



**Universidad
Europea** VALENCIA

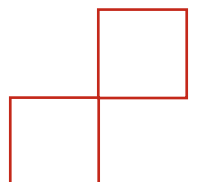
Grado en ODONTOLOGÍA

Trabajo Fin de Grado

Curso 2021-22

**Análisis transcriptómico en la cavidad
bucal para el cáncer oral en humanos
mediante *microarrays***

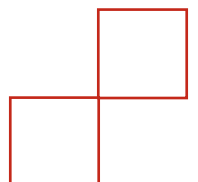
**Presentado por: Gaia Gasparroni
Tutor: Marta Serna Garcia**





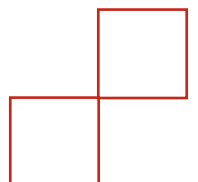
INDICES DE CONTENIDOS

1. LISTADO DE SÍMBOLOS Y SIGLAS	4
2. RESUMEN	4
3. PALABRAS CLAVE	5
4. INTRODUCCIÓN	5
4.1 Cáncer oral	5
4.2 Epidemiología cáncer oral en el mundo	6
4.3 Factores de riesgo del cáncer oral	6
4.4 Definición de genómica	7
4.5 El transcriptoma y su análisis	7
4.6 Que son los microarrays o biochips de expresión	8
4.7 Tipos de biochips de expresión	9
4.8 Microarrays de expresión affymetrix genechip®	9
4.9 Microarray genechip™ human genome u133a 2.0 array	11
5. JUSTIFICACIÓN, HIPÓTESIS Y OBJETIVOS	11
5.1 Hipótesis	11
5.2 Objetivo General	12
5.3 Objetivos Específicos	12
6. MATERIALE Y METODOS	12
6.1 Análisis transcriptómica	12
6.1.1 Diseño del estudio	12
6.1.2 Análisis de expresión diferencial	13
6.1.3 Análisis de enriquecimiento	13
7. RESULTADOS	14
7.1. Análisis exploratorio de los datos	14





7.2 Análisis diferencial.....	15
7.3 Genes diferenciales comunes entre la cohorte 1 y cohorte 2....	16
7.4 Análisis funcional.....	18
8. DISCUSIÓN.....	22
9. CONCLUSIÒN.....	25
10. BIBLIOGRAFIA.....	26





1. LISTADO DE SÍMBOLOS Y SIGLAS

SAM: Análisis de Significación de Microarrays

RP: Análisis de Producto de Rango

COCE: Carcinoma oral de células escamosas primario

GEO: Gene Expresión Omnibus

TAC: *Transcriptome Analysis Console*

KEGG: Enciclopedia de Genes y Genomas de Kyoto

PCR: Reacción en Cadena de la Polimersa

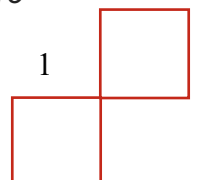
PCA: Análisis de componentes principales

RMA: *Robust Multichip Analysis*

DAVID: *Database for Annotation, Visualization and Integrated Discovery*

2. RESUMEN

El cáncer oral es el más frecuente de los cánceres que de cabeza y cuello y es una enfermedad multifactorial causada por alteraciones en los genes mediante una amplia gama de mecanismos. Las herramientas como los *microarrays* y su posterior análisis bioinformático nos permiten interpretar y analizar los datos obtenidos de una secuenciación de RNA de manera robusta. El objetivo de nuestro estudio fue estudiar el tejido de la cavidad oral en humanos mediante la técnica de *microarrays*, con la finalidad de identificar los procesos biológicos y genes de interés para en un futuro usar como biomarcadores. Fueron comparados dos grupos divergentes: cáncer (C) *versus* control (N) para dos cohortes diferentes (cohorte 1 y cohorte 2) en la cavidad oral en humanos. Los datos brutos de las dos cohortes (.cel) fueron obtenidos de una base de datos de acceso público (*ArrayExpress*). Se realizó un análisis diferencial mediante el *software* TAC (*Transcriptome Analysis Console*) considerando $p \text{ value} \leq 0,05$. El análisis de enriquecimiento de los genes diferenciales encontrados fue mediante el *software* DAVID (*Database for Annotation, Visualization and Integrated Discovery*) y la visualización de los procesos biológicos de interés y genes relevantes mediante el *software*





Cytoscape. Fueron detectados 4.083 genes diferenciales para la cohorte 1 y 5.330 para la cohorte 2 siendo 2.072 los genes comunes para las dos cohortes. Se realizó un filtro de *fold change* de $|5|$ siendo 31 genes sobreexpresados y 72 infraexpresados para el grupo de cáncer oral. Destacamos los genes: SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 todos sobreexpresados en cáncer. Se ha visto como proceso biológico más significativo es “extracelular matrix organization” que está involucrado la familia de MMP. Estos genes podrían ser futuros biomarcadores tempranos para diagnosticar el cáncer oral y pudiendo considerarse como dianas terapéuticas.

3. PALABRAS CLAVE

Cancer oral, microarrays, expresión diferencial.

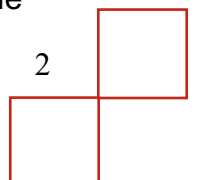
4. INTRODUCCIÒN

4.1 Cáncer oral

El cáncer oral es el más frecuente de los cánceres que de cabeza y cuello. El cáncer de cavidad oral se encuentra dentro de las 10 neoplasias malignas más frecuentes en nuestro país y se observa un aumento de los casos diagnosticados de cáncer oral en los últimos 10 años¹.

Cuando se encuentran en fases tempranas de desarrollo, los cánceres orales tienen una tasa de supervivencia del 80 al 90%. Desafortunadamente la mayoría se encuentran como cánceres de etapa tardía y la tasa de mortalidad es muy alta de aproximadamente 43% a los cinco años del diagnóstico¹. La mayoría de estos cánceres son difíciles de descubrir y esto es debido a la falta de conciencia pública junto con la falta de un programa nacional de detección precoz de cáncer oral, llevado a cabo por profesionales médicos y odontólogos¹.

Así mismo una deficiencia en el conocimiento de esta enfermedad en los médicos y odontólogos también contribuiría en el retraso del tratamiento de





los pacientes que la padecen. De esta manera, se resalta la necesidad del conocimiento de su diagnóstico dentro de los conocimientos básicos del odontólogo general¹.

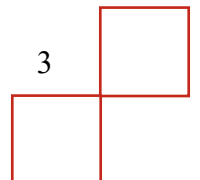
4.2 Epidemiología cáncer oral en el mundo

El cáncer bucal representa la undécima neoplasia más incidente en el mundo, con más de 202.000 con una relación hombre: mujer de 2:1. En el caso particular de España, el cáncer bucal ocupa el puesto 16 entre todas las neoplasias por incidencia y el puesto 19 por mortalidad². En cuanto a la edad, el 90% de los cánceres orales se diagnostican en mayores de 40 años, y más del 50% en individuos de más de 65 años³. Recientemente se ha observado un aumento en la incidencia de esta enfermedad en menores de 40 años lo que podría estar asociado a una infección por el virus del papiloma humano. Las principales localizaciones son lengua y el paladar de boca en Europa y el norte de América y la mucosa bucal en la India. En referencia a la histología, el 90% de los carcinomas orales son del tipo carcinoma oral de células escamosas. La supervivencia a los 5 años es del 40-56% aunque varía según el tipo de cáncer: 70% cáncer labial, 30% otras localizaciones orales³.

4.3 Factores de riesgo del cáncer oral

El cáncer es una enfermedad multifactorial causada por alteraciones en los genes mediante una amplia gama de mecanismos que involucran también factores no genéticos. En la carcinogénesis son relevantes, también, algunos factores como las radiaciones, las infecciones, la dieta, la irritación crónica, la exposición excesiva a la luz solar y los estados de inmunodepresión, el hábito de fumar y el consumo de alcohol⁴.

El desarrollo del cáncer bucal es un proceso en múltiples etapas que implica la acumulación de alteraciones genéticas y epigenéticas en genes regulatorios clave. Los mecanismos genéticos básicos en el cáncer son la sobreexpresión de oncogenes y el silenciamiento de genes supresores





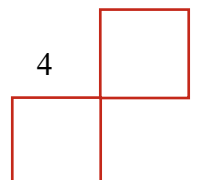
tumorales. Las variaciones individuales en la susceptibilidad al carcinoma de células escamosas bucal se atribuyen a complejas interacciones entre factores genéticos y ambientales, aunque los mecanismos convergen en procesos inflamatorios. La inflamación se relaciona con alteraciones de la expresión de oncogenes y genes supresores tumorales⁵.

4.4 Definición de genómica

La genómica es el área de la biología que investiga cómo se constituye y funciona un genoma. El campo al que pertenece la genómica es la biología molecular, una rama de la biología que se centra en los procesos que tienen lugar en los organismos vivos observándolos desde una perspectiva molecular. En otros términos, su meta es dar explicación a los fenómenos de los seres vivos partiendo de sus propiedades macromoleculares. Las tareas de la genómica son caracterizar y cuantificar los genes. Entonces la genómica se centra en la secuenciación de ADN y RNA para examinar y ensamblar la estructura de un genoma completo y para poder garantizar también su función. Hoy en día se han secuenciado pocos genomas, por eso la genómica tiene un gran futuro por delante. Este conocimiento es un arma a doble hilo. Por un lado, permite de comprender nuestro genoma, su función y su evolución y se podrá usar en el ámbito de la salud para potenciar los tratamientos en pacientes de todas las edades. Por otro lado, podría evidenciar la presencia de ciertos trastornos y enfermedades⁶.

4.5 El transcriptoma y su análisis

El genoma humano está compuesto de ADN que es una molécula larga que contiene las instrucciones necesarias para producir y mantener las células. Para que las instrucciones puedan llevarse a la práctica, el ADN debe "leerse", transcribirse y copiarse para crear ARN. Estas "lecturas" de genes se llaman transcritos. Entonces podemos decir que un transcriptoma es una colección de todas las lecturas de genes presentes en una célula. Los transcritos pudieran



servir para influir en la estructura celular y regular los genes, estableciendo así el Dogma Central de la Biología Molecular ⁷(Figura 1).

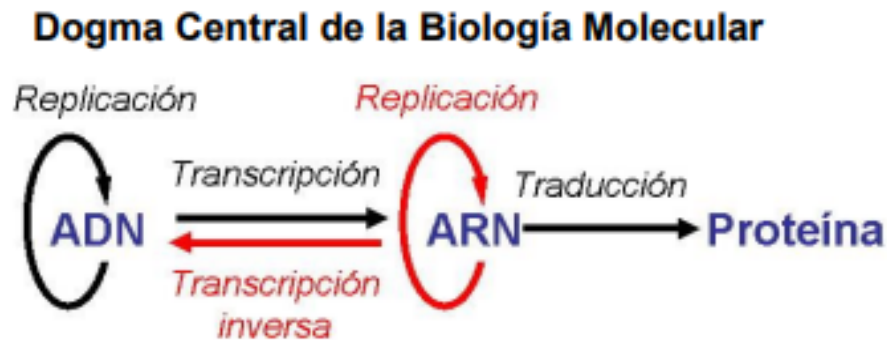
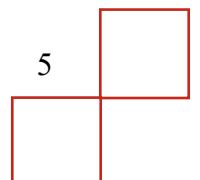


Figura 1. Dogma Central de la Biología Molecular. Modificado de la propuesta inicial de Crick (1970)¹².

Definimos una secuencia de ARN como un reflejo de la secuencia del ADN de la que fue transcrito y por estos investigadores que pueden determinar cuándo y dónde está activado o desactivado cada gen en las células y los tejidos de un organismo⁷.

4.6 *Microarrays* o biochips de expresión

Un *microarray*, por definición, consiste en un gran número de moléculas de ADN ordenadas sobre un sustrato sólido de manera que formen una matriz de secuencias en dos dimensiones. Por esto los oligonucleótidos es la denominación que se da a estos fragmentos de material genético pueden ser secuencias cortas o de mayor tamaño, cDNA (ADN complementario, sintetizado a partir de mRNA), o bien productos de PCR. La denominación de sonda, entonces, deriva de estos fragmentos de ADN de una sola hebra inmovilizados en el soporte⁷. Los ácidos nucleicos de las muestras a analizar se marcan por diversos métodos, como: métodos enzimáticos, métodos fluorescentes, etc. Sucesivamente se incuban sobre el panel de sondas, permitiendo la hibridación de secuencias homologas. El término hibridación quiere definir el reconocimiento y unión entre moléculas complementarias y su homologación. Durante la hibridación, las muestras de material genético marcadas se unirán a sus complementarias inmovilizadas en el soporte del chip, permitiendo la identificación y cuantificación del ADN presente en la muestra. Con



posterioridad, el escáner y las herramientas informáticas nos permiten interpretar y analizar los datos obtenidos⁹.

4.7 Tipos de biochips de expresión

Son disponibles tres diferentes tipos de biochips, que son: *microarrays* de ADN, microchips de microfluidos y *microarrays* de proteínas⁹. El *microarray* de ADN o biochip de ADN es constituido por un conjunto de pequeños trozos de ADN fijadas sobre una superficie resistente. Es utilizado para calcular los niveles de expresión de una gran cantidad de genes. Cada marca de ADN comprende determinados genes particulares que se denominan sondas. La hibridación sonda-diana se obtiene mediante el reconocimiento de dianas marcadas para decidir la cantidad relativa de series de ácidos nucleicos en la diana. De otra parte, los biochips de microfluidos son una opción para los laboratorios bioquímicos habituales. Estos chips se están volviendo más complejos al usar miles de componentes, pero esos componentes se diseñan físicamente y se denominan plan de abajo hacia arriba totalmente personalizado. Por último, lo *microarrays* de proteínas o chips de proteínas son utilizados para seguir las acciones y las conexiones de las proteínas, y para conocer su función a gran escala. Estos son automatizados, rápidos, económicos, muy sensibles, consumen menos cantidad de muestras. Este tipo de tecnología se ha convertido en los *microarrays* más utilizados⁹.

4.8 *Microarrays* de expresión *affymetrix genechip*[®]

Los *GeneChip* de *Affymetrix* son *microarrays* de expresión de *Affymetrix* que se llaman *GeneChip*[®] (Figura 2). Para poder determinar el nivel de expresión de un determinado gene estos tipos de *microarrays* utilizan la atracción química que se ha de forma natural entre el DNA y las moléculas de DNA de cadena única o RNA⁹. Es importante decir, entonces, que tiene que existir una perfecta complementariedad entre dos cadenas de nucleótidos para que ambas las cadenas se mantendrán unidas. Entonces para poder medir la función de cada gen los *microarrays* tienen por primero sintetizar la cadena de DNA

denominada sonda en la superficie de un chip de cuarzo. Estas cadenas, de 25 pb serán específica de un determinado gen específico del genoma. La superficie de un *microarrays* de *Affymetrix* es de aproximadamente 1 cm por 1 cm. Dicha superficie está dividida formando una cuadrícula donde cada partición tiene unas dimensiones de 5 μm por 5 μm y contiene millones de réplicas de un solo tipo de sonda. Por otra parte, pueden llegar a contener hasta aproximadamente 6,5 millones de sondas diferentes representando todo el genoma¹⁰.

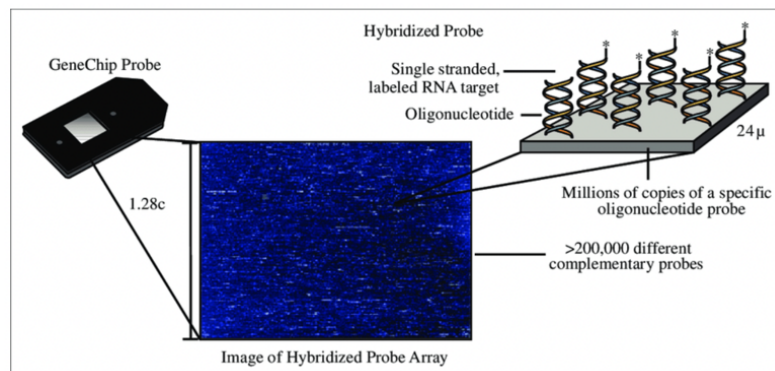


Figura 2. Superficie de un biochip. Imagen de *Affymetrix*¹³.

