



**Universidad de Oviedo**

**Instituto Universitario de Oncología del Principado de Asturias**

**Máster en Biomedicina y Oncología Molecular**

# **Análisis de haplotipos en familias con alteraciones en genes implicados en cáncer hereditario**

**Raquel Capín Pérez**

**Julio 2014**

**Trabajo Fin de Máster**



# ÍNDICE

Resumen.....	1
1. Introducción.....	2
1.1.  Cáncer hereditario.....	3
1.2.  Cáncer de mama y <i>BRCA</i> .....	4
1.3.  Funciones de <i>BRCA</i> .....	5
1.4.  Cáncer de mama hereditario y mutaciones fundadoras.....	7
1.5.  Haplotipos.....	8
1.6.  Cáncer de mama hereditario en España.....	8
2. Objetivos.....	11
3. Material y Métodos.....	12
3.1.  Pacientes.....	12
3.2.  Determinación de mutaciones en <i>BRCA2</i> .....	13
3.3.  Genotipado de microsatélites y SNPs.....	13
3.4.  Análisis de los datos de genotipado y generación de haplotipos.....	18
3.5.  Cálculo de la antigüedad de la mutación.....	19
4. Resultados.....	21
4.1.  Genotipado y obtención de haplotipos .....	21
4.2.  Cálculo de la edad de la mutación.....	28
5. Discusión.....	31
6. Conclusiones.....	37
7. Bibliografía.....	38

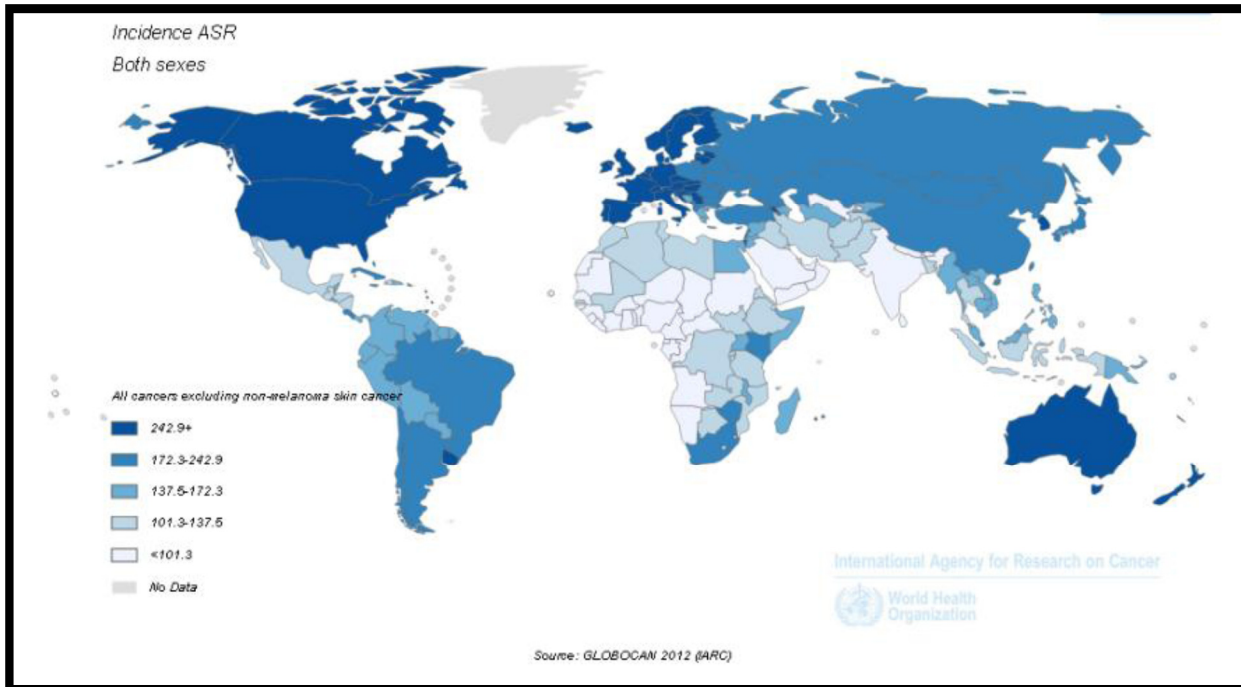
## Resumen

Las alteraciones en *BRCA1* y *BRCA2* alcanzan el 20-40% en los casos de cáncer de mama hereditario, por lo que se las considera los factores de riesgo genético más significativos para identificar los casos de cáncer de mama y/u ovario. Algunas mutaciones en estos genes parecen distribuirse de forma particular en cada población y comunidades étnicas. En un estudio previo sobre cáncer de mama hereditario en la población asturiana se describieron cinco mutaciones noveles en los genes *BRCA1* y *BRCA2*. El objetivo del presente trabajo ha sido determinar los haplotipos en el entorno del gen *BRCA2* en las familias con las dos mutaciones novel en este gen, con el fin de conocer si existe un haplotipo único relacionado con cada mutación que sugiera un efecto fundador. Se analizaron un total de 9 marcadores (6 SNPs y 3 microstélites) en una región genómica cercana a *BRCA2* en 65 individuos pertenecientes a 9 familias independientes y 42 individuos control. A partir de los datos de genotipado se dedujeron los haplotipos, utilizando el programa PHASE2.0. Se ha podido deducir un haplotipo específico de los portadores de la mutación 1, presente en 23 de 24 portadores. Se ha detectado un haplotipo relacionado con la mutación 2, el cual presenta varios eventos de recombinación dentro de las 3 familias estudiadas, por lo que no se puede establecer un haplotipo fundador con seguridad. La identificación de las mutaciones fundadoras y de aquellas que mantienen una mayor prevalencia en cada población y el conocimiento de su distribución geográfica pueden ser de utilidad a la hora de diseñar métodos de *screening* o cribado mutacional eficientes y de ofrecer consejo genético a las familias.

## 1. Introducción

El cáncer se constituye, junto con las enfermedades cardiovasculares, como la principal causa de muerte en los países industrializados. La aparición de células cancerosas es, generalmente, producto de una acumulación progresiva de mutaciones en el DNA, tanto genéticas como epigenéticas, que otorgan a estas células la capacidad de crecer de forma descontrolada y excesiva, invadiendo y dañando tejidos y órganos circundantes, así como también pueden adquirir capacidad invasiva que les permita migrar a zonas distantes donde metastatizar. Estas características ofrecen a las células tumorales una cierta ventaja selectiva ante aquellas células sanas; esto se debe a que presentan mutaciones en genes tumorales implicados en el control del crecimiento y la reparación de daños a nivel genómico, principalmente. Muchos de los distintos tipos de cáncer se pueden prevenir evitando la exposición a agentes carcinogénicos, como el humo del tabaco; además, existe una proporción significativa de casos en los que se puede curar si son detectados en estadios tempranos, mediante cirugía, radioterapia y quimioterapia (fuente: Organización Mundial de la Salud).

En la actualidad, la tasa global de incidencia del cáncer estandarizada por edad es un 25% mayor en hombres que en mujeres, con tasa de 205 y 165 por 100000, respectivamente. La mortalidad en hombres en los países más desarrollados es un 15% mayor que en los menos industrializados, y un 8% mayor en el caso de las mujeres (Figura 1.1). Asimismo, en hombres los tipos de tumores más frecuentes son de pulmón, próstata, colon, estómago e hígado, mientras que las mujeres presentan mayor incidencia en cáncer de mama, colon, cérvix, pulmón y útero (fuente: GLOBOCAN (IARC) 2012). Es por esto por lo que ha cobrado cada vez más relevancia la investigación oncológica, *screening* y seguimiento de los pacientes afectados y de la población en riesgo; en el caso de las mujeres, en particular, los esfuerzos se han centrado en la detección y tratamiento del cáncer de mama (Cáncer hereditario, SEOM, 2006; Lakhani *et al.*, 1998).



**Figura 1.1.** Estimación de la incidencia de cáncer a nivel mundial (fuente: GLOBOCAN (IARC) 2012).

### 1.1. CÁNCER HEREDITARIO

La mayoría de los tumores se desarrolla por la aparición de distintas mutaciones en células somáticas, pero solo en caso de que se produzca una mutación en línea germinal tendrá lugar el conocido como cáncer hereditario. Éste se desarrolla debido a la adquisición de una copia mutada de un determinado gen a través de los progenitores. Los genes causantes del cáncer hereditario por una mutación son proto-oncogenes, genes supresores de tumores o genes encargados de la reparación, integridad o estabilidad del ADN, siendo estas mutaciones altamente penetrantes. De esta forma, individuos que presenten tan solo una copia normal de un gen implicado en el desarrollo tumoral tendrán una fuerte predisposición genética a desarrollar determinados tipos de cáncer a una edad temprana. La identificación de las familias afectadas por un síndrome de predisposición al cáncer (SPC) es muy importante para un adecuado asesoramiento genético, ya que se estima que entre el 5 y el 10% de los tumores tienen como causa la herencia familiar (Cáncer hereditario, SEOM, 2006).

## 1.2. CÁNCER DE MAMA Y BRCA

El cáncer de mama se establece como la afección maligna más común entre las mujeres de países industrializados, diferenciándose cuatro subgrupos en función de los perfiles de expresión génica y sus características fenotípicas diferenciales: luminal, HER2, normal y *basal-like* (o triple negativo). Esta división se realizó en un principio mediante estudios inmunohistoquímicos y, posteriormente, mediante estudios moleculares. Dependiendo del fenotipo tumoral la respuesta a los tratamientos es diferente, por lo que la clasificación tumoral es relevante en la toma de decisiones terapéuticas (De Abreu *et al.*, 2013).

Los principales genes de susceptibilidad, que incrementan diez veces el riesgo de desarrollar cáncer de mama/ovario respecto al resto de la población, son *BRCA1*, *BRCA2* y *TP53*. A día de hoy, los genes asociados a un mayor número de casos de cáncer de mama y ovario hereditario son *BRCA1* y *BRCA2*. Estos genes apenas se encuentran mutados en los casos de cáncer de mama esporádico; en cambio, hasta el 40% de los casos de cáncer de mama y ovario familiar presentan mutaciones en uno de estos dos genes. La mayoría de los casos dan lugar a carcinomas ductales invasivos, mientras que los tumores de mama no debidos a estos genes tienden a presentar grados histológicos e índices mitóticos menores que los asociados a la mutación (Lakhani *et al.*, 1998).

Desde un punto de vista histórico, la demostración de la existencia de un supuesto gen causante de la predisposición genética para el desarrollo temprano de cáncer de mama hereditario, *BRCA1*, fue llevada a cabo por más de una docena de laboratorios entre 1974 y 1990, estimando su localización en el locus 17q21 (King, 2014; Hall *et al.* 1990). Durante los siguientes 4 años, múltiples laboratorios intentaron determinar la localización exacta y lograr la clonación posicional del gen de susceptibilidad. El grupo de M.C. King estudió 1579 familias y analizó los resultados obtenidos mediante análisis de segregación complejos, obteniendo como resultado un supuesto gen de susceptibilidad al cáncer de alta penetrancia y de herencia autosómica dominante. En este momento, el Proyecto Genoma Humano aún no había comenzado y, por ello, la búsqueda del gen se estaba llevando a cabo sin disponer de una secuencia genómica de referencia, lo que hacía un reto pretender mapearlo. En cuanto se logró reducir la región donde

se encontraba el supuesto gen, se pudo comenzar a clonar en cromosomas artificiales de levadura (YACs) y bacterias (BACs) y cósmidos. Finalmente, se lograron secuenciar 15 genes comprendidos en una región de 1Mb. De estos 15 genes, 5 ya se conocían pero no habían sido mapeados previamente y los otros 10 no se conocían, pero ninguno era *BRCA1*. A pesar de que el laboratorio de King fuera el primero en evidenciar la existencia de dicho gen, el gen fue clonado finalmente por el grupo de Skolnick de la Universidad de Utah en 1994 (Miki *et al.*, 1994). Ese mismo año, Michael Stratton y Richard Wooster clonaron el gen *BRCA2* en el *Institute of Cancer Research* en Reino Unido (King, 2014).

### 1.3. FUNCIONES DE *BRCA*

Desde su clonación hasta la actualidad, son numerosos los estudios encaminados a descifrar las funciones biológicas de *BRCA1* y *BRCA2*. Ambas proteínas son las encargadas de mantener la estructura e integridad numérica de los cromosomas durante el ciclo celular, por lo que funcionalmente se les puede considerar supresores tumorales (Venkitaraman, 2014). La mutagénesis dirigida de ambas copias de *BRCA1* o *BRCA2* en línea germinal de ratón demostró la letalidad temprana embrional e impidió la proliferación celular (Liu *et al.*, 1996; Sharan *et al.*, 1997; Connor *et al.*, 1997), además de presentar una hipersensibilidad a las genotoxinas que afectaban a la migración de estas proteínas al núcleo (Sharan *et al.*, 1997; Patel *et al.*, 1998; Scully *et al.*, 1997; Chen *et al.*, 1998). Las células deficientes en *BRCA2* muestran translocaciones, grandes deleciones o fusiones que afectan a varios cromosomas no homólogos, y tienen como consecuencia la mala segregación de los mismos (Yu *et al.*, 2000); las células deficientes para *BRCA1* presentan efectos similares (Shen *et al.*, 1998). Estos datos indican que la inactivación de *BRCA* fomenta la carcinogénesis por inestabilidad cromosomal.

