

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTAMENTO DE BIOTECNOLOGÍA



*Un enfoque de biología de sistemas para
integrar la investigación sobre los
mecanismos moleculares, celulares y
tisulares del daño del ADN para mejorar la
predicción del riesgo de cáncer.*

TRABAJO FIN DE MÁSTER EN BIOTECNOLOGÍA BIOMÉDICA

ALUMNA: MARTA MARTÍNEZ SANTA

TUTOR UPV: RAFAEL SIRERA PÉREZ

COTUTOR EXTERNO: JOAN CLIMENT BATALLER

Curso académico: 2019/2020

VALENCIA, DICIEMBRE 2020

ÍNDICE

1. Introducción.....	3
1.1. Incidencia, mortalidad y supervivencia en cáncer de mama	3
1.2. Clasificación del cáncer de mama	4
1.3. Sistemas modelo en cáncer de mama	9
1.4. Deleción de <i>TP53</i> y daño en el DNA en cáncer de mama.....	10
2. Objetivos	12
3. Materiales y métodos.....	13
3.1. Modelo animal (generado por el <i>Lawrence Berkeley National Laboratory</i>).....	13
3.2. Inmunohistoquímica (IHC).....	13
3.3. Análisis de RNA-seq	14
3.4. Análisis estadístico	17
4. Resultados y discusión	18
4.1. Comparación del crecimiento tumoral entre grupos.....	18
4.2. Clasificación histológica de los tumores	19
4.3. Caracterización molecular de los tumores.....	20
4.4. Calidad de las secuencias Fasta	27
4.5. Resultados del alineamiento y el conteo.....	28
4.6. Análisis de expresión diferencial irradiados vs. no irradiados	28
4.7. Análisis de enriquecimiento en categorías funcionales	35
5. Conclusión.....	39
6. Anexos.....	40
Anexo I. Especificaciones de los anticuerpos primarios empleados para la IHC.....	40
Anexo II. Sistema de puntuación utilizado en la evaluación de los distintos biomarcadores por IHC.	41
Anexo III. Códigos utilizados en el análisis de RNA-seq.	42
Anexo IV. Ejemplos de inmunotinción positiva para los biomarcadores analizados en los TMAs.	50
Anexo V. Correlaciones detectadas entre los distintos biomarcadores analizados.	51
7. Bibliografía.....	54

1. INTRODUCCIÓN

1.1. INCIDENCIA, MORTALIDAD Y SUPERVIVENCIA EN CÁNCER DE MAMA

El cáncer de mama es el tumor más frecuente entre las mujeres residentes en países industrializados y constituye la primera causa de muerte por cáncer en mujeres de todo el mundo (*Figura 1*). Cada año se diagnostican más de 2 millones de nuevos casos de cáncer de mama a nivel global, lo que supone prácticamente un cuarto de los cerca de 8,6 millones de cánceres diagnosticados anualmente en mujeres (Bray et al., 2018) (*Figura 1A*). Se estima que, solo en España, durante el año 2020 se habrán diagnosticado cerca de 32.953 casos nuevos (Santaballa, 2020).

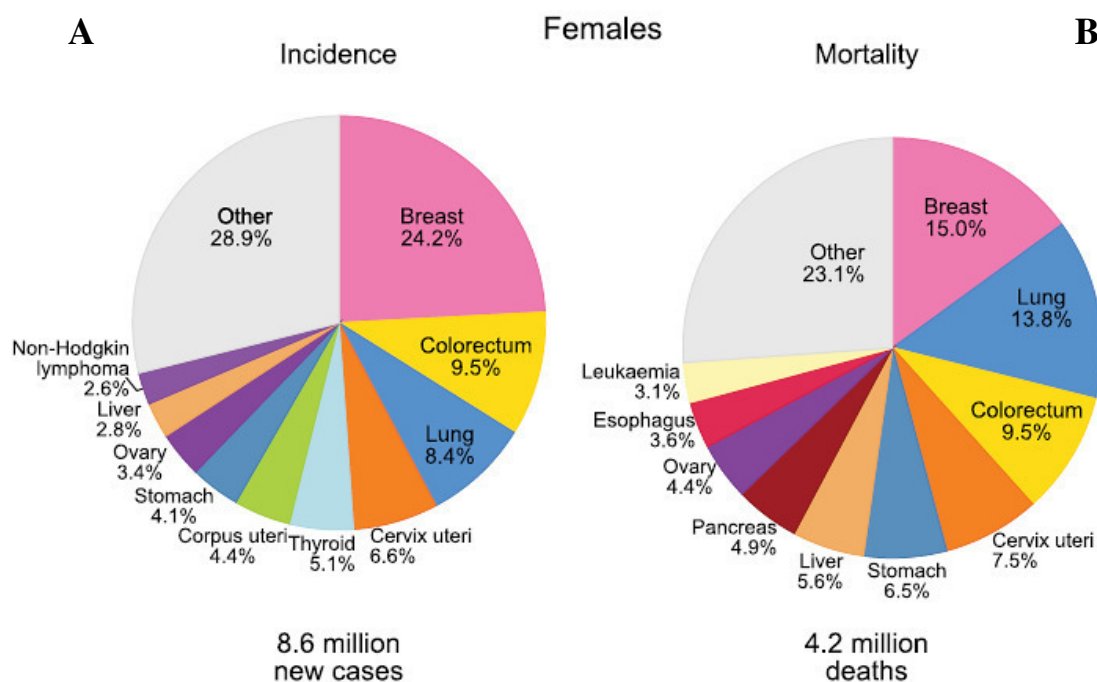


Figura 1. Incidencia (A) y mortalidad (B) estimadas (en %) de los cánceres más comunes en mujeres para el año 2018 a nivel mundial (Bray et al., *CA: a Cancer Journal for Clinicians*, 2018).

Se calcula que, aproximadamente, 1 de cada 8 mujeres, es decir, un 13% de la población femenina, será diagnosticada de un cáncer de mama invasivo a lo largo de su vida; y 1 de cada 38 (alrededor de un 3%) morirá como consecuencia del cáncer que padecía (American Cancer Society, 2020).

En los países más desarrollados, gracias a las pruebas de detección temprana, a un mejor acceso a la información y a las mejoras terapéuticas, las tasas de mortalidad por cáncer de mama se han mantenido estables en las mujeres menores de 50 años desde el año 2007, y han disminuido en las mujeres de edad más avanzada, hasta un 1,3%, entre 2013 y 2017 (American Cancer Society, 2020). Aun así, alrededor de un 20-30% de las pacientes desarrolla una metástasis que resulta incurable (Eroles et al., 2012). Se estima que en 2018 murieron 627.000 mujeres en el mundo por cáncer de mama, aproximadamente un 15% de todas las muertes femeninas por cáncer (World Health Organization, 2020) (*Figura 1B*). En España, aunque la mortalidad por cáncer de

mama ha descendido en los últimos años gracias a los programas de cribado y a la mejora de los tratamientos, el cáncer de mama sigue siendo la primera causa de muerte por cáncer en mujeres, con 6.534 fallecimientos en 2018 (Santaballa, 2020).

Aunque las causas desencadenantes de un cáncer de mama todavía no están muy claras, sí se han identificado numerosos factores de riesgo que aumentan las probabilidades de padecerlo. Algunos están relacionados con el estilo de vida, como la obesidad, el tabaquismo, la ingesta de alcohol, la inactividad física, el uso de terapia hormonal sustitutiva o la decisión de no tener descendencia; mientras que hay otros factores sobre los que no podemos actuar, como son una historia familiar previa de cáncer de mama, predisposición genética por mutaciones heredadas en los genes *BRCA1* y *BRCA2*, antecedentes personales de cáncer de mama, niveles de hormonas endógenas elevados, aparición temprana de la primera menstruación, menopausia tardía, exposición a radiaciones ionizantes y otros carcinógenos, así como la edad y el origen étnico (American Cancer Society, 2020). La incidencia de cáncer de mama aumenta con la edad, situándose la media de diagnóstico en los 62 años (American Cancer Society, 2019). No obstante, aproximadamente un 10% de los casos se diagnostican en mujeres menores de 40 años (Santaballa, 2020). Respecto a la etnia, la incidencia es mucho mayor entre mujeres de raza blanca, mientras que la tasa de mortalidad es superior en mujeres de raza negra (American Cancer Society, 2019).

La supervivencia media relativa de las mujeres a los 5 años tras el diagnóstico es del 91%, aunque esta cifra depende del estadio en el que se encuentre el cáncer en el momento del diagnóstico (99% para el estadio I *versus* 27% para el estadio IV), así como del subtipo molecular de cáncer (92% para el subtipo luminal A *versus* 77% para el cáncer de mama triple negativo) (American Cancer Society, 2019).

1.2. CLASIFICACIÓN DEL CÁNCER DE MAMA

El cáncer de mama no es una enfermedad única, sino que comprende un abanico muy heterogéneo de entidades biológicamente diferentes con características genéticas, histopatológicas e implicaciones clínicas muy divergentes, que exhiben comportamientos muy diferentes y hacen que su pronóstico y su respuesta terapéutica sean distintos. Esto hace que la agrupación del cáncer de mama en subtipos sea necesaria a la hora de decidir el tratamiento más adecuado para cada paciente en particular (Dai et al., 2015).

Existen distintas clasificaciones del cáncer de mama en función de diversos parámetros:

1.2.1. Clasificación histopatológica

La clasificación histológica se basa en el tamaño, forma y organización de las células tumorales cuando se observan al microscopio. El cáncer de mama puede iniciarse en diferentes partes del tejido mamario. Dependiendo del tipo de tejido en el que se origina el tumor, el cáncer

de mama se clasifica en, al menos, 21 subtipos histológicos distintos que difieren en su forma de presentación, respuesta al tratamiento y resultado clínico (American Cancer Society, 2019).

El cáncer de mama se puede dividir en dos grupos principales: los carcinomas y los sarcomas. La mayoría de los cánceres de mama son carcinomas, y más concretamente adenocarcinomas, es decir, tumores que se originan en las células epiteliales que revisten los ductos por los que fluye la leche desde los lóbulos hasta el pezón (carcinomas ductales) o en los lóbulos, donde se encuentran las glándulas encargadas de producir la leche durante la lactancia (carcinomas lobulares). Otros tipos de carcinomas menos frecuentes son los carcinomas medular, mucinoso, tubular, papilar, mixto (tiene características de ductal y de lobulillar) o metaplásico (incluyendo los carcinomas de células escamosas y de células fusiformes), etc. Respecto a los sarcomas, son cánceres raros (representan menos del 1% de los cánceres de mama) que pueden originarse en las células musculares, en el tejido adiposo o en el tejido conectivo del seno (tumores filodes), o en las células que revisten los vasos sanguíneos o linfáticos cercanos a la mama (angiosarcomas) (American Cancer Society, 2020).

En función del tipo de células afectadas y del poder infiltrante del tumor, el cáncer de mama se puede clasificar en diferentes subtipos (American Cancer Society, 2020):

- **Carcinomas no invasivos o *in situ***: son aquellos que no se han extendido más allá de las células donde se han originado.
 - **Carcinoma lobulillar *in situ* (LCIS)**: las células cancerosas se originan en el revestimiento de los lobulillos que contienen las glándulas productoras de leche del seno, pero no atraviesan la pared de los mismos. Es considerado un carcinoma benigno y sin potencial de progresar a cáncer invasivo (American Cancer Society, 2019). Sin embargo, aumenta el riesgo de padecer cáncer de mama en cualquiera de los senos.
 - **Carcinoma ductal *in situ* (DCIS)**: constituye el 16% de todos los casos de cáncer de mama diagnosticados (American Cancer Society, 2019). Las células que revisten los ductos han sufrido una transformación a células cancerosas, pero no se han propagado hacia el tejido mamario adyacente. Sin embargo, si no se trata, el DCIS tiene la capacidad de progresar a un cáncer invasivo, extendiéndose hacia tejidos cercanos y, a partir de ahí, causar metástasis en otras partes del cuerpo.
- **Carcinomas invasivos**: constituyen alrededor del 81% de los casos y son aquellos en los que el cáncer se ha extendido hacia el tejido mamario adyacente. Los tipos más comunes son:
 - **Carcinoma lobulillar invasivo (ILC)**. Aproximadamente 1 de cada 10 cánceres de seno invasivos son ILCs. El ILC comienza en las glándulas productoras de leche y puede atravesar la pared de los lobulillos e invadir el tejido mamario adyacente. En este punto puede tener la capacidad de propagarse hacia otras partes del cuerpo a través del sistema linfático y el torrente sanguíneo.
 - **Carcinoma ductal invasivo (IDC)**: es el tipo más común de cáncer de mama, representando cerca del 80% de todos los carcinomas invasivos. El IDC comienza en las células que revisten los conductos de leche en el seno y es capaz de invadir la pared de

los mismos y crecer en los tejidos mamarios cercanos. Al igual que el ILC, puede hacer metástasis a otras partes del cuerpo.

- Tipos especiales de carcinoma de mama invasivo: son poco comunes, ya que cada tipo constituye menos del 5% de todos los cánceres de mama. El cáncer suele nombrarse de acuerdo a la manera en que las células están agrupadas a vista de microscopio. Engloba los ya citados carcinoma medular, mucinoso, tubular, papilar, mixto, metaplásico, etc.

1.2.2. Clasificación molecular

Las técnicas inmunohistoquímicas permiten clasificar el cáncer de mama en varios subgrupos en función de la expresión de diversos genes y receptores que determinan las mejores opciones de tratamiento y ayudan a predecir el pronóstico de la enfermedad (*Figura 2*), el riesgo de recurrencia, de metástasis a distancia y la supervivencia (Majumder, Jagani y Basu, 2020).

Actualmente, atendiendo a la expresión de receptores hormonales (HR) (receptor de estrógenos, ER, y receptor de progesterona, PR), del receptor 2 del factor de crecimiento epidérmico humano (HER2 ó ErbB2) y del marcador de proliferación Ki67 (antígeno nuclear que se expresa en todas las fases del ciclo celular, excepto en la fase G0) (Majumder, Jagani and Basu, 2020), el cáncer de mama se clasifica en cuatro subtipos moleculares principales:

- **Luminal A.** Son tumores ER+ y/o PR+, HER2-, que tienen un índice de proliferación bajo (Ki67 < 14%). Es el subtipo más común de cáncer de mama y el menos agresivo debido a su baja tasa de crecimiento (American Cancer Society, 2019). Suelen presentar un pronóstico más favorable porque generalmente responden a terapia hormonal, que impide que el estrógeno y la progesterona se unan a sus receptores y estimulen el crecimiento y la propagación del cáncer (American Cancer Society, 2020).
- **Luminal B.** Son tumores ER+ y/o PR+, HER2- y con un índice Ki67 \geq 14%; o tumores ER+ y/o PR+, HER2+ y de cualquier Ki67. Están caracterizados por presentar un índice proliferativo bastante elevado, sean o no HER2+, lo que los hace más agresivos y, por tanto, están asociados a un peor pronóstico en comparación con los del tipo luminal A (American Cancer Society, 2019). La estrategia terapéutica más exitosa para estos casos combina la quimioterapia y el tratamiento hormonal (Dai et al., 2015).
- **Enriquecido en HER2/ErbB2 (ER-PR-HER2+).** Son cánceres que no tienen receptores de estrógeno ni de progesterona y que sobreexpresan la proteína HER2, promotora del crecimiento celular. Se asocian con índices proliferativos aumentados (Ki67 \geq 14%), metástasis y recurrencia (Fallah et al., 2017). No obstante, en los últimos años, el desarrollo de terapias dirigidas contra el receptor HER2 ha mejorado el pronóstico de estos pacientes (American Cancer Society, 2020).
- **Triple negativo (TNBC).** Constituyen cerca del 10-15% de los casos de cáncer de mama y se caracterizan por la ausencia de expresión de receptores hormonales y de amplificación de la proteína HER2 (ER-PR-HER2-). A menudo, se asocian con mutaciones en el gen *BRCA1*.

El cáncer de mama triple negativo crece y se propaga más rápidamente ($Ki67 > 14\%$) que los otros tipos de cáncer descritos. Es el subtipo con el peor pronóstico, pues no tiene un objetivo farmacológico definido. Dado que las células cancerosas carecen de estos receptores, las opciones de tratamiento son limitadas, ni la terapia hormonal ni los tratamientos anti-HER2 son útiles, por lo que la quimioterapia es la principal opción de tratamiento sistémico. Recientemente, se ha descrito que aproximadamente un 20% de los TNBC expresan PD-L1, una proteína transmembrana que ejerce una acción supresora sobre el sistema inmune y cuya expresión en las células tumorales se asocia con su evasión del sistema inmune (Guo et al., 2016), por lo que constituye una potente diana terapéutica para la inmunoterapia (American Cancer Society, 2020).

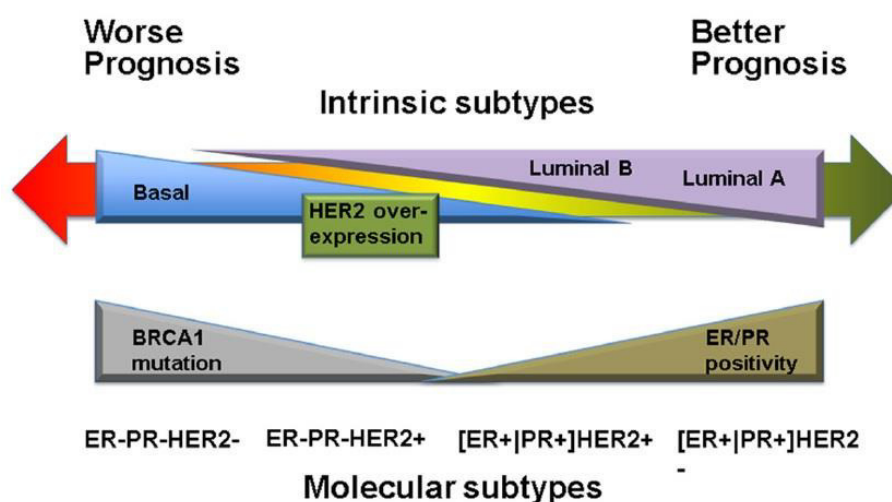


Figura 2. Pronóstico esperado en pacientes con cáncer de mama en función del subtipo molecular (Dai et al., *American Journal of Cancer Research*, 2015).

1.2.3. Clasificación en función del grado tumoral

El grado del cáncer hace referencia a cómo de similar es en apariencia, a nivel microscópico, el tejido tumoral al tejido mamario normal. Conocer el grado de un cáncer de mama es importante para entender cómo de rápido es probable que crezca y se propague, decidir qué tratamientos podrían ser más eficaces y predecir su resultado clínico. Atendiendo al grado de diferenciación de las células tumorales, el cáncer de mama se clasifica en 3 subgrupos (American Cancer Society, 2020):

- **Grado 1 o bien diferenciado:** hace referencia a un cáncer de crecimiento más lento, con menos capacidad de propagación y en el que las células se parecen más al tejido mamario normal.
- **Grado 2 o moderadamente diferenciado:** la velocidad de crecimiento de las células es mayor que en el caso anterior y presentan un aspecto menos diferenciado.

