



# **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**DOCTORADO EN CIENCIAS BIOMÉDICAS  
FACULTAD DE MEDICINA  
INSTITUTO NACIONAL DE MEDICINA GENÓMICA**

**ANÁLISIS DE PERFILES DE EXPRESIÓN DE RNAs LARGOS NO  
CODIFICANTES Y SU RELACIÓN CON CARACTERÍSTICAS CLÍNICAS Y  
BIOLÓGICAS DE TUMORES DE MAMA.**

## **TESIS**

**QUE PARA OPTAR POR EL GRADO DE:  
DOCTOR EN CIENCIAS BIOMÉDICAS**

**PRESENTA:  
M en C ALBERTO CEDRO TANDA**

**DIRECTOR DE TESIS  
DR. ALFREDO HIDALGO MIRANDA  
FACULTAD DE MEDICINA**

**COMITÉ TUTOR**

**DR. ALEJANDRO M. GARCÍA CARRANCA  
INSTITUTO DE INVESTIGACIONES  
BIOMÉDICAS**

**DR. CARLOS G. PÉREZ PLASCENCIA  
FACULTAD DE ESTUDIOS SUPERIORES  
IZTACALA**

**CIUDAD DE MÉXICO, DICIEMBRE DE 2020**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

Dedico este trabajo a todos los pacientes oncológicos, que la investigación aquí plasmada contribuya a seguir desenmascarando al tenaz *carcinos...*

## **AGRADECIMIENTOS**

A mi familia, por su ejemplo y cariño.

Al Dr. Alfredo Hidalgo, por su guía durante mi doctorado.

Al INMEGEN, por brindarme recursos e infraestructura para llevar a término mi proyecto de investigación.

A mis amigos que se convirtieron en parte de mi familia, que enriquecieron mi pensamiento y que me ayudaron a ser mejor persona.

A mis colegas que me apoyaron durante mi estancia en el laboratorio.

## RESUMEN

El cáncer de mama es la neoplasia más comúnmente diagnosticada en mujeres en todo el mundo, es reconocida como una patología heterogénea y se clasifica en cuatro subtipos moleculares: Luminal A, Luminal B, HER2-enriched y Basal-like, cada uno con diferentes características biológicas y clínicas. Los RNAs largos no codificantes (lncRNAs) representan el 33% del transcriptoma humano y desempeñan funciones críticas en la carcinogénesis de mama, pero la mayoría de sus funciones aún se desconocen. Por lo tanto, la investigación del cáncer podría beneficiarse de la exploración de la biología de los lncRNAs en esta neoplasia. En este trabajo caracterizamos los perfiles de expresión de lncRNAs en 75 tumores de mama pertenecientes a los cuatro subtipos moleculares utilizando microarreglos. Para inferir el papel biológico de los lncRNA desregulados en los subtipos moleculares, realizamos análisis de coexpresión de lncRNA-mRNA y análisis de gene ontology. Identificamos 307 lncRNA desregulados en el tumor en comparación con el tejido normal y 354 lncRNA desregulados entre los diferentes subtipos moleculares. A través del análisis de coexpresión entre los lncRNA y los mRNAs, junto con el análisis de enriquecimiento de genes, inferimos la función potencial de los lncRNA más desregulados en cada subtipo molecular y validamos de forma independiente nuestros resultados utilizando los datos de TCGA. Se observó sobreexpresión del lncRNA AC009283.1 en el subtipo HER2-enriched, localizado en una zona de amplificación en el cromosoma 17q12. Se investigó el papel funcional del lncRNA AC009283.1 mediante ensayos de pérdida de función *in vitro* y se determinó su impacto en la expresión génica global. Estos estudios mostraron que AC009283.1 regula genes implicados en la proliferación, el ciclo celular y la apoptosis en un modelo celular HER2-enriched. Además, confirmamos estos hallazgos a través del análisis de ssGSEA y CEMITool en una cohorte independiente de tumores de mama HER2 positivos. Nuestros hallazgos sugieren un amplio repertorio de funciones biológicas para los lncRNAs en cada subtipo molecular de cáncer de mama y proporcionan una base para su estudio biológico y funcional, como se realizó para AC009283.1, demostrando que es un regulador potencial de la proliferación y apoptosis en el subtipo HER2-enriched.

## ABSTRACT

Breast cancer is the most commonly diagnosed neoplasm in women worldwide with a well-recognized heterogeneous pathology, classified into four molecular subtypes: Luminal A, Luminal B, HER2-enriched and Basal-like, each one with different biological and clinical characteristics. Long non-coding RNAs (lncRNAs) represent 33% of the human transcriptome and play critical roles in breast carcinogenesis, but most of their functions are still unknown. Therefore, cancer research could benefit from continued exploration into the biology of lncRNAs in this neoplasm. We characterized lncRNA expression portraits in 74 breast tumors belonging to the four molecular subtypes using transcriptome microarrays. To infer the biological role of the deregulated lncRNAs in the molecular subtypes, we performed co-expression analysis of lncRNA–mRNA and gene ontology analysis. We identified 307 deregulated lncRNAs in tumor compared to normal tissue and 354 deregulated lncRNAs among the different molecular subtypes. Through co-expression analysis between lncRNAs and protein-coding genes, along with gene enrichment analysis, we inferred the potential function of the most deregulated lncRNAs in each molecular subtype, and independently validated our results taking advantage of TCGA data. Overexpression of the AC009283.1 was observed in the HER2-enriched subtype and it is localized in an amplification zone at chromosome 17q12. The functional role of lncRNA AC009283.1 was examined through loss of function assays *in vitro* and determining its impact on global gene expression. These studies revealed that AC009283.1 regulates genes involved in proliferation, cell cycle and apoptosis in a HER2 cellular model. We further confirmed these findings through ssGSEA and CEMITool analysis in an independent HER2-amplified breast cancer cohort. Our findings suggest a wide range of biological functions for lncRNAs in each breast cancer molecular subtype and provide a basis for their biological and functional study, as was conducted for AC009283.1, showing it to be a potential regulator of proliferation and apoptosis in the HER2-enriched subtype.

## INDICE

<b>Introducción</b> .....	<b>1</b>
EL CÁNCER DE MAMA COMO UN RETO DE SALUD MUNDIAL .....	1
BIOLOGÍA MOLECULAR DEL CÁNCER. ....	1
BIOLOGÍA MOLECULAR DEL CÁNCER DE MAMA .....	3
CLASIFICACIÓN Y FUNCIÓN DE LOS LNCRNAS .....	4
LNCRNAS Y SU ASOCIACIÓN CON CÁNCER.....	6
<b>Justificación, Hipótesis</b> .....	<b>8</b>
<b>Objetivo general</b> .....	<b>9</b>
<b>Objetivos particulares</b> .....	<b>9</b>
<b>Materiales y métodos</b> .....	<b>9</b>
<b>Resultados</b> .....	<b>17</b>
LOS SUBTIPOS MOLECULARES ESTÁN ASOCIADOS A DISTINTOS DESENLACES CLÍNICOS	17
LA EXPRESIÓN DE LNCRNAS ESTA DESREGULADA EN CÁNCER DE MAMA .....	19
PERFIL DE EXPRESIÓN DE LNCRNAS EN SUBTIPOS MOLECULARES.....	20
PATRONES DE EXPRESIÓN DE LNCRNAS EN SUBTIPO LUMINAL A .....	22
PATRONES DE EXPRESIÓN DE LNCRNAS EN SUBTIPO LUMINAL B .....	22
PATRONES DE EXPRESIÓN DE LNCRNAS EN SUBTIPO BASAL-LIKE .....	22
PATRONES DE EXPRESIÓN DE LNCRNAS EN SUBTIPO HER2-ENRICHED .....	23
EL LNCRNA AC009283.1 EN TUMORES DEL SUBTIPO HER2-ENRICHED .....	24
EL SILENCIAMIENTO DE AC009283.1 MODIFICA EL PERFIL TRANSCRIPCIONAL DE LA LÍNEA CELULAR SK-BR-3 DE FENOTIPO HER2-ENRICHED.....	28
EL SILENCIAMIENTO DE AC009283.1 INHIBE LA PROLIFERACIÓN E INDUCE APOPTOSIS EN LA LÍNEA CELULAR SK-BR-3 DE FENOTIPO HER2-ENRICHED .....	30
AC009283.1 SE ENCUENTRA SOBREPRESADO EN TUMORES HER2 POSITIVOS Y REGULA PROCESOS ONCOGÉNICOS.....	33
ANÁLISIS MODULAR DE EXPRESIÓN DE AC009283.1 EN TUMORES HER2 POSITIVOS ....	36
<b>Discusión</b> .....	<b>38</b>
<b>Conclusiones y perspectivas</b> .....	<b>46</b>
<b>Referencias</b> .....	<b>47</b>

## **INTRODUCCIÓN**

### **El cáncer de mama como un reto de salud mundial**

Las neoplasias se están incrementado significativamente de forma global, siendo uno de los principales retos de salud pública que enfrentan los países. Al año 2018 37 millones de personas vivían con cáncer (prevalencia a 5 años), 31.5 millones fueron diagnosticados y 8.8 millones murieron a causa de esta enfermedad, posicionando al cáncer como la segunda causa de muerte en el mundo. La incidencia y mortalidad por neoplasias es proporcional al índice humano de desarrollo [1].

El cáncer de mama (CM) es uno de los tumores más comunes en mujeres de todo el mundo ya que 1 de 4 tumores femeninos diagnosticados corresponde a este órgano. En 2018, 2,088,840 mujeres fueron diagnosticadas con esta neoplasia y 626,679 fallecieron, se estima que en el 2030 una de cada ocho mujeres en el mundo desarrolle CM. La incidencia de esta enfermedad está aumentando particularmente en países con economías emergentes como México, en donde cada día 76 mujeres reciben el diagnóstico de esta neoplasia, acumulando 27,830 casos en 2018. Lamentablemente 6,884 mujeres fallecen debido a esta enfermedad, representando la segunda causa de muerte en mujeres mexicanas [1].

El sistema de salud público y privado mexicano cuenta con métodos de diagnóstico y tratamiento para el CM, desarrollados a partir de la investigación básica y clínica, sin embargo, se sigue indagando en los factores que determinan su aparición, desarrollo y progresión, por lo que sigue considerándose un reto social, médico y científico.

### **Biología molecular del cáncer**



La biología molecular a logrado importantes avances en la comprensión del cáncer. Nos encontramos en el mayor periodo de desarrollo de la ciencia y de la tecnología molecular y genómica, esto ha contribuido a develar mecanismos biológicos y moleculares y a generar aplicaciones en la clínica.

El cáncer es un conjunto de enfermedades de origen multifactorial cuyas principales capacidades biológicas son: mantenimiento de la señalización de proliferación celular, evasión de la muerte celular, ilimitado potencial replicativo, inducción de angiogénesis y activación de la invasión y metástasis, aunado a esto la inestabilidad genómica genera diversidad genética que acelera la adquisición de nuevas alteraciones a través del tiempo [2]. Así mismo, la reprogramación del metabolismo energético y la evasión de la destrucción inmunológica son características importantes para el desarrollo neoplásico [3] (Figura 1).



**Figura 1. Características distintivas del cáncer.** Las células cancerosas adquieren capacidades que les hacen tener un fenotipo tumoral determinante para su desarrollo y supervivencia.

La iniciación y progresión del cáncer se deben a múltiples procesos moleculares, las células cancerosas comparadas con las células normales presentan diferencias a nivel genómico (amplificaciones, deleciones, inserciones, mutaciones y diversos re-arreglos cromosómicos); a nivel transcriptómico (transcripción aberrante de genes) y epigenético (metilación de DNA y modificaciones de histonas). Estos cambios juegan un papel importante en el control de la regulación de la maquinaria molecular de las células neoplásicas.

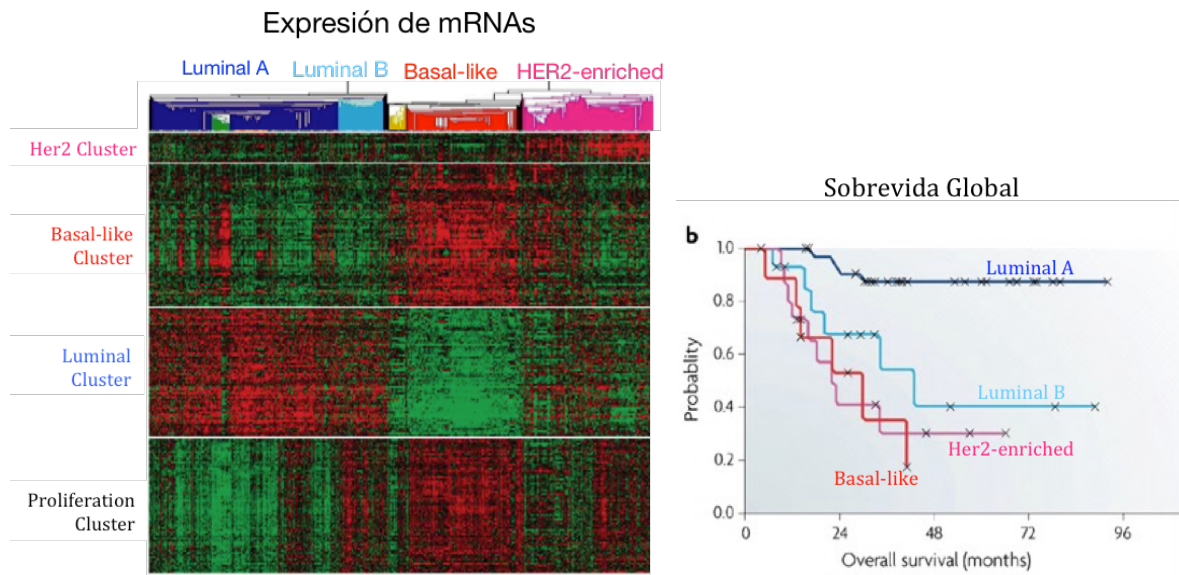
### **Biología molecular del cáncer de mama**

El CM es una enfermedad compleja y heterogénea. Los tumores de mama se clasifican en grupos clínicamente relevantes según la expresión de tres marcadores inmunohistoquímicos: receptores de estrógeno, progesterona y HER2 (receptor del factor de crecimiento epidérmico humano 2) [4]. Los pacientes cuyos receptores hormonales son positivos tienen un mejor pronóstico en comparación con los que son receptores negativos, además los receptores son de gran valor en la elección de terapia adyuvante y para el tratamiento de enfermedad avanzada.

La presencia de receptores HER2 en celular tumorales ha tenido implicaciones en la selección de tratamiento, el pronóstico y la monitorización. Las pacientes con receptor HER2 positivo se han asociado a un fenotipo muy agresivo, incremento de recurrencia y disminución de supervivencia, en general tienen mal pronóstico.

Los tumores de mama pueden subdividirse aún más según las firmas de expresión de RNA mensajero (mRNA). Una de estas firmas intrínsecas es PAM50, identifica subgrupos basados en la expresión de 50 genes Luminal A, Luminal B, HER2-enriched y basal-like [5, 6]. Más importante aún, cada uno de estos subtipos de tumores se ha asociado con comportamientos biológicos y clínicos particulares. En comparación con los subtipos luminales, los subtipos basal-like y presentan un peor pronóstico a cinco años, y el subtipo HER2-enriched tiene la peor sobrevida

libre de recaída [7]. La clasificación de subtipos moleculares ha ayudado a los oncólogos a seleccionar la terapia mas adecuada para los pacientes. Varias líneas de evidencia indican que estos subtipos moleculares reflejan no solo la alteración de los niveles de expresión de mRNA, sino también las alteraciones genéticas y epigenéticas que contribuyen al establecimiento y mantenimiento de diferentes procesos oncogénicos [8].



**Figura 2. Clasificación de subtipo moleculares de CM.** Obtenidos a partir de su perfil de expresión de mRNA. La clasificación se realizó usando el algoritmo PAM50 para obtener cuatro subtipos moleculares clínicamente relevantes.

