<i>Tubulin Beta 3 Class III</i>	2.060	0.614
<b>TUBG1</b>	<i>Tubulin Gamma 1</i>	1.067	0.207
<b>UHRF1</b>	<i>Ubiquitin Like With PHD And Ring Finger Domains 1</i>	1.197	0.104
<b>USB1</b>	<i>U6 SnRNA Biogenesis Phosphodiesterase 1</i>	1.561	0.175
<b>WDR5</b>	<i>WD Repeat Domain 5</i>	0.841	0.056
<b>WDR53</b>	<i>WD Repeat Domain 53</i>	1.154	0.147
<b>WDR62</b>	<i>WD Repeat Domain 62</i>	0.998	0.284
<b>WDR66</b>	<i>WD Repeat Domain 66</i>	1.971	0.665
<b>YWHAG</b>	<i>Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Gamma</i>	0.679	0.195
<b>ZNF622</b>	<i>Zinc Finger Protein 622</i>	0.544	0.073

\*The *Fold-change* estimados (cambios de expresión relativa de los genes).

**Tabla 2:** Genes comunes sobreexpresado entre pacientes con cáncer y fibrosis



<i>ID</i>	<i>Nombre gen</i>	<i>Fold change (log2)* C vs N</i>	<i>Fold change (log2)* F vs N</i>
<i>RP11-23P13.6</i>	<i>Long-non-coding</i>	1.291	1.366
<i>STARD 4</i>	<i>StAR Related Lipid Transfer Domain Containing 4</i>	1.292	1.150
<i>RP11-420K8.1</i>	<i>Long-non-coding RNA</i>	1.293	1.538
<i>NUS1</i>	<i>NUS1 Dehydrodolichyl Diphosphate Synthase Subunit</i>	1.294	0.614

**Tabla 3:** Gen común sobreexpresados entre las 3 patologías

<i>ID</i>	<i>Nombre gen</i>	<i>Fold change (log2)* C vs N</i>	<i>Fold change (log2)* F vs N</i>	<i>Fold change (log2)* L vs N</i>
<i>EPRS1</i>	<i>Glutamyl-prolyl-tRNA synthetase</i>	0.653	0.492	0.091

**METAANÁLISIS DE EXPRESIÓN GÉNICA EN PACIENTES CON CÁNCER,  
FIBROSIS Y LEUCOPLASIAS EN LA CAVIDAD ORAL MEDIANTE DATOS  
PÚBLICOS DE RNA-SEQ**

**Título corto: Metaanálisis de expresión génica en pacientes con cáncer y  
enfermedades pre-cancerígenas orales con datos público de RNA-seq**

**Autores:**

**Agnese Formaggio<sup>1</sup>, Marta Serna-García<sup>2</sup>**

*<sup>1</sup> 5th year student of the Dentistry degree at European University of Valencia,  
Valencia, Spain.*

*<sup>2</sup> Department of Dentistry, Faculty of Health Sciences, European University of  
Valencia, Valencia, Spain.*

**Correspondencia**

Marta Serna Garcia

Paseo de Alameda 7, Valencia

46010, Valencia

marta.serna@universidadeuropes.es

## Resumen

**Introducción:** El RNA-seq es una herramienta que descubre nuevos biomarcadores de cáncer oral y lesiones pre-cancerígenas como la leucoplasia y fibrosis oral, además, estudia genes relevantes y conoce la estructura biológica pudiendo ayudar a desarrollar terapias. El objetivo fue realizar un estudio transcriptómico e identificar los procesos biológicos en la cavidad oral mediante RNA-seq para muestras obtenidas de bases de datos públicas de grupos divergentes; muestras de pacientes con lesiones pre-cancerosas: Leucoplasia (L) y Fibrosis oral (F) y Cáncer oral (C) comparada con pacientes sin patología (N).

**Materiales y métodos:** Seleccionaron tres estudios de la base de datos pública *GeneExpression Omnibus* (*GSE125866*<sup>(2019)</sup>, *GSE20116*<sup>(2010)</sup> y *GSE131568*<sup>(2020)</sup>). Todos los archivos brutos se analizaron con el mismo flujo de trabajo. La calidad de lecturas se verificó mediante el *Fastqc software* y se filtraron con *Trimmomatic software*. Se alineó con genoma *Homo Sapiens* con *Kallisto software*. *EdgeR software* y *Limma software* normalizaron las lecturas e identificaron genes expresados diferencialmente ( $FDR \leq 0.05$ ). Para determinar procesos biológicos significativos fue usado *DAVID software* ( $p \text{ value} \leq 0,05$ ). *Cytoscape software* fue utilizado para términos biológicos.

**Resultados:** Se detectaron 800 genes infraexpresados y 608 sobreexpresados (C vsN), 102 genes infraexpresado y 8 sobreexpresado (FvsN) y 2567 genes infraexpresado y 2603 sobreexpresado (LvsN). 154 genes comunes sobreexpresados (C y L), un gen (L y F), 4 genes (C y F) 148 genes comunes infraexpresado (C y L) y 104 genes (F y L) y 1 gen (C y F). EPRS fue el único gen común en las tres patologías ( $Foldchange 0.653$ ). De los 49 genes destacamos 8 involucrados con el tejido epitelial.

**Conclusiones:** Los genes ACTIN1, LIMK1, CORO1C, INF2, CFL1, FSCN1, MYO1B y SH3D21 podrían influenciar la integridad del tejido epitelial y ayudar al avance de lesiones precancerosas y cáncer. El EPRS es un posible biomarcador temprano.

**Palabras claves:** RNA-seq, cáncer oral, fibrosis, leucoplasia, tejido epitelial, análisis transcriptómica

## Introducción

El cáncer es un problema de salud a nivel mundial y es una de las causas más importantes de mortalidad (1). Uno de los cánceres orales más frecuente es el de cabeza y cuello; se encuentra dentro de las 10 neoplasias malignas más comunes y se observa un aumento de los casos diagnosticados en los últimos 10 años (2).