- **Grado 3 o pobremente diferenciado:** las células cancerosas son muy diferentes a las células normales y muestran un crecimiento más rápido, por lo que es más probable que el tumor se propague.

1.2.4. Clasificación en función del estadio del cáncer

El pronóstico del cáncer de mama depende, principalmente, de la extensión de la enfermedad en el momento del diagnóstico (Santaballa, 2020). El estadio de un cáncer describe cuánto cáncer hay en el cuerpo, es decir, si se ha propagado y a qué distancia, y ayuda a determinar su gravedad, así como las mejores opciones de tratamiento (American Cancer Society, 2020).

Para el proceso de estadificación se emplea el sistema TNM del *American Joint Committee on Cancer* (AJCC), que clasifica el cáncer de mama en 5 estadios diferentes (del 0 al IV), basándose en el análisis de 3 parámetros principales: el tamaño del tumor (T) y su extensión a los ganglios linfáticos regionales (N) o a otras partes del cuerpo (M). También puede emplearse el sistema de estadificación utilizado por el *Surveillance, Epidemiology and End Results* (SEER) Program, que divide el cáncer en 4 etapas (*in situ*, localizado, regional y distante) (American Cancer Society, 2020).

Teniendo en cuenta estos dos sistemas, y en función de la extensión del tumor, la *American Cancer Society* clasifica el cáncer de mama en:

- **Estadio 0 o *in situ*:** las células tumorales se limitan a la capa de células donde se originaron.
- **Estadio I o localizado:** el tumor, de pequeño tamaño, se ha extendido por el seno, pero no se ha diseminado fuera de la mama.
- **Estadio II:** son tumores más grandes que los del estadio I, pero que no se han propagado fuera del tejido mamario o cánceres pequeños que han alcanzado algún ganglio linfático adyacente.
- **Estadio III o regional:** las células tumorales se han diseminado a tejidos circundantes y/o varios ganglios linfáticos cercanos.
- **Estadio IV o distante:** el tumor ha metastatizado a órganos distantes.

En el momento del diagnóstico, aproximadamente el 64% de los casos se encuentra en un estadio local; el 27%, en estadio regional; y un 6% ha desarrollado metástasis a distancia (American Cancer Society, 2019).

1.3. SISTEMAS MODELO EN CÁNCER DE MAMA

Tabla 1. Principales modelos murinos empleados en la investigación del cáncer de mama.

MODELO	DESCRIPCIÓN	VENTAJAS/APLICACIONES	INCONVENIENTES
Xenotrasplante derivado de línea celular (CDX): CDX ectópico	Inoculación subcutánea de células de cáncer de mama humano en un hospedador murino inmunodeprimido.	-Fácil, rápido y económico. <u>Aplicaciones:</u> Evaluar la genética del cáncer y monitorizar el crecimiento tumoral.	-Hospedador inmunodeficiente: limita estudios de inmunoterapia. -No simula microambiente tumoral. -Salto de especie entre las células tumorales implantadas (humano) y el estroma (ratón). -Homogeneidad tumoral. -No metástasis. -No todas las líneas celulares pueden establecerse <i>in vivo</i> . -Presión selectiva por crecimiento celular <i>in vitro</i> .
	CDX ortotópico Las células tumorales se trasplantan directamente en el epitelio mamario del ratón.	-Microambiente apropiado. <u>Aplicaciones:</u> Estudiar fenotipos malignos y metastásicos.	-Complejidad técnica. -Hospedador inmunocomprometido. -Restricciones de especie.
Xenotrasplante derivado de paciente (PDX)	Un fragmento de tumor de paciente se implanta quirúrgicamente en el epitelio mamario de un hospedador murino inmunocomprometido.	-Refleja la heterogeneidad del cáncer de mama humano. -Mantiene cinética de crecimiento, características histológicas y perfil genético del tumor primario. -Retiene capacidad invasiva, metastásica y respuesta clínica a tratamientos farmacológicos. -No requiere cultivo <i>in vitro</i> . <u>Aplicaciones:</u> Identificación de biomarcadores de respuesta a tratamiento y estudios de resistencia terapéutica.	-Hospedador inmunodeficiente. -Invasividad del procedimiento quirúrgico. -Salto de especie entre las células tumorales y el estroma implantados (humano) y el estroma infiltrante (ratón). -Acceso a muestras de pacientes. -Tiempo prolongado y coste elevado.
Modelo singénico	Implantación de un tumor mamario de ratón en hospedador murino inmunocompetente.	-No restricciones de especie, no rechazo. -Sistema inmune intacto y microambiente tumoral específico de especie. <u>Aplicaciones:</u> Estudiar biología del cáncer, y evaluar eficacia y toxicidad de pequeñas moléculas antitumorales.	-Microambiente murino. -La especificidad de especie de los anticuerpos impide su evaluación en estos modelos.
GEMM: modelo murino modificado genéticamente	Inducción de tumores mamaros a través de la inactivación de genes supresores de tumores (<i>Trp53</i> , <i>Brca1</i>) o sobreexpresión de oncogenes (<i>ErbB2</i> , <i>Myc</i>) específicamente dentro del epitelio mamario del ratón.	-Formación espontánea del tumor en tejido apropiado, en el contexto de un sistema inmune intacto y un microambiente tumoral nativo. -Imita perfil genético y características fisiopatológicas de muchos subtipos de cáncer de mama humano. -Modela todas las etapas del tumor, incluida la metástasis. <u>Aplicaciones:</u> Estudiar desarrollo y progresión del tumor, su interacción con el microambiente, su respuesta a determinados fármacos, y definir mecanismos moleculares de resistencia.	-Microambiente murino. -No siempre reflejan la patología de la enfermedad humana. -Tiempo prolongado y coste elevado. -La especificidad de especie de los anticuerpos impide su evaluación en estos modelos.

A pesar de los recientes avances en el diagnóstico y tratamiento del cáncer de mama, investigaciones adicionales sobre los procesos de tumorigénesis, metástasis y resistencia a fármacos siguen siendo necesarias (M. K. Park, C. H. Lee y H. Lee, 2018). No obstante, la investigación del cáncer de mama se ve obstaculizada por la falta de modelos *in vivo* que recapitulen con precisión las características clínicas más relevantes de la enfermedad humana (Manning, Buck y Cook, 2016). Sin embargo, aunque todos los sistemas modelo constituyen una aproximación del cáncer de mama humano, la investigación en modelos animales ha sido fundamental para comprender los mecanismos moleculares implicados en el desarrollo y progresión de la enfermedad y ofrecer mejores terapias para ella. A día de hoy, hay disponible una gran cantidad de modelos que simulan diferentes subtipos y etapas de la enfermedad (Holen et al., 2017).

El ratón de laboratorio *Mus musculus* representa un buen sistema modelo para simular toda la secuencia de eventos que conducen al cáncer de mama en humanos y permite dilucidar algunos aspectos de la biología del cáncer que pueden no ser abordables en los pacientes. Entre las ventajas del ratón como sistema modelo encontramos que es un mamífero de pequeño tamaño, lo que facilita su manejo y permite un alojamiento económico, se reproduce rápidamente, tiene una esperanza de vida relativamente larga, y conocemos la secuencia completa de su genoma. Además, los ratones comparten muchas similitudes fisiológicas con los humanos y se usan comúnmente en los estudios de farmacocinética y toxicidad de fármacos. Irónicamente, para los estudios de imagen, el tamaño del ratón puede ser una limitación cuando el tumor es más pequeño que la resolución efectiva de la modalidad de imagen elegida (Manning et al., 2016).

En la última década se ha observado un crecimiento exponencial en el número de ratones utilizados para la investigación del cáncer de mama. En la *Tabla 1* se describen los modelos murinos *in vivo* más empleados, sus ventajas, sus limitaciones y sus principales aplicaciones (Manning et al., 2016; Holen et al., 2017; M. K. Park et al., 2018).

En resumen, aunque no existe un modelo preclínico perfecto para la investigación del cáncer de mama humano, los modelos murinos representan un escalón esencial entre los sistemas *in vitro* y los estudios clínicos, que permiten estudiar el inicio y progresión del cáncer, la invasión y la metástasis, así como evaluar nuevas terapias e investigar posibles mecanismos de resistencia (Holen et al., 2017). Además, constituyen una plataforma muy útil para la comprensión, caracterización y clasificación de los diferentes subtipos de cáncer de mama.

1.4. DELECIÓN DE *TP53* Y DAÑO EN EL DNA EN CÁNCER DE MAMA

El gen *TP53* es uno de los supresores tumorales más comúnmente mutado en los cánceres humanos. La delección del gen *TP53* conlleva la pérdida de las funciones de la proteína p53, encargada de mantener la integridad genómica de la célula. Cuando reconoce un daño en el DNA, p53 detiene el ciclo celular en el punto de control de la fase G1/S mediante la inducción del inhibidor de cinasa dependiente de ciclina p21. Esto permite a la célula tener tiempo para reparar el DNA y evitar la replicación de los cromosomas dañados. Si el daño es muy severo y no puede

ser reparado, p53 desencadena la apoptosis celular para evitar la proliferación de células que contienen DNA aberrante (Weiss et al., 2010).

Se ha observado que las células epiteliales mamarias *knockout* para el gen *TP53* tienen la capacidad de crecer en ausencia del factor de crecimiento epidérmico (EGF) y presentan defectos en los *checkpoints* de respuesta a daño en el DNA del ciclo celular. Además, la pérdida de p53 induce inestabilidad cromosómica, que conduce a transformación celular y selección de clones con fenotipos variables (Weiss et al., 2010).

Del mismo modo, ratones con una deleción homocigota en la línea germinal para el gen *Trp53* son susceptibles al desarrollo de neoplasias espontáneas tempranas (Weiss et al., 2010). De manera que, el epitelio mamario de un ratón $p53^{-/-}$ se puede trasplantar ortotópicamente en una hembra receptora, que acabará desarrollando un tumor mamario. Una característica de este modelo de trasplante singénico es la naturaleza agresiva del tumor, así como la variedad de fenotipos histológicos y subtipos moleculares que simula (Usary et al., 2016), por lo que, este sistema puede ser útil para comprender mejor los cambios fenotípicos específicos del cáncer de mama que acompañan a la pérdida de p53, así como para ayudar a proporcionar futuras estrategias terapéuticas para tumores malignos humanos que albergan p53 aberrante (Weiss et al., 2010).

2. OBJETIVOS

El presente proyecto es la continuación de un trabajo previo inacabado del *Lawrence Berkeley National Laboratory* en el que se utilizó un modelo de trasplante singénico de glándula mamaria p53^{-/-} de un ratón modificado genéticamente al que se había eliminado el gen *Trp53* en la línea germinal a un ratón *wild-type* que fue sometido posteriormente a radiación para estudiar los efectos del daño en el DNA en cáncer de mama.

El objetivo principal del trabajo es comprobar si la radiación no dirigida sobre el microambiente tumoral tiene algún efecto en el proceso de carcinogénesis en un epitelio mamario p53^{-/-}, mediante la consecución de dos objetivos secundarios:

- I. Caracterización molecular de los tumores irradiados y no irradiados en base a la expresión de 8 biomarcadores seleccionados como relevantes en la aparición, evolución y pronóstico del cáncer de mama mediante inmunohistoquímica (IHC). Análisis de las diferencias de expresión entre los dos grupos y estudio de las correlaciones entre biomarcadores.
- II. Análisis de expresión génica diferencial entre los tumores originados en ratones irradiados vs. ratones no irradiados mediante un análisis de RNA-seq. Estudio del papel de los genes diferencialmente expresados en el inicio, desarrollo o resultado clínico del cáncer.

3. MATERIALES Y MÉTODOS

3.1. MODELO ANIMAL (generado por el *Lawrence Berkeley National Laboratory*)

Se emplearon ratones hembra de la cepa BALB/c, que son sensibles a cancerígenos y, por tanto, susceptibles al desarrollo de cáncer (Zhang et al., 2015). A las 3 semanas de edad (justo antes del inicio de la pubertad y de que el árbol ductal empiece a extenderse por toda la mama) (Plante, Stewart y Laird, 2011), se les extirparon quirúrgicamente las glándulas mamarias inguinales del epitelio endógeno, y se les realizó un trasplante ortotópico con un fragmento de tumor mamario de un ratón singénico *knockout* para el gen *Trp53*, implicado en la reparación del DNA dañado, lo que los hace más susceptibles a la transformación oncogénica. Tras el trasplante, los ratones fueron divididos en dos grupos: unos recibieron radiación de 10 cGy (centigray) (dosis óptima para evitar alteraciones en la función ovárica y, por tanto, en los resultados, y para simular las exposiciones humanas a la radiación) (Nguyen et al., 2011) y otros que no fueron irradiados (*sham*). El desarrollo de los tumores se monitorizó por palpación hasta que alcanzaron un tamaño de 10 mm³, fecha en la que los ratones fueron sacrificados (*Figura 3*).

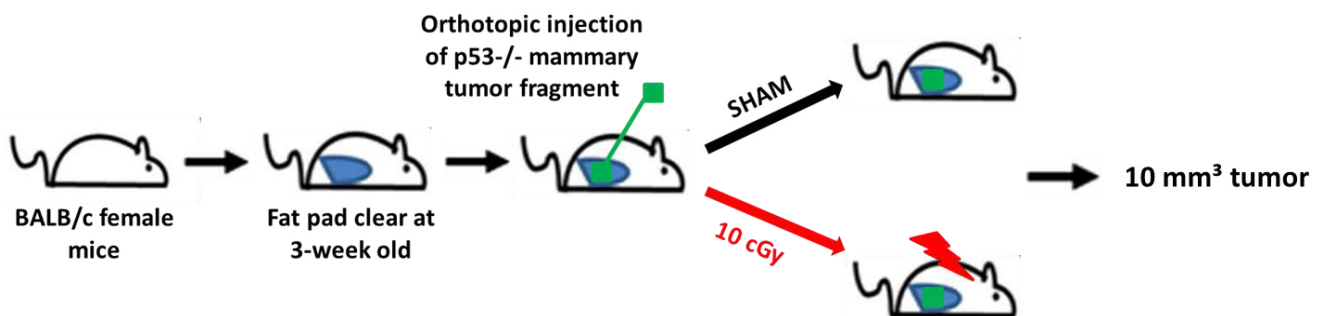


Figura 3. Esquema resumen del protocolo empleado para la inducción de tumores mamarios en ratones BALB/c.

3.2. INMUNOHISTOQUÍMICA (IHC)

Con el fin de estudiar el patrón de expresión de 8 de los biomarcadores de mayor relevancia en cáncer de mama (ER, PR, HER2, Ki67, c-Myc, CK14, CD14 y E-cadherina) se llevó a cabo un análisis inmunohistoquímico que nos permitía determinar tanto la localización subcelular donde se expresaba la proteína de interés, como la intensidad de expresión de la misma. La evaluación de la expresión de estos marcadores se realizó en 148 tumores de ratón, de los cuales 67 habían sido irradiados y 81 eran controles no irradiados.

El análisis inmunohistoquímico de los tumores se agilizó con la construcción, por parte del *Lawrence Berkeley National Laboratory*, de 6 TMAs (del inglés, *Tissue MicroArrays*), bloques de parafina en los que cada tumor estaba representado por duplicado. Con la ayuda de un micrótopo, se realizaron tantos cortes de 5 µm como fueron necesarios. Los cortes se desparafinaron en estufa a 37°C durante 24 horas. A continuación, se incubaron 20 minutos a temperatura ambiente en una solución de xilol concentrada, y seguidamente, se pasaron por xiloles a concentración más baja.

Con el fin de rehidratar las muestras, los cortes se pasaron por soluciones con porcentajes de alcohol decrecientes (90°, 80° y 70°, respectivamente), lavándolos finalmente con agua destilada.

El resto del proceso de IHC se realizó de forma automatizada con el sistema de inmunotinción VENTANA BenchMark de Roche, que lleva a cabo desde el proceso de desenmascaramiento antigénico por calor hasta la contratinción de los núcleos con hematoxilina. Los anticuerpos primarios utilizados, con sus correspondientes diluciones y kit de detección se relacionan en el **Anexo I**.

Para una mejor conservación de los tejidos, se procedió a la deshidratación de las muestras mediante sumersión en soluciones con porcentajes crecientes de alcohol (70°, 80° y 90° respectivamente) y, a continuación, en xilol. Finalmente, se montaron las muestras con un cubreobjetos con la ayuda del medio de montaje Entellan® de Sigma-Aldrich, quedando los tejidos fijados y evitando así su degradación con el tiempo.

La evaluación de los distintos biomarcadores fue realizada por un médico especialista en Anatomía Patológica con el dispositivo de obtención de microimágenes digitales Leica DMD108. La anotación de cada biomarcador se realizó utilizando los sistemas de puntuación descritos en la literatura (ver **Anexo II**).

3.3. ANÁLISIS DE RNA-SEQ

Para comprobar si los tumores que surgen en los hospedadores que reciben radiación de 10 cGy son diferentes a los que ocurren en los hospedadores *sham* en cuanto a los niveles de expresión de distintos genes, analizamos el RNA total de 162 tumores, de los cuales 73 surgieron en ratones irradiados y 89, en ratones no irradiados. Para ello, utilizamos las secuencias en formato FASTA procedentes de la extracción, purificación, conversión a cDNA, amplificación y secuenciación masiva del RNA aislado del ganglio linfático de las glándulas mamarias inguinales de cada ratón que nos proporcionó el *Lawrence Berkeley National Laboratory*. Todos los códigos utilizados en el análisis de RNA-seq se muestran en el **Anexo III**.

3.3.1. Control de calidad de secuencias FASTA

Los secuenciadores de nueva generación son capaces de producir decenas de millones de secuencias en una sola ejecución, pudiendo introducir errores en las secuencias que generan (MU DNA Core Facility, 2015). Las lecturas que contengan múltiples errores de secuenciación dificultarán la alineación contra el genoma de referencia y nos proporcionarán menos información biológica (University of Oregon, n.d.). Por este motivo, el análisis de datos de RNA-seq comienza siempre con la evaluación de la calidad de las lecturas obtenidas tras el proceso de secuenciación masiva y la filtración de aquellas con calidad aceptable para alinear contra el genoma de referencia.

El control de calidad de las secuencias FASTA se llevó a cabo con el programa FastQC (ver **Anexo III sección 1**), una herramienta que detecta si hay regiones problemáticas (originadas por el propio secuenciador o durante la preparación de las muestras) en los datos importados que puedan afectar al análisis posterior de las secuencias, y asigna una puntuación de calidad (Q) a todas las bases en cada posición de la lectura. Este valor Q estima la probabilidad de cada nucleótido secuenciado de ser erróneo, de manera que cuanto mayor sea la puntuación, menor será la probabilidad de error y mayor la calidad de la secuencia (MU DNA Core Facility, 2015).

3.3.2. Alineamiento y conteo

Las lecturas obtenidas tras el proceso de secuenciación masiva son secuencias *paired-end*, es decir, lecturas en las que la secuenciación se ha llevado a cabo desde los dos extremos del DNA molde con el fin de aumentar la cobertura de secuencia del cDNA de partida, por lo que de cada fragmento de DNA molde se obtienen dos secuencias: una directa o *forward* y una inversa o *reverse*, que son alineadas como pares de lecturas, lo que permite un alineamiento más preciso y con menor tasa de error (Illumina, 2017).

