



UNIVERSIDAD NACIONAL DE COLOMBIA

Asociación de variantes genéticas en el gen *DEAR1* con cáncer de seno y desenlace clínico en población colombiana

Ángela Patricia Beltrán López

Universidad Nacional de Colombia
Facultad de ciencias, Instituto de Biotecnología
Bogotá, Colombia
2017

Asociación de variantes genéticas en el gen *DEAR1* con cáncer de seno y desenlace clínico en población colombiana

Ángela Patricia Beltrán López

Trabajo de investigación presentado como requisito parcial para optar al título de:

Doctor en Biotecnología

Director (a):

PhD., Fabio Ancízar Aristizábal Gutiérrez.

Línea de investigación:

Marcadores moleculares en cáncer de seno

Grupo de Investigación:

Farmacogenética del cáncer

Universidad Nacional de Colombia

Facultad de ciencias, Instituto de Biotecnología

Bogotá, Colombia

2017

A Dios quien me abrió las puertas para seguir adelante con este trabajo y me dio la sabiduría para culminarlo con éxito.

A mis padres y hermanos quienes siempre me apoyaron y nunca me dejaron desfallecer.

A mi amor quien siempre fue paciente y creyó en mí.

A toda mi familia en especial a mi abuelo quien siempre se preocupó por mí.

Agradecimientos

Quiero expresar mis más sinceros agradecimientos a todas las instituciones y personas que hicieron posible el desarrollo de este trabajo, especialmente:

Al programa crédito- beca de COLCIENCIAS doctorados nacionales por creer en mi trabajo y darme la oportunidad de escalar un peldaño más en mi vida profesional.

A la Universidad Nacional de Colombia, Doctorado en Biotecnología por contribuir en mi desarrollo personal y profesional.

A la Pontificia Universidad Javeriana, Universidad de la Sabana y Hospital San Ignacio por permitirme el ingreso a sus instalaciones para el desarrollo de mi proyecto de investigación.

Al instituto de investigación en cáncer de Alemania DKFZ por su apoyo para la ejecución de las pruebas moleculares, especialmente a la Dra. Ute Hamann, por la oportunidad de hacer parte de su grupo de investigación.

A mi director el Doctor Fabio Ancizar Aristizabal, director del Instituto de Biotecnología de la Universidad Nacional de Colombia por su confianza y apoyo en todo momento durante el desarrollo de mi doctorado, por su preocupación en todo momento y su paciencia y celeridad para tomar decisiones.

A mi asesor el Doctor Ignacio Briceño, profesor asociado, Facultad de Medicina, Universidad de la Sabana, por su apoyo incondicional, disposición y colaboración en el desarrollo del proyecto de investigación.

A la Dra. Diana María Torres, por sus enseñanzas tanto académicas como personales, por sus consejos y su comprensión.

A mi amiga Viviana Ariza estudiante del Doctorado en Biotecnología de la Universidad Nacional de Colombia por su compañía y consejos durante el desarrollo de mi trabajo.

A mi amiga Yamile Bernal, asistente del Instituto de genética de la Pontificia Universidad Javeriana. Por toda su colaboración dentro de este proceso.

A mis compañeros de laboratorio en el Medical Research Center Universidad de la Sabana, Campus Biomédico: Daniel, Julián, Estefanía, Wendy y Andrés por su colaboración, tiempo y por compartir su conocimiento conmigo.

Al Doctor Carlos Guerrero, profesor Facultad de Medicina, Universidad Nacional de Colombia por su preocupación, su colaboración y sus críticas en el desarrollo de mi trabajo de investigación.

A Raquel Noguera asistente del Doctorado en Biotecnología de la Universidad Nacional de Colombia por su celeridad en todos los procesos administrativos asociados con el desarrollo de mi tesis doctoral.

A mis profesores, quienes me dieron las pautas para fortalecer mi proyecto de investigación, por su amistad y confianza.

A mis padres, por haberme inculcado el valor de la educación, por ser la razón de querer ser mejor cada día, gracias por estar siempre ahí de manera incondicional y perdonar mis ausencias. Por respetar siempre mis elecciones y no dejarme dar por vencida jamás.

A Javi, no alcanzan las palabras para agradecerte todos estos años de aguante y apoyo incondicionales, por compartir tu vida conmigo, por estar siempre ahí. Te amo.

A mi amado abuelo, gracias por tanta preocupación, finalmente lo logré.

A todos, infinitas gracias por ayudarme a llegar hasta acá.

Resumen

El cáncer de seno es el cáncer más común y también es la primera causa de muerte por cáncer en mujeres a nivel mundial. En Colombia es la primera causa de muerte por cáncer en mujeres, de ahí la importancia en la identificación de genes que puedan estar implicados en el desarrollo y progresión de la enfermedad. En este trabajo se llevó a cabo un estudio de asociación tipo gen candidato en el gen *DEAR1*. El gen *DEAR1*, codifica para un miembro de la subfamilia TRIM de proteínas “finger RING” (TRIM 62), está ubicado dentro del cromosoma 1p35.1, presenta mutaciones y deleciones homocigotas en cáncer de seno y su expresión esta disminuida o pérdida en DCIS (carcinoma ductal *in situ*). *DEAR1* es una E3 ubiquitin ligasa, involucrada en el proceso de morfogénesis acinar de la glándula mamaria. Es un supresor tumoral que actúa como un regulador dominante de polaridad, arquitectura del tejido (Lott ST *et al.*, 2009) y como un regulador del proceso de transición epitelio mesenquimal dirigido por *TGFβ* (Chen N *et al.*, 2013). El objetivo principal del estudio fue determinar la asociación de variantes moleculares en el gen *DEAR1* con cáncer de seno, parámetros histopatológicos y sobrevida usando una aproximación “tagging SNP” en una muestra de población Colombiana. Se realizó un estudio caso – control que incluyo 1023 pacientes afectadas con cáncer de seno (casos) y 1023 pacientes sin la enfermedad (controles), se incluyeron datos epidemiológicos, histopatológicos y de sobrevida. Se realizó la genotipificación de cuatro SNPs rs584298, rs2927970, rs59983645 y rs599167 que capturaron el 93,8% de la variación genética en la región investigada considerando un MAF \geq 5% y $r^2\geq$ 0,8, para su genotipificación se utilizaron dos aproximaciones metodológicas iPLEX y KASP. Se realizó el análisis de asociación por regresión condicional logística para seis variables de riesgo epidemiológico y variables patológicas. Los resultados revelaron una asociación entre el tagSNP rs584298 genotipos AG y GG con cáncer de seno [$p=0.048$ y $p=0.004$ respectivamente] y el rs584298 (genotipo AG $p=0.015$) mostró asociación con la expresión de PR y el genotipo AA ($p=0.048$) mostró asociación con la expresión de HER2. La secuenciación de la región codificante del gen *DEAR1* en tejidos tumorales de cáncer de seno y tejidos normales embebidos en parafina fue realizada con el fin de identificar nuevas variantes. Los análisis de secuencia realizados con el software CLC genomics workbench revelaron una baja longitud de distribución de las lecturas (1000pb). En el análisis se evidenció la presencia de una variante ya anotada en el NCBI rs622407 que se localiza en el dominio RING del gen *TRIM62*, es un SNP sinónimo que no altera la funcionalidad de la proteína. Estos resultados sugieren que variantes en *DEAR1* están asociadas con cáncer de seno en una muestra de la población colombiana lo cual se traduciría en el posible uso de estas variantes como factores pronósticos de la enfermedad.

Palabras clave: cáncer de seno, morfogénesis, DEAR1/TRIM62, factores de riesgo genético

Abstract

Breast cancer is the most common cancer and is the leading cause of cancer death in women worldwide. In Colombia, it is the leading cause of cancer death in women. For this reason is important to identify genes that may be involved in the development and progression of the disease. In this work, a candidate gene association study was carried out in the *DEAR1* gene. The *DEAR1* gene, coding for a member of the subfamily TRIM of finger RING proteins (TRIM 62), is located within chromosome 1p35.1, that is mutated and homozygously deleted in breast cancer and whose expression is down regulated/lost in DCIS. *DEAR1* is an E3 ubiquitin ligase, involved in the process of acinar morphogenesis of the mammary gland. Previous work has shown that *DEAR1* is a novel tumor suppressor that acts as a dominant regulator of polarity, tissue architecture, (Lott ST et al., 2009) and TGF β -driven epithelial-mesenchymal transition (EMT) (Chen N et al., 2013). The main objective of this study was to determine the association of molecular variants in the *DEAR1* gene with breast cancer, histopathological parameters and survival using a tagging SNP approach in a Colombian population sample. A case - control study was carried out, including 1023 patients with breast cancer (cases) and 1023 patients without the disease (controls), including epidemiological, histopathological and survival data. Genotyping of four SNPs rs584298, rs2927970, rs59983645 and rs599167, which captured 93.8% of the genetic variation in the region under investigation considering an MAF $\geq 5\%$ and $r^2 \geq 0.8$, were performed for genotyping two methodological strategies IPLEX and KASP. The association analysis was performed by logistic conditional regression for six variables of epidemiological risk and pathological variables. Associations were found between tagSNPs and breast cancer adjusted for the epidemiological risk factors rs584298 genotypes AG and GG ($p=0.048$ and $p= 0.004$, respectively). Analysis of the disease characteristics showed that SNPs rs584298 (genotype AG) ($p = 0.015$) show association with PR status and (genotype AA) ($p= 0.048$) show association with HER2 status. Sequencing of the coding region of the *DEAR1* gene in breast tumor tissues and normal tissues embedded in paraffin was performed in order to identify new variants. Sequence analyzes performed with the CLC software genomics workbench revealed a low reading distribution length (1000pb). In the analysis, the presence of a variant already noted in the NCBI rs622407 that is located in the RING domain of the *TRIM62* gene shown to be a synonymous SNP that does not alter the functionality of the protein. These results suggest that variants in *DEAR1* are associated with breast cancer in a sample of the Colombian population, which would translate into the possible use of these variants as prognostic factors for the disease.

Keywords: Breast cancer, morphogenesis, cell polarity, *DEAR1/TRIM62*, genetic risk factors.

Contenido

Resumen	VII
Abstract	VIII
Lista de tablas	XIII
Lista de figuras	XV
Lista de abreviaturas	XVI
Introducción	1
1.Marco teórico	5
1.1 Cáncer de seno.....	5
1.1.1 Características generales del cáncer de seno	7
1.1.2 Factores de riesgo asociados a cáncer de seno.....	16
1.1.3 Genética del cáncer de seno	20
1.2 Ubiquitinación y su rol en cáncer de seno	22
1.2.1 Ubiquitinación y degradación proteasomal	23
1.2.2 Ubiquitin ligasas en cáncer de seno	24
1.3 Dear1: Clasificación, Estructura Y Asociación Con Cáncer De Seno	27
1.3.1 Clasificación Familia de proteínas RBCC/TRIM.....	27
1.3.2 Descripción del gen DEAR1	28
1.3.3 <i>DEAR1</i> y su relación con Cáncer de seno.....	28
1.4 Estudios de asociación en cáncer de seno	31
1.4.1 Polimorfismos de único nucleótido (SNPs) y su asociación con cáncer de seno	31
1.4.2 Estudios de Asociación gen candidato.	32
1.5 Estrategias metodológicas para el análisis de SNPs.....	34
1.5.1 Métodos para el análisis de SNPs.....	34
1.5.2 Plataformas comerciales para el análisis de SNPs.....	36
2.METODOLOGÍA	39

2.1 Selección de SNPs candidatos (TagSNPs).....	39
2.1.1 Análisis genéticos- Genotipificación Tag SNPs	39
2.1.2 Descripción de la Población de estudio	39
2.1.3 Extracción de ADN	41
2.1.4 Cuantificación de ADN	42
2.1.5 Genotipificación.....	42
2.1.6 Control de Calidad.....	43
2.1.7 Protocolo de genotipificación.....	43
2.1.8 Estimación de las frecuencias alélicas y genotípicas para los tag SNPs en TRIM62	49
2.2 Secuenciación de tejido tumoral	50
2.2.1 Amplificación de ADN de tejido tumoral y normal	51
2.2.2 Análisis de secuencias	54
2.3 Equilibrio Hardy-Weinberg	¡Error! Marcador no definido.
2.4 Análisis estadísticos.....	58
2.4.1 Asociación entre genotipos de TRIM62 con las variables epidemiológicas e histopatológicas.....	57
2.4.2 Análisis de supervivencia	58
2.5 Consideraciones éticas.....	59
3.Resultados y discusión.....	60
3.1 Descripción de la población de estudio	60
3.1.1 Casos y controles.....	60
3.1.2 Tejidos embebidos en parafina.....	66
3.2 Selección de los “Tag SNPS”.....	68
3.2.1 Primera aproximación “tagging SNPs” comparando dos poblaciones.....	68
3.2.2 Segunda aproximación “Tagging SNPs” sobre la población CLM.....	70
3.2.3 Tercera aproximación tag SNP sobre el poder para población CLM.....	71
3.3 Genotipificación Tag SNPs en casos y controles	73
3.3.1 Descripción genotipos y frecuencias alélicas.....	73
3.3.2 Equilibrio de Hardy Weinberg	75
3.3.3 Análisis de asociación de los genotipos y las variables epidemiológicas con cáncer de seno.....	80
3.3.4 Análisis de asociación de los genotipos con las variables histopatológicas del tumor.....	82

3.3.5 Análisis de supervivencia	83
3.4 Secuenciación de los tejidos embebidos en parafina	84
3.4.1 Análisis de las secuencias obtenidas.....	84
4.Consideraciones finales	91
5.Conclusiones y recomendaciones.....	97
5.1 Conclusiones	97
5.2 Recomendaciones	98
6.Referencias.....	100
A. Anexo: Consentimiento informado.....	110
B. Anexo: Cuestionario sobre factores potenciales de riesgo para pacientes con cáncer de seno	114
C. Anexo: Cuestionario sobre factores potenciales de riesgo para controles.....	124
D. Anexo: Cuestionario sobre características clínicas e histopatológicas	137
E. Anexo: Protocolo de extracción de ADN desde sangre periférica	139
F. Anexo: Extracción de ADN desde tejidos embebidos en parafina.	141
G. Anexo: Representación gráfica de los resultados de genotipificación.....	144
H. Anexo: Análisis de colinealidad.....	146
I. Anexo: Análisis de supervivencia	149
J. Anexo: Modelo final con selección de variables usando la aproximación Stepwise.....	152

Lista de tablas

Tabla 1-1: Características de subtipos moleculares de cáncer de seno.	15
Tabla 1-2: Factores de riesgo asociados con cáncer de seno.....	19
Tabla 1-3: E3 ligasas asociadas con cáncer de seno como oncogenes,.....	25
Tabla 1-4 : E3 ligasas asociadas con cáncer de seno como supresores tumorales	26
Tabla 2-1: Secuencia cebadores ensayo IPLEX	44
Tabla 2-2: Mezcla de reacción y condiciones de incubación usadas para la reacción con SAP de los SNPs rs2927970, rs584298, rs59983645.....	45
Tabla 2-3: Cebadores de extensión usados para la reacción iPLEX de los SNPs rs2927970, rs584298 y rs59983645	45
Tabla 2-4: Mezcla de reacción usada para la reacción iPLEX de los SNPs rs2927970, rs584298 y rs59983645.....	46
Tabla 2-5: Condiciones de amplificación para la reacción IPLEX.....	46
Tabla 2-6: Cebadores ensayo KASP	48
Tabla 2-7: Mezcla de reacción para el ensayo KASP del SNP rs599167	48
Tabla 2-8: Condiciones de PCR para el ensayo KASP del SNP rs599167.....	49
Tabla 2-9: Valores de excitación y emisión de los fluoróforos usados en el ensayo KASP.	49
Tabla 2-10: Genotipos y alelos para cada uno de los tag SNPs analizados.....	50
Tabla 2-11: Mezcla de reacción empleada en la amplificación de exones TRIM62.....	51
Tabla 2-12: Cebadores usados para la amplificación de los exones del gen TRIM62....	51
Tabla 2-13: Condiciones de PCR empleadas en la amplificación de los exones TRIM62	52