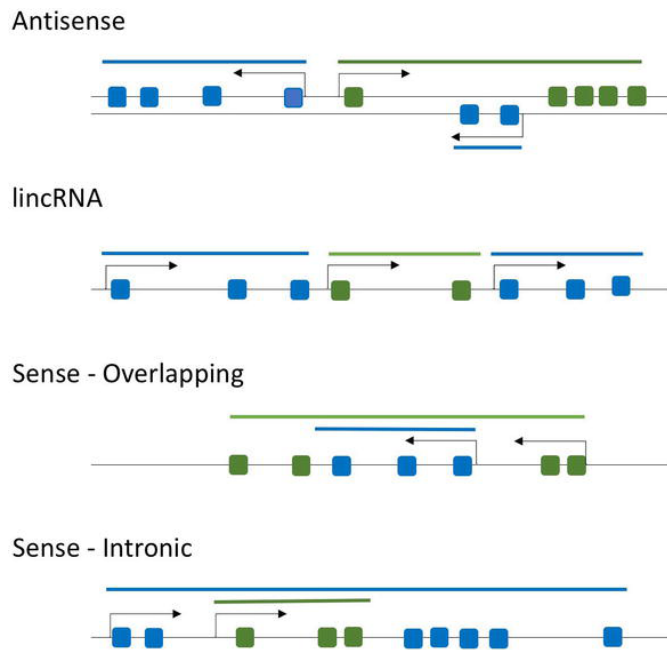
La clasificación del CM en subtipos moleculares se basó en la parte codificante del transcriptoma, que representa el 33% del transcriptoma humano, el 67% restante se compone por transcritos no codificantes, que a su vez se divide en dos clases: RNAs pequeños no codificantes de <200 nucleótidos (miRNAs, piRNAs, piwiRNAs) y los RNAs largos no codificantes (lncRNAs) de mas de 200 nucleótidos de longitud hasta 10kb [9]

### Clasificación y función de los lncRNAs

Los lncRNAs son transcritos sin potencial codificante, cuya longitud es mayor a 200 nucleótidos, son transcritos mediante la Polimerasa II presentan adición del nucleótido 7-metil guanosina o también llamado cap en el extremo 5', y

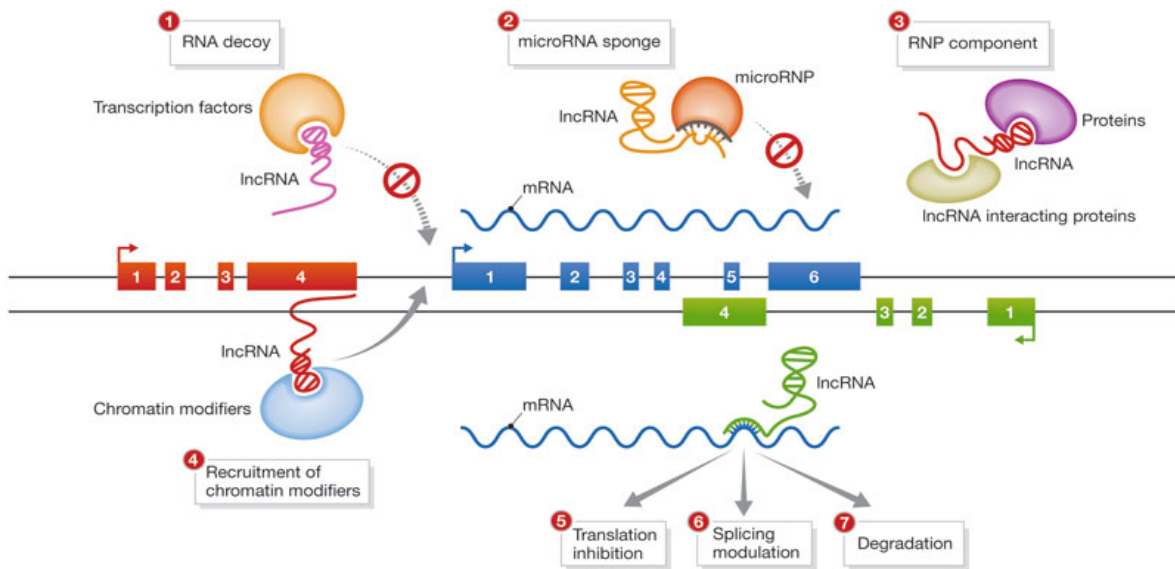
poliadenilación en el extremo 3', tienen proceso de corte y empalme. Se clasifican de acuerdo a su posición genómica y función biológica.

Los lncRNAs se localizan en regiones intergénicas (lincRNAs) e intragénicas. Los lncRNAs intragénicos pueden ser antisentidos, sentido sobrelapados y sentido intrónicos, mientras que los lincRNAs se encuentran entre los genes codificantes, y son los mas comunes a lo largo del genoma (Figura 3).



**Figura 3. Clasificación de lncRNAs por su localización.** Existen lncRNAs antisentidos, lincRNAs, sentido sobrelapados y sentido intrónicos. Los bloques azules representan a los exones de mRNAs y los verdes a los exones de los lncRNAs. La dirección de la transcripción esta indicada por las flechas.

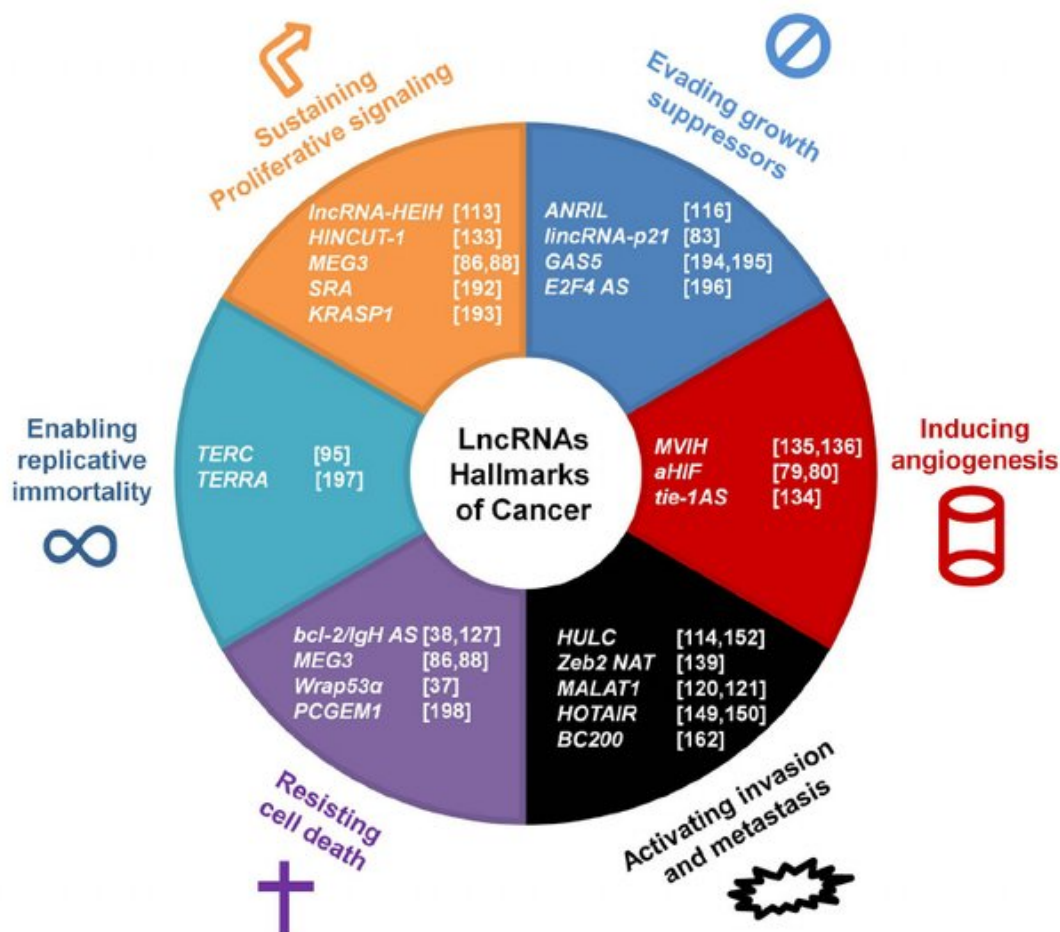
Las funciones de los lncRNAs son diversas, dependiendo del contexto celular e histológico en donde se presenten. Se sabe que las funciones de regulación de los lncRNAs se deben a su capacidad para interactuar con proteínas, DNA y RNA siendo así idóneos para el remodelado de la cromatina, la regulación de dominios génicos, traducción, etc. En la figura 4 se muestra como los lncRNAs regulan la expresión génica a través de siete mecanismos actuando a nivel pre y postranscripcional.



**Figura 4. Mecanismos de acción de los lncRNAs.** 1. Señuelo. 2. Esponja de microRNAs. 3. Componente de ribonucleoproteínas. 4. Reclutamiento de modificadores de la cromatina. 5. Inhibición de la traducción. 6. Modulación del corte y empalme y 7. Degradación.

## **lncRNAs y su asociación con cáncer**

Los lncRNAs tienen importantes papeles biológicos: regulan la expresión génica a nivel epigenético, transcripcional y post-transcripcional. Existe evidencia de que los perfiles de expresión de los lncRNAs en células cancerosas son diferentes a los encontrados en células del tejido adyacente normal. [10]. Cada vez hay mas evidencia de que los lncRNA juegan un papel en la aparición y desarrollo del cáncer, de hecho, los lncRNA están implicados en las características distintivas del cáncer que se propusieron por primera vez en 2001 (Figura 5).



**Figura 5. Los lncRNAs impactan en las características distintivas del cáncer.** Las seis características del cáncer se muestran con lncRNAs que están implicados en la aparición y progresión del cáncer. Las referencias se enumeran entre paréntesis. Imagen tomada de [A pathophysiological view of the long non-coding RNA world.](#)

En CM los lncRNAs emergen como reguladores maestros de la biología tumoral con funciones oncogénicas y supresoras de tumor relacionadas a la iniciación y progresión de la enfermedad. Algunos ejemplos muy estudiados de lncRNAs oncogénicos son los siguientes: H19, SRA, LSINCT5, Zfas1, SMAD7, LOC554202 y UCA1 involucrados en la proliferación y apoptosis; mientras que HOTAIR en invasión y metástasis. Respecto a la troncalidad en CM, los lncRNAs SOX2OT y FAL1 tienen papeles importantes como reguladores y promotores de este fenotipo. Por otro lado, los lncRNAs con capacidad de supresores tumorales son: GAS5 y

XIST involucrados en proliferación/apoptosis e invasión/metástasis, respectivamente. PMC5651513. DOI 10.1007/s12282-014-0554-y Además se ha observado que HOTAIR, MALAT-1, lincRNAp21 y GAS5, están asociados a características clínicas como la sobrevida global, sobrevida libre de progresión y riesgo de presentar metástasis [11, 12].

Los análisis de la expresión de lncRNAs en CM han revelado que existe un patrón diferencial entre tumores definidos por IHC (HER2 positivos y triples negativos) [13]. Otros estudios han evaluado la expresión aberrante de lncRNAs en subtipos tumorales de CM identificando patrones de expresión en cada subtipo [14, 15]. Aun existe poca información acerca del impacto de la expresión desregulada de los lncRNAs en cada subtipo molecular de CM, particularmente en aquellos con mal pronóstico como el subtipo molecular HER2-enriched.

## **JUSTIFICACIÓN**

Para continuar dilucidando el panorama molecular del cáncer de mama es importante identificar, estudiar y caracterizar a los lncRNAs mediante tecnologías de alto rendimiento. En este trabajo se generó nuevo conocimiento acerca de los lncRNAs desregulados en cáncer de mama en los subtipos moleculares, centrando esfuerzos en los tumores del subtipo HER2-enriched ya que estos pacientes presentan la menor tasa de sobrevida global.

De tal forma que estamos contribuyendo a describir la función biológica de un lncRNA y como su desregulación impacta en diversos procesos biológicos.

## **HIPÓTESIS**

Los lncRNAs tendrán un perfil de expresión diferente entre subtipos tumorales de mama los cuales se relacionarán con sus características biológicas.

## **OBJETIVO GENERAL**

Analizar los perfiles de expresión de RNAs largos no codificantes y su relación con características clínicas y biológicas de tumores de mama.

## **OBJETIVOS PARTICULARES**

- Explorar el panorama transcripcional de los lncRNAs en los cuatro subtipos moleculares de CM y validarlo en una cohorte independiente.
- Identificar lncRNAs sobrepresados en cada subtipo molecular e inferir su relevancia biológica en el fenotipo tumoral.
- Seleccionar un lncRNA en el subtipo molecular HER2-enriched y describir su función biológica mediante abordajes *in vitro* e *in silico*.

## **MATERIALES Y MÉTODOS**

*Banco de tumores de mama.* Se utilizaron 75 muestras de tumores de mama (Luminal A =24, Luminal B=23, HER2-enriched=14 y Basal-like=13) del biobanco de tumores de mama del Laboratorio de Genómica del Cáncer, INMEGEN. Estos tumores fueron colectados en el Instituto de Enfermedades de la Mama, FUCAM entre los años 2008-2015. Ninguna de las pacientes recibió quimioterapia neoadyuvante. Se utilizó también tejido adyacente a 2cm del margen del tumor. Todos los tejidos fueron congelados en nitrógeno líquido y almacenados a -80°C hasta su uso. La evaluación histológica se realizó por dos patólogos, solo se utilizaron muestras de mas de 80% de células tumorales para el análisis. Este estudio fue aprobado por el comité de ética e investigación del Instituto Nacional de Medicina Genómica (CE2009/11). Este protocolo de investigación fue realizado de acuerdo a las guías y regulación internacionales de estudios en sujetos humanos. Cada paciente firmó el consentimiento informado para aceptar participar en el protocolo de investigación.



*Análisis de sobrevida global.* Se utilizaron datos clínicos de las bases de datos de The Cancer Genome Atlas (TCGA) y METABRIC. Los datos clínicos fueron analizados usando regresión multivariada Cox, el riesgo relativo también fue calculado. La sobrevida libre de enfermedad de la cohorte mexicana fue analizada con el modelo Kaplan-Meier. El nivel de significancia para todas las pruebas fue <0.05. Se utilizó el software PASW (SPSS, IBM).

Modelos celulares. Las líneas celulares de CM incluidas fueron: MCF7, ZR-75-1, MDA-MB-361, SK-BR-3, MDA-MB-468, HCC1187, HS578T, MDA-MB-231, MDA-MB-453 y la línea no tumorigénica MCF10A, fueron adquiridas de American Type Culture Collection (ATCC, Manassas, VA, USA). Las líneas celulares se crecieron de acuerdo a las guías de ATCC., Fueron suplementadas con 10% de suero fetal bovino en una incubadora a 37 °C con 5% de CO<sub>2</sub>.

Extracción de RNA. Los tejidos tumorales congelado fueron disgregados usando un TissueRuptor (Qiagen Inc, Valencia, CA) y la extracción fue realizada con DNA-RNA AllPrep System (Qiagen Inc, Valencia, CA). El RNA de las líneas celulares fue extraído con Trizol (Invitrogen). La concentración del RNA fue evaluada por espectrofotometría (NanoDrop Technologies, Wilmington, Delaware), Y la integridad del RNA fue analizada con BioAnalyzer 2100 (Agilent Technologies, Palo Alto, CA). Sólo las muestras con un rin mayor a ocho fueron usadas para el análisis de microarreglos. El RNA total fue almacenado a -80 °C hasta su procesamiento.

*Reacción en cadena de la polimerasa acoplada a retrotranscripción (RT-qPCR).* El cDNA fue sintetizado usando SuperScript III RT-PCR (Invitrogen) siguiendo las recomendaciones del fabricante. 100 nanogramos (ng) de RNA total de cada línea celular fueron utilizadas para sintetizar cDNA en una reacción final de 20µL. La mezcla para PCR contuvo: 1µL of cDNA, 5 µL 2X TaqMan Universal Master Mix (Applied Biosystems, Cat 4304437), 0.5 µL TaqMan probes (Hs\_01590788 custom from AC009283.1) y 3.5 µL de agua grado biología molecular. Los genes *GAPDH*

(Hs99999905) y *SCARNA* (Hs03391742\_cn) fueron usados como controles endógenos.

*Microarreglos de expresión.* Se utilizó el microarreglo Human Transcriptome Array 2.0 HTA2.0 (Affymetrix, Inc, Santa Clara, CA) el cual evalúa la expresión de mRNAs y lncRNAs. Se utilizaron 200ng de RNA total usando el kit de hibridación se WT Plus Reagent Kit protocol, siguiendo las instrucciones del fabricante. El procesamiento se realizó en la Unidad de Microarreglos del INMEGEN.

*Análisis de microarreglos.* Los datos de expresión fueron analizados con los programas: Affymetrix's Expression y Transcriptome Analysis. Los datos crudos de los microarreglos se encuentran públicamente disponibles en GEO con número de acceso: GSE134254. De acuerdo a la anotación de Affymetrix, el microarreglo HTA2.0 contiene 44,699 genes codificantes y 22,829 genes no codificantes, sin embargo solo el 62% de los genes (codificantes y no codificantes) tenían anotado un nombre del gen, el resto (38%) no tenían anotado ningún gen ya sea codificante o no codificante. Para poder tener una anotación del microarreglo de mayor calidad, decidimos re-anotarlo con la herramienta BioMart de Ensemble, de esta forma identificamos que el microarreglo es capaz de detectar 10,153 lncRNAs. (supplementary data 1.4).

*Clasificación de subtipos moleculares.* Las muestras tumorales de pacientes mexicanas fueron clasificadas en subtipos moleculares de acuerdo a su perfil de expresión de mRNA, utilizando el algoritmo PAM50 [5].

