UNIVERSIDAD TÉCNICA DE AMBATO

FACULTAD DE CIENCIA E INGENIERÍA EN ALIMENTOS Y BIOTECNOLOGÍA

CARRERA DE BIOTECNOLOGÍA

Tema: Evaluación *in silico* de la influencia de la heterogeneidad tumoral en la evolución clonal del cáncer de mama

---

Informe Final del Trabajo de Titulación, opción Proyecto de Investigación, previo a la obtención del Título de Ingeniera Biotecnóloga, otorgado por la Universidad Técnica de Ambato, a través de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

Autor: Samantha Lizbeth Caicedo Albán

Tutor: Mg. Cristian Fernando Galarza Galarza

Ambato – Ecuador

Febrero – 2025

APROBACIÓN DEL TUTOR

Mg. Cristian Fernando Galarza Galarza

CERTIFICA:

Que el presente Informe Final del Trabajo de Titulación ha sido prolijamente revisado. Por lo tanto, autorizo la presentación de este Informe Final del Trabajo de Titulación, opción Proyecto de Investigación el mismo que responde a las normas establecidas en el reglamento de Títulos y Grados de la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología.

Ambato, 20 de enero de 2025



Mg. Cristian Fernando Galarza Galarza

1803160272

TUTOR

AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Samantha Lizbeth Caicedo Albán, manifiesto que los resultados obtenidos en el presente Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación, previo a la obtención del título de Ingeniera Biotecnóloga, son absolutamente originales, auténticos y personales; a excepción de las citas bibliográficas.



Samantha Lizbeth Caicedo Albán

1805781182

AUTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que haga de este Informe Final del Trabajo de Titulación o parte de él, un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos en línea patrimoniales de mi Informe Final del Trabajo de Titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor.



Samantha Lizbeth Caicedo Albán

1805781182

AUTOR

APROBACIÓN DEL TRIBUNAL DE GRADO

Los suscritos Profesores Calificadores, aprueban el presente Informe Final del Trabajo de Titulación, Opción Proyecto de Investigación, el mismo que ha sido elaborado de conformidad con las disposiciones emitidas por la Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología de la Universidad Técnica de Ambato.

Para constancia firman:

Firmado electrónicamente por:
ESTEBAN MAURICIO FUENTES PEREZ

Presidente del Tribunal

Firmado electrónicamente por:
HELENA MARITZA DE LA TORRE OLVERA

Dra. Helena Maritza De la Torre Olvera
1309651998

Firmado electrónicamente por:
DANIEL ALFONSO CABRERA VALLE

Mg. Daniel Alfonso Cabrera Valle
1802561595

Ambato, 3 de Febrero del 2025

ÍNDICE GENERAL DE CONTENIDOS

APROBACIÓN DEL TUTOR ..... **II**

AUTORÍA DEL TRABAJO DE TITULACIÓN ..... **III**

DERECHOS DE AUTOR ..... **IV**

APROBACIÓN DEL TRIBUNAL DE GRADO ..... **V**

ÍNDICE GENERAL DE CONTENIDOS ..... **VI**

ÍNDICE DE TABLAS ..... **VIII**

ÍNDICE DE FIGURAS ..... **IX**

RESUMEN EJECUTIVO ..... **X**

ABSTRACT ..... **XI**

CAPÍTULO I.- MARCO TEÓRICO **1**

1.1. ANTECEDENTES INVESTIGATIVOS 1

1.1.1. Justificación 1

1.1.2. Cáncer de mama 2

1.1.2.1. Clasificación del cáncer de mama 2

1.1.2.2. Epidemiología 4

1.1.2.3. Factores de riesgo 4

1.1.2.4. Diagnóstico y tratamientos contra el cáncer de mama 5

1.1.2.5. Metástasis 6

1.1.3. Impacto de la heterogeneidad tumoral en la evolución clonal 8

1.1.3.1. Heterogeneidad tumoral en el cáncer de mama 8

1.1.3.2. Modelo de la Evolución clonal 8

1.1.3.3. Hipótesis de las células madre cancerosas (CMC) 9

1.1.4. Tecnologías aplicadas en la evaluación de la heterogeneidad tumoral 9

1.1.4.1. Secuenciación de próxima generación 9

1.1.4.2. Modelado in silico 10

1.2. OBJETIVOS 13

1.2.1. Objetivo General 13

1.2.2. Objetivos Específicos 13

CAPÍTULO II.- METODOLOGÍA **14**

2.1. MATERIALES 14

2.2. MÉTODOS 15

2.2.1. Identificación de la variabilidad genética asociada a la heterogeneidad tumoral en el cáncer de mama en diferentes muestras. 15
2.2.1.1. Selección de bibliografía y búsqueda de conjuntos de datos génicos relacionados a la enfermedad. 15
2.2.1.2. Control de calidad y procesamiento previo de datos genéticos. 16
2.2.1.3. Análisis de Expresión Diferencial (DEG). 16
2.2.2. Identificación de las funciones biológicas que se degradan en diferentes estadios de cáncer de mama. 17
2.2.2.1. Análisis funcional de genes diferencialmente expresados en el cáncer de mama. 17
2.2.2.2. Construcción de redes de interacción y coexpresión génica. 17
2.2.3. Evaluación de la influencia de los tipos de tumores en diferentes estadios del cáncer de mama sobre diferentes poblaciones. 18
2.2.3.1. Análisis de variantes genéticas de la heterogeneidad tumoral. 18
2.2.3.2. Evaluación de la influencia de las variantes genéticas de genes diferencialmente expresados en la evolución clonal del cáncer de mama. 18
CAPÍTULO III.- RESULTADOS Y DISCUSIÓN **19**
3.1. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS 19
3.1.1. Identificación de la variabilidad genética asociada a la heterogeneidad tumoral en el cáncer de mama. 19
3.1.2. Examinación de las funciones biológicas afectadas en el desarrollo del CM. 23
3.1.3. Evaluación de la influencia de los tipos de tumores en diferentes estadios del cáncer de mama. 35
CAPÍTULO IV.- CONCLUSIONES Y RECOMENDACIONES **39**
4.1. CONCLUSIONES 39
4.2. RECOMENDACIONES 40
REFERENCIAS BIBLIOGRÁFICAS **41**
ANEXOS **56**

ÍNDICE DE TABLAS

Tabla 1. Recursos bioinformáticos. 14

Tabla 2. Equipos. 14

Tabla 3. Muestras seleccionadas. 15

Tabla 4. Datos de tumores obtenidos de pacientes con cáncer de mama. 19

Tabla 5. Genes diferencialmente expresados. 21

Tabla 6. Identificación funcional de los genes diferencialmente expresados. 24

Tabla 7. Identificación de genes presentes en las redes de interacción. 29

Tabla 8. Identificación de genes exclusivos. 35

Tabla 9. Identificación de variantes. 37

ÍNDICE DE FIGURAS

Figura 1. Clasificación tradicional del cáncer de mama. 2

Figura 2. Características de los subtipos moleculares del cáncer de mama. 3

Figura 3. Papel de EMT en la metástasis tumoral. 6

Figura 4. Control de calidad de los datos. 20

Figura 5. Representación de los genes diferencialmente expresados. 23

Figura 6. Redes de interacción de los genes diferencialmente expresados correspondientes al grupo TP y GL. 26

Figura 7. Rutas biológicas compartidas entre los genes diferencialmente expresados. 27

Figura 8. Identificación de mecanismos moleculares compartidos por los genes relacionados a los genes diferencialmente expresados identificados en TP. 32

Figura 9. Identificación de mecanismos moleculares compartidos por los genes relacionados a los genes diferencialmente expresados identificados en ganglios linfáticos. 34

RESUMEN EJECUTIVO

En el presente estudio se analizó la influencia de la heterogeneidad tumoral sobre la evolución clonal en el cáncer de mama (CM) puesto que ha sido una de las principales causas por las que los tratamientos para el control de la enfermedad han fracasado. Por esta razón, este proyecto busca comprender la influencia de la heterogeneidad tumoral en la progresión del cáncer y desarrollo de la metástasis planteando un punto de partida para un tratamiento personalizado. Se trabajaron con datos pertenecientes al cáncer de mama, con secuencias de muestras pertenecientes a tumores primarios y tumores en los ganglios linfáticos. Se aplicaron técnicas bioinformáticas para su procesamiento, se empleó un análisis de expresión diferencial para identificar genes que se expresaban en cada una de las condiciones experimentales. Se representaron redes de interacción para identificar las diferentes interacciones de los genes élite con otros genes, se validó la funcionalidad de estos utilizando la técnica de enriquecimiento funcional para identificar el rol que cumplen estos genes en la integridad y homeostasis celular. Con los genes seleccionados se realizó una búsqueda de variantes para identificar la influencia que tienen estos genes en el desarrollo y progresión de la metástasis. Se evalúo la heterogeneidad tumoral sobre la evolución clonal analizando de la influencia de las variantes de cada tipo de tumor, observando que GL tiene mayor incidencia sobre la metástasis en comparación con TP.

Palabras clave: Cáncer de mama (CM), metástasis, evolución clonal, heterogeneidad tumoral.

ABSTRACT

In the present study, the influence of tumor heterogeneity on clonal evolution in breast cancer (BC) was analyzed, since it has been one of the main reasons why treatments for controlling the disease have failed. For this reason, this project seeks to understand the influence of tumor heterogeneity on cancer progression and development of metastasis, proposing a starting point for personalized treatment. Data pertaining to breast cancer were worked with, with sequences of samples belonging to primary tumors and tumors in the lymph nodes. Bioinformatics techniques were applied for processing, and a differential expression analysis was used to identify genes that were expressed in each of the experimental conditions. Interaction networks were represented to identify the different interactions of elite genes with other genes, and the functionality of these genes was validated using the functional enrichment technique to identify the role that these genes play in cell integrity and homeostasis. A search for variants was performed with the selected genes to identify the influence that these genes have on the development and progression of metastasis. Tumor heterogeneity on clonal evolution was evaluated by analyzing the influence of the variants of each type of tumor, observing that GL has a higher incidence on metastasis compared to TP.

Keywords: Breast cancer (BC), metastasis, clonal evolution, tumor heterogeneity.

CAPÍTULO I.- MARCO TEÓRICO

1.1. Antecedentes investigativos

1.1.1. Justificación

El cáncer de mama se ha convertido en una neoplasia maligna con mayor incidencia y una de las principales causas de muerte entre las mujeres de todo el mundo **(Yin et al., 2024).** En los países en vías de desarrollo alrededor de 2.3 millones de mujeres padecen esta enfermedad **(Arnold et al., 2022).** A pesar de los esfuerzos realizados enfocados en su tratamiento y prevención, las evidencias encontradas no son suficientes para comprender la etiología y progresión de la enfermedad.

En Ecuador aún se carece de recursos para un diagnóstico y tratamiento adecuado, por esta razón, en el año 2020 se han reportado 38,2 casos por cada 100,000 habitantes con un porcentaje de mortalidad del 5,1%, por lo que se sitúa en el puesto 161 a nivel mundial en letalidad **(Loor & Recalde, 2024).**

Dentro de los principales problemas del cáncer de mama, la recurrencia metastásica es una de las más importantes, debido a que los diferentes tipos de tumores poseen tropismos orgánicos distintos, los cuales se diseminan a diferentes órganos del cuerpo, tales como huesos, pulmones, hígado, ganglios linfáticos y cerebro **(Riggio et al., 2020).** Esto es la base de la aparición de la heterogeneidad intratumoral, la cual suele originarse en mutaciones genéticas, modificaciones epigenéticas y microambientales. Los cambios epigenéticos y microambientales son adaptativos, mientras que las mutaciones genéticas son permanentes **(Biswas & De, 2021).**

En la actualidad, han aparecido diferentes tecnologías que abren una brecha importante en el tratamiento de diversas enfermedades. La terapia génica emerge como una herramienta para intentar caracterizar esta heterogeneidad tumoral. Las técnicas de secuenciación de próxima generación (NGS) han demostrado ser las más idóneas para este tipo de análisis, ya que permiten determina perfiles de expresión génica con mayor precisión y eficiencia **(Kim et al., 2020).** Las secuencias obtenidas a partir de NGS se utilizan como materia prima para implementar la medicina personalizada. Uno de sus enfoques, busca identificar el espectro único de mutaciones que da lugar a los tumores individuales y a las subpoblaciones tumorales en pacientes con la enfermedad.

El presente proyecto tiene como finalidad evaluar los factores genéticos que influyen en la variabilidad de tumores presentes en el cáncer de mama, ya que esta variabilidad puede repercutir en la progresión de la enfermedad. Por tal motivo, se buscará identificar la influencia que tienen los tipos de tumores en dicha progresión, con el propósito de sentar las bases para la implementación de la medicina

personalizada en el tratamiento del cáncer de mama, como una alternativa para mejorar la efectividad de los tratamientos actualmente disponibles.

1.1.2. Cáncer de mama

El cáncer de mama tiene una historia muy larga, ya que fue informado por primera vez por los antiguos egipcios hace más de 3500 años, aproximadamente en el año 1500 a. C. Hoy en día, es el segundo tipo de cáncer más frecuente y es una de las principales causas de muertes relacionadas con el cáncer en mujeres del todo el mundo **(Brushan et al., 2021).** El cáncer es producido por daño genético en las células que exhiben división celular anormal y mutación. Es una neoplasia maligna relacionada con los cambios hormonales en el cuerpo ya que los tejidos glandulares de las mamas son susceptibles a dichos cambios. Las hormonas sexuales son las responsables de la formación de los canales mamarios, la expansión del tejido y los brotes, por lo que derivan factores de riesgo para desarrollar cáncer de mama **(Rasul et al., 2022).**

1.1.2.1. Clasificación del cáncer de mama

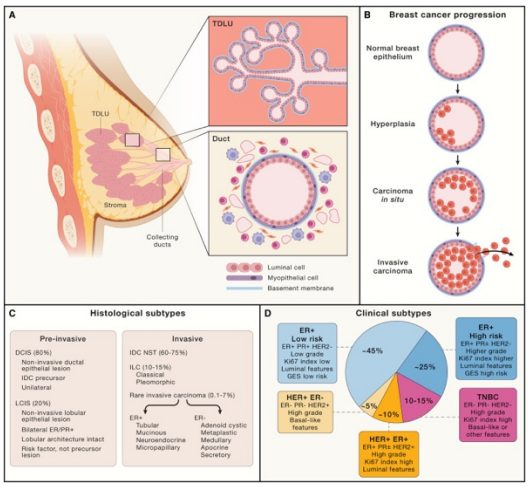


Figura 1. Clasificación tradicional del cáncer de mama.
Fuente: **(Nolan et al., 2023)**

1.1.2.1.1. Clasificación histopatológica

Los tumores se clasifican como carcinomas preinvasivos (*in situ*) y carcinomas invasivos, dependiendo de la propagación de células malignas desde los lóbulos o conductos mamarios al estroma circundante. El cáncer preinvasivo permanece en una ubicación particular de la mama sin propagarse a tejido mamario circundante, además, solo el 10% al 30% de los casos progresan a cáncer invasivo. Se clasifica

en carcinoma ductal in situ (CDIS) y carcinoma lobulillar in situ (LCIS). CDIS es una forma preinvasiva de cáncer de mama donde el tumor surge de las células epiteliales que cubren el conducto galactóforo. La masa celular crece dentro del conducto, pero con el tiempo puede atravesar sus paredes e ir a tejido circundante. LCIS se origina de células adenomatosas dentro de los lóbulos mamarios productores de leche y permanecen dentro de los lóbulos. Es detectable únicamente por mamografía o biopsia. Los carcinomas invasivos crecen en tejido mamario circundante, se propagan a los ganglios linfáticos y a otros órganos. Se clasifican en carcinoma ductal invasivo (IDC) y carcinoma lobulillar invasivo (ILC). IDC representa entre el 60% al 75% de los casos siendo el más común. ILC está presente del 10% al 15% de los tumores. Ambos difieren en su patología, perfiles genómicos, organotropismo metastásico y respuesta al tratamiento **(Beňačka et al., 2022; Nolan et al., 2023).**

1.1.2.1.2. Clasificación molecular

El cáncer de mama posee cuatro subtipos moleculares clínicamente relevantes: Luminal A, Luminal B, HER2 y Triple Negativo. Los genes responsables de la segregación de los subtipos moleculares están relacionados con la expresión de los receptores de estrógenos (ER), receptores de progesterona (PR), HER2 (Receptor 2 del factor de crecimiento epidérmico humano) y los reguladores de la proliferación celular (Ki-67) **(Nascimento & Otoni, 2020).** Luminal A (HER2- y HR+) presenta niveles bajos de la proteína K-67, por lo que tiende a proliferar lentamente con un pronóstico favorable. Luminal B (HR+ y HER2+/HER2-) tiene niveles elevados de la proteína Ki-67 por lo que su pronóstico no es tan favorable. HER2 (HER2+ y HR-) prolifera a un ritmo más rápido que los cánceres luminales. El cáncer de mama triple Negativo (HR- y HER2-) tiene mayor incidencia en mujeres con mutación en el gen BRCA1 **(Brushan et al., 2021)**.

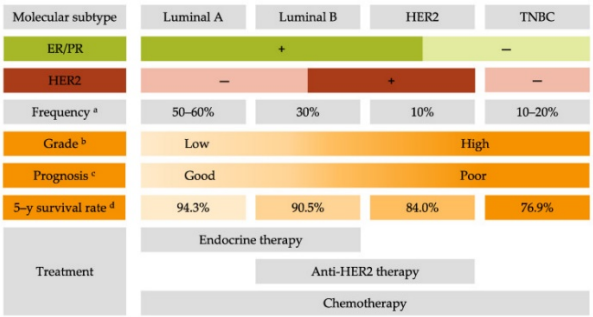


Figura 2. Características de los subtipos moleculares del cáncer de mama.
Fuente**: (Burguin et al., 2021)**

1.1.2.2. Epidemiología

La incidencia del cáncer de mama aumenta conforme al índice de desarrollo de un país, junto con la disponibilidad y utilización de pruebas de detección. Desde 1997, países con alto índice de desarrollo incluyendo a Estados Unidos, Europa y Oceanía, han tenido un aumento anual de la tasa de incidencia de cáncer de mama de menos del 0,5 %. Durante el mismo período, las regiones con un índice de desarrollo humano medio y bajo, como Sudamérica y África, han tenido una mayor tasa de aumento, aproximadamente entre el 3 % y 5 % respectivamente. Los países asiáticos, como Corea y Japón, aunque se consideran que tienen un alto índice de desarrollo humano, también han experimentado un rápido aumento de la tasa de incidencia **(Arzanova & Mayrovitz, 2022).**

Desde la creación del Registro Nacional de Tumores (RNT), Ecuador ha dado seguimiento a la evolución de la epidemiología de los casos de cáncer, reportando entre 1985 y 2013 una incidencia de CM de 25,4 a 38,8 por 100,000 habitantes, representando un porcentaje de cambio anual de 1,9. Esto se explica en gran medida a la inclusión de la mujer en la economía nacional, conllevando a un aumento de la ventana de riesgo estrogénico y a su vez en embarazos tardíos, menos embarazos, lactancia reducida o sobrecarga calórica **(Cedeño et al., 2024).**

1.1.2.3. Factores de riesgo

1.1.2.3.1. Factores genéticos

Los factores genéticos se manifiestan principalmente por predisposición genética en los genes BRCA1/2. Los cánceres de mama asociados a BRCA1 se presentan con un fenotipo triple negativo, mientras que los casos ER-positivos, HER2-negativos se manifiesta en portadores de mutaciones BRCA2 y PALB2 **(Hong & Xu, 2022).** BRCA1 y BRCA2 participan en la reparación de roturas de la doble hebra de ADN mediante recombinación homóloga. Las mutaciones hereditarias en estos genes representan aproximadamente el 2,5% de todos los cánceres de mama en mujeres con una historia familiar de la enfermedad, y contienen alrededor de un 70% de riesgo de cáncer de mama para mujeres con 80 años de edad. Otros genes de predisposición al cáncer de mama, de penetrancia alta y moderada, incluyen CDH1, PTEN, LKB1, TP53, CHEK2, ATM, NBN y pareja localizadora de BRCA2 (PALB2); sin embargo, las mutaciones en sus respectivas líneas germinales son poco comunes **(Britt et al., 2020).**

1.1.2.3.2. Factores exógenos

Se conocen varios factores de riesgo que favorecen al desarrollo del cáncer de mama, siendo los más importantes el género, la edad y el grado de desarrollo económico de un determinado país. También se

consideran los factores hormonales, relacionados con el tiempo de exposición a los estrógenos y los factores procreativos (número de hijos nacidos, lactancia materna, etc.). De igual manera, se incluye la terapia de reemplazo hormonal, la dieta inadecuada y la obesidad resultante. En una edad temprana, se atribuyen factores como la anticoncepción hormonal, el consumo de alcohol y la exposición a radiaciones ionizantes **(Smolarz et al., 2022).** La edad avanzada es otro factor importante a considerar ya que el 80% de las pacientes con cáncer de mama son mayores de 50 años, influye debido a la acumulación de alteraciones celulares y a la exposición a carcinógenos potenciales. Para prevenir el cáncer es crucial hacer actividad física para reducir la exposición a hormonas sexuales endógenas y evitar la obesidad ya que esta conlleva a peores resultados clínicos porque empeora el estado inflamatorio y afecta los niveles de hormonas circundantes **(Lukasiewicz et al., 2021).**