Tabla 3-1: Estadística descriptiva variables categóricas de riesgo epidemiológico incluidas en el análisis.....	61
Tabla 3-2: Estadística descriptiva variables continuas incluidas en el estudio	62
Tabla 3-3: Estadística descriptiva variables histopatológicas categóricas.....	63
Tabla 3-4: Estadística descriptiva variables histopatológicas continuas.....	64
Tabla 3-5: Estadística descriptiva variables seguimiento	65
Tabla 3-6: Variables histopatológicas de los tejidos tumorales	66
Tabla 3-7: Variables histopatológicas continuas tejidos tumorales.....	67
Tabla 3-8: Alelos capturados por los "Tag SNPs" segunda aproximación.....	70
Tabla 3-9: Alelos capturados por los "TagSNPs" tercera aproximación.....	72
Tabla 3-10: Frecuencias de genotipos de los 4 Tag SNPs en casos y controles.....	74
Tabla 3-11: Prueba para la desviación del equilibrio de Hardy Weinberg rs584298.....	76
Tabla 3-12: Prueba para la desviación del equilibrio de Hardy Weinberg rs2927970.....	77
Tabla 3-13: Prueba para la desviación del equilibrio de Hardy Weinberg rs59983645... ..	78
Tabla 3-14: Prueba para la desviación del equilibrio de Hardy Weinberg rs599167.....	79
Tabla 3-15: Estadísticos del análisis Stepwise para el modelo final de regresión.....	80
Tabla 3-16: Distribución alélica rs622407 en muestras de tejido tumoral	88
Tabla 3-17: Prueba de desviación del equilibrio de Hardy-Weinberg para los tejidos embebidos en parafina	90

Lista de figuras

Figura 1-1: Tasas de incidencia y mortalidad de cáncer de seno estandarizados por edad a nivel mundial.	6
Figura 1-2: Tasas de incidencia y mortalidad de cáncer de seno estandarizadas por edad en Colombia.	6
Figura 1-3: Anatomía del seno.	8
Figura 1-4: Localización cromosómica de <i>DEAR1</i> y representación gráfica de los exones y dominios de la proteína (RING:Really interresting new gene;Znf_BB: B-Box-type zinc finger, BBC: B-Box dominio C-terminal; dominio PRY; dominio SPRY:)	28
Figura 1-5: Hibridación alelo específica..	35
Figura 1-6: Extensión de primer alelo-específica.	35
Figura 1-7: Ligación de oligonucleótidos.....	36
Figura 1-8: Ensayo TaqMan..	37
Figura 3-1 : Longitud de distribución de las secuencias.....	85
Figura 3-2 : Porcentaje de GC	85
Figura 3-3: Contenido de bases ambiguas "Ns"	86
Figura 3-4: Distribución de la calidad PHRED score.....	86
Figura 3-5: Cubrimiento de las secuencias.....	87
Figura 3-6 : SNP en la posición 722	88

Lista de abreviaturas

ADN	ácido desoxirribonucleico
AR	receptor de andrógenos
ASO	Oligonucleótido alelo específico
ASR	tasas estandarizadas por edad
ATM	Gen ataxia telangiectasia mutado,
BCAC	consorcio de cáncer de seno
BRCA1	Gen cáncer de seno 1
BRCA2	Gen cáncer de seno 2
BRIP1	proteína de interacción con BRCA1, C-terminal helicasa 1
CASP8	Caspasa 8
CDH1	Caderina 1
CEU	Población incluida en el consorcio HapMap. Residentes de Utah con ancestría del norte y este de Europa.
CHEK2	Serina treonina quinasa no específica
CHFR	Punto de control con dominios "Forkhead" y RING
CHIP	E3 ubiquitin ligasa homólogo STIP1 y proteína conteniendo U-box 1
CIMBA	Consorcio de investigadores de modificadores de BRCA1/2
CLM	Población desde el proyecto de los 1000 genomas. Población de Medellín Colombia
DEAR1	Nombre del gen de interés "Ductal Epithelium Ring Chromosome 1"
DL	desequilibrio de ligamiento
dNTPs	Desoxynucleotidos trifosfato
E6-AP	Proteína asociada a E6
EDTA	Ácido etilendiaminotetraacético
EGFR	Receptor del factor de crecimiento epidermal
EMT	Transición epitelio mesenquimal
ER	receptor de estrógenos
FGFR2	Receptor del factor de crecimiento 2
FRET	Transmisión de energía de resonancia
H&E	Hematoxilina- eosina
HapMap	Mapa de Haplotipos
HECT	Dominio homólogo a E6-AP carboxy terminal
HER1	receptor 2 de factor de crecimiento epidérmico humano
HER2	receptor 2 de factor de crecimiento epidérmico humano

HUSI	Hospital Universitario San Ignacio
HWE	Equilibrio de Hardy Weinberg
IC	Intervalos de confianza
IDC-NST	Carcinoma in situ- de tipo no especial
IHC	Inmunohistoquímica
IMC	Índice de masa corporal
KASP	PCR competitiva alelo- específica
KIT	receptor tirosina quinasa
LOH	Pérdida de la heterocigosidad
LSO	Oligonucleótido locus específico
LSP1	Proteína específica de linfocitos 1
MAF	Frecuencia Alélica Menor
MALDI-TOF	desorción/ionización mediante láser asistida por Matriz), acoplada a un analizador TOF (tiempo de vuelo),
MAPK	proteína quinasas activadas por mitógenos
mRNA	Ácido ribonucleico mensajero
MYC	Protooncogen
NF-$\kappa$$\beta$	Factor nuclear kappa-beta
NTCs	Control
OLA	Ligación de oligonucleótidos
OR	Razón de odds
P27Kip1	El inhibidor 1B de quinasa dependiente de ciclina
P53	proteína supresora de tumores de 53 kd
PALB2	Compañero y localizador de BRCA2 "Partner and Localizer of BRCA2"
PCR	Reacción en cadena de la polimerasa
PI3K	Fosfoinositol 3-quinasas
PTEN	Fosfatidilinositol-3,4,5-trisfosfato 3-fosfatasa
RAD51	Proteína de reparación del ADN 1
RhoA	Miembro de la familia de genes homólogos Ras A
RING	Dominio de las proteínas TRIM " Really New Interesting Gene"
RP	receptor de progesterona
RR	riesgo relativo
SIAH1	Siete en ausencia del homologo 1
SMAD3	Miembro de la Familia SMAD 3 "Mothers Against DPP Homolog 3"
SNPs	Polimorfismo de único nucleótido
SSH	Hibridación supresora substractiva
STAT1	Transductor de señal y activador de la transcripción1
STK11	Serina/treonina quinasa 11
TDLU	Unidades lobulares ductales terminales
TEB	Yemas terminales "Terminal End Buds"
TGFβ1	Factor de crecimiento transformante 1
TOX	Proteína Tox del grupo de alta movilidad asociada a la selección de timocitos
TRIM	Familia de proteínas motivos tripartito "Tripartite Motif"

Introducción

El cáncer de seno es una enfermedad con altas de tasas de incidencia y mortalidad a nivel mundial y es la primera causa de muerte en mujeres por cáncer en el mundo, razón por la cual en los últimos años se han realizado diferentes esfuerzos a nivel genómico, transcritoómico, proteómico y epigenómico con el fin de elucidar otros genes que puedan estar implicados en el desarrollo del mismo y que puedan ser útiles como marcadores pronósticos y predictivos de la enfermedad, así como blancos terapéuticos que lleven al desarrollo de nuevos fármacos y así evolucionar hacia una medicina personalizada y de precisión.

Los estudios a nivel genómico a través del tiempo han permitido identificar genes de alta penetrancia implicados en el desarrollo de la enfermedad principalmente en familias con componente hereditario. Los estudios de "*asociación genética*" buscan establecer la relación estadística entre variables genéticas poblacionales y un fenotipo determinado (ejemplo: rasgo, riesgo de enfermedad, etc.). Estos están siendo utilizados para descubrir el componente genético que subyace a las enfermedades de alta prevalencia como el cáncer de seno. Se trata generalmente de estudios de cohortes prospectivas o de tipo casos-contróles en los cuales se establece el peso relativo del componente genómico con respecto a otros factores asociados con el riesgo de desarrollar la enfermedad.

Habitualmente, se utilizan como marcadores genéticos a los polimorfismos de único nucleótido (SNPs). Estas variaciones pueden ser en sí mismas funcionales y estar relacionadas a la fisiopatología de la enfermedad, pero en la mayoría de los casos, son utilizadas para el mapeo y ubicación de los verdaderos sitios relevantes. Los dos acercamientos posibles son el del "*gen candidato*" cuando existe evidencia previa de funcionalidad de la variante, o el de la "asociación indirecta".

Actualmente están en curso innumerables estudios de asociación basados en la diferente distribución de frecuencias de variantes genéticas en grupos de individuos afectados con respecto a individuos sanos, existen plataformas que permiten analizar múltiples polimorfismos genéticos, estos avances en las técnicas de genotipado van ligados también a un mayor conocimiento de la arquitectura global del genoma humano, así como de la distribución y localización de los polimorfismos genéticos a lo largo del mismo. Como resultado del International Haplotype Mapping Project (International HapMap Consortium) ("The International HapMap Project," 2003a) las bases de datos públicas contienen datos sobre las frecuencias verificadas de más de 2 millones de SNPs. La intención es caracterizar el patrón de desequilibrio de ligamiento (DL) a lo largo del genoma para facilitar la selección de los SNPs con mayor información (tag SNP) para ser incluidos en estudios genéticos. Estos "tag SNP" permiten capturar la variabilidad de cada uno de los bloques haplotípicos sin necesidad de genotipar todos los SNP que forman parte de la región, lo que permite estudiar la variabilidad genética en amplias zonas del genoma. Los resultados de estos estudios con multitud de polimorfismos genéticos requieren herramientas estadísticas que permitan su análisis. En este sentido se están desarrollando algoritmos apropiados que permitan analizar los datos teniendo en cuenta factores determinantes como las relaciones de DL, el equilibrio de Hardy-Weinberg (HWE) o las correcciones por múltiples comparaciones.

Es claro que el proceso de carcinogénesis es un proceso de múltiples pasos donde genes supresores de tumor pierden su función y oncogenes se sobre expresan para dar lugar al proceso de transformación. La iniciación y progresión del cáncer de seno involucra diferentes mecanismos, principalmente relacionados con control del ciclo celular, proliferación, diferenciación y arquitectura del tejido, por lo cual la búsqueda de genes con funciones pleiotrópicas relacionadas principalmente con estos mecanismos de control es de gran interés.

El gen *DEAR1* es un miembro de la familia de proteínas TRIM la cual se caracteriza por participar principalmente en procesos de ubiquitinación, este grupo de proteínas se ha visto implicado en la regulación de diferentes procesos fundamentales en la célula como progresión del ciclo celular, apoptosis, regulación transcripcional. Además, se ha observado su relevancia clínica principalmente asociada con inmunidad, cáncer y en el

desarrollo de enfermedades genéticas. El estudio llevado a cabo por el grupo de investigación de la Dra. Killary del MD Anderson cáncer Center (Lott et al., 2009b), describió por primera vez este gen y se mostró su asociación como regulador dominante de morfogénesis acinar en tejido mamario y como predictor independiente de supervivencia libre de enfermedad a edad temprana en cáncer de seno (Lott et al., 2009a). Otros estudios han enmarcado a *TRIM62* como factor pronóstico adverso en leucemia mieloide aguda (Quintas-Cardama et al., 2015) y como un regulador clave en el proceso EMT (transición epitelio mesenquimal) en cáncer de pulmón y mama, promoviendo metástasis e invasividad a otros órganos (N. Chen et al., 2013; Quintas-Cardama et al., 2014) . Hasta el momento no se han llevado a cabo otro tipo de estudios, razón por la cual realizamos un estudio caso-control en una muestra de población colombiana para así determinar la asociación de variantes genéticas con una frecuencia alélica menor de 0.05 (MAF > 5%) en el gen *TRIM62* con cáncer de seno en nuestra población y saber si este puede ser un marcador pronóstico útil de la enfermedad.

OBJETIVOS

Objetivo general

Determinar si el gen *DEAR1* actúa como un factor de riesgo de baja penetrancia para cáncer de seno y está asociado con desenlace clínico en población colombiana

Objetivos específicos

Evaluar la asociación de variantes genéticas en el gen *DEAR1* con riesgo a cáncer de seno usando una aproximación de etiquetado de SNPs (tagging SNP) y datos desde el proyecto internacional HapMap (CEU) y el proyecto de los 1000 genomas (CLM).

Evaluar si variantes genéticas en *DEAR1* están asociadas con parámetros histopatológicos del tumor y sobrevida.

Identificar mutaciones somáticas en *DEAR1* por secuenciación tanto de la región promotora como codificante de al menos 20 tumores de seno.

1. Marco teórico

1.1 Cáncer de seno

El cáncer de seno es el más frecuente en mujeres a nivel mundial. Con un número estimado de 1.67 millones nuevos casos de cáncer de seno diagnosticados en 2012 y 522.000 muertes causadas por esta enfermedad (Ferlay et al., 2015). Actualmente el cáncer de seno es la primera causa de muerte en mujeres por cáncer en regiones menos desarrolladas (324.000 muertes, 14.3%). Alrededor de la mitad de los casos de cáncer de seno y 60% de las muertes se estima que ocurren en países en vía de desarrollo. Las tasas de incidencia son más altas en el norte y occidente de Europa, Australia/ nueva Zelanda y Norteamérica; intermedias en Suramérica, el Caribe, Norte de África; y bajas en África subsahariana y Asia (Ferlay et al., 2015).

En población colombiana el cáncer de seno es el más frecuente en mujeres, con tasas intermedias de incidencia y mortalidad, de acuerdo a estadísticas de GLOBOCAN, la tasa de incidencia en Colombia fue de 23.4% (8.686 casos) y la tasa de mortalidad 13.9% (2.649 muertes) (Ferlay et al., 2015). Los departamentos que muestran el mayor número de casos anuales según estadísticas del grupo de vigilancia epidemiológica, (INCC- instituto nacional de cancerología), son en su orden, Bogotá, Antioquia y Valle del cauca. Las tasas de incidencia y mortalidad para cáncer de seno estandarizadas por edad (ASR) en el mundo y las tasas de incidencia y mortalidad en Colombia son mostradas en las figuras 1-1 y 1-2, respectivamente.

Figura 1-1: “Tasas de incidencia y mortalidad de cáncer de seno estandarizados por edad a nivel mundial”. Fuente: (Ferlay et al., 2015)

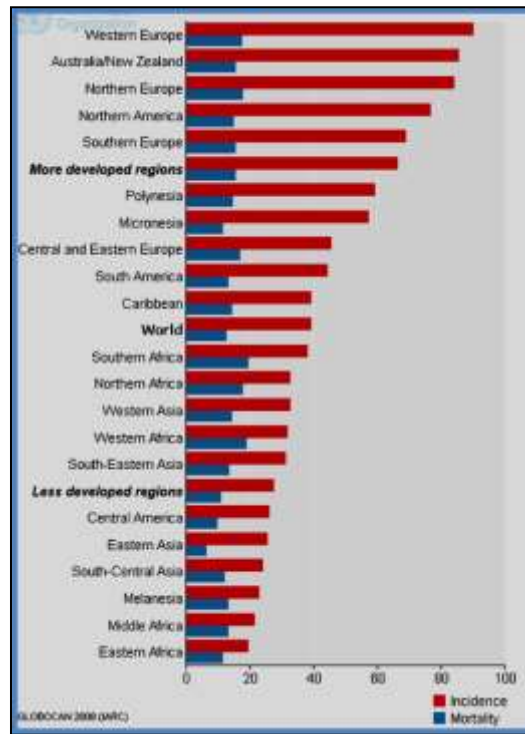
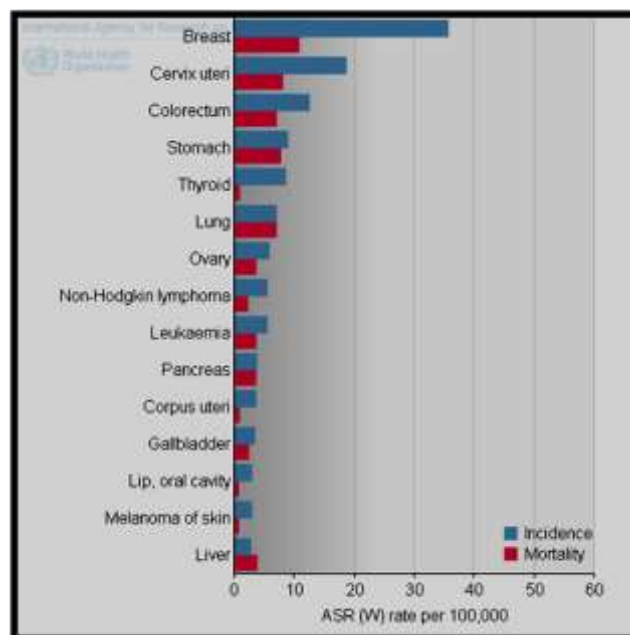


Figura 1-2: Tasas de incidencia y mortalidad de cáncer de seno estandarizadas por edad en Colombia. Fuente: (Ferlay et al., 2015)



1.1.1 Características generales del cáncer de seno

El cáncer de seno es una enfermedad genética. Su desarrollo ha sido descrito como un proceso de múltiples pasos, los cuales están asociados con transformación de células normales a malignas atravesando por hiperplasia, cambios pre-malignos y carcinoma *in situ*, en el cual cada paso se correlaciona con una o varias mutaciones en genes regulatorios fundamentales. Estos genes juegan un papel importante en el mantenimiento del balance fisiológico ya que participan en procesos como proliferación, apoptosis y diferenciación o en la expresión de receptores esteroides, moléculas de adhesión celular y factores angiogénicos (Kenemans, Verstraeten, & Verheijen, 2004). Además, dado que el seno es un órgano dinámico el cual es continuamente remodelado por hormonas y factores de crecimiento, es posible que esta habilidad dinámica intrínseca de las células del seno, las haga especialmente susceptibles a carcinogénesis (Howard & Gusterson, 2000).

