

Estudio del perfil mutacional somático y su impacto en el pronóstico de pacientes con diferentes subtipos de cáncer de mama

Nadia Moltó Lliso

Máster Universitario en Bioinformática y Bioestadística

Área 4. Subárea 1: Estudios genéticos de enfermedades humanas

Tutor UOC: Dra. Helena Brunel Montaner

Profesor responsable: Antoni Pérez Navarro

08/06/2021



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Estudio del perfil mutacional somático y su impacto en el pronóstico de pacientes con diferentes subtipos de cáncer de mama
Nombre del autor:	Nadia Moltó Lliso
Nombre del consultor/a:	Helena Brunel Montaner
Nombre del PRA:	Antoni Pérez Navarro
Fecha de entrega (mm/aaaa):	08/06/2021
Titulación:	Máster universitario en Bioinformática y Bioestadística
Área del Trabajo Final:	Área 4. Subárea 1: Estudios genéticos de enfermedades humanas
Idioma del trabajo:	Español
Número de créditos:	15
Palabras clave	Cáncer de mama, mutación, subtipos, caracterización
Resumen del Trabajo:	
<p>Uno de los puntos clave a la hora de establecer el pronóstico de los pacientes con tumores mamarios reside en el subtipo de cáncer de mama que padecen.</p> <p>El presente estudio parte de la iniciativa de poder consolidar la importancia de las mutaciones somáticas, así como de su perfil mutacional en los pacientes pertenecientes a diferentes subtipos, para alcanzar el objetivo principal de evaluar el impacto que ejercen los genes más representativos a nivel mutacional sobre la caracterización de estos pacientes.</p> <p>Mediante el empleo de representaciones gráficas y análisis estadísticos para los genes mutados en diferentes subtipos de cáncer de mama, se ha podido comprobar como la clasificación PAM50 presenta un patrón mutacional característico para algunos de sus subtipos, pudiendo ser una clasificación interesante a la hora de abordar el objetivo inicial. A través de esta clasificación, se ha visto como determinadas mutaciones en los genes TP53, PIK3CA, AHNK2 y GATA3 influyen de forma directa en la supervivencia de determinados pacientes y pueden ejercer un papel importante a la hora de pronosticar el cáncer de mama. Se ha podido determinar que todos estos genes se encuentran directamente relacionados con rutas que influyen positiva o negativamente en el cáncer de mama y cuyas mutaciones presentan un perfil característico acorde a la intervención sobre los procesos biológicos que realizan. Así pues, el estudio de las mutaciones somáticas ha demostrado tener un potencial notorio que puede y debe ser explotado a la hora de establecer pronósticos en pacientes con cáncer de mama.</p>	

Abstract:

One of the key points when establishing the prognosis of patients with breast tumors lies in the subtype of the breast cancer they suffer from.

The following study is based on the initiative to consolidate the importance of somatic mutations, as well as their mutational profile, in patients belonging to different subtypes. This is done in order to achieve the main objective of evaluating the impact of the most representative genes at the mutational level on the characterization of these patients.

It has been possible to verify, through the use of graphical representations and statistical analysis for the mutated genes in different subtypes of breast cancer, how the PAM50 classification presents a characteristic mutational pattern for some of its subtypes. Which could be a compelling classification when addressing the initial goal. Through this classification, it has been proved how certain mutations in the genes TP53, PIK3CA, AHNK2 and GATA3 directly influence the survival of some patients and can play an important role in predicting breast cancer. It has been possible to determine that all these genes are directly related to pathways that influence breast cancer in a positive or negative way and whose mutations present a characteristic profile according to the intervention on the biological processes carried out by them. Therefore, the study of somatic mutations has shown a remarkable potential that can and ought to be exploited when establishing prognoses in patients with breast cancer.

Índice

1. Introducción	2
1.2 Objetivos del Trabajo	3
1.3 Enfoque y método seguido	5
1.4 Planificación del Trabajo	7
1.5 Breve resumen de productos obtenidos	9
1.6 Breve descripción de los otros capítulos de la memoria	10
2. Estado del arte	11
2.1 Antecedentes	11
2.2 Clasificación del cáncer de mama	11
2.3 Desarrollo del cáncer de mama	13
2.4 Tratamientos frente al cáncer de mama	13
3. Metodología	15
3.1 Pacientes y datos clínicos	15
3.2 Clasificación de los pacientes	16
3.3 Análisis de las mutaciones	17
3.4 Análisis estadístico	18
3.5 Caracterización del perfil mutacional en los subtipos de PAM50	18
3.6 Análisis de supervivencia	18
3.7 Pathway analysis	19
3.8 Interpretación biológica	19
4. Resultados	20
4.1 Mutaciones en los subtipos de pacientes con cáncer de mama	20
4.2 Test estadístico	24
4.3 Caracterización del perfil mutacional en los subtipos de PAM50	25
4.4 Estudio de supervivencia en la clasificación PAM50	8
4.5 Pathway analysis	32
5. Discusión	38
6. Conclusiones	45
7. Glosario	46
8. Bibliografía	47

Lista de figuras

Figura 1: Gráfico de Gantt. Descripción y planificación de las tareas a realizar durante el proyecto.

Figura 2: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación PAM50.

Figura 3.1: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación histológica.

Figura 3.2: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación histológica.

Figura 4: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación Claudin subtype.

Figura 5: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para los pacientes positivos y negativos en la expresión de los receptores de estrógenos ER, PR y HER2.

Figura 6. Histograma de las mutaciones que presenta el gen TP53 en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

Figura 7. Histograma de las mutaciones que presenta el gen PIK3CA en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

Figura 8. Histograma de las mutaciones que presenta el gen AHNK2 en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

Figura 9. Histograma de las mutaciones que presenta el gen GATA3 en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

Figura 10. Curvas de supervivencia para los pacientes clasificados según PAM50.

Figura 11a. Supervivencia de los pacientes Luminales A cuando presentan mutación (m) con respecto a cuando no la presentan(nm) para el gen TP53.

Figura 11b. Supervivencia en los pacientes Luminales B cuando presentan mutación (m) con respecto a cuando no la presentan (nm) en los genes TP53.

Figura 11c. Supervivencia en los pacientes pertenecientes al subtipo Normal cuando presentan mutación (m) con respecto a cuando no la presentan (nm) en los genes TP53.

Figura 12a. Supervivencia en los pacientes Luminales B cuando presentan mutación en el gen PIK3CA (m) con respecto a cuando no presentan mutación en este gen (nm).

Figura 12b. Supervivencia en los pacientes pertenecientes al subtipo Basal, cuando presentan mutación en el gen PIK3CA (m) con respecto a cuando no presentan mutación en este gen (nm).

Figura 13. Supervivencia en los pacientes dentro del subgrupo Basal para la mutación en el gen AHNAK2.

Figura 14a. Supervivencia en los pacientes clasificados dentro del subtipo Luminal A cuando presentan mutación en el gen GATA3 (m), con respecto a cuando no la presentan (m)

Figura 14b. Supervivencia en los pacientes clasificados dentro del subtipo Luminal B cuando presentan mutación en el gen GATA3 (m), con respecto a cuando no la presentan (m)

Figura 14c. Supervivencia en los pacientes clasificados dentro del subtipo Normal cuando presentan mutación en el gen GATA3 (m), con respecto a cuando no la presentan (m)

Figura 15. Pathway analysis del gen PIK3CA y TP53 mediante el uso de la base de datos KEGG.

Figura 16. Pathway analysis realizado para los genes TP53, PIK3CA y AHNAK2, respectivamente, mediante uso de String.

Figura 17. Pathway analysis realizado para el gen GATA3 mediante uso de String.

Lista de tablas

Tabla 1. Resumen del contenido de los datos contenidos en el estudio que se toma como referencia.

Tabla 2. Número de pacientes para cada subtipo de la clasificación PAM50.

Tabla 3. Número de pacientes para cada subtipo de la clasificación histológica.

Tabla 4. Número de pacientes para cada subtipo de la clasificación claudin-subtype.

Tabla 5. Número de pacientes pertenecientes a cada grupo formado por la expresión de los receptores hormonales incluidos en el estudio.

Tabla 6. Resultados obtenidos para la prueba chi-cuadrado en cada clasificación para los siete genes mayoritariamente mutados en cada una de ellas.

Tabla 7. Resultados de los test chi-cuadrado realizados para comparar los diferentes tipos de mutaciones en distintos genes para los subtipos de la clasificación PAM50.

Tabla 8. Genes que, cuando presentan mutación, influyen en la supervivencia de los pacientes pertenecientes a cada subtipo de cáncer de mama estudiado en la clasificación PAM50.

1 Introducción

1.1 Contexto y justificación del Trabajo

El cáncer es una de las principales causas de muerte en todo el mundo, en concreto, el cáncer de mama es la principal causa de muerte por cáncer en mujeres españolas [1], es decir, es el cáncer más letal en el género femenino en la actualidad, además, aunque en menor medida, también está presente en hombres, convirtiéndose de esta forma en una enfermedad que puede presentarse en la totalidad de la población y para la cual no existe un tratamiento que garantice una total supervivencia en todos los casos.

Se trata de una enfermedad genética multifactorial y altamente heterogénea en la que influyen factores genéticos, en algunos casos heredables, así como otros factores no genéticos (asociados a hábitos de vida, edad, sexo, exposición a agentes mutágenos, etc.) [2]. Se estima que, en el mundo, aproximadamente 1 de cada 8 mujeres padecerá cáncer de mama, de todos estos casos se estima que entre el 5-10% se deberá a factores genéticos y aproximadamente el 90-95% se deberá a factores ambientales, así como a hábitos de vida [3].

Todo ello, junto con el aumento en la incidencia de esta enfermedad entre la población [4], hace que sea palpable la necesidad de disponer de tratamientos efectivos capaces de aumentar la supervivencia para estos pacientes de forma que, tras un diagnóstico, pueda asignarse un tratamiento que garantice una efectividad.

La cantidad de pacientes con cáncer de mama, junto con la enorme necesidad de indagar en el descubrimiento de soluciones efectivas frente a ello, han permitido que la comunidad científica se vuelque en la investigación de esta variedad de cáncer y sus posibles tratamientos. Es por ello, que en la actualidad no resulta difícil encontrar gran cantidad de información variada dentro de este ámbito. No obstante, los tumores que caracterizan este cáncer son muy heterogéneos tanto entre pacientes (heterogeneidad intertumoral), como dentro de un mismo tumor (heterogeneidad intratumoral) [5], siendo éste, el principal obstáculo a la hora de establecer un sistema de tratamientos efectivos y personalizados para cada caso.

La evidente necesidad de descubrir nuevas vías de tratamiento que permitan acometer el cáncer de mama ha generado diferentes revelaciones con respecto a nuevos tratamientos y clasificaciones basadas en características que difieren entre tumores. Dichas clasificaciones se basan en distintas características tumorales, ya sean histológicas, temporales, moleculares, perfiles de expresión, presencia/ausencia de receptores hormonales u otros biomarcadores [6]. Esto ha facilitado que algunas tipologías de cáncer de mama ya dispongan de algunos tratamientos más o menos eficaces [7] pero, aun así, la respuesta que se genera en estos pacientes al ser tratados es

muy heterogénea entre ellos, aunque el cáncer que padezcan esté clasificado dentro de un mismo grupo, como pudiera ser el caso de los pacientes que se encuentran en los grupos HER+ y HER- o luminales y no luminales [8]. Todo ello reafirma la enorme necesidad de focalizar las nuevas vías de investigación en la exploración de nuevos marcadores que permitan establecer una clasificación que, al fin, consiga predecir la respuesta al tratamiento en los pacientes diagnosticados con un determinado tipo de cáncer de mama.

Son abundantes los artículos que profundizan en las diferentes mutaciones que aparecen en los tumores de pacientes con cáncer de mama, tanto a nivel somático como hereditario. Las mutaciones hereditarias son aquellas que se producen en el ADN de las células germinales y que los individuos portadores de las mismas pueden transmitir a su descendencia. Se ha visto que, en aquellos tumores desarrollados debido a factores hereditarios, juegan un papel importante, principalmente, las mutaciones desarrolladas en los genes BRCA1 y BRCA2, que participan en múltiples vías esenciales de reparación del ADN [9]. Cabe decir que aún se encuentra en vías de investigación el descubrimiento de otros genes que puedan ser partícipes de ello.

Las mutaciones somáticas, en cambio, son aquellas que, debido a diferentes factores de la vida del individuo, pueden producirse en múltiples genes de células no sexuales y que no son heredables pero que sí pueden producir la proliferación desmesurada de las células generando tumores, entre ellos, y siguiendo con el hilo del tema tratado, los mamarios.

Tal y como se ha comentado anteriormente, entre el 90-95% de los tumores desarrollados entre pacientes con cáncer de mama se deben a mutaciones de tipo somático en una gran variedad de genes que aún se encuentra en vías de investigación debido a que no se ha conseguido por el momento establecer un patrón mutacional que permita asociar cada tratamiento con cada tipo de mutación.

Todo ello genera indicios sobre el posible papel clave que puede tener el establecimiento del perfil mutacional en los tumores de estos pacientes para poder determinar un tipo de tratamiento o un pronóstico. Ello puede llegar a ser una vía de abordaje muy importante y aparentemente necesaria a la hora de establecer una clasificación o una caracterización de los diferentes subtipos que permita llegar al objetivo principal, que es el de conseguir tratar a los pacientes con tratamientos ajustados a cada caso y así alcanzar un aumento en la supervivencia, o incluso con una mirada optimista hacia el futuro, una supervivencia total.

1.2 Objetivos del Trabajo

El propósito general del presente Proyecto Final de Máster es el estudio del perfil mutacional somático y la caracterización a nivel mutacional en diferentes pacientes con cáncer de mama pertenecientes a diferentes subtipos establecidos hasta el momento. De esta forma, se pretende observar, partiendo de un estudio estadístico, el posible impacto que pueden tener las mutaciones más presentes en los tumores de los pacientes que padecen esta enfermedad sobre el pronóstico y la caracterización de los pacientes pertenecientes a cada subtipo. Así pues, partiendo del objetivo principal mencionado, se espera poder evaluar si existen diferencias significativas en cuanto a mutaciones entre los subtipos de cáncer que forman la clasificación estudiada, poder establecer si ello influye en la supervivencia de los pacientes y si a nivel global tiene un valor biológico robusto que pueda aportar más información a los datos que se disponen actualmente sobre el cáncer de mama.

Cada uno de los pasos que se seguirán, irán correspondidos con el anterior y será necesario ir avanzando en la exploración de los datos para confeccionar el siguiente paso en cada caso, el motivo de ello es, tal y como se menciona anteriormente, que el presente estudio es mayormente exploratorio.

Por tal de conseguir dicho propósito se parte de diferentes objetivos:

- Desarrollo de scripts para la obtención e integración de datos clínicos y moleculares a partir de una base de datos pública que contiene información de pacientes con cáncer de mama.
- Identificación y análisis estadístico de los patrones que siguen los perfiles mutacionales en cada uno de los grupos de pacientes.
- Para aquellos grupos de tumores en los que haya sido posible identificar un patrón concreto de mutaciones somáticas, evaluar el impacto de éstas en la supervivencia.
- Asociar la información obtenida hasta el momento con una base biológica basándose en bibliografía científica y bases de datos biológicas.

1.3 Enfoque y método seguido

El enfoque del presente trabajo se centra en la exploración de los datos del estudio METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) llevado a cabo por la Universidad de Cambridge [10]. Se pretende emplear los archivos proporcionados en esta base de datos y realizar un análisis para comprobar si los datos mutacionales en tumores de pacientes con cáncer de mama juegan un papel importante en la subclasificación de los tipos de este cáncer y la supervivencia de los pacientes que lo padecen. Para ello se va a contar con el uso del programa RStudio versión 3.6.3.

Los datos con los que se va a trabajar parten del estudio mencionado y se encuentran recogidos en ficheros con formato .xlsx, .csv y .txt. Este trabajo se focaliza en asociar los datos clínicos de los pacientes con los datos de mutaciones obtenidos a partir de la secuenciación de 173 genes.

El método seguido en este proyecto va a ser flexible y adaptable a los resultados que se vayan obteniendo en cada paso, pero focalizado a un único objetivo, que es el de integrar los datos clínicos y de mutaciones para obtener un patrón mutacional que permita obtener información sobre datos de supervivencia, clasificaciones y su valor biológico en los subtipos de cáncer de mama. Será de gran importancia la interpretación biológica propuesta en cada una de las fases seguidas en el punto anterior y que realmente será la que marcará el método del proyecto.

En primer lugar, se pretende obtener una representación gráfica de aquellos genes que se encuentran mutados en, al menos, el 10% de los pacientes para cada una de las clasificaciones tumorales establecidas en el estudio base.

Una vez obtenidas las representaciones, se evaluará mediante test estadísticos basados en pruebas chi-cuadrado si hay asociación entre las mutaciones y la diferenciación de los subgrupos establecidos en cada clasificación. Para ello, se seleccionarán los genes más representativos en cada una de las clasificaciones y se testarán.

Se estudiará el perfil mutacional presente en aquellos subtipos de cáncer de mama con los que se haya decidido seguir en base a los resultados obtenidos previamente y al valor bibliográfico y clínico que ocupen en el campo estudiado.

A continuación, se estudiará la supervivencia de los pacientes que presentan mutación y los que no, y que pertenezcan a aquellos subtipos seleccionados. Con ello, se relacionará el impacto en la supervivencia de las mutaciones observadas con el perfil de cada uno de los subtipos.

Una vez establecido todo lo anterior, será necesario realizar una valoración e interpretación a nivel biológico de lo obtenido basándose en

el análisis de rutas biológicas y estudios recientes y que sigan líneas de investigación similares.

1.4 Planificación del Trabajo

Para desarrollar el proyecto, primero se partirá de los ficheros descritos en el punto anterior. Se trata de ficheros en diferentes formatos que hay que explorar y transportar a tablas, con ello se pretende poder visualizar los datos de una manera más intuitiva con el objetivo de poder manipularlos y adaptarlos al análisis a realizar. Para ello, será necesario elaborar un script con RStudio y procesar los datos a partir del mismo, con ello se compondrá la base a partir de la cual sucederán el resto de fases del proyecto.