1.1.2.4. Diagnóstico y tratamientos contra el cáncer de mama

La obtención de imágenes de mama se utiliza exclusivamente para la detección, el diagnóstico y el tratamiento clínico de los cánceres. Su enfoque es la detección temprana de la enfermedad y la evaluación precisa de las lesiones. Como modalidad de su obtención se utiliza la ecografía, la biopsia guiada por imágenes y el diagnóstico de los ganglios linfáticos. Dependiendo de la extensión del cáncer de mama, se determina la necesidad de terapia sistémica preoperatoria (neoadyuvante). Actualmente se necesitan terapias dirigidas y efectivas con efectos secundarios mínimos fuera del objetivo para el tratamiento del cáncer de mama **(Brushan et al., 2021).** Los tratamientos comunes para todos los subtipos de cáncer de mama son la cirugía para la escisión total de la mama o la extirpación de los ganglios linfáticos con el fin de determinar la propagación de células cancerosas. Por otro lado, la radioterapia ofrece una disminución del riesgo de recurrencia y de mortalidad, y puede ser un complemento de refuerzo. Sin embargo, presenta efectos secundarios como la cardiotoxicidad. La quimioterapia es la más común y puede ser administrada de forma neoadyuvante (quimioterapia administrada antes de cirugía) o adyuvante (quimioterapia administrada después de cirugía). Consiste en la administración de varias familiar de fármacos citotóxicos con el fin de causar roturas en el ADN, detención del ciclo celular, inhibición de biosíntesis, entre otros mecanismos **(Burguin et al., 2021).**

Las terapias más efectivas para el cáncer de mama son aquellas que se dirigen específicamente a los receptores ER y HER2 ya que el curso de esta enfermedad esta influenciado en gran medida por estrógeno y ER. Este tipo de terapias consiste en inhibir objetivos que cumplen funciones en varias vías de carcinogénesis para detener la progresión del cáncer **(Ye et al., 2023).** Los inhibidores de la quinasa dependiente de ciclina CDK4 y CDK6, los inhibidores de PI3K, los inhibidores de PARP y la inmunoterapia anti-PD-L1 son ejemplos de opciones de tratamiento para el cáncer de mama metastásico. Para los cánceres HR se utiliza la terapia endócrina (moduladores selectivos del ER, los desreguladores selectivos de ER y los inhibidores de aromatasa) la cual busca inhibir la síntesis de estrógenos,

degradarlos o inactivarlos. TNBC actualmente no tiene terapia dirigida aprobada debido a que es cáncer con alta proliferación celular **(Guo et al., 2023).** Por otro lado, existen dos terapias efectivas dirigidas a HER2 las cuales incluyen anticuerpos monoclonales e inhibidores de la tirosina quinasa. Los anticuerpos monoclonales se dirigen a las regiones de unión del receptor HER2 bloqueando su señalización descendente. Los inhibidores de la tirosina quinasa son moléculas pequeñas que bloquean las regiones de unión de ATP del receptor de HER2 previniendo la fosforilación y a su vez, la señalización descendente **(Lau et al., 2022).**

1.1.2.5. Metástasis

La metástasis es la principal causa de mortalidad en la mayoría de los cánceres. Predecir el cáncer de mama antes de la metástasis puede aumentar las tasas de supervivencia, sin embargo, a menudo esta enfermedad es diagnosticada después de que se haya dado la metástasis, razón por la que el cáncer de mama tiene un mal pronóstico debido a que puede propagarse a los ganglios linfáticos u órganos distantes, siendo esto un impedimento para su tratamiento **(Park et al., 2022).** La metástasis es un proceso selectivo de órganos de varios pasos que inicia con el escape de células tumorales del tumor primario y termina con la colonización de tumores secundarios en los sitios distantes. Los pasos en la metástasis son: escape de células cancerosas del tumor primario, invasión, mantenimiento de la supervivencia, extravasación (siembra de sitio secundario) y crecimiento (colonización) **(Majidpoor & Mortezaee, 2021).**

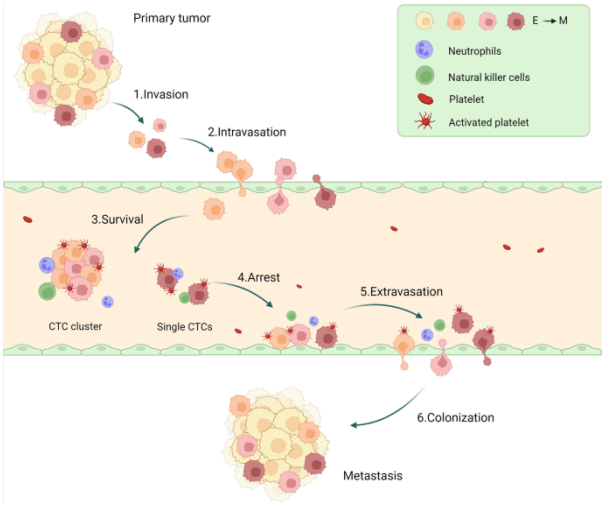


Figura 3. Papel de EMT en la metástasis tumoral.
Fuente: **(Huang et al., 2022)**

1.1.2.5.1. Escape de células cancerosas del tumor primario/invasión

El primer paso de la metástasis es cuando las células tumorales se desprenden del lecho tumoral y migran del estroma al torrente sanguíneo. Para ello, las células tumorales deben reducir su adherencia celular a través de transiciones epiteliales-mesenquimales (EMT). La EMT consiste en la pérdida de rasgos epiteliales (polaridad celular y uniones célula-célula) y la adquisición de características mesenquimales (morfología fibroblástica en forma de uso) con el fin de aumentar la movilidad celular hacia vasos sanguíneos intratumorales, caracterizados por una mayor permeabilidad, permitiendo que las células cancerosas ingresen fácilmente a la circulación sistémica **(Wang et al., 2021).** Durante la diseminación, las células tumorales que albergan mutaciones impulsoras oncogénicas invaden a través de la membrana basal hacia capaz tisulares más profundas, seguido por la invasión en los vasos sanguíneos proximales o linfáticos distribuyéndose a otros tejidos de órganos distales **(Gerstberger et al., 2023).**

1.1.2.5.2. Mantenimiento de la supervivencia

Tanto las células que se encuentran en circulación como las células que residen por primera vez dentro del sitio o sitios secundarios de metástasis se encuentran continuamente bajo la exposición continua de señales destructivas. En circulación, las células tumorales circulantes (CTC) se enfrentan a factores de estrés ambiental, estrés oxidativo, fuerzas de corte y un asalto impuesto por el sistema inmunológico **(Majidpoor & Mortezaee, 2021).** Como mecanismo de defensa, las células tumorales adoptan una variedad de vías dependientes e independientes de las células huésped para defenderse durante el tiempo limitado de la fase circulatoria **(Khan et al., 2022).** Posteriormente, solo un poco de células permanecen vivas y dan el siguiente paso.

1.1.2.5.3. Extravasación

Al salir del sistema circulatorio, las células cancerosas se adhieren al endotelio vascular en un sitio secundario metastásico distante y abandonan la circulación pasando de ser células tumorales circulantes (CTC) a células tumorales diseminativas (DTC). A pesar de ser un número pequeño de DTC, son altamente competentes para tomar un destino metastásico efectivo, siendo los sitios con mayor riesgo de metástasis el hueso (65%), el pulmón (31%), el hígado (26%) y el cerebro (8,8%) (**Giordano et al., 2020; Majidpoor & Mortezaee, 2021).**

1.1.2.5.4. Crecimiento

Para lograr el desarrollo de tumores metastásicos, los CTC encuentran un microambiente receptivo en órganos objetivo. La especificidad del microambiente de nicho premetastásico es fundamental para el

destino del crecimiento de CTC **(Giordano et al., 2020).** Las células tumorales eventualmente colonizan a través de transiciones mesenquimales-epiteliales (MET), donde las células metastásicas recuperan características epiteliales generando tumores secundarios distales, histopatológicamente similares al tumor primario **(Huang et al., 2022).**

1.1.3. Impacto de la heterogeneidad tumoral en la evolución clonal

1.1.3.1. Heterogeneidad tumoral en el cáncer de mama

El cáncer de mama es una enfermedad muy heterogénea ya que es causada por una variedad de alteraciones genéticas distintas en las células epiteliales mamarias, por lo que la enfermedad se manifiesta de diferente manera en pacientes individuales. Estas diferencias entre pacientes, que constituye la heterogeneidad intertumoral, afectan el pronóstico del paciente y las opciones de tratamiento. Además, existe una enorme diversidad en las subpoblaciones de células tumorales dentro de un tumor; esto se conoce como heterogeneidad intratumoral. Ambos tipos de heterogeneidad se dan como consecuencia de diferentes parámetros intrínsecos de las células cancerosas, como el perfil genético, la interacción entre el genoma, transcriptoma y proteoma, las capacidades de migración e invasión, la ploriferación, la pluripotencialidad y la plasticidad celular **(Lüönd et al., 2021).**

1.1.3.2. Modelo de la Evolución clonal

Peter Nowell en 1976 describió el concepto de evolución clonal, donde los subclones de células con una ventaja de crecimiento en un determinado ecosistema persistirán, sobrevivirán y finalmente se expandirán, mientras que los subclones menos aptos no podrán competir con los demás y se extinguirán **(Lüönd et al., 2021).**

El modelo de la evolución clonal afirma que las células cancerosas con el tiempo adquieren varias combinaciones de mutaciones dentro de un tumor y que la deriva genética y la selección natural gradual de las células más aptas y agresivas impulsan la progresión del tumor. Menciona que el inicio del tumor tiene lugar una vez que se producen múltiples mutaciones en una única célula aleatoria, proporcionándole ventaja de crecimiento selectivo sobre las células normales adyacentes. A medida que el tumor progresa, la inestabilidad genética y la proliferación descontrolada permiten la producción de células con mutaciones adicionales y, por tanto, nuevas características que les proporcionan una ventaja de crecimiento sobre otras células tumorales **(Campbell & Polyak, 2007).**

1.1.3.3. Hipótesis de las células madre cancerosas (CMC)

Establece que un subconjunto particular de células tumorales con propiedades similares a las de las células madre, llamadas "células madre cancerosas", impulsan el inicio, la progresión y la recurrencia del tumor. Estas células tienen la capacidad de autorrenovarse indefinidamente y diferenciarse, como las células madre adultas normales. Su autorrenovación y diferenciación conducen a la producción de todos los tipos de células de un tumor, generando heterogeneidad tumoral. Además, menciona que la progresión del tumor es el resultado de la diseminación metastásica de estas células, y la recurrencia del cáncer es causada por su resistencia a la terapia **(Campbell & Polyak, 2007).**

La heterogeneidad del cáncer de mama se atribuye a la compleja jerarquía de diferenciación mamaria; es decir, existe una amplia diversidad de las células progenitoras mamarias normales debido a factores espaciales y temporales. Por ello, múltiples estados celulares pueden conferir un fenotipo de células madre cancerosas (CSC), impulsando la proliferación tumoral, la invasión, la metástasis, la resistencia a la terapia farmacológica y la enfermedad recurrente. Esto ocurre mediante una transición epitelial-mesenquimal, donde las células pasan de un estado epitelial diferenciado, caracterizado por la adhesión célula-célula y la falta de motilidad, a un estado mesenquimal, caracterizado por la motilidad y la invasividad. Las células tumorales metastásicas presentan un fenotipo epitelial con una transición mesenquimal **(Turner et al., 2021).**

1.1.4. Tecnologías aplicadas en la evaluación de la heterogeneidad tumoral

1.1.4.1. Secuenciación de próxima generación

La secuenciación masiva de ARN (RNA-seq) permite descubrir nuevas vías moleculares y genes expresados diferencialmente asociados con distintas etapas de la progresión del cáncer de mama. Sin embargo, la secuenciación de ARN de una sola célula (scRNA-seq) ofrece un aumento en la resolución celular, ya que, además de capturar la complejidad de la heterogeneidad celular, facilita la identificación de vías moleculares adicionales y biomarcadores específicos de las células **(Ionkina et al., 2021).** La scRNA-seq permite una comprensión más profunda de la diversidad de estados celulares y de la heterogeneidad de las poblaciones celulares, siendo una herramienta útil para evaluar los tipos de células dentro y alrededor de los tumores de mama. El desarrollo de la transcriptómica de células individuales revela los complejos mecanismos del desarrollo tumoral, la metástasis y la resistencia a los fármacos, contribuyendo a las terapias dirigidas **(Ding et al., 2020).**

1.1.4.2. Modelado *in silico*

Los modelos computacionales pueden simular dinámicas complejas impulsadas por vías de señalización intrincadas, con retroalimentaciones y diálogos cruzados. Son herramientas que permiten descifrar los mecanismos cooperativos en el contexto de la biología del cáncer. Explican los fenotipos de esta enfermedad a nivel molecular, integrando interacciones documentadas en la literatura y datos generados por tecnologías ómicas de alto rendimiento (como genómico, transcriptómico, etc.) **(Selvaggio et al., 2021).** Los métodos in silico son rentables y los resultados son más precisos, permitiendo analizar el mecanismo de acción relacionado con el cáncer y diseñar medicamentos con mayor eficiencia **(Shrihastini et al., 2021).**

1.1.4.3. Herramientas utilizadas en la evaluación de la heterogeneidad tumoral

Los rápidos y recientes avances en las estrategias de perfil molecular junto con las técnicas analíticas basadas en el desarrollo de la secuenciación de alto rendimiento y la espectrometría de masas (EM) han generado conjuntos de datos multiómicos a gran escala, que surgen principalmente de las tres grandes ómicas según el dogma central de la biología molecular: genómica, transcriptómica y proteómica **(Neagu et al., 2023).** Los repositorios de datos ómicos en línea asociados al cáncer desempeñan un papel fundamental en la comprensión de los procesos y mecanismos celulares asociados a esta enfermedad. Este recurso permite integrar diferentes conjuntos de datos de forma sistemática para generar y validar hipótesis de manera rápida. Ómnibus de expresión genética (GEO) es un recurso de datos de acceso abierto el cual alberga aproximadamente 291 conjuntos de datos de pacientes que abarcan 48 tipos de cáncer diferentes donde el cáncer de mama y el cáncer cerebral son las categorías predominantes con 39 y 25 proyectos respectivamente **(Das et al., 2020).**

FASTQC es una herramienta de control de calidad que involucra una serie de pasos de preprocesamiento con el fin de garantizar la limpieza y precisión de los mismos. El preprocesamiento consiste en eliminar contaminación proveniente de adaptadores, filtrar lecturas de baja calidad y corregir bases representadas de forma errónea. De este proceso dependió en gran medida los resultados del análisis final **(Chen, 2023).** FASTQC es catalogada como la más utilizada para el control de calidad de datos de secuenciación; sus informes de análisis constan de diez módulos donde se resume el contenido de un archivo de secuenciación **(Smith & De Sena, 2021).** Los errores detectados son corregidos con Trimmomatic, una herramienta de línea de comandos multiproceso que permite el recorte de secuencias con baja calidad y la eliminación de adaptadores de los datos FASTQ para las lecturas de extremos emparejados **(Maran & Davis, 2022).**

HISAT2 es uno de los programas de comparación de secuencias más populares gracias a su velocidad de procesamiento, precisión y sensibilidad **(Ibrahim et al., 2022).** Cuya función radica en determinar el lugar donde se alinearon pequeños fragmentos de secuencias con el genoma de referencia mediante su fragmentación en componentes más pequeños, proceso conocido como indexación. La indexación del genoma de referencia permitió que el alineador pueda encontrar de manera mas eficiente coincidencias exactas con lecturas secuenciadas en una estructura de datos dada y no escanear necesariamente todo el genoma de referencia para cada lectura **(Musich et al., 2021).**

SAMTools es una herramienta ejecutable que ofrece varios comandos para trabajar con datos de alineación. Es un programa ampliamente usado para procesar y analizar datos de secuenciación de alto rendimiento. Fue utilizada para la conversión y manipulación de formatos de archivo en formato SAM y BAM (**Danecek et al., 2021)**. El archivo de formato de alineación/mapa de secuencia (SAM) es un formato de archivo de texto donde se almacena información de la alineación de las lecturas al genoma de referencia. El formato SAM junto con herramientas relacionadas permitieron construir una estructura modular para el análisis de secuenciación del genoma. Por otro lado, la alineación/mapa binario (BAM) es el formato de archivo binario de SAM y se encuentra comprimido. Ordenar el contenido del archivo organizando contenidos similares y mejorando su comprensión **(Liu et al., 2023).**

FeatureCounts es una herramienta de conteo de genes que emplea archivos de lecturas alineadas en formato SAM o BAM y una lista de características genómicas en formato de características generales (GFF3) y el genoma de referencia con el que se realice los análisis. Las lecturas representan un fragmento de ARN, que son contados por dicha herramienta. Como resultado, genera una tabla con los nombres de las secuencias de referencia y sus características ordenadas de acuerdo a su posición de inicio **(Liao et al., 2014).**

R es un software con una curva de aprendizaje fácil debido por su funcionalidad de gráficos integrada. Su practicidad lo convierte en el primer lenguaje de programación ideal para aprender. Su funcionalidad radica en la interacción de datos, análisis y visualización de resultados **(Giorgi et al., 2022)**.

Cytoscape es una herramienta que permite la construcción de redes, un software de código abierto que permite integrar y visualizar redes de interacción molecular **(Yang et al., 2022).** Se puede integrar con otros complementos y aplicaciones, convirtiéndola en una plataforma flexible para análisis y visualización de redes complejas **(Mousavian et al., 2021)**. Por otro lado, Metascape permite el análisis de enriquecimiento funcional y de vías, permite analizar la clasificación biológica de genes **(Nan et al., 2021)**.

FreeBayes permite un preprocesamiento computacional con el fin de adaptar los datos de alineamiento de secuencias, permitiendo el análisis posterior por hebra para la identificación de variables. Tiene la función de llamada de variantes como un objetivo independiente para la facilidad del análisis, esto con el fin de optimizar la precisión y sensibilidad en el descubrimiento de alteraciones genéticas **(Nunn et al., 2022)**. En el análisis de variantes también se usa VCFfilter, se usa para filtrar variantes irrelevantes tomando en cuenta una secuencia de filtros predefinida, esta es aplicada a las variantes de entrada **(Eidi et al., 2024)**. SnpEff permite anotar variantes identificadas por muestra y estimar con una predicción simple de sus efectos funcionales, las cuales pueden ser comparadas entre muestras de análisis **(Kutschera et al., 2022)**.

1.2. Objetivos

1.2.1. Objetivo General

Evaluar *in silico* la influencia de la heterogeneidad tumoral en la evolución clonal del cáncer de mama.

1.2.2. Objetivos Específicos

- Identificar la variabilidad genética asociada a la heterogeneidad tumoral en el cáncer de mama en diferentes muestras.
- Examinar las funciones biológicas que se degradan en diferentes estadios de cáncer de mama.
- Evaluar la influencia de los tipos de tumores en diferentes estadios del cáncer de mama sobre diferentes poblaciones.

CAPÍTULO II.- METODOLOGÍA

2.1. Materiales

Este proyecto investigativo tuvo un enfoque *in silico*, por lo que se emplearon los siguientes materiales:

Tabla 1.

Recursos bioinformáticos.

Recursos	**Enlaces**
Centro Nacional de Información Biotecnológica (NCBI)	https://www.ncbi.nlm.nih.gov/
Catálogo de mutaciones somáticas en el cáncer (COSMIC)	https://cancer.sanger.ac.uk/cosmic
Metascape	http://metascape.org
Cytoscape	https://cytoscape.org/
R	https://cran.r-project.org/
Ensembl	https://grch37.ensembl.org/index.html
Galaxy	https://usegalaxy.org/
Cancer Genome Interpreter (CGI)	https://www.cancergenomeinterpreter.org/home

Tabla 2.

Equipos.

Recursos	**Cantidad**
Computadora portátil	1
Servicio de internet o Router	1

2.2. Métodos

2.2.1. Identificación de la variabilidad genética asociada a la heterogeneidad tumoral en el cáncer de mama en diferentes muestras.

2.2.1.1. Selección de bibliografía y búsqueda de conjuntos de datos génicos relacionados a la enfermedad.

Para la búsqueda de muestras referentes al cáncer de mama, se utilizó la base de datos Gene Expression Omnibus (GEO), donde se emplearon los siguientes criterios de búsqueda: Metastatic Breast Cancer, Next Generation Sequencing (NGS), y añadiendo secuencias que provengan de tecnologías ARN (ARN-Seq), se acotó la fecha de publicación a los últimos 5 años. Con estos criterios, se seleccionó el conjunto de datos identificado con el código GSE183947 "*Identification of five cytotoxicity-related genes involved in the progression of breast cancer*". Este proyecto posee las siguientes muestras, en las cuales se puede identificar el diseño experimental aplicado, se observa muestras de control, tumor primario y tumor en ganglios linfáticos.