A día de hoy, aún no se conoce con fiabilidad cómo *BRCA1* y *BRCA2* protegen la integridad de los cromosomas, aunque se han observado múltiples interacciones de estas proteínas con otras. Se cree por ello que puedan pertenecer a un pequeño conjunto de proteínas que sirven como centros activos dinámicos para múltiples complejos macromoleculares. Se ha sugerido que estas grandes proteínas actúan como entidades segmentadas, donde la distintas regiones, y a veces



regiones intrínsecamente desordenadas, pueden interactuar físicamente de diferentes formas para llevar a cabo distintas funciones biológicas; aunque no todas sus funciones tienen por qué ser relevantes para la supresión tumoral (Venkitaraman, 2014).

Las aberraciones cromosómicas estructurales se acumulan de forma espontánea durante la división de las células deficientes en *BRCA*, donde las rupturas de doble hebra de DNA (*double strand breaks*) y los huecos de las hebras hija (*double strand gaps*) son las principales lesiones celulares (Cox, 2000). La recombinación homóloga es la principal solución para estos problemas. Tanto *BRCA1* como *BRCA2* son esenciales para que la recombinación homóloga sea eficiente en las células de mamíferos (Moynahan *et al.*, 1999; Moynahan *et al.*, 2001). En su ausencia, las roturas en el DNA asociadas a la replicación pueden ser reparadas por mecanismos propensos a errores como los de unión de extremos no homólogos (*non homologous end joining*) y microhomología mediada por unión de extremos (Patel *et al.*, 1998; Yu *et al.*, 2000; Moynahan *et al.*, 2001), mecanismos que ligan los extremos rotos, a menudo mediante microhomologías. De esta forma, aquellos tumores que tengan mutaciones en homocigosis para los genes de interés, además de presentar grandes reordenamientos estructurales a nivel cromosomal, también tendrán muchas deleciones cortas (50 pb) por la superposición de microhomologías en los puntos de rotura (Alexandrov *et al.*, 2013).

Cabe señalar que se pueden observar características particulares ante la pérdida de heterocigosidad en los cánceres cuya mutación está en el gen *BRCA1* frente a los que al presentan en *BRCA2*. Los portadores de una mutación en *BRCA1* presentan un marcado exceso en cánceres de ovario y mama en mujeres, mientras que aquellas mutaciones hereditarias en *BRCA2*, además, tienen una importante predisposición a cánceres de mama en hombres, páncreas, próstata y otros órganos (Breast Cancer Linkage Consortium, 1999; Thompson & Easton, 2001). Por otra parte, los cánceres de mama debidos a deficiencias en *BRCA1* tienden a presentar una predominancia en las características del subgrupo *basal-like*, así como una falta de expresión del receptor de estrógenos que no se observa en aquellos tumores con deficiencias en *BRCA2* (Lucas *et al.*, 2013).

#### 1.4. CÁNCER DE MAMA HEREDITARIO Y MUTACIONES FUNDADORAS

Las alteraciones en *BRCA1* y *BRCA2* alcanzan el 20-40% en los casos de cáncer de mama hereditario (Wooster & Weber, 2003; Fackenthal & Olopade, 2007), lo que convierte a las mutaciones en dichos genes en los factores de riesgo genético más significativos para identificar los casos de cáncer de mama y/u ovario hasta la fecha. El riesgo de desarrollar cáncer de mama asociado a una mutación de *BRCA* en personas con riesgo genético a la edad de 70 años asciende al 85% y para cáncer de ovario es del 44% (Ford *et al.*, 1998).

La mayoría de las mutaciones descritas hasta la fecha en *BRCA1* y *BRCA2* son debidas a cambios en la pauta de lectura, mutaciones *nonsense* y alteraciones en los sitios de *splice* que provocan el truncamiento de las proteínas *BRCA1* y *BRCA2*. Se han descrito más de 3000 mutaciones en estos dos genes y, aproximadamente, el 60% de las alteraciones se han detectado en una sola familia (fuente: *Breast Information Core, BIC*). Algunas mutaciones en estos genes, sin embargo, parecen distribuirse de forma particular en cada población y comunidades étnicas (Szabo & King, 1997). Así, se han descrito mutaciones fundadoras en distintas poblaciones, estando entre las más estudiadas las procedentes de los judíos Ashkenazi (Rubinstein, 2004). También se han descrito mutaciones fundadoras en distintas regiones europeas, americanas y asiáticas (Campos *et al.*, 2003; Ferla *et al.*, 2007; Janavicius, 2010; Kwong *et al.*, 2012). La identificación y determinación de la distribución geográfica de las mutaciones fundadoras y de aquellas que mantienen una mayor prevalencia pueden ser de utilidad a la hora de diseñar métodos de *screening* o cribado mutacional eficientes y de ofrecer consejo genético a las familias (Díez *et al.*, 2010; Janavicius *et al.*, 2013).

El efecto fundador es detectable mediante estudios del desequilibrio de ligamiento (DL) existente en determinadas regiones del genoma. El desequilibrio de ligamiento o asociación alélica se define como un exceso en la co-aparición, dependencia estadística, de dos alelos por encima de lo esperado si los dos alelos fueran independientes (Shifman & Darvasi, 2001). De esta forma, en una población aislada que se haya constituido a partir de unos pocos individuos fundadores, la deriva genética reducirá el número de mutaciones causantes de enfermedad en dicha población (Daly *et al.*, 2001). Dichas mutaciones se convierten en "mutaciones fundadoras" dado que todos

los individuos de la población aislada portadores de la mutación serán descendientes del individuo fundador.

### 1.5. HAPLOTIPOS

La combinación particular de los alelos de los loci junto con la mutación/alelo de la enfermedad de estudio constituyen un haplotipo (Genotipo haploide) fundador; un haplotipo es un conjunto de alelos en fase. Al conjunto de haplotipos presentes en la población se le conoce como haplogrupo (Gabriel *et al.*, 2002; Phillips *et al.*, 2003). La existencia de este desequilibrio de ligamiento entre distintos marcadores hace posible conocer la historia de la población y la localización de las variantes genéticas subyacentes a caracteres complejos (Zhao *et al.*, 2003).

Cabe decir que, a medida que se va transmitiendo, el cromosoma que contiene la mutación fundadora con su respectivo haplotipo irá sufriendo cambios exponenciales en cada generación, al verse afectado por distintos factores como la edad de las variantes, la historia poblacional, las tasas de recombinación, la conversión génica y la selección natural entre otros factores (Zhao *et al.*, 2003). De esta forma, el DL en el que se encuentran la mutación y los marcadores específicos se irá deteriorando a causa de la frecuencia de recombinación. Es por esto que, en poblaciones donde la mutación de estudio se haya establecido recientemente, los individuos portadores de dicha mutación compartirán una región cromosomal de mayor tamaño (millones de pares de bases) que en aquellas poblaciones donde haya existido durante más tiempo una mutación concreta.

### 1.6. CÁNCER DE MAMA HEREDITARIO EN ESPAÑA

En España, se ha descrito el espectro mutacional de *BRCA1* y *BRCA2* en familias con síndrome de cáncer de mama y ovario procedentes de numerosas regiones. Se han encontrado distintas frecuencias de mutaciones descritas y también nóveles en Galicia (Vega *et al.*, 2002), Salamanca (Salaza *et al.*, 2006), Castilla y León (Infante *et al.*, 2006), País Vasco (Beristain *et al.*, 2007), Aragón (Miramar *et al.*, 2008), Cataluña (Llort *et al.*, 2002), Valencia (Esteban-Cardena *et al.*,

2008) y Asturias (Blay *et al.*, 2013). Las discrepancias entre las distintas regiones reflejan la heterogeneidad geográfica de las familias, aumentando la variedad de mutaciones que se pueden encontrar en las poblaciones españolas y modificando las frecuencias relativas de mutaciones particulares dentro de cada grupo de estudio (Díez *et al.*, 2010).

En el caso de Asturias, se ha llevado a cabo un estudio mutacional en 256 familias de alto riesgo. Los criterios de selección para llevar a cabo el estudio en *BRCA1* y 2 de los casos índices, afectados de la enfermedad en estas familias fueron uno de los siguientes:

1. Que en la familia haya tres o más miembros con cáncer de mama y/o cáncer de ovario.
2. Pertener a una familia con dos afectadas con cáncer de ovario.
3. Pertener a una familia con una afectada por cáncer de ovario y una con cáncer de mama.
4. Uno de los miembros de la familia con un cáncer de mama masculino y uno o más con cáncer de mama y/o cáncer de ovario.
5. Pertener a una familia con dos afectadas con cáncer de mama antes de los 50 años.
6. Pertener a una familia con una afectada con cáncer de mama bilateral y otra con cáncer de mama, siendo la edad de aparición del cáncer en al menos una de ellas antes de los 50 años.
7. Una única afectada ya sea con: cáncer de mama bilateral antes de los 40 años, cáncer de mama u ovario antes de los 30 años.
8. Probandos afectados de familias cerca de cumplir alguno de los criterios anteriores, con tumores de fenotipo triple negativo.

De las 256 familias estudiadas, 39 presentaron mutación en *BRCA1* y 20 en *BRCA2*. La mutación en *BRCA1* más frecuentemente detectada es c.211A>G (p.Arg71Gly). Esta es una mutación de origen fundador en Galicia (Vega *et al.*, 2001; Vega *et al.*, 2002), presente en 9 de las 39 familias (23% de las que presentan este gen mutado) y se explica por la proximidad de Galicia a Asturias y la ascendencia gallega de las personas estudiadas. Otra de las mutaciones en *BRCA1*

frecuentemente detectada en Asturias es c.3331\_3334delCAAG (p.Gln1111Asnfs), afectando a 6 familias (15%). Es una mutación ya descrita en 1996 en una familia canadiense (Durocher *et al.*, 1996), pero posteriores estudios sugieren que su origen pueda ser español (Torres *et al.*, 2007). Las nuevas mutaciones descritas en *BRCA1* en las familias de Asturias son: c.2900\_2901dupCT (p.Pro968Leufs), presente en 8 familias (20% de las familias con *BRCA1* mutado) aparentemente no relacionadas; c.1674delA (p.Gly559Valfs) y c.1965C>A (p.Tyr655\*), presentes en una familia cada una.

En relación a *BRCA2*, las dos mutaciones más frecuentes que se han encontrado en las familias analizadas en Asturias son c.4030\_4035delinsC (p.Asn1344Hisfs), no descrita anteriormente y que está presente en 6 de las 20 familias con *BRCA2* mutado, y c.2095C>T (p.Gln699\*), que también es una mutación nueva y está presente en 3 de las 20 familias ya mencionadas.

Es posible que la existencia de estas mutaciones no descritas con anterioridad y solo presentes en Asturias en un número alto de familias pueda deberse al histórico aislamiento geográfico de la región. En particular, las ocho familias con la mutación en *BRCA1* c.2900\_2901dupCT (p.Pro968Leufs) tienen un origen geográfico en una zona concreta de Asturias, en una franja geográfica relacionada con la calzada romana de La Mesa. Por otra parte, en relación a la mutación p.Asn1344Hisfs de *BRCA2*, se puede destacar que las 6 familias portadoras de esta alteración tienen su origen en una zona reducida del occidente asturiano, entre dos valles (ríos Eo y Navia) y donde, además, existe un dialecto propio (eonaviego) que apoya la existencia de aislamiento de dicha población (Blay *et al.*, 2013). Estos datos sugieren que estas mutaciones sean fundadoras en esta población.

De acuerdo con la hipótesis anterior, el objetivo del presente trabajo ha sido determinar los haplotipos en el entorno del gen *BRCA2* en las familias con las dos mutaciones novel en este gen, con el fin de conocer si existe un haplotipo único relacionado con cada mutación que sugiera un efecto fundador.