Las secuencias que presentaron una buena puntuación de calidad por nucleótido se mapearon contra el genoma de referencia de *Mus musculus* con el programa de alineamiento HISAT2. En el **Anexo III sección 2** se puede consultar el código empleado para generar los alineamientos.

HISAT2, al igual que otros *softwares* de alineamiento, actúa en dos pasos principales: en primer lugar, convierte el genoma de referencia en una “referencia” indexada y, a continuación, mapea las lecturas de interés contra este genoma indexado (University of Oregon, n.d.). Para ello, HISAT2 emplea tanto un índice FM (Ferragina-Manzini) global basado en grafos (GFM) que representa un genoma completo de la población general, como un conjunto de pequeños índices GFM que cubren colectivamente todo el genoma y sus variantes. El empleo de estos índices FM hace posible la compresión del genoma de referencia, reduciendo el espacio de almacenamiento requerido en el proceso de alineamiento. Además, dado que en la estructura de los grafos, cada ruta define una serie de bases que ocurren en un determinado orden en el genoma de referencia o en una de sus variantes, los índices GFM también permiten encontrar de manera eficiente el número de ocurrencias de un patrón determinado dentro del genoma comprimido, así como ubicar la posición de cada coincidencia, minimizando el tiempo del proceso de alineamiento (Kim et al., 2019).

El programa HISAT2 devuelve los resultados de los alineamientos en archivos con formato SAM, en los que cada alineamiento se representa en una sola línea. Cada columna del fichero SAM incluye un tipo de información asociada a cada alineamiento, como el nombre de la lectura, la posición del alineamiento, una puntuación numérica de la calidad del alineamiento, etc.

Las lecturas alineadas de las 162 muestras tumorales fueron asignadas a genes específicos empleando la función `featureCounts` del paquete `Rsubread`, el genoma de referencia de ratón “mm10” y la opción “`isPairedEnd`” seleccionada. La función `featureCounts` reconoce automáticamente los archivos en formato SAM y asigna cada alineamiento del fichero a un gen concreto. Para ello, dispone de la anotación del genoma de referencia, en este caso mm10, que

incluye las coordenadas cromosómicas de los exones de cada gen, por lo que solo tendrá en cuenta para el conteo aquellas lecturas que hayan mapeado contra exones. En el **Anexo III sección 3** se puede consultar el código de R empleado para efectuar el conteo de las secuencias.

3.3.3. Análisis de expresión diferencial

Una vez realizado el alineamiento, el conteo de las lecturas mapeadas contra cada región del genoma nos permite determinar el número de secuencias que han sido alineadas únicamente a un gen específico y, por tanto, estimar su nivel de expresión. No obstante, para realizar un análisis apropiado de expresión génica mediante RNA-seq, lo usual es llevar a cabo previamente un proceso de normalización de los datos de conteo para corregir una serie de sesgos específicos de bibliotecas y/o genes que introducen diferencias entre las muestras que estamos estudiando y dentro de las mismas. Una de las fuentes que más variabilidad genera entre muestras son las variaciones en la profundidad de secuenciación de cada muestra y, por tanto, en el tamaño de la biblioteca. De hecho, bibliotecas de tamaño más grande se asocian a recuentos más altos. Del mismo modo, las variaciones en la longitud de los genes influyen en el recuento de lecturas dentro de una misma muestra, ya que se observará mayor número de lecturas para genes más largos (Dillies et al., 2012).

Para el análisis de genes que se expresan diferencialmente en los tumores de ratones irradiados en comparación con los tumores de los ratones no irradiados utilizamos el programa DESeq2, incluido en el entorno de programación R. El paquete DESeq2 de Bioconductor está diseñado para la normalización, visualización y análisis diferencial de los datos de conteo de RNA-seq, basándose en el uso de modelos lineales generalizados con distribución binomial negativa (Bioconductor, 2020). El método de normalización de los datos de conteo empleado por DESeq2 se basa en la hipótesis de que la mayoría de los genes no se expresan diferencialmente, por lo que estos genes deberían tener recuentos de lecturas similares en todas las muestras. Para ello, el programa crea una muestra de pseudo-referencia cuyos datos de recuento equivalen a la media geométrica de todas las muestras. A continuación, para cada muestra, calcula el ratio n° lecturas gen X / n° lecturas gen X en la referencia (siendo X cualquier gen). Como no todos los genes se expresan diferencialmente, la mayoría de los genes de una misma muestra presentarán ratios muy similares y próximos a 1. La mediana de los ratios de todos los genes de una misma muestra se toma como factor de normalización para todas las lecturas de dicha muestra. Finalmente, para calcular los valores de recuento normalizados para cada gen en cada muestra, el programa divide cada valor de conteo sin procesar entre el factor de normalización calculado para esa muestra (Harvard Chan Bioinformatics Core, n.d.); de manera que, ya podemos disponer de datos de conteo normalizados para hacer comparaciones precisas de los niveles de expresión génica entre y dentro de las muestras de interés.

En el **Anexo III sección 4** se puede consultar el código de R empleado para efectuar los análisis de expresión diferencial.

3.3.4. Análisis de enriquecimiento en categorías funcionales

Los resultados del análisis de expresión diferencial se emplearon para determinar si los genes sobreexpresados e infraexpresados se encuentran enriquecidos en *pathways* o categorías funcionales concretas. Para ello, empleamos el método Gene Set Enrichment Analysis (GSEA). Este método parte de una lista de genes ordenada en función de los valores de *fold change* obtenidos en el análisis de expresión diferencial previo y de una base de datos de *pathways* o categorías funcionales. Básicamente, el método permite identificar si los genes asociados a un determinado *pathway* o categoría funcional tienden a estar situados en los extremos superiores o inferiores de la lista ordenada de genes. Esto se consigue mediante el cómputo de un *score* de enriquecimiento que se obtiene empleando un test similar al test ponderado de Kolmogorov-Smirnov. La información sobre *pathways* y categorías funcionales se obtuvo en la siguiente dirección: <http://bioinf.wehi.edu.au/MSigDB/v6.1/>. Como fuente de información de *pathways* seleccionamos los *pathways* canónicos C2. Esta colección de *pathways* aglutina la información de múltiples bases de datos, que incluye Reactome, KEGG y Biocarta, entre otras.

En el **Anexo III sección 4** se puede consultar el código de R empleado para efectuar el análisis de enriquecimiento en categorías funcionales.

3.4. ANÁLISIS ESTADÍSTICO

Además de los datos obtenidos por inmunohistoquímica y mediante el análisis de RNA-seq, para la realización de este proyecto disponíamos de otros datos clínicos (días de irradiación hasta que los tumores alcanzaron un tamaño de 10 mm³ y clasificación histológica de cada muestra tumoral) obtenidos previamente sobre las mismas muestras de ratón por el *Lawrence Berkeley National Laboratory*.

El análisis estadístico de los resultados de expresión de los distintos biomarcadores por IHC y de los datos cedidos por el *Lawrence Berkeley National Laboratory* se llevó a cabo con el *software* IBM SPSS Statistics, incluyendo prueba de la mediana, prueba exacta de Fisher y análisis de correlaciones bivariadas, y considerándose como significativas las diferencias cuando el p-valor era inferior a 0,05.

4. RESULTADOS Y DISCUSIÓN

4.1. COMPARACIÓN DEL CRECIMIENTO TUMORAL ENTRE GRUPOS

Tras la inducción tumoral, los ratones fueron mantenidos hasta que el tumor alcanzó el tamaño máximo permitido según los criterios éticos establecidos para evitar el sufrimiento del animal (10 mm^3). En general, los 129 tumores de los que disponíamos de datos tardaron, de media, 406 días en alcanzar los 10 mm^3 y no se observaron diferencias entre las medias de los dos grupos. No obstante, observamos que la mediana de días de irradiación hasta alcanzar los 10 mm^3 era de 420 días en los ratones que habían recibido radiación 10 cGy, frente a los 382 días de los ratones *sham* (Tabla 2; Figura 4).

Tabla 2. Datos de media y mediana de días de irradiación hasta que los tumores alcanzaron un tamaño de 10 mm^3 .

	Días IR- 10 mm^3	
	Media	Mediana
General	406	405
<i>Sham</i>	403	382
10 cGy	409	420

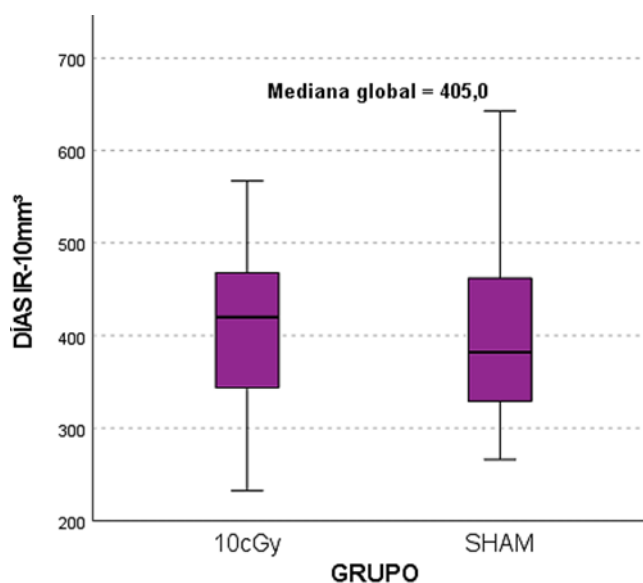


Figura 4. Distribución de los días de irradiación hasta los 10 mm^3 de tumor en función del grupo. Se muestran la mediana y los percentiles 25 y 75 de cada grupo.

La prueba de la mediana, que compara las medianas de ambos grupos y determina si existen diferencias significativas entre ellas, nos llevó a rechazar la hipótesis de que existen diferencias en el tiempo que tardan en desarrollarse los tumores en función de si los ratones reciben o no radiación post-trasplante ($p\text{-valor} = 0,333$) (Figura 5) y a concluir que el daño en el DNA a través de la exposición a radiación ionizante no tiene ningún efecto en la tasa de crecimiento de tumores $p53^{-/-}$. Este resultado difiere del reportado por Nguyen et al. (2011) en el que, empleando un modelo idéntico de ratón, pero en el que la radiación se administraba previa al trasplante, se observaron tasas de crecimiento tumoral mayores en los ratones irradiados frente a los ratones control (*sham*).

Frecuencias

		GROUP cod	
		SHAM	10 cGy
DÍAS IR-10mm ³	> Mediana	29	35
	<= Mediana	36	29

Estadísticos de prueba^a

		DÍAS IR-10mm ³
N		129
Mediana		405,00
Chi-cuadrado		1,309
gl		1
Sig. asin.		,253
Corrección de continuidad de Yates	Chi-cuadrado	,937
	gl	1
	Sig. asin.	,333

a. Variable de agrupación: GROUP cod

Figura 5. Resultados de la prueba de la mediana. Las diferencias entre las medianas de los grupos *sham* y 10 cGy no son significativas (p-valor > 0,05).

4.2. CLASIFICACIÓN HISTOLÓGICA DE LOS TUMORES

Los tumores fueron sometidos a tinción de hematoxilina-eosina por parte del *Lawrence Berkeley National Laboratory* para su clasificación histológica en 3 subtipos principales: adenocarcinomas, carcinomas de células escamosas y carcinomas de células fusiformes. De los 148 tumores analizados, 74, es decir, el 50% eran adenocarcinomas, 53 carcinomas de células escamosas y tan solo 8 (5,4%) se clasificaron como carcinomas de células fusiformes (*Tabla 3*).

Tabla 3. Porcentaje de tumores que pertenecen a cada subtipo histológico, y número de ratones *sham* e irradiados que desarrollaron cada uno de los subtipos tumorales.

Subtipo histológico	Nº tumores	Frecuencia	Condición	Nº tumores
Adenocarcinoma	74	50%	<i>Sham</i>	39
			10 cGy	35
Carcinoma de células escamosas	53	35,8%	<i>Sham</i>	31
			10 cGy	22
Carcinoma de células fusiformes	8	5,4%	<i>Sham</i>	3
			10 cGy	5
Otros	13	8,8%	<i>Sham</i>	8
			10 cGy	5
TOTAL	148			

Estos resultados difieren considerablemente de los obtenidos por Jerry et al. (2000), trabajo en el que, empleando el mismo modelo de trasplante de epitelio mamario p53^{-/-} en ratones BALB/c *wild type*, el 100% de los tumores originados fueron adenocarcinomas. Sin embargo, en nuestros resultados observamos mayor variación histológica, similar a la obtenida por Zhang et al. (2015), y más parecida a la que ocurre en el cáncer de mama humano. De hecho, nuestros resultados coinciden con los datos publicados por la *American Cancer Society* (2020) en cuanto a porcentajes de subtipos histológicos, ya que prácticamente la totalidad de los tumores analizados son carcinomas y, dentro de ellos, la mayoría son adenocarcinomas. Además, observamos que únicamente un 5,4% de los tumores son carcinomas de células fusiformes, similar a lo que ocurre en humanos (< 5% del total). No obstante, vimos un claro incremento en el número de carcinomas de células escamosas (35,8% de los casos), respecto a lo que se esperaría en humanos, donde constituye un subtipo minoritario con menos del 5% de los casos totales.

Cabe destacar que no se observaron diferencias significativas entre los grupos *sham* y 10 cGy en cuanto a estos porcentajes (*Tabla 4*, p-valores no mostrados), por lo que podemos inferir que el subtipo tumoral no se asocia con el estado de irradiación del hospedador, al igual que reportaron Nguyen et al. (2011) y Zhang et al. (2015).

Tabla 4. Porcentaje de ratones irradiados y no irradiados que desarrollaron cada subtipo histológico.

	Adenocarcinoma	Carcinoma de células escamosas	Carcinoma de células fusiformes	Otros
Sham (n=81)	39 (48%)	31 (38%)	3 (4%)	8 (10%)
10 cGy (n=67)	35 (52%)	22 (33%)	5 (7,5%)	5 (7,5%)

4.3. CARACTERIZACIÓN MOLECULAR DE LOS TUMORES

De los 148 *cores* de TMAs analizados por inmunohistoquímica, 17 se descartaron del estudio por ser tejido mamario normal, sangre, tejido necrótico o papilomas. Los restantes 131 tumores se distribuyeron en 72 originados en ratones *sham* y 59, en irradiados. Tras la evaluación de los anticuerpos, observamos que el proceso de inmunotinción había funcionado en 7 de los 8 biomarcadores analizados. No detectamos señal correspondiente al receptor de estrógenos (ER) en ninguno de los 131 tumores analizados, lo que nos llevó a suponer que el anticuerpo primario había fallado en nuestras muestras, quizás, por un problema de especificidad. Este hecho nos impidió poder llevar a cabo la clasificación molecular de los tumores surgidos en ratones no irradiados e irradiados, y estudiar si había diferencias entre ellos.

Los resultados de expresión obtenidos para cada uno de los biomarcadores evaluados por IHC se muestran en la *Tabla 5*. Los sistemas de puntuación empleados para considerar positivo cada uno de ellos se resumen en el **Anexo II** y algunos ejemplos de inmunotinción positiva representativa de cada biomarcador se ilustran en el **Anexo IV**.

Tabla 5. Resumen de los resultados de expresión por IHC. Se especifican la localización subcelular donde se expresa cada biomarcador, la frecuencia con la que se expresan y el nº de muestras tumorales sham o 10 cGy en las que se expresa cada biomarcador.

Biomarcador	Localización		Valor	n	Frecuencia		n según condición			
	Esperada	Observada					Sham (n=72)		10 cGy (n=59)	
c-Myc	Núcleo	Núcleo	Positivo	30	22,9%		18		12	
			Negativo	100	76,3%		53		47	
			N/A	1	0,8%		1		-	
Ki67	Núcleo	Núcleo	> 14%	82	62,6%		44		38	
			≤ 14%	48	36,6%		27		21	
			N/A	1	0,8%		1		-	
CK14	Citoplasma	Citoplasma	Positivo	80	61,1%		44		36	
			Negativo	49	37,4%		27		22	
			N/A	2	1,5%		1		1	
ER	Núcleo	Non-signal	-	-	-		-		-	
PR	Núcleo	Núcleo	Positivo	4	3%		2		2	
			Negativo	126	96,2%		69		57	
			N/A	1	0,8%		1		-	
		Citoplasma	Positivo	25	19,1%		18		7	
			Negativo	105	80,1%		53		52	
			N/A	1	0,8%		1		-	
HER2	Membrana	Membrana	Positivo	4	3%	11,4%	2	5	2	10
			Intermedio	11	8,4%		3		8	
			Negativo	113	86,3%		67		46	
			N/A	3	2,3%		-		3	
		Nuc/Cit	Positivo	1	0,8%	18,4%	1	11	-	13
			Intermedio	23	17,6%		10		13	
			Negativo	103	78,6%		60		43	
			N/A	4	3%		1		3	
CD14	Núcleo	Núcleo	Positivo	59	45%		34		25	
			Negativo	71	54,2%		38		33	
			N/A	1	0,8%		-		1	
E-cadherina	Membrana	Citoplasma	Positivo	67	51%		37		30	
			Negativo	59	45%		34		25	
			N/A	5	4%		1		4	

Al analizar la localización subcelular de los biomarcadores por IHC, vimos que algunos de ellos presentaban un patrón de distribución diferente al esperado, como son HER2, PR y E-cadherina (Figura 6).

Observamos que la proteína HER2 se expresaba tanto en la membrana plasmática (MemErbB-2) como en el núcleo celular (NuclErbB-2) (Figura 6A), al igual que se ha visto en humanos (Wang et al., 2004; Schillaci et al., 2012). El patrón de expresión nuclear de HER2 se ha descrito como un factor pronóstico independiente de resultado clínico desfavorable en tumores de mama positivos para MemErbB-2, ya que HER2 en el núcleo celular actúa como factor de transcripción activador de COX-2, enzima implicada en un aumento del potencial antiapoptótico, proangiogénico y metastásico en las células cancerosas (Wang et al., 2004). Asimismo, NuclErbB-2 también puede ensamblar un complejo transcripcional en el que funciona como co-activador de Stat3, promoviendo la expresión de ciclina D1 e induciendo la proliferación celular (Schillaci et al., 2012). En efecto, todas nuestras muestras positivas para MemErbB-2 y NuclErbB-2 mostraban un índice de proliferación elevado (Ki67 > 14%). Si bien, no encontramos literatura referente a la expresión exclusiva de HER2 en el núcleo celular, como observamos en

algunas de nuestras muestras, y si ésta tiene algún significado en cáncer de mama. En cualquier caso, no percibimos diferencias entre las muestras que únicamente expresaban NuclErbB-2 y las muestras MembErbB-2 y NuclErbB-2 positivas en cuanto a la expresión del resto de biomarcadores analizados.

