



UNIVERSIDAD DE SANTIAGO DE COMPOSTELA  
FACULTAD DE MEDICINA Y ODONTOLOGÍA

TRABAJO FIN DE GRADO DE MEDICINA

**ESTIMACIÓN DEL RIESGO POLIGÉNICO DE CÁNCER DE  
MAMA EN GALICIA**

AUTORA: Sobrino Robelo, Olalla

TUTOR: Carracedo Álvarez, Ángel María

COTUTORA: Gago Domínguez, Manuela

Departamento: Medicina Genómica y Molecular

Curso académico: 2019-2020

Convocatoria: Junio

## **Agradecimientos**

Mi más sincero agradecimiento al Dr. Carracedo y a la Dra. Gago por su tutela y gran ayuda en este Trabajo de Fin de Grado. A Juan Ansedo, por su impagable papel de guía en la aproximación al trabajo de laboratorio de genómica y a María Torres, siempre protectora del buen orden de las cosas.

A mi querido tío, por encaminarme hacia la pasión de mi vida cuando me sentía perdida.

A mi familia, por su apoyo incansable durante las largas horas de estudio.

A mis amigos, por hacer inmejorables estos años a vuestro lado.

A todos, gracias.

## Índice

Clave de abreviaturas.....	5
Resumen.....	8
Abstract.....	9
Resumo.....	10
1. Introducción.....	12
1.1. Generalidades de la glándula mamaria.....	12
1.2. El cáncer de mama. Subtipos.....	13
1.2.1. Subtipos histológicos.....	13
1.2.2. Subtipos moleculares.....	14
1.3. Epidemiología del cáncer de mama.....	16
1.4. Factores de riesgo del cáncer de mama.....	17
1.4.1. Factores de riesgo no modificables.....	17
1.4.1.1. Sexo.....	17
1.4.1.2. Edad.....	18
1.4.1.3. Ancestralidad.....	18
1.4.1.4. Vida fértil prolongada.....	18
1.4.1.5. Densidad mamaria.....	18
1.4.1.6. Alteraciones mamarias benignas/premalignas.....	19
1.4.1.7. Antecedentes personales de cáncer de mama.....	19
1.4.1.8. Antecedentes familiares de cáncer de mama.....	19
1.4.2. Factores de riesgo modificables.....	19
1.4.2.1. Nivel socioeconómico.....	19
1.4.2.2. Lugar de residencia.....	20
1.4.2.3. Hábitos reproductivos.....	20
1.4.2.4. Ejercicio físico e índice de masa corporal.....	20
1.4.2.5. Hábitos tóxicos, alcohol y tabaco.....	21
1.4.2.6. Exposición a estrógenos exógenos.....	21
1.4.2.7. Exposición a radiaciones.....	22
1.5. Carga genética del cáncer de mama.....	22
1.5.1. Genes de alta penetrancia.....	24
1.5.1.1. BRCA1 y BRCA2.....	24
1.5.1.2. TP53.....	24
1.5.1.3. PTEN.....	25
1.5.1.4. SKT-11.....	25
1.5.2. Genes de moderada penetrancia.....	26
1.5.2.1. ATM.....	26
1.5.2.2. CHEK2.....	26
1.5.2.3. PALB2.....	27
1.5.3. Genes de baja penetrancia.....	28

1.6. Predicción de riesgo.....	29
1.7. Análisis de la variación genética.....	30
1.7.1. Secuenciación.....	30
1.7.2. Genotipado.....	31
1.7.2.1. Reacción en cadena de la polimerasa (PCR) .....	31
1.7.2.2. Técnica iPLEX.....	32
2. Hipótesis.....	35
3. Objetivos.....	35
4. Material y métodos.....	35
4.1. Población de estudio: Cohorte Breogán.....	35
4.1.1. Información sobre factores de riesgo epidemiológicos.....	36
4.1.2. Información clínico-patológica.....	36
4.2. Diseño de panel de SNPs.....	37
4.3. Desarrollo de los ensayos de genotipado.....	39
4.3.1. Protocolo de extracción de ADN.....	39
4.3.2. Determinación de la concentración de ADN.....	40
4.3.3. Normalización de las muestras.....	40
4.3.4. PCR múltiple.....	41
4.3.5. Reacción de purificación con fosfatasa alcalina SAP.....	41
4.3.6. Reacción de extensión de base única (Reacción iPLEX) .....	41
4.3.7. Purificación de los productos de extensión.....	41
4.3.8. Transferencia de la muestra.....	42
4.3.9. Lectura de los productos de extensión.....	42
4.4. Análisis estadístico.....	42
5. Resultados.....	43
5.1. Diseño de los ensayos de genotipado.....	43
5.2. Datos epidemiológicos y del tipo de tumor.....	43
5.3. Evaluación de los SNPs en sus 71 loci.....	44
5.4. Evaluación de la asociación según subtipo molecular ER/PR/HER2 (+/-).....	44
5.5. Análisis del score de riesgo poligénico (PRS) para el cáncer de mama en general.....	46
6. Discusión.....	47
7. Perspectivas futuras.....	50
8. Conclusiones.....	51
9. Anexo.....	53
10. Referencias bibliográficas.....	56

## Clave de abreviaturas

- ADAM29: Proteasa con dominio metaloproteasa y desintegrina 29.
- ADS: Software de diseño de ensayos de laboratorio de AgenaCx Bioscience (*AgenaCx Assay Design Software V2*).
- AECC: Asociación Española contra el Cáncer.
- ADN: Ácido Desoxirribonucleico.
- ATM: Ataxia telangiectasia mutado (gen).
- AUC: Área bajo la curva (*Area under the ROC Curve*).
- BCAC: Consorcio Internacional de Cáncer de Mama (*Breast Cancer Association Consortium*).
- BOADICEA: Algoritmo de estimación de riesgo de desarrollo de cáncer de mama y ovario y de estado de portador de mutaciones de riesgo (*Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm*).
- BRCA1: Gen de susceptibilidad al cáncer de mama tipo 1 (*BRest CAncer susceptibility 1*).
- BRCA2: Gen de susceptibilidad al cáncer de mama tipo 2 (*BRest CAncer susceptibility 2*).
- BREOGAN: Consorcio Gallego de Cáncer de Mama (*BRest Oncology GALician Network*).
- BRIP1: Gen Helicasa que interacciona con el extremo C-terminal de BRCA1 (*BRCA1 Interacting Protein C-Terminal Helicase 1*).
- C-KIT: Receptor transmembrana con actividad tirosin-quinasa.
- CDH1: Gen de la cadherina 1 (*cadherine 1*).
- CDI: Carcinoma ductal infiltrante.
- CEU: Cohorte de referencia del proyecto HapMap de población de ascendencia europea (*CEU: Northern Europeans from Utah*).
- CHEK2: Gen serina/treonina protein-quinasa CHK2.
- CI: Intervalo de confianza.
- CK5: Citoqueratina básica 5 de expresión epitelial.
- CK17: Citoqueratina ácida 17 de expresión epitelial.
- CLI: Carcinoma lobulillar infiltrante.
- CM: Carcinoma medular.
- COGS: Estudio Oncológico Colaborativo de Genes-Ambiente (*Collaborative Oncological Gene-Environment Study*).
- dbSNP: Base de datos de polimorfismos de un solo nucleótido (*Single Nucleotide Polymorphism Database*).
- ddNTP: didesoxinucleótido.
- dNTP: desoxirribonucleótido.
- EBF1: Gen del factor de transcripción temprano de célula B (*Early B-cell Factor-1*).
- ER: Expresión de receptores de estrógenos en la superficie celular (*Estrogen Receptors*).
- FGFR2: Receptor 2 del factor de crecimiento de fibroblastos (*Fibroblast growth factor receptor 2*).

FTO: Gen codificante de la proteína asociada con la masa grasa y obesidad (*Fat mass and obesity-associated protein*).

GWAS: Estudios de asociación de genoma completo (*Genome-wide association studies*).

HapMap: *Haplotype Map Project*.

HER2: Receptor 2 de factor de crecimiento epidérmico humano (*Human Epidermal Growth Factor Receptor 2*).

IANUS: Sistema gallego de historia clínica electrónica.

iCOGS: Array de Illumina customizado para el COGS.

IDH: Índice de desarrollo humano.

IMC: Índice de masa corporal.

KI67: Marcador de proliferación celular.

LKB1: Proteína quinasa hepática B1 (*liver kinase B1*).

LSP1: Proteína específica de linfocito 1 (*Lymphocyte Specific Protein 1*).

MAF: Frecuencia del alelo menos común (*Minor allele frequency*).

MALDI-TOF: Espectrometría de masas basada en la desorción/ionización mediante láser asistida por matriz acoplada a un analizador de tiempo de vuelo (*Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry*).

MAP3K1: Gen MAP quinasa quinasa quinasa 1 (*mitogen-activated kinase 1*).

MLH1: Gen reparador de ADN *MutL Homolog 1*.

MPS: Secuenciación paralela masiva (*Massive parallel sequencing*).

MQ: Agua purificada utilizando un sistema de intercambio de iones (*Milli-Q water*).

MSH2: Gen reparador de ADN *MutS Homolog 2*.

MSH6: Gen reparador de ADN *MutS Homolog 6*.

OR: Riesgo relativo (*Odds ratio*).

P53: Proteína tumoral 53 (*tumor protein p53*).

PALB2: Gen acompañante y localizador de BRCA2 (*Partner and localizer of BRCA2*).

PCR: Reacción en cadena de la polimerasa (*polymerase chain reaction*).

PI3K/AKT: Fosfatidilinositol 3'-quinasa /proteín-quinasa B.

PR: Expresión de receptores de progesterona en la superficie celular (*Progesterone receptors*).

PRS: Score de riesgo poligénico (*polygenic risk score*).

PTEN: Gen homólogo de la fosfatasa y tensina (*phosphatase and tensin homolog*).

RAD51: Gen RAD51 recombinasa.

RAD51L1: Gen 2 homólogo a RAD51.

RDT: Radioterapia.

RH: Recombinación homóloga.

RS: Referencia específica de los SNPs.

SAP: Fosfatasa alcalina de camarón (*Shrimp Alkaline Phosphatase*).

SAS: Sistema de análisis estadístico (*Statistical Analysis System*).

SKT-11: Serín-treonin quinasa 11, previamente conocida como LKB1.

SNP: Polimorfismo de nucleótido único (*Single Nucleotide Polymorphism*).

SNV: Variación nucleotídica simple de baja frecuencia (*Single Nucleotide Variant*).

TCF7L2: Factor de transcripción 7 similar a 2 (*Transcription Factor 7 Like 2*).

THS: Terapia hormonal sustitutiva.

TNBC: Tumores triple negativo (ER negativo, PR negativo, HER2 negativo).

TNRC9: Gen *trinucleotide-repeat-containing-9*.

TOX3: Gen *TOX High mobility group box family member 3*.

TP53: Gen codificante de proteína tumoral 53 (*tumor protein p53*).

WNT: vía de transducción de señales mediada por proteína WNT (*Wingless e Int*).

## Resumen

El cáncer de mama constituye el tumor maligno más frecuente en las mujeres de todo el mundo. A pesar de la reducción considerable de su mortalidad a 5 años (<20%) asociada al gran avance de las técnicas de diagnóstico precoz y terapéuticas, este tumor sigue siendo un importante problema en la salud de la población que requiere atención focalizada.

Hoy en día sabemos que existen factores epidemiológicos que predisponen al desarrollo de cáncer de mama, así como alteraciones a nivel genético que también aumentan su riesgo. Dentro de estas alteraciones genéticas, las que parecen tener gran relevancia a nivel poblacional son los polimorfismos de nucleótido único o SNPs, cuyo efecto individual es despreciable, pero se torna multiplicativo cuando presentados de forma conjunta, pudiendo llegar a explicar cerca del 20% de la heredabilidad del cáncer de mama.

El objetivo de esta investigación fue analizar la frecuencia de los SNPs que más contribuyen al desarrollo del cáncer de mama en la población gallega e investigar si su combinación en forma de puntuación de riesgo poligénico (PRS) ayuda a estratificar a las mujeres según su riesgo de padecer cáncer de mama. A este efecto se diseñó con el software de Agena Bioscience un panel de SNPs a partir de 77 SNPs de riesgo definidos por Mavaddat et al. en 2015 y se analizó por medio de la tecnología de espectrometría de masas MALDI-TOF su presencia en la población gallega a través de la cohorte BREOGAN. La posible asociación entre los SNPs y el riesgo de cáncer de mama fue valorada utilizando riesgos relativos (OR) e intervalos de confianza del 95% derivados de modelos de regresión logística, ajustados por edad y lugar de estudio. Para evaluar su efecto combinado, creamos un score de riesgo poligénico ponderado para cada participante del estudio, resultado del sumatorio del número de alelos de riesgo de cada SNP multiplicado por el peso de ese SNP (escala logarítmica del OR por alelo derivado).

De los 71 SNPs que pasaron el control de calidad, se observó que 54 presentan unos ORs en la misma dirección que en poblaciones europeas, sin embargo, solo 12 de ellos mostraron nominalmente una asociación significativa estadística ( $p \leq 0.05$ ). Por otra parte, 10 SNPs no denotaron una asociación con el cáncer de mama de manera general, pero sí se demostró una relación estadísticamente significativa en el análisis por subtipos moleculares (3 con tumores ER+, 6 con tumores ER- y 1 con tumores triple negativo), en la misma dirección que estudios previos. A partir de los SNPs con asociación estadísticamente significativa, creamos nuestro PRS, que mostró una asociación creciente entre los ORs para desarrollar cáncer de mama y los percentiles del PRS. Así vemos que las mujeres en el percentil más bajo tienen un OR significativamente menor comparado con las mujeres del percentil más alto (OR=0.17 (0.05-0.50) vs OR=3.86 (1.50-9.94), respectivamente). Estos resultados fueron corroborados por un análisis de validación independiente, que encontró estimaciones de OR similares a las derivadas en este estudio.

Nuestros resultados demuestran la existencia de relación significativa entre determinados SNPs y el cáncer de mama en la población gallega, así como su capacidad de predicción de riesgo cuando son combinados en forma de PRS. Este PRS, ampliado a todos los SNPs conocidos que muestren asociación con el cáncer de mama y usado en combinación con factores

epidemiológicos de riesgo, se perfila como el mejor método de estratificación y predicción de riesgo en la población, tal y como se describe en modelos como BOADICEA (desarrollado para la población inglesa), pudiendo valorarse en un futuro su implantación como proyecto dentro del programa gallego de cribado de cáncer de mama para el seguimiento precoz de las mujeres con alto riesgo estimado.

## Abstract

Breast cancer is the most frequent malignant tumor in women worldwide. Despite the considerable decrease in its 5-year mortality rate (<20%) associated with the constant improvement of early diagnostic and therapeutic techniques, this tumor continues to be a major problem for the population's health that requires focused attention.

Nowadays we know that there are epidemiological factors that predispose to the development of breast cancer, as well as alterations at a genetic level that also increase its risk. Among these genetic alterations, the ones that seem to have the most relevance at a population level are the single nucleotide polymorphisms or SNPs, whose individual effect is disregardable, but becomes multiplicative when presented together, and may explain about 20% of the heritability of breast cancer.

The aim of this research was to analyse the frequency of the SNPs that contribute most to the appearance of breast cancer in the Galician population and to investigate whether their combination in the form of a polygenic risk score (PRS) helps to stratify women according to their risk of suffering from breast cancer. To this effect, a panel of SNPs was designed with the Agena Bioscience software based on 77 risk SNPs defined by Mavaddat et al. in 2015 and their presence in the Galician population was analysed using the MALDI-TOF mass spectrometry technology through the BREOGAN cohort. The possible association between SNPs and breast cancer risk was assessed using odds ratios (OR) and 95% confidence intervals derived from logistic regression models, adjusted by age and place of study. To evaluate their combined effect, we created a weighted polygenic risk score for each study participant, resulting from the addition of the number of risk alleles of each SNP multiplied by the weight of the same SNP (logarithmic scale of the OR by derived allele).

Of the 71 SNPs that were found to pass the quality control, 54 showed ORs in the same direction as they had previously shown in European populations, however, only 12 of these nominally showed a statistically significant association ( $p \leq 0.05$ ). On the other hand, 10 SNPs did not denote an association with breast cancer as a whole, but did reveal a statistically significant association in the analysis by molecular subtype (two with ER+ tumors, six with ER- tumors and one with triple negative tumors), in the same direction as previous studies. Out of the SNPs with statistically significant association, we created our PRS, which showed a growing association between ORs for developing breast cancer and the percentiles of the PRS. Thus, we see that women in the lowest percentile have a significantly lower OR compared to women in the higher percentile (OR=0.17 (0.05-0.50) vs OR=3.86 (1.50-9.94), respectively). These results were corroborated by an independent validation analysis, which found similar OR estimates to those derived in this study.

Our results demonstrate the existence of a significant relationship between certain SNPs and breast cancer in the Galician population and their ability to predict risk when combined in the form of a PRS. This PRS, extended to all known SNPs that have shown association with breast cancer and used in combination with epidemiological risk factors, is revealed as the best method of stratification and risk prediction in the population, as described in models such as BOADICEA (developed for the British population). In the future, its implementation could be considered, as a project within the Galicia breast cancer screening programme for the early follow-up of women with an estimated high risk.

## Resumo

O cancro de mama constitúe o tumor maligno máis frecuente nas mulleres de todo o mundo. A pesar da redución considerable da súa mortalidade a 5 anos (<20%) asociada ao gran avance das técnicas de diagnóstico precoz e terapéuticas, este tumor segue sendo un importante problema na saúde da poboación que require atención focalizada.

Hoxe en día sabemos que existen factores epidemiolóxicos que predispoñen ao desenvolvemento de cancro de mama, así como alteracións a nivel xenético que tamén aumentan o seu risco. Dentro destas alteracións xenéticas, as que parecen ter gran relevancia a nivel poboacional son os polimorfismos de nucleótido único ou SNPs, cuxo efecto individual é despreziable, pero tórnase multiplicativo cando presentados de forma conxunta, podendo chegar a explicar preto do 20% da herdanza do cancro de mama.

O obxectivo desta investigación foi analizar a frecuencia dos SNPs que máis contribúen ao desenvolvemento do cancro de mama na poboación galega e investigar se a súa combinación en forma de puntuación de risco polixénico (PRS) axuda a estratificar ás mulleres segundo o seu risco de padecer cancro de mama. A este efecto deseñouse co software de Agena Bioscience un panel de SNPs a partir de 77 SNPs de risco definidos por Mavaddat et al. en 2015 e analizouse por medio da tecnoloxía de espectrometría de masas MALDI- TOF a súa presenza na poboación galega a través da cohorte BREOGAN. A posible asociación entre os SNPs e o risco de cancro de mama foi valorado utilizando riscos relativos (OR) e intervalos de confianza do 95% derivados de modelos de regresión loxística, axustados por idade e lugar de estudo. Para avaliar o seu efecto combinado, creamos un score de risco polixénico ponderado para cada participante do estudo, resultado do sumatorio do número de alelos de risco de cada SNP multiplicado polo peso dese SNP (escala logarítmica do OR por alelo derivado).

Dos 71 SNPs que pasaron o control de calidade, observouse que 54 presentan uns ORs na mesma dirección que en poboacións europeas, con todo, só 12 deles mostraron nominalmente unha asociación estatisticamente significativa ( $p \leq 0.05$ ). Por outra banda, 10 SNPs non denotaron unha asociación co cancro de mama de maneira xeral, pero si se demostrou unha relación estatisticamente significativa na análise por subtipos moleculares (3 con tumores ER+, 6 con tumores ER- e 1 con tumores triplo negativo), na mesma dirección que estudos previos. A partir dos SNPs con asociación estatisticamente significativa, creamos o noso PRS, que mostrou unha asociación crecente entre os ORs para desenvolver cancro de mama e os percentiles do PRS. Así vemos que as mulleres no percentil máis baixo teñen un OR significativamente menor comparado coas mulleres do percentil máis alto (OR=0.17 (0.05-

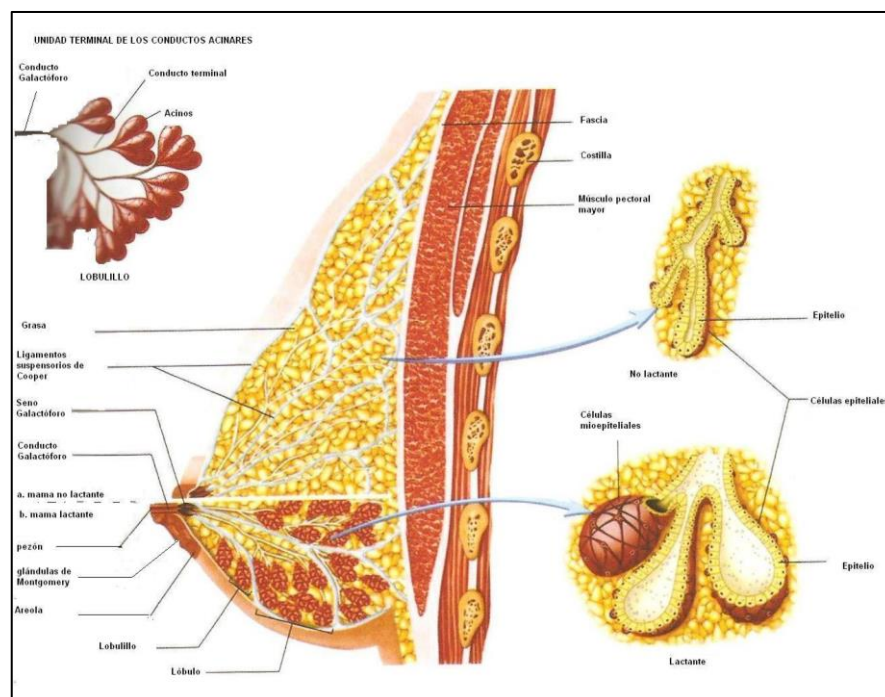
0.50) vs OR=3.86 (1.50-9.94), respectivamente). Estes resultados foron corroborados por unha análise de validación independente, que atopou estimacións de OR similares ás derivadas neste estudo.

Os nosos resultados demostran a existencia de relación significativa entre determinados SNPs e o cancro de mama na poboación galega, así como a súa capacidade de predición de risco cando son combinados en forma de PRS. Este PRS, ampliado a todos os SNPs coñecidos que mostren asociación co cancro de mama e usado en combinación con factores epidemiolóxicos de risco, perfílase como o mellor método de estratificación e predición de risco na poboación, tal e como se describe en modelos como BOADICEA (desenvolvido para a poboación inglesa), podendo valorarse nun futuro a súa implantación como proxecto dentro do programa galego de cribado de cancro de mama para o seguimento precoz das mulleres con alto risco estimado.

## 1-Introducción

### 1.1-GENERALIDADES DE LA GLÁNDULA MAMARIA

Denominamos mamas a las estructuras bilaterales situadas en la región anterosuperior del tórax. Si bien se desarrollan de igual forma tanto en hombres como en mujeres durante la infancia, la llegada de la pubertad determina un gran crecimiento de la mama en la mujer, dependiente del estallido de hormonas sexuales que ocurre durante este periodo. Durante toda su vida fértil, las mamas se verán influenciadas por el efecto de estas hormonas (estrógeno y progesterona principalmente), ya sea por los ciclos menstruales o durante el embarazo y lactancia. La entrada en la menopausia determina el inicio de su atrofia, de la mano del cese de estímulos hormonales característicos de este periodo vital.



**Figura 1.** Estructura del tejido mamario  
(1)

Las mamas se organizan principalmente en dos tipos de tejidos: tejido fibroadiposo, que constituye alrededor del 90% del volumen del órgano en mujeres no lactantes y que sirve de sostén; y tejido glandular, responsable de la producción de la leche materna. Cada mama consta de alrededor de 15-20 lóbulos mamarios, unidades glandulares independientes formadas por glándulas ducto-acinares (Figura 1). La secreción que producen dichas glándulas es vertida a ductos que irán confluyendo en otros cada vez de mayor calibre hasta llegar a los conductos galactóforos, situados justo detrás del pezón, que desembocan al exterior, permitiendo así la salida de la leche durante la lactancia (2).

## 1.2-EL CÁNCER DE MAMA. SUBTIPOS.

Los órganos son unidades estructurales que realizan una función determinada dentro del cuerpo humano. Cada órgano está constituido a su vez por uno o más tejidos, que son agrupaciones de células ordenadas regularmente, que trabajan de forma coordinada y con un comportamiento fisiológico común. El mantenimiento de la integridad de estos tejidos se consigue gracias a un adecuado equilibrio entre la proliferación celular y la muerte celular programada. Llamamos tumor a la proliferación tisular anómala y descontrolada, resultado de un acúmulo de alteraciones en el material genético (ADN) de las células que lo componen, de forma que se bloquea el proceso de regulación del ciclo celular. Esta proliferación de células puede ser autolimitada y no progresar hacia tejidos adyacentes u otras regiones corporales, constituyendo lo que conocemos como tumor benigno. Cuando el crecimiento del tumor es descontrolado e invade otros tejidos sanos del organismo (ya contiguos a él, ya a distancia) entonces es cuando hablamos de un tumor maligno. Este proceso de desregulación del crecimiento de los tejidos resultado de alteraciones en el material genético de las células que lo componen se conoce como oncogénesis, y puede suceder en cualquier tipo de tejido. Los tumores se engloban en dos categorías principales según el tipo de tejido del que provienen: carcinomas, tumores que nacen a partir de epitelios (tejidos que recubren las superficies tanto externas como internas del cuerpo) y sarcomas, que surgen de los tejidos de conjunción, que dan soporte al organismo.