*Expresión diferencial.* Primeramente se comparó el perfil de expresión génica (lncRNAs y mRNAs) de tumores (n=75) respecto a tejidos adyacentes (n=13). Posteriormente se comparó la expresión génica de cada subtipo molecular respecto al resto de los subtipos. Se utilizó la prueba de ANOVA de una vía. Se consideraron a los genes diferencialmente expresados cuando tenían un fold

change  $>2.0$ ,  $<-2.0$ , un p-value de ANOVA  $<0.05$ , y un FDR (false discovery rate)  $<0.05$ .

*Validación de perfiles de expresión de lncRNAs en cohortes independientes.* Para validar nuestros hallazgos usamos datos de expresión de RNA-Seq de lncRNAs nivel 3 de TCGA. Para la validación del perfil de expresión de tumores respecto a tejido adyacente se usaron datos de RNA-Seq de The Cancer lncRNome Atlas (TCLA) [16]. Para la validación de los perfiles de expresión de subtipos moleculares usamos datos de TCGA en the Atlas of Non-Coding RNA (TANRIC).

*Análisis de coexpresión.* Para identificar la coexpresión de los pares lncRNA-mRNA en datos de nuestra cohorte y de TCGA usamos correlación de Pearson empleando la herramienta. El valor de corte del coeficiente Pearson fue de 0.4 y un p-value  $<0.05$ . El análisis de Gene ontology (GO) para los mRNAs coexpresados con los lncRNAs de interés fue realizado con la aplicación ClueGO app en Cytoscape.

*Análisis de datos genómicos para tumores HER2.* Para RNA-Seq se descargaron los conteos crudos (HT-Seq) de la base de datos de Xena (<https://xenabrowser.net/datapages/>). Los datos fueron anotados con biomaRt[17], Gencode v33 (<https://www.gencodegenes.org/human/>) y LNCipedia [18]. Para el análisis de expresión diferencial, se consideró un gen expresado si tuvo más de 10 conteos, la y expresión diferencial se realizaron con DESeq2 en DESeq2 package en R[18], se analizaron independientemente los mRNAs y los lncRNAs. Para el cálculo de número de copias se descargaron los archivos de segmentación de la base de datos Xena y para determinar las regiones significativamente amplificadas se realizó GISTIC GISTIC (Genomic Identification of Significant Targets in Cancer) 2.0 [19, 20] el cual fue hecho en el servidor de Genepattern (<https://cloud.genepattern.org/gp/pages/index.jsf>). Para el análisis de microarreglos de expresión se descargaron seis proyectos (GSE20711, GSE29431, GSE5460, GSE17907, GSE29044, and GSE66305) desde Gene Expression Omnibus para el

microarreglo Human Genome U133 Plus 2.0 de Affymetrix, se seleccionaron solo los tumores HER2+ por IHC.

*Genes contenidos dentro del amplicon 17q-21.* Se utilizó el análisis de GISTIC para identificar los genes contenidos en el amplicon 17q-21. Después se realizó un análisis de neighboring core usando GenomicRanges package [21]. Se definieron dos categorías de acuerdo a la proximidad de los genes core: 1) genes cercanos: el vecino mas cercano y 2) genes distantes: el vecino mas distante. Se consideraron los genes solo si se expresaban en al menos el 50% de las muestras.

*Identificación de módulos de coexpresión y red de consenso.* Los tumores HER2+ con diferentes niveles de expresión de AC009283.1 fueron clasificados como: baja expresión (AC009283.1 niveles <1er cuantil), expresión media (>1er cuantil y <3er cuantil) y alta expresión (>3er cuantil). Los conteos normalizados con DeSeq2 de los genes codificantes y de AC009283.1 fueron analizados en la herramienta CEMiTool [22] en ambiente R usando parámetros predeterminados. Se evaluó el coeficiente de correlación de Pearson entre cada par de genes de los diferentes niveles de expresión de AC009283.1. Los genes observados en cada módulo de CEMiTool fueron conectados en una red, usando solo los genes con Pearson >30% y P <0.05. Para cada módulo de CEMiTool se realizo un GSEA usando los genes significativos en cada módulo [23] se utilizaron las siguiente bases para alimentar el GSEA: C1-Hallmark gene sets, C2-Kegg curated genes, and C5GO-BP gene sets obtenidas de Molecular Signature Database (MSigDB) (<http://www.broad.mit.edu/gsea/>).

*Implementación de ssGSEA.* Se calculó el puntaje individual de enriquecimiento de genes para cada tumor por separado implementado en el paquete GSVA en ambiente R [24]. El grupo de genes evaluados fue: proliferation UNC337 intrinsic clustering, hallmark apoptosis, GO regulation of cell cycle g2 m phase transition,

GO negative regulation of cell cycle arrest y GO cell cycle g2 m phase transition. Los puntajes de ssGSEA fueron graficados en la paquetería ggplot en ambiente R.

*Fraccionamiento celular.* Se realizó separación de núcleo y citoplasma de la línea celular SKBR-3 utilizando el kit Ambion PARIS (Thermo Fisher, Cat AM1921) de acuerdo a las instrucciones el fabricante. Se realizó RT-qPCR para los genes GAPDH, MALAT1 y SCARNA5.

*Silenciamiento del lncRNA AC009283.1.* Se diseñaron short hairping RNAs (shRNAs) específicamente dirigidos hacia el lncRNA AC009283.1 usando el sistema de silenciamiento BLOCK-iT U6 RNAi Entry Vector (Invitrogen, Cat K494500) siguiendo las instrucciones del fabricante. Se generó un oligo de doble cadena con las regiones complementarias al lncRNA y posteriormente se ligó en el vector pENTR /U6, con este vector se transformaron bacterias competentes *E. coli* One Shot TOP10. Para corroborar la presencia y correcta orientación del vector con el shRNA usamos secuenciación Sanger.

Se usaron dos diseños de shRNA para el silenciamiento de AC009283.1 (shRNA1 and shRNA2). La secuencias del shRNA1 fueron: top 5'CACCGAGGAAGTAGGTTAGGGAATGCGAACATTCCCTAACCTACTTCCTC3', bottom

5'AAAAGAGGAAGTAGGTTAGGGAATGTTTCGCATTCCCTAACCTACTTCCTC3' y del shRNA2: top 5'CACCGCAAATAGGTGTCTCATAGCTCGAAAGCTATGAGACACCTATTTGC3' and bottom 5'AAAAGCAAATAGGTGTCTCATAGCTTTTCGAGCTATGAGACACCTATTTGC3'.

La secuencia del shRNA control negativo (NC) fueron: top: 5'CACCGGAATTACGGAGTCTTCTTCGCGAACGAAGAAGACTCCGTAATTCC3' and bottom: 5'AAAAGGAATTACGGAGTCTTCTTCGTTTCGCGAAGAAGACTCCGTAATTCC3').

El silenciamiento del AC009283.1 en la línea celular SKBR-3 se realizó con el shRNA1 y el shRNA2 usando el kit Xfect Transfection Reagent Protocol-At-A-Glance (Clontech, Palo Alto, CA, Cat 631318),  $3.5 \times 10^5$  células SK-BR-3 fueron sembradas en placas de cultivo celular de 6 pozos, 3 $\mu$ g de vector con cada uno de los shRNAs fueron transfectados durante 48 horas. El transfectante con los shRNAs fue removido completamente y reemplazado por medio McCoy5A (ATCC). Después de 96h se evaluó la expresión de AC009283.1 mediante RT-qPCR. El RNA fue extraído con el uso de Trizol (Thermo Scientific, Cat 15596026) y almacenado  $-80^{\circ}\text{C}$ . Se realizaron tres replicas biológicas y con sus respectivas réplicas técnicas.

El efecto del silenciamiento de AC009283.1 en el transcriptoma de la línea celular SK-BR-3 fue analizado con el microarreglo HTA2.0 (Affymetrix, Santa Clara, USA) como se describió arriba. Los datos crudos de los microarreglos se encuentran públicamente disponibles en GEO con número de acceso: GSE134359. Realizamos tres microarreglos para el shRNA control (NC) y tres para el shRNA2 para el silenciamiento de AC009283.1. Se realizó expresión diferencial, los genes con fold change  $>1.2$  y  $<-1.2$  y p-value  $<0.05$  fueron considerados como significantes y seleccionados para el análisis de procesos biológicos usando Ingenuity Pathway Analysis (IPA). Para explorar el efecto de AC009283.1 en muestras de tumores realizamos análisis de expresión diferencial de los tumores HER2-enriched de TCGA con alta y baja expresión de AC009283.1 (primer y cuarto cuartil de expresión). Las tablas de conteos de RNA-Seq fueron obtenidos del portal de TCGA, la normalización y la expresión diferencial se realizó con DESeq2, los genes con fold change  $>1.2$  y  $<-1.2$  y p-value  $<0.05$  fueron considerados como significantes y usados para el análisis de procesos biológicos con IPA.

Proliferación celular. Se realizó en las células SK-BR-3 transfectadas con el shRNA2 y con el shRNA NC para evaluar el efecto del silenciamiento de AC009283.1. Se sembraron  $7.7 \times 10^4$  células SK-BR-3 en placas de cultivo de 24

pozos. Se utilizaron dos métodos, en el primer método se contabilizaron las células teñidas con azul tripano usando el contador automático TC20 Automated Cell Counter (BIO-RAD), en el segundo método se usó el kit Cell Trace CFSE Cell Proliferation (Invitrogen, Cat C34554) y el citómetro de flujo (Attune, Applied Biosystems). La proliferación celular se midió a las 24, 48, 72 y 96hrs, con tres réplicas biológicas y sus respectivas réplicas técnicas.

Ensayo de apoptosis. Se realizó en las células SK-BR-3 transfectadas con el shRNA2 y con el shRNA NC para evaluar el efecto del silenciamiento de AC009283.1. Se sembraron  $7.7 \times 10^4$  células SK-BR-3 en placas de cultivo de 24 pozos. Se utilizaron dos métodos, el primero fue mediante citometría de flujo (Attune, Applied Biosystems) usando con Anexina V y Ioduro de propidio IP (FITC Annexin V/ Dead cell Apoptosis kit, Invitrogen, Carlsbad, CA, USA), en este método las células fueron tratadas con QBS 20 $\mu$ M por 24hrs, después fueron teñidas con Anexina V y IP por 15 minutos y analizadas en citómetro de flujo. Las células se categorizaron en apoptosis temprana y tardía. El segundo ensayo de apoptosis fue realizado usando caspasa 3 siguiendo las recomendaciones del fabricante (Merck #235419). Usamos QBS 50 $\mu$ M como control positivo de apoptosis y el software FlowJo T v10.0 para el análisis. La apoptosis se midió a las 96hrs, con tres réplicas biológicas y sus respectivas réplicas técnicas.

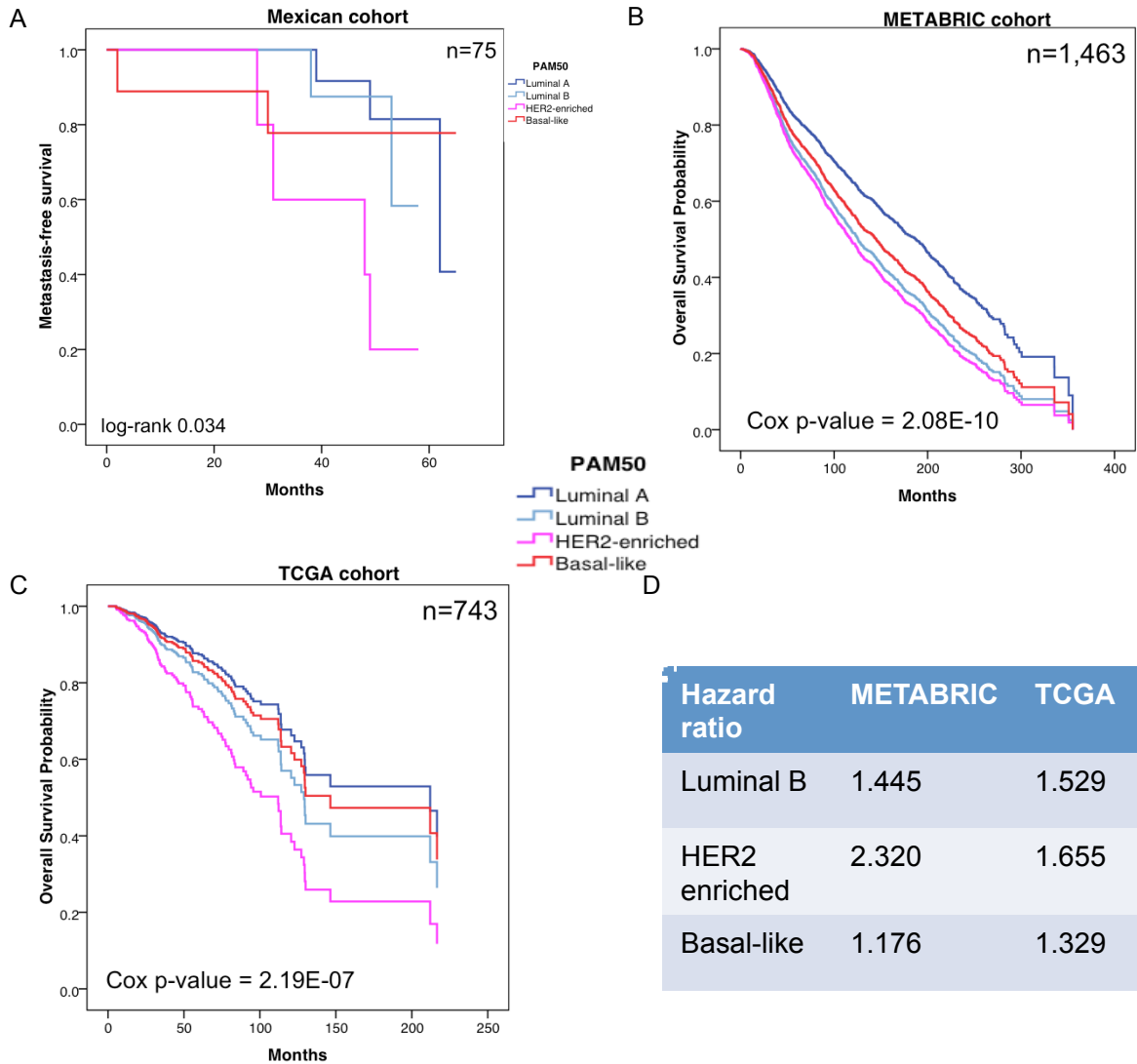
Análisis estadístico. La significancia estadística fue calculada con GraphPadPrism (version 6, San Diego, USA). ANOVA y t de Student fueron realizadas para todas las comparaciones. La correlación entre variables fue determinada mediante coeficiente de Pearson. El método de Kaplan-Meier fue realizado mediante log-rank test y COX. Un p-value <0.05 fue considerado como significativo (\* $p$ <0.05, \*\* $p$ <0.005, \*\*\* $p$ < 0.0005).

## RESULTADOS

### Los subtipos moleculares están asociados a distintos desenlaces clínicos

La incidencia de subtipos moleculares en nuestra cohorte fue de: 32.4% Luminal A (n=24), 31.1% Luminal B (n=23), 18.9% HER2-enriched (n=14), y 17.6% Basal-like (n=13). El desenlace clínico es diferente entre cada subtipo molecular, en nuestra cohorte observamos que el subtipo molecular HER2-enriched mostró la mayor probabilidad de metástasis (log-rank  $p=0.034$ ). En dos cohortes independientes observamos que el subtipo HER2-enriched mostró la menor tasa de supervivencia global: METABRIC (n=1,463) (95% CI [1.417-2.368] Cox p-value=  $2.08E-10$ ) y TCGA (n=743) (95% CI [0.97-4.01] Cox p-value=  $2.19E-7$ ) (Figura 1). Las características clínicas y patológicas de nuestra cohorte se encuentran en la Tabla 1.





**Figura 1. Desenlaces clínicos en subtipos moleculares de CM en tres cohortes.** A. Sobrevida libre de metástasis en la cohorte mexicana de este estudio. B. Sobrevida global en la cohorte de METABRIC. C. Sobrevida global en la cohorte de TCGA. D. Hazard ratio en subtipos moleculares Luminal B, HER2-enriched y Basal-like usando a Luminal A como referencia.