## 2. Objetivos

A partir de un estudio previo sobre cáncer de mama hereditario en la población asturiana (Blay *et al.*, 2013) en el que se reportan cinco mutaciones nóveles en los genes *BRCA1* y *BRCA2*, se considera que el estudio y determinación de los haplotipos asociados a dichas mutaciones podría resultar de interés para demostrar el efecto fundador de las mutaciones mencionadas, poniendo de manifiesto el aislamiento geográfico y genético que ha sufrido Asturias a lo largo de la Historia. Por ello, se propone el estudio de los haplotipos de las dos mutaciones presentes en el gen *BRCA2*, p.Asn1344Hisfs (6 familias) y p.Gln699\* (3 familias), junto al estudio de una población control. Asimismo, también se pretende calcular cuántas generaciones transcurrieron desde que surgieron ambas mutaciones. Específicamente, los objetivos que se han planteado para el presente proyecto son:

1. Seleccionar una serie de marcadores moleculares en el entorno del gen *BRCA2*.
2. Realizar un estudio de genotipado con estos marcadores en individuos portadores de las mutaciones de *BRCA2* p.Asn1344Hisfs y p.Gln699\*, en familiares no portadores y en población control.
3. Analizar los datos de genotipado, inferir los haplotipos de cada individuo y determinar el haplotipo fundador.
4. Calcular la antigüedad de las mutaciones a partir de los datos obtenidos.

### 3. Material y métodos

#### 3.1. PACIENTES

Se han analizado 6 familias en las que se identificó la mutación que denominaremos Mutación 1 [*BRCA2* c.4030\_4035delinsC (p.Asn1344Hisfs)]. Dentro de estas 6 familias se estudiaron 24 portadores de la mutación y 20 familiares de los afectados no portadores de la misma. Para la Mutación 2, [*BRCA2* c.2095C>T (p.Gln699\*)], se han estudiado 3 familias, con 11 afectados por la mutación, y 10 individuos familiares no portadores de dicha mutación. En la tabla 3.1 se resumen los casos analizados por familias. Todas las personas analizadas habían firmado el consentimiento informado para realizar estudios genéticos y para participar en estudios de investigación derivados de los mismos, tras asesoramiento en la consulta de Cáncer Familiar del HUCA.

MUTACIÓN	FAMILIA	INDIVIDUOS CON MUTACIÓN	FAMILIARES NO PORTADORES
1	27	4	4
	208	8	4
	331	3	0
	424	5	5
	609	3	4
	795	1	3
2	176	5	5
	293	2	0
	696	4	5

**Tabla 3.1.** Familias de estudio.

De forma adicional, se analizaron 42 muestras de ADN de donantes sanos no relacionados residentes en Asturias, utilizadas como población control.

### 3.2. DETERMINACIÓN DE MUTACIONES EN *BRCA2*

El DNA fue obtenido a partir de sangre periférica utilizando el método DNAzol<sup>®</sup> (MRC, USA), siguiendo las instrucciones del fabricante.

Tras llevar a cabo una PCR de amplificación de la región específica de estudio del gen *BRCA2* se realizó la secuenciación del fragmento amplificado mediante el método de Sanger con el kit comercial BigDye<sup>®</sup> 3.1 (Applied Biosystems, Foster City, USA). Los fragmentos secuenciados se purificaron con BigDye<sup>®</sup> Xterminator<sup>™</sup> Purification Kit (Applied Biosystems, Foster City, USA) y se analizaron mediante electroforesis capilar con el secuenciador ABI PRISM<sup>™</sup> 310 Genetic Analyser (Applied Biosystems, Foster City, USA). La detección de las mutaciones se llevó a cabo o bien de forma manual o utilizando el software Mutation Surveyor<sup>®</sup> 4.0.

### 3.3. GENOTIPADO DE MICROSATELITES Y SNPs

Se analizaron un total de 9 marcadores en la región genómica cercana a *BRCA2*: 3 microsatélites y 6 SNPs. Todas las posiciones cromosómicas mencionadas en este estudio están referidas a la versión del genoma humano hg19 (GRCh37).

#### Selección de microsatélites

Se eligieron tres microsatélites, o marcadores de repeticiones cortas en tándem (STR *Short Tandem Repeats*), descritos por Neuhausen (Neuhausen *et al.*, 1998) en el análisis de haplotipos relacionados con mutaciones en *BRCA2*. Estos marcadores fueron:

STR	Localización cromosómica	Repetición	Oligonucleótidos	
			Directo	Reverso
D13S260	32436833	AC	agatattgtctccgttccatga	cccagatataaggacctggcta
D13S1698	32704522	AC	gtccataccactaagtctgac	aacctcaggctaatagtctca
D13S1695	33523563	TG	agaatcattgcctactta	gataacttaccagcatgtga



**Tabla 3.2.** Microsatélites seleccionados para el estudio y oligonucleótidos utilizados para la amplificación de los mismos.

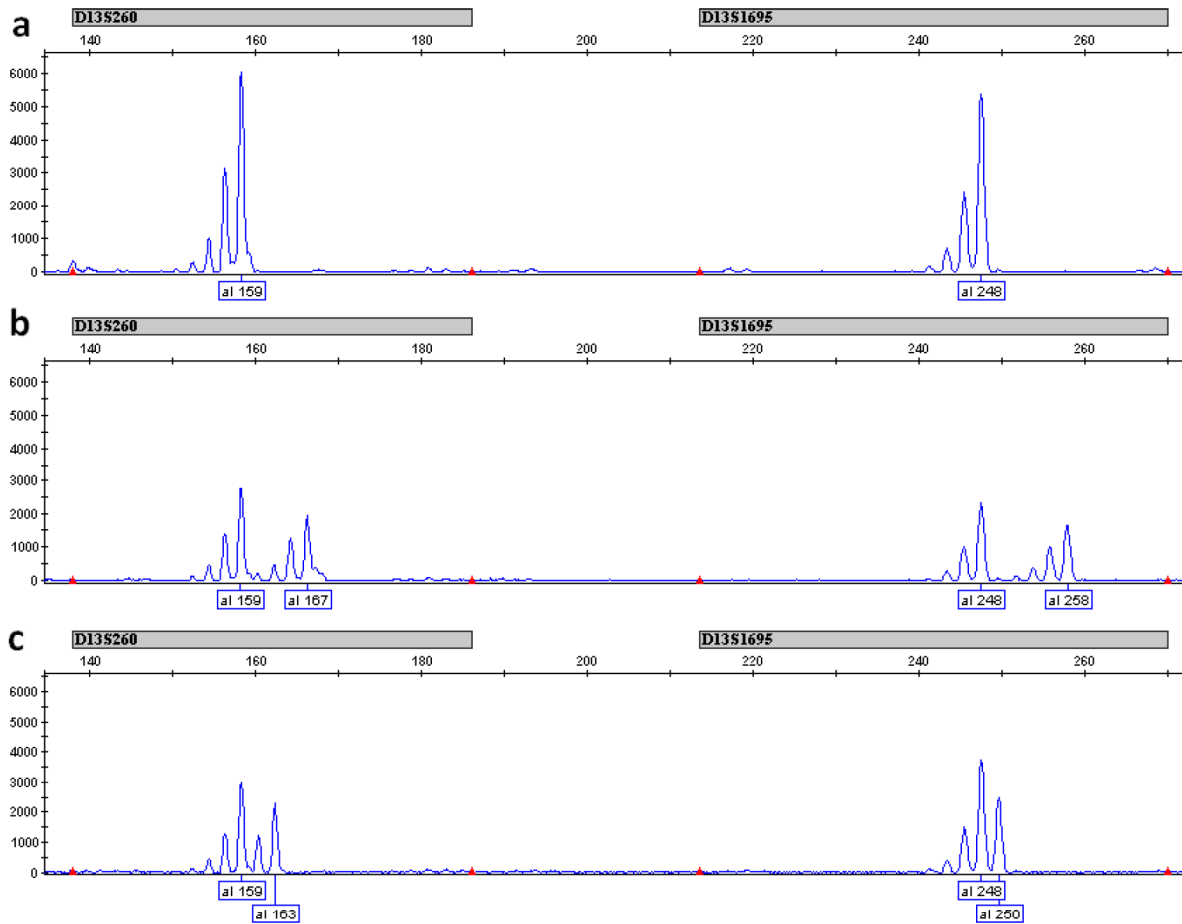
Para la amplificación de los fragmentos se utilizaron las parejas de oligonucleótidos descritas, estando una de ellas marcada con un fluoróforo (6-FAM o HEX). La reacción de amplificación se realizó en un volumen de 15 µl, utilizando 60 ng de DNA genómico por reacción, 2,5 mM de MgCl<sub>2</sub>, 0,025mM de dNTPs y 0,04U de Taq polimerasa por reacción.

El programa de amplificación de PCR para estos marcadores fue el siguiente:

Activación	Desnaturalización	Alineamiento	Elongación	Desnat.	Alin.	Elong.		
95°C	94°C	55°C	72°C	89°C	55°C	72°C	72°C	60°C
5 min	15 s	15 s	15 s	15 s	15 s	15 s	10 min	30 min
	x10			x20				

**Tabla 3.3.** Programa PCR microsatélites.

Una vez finalizada la reacción, se analizaron los fragmentos amplificados mediante electroforesis capilar en un ABI PRISM™ 310 Genetic Analyser (Applied Biosystems, Foster City, USA). Los resultados se analizaron utilizando el software *Genemapper* 4.0. Un ejemplo de genotipado de los microsatélites analizados se presenta en la figura 3.1.



**Figura 3.1.** Muestra de resultados de análisis de dos microsatélites (D13S260 y D13S1695) en tres individuos (a, b, c) utilizando el software *Genemapper 4.0*. El individuo **a** es homocigoto para ambos microsatélites, mientras que los individuos **b** y **c** son heterocigotos para ambos marcadores genéticos.

### Selección de SNPs

A partir de los datos obtenidos con el chip *Affymetrix Genome-Wide Human SNP Array 6.0* en 4 casos índice con la mutación 1 de *BRCA2* y un caso control (datos aportados por el Dr. X. Puente de la Universidad de Oviedo, analizados por la Dra. L. Fernández en nuestro laboratorio), se procedió a seleccionar 6 SNPs en las proximidades del gen *BRCA2*, cuya frecuencia del alelo menor (MAF, *minor allele frequency*) fuera lo más baja posible (en nuestro caso,  $MAF < 0,30$ ) y dicho alelo, en la medida de lo posible, no apareciera en ninguno de los otros haplotipos. Tras realizar las modificaciones oportunas en el formato del archivo procedente del chip (facilitado

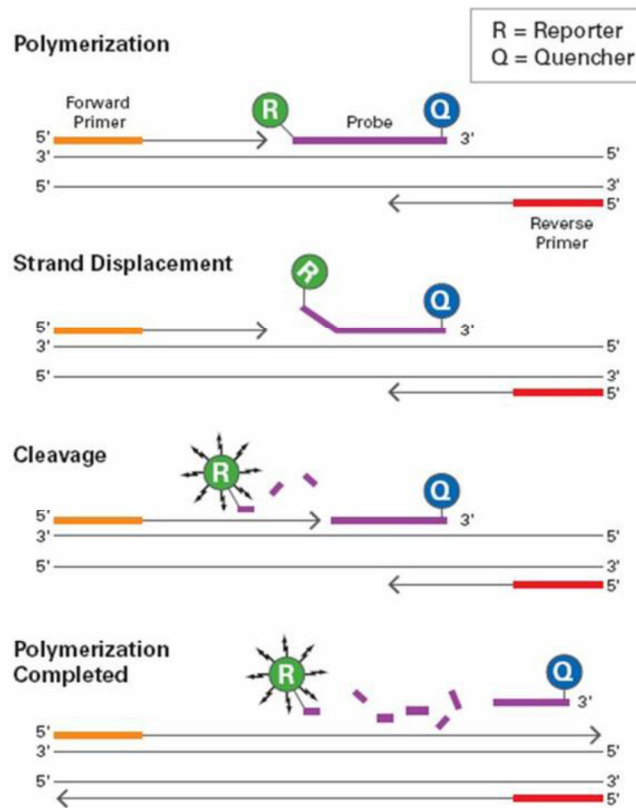
por R. Valdés, Universidad de Oviedo), estos datos se analizaron con el software SHAPEIT2, a fin de establecer la fase de los haplotipos en las cinco muestras. Estos haplotipos se fueron extendiendo de forma manual a partir de la mutación hasta aquellas posiciones en las que se rompía el desequilibrio de ligamiento. De esta manera, se encontró que el haplotipo compartido por los cuatro pacientes portadores de la mutación tiene un tamaño aproximado de 1 Mb. Dentro de esta región se seleccionaron 6 SNPs (tabla 3.4) para analizar los haplotipos tanto en individuos con mutación como en familiares o controles no afectados.