Los estudios de las enfermedades pre-cáncer y cáncerosas oral son de suma importancia debido a un aumento en el número de casos en los últimos años (3) por lo que su conocimiento, prevención y diagnóstico se va enfatizando cada día más para los odontólogos, médicos y científicos; Gracias a la tecnología RNA seq, existen de números estudios para estas patologías (4–7) pero aún se necesita delucitar más genes que podrían ser responsable de estas patologías, así como entender mejor las funciones biológicas pudiendo en un futuro ayudar a las terapias génicas. De esta manera, identificar genes que están siendo activados comunes a diferentes lesiones precancerosas y también con el cáncer oral comparado con persona sin patologías, podrían ser la clave para ayudar a la detección precoz de estas patologías.

Este trabajo nos permitirá analizar la expresión de los genes e identificar los procesos biológicos de interés en la cavidad oral en humanos mediante la técnica de RNA-seq para diferentes muestras (de pacientes con lesiones precancerosas comparada con muestras de pacientes sin patologías), se detectaran los genes diferenciales (sobre e infra expresados) comunes de pacientes con cáncer oral *versus* tejido sano, fibrosis *versus* tejido sano y Leucoplasia *versus* tejido sano, se realizaran un análisis de enriquecimiento para identificar procesos biológicos significativos para cáncer oral, fibrosis y Leucoplasia y se identificaran los genes diferenciales comunes más relevantes asociados a cáncer y lesiones precancerosas para en un futuro podrían utilizarse como biomarcadores.

## Materiales y métodos

### - **Diseño del estudio**

Fueron seleccionados diferentes estudios de datos brutos de secuenciación del RNA de muestras de pacientes a través de la técnica de RNA-seq, en las bases de datos públicas de Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), repositorio de datos de genómica funcional pública, datos basados en matrices y secuencias y BioStudies (<https://www.ebi.ac.uk/biostudies/arrayexpress>). Fueron seleccionados tres estudios con el número de acceso a la base de datos de GSE125866, GSE20116 y GSE131568. El estudio **GSE125866** tenemos n=40 muestras de tejido de la cavidad oral, de las cuales n=10 son de pacientes que presentan fibrosis, n=20 muestras de pacientes que presentan cáncer y n=10 de pacientes que no muestran patología y la consideramos muestras control (normal). El estudio **GSE20116** tenemos n=20 muestras de tejido de la cavidad oral, de las cuales n=6 muestras de pacientes que presentan cáncer y n=14 de pacientes que no muestran patología y la consideramos muestras control (normal). El estudio **GSE131568** tenemos n=26 muestras de tejido de la cavidad oral, de las cuales n=6 muestras de pacientes que presentan Leucoplasia y n=20 de pacientes que no muestran patología y la considerados muestras control (normal).

Realizamos 4 análisis diferenciales con los datos brutos de las muestras de los pacientes de los 3 estudios seleccionados. Para el primer análisis seleccionamos las muestras del estudio GSE125866, 10 muestras de fibrosis las comparamos con 10 muestras control. El segundo análisis seleccionamos las muestras del estudio GSE125866, 10 muestras de cáncer las comparamos con 10 muestras control. El tercer análisis seleccionamos las muestras del estudio GSE20116, 6 muestras de cáncer las comparamos con 14 muestras control. El cuarto análisis seleccionamos las muestras del estudio GSE131568, 6 muestras de Leucoplasia las comparamos con 20 muestras control.

### - **Análisis de expresión diferencial**

Todos los archivos brutos correspondientes a cada muestra de los estudios se analizaron con el mismo flujo de trabajo.

La calidad de las lecturas de RNA-Seq (índices de calidad, contenido de GC, contenido de N, distribuciones de longitud, duplicación, secuencias sobrerrepresentadas y contenido de K-mer) se verificó mediante el software Fastqc (v.0.11.4) (8).

Las lecturas de baja calidad (secuencia del adaptador y lecturas que contienen poli-N) de datos sin procesar usando Trimmomatic v.0.36 (61) con los siguientes parámetros: `java -jar /usr/local/bin/trimmomatic-0.36.jar PE -phred33 <IN_R1.fastq.gz> <IN_R2.fastq.gz> <OUT_R1.fastq.gz> <OUT_R1_UN.fastq.gz> <OUT_R2.fastq.gz> <OUT_R2_UN.fastq.gz> LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50`.

Todo el análisis posterior se basó en los datos recortados con lecturas de alta calidad, se utilizó KALLISTO (9) para alinear las lecturas recortadas de extremos emparejados con el genoma de referencia Homo Sapiens depositado en el Centro Nacional de Información Biotecnológica (NCBI) (<https://www.ncbi.nlm.nih.gov/>).

Este recuento de lecturas se llevó a cabo utilizando el script interno y Ensembl (Índice de [/pub/release-105/gtf/homo\\_sapiens/](http://pub/release-105/gtf/homo_sapiens/) (ensembl.org)) para las anotaciones de los genes.

Los paquetes EdgeR y Limma disponibles del proyecto Bioconductor se utilizaron para estimar lecturas de CPM normalizadas (recuentos por millón) e identificar genes expresados diferencialmente (DEG) entre los grupos.

Se utilizó EdgeR para normalizar los datos y el software Limma para los métodos estadísticos (modelado lineal y Bayes empírico), siguiendo la metodología de Law et al. (10).

Consideramos que un gen se expresa diferencialmente cuando el valor de FDR (false positive) es  $\leq 0.05$

#### - **Análisis de enriquecimiento**

Se realizó análisis de enriquecimiento de los genes diferenciales mediante el software DAVID v.6.8 (<https://david.ncifcrf.gov>) usando la base de datos de la

Enciclopedia de Genes y Genomas de Kyoto (KEGG) ( $p \text{ value} \leq 0,05$ ). Se utilizó el ClueGO, un complemento de Cytoscape software que visualiza los términos biológicos no redundantes para grandes grupos de genes en una red agrupada funcionalmente (11).