<b>Características clínicas</b>		<b>n = 75</b>
<b>Edad (años)</b>		
<40	4	5.3%
40-60	36	75%
60-80	20	26.7%
>80	3	4%
Desconocido	12	16%
<b>Receptores hormonales (IHC)</b>		
<b>RE</b>		
Positivo	50	66.7%
Negativo	20	26.7%
Desconocido	5	6.7%
<b>RP</b>		
Positivo	37	49.3%
Negativo	33	44%
Desconocido	5	6.7%
<b>HER2</b>		
Positivo	12	16%
Negativo	55	73.3%
Desconocido	8	10.7%
<b>Subtipo molecular (PAM50)</b>		
Luminal A	24	32%
Luminal B	24	32%
HER2-enriched	14	19%
Basal-like	13	17%
<b>Etapa clínica</b>		
I	5	6.7%
II	50	66.7%
III	6	8%
IV	1	1.3%
Desconocido	13	17.3%
<b>Grado tumoral</b>		
G1	3	4%
G2	37	49.3%
G3	14	18.7%
G4	0	0%
Desconocido	21	28%
<b>Diagnóstico histopatológico</b>		
Carcinoma ductal invasivo	56	74.7%
Otros	6	8%
Desconocido	13	17.3%
<b>Sobrevida global</b>		
<5 años	52	69.3%
>5 años	5	6.7%
Desconocido	18	24%
<b>Metástasis</b>		
Yes	11	14.7%
No	37	49.3%
Desconocido	27	36%

**Tabla 1. Características clínicas de la cohorte en este estudio.**

## **La expresión de lncRNAs esta desregulada en cáncer de mama**

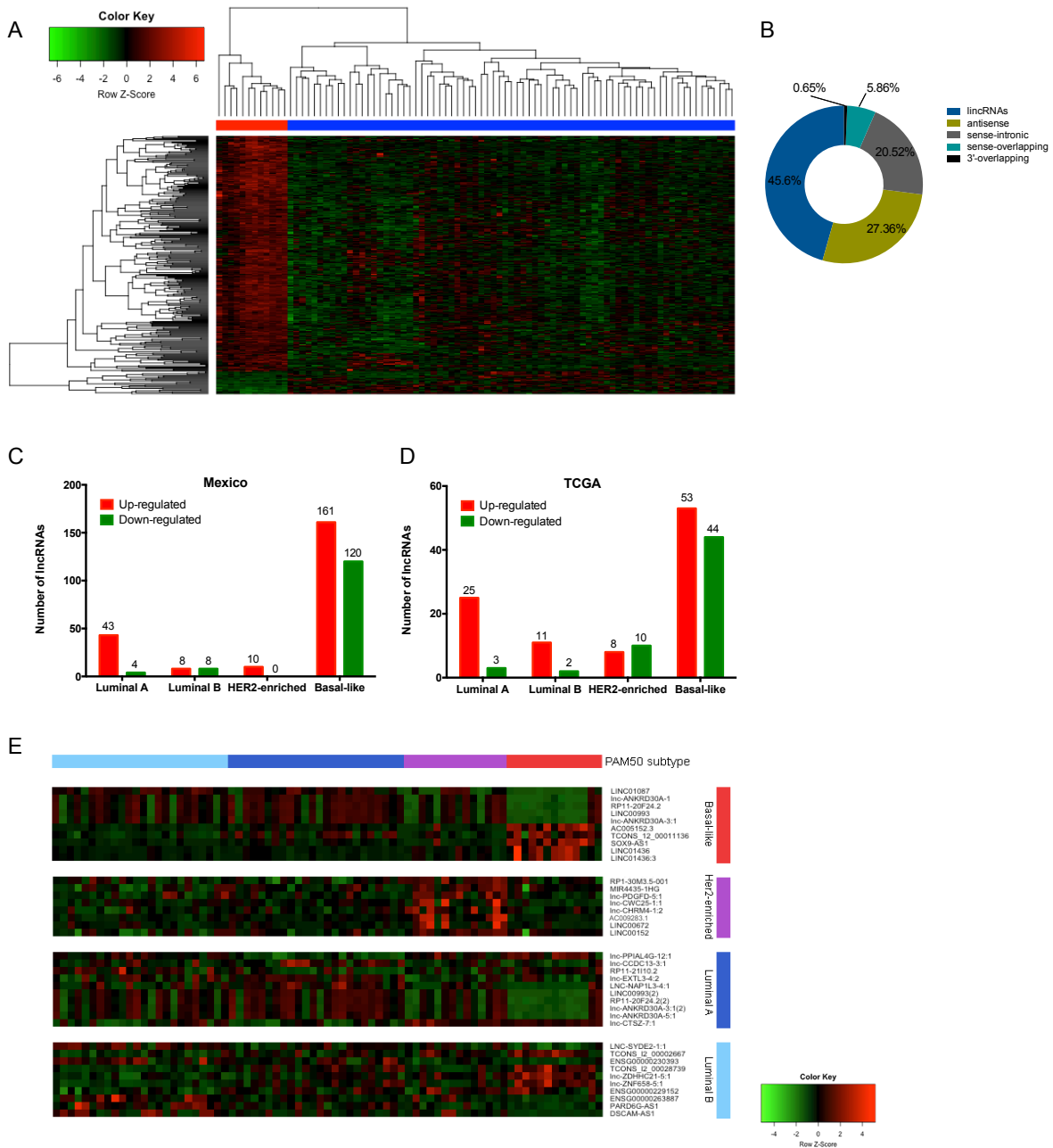
El análisis del panorama transcripcional entre 74 tumores de mama y 12 tejidos adyacentes reveló que 307 lncRNAs están diferencialmente expresados (27 sobrexpresados y 280 subexpresados), con un fold change >2.0 y <-2.0, (false discovery rate) FDR <0.05. El análisis del agrupamiento jerárquico basado en los 307 lncRNAs diferencialmente expresados fue optimo para distinguir a los tumores

de los tejidos adyacentes, como se muestra en el mapa de calor de la figura 2A). Los lncRNAs alterados fueron clasificados en las siguientes categorías de acuerdo a su posición genómica: lincRNAs (45.6%), antisentido (27.36%), intrónicos sentido (20.52%), sobrepuesto sentido (5.86%) y 3'-sobrepuesto (0.65%) (Figure 2B).

Varios lncRNAs alterados en este estudio han sido previamente reportados y asociados a la carcinogénesis de la mama, tales como: HOTAIR, HOTAIRM1, XIST, PANDAR, EPB41L4A-AS1, BC040587, FGF14-AS2, DSCAM-AS1, LINC00472, MIR31HG y FAM83H-AS1.

### **Perfil de expresión de lncRNAs en subtipos moleculares**

Para continuar entendiendo la biología molecular de los subtipos, exploramos el panorama transcripcional de los lncRNAs en los subtipos PAM50. Encontramos 354 lncRNAs diferencialmente expresados: 47 en Luminal A, 16 en Luminal B, 10 en HER2-enriched y 281 en Basal-like. (Figura 2C). Los datos de TCGA mostraron una proporción similar de lncRNAs diferencialmente expresados en cada subtipo molecular (Figura 2D). La figura 2E muestra los 10 lncRNAs con mayor fold change en cada subtipo molecular, 5 sobrepresados y 5 subexpresados en Basal-like, Luminal A, Luminal B, y 8 sobre expresados en HER2-enriched.



**Figure 1. Perfiles de expresión de lncRNAs en tumores de mama.** **A.** Agrupamiento jerárquico y mapa de calor de 307 lncRNAs con expresión diferencial entre tumores (barra azul) y tejido normal (barra roja). **B.** Tipo de lncRNAs desregulados de acuerdo a su clasificación por posición genómica. **C.** lncRNAs desregulados en subtipos moleculares en la cohorte mexicana (fold change >1.5, <-1.5, FDR <0.05) y **D.** en la cohorte de TCGA (fold change log 1.0. adj. P-value <0.05). **E.** Los 10 lncRNAs más desregulados a través de los subtipos moleculares en la cohorte mexicana Basal-like en rojo, Her2-enriched en morado, Luminal A en azul cielo and Luminal B en azul marino.

### **Perfil de expresión de lncRNAs en subtipo Luminal A**

En nuestra cohorte identificamos 47 lncRNAs diferencialmente expresados (43 sobreexpresados y 4 subexpresados). Para definir más robustamente los lncRNAs desregulados en este subtipo examinamos la cohorte de TCGA. Encontramos cuatro lncRNAs sobreexpresados en nuestra cohorte y en TCGA: LINC00504, LINC00993, MIR205HG y PCAT18. Para averiguar la posible función biológica de estos 4 lncRNAs investigamos los patrones de coexpresión lncRNA-mRNA encontramos en ambas cohortes 340 mRNAs significativamente correlacionados con estos 4 lncRNAs ( $R > 0.3$ ,  $p\text{-value} < 0.05$ ). El análisis de Gene Ontology (GO) de estos 340 mRNAs mostró su asociación con regulación positiva de la transcripción de DNA en respuesta a estrés, movimiento y organización basado en microtúbulos, entre otros (Figure 3A).

### **Perfil de expresión de lncRNAs en subtipo Luminal B**

Observamos que en nuestra cohorte existen 16 lncRNAs desregulados, 8 con sobreexpresión y 8 con subexpresión. En ambas cohortes (este estudio y TCGA) encontramos 3 lncRNAs sobreexpresados: LINC01087, RP11-379F12.4 y GATA3-AS1. Estos 3 lncRNAs tuvieron correlación significativa con la expresión de 70 mRNAs en esta cohorte y en TCGA. De acuerdo con el análisis de GO, estos mRNAs participan en la regulación de la acetilación de histona H3-K14 y la diferenciación de células progenitoras (Figure 3B). Además en la cohorte de TCGA, la sobreexpresión de LINC01087 y RP11-379F12.4 estuvo asociada al incremento en la supervivencia global (log-rank test  $< 0.05$ ).

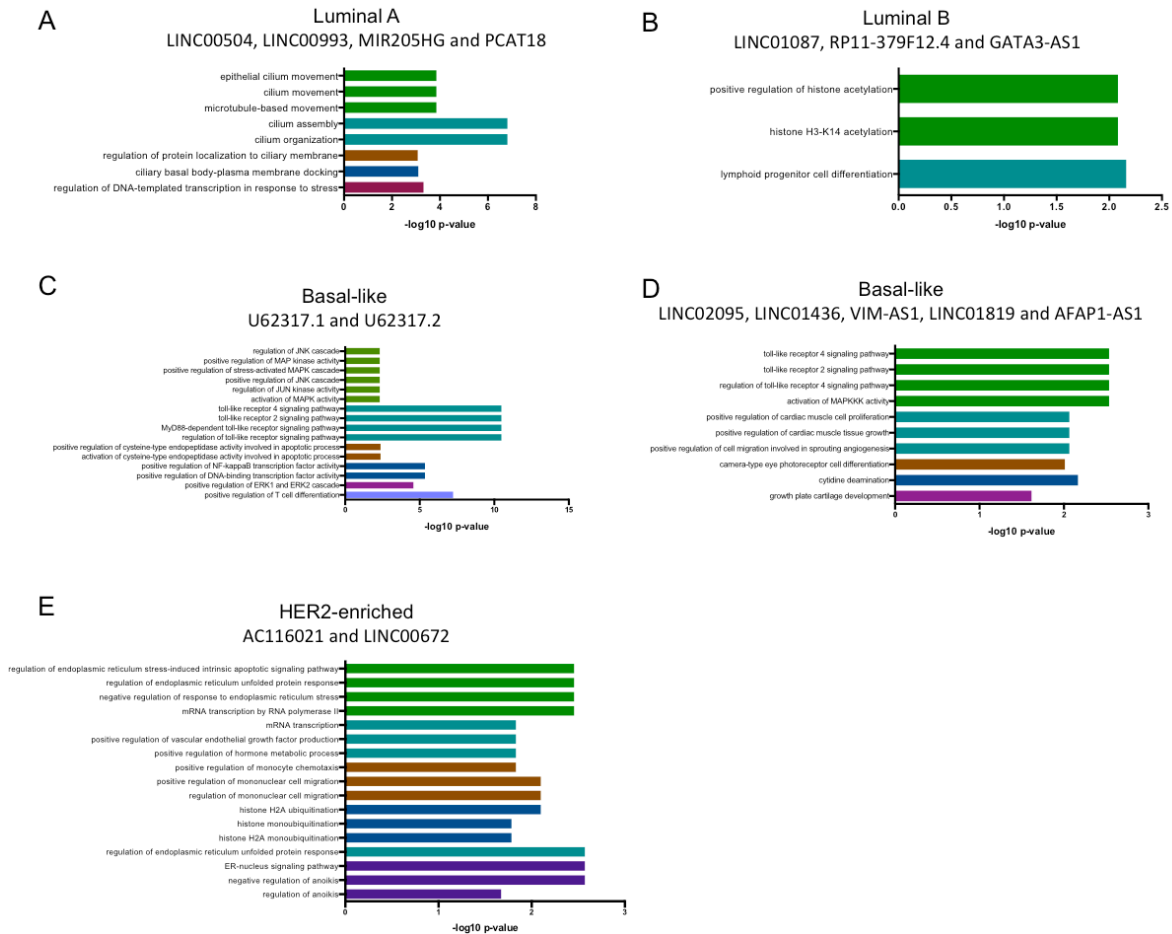
### **Perfil de expresión de lncRNAs en subtipo Basal-like**

El subtipo molecular Basal-like mostró el mayor número de lncRNAs diferencialmente expresados, con 161 sobreexpresados y 120 subexpresados en

nuestra cohorte. De los cuales solo en un bajo numero se ha estudiado su implicación en cáncer, tales como CASC15, LINC00662, LINC01272, LINC00857, LINC00092, LINC00673, sobreexpresados en esta cohorte y LINC00657, LINC01087 y LINC00993, subexpresados en esta cohorte. Estudiamos mas a detalle a U62317.1 y U62317.2 ya que son lncRNAs sobreexpresados en nuestra cohorte y en TCGA, su expresión correlaciona con la de 419 mRNAs que en conjunto están implicados en regulación positiva de la cascada de JNK, activación de endopeptidasas tipo cisteína, regulación positiva de NF-kappaB, vía de MAPK3, MAPK1 y regulación positiva de células T (Figura 3C). La sobreexpresión de U62317.2 fue asociada con una mejor sobrevida global de los pacientes (log-rank, p-value=0.04378). Adicionalmente, otros 5 lncRNAs tuvieron sobreexpresión en ambas cohortes: LINC02095, LINC01436, VIM-AS1, LINC01819 y AFAP1-AS1, estos lncRNAs tuvieron coexpresión con 103 mRNAs en ambas cohortes, dichos mRNAs están asociados a la activación de la vía MAPKKK y la regulación positiva de la proliferación y diferenciación (Figure 3D).

### **Perfil de expresión de lncRNAs en subtipo HER2-enriched**

En nuestra cohorte detectamos 8 lncRNAs sobreexpresados que no han sido asociados previamente en la literatura a CM. Dos de ellos (AC116021 y LINC00672) están sobreexpresados en nuestra cohorte y en TCGA, tienen correlación en expresión con 90 mRNAs. El análisis de GO reveló que están asociados a la regulación de la transcripción de Pol II, respuesta a estrés de retículo endoplásmico, regulación positiva de producción de VEGF, monoubiquitinación de histonas y regulación negativa de anoikis (Figure 3E).



**Figure 3. Análisis de gene ontology (GO) de los lncRNAs con mayor tasa de sobreexpresión en nuestra cohorte y en TCGA.** GO fue realizado usando el grupo de mRNAs que tuvieron alta correlación con los lncRNAs sobreexpresados en cada subtipo molecular.