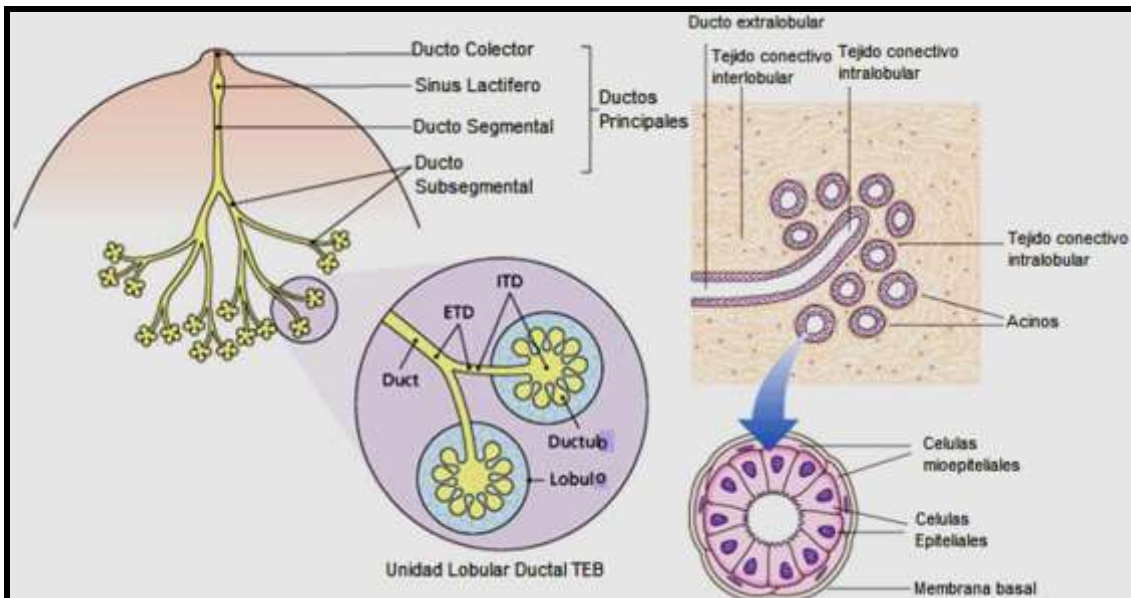
Anatomía de la glándula mamaria

La glándula mamaria no está formada completamente al nacer, por lo cual, el desarrollo mamario ocurre en tres etapas distintas y reguladas diferencialmente que son la etapa embrionaria donde se encuentra una estructura ductal rudimentaria, después del nacimiento, la glándula rudimentaria entra en una fase de quiescencia morfogénica hasta la etapa de la pubertad considerada como la etapa más importante de la morfogénesis mamaria (Anderson, 2002). El estímulo por niveles elevados de hormonas como estrógenos, promueve la proliferación y crecimiento de los ductos rudimentarios llamados yemas terminales (Terminal end buds, TEB) dentro de distintas capas epiteliales. Los ductos primarios que alcanzan el pezón dan lugar a un patrón de ramificación compleja de ductos auxiliares que guían a la formación de los ductos segmentales y a pequeños ductos subsegmentales. Los ductos subsegmentales guían a la formación de ductos terminales que dan lugar a los conductillos llamados acinos (Howard & Gusterson, 2000). Una colección de acinos derivados desde un ducto terminal embebidas en un estroma complejo, (la almohadilla de grasa mamaria, “fat pad mammary”) la cual contiene fibroblastos, adipocitos, vasos sanguíneos, nervios y células inmunes, importantes para el desarrollo y función normal de la glándula mamaria se conoce como unidad lobular ductal terminal (terminal ductal lobular units, TDLU), la cual es considerada como la unidad funcional del seno (Gjorevski & Nelson, 2011) formada por

una capa continúa de células epiteliales lumbales, rodeada por una capa de células mioepiteliales, estas células mioepiteliales están en contacto directo con la membrana basal y por último en la etapa de gestación, el epitelio luminal prolifera y se diferencia dentro de alveolos secretores productores de leche (Gjorevski & Nelson, 2011).

Diferentes estudios histológicos han mostrado que la mayoría de los tumores de seno derivan desde las TLDU y tienen características morfológicas de células epiteliales lumbales (Allred, Mohsin, & Fuqua, 2001). Además, la mayoría de tumores de seno retienen las características bioquímicas de células lumbales. Los tumores humanos también contienen receptores para estradiol y progesterona que, en el seno normal, son expresados únicamente en el compartimiento de las células lumbales. Esto sugiere que las células epiteliales lumbales son los blancos primarios para transformación maligna y subsecuente formación del tumor (Anderson, 2002). Además, el compartimiento de células lumbales en el tejido mamario posee alta heterogeneidad donde diferentes niveles de diferenciación de células lumbales y capacidades proliferativas pueden ser identificados. (Shehata et al., 2012).

Figura 1-3: Anatomía del seno. Modificado de Rosai & Ackerman's. Surgical Pathology 9e.



Tipos cáncer de seno

Hay tres tipos de cáncer de seno, el cáncer de seno de tipo esporádico, de tipo familiar y de tipo hereditario.

Cáncer de seno esporádico

El cáncer de seno esporádico representa la mayoría de casos por cáncer de seno en el mundo y se originan por mutaciones somáticas. Se define como un tipo de cáncer donde el paciente no presenta antecedentes familiares en al menos dos generaciones, incluyendo hermanos, descendientes, padres, tíos y tías, y abuelos tanto de la línea paterna como materna, incluso si hay algún caso en la familia no hay un patrón particular de herencia. Se presentan en la edad adulta (Berliner & Fay, 2007)

Cáncer de seno de tipo familiar

Se caracteriza por ser un cáncer donde la paciente afectada presenta algún antecedente familiar, que incluye uno o más familiares de primer o segundo grado con esta enfermedad, que no se ajustan a la definición de cáncer de mama hereditario y presenta una edad variable de inicio (Berliner & Fay, 2007).

Cáncer de seno de tipo hereditario

El cáncer de seno hereditario se define por la identificación de una mutación en la línea germinal en pacientes con cáncer de mama, antecedentes familiares de la enfermedad o cánceres relacionados (p. ej., ovario, colon), con alta frecuencia y distribución en la genealogía, compatible con un factor de susceptibilidad de cáncer autosómico dominante (López, 2000). Aproximadamente 5-10% de todos los cánceres de seno son hereditarios (Hofmann & Schlag, 2000). Otros factores que apoyan la clasificación del cáncer de mama hereditario incluyen inicio a una edad temprana, cáncer bilateral de la mama y otros cánceres primarios múltiples (Berliner & Fay, 2007).

Clasificación del cáncer de seno

El cáncer de seno puede ser clasificado de acuerdo al tipo histológico, grado histológico, el estado del tumor y por subtipos moleculares.

Tipos histológicos

Los carcinomas de seno están divididos dentro de carcinoma *in situ* (ductal y lobular) y carcinomas invasivos, los cuales son agrupados dentro de carcinoma ductal de tipo no especial (IDC-NST) representando el 60-75% de todos los cánceres de seno o de tipo especial, de los cuales el más común es el lobular seguido por tubular, papilar y de tipo mucinoso (Michelle Alizart, 2012).

Grado histológico

El método de graduación histológica del carcinoma de mama más comúnmente utilizado en nuestro medio es la modificación de Nottingham (Elston & Ellis, 1991), al sistema de Scarff-Bloom-Richardson (Bloom HJG, 1957). El método se basa en la asignación de una puntuación (1 a 3) a cada una de las siguientes características tumorales: formación de túbulos, pleomorfismo nuclear y número de mitosis. En la evaluación de la formación de túbulos sólo se consideran las estructuras con una clara luz central y se puntúa: 1: formación de túbulos mayor del 75% del área tumoral; 2: 10-75% del tumor; 3: menos del 10%).

El pleomorfismo nuclear se valora en relación a la variabilidad de la forma y el tamaño de los núcleos neoplásicos en comparación con los normales. La irregularidad de los contornos nucleares y el número y tamaño del nucléolo se evalúan como características adicionales. La puntuación asignada sería: 1: núcleos pequeños, regulares y uniformes con mínima variación de forma y tamaño; 2: moderada variación de forma y tamaño; 3: marcada variación de forma y tamaño.

El número de mitosis se debe evaluar en la periferia de los tumores y se deben contar exclusivamente mitosis claras, obviando aquellas imágenes dudosas que pueden corresponder a núcleos pícnóticos en apoptosis.

Una de las modificaciones más importantes en la evaluación del grado histológico en el método de Nottingham es la evaluación cuantitativa de las mitosis. Una vez sumadas las puntuaciones establecidas para cada característica tumoral, el carcinoma se considera: grado 1 (bien diferenciado), puntuación 3-5, grado 2 (moderadamente diferenciado), puntuación 6-7 y grado 3 (pobremente diferenciado) puntuación 8-9. Los tumores grado 1

tienen un mejor pronóstico en cuanto a la sobrevida libre de recurrencia y sobrevida total comparado con los tumores grado 2 y grado 3 (Metzger Filho, Ignatiadis, & Sotiriou, 2011).

Clasificación TNM

Otro parámetro importante en la clasificación del cáncer de seno es la clasificación **T** (tamaño del tumor); **N** (nódulos linfoides afectados); **M** (metástasis), fue creada para permitir la definición de categorías indicando el grado de extensión local, regional y extensión general a el tiempo de un tratamiento primario, proveyendo una descripción objetiva y permanente. Varias categorías con similar valor pronóstico pueden ser agrupadas juntas para definir los estados de la enfermedad (Veronesi, Viale, Rotmensz, & Goldhirsch, 2006). Los estadios de la enfermedad se determinan agrupando las categorías TNM.

Receptores hormonales

Los receptores de estrógenos (ER) y receptores de progesterona (PR) son evaluados rutinariamente en tejido tumoral mamario, por su habilidad para predecir beneficio de la terapia endocrina y guiar las decisiones terapéuticas (Anderson, 2002).

Tanto RE como RP son factores de transcripción nuclear dependientes de ligando que pertenecen a la familia de receptores de hormonas nucleares. Este tipo de receptores tienen varias características estructurales comunes. Estas incluyen un dominio de unión al ADN y un dominio de unión al ligando (Olefsky, 2001).

La relación entre la respuesta a terapia hormonal y el uso de RP y RE como marcadores predictivos se estableció gracias al estudio realizado por McGuire et al. (1975), donde mostró que la proteína ER estaba presente en 50% a 70% de cánceres de seno invasivos en un análisis hecho sobre 400 pacientes con cáncer de seno avanzado, además el estudio también evidenció que el 50% al 60% de mujeres que tenían tumores ER positivos respondían a la terapia endocrina comparado con un 5% -10% de tumores ER-negativos. El mismo grupo de investigación en un estudio posterior mostró que el 70% al 80% de cánceres de seno con RE y RP positivos respondían mejor a la terapia hormonal (McGuire WL, 1975), (McGuire WL, 1980). Debido a la diferencia en la respuesta a la

terapia hormonal en cánceres de seno con receptores esteroides positivos y negativos, el análisis inmunohistoquímico de estos marcadores predictivos se realiza en todos los cánceres de seno primarios, sobre bloques de tumor.

HER2

La proteína HER2 (human epidermal receptor 2), es un miembro de la subclase 1 de la superfamilia de receptores tirosina quinasa. Otros miembros de esta familia incluyen el receptor del factor de crecimiento epidermal (HER1/EGFR), HER3 y HER4. Todos los miembros de esta familia poseen un dominio de unión a ligando extracelular, una región transmembranal y un dominio citoplasmático con actividad tirosina quinasa. Aunque estos receptores comparten una estructura común, solo se han encontrado ligandos para HER-1, HER-3, y HER-4. Parece ser que HER2 es un receptor huérfano ya que no posee un ligando de unión directo. HER2 de cualquier modo, puede mediar la señalización intracelular por heterodimerización con otros miembros de la familia HER. Después de la heterodimerización, inicia la vía de señalización intracelular de MAPK, PI3K y fosfolipasa C (Olayioye, Neve, Lane, & Hynes, 2000).

En cáncer de seno humano, el incremento en el número de copias del gen *HER2* se ha encontrado en 15-30% de tumores invasivos primarios, esta se ha correlacionado generalmente con pronóstico adverso, particularmente en pacientes con cáncer de seno con nódulos linfoides positivos

La sobre-expresión de HER-2 se observa en carcinomas de mama poco diferenciados con alto grado histológico, algunos de ellos se asocian a carcinomas ductales *in situ* de tipo comedocarcinoma y habitualmente son negativos a receptores hormonales (Winston, Ramanaryanan, & Levine, 2004),(Pérez Sánchez VM, 2008)

Clasificación molecular del cáncer de seno

En el año 2000, Perou y colaboradores clasificaron el cáncer de seno dentro de subtipos intrínsecos (genes que varían principalmente entre tumores desde diferentes pacientes comparados con muestras tumorales del mismo paciente) basados sobre perfiles de

expresión de genes. Los análisis de expresión génica diferencial se realizaron en 38 muestras tumorales de cáncer de seno invasivo, 1 carcinoma ductal *in situ*, 1 fibroadenoma y 3 muestras de seno normales. Un número de replicados desde el mismo paciente definió la lista de genes intrínsecos. El análisis de agrupamiento jerárquico con esta "lista de genes intrínsecos" reveló que la principal división recaía sobre el estado del receptor de estrógenos (cánceres de seno ER positivo y cánceres de seno ER negativos) (Weigelt B, 2010), y la existencia de cinco subtipos moleculares de cáncer de seno: luminal A, luminal B, parecido al seno normal (normal breast-like), HER2 y basal (basal-like) (Perou et al., 2000; Sorlie *et al.*, 2001).

Los tumores de tipo luminal A muestran alta expresión de ER, tienen bajas tasas de proliferación, tienden a ser de bajo grado histológico y tienen el mejor pronóstico. Los tumores de tipo luminal B muestran una baja expresión de genes asociados con la vía activa de ER, tienden a tener un grado histológico alto, tienen altas tasas de proliferación y son de mal pronóstico comparado con los del tipo luminal A (Michelle Alizart, 2012).