### **Resultados**

Se han encontrado 800 genes infraexpresado y 608 muestra sobreexpresado en el primer estudio (cáncer vs pacientes sanos), 102 genes infra expresado y 8 sobre expresado en el segundo estudio (fibrosis vs pacientes sanos y 2567 genes infra expresado y 2603 muestras sobre expresado en el tercer estudio (leucoplasia vs pacientes sanos).

Todos los genes diferenciales obtenidos se aplicó la corrección FDR (*False Discovery Rate*) menor que 0,05.

Se han detectado un total de 160 genes comunes sobre expresados las diferentes patologías. Se ha identificado un solo gen sobreexpresado común entre las tres patologías, 154 genes comunes entre el cáncer oral y la leucoplasia, un gen común entre fibrosis y leucoplasia y 4 genes comunes entre cáncer oral y fibrosis.

Se ha detectado un total de 148 genes comunes infraexpresados comparando las diferentes patologías. No se han detectado genes infraexpresados comunes entre las tres patologías; entre el cáncer oral y la leucoplasia encontramos 104 genes comunes, entre fibrosis y leucoplasia 1 y entre cáncer oral y fibrosis 43.

El único gen sobre expresado común encontrado entre paciente que presentan cáncer oral, fibrosis o leucoplasia, comparados con pacientes sanos fue el EPRS1 (glutamyl-prolyl-tRNA synthetase 1) con un Fold change de 0.653 en el estudio entre cáncer (C) y paciente sano (N), un Fold change de 0.492 en el estudio entre fibrosis (F) y paciente sano (N) y un Fold change de 0.091 en el último estudio entre leucoplasia (L) y paciente sano (N) (Tabla 3). Destacamos 49 genes involucrados en los procesos biológicos, en la función molecular y en el componente molecular de interés en el estudio ( $p \text{ valor} < 0,05$ ). De los 49 genes encontrados destacamos los genes: *INF2*, *ACTN1*, *CFL1*, *LIMK1*,

*FSCN1, MYO1B, SH3D21, CORO1C*) siendo involucrados directamente en la formación del tejido epitelial.

### **Discusión**

El gen EPRS1 (Glutamyl-prolyl-tRNA synthetase 1) es considerado lo con más relevancia por encontrarse sobreexpresado en muestras asociadas al cáncer oral, fibrosis oral y leucoplasia oral. Este gen es un gen codificador de proteínas. En el estudio de Krishnendu Khan y cols. el gen EPRS1 presenta una mayor resistencia al tamoxifeno que son moduladores selectivos de receptores de estrógeno utilizado como medicamentos contra el cáncer, y su vez un mal pronóstico en las salas de emergencia (12).

En otro estudio se observó EPRS1 estaba sobreexpresado en el tejido tumoral de mama en comparación con el tejido normal y se asoció con una mayor destrucción celular y una peor supervivencia de los pacientes (13) también en cáncer de pulmón promueve la extinción y la invasión de las células cancerígenas, y que su inhibición reduzca significativamente el crecimiento tumoral y la capacidad metastásica (14). Tannous et al. en el 2015 investigaron los perfiles de expresión génica en muestras de leucoplasia oral y cáncer oral donde encontraron el gen EPIRS más elevado (15). En un estudio (16) con células epiteliales de la piel y cabello se vio que la pérdida de EPRS1 en células epiteliales condujo a alteraciones en el desarrollo y función de la piel.

Este gen también se vio relacionado en la integridad de la barrera intestinal al regular la expresión de genes relacionados con la función de barrera y la respuesta inflamatoria (17). Sin embargo, se necesitan más estudios para comprender completamente los mecanismos y la implicación precisa de EPRS1 en el tejido epitelial.

Otro gen relevante es el Las Actinas (ACTN) -1, denominada también “actinas no musculares”, son las principales proteínas del citoesqueleto y realizan importantes funciones como la regulación de la citocinesis, la adhesión celular, la migración y median la función del sarcómero (18). En nuestro estudio se mostró que los niveles de proteína Actina 1 del tejido



enfermo estaban mayores respecto a lo del tejido sano; lo que indica que la proteína ACTN1 están mayormente expresada en paciente que padece de cáncer oral; De acuerdo con el estudio de Guo-feng Xie y cols. los cuales consideran el ACTIN 1 una de los principales marcadores de los tejidos que presentaban el cáncer carcinoma oral de la célula escamosa (OSCC), además, se vieron los niveles aumentados estaban asociado a metástasis de los ganglios con un mal pronóstico (19). De manera similar el ACTIN 1 está relacionado también a otras tipologías de cáncer, como por ejemplo el cáncer de mama o el carcinoma hepatocelular (20); También está respaldado por parte de Xu y cols. Los cuales con su estudio encontraron también los niveles de ACTIN1 en el tejido con OSCC que en el tejido normal. Otro gen sobreexpresado que se ha detectado relevante en nuestro estudio es el LIM Domain Kinase 1 (LMNK1), es una serina quinasa que cataliza la fosforilación del grupo hidróxido de la serina, la cuales regulan la polimerización de la actina; Este gen tiene la función de regular los movimientos celulares y el ciclo celular (21).

De acuerdo con nuestro estudio también en el estudio de Bingxia Shi y cols. se ha detectado sobreexpresado el gen en el cáncer oral, sobre todo en el cáncer OSCC, el cual se afirma que en la tumorigenesis y metastasis del cáncer está afectada cuando LIMK1, además en este estudio se ha detectado también una mayor activación del gen LIM Domain Kinase 2 (LIMK2): En ambos los genes se desconocen el papel desenvuelve (22).

El tercer gen sobreexpresado es Coronin 1C (COROC1), gen que ejecuta como una plataforma estable para la interacción con otras proteínas, conforme con el estudio de Xiaonfan Dai y cols. los cuales afirman que, debido a su función de motilidad celular, el gen CORO1C juega un papel fundamental para la metástasis de tumores malignos demostrando la función de gen en el cáncer (23).

