



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

“ANÁLISIS DE LA EXPRESIÓN DEL RNA LARGO
NO CODIFICANTE *LINC01614* Y SU ASOCIACIÓN
CON LA SOBREVIDA GLOBAL EN PACIENTES
MEXICANAS CON CÁNCER DE MAMA”

T E S I S

QUE PARA OBTENER EL TÍTULO DE
QUÍMICA FARMACÉUTICA BIÓLOGA

PRESENTA

ILSE ALEJANDRA COLORADO GONZÁLEZ

ASESOR

DR. ALFREDO HIDALGO MIRANDA



CIUDAD UNIVERSITARIA, CDMX

2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

JURADO ASIGNADO

PRESIDENTE: Profesor: DR. FRANCISCO JAVIER PLASENCIA DE LA PARRA

VOCAL: Profesor: DRA. MARÍA ELENA IBARRA RUBIO

SECRETARIO: Profesor: DR. ALFREDO HIDALGO MIRANDA

1er. SUPLENTE: Profesor: DRA. LIZBETH ESMERALDA GARCÍA VELÁZQUEZ

2º SUPLENTE: Profesor: DR. EDGAR RICARDO VÁZQUEZ MARTÍNEZ

SITIO DONDE SE DESARROLLÓ EL TEMA:

**LABORATORIO DE GENÓMICA DEL CÁNCER
INSTITUTO NACIONAL DE MEDICINA GENÓMICA (INMEGEN)**

ASESOR DEL TEMA:

DR. ALFREDO HIDALGO MIRANDA



SUPERVISOR TÉCNICO:

DRA. MAGDALENA RÍOS ROMERO



SUSTENTANTE:

ILSE ALEJANDRA COLORADO GONZÁLEZ



CONTENIDO

	Página
RESUMEN	1
1. ABREVIATURAS Y SIGLAS	2
2. INTRODUCCIÓN	3
2.1. Anatomía de la mama	3
2.2. Generalidades del cáncer	4
2.3. Cáncer de mama	5
2.3.1. Tipos histológicos del cáncer de mama	5
2.3.2. Epidemiología del cáncer de mama	6
2.3.3. Factores de riesgo	8
2.3.4. Diagnóstico	9
2.4. Clasificación molecular del cáncer de mama	10
2.4.1. Características moleculares del cáncer de mama	12
2.4.2. Alteraciones moleculares del cáncer de mama	12
2.4.3. Características de los subtipos intrínsecos	13
2.4.4. Pruebas genómicas con genes codificantes	15
2.5. El descubrimiento de los RNAs no codificantes	17
2.5.1. Clasificación de los RNAs no codificantes	18
2.6. Generalidades de los RNAs largos no codificantes	18
2.6.1. Clasificación de los RNAs largos no codificantes	21
2.6.2. Función de los RNAs lagos no codificantes	23
2.7. El papel de los RNAs largos no codificantes en el cáncer	25
2.7.1. RNA largos no codificantes como biomarcadores del cáncer de mama	27
2.8. El RNA largo no codificante <i>LINC01614</i>	29
2.8.1. Características del <i>LINC01614</i>	30
2.8.2. Localización celular del <i>LINC01614</i>	31
3. ANTECEDENTES PARTICULARES	33
4. PLANTEAMIENTO DEL PROBLEMA	36

5. JUSTIFICACIÓN	37
6. HIPÓTESIS	38
7. OBJETIVO GENERAL	39
7.1. Objetivos particulares	39
8. METODOLOGÍA	40
8.1. Caracterización <i>in silico</i>	40
8.1.1. Análisis del potencial codificante del <i>LINC01614</i>	40
8.1.2. Predicción de estructuras secundarias del <i>LINC01614</i>	40
8.2 Análisis bioinformático	40
8.2.1. Análisis de supervivencia global en muestras independientes	40
8.2.2. Análisis de expresión diferencial	41
8.2.2.1. Selección de datos del análisis de expresión diferencial	41
8.3. Validación experimental	42
8.3.1. Selección de muestras del biobanco FUCAM-INMEGEN	42
8.3.1.1. Criterios de inclusión, exclusión y eliminación	42
8.3.2. Extracción de RNA total de las muestras de tumores de mama	43
8.3.3. Cuantificación de RNA total	43
8.3.4. Síntesis de cDNA	44
8.3.5. Cuantificación de la expresión del <i>LINC01614</i> mediante reacción en cadena de la polimerasa cuantitativa	45
8.3.6. Análisis estadístico	46
9. RESULTADOS	45
9.1. El <i>LINC01614</i> es un transcrito no codificante	47
9.2. Estructura secundaria del <i>LINC01614</i>	48
9.3. La expresión del <i>LINC01614</i> se asocia con la supervivencia global en muestras independientes de cáncer de mama	50
9.4. El <i>LINC01614</i> puede regular la expresión de transcritos implicados en el desarrollo del cáncer	52
Validación experimental	55
9.5. Características clínico-patológicas de la cohorte FUCAM-INMEGEN ...	56
9.6. Los niveles de expresión del <i>LINC01614</i> varían entre las pacientes Luminal y no Luminal	57

9.7. La disminución de la supervivencia global se asocia con la sobre-expresión del <i>LINC01614</i> en pacientes mexicanas con CaMa Luminal B	58
9.8. La supervivencia global del CaMa tiende a disminuir en pacientes jóvenes con sobre-expresión del <i>LINC01614</i>	62
9.9. La supervivencia global del CaMa disminuye en pacientes con tumores de bajo grado y sobre-expresión del <i>LINC01614</i>	63
10. DISCUSIÓN	64
11. CONCLUSIONES	73
12. PERSPECTIVAS	74
13. REFERENCIAS BIBLIOGRÁFICAS	73

ÍNDICE DE FIGURAS

	Página
Figura 1. Anatomía normal de la mama	3
Figura 2. Características distintivas del cáncer en tumores humanos	5
Figura 3. Estadísticas del cáncer de mama	7
Figura 4. Mutaciones moleculares en el cáncer de mama	13
Figura 5. Proporción de secuencias codificantes y no codificantes en el genoma humano	17
Figura 6. Biogénesis y procesamiento de los lncRNAs	20
Figura 7. Interactoma de los lncRNAs	21
Figura 8. Clasificación y procesos transcripcionales de los lncRNAs	23
Figura 9. Mecanismos de acción asociados a la función de los lncRNAs	24
Figura 10. lncRNAs y su asociación con el cáncer	26
Figura 11. Mapa de localización genómica del <i>LINC01614</i>	29
Figura 12. La localización celular del <i>LINC01614</i> es predominantemente nuclear	32
Figura 13. Los niveles de expresión del candidato <i>LINC01614</i> se asocian con la supervivencia global de las pacientes con CaMa	34
Figura 14. La alta expresión del <i>LINC01614</i> se relaciona con una menor probabilidad de supervivencia en pacientes con CaMa	35
Figura 15. <i>LINC01614</i> es un transcrito no codificante	48
Figura 16. Estructuras secundarias del <i>LINC01614</i>	49
Figura 17. La sobre-expresión del <i>LINC01614</i> se asocia con un menor tiempo de supervivencia global, especialmente en pacientes con CaMa Luminal B	51
Figura 18. Los niveles de expresión del <i>LINC01614</i> se correlacionan con un grupo de transcritos diferencialmente expresados	53
Figura 19. Niveles de expresión del <i>LINC01614</i> en la cohorte FUCAM-INMEGEN	58

Figura 20. La sobre-expresión del <i>LINC01614</i> es desfavorable para la supervivencia global de las pacientes mexicanas con CaMa, especialmente del subtipo Luminal B	61
Figura 21. La sobre-expresión del <i>LINC01614</i> tiende a ser un factor de mal pronóstico cuando las pacientes con CaMa tienen 50 años o menos	62
Figura 22. La sobre-expresión del <i>LINC01614</i> se asocia con un menor tiempo de supervivencia en pacientes con mejor desarrollo de la enfermedad	63

ÍNDICE DE TABLAS

	Página
Tabla 1. Clasificación molecular del CaMa y su aproximación por IHQ	11
Tabla 2. Pruebas de expresión génica para cáncer de mama empleadas en la práctica clínica	16
Tabla 3. Composición de la mezcla de reacción empleada para la síntesis de cDNA	44
Tabla 4. Condiciones del termociclador para la síntesis de cDNA	44
Tabla 5. Composición de la mezcla de reacción empleada para la reacción de qPCR	45
Tabla 6. Condiciones del termociclador para la reacción de qPCR	45
Tabla 7. Clasificación por subtipo molecular de las muestras independientes del GEO	50
Tabla 8. Principales transcritos con expresión diferencial en CaMa Luminal B con alta y baja expresión de <i>LINC01614</i>	54
Tabla 9. Características clínico-patológicas de la cohorte de pacientes FUCAM-INMEGEN	55
Tabla 10. Asociación clínica entre la expresión del <i>LINC01614</i> y las características clínico-patológicas en la cohorte FUCAM-INMEGEN	59

RESUMEN

El CaMa es el tumor más frecuente y es la principal causa de muerte en mujeres a nivel mundial. En México, la enfermedad constituye un problema relevante para la salud pública debido al incremento constante en las tasas de incidencia y mortalidad. En los últimos años, las tecnologías genómicas han permitido mejorar el diagnóstico y el pronóstico de las pacientes mediante el análisis de expresión de genes codificantes como biomarcadores. Sin embargo, estos genes constituyen un porcentaje muy pequeño del genoma (1-2%). Otros genes, como aquellos que corresponden a RNAs no codificantes, pueden aportar información relevante en estas patologías. Uno de éstos, los RNAs largos no codificantes (lncRNAs), son transcritos mayores a 200 nt que se caracterizan por regular la expresión génica desde el nivel epigenético hasta el nivel post-traducciona. Diversos estudios concuerdan con que numerosos lncRNAs se encuentran desregulados en el cáncer, presentando perfiles de expresión característicos. En el Laboratorio de Genómica del Cáncer, se identificó que los niveles de expresión del *LINC01614* se encuentran desregulados en tumores de mama. En el presente trabajo, se examinó el patrón de expresión del *LINC01614* en muestras independientes de tumores de mama, utilizando bases de datos públicas. Los resultados se validaron mediante un análisis de PCR en tiempo real en 97 muestras de tumores de mama de pacientes mexicanas. Como resultado del análisis de supervivencia, se identificó que la sobre-expresión del *LINC01614* se asocia con un pronóstico significativamente desfavorable, especialmente en CaMa Luminal B. Mediante un análisis de expresión diferencial, se identificó que en este subtipo, el *LINC01614* potencialmente regula los niveles de expresión de genes que participan en procesos relacionados al cáncer. La regulación positiva del *LINC01614* también se asoció con una menor supervivencia global en las pacientes con tumores de bajo grado (\leq II). En conclusión, los resultados sugieren que en pacientes mexicanas, los niveles de expresión del *LINC01614* influyen en la supervivencia global del CaMa. Esto permite resaltar su papel potencial como biomarcador de pronóstico de la enfermedad, tal como lo reportan otros grupos de investigación en otros tipos tumorales.

1. ABREVIATURAS Y SIGLAS

Aa	Aminoácido.
CaMa	Cáncer de mama.
cDNA	DNA complementario.
DNA	Ácido desoxirribonucleico.
ER	Receptor de Estrógenos (del inglés <i>Estrogen Receptor</i>).
eRNA	RNA potenciador (del inglés <i>Enhancer RNA</i>).
FDR	Tasa de falsos descubrimientos (del inglés <i>False Discovery Rate</i>).
GEO	<i>Gene Expression Omnibus</i> .
HER2	Receptor 2 del Factor de Crecimiento Epidérmico Humano (del inglés <i>HumanEpidermal Growth Factor Receptor 2</i>).
IARC	Agencia Internacional de la Investigación en Cáncer (<i>International Agency for Research on Cancer</i> , por sus siglas en inglés).
IHQ	Inmunohistoquímica.
Kb	Kilobase (1kb = 1000 pb de DNA ó RNA).
lincRNA	RNA largo no codificante intergénico (del inglés <i>Long intergenic non-coding RNA</i>).
lncRNA	RNA largo no codificante (del inglés <i>Long non-coding RNA</i>).
miRNA	Micro RNA.
mRNA	RNA mensajero.
NCBI	Centro Nacional para la Información Biotecnológica (<i>National Center for Biotechnology Information</i> , por sus siglas en inglés).
ncRNA	RNA no codificante (del inglés <i>Non-coding RNA</i>).
nt	Nucleótidos.
ORF	Marco abierto de lectura (del inglés <i>Open Reading Frame</i>).
piRNA	RNA asociado a Piwi (del inglés <i>Piwi-interacting RNA</i>).
PR	Receptor de Progesterona (del inglés <i>Progesterone Receptor</i>).
qPCR	Reacción en cadena de la polimerasa cuantitativa.
RNA	Ácido ribonucleico (del inglés <i>Ribonucleic acid</i>).
siRNA	RNA pequeño de interferencia (del inglés <i>Small interfering RNA</i>).
snRNA	RNA pequeño nuclear (del inglés <i>Small nuclear RNA</i>).
snoRNA	RNA pequeño nucleolar (del inglés <i>Small nucleolar RNA</i>).
RT-PCR	Reacción en cadena de la polimerasa con transcriptasa inversa.
TCGA	El Atlas del Genoma del Cáncer (<i>The Cancer Genome Atlas</i> , por sus siglas en inglés)

2. INTRODUCCIÓN

2.1. Anatomía de la mama

La mama se localiza en la pared torácica anterior, la cual se extiende desde la segunda hasta la sexta costilla (Drew, Cawthorn y Michell, 2007). Ésta se conforma principalmente de tejido adiposo y tejido conectivo, y contiene en menor proporción glándulas. En el tejido glandular, se encuentran los conductos galactóforos y las unidades ductal-lobular terminales (TDLU), que son la unidad morfológica y funcional de la glándula mamaria (Pinamonti y Zanconati, 2018). Cada TDLU comprende un lóbulo, que es un grupo de 10–100 acinos en forma de saco con función secretora, y un conducto terminal que lo conecta al sistema de conductos. En general, la mama está compuesta por 15 a 20 lóbulos (Pinamonti y Zanconati, 2018; Pandya y Moore, 2011). Próximos a la glándula mamaria, se localizan los ganglios linfáticos, un sitio crucial para la interacción del tumor con las células del sistema inmunológico, y una puerta de entrada para la diseminación de las células tumorales a otros sitios metastásicos (Figura 1) (Chatterjee *et al.* 2018).

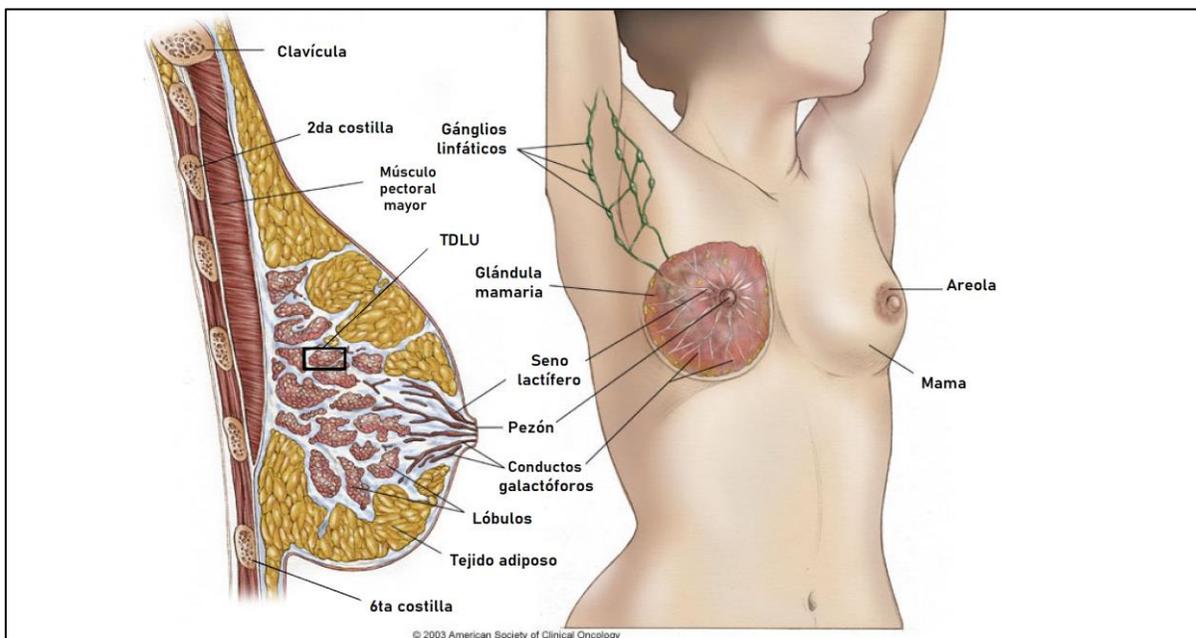


Figura 1. Anatomía normal de la mama. En la parte externa de la mama se muestra el pezón y la areola; al interior, se muestra el seno lactífero, los conductos galactóforos y las unidades ductal-lobular terminales (TDLU). Localizados junto a la glándula mamaria, los ganglios linfáticos son estructuras de importancia inmunológica. *Imagen modificada de A. Society of Clinical Oncology (2003).*

2.2. Generalidades del cáncer

En diferentes áreas de la glándula mamaria como los conductos, los lóbulos o el tejido entre ellos, se pueden desarrollar un conjunto de neoplasias que engloban al cáncer de mama (Feng *et al.*, 2018). Durante las últimas dos décadas, se ha establecido que el cáncer de mama surge por la acumulación de alteraciones genéticas y/o epigenéticas en diferentes genes relacionados con el cáncer (Campoy *et al.*, 2018).

De acuerdo con la Organización Mundial de la Salud (OMS), el cáncer se define como un proceso de crecimiento y diseminación descontrolado de células que puede aparecer en cualquier sitio del cuerpo, y que suele invadir el tejido circundante provocando metástasis en puntos distantes del organismo. El mecanismo de desarrollo del cáncer está asociado con la acumulación y selección sucesiva de alteraciones que permiten a las células sobrevivir, replicarse y evadir mecanismos reguladores de apoptosis, proliferación y del ciclo celular (Willingham *et al.*, 2004).

En particular, las células cancerosas son células transformadas, las cuáles han acumulado múltiples alteraciones genéticas que les permiten formar tumores con base en las capacidades que adquieren (Cree, 2011; Cahill *et al.*, 1999); estas son conocidas como “Características distintivas del Cáncer” (del inglés *Hallmarks of Cancer*) y fueron descritas en los años 2000 y 2011, por Hanahan y Weinberg (Figura 2). Las características distintivas incluyen capacidades como la señalización proliferativa sostenida, la insensibilidad a las señales inhibitoras del crecimiento, la evasión de la muerte celular programada, el potencial replicativo ilimitado, la inducción de la angiogénesis, la invasión de tejidos y metástasis, la desregulación del metabolismo energético y la evasión del sistema inmunológico. Adicionalmente, la inestabilidad genómica y la inflamación generada por las células del sistema inmune innato, son dos características que actúan favoreciendo la aparición de las ocho características anteriores.

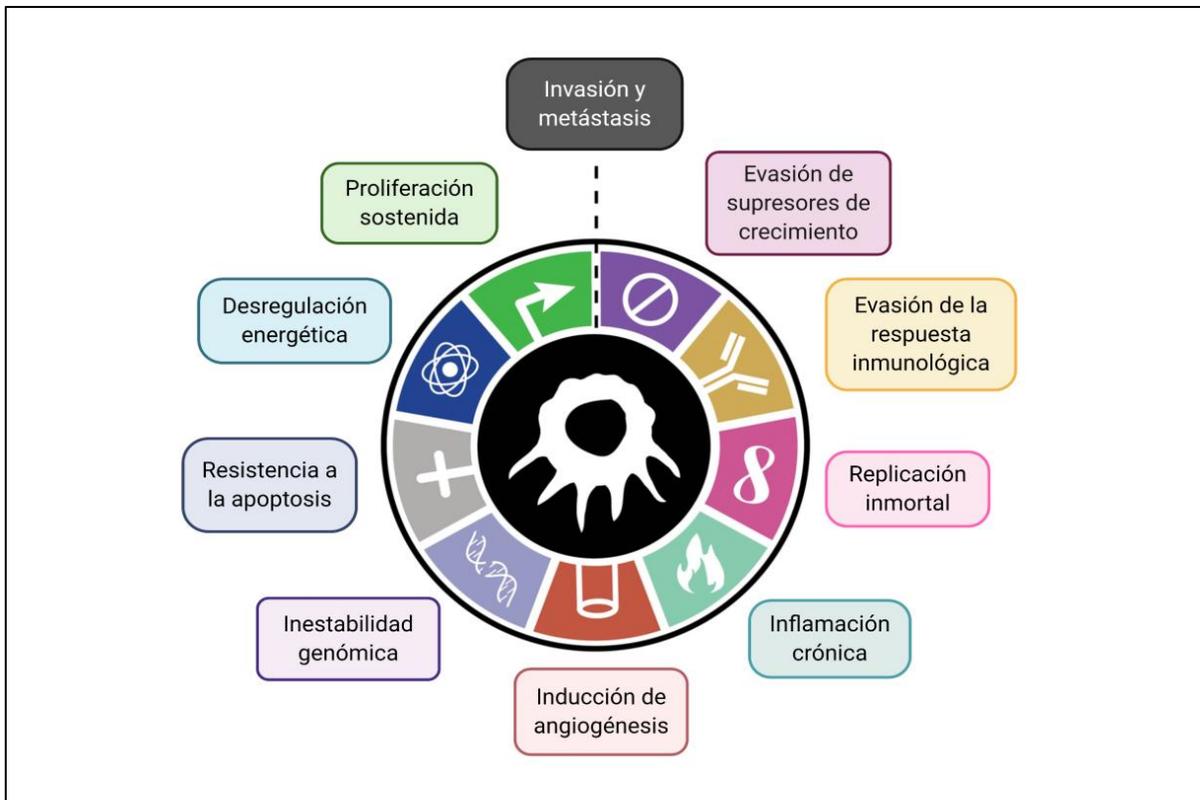


Figura 2. Características distintivas del cáncer en tumores humanos. Cuando las células normales evolucionan progresivamente a un estado neoplásico, adquieren capacidades funcionales que le permiten a las células tumorales sobrevivir, proliferar y diseminarse. *Imagen modificada de Meirson, Gil-Henn y Samson, 2019.*

2.3. Cáncer de mama

El cáncer de mama (CaMa) se caracteriza por el crecimiento de células malignas en los tejidos de la mama (Brandan y Villaseñor, 2006). Puesto que la mama se conforma de tejido altamente vascularizado, en ella se crea un ambiente favorable para la formación y progresión del tumor (Dellaire, Berman y Arceci, 2014).

2.3.1. Tipos histológicos del cáncer de mama

El 80-85% de los casos de CaMa surgen en la capa de células epiteliales situada en los conductos mamarios, rica en receptores de estrógenos (ER); estos tumores se conocen como carcinomas ductales. En contraste, los carcinomas lobulares crecen en la región lobular de la mama y son menos prevalentes, comprendiendo del 10 al 15% de los casos (Dillon, Guidi y Schnitt, 2014; Dellaire, Berman y Arceci, 2014). A

diferencia de los primeros, las mujeres con carcinoma lobular tienen hasta un 30% de riesgo de desarrollar un carcinoma invasivo, principalmente del tipo ductal y por lo general, en ambos senos. (Rosen *et al.*, 1978). Otros tipos de CaMa como el carcinoma mixto y el cáncer de mama inflamatorio constituyen el porcentaje restante de los casos. (Dillon, Guidi y Schnitt, 2014).

Para desarrollar un carcinoma invasivo, se requiere que ocurran una serie de alteraciones genéticas. Inicialmente, las alteraciones dan lugar a una hiperplasia (aumento en el número de células) y posteriormente, ocurre una hiperplasia atípica, en la cual hay una aparición de células anormales de la mama, así como la formación de un carcinoma *in situ* (CIS) con fenotipo no invasivo. Generalmente este permanece confinado a los lóbulos o a los conductos mamarios y no invade los tejidos circundantes. No obstante, en muchos casos, los tumores pueden adquirir alteraciones genéticas adicionales que resultan en la progresión a un carcinoma invasivo (IC). La diseminación de las células tumorales a través de los vasos linfáticos y la circulación sanguínea, culmina en metástasis hacia los ganglios linfáticos y órganos distantes como los pulmones, los huesos, el cerebro y la piel (Bartella *et al.*, 2007; Dellaire, Berman y Arceci, 2014).

2.3.2. Epidemiología del cáncer de mama

A nivel mundial, el CaMa es el tumor maligno más frecuente en las mujeres y es la primera causa de muerte por cáncer (Sánchez *et al.*, 2017). En las últimas décadas, la incidencia global del CaMa ha mostrado un aumento constante, con un incremento anual de 3.1%, iniciando con 641,000 casos en 1980 y alcanzando más de 1.6 millones en 2010 (Fray *et al.*, 2015; Fray *et al.*, 2018). De acuerdo a las estadísticas de GLOBOCAN (2018) reportadas por la Agencia Internacional de Investigación del Cáncer (IARC, por sus siglas en inglés), en ese año se diagnosticaron cerca de 2.1 millones de nuevos casos, y 626,679 pacientes fallecieron a causa de la enfermedad (Figura 3-A).

En México, el CaMa también ha tenido un incremento constante tanto en la incidencia como en la mortalidad en las últimas tres décadas (Sánchez *et al.*, 2017). En 2018, la IARC reportó que México se posiciona en el segundo lugar de los países de América Latina y el Caribe en dichos rubros, registrándose 27,283 nuevos casos diagnosticados, así como 6,884 decesos por la enfermedad (Figura 3-B). El Instituto Nacional de Estadística y Geografía (INEGI) indicó que en la mujer mexicana esta patología se convirtió en la tercera causa de muerte, después de enfermedades del corazón y la diabetes, y en la primera causa por cáncer.

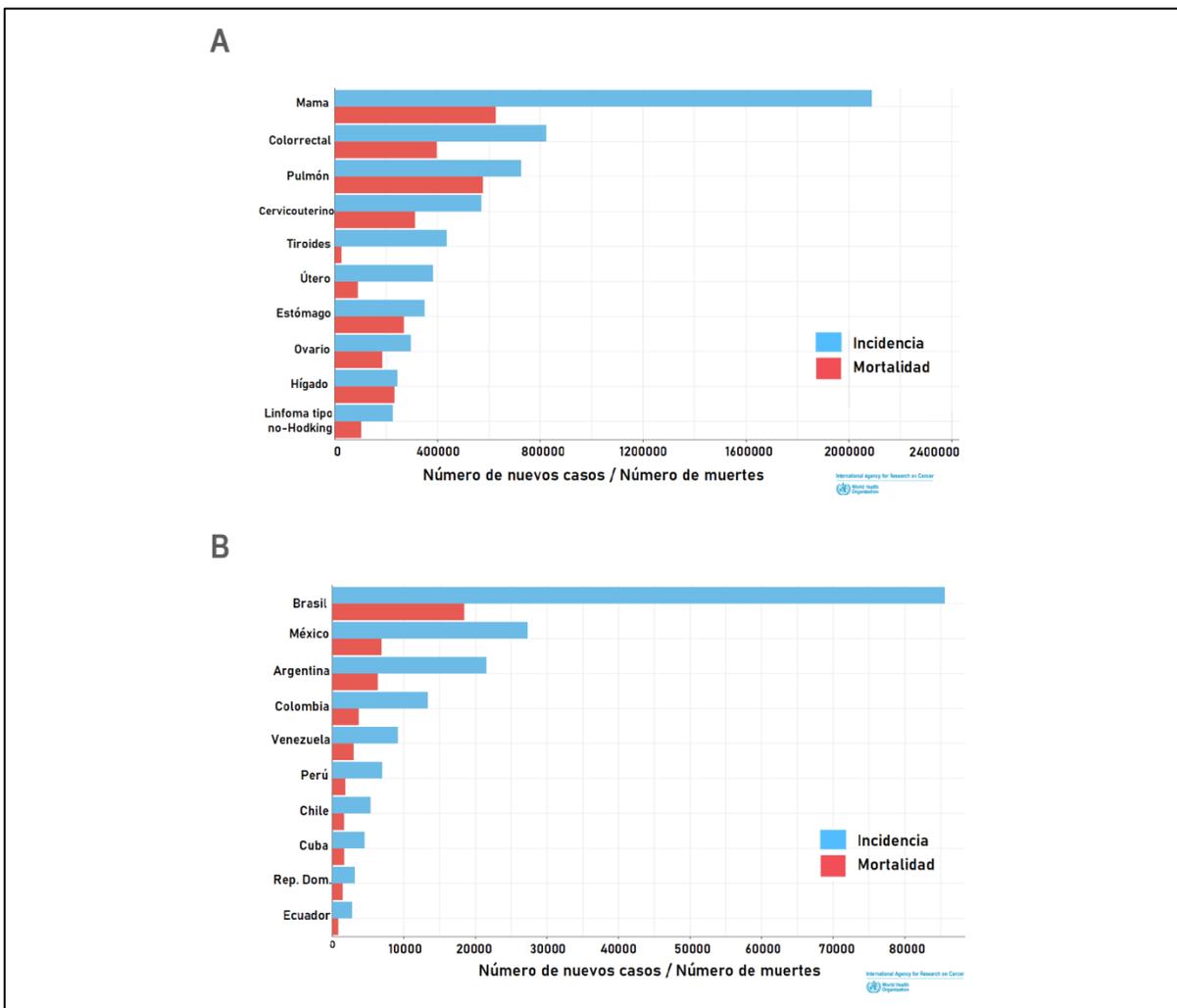


Figura 3. Estadísticas del cáncer de mama. A) El número estimado de nuevos casos y muertes a nivel mundial apunta a que el CaMa es el tumor más frecuente y es la principal causa de muerte por cáncer en mujeres. **B)** Dentro de los países de América Latina y el Caribe, México ocupa el segundo lugar en incidencia y mortalidad por CaMa. *Imagen modificada de Cancer Today, GLOBOCAN (2018).*

2.3.3. Factores de riesgo

El CaMa se asocia con factores ambientales, reproductivos y de estilo de vida, algunos de los cuales son potencialmente modificables, mientras que menos del 10% de los casos pueden atribuirse a una mutación genética heredada (Rojas y Stuckey, 2017). En su mayoría, el CaMa ocurre en la mujer y el número de casos es 100 veces mayor en mujeres que en hombres. Además del sexo, el envejecimiento es uno de los factores de riesgo más importantes para la enfermedad ya que la incidencia está altamente relacionada con el aumento de la edad, particularmente en mujeres de 50 años o más (Siegel *et al.*, 2017). Adicionalmente, se ha descrito que los factores reproductivos como la menarquia temprana, la menopausia tardía, la edad avanzada al tener el primer embarazo, la baja paridad y la ausencia de lactancia, pueden aumentar el riesgo del CaMa. Cada año de retraso en la menopausia aumenta el riesgo en un 3%, mientras que cada año de retraso en la menarquía o cada parto adicional disminuye el riesgo en un 5% y 10%, respectivamente (Washbrook, 2006; Dall y Britt, 2017). Las mujeres que han tenido un hijo antes de cumplir los 20 años tienen un menor riesgo que aquellas que no han dado a luz o que han tenido a su primer hijo después de los 30 años (Franca *et al.*, 2012). La exposición a estrógenos endógenos y exógenos también se asocia con un mayor riesgo de la enfermedad, especialmente de manera prolongada. En el caso de los estrógenos exógenos, las principales fuentes son los anticonceptivos orales y la terapia de reemplazo hormonal (TRH) (Sun *et al.*, 2017).