SNP	Localización cromosómica	Alelos	Minor Allele Frequency
rs455805	13:32588825	A/G	G = 0,1377
rs17077090	13:32688074	C/T	T = 0,0358
rs9315161	13:32789242	C/T	C = 0,2470
rs206085	13:32842335	C/T	T = 0,1483
rs9943888	13:32928202	A/G	G = 0,2309
rs3848093	13:33496899	C/G	C = 0,2507

**Tabla 3.4.** SNPs seleccionados y características.

Para el análisis de los SNPs en las muestras a estudio, se utilizó la técnica de PCR cuantitativa (q-PCR) mediante un ensayo comercial con sondas TaqMan para discriminación alélica *Taqman*<sup>®</sup> *SNP Genotyping Assays* de Applied Biosystems. Estos ensayos constan de dos oligos específicos que flanquean a un SNP y de dos sondas fluorescentes específicas para cada alelo, cada una marcada con distinto fluorocromo. Cada sonda consta de un *quencher* no fluorescente (NFQ) en el extremo 3' y de un fluoróforo (*reporter*) en el extremo 5', siendo el quencher el que evita la liberación de fluorescencia si la sonda no es degradada. Estos ensayos se basan en la actividad exonucleasa 5'-3' de la Taq Polimerasa, que romperá la sonda que haya hibridado, liberando el fluoróforo (ver figura 3.2). Esta fluorescencia es cuantificada con el software del termociclador (7500 Real Time PCR System, Applied Biosystems) y es proporcional al número de copias del gen

en estudio presente en la muestra. Así, el programa de análisis mostrará los resultados indicando si el DNA analizado era heterocigoto u homocigoto para un SNP (fuente: Applied Biosystems).



**Figura 3.2.** Principio del método TaqMan.

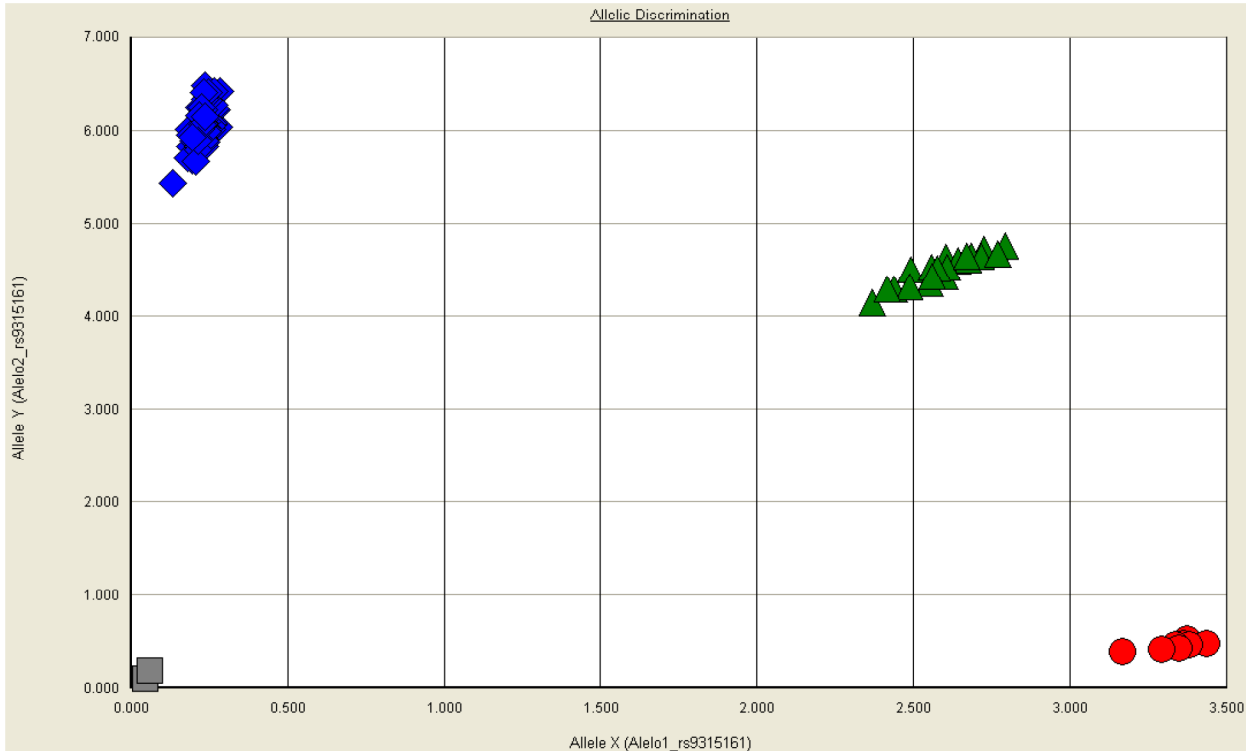
Los ensayos se realizaron siguiendo las pautas establecidas por la casa comercial, Applied Biosystems, haciendo duplicados para cada muestra. Cada ensayo se realizó en 10 µl, utilizando 20 ng de DNA por reacción.






El programa de amplificación en el termociclador 7500 Real Time PCR System (ABI7500) fue el siguiente:

Activación AmpliTaq	PCR	
Hold	x40	
	Desnaturalización	Alineamiento / Elongación
95°C	92°C	60°C
10 min	15 s	1 min

**Tabla 3.5.** Programa q-PCR.

Un ejemplo de los resultados obtenidos en el análisis de SNPs se representa en la figura 3.3. Además de obtener una gráfica de discriminación alélica, el programa de análisis genera un informe detallado para cada muestra.



**Figura 3.3.** Gráfica de discriminación alélica. En ella se muestran con el símbolo  aquellas muestras homocigotas para el alelo X, con  las homocigotas para el alelo Y,  representan las muestras heterocigotas y  para aquellas muestras control (*NTC, non template control*). En caso de existir alguna muestra que no fuera posible situar en ninguno de los grupos anteriormente mencionados se mostraría como indeterminada mediante el símbolo .

### 3.4. ANÁLISIS DE LOS DATOS DE GENOTIPADO Y GENERACIÓN DE HAPLOTIPOS

La obtención de haplotipos a partir de los resultados de genotipado, tanto de los SNPs como los microsatélites, fue realizado utilizando el software PHASE 2.0 (University of Washington, Seattle, 2004). Este programa implementa un método estadístico bayesiano para la reconstrucción de haplotipos a partir de los datos de los genotipos de una población. Permite el análisis de SNPs,

microsatélites y loci multi-alélicos, en cualquier combinación y permite la falta de datos (Stephens *et al.*, 2001; Stephens *et al.*, 2003).

Se compararon los datos de varias formas: entre el grupo de pacientes portadores de la mutación y los donantes no relacionados; entre el grupo de pacientes portadores de la mutación y sus familiares no portadores y entre los datos de los controles donantes no relacionados y los familiares no portadores de la mutación. El programa se ejecutó 5 veces para cada prueba. En caso de existir diferencias en los resultados se tomó la mediana como resultado final.

Para ejecutar cada prueba con este programa es necesario especificar ambos grupos en el archivo de entrada. El programa se ejecuta entonces con la siguiente línea de comando:

```
>PHASE.exe -c archivoin.txt archivoout.txt -5x
```

El programa genera una lista de archivos, siendo los más representativos los siguientes: un archivo resumen donde se describen los haplotipos que ha determinado para cada individuo; un archivo con las frecuencias de cada haplotipo en la muestra de la población estudiada total y relativa para cada grupo; otro archivo muestra las distintas parejas de haplotipos posibles para cada muestra con su respectiva probabilidad; y por último, el archivo donde se indica la significación estadística (p-valor) de la comparación de los dos grupos estudiados.

Con los datos obtenidos por el programa se otorgó a cada individuo analizado su haplotipo correspondiente. Estos haplotipos se ajustaron posteriormente en algunos casos de forma manual al analizar los individuos dentro de cada pedigrí familiar.

### 3.5. CÁLCULO DE LA ANTIGÜEDAD DE LA MUTACIÓN

Se utilizó el método descrito por Peixoto *et al.* (2011) en el que estiman la edad de una mutación a partir de la variación acumulada desde el haplotipo ancestral. El método tiene en cuenta tanto las tasas de recombinación como de mutación en la generación de variación. Así, la probabilidad de cambio por generación ( $\epsilon$ ) viene dada por la fórmula:

$$\epsilon = 1 - [(1 - c) \cdot (1 - \mu)]$$

donde  $c$  es la tasa de recombinación y  $\mu$  es la tasa de mutación.

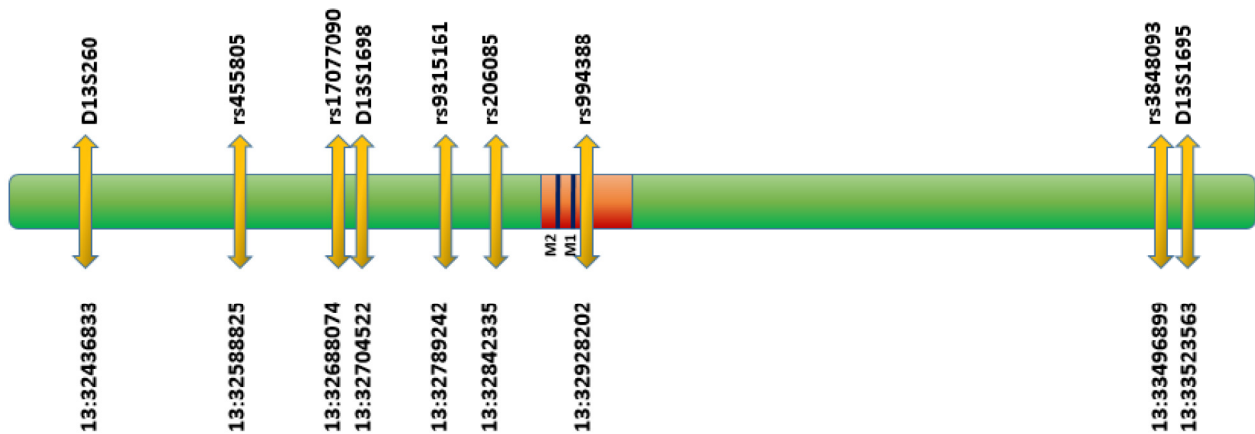
La estimación de  $\mu$  varía si se están analizando microsatélites o SNPs. Se considera  $\mu = 7,8 \cdot 10^{-4}$  (Gyapay *et al.*, 1994) para microsatélites de dinucleótidos y  $\mu = 10^{-6}$  (Weber & Wong, 1993) para SNPs.

La tasa de recombinación ( $c$ ) se puede estimar a partir de la distancia física entre los dos marcadores más distantes, usando un factor de conversión calculado con el programa “*Rutgers Map Interpolator*” (<http://compgen.rutgers.edu/old/map-interpolator/>). Con estos datos se puede calcular  $\epsilon$  y con ello el tiempo transcurrido desde la aparición de la mutación mediante la fórmula  $\lambda = \epsilon \cdot t$ , donde  $\lambda$  es la media de eventos (recombinación y mutación) hallados en el total de las familias estudiadas y  $t$  el tiempo (en generaciones) que hace que surgió dicha mutación.

Conociendo  $t$  se puede calcular el número de años (20-25 años/generación) que transcurrieron desde que apareció el individuo fundador.

## 4. Resultados

Se han genotipado 9 marcadores (3 microsatélites y 6 SNPs) para un total de 65 individuos y 42 controles con el fin de determinar si existe un haplotipo común para cada una de las dos mutaciones a estudiar: p.Asn1344Hisfs (Mut. 1) y p.Gln699\* (Mut. 2). Ambas mutaciones en el gen *BRCA2* producen el truncamiento de la proteína y están asociadas a cáncer de mama y ovario hereditario; los marcadores empleados en el estudio se encuentran en las proximidades del gen antes mencionado y la determinación de un haplotipo común podría poner de manifiesto la existencia de un ancestro común para la mutación que lo presente. La localización de todos los marcadores respecto al gen *BRCA2* se muestra en la figura 4.1.



**Figura 4.1.** Posición cromosómica de los marcadores analizados en relación con la posición del gen *BRCA2* (en rojo). El nombre del marcador se describe en la parte superior de la figura y su posición cromosómica en la parte inferior. Se han marcado como M1 Y M2 las mutaciones estudiadas en *BRCA2*.

### 4.1. Genotipado y obtención de haplotipos

Una vez realizadas las PCR para los microsatélites, el análisis mediante electroforesis capilar y los ensayos de discriminación alélica para los SNPs mediante PCR cuantitativa en tiempo real, se obtuvieron los genotipos para cada uno de los 107 individuos analizados. Estos genotipos se presentan en el Anexo 1.



Para la obtención de los haplotipos correspondientes, se analizaron los datos genotípicos con el programa PHASE 2.0 realizando las siguientes comparaciones:

- 1.1 Individuos mutados (casos) de la Mutación 1 frente a controles (donantes no relacionados).
- 1.2 Casos Mutación 1 frente a familiares no mutados de éstos.
- 1.3 Familiares no mutados (Mutación 1) frente a controles (donantes no relacionados).
- 2.1 Casos Mutación 2 frente a controles (donantes no relacionados).
- 2.2 Casos Mutación 2 frente a familiares no mutados de éstos.
- 2.3 Familiares no mutados (Mutación 2) frente a controles (donantes no relacionados).