La fabricación de estos tipos de *microarrays* se hace mediante la técnica de fotolitografía. El proceso de fabricación utiliza una máscara para poder exponer selectivamente la luz sobre una oblea de silicio, esta luz ultravioleta, va eliminando los grupos protectores fotolábiles en posiciones definidas. Las zonas desprotegidas son donde se van a ir añadiendo las bases de interés. De modo tal que aquellas cadenas no “desprotegidas” no serán capaces de recibir un nuevo nucleótido. Este tipo de proceso es muy meticuloso y preciso porque es importante que los nucleótidos sean añadidos en el orden correcto para formar la sonda correcta. Una vez diseñado el chip, se lleva a cabo la extracción del RNA total de la muestra biológica y que depende del protocolo que especifique el biochip *GeneChip*[®] utilizado. En seguida se realizará la hibridación de la muestra preparada que se pondrá en el *microarray*. Durante este proceso las cadenas de RNA o sDNA de la muestra intentan buscar sondas complementarias a lo largo del *microarray*. Las hibridaciones no específicas que se crearan durante la hibridación se eliminarán mediante el lavado del chip. El último paso del procedimiento es el escaneo y lectura del *microarray*¹⁰.

4.9 Microarray genechip™ human genome u133a 2.0 array

Los nuevos *microarray* que se encuentran en comercio por *Affymetrix* nos dan una cobertura global. La principal ventaja residía en la alta densidad de integración de material biológico que se consigue inmovilizar entonces analizar simultáneamente miles de genes. Esta tecnología actualmente se está aplicando en la detección de mutaciones y polimorfismos, secuenciación, seguimiento de terapia, medicina preventiva, screening y toxicología de fármacos, y diagnóstico molecular. Un fragmento de ADN con una utilidad determinada, como puede ser una sonda para el diagnóstico de una enfermedad específica, es potencialmente patentable siempre que cumpla los tres criterios fundamentales de una patente: novedad, invención y aplicación industrial.

La matriz *GeneChip™ Human Genome U133A 2.0* es una matriz única que se puede utilizar para explorar la biología humana y los procesos de enfermedades¹¹. Proporciona cobertura de genes bien fundamentados en el genoma humano transcrito en una única matriz. Analiza el nivel de expresión de 18.400 transcripciones y variantes, incluidos 14.500 genes humanos bien caracterizados. Es compuesto por más de 22,000 conjuntos de sondas y 500,000 características distintas de oligonucleótidos y por último usa el poder del conjunto de sondas y obtiene múltiples mediciones independientes para cada transcripción que brinden la mayor precisión y reproducibilidad de cualquier plataforma de *microarrays*¹¹.

5. JUSTIFICACIÓN, HIPÓTESIS Y OBJETIVOS

5.1 Hipótesis

Determinar la existencia de un cambio significativo en la expresión de los genes en muestras con cáncer respecto a muestras normales comparando los genes de la cohorte 1 con los de la cohorte 2.



5.2 Objetivo General

Estudiar el tejido de la cavidad oral en humanos mediante la técnica de *microarrays* para dos cohortes diferentes clasificados en dos grupos de muestras divergentes (cáncer oral vs la mucosa "normal") para identificar los procesos biológicos y genes de interés para en un futuro usar como biomarcadores.

5.3 Objetivos Específicos

Identificar genes diferenciales en tejido de la cavidad oral en humanos comunes para ambas cohortes asociados a cáncer oral.

Efectuar un análisis de enriquecimiento para identificar procesos biológicos significativos de interés para el cáncer oral.

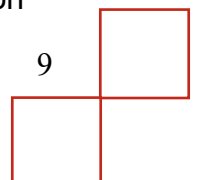
6. MATERIALES Y MÉTODOS

6.1 Análisis transcriptómica

6.1.1 Diseño del estudio

Los datos brutos (.cel) fueron elegidos de *ArrayExpress*, una base de datos de acceso público. El número de acceso a la base de datos de *ArrayExpress* es E-GEOD-42743 (cohorte 1) y E-GEOD-30784 (cohorte 2). La búsqueda fue realizada aplicando un filtro del organismo; *homo sapiens sapiens* y mediante la palabra clave "cáncer oral". El resultado de nuestra búsqueda ha sido de 219 experimentos. Siendo elegidos 2, descartando los experimentos con .cel corruptos.

Para la cohorte 1, se realizó un análisis comparativo de la expresión génica utilizando la plataforma Affymetrix mediante el *GeneChip™ Human Genome U133A 2.0* para comparar las diferencias entre las muestras de cáncer oral y la mucosa "normal". Los controles fueron 29 casos, mientras que los casos fueron 74, de total de 103 muestras. Para la cohorte 2, los casos fueron





pacientes con COCE primario (carcinoma oral de células escamosas) de la mucosa oral. Los controles fueron pacientes que se sometieron a amigdalectomía o cirugía oral para el tratamiento de enfermedades distintas del cáncer. Se obtuvieron un total de 229 muestras, los controles fueron 45 casos, mientras que los casos fueron 167. Se realizó también aquí un análisis comparativo de la expresión génica utilizando la plataforma *Affymetrix* mediante el *GeneChip™ Human Genome U133A 2.0* para comparar las diferencias entre las muestras de casos y control.

6.1.2 Análisis de expresión diferencial

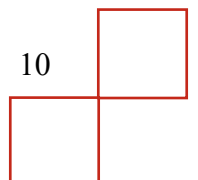
Todos los archivos brutos *.cel* para la cohorte 1 and cohorte 2 fueron comparados usando dos grupos: cáncer oral (C) y el control (N), analizándose todas las muestras con el mismo flujo de trabajo.

El software TAC (*Transcriptome Analysis Console*) (*Affymetrix*) realizó el análisis diferencial entre los grupos C vs N para las dos cohortes diferentes. Los datos sin procesar se preprocesaron, incluida la corrección de fondo, la normalización mediante un análisis robusto de promedio de múltiples matrices (RMA, *Robust Multichip Analysis*). El análisis ANOVA unidireccional se realizó para identificar genes expresados diferencialmente. Un $p \text{ value} \leq 0,05$ se consideró estadísticamente significativo.

También se realizó un análisis de componentes principales (PCA) para determinar la variabilidad en los conjuntos de datos. PCA reduce la complejidad de los datos de alta dimensión y los simplifica, identificando patrones de expresión y fuentes de variabilidad en un gran conjunto de datos de forma tridimensional¹⁴. La distancia entre cualquier par de puntos está relacionada con la similitud entre dos muestras en un espacio de alta dimensión, en este caso cada variable correspondiente a un archivo *.cel* unidimensional.

6.1.3 Análisis de enriquecimiento

Los 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) para mapear las rutas enriquecidas de genes usando la base de datos de la Enciclopedia de Genes y Genomas de Kyoto (KEGG), se determinó $p\text{ value} \leq 0,05$ 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 funcionalmente¹⁵.

7. RESULTADOS

7.1. Análisis exploratorio de los datos

El análisis de componentes principales (ACP) no supervisado permite observar la distribución en el espacio de las muestras de tejido de la cavidad oral en humanos en cuenta las tres componentes principales. En la cohorte 1 explican el 45,6 % de la variabilidad de las muestras (Figura 3.a, b). En el análisis ACP de las 103 muestras analizadas, 74 muestras pertenecientes el grupo con cáncer (C) y 29 muestras pertenecientes al grupo normal (N), se observa que las muestras se diferencian desde el punto de vista transcriptómico.

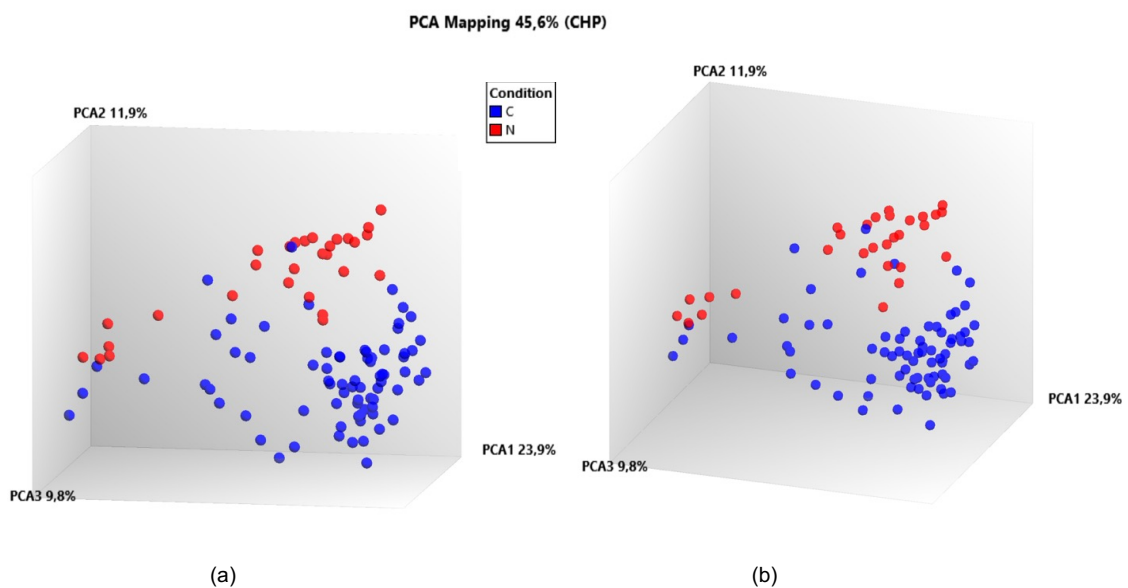


Figura 3. En las figuras se observan imágenes del ACP de la cohorte 1, pero en diferentes planos tridimensionales: (a) Esta figura nos da una visión del ACP más frontal. (b) Visión del ACP ligeramente lateralizada¹⁶.

El ACP de la cohorte 2 permite observar la distribución del el 35,6 % de la variabilidad de las muestras (Figura 4). En el análisis ACP de las 212 muestras analizadas (167 las muestras pertenecientes a las hembras de la población cáncer (C) y 45 muestras pertenecientes a las hembras de la población no cáncer (N) se observa que las muestras se diferencian desde el punto de vista transcriptómico.

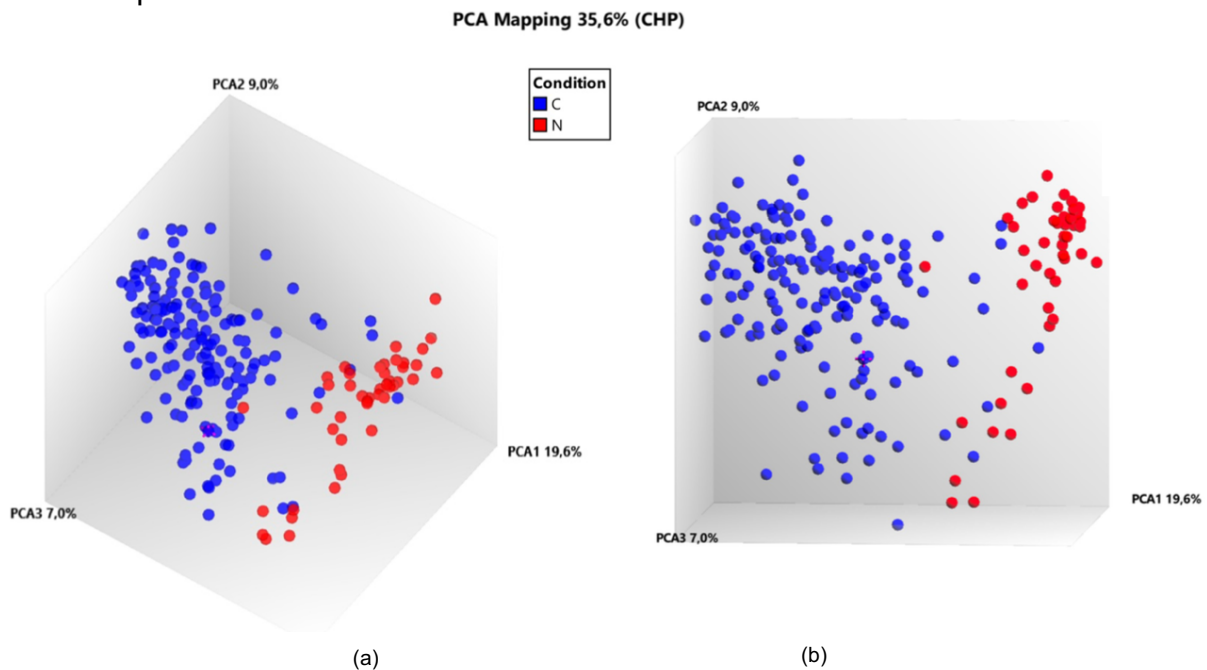
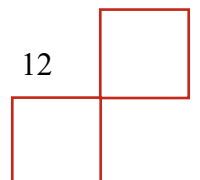


Figura 4. En las figuras se observan imágenes del ACP de la cohorte 2, pero en diferentes planos tridimensionales: (a) Esta figura nos da una visión del ACP desde arriba. (b) Visión del ACP lateralizada¹⁶.

7.2 Análisis diferencial

Se realizó el análisis de la varianza (ANOVA) de la expresión de los genes que se expresaban de manera diferencial en cohorte 1 y la cohorte 2 cada uno dividido en dos grupos, cáncer vs normal mediante el uso del software TAC. El número de genes diferencialmente expresados obtenidos para la cohorte 1 ha sido de 4.083 y 5.330 genes diferenciales para la cohorte 2 ($p \text{ value} \leq 0,05$).



7.3 Genes diferenciales comunes entre la cohorte 1 y cohorte 2

Hemos encontrado 103 (Tabla 1, Figura 5) genes comunes entre las dos cohortes realizando un filtro por cambio de expresión (*fold change*) $|5|$ y despreciados aquellos genes que no tienen una anotación funcional hasta el momento. Se observó que 31 genes tienen unos valores de cambio de expresión positivos (sobreexpresados) (Tabla 1) y 72 genes con valores de cambio de expresión negativos (infraexpresados) para el grupo de cáncer oral.

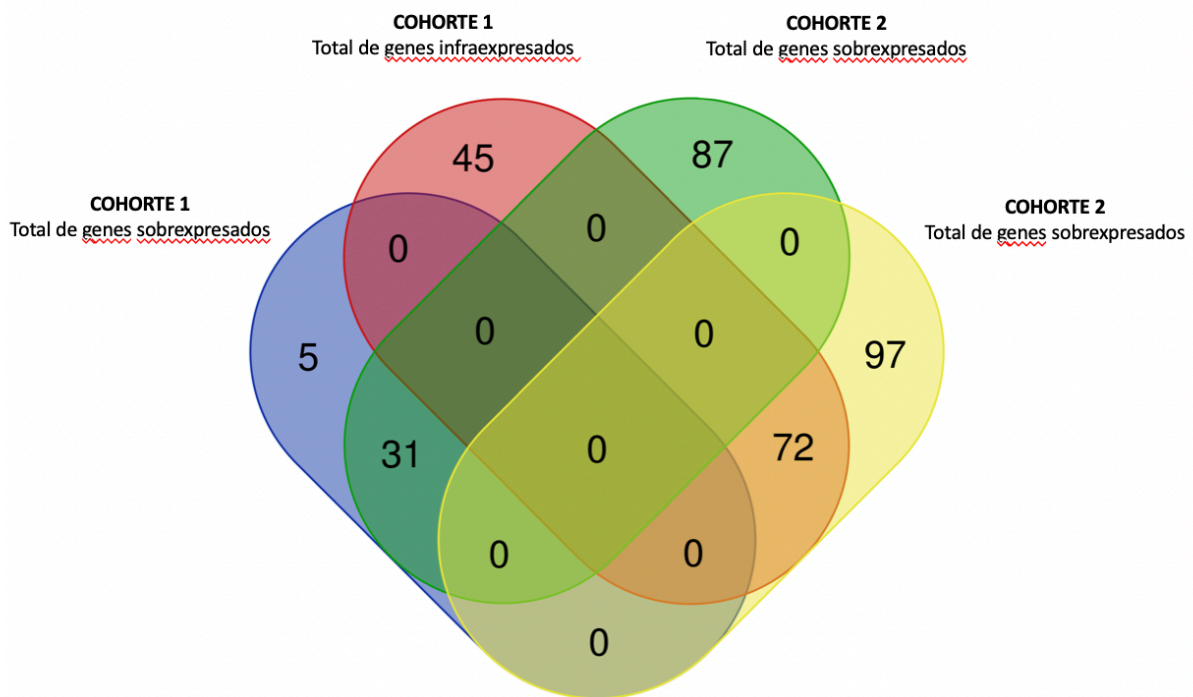
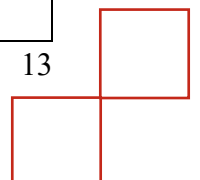
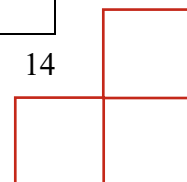


Figura 5. Venn diagram entre genes sobreexpresados y genes infraexpresados entre cohorte 1 y cohorte 2 realizando el filtro por *fold change* $|5|$. Fueron 42 genes sobreexpresados y 138 genes infraexpresados para la cohorte 1 y 159 genes sobreexpresados y 205 genes infraexpresados para la cohorte 2. Los datos obtenidos han sido 31 genes comunes sobreexpresados y 72 genes comunes infraexpresados entre las dos cohortes¹⁷.