El gen INF2 (Inverted Formin 2) es el cuarto gen sobreexpresado que encontramos en nuestro estudio; de acuerdo con nuestro estudio, Vanina D Heuser et cols. Encontró actividad del INF2 sobreelevada en el cáncer de mama de tipo basal, forma de cáncer más agresiva, además en este estudio

se notó que la eliminación de INF2 iba a reducir la proliferación de las células tumorales (24) .

El gen CFL1 hace de función como receptor juega un el importante en la proliferación y migración celular en el cuales encontramos funciones en la aparición, desarrollo, infiltración y metástasis de lesiones tumorales. En nuestro estudio encontramos niveles del gen sobreexpresado en el tejido epitelial de la cavidad oral en pacientes que presentan lesiones malignas. Dritan Turhani y cols. Encontraron CFL1 altamente expresado en el cáncer carcinoma oral de la célula escamosa (25) además, Min- li y cols. encontraron relaciones en los cánceres al páncreas, mama, vesícula biliar y demostraron que la inhibición de la actividad de CFL1 en un tumor maligno puede ser un objetivo potencial para inhibir la progresión tumoral y la metástasis (26).

El gen FSCN1 está centrada en el papel de micropuntas, filopodios, pseudópodos lamelares y otras protuberancias basadas en actina debajo de la membrana plasmática, que son importantes para procesos que incluyen la migración celular y la adhesión a la matriz celular, Zhongxun Li y cols ha indicado una expresión altamente significativa en los tejido de cáncer de la células escamosas de la Laringe, además se descubrió que el FSCN1 unido a otra proteína estaba involucrado en la progresión maligna de dicho cáncer; Además de estar involucrado en cáncer de laringe también lo encontramos sobreexpresado en cáncer respiratorios, del aparato digestivo, de sistema Urinario y del sistema Genital (27). Lii C y cols. considera que la inhibición del FSCN1 con fármacos contra el cáncer tiene un potencial significativo para el tratamiento y para aplicaciones clínicas (28).

MYO1B es considerado importante para la migración de la células y la motilidad celulares; en nuestro estudio encontramos nivéleles elevados de acuerdo con Lang Xie y cols lo considera un biomarcador fundamental para la metástasis del cáncer (29).

Por último, tenemos el gen S3D21 con una elevada actividad en nuestro estudio. Masoundi M y cols. considera que, aunque no sea un gen muy estudiado y que no se ha encontrado mucho informe para el estudio, el

S3D21 puede ser relacionados con la activación y sensibilización de las células del cáncer, denominándose un gen de supervivencia para el cáncer de los pulmones (30). En conclusión, se encontraron un total de 160 genes comunes sobreexpresados en paciente que presenta dos de las tres patologías; 154 genes fueron comunes sobreexpresados (en C y L), un gen (en L y F), 4 genes (en C y F.) Un total de 148 fueron comunes infraexpresados dos de las tres patologías; 104 genes (en C y L) y 1 gene (en F y L) y 43 genes (en C y F). Encontramos procesos biológicos significativos involucrados 49 genes de los 160 genes comunes sobreexpresados presentes en dos de las tres patologías (F,C y L) destacando los genes ACTIN1, LIMK1, CORO1C, INF2, CFL1, FSCN1, MYO1B y SH3D21 teniendo relación con procesos biológicos que podrían estar involucrados en el desequilibrio de la integridad del tejido epitelial de la cavidad oral El único gen encontrado sobreexpresado común en las tres patologías fue EPRS1, que podría en un futuro considerarse como un biomarcador temprano de Cáncer Oral.

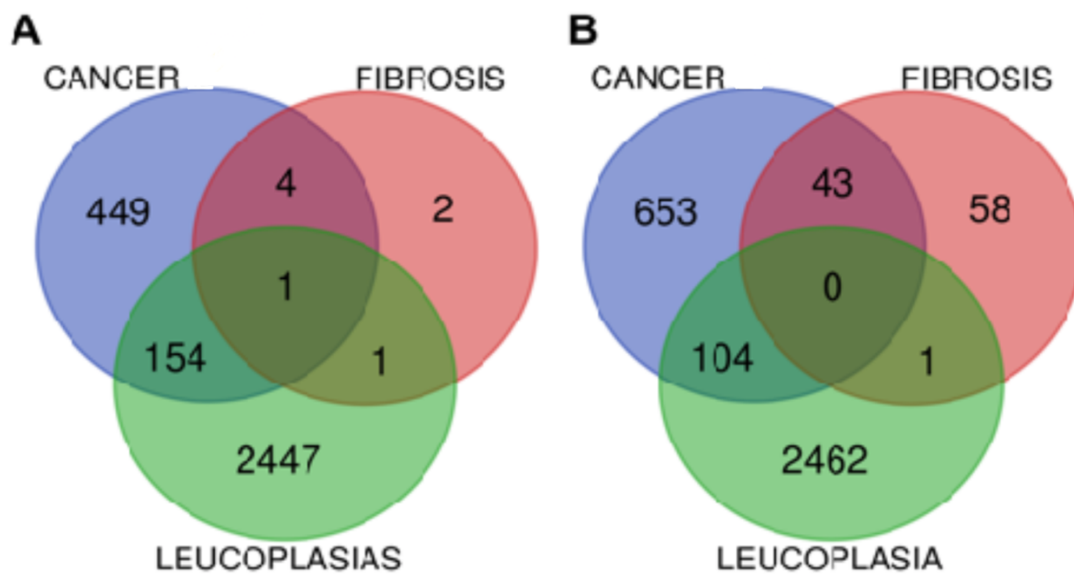
### **Bibliografía**

1. Odontoestomatología AE, Antón MSM, Pérez SM. Revisión de la literatura. Av Odontoestomatol. 2015;31(4):247–59.
2. Izaguirre Perez PM, Benevente Lipa LA. Nivel de conocimiento sobre el cancer oral de estudiante del ultimo ano de carrera professional de odontologia de la universidades de Lima. 2012.
3. Andrés U, Chile B. Epidemiological characterization of oral cancer. Literature review.2023;12(1).
4. Nookaew I, Papini M, Pornputtpong N, Scalcinati G, Fagerberg L, Uhlén M, et al. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. Nucleic Acids Res.2012;40(20):10084–97.
5. Smith CC, Bixby LM, Miller KL, Selitsky SR, Bortone DS, Hoadley KA, et al. Using RNA Sequencing to Characterize the Tumor Microenvironment. Methods Mol Biol.2020;20:245–72.