Por otra parte, se estima que cerca del 20% de los casos en todo el mundo se pueden atribuir a factores de riesgo modificables, incluida la obesidad, la inactividad física y el consumo de alcohol, ofreciendo potencialmente reducir la carga de la enfermedad al promover un estilo de vida saludable (Danaei, 2005).

Respecto a la predisposición genética, las mujeres con antecedentes de CaMa en familiares de primer grado tienen de 2 a 4 veces más probabilidades de desarrollar el cáncer, especialmente aquellas que son portadoras de mutaciones en los genes *BRCA1* (17q21) o *BRCA2* (13q13) (Key *et al.*, 2001; Goldgar *et al.*, 1994),

los cuales codifican para proteínas involucradas en la reparación del DNA a través del mecanismo de recombinación homóloga. Las mutaciones que se presentan en estos genes supresores de tumores muestran un patrón de herencia autosómico dominante. Una mutación de delección en cualquiera de ellos aumenta hasta 10 veces la probabilidad de desarrollar CaMa (Huen *et al.*, 2009; Chen *et al.*, 1999).

2.3.4. Diagnóstico

Generalmente, el diagnóstico del CaMa se basa en una prueba triple que comprende un examen clínico, estudios de imagen como la mastografía, el ultrasonido mamario (US), la resonancia magnética (RM) o la tomografía por emisión de positrones (PET), y una biopsia con aguja (Irwig, Macaskill y Houssami, 2002; Sánchez *et al.*, 2017). Esta última es la mejor técnica para diagnosticar el CaMa, ya que se requiere la confirmación de la enfermedad mediante criterios histológicos y citológicos (Akram *et al.*, 2017; Ellis *et al.*, 2012). Dentro de ellos, el grado tumoral evalúa la histopatología de los tumores a través de la caracterización del estadio de diferenciación, el pleomorfismo nuclear y la capacidad proliferativa, en tanto que la etapa clínica define el comportamiento de un tumor de acuerdo al tamaño y la ubicación del tumor, la afectación de los ganglios linfáticos y la presencia o ausencia de metástasis (Elston, 1999). Conjuntamente, el análisis por IHQ se utiliza para caracterizar la expresión de tres receptores, el receptor de estrógenos (ER), el receptor de progesterona (PR) y el receptor 2 del factor de crecimiento epidérmico humano (HER2), los cuales son los tres principales biomarcadores utilizados en la práctica clínica para clasificar el CaMa (Dellaire, Berman y Arceci, 2014).

La combinación del estado de los receptores con el grado histológico y el estadio del tumor es utilizada para predecir el pronóstico de las pacientes y su respuesta al tratamiento. Sin embargo, se ha observado que las pacientes presentan diferentes resultados clínicos y responden de maneras distintas a los tratamientos, aun cuando el diagnóstico histopatológico es el mismo. Esto podría atribuirse principalmente a que la clasificación basada en receptores hormonales no considera las bases moleculares y genómicas del CaMa (Dellaire, Berman y Arceci, 2014).

2.4. Clasificación molecular del cáncer de mama

Como consecuencia de la recurrencia y la resistencia a las terapias actuales, existe en la actualidad una alta incidencia en la mortalidad de las pacientes con CaMa (Suresh, 2018). La dificultad para lograr un diagnóstico eficaz y un tratamiento exitoso se debe principalmente a que el CaMa es una enfermedad compleja que incluye distintas entidades clínicas, morfológicas y moleculares, de modo que esta heterogeneidad tumoral no se puede explicar únicamente mediante los parámetros clínicos y los biomarcadores habitualmente utilizados en la clínica (Eroles *et al.*, 2012).

En el año 2000, Perou y colaboradores publicaron el primer artículo que clasifica al CaMa en subtipos intrínsecos. En éste, se postula que los tumores de mama pueden diferenciarse de acuerdo con sus patrones de expresión génica, permitiendo establecer las bases de la actual clasificación molecular. Inicialmente, en ese estudio se revelaron cuatro subtipos moleculares: *Luminal*, *HER2*, *Basal* y *Similar a mama normal*. Trabajos posteriores en una cohorte de pacientes más amplia, demostraron que el subtipo Luminal se subdivide en por lo menos dos grupos: *Luminal A* y *Luminal B*. Además, se identificó que cada subtipo molecular se asocia con un pronóstico diferente (Perou *et al.*, 2000; Eroles *et al.*, 2012; Sotiriou *et al.*, 2003). Posteriormente, la clasificación intrínseca fue refinada hasta la obtención de una lista de 50 genes llamada *PAM50*. Esta es actualmente comercializada como una herramienta diagnóstica que evalúa la expresión de dichos genes y cuenta con la suficiente información para distinguir entre los cinco subtipos moleculares del CaMa (Parker *et al.*, 2009).

De acuerdo con lo publicado en el Consenso Mexicano sobre Diagnóstico y Tratamiento Mamario (2019), los subtipos intrínsecos de la clasificación molecular se pueden aproximar mediante la técnica de IHQ (Tabla 1) (Sánchez *et al.*, 2017; Rakha *et al.*, 2007; Geyer, Marchiò y Reis-Filho, 2009; Bhargava *et al.*, 2009; Prat *et al.*, 2013).

Tabla 1. Clasificación molecular del CaMa y su aproximación por IHQ.

SUBTIPO MOLECULAR	APROXIMACIÓN POR INMUNOHISTOQUÍMICA
Luminal A	ER (+) PR >20% Ki67 <20% HER2 (-)
Luminal B	ER (+) PR <20% Ki67 >20% HER2 (-) ó (+)
HER2	ER (-) PR (-) HER2 (+)
Triple negativo	ER (-) PR (-) HER2 (-)

Ki67= Antígeno Ki-67, marcador de proliferación celular. La investigación realizada en torno a los cinco subtipos moleculares del CaMa (*Luminal A*, *Luminal B*, *HER2*, *Basal* y *Similar a mama normal*) definidos mediante análisis genómicos, ha permitido aproximarse a esta clasificación mediante metodologías más accesibles como la inmunohistoquímica, empleando biomarcadores rutinarios en la práctica clínica como ER, PR y HER2. Los tumores *Luminal A* y *Luminal B* se caracterizan por la expresión de los receptores hormonales ER y PR, y se diferencian entre sí por la expresión del marcador Ki67 y la presencia o ausencia de HER2. El subtipo *HER2* presenta enriquecimiento de la proteína HER2 pero no expresa receptores hormonales. Por último, la mayoría de los tumores basales no expresan los receptores hormonales ni la proteína HER2, de modo que estos tumores generalmente son conocidos como triple negativo.

En estudios posteriores, se describió un nuevo subtipo intrínseco denominado *Bajo en Claudina*, el cual consiste principalmente en tumores triple negativos que expresan niveles bajos de la proteína E-cadherina. De acuerdo con ello, las clasificaciones más recientes de CaMa definen la existencia de seis subtipos moleculares: *Luminal A*, *Luminal B*, *HER2 enriquecido*, *Basal*, *Similar a mama normal* y *Bajo en Claudina* (Prat *et al.*, 2010; Perou *et al.*, 2010). Sin embargo, aún existen dudas sobre la existencia del subtipo *Similar a mama normal*. Algunos autores sugieren que podría tratarse de un artefacto técnico, resultado de la contaminación con células normales de la mama (Weigelt *et al.*, 2010; Bastien *et al.*, 2012).

2.4.1. Características moleculares del cáncer de mama

El mecanismo exacto de la carcinogénesis en la glándula mamaria aún es desconocido. Recientemente, los estudios se han enfocado en caracterizar molecularmente la enfermedad a modo de comprender y refinar el proceso de formación y progresión de los tumores (Harbeck *et al.*, 2019). La evidencia sugiere que el CaMa se desarrolla a través de dos vías que son divergentes entre sí. La primera de ellas, conocida como la vía de bajo grado, se caracteriza por la ganancia de la región 1q, la pérdida de la región 16q y, con baja frecuencia, la amplificación de la región 17q12 (dentro de la cual se localiza el gen *ERBB2* que codifica para la proteína HER2), además de una firma de expresión génica en donde la mayoría de los genes están asociados con el fenotipo ER y un bajo grado tumoral. En contraste, la vía denominada vía de alto grado, se distingue por la pérdida de la región 13q, la ganancia de la región 11q13, la amplificación de la región 17q12, y un perfil de expresión de genes que se encuentran involucrados en el ciclo celular y en la proliferación celular. A la primera vía pertenecen principalmente los tumores *Luminal A* y, en menor medida, los tumores *Luminal B*, mientras que dentro de la segunda vía se incluyen los tumores *HER2 enriquecido* y los tumores *Basales* (Ellis *et al.*, 2012; Lopez-Garcia *et al.*, 2010).

2.4.2. Alteraciones moleculares del cáncer de mama

En un estudio realizado en CaMa con diagnóstico temprano, se detectó que los genes que se encuentran con mayor frecuencia mutados en las células tumorales son *TP53* (41%), *PIK3CA* (30%), *MYC* (20%), *PTEN* (16%), *CCND1* (16%), *ERBB2* (13%), *FGFR1* (11%) y *GATA3* (10%) (Figura 4) (Nik-Zainal *et al.*, 2016). Estos genes codifican para proteínas que regulan el ciclo celular, por ejemplo, p53 y la ciclina D1, y que inhiben vías oncogénicas importantes, como MYC, HER2 y FGFR1 (Yates y Desmedt, 2017). En general, los tumores Luminal A tienen una alta prevalencia de mutaciones en el gen *PIK3CA* (49%), mientras que un sello distintivo de los tumores basales (84%) es la alta frecuencia de mutaciones en el gen *TP53* (Harbeck *et al.*, 2019).

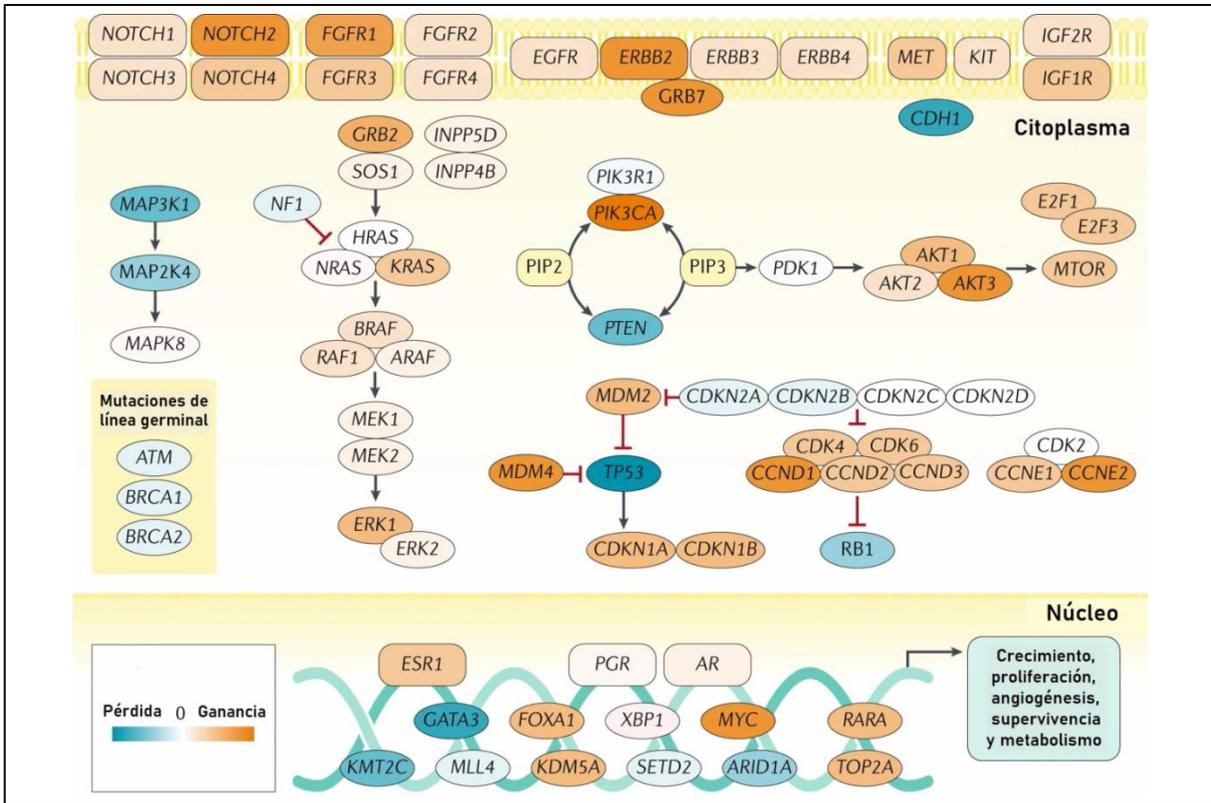


Figura 4. Alteraciones moleculares en el cáncer de mama. En el cáncer, la información genética de las células acumula mutaciones hasta presentar un fenotipo que les permite evadir el control del crecimiento, promover la proliferación, favorecer la angiogénesis y la supervivencia, así como modificar el metabolismo. En la imagen cada gen está sombreado de acuerdo con la frecuencia de alteración en los tumores de mama. El color naranja indica un alto nivel de amplificación y/o mutaciones de ganancia de función, en tanto que el color azul representa deleciones y/o mutaciones de pérdida de función. Las alteraciones se presentan predominantemente en los genes *TP53*, *PIK3CA*, *PTEN*, *CCND1*, *ERBB2* y *FGFR1*, *MYC* y *GATA3*. Imagen modificada de Harbeck et al., 2019.

2.4.3. Características de los subtipos intrínsecos

Luminal A. Es el subtipo más común y representa el 50-60% de los casos. El 85% de estos tumores se caracterizan por un bajo grado tumoral (De Kruijf et al., 2014). Presentan niveles altos de ER (ER+), y niveles bajos de genes relacionados con la proliferación celular (Yersal, 2014). Las pacientes tienen buen pronóstico (probabilidad de supervivencia, 93%) (Howlander et al., 2018), y la tasa de recaída es significativamente menor que la de otros subtipos (<5%) (Shim et al., 2014). La metástasis es común en los huesos (70%), mientras que en el hígado, los pulmones y el cerebro, ocurre en menos del 10% de las pacientes (Kennecke et al., 2010).

Luminal B. Estos tumores comprenden entre el 15-20% de los casos. Tienen un fenotipo agresivo, un grado tumoral más alto (93% de los casos) y un peor pronóstico (probabilidad de supervivencia, 90%) (Creighton, 2012; De Kruijf *et al.*, 2014; Howlader *et al.*, 2018). Se diferencian del CaMa Luminal A en la alta expresión de genes relacionados con la proliferación celular como: Homólogo de oncogén viral de mieloblastosis (*v-MYB*), Gamma glutamil hidrolasa (*GGH*), Proteína transmembranal asociada a lisosomas 4- β (*LAPTMB4*), Proteína de unión al elemento de nucleasa sensible-1 (*NSEP1*) y Ciclina E1 (*CCNE1*). Algunos muestran un incremento en la expresión de genes que codifican para receptores de factores de crecimiento como: Receptor 2 del factor de crecimiento epidérmico humano (*HER2*), Receptor 1 del factor de crecimiento de fibroblastos (*FGFR1*) y Receptor 1 del factor de crecimiento similar a la insulina (*IGF-1R*) (Reis-Filho *et al.*, 2010). Aproximadamente el 20%-30% de los tumores HER2+ son clasificados en este subtipo (Tran y Bedard, 2011).

La supervivencia global de las pacientes no tratadas es similar a la de pacientes del grupo no Luminal (HER2 y Basal), que se reconocen ampliamente como tumores de alto riesgo (Hu *et al.*, 2006). Además, presentan tasas de recurrencia más altas (8%), en comparación con el subtipo Luminal A (Hu *et al.*, 2006; Shim *et al.*, 2014).

HER 2 enriquecido. En este subtipo se ubica el 15-20% de los casos. La positividad a HER2 confiere un comportamiento biológico y clínico más agresivo (probabilidad de supervivencia, 83%) (Yersal, 2014; Howlader *et al.*, 2018). Morfológicamente, los tumores son muy proliferativos; el 75% de ellos tienen un alto grado tumoral y más del 40% tienen mutaciones en p53 (Tsutsui *et al.*, 2003). En ausencia de tratamiento, los tumores HER2 positivos se asocian con un pronóstico desfavorable (Yersal, 2014).

Basal. Representan del 8-37% de todos los casos. Se asocia con un alto grado tumoral (93% de los casos) (De Kruijf *et al.*, 2014) e índices mitóticos y proliferativos excepcionalmente altos (ej. Ki67>50%) (Rakha *et al.*, 2009; Wang *et al.*, 2016). La mayoría de estos tumores tienen un comportamiento clínico altamente agresivo

(probabilidad de supervivencia, 77%) (Howlander *et al.*, 2018) y alta tasa de metástasis a los huesos (37%) y los pulmones (32%) (Wu *et al.*, 2017). Dentro de este subtipo, se ubican cerca del 75% de las pacientes con una mutación en el gen *BRCA1* (Foulkes *et al.*, 2004).

2.4.4. Pruebas genómicas con genes codificantes

Para evaluar parámetros como el riesgo de recaída o la mortalidad de las pacientes en estadios tempranos del CaMa, se desarrollaron pruebas genómicas basadas en una firma predictiva (Tabla 2). Estas pruebas determinan la expresión simultánea de un gran número de genes que juegan un papel importante en el pronóstico de la enfermedad, así como en la respuesta a un determinado tratamiento. Entre las pruebas más utilizadas se encuentran MammaPrint®, Oncotype DX®, EndoPredict® y Prosigna™. Éstas dos últimas combinan los datos de expresión génica con los parámetros clínico-patológicos de las pacientes, para así mejorar la estimación del riesgo de recaída (Martínez-Férez *et al.*, 2018; Gökmen-Polar y Badve, 2012).

El primer ensayo de expresión génica fue desarrollado en el año 2002. La firma predictiva Mammaprint® evalúa la expresión de 70 genes y distingue entre las pacientes con alta probabilidad de supervivencia sin metástasis y las pacientes con riesgo de desarrollar metástasis a distancia (dentro de los primeros 5 años posteriores al diagnóstico). Dentro de las funciones de estos genes se encuentran procesos asociados al cáncer, como: la apoptosis, la autosuficiencia en las señales de crecimiento, la insensibilidad a la inhibición de crecimiento, el potencial replicativo ilimitado y la invasión y metástasis (Giuliano *et al.*, 2017; Van 't Veer *et al.*, 2002).

Por otra parte, el ensayo de expresión génica más ampliamente utilizado es la firma de 21 genes Oncotype DX® (Sinn *et al.*, 2013). Ésta prueba proporciona información del riesgo de recaída en 5 o 10 años. Las pacientes se estratifican en tres grupos: *Bajo riesgo* (el beneficio clínico de la quimioterapia es mínimo), *Riesgo intermedio* (no está claro si el efecto benéfico de la quimioterapia supera los riesgos

y las complicaciones de sus efectos tóxicos) y *Alto riesgo* (hay una alta probabilidad de recurrencia, y los beneficios de la quimioterapia son mayores a los riesgos) (Sparano *et al.*, 2015). Los genes evaluados se encuentran implicados en la señalización estrogénica (*ESR1*, *PGR*, *BCL2*, *SCUBE2*), relacionados con HER2 (*GRB7* y *ERBB2*) y asociados a la proliferación celular (*MKI67*, *AURKA*, *BIRC5*, *CCNB1* y *MYBL2*) (Lal *et al.*, 2017).

Tabla 2. Pruebas de expresión génica para cáncer de mama empleadas en la práctica clínica.

PRUEBA CARACT.	MammaPrint®	Oncotype DX®	EndoPredict®	Prosigna™ (PAM50)
Compañía	Agendia (Irvine,CA)	Genomic Health (Redwood City, CA)	Myriad/Sividon Diagnostics	NanoString Technologies (Seattle, WA)
Muestra	Tejido fresco, congelado o FFPE	FFPE	FFPE	FFPE
No. de genes evaluados	70	- 16 relacionados con cáncer - 5 control	- 8 relacionados con cáncer - 3 control	- 50 relacionados con cáncer - 5 control
Tecnología	Microarreglos	RT-PCR cuantitativa	RT-PCR cuantitativa	RT-PCR cuantitativa
Valor predictivo	-	+	-	-
Valor pronóstico	+	+	+	+

FFPE= Tejido embebido en parafina y fijado con formalina.

En conclusión, el análisis genómico del CaMa ha favorecido el desarrollo de diversas herramientas de predicción de riesgo y respuesta al tratamiento. En específico, los perfiles de expresión génica han permitido generar una mejor clasificación de los tumores, demostrando su utilidad para detectar patrones de expresión que tienen un mayor valor predictivo que los parámetros clínicos tradicionales (Hidalgo-Miranda y Jiménez-Sánchez, 2009).

2.5. El descubrimiento de los RNAs no codificantes

A mediados del siglo pasado, el RNA había sido considerado exclusivamente una molécula intermediaria en el flujo de la información genética del DNA a la proteína (Morris y Mattick, 2014). Recientemente, se ha demostrado que del 93% del genoma humano que se transcribe en RNA, únicamente 1-2% de estas secuencias de nucleótidos codifican proteínas (Figura 5). Por lo tanto, el 98% restante corresponde a transcritos que carecen parcial o completamente de la capacidad para codificar proteínas (Dunham et al., 2012; Kapranov et al., 2007).

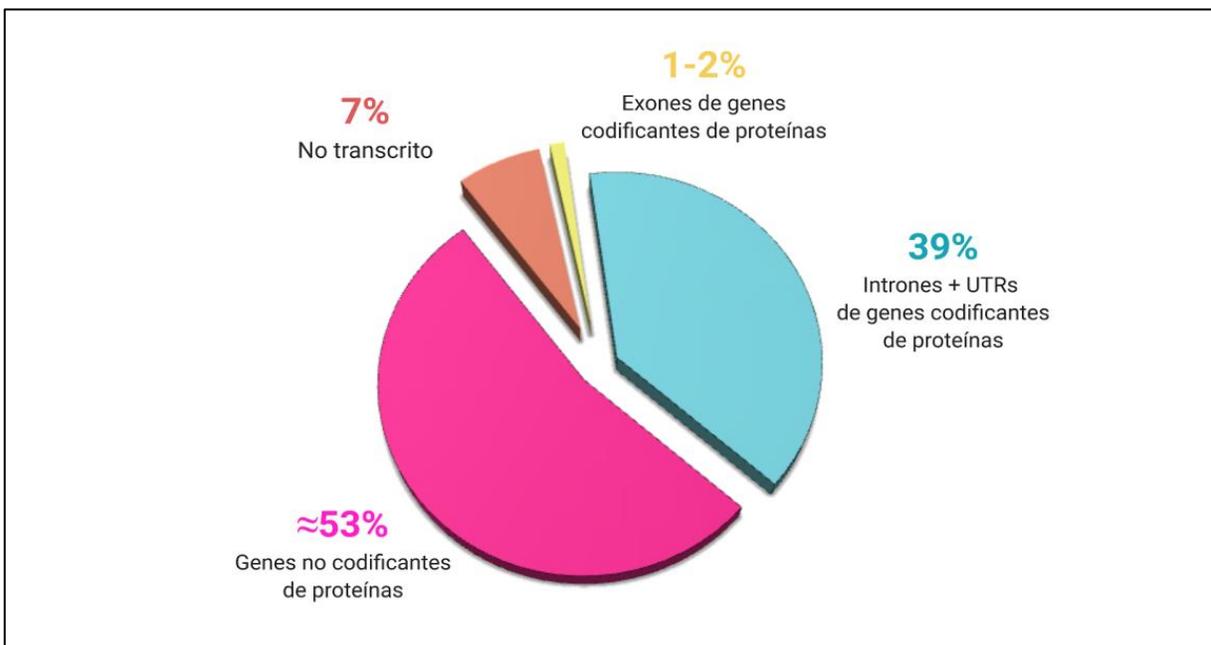


Figura 5. Proporción de secuencias codificantes y no codificantes en el genoma humano. Los datos publicados por el Proyecto ENCODE (*Encyclopedia of DNA Elements*) indican que el 93% del DNA se transcribe activamente. De estos transcritos, el 39% del genoma representa a los genes codificantes de proteínas, abarcando desde el promotor hasta la señal de poliadenilación, pero sin considerar un 1-2% que incluye a los exones de los genes codificantes de proteínas. El porcentaje restante, cercano a 53%, corresponde a regiones donde no hay presencia de genes codificantes. *Figura creada en BioRender.*

En el pasado, las secuencias no codificantes se asociaban con el término "basura evolutiva". Las moléculas que se generan a partir de su transcripción, actualmente son llamadas RNAs no codificantes (ncRNAs), y se caracterizan por regular mecanismos moleculares importantes (Beermann *et al.*, 2016).

Uno de los descubrimientos relevantes en torno a estas moléculas indica que el tamaño de la porción no codificante del genoma está relacionado con la complejidad de un organismo. Este argumento se basa en la observación de que la porción codificante de proteínas se ha mantenido constante a través del tiempo, mientras que la porción de RNAs no codificantes ha aumentado significativamente durante la evolución de los organismos (Mattick, 2004; Taft *et al.*, 2007).

2.5.1. Clasificación de los RNAs no codificantes

De acuerdo con lo reportado por el Proyecto GENCODE, más del 60% del transcriptoma humano está constituido por RNAs no codificantes. Su clasificación se basa en dos categorías principales: RNAs no codificantes estructurales y RNAs no codificantes reguladores (Ponting, Oliver y Reik, 2009). Los estructurales comprenden a los RNAs ribosomales (rRNAs) y los RNAs de transferencia (tRNAs). Los reguladores se dividen en tres grupos, de acuerdo con su tamaño: pequeños, medianos y largos (Dahariya *et al.*, 2019).

Los ncRNAs con un tamaño de 20-50 nucleótidos (nt) se denominan RNAs pequeños no codificantes, e incluyen principalmente a los miRNAs, siRNAs y piRNAs. Los RNAs pequeños nucleares (snRNAs) y los RNAs pequeños nucleolares (snoRNAs) se clasifican como RNAs medianos no codificantes, con un tamaño de 50-200 nt. Por último, los ncRNAs de más de 200 nt de longitud son conocidos como RNAs largos no codificantes (lncRNAs), y conforman cerca del 80% de todos los ncRNAs (Beermann *et al.*, 2016; Kazimierczyk *et al.*, 2020).

2.6. Generalidades de los RNAs largos no codificantes

Los lncRNAs representan un grupo muy heterogéneo de RNAs reguladores en relación a sus características, localización y mecanismos de acción. Aunque son pocas las similitudes estructurales y funcionales entre todos ellos, se definen como transcritos mayores a 200 nt de longitud de hasta 100 kb, con ausencia de potencial codificante (Kung, Colognori y Lee, 2013; Bassett *et al.*, 2014).

En general, los lncRNAs carecen de potencial codificante para proteínas, sin embargo, análisis recientes indican que algunos de estos transcritos pueden generar péptidos pequeños (Ingolia, Lareau y Weissman, 2011). Sin embargo, esto no excluye la potencial doble naturaleza del lncRNA, de modo que podría existir como un transcrito o como un producto proteico (Rinn y Chang, 2012). Aunado a ello, se ha reportado que es posible que muchas especies de lncRNAs se encuentren mal clasificadas, siendo que realmente pueden codificar proteínas funcionalmente relevantes (Anderson *et al.*, 2015).

La biogénesis de estas moléculas se produce en el núcleo (Beermann *et al.*, 2016) (Figura 6). Los lncRNAs son sintetizados por la RNA polimerasa II y son procesados de forma semejante a los mRNAs, con *capping*, *splicing* y poliadenilación (Ulitsky y Bartel, 2013). Así mismo, algunos lncRNAs se transcriben por la RNA polimerasa III (Dieci *et al.*, 2007). Generalmente presentan en el extremo 5' la adición del nucleótido modificado 7-metilguanosina (cap) y el 60% presenta poliadenilación en el extremo 3' (Yang *et al.*, 2011; Cheng *et al.*, 2005). Los lncRNAs que muestran la señal de poliadenilación tienen una mayor estabilidad que aquellos que se encuentran pobremente poliadenilados o sin poliadenilación (Wilusz *et al.*, 2012). La mayoría de los lncRNAs se procesan por reacciones de *splicing* (98%), mostrando una tendencia a presentar dos exones (42%), tres exones (26%), cuatro exones (15%) y cinco exones (7%) (Derrien *et al.*, 2012).

El paso final en la síntesis es la formación de estructuras termodinámicamente estables. A nivel de la estructura primaria, la actividad funcional se basa en el apareamiento de bases de Watson-Crick, que facilita las interacciones directas con moléculas de DNA y otras especies de RNAs como los mRNAs y los miRNAs. Las estructuras secundarias incluyen hélices, horquillas, protuberancias y pseudonudos. Éstas se forman mediante el apareamiento de bases o a través de interacciones con la ribosa de la estructura del RNA, permitiendo la formación de complejos con proteínas, péptidos y compuestos de bajo peso molecular (Novikova *et al.*, 2013; Cruz y Westhof, 2009; Mercer y Mattick, 2013; Kazimierczyk *et al.*, 2020).