Con los datos en un formato manipulable, se analizarán las diferentes variables contenidas en el estudio de partida y se interpretarán según las indicaciones de los autores. Entre estas variables se encuentran las clasificaciones consideradas en el estudio base, las cuales son el epicentro del presente proyecto.

Con toda la información organizada y entendida, se representarán aquellos genes que aparezcan mutados en, al menos, el 10% de los pacientes estudiados de forma que permita visualizar si corroboran las clasificaciones o si permiten obtener nueva información al respecto.

A partir de los gráficos obtenidos, se realizará un estudio exploratorio que irá adaptándose a los resultados obtenidos en cada momento para, finalmente, realizar una interpretación biológica relacionando datos mutacionales, clínicos, otros datos moleculares y las conclusiones obtenidas en estudios que sigan líneas de investigación en común. Se pretende que con todos los datos en un conjunto se puedan establecer conclusiones que confluyan toda la información obtenida.

De forma paralela, se irán redactando diferentes informes o Pruebas de Evaluación Continua (PEC) en las que se describan los pasos llevados a cabo hasta el momento de la entrega de las mismas y que constituirán la arquitectura del presente proyecto.

Las tareas a realizar para conseguir el objetivo propuesto son:

- Planteamiento del proyecto: Búsqueda de bibliografía de estudios relacionados con el cáncer de mama, obtención de información sobre la problemática actual de esta enfermedad, tipos de clasificaciones descritas hasta el momento, mutaciones más comunes que presentan los pacientes y obtención de los ficheros de trabajo e interpretación de las variables que incluyen [10].
- Lectura de datos mediante uso de R: Interpretación de los datos proporcionados, acondicionamiento de los mismos en tablas, relación de datos clínicos y mutacionales, interpretación de las diferentes clasificaciones de cáncer de mama incluidas en el proyecto.
- Propuesta del proyecto a partir de toda la información obtenida hasta el momento.

- Representación con R: Establecimiento de las diferentes clasificaciones y representación de aquellos genes que aparecen mutados en, al menos, el 10% de los pacientes estudiados con el objetivo de obtener un patrón mutacional.
- Realización de test estadísticos basados en pruebas chi-cuadrado.
- Representación de curvas de supervivencia para aquellas clasificaciones más significativas.
- Identificación de patrones que relacionen las clasificaciones con los datos de mutaciones, supervivencia y tratamiento de la enfermedad, entre otras variables presentes en los datos proporcionados.
- Interpretación biológica de toda la información obtenida hasta el momento de forma que permita generar nuevas clasificaciones, nuevos patrones de mutación y/o comportamiento de esta enfermedad.
- Dilucidación de los resultados obtenidos y las conclusiones que permiten obtener.

Cada uno de las tareas llevadas a cabo se presentan a continuación en formato de diagrama de Gantt (figura 1).

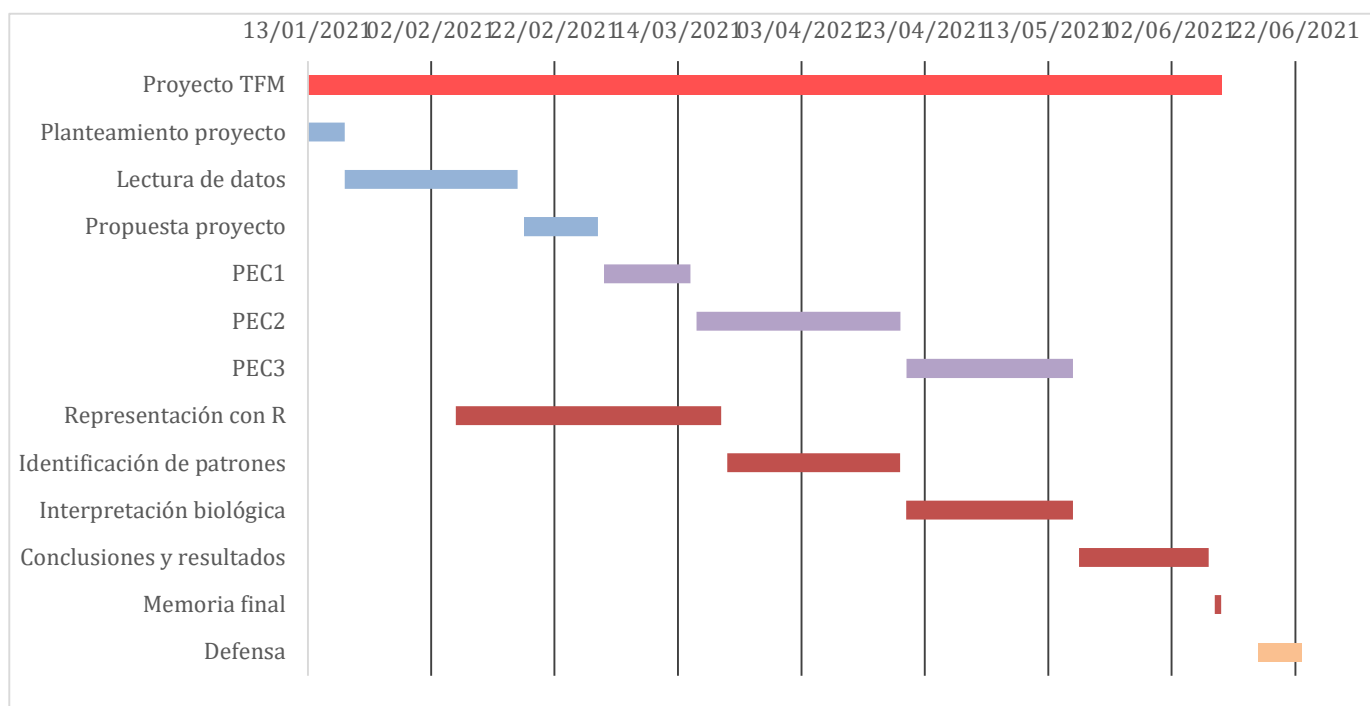


Figura 1: Gráfico de Gantt. Descripción y planificación de las tareas a realizar durante el proyecto.

1.5 Breve resumen de contribuciones y productos obtenidos

Documentos obtenidos:

- Propuesta del Proyecto Final de Máster (PEC0)
- Pruebas de Evaluación continua (PEC1, PEC2, PEC3).
- Memoria final.

Resultados obtenidos:

- Repositorio de GitHub (<https://github.com/nadiamolto/TFM.git>)
- Script en RMarkdown en el que se desarrolla la totalidad del proyecto

Adicionalmente, se espera obtener la caracterización clínica y molecular de subgrupos de tumores mamarios por tal de aportar más información a las líneas de investigación centradas en el cáncer de mama.

1.6 Breve descripción de los otros capítulos de la memoria

El proyecto contiene los siguientes puntos, tal y como se describe en el índice inicial:

- Introducción: Se expone, en el contexto de la actual problemática, la importancia que ocupa el cáncer de mama, así como los principales obstáculos que se encuentran las líneas de investigación centradas en el mismo. Se aporta, además, los aspectos más importantes que permiten introducir el ámbito del presente proyecto, los objetivos que lo marcan y el enfoque y método que va a permitir llevarlo a cabo.
- Estado del arte: Se expone la importancia que abarca el presente proyecto y con qué aportaciones contribuye en el estudio del cáncer de mama.
- Metodología: Se abarcan los métodos seguidos justificando su uso y lo que se pretende obtener.
- Resultados: En este apartado se muestran los resultados obtenidos y se relacionarán con la metodología empleada para su obtención.
- Discusión: Se debate si se ha conseguido llegar al objetivo marcado, se tratan los puntos fuertes y débiles del proyecto y se relacionan los resultados con bibliografía de forma que se consigan obtener razonamientos claros y concisos.
- Conclusiones obtenidas a partir de todo lo anterior.

2 Estado del arte

2.1 Antecedentes

A pesar de que, como se ha indicado previamente, el cáncer de mama es una de las enfermedades más conocidas y recurrentes en la actualidad, aún no se dispone de pronósticos y tratamientos adecuados para cada uno de los pacientes que la padecen.

Si bien, por su notoria recurrencia, es conocido que existe una necesidad social en el descubrimiento de tratamientos más efectivos adecuados a cada caso.

Se ha visto que la recurrencia de este tipo de cáncer es cada vez mayor, especialmente entre mujeres, con lo cual no es difícil encontrar cada vez más estudios que tratan de obtener nueva información con respecto a esta enfermedad que, en muchos casos, llega a ser mortal.

2.2. Clasificaciones de cáncer de mama

El cáncer es una enfermedad muy heterogénea, ya sea entre los diferentes subtipos establecidos o dentro de los mismos [11]. Por ello, la clasificación de los pacientes que padecen este tipo de cáncer ocupa una posición primordial a la hora de establecer nuevos conocimientos que permitan obtener tratamientos cada vez más eficaces.

Actualmente, hay establecidas distintas clasificaciones atendiendo a las características tumorales que se consideren. Las clasificaciones más ampliamente empleadas a nivel clínico (y las incluidas en el presente estudio) son:

PAM50

La clasificación PAM50 fue propuesta Por Perou et al. En el año 2000 [12]. Se trata de una clasificación que ha ido evolucionando y que es ampliamente usada en la actualidad, además de un foco de investigación en la búsqueda de nuevos tratamientos y alternativas en el cáncer de mama.

Inicialmente estaba compuesta por 4 subtipos (Basal, Normal, Luminales y Her2), cuya diferenciación se basaba en los niveles de expresión de cuatro marcadores (receptor de estrógeno (ER), receptor de progesterona (PR), receptor 2 del factor de crecimiento epidérmico humano (HER2) y Ki-67) a partir de muestras tomadas de tejidos tumorales. Para esta clasificación inicial, no se consideraron aspectos clínicos o de pronóstico [13]. En estudios posteriores se introdujeron los subtipos Luminales A y Luminales B dentro del subtipo de los Luminales.

La clasificación PAM50 estaría compuesta por los subtipos: Luminal A (ER+, PR+, HER2- y un nivel bajo de Ki67), Luminal B (ER+,PR+, HER2+ o HER2- y un nivel alto de Ki67), HER2 “enriched” (ER-, PR- y HER2+), Basal-like (ER-, PR-, HER2-) y Normal- Breast (se trata de un grupo bastante inespecífico) [14].

A posteriori, se han desarrollado plataformas que permiten determinar, mediante pruebas de expresión, estos subtipos a partir de la expresión de 50 genes con el uso de microarray, como es el caso de Prosigna™ [15].

Se ha demostrado que, en algunos casos, permite establecer si un tratamiento será más efectivo o menos, como es el caso de la hormonoterapia y la quimioterapia (o una combinación de ambas), así como determinar si un pronóstico será mayor o menormente favorable. El motivo es que esta clasificación se basa en la expresión de determinados receptores hormonales como son ER, PR o HER2 y, adicionalmente, tiene en cuenta factores como el Ki67, que es un antígeno nuclear expresado en células en proliferación [16]. Esto permite, a través de datos de expresión, obtener información a la hora de determinar un pronóstico.

HISTOLÓGICA

Aunque el establecimiento molecular y perfiles de expresión de los diferentes tipos tumorales han desvelado datos clínicamente relevantes que han mejorado nuestra comprensión con respecto al cáncer de mama, uno de los factores establecidos más empleados para el pronóstico de este tipo de cáncer es el estado histológico del tejido tumoral y su entorno [17]. Con ello, se hace visible que el establecimiento de las clasificaciones no es independiente y que muchas veces es importante considerar varios aspectos y subtipos en los pacientes.

La clasificación histológica, se basa en el estado a nivel histológico de los tumores mamarios. Los subtipos presentes en esta clasificación son: BENIGN (tumor benigno), IDC (Invasive Ductal Carcinoma, es el más común y, aproximadamente el 80% de los casos de cáncer), IDC-MUC (Invasive Ductal carcinoma Mucinous, es un tipo de carcinoma ductal que empieza en el conducto lácteo y se propaga hacia fuera. En este caso hacia las células se recubren de mucina), IDC-MED (Invasive Ductal carcinoma Medullary, este tipo también empieza en el conducto lácteo pero la masa celular es distinta a la anterior, el tumor es una masa suave similar a la médula), IDC-TUB (Invasive Ductal Carcinoma Tubular, al igual que los ductales, empieza en el conducto lácteo, en este caso son pequeños y formados por estructuras tubulares), DCIS (Ductal Carcinoma in situ, este tipo no es invasivo porque empieza en el conducto lácteo pero no se extiende al exterior, aun así puede desencadenar un cáncer invasivo si no se trata a tiempo), PHYL (Phylloides, suele ser no invasivo pero presenta un crecimiento rápido, es tipo de tumor poco común), ILC (Invasive Lobular Carcinoma, en este caso el tumor empieza en los “lobulillos” productores de leche y se extiende hacia los tejidos adyacentes), IDC+ILC (el tumor contiene tanto el tipo invasivo ductal, como el tipo invasivo lobular), OTHER INVASIVE (otros tipos de tumores invasivos no mencionados), INVASIVE TUMOUR (algunos tipos de cancer invasivo se clasifican como NST que significa non special type) [18,19].

CLAUDIN SUBTYPE

En cierta bibliografía científica, el subtipo Claudin-low se incluye dentro de la clasificación PAM50, no obstante, es otra de las clasificaciones que se toman clínicamente en consideración y que se incluyen en el presente estudio. Esta clasificación se asemeja mucho a la clasificación PAM50, los subtipos que la componen son: Luminal A, Luminal B, Her2, Claudin-low y Normal (muy similar al subtipo Basal -like de PAM50 ya que muchos pacientes coinciden en estos dos subtipos para una clasificación y para la otra) [20]. El subtipo Claudin-low, se caracteriza por presentar niveles de expresión bajos de genes relacionados con las uniones estrechas entre células y la adhesión intercelular [21].

2.3. Desarrollo del cáncer de mama

Los factores asociados al desarrollo de cáncer de mama pueden estar relacionados con el sexo, siendo mayormente frecuente en mujeres que en hombres [22] o la edad, ya que la incidencia suele estar entorno a mujeres mayores de 40 años con mayor frecuencia [23]. También se relaciona con el estilo de vida y la alimentación [24], así como con factores hereditarios [25].

A nivel genético, se ha visto como las mutaciones influyen de manera directa en el desarrollo de los diferentes tipos de cáncer existentes. Se han encasillado las mutaciones encontradas en pacientes con cáncer como conductoras (aquellas que están relacionadas con la oncogénesis) y pasajeras (aquellas presentes en los pacientes pero que no se encuentran relacionadas con el desarrollo de cáncer) [26]. En este punto, es lógico interpretar que aquellas mutaciones conductoras serán las que influenciarán en la supervivencia del paciente al estar relacionadas con el crecimiento tumoral y, por tanto, en el desarrollo de cáncer y su impacto biológico y clínico.

Siguiendo esta línea, se ha establecido que existen mutaciones conductoras heredables cuando se producen en determinados genes. En este caso se encontrarían las mutaciones que se dan en los genes BRCA1 y BRCA2, que están directamente relacionadas con el desarrollo cáncer de mama y ovario hereditarios [27].

Por otro lado, a asociación de los hábitos de vida, la obesidad, el sexo, la alimentación o la edad con el cáncer de mama podrían estar relacionados con la aparición de un mayor número de mutaciones somáticas al someter el organismo a condiciones de envejecimiento, estrés celular o acumulación de toxinas y moléculas oxidantes que, en algunos casos podrían conllevar a la aparición de mutaciones. Estas mutaciones podrían llegar a ser conductoras y producir el desarrollo de tumores, como sería el mamario [28, 29,30].

2.4. Tratamientos frente al cáncer de mama

Antiguamente, las bases para el establecimiento del tratamiento en pacientes con cáncer de mama, se basaba en datos del grado histopatológico y estadio TMN (tumor, nodo o metástasis y sus niveles) [31]. Actualmente, son diversas las opciones que pueden emplearse para tratar este tipo de cáncer, como vendrían siendo la quimioterapia la hormonoterapia, radioterapia, cirugía o alguna combinación de éstas [32]. Así pues, en los últimos años, por tal de establecer los diferentes tipos de tratamientos posibles, se analizan factores como la edad, el estadio histológico, así como la expresión de marcadores como serían los receptores hormonales o factores de crecimiento [33].

Es por ello que, actualmente, existen muchos más factores que pueden tomarse en consideración para determinar un tratamiento, atendiendo al subtipo de cáncer que presente el paciente en cuestión. Por ello, es de gran importancia centrar gran parte de los estudios llevados a cabo en este ámbito en explotar la información disponible sobre clasificaciones, subtipos, datos de expresión, moleculares y/o mutacionales, para poder disponer de todos estos conocimientos y, en conjunto, ser capaces establecer nuevos tratamientos o ajustar los existentes a cada paciente.

Por todo lo mencionado, ya no es apropiado considerar el cáncer de mama como una única enfermedad, sino más bien como un conjunto de factores que convergen en el desarrollo del tumor o tumores mamarios [34]. Es importante centrar los nuevos estudios en descubrir qué factores intervienen en el desarrollo del mismo y descubrir nuevas vías que aborden la caracterización molecular de cada paciente de forma individualizada.

Debido a que se ha demostrado que las mutaciones presentan una relación directa con el desarrollo de cáncer de mama, el estudio mutacional somático puede considerarse uno de los pilares que sustentan las nuevas vías de investigación en el cáncer de mama. Se considera necesario que, junto con los datos de los que se disponen hasta el momento, puedan aunarse para así, poder caracterizar o complementar la información de cada uno de los pacientes, por tal de conseguir obtener tratamientos con un porcentaje de efectividad elevado y reducir así, la alta tasa de mortalidad que este cáncer tiene asociada.

3 Metodología

3.1. Pacientes y datos clínicos

Los datos a partir de los cuales se desarrolla el presente estudio parten de la base de datos METABRIC (consorcio internacional de cáncer de mama), se trata de una base de datos que alberga múltiple información en cuanto a datos de expresión, características clínicas o perfil de mutaciones procedentes de datos de tumores de mama pertenecientes a pacientes que participan en el proyecto.