ALIAS	NOMBRE GENES	FC* COHORTE 1	FC* COHORTE 2
CDH3	Caderina-3	5,07	9,18
ADAM12	ADAM Metallopeptidase Domain 12	4,42	4,13
BST2	Bone Marrow Stromal Cell Antigen 2	5,2	6,99



COL11A1	Collagen, type XI, alpha 1	3,53	12,57
COL12A1	Collagen, type XI, alpha 1	2,39	8,13
CTHRC1	Collagen Triple Helix Repeat Containing 1	8,34	8,34
CXCL10	C-X-C Motif Chemokine Ligand 10	6,22	20
CXCL11	C-X-C Motif Chemokine Ligand 11	11,88	10,62
CXCL13	C-X-C Motif Chemokine Ligand 13	7,39	54,81
CXCL8	C-X-C Motif Chemokine Ligand 8	5,64	40,55
HMGA2	High Mobility Group AT-Hook 2	5,17	7,21
IFI27	interferon, alpha-inducible protein 27	5,92	6,88
INHBA	Inhibin Subunit Beta A	32,55	1,6
ISG15	ISG15 Ubiquitin Like Modifier	9,49	11,61
KRT17	Keratin 17	3,35	24,43
LAMC2	Laminin Subunit Gamma 2	2,91	17,99
MMP1	Matrix Metallopeptidase 1	8,34	358,93
MMP10	Matrix Metallopeptidase 10	31,84	36,29
MMP12	Matrix Metallopeptidase 12	20,96	109,44
MMP13	Matrix metallopeptidase 13	8,34	8,34
MMP3	Matrix Metallopeptidase 3	21,72	72,22
MMP9	Matrix Metallopeptidase 9	7,71	13,64
PLAU	Plasminogen Activator, Urokinase	2,15	8,93
PMEPA1	Prostate Transmembrane Protein, Androgen Induced 1	5,62	6,62
POSTN	Periostin	5,9	22,56
PTH1H	Parathyroid Hormone Like Hormone	37,32	32,38
RSAD2	Radical S-Adenosyl Methionine Domain Containing 2	5,1	8,76
S100A7A	S100 Calcium Binding Protein A7A	7,01	9,24
SPP1	Secreted Phosphoprotein 1	24,09	20,58



TNC	Tenascin C	6,79	8,37
WDR66	Cilia And Flagella Associated Protein 251	6,9	-1,65

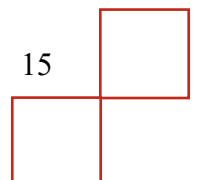
Tabla 1. Tabla de los 31 genes diferenciales sobreexpresados comunes entre las dos cohortes en el grupo de cáncer oral.

*fold change estimado (expresión relativa) referente al grupo de cáncer oral.

7.4 Análisis funcional

Ha sido realizada un análisis de enriquecimiento de los 31 genes comunes entre las dos cohortes sobre expresados para el grupo de cáncer oral (Tabla 2).

<i>Termino GO</i>	<i>Procesos biologicos</i>	<i>PValue</i>	<i>Genes</i>
GO:0030198	extracellular matrix organization	0,000000003	MMP12, POSTN, MMP13, MMP1, COL11A1, MMP3, MMP9, MMP10
GO:0030574	collagen catabolic process	0,000000003	MMP12, MMP13, MMP1, MMP3, MMP9, MMP10
GO:0022617	extracellular matrix disassembly	0,000000006	MMP12, MMP13, MMP1, MMP3, MMP9, MMP10
GO:0035987	endodermal cell differentiation	0,0000000110	COL11A1, COL12A1, HMGA2, INHBA, MMP9
GO:1904645	response to beta-amyloid	0,0000001238	MMP12, MMP13, MMP3, MMP9
GO:0006508	proteolysis	0,0000001492	MMP12, MMP13, PLAU, MMP1, ADAM12, MMP3, MMP9, MMP10
GO:0071492	cellular response to UV-A	0,0000111239	MMP1, MMP3, MMP9
GO:0007155	cell adhesion	0,0000124823	POSTN, CDH3, ADAM12, COL12A1, SPP1, TNC, LAMC2

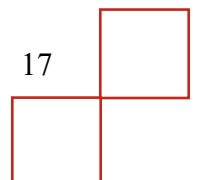




GO:0031640	killing of cells of other organism	0,000141202	CXCL10, CXCL11, CXCL8, CXCL13
GO:0070098	chemokine-mediated signaling pathway	0,000141202	CXCL10, CXCL11, CXCL8, CXCL13
GO:0030593	neutrophil chemotaxis	0,000225790	CXCL10, CXCL11, CXCL8, CXCL13
GO:0007267	Cell-cell signaling	0,000324099	CXCL10, CXCL11, INHBA, CXCL13, PTHLH
GO:0061844	antimicrobial humoral immune response mediated by antimicrobial peptide	0,000521375	CXCL10, CXCL11, CXCL8, CXCL13
GO:0009615	response to virus	0,00054981	BST2, RSAD2, HMGA2, ISG15
GO:0006935	chemotaxis	0,000741411	CXCL10, CXCL11, CXCL8, PLAU
GO:0042127	regulation of cell proliferation	0,001699287	CXCL10, CXCL11, PLAU, TNC
GO:0045071	negative regulation of viral genome replication	0,002028474	BST2, RSAD2, ISG15
GO:0071222	cellular response to lipopolysaccharide	0,002700199	CXCL10, CXCL11, CXCL8, CXCL13
GO:0006954	inflammatory response	0,002790922	CXCL10, CXCL11, CXCL8, SPP1, CXCL13
GO:0030155	regulation of cell adhesion	0,003538632	CXCL8, PLAU, TNC
GO:0051607	defense response to virus	0,004473902	BST2, IFI27, RSAD2, ISG15
GO:0010628	positive regulation of gene expression	0,006136959	CXCL8, CDH3, TNC, HMGA2, INHBA



GO:0001503	ossification	0,006756389	COL11A1, SPP1, MMP9
GO:0042060	wound healing	0,009782769	POSTN, CDH3, TNC
GO:0001649	osteoblast differentiation	0,014152188	SPP1, TNC, CTHRC1
GO:0007189	adenylate cyclase-activating G-protein coupled receptor signaling pathway	0,015243610	CXCL10, CXCL11, PTHLH
GO:0048333	mesodermal cell differentiation	0,017315138	HMGA2, INHBA
GO:0010818	T cell chemotaxis	0,023022334	CXCL10, CXCL11
GO:0045766	positive regulation of angiogenesis	0,023574048	CXCL8, ADAM12, HMGA2
GO:0044267	cellular protein metabolic process	0,024444131	MMP13, MMP1
GO:0033280	response to vitamin D	0,025863933	CXCL10, SPP1
GO:0060339	negative regulation of type I interferon-mediated signaling pathway	0,028697561	MMP12, ISG15
GO:0051281	positive regulation of release of sequestered calcium ion into cytosol	0,044141000	CXCL10, CXCL11
GO:0008284	positive regulation of cell proliferation	0,044542373	CXCL10, TNC, LAMC2, PTHLH
GO:0044344	cellular response to fibroblast	0,046923341	POSTN, CXCL8



	growth factor stimulus		
GO:0045648	positive regulation of erythrocyte differentiation	0,048311582	ISG15, INHBA
GO:0009888	tissue development	0,049697872	POSTN, LAMC2

Tabla 2. Procesos biológicos significativos involucrados los 31 genes diferenciales sobreexpresados comunes entre las dos cohortes. Análisis realizado mediante el *software* DAVID.

Destacamos que los 15 procesos biológicos más significativos fueron: Caderina-3, C-X-C Motif Chemokine Ligand 10, ADAM Metallopeptidase Domain 12, Plasminogen Activator, Urokinase, Radical S-Adenosyl Methionine Domain Containing 2, Matrix metallopeptidase 13, Tenascin C, Matrix Metallopeptidase 3, C-X-C Motif Chemokine Ligand 13, Laminin Subunit Gamma 2, Bone Marrow Stromal Cell Antigen 2, Inhibin Subunit Beta A, Secreted Phosphoprotein 1, Periostin y S100 Calcium Binding Protein A7A; de los cuales están involucrados genes: SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 (Figura 6).

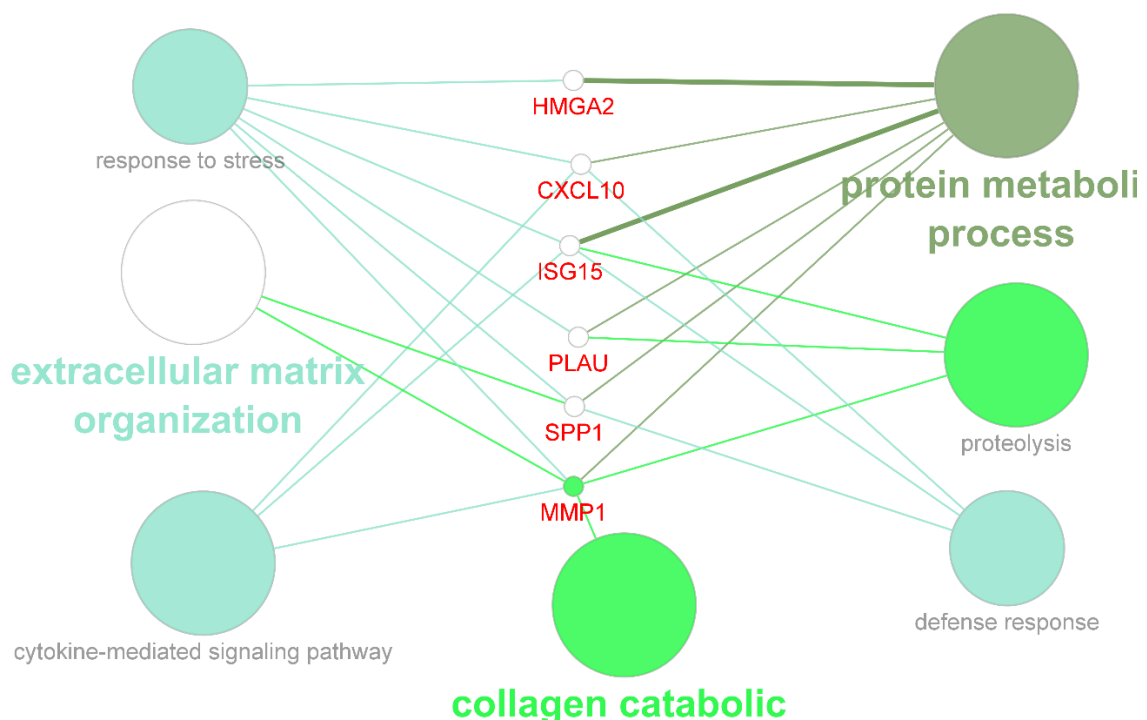


Figura 6. Procesos biológicos relacionados con los genes SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 mediante *software* cytoscape¹⁸.

8. DISCUSIÓN

El COCE sigue siendo uno de los cánceres más prevalentes en todo el mundo con un número creciente de muertos. Las causas principales atribuida al crecimiento de COCE son la falta de medios de diagnóstico temprano, la alta tendencia metastásica y la resistencia a los medicamentos¹⁹. Entonces se necesita la exploración de biomarcadores confiables de diagnóstico temprano y pronóstico de COCE. En nuestro estudio hemos detectado 6 genes sobreexpresados SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 en el grupo de cáncer oral comunes en las dos cohortes, importantes para el diagnóstico precoz y para la detección del cáncer oral. Se ha observado como estos genes son relacionándolos con cáncer oral.

El gen SPP1 (fosfoproteína 1 secretada) codifica la osteopontina (OPN). La osteopontina es una fosfoglicoproteína matriz celular secretada no colágena, que facilita las interacciones célula-matriz y promueve la progresión tumoral. Además, este gen regula: la proliferación celular, la supervivencia, la resistencia a los fármacos, la invasión y el comportamiento. Su sobreexpresión está directamente relacionada con muchas patologías, incluido los varios tipos de cáncer²⁰. En un estudio es particular encontraron que el gen SPP1 sobreexpresado relacionando con la carcinogénesis y la progresión de COCE²¹. Este gen también se vio sobreexpresado, activando la invasión y metástasis de los ganglios linfáticos en cáncer oral²¹. De manera similar, el silenciamiento de SPP1 inhibió la progresión del cáncer de ovario al impedir la vía de señalización de ITGβ1/FAK/AKT, lo que confirma su papel en la carcinogénesis²². En conclusión, SPP1 se sobreexpresa en COCE y desempeña un papel clave en la carcinogénesis y la progresión en COCE.

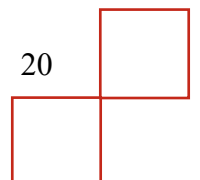
Otro gen relacionado con el cancer oral es PLAU. Este gen es el receptor de uroquinasa activador de plasminógeno, que es una de las proteínas de membrana ancladas a glicosil-fosfatidilinositol (GPI)²³ viendo que inactivamos este gen podría inhibir la proliferación y la metástasis del cáncer en varios tipos de cánceres²⁴. PLAU es una es clave para la disolución de fibrina, la



degradación de la matriz extracelular, la invasión, la metástasis, la angiogénesis y el escape inmunitario de las células tumorales²⁵. La desregulación transcripcional de este gen sugiere su papel dominante en la progresión del cáncer.

Otro gen involucrado es el CXCL10. Rescontramos el alto nivel de este gen en los tejidos tumorales atrae las células T CXCR3+ Th1 y CD8+ a los sitios tumorales inflamados y estimula la polarización de las células T en células T efectoras para promover la destrucción de las células cancerosas. Además, CXCL10 participa en la inducción de la apoptosis en células cancerosas. Las altas expresiones de CXCL10 se correlacionan con la activación de las vías de células T y la infiltración tumoral de células T Th1 y CD8+ son biomarcadores de buen pronóstico para los pacientes con cáncer oral²⁶. Entonces la alta expresión de CXCL10 puede predecir un resultado favorable en pacientes con cáncer oral.

Otro gen involucrado es el ISG15. El ISG15 es una proteína de bajo peso molecular cuya expresión es inducida por interferón. Este gen participa en la regulación de varios procesos biológicos a través de la unión a proteínas diana. Se han identificado proteínas diana ISG15 en células tumorales o células infectadas por virus, lo que sugiere que ISG15 desempeña funciones importantes en la regulación inmunitaria y el desarrollo tumoral²⁷. Previamente se ha demostrado que ISG15 se expresa diferencialmente en diferentes células tumorales y diferentes líneas celulares del mismo origen histológico. Se demostró que la expresión de miR-138 fue significativamente menor en los tejidos cancerosos que en los tejidos adyacentes, lo que indica que la expresión de ISG15 está asociada con la expresión de miR-138. Esto sugiere que miR-138 puede regular los procesos biológicos de COCE mediante la represión de la expresión de ISG15. MiR-138 puede desempeñar un papel nel regular los factores en la vía de la ubiquitina-proteasoma²⁸. Por esto el gen ISG15 puede tener aplicaciones potenciales como un nuevo biomarcador para el diagnóstico temprano y el pronóstico de COCE.

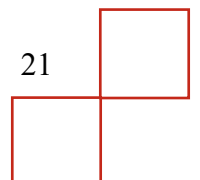




En nuestro estudio de expresión diferencial se encontraron varios genes que pertenecientes a la familia de metaloproteinasas de matriz (MMP): MMP1, MMP3, MMP9, MMP10, MMP12, MMP13. La medición de algunas MMP en pacientes con COCE puede examinarse para el diagnóstico y pronóstico de esta enfermedad²⁹. Las metaloproteinasas de matriz (MMP) se han estudiado con frecuencia como posibles biomarcadores de cáncer, incluso en COCE, y se han asociado con invasión tumoral y metástasis³⁰. La sobreexpresión de los genes MMP1, MMP2, MMP3 y MMP9 se activaron en el tejido tumoral COCE en comparación con el tejido de control³¹. En particular, MMP1 surgió como un marcador excepcionalmente fuerte con respecto a los otros genes de su familia. El nivel de expresión fue de 200 veces mayor en el COCE primario en comparación con el epitelio oral de los controles. Además, la expresión elevada de MMP1 a lo largo de diferentes etapas de COCE y la expresión es creciente, correspondiente a la progresión de tejido normal a displasia a COCE³². Este gen es un potencial candidato como biomarcador para para el desarrollo de COCE.