6. Wu J, Ding Y, Wang J, Lyu F, Tang Q, Song J, et al. Single-cell RNA sequencing in oral science: Current awareness and perspectives. *Cell Prolif.*2022;55(10):13287.
7. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the Impact of Smoking and Lung Cancer on the Airway Transcriptome Using RNA-Seq. *Cancer Prevention Research.* 2011;4(6):803–17.
8. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.*2018;7:1338.
9. Eduardo A, Cubillos R, Laura B, Jiménez P, Adriana J, Bernal G. Analizando datos de RNA-seq en procariotas: una revisión para no expert. *Acta Biolo Colomb.*2014 [cited 2023 Feb 15];19(2):131–42.
10. Smyth GK, Ritchie ME, Law CW, Alhamdoosh M, Su S, Dong X, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res.*2016;5.
11. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*2003;13(11):2498–504
12. Khan K, Gogonea V, Fox PL. Aminoacyl-tRNA synthetases of the multi-tRNA synthetase complex and their role in tumorigenesis. *Translational Oncology.* Neoplasia Press, Inc.2022;19
13. Berral AV, Murua CM, de la Vega FA, García IH, Rosino AA, de Liaño AD, et al. Situación actual en el tratamiento del cáncer gástrico. *Revista Española de Enfermedades Digestivas [Internet].* 2012 [cited 2023 Jun 3];104(3):134–41.
14. Castañeda CA, Casavilca S, Orrego E, García-Corrochano P, Deza P, Heinike H, et al. Glioblastoma: Análisis molecular y sus implicancias clínicas. *Rev Peru Med Exp Salud Publica.*2015;32(2):316–25.
15. Martínez-Sahuquillo Marquez A, Gallardo Castillo I, et al. La leucoplasia oral: Su implicación como lesión precancerosa. *Av Odontostomatol.*2008;24(1).
16. Fuchs-Telem D, Stewart H, Rapaport D, Nousbeck J, Gat A, Gini M, et al. CEDNIK syndrome results from loss-of-function mutations in SNAP29. *Br J Dermatol.*2011;164(3):610–6.
17. Kim S, You S, Hwang D. Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nature Reviews Cancer.*2011;11(10):708–18.
18. Foley KS, Young PW. The non-muscle functions of actinins: An update. *Biochemical Journal.*2014;459:1–13.
19. Xie GF, Zhao LD, Chen Q, Tang DX, Chen QY, Lu HF, et al. High ACTN1 Is Associated with Poor Prognosis, and ACTN1 Silencing Suppresses Cell Proliferation and Metastasis in Oral Squamous Cell Carcinoma. *Drug Des Devel Ther.*2020;14:1717–27.

20. Chen Q, Zhou XW, Zhang AJ, He K. ACTN1 supports tumor growth by inhibiting Hippo signaling in hepatocellular carcinoma. *J Exp Clin Cancer Res.*2021;40(1).
21. Su J, Zhou Y, Pan Z, Shi L, Yang J, Liao A, et al. Downregulation of LIMK1–ADF/cofilin by DADS inhibits the migration and invasion of colon cancer. *Scientific Reports.*2017;7(1):1–12.
22. Xu J, Lv H, Zhang B, Xu F, Zhu H, Chen B, et al. miR-30b-5p acts as a tumor suppressor microRNA in esophageal squamous cell carcinoma. *J Thorac Dis.*2019;11(7):3015–29.
23. Dai X, Liang Z, Liu L, Guo K, Xu S, Wang H. Silencing of MALAT1 inhibits migration and invasion by sponging miR-1-3p in prostate cancer cells. *Mol Med Rep.*2019;20(4):3499–508.
24. Heuser VD, Mansuri N, Mogg J, Kurki S, Repo H, Kronqvist P, et al. Formin Proteins FHOD1 and INF2 in Triple-Negative Breast Cancer: Association With Basal Markers and Functional Activities. *Breast Cancer.*2018;12.
25. Turhani D, Krapfenbauer K, Thurnher D, Langen H, Fountoulakis M. Identification of differentially expressed, tumor-associated proteins in oral squamous cell carcinoma by proteomic analysis. *Electrophoresis.* 2006;27(7):1417–23.
26. Wang L, Xiong L, Wu Z, Miao X, Liu Z, Li D, et al. Expression of UGP2 and CFL1 expression levels in benign and malignant pancreatic lesions and their clinicopathological significance. *World J Surg Oncol.*2018;16(1).
27. Li Z, Shi J, Zhang N, Zheng X, Jin Y, Wen S, et al. FSCN1 acts as a promising therapeutic target in the blockade of tumor cell motility: a review of its function, mechanism, and clinical significance. *J Cancer.*2022;13(8):2528.
28. Lii CK, Chang JW, Chen JJ, Chen HW, Liu KL, Yeh SL, et al. Docosahexaenoic acid inhibits 12-O-tetradecanoylphorbol-13- acetate-induced fascin-1-dependent breast cancer cell migration by suppressing the PKC $\delta$ - and Wnt-1/ $\beta$ -catenin-mediated pathways. *Oncotarget.* 2016;7(18):25162–79.
29. Xie L, Huang H, Zheng Z, Yang Q, Wang S, Chen Y, et al. MYO1B enhances colorectal cancer metastasis by promoting the F-actin rearrangement and focal adhesion assembly via RhoA/ROCK/FAK signaling. *Ann Transl Med.*2021;9(20):1543–1543.
30. Masoudi M, Seki M, Yazdanparast R, Yachie N, Aburatani H. A genome-scale CRISPR/Cas9 knockout screening reveals SH3D21 as a sensitizer for gemcitabine. *Sci Rep.*2019;9(1).