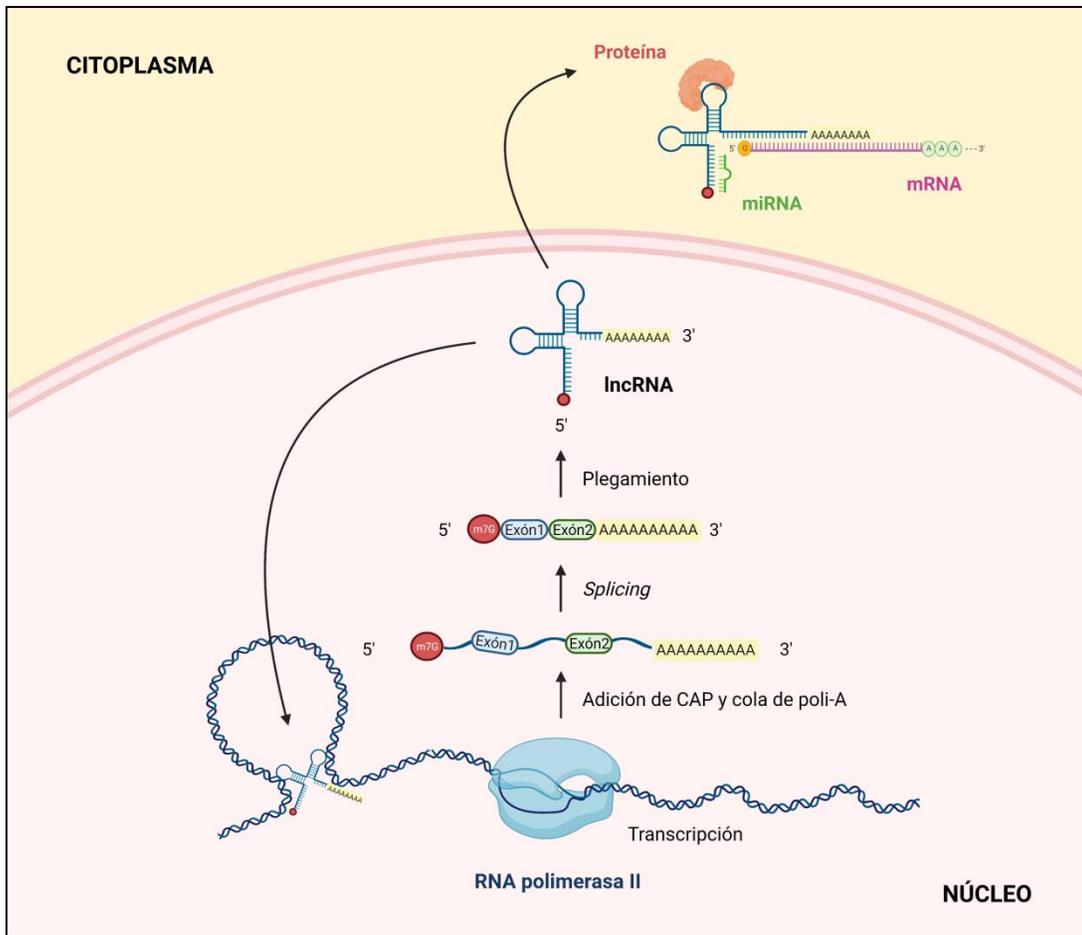


Figura 6. Biogénesis y procesamiento de los lncRNAs. El núcleo es el sitio de la biogénesis y el procesamiento de los lncRNAs. Estos eventos comparten muchas características con el de los RNAs codificantes para proteínas. La mayoría se transcriben por la RNA pol II, mientras que otros albergan estructuras que tienen más probabilidades de ser transcritas por la RNA pol III. Al igual que los mRNAs, se les añade una 7-metilguanosa en el extremo 5', se poliadenilan en 3' y se procesan post-transcripcionalmente (*splicing*). Posteriormente, los lncRNAs forman estructuras termodinámicamente estables y son liberados al citoplasma o bien, son retenidos en el núcleo y reclutados hacia la cromatina. *Figura creada en Biorender.*

Como consecuencia de su interacción con diversas biomoléculas y compuestos (Figura 7), los lncRNAs pueden regular la expresión génica en los niveles epigenético, transcripcional, post-transcripcional, traduccional y post-traduccional. Esto les permite desempeñar papeles importantes en procesos biológicos esenciales como la remodelación de la cromatina, la activación y represión transcripcional, el procesamiento de RNAs, así como la traducción de los mRNAs (Zhang *et al.*, 2019).

Pese a su amplia actividad funcional, los lncRNAs aún constituyen un componente del genoma que se encuentra poco explorado. El descubrimiento de los lncRNAs y su estudio, tiene una gran relevancia para el entendimiento de la biología humana y múltiples procesos patológicos (Guigó, 2013).

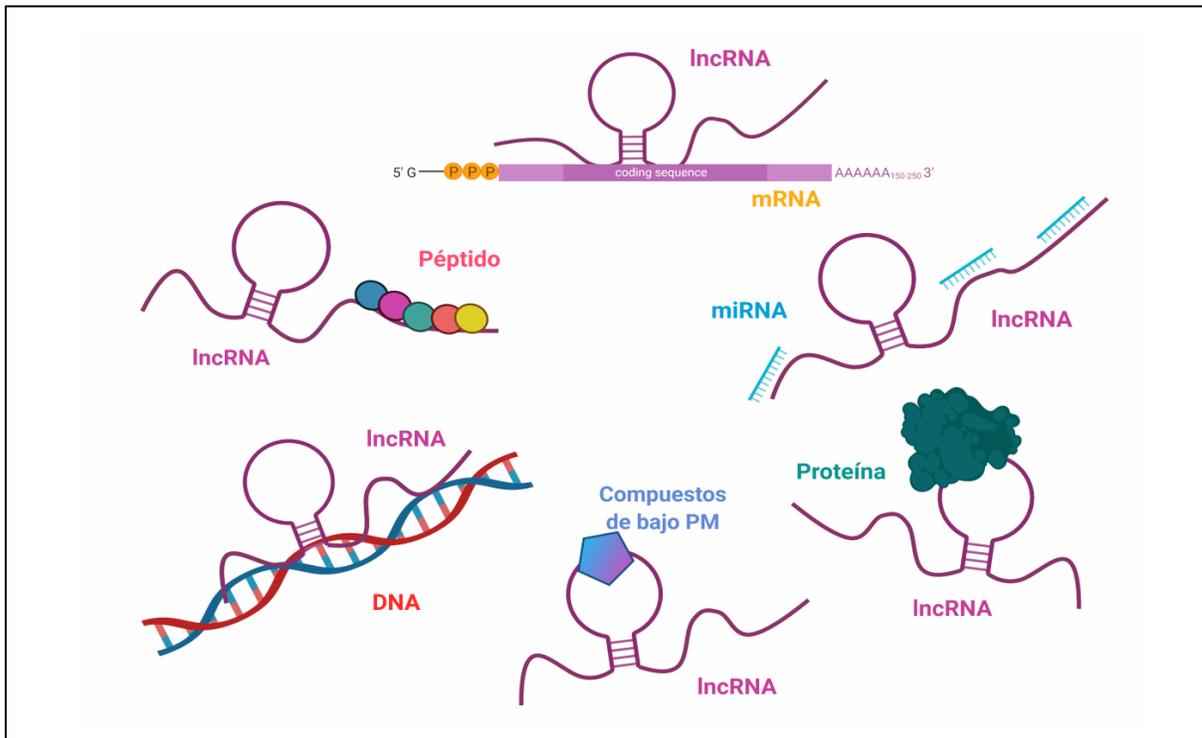


Figura 7. Interactoma de los lncRNAs. La interacción de los lncRNAs con una amplia gama de biomoléculas y compuestos, tales como el DNA, los mRNAs, los miRNAs, las proteínas, los péptidos, y compuestos de bajo peso molecular, forma el interactoma de estas moléculas, el cual está involucrado en muchos procesos celulares fundamentales para la vida. *Figura creada en BioRender.*

2.6.1. Clasificación de los RNAs largos no codificantes

Con base en su localización en el genoma, los RNAs largos no codificantes se clasifican en génicos e intergénicos. Los lncRNAs génicos se dividen en tres grupos: exónicos (16%), cuando la secuencia de nucleótidos del lncRNA coincide con la posición genómica del exón de un gen codificante; intrónicos (19%), si el lncRNA se localiza dentro del intrón de un gen codificante; y sobrelapados (1%), cuando el gen codificante se ubica al interior de un intrón del lncRNA. La mayoría de los lncRNAs no intersectan con ningún gen codificante, por lo que se agrupan en el grupo de los lncRNAs intergénicos (64%) (Derrien *et al.*, 2012).

La clasificación GENCODE es otro tipo de agrupación que se basa en la asociación de los lncRNAs con los genes codificantes. De acuerdo con la posición en la que se relacionan entre sí y sus procesos transcripcionales, se pueden clasificar en: sentido, antisentido, intrónicos, bidireccionales, intergénicos y potenciadores (Figura 8) (Dahariya *et al.*, 2019; Beermann *et al.*, 2016).

Los lncRNAs sentido se transcriben a partir de la cadena sentido en la que se localizan los genes que codifican proteínas, conteniendo en su estructura los exones de dichos genes. Este tipo de lncRNAs se empalman parcialmente con el gen codificante, o cubren la secuencia completa del gen codificante a través de un intrón. Los lncRNAs antisentido, por el contrario, se sintetizan a partir de la cadena antisentido a los genes que codifican proteínas, y se superponen con las regiones exónicas, intrónicas, o cubren toda la secuencia codificante con un intrón de su estructura (Ma, Bajic y Zhang, 2013).

Los lncRNAs que se transcriben a partir de las regiones intergénicas se denominan lncRNAs intergénicos (lincRNAs), en tanto que los que provienen de los intrones de los genes que codifican proteínas se denominan lncRNAs intrónicos (Ma, Bajic y Zhang, 2013). Los lincRNAs representan el grupo más grande de los lncRNAs; se ubican entre dos genes que codifican proteínas pero están al menos a 1 kb de distancia del gen codificante más cercano y son unidades transcripcionalmente independientes (Ponting, Oliver y Reik, 2009). Los lncRNAs intrónicos son aquellos que empalman su secuencia con las regiones intrónicas de un segundo transcrito (gen codificante para proteína), ya sea en la orientación sentido o bien, antisentido (Losko, Kotlinowski y Jura, 2016).

Por su parte, los lncRNAs bidireccionales se localizan en la cadena opuesta al gen que codifica proteínas, cuya transcripción inicia a menos de 1 kb de distancia. Ambos genes comparten un promotor bidireccional en común (Bayoglu y Cengiz, 2017). Por último, los lncRNAs *enhancers* (eRNAs) se transcriben a partir de las regiones genómicas conocidas como *enhancers* o de alguna otra región que funja como elemento de regulación (Zhang *et al.*, 2019).

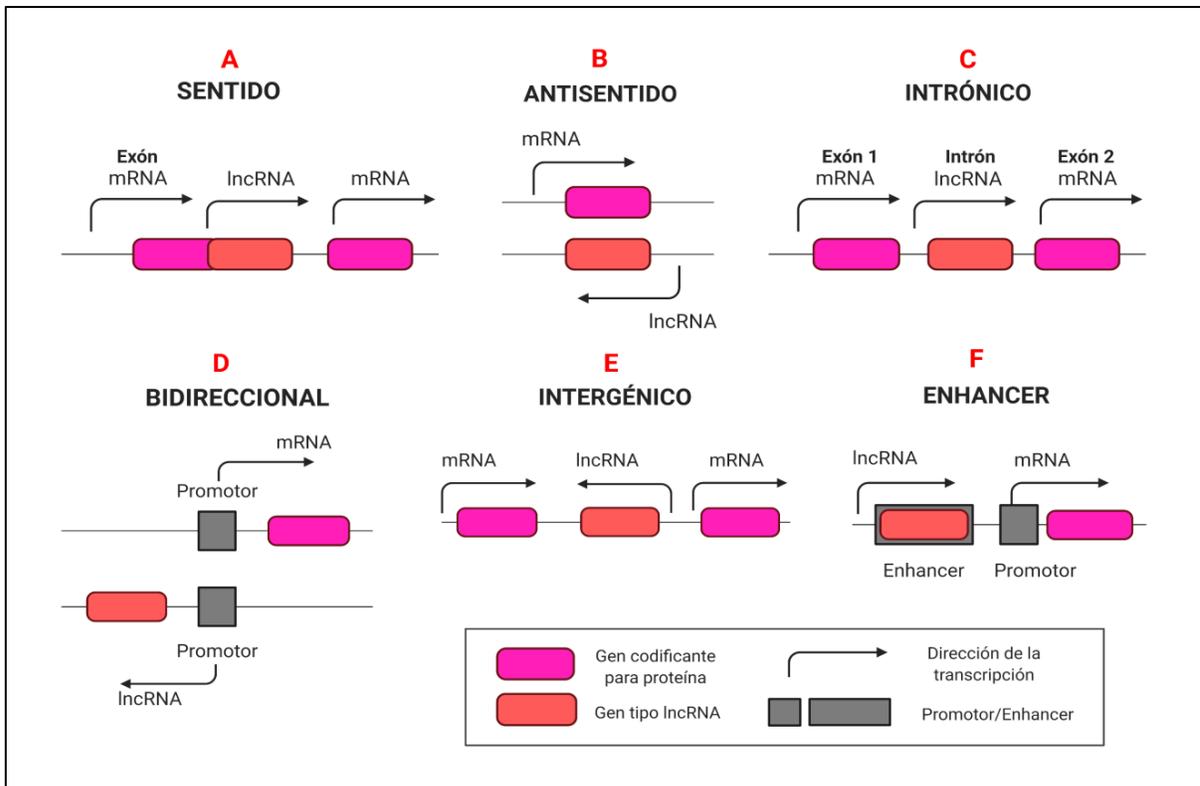


Figura 8. Clasificación y procesos transcripcionales de los lncRNAs. **A)** Los lncRNAs sentido se ubican dentro de la misma cadena que el gen codificante para proteína y se pueden superponer a uno o más exones de este. **B)** Los transcritos de los lncRNAs antisentido se localizan en la cadena complementaria al gen codificante para proteína, y complementan parcial o totalmente dicha secuencia. **C)** Los lncRNAs intrónicos se generan a partir de los intrones de los genes codificantes. **D)** Los lncRNAs bidireccionales comparten la misma región promotora con los genes codificantes para proteínas, pero se transcriben en la dirección opuesta. **E)** Los lncRNAs intergénicos (lincRNAs) se localizan entre los genes codificantes para proteínas y se transcriben de forma independiente. **F)** Los lncRNAs *enhancers* o potenciadores (eRNAs) se producen a partir de las regiones *enhancers* de los genes codificantes para proteínas. *Figura creada en BioRender.*

2.6.2. Función de los RNAs largos no codificantes

En el núcleo, los lncRNAs están funcionalmente implicados en procesos de regulación génica, los cuales incluyen activación y represión de la transcripción, regulación a nivel epigenético y procesamiento alternativo de los mRNAs. Los lncRNAs localizados en el citoplasma están involucrados en la regulación de la estabilidad y la degradación de los mRNAs. Además, pueden participar en la traducción y controlar algunas de las modificaciones post-traduccionales de las proteínas (Figura 9) (Hombach y Kretz, 2016; Kung, Colognori y Lee, 2013; Batista y Chang, 2013; Zhang *et al.*, 2019).

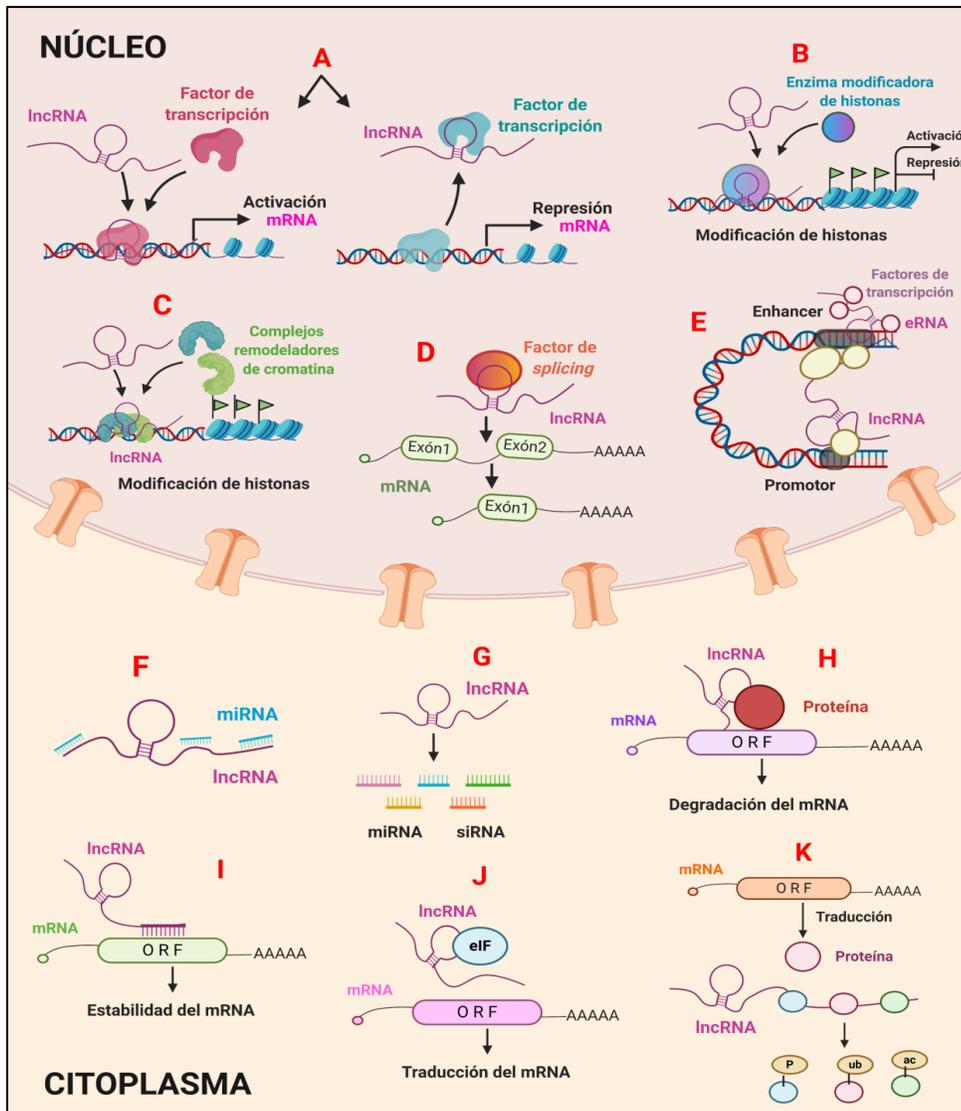


Figura 9. Mecanismos de acción asociados a la función de los lncRNAs. **A-E: Funciones de los lncRNAs en el núcleo.** **A)** Los lncRNAs se unen a factores de transcripción para activar o reprimir la expresión génica. **B)** Los lncRNAs interactúan con enzimas modificadoras de histonas para activar o reprimir la transcripción génica. **C)** Muchos lncRNAs regulan la transcripción actuando como andamios moleculares para favorecer las interacciones proteína-proteína de los complejos remodeladores de cromatina. **D)** Los lncRNAs interactúan con factores de *splicing* para regular el procesamiento alternativo de los mRNAs. **E)** Los transcritos de *enhancers* pueden generar lncRNAs *enhancers* (eRNAs) o lncRNAs asociados a *enhancers*; los primeros pueden captar factores de transcripción en el *enhancer*, mientras que los segundos regulan la transcripción en *cis* y en *trans*. **F-K: Funciones de los lncRNAs en el citoplasma.** **F)** Los lncRNAs actúan como esponjas de miRNAs, secuestrando estas moléculas lejos de sus mRNAs blanco, y favoreciendo la regulación de la expresión de los genes. **G)** Los lncRNAs pueden ser precursores de miRNAs y siRNAs. **H)** Mediante la unión a proteínas, los lncRNAs están implicados en la degradación de los mRNAs. **I)** Los lncRNAs se unen directamente al mRNA y regulan su estabilidad. **J)** Los lncRNAs se unen al complejo de iniciación de la traducción eucariota eIF para regular la traducción del mRNA. **K)** Los lncRNAs interactúan con las proteínas y controlan su fosforilación, acetilación y ubiquitinación, a nivel post-traduccional. *Figura creada en BioRender.*

La primera función estudiada de los lncRNAs es la relacionada con la participación en la inactivación del cromosoma X. El actor principal de este proceso es *XIST*. Éste lncRNA es encargado del reclutamiento de complejos remodeladores de cromatina para el silenciamiento del cromosoma X que transcribe primero este gen. Por el contrario, *TSIX* es el lncRNA involucrado en mantener activo al cromosoma X (Lee, Davidow y Warshawsky, 1999).

En la regulación de la transcripción, los lncRNAs activan o reprimen la expresión de sus genes blanco a nivel *cis* o *trans*. Al actuar en *cis*, silencian o potencian la expresión de genes adyacentes que se ubican en el mismo cromosoma. Por ejemplo, el gen del lncRNA *HOTTIP* está presente en el grupo de genes HOXA, y su expresión intensifica la expresión de otros genes componentes del mismo grupo. También pueden actuar en *trans*, silenciando o potenciando la expresión de genes en diferentes cromosomas. Durante la activación de la expresión de genes involucrados en el desarrollo de la retina, *Six3OS* es un ejemplo de lncRNA que actúa bajo este mecanismo (Wang *et al.*, 2011; Rapicavoli *et al.*, 2011).

2.7. El papel de los RNAs largos no codificantes en el cáncer

Aun cuando los lncRNAs son las especies de ncRNAs menos estudiadas, recientemente las investigaciones han demostrado que son reguladores maestros en múltiples vías de señalización celular tales como la vía TGF- β , la vía Wnt/ β -catenina, la vía fosfatidilinositol 3 cinasa (PI3K/Akt) y la vía del factor nuclear κ B (NF- κ B). Estas vías se encuentran involucradas en diversos procesos como la proliferación celular, la apoptosis y la inflamación (Abolghasemi *et al.*, 2020).

La desregulación de los lncRNAs en diferentes vías de señalización celular está relacionada con múltiples patologías humanas, entre las que destaca el cáncer (Tang *et al.*, 2013). De acuerdo con la base de datos Lnc2Cancer (<http://www.bio-bigdata.com/lnc2cancer/>), que proporciona información experimentalmente respaldada de la relación entre algunos lncRNAs y 165 variedades de cáncer humano, existen por lo menos 1614 lncRNAs humanos descritos en la literatura asociados al cáncer (Figura 10). Estas moléculas pueden desempeñar un papel

oncogénico al aumentar la proliferación celular, favorecer la evasión de la apoptosis, inducir resistencia a los fármacos y aumentar las tasas de migración y metástasis. No obstante, se ha demostrado que algunos lncRNAs poseen actividades supresoras de tumores (Abolghasemi *et al.*, 2020).

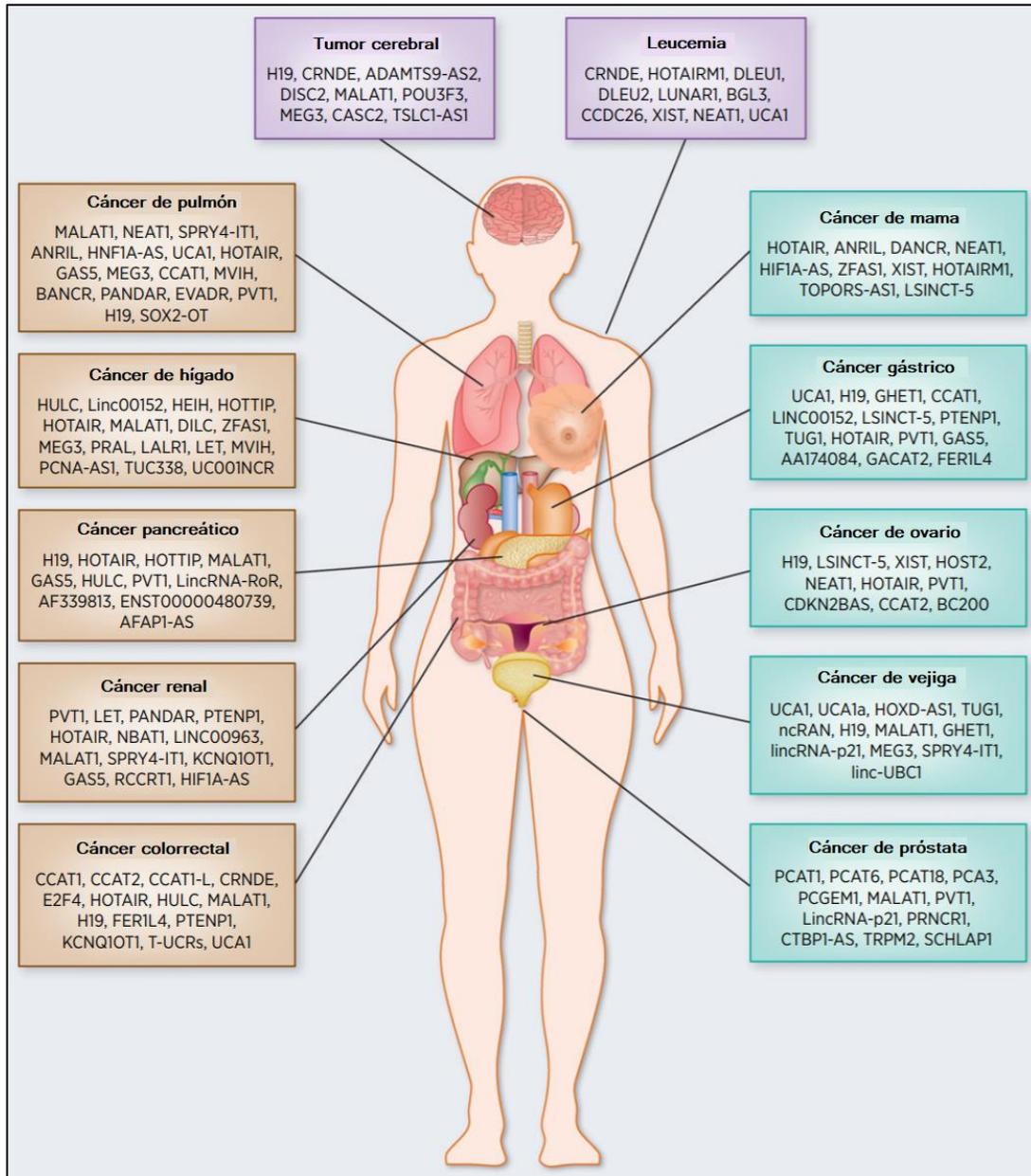


Figura 10. LncRNAs y su asociación con el cáncer. Se ha identificado que algunos lncRNAs se encuentran involucrados en alteraciones del ciclo celular, la evasión de la apoptosis y la metástasis, generando distintos tipos de cáncer en órganos como el cerebro, el pulmón, el hígado, el páncreas, el riñón, el colon, la próstata, las mamas y los ovarios, y en tejidos como la sangre. *Imagen modificada de Bhan, Soleimani y Mandal, 2017.*

Por ejemplo, se requiere del lncRNA *PVT1* para tener altos niveles del proto-oncogen c-Myc. La expresión de ambos está altamente correlacionada con la formación de tumores primarios, mientras que la supresión de *PVT1* inhibe la capacidad tumorigénica de c-Myc (Tseng *et al.*, 2014). *MALAT1* es otro lncRNA altamente estudiado, el cual se identificó por primera vez en cáncer de pulmón. Este transcrito es capaz de promover la formación de tumores a través de la vía Wnt/ β -catenina, la vía PI3K/AKT, la vía ERK / MAPK, y la angiogénesis (Ji *et al.*, 2003; Liu *et al.*, 2017). En cáncer de próstata, *PCAT-1* y *PCGEM1* actúan como reguladores de la proliferación celular. Particularmente, *PCGEM1* se caracteriza por inhibir el proceso de apoptosis inducido por fármacos empleados en el tratamiento del cáncer (Prensner *et al.*, 2011; Petrovics *et al.*, 2004; Fu *et al.*, 2006). *GAS5* es un lncRNA significativamente sub-expresado, identificado en una variedad de tumores. Tiene una correlación negativa con miR-21, un oncogén importante en algunos tumores malignos. Al unirse a miR-21, inhibe la proliferación celular y la invasión de los tejidos. Esta característica apunta a que podría ser un lncRNA supresor de tumores (Yang *et al.*, 2015; Zhang *et al.*, 2013; Chi *et al.*, 2019).

2.7.1. RNA largos no codificantes como biomarcadores del cáncer de mama

Según el Grupo de Trabajo de Biomarcadores del Instituto Nacional de Salud (NIH, por sus siglas en inglés), el término biomarcador se refiere a una característica que se puede medir objetivamente y evaluar como un indicador de un proceso biológico normal, estado patogénico o de respuesta a un tratamiento farmacológico (Biomarkers Definitions Working Group, 2001). Otra definición dada por la Organización Mundial de la Salud (OMS) indica que es casi cualquier medida que refleje una interacción entre un sistema biológico y un peligro potencial, el cual puede ser químico, físico o biológico. Esta respuesta medida puede ser funcional o fisiológica, y ocurre a nivel celular o molecular (World Health Organization, 1993).

Con base en sus aplicaciones, se han definido diferentes tipos de biomarcadores. Por ejemplo, un biomarcador de diagnóstico se utiliza cuando se desea detectar o confirmar la presencia de una enfermedad o condición de interés, así como identificar

a un individuo con un subtipo de la enfermedad. Por otro lado, se emplea un biomarcador de pronóstico para encontrar la probabilidad de un evento clínico, recurrencia o progresión en pacientes con una enfermedad o condición médica de interés (Califf, 2018).