En la mama el cáncer no constituye una entidad única, si no que podemos distinguir tantos tipos de cáncer de mama como estirpes celulares puedan originarlo. La gran mayoría de los tumores de mama son carcinomas, derivados de las células epiteliales que constituyen los ductos y acinos que forman las glándulas mamarias, aunque también podemos encontrar sarcomas originados a partir del tejido fibroadiposo de sostén. Dada la elevada proporción de carcinomas respecto a los sarcomas, nos centraremos en la descripción de los primeros.

Atendiendo a la localización del carcinoma, podemos diferenciar dos tipos de alteraciones: aquellas lesiones cancerígenas **in situ** y aquellas que son **invasivas**. El carcinoma in situ se caracteriza porque la proliferación celular, aunque descontrolada y atípica, se encuentra confinada al mismo epitelio (no sobrepasa la membrana basal) de forma que aún no afecta a los tejidos subyacentes. En el carcinoma invasivo, por contra, las células malignas han proliferado suficiente como para sobrepasar esta membrana basal e infiltrar los tejidos que rodean a los ductos y acinos (tejido de sostén, vasos sanguíneos...) (3).

### 1.2.1 Subtipos histológicos

Subtipos principales de carcinomas infiltrantes de mama en función de su histología:

- Carcinoma ductal infiltrante (CDI): tipo de tumor de mama más común, supone aproximadamente el 90% de todos los casos de cáncer de mama. Se origina a partir de las células que recubren el interior de los conductos que llevan la leche materna desde las glándulas mamarias hasta el pezón. Pueden crecer de forma bien diferenciada, formando túbulos y con escasas atipias celulares a pesar del comportamiento invasivo, o como tumores poco diferenciados, formando masas sólidas que no se asemejan a ninguna estructura histológica normal y con marcado pleomorfismo.

- Carcinoma lobulillar infiltrante (CLI): representa cerca del 5% de los casos de cáncer de seno, y surgen de la malignización de las células de los acinos glandulares mamarios. Suelen presentar un patrón de infiltración en células aisladas o en grupos laxos, con escasas atipias, pero su pronóstico tiende a no ser muy bueno, al presentarse con relativa frecuencia en formas multicéntricas (más de un foco cancerígeno en la misma mama) y/o bilaterales (en ambos senos), además de tener alta incidencia de afectación ganglionar precoz.
- Carcinoma medular (CM): masas bien circunscritas formadas por células pobremente diferenciadas y una periferia donde se observa una infiltración linfocitaria importante (respuesta linfoide del organismo ante la proliferación tumoral). Suponen el 5% de los tumores de mama, y tienen un pronóstico relativamente bueno.
- Carcinoma coloide o mucinoso: aunque lo diferenciamos como un tipo de carcinoma diferente, se trata de una forma poco frecuente de carcinoma ductal invasivo, que supone sobre un 2%-3% del total de tumores malignos de mama. Se caracteriza por un gran acúmulo de mucina que producen las células tumorales y en el que se encuentran sumergidas. Suele afectar a mujeres postmenopáusicas y tiene un pronóstico relativamente bueno.
- Carcinoma tubular: tumor histológicamente constituido por túbulos bien formados revestidos por una sola capa de células bien diferenciadas, aunque sin membrana basal. Representan un 2% de todos los casos de cáncer de mama y, pese a ser de buen pronóstico, suponen un reto diagnóstico al poder llegar a confundirse con proliferaciones benignas.
- Enfermedad de Paget de la mama: se presenta en el pezón como una placa eritematosa y costrosa, ocasionalmente pruriginosa, que simula un eczema. Realmente se trata de la expresión cutánea de un carcinoma ductal (in situ) en el que las células malignas migran por el sistema canalicular hacia el pezón, llegando hasta la piel.
- Carcinoma inflamatorio: forma de presentación infrecuente y muy agresiva de cáncer de mama. La mama se caracteriza por ser dolorosa, con una apariencia edematosa, eritematosa e indurada (piel de naranja). Está causado por la infiltración de las células cancerígenas de los vasos linfáticos de la piel de forma que bloquean su drenaje.

### 1.2.2 Subtipos moleculares

La clasificación histológica del cáncer de mama hoy en día no es la única con relevancia clínica. Cerca del 90% de los tumores de mama caen dentro del subtipo de carcinoma ductal según su histología, pero pueden llegar a mostrar entre sí una gran variabilidad respecto a morfología tumoral, mecanismos moleculares y pronóstico (4). Dado que el ideal a la hora de manejar al paciente oncológico es el conocimiento preciso de la mecánica tumoral que permita un tratamiento específico dirigido a las alteraciones moleculares subyacentes, en las últimas décadas se hace patente la necesidad de un análisis de la expresión génica diferencial en los tumores de mama que permita conocer la causa de la heterogeneidad en el tratamiento y pronóstico de los tumores de mama que la clasificación histológica no expone. El artículo

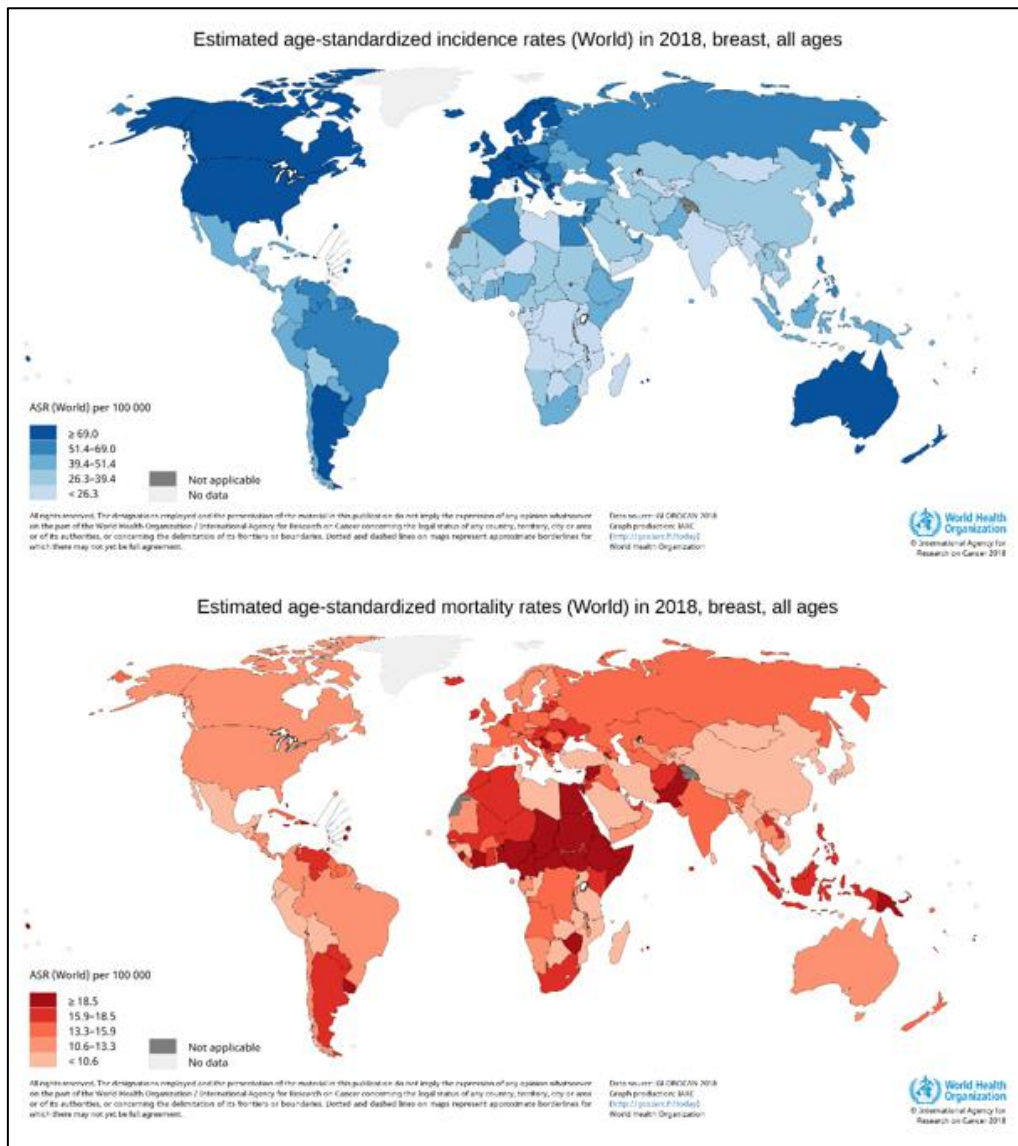
pionero que llevó a la identificación de los 5 subtipos moleculares de cáncer de mama que conocemos hoy en día fue publicado en el año 2000 por el grupo de Perou et al. (5). Estudios posteriores muestran que además la caracterización de cada subtipo determina diferencias en el pronóstico vital del paciente que lo presenta. En la Tabla 1 se recogen las características determinantes de los subtipos moleculares de mama que se definen en la actualidad.

SUBTIPOS	EXPRESIÓN GÉNICA
<b>Luminal A</b>	Receptores de Estrógenos (ER) + Receptores de Progesterona (PR) + ( $\geq 20$ ) Receptor 2 de factor de crecimiento epidérmico humano (HER2) - Índice de proliferación bajo Ki67 < 20 Subtipo más frecuente (67%) Buen pronóstico, riesgo de recurrencia bajo
<b>Luminal B</b>	ER+ PR < 20/- HER2 +/- Ki67 > 20, índice de proliferación alto Peor pronóstico que el luminal A Riesgo de recurrencia alto
<b>HER2+</b>	ER- PR- HER2+ (sobre expresado o amplificado) Cerca del 25% de los tumores malignos de mama son HER2+. Subtipo molecular de crecimiento rápido y por tanto agresivo, pero el desarrollo de tratamientos específicos altamente efectivos contra la proteína HER2 han mejorado su tasa de supervivencia.
<b>TNBC (triple negativo)</b>	ER- PR- HER2- Expresa genes asociados a células mioepiteliales cuyo inmunofenotipo corresponde a CK5, CK17, c-kit (receptor con actividad tirosin-quinasa), factores de crecimiento de hepatocito e insulina, calponina 1, caveolina y laminina. Es el subtipo más agresivo.
<b>Normal-like</b>	Con frecuencia ER -, PR -, HER2- Gran expresión de genes característicos de células epiteliales basales y adiposas y escasa expresión de genes de células epiteliales lumbinales.

**Tabla 1.** Subtipos moleculares de cáncer de mama (modificado de (6))

### 1.3-EPIDEMIOLOGÍA DEL CÁNCER DE MAMA

El cáncer de mama es el segundo cáncer más frecuente en el mundo por detrás del cáncer de pulmón, pero sin embargo constituye el tumor maligno más frecuente con diferencia en las mujeres de todo el mundo, siendo responsable de un 24% de todos los nuevos casos de cáncer en el 2018. (7) La incidencia varían enormemente de unas zonas geográficas a otras, encontrándose las tasas más bajas en el Sudeste Asiático, África subsahariana y Centroamérica (con una incidencia de menos de 40 casos por 100,000) y las más altas en Australia/Nueva Zelanda, Oeste y Norte de Europa y Norteamérica (con una incidencia de más de 80 casos por 100,000). Si bien, como muestran los datos (Figura 2), se trata de un cáncer propio de los países desarrollados, hoy en día su incidencia está en ascenso en las regiones en vías de desarrollo y es precisamente en estos lugares donde su tasa de mortalidad es mayor, resultado del escaso acceso a una sanidad que permita un diagnóstico y tratamiento óptimos.



**Figura 2. Incidencia (en azul) y mortalidad (en rojo) del cáncer de mama en el mundo en el año 2018 (8)**

En España, siguiendo la tendencia global, el cáncer de mama es el tumor más frecuente en mujeres. Según datos del Observatorio del Cáncer AECC (Asociación Española Contra el Cáncer) (7,9) en el 2019 se detectaron 33,307 nuevos casos de cáncer de mama en nuestro país (139 casos por 100.000 mujeres), lo que supone el 12% del total de nuevos casos oncológicos detectados dicho año en ambos sexos y cerca del 30% de todos los tumores malignos en mujeres. El diagnóstico ocurre mayoritariamente en mujeres dentro de la franja de edad de los 35 a los 85 años, con un pico máximo entre los 45 y los 65 años. Desglosando por comunidades autónomas, existe una variación notable en las distintas tasas de incidencia, encontrándose Galicia ligeramente por encima de la media nacional con 156 casos por cada 100.000 mujeres. Esta mayor proporción de mujeres afectadas se pone en relación con el envejecimiento de la población gallega en comparación con otras poblaciones más jóvenes (como por ejemplo en la comunidad murciana, con 126 casos por cada 100.000 mujeres).

El cáncer de mama constituye la primera causa de muerte oncológica entre las mujeres del mundo (14% de todas las defunciones por tumores malignos). Considerando ambos sexos, solo es superado en mortalidad por los cánceres de pulmón, estómago, colorrectal e hígado. España cuenta con una de las tasas de mortalidad por cáncer de mama más bajas de Europa, con 28 fallecimientos por cada 100.000 mujeres. A pesar de ello, sigue suponiendo el 16% de todos los fallecimientos femeninos por cáncer en nuestro país, y el 3,3% del total de muertes entre las mujeres (10). Afortunadamente, las mejoras en diagnóstico y terapéutica han conseguido disminuir la mortalidad a menos del 20% a los 5 años del diagnóstico. Aun así, sigue siendo un problema muy importante en la salud de la población, por lo que en el 2003 la Unión Europea establece la necesidad de implantar un protocolo de cribado poblacional del cáncer de mama. Este se lleva a cabo mediante la realización de una mamografía cada dos años a las mujeres entre los 50 y 69 años, rango de edad donde se ha confirmado el efecto protector del diagnóstico por técnica de imagen (11).

#### **1.4-FACTORES DE RIESGO DEL CÁNCER DE MAMA**

A pesar de que en la mayoría de los casos de cáncer de mama la etiología no llega a conocerse, considerando la importancia a nivel global que tiene este cáncer en la salud de la población, se han llevado a cabo múltiples estudios para intentar determinar qué elementos predisponen a las mujeres al desarrollo de dicho cáncer. Gracias a ellos, hoy en día se conocen una gran diversidad de variables que en conjunto pueden llegar a predisponer al desarrollo del cáncer de mama en aquellas mujeres que los reúnen. Para una mejor comprensión, dividiremos estos factores de riesgo en modificables y no modificables, que pueden a su vez ser genéticos o no genéticos.

##### **1.4.1-Factores de riesgo no modificables:**

###### **1.4.1.1-Sexo:**

El 99% de los cánceres de mama aparecen en mujeres. Si bien el cáncer de mama puede darse en los varones, la gran actividad hormonal a la que se ven sometidas las mujeres durante su vida fértil las hace mucho más propensas al desarrollo de este tumor.

#### 1.4.1.2-Edad:

Igual que muchos otros tipos de cáncer, la incidencia del cáncer de mama aumenta proporcionalmente con la edad (12). El envejecimiento se asocia fuertemente con la aparición de nuevos casos, relacionado con la acumulación de mutaciones, empeoramiento de los mecanismos de control celular y exposición acumulada a otros factores de riesgo. En todo el mundo las incidencias se incrementan notablemente a partir de los 40 años, con un pico máximo en los 60 años. En España de acuerdo con estos datos, la franja de mayor incidencia de cáncer de mama ocurre entre los 45 y los 65 años de edad, coincidiendo con las fases finales de la etapa reproductiva femenina (perimenopausia).

#### 1.4.1.3-Ancestralidad:

Existen disparidades notables tanto en la incidencia como en la supervivencia a 5 años de las mujeres afectas de cáncer de mama según su etnia. Se ha visto que las mujeres europeas constituyen el grupo poblacional más propenso a desarrollar cáncer de mama, mientras que las mujeres afroamericanas tienen las tasas de mortalidad a 5 años más elevadas, independientemente del estadio del tumor al diagnóstico. Según datos recogidos en Estados Unidos por la Sociedad Americana contra el Cáncer, en el 2012 la tasa de mortalidad en mujeres afroamericanas eran un 42% superior a la registrada en las mujeres europeas. Contrastando con estos datos, las poblaciones asiáticas parecen ser las que cuentan con la menor incidencia y mortalidad por cáncer de mama (13).

#### 1.4.1.4-Vida fértil prolongada:

Dado que los estrógenos son los principales encargados de la estimulación del crecimiento de las células mamarias, una exposición prolongada a dichas hormonas aumentaría el riesgo de que, durante estos ciclos de crecimiento e involución celular, algunas células puedan mutar y tornarse cancerígenas. Por tanto, la menarquia precoz (primera menstruación antes de los 12 años) y la menopausia tardía (última menstruación pasados los 55 años) serían factores de riesgo para el cáncer de mama. Los datos corroboran estas premisas, pues se ha visto que tanto la menarquia precoz como la menopausia tardía suponen, para la mujer que las presenta, un riesgo superior a la media (20% y 12% respectivamente) de desarrollar cáncer de mama (14).

#### 1.4.1.5-Densidad mamaria:

Un tejido mamario denso se asocia con mayor riesgo de cáncer de mama. Comparado con mujeres con densidad mamaria de menos del 5%, aquellas con densidades mamarias de 50%-75% y de más del 75% presentan un riesgo aproximadamente 3 y 6 veces mayor, respectivamente. (15) La densidad del tejido mamario refleja la proporción de tejido glandular con respecto al tejido fibroadiposo de la mama. Considerando que la variedad de cáncer de mama más común (el carcinoma ductal de mama) se origina a partir de las células epiteliales de los conductos glandulares, una alta proporción de tejido glandular supondría un mayor número de células susceptibles de mutagénesis y evolución maligna. Además, este aumento de tejido glandular supone una dificultad añadida para el diagnóstico precoz de cáncer por medio de mamografías (técnica de imagen utilizada de forma habitual en el screening poblacional), por la difícil interpretación de las imágenes que se obtienen.

#### 1.4.1.6-Alteraciones mamarias benignas/premalignas:

No toda la patología benigna de mama implica un mayor riesgo de cáncer, pero sí está demostrado un riesgo dos veces mayor en el caso de las lesiones mamarias proliferativas, especialmente aquellas que presentan atipias celulares (hiperplasia ductal atípica e hiperplasia lobulillar atípica) (16). Por otra parte, las mujeres que ya han sido diagnosticadas de una alteración histológica precursora de cáncer de mama como son el carcinoma ductal in situ o el carcinoma lobulillar in situ tienen, si no reciben tratamiento alguno, sobre 9 y 10 veces (respectivamente) más riesgo de desarrollar un cáncer de mama invasivo en comparación con las mujeres sin este diagnóstico (14).

#### 1.4.1.7-Antecedentes personales de cáncer de mama:

Las mujeres con una historia previa de cáncer de mama tienen un riesgo aumentado de desarrollar algunos tipos de cáncer, como son, entre otros, un segundo cáncer de seno, cáncer de esófago, cáncer de pulmón, cáncer de útero o cáncer de tiroides. El cáncer de mama al que nos referimos no se trataría de una recurrencia del cáncer primario, si no de un nuevo cáncer de mama independiente del inicial, y que puede desarrollarse en la misma mama que la primera ocasión o en la contralateral. Este incremento del riesgo ocurre especialmente en las mujeres menores de 50 años al primer diagnóstico (25-44 años), equiparándose al de la población normal en las mujeres que rondan los 65 años al diagnóstico (17).

#### 1.4.1.8-Antecedentes familiares de cáncer de mama:

El riesgo de cáncer de mama es más alto en mujeres con familiares afectados por este tumor maligno, condicionado por la carga genética heredada de los progenitores. Este incremento de riesgo es mayor cuanto más temprana es la aparición del cáncer y más cercano el grado de parentesco del familiar, así como del número de familiares afectados. Comparado con mujeres sin familiares conocidos afectados, el riesgo para uno, dos y tres o más familiares de primer grado con cáncer de seno es de 1.80 (Intervalo de confianza, CI, al 99% de 1.69-1.91), 2.93 (2.36-3.64), y 3.90 (2.03-7.49), respectivamente (18). Es en este apartado y en la inclusión de factores genéticos como marcadores de riesgo para el ajuste personalizado del screening de cáncer de mama en la población, en lo que se centrará este trabajo.

### **1.4.2-Factores de riesgo modificables**

#### 1.4.2.1-Nivel socioeconómico:

Cómo se puede observar a partir de los datos epidemiológicos del cáncer de mama a nivel mundial, su incidencia es mayor en países con un Índice de desarrollo humano (IDH) más alto. Este índice es un parámetro elaborado por el Programa de las Naciones Unidas para el Desarrollo que pretende indicar de forma sintética el grado de desarrollo humano por país, entendido en las dimensiones de salud, educación y riqueza. Se elabora a partir de los datos de esperanza de vida al nacer, el promedio de escolarización de la población y la renta per cápita de cada país (19). Por otro lado, se ha visto que la mortalidad del cáncer de mama es significativamente más alta en los países con un IDH bajo e incluso medio en comparación con aquellos con índices de desarrollo humano muy elevado (y por tanto incidencias muy altas).

#### 1.4.2.2-Lugar de residencia:

Existe disparidad en los resultados de los estudios que intentan valorar la relación entre el tipo de ambiente residencial y la incidencia de cáncer de mama. Aunque parece que en la mayoría de los casos se apunta discretamente a una mayor incidencia de cáncer de mama en las mujeres que viven en ambientes urbanos respecto a las del rural, esta relación estaría mediada por el aumento del nivel socioeconómico y la mayor densidad de médicos de atención primaria (capaces de orientar hacia un diagnóstico) que supone la vida en la ciudad y no tanto por una relación causal directa (20).

#### 1.4.2.3- Hábitos reproductivos:

Existe una fuerte relación entre la edad de la mujer al primer embarazo a término y su riesgo de desarrollar cáncer de mama. Se estima que las mujeres nulíparas o que tienen su primera gestación a término con más de 30 años tienen cerca de 3 veces más riesgo de desarrollar cáncer de mama a largo plazo que las mujeres que tienen su primer hijo antes de los 18 años (21). Además, se observa una importante disminución del riesgo a largo plazo cuanto mayor sea el número de embarazos y menor el tiempo entre gestaciones (paridad elevada). Aunque se sabe que existe un aumento transitorio del riesgo de cáncer de mama postparto en gestantes mayores de 25 años, el riesgo acumulado a largo plazo sigue resultando menor en las mujeres multíparas (22). Por otra parte, diversas investigaciones han demostrado que la lactancia materna disminuye notablemente el riesgo de cáncer de seno entre las mujeres que han tenido hijos. En concreto, un importante estudio en 2002 con cerca de 50.000 casos de cáncer de mama describió una reducción de un 4,3% del riesgo por cada 12 meses que una mujer da de mamar (23). En la misma línea, un gran metaanálisis en 2013 de la evidencia recogida hasta ese momento, describe una disminución de un 14% del riesgo de cáncer de mama en mujeres lactantes en comparación con aquellas que nunca lo hicieron (24), remarcando así el importante efecto protector de la lactancia materna.

#### 1.4.2.4- Ejercicio físico e índice de masa corporal:

Especialmente en las mujeres postmenopáusicas, un índice de masa corporal (IMC) elevado se asocia con un aumento en el riesgo del cáncer de mama. La mayor cantidad de tejido graso, principal fuente de estrógenos en la menopausia tras el fallo ovárico correspondiente, condiciona un alto nivel de estrógenos en la mujer, que como ya se ha descrito es responsable de una mayor probabilidad de padecer cáncer de mama por su efecto sobre el tejido mamario. Los datos avalan esta conclusión, al demostrarse en postmenopáusicas con  $IMC \geq 28 \text{ kg/m}^2$  un riesgo 1,26 veces mayor, comparado al de las postmenopáusicas con un peso corporal adecuado (25). Por contrapartida, se ha encontrado en diversos estudios una relación inversa entre el peso y el cáncer de mama en la etapa fértil de la mujer, sugiriendo que la obesidad pueda tener un cierto efecto protector frente al cáncer de mama en la etapa premenopáusica. Esta aparente incongruencia podría explicarse por la falta de ovulación (y por tanto bajo nivel de estrógenos) que experimentan frecuentemente las mujeres con sobrepeso/obesidad (ej. Síndrome de Ovario Poliquístico). Siguiendo el enfoque expuesto hasta ahora, el ejercicio físico será un factor protector importante al reducir la cantidad de grasa corporal y a su vez los niveles de estrógenos en el organismo. La literatura así lo confirma, indicando una reducción de un 10%-25% del riesgo de cáncer de mama en aquellas mujeres que andan una media de 7 horas a la semana, con un efecto protector aún mayor en la menopausia. Además de esta reducción del riesgo, se ha descrito también una reducción de un 26-40% de la mortalidad y recurrencia de cáncer de

mama entre las mujeres ya diagnosticadas que realizan ejercicio en comparación con sus semejantes que llevan una vida más sedentaria (26).