El gen HMGA2 es un factor de transcripción que codifica a proteína del grupo A2 de alta movilidad, es una proteína asociada a la cromatina y permanece ausente en las células somáticas normales. La desregulación de este gen estimula su represión en varios tumores y se encontró que está relacionada con la supervivencia celular, la regulación del ciclo celular, la resistencia a los medicamentos y la disminución de la supervivencia del paciente³³. HMGA2 se expresa mucho durante la oncogénesis. Los estudios han demostrado que HMGA2 aumenta la progresión del cáncer a través de la activación de varias vías³⁴. Se ha observado sobreexpresado en el COCE humano y se sugirió que su expresión actuaba como una herramienta predictiva y pronóstica útil en el tratamiento clínico de los carcinomas orales³⁵.

Por todo ello nuestro análisis mostró que el grupo de cáncer oral mostro genes y rutas biológicas que podría tener una implicación directa con la progresión, metástasis e invasión del cáncer oral. Así como reafirmamos la identificación genes centrales en dos cohortes diferentes validando nuestros

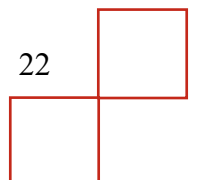




resultados pudiendo proporcionar pistas valiosas para la terapia dirigida en el COCE.

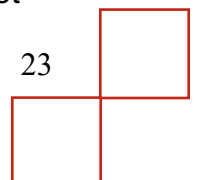
9. CONCLUSIÒN

Fueron identificados 4.083 genes diferenciales para la cohorte 1 y 5.330 para la cohorte 2 en tejido mucoso para càncer oral en humanos, obteniendo un total de 2780 genes comunes entre las dos cohortes. Encontramos 103 genes diferenciales comunes (*fold change* [5]) de los cuales fueron 31 genes sobreexpresados y 72 infraexpresados para el grupo de càncer oral. Mediante el anàlisis de enriquecimiento de 31 genes diferenciales sobreexpresados para grupo de càncer oral se observó “extracelular matrix organization” (GO:0030198) como el proceso biològico más significativos. Destacamos los genes: SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2, que podrían ser futuros biomarcadores para diagnosticar el càncer oral y pudiendo considerarse como dianas terapéuticas.

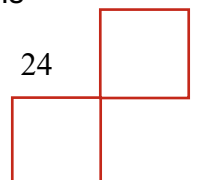


10. BIBLIOGRAFIA

1. Izaguirre Pérez PM. Nivel de conocimiento sobre cáncer oral de estudiantes del último año de la carrera profesional de odontología de tres universidades de Lima, en el año 2012. Lima – Perú; 2012.
2. Varela-Centelles P, Seoane J, Ulloa-Morales Y, Estany-Gestal A, Blanco-Hortas A, -García-Pola M-J, et al. Oral cancer awareness in North-Western Spain: a population-based study. *Med Oral Patol Oral Cir Bucal*. 2021;26(4):e518–25.
3. García-García V, Bascones Martínez A. Cáncer oral: Puesta al día. *Av Odontoestomatol*. 2009;25(5):239–48.
4. Equipo de redactores y equipo de editores médicos de la Sociedad Americana Contra El Cáncer. ¿En qué consisten los tipos de cáncer de orofaringe y de cavidad oral? 2018.
5. Antonio Miguel Cruz P, Aluett Niño Peña I, Kelia Batista Marrero I, Angulo Holguín M. Factores de riesgo de cáncer bucal Risk factors for oral cancer. Vol. 53, *Revista Cubana de Estomatol*. 2016.
6. Serna García M. Estudio del tejido ovàrico para mejorar la producció n a través del tamaño de camada en conejos. 2016 Dec.
7. Megía González R. De gen a carácter: el dogma central de la biología molecular. *El Blog de Genotipia*. 2020.
8. National Human Genome Research Institute. Tecnología de microarrays (chips de ADN o ARN). 2022.
9. González Maestre J. Anàlisis de datis de microarrays. 2010 Apr.
10. López Marta, Mallorquín Paloma, Vega García Miguel. Microarrays y biochips de ADN: informe de vigilancia tecnológica. *Genoma España*; 2002.
11. ThermoFisher Scientific. GeneChip™ Human Genome U133 Plus 2.0 Array.
12. Acosta Chávez J. Dogma central de la Biología Molecular. *Reeditor.com*
13. Burguete A. Figure 2 – uploaded by Ana Burguete. *researchgate.net*

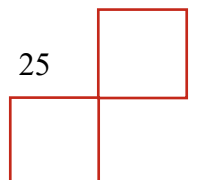


14. ThermoFisher.com. Transcriptome Analysis Console (TAC) 4.0.2.2019.
15. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091–3.
16. Transcriptome analysis console (TAC) software - UK.
17. Sterck L. Draw Venn diagram. Ugent.be.
18. Ono K. Cytoscape. Cytoscape.org.
19. Choi HS, Kim Y-K. Upregulation of MDR- and EMT- related molecules in cisplatin-resistant human oral squamous cell carcinoma cell lines. *Int J MolSci*. 2019;20.
20. Shevde LA, Samant RS. Role of osteopontin in the pathophysiology of cancer. *Matrix Biol*. 2014;37:131–41.
21. Huang C-F, Yu G-T, Wang W-M, Liu B, Sun Z-J. Prognostic and predictive values of SPP1, PAI and caveolin-1 in patients with oral squamous cell carcinoma. *Int J Clin Exp Pathol*. 2014;7(9):6032–9.
22. Zeng B, Zhou M, Wu H, Xiong Z. SPP1 promotes ovarian cancer progression via Integrin β 1/FAK/AKT signaling pathway. *Onco Targets Ther*. 2018;11:1333–43.
23. Ploug M, Ronne E, Behrendt N, Jensen AL, Blasi F, Dano K. Cellular receptor for urokinase plasminogen activator. Carboxyl-terminal processing and membrane anchoring by glycosyl-phosphatidylinositol. *J BiolChem*. 1991;266(3):1926–33.
24. Zhou J, Kwak KJ, Wu Z, Yang D, Li J, Chang M, Song Y, Zeng H, Lee LJ, Hu J, et al. PLAUR Confers Resistance to Gefitinib Through EGFR/P-AKT/Survivin Signaling Pathway. *Cell PhysiolBiochem*. 2018;47(5):1909–24.
25. He F, Chen H, Probst-Kepper M et al (2012) PLAU inferred from a correlation network is critical for suppressor function of regulatory T cells. *MolSystBiol* 8:624.
26. Huang Y-C, Huang J-L, Tseng L-C, Yu P-H, Chen S-Y, Lin C-S. High expression of interferon pathway genes CXCL10 and STAT2 is





- associated with activated T-cell signature and better outcome of oral cancer patients. *J Pers Med.* 2022; 12(2):140.
27. Andersen JB, Hassel BA. The interferon regulated ubiquitin-like protein, ISG15, in tumorigenesis: friend or foe? *Cytokine Growth Factor Rev.* 2006;17(6):411–21.
 28. Zhang Q, He Y, Nie M, Cai W. Roles of miR-138 and ISG15 in oral squamous cell carcinoma. *Exp Ther Med.* 2017;14(3):2329–34.
 29. Mishev G, Deliverska E, Hlushchuk R, Velinov N, Aebersold D, Weinstein F, et al. Prognostic value of matrix metalloproteinases in oral squamous cell carcinoma. *Biotechnol Biotechnol Equip.* 2014;28(6):1138–49.
 30. Ziober FA, Patel CR, Gimotty F, Weber P. Feldman milímetro, et al. Identificación de una firma genética para la detección rápida del carcinoma de células escamosas orales. *Clin Cancer Res.* 2006;12:5960–71.
 31. Baker EA, Leaper DJ, Hayter JP, Dickenson AJ. The matrix metalloproteinase system in oral squamous cell carcinoma. *Br J Oral Maxillofac Surg.* 2006;44(6):482–6.
 32. Stott-Miller M, Houck JR, Lohavanichbutr P, Méndez E, Upton MP, Futran ND, et al. Tumor and salivary matrix metalloproteinase levels are strong diagnostic markers of oral squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2011;20(12):2628–36.
 33. Yadav M, Pradhan D, Singh RP. Integrated analysis and identification of nine-gene signature associated to oral squamous cell carcinoma pathogenesis. *3 Biotech.* 2021;11(5):215.
 34. Mansoori B, Mohammadi A, Ditzel HJ, Duijf PHG, Khaze V, Gjerstorff MF, et al. HMGA2 as a critical regulator in cancer development. *Genes (Basel).* 2021;12(2):269.
 35. Sterenczak KA, Eckardt A, Kampmann A, Willenbrock S, Eberle N, Länger F, et al. HMGA1 and HMGA2 expression and comparative analyses of HMGA2, Lin28 and let-7 miRNAs in oral squamous cell carcinoma. *BMC Cancer.* 2014;14(1):694.



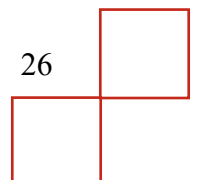
11. ANEXO

Anexo 1. Tabla

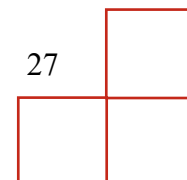
Anexo 2. Paper

A1. Tabla de los 103 genes diferenciales: 31 genes comunes sobreexpresados entre las dos cohortes y 72 genes comunes infraexpresados entre las dos cohortes en el grupo de cáncer oral.

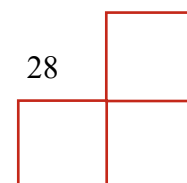
Names	Total	Elements	Nombre	Cohorte1			Cohorte 2		
				FC	PV	FD R	FC	PV	FD R
C1- sobre expres ado C2- sobre excpre sado	31	CDH3	cadherin 3, type 1, P-cadherin (placental)	5,0 7	8,7 5E- 11	1,1 4E- 08	9,1 8	2,0 0E- 43	2,0 0E- 43
		CXCL1 0	chemokine (C-X-C motif) ligand 10	6,2 2	8,2 1E- 07	2,0 1E- 05	20	2,2 5E- 23	2,2 5E- 23
		ADAM1 2	ADAM metallopeptidase domain 12	4,4 2	6,2 7E- 13	2,0 3E- 10	4,1 3	3,8 4E- 21	1,4 3E- 19
		PLAU	plasminogen activator, urokinase	2,1 5	9,3 1E- 08	3,3 8E- 06	8,9 3	9,7 4E- 41	5,7 2E- 38
		RSAD2	radical S-adenosyl methionine domain containing 2	5,1	7,7 0E- 08	2,8 9E- 06	8,7 6	3,0 6E- 20	1,0 1E- 18
		MMP13	matrix metallopeptidase 13	8,3 4	2,2 0E- 10	2,4 0E- 08	8,3 4	2,8 0E- 15	4,9 1E- 14
		TNC	tenascin C	6,7 9	2,5 5E- 13	1,0 3E- 10	8,3 7	1,0 0E- 34	2,2 5E- 32
		MMP3	matrix metallopeptidase 3	21, 72	1,0 4E- 10	1,3 2E- 08	72, 22	4,4 3E- 33	7,8 3E- 31



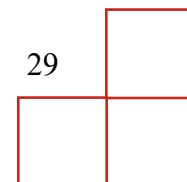
		CXCL1 3	chemokine (C-X-C motif) ligand 13	7,3 9	1,8 5E- 08	8,8 8E- 07	54, 81	1,9 7E- 18	5,1 7E- 17
		LAMC2	laminin, gamma 2	12, 53	2,9 3E- 16	4,8 5E- 13	17, 99	3,5 3E- 37	1,2 1E- 34
		BST2	bone marrow stromal cell antigen 2	5,2	1,7 5E- 07	5,7 0E- 06	6,9 9	3,5 9E- 27	2,9 3E- 25
		INHBA	inhibin beta A	32, 55	2,8 0E- 20	4,4 4E- 16	1,6	4,2 2E- 10	3,6 0E- 09
		SPP1	secreted phosphoprotein 1	24, 09	4,6 3E- 13	1,5 9E- 10	20, 58	1,8 7E- 27	1,5 8E- 25
		POSTN	periostin, osteoblast specific factor	5,9	5,5 9E- 11	7,9 1E- 09	22, 56	8,6 9E- 37	2,8 1E- 34
		S100A7 A	S100 calcium binding protein A7	4,1	0,0 242	0,0 726	9,2 4	9,4 2E- 15	1,5 4E- 13
		PMEPA 1	prostate transmembrane protein, androgen induced 1	3,5 1	3,1 5E- 13	1,2 1E- 10	6,6 2	4,0 8E- 37	1,3 9E- 34
		COL11 A1	collagen, type XI, alpha 1	10, 42	8,2 3E- 12	1,6 9E- 09	12, 27	1,1 1E- 17	2,6 5E- 16
		PTHLH	Parathyroid hormone-like hormone	37, 32	2,3 0E- 15	2,3 7E- 12	32, 28	2,0 2E- 28	1,9 7E- 26
		CTHRC 1	collagen triple helix repeat containing 1	8,6 4	1,9 3E- 17	7,0 3E- 14	16, 34	4,4 6E- 43	3,4 3E- 40
		MMP10	matrix metallopeptidase 10	31, 84	7,6 3E- 14	4,2 1E- 11	36, 29	5,2 2E- 27	4,2 1E- 25
		CXCL1 1	chemokine (C-X-C motif) ligand 11	11, 88	2,3 5E- 08	1,0 8E- 06	10, 62	9,6 3E- 15	1,5 7E- 13
		MMP1	matrix metallopeptidase 1	99, 02	1,3 7E- 17	5,3 3E- 14	358 ,93	1,4 0E- 45	1,4 0E- 45



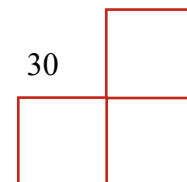
		MMP12	matrix metalloproteinase 12	20,96	2,40E-12	6,17E-10	109,44	7,43E-44	6,38E-41
		CXCL8	chemokine (C-X-C motif) ligand 8	5,64	4,46E-06	7,91E-05	40,55	9,49E-35	2,16E-32
		COL12A1	collagen, type XII, alpha 1	2,39	2,91E-05	0,0004	8,13	5,35E-28	4,96E-26
		WDR66	WD repeat domain 66	6,9	3,62E-12	8,74E-10	3,03	4,67E-15	7,92E-14
		HMGA2	high mobility group AT-hook 2	5,17	2,53E-10	2,70E-08	1,83	5,58E-12	6,27E-11
		KRT17	keratin 17, type I	3,35	5,41E-06	9,31E-05	24,43	1,40E-45	1,40E-45
		ISG15	ISG15 ubiquitin-like modifier	9,49	3,24E-13	1,21E-10	11,61	2,43E-31	3,45E-29
		MMP9	matrix metalloproteinase 9	7,71	1,93E-11	3,23E-09	13,64	6,01E-28	5,53E-26
		IFI27	interferon, alpha-inducible protein 27	5,92	2,79E-09	1,91E-07	6,88	2,09E-26	1,56E-24
C1- infra expresado C2- Infra expresado		EMP1	epithelial membrane protein 1	- 3,49	7,39E-10	6,42E-08	- 4,11	8,90E-20	2,76E-18
		TMPRSS11B	transmembrane serine protease, 11B	- 103,43	4,49E-14	2,73E-11	- 86,03	2,04E-28	1,98E-26
		PTN	pleiotrophin	- 3,56	1,61E-10	1,89E-08	- 5,35	1,10E-19	3,36E-18
		PPL	periplakin	-5,5	2,82E-09	1,93E-07	- 5,67	7,99E-28	7,18E-26