Tabla 3.
Muestras seleccionadas.

SRA ID	Estado
SRR15852416	Tumor a los ganglios linfáticos
SRR15852394	Tumor a los ganglios linfáticos
SRR15852401	Tumor a los ganglios linfáticos
SRR15852399	Tumor primario
SRR15852407	Tumor primario
SRR15852413	Tumor primario
SRR15852443	Tejido mamario sano
SRR15852427	Tejido mamario sano
SRR15852426	Tejido mamario sano

Nota. Se seleccionaron muestras representativas de manera aleatoria correspondiente a cada tipo de tejido.

Para la descarga de los archivos sin procesar, se utilizó la herramienta SRA Toolkit **(Anexo 1)** la cual fue eficiente para la adquisición de datos de secuenciación ya que los datos no se corrompieron durante

su descarga como con otros métodos utilizados. Permitió la conversión de los archivos RNA-Seq sin procesar al formato FASTQ para el análisis de secuencias **(Sola et al., 2021)**.

2.2.1.2. Control de calidad y procesamiento previo de datos genéticos.

Los datos crudos obtenidos en formato .FASTQ del estudio pasaron por un control de calidad y preprocesamiento utilizando FASTQC, esto con el fin de garantizar la limpieza y precisión de los mismos. Finalizado el análisis, se visualizó un informe de resultados el cual constaba de diez módulos donde se resumía el contenido del archivo de secuenciación **(Smith & De Sena, 2021).** En el control de calidad se identificó contaminación proveniente de adaptadores y bases representadas de forma errónea, por lo que se usó la herramienta Trimmomatic donde las lecturas Paired End (PE) fueron filtradas con el fin de eliminar adaptadores y bases líderes y finales de baja calidad utilizando parámetros preestablecidos (parámetros: PE ILLUMINACLIP:TrueSeq-PE.fa:2:30:10:4 LEADING:3 TRAILING:3) **(Coupé et al., 2023)**. El archivo .fa el cual contenía los adaptadores para el recorte se presenta en el **Anexo 8**. Una vez finalizado, se vuelve a realizar un control de calidad para verificar que la calidad de las secuencias haya mejorado.

Se realizó un alineamiento con los archivos de salida *paired* ya que los *unpaired* fueron utilizados como un indicador falso por defecto, donde su lectura asociada es la emparejada y con la que se suele trabajar **(Mrozek et al., 2021)**. Las lecturas fueron alineadas al genoma de referencia GRCh37 (https://ftp.ensembl.org/pub/grch37/cu2rrent/fasta/homo_sapiens/), comúnmente conocido como hg19, mediante HISAT2 **(Anexo 11)**, donde se realizó una indexación del genoma de referencia para que el alineador pueda encontrar de manera más eficiente coincidencias exactas con lecturas secuenciadas en la estructura de datos dada **(Musich et al., 2021).**

Se utilizó la herramienta SAMtools para la conversión y manipulación de formatos SAM y BAM (**Danecek et al., 2021)**. Se realizó el conteo de genes mediante FeatureCounts donde se utilizaron los archivos de lecturas alineadas en formato SAM y una lista de características genómicas en formato de características generales (GFF3), ambos pertenecientes al genoma de referencia GRCh37.

2.2.1.3. Análisis de Expresión Diferencial (DEG).

Se definió el entorno de trabajo en R y se cargaron las librerías visualizadas en el **Anexo 16**, junto con la matriz de expresión. Se definieron las condiciones del estudio (“GL” y “Control” o “TP” y “Control”). Se utilizó la función DESeqDataSetFromMatrix para construir un objeto DESeqDataSet (dds) a partir de la matriz se expresión y se utilizó DESeq() para un análisis de varianza **(Feng et al., 2024)**.

Genes en el grupo TP y GL con log2BaseMean > 0 y padjust < 0.01 en relación con el control se consideraron como genes significativos **(Yang et al., 2023).** Estos fueron representados mediante un Volcano plot con el fin de determinar si los genes se regulaban positiva o negativamente, además de observar la significancia con la que se expresaron.

2.2.2. Identificación de las funciones biológicas que se degradan en diferentes estadios de cáncer de mama.

2.2.2.1. Análisis funcional de genes diferencialmente expresados en el cáncer de mama.

Se utilizó Metascape (http://metascape.org) para visualizar las vías de señalización clave y las firmas funcionales de los genes significativos **(Zhang et al., 2021).** Se obtuvo un archivo zip que contenía toda la información funcional referente a los genes. Los resultados completos de Metascape se pueden encontrar en el **Anexo 25** y **Anexo 26**.

Se utilizó "KEGG" (https://www.kegg.jp/) para identificar vías compartidas por los genes y su relación con funciones sistémicas de la célula en función de la interacción molecular **(Kanehisa et al., 2013).** Ambas herramientas permitieron dilucidar vías de señalización y mecanismos moleculares de los genes **(Zheng et al., 2021).**

2.2.2.2. Construcción de redes de interacción y coexpresión génica.

Se utilizó Cytoscape (https://cytoscape.org/) para visualizar redes de interacción molecular que compartían los genes encontrados en el análisis de expresión diferencial con otros genes cuyas funciones eran similar a las de los genes significativos **(Yang et al., 2022).** Se utilizó su complemento "GeneMANIA", el cual permitió una búsqueda de redes de interacción en bases de datos públicas, lo que permitió el análisis y la visualización de redes biológicas **(Shayakhmetov, 2024)**. En la parte inferior derecha del entorno de Cytoscape, se encontraban las opciones Node Table, Edge Table, Network Table y Unassigned Tables las cuales permitían filtrar los términos de enriquecimiento, eliminar términos redundantes, visualizar términos enriquecidos seleccionados en los nodos y exportar los resultados a un archivo **(Doncheva et al., 2023)**. En Node Table se observaron los genes (nodos) con sus anotaciones funcionales, mientras que en Edge Table se identificó el tipo de interacción (aristas) que compartían los genes. Se seleccionaron los genes de acuerdo a la información de Node Table tomando en cuenta el Score donde 14 genes de GL y 20 genes de TP tenían las puntuaciones Score.

2.2.3. Evaluación de la influencia de los tipos de tumores en diferentes estadios del cáncer de mama sobre diferentes poblaciones.

2.2.3.1. Análisis de variantes genéticas de la heterogeneidad tumoral.

Los archivos BAM de TP y GL junto con el genoma de referencia GRCh37 fueron cargados y procesados para la detección de SNP e indel mediante el software de llamada de variantes VarScan, perteneciente a Galaxy (https://usegalaxy.org/) donde los archivos SNP sin procesar se concatenaron en un archivo de formato de llamada de variante única (VCF) que contenía la ubicación de la variante en el genoma de referencia, las bases que se desvían de la referencia y la confianza estadística con la que se llamó a la variante **(Müller et al., 2017)**.Y se filtraron con VCFfilter donde los verdaderos positivos, falsos positivos y falsos negativos se evaluaron en respuesta a la variación de umbrales de filtrado como la profundidad de secuenciación (número total de lecturas mapeadas en un determinado locus) DP > 10 y la calidad de la llamada de variante Q > 30 **(Nunn et al., 2022)**.

Las variantes fueron anotadas mediante SnpEff, donde se ingresaron los VCF filtrados y se seleccionó el genoma de referencia GRCh37, esto con el fin de obtener información funcional que permita dilucidar las funciones multifacéticas de las variantes genéticas **(Zhou et al., 2023)**. Los resultados se pueden observar en el **Anexo 28.** Los archivos VCF de TP y GL fueron descargados y se verificó que su formato sea correcto empleando el script del **Anexo 27**. El VCF corregido se cargó a la página Cancer Genome Interpreter (CGI) (https://www.cancergenomeinterpreter.org/home) y se seleccionó el genoma de referencia hg37. Se determinó el impacto funcional de las variantes encontradas en el desarrollo del CM con una interpretación clínica de las alteraciones genómicas.

2.2.3.2. Evaluación de la influencia de las variantes genéticas de genes diferencialmente expresados en la evolución clonal del cáncer de mama.

Ya identificadas las variantes de los genes seleccionados en el análisis de expresión diferencial, se clasificaron dependiendo de si eran exclusivos de un grupo (TP o GL) o se encontraban en ambos; las variantes de los genes encontrados en ambos grupos fueron descartadas y aquellas cuyo gen era exclusivo de un grupo fueron tomadas en cuenta.

De determinó el nivel de implicación del gen junto con sus variantes en el desarrollo del cáncer, además de la función biológica que desempeñaba y cómo esta era alterada para la predisposición de la metástasis.

CAPÍTULO III.- RESULTADOS Y DISCUSIÓN

3.1. Análisis y discusión de los resultados

3.1.1. Identificación de la variabilidad genética asociada a la heterogeneidad tumoral en el cáncer de mama.

Comprender los eventos genéticos que conducen a la iniciación y progresión del cáncer sigue siendo uno de los mayores desafíos en la biología del cáncer. Sin embargo, la existencia de bases de datos oncogénicas brinda herramientas para la identificación de variantes causales que intervienen en el cáncer **(Wang et al., 2021).** La aparición de conjunto de datos multiómicos a gran escala ha permitido identificar aberraciones moleculares a nivel de ADN, ARN, proteínas y/o epigenéticos facilitando la caracterización de cánceres humanos **(Tang et al., 2022).** El análisis bioinformático, una estrategia valiosa para el análisis de grandes bases de datos, incluyendo información genética compleja. Por lo que, en este proyecto busca comprender la influencia de heterogeneidad en la evolución clonal **(Zeng et al., 2021).** Para ello, los datos fueron obtenidos de NCBI, correspondiente a 30 pacientes con cáncer de mama, 15 de ellos presentaban metástasis a los ganglios linfáticos y los 15 restantes no presentaban metástasis. Se recolectó de manera aleatoria muestras representativas, las cuales se muestran a continuación:

Tabla 4.
Datos de tumores obtenidos de pacientes con cáncer de mama.

	Run	BioSample	GEO_Accession	Metastasis	Tissue
	SRR15852416	SAMN21394931	GSM5574708	Yes	Breast tumor
GL	SRR15852394	SAMN21395377	GSM5574686	Yes	Breast tumor
	SRR15852401	SAMN21394916	GSM5574693	Yes	Breast tumor
	SRR15852399	SAMN21394914	GSM5574691	No	Breast tumor
TP	SRR15852407	SAMN21394922	GSM5574699	No	Breast tumor
	SRR15852413	SAMN21394928	GSM5574705	No	Breast tumor
	SRR15852443	SAMN21395382	GSM5574735	No	Normal breast tissue
C	SRR15852427	SAMN21394942	GSM5574719	No	Normal breast tissue
	SRR15852426	SAMN21394941	GSM5574718	No	Normal breast tissue

Nota. Para la búsqueda y selección de datos se verificó que los mismos fueran obtenidos mediante secuenciación de próxima generación (NGS), específicamente datos RNA-Seq, correspondientes a cáncer de mama. Además, se eligieron datos publicados los últimos cinco años con el fin de evitar errores en los análisis posteriores.

Los datos que permitieron este estudio fueron obtenidos mediante RNA-Seq, un enfoque transcriptómico revolucionario de alto rendimiento, que ha superado a otros métodos tradicionales por su velocidad, sensibilidad y rendimiento **(Negi et al., 2022).** El estudio del ARN ha permitido comprender la función genómica junto con la causa y el desarrollo de las enfermedades. Se han convertido en biomoléculas de interés debido a que son intermediarios entre los genes y proteínas, razón por la que los fragmentos de ARN proporcionan información directa sobre la actividad genética **(Ergin et al., 2022).**

Los datos NGS típicos generados por la secuenciación de Ilumina se entregan en archivos con formato FASTQ, el cual contiene las secuencias de ADN inferidas con sus puntuaciones de calidad. La mayoría de aplicaciones requieren que se extraiga y cuenten características de los archivos FASTQ para su análisis posterior **(Bravo et al., 2022).** Para su correcto procesamiento, se requieren análisis de calidad lo cuales consisten en una serie de pasos de control donde se realiza un análisis riguroso de los datos para alcanzar una detección temprana de sesgos o problemas técnicos. De esa manera se garantiza la consistencia e integridad de los resultados obtenidos **(Vashishtha et al., 2024).** Para el análisis de control de calidad se empleó la herramienta FASTQC donde se obtuvieron los siguientes resultados:

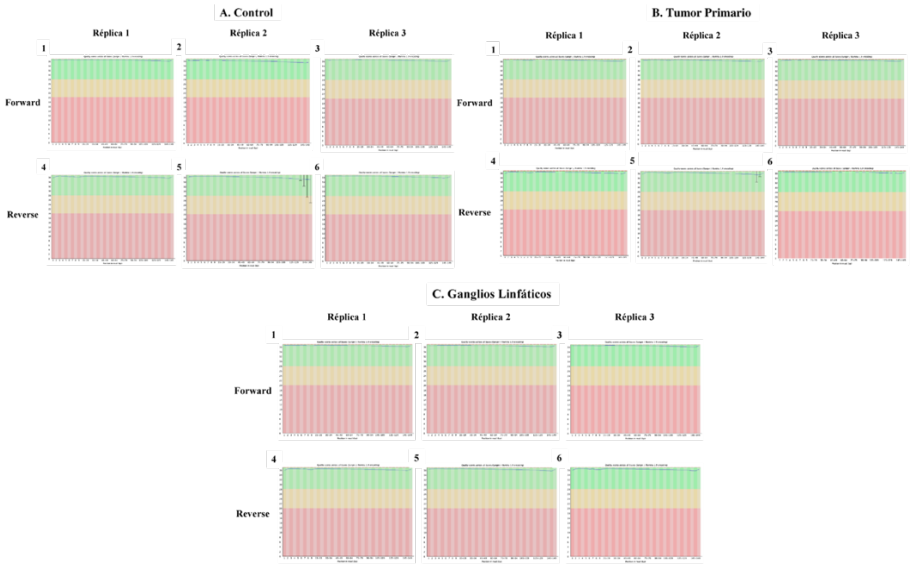


Figura 4. Control de calidad de los datos.

Nota. El eje *y* representa el phred Score el cual se encuentra entre una media de 36. Mientras que el eje *x* ilustra la posición de la lectura en pares de bases (bp). También se pueden observar diferentes colores los cuales representan el nivel de calidad de los datos donde el verde resalta regiones de alta calidad, el amarillo regiones de calidad moderada y el rojo regiones de baja calidad.

La calidad de las lecturas se evalúa en función de la puntuación phred que se asigna a cada nucleótido dentro de la lectura. Cuanto mayor sea la puntuación phred, mejor será la calidad de los datos. Como se puede observar en la **Figura 4,** la mayoría de los datos se encuentran en una región de alta calidad, sin embargo, en la figura A5, algunos reads se encuentran en una región de calidad moderada con un Phred entre 25 y 28, el cual es un valor aceptable según **Gondane & Itkonen, (2023)**.

La mayoría de los conjuntos de datos RNA-Seq se utilizan principalmente para el análisis de expresión diferencial debido a que proporciona información confiable respecto a la influencia de la expresión de genes sobre un fenotipo **(Muzellec et al., 2023).** En este caso los genes identificados fueron:

Tabla 5.
Genes diferencialmente expresados.

Grupo	Nombre	Símbolo	Gene ID
TP	Inmunoglobulina kappa variable 3-11	IGKV3-11	28914
	Dominio kelch que contiene 7B	KLHDC7B	113730
	LON peptidasa dominio N-terminal y dedo anular 2	LONRF2	164832
GL	Distonina	DST	667
	Factor de transcripción de caja SRY 11	SOX11	6664
	Dominio de trombospondina tipo 1 que contiene 4	THSD4	79875

Nota. El análisis de expresión diferencial se lo realizó en R aplicando los criterios log2BaseMean > 0 y padjust < 0.01.

Los genes ilustrados en la **Tabla 2** desempeñan diferentes funciones para mantener la integridad de diversos procesos celulares. Algunos de ellos codifican proteínas presentes en la remodelación de la ECM, donde una serie de enzimas de remodelación como oxidasas y proteasas interactúan con receptores de la superficie celular y factores unidos a la ECM con el fin de mediar la adhesión celular y la señalización celular, regulando de esa manera diferentes procesos como la proliferación, la diferenciación, la migración y la apoptosis **(Winkler et al., 2020).** En tumor primario, el gen IGKV3-11 desempeña un papel importante en la respuesta inmune especialmente en enfermedades autoinmunes

y xenotrasplantes. Su función radica en la codificación de cadenas ligeras de inmunoglobulinas que son cruciales para la especificidad de los anticuerpos dirigidos al colágeno de tipo IV y II, componentes necesarios en la ECM **(Clark et al., 2017)**. KLHDC7B codifica una proteína con un dominio kelch (KLHL). Las KLHL regulan diversos procesos fisiológicos y patológicos mediante la ubiquitinación de sus respectivos sustratos. Este proceso consiste en la degradación de proteínas con el fin de regular procesos celulares como la proliferación celular, diferenciación, apoptosis, migración e invasión **(Zhou et al., 2024).** LONRF2 contiene el dominio N-terminal de la peptidasa LON y el dedo anular 2. Las proteasas Lon junto con chaperonas moleculares y moléculas reguladoras, forman un sistema de control de calidad de proteínas en las mitocondrias y otros orgánulos especializados de los organismos superiores **(Wlodawer et al.,2022).** Por otro lado, en ganglios linfáticos, el gen DST codifica una proteína plaquina. Su función radica en mantener la integridad del citoesqueleto y brindar soporte organizativo a los complejos de adhesión celular actuando como adaptadores para las proteínas de señalización que regulan las conexiones célula-matriz extracelular, la conexión célula-célula, la migración e invasión celular, la diferenciación y respuestas al estrés. Estas características hacen de esta familia de proteínas un objetivo para la investigación del cáncer **(Wesley et al., 2021).** Por otro lado, las proteínas SOX, como SOX11, llevan un dominio HGM que permite su unión específica a dominios adicionales involucrados en la regulación transcripcional. Los TF SOX están presentes en diversos procesos celulares como el mantenimiento de la pluripotencia de las células madre, la proliferación celular, el destino celular y la diferenciación celular terminal en tejidos y órganos. Además, cumplen otras funciones como la supervivencia, la regeneración, la muerte celular y el mantenimiento de la homeostasis en tejidos adultos **(Stevanovic et al., 2021).** Finalmente, el gen THSD4 contiene dominios tipo trombospondina, quienes conforman una familia de proteínas de la matriz extracelular. Regulan dinámicamente el comportamiento celular y la respuesta a los estímulos influyendo directa e indirectamente en procesos biológicos como la respuesta inmunitaria, la angiogénesis y la progresión del cáncer **(Morandi et al., 2021).**

La comparación de un conjunto de datos permite identificar funciones junto con vías críticas de los genes y la mejor forma de hacerlo es de una manera visual, sin embargo, visualizar datos de expresión diferencial puede ser un desafío debido a su complejidad. Para la representación de los genes diferencialmente expresados, una interfaz gráfica fácil de interpretar es una guía a través de los diferentes pasos del análisis, debido a que permite visualizar los resultados intermedios y finales generando figuras de calidad como mapas de calor, gráficos de volcanes, histogramas, diagramas de Venn y perfiles de expresión. Convencionalmente, estos conjuntos de datos suelen representarse como diagramas de dispersión, comúnmente conocidos como diagramas de volcán **(Guo et al., 2021; Mullan et al., 2021).**

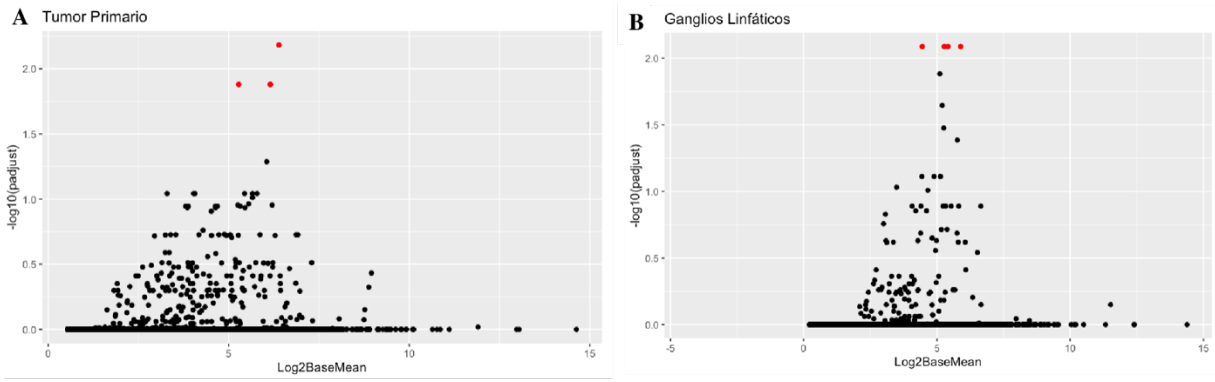


Figura 5. Representación de los genes diferencialmente expresados.