#### 1.4.2.5-Hábitos tóxicos, alcohol y tabaco:

Es ampliamente conocido el papel del alcohol en el desarrollo de diversos tipos de cáncer, como los de cabeza y cuello, esófago o hígado. Numerosos estudios han encontrado también una relación entre el consumo de alcohol y el incremento del riesgo de cáncer de mama. Por cada 10 gramos de alcohol/día se calcula un incremento del riesgo de un 7%-10%, tanto en mujeres premenopáusicas como postmenopáusicas. Parece tratarse de una relación dosis-dependiente, aunque algunos estudios hablan más de un incremento del riesgo por consumo acumulado durante largos periodos de tiempo, y otros lo asocian al consumo excesivo en breves periodos de tiempo (*binge drinking*, consumo de más de 4/5 bebidas alcohólicas en un periodo de dos horas). Sea como fuere, la relación entre el hábito alcohólico y el cáncer de mama podría explicarse por la capacidad del alcohol para elevar los niveles de estrógeno en sangre (27). Si bien la relación del alcohol con el cáncer de mama ha sido de sobra probada, la asociación entre el tabaco y el aumento del riesgo de cáncer de mama aún no posee una evidencia excesivamente fuerte a su favor, encontrándose bastante heterogeneidad en los resultados de los distintos estudios, si bien un gran metaanálisis en el 2015 de la evidencia disponible hasta el momento apoya la existencia de relación entre el hábito tabáquico y un moderado incremento del riesgo de cáncer de mama. El tabaquismo pasivo también parece estar asociado con un moderado aumento del riesgo, pero la fuerza de estos resultados es un poco menor, dada la dificultad de evaluar la exposición ambiental al tabaco y los escasos estudios de los que se dispone (28).

#### 1.4.2.6-Exposición a estrógenos exógenos:

Con la menopausia y la caída fisiológica de los niveles hormonales, las mujeres experimentan una serie de síntomas y signos característicos a corto, medio y largo plazo, entre los que destacan los síntomas neurovegetativos como los sofocos, la atrofia urogenital y osteoporosis por pérdida de masa ósea. Para paliar estos síntomas, se puede recurrir a la utilización de terapia hormonal sustitutiva, THS, a pesar de que su uso esté sujeto a debate, al tratarse la menopausia de un proceso fisiológico y no patológico. La THS se puede usar en dos pautas, monoterapia estrogénica o en un tratamiento combinado de estrógenos asociados a gestágenos. El uso de THS en pauta combinada aumenta el riesgo de desarrollar cáncer de seno. Este riesgo se incrementa de forma directamente proporcional con el tiempo de uso de la terapia y con lo temprano de su implementación desde el inicio de la menopausia. Este riesgo parece, sin embargo, que es solo transitorio, reduciéndose e igualándose al de las mujeres postmenopáusicas sin uso de THS combinada tras 5 años de cese del tratamiento (14). Los resultados sobre la relación entre el cáncer de mama y la THS exclusivamente con estrógenos son muy dispares entre los distintos estudios, aunque sí se ha confirmado que incrementa el riesgo de cáncer de endometrio en sus usuarias (29). La exposición a hormonas exógenas como el dietilestilbestrol, fármaco usado a mediados del siglo pasado como antiabortivo, también se ha demostrado que aumenta el riesgo de cáncer de mama. En el caso del dietilestilbestrol, los estudios demostraron un aumento en la incidencia de cáncer de mama especialmente después de los 40 años tanto en las mujeres que lo recibieron como en las hijas expuestas al fármaco intraútero (30).

#### 1.4.2.7-Exposición a radiaciones:

Las mujeres tratadas con radioterapia (RDT) en la edad pediátrica y adolescencia, especialmente con terapias localizadas en la región torácica (modalidad típicamente utilizada en los esquemas terapéuticos de linfomas de Hodgkin mediastínicos) tienen un riesgo elevado de desarrollar cáncer de mama. Este incremento del riesgo aparece muy pronto (a los 8 años post-radiación), y no se estabiliza con el paso de los años. La incidencia acumulada de cáncer de mama en estas mujeres cuando alcanzan los 40-45 años, oscila entre el 13% y 20%, siendo llamativamente mayor a la que presenta la población general de mujeres a esa edad (incidencia acumulada de solo el 1%) (31).

### 1.5-CARGA GENÉTICA DEL CÁNCER DE MAMA

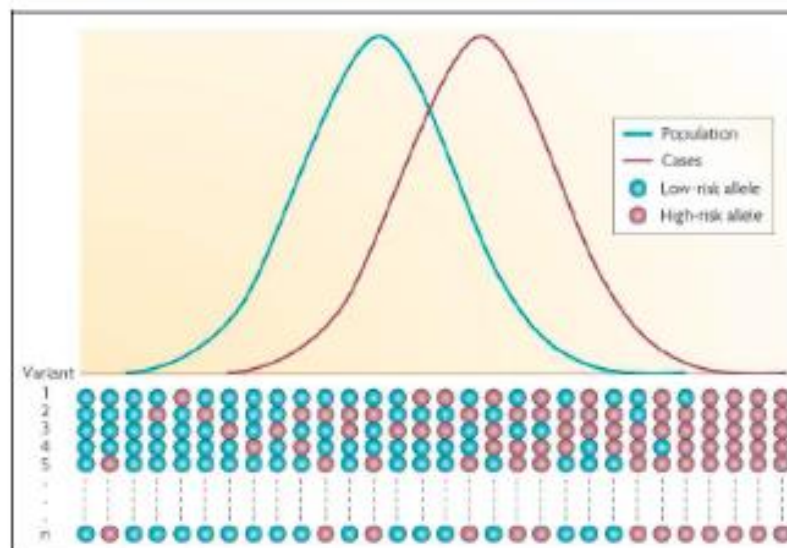
Como se ha mencionado previamente, uno de los factores de riesgo más importantes para el desarrollo de cáncer de mama es la presencia de historia familiar. Se estima que entre 5% y 10% de todos los casos de cáncer de seno se deben al carácter hereditario de mutaciones en línea germinal de genes de transmisión autosómica dominante. Se detectan en familias con casos de cáncer de mama en todas las generaciones, caracterizados por una presentación a edades tempranas, en ocasiones en forma de tumores bilaterales y que pueden asociarse a otros tumores malignos, el más frecuente el cáncer de ovario. Por otro lado, aproximadamente del 15% al 20% de los casos de cáncer de mama diagnosticados se presentan en mujeres cuya historia familiar es significativa (dos o más parientes de primer o segundo grado afectos, a edades más elevadas que en el cáncer de mama hereditario). Estas personas tienen un riesgo más elevado de padecer la enfermedad, sin embargo, no existe un patrón de herencia bien definido, ni un gen concreto mutado que pueda ser responsabilizado por la predisposición al cáncer. No se puede hablar de cáncer de mama hereditario en sentido estricto, y se denomina cáncer de mama de agregación familiar (32).

La totalidad de la heredabilidad del cáncer de mama se estima que es alrededor del 30%, un poco inferior a la del cáncer colorrectal(33), y entre el 5% y el 10% corresponden a formas familiares.

Tras años de estudios de familias afectas, dificultados por la concurrencia de casos de cáncer de mama esporádicos (la forma de presentación más frecuente) y hereditarios dentro de la misma familia, así como la variabilidad genética de la enfermedad, se consigue la identificación en 1990 y 1994 respectivamente de los genes BRCA1 (*BR*east *C*ancer *s*usceptibility *1*, MIM 113705) (34) y BCRA2 (*BR*east *C*ancer *s*usceptibility *2*, MIM 600185) (35). Aproximadamente el 30% de los casos de cáncer de mama hereditarios son atribuidos a mutaciones en línea germinal de estos dos genes de alta penetrancia. Sin embargo, la evolución de la tecnología de secuenciación de ADN ha permitido el estudio a gran escala en búsqueda de otros genes de predisposición al cáncer de mama que expliquen los casos de cáncer hereditario que BRCA1 y BRCA2 no justifican. Así, han sido identificados nuevos genes de susceptibilidad tanto de alta penetrancia (TP53, PTEN, SKT-11) como moderada penetrancia (ATM, CHEK2, PALB2). A pesar de la relevancia individual que pueda suponer a la mujer que presente mutaciones en estos genes, que además asocian no solo cáncer de mama si no todo un conglomerado sindrómico, aparecen de forma tan poco frecuente en la población que solo justifican una proporción muy pequeña de todos los casos de cáncer de mama (entre 2% y 5%) (36). En algunas poblaciones, como judíos askenazis, encontramos una frecuencia incrementada de algunas variantes por efecto fundador (34). Lo mismo ocurre en la población

gallega, donde la mutación en BRCA1, **c.211G>A** representa más algo más del 50% del cáncer de mama familiar en nuestra comunidad (37).

Esta situación lleva a la propuesta presentada por el equipo de Antoniou AC et al. (38) de un modelo de riesgo poligénico. Este modelo sugiere que la presencia conjunta de múltiples variantes genéticas comunes cuyo efecto individual es mínimo (baja penetrancia) tenga un efecto multiplicativo sobre el riesgo de cáncer de mama y sean por tanto las responsables del cáncer de mama de agregación familiar en las familias BRCA1/2 negativas (39).



**Figura 3.** Modelo poligénico de susceptibilidad al cáncer (40)

Actualmente la idea planteada por Antoniou AC et al. se ha impuesto como la dominante, de forma que se establece que la predisposición heredada al cáncer de mama depende de la participación de tres categorías de alelos de susceptibilidad, en función del riesgo relativo que su presencia confiere:

- Genes de alta penetrancia: Confieren un riesgo relativo elevado de padecer cáncer de mama (mayor al 5%). En la actualidad, los genes que dominan esta categoría son BRCA1 y BCRA2, que como ya hemos descrito asocian la mayor proporción de cáncer de mama y/o ovario hereditario (cerca del 25%). Estudios posteriores demuestran la existencia de otros genes, como son TP53, PTEN y SKT-11, cuya mutación también confiere un elevado riesgo de cáncer de mama (entre otros tumores), pero cuya frecuencia es tan baja que justifican muy pocos casos de cáncer de mama en la población.
- Genes de moderada penetrancia: identificados a partir de su interacción con los genes BRCA1 y BRCA2 o por su participación en las mismas vías de reparación del ADN. Su mutación confiere un riesgo moderado de padecer cáncer de mama (entre el 1'5% y el 5% de riesgo relativo). Destacamos ATM, CHEK2 y PALB2. Son poco frecuentes en la población (<1%), y justifican sobre un 2%-5% de los casos de cáncer de mama hereditarios.
- Genes de baja penetrancia: Identificados recientemente mediante GWAs (Estudios de asociación de genoma completo). Se trata de polimorfismos de un solo nucleótido (SNPs) que se encuentran de forma relativamente frecuente en la población general

(frecuencia entre el 5% y 50%) y que de forma individual suponen un escaso incremento del riesgo de cáncer de mama, pero cuyo efecto sobre el riesgo es multiplicativo cuando se presentan de forma conjunta. Pueden llegar a explicar hasta un 18% de la heredabilidad(41). Para poder hablar de un SNP, la variante debe darse en al menos un 1% de la población.

A continuación, hablaremos brevemente de los principales genes dentro de cada categoría:

### **1.5.1-Genes de alta penetrancia**

#### **1.5.1.1-BRCA1 y BRCA2:**

BRCA1 y BCRA2 son dos genes supresores de tumores localizados en los cromosomas 17q21 y 13q12, respectivamente. Las proteínas BRCA1 y BRCA2 no mutadas intervienen en el mantenimiento de la estabilidad del genoma a través de múltiples procesos celulares. El principal mecanismo es la recombinación homóloga (RH), sistema que permite la reparación de las roturas en la doble hebra de ADN de una forma exacta en las células en proliferación (se utiliza la cromátide hermana intacta como modelo complementario para reparar la rotura, conservando la secuencia nucleotídica indemne). BRCA1 también está implicado en la reparación de las roturas dobles de ADN por medio de la unión de extremos no homólogos, mecanismo que modifica y une cada una de las cadenas de ADN sin tener en consideración la homología de las secuencias de sus extremos, generando así pequeñas deleciones o inserciones, lo cual supone una alta propensión a generar errores (42). Si BRCA1/2 se encuentran mutados y por tanto la célula carece de las proteínas correspondientes o estas no son funcionales, la reparación de la doble hebra de ADN puede no ocurrir o realizarse de forma incorrecta, pudiendo resultar en pérdida masiva de información genética y/o reordenamientos que supongan una transformación neoplásica e incluso la muerte celular. La herencia por parte de un progenitor de BRCA1/2 mutado no implica directamente neoplasia, pues el alelo sano heredado del otro progenitor es suficiente para mantener la funcionalidad de las proteínas BRCA1/2. Es necesario una segunda agresión al genoma que suponga la pérdida del alelo sano para el desarrollo de la neoplasia. Por otra parte, considerando que estos dos genes se expresan en casi todas las células del cuerpo, resulta destacable que, sin embargo, su mutación cause de forma característica tumores en tejidos específicos (mama, ovario los principales). Aunque esta relación sigue siendo estudiada hoy en día, estudios ya han sugerido el papel de BRCA1 en la morfogénesis ductal del pecho en desarrollo, así como una posible función de BRCA1/2 como reguladores del crecimiento celular inducido por estrógenos. (43).

#### **1.5.1.2-TP53:**

Gen supresor de tumores de gran importancia, localizado en el cromosoma 17p13.1, que se conoce como “el guardián del genoma” por su importante papel en la regulación del crecimiento y división celular. La proteína que codifica, p53, actúa como punto de control del ciclo celular de forma que bloquea la división si la molécula de ADN no se ha replicado correctamente/ está dañada. En función del tipo y extensión de la alteración que detecte, activará la expresión en cascada de genes de reparación del ADN o inducirá la apoptosis celular. La mutación en línea germinal de TP53 causa un síndrome de predisposición familiar al cáncer, conocido como Síndrome de Li-Fraumeni, con una frecuencia muy baja en la población (1:17.000- 1:23.000) (44). Las personas con este síndrome presentan un alto riesgo de

desarrollar un amplio espectro de tumores malignos, entre los que se incluyen sarcomas, leucemias, carcinomas adrenocorticales, cáncer de pulmón broncoalveolar y cáncer de mama, con una mediana de edad de 25 años al diagnóstico del primer evento cancerígeno. Si bien dada su baja frecuencia en la población la mutación germinal de TP53 solo justifica un 1% de todos los casos diagnosticados de cáncer de mama, las mujeres portadoras presentarán un riesgo 60 veces mayor que la población general de desarrollo temprano de cáncer de mama (entre la tercera y cuarta década de la vida) así como una mayor proporción de casos de enfermedad avanzada al diagnóstico (36).

#### 1.5.1.3-PTEN:

Gen supresor de tumores localizado en el cromosoma 10q23.31. Codifica para la enzima fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa (PTEN), reguladora negativa de la vía de señalización PI3K/AKT (que transfiere de forma normal señales inductoras de crecimiento y división al interior de la célula). Interviene así en el control del ciclo celular bloqueando el crecimiento celular descontrolado (45). El Síndrome de Cowden está causado por mutaciones en línea germinal de PTEN. Se trata de un síndrome de herencia autosómica dominante de penetrancia incompleta poco frecuente en la población (1:200.000 aprox.) que se caracteriza por el desarrollo de múltiples hamartomas en varios tejidos (incluyendo piel, cerebro, tiroides, mamas y tracto gastrointestinal) con un alto riesgo de malignización. El cáncer de mama constituye la principal neoplasia que desarrollan las personas afectas (aunque suponen menos del 1% de los casos de cáncer de mama en la población), de forma que las mujeres con este síndrome presentan un riesgo vital del 50% de desarrollar cáncer de mama, que además es diagnosticado a edades mucho más precoces que los casos esporádicos (36-46 años)(36).

#### 1.5.1.3-SKT-11:

Gen supresor de tumores ubicado en el cromosoma 19p13.3, previamente conocido como *Liver Kinase B1* (LKB1). Se expresa de forma ubicua en todo el organismo, conformando una serín-treonin quinasa que participa en las vías de apoptosis celular. La mutación de SKT-11 en línea germinal causa el Síndrome de Peutz-Jeghers, síndrome polipomatoso de baja frecuencia poblacional con patrón de herencia autosómico dominante caracterizado por la formación de hamartomas polipomatosos por todo el tracto gastrointestinal (predominantemente intestino delgado), pigmentación mucocutánea y susceptibilidad aumentada a desarrollar patología maligna.(46) El cáncer de mama se encuentra entre los principales tumores malignos a los que predispone este síndrome, junto al colorrectal, de intestino delgado, páncreas, estómago y el cáncer ovárico (similar a lo que ocurría en las mutaciones BRCA1/2). El riesgo de desarrollar cáncer de mama es más bajo que el que supone tener los síndromes tumorales previamente descritos (36). Aun así, se estima un riesgo relativo 15 veces más alto de cáncer mamario a las mujeres que padecen esta condición en comparación con la población general(47).

Se recogen en la Tabla 2 las manifestaciones principales en forma de tumores y/u otras características clínicas que determinan los síndromes hereditarios ocasionados por la presencia de mutaciones en estos genes de alta penetrancia, junto con su patrón de herencia y ubicación en el genoma.

SÍNDROME	GEN	CROMOSOMA	CARACTERÍSTICAS
Mama/Ovario hereditario	BRCA1	17q21	Cáncer de mama y ovario. Herencia autosómica dominante.
Mama/Ovario hereditario	BRCA2	13q12.3	Cáncer de mama, ovario, próstata, páncreas. Herencia autosómica dominante.
Síndrome de Li-Fraumeni	TP53	17p13.1	Cáncer de mama, cerebro, sarcoma de tejidos blandos, leucemia aguda, osteosarcoma. Herencia autosómica dominante.
Síndrome de Cowden	PTEN	10q23.3	Cáncer de mama, ovario, tiroides y colon. Hamartomas múltiples. Macrocefalia. Anomalías vasculares. Herencia autosómica dominante.
Síndrome de Peutz-Jeghers	SKT-11	19p13.3	Cáncer gastrointestinal, cáncer de mama y útero. Hamartomas en intestino, pigmentación mucosa. Herencia autosómica dominante.

**Tabla 2.** Síndromes hereditarios de cáncer de mama (modificado de(47))

### 1.5.2-Genes de moderada penetrancia

#### 1.5.2.1-ATM:

Gen supresor tumoral ubicado en la región cromosómica 11q22.3. Su producto es una quinasa de treonina/serina que interviene en las vías de reparación de roturas de doble hebra de ADN y control del ciclo celular. Su mutación en ambos alelos se expresa como Ataxia Telangiectasia, síndrome caracterizado por déficits en la respuesta inmunológica frente a patógenos, ataxia cerebelosa, fragilidad vascular y predisposición a tumores como el cáncer de mama(36). Los portadores en heterocigosis de un alelo mutado de ATM presentarán un riesgo dos veces mayor de padecer cáncer de mama respecto a la población general (47).

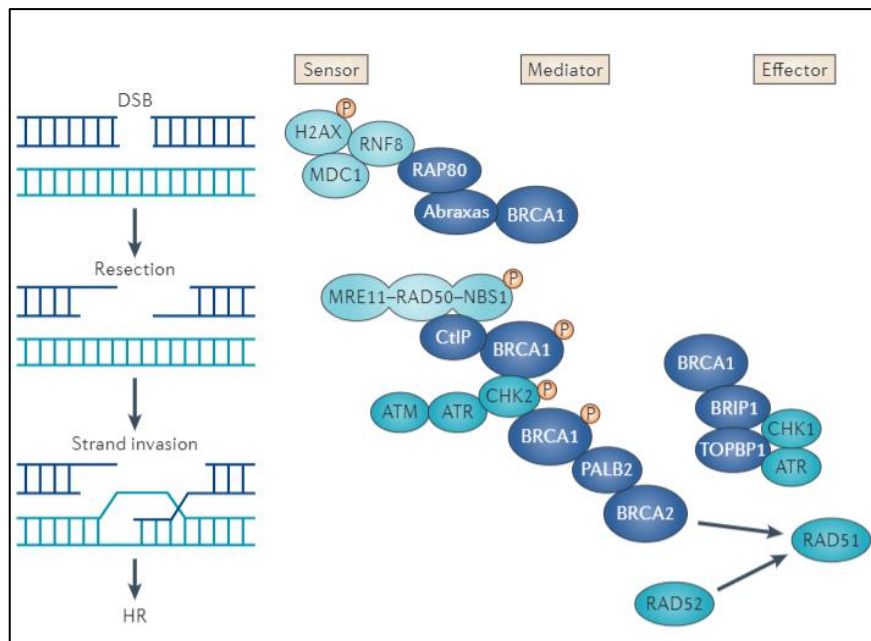
#### 1.5.2.2-CHEK2:

Gen supresor tumoral (cromosoma. 22q12.1), codifica una enzima serin/treonin quinasa que participa en la regulación del ciclo celular al ser activada por ATM ante un daño de doble cadena de ADN, interviniendo en la cascada de reacciones que permite bloquear el ciclo celular hasta que la rotura del ADN es solucionada (48). Determinadas mutaciones en CHEK2 han sido asociadas con un incremento del riesgo de cáncer de mama. Destacamos la variante 1100delC, particularmente frecuente en el norte de Europa, cuya presencia se ha demostrado que aumenta el riesgo de cáncer de mama 2-3 veces (36).

### 1.5.2.3-PALB2:

Gen en cromosoma 16p12.2 que expresa una proteína participante en el mecanismo de reparación de ADN por recombinación homóloga, en estrecha asociación con BRCA2. La expresión del gen mutado en homocigosis causa la Anemia de Fanconi, una afectación autosómica recesiva caracterizada por fallo medular, riesgo incrementado de tumores y anomalías en el desarrollo. La presencia de PALB2 mutado en heterocigosis se evidenció en aprox. un 1% de individuos con cáncer hereditario BRCA1/2 negativo, suponiendo un riesgo 2 veces superior a la población general y encuadrándose entre los genes de penetrancia moderada (36). A pesar de esta categorización inicial, recientes metaanálisis orientan a un riesgo relativo de 5.3 en portadores monoalélicos de mutaciones de PALB2 que supongan pérdida de función de la proteína, por lo que podría revalorarse su posición entre los genes de alto riesgo (48).

Otros genes de moderada o alta penetrancia como como CDH1, RAD51 y BRIP1 o genes reparadores como MLH1, MSH2 y MSH6 también aparecen involucrados en casos de cáncer de mama familiar con cierta frecuencia. La estrategia actual busca analizarlos todos conjuntamente en un panel mediante secuenciación de nueva generación.



**Figura 4.** Mecanismo molecular de reparación homóloga del ADN. (49)

En la figura 4 se esquematiza el mecanismo de reparación homóloga de la hebra de ADN, importante proceso en el que intervienen tanto BRCA1 y BRCA2 como parte de los productos de transcripción de los genes que hemos ido explicando hasta el momento, y cuya alteración explica la inestabilidad genómica celular que sufren los pacientes portadores de mutaciones en dichos genes.

### 1.5.3-Genes de baja penetrancia:

Como se ha expuesto previamente, en conjunto los genes de alta y moderada penetrancia justifican poco más del 30% del riesgo familiar de cáncer de mama. El restante riesgo heredado de cáncer de mama, como planteaban Antoniou AC et al. se debería a la presencia conjunta de múltiples variantes genéticas comunes cuya influencia en la susceptibilidad genética del portador sea baja de forma individual pero relevante en conjunto.

En las últimas dos décadas, numerosos grupos investigadores orientaron sus esfuerzos hacia la identificación de estas supuestas variantes del genoma. Es a partir del año 2005, con la constitución de grandes grupos colaborativos de investigadores como el *Breast Cancer Association Consortium* (BCAC) que se empiezan a llevar a cabo estudios a gran escala y con mayor poder estadístico, permitiendo identificar los primeros SNPs ligados a un incremento del riesgo de cáncer de mama. Estos estudios a gran escala los conocemos por el nombre de GWAS (*genome wide association studies*), que se basan en el escáner de marcadores a lo largo del genoma completo de muchos individuos (casos y controles), buscando variantes genéticas que se presenten de forma persistente y predominante en los individuos afectados y que por tanto puedan asociarse causalmente con la enfermedad a estudio, en este caso el cáncer de mama. En 2007 Easton et al. identifican 5 loci asociados a riesgo de cáncer de mama (FGFR2, 8q24, LSP1, TNRC9 y MAP3K1) utilizando este nuevo sistema de GWAS en población de ascendencia europea (41).