El presente trabajo toma los datos de esta base de datos [10]. Los datos con los que se trabaja se obtienen a partir de diferentes ficheros publicados por METABRIC que contienen datos clínicos y mutacionales de un total de 1904 pacientes con cáncer de mama. Mediante R se realiza una integración de los datos en un mismo formato por tal de facilitar el posterior análisis exploratorio y estadístico. Los pacientes que aparecen en dicha tabla (disponible en el script publicado en GitHub), se encuentran identificados mediante un código de dos letras y 4 dígitos (MB-XXXX). En el fichero creado, dicho código dispone de información asociada, tal como: edad de diagnóstico, tiempo transcurrido hasta la muerte, tipo de subgrupo de cáncer de mama al que pertenece o si dispone de mutación o no en cada uno de los genes analizados (174 en total), así como el tipo de mutación (ver script publicado en GitHub).

A partir de este fichero, se realiza, a posteriori, un análisis exploratorio para relacionar los datos que se consideren de interés por tal de conseguir el objetivo previsto e indicado anteriormente.

A continuación, se presenta una tabla resumen de algunas de las variables que aparecen en la tabla a partir de la cual se inicia el proyecto.

Tabla 1. Resumen del contenido de los datos incluidos en el estudio que se toma como referencia.

Número de pacientes	Edad media de diagnóstico	Desviación estándar de la edad de diagnóstico	Número de genes estudiados	Número de variables estudiadas
1980	61,80	12,98	174	37

Uno de los datos que parece ser remarcable en la tabla anterior (tabla 1), es la edad media del diagnóstico de la enfermedad, así como su desviación estándar, ya que aparentemente, se sitúa en personas comprendidas entre los 50 y los 70 años aproximadamente.

Adicionalmente, con respecto a los datos facilitados por dicho estudio, se ha podido comprobar que, entre los genes incluidos, hay algunos cuya relación con el cáncer de mama se ha relacionado en numerosa bibliografía científica y los cuales se consideran importantes a la hora de llevar a cabo el presente proyecto (en los siguientes capítulos se dará mención a dichos genes).

Para el desarrollo de los siguientes puntos se hace uso del programa RStudio en su versión 3.6.3.

El script donde se desarrolla el trabajo realizado con R puede encontrarse en el repositorio de GitHub: <https://github.com/nadiamolto/TFM.git>

3.2. Clasificación de los pacientes

Todos los pacientes analizados han presentado o presentan tumores mamarios, cada uno de ellos dispone de un perfil clínico distinto con diferentes datos tomados desde el diagnóstico hasta su fallecimiento o recuperación. Puesto que cada clasificación considera unas características biológicas distintas, un mismo paciente puede formar parte de diferentes subgrupos.

En concreto, las principales clasificaciones de cáncer de mama que se toman en consideración para el desarrollo del presente proyecto son:

- **PAM50.** A continuación en la Tabla 2 se muestra el número de pacientes que se incluyen en el estudio para cada uno de los subtipos comprendidos para esta clasificación.

Tabla 2. Número de pacientes para cada subtipo de la clasificación PAM50.

Clasificación PAM50				
Luminal A	Luminal B	Her2	Basal-like	Normal-breast
696	474	236	308	184

- **Histológica.** A continuación, en la tabla 3, se muestra el número de pacientes que se incluyen en el estudio para cada uno de los subtipos que conforman la clasificación histológica.

Tabla 3. Número de pacientes para cada subtipo de la clasificación histológica.

Clasificación Histológica				
Benign	DCIS	IDC	IDC-MED	IDC-MUC
1	2	1503	31	42
IDC-TUB	IDC+ILC	ILC	INVASIVE TUMOUR	OTHER INVASIVE
67	87	141	9	19
PHYL				
0				

- **Claudin subtype.** A continuación, se muestra la tabla 4, en la que se incluyen el número de pacientes para cada subtipo de la clasificación Claudin subtype incluida en el estudio.

Tabla 4. Número de pacientes para cada subtipo de la clasificación Claudin subtype.

Clasificación Claudin subtype					
Luminal A	Luminal B	Her2	Basal-like	Normal-breast	Claudin-low
679	461	220	199	140	199

- Adicionalmente, en el estudio base se incluye la clasificación de los pacientes atendiendo a la expresión de los receptores hormonales (como son ER, PR y HER2), entre otros parámetros estudiados. Se muestra a continuación en la tabla 5 el número de pacientes que se incluyen dentro de estos parámetros.

Tabla 5. Número de pacientes pertenecientes a cada grupo formado por la expresión de los receptores hormonales incluidos en el estudio.

Receptores hormonales					
ER+	ER-	PR+	PR-	HER2+	HER2-
1459	445	1009	895	236	1668

3.3. Análisis de mutaciones

Una vez consideradas las clasificaciones a estudiar y que se mencionan en el punto anterior, se realiza una representación gráfica de tipo histograma en la que se muestran aquellos genes que aparecen mutados en, al menos, el 10% de los pacientes para cada uno de los subgrupos que componen las clasificaciones.

Se parte de la premisa de que los genes que más mutaciones presenten entre los pacientes serán los que muestren una variación remarcable entre los diferentes subtipos, y por tanto intervendrán directamente en la diferenciación de éstos. Estas mutaciones serán las que se espera que tengan una relación de causalidad con respecto al cáncer de mama y por tanto deberán tener influencia sobre los parámetros clínicos tales como la supervivencia y la clasificación de los pacientes.

Del total de los 174 genes incluidos en el estudio METABRIC del cual se parte en el presente proyecto, y partiendo de los resultados obtenidos en el histograma realizado para cada una de las clasificaciones, se obtienen los diferentes genes mayormente mutados entre los pacientes que conforman cada clasificación. De estos genes, se selecciona un total de 7 genes para pasar al siguiente paso. Se decide cribar este número de genes, debido a que se realiza

un análisis estadístico en el cual la media total de genes que aparentemente están diferencialmente mutados se encuentra en torno a 7.

3.4. Análisis estadístico

A continuación, se realiza una prueba estadística basada en la prueba chi-cuadrado. Con ello, se pretende valorar si los genes seleccionados en el paso anterior presentan diferencias significativas a nivel mutacional que varíen entre los subtipos de cada clasificación. La prueba realizada permitirá confirmar la hipótesis de partida en la cual existe una asociación entre los genes que se han representado como mayormente mutados y los subgrupos de las clasificaciones estudiadas. Se aceptará así un nivel de confianza del 5%, de modo que si en alguna de las clasificaciones y tras realizar el test se obtiene un p-valor superior a 0.05 se considerará que los genes estudiados en ese caso no tienen una relación asociativa con respecto a los subtipos que componen la clasificación en cuestión.

En caso de que el test estadístico muestre valores significativos en alguna de las clasificaciones, se pasará al siguiente paso.

3.5. Caracterización del perfil mutacional en los subtipos de PAM50

Posteriormente, se procede a realizar un estudio del perfil mutacional de los pacientes pertenecientes a cada uno de los subtipos que componen la clasificación PAM50 y de los genes que en el punto anterior surgieron como significativos. En este caso, se realiza un recuento de cada tipo de mutación presente en los genes analizados, una representación en forma de histograma. Finalmente, se realizan test chi-cuadrado para cada gen, para evaluar si la distribución de los diferentes tipos de mutaciones presenta diferencias significativas entre subtipos de cáncer de mama.

3.6. Análisis de supervivencia

El enfoque del presente proyecto desde el inicio es el de explorar las clasificaciones del cáncer de mama y el papel que pueden llegar a ejercer las mutaciones en determinados genes sobre las mismas. Un parámetro fundamental para poder obtener una aproximación al papel que pueden ejercer ciertas mutaciones es el de la supervivencia, se trata de un análisis ampliamente extendido entre los estudios del cáncer. Por ello, se estima apropiado emplearlo para observar si realmente estas mutaciones influyen en la supervivencia de los pacientes con respecto a aquellos que no presentan mutaciones.

Desde este punto se siguen los análisis empleando la clasificación PAM50 por razones que se tratarán en capítulos posteriores.

Para el desarrollo de este punto, se hará uso del estimador Kaplan- Meier, se trata de un método no paramétrico altamente empleado en los estudios de supervivencia. Este método permite obtener una representación gráfica en la cual se asocia la probabilidad de supervivencia de los pacientes en el tiempo en base a los datos facilitados en el estudio METABRIC, los cuales incluyen el tiempo transcurrido desde la detección del cáncer hasta el fallecimiento de los pacientes o su retirada del estudio.

3.7. Pathway analysis

Obtenidos los datos sobre los genes que, a nivel estadístico, parecen tener influencia sobre la supervivencia de los pacientes, es interesante ver en qué rutas moleculares están involucrados y si estas rutas representan algún eslabón en el desarrollo de cáncer de mama.

Para ello, se hace uso de la KEGG PATHWAY DATABASE [35] y String [36] que permiten estudiar cada gen viendo los procesos biológicos en los cuales se encuentran involucrados, así como su interacción con otros genes y proteínas que pueden estar relacionados con los mismos procesos o similares.

De forma adicional, se realizó un análisis en cuanto a ontologías genéticas mediante el uso de la base de datos Gene Ontology [37] de forma que reforzara todos los datos obtenidos a partir del pathway analysis.

3.8. Interpretación biológica

Es necesario remarcar que el presente estudio tiene un carácter exploratorio, en el cual se avanza a medida y según los resultados que se van obteniendo. Se ha considerado que éste sería el método óptimo para llegar al objetivo del proyecto, de forma que pueda ir adaptándose cada uno de los pasos seguidos acorde a lo obtenido en los puntos anteriores. Los métodos empleados parten de la necesidad de explorar de una forma adaptable los datos de los que se dispone y valorar de forma detallada los resultados obtenidos.

Como se observará en el siguiente punto, la parte gráfica, visual y representativa de los datos juega un papel crucial a la hora de obtener las conclusiones.

Por todo ello y por tratarse de un estudio exploratorio, será de suma importancia asignarle una interpretación biológica a todo lo obtenido en el mismo. Así pues, con toda la información recogida hasta el momento, será necesario establecer unas conclusiones sólidas basadas en una interpretación biológica de los resultados obtenidos. Los genes analizados podrán tener una importancia destacable en cuanto a la clasificación existente del cáncer de mama y la supervivencia de los pacientes que lo padecen, tal y como se espera.

Para ello, será necesario hacer uso de bases de datos científicas con estudios que sigan líneas de investigación similares para así obtener un razonamiento de lo obtenido y aportar la nueva información obtenida en el transcurso del presente proyecto.

4 Resultados

4.1. Mutaciones en los subtipos de pacientes con cáncer de mama

Tal y como se menciona en el apartado anterior, se representaron gráficamente aquellos genes que presentaban mutaciones en, al menos, el 10% de los pacientes para cada una de las clasificaciones, obteniendo de esta forma diferentes gráficos de barras normalizados en los cuales se observa de una forma más intuitiva aquellos genes que se hallan mayor y menormente mutados para cada uno de los subgrupos analizados en el presente estudio.

Como se aprecia en la Figura 2, en la clasificación PAM50 los gráficos de barras dieron como resultado un conjunto de diferentes genes, entre los cuales existen algunos que coinciden entre subgrupos, otros que también lo hacen, pero en menor medida y otros que están mutados mayormente en un subgrupo con respecto al resto de subgrupos para esta misma clasificación.

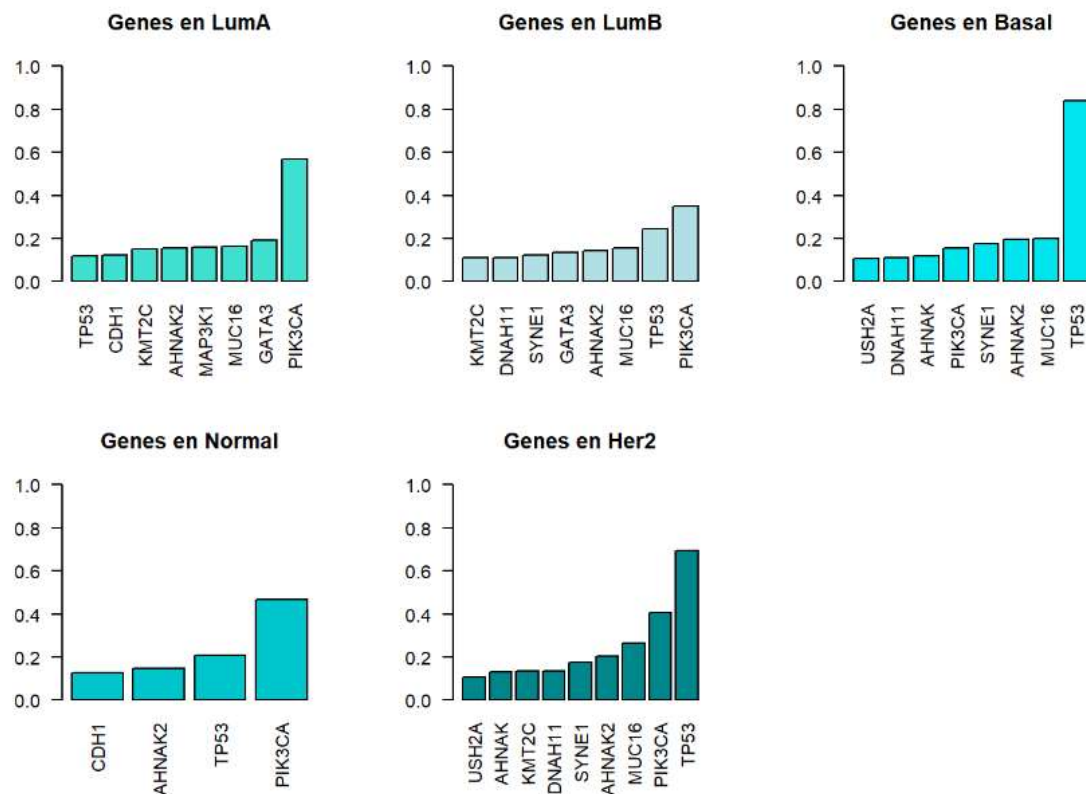


Figura 2: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación PAM50.

Partiendo del histograma mostrado, en la clasificación PAM50 se seleccionaron los siguientes genes como los más representativos de esta clasificación: *TP53*, *PIK3CA*, *SYNE1*, *GATA3*, *MUC16*, *AHNK2*, *CDH1*.

Con estos genes se procedió a realizar el análisis estadístico que se mostrará a continuación, considerando que podrían jugar un papel importante dentro de la clasificación PAM50 a nivel mutacional.

Para la clasificación histológica, se siguió el mismo proceso, de forma que se obtuvieron, igualmente, un conjunto de gráficos de barras normalizados con aquellos genes mutados en el 10% de los pacientes estudiados (Figura 3.1 y 3.2).

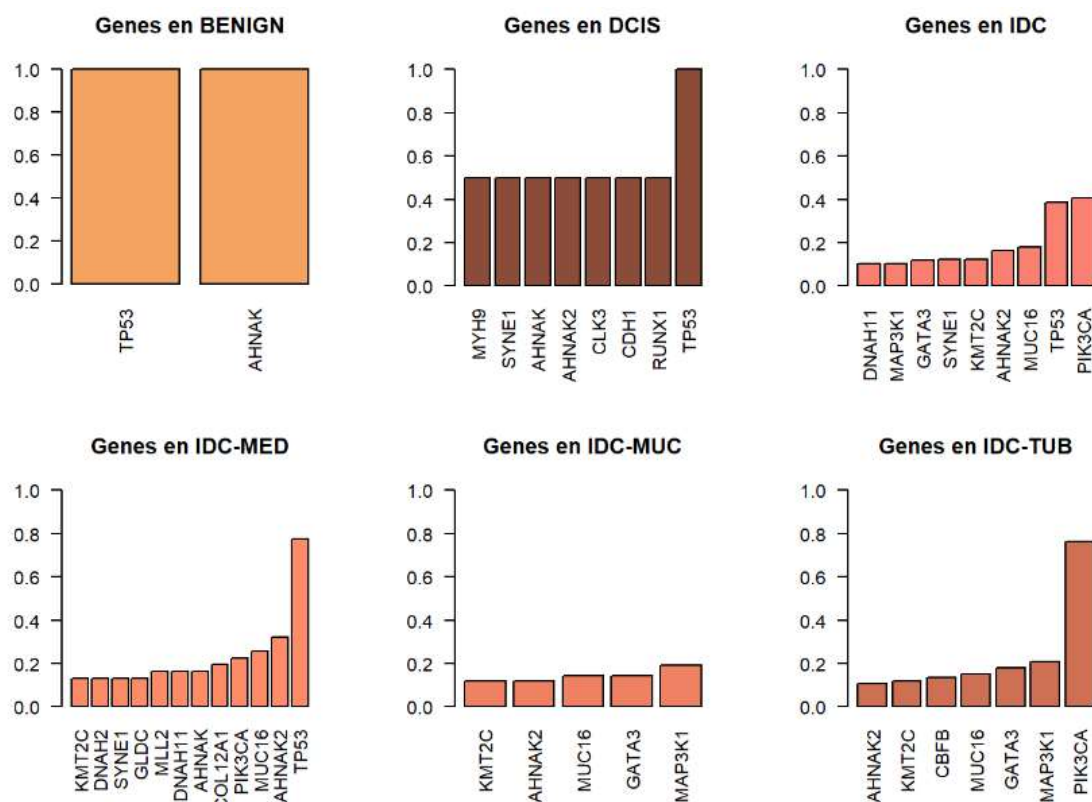


Figura 3.1: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación histológica.

Para esta clasificación histológica se tomaron como genes más representativos: *TP53*, *PIK3CA*, *MUC16*, *GATA3*, *AHNAK2*, *KMT2C*, *AHNAK*.