Nota. Los genes identificados en el análisis de expresión diferencial están representados por los puntos rojos que se encuentran en la parte superior de la imagen. Los conjuntos de datos de expresión diferencial pueden contener miles de características a considerar. En el gráfico de volcán se representa cada característica como un cambio de pliegue transformado en log2 a lo largo del eje *x* y el log10 negativo de un valor *p* a lo largo del eje *y* **(OConnell, 2024).**

Un gráfico de volcán muestra una relación entre el cambio y la significancia estadística (valor *p*) de los genes. Los genes con un alto cambio y valores *p* bajos se consideran significativos, indicando una posible relevancia biológica, mientras que los genes ubicados cerca del centro son menos significativos **(Naumov et al., 2017).** En la **Figura 5**, cada gen es representado por un punto, los puntos rojos representan genes con alta significancia reflejado por el eje *y*, y aumentados en la enfermedad ya que se encuentran alejados del origen en una escala positiva, convirtiéndose en los genes de interés de este estudio. En la Figura 5A correspondiente a TP son los genes IGKV3-11, KLHDC7B y LONRF2, y los genes representados en la Figura 5B son DST, SOX11 y THSD4, pertenecientes a GL, donde su implicación en la enfermedad se explicará más adelante.

3.1.2. Examinación de las funciones biológicas afectadas en el desarrollo del CM.

Para dilucidar la patogénesis de la enfermedad, es importante comprender la función de los genes identificados, especialmente en condiciones asociadas con la enfermedad **(Ota et al., 2021).** A continuación, se presenta la función biológica identificada mediante Metascape de los genes identificados:

Tabla 6.

Identificación funcional de los genes diferencialmente expresados.

Grupo	Gen	Función
TP	IGKV3-11	Se prevé que participe en la respuesta inmunitaria.
	KLHDC7B	Está involucrado en varios procesos celulares, incluyendo la organización del citoesqueleto, la degradación de proteínas y la transducción de señales.
	LONRF2	Se prevé que tenga actividad de unión a iones metálicos.
GL	DST	Codifica una proteína plaquina que participa en las uniones de adhesión y se expresa en tejido epitelial.
	SOX11	Codifica una proteína que actúa como regulador transcripcional ya que forma complejos con otras proteínas.
	THSD4	Se prevé que este gen actúe como componente de la matriz extracelular.

Nota. Los genes ilustrados tienen gran importancia en el correcto desarrollo de múltiples procesos biológicos, sin embargo, dichas funciones pueden jugar un papel en contra en la progresión del cáncer.

La metástasis es un proceso complejo que involucra varios tipos de células que componen un tumor y el microambiente en donde se desarrolla. Se identificó que en el cáncer primario se dan mutaciones impulsoras que contribuyen a la proliferación y supervivencia de las células cancerosas **(Cha et al., 2021).** Además, los genes impulsores del cáncer participan en la transformación de células sanas a células cancerosas mediante la ganancia o pérdida de su función. **Sinkala, (2022),** menciona que aquellos genes impulsores que mutan con mayor frecuencia son los genes supresores de tumores (94%), seguidos de los oncogenes (93%), los factores de transcripción (72%), las quinasas (64%), los receptores de la superficie celular (63%) y las fosfatasas (22%). Un ejemplo de gen impulsor es el gen KLHDC7B**.** Esto lo confirman **Luo et al., (2023),** quienes mencionan que sus altos niveles han sido asociados con un mayor porcentaje de células Ki67 (un marcador de proliferación celular). Es decir, cuando el gen se regula positivamente, la proliferación de las células cancerosas aumenta y la apoptosis disminuye. De igual manera, se presenta el gen IGKV3-11. El papel que desempeña en el desarrollo del CM es desconocido, sin embargo, se conoce que interactúa con el colágeno tipo II, quien apoya la salud y la integridad de los condrocitos, mismos que a su vez contribuyen a la formación de fibroblastos. **Lujano et al., (2024)** mencionan que a pesar de que los fibroblastos se encuentren inactivos, están preparados para su activación sea de manera necesaria o indirecta, lo que podría inducir enfermedades como la fibrosis o el cáncer ya que su activación permite la proliferación de los fibroblastos, la migración, la remodelación de la ECM e interacciones paracrinas con otras células como las células endoteliales o

inmunitarias. También se puede observar el gen LONRF2, a pesar de que su información es limitada, estudios recientes sugieren que tiene un papel en el desarrollo y la progresión del cáncer de mama.

Por otro lado, **Martínez et al., (2021)** mencionan que los tumores primarios comparten muchas mutaciones con sus sitios metastásicos, sin embargo, existe una diferencia en el número y tipo de genes afectados como resultado de la evolución del tumor. Además, afirma que los ganglios linfáticos presentan un menor número de genes conductores alterados a diferencia de tumor primario. En ganglios linfáticos se identificó el factor de transcripción SOX11 donde **Sun et al., (2022)** afirman su papel en la carcinogénesis. Mencionan que su expresión disfuncional se correlaciona con una mayor supervivencia de las células cancerosas, la inhibición de la diferenciación celular y la progresión tumoral, convirtiéndolo en biomarcador potencial en el cáncer. Además, en la **Tabla 3,** el gen DST es asociado con una función estructural, sin embargo, en el contexto de la carcinogénesis, **Qiu et al., (2022)** mencionan que este gen tiene un papel en el microambiente inmunológico del CM debido a que se relaciona con múltiples puntos de control inmunes. Por lo que DST desempeña un papel fundamental en la progresión y desarrollo de la enfermedad mediante la regulación de la respuesta inmune. El gen THSD4 al igual que DST, cumple una función estructural, sin embargo, **Majed & Mustafa, (2021)** mencionan que en el CM su actividad depende de la regulación del gen miR-1275, si su regulación es negativa, promueve la proliferación de células cancerosas, la diferenciación celular, el crecimiento tumoral, la invasión, la migración y la inhibición de la apoptosis a través de THSD4 junto con otros genes. A pesar de que el papel de THSD4 no ha sido identificado, se tiene conocimiento de que contribuye al crecimiento tumoral en una variedad de cánceres, siendo uno de ellos el CM.

La inteligencia artificial se ha convertido en un método eficaz para el procesamiento de datos de redes biológicas que son construidas por programas para facilitar la clasificación, agrupamiento y predicción en redes biológicas ya que permiten abordar la complejidad del cáncer que surge de interacciones entre los genes y sus productos con el fin de mejorar la comprensión de la carcinogénesis y explorar objetivos anticancerígenos **(You et al., 2022).** Fenotipos complejos como el cáncer implican relaciones funcionales entre genes altamente conectados donde interacciones construidas a partir de varios conjuntos de datos y métodos aumentan la confiabilidad y la cobertura de una red genética para su correcto estudio (**Kim et al., 2022)**. A continuación, se obtuvieron redes de interacción aplicadas a ambos grupos donde se observan los siguientes resultados:

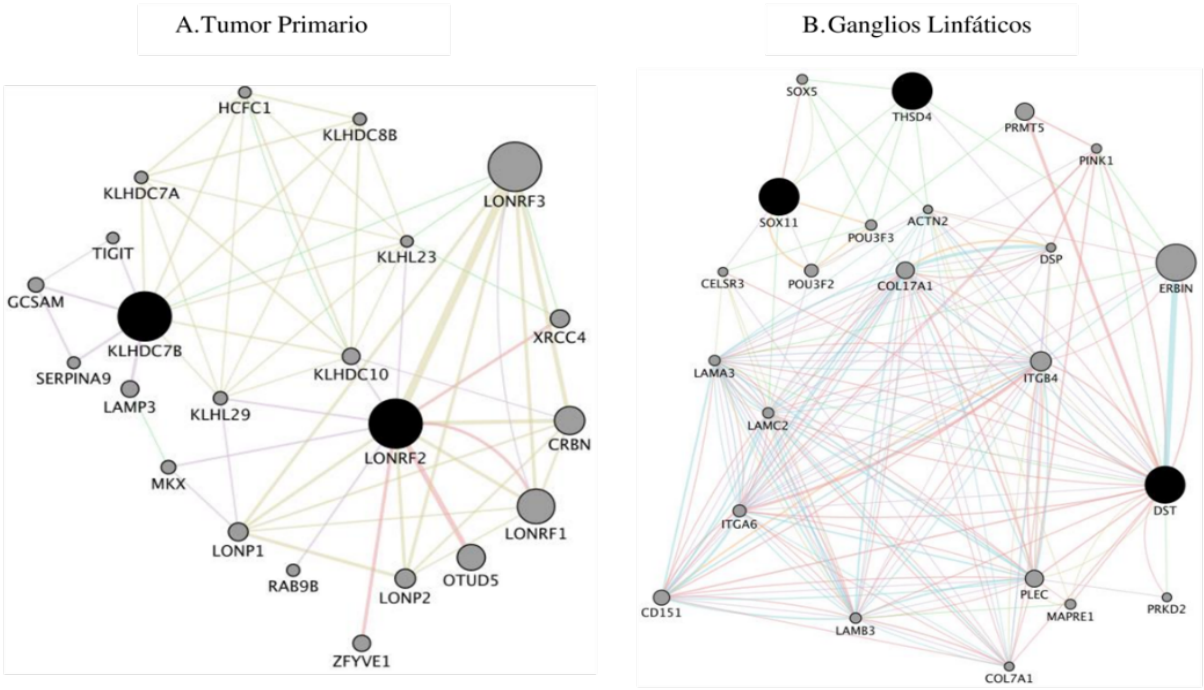


Figura 6. Redes de interacción de los genes diferencialmente expresados correspondientes al grupo TP y GL.

Nota. Los nodos representan genes y las aristas de diferentes colores corresponden al tipo de relación. En TP existen 15 aristas referentes a coexpresión (líneas moradas), 6 de interacción genética (líneas verdes), 4 de interacción física (líneas rosadas) y 34 aristas de dominios proteicos compartidos (líneas cafés). Mientras que en GL existen 37 aristas de coexpresión (líneas moradas), 19 de interacción genética (líneas verdes), 39 rutas compartidas (líneas celestes), 56 de interacción física (líneas rosadas), 6 de co-localización (líneas azules) y 17 aristas de dominios proteicos compartidos (líneas cafés).

El cáncer es reconocido como un proceso evolutivo donde las alteraciones genéticas proporcionan el sustrato para la evolución. La adaptación del tumor al ambiente que le rodea le confiere cambios aleatorios en las frecuencias de mutación **(Rogiers et al., 2022).** Es una de las razones por las que en la **Figura 6** se observan genes completamente distintos en cada grupo. Además, existe un mayor número interacciones entre los genes de GL. Esto según **Patel et al., (2020)** es debido a que la metástasis es más compleja debido a una mayor carga mutacional y diversidad clonal en comparación con cánceres de mama en etapa temprana. También menciona que las alteraciones genéticas son únicas en el sitio metastásico en comparación con su respectivo tumor primario. **Zhou & Zhang, (2024)** afirman que las células metastásicas adquieren cambios evolutivos que mejoran su supervivencia y proliferación en nuevos entornos, presentando un panorama genético más complejo. Además, menciona que el tumor primario presenta perfiles genéticos más estables, limitando su adaptabilidad.

La construcción de redes contribuye a dilucidar asociaciones entre el origen y el mecanismo molecular de las enfermedades, importante para la investigación médica. También permite el descubrimiento de relaciones entre enfermedades humanas, información que se puede aplicar para la reutilización de medicamentos **(Yu & Gao, 2019).** Las vías biológicas compartidas por los genes son cruciales en el estudio de enfermedades debido a que revelan interacciones moleculares, mejoran los tratamientos clínicos y ayudan en el descubrimiento de objetivos farmacológicos y biomarcadores, favoreciendo un correcto pronóstico de enfermedades **(Zheng et al., 2018).**

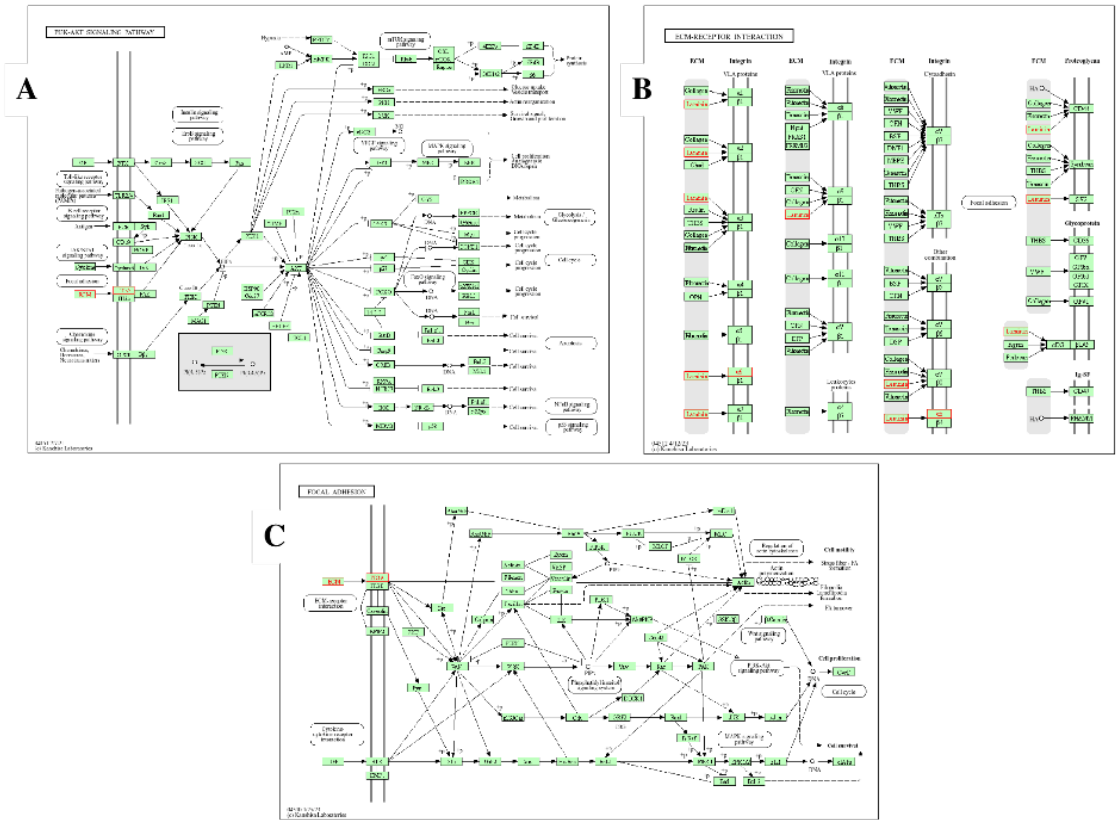


Figura 7. Rutas biológicas compartidas entre los genes diferencialmente expresados.

Nota. Las rutas biológicas identificadas fueron: A. Vías de señalización PI3K-AKT. B. Interacción del receptor del ECM. C. Adhesión focal. Los genes en rojo corresponden a LAMA3, LAMB3, LAMC2 e ITGA6.

La sobreexpresión de proteínas en los componentes de la ECM y la alteración del entorno de la ECM, conduce a la progresión tumoral en varios tipos de cáncer. Los genes LAMA3, LAMB3 y LAMC2 representados como ECM y laminin en la **Figura 7**, codifican la laminina-332, promotora de la progresión tumoral, ya que interactúa con receptores de la superficie celular, como las integrinas (ITGA6) y afecta la vía de señalización PI3K-AKT al reclutar tirosina quinasas receptoras

(Wattanathavorn et al., 2024). La vía de señalización PI3K/AKT es de interés en el cáncer debido a su implicación en la EMT. Además, está presente en varios procesos celulares tales como el metabolismo de la glucosa, la apoptosis, la proliferación celular y la migración celular **(Qu et al., 2021).**

Por otro lado, estos genes también se encuentran conformando la ECM. Esto lo confirma **(Skowron et al., 2021)** quien menciona que las lamininas son componentes insolubles de la ECM ya que se encuentran unidas a las membranas celulares a través de receptores de integrina como ITGA6, los cuales son responsables de la transducción de señales desde el microambiente hacia la célula. Las interacciones laminina-integrina inician vías de señalización que promueven la supervivencia celular y la autorrenovación. En el cáncer, los genes diferencialmente expresados influyen remodelación de la ECM normal para dar lugar a una ECM tumoral, un paso clave en la tumorigénesis y la metástasis. A través de este proceso, se modulan características definidas del cáncer, promoviendo el crecimiento, la supervivencia y la invasión celular **(Lepucki et al., 2022).**

La adhesión focal también cumple con diversas funciones biológicas. Las proteínas extracelulares, lamininas, junto con sus receptores transmembrana, integrinas, median la comunicación celular, la adhesión celular y el enlace con los receptores de la superficie celular, importante para la generación de señales que regulan la forma y la función de los tejidos **(Calaf et al., 2022).** En la tumorigénesis, **Nathanson et al., (2021)** mencionan que moléculas de adhesión se expresan y desexpresan, haciendo referencia a la EMT y MET al desprendimiento y colonización de las células cancerosas. Además, afirman que la adhesión célula-célula contribuye a los grupos de células tumorales a evadir con éxito el ataque del sistema inmunológico ya que son más resistentes a las células NK en comparación con células individuales. Esto se debe a que varias moléculas de adhesión célula-célula son ligandos inhibidores de las células NK, razón por la que las células tumorales solitarias o células con una EMT completa son susceptibles.

Del análisis de redes de interacción se obtuvieron los siguientes genes:

Tabla 7.

Identificación de genes presentes en las redes de interacción.

Tumor Primario				**Ganglios Linfáticos**			
Nombre	**Símbolo**	**Gene ID**	**Score**	**Nombre**	**Símbolo**	**Gene ID**	**Score**
Inmunorreceptor de células T con dominios Ig e ITIM	TIGIT	201633	1.28E-4	Actinina alfa 2	ACTN2	88	3.93E-4
Kelch como miembro de la familia n° 23	KLHL23	151230	1.29E-4	Proteína quinasa D2	PRKD2	25865	4.30E-4
	HCFC1	3054	1.30E-4	Desmoplaquina	DSP	1832	4.31E-4
Miembro 9 de la familia A de las serpinas	SERPINA9	327657	1.35E-4	Cadena alfa 1 de colágeno tipo VII	COL7A1	1294	4.63E-4
Miembro de la familia de oncogenes RAS	RAB9B	51209	1.42E-4	Receptor 3 de tipo G de siete pasos de cadherina EGF LAG	CELSR3	1951	4.63E-4
Dominio kelch que contiene 8B	KLHDC8B	200942	1.42E-4	Quinasa 1 inducida por PTEN	PINK1	65018	4.94E-4
Dominio Kelch que contiene 7A	KLHDC7A	127707	1.43E-4	Factor de transcripción 5 de la caja SRY	SOX5	6660	5.74E-4
Kelch como miembro de la familia 29	KLHL29	114818	1.62E-4	Miembro 1 de la familia de proteínas asociadas a microtúbulos RP/EB	MAPRE1	22919	6.07E-4
Mohawk homeobox	MKX	283078	1.69E-4	Subunidad alfa 3 de laminina	LAMA3	3909	6.55E-4

Señalización y motilidad asociadas al centro germinal	GCSAM	257144	1.88E-4	Subunidad beta 3 de laminina	LAMB3	3914	6.62E-4
Dedo de zinc tipo FYVE que contiene 1	ZFYVE1	53349	2.22E-4	Subunidad gamma 2 de laminina	LAMC2	5455	6.66E-4
Dominio kelch que contiene 10	KLHDC10	23008	2.24E-4	POU clase 3 homeobox 3	POU3F3	3918	6.65E-4
Proteína 3 de membrana asociada a lisosomas	LAMP3	27074	2.26E-4	Subunidad alfa 6 de integrina	ITGA6	3655	8.95E-4
Reparación de rayos X con complemento cruzado 4	XRCC4	7518	2.53E-4	POU clase 3 homeobox 2	POU3F2	5454	9,71E-4
Peptidasa lon 1, mitocondrial	LONP1	9361	2.59E-4				
Peptidasa lon 2, peroxisomal	LONP2	83752	2.76E-4				
Desubiquitinasa 5 de OTU	OTUD5	55593	4.12E-4				
Cereblon	CRBN	51185	4.54E-4				
Dominio N-terminal de la peptidasa LON y dedo anular 1	LONRF1	91694	5.82E-4				
Dominio N-terminal de la peptidasa LON y dedo anular 3	LONRF3	79836	8.64E-4				

Nota. Se seleccionaron 20 genes de TP y 14 genes de GL, según el Score que presentaban.