Hasta el presente, el avance de la tecnología y las plataformas de genotipado que dan cabida al estudio de grandes volúmenes de datos han permitido la reafirmación de la significancia de los SNPs ya conocidos en muestras de poblaciones cada vez mayores, así como el creciente descubrimiento de nuevos SNPs asociados con el cáncer de mama. Siendo así, hoy en día se han descrito más de 70 SNPs fuertemente asociados con el cáncer de mama ER+ (refiriéndonos a población caucásica), y cerca de 300 SNPs con una asociación más débil, identificados en un estudio a gran escala llevado a cabo por la BCAC, como parte de un consorcio internacional llamado COGS (Estudio Oncológico Colaborativo de Genes-Ambiente), en el que se genotiparon sobre 80.000 mujeres (40.000 casos de cáncer de mama y 40.000 controles) usando un array customizado, el iCOGs (*Illumina custom array for the Collaborative Oncological Gene-environment Study*) que incluye cerca de 200.000 de SNPs.(50)

La mayoría de los SNPs asociados con cáncer de mama han sido descubiertos en poblaciones de ascendencia europea. Esto supone una dificultad para su utilización práctica en poblaciones no europeas, pues se ha visto que ciertos SNPs que suponen un incremento del riesgo de cáncer de mama en europeos, no tienen significación estadística (y algunos hasta pueden tener un efecto protector) en poblaciones japonesas o africanas (41). En este trabajo nos centraremos únicamente en los resultados obtenidos para población europea, pero ya existen diversos grupos de trabajo que centran su atención en el estudio de la genética del cáncer en otras poblaciones en aras de proporcionar una imagen global que permita en un futuro al paciente, independientemente de su etnia, una valoración y estimación de su riesgo de cáncer de mama de cara al manejo clínico de esta patología.

## 1.6-PREDICCIÓN DE RIESGO

Resumiendo lo relatado hasta este punto, observamos que tanto la genética como los factores ambientales son determinantes en la etiología del cáncer de mama.

Centrándonos en la parte genética, existe una gran variedad de SNPs que aparecen de forma común en la población y que tienen cierta relación con el incremento de riesgo de cáncer de mama. La presencia de cada uno de forma individual no tiene un efecto suficientemente potente como para determinar el riesgo de desarrollar cáncer de mama, pero el efecto combinado de múltiples SNPs podría proporcionar valiosa información sobre el riesgo de cáncer de mama en los individuos. Siguiendo esta teoría, numerosos trabajos plantean el cálculo de un score de riesgo poligénico (PRS) basado en los SNPs que se han ido identificando, y que ha demostrado con éxito la habilidad para la estratificación de las mujeres en función de su riesgo de cáncer de mama. En el estudio hasta la fecha más reciente, publicado por los investigadores del BCAC en enero de 2019 (51), se ha desarrollado un PRS para la predicción de cáncer de mama optimizado según la expresión de receptores de estrógenos del tumor (ER+/-). Los 313 SNPs a partir de los cuales se crea este PRS fueron seleccionados de la mayor base de datos de GWAS disponible de población europea, incluyendo casos y controles de 69 estudios (94.075 casos y 75.017 controles). Este PRS, posteriormente validado en estudios prospectivos en población de Reino Unido, demuestra una capacidad de estimación de riesgo de enfermedad muy alta en los extremos de la distribución. De esta forma, las mujeres en el percentil más alto tienen un riesgo 4.37 y 2.78 mayor a la media, y las mujeres en el percentil más bajo su riesgo es 0.16 y 0.27 el de la media (de desarrollar cáncer de mama ER+ y ER- respectivamente).

Por otro lado, no podemos olvidar que la susceptibilidad genética al cáncer de mama no depende únicamente de estas variantes genéticas comunes en forma de SNPs, sino también de la presencia de variantes menos prevalentes de alta y moderada penetrancia como BRCA1, BRCA2, PTEN, CHEK2 o ATM. Aun así, a pesar del avance en el descubrimiento y descripción todos estos determinantes genéticos, parece que solo explican alrededor del 35% de la agregación familiar observada (52).

Dejando aparte los determinantes genéticos, otros factores epidemiológicos como la densidad mamaria, el IMC, la actividad física, consumo de alcohol, factores reproductivos etc. también están asociados, como hemos descrito en la introducción de este trabajo, con un incremento en el riesgo de cáncer de mama. Similar a lo que ocurre con los SNPs, cada uno de estos factores de riesgo de forma individual suponen un incremento de riesgo de cáncer de mama no más que modesto, pero en conjunto, y asociados a los factores de riesgo genéticos (expresados en forma de PRS) conocidos, podrían aportar una información ya más relevante a nivel clínico.

En búsqueda de la creación de un sistema de predicción de riesgo que incluya todos los factores etiológicos conocidos, se crea el modelo BOADICEA, algoritmo que utilizando información de todos los factores de riesgo conocidos del cáncer de mama (epidemiológicos y genéticos, basados sobre la población de Reino Unido), pretende actuar como herramienta de predicción de riesgo y estimar las posibilidades de una mujer de desarrollar cáncer de mama (u ovario). En el método que se desarrolla en BOADICEA, se valora la capacidad de estratificación según la combinación de los distintos factores, demostrando como se esperaba que la variación en el riesgo (es decir, la amplitud de la distribución y por tanto la capacidad discriminatoria del modelo) es mayor cuando se incluyen todos los factores de riesgo conocidos en el algoritmo y que, cuando se consideran de forma independiente, el que posee mayor poder

de discriminación es el PRS, como se venía indicando previamente, seguido por la densidad mamaria(52).

Estos datos suponen que el PRS es una herramienta útil en la predicción del riesgo de cáncer de mama, sugiriendo la promesa de una mejora en las medidas de screening o prevención del cáncer de mama, ya que su implementación permitiría la focalización de los esfuerzos sanitarios cara a las mujeres con mayor posibilidad de beneficio.

## **1.7-ANÁLISIS DE LA VARIACIÓN GENÉTICA**

Como hemos ido explicando hasta ahora, el llegar a conocer las múltiples variantes genéticas comunes, cuyo efecto individual es mínimo pero que en conjunto alcanzan un considerable poder predictivo de riesgo de cáncer de mama (junto con otros factores determinantes de riesgo), resulta de gran atractivo en vistas a mejorar las medidas de cribado y tratamiento precoz de esta enfermedad.

Las técnicas de análisis de variación de ADN difieren según se trate de una variación nucleotídica simple de baja frecuencia (SNVs) o variación nucleotídica polimórfica (SNPs).

### **1.7.1-Secuenciación**

Las variaciones nucleotídicas simples de baja frecuencia (SNVs) a diferencia de los SNPs carecen de limitación de frecuencia poblacional, y para su estudio son precisas técnicas de secuenciación de ADN. La secuenciación de ADN es el proceso que nos da a conocer el orden en el que aparecen los nucleótidos de una molécula de ADN y que por tanto nos permitirá identificar, comparando muestras de miles de individuos, las diferencias en la secuencia de bases entre unos y otros para luego valorar su posible implicación clínica.

La primera técnica desarrollada para el estudio de las bases nitrogenadas del ADN la describe en 1977 el bioquímico británico Fred Sanger(53). Este sistema revolucionario, basado en la replicación controlada del ADN celular y marcaje con radioisótopos de las bases incorporadas para poder identificar cada una, supuso el inicio de una nueva era en la historia de la medicina, donde el estudio del código genético que determina las características individuales de cada persona podría ser desentrañado. Utilizando el método Sanger, en el año 2003 se consigue secuenciar el primer ADN humano completo, en el contexto del Proyecto Genoma Humano. Pese a la enorme importancia que supuso esta primera técnica de estudio de ADN, rápidamente se hace patente la necesidad de nuevos sistemas que permitan agilizar y abaratar el trabajo de laboratorio (el trabajo realizado por el Proyecto Genoma Humano requirió de la colaboración de cientos de equipos de investigadores, y tardó cerca de 15 años en completarse dado su carácter totalmente manual, costando aproximadamente 3.000 millones de dólares), de forma que el estudio del genotipo sea accesible y resulte factible su análisis en los laboratorios de investigación sin esfuerzos inasumibles.

Gracias al avance de la tecnología y el desarrollo de la bioinformática, los métodos de secuenciación han ido evolucionando desde el primitivo método Sanger, surgiendo sucesivamente nuevas generaciones de secuenciadores cada vez más eficientes (54). El culmen de esta mejora tecnológica llega con el desarrollo de la conocida como secuenciación de nueva

generación o secuenciación masiva paralela (*Massive parallel sequencing*, MPS)(55), que actualmente es la generación de uso predominante en el campo de la investigación. Esta generación de plataformas de secuenciación se caracteriza por su capacidad para secuenciar de forma paralela millones de fragmentos de ADN, de forma que permiten un análisis masivo de material genético en un único experimento, con el consecuente abaratamiento de la técnica y aumento de velocidad de procesado(56). Hoy en día existen numerosas plataformas de secuenciado masivo, tanto de segunda como de tercera generación. El avance inexorable es impresionante, y los actuales secuenciadores de nueva generación permiten el análisis simultáneo de exomas y genomas completos con alta profundidad de lectura y cobertura, a costes muy bajos (un exoma a una profundidad de lectura de 90x se hace actualmente por menos de 500 euros, y un genoma por menos de 1.000).

### 1.7.2-Genotipado

Para el estudio de la variación común (SNPs) son precisas tecnologías de genotipado. Estas se dividen a su vez en tecnologías de alto rendimiento, que permiten el análisis simultáneo de centenares de miles o millones de SNPs y entre las que destacan las tecnologías de *Illumina* y *Affymetrix*; y las de mediano rendimiento, entre las que destaca la tecnología iPLEX de la plataforma de *MassARRAY Agena Bioscience* (San Diego, CA, USA; previamente *Sequenom* Inc.) (57), utilizada en este trabajo y basada en un sistema de detección de espectrometría de masas MALDI-TOF (*matrix assisted desorption/ionization time of light*)

Esta tecnología, desarrollada primariamente en el campo de la microbiología para la identificación de microorganismos, se ha revelado, con el avance del estudio del ADN, como una herramienta para diferenciar genotipos de gran utilidad debido a su método tan eficaz para el genotipado de alto rendimiento que permite la determinación de múltiples polimorfismos en la misma reacción (58). La técnica consiste en una fase inicial, en la que se hace una PCR locus-específica de la región que contiene el polimorfismo a identificar, seguida de una extensión de base única ddNTP (didesoxinucleótido que bloquea la adhesión de más desoxirribonucleótidos) a partir de un cebador oligonucleotídico complementario a la secuencia de ADN inmediatamente anterior (*upstream*) al SNP en cuestión. Utilizando la espectrometría de masas MALDI-TOF determinamos, a partir de la masa del cebador (en inglés *primer*) extendido, la base que se ha añadido en la reacción y, por tanto, la base que compone el SNP inicial (59).

#### 1.7.2.1-Reacción en cadena de la polimerasa (PCR):

Para poder entender el procedimiento desarrollado primero debemos comprender la PCR. La PCR es una técnica de laboratorio que utiliza una enzima ADN polimerasa para hacer múltiples copias de una única molécula de ADN, y así poder disponer de gran cantidad de material y proceder a su estudio. Esto es posible gracias a la capacidad de esta enzima de sintetizar una cadena complementaria a una secuencia de ADN, si se proporciona una mezcla con desoxirribonucleótidos libres y cebadores de ADN (pequeñas secuencias complementarias a un segmento de ADN, que marca el punto desde el que la ADN polimerasa comienza a trabajar para copiar la región de ADN que nos interesa). Se coloca la mezcla en un tubo de laboratorio, junto con los cofactores necesarios para que la enzima funcione ( $Mg^{2+}$ , etc.) y se somete repetidamente a ciclos de cambios de temperatura, que permiten la síntesis del ADN. Los pasos básicos de una PCR son(60):

1-Desnaturalización: se calienta la reacción hasta una temperatura de 94°C, a la cual se desnaturaliza la doble hebra de ADN y se separan las cadenas de ADN, de forma que quedan dos moldes de cadenas sencillas para el siguiente paso.

2-Templado: se baja la temperatura de la reacción hasta 40-55°C, de forma que los cebadores puedan unirse a sus secuencias complementarias en cada una de las cadenas simples del ADN desnaturalizado.

3-Extensión: la enzima sintetiza cadenas complementarias de ADN a lo largo de las dos cadenas originales desnaturalizadas, partiendo de los cebadores primarios unidos a su región complementaria. La ADN polimerasa que se suele utilizar hoy en día en la PCR se denomina Taq polimerasa, extraída de la bacteria *Thermus aquaticus*. Su uso preferente se debe a su capacidad de resistir a altas temperaturas y mantener su función (termoestable), lo cual permite llevar a cabo el proceso de desnaturalización de la doble hebra de ADN a 94°C sin tener que añadir más enzima en cada ciclo por dejar de ser funcional. Para el proceso de extensión se recalienta la reacción a 70°C, pues la Taq polimerasa presenta su máxima actividad a esta temperatura.

Este ciclo se repite N veces, de forma que acabamos consiguiendo miles de millones de copias de la región a estudio a partir de una única muestra de ADN.

Se ha desarrollado una variante de la PCR tradicional conocida como PCR múltiple, que es la que utilizamos en nuestro estudio. Dicho sistema permite el amplificado de varias regiones del ADN de manera simultánea en la misma reacción. Se consigue incluyendo varios conjuntos de cebadores o *primers* específicos para distintas secuencias de ADN junto con la molécula de base, la ADN polimerasa y los desoxirribonucleótidos libres, de forma que en la misma reacción se generan cadenas de ADN complementarias a las regiones de ADN adyacentes a los cebadores incluidos. Esto supone la posibilidad de ampliar distintas secciones clave del ADN que nos interesen según el estudio que queramos realizar (Estudio de polimorfismos, mutaciones puntuales, etc.).

#### 1.7.2.2-Técnica iPLEX:

Entendidos los conceptos básicos que permiten el poder estudiar el ADN, podemos explicar los conceptos que sustentan las fases del trabajo de laboratorio utilizadas para este estudio(57), ejemplificados en la figura 5:

1- PCR múltiple: amplificación de los segmentos de ADN que contienen los SNPs que nos interesa estudiar. Como ya hemos mencionado previamente, se realiza sometiendo a ciclos (aproximadamente 45 ciclos) de variación de temperatura a una mezcla compuesta de agua purificada, la molécula de ADN base suspendida, Taq polimerasa (*Holstar Taq®*), *buffer* para la PCR (solución tampón que mejora la actividad de la reacción múltiple, contiene cofactores como MgCl<sub>2</sub> necesarios para el funcionamiento de la enzima polimerasa), la mezcla con los cebadores que determinan los extremos de la región de ADN a replicar y una solución de desoxirribonucleótidos libres (dNTP).

2- Purificación con fosfatasa alcalina: una vez realizada la amplificación por PCR, se procede a incorporar a la mezcla una solución que contiene la enzima fosfatasa alcalina, cuya función es la de defosforilar los dNTP que no han sido incorporados en la creación de

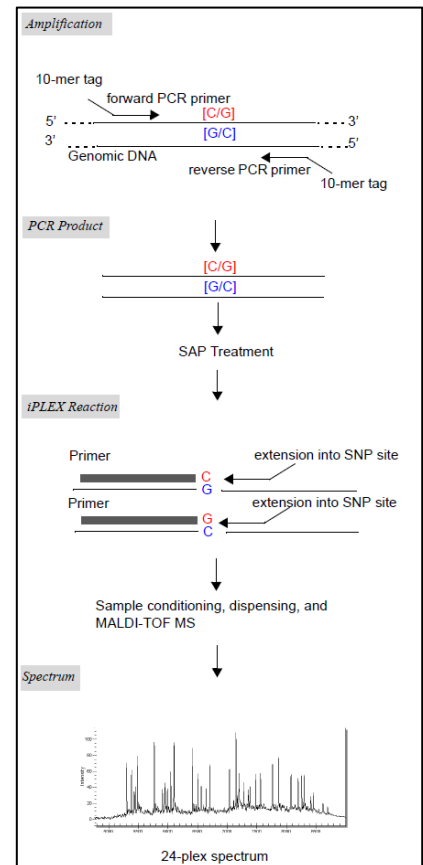
las nuevas hebras de ADN. Así se impide su interferencia en el siguiente paso del proceso. La mezcla con la fosfatasa alcalina añadida debe incubarse en un termociclador estándar.

**3- Reacción de extensión de una única base (Reacción iPLEX):** Incorporamos a la solución que contiene los productos de la PCR múltiple una mezcla de cebadores de extensión diseñados para unirse complementariamente a los fragmentos de ADN ampliados, justo en la secuencia nucleotídica aguas arriba del polimorfismo desconocido. Además de estos cebadores, se incluye una enzima de extensión (*iPLEX enzyme*), una solución tampón como en la PCR y didesoxinucleótidos libres (ddNTPs). Los ddNTPs resultan de quitar un grupo hidroxilo en el carbono C3' de un dNTP, lo cual implica que una vez incorporados a la secuencia actúan como base terminal y bloquean la extensión. Esto permite que el cebador sea extendido y terminado por un único ddNTP, que será complementario a la posición del SNP, diana de nuestro estudio. Igual que en la PCR múltiple, se somete toda esta mezcla a ciclos de variación de temperatura (programa de 200 ciclos cortos) en un termociclador estándar para que suceda la reacción de extensión.

**4- Incorporación de resina:** este paso tiene como objetivo desalinizar la solución con los productos de extensión. Esto se consigue incorporando una resina pura que actúa como quelante de sales, capturando iones que pueda haber en la mezcla y que generarían interferencias en el análisis de las secuencias extendidas por medio de la técnica de MALDI-TOF.

**5- Tránsito de la muestra a un SpectroCHIP Bioarray:** Con la ayuda de un nanodispensador se transfieren los productos de reacción a un *SpectroCHIPArray*. Este chip es un pequeño soporte con 384 pocillos en los que se mezcla la muestra con una matriz orgánica fotosensible. El papel de esta matriz es estabilizar la mezcla y absorber y transmitir energía a las moléculas de la muestra, dotándolas de una carga eléctrica (ionización de las moléculas sin destrucción de su estructura básica).

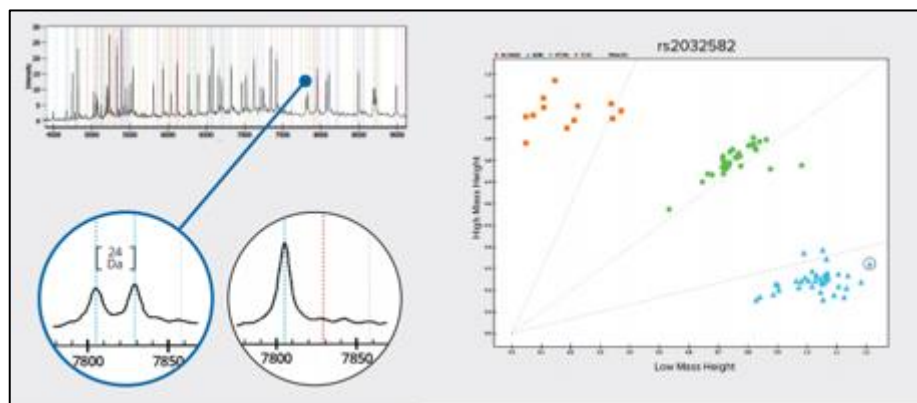
**6- Análisis por espectrometría de masas:** el Chip se carga dentro del equipo de *MassARRAY*, que mediante la técnica de MALDI-TOF identifica el SNP desconocido. El funcionamiento de este sistema se basa en un láser ultravioleta con longitud de onda de 337nm (61) que irradia la matriz que contiene los productos de extensión, en los que induce desorción e ionización, preservando su estructura molecular. Las moléculas cargadas positivamente se volatilizan (desorción) y son aceleradas en un tubo mediante la aplicación de un campo electromagnético hacia un detector. El analizador tipo TOF (tiempo de vuelo) discrimina unos iones de otros en función del tiempo que tarda cada uno en alcanzar el detector. Este parámetro se puede poner en relación con su ratio masa/carga, de forma que los iones más pequeños tardan menos en alcanzar el detector, y los de mayor masa vuelan más lento. El detector registra el tiempo de vuelo de cada producto de extensión (cebador



**Figura 5. Esquema del proceso de genotipado a través de la tecnología iPLEX MassARRAY de Agena Bioscience (modificada de (57))**

+ ddNTP), y el sistema deduce a partir de la masa del conjunto, la masa del ddNTP añadido (pues del cebador conocemos su masa desde la fase de diseño) y por tanto determina la base del SNP inicial que desconocíamos. El proceso completo de análisis de las 384 muestras, desde que el láser incide en el Chip hasta la detección de la señal emitida, tiene lugar en aproximadamente 50 minutos. El detector transfiere automáticamente los datos adquiridos a una base de datos que permite al software el análisis y producción de resultados.

7- Expresión de los resultados: El software de análisis de *MassARRAY* genera informes detallando los resultados del genotipado (Figura 6). Proporciona en una gráfica el espectro obtenido del MALDI-TOF, donde se ven la intensidad de las señales de los analitos en relación con el rango de masas, que determina el ddNTP incorporado y por tanto el SNP a estudiar. El sistema iPLEX incorpora ddNTP con masas modificadas, de forma que permite una mayor distinción por masa de los analitos y por tanto mayor diferenciación entre los posibles alelos, resultando en una detección más sensible. El analizador es diseñado para ser capaz de distinguir analitos con una diferencia de masa de tan solo 16Da. En la gráfica de resultados deben ser valorados 3 picos: el pico del *primer* de extensión sin extender (debería no existir si se ha realizado bien la técnica, pues todos los cebadores deberían estar extendidos con su base correspondiente), un segundo pico para una posible base del alelo materno y un tercer pico para otra base del alelo paterno. Si la muestra resulta homocigota para el SNP en concreto, solo veremos un único pico de la misma base, que englobaría ambos alelos. Por otra parte, el software también proporciona una gráfica de conglomerados (*cluster plot*), que muestra los genotipos de los SNPs de todas las muestras analizadas, que se revisará para verificar que se cumplen los criterios de calidad del genotipado (buena diferenciación de los 3 *clusters*, homogeneidad dentro de cada *cluster*, 3 *clusters* alejados del origen de coordenadas).



**Figura 6.** Panel de resultados del software de Agena MassArray iPLEX (57)

## **2-Hipótesis:**

La predicción del riesgo individual y poblacional de padecer cáncer de mama y de su agresividad, pasa por asociar factores genéticos y ambientales, estos últimos estudiados con criterios epidemiológicos, y analizarlos mediante PRS.

Se han definido ya por consorcios internacionales tras estudios masivos de GWAS los SNPs asociados al riesgo genético y han demostrado su utilidad en poblaciones como la británica.

Las variaciones poblacionales en el riesgo genético tienen una influencia importante en la estima del riesgo (62–64) y desconocemos su frecuencia y contribución al riesgo en poblaciones españolas y más concretamente en población gallega.

## **3-Objetivos:**

**1-** Seleccionar los SNPs más adecuados para cubrir la máxima varianza poligénica del cáncer de mama en la población gallega.

**2-** Diseñar un panel con los SNP que más contribuyen al riesgo de cáncer de mama y analizar su frecuencia en la cohorte BREOGAN de cáncer de mama (65) mediante tecnología de Arrays con espectrometría de masas MALDI-TOF.

**3-** Investigar si una puntuación de riesgo poligénico (PRS), creado combinando los pequeños efectos de los loci de genes asociados a susceptibilidad común, ayuda a discriminar entre mujeres con alto y bajo grado de riesgo de cáncer de mama en la población gallega.

## **4- Material y métodos**

### **4.1- POBLACIÓN DE ESTUDIO: COHORTE BREOGÁN**

Como hemos venido comentando, los GWAS desarrollados hasta ahora han incluido principalmente poblaciones europeas o asiáticas, identificando hasta 313 loci de susceptibilidad genética, que han permitido el desarrollo de un PRS. Este score permite estratificar a las mujeres en grupos de riesgo, de forma que las diferencias de probabilidad de desarrollo de cáncer en los extremos de la distribución llegan a ser muy significativas, con las consecuencias a nivel asistencial (*screening*) que esto puede conllevar. Sin embargo, existe una amplia diferencia en la estructura genética de los genomas de las poblaciones de distinta ascendencia. Estas diferencias justifican la necesidad de evaluar en la población gallega estos SNPs identificados como loci de susceptibilidad al cáncer de mama, con el objetivo de valorar su utilidad en nuestro

entorno. La presencia de estos SNPs fue estudiada en las muestras recogidas en el proyecto BREGAN (*BReast Oncology Galicia Network*).

BREGAN es el Consorcio Gallego de Cáncer de Mama, que nace en el 2010 con la intención de aumentar la información sobre el cáncer de mama en la población gallega, para facilitar su prevención primaria y secundaria. Su base de datos parte de un estudio de casos y controles llevado a cabo en las ciudades de Vigo y Santiago de Compostela, englobando un área sanitaria que incluye aproximadamente un millón de habitantes. Los métodos de recolección de datos han sido descritos ya previamente (66–70). Los casos suponen 1.399 mujeres diagnosticadas de cáncer de mama invasivo entre el año 2000 y 2014 en los hospitales universitarios de Vigo y Santiago (CHUVI y CHUS respectivamente). Los controles son 918 mujeres libres de cáncer (excluyendo cáncer de piel no melanoma), frecuentemente apareados a los casos en lugar de residencia y edad (+/- 5 años). Todas las participantes firmaron un consentimiento informado antes de su inclusión en el estudio, que fue llevado a cabo de acuerdo a los Principios de Helsinki de 1975, revisados en 1983.