En la tabla 4.1 se presentan los resultados de significación estadística (p-valor) de cada una de las comparaciones entre los grupos descritos anteriormente. Se observa una diferencia significativa en los haplotipos de los casos con mutación frente a los controles. En los casos correspondientes a la Mutación 1 también se observa una diferencia significativa entre los haplotipos de los casos y de los familiares no afectados. Esto no ocurre entre los casos de la Mutación 2 y los familiares no afectados. No existen diferencias significativas entre el grupo control y los familiares no afectados de portadores de la Mutación 1, así como tampoco entre los portadores de la Mutación 2 y sus familiares.

Mutación	Interacción	p-valor
1	1.1 Casos vs Controles	0,01
	1.2 Casos vs Familiares	0,03
	1.3 Familiares vs Controles	0,16
2	2.1 Casos vs Controles	0,01
	2.2 Casos vs Familiares	0,2
	2.3 Familiares vs Controles	0,01

**Tabla 4.1.** Significancia de los grupos comparados.

Asimismo, se han analizado los haplotipos resultantes de la comparación para los casos de la Mutación 1 frente a la suma de controles y familiares juntos, existiendo también diferencias significativas entre estos grupos así comparados. Se identifica un haplotipo único para 23 de los 24 individuos portadores de la mutación 1, y además, el único que difería presentaba en segundo lugar con más probabilidades el haplotipo común de la mutación.

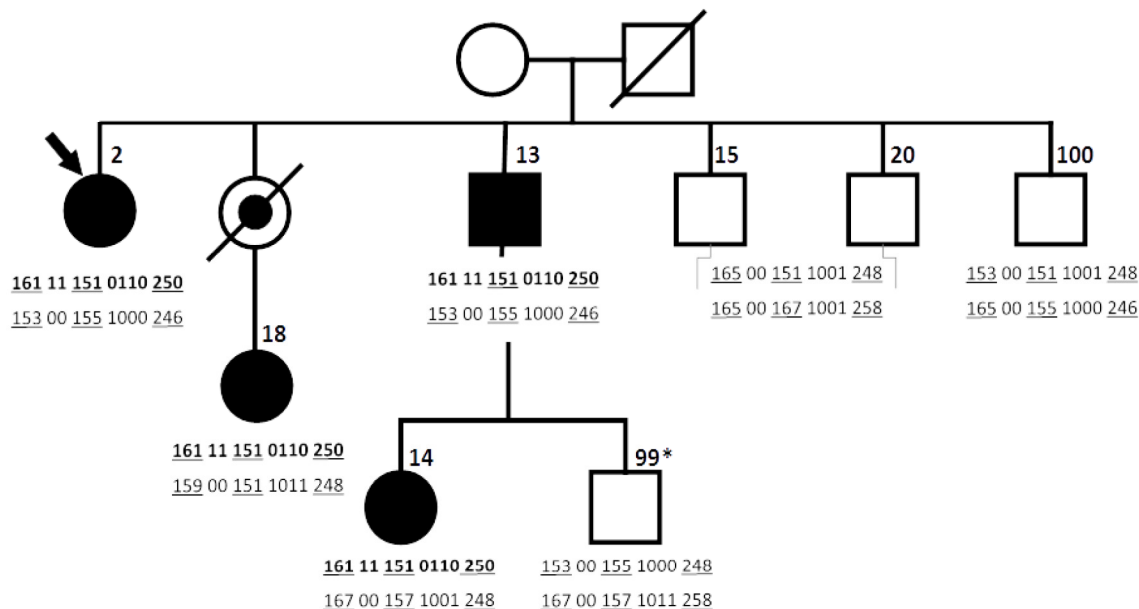
Los haplotipos obtenidos con el programa PHASE 2.0 para todos los portadores de la mutación y familiares estudiados en el análisis se muestran en las siguientes figuras 4.2 y 4.3 en árboles genealógicos. Todos aquellos que presenten un asterisco (\*) junto al número de muestra (en la parte superior derecha de cada individuo) hacen referencia a aquellos cuyo haplotipo ha tenido que ser ajustado en función de sus familiares, es decir, no era el haplotipo más probable según el programa PHASE 2.0. Los loci se muestran en orden creciente de sus posiciones cromosómicas.

El haplotipo correspondiente a la Mutación 1 se muestra en negrita, siendo [**161 11 151 0110 250**] donde “161”, “151” y “250” hacen referencia a los microsatélites, y el código binario a los SNPs, los cuales se especifican en la tabla 4.2. Este haplotipo es común para 23 de los 24 portadores de la Mutación 1 y no está presente en ninguno de los familiares no afectados o individuos control.

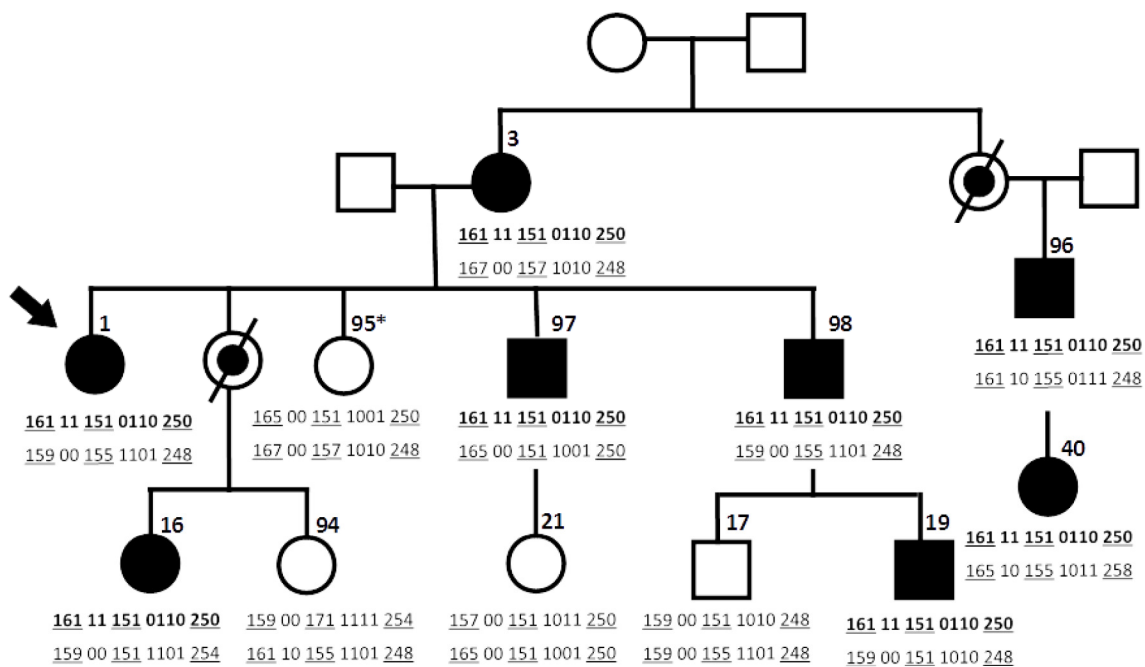
SNP	0	1
rs455805	A	G
rs17077090	C	T
rs9315161	C	T
rs206085	C	T
rs9943888	A	G
rs3848093	C	G