En el grupo de ER- negativo los tres subtipos identificados fueron "normal breast-like", HER2 y "basal-like". El subtipo "normal breast-like" muestra expresión de genes asociados con el tejido adiposo. Tumores de este subtipo están pobremente caracterizados y su significancia clínica todavía no ha sido determinada. Los subtipos HER2 y "basal-like" están asociados con un comportamiento clínico agresivo. Tumores HER2 sobreexpresan HER2 y genes asociados con la vía HER2. El subtipo "basal-like" se caracteriza por la expresión de genes encontrados usualmente en células basales/mioepiteliales normales del seno, incluyendo citoqueratinas 5/6 y 14, p-caderina y *EGFR*, son frecuentemente triple negativos por inmunohistoquímica. Histológicamente los tumores "basal-like" tienen alto grado histológico y altas tasas de proliferación (Eroles, Bosch, Perez-Fidalgo, & Lluch, 2012). Recientemente tres subtipos moleculares adicionales al grupo ER- negativo fueron descritos. El subtipo molecular apocrino (molecular apocrine) que se caracteriza por la expresión del receptor de andrógenos (AR) (Farmer *et al.*, 2005). El subtipo interferón el cual se caracteriza por alta expresión de genes reguladores de interferón, incluyendo STAT1 (Hu *et al.*, 2006) y el subgrupo de claudina-baja (claudin-low), el cual agrupa tumores que tienen características transcriptómicas sugestivas de un fenotipo mesénquimal y de células madre mamarias (Hennessy *et al.*, 2009). Algunas de las

características principales que definen cada uno de estos subtipos son mostradas en la tabla1-

Tabla 1-0-1: Características de subtipos moleculares de cáncer de seno. Fuente:(Michelle Alizart, 2012). Molecular classification of breast carcinoma

Receptor de estrógenos	Subtipo molecular	Expresión HER2	Ki-67 IHC	Grado histológico	Características adicionales	Tipos histológicos comunes
RE- positivo	Luminal A	HER2-	Bajo	1 o 2	Citoqueratina Luminal +; e- Caderina +/-	IDC-NST; Lobular; tubular; mucinoso; neuroendocrino; Cribiforme; micropapilar.
	Luminal B	HER2-/+	Alto	2 o 3	Citoqueratina Luminal +; mutaciones en p53	IDC-NST ; Micropapilar
RE-negativo	HER2	HER2 +	Alto	2 o 3	Mutaciones p53	IDC-NST ;Apocrino; micropapilar; lobular pleomórfico
	Basal	HER2 -	Alto	3	Citoqueratina basal +; mutaciones en p53; perdida de reparación del ADN; EGFR+/-; KIT+/-	IDC- NST;medular;metaplásico; quístico adenoide; secretorio
	Apocrino	HER2-	Alto	2 o 3	Receptor de Andrógeno +	IDC-NST; Apocrino; pleomórfico lobular
	Claudina	HER2-	Alto	3	CSCL; EMT-L; E-Caderina :baja	IDC-NST; Medular; Metaplásico
	Relacionado a Interferón	HER2-	Alto	3	STAT1	IDC-NST; Medular

1.1.2 Factores de riesgo asociados a cáncer de seno

Los factores de riesgo se definen como una variable que pueda modificar la probabilidad de contraer una enfermedad. La etiología del cáncer de seno involucra tanto factores genéticos como no genéticos. Para cáncer de seno se han definido los siguientes factores de riesgo no genéticos:

Factores reproductivos

Factores hormonales y reproductivos incluyendo la edad de la menarquía, paridad, edad del primer embarazo, amamantar, uso de anticonceptivos orales y terapia de reemplazo hormonal han sido asociados con riesgo a cáncer de seno.

La menarquía a edad temprana se considera como un factor de riesgo importante (< 11 años)(McPherson, Steel, & Dixon, 2000), mientras que la edad de la menarquía a los 15 años o a una edad mayor se asocia con un riesgo reducido para cáncer de seno (riesgo relativo (RR), 0,87[IC, 0.78 a 0.97]) (Nelson *et al.*, 2012).

Mujeres nulíparas tienen un riesgo significativamente alto para cáncer de seno comparado con mujeres que han tenido algún embarazo. De igual manera varios estudios han reportado que el riesgo para cáncer de seno disminuye significativamente para mujeres con 3 o más embarazos comparado con mujeres nulíparas (Nelson *et al.*, 2012). Además en un reciente metaanálisis se asoció la paridad con un 25% en la reducción del riesgo de desarrollar cáncer de seno de subtipo luminal A (OR 0.75; 95% CI, 0.70–0.81; $p < 0.0001$)(Lambertini *et al.*, 2016).

La edad tardía del primer embarazo incrementa la incidencia del cáncer de seno. La edad avanzada en el primer nacimiento en el metaanálisis llevado a cabo por Lambertini (2016) se asoció con un mayor riesgo de desarrollar cáncer de seno de subtipo luminal (OR 1,15; IC del 95 %, 1,00 a 1,32; $p = 0,05$)(Lambertini *et al.*, 2016). El grupo de mayor riesgo lo constituyen aquellas mujeres que han tenido su primer hijo después de los 35-40 años (McPherson *et al.*, 2000).

Amamantar se ha asociado con un riesgo reducido para cáncer de seno, particularmente cuando se realiza por al menos 12 meses o por periodos más prolongados (Lee *et al.*, 2003), además se reduce el riesgo de desarrollar cáncer de seno de subtipo luminal (OR 1.15; 95% CI, 1.00–1.32; $p = 0.05$) y de subtipo triple negativo (OR 0.77; 95% CI, 0.66–0.88; $p = 0.003$) (Lambertini *et al.*, 2016).

Se ha visto un aumento del riesgo en mujeres en las cuales se retrasa la menopausia, este aumento del riesgo se ha estimado en un 3% por cada año de retraso. Por el contrario, en pacientes que han sido ooforectomizadas antes de los 45 años, se ha visto una disminución del riesgo de aproximadamente un 20% (McPherson *et al.*, 2000).

La exposición y duración a hormonas endógenas (tiempo entre la edad de la menarquía y edad de la menopausia) y exógenas (anticonceptivos orales y terapia de reemplazo hormonal) se ha considerado como un factor de riesgo para cáncer de seno. El uso de anticonceptivos orales puede producir un riesgo leve (RR 1.24) (Washbrook, 2006), donde el mayor riesgo se ha observado entre mujeres con un consumo prolongado de estos (>4 años). La duración del uso, la edad del primer uso o la dosis del anticonceptivo oral tienen un pequeño efecto sobre el riesgo de cáncer de seno (Hulka & Moorman, 2008). El uso de terapia hormonal de reemplazo está asociado con un modesto incremento en el riesgo de cáncer de seno. El uso a largo plazo (5 años o más) está asociado con un 30% a 50% de incremento del riesgo a cáncer de seno (Hulka & Moorman, 2008).

Estilo de vida

Algunos factores asociados al estilo de vida se han vinculado con el riesgo a cáncer de seno. Dentro de estos se encuentran el consumo de alcohol el cual ha sido relacionado con un aumento del riesgo de cáncer de mama, algunos estudios han mostrado un incremento del riesgo hasta del 30% entre consumidores, y por el aumento diario de 10 g de alcohol, el riesgo incrementa en aproximadamente el 7% (Washbrook, 2006). El consumo de alcohol también se ha asociado con el riesgo de desarrollar alguno de los subtipos de cáncer de seno Luminal A, luminal B y cáncer de seno triple negativo (OR 2.16, 95% CI 1.55–3.02; OR 1.98, 95 % CI 1.11–3.53, y OR 1.93, 95 % CI 1.07–3.47) respectivamente (Gago-Dominguez *et al.*, 2016).

El consumo de cigarrillo en un reciente estudio prospectivo encontró un incremento del riesgo de cáncer de seno entre fumadores activos comparado con no fumadores, entre mujeres quienes iniciaban el consumo de cigarrillo a edad temprana, entre mujeres quienes iniciaban el consumo al menos 5 años antes de su primer embarazo y en mujeres quienes tenían una larga duración o una gran intensidad en el consumo de cigarrillo (Reynolds *et al.*, 2004).

Parece existir una relación entre índice de masa corporal (IMC) y riesgo de cáncer de mama para las mujeres postmenopáusicas, mientras que para las mujeres premenopáusicas parece existir la relación contraria (Lahmann *et al.*, 2004). El RR para mujeres postmenopáusicas con un IMC mayor a 21 se ha estimado en 1.26, mientras que el RR para las mujeres premenopáusicas con un IMC mayor de 31 es 0.55. Incluso se ha visto que aquellas mujeres que ganan peso cuando ya son adultas, aumentan su riesgo de desarrollar un cáncer de mama en el periodo postmenopáusico (Gomez, 2006).

Lesiones benignas en el seno como factores de riesgo

Algunas lesiones benignas en el seno son reconocidos factores de riesgo para presentar cáncer de seno y por lo tanto son consideradas como lesiones precursoras. Mujeres con hiperplasia epitelial atípica tienen cuatro a cinco veces más riesgo de desarrollar cáncer de seno que las que no tiene ningún cambio proliferativo en sus senos. Mujeres con este tipo de cambios mamarios además de historia familiar de cáncer de seno en un pariente de primer grado tiene un mayor incremento en el riesgo. Además, aquellas mujeres que presentan en sus senos quistes palpables, fibroadenomas, y papilomas ductales, tienen un riesgo ligeramente mayor que aquellas mujeres que no han presentado ninguno de estos cambios (McPherson *et al.*, 2000).

Historia familiar

Una historia familiar de cáncer de seno es un factor de riesgo bien establecido para esta enfermedad. Una mujer con un pariente en primer grado afectado por cáncer de seno tiene aproximadamente dos veces más riesgo de desarrollar la enfermedad (Thompson & Easton, 2004). Teniendo múltiples parientes afectados, particularmente con inicio temprano de la enfermedad o cáncer de seno bilateral el riesgo se incrementa aún más (McPherson *et al.*, 2000; Thompson & Easton, 2004). Mujeres con uno, dos, tres o más

parientes en primer grado con cáncer de seno tienen un RR de 1,80, 2,93 y 3,90 respectivamente, comparado con mujeres sin parientes afectados (Beral *et al.*, 2001). Cada mujer tiene su propio perfil de riesgo el cual puede ser alto o bajo dependiendo de la presencia o ausencia de dichos factores de riesgo. Por ello se debe valorar e individualizar, lo más exactamente posible, el perfil de riesgo de una paciente a fin de poder asegurarle la mejor relación costo-beneficio de las alternativas de vigilancia y tratamiento disponibles, Además se debe tener en cuenta la etnicidad de las pacientes ya que diferentes factores de riesgo pueden variar con respecto a este factor epidemiológico. Los factores de riesgo descritos anteriormente junto con los valores de riesgo relativo para cada caso son mostrados en la tabla 1-2.

Tabla 1-0-2: Factores de riesgo asociados con cancer de seno. Fuente: The lancet, Breast cancer. Veronesi, *et al.*, 2006

Variable	Grupos de alto riesgo	RR*
Edad	Población de mayor edad	10
Edad de la menarquia	Antes de los 11 años	3
Edad primer embarazo	Primer hijo después de los 40 años	3
Edad de la menopausia	Después de los 54 años	2
Historia Familiar	Primer grado	2
Exposición a radiación ionizante	Exposición anormal en niños de 10 años y más Irradiación de la pared torácica	3
Enfermedad benigna previa	Hiperplasia con atipias	4 a 5
Índice de masa corporal después de la menopausia	Alto índice de masa corporal	2
Uso de terapia hormonal de reemplazo	Uso por más de 5 años	1,6

1.1.3 Genética del cáncer de seno

La relación entre la ocurrencia de cáncer y la existencia de alteraciones genéticas está bien establecida, en la mayoría de cánceres estas alteraciones son principalmente somáticas. En un 4-10% de los casos de cáncer de seno, una mutación en la línea germinal, predispone al desarrollo de la enfermedad.

– genes de alta penetrancia

Los principales genes de alta penetrancia en los cuales mutaciones incrementan la susceptibilidad a cáncer de seno son el gen de susceptibilidad a cáncer de seno 1 (*BRCA1*: breast cáncer 1)(Hall *et al.*, 1990) y el gen de susceptibilidad a cáncer de seno 2 (*BRCA2*: breast cáncer 2) (Wooster *et al.*, 1995). Una mutación en *BRCA1* o *BRCA2* confiere un riesgo incrementado de cáncer de seno y otros cánceres. Adicionalmente a *BRCA1/ BRCA2* mutaciones en otros genes poco comunes, pero altamente penetrante incluyen *PTEN* (FitzGerald *et al.*, 1998; Garber *et al.*, 1991; Tan *et al.*, 2012), *TP53* (Birch *et al.*, 2001; Garber *et al.*, 1991), *CDH1* (Pharoah, Guilford, & Caldas, 2001), y *STK11* (Boardman *et al.*, 1998; Lim *et al.*, 2004), cada uno está asociado con un síndrome clínico distinto.

En conjunto con *BRCA1* y *BRCA2*, se estima que los genes de alta penetrancia conocidos representan no más del 25% de los casos sobre la base de estudios previos y modelización matemática (Antoniou & Easton, 2006; Walsh *et al.*, 2006).

– Genes de penetrancia intermedia

Se han identificado otros genes implicados en el proceso de reparación del ADN que interactúan con *BRCA1*, *BRCA2*, y / o las vías BRCA, y confieren un aumento de dos veces en el riesgo de cáncer de mama. Estos incluyen *CHEK2* [(Meijers-Heijboer *et al.*, 2002), *BRIP1* (*BACH1*) (Seal *et al.*, 2006), *ATM* (Renwick *et al.*, 2006) y *PALB2* (Rahman *et al.*, 2007). Mutaciones germinales en estos genes se presentan en aproximadamente un 5% de los casos de cáncer de seno hereditario. Las mutaciones en estos genes de penetrancia moderada pueden ser mediados por factores ambientales, disminuyendo su impacto genético (Gupta, Jaworska-Bieniek, Lubinski, & Jakubowska, 2013)

– Genes de baja penetrancia

Recientemente se han publicado estudios que reportan diferentes genes de baja penetrancia que están asociados con riesgo de cáncer de seno (Ripperger, Gadzicki, Meindl, & Schlegelberger, 2009).

En cáncer de seno, la mayoría de los casos donde no hay una mutación identificada en genes de alta penetrancia o de penetrancia intermedia (mencionados anteriormente), el riesgo podría ser explicado por un modelo poligénico, el cual involucra una combinación de muchas variantes individuales asociadas con riesgo mínimo, que se conocen como polimorfismos de baja penetrancia (Turnbull & Rahman, 2008). Actualmente estudios basados en poblaciones han aumentado considerablemente, debido en gran parte a la accesibilidad a bases de datos genómicas, como por ejemplo el proyecto internacional HapMap, el cual ha favorecido sustancialmente los estudios de asociación indirectos al tener como objetivo determinar los patrones comunes de variación en la secuencia de ADN en el genoma humano ("The International HapMap Project," 2003b; Vega *et al.*, 2009) así como el proyecto de los 1000 genomas el cual tiene como objetivo principal describir la mayoría de la variación genética que ocurre en una población con una frecuencia mayor al 1% ("A map of human genome variation from population-scale sequencing," 2010)

Algunos de los estudios que han mostrado asociación de SNPs con el riesgo de cáncer de seno se han basado en la metodología del "gen candidato" y estudios de GWAS (Mavaddat, Antoniou, Easton, & Garcia-Closas, 2010). La mayoría de loci identificados a través de la aproximación del gen candidato, no han sido confirmados. La más convincente asociación ha sido para una variante codificante (D302H) en el gen *CASP8* (caspasa 8) que se asoció con reducción del riesgo en cáncer de seno (Cox *et al.*, 2007). Otra variante común con evidencia más débil para una asociación con riesgo es una variante codificante (L10P) en el gen *TGFB1* (factor de crecimiento transformante- β), donde la asociación fue restringida a tumores con receptor de progesterona negativo (Cox *et al.*, 2007). En contraste a los estudios de genes candidatos, los estudios GWAS han guiado al descubrimiento de marcadores genéticos localizados en 12 loci de susceptibilidad, algunas de estas regiones contienen o son cercanas a genes conocidos entre estos se encuentran los genes *FGFR2* (Fibroblast growth factor receptor 2), el SNP

rs3803662 en 16q12 está localizado a 8Kb de *TOX3* (TOX high mobility group box family member 3)(Easton *et al.*, 2007), en un estudio anterior se sugirió que la expresión *TOX3/TNRC9* es predictiva de metástasis de cáncer de seno a hueso (Smid *et al.*, 2006). La variante en 5q11 se encuentra en un bloque de desequilibrio de ligamiento conteniendo el gen *MAP3K1* (Mitogen-activated protein kinase kinase kinase 1), este gen codifica un componente de la vía de señalización MAPK involucrada en señalización celular, proliferación y apoptosis (Easton *et al.*, 2007). Otros “tag SNP” asociados con cáncer de seno se encuentran en *LSP1* (Lymphocyte-specific protein 1)(Easton *et al.*, 2007) y *RAD51L1* (DNA repair protein RAD51 homolog 2)(Thomas *et al.*, 2009), mientras otros están localizados en regiones no génicas lejos del gen más cercano, por ejemplo sobre 8q24 se encontró asociación para un SNP dentro de un bloque de 110Kb de desequilibrio de ligamiento, ubicado en esta región donde no hay genes conocidos, el gen cercano al locus de predisposición a cáncer de seno es el protooncogen *MYC* (Easton *et al.*, 2007; Garcia-Closas *et al.*, 2008). Se podría especular que la susceptibilidad ocurre a través de algún mecanismo de activación de este gen. En un estudio realizado en Islandia se encontró evidencia para asociación con cáncer de seno en una variante sobre 2q35, sin embargo, otros estudios no han encontrado esta asociación, excepto un estudio realizado en el Reino unido donde se replicó la asociación en casos de cáncer de seno familiar (Turnbull & Rahman, 2008). En el estudio realizado por Michailidou (2013) donde se incluyó únicamente población de ancestría europea, se encontraron 41 SNPs para los cuales la asociación con riesgo a cáncer de seno fue estadísticamente significativa $p \leq 5 \times 10^{-8}$. (Michailidou *et al.*, 2013) hasta el momento en esta población se han descrito 79 loci asociados con cáncer de seno.