#### **4.1.1-Información sobre factores de riesgo epidemiológico**

La información epidemiológica acerca de factores de riesgo de las participantes de la cohorte fue recogida mediante un cuestionario de factores de riesgo, elaborado a partir del *Ella Binational Breast Cancer Study* (71), de forma que se ajustase a la población de Galicia. Las respuestas a dicho cuestionario fueron obtenidas por facultativos a partir de la información registrada en la historia clínica electrónica (IANUS). Las variables estudiadas fueron: nivel educacional [categorizado en: sin estudios, educación primaria, educación secundaria, grado universitario (a su vez dividido en grados de 3 años y 5 años) y doctorado], lactancia materna [categorizado en: nunca lactante, lactancia de duración menor a la media (7 meses), lactancia materna de duración mayor a la media (más de 7 meses)], edad de menarquia, edad de primer embarazo a término, embarazo [categorizado en nunca embarazada vs algún embarazo], edad al diagnóstico, edad de entrada en menopausia, fertilidad al diagnóstico [categorizado en pre y postmenopausia], número de embarazos [categorizado en ninguno, 1-2, ≥3], uso de anticoncepción [si/no], tabaquismo [categorizado en: nunca fumadora, exfumadora, fumadora actual], historia familiar [categorizado en sin historia familiar vs uno o más parientes de primer grado afectos de cáncer de mama u ovario].

#### **4.1.2-Información clínico-patológica**

La información histopatológica fue extraída a partir de los datos registrados en la historia clínica electrónica. De las participantes se recogieron muestras de sangre, orina y, en los casos diagnosticados de cáncer, también tejido tumoral.

El estatus ER, PR y HER2 [Categorizado en positivo y negativo], el grado histológico [categorizada en I (bien diferenciado), II (moderadamente diferenciado) y III (pobremente diferenciado o indiferenciado)], el tipo histológico [categorizado como carcinoma ductal invasivo, carcinoma invasivo lobulillar y otros], y tamaño tumoral (en mm) son las principales características que se estudian de las muestras tumorales.

En cada tumor, se cortan secciones histológicas de 4 µm que se tiñen con tinción de hematoxilina-eosina para su estudio histopatológico, siguiendo los criterios de la Organización

Mundial de la Salud (72). La gradación histológica fue realizada usando la modificación de Nottingham del sistema Bloom-Richardson (73).

Los análisis inmunohistoquímicos del material embebido en parafina para determinar el estado de ER y PR de cada tumor fueron realizados siguiendo los procedimientos estándar de los hospitales gallegos, usando un kit universal de anticuerpos que utiliza un polímero marcado de dextrano conjugado con peroxidasa (*EnVision®*, *Peroxidase/DAB*; *Dako, Glostrup, Denmark*), con anticuerpos para ER (*clone 6F11, dilution 1:50, water bath; Novocastra, Newcastle-upon-Tyne, UK*) y PR (*clone PgR 636, dilution 1:50, water bath; Dako, Glostrup, Denmark*). De forma simultánea a las muestras de interés, se analizaron controles positivos y negativos con los susodichos anticuerpos, con resultados satisfactorios. Las células fueron consideradas inmunopositivas cuando, independientemente de la intensidad de la señal, se detecta marcaje nuclear (tanto difuso como puntiforme) con los anticuerpos. El conteo de las células positivas fue realizado por dos observadores de forma independiente. Cuando el porcentaje de núcleos marcados llega a ser  $\geq 10\%$  se considera la positividad para los receptores ER o PR.

El estudio de inmunohistoquímica se realizó también para la determinación del estatus de HER2 (marcaje de membrana). La expresión de HER2 se categorizó en: Baja expresión de HER2, los casos inmunonegativos (0+) o débilmente positivos (1+); y Sobreexpresión de HER2 (HER2+), los casos con fuerte marcaje inmunológico de membrana (3+). En los casos donde las muestras presentan una positividad de membrana moderada (2+) se realiza una segunda técnica de análisis, utilizando hibridación in situ fluorescente, de forma que aquellas muestras cuya relación de número de copias *ceb-B2/centrómero 17* es  $>2.0$  son consideradas HER2+.

De las 1.766 mujeres participantes en el estudio, 100 presentaban estatus ER desconocido, 114 estatus PR desconocido y 340 estatus HER2 desconocido. 184 mujeres presentaron grado histológico desconocido, 14 tipo histológico desconocido y 144 tamaño tumoral desconocido. De 62 mujeres se desconoce la edad de la menarquia y de 1.443 mujeres los antecedentes de gestaciones previas.

#### **4.2-DISEÑO DE PANEL DE SNPS**

El objetivo del estudio es la valoración de la utilidad de un PRS que estratifique a las mujeres gallegas según su riesgo de cáncer de mama con unas diferencias lo suficientemente relevantes como para que sea útil a nivel clínico y preventivo.

Para la elaboración de nuestro panel de SNPs a analizar en la población gallega, recurrimos al último artículo publicado por Mavaddat et al. (74) sobre la elaboración de scores de riesgo poligénico (enero de 2019). El objetivo de este trabajo es el desarrollo de un PRS para la predicción de cáncer de mama a partir de la base de datos de GWAS más amplia que se dispone hoy en día, y posteriormente validarlo en sets de casos y controles de 10 estudios prospectivos en población de UK. El estudio demuestra la creación de un PRS a partir de 313SNPs cuyo poder predictivo del riesgo de cáncer de mama (optimizado según el subtipo hormonal específico) es significativo y consistente en todos los estudios prospectivos en los que fue validado. El riesgo relativo (*odds ratio*, OR) para el total de la enfermedad por una desviación estándar (SD) fue de 1.61 (95% CI:1.57-1.65) con un área debajo de la curva (AUC): 0,630 (95% CI). Además de demostrar la relevancia significativa de este nuevo PRS desarrollado, el estudio también lo compara con un PRS previamente desarrollado y publicado en el 2015 por

el mismo grupo de trabajo(50), basado en 77 SNPs de susceptibilidad de cáncer de mama bien reconocidos, cuyo OR=1.46 (95%CI:1.42-1.49) con AUC: 0.603 (95%CI). Se observa (Tabla 3) que el poder predictivo del PRS con 313SNPs es superior al del PRS con 77SNPs, pero el incremento de su poder estadístico no es tan elevado en comparación con el gran incremento en el número de SNPs necesarios para poder alcanzar esta mayor capacidad de estratificación. Por esta escasa diferencia estadística decidimos diseñar nuestro panel de SNPs a estudiar en la población gallega a partir de los 77 SNPs definidos en 2015.

	Validation Set			Prospective Test Set		
	OR <sup>a</sup>	95% CI	AUC	OR <sup>a</sup>	95% CI	AUC
<b>77 SNP PRS (PRS<sub>77</sub>)</b>						
Overall BC	1.49	1.44–1.56	0.612	1.46	1.42–1.49	0.603
ER-positive	1.56	1.49–1.63	0.623	1.52	1.48–1.56	0.615
ER-negative	1.40	1.30–1.50	0.596	1.35	1.27–1.43	0.584
<b>313 SNP PRS (PRS<sub>313</sub>)</b>						
Overall BC	1.65	1.59–1.72	0.639	1.61	1.57–1.65	0.630
ER-positive	1.74	1.66–1.82	0.651	1.68	1.63–1.73	0.641
ER-negative	1.47	1.37–1.58	0.611	1.45	1.37–1.53	0.601

**Tabla 3.** Comparativa de OR y AUC obtenidos para PRS<sub>77</sub> y PRS<sub>313</sub>

Esta técnica requiere de un diseño previo cuidadoso que se realiza con el software online de *AgenaCx Assay Design Software V2 (ADS)*. Este programa a partir de la información que le proporcionamos de los SNPs que queremos estudiar, genera las secuencias de cebadores que emplearíamos para ello, y valora *in silico* qué SNPs pueden amplificarse en la misma reacción y cuál es la secuencia de cebador más adecuada, considerando que la utilización de varios cebadores en la misma reacción puede presentar problemas de compatibilidad y alterar los resultados (formación de dímeros de cebadores, interferencias entre ellos, distintas temperaturas de hibridación que impidan su funcionamiento en la misma reacción de amplificación, etc). Una vez diseñado el trabajo, se procede en el laboratorio al desarrollo de la técnica, que ejecutamos con el sistema de genotipado iPLEX.

En primer lugar, para poder hacer nuestro propio panel de SNPs recurrimos a través del software ADS a la **base de datos dbSNP**. Esta plataforma tiene registrados SNPs, microsatélites y pequeñas inserciones y deleciones junto con datos sobre publicaciones en las que se han descrito, frecuencia poblacional, localización en el genoma y consecuencia molecular para polimorfismos comunes. Cada SNP tiene una nomenclatura estandarizada en forma de referencia específica (rs) que sirve de marcador para poder registrarlos en los distintos programas de diseño de trabajo de laboratorio. En el software *AgenaCx* registramos la referencia de cada uno de los 77SNPs que queremos analizar, junto con  $\pm 100$  pares de bases de la región flanqueante de los mismos, todo ello obtenido a partir de la versión del genoma de referencia **Hg 38**, producido en diciembre del 2013 por el Genome Reference Consortium (75). A continuación, tomando como partida la población de referencia europea (CEU, del proyecto internacional para crear un mapa de haplotipos del genoma humano conocido como *HapMap*), se anotan los SNPs secundarios contenidos en el conjunto de secuencias de ADN extraídas que presentaran polimorfismos con una frecuencia del alelo menor  $>0.01\%$  (MAF $>0.01$ , *Minor allele frequency*). Así, identificamos las regiones potencialmente polimórficas con la intención de evitar diseñar cebadores sobre las mismas.

Una vez definidos todos los parámetros descritos hasta el momento, el software ADS trató de agrupar los 77SNPs seleccionados en el menor número de reacciones independientes, teniendo en cuenta la compatibilidad de la secuencia de los cebadores. Son descartados aquellos cebadores con una probabilidad de formación de horquillas y dímeros in silico superior al 40%.

#### 4.3- DESARROLLO DE LOS ENSAYOS DE GENOTIPADO

A partir de las muestras de ADN buscamos identificar las regiones donde se encuentran los SNPs a estudio. Para conocer esta información que nos falta en el laboratorio, seleccionamos el método de genotipado basado en la tecnología iPLEX (57) de la plataforma de *MassARRAY Agena Bioscience*, que permite el análisis múltiple de variantes puntuales del ADN, diferenciadas alélicamente por medio de una reacción de extensión de base única. El producto de esta reacción fue identificado empleando el espectrómetro de masas *MassARRAY4 (MA4, Agena Bioscience)*.

##### 4.3.1-Protocolo de extracción de ADN

Aplicamos a las muestras de sangre obtenidas de la cohorte BREOGAN el protocolo de extracción y purificación de ADN de *Chemagic Magnetic Separation Module I: Chemagic DNA Blood 200 EB 50-200 µl V071212.che*, usando el programa Chemagic IPK del robot Chemagen.

El sistema de Chemagic incluye una rejilla que permite ubicar los distintos componentes en posiciones numeradas de forma que se agilice el proceso y eviten errores de técnica:

- **Posición 1:** Rack con puntas desechables.
- **Posición 2:** Placa con pocillos de 150 µl de la suspensión de las perlas magnéticas.
- **Posición 3:** Placa con 96 pocillos donde incluiremos:
  - 200 µl de Sangre
  - 500 µl de solución tampón de lisis 1 (*Buffer Lysis 1*)
  - 10 µl de Proteasa
  - 1050 µl de solución tampón de ligado 2(*Binding Buffer 2*)
- **Posición 4:** Placa con pocillos con 1000 µl de solución tampón de lavado 3 (*Wash Buffer 3*).
- **Posición 5:** Placa con pocillos con 1000 µl de solución tampón de lavado 4 (*Wash Buffer 4*).
- **Posición 6:** Placa con pocillos con 1000 µl de solución tampón de lavado 5 (*Wash Buffer 5*).
- **Posición 7:** Placa con pocillos con 1000 µl de solución tampón de lavado 6 (*Wash Buffer 6*).
- **Posición 8:** Placa con pocillos con 50-200 µl de solución tampón de elución 7 (*Elution Buffer 7*).
- **Posición 9:** Placa con pocillos vacíos para el desechado de las puntas de pipeteado.

Antes de llenar los pocillos con sus correspondientes contenidos, marcamos cada uno con etiquetas que permitan identificarlos. Una vez hecho esto, utilizando pipetas desechables procedemos a dispensar los tampones de lavado, el tampón de elución 7 y la suspensión de perlas magnéticas, todo ello incluido en el kit de extracción de ADN, en sus correspondientes posiciones predeterminadas. En la posición 3, llenamos los 96 pocillos con 200  $\mu\text{l}$  de sangre cada uno y posteriormente añadimos 10  $\mu\text{l}$  de Proteasa seguido inmediatamente después de 500  $\mu\text{l}$  de solución tampón de lisis 1. Una vez llenados todos los pocillos y verificada su correcta situación en la rejilla del sistema Chemagic, lo introducimos en la máquina de análisis, cerramos herméticamente e iniciamos el proceso. La máquina automáticamente mezcla la solución con sangre y la lisa. Tan pronto la lisis es completada, el procedimiento se interrumpe y extraemos la placa de muestras del sistema para añadir 1050  $\mu\text{l}$  de solución tampón de ligado 2 a cada solución de lisado. Volvemos a introducir la placa y reiniciamos el funcionamiento de la máquina, que continúa el proceso de aislamiento del ADN basado en la separación magnética utilizando las perlas magnéticas que hemos explicado previamente en la introducción. Este sistema es muy rápido, y tras un breve periodo la máquina nos proporciona ya el ADN purificado del resto de componentes celulares y listo para proceder a su cuantificación.

#### **4.3.2- Determinación de la concentración de ADN**

Una vez obtenido el ADN purificado a partir de las muestras de sangre, se procede a su homogeneización durante media hora a una temperatura de 36-37°C, con el objetivo de separar el material genético del sobrenadante restante. Seguidamente se determina la concentración de ADN de cada muestra en  $\text{ng}/\mu\text{L}$ . Esto se realiza con un marcador fluorescente que se une de forma selectiva a la doble hebra de ADN, obviando nucleótidos sueltos, ADN de cadena simple y ARN. El analito empleado fue el tinte *PicoGreen* y el sistema de cuantificación de la señal fluorescente, un fluorímetro *Oubit 2.0*.

Todo el procedimiento tuvo lugar siguiendo las normas de protocolo establecidas del fluorímetro, además de utilizando controles negativos y muestras cuya concentración era conocida ya de antemano.

#### **4.3.3- Normalización de las muestras**

Con el objetivo de tener una concentración similar de ADN en todas las muestras recogidas (10  $\text{ng}/\mu\text{L}$  en 30  $\mu\text{L}$ ), se procede a su normalización.

$$x \mu\text{L} (C \text{ ng}/1 \mu\text{L}) = 300 \text{ ng} \rightarrow x \mu\text{L} = (300 \text{ ng}/C \text{ ng}) 1 \mu\text{L}$$

La formulación previa muestra los cálculos necesarios para realizar este proceso, siendo  $C \text{ ng}/1 \mu\text{L}$  la concentración inicial del ADN (conocida) y  $x \mu\text{L}$  el volumen necesario de muestra a extraer para su normalización.

Las muestras ya normalizadas están preparadas para su estudio.

#### **4.3.4- PCR Múltiple**

Amplificación de los segmentos de ADN que contienen los SNPs que nos interesa estudiar. Para llevar a cabo esta reacción, se emplean 2 cebadores por SNP, uno *forward* (que se une en el extremo 3' de la secuencia a estudio) y otro *reverse* (inverso), que determinan el inicio y fin de la región a amplificar.

En este estudio las condiciones de mezcla de los reactivos fueron los siguientes: 2µL de ADN genómico (10-15ng/µL) en un total de 5µL conteniendo 1,42µL de H<sub>2</sub>O destilada; 0,5µL 10X de solución tampón *Holstar* (solución que mejora la actividad de la reacción múltiple al contener cofactores como MgCl<sub>2</sub> necesarios para el funcionamiento de la enzima polimerasa); 0,04µL de dNTP; 0,04µL de Taq polimerasa (*Holstar Taq*<sup>®</sup>) y 1µL de solución con la mezcla de cebadores que participan en la PCR.

El protocolo del termociclador comprendió: un primer ciclo de desnaturalización inicial de la doble hebra de ADN a 94°C durante 15 minutos y posteriormente 45 ciclos de 94°C durante 20 segundos, 62°C durante 30 segundos y 72°C durante 1 minuto, seguido de una extensión final a 72°C durante 3 minutos.

#### **4.3.5-Reacción de purificación con fosfatasa alcalina SAP**

Una vez realizada la amplificación por PCR, los productos amplificados son tratados con una solución de SAP (del inglés *Shrimp Alkaline Phosphatase*, fosfatasa alcalina de camarón. Para ello se preparamos una reacción con 1,7µL de H<sub>2</sub>O destilada y 0,3U de SAP. De esta mezcla se dispensa 2µL sobre el producto de PCR, y posteriormente se incubaba a 37°C durante 20 minutos para favorecer la actividad de la enzima, y acto seguido a 80°C durante 5 minutos para inactivar la SAP y finalizar el proceso.

#### **4.3.6-Reacción de extensión de base única (Reacción iPLEX)**

Para llevar a cabo esta reacción de extensión, incorporamos sobre el producto de PCR-SAP un volumen de 2µL de solución de extensión, que se prepara con las siguientes características: 0,755µL de H<sub>2</sub>O destilada, 0,2µL de solución tampón iPLEX, 0,2µL de mezcla iPLEX (que incluye los ddNTPs libres que se incorporan); 0,804µL de solución de mezcla de cebadores de extensión (7µM:14 µM) y 0,041µL de enzima iPLEX.

El protocolo del termociclador comprendió: desnaturalización inicial a 94°C durante 15 minutos, 45 ciclos de 94°C durante 20 segundos, 62°C durante 30 segundos y 72°C durante 1 minuto, seguido de una extensión final a 72°C durante 3 minutos.

#### **4.3.7-Purificación de los productos de extensión**

En este paso procedemos a purificar los productos de extensión obtenidos de la reacción iPLEX, con el objetivo de eliminar las sales presentes en la solución. Esto se realiza diluyendo las muestras en 16µL de agua *MQ* (*Milli-Q water*, agua purificada mediante tecnología de electrodesionización) y 6mg de resina (*Clean Resin*), manteniendo la mezcla en rotación durante 15 minutos. Posteriormente se centrifuga a 3800rpm durante 5 minutos.

#### 4.3.8-Trasferencia de la muestra

Con el equipo *RS1000 Nanodispenser* de *Agena Bioscience*, se dispensan automáticamente entre 10nL y 15nL de cada muestra de los productos de extensión purificados a un *SpectroCHIPArray* (*Agena Bioscience*).

#### 4.3.9-Lectura de los productos de extensión

El Chip se carga dentro del equipo *MassARRAYAnalyzer 4*, que mediante la técnica de MALDI-TOF identifica el SNP desconocido (61). El proceso completo de análisis de las 384 muestras desde que el láser incide en el Chip hasta la detección de la señal emitida tiene lugar en aproximadamente 50 minutos. El detector transfiere automáticamente los datos adquiridos a una base de datos que permite al software el análisis y producción de resultados.

### 4.7- ANÁLISIS ESTADÍSTICO

La posible asociación entre los SNPs individuales y el riesgo de cáncer de mama fue valorada utilizando riesgos relativos (*odds ratios*, ORs) e intervalos de confianza del 95% (CIs) derivados de modelos de regresión logística, y ajustados por edad y lugar de estudio. En el estudio actual, los datos de ER, PR y HER2 están disponibles en 564, 555 y 250 mujeres con cáncer de mama, respectivamente. Los análisis de subgrupo fueron realizados dentro de los subtipos de cáncer de mama, incluyendo ER +, ER- y ER-/PR-/HER2-.

Para evaluar los efectos combinados de los SNPs en el riesgo de cáncer de mama, hemos creado un Score de riesgo poligénico ponderado (*weighted polygenic risk score*, PRS) para cada participante del estudio, multiplicando el número de alelos de riesgo (0/1/2) de cada SNP por el peso para ese SNP (escala logarítmica del OR por alelo derivado del estudio actual), y posteriormente sumándolos todos en conjunto. Dado que los datos para un set completo de SNPs solo estaban disponibles para los casos y controles genotipados usando el *iCOGS SNP array* (recordamos que se trata de un array customizado diseñado para el estudio de SNPs determinantes en tumores hormonales como el cáncer de mama, y del que se valió el gran consorcio internacional BCAC para la descripción de más de 300 SNPs de riesgo de cáncer de mama), se realizó un análisis PRS entre dichos sujetos.

Hemos creado un PRS usando los SNPs que en este estudio mostraron una asociación estadísticamente significativa con el riesgo de cáncer de mama. En el presente estudio, diez SNPs se asociaron con cáncer de mama ( $P < 0.1$ ) mostrando una asociación en dirección coherente con la información recogida en trabajos previos. El SNP rs1219648 no fue incluido en el array de SNPs iCOGS y solo 66% de las mujeres incluidas entre los sujetos del COGS tenían datos disponibles para este SNP, genotipados en nuestros proyectos previos a través de la tecnología Agena usada en este mismo trabajo. Sin embargo, no alcanzaron  $p < 0.1$  en las muestras analizadas utilizando el array de iCOGS, y por tanto no fue incluido en el análisis de PRS. Lo mismo sucedió con el SNP rs13387042, que por tanto también fue excluido del PRS score. Los 8 restantes SNPs, rs10069690, rs999737, rs8170, rs17817449, rs13329835, rs1045485, rs4849887, y rs4808801 si fueron incluidos en el análisis de PRS. Todos los análisis estadísticos fueron desarrollados en el software SAS (*Statistical Analysis System*), versión 9.3, con la utilización de test de dos colas (prueba bilateral).

## 5- Resultados

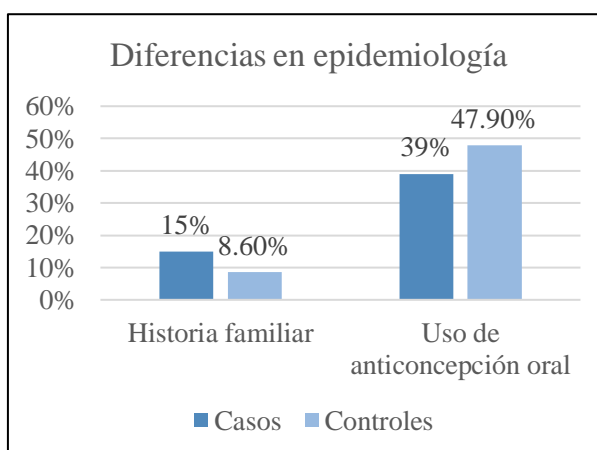
### 5.1-DISEÑO DE LOS ENSAYOS DE GENOTIPADO

El conjunto de los 77 SNPs seleccionados e incluidos en nuestro diseño de genotipado quedaron agrupados en 3 ensayos (W) o reacciones independientes (Anexo 1). El ensayo 1 (W1) permitió genotipar 32 SNPs simultáneamente y el ensayo W2 integró el genotipado de 31 SNPs, dejando el ensayo W3 conformado por los 14 SNPs restantes.

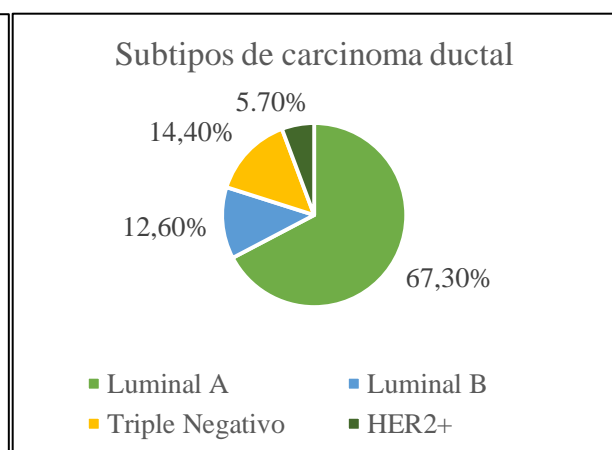
Revisamos la gráfica de resultados que genera de forma automática el software de *MassARRAY*, evaluando la calidad del estudio realizado. De los 77 SNPs, analizados en 3 multiplexes, 6 no pasaron controles de calidad. En 71 los resultados fueron perfectos y se separaban en *clusters* claros, pasando todos ellos el control de calidad. Los 6 SNPs que no pasaron el control de calidad ya habían dado problemas en estudios previos de genotipado realizados en nuestro laboratorio. La tasa de éxito del genotipado fue del 92.2%.

### 5.2-DATOS EPIDEMIOLÓGICOS Y DE TIPO DE TUMOR

Un 15% de los casos vs 8.6% de los controles presentaron historia familiar de cáncer de mama u ovario, y 39% de los casos frente a 47.9% de los controles mostraron uso previo de anticoncepción oral. Casos y controles presentaron proporciones similares en otros factores de riesgo incluyendo IMC, uso de terapia hormonal sustitutiva y demás características de corte hormonal y reproductivo. Concordando con la literatura médica actual, el carcinoma ductal invasivo fue el tipo histológico más común diagnosticado (85.2%) y la distribución de los 4 grandes subtipos moleculares de cáncer de mama fue de 67.3%, 12.6%, 14.4% y 5.7%, para Luminal A, Luminal B, Triple Negativo (TNBC) y HER2+, respectivamente. La media de tamaño tumoral fue de 2.4cm.