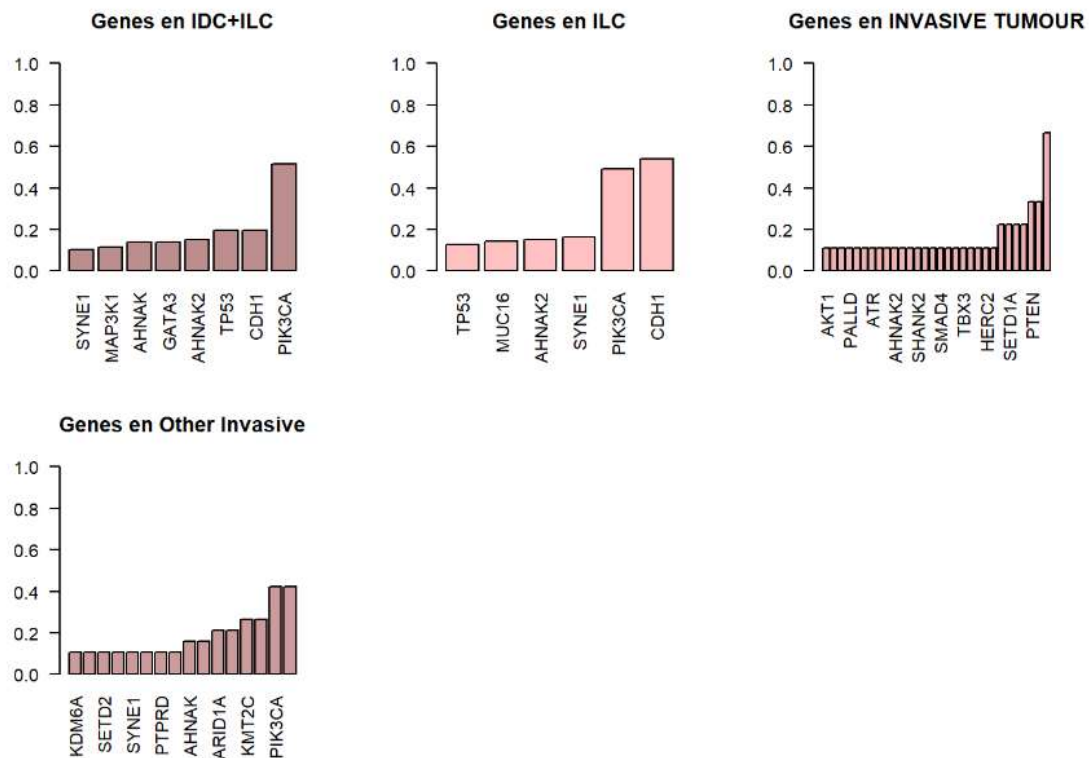


Figura 3.2: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación histológica.

Se siguió el mismo procedimiento que en las clasificaciones anteriores para la clasificación Claudin subtype, obteniendo los gráficos de barras normalizados con los genes que aparecen mayormente mutados entre los subgrupos que la conforman (Figura 4).

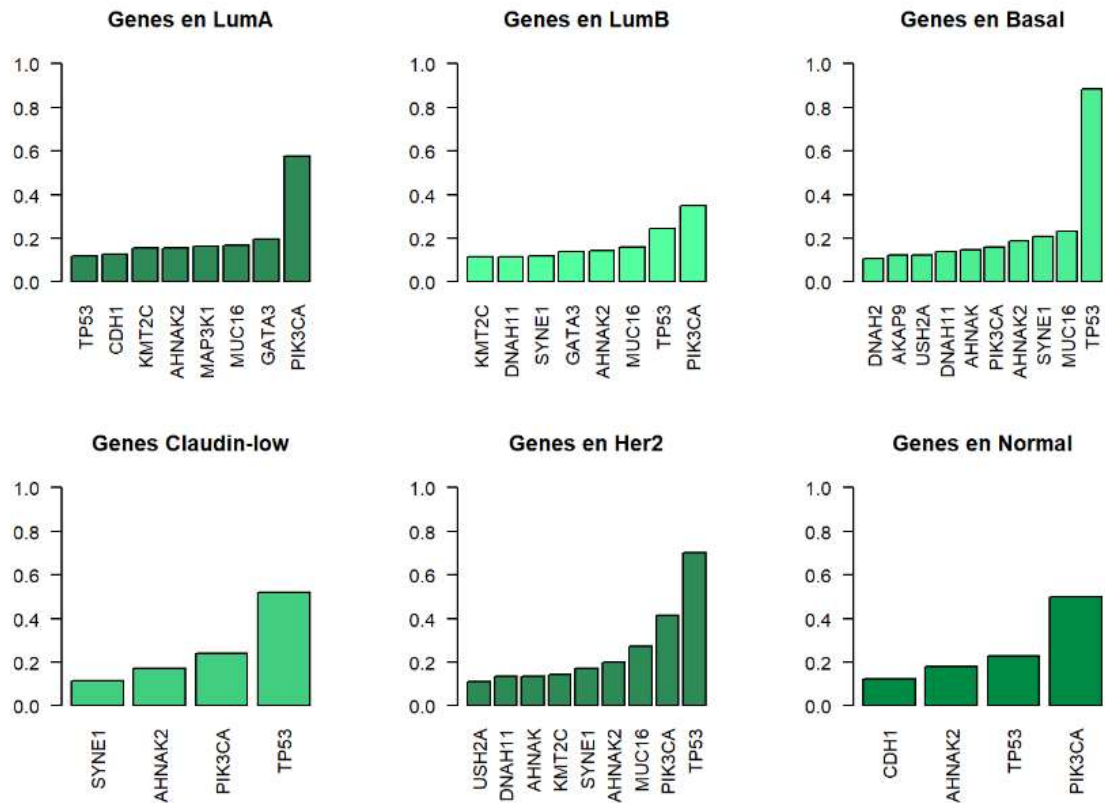


Figura 4: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para la clasificación Claudin subtype.

Para esta clasificación, se tomaron como genes más representativos: *PIK3CA*, *AHNAK2*, *SYNE1*, *TP53*, *MUC16*, *GATA3*, *DNAH11*.

Finalmente, siguiendo la misma línea que en los puntos anteriores, se realizó el histograma normalizado para los genes mayormente mutados en aquellos grupos de pacientes clasificados como positivos y negativos para los receptores de estrógenos ER, PR y HER2, tal y como se muestra a continuación en la figura5.

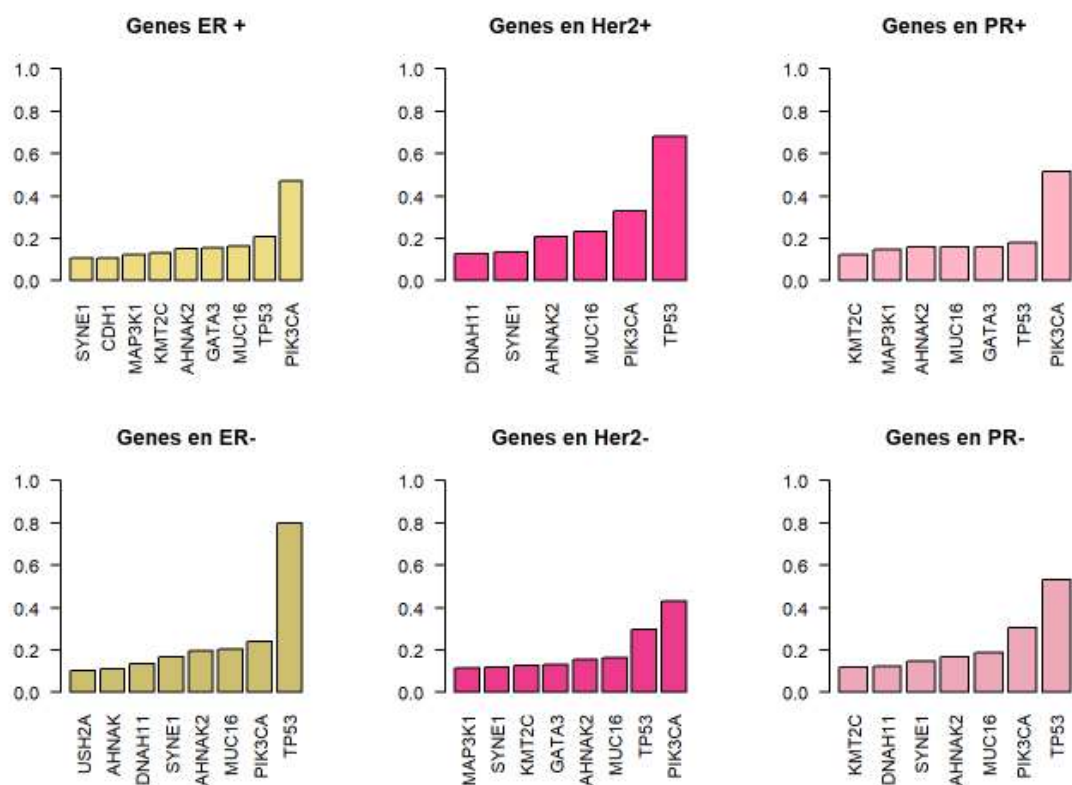


Figura 5: Gráficos de barras normalizado donde se representan las mutaciones presentes en, al menos, el 10% de los pacientes para los pacientes positivos y negativos en la expresión de los receptores de estrógenos ER, PR y HER2.

Finalmente, para estas tres clasificaciones se tomaron como genes más representativos: *TP53*, *PIK3CA*, *GATA3*, *MUC16*, *AHNAK2*, *SYNE 1* Y *KMT2C*.

4.2. Test estadístico

Tras obtener los genes que aparecen mayormente representados y, por tanto, mutados en cada uno de los subgrupos, se realizó un test Chi-cuadrado para cada clasificación, incluyendo en cada test todos los subgrupos de cada una de las clasificaciones, así como los genes más representativos en términos mutacionales, obteniéndose en todos los casos p-valores muy bajos y por tanto significativos para el nivel de confianza indicado previamente. A continuación, en la tabla 5 se muestra un resumen de los genes estudiados para cada clasificación, así como los resultados obtenidos en el test chi cuadrado.

Tabla 6. Resultados obtenidos para la prueba chi-cuadrado en cada clasificación para los siete genes mayoritariamente mutados en cada una de ellas.

PAM 50							
Genes estudiados	TP53	PIK3CA	GATA3	MUC16	AHNAK2	CDH1	SYNE1
P-valor obtenido	< 2,2e-16						
Histológica							
Genes estudiados	TP53	PIK3CA	GATA3	MUC16	AHNAK2	AHNAK	KMT2C
P-valor obtenido	6,38e-12						
Claudin subtype							
Genes estudiados	TP53	PIK3CA	GATA3	MUC16	AHNAK2	SYNE1	DNAH11
P-valor obtenido	<2,2e-16						
HER2+ y HER2-							
Genes estudiados	TP53	PIK3CA	GATA3	MUC16	AHNAK2	SYNE1	KMT2C
P-valor obtenido	< 2,2e-16						
ER+ y ER-							
Genes estudiados	TP53	PIK3CA	GATA3	MUC16	AHNAK2	SYNE1	KMT2C
P-valor obtenido	< 2,2e-16						
PR+ y PR-							
Genes estudiados	TP53	PIK3CA	GATA3	MUC16	AHNAK2	SYNE1	KMT2C
P-valor obtenido	< 2,2e-16						

4.3. Caracterización del perfil mutacional en los subtipos de PAM50

A partir de este punto, los estudios se centran en la clasificación PAM50 por motivos que se concretarán posteriormente.

De esta forma, tras una visualización global de los gráficos mostrados previamente, se realizó un estudio del perfil mutacional de aquellos genes pertenecientes a la clasificación PAM50 que parecen mostrar un patrón mutacional que se adecúa a la subclasificación de los tipos de cáncer de mama para corroborar lo observado.

En las figuras 6, 7 8 y 9 se observan los gráficos obtenidos en los que aparecen diferentes mutaciones, como serían:

Frameshit_indel: Inserción o delección en el ADN de un número de nucleótidos que no es múltiple de 3, por lo que alteran los codones que suceden el lugar de la inserción o delección.

Inframe_indel: Inserción o delección en el ADN de un número de nucleótidos múltiple de 3.

Missense_SNV: "Single nucleotid variation", esta mutación produciría un cambio en el aminoácido producido tras la traducción del nucleótido mutado.

Nonsense_SNV: En este caso, la mutación produciría un codón de SOP en la proteína resultante.

Splice_mut: Esta mutación produciría un cambio en los nucleótidos ubicados en una posición donde, al traducir la proteína, se da lugar el splicing.

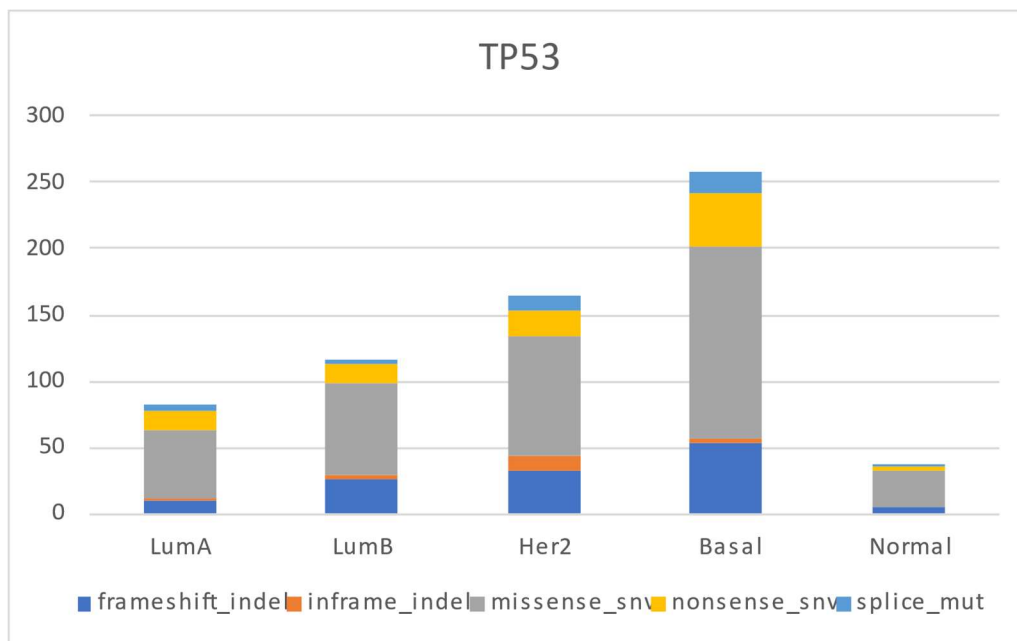


Figura 6. Histograma de las mutaciones que presenta el gen TP53 en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

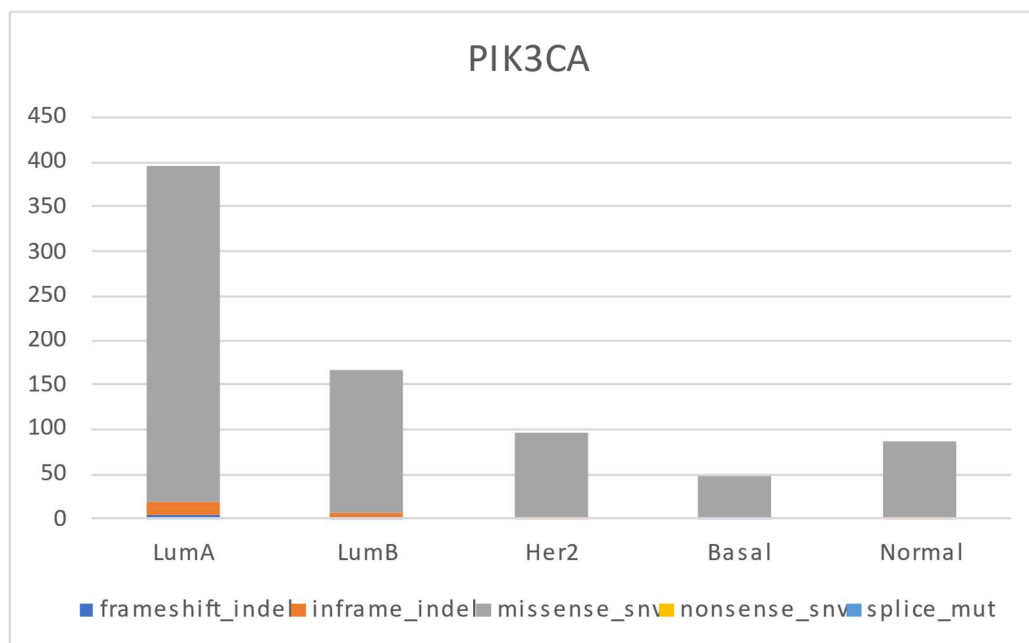


Figura 7. Histograma de las mutaciones que presenta el gen PIK3CA en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

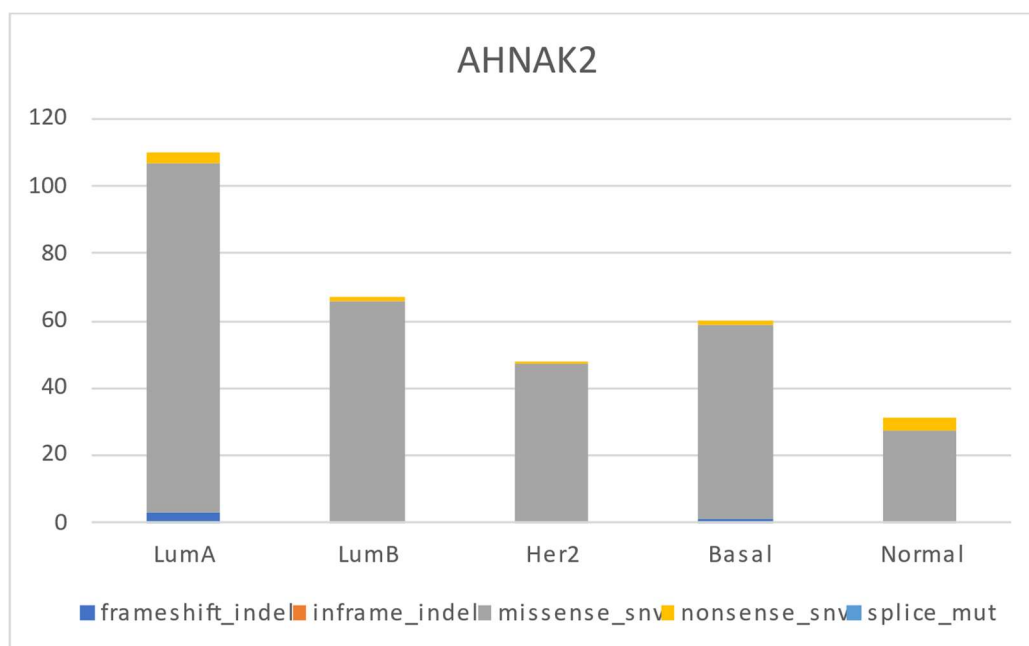


Figura 8. Histograma de las mutaciones que presenta el gen AHNAK2 en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