	ABCA8	ATP binding cassette subfamily A member 8	- 18, 19	1,9 4E-14	1,4 3E-11	- 10, 58	1,5 7E-25	1,0 4E-23
	PITX2	paired-like homeodomain 2	- 8,2 3	6,3 7E-15	5,5 3E-12	- 5,9 3	2,6 9E-07	1,4 9E-06
	CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	- 3,6 9	1,0 4E-11	1,9 6E-09	- 9,5 2	1,5 6E-35	4,1 5E-33
	CHRD1	chordin-like 1	-8,5	2,6 3E-10	2,7 9E-08	- 8,7 3	7,3 1E-23	3,3 6E-21
	TTN	titin	- 4,4 9	0,0 061	0,0 256	- 1,7 7	1,3 0E-07	7,6 1E-07
	ALOX12	arachidonate 12-lipoxygenase	- 6,0 2	1,2 2E-10	1,5 0E-08	- 5,6 8	3,8 3E-18	9,7 3E-17
	FLG	filaggrin	- 9,0 8	8,7 0E-05	0,0 009	- 22, 39	2,7 1E-15	4,7 7E-14
	ADH7	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	- 15, 95	7,6 8E-12	1,6 1E-09	- 10, 78	1,5 8E-21	6,1 4E-20
	ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	- 32, 03	6,2 9E-11	8,7 0E-09	- 15, 12	2,4 4E-36	7,4 2E-34
	FAM3B	family with sequence similarity 3, member B	- 14, 43	3,0 7E-16	4,9 4E-13	- 17, 32	3,1 0E-25	1,9 8E-23
	FCER1A	Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide	- 7,6 2	7,5 8E-13	2,3 5E-10	- 5,4 8	8,2 9E-25	4,9 7E-23
	ENDOU	endonuclease, polyU-specific	- 17, 22	1,4 1E-13	6,5 1E-11	- 26, 12	2,8 3E-40	1,5 0E-37
	OGN	osteoglycin	-2,2	0,0 004	0,0 03	- 5,1 7	7,1 7E-11	6,8 5E-10

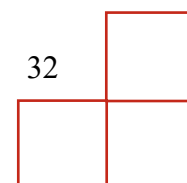


	SCARA 5	scavenger receptor class A, member 5	-4,6	4,6 3E- 13	1,5 9E- 10	- 4,2 5	8,2 2E- 25	4,9 3E- 23
	CLCA4	chloride channel accessory 4	- 14, 05	1,8 3E- 07	5,8 8E- 06	- 24, 01	2,7 8E- 25	1,7 8E- 23
	ATP1A 2	ATPase, Na ⁺ /K ⁺ transporting, alpha 2 polypeptide	- 3,2 3	0,0 047	0,0 21	- 6,3 5	5,5 5E- 14	8,2 0E- 13
	CSTB	cystatin B (stefin B)	- 8,8 7	1,5 7E- 11	2,7 7E- 09	- 8,0 2	1,0 6E- 33	2,0 5E- 31
	ID4	inhibitor of DNA binding 4, dominant negative helix-loop- helix protein	- 1,9 4	8,9 5E- 09	4,9 6E- 07	- 6,5 7	2,7 6E- 30	3,4 6E- 28
	SLC16 A6	solute carrier family 16, member 6	- 2,7 2	4,4 4E- 14	2,7 2E- 11	- 3,8 3	3,4 7E- 27	2,8 3E- 25
	KRT4	keratin 4, type II	- 58, 99	6,9 9E- 11	9,4 8E- 09	- 48, 05	4,2 6E- 23	2,0 4E- 21
	PPP1R 3C	protein phosphatase 1, regulatory subunit 3C	- 1,7 6	1,0 4E- 06	2,4 3E- 05	- 15, 07	2,8 5E- 37	9,9 7E- 35
	SLURP 1	secreted LY6/PLAUR domain containing 1	-5,1	8,4 2E- 08	3,1 1E- 06	- 12, 35	8,8 7E- 19	2,4 3E- 17
	KLK13	kallikrein related peptidase 13	- 6,9 3	2,0 5E- 07	6,4 9E- 06	- 4,7 2	1,1 1E- 16	2,3 6E- 15
	KRT13	keratin 13, type I	- 31, 88	3,5 7E- 07	1,0 2E- 05	- 8,7 3	1,9 1E- 13	2,6 0E- 12
	MB	myoglobin	- 6,7 7	0,0 003	0,0 021	- 7,8 4	6,9 3E- 08	4,2 1E- 07
	TGM3	transglutaminase 3	- 96, 84	3,0 9E- 12	7,6 7E- 10	- 26, 21	2,7 7E- 22	1,1 7E- 20
	ACTA1	actin, alpha 1, skeletal muscle	- 7,2 4	0,0 017	0,0 095	- 24, 26	1,2 5E- 06	6,2 1E- 06

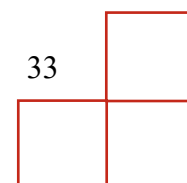


	A2ML1	alpha-2-macroglobulin-like 1	- 11,73	3,59E-08	1,54E-06	- 9,11	9,90E-23	4,47E-21
	GALNT12	polypeptide N-acetylgalactosaminyltransferase 12	- 1,98	5,66E-10	5,18E-08	- 6,48	1,33E-34	2,90E-32
	EXPH5	exophilin 5	- 4,38	5,83E-06	9,90E-05	- 6,33	3,14E-28	2,98E-26
	SPINK7	serine peptidase inhibitor, Kazal type 7 (putative)	- 36,26	4,48E-08	1,85E-06	- 12,78	7,25E-15	1,20E-13
	CYP2C18	cytochrome P450, family 2, subfamily C, polypeptide 18	-6,8	1,08E-07	3,79E-06	- 6,51	4,94E-23	2,34E-21
	ALDH3A1	aldehyde dehydrogenase 3 family, member A1	- 10,01	3,97E-10	3,91E-08	- 7,61	2,41E-17	5,53E-16
	CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5	- 5,68	0,0001	0,0012	- 8,61	1,61E-17	3,75E-16
	CNFN	cornifelin	- 6,78	0,007	0,0284	- 5,72	1,92E-13	2,61E-12
	SCIN	scinderin	- 6,78	0,007	0,0284	- 12,29	1,40E-45	1,40E-45
	KLHL41	kelch-like family member 42	- 2,07	0,0347	0,0951	- 22,58	1,59E-06	7,78E-06
	PYGM	phosphorylase, glycogen, muscle	- 9,66	6,42E-06	0,0001	- 5,05	3,36E-18	8,63E-17
	TMPRSS11E	transmembrane protease, serine 11E	- 8,13	2,20E-06	4,48E-05	- 10,65	3,81E-19	1,10E-17
	ACPP	acid phosphatase, prostate	- 1,59	9,24E-06	0,0001	- 10,57	8,22E-39	3,71E-36

		HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	- 6,2 9	3,2 3E- 13	1,2 1E- 10	- 3,9 3	1,5 9E- 20	5,4 0E- 19
		CAPN1 4	calpain, small subunit 2	- 8,6 7	7,6 7E- 09	4,3 8E- 07	- 37, 45	1,1 1E- 28	1,1 2E- 26
		PAX9	paired box 9	- 2,6 8	4,3 5E- 15	3,9 8E- 12	- 4,4 9	5,1 7E- 33	9,0 2E- 31
		MYBPC 1	myosin binding protein C, slow type	- 6,5 1	0,0 005	0,0 035	- 9,0 4	8,6 1E- 07	4,4 0E- 06
		TNNC2	troponin C type 2 (fast)	- 9,0 5	0,0 005	0,0 037	- 7,8 7	8,4 2E- 09	5,9 1E- 08
		RHCG	Rh family, C glycoprotein	- 13, 12	1,0 4E- 06	2,4 3E- 05	- 10, 69	2,3 1E- 24	1,3 1E- 22
		SCEL	sciellin	- 18, 04	1,7 0E- 06	3,6 0E- 05	- 14, 06	4,5 4E- 32	6,9 3E- 30
		ALDH1 A1	aldehyde dehydrogenase 1 family, member A1	- 12, 01	3,5 2E- 14	2,3 2E- 11	- 7,8 8	3,9 8E- 16	7,8 0E- 15
		GBP6	guanylate binding protein family, member 6	- 14, 21	9,6 6E- 10	7,9 4E- 08	- 11, 01	9,0 0E- 29	9,2 4E- 27
		CKM	creatine kinase, muscle	-7,5	0,0 019	0,0 102	- 13, 65	2,4 5E- 08	1,6 0E- 07
		KRT78	keratin 78, type II	- 10, 9	1,0 5E- 11	1,9 8E- 09	- 27, 23	2,2 7E- 41	1,4 3E- 38
		MAL	mal, T-cell differentiation protein	- 79, 44	1,0 8E- 12	3,2 0E- 10	- 49, 38	3,8 6E- 29	4,1 3E- 27
		BNIP1	BCL2/adenovirus E1B 19kD	- 5,6 9	2,0 7E- 05	0,0 003	- 4,1 1	8,4 3E- 26	5,8 2E- 24



			interacting protein like						
		GPD1L	glycerol-3-phosphate dehydrogenase 1-like	- 5,7 4	4,5 5E- 14	2,7 3E- 11	- 5,8 7	2,4 5E- 43	2,0 3E- 40
		VSIG10L	V-set and immunoglobulin domain containing 10 like	- 9,0 1	2,7 4E- 07	8,1 8E- 06	- 9,5 2	2,1 8E- 23	1,0 8E- 21
		CRNN	cornulin	- 87, 73	5,0 8E- 14	3,0 2E- 11	- 54, 37	6,0 9E- 25	3,7 3E- 23
		HRASLS	HRAS-like suppressor	- 3,9 9	3,1 1E- 10	3,1 9E- 08	- 6,1 3	2,5 8E- 16	5,2 0E- 15
		CRISP3	cysteine-rich secretory protein 3	- 91, 32	1,6 4E- 12	4,4 4E- 10	- 304 ,47	6,4 0E- 33	1,1 0E- 30
		SAMD5	sterile alpha motif domain containing 5	- 6,8 3	7,1 3E- 15	6,0 6E- 12	- 2,3 3	8,9 0E- 25	5,3 1E- 23
		HLF	hepatic leukemia factor	- 7,4 2	1,5 3E- 20	4,1 7E- 16	- 8,2 4	8,8 0E- 39	3,8 7E- 36
		CFD	complement factor D (adipsin)	- 11, 28	1,3 4E- 09	1,0 4E- 07	- 6,4 3	7,1 5E- 25	4,3 4E- 23
		MYH2	myosin, heavy chain 2, skeletal muscle, adult	- 11, 35	0,0 003	0,0 023	- 41, 02	7,4 6E- 08	4,5 2E- 07
		CYP4B1	cytochrome P450, family 4, subfamily B, polypeptide 1	- 7,2 3	3,5 4E- 18	1,9 3E- 14	- 20, 28	8,2 8E- 27	6,5 2E- 25
		CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7	- 7,1 1	1,3 5E- 06	3,0 0E- 05	- 4,0 3	1,5 4E- 12	1,8 7E- 11
		HOPX	HOP homeobox	- 2,4 2	8,0 4E- 09	4,5 3E- 07	- 6,6 4	4,0 3E- 21	1,4 9E- 19





		MUC15	mucin 15, cell surface associated	- 6,1 6	2,2 4E- 07	6,9 3E- 06	- 6,0 9	1,7 9E- 20	6,0 5E- 19
		SPINK5	serine peptidase inhibitor, Kazal type 5	- 10, 52	5,4 9E- 07	1,4 5E- 05	- 11, 5	9,4 8E- 24	4,9 2E- 22
		CRCT1	cysteine rich C-terminal 1	- 9,7 8	0,0 002	0,0 018	- 6,8 3	1,2 6E- 14	2,0 2E- 13



Análisis transcriptómico en la cavidad bucal para el cáncer oral en humanos mediante *microarrays*

Gaia Gasparroni¹ and Marta Serna-Garcia¹

¹ Universidad Europea de Valencia. Faculty of Health Sciences; gaia-1996@live.it

² marta.serna@universidadeuropea.es; Tel.: +34610819223

Resumen: (1) Introducción: El cáncer oral es el más frecuente de los cánceres que de cabeza y cuello. Los *microarrays* nos permiten interpretar y analizar los datos obtenidos de una secuenciación de RNA. (2) Objetivo: estudiar el tejido de la cavidad oral en humanos mediante la técnica de *microarrays*, con la finalidad de identificar los procesos biológicos y genes de interés para en un futuro usar como biomarcadores. (3) Materiales y métodos: Fueron comparados dos grupos divergentes: cáncer (C) *versus* control (N) para dos cohortes diferentes, los datos brutos (.cel) fueron obtenidos de una base de datos de acceso público (*ArrayExpress*). Se realizó un análisis diferencial mediante el software TAC (*Transcriptome Analysis Console*) considerando *p value* ≤ 0,05. El análisis de enriquecimiento de los genes diferenciales encontrados fue mediante el software DAVID (*Database for Annotation, Visualization and Integrated Discovery*) y la visualización de los procesos biológicos y genes relevantes mediante el software Cytoscape. (4) Resultados: Fueron detectados 4.083 genes diferenciales para la cohorte 1 y 5.330 para la cohorte 2, 2072 los genes comunes para las dos cohortes. Se realizó un filtro de *fold change* de |5| siendo 31 genes sobreexpresados y 72 infraexpresados para el grupo de cáncer oral. Destacamos los genes: SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 todos sobreexpresados en cáncer oral de los cuales los genes. (5) Conclusión: Estos genes podrían ser futuros biomarcadores tempranos para diagnosticar el cáncer oral.

Palabras clave: Cancer oral, microarrays, expresión diferencial.

Citaci3n: Gasparroni, Gaia; Serna-Garica, Marta; Análisis transcript3mico en la cavidad bucal para el c3ncer oral en humanos mediante *microarrays* *Dent. J.* **2022**, *10*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor: Firstname Last-name

Received: date
Accepted: date
Published: date

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms

1. Introducción

El cáncer oral es el más frecuente de los cánceres que de cabeza y cuello. El cáncer de cavidad oral se encuentra dentro de las 10 neoplasias malignas más frecuentes en nuestro país y se observa un aumento de los casos diagnosticados de cáncer oral en los últimos 10 años [1]. La gravedad de este tipo de cancer se clasifica también en base a la fase en el cual se encuentra. Cuando se encuentran en fases tempranas de desarrollo, los cánceres orales tienen una tasa de supervivencia del 80 al 90%. Desafortunadamente la mayoría se encuentran como cánceres de etapa tardía y la tasa de mortalidad es muy alta de aproximadamente 43% a los cinco años del diagnóstico [1]. La mayoría de estos cánceres son difíciles de descubrir y esto es debido a la falta de conciencia pública junto con la falta de un programa nacional de detección precoz de cáncer oral, llevado a cabo por profesionales médicos y odontólogos [1]. De esta manera, se resalta la necesidad del conocimiento de su diagnóstico dentro de los conocimientos básicos del odontólogo general [1]. Por esto, el nuestro objetivo es estudiar el tejido de la cavidad oral en

humanos mediante la técnica de *microarrays*, con la finalidad de identificar los procesos biológicos y genes de interés para en un futuro usar como biomarcadores.

El cáncer bucal representa la undécima neoplasia más incidente en el mundo, con más de 202.000 con una relación hombre:mujer de 2:1. En el caso particular de España, el cáncer bucal ocupa el puesto 16 entre todas las neoplasias por incidencia y el puesto 19 por mortalidad [2]. En cuanto a la edad, el 90% de los cánceres orales se diagnostican en mayores de 40 años, y más del 50% en individuos de más de 65 años [3]. Recientemente se ha observado un aumento en la incidencia de esta enfermedad en menores de 40 años lo que podría estar asociado a una infección por el virus del papiloma humano. Las principales localizaciones son lengua y el paladar de boca en Europa y el norte de América y la mucosa bucal en la India. En referencia a la histología, el 90% de los carcinomas orales son del tipo carcinoma oral de células escamosas. La supervivencia a los 5 años es del 40-56% aunque varía según el tipo de cáncer: 70% cáncer labial, 30% otras localizaciones orales [3].

El cáncer es una enfermedad multifactorial causada por alteraciones en los genes mediante una amplia gama de mecanismos que involucran también factores no genéticos. En la carcinogénesis son relevantes, también, algunos factores como las radiaciones, las infecciones, la dieta, la irritación crónica, la exposición excesiva a la luz solar y los estados de inmunodepresión, el hábito de fumar y el consumo de alcohol [4]. El desarrollo del cáncer bucal es un proceso en múltiples etapas que implica la acumulación de alteraciones genéticas y epigenéticas en genes regulatorios clave. Los mecanismos genéticos básicos en el cáncer son la sobreexpresión de oncogenes y el silenciamiento de genes supresores tumorales. Las variaciones individuales en la susceptibilidad al carcinoma de células escamosas bucal se atribuyen a complejas interacciones entre factores genéticos y ambientales, aunque los mecanismos convergen en procesos inflamatorios. La inflamación se relaciona con alteraciones de la expresión de oncogenes y genes supresores tumorales [5].

2. Materiales and Metodos

2.1 Análisis transcriptòmica

2.1.1 Diseño del estudio

Los datos brutos (*.cel*) fueron elegidos de *ArrayExpress*, una base de datos de acceso público. El número de acceso a la base de datos de *ArrayExpress* es E-GEOD-42743 (cohorte 1) y E-GEOD-30784 (cohorte 2). La búsqueda fue realizada aplicando un filtro del organismo; *homo sapiens sapiens* y mediante la palabra clave "cáncer oral". El resultado de nuestra búsqueda ha sido de 219 experimentos. Siendo elegidos 2, descartando los experimentos con *.cel* corruptos.