**Figura 7.** Representación gráfica de las diferencias epidemiológicas halladas entre casos y controles



**Figura 8.** Subtipos moleculares de carcinoma ductal diagnosticados entre los casos a estudio

### 5.3-EVALUACIÓN DE LOS SNPs EN SUS 71 LOCI

De los 71 SNPs investigados que pasaron el control de calidad, 54 tenían un riesgo relativo (OR) de cáncer de mama en la misma dirección que en los europeos. Sin embargo, solo 12 de ellos mostraron nominalmente una asociación estadísticamente significativa ( $p \leq 0.05$ ) (Tabla 4). El OR (95% CI) fue de 0.8 (0.66-0.96) para el SNP rs6828523, ubicado en el gen 4q34.1/ADAM29 ( $p=0.014$ ), 1.15 (1.03-1.29) para el SNP rs1432679 localizado en 5q33.3/EBF1 ( $p=0.012$ ), 0.86 (0.77-0.95) para el SNP rs7904519 en 10q25.2/TCF7L2 ( $p=0.0052$ ), 0.87 (0.77-0.98) para el SNP rs11199914, situado en 10q26.12 ( $p=0.022$ ), 1.18 (1.04-1.33) para el SNP rs12422552, ubicado en 12p13.1 ( $p=0.008$ ), 1.2 (1.06-1.36) para el SNP rs10941679, localizado en 5p12 ( $p=0.0045$ ), 1.14 (1.01-1.28) para el SNP rs10069690 en 5p15.33/TERT ( $p=0.039$ ), 0.87 (0.78-0.97) para el SNP rs865686 situado en 9q31.2 ( $p=0.015$ ), 1.53 (1.37-1.72) para el SNP rs2981579 ubicado en 10q26.13/FGFR2 ( $p=7.8.E^{-14}$ ), 0.82 (0.72-0.93) para el SNP rs999737 localizado en 14q24.1/RAD51L1 ( $p=0.0028$ ), 0.88 (0.79-0.98) para el SNP rs17817449 en 16q12.2/FTO ( $p=0.024$ ), y de 1.26 (1.12-1.42) para el SNP rs3803662 situado en 16q12.1/TOX3 ( $p=0.00013$ ).

SNP	Chr/Gene	Alleles		Cases		Controls		OR (95% CI)	pvalue
		Common/Minor		N	MAF	N	MAF		
rs6828523	4q34.1/ADAM29	C/A		1399	9.47%	1806	11.63%	0.8 (0.66-0.96)	0.014
rs10941679	5p12	A/G		1324	28.13%	1803	24.27%	1.2 (1.06-1.36)	0.0045
rs10069690	5p15.33/TERT	G/A		1392	32.00%	1800	30.00%	1.14 (1.01-1.28)	0.039
rs1432679	5q33.3/EBF1	A/G		1398	49.43%	1806	47.37%	1.15 (1.03-1.29)	0.012
rs865686	9q31.2	A/C		1397	35.15%	1805	38.14%	0.87 (0.78-0.97)	0.015
rs7904519	10q25.2/TCF7L2	G/A		1399	45.1%	1804	47.75%	0.86 (0.77-0.95)	0.0052
rs11199914	10q26.12	G/A		1398	31.29%	1806	34.05%	0.87 (0.77-0.98)	0.022
rs2981579	10q26.13/FGFR2	G/A		1398	50.32%	1805	40.33%	1.53 (1.37-1.72)	7.86E-14
rs12422552	12p13.1	G/C		1398	29.97%	1803	26.93%	1.18 (1.04-1.33)	0.0080
rs999737	14q24.1/RAD51L1	G/A		1396	21.38%	1789	24.62%	0.82 (0.72-0.93)	0.0028
rs17817449	16q12.2/FTO	A/C		1399	40.35%	1805	42.30%	0.88 (0.79-0.98)	0.024
rs3803662	16q12.1/TOX3	G/A		1395	34.12%	1806	29.57%	1.26 (1.12-1.42)	0.00013

**Tabla 4.** Riesgo de cáncer de mama y loci de susceptibilidad definidos con  $p \leq 0.05$

Se encontraron a mayores seis SNPs que se asocian marginalmente con el riesgo de cáncer de mama ( $p < 0.10$ ). El OR (95% CI) fue de 1.11 (1-1.25) para el SNP rs2046210 de localización 6q25.1/ESR1 ( $p=0,059$ ), 0.86 (0.74-1.01) para el SNP rs6472903 localizado en 8q21.11 ( $p=0.062$ ), 1.12 (0.98-1.29) para el SNP rs11780156 en 8q24.21 ( $p=0,10$ ), 0.91 (0.81-1.01) para el SNP rs2380205 situado en 10p15.1/ANKRD16 ( $p=0,086$ ), 0.9 (0.8-1) para el SNP rs4808801 en 19p13.11/SSBP4 ( $p=0,057$ ) y de 1.11 (0.99-1.24) para el SNP rs3760982 ubicado en 19q13.31 ( $p=0.062$ ).

### 5.4-EVALUACIÓN DE LA ASOCIACIÓN SEGÚN SUBTIPO MOLECULAR ER/PR/HER2 (+/-)

Los resultados de asociación entre los distintos SNPs con el riesgo de cáncer de mama ajustado según la positividad de ER se presentan en la Tabla 5. Solo los SNPs que presentaron una asociación estadísticamente significativa ( $p \leq 0.05$ ) o marginalmente significativa ( $p \leq 0.1$ ) con los subtipos de cáncer determinados en este apartado son representados en dicha tabla. Diez SNPs que no fueron asociados con el riesgo de cáncer de mama de manera general han mostrado una relación de significación estadística ( $p \leq 0.05$ ) en el análisis por subtipos (3 con ER+, 6 con ER- y 1 con TNBC), en la misma dirección que ya habían indicado estudios previos. Los SNPs rs13329835 en 16q23.2/CDYL2, rs3760982 en 19q13.31 y rs13329835 en 16q23.2/CDYL2 se asocian con el cáncer de mama ER+, con un OR (95% CI) de 1.15 (1-1.32), 1.15 (1.02-1.29), y

1.15 (1-1.32), respectivamente. Los SNPs rs11780156 en 8q24.21, rs2046210 en 6q25.1/ESR1, rs2363956 en 19p13.11/MERIT40, rs4245739 en 1q32.1/MDM4, rs6472903 en 8q21.11, y rs6678914 en 1q32.1/LGR6 presentan una asociación por el contrario con el cáncer de mama ER-, con un OR (95% CI) de 1.29 (1.03-1.61), 1.21 (1-1.46), 0.81 (0.67-0.98), 1.31 (1.08-1.6), 0.73 (0.55-0.98), y 0.82 (0.67-0.99), respectivamente. Finalmente, el SNP rs720475 en 7q35 se asocia con tumores de mama triple negativo (TNBC, ER-/PR-/HER2-), con un OR (95% CI) de 1.38 (1.07-1.79).

Se observaron además diferencias en la fuerza de asociación de los distintos SNPs con el cáncer de mama, en función del determinado subtipo molecular de cáncer de mama al que nos refiramos. Los SNPs rs6828523 en 4q34.1/ADAM29, rs10941679 en 5p12, rs1432679 en 5q33.3/EBF1, rs865686 en 9q31.2, rs2981579 en 10q26.13/FGFR2, rs12422552 en 12p13.1, rs999737 en 14q24.1/RAD51L1, rs17817449 en 16q12.2/FTO, rs3803662 en 16q12.1/TOX3, rs13329835 en 16q23.2/CDYL2, rs4808801 en 19p13.11/SSBP4 y rs3760982 en 19q13.31, mostraron una asociación con tumores ER+ con unos OR (95% CI) de 0.78 (0.64-0.95), 1.24 (1.09-1.42), 1.15 (1.02-1.29), 0.86 (0.76-0.98), 1.62 (1.43-1.83), 1.17 (1.03-1.34), 0.81 (0.7-0.94), 0.88 (0.78-1), 1.26 (1.11-1.43), 1.15 (1-1.32), 0.88 (0.78-1), y 1.15 (1.02-1.29), respectivamente, pero no con tumores con ER-. De forma similar, los SNPs rs1199914 en 10q26.12 y rs2363956 en 19p13.11/MERIT40 se asocian con tumores ER- con un OR (95% CI) respectivo de 0.81 (0.67-0.98) y 0.76 (0.62-0.93), pero no con tumores ER+.

Los SNPs rs10069690 en 5p15/TERT y rs999737 en 14q24/RAD51L1 son los que mostraron la asociación más fuerte con los tumores TNBC (triple negativo, ER-/PR-/HER2-), aunque su asociación con cáncer de mama ER+ y ER- también fue puesta en evidencia (Tabla 5).

SNP	Chr/Gene	ER+ (N= 1043)		ER- (N= 278)		ER-/PR-/HER2- (N= 157)	
		log-Additive OR & 95%CI	pvalue	log-Additive OR & 95%CI	pvalue	log-Additive OR & 95%CI	pvalue
rs4245739	1q32.1/MDM4	0.96(0.84-1.09)	0.55	1.31(1.08-1.6)	<b>0.0065</b>	1.29(1.01-1.66)	<b>0.044</b>
rs6678914	1q32.1/LGR6	1.06(0.94-1.2)	0.34	0.82(0.67-0.99)	<b>0.043</b>	0.72(0.56-0.93)	<b>0.0097</b>
rs6828523	4q34.1/ADAM29	0.78(0.64-0.95)	<b>0.015</b>	0.83(0.61-1.14)	0.25	0.9(0.6-1.34)	0.59
rs10941679	5p12	1.24(1.09-1.42)	<b>0.0014</b>	1.03(0.82-1.3)	0.80	1.22(0.91-1.64)	0.20
rs10069690	5p15.33/TERT	1.1(0.96-1.25)	0.16	1.19(0.97-1.45)	0.098	1.38(1.07-1.77)	<b>0.014</b>
rs1432679	5q33.3/EBF1	1.15(1.02-1.29)	<b>0.026</b>	1.15(0.95-1.39)	0.14	1.44(1.12-1.83)	<b>0.0036</b>
rs2046210	6q25.1/ESR1	1.09(0.96-1.23)	0.19	1.21(1-1.46)	<b>0.048</b>	1.13(0.88-1.44)	0.34
rs720475	7q35	1.03(0.9-1.18)	0.68	1.2(0.98-1.48)	0.081	1.38(1.07-1.79)	<b>0.015</b>
rs6472903	8q21.11	0.91(0.77-1.08)	0.26	0.73(0.55-0.98)	<b>0.029</b>	0.74(0.52-1.07)	0.10
rs11780156	8q24.21	1.03(0.89-1.2)	0.65	1.29(1.03-1.61)	<b>0.027</b>	1.27(0.96-1.68)	0.098
rs865686	9q31.2	0.86(0.76-0.98)	<b>0.018</b>	0.87(0.72-1.05)	0.15	0.91(0.72-1.16)	0.46
rs7904519	10q25.2/TCF7L2	0.83(0.74-0.94)	<b>0.0024</b>	0.83(0.69-0.99)	<b>0.040</b>	0.91(0.72-1.14)	0.41
rs2981579	10q26.13/FGFR2	1.62(1.43-1.83)	<b>8.81E-15</b>	1.2(0.99-1.45)	0.064	1.1(0.86-1.41)	0.44
rs11199914	10q26.12	0.88(0.78-1)	0.058	0.76(0.62-0.93)	<b>0.0074</b>	0.79(0.61-1.02)	0.072
rs12422552	12p13.1	1.17(1.03-1.34)	<b>0.019</b>	1.23(1-1.51)	0.056	1.22(0.94-1.59)	0.14
rs999737	14q24.1/RAD51L1	0.81(0.7-0.94)	<b>0.0038</b>	0.84(0.67-1.04)	0.11	0.84(0.63-1.12)	0.23
rs17817449	16q12.2/FTO	0.88(0.78-1)	<b>0.047</b>	0.83(0.69-1.01)	0.062	0.87(0.68-1.11)	0.26
rs3803662	16q12.1/TOX3	1.26(1.11-1.43)	<b>0.00030</b>	1.11(0.91-1.36)	0.31	1.07(0.83-1.38)	0.62
rs13329835	16q23.2/CDYL2	1.15(1-1.32)	<b>0.046</b>	0.97(0.78-1.21)	0.81	1.1(0.84-1.45)	0.50
rs4808801	19p13.11/SSBP4	0.88(0.78-1)	<b>0.050</b>	0.94(0.77-1.13)	0.50	0.97(0.76-1.23)	783
rs2363956	19p13.11/MERIT40	0.96(0.86-1.09)	0.55	0.81(0.67-0.98)	<b>0.032</b>	0.79(0.62-1.01)	0.054
rs3760982	19q13.31	1.15(1.02-1.29)	<b>0.025</b>	0.98(0.81-1.18)	0.84	0.88(0.69-1.12)	0.29

**Tabla 5.** Riesgo de cáncer de mama según subtipos ER, PR, HER2. Se representan los SNPs que mostraron una asociación estadísticamente significativa (p<0.05) con alguno de los 3 subtipos de cáncer de mama definidos. Los resultados con un valor p<0.05 se resaltan en negrita.

## 5.5-ANÁLISIS DEL SCORE DE RIESGO POLIGÉNICO (GENETIC RISK SCORE, GRS) PARA EL CÁNCER DE MAMA EN GENERAL

Examinamos la asociación entre los 71 SNPs establecidos (que pasan el control de calidad) en forma de PRS y el riesgo de cáncer de mama. Los OR (95% CI) para el riesgo global de cáncer de mama en quintiles crecientes del PRS fueron de 1.75 (1.30–2.37), 1.56 (1.15–2.11), 2.02 (1.50–2.74) y 2.63 (1.96–3.52) ( $p=7.8 \times 10^{-10}$ ). Estas asociaciones crecientes también se observaron para los subtipos moleculares principales de cáncer de mama.

Los ORs para desarrollar cáncer de mama por percentil del PRS, comparados con el de las mujeres ubicadas en el quintil central (del 40° al 60° percentil) se muestran en la Tabla 6. Para las mujeres ubicadas en el percentil más bajo (1%) de la distribución del PRS, el OR estimado (95% CI), en comparación con aquellas mujeres situadas en el quintil central es de 0.17 (0.05-0.50). En contraste, las mujeres en el percentil de riesgo más alto según la distribución del PRS muestran un OR comparado con las mujeres del quintil central de 3.86 (95% CI: 1.50-9.94), con  $p= 1.69 \times 10^{-24}$ .

El análisis de validación de los datos obtenidos por COGS (50), que incluía solo un estudio a gran escala (pKARMA) el cual no había formado parte de ningún análisis de descubrimiento de SNPs, determinó unas estimaciones de OR similares a aquellas descritas en los estudios restantes, a excepción de las categorías del 60% al 80% y del 90% al 95%, para las cuales las estimaciones concluidas por pKARMA fueron más elevadas. El log OR por unidad de desviación estándar (SD) fue también similar al del estudio pKARMA independiente (log OR por unidad de SD=0.4).

GRS percentille %	Controls	Cases	OR (95% CI)	All studies in iCOGs excluding pKARMA	pKARMA only
	N=1696	N=1282		OR (95% CI)	OR (95% CI)
<1	26	4	0.17 (0.05-0.5)	0.29 (0.23-0.37)	0.48 (0.28-0.83)
[1.5)	93	26	0.34 (0.21-0.56)	0.42 (0.37-0.47)	0.48 (0.36-0.63)
[5.10)	106	43	0.46 (0.3-0.7)	0.55 (0.5-0.61)	0.58 (0.45-0.74)
[10.20)	195	103	0.54 (0.39-0.74)	0.65 (0.6-0.7)	0.68 (0.57-0.81)
[20.40)	382	213	0.68 (0.52-0.88)	0.8 (0.76-0.85)	0.81 (0.71-0.94)
[40.60)	331	265	1 (Reference)	1 (Reference)	1 (Reference)
[60.80)	319	276	1.09 (0.85-1.41)	1.18 (1.12-1.24)	1.35 (1.19-1.54)
[80.90)	136	162	1.4 (1.02-1.91)	1.48 (1.39-1.57)	1.56 (1.34-1.82)
[90.95)	59	90	1.99 (1.31-3.01)	1.69 (1.56-1.82)	2.05 (1.70-2.47)
[95.99)	42	77	2.53 (1.59-4.03)	2.2 (2.03-2.38)	2.12 (1.73-2.59)
>99	7	23	3.86 (1.5-9.94)	2.81 (2.43-3.24)	3.06 (2.16-4.34)
P for trend			1.69E-24		

**Tabla 6.** Comparativa de los ORs estimados por percentil del PRS entre los estudios incluidos en iCOG (excluyendo pKARMA) v el estudio de validación pKARMA

De forma similar, en el estudio actual, el análisis de validación de los datos de COGS que incluyen el estudio BREOGAN en el que se basa el presente estudio, que no formaba parte de ningún análisis de descubrimiento de SNPs (es decir, un análisis totalmente independiente del proceso de estudio sobre el que se basan nuestros trabajos) encontró estimaciones de OR similares a aquellas en los estudios restantes de COGS, excepto por los intervalos del 90% al 95%, y >99%, para las cuales las estimaciones elaboradas en el estudio BREOGAN fueron mayores.

## 6- Discusión

En este estudio tratamos de ver el riesgo genético que confieren un grupo de SNPs, seleccionados de los más de 300 que se han encontrado asociados.

El riesgo de cáncer de mama está influido por un pequeño número de infrecuentes variantes genéticas en genes de susceptibilidad, como son BRCA1, BRCA2, PALB2, ATM y CHEK2, pero a su vez también está determinado por un gran número de variantes genéticas comunes (76). Más de 100 loci comunes de susceptibilidad (77,78) han sido identificados hasta la fecha, principalmente por medio de estudios GWAS y que pueden explicar aproximadamente un 18% de la heredabilidad.

Uno de los objetivos de los estudios GWAS es identificar SNPs que puedan ser utilizados para la creación de modelos de predicción de riesgo capaces de identificar mujeres con alto riesgo de desarrollar cáncer de mama y consecuentemente mejorar la detección precoz y prevención de la enfermedad. Esto es llevado a cabo básicamente mediante la combinación de los pequeños efectos individuales que determinan cada uno de los loci de susceptibilidad común en un score de riesgo genético, de forma global y condicionado según el estatus de expresión de receptores hormonales. En este estudio identificamos loci comunes de susceptibilidad genética específicamente asociados con el cáncer de mama en mujeres de Galicia, tanto relacionándolos con el riesgo de cáncer de mama en general como su influencia en el desarrollo de determinados subtipos específicos de tumor mamario, caracterizados por la expresión de receptores hormonales que muestran las células cancerígenas. Esta matización del riesgo conferido por los loci comunes de susceptibilidad según subtipo hormonal la realizamos de acuerdo a la evidencia científica de la literatura médica con la que contamos, que ha orientado en esta dirección previamente (76,79).

Se ha demostrado recientemente que la combinación de los pequeños efectos individuales de las variantes genéticas comunes en un score de riesgo genético común puede ser utilizado para desarrollar modelos de predicción de riesgo, que a su vez servirían para poder discriminar a aquellas mujeres con bajo y alto riesgo de desarrollar cáncer de mama, y por tanto mejorar la detección precoz y prevención de esta patología(50).

Algunas de estas variantes genéticas se sitúan en proximidad o están contenidas en genes ya conocidos como **FGFR2**, **FTO**, **TOX3**, **LSP1**(80), **RAD51L1**(81), **ADAM29**, **EBF1** o **TCF7L2**. En contraposición, otros SNPs de los estudiados se encuentran ubicados en regiones no génicas, en posiciones muy alejadas de los genes conocidos más próximos. Como ejemplo destacamos posiciones como **8q24**(80), y **2q35**(82,83). El conocimiento actual de las potenciales funciones biológicas de los SNPs que asociamos al cáncer de mama ha sido recogido y compendiado en el trabajo elaborado por Ghoussani y Pharoah (84).

Las variantes ubicadas en **10q26.13** se localizan en el intrón 2 del gen **FGFR2**, que codifica un receptor tirosin-quinasa de la familia de los receptores de factores de crecimiento de fibroblastos. Se sabe que los factores de crecimiento de fibroblastos pueden actuar como oncogenes en el desarrollo de cáncer de mama en roedores (85) y además se han observado amplificaciones y sobreexpresión del receptor FGFR2 en tumores mamarios en humanos, en

un porcentaje entre el 5% y 10% (86,87). Por medio de mapeo genético y el desarrollo de diversos estudios funcionales se ha llegado a concluir que el SNP rs2981579 estudiado en este trabajo es, con alta probabilidad, la variante causal que determina esta mutación en FGFR2 asociada con el desarrollo de tumores de mama(50,88,89).

Dos SNPs que determinamos asociados al incremento del riesgo de cáncer de mama subtipo hormonal- específico, rs7904519 (se asocia a ER+ y ER-) y rs17817449 (se asocia a ER+ exclusivamente), se localizan en los genes **TCF7/L2 (10q25.2–q25.3)** y **FTO (16q12.2)** respectivamente. Estos genes han sido previamente asociados por medio de GWAS con el desarrollo de Diabetes Mellitus tipo 2 así como de obesidad/alto índice de masa corporal(90,91). TCF/L2 parece actuar a su vez como un protooncogén involucrado en la vía de señalización WNT y en la formación de tumores de diversas estirpes (principalmente colorrectal, pero también endometrial, basocelular, de mama u gástrico)(92).

El SNP rs3803662 localizado en **16q12.1** está situado a tan solo 8 kilobases del gen **TOX3** (*TOX high mobility group box family member 3*), codificante para una proteína que se piensa actúa como un factor de transcripción capaz de modificar la estructura de la cromatina(93). Una sobreexpresión de este gen TOX3 ha sido asociado previamente con el desarrollo de metástasis óseas de cáncer de mama (83,94). En nuestro estudio hallamos una asociación entre este SNP y específicamente el subtipo tumoral ER+, en concordancia con la evidencia demostrada en trabajos previos.

El SNP rs10941679 en **5p12**(95) está localizado en la región próxima al gen **MRPS30** (*Mitochondrial Ribosomal Protein S30*), codificante para la conocida como proteína de muerte celular programada. Esta proteína resulta un elemento clave en las vías de señalización intracelular responsables de la apoptosis celular regulada(96,97). En un análisis reciente llevado a cabo por Ghoussaini M et al. (98) se demuestra la relación de la variante rs10941679 con el desarrollo de carcinomas con características tumorales generalmente determinantes de una mejor prognosis. Así mismo, en su artículo asocian este SNP con los tumores ER+, suponiendo así su presencia un mayor riesgo de desarrollo de este tipo de cáncer de mama vs resto de subtipos humorales. Consistentemente con estos hallazgos, en nuestro estudio encontramos que este loci de susceptibilidad está efectivamente asociado con los tumores ER+ ( $p < 0.0014$ ) poco avanzados.

El SNP en **9q31.2**, rs865686, se localiza en una región de 17kb sin genes conocidos hasta el momento. Los genes más próximos a dicha área son **KLF4** (*Kruppel-Like Factor 4*, 636kb centromérico), **RAD23B** (799kb centromérico), y **ACTL7A** (*Actin-Like 7A*, 736kb telomérico). Un reciente estudio de replicación de Warren et al. (99) realizado dentro del BCAC proporcionó una estimación precisa de la magnitud de la influencia de este SNP en cuestión con el incremento del riesgo de cáncer de mama. Este trabajo fue el primero en demostrar que el SNP rs865686 está asociado con el desarrollo de un subtipo hormonal específico de cáncer de mama, el ER+. Curiosamente, un estudio posterior llevado a cabo en población afroamericana (100) choca diametralmente con las conclusiones previas expuestas, asociando este mismo loci de forma general con todos los subtipos de cáncer de mama independientemente de su estatus hormonal, y con una fuerza de asociación mayor precisamente con los tumores de mama ER- en vez de los ER+ como sugerían Warren et al. Frente estos resultados contradictorios, cabe valorar los derivados del segundo estudio como menos fiables, pues el diseño realizado presentaba bajo poder de discriminación entre subtipos hormonales (número de sujetos mucho menor frente al primer estudio:  $n = 2.508$  casos y 2.745 controles, vs  $n = 11.781$  casos y 12.378 controles). Concordando con lo establecido por Warren et al., en nuestro estudio encontramos

una relación estadísticamente significativa ( $p=0.018$ ) entre este SNP y los tumores de mama ER+.

El SNP [rs12422552](#) se ubica en **12p13.1**, una región sin ningún gen aparentemente relevante para el cáncer de mama en sus proximidades. A pesar de esta ubicación intergénica poco sugestiva, encontramos evidencia científica previa (76) que asocia la dotación de esta variable a un incremento de riesgo de cáncer de mama. Nuestro estudio apoya dichas conclusiones, y sugiere una asociación estadísticamente significativa entre este SNP y el desarrollo de tumores ER+ ( $p=0.019$ ).

El SNP [rs6828523](#) se ubica en **4q34.1**, en un intrón del gen **ADAM29**. Este gen codifica una proteína de la familia ADAM (desintegrina-metaloproteasa), que se ha visto participa en numerosos procesos biológicos de interacción intercelular y célula-matriz. Recientes GWAS realizados han demostrado con significancia estadística el perfil de ADM29 como locus de susceptibilidad y por tanto factor de riesgo para el desarrollo de cáncer de mama(101). Las metaloproteasas se involucran en la modificación de la matriz extracelular y disponibilidad de factores de crecimiento, y recientemente se han relacionado sus cambios de expresividad con la progresión del cáncer de mama(102). En nuestro estudio exponemos una posible asociación entre rs6828523 y el desarrollo de cáncer de mama ER+.