**Tabla 4.2.** Leyenda del código binario de SNPs.

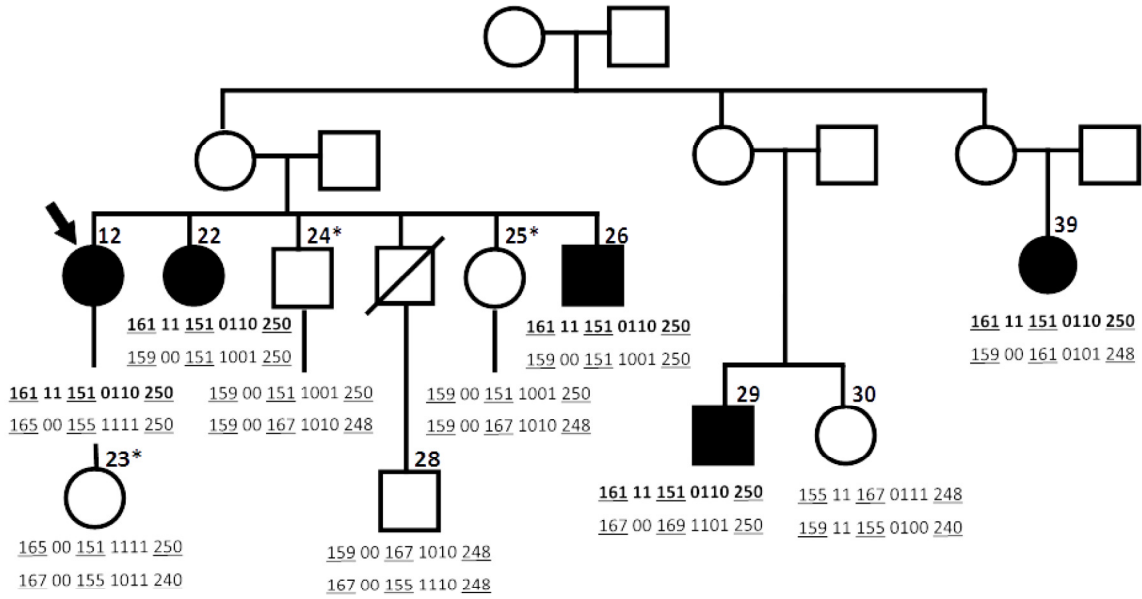
### Familia 27



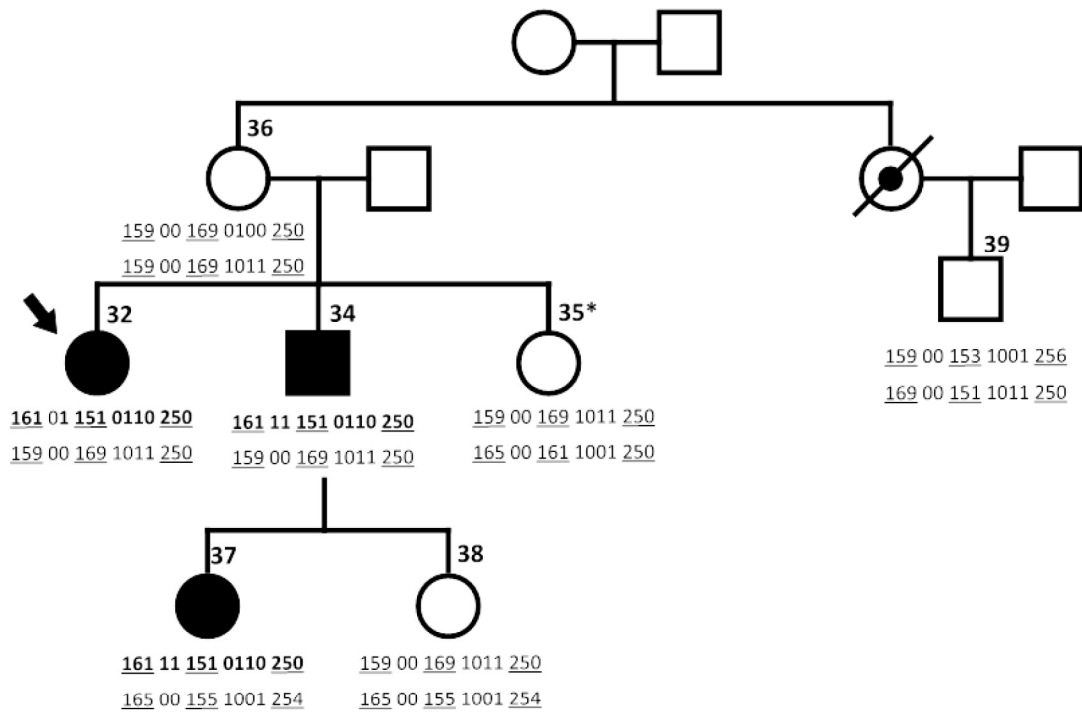
### Familia 208



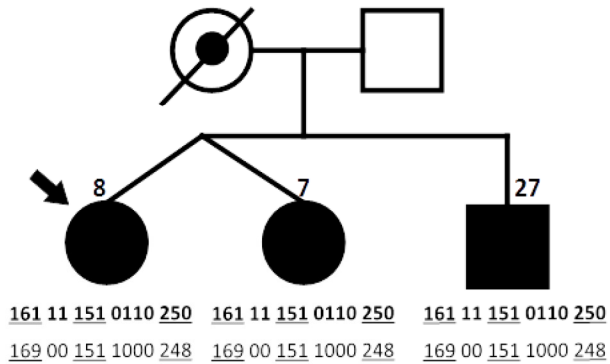
### Familia 424



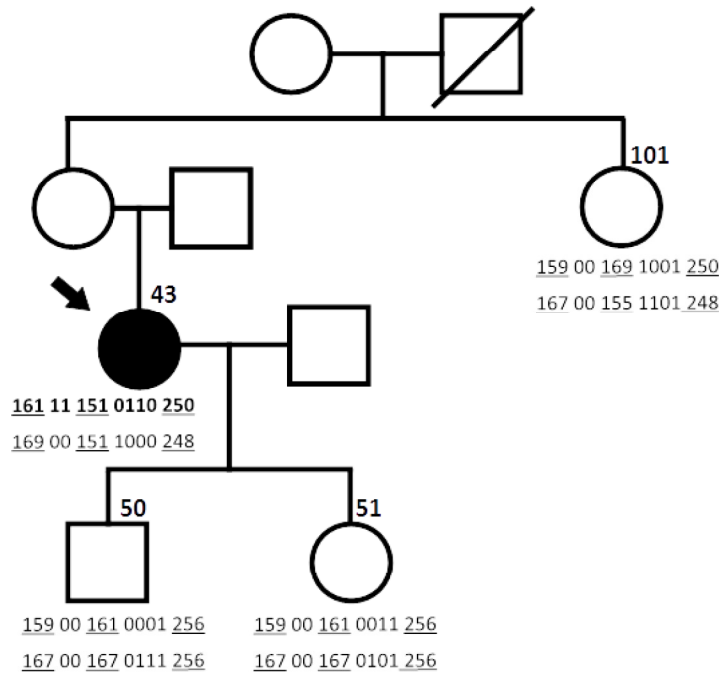
### Familia 609



### Familia 331

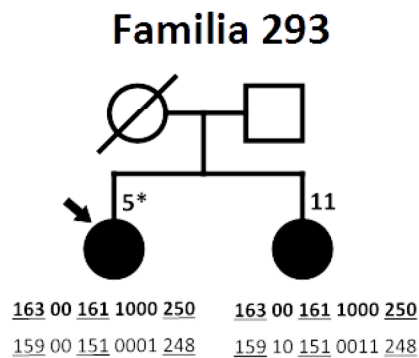
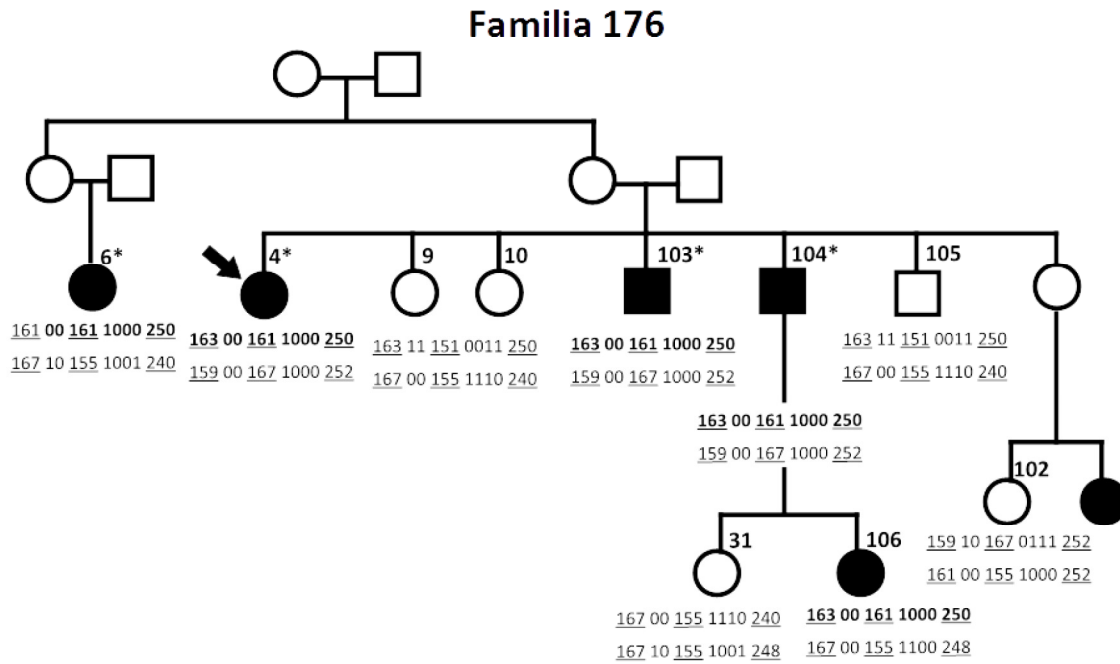


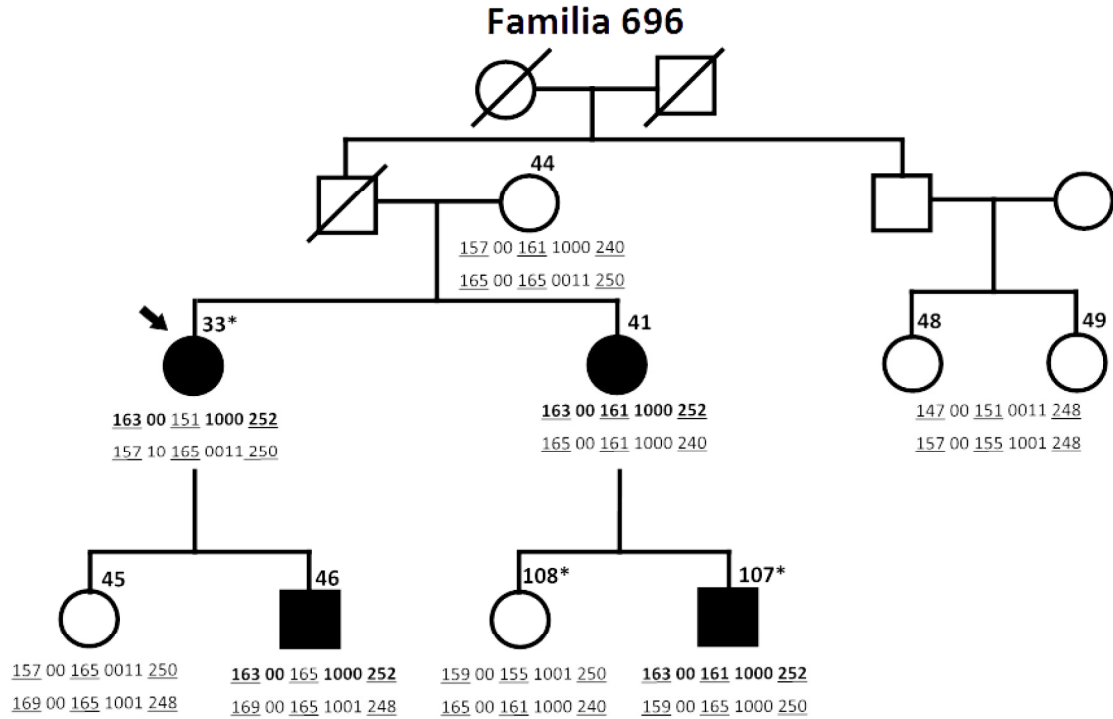
### Familia 795



**Figura 4.2.** Árboles genealógicos de las familias con Mutación 1. Los números en la parte superior derecha corresponden al nº de muestra analizada. La flecha corresponde al caso índice de la familia. El asterisco al lado del nº de muestra indica que el haplotipo se ajustó manualmente en relación con los datos del resto de familiares. Los miembros portadores de la Mutación 1 en *BRCA2* se muestran en color negro. Existe un haplotipo idéntico en todos los portadores de mutación, excepto en el caso índice de la familia 609, donde se ha producido una mutación en uno de los SNPs.

En relación a los haplotipos obtenidos en los casos y familiares de la Mutación 2, no se observa un haplotipo único definido para los individuos con mutación, dado que los SNPs no son informativos y los microsatélites no se mantienen en todos los portadores de mutación, aunque en este caso contamos con menos de la mitad de portadores y familias afectas que en la primera mutación. El haplotipo que es compartido por un mayor número de portadores de la mutación 2 es 163 00 161 1000 250.





**Figura 4.3.** Árboles genealógicos de las familias con la Mutación 2. Los números en la parte superior derecha corresponden al nº de muestra analizada. La flecha corresponde al caso índice de la familia. El asterisco al lado del nº de muestra indica que el haplotipo se ajustó manualmente en relación con los datos del resto de familiares. Los miembros portadores de la Mutación 1 en BRCA2 se muestran en color negro. Se observan incongruencias entre los pacientes 44, 33 y 41; y eventos de recombinación en el caso del tercer microsatélite (fragmento de 252 bases en lugar de 250). También se observa recombinación para el segundo microsatélite en los pacientes 33 y 46.

#### 4.2. Cálculo de la edad de la mutación

Para llevar a cabo el cálculo de la antigüedad de las mutaciones se tuvieron en cuenta los siguientes parámetros (Peixoto *et al.*, 2011):

- $\epsilon$  = probabilidad de cambio por generación
- $c$  = tasa de recombinación
- $\mu$  = tasa de mutación
- $\lambda$  = media de los eventos de recombinación/mutación (eventos/nº familias)
- $t$  = nº de generaciones (20-25 años/generación)

$$\varepsilon = 1 - [(1 - c) \cdot (1 - \mu)]$$

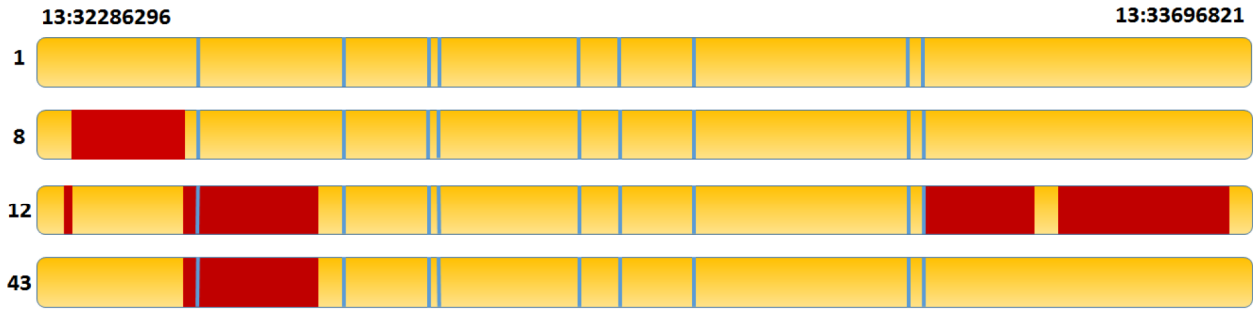
$$t = \lambda / \varepsilon$$

Calculamos de dos formas diferentes la edad de la mutación 1: por un lado teniendo en cuenta la los datos de los haplotipos encontrados en un total de seis familias analizadas, y, por otro lado, considerando bloques de recombinación inferidos a partir de una ampliación del estudio de los datos del chip *Affymetrix Genome-Wide Human SNP Array 6.0* con cuatro casos índice y un control.

En la primera aproximación, tuvimos en cuenta la existencia de un único evento de mutación, correspondiente a la variación encontrada para el SNP rs455805 en la paciente 32 (familia 609) en las seis familias analizadas. Para el cálculo de  $\mu$  sumamos las tasas de mutación de los marcadores seleccionados siendo  $\mu=7,8 \cdot 10^{-4}$  y  $\mu=10^{-6}$  la tasa de mutación de microsatélites y SNPs respectivamente. Mediante la aplicación *Rutgers Map Interpolator* calculamos la tasa de recombinación de la región comprendida entre las posiciones cromosómicas 13:32436834 y 13:33523563, siendo ésta  $c=0,019905$ . De esta manera, el resultado obtenido fue 7,5 generaciones (150 años) transcurridos desde el evento fundador ( $c=0,019905$ ,  $\mu=0,002346$ ,  $\lambda=1/6$ , 20 años/generación).

A fin de tener en cuenta no sólo este único evento mutacional, sino también las posibles recombinaciones que se hubieran producido a lo largo de las generaciones, procedimos a aumentar la ventana de análisis, extendiendo los haplotipos más allá de la región compartida por todos los casos para identificar grandes bloques de recombinación. De esta manera identificamos cinco bloques en los haplotipos, con tamaño variable en cuanto a número de SNPs y que están reflejados en la figura 4.4. El primer bloque que aparece en el individuo 12 fue descartado para los cálculos, al asumir que se trata de un evento mutacional más que de un gran evento de recombinación. Mediante la aplicación *Rutgers Map Interpolator* hemos calculado la tasa de recombinación de la región comprendida entre las posiciones cromosómicas 13:32286296 y 13:33696821, siendo ésta  $c=0,02502$ . El resultado derivado de esta estimación fue de 40 generaciones (800 años) transcurridos desde el evento fundador ( $c=0,02502$ ,  $\mu=4 \cdot 10^{-6}$ ,  $\lambda=1$  (4/4), 20 años/generación).