La desregulación de los lncRNAs y su contribución significativa a la progresión del CaMa, sugiere que estas moléculas se pueden emplear como biomarcadores de la enfermedad (Abolghasemi *et al.*, 2020). *HOTAIR* es uno de los primeros lncRNAs reportados que está involucrado en el desarrollo del cáncer y la metástasis. Este lncRNA es un transcrito intergénico de 2,158 nt de longitud, compuesto por 6 exones. Se sintetiza a partir de la hebra antisentido del gen *HOXC* (12Q13.13), y se procesa post-transcripcionalmente por *splicing* y la adición de una cola de poli-A (Rinn *et al.*, 2007). En CaMa, es considerado un biomarcador de predicción de la progresión metastásica y la sobrevida global en etapas tempranas de la enfermedad. Su mecanismo de acción es a través de la represión de la expresión a nivel de la cromatina, actuando como andamio molecular para permitir las interacciones proteína-proteína del complejo remodelador de cromatina, el Complejo Represor Polycomb 2 (PRC2, por sus siglas en inglés), que cataliza la marca epigenética de represión H3K27me3. La evidencia señala que la sobre-expresión del transcrito en tumores primarios de mama es un potente predictor de metástasis y muerte para el pronóstico de las pacientes (Gupta *et al.*, 2010; Portoso *et al.*, 2017).

Recientemente se identificó que *MALAT1* está subexpresado en CaMa, lo cual se asocia con una disminución de la sobrevida global de las pacientes. Al unirse al factor de transcripción TEAD, *MALAT1* suprime la expresión de genes asociados a la metástasis y la migración celular. Para regular la transcripción génica, utiliza un mecanismo de unión a enzimas modificadoras de histonas y factores de transcripción. Además, a nivel post-transcripcional regula la expresión de mRNAs actuando como esponja de miRNAs. Estas características han respaldado la idea de que *MALAT1* es un gen supresor de tumores, y permite su postulación como biomarcador de pronóstico (Chen, Zhu y Jin, 2020).

2.8. El RNA largo no codificante *LINC01614*

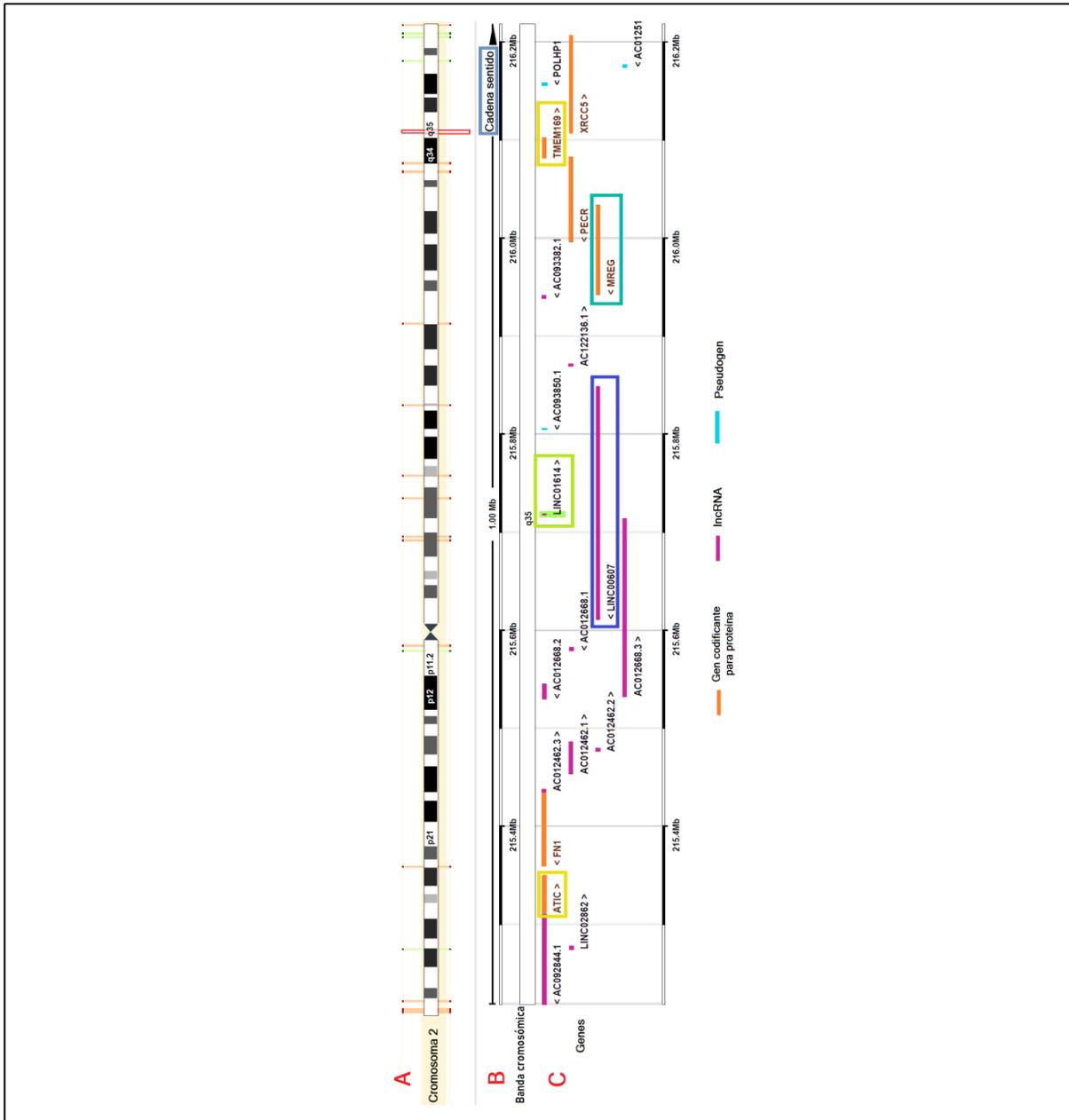


Figura 11. Mapa de localización genómica del *LINC01614*. A) En el recuadro rojo, se muestra que este lncRNA se localiza en el cromosoma 2, dentro de la banda 35 del brazo largo. B) En color gris, se indica que el lncRNA se sintetiza a partir de la cadena sentido. C) Dentro de la banda 35, se ubican diversos genes. En el recuadro verde, se señala la ubicación del *LINC01614*, el cual se caracteriza por ser un lncRNA intergénico. En los recuadros amarillos, se muestran los genes codificantes para proteínas adyacentes, *ATIC* y *TMEM169*, localizados en la cadena sentido. En el recuadro azul turquesa se marca el gen codificante para proteína, *MREG*, ubicado en la cadena antisentido. En el recuadro azul rey se señala que en la cadena complementaria se localiza el lncRNA *LINC00607*. Fuente: Ensembl Genome Browser.

Dependiendo de la base de datos o la literatura en la cual se reporte, el RNA largo no codificante *LINC01614* recibe otros nombres como: *LCAL4*, *TCONS_00003105*, *ENSG00000230838*, y lncRNA 4 asociado a cáncer de pulmón (del inglés *Lung cancer associated lncRNA 4*).

2.8.1. Características del *LINC01614*

La siguiente descripción del gen del *LINC01614* y la caracterización de la región en donde se localiza, se llevó a cabo empleando la plataforma *Gene* de la base de datos públicos *National Center for Biotechnology Information* (NCBI, <https://www.ncbi.nlm.nih.gov/gene>) y el visualizador del genoma *Ensembl Genome Browser* (<https://www.ensembl.org/index.html>) en su versión GRCh38.p13.

Su localización citogenética, 2q35, indica que se ubica en la banda 35 en el brazo largo del cromosoma 2 (Figura 11-A). Según el ensamblaje GRCh38.p13, las coordenadas del *LINC01614* son Chr2: 215,718,043-215,719,424, lo cual señala que su localización física abarca desde el nucleótido 215,718,043 hasta el nucleótido 215,719,424 del cromosoma 2 humano. En términos de tamaño, posee una longitud de 1382 nt. Se conforma por dos exones y un intrón, y se transcribe a partir de la cadena sentido (Figura 11-B). A nivel de transcrito, únicamente se ha descrito su procesamiento en la isoforma *LINC01614-201*. Ésta tiene una longitud de 648 nt y se caracteriza por contener ambos exones en su estructura.

El gen del lncRNA se ubica entre dos genes codificantes para proteína, sin intersectar con ellos. De acuerdo con la clasificación de los lncRNAs basada en la localización genómica (Derrien *et al.*, 2012), *LINC01614* se clasifica como un RNA largo no codificante intergénico (lincRNA). En la cadena sentido, los dos genes codificantes para proteínas adyacentes a *LINC01614* son el gen *ATIC* y el gen *TMEM169*. El primero codifica para la enzima 5-aminoimidazol-4-carboxamida ribonucleótido formiltransferasa/IMP ciclohidrolasa que participa en la de síntesis *de novo* de purinas y se localiza a 370 kb río arriba; el segundo se encuentra a 360 kb río abajo y codifica para la proteína transmembranal 169. En la cadena antisentido,

MREG es el gen codificante para proteína más cercano. Éste se ubica a 225 kb río abajo y da origen a la proteína melanoregulina (Figura 11-C).

La región del cromosoma en la que se localiza *LINC01614* destaca por la presencia de una mayor proporción de especies de lncRNAs en relación con la proporción de genes codificantes para proteínas. En la cadena complementaria al gen, se localiza el lncRNA *LINC00607* (Figura 11-C). De acuerdo con su estructura, *LINC01614* se empalma con uno de sus intrones.

2.8.2. Localización celular del *LINC01614*

Muchas especies de lncRNAs tienen localizaciones celulares muy bien definidas. De acuerdo con algunos autores, los lncRNAs se encuentran más enriquecidos en el núcleo en relación con el citoplasma (Derrien *et al.*, 2012; Quinn y Chang, 2016). Ésto respalda la teoría de que muchos de ellos están involucrados en la regulación epigenética de la cromatina (Wang *et al.*, 2017).

Para identificar su localización celular, se utilizó la base de datos *IncATLAS* (<http://Incatlas.org.eu>). Particularmente, en la línea celular HUVEC (células endoteliales de venas de cordón umbilical humano), se observa que *LINC01614* se expresa en ambos compartimientos, sin embargo, el valor RCI obtenido (-1.29) denota su tendencia a localizarse en el núcleo, en tanto que el color de la barra (marrón oscuro) indica que su nivel de expresión nuclear es cerca de 100 veces mayor al observado en citoplasma. En el caso de la línea celular H1.hESC (células troncales embrionarias), *LINC01614* se expresa únicamente en núcleo, de modo que no es posible calcular el valor RCI y se representa en líneas punteadas (Figura 12).

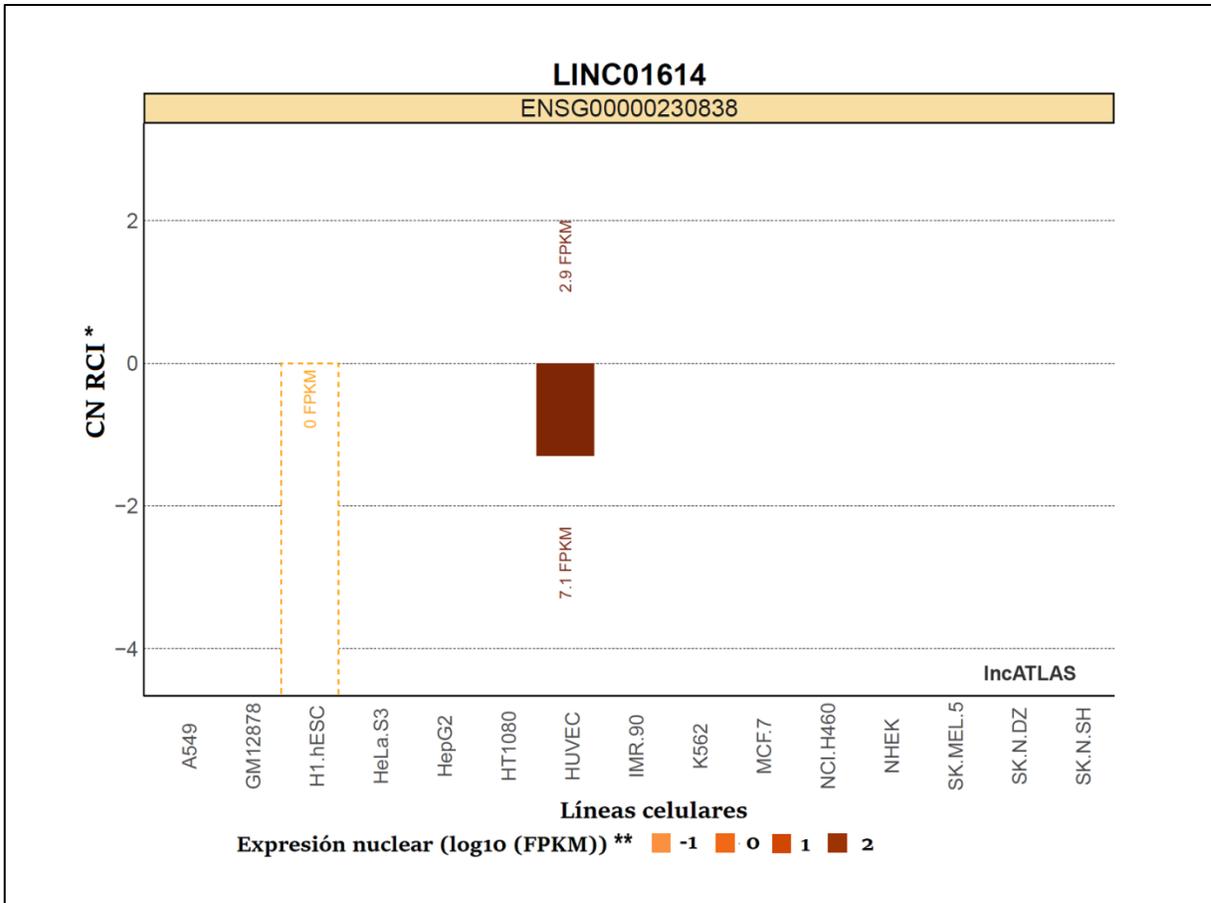


Figura 12. La localización celular del *LINC01614* es predominantemente nuclear. En línea celular HUVEC, *LINC01614* se localiza mayoritariamente en el núcleo, respecto al citoplasma. En la línea celular H1.hESC, la localización del transcrito es únicamente nuclear. *Fuente: IncATLAS Database.*

*El Índice de Concentración Relativa Citoplasma-Núcleo (CNRCI, por sus siglas en inglés), se calcula a partir del logaritmo base 2 del cociente de la expresión citoplasmática (en FPKM) y la expresión nuclear (en FPKM). El desplazamiento de la barra indica la tendencia del lncRNA a localizarse ya sea en el citoplasma o en el núcleo; un valor positivo indica que está localizado preferencialmente en citoplasma, mientras que un valor negativo señala que la localización es mayoritaria en el núcleo.

** El color y el valor de expresión en FPKM (fragments per kilobase per million mapped) muestran el nivel de expresión en el compartimiento.

3. ANTECEDENTES PARTICULARES

En oncología, la supervivencia global es un término que señala el porcentaje de individuos de un grupo que sobreviven después de un período de tiempo definido, tras ser diagnosticados con algún tipo de cáncer (Driscoll y Rixe, 2009). Para su cálculo, se realiza el análisis de supervivencia, un método estadístico que permite examinar los cambios a lo largo del tiempo de un evento específico (Dudley, Wickham y Coombs, 2016). La importancia de este estudio recae en que a través de él se puede comprender como los niveles de expresión de un transcrito, como el RNA largo no codificante *LINC01614*, influyen en la progresión del CaMa de acuerdo con las características de una paciente.

Dentro de sus múltiples líneas de investigación, el Laboratorio de Genómica del Cáncer del Instituto Nacional de Medicina Genómica (INMEGEN) participa en el proyecto titulado “Análisis de alta resolución de alteraciones somáticas en el genoma y transcriptoma de tumores de mama en pacientes mexicanas”. A continuación, se detalla cómo se llevó a cabo la selección del *LINC01614* como gen candidato asociado a la supervivencia global de las pacientes con CaMa, y los estudios posteriores realizados en torno al gen.

En un primer estudio, se realizó minería en la base de datos *The Cancer LncRNome Atlas* (TCLA, <http://fcgportal.org/TCLA/>). Con los datos de secuenciación obtenidos, se llevó a cabo un análisis de expresión diferencial, en el cual se detectó un grupo de lncRNAs que se encontraban desregulados en tumores de mama, así como en otros tipos de cáncer. Adicionalmente, los transcritos se mostraban diferencialmente expresados respecto al tejido mamario normal adyacente. El estudio se realizó considerando dos parámetros: 1) *FDR* (del inglés *False Discovery Rate*) y 2) tasa de cambio en la expresión (*Fold-change*).

Inicialmente, se identificaron 1144 y 1109 lncRNAs sobre-expresados y sub-expresados, respectivamente. La selección de lncRNAs se refinó, empleando la plataforma *The Atlas of Non-Coding RNAs in Cancer* (TANRIC, <https://www.tanric.org/>), hasta la obtención de 18 transcritos. De ellos, once se encontraban sobre-expresados, mientras que los siete restantes se mostraban sub-expresados. Lo anterior se determinó bajo el criterio de que estos lncRNAs estuvieran asociados con la supervivencia global del paciente y que cumplieran con los valores de Cox $p < 0.05$ y Log-Rank $p < 0.04$ en el análisis de supervivencia. Entre ellos, *LINC01614* se caracterizó por presentar una asociación significativa con la supervivencia global de las pacientes (Figura 13). Los detalles de este trabajo se encuentran descritos en la tesis “Análisis de la expresión de un RNA largo no codificante alterado en cáncer de mama, empleando bases de datos genómicos”, Padilla Palma (2018).

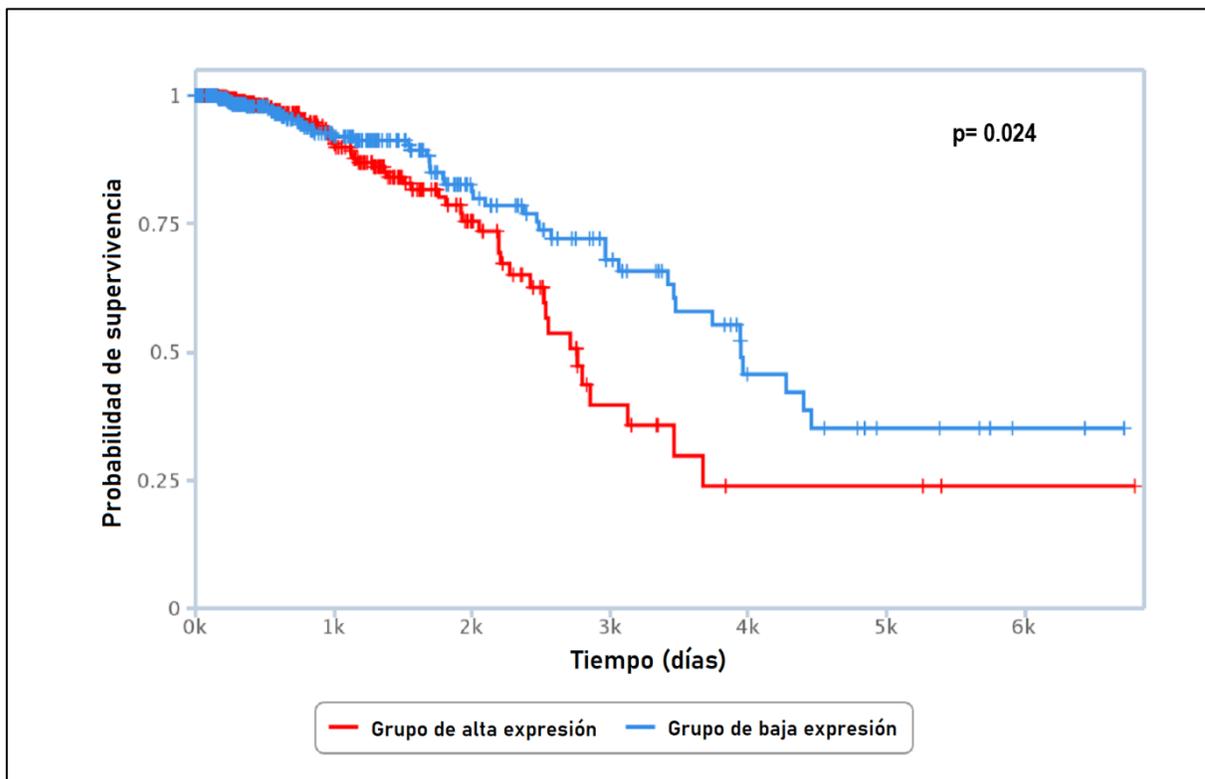


Figura 13. Los niveles de expresión del candidato *LINC01614* se asocian con la supervivencia global de las pacientes con CaMa. Curva de Kaplan-Meier obtenida a partir de una cohorte de bases de datos públicos. En las pacientes de esta cohorte, se identificó que la sobre-expresión de *LINC01614* se asocia significativamente con una menor probabilidad de supervivencia ($p = 0.024$). Imagen modificada de TANRIC.

Posteriormente, se realizó un segundo estudio, en el cual se analizaron 57 tumores de mama derivados de bloques de parafina que fueron colectados del Hospital Central Sur de Alta Especialidad (HCSAE) de Petróleos Mexicanos. De estos tumores, se seleccionaron 33 muestras que cumplían con los criterios de inclusión, exclusión y eliminación, y se evaluó la expresión del gen mediante RT-PCR en tiempo real. Los resultados de la validación en esta cohorte indicaron que existía un comportamiento semejante al obtenido en el trabajo bioinformático realizado anteriormente (Figura 14). Esta investigación se llevó a cabo por el Médico Especialista en Anatomía Patológica Alejandro Martínez Aguirre (2019), bajo el título de tesis “Análisis de expresión de RNA largo no codificante (*LINC-01614*) en carcinoma ductal infiltrante de la mama y su asociación al curso clínico”.

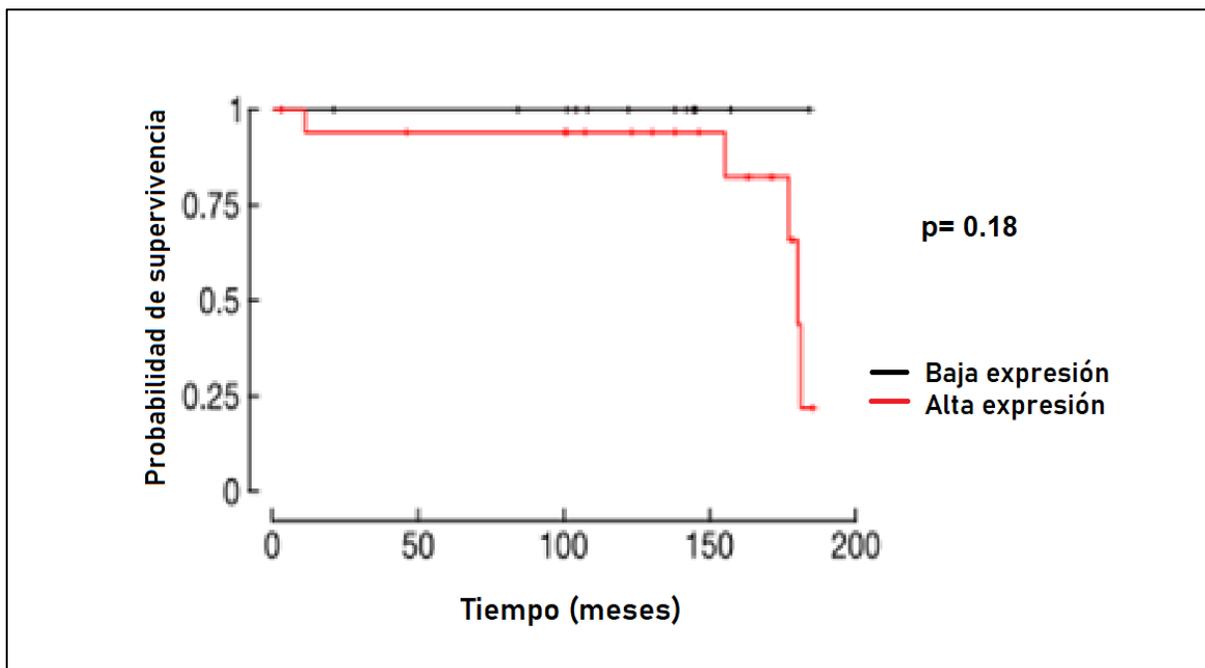


Figura 14. La alta expresión del *LINC01614* se relaciona con una menor probabilidad de supervivencia en pacientes con CaMa. Curva de Kaplan-Meier obtenida a partir de la cohorte del HCSAE de PEMEX. La probabilidad de supervivencia de las pacientes disminuye cuando el transcrito *LINC01614* se encuentra sobre-expresado en tumores de mama, aunque los datos no muestran una significancia estadística ($p= 0.18$). *Imagen modificada de Martínez-Aguirre, 2019.*

4. PLANTEAMIENTO DEL PROBLEMA

El CaMa es el tumor maligno más frecuente y mortal en mujeres de todo el mundo. México se posiciona en el segundo lugar de los países de América Latina y el Caribe en incidencia y mortalidad, lo que constituye un problema relevante para la salud pública. La heterogeneidad celular y molecular de la enfermedad impide ofrecer a las pacientes un diagnóstico y pronóstico satisfactorio. En el pasado, los estudios en torno a sus bases moleculares se han centrado en los genes codificantes, sin embargo, estos constituyen un porcentaje pequeño del genoma (1-2%). Lo anterior ofrece un argumento fuerte y sólido para buscar otros biomarcadores. Entre éstos, los RNAs no codificantes podrían estar participando en el desarrollo y la progresión de los tumores. Los lncRNAs son reguladores maestros de distintas vías de señalización celular debido a su capacidad para afectar la expresión de miles de genes. Con el desarrollo de las tecnologías genómicas, se ha revelado que algunos lncRNAs están desregulados en el cáncer. Esta evidencia ha abierto un nuevo panorama hacia la búsqueda de lncRNAs que juegan un papel importante en la carcinogénesis, y que puedan constituir potenciales herramientas de diagnóstico y pronóstico en distintas variedades de cáncer humano, incluido el CaMa.

5. JUSTIFICACIÓN

Este trabajo se centra en analizar la expresión del transcrito *LINC01614* en muestras de tejido tumoral de pacientes con CaMa, con la intención de validar si el RNA largo no codificante se encuentra desregulado en población mexicana, conforme a lo observado en las bases de datos públicas. Así mismo, permitirá determinar si existe una asociación significativa entre la expresión de *LINC01614* y la supervivencia global de las pacientes. Esto podría identificar una herramienta de pronóstico útil que beneficie a las pacientes y favorezca el avance hacia el empleo de genes no codificantes como biomarcadores de la enfermedad.

6. HIPÓTESIS

En población mexicana, los niveles de expresión del LINC01614 se asociarán con la supervivencia global de las pacientes, indicando su potencial utilidad como biomarcador de mal pronóstico clínico en pacientes Luminal B.

7. OBJETIVO GENERAL

Evaluar los niveles de expresión del RNA largo no codificante *LINC01614* en muestras de pacientes mexicanas con CaMa y correlacionar su tasa de expresión con la supervivencia global.

7.1. Objetivos particulares

1. Caracterizar *in silico* el potencial codificante y el plegamiento del *LINC01614*.
2. Analizar *in silico* la tasa de expresión del *LINC01614* en muestras de tumores de mama independientes.
3. Identificar genes diferencialmente expresados entre los niveles bajos y altos del *LINC01614*.
4. Identificar muestras de tumores de mama del biobanco FUCAM-INMEGEN que cumplan con los criterios de inclusión, exclusión y eliminación del estudio.
5. Analizar la tasa de expresión del *LINC01614* en muestras de tumores de mama mediante RT-PCR en tiempo real.
6. Correlacionar los niveles de expresión de *LINC01614* con la supervivencia global y las características clínico-patológicas de los tumores de mama analizados.

8. METODOLOGÍA

8.1. Caracterización *in silico*

8.1.1. Análisis del potencial codificante del *LINC01614*

La determinación del potencial codificante se llevó a cabo con la herramienta *Coding Potential Calculator 2* (CPC 2.0, <http://cpc2.cbi.pku.edu.cn/index.php>). Este análisis se basa en la búsqueda de marcos abiertos de lectura en las secuencias de los transcritos. El algoritmo del programa generó un valor denominado *Potencial Codificante*, el cual predice si un transcrito es codificante o no codificante. Adicionalmente, el software obtuvo la longitud del péptido hipotético generado, mediante la localización de un codón de inicio y un codón de paro. Para ejecutar el análisis, se seleccionaron dos genes de referencia: una secuencia codificante y una secuencia no codificante. Estos genes se emplearon para validar los resultados obtenidos, ya que cumplen con ser ampliamente caracterizados en la literatura.

8.1.2. Predicción de estructuras secundarias del *LINC01614*

Para obtener la predicción de las posibles estructuras secundarias, se utilizó el servidor *Mfold* del Instituto de RNA, Colegio de Artes y Ciencias (<http://unafold.rna.albany.edu/?q=mfold>).

8.2. Análisis bioinformático

8.2.1. Análisis de supervivencia global en muestras independientes

Se identificó la expresión de *LINC01614* en una cohorte independiente de tumores de mama empleando los datos de microarreglos procedentes de la base de datos *Gene Expression Omnibus* (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). Dicha cohorte se compone de 4 conjuntos de datos, con las claves de acceso: GSE16446, GSE20685, GSE20711 y GSE42568, analizados en el arreglo *Affymetrix Human Genome U133 Plus 2.0* (Thermo Fischer Scientific).

Empleando la plataforma Kaplan-Meier Plotter (*KM Plotter*, <http://kmplot.com/analysis/index.php?p=service&start=1>), se analizaron los valores de expresión en la cohorte, seleccionando el análisis de supervivencia global. El software generó la gráfica de Kaplan-Meier de supervivencia global, y realizó la prueba de hipótesis de Log-Rank.

8.2.2. Análisis de expresión diferencial

Se realizó un análisis de expresión diferencial en 109 muestras de CaMa Luminal B. Para ello, se utilizaron datos de secuenciación por RNA-Seq de la base de datos *The Cancer Genome Atlas* (TCGA, <https://portal.gdc.cancer.gov>). Las muestras se estratificaron mediante el cálculo del primer cuartil (percentil 25) y el tercer cuartil (percentil 75). Con ello se obtuvieron dos grupos: 1) muestras con sub-expresión de *LINC01614* y 2) muestras con sobre-expresión de *LINC01614*, de manera que cada grupo representa los niveles de expresión más bajos y más altos del transcrito, respectivamente.