Además de la expresión en membrana y/o nuclear, observamos que HER2 también se expresaba en el citoplasma (c-ErbB-2) de algunas células junto con NuclErbB-2 o MembErbB-2 (*Figura 6B*). La sobreexpresión de c-ErbB-2 se ha descrito escasamente en varios tipos de cáncer humano (neoplasia de tiroides, carcinoma de páncreas, tumores suprarrenales o cáncer de próstata) y parece estar relacionada con la diferenciación neuroendocrina en cáncer de mama, ya que se ha constatado co-expresión de c-Erb-2 con varios marcadores neuroendocrinos, como Sinaptofisina, Cromogranina A, Enolasa neuroespecífica (NSE) y CD56 (Horiguchi et al., 2010). Aunque los tumores neuroendocrinos tienden a ser menos agresivos que los carcinomas infiltrantes (Puente et al., 2015), las implicaciones clínicas y biológicas de c-Erb-2 (asociado o no a la expresión de marcadores neuroendocrinos) en cáncer de mama todavía no se conocen (Horiguchi et al., 2010).

Respecto al receptor de progesterona, observamos que casi el 20% de los tumores lo expresaban a nivel citoplásmico (*Figura 6C*). Este patrón de distribución ha sido reportado tanto en cáncer de endometrio (Leslie et al., 2005) como en cáncer de mama humano (Graham et al., 1996), y parece que podría deberse a un aumento de la expresión de la isoforma B del receptor (PRB), que se localiza principalmente en citoplasma, frente a la isoforma A (PRA), que se expresa mayoritariamente en el núcleo celular, incluso en ausencia de ligando (Lim et al., 1999; Leslie et al., 2005). El receptor de progesterona, al igual que el receptor de estrógenos, juega un papel muy importante en la aparición, el desarrollo y el pronóstico del cáncer de mama. De hecho, los tumores que expresan PR y/o ER en el núcleo celular se han asociado a un mejor pronóstico debido a una mejor respuesta a la terapia hormonal. Se ha descrito que un elevado porcentaje de tumores de mama PR positivos muestran un nivel muy bajo de PRB y, en consecuencia, una alta proporción PRA:PRB (Lim et al., 1999), lo que nos lleva a suponer que los receptores PRA son los responsables de responder a la terapia endocrina. En efecto, los tumores que expresan PRA muestran un menor tamaño que aquellos que expresan PRB (Ordóñez, Bastidas and Castro, 2013). Es por ello que, una relación PRA:PRB alterada podría conducir a un pronóstico desfavorable debido a la mala respuesta de algunos pacientes a la terapia hormonal (Lim et al., 1999). Resultados similares se obtuvieron al analizar la expresión citosólica de los receptores de estrógenos ER α (Li et al., 2015) y ER β 2 (Shaaban et al., 2008) en cáncer de mama, que se asociaba a mayor tamaño tumoral, estadíos más avanzados, mayor riesgo de recaída y menor supervivencia. Efectivamente, observamos que en nuestras muestras, el 76% de los tumores con expresión citosólica de PR presentaban un índice proliferativo elevado (Ki67 > 14%), resultados congruentes con los descritos en la literatura (Lim et al., 1999).

En cuanto a la expresión de E-cadherina, vimos que mostraba un patrón de distribución principalmente citoplásmico (*Figura 6D*), que ha sido sumamente descrito en humanos (Acs et al., 2001; Harigopal et al., 2005; Younis, El Sakka and Haque, 2007). La E-cadherina es una proteína transmembrana de adhesión celular implicada en la regulación de procesos como polaridad, diferenciación, proliferación y migración celular a través de su asociación con el citoesqueleto de actina por medio de cateninas (Buda and Pignatelli, 2011). La pérdida de expresión o la disminución del número de moléculas funcionales de E-cadherina se relaciona con un aumento de

la motilidad celular y, por tanto, una mayor invasividad, lo que explica su papel como supresor tumoral. La acumulación de E-cadherina en el citoplasma de las células tumorales podría ser el resultado de una mutación que impide que la proteína sea correctamente procesada en el aparato de Golgi. En humanos, la pérdida completa de expresión de E-cadherina en membrana con o sin la expresión citoplásmica de E-cadherina está asociada al desarrollo de carcinomas lobulares, mientras que los carcinomas ductales se caracterizan por la expresión de E-cadherina membranosa (Acs et al., 2001). Teniendo en cuenta las diferencias en la distribución de la E-cadherina, el patrón de tinción de E-cadherina por IHC nos permite clasificar los tumores en lobulares o ductales. De acuerdo con este criterio, la mayor parte de los tumores de nuestro estudio se clasificaron como carcinomas lobulares y una minoría, como carcinomas mixtos (que presentan pérdida casi completa de la expresión de E-cadherina en membrana (Acs et al., 2001)), a diferencia de lo que ocurre en humanos, donde representan cerca del 10% y menos del 5%, respectivamente (American Cancer Society, 2020). Además, según Derksen et al. (2006), la pérdida combinada de E-cadherina funcional y p53 específicamente en el epitelio mamario resulta en el desarrollo acelerado de carcinomas mamarios invasivos y metastásicos, que muestran un gran parecido con los ILCs humanos, por lo que el modelo de ratón empleado en este trabajo podría constituir un buen modelo animal para el estudio del ILC humano.

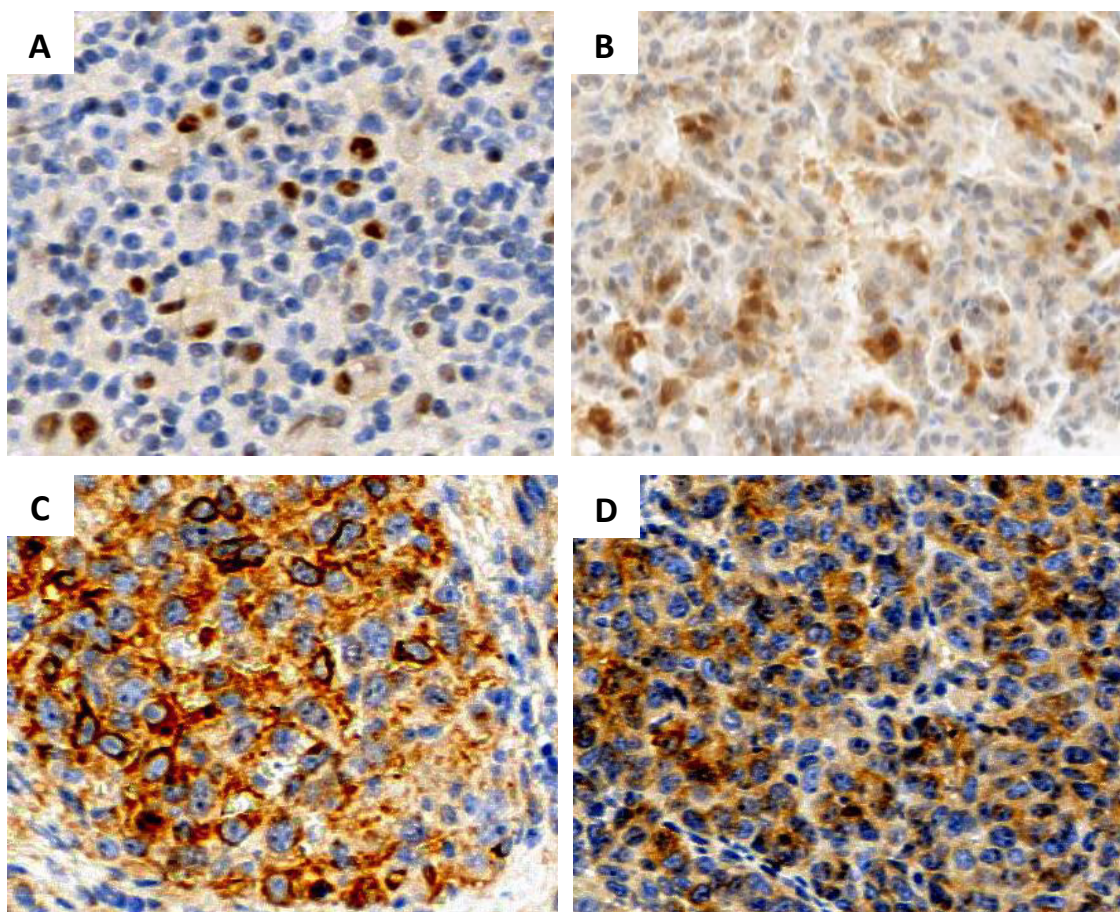


Figura 6. Tinción aberrante de algunos biomarcadores por IHC. **A)** Distribución nuclear de HER2. **B)** Co-expresión de HER2 en núcleo y citoplasma. **C)** Distribución de PR en citoplasma y membrana. **D)** Expresión citosólica de E-cadherina.

En la *Tabla 6* se muestran los porcentajes de tumores irradiados y no irradiados que expresaron cada uno de los biomarcadores analizados. Cabe destacar que no se encontraron diferencias significativas en la expresión de los distintos marcadores entre los grupos, excepto para HER2 de membrana, cuya expresión no se distribuye del mismo modo entre ratones no irradiados e irradiados (p -valor = 0,022) (*Figura 7*). Los tumores irradiados muestran mayor expresión de MembErbB-2, lo que podría explicarse por efecto de la radiación, que introduce mutaciones en el protooncogén *HER2*, amplificándolo y conduciendo a la sobreexpresión de la oncoproteína HER2 en la membrana y a una tasa de crecimiento acelerada. Esta misma hipótesis fue comprobada por Castiglioni et al. (2007) en tumores de mama humano, donde observaron que la radiación administrada antes de la pubertad (como tratamiento de cánceres pediátricos) es un factor de riesgo para el desarrollo de tumores HER2+.

Tabla 6. Porcentaje de tumores *sham* o 10 cGy que expresan cada biomarcador.

	c-Myc		Ki67		CK14		PR nuclear		PR citoplásmico		HER2 membrana		HER2 nuc/cit		CD14		E-cadherina	
	+	-	≤14%	>14%	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Sham	25%	75%	38%	62%	62%	38%	3%	97%	25%	75%	7%	93%	15,5%	84,5%	47%	53%	52%	48%
10 cGy	20%	80%	35,6%	64,4%	62%	38%	3,4%	96,6%	12%	88%	18%	82%	23%	77%	43%	57%	54,5%	45,5%

		HER2/ErbB2 Mb			Total
		N/A	NEGATIVE	POSITIVE	
GRUPO	10cGy	3	46	10	59
	SHAM	0	67	5	72
Total		3	113	15	131

Pruebas de chi-cuadrado

	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)
Chi-cuadrado de Pearson	7,352 ^a	2	,025	,017
Razón de verosimilitud	8,491	2	,014	,020
Prueba exacta de Fisher-Freeman-Halton	6,857			,022
N de casos válidos	131			

a. 2 casillas (33,3%) han esperado un recuento menor que 5. El recuento mínimo esperado es 1,35.

Figura 7. Resultado prueba exacta de Fisher. p -valor < 0,05 rechaza la hipótesis de expresión independiente de MembErbB-2 entre grupos.

Por otra parte, el análisis estadístico de las frecuencias de expresión de los distintos biomarcadores arrojó varias correlaciones interesantes con un umbral de significatividad p -valor < 0,05. Los resultados de los análisis de correlación entre los distintos biomarcadores se resumen en el **Anexo V**.

Obtuvimos que la expresión de CD14 se correlaciona con Ki67 con un p-valor = 0,002. Esta asociación entre el marcador de proliferación Ki67 y el marcador de células madre CD14 ha sido previamente descrita por Cheah et al. (2015) en líneas celulares de cáncer de vejiga humanas y de ratón. CD14 es un conocido marcador de monocitos/macrófagos implicado en la fagocitosis de células apoptóticas, pero también se ha descrito su expresión en células luminales del epitelio mamario (Lobba et al., 2012). La expresión de CD14 en las células cancerosas estimula la secreción de varios mediadores de la inflamación, incluidas citocinas (principalmente IL-6) y factores de crecimiento que promueven la proliferación y el crecimiento tumoral, así como factores angiogénicos que activan la angiogénesis para favorecer el establecimiento y mantenimiento de un microambiente tumoral inflamatorio. Además, algunos de esos factores inflamatorios secretados por las células CD14+ tienen propiedades quimiotácticas que inducen la migración y el reclutamiento de macrófagos, monocitos y neutrófilos en el microambiente tumoral, donde los macrófagos se polarizan hacia el fenotipo M2 (más tumorigénico) y los monocitos presentan una reducción en la expresión de moléculas MHC II, suprimiendo localmente la respuesta inmune del hospedador. Todo este proceso inflamatorio promovido por el propio tumor hace que los tumores surgidos a partir de células CD14+, a diferencia de los CD14-, sean tumores más vascularizados, con mayor infiltrado de células mieloides y con características inmunosupresoras que, en conjunto, favorecen la proliferación celular (Ki67>14%) y el crecimiento tumoral (Cheah et al., 2015). Sin embargo, en nuestras muestras no observamos diferencias en la media de días que estos ratones tardaron en alcanzar un tamaño tumoral de 10 mm³ respecto a la media general.

Al estudiar esta correlación por separado en los dos grupos de ratones, vimos que solo era significativa en los tumores irradiados. Esta diferencia podría explicarse porque la radiación introduce daño en el DNA, lo que lleva a la liberación de DAMPs (patrones moleculares asociados al daño), que actúan como ligandos endógenos de CD14 en las células tumorales, activándolo e induciendo toda la cascada de señalización que lleva al desarrollo de un microambiente inflamatorio e inmunosupresor pro-tumoral (Cheah et al., 2015).

También en los tumores irradiados se observó una asociación significativa (p-valor < 0,001) entre la expresión de CD14 y la pérdida de expresión de E-cadherina en membrana. Esta correlación podría deberse únicamente al efecto de la radiación sobre el DNA. Como se ha descrito previamente, la radiación a la que fueron sometidos los ratones puede introducir daños en el DNA que llevan a la liberación de DAMPs, que activan el receptor CD14, promoviendo el establecimiento de un microambiente inflamatorio y el crecimiento tumoral. Por otra parte, la radiación es un agente mutágeno que podría ser la responsable de introducir mutaciones en el gen *Cdh1* (que codifica para la E-cadherina), impidiendo que la proteína pueda translocarse a la membrana para realizar su función. Resultados similares a los propuestos se han descrito en células mamarias epiteliales humanas, que disminuían la expresión de E-cadherina en membrana tras su exposición a radiación ionizante (C. C. Park, et al., 2003). La desregulación de ambas moléculas por efecto de la radiación explicaría el desarrollo de tumores con alta tasa de proliferación e invasividad, asociados a un peor pronóstico.

Asimismo, observamos una correlación significativa (p-valor < 0,001) entre la expresión de CK14 y E-cadherina, que se mantenía tanto en los tumores irradiados como en los no irradiados al analizarlos por separado. Las citoqueratinas (CK) son filamentos intermedios que conectan la membrana nuclear con la membrana plasmática, por lo que están implicadas en el mantenimiento

de la integridad celular y en la regulación de procesos como crecimiento, proliferación, migración y apoptosis (Nieves et al., 2014). En concreto, la CK14 se expresa en las células epiteliales basales y es considerada un marcador de células epiteliales, al igual que la E-cadherina. La pérdida de estas proteínas se relaciona con una pérdida de las características epiteliales de la célula, la adquisición de un fenotipo mesenquimal (transición epitelio-mesénquima, EMT) y un aumento en la capacidad migratoria de las células, ya que la pérdida de E-cadherina y CK14 conlleva la pérdida de la adhesión intercelular y la desregulación de la proliferación y la migración celular. Este fenotipo mesenquimal se asocia a un aumento del potencial invasivo y metastásico del tumor, por lo que se espera que estos tumores sean más agresivos y tengan un pronóstico desfavorable (Sarrió et al., 2008).

Otra de las correlaciones encontradas fue la co-expresión de c-Myc y Ki67 (p-valor = 0,029). El protooncogén *myc* codifica para el factor de transcripción c-Myc, implicado en la regulación de múltiples procesos como el ciclo celular, metabolismo celular, angiogénesis, adhesión celular, diferenciación celular, reparación del DNA y muerte celular. La amplificación del protooncogén *myc* en las células epiteliales, descrita en varios tipos de cáncer de mama conduce a la sobreexpresión de la oncoproteína c-Myc y a una sobreactivación del ciclo celular alterando la tasa de proliferación celular (Ki67 > 14%), lo que promueve la inestabilidad genómica y acelera la transformación neoplásica (Ospina and Muñetón, 2011). De hecho, observamos que la media de días hasta que estos tumores alcanzaron los 10 mm³ estaba por debajo de la media general (387 vs. 406 días).

En línea con la anterior, se puede explicar la correlación entre c-Myc y E-cadherina (p-valor = 0,018), más significativa en ratones *sham*. Varios autores han descrito que la sobreexpresión de c-Myc induce la represión de E-cadherina y estimula la transición epitelio-mesénquima (EMT) en cáncer de mama (Cho et al., 2010; Ma et al., 2010; Liu et al., 2013). La represión de E-cadherina por sobreexpresión de c-Myc se puede explicar por dos vías: en primer lugar, la sobreexpresión de c-Myc activa la cinasa ERK, que fosforila a la cinasa GSK-3 β , inactivándola y, por tanto, induciendo la activación constitutiva de la proteína snail, implicada en la supresión de la E-cadherina (Cho et al., 2010). La segunda de las vías explica la represión de E-cadherina a través de un mecanismo post-transcripcional: la sobreexpresión de c-Myc activa la transcripción del microRNA-9, que reconoce y degrada el mRNA *CDHI*, inhibiendo la expresión de E-cadherina para la que codifica. Todo ello conduce a una mayor motilidad celular, invasividad y metástasis (Ma et al., 2010; Liu et al., 2013).

Por último, se vio que existía una correlación significativa entre la expresión de CD14 y PR nuclear y PR citoplásmico (p-valor = 0,27 y p-valor < 0,001, respectivamente), que fue más significativa en los tumores *sham*. La relación entre la expresión de CD14 en las células tumorales que no expresan PR nuclear fue descrita por Mehmeti et al. (2015), trabajo en el que observaron que CD14 se expresaba como co-receptor de TLR4 en líneas celulares de cáncer de mama triple negativo. La señalización mediada por los TLRs conduce a la activación del factor nuclear kappa B (NF κ B) y una posterior expresión de genes pro-inflamatorios, que inducen la proliferación celular y el reclutamiento de células mieloides, y resultan en un peor pronóstico. Se ha descrito que NF κ B inhibe la expresión de ER α en las células tumorales de mama, pero no se ha reportado cuál podría ser el papel de CD14 sobre PR. En nuestros datos, el 82% de los tumores que fueron CD14+PR- fueron también HER2-, lo que nos lleva a pensar que la sobreexpresión de CD14 podría ser un marcador de cáncer de mama triple negativo. Además, el 80% de los tumores

CD14+PR-HER2- presentaron un índice de proliferación elevado ($Ki67 > 14\%$), que se asocia también a los TNBC. No obstante, sería necesario repetir el análisis inmunohistoquímico con anti-ER para conocer realmente qué porcentaje de estos tumores son ER-.