1.2 Ubiquitinación y su rol en cáncer de seno

La homeostasis de proteínas en células de mamíferos está estrechamente controlada por la síntesis de proteínas y la degradación. El recambio rápido de proteínas es una característica típica de muchas proteínas importantes que regulan el ciclo celular, apoptosis, y la transcripción, tales como p53, p27kip1, y Myc. La degradación selectiva de estas proteínas a través del sistema ubiquitin-proteasoma juega un papel esencial en el crecimiento normal de las células y la diferenciación, mientras que la acumulación

anormal o degradación hiperactiva de estas proteínas reguladoras están asociadas con carcinogénesis (C. Chen, Seth, & Aplin, 2006).

1.2.1 Ubiquitinación y degradación proteasomal

La vía de degradación de proteínas ubiquitin-proteasoma es un mecanismo de control celular estricto de moléculas regulatorias de vida media. Esta vía de degradación proteolítica es importante para mantener la fisiología normal de la célula. El sistema se compone de ubiquitina, un complejo de tres enzimas de ubiquitinación, las proteínas intracelulares blancas de Ubiquitinación, y el proteasoma (Ciechanover, 2005).

Las proteínas degradadas a través del proteasoma son primero marcadas por una cadena de poliubiquitina en lisina 48 (K48), a través de la acción de tres clases de enzimas: la enzima activadora de ubiquitin (E1), enzima conjugadora de ubiquitin (E2), y la enzima ubiquitin ligasa (E3).

La ubiquitina es una proteína pequeña conservada con 76 aminoácidos. E1 se une y activa la ubiquitina por medio de la adenilación de ubiquitina, luego la molécula ubiquitina es transferida a una E2. E2 trabaja en conjunto con la ubiquitina ligasa E3, que es responsable de conferir especificidad de sustrato en la reacción. E3 media la transferencia de ubiquitina a una lisina interna de la proteína diana. Tras la monoubiquitinación, la segunda molécula de ubiquitina puede conjugarse con la anterior a través de un enlace iso-peptídico entre Gly76 de una nueva molécula de ubiquitina y los grupos ϵ -NH₂ de uno de las siete lisinas (K6, K11, K27, K29, K33, K48 y K63) de la ubiquitina conjugada anteriormente. Las cadenas de poliubiquitina predominantes en las células son a través de los puntos de ramificación K48 y K63. Como se mencionó anteriormente, las proteínas etiquetadas con una cadena K48 poliubiquitina son capaces de ser degradados rápidamente por un gran complejo de proteasas citosólicas llamado el proteasoma 26S. Por otra parte, las proteínas con una cadena de poliubiquitina K63 no serán objeto de degradación, sino más bien se ha demostrado que desempeñan un papel importante en la transducción de la señalización del factor nuclear kappa-Beta (NF- κ B) (Z. J. Chen, 2005).

1.2.2 Ubiquitin ligasas en cáncer de seno

E3 lleva a cabo el paso clave que limita la velocidad en la proteólisis mediada por ubiquitina. En células de mamíferos, abundantes ubiquitin ligasas garantizan el reconocimiento específico de sustrato. Las E3 ubiquitin ligasas están divididas en dos clases principales: La primer clase corresponde a la Proteína asociada a E6 (E6-AP) y un dominio C terminal HECT (homologous to E6-AP carboxy terminus) y la segunda clase corresponde a las RING- E3 ligasas (really interesting new gene). Estas se diferencian en la forma como catalizan la reacción de transferencia de ubiquitina al sustrato (Marblestone, LaRocque, Mattern, & Leach, 2012). Las enzimas E3 ligasas de la familia RING catalizan la transferencia directa de ubiquitina desde la enzima E2 al sustrato (Deshaies & Joazeiro, 2009). En contraste las enzimas de clase HECT ubiquitinan el sustrato en dos pasos secuenciales donde primero la ubiquitina es transferida desde la E2 al sitio activo cisteína de la E3 y entonces es transferida desde la E3 al sustrato. (Berndsen & Wolberger, 2014; Huibregtse, Scheffner, Beaudenon, & Howley, 1995).

Otra clase de E3 ligasas es RBR (RING-between RING- RING), estas representan una familia adicional de E3 que combina características de las familias HECT y RING ligasas. Para la ubiquitinación de los sustratos reclutan la molécula de ubiquitina conjugada con la E2 por un dominio RING N-terminal y entonces transfieren la ubiquitina dentro de un residuo cisteína catalítico c-terminal tipo HECT de la E3 antes de realizar la transferencia final sobre el sustrato (Spratt, Walden, & Shaw, 2014).

Las enzimas E3s ligasas pueden actuar como oncogenes o genes supresores de tumor, la frecuente desregulación de E3 ha sido mostrada en carcinogénesis. La inhibición de las E3 sobreexpresadas en cáncer de seno guían a la detención del crecimiento o apoptosis. En contraste las E3 ligasas que funcionan como supresores tumorales frecuentemente sufren mutaciones inactivadoras o baja expresión. Varias E3 como por ejemplo CHFR, SIAH1, y CHIP, pueden jugar un rol significativo en la regulación de tumorigenesis mamaria. Además de mutaciones, perdida del número de copia y alteraciones epigenéticas también contribuyen a la inactivación de estos supresores de tumor. Algunas E3 asociadas con cáncer de seno son mostradas en la tabla 1-3 y 1- 4. (Kitagawa, Kotake, & Kitagawa, 2009)

Tabla 1-0-3: E3 ligasas asociadas con cáncer de seno como oncogenes (C. Chen *et al.*, 2006)

E3	Gen locus	Función	Alteración en cáncer de seno
ARF-BP1	Xp11.22	Promover el crecimiento celular y la antiapoptosis	Sobreexpresión
WWP1	8q21	Transcripción y transducción de señales	Amplificación y sobreexpresión
Mdm2	12q14.3-q15	Antiapoptosis	Amplificación, mutación, sobreexpresión, y splicing aberrante del ARN
RNF11	1p31-32	Transducción de señales	Sobreexpresión
EFP	17q23.1	Ciclo celular	Sobreexpresión
XIAP	Xq25	Antiapoptosis	
BCA2	1q21.1		Amplificación y Sobreexpresión
Skp2	5p13	Ciclo celular, transcripción	Amplificación y Sobreexpresión
β -TrcP	10q24.32	Ciclo celular, transcripción	Amplificación y Sobreexpresión
Cul4A	13q34	Reparación del ADN Y antiapoptosis	Amplificación y Sobreexpresión

Tabla 1-4 : E3 ligasas asociadas con cáncer de seno como supresores tumorales (C. Chen et al., 2006)

E3	Locus	Función	Blancos principales	Alteración en cáncer de seno
BRCA1	17q21	Reparación del ADN, estabilidad genómica	H2A, Rpb1, γ -tubulin	Mutación
BARD1	2q34-q35	Reparación del ADN, apoptosis	H2A, Rpb1, γ -tubulin	Mutación, expresión aberrante
SIAH1	16q12-13	Ciclo celular, Transcripción y apoptosis	β -Catenin, OBF-1, CtIP, Kid, Numb	
Parkin	6q25		α SP22, p38/JTV1, Pael-R	Delección y pérdida de la expresión
CHFR	12q24.33	Ciclo celular	Plk1, Aurora A	Pérdida de la expresión por metilación del promotor
Cul5	11q22-23			Pérdida de la expresión
Fbw7	4q31	Ciclo celular	Cyclin E, Myc, Notch	Mutación y pérdida de la expresión
CHIP	16q13.3	Señalización	ErbB2, RE α , AR, Smads	

1.3 Dear1: Clasificación, Estructura Y Asociación Con Cáncer De Seno

1.3.1 Clasificación Familia de proteínas RBCC/TRIM

Las proteínas TRIM/RBCC están definidas por la presencia de un motivo tripartito conformado de un dominio RING (really interesting new gene), uno o 2 motivos “B-box” y un motivo estructural “coiled-coil”. El dominio RING está definido por una disposición regular de residuos cisteína e histidina que coordinan la unión a 2 átomos de zinc (Freemont, 1993). Funcionalmente, el dominio RING está involucrado en mediar la transferencia de ubiquitin tanto a sustratos heterólogos como a las mismas proteínas RING. En la mayoría de proteínas TRIM el dominio RING está asociado con una función como E3 ubiquitin ligasa. El dominio “B-box” es otro motivo de unión a zinc que puede ser del tipo “B-box1” y “B-box2”, los cuales tienen un similar pero diferente patrón de residuos cisteína e histidina (Meroni & Diez-Roux, 2005). La región “coiled-coil” está involucrada principalmente en homo interacciones, promoviendo la formación de complejos de alto peso molecular y en la definición de compartimentos subcelulares (Reymond *et al.*, 2001).

Las proteínas TRIM/RBCC están involucradas en un amplio rango de procesos biológicos que incluyen procesos celulares fundamentales como apoptosis, transcripción, diferenciación y regulación de la progresión del ciclo celular (Uchil, Quinlan, Chan, Luna, & Mothes, 2008). Varios miembros de la familia TRIM están implicados en diferentes condiciones patológicas incluyendo enfermedades genéticas mendelianas (fiebre mediterránea familiar), procesos oncogénicos, donde funcionan como reguladores transcripcionales, reguladores del ciclo celular e intervienen en procesos de infección Viral (Meroni & Diez-Roux, 2005).

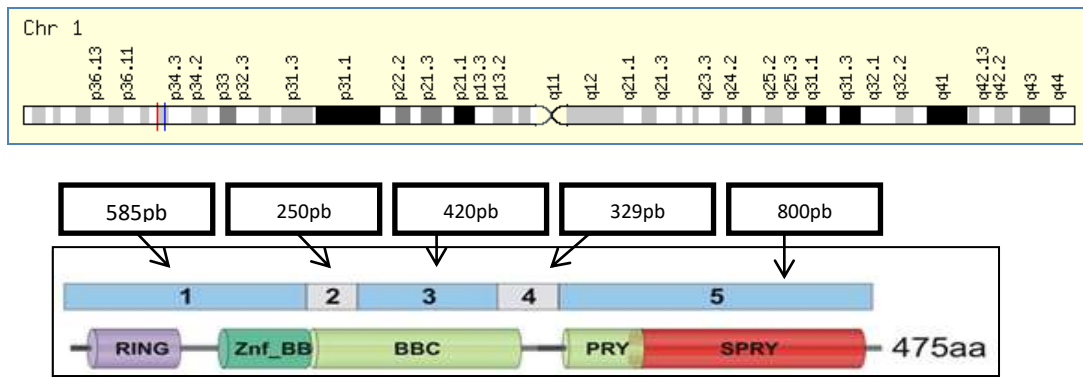
Algunas proteínas TRIM están involucradas en el desarrollo de cáncer de seno. *TRIM24*, también conocido como *TIF1 α* se ha asociado con un pobre pronóstico cuando está sobreexpresada. *TRIM27* también conocido como *RFP* puede promover o suprimir el desarrollo de tumores y su expresión se correlaciona con expresión de la proteína ERBB2 y la amplificación del gen *ERBB2* en cáncer de seno. Por último, el gen *TRIM25* tiene un elemento de respuesta a estrógenos, por lo cual su expresión es estimulada por los mismos. *TRIM25* presenta altos niveles de expresión en órganos reproductivos

femeninos, que incluye glándula mamaria y útero, además se ha observado el mismo comportamiento en cáncer de seno y ovario, donde se ha correlacionado con pobre pronóstico (Hatakeyama, 2011).

1.3.2 Descripción del gen *DEAR1*

El gen *DEAR1* también conocido como *TRIM62*, está localizado en el cromosoma 1 (1p35.1), en la posición 33611003-33647671. Está constituido por 5 exones que codifican para una proteína de 475 aminoácidos. *DEAR1* tiene cinco dominios, los cuales son característicos de la subfamilia de proteínas RBCC/TRIM de la familia RING finger” con los dominios “PRY” y “SPRY” presentes dentro de la región carboxilo terminal (Lott *et al.*, 2009b) (Figura 3). El gen *TRIM62* es altamente conservado, encontrándose homología en chimpancé, perro, vaca, ratón, rata, pollo y el pez cebra. Adicionalmente *TRIM62* tiene tres parálogos, los genes *TRIM35*, *TRIM69*, and *TRIM74*.

Figura 1-4: Localización cromosómica de *DEAR1* y representación gráfica de los exones y dominios de la proteína (RING:Really interesting new gene;Znf_BB: B-Box-type zinc finger, BBC: B-Box dominio C-terminal; dominio PRY; dominio SPRY:)



1.3.3 *DEAR1* y su relación con Cáncer de seno

La iniciación y progresión de cáncer de seno involucra una interrupción de vías celulares que intervienen en procesos de proliferación, diferenciación, muerte celular programada y daño de vías de señalización extracelular que están involucradas en diferenciación y

arquitectura del tejido mamario (Fish & Molitoris, 1994). Simultáneamente con la iniciación de tumorigénesis, las células epiteliales en la glándula mamaria pierden la habilidad de mantener un tejido polarizado lo cual es crítico para la función de las células epiteliales mamarias (Muthuswamy, 2009).

DEAR1 (Ductal Epithelium-Associated RING chromosome 1, TRIM 62), fue identificado mediante la técnica SSH (suppression subtractive hybridization) en una región del cromosoma 1, que a menudo muestra LOH (pérdida de la heterocigosidad) en cánceres de seno familiar (Lott et al., 2009b). Este gen codifica para un miembro de la subfamilia TRIM (tripartite motif) de proteínas "RING finger".

En el seno, *DEAR1* está expresado tanto en epitelio ductal como en el glandular y su expresión se encuentra reducida o pérdida completamente en varios carcinomas *in situ* y en cáncer de seno avanzado. Varias mutaciones sin sentido y deleciones homocigotas en el gen *DEAR1* fueron identificadas en líneas celulares de cáncer de seno y en tumores de seno. Adicionalmente al incorporar el gen *DEAR1* de tipo salvaje en líneas celulares que tenían una mutación sin sentido asociada a la alteración de la estructura acinar, se logra restaurar la estructura acinar normal. Disminución estable de la expresión de *DEAR1* en células epiteliales mamarias humanas inmortalizadas recapitulan el crecimiento en cultivos 3D de líneas celulares de cáncer de seno conteniendo *DEAR1* mutado con interrupción de la arquitectura del tejido, pérdida de la polaridad y formación de lumen. Estos hallazgos indican que *DEAR1* es un regulador dominante de morfogénesis acinar. Además al realizar análisis por IHC en tejido tumoral mamario la expresión de *DEAR1* estaba pérdida frecuentemente en pacientes con cáncer de seno de inicio temprano y la pérdida de la expresión de esta proteína se correlacionó con una reducción de la supervivencia libre de recurrencia local, fuerte historia familiar de cáncer de seno y un fenotipo triple negativo (es decir, negativo para la expresión de receptor de estrógenos (ER), receptor de progesterona (PR) y factor de crecimiento epidérmico receptor 1 (HER2)). La identificación de mutaciones somáticas y en la línea germinal sugiere que *DEAR1* puede jugar un rol en el desarrollo de cáncer de seno tanto esporádico como hereditario (Lott et al., 2009b).

Además del rol evidente como regulador dominante de morfogénesis acinar, el gen *DEAR1* se ha asociado como regulador de la estabilidad de *p27* un inhibidor de kinasas dependiente de ciclinas (iCDKs). *p27* es un supresor de tumor que tiene como función regular el ciclo celular al inhibir la entrada en la fase G1 del ciclo celular (Roberts, Eide, & Faltermier, 2009). Típicamente, las células tumorales exhiben una disminución en la expresión de *p27* en el núcleo, aumentando la cantidad de *p27* en el citoplasma. Mientras *p27* nuclear se comporta como supresor tumoral al inhibir la proliferación celular, nueva evidencia sugiere que *p27* citoplasmático tiene una acción oncogénica: este interactúa con *RhoA* (Ras homolog gene family, member A) para promover la migración celular y la movilidad, por lo cual *p27* citoplasmático podría ser un indicador de invasividad tumoral y potencial metastásico. Con el objetivo de comprender el mecanismo por el cual *p27* no se encuentra localizado correctamente en cánceres de seno y de esta manera revelar algunas vías de transformación oncogénica y proveer nuevos marcadores para predecir la respuesta clínica a terapias específicas. El grupo de investigación del Dr. Roberts y colaboradores desarrollo anticuerpos para la detección de *p27* citoplasmático y para la identificación de cánceres de seno positivos para HER2 que exhibían cantidades incrementadas de *p27* citoplasmático, encontraron una E3 ubiquitin ligasa, TRIM62 que regula la estabilidad de *p27* y se encontró sobreexpresada en cánceres de seno HER2 positivo. La disminución en la expresión de TRIM62 por RNA de interferencia en líneas celulares de cáncer HER2+ no solo induce el bloqueo del ciclo celular, sino que también causa la relocalización completa de *p27* al núcleo. Además de esta evidencia también investigaron el rol de TRIM62 como modulador de la respuesta de células de cáncer de seno a la terapia anti-HER2, teniendo como antecedente que la inhibición farmacológica de HER2 muestra un incremento en los niveles de la proteína *p27* en el núcleo. Estos resultados mostraron que la disminución en la expresión de TRIM62 incrementa la sensibilidad celular al tratamiento farmacológico anti-HER2. Lo cual sugiere que TRIM62 podría ser un biomarcador potencial para predecir respuesta biológica de células de cáncer de seno a terapias anti-HER2 (Roberts *et al.*, 2009).