En tumor primario, los genes SERPINA9, LAMP3, HCFC1 se relacionan con IGKV3-11 debido a que cumplen un papel en el sistema inmunitario. **Wang et al., (2024)** mencionan que las serpinas inhibidoras desempeñan un papel en la respuesta inflamatoria, la activación del complemento y la regulación inmunitaria. **Tanaka at al., (2022)** afirman que la expresión de LAMP3 se limita a células específicas siendo una de ellas las células dendríticas, componentes clave del sistema inmunológico. Finalmente, **Makhijani et al., (2023**) mencionan una implicación de las vías de señalización aguas abajo del receptor de insulina (InsR) en la conformación de la respuesta inmunitaria. InsR activado se colocaliza con el factor de la célula huésped-1 (HCF-1) con el fin de mediar la actividad transcripcional directa.

Los genes KLHL23, KLHL29, KLHDC8B, KLHDC10, RAB9B, OTUD5 mostraron relación con KLHDC7B ya que cumplen funciones regulatorias. La familia de genes tipo Kelch (klhl) codifican proteínas adaptadoras de sustrato para la ubiquitinación de Cullin-E3. **Peng et al., (2023)** confirman que RAB participa en la regulación del tráfico vesicular, incluyendo la formación de vesículas, el transporte, el acoplamiento de la membrana y la fusión **(Jiang et al., 2022).** Finalmente, OTUD5 es un miembro de la familia de enzimas desubiquinantes OTU. Su función radica en la desubiquitinación de proteínas clave en varias vías de señalización celular y el mantenimiento de las funciones biológicas **(Fu et al., 2023).**

Los genes OTUD5, LONP1, LONP2, CRBN, KLHL23, KLHL29, KLHDC8B, KLHDC10 se asocian con LONRF2 debido a que mantienen la homeostasis proteica. LONP1 es una parte esencial del sistema de control de calidad de las proteínas de los orgánulos, específicamente de la matriz mitocondrial **(Pollecker et al., 2021).** Mientras que LONP2 degrada proteínas dañadas por H_2O_2 que escapan de la reducción, manteniendo la homeostasis de las proteínas peroxisomales y evitando la acumulación del daño oxidativo **(Pomatto et al., 2017).** CRBN es un receptor de reconocimiento de sustrato para la ubiquitina ligasa Cullin 4 RING E3 (CRL4), por lo que participa en el proceso de ubiquitinación **(Yamamoto et al., 2022).**

En ganglios linfáticos, existen genes relacionados con la ECM y adhesión celular tales como COL7A1, DST, THSD4, LAMA3, LAMB3, LAMC2 e ITGA6. Las proteínas de colágeno son los principales componentes estructurales funcionales de la ECM. El colágeno tipo VII es una proteína que pertenece a la familia de los colágenos. La fibrilla de colágeno 7 funciona como un anclaje entre la membrana basal y las células proximales de los epitelios escamosos estratificados **(Koca et al., 2023).** THSD4 es un componente de la ECM, promueve el ensamblaje de la matriz de fibrilina-1 **(Elbitar et al., 2020).**

Los genes SOX5, SOX11, POU3F2 y POU3F3 son factores de transcripción, genes implicados en la señalización neuronal y desarrollo. SOX5 y SOX11 pertenecen a la familia de factores de transcripción SOX con un dominio HMG. Son proteínas que se unen al ADN y desempeñan un papel clave en la

transcripción genética **(Zhang & Hou, 2021).** Finalmente, POU3F2 y POU3F3 pertenecen a la familia POU de factores de transcripción. Tienen como función inducir la pluripotencia celular y la activación del genoma cigótico **(Bakhmet & Tomilin, 2022).**

Además, en la **Tabla 6**, se puede observar una columna denominada "Score", suele usarse en la identificación de candidatos interactores en estudios de interacción. Es la puntuación que se le da a los datos con el fin de determinar la precisión de la interacción identificada. Esto se consigue combinando varios parámetros como la reproducibilidad, la especificidad y la abundancia de cada gen o proteína detectado **(Richards et al., 2021).**

La metástasis es una serie de eventos biológicos que consisten en la diseminación, entrada y salida de latencia de manera variable y la colonización de órganos distales. El éxito de estos eventos depende de la selección clonal, el potencial de las células metastásicas para una transición dinámica y la capacidad de las células tumorales para apropiarse del entorno inmunológico **(Gerstberger et al., 2023)**. Eso es determinado por los mecanismos moleculares que subyacen en el CM.

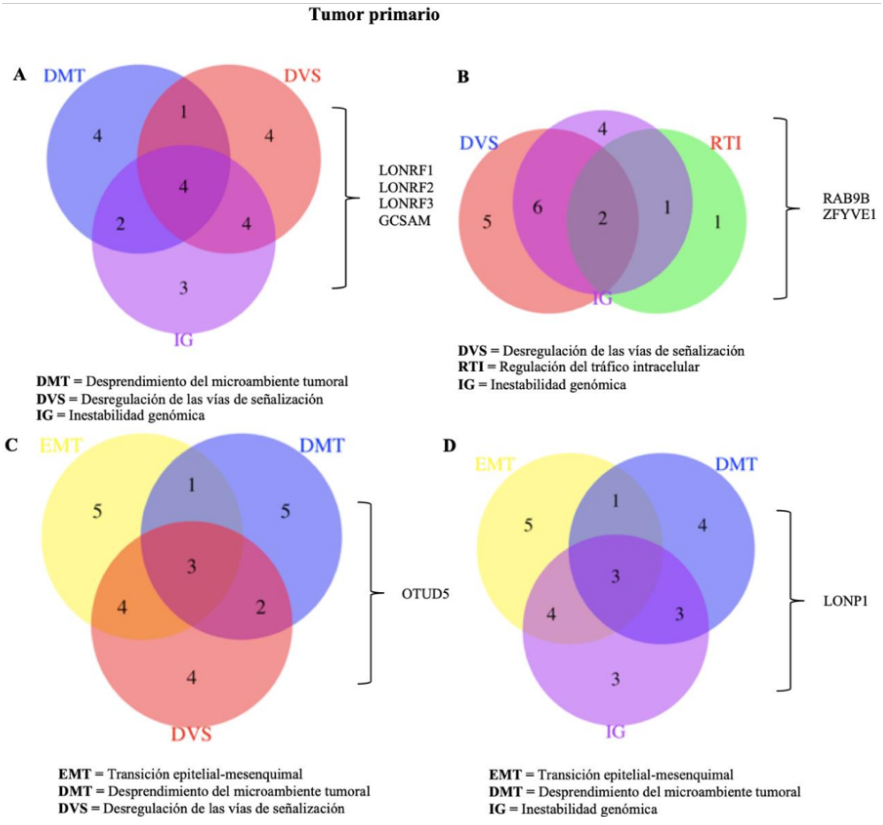


Figura 8. Identificación de mecanismos moleculares compartidos por los genes relacionados a los genes diferencialmente expresados identificados en TP.

Nota. Existen 8 genes pertenecientes al grupo TP que comparten funciones y mecanismos moleculares en la progresión del cáncer.

En tumor primario, LONRF1, LONRF2, LONRF3 junto con OTUD5 median el proceso de ubiquitinación y desubiquitinación, su desregulación puede traer efectos negativos sobre la integridad celular. Puede promover un microambiente tumoral por un desequilibrio en la modulación de citocinas pro- y antitumorales **(Chang & Ding, 2018)**. También puede causar una desactivación o activación aberrante de las vías de señalización, la acumulación de proteínas dañadas o mal plegadas, la deslocalización o el mal tráfico de proteínas **(Pruneda et al., 2021)**. Además, puede contribuir al inicio y la progresión de la transformación neoplásica dando lugar a inestabilidad genómica **(Morgan & Crawford, 2021)**. El gen GCSAM induce la fosforilación de Syk, y regulan negativamente la motilidad celular mediante la activación de Rho. Además, está involucrado en la señalización de células B dentro de los centros germinales **(SoRelle et al., 2022).** El microambiente tumoral es una red compleja de diversas interacciones entre células inmunes y componentes no celulares como la ECM, exosomas e interleucinas. En un microambiente neoplásico fuertemente inmunosupresor, las células del sistema inmune sufren una reprogramación dejando de cumplir su función original **(Wojas et al., 2022).** Las aberraciones epigenéticas, la desregulación de las vías de señalización y el metabolismo alterado, desempeñan un papel importante en la composición y estado funcional del paisaje inmunológico **(Weverwijk & Visser, 2023).** RAB9B junto con ZFYVE1 median el transporte de vesículas. Los microARN endosomales pueden afectar la carcinogénesis y la progresión del cáncer. Las endosomas tienen un impacto regulador en la formación de tumores y el cambio de microambiente tumoral al transportar circARN a las células tumorales, contribuyendo también a la inestabilidad genómica **(Nedaeinia et al., 2024).** Los metabolitos, además de proporcionar sustratos para procesos biológicos; también proporcionan señales críticas, ya sea mediante efectos de las vías metabólicas o por la modulación de otras proteínas reguladoras **(Baker & Rutter, 2023).** Por otro lado, LONP1 puede impulsar la EMT al aumentar la producción de especies reactivas de oxígeno mitocondriales, activando al mismo tiempo las vías ERK y p38 MAPK. Su actividad se encuentra mediada por Akt, contribuyendo al crecimiento del tumor primario y metástasis. Además, permite la supervivencia de las células cancerosas en respuesta a las condiciones de estrés presentes con frecuencia en el microambiente tumoral **(Zanini et al., 2022).** LONP1 mitocondrial pertenece a la familia de ATPasas asociadas con diversas actividades celulares que gobiernan varios procesos celulares como la replicación del ADN, la expresión génica, la fusión de membranas y la degradación de proteínas **(Shetty et al., 2024).**

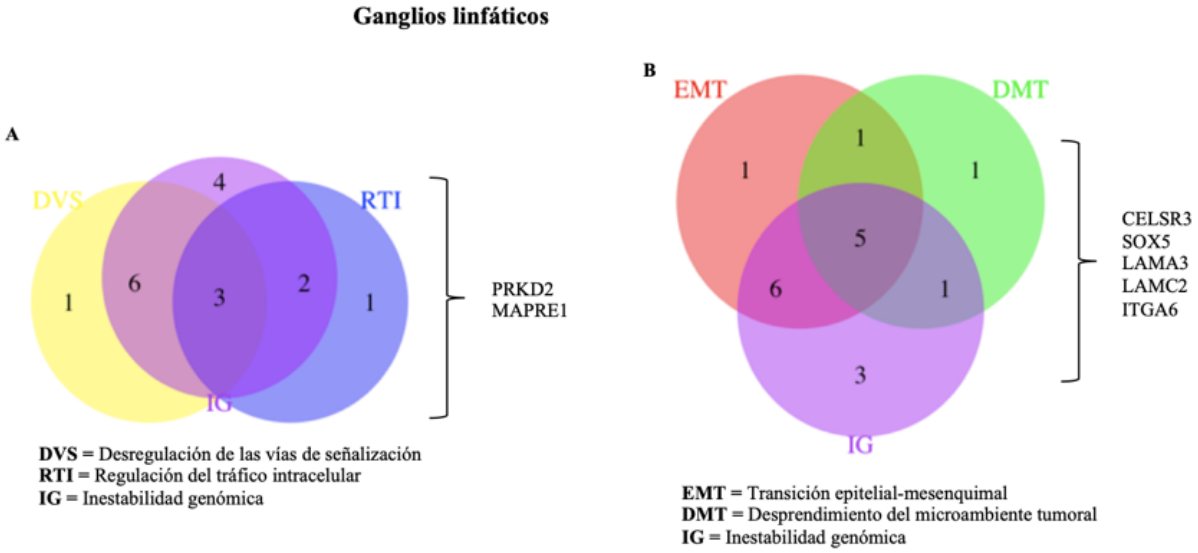


Figura 9. Identificación de mecanismos moleculares compartidos por los genes relacionados a los genes diferencialmente expresados identificados en ganglios linfáticos.

Nota. Existen siete genes pertenecientes al grupo GL que comparten funciones y mecanismos moleculares en la progresión del cáncer.

El tráfico de membrana se coordina mediante una red compleja de vías de señalización para mantener la homeostasis celular. La desregulación de las proteínas y las vías de señalización promueven la expresión aberrante de proteínas transmembrana, la clasificación y el transporte de carga, impulsando el desarrollo de enfermedades **(Gutiérrez et al., 2023).** PRKD2 en el CM afecta la resistencia a fármacos al regular la proliferación de células tumorales, la apoptosis, la metástasis y la invasión. Promueve la sobreexpresión del gen CDKN1A mediante vías dependientes o independientes de p53 y la regulación integrada de la señal extracelular, afectando la sensibilidad celular **(Xu et al., 2022).** El factor de crecimiento transformante-beta (TGF-β) esculpe el microambiente tumoral local y apoya la inmunosupresión. Su firma proteica se relaciona con los genes CLIC, MAPRE1, SERPINA3, quienes promueven la progresión del tumor **(Kvokačková et al., 2021).** Varios factores de transcripción de la familia SOX funcionan como reguladores clave del destino de las células madre/progenitoras en la glándula mamaria. Actúan modulando múltiples vías de señalización oncogénica en el CM. Además, pueden moldear el microambiente inmunológico del tumor **(Liu & Guo, 2021).** SOX5 es un miembro importante de familia SOX. Actúa como un oncogén en niveles desregulados, promoviendo la aparición y el desarrollo de tumores; además, mantiene los fenotipos de las células cancerosas **(Xue et al., 2024)**. Finalmente, las lamininas (LAMA3 y LAMC2) al relacionarse con integrinas (ITGA6), adquieren cierta protección celular de adquirir un fenotipo invasivo, sin embargo, las células transformadas aprovechan a las lamininas para la transmisión de señales mecánicas que promueven la EMT, la invasión celular y

la metástasis. Además, el factor de crecimiento transformante, un inductor de EMT, altera el patrón de expresión de integrinas en células cancerosas, suprimiendo su adhesión a las lamininas **(Nonnast et al., 2024).** Los genes diferencialmente expresados desempeñan un papel en la remodelación de la matriz extracelular normal y la aparición de la matriz extracelular tumoral modulando la composición del microambiente tumoral, un factor clave para la progresión maligna **(Lepucki et al., 2022).** CELSR3 también cumple una función de adhesión celular, además de la comunicación mediada por contacto y en las interacciones receptor-ligando. Regula PCP mediante la vía de señalización Wnt/PCP donde la desregulación de sus componentes centrales contribuye a varios tipos de tumores sólidos **(Zheng et al., 2022).**

3.1.3. Evaluación de la influencia de los tipos de tumores en diferentes estadios del cáncer de mama.

La variación genética hace referencia a un cambio simultáneo de alelos de un gen en el genoma. Se da entre poblaciones, dependiendo de los portadores individuales y compara diferentes sistemas biológicos en el mismo período de tiempo. Esta variación genética es causada por mutaciones aleatorias, las cuales son cambios constantes cualitativos o cuantitativos en el genoma de un individuo y población, pertenecientes a un determinado organismo **(Franjić, 2023).** La búsqueda de variantes genéticas en CGI permitieron obtener los siguientes resultados:

Tabla 8.
Identificación de genes exclusivos.

			MUESTRAS	
			TP	**GL**
GENES	**TP**	IGKV3-11	-	-
		KLHDC7B	SI	SI
		LONRF2	-	-
	GL	DST	SI	SI
		SOX11	-	SI
		THSD4	SI	SI

Nota. Los genes KLHDC7B, DST y THSD4 presentaron variantes tanto en TP como en GL, IGKV3-11 y LONRF2 carecían de variantes y SOX11 se identificó como un gen exclusivo de GL.

En la **Tabla 8,** SOX11 se identificó como un gen exclusivo de GL. Los TF SOX realizan múltiples procesos celulares durante el desarrollo como el mantenimiento de la pluripotencia de las células madre, la proliferación celular, el destino celular y la diferenciación celular terminal en tejidos y órganos. Además, influye en la supervivencia, la regeneración, la muerte celular, y controlan la homeostasis en los tejidos adultos **(Stevanovic et al., 2021)**. Las funciones en la plasticidad celular lo involucran en la EMT, un proceso donde células epiteliales adquieren características indiferenciadas en determinadas condiciones. Varios factores de transcripción influyen en la EMT tales como SNAIL, TWIST1, ZEB y SLUG. Los TF SOX median la EMT de los tumores activando ALK directamente o mediante la regulación positiva de n-MYC, donde se activan las cascadas de transducción descendentes las cuales contienen AKT, NF- κB y TWIST1 conllevando a la EMT **(Jiang et al., 2024)**. Como consecuencia, las células cancerosas expresan marcadores epiteliales y mesenquimales, proveyéndoles una ventaja de supervivencia para mantenerse dentro del huésped **(Debnath et al., 2022)**.

Las proteínas SOX controlan múltiples procesos esenciales del desarrollo y homeostáticos, por lo que su disfunción y mutación puede conducir a una variedad de enfermedades hereditarias, incluyendo el cáncer. El ARNm de SOX11 es una de las transcripciones aumentadas con mayor frecuencia en varios cánceres, incluyendo el CM, ya que varios de sus mecanismos reguladores, como la metilación del promotor, las modificaciones de histonas y la interferencia de microARN, regulan su expresión y actividad. Actualmente se lo ha catalogado como un oncogén debido a sus efectos proliferativos y anti apoptóticos **(Yang et al., 2019)**. Sus altos niveles de expresión se asocian con metástasis distante y un mal pronóstico. Esto debido a que es un gen característico de las células madre mesenquimales (MSC) y un biomarcador potencial para las MSC humanas progenitoras tempranas, razón por la que su expresión disfuncional se ha correlacionado con una mayor supervivencia de las células cancerosas, la inhibición de la diferenciación celular y la progresión tumoral. Las mutaciones de SOX11 son poco comunes en cánceres primarios, sin embargo, alteraciones genéticas en cánceres metastásicos siguen siendo desconocidas **(Sun et al., 2022)**.

Las variaciones genéticas incluyen variaciones de la línea germinal, mutaciones somáticas y diversidades en los genes receptores de células inmunes reorganizadas, células T y células B. Según investigaciones médicas, también se clasifican en RFLP (polimorfismo de longitud de fragmentos de restricción), VNTR (número variable de repeticiones en tándem), STR (repeticiones cortas en tándem o microsatélites), SNP (polimorfismo de un solo nucleótido), CNV (variación del número de copias) y SV (variaciones estructurales) **(Kamatani & Nakamura, 2021).** A continuación, se muestran las variantes encontradas en tejido mamario.

Tabla 9.

Identificación de variantes.

Group	Gen	Protein Change	Oncogenicity	Mutation	Consequence
	SOX11	T44M	Passenger	chr2:5832984C>T	Missense variant
	SOX11	T280M	Passenger	chr2:5833692C>T	Missense variant
GL	SOX11	3 prime UTR variant	Non-affecting	chr2:5835977 C>T	3 prime UTR variant

Nota. Se identificaron 3 variantes correspondientes al gen SOX11 donde todas presentan la transición C > T y dos de ellas poseen mutaciones no sinónimas.

En la **Tabla 9** se puede observar mutaciones puntuales SNP donde se da la sustitución de un solo nucleótido por otro en una ubicación específica dentro del genoma. Este tipo de mutaciones representa más del 90% de todas las variaciones genéticas humanas, por lo que son consideradas importantes en la variación genética **(Sameer et al., 2021)**. Los cambios de nucleótidos entre dos purinas (A y G) y entre dos pirimidinas (C y T) se conocen como transiciones, mientras que los cambios entre una purina y una pirimidina se conocen como transversiones. Las transiciones suelen ser más frecuentes debido a que requieren una distorsión menor de la estructura del ADN y ocurren por desaminación, un cambio químico común de los nucleótidos **(Zou & Zhang, 2021).** También se puede observar que son sustituciones missense pertenecientes a mutaciones no sinónimas las cuales se dividen en dos categorías: mutación missense quien produce cambios de un solo aminoácido, o mutación nonsense que produce proteínas truncadas o más largas. La mutación missense genera variantes proteicas con una sola variación de aminoácido (SAV), son de interés debido a que son capaces de comprometer la integridad de la estabilidad o función de la proteína conduciendo a su degradación por el sistema ubiquitina-proteasoma y a su vez, a una cantidad celular insuficiente del SAV, que puede ser causa de enfermedad **(Petrosino et al., 2021).**

Por otro lado, se puede observar en la columna "Protein Change" que se da el cambio de una treonina a una metionina, **Sedillo & Cryns, (2022)** afirman que las alteraciones oncogénicas de las células cancerosas cambian la utilización de nutrientes, lo que conlleva a estados metabólicos alterados con el fin de promover una mayor proliferación y crecimiento. La metionina es el aminoácido iniciador de la síntesis de proteínas, la fuente de metilo para la estructura carbonatada para varios aspectos de la respuesta antioxidante celular y la biosíntesis de nucleótidos. A diferencia de las células normales, las células cancerosas tienen una alta actividad del ciclo de la metionina ya que dependen de la metionina

exógena para el crecimiento continuo. Por otro lado, treonina juega un papel importante en la regulación fisiológica, particularmente en el crecimiento y la proliferación celular. También actúa como una molécula de señalización regulando la vía de síntesis de proteínas, por lo que su alteración puede comprometer la estabilidad e integridad de la síntesis proteica **(Tang et al., 2021)**.