**Fig 1.A.** Venn Diagramm de los genes comunes entre sobreexpresados entre Leucoplasia, Cáncer Oral y Fibrosis. **1.B.** Venn Diagramm de los genes comunes infraexpresados entre Leucoplasia, Cáncer Oral, Fibrosis



**Tabla 1:** Genes comunes sobreexpresado entre cáncer oral y leucoplasia.

<b>ID</b>	<b>Nombre del Gen</b>	<b>Fold change (log2)* C vs N</b>	<b>Fold change (log2)* L vs N</b>
<b>ACLY</b>	<i>ATP Citrate Lyase</i>	1.057	0.087
<b>ACTN1</b>	<i>Actinin Alpha 1</i>	1.340	0.074
<b>ADAM19</b>	<i>ADAM Metallopeptidase Domain 19</i>	1.342	0.201
<b>ADM</b>	<i>Adrenomedullin</i>	1.708	0.263
<b>AEN</b>	<i>Apoptosis Enhancing Nuclease</i>	1.371	0.339
<b>ALYREF</b>	<i>Aly/REF Export Factor</i>	1.385	0.117
<b>ATAD3A</b>	<i>ATPase Family AAA Domain Containing 3A</i>	1.434	0.159
<b>ATF5</b>	<i>Activating Transcription Factor 5</i>	1.156	0.156
<b>BOP1</b>	<i>Ribosomal Biogenesis Factor</i>	1.362	0.159
<b>C1orf216</b>	<i>Chromosome 1 Open Reading Frame 216</i>	1.091	0.481
<b>CARD10</b>	<i>Caspase Recruitment Domain Family Member 10</i>	1.377	0.174

<b>CARS</b>	<i>Nuclear Receptor Subfamily 1 Group 1 Member 3</i>	1.024	0.062
<b>CCT3</b>	<i>Chaperonin Containing TCP1 Subunit 3</i>	1.106	0.135
<b>CCT4</b>	<i>Chaperonin Containing TCP1 Subunit 4</i>	0.449	0.042
<b>CDV3</b>	<i>CDV3 Homolog</i>	0.37	0.158
<b>CEBPB</b>	<i>CCAAT Enhancer Binding Protein Beta</i>	1.581	0.332
<b>CFL1</b>	<i>Cofilin 1</i>	0.921	0.068
<b>CHST2</b>	<i>Carbohydrate Sulfotransferase 2</i>	3.001	0.781
<b>CHST3</b>	<i>Carbohydrate Sulfotransferase 3</i>	1.069	0.351
<b>CLIC4</b>	<i>Chloride Intracellular Channel 4</i>	0.728	0.138
<b>CLSPN</b>	<i>Claspin</i>	2.011	0.203
<b>COL27A1</b>	<i>Collagen Type XXVII Alpha 1 Chain</i>	1.357	0.114
<b>COL4A1</b>	<i>Collagen Type IV Alpha 1 Chain</i>	1.386	0.02
<b>CORO1C</b>	<i>Coronin 1C</i>	0.676	0.08
<b>CPA4</b>	<i>Carboxypeptidase A4</i>	1.993	0.329
<b>CTPS1</b>	<i>CTP Synthase 1</i>	1.327	0.077
<b>CTSL</b>	<i>Cathepsin L</i>	1.453	0.222
<b>CYP19A1</b>	<i>Cytochrome P450 Family 19 Subfamily A Member 1</i>	3.076	1.204
<b>CYP24A1</b>	<i>Cytochrome P450 Family 24 Subfamily A Member 1</i>	3.486	1.103
<b>DAP3</b>	<i>Death Associated Protein 3</i>	0.508	0.085
<b>DAPK3</b>	<i>Death Associated Protein Kinase 3</i>	1.46	0.104
<b>DDX21</b>	<i>DEXD-Box Helicase 21</i>	1.452	0.116
<b>DDX23</b>	<i>DEAD-Box Helicase 23</i>	0.802	0.048
<b>DESI1</b>	<i>Desumoylating Isopeptidase 1</i>	0.838	0.147
<b>DKC1</b>	<i>Dyskerin Pseudouridine Synthase 1</i>	0.739	0.172
<b>EBNA1BP2</b>	<i>EBNA1 Binding Protein 2</i>	0.999	0.144
<b>EEF1D</b>	<i>Eukaryotic Translation Elongation Factor 1 Delta</i>	0.704	0.06
<b>EFNB1</b>	<i>Ephrin B1</i>	1.484	0.3
<b>EHD1</b>	<i>EH Domain Containing 1</i>	1.133	0.136
<b>EIF3B</b>	<i>Eukaryotic Translation Initiation Factor 3 Subunit B</i>	0.684	0.096
<b>EIF4A1</b>	<i>Eukaryotic Translation Initiation Factor 4A1</i>	0.899	0.055
<b>ELAC2</b>	<i>ElaC Ribonuclease Z 2</i>	0.685	0.061
<b>ENTPD7</b>	<i>Ectonucleoside Triphosphate Diphosphohydrolase 7</i>	1.183	0.089
<b>EPHB2</b>	<i>EPH Receptor B2</i>	1.727	0.615
<b>EPPK1</b>	<i>Epiplakin 1</i>	3.504	0.196
<b>ETV4</b>	<i>ETS Variant Transcription Factor 4</i>	1.528	0.59
<b>FBXO45</b>	<i>F-Box Protein 45</i>	2.134	0.218
<b>FMNL2</b>	<i>Formin Like 2</i>	1.37	0.084