4.4. CALIDAD DE LAS SECUENCIAS FASTA

Tras evaluar la calidad de las lecturas resultantes del proceso de secuenciación, se observó que la calidad de las bases iba disminuyendo conforme se avanzaba en la posición de la secuencia (*Figura 8*). Esta pérdida de calidad se debe a que el proceso de secuenciación por síntesis se hace nucleótido por nucleótido, es decir, por ciclos. En cada ciclo se añaden los 4 nucleótidos unidos reversiblemente a terminadores fluorescentes (cada uno de los cuales emite a una longitud de onda diferente) que impiden la unión de más de un nucleótido a la hebra que se está sintetizando. Cuando se ha incorporado una base, se lleva a cabo un lavado para eliminar dichos terminadores y los nucleótidos no incorporados y se vuelve a añadir el mix de 4 nucleótidos para permitir la incorporación de la siguiente base, y así sucesivamente. Un mal lavado de la celda de flujo donde se lleva a cabo la secuenciación puede no eliminar correctamente el terminador de alguna de las secuencias o los nucleótidos no incorporados, lo que puede resultar en la pérdida o ganancia, respectivamente, de un ciclo respecto al resto de secuencias del *cluster* (Pfeiffer et al., 2018). Esta desincronización hace que haya una contaminación en la señal fluorescente detectada por el secuenciador, lo que disminuye la calidad de la base leída.

En nuestro caso, como se observa en la *Figura 8*, hay una tendencia a la disminución de la calidad de las lecturas al final de las secuencias, por lo que se decidió eliminar las últimas 25 bases del extremo 3' de cada una de ellas para evitar problemas potenciales en el subsiguiente proceso de alineamiento.

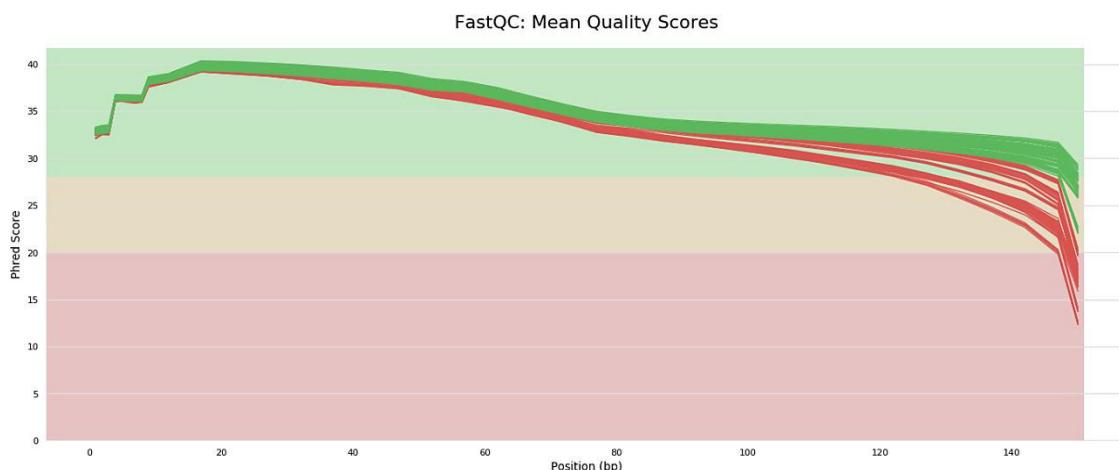


Figura 8. Representación gráfica de la calidad promedio de las lecturas en función de su posición en la secuencia nucleotídica.

4.5. RESULTADOS DEL ALINEAMIENTO Y EL CONTEO

Los resultados del proceso de recuento se presentan en una matriz en la que cada fila representa un gen y cada columna, una muestra. Las entradas de la matriz indican el número de lecturas que se han asignado de forma inequívoca a un gen concreto en una muestra determinada.

El promedio de secuencias alineadas por cada muestra fue de 18675639,59 de las cuales una media de 12926181,5 fue asignada a genes. El resto de secuencias alineadas fueron descartadas del estudio porque mapeaban contra múltiples puntos del genoma o contra regiones sin contenido genético.

4.6. ANÁLISIS DE EXPRESIÓN DIFERENCIAL IRRADIADOS VS. NO IRRADIADOS

Una vez disponíamos de los datos de recuento normalizados, los niveles de expresión de los RNAs mensajeros (mRNAs) fueron comparados entre el grupo de tumores irradiados y el grupo de tumores no irradiados. De las 162 muestras disponibles, 73 pertenecían a tumores irradiados, mientras que 89 correspondían a tumores sin irradiar. Aquellos genes que presentaron un promedio menor a 10 conteos por muestra fueron excluidos del análisis. Tras el filtrado, la expresión diferencial de un total de 15.299 genes fue evaluada empleando el paquete de R DeSeq2. Se observaron 135 genes diferencialmente expresados con p-valores ajustados inferiores a 0,05, de los cuales 40 estaban sobreexpresados y 95, infraexpresados en tumores irradiados (*Figuras 9A y 9C*).

En la *Tabla 7* se pueden consultar los resultados del análisis de expresión diferencial para aquellos genes que presentaron un p-valor ajustado menor a 0,05. Los genes sobreexpresados en tumores irradiados fueron los siguientes: Smim38, Upk3b, Cpne7, Tafa5, Sobp, Bmx, Cdh6, Crym, Dlx3, Dlx4, Drd4, Fgf4, Fgf8, Gpr83, Htr5b, Lepr, Nppc, Rad52, Frzb, Ctr9, A2m, Otoa, Stag3, Il17b, Ltb4r2, Otor, Ttyh1, Krtdap, Cntnap2, Catsperz, Serpina9, Rspo3, Kremen2, Anks1b, Notum, Pi15, Fam90a1a y Fgg, mientras que los genes infraexpresados en las muestras irradiadas fueron: Tmem168, Plin1, Acly, Ces1d, Osr2, Aldh111, Lrrc39, Timp4, Lyve1, Adipoq, Ancy8, Cfd, Adrb3, Pnpla3, Aqp7, Mmp28, Bmpr1b, Car3, Car4, Runx1t1, Cldn5, Cyp2e1, Cyp2f2, Ackr1, Opn3, Hnmt, F2r13, Lpin1, Cidec, Gsta3, Irf4, Kcnab2, Kcnk3, Napsa, Kdr, Serpina3c, Lipe, Lpl, Ogn, Pck1, Pfkfb1, Pitx1, Ptprd, Rxrg, Six1, Slc22a3, Serpina3n, Syp, Pank3, Acsl6, Tshr, Ikzf3, Bpifb5, Insig1, Atp1a3, Nbeal2, Btnl9, Pigz, Sspo, Myo16, Slc36a2, Abcd2, Wdr3, Nat8l, A530016L24Rik, Ric3, Slc7a10, Sfrp5, Dpt, Retnla, Retn, Plin4, Ppp1r1a, Smpd3, Dhhr7, Abhd15, Agpat2, Ccdc80, Dhhdh, Apol6, Ces2g, Kbtbd11, Mrap y Htra3.

En términos generales, los perfiles transcriptómicos de las muestras irradiadas y no irradiadas no presentaron demasiadas diferencias. Este patrón se puede observar en la *Figura 9B*, que contiene la representación gráfica de las dos primeras componentes principales de los datos de expresión. Como se puede apreciar en el gráfico, no existe una agrupación clara entre las muestras irradiadas y las no irradiadas.

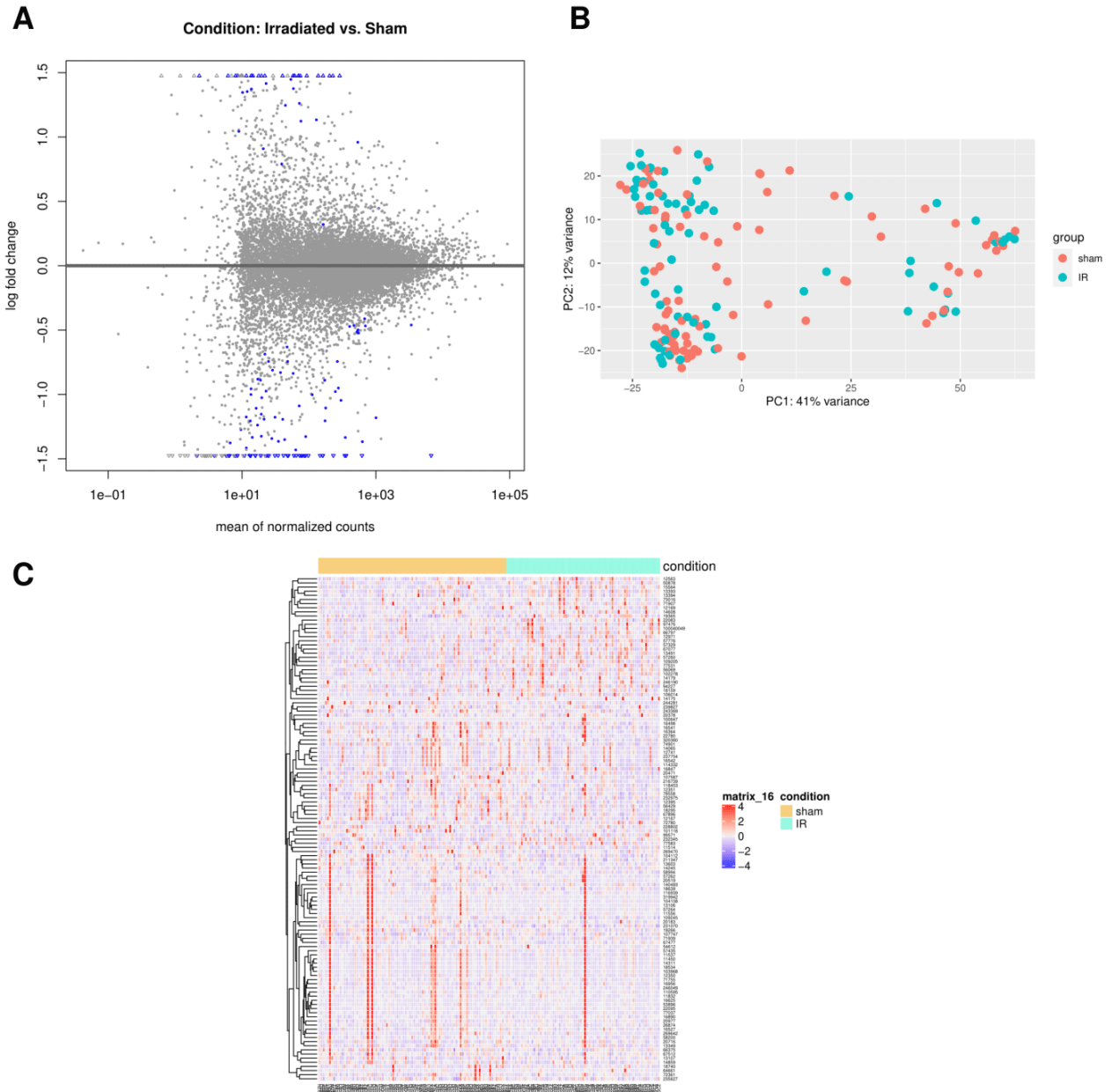


Figura 9. **A)** Gráfico MA que muestra el logaritmo del *fold change* en el eje y contra el promedio de los conteos normalizados en el eje x. Los puntos azules representan genes diferencialmente expresados con p-valores ajustados menores de 0,05. **B)** Análisis de componentes principales de los datos de expresión. Las dos primeras componentes explican el 41% y el 12% de la variabilidad de los datos, respectivamente. Los puntos rojos señalan las muestras no irradiadas, mientras que los puntos azules indican las muestras irradiadas. **C)** *Heatmap* que muestra los 135 genes diferencialmente expresados. La barra superior amarilla señala las muestras no irradiadas, mientras que la barra azul celeste señala las muestras irradiadas.

Tabla 7. Genes diferencialmente expresados entre los tumores irradiados y los no irradiados.

Entrez	baseMean	log2FoldChange	p-value	padj	external_gene_name
100040049	74.3	1.61	1.77e-04	2.68e-02	Smim38
100647	2.29	2.98	1.66e-05	4.69e-03	Upk3b
101118	539.74	-0.5	2.13e-04	2.96e-02	Tmem168
102278	53.67	1.45	1.97e-04	2.86e-02	Cpne7
103968	59.58	-3.54	6.84e-11	3.49e-07	Plin1
104112	3375.32	-0.46	3.93e-04	4.56e-02	Acly
104158	18.89	-3.38	2.09e-09	4.57e-06	Ces1d
106014	44.37	1.25	3.30e-04	4.10e-02	Tafa5
107587	17.05	-1.24	2.60e-04	3.40e-02	Osr2
107747	177.78	-1.11	6.62e-05	1.30e-02	Aldh111
109205	75.71	1.12	9.03e-05	1.62e-02	Sobp
109245	21.95	-0.69	2.10e-04	2.96e-02	Lrrc39
110595	8.54	-1.64	9.35e-05	1.66e-02	Timp4
114332	160.28	-1.73	5.00e-06	2.07e-03	Lyve1
11450	64.8	-3.01	6.79e-10	2.08e-06	Adipoq
11514	72.38	-1.55	1.91e-04	2.80e-02	Adcy8
11537	231.33	-2.49	1.64e-09	4.18e-06	Cfd
11556	13.32	-2.39	4.74e-05	1.00e-02	Adrb3
116939	21.32	-1.19	1.52e-04	2.48e-02	Pnpla3
11832	10.88	-1.54	1.53e-04	2.48e-02	Aqp7
118453	31.08	-1.18	1.29e-05	4.08e-03	Mmp28
12167	19.37	-1.15	8.53e-05	1.57e-02	Bmpr1b

12169	13.8	1.37	1.39e-05	4.08e-03	Bmx
12350	161.4	-3.45	6.75e-12	5.17e-08	Car3
12351	22.92	-1.88	2.29e-06	1.21e-03	Car4
12395	16.25	-1.11	1.14e-04	1.94e-02	Runx1t1
12563	11.92	1.35	3.65e-04	4.36e-02	Cdh6
12741	1003.14	-1.18	2.42e-04	3.24e-02	Cldn5
12971	227.3	1.54	1.02e-05	3.53e-03	Crym
13106	30.97	-4.74	5.72e-12	5.17e-08	Cyp2e1
13107	97.76	-2.02	1.35e-08	1.47e-05	Cyp2f2
13349	11.58	-1.42	3.83e-04	4.47e-02	Ackr1
13393	203.17	1.59	4.70e-06	2.00e-03	Dlx3
13394	20.78	0.91	2.11e-04	2.96e-02	Dlx4
13491	58.51	1.38	2.87e-04	3.66e-02	Drd4
13603	37.55	-0.83	3.08e-06	1.52e-03	Opn3
140483	22.84	-0.97	3.71e-04	4.36e-02	Hnmt
14065	71.52	-1.09	4.24e-04	4.87e-02	F2r13
14175	6.18	4.41	1.23e-08	1.45e-05	Fgf4
14179	9.99	2.75	2.28e-07	1.52e-04	Fgf8
14245	182.88	-0.66	5.25e-05	1.06e-02	Lpin1
14311	85.04	-2.52	3.70e-08	3.33e-05	Cidec
14608	13.77	2.18	7.27e-06	2.78e-03	Gpr83
14859	79.45	-1.54	2.32e-05	6.23e-03	Gsta3
15564	17.67	1.62	3.42e-04	4.20e-02	Htr5b
16364	14.21	-1.33	8.11e-06	2.97e-03	Irf4
16498	247.88	-0.98	4.09e-06	1.89e-03	Kcnab2

16527	27.89	-1.34	7.86e-05	1.47e-02	Kcnk3
16541	111.52	-1	1.28e-05	4.08e-03	Napsa
16542	626.43	-1.37	1.54e-06	8.42e-04	Kdr
16625	3.85	-2.41	4.26e-05	9.44e-03	Serpina3c
16847	39.13	0.79	2.56e-04	3.38e-02	Lepr
16890	265.44	-0.74	2.78e-05	6.87e-03	Lipe
16956	349.46	-1.33	5.33e-07	3.26e-04	Lpl
18159	40.26	1.55	2.46e-05	6.32e-03	Nppc
18295	48.97	-2.09	3.42e-09	5.82e-06	Ogn
18534	34.58	-2.36	5.14e-05	1.05e-02	Pck1
18639	47.93	-0.74	2.38e-05	6.28e-03	Pfkfb1
18740	41.7	-1.77	1.21e-05	4.04e-03	Pitx1
19266	34.99	-1.36	3.32e-05	7.67e-03	Ptprd
19365	163.97	0.32	1.00e-04	1.75e-02	Rad52
20183	7.97	-1.83	2.02e-05	5.63e-03	Rxrg
20378	8.95	1.04	7.64e-05	1.46e-02	Frzb
20471	63.36	-1.43	1.38e-05	4.08e-03	Six1
20519	5.78	-2.22	9.89e-05	1.74e-02	Slc22a3
20716	173.81	-1.2	3.63e-04	4.36e-02	Serpina3n
20977	28.5	-0.81	5.06e-05	1.05e-02	Syp
211347	696.48	-0.47	7.18e-05	1.39e-02	Pank3
216739	24.66	-1.67	2.97e-05	7.11e-03	Acsl6
22083	1100.64	0.45	4.29e-04	4.87e-02	Ctr9
22095	2.92	-1.85	4.77e-05	1.00e-02	Tshr

22780	13.59	-1	2.00e-04	2.86e-02	Ikzf3
228802	88.91	-2.47	2.90e-09	5.54e-06	Bpifb5
231070	551.52	-0.52	3.10e-05	7.30e-03	Insig1
232345	61.2	1.48	7.76e-05	1.47e-02	A2m
232975	172.25	-0.89	1.76e-04	2.68e-02	Atp1a3
235627	523.86	-0.51	3.63e-04	4.36e-02	Nbeal2
237754	89.36	-1.33	1.73e-04	2.68e-02	Btm19
239827	17.2	-0.88	1.54e-04	2.48e-02	Pigz
243369	19.72	-1.03	2.81e-04	3.62e-02	Sspo
244281	6.48	-1.93	1.11e-05	3.77e-03	Myo16
246049	15.86	-1.81	5.30e-08	4.27e-05	Slc36a2
246190	13.72	1.67	1.24e-07	8.80e-05	Otoa
26874	43.27	-1.35	3.18e-07	2.02e-04	Abcd2
269470	678.62	-0.41	2.36e-04	3.19e-02	Wdr3
269642	47.65	-1.59	6.88e-06	2.70e-03	Nat81
319942	6.23	-3.08	2.93e-08	2.99e-05	A530016L24Rik
320360	13.29	-1.21	3.07e-04	3.88e-02	Ric3
50878	69.84	1.85	2.50e-06	1.27e-03	Stag3
53896	15.76	-3.19	9.05e-09	1.15e-05	Slc7a10
54612	6.64	-1.38	2.82e-04	3.62e-02	Sfrp5
56069	287.18	2.04	4.32e-06	1.89e-03	Il17b
56429	358.47	-1.65	9.32e-06	3.32e-03	Dpt
57260	19.54	1.65	1.38e-05	4.08e-03	Ltb4r2
57262	12.71	-1.47	1.71e-04	2.67e-02	Retnla
57264	24.57	-2.15	3.47e-06	1.66e-03	Retn

57329	14.52	2.75	3.43e-04	4.20e-02	Otor
57435	76.72	-1.59	3.95e-05	8.88e-03	Plin4
57776	128.95	1.13	1.32e-05	4.08e-03	Ttyh1
58200	15.88	-1.6	4.99e-08	4.24e-05	Ppp1r1a
58994	18.86	-0.89	8.16e-06	2.97e-03	Smpd3
64661	8.01	2.15	1.91e-04	2.80e-02	Krtdap
66375	470.16	-0.46	2.45e-04	3.25e-02	Dhrs7
66797	69.66	1.67	4.48e-05	9.65e-03	Cntnap2
67077	71.89	1.26	1.25e-04	2.09e-02	Catsperz
67477	58.04	-0.84	5.51e-06	2.22e-03	Abhd15
67512	404.39	-0.47	3.67e-04	4.36e-02	Agpat2
67896	618.86	-1.7	1.98e-10	7.57e-07	Ccdc80
71755	47.14	-0.63	3.09e-04	3.88e-02	Dhdh
71907	8.55	2.06	3.36e-05	7.67e-03	Serpina9
71939	40.51	-1.19	2.48e-05	6.32e-03	Apol6
72361	19.44	-1.33	2.23e-04	3.08e-02	Ces2g
72780	21.89	1.63	4.22e-06	1.89e-03	Rspo3
73016	11.59	1.49	2.32e-04	3.16e-02	Kremen2
74901	300.97	-1.05	1.32e-04	2.20e-02	Kbtbd11
77037	13.13	-1.52	4.30e-04	4.87e-02	Mrap
77531	22.9	1.42	1.85e-04	2.77e-02	Anks1b
77583	162.53	1.83	9.36e-07	5.36e-04	Notum
78558	274.93	-0.95	4.46e-05	9.65e-03	Htra3
94227	137.3	1.69	2.12e-05	5.79e-03	Pi15

97476	57.81	3.94	5.11e-09	7.82e-06	Fam90a1a
99571	92.71	1.85	1.11e-04	1.91e-02	Fgg

4.7. ANÁLISIS DE ENRIQUECIMIENTO EN CATEGORÍAS FUNCIONALES

Tras el análisis de enriquecimiento, observamos que 31 *pathways* presentaban p-valores ajustados inferiores a 0,05, de los cuales 21 estaban sobreexpresados en tumores irradiados y 10, infraexpresados. Entre los *pathways* sobreexpresados encontramos iniciación y elongación de la traducción, ribosoma, degradación del mRNA mediada por mutación terminadora y meiosis, entre otros. Por contra, entre los *pathways* downregulados observamos funciones relacionadas con la señalización a través de PPAR, el metabolismo de triglicéridos o la vía de señalización de la insulina. En la *Tabla 8* se muestran todos los *pathways* diferencialmente expresados de manera significativa entre tumores irradiados y no irradiados.