Este análisis se llevó a cabo en la plataforma *GenePattern* (<http://software.broadinstitute.org/cancer/software/genepattern/>), con la herramienta *DESeq2*. El software genera una lista de transcritos diferencialmente expresados que incluye la siguiente información: el ID *Ensembl*, el número promedio de lecturas, la tasa de cambio en la expresión en logaritmo base 2 ($\log_2\text{FoldChange}$), el error estándar, el valor p y el valor p ajustado o FDR (del inglés *False Discovery Rate*).

8.2.2.1. Selección de datos del análisis de expresión diferencial

La selección de los transcritos diferencialmente expresados se basó en dos criterios: 1) Tasa de cambio en la expresión ($\log_2\text{FoldChange}$) con un valor >2.0 , en el caso de genes con alta expresión, y valores <-2.0 , en genes con baja expresión; y 2) Valor p ajustado o FDR estadísticamente significativo ($p < 0.05$).

8.3. Validación experimental

8.3.1. Selección de muestras del biobanco FUCAM-INMEGEN

El Laboratorio de Genómica del Cáncer del INMEGEN, en colaboración con la Fundación Mexicana de Fomento para la Prevención Oportuna del Cáncer de Mama A.C. (FUCAM), cuenta con un biobanco de muestras de tumores de mama, el cual comprende biopsias de tejido fresco (normal adyacente y tumor). Éstas fueron obtenidas de pacientes que acudieron al FUCAM, quienes previamente firmaron el consentimiento informado correspondiente al proyecto “Análisis de alta resolución de alteraciones somáticas en el genoma y transcriptoma de tumores de mama en pacientes mexicanas” aprobado por el Centro de Estudios de los Aspectos Éticos, Legales y Sociales del INMEGEN (Número de registro: CE2009/11). En el presente estudio, se preseleccionaron 170 muestras con diagnóstico de la enfermedad en el periodo comprendido entre los años 2008 y 2016. A partir de ellas, se seleccionaron 97 muestras que cumplen con los criterios de inclusión, exclusión y eliminación detallados a continuación.

8.3.1.1. Criterios de inclusión, exclusión y eliminación

Criterios de inclusión

- Pacientes de nacionalidad mexicana.
- Pacientes con diagnóstico confirmado de CaMa.
- Pacientes de todas las etapas clínicas (I a IV).
- Pacientes sin tratamiento previo que acuden por primera vez al FUCAM para iniciar tratamiento.
- Pacientes que proporcionen la firma del consentimiento informado.

Criterios de exclusión

- Pacientes que hayan tenido una neoplasia previa.
- Pacientes que tengan otra neoplasia primaria.

- Pacientes con infección sistémica aguda grave.
- Pacientes con embarazo o que se encuentren lactando.

Criterios de eliminación

- Pacientes con tratamiento neoadyuvante.
- Pacientes que no cumplen con un tiempo de seguimiento de 36 meses, a excepción de las pacientes que hayan fallecido a causa de CaMa.
- Material biológico colectado (biopsia) insuficiente para los estudios moleculares.

8.3.2. Extracción de RNA total de las muestras de tumores de mama

Se realizó la extracción de RNA total de las muestras de tumores de mama a partir de tejido fresco, empleando el kit de trabajo AllPrep® DNA/RNA Mini Kit (Qiagen® Inc, Valencia, CA) de acuerdo a las instrucciones del fabricante. La cantidad de muestra empleada se estandarizó a 20-30 mg de tumor, y previo a la extracción, las muestras se sometieron a un proceso de ruptura y homogeneización mecánica empleando un disruptor de tejidos y buffer de lisis. Éste último se compone principalmente de tiocionato de guanidino y β -mercaptoetanol, agentes inactivadores de las RNasas, esenciales para la obtención de RNA de alta calidad.

8.3.3. Cuantificación de RNA total

La concentración de RNA total en cada muestra se determinó mediante espectrofotometría, utilizando el instrumento NanoDrop™ One (ThermoFisher™ Scientific, Waltham, Massachusetts, USA). Así mismo, se evaluó la calidad del RNA aislado a través de los criterios de pureza dados por los ratios de absorbancia 260/280 y 260/230, útiles para detectar contaminación por proteínas, y por sales y compuestos orgánicos, respectivamente. Posteriormente, se almacenó a -80 °C hasta su procesamiento.

8.3.4. Síntesis de cDNA

Para la cuantificación relativa del RNA largo no codificante *LINC01614* mediante la técnica de PCR tiempo real, fue necesario obtener previamente cDNA a partir del RNA extraído de las muestras de tumores de mama. Para ello, se llevó a cabo la técnica de RT-PCR, de acuerdo a lo descrito en el protocolo del kit High Capacity cDNA Reverse Transcription (Applied Biosystems™, ThermoFisher™ Scientific, Waltham, Massachusetts, USA).

Previo a la reacción, se realizaron diluciones de las muestras hasta obtener una concentración final de 10.0 ng/μL (100 ng de RNA / 10 μL H₂O). La reacción se realizó en un volumen final de 20.0 μL, que incluye 10.0 μL de la mezcla de reacción preparada acorde a lo indicado en Tabla 3 y 10.0 μL de la dilución de cada muestra. Dicha reacción fue procesada en el equipo GeneAmp PCR System 9700 (Applied Biosystems™, ThermoFisher™ Scientific, Waltham, Massachusetts, USA), siguiendo las condiciones mostradas en la Tabla 4.

Tabla 3. Composición de la mezcla de reacción empleada para la síntesis de cDNA.

Componente de la reacción	Volumen por reacción (μL)
Solución amortiguadora RT	2.0
DNTPs	0.8
<i>Random primers</i>	2.0
<i>MultiScribe</i> ™ Transcriptasa Reversa	1.0
Agua libre de nucleasas	4.2
Volumen total por reacción	10.0

Tabla 4. Condiciones del termociclador para la síntesis de cDNA.

Condición	Fase 1	Fase 2	Fase 3	Fase 4
Temperatura (°C)	25	37	80	4
Tiempo (min)	10	120	5	∞

8.3.5. Cuantificación de la expresión del *LINC01614* mediante reacción en cadena de la polimerasa cuantitativa

Se evaluó la expresión de *LINC01614* con base en el protocolo TaqMan® Non-Coding RNA Assay (Applied Biosystems™, ThermoFisher™ Scientific, Waltham, Massachusetts, USA), usando el gen de expresión constitutiva *SCARNA5* (Hs03298717_s1, Thermo Fisher™ Scientific, Waltham, Massachusetts, USA) como gen de referencia en la reacción. El ensayo se realizó en un volumen final de 10.0 µL, que incluyen 9.0 µL de mezcla de reacción preparada de acuerdo a la Tabla 5, y 1.0 µL de cDNA sintetizado en el apartado 8.2.4., por triplicado. El ensayo se procesó en el equipo QuantStudio™ 5 Real-Time PCR System (Applied Biosystems™, ThermoFisher™ Scientific, Waltham, Massachusetts, USA) siguiendo las condiciones descritas en la Tabla 6.

Tabla 5. Composición de la mezcla de reacción empleada para la reacción de qPCR.

Componente de la reacción	Volumen por reacción (µL)
TaqMan® Universal Master Mix	5.0
Agua libre de nucleasas	3.5
Sonda TaqMan®	0.5
Volumen total por reacción	9.0

Tabla 6. Condiciones del termociclador para la reacción de qPCR.

Fase de la reacción		Temperatura (°C)	Tiempo	Ciclos
Hold	Incubación de UNG	50	2 minutos	1
	Activación enzimática	95	10 minutos	1
PCR	Desnaturalización	95	15 segundos	40
	Alineamiento/Extensión	60	1 minuto	

La cuantificación de la expresión del transcrito en las muestras se realizó a través del método de cuantificación comparativo $2^{-\Delta Ct}$, en donde:

$$\Delta Ct = Ct_{gen\ de\ interés} - Ct_{gen\ de\ referencia}$$

El cálculo de la cuantificación se limita a esta fórmula ya que no se cuenta con un tejido control para cada muestra.

8.3.6. Análisis estadístico

El análisis estadístico se realizó con el programa IBM SPSS Statistics® versión 25. Para la comparación de las medias de expresión entre más de dos grupos se utilizó la prueba de ANOVA y la prueba post hoc de Tukey. La prueba de t de Student para muestras independientes se usó cuando se comparó la expresión entre dos grupos. En variables categóricas, se utilizó la prueba Chi-cuadrado de Pearson o la prueba exacta de Fisher, según el caso, para buscar asociaciones entre el nivel de expresión y las características clínico-patológicas de las pacientes. Por último, para el análisis de supervivencia se realizaron las gráficas de Kaplan-Meier junto con la prueba de Log rank. Las muestras se estratificaron en las categorías de baja expresión y alta expresión de *LINC01614* con base en la mediana de expresión. En todos los casos se consideró como estadísticamente significativo el valor $p < 0.05$, el cual se determinó a una vía.

9. RESULTADOS

9.1. El *LINC01614* es un transcrito no codificante

Se determinó el potencial codificante de *LINC01614*, para confirmar su naturaleza no codificante. El análisis se realizó con dos genes de referencia: Alfa actina 1 (*ACTA1*) y RNA transcrito antisentido intergénico HOX (*HOTAIR*). Al ingresar la secuencia del gen *ACTA1*, el software indicó que éste codifica para un péptido con longitud de 378 aa, mientras que su valor de potencial codificante fue el máximo posible (100%). Este resultado indica que la herramienta es útil para distinguir a una secuencia codificante. Por el contrario, la secuencia del gen *HOTAIR* arroja un potencial codificante de 18.49% y un marco de lectura abierto de 48 aa. Este gen se predice como un transcrito no codificante y valida la capacidad del algoritmo para identificar una secuencia no codificante. Así mismo, los resultados de *HOTAIR* representan una baja probabilidad codificante para cualquier gen de prueba (Kang *et al.*, 2017).

En el caso de la secuencia de interés, se obtuvo un potencial codificante de 3.94%, valor ubicado por debajo del obtenido para *HOTAIR*. El marco de lectura abierto tiene una longitud de 59 aa, lo cual cumple con la observación de que las secuencias no codificantes no generan péptidos mayores a 200 aa (Kang *et al.*, 2017). De acuerdo con las características de su secuencia, los resultados sugieren que *LINC01614* es un transcrito no codificante, que presenta baja probabilidad de generar un péptido (Figura 15).

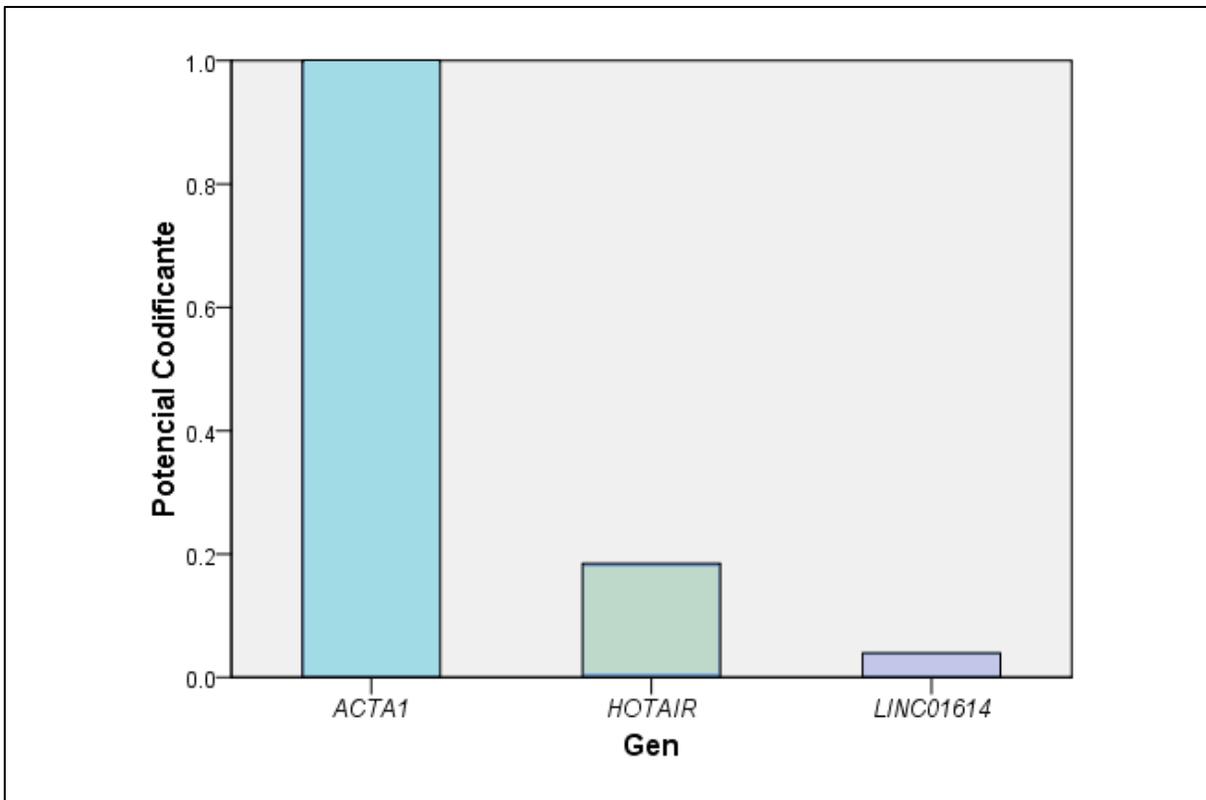


Figura 15. *LINC01614* es un transcrito no codificante. La gráfica de barras muestra los resultados obtenidos del análisis de potencial codificante realizado para tres genes diferentes. El transcrito *LINC01614* presenta un valor de potencial codificante de 0.0394 (3.94%), lo que demuestra que su secuencia es no codificante. El potencial codificante obtenido para *ACTA1* es 1.0 (100%), que indica la máxima probabilidad de generar un péptido. Para *HOTAIR*, el valor del potencial codificante fue 0.1849 (18.49%). Estos dos últimos genes se emplearon como secuencias de referencia para validar los resultados del software utilizado.

9.2. Estructura secundaria del *LINC01614*

En la literatura se ha descrito que los lncRNAs generan estructuras secundarias termodinámicamente estables, las cuales pueden determinar importantemente su función celular (Liu, Somarowthu y Pyle, 2017). Se empleó la herramienta *Mfold* (<http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form>) para predecir la estructura secundaria de *LINC01614*. A partir de la secuencia de 648 nt de la isoforma *LINC01614-201*, el algoritmo determinó 18 posibles estructuras para este RNA. En la figura 16 se presentan tres de ellas, que mostraron ser las más termodinámicamente estables (variación de energía libre de Gibbs, $\Delta G < 0$).

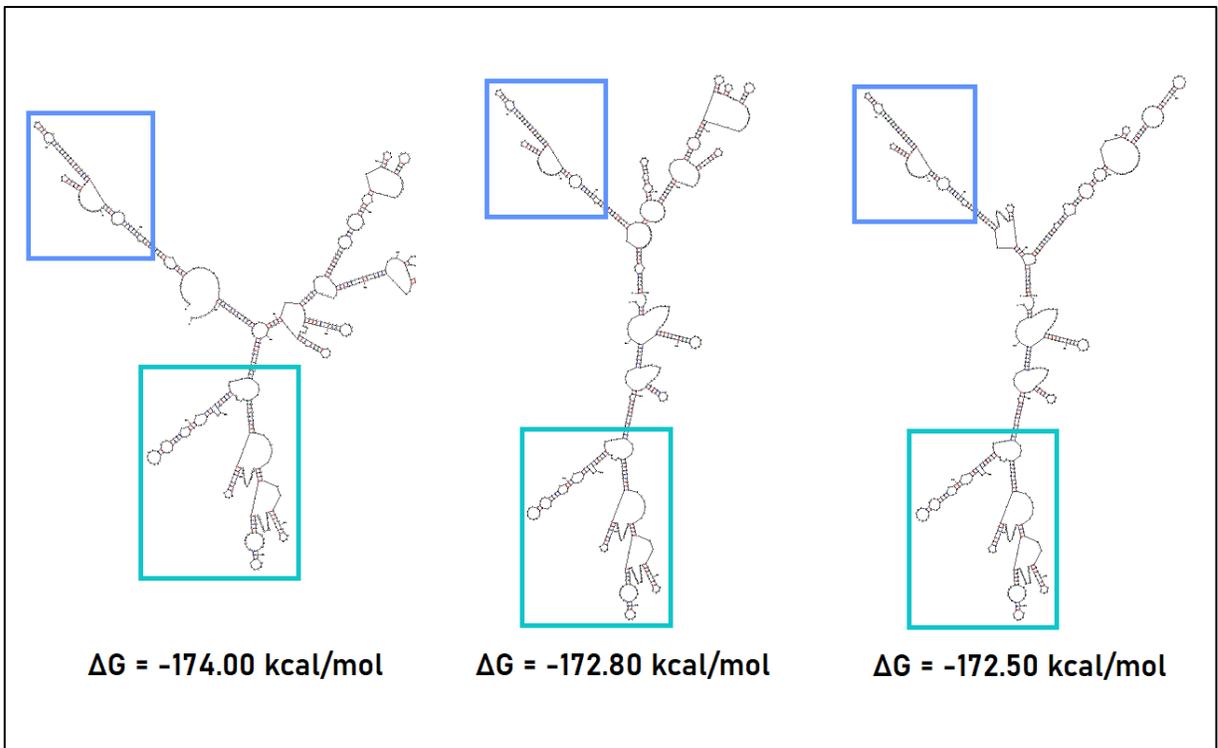


Figura 16. Estructuras secundarias del *LINC01614*. Con base en la secuencia del transcrito, se muestran tres estructuras secundarias probables para el plegamiento de la isoforma *LINC01614-201*. Los recuadros de color azul y verde señalan como para diferentes estructuras secundarias, existen dominios idénticos entre sí, que pueden estar asociados a una función específica importante para su mecanismo de acción.

Hasta este punto, los análisis realizados se enfocaron en aproximaciones computacionales que permitieron ampliar la caracterización del lncRNA *LINC01614*. Sin embargo, el objetivo central de esta investigación se centra en el análisis de los niveles expresión de *LINC0614* en tumores de mama de pacientes mexicanas y la asociación que existe con la supervivencia global. Para alcanzar este objetivo, un paso crucial es el estudio del lncRNA en una cohorte independiente de bases de datos públicos. La disponibilidad de un amplio número de muestras permite validar los resultados obtenidos en la cohorte mexicana.

9.3. La expresión del *LINC01614* se asocia con la supervivencia global en muestras independientes de cáncer de mama

Se identificó la expresión del transcrito *LINC01614* en una cohorte independiente de tumores de mama, empleando los datos de microarreglos procedentes de GEO DataSets. La cohorte comprende 626 muestras de tumores de mama, que incluyen los subtipos moleculares Luminal A, Luminal B, HER2 y Basal, de acuerdo a la proporción mostrada en la tabla 7.

Tabla 7. Clasificación por subtipo molecular de las muestras independientes del GEO. En paréntesis se indica el % de cada subtipo en la cohorte analizada.

SUBTIPO MOLECULAR				
Luminal A	Luminal B	HER2	Basal	TOTAL
271 (43.3%)	129 (20.6%)	73 (11.7%)	153 (24.4%)	626

El análisis de la supervivencia se realizó con la plataforma KM Plotter (<https://kmplot.com/analysis/>), mediante la cual se generaron las gráficas de Kaplan Meier y se llevó a cabo la prueba estadística no paramétrica correspondiente. La gráfica de Kaplan-Meier se compone de dos curvas, de las cuáles, una corresponde a las pacientes que presentan sobre-expresión de *LINC01614* (curva roja), en tanto que la segunda indica a las pacientes que muestran sub-expresión de *LINC01614* (curva negra).

En la primera gráfica de supervivencia global, que incluye todos los subtipos moleculares (figura 17-A), se observa que durante los primeros 50 meses del estudio, ambos grupos de pacientes comienzan a disminuir su probabilidad de supervivencia, sin embargo, esta disminución es más notoria para las pacientes con alta expresión. Al transcurrir 100 meses, la diferencia en las curvas es más evidente. Las pacientes con sub-expresión muestran una mayor probabilidad de supervivencia, con respecto a las pacientes con sobre-expresión. Al final del estudio, se observa que el 75% de las pacientes con sub-expresión sobrevive, mientras que únicamente el 58% de las que presentan sobre-expresión lo hace.

En general, el comportamiento observado sugiere que las pacientes con sobre-expresión de *LINC01614* tienen una menor probabilidad de supervivencia. Para corroborar lo anterior, el programa realiza la prueba de hipótesis de Log-Rank. El valor p obtenido indica que existe una asociación significativa entre la expresión del lncRNA y la supervivencia global ($p= 0.0264$). Esto demuestra que los niveles de expresión de *LINC01614* sí influyen en la supervivencia de las pacientes, de tal manera que la alta expresión se asocia con un pronóstico desfavorable en el CaMa.

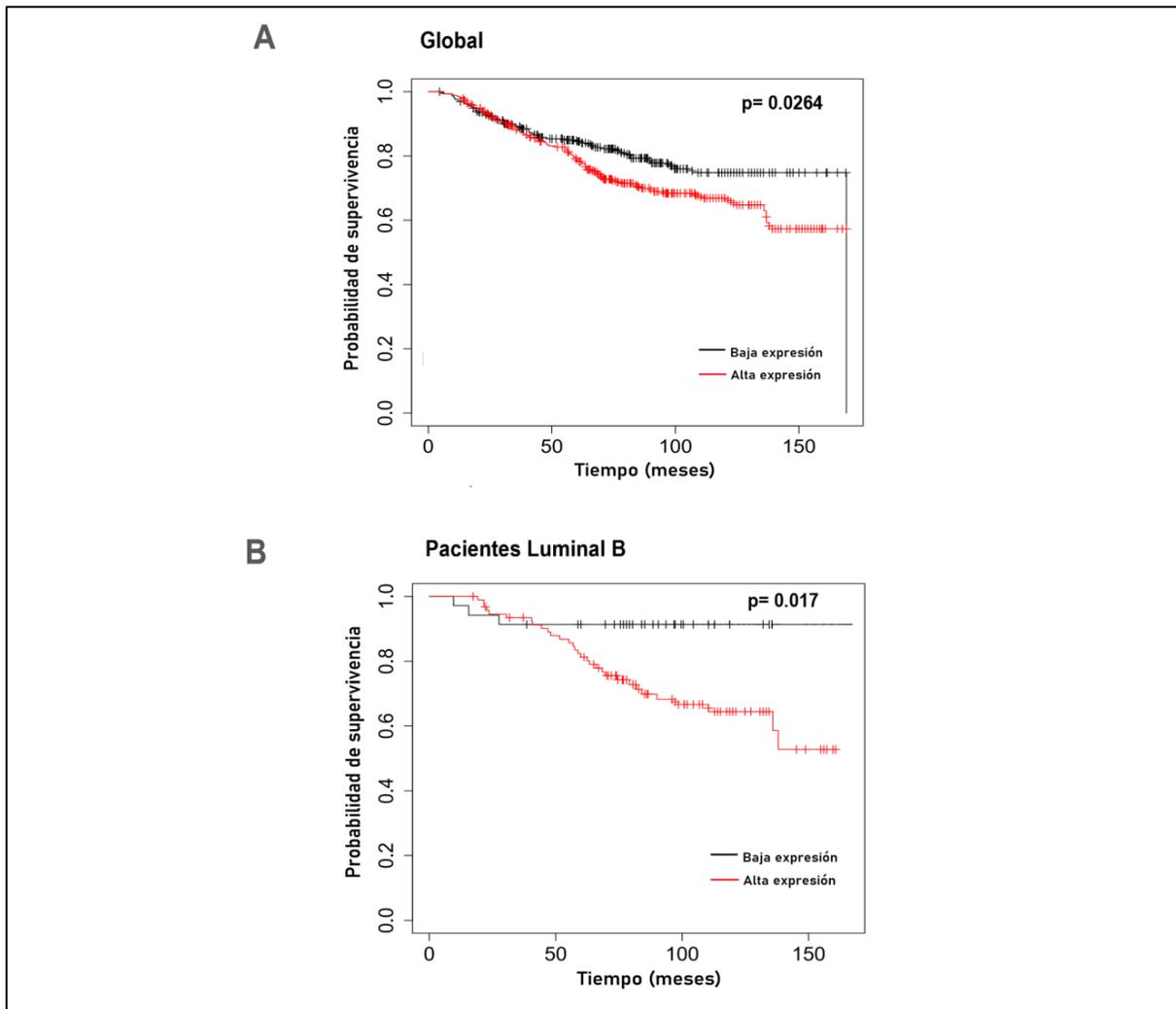


Figura 17. La sobre-expresión del *LINC01614* se asocia con un menor tiempo de supervivencia global, especialmente en pacientes con CaMa Luminal B. A) La gráfica de Kaplan-Meier muestra que en las pacientes con CaMa, la alta expresión de *LINC01614* se asocia con un pronóstico desfavorable ($p=0.0264$). B) La gráfica de Kaplan Meier indica que en pacientes con CaMa Luminal B, la probabilidad de supervivencia global es significativamente menor cuando sobre-expresan el transcrito *LINC01614* ($p=0.017$).

Posteriormente, se analizó individualmente cada uno de los cuatro subtipos moleculares que conforman la cohorte. Se obtuvieron las gráficas de supervivencia global de Kaplan Meier, así como el valor de p de la prueba de Log-Rank para cada subtipo (LumA, $p=0.061$; LumB, $p=0.017$; HER2, $p=0.20$; Basal, $p=0.13$). El subtipo Luminal B fue el único en presentar una asociación significativa entre el nivel de expresión de *LINC01614* y la supervivencia global. Estas pacientes presentaron un valor p menor ($p=0.017$) al que se obtuvo en el análisis general. La gráfica de estas pacientes (figura 17-B) muestra que el grupo de pacientes con sub-expresión de *LINC01614* permanece con una probabilidad de supervivencia mayor al 90% a lo largo del estudio, mientras que el grupo que sobre-expresa el transcrito tiende a ir disminuyendo su probabilidad de supervivencia a mayor número de meses transcurridos, hasta alcanzar una probabilidad de 52% cerca de los 150 meses.

9.4. El *LINC01614* puede regular la expresión de transcritos implicados en el desarrollo del cáncer

Con base en la asociación observada entre la alta expresión del *LINC01614* y la disminución de la supervivencia global (cohorte GEO), se decidió identificar si otros transcritos se encuentran desregulados en dichas condiciones. Para ello, se compararon los niveles de expresión de un amplio número de transcritos mediante el análisis de expresión diferencial. La comparación se llevó a cabo en muestras Luminal B, las cuáles se estratificaron en dos grupos: alta expresión del *LINC01614* y baja expresión del *LINC01614*. De acuerdo con los criterios establecidos para el análisis ($\log_2FC >2.0$ ó <-2.0 , y valor p FDR <0.05), se encontraron 1682 transcritos diferencialmente expresados. De éstos, la gran mayoría se encontraban regulados negativamente ($n=1387$, 82%) (Figura 18).

Los resultados de este primer acercamiento permiten identificar diferencias significativas en el transcriptoma de las pacientes con CaMa Luminal B, cuando se consideran los niveles de expresión del *LINC01614*. Además, sugieren la existencia de un conjunto de transcritos relacionados con la sobre-expresión del *LINC01614* y su asociación al proceso patológico del CaMa.

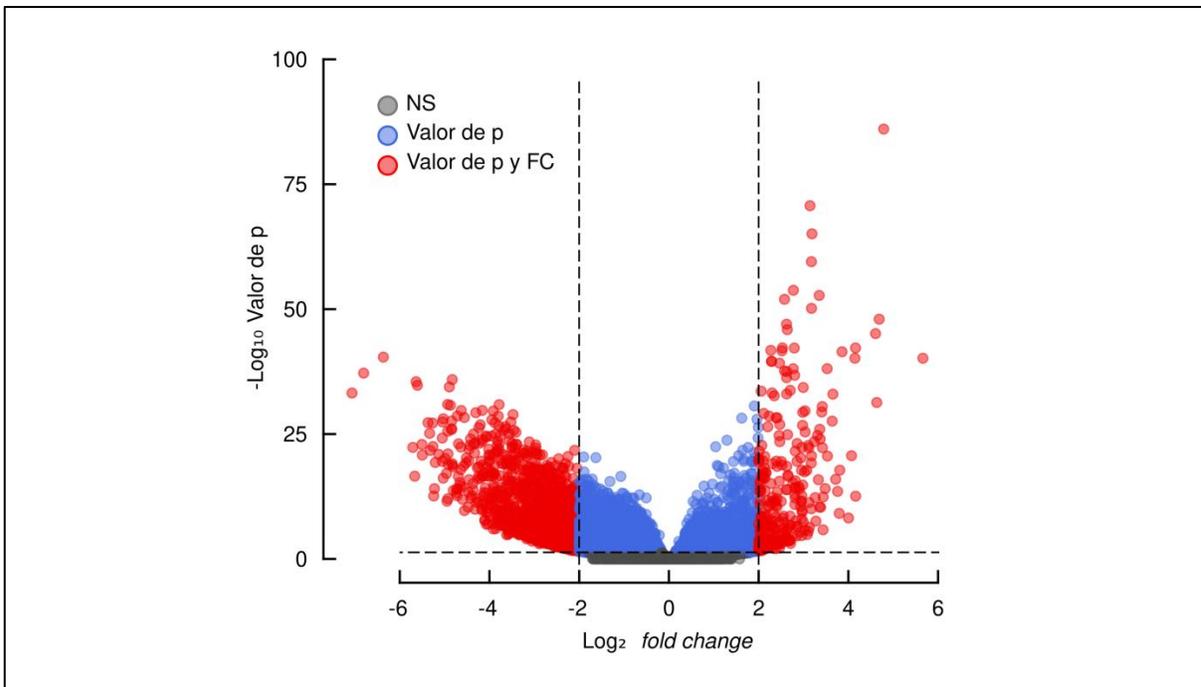


Figura 18. Los niveles de expresión del *LINC01614* se correlacionan con un grupo de transcritos diferencialmente expresados. El gráfico de volcán muestra que existen transcritos diferencialmente expresados entre las muestras Luminal B con niveles de expresión altos y bajos de *LINC01614*. Los puntos en gris indican a los transcritos sin significancia estadística; los puntos en azul señalan a los transcritos con valor p FDR significativo; los puntos en rojo muestran a los transcritos con tasa de cambio en la expresión (*Fold Change*) y valor p FDR significativo. En éstos últimos, en el cuadrante izquierdo se representan a los transcritos diferencialmente sub-expresados, mientras que en el cuadrante derecho se muestran los diferencialmente sobre-expresados.