**Figura 4.4.** Representación gráfica de una ventana cromosómica localizada en el cromosoma 13 entre las posiciones 32286296 y 33696821 (1,41 Mb), región dentro de la cual se encuentra *BRCA2*. Los números de la izquierda hacen referencia al número de muestra analizado mediante el chip. Las zonas amarillas representan el posible haplotipo “ancestral” siendo las bandas rojas aquellas regiones donde éste último varía y, por lo tanto, se estima que haya existido recombinación cromosómica. Esta estimación se ha llevado a cabo mediante la extensión de forma manual de los haplotipos. Las bandas azules representan la ubicación de los nueve marcadores genotipados.

En el caso de la mutación 2, se tienen muy pocas familias y con algunas incoherencias en los resultados, lo que hace que el cálculo de la edad de la mutación no tenga suficiente credibilidad. Aun así, se ha realizado el cálculo de forma teórica, teniendo en cuenta dos eventos de recombinación. Por un lado, el microsatélite D13S260 ha recombinado/mutado en la paciente número 6 y, por otro, el microsatélite D13S1695 que también difiere del supuesto haplotipo fundador en la familia 696. De este modo, el resultado obtenido sugiere que la mutación fundadora podría haber surgido hace, aproximadamente, 30 generaciones (600 años) ( $c=0,019905$ ,  $\mu=0,002346$ ,  $\lambda=2/3$ , 20 años/generación).

## 5. Discusión

En este estudio se ha llevado a cabo el análisis de haplotipos en relación a un gen de susceptibilidad a cáncer de mama familiar como es *BRCA2*, el cual presenta dos mutaciones restringidas a la población asturiana descritas en el artículo de Blay *et al.*, 2013. La mutación 1 se encuentra en un área muy reducida del occidente asturiano, delimitada por los ríos Eo y Navia, lo que pudo generar el aislamiento de la población. Este hecho nos ofrece un primer indicio sobre la posible existencia de un ancestro común. La mutación 2 se encuentra en un área de mayor extensión y en menor número de familias, por lo que la demostración de dicho ancestro común para esta mutación ha resultado menos precisa.

### **Selección de microsatélites para el estudio de los haplotipos**

La selección de los tres microsatélites se llevó a cabo a través del análisis de la literatura hasta la fecha, siendo seleccionados tres de los microsatélites empleados en el estudio de Neuhausen *et al.* (1998), en el cual los marcadores D13S260, D13S1698 y D13S1695 se encontraban dentro del mismo núcleo de microsatélites en fase.

La determinación de haplotipos mediante microsatélites ha sido la forma tradicional de llevar a cabo este tipo de estudio, siendo un método asequible aunque laborioso, ya que puede haber múltiples factores que afecten al resultado final. La identificación de genotipos en heterocigosis puede resultar compleja, debido a los numerosos picos satélites que se producen en la amplificación por PCR de las repeticiones de dinucleótidos. En nuestro estudio algunos de los genotipos inicialmente establecidos sin tener en cuenta las asignaciones de los pedigrís familiares fueron redefinidos posteriormente al observar inconsistencias en los datos de las familias, siendo la mayoría de los errores debidos a la asignación inicial de un genotipo como homocigoto cuando en realidad era heterocigoto (figura 2.1).

La asignación del tamaño del fragmento amplificado también puede resultar confusa. En nuestro caso, hemos utilizado siempre como controles para la asignación de tamaños dos muestras de DNA presentes en todas las amplificaciones que hemos realizado. Hemos observado que no

siempre el resultado se ajustaba al inicial sino que el tamaño de los fragmentos resultantes tras cada PCR podía variar más de una base, por lo que se ha tenido que duplicar el panel de análisis para estandarizar el proceso. No obstante, no se deben descartar errores derivados de la técnica, que puedan haber dado lugar a genotipos erróneos.

## Selección de SNPs

Todos los avances derivados del Proyecto Genoma Humano y el abaratamiento de costes en tecnología de genotipado, han permitido afrontar múltiples proyectos para un estudio más en profundidad del genoma humano, entre los que se encontraba el Proyecto Internacional HapMap, el cual tenía como objetivo establecer el catálogo de haplotipos presentes en las distintas poblaciones humanas a fin de facilitar posteriores estudios de asociación entre variantes genéticas particulares y enfermedades. Para ello, se estudiaron en torno a 10 millones de SNPs (aproximadamente 1 cada 1200 bases) y se determinaron los haplotipos creados a causa del desequilibrio de ligamiento entre distintos SNPs.

Hemos utilizados los datos que teníamos disponibles del genotipado de cuatro casos índice portadores de la Mutación 1 junto con un paciente control que no presentaba *BRCA2* mutado. Dicho genotipado se llevó a cabo mediante el ensayo de *Affymetrix "Genome-Wide Human SNP Array 6.0"*. Este ensayo caracteriza 1,8 millones de marcadores genéticos entre los que se encuentran más de 906600 SNPs y más de 946000 variantes en el número de copias (CNVs).

En estudios como el nuestro, la reconstrucción de haplotipos y su fase partiendo únicamente de datos genotípicos procedentes de un número pequeño de muestras presenta limitaciones. Una estima adecuada de la fase de los haplotipos depende del número de muestras, densidad de marcadores, disponibilidad de datos familiares, estructura poblacional y calidad del genotipado (Browning & Browning, 2011). En nuestro estudio inicial, partimos de 4 casos no emparentados y un control. SHAPEIT2 nos permitió inferir los haplotipos y su fase al no estar basado únicamente en análisis de datos procedentes de pedigrís, sino que su algoritmo hace uso de datos genotípicos a nivel poblacional obtenidos de los proyectos HapMap y 1000 Genomes. De esta manera, el programa realiza interacciones del algoritmo de Markov Chain Monte Carlo (MCMC) a fin de

estimar cuáles son los haplotipos con mayor probabilidad, teniendo en cuenta su frecuencia poblacional global. En caso de disponer de datos familiares, en una segunda etapa los incluiría en el análisis en vistas a aumentar la precisión de la predicción. Los algoritmos que emplea están basados en el modelo genético de coalescencia con recombinación desarrollado por Stephens *et al.* (2001) en el programa PHASE 2.0, que en este caso han implementado para aumentar la cantidad y velocidad de procesamiento de datos, de manera que sea compatible con plataformas de genotipado de alto rendimiento.

Como entrada del programa se emplearon el conjunto de genotipos procedentes del array y el mapa genético de recombinación, produciendo como salida el conjunto de haplotipos en fase.

Los haplotipos obtenidos se analizaron manualmente a partir de la posición de la mutación hasta que se observó la ruptura del desequilibrio de ligamiento. Dicho haplotipo tenía un tamaño aproximado de 1 Mb. Para llevar a cabo la selección de los 6 SNPs se tomaron aquellos que se encontraran dentro de este marco, cuya frecuencia menor de alelo (MAF) fuera lo más baja posible ( $MAF < 0,3$ ) y dicho alelo, en la medida de lo posible, no se encontrase en el haplotipo control. Con los seis SNPs y los tres microsatélites seleccionados se estudiaron los casos con mutación en *BRCA2* y los controles no mutados (familiares no afectados y otros individuos control no relacionados) con el fin de estimar si los individuos con mutación compartían un haplotipo no presente en los individuos no mutados.

### **Identificación de haplotipos**

Para la estimación de la frecuencia de los distintos haplotipos y de qué pareja de haplotipos porta cada individuo se pueden llevar a cabo distintos análisis. Muchos grupos aplican el algoritmo de Expectación-Maximización (EM) para obtener la estimación de máxima probabilidad de las frecuencias de haplotipos en la muestra (Dempster *et al.*, 1977; Excoffier & Slatkin, 1995).

Por otra parte, se encuentran los modelos coalescentes y el método MCMC, utilizados por el programa PHASE 2.0 (Stephens *et al.*, 2001; Lin *et al.*, 2002), que asignan fases a cada individuo y estiman la frecuencia de los haplotipos. Estos algoritmos utilizan modelos de genética de

poblaciones para relacionar distintos patrones de haplotipos, de tal manera que si un haplotipo es más similar a los patrones de haplotipos observados más comunes, éste es más probable que sea inferido que aquellos que sean menos similares. Una ventaja de este programa es que además de mostrar la pareja de haplotipos más probable para cada individuo, también genera un documento con las probabilidades de cada pareja de haplotipos, con las distintas combinaciones de alelos, para cada individuo; de esta forma, no se pierde información y se puede evaluar la eficiencia del análisis.

En nuestro estudio, utilizando el programa PHASE 2.0, hemos podido identificar un haplotipo en torno a *BRCA2* para la mutación 1, donde hemos comprobado que los pacientes afectados portadores de la misma comparten al menos 1 Mb. El haplotipo común a todos estos individuos es el 161 11 151 0110 250 (o bien, 161 GT 151 CTGC 250) correspondiente a los 9 marcadores estudiados y no está presente en ningún individuo sano no portador de la mutación, lo que nos sugiere la existencia de un ancestro común para esta mutación.

Al contrario, en la mutación 2, para los mismos marcadores, hemos observado cómo tan solo una parte de los individuos afectados comparten un haplotipo, 6 de 11, por lo que éste no está asociado claramente a la mutación, siendo el haplotipo que más individuos presentan 163 00 161 1000 250 (163 AA 161 TCAC 250). Asumiendo este haplotipo como el asociado a la mutación, se observan eventos de recombinación en los tres microsatélites y siendo muy frecuente en la población sana la combinación de alelos de SNPs asociada a los individuos portadores de la misma. Esto no nos permite determinar si estas familias comparten un ancestro común, pero en caso afirmativo esta mutación sería más antigua que la mutación 1, ya que se han dado más eventos de recombinación en este caso. Los eventos de recombinación han sido:

- Familia 176, la paciente 6 presenta un evento de recombinación para el primer microsatélite, D13S270.
- Familia 696, todos los afectados presentan recombinación para el tercer microsatélite, D13S1695. La paciente 33 manifiesta recombinación, además, para el segundo microsatélite, D13S1698.

Asimismo, hemos detectado incoherencias a nivel genotípico en las familias con la Mutación 2 en las relaciones padre/madre-hijos. Estas incoherencias podrían deberse a distintos factores, bien a que los datos de las familias son erróneos o a que el método no es lo suficientemente robusto para determinar con precisión los genotipos, tal como ya se ha señalado anteriormente en relación sobre todo a los microsatélites. Es por ello que los antes mencionados eventos de recombinación no pueden tomarse como tales de forma definitiva hasta demostrar que no se deben a errores del método.

### **Cálculo de la antigüedad de la mutación**

La edad de una mutación fundadora se puede estimar de forma teórica determinando la longitud del haplotipo en la región de la mutación, el cual se reduce en tamaño con cada generación, debido a efectos de recombinación y mutación. Se han propuesto diferentes métodos para estimar la antigüedad de una mutación (Greenwood *et al.*, 2010; Peixoto *et al.*, 2011).

En lo referente al estudio de la edad de la mutación que hemos denominado Mutación 1 de *BRCA2*, la hemos calculado teniendo en cuenta los 9 marcadores que nosotros mismos hemos genotipado. Por otra parte también la hemos calculado teniendo en cuenta los eventos de recombinación que hemos observado a partir de los datos obtenidos.

Para calcularla lo hemos hecho teniendo en cuenta tanto la única variación encontrada en uno de esos marcadores en un individuo como también analizando los bloques de recombinación observados tras la extensión de los haplotipos a partir del gen de estudio.

En el primer caso hemos obtenido una edad de unos 150 años mientras que en el segundo, estudiando los bloques de recombinación, hemos obtenido una edad de 800 años. Esta diferencia tan notable se debe a que los métodos de cálculo de la antigüedad de las mutaciones tienen en cuenta muchas asunciones. Cabe destacar que en el artículo de Greenwood *et al.* (2010) hicieron un estudio para comparar los distintos métodos de análisis de edad de mutaciones fundadoras, obteniendo resultados que diferían hasta en 80 generaciones dependiendo del método

empleado y marcadores. Por ello, la robustez de estos cálculos es incierta, y no se deben tomar como datos fidedignos.

Por otro lado, la edad de la mutación 2 se ha estimado en 600 años, siendo esperable en nuestra opinión que fuera más antigua que la mutación 1 debido a la presencia de un mayor número de variaciones en el haplotipo estudiado en los portadores de la mutación. Cabe decir que en el caso de esta mutación, los SNPs no aportan información específica de la misma, dado que el conjunto de alelos de SNPs es muy frecuente en la población general y el número de familias estudiadas fue muy bajo, lo que no permite sacar conclusiones con fiabilidad.

Una mejor aproximación para el cálculo de la edad de estas mutaciones se podría obtener seleccionando SNPs marcadores dentro de estos bloques de recombinación para su genotipado y análisis en el resto de familias reclutadas, junto a otros micrósatelites más alejados de las mutaciones en estudio, con la finalidad de dibujar un patrón más preciso de los cambios que fue sufriendo el haplotipo ancestral hasta nuestros días. En el presente estudio no fue posible realizar esta extensión del análisis debido a restricciones temporales para la presentación de la presente memoria.