Tabla 8. *Pathways* diferencialmente expresados en tumores irradiados y no irradiados. Las dos primeras columnas contienen los valores para el *score* de enriquecimiento y el *score* de enriquecimiento ajustado. Los valores positivos indican sobreexpresión en tumores irradiados, mientras que valores negativos indican infraexpresión en tumores irradiados.

PATHWAY	ES	NES	p-val	padj
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION	0.71	2.34	3.15e-10	4.17e-07
KEGG_RIBOSOME	0.72	2.32	8.57e-10	4.53e-07
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_A CID_DEFICIENCY	0.68	2.28	1.23e-08	4.86e-06
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION	0.66	2.25	5.26e-10	4.17e-07
REACTOME_NONSENSE_MEDIATED_DECAY_NMD	0.62	2.11	4.16e-08	1.10e-05
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PRO TEIN_TARGETING_TO_MEMBRANE	0.62	2.1	1.16e-07	2.63e-05
REACTOME_MEIOSIS	0.62	1.98	1.08e-05	1.31e-03
REACTOME_MEIOTIC_SYNAPSIS	0.65	1.96	8.64e-05	8.56e-03
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDIN G_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUB SEQUENT_BINDING_TO_43S	0.62	1.95	8.32e-05	8.56e-03

REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_A ND_ROBOS	0.53	1.93	3.25e-07	6.43e-05
REACTOME_NEGATIVE_REGULATION_OF_FGFR3_SIGNA LING	0.75	1.92	1.91e-04	1.53e-02
REACTOME_NEGATIVE_REGULATION_OF_FGFR4_SIGNA LING	0.73	1.91	4.15e-04	2.86e-02
REACTOME_INFLUENZA_INFECTION	0.54	1.91	1.39e-06	2.45e-04
REACTOME_FGFR1_MUTANT_RECEPTOR_ACTIVATION	0.69	1.88	6.87e-04	4.19e-02
REACTOME_REPRODUCTION	0.57	1.85	1.93e-04	1.53e-02
REACTOME_MEIOTIC_RECOMBINATION	0.64	1.85	8.88e-04	4.80e-02
REACTOME_RRNA_PROCESSING	0.49	1.81	7.56e-06	1.09e-03
REACTOME_SELENOAMINO_ACID_METABOLISM	0.51	1.73	2.70e-04	2.04e-02
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_ MATRIX	0.46	1.62	9.34e-04	4.80e-02
REACTOME_TRANSLATION	0.41	1.57	1.45e-04	1.28e-02
NABA_MATRISOME_ASSOCIATED	0.34	1.37	6.80e-04	4.19e-02
REACTOME_CLASS_A_1_RHODOPSIN_LIKE_RECEPTORS	-0.48	-1.61	8.42e-04	4.80e-02
KEGG_CHEMOKINE_SIGNALING_PATHWAY	-0.49	-1.63	3.89e-04	2.80e-02
REACTOME_BIOLOGICAL_OXIDATIONS	-0.5	-1.66	6.84e-04	4.19e-02
REACTOME_INTEGRATION_OF_ENERGY_METABOLISM	-0.53	-1.68	9.39e-04	4.80e-02
KEGG_HEMATOPOIETIC_CELL_LINEAGE	-0.56	-1.74	9.16e-04	4.80e-02
KEGG_INSULIN_SIGNALING_PATHWAY	-0.55	-1.78	1.36e-04	1.27e-02
REACTOME_TRANSCRIPTIONAL_REGULATION_OF_WHIT E_ADIPOCYTE_DIFFERENTIATION	-0.62	-1.93	8.27e-06	1.09e-03
REACTOME_TRIGLYCERIDE_CATABOLISM	-0.84	-1.94	2.16e-05	2.44e-03
REACTOME_TRIGLYCERIDE_METABOLISM	-0.82	-2.08	4.76e-06	7.54e-04
KEGG_PPAR_SIGNALING_PATHWAY	-0.77	-2.23	2.58e-08	8.16e-06

Cuando analizamos los *pathways* sobreexpresados, observamos que gran parte de las vías de señalización sobreactivadas en tumores irradiados estaban implicadas en el proceso de traducción del mRNA (iniciación, elongación o formación del ribosoma). Varios autores han descrito previamente que la traducción desregulada del mRNA es una característica frecuente de las neoplasias, ya que muchos factores implicados en dichas *pathways* se encuentran sobreexpresados en la mayoría de tumores, incluido el cáncer de mama. La sobreactivación del proceso de traducción conlleva un aumento remarcable en la síntesis de proteínas que se ha asociado a un incremento anormal de la proliferación celular, de la angiogénesis y a alteraciones en la respuesta inmune que, en conjunto, desencadenan la transformación celular y la progresión oncogénica (Bhat et al., 2015; Ali et al., 2017).

También observamos que los tumores irradiados estaban enriquecidos en vías de degradación de matriz extracelular. La degradación de la matriz extracelular está mediada por metaloproteinasas de la matriz (MMP). La liberación de estas MMPs es el resultado de comunicaciones cruzadas entre las células tumorales con el resto de células dentro del microambiente tumoral, incluidos fibroblastos asociados al cáncer (CAFs), macrófagos tipo 2 (M2) y células madre cancerosas. La escisión de los componentes de la matriz por acción de las MMPs es un proceso que conduce a la desregulación de su composición, estructura y abundancia, así como a la liberación de moléculas biológicamente activas, como los factores de crecimiento. Por ende, la degradación de la matriz extracelular en los tumores irradiados conduce a un aumento de la proliferación, migración e invasión, así como a la inducción de la angiogénesis y supervivencia tumoral (Najafi, Farhood and Mortezaee, 2019).

Respecto a la sobreactivación de la meiosis, cabe recordar que una característica de muchos cánceres es que activan genes que normalmente están asociados con distintos estados del desarrollo, incluidos genes específicos de células germinales. En la actualidad, está emergiendo una fuerte evidencia de que la activación inapropiada de genes específicos de la línea germinal en tejido tumoral conduce a inestabilidad genómica, lo que favorece el desarrollo, mantenimiento y diseminación del cáncer (McFarlane and Wakeman, 2017).

Por último, aunque tradicionalmente se ha conocido la vía de degradación del mRNA mediada por mutaciones terminadoras (NMD, del inglés *Nonsense-Mediated mRNA Decay*) como un mecanismo de control de calidad que degrada los transcritos aberrantes derivados de genes mutados o de errores en el procesamiento del mRNA, en los últimos años, se ha reportado que la vía NMD también está implicada en la degradación de muchos mRNAs que codifican proteínas funcionales, por lo que se establece como un mecanismo post-transcripcional de regulación de la expresión génica. Se ha descrito que las células tumorales se aprovechan de la función dual de esta vía para, por una parte, degradar los mRNAs de genes supresores de tumores por adquisición selectiva de codones de parada prematuros (PTCs, del inglés *Premature Termination Codon*), lo que favorece el crecimiento sin restricciones del tumor; y, por otra parte, ajustar la actividad NMD para favorecer la regulación positiva de genes necesarios para la adaptación al microambiente tumoral (Fernandes et al., 2019).

Por otro lado, centrándonos en los *pathways* infraexpresados, la activación reducida de la vía de señalización de los PPAR (receptores activados por proliferadores de peroxisomas), en

concreto a través del receptor PPAR- γ se ha asociado a desdiferenciación celular y mayor invasividad tumoral en líneas celulares de cáncer de mama. Los PPAR- γ pueden ser activados por ligandos naturales como ácidos grasos y sus derivados. Una vez activados, están implicados en la diferenciación adipogénica de las células, lo que resulta en la inhibición del crecimiento y la proliferación celular (Vella et al., 2017). Dado que el metabolismo de los triglicéridos está afectado en los tumores irradiados, la cantidad de ácidos grasos disponibles para activar los PPAR se ve reducida y, consecuentemente, la vía de diferenciación a adipocitos resulta infractivada en estos tumores, desencadenando un fenotipo más agresivo y metastásico.

5. CONCLUSIÓN

El objetivo del presente proyecto era comprobar si la radiación no dirigida sobre el microambiente tumoral tenía algún efecto en el proceso de carcinogénesis en un epitelio mamario preparado para sufrir una transformación neoplásica mediante la pérdida genética de p53.

El 100% de los ratones incluidos en el estudio desarrollaron un tumor mamario de 10 mm³, independientemente de la condición de irradiación, lo que demuestra que la pérdida de p53 es suficiente para inducir la transformación neoplásica. No obstante, observamos que, además, la radiación ionizante, un conocido carcinógeno, era capaz de acelerar el proceso de tumorigénesis actuando tanto sobre las propias células tumorales como sobre la señalización del microambiente tumoral.

Mediante IHC comprobamos que el daño en el DNA producido por la radiación ionizante promueve la adquisición de mutaciones con alto potencial carcinogénico, como la amplificación del oncogén HER2 o la inactivación del supresor tumoral E-cadherina, y estimula la sobreexpresión de CD14 por liberación de DAMPs, promoviendo así el establecimiento de un microambiente tumoral inflamatorio y más proliferativo. De igual manera, tras el análisis de expresión génica diferencial por RNA-seq, observamos que tanto las vías de señalización sobreactivadas como las reducidas en ratones irradiados conducían, mayoritariamente, al desarrollo de tumores más agresivos, por alteraciones en las uniones celulares, inducción de angiogénesis o desregulación del ciclo celular. En general, estos resultados sostienen la idea de que la radiación ionizante está asociada al desarrollo de tumores caracterizados por una tasa de crecimiento acelerada, mayor invasividad y un elevado potencial metastásico.

De modo que, podemos concluir que tanto las células tumorales como el microambiente tumoral parecen estar afectados por los efectos de la radiación y que ésta podría jugar un papel importante en el proceso de carcinogénesis. Sin embargo, no pudimos encontrar diferencias significativas en la tasa de crecimiento tumoral entre los ratones irradiados y los ratones no irradiados, lo cual podría deberse al pequeño tamaño muestral (n) que resultaba al dividir los ratones en dos grupos: irradiados y controles no irradiados. Pese a ello, nuestros resultados tienden a mostrar ligeras diferencias de expresión en los tumores tratados con radiación 10 cGy, que deberían ser confirmadas en estudios con una n mayor y que podrían esclarecer los efectos de la radiación en el proceso de tumorigénesis.

6. ANEXOS

ANEXO I. Especificaciones de los anticuerpos primarios empleados para la IHC.

Anticuerpo	Clona	Compañía	Dilución	Kit de detección
ER: Estrogen Receptor Rabbit Monoclonal Primary Antibody, CONFIRM	SP1	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit
PR: Progesterone Receptor Rabbit Monoclonal Primary Antibody, CONFIRM	1E2	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit
HER2: HER-2/neu Rabbit Monoclonal Primary Antibody, PATHWAY	4B5	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit
Ki67: Ki67 Rabbit Monoclonal Primary Antibody, CONFIRM	30-9	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit
c-Myc: c-MYC Rabbit Monoclonal Primary Antibody	Y69	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit
CK14: Cytokeratin 14 Rabbit Monoclonal Primary Antibody	SP53	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit
CD14: CD14 Rabbit Monoclonal Primary Antibody	EPR3653	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit
E-Cadherina: E-cadherin Antibody	EP700Y	Ventana Roche	Pre-diluido	ultraView Universal DAB Detection Kit

ANEXO II. Sistema de puntuación utilizado en la evaluación de los distintos biomarcadores por IHC.

Especificidad del anticuerpo	Patrón de tinción observado	Valoración
ER(Estrogen Receptor)	Nuclear	Allred score: valores comprendidos entre 0-2 (negativos) vs 3-8 (positivos) (Campbell et al., 2016)
PR (Progesterone Receptor)	Nuclear/Citoplásmico	Allred score: valores comprendidos entre 0-2 (negativos) vs 3-8 (positivos) (Campbell et al., 2016)
HER2 (Human epidermal receptor type 2)	Membranoso/Nuclear	Para valorar la expresión de HER2 se distinguen cuatro grupos: 0 (no se observa tinción de membrana o ésta es incompleta y apenas perceptible en $\leq 10\%$ de las células tumorales); 1 (se observa tinción tenue e incompleta de membrana y apenas perceptible en $> 10\%$ de células tumorales); 2 (tinción débil pero completa de membrana en $> 10\%$ de las células tumorales); 3 (tinción de membrana completa e intensa en $> 10\%$ de las células tumorales). Los grupos 0 y 1 se consideran negativos y los grupos 2 y 3 se consideran positivos (Mohammadizadeh et al., 2009).
Ki67	Nuclear	$<14\%$ vs. $\geq 14\%$ de células teñidas (<i>cut-off</i> establecido en el consenso de St. Gallen de 2013) (Harbeck et al., 2013).
c-Myc	Nuclear	Se han descrito dos <i>scores</i> para considerar una tinción positiva para c-Myc, uno de ellos basado en la intensidad de tinción de las células (0: no se observa tinción; 1: tinción débil; 2: tinción moderada; 3: tinción fuerte) y otro <i>score</i> basado en el porcentaje de células teñidas (0: 0%; 1: 1-25%; 2: 26-50%; 3: 51-75%; 4: 76-100%). Se considera positivo un <i>score</i> de intensidad >1 ó un <i>score</i> de porcentaje >3 (Blancato et al., 2004).
CK14	Citoplásmico	Los tumores se clasifican en función de dos <i>scores</i> : la intensidad de tinción (0: no tinción; 1: tinción débil; 2: tinción moderada/fuerte) y el porcentaje de células que muestran expresión de CK14 (0: 0%; 1: $\leq 10\%$; 2: $>10\%$). La multiplicación de estos dos parámetros da un <i>score</i> final: ≤ 1 se considera negativo y ≥ 2 se considera positivo (Majumder, Jagani and Basu, 2020).
CD14	Nuclear	Al igual que ocurre con otros marcadores de células madre de cáncer de mama, como CD44 y CD24, consideramos expresión positiva de CD14 en aquellas muestras en las que $>10\%$ de las células tumorales estaban teñidas (Katayama et al., 2017).
E-Cadherina	Citoplásmico/ Membranoso	La expresión de E-cadherina se puntuó como 0 o negativa (no tinción); 1 (tinción débil y en $<10\%$ de las células tumorales); 2 (tinción moderada y en $>10\%$ de las células); 3 (tinción intensa en la mayor parte de las células). Los <i>scores</i> 1, 2 y 3 se consideraron positivos (Younis, El Sakka and Haque, 2007).

ANEXO III. Códigos utilizados en el análisis de RNA-seq.