La pérdida de la función de *DEAR1* en presencia de *TGFβ* resulta en insuficiencia de morfogénesis acinar, regulación positiva de los marcadores de EMT, resistencia a anoikis, migración e invasión. Además, *DEAR1* bloquea la señalización de *TGFβ-SMAD3* resultando en una disminución de *SMAD3* nuclear fosforilada mediante la unión a y la

promoción de la ubiquitinación de SMAD3, el principal efector de EMT inducida por *TGFβ* (N.Chen *et al.*, 2013).

1.4 Estudios de asociación en cáncer de seno

1.4.1 Polimorfismos de único nucleótido (SNPs) y su asociación con cáncer de seno

Las secuencias completas de dos genomas humanos son por término medio un 99.9% idénticas (www.wellcome.ac.uk/genome). El 0.1% restante varía entre cada individuo, siendo las variaciones más comunes las que se conocen como SNPs o polimorfismos de una sola base. En el primer análisis del genoma humano se detectaron aproximadamente 1.42 millones de SNPs (Lander *et al.*, 2001; Venter *et al.*, 2001). Se ha estimado que el genoma humano contiene al menos 11 millones de SNPs con un promedio de 7 millones de estos presentándose con MAF de 5% y los restantes teniendo MAF entre 1% y 5% (Frazer *et al.*, 2009). Estos pueden presentarse en la porción codificante de los genes (exones), interrumpiendo secuencias (intrones) o en regiones entre dos genes (regiones intergénicas) (Harley and Narod, 2009).

Los SNPs aparecen en cualquier parte del genoma, pero el análisis de la distribución de los polimorfismos a lo largo del genoma humano muestra variaciones significativas en la densidad de polimorfismos y en la distribución de las frecuencias alélicas. Chakravarti (Chakravarti, 1999) mostró una diferencia entre la densidad de SNPs en regiones exónicas y en regiones intergénicas o intrónicas, ya que aparecen en intervalos medios de 1.2Kb en las primeras y en intervalos de 0.9Kb en las segundas.

Atendiendo a sus posibles efectos funcionales, las variaciones se pueden subdividir en distintas clases. Si un SNP está localizado en una región codificante, la variación puede resultar en un cambio de aminoácido y alterar la secuencia de la proteína (SNP no - sinónimo). El SNP también puede ser funcional, aunque no produzca ningún cambio de aminoácido (SNP sinónimo), ya que pueden alterar la estabilidad del mRNA (Capon *et al.*, 2006). Los SNPs en las regiones intergénicas o intrones también pueden ser funcionales si alteran los sitios de empalme (splicing), o si interrumpen o crean nuevos sitios de unión

a factores de transcripción o sitios que actúan como potenciadores o silenciadores de la transcripción. En enfermedades complejas, probablemente lo más común sean los cambios sutiles como las sustituciones sinónimas y SNPs en zonas intergénicas e intrónicas, donde la mutación sólo aumenta la susceptibilidad a la enfermedad, pero no causa directamente la enfermedad.

Actualmente los SNPs están siendo estudiados intensamente con el fin de entender las bases biológicas de enfermedades complejas, además el interés también recae en el hecho que los SNPs podrían ser usados como marcadores para la identificación de genes que predisponen a desordenes multifactoriales como las enfermedades neoplásicas por el uso del desequilibrio de ligamiento (Chan *et al.*, 2012).

Gracias a la terminación del proyecto genoma humano, las investigaciones actuales están dirigidas al estudio de la variación genética individual. Varios métodos han sido utilizados para identificar alelos de susceptibilidad al cáncer de mama, y estos han revelado loci que se pueden clasificar en tres grupos en función de su frecuencia en la población y los riesgos que le confiere; alelos de alta penetrancia, que son muy poco frecuentes en la población (MAF <0,005) que confieren un riesgo mayor de diez veces; alelos raros, moderada penetrancia (MAF: 0,005-0,01), que le confieren un aumento de dos a cuatro veces en el riesgo; y alelos de baja penetrancia (MAF > 0,05) que confieren un aumento de menos de 1,5 veces en el riesgo (Ghoussaini & Pharoah, 2009).

En solo un 25% de los casos de agregación familiar para cáncer de seno se han identificado genes conocidos que se sabe aumentan el riesgo para esta enfermedad (Thompson & Easton, 2004). Sugiriendo que la mayoría del riesgo de cáncer de seno, puede involucrar una combinación de múltiples alelos de susceptibilidad de baja penetrancia, cada uno confiriendo un pequeño efecto (Antoniou, Easton, 2006).

1.4.2 Estudios de Asociación gen candidato

La identificación de variantes causales en regiones cromosómicas por análisis de ligamiento ha sido muy útil para muchos desordenes raros que involucran un único gen. Por el contrario, los estudios de ligamiento han tenido mucho menos éxito en la

localización de variantes genéticas comunes que influyen en enfermedades complejas, ya que cada variante individual contribuye modestamente al riesgo de la enfermedad. Una aproximación complementaria para identificar estos factores de riesgo genéticos es la búsqueda de asociación entre una variante específica y una enfermedad, por comparación de un grupo de individuos afectados (casos) con un grupo de individuos no afectados (controles) (Botstein & Risch, 2003).

Un acercamiento para realizar estudios de asociación se conoce como “aproximación directa” la cual implica hacer una correlación de cada variante putativa causal con la enfermedad. Un enfoque alternativo ha sido propuesto donde un conjunto de variantes podrían servir como marcadores genéticos para detectar la asociación entre una determinada región genómica y la enfermedad, por lo cual la búsqueda de la variante causal podría limitarse a las regiones que muestran asociación con la enfermedad, este se conoce como aproximación indirecta (Kruglyak, Nickerson, 2001) o aproximación “Tag SNP” el cual tiene en cuenta el desequilibrio de ligamiento entre los polimorfismos de un haplotipo y por lo tanto no es necesario evaluar todos los polimorfismos de un único nucleótido (SNPs) de un gen, un número reducido de marcadores SNPs puede ser usado para realizar el seguimiento indirecto de variantes causantes de la enfermedad. En la selección de los tag SNPs se debe tener en cuenta el coeficiente de correlación r^2 que es la medida para LD el cual mide la correlación entre dos loci con alelos en asociación. Un gen en un locus con SNPs en LD tiene una alta probabilidad de ser conservado en poblaciones con recombinación ocurriendo a cada lado de este. Los análisis de LD dentro de SNPs candidatos son una invaluable herramienta para reducir los límites de la región genómica susceptible a estar asociada con la enfermedad; porque estos SNPs en su mayoría son heredados juntos y muestran frecuencias similares en individuos afectados de una población (Patnala, Clements, & Batra, 2013).

Los SNP tienen el potencial de ser usados como marcadores de riesgo de enfermedad, incluso en ausencia de la comprensión funcional del SNP. Mientras que los SNPs pueden no estar asociados con riesgo de cáncer al mismo grado que una mutación altamente penetrante, seguirán contribuyendo de manera significativa a la comprensión de un camino o proceso en la biología del cáncer. Los SNPs Pueden conferir cambios sutiles aún desconocidos en la función génica, afectando procesos celulares fundamentales como transcripción, empalme intrón - exón o plegado de proteínas que, podría tener un

efecto significativo sobre el riesgo o el resultado de la enfermedad (Savage & Chanock, 2006).

1.5 Estrategias metodológicas para el análisis de SNPs

La genotipificación de SNPs es uno de los principales componentes de los estudios de asociación genética y el método apropiado de genotipificación es fundamental para el éxito del estudio.

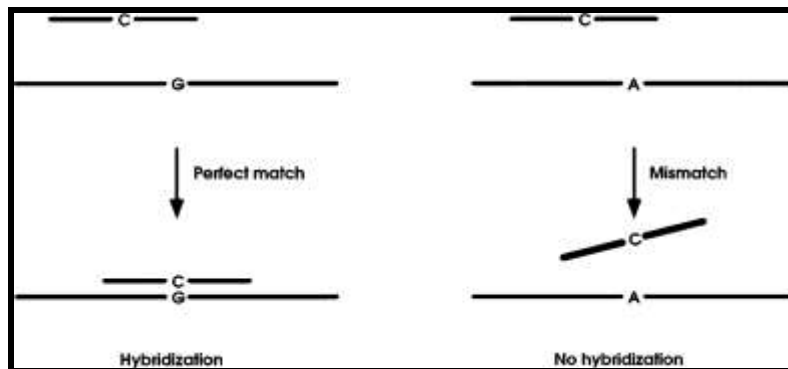
Los métodos más populares para genotipificar SNPs son básicamente: hibridación alelo-específica, incorporación de nucleótidos alelo-específica, ligación de oligonucleótidos alelo-específica y escisión invasiva alelo-específica donde cada una de ellas presenta ventajas y desventajas (Kwok, 2000).

1.5.1 Métodos para el análisis de SNPs

Hibridación alelo-específica

Con la aproximación de hibridación, dos sondas alelo-específicas son diseñadas para hibridar a la secuencia blanco cuando haya un emparejamiento perfecto. Bajo condiciones óptimas la falta de coincidencia en una sola base desestabiliza la hibridación previniendo así que la sonda se una a la secuencia blanco. Como en este proceso no hay enzimas involucradas, el desafío en este tipo de aproximaciones metodológicas para asegurar una discriminación alélica robusta se encuentra en el diseño de la sonda (Kwok, 2001). El esquema del método de hibridación alelo específica es mostrado en la Figura 1-5.

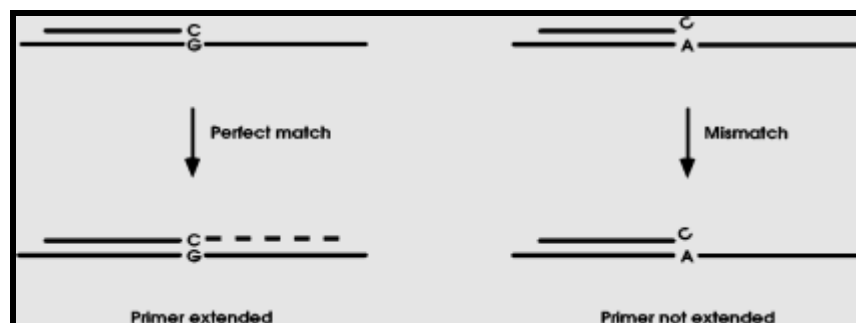
Figura 1-5: Hibridación alelo específica. Fuente: Annu. Rev. Genomics Hum. Genet. 2001. 2:235–58 Methods for genotyping single nucleotide polymorphisms.



Incorporación de nucleótidos alelo-específica

Es un mecanismo de discriminación alélica muy robusto. Es altamente flexible y requiere un número muy pequeño de sondas/primer. Una de las metodologías más usadas es la extensión de sonda alelo-específica. El producto de la PCR conteniendo el sitio polimórfico sirve como plantilla, y el 3' final de la sonda contiene la base alélica. El cebador es extendido sólo si la base en 3' es complementaria al alelo presente en el ADN blanco (Bayés & Gut, 2011). El esquema sobre la metodología de genotipificación por incorporación de nucleótidos alelos-específica es mostrado en la figura 1-6.

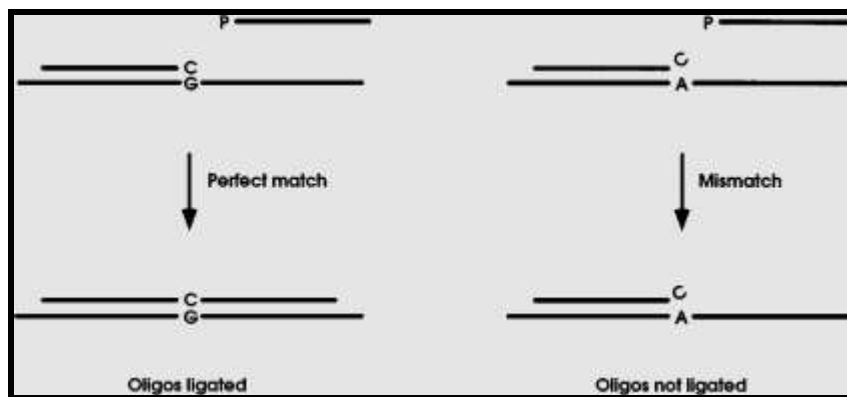
Figura 1-6: Extensión de primer alelo-específica. Fuente: Annu. Rev. Genomics Hum. Genet. 2001. 2:235–58 Methods for genotyping single nucleotide polymorphisms.



Ligación de oligonucleótidos (OLA)

OLA se basa en la especificidad que tiene la DNA ligasa en la reparación de mellas (nicks) en la molécula de ADN. Esta aproximación utiliza dos oligonucleótidos adyacentes el uno al otro, son ligados enzimáticamente por una DNA ligasa cuando las bases siguientes a la posición de ligación son totalmente complementarias a la plantilla de ADN (Jarvius, Nilsson, & Landegren, 2003). Por lo tanto uno puede inferir el alelo presente en el ADN blanco al determinar si la ligación ha ocurrido. Un ejemplo del procedimiento de genotipificación OLA es mostrado en la figura 1-7.

Figura 1-7 Ligación de oligonucleótidos. Fuente: Annu. Rev. Genomics Hum. Genet. 2001. 2:235–58 Methods for genotyping single nucleotide polymorphisms.



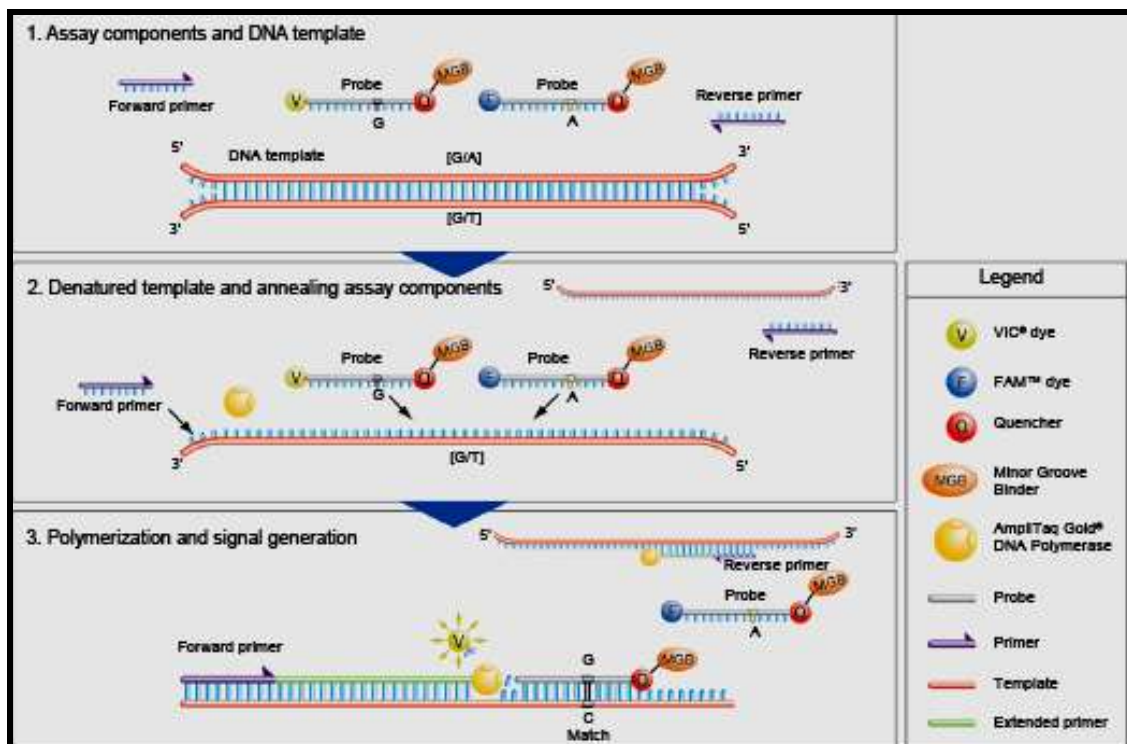
Los procedimientos pueden ser llevados a cabo en reacciones homogéneas las cuales son generalmente robustas y altamente flexibles, o reacciones en fase sólida donde muchos marcadores pueden ser evaluados al mismo tiempo. Los mecanismos de detección incluyen espectrometría de masas, análisis de fluorescencia y detección por quimioluminiscencia (Syvanen, 2001).