## 6. Conclusiones

De acuerdo a los resultados obtenidos en el presente estudio, extraemos las siguientes conclusiones:

1. Se ha logrado determinar un haplotipo fundador constituido por 9 marcadores (6 SNPs y 3 microsatélites) localizados en una ventana de 1Mb en torno al gen *BRCA2* y asociado a la mutación 1, p.Asn1344Hisfs, siendo éste 161 11 151 0110 250.
2. Se ha detectado un haplotipo relacionado con la mutación 2, p.Gln699\*, siendo éste 163 00 161 1000 250 (presente en 6 de los 11 casos de estudio) y el cual presenta varios eventos de recombinación dentro de las 3 familias estudiadas, por lo que no se puede establecer un haplotipo fundador con seguridad.
3. No se ha podido establecer la edad de las mutaciones con fiabilidad debido a la falta de datos. Para ello deberían haberse analizado marcadores más alejados del haplotipo asociado con la mutación, extendiendo la ventana de análisis a fin de tener en cuenta los eventos de recombinación y mutación que fueron produciéndose en el haplotipo ancestral durante las sucesivas generaciones.



## 7. Bibliografía

Alexandrov LB, Nik-Zainal Serena, Wedge DC *et al.* **Signatures of mutational processes in human cancer.** *Nature* 2013; 500:415-421.

Alonso A, Benavides MM, Blanco I *et al.* **Cáncer hereditario.** Sociedad Española de Oncología Médica (SEOM). Ed.: Dispublic S.L., Madrid, 2006.

Blay P, Santamaría I, Pitiot AS, *et al.* **Mutational analysis of *BRCA1* and *BRCA2* in hereditary breast and ovarian cancer families from Asturias (Northern Spain).** *BMC Cancer* 2013; 13:243.

Beristain E, Martínez-Bouzas C, Guerra I *et al.* **Differences in the frequency and distribution of *BRCA1* and *BRCA2* mutations in breast/ovarian cancer cases from the Basque country with respect to the Spanish population: implications for genetic counseling.** *Breast Cancer Res Treat* 2007; 106:255-262.

**Breast Cancer Linkage Consortium.** *J. Natl. Cancer Inst.* 1999; 91:1310-1316.

Browning SR, Browning BL. **Haplotype phasing: existing methods and new developments.** *Nat Rev Genet* 2011; 12:703-714.

Campos B, Díez O, Odefrey F, *et al.* **Haplotype analysis of the *BRCA2* 9254delATCAT recurrent mutation in breast/ovarian cancer families from Spain.** *Hum Mutat* 2003; 21:452.

Chen J, Silver DP, Walpita D *et al.* **Stable interaction between the products of the *BRCA1* and *BRCA2* tumor suppressor genes in mitotic and meiotic cells.** *Mol Cell* 1998; 2:317-328.

Cox MM. **The importance of repairing stalled replication forks.** *Nature* 2000; 404:37-41.

- Connor F, Bertwistle D, Mee PJ *et al.* **Tumorigenesis and a DNA repair defect in mice with a truncating *BRCA2* mutation.** *Nat Genet* 1997; 17:423-430.
- Daly MJ, Rioux JD, Schaffner SF, *et al.* **High resolution haplotype structure in the human genome.** *Nat Genet* 2001; 29:2225-2229.
- De Abreu FB, Wells WA, Tsongalis GJ. **The emerging role of the molecular diagnostics laboratory in breast cancer personalized medicine.** *Am J Pathol* 2013; 183: 1075-1083.
- Delaneau O, Howie B, Cox A, *et al.* **Haplotype estimation using sequence reads.** *Am J Hum Genet* 2013; 93(4):687-696.
- Dempster AP, Laird NM, Rubin DB. **Maximum likelihood from incomplete data via the EM algorithm.** *J Roy Statist Soc B* 1977; 39:1-38.
- Díez O, Gutiérrez-Enríquez S, Balmaña J. **Heterogeneous prevalence of recurrent *BRCA1* and *BRCA2* mutations in Spain according to the geographical area: implications for genetic testing.** *Familial Cancer* 2010; 9:187-191.
- Durocher F, Tonin P, Shattuck-Eidens D, Skolnick M *et al.* **Mutation analysis of the *BRCA1* gene in 23 families with cases of cancer of the breast, ovary, and multiple other sites.** *J Med Genet* 1996; 33(10):814-819.
- Esteban-Cardenosa E, Bolufer P, Palanca S *et al.* ***BRCA1* and *BRCA2* mutations in families studied in the program of genetic counselling in cancer of the Valencian community (Spain).** *Med Clin (Barc)* 2008; 130:121-126.
- Excoffier L, Slatkin M. **Maximization-likelihood estimation of molecular haplotype frequencies in a diploid population.** *Mol Biol Evol* 1995; 12:921-927.

- Fackenthal JD, Olopade OI. **Breast cancer risk associated with *BRCA1* and *BRCA2* in diverse populations.** *Nat Rev Cancer* 2007; 7(12):937-948.
- Ferla R, Calò V, Cascio S *et al.* **Founder mutations in *BRCA1* and *BRCA2* genes.** *Ann Oncol* 2007; 18(6):vi93-vi98.
- Ford D, Easton DF, Stratton M, *et al.* **Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families.** The Breast Cancer Consortium. *Am J Hum Genet* 1998; 62:676-89.
- Gabriel SB, Schaffner SF, Nguyen H, *et al.* **The structure of haplotype blocks in the human genome.** *Science* (Wash DC) 2002; 296:2225-2229.
- Greenwood CMT, Sun S, Veenstra J *et al.* **How old is this mutation? – a study of three Ashkenazi Jewish founder mutations.** *BMC genet* 2010; 11:39.
- Gyapay G, Morissette J, Vignal A *et al.* **The 1993-94 genethon human genetic linkage map.** *Nat Genet* 1994; 7.
- Hall JM, Newman B, Morrow JE, *et al.* **Linkage of early-onset familial breast cancer to chromosome 17q21.** *Science* 1990; 250:1684-1689.
- Infante M, Duran M, Esteban-Cardenosa E *et al.* **High proportion of novel mutations of *BRCA1* and *BRCA2* in breast/ovarian cancer patients from Castilla-León (central Spain).** *J Hum Genet* 2006; 51:611-617.
- Janavicius R. **Founder *BRCA1/2* mutations in the Europe: implications for hereditary breast-ovarian cancer prevention and control.** *EPMA J* 2010; 1(3):397-412.

- Janavicius R, Rudaitis V, Feng BJ, *et al.* **Haplotype analysis and ancient origin of the *BRCA1* C.4035delA Baltic founder mutation.** *European Journal of Medical Genetics* 2013; 56:125-130.
- King MC. **“The race” to clone *BRCA1*.** *Science* 2014; 343:1462-1465.
- Kwong A, Ng EKO, Wong CLP *et al.* **Identification of *BRCA1/2* founder mutations in Southern Chinese breast cancer patients using gene sequencing and high resolution DNA melting analysis.** *PLOS One* 2012; DOI: 10.1371/journal.pone.0043994.
- Lakhani SR, Jacquemier J, Sloane JP *et al.* **Multifactorial analysis of differences between sporadic breast cancers and cancers involving *BRCA1* and *BRCA2* mutations.** *J Natl Cancer Inst* 1998; 90:1138-1145.
- Lin S, Cutler DJ, Zwick ME, Chakravarti A. **Haplotype inference in random population samples.** *Am J Hum Genet* 2002; 71:1129-1137.
- Liu CY, Flesken-Nikitin A, Li S *et al.* **Inactivation of the mouse *BRCA1* gene leads to failure in the early postimplantation development.** *Genes Dev* 1996; 10:1835-1843.
- Llort G, Yagüe C, Peris M *et al.* **Low frequency of recurrent *BRCA1* and *BRCA2*. Mutations in Spain.** *Hum Mutat* 2002; 19(3):307.
- Lucas AL, Shakya R, Lipsyc MD *et al.* **High prevalence of *BRCA1* and *BRCA2* germline mutations with loss of heterozygosity in a series of resected pancreatic adenocarcinoma and other neoplastic lesions.** *Clin. Cancer Res* 2013; 19:3396-3403.
- Miki Y, Swensen J, Shattuck-Eidens D *et al.* **A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*.** *Science* 1994; 266(5182):66-71.

- Miramar MD, Calvo MT, Rodriguez A *et al.* **Genetic analysis of *BRCA1* and *BRCA2* in breast/ovarian cancer families from Aragon (Spain): two novel truncating mutations and a large genomic deletion in *BRCA1*.** *Breast Cancer Res Treat* 2008; 112:353-358.
- Moynahan ME, Chiu JW, Koller BH, Jasin M. ***BRCA1* controls homology-directed DNA repair.** *Mol Cell* 1999; 4:511-518.
- Moynahan ME, Pierce AJ, Jasin M. ***BRCA2* is required for homology-directed repair of chromosomal breaks.** *Mol Cell* 2001; 7:263-272.
- Neuhausen SL, Godwin AK, Gershoni-Baruch R, *et al.* **Haplotype and phenotype analysis of nine recurrent *BRCA2* mutations in 111 families: results of an international study.** *Am J Hum Genet* 1998; 62:1381-1388.
- Patel KJ, Yu VP, Lee H *et al.* **Involvement of *BRCA2* in DNA repair.** *Mol Cell* 1998; 1:347-357.
- Peixoto A, Santos C, Pinheiro M *et al.* **International distribution and age estimation of the Portuguese *BRCA2* c.156\_157insAlu founder mutation.** *Breast Cancer Res Treat* 2011; 127:671-679.
- Phillips MS, Lawrence R, Sachidanandam R, *et al.* **Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots.** *Nat Genet* 2003; 33:382-387.
- Rubinstein W. **Hereditary breast cancer in Jews.** *Familial Cancer* 2004; 3:249-257.
- Salazar R, Cruz-Hernandez JJ, Sanchez-Valdivieso E *et al.* ***BRCA1-2* mutations in breast cancer: identification of nine new variants of *BRCA1-2* genes in a population from central Western Spain.** *Cancer Lett* 2006; 233:172-177.

- Scully R, Che J, Plug A *et al.* **Association of *BRCA1* with Rad51 in mitotic and meiotic cells.** *Cell* 1997; 88:265-275.
- Sharan SK, Moritatsu M, Albrecht U *et al.* **Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking *BRCA2*.** *Nature* 1997; 386:804-810.
- Shen SX, Weaver Z, Xu X *et al.* **A targeted disruption of the murine *BRCA1* gene causes gamma-irradiation hypersensitivity and genetic instability.** *Oncogene* 1998; 17:3115-3124.
- Shifman S, Darvasi A. **The value of isolated populations.** *Nat Genet* 2001; 28:309-10.
- Stephens M, Smith NJ, Donnelly P. **A new statistical method for haplotype reconstruction from population data.** *Am J Hum Genet* 2001; 68:978-989.
- Stephens M, Donnelly P. **A comparison of bayesian methods for haplotype reconstruction.** *Am J Hum Genet* 2003; 73:1162-1169.
- Szabo CI, King MC. **Population genetics of *BRCA1* and *BRCA2*.** *Am J Hum Genet* 1997; 60(5):1013-1020.
- Thompson D, Easton D. **Breast Cancer Linkage Consortium.** *Am. J. Hum. Genet.* 2001;68:410-419.
- Torres D, Rashid MU, Gil F, *et al.* **High proportion of *BRCA1/2* founder mutations in Hispanic breast/ovarian cancer families from Colombia.** *Breast Cancer Res Treat* 2007; 103(2):225-232.
- Vega A, Campos B, Bressac-De-Paillerets B *et al.* **The R71G *BRCA1* is a founder Spanish mutation and leads to aberrant splicing of the transcript.** *Hum Mutat* 2001; 17(6):520-521.

- Vega A, Torres M, Martínez JI *et al.* **Analysis of *BRCA1* and *BRCA2* in breast and breast/ovarian cancer families shows population substructure in the Iberian peninsula.** *Ann Hum Genet* 2002; 66(Pt 1):29-36.
- Venkitaraman AR. **Cancer suppression by the chromosome custodians, *BRCA1* and *BRCA2*.** *Science* 2014; 343:1470-1475.
- Weber JL, Wong C. **Mutation of human short tandem repeats.** *Hum Mol Genet* 1993; 2(8):1123-1128.
- Wooster R, Weber BL. **Breast and ovarian cancer.** *N Engl J Med* 2003; 348(23):2339-2347.
- Yu VP, Kochler M, Steinlein C *et al.* **Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following *BRCA2* inactivation.** *Genes Dev* 2000; 14:1400-1406.
- Zhao H, Pfeiffer R, Gail MH. **Haplotype analysis in population genetics and association studies.** *Pharmacogenomics* 2003; 4(2):171-178.