Posteriormente, se seleccionaron los veinte principales transcritos con expresión diferencial, y se realizó una búsqueda en la literatura de sus principales funciones. Para este estudio, nos interesamos principalmente en aquellas funciones relacionadas con el desarrollo y la progresión tumoral (Tabla 8).

Dentro de los transcritos regulados positivamente, se identificó a *COL11A1*, *COL10A1*, *KLKP1*, *LINC01705*, *MMP13*, los cuales tienen papeles importantes en procesos de proliferación celular, migración celular, invasión, degradación de la matriz extracelular, angiogénesis y en la transición epitelial-mesenquimal. Por el contrario, los transcritos implicados en procesos de apoptosis, regulación de la proliferación, reconocimiento inmunológico y metabolismo de lípidos se mostraron regulados negativamente, por ejemplo, *CA4*, *LEP*, *CD300LG*, *PLIN4* y *SCARA5*.

Tabla 8. Principales transcritos con expresión diferencial en CaMa Luminal B con alta y baja expresión de *LINC01614*.

Símbolo	Descripción	Función	log2FC	Valor p FDR
COL11A1	Cadena alfa-1 de colágeno de tipo XI	Proliferación, migración e invasión celular	4.69	1.0E-48
COL10A1	Cadena alfa-1 de colágeno de tipo X	Angiogénesis; transición epitelial-mesenquimal	4.61	7.87E-46
LINC01929	RNA largo no codificante intergénico 1929	/	4.16	5.75E-43
KLKP1	Pseudogen de calicreína 1	Proliferación e invasión celular	3.86	3.44E-42
CD177	Antígeno de neutrófilos	Regulador de vía de señalización Wnt/ β -catenina	5.66	6.75E-41
LINC01705	RNA largo no codificante intergénico 1705	Proliferación y migración celular	4.15	6.75E-41
LINC01429	RNA largo no codificante intergénico 1429	/	3.53	8.28E-39
SLC24A2	<i>Acarreador de solutos familia 24, miembro 2</i>	Transporte transmembranaral	3.66	1.03E-33
MMP13	Metaloproteinasa de la matriz extracelular 13	Degradación de matriz extracelular, angiogénesis	4.63	4.93E-32
AP001434.1	RNA largo no codificante	/	3.64	2.64E-28
CA4	Anhidrasa carbónica 4	Apoptosis	-6.36	4.00E-41
LEP	Leptina	Proliferación, migración e invasión celular	-6.80	6.40E-38
CD300LG	Molécula de adhesión	Reconocimiento inmunológico	-5.63	3.34E-36
PLIN4	Perilipina-4	Regulación del metabolismo lipídico	-5.60	1.78E-35
MYOC	Miocilina	Estructura muscular	-7.06	6.22E-34
KCNJ16	Canal de potasio Subfamilia J Miembro 16	Proliferación celular y angiogénesis	-5.37	5.33E-28
STAC2	Proteína con dominio rico en Cisteína y SH3	/	-5.27	6.33E-28
ANGPTL7	Proteína similar a Angiopoyetina-7	Angiogénesis	-5.33	7.34E-26
SCARA5	Receptor <i>scavenger</i> Clase A Miembro 5	Apoptosis	-5.05	6.45E-25
ADGRD2	Receptor D2 acoplado a proteína G	Proliferación, migración e invasión celular	-5.50	1.22E-23

Validación experimental

Tabla 9. Características clínico-patológicas de la cohorte de pacientes FUCAM-INMEGEN.

Característica	Total n= 97	Luminal A n= 44 (45.4%)	Luminal B n= 37 (38.1%)	HER2 n= 7 (7.2%)	Triple negativo n= 9 (9.3%)	Valor p
Edad						
≤50 años	47	15	21	6	5	0.017*
>50 años	50	29	16	1	4	
Grado del tumor						
≤II	59	33	22	2	2	4.2E-5*
>II	29	4	14	5	6	
Desconocido	9	7	1	0	1	
Tamaño tumoral						
≤ 20 mm	30	19	10	1	0	0.018*
> 20 mm	67	25	27	6	9	
Etapa clínica						
≤II	87	43	32	6	6	0.011*
>II	10	1	5	1	3	
ER						
Positivo	77	41	34	2	0	3.4E-11 *
Negativo	20	3	3	5	9	
PR						
Positivo	63	41	19	3	0	1.3E-9 *
Negativo	34	3	18	4	9	
HER2						
Positivo	16	3	6	7	0	5.0E-7 *
Negativo	79	39	31	0	9	
Desconocido	2	2	0	0	0	
Ki-67						
≤20%	48	27	17	3	1	5.0E-4*
>20%	10	0	7	2	1	
Desconocido	39	17	13	2	7	
Recurrencia						
Si	7	1	3	1	2	0.051*
No	87	40	34	6	7	
Desconocido	3	3	0	0	0	
Metástasis						
Si	11	3	5	1	2	0.178*
No	86	41	32	6	7	
Evento de muerte						
Si	11	4	4	2	1	0.210*
No	86	40	33	5	8	

Ki-67= Antígeno Ki-67, marcador de proliferación celular. *= Prueba exacta de Fisher. La significancia estadística se determinó mediante un análisis a una vía. Los datos Desconocido no fueron incluidos.

En este trabajo, se incluyeron 97 muestras seleccionadas a partir del biobanco INMEGEN-FUCAM, las cuales fueron previamente clasificadas en cuatro subtipos con base en el diagnóstico por inmunohistoquímica. De acuerdo con esta clasificación, se observa que el subtipo Luminal A es el más frecuente en las pacientes (45.4%), seguido por el subtipo Luminal B (38.1%). En contraste, el 9.3% de los pacientes se clasifican en el subtipo basal, mientras que el subtipo HER2 es el menos frecuente, presentándose únicamente en el 7.2 % de las pacientes.

9.5. Características clínico-patológicas de la cohorte FUCAM-INMEGEN

Para estudiar la relevancia clínica del *LINC011614*, en primer lugar se evaluaron las características clínico-patológicas de la cohorte. Este análisis se realizó con el objetivo de estratificar a las pacientes y determinar su asociación con la supervivencia global. La primera de estas características, la edad, muestra que existe una proporción similar entre las pacientes con 50 años o menos (48%) y las pacientes mayores a 50 años (52%). En las primeras, el subtipo Luminal B es predominante en el 45% de las pacientes, mientras que en las segundas, el 58% de las pacientes se clasificó como Luminal A.

Respecto a las características del tumor, la mayoría de las muestras se encuentran en un grado tumoral menor o igual a II (61%), y se caracterizaron por tener un tamaño tumoral mayor a 20 mm (69%). Además, el 90% de las pacientes se localizan en una etapa clínica menor o igual a II, indicando que el diagnóstico se realizó en estadios tempranos de la enfermedad.

En relación a los receptores hormonales, el 79% de las pacientes es positiva para el receptor de estrógenos (ER), en tanto que el 65% de las pacientes es positiva para el receptor de progesterona (PR); la positividad de ambos receptores se observa predominantemente en las pacientes Luminales. En el caso de HER2, 16% de las pacientes son positivas para este receptor. De ellas, el 19% corresponde a las pacientes Luminal A, 37% a las pacientes Luminal B, y 44% a las pacientes HER2.

Por otra parte, la evaluación de la proliferación celular mediante el marcador Ki67 indicó que el 49% de los tumores presentan una tasa de proliferación menor o igual al 20%, mientras que un porcentaje menor (10%) presenta una tasa de proliferación mayor al 20%. De estos últimos, el 70% corresponde a las pacientes Luminal B. La recurrencia y la metástasis a otros órganos fueron eventos poco frecuentes en la cohorte, presentándose en un 7% y 11%, respectivamente. Sin embargo, ambos fenómenos afectaron al 22% de las pacientes triple negativo.

Por último, el evento de interés para esta investigación, el fallecimiento, ocurrió en el 11% de las pacientes. Dentro de las pacientes Luminal A, el 9% presentó el evento; en el caso de los subtipos Luminal B y triple negativo, ambos grupos presentaron el evento en el 11% de las pacientes; en tanto que en las pacientes HER2 se observó que en el 29% de ellas ocurrió el evento. Todo lo descrito en este apartado se presenta en la Tabla 9.

9.6. Los niveles de expresión del *LINC01614* varían entre las pacientes Luminal y no Luminal

Para determinar la expresión del *LINC01614* en la cohorte de pacientes FUCAM-INMEGEN, se realizó la cuantificación del transcrito en las muestras de tumores mediante reacción en cadena de la polimerasa cuantitativa (qPCR). Al comparar los datos de expresión entre los cuatro subtipos clasificados por inmunohistoquímica, se identificó mediante la prueba de ANOVA que no existe diferencia significativa entre la expresión del transcrito y los subtipos del CaMa ($p= 0.118$) (Figura 19-A).

Sin embargo, se observó que las muestras de los subtipos Luminal A y Luminal B presentan una expresión del transcrito semejante entre sí ($p= 0.499$), lo cual también se cumple en los subtipos HER2 y Triple negativo ($p= 0.490$) (Resultados no mostrados). Por otra parte, se demostró que la distribución de la expresión del transcrito en el subtipo Luminal A es ligeramente diferente a la del subtipo Triple negativo aunque no significativa ($p=0.134$), comportamiento que también se observa entre el subtipo Luminal B y el subtipo Triple negativo ($p= 0.186$) (Resultados no

mostrados). Este hallazgo, en conjunto con el hecho de que el número de casos de los subtipos HER2 y Triple negativo son bajos ($n= 7$ y $n= 9$, respectivamente), condujo a agrupar a las muestras en dos grupos. Con base en qué tan similares son entre sí estadísticamente, se generó una nueva clasificación que las divide en pacientes Luminal (Luminal A y Luminal B) y pacientes No Luminal (HER2 y Triple negativo). La prueba de t de Student evidenció que existe una diferencia estadísticamente significativa entre la expresión del *LINC01614* de ambos grupos ($p=0.022$) (Figura 19-B).

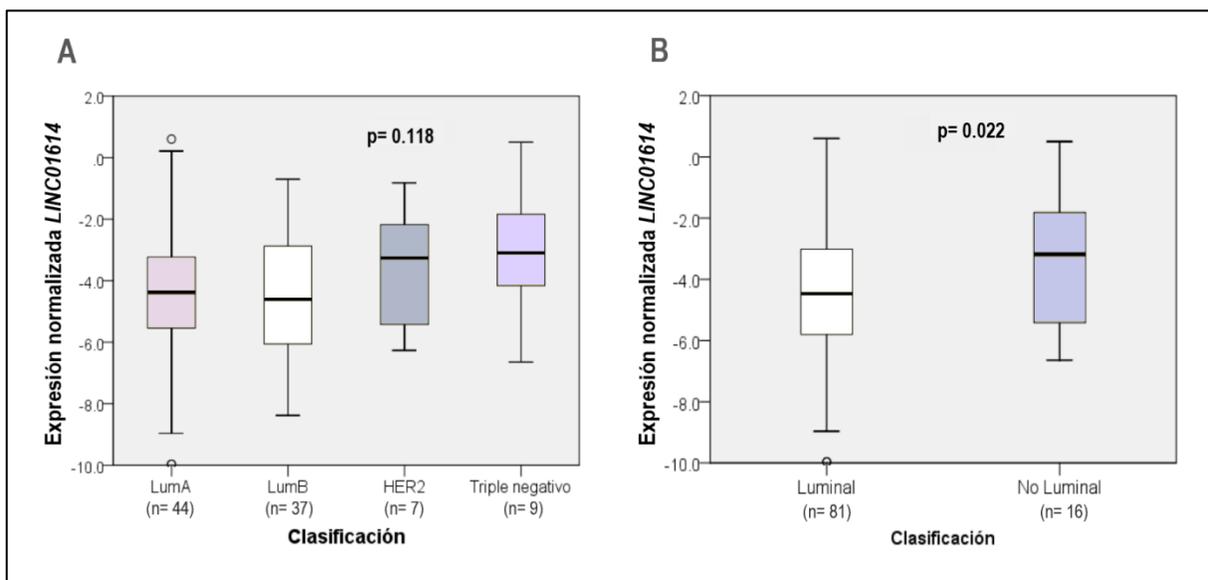


Figura 19. Niveles de expresión del *LINC01614* en la cohorte FUCAM-INMEGEN. A) Expresión del lncRNA *LINC01614* en las muestras de tumores de mama clasificadas por inmunohistoquímica ($p=0.118$). **B)** Nivel de expresión del lncRNA *LINC01614* significativamente diferente entre las pacientes de los grupos Luminal y No Luminal ($p= 0.022$). La significancia estadística se determinó mediante un análisis a una vía.

9.7. La disminución de la supervivencia global se asocia con la sobre-expresión del *LINC01614* en pacientes mexicanas con CaMa Luminal B

A modo de estudiar la relevancia clínica de *LINC01614* en pacientes mexicanas con CaMa, se estratificó la expresión del transcrito de cada muestra en las categorías de baja expresión y alta expresión tomando como referencia la mediana de los datos. Como se indica en la Tabla 10, los resultados de la prueba de Chi-

cuadrado o prueba exacta de Fisher (según cada caso) revelaron que existe una asociación clínica entre el nivel de expresión del *LINC01614* y los subtipos del CaMa clasificados por inmunohistoquímica ($p= 0.035$), así como con los grupos Luminal o No Luminal ($p= 0.016$). Esta asociación también se observó en variables como la edad ($p= 0.042$), el grado tumoral ($p= 0.002$), y la recurrencia ($p=0.046$).

Tabla 10. Asociación clínica entre la expresión del *LINC01614* y las características clínico-patológicas en la cohorte FUCAM-INMEGEN.

Variable	Expresión de <i>LINC01614</i>		Valor p
	Baja	Alta	
Subtipo por IHQ			
Luminal A	22	22	0.035*
Luminal B	22	15	
HER2	3	4	
Triple negativo	1	8	
Grupo			
Luminal	44	37	0.016~
No Luminal	4	12	
Edad			
≤50 años	19	28	0.042~
>50 años	29	21	
Grado del tumor			
≤II	32	27	0.002~
>II	9	20	
Tamaño tumoral			
≤ 20 mm	16	14	0.306~
> 20 mm	32	35	
Etapas clínicas			
≤II	42	45	0.262*
>II	6	4	
ER			
Positivo	41	36	0.073~
Negativo	7	13	
PR			
Positivo	32	31	0.219~
Negativo	16	18	
HER2			
Positivo	6	10	0.147~
Negativo	41	38	
Ki-67			
≤ 20%	26	22	0.250*
> 20%	4	6	
Recurrencia			
Si	2	5	0.046*
No	43	44	
Metástasis			
Si	6	5	0.388~
No	42	44	

Ki-67= Antígeno Ki-67, marcador de proliferación celular. ~ = Prueba de Chi-cuadrado. * = Prueba exacta de Fisher. La significancia estadística se determinó mediante un análisis a una vía.

Después, se realizó el análisis de supervivencia y la prueba estadística de Log-rank para explorar el efecto de la expresión del transcrito en la supervivencia global. Inicialmente, se realizó un análisis general que incluía a todas las pacientes de la cohorte, en el cual se observa una tendencia a que las pacientes con sobre-expresión del transcrito presenten una menor supervivencia global ($p=0.063$) (Figura 20-A). Específicamente, aquellas con sub-expresión del transcrito presentan un 94% de probabilidad de supervivencia global, mientras que en las pacientes con sobre-expresión esta probabilidad disminuye al 84%. Ésto se relaciona con un tiempo de supervivencia promedio de 114 meses y 108 meses, respectivamente. Lo anterior indica que el pronóstico de vida para las pacientes con alta expresión de *LINC01614* es 6 meses mayor en relación a las que presentan baja expresión.

Al estratificar a las pacientes en los grupos Luminal y no Luminal, se observó que los altos niveles de expresión de *LINC01614* se asocian con un tiempo de supervivencia significativamente más corto en pacientes con fenotipo Luminal ($p=0.039$) (Figura 20-B). En este caso, la sub-expresión del transcrito se asocia con un pronóstico de vida de 116 meses, mientras que en el escenario de la sobre-expresión, este pronóstico disminuye a 108 meses.

Por último, uno de los cuestionamientos de mayor interés para esta investigación fue identificar si en alguno de los cuatros subtipos de la clasificación por IHQ los niveles de expresión del transcrito influyen en la supervivencia global. Los resultados mostraron que las pacientes Luminal B tienen una probabilidad de supervivencia global significativamente más baja cuando sobre-expresan a *LINC01614* ($p= 0.045$) (Figura 20-C). Dicha asociación no fue observada en ninguno de los otros subtipos (LumA, $p=0.186$; HER2, $p= 0.124$; Basal, $p= 0.353$, Resultado no mostrados). En el grupo de pacientes Luminal B, la baja expresión del *LINC01614* se asocia con un 96% de probabilidad de supervivencia, valor que disminuye a 80% cuando el transcrito se expresa en niveles altos. Por lo tanto, la sobre-expresión del lncRNA *LINC01614* es un marcador de mal pronóstico clínico en pacientes con CaMa Luminal B.

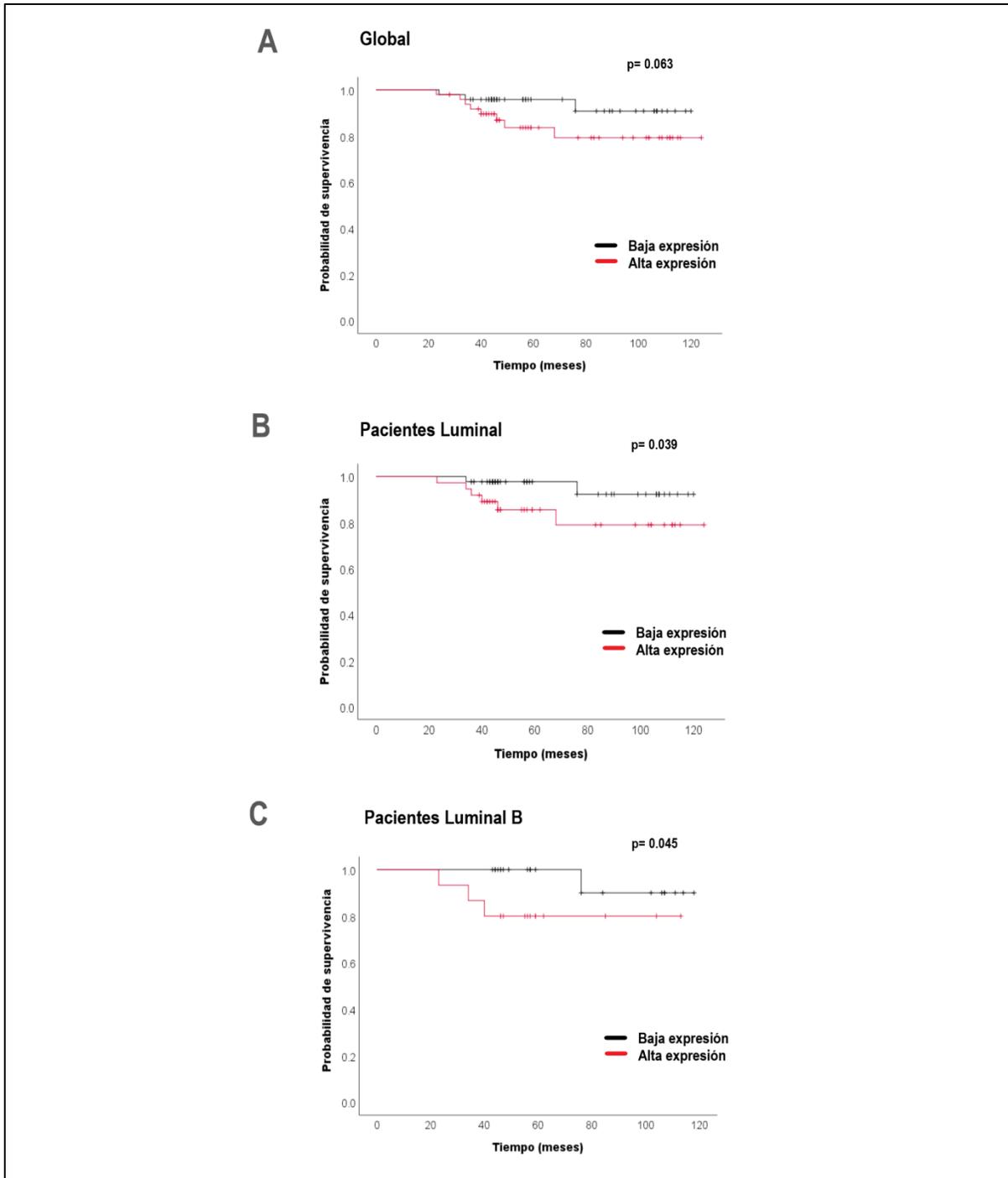


Figura 20. La sobre-expresión del *LINC01614* es desfavorable para la supervivencia global de las pacientes mexicanas con CaMa, especialmente del subtipo Luminal B. **A)** La gráfica de Kaplan-Meier indica que las pacientes con sobre-expresión de *LINC01614* tienden a presentar una menor probabilidad de supervivencia global ($p= 0.063$). **B)** En las pacientes del grupo Luminal, la sobre-expresión de *LINC01614* se asocia significativamente con una menor probabilidad de supervivencia global ($p= 0.039$). **C)** Las pacientes del subtipo Luminal B presentan un tiempo de supervivencia significativamente más corto cuando sobre-expresan el transcrito *LINC01614* ($p=0.045$). La significancia estadística se determinó mediante un análisis a una vía.

9.8. La supervivencia global del CaMa tiende a disminuir en pacientes jóvenes con sobre-expresión del *LINC01614*

Posteriormente, el análisis se enfocó hacia la búsqueda de características clínico-patológicas significativamente relacionadas con la supervivencia global. En este sentido, se observó que las pacientes de 50 años o menores presentan un ligero incremento en la expresión del *LINC01614* respecto a las pacientes mayores de 50 años. Dicho hallazgo fue confirmado mediante la prueba de t de Student, la cual reveló que los niveles de expresión entre ambos grupos son significativamente diferentes ($p= 0.0025$) (Figura 21-A). Al realizar el análisis de supervivencia de Kaplan Meier, se observó que las pacientes con menor probabilidad de supervivencia son aquellas que presentan alta expresión del *LINC01614*. Aunque ésta no es significativa, la tendencia señala que los niveles de expresión son un factor de interés para el pronóstico únicamente cuando las pacientes tienen 50 años o menos ($p=0.062$) (Figura 21-B). En este grupo, las pacientes con sub-expresión muestran una probabilidad de supervivencia del 96%, mientras que la probabilidad disminuye a 80% en las pacientes con el patrón de sobre-expresión.

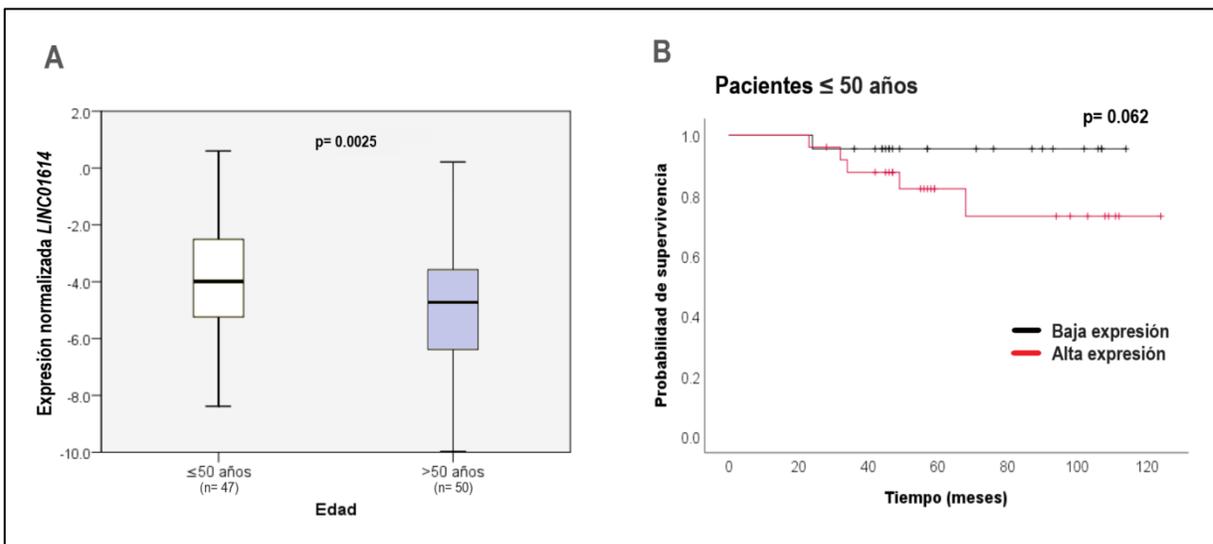


Figura 21. La sobre-expresión del *LINC01614* tiende a ser un factor de mal pronóstico cuando las pacientes con CaMa tienen 50 años o menos. A) El nivel de expresión del *LINC01614* es significativamente diferente entre las pacientes de 50 años o menos y las pacientes mayores de 50 años ($p= 0.0025$). **B)** La gráfica de Kaplan-Meier indica que las pacientes con una edad de 50 años o menos, tienden a presentar una menor probabilidad de supervivencia global cuando sobre-expresan el lncRNA *LINC01614* ($p= 0.062$). La significancia estadística se determinó mediante un análisis a una vía.

9.9. La supervivencia global del CaMa disminuye en pacientes con tumores de bajo grado y sobre-expresión del *LINC01614*

Por otra parte, se detectó una asociación estadísticamente significativa entre la expresión de *LINC01614* y la probabilidad de supervivencia global, en las pacientes con tumores de bajo grado (\leq II) ($p= 0.0045$) (Figura 22-A). En estas pacientes, la probabilidad de supervivencia es la más alta (100%) cuando los tumores presentan sub-expresión del transcrito, sin embargo, en el panorama contrario, la sobre-expresión se asocia a una probabilidad de supervivencia del 80%. Adicionalmente, se observó que, en el grupo de pacientes que no presentaron evento de recaída, el tiempo de supervivencia media fue significativamente más corto cuando el *LINC01416* se encuentra sobre-expresado. (112 meses), respecto a cuándo su expresión es baja (118 meses) ($p=0.047$) (Figura 22-B).

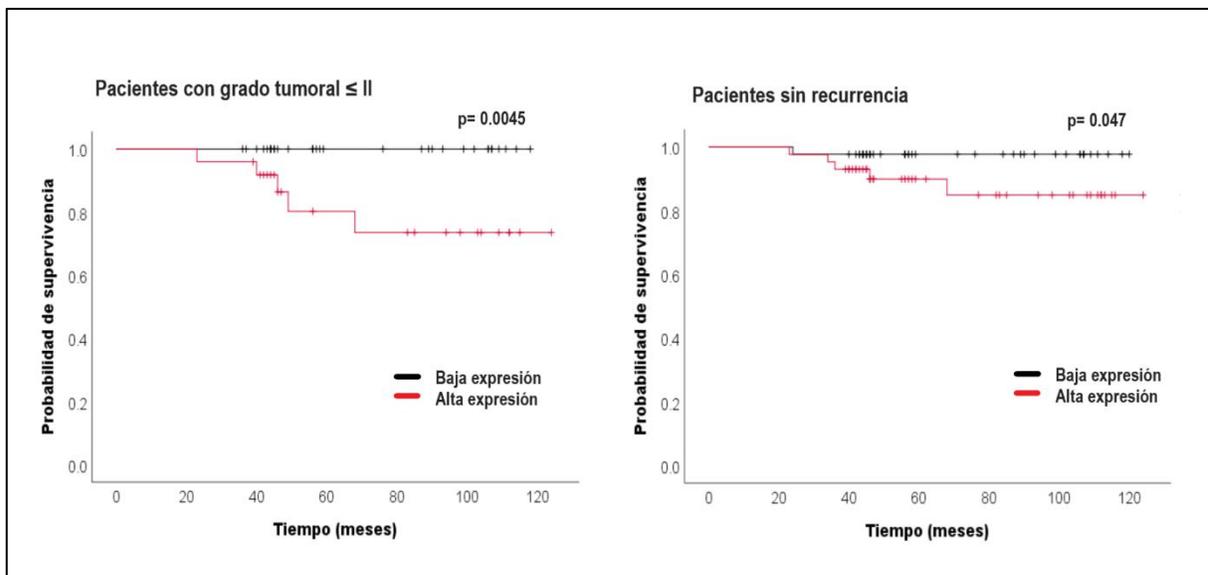


Figura 22. La sobre-expresión del *LINC01614* se asocia con un menor tiempo de supervivencia en pacientes con mejor desarrollo de la enfermedad. A) La gráfica de Kaplan-Meier indica que las pacientes de bajo grado tumoral tienen una probabilidad de supervivencia global significativamente menor cuando sobre-expresan el transcrito *LINC01614* ($p= 0.0045$). **B)** La gráfica de Kaplan-Meier señala que la supervivencia global se asocia significativamente con los niveles de expresión de *LINC01614* en pacientes sin recurrencia ($p= 0.047$). La significancia estadística se determinó mediante un análisis a una vía.

10. DISCUSIÓN

El desarrollo de las tecnologías de secuenciación de nueva generación y su aplicación al estudio del transcriptoma de células tumorales ha facilitado la identificación de RNAs no codificantes de relevancia biológica en el cáncer (Ratti *et al.*, 2020). En la literatura existe amplia evidencia que revela que los RNAs largos no codificantes son reguladores relevantes para promover o inhibir el desarrollo de los tumores mediante diversos mecanismos de acción (Chen *et al.*, 2016; Bin *et al.*, 2018). Actualmente, estas moléculas son el centro de atención de grupos de investigación, los cuales dirigen sus esfuerzos a explorar sus capacidades de regulación en la transformación y progresión neoplásica, así como en procesos celulares de proliferación, migración y estabilidad del DNA (Rathinasamy y Velmurugan, 2018).

Actualmente, la mayoría de los biomarcadores de cáncer se basan en genes codificantes para proteínas. Sin embargo, diversos estudios muestran que numerosos lncRNAs que se encuentran desregulados en el cáncer presentan perfiles de expresión característicos de acuerdo con la ubicación, el grado tumoral y la progresión de la enfermedad (Chandra Gupta y Nandan Tripathi, 2016; Sánchez y Huarte, 2013). El análisis de los datos de expresión, disponibles públicamente, se puede integrar con la información clínico-patológica de pacientes con la enfermedad a modo de demostrar su potencial como marcadores biológicos (Du *et al.*, 2013).