Actualmente no existe información que respalde la implicación de las variantes mencionadas en el cáncer de mama. A pesar de eso, esta investigación puede ser un punto de partida para estudios futuros que permitan comprender el impacto funcional de dichas variantes.

CAPÍTULO IV.- CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones

Se identificó la variabilidad genética asociada a la heterogeneidad tumoral mediante un DEG donde los genes IGKV3-11, KLHDC7B y LONRF2 para muestras de tumor primario y los genes DTS, SOX11, THSD4 para muestras de ganglios linfáticos, se expresaron significativamente bajo los criterios log2BaseMean > 0 y padjust < 0.01. Además, los análisis de control de calidad aseguraron que las secuencias utilizadas sean fiables, garantizando resultados confiables.

Se examinaron las funciones biológicas afectadas en el desarrollo del CM mediante un análisis de enriquecimiento funcional donde los resultados reflejaron genes codificadores de proteasas (KLHDC7B y LONRF2) y mediadores del sistema inmunitario (IGKV3-11) en TP y genes con funciones estructurales (DTS y THSD4) y factores de transcripción (SOX11) en GL. Además, mediante herramientas gráficas se visualizaron redes de interacción que permitieron asociar a los genes con funciones celulares mediante redes de interacción molecular, lo que permitió establecer mecanismos moleculares de la metástasis. Un ejemplo es la EMT donde células epiteliales adquieren características mesenquimales para la migración celular lo que implica una serie de cambios tanto en el microambiente donde se está desarrollando el tumor como de las funciones que cumplen ciertas proteínas que promueven la proliferación de células tumorales. Tanto proteasas como proteínas con funciones estructurales (Ej. Adhesión celular) pueden involucrarse en la remodelación de la ECM con el fin de promover un microambiente tumoral que garantice la supervivencia y proliferación de células cancerosas.

Se evaluó la influencia de los tipos de tumores en diferentes estadios del cáncer de mama sobre diferentes poblaciones mediante el análisis de variantes genéticas, como resultado se obtuvo que SOX11 fue un gen exclusivo de GL, promotor de metástasis debido a que es un gen multifacético por lo que su desregulación altera múltiples procesos celulares. Además, su alta expresión se la ha relacionado con malos pronósticos en el CM.

La comparación entre TP y GL permitió evaluar la influencia de la heterogeneidad tumoral en la evolución clonal mediante la identificación de SOX11 como un gen exclusivo. A pesar de que hubo genes como KLHDC7B, DST y THSD4 que se expresaron en ambos grupos en el análisis de variantes, se encontró que SOX11 son poco comunes en cánceres primarios, sin embargo, se requieren estudios que permitan comprender las alteraciones genéticas que presenta en cánceres metastásicos.

4.2. Recomendaciones

- Aumentar el tamaño de la muestra implementando otros sitios metastásicos para la identificación de nuevos biomarcadores que promuevan la medicina personalizada, y a su vez, permitan una mejor comprensión de la heterogeneidad tumoral.
- Trabajar con secuencias obtenidas de hospitales del Ecuador, ya que los datos con lo que se trabajó son extranjeros, por lo que el ambiente y el estilo de vida influirán en los resultados obtenidos en este estudio.
- Investigar variantes de genes impulsores para comprender su papel en la metástasis y su afección en la estructura y función proteica.
- Fomentar la educación sobre herramientas bioinformáticas en estudios biomédicos para obtener una mejor comprensión de la relación existente entre la genética y la salud.

REFERENCIAS BIBLIOGRÁFICAS

Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast : Official Journal of the European Society of Mastology*, *66*, 15. https://doi.org/10.1016/J.BREAST.2022.08.010

Arzanova, E., & Mayrovitz, H. (2022). The Epidemiology of Breast Cancer. *Exon Publications*, *23*, 1–19. https://doi.org/10.36255/EXON-PUBLICATIONS-BREAST-CANCER-EPIDEMIOLOGY

Baker, S. A., & Rutter, J. (2023). Metabolites as signalling molecules. *Nature Reviews Molecular Cell Biology 2023 24:5*, *24*(5), 355–374. https://doi.org/10.1038/s41580-022-00572-w

Bakhmet, E. I., & Tomilin, A. N. (2022). The functional diversity of the POUV-class proteins across vertebrates. Open Biology, 12(6). https://doi.org/10.1098/RSOB.220065/

Beňačka, R., Szabóová, D., Guľašová, Z., Hertelyová, Z., & Radoňák, J. (2022). Classic and New Markers in Diagnostics and Classification of Breast Cancer. *Cancers 2022, Vol. 14, Page 5444*, *14*(21), 5444. https://doi.org/10.3390/CANCERS14215444

Bhushan, A., Gonsalves, A., & Menon, J. U. (2021). Current State of Breast Cancer Diagnosis, Treatment, and Theranostics. *Pharmaceutics 2021, Vol. 13, Page 723*, *13*(5), 723. https://doi.org/10.3390/PHARMACEUTICS13050723

Biswas, A., & De, S. (2021). Dynamic Tumor Heterogeneity and Cancer Progression: Drivers of dynamic intratumor heterogeneity and phenotypic plasticity. *American Journal of Physiology - Cell Physiology*, *320*(5), C750. https://doi.org/10.1152/AJPCELL.00575.2020

Bravo, A. M., Typas, A., & Veening, J. W. (2022). 2FAST2Q: a general-purpose sequence search and counting program for FASTQ files. *PeerJ*, *10*, e14041. https://doi.org/10.7717/PEERJ.14041/SUPP-11

Burguin, A., Diorio, C., & Durocher, F. (2021). Breast Cancer Treatments: Updates and New Challenges. *Journal of Personalized Medicine 2021, Vol. 11, Page 808*, *11*(8), 808. https://doi.org/10.3390/JPM11080808

Calaf, G. M., Crispin, L. A., Muñoz, J. P., Aguayo, F., Narayan, G., & Roy, D. (2022). Cell Adhesion Molecules Affected by Ionizing Radiation and Estrogen in an Experimental Breast Cancer Model. *International Journal of Molecular Sciences 2022, Vol. 23, Page 12674*, *23*(20), 12674. https://doi.org/10.3390/IJMS232012674

Campbell, L. L., & Polyak, K. (2007). Breast Tumor Heterogeneity: Cancer Stem Cells or Clonal Evolution? *Cell Cycle*, *6*(19), 2332–2338. https://doi.org/10.4161/CC.6.19.4914

Cedeño, M. M. C., Macias, J. R. P., Zambrano, A. P. E., Basurto, C. S. C., Chóez, K. S. V., Velez, J. N. M., Pita, J. G. M., Reyna, M. A. V., & Bastidas, S. E. E. (2024). Impact of primary prevention in the early diagnosis and mortality of breast cancer in Ecuador. *Revista Latinoamericana de Hipertension*, *19*(3), 129–133. https://doi.org/10.5281/ZENODO.10980345

Cha, S., Lee, E., & Won, H. H. (2021). Comprehensive characterization of distinct genetic alterations in metastatic breast cancer across various metastatic sites. *Npj Breast Cancer 2021 7:1*, *7*(1), 1–11. https://doi.org/10.1038/s41523-021-00303-y

Chang, S. C., & Ding, J. L. (2018). Ubiquitination and SUMOylation in the chronic inflammatory tumor microenvironment. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1870(2), 165–175. https://doi.org/10.1016/J.BBCAN.2018.08.002

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *IMeta*, *2*(2), e107. https://doi.org/10.1002/IMT2.107

Clark, A. G., Worni-Schudel, I. M., Korte, F. M., & Foster, M. H. (2017). A murine Ig light chain transgene reveals IGKV3 gene contributions to anti-collagen types IV and II specificities. *Molecular Immunology*, *91*, 49–56. https://doi.org/10.1016/J.MOLIMM.2017.08.015

Clevenger, J., Chavarro, C., Pearl, S. A., Ozias-Akins, P., & Jackson, S. A. (2015). Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. *Molecular Plant*, *8*(6), 831–846. https://doi.org/10.1016/J.MOLP.2015.02.002

Coupé, S., Giantsis, I. A., Vázquez Luis, M., Scarpa, F., Foulquié, M., Prévot, J. M., Casu, M., Lattos, A., Michaelidis, B., Sanna, D., García-March, J. R., Tena-Medialdea, J., Vicente, N., & Bunet, R. (2023). The characterization of toll-like receptor repertoire in Pinna nobilis after mass mortality events suggests adaptive introgression. *Ecology and Evolution*, *13*(8), e10383. https://doi.org/10.1002/ECE3.10383

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

Das, T., Andrieux, G., Ahmed, M., & Chakraborty, S. (2020). Integration of Online Omics-Data Resources for Cancer Research. *Frontiers in Genetics*, *11*, 578345. https://doi.org/10.3389/FGENE.2020.578345/BIBTEX

Debnath, P., Huirem, R. S., Dutta, P., & Palchaudhuri, S. (2022). Epithelial-mesenchymal transition and its transcription factors. Bioscience Reports, 42(1). https://doi.org/10.1042/BSR20211754/230017

Ding, S., Chen, X., & Shen, K. (2020). Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. *Cancer Communications*, *40*(8), 329–344. https://doi.org/10.1002/CAC2.12078

Doncheva, N. T., Morris, J. H., Holze, H., Kirsch, R., Nastou, K. C., Cuesta-Astroz, Y., Rattei, T., Szklarczyk, D., von Mering, C., & Jensen, L. J. (2023). Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks. *Journal of Proteome Research*, *22*(2), 637–646. https://doi.org/10.1021/ACS.JPROTEOME.2C00651/ASSET/IMAGES/LARGE/PR2C00651_0004.JPEG

Eidi, M., Abdolalizadeh, S., Moeini, S., Garshasbi, M., & Zahiri, J. (2024). 123VCF: an intuitive and efficient tool for filtering VCF files. *BMC Bioinformatics*, *25*(1), 1–7. https://doi.org/10.1186/S12859-024-05661-5/TABLES/2

Elbitar, S., Renard, M., Arnaud, P., Hanna, N., Jacob, M. P., Guo, D. C., Tsutsui, K., Gross, M. S., Kessler, K., Tosolini, L., Dattilo, V., Dupont, S., Jonquet, J., Langeois, M., Benarroch, L., Aubart, M., Ghaleb, Y., Abou Khalil, Y., Varret, M., … Abifadel, M. (2020). Pathogenic variants in THSD4, encoding the ADAMTS-like 6 protein, predispose to inherited thoracic aortic aneurysm. Genetics in Medicine 2020 23:1, 23(1), 111–122. https://doi.org/10.1038/s41436-020-00947-4

Ergin, S., Kherad, N., & Alagoz, M. (2022). RNA sequencing and its applications in cancer and rare diseases. Molecular Biology Reports, 49(3), 2325–2333. https://doi.org/10.1007/S11033-021-06963-0/TABLES/1

Feng, S., Wang, Z., Jin, Y., & Xu, S. (2024). TabDEG: Classifying differentially expressed genes from RNA-seq data based on feature extraction and deep learning framework. *PLOS ONE*, *19*(7), e0305857. https://doi.org/10.1371/JOURNAL.PONE.0305857

Franjić, S. (2023). A few words about genetic variation. *International Journal of Immunology Research*, *5*(2), 01–05. https://doi.org/10.33545/26648865.2023.V5.I2A.17

Frasca, A., Pavlidou, E., Bizzotto, M., Gao, Y., Balestra, D., Pinotti, M., Dahl, H. A., Mazarakis, N. D., Landsberger, N., & Kinali, M. (2022). Not Just Loss-of-Function Variations Identification of a Hypermorphic Variant in a Patient With a CDKL5 Missense Substitution. *Neurology: Genetics*, *8*(2). https://doi.org/10.1212/NXG.0000000000000666/ASSET/0E576F68-195A-45A2-B074-76BBB644B32C/ASSETS/GRAPHIC/NXG.0000000000000666TU1.JPEG

Fu, L., Lu, K., Jiao, Q., Chen, X., & Jia, F. (2023). The Regulation and Double-Edged Roles of the Deubiquitinase OTUD5. Cells 2023, Vol. 12, Page 1161, 12(8), 1161. https://doi.org/10.3390/CELLS12081161

Gerstberger, S., Jiang, Q., & Ganesh, K. (2023). Metastasis. *Cell*, *186*(8), 1564–1579. https://doi.org/10.1016/J.CELL.2023.03.003/ASSET/467121B0-7D42-4387-8CE9-F4A6E9442799/MAIN.ASSETS/GR5.JPG

Giordano, C., Camera, G. La, Gelsomino, L., Barone, I., Bonofiglio, D., Andò, S., & Catalano, S. (2020). The Biology of Exosomes in Breast Cancer Progression: Dissemination, Immune Evasion and Metastatic Colonization. *Cancers 2020, Vol. 12, Page 2179*, *12*(8), 2179. https://doi.org/10.3390/CANCERS12082179

Giorgi, F. M., Ceraolo, C., & Mercatelli, D. (2022). The R Language: An Engine for Bioinformatics and Data Science. *Life*, *12*(5), 648. https://doi.org/10.3390/LIFE12050648/S1

Gondane, A., & Itkonen, H. M. (2023). Revealing the History and Mystery of RNA-Seq. Current Issues in Molecular Biology, 45(3), 1860–1874. https://doi.org/10.3390/CIMB45030120/S1

Guo, L., Kong, D., Liu, J., Zhan, L., Luo, L., Zheng, W., Zheng, Q., Chen, C., & Sun, S. (2023). Breast cancer heterogeneity and its implication in personalized precision therapy. *Experimental Hematology & Oncology 2023 12:1*, *12*(1), 1–27. https://doi.org/10.1186/S40164-022-00363-1

Guo, W., Tzioutziou, N. A., Stephen, G., Milne, I., Calixto, C. P. G., Waugh, R., Brown, J. W. S., & Zhang, R. (2021). 3D RNA-seq: a powerful and flexible tool for rapid and accurate differential expression and alternative splicing analysis of RNA-seq data for biologists. *RNA Biology*, *18*(11), 1574–1587. https://doi.org/10.1080/15476286.2020.1858253

Gutiérrez-Galindo, E., Yilmaz, Z. H., & Hausser, A. (2023). Membrane trafficking in breast cancer progression: protein kinase D comes into play. *Frontiers in Cell and Developmental Biology*, *11*, 1173387. https://doi.org/10.3389/FCELL.2023.1173387/BIBTEX

Herreros, E., Janssens, X., Pepe, D., & De Keersmaecker, K. (2022). SNPs ability to influence disease risk: Breaking the silence on synonymous mutations in cancer. *Single Nucleotide Polymorphisms: Human Variation and a Coming Revolution in Biology and Medicine*, 77–96. https://doi.org/10.1007/978-3-031-05616-1_5/FIGURES/2

Huang, H., Du, J., Jin, B., Pang, L., Duan, N., Huang, C., Hou, J., Yu, W., Hao, H., & Li, H. (2021). Combination of Urine Exosomal mRNAs and lncRNAs as Novel Diagnostic Biomarkers for Bladder Cancer. *Frontiers in Oncology*, *11*, 667212. https://doi.org/10.3389/FONC.2021.667212/BIBTEX

Huang, Y., Hong, W., & Wei, X. (2022). The molecular mechanisms and therapeutic strategies of EMT in tumor progression and metastasis. *Journal of Hematology & Oncology 2022 15:1*, *15*(1), 1–27. https://doi.org/10.1186/S13045-022-01347-8

Hussen, B. M., Abdullah, S. T., Salihi, A., Sabir, D. K., Sidiq, K. R., Rasul, M. F., Hidayat, H. J., Ghafouri-Fard, S., Taheri, M., & Jamali, E. (2022). The emerging roles of NGS in clinical oncology and personalized medicine. Pathology - Research and Practice, 230, 153760. https://doi.org/10.1016/J.PRP.2022.153760

Ibrahim, M. K., Kalsom Yusof, U., Abdullah -, R., Nasser Boraik Ali, A., Pyar Ali Hassan, H., Bahamish -, H., Guo, J., Gao, J., & Liu, Z. (2022). HISAT2 Parallelization Method Based on Spark Cluster. *Journal of Physics: Conference Series*, *2179*(1), 012038. https://doi.org/10.1088/1742-6596/2179/1/012038

Ionkina, A. A., Balderrama-Gutierrez, G., Ibanez, K. J., Phan, S. H. D., Cortez, A. N., Mortazavi, A., & Prescher, J. A. (2021). Transcriptome analysis of heterogeneity in mouse model of metastatic breast cancer. *Breast Cancer Research*, *23*(1), 1–16. https://doi.org/10.1186/S13058-021-01468-X/FIGURES/7

Jiang, J., Wang, Y., Sun, M., Luo, X., Zhang, Z., Wang, Y., Li, S., Hu, D., Zhang, J., Wu, Z., Chen, X., Zhang, B., Xu, X., Wang, S., Xu, S., Huang, W., & Xia, L. (2024). SOX on tumors, a comfort or a constraint? *Cell Death Discovery 2024 10:1*, *10*(1), 1–18. https://doi.org/10.1038/s41420-024-01834-6

Jiang, X., Yang, L., Gao, Q., Liu, Y., Feng, X., Ye, S., & Yang, Z. (2022). The Role of RAB GTPases and Its Potential in Predicting Immunotherapy Response and Prognosis in Colorectal Cancer. Frontiers in Genetics, 13, 828373. https://doi.org/10.3389/FGENE.2022.828373/BIBTEX

Kamatani, Y., & Nakamura, Y. (2021). Genetic variations in medical research in the past, at present and in the future. *Proceedings of the Japan Academy, Series B*, *97*(6), 324–335. https://doi.org/10.2183/PJAB.97.018

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., & Ishiguro-Watanabe, M. (2013). KEGG: biological systems database as a model of the real world. *Nucleic Acids Research*, *1*(1256879), 13–14. https://doi.org/10.1093/NAR/GKAE909

Khan, S. U., Fatima, K., & Malik, F. (2022). Understanding the cell survival mechanism of anoikis-resistant cancer cells during different steps of metastasis. *Clinical and Experimental Metastasis*, *39*(5), 715–726. https://doi.org/10.1007/S10585-022-10172-9/TABLES/1

Kim, C. Y., Baek, S., Cha, J., Yang, S., Kim, E., Marcotte, E. M., Hart, T., & Lee, I. (2022). HumanNet v3: an improved database of human gene networks for disease research. *Nucleic Acids Research*, *50*(D1), D632–D639. https://doi.org/10.1093/NAR/GKAB

Kim, G., Kim, N., Lee, J., Park, M., & Kang, K. (2020). Differentially Expressed Genes in Matched Normal, Cancer, and Lymph Node Metastases Predict Clinical Outcomes in Patients With Breast Cancer. *Applied Immunohistochemistry & Molecular Morphology*, *28*(2), 111. https://doi.org/10.1097/PAI.0000000000000717

Koca, D., Séraudie, I., Jardillier, R., Cochet, C., Filhol, O., & Guyon, L. (2023). COL7A1 Expression Improves Prognosis Prediction for Patients with Clear Cell Renal Cell Carcinoma Atop of Stage. Cancers, 15(10), 2701. https://doi.org/10.3390/CANCERS15102701/S1

Kong, D., Mao, J. H., Li, H., Wang, J. Y., Li, Y. Y., Wu, X. C., Re, G. F., Luo, H. Y., Kuang, Y. Q., & Wang, K. H. (2022). Effects and associated transcriptomic landscape changes of methamphetamine on immune cells. *BMC Medical Genomics*, *15*(1), 1–15. https://doi.org/10.1186/S12920-022-01295-9/FIGURES/6

Kutschera, V. E., Kierczak, M., van der Valk, T., von Seth, J., Dussex, N., Lord, E., Dehasque, M., Stanton, D. W. G., Khoonsari, P. E., Nystedt, B., Dalén, L., & Díez-del-Molino, D. (2022). GenErode: a bioinformatics pipeline to investigate genome erosion in endangered and extinct species. *BMC Bioinformatics*, *23*(1), 1–17. https://doi.org/10.1186/S12859-022-04757-0/FIGURES/3

Kvokačková, B., Remšík, J., Jolly, M. K., & Souček, K. (2021). Phenotypic Heterogeneity of Triple-Negative Breast Cancer Mediated by Epithelial–Mesenchymal Plasticity. *Cancers 2021, Vol. 13, Page 2188*, *13*(9), 2188. https://doi.org/10.3390/CANCERS13092188

Lau, K. H. ;, Tan, A. M. ;, Shi, Y., Lau, K. H., Tan, A. M., & Shi, Y. (2022). New and Emerging Targeted Therapies for Advanced Breast Cancer. *International Journal of Molecular Sciences 2022, Vol. 23, Page 2288*, *23*(4), 2288. https://doi.org/10.3390/IJMS23042288