Para la cohorte 1, se realizó un análisis comparativo de la expresión génica utilizando la plataforma Affymetrix mediante el *GeneChip™ Human Genome U133A 2.0* para comparar las diferencias entre las muestras de cáncer oral y la mucosa "normal". Los controles fueron 29 casos, mientras que los casos fueron 74, de total de 103 muestras. Para

la cohorte 2, los casos fueron pacientes con COCE primario (carcinoma oral de células escamosas) de la mucosa oral. Los controles fueron pacientes que se sometieron a amigdalectomía o cirugía oral para el tratamiento de enfermedades distintas del cáncer. Se obtuvieron un total de 229 muestras, los controles fueron 45 casos, mientras que los casos fueron 167. Se realizó también aquí un análisis comparativo de la expresión génica utilizando la plataforma *Affymetrix* mediante el *GeneChip™ Human Genome U133A 2.0* para comparar las diferencias entre las muestras de casos y control.

2.1.2 Análisis de expresión diferencial

Todos los archivos brutos *.cel* para la cohorte 1 and cohorte 2 fueron comparados usando dos grupos: cáncer oral (C) y el control (N), analizándose todas las muestras con el mismo flujo de trabajo. El *software* TAC (*Transcriptome Analysis Console*) (*Affymetrix*) realizó el análisis diferencial entre los grupos C vs N para las dos cohortes diferentes. Los datos sin procesar se preprocesaron, incluida la corrección de fondo, la normalización mediante un análisis robusto de promedio de múltiples matrices (RMA, *Robust Multichip Analysis*). El análisis ANOVA unidireccional se realizó para identificar genes expresados diferencialmente. Un $p\text{ value} \leq 0,05$ se consideró estadísticamente significativo. También se realizó un análisis de componentes principales (PCA) para determinar la variabilidad en los conjuntos de datos. PCA reduce la complejidad de los datos de alta dimensión y los simplifica, identificando patrones de expresión y fuentes de variabilidad en un gran conjunto de datos de forma tridimensional [6]. La distancia entre cualquier par de puntos está relacionada con la similitud entre dos muestras en un espacio de alta dimensión, en este caso cada variable correspondiente a un archivo *.cel* unidimensional.

2.1.3 Análisis de enriquecimiento

Los 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) para mapear las rutas enriquecidas de genes usando la base de datos de la Enciclopedia de Genes y Genomas de Kyoto (KEGG), se determinó $p\text{ value} \leq 0,05$ 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 [6].

3. Resultados

3.1 Análisis exploratorio de los datos

El análisis de componentes principales (ACP) no supervisado permite observar la distribución en el espacio de las muestras de tejido de la cavidad oral en humanos en cuenta las tres componentes principales. En la cohorte 1 explican el 45,6 % de la variabilidad de las muestras (Figura 3.a, b). En el análisis ACP de las 103 muestras analizadas, 74 muestras pertenecientes el grupo con cáncer (C) y 29 muestras pertenecientes al grupo normal (N), se observa que las muestras se diferencian desde el punto de vista transcriptómico.

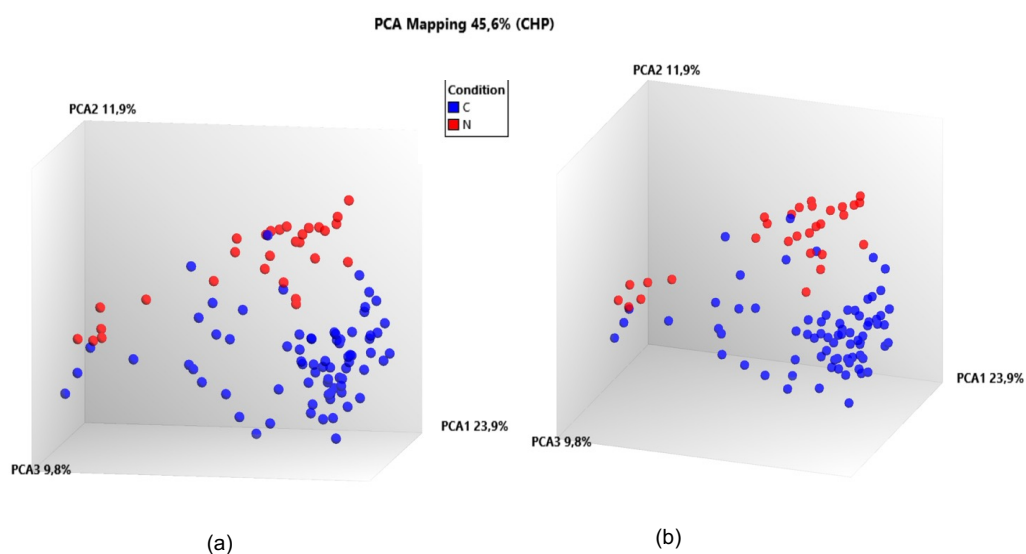


Figura 3. En las figuras se observan imágenes del ACP de la cohorte 1, pero en diferentes planos tridimensionales: (a) Esta figura nos da una visión del ACP más frontal. (b) Visión del ACP ligeramente lateralizada [7].

El ACP de la cohorte 2 permite observar la distribución del el 35,6 % de la variabilidad de las muestras (Figura 4). En el análisis ACP de las 212 muestras analizadas (167 las muestras pertenecientes a las hembras de la población cáncer (C) y 45 muestras pertenecientes a las hembras de la población no cáncer (N) se observa que las muestras se diferencian desde el punto de vista transcriptómico.

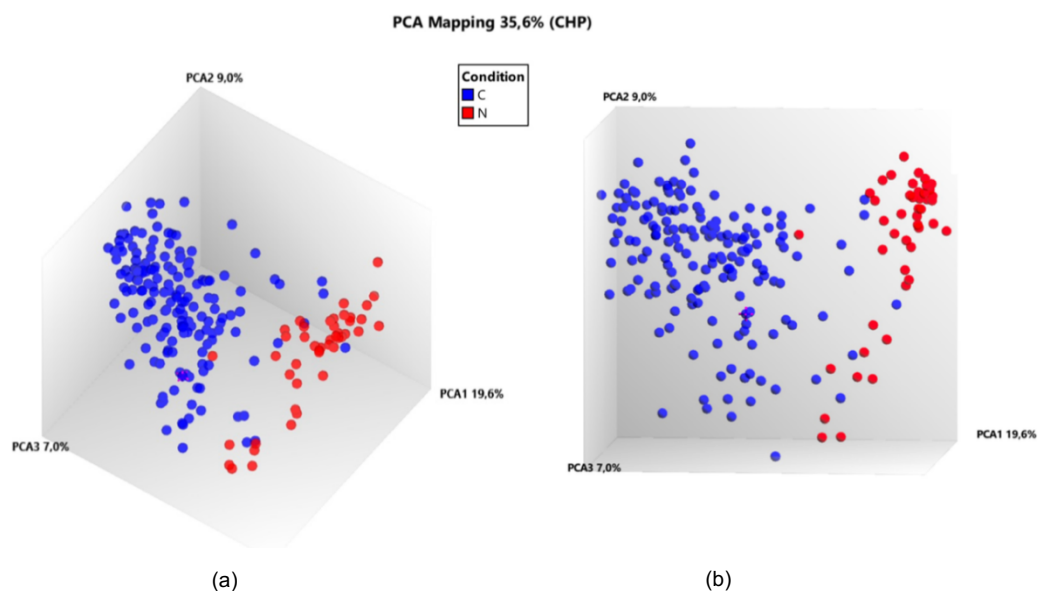


Figura 4. En las figuras se observan imágenes del ACP de la cohorte 2, pero en diferentes planos tridimensionales: (a) Esta figura nos da una visión del ACP desde arriba. (b) Visión del ACP lateralizada [7].

3.2 Análisis diferencial

Se realizó el análisis de la varianza (ANOVA) de la expresión de los genes que se expresaban de manera diferencial en cohorte 1 y la cohorte 2 cada uno dividido en dos grupos, cáncer *vs* normal mediante el uso del software TAC. El número de genes diferencialmente expresados obtenidos para la cohorte 1 ha sido de 4.083 y 5.330 genes diferenciales para la cohorte 2 ($p \text{ value} \leq 0,05$).

3.3 Genes diferenciales comunes entre la cohorte 1 y cohorte 2

Hemos encontrado 103 (Tabla A1, Figura 5) genes comunes entre las dos cohortes realizando un filtro por cambio de expresión (*fold change*) $| 5 |$ y despreciados aquellos genes que no tienen una anotación funcional hasta el momento. Se observó que 31 genes tienen unos valores de cambio de expresión positivos (sobrexpresados) (Tabla 1) y 72 genes con valores de cambio de expresión negativos (infraexpresados) para el grupo de cáncer oral.

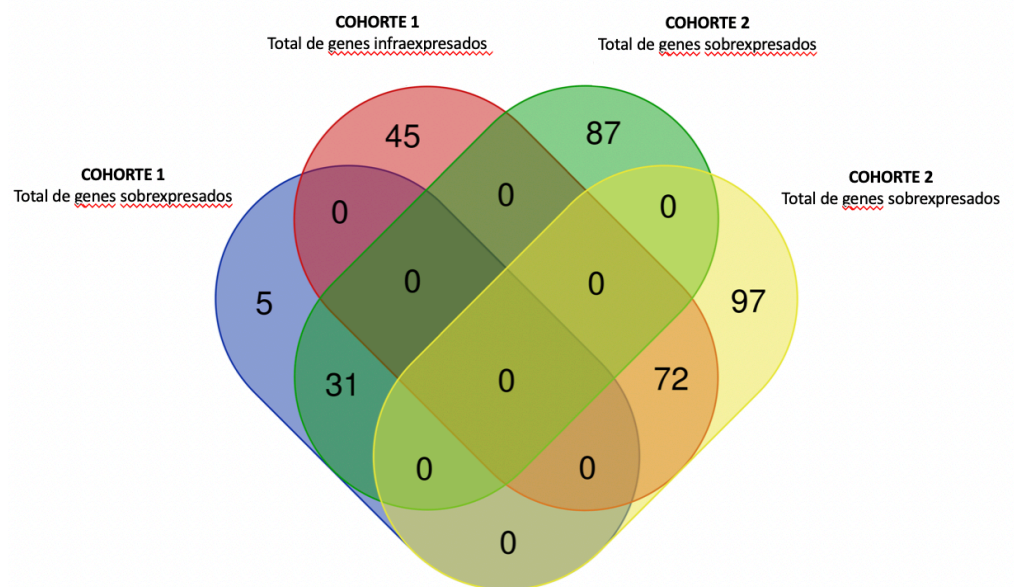


Figura 5. Venn diagram entre genes sobreexpresados y genes infraexpresados entre cohorte 1 y cohorte 2 realizando el filtro por *fold change* $| 5 |$. Fueron 42 genes sobreexpresados y 138 genes infraexpresados para la cohorte 1 y 159 genes sobreexpresados y 205 genes infraexpresados para la cohorte 2. Los datos obtenidos han sido 31 genes comunes sobreexpresados y 72 genes comunes infraexpresados entre las dos cohortes [8].

Table 1. Tabla de los 31 genes diferenciales sobreexpresados comunes entre las dos cohortes en el grupo de cáncer oral.

ALIAS	NOMBRE GENES	FC* COHORTE 1	FC* COHORTE 2
CDH3	Caderina-3	5,07	9,18
ADAM12	ADAM Metallopeptidase Domain 12	4,42	4,13

BST2	Bone Marrow Stromal Cell Antigen 2	5,2	6,99
COL11A1	Collagen, type XI, alpha 1	3,53	12,57
COL12A1	Collagen, type XI, alpha 1	2,39	8,13
CTHRC1	Collagen Triple Helix Repeat Containing 1	8,34	8,34
CXCL10	C-X-C Motif Chemokine Ligand 10	6,22	20
CXCL11	C-X-C Motif Chemokine Ligand 11	11,88	10,62
CXCL13	C-X-C Motif Chemokine Ligand 13	7,39	54,81
CXCL8	C-X-C Motif Chemokine Ligand 8	5,64	40,55
HMGA2	High Mobility Group AT-Hook 2	5,17	7,21
IFI27	interferon, alpha-inducible protein 27	5,92	6,88
INHBA	Inhibin Subunit Beta A	32,55	1,6
ISG15	ISG15 Ubiquitin Like Modifier	9,49	11,61
KRT17	Keratin 17	3,35	24,43
LAMC2	Laminin Subunit Gamma 2	2,91	17,99
MMP1	Matrix Metalloproteinase 1	8,34	358,93
MMP10	Matrix Metalloproteinase 10	31,84	36,29
MMP12	Matrix Metalloproteinase 12	20,96	109,44
MMP13	Matrix metalloproteinase 13	8,34	8,34
MMP3	Matrix Metalloproteinase 3	21,72	72,22
MMP9	Matrix Metalloproteinase 9	7,71	13,64
PLAU	Plasminogen Activator, Urokinase	2,15	8,93
PMEPA1	Prostate Transmembrane Protein, Androgen Induced 1	5,62	6,62
POSTN	Periostin	5,9	22,56
PTH1H	Parathyroid Hormone Like Hormone	37,32	32,38
RSAD2	Radical S-Adenosyl Methionine Domain Containing 2	5,1	8,76

S100A7A	S100 Calcium Binding Protein A7A	7,01	9,24
SPP1	Secreted Phosphoprotein 1	24,09	20,58
TNC	Tenascin C	6,79	8,37
WDR66	Cilia And Flagella Associated Protein 251	6,9	-1,65

* fold change estimado (expresión relativa) referente al grupo de cáncer oral.

3.4 Functional analysis

Ha sido realizada un análisis de enriquecimiento de los 31 genes comunes entre las dos cohortes sobreexpresados para el grupo de cáncer oral (Tabla 2).

Tabla 2. Procesos biológicos significativos involucrados los 31 genes diferenciales sobreexpresados comunes entre las dos cohortes. Análisis realizado mediante el *software* DAVID.