En el locus **5p15** donde se ubica el gen **TERT-CLPTM1L** encontramos el SNP [rs10069690](#). Este gen codifica la subunidad catalítica de una telomerasa, cuya función es el control de la longitud de los telómeros, desempeñando un papel fundamental en la senescencia celular. La desregulación de su funcionalidad supone una vía hacia la carcinogénesis muy importante en tumores como el de mama, seroso de ovario, el gástrico o las leucemias, al determinar una gran inestabilidad genómica celular(103). De forma similar a lo concluido por Haiman C et al. (104), determinamos la fuerte asociación [OR= 1.38 (1.07-1.77)] de esta variante con el incremento del riesgo de desarrollo de cáncer de mama triple negativo (TNBC).

**EBF1** (*Early B-Cell Factor 1*) es un gen localizado en **5q33.3**, que codifica un factor de transcripción conocido como COE1. Este factor de transcripción actúa como elemento clave en el control de las vías de transducción que condicionan la diferenciación, activación y correcta funcionalidad de los linfocitos B específicos(105). Además de esta función principal, COE1 está implicada en ciertas vías de regulación de la adipogénesis, y parece tener un papel en el desarrollo de abscesos mamarios (106). En nuestro estudio valoramos la variante [rs1432679](#) localizada en el gen EBF1, demostrando una asociación positiva entre la presencia de dicho SNP y el desarrollo de tumores de mama ER+ y TNBC.

La variante [rs999737](#) se ubica en **14q24.1**, en una región no codificante pero muy próxima a un gen conocido previamente, el **RAD51L1** (81). Este gen codifica una proteína altamente conservada a lo largo de la evolución, que resulta imprescindible en el proceso de reparación de roturas de la doble hebra de ADN por medio de recombinación homóloga. Mutaciones en dicho gen se asocian consecuentemente con un alto riesgo de desarrollo tumoral (ante la inestabilidad genética que suponen los errores de ADN incorregibles), destacando entre ellos los tumores escamosos de recto, leiomiomas y tumores de mama (107). En nuestro estudio demostramos que la presencia de la variante rs999737 en las proximidades del gen RAD51L1 se asocia con un incremento del riesgo de cáncer de mama ER+, si bien la relación causal directa entre la aparición de esta variante y la alteración del gen próximo a ella debe ser estudiada en un futuro para poderse confirmar/descartar esta teoría.

El SNP rs4245739 ubicado en **1q32.1** se localiza en el oncogén **MDM4** (*Murine Double Minute 4*). Este gen codifica para una proteína que actúa como inhibidora de TP73 y TP53, responsables del principal sistema supresor de tumores de la célula, regulando el ciclo celular y la apoptosis. En un GWAS basado en el COGS (79) identificaron este SNP como uno de los cuatro loci asociados específicamente con tumores de mama ER negativos. Así mismo, la variante rs4245739 en el locus de MDM4 parecen mostrar también una asociación con el cáncer de mama en mujeres portadoras de BRCA1, que se presentan predominantemente tumores ER negativos. En la misma dirección que estos estudios previos, los resultados de nuestro trabajo también orientan a una asociación exclusiva de rs4245739 con cáncer de mama ER-, sin relación estadísticamente significativa con los tumores ER+ ( $p=0.55$ ).

El SNP rs2046210 está localizado en **6q25.1**, aproximadamente 29kb *upstream* del gen **ESR1** (*Estrogen Receptor 1*), de gran importancia en el desarrollo sexual y función reproductiva, así como en el desarrollo de tumores de tejidos hormono-dependientes (cáncer de mama, endometrio, ovario). Un GWAS realizado por Zheng et al. (108) en más de 31.000 participante encontró una asociación positiva entre la presencia de rs2046210 y el incremento de riesgo de cáncer de mama en las mujeres de ascendencia china, japonesa y europea, descartando a su vez una relación entre este SNP y el cáncer de mama en mujeres afro-americanas. En nuestro estudio, demostramos una pequeña relación en la población gallega entre rs2046210 y el riesgo de desarrollo de tumores ER-.

Globalmente, el conjunto de los loci de susceptibilidad estudiados parece estar involucrado en múltiples e intrincadas vías de señalización celular. A excepción de FGFR2, la relación entre las variantes causales y los mecanismos biológicos responsables de su efecto clínico continúan siendo actualmente ampliamente desconocidos. Es preciso mantener el empeño científico en este ámbito y persistir con la realización de mayores trabajos de secuenciación, mapeo genómico fino y estudios funcionales para llegar a revelar finalmente los mecanismos biológicos responsables ocultos.

## 7-Perspectivas futuras

La estimación del riesgo de cáncer de mama es vital para poder realizar políticas sanitarias adecuadas, particularmente en lo que se refiere al cribado.

Hoy en día estamos aún lejos de poder predecir el riesgo individual de forma precisa e inequívoca, pero los conocimientos de los que si disponemos pueden servir ya para estimar grupos de mayor y menor riesgo entre las mujeres, y tal vez en base a ellos cambiar las estrategias de cribado hacia unas técnicas más eficaces.

La estimación del riesgo pasa por calcular puntuaciones de riesgo poligénico y combinarlas con marcadores de riesgo epidemiológicos, analizados en conjunto con modelos como BOADICEA (*Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm*) que está demostrando ser el mejor modelo en predicción hasta la fecha. BOADICEA permite calcular una estimación del riesgo de desarrollar cáncer de mama teniendo en cuenta

no solo datos genéticos mediante un PRS (*polygenic risk score*), sino también considerando factores reproductivos y de estilo de vida.

El panel que aquí hemos usado seguramente habría de ser extendido a todos los SNPs conocidos que han mostrado asociación, y datos recientes aún no publicados del consorcio B-CAST sugieren que, sobre todo el valor predictivo en personas de más de 50 años, sería de gran utilidad en la estratificación de riesgo.

Por todo lo previamente expuesto, sería ideal llegar a realizar un proyecto piloto unido al programa gallego de cribado de cáncer de mama en el que se valore, y se hagan los correspondientes estudios de eficiencia y coste/economía, el valor de la utilización del score de riesgo tal y como se describe en BOADICEA, y se implantase el mismo en un grupo de personas más jóvenes que la edad de corte del cribado actual (50 años) para, al menos, hacer el correspondiente seguimiento precoz en aquellas personas que presentasen un alto riesgo.

## 8-Conclusiones

**Primera:** La tecnología Agena de espectrometría de masas MALDIFOF es una tecnología de gran eficacia para el análisis de un conjunto de SNPs a mediana escala (decenas o centenares) como el del presente trabajo.

**Segunda:** De los 77 SNPs seleccionados (de entre los más de 300 asociados a cáncer de mama), 71 pasaron el control de calidad, determinando una tasa de éxito del genotipado del 92.2%.

**Tercera:** De los 71 SNPs investigados que pasaron el control de calidad, 54 presentan un OR de cáncer de mama en la misma dirección que en los europeos. Sin embargo, solo 12 de ellos mostraron nominalmente una asociación estadísticamente significativa ( $p \leq 0.05$ ).

**Cuarta:** El OR (95% CI) fue de 0.8 (0.66-0.96) para el SNP rs6828523, ubicado en el gen 4q34.1/ADAM29 ( $p=0.014$ ), 1.15 (1.03-1.29) para el SNP rs1432679 localizado en 5q33.3/EBF1 ( $p=0.012$ ), 0.86 (0.77-0.95) para el SNP rs7904519 en 10q25.2/TCF7L2 ( $p=0.0052$ ), 0.87 (0.77-0.98) para el SNP rs11199914, situado en 10q26.12 ( $p=0.022$ ), 1.18 (1.04-1.33) para el SNP rs12422552, ubicado en 12p13.1 ( $p=0.008$ ), 1.2 (1.06-1.36) para el SNP rs10941679, localizado en 5p12 ( $p=0.0045$ ), 1.14 (1.01-1.28) para el SNP rs10069690 en 5p15.33/TERT ( $p=0.039$ ), 0.87 (0.78-0.97) para el SNP rs865686 situado en 9q31.2 ( $p=0.015$ ), 1.53 (1.37-1.72) para el SNP rs2981579 ubicado en 10q26.13/FGFR2 ( $p=7.8.E^{-14}$ ), 0.82 (0.72-0.93) para el SNP rs999737 localizado en 14q24.1/RAD51L1 ( $p=0.0028$ ), 0.88 (0.79-0.98) para el SNP rs17817449 en 16q12.2/FTO ( $p=0.024$ ), y de 1.26 (1.12-1.42) para el SNP rs3803662 situado en 16q12.1/TOX3 ( $p=0.00013$ ).

**Quinta:** Algunas de las variantes genéticas estudiadas se sitúan en proximidad o están contenidas en genes ya conocidos, como FGFR2, FTO, TOX3, LSP1, RAD51L1, ADAM29, EBF1 o TCF7L2. Contrastando con estos hallazgos, encontramos que otros SNPs de los estudiados se ubican en regiones no génicas, alejados de los genes conocidos más próximos. La dirección del efecto (protector o de riesgo) está en la dirección obtenida en previos estudios para poblaciones europeas.

**Sexta:** Diez SNPs que no fueron asociados con el riesgo de cáncer de mama de manera general han mostrado una relación de significación estadística ( $p \leq 0.05$ ) en el análisis por subtipos, en la misma dirección que ya habían indicado estudios previos: los SNPs rs13329835 en 16q23.2/CDYL2, rs3760982 en 19q13.31 y rs13329835 en 16q23.2/CDYL2 se asocian con los tumores ER+; los SNPs rs11780156 en 8q24.21, rs2046210 en 6q25.1/ESR1, rs2363956 en 19p13.11/MERIT40, rs4245739 en 1q32.1/MDM4, rs6472903 en 8q21.11, y rs6678914 en 1q32.1/LGR6 con tumores ER- y el SNP rs720475 con el cáncer de mama triple negativo.

**Séptima:** La estimación del riesgo de desarrollar cáncer de mama pasa por el cálculo de puntuaciones de riesgo poligénico (PRS) y su combinación con marcadores de riesgo epidemiológicos, de forma que se analicen conjuntamente por medio de modelos como BOADICEA (*Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm*), que demuestra ser el mejor modelo en predicción de riesgo hasta la fecha. La creación de un modelo como BOADICEA ajustado a la población gallega supondría una potente herramienta clínica para la mejora de la prevención secundaria (*screening*) y el tratamiento precoz de las mujeres afectas por este tumor tan prevalente.

## 9-Anexo

Ensayo	SNP_ID	Cebador PCR_F	Cebador PCR_R	Cebador EXTENSIÓN	Polimorfismo
W1	rs554219	ACGTTGGATGACAGAGGGA GTGTGGAAGTG	ACGTTGGATGATTCCACT CCCATGAAGGTC	TGACTGTGGGTTTCC	G/C
W1	rs11780156	ACGTTGGATGCTCACTTACG AAGGTATGGG	ACGTTGGATGACTGGTCC CAGATAAGGTGA	TGAAGGGTGGTGGGA	T/C
W1	rs889312	ACGTTGGATGGAGATGATCT CTGAGATGCC	ACGTTGGATGTATGGGAA GGAGTCGTTGAG	gCCTGCTGGAGAAAGG	C/A
W1	rs3803662	ACGTTGGATGTTTTCTCTCC TTAATGCCTC	ACGTTGGATGAGACCCAG TACTTTCTTCGC	ATGCCTCTATAGCTGTC	G/A
W1	rs2981579	ACGTTGGATGGGTTGAAGG TCACTTGCTTG	ACGTTGGATGTTTCATATT TCCACCTGCCCG	TTTGGATACGACCTCTG	G/A
W1	rs11249433	ACGTTGGATGACATGGAATC CAAAACTGTG	ACGTTGGATGTGTGCTAA GGAGAAGCAACC	gaCAGGGCTGGGTTTAA	G/A
W1	rs7726159	ACGTTGGATGTGGAGCTCAC GTTCTCTTAC	ACGTTGGATGCCCTTCTG CTTTACAAACCC	atGAGTTTGTGCCAAGT G	C/A
W1	rs10941679	ACGTTGGATGATGATCGGTT GGGCTGTAAG	ACGTTGGATGGCCAGTAA AATGTGGGATGC	TTATGCTGTGTTCTTTC CA	G/A
W1	rs10472076	ACGTTGGATGAGGTCCAGCC GCCAAGTTAA	ACGTTGGATGCCTCATGT CCAAGCTCTTTG	CAGCAGCTGTCCAGACA TC	T/C
W1	rs8170	ACGTTGGATGAGTTTCGCCA TGCAGTTGTG	ACGTTGGATGTGGATACC AAGGGTACCAG	ccccCCAGTGCCACCTCA TA	G/A
W1	rs7904519	ACGTTGGATGGCAGAGCGT GGACATTTTTTC	ACGTTGGATGAAGAAGG GTGGTTAGACAGG	gttgCAGCTGAATGCCGA CC	G/A
W1	rs4849887	ACGTTGGATGTTAAACCCTG GGTGTGATGC	ACGTTGGATGTGGTTGA GGGGATTACCTAC	aCAGCTGTGATCTGGCA ATG	C/T
W1	rs11571833	ACGTTGGATGCTGAATTCTC CTCAGATGAC	ACGTTGGATGAATGCAAG TTCTTCGTCAGC	TCTCCTCAGATGACTCC ATTT	A/T
W1	rs1436904	ACGTTGGATGGCAAGACAG GAGACAGATTC	ACGTTGGATGGTCTTCCC TGCAAGACTATG	tgtgaAGATTCTGCAGGG TGC	G/T
W1	rs1011970	ACGTTGGATGAGGGAAGAT ACAGGTGGAAC	ACGTTGGATGCTCTAGAA CTGATAGGGAGC	gggtGAACTGGGCCAGT GTTT	G/T
W1	rs17879961	ACGTTGGATGAGGAAGTGG GTCCTAAAAAC	ACGTTGGATGCAAAGGTT CCATTGCCACTG	GTGGGTCTAAAAACTC TTACA	G/A
W1	rs11242675	ACGTTGGATGTCTAGAGGA GACTACAACGC	ACGTTGGATGAGCAAACA GTGTCACCGTCC	gggaCTGGGAGACGTGC ATGTG	C/T
W1	rs2943559	ACGTTGGATGAAAAGTGTG AAGAATTCCCG	ACGTTGGATGCTCCGTAA TGCAAGATTTCC	tTGTATATTTGCAACTCT TGTAT	G/A
W1	rs865686	ACGTTGGATGGTCTCATTAG CCAGGAAACC	ACGTTGGATGCGAGACT GTGGTATTCTAGG	gggatGAAACCAAGGTTT CCAAC	G/T
W1	rs2380205	ACGTTGGATGATCTGTCCAG GAGGACAATC	ACGTTGGATGATTTCAC CAGGCACGTTTT	ccataTGTGTAGTCATTT GCCCTA	C/T
W1	rs6678914	ACGTTGGATGGTGTCTATCTT GTGTGTTTGG	ACGTTGGATGCTCCCCAC TTTAGATGGCAC	aatCTAGTATTCAAGTCA TCTCTG	A/G
W1	rs11075995	ACGTTGGATGTGCTTTGAGT ATACTGGGGC	ACGTTGGATGTGAAGCCA TGAGCAGCAACT	ggggTGGGGCCTCAGGG ACTTGCT	A/T
W1	rs13281615	ACGTTGGATGAAGACACGTG GAATTTACTC	ACGTTGGATGCCAAACCC CCTACTCAGAAT	GTAACATGAATCTCAT CAAAAGAA	A/G
W1	rs11814448	ACGTTGGATGAAGTATGTGA ATGTCTCACC	ACGTTGGATGTTGAGCCC CAGAAGCTAATT	AAAAATAAATTCTAAA GCTGAAAA	C/A
W1	rs720475	ACGTTGGATGCAGTGAGCAA GTAAACAGCG	ACGTTGGATGCAAACAAC TACAAGGAACCC	ccaacTAACAGGTTTCTAG TTACTACA	G/A
W1	rs2823093	ACGTTGGATGTCAAGTCAAG GAACAAAGGG	ACGTTGGATGAAGTGCT GGAGGAAGTCAAG	ccccTTCACAAAGAATTT GACTGGCT	A/G
W1	rs6504950	ACGTTGGATGCCAGGGTTT GTCTACCAAAG	ACGTTGGATGCTGAATCA CTCCTTGCCAAC	aggggTGTCTACCAAAGG CAGGATAC	G/A

W1	rs16857609	ACGTTGGATGCTCAAGAAAGGGATGCCAAC	ACGTTGGATGCAGAGACCGGAAGGTATTCT	acctGTCTGCTTTGTCTGGAGACTTC	C/T
W1	rs10995190	ACGTTGGATGCAATGGTTGTGTCCAAGTGC	ACGTTGGATGTGTTCTGATTGGCTTGCCAC	caagGGTTGTGTCCAAGTGCATATTGA	A/G
W1	rs4245739	ACGTTGGATGAGCATAATGGTAGTACGAAC	ACGTTGGATGAACATTCTCTGACAGGTTGG	ccccCGAACATAAAAATGCATTTATTC	C/A
W1	rs704010	ACGTTGGATGGTAAGAGTCTGGGCAGCTTG	ACGTTGGATGCTACTGCCACGTCTTACAAC	tctACCTGAGACCTGACCTGAAAATAGC	C/T
W1	rs7072776	ACGTTGGATGGCAGTACTCACACTTGGTAG	ACGTTGGATGAGACCACTAAGTAGATTGGG	ggggTTTAACTAGAGGCCATATCCATGAA	A/G
W2	rs78540526	ACGTTGGATGATGTTCCCATGGACCTGAGC	ACGTTGGATGCCTCTAAACTGACAGCAGTTG	AGCTCTTCCCAGCAC	T/C
W2	rs10069690	ACGTTGGATGCTGTTTGAAACGGGTTCTCTG	ACGTTGGATGCGTCATCTGAGGAGAGTGTG	GGGATCCTCATGCCA	C/T
W2	rs941764	ACGTTGGATGTGTGAGACTTCTTGTGC	ACGTTGGATGCCAACAGTCTGGCAGCATC	CGGAATGAGCCTGAC	G/A
W2	rs2016394	ACGTTGGATGTCTCTCCCCAAGAGAGTCC	ACGTTGGATGTCTGTTCAAGGAAATCTTCCC	tAGAGTCTTGAAGGG	A/G
W2	rs11199914	ACGTTGGATGACAGTGGGTGGAATAAGCAG	ACGTTGGATGGGAGCTCGTAGATATCAAG	CCACATTCAAGTCTGT	C/T
W2	rs1292011	ACGTTGGATGCAAGTACACCGCTTGCCTTC	ACGTTGGATGACTCCCAAGATCCTAAGTCC	taTCCTGCTGCCTCTGA	A/G
W2	rs616488	ACGTTGGATGTAGGGACAACTAGCTCCTG	ACGTTGGATGTCTTGGTGCTCTGAACAAG	cGGGGTACAGATGAGTG	G/A
W2	rs13387042	ACGTTGGATGGAAGGAAGATTTCGATTCAAC	ACGTTGGATGGAACAGCTAAACCAGAACAG	ATCCTTGGTTTCTGTAGC	G/A
W2	rs12710696	ACGTTGGATGAAGAAGGCGCCTCACATAGT	ACGTTGGATGTCCGCTCAGGAATGATGC	tGTCACATTGTGGTTACC	T/C
W2	rs6762644	ACGTTGGATGATTCTGAGTCCCAAATCG	ACGTTGGATGGGATGGATGTAAAAGAGGC	cccAAATCGATCCTGCC	G/A
W2	rs2588809	ACGTTGGATGACAGTGAGCCAGAACAACAG	ACGTTGGATGCTTGCCCTCACTTGACTTGG	GCCAGAACAAGATGATA	C/T
W2	rs999737	ACGTTGGATGGGTCCCCTTACATGATATG	ACGTTGGATGCACCAAGACTTATGGACAG	CACATGATGAATGGGCG	C/T
W2	rs1353747	ACGTTGGATGTTGGTTACAAAGAAGCTGGG	ACGTTGGATGTGAGAAAGTTGGAGTGCAGG	cctcaCCTCTGAGGTTCCGC	G/T
W2	rs10759243	ACGTTGGATGCCAGCCCCCTTTTATTTTC	ACGTTGGATGTGGGTAACACAGCAAAATG	CCCTTTTTATTTTCTTGTAC	C/A
W2	rs6001930	ACGTTGGATGAGCAGAAATGGAAACCCAGC	ACGTTGGATGACATACAGAGAACTGGTGTG	ATGTATAGCCTGAATCTTCAA	C/T
W2	rs13329835	ACGTTGGATGGAGCTTTGGGTATGTCATGG	ACGTTGGATGCCTAGCAACACACCAGTGAG	GGTATGTCATGGATCTCTCTG	G/A
W2	rs12662670	ACGTTGGATGACTTCTAACGAAGGCAGAGC	ACGTTGGATGTGGCACACATGCATGACAGC	GAGCAAAAAGAAATGGAACACT	G/T
W2	rs2236007	ACGTTGGATGTCCATCACCGACCAAGGTAG	ACGTTGGATGTGCAGATCTGGGACCTCC	ggaggGTGGATGTGCAGCGTCT	A/G
W2	rs3760982	ACGTTGGATGTACACCAGCACAGTAACG	ACGTTGGATGATGTGGTCTTTGTGGAAGG	caccCAGTACGCAAAACACATA	A/G
W2	rs527616	ACGTTGGATGATGGAATGCCCTTAGGAC	ACGTTGGATGTTACACGAGACTGAGCCAAC	aCTTAGGACAAGTCTAACTAGG	C/G
W2	rs6472903	ACGTTGGATGTGAAAAGACAAGCTGCCAGG	ACGTTGGATGGACCTCAAAAATGTCTTC	ccctCTGCCAGGATATAATACTTG	G/T
W2	rs17817449	ACGTTGGATGCTATTAAGGAGCTGGACTG	ACGTTGGATGAGAAGAGTGATCCCTTTGTG	GGAGCTGGACTGTTAAATTAAC	G/T
W2	rs4808801	ACGTTGGATGGGGAAGCAGCTGCATAATGA	ACGTTGGATGTGCAGGCAGATTGTCCCTTC	gTAATGAATATAGAAGCAAGAAGC	A/G
W2	rs6828523	ACGTTGGATGGGCATGGATTTAATAACGAGG	ACGTTGGATGTCTCCATCCCCTTTCTCC	gaagAACGAGGTGTTAATATTGGTT	C/A
W2	rs3817198	ACGTTGGATGTCTCACCTGTACCAGATTC	ACGTTGGATGCTTGAGCTTCCCTAGTGGAG	tcaaCTGATACCAGATTCAACTCTC	C/T
W2	rs4973768	ACGTTGGATGCACTGTCTCAATGAATGC	ACGTTGGATGACTCCATTAAAGACAAAGG	cTCACTTAAAACAAGCAGTTAATTAC	C/T

W2	rs9790 517	ACGTTGGATGTAACCCCATG TCTTGGTTTC	ACGTTGGATGTACCATGC TAAGCACTTTCC	gttgtTTTCTTCCTTAGTA AAGTGAG	T/C
W2	rs1550 623	ACGTTGGATGGGAATTCTTT GTGCAAGGTG	ACGTTGGATGGTTAAAGA GCTTTGAATTCCC	cAGGTATAGAGAAACAC AGAATTCTTA	A/G
W2	rs1735 6907	ACGTTGGATGTGCTGCCAGA TGATCACTTC	ACGTTGGATGTTGTCAGA TCCAGCAGAGAC	ataACTTCATAATGGGG ATTAGATGGT	A/G
W2	rs2046 210	ACGTTGGATGCCTCACACAT ACATACAGTC	ACGTTGGATGTGAAACCA TCAGGGTGCCTC	tgcaTCACACATACATAC AGTCACATAC	A/G
W3	rs1155 2449	ACGTTGGATGCAATTACAGC CCACCTCTTG	ACGTTGGATGTTGAAGTC CCAACCCAGATG	CACCTCTTGATCGT	C/G/T
W3	rs1242 2552	ACGTTGGATGTGAGGGTAA CATCGCGGTAG	ACGTTGGATGCACTGATG GAAAATCGTCGG	GCGGTAGTCAGTCCA	C/G
W3	rs9693 444	ACGTTGGATGTTGCCAGGCA AACTCATGAC	ACGTTGGATGTGCCGTGC ACCCTAAGTTAC	ATGACTAAGGGCAGC	C/A
W3	rs1323 90	ACGTTGGATGACTCTTGCA ATGCCAGTTC	ACGTTGGATGGCTGATG CTCAGAAAAGTGC	TTCAATCCCCCATTTC	C/T
W3	rs1045 485	ACGTTGGATGTGTGGTCCAT GAGTTGGTAG	ACGTTGGATGTGACCACG ACCTTTGAAGAG	CTCTACTGTGCAGTCAT	G/C
W3	rs1752 9111	ACGTTGGATGCTTATTTTTTC CCGGTCTTAG	ACGTTGGATGAGAGCTG TCAGTAGAGTTGC	CGGTCTTAGACCAAGCT A	C/T
W3	rs2042 47	ACGTTGGATGTCTGCTTCAG CTGCAATTAG	ACGTTGGATGCACAGACA ATGCTACCTCTC	ACTTTGAAACAGGCATC A	A/G
W3	rs2363 956	ACGTTGGATGAGCCATGCAG AGGTGACAAC	ACGTTGGATGGTTTGTC ACAGTTTCAAGG	ggAACAGGGACATTGGC T	G/T
W3	rs7591 5166	ACGTTGGATGTCCTCTTCCT AGGACACTTG	ACGTTGGATGTTGCAGTG AAGTGAAGGCTC	ctctGGCCCACCCTGTTA A	C/A
W3	rs1249 3607	ACGTTGGATGATGTATATGA CCTCAAGGGC	ACGTTGGATGAGCCATCG ACTCCCTATTAG	GCTACCTGGAATTGTTA AG	C/G
W3	rs1182 0646	ACGTTGGATGTGCTCCTGGA AGAAGGTAG	ACGTTGGATGGTCATATG ACTCTGGAGGTG	GGACTTGAAATCAAAAG ACA	C/T
W3	rs3903 072	ACGTTGGATGCATTCATCCA CCATATCCCG	ACGTTGGATGCTTCTTTA GTCCTCTCCAG	ATCCATCACTCCTGACC TAAT	G/T
W3	rs1432 679	ACGTTGGATGACTTCGCAAT TGCCACCAAG	ACGTTGGATGGTGAGGG CTACATCATGCAG	aatttCCACCAAGACCAT GGC	C/T
W3	rs1077 1399	ACGTTGGATGTGGTAGGAC ATACTGGTGAG	ACGTTGGATGAAAAATCA GTGCTCCCATCC	TGAGAACATAAATTATCT ATCAGAA	G/A

**Tabla 7.** Diseño del panel de SNPs por el software de Agena Bioscience. Recoge las 3 reacciones independientes para genotipar los 77SNP, el listado de SNP que conforman cada uno de los 3 ensayos, la secuencia de los cebadores y los polimorfismos de los SNP.