<b>FSCN1</b>	<i>Fascin Actin-Bundling Protein 1</i>	1.996	0.224
<b>FSD1L</b>	<i>Fibronectin Type III And SPRY Domain Containing 1 Like</i>	0.831	0.328
<b>FTSJ1</b>	<i>FtsJ RNA 2'-O-Methyltransferase 1</i>	0.661	0.126
<b>GALNT18</b>	<i>Polypeptide N-Acetylgalactosaminyltransferase 18</i>	1.001	0.17
<b>GALNT6</b>	<i>Polypeptide N-Acetylgalactosaminyltransferase 6</i>	2.704	0.434
<b>GPATCH4</b>	<i>G-Patch Domain Containing 4 (Gene/Pseudogene)</i>	0.934	0.143
<b>GPR153</b>	<i>G Protein-Coupled Receptor 153</i>	1.377	0.237
<b>HARS</b>	<i>Histidyl-tRNA synthetase</i>	0.507	0.057
<b>HOMER3</b>	<i>Homer Scaffold Protein 3</i>	1.873	0.301
<b>HSP90B1</b>	<i>Heat Shock Protein 90 Beta Family Member 1</i>	0.730	0.053
<b>HSPA5</b>	<i>Heat Shock Protein Family A (Hsp70) Member 5</i>	1.242	0.068
<b>HSPH1</b>	<i>Heat Shock Protein Family H (Hsp110) Member 1</i>	1.199	0.127
<b>HYOU1</b>	<i>Hypoxia Up-Regulated 1</i>	0.682	0.106
<b>INF2</b>	<i>Inverted Formin 2</i>	0.734	0.115
<b>ITGA3</b>	<i>Integrin Subunit Alpha 3</i>	1.710	0.092
<b>KIAA0930</b>	<i>KIAA0930</i>	0.740	0.223
<b>KLF7</b>	<i>KLF Transcription Factor 7</i>	1.114	0.193
<b>KRT17</b>	<i>Keratin 17</i>	5.685	0.683
<b>LAMC2</b>	<i>Laminin Subunit Gamma 2</i>	3.083	0.101
<b>LIMK1</b>	<i>LIM Domain Kinase 1</i>	1.278	0.225
<b>LMNB2</b>	<i>Lamin B2</i>	1.505	0.111
<b>LRP8</b>	<i>LDL Receptor Related Protein 8</i>	1.381	0.338
<b>LRRC59</b>	<i>Leucine Rich Repeat Containing 59</i>	1.000	0.171
<b>LY6K</b>	<i>Lymphocyte Antigen 6 Family Member K</i>	2.447	0.696
<b>LYAR</b>	<i>Ly1 Antibody Reactive</i>	1.209	0.188
<b>MDFI</b>	<i>MyoD Family Inhibitor</i>	1.698	0.203
<b>MET</b>	<i>MET Proto-Oncogene, Receptor Tyrosine Kinase</i>	1.051	0.144
<b>MICAL2</b>	<i>Microtubule Associated Monooxygenase, Calponin And LIM Domain Containing 2</i>	1.748	0.085
<b>MRPL14</b>	<i>Mitochondrial Ribosomal Protein L14</i>	1.042	0.156
<b>MSN</b>	<i>Moesin</i>	0.726	0.228
<b>MTHFD1L</b>	<i>Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1 Like</i>	1.549	0.175
<b>MTHFD2</b>	<i>Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 2, Methenyltetrahydrofolate Cyclohydrolase</i>	1.017	0.156
<b>MYBL2</b>	<i>MYB Proto-Oncogene Like 2</i>	1.55	0.13
<b>MYO1B</b>	<i>Myosin IB</i>	1.978	0.233



<b>NACC1</b>	<i>Nucleus Accumbens Associated 1</i>	0.933	0.135
<b>NADK</b>	<i>NAD Kinase</i>	0.79	0.117
<b>NCF2</b>	<i>Neutrophil Cytosolic Factor 2</i>	1.123	0.385
<b>NKRF</b>	<i>NFKB Repressing Factor</i>	0.729	0.185
<b>NOC2L</b>	<i>NOC2 Like Nucleolar Associated Transcriptional Repressor</i>	0.907	0.052
<b>NOMO1</b>	<i>NODAL Modulator 1</i>	0.881	0.082
<b>NOP56</b>	<i>NOP56 Ribonucleoprotein</i>	0.964	0.079
<b>NPLOC4</b>	<i>NPL4 Homolog, Ubiquitin Recognition Factor</i>	0.080	0.091
<b>NRIP1</b>	<i>Nuclear Receptor Interacting Protein 2</i>	1.224	0.088
<b>NSUN2</b>	<i>NOP2/Sun RNA Methyltransferase 2</i>	0.944	0.063
<b>NUP62</b>	<i>Nucleoporin 62</i>	0.587	0.076
<b>OAS3</b>	<i>2'-5'-Oligoadenylate Synthetase 3</i>	1.877	0.094
<b>PA2G4</b>	<i>Proliferation-Associated 2G4</i>	0.869	0.111
<b>PDCD11</b>	<i>Programmed Cell Death 11</i>	1.149	0.068
<b>PFDN2</b>	<i>Prefoldin Subunit 2</i>	1.101	0.359
<b>PLAU</b>	<i>Plasminogen Activator, Urokinase</i>	2.666	0.096
<b>PLCD3</b>	<i>Phospholipase C Delta 3</i>	1.228	0.085
<b>PLEC</b>	<i>Plectin</i>	1.528	0.068
<b>PLOD3</b>	<i>Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 3</i>	0.831	0.194
<b>PMEPA1</b>	<i>Prostate Transmembrane Protein, Androgen Induced 1</i>	1.725	0.238
<b>PNLIPRP3</b>	<i>Pancreatic Lipase Related Protein 3</i>	4.07	2.771
<b>PNP</b>	<i>Purine Nucleoside Phosphorylase</i>	1.132	0.109
<b>PPP1R14C</b>	<i>Protein Phosphatase 1 Regulatory Inhibitor Subunit 14C</i>	1.805	0.239
<b>PPP2R5B</b>	<i>Protein Phosphatase 2 Regulatory Subunit B'Beta</i>	1.053	0.154
<b>PRNP</b>	<i>Prion Protein</i>	1.164	0.253
<b>PSMB7</b>	<i>Proteasome 26S Subunit, Non-ATPase 7</i>	0.822	0.71
<b>PSMC3</b>	<i>Proteasome 26S Subunit, ATPase 3</i>	1.387	0.102
<b>PSMD2</b>	<i>Proteasome 26S Subunit Ubiquitin Receptor, Non-ATPase 2</i>	1.095	0.98
<b>PSMD3</b>	<i>Proteasome 26S Subunit, Non-ATPase 3</i>	0.607	0.103
<b>PTH1H</b>	<i>Parathyroid Hormone Like Hormone</i>	2.367	0.539
<b>PTPRZ1</b>	<i>Protein Tyrosine Phosphatase Receptor Type Z1</i>	2.561	0.062
<b>PTTG1IP</b>	<i>PTTG1 Interacting Protein</i>	0.728	0.129
<b>PUS1</b>	<i>Pseudouridine Synthase 1</i>	0.954	0.171
<b>RAD23A</b>	<i>RAD23 Homolog A, Nucleotide Excision Repair Protein</i>	0.834	0.157
<b>RGS20</b>	<i>Regulator Of G Protein Signaling 20</i>	2.397	0.654
<b>RRAS2</b>	<i>RAS Related 2</i>	0.835	0.171