Term GO	Biological Process	PValue	Genes
GO:0030198	extracellular matrix organization	0,000000003	MMP12, POSTN, MMP13, MMP1, COL11A1, MMP3, MMP9, MMP10
GO:0030574	collagen catabolic process	0,000000003	MMP12, MMP13, MMP1, MMP3, MMP9, MMP10
GO:0022617	extracellular matrix disassembly	0,000000006	MMP12, MMP13, MMP1, MMP3, MMP9, MMP10
GO:0035987	endodermal cell differentiation	0,000000110	COL11A1, COL12A1, HMGA2, INHBA, MMP9
GO:1904645	response to beta-amyloid	0,000001238	MMP12, MMP13, MMP3, MMP9
GO:0006508	proteolysis	0,000001492	MMP12, MMP13, PLAU, MMP1, ADAM12, MMP3, MMP9, MMP10
GO:0071492	cellular response to UV-A	0,000111239	MMP1, MMP3, MMP9
GO:0007155	cell adhesion	0,000124823	POSTN, CDH3, ADAM12, COL12A1, SPP1, TNC, LAMC2
GO:0031640	killing of cells of other organism	0,000141202	CXCL10, CXCL11, CXCL8, CXCL13

GO:0070098	chemokine-mediated signaling pathway	0,000141202	CXCL10, CXCL11, CXCL8, CXCL13
GO:0030593	neutrophil chemotaxis	0,000225790	CXCL10, CXCL11, CXCL8, CXCL13
GO:0007267	Cell-cell signaling	0,000324099	CXCL10, CXCL11, INHBA, CXCL13, PTHLH
GO:0061844	antimicrobial humoral immune response mediated by antimicrobial peptide	0,000521375	CXCL10, CXCL11, CXCL8, CXCL13
GO:0009615	response to virus	0,00054981	BST2, RSAD2, HMGA2, ISG15
GO:0006935	chemotaxis	0,000741411	CXCL10, CXCL11, CXCL8, PLAU
GO:0042127	regulation of cell proliferation	0,001699287	CXCL10, CXCL11, PLAU, TNC
GO:0045071	negative regulation of viral genome replication	0,002028474	BST2, RSAD2, ISG15
GO:0071222	cellular response to lipopolysaccharide	0,002700199	CXCL10, CXCL11, CXCL8, CXCL13
GO:0006954	inflammatory response	0,002790922	CXCL10, CXCL11, CXCL8, SPP1, CXCL13
GO:0030155	regulation of cell adhesion	0,003538632	CXCL8, PLAU, TNC
GO:0051607	defense response to virus	0,004473902	BST2, IFI27, RSAD2, ISG15
GO:0010628	positive regulation of gene expression	0,006136959	CXCL8, CDH3, TNC, HMGA2, INHBA
GO:0001503	ossification	0,006756389	COL11A1, SPP1, MMP9
GO:0042060	wound healing	0,009782769	POSTN, CDH3, TNC
GO:0001649	osteoblast differentiation	0,014152188	SPP1, TNC, CTHRC1
GO:0007189	adenylate cyclase-activating G-protein coupled receptor signaling pathway	0,015243610	CXCL10, CXCL11, PTHLH
GO:0048333	mesodermal cell differentiation	0,017315138	HMGA2, INHBA
GO:0010818	T cell chemotaxis	0,023022334	CXCL10, CXCL11
GO:0045766	positive regulation of angiogenesis	0,023574048	CXCL8, ADAM12, HMGA2
GO:0044267	cellular protein metabolic process	0,024444131	MMP13, MMP1
GO:0033280	response to vitamin D	0,025863933	CXCL10, SPP1

GO:0060339	negative regulation of type I interferon-mediated signaling pathway	0,028697561	MMP12, ISG15
GO:0051281	positive regulation of release of sequestered calcium ion into cytosol	0,044141000	CXCL10, CXCL11
GO:0008284	positive regulation of cell proliferation	0,044542373	CXCL10, TNC, LAMC2, PTHLH
GO:0044344	cellular response to fibroblast growth factor stimulus	0,046923341	POSTN, CXCL8
GO:0045648	positive regulation of erythrocyte differentiation	0,048311582	ISG15, INHBA
GO:0009888	tissue development	0,049697872	POSTN, LAMC2

Destacamos que los 15 procesos biológicos más significativos fueron: Caderina-3, C-X-C Motif Chemokine Ligand 10, ADAM Metallopeptidase Domain 12, Plasminogen Activator, Urokinase, Radical S-Adenosyl Methionine Domain Containing 2, Matrix metallopeptidase 13, Tenascin C, Matrix Metallopeptidase 3, C-X-C Motif Chemokine Ligand 13, Laminin Subunit Gamma 2, Bone Marrow Stromal Cell Antigen 2, Inhibin Subunit Beta A, Secreted Phosphoprotein 1, *Periostin* y S100 Calcium Binding Protein A7A; de los cuales están involucrados genes: SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 (Figura 6). procesos biológicos más significativos fueron como: organización de la matriz extracelular, proceso catabólico del colágeno, desmontaje de la matriz extracelular, diferenciación de células endodérmicas y respuesta a beta-amiloide; de los cuales están involucrados genes: SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 (Figura 6).

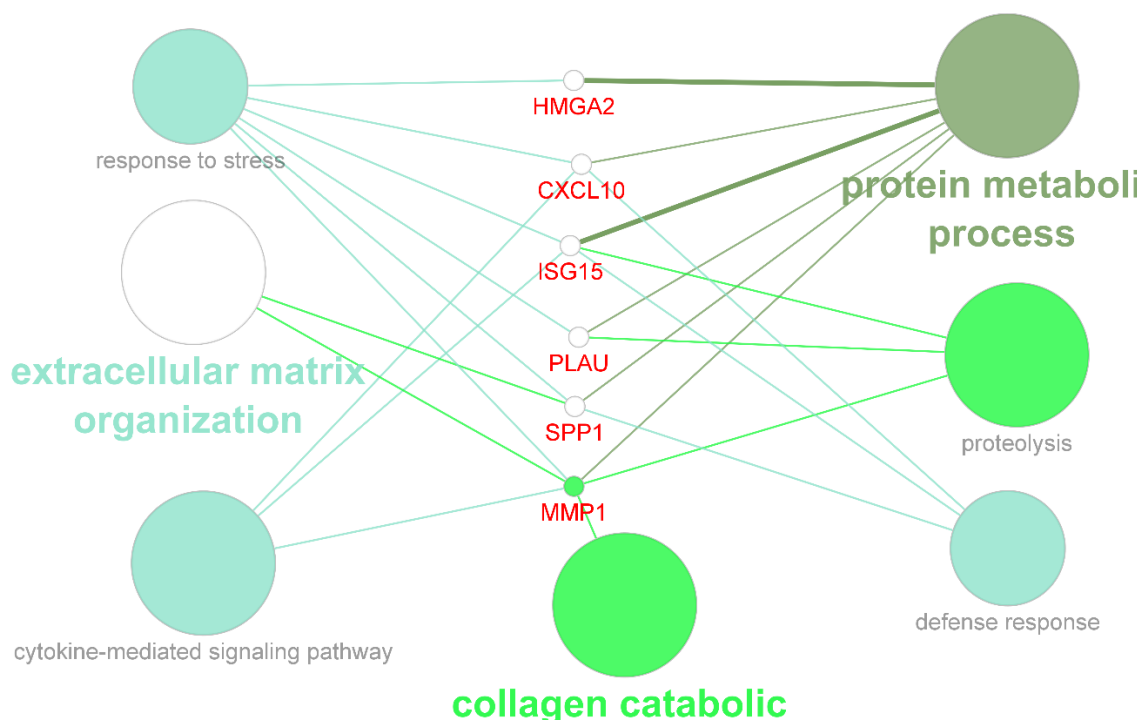


Figura 6. Procesos biológicos relacionados con los genes SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 mediante *software* cytoscape [9].

5. Discusión

El COCE sigue siendo uno de los cánceres más prevalentes en todo el mundo con un número creciente de muertos. Las causas principales atribuida al crecimiento de COCE son la falta de medios de diagnóstico temprano, la alta tendencia metastásica y la resistencia a los medicamentos [10]. Entonces se necesita la exploración de biomarcadores confiables de diagnóstico temprano y pronóstico de COCE. En nuestro estudio hemos detectado 6 genes sobreexpresados SPP1, PLAU, CXCL10, ISG15, MMP1 y HMGA2 en el grupo de cáncer oral comunes en las dos cohortes, importantes para el diagnóstico precoz y para la detección del cáncer oral. Se ha observado como estos genes son relacionándolos con cáncer oral.

El gen SPP1 (fosfoproteína 1 secretada) codifica la osteopontina (OPN). La osteopontina es una fosfoglicoproteína matricelular secretada no colágena, que facilita las interacciones célula-matriz y promueve la progresión tumoral. Además, este gen regula: la proliferación celular, la supervivencia, la resistencia a los fármacos, la invasión y el comportamiento. Su sobreexpresión está directamente relacionada con muchas patologías, incluido los varios tipos de cáncer [11]. En un estudio es particular encontraron que el gen SPP1 sobreexpresado relacionando con la carcinogénesis y la progresión de COCE [12]. Este gen también se vio sobreexpresado, activando la invasión y metástasis de los ganglios linfáticos en cáncer oral [12]. De manera similar, el silenciamiento de SPP1 inhibió la progresión del cáncer de ovario al impedir la vía de señalización de ITG β 1/FAK/AKT, lo que confirma su papel en la carcinogénesis [13]. En conclusión, SPP1 se sobreexpresa en COCE y desempeña un papel clave en la carcinogénesis y la progresión en COCE.

Otro gen relacionado con el cancer oral es PLAU. Esto gen es el receptor de uroquinasa activador de plasminógeno, que es una de las proteínas de membrana ancladas a glicosilfosfatidilinositol (GPI) [14] viendo que inactivamos este gen podría inhibir la proliferación y la metástasis del cáncer en varios tipos de cánceres [15]. PLAU es una es clave para la disolución de fibrina, la degradación de la matriz extracelular, la invasión, la metástasis, la angiogénesis y el escape inmunitario de las células tumorales [16]. La desregulación transcripcional de este gen sugiere su papel dominante en la progresión del cáncer.

Otro gen involucrado es el CXCL10. Rescontramos el alto nivel de esto gen en los tejidos tumorales atrae las células T CXCR3⁺ Th1 y CD8⁺ a los sitios tumorales inflamados y estimula la polarización de las células T en células T efectoras para promover la destrucción de las células cancerosas. Además, CXCL10 participa en la inducción de la apoptosis en células cancerosas. Las altas expresiones de CXCL10 se correlacionan con la activación de las vías de células T y la infiltración tumoral de células T Th1 y CD8⁺ son biomarcadores de buen pronóstico para los pacientes con cáncer oral [17]. Entonces la alta expresión de CXCL10 puede predecir un resultado favorable en pacientes con cáncer oral.

Otro gen involucrado es el ISG15. El ISG15 es una proteína de bajo peso molecular cuya expresión es inducida por interferón. Esto gen participa en la regulación de varios procesos biológicos a través de la unión a proteínas diana. Se han identificado proteínas diana ISG15 en células tumorales o células infectadas por virus, lo que sugiere que ISG15 desempeña funciones importantes en la regulación inmunitaria y el desarrollo tumoral [18].

Previamente se ha demostrado que ISG15 se expresa diferencialmente en diferentes células tumorales y diferentes líneas celulares del mismo origen histológico. Se demostró que la expresión de miR-138 fue significativamente menor en los tejidos cancerosos que en los tejidos adyacentes, lo que indica que la expresión de ISG15 está asociada con la expresión de miR-138. Esto sugiere que miR-138 puede regular los procesos biológicos de COCE mediante la represión de la expresión de ISG15. MiR-138 puede desempeñar un papel nel regular los factores en la vía de la ubiquitina-proteasoma [19]. Por esto el gen ISG15 puede tener aplicaciones potenciales como un nuevo biomarcador para el diagnóstico temprano y el pronóstico de COCE.

En nuestro estudio de expresión diferencial se encontraron varios genes que pertenecientes a la familia de metaloproteinasas de matriz (MMP): MMP1, MMP3, MMP9, MMP10, MMP12, MMP13. La medición de algunas MMP en pacientes con COCE puede examinarse para el diagnóstico y pronóstico de esta enfermedad [20]. Las metaloproteinasas de matriz (MMP) se han estudiado con frecuencia como posibles biomarcadores de cáncer, incluso en COCE, y se han asociado con invasión tumoral y metástasis [21]. La sobreexpresión de los genes MMP1, MMP2, MMP3 y MMP9 se activaron en el tejido tumoral COCE en comparación con el tejido de control [22]. En particular, MMP1 surgió como un marcador excepcionalmente fuerte con respecto a los otros genes de su familia. El nivel de expresión fue de 200 veces mayor en el COCE primario en comparación con el epitelio oral de los controles. Además, la expresión elevada de MMP1 a lo largo de diferentes etapas de COCE y la expresión es creciente, correspondiente a la progresión de tejido normal a displasia a COCE [23]. Este gen es un potencial candidato como biomarcador para para el desarrollo de COCE.

El gen HMGA2 es un factor de transcripción que codifica a proteína del grupo A2 de alta movilidad, es una proteína asociada a la cromatina y permanece ausente en las células somáticas normales. La desregulación de este gen estimula su represión en varios tumores y se encontró que está relacionada con la supervivencia celular, la regulación del ciclo celular, la resistencia a los medicamentos y la disminución de la supervivencia del paciente [24]. HMGA2 se expresa mucho durante la oncogénesis. Los estudios han demostrado que HMGA2 aumenta la progresión del cáncer a través de la activación de varias vías [25]. Se ha observado sobreexpresado en el COCE humano y se sugirió que su expresión actuaba como una herramienta predictiva y pronóstica útil en el tratamiento clínico de los carcinomas orales [26].

Por todo ello nuestro análisis mostró que el grupo de cáncer oral mostro genes y rutas biológicas que podría tener una implicación directa con la progresión, metástasis e invasión del cáncer oral. Así como reafirmamos la identificación genes centrales en dos cohortes diferentes validando nuestros resultados pudiendo proporcionar pistas valiosas para la terapia dirigida en el COCE.

6. Conclusión

Fueron identificados 4.083 genes diferenciales para la cohorte 1 y 5.330 para la cohorte 2 en tejido mucoso para cáncer oral en humanos, obteniendo un total de 2780 genes comunes entre las dos cohortes. Encontramos 103 genes diferenciales comunes (*fold change*

151) de los cuales fueron 31 genes sobreexpresados y 72 infraexpresados para el grupo de 269
 cáncer oral. Mediante el análisis de enriquecimiento de 31 genes diferenciales 270
 sobreexpresados para grupo de cáncer oral se observó “extracelular matrix organization” 271
 (GO:0030198) como el proceso biológico más significativos. Destacamos los genes: SPP1, 272
 PLAU, CXCL10, ISG15, MMP1 y HMGA2, que podrían ser futuros biomarcadores para 273
 diagnosticar el cáncer oral y pudiendo considerarse como dianas terapéuticas. 274
 275

7. Patentes 276

Contribuciones de los autores: Conceptualización, M.S. y G. G.; metodología, G.G.; análisis formal, 277
 G.G.; investigación, MS; recursos MS; redacción—preparación del borrador original, MS y G.G.; 278

Financiamiento: Esta investigación no recibió financiamiento externo 279

Declaración de la Junta de Revisión Institucional: No aplicable para estudios que no involucren 280
 humanos o animales. 281

Declaración de consentimiento informado: No aplicable. 282

Declaración de disponibilidad de datos: Los conjuntos de datos utilizados y/o analizados durante 283
 el presente estudio son 284

disponible en <https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-42743/samples/> y 285

<https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-30784/samples/> 286

Agradecimientos: Los autores desean agradecer a la Universidad Europea de Valencia. 287

Conflictos de interés: Los autores declaran no tener ningún conflicto de interés. 288

Apéndice A 289

Tabla A1. Tabla de los 103 genes diferenciales: 31 genes comunes sobreexpresados entre las 290
 dos cohortes y 72 genes comunes infraexpresados entre las dos cohortes en el grupo de cáncer oral. 291

Names	Total	Elements	Nombre	Cohorte1			Cohorte 2		
				FC	PV	FDR	FC	PV	FDR
C1- sobre expresado C2-sobre expresado	31	CDH3	cadherin 3, type 1, P-cadherin (placental)	5,07	8,75E- 11	1,14E- 08	9,18	2,00E- 43	2,00E- 43
		CXCL10	chemokine (C-X-C motif) ligand 10	6,22	8,21E- 07	2,01E- 05	20	2,25E- 23	2,25E- 23
		ADAM12	ADAM metallopeptidase domain 12	4,42	6,27E- 13	2,03E- 10	4,13	3,84E- 21	1,43E- 19
		PLAU	plasminogen activator, urokinase	2,15	9,31E- 08	3,38E- 06	8,93	9,74E- 41	5,72E- 38

RSAD2	radical S-adenosyl methionine domain containing 2	5,1	7,70E-08	2,89E-06	8,76	3,06E-20	1,01E-18
MMP13	matrix metalloproteinase 13	8,34	2,20E-10	2,40E-08	8,34	2,80E-15	4,91E-14
TNC	tenascin C	6,79	2,55E-13	1,03E-10	8,37	1,00E-34	2,25E-32
MMP3	matrix metalloproteinase 3	21,72	1,04E-10	1,32E-08	72,22	4,43E-33	7,83E-31
CXCL13	chemokine (C-X-C motif) ligand 13	7,39	1,85E-08	8,88E-07	54,81	1,97E-18	5,17E-17
LAMC2	laminin, gamma 2	12,53	2,93E-16	4,85E-13	17,99	3,53E-37	1,21E-34
BST2	bone marrow stromal cell antigen 2	5,2	1,75E-07	5,70E-06	6,99	3,59E-27	2,93E-25
INHBA	inhibin beta A	32,55	2,80E-20	4,44E-16	1,6	4,22E-10	3,60E-09
SPP1	secreted phosphoprotein 1	24,09	4,63E-13	1,59E-10	20,58	1,87E-27	1,58E-25
POSTN	periostin, osteoblast specific factor	5,9	5,59E-11	7,91E-09	22,56	8,69E-37	2,81E-34
S100A7A	S100 calcium binding protein A7	4,1	0,0242	0,0726	9,24	9,42E-15	1,54E-13
PMEPA1	prostate transmembrane protein, androgen induced 1	3,51	3,15E-13	1,21E-10	6,62	4,08E-37	1,39E-34
COL11A1	collagen, type XI, alpha 1	10,42	8,23E-12	1,69E-09	12,27	1,11E-17	2,65E-16
PTH1LH	Parathyroid hormone-like hormone	37,32	2,30E-15	2,37E-12	32,28	2,02E-28	1,97E-26
CTHRC1	collagen triple helix repeat containing 1	8,64	1,93E-17	7,03E-14	16,34	4,46E-43	3,43E-40
MMP10	matrix metalloproteinase 10	31,84	7,63E-14	4,21E-11	36,29	5,22E-27	4,21E-25