Sección 1: Código bash empleado para llevar a cabo el control de calidad de los ficheros fastq.

```
DEST_DIR=/Destination_directory/
FILES=/Directory_Containing_Files/*
for f in $FILES
do
    echo $f
    var_temp=$(
    echo $f | cut -d '/' -f 6 | cut -d '.' -f 1
    )
    echo $var_temp
    FILE_NAME=$(
    echo $f | cut -d '/' -f 6
    )
    NEW_FOLDER=$(echo $DEST_DIR$var_temp)
    echo "The new folder to create is..."
    echo $NEW_FOLDER
    echo "The file to transfer is..."
    echo $f
    echo "The name of the file is..."
    echo $FILE_NAME
    echo "The unzipping file plus dir is..."
    FILE_TO_UNZIP=$(echo $NEW_FOLDER/'/$FILE_NAME)
    echo "The unzipping file in a variable is..."
    echo $FILE_TO_UNZIP
    #echo "copy the file to the new directory..."
    mkdir -p $NEW_FOLDER
    #cp $f $NEW_FOLDER
    #echo "Uncompressing file..."
    #gunzip -d $FILE_TO_UNZIP
    #echo "Removing compressed file..."
    #rm -f $FILE_TO_UNZIP
    fastqc $f --outdir=$NEW_FOLDER
done
```

Sección 2: Código bash utilizado para generar los alineamientos.

```
DEST_DIR=/Destination_directory/
FILES=/Directory_Containing_Files/*
for f in $FILES
do
    echo $f
    var_temp=$(
    echo $f | cut -d'/' -f 6 | cut -d'.' -f 1
    )
    echo $var_temp
    FILE_NAME=$(
    echo $f | cut -d'/' -f 6
    )
    echo "The name of the samples is..."
    echo $var_temp
    NEW_FOLDER=$(echo $DEST_DIR$var_temp)
    echo "The new folder to create is..."
    echo $NEW_FOLDER
    echo "The file to transfer is..."
    echo $f
    echo "The name of the file is..."
    echo $FILE_NAME
    echo "The unzipping file plus dir is..."
    FILE_TO_UNZIP=$(echo $NEW_FOLDER'/'$FILE_NAME)
    echo $FILE_TO_UNZIP
    echo "copy the file to the new directory..."
    #mkdir -p $NEW_FOLDER
    cp $f $NEW_FOLDER
    echo "Uncompressing file..."
    gunzip -d $FILE_TO_UNZIP
    echo "Removing compressed file..."
    rm -f $FILE_TO_UNZIP
    echo "The unzipped file name is..."
    UNZIPPED_FILE=${FILE_TO_UNZIP::-3}
    echo $UNZIPPED_FILE
    echo "This is the forward file name..."
    FORWARD_FILE=$(echo $NEW_FOLDER$'/forward.fastq')
    echo $FORWARD_FILE
    echo "This is the reverse file name..."
    REVERSE_FILE=$(echo $NEW_FOLDER$'/reverse.fastq')
    echo $REVERSE_FILE
    echo "Deinterleaving data..."
    echo 'deinterleave_fastq.sh < '$UNZIPPED_FILE' '$FORWARD_FILE'
```

```
'$REVERSE_FILE
  deinterleave_fastq.sh < $UNZIPPED_FILE $FORWARD_FILE
$REVERSE_FILE
  deinterleave_fastq.sh < $UNZIPPED_FILE $FORWARD_FILE
$REVERSE_FILE
  #fastqc $f --outdir=$NEW_FOLDER
  echo "The name of the summary statistics file is..."
  SUMMARY_FILE=$(echo $NEW_FOLDER$'/Summary_File.txt')
  echo "The name of the output SAM is..."
  OUTPUT_SAM=$(echo $NEW_FOLDER$'/${var_temp}_SAM_out.sam')
  echo $OUTPUT_SAM
  echo $SUMMARY_FILE
  echo "Running alignment..."
  hisat2 -x genome -1 $FORWARD_FILE -2 $REVERSE_FILE -S $OUTPUT_SAM
-3 25 -p 8 --summary-file $SUMMARY_FILE
  echo "Removing data..."
  rm -f $UNZIPPED_FILE
  rm -f $FORWARD_FILE
  rm -f $REVERSE_FILE
  rm -f $f
done
```

Sección 3: Código de R empleado para generar los conteos a nivel de gen.

```
directories_to_find_files <- dir("/PATH_TO_SAM_FILES/",full.names
= T)
file_dir_names <- c()
for(i in 1:length(directories_to_find_files)){
  file_dir_names <-
c(file_dir_names,dir(directories_to_find_files[i],full.names =
T)[grepl(".sam",dir(directories_to_find_files[i],full.names =
T))])
}

Gene_Counts <- featureCounts(file_dir_names,annot.inbuilt =
"mm10",isPairedEnd = TRUE,useMetaFeatures = TRUE)
Gene_Counting_Stats <- t(Gene_Counts$stat)
for(i in 1:ncol(Gene_Counting_Stats)){
  Gene_Counting_Stats[,i] <-
as.numeric(as.character(trimws(Gene_Counting_Stats[,i])))
}
colnames(Gene_Counting_Stats) <- Gene_Counts$stat$Status
Gene_Counting_Stats <- Gene_Counting_Stats[-1,]
rownames(Gene_Counting_Stats) <-
unlist(lapply(strsplit(rownames(Gene_Counting_Stats),"\\."),functi
on(x) x[[8]]))

library(cliplr)

save(file = "/PATH_TO_COUNTS/Gene_Level_Counts.Rda",Gene_Counts)
Gene_Counts <-
get(load(file="/PATH_TO_COUNTS/Gene_Level_Counts.Rda"))
help(featureCounts)
```

Sección 4: Código de R empleado para efectuar los análisis de expresión diferencial y enriquecimiento en categorías funcionales.

```
options(stringsAsFactors = F)
library(DESeq2)

#Read counts.

counts_marta <- get(load(file =
"/DIR_ANALYSIS/Gene_Level_Counts.Rda" ))
names_to_replace <-
unlist(lapply(strsplit(colnames(counts_marta$counts), "\\."), function(x) x[8])))
colnames(counts_marta$counts) <- names_to_replace
colnames(counts_marta$stat)[2:163] <- names_to_replace

# Loading sample treatment.

Sample_Treatment <- read.csv(file =
"/DIR_ANALYSIS//Sample_Codes.txt", sep= "\t")

intersected <-
intersect(colnames(counts_marta$counts), Sample_Treatment$Library)
counts_marta$counts <- counts_marta$counts[,intersected]
Sample_Treatment <- Sample_Treatment[Sample_Treatment$Library %in%
intersected,]
countData <- counts_marta$counts
colData <- Sample_Treatment[,c("Library", "Treatment")]
rownames(colData) <- colData$Library
colnames(colData)[2] <- "condition"
colData <- colData[,-1, drop = F]
colData <- colData[colnames(countData),, drop=F]

# Constructing DeSeq2 object.

dataset <- DESeqDataSetFromMatrix(countData = countData, colData =
colData, design = ~condition)

# We carried out minimal filtering of the genes with low counts.

keep <- (rowSums(counts(dataset))/ncol(dataset)) > 10
table(keep)
dataset <- dataset[keep,]
dataset$condition <- relevel(dataset$condition, ref = "sham")
table(dataset$condition)
sum(table(dataset$condition))
```

```

# Running the DESeq2 algorithm and performing the two-class
comparison.

dds <- DESeq(dataset)
summary(dds)
result <- results(dds, contrast=c('condition','IR','sham'),alpha =
0.05)
result <- result[complete.cases(result),] #remove any rows with
NA

# Creating an annotated results dataframe.

df_results <- data.frame(result)
View(df_results)
rownames(df_results)
df_results$Entrez <- rownames(df_results)
#Annotating results using biomaRt.

library(biomaRt)
library(biomaRt)
library(DESeq2)
library(tidyverse)

listMarts()
ensembl=useMart("ENSEMBL_MART_ENSEMBL")
listDatasets(ensembl) %>% filter(str_detect(description, "Mouse"))
ensembl = useDataset("mmusculus_gene_ensembl", mart=ensembl)
ourFilterType <- "entrezgene_id"
filterValues <- rownames(df_results)
attributeNames <- c('ensembl_gene_id', 'entrezgene_id',
'external_gene_name')
annot <- getBM(attributes=attributeNames,filters =
ourFilterType,values = filterValues, mart = ensembl)

df_results <- merge(df_results,annot,by.x = "Entrez", by.y =
"entrezgene_id")
df_results_filt <- df_results[df_results$padj < 0.05,]
colnames(df_results_filt)

source("/DIR_ANALYSIS/Misc.R")
df_results_filt <-
df_results_filt[,c("Entrez","baseMean","log2FoldChange","pvalue","
padj","external_gene_name")]
df_results_filt$baseMean <- round(df_results_filt$baseMean,digits
= 2)
df_results_filt$log2FoldChange <-
round(df_results_filt$log2FoldChange,digits = 2)
df_results_filt$pvalue <- formatC(df_results_filt$pvalue,format =
"e",digits = 2)
df_results_filt$padj <- formatC(df_results_filt$padj,format =
"e",digits = 2)

```

```

copy.table(df_results_filt)
copy.table(paste(df_results_filt[df_results_filt$log2FoldChange >
0,"external_gene_name"],collapse = ", "))
copy.table(paste(df_results_filt[df_results_filt$log2FoldChange <
0,"external_gene_name"],collapse = ", "))
copy.table(df_results_filt$external_gene_name)
# Plotting graphs

library(ComplexHeatmap)
library(circlize)

sigGenes <- df_results[df_results$padj < 0.05,"Entrez"]
vst <- vst(dataset, blind=TRUE)
vst <- vst[,order(colData(vst)$condition)]
plotDat <- assay(vst)[sigGenes,]
z.mat <- t(scale(t(plotDat), center=TRUE, scale=TRUE))
hal = HeatmapAnnotation(df = colData(vst)[,c("condition"),drop=F])
pdf(file = "C:/Users/jjj/Desktop/Analisis_Marta/Heatmap.pdf")
Heatmap(z.mat,top_annotation = hal,cluster_columns = FALSE,rect_gp
= gpar(col = "white", lwd=0.5), row_names_gp = gpar(fontsize =
4),column_names_gp = gpar(fontsize = 4) )
dev.off()

pdf(file = "/DIR_ANALYSIS//plotMA.pdf")
plotMA(result, main=paste0('Condition: Irradiated vs. Sham'))
#, ylim=c(-5,5)
dev.off()
vst <- vst(dataset, blind=TRUE)
pdf(file = "C/DIR_ANALYSIS//plotPCA.pdf")
plotPCA(vst, intgroup = 'condition')
dev.off()
View(data.frame(result))

plotCounts(dataset, gene=which.min(result$pvalue),
intgroup='condition', pch = 19)

library(cowplot)

fig_lab_size <- 8

heat_map <- ggdraw() + draw_image("/DIR_ANALYSIS//Heatmap-1.png")
plotMA <- ggdraw() + draw_image("C/DIR_ANALYSIS//plotMA-1.png")
plotPCA <- ggdraw() + draw_image("/DIR_ANALYSIS//plotPCA-1.png")

plot_grid()
top_row <- plot_grid(plotMA,plotPCA,ncol = 2,labels = c('A','B'))
bottom_row <- plot_grid(heat_map,labels = c("C"))
pdf(file = "/DIR_ANALYSIS//Combined_figure.pdf" )
plot_grid(top_row,bottom_row,nrow = 2)
dev.off()

```



```

# Gene set enrichment analysis.

library(clusterProfiler)

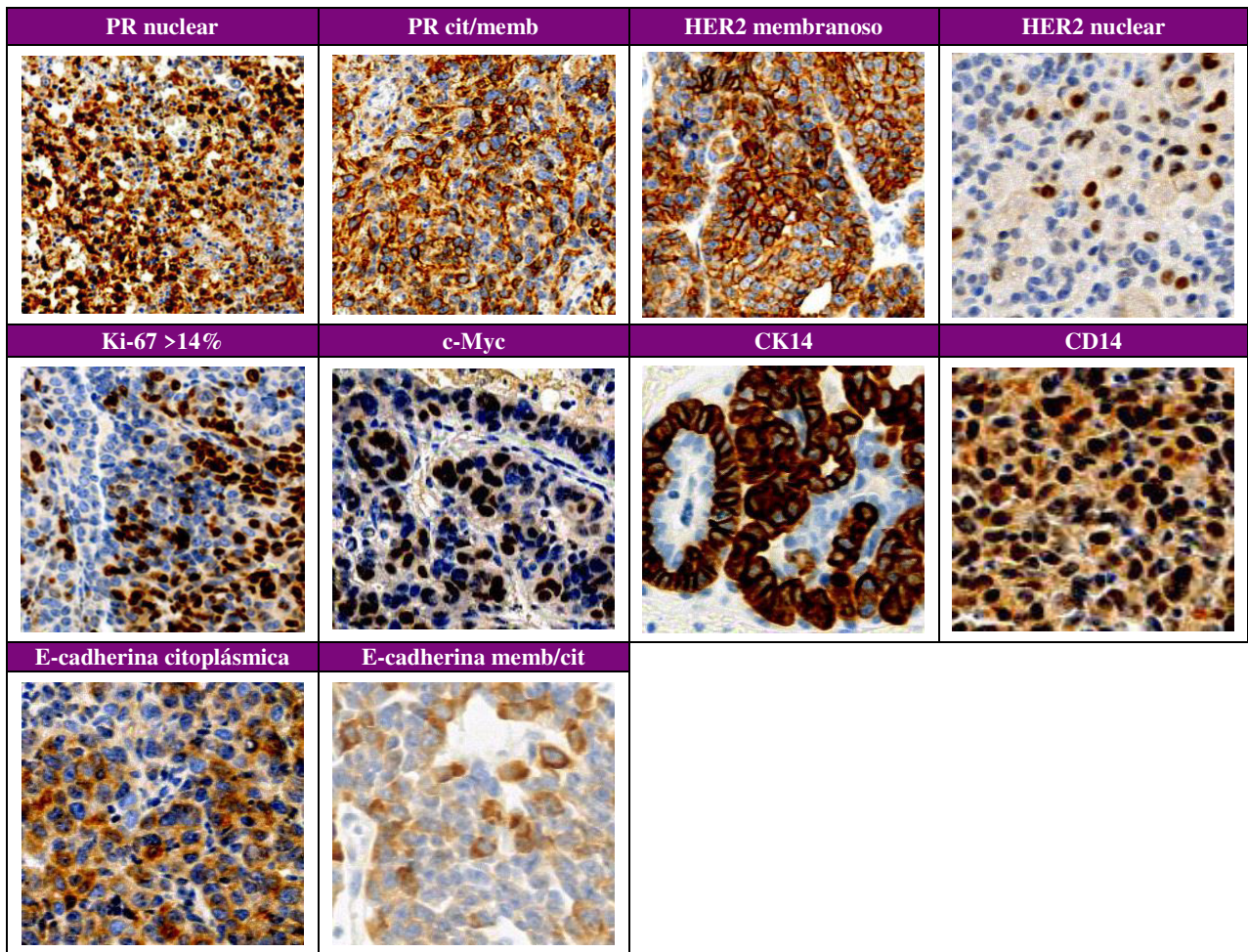
df_results
fold_vector <- df_results$log2FoldChange
names(fold_vector) <- df_results$Entrez
c2_cp <- readRDS(file = "/DIR_ANALYSIS/Mm.c2.cp.v7.1.entrez.rds")
#c2_cp <- get(load(file =
"C:/Users/jjj/Desktop/Analisis_Marta/Mm.c5.all.entrez.RData"))

library(fgsea)
fgseaRes <- fgsea(pathways = c2_cp,stats= fold_vector,minSize =
15, maxSize = 500)
View(fgseaRes[order(pval), ])

df_results <- data.frame(fgseaRes[order(pval), ])
table(df_results$padj < 0.05)
table(df_results$padj < 0.05 & df_results$NES > 0)
View(df_results[df_results$padj < 0.05 & df_results$NES > 0,])
View(df_results[df_results$padj < 0.05 & df_results$NES < 0,])
table(df_results$padj < 0.05)
df_ord <- df_results[df_results$padj < 0.05,]
df_ord
colnames(df_ord)
df_ord <- df_ord[,c("pathway","ES","NES","pval","padj")]
df_ord <- df_ord[order(df_ord$NES,decreasing = T),]
df_ord$ES <- round(df_ord$ES,digits = 2)
df_ord$NES <- round(df_ord$NES,digits = 2)
df_ord$pval <- formatC(df_ord$pval,format = "e",digits = 2)
df_ord$padj <- formatC(df_ord$padj,format = "e",digits = 2)
copy.table(df_ord)

```

ANEXO IV. Ejemplos de inmunotinción positiva para los biomarcadores analizados en los TMAs.



ANEXO V. Correlaciones detectadas entre los distintos biomarcadores analizados en la población en general (**A**), en los ratones irradiados (**B**) y en los ratones no irradiados (**C**).

		PR nuc	PR cyto	HER2 mb	HER2 cyto	Ki67	c-Myc	CK14	CD14	E-cadh	
A	PR nuc	Correlación de Pearson	1	,365**	,042	-,086	,044	-,098	,046	,195*	,042
		Sig. (bilateral)		<,001	,643	,339	,619	,269	,603	,027	,639
		N	130	130	127	126	130	130	128	129	126
	PR cyto	Correlación de Pearson	,365**	1	-,034	,049	,131	-,036	,000	,298**	-,112
		Sig. (bilateral)	<,001		,708	,582	,138	,687	1,000	<,001	,212
		N	130	130	127	126	130	130	128	129	126
	HER2 mb	Correlación de Pearson	,042	-,034	1	,124	,113	-,095	,118	,092	,001
		Sig. (bilateral)	,643	,708		,165	,204	,286	,185	,303	,987
		N	127	127	128	127	127	127	127	128	124
	HER2 cyto	Correlación de Pearson	-,086	,049	,124	1	,091	-,117	,138	-,001	,077
		Sig. (bilateral)	,339	,582	,165		,310	,193	,124	,992	,399
		N	126	126	127	127	126	126	126	127	123
	Ki67	Correlación de Pearson	,044	,131	,113	,091	1	,192*	,100	,275**	,081
		Sig. (bilateral)	,619	,138	,204	,310		,029	,261	,002	,366
		N	130	130	127	126	130	130	128	129	126
	c-Myc	Correlación de Pearson	-,098	-,036	-,095	-,117	,192*	1	,149	,158	,211*
		Sig. (bilateral)	,269	,687	,286	,193	,029		,092	,074	,018
		N	130	130	127	126	130	130	128	129	126
	CK14	Correlación de Pearson	,046	,000	,118	,138	,100	,149	1	,077	,351**
		Sig. (bilateral)	,603	1,000	,185	,124	,261	,092		,387	<,001
		N	128	128	127	126	128	128	129	128	124
	CD14	Correlación de Pearson	,195*	,298**	,092	-,001	,275**	,158	,077	1	,324**
		Sig. (bilateral)	,027	<,001	,303	,992	,002	,074	,387		<,001
		N	129	129	128	127	129	129	128	130	126
	E-cadh	Correlación de Pearson	,042	-,112	,001	,077	,081	,211*	,351**	,324**	1
		Sig. (bilateral)	,639	,212	,987	,399	,366	,018	<,001	<,001	
		N	126	126	124	123	126	126	124	126	126

** . La correlación es significativa en el nivel 0,01 (bilateral).

* . La correlación es significativa en el nivel 0,05 (bilateral).

Correlaciones 10 cGy

B

		PR nuc	PR cyto	HER2 mb	HER2 cyto	Ki67	c-Myc	CK14	CD14	E-cadh
PR nuc	Correlación de Pearson	1	,511**	,112	-,106	,139	-,095	,148	,217	,124
	Sig. (bilateral)		<,001	,410	,438	,293	,476	,268	,102	,366
	N	59	59	56	56	59	59	58	58	55
PR cyto	Correlación de Pearson	,511**	1	,165	-,080	,163	,075	,071	,212	-,032
	Sig. (bilateral)	<,001		,223	,558	,217	,572	,594	,110	,817
	N	59	59	56	56	59	59	58	58	55
HER2 mb	Correlación de Pearson	,112	,165	1	,191	,098	-,109	,038	,154	-,026
	Sig. (bilateral)	,410	,223		,158	,473	,425	,783	,259	,855
	N	56	56	56	56	56	56	56	56	53
HER2 cyto	Correlación de Pearson	-,106	-,080	,191	1	,145	-,256	,076	,143	,099
	Sig. (bilateral)	,438	,558	,158		,286	,056	,575	,294	,479
	N	56	56	56	56	56	56	56	56	53
Ki67	Correlación de Pearson	,139	,163	,098	,145	1	,200	,224	,338**	,258
	Sig. (bilateral)	,293	,217	,473	,286		,129	,090	,009	,057
	N	59	59	56	56	59	59	58	58	55
c-Myc	Correlación de Pearson	-,095	,075	-,109	-,256	,200	1	,106	,157	,091
	Sig. (bilateral)	,476	,572	,425	,056	,129		,427	,239	,507
	N	59	59	56	56	59	59	58	58	55
CK14	Correlación de Pearson	,148	,071	,038	,076	,224	,106	1	,209	,288*
	Sig. (bilateral)	,268	,594	,783	,575	,090	,427		,118	,035
	N	58	58	56	56	58	58	58	57	54
CD14	Correlación de Pearson	,217	,212	,154	,143	,338**	,157	,209	1	,582**
	Sig. (bilateral)	,102	,110	,259	,294	,009	,239	,118		<,001
	N	58	58	56	56	58	58	57	58	55
E-cadh	Correlación de Pearson	,124	-,032	-,026	,099	,258	,091	,288*	,582**	1
	Sig. (bilateral)	,366	,817	,855	,479	,057	,507	,035	<,001	
	N	55	55	53	53	55	55	54	55	55

** La correlación es significativa en el nivel 0,01 (bilateral).

* La correlación es significativa en el nivel 0,05 (bilateral).