## 10-Referencias bibliográficas

1. La Mama [Internet]. [cited 2020 May 24]. Available from: <https://www.aego.es/otra-informacion/generar-automaticamente-desde-el-titulo>
2. Cormack DH. Essential Histology. 2nd Revised edition. Philadelphia: Lippincott Williams & Wilkins; 2001. 465–467.
3. Cotran R, Kumar V, Collins T. Patología estructural y funcional. sexta edición. Madrid: McGraw-Hill Interamericana; 2000. 1137–1162.
4. Provenzano E, Ulaner GA, Chin SF. Molecular Classification of Breast Cancer. Vol. 13, PET Clinics. W.B. Saunders; 2018. p. 325–38.
5. Perou CM, Sørile T, Eisen MB, van de Rijn M, Jeffrey SS, Ress CA, et al. Molecular portraits of human breast tumours. Nature. 2000 Aug 17;406(6797):747–52.
6. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, et al. Personalizing the treatment of women with early breast cancer: Highlights of the st gallen international expert consensus on the primary therapy of early breast Cancer 2013. Annals of Oncology. 2013 Sep 1;24(9):2206–23.
7. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. International Journal of Cancer [Internet]. 2019 Apr 15 [cited 2020 Jan 27];144(8):1941–53. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.31937>
8. Age standardized (World) incidence rates, breast, all ages [Internet]. 2018. Available from: <http://gco.iarc.fr/today>
9. Observatorio AECC | Observatorio AECC [Internet]. [cited 2020 Jan 27]. Available from: [http://observatorio.aecc.es/es/?\\_ga=2.3094131.2134996838.1580151588-1259126227.1579031518#datos-informes](http://observatorio.aecc.es/es/?_ga=2.3094131.2134996838.1580151588-1259126227.1579031518#datos-informes)
10. AEAC ISCIII. La situación del cáncer en España. Área de Epidemiología Ambiental y Cáncer, Centro Nacional de Epidemiología, Instituto de Salud Carlos III. Planificación Sanitaria. Madrid: Centro de Publicaciones Ministerio de Sanidad y Consumo; 2005. 188.
11. Paci E, Broeders M, Hofvind S, Duffy SW. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. Journal of Medical Screening [Internet]. 2012 Sep [cited 2020 Feb 7];19(SUPPL. 1):5–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22972806>

12. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends - An update. Vol. 25, *Cancer Epidemiology Biomarkers and Prevention*. American Association for Cancer Research Inc.; 2016. p. 16–27.
13. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: A Cancer Journal for Clinicians* [Internet]. 2016 Jan 1 [cited 2020 Feb 7];66(1):31–42. Available from: <http://doi.wiley.com/10.3322/caac.21320>
14. Winters S, Martin C, Murphy D, Shokar NK. Breast Cancer Epidemiology, Prevention, and Screening. In: *Progress in Molecular Biology and Translational Science* [Internet]. Elsevier B.V.; 2017 [cited 2020 Feb 7]. p. 1–32. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1877117317301126>
15. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis. *Cancer Epidemiology Biomarkers and Prevention*. 2006 Jun 1;15(6):1159–69.
16. Byrne C, Schairer C, Brinton LA, Wolfe J, Parekh N, Salane M, et al. Effects of mammographic density and benign breast disease on breast cancer risk (United States). *Cancer Causes and Control*. 2001;12(2):103–10.
17. Bertelsen L, Mellekjær L, Christensen J, Rawal R, Olsen JH. Age-specific incidence of breast cancer in breast cancer survivors and their first-degree relatives. *Epidemiology* [Internet]. 2009 Mar [cited 2020 Feb 8];20(2):175–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19057388>
18. Beral V, Bull D, Doll R, Peto R, Reeves G, Skegg D, et al. Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet*. 2001 Oct 27;358(9291):1389–99.
19. Índice de Desarrollo Humano | PNUD [Internet]. [cited 2020 Feb 9]. Available from: <https://desarrollohumano.org.gt/desarrollo-humano/calculo-de-idh/>
20. Moss JL, Liu B, Feuer EJ. Urban/Rural Differences in Breast and Cervical Cancer Incidence: The Mediating Roles of Socioeconomic Status and Provider Density. *Women's Health Issues* [Internet]. 2017 Nov 1 [cited 2020 Feb 9];27(6):683–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29108988>
21. Trapido EJ. Age at first birth, parity, and breast cancer risk. *Cancer*. 1983;51(5):946–8.
22. Albrektsen G, Heuch I, Hansen S, Kvåle G. Breast cancer risk by age at birth, time since birth and time intervals between births: Exploring interaction effects. *British Journal of Cancer*. 2005 Jan 17;92(1):167–75.
23. Beral V, Bull D, Doll R, Peto R, Reeves G. Breast cancer and breastfeeding: Collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50 302 women with breast cancer and 96 973 women without the

- disease. *Lancet* [Internet]. 2002 Jul 20 [cited 2020 Feb 9];360(9328):187–95. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12133652>
24. Anothaisintawee T, Wiratkapun C, Lerdsitthichai P, Kasamesup V, Wongwaisayawan S, Srinakaran J, et al. Risk factors of breast cancer: A systematic review and meta-analysis [Internet]. Vol. 25, *Asia-Pacific Journal of Public Health*. 2013 [cited 2020 Feb 9]. p. 368–87. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23709491>
  25. Carmichael AR, Bates T. Obesity and breast cancer: A review of the literature. Vol. 13, *Breast*. Churchill Livingstone; 2004. p. 85–92.
  26. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: The evidence. Vol. 174, *CMAJ. Canadian Medical Association*; 2006. p. 801–9.
  27. Liu Y, Nguyen N, Colditz GA. Links between Alcohol Consumption and Breast Cancer: A Look at the Evidence. *Women’s Health* [Internet]. 2015 Jan [cited 2020 Jan 15];11(1):65–77. Available from: <http://journals.sagepub.com/doi/10.2217/WHE.14.62>
  28. Macacu A, Autier P, Boniol M, Boyle P. Active and passive smoking and risk of breast cancer: a meta-analysis. Vol. 154, *Breast Cancer Research and Treatment*. Springer New York LLC; 2015. p. 213–24.
  29. Marjoribanks J, Farquhar C, Roberts H, Lethaby A, Lee J. Long-term hormone therapy for perimenopausal and postmenopausal women. Vol. 2017, *Cochrane Database of Systematic Reviews*. John Wiley and Sons Ltd; 2017.
  30. Palmer JR, Wise LA, Hatch EE, Troisi R, Titus-Ernstoff L, Strohsnitter W, et al. Prenatal diethylstilbestrol exposure and risk of breast cancer. *Cancer Epidemiology Biomarkers and Prevention* [Internet]. 2006 Aug [cited 2020 Feb 11];15(8):1509–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16896041>
  31. Henderson TO, Amsterdam A, Bhatia S, Hudson MM, Meadows AT, Neglia JP, et al. Surveillance for Breast Cancer in Women Treated with Chest Radiation for a Childhood, Adolescent or Young Adult Cancer: A Report from the Children’s Oncology Group. *Annals of internal medicine*. 2010;152(7):444.
  32. Arver B, Du Q, Chen J, Luo L, Lindblom A. Hereditary breast cancer: A review. *Seminars in Cancer Biology*. 2000 Aug 1;10(4):271–88.
  33. Möller S, Mucci LA, Harris JR, Scheike T, Holst K, Halekoh U, et al. The Heritability of Breast Cancer among Women in the Nordic Twin Study of Cancer. *Cancer Epidemiology Biomarkers & Prevention* [Internet]. 2016 Jan 1;25(1):145. Available from: <http://cebp.aacrjournals.org/content/25/1/145.abstract>
  34. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*. 1990;250(4988):1684–9.

35. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*. 1994 Sep 30;265(5181):2088–90.
36. Economopoulou P, Dimitriadis G, Psyrris A. Beyond BRCA: New hereditary breast cancer susceptibility genes [Internet]. Vol. 41, *Cancer Treatment Reviews*. W.B. Saunders Ltd; 2015 [cited 2020 Feb 14]. p. 1–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25467110>
37. Vega A, Campos B, Bressac-de-Paillerets B, Bond PM, Janin N, Douglas FS, et al. The R71G BRCA1 is a founder Spanish mutation and leads to aberrant splicing of the transcript. *Human Mutation* [Internet]. 2001 Jun 1;17(6):520–1. Available from: <https://doi.org/10.1002/humu.1136>
38. Antoniou AC, Pharoah PDP, McMullan G, Day NE, Ponder BAJ, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic Epidemiology* [Internet]. 2001 Jul [cited 2020 Feb 14];21(1):1–18. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11443730>
39. Antoniou AC, Pharoah PDP, McMullan G, Day NE, Stratton MR, Peto J, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *British Journal of Cancer* [Internet]. 2002 Jan 25 [cited 2020 Feb 14];86(1):76–83. Available from: <http://www.nature.com/articles/6600008>
40. Fletcher O, Houlston RS. Architecture of inherited susceptibility to common cancer. Vol. 10, *Nature Reviews Cancer*. 2010. p. 353–61.
41. Lilyquist J, Ruddy KJ, Vachon CM, Couch FJ. Common genetic variation and breast cancer Risk—Past, present, and future [Internet]. Vol. 27, *Cancer Epidemiology Biomarkers and Prevention*. American Association for Cancer Research Inc.; 2018 [cited 2020 Feb 24]. p. 380–94. Available from: <http://cebp.aacrjournals.org/lookup/doi/10.1158/1055-9965.EPI-17-1144>
42. Stoppa-Lyonnet D. The biological effects and clinical implications of BRCA mutations: Where do we go from here? Vol. 24, *European Journal of Human Genetics*. Nature Publishing Group; 2016. p. S3–9.
43. Venkitaraman AR. Functions of BRCA1 and BRCA2 in the biological response to DNA damage [Internet]. Vol. 114, *Journal of Cell Science*. 2001 [cited 2020 Feb 18]. Available from: [http://www.nhgri.nih.gov/Intramural\\_research/](http://www.nhgri.nih.gov/Intramural_research/)
44. Schon K, Tischkowitz M. Clinical implications of germline mutations in breast cancer: TP53 [Internet]. Vol. 167, *Breast Cancer Research and Treatment*. Springer New York LLC; 2018 [cited 2020 Feb 18]. p. 417–23. Available from: <http://link.springer.com/10.1007/s10549-017-4531-y>
45. Li S, Shen Y, Wang M, Yang J, Lv M, Li P, et al. Loss of PTEN expression in breast cancer: Association with clinicopathological characteristics and prognosis. *Oncotarget*

- [Internet]. 2017 May 9 [cited 2020 Feb 18];8(19):32043–54. Available from: <http://www.oncotarget.com/fulltext/16761>
46. Daniell J, Plazzer JP, Perera A, Macrae F. An exploration of genotype-phenotype link between Peutz-Jeghers syndrome and STK11: a review. Vol. 17, *Familial Cancer*. Springer Netherlands; 2018. p. 421–7.
  47. Miguel-Soca P, Argüelles González I, Peña González M. Factores genéticos en la carcinogénesis mamaria. *Revista Finlay*. 2016;6:299–316.
  48. Wendt C, Margolin S. Identifying breast cancer susceptibility genes – a review of the genetic background in familial breast cancer. *Acta Oncologica* [Internet]. 2019 Feb 1 [cited 2020 Feb 19];58(2):135–46. Available from: <https://www.tandfonline.com/doi/full/10.1080/0284186X.2018.1529428>
  49. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: Different roles in a common pathway of genome protection. Vol. 12, *Nature Reviews Cancer*. 2012. p. 68–78.
  50. Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute*. 2015 May 1;107(5).
  51. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*. 2019 Jan 3;104(1):21–34.
  52. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *GENETICS in MEDICINE* [Internet]. 2019;21:1708–18. Available from: <https://doi.org/10.1038/s41436->
  53. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* [Internet]. 1977 Dec;74(12):5463–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/271968>
  54. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Vol. 107, *Genomics*. Academic Press Inc.; 2016. p. 1–8.
  55. Shendure J, Ji H. Next-generation DNA sequencing. Vol. 26, *Nature Biotechnology*. 2008. p. 1135–45.
  56. Rodríguez-Santiago B, Armengol L. Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. *Diagnostico Prenatal*. 2012 Apr 1;23(2):56–66.
  57. Oeth P, Beaulieu M, Park C, Kosman D, del Mistro G, van den Boom D, et al. iPLEX™ Assay: Increased Plexing Efficiency and Flexibility for MassARRAY® System Through Single Base Primer Extension with Mass-Modified Terminators. *Sequenom Application Note Doc*. 2007;No. 8876-006(R05 CO):070145.

58. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: Practice, problems and promise. *Human Mutation* [Internet]. 2001 Apr [cited 2020 Apr 9];17(4):296–304. Available from: <http://doi.wiley.com/10.1002/humu.27>
59. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the sequenom massARRAY iPLEX Platform. *Current Protocols in Human Genetics*. 2009.
60. Peake I. The polymerase chain reaction. *Journal of Clinical Pathology*. 1989;42(7):673–6.
61. Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Research*. 1997;7(4):378–88.
62. Hines LM, Risendal B, Slattery ML, Baumgartner KB, Giuliano AR, Byers T. Differences in estrogen receptor subtype according to family history of breast cancer among hispanic, but not non-hispanic white women. *Cancer Epidemiology Biomarkers and Prevention*. 2008 Oct;17(10):2700–6.
63. Fejerman L, John EM, Huntsman S, Beckman K, Choudhry S, Perez-Stable E, et al. Genetic ancestry and risk of breast cancer among U.S. Latinas. *Cancer Research*. 2008 Dec 1;68(23):9723–8.
64. Fejerman L, Hu D, Huntsman S, John EM, Stern MC, Haiman CA, et al. Genetic ancestry and risk of mortality among U.S. latinas with breast cancer. *Cancer Research*. 2013 Dec 15;73(24):7243–53.
65. Proyecto Breogan – Proyecto Breogan [Internet]. [cited 2020 May 25]. Available from: <http://proyectobreogan.es/#que-hacemos>
66. Cruz GI, Martínez ME, Natarajan L, Wertheim BC, Gago-Dominguez M, Bondy M, et al. Hypothesized role of pregnancy hormones on HER2+ breast tumor development. *Breast Cancer Research and Treatment*. 2013 Jan;137(1):237–46.
67. Jiang X, Castelao JE, Chavez-Uribe E, Rodriguez BF, Muñoz CC, Redondo CM, et al. Family history and breast cancer hormone receptor status in a Spanish cohort. *PLoS ONE*. 2012 Jan 6;7(1).
68. Redondo CM, Gago-Domínguez M, Ponte SM, Castelo ME, Jiang X, García AA, et al. Breast feeding, parity and breast cancer subtypes in a Spanish cohort. *PLoS ONE*. 2012 Jul 11;7(7).
69. Ali AMG, Schmidt MK, Bolla MK, Wang Q, Gago-Dominguez M, Esteban Castelao J, et al. Alcohol consumption and survival after a breast cancer diagnosis: A literature-based meta-analysis and collaborative analysis of data for 29,239 cases. *Cancer Epidemiology Biomarkers and Prevention*. 2014;23(6):934–45.
70. Rudolph A, Milne RL, Truong T, Knight JA, Seibold P, Flesch-Janys D, et al. Investigation of gene-environment interactions between 47 newly identified breast

- cancer susceptibility loci and environmental risk factors. *International Journal of Cancer*. 2015 Mar 15;136(6):E685–96.
71. Martínez ME, Gutiérrez-Millan LE, Bondy M, Daneri-Navarro A, Meza-Montenegro MM, Anduro-Corona I, et al. Comparative study of breast cancer in Mexican and Mexican-American women. *Health* [Internet]. 2010 Sep 29 [cited 2020 Mar 31];02(09):1040–8. Available from: <http://www.scirp.org/journal/doi.aspx?DOI=10.4236/health.2010.29153>
  72. Dabbs DJ, Schnitt SJ, Geyer FC, Weigelt B, Baehner FL, Decker T, et al. Lobular neoplasia of the breast revisited with emphasis on the role of e-cadherin immunohistochemistry. *American Journal of Surgical Pathology*. 2013 Jul;37(7).
  73. HF F, RA W, KW B, DW F, MJ G, JC B, et al. Interobserver Reproducibility of the Nottingham Modification of the Bloom and Richardson Histologic Grading Scheme for Infiltrating Ductal Carcinoma. *American journal of clinical pathology*. 1995;103(2).
  74. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*. 2019 Jan 3;104(1):21–34.
  75. GRCh38 - hg38 - Genome - Assembly - NCBI [Internet]. [cited 2020 May 25]. Available from: [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.26/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/)
  76. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*. 2013 Apr;45(4):353–61.
  77. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature Genetics*. 2015 Apr 28;47(4):373–80.
  78. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017 Nov 2;551(7678):92–4.
  79. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nature genetics* [Internet]. 2013 Apr;45(4):392-398e3982. Available from: <https://pubmed.ncbi.nlm.nih.gov/23535733>
  80. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007 Jun 28;447(7148):1087–93.
  81. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nature Genetics*. 2009 May;41(5):579–84.
  82. Milne RL, Benítez J, Nevanlinna H, Heikkinen T, Aittomäki K, Blomqvist C, et al. Risk of Estrogen Receptor–Positive and –Negative Breast Cancer and Single–

- Nucleotide Polymorphism 2q35-rs13387042. JNCI: Journal of the National Cancer Institute [Internet]. 2009 Jul 15;101(14):1012–8. Available from: <https://doi.org/10.1093/jnci/djp167>
83. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor–positive breast cancer. *Nature Genetics* [Internet]. 2007;39(7):865–9. Available from: <https://doi.org/10.1038/ng2064>
  84. Ghossaini M, Pharoah PDP. Polygenic susceptibility to breast cancer: Current state-of-the-art. Vol. 5, *Future Oncology*. Future Oncol; 2009. p. 689–701.
  85. Dickson C, Spencer-Dene B, Dillon C, Fantl V. Tyrosine kinase signalling in breast cancer fibroblast growth factors and their receptors. Vol. 2, *Breast Cancer Research*. Breast Cancer Res; 2000. p. 191–6.
  86. Moffa AB, Tannheimer SL, Ethier SP. Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. *Molecular Cancer Research*. 2004 Nov;2(11):643–52.
  87. Adnane J, Gaudray P, Dionne CA, Crumley G, Jaye M, Schlessinger J, et al. BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. *Oncogene* [Internet]. 1991 Apr [cited 2020 May 20];6(4):659–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1851551>
  88. Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF, Caldas C, et al. Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. *PLoS Biology*. 2008 May;6(5):1098–103.
  89. Udler MS, Meyer KB, Pooley KA, Karlins E, Struewing JP, Zhang J, et al. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Human Molecular Genetics* [Internet]. 2009 Feb 17;18(9):1692–703. Available from: <https://doi.org/10.1093/hmg/ddp078>
  90. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, NY)* [Internet]. 2007;316(5826):889–94. Available from: <http://europepmc.org/abstract/MED/17434869>
  91. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* [Internet]. 2007;445(7130):881–5. Available from: <https://doi.org/10.1038/nature05616>
  92. Jingushi K, Takahashi-Yanaga F, Yoshihara T, Shiraishi F, Watanabe Y, Hirata M, et al. DIF-1 inhibits the Wnt/ $\beta$ -catenin signaling pathway by inhibiting TCF7L2 expression in colon cancer cell lines. *Biochemical Pharmacology*. 2012 Jan 1;83(1):47–56.

93. O'Flaherty E, Kaye J. TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics* [Internet]. 2003 Apr 2 [cited 2020 May 20];4(1):13. Available from: <https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-4-13>
94. Smid M, Wang Y, Klijn JGM, Sieuwerts AM, Zhang Y, Atkins D, et al. Genes Associated With Breast Cancer Metastatic to Bone. *Journal of Clinical Oncology* [Internet]. 2006 May 20;24(15):2261–7. Available from: <https://doi.org/10.1200/JCO.2005.03.8802>
95. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genetics*. 2008 Jun;40(6):703–6.
96. Koc EC, Burkhardt W, Blackburn K, Moseley A, Spremulli LL. The small subunit of the mammalian mitochondrial ribosome: Identification of the full complement of ribosomal proteins present. *Journal of Biological Chemistry*. 2001 Jun 1;276(22):19363–74.
97. Koc EC, Burkhardt W, Blackburn K, Moyer MB, Schlatzer DM, Moseley A, et al. The large subunit of the mammalian mitochondrial ribosome: Analysis of the complement of ribosomal proteins present. *Journal of Biological Chemistry*. 2001 Nov 23;276(47):43958–69.
98. Ghossaini M, French JD, Michailidou K, Nord S, Beesley J, Canisus S, et al. Evidence that the 5p12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGF10 and MRPS30 Regulation. *American Journal of Human Genetics*. 2016 Oct 6;99(4):903–11.
99. Warren H, Dudbridge F, Fletcher O, Orr N, Johnson N, Hopper JL, et al. 9q31.2-rs865686 as a susceptibility locus for estrogen receptor-positive breast cancer: Evidence from the Breast Cancer Association Consortium. *Cancer Epidemiology Biomarkers and Prevention*. 2012;21(10):1783–91.
100. Palmer JR, Ambrosone CB, Olshan AF. A collaborative study of the etiology of breast cancer subtypes in African American women: The AMBER consortium. *Cancer Causes and Control*. 2014;25(3):309–19.
101. Rudolph A, Milne RL, Truong T, Knight JA, Seibold P, Flesch-Janys D, et al. Investigation of gene-environment interactions between 47 newly identified breast cancer susceptibility loci and environmental risk factors. *International journal of cancer* [Internet]. 2014/09/23. 2015 Mar 15;136(6):E685–96. Available from: <https://pubmed.ncbi.nlm.nih.gov/25227710>
102. Hojilla C v, Wood GA, Khokha R. Inflammation and breast cancer: metalloproteinases as common effectors of inflammation and extracellular matrix breakdown in breast cancer. *Breast cancer research : BCR* [Internet]. 2008/03/31. 2008;10(2):205. Available from: <https://pubmed.ncbi.nlm.nih.gov/18394187>

103. Baird DM. Variation at the TERT locus and predisposition for cancer. *Expert Reviews in Molecular Medicine* [Internet]. 2010/05/18. 2010;12:e16. Available from: <https://www.cambridge.org/core/article/variation-at-the-tert-locus-and-predisposition-for-cancer/D9F3BB87B9AB2E9E82E6F8811B575FAF>
104. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor $\alpha$ -negative breast cancer. *Nature Genetics*. 2011 Dec;43(12):1210–4.
105. Milatovich A, Qiu RG, Grosschedl R, Francke U. Gene for a tissue-specific transcriptional activator (EBF or Olf-1), expressed in early B lymphocytes, adipocytes, and olfactory neurons, is located on human Chromosome 5, band q34, and proximal mouse Chromosome 11. *Mammalian Genome*. 1994 Apr;5(4):211–5.
106. EBF1 Gene - GeneCards | COE1 Protein | COE1 Antibody [Internet]. [cited 2020 May 25]. Available from: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=EBF1&keywords=ebf1>
107. RAD51B Gene - GeneCards | RA51B Protein | RA51B Antibody [Internet]. [cited 2020 May 25]. Available from: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=RAD51B>
108. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang Y bin, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nature Genetics*. 2009 Mar;41(3):324–8.