### El lncRNA AC009283.1 en tumores del subtipo HER2-enriched

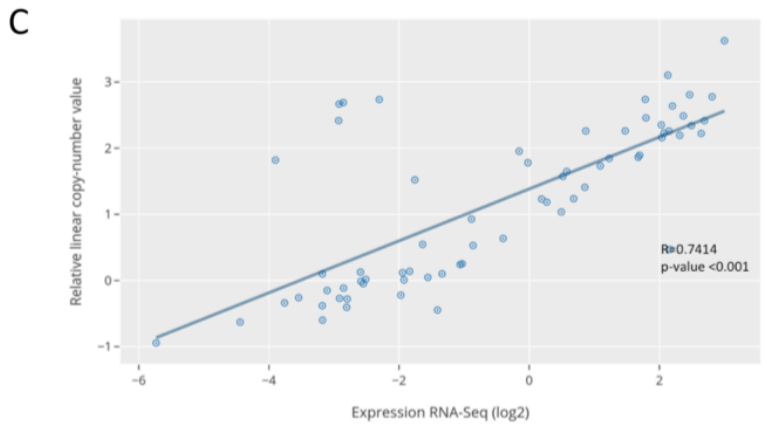
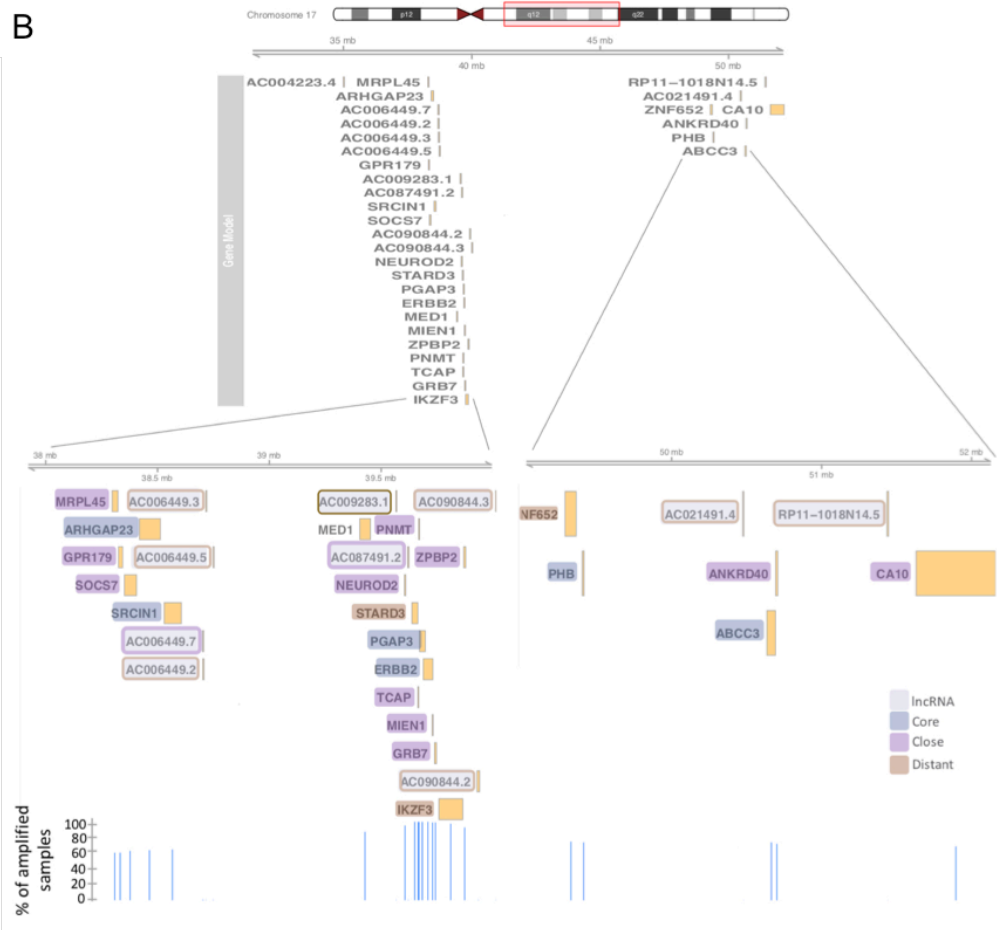
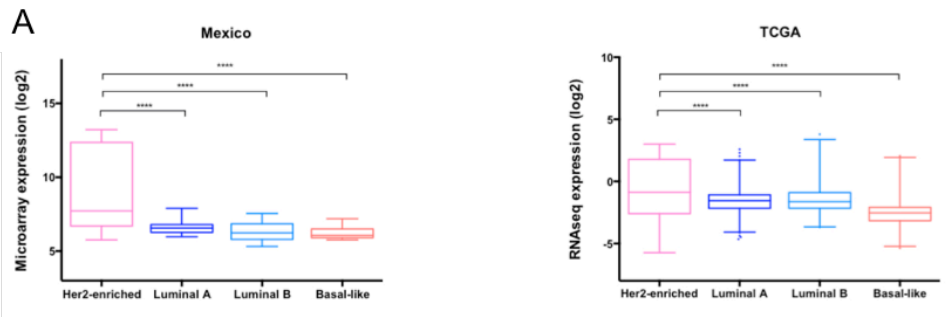
Como lo comentamos anteriormente, los pacientes con CM del subtipo molecular HER2-enriched muestran la menor probabilidad de sobrevivir respecto a los otros subtipos moleculares, en las cohortes de TCGA (n=743) como en METABRIC (n=1,463), por lo que elegimos este subtipo molecular para una caracterización mas amplia de los lncRNAs alterados. Particularmente nos enfocamos en el lncRNA AC009283.1 que presentó el nivel mas alto de expresión tanto en nuestra cohorte (fold change 4.02, p-value <0.0001) como en TCGA (fold change 2.46 p-value <0.0001) (Figure 4A). No hay reportes previos de la función de este lncRNA,

por lo que fue seleccionado para su descripción funcional en el subtipo molecular HER2-enriched.

AC009283.1 es un lncRNA intergénico localizado en chr17:39,566,915-39,567,559 (GRCh38/hg38) anotado en la base ENSEMBL como ENSG00000273576. Algunos nombres alternativos que han sido reportados en otras bases son: HSALNG0116248, NONHSAG021701, lnc-CDK12-1 y RP11-390P24.1. El potencial codificante de AC009283.1 es 0.113, similar a otros lncRNAs bien conocidos como HOTAIR, MALAT-1 y PANDAR (0.09, 0.1295 y 0.058, respectivamente).

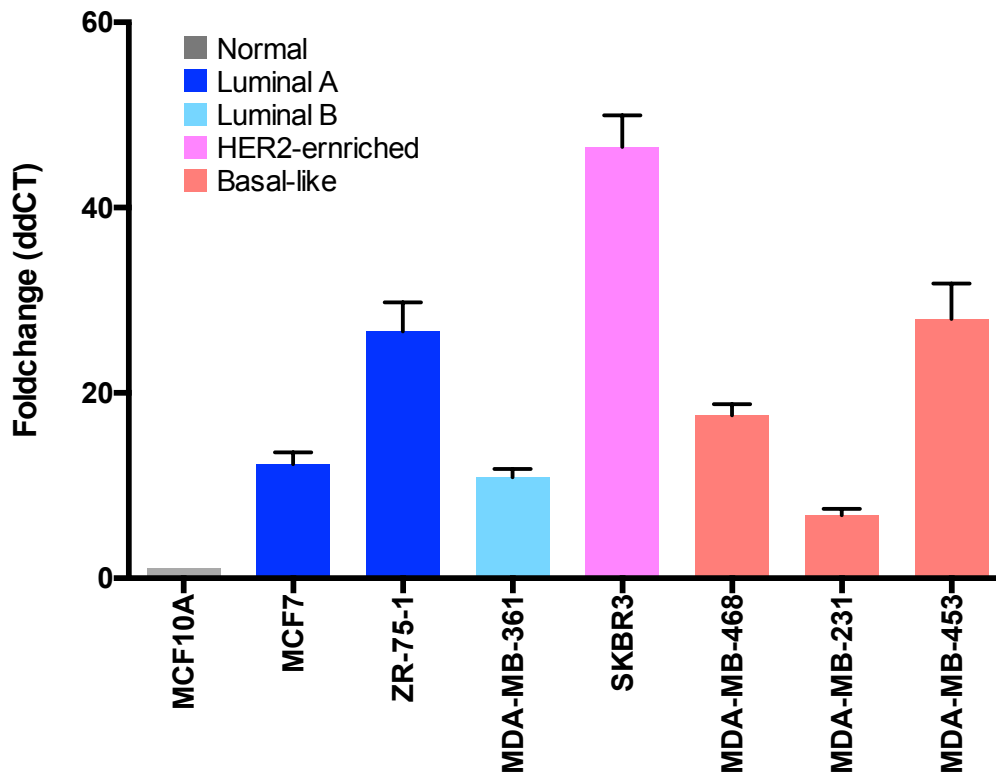
Un análisis genómico mostró que AC009283.1 se encuentra adyacente al gen ERBB2 y que co-amplifica con la región central del amplicón HER2 17q12 (Figura 4B). Un análisis de alteraciones en número de copias en tumores HER2 positivos de TCGA demostró que el amplicón 17q12 contiene genes codificantes y no codificantes amplificados (Figure 4B), dentro de los cuales se encuentra AC009283.1 cuyos niveles de expresión estuvieron significativamente correlacionados con su número de copias (Figure 4C). Estos datos sugieren que AC009283.1 es un importante gen en los tumores HER2-enriched, por lo que se estudió como la alteración de la expresión de AC009283.1 afecta la biología de una línea celular del subtipo HER2-enriched.





**Figure 4. Expresión y contexto del AC009283.1 en el amplicón 17q12. A.** Expresión de AC009283.1 en los subtipos moleculares en la cohorte mexicana (n=75, evaluado por microarreglos) y en la cohorte de TCGA/TANRIC (n=853, evaluado por RNA-Seq **B.** Mapa genómico del amplicón 17q12 denotando los lncRNAs contenidos dentro. **C.** Análisis de correlación de expresión y número de copias AC009283.1 en tumores HER2-enriched de TCGA.

La expresión de AC009283.1 fue evaluada en un panel de siete líneas celulares, la línea celular SK-BR-3 mostró la mayor tasa de expresión de AC009283.1, además posee una amplificación en la zona 17q12 y de acuerdo a PAM50 se clasifica como subtipo molecular HER2-enriched, por lo que elegimos esta línea celular como el modelo para el estudio funcional de AC009283.1 (Figura 5).



**Figure 4.** Tasa de expresión de AC009283.1 en un panel de líneas celulares de CM y una línea celular no transformada (MCF10A).

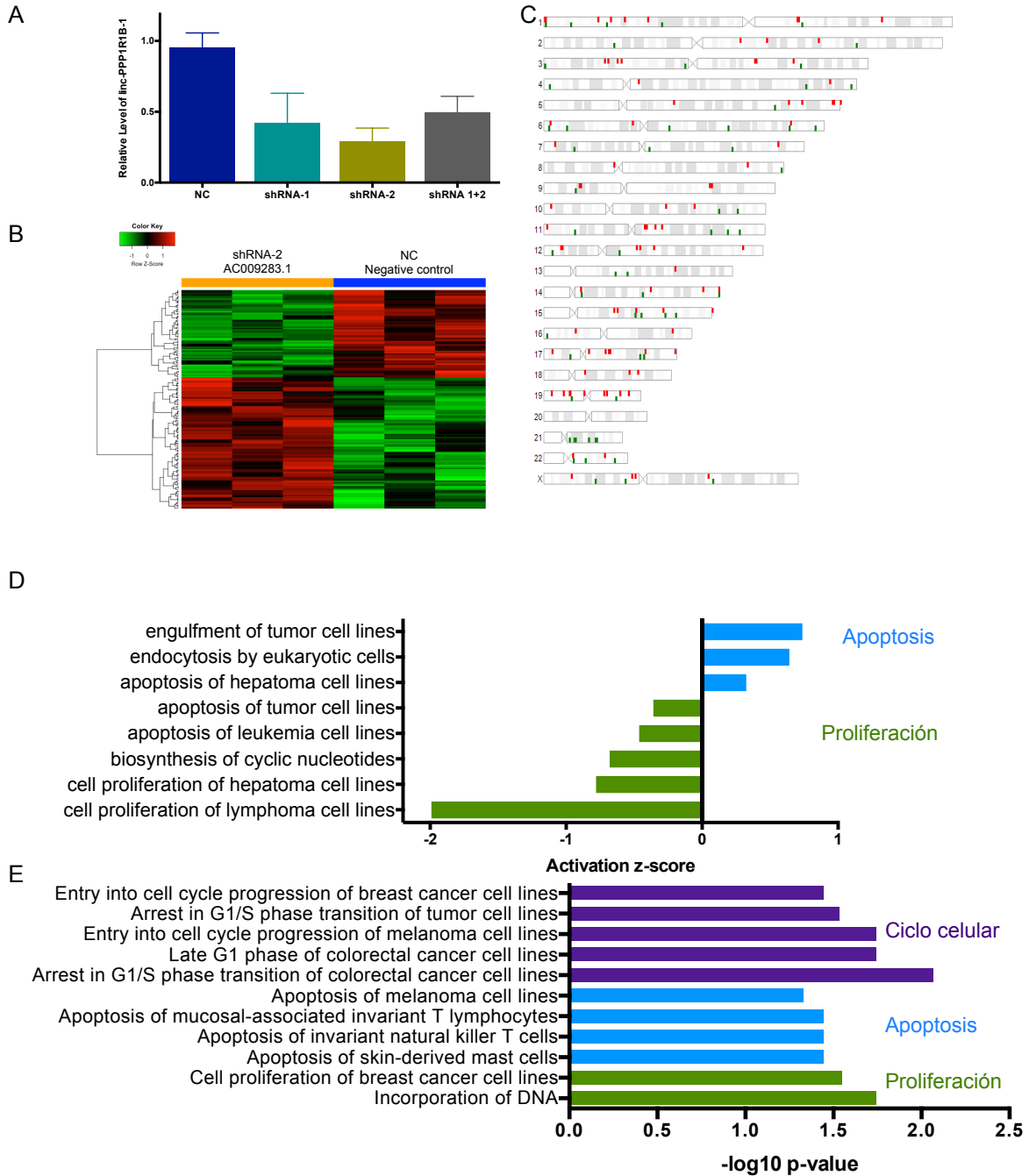
Para inferir el papel regulador de AC009283.1 comenzamos realizando un fraccionamiento celular, de manera preferencial el lncRNA se localiza en el núcleo lo cual sugiere una función predominantemente como regulador transcripcional.

## **El silenciamiento de AC009283.1 modifica el perfil transcripcional de la línea celular SK-BR-3 de fenotipo HER2-enriched**

Para explorar la función de AC009283.1 en el subtipo molecular HER2-enriched, silenciamos su expresión empleando dos diseños de shRNA (shRNA1 y shRNA2). Ambas secuencias se dirigen al exón 10 y 14 de las 14 isoformas reportadas para AC009283.1. Después de 96h de transfección usando el shRNA2, la expresión del lncRNA se redujo al 30% respecto al control, por lo que este diseño fue seleccionado para los experimentos posteriores (Figure 6A).

Debido a que los lncRNAs son capaces de regular la expresión de genes vecinos (*cis*) o distantes (*trans*), usamos un abordaje de microarreglos para evaluar las alteraciones en el transcriptoma después del silenciamiento del AC009283.1. Observamos que la tasa de expresión de 158 genes de la línea celular SK-BR-3 se modificó después del silenciamiento del AC009283.1, estos genes se encuentran distribuidos en todo el genoma (*trans*) con valores de fold change >1.2 y <-1.2 (ANOVA p-value <0.05) (Figuras 6B y 6C).

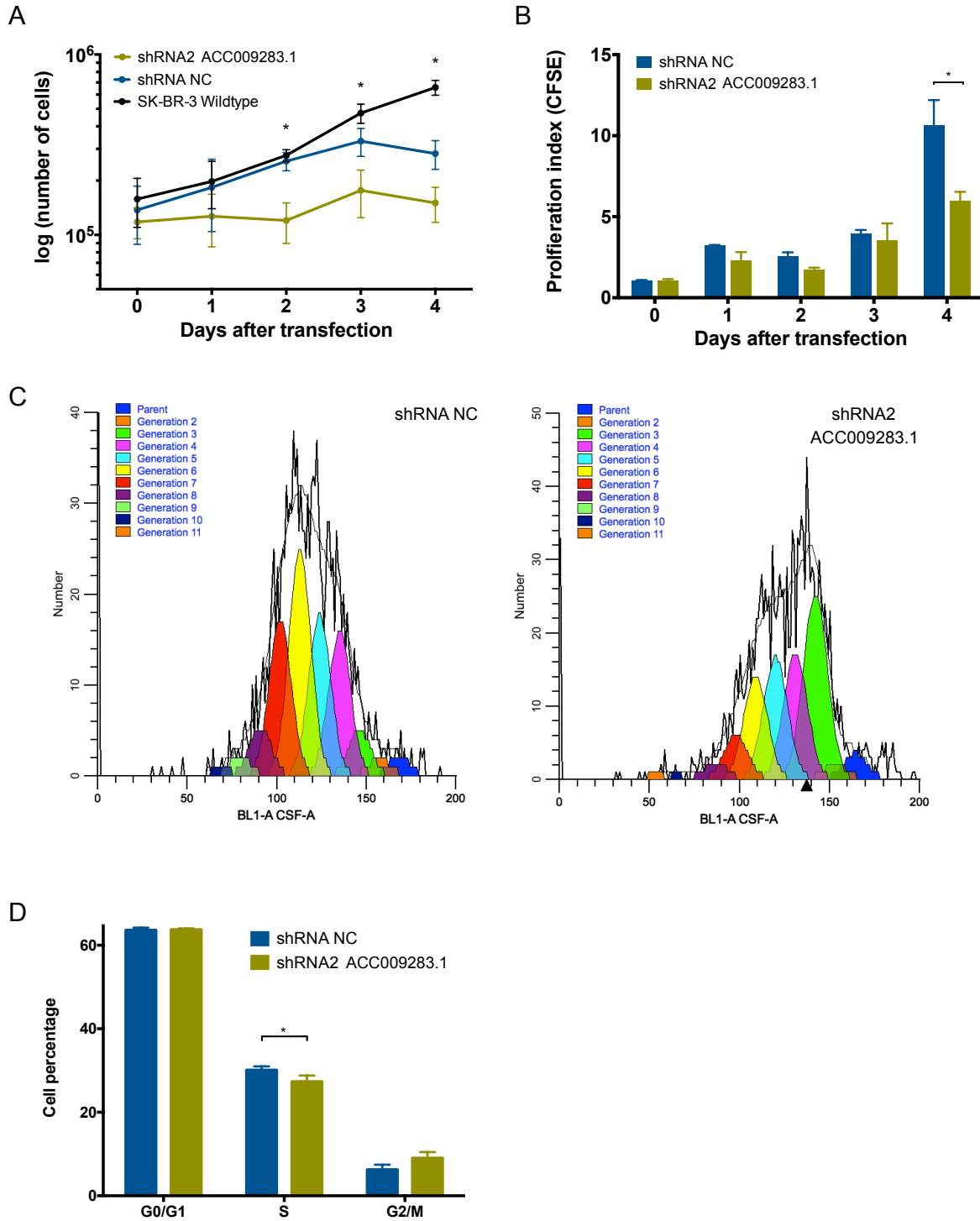
Dentro de estos 158 genes ya hay algunos reportados como asociados a procesos carcinogénicos: *CASCA4*, *NOCTH3*, *TNFa*, *FOSB*, *BLC2A1*, *DNML1*, y *KLF6*. Para explorar el efecto del silenciamiento de AC009283.1 en los procesos biológicos realizamos un enriquecimiento de vías usando los 158 genes diferencialmente expresados y observamos que están involucrados en las vías de proliferación y apoptosis (Figure 6D). Otras vías asociadas a la oncogénesis resultaron alteradas, tales como fosforilación oxidativa, disfunción mitocondrial, señalización e MAPK, entre otras. Interesantemente las vías de proliferación, apoptosis y ciclo celular estuvieron enriquecidas en los tumores de pacientes TCGA con alta expresión de AC009283.1 (Figure 6E). Estos resultados sugieren que AC009283.1 tienen un papel regulando la expresión de genes importantes asociados a vías carcinogénicas.