Algunos lncRNAs ya son utilizados en combinación con biomarcadores de genes codificantes, para mejorar la especificidad y la sensibilidad del diagnóstico de algunas enfermedades. Por ejemplo, el lncRNA *PCA3*, el cual se detecta en orina para el diagnóstico de cáncer de próstata, aumenta la precisión y la sensibilidad de la prueba respecto a la cuantificación de la proteína PSA sola (Wang *et al.*, 2018). En general, los lncRNAs han demostrado que presentan alta especificidad, alta estabilidad, y fácil detección en tejidos y en fluidos corporales como suero, plasma, orina y saliva, lo que ha aumentado el interés en su estudio como biomarcadores de cáncer (Chandra Gupta y Nandan Tripathi, 2016).

En los estudios de expresión diferencial realizados en el laboratorio de Genómica del Cáncer, se identificó que el *LINC01614* es un transcrito cuya expresión se encuentra desregulada en diferentes tumores. En la literatura se ha descrito esta desregulación en por lo menos tres tipos tumorales distintos: cáncer de pulmón, cáncer de mama y gliomas de alto grado tumoral. En un reporte se identificó que el *LINC01614* se encuentra desregulado en muestras de cáncer de pulmón de células no pequeñas, principalmente en las etapas tumorales más avanzadas. Además, los autores describen que la disminución de la supervivencia de las pacientes se asocia con la alta expresión del *LINC01614*. En conjunto, estos hallazgos condujeron a postularlo como un potencial biomarcador de diagnóstico en este tipo de cáncer (Sun y Ling, 2019). Por otra parte, dos trabajos de investigación independientes en cáncer de mama encontraron que la sobre-expresión del *LINC01614* se asocia significativamente con una menor supervivencia global (Vishnubalaji *et al.*, 2019; Wang *et al.*, 2019). Finalmente, en una investigación más reciente realizada con datos bioinformáticos de muestras de glioma, se observó que el *LINC01614* es una potencial herramienta de pronóstico, al hallar que la alta expresión favorece una menor supervivencia global de los pacientes (Wang, Wu y Guo, 2020).

Respecto a la función del *LINC01614*, estudios recientes han descrito algunas de sus funciones biológicas en el desarrollo del cáncer. Por ejemplo, en gliomas se identificó que el factor de transcripción SP1 regula positivamente al *LINC01614*, lo cual favorece que se una competitivamente al supresor tumoral miR-383, e inhibe la regulación de la expresión del transcrito *ADAM12* (Wang, Wu y Guo, 2020). Este RNA codifica para una metaloproteinasa altamente asociada con la progresión del cáncer (Kveiborg *et al.*, 2008). De este modo, la sobre-expresión del *LINC01614* aumenta los niveles de *ADAM12* mediante el secuestro de miR-383. Así mismo, se demostró que el *LINC01614* promueve el crecimiento tumoral y la metástasis, al inhibir la expresión de la caspasa 9 y la caspasa 3. Ésto evita que las células de glioma entren en proceso de apoptosis (Wang, Wu y Guo, 2020). De manera interesante, en adenocarcinoma de pulmón se reportó que el *LINC01614* actúa bajo este mismo mecanismo de competidor endógeno observado en gliomas. El

LINC01614 secuestra a miR-217, de forma que afecta su capacidad para unirse a FOXP1 (su transcrito blanco). Éste último promueve la proliferación celular y conduce al desarrollo y la progresión del cáncer de pulmón (Liu *et al.*, 2018).

Todo lo anterior sugiere que el mecanismo de acción por el cual la sobre-expresión del *LINC01614* se asocia a la progresión del cáncer en diferentes tipos tumorales es a través de su papel como esponja de miRNAs (Wang, Wu y Guo, 2020; Liu *et al.*, 2018). Este tipo de regulación ocurre en el citoplasma a nivel post-transcripcional y evita que los miRNAs degraden a sus transcritos blanco, lo cual es posible cuando el competidor endógeno presenta sitios de unión semejantes al mRNA (Militello *et al.*, 2016; Sen *et al.*, 2014). Esto concuerda con el trabajo realizado en células de glioma (Wang, Wu y Guo, 2020), en el cual se mostró que el *LINC01614* se expresa principalmente en el citoplasma. De acuerdo con *IncATLAS Database*, el transcrito se expresa mayoritariamente en el núcleo en la línea celular HUVEC, mientras que en la línea celular h1.hESC su expresión es exclusivamente nuclear. Los datos sugieren que el *LINC01614* presenta otros mecanismos de acción dentro del núcleo, mismos que aún no se han dilucidado. Además, la evidencia indica que la expresión citoplasma/núcleo es tejido o contexto celular dependiente, como se ha reportado anteriormente (Han *et al.*, 2018; Bala y Johnson, 2018). En tumores de mama, se reporta que durante la sobre-expresión del *LINC01614* se activa predominantemente la vía del factor de crecimiento transformante β (TGF- β) y la señalización por cinasas de adhesión focal (FAK) (Vishnubalaji *et al.*, 2019). No obstante, los detalles de su mecanismo de acción molecular o sus moléculas blanco aún no se han dilucidado en este tipo tumoral.

En cuanto a su plegamiento, se identificaron dominios idénticos en las estructuras secundarias predichas computacionalmente para la secuencia del *LINC01614*. Los dominios conservados potencialmente están relacionados con la función citoplasmática de secuestro de miRNAs. Sin embargo, es necesario validar computacionalmente y/o experimentalmente si en estos dominios existen secuencias complementarias a miR-383 y miR-217, que le permitan ejercer dicha función.

Tras este análisis, la investigación se dirigió hacia el estudio de los niveles expresión del *LINC0614* en tumores de mama y la asociación que muestran con la supervivencia global. En la cohorte mexicana, se encontró una prevalencia del subtipo Luminal A (45%), Luminal B (37%), HER2 (7%) y Triple negativo (9%), ligeramente diferente a las proporciones identificadas en otras investigaciones realizadas en población mexicana. En un estudio se reportó que las prevalencias para cada subtipo son 65%, 12%, 9% y 14%, respectivamente (Pérez-Rodríguez, 2015). En otro trabajo, se indicó que el 70%, 17%, 5% y 8% corresponde a cada subtipo (Valle-Solís *et al.*, 2019). Por otro lado, respecto a la cohorte del GEO, las proporciones para cada uno son 43%, 21%, 12% y 24%. A pesar de las diferencias, se observa que en todos los grupos se cumple con que el subtipo Luminal A es el más frecuente, y que en el subtipo HER2 se ubica el menor número de casos.

Adicionalmente, se observa que los datos experimentales se comportan de la manera esperada, dado que los subtipos clasificados por IHQ muestran asociaciones significativas con la mayoría de las características clínico-patológicas, las cuales ya se encuentran ampliamente descritas en la literatura. Por ejemplo, se ha demostrado que las pacientes mayores a 50 años se relacionan con los subtipos Luminal A y Luminal B, ya que los receptores hormonales (ER y PR) positivos son más frecuentes en las pacientes posmenopáusicas (Mills *et al.*, 2019; Ellingjord-Dale *et al.*, 2017). En contraste, en las pacientes menores a 50 años prevalecen los subtipos HER2 y Triple negativo, pues en la premenopausia estos receptores son menos frecuentes (Pérez-Rodríguez, 2015). Otro ejemplo es el grado tumoral, el cual refleja la agresividad de la enfermedad (Harbeck *et al.*, 2019). En el caso del subtipo Luminal A, por lo general el grado del tumor es bajo (85% de los casos) (De Kruijf *et al.*, 2014), lo que es indicativo de un buen pronóstico para las pacientes (probabilidad de supervivencia, 93%) (Howlader *et al.*, 2018), mientras que en las pacientes Luminal B el grado tumoral tiende a ser intermedio o alto (93% de los casos) (De Kruijf *et al.*, 2014), lo cual les confiere un peor pronóstico (probabilidad de supervivencia, 90%) (Howlader *et al.*, 2018). Por su parte, los subtipos HER2 y Triple negativo se caracterizan por un grado tumoral alto, que se asocia con un mal

pronóstico. Éste comportamiento es particularmente observado en el subtipo triple negativo (Luminal A > Luminal B > HER2 > Triple negativo) (Harbeck *et al.*, 2019).

Cuando se comparó la expresión del *LINC01614* entre los subtipos por IHQ, se observó que las pacientes Luminal A y Luminal B, lo mismo que las pacientes HER2 y Triple negativo, mostraron niveles de expresión semejantes del transcrito ($p=0.499$ y $p=0.490$, respectivamente). Numerosos reportes han mostrado que estos subtipos comparten ciertas similitudes, tales como la expresión de receptores hormonales, la expresión de genes relacionados con la proliferación celular y el grado tumoral, por mencionar algunas (Zepeda-Castilla *et al.*, 2008; Harbeck *et al.*, 2019). La diferencia estadísticamente significativa entre los niveles de expresión del *LINC01614* de los grupos Luminal y No Luminal ($p=0.022$), sugiere que el transcrito promueve el desarrollo y la progresión del CaMa a través de mecanismos de acción distintos, de acuerdo con el subtipo inmunohistoquímico.

El análisis de supervivencia global realizado en la cohorte del GEO, mostró que en pacientes con CaMa, la alta expresión del *LINC01614* se asocia significativamente con un menor tiempo de supervivencia global ($p=0.0264$), lo cual es consistente con los resultados observados en la cohorte de pacientes mexicanas. Si bien no se encontró una asociación significativa, se observó la misma tendencia, la cual indica que la sobre-expresión del *LINC01614* es de mal pronóstico clínico ($p=0.063$). Adicionalmente, en ambas cohortes se identificaron diferencias en la supervivencia global con base en el subtipo inmunohistoquímico. En pacientes del subtipo Luminal B, analizar la expresión del *LINC01614* es especialmente útil para predecir el tiempo de supervivencia global (Cohorte GEO, $p= 0.017$; Cohorte FUCAM-INMEGEN, $p= 0.045$). Las muestras de tumores agrupadas dentro del grupo Luminal mostraron una asociación significativa entre la supervivencia global y los niveles de expresión del transcrito ($p=0.039$). Este hallazgo señala que el nivel de expresión del *LINC01614* probablemente no es un factor que influya en el pronóstico de las pacientes con tumores del grupo no Luminal (HER2 y Triple negativo).

Los resultados del análisis por subtipo inmunohistoquímico demuestran que el tiempo de supervivencia global promedio no depende exclusivamente del subtipo del tumor, si no que existen otros factores génicos poco explorados que juegan un papel importante en el desarrollo de la enfermedad.

Dada la asociación existente entre los niveles de expresión del *LINC01614* y la supervivencia global en el subtipo Luminal B, se identificaron los transcritos diferencialmente expresados entre los grupos con alta y baja expresión. Los transcritos que se muestran regulados positivamente están involucrados en procesos relacionados con la proliferación celular, la migración celular, la invasión, la degradación de la matriz extracelular, la angiogénesis, y la transición epitelial-mesenquimal. Se encontró que algunos tipos de colágeno, como *COL10A1* y *COL11A1*, están sobre-expresados, lo cual también se reportó en diferentes tipos tumorales, incluido el CaMa (Chapman *et al.*, 2012; Makoukji *et al.*, 2016). El colágeno incrementa la angiogénesis así como la resistencia a la apoptosis, lo que promueve la progresión tumoral (Januchowski *et al.*, 2016; Chapman *et al.*, 2012). La sobre-expresión de la metaloproteinasa de matriz *MMP3*, favorece la capacidad de invasión de las células cancerosas a través de la degradación de la matriz extracelular (Wu *et al.*, 2013; Curran y Murray, 1999). En general, los procesos se relacionan con la transición de las células epiteliales al fenotipo mesenquimal. Éste último se caracteriza por el aumento en la migración, la invasividad, la resistencia a la apoptosis y la producción elevada de componentes de la matriz extracelular (Lamouille, Xu y Derynck, 2004).

En el grupo con el comportamiento contrario (regulados negativamente), se identificó el 82% de los transcritos diferencialmente expresados. Dentro de ellos, se identificaron genes involucrados en el proceso de apoptosis, la regulación de la proliferación, el reconocimiento inmunológico, el metabolismo de lípidos y la angiogénesis. Por ejemplo, *CA4* y *SCARA5* se han reconocido en diversos tipos de cáncer por sus papeles de supresores tumorales, promoviendo la apoptosis (Zhang *et al.*, 2015; You *et al.*, 2017). El primero de ellos también disminuye la expresión de

proteínas relacionadas con la transición epitelial-mesenquimal (Liang *et al.*, 2016). La baja expresión del antígeno *CD300LG* en tejidos tumorales se asocia con una disminución de la función citotóxica de las células del sistema inmunológico, lo cual les permite a las células tumorales el escape inmunológico (Zhai *et al.*, 2016). *KCNJ16* es un transcrito que se ha reportado subexpresado en diversos tipos de cáncer. Aunque se desconoce su mecanismo de acción, la evidencia indica que este canal de potasio puede desempeñar un papel importante en la carcinogénesis (Zhang *et al.*, 2020).

Los resultados señalan que la alta expresión del *LINC0614* en el CaMa Luminal B puede potencialmente estar alterando los niveles de expresión de genes relevantes que participan en procesos como la proliferación celular, la migración celular, la apoptosis, la angiogénesis, el reconocimiento inmunológico y la transición epitelial-mesenquimal. La combinación del efecto que ejercen todos estos genes desregulados podría estar favoreciendo el desarrollo y la progresión del CaMa, lo cual nos brinda un acercamiento al posible papel regulador del *LINC0614* en múltiples vías relacionadas al cáncer.

Posteriormente, el análisis de supervivencia global se estratificó mediante las características clínico-patológicas de la cohorte, con la finalidad de precisar en cuáles pacientes representaría un mayor beneficio clínico medir los niveles de expresión del *LINC0614*. En las pacientes de 50 años o menos, se observó una tendencia a la disminución de la supervivencia global cuando el *LINC0614* se sobreexpresa ($p=0.062$). Tal como se mencionó anteriormente, los subtipos Luminal están mayormente relacionados con mujeres que se encuentran en la posmenopausia, es decir, mujeres mayores de 50 años. Aunque éste resultado no alcanza una significancia estadística, nos indica que la regulación positiva del *LINC0614* puede ser un factor de riesgo para la supervivencia de las pacientes únicamente cuando la enfermedad se presenta en una edad temprana (≤ 50 años). Por el contrario, en una edad avanzada (> 50 años), los niveles de expresión del *LINC0614* pueden no ser relevantes para el pronóstico de las pacientes.

Por otro lado, se identificó que en tumores de bajo grado (\leq II) existe una asociación estadísticamente significativa entre los niveles de expresión del *LINC01614* y la probabilidad de supervivencia global ($p=0.0045$). Este resultado se relaciona con otro de los hallazgos del estudio, en el cual se reveló que en aquellas pacientes sin evento de recurrencia durante el estudio, el tiempo de supervivencia fue significativamente más corto cuando los tumores presentaban un patrón de sobre-expresión del transcrito ($p= 0.047$). El grado del tumor es un parámetro que mide el grado de diferenciación del tejido tumoral, y refleja cuán estrechamente relacionado está a las células epiteliales de la mama normal. Es decir, un tumor bien diferenciado o de bajo grado tumoral, es aquel que muestra una alta homología con la unidad ductal-lobular terminal de la mama normal (Pereira *et al.*, 1995; Rakha *et al.*, 2010). La recurrencia es un término que indica la reaparición del CaMa después de concluir el tratamiento primario (Ahmad, 2013). Ambos resultados se relacionan entre sí, en que las pacientes con bajo grado tumoral tienen probabilidades de recaídas significativamente menores que aquellas con un alto grado tumoral (Rakha *et al.*, 2010; Liang *et al.*, 2020). Los tumores de alto grado suelen ser resistentes a las terapias convencionales, ya sea por la presencia de células troncales tumorales, o bien, porque las células del tumor adquieren un fenotipo más agresivo, lo que les permite sobrevivir y multiplicarse para regenerar el tumor (Ahmad, 2013). En consecuencia, en tumores de alto grado, la enfermedad tiende a reaparecer.

Con ello se puede concluir que en pacientes con tumores bien diferenciados, es potencialmente útil identificar si existe una baja o alta expresión del *LINC01614* para estimar el pronóstico de la supervivencia global. Cabe resaltar que aunque la literatura indica que la mayoría de los tumores Luminal B son de grado intermedio o alto (93%) (De Kruijf *et al.*, 2014), las estadísticas de la cohorte FUCAM-INMEGEN muestran un comportamiento diferente. En este caso, las pacientes Luminal B tienen tumores predominantemente de grado bajo o intermedio (59%).

Por último, nuestros resultados mostraron que aún en pacientes sin recurrencia, existen diferencias en el tiempo de supervivencia promedio. Específicamente, en pacientes con sobre-expresión del *LINC01614*, la supervivencia global promedio fue 6 meses menor respecto a las pacientes con sub-expresión. Este dato sugiere que la cuantificación de los niveles de expresión de *LINC01614* en un tumor diferenciado podría ser determinante para conocer con precisión el pronóstico de una paciente.

Las aportaciones de este trabajo permitieron discutir cómo la combinación de características clínico-patológicas con la información proporcionada por la expresión de un RNA particular, puede ayudar a comprender mejor la complejidad biológica de una enfermedad como el CaMa. Además, resaltaron el papel potencial que presenta *LINC01614* en tumores de mama como biomarcador para la predicción del pronóstico. Ésto nos permite aportar más pruebas experimentales que respaldan lo postulado por otros grupos de investigación en otros tipos tumorales.

En general, este estudio muestra que los niveles de expresión del *LINC01614* se asocian con la probabilidad de supervivencia global de las pacientes de acuerdo con el subtipo de la clasificación por IHQ. Además, señala que el grado tumoral es un factor importante e independiente a tomar en cuenta para definir con mayor precisión el pronóstico del CaMa en población mexicana. En específico, se muestra cuáles pacientes se pueden ver mayormente beneficiadas con la determinación de los niveles de expresión del *LINC01614*. Éstas pacientes son aquellas que en la {evaluación histológica muestren un bajo grado tumoral (\leq II)}. Así mismo, que se clasifiquen por IHQ en uno de los subtipos Luminal, particularmente en el subtipo Luminal B. Sin embargo, es importante mencionar que este estudio también presenta ciertas limitaciones, principalmente ocasionadas por la baja proporción de muestras de los subtipos inmunohistoquímicos HER2 y Triple negativo.

11. CONCLUSIONES

Los resultados de esta investigación permiten concluir lo siguiente:

- La sobre-expresión del *LINC01614*, en muestras independientes de tumores de mama, se asocia con un pronóstico significativamente desfavorable. Esta asociación fue particularmente observada en pacientes con CaMa Luminal B, lo cual se validó en una cohorte mexicana.
- El grado tumoral es una característica clínico-patológica que se asocia significativamente con los niveles de expresión del *LINC01614*. La sobre-expresión del transcrito pronostica una menor supervivencia global en las pacientes con tumores de mama de bajo grado (\leq II).
- En el CaMa Luminal B, el *LINC01614* potencialmente regula los niveles de expresión de genes que participan en procesos relacionados al cáncer.
- En población mexicana, el *LINC01614* tiene un papel potencial como biomarcador de pronóstico del CaMa.

12. PERSPECTIVAS

Queda como propuesta para continuar el proyecto lo siguiente:

- Llevar a cabo un ensayo de fraccionamiento subcelular en líneas celulares de CaMa. La determinación de la localización celular del *LINC01614* puede relacionarse con el mecanismo de acción asociado al desarrollo del cáncer en tumores de mama y sus potenciales moléculas blanco.
- Ampliar el número de muestras de la cohorte FUCAM-INMEGEN, en particular de los subtipos HER2 y Triple Negativo. Ésto con la finalidad de alcanzar la significancia estadística en la asociación entre los niveles de expresión del *LINC01614* y la supervivencia global, tal como se observa en la cohorte de la base de datos GEO.
- A partir de los resultados del análisis de expresión diferencial, realizar un análisis de enriquecimiento funcional de los transcritos diferencialmente expresados en las pacientes Luminal B con alta y baja expresión del *LINC01614*.

13. REFERENCIAS BIBLIOGRÁFICAS

- Abolghasemi, M., Tehrani, S. S., Yousefi, T., Karimian, A., Rameshknia, V. *et al.* (2020). Critical roles of long noncoding RNAs in breast cancer. *Journal of Cellular Physiology*. doi:10.1002/jcp.29442.
- Ahmad, A. (2013). *Pathways to Breast Cancer Recurrence*. *ISRN Oncology*, 1–16. doi:10.1155/2013/290568.
- Akram, M., Iqbal, M., Daniyal, M. & Khan, A. (2017). Awareness and current knowledge of breast cancer. *Biological Research*, 50(1). doi:10.1186/s40659-017-0140-9.
- Anderson, D. M., Anderson, K. M. *et al.* (2015). A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell*, 160(4), 595–606. doi:10.1016/j.cell.2015.01.009.
- Balas, M. M., & Johnson, A. M. (2018). Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-Coding RNA Research*. doi:10.1016/j.ncrna.2018.03.001.
- Bartella, L., Smith, C., Dershaw, D. & Liberman, L. (2007). Imaging Breast Cancer. *Radiologic Clinics of North America*, 45(1), 45–67. doi:10.1016/j.rcl.2006.10.007.
- Bassett, A. R., Akhtar, A., Barlow, D. P., Bird, A. P., Brockdorff, N., Duboule, D., Ponting, C. P. *et al.* (2014). Considerations when investigating lncRNA function in vivo. *eLife*, 3. doi:10.7554/eLife.03058.
- Bastien, R. R., Rodríguez, Á. *et al.* (2012). PAM50 Breast Cancer Subtyping by RT-qPCR and Concordance with Standard Clinical Molecular Markers. *BMC Medical Genomics*, 5(1). doi:10.1186/1755-8794-5-44.
- Batista, P. J., & Chang, H. Y. (2013). Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell*, 152(6), 1298–1307. doi:10.1016/j.cell.2013.02.012.
- Bayoglu, B. & Cengiz, M. (2017). The roles of the long non-coding RNA transcripts in cardiovascular diseases. *Bezmialem Science*, 5: 74-9. doi:10.14235/bs.2016.96.
- Beermann, J., Piccoli, M., Viereck, J. & Thum, T. (2016). Non-coding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. *Physiological Reviews*, 96(4), 1297-1325.
- Bhan, A., Soleimani, M., & Mandal, S. S. (2017). Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Research*, 77(15), 3965–3981. doi:10.1158/0008-5472.can-16-2634.
- Bhargava, R., Striebel, J., Beriwal, S., Flickinger, J. C., & Dabbs, D. J. (2009). Prevalence, morphologic features and proliferation indices of breast carcinoma molecular classes using immunohistochemical surrogate markers. *International journal of clinical and experimental pathology*, 2(5), 444–455.
- Bin, X., Hongjian, Y., Xiping, Z., Bo, C., Shifeng, Y., & Binbin, T. (2018). Research progresses in roles of lncRNA and its relationships with breast cancer. *Cancer Cell International*, 18(1).
- Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. (2001). *Clinical Pharmacology & Therapeutics*, 69(3), 89–95. doi:10.1067/mcp.2001.113989.
- Bombonati, A. & Sgroi, D. (2010). The molecular pathology of breast cancer progression. *The Journal of Pathology*, 223(2), 308–318. doi:10.1002/path.2808.
- Brandan, M. & Villaseñor, N. (2006) Detección del cáncer de mama: estado de la mamografía en México. *Cancerología*, 1, 147–162.
- Bray, F., Ferlay, J., *et al.* (2015). Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. *International Journal of Cancer*, 137(9), 2060–2071.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394-424. doi:10.3322/caac.21492.
- Cahill, D., Kinzler, K., Vogelstein, B. & Lengauer, C. (1999). Genetic instability and Darwinian selection in tumours. *Trends in Cell Biology*. 9 (12), M57–M60.
- Califf, R. (2018). Biomarker definitions and their applications. *Experimental Biology and Medicine*, 243(3), 213-221.
- Cancer Today by International Agency for Research on Cancer 2020. Disponible en: <http://gco.iarc.fr/> (Consultado el 02 de Febrero de 2020).
- Campoy, E. M., Laurito, S. R. *et al.* (2016). Asymmetric Cancer Hallmarks in Breast Tumors on Different Sides of the Body. *PLOS ONE*, 11(7), e0157416. doi:10.1371/journal.pone.0157416.
- Chandra Gupta, S., & Nandan Tripathi, Y. (2016). Potential of long non-coding RNAs in cancer patients: From biomarkers to therapeutic targets. *International Journal of Cancer*, 140(9), 1955–1967.
- Chapman, K. B., Prendes, M. J. *et al.* (2012). COL10A1 expression is elevated in diverse solid tumor types and is associated with tumor vasculature. *Future Oncology*, 8(8), 1031–1040. doi:10.2217/fon.12.79.
- Chatterjee, G., Pai, T., Hardiman, T., Grigoriadis, A. *et al.* (2018). Molecular patterns of cancer colonisation in lymph nodes of breast cancer patients. *Breast Cancer Research*, 20(1). doi:10.1186/s13058-018-1070-3.

- Chauhan, A., Sachan, P. K., Chauhan, N. (2016). A study of correlation between molecular subtypes of breast cancer and site of metastasis. *Clinical Cancer Investigation Journal*, 5, 393-7.
- Chen J., Silver D., Cantor, S., Livingston, D. & Scully, R. (1999) BRCA 1, BRCA2, and Rad51 operative in a common DNA damage response pathway. *Cancer Research*, 59(7), 1752–1756.
- Chen, Q., Zhu, C., & Jin, Y. (2020). The Oncogenic and Tumor Suppressive Functions of the Long Noncoding RNA MALAT1: An Emerging Controversy. *Frontiers in genetics*, 11, 93. doi:10.3389/fgene.2020.00093.
- Chen, D., Sun, Q. *et al.* (2016). Genome-wide analysis of long noncoding RNA (lncRNA) expression in colorectal cancer tissues from patients with liver metastasis. *Cancer Medicine*, 5(7), 1629–1639. doi:10.1002/cam4.738.
- Cheng, J., Kapranov, P., *et al.* (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, 308(5725), 1149–1154. doi:10.1126/science.1108625.
- Chi, Y., Wang, Di., Wang, J., Yu, W., & Yang., Y. (2019). Long Non-Coding RNA in the Pathogenesis of Cancers. *Cells*, 8(9), 1015. doi:10.3390/cells8091015.
- Cree, I. (2011). Cancer Biology. *Cancer Cell Culture*, 1–11. doi:10.1007/978-1-61779-080-5_1.
- Creighton C. J. (2012). The molecular profile of luminal B breast cancer. *Biologics: targets & therapy*, 6, 289–297.
- Cruz, J. A., & Westhof, E. (2009). The Dynamic Landscapes of RNA Architecture. *Cell*, 136(4), 604–609. doi:10.1016/j.cell.2009.02.003.
- Curran, S., & Murray, G. I. (1999). Matrix metalloproteinases in tumor invasion and metastasis. *The Journal of pathology*, 189(3), 300–308.
- Dahariya, S., Paddibhatla, I. *et al.* (2019). Long non-coding RNA: Classification, biogenesis and functions in blood cells. *Molecular Immunology*, 112, 82-92. doi:10.1016/j.molimm.2019.04.011.
- Dall, G. & Britt, K. (2017). Estrogen Effects on the Mammary Gland in Early and Late Life and Breast Cancer Risk. *Frontiers in Oncology*, 110(7). doi:10.3389/fonc.2017.00110.
- Danaei, G., Vander Hoorn, S., Lopez, A., Murray, C. & Ezzati, M. (2005). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *The Lancet*, 366(9499), 1784–1793. doi:10.1016/s0140-6736(05)67725-2.
- De Kruijf, E. M., Bastiaannet, E. *et al.* (2014). Comparison of frequencies and prognostic effect of molecular subtypes between young and elderly breast cancer patients. *Molecular Oncology*, 8(5), 1014–1025.
- Dellaire, G., Berman, J. & Arceci, R. (2014) Breast Cancer Genomics in *Cancer genomics: from bench to personalized medicine*. Academic Press. ISBN 9780123972743.
- Derrien, T., Johnson, R. *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, 22(9), 1775–1789.
- Dieci, G., Fiorino, G., Castelnovo, M., Teichmann, M., & Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends in Genetics*, 23(12), 614–622. doi:10.1016/j.tig.2007.09.001.
- Dillon, D., Guidi, A. & Schnitt, S. (2014) *Pathology of invasive breast cancer*. Diseases of the Breast. 5th ed. Philadelphia, PA: Wolters Kluwer Health.
- Drew, P., Cawthorn, S. & Michell, M. (2007) *Interventional Ultrasound of the Breast - Anatomy of the Breast by Menos Lagopoulos*. London, UK: Informa Healthcare, 11–22.
- Driscoll, J. J., & Rixe, O. (2009). Overall Survival: Still the Gold Standard. *The Cancer Journal*, 15(5), 401–405. doi:10.1097/ppo.0b013e3181bdc2e0.
- Du, Z., Fei, T. *et al.* (2013). Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nature Structural & Molecular Biology*, 20(7), 908–913. doi:10.1038/nsmb.2591.
- Dudley, W., Wickham, R., Coombs, N. (2016). An Introduction to Survival Statistics: Kaplan-Meier Analysis. *Journal of Advanced Practitioner in Oncology*, 7(1), 91-100. doi: 10.6004/jadpro.2016.7.1.8.
- Dunham, I., Kundaje, A., Aldred, S. *et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. doi:10.1038/nature11247.
- Ehinger, A., Malmström, P. *et al.* (2016). Histological grade provides significant prognostic information in addition to breast cancer subtypes defined according to St Gallen 2013. *Acta Oncologica*, 56(1), 68–74.
- Ellingjord-Dale, M., Vos, L. *et al.* (2017) Parity, hormones and breast cancer subtypes - results from a large nested case-control study in a national screening program. *Breast Cancer Research* 19, 10.
- Ellis, M., Ding, L., Shen, D., Luo, J., Suman, V., Wallis, J., *et al.* (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*, 486(7403), 353–360. doi:10.1038/nature11143.