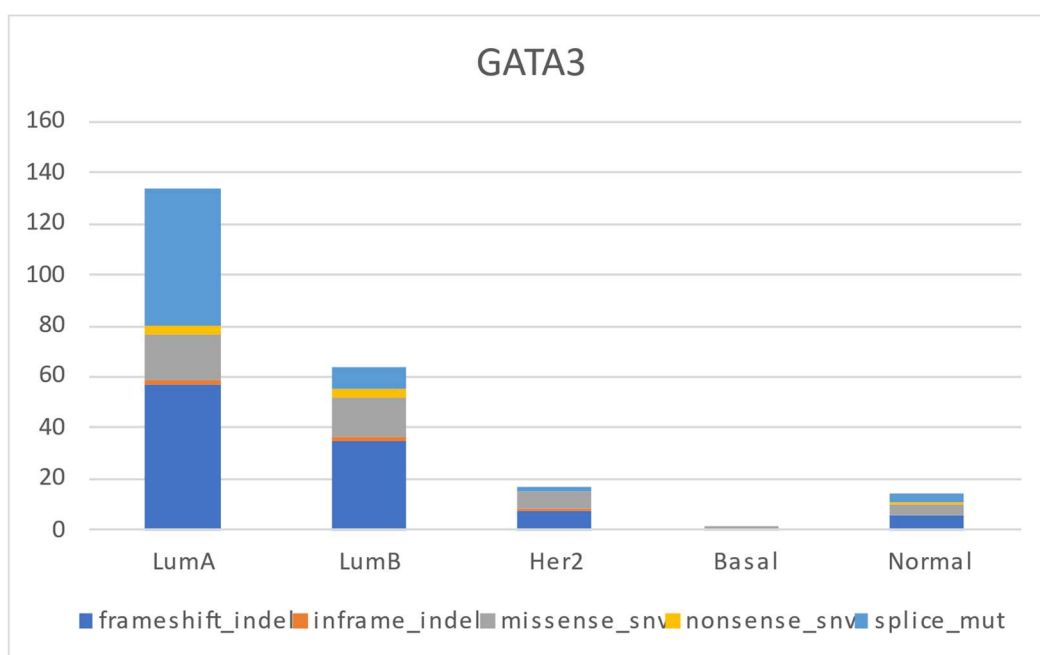


Figura 9. Histograma de las mutaciones que presenta el gen GATA3 en los pacientes que forman parte de cada uno de los subtipos pertenecientes a la clasificación PAM50.

A continuación, y siguiendo con la línea del presente apartado, se muestra una tabla resumen (tabla 7) donde pueden observarse los resultados del test chi-cuadrado realizado para establecer si existe significancia en las diferencias observadas en los gráficos anteriores.

Tabla 7. Resultados de los test chi-cuadrado realizados para comparar los diferentes tipos de mutaciones en distintos genes para los subtipos de la clasificación PAM50.

	TP53	PIK3CA	GATA3	AHNAK2
p-valor	0,17	0,47	0,02	0,07

4.4. Estudio de supervivencia en la clasificación PAM50

Por tal de ampliar la información obtenida hasta el momento y tomando en consideración el procedimiento habitual en la bibliografía sobre estudios en el cáncer de mama que toman una dinámica similar, se procedió con el estudio de la supervivencia de los pacientes para diferentes subtipos.

Inicialmente y a nivel exploratorio, se representó la supervivencia de los diferentes subgrupos que componen la clasificación PAM50, tal y como se puede observar en la figura 10.

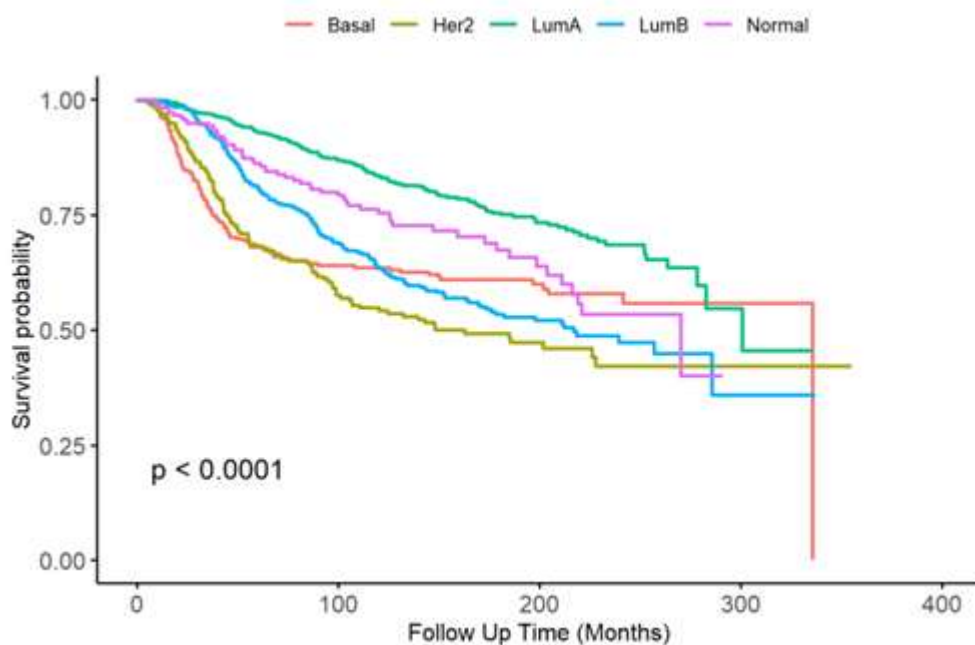


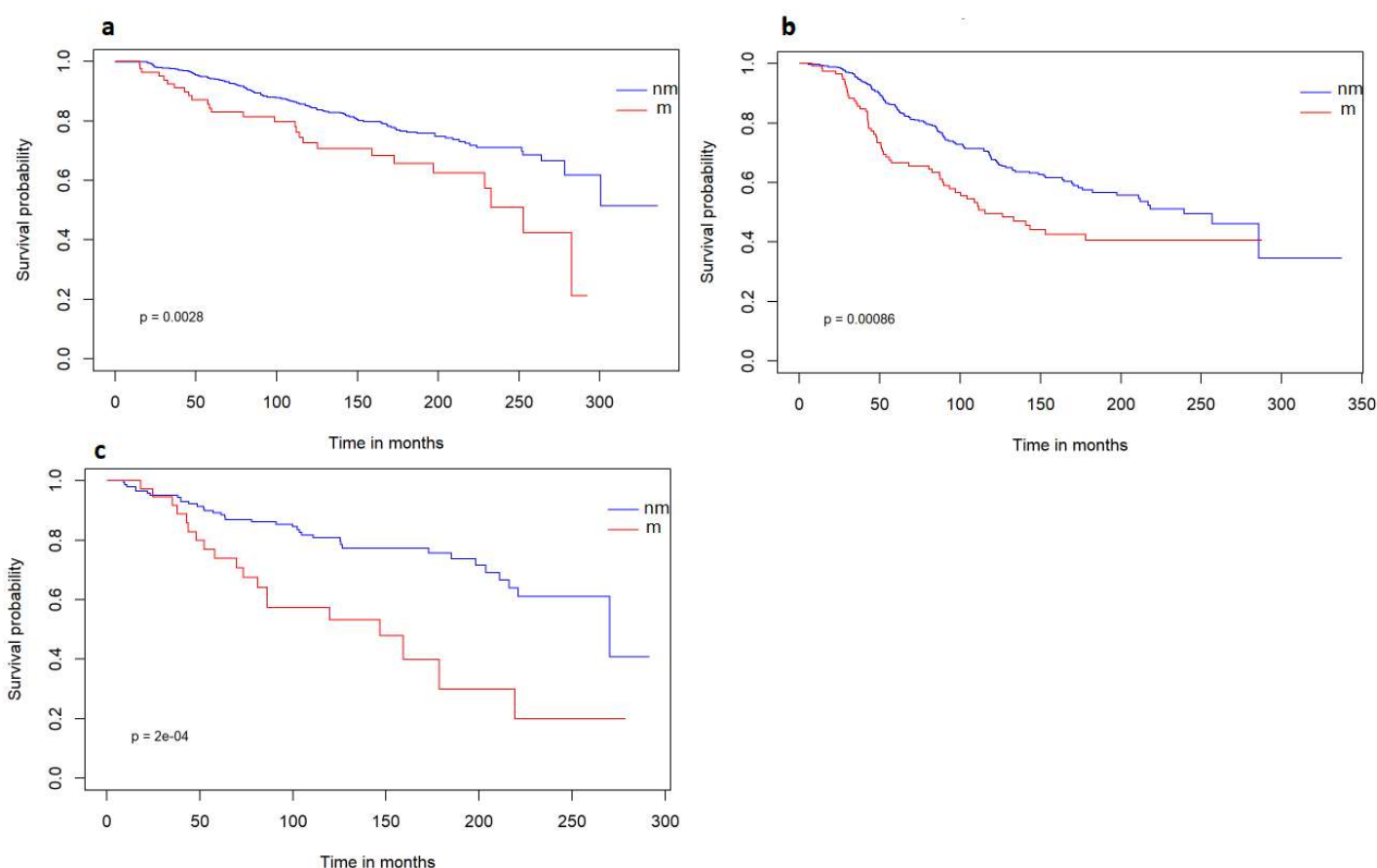
Figura 10. Curvas de supervivencia para los pacientes clasificados según PAM50.

Como se puede observar en la figura 10, hay una diferencia significativa en cuanto a la supervivencia de los pacientes que pertenecen a cada uno de los subtipos de la clasificación. Esto tiene sentido puesto que esta clasificación se basa en la expresión diferencial de distintos genes que es diferente entre los subtipos y que además presentan diferentes niveles de expresión para diferentes receptores hormonales.

Seguidamente, se realizó una representación gráfica de la supervivencia de los pacientes pertenecientes a cada uno de los subtipos que conforman la clasificación PAM50, individualizando por cada uno de los genes que salieron significativos en las representaciones gráficas llevadas a cabo previamente.

Se muestran a continuación las representaciones gráficas de la supervivencia de los pacientes para aquellos genes en los cuales presentan una diferencia estadísticamente significativa están mutados con respecto a cuando no lo están. El resto de las representaciones pueden ser consultadas a través del script contenido en el enlace de GitHub indicado en puntos anteriores.

A continuación, en las figuras 11a, 11b y 11c se muestra, de igual forma, las curvas de supervivencia para los pacientes que presentan diferencias significativas en cuanto a supervivencia cuando el gen TP53 contiene algún tipo de mutación y cuando no.



TP53

Figura 11a. Supervivencia en los pacientes Luminales A cuando presentan mutación (m) con respecto a cuando no la presentan(nm) para el gen TP53.

Figura 11b. Supervivencia en los pacientes Luminales B cuando presentan mutación (m) con respecto a cuando no la presentan (nm) en los genes TP53.

Figura 11c. Supervivencia en los pacientes pertenecientes al subtipo Normal cuando presentan mutación (m) con respecto a cuando no la presentan (nm) en los genes TP53.

En la figura 12 puede observarse, de igual forma, las curvas de supervivencia para los pacientes pertenecientes a los subtipos Luminal B y Basal cuando

presentan el gen PIK3CA mutado con respecto a cuando no lo presentan. Estos dos subtipos han sido los únicos del total de los subtipos estudiados que han mostrado diferencias significativas en la supervivencia para este gen.

PIK3CA

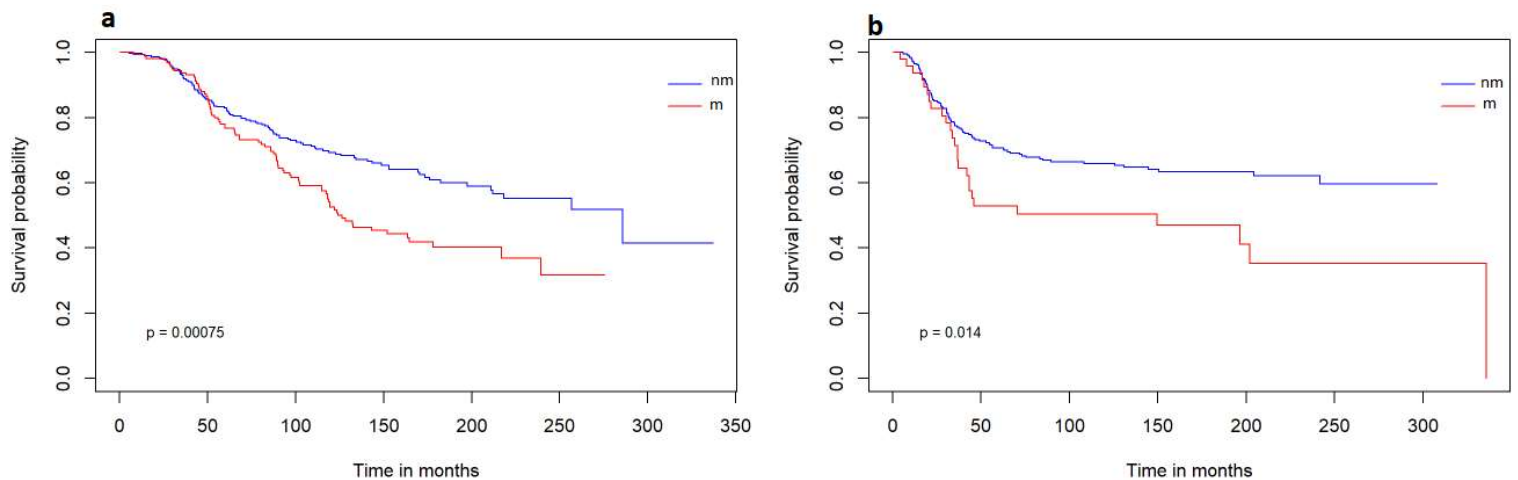


Figura 12a. Supervivencia en los pacientes Luminales B cuando presentan mutación en el gen PIK3CA (m) con respecto a cuando no presentan mutación en este gen (nm).

Figura 12b. Supervivencia en los pacientes pertenecientes al subtipo Basal, cuando presentan mutación en el gen PIK3CA (m) con respecto a cuando no presentan mutación en este gen (nm).

Seguidamente, en la figura 13, pueden observarse las curvas de supervivencia para los pacientes pertenecientes al subtipo Her2 cuando presentan mutación en el gen AHNAK2 con respecto a cuando no la presentan

AHNAK2

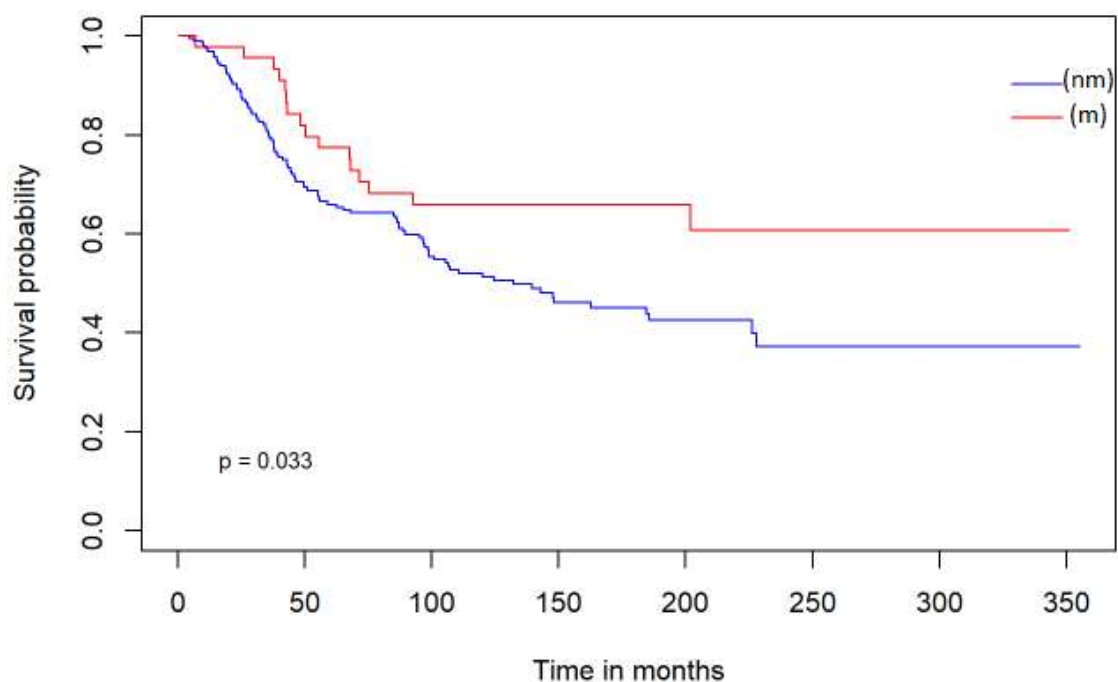


Figura 13. Supervivencia en los pacientes dentro del subgrupo Basal para la mutación en el gen AHNAK2.

En la figura 14 se observa, posteriormente, las curvas de supervivencia para los pacientes que presentan mutación en el gen GATA3 y cuya supervivencia ha sido estadísticamente significativa comparando la presencia y ausencia de mutación en dicho gen.