**Figura 6. Silenciamiento de AC009283.1 en la línea celular SK-BR-3 y su efecto en el transcriptoma. A.** RT-qPCR del silenciamiento del AC009283.1 usando shRNAs. **B.** Mapa de calor supervisado mostrando 158 genes diferencialmente expresados después del silenciamiento del AC009283.1, como control se usó el shRNA NC (fold change <-1.2, >1.2 p-value <0.05). **C.** Distribución de estos genes a lo largo del genoma. **D.** Análisis IPA de los 158 genes diferencialmente expresados mostrando enriquecimiento en los procesos de apoptosis y proliferación celular. **E.** Análisis IPA de los genes diferencialmente expresados entre tumores HER2-enriched con alta y baja tasa de expresión de AC009283.1.

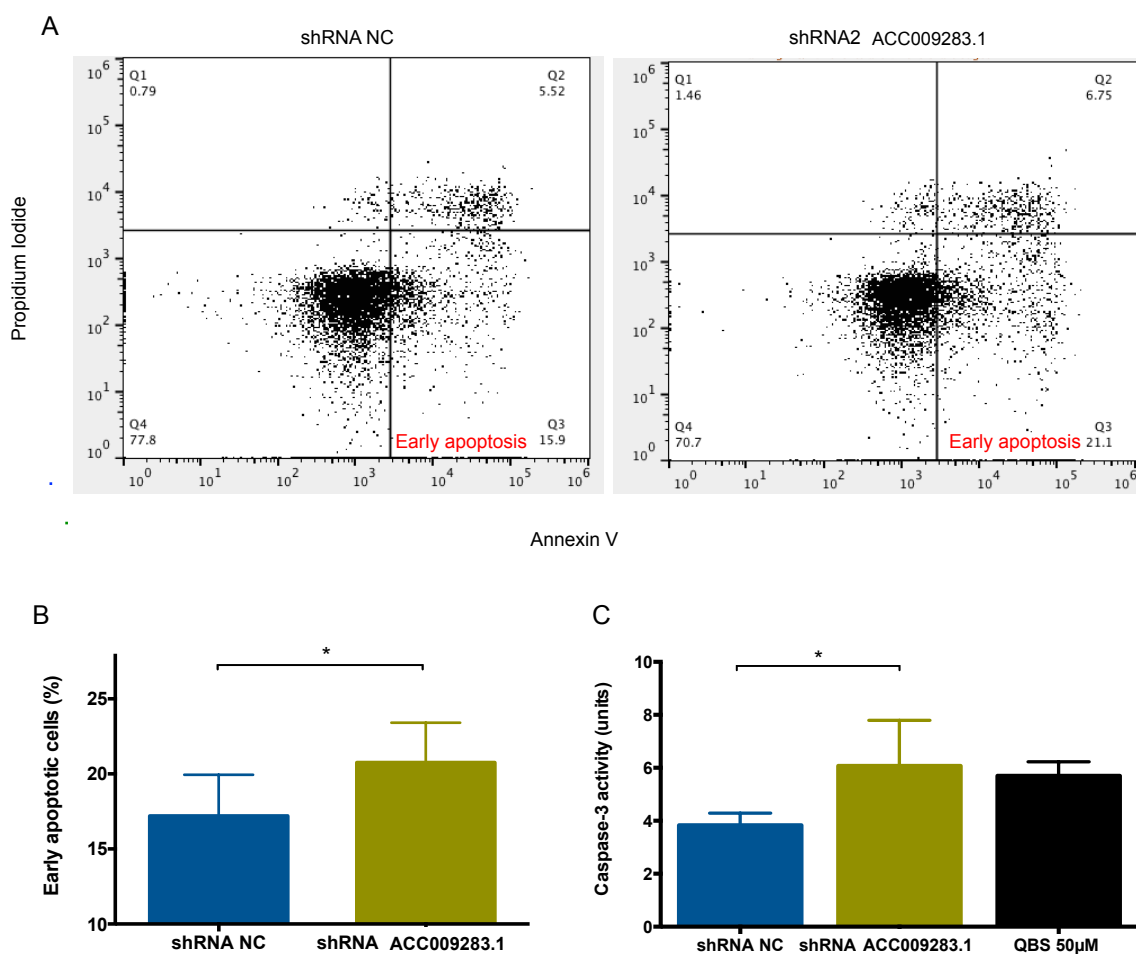
## **El silenciamiento de AC009283.1 inhibe la proliferación e induce apoptosis en la línea celular SK-BR-3 de fenotipo HER2-enriched**

Para validar los hallazgos del enriquecimiento de vías, realizamos ensayos *in vitro* de proliferación y apoptosis en la línea celular SK-BR-3 usando el shRNA2 para inhibir la expresión de AC009283.1 y el shRNA NC como control de expresión basal. El conteo celular automático mostro que el silenciamiento de AC009283.1 disminuye la proliferación comparado con el grupo control, observamos un conteo celular significativamente menor a partir de las 48hrs después de la transfección (Figure 7A), confirmamos este hallazgo con otro ensayo de proliferación basado en citometría de flujo con CFSE, mostrando que el índice de proliferación fue menor en las células con el silenciamiento del AC009283.1 (6.01 vs 10.68 en las células control) (Figure 7B). En la figura 7C se muestra una gráfica representativa del descenso de los índices de proliferación en el ensayo CFSE en células SK-BR-3 con el silenciamiento de AC009283.1. Para comprender mejor a que nivel el cambio en la proliferación estaba siendo regulado, realizamos un ensayo de ciclo celular a las 96hrs después de la transfección, observando un arresto en la fase S del ciclo celular en las células con baja expresión del AC009283.1 (Figure 7D). Estos resultados sugieren que AC009283.1 regula la proliferación celular en la línea SK-BR-3 mediante el arresto en la fase S del ciclo celular.



**Figura 7. Ensayos de proliferación y ciclo celular tras el silenciamiento de AC009283.1 en la línea celular SK-BR-3. A.** Las células fueron contabilizadas cada día por cuatro días, se comparó el número de células entre las condiciones shRNA2 y shRNA NC. **B.** Se midió en índice de proliferación usando el ensayo de CFSE por citometría de flujo. **C.** Imágenes representativas de ensayos de CFSE, se muestran las generaciones celulares en colores diferentes. **D.** Ensayo de ciclo celular usando yoduro de propidio en citometría de flujo con las fases del ciclo celular en el eje de las X.

Adicionalmente realizamos dos ensayos de apoptosis, en el primero usando Anexina y IP, a través de citometría de flujo, se observó que las células SK-BR-3 con silenciamiento de AC009283.1 aumentaron significativamente su porcentaje de apoptosis temprana (21.3% vs 15.9% en el control) (Figuras 8A y 8B). También evaluamos la actividad enzimática de caspasa-3, notando un aumento significativo a las 96hrs en las células con el silenciamiento del AC009283.1 comparado con las células control (Figure 8C). Esto sugiere que AC009283.1 regula la apoptosis en la células de CM HER2-enriched.



**Figura 8. Ensayos de apoptosis tras el silenciamiento de AC009283.1 en la línea celular SK-BR-3.** **A.** Evaluación de apoptosis mediante citometría de flujo con Anexina V y yoduro de propidio. **B.** Representación del porcentaje de células en apoptosis tardía tras el silenciamiento de AC009283.1. **C.** Análisis cuantitativo de actividad de caspasa 3, se uso QBS 50uM como control positivo de apoptosis.

### **AC009283.1 se encuentra sobreexpresado en tumores HER2 positivos y regula procesos oncogénicos**

Con base en la correlación de la expresión y de número de copias de AC009283.1 en la región 17q12 de HER2, creemos que la sobreexpresión de este lncRNA no es única del subtipo molecular HER2-enriched sino también de los tumores HER2 positivos con amplificación en 17q12.

Debido a que PAM50 no clasifica como HER2-enriched a todos los tumores con amplificaciones en HER2, expandimos nuestro análisis a los tumores HER2 positivos con amplificación en HER2 evaluada a través de IHC, usando bases de datos públicamente disponibles de TCGA y Gene Expression Omnibus (GEO).

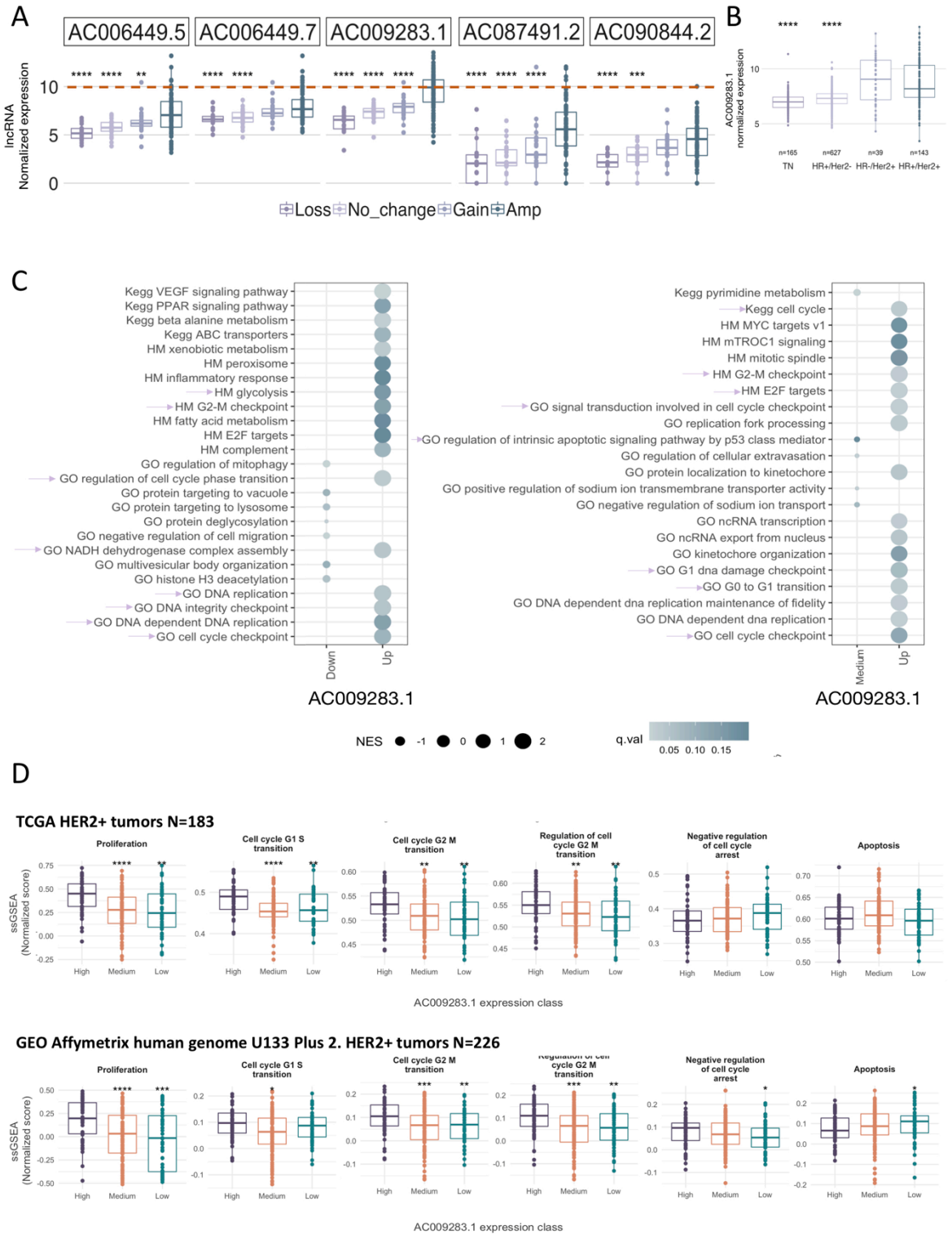
De todos los lncRNAs evaluados dentro del amplicón 17q12, el lncRNA AC009283.1 presentó la mayor tasa de expresión en los tumores HER2 positivos comparado con los tumores HER2 positivos con bajo número de copias en 17q12 o pérdida de esta amplificación (Figure 9A). De acuerdo con estos datos encontramos una sobreexpresión significativa de AC009283.1 en tumores HER2 positivos evaluados por IHC independientemente del estado del receptor hormonal (Figure 9B).

Después de determinar la relevante correlación entre la expresión de AC009283.1 y la amplificación de HER2, realizamos un análisis de enriquecimiento de genes (GSEA) para identificar procesos biológicos relevantes enriquecidos en tumores HER2 positivos con diversos niveles de expresión de AC009283.1 (alto, mediano y bajo). Como se muestra en la figura 9C, el GSEA confirmó los hallazgos previos obtenidos en los ensayos *in vitro*. Encontramos que los tumores HER2 positivos con alta expresión de AC009283.1 tienen un enriquecimiento significativo de los siguientes procesos: check point de la fase G2-M y G0-G1 del ciclo celular,



regulación de la transición del ciclo celular, replicación de DNA, y regulación de la apoptosis intrínseca.

Para validar estas observaciones en una caracterización mas profunda, realizamos single-sample GSEA (ssGSEA) en 409 tumores de mama HER2 positivos obtenidos de TCGA (n=183) y de GEO (n=226). Observamos que los tumores con alta expresión de AC009283.1 respecto a los de baja expresión de AC009283.1 tuvieron predominantemente enriquecimiento en las vías de proliferación, ciclo celular, transición G1-S y G2-M, y regulación negativa del arresto en ciclo celular (Figure 9D).