<b>RRS1</b>	<i>Ribosome Biogenesis Regulator 1 Homolog</i>	1.374	0.413
<b>SDF4</b>	<i>Stromal Cell Derived Factor 4</i>	0.955	0.065
<b>SDK2</b>	<i>Sidekick Cell Adhesion Molecule 2</i>	1.792	0.564
<b>SEMA3C</b>	<i>Semaphorin 3C</i>	1.789	0.337
<b>SERPINB7</b>	<i>Serpin Family B Member 7</i>	3.093	0.737
<b>SH2D2A</b>	<i>SH2 Domain Containing 2A</i>	1.402	0.301
<b>SH3D21</b>	<i>SH3 Domain Containing 21</i>	1.216	0.112
<b>SLC15A1</b>	<i>Solute Carrier Family 15 Member 1</i>	2.131	0.258
<b>SLC16A1</b>	<i>Transmembrane Channel Like 7</i>	1.861	0.358
<b>SLC28A3</b>	<i>Solute Carrier Family 28 Member 3</i>	3.059	0.433
<b>SLC3A2</b>	<i>Solute Carrier Family 3 Member 2</i>	1.411	0.08
<b>SMOX</b>	<i>Spermine Oxidase</i>	2.494	0.339
<b>SQLE</b>	<i>Squalene Epoxidase</i>	1.701	0.138
<b>STIP1</b>	<i>Stress Induced Phosphoprotein 1</i>	1.068	0.126
<b>STRIP2</b>	<i>Striatin Interacting Protein 2</i>	2.287	0.633
<b>STX1A</b>	<i>Syntaxin 1A</i>	1.537	0.613
<b>SZRD1</b>	<i>SUZ RNA Binding Domain Containing 1</i>	0.438	0.138
<b>TGM2</b>	<i>Transglutaminase 2</i>	2.319	0.414
<b>THOC5</b>	<i>THO Complex Subunit 5</i>	0.468	0.066
<b>TIMM17A</b>	<i>Translocase Of Inner Mitochondrial Membrane 17A</i>	0.781	0.143
<b>TMC7</b>	<i>Transmembrane Channel Like 7</i>	1.244	0.602
<b>TMEM184B</b>	<i>Transmembrane Protein 184B</i>	1.027	0.17
<b>TNC</b>	<i>Tenascin C</i>	2.178	0.214
<b>TOP1MT</b>	<i>DNA Topoisomerase I Mitochondrial</i>	1.18	0.155
<b>TRMT6</b>	<i>TRNA Methyltransferase 6 Non-Catalytic Subunit</i>	0.966	0.108
<b>TRPV3</b>	<i>Transient Receptor Potential Cation Channel Subfamily V Member 3</i>	2.871	0.366
<b>TUBB3</b>	<i>Tubulin Beta 3 Class III</i>	2.060	0.614
<b>TUBG1</b>	<i>Tubulin Gamma 1</i>	1.067	0.207
<b>UHRF1</b>	<i>Ubiquitin Like With PHD And Ring Finger Domains 1</i>	1.197	0.104
<b>USB1</b>	<i>U6 SnRNA Biogenesis Phosphodiesterase 1</i>	1.561	0.175
<b>WDR5</b>	<i>WD Repeat Domain 5</i>	0.841	0.056
<b>WDR53</b>	<i>WD Repeat Domain 53</i>	1.154	0.147
<b>WDR62</b>	<i>WD Repeat Domain 62</i>	0.998	0.284
<b>WDR66</b>	<i>WD Repeat Domain 66</i>	1.971	0.665
<b>YWHAG</b>	<i>Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Gamma</i>	0.679	0.195
<b>ZNF622</b>	<i>Zinc Finger Protein 622</i>	0.544	0.073

\*The *Fold-change* estimados (cambios de expresión relativa de los genes).

**Tabla 2:** Genes comunes sobreexpresado entre pacientes con cáncer y fibrosis

<b>ID</b>	<b>Nombre gen</b>	<b>Fold change (log2)* C vs N</b>	<b>Fold change (log2)* F vs N</b>
<b>RP11- 23P13.6</b>	<i>Long-non-coding</i>	1.291	1.366
<b>STARD 4</b>	<i>StAR Related Lipid Transfer Domain Containing 4</i>	1.292	1.150
<b>RP11- 420K8.1</b>	<i>Long-non-coding RNA</i>	1.293	1.538
<b>NUS1</b>	<i>NUS1 Dehydrodolichyl Diphosphate Synthase Subunit</i>	1.294	0.614

**Tabla 3:** Gen común sobreexpresados entre las 3 patologías

<b>ID</b>	<b>Nombre gen</b>	<b>Fold change (log2)* C vs N</b>	<b>Fold change (log2)* F vs N</b>	<b>Fold change (log2)* L vs N</b>
<b>EPRS1</b>	<i>Glutamyl-prolyl-tRNA synthetase</i>	0.653	0.492	0.091