	CXCL11	chemokine (C-X-C motif) ligand 11	11,88	2,35E-08	1,08E-06	10,62	9,63E-15	1,57E-13
	MMP1	matrix metalloproteinase 1	99,02	1,37E-17	5,33E-14	358,93	1,40E-45	1,40E-45
	MMP12	matrix metalloproteinase 12	20,96	2,40E-12	6,17E-10	109,44	7,43E-44	6,38E-41
	CXCL8	chemokine (C-X-C motif) ligand 8	5,64	4,46E-06	7,91E-05	40,55	9,49E-35	2,16E-32
	COL12A1	collagen, type XII, alpha 1	2,39	2,91E-05	0,0004	8,13	5,35E-28	4,96E-26
	WDR66	WD repeat domain 66	6,9	3,62E-12	8,74E-10	3,03	4,67E-15	7,92E-14
	HMGA2	high mobility group AT-hook 2	5,17	2,53E-10	2,70E-08	1,83	5,58E-12	6,27E-11
	KRT17	keratin 17, type I	3,35	5,41E-06	9,31E-05	24,43	1,40E-45	1,40E-45
	ISG15	ISG15 ubiquitin-like modifier	9,49	3,24E-13	1,21E-10	11,61	2,43E-31	3,45E-29
	MMP9	matrix metalloproteinase 9	7,71	1,93E-11	3,23E-09	13,64	6,01E-28	5,53E-26
	IFI27	interferon, alpha-inducible protein 27	5,92	2,79E-09	1,91E-07	6,88	2,09E-26	1,56E-24
C1- infra expresado C2- Infra expresado	EMP1	epithelial membrane protein 1	-3,49	7,39E-10	6,42E-08	-4,11	8,90E-20	2,76E-18
	TMPRSS11B	transmembrane protease, serine 11B	- 103,43	4,49E-14	2,73E-11	-86,03	2,04E-28	1,98E-26
	PTN	pleiotrophin	-3,56	1,61E-10	1,89E-08	-5,35	1,10E-19	3,36E-18
	PPL	periplakin	-5,5	2,82E-09	1,93E-07	-5,67	7,99E-28	7,18E-26
	ABCA8	ATP binding cassette subfamily A member 8	-18,19	1,94E-14	1,43E-11	-10,58	1,57E-25	1,04E-23

PITX2	paired-like homeodomain 2	-8,23	6,37E-15	5,53E-12	-5,93	2,69E-07	1,49E-06
CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	-3,69	1,04E-11	1,96E-09	-9,52	1,56E-35	4,15E-33
CHRD1	chordin-like 1	-8,5	2,63E-10	2,79E-08	-8,73	7,31E-23	3,36E-21
TTN	titin	-4,49	0,0061	0,0256	-1,77	1,30E-07	7,61E-07
ALOX12	arachidonate 12-lipoxygenase	-6,02	1,22E-10	1,50E-08	-5,68	3,83E-18	9,73E-17
FLG	filaggrin	-9,08	8,70E-05	0,0009	-22,39	2,71E-15	4,77E-14
ADH7	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	-15,95	7,68E-12	1,61E-09	-10,78	1,58E-21	6,14E-20
ADH1B	alcohol dehydrogenase 1B (class I), beta polypeptide	-32,03	6,29E-11	8,70E-09	-15,12	2,44E-36	7,42E-34
FAM3B	family with sequence similarity 3, member B	-14,43	3,07E-16	4,94E-13	-17,32	3,10E-25	1,98E-23
FCER1A	Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide	-7,62	7,58E-13	2,35E-10	-5,48	8,29E-25	4,97E-23
ENDOU	endonuclease, polyU-specific	-17,22	1,41E-13	6,51E-11	-26,12	2,83E-40	1,50E-37
OGN	osteoglycin	-2,2	0,0004	0,003	-5,17	7,17E-11	6,85E-10
SCARA5	scavenger receptor class A, member 5	-4,6	4,63E-13	1,59E-10	-4,25	8,22E-25	4,93E-23
CLCA4	chloride channel accessory 4	-14,05	1,83E-07	5,88E-06	-24,01	2,78E-25	1,78E-23
ATP1A2	ATPase, Na ⁺ /K ⁺ transporting, alpha 2 polypeptide	-3,23	0,0047	0,021	-6,35	5,55E-14	8,20E-13
CSTB	cystatin B (stefin B)	-8,87	1,57E-11	2,77E-09	-8,02	1,06E-33	2,05E-31
ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	-1,94	8,95E-09	4,96E-07	-6,57	2,76E-30	3,46E-28

SLC16A6	solute carrier family 16, member 6	-2,72	4,44E-14	2,72E-11	-3,83	3,47E-27	2,83E-25
KRT4	keratin 4, type II	-58,99	6,99E-11	9,48E-09	-48,05	4,26E-23	2,04E-21
PPP1R3C	protein phosphatase 1, regulatory subunit 3C	-1,76	1,04E-06	2,43E-05	-15,07	2,85E-37	9,97E-35
SLURP1	secreted LY6/PLAUR domain containing 1	-5,1	8,42E-08	3,11E-06	-12,35	8,87E-19	2,43E-17
KLK13	kallikrein related peptidase 13	-6,93	2,05E-07	6,49E-06	-4,72	1,11E-16	2,36E-15
KRT13	keratin 13, type I	-31,88	3,57E-07	1,02E-05	-8,73	1,91E-13	2,60E-12
MB	myoglobin	-6,77	0,0003	0,0021	-7,84	6,93E-08	4,21E-07
TGM3	transglutaminase 3	-96,84	3,09E-12	7,67E-10	-26,21	2,77E-22	1,17E-20
ACTA1	actin, alpha 1, skeletal muscle	-7,24	0,0017	0,0095	-24,26	1,25E-06	6,21E-06
A2ML1	alpha-2-macroglobulin-like 1	-11,73	3,59E-08	1,54E-06	-9,11	9,90E-23	4,47E-21
GALNT12	polypeptide N-acetylgalactosaminyltransferase 12	-1,98	5,66E-10	5,18E-08	-6,48	1,33E-34	2,90E-32
EXPH5	exophilin 5	-4,38	5,83E-06	9,90E-05	-6,33	3,14E-28	2,98E-26
SPINK7	serine peptidase inhibitor, Kazal type 7 (putative)	-36,26	4,48E-08	1,85E-06	-12,78	7,25E-15	1,20E-13
CYP2C18	cytochrome P450, family 2, subfamily C, polypeptide 18	-6,8	1,08E-07	3,79E-06	-6,51	4,94E-23	2,34E-21
ALDH3A1	aldehyde dehydrogenase 3 family, member A1	-10,01	3,97E-10	3,91E-08	-7,61	2,41E-17	5,53E-16
CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5	-5,68	0,0001	0,0012	-8,61	1,61E-17	3,75E-16
CNFN	cornifelin	-6,78	0,007	0,0284	-5,72	1,92E-13	2,61E-12
SCIN	scinderin	-6,78	0,007	0,0284	-12,29	1,40E-45	1,40E-45

KLHL41	kelch-like family member 42	-2,07	0,0347	0,0951	-22,58	1,59E-06	7,78E-06
PYGM	phosphorylase, glycogen, muscle	-9,66	6,42E-06	0,0001	-5,05	3,36E-18	8,63E-17
TMPRSS11E	transmembrane protease, serine 11E	-8,13	2,20E-06	4,48E-05	-10,65	3,81E-19	1,10E-17
ACPP	acid phosphatase, prostate	-1,59	9,24E-06	0,0001	-10,57	8,22E-39	3,71E-36
HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	-6,29	3,23E-13	1,21E-10	-3,93	1,59E-20	5,40E-19
CAPN14	calpain, small subunit 2	-8,67	7,67E-09	4,38E-07	-37,45	1,11E-28	1,12E-26
PAX9	paired box 9	-2,68	4,35E-15	3,98E-12	-4,49	5,17E-33	9,02E-31
MYBPC1	myosin binding protein C, slow type	-6,51	0,0005	0,0035	-9,04	8,61E-07	4,40E-06
TNNC2	troponin C type 2 (fast	-9,05	0,0005	0,0037	-7,87	8,42E-09	5,91E-08
RHCG	Rh family, C glycoprotein	-13,12	1,04E-06	2,43E-05	-10,69	2,31E-24	1,31E-22
SCEL	sciellin	-18,04	1,70E-06	3,60E-05	-14,06	4,54E-32	6,93E-30
ALDH1A1	aldehyde dehydrogenase 1 family, member A1	-12,01	3,52E-14	2,32E-11	-7,88	3,98E-16	7,80E-15
GBP6	guanylate binding protein family, member 6	-14,21	9,66E-10	7,94E-08	-11,01	9,00E-29	9,24E-27
CKM	creatine kinase, muscle	-7,5	0,0019	0,0102	-13,65	2,45E-08	1,60E-07
KRT78	keratin 78, type II	-10,9	1,05E-11	1,98E-09	-27,23	2,27E-41	1,43E-38

MAL	mal, T-cell differentiation protein	-79,44	1,08E-12	3,20E-10	-49,38	3,86E-29	4,13E-27
BNIP1L	BCL2/adenovirus E1B 19kD interacting protein like	-5,69	2,07E-05	0,0003	-4,11	8,43E-26	5,82E-24
GPD1L	glycerol-3-phosphate dehydrogenase 1-like	-5,74	4,55E-14	2,73E-11	-5,87	2,45E-43	2,03E-40
VSIG10L	V-set and immunoglobulin domain containing 10 like	-9,01	2,74E-07	8,18E-06	-9,52	2,18E-23	1,08E-21
CRNN	cornulin	-87,73	5,08E-14	3,02E-11	-54,37	6,09E-25	3,73E-23
HRASLS	HRAS-like suppressor	-3,99	3,11E-10	3,19E-08	-6,13	2,58E-16	5,20E-15
CRISP3	cysteine-rich secretory protein 3	-91,32	1,64E-12	4,44E-10	-	6,40E-33	1,10E-30
SAMD5	sterile alpha motif domain containing 5	-6,83	7,13E-15	6,06E-12	-2,33	8,90E-25	5,31E-23
HLF	hepatic leukemia factor	-7,42	1,53E-20	4,17E-16	-8,24	8,80E-39	3,87E-36
CFD	complement factor D (adipsin)	-11,28	1,34E-09	1,04E-07	-6,43	7,15E-25	4,34E-23
MYH2	myosin, heavy chain 2, skeletal muscle, adult	-11,35	0,0003	0,0023	-41,02	7,46E-08	4,52E-07
CYP4B1	cytochrome P450, family 4, subfamily B, polypeptide 1	-7,23	3,54E-18	1,93E-14	-20,28	8,28E-27	6,52E-25
CEACAM7	carcinoembryonic antigen-related cell adhesion molecule 7	-7,11	1,35E-06	3,00E-05	-4,03	1,54E-12	1,87E-11
HOPX	HOP homeobox	-2,42	8,04E-09	4,53E-07	-6,64	4,03E-21	1,49E-19
MUC15	mucin 15, cell surface associated	-6,16	2,24E-07	6,93E-06	-6,09	1,79E-20	6,05E-19
SPINK5	serine peptidase inhibitor, Kazal type 5	-10,52	5,49E-07	1,45E-05	-11,5	9,48E-24	4,92E-22

CRCT1	cysteine rich C-terminal 1	-9,78	0,0002	0,0018	-6,83	1,26E-	2,02E-
						14	13

Referencias

- Izaguirre Pérez PM. Nivel de conocimiento sobre cáncer oral de estudiantes del último año de la carrera profesional de odontología de tres universidades de Lima, en el año 2012. Lima – Perú; 2012.
- García-García V, Bascones Martínez A. Cáncer oral: Puesta al día. *Av Odontostomatol.* 2009;25(5):239–48.
- Equipo de redactores y equipo de editores médicos de la Sociedad Americana Contra El Cáncer. ¿En qué consisten los tipos de cáncer de orofaringe y de cavidad oral? 2018.
- Antonio Miguel Cruz P, Aluett Niño Peña I, Kelia Batista Marrero I, Angulo Holguín M. Factores de riesgo de cáncer bucal Risk factors for oral cancer. Vol. 53, *Revista Cubana de Estomatol.* 2016.
- Serna García M. Estudio del tejido ovárico para mejorar la producción a través del tamaño de camada en conejos. 2016 Dec.
- Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009;25(8):1091–3.
- Transcriptome analysis console (TAC) software - UK.
- Sterck L. Draw Venn diagram. Ugent.be.
- Ono K. Cytoscape. Cytoscape.org.
- Choi HS, Kim Y-K. Upregulation of MDR- and EMT- related molecules in cisplatin-resistant human oral squamous cell carcinoma cell lines. *Int J MolSci.* 2019;20.
- Shevde LA, Samant RS. Role of osteopontin in the pathophysiology of cancer. *Matrix Biol.* 2014;37:131–41.
- Huang C-F, Yu G-T, Wang W-M, Liu B, Sun Z-J. Prognostic and predictive values of SPP1, PAI and caveolin-1 in patients with oral squamous cell carcinoma. *Int J Clin Exp Pathol.* 2014;7(9):6032–9.
- Zeng B, Zhou M, Wu H, Xiong Z. SPP1 promotes ovarian cancer progression via Integrin β 1/FAK/AKT signaling pathway. *Onco Targets Ther.* 2018;11:1333–43.
- Ploug M, Ronne E, Behrendt N, Jensen AL, Blasi F, Dano K. Cellular receptor for urokinase plasminogen activator. Carboxyl-terminal processing and membrane anchoring by glycosyl-phosphatidylinositol. *J BiolChem.* 1991;266(3):1926–33.
- Zhou J, Kwak KJ, Wu Z, Yang D, Li J, Chang M, Song Y, Zeng H, Lee LJ, Hu J, et al. PLAUR Confers Resistance to Gefitinib Through EGFR/P-AKT/Survivin Signaling Pathway. *Cell PhysiolBiochem.* 2018;47(5):1909–24.
- He F, Chen H, Probst-Kepper M et al (2012) PLAU inferred from a correlation network is critical for suppressor function of regulatory T cells. *MolSystBiol* 8:624.
- Huang Y-C, Huang J-L, Tseng L-C, Yu P-H, Chen S-Y, Lin C-S. High expression of interferon pathway genes CXCL10 and STAT2 is associated with activated T-cell signature and better outcome of oral cancer patients. *J Pers Med.* 2022; 12(2):140.
- Andersen JB, Hassel BA. The interferon regulated ubiquitin-like protein, ISG15, in tumorigenesis: friend or foe? *Cytokine Growth Factor Rev.* 2006;17(6):411–21.
- Zhang Q, He Y, Nie M, Cai W. Roles of miR-138 and ISG15 in oral squamous cell carcinoma. *Exp Ther Med.* 2017;14(3):2329–34.
- Mishev G, Deliverska E, Hlushchuk R, Velinov N, Aebersold D, Weinstein F, et al. Prognostic value of matrix metalloproteinases in oral squamous cell carcinoma. *Biotechnol Biotechnol Equip.* 2014;28(6):1138–49.

-
21. Ziober FA, Patel CR, Gimotty F, Weber P, Feldman milímetro, et al Identificación de una firma genética para la detección rápida del carcinoma de células escamosas orales. *Clin Cancer Res.* 2006;12:5960–71. 331
332
 22. Baker EA, Leaper DJ, Hayter JP, Dickenson AJ. The matrix metalloproteinase system in oral squamous cell carcinoma. *Br J Oral Maxillofac Surg.* 2006;44(6):482–6. 333
334
 23. Stott-Miller M, Houck JR, Lohavanichbutr P, Méndez E, Upton MP, Futran ND, et al. Tumor and salivary matrix metalloproteinase levels are strong diagnostic markers of oral squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2011;20(12):2628–36. 335
336
337
 24. Yadav M, Pradhan D, Singh RP. Integrated analysis and identification of nine-gene signature associated to oral squamous cell carcinoma pathogenesis. *3 Biotech.* 2021;11(5):215. 338
339
 25. Mansoori B, Mohammadi A, Ditzel HJ, Duijf PHG, Khaze V, Gjerstorff MF, et al. HMGA2 as a critical regulator in cancer development. *Genes (Basel).* 2021;12(2):269. 340
341
 26. Sterenczak KA, Eckardt A, Kampmann A, Willenbrock S, Eberle N, Länger F, et al. HMGA1 and HMGA2 expression and comparative analyses of HMGA2, Lin28 and let-7 miRNAs in oral squamous cell carcinoma. *BMC Cancer.* 2014;14(1):694. 342
343
344
345