Correlaciones SHAM

C

		PR nuc	PR cyto	HER2 mb	HER2 cyto	Ki67	c-Myc	CK14	CD14	E-cadh
PR nuc	Correlación de Pearson	1	,292*	-,044	-,071	-,042	-,099	-,046	,178	-,007
	Sig. (bilateral)		,013	,715	,558	,728	,410	,708	,138	,952
	N	71	71	71	70	71	71	70	71	71
PR cyto	Correlación de Pearson	,292*	1	-,151	,151	,123	-,116	-,047	,349**	-,154
	Sig. (bilateral)	,013		,209	,211	,307	,334	,698	,003	,199
	N	71	71	71	70	71	71	70	71	71
HER2 mb	Correlación de Pearson	-,044	-,151	1	,017	,127	-,066	,203	,051	,026
	Sig. (bilateral)	,715	,209		,886	,293	,586	,090	,670	,829
	N	71	71	72	71	71	71	71	72	71
HER2 cyto	Correlación de Pearson	-,071	,151	,017	1	,045	-,007	,187	-,108	,057
	Sig. (bilateral)	,558	,211	,886		,714	,956	,122	,368	,637
	N	70	70	71	71	70	70	70	71	70
Ki67	Correlación de Pearson	-,042	,123	,127	,045	1	,190	-,002	,228	-,054
	Sig. (bilateral)	,728	,307	,293	,714		,113	,989	,056	,655
	N	71	71	71	70	71	71	70	71	71
c-Myc	Correlación de Pearson	-,099	-,116	-,066	-,007	,190	1	,182	,154	,299*
	Sig. (bilateral)	,410	,334	,586	,956	,113		,132	,199	,011
	N	71	71	71	70	71	71	70	71	71
CK14	Correlación de Pearson	-,046	-,047	,203	,187	-,002	,182	1	-,026	,399**
	Sig. (bilateral)	,708	,698	,090	,122	,989	,132		,828	<,001
	N	70	70	71	70	70	70	71	71	70
CD14	Correlación de Pearson	,178	,349**	,051	-,108	,228	,154	-,026	1	,129
	Sig. (bilateral)	,138	,003	,670	,368	,056	,199	,828		,284
	N	71	71	72	71	71	71	71	72	71
E-cadh	Correlación de Pearson	-,007	-,154	,026	,057	-,054	,299*	,399**	,129	1
	Sig. (bilateral)	,952	,199	,829	,637	,655	,011	<,001	,284	
	N	71	71	71	70	71	71	70	71	71

*. La correlación es significativa en el nivel 0,05 (bilateral).

** La correlación es significativa en el nivel 0,01 (bilateral).

7. BIBLIOGRAFÍA

- Acs, G., Lawton, T. J., Rebbeck, T. R., LiVolsi, V. A., and Zhang, P. J. (2001). Differential expression of E-cadherin in lobular and ductal neoplasms of the breast and its biologic and diagnostic implications. *American journal of clinical pathology*, 115(1), 85-98.
- Ali, M. U., Ur Rahman, M. S., Jia, Z., and Jiang, C. (2017). Eukaryotic translation initiation factors and cancer. *Tumor Biology*, 39(6).
- American Cancer Society (2020). *Cáncer de seno*. Disponible online: <https://www.cancer.org/es/cancer/cancer-de-seno.html> [Acceso: 21-Marzo-2020].
- American Cancer Society (2019). Breast Cancer Facts & Figures 2019-2020. *American Cancer Society*, 1-44. Disponible online: <https://www.cancer.org/content/dam/cancer.org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf> [Acceso: 21-Marzo-2020].
- Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J., and Topisirovic, I. (2015). Targeting the translation machinery in cancer. *Nature reviews Drug discovery*, 14(4), 261-278.
- Bioconductor (2020). *Deseq2: Differential gene expression analysis based on the negative binomial distribution*. Disponible online: <https://bioconductor.org/packages/release/bioc/html/DESeq2.html> [Acceso: 14-Julio-2020].
- Blancato, J., Singh, B., Liu, A., Liao, D. J., and Dickson, R. B. (2004). Correlation of amplification and overexpression of the c-myc oncogene in high-grade breast cancer: FISH, in situ hybridisation and immunohistochemical analyses. *British journal of cancer*, 90(8), 1612-1619.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians*, 68(6), 394-424.
- Buda, A., and Pignatelli, M. (2011). E-cadherin and the cytoskeletal network in colorectal cancer development and metastasis. *Cell communication & adhesion*, 18(6), 133-143.
- Campbell, E. J., Tesson, M., Doogan, F., Mohammed, Z. M., Mallon, E., and Edwards, J. (2016). The combined endocrine receptor in breast cancer, a novel approach to traditional hormone receptor interpretation and a better discriminator of outcome than ER and PR alone. *British Journal of Cancer*, 115(8), 967-973.
- Castiglioni, F., Terenziani, M., Carcangiu, M. L., Miliano, R., Aiello, P., Bertola, L., Triulzi, T., Gasparini, P., Camerini, T., Sozzi, G., Fossati-Bellani, F., Ménard, S. and Tagliabue, E. (2007). Radiation effects on development of HER2-positive breast carcinomas. *Clinical Cancer Research*, 13(1), 46-51.
- Cheah, M. T., Chen, J. Y., Sahoo, D., Contreras-Trujillo, H., Volkmer, A. K., Scheeren, F. A., Volkmer, J. P. and Weissman, I. L. (2015). CD14-expressing cancer cells establish the inflammatory and proliferative tumor microenvironment in bladder cancer. *Proceedings of the National Academy of Sciences*, 112(15), 4725-4730.

- Cho, K. B., Cho, M. K., Lee, W. Y., and Kang, K. W. (2010). Overexpression of c-myc induces epithelial mesenchymal transition in mammary epithelial cells. *Cancer letters*, 293(2), 230-239.
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., and Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *American Journal of Cancer Research*, 5(10), 2929-43.
- Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M. and Jaffrézic, F. (2012). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics*, 14(6), 671-683.
- Derksen, P. W. B., Liu, X., Saridin, F., van der Gulden, H., Zevenhoven, J., Evers, B., van Beijnum, J. R., Griffioen, A. W., Vink, J., Krimpenfort, P., Peterse, J. L., Cardiff, R. D., Berns, A. and Jonkers, J. (2006). Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis. *Cancer cell*, 10(5), 437-449.
- Eroles, P., Bosch, A., Pérez-Fidalgo, J. A., and Lluch, A. (2012). Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer treatment reviews*, 38(6), 698-707.
- Fallah, Y., Brundage, J., Allegakoen, P. and Shajahan-Haq, A. N. (2017). MYC-driven pathways in breast cancer subtypes. *Biomolecules*, 7(3), 53.
- Fernandes, R., Nogueira, G., da Costa, P. J., Pinto, F., and Romão, L. (2019). Nonsense-mediated mRNA decay in development, stress and cancer. In *The mRNA Metabolism in Human Disease* (pp. 41-83). Springer, Cham.
- Graham, J. D., Yeates, C., Balleine, R. L., Harvey, S. S., Milliken, J. S., Bilous, A. M., and Clarke, C. L. (1996). Progesterone receptor A and B protein expression in human breast cancer. *The Journal of steroid biochemistry and molecular biology*, 56(1-6), 93-98.
- Guo, L., Li, W., Zhu, X., Ling, Y., Qiu, T., Dong, L., Fang, Y., Yang, H. and Ying, J. (2016). PD-L1 expression and CD274 gene alteration in triple-negative breast cancer: implication for prognostic biomarker. *SpringerPlus*, 5(1), 1-8.
- Harbeck, N., Thomssen, C., and Gnant, M. (2013). St. Gallen 2013: brief preliminary summary of the consensus discussion. *Breast care*, 8(2), 102-109.
- Harigopal, M., Shin, S. J., Murray, M. P., Tickoo, S. K., Brogi, E., and Rosen, P. P. (2005). Aberrant E-cadherin staining patterns in invasive mammary carcinoma. *World Journal of Surgical Oncology*, 3(1), 73.
- Harvard Chan Bioinformatics Core. (n.d.). *Introduction to DGE: Count normalization with DESeq2*. GitHub. Disponible online: https://hbctraining.github.io/DGE_workshop/lessons/02_DGE_count_normalization.html [Acceso: 9-Noviembre-2020].
- Holen, I., Speirs, V., Morrissey, B. and Blyth, K. (2017). In vivo models in breast cancer research: progress, challenges and future directions. *Disease models & mechanisms*, 10(4), 359-371.

- Horiguchi, S. I., Hishima, T., Hayashi, Y., Shiozawa, Y., Horiguchi, K., Kuroi, K., Toi, M. Funata, N. and Eishi, Y. (2010). HER-2/neu cytoplasmic staining is correlated with neuroendocrine differentiation in breast carcinoma. *Journal of medical and dental sciences*, 57(2), 155-163.
- Illumina. (2017). *An introduction to Next-Generation Sequencing Technology*. Disponible online: https://emea.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf [Acceso: 6-Abril-2020].
- Jerry, D. J., Kittrell, F. S., Kuperwasser, C., Laucirica, R., Dickinson, E. S., Bonilla, P. J., Butel, J. S. and Medina, D. (2000). A mammary-specific model demonstrates the role of the p53 tumor suppressor gene in tumor development. *Oncogene*, 19(8), 1052-1058.
- Katayama, A., Handa, T., Komatsu, K., Togo, M., Horiguchi, J., Nishiyama, M., and Oyama, T. (2017). Expression patterns of claudins in patients with triple-negative breast cancer are associated with nodal metastasis and worse outcome. *Pathology international*, 67(8), 404-413.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, 37(8), 907-915.
- Leslie, K. K., Stein, M. P., Kumar, N. S., Dai, D., Stephens, J., Wandinger-Ness, A., and Glueck, D. H. (2005). Progesterone receptor isoform identification and subcellular localization in endometrial cancer. *Gynecologic oncology*, 96(1), 32-41.
- Li, L., Wang, Q., Lv, X., Sha, L., Qin, H., Wang, L., and Li, L. (2015). Expression and localization of estrogen receptor in human breast cancer and its clinical significance. *Cell biochemistry and biophysics*, 71(1), 63-68.
- Lim, C. S., Baumann, C. T., Htun, H., Xian, W., Irie, M., Smith, C. L., and Hager, G. L. (1999). Differential localization and activity of the A-and B-forms of the human progesterone receptor using green fluorescent protein chimeras. *Molecular endocrinology*, 13(3), 366-375.
- Liu, M., Zhu, H., Yang, S., Wang, Z., Bai, J., and Xu, N. (2013). c-Myc suppressed E-cadherin through miR-9 at the post-transcriptional level. *Cell biology international*, 37(3), 197-202.
- Lobba, A. R. M., Forni, M. F., Carreira, A. C. O., and Sogayar, M. C. (2012). Differential expression of CD90 and CD14 stem cell markers in malignant breast cancer cell lines. *Cytometry Part A*, 81(12), 1084-1091.
- Ma, L., Young, J., Prabhala, H., Pan, E., Mestdagh, P., Muth, D., Teruya-Feldstein, J., Reinhardt, F., Onder, T. T., Valastyan, S., Westermann, F., Speleman, F., Vandesompele, J. and Weinberg, R. A. (2010). miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nature cell biology*, 12(3), 247-256.
- Majumder, A., Jagani, R. and Basu, A. (2020). Double-positive in triple-negative? How significant is basal cytokeratin expression in breast cancer?. *Medical Journal Armed Forces India*, 76(1), 63-70.
- Manning, H. C., Buck, J. R. and Cook, R. S. (2016). Mouse models of breast cancer: platforms for discovering precision imaging diagnostics and future cancer medicine. *Journal of nuclear medicine: official publication, Society of Nuclear Medicine*, 57(Suppl 1), 60S-68S.

- McFarlane, R. J., and Wakeman, J. A. (2017). Meiosis-like functions in oncogenesis: a new view of cancer. *Cancer research*, 77(21), 5712-5716.
- Mehmeti, M., Allaoui, R., Bergenfelz, C., Saal, L. H., Ethier, S. P., Johansson, M. E., Jirström, K. and Leandersson, K. (2015). Expression of functional toll like receptor 4 in estrogen receptor/progesterone receptor-negative breast cancer. *Breast Cancer Research*, 17(1), 130.
- Mohammadizadeh, F., Naimi, A., Rajabi, P., Ghasemibasir, H., and Eftekhari, A. (2009). Expression of basal and luminal cytokeratins in breast cancer and their correlation with clinicopathological prognostic variables. *Indian journal of medical sciences*, 63(4), 152-162.
- MU DNA Core Facility. (2015). *FastQC Manual*. University of Missouri. Disponible en: https://dnacore.missouri.edu/PDF/FastQC_Manual.pdf [Acceso: 4-Abril-2020]
- Najafi, M., Farhood, B., and Mortezaee, K. (2019). Extracellular matrix (ECM) stiffness and degradation as cancer drivers. *Journal of cellular biochemistry*, 120(3), 2782-2790.
- Nguyen, D. H., Oketch-Rabah, H. A., Illa-Bochaca, I., Geyer, F. C., Reis-Filho, J. S., Mao, J., Ravani, S. A., Zavadil, J., Borowsky, A. D., Jerry, D. J., Dunphy, K. A., Seo, J. H., Haslam, S., Medina, D. and Barcellos-Hoff, M. H. (2011). Radiation Acts on the Microenvironment to Affect Breast Carcinogenesis by Distinct Mechanisms that Decrease Cancer Latency and Affect Tumor Type. *Cancer Cell*, 19(5), 640-651.
- Nieves, S., Apellaniz, D., Tapia, G., Maglia, A., Mosqueda-Taylor, A. and Bologna-Molina, R. (2014). Cytokeratins 14 and 19 in odontogenic cysts and tumors: A review. *Odontoestomatología*, 16, 44-55.
- Ordóñez, N. M., Bastidas, B. E., and Castro, L. I. (2013). Importancia de los receptores hormonales en el cáncer de mama. *Revista Facultad Ciencias de la Salud: Universidad del Cauca*, 15(3), 25-30.
- Ospina, M., and Muñetón, C. M. (2011). Alteraciones del gen c-Myc en la oncogénesis. *Iatreia*, 24(4), 389-401.
- Park, C. C., Henshall-Powell, R. L., Erickson, A. C., Talhouk, R., Parvin, B., Bissell, M. J., and Barcellos-Hoff, M. H. (2003). Ionizing radiation induces heritable disruption of epithelial cell interactions. *Proceedings of the National Academy of Sciences*, 100(19), 10728-10733.
- Park, M. K., Lee, C. H. and Lee, H. (2018). Mouse models of breast cancer in preclinical research. *Laboratory animal research*, 34(4), 160-165.
- Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., and Mayer, G. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific reports*, 8(1), 1-14.
- Plante, I., Stewart, M. K. and Laird, D. W. (2011). Evaluation of mammary gland development and function in mouse models. *JoVE (Journal of Visualized Experiments)*, (53), e2828.
- Puente, R., Bas, E., Peiro, F., Fitor, M., and Ortolá, L. (2015). Cáncer neuroendocrino de mama. A propósito de un caso. *Rev. senol. patol. mamar.(Ed. impr.)*, 143-144.

- Santaballa, A. (2020). *Cáncer de mama*. SEOM: Sociedad Española de Oncología Médica. Disponible online: <https://seom.org/info-sobre-el-cancer/cancer-de-mama?showall=1> [Acceso: 20-Marzo-2020].
- Sarrió, D., Rodríguez-Pinilla, S. M., Hardisson, D., Cano, A., Moreno-Bueno, G., and Palacios, J. (2008). Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer research*, 68(4), 989-997.
- Schillaci, R., Guzmán, P., Cayrol, F., Beguelin, W., Flaqué, M. C. D., Proietti, C. J., Pineda, V., Palazzi, J., Frahm, I., Charreau, E. H., Maronna, E., Roa, J. C. and Elizalde, P. V.. (2012). Clinical relevance of ErbB-2/HER2 nuclear expression in breast cancer. *BMC cancer*, 12(1), 74.
- Shaaban, A. M., Green, A. R., Karthik, S., Alizadeh, Y., Hughes, T. A., Harkins, L., Ellis, I. O., Robertson, J. F., Paish, E. C., Saunders, P. T. K., Groome, N. P., and Speirs, V. (2008). Nuclear and Cytoplasmic Expression of ERB1, ERB2, and ERB5 Identifies Distinct Prognostic Outcome for Breast Cancer Patients. *Clin Cancer Res*, 2008(14), 5228-5235.
- University of Oregon. (n.d.). *RNAseqlopedia*. Disponible en: <https://rnaseq.uoregon.edu/#exp-design> [Acceso: 29-Junio-2020]
- Usary, J., Darr, D. B., Pfefferle, A., and Perou, C. M. (2016). Overview of genetically engineered mouse models of distinct breast cancer subtypes. *Current protocols in pharmacology*, 72(1), 14-38.
- Vella, V., Nicolosi, M. L., Giuliano, S., Bellomo, M., Belfiore, A., and Malaguarnera, R. (2017). PPAR- γ agonists as antineoplastic agents in cancers with dysregulated IGF axis. *Frontiers in Endocrinology*, 8, 31.
- Wang, S. C., Lien, H. C., Xia, W., Chen, I. F., Lo, H. W., Wang, Z., Ali-Seyed, M., Lee, D. F., Bartholomeusz, G., Ou-Yang, F., Giri, D. K. and Hung, M. C. (2004). Binding at and transactivation of the COX-2 promoter by nuclear tyrosine kinase receptor ErbB-2. *Cancer cell*, 6(3), 251-261.
- Weiss, M. B., Vitolo, M. I., Mohseni, M., Rosen, D. M., Denmeade, S. R., Park, B. H., Weber, D. J. and Bachman, K. E. (2010). Deletion of p53 in human mammary epithelial cells causes chromosomal instability and altered therapeutic response. *Oncogene*, 29(33), 4715-4724.
- World Health Organization (2020). *Cancer: Breast cancer*. Disponible online: <https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/> [Acceso: 18-Marzo-2020].
- Younis, L. K., El Sakka, H., and Haque, I. (2007). The prognostic value of E-cadherin expression in breast cancer. *International journal of health sciences*, 1(1), 43-51.
- Zhang, P., Lo, A., Huang, Y., Huang, G., Liang, G., Mott, J., Karpen, G. H., Blakely, E. A., Bissell, M. J., Barcellos-Hoff, M. H., Snijders, A. M. and Mao, J. H. (2015). Identification of genetic loci that control mammary tumor susceptibility through the host microenvironment. *Scientific reports*, 5(1), 8919.