GATA3

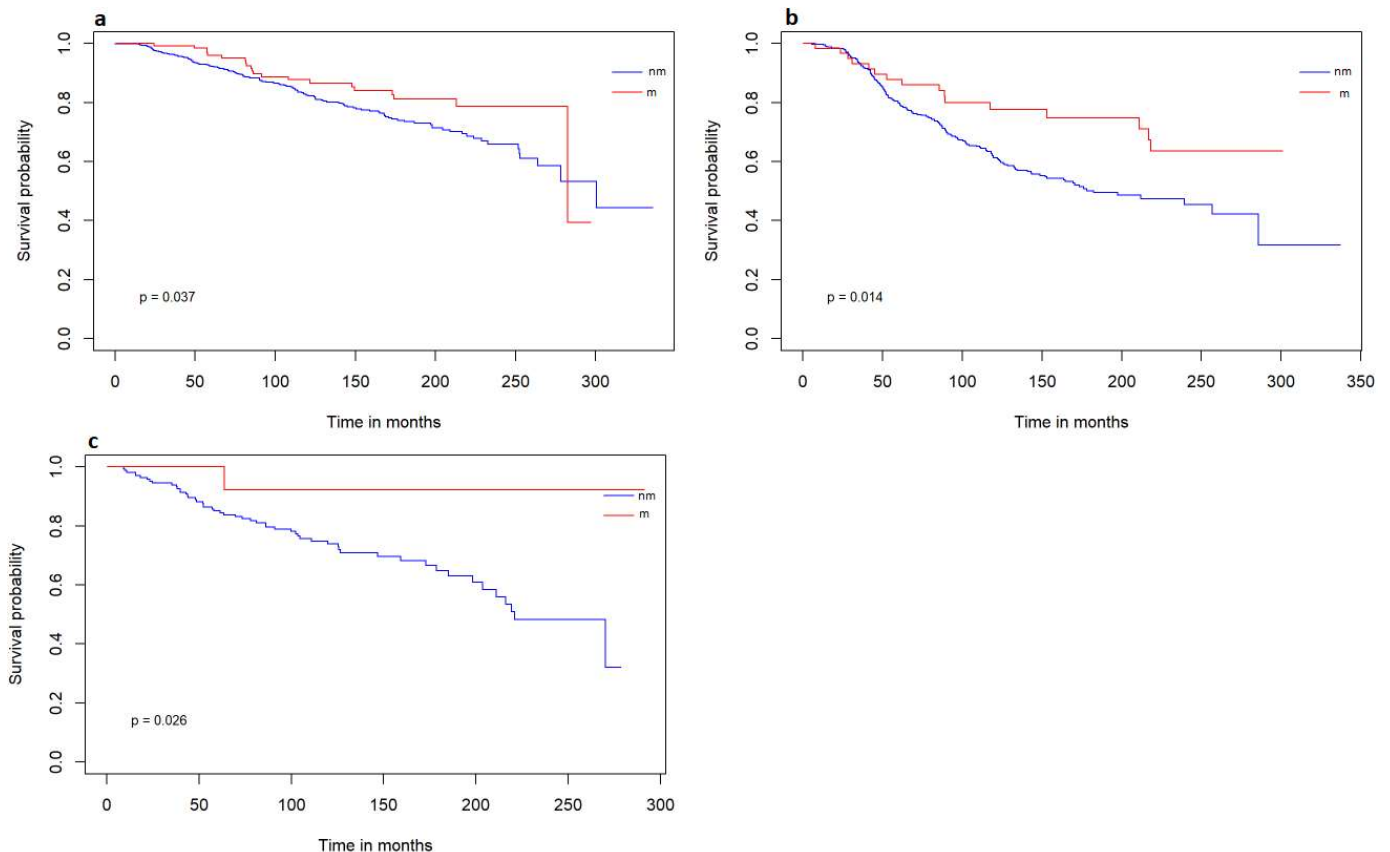


Figura 14a. Supervivencia en los pacientes clasificados dentro del subtipo Luminal A cuando presentan mutación en el gen GATA3 (m), con respecto a cuando no la presentan (nm)

Figura 14b. Supervivencia en los pacientes clasificados dentro del subtipo Luminal B cuando presentan mutación en el gen GATA3 (m), con respecto a cuando no la presentan (nm)

Figura 14c. Supervivencia en los pacientes clasificados dentro del subtipo Normal cuando presentan mutación en el gen GATA3 (m), con respecto a cuando no la presentan (nm)

A continuación, se muestra una tabla resumen (tabla 7) en la que se indican aquellos genes para cada subtipo de la clasificación PAM50 en los que sí hay diferencias significativas en la supervivencia de los pacientes cuando los presentan mutados y cuando no.

Tabla 8. Genes que, cuando presentan mutación, influyen en la supervivencia de los pacientes pertenecientes a cada subtipo de cáncer de mama estudiado en la clasificación PAM50

	Luminal A	Luminal B	Her2	Normal	Basal
TP53	X	X		X	
PIK3CA		X			X
GATA3	X	X		X	
AHNAK2			X		

4.5. Pathway analysis

Vista la supervivencia de los pacientes cuando presentan mutación en los genes estudiados y cuando no la presentan, y pudiendo observar que, realmente, algunos de los genes sí influyen en la misma, resulta interesante estudiar las rutas moleculares y las funciones biológicas en las cuales se encuentran implicados dichos genes.

Se obtuvieron de esta forma una serie de relaciones en procesos biológicos con los genes que hasta el momento han demostrado influir en la supervivencia de los pacientes.

Rutas biológicas en las que se encuentran implicados los genes estudiados

En la figura 15 se muestra un esquema en el cual se observa como el gen TP53 participa, junto con el gen PIK3CA, en la regulación de diferentes procesos que presentan relación directa e indirectamente con desarrollo de cáncer de mama. En este gráfico se identifican genes y proteínas relacionadas con tumorigenesis, supresión de tumores, rutas de reparación del ADN o ciclo celular. Las funciones de estos genes, así como su implicación en estos procesos se discutirá en el apartado de Discusión. Dicho esquema es una extracción del Pathway analysis realizado mediante el empleo de la base de datos KEGG, en el cual se pudo observar la participación de dichos genes en el proceso de cáncer de mama.

Adicionalmente, se obtuvo mediante el empleo del programa String las rutas incluidas en la figura 16, donde puede observarse las diferentes rutas funcionales con elementos en común con la figura 15 para los genes TP53, PIK3CA y AHNAK2.

BREAST CANCER

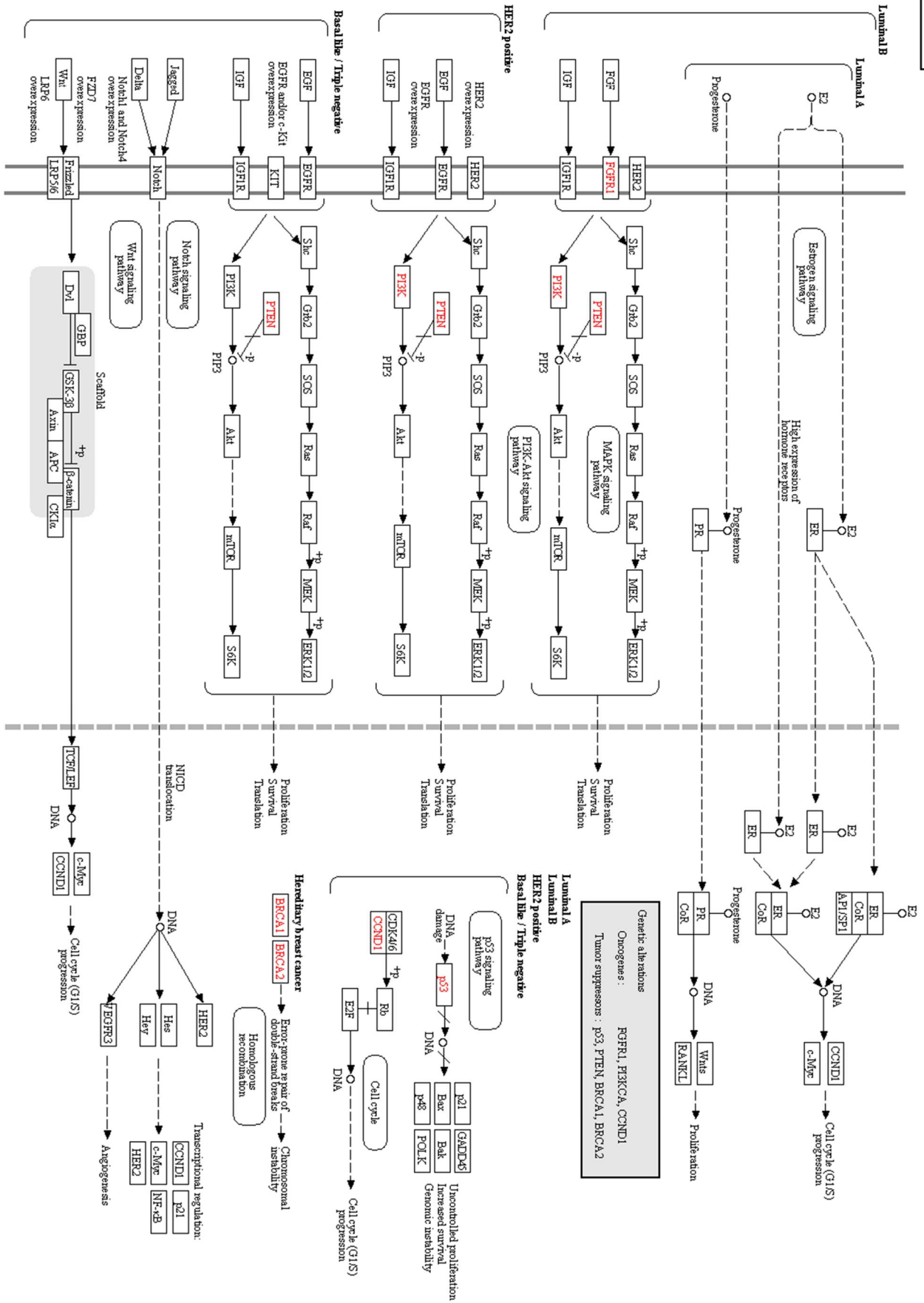


Figura 15. Pathway análisis del gen PIK3CA y TP53 mediante el uso de la base de datos KEGG.

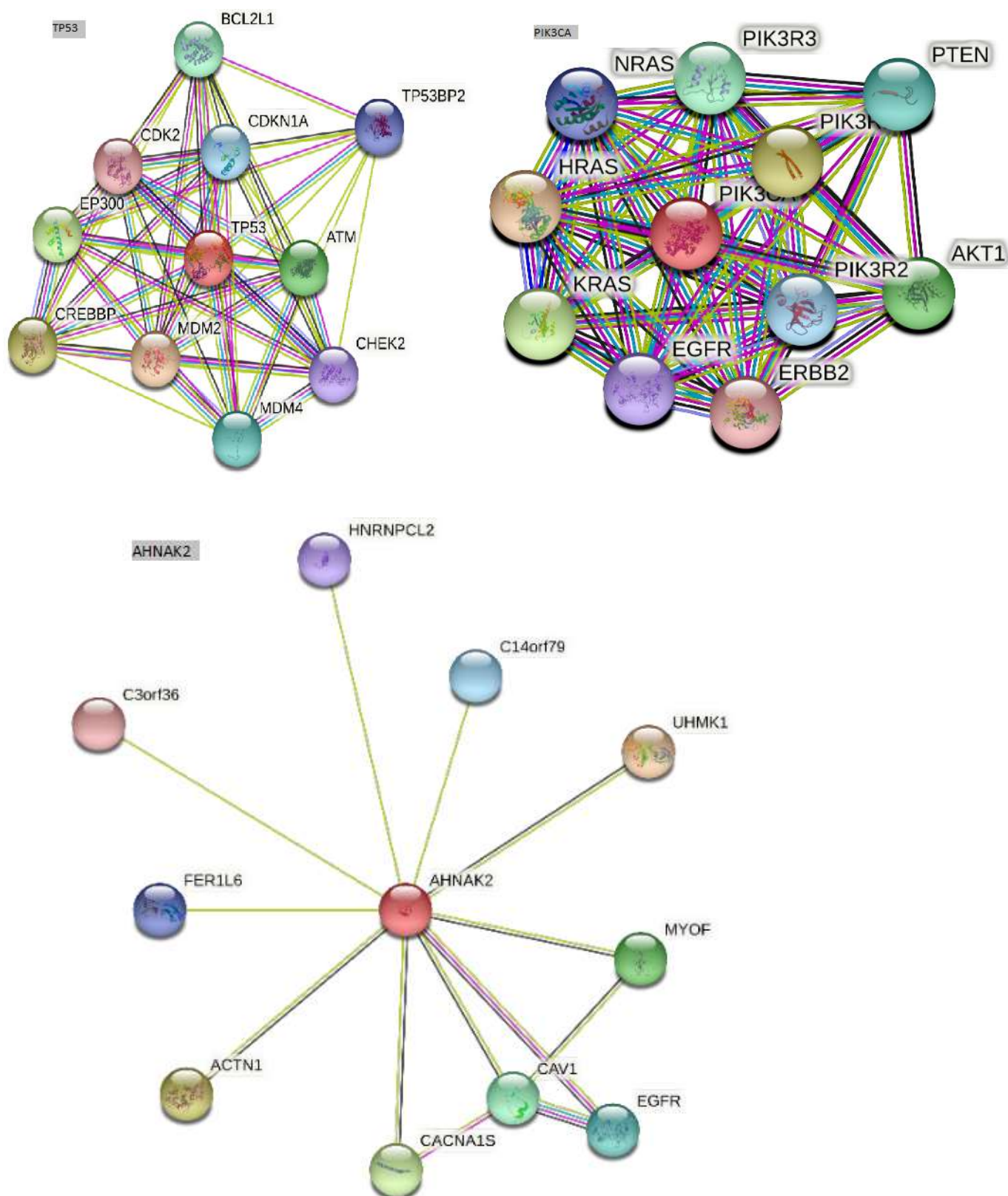


Figura 16. Pathway análisis realizado para los genes TP53, PIK3CA y AHNAK2, respectivamente, mediante uso de String.

Lo mismo se realizó para el gen GATA3 (figura 17), donde se pudo relacionar con los diferentes componentes biológicos que pueden verse a continuación.

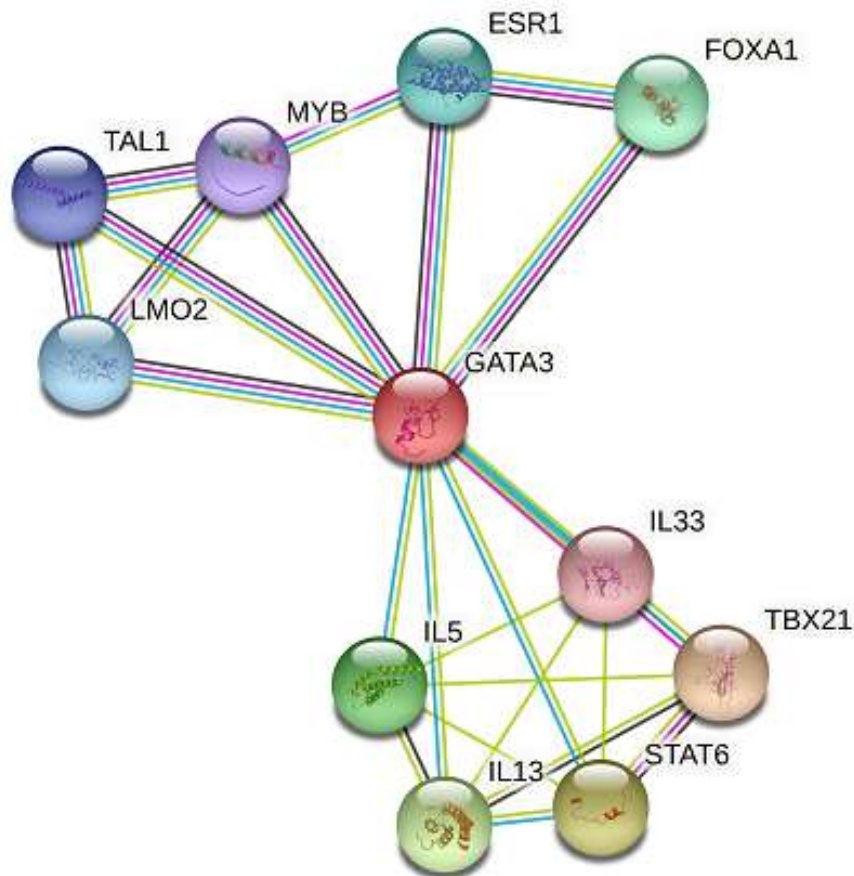


Figura 17. Pathway análisis realizado para el gen GATA3 mediante uso de String.

Análisis de ontologías genéticas

Se indican a continuación los resultados obtenidos a partir del estudio de ontologías para los genes que ocupan el presente punto:

- TP53 → En el estudio de ontologías genéticas se vio que dicho gen está relacionado con un proceso biológico identificado mediante el término GO:0030330 cuya descripción incluye: “DNA damage response, signal transduction by p53 class mediator”.
- PIK3CA → Mediante el análisis de ontologías genéticas, se obtuvo que dicho gen está relacionado con un componente celular con término GO:0005942, y cuyo nombre es fosfatidilinositol-3 kinasa. Este gen codifica para la subunidad p110 α del complejo kinasa mencionado, que a su vez participa en la ruta de señalización PTEN cuyo valor biológico se explicará en el apartado de Discusión.

- AHNAK2→Este gen, en el análisis de ontologías se relaciona con diversos procesos, entre ellos la reparación de la membrana celular (GO:0001778) o producción de proteínas de unión (GO:0005515).
- GATA3→ Este gen codifica para un factor de transcripción que participa en múltiples procesos biológicos relacionados con la producción y diferenciación de células T (GO:0030217).

5 Discusión

A continuación, se van a discutir los resultados obtenidos en cada uno de los puntos anteriores. Con ello, se pretende integrar toda la información obtenida y darle un valor biológico y clínico, para así aportar nuevos conocimientos en la investigación contra el cáncer de mama.

Estudio de las mutaciones somáticas en las diferentes clasificaciones de cáncer de mama

En primer lugar, se representaron para diferentes clasificaciones (PAM50, Histológica, Claudin-Subtype, ER, PR y HER2) un conjunto de histogramas en los que se pudo apreciar qué genes aparecían mutados mayormente en, al menos, el 10% de los pacientes que formaban parte del estudio inicial de METABRIC. Con ello, se pudo observar un conjunto de genes que se encontraban mutados en todas las clasificaciones estudiadas (sin profundizar en los subtipos de las mismas). Estos genes son: TP53, PIK3CA, AHNK2 y GATA3 (tabla 6). Esto cobra sentido porque, tal y como se comentará a continuación, estos genes participan en procesos involucrados en el cáncer de mama e influyen significativamente en la supervivencia de los pacientes.