1.5.2 Plataformas comerciales para el análisis de SNPs

Las plataformas comerciales para la genotipificación de SNPs, ofrecen desde plataformas de genotipificación de SNPs individuales hasta plataformas para genotipificación del genoma total. Dentro de estas encontramos el ensayo TaqMan (life technologies applied biosystems), se basa en la hibridación alelo-específica acoplada a la actividad 5´nucleasa de la taq polimerasa durante la PCR (Livak, 2003). La detección es hecha por la medida

de la disminución de FRET (fluorescence resonance energy transfer) desde un fluoróforo donador a una molécula inhibidora de fluorescencia (quencher). Las sondas TaqMan son sondas alelo-específicas marcadas con un reportero fluorescente en el extremo 5' y un "quencher" en el extremo 3' que virtualmente elimina la fluorescencia en la sonda intacta. Cada ensayo usa dos sondas TaqMan que difieren en el sitio del SNP, y un par de cebadores. Durante la PCR la hibridación exitosa de la sonda TaqMan debido al emparejamiento con uno de los alelos del SNP resulta en su degradación por la actividad nucleasa 5' a 3' de la DNA polimerasa empleada, por lo cual el marcador fluorescente y el quencher son separados, lo cual promueve la fluorescencia (Engle, Simpson, & Landers, 2006). El principio de genotipificación por TaqMan es mostrado en la figura 1-8.

Figura 1-8: Ensayo TaqMan. Fuente: TaqMan® SNP Genotyping Assays. www.lifetechnologies.com/snpcadt.



El ensayo iPLEX Gold (sequenom) es un método para detectar inserciones, deleciones, sustituciones y otros polimorfismos que combinan PCR multiPLEX seguido por una extensión de única base y detección por espectrometría de masas (Oeth, del Mistro, Marnellos, Shi, & van den Boom, 2009). Después de la PCR, los nucleótidos restantes

son inactivados usando fosfatasa alcalina, la cual cliva un fosfato en los dNTPs no incorporados, convirtiendo estos en dDNPs lo cual los hace no disponibles para reacciones de polimerización futuras. Luego, un paso de extensión de cebador es realizado. Los productos de extensión son transferidos dentro de una microplaca conteniendo 384 puntos en la matriz. Los productos de extensión alelo-específico de diferentes masas son analizados usando espectrometría de masas. Este es un método muy sensible que genera datos de alta precisión (Bayés & Gut, 2011).

El Ensayo GoldenGate (Illumina) combina la metodología por extensión de cebador alelo-específica y ligación para generar productos alelo-específicos seguido de la amplificación por PCR con cebadores universales (Shen *et al.*, 2005). Tres oligonucleótidos son diseñados para cada SNP, dos de los cuales son alelo-específicos (ASO) con el SNP en su extremo 3', y un oligonucleótido locus-específico (LSO) que hibrida varias bases corriente abajo del sitio del SNP. En el protocolo, durante el proceso de hibridación, los oligonucleótidos hibridan al ADN genómico que ha sido primero inmovilizado sobre un soporte sólido. El ASO complementario es extendido y ligado al LSO, resultando así en una alta especificidad del locus. Los productos ligados son entonces amplificados usando cebadores universales P1, P2 y P3. Los cebadores P1 y P2 son específicas para cada ASO y portan un marcador fluorescente que es usado para la identificación de los alelos (Bayés & Gut, 2011).

El ensayo de discriminación alélica usando la tecnología GeneChip (Affymetrix) es realizado por hibridación directa de ADN marcado a las matrices que contienen oligonucleótidos alelo-específicos (Engle *et al.*, 2006). Otras plataformas populares para la genotipificación de SNPs son SNPstream (Beckman coulter) basado en la técnica de extensión de base única (Syvanen. 2001) y pirosecuenciación (Qiagen) (Ragoussis. 2009).

2. METODOLOGÍA

2.1 Selección de SNPs candidatos (TagSNPs)

Para definir el set de etiquetas de SNPs, se usó el software Tagger (Bakker *et al.*, 2005) en Haploview 4.2 utilizando los datos de genotipos de CEU y CLM desde el proyecto internacional HapMap (Gibbs *et al.*, 2003) y el proyecto de los 1000 genomas. El gen *DEAR1* tiene 4 transcritos codificando para proteína: ENST00000291416 (3.437 bp), ENST00000373432 (2.246 bp), ENST00000373430 (1.823 bp) y ENST00000543586 (1.638 bp). Variantes comunes con una frecuencia alélica menor (MAF) de $\geq 0,05$ fueron capturadas usando un coeficiente de correlación (r^2) entre etiquetas de SNPs y SNPs no tipificados mayor a 0,8. Se seleccionaron 4 tag SNPs

2.1.1 Análisis genéticos- Genotipificación Tag SNPs

El estudio que se llevó a cabo fue un estudio de casos y controles. Las muestras biológicas incluidas en el estudio fueron:

- A. Muestras de ADN aisladas desde sangre periférica de pacientes con cáncer de seno y de pacientes controles (1023).
- B. Muestras de ADN aisladas desde tejidos tumorales de cáncer de mama embebidos en parafina (60).

2.1.2 Descripción de la Población de estudio

– Descripción población de estudio casos y controles

La población estudio fue tomada de un estudio previo titulado "factores genéticos y no genéticos de cáncer de seno y ovario en Colombia" realizado en el instituto de genética humana de la Pontificia Universidad Javeriana (Torres *et al.*, 2007). La población incluyó

1.023 casos y 1.023 controles. Los casos fueron reclutados principalmente en los hospitales de Neiva, Villavicencio y Bogotá-Colombia entre enero de 2004 y enero 2011. Los criterios de inclusión de los casos fueron mujeres con diagnóstico de cáncer de seno desde el primero de enero de 2004 con o sin antecedentes familiares de cáncer de seno de cualquier edad, que quisieran participar voluntariamente en el estudio.

Los controles fueron mujeres, que participaron en el Programa Nacional de citologías cervicales (test de Papanicolaou), mujeres donantes sanas, sin relación con los casos, adicionalmente se tuvo en cuenta que no presentaran antecedentes de cáncer de seno en la familia en al menos dos generaciones. Los controles fueron emparejados a los casos por la misma edad o una diferencia no mayor a 2 años. Los controles fueron recolectados entre Junio de 2007 y abril de 2012.

Todas las mujeres participantes en el estudio tanto casos como controles firmaron un consentimiento informado (anexo A), llenaron el cuestionario sobre factores de riesgo (anexo B y anexo C), y se les tomó una muestra de sangre de 8ml para la posterior extracción de ADN y conservación de la muestra para estudios posteriores. Adicionalmente se recolectó la información sobre las características clínicas e histopatológicas del tumor incluyendo: histología, grado histológico, tamaño tumoral estado de los ganglios linfáticos, estado del receptor de estrógeno (ER), estado del receptor de progesterona (PR), estado del receptor de factor de crecimiento epidérmico 2 (HER2) y otros parámetros (anexo D).

Datos sobre el seguimiento de las pacientes de cáncer de seno incluyendo fecha del último seguimiento, fecha de recaída loco-regional, fecha de metástasis distantes, fecha de cáncer de seno contralateral, fecha de la muerte y causa de la muerte, fueron recolectados durante el curso de este estudio.

El estudio de casos y controles en Colombia fue aprobado por el comité de ética de la Pontificia Universidad Javeriana. Actualmente la población de estudio está incluida en el consorcio BCAC (breast cancer association consortium).

– **Descripción población de estudio tejidos tumorales**

Los tejidos embebidos en parafina fueron recolectados en el HUSI (Hospital Universitario San Ignacio), algunos tejidos fueron enviados desde el hospital universitario de Neiva. Los

casos con cáncer de seno de inicio temprano y/o historia familiar de cáncer de seno y/o fenotipo triple negativo fueron incluidos, así como algunos casos de pacientes con fenotipo positivo para alguno de los tres receptores hormonales, los controles incluidos fueron tejidos de pacientes con patología benigna de la mama. Se realizó la búsqueda de los casos y controles en la base de datos del departamento de patología del HUSI. La búsqueda en la base de datos se realizó filtrando por glándula mamaria y se tomaron en cuenta los años desde el 2010 hacia atrás para poder contar con los 5 años de seguimiento. Se incluyeron 37 casos con fenotipo triple negativo, 14 casos con fenotipo positivo y 21 controles. Luego de obtener los bloques de parafina se procedió a realizar cortes de cada uno de estos sobre las láminas y se realizó la coloración de hematoxilina-eosina para identificar la zona tumoral y corroborar el diagnóstico. Todas las muestras tumorales fueron examinadas por un patólogo y las muestras de tumor con al menos 70% de células tumorales fueron incluidas.

Para la recolección de los tejidos mamarios provenientes de pacientes sanos, se realizó la búsqueda en la base de datos de pacientes que hubieran consultado por problemas estéticos y que hubieran sido sometidas a cirugías de reducción mamaria que cumplieran con los requisitos establecidos, como no tener diagnóstico de cáncer de seno, no presentar antecedentes familiares de cáncer de seno, no haber presentado lesiones precursoras en el seno y se hizo un emparejamiento por edad con los casos. Los bloques embebidos en parafina de las pacientes control fueron evaluados por un patólogo previo a la extracción de ADN para asegurar que el tejido mamario era normal.

El uso de los bloques de parafina y el uso de los datos e información de los pacientes fue autorizado por el departamento de patología del HUSI.

2.1.3 Extracción de ADN

Sangre periférica

El ADN genómico fue extraído usando el método de extracción salina descrito previamente por (Miller, Dykes, & Polesky, 1988).

Recolección y procesamiento de las muestras

Se extrajeron 8 ml de sangre de cada participante del estudio de casos-contróles de cáncer en Colombia, posterior a la firma del consentimiento informado se procedió a realizar venopunción para la extracción de sangre. Las muestras fueron tomadas en tubos vacutainer™ con anticoagulante EDTA. Las muestras de DNA de todos los participantes están disponibles y son conservadas a -20°C. Posterior a la recolección, las muestras fueron procesadas inmediatamente siguiendo el protocolo descrito en el anexo E.

- **Tejido tumoral**

Como previamente se mencionó, las muestras de cáncer fueron seleccionadas desde el archivo del HUSI. La extracción de ADN, se realizó sobre las muestras que contenían más del 70% de células tumorales. De cada bloque de tejido embebido en parafina se obtuvieron 4 cortes histopatológicos, uno de 3 µm y tres cortes de 10 µm. El de 3 µm de espesor se tiñó con H&E y se usó para identificar las áreas con tejido tumoral y señalar las áreas a microdisectar. Los cortes de 10 µm se emplearon para realizar la microdisección manual (Alvarez & Aristizábal, 2011). La extracción del ADN de tejido tumoral se realizó siguiendo el protocolo de ZYMO (pinpoint slide). La extracción de ADN de tejido mamario normal se realizó según protocolo de Rivero y colaboradores (2006). (Anexo F protocolo para la extracción de DNA tumoral y normal).

2.1.4 Cuantificación de ADN

Las muestras de ADN tanto de sangre periférica y tejido tumoral fueron cuantificadas por espectrofotometría en el equipo nanodrop, usando 2µL de cada muestra de ADN de acuerdo al protocolo establecido por Thermo Scientific (Dreieich, Germany).

2.1.5 Genotipificación

La genotipificación de los tag SNPs rs584298, rs2927970 y rs59983645 fue llevada a cabo por la técnica de MALDI-TOF (Matrix-assisted laser desorption/ionization) utilizando el sistema MassArray IPLEX de Sequenom y el sistema IPLEX. Los datos obtenidos fueron analizados mediante el software TyperAnalyzer de Sequenom versión 4.0.20. <http://agenabio.com/products/massarray-system/>. La genotipificación del tag SNP rs599167 fue realizada usando el ensayo de genotipificación KASP y los datos obtenidos fueron analizados utilizando el software SNPviewer versión 4.0.0.0.

http://www.lgcgroup.com/products/kasp-genotyping_chemistry/overview/#.VhY0UEa6KSo.

Todos los ensayos se realizaron en placas de 384 pocillos. Los cebadores para cada tag SNP fueron diseñados utilizando el software Sequenom MassArray Assay Design.

2.1.6 Control de Calidad

Los controles de calidad estuvieron de acuerdo al consorcio BCAC (breast cancer association consortium) y CIMBA (The Consortium of Investigators of Modifiers of BRCA1/2). Los controles de calidad para las muestras analizadas fueron:

- Muestras duplicadas para MALDI-TOF: 5.8 % de muestras duplicadas (59 casos y 60 controles) fueron genotipificados. La tasa de concordancia fue 100 % para rs584298, rs59983645 y 99.2 % para rs2927970. Al menos 5% de duplicados con una concordancia ≥ 98 % fue requerida.
- Muestras duplicadas ensayo KASP: 6.0% de muestras por duplicado (61 casos y 61 controles) fueron genotipificadas y la tasa de concordancia fue del 100%. Al menos 5% de duplicados con una concordancia ≥ 98 % fue requerida.

Se incluyeron 4 controles NTC (no template control) por cada placa de 384 pocillos. Un mínimo de ≥ 2 NTCs fueron usados por placa. Las muestras de ADN que no mostraron amplificación para más de 2 de los 4 SNPs analizados fueron excluidas (8 muestras).

Después de la exclusión de estas muestras la tasa de llamada “call rate” para SNPs fue rs584298: 98.5%; rs2927970:98.7%; rs59983645:98.3%; rs599167:95.6%. La tasa mínima de llamadas requeridas fue ≥ 95 %.

2.1.7 Protocolo de genotipificación

- IPLEX

El ensayo IPLEX se basa en una reacción de PCR locus- específica, seguida por una extensión de única base usando terminadores dideoxynucleótidos modificados acoplados a un cebador el cual anilla inmediatamente corriente arriba del sitio polimórfico de interés. La identificación se realiza de acuerdo a las distintas masas de los cebadores extendidos usando MALDI-TOF.

- ✓ Amplificación: Para el paso de amplificación se utilizaron los cebadores mostrados en la tabla 2-1

Tabla 2-1 Secuencia cebadores ensayo IPLEX y reacción de amplificación

Cebadores	Secuencia de las cebadores
rs584298: 5' – 3' Cebador directo Cebador Reverso	ACGTTGGATGTCCTTCCCTTCTCAGATGGC ACGTTGGATGTGTTTCTGTCCTGGCAATC
rs2927970: 5' – 3' Cebador directo Cebador Reverso	ACGTTGGATGGAGTCAGAAAGACCACTTTG ACGTTGGATGGGCTGAGAGATTGAGTAACC
rs59983645: 5' – 3' Cebador directo Cebador Reverso	ACGTTGGATGAGTGCATTGCTCTGGGTCAC ACGTTGGATGGCCCCCTGCTAATCTGAATC

Reactivos	Concentración en 5 µL	Volumen (1 Reacción)
Agua (Grado HPLC)		1,8 µL
Buffer PCR 10x (20mM MgCl ₂)	1x	0,5 µL
MgCl ₂ (25mM)	2mM	0,4 µL
Mezcla DNTPs(25mM cada uno)	500 µM	0,1 µL
Mezcla Cebadores (500nM cada uno)	100 nM	1,0 µL
Enzima PCR	1.0 U/rxn	0,2 µL
Total		4,0 µL

En el volumen de reacción final se agregó un excedente del 30%, la cantidad de ADN que se utilizó fue de 1 µL (5-10 ng/ µL). El programa usado para la amplificación en el termociclador fue denaturación inicial 94°C por 2 minutos, seguido de una denaturación de 94°C por 30 segundos, anillamiento 56°C por 30 segundos y extensión 72°C por 1 minuto durante 45 ciclos. Extensión final 72°C por 5 minutos y refrigeración 4°C ∞. Luego de obtener el producto de PCR este debe ser sometido a un tratamiento con SAP

(fosfatasa alcalina) con el fin de defosforilar dNTPs no incorporados por el clivaje de los grupos fosfato en la posición 5'. La mezcla usada para llevar a cabo la reacción de limpieza con SAP es mostrada en la tabla 2-2.