**Figura 9. Expresión de AC009283.1 en tumores HER2 positivos y su efecto en procesos biológicos. A.** Top de los 5 lncRNAs más expresados en 17q12-21.3 respecto a su estado de número de copias. **B.** Expresión de AC009283.1 en subtipos por IHQ de CM. **C.** Gráfico de burbujas representando el GSEA de tumores con alta expresión del AC009283.1 respecto a los de mediana y baja expresión, las flechas rosas

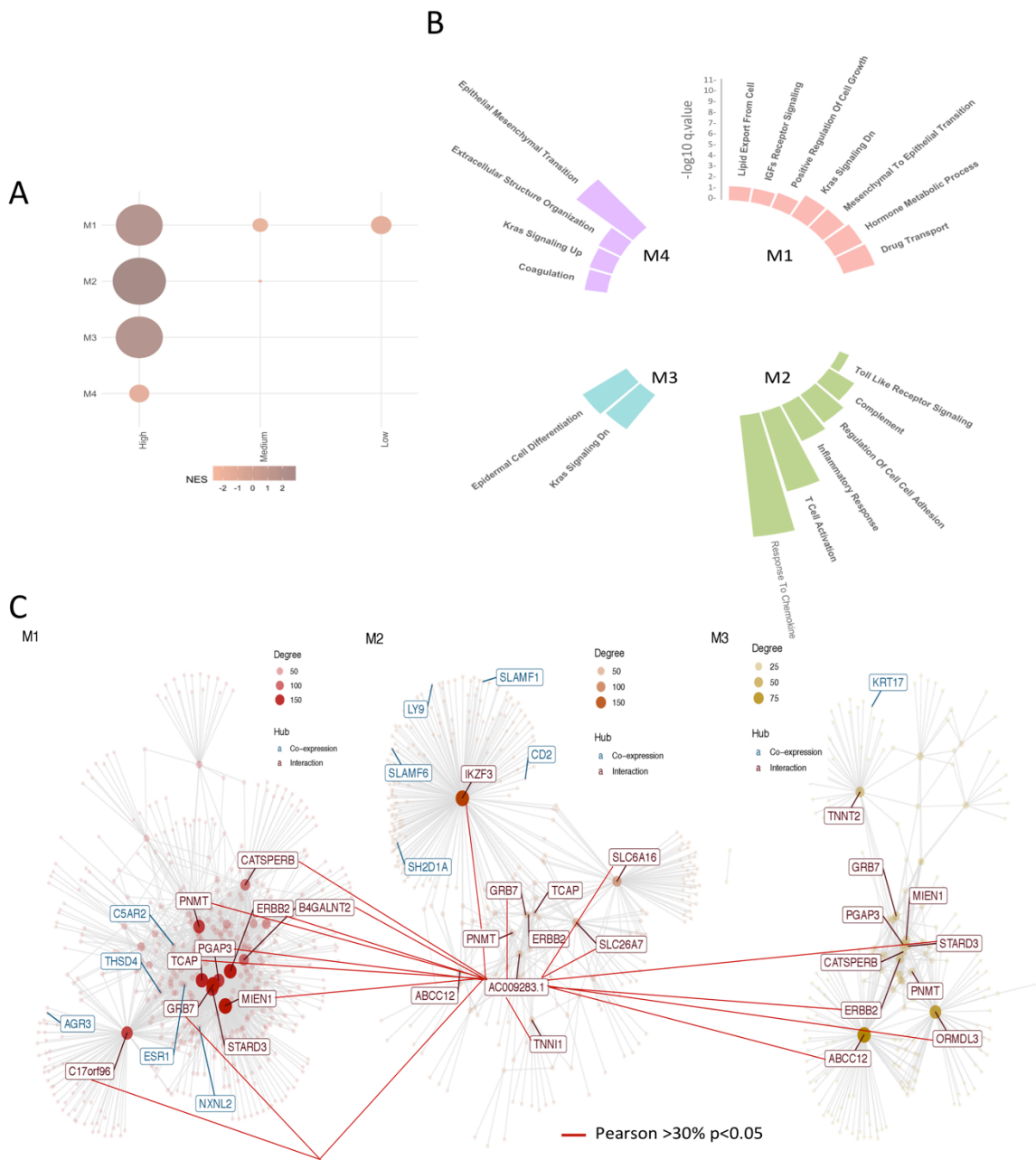
muestran procesos biológicos previamente reportados en los ensayos de silenciamiento *in vitro*. **D.** Análisis de ssGSEA para la expresión alta, mediana y baja de AC009283.1 mostrando enriquecimiento en proliferación celular, ciclo celular y apoptosis.

### **Análisis modular de expresión de AC009283.1 en tumores HER2 positivos**

Evaluar la coexpresión puede ayudarnos a identificar conexiones funcionales entre AC009283.1 y los mRNAs en los tumores HER2 positivos. Por lo que realizamos un análisis de redes de coexpresión con la herramienta CEMiTool usando perfiles de expresión de tumores HER2 positivos divididos de acuerdo al nivel de expresión del AC009283.1 (alto, mediano y bajo). Encontramos cuatro módulos, de los cuales M1, M2 y M3 fueron altamente enriquecidos con la alta expresión de AC009283.1 (Figure 10A), el análisis de GSEA mostró una representación significativa de metabolismo del crecimiento en M1, señalización inmune en M2, y vías oncogénicas como KRAS en M3, todos los módulos tienen procesos asociados con la proliferación celular, progresión en el ciclo celular y apoptosis (Figure 10B), lo que corrobora nuestros resultados *in vitro*.

Cuando integramos la información de coexpresión para identificar reguladores maestros y conexiones entre los módulos, interesantemente se observó la interacción de AC009283.1 en la red que contiene los módulos M1, M2 y M3 (Figure 10C). Este análisis muestra que el lncRNA AC009283.1 puede revelar las posibles relaciones transregulatorias dentro de la red de respuesta de tumores HER2 positivos.

En general, estos datos sugirieron que la sobreexpresión de AC009283.1 está asociada con la proliferación celular, la progresión del ciclo celular y la apoptosis, y además muestran que el papel biológico de AC009283.1 en el CM enriquecido con HER2 probablemente esté modulado y asociado con la amplificación en 17q12.



**Figure 8. Análisis de expresión modular en tumores HER2. A.** Módulos de coexpresión detectados por CEMiTool y su enriquecimiento en tumores con alta, mediana y baja expresión de AC009283.1. **B.** Representación de los procesos biológicos para cada módulo de coexpresión. **C.** Red con el lncRNA AC009283.1 en el centro y su relación de coexpresión con los genes de los módulos 1, 2 y 3.

## DISCUSIÓN

Hace algunas décadas, las secuencias no codificantes del genoma (98%) se consideraban no funcionales y la búsqueda de oncogenes y supresores tumorales se centraba solo en secuencias codificantes de proteínas (2% del genoma). Ahora, con el uso de tecnologías de alto rendimiento como la secuenciación de nueva generación (NGS), sabemos que alrededor del 70% del genoma humano se transcribe, incluidas las regiones no codificantes, las cuales desempeñan un papel importante en la regulación de la expresión génica [25]. Un 33% de estas regiones no codificantes comprenden a los lncRNAs, y su desregulación se asocia a diferentes tipos de cáncer [10]. Este estudio muestra una caracterización del perfil transcripcional de tumores de mama, e identifica la expresión desregulada de AC009283.1 en el subtipo HER2-enriched.

Encontramos 307 lncRNA expresados diferencialmente en los tumores de mama en comparación con el tejido adyacente, este patrón de expresión fue capaz de distinguir el tejido tumoral del adyacente. Varios estudios han informado de un pequeño conjunto de lncRNAs desregulados en CM, sin embargo, se desconoce la función biológica de la mayoría.

Cuando validamos nuestros perfiles de expresión génica utilizando datos de TCGA, encontramos una concordancia del 33.8% en los lncRNAs desregulados en ambas cohortes. Los niveles de concordancia entre los pacientes de esta cohorte y la cohorte TCGA podrían estar influenciados por tres variables: 1) Disparidades en el número de muestras (75 para la cohorte mexicana y 990 para TCGA), 2) Las tecnologías de detección utilizadas para obtener datos de expresión génica: microarreglos para nuestra cohorte y RNA-seq para la cohorte TCGA), y 3) Los factores biológicos intrínsecos de las muestras, como el porcentaje de pureza celular tumoral. A pesar de esto los lncRNA concordantes entre ambas cohortes pueden tener un papel biológico relevante en el CM, ya que están presentes en dos cohortes independientes. Además, estos datos sugieren que los microarreglos

y el RNA-seq pueden ser métodos complementarios para identificar lncRNAs desregulados en CM [26, 27].

Para adquirir información sobre el papel biológico potencial de los lncRNAs alterados en CM se realizó un análisis de coexpresión de lncRNA-mRNA, en nuestra cohorte y en la de TCGA. Es de destacar que el análisis de coexpresión se ha descrito como un enfoque sólido para inferir la función biológica de los lncRNAs [28].

En el subtipo Luminal A, los lncRNAs: LINC00504, LINC00993, MIR205HG y PCAT18 se encontraron sobre expresados en ambas cohortes (México y TCGA). No se ha reportado su función biológica, pero de acuerdo al análisis de coexpresión (lncRNA-mRNA) se asocian con la regulación positiva de la transcripción, un proceso que regula la frecuencia, velocidad y extensión de la transcripción de la ARN polimerasa II, y que se ha convertido en un blanco para la terapia del cáncer [29]. Esta observación también podría estar relacionada con el hecho de que se ha observado que en la línea celular MCF7 de subtipo Luminal A el estrógeno es un potente estimulador de la transcripción de la ARN polimerasa II [30]. Estos cuatro lncRNAs (LINC00504, LINC00993, MIR205HG y PCAT18) sobreexpresados en los tumores Luminal A en conjunto también regulan el movimiento basado en los microtúbulos, un proceso que resulta en el movimiento de los orgánulos a través de la polimerización o despolimerización de los microtúbulos, este proceso también es un blanco prometedor para los medicamentos contra el CM [31]. Otro proceso biológico enriquecido fue el acoplamiento y la organización ciliar. Se ha informado que los cilios en los tumores de mama son más frecuentes en tumores más diferenciados como los del subtipo luminal A [32]. Además, se observó que el promotor de MIR205HG muestra un sitio de unión para el receptor de estrógeno en la línea celular Luminal A MCF7, según los ensayos de secuenciación de inmunoprecipitación de cromatina (ChIP-Seq) [33], y el lncRNA de PCAT18 se asocia significativamente con la señalización del receptor de andrógenos (AR) en el cáncer de próstata [34]. Estos datos

sugieren una posible asociación de estos cuatro lncRNAs en la biología tumoral y la señalización hormonal en el subtipo molecular Luminal A.

En el subtipo Luminal B, observamos sobreexpresión de DSCAM-AS1 del cual se ha informado previamente que regula la progresión del tumor y la resistencia al tamoxifeno. Luminal B es el tumor positivo para receptor de estrógeno y progesterona más agresivo clínicamente, y DSCAM-AS1 es un discriminador importante de los subtipos luminales en CM [35]. Otros tres lncRNA sobreexpresados en nuestra cohorte y en TCGA fueron LINC0187, RP11-379F12.4 y GATA3-AS1. Hasta ahora, no hay informes de su función biológica en el cáncer. El análisis de coexpresión (lncRNA-mRNA) los asoció con la acetilación de histona H3K14 (H3K14ac). La acetilación de residuos de lisina específicos de histonas juega un papel clave en la regulación de la expresión génica, de hecho, H3K14ac es una marca de histona activa y se correlaciona con la magnitud de la expresión génica, probablemente promoviendo la expresión de oncogenes [36].

En los tumores basal-like, un subtipo con mayor alteraciones no solo transcripciones sino estructurales y mutacionales, observamos 280 lncRNAs desregulados, lo que es más importante, la función de la mayoría de ellos aún se desconoce. Esto pone de manifiesto la falta de conocimiento sobre la función de los lncRNA desregulados en el fenotipo tumoral basal-like, uno de los subtipos más agresivos que carecen de terapia dirigida. La regulación positiva de los lncRNAs U62317.1 y U62317.2 (en nuestra cohorte y TCGA) se correlaciona con genes implicados en la regulación positiva de la cascada JNK, una cascada de proteína quinasa intracelular. Se sabe que la activación persistente de las JNK está involucrada en el desarrollo y la progresión del cáncer [37]. Otra vía enriquecida con estos dos lncRNAs es la activación de endopeptidasas de tipo cisteína implicadas en el proceso apoptótico, como las caspasas, enzimas conocidas por su papel en el inicio y ejecución de la apoptosis [38]. Además, estos dos lncRNA están asociados con la regulación positiva de NF- $\kappa$ B; La vía desregulada de NF- $\kappa$ B conduce a la alteración del equilibrio entre la proliferación

celular y la muerte a través de la regulación positiva de proteínas antiapoptóticas [39].

Otros cuatro lncRNA (LINC02095, LINC01436, VIM-AS1, LINC01819 y AFAP1-AS1) también se regularon a la alza en nuestra cohorte y la cohorte TCGA. Tienen una correlación de expresión con 103 mRNAs en ambas cohortes, asociado con la activación de MAPKKK (MAP3K1), que regula la supervivencia celular, apoptosis y migración celular en múltiples tipos de células tumorales [40].

En el subtipo HER2-enriched, 8 lncRNA están sobreexpresados, de los cuales, solo se conoce la función de LINC00152, ya que regula la proliferación celular, promueve la detención del ciclo celular en la fase G1, desencadena la apoptosis tardía, reduce el programa de transición epitelial mesenquimal, y suprime la migración e invasión celular en el cáncer gástrico [41]. MIR4435-1HG está regulado al alza en tumores HER2-enriched y se ha asociado con una supervivencia global reducida y supervivencia libre de enfermedad en el cáncer de cabeza y cuello [42]. Es de destacar que las funciones de LINC00152 y MIR4435-1HG, mencionadas anteriormente, se han descrito en modelos de cáncer gástrico y de cabeza y cuello, pero no en CM.

No hay informes de la función biológica en el cáncer de AC116021 y LINC00672, ambos se encontraron co-expresados con 90 mRNAs en nuestra cohorte y en TCGA, y según GO están asociados con la regulación positiva de la producción de VEGF y regulación negativa de anoikis. Se ha documentado que en tumores con HER2 sobreexpresado se induce la expresión y secreción de VEGF, y la expresión aumenta cuando los tumores se vuelven resistentes al tratamiento [43]. Además, la regulación negativa de anoikis también es una vía enriquecida. Anoikis es un sello distintivo de las células metastásicas y se ha informado que las células HER2 son resistentes a este proceso [44]. Ninguno de estos lncRNAs sobreexpresados han sido descritos previamente en la literatura como alterados en el subtipo HER2-enriched.



El análisis de coexpresión de lncRNA-mRNA destaca funciones importantes de lncRNAs específicos para cada subtipo molecular. Algunos grupos de lncRNAs alterados en los subtipos están involucrados en vías comunes (transcripción de la RNA polimerasa, acetilación de histonas, etc.), mientras que otros son más especializados (en el subtipo HER2, los relacionados con la señalización de VEGF), lo que explica la alta heterogeneidad de la biología molecular del CM. Esto sugiere que la sobreexpresión de algunos lncRNAs puede regular los procesos asociados con la carcinogénesis, lo que los convierte en objetivos interesantes para futuros estudios para comprender la biología de los subtipos moleculares de CM.

En este trabajo nos enfocamos al estudio del subtipo molecular HER2-enriched, caracterizado por la amplificación de genes contenidos en la región 17q12, así como por la alta tasa de expresión del gen ERBB2 y otros cinco genes promotores del cáncer (GRB7, PNMT, STARD3, PGAP3 y MGC1483) [45]. Los tumores HER2-enriched también son uno de los subtipos que exhiben la supervivencia libre de progresión más baja [46] y la tasa más alta de metástasis en la cohorte mexicana y menor sobrevida global en TGCA y METABRIC. Por tanto, la caracterización biológica de los tumores HER2-enriched es una tarea relevante, especialmente para descifrar el papel de los genes ubicados dentro de la zona de amplificación de HER2.

Pocos informes se han centrado en el estudio de lncRNAs en el subtipo molecular HER2-enriched [47]. Hasta la fecha, no se ha investigado el papel funcional del AC009283.1, un lncRNA sobreexpresado en el subtipo HER2-enriched y localizado en una región amplificada que se ha documentado como necesaria para el proceso carcinogénico en este subtipo. Para comprender mejor la función del AC009283.1, evaluamos el efecto de su silenciamiento en el transcriptoma de las células SK-BR-3 del subtipo HER2-enriched. Hasta donde sabemos el papel

biológico de AC009283.1 no se ha informado aún en el CM, por lo que este es el primer estudio en describirlo.

El análisis de microarreglos reveló que el silenciamiento de AC009283.1 provoca la desregulación de 158 genes, algunos asociados con la proliferación, el ciclo celular y la apoptosis, como se mostró en los análisis de enriquecimiento. Este análisis encontró que el silenciamiento de AC009283.1 alteró la expresión de NOTCH3, TNFa y FOSB, genes que se habían sugerido previamente como impulsores de la proliferación y el ciclo celular en el cáncer.

La regulación positiva de NOTCH3 induce la detención del ciclo celular en la fase G0 / G1-S e inhibe la proliferación y formación de colonias en las líneas celulares de CM [48]. NOTCH3 se sobrepresó por el silenciamiento de AC009283.1 y observamos la detención de células en la fase S del ciclo celular. La expresión del gen de TNFa fue regulada negativamente por la caída de AC009283.1 y también se ha demostrado que la regulación positiva de TNFa aumenta la proliferación celular [49]. El oncógeno FOSB se incrementó después del silenciamiento de AC009283.1, y FOSB se ha asociado con la proliferación celular reducida en líneas celulares de cáncer gástrico [50].

En conclusión, los ensayos *in vitro* (recuento celular automatizado, índice de proliferación y ensayo de ciclo celular por citometría de flujo) mostraron que el silenciamiento de AC009283.1 redujo la proliferación y disminuyó la fase S en las células SK-BR-3.

Este análisis de transcriptoma después del silenciamiento de AC009283.1 mostró que la expresión de BCL2A1, DNML1 y KLF6 se modificó, genes previamente reportados como asociados con la apoptosis.