De forma paralela, se vio como otro grupo de genes, tales como AHNK o KMT2C (tabla 6), aparecían mutados solo en determinadas clasificaciones, pero, como se tratará posteriormente, no se vio que influyeran en la supervivencia de los pacientes. Con ello, no parecían tener relevancia en el objetivo que ocupa el presente proyecto.

El siguiente paso, fue seleccionar un conjunto de genes para cada clasificación. Los genes seleccionados fueron aquellos que, según lo obtenido en los gráficos previos, habían demostrado estar mayormente mutados en los pacientes de cada una de las clasificaciones estudiadas. Con estos genes, se realizó una prueba estadística de tipo chi-cuadrado para cada uno de los subtipos que componen cada clasificación estudiada. El nivel de confianza establecido fue del 95%. El objetivo de ello fue establecer (mediante la obtención de p-valores inferiores a 0.05) que las mutaciones en los genes seleccionados para los subtipos que conforman cada clasificación influían significativamente en la diferenciación de dichos subtipos o que, al menos, presentaban variaciones significativas entre ellos.

Los resultados obtenidos en este punto dieron una validación de los subtipos estudiados desde una vista a nivel mutacional, lo cual resulta interesante ya que estas clasificaciones no se establecen en base a las mutaciones que contienen los genes estudiados. El enfoque mutacional demostró, de esta forma, ser una buena herramienta para diferenciar subtipos de pacientes con cáncer de mama.

Así pues, tras una visión a nivel global de los resultados obtenidos en este proceso exploratorio, se consideró oportuno centrar los siguientes pasos en la clasificación PAM50. La clasificación PAM50 es una de las más ampliamente usadas para el pronóstico de pacientes con cáncer de mama, especialmente para la determinación del uso de tratamientos hormonales debido a las características que presentan sus subtipos. Se ha visto en estudios recientes que esta clasificación proporciona información relevante en cuanto al pronóstico de pacientes con cáncer de mama a nivel clínico-patológico, se trata de una clasificación aparentemente útil para la estimación de la supervivencia [38]. Esto se confirma en los resultados obtenidos en el presente estudio y que se contrastan a continuación. A pesar de lo mencionado, hay diversos estudios que actualmente siguen en la búsqueda de tratamientos funcionales según qué subtipos de esta clasificación presentan los pacientes [39, 40, 41] y esto da una aproximación sobre el amplio abanico de opciones que esta clasificación puede ser capaz de proporcionar a la hora de establecer un pronóstico y un tratamiento.

La exploración de la clasificación PAM50 a nivel mutacional en diferentes genes puede ser un abordaje interesante debido a que, como se ha mencionado previamente, esta clasificación se basaría en los niveles de expresión de 50 genes, pero no en el perfil mutacional somático de los pacientes.

Adicionalmente, resulta interesante esta clasificación ya que, tras una exploración inicial de los genes que aparecían mutados en todas las clasificaciones, parecía haber un patrón común mutacional en los genes TP53, PIK3CA, AHNK2 y GATA3 en todas ellas. Sin embargo, pudo observarse como el gen GATA3, para la clasificación PAM50, era un gen característico en los subtipos luminales A y B. Esto parece tener suma importancia siguiendo con la dinámica que toman los estudios que abordan esta clasificación, tal y como se menciona en las líneas anteriores.

Por tal de confirmar lo observado y siguiendo con la exploración que caracteriza el proyecto, se realizó un estudio del perfil mutacional de los subtipos que conforman la clasificación PAM50 para los cuatro genes indicados (figuras 6, 7, 8 y 9). De esta forma, pudo confirmarse que existe un patrón diferente entre las mutaciones de estos grupos de genes, ya que los genes TP53, PIK3CA y AHNK2 presentaban mayoritariamente un tipo de mutación (missense_SNV) y el gen GATA3 otro distinto (splice_mut e indel), tema en el cual se profundizará a posteriori. Esto parece cobrar cierta importancia, ya que confirma los patrones observados hasta el momento y parece ser un punto de inflexión a la hora de clasificar estos subtipos, abriendo una ventana a la hora de estudiar los pacientes luminales.

En las representaciones mostradas previamente, se ha podido determinar como el gen TP53 tiene un papel notorio para todos y cada uno de los subtipos de cada clasificación, por lo que se espera que el estudio de éste, junto con la clasificación PAM50 pueda aportar información de gran interés. Además, se ha visto que, con el estudio de la mutación en el gen TP53, podría alcanzarse con más facilidad un pronóstico en los pacientes que padecen este tipo de cáncer, aportando un valor predictivo adicional al proporcionado por la clasificación PAM50. Los autores de este estudio proponen ampliar esta información sobre

mutaciones en el gen TP53 para poder integrarla en el pronóstico basado en la clasificación PAM50 [42].

Por todo lo mencionado, se consideró apropiado centrar los siguientes pasos del presente proyecto en esta clasificación, dejando abiertas las vías de abordaje para el resto de las clasificaciones, ya que efectivamente se demostró que las mutaciones en los genes estudiados influyen en la diferenciación de todos los subtipos analizados.

Así pues, centrándonos en la clasificación PAM50, se realizó el estudio de la supervivencia de los pacientes pertenecientes a cada subtipo que la compone. De esta forma, se apreció que la supervivencia de los pacientes era significativamente diferente entre los subtipos y, por tanto, su pronóstico también.

Según la figura 10, el pronóstico era más favorable para los pacientes que pertenecían al subtipo Luminal A y se determinó un peor pronóstico para los pacientes que formaban parte del subtipo Her2. Esto cobra cierto sentido ya que, aunque se ha visto que la supervivencia es una variable multifactorial, en la que participan factores como la edad, el sexo o el diagnóstico tardío [43], los pacientes incluidos en subtipo Luminal A, suelen relacionarse con una mayor supervivencia debido a que presentan (en la mayoría de los casos) una buena respuesta a tratamientos endocrinos, a diferencia de los subtipos Her2 frente a los que se suele actuar con quimioterapia [44].

Por otro lado, se ha visto que los pacientes luminales, en general, presentan un mejor pronóstico debido a la temprana y más fácil diferenciación de los tumores mamarios que favorecen un diagnóstico y tratamiento más temprano que en los Her2 [45]. Paralelamente, se ha visto que los pacientes Luminales A, en contraposición con los pacientes Her2, presentan una menor probabilidad de metástasis y en caso de desarrollo de la misma, una menor agresividad [46]. Por tanto, todos estos factores contribuyen al aumento de supervivencia de los pacientes Luminales A con respecto a los pacientes Her2.

Por todo lo expuesto y porque los subtipos de pacientes, tal y como se menciona anteriormente, presentan diferentes marcadores moleculares que los caracterizan, no resulta extraño lo obtenido en los resultados del estudio de la supervivencia para la clasificación PAM50, siendo, según lo obtenido, los pacientes Luminales A los que mejor pronóstico presentan y los Her2 los que peor pronóstico tienen.

Seguidamente, se pasó a la siguiente fase, en la que se representó la supervivencia de los pacientes pertenecientes a cada uno de los subtipos PAM50 cuando presentaban mutación en los genes seleccionados, con respecto a cuando no presentaban mutación en dichos genes. De esta forma, se pudo ver, tal y como se describe en la tabla 8 mostrada anteriormente, que no todos los genes estudiados influían de forma significativa en la supervivencia de los pacientes al presentar mutación, ni todos los genes que sí influían en la supervivencia cuando estaban mutados coincidían en todos los subtipos, siendo algo que característico de cada uno de los subtipos que aportaría un gran valor al presente estudio.

Se tratan, a continuación, aquellos genes que han demostrado presentar un papel importante, a nivel mutacional, en los subtipos de la clasificación PAM50:

El gen TP53 y su implicación en el cáncer de mama

Tras la representación de las curvas de supervivencia, se pudo ver como las mutaciones en el gen TP53 influyen en la supervivencia de los pacientes que forman parte de los subtipos Luminales A, Luminales B y Normal, disminuyendo la supervivencia de estos pacientes al presentar mutación (figura 11).

En primer lugar, se tratará el motivo por el cual afecta únicamente a estos tres subtipos: Los luminales comparten ciertas características moleculares como ser positivos para el receptor de estrógenos o de progesterona, cosa que podría justificar el papel de este gen sobre ambos subtipos. Sin embargo, también sería atribuible su influencia en el grupo Normal, a que este último se trata de un subtipo muy heterogéneo y puede compartir características con todos y cada uno de los subtipos existentes en esta clasificación.

Por otro lado, en todos los casos, la presencia de mutación en el gen TP53 induce a una menor supervivencia. Además, se ve oportuno considerar que, para este gen, tal y como establece el estudio del perfil mutacional, la mayoría de las mutaciones son de tipo missense_snv (figura 6), es decir, una mutación que produce un cambio de aminoácido en la proteína final que codifica el gen.

Con esta información, y teniendo en consideración el análisis de rutas moleculares en las que se encuentra implicado este gen (figura 15 y 16) y de ontologías genéticas, donde se determina su papel principal en procesos de reparación de ADN y supresión de tumores, puede afirmarse que el gen TP53 codifica para la proteína p53, la cual se trata de un factor de transcripción multifuncional que participa en procesos como estabilizador genómico, regulador de apoptosis celular o regulador del ciclo celular [47]. Este hecho queda patente en las figuras 15 y 16, donde puede observarse su papel regulador en proteínas como GADD45, p48, p21, POLK, Bax o Bak, las cuales participan en el normal funcionamiento del sistema inmune, actuando contra la oncogénesis [48], reparación del ADN [49], transcripción o regulación de la muerte celular [50].

El estudio realizado previamente para el establecimiento del perfil mutacional en este gen (figura 6) confirma, tal y como se menciona anteriormente, que la mutación más común en los pacientes que presentan mutación en este gen es la de tipo SNV. Una mutación de este tipo en el gen TP53, la cual puede dar como resultado una proteína anómala en el producto final del mismo, puede suponer un truncamiento en los procesos mencionados y regulados por dicha proteína, los cuales pueden participar en el desarrollo del tumor mamario al restringir o modificar las funciones reguladoras de las proteínas implicadas en las rutas biológicas mencionadas, como las que aparecen en las figuras 15 y 16. Esto, como es lógico, puede influenciar de forma muy notable en el desarrollo del cáncer de mama, actuando como agravante (o más bien no evitando su desarrollo o empeoramiento) y, por supuesto, produciendo un descenso en el parámetro supervivencia cuando ello ocurre [51].

Adicionalmente, cabe mencionar que, en algunos estudios se apunta que el gen TP53 puede ser un buen marcador (cuando está mutado) para el diagnóstico de pacientes ER+, como pueden ser los Luminales A y los Luminales B [52], lo cual resulta interesante en el tema que el presente proyecto abarca y cuyo resultado hasta el momento, reafirma.

EL gen PIK3CA y su implicación en el cáncer de mama

Puede observarse en los gráficos de supervivencia presentados (figura 12), como las mutaciones en el gen PIK3CA influyen notablemente en la variación de la supervivencia de los pacientes Luminales B y Basal. Al nivel de la información que ocupa el presente proyecto, este dato no llegaría a ser vinculante, puesto que estos subtipos no comparten ningún parámetro en común, no obstante, ello no desvincula que esta información llegue suponer una nueva vía de abordaje por tal de determinar características comunes entre estos grupos de pacientes.

A pesar de ello, se observa como mutaciones en el gen PIK3CA sí influyen en la supervivencia de los pacientes pertenecientes a los subtipos mencionados y se ve como es superior cuando no presentan mutación.

Atendiendo al establecimiento del perfil mutacional del gen PIK3CA en el presente proyecto (figura 7), la mutación más presente en este gen sería, tal y como ocurre en el gen TP53, de tipo missense_snv, lo cual produciría como resultado del proceso de síntesis proteica, una proteína anómala con posible alteración o rescisión de sus funciones.

Si se toma en consideración el análisis de rutas biológicas (figura 15 y 16) y ontologías genéticas para el gen PIK3CA, se determina que produce una proteína kinasa que participa en la ruta de señalización PTEN. Esta ruta de señalización está relacionada nuevamente con procesos de la misma índole que el gen TP53, el cual también se encuentra implicado en esta ruta (figura 15). Ambos participan en procesos relacionados con apoptosis, supresión de tumores y regulación de oncogenes que participan en el cáncer de mama en concreto.

Con todo ello, se podría deducir que una mutación inhibitoria o alterante en el resultado de la proteína codificada por el gen PIK3CA, podría producir una alteración o agravación en los procesos en los cuales se encuentra implicado, favoreciendo eventos como la proliferación celular y el desarrollo de tumores mamarios.

Además, se ha visto como mutaciones influyentes en la ruta de señalización PTEN, están relacionadas con el desarrollo de diferentes tipos de cáncer [53, 54], siendo el cáncer de mama uno más en los que podría ocurrir esto y pudiendo estar implicado en ello el gen PIK3CA.

El gen AHNK2 y su implicación en el cáncer de mama

Se vio también en la figura 13 como el gen AHNK2 influye en la supervivencia de los pacientes pertenecientes al grupo Her2 cuando presenta mutación. En

este caso, se observa como la presencia de mutación en este gen contribuye al aumento de la supervivencia de los pacientes.

En el estudio del perfil de mutaciones realizado (figura 8) puede verse como las mutaciones en este gen son, mayoritariamente, del tipo missense_snv, tal y como ocurre en los casos anteriores, pero con la diferencia de que, en este caso, parece producir un aumento de la supervivencia. Una modificación de este tipo en el gen puede producir como resultado una inactivación o una modificación de su función, tal y como ocurre con los genes TP53 y PIK3CA.

Tal y como puede observarse en el análisis de rutas biológicas (figura 16) y ontologías genéticas, este gen interactúa con el factor de crecimiento EGFR, que a su vez participa en las rutas biológicas en las que participan también los genes TP53 y PIK3CA (figura 15 y 16). Todo lo mencionado sería un indicio del papel importante que podría asumir en el cáncer de mama este gen, ya que algunos estudios lo asocian con la tumorigénesis cuando se produce una inactivación inapropiada del mismo [55]. El gen AHNK2 ha sido en numerosas ocasiones relacionado con diferentes tipos de cáncer [56,57] y en este caso, según lo mencionado, parece tener también una cierta relación con el cáncer de mama de forma directa o indirecta.

Se trata de un gen que ejerce un papel fundamental en la migración celular y que, tal y como se comenta en las líneas anteriores, se ha visto involucrado en diferentes procesos cancerígenos. En algunos estudios, se ha relacionado sus altos niveles de expresión con un mal pronóstico de cáncer de pulmón y cuyos autores plantean que una baja expresión del mismo podría reducir la severidad del cáncer [58]. Con ello, se plantea la opción de que una mutación en el gen AHNK2 influya de forma similar a una baja expresión del mismo, con lo cual se vería relacionado con un proceso en el que la severidad del cáncer disminuyera, tal y como lo plantean los autores del estudio indicado arriba para el cáncer de pulmón. Ello conllevaría a concluir que, en este caso, muy improbablemente el gen es el detonante para el desarrollo del cáncer de mama, sino que, más bien, ejerce una función “terapéutica” en su pronóstico.

Es decisivo para la determinación de su asociación con el cáncer de mama, el hecho de que participe en rutas moleculares relacionadas con TP53 y PIK3CA, los cuales, está demostrado (como se menciona anteriormente) que participan en el desarrollo de cáncer de mama.

En el tema que ocupa, también habría que tomar en consideración que el grupo Her2, en el cual se ha visto que influye este gen, es muy heterogéneo y que, tal y como plantea un estudio reciente, las características moleculares de algunas variantes de este subtipo de cáncer aún están pendientes de elucidar [59]. Esto abriría una posible vía de investigación en la relación del gen AHNK2 y el subtipo Her2, cuya heterogeneidad podría afectar sobre los resultados de supervivencia observados.

EL gen GATA3 y su implicación en el cáncer de mama

Finalmente, la mutación en el gen GATA3, tal y como se observa en la figura 14, juega un papel importante en la supervivencia de los pacientes pertenecientes a los subtipos LumA, LumB y Normal. Así pues, tal y como

ocurre en el caso anterior, la presencia de mutación parece influir positivamente en la supervivencia (figura 14).

Atendiendo a la caracterización del perfil mutacional (figura 9), las mutaciones más comunes en este gen son de tipo indel y splicing, mutaciones que también producen variaciones significativas en las proteínas finales, pero siendo un patrón diferente al que se había observado hasta el momento en el que, en todos los casos, el perfil mutacional mayoritario era el SNV.

En el análisis de rutas biológicas para este gen (figura 17), se observa una relación molecular del mismo con el gen FOXA1, cuya mutación se relaciona con un efecto “terapéutico” en los pacientes con cáncer de mama cuando los pacientes son positivos para el receptor de estrógenos (ER+), como sería el caso de los grupos cuya supervivencia se ve afectada con la presencia de esta mutación en el gen GATA3 [60,61]. Así pues, la mutación en este gen podría influir en la expresión de FOXA1 (al estar molecularmente relacionados), siendo este hecho análogo a lo ocurrido cuando es el propio gen FOXA1 el que se encuentra mutado. De esta forma, podría considerarse que ejerce un efecto positivo para los grupos de pacientes ER+, como son los Luminales A, Luminales B y Normal, tanto en presencia de mutación en el gen GATA3 como el gen FOXA.

Este gen parece ocupar una posición importante en los pacientes luminales y este estudio espera ser un primer paso al desarrollo de nuevas vías que puedan abordarlo.