Lepucki, A., Orlińska, K., Mielczarek-Palacz, A., Kabut, J., Olczyk, P., & Komosińska-Vassev, K. (2022). The Role of Extracellular Matrix Proteins in Breast Cancer. *Journal of Clinical Medicine 2022, Vol. 11, Page 1250*, *11*(5), 1250. https://doi.org/10.3390/JCM11051250

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

Li, D., Wang, T. W., Aratani, S., Omori, S., Tamatani, M., Johmura, Y., & Nakanishi, M. (2023). Transcriptomic characterization of Lonrf1 at the single-cell level under pathophysiological conditions. *The Journal of Biochemistry*, *173*(6), 459–469. https://doi.org/10.1093/JB/MVAD021

Li, Q., Li, J., Yu, C. peng, Chang, S., Xie, L. ling, & Wang, S. (2021). Synonymous mutations that regulate translation speed might play a non-negligible role in liver cancer development. *BMC Cancer*, *21*(1), 1–13. https://doi.org/10.1186/S12885-021-08131-W/FIGURES/7

Liu, Y., & Guo, W. (2021). SOX factors as cell-state regulators in the mammary gland and breast cancer. *Seminars in Cell & Developmental Biology*, *114*, 126–133. https://doi.org/10.1016/J.SEMCDB.2021.01.002

Loor, C., & Recalde, R. (2024). Characterization of breast cancer detection programs in the Americas region with a focus on Ecuador. In *Salud, Ciencia y Tecnologia* (Vol. 4). Editorial Salud, Ciencia y Tecnologia. https://doi.org/10.56294/saludcyt2024741

Lujano Olazaba, O., Farrow, J., & Monkkonen, T. (2024). Fibroblast heterogeneity and functions: insights from single-cell sequencing in wound healing, breast cancer, ovarian cancer and melanoma. *Frontiers in Genetics*, *15*, 1304853. https://doi.org/10.3389/FGENE.2024.1304853/BIBTEX

Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers 2021, Vol. 13, Page 4287*, *13*(17), 4287. https://doi.org/10.3390/CANCERS13174287

Luo, H., Hong, R., Xu, Y., Zheng, Q., Xia, W., Lu, Q., Jiang, K., Xu, F., Chen, M., Shi, D., Deng, W., & Wang, S. (2023). Construction and validation of a macrophage polarization-related prognostic index to predict the overall survival in patients with early-stage triple-negative breast cancer. *Gland Surgery*, *12*(2), 225–242. https://doi.org/10.21037/GS-23-6/COIF

Lüönd, F., Tiede, S., & Christofori, G. (2021). Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. *British Journal of Cancer 2021 125:2*, *125*(2), 164–175. https://doi.org/10.1038/s41416-021-01328-7

Majed, S. O., & Mustafa, S. A. (2021). MACE-Seq-based coding RNA and TrueQuant-based small RNA profile in breast cancer: tumor-suppressive miRNA-1275 identified as a novel marker. *BMC Cancer*, *21*(1), 1–13. https://doi.org/10.1186/S12885-021-08218-4/FIGURES/7

Majidpoor, J., & Mortezaee, K. (2021). Steps in metastasis: an updated review. *Medical Oncology 2021 38:1*, *38*(1), 1–17. https://doi.org/10.1007/S12032-020-01447-W

Makhijani, P., Basso, P. J., Chan, Y. T., Chen, N., Baechle, J., Khan, S., Furman, D., Tsai, S., & Winer, D. A. (2023). Regulation of the immune system by the insulin receptor in health and disease. *Frontiers in Endocrinology*, *14*, 1128622. https://doi.org/10.3389/FENDO.2023.1128622/BIBTEX

Maran, M. I. J., & Davis G., D. J. (2022). Benefits of merging paired-end reads before pre-processing environmental metagenomics data. *Marine Genomics*, *61*, 100914. https://doi.org/10.1016/J.MARGEN.2021.100914

Martínez-Gregorio, H., Rojas-Jiménez, E., Mejía-Gómez, J. C., Díaz-Velásquez, C., Quezada-Urban, R., Vallejo-Lecuona, F., de la Cruz-Montoya, A., Porras-Reyes, F. I., Pérez-Sánchez, V. M., Maldonado-Martínez, H. A., Robles-Estrada, M., Bargalló-Rocha, E., Cabrera-Galeana, P., Ramos-Ramírez, M., Chirino, Y. I., Herrera, L. A., Terrazas, L. I., Frecha, C., Oliver, J., … Vaca-Paniagua, F. (2021). The evolution of clinically aggressive triple-negative breast cancer shows a large mutational diversity and early metastasis to lymph nodes. *Cancers*, *13*(20), 5091. https://doi.org/10.3390/CANCERS13205091/S1

Martins, Y. C., Bhawsar, P. M., Balasubramanian, J. B., Russ, D., Wong, W. S., Maass, W., & Almeida, J. S. (2024). FAIR privacy-preserving operation of large genomic variant calling format (VCF) data without download or installation. *AMIA Summits on Translational Science Proceedings*, *2024*, 65.

Morandi, V., Petrik, J., & Lawler, J. (2021). Endothelial Cell Behavior Is Determined by Receptor Clustering Induced by Thrombospondin-1. *Frontiers in Cell and Developmental Biology*, *9*, 664696. https://doi.org/10.3389/FCELL.2021.664696/BIBTEX

Morgan, J. J., & Crawford, L. J. (2021). The Ubiquitin Proteasome System in Genome Stability and Cancer. *Cancers 2021, Vol. 13, Page 2235*, *13*(9), 2235. https://doi.org/10.3390/CANCERS13092235

Mousavian, Z., Khodabandeh, M., Sharifi-Zarchi, A., Nadafian, A., & Mahmoudi, A. (2021). StrongestPath: a Cytoscape application for protein-protein interaction analysis. *BMC Bioinformatics*, *22*(1), 352. https://doi.org/10.1186/S12859-021-04230-4/TABLES/3

Mrozek, D., Stępień, K., Grzesik, P., & Małysiak-Mrozek, B. (2021). A Large-Scale and Serverless Computational Approach for Improving Quality of NGS Data Supporting Big Multi-Omics Data Analyses. *Frontiers in Genetics*, *12*, 699280. https://doi.org/10.3389/FGENE.2021.699280/BIBTEX

Mullan, K. A., Bramberger, L. M., Munday, P. R., Goncalves, G., Revote, J., Mifsud, N. A., Illing, P. T., Anderson, A., Kwan, P., Purcell, A. W., & Li, C. (2021). ggVolcanoR: A Shiny app for customizable visualization of differential expression datasets. *Computational and Structural Biotechnology Journal*, *19*, 5735–5740. https://doi.org/10.1016/j.csbj.2021.10.020

Müller, H., Jimenez-Heredia, R., Krolo, A., Hirschmugl, T., Dmytrus, J., Boztug, K., & Bock, C. (2017). VCF.Filter: interactive prioritization of disease-linked genetic variants from sequencing data. *Nucleic Acids Research*, *45*(W1), W567–W572. https://doi.org/10.1093/NAR/GKX425

Muzellec, B., Teleńczuk, M., Cabeli, V., & Andreux, M. (2023). PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics*, *39*(9). https://doi.org/10.1093/BIOINFORMATICS/BTAD547

Nan, B. Y., Xiong, G. F., Zhao, Z. R., Gu, X., & Huang, X. S. (2021). Comprehensive Identification of Potential Crucial Genes and miRNA-mRNA Regulatory Networks in Papillary Thyroid Cancer. *BioMed Research International*, *2021*(1), 6752141. https://doi.org/10.1155/2021/6752141

Nathanson, S. D., Detmar, M., Padera, T. P., Yates, L. R., Welch, D. R., Beadnell, T. C., Scheid, A. D., Wrenn, E. D., & Cheung, K. (2021). Mechanisms of breast cancer metastasis. *Clinical & Experimental Metastasis 2021 39:1*, *39*(1), 117–137. https://doi.org/10.1007/S10585-021-10090-2

Naumov, V., Balashov, I., Lagutin, V., Borovikov, P., & Alexeev, A. (2017). VolcanoR - web service to produce volcano plots and do basic enrichment analysis. *BioRxiv*, 165100. https://doi.org/10.1101/165100

Neagu, A. N., Whitham, D., Bruno, P., Morrissiey, H., Darie, C. A., & Darie, C. C. (2023). Omics-Based Investigations of Breast Cancer. *Molecules 2023, Vol. 28, Page 4768*, *28*(12), 4768. https://doi.org/10.3390/MOLECULES28124768

Nedaeinia, R., Najafgholian, S., Salehi, R., Goli, M., Ranjbar, M., Nickho, H., Haghjooy Javanmard, S., A. Ferns, G., & Manian, M. (2024). The role of cancer-associated fibroblasts and exosomal miRNAs-mediated intercellular communication in the tumor microenvironment and the biology of carcinogenesis: a systematic review. *Cell Death Discovery 2024 10:1*, *10*(1), 1–20. https://doi.org/10.1038/s41420-024-02146-5

Negi, A., Shukla, A., Jaiswar, A., & Shrinet, J. (2022). Applications and challenges of microarray and RNA-sequencing. Bioinformatics: Methods and Applications, 91–103. https://doi.org/10.1016/B978-0-323-89775-4.00016-X

Nolan, E., Lindeman, G. J., & Visvader, J. E. (2023). Deciphering breast cancer: from biology to the clinic. *Cell*, *186*(8), 1708–1728. https://doi.org/10.1016/J.CELL.2023.01.040

Nonnast, E., Mira, E., & Mañes, S. (2024). Biomechanical properties of laminins and their impact on cancer progression. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1879*(6), 189181. https://doi.org/10.1016/J.BBCAN.2024.189181

Nunn, A., Otto, C., Fasold, M., Stadler, P. F., & Langenberger, D. (2022). Manipulating base quality scores enables variant calling from bisulfite sequencing alignments using conventional bayesian approaches. *BMC Genomics*, *23*(1), 1–10. https://doi.org/10.1186/S12864-022-08691-6/FIGURES/4

OConnell, T. M. (2024). Pathway Volcano: An interactive tool for pathway guided visualization of differential expression data. *BioRxiv*, 2024.11.21.624662. https://doi.org/10.1101/2024.11.21.624662

Ota, M., Nagafuchi, Y., Hatano, H., Ishigaki, K., Terao, C., Takeshima, Y., Yanaoka, H., Kobayashi, S., Okubo, M., Shirai, H., Sugimori, Y., Maeda, J., Nakano, M., Yamada, S., Yoshida, R., Tsuchiya, H., Tsuchida, Y., Akizuki, S., Yoshifuji, H., … Fujio, K. (2021). Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell*, *184*(11), 3006-3021.e17. https://doi.org/10.1016/J.CELL.2021.03.056/ASSET/F8CECE9F-4A15-4998-B02A-01FBC72504A1/MAIN.ASSETS/GR7.JPG

Park, M., Kim, D., Ko, S., Kim, A., Mo, K., & Yoon, H. (2022). Breast Cancer Metastasis: Mechanisms and Therapeutic Implications. *International Journal of Molecular Sciences 2022, Vol. 23, Page 6806*, *23*(12), 6806. https://doi.org/10.3390/IJMS23126806

Patel, S. A., Rodrigues, P., Wesolowski, L., & Vanharanta, S. (2020). Genomic control of metastasis. *British Journal of Cancer 2020 124:1*, *124*(1), 3–12. https://doi.org/10.1038/s41416-020-01127-6

Peng, F., Zhou, L., Lu, W., Gan, R., Lu, M., Li, Z., Zhang, X., Wang, Y., & Gui, J. (2023). Genomic and Transcriptional Profiles of Kelch-like (klhl) Gene Family in Polyploid Carassius Complex. International Journal of Molecular Sciences, 24(9), 8367. https://doi.org/10.3390/IJMS24098367/S1

Petrosino, M., Novak, L., Pasquo, A., Chiaraluce, R., Turina, P., Capriotti, E., & Consalvi, V. (2021). Analysis and interpretation of the impact of missense variants in cancer. *International Journal of Molecular Sciences*, *22*(11), 5416. https://doi.org/10.3390/IJMS22115416/S1

Pollecker, K., Sylvester, M., & Voos, W. (2021). Proteomic analysis demonstrates the role of the quality control protease LONP1 in mitochondrial protein aggregation. Journal of Biological Chemistry, 279(4), 101134. https://doi.org/10.1016/J.JBC.2021.101134/ATTACHMENT/DE01DA8E-B985-4A80-8005-E8A976BF1B28/MMC5.XLSX

Pomatto, L. C. D., Raynes, R., & Davies, K. J. A. (2017). The peroxisomal Lon protease LonP2 in aging and disease: functions and comparisons with mitochondrial Lon protease LonP1. Biological Reviews, 92(2), 739–753. https://doi.org/10.1111/BRV.12253

Pruneda, J. N., Damgaard, R. B., Pruneda, J. N., & Damgaard, R. B. (2021). Ester-linked ubiquitination by HOIL-1 controls immune signalling by shaping the linear ubiquitin landscape. *The FEBS Journal*, *288*(20), 5903–5908. https://doi.org/10.1111/FEBS.16118

Qian, F., Kong, W., & Wang, S. (2022). Exploring autophagy-related prognostic genes of Alzheimer's disease based on pathway crosstalk analysis. *Bosnian Journal of Basic Medical Sciences*, *22*(5), 751. https://doi.org/10.17305/BJBMS.2021.7019

Qiu, X., Li, X., Yan, Y., Cai, Y., Liang, Q., Peng, B., Xu, Z., Xiao, M., Xia, F., & Peng, J. (2022). Identification of m6A-Associated Gene DST as a Prognostic and Immune-Associated Biomarker

in Breast Cancer Patients. *International Journal of General Medicine*, *15*, 523–534. https://doi.org/10.2147/IJGM.S344146

Qu, J., Li, J., Zhang, Y., He, R., Liu, X., Gong, K., Duan, L., Luo, W., Hu, Z., Wang, G., Xia, C., & Luo, D. (2021). AKR1B10 promotes breast cancer cell proliferation and migration via the PI3K/AKT/NF-κB signaling pathway. *Cell and Bioscience*, *11*(1), 1–13. https://doi.org/10.1186/S13578-021-00677-3/FIGURES/6

Rao, X. (2024). Plant Transcriptome Analysis with HISAT–StringTie–Ballgown and TopHat–Cufflinks Pipelines. *Methods in Molecular Biology*, *2812*, 203–213. https://doi.org/10.1007/978-1-0716-3886-6_11

Rasul, H. O., Aziz, B. K., Ghafour, D. D., & Kivrak, A. (2022). In silico molecular docking and dynamic simulation of eugenol compounds against breast cancer. *Journal of Molecular Modeling*, *28*(1), 1–18. https://doi.org/10.1007/S00894-021-05010-W/FIGURES/17

Richards, A. L., Eckhardt, M., & Krogan, N. J. (2021). Mass spectrometry-based protein–protein interaction networks for the study of human diseases. Molecular Systems Biology, 17(1), 8792. https://doi.org/10.15252/MSB.20188792/ASSET/801FB2B6-84BF-4997-A2EA-1B57BAC8BD9F/ASSETS/GRAPHIC/MSB20188792-FIG-0004-M.PN

Riggio, A., Varley, K., & Welm, A. (2020). The lingering mysteries of metastatic recurrence in breast cancer. *British Journal of Cancer 2020 124:1*, *124*(1), 13–26. https://doi.org/10.1038/s41416-020-01161-4

Rogiers, A., Lobon, I., Spain, L., & Turajlic, S. (2022). The Genetic Evolution of Metastasis. *Cancer Research*, *82*(10), 1849–1857. https://doi.org/10.1158/0008-5472.CAN-21-3863/694766/P/THE-GENETIC-EVOLUTION-OF-METASTASISTHE-GENETIC

Sameer, A. S., Banday, M. Z., & Nissar, S. (2021). Mutations and Polymorphisms: What Is The Difference? *Genetic Polymorphism and Cancer Susceptibility*, 1–21. https://doi.org/10.1007/978-981-33-6699-2_1

Sedillo, J. C., & Cryns, V. L. (2022). Targeting the methionine addiction of cancer. *American Journal of Cancer Research*, *12*(5), 2249. https://pmc.ncbi.nlm.nih.gov/articles/PMC9185618/

Selvaggio, G., Chaouiya, C., & Janody, F. (2021). In silico logical modelling to uncover cooperative interactions in cancer. *International Journal of Molecular Sciences*, *22*(9), 4897. https://doi.org/10.3390/IJMS22094897/S1

Shayakhmetov, R. D. (2024). Tools for gene regulatory network analysis and network and pathway visualization. Current Research, 67.

Shetty, R., Noland, R., Nandi, G., & Suzuki, C. K. (2024). Powering down the mitochondrial LonP1 protease: a novel strategy for anticancer therapeutics. *Expert Opinion on Therapeutic Targets*, *28*(1–2), 9–15. https://doi.org/10.1080/14728222.2023.2298358

Shrihastini, V., Muthuramalingam, P., Adarshan, S., Sujitha, M., Chen, J. T., Shin, H., & Ramesh, M. (2021). Plant Derived Bioactive Compounds, Their Anti-Cancer Effects and In Silico Approaches

as an Alternative Target Treatment Strategy for Breast Cancer: An Updated Overview. *Cancers 2021, Vol. 13, Page 6222*, *13*(24), 6222. https://doi.org/10.3390/CANCERS13246222

Sinkala, M. (2022). Mutational Landscape of Cancer-Driver Genes Across Human Cancers. *BioRxiv*, 2022.09.11.507448. https://doi.org/10.1101/2022.09.11.507448

Skowron-Kandzia, K., Tomsia, M., Koryciak-Komarska, H., Plewka, D., Wieczorek, P., & Czekaj, P. (2021). Gene Expression in Amnion-Derived Cells Cultured on Recombinant Laminin 332—A Preliminary Study. *Frontiers in Medicine*, *8*, 719899. https://doi.org/10.3389/FMED.2021.719899/BIBTEX

Smith, A. D., & De Sena, G. (2021). Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Research*, *8*, 1874. https://doi.org/10.12688/F1000RESEARCH.21142.2

Sola-Leyva, A., Andrés-Leóe, E., Molina, N. M., Terron-Camero, L. C., Plaza-Diaz, J., Sáez-Lara, M. J., Gonzalvo, M. C., Sánchez, R., Ruiz, S., Martinez, L., & Altmae, S. (2021). Mapping the entire functionally active endometrial microbiota. *Human Reproduction*, *36*(4), 1021–1031. https://doi.org/10.1093/HUMREP/DEAA372

Sondka, Z., Dhir, N. B., Carvalho-Silva, D., Jupe, S., Madhumita, McLaren, K., Starkey, M., Ward, S., Wilding, J., Ahmed, M., Argasinska, J., Beare, D., Chawla, M. S., Duke, S., Fasanella, I., Neogi, A. G., Haller, S., Hetenyi, B., Hodges, L., … Teague, J. (2024). COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Research*, *52*(D1), D1210–D1217. https://doi.org/10.1093/NAR/GKAD986

SoRelle, E. D., Reinoso-Vizcaino, N. M., Horn, G. Q., & Luftig, M. A. (2022). Epstein-Barr virus perpetuates B cell germinal center dynamics and generation of autoimmune-associated phenotypes in vitro. Frontiers in Immunology, 13, 1001145. https://doi.org/10.3389/FIMMU.2022.1001145/BIBTEX

Stevanovic, M., Drakulic, D., Lazic, A., Ninkovic, D. S., Schwirtlich, M., & Mojsin, M. (2021). SOX Transcription Factors as Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and Adult Neurogenesis. *Frontiers in Molecular Neuroscience*, *14*, 654031. https://doi.org/10.3389/FNMOL.2021.654031/BIBTEX

Sun, Q., Du, J., Dong, J., Pan, S., Jin, H., Han, X., & Zhang, J. (2022). Systematic Investigation of the Multifaceted Role of SOX11 in Cancer. *Cancers*, *14*(24), 6103. https://doi.org/10.3390/CANCERS14246103/S1

Tanaka, T., Warner, B. M., Michael, D. G., Nakamura, H., Odani, T., Yin, H., Atsumi, T., Noguchi, M., & Chiorini, J. A. (2022). LAMP3 inhibits autophagy and contributes to cell death by lysosomal membrane permeabilization. Autophagy, 18(7), 1629–1647. https://doi.org/10.1080/15548627.2021.1995150

Tang, G., Cho, M., & Wang, X. (2022). OncoDB: an interactive online database for analysis of gene expression and viral infection in cancer. Nucleic Acids Research, 50(D1), D1334–D1339. https://doi.org/10.1093/NAR/GKAB970

Tang, Q., Tan, P., Ma, N., & Ma, X. (2021). Physiological Functions of Threonine in Animals: Beyond Nutrition Metabolism. *Nutrients 2021, Vol. 13, Page 2592*, *13*(8), 2592. https://doi.org/10.3390/NU13082592