BCL2A1, un gen antiapoptótico que se sobreexpresa en una variedad de células cancerosas, incluidas neoplasias hematológicas y tumores sólidos [51]

y DNML1 participa en la división peroxisomal, mitocondrial y en la regulación de la apoptosis. En un trabajo previo, la caída de la expresión de DNML1 provocó un aumento significativo de la apoptosis [52]. Ambos genes (BCL2A1 y DNML1) redujeron su expresión después del silenciamiento AC009283.1 en la línea celular SK-BR-3.

Finalmente, KLF6 es un supresor de tumores que está regulado negativamente en varios tipos de cánceres, suprime el crecimiento tumoral a través de la activación de p21 e induce apoptosis en el cáncer de próstata [53]. El silenciamiento de AC009283.1 provocó una sobreexpresión de KLF6.

En conclusión, el aumento de la apoptosis en las células SK-BR-3 con silenciamiento de AC009283.1 puede deberse a la modulación de los genes BCL2A1, DNML1 y KLF6.

Todo esto indica que el análisis de enriquecimiento de vías a partir de los microarreglos y los ensayos *in vitro* demuestran que el silenciamiento de AC009283.1 tiene un efecto antiproliferativo y aumenta la apoptosis en la línea celular SK-BR-3. Por tanto, la sobreexpresión de AC009283.1 en los tumores de mama induce la evasión de la apoptosis y la proliferación sostenida, siendo ambos procesos distintivos del cáncer [3].

La doble regulación de la proliferación y la apoptosis que observamos en este trabajo parece ser común para los lncRNAs alterados en cáncer. Un ejemplo es LSINCT5 que regula 36 genes que codifican proteínas en el CM y cáncer de ovario 45, GAS5 en el cáncer de pulmón [54], y HOXA11-AS en cáncer gástrico a través de la regulación de miR-1297 [55] y UCA1 en cáncer colorrectal [56]. En este contexto, nuestros resultados sugieren que AC009283.1 contribuye al fenotipo maligno en el subtipo HER2-enriched a través de la interacción con genes clave, conduciendo a una mayor proliferación y resistencia a la muerte por apoptosis.

Para validar aún más el papel biológico funcional de AC009283.1 en tumores HER2-enriched impulsados por la amplificación de HER2, realizamos un GSEA y ssGSEA en dos cohortes independientes de CM. El análisis de GSEA muestra que la sobreexpresión de AC009283.1 está asociada con el enriquecimiento de la proliferación celular de progresión del ciclo celular. Estos datos fueron confirmados por ssGSEA. Adicionalmente en el análisis de coexpresión de Cemitool encontramos módulos significativamente enriquecidos en tumores HER2 positivos que sobreexpresan AC009283.1, dichos módulos estuvieron relacionados con la proliferación y el metabolismo celular. Algunos procesos inmunológicos también estuvieron sobrerrepresentados. Dado que el subtipo HER2 ha sido descrito como un tumor inmunogénico, en futuros estudios debería incluirse un estudio detallado de este fenotipo y su relación con AC009283.1 [57]. Esto nos indica que los procesos celulares son impulsados por múltiples moléculas que interactúan entre sí (lncRNAs y mRNAs) ya que los genes que participan en la misma señalización y que comparten funciones similares tenderán a coexpresarse [58].

Los datos resultantes de diversas estrategias computacionales sugieren que AC009283.1 es un potencial impulsor del fenotipo maligno en tumores enriquecidos y amplificados con HER2, lo que confirma nuestros resultados de los ensayos de silenciamiento *in vitro*. Falta por conocer el mecanismo molecular por el cual AC009283.1 regula la expresión de genes blanco implicados en la proliferación, el ciclo celular y la apoptosis. Estudios anteriores han informado que una gran parte de los lncRNA regulan la transcripción a través de la interacción y modulación de la cromatina [11], encontramos que el AC009283.1 está enriquecido en el núcleo celular, por lo que es probable que sus funciones tengan lugar en la cromatina. Por lo tanto, se necesitan más estudios para comprender el mecanismo molecular de AC009283.1 en CM subtipo HER2.

## CONCLUSIONES

Observamos que la expresión de 307 lncRNAs está desregulada en tumores de mama respecto a su tejido normal y que entre los subtipos moleculares hay 355 lncRNAs con expresión desregulada. El análisis de co-expresión lncRNA-mRNA entre subtipos moleculares demostró que los lncRNAs están involucrados en procesos biológicos comunes (transcripción, acetilación, etc.), pero también en otros más especializados (señalización de VEGF, regulación positiva de NF-kappaB, activación de la vía MAPKKK, etc.). El subtipo molecular HER2-enriched, uno de los más agresivos clínicamente, tuvo sobreexpresión del lncRNA AC009283.1. El silenciamiento de este lncRNA *in vitro* mostró su papel como regulador de 158 genes involucrados en proliferación, ciclo celular y apoptosis. Estos procesos biológicos fueron validados a través de ensayos *in vitro* que evalúan dichos procesos, confirmando el papel del lncRNA AC009283.1 como su potencial regulador. Validamos aún más el papel biológico del lncRNA a través de GSEA y ssGSEA en dos cohortes independientes de tumores HER2+, lo que demostró que el lncRNA AC009283.1 es un impulsor del fenotipo maligno en tumores HER2. Por lo tanto, la alta expresión del lncRNA AC009283.1 puede estar asociada con la carcinogénesis y sugerimos que juega un papel importante en el CM del subtipo HER2-enriched.

## PERSPECTIVAS

Este trabajo sugiere que la sobreexpresión y subexpresión de los lncRNAs aquí reportados puede estar regulando los procesos asociados con la carcinogénesis, lo que los convierte en objetivos interesantes para futuros estudios para comprender la biología de los subtipos moleculares de CM.

Falta por conocer el mecanismo molecular por el cual el lncRNA AC009283.1 regula la expresión de genes blanco implicados en la proliferación, el ciclo celular y la apoptosis.

## REFERENCIAS

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018.
2. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100**(1):57-70.
3. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**(5):646-674.
4. Allred DC: **Issues and updates: evaluating estrogen receptor-alpha, progesterone receptor, and HER2 in breast cancer.** *Mod Pathol* 2010, **23 Suppl 2**:S52-59.
5. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al*: **Supervised risk predictor of breast cancer based on intrinsic subtypes.** *J Clin Oncol* 2009, **27**(8):1160-1167.
6. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**(6797):747-752.
7. Haque R, Ahmed SA, Inzhakova G, Shi J, Avila C, Polikoff J, Bernstein L, Enger SM, Press MF: **Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades.** *Cancer Epidemiol Biomarkers Prev* 2012, **21**(10):1848-1855.
8. Cancer Genome Atlas N: **Comprehensive molecular portraits of human breast tumours.** *Nature* 2012, **490**(7418):61-70.
9. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J *et al*: **GENCODE reference annotation for the human and mouse genomes.** *Nucleic Acids Res* 2019, **47**(D1):D766-D773.
10. Gibb EA, Brown CJ, Lam WL: **The functional role of long non-coding RNA in human carcinomas.** *Mol Cancer* 2011, **10**:38.
11. Liu Y, Sharma S, Watabe K: **Roles of lncRNA in breast cancer.** *Front Biosci (Schol Ed)* 2015, **7**:94-108.
12. Ye N, Wang B, Quan ZF, Cao SJ, Wen XT, Huang Y, Huang XB, Wu R, Ma XP, Yan QG: **Functional roles of long non-coding RNA in human breast cancer.** *Asian Pac J Cancer Prev* 2014, **15**(15):5993-5997.
13. Liu YR, Jiang YZ, Xu XE, Yu KD, Jin X, Hu X, Zuo WJ, Hao S, Wu J, Liu GY *et al*: **Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer.** *Breast Cancer Res* 2016, **18**(1):33.
14. Zhao W, Luo J, Jiao S: **Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications.** *Sci Rep* 2014, **4**:6591.
15. Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, Weinstein JN, Spano JP, Meric-Bernstam F, Khayat D *et al*: **Comprehensive analysis of long**

- non-coding RNAs in human breast cancer clinical subtypes.** *Oncotarget* 2014, **5**(20):9864-9876.
16. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, Zhang Y, Yang L, Shan W, He Q *et al*: **Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers.** *Cancer Cell* 2015, **28**(4):529-540.
  17. Durinck S, Spellman PT, Birney E, Huber W: **Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.** *Nat Protoc* 2009, **4**(8):1184-1191.
  18. Volders PJ, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, Vandesompele J: **LNCipedia 5: towards a reference set of human long non-coding RNAs.** *Nucleic Acids Res* 2019, **47**(D1):D135-D139.
  19. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S *et al*: **Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.** *Proc Natl Acad Sci U S A* 2007, **104**(50):20007-20012.
  20. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M *et al*: **The landscape of somatic copy-number alteration across human cancers.** *Nature* 2010, **463**(7283):899-905.
  21. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ: **Software for computing and annotating genomic ranges.** *PLoS Comput Biol* 2013, **9**(8):e1003118.
  22. Russo PST, Ferreira GR, Cardozo LE, Burger MC, Arias-Carrasco R, Maruyama SR, Hirata TDC, Lima DS, Passos FM, Fukutani KF *et al*: **CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses.** *BMC Bioinformatics* 2018, **19**(1):56.
  23. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P: **The Molecular Signatures Database (MSigDB) hallmark gene set collection.** *Cell Syst* 2015, **1**(6):417-425.
  24. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C *et al*: **Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1.** *Nature* 2009, **462**(7269):108-112.
  25. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG *et al*: **The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.** *Genome Res* 2012, **22**(9):1775-1789.
  26. Guo Y, Sheng Q, Li J, Ye F, Samuels DC, Shyr Y: **Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data.** *PLoS One* 2013, **8**(8):e71462.
  27. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res* 2008, **18**(9):1509-1517.
  28. Guo Q, Cheng Y, Liang T, He Y, Ren C, Sun L, Zhang G: **Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-**

- associated lncRNA biomarkers in ovarian cancer malignant progression. *Sci Rep* 2015, **5**:17683.**
29. Villicana C, Cruz G, Zurita M: **The basal transcription machinery as a target for cancer therapy.** *Cancer Cell Int* 2014, **14**(1):18.
  30. Wu J, Cai Y, Zhao G: **Bioinformatic analysis of changes in RNA polymerase II transcription stimulated by estradiol in MCF7 cells.** *Neoplasma* 2018, **65**(1):14-20.
  31. Higa GM: **The microtubule as a breast cancer target.** *Breast Cancer* 2011, **18**(2):103-119.
  32. Menzl I, Lebeau L, Pandey R, Hassounah NB, Li FW, Nagle R, Weihs K, McDermott KM: **Loss of primary cilia occurs early in breast cancer development.** *Cilia* 2014, **3**:7.
  33. Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, Iyer MK, Pitchiaya S, Malik R, Hosono Y *et al*: **The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression.** *Nat Commun* 2016, **7**:12791.
  34. Crea F, Watahiki A, Quagliata L, Xue H, Pikor L, Parolia A, Wang Y, Lin D, Lam WL, Farrar WL *et al*: **Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer.** *Oncotarget* 2014, **5**(3):764-774.
  35. Ma Y, Bu D, Long J, Chai W, Dong J: **LncRNA DSCAM-AS1 acts as a sponge of miR-137 to enhance Tamoxifen resistance in breast cancer.** *J Cell Physiol* 2018.
  36. Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L: **H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells.** *BMC Genomics* 2012, **13**:424.
  37. Bubici C, Papa S: **JNK signalling in cancer: in need of new, smarter therapeutic targets.** *Br J Pharmacol* 2014, **171**(1):24-37.
  38. Krepela E: **Cysteine proteinases in tumor cell growth and apoptosis.** *Neoplasma* 2001, **48**(5):332-349.
  39. Biswas DK, Shi Q, Baily S, Strickland I, Ghosh S, Pardee AB, Iglehart JD: **NF-kappa B activation in human breast cancer specimens and its role in cell proliferation and apoptosis.** *Proc Natl Acad Sci U S A* 2004, **101**(27):10137-10142.
  40. Pham TT, Angus SP, Johnson GL: **MAP3K1: Genomic Alterations in Cancer and Function in Promoting Cell Survival or Apoptosis.** *Genes Cancer* 2013, **4**(11-12):419-426.
  41. Zhao J, Liu Y, Zhang W, Zhou Z, Wu J, Cui P, Zhang Y, Huang G: **Long non-coding RNA Linc00152 is involved in cell cycle arrest, apoptosis, epithelial to mesenchymal transition, cell migration and invasion in gastric cancer.** *Cell Cycle* 2015, **14**(19):3112-3123.
  42. Nohata N, Abba MC, Gutkind JS: **Unraveling the oral cancer lncRNAome: Identification of novel lncRNAs associated with malignant progression and HPV infection.** *Oral Oncol* 2016, **59**:58-66.
  43. Schoppmann SF, Tamandl D, Roberts L, Jomrich G, Schoppmann A, Zwirtek R, Dubsy P, Gnant M, Jakesz R, Birner P: **HER2/neu expression**



- correlates with vascular endothelial growth factor-C and lymphangiogenesis in lymph node-positive breast cancer.** *Ann Oncol* 2010, **21**(5):955-960.
44. Hou J, Zhou Z, Chen X, Zhao R, Yang Z, Wei N, Ni Q, Feng Y, Yu X, Ma J *et al*: **HER2 reduces breast cancer radiosensitivity by activating focal adhesion kinase in vitro and in vivo.** *Oncotarget* 2016, **7**(29):45186-45198.
  45. Hongisto V, Aure MR, Makela R, Sahlberg KK: **The HER2 amplicon includes several genes required for the growth and survival of HER2 positive breast cancer cells - A data description.** *Genom Data* 2014, **2**:249-253.
  46. Sontrop HMJ, Reinders MJT, Moerland PD: **Breast cancer subtype predictors revisited: from consensus to concordance?** *BMC Med Genomics* 2016, **9**(1):26.
  47. Yang F, Lyu S, Dong S, Liu Y, Zhang X, Wang O: **Expression profile analysis of long noncoding RNA in HER-2-enriched subtype breast cancer by next-generation sequencing and bioinformatics.** *Onco Targets Ther* 2016, **9**:761-772.
  48. Chen CF, Dou XW, Liang YK, Lin HY, Bai JW, Zhang XX, Wei XL, Li YC, Zhang GJ: **Notch3 overexpression causes arrest of cell cycle progression by inducing Cdh1 expression in human breast cancer cells.** *Cell Cycle* 2016, **15**(3):432-440.
  49. Wang XH, Hong X, Zhu L, Wang YT, Bao JP, Liu L, Wang F, Wu XT: **Tumor necrosis factor alpha promotes the proliferation of human nucleus pulposus cells via nuclear factor-kappaB, c-Jun N-terminal kinase, and p38 mitogen-activated protein kinase.** *Exp Biol Med (Maywood)* 2015, **240**(4):411-417.
  50. Tang C, Jiang Y, Shao W, Shi W, Gao X, Qin W, Jiang T, Wang F, Feng S: **Abnormal expression of FOSB correlates with tumor progression and poor survival in patients with gastric cancer.** *Int J Oncol* 2016, **49**(4):1489-1496.
  51. Vogler M: **BCL2A1: the underdog in the BCL2 family.** *Cell Death Differ* 2012, **19**(1):67-74.
  52. Xu XW, Yang XM, Zhao WJ, Zhou L, Li DC, Zheng YH: **DNM1L, a key prognostic predictor for gastric adenocarcinoma, is involved in cell proliferation, invasion, and apoptosis.** *Oncol Lett* 2018, **16**(3):3635-3641.
  53. Huang X, Li X, Guo B: **KLF6 induces apoptosis in prostate cancer cells through up-regulation of ATF3.** *J Biol Chem* 2008, **283**(44):29795-29801.
  54. Shi X, Sun M, Liu H, Yao Y, Kong R, Chen F, Song Y: **A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer.** *Mol Carcinog* 2015, **54** Suppl 1:E1-E12.
  55. Yu W, Peng W, Jiang H, Sha H, Li J: **LncRNA HOXA11-AS promotes proliferation and invasion by targeting miR-124 in human non-small cell lung cancer cells.** *Tumour Biol* 2017, **39**(10):1010428317721440.
  56. Han Y, Yang YN, Yuan HH, Zhang TT, Sui H, Wei XL, Liu L, Huang P, Zhang WJ, Bai YX: **UCA1, a long non-coding RNA up-regulated in**

- colorectal cancer influences cell proliferation, apoptosis and cell cycle distribution.** *Pathology* 2014, **46**(5):396-401.
57. Griguolo G, Pascual T, Dieci MV, Guarneri V, Prat A: **Interaction of host immunity with HER2-targeted treatment and tumor heterogeneity in HER2-positive breast cancer.** *J Immunother Cancer* 2019, **7**(1):90.
58. Wang J, Xia S, Arand B, Zhu H, Machiraju R, Huang K, Ji H, Qian J: **Single-Cell Co-expression Analysis Reveals Distinct Functional Modules, Co-regulation Mechanisms and Clinical Outcomes.** *PLoS Comput Biol* 2016, **12**(4):e1004892.