6 Conclusiones

6.1 Conclusiones del estudio

Con este estudio exploratorio, cuya finalidad ha residido en todo momento en extraer información sobre las clasificaciones y los subtipos de cáncer de mama establecidos hasta el momento a partir de datos mutacionales somáticos, se han podido concluir los siguientes puntos:

- A partir del estudio de las mutaciones somáticas y del perfil mutacional, se han observado diferentes patrones mutacionales en los pacientes que pertenecen a las diferentes clasificaciones y subtipos de cáncer de mama estudiados.
- Se ha visto como las mutaciones somáticas estudiadas influyen en la supervivencia de los pacientes y, por tanto, en su pronóstico. Por ello, el estudio mutacional somático puede abrir nuevos caminos a la hora de determinar clasificaciones y tratamientos para poder mejorar el pronóstico de los pacientes que padecen este tipo de cáncer.
- La clasificación PAM50 ocupa un lugar muy interesante a la hora de abordar el estudio mutacional de los pacientes que se encuentran clasificados en cada uno de sus subtipos. Se ha observado un patrón mutacional que aporta datos relevantes, en el que han participado los genes TP53, PIK3CA, AHNK2 y GATA3.
- Las mutaciones en los genes TP53, PIK3CA y AHNK2 presentan un perfil mutacional similar en la clasificación PAM50, donde se observan (de forma significativamente mayoritaria) mutaciones de tipo SNV en todos los casos para los subtipos que la conforman.
- Los genes TP53, PIK3CA y AHNK2 se encuentran conectados en procesos biológicos, ya que tanto el gen TP53 como el gen PIK3CA participan en la ruta de señalización PTEN, relacionada con procesos celulares anti cancerígenos. Estos dos genes también se relacionan directamente con rutas biológicas implicadas en el cáncer de mama. El gen PIK3CA y el gen AHNK2 intervienen en la función del factor de crecimiento EGFR, involucrado en la ruta biológica de desarrollo de cáncer de mama. Así pues, podría establecerse que una mutación en alguno de estos tres genes podría ser un indicador para el pronóstico o subclasificación de los pacientes que padecen cáncer de mama, siendo, de todas ellas, una mutación favorable en cuanto a

supervivencia, la producida en el gen AHNAK2 y desfavorables las producidas en los genes TP53 y PIK3CA.

- La mutación en el gen GATA3 parece ser un buen indicador para los pacientes positivos en el receptor de estrógenos, como serían los pacientes Luminales. Este gen presenta un perfil mutacional diferente al de los anteriores, donde abundan mutaciones de tipo splicing o indel (inserción-delección). Parece tener un efecto positivo en dichos pacientes al estar mutado, tal y como ocurriría con el gen FOXA1, con el cual se encuentra molecularmente relacionado. Este gen parece presentar resultados prometedores en cuanto a su análisis mutacional a la hora de clasificar o pronosticar los pacientes Luminales.
- Debe otorgársele una gran importancia clínica y científica al estudio de nuevas vías de investigación que aborden el cáncer de mama y a la continuidad y seguimiento de las que ya se encuentran iniciadas, siempre desde diferentes perspectivas biológicas y moleculares, como es el caso de las mutaciones. Ello se debe a que hay una necesidad incipiente en la población a la hora de reducir la mortalidad de esta enfermedad tan recurrente que afecta a un gran número de personas cada año con pronósticos que no siempre son alentadores. Con ello, queda demostrado que aun quedan muchos conocimientos que sacar a la luz y que la bioinformática puede y debe jugar un papel muy importante en ello.

Con estas conclusiones que se han podido establecer, se ha conseguido alcanzar el objetivo principal del proyecto que, como se ha mencionado desde el inicio, es exploratorio. Se ha podido determinar cómo las mutaciones en determinados genes sí influyen en la supervivencia de los pacientes para diferentes clasificaciones de cáncer de mama. Además, se han encontrado patrones en determinados genes que, cuando presentan mutación, pueden aportar información de valor a la hora de establecer un pronóstico, como sería la influencia de éstos en la supervivencia, una aproximación al tipo de cáncer que puede presentar el paciente en cuestión, así como las rutas biológicas que puede presentar alteradas.

6.2 Líneas de futuro

Aunque con el presente trabajo se han conseguido establecer una serie de conclusiones que cobran cierto sentido en el ámbito mutacional del cáncer de mama, no resulta sorprendente afirmar que estos genes pueden presentar muchas más prestaciones con un análisis de los mismos más concreto, donde partiendo de lo obtenido hasta este punto, se consideren de forma individual aquellos subtipos de pacientes que han presentado variaciones en su supervivencia al estar dichos genes mutados y se analicen otras características que puedan aportar más información al pronóstico, como los diferentes tratamientos empleados.

También puede aportar información de valor el estudio mutacional en el resto de las clasificaciones que han sido descartadas en el presente proyecto y ver qué tipo de genes pueden tener una papel importante (al encontrarse mutados) en cada uno de los diferentes subtipos, pudiendo relacionarse dicha información con la obtenida hasta el momento.

La elaboración de clasificaciones establecidas a partir de un patrón mutacional somático sería una buena forma de abordar el estudio del cáncer de mama, ya que se ha visto como determinadas mutaciones influyen significativamente en la supervivencia de los pacientes y como participan en rutas biológicas importantes para su desarrollo.

6.3 Seguimiento de la planificación

Como se menciona en diferentes ocasiones a lo largo del proyecto, el objetivo ha sido el de explorar los datos y explotar los resultados que pudieran obtenerse con ello. Para ello, desde el principio se ha planteado una planificación variable y adaptable a los resultados obtenidos a lo largo de las diferentes fases del trabajo. Con ello, el objetivo principal se ha conseguido.

7 Glosario

METABRIC: Molecular Taxonomy of Breast Cancer International Consortium.

SNV: Single nucleotid variation.

Indel: Inserción – delección.

LumA: Luminal A.

LumB: Luminal B.

GO: Gene Ontology.

8 Bibliografía

1. Llorc, G., Chirivella, I., Morales, R., Serrano, R., Sanchez, A. B., Teulé, A., ... & SEOM Hereditary Cancer Working Group. (2015). SEOM clinical guidelines in Hereditary Breast and ovarian cancer. *Clinical and Translational Oncology*, 17(12), 956-961.
2. Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., ... & Zhu, H. P. (2017). Risk factors and preventions of breast cancer. *International journal of biological sciences*, 13(11), 1387.
3. Kolak, A., Kamińska, M., Sygit, K., Budny, A., Surdyka, D., Kukielka-Budny, B., & Burdan, F. (2017). Primary and secondary prevention of breast cancer. *Ann Agric Environ Med*, 24(4), 549-553.
4. American Society of Clinical Oncology. (2008). Cancer. net.Accession: June 2021
5. Bruna, A., Rueda, O. M., Greenwood, W., Batra, A. S., Callari, M., Batra, R. N., ... & Caldas, C. (2016). A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. *Cell*, 167(1), 260-274.
6. Turashvili, G., & Brogi, E. (2017). Tumor heterogeneity in breast cancer. *Frontiers in medicine*, 4, 227.
7. Maughan, K. L., Lutterbie, M. A., & Ham, P. (2010). Treatment of breast cancer. *American family physician*, 81(11), 1339-1346.
8. Blows, F. M., Driver, K. E., Schmidt, M. K., Brooks, A., Van Leeuwen, F. E., Wesseling, J., ... & Huntsman, D. (2010). Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS med*, 7(5), e1000279.
9. Yoshida, K., & Miki, Y. (2004). Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer science*, 95(11), 866-871.
10. Pereira, B., Chin, S. F., Rueda, O. M., Vollan, H. K. M., Provenzano, E., Bardwell, H. A., ... & Caldas, C. (2016). The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature communications*, 7(1), 1-16.
11. Yeo, S. K., & Guan, J. L. (2017). Breast cancer: multiple subtypes within a tumor?. *Trends in cancer*, 3(11), 753-760.
12. Perou, C. M., Sørlie, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Rees, C. A., ... & Botstein, D. (2000). Molecular portraits of human breast tumours. *nature*, 406(6797), 747-752.
13. Nielsen, T., Wallden, B., Schaper, C., Ferree, S., Liu, S., Gao, D., ... & Storhoff, J. (2014). Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC cancer*, 14(1), 1-14.

14. Pellegrino, B., Hlavata, Z., Migali, C., De Silva, P., Aiello, M., Willard-Gallo, K., ... & Solinas, C. (2021). Luminal Breast Cancer: Risk of Recurrence and Tumor-Associated Immune Suppression. *Molecular Diagnosis & Therapy*, 1-16.
15. Wallden, B., Storhoff, J., Nielsen, T., Dowidar, N., Schaper, C., Ferree, S., ... & Parker, J. S. (2015). Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC medical genomics*, 8(1), 1-14.
16. Inwald, E. C., Klinkhammer-Schalke, M., Hofstädter, F., Zeman, F., Koller, M., Gerstenhauer, M., & Ortmann, O. (2013). Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast cancer research and treatment*, 139(2), 539-552.
17. Rakha, E. A., Reis-Filho, J. S., Baehner, F., Dabbs, D. J., Decker, T., Eusebi, V., ... & Ellis, I. O. (2010). Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Research*, 12(4), 1-12.
18. Christgen, M., Länger, F., & Kreipe, H. (2016). Histological grading of breast cancer. *Der Pathologe*, 37(4), 328-336.
19. Weigelt, B., Geyer, F. C., & Reis-Filho, J. S. (2010). Histological types of breast cancer: how special are they?. *Molecular oncology*, 4(3), 192-208.
20. Fougner, C., Bergholtz, H., Norum, J. H., & Sørli, T. (2020). Re-definition of claudin-low as a breast cancer phenotype. *Nature communications*, 11(1), 1-11.
21. Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., ... & Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research*, 12(5), 1-18.
22. Elbachiri, M., Fatima, S., Bouchbika, Z., Bencheikroun, N., Jouhadi, H., Tawfiq, N., ... & Benider, A. (2017). Breast cancer in men: about 40 cases and literature review. *The Pan African medical journal*, 28, 287-287.
23. Radecka, B., & Litwiniuk, M. (2016). Breast cancer in young women. *Ginekologia polska*, 87(9), 659-663.
24. Alegre, M. M., Knowles, M. H., Robison, R. A., & O'Neill, K. L. (2013). Mechanics behind breast cancer prevention-focus on obesity, exercise and dietary fat. *Asian Pacific Journal of Cancer Prevention*, 14(4), 2207-2212.
25. Kelsey, J. L., & Bernstein, L. (1996). Epidemiology and prevention of breast cancer. *Annual review of public health*, 17(1), 47-67.
26. Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, 458(7239), 719-724.
27. Del Valle, S. A. C., & VILLEGAS, L. J. J. G. (2012). Cáncer de mama asociado a mutaciones genéticas de los BRCA 1 y 2/Breast cancer associated to BRCA 1 and 2 genetic mutationsLOS BRCA 1 y 2. *CES Medicina*, 26(2), 185-200.
28. Locken-Castilla, A., Pacheco-Pantoja, E. L., Rodríguez-Brito, F., May-Kim, S., López-Rivas, V., & Ceballos-Cruz, A. (2019). Smoking index, lifestyle

- factors, and genomic instability assessed by single-cell gel electrophoresis: a cross-sectional study in subjects from Yucatan, Mexico. *Clinical epigenetics*, 11(1), 1-14.
29. Gumeni, S., Evangelakou, Z., Gorgoulis, V. G., & Trougakos, I. P. (2017). Proteome stability as a key factor of genome integrity. *International journal of molecular sciences*, 18(10), 2036.
 30. Shapira, N. (2017). The potential contribution of dietary factors to breast cancer prevention. *European Journal of Cancer Prevention*, 26(5), 385.
 31. Li, J., Chen, Z., Su, K., & Zeng, J. (2015). Clinicopathological classification and traditional prognostic indicators of breast cancer. *International journal of clinical and experimental pathology*, 8(7), 8500.
 32. Fisusi, F. A., & Akala, E. O. (2019). Drug combinations in breast cancer therapy. *Pharmaceutical nanotechnology*, 7(1), 3-23.
 33. Cain, H., Macpherson, I. R., Beresford, M., Pinder, S. E., Pong, J., & Dixon, J. M. (2017). Neoadjuvant therapy in early breast cancer: treatment considerations and common debates in practice. *Clinical Oncology*, 29(10), 642-652.
 34. Goldhirsch, A., Wood, W. C., Coates, A. S., Gelber, R. D., Thürlimann, B., & Senn, H. J. (2011). Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of oncology*, 22(8), 1736-1747.
 35. KEGG: Kyoto Encyclopedia of Genes and Genomes <https://www.genome.jp/kegg/> Accession June 2021.
 36. STRING: functional protein association networks <https://string-db.org/> Accession June 2021.
 37. 2021Gene Ontology Resource <http://geneontology.org/> Accession June 2021.
 38. Wallden, B., Storhoff, J., Nielsen, T., Dowidar, N., Schaper, C., Ferree, S., ... & Parker, J. S. (2015). Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC medical genomics*, 8(1), 1-14.
 39. Prat, A., Bianchini, G., Thomas, M., Belousov, A., Cheang, M. C., Koehler, A., ... & Baselga, J. (2014). based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clinical Cancer Research*, 20(2), 511-521.
 40. Chia, S. K., Bramwell, V. H., Tu, D., Shepherd, L. E., Jiang, S., Vickery, T., ... & Nielsen, T. O. (2012). A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clinical cancer research*, 18(16), 4465-4472.
 41. Gluz, O., Kolberg-Liedtke, C., Prat, A., Christgen, M., Gebauer, D., Kates, R., ... & Harbeck, N. (2020). Efficacy of deescalated chemotherapy according to PAM50 subtypes, immune and proliferation genes in triple-negative early breast cancer: Primary translational analysis of the WSG-ADAPT-TN trial. *International journal of cancer*, 146(1), 262-271.

42. Lin, C. H., Chen, I. C., Huang, C. S., Hu, F. C., Kuo, W. H., Kuo, K. T., ... & Cheng, A. L. (2015). TP53 mutational analysis enhances the prognostic accuracy of IHC4 and PAM50 assays. *Scientific reports*, 5(1), 1-9.
43. Rueda, O. M., Sammut, S. J., Seoane, J. A., Chin, S. F., Caswell-Jin, J. L., Callari, M., ... & Curtis, C. (2019). Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature*, 567(7748), 399-404.
44. Liu, M. C., Pitcher, B. N., Mardis, E. R., Davies, S. R., Friedman, P. N., Snider, J. E., ... & Barry, W. T. (2016). PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline-and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *NPJ breast cancer*, 2(1), 1-8.
45. Bueno, G. A. M. (2017). Características clínicas y pronósticas de los subtipos moleculares de cáncer de mama determinados por inmunohistoquímica. Arequipa, Perú. *Revista Peruana de Medicina Experimental y Salud Pública*, 34, 472-477.
46. Molnár, I. A., Molnár, B. Á., Vízkeleti, L., Fekete, K., Tamás, J., Deák, P., ... & Tőkés, A. M. (2017). Breast carcinoma subtypes show different patterns of metastatic behavior. *Virchows Archiv*, 470(3), 275-283.
47. Elledge, R. M., & Allred, D. C. (1998). Prognostic and predictive value of p53 and p21 in breast cancer. *Breast cancer research and treatment*, 52(1), 79-98.
48. Schmitz, I. (2013). Gadd45 proteins in immunity. *Gadd45 Stress Sensor Genes*, 51-68.
49. Itoh, T., O'Shea, C., & Linn, S. (2003). Impaired regulation of tumor suppressor p53 caused by mutations in the xeroderma pigmentosum DDB2 gene: mutual regulatory interactions between p48DDB2 and p53. *Molecular and cellular biology*, 23(21), 7540.
50. Itoh, T., O'Shea, C., & Linn, S. (2003). Impaired regulation of tumor suppressor p53 caused by mutations in the xeroderma pigmentosum DDB2 gene: mutual regulatory interactions between p48DDB2 and p53. *Molecular and cellular biology*, 23(21), 7540.
51. Hashimoto, N., Nagano, H., & Tanaka, T. (2019). The role of tumor suppressor p53 in metabolism and energy regulation, and its implication in cancer and lifestyle-related diseases. *Endocrine journal*, EJ18-0565.
52. Griffith, O. L., Spies, N. C., Anurag, M., Griffith, M., Luo, J., Tu, D., ... & Ellis, M. J. (2018). The prognostic effects of somatic mutations in ER-positive breast cancer. *Nature communications*, 9(1), 1-16.
53. Carnero, A., Blanco-Aparicio, C., Renner, O., Link, W., & Leal, J. F. (2008). The PTEN/PI3K/AKT signalling pathway in cancer, therapeutic implications. *Current cancer drug targets*, 8(3), 187-198.
54. Jiang, B. H., & Liu, L. Z. (2009). PI3K/PTEN signaling in angiogenesis and tumorigenesis. *Advances in cancer research*, 102, 19-65.
55. Sigismund, S., Avanzato, D., & Lanzetti, L. (2018). Emerging functions of the EGFR in cancer. *Molecular oncology*, 12(1), 3-20.
56. Zheng, L., Li, S., Zheng, X., Guo, R., & Qu, W. (2021). AHNK2 is a novel prognostic marker and correlates with immune infiltration in papillary thyroid

- cancer: Evidence from integrated analysis. *International Immunopharmacology*, 90, 107185.
57. Zhang, S., Lu, Y., Qi, L., Wang, H., Wang, Z., & Cai, Z. (2020). AHNAK2 is associated with poor prognosis and cell migration in lung adenocarcinoma. *BioMed Research International*, 2020.
 58. Wang, D. W., Zheng, H. Z., Cha, N., Zhang, X. J., Zheng, M., Chen, M. M., & Tian, L. X. (2020). Down-Regulation of AHNAK2 Inhibits Cell Proliferation, Migration and Invasion Through Inactivating the MAPK Pathway in Lung Adenocarcinoma. *Technology in cancer research & treatment*, 19, 1533033820957006.
 59. Schettini, F., Chic, N., Brasó-Maristany, F., Paré, L., Pascual, T., Conte, B., ... & Prat, A. (2021). Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. *npj Breast Cancer*, 7(1), 1-13.
 60. Arruabarrena-Aristorena, A., Maag, J. L., Kittane, S., Cai, Y., Karthaus, W. R., Ladewig, E., ... & Scaltriti, M. (2020). FOXA1 mutations reveal distinct chromatin profiles and influence therapeutic response in breast cancer. *Cancer Cell*, 38(4), 534-550.
 61. Fu, X., Pereira, R., De Angelis, C., Veeraraghavan, J., Nanda, S., Qin, L., ... & Schiff, R. (2019). FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. *Proceedings of the National Academy of Sciences*, 116(52), 26823-26834.