- Ellis, M. J., Tao, Y. *et al.* (2008). Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. *Journal of the National Cancer Institute*, 100(19), 1380–1388. doi:10.1093/jnci/djn309.
- Elston, C. (1999). Pathological prognostic factors in breast cancer. *Critical Reviews in Oncology/Hematology*, 31(3), 209–223. doi:10.1016/s1040-8428(99)00034-7.
- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489: 57–74. doi:10.1038/nature11247.
- Ensembl Genome Browser. Disponible: <https://www.ensembl.org/index.html> (Consultado 07 de Abril de 2020).
- Eroles, P., Bosch, A., Pérez-Fidalgo, J. & Lluch, A. (2012). Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. *Cancer Treatment Reviews*, 38(6), 698–707.
- Feng, Y., Spezia, M. *et al.* (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & Diseases*, 5(2), 77–106. doi:10.1016/j.gendis.2018.05.001.
- Foulkes, W. D., Brunet, J. S. *et al.* (2004). The prognostic implication of the basal-like phenotype of BRCA1-related breast cancer. *Cancer research*, 64(3), 830–835. doi:10.1158/0008-5472.can-03-2970.
- Franca, A., Ferreira, M., Franca, J., Franca, E. & Honorio-Franca, A. (2012) Breastfeeding and its relationship with reduction of breast cancer: a review. *Asian Pacific Journal of Cancer Prevention*, 13(11), 5327–32.
- Fu, X., Ravindranath, L., Tran, N., Petrovics, G., & Srivastava, S. (2006). Regulation of Apoptosis by a Prostate-Specific and Prostate Cancer-Associated Noncoding Gene, PCGEM1. *DNA and Cell Biology*, 25(3), 135–141.
- Gad, A. A., & Balenga, N. (2020). The Emerging Role of Adhesion GPCRs in Cancer. *ACS Pharmacology & Translational Science*. doi:10.1021/acspsci.9b00093.
- GENCODE: the reference human genome annotation for The ENCODE Project. Disponible en: <https://www.encodegenes.org/pages/publications.html/> (Consultado el 05 de Marzo de 2020).
- Geyer, F. C., Marchiò, C., & Reis-Filho, J. S. (2009). The role of molecular analysis in breast cancer. *Pathology*, 41(1), 77–88. doi:10.1080/00313020802563536.
- Giuliano, A. E., Connolly, J. L. *et al.* (2017). Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA: A Cancer Journal for Clinicians*, 67(4), 290–303.
- Gökmen-Polar, Y. & Badve, S. (2012) Molecular profiling assays in breast cancer: are we ready for prime time?. *Oncology (Williston Park)*, 26(4), 350–357.
- Goldgar, D., Easton, D., Cannon-Albright, L. & Skolnick, M. (1994). Systematic Population-Based Assessment of Cancer Risk in First-Degree Relatives of Cancer Probands. *JNCI: Journal of the National Cancer Institute*, 86(21), 1600–1608. doi:10.1093/jnci/86.21.1600.
- Guigó, R. (2013). *Handbook of Systems Biology*. Chapter 2 - The Coding and the Non-coding Transcriptome. doi:10.1016/B978-0-12-385944-0.00002-2.
- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., Chang, H. Y. *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 464(7291), 1071–1076. doi:10.1038/nature08975.
- Han, X., Luo, S. *et al.* (2018). Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. *Journal of Molecular Cell Biology*, 10(2), 175–178.
- Hanahan, D., & Weinberg, R. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. doi:10.1016/j.cell.2011.02.013.
- Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., *et al.* (2019). Breast cancer. *Nature Reviews Disease Primers*, 5(66). doi:10.1038/s41572-019-0111-2.
- Hidalgo-Miranda, A. & Jiménez-Sánchez, G. (2009) Bases genómicas del cáncer de mama: avances hacia la medicina personalizada. *Salud Publica Mex* 2009;51 supl 2:S197-S207.
- Hombach, S., & Kretz, M. (2016). Non-coding RNAs: Classification, Biology and Functioning. *Non-Coding RNAs in Colorectal Cancer*, 3–17. doi:10.1007/978-3-319-42059-2_1.
- Howlader, N., Cronin, K. A., Kurian, A. W., & Andridge, R. (2018). Differences in Breast Cancer Survival by Molecular Subtypes in the United States. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research*, 27(6), 619–626. doi.org/10.1158/1055-9965.EPI-17-0627.
- Hu, Z., Fan, C. *et al.* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC genomics*, 7, 96. doi:10.1186/1471-2164-7-96.
- Huang, H., Li, T. *et al.* (2018). High expression of COL10A1 is associated with poor prognosis in colorectal cancer. *OncoTargets and Therapy*, Volume 11, 1571–1581. doi:10.2147/ott.s160196.

- Huen, M., Sy, S. & Chen, J. (2009). BRCA1 and its toolbox for the maintenance of genome integrity. *Nature Reviews Molecular Cell Biology*, 11(2), 138–148. doi:10.1038/nrm2831.
- Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*, 147(4), 789–802. doi:10.1016/j.cell.2011.10.002.
- Instituto Nacional de Estadística y Geografía (INEGI). Causas de defunción 2014. Disponible en: <https://www.inegi.org.mx/> (Consultado el 02 de Febrero de 2020).
- Irwig, L., Macaskill, P. & Houssami, N. (2002). Evidence relevant to the investigation of breast symptoms: the triple test. *The Breast*, 11(3), 215–220. doi:10.1054/brst.2001.0409.
- Januchowski, R., Zabel, M. *et al.* (2016). Increased Expression of Several Collagen Genes is Associated with Drug Resistance in Ovarian Cancer Cell Lines. *Journal of Cancer*, 7(10), 1295–1310. doi:10.7150/jca.15371
- Ji, P., Diederichs, S. *et al.* (2003). MALAT-1, a novel noncoding RNA and thymosin β 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, 22(39), 8031–8041. doi:10.1038/sj.onc.1206928.
- Kang, Y. J. *et al.* (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic acids research*, 45(W1), W12–W16. doi:org/10.1093/nar/gkx428.
- Kapranov, P., Cheng, J. *et al.* (2007). RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science*, 316(5830), 1484–1488. doi:10.1126/science.1138341.
- Kazimierczyk, M., Kasproicz, M., Kasprzyk, M. & Wrzesinski, J. (2020). Human Long Noncoding RNA Interactome: Detection, Characterization and Function. *International Journal of Molecular Sciences*, 21.
- Kennecke, H., Yerushalmi, R. *et al.* (2010). *Metastatic Behavior of Breast Cancer Subtypes*. *Journal of Clinical Oncology*, 28(20), 3271–3277. doi:10.1200/jco.2009.25.9820.
- Key, T., Verkasalo, P. & Banks, E. (2001). Epidemiology of breast cancer. *The Lancet Oncology*, 2(3), 133–140. doi:10.1016/s1470-2045(00)00254-0.
- Kluz, P. N., Kolb, R., Xie, Q. *et al.* (2020). Cancer cell-intrinsic function of CD177 in attenuating β -catenin signaling. *Oncogene* 39, 2877–2889. doi:10.1038/s41388-020-1203-x.
- Kung, J. T. Y., Colognori, D., & Lee, J. T. (2013). Long Noncoding RNAs: Past, Present, and Future. *Genetics*, 193(3), 651–669. doi:10.1534/genetics.112.146704.
- Kudo, Y., Iizuka, S. *et al.* (2012). Matrix Metalloproteinase-13 (MMP-13) Directly and Indirectly Promotes Tumor Angiogenesis. *Journal of Biological Chemistry*, 287(46), 38716–38728. doi:10.1074/jbc.m112.373159.
- Kveiborg, M., Albrechtsen, R., Couchman, J. R., & Wewer, U. M. (2008). Cellular roles of ADAM12 in health and disease. *The International Journal of Biochemistry & Cell Biology*, 40(9), 1685–1702.
- Lal, S., McCart Reed, A., de Luca, X., & Simpson, P. (2017). Molecular signatures in breast cancer. *Methods*, 131, 135-146. doi:10.1016/j.ymeth.2017.06.032.
- Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, 15(3), 178–196. doi:10.1038/nrm3758.
- Lee, J., Davidow, L. S., & Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics*, 21(4), 400–404. doi:10.1038/7734.
- Li, J., Han, L., Roebuck, P., Yuan, Y., Weinstein, J. N., & Liang, H. (2015). TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer Research*, 75(18), 3728-3737.
- Liang, Q., Ma, D., Gao, R. *et al.* (2020) Effect of Ki-67 Expression Levels and Histological Grade on Breast Cancer Early Relapse in Patients with Different Immunohistochemical-based Subtypes. *Scientific Reports*, 10, 7648.
- Liang, W., Lai, Y. *et al.* (2016). Combretastatin A4 Regulates Proliferation, Migration, Invasion, and Apoptosis of Thyroid Cancer Cells via PI3K/Akt Signaling Pathway. *Medical science monitor: International Medical Journal of Experimental and Clinical Research*, 22, 4911–4917. doi:10.12659/msm.898545.
- Liu, A.-N., Qu, H.-J., Yu, C.-Y., & Sun, P. (2018). Knockdown of LINC01614 inhibits lung adenocarcinoma cell progression by up-regulating miR-217 and down-regulating FOXP1. *Journal of Cellular and Molecular Medicine*, 22(9), 4034–4044. doi:10.1111/jcmm.13483.
- Liu, J., Peng, W. X., Mo, Y. Y., Luo, D. (2017) MALAT1-mediated tumorigenesis. *Frontiers in Bioscience*, 22:66–80.
- Liu, F., Somarowthu, S., & Pyle, A. M. (2017). Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nature chemical biology*, 13(3), 282–289. doi:10.1038/nchembio.2272.
- Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Research*, 2018. Disponible en: <http://www.bio-bigdata.com/lnc2cancer/> (Consultado el 26 de Marzo de 2020).

- Lopez-Garcia, M., Geyer, F., Lacroix-Triki, M., Marchió, C. & Reis-Filho, J. (2010). Breast cancer precursors revisited: molecular features and progression pathways. *Histopathology*, 57(2), 171–192.
- Losko, M., Kotlinowski, J., & Jura, J. (2016). Long Noncoding RNAs in Metabolic Syndrome Related Disorders. *Mediators of Inflammation*, 2016, 1–12. doi:10.1155/2016/5365209.
- Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA biology*, 10(6), 925–933. <https://doi.org/10.4161/rna.24604>
- Makoukji, J., Makhoul, N. J., Khalil, M., El-Sitt, S., Aldin, E. S., Jabbour, M., Tfayli, A. (2016). Gene expression profiling of breast cancer in Lebanese women. *Scientific Reports*, 6(1). doi:10.1038/srep36639.
- Martínez-Aguirre, A. (2019). *Análisis de expresión de RNA largo no codificante (LINC-01614) en carcinoma ductal infiltrante de la mama y su asociación al curso clínico*. Tesis de especialidad. Universidad Nacional Autónoma de México, México.
- Martínez-Férez, I., Viguera-Guerra, I., Lozano, M., et al. (2018) *Plataformas genómicas de carácter pronóstico-predictivo en el cáncer de mama: actualización de la evidencia. Revisión Sistemática*. Sevilla: Agencia de Evaluación de Tecnologías Sanitarias de Andalucía.
- Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Pulido, T., Guigo, R., Johnson, R. (2017) LncATLAS database for subcellular localization of long noncoding RNAs. *RNA*, 23(7), 1080-1087.
- Mattick, J. S. (2004). *RNA regulation: a new genetics?* *Nature Reviews Genetics*, 5(4), 316–323. doi:10.1038/nrg1321.
- Meirson, T., Gil-Henn, H., & Samson, A. O. (2019). *Invasion and metastasis: the elusive hallmark of cancer. Oncogene*. doi:10.1038/s41388-019-1110-1.
- Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*, 20(3), 300–307. doi:10.1038/nsmb.2480.
- Militello, G., Weirick, T., John, D., Döring, C., Dimmeler, S., & Uchida, S. (2016). *Screening and validation of lncRNAs and circRNAs as miRNA sponges. Briefings in Bioinformatics, bbw053*. doi:10.1093/bib/bbw053.
- Mills, M. et al. (2019). The prevalence of luminal B subtype is higher in older postmenopausal women with ER+/HER2– breast cancer and is associated with inferior outcomes. *Journal of Geriatric Oncology*, 10(6), S19–S20. doi:10.1016/s1879-4068(19)31150-6.
- Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, 15(6), 423–437. doi:10.1038/nrg3722.
- National Center for Biotechnology Information (NCBI). Disponible en: <https://www.ncbi.nlm.nih.gov>. (Consultado el 07 de Abril de 2020).
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534(7605), 47–54.
- Novikova, I. V., Hennelly, S. P., Tung, C.-S., & Sanbonmatsu, K. Y. (2013). Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs. *Journal of Molecular Biology*, 425(19), 3731–3746.
- Organización Mundial de la Salud (OMS). Temas de Salud: Cáncer. Disponible en: <https://www.who.int/topics/cancer/es/> (Consultado el 16 de Febrero de 2020).
- Padilla-Palma, V. (2018). *Análisis de la expresión de un RNA largo no codificante alterado en cáncer de mama, empleando bases de datos genómicos*. Tesis de pregrado. Universidad Nacional Autónoma de México.
- Pandya, S. & Moore, R. (2011). Breast Development and Anatomy. *Clinical Obstetrics and Gynecology*, 54(1), 91–95. doi:10.1097/grf.0b013e318207ffe9.
- Park, J., & Scherer, P. E. (2011). Leptin and cancer: from cancer stem cells to metastasis. *Endocrine Related Cancer*, 18(4), C25–C29. doi:10.1530/erc-11-0163.
- Parker, J., Mullins, M., Cheang, M., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Journal of Clinical Oncology*, 27(8), 1160–1167.
- Parri, M., Pietrovito, L. et al. (2014). Angiopoietin-like 7, a novel pro-angiogenic factor over-expressed in cancer. *Angiogenesis*, 17(4), 881–896. doi:10.1007/s10456-014-9435-4.
- Pereira, H., Pinder, S. E., Ellis, I. O. et al. (1995). Pathological prognostic factors in breast cancer. IV: Should you be a typer or a grader? A comparative study of two histological prognostic features in operable breast carcinoma. *Histopathology*, 27(3), 219–226. doi:10.1111/j.1365-2559.1995.tb00213.x.
- Pérez-Rodríguez, G. (2015). Prevalencia de subtipos por inmunohistoquímica del cáncer de mama en pacientes del Hospital General Regional 72, Instituto Mexicano del Seguro Social. *Cirugía y Cirujanos*, 83(3), 193-198.
- Perou, C. M., Sørlie, T., Eisen, M., van de Rijn, M., Jeffrey, S., Rees, C., et al. (2000). Molecular portraits of human breast tumours. *Nature*, 406(6797), 747–752. doi:10.1038/35021093.

- Perou, C., Parker, J., Prat, A., Ellis, M., & Bernard, P. (2010). Clinical implementation of the intrinsic subtypes of breast cancer. *The Lancet Oncology*, 11(8), 718–719. doi:10.1016/s1470-2045(10)70176-5.
- Petrovics, G., Zhang, W., Makarem, M., Street, J. P., Connelly, R., Sun, L., Srivastava, S. *et al.* (2004). Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. *Oncogene*, 23(2), 605–611.
- Pinamonti, M. & Zanconati, F. (2018) Breast Cytopathology. Assessing the Value of FNAC in the Diagnosis of Breast Lesions. *Monographs in Clinical Cytology*, 24, 20–24. doi:10.1159/000479764.
- Ponting, C. P., Oliver, P. L., & Reik, W. (2009). *Evolution and Functions of Long Noncoding RNAs*. *Cell*, 136(4), 629–641. doi:10.1016/j.cell.2009.02.006.
- Portoso, M., Ragazzini, R., Brenčić, Ž., Moiani, A., Michaud, A., Vassilev, I., Margueron, R. *et al.* (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *The EMBO Journal*, 36(8), 981–994.
- Prat, A., Adamo, B., Cheang, M. C. U., Anders, C. K., Carey, L. A., & Perou, C. M. (2013). Molecular Characterization of Basal-Like and Non-Basal-Like Triple-Negative Breast Cancer. *The Oncologist*, 18(2), 123–133. doi:10.1634/theoncologist.2012-0397.
- Prat, A., Parker, J., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J., *et al.* (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research*, 12(5). doi:10.1186/bcr2635.
- Prensner, J. R., Iyer, M. K., Balbin, O. A., Dhanasekaran, S. M., Chinnaiyan, A. M. *et al.* (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature Biotechnology*, 29(8), 742–749. doi:10.1038/nbt.1914.
- Quinn, J. J., & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics*, 17(1), 47–62. doi:10.1038/nrg.2015.10.
- Rakha, E. A., El-Sayed, M. E., Reis-Filho, J. S., & Ellis, I. O. (2007). *Expression profiling technology: its contribution to our understanding of breast cancer*. *Histopathology*, 52(1), 67–81.
- Rakha, E. A., Elsheikh, S. E., *et al.* (2009). Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. *Clinical Cancer Research: an official journal of the American Association for Cancer Research*, 15(7), 2302–2310. doi:10.1158/1078-0432.CCR-08-2132.
- Rakha, E.A., Reis-Filho, J.S., Baehner, F. *et al.* (2010). Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Research*, 12, 207. doi:10.1186/bcr2607.
- Rapicavoli, N. A., Poth, E. M., Zhu, H., & Blackshaw, S. (2011). The long noncoding RNA Six3OS acts in trans to regulate retinal development by modulating Six3 activity. *Neural Development*, 6(1), 32.
- Rathinasamy, B., & Velmurugan, B. K. (2018). *Role of lncRNAs in the cancer development and progression and their regulation by various phytochemicals*. *Biomedicine & Pharmacotherapy*, 102, 242–248.
- Ratti M, Lampis A, Ghidini M, *et al.* MicroRNAs (miRNAs) and Long Non-Coding RNAs (lncRNAs) as New Tools for Cancer Therapy: First Steps from Bench to Bedside. *Targeted Oncology*, 15(3): 261-278.
- Reis-Filho, J. S., Weigelt, B., Fumagalli, D., & Sotiriou, C. (2010). Molecular profiling: moving away from tumor philately. *Science translational medicine*, 2(47), 47ps43. doi:10.1126/scitranslmed.3001329.
- Rinn, J. L., Kertesz, M., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 129(7), 1311–1323. doi:org/10.1016/j.cell.2007.05.022.
- Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry*, 81, 145–166. https://doi.org/10.1146/annurev-biochem-051410-092902.
- Rojas, K. & Stuckey, A. (2016). Breast Cancer Epidemiology and Risk Factors. *Clinical Obstetrics and Gynecology*, 59(4), 651–672. doi:10.1097/grf.0000000000000239.
- Rosen, P., Kosloff, C., Lieberman, P., *et al.* (1978) Lobular carcinoma in situ of the breast. Detailed analysis of 99 patients with average follow-up 24 years. *The American Journal of Surgical Pathology*, 2(3), 225–251.
- Sánchez, J., Rocha, J., Piña, V., Sánchez, G., Erazo Valle-Solis, A., Balcázar, C., Queijeiro, M. (2017). Consenso Mexicano sobre diagnóstico y tratamiento del cáncer mamario. Octava revisión. *Gaceta Mexicana de Oncología*, 16, 7-78. doi:10.24875/j.gamo.M18000133.
- Sánchez, Y., & Huarte, M. (2013). Long Non-Coding RNAs: Challenges for Diagnosis and Therapies. *Nucleic Acid Therapeutics*, 23(1), 15–20. doi:10.1089/nat.2012.0414.
- Sen, R., Ghosal, S., Das, S., Balti, S., & Chakrabarti, J. (2014). Competing Endogenous RNA: The Key to Posttranscriptional Regulation. *The Scientific World Journal*, 2014, 1–6. doi:10.1155/2014/896206.
- Shen, L., Yang, M., Lin, Q., Zhang, Z., Zhu, B., & Miao, C. (2016). COL11A1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. *Oncology Reports*, 36, 877-885. doi:10.3892/or.2016.4869.

- Shim, H. J., Kim, S. H. *et al.* (2014). Breast cancer recurrence according to molecular subtype. *Asian Pacific Journal of Cancer Prevention*, 15(14), 5539–5544. doi:10.7314/apjcp.2014.15.14.5539.
- Siegel, R., Miller, K. & Jemal, A. (2017). Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, 67(1), 7–30. doi:10.3322/caac.21387.
- Sinn, P., Aulmann, S., Wirtz, R., Schott, S., Marmé, F., Varga, Z., Schneeweiss, A. (2013). Multigene Assays for Classification, Prognosis, and Prediction in Breast Cancer: a Critical Review on the Background and Clinical Utility. *Geburtshilfe Und Frauenheilkunde*, 73(09), 932–940. doi:10.1055/s-0033-1350831.
- Sotiriou, C., Neo, S., McShane, L., Korn, E., Long, P., Jazaeri, A., *et al.* (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences*, 100(18), 10393–10398. doi:10.1073/pnas.1732912100.
- Sparano, J., Gray, R. J., Makower, D. F., Pritchard, K. I., Albain, K. S., Hayes, D. F., Sledge, G. (2015). Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *New England Journal of Medicine*, 373(21), 2005–2014. doi:10.1056/nejmoa1510764.
- Sun, Y., & Ling, C. (2019). Analysis of the long non-coding RNA LINC01614 in non-small cell lung cancer. *Medicine*, 98(30), e16437. doi:10.1097/md.00000000000016437.
- Sun, Y., Zhao, Z., Yang, Z., Xu, F., Lu, H., Zhu, Z. & Zhu, H. (2017). Risk Factors and Preventions of Breast Cancer. *International Journal of Biological Sciences*, 13(11), 1387–1397. doi:10.7150/ijbs.21635.
- Suresh, P. (2018). Breast cancer heterogeneity: a focus on epigenetics and in vitro 3D model systems. *Cell Journal*, 20(3), 302-311. doi:10.22074/cellj.2018.5442.
- Taft, R. J., Pheasant, M., & Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays*, 29(3), 288–299. doi:10.1002/bies.20544.
- Tang, J.-Y., Lee, J.-C., Chang, Y.-T., Hou, M.-F., Huang, H.-W., Liaw, C.-C., & Chang, H.-W. (2013). Long Noncoding RNAs-Related Diseases, Cancers, and Drugs. *The Scientific World Journal*, 2013, 1–7.
- Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology*, 19(1A): A68-77. doi:10.5114/wo.2014.47136.
- Tsutsui, S., Ohno, S., Murakami, S., Kataoka, A., Kinoshita, J., & Hachitanda, Y. (2003). Prognostic significance of the coexpression of p53 protein and c-erbB2 in breast cancer. *American journal of surgery*, 185(2), 165–167.
- Tran, B., & Bedard, P. L. (2011). Luminal-B breast cancer and novel therapeutic targets. *Breast Cancer Research*, 13(6). doi:10.1186/bcr2904.
- Tseng, Y.-Y., Moriarity, B. S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Bagchi, A. *et al.* (2014). PVT1 dependence in cancer with MYC copy-number increase. *Nature*, 512(7512), 82–86.
- Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: Genomics, Evolution, and Mechanisms. *Cell*, 154(1), 26–46. doi:10.1016/j.cell.2013.06.020.
- Valle-Solís A., Miranda-Aguirre, A., Cárdenas-Cárdenas, E. *et al.* (2019). Supervivencia en cáncer de mama por subtipo mediante inmunohistoquímica: Un estudio retrospectivo. *Gaceta Médica de México*, 155(1), 50-55.
- Van 't Veer, L. J., Dai, H., Van de Vijver, M. J., Mao, M., Friend, S. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871), 530-536. doi:10.1038/415530a.
- Vishnubalaji, R., Shaath, H., Elkord, E. & Alajez, N. (2019) Long non-coding RNA (lncRNA) transcriptional landscape in breast cancer identifies LINC01614 as non-favorable prognostic biomarker regulated by TGFβ and focal adhesion kinase (FAK) signaling. *Cell Death Discovery*. doi:10.1038/s41420-019-0190-6.
- Wang, K. C., Yang, Y. W., Liu, B., Chang, H. Y. *et al.* (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 472(7341), 120–124. doi:10.1038/nature09819.
- Wang, R. X., Chen, S., Jin, X., & Shao, Z. M. (2016). Value of Ki-67 expression in triple-negative breast cancer before and after neoadjuvant chemotherapy with weekly paclitaxel plus carboplatin. *Scientific reports*, 6, 30.
- Wang, C., Wang, L. *et al.* (2017). LncRNA Structural Characteristics in Epigenetic Regulation. *International journal of molecular sciences*, 18(12), 2659. doi:10.3390/ijms18122659.
- Wang, Y.-H., Ji, J., Wang, B.-C., Chen, H., Yang, Z.-H., Wang, K., Zhang, X.-L. *et al.* (2018). Tumor-Derived Exosomal Long Noncoding RNAs as Promising Diagnostic Biomarkers for Prostate Cancer. *Cellular Physiology and Biochemistry*, 46(2), 532–545. doi:10.1159/000488620.
- Wang, Y., Song, B., Zhu, L. & Zhang, X. (2019) Long non-coding RNA, LINC01614 as a potential biomarker for prognostic prediction in breast cancer. *PeerJ*, 7:e7976. doi:10.7717/peerj.7976.
- Wang, H., Wu, J., & Guo, W. (2020). SP1-Mediated Upregulation of lncRNA LINC01614 Functions as a ceRNA for miR-383 to Facilitate Glioma Progression Through Regulation of ADAM12. *Oncotargets and Therapy*, Volume 13, 4305–4318. doi:10.2147/ott.s242854.

- Washbrook, E. (2006). Risk factors and epidemiology of breast cancer. *Women's Health Medicine*, 3(1), 8–14. doi:10.1383/wohm.2006.3.1.8.
- Weigelt, B., Mackay, A. *et al.* (2010). Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *The Lancet Oncology*, 11(4), 339–349. doi:10.1016/s1470-2045(10)70008-5.
- World Health Organization (1993). Biomarkers and risk assessment: concepts and principles. *Environmental health criteria*, 1551.
- Willingham, A., Deveraux, Q., Hampton, G. & Aza-Blanc, P. (2004). RNAi and HTS: exploring cancer by systematic loss-of-function. *Oncogene*, 23(51), 8392–8400. doi:10.1038/sj.onc.1208217.
- Wilusz, J. E., JnBaptiste, C. K., Lu, L. Y., Kuhn, C.-D., Joshua-Tor, L., & Sharp, P. A. (2012). A triple hélix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes & Development*, 26(21), 2392–2407.
- Wu, Y.-H., Chang, T.-H. *et al.* (2013). *COL11A1* promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene*, 33(26), 3432–3440. doi:10.1038/onc.2013.307.
- Wu, Q., Li, J. *et al.* (2017). Breast cancer subtypes predict the preferential site of distant metastases: a SEER based study. *Oncotarget*, 8(17), 27990–27996. doi:10.18632/oncotarget.15856.
- Yang, C. H., Pfeffer, S. R., Sims, M., Yue, J., Wang, Y., Linga, V. G., Pfeffer, L. M. *et al.* (2015). The Oncogenic MicroRNA-21 Inhibits the Tumor Suppressive Activity of FBXO11 to Promote Tumorigenesis. *Journal of Biological Chemistry*, 290(10), 6037–6046. doi:10.1074/jbc.m114.632125.
- Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G., & Chen, L.-L. (2011). Genomewide characterization of non-polyadenylated RNAs. *Genome Biology*, 12(2), R16. doi:10.1186/gb-2011-12-2-r16.
- Yates, L. & Desmedt, C. (2017). Translational Genomics: Practical Applications of the Genomic Revolution in Breast Cancer. *Clinical Cancer Research*, 23(11), 2630–2639. doi:10.1158/1078-0432.ccr-16-2548.
- Yersal, O. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World Journal of Clinical Oncology*, 5(3), 412. doi:10.5306/wjco.v5.i3.412.
- You, K., Su, F., Liu, L., Lv, X., Zhang, J., Zhang, Y., & Liu, B. (2017). SCARA5 plays a critical role in the progression and metastasis of breast cancer by inactivating the ERK1/2, STAT3, and AKT signaling pathways. *Molecular and Cellular Biochemistry*, 435(1-2), 47–58. doi:10.1007/s11010-017-3055-4.
- Zepeda-Castilla, E., Recinos-Money, E., Cuéllar-Hubbe, M., Robles-Vidal, C., Maafs-Molina, E. (2008). Clasificación molecular del cáncer de mama. *Cirugía y cirujanos*, 76(1), 87-93.
- Zhai, S., Xu, H. *et al.* (2016). Expression Depression of CD300LG-γ in Human Pulmonary Carcinoma. *Monoclonal Antibodies in Immunodiagnosis and Immunotherapy*, 35(2), 94–99. doi:10.1089/mab.2015.0043.
- Zhang, J., Han, J. *et al.* (2020). Inwardly rectifying potassium channel 5.1: Structure, function, and possible roles in diseases. *Genes & Diseases*. doi:10.1016/j.gendis.2020.03.006.
- Zhang, J., Tsoi, H. *et al.* (2015). Carbonic anhydrase IV inhibits colon cancer development by inhibiting the Wnt signalling pathway through targeting the WTAP–WT1–TBL1 axis. *Gut*, 65(9), 1482–1493.
- Zhang, P., Meng, L., Song, L., Du, J., Du, S., Cui, W., Liu, C., & Li, F. (2018). Roles of Perilipins in Diseases and Cancers. *Current genomics*, 19(4), 247–257. doi:10.2174/1389202918666170915155948.
- Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., & Shen, F. (2019). Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. *International Journal of Molecular Sciences*, 20(22), 5573. doi:10.3390/ijms20225573.
- Zhang, Z., Zhu, Z., Watabe, K., Zhang, X., Bai, C., Xu, M., Mo, Y.-Y. *et al.* (2013). Negative regulation of lncRNA GAS5 by miR-21. *Cell Death & Differentiation*, 20(11), 1558–1568. doi:10.1038/cdd.2013.110.
- Zhang, Z., Lee, J.-H., Ruan, H., Ye, Y., Krakowiak, J., Hu, Q., Han, L. (2019). Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nature Communications*, 10(1). doi:10.1038/s41467-019-12543-5.
- Zhao, B., Xu, H., Ai, X., Adalat, Y., Tong, Y., Zhang, J., & Yang, S. (2018). Expression profiles of long noncoding RNAs in lung adenocarcinoma. *OncoTargets and therapy*, 11, 5383–5390. doi:10.2147/OTT.S167633.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13), 3406-3415.