Tsang, S. M., Kim, H., Oliemuller, E., Newman, R., Boateng, N. A., Guppy, N., & Howard, B. A. (2021). Sox11 regulates mammary tumour-initiating and metastatic capacity in Brca1-deficient mouse mammary tumour cells. *DMM Disease Models and Mechanisms*, *14*(5). https://doi.org/10.1242/DMM.046037/261799

Turner, K. M., Yeo, S. K., Holm, T. M., Shaughnessy, E., & Guan, J. L. (2021). Heterogeneity within molecular subtypes of breast cancer. *American Journal of Physiology - Cell Physiology*, *321*(2), C343–C354. https://doi.org/10.1152/AJPCELL.00109.2021/ASSET/IMAGES/LARGE/AJPCELL.00109.2021_F003.JPEG

Vashishtha, K., Gaud, C., Andrews, S., & Krueger, C. (2024). Librarian: A quality control tool to analyse sequencing library compositions. *F1000Research*, *11*, 1122. https://doi.org/10.12688/F1000RESEARCH.125325.2

Wang, J., Li, J., Zhou, L., Hou, H., & Zhang, K. (2024). Regulation of epidermal barrier function and pathogenesis of psoriasis by serine protease inhibitors. *Frontiers in Immunology*, *15*, 1498067. https://doi.org/10.3389/FIMMU.2024.1498067/BIBTEX

Wang, L., Zhang, S., & Wang, X. (2021). The Metabolic Mechanisms of Breast Cancer Metastasis. *Frontiers in Oncology*, *10*, 602416. https://doi.org/10.3389/FONC.2020.602416/BIBTEX

Wang, T., Ruan, S., Zhao, X., Shi, X., Teng, H., Zhong, J., You, M., Xia, K., Sun, Z., & Mao, F. (2021). OncoVar: an integrated database and analysis platform for oncogenic driver variants in cancers. *Nucleic Acids Research*, *49*(D1), D1289–D1301. https://doi.org/10.1093/NAR/GKAA1033

Wesley, T., Berzins, S., Kannourakis, G., & Ahmed, N. (2021). The attributes of plakins in cancer and disease: perspectives on ovarian cancer progression, chemoresistance and recurrence. *Cell Communication and Signaling 2021 19:1*, *19*(1), 1–21. https://doi.org/10.1186/S12964-021-00726-X

Weverwijk, A., & Visser, K. E. (2023). Mechanisms driving the immunoregulatory function of cancer cells. *Nature Reviews Cancer 2023 23:4*, *23*(4), 193–215. https://doi.org/10.1038/s41568-022-00544-4

Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., & Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nature Communications 2020 11:1*, *11*(1), 1–19. https://doi.org/10.1038/s41467-020-18

Wlodawer, A., Sekula, B., Gustchina, A., & Rotanova, T. V. (2022). Structure and the Mode of Activity of Lon Proteases from Diverse Organisms. *Journal of Molecular Biology*, *434*(7), 167504. https://doi.org/10.1016/J.JMB.2022.167504

Wojas-Krawczyk, A. ;, Krawczyk, K. ;, Milanowski, P. ;, Bo˙, A., Wojas-Krawczyk, K., Krawczyk, P., & Milanowski, J. (2022). Tumor Microenvironment—A Short Review of Cellular and Interaction Diversity. *Biology 2022, Vol. 11, Page 929*, *11*(6), 929. https://doi.org/10.3390/BIOLOGY11060929

Xu, Y., Jin, Y., Gao, S., Wang, Y., Qu, C., Wu, Y., Ding, N., Dai, Y., Jiang, L., & Liu, S. (2022). Prognostic Signature and Therapeutic Value Based on Membrane Lipid Biosynthesis-Related Genes in Breast Cancer. *Journal of Oncology*, *2022*(1), 7204415. https://doi.org/10.1155/2022/7204415

Xue, J. Di, Xiang, W. F., Cai, M. Q., & Lv, X. Y. (2024). Biological functions and therapeutic potential of SRY related high mobility group box 5 in human cancer. *Frontiers in Oncology*, *14*, 1332148. https://doi.org/10.3389/FONC.2024.1332148/BIBTEX

Yamamoto, J., Ito, T., Yamaguchi, Y., & Handa, H. (2022). Discovery of CRBN as a target of thalidomide: a breakthrough for progress in the development of protein degraders. Chemical Society Reviews, 51(15), 6234–6250. https://doi.org/10.1039/D2CS00116K

Yang, Z., Jiang, S., Lu, C., Ji, T., Yang, W., Li, T., Lv, J., Hu, W., Yang, Y., & Jin, Z. (2019). SOX11: friend or foe in tumor prevention and carcinogenesis? *Therapeutic Advances in Medical Oncology*, *11*. https://doi.org/10.1177/1758835919853449/ASSET/IMAGES/LARGE/10.1177_1758835919853449-FIG3.JPEG

Yang, Y., Xiao, B., Feng, X., Chen, Y., Wang, Q., Fang, J., Zhou, P., Wei, X., & Cheng, L. (2023). Identification of hub genes and key signaling pathways by weighted gene co-expression network analysis for human aortic stenosis and insufficiency. *Frontiers in Cardiovascular Medicine*, *10*, 857578. https://doi.org/10.3389/FCVM.2023.857578/BIBTEX

Yang, H., Xiong, B., Xiong, T., Wang, D., Yu, W., Liu, B., & She, Q. (2022). Identification of key genes and mechanisms of epicardial adipose tissue in patients with diabetes through bioinformatic analysis. *Frontiers in Cardiovascular Medicine*, *9*, 927397. https://doi.org/10.3389/FCVM.2022.927397/BIBTEX

Ye, F., Dewanjee, S., Li, Y., Jha, N. K., Chen, Z. S., Kumar, A., Vishakha, Behl, T., Jha, S. K., & Tang, H. (2023). Advancements in clinical aspects of targeted therapy and immunotherapy in breast cancer. *Molecular Cancer 2023 22:1*, *22*(1), 1–40. https://doi.org/10.1186/S12943-023-01805-Y

Yin, G., Liu, L., Yu, T., Yu, L., Feng, M., Zhou, C., Wang, X., Teng, G., Ma, Z., Zhou, W., Ye, C., Zhang, J., Ji, C., Zhao, L., Zhou, P., Guo, Y., Meng, X., Fu, Q., Zhang, Q., … Yu, Z. (2024). Genomic and transcriptomic analysis of breast cancer identifies novel signatures associated with response to neoadjuvant chemotherapy. *Genome Medicine*, *16*(1), 1–19. https://doi.org/10.1186/S13073-024-01286-8/FIGURES/6

You, Y., Lai, X., Pan, Y., Zheng, H., Vera, J., Liu, S., Deng, S., & Zhang, L. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal Transduction and Targeted Therapy 2022 7:1*, *7*(1), 1–24. https://doi.org/10.1038/s41392-022-00994-0

Yu, L., & Gao, L. (2019). Human pathway-based disease network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *16*(4), 1240–1249. https://doi.org/10.1109/TCBB.2017.2774802

Zanini, G., Selleri, V., De Gaetano, A., Gibellini, L., Malerba, M., Mattioli, A. V., Nasi, M., Apostolova, N., & Pinti, M. (2022). Differential Expression of Lonp1 Isoforms in Cancer Cells. *Cells*, *11*(23), 3940. https://doi.org/10.3390/CELLS11233940/S1

Zeng, X., Shi, G., He, Q., & Zhu, P. (2021). Screening and predicted value of potential biomarkers for breast cancer using bioinformatics analysis. Scientific Reports 2021 11:1, 11(1), 1–11. https://doi.org/10.1038/s41598-021-00268-9

Zhang, Z., Richmond, A., & Yan, C. (2022). Immunomodulatory Properties of PI3K/AKT/mTOR and MAPK/MEK/ERK Inhibition Augment Response to Immune Checkpoint Blockade in Melanoma and Triple-Negative Breast Cancer. *International Journal of Molecular Sciences 2022, Vol. 23, Page 7353*, *23*(13), 7353. https://doi.org/10.3390/IJMS23137353

Zhang, B., Yuan, L., Chen, G., Chen, X., Yang, X., Fan, T., ... & Chen, Z. (2021). Deciphering Expression Profiling and Functional Signatures of Individualized Obesity-Associated Genes in Ossification of Ligamentum Flavum Through RNA-Sequence Data Mining.

Zheng, K., Lan, T., Li, G. ping, Huang, L., Chen, Y. peng, Su, B. H., Zhang, S., & Zheng, D. L. (2022). Evaluated expression of CELSR3 in oral squamous cell carcinoma is associated with perineural invasion and poor prognosis. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *133*(5), 564–573. https://doi.org/10.1016/J.OOOO.2021.10.016

Zhang, Z., Richmond, A., & Yan, C. (2022). Immunomodulatory Properties of PI3K/AKT/mTOR and MAPK/MEK/ERK Inhibition Augment Response to Immune Checkpoint Blockade in Melanoma and Triple-Negative Breast Cancer. *International Journal of Molecular Sciences 2022, Vol. 23, Page 7353*, *23*(13), 7353. https://doi.org/10.3390/IJMS23137353

Zheng, F., Wei, L., Zhao, L., & Ni, F. (2018). Pathway Network Analysis of Complex Diseases Based on Multiple Biological Networks. *BioMed Research International*, *2018*(1), 5670210. https://doi.org/10.1155/2018/5670210

Zheng, J., Zhao, Z., Wan, J., Guo, M., Wang, Y., Yang, Z., Li, Z., Ming, L., & Qin, Z. (2021). N-6 methylation-related lncRNA is potential signature in lung adenocarcinoma and influences tumor microenvironment. *Journal of Clinical Laboratory Analysis*, *35*(11), e23951. https://doi.org/10.1002/JCLA.23951

Zhou, H., Arapoglou, T., Li, X., Li, Z., Zheng, X., Moore, J., Asok, A., Kumar, S., Blue, E. E., Buyske, S., Cox, N., Felsenfeld, A., Gerstein, M., Kenny, E., Li, B., Matise, T., Philippakis, A., Rehm, H. L., Sofia, H. J., … Lin, X. (2023). FAVOR: functional annotation of variants online resource and

annotator for variation across the human genome. *Nucleic Acids Research*, *51*(D1), D1300–D1311. https://doi.org/10.1093/NAR/GKAC966

Zhou, W., & Zhang, J. (2024). Pan-cancer study on transition of signaling systems from primary to metastatic tumors. *BioRxiv*, 2024.11.11.619269. https://doi.org/10.1101/2024.11.11.619269

Zhou, Y., Zhang, Q., Zhao, Z., Hu, X., You, Q., & Jiang, Z. (2024). Targeting kelch-like (KLHL) proteins: achievements, challenges and perspectives. *European Journal of Medicinal Chemistry*, *269*, 116270. https://doi.org/10.1016/J.EJMECH.2024.116270

Zou, Z., & Zhang, J. (2021). Are Nonsynonymous Transversions Generally More Deleterious than Nonsynonymous Transitions? *Molecular Biology and Evolution*, *38*(1), 181–191. https://doi.org/10.1093/MOLBEV/MSAA200

ANEXOS

Anexo 1. Instalación de SRA Toolkit.

```
mkdir SRA_Toolkit
cd SRA_Toolkit
ls
wget --output-document sratoolkit.tar.gz https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/current/sratoolkit.current-ubuntu64.tar.gz
gunzip sratoolkit.current-ubuntu64.tar.gz
echo 'export PATH= E:/Proyecto/HerramientasBioinformaticas:$PATH' >> ~/.zshrc
source ~/.zshrc
vdb-config –interactive
```

Anexo 2. Descarga de los FASTQ.

```
Prefetch SRRXXXXXXX
fastq-dump --split-files SRRXXXXXXX
```

Anexo 3. Descarga de genoma de referencia y anotaciones.

Ensembl FTP Server:

ftp://ftp.ensembl.org/pub/

Genoma de referencia:

```
wget ftp://ftp.ensembl.org/pub/release-87/fasta/homo_sapiens/dna/Homo_sapiens.GRCh37.87.dna.primary_assembly.fa.gz
```

Anotaciones:

```
wget ftp://ftp.ensembl.org/pub/release-87/gff3/
```

Anexo 4. Instalación de FASTQC.

```
/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"
echo 'eval "$(/E:/Proyecto/homebrew/bin/brew shellenv)"' >> ~/.zshrc
source ~/.zshrc
brew --version
brew install FASTQC
```

Anexo 5. Instalación de Trimmomatic.

```
brew install trimmomatic
trimmomatic --version
brew --version
```

java -jar trimmomatic-0.39.jar

Anexo 6. Análisis de calidad.
Fastqc SRRXXXXXXX

Anexo 7. Reporte de análisis de calidad de datos crudos.
https://utaedu-my.sharepoint.com/:f:/g/personal/scaicedo1182_uta_edu_ec/En01eM6uvWxIjjyytayawnEBK1wH7Hnic82mIpUfNoL4HA?e=babDXL

Anexo 8. Recorte y mejoramiento de secuencias.
java -jar /Users/samantha/trimming/Trimmomatic-0.39/trimmomatic-0.39.jar PE SRRXXXXXXXX_1.fastq SRRXXXXXXXX_2.fastq forward_paired_1.fastq forward_unpaired_1.fastq reverse_paired_1.fastq reverse_unpaired_1.fastq ILLUMINACLIP:/Users/samantha/Desktop/Adapters.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:150

Anexo 9. Reporte de análisis de calidad de datos mejorados.
https://utaedu-my.sharepoint.com/:f:/g/personal/scaicedo1182_uta_edu_ec/EgL9mjIhBWtKonusNk39dPMBD7mOzOnJ0gfCCAEW8gJ4YA?e=kDLxrM

Anexo 8. Archivo de adaptadores.
https://utaedu-my.sharepoint.com/:b:/g/personal/scaicedo1182_uta_edu_ec/EQfeZ0ouyzBCrFnaaSD4dHUB-2FMkenzabDWawaX9Fvvfw?e=YYCsxW

Anexo 9. Instalación de hisat2, samtools, subread y feature counts
brew install hisat2
brew install samtools
brew install subread

Anexo 10. Creación de índices.
hisat2-build Homo_sapiens.GRCh37.dna.alt.fa Homo_sapiens.GRCh37.dna.alt_index

Anexo 11. Alineamiento de secuencias con el genoma de referencia.

hisat2 -x Homo_sapiens.GRCh37.dna.alt_index -1 /Volumes/Samy/Proyecto/GL_M1/output_forward_paired.fastq -2 /Volumes/Samy/Proyecto/GL_M1/output_reverse_paired.fastq -S GL_M1.sam

Anexo 12. Conversión del formato de archivo SAM en archivo BAM.

samtools view -bS GL_M1.sam | samtools sort -o GL_M1_sorted.bam

Anexo 13. Generación de la matriz de expresión.

featureCounts -p -a /Volumes/Samy/Proyecto/Referencia/GRCh37/Homo_sapiens.GRCh37.87.gff3 -o GLvsC_Conteo.txt -t exon -g Name -s 2 *.bam

Anexo 14. Matriz de expresión.

https://utaedu-my.sharepoint.com/:f:/g/personal/scaicedo1182_uta_edu_ec/Ek12WDrCuqNDsa1Oq8lG_EUBeGWQilx08rC9l-8z9I4iMg?e=qb997S

Anexo 15. Establecimiento del repositorio de trabajo.

setwd("/Volumes/Samy/Proyecto/VS/GL_vs_Control")
getwd()

Anexo 16. Librerías.

library(DESeq2)
library(ggplot2)
library(reshape2)
library(utils)
library(magrittr)
library(dplyr)
library(factoextra)
library(clusterProfiler)
library(EnsDb.Hsapiens.v86)
library(org.Hs.eg.db)
library(GEOquery)
library(biomaRt)
library(tibble)
library(igraph)
library(BSgenome.Hsapiens.UCSC.hg19)
library(TxDb.Hsapiens.UCSC.hg19.knownGene)

```
library(GenomicFeatures)
library(pheatmap)
library(rentrez)
library(GOstats)
library(AnnotationDbi)
```

Anexo 17. Carga de matriz de expresión al entorno R.

```
countData_GLvsC <- read.table("/Volumes/Samy/Proyecto/VS/GL_vs_Control/GLvsC_Conteo.txt", header = TRUE, sep = "\t", row.names = 1)
view(countData_GLvsC)
```

Anexo 18. Establecimiento de las condiciones de estudio.

```
colData_GLvsC <- data.frame(condition = c("GL", "GL", "GL", "Control", "Control", "Control"), row.names = colnames(countData_GLvsC))
colData_GLvsC$condition <- as.factor(colData_GLvsC$condition)
view(colData_GLvsC)
```

Anexo 19. Identificación de genes significativos.

```
dds_GLvsC <- DESeqDataSetFromMatrix(countData = countData_GLvsC, colData = colData_GLvsC, design = ~ condition)
dds_GLvsC <- DESeq(dds_GLvsC)
result_GLvsC <- results(dds_GLvsC)
result_GLvsC <- data.frame(GENEID = row.names(result_GLvsC), log2BaseMean = log2(result_GLvsC$baseMean), log2Ratio = result_GLvsC$log2FoldChange, STDERR_log2Ratio = result_GLvsC$lfcSE, pvalue = result_GLvsC$pvalue, padjust = result_GLvsC$padj)
result_GLvsC <- na.omit(result_GLvsC)
view(result_GLvsC)
sign_GLvsC_exp <- subset(result_GLvsC,result_GLvsC$log2BaseMean > 0 & result_GLvsC$padjust < 0.01)
view(sign_GLvsC_exp)
```

Anexo 20. Representación de los genes diferencialmente expresados en Volcano Plot.

```
ggplot(result_TPvsC, aes(x = log2BaseMean, y = -log10(padjust))) +
  geom_point() +
  geom_point(data = sign_TPvsC_exp, aes(x = log2BaseMean, y = -log10(padjust)), color = 'red') +
  labs(title = "Tumor Primario", x = "Log2BaseMean", y = "-log10(padjust)")
```

Anexo 21. Conversión de Exon ID a Ensembl Gene ID

```
exon_ids_GLvsControl <- c("ENSE00001535281", "ENSE00001282530", "ENSE00001946367", "ENSE00001002414")
ensembl <- useMart("ENSEMBL_MART_ENSEMBL",
                   dataset = "hsapiens_gene_ensembl",
                   host = "https://grch37.ensembl.org")
conversion <- getBM(attributes = c("ensembl_exon_id", "ensembl_gene_id"),
                    filters = "ensembl_exon_id",
                    values = exon_ids_GLvsControl,
                    mart = ensembl)
view(conversion)
```

Anexo 22. Conversión de Ensembl Gene ID a ENTREZ IDs

```
ensembl <- useMart("ENSEMBL_MART_ENSEMBL",
                   dataset = "hsapiens_gene_ensembl",
                   host = "https://grch37.ensembl.org")
gene_ids <- c("ENSG00000151914", "ENSG00000176887", "ENSG00000187720", "ENSG00000151914")
conversion_entrez <- getBM(attributes = c("ensembl_gene_id", "entrezgene_id"),
                    filters = "ensembl_gene_id",
                    values = gene_ids,
                    mart = ensembl)
print(conversion_entrez)
```

Anexo 23. Redes de interacción.

https://utaedu-my.sharepoint.com/:f:/g/personal/scaicedo1182_uta_edu_ec/Eq3seo1Y_71NrqRLzNVasxkBm4Liss0BGSCyIOvKtS44tw?e=UVcsK0

Anexo 25. Informe de Enriquecimiento Funcional de los genes diferencialmente expresados.

https://utaedu-my.sharepoint.com/:f:/g/personal/scaicedo1182_uta_edu_ec/Ei74KxjFhYBHvX52fN97YUEB4Kqopm8M9cFkvdUUvqhapg?e=OiFJ5b

Anexo 26. Informe de Enriquecimiento Funcional de genes identificados en Cytoscape.

https://utaedu-my.sharepoint.com/:f:/g/personal/scaicedo1182_uta_edu_ec/EqnIspphXh5KvGoJ7h4VA_IBYFy1ae9fvthBvtE3NrwKfg?e=DN8oVR

Anexo 27. Instalación de vcftools

brew install vcftools

vcftools --version

Anexo 27. Corrección del formato VCF.

vcf-validator /Volumes/Samy/VCF/GL.vcf

grep -v -E 'GL000191.1|GL000192.1|GL000194.1|GL000195.1|GL000199.1|GL000205.1|GL000217.1|GL000218.1|GL000219.1|GL000220.1|GL000221.1|GL000223.1|GL000224.1|GL000241.1|GL000242.1' /Volumes/Samy/VCF/GL.vcf > /Volumes/Samy/VCF/GL_corregido.vcf

vcf-validator /Volumes/Samy/VCF/GL_corregido.vcf

Anexo 28. Archivo SnpEff eff.

https://utaedu-my.sharepoint.com/:f:/g/personal/scaicedo1182_uta_edu_ec/Enfe6W5Cm7VDpzrrJWs9P5YBimNMqIL3nrgLibJeDssLuQ?e=AJBe1H