Tabla 2-2: Mezcla de reacción y condiciones de incubación usadas para la reacción con SAP de los SNPs rs2927970, rs584298, rs59983645

Reacción SAP		
Reactivos (Se colocó un exceso en la reacción del 30%)	1	volumen/muestra (µl)
Agua (grado HPLC)	1,53	1,99
Buffer SAP (10x)	0,17	0,22
Enzima SAP (1,7 U/µl)	0,3	0,39
Volumen Total	2,0	2,6
Condiciones de Incubación		
Condiciones SAP		
Temp. C°	Tiempo	
37	40 min.	
85	5 min.	
4	∞	

Luego de realizar el paso de reacción SAP, se procedió a realizar el paso de extensión del cebador o reacción iPLEX, este es el método universal para la detección de polimorfismo de una sola base. La mezcla para la reacción de extensión contiene el cebador de extensión, buffer, enzima y ddNTPs con la masa modificada, esta es adicionada al producto de amplificación. Durante la reacción iPLEX, el cebador es extendido por un nucleótido con la masa modificada dependiendo sobre el alelo y el diseño del ensayo. El protocolo y los cebadores usados para el ensayo IPLEX son mostrados en la tabla 2-3 y 2-4.

Tabla 2-3: Cebadores de extensión usados para la reacción iPLEX de los SNPs rs2927970, rs584298 y rs59983645

Cebadores	Secuencia de los cebadores
rs584298: 5' – 3' Cebador de extensión	CCAGAGCAGGAAGAG
rs2927970: 5' – 3' Cebador de extensión	ggggaCCAAGATCACACTGCAAGTAGGA
rs59983645: 5' – 3' Cebador de extensión	ttagGAATCCAGTCATTTTGAGTCCC

Tabla 2-4: Mezcla de reacción usada para la reacción iPLEX de los SNPs rs2927970, rs584298 y rs59983645✓ **rs2927970**

Reacción iPLEX 19 - 36 plex		
Reactivos (Exceso del 30%)	1	volumen/muestra (µl)
Agua (grado HPLC)	0,619	0,8
Buffer iPLEX (10x)	0,2	0,3
Mix iPLEX extensión	0,2	0,3
Mix cebadores (7; 9,33; 11,66; 14 µM)	0,94	1,2
enzima iPLEX	0,041	0,1
Volumen total	2,0	2,6

✓ **rs584298 y rs59983645**

Reacción iPLEX 19 - 36 plex		
Reactivos (Exceso del 30%)	1	volumen/muestra (µl)
Agua (grado HPLC)	0,7395	0,96135
Buffer iPLEX (10x)	0,2	0,26
Mix iPLEX extensión	0,1	0,13
Mix cebadores (7; 9,33; 11,66; 14 µM)	0,94	1,222
enzima iPLEX	0,0205	0,02665
Volumen total	2,0	2,6

Luego se realizó el proceso de termociclado de las muestras al colocar las placas en el ABI y siguiendo el protocolo mostrado en la tabla 2-5.

Tabla 2-5: Condiciones de amplificación para la reacción iPLEX

Condiciones de PCR iPLEX ("EXTPCR")	Temp. C°	Tiempo
Denaturación inicial	94	30 sec.
1-Denaturación	94	5 sec.
2-Anillamiento	52	5 sec.
3-Extensión	80	5 sec.
Repetir paso 3 por 4 ciclos mas		
Repetir paso 2 por 39 ciclos mas		

Extensión final	72	3 min.
Refrigeración	4	∞

El siguiente paso fue realizar la limpieza de los productos de amplificación de la reacción iPLEX para lo cual se usó una resina (SpectroCLEAN) de Sequenom, esta es una resina catiónica pretratada con reactivos ácidos. Una porción de la resina es adicionada directamente sobre los productos de la reacción iPLEX con el fin de remover sales como iones Na^+ , K^+ y Mg^{2+} . Por último se realizó la detección de los productos de la reacción iPLEX por espectrometría de masas.

- KASP

El ensayo KASP (Kompetitive allele specific PCR) se basa en una PCR competitiva específica de alelo, con un sistema reportero fluorescente para la identificación y medida de la variación genética que se desea estudiar. El ensayo KASP usa tres componentes que son la muestra de ADN con el SNP de interés; el mix KASP que contiene dos diferentes cebadores alelo específicos en sentido directo (uno para cada alelo del SNP), con secuencias específicas en 5' y un cebador en sentido reverso. También incluye el KASP master mix donde se encuentra el casete FRET más la Taq polimerasa en una solución buffer optimizada. Los cebadores usados para la amplificación del SNP rs599167 son mostrados en la tabla 2-6.

Tabla 2-6: Cebadores ensayo KASP

Cebadores	Secuencia
Primer alelo 1- FAM	CAACCACCCCCAGAGGCTCA
Primer alelo 2- HEX	CAACCACCCCCAGAGGCTCT
Primer reverse común	TTCCCAAAGGGTGGGCCTTTTGAAT

Las muestras fueron amplificadas por PCR usando la mezcla de reacción y las condiciones de amplificación mostradas en las tablas 2-7 y 2-8

Tabla 2-7: Mezcla de reacción para el ensayo KASP del SNP rs599167

Reactivos	KASP Genotipificación
DNA (5-50ng)	2,5
2x Master Mix	2,5
Primer mix	0,07
H ₂ O	N/A
Volumen total	5

Tabla 2-8: Condiciones de PCR para el ensayo KASP del SNP rs599167

Condiciones de PCR	
94° C 15 min	Activación polimerasa
94 ° C 20 seg.	10 ciclos
61-55°C 60 seg. (disminuyendo 0.6°C por ciclo)	
94°C 20 seg.	26 ciclos
55°C 60 seg.	

Por último, se realizó la lectura de las placas teniendo en cuenta los valores de emisión y excitación de los fluoróforos que son mostrados en la tabla 2-9.

Tabla 2-9: Valores de excitación y emisión de los fluoróforos usados en el ensayo KASP.

Fluoróforo	Excitación (nm)	Emisión (nm)
FAM	485	520
HEX	535	556
ROX	575	610

El marcador ROX se usa como un fluoróforo de referencia y también permite la normalización de las variaciones en la señal causadas por diferencias en el volumen de líquido de pocillo a pocillo en la placa de reacción. Por último, se realizó la visualización gráfica de los resultados.

2.1.8 Estimación de las frecuencias alélicas y genotípicas para los tag SNPs en TRIM62

Una vez se analizaron los perfiles del proceso de genotipificación, se calcularon las frecuencias genotípicas y alélicas teniendo en cuenta los posibles genotipos y alelos en la población, los datos de las frecuencias genotípicas y alélicas son mostrados en la tabla 2-10.

Tabla 2-10: Genotipos y alelos para cada uno de los tag SNPs analizados

Tag SNPs	Genotipos	Alelos
rs584298	GG AG AA	G Silvestre (p) A Mutado (q)
rs59983645	CC CT TT	C Silvestre (p) T Mutado (q)
rs2927970	CC TC TT	C Silvestre (p) T Mutado (q)
rs599167	AA TA TT	A Silvestre (p) T Mutado (q)

Las frecuencias genotípicas se calcularon por conteo directo y las frecuencias alélicas se estimaron empleando el método descrito por Hartl (Hartl, 2000).

Frecuencia alelo silvestre (p):

$$p = F_{a/a} + \frac{1}{2} F_{a/b}$$

Frecuencia alelo mutado (q):

$$q = F_{b/b} + \frac{1}{2} F_{a/b}, \text{ donde;}$$

F a/a = Frecuencia genotipo homocigótico silvestre

F b/b = Frecuencia genotipo Homocigóticos mutado

F a/b = Frecuencia genotipo Heterocigótico

2.2 Secuenciación de tejido tumoral

La secuenciación directa de la región codificante del gen *DEAR1* se realizó en 60 muestras de ADN de tejido embebido en parafina diferentes a las muestras incluidas en la cohorte de asociación, 11 muestras provenientes de pacientes Neiva-Huila que habían sido incluidas en la cohorte inicial fueron incluidas para este análisis.

2.2.1 Amplificación de ADN de tejido tumoral y normal

Para enviar las muestras al servicio de secuenciación se realizó primero la amplificación por PCR del ADN extraído desde los tejidos normales y tumorales. Se diseñaron varios pares de cebadores para cada uno de los exones con el objetivo de obtener fragmentos de un tamaño no mayor a 300bp, debido a la dificultad para la amplificación del ADN en muestras provenientes de tejidos embebidos en parafina.

Las condiciones de la reacción como el protocolo de amplificación para cada exón son mostradas en las tablas 2-11, 2-12 y 2-13.

Tabla 2-11: Mezcla de reacción empleada en la amplificación de exones TRIM62

Reactivos	1x μL	[] inicial	[] final	
Buffer	8	5X	1X	
DNTPs	5	2mM	0,25mM	
Cebador Directo	1	10μM	0,25 μM	
Cebador Reverso	1	10μM	0,25 μM	
[] Molar	1mM	2mM	3Mm	4mM
MgCl2 (μL)	1,6	3,2	4,8	6,4
Taq (μL)	0,2	0,2	0,2	0,2
DNA (μL)	4	4	4	4
Agua (μL)	19,2	17,6	16	14,4

Tabla 2-12: Cebadores usados para la amplificación de los exones del gen TRIM62

Exón <i>TRIM62</i>	Tamaño del Exón	Secuencia de los cebadores	Tamaño del fragmento
Exón 1	585 bp	Cebador directo 1a: 5'-TAGTCTCCTACCCCTGCCTGT-3'	243 bp
		Cebador reverso 1a: 5'-CAGTGCTCCGTGATGCAG-3'	
		Cebador directo 1b: 5'-GCCTGAGCATCTACCAGGAC-3'	200 bp
		Cebador reverso 1b: 5'-CAGCGGGAAGGAGCTGTAG-3'	
		Cebador directo 1c: 5'-AAGCTGGCCAACATCGTG-3'	324 bp
		Cebador reverso 1c: 5'-CCCCAGGACCATTCTGACT-3'	

Exón TRIM62	Tamaño del exón	Secuencia de los cebadores	Tamaño del fragmento
Exón 2	210 bp	Cebador directo 2: 5'-CAGTGGTCAGGGCTGAATG-3' Cebador reverso 2: 5'-AGCCCTGCCCTCATCTCT-3'	218 bp
Exón 3	420 bp	Cebador directo 3a: 5'-CAGAGATCACCCCATCTGCT-3' Cebador reverso 3a: 5'-GGCTGTAGCGCTGGACTTT-3'	210 bp
		Cebador directo 3b: 5'-CCATGCTAGAGGAGCTGGAG-3' Cebador reverso 3b : 5'-CTGCTAAGGATCCCATCTGC-3'	272 bp
Exón 4	200 bp	Cebador directo 4: 5'-AGGAATCCAGTCCCTGGTG-3' Cebador reverso 4: 5'-ACATGCCCCACCTAGCTCT-3'	205 bp
Exón 5	800 bp	Cebador directo 5a: 5'-CAGA ACTCTGCCCTCAGGAA-3' Cebador reverso 5a: 5'-GGTCTTCTCCGCCACCAC-3'	292 bp
		Cebador directo 5b: 5'-GTGTCGGTGCTGGGTTCT-3' Cebador reverso 5b: 5'-CAGGACATGTCATCAGCATTG-3'	299 bp
		Cebador directo 5c: 5'-ACAAGGTGGGTGTCTTCCTG-3' Cebador reverso 5c: 5'-GCTCTTGCAGGTGGCAGT-3'	240 bp

Tabla 2-13 Condiciones de PCR empleadas en la amplificación de los exones TRIM62

Pasos	Temperatura	Tiempo	Numero de ciclos
Denaturación inicial	95 °C	05:00	1X
Denaturación Anillamiento Extensión	95° C	00:30	32X
	Tm	00:30	
	72 °C	00:30	
Extensión Final	72°C	10:00	1X
Enfriamiento	4°C	∞	1X

Cebadores	Tm Estandarizada	Concentración MgCl₂
Cebador directo 1a Cebador reverso 1a	61.3	1mM
Cebador directo 1b Cebador Reverso 1b	63	1mM
Cebador directo 1c Cebador Reverso 1c	63	2mM
Cebador directo 2 Cebador Reverso 2	61	2mM
Cebador directo 3a Cebador Reverso 3a	63	4mM
Cebador directo 3b Cebador Reverso 3b	62	3mM
Cebador directo 4 Cebador Reverso 4	63	1mM
Cebador directo 5a Cebador Reverso 5a	65	1mM
Cebador directo 5b Cebador Reverso 5b	59	1mM
Cebador directo 5c Cebador Reverso 5c	64	1mM

En cada una de las muestras se realizó reamplificación de los productos de PCR. Antes de enviar las muestras al servicio de secuenciación se verificaron las condiciones de la amplificación, se analizaron 10 µL del producto de PCR mediante electroforesis en gel de agarosa al 1,5% y 2% dependiendo del tamaño de banda esperado. Para visualizar los fragmentos a través de la luz ultravioleta la agarosa fue teñida con SYBR® Safe siguiendo las recomendaciones del fabricante. Una vez se verificó la amplificación de los exones, el producto de PCR fue enviado al servicio de secuenciación en MacroGen de acuerdo a las instrucciones provistas. Se realizó la purificación de las muestras y posterior secuenciación por duplicado y bidireccional de las muestras.

2.2.2 Análisis de secuencias

El análisis de las secuencias fue llevado a cabo usando el software CLC genomics workbench versión 8.0.

– Control de calidad de la descarga

Los datos de secuenciación fueron descargados desde el sitio web de Macrogen Para comprobar que la descarga está completa, utilizamos el comando md5sum –cmd5.txt en Linux, comprobando que el número que obtenemos y el que consta en el archivo md5.txt de la descarga son el mismo.

– Comprobación de la calidad de los datos

El formato en el que se reciben desde el laboratorio las lecturas de la secuenciación es en formato ABI. (*ab.1) Este tipo de archivos puede ser importado directamente al CLC WORKBENCH, luego las secuencias son sometidas a los puntajes de calidad. Donde para cada base nucleotídica correctamente alineada el programa nos arroja una probabilidad. De esta manera una probabilidad de 1 significa alineación perfecta a la secuencia de referencia.

Dentro del proceso de comprobación de la calidad de los datos se evalúan varios parámetros como:

- La distribución de la calidad promedio de la secuencia: gráfica que representa la distribución de la calidad de las lecturas. Si una gran cantidad de secuencias en un ciclo determinado son de baja calidad, puede deberse a un problema sistemático del proceso de secuenciación.
- Calidad de la secuencia por base: gráfica que muestra la calidad de las bases según la posición de lectura; si la calidad de las lecturas en las distintas posiciones es alta y homogénea (rango intercuartílico 25-75%) las secuencias son de excelente calidad, por el contrario si la calidad de las lecturas largas es mala, observándose un aumento de variabilidad en las lecturas de las últimas posiciones las secuencias serían de mala calidad.

- Contenido de GC por base. Muestra la distribución de GC por posición de lectura. La aparición de desviaciones puede deberse a la presencia de alguna secuencia sobrerrepresentada.
- Contribuciones de nucleótidos. Gráfica que muestra la distribución de los diferentes nucleótidos a lo largo de la secuencia.
- Cubrimiento. Gráfica que muestra el número de secuencias que cubren las bases individuales a lo largo de las lecturas.
- Contenido de Ns por base. Se utiliza N para marcar las bases de las que no se ha obtenido lectura. Cuantas más Ns contengan los datos su calidad será menor
- Longitud de distribución. Gráfica que representa la longitud de las secuencias en pares de bases con respecto al porcentaje de secuencias.

Luego de comprobar la calidad de las secuencias se procedió a realizar el ensamblaje de las mismas y el ensamblaje a la secuencia de referencia. Primero se realizó un “TRIM sequence” o cortado de la secuencia donde se remueven todas aquellas pares de bases que no muestran una buena calidad por lo cual esto crea anotaciones sobre las secuencias removidas que no serán tenidas en cuenta en el proceso de ensamblaje.

Para este proceso se tienen en cuenta dos parámetros:

- Puntajes de calidad que en este caso lo definimos como 0.02
- Uso de nucleótidos ambiguos el cual fue definido como 2 residuos.

Luego de obtener las secuencias limpias, se realizó el ensamblaje de las secuencias dentro de contigs, los cuales son segmentos de ADN superpuestos, donde juntos representan una región consenso de ADN; para lo cual se tuvieron en cuenta los siguientes parámetros:

- Mínima longitud de lectura en el alineamiento: **30**, este parámetro se refiere al número mínimo de nucleótidos en una lectura el cual debe ser alineado exitosamente al contig.
- Astringencia del alineamiento: **media**, este parámetro especifica la rigurosidad de la función de puntuación utilizada en la etapa de alineamiento. Por lo tanto un alto nivel. .

de astringencia producirá contigs con pocas ambigüedades, pero también omitirá más lecturas de secuencias y generará mayor número de contigs de menor tamaño. Al obtener las secuencias organizadas en contigs, se procedió a realizar el ensamblaje a la secuencia de referencia, para esto se descargó la secuencia del cDNA del gen de interés desde la base de datos del NCBI con número de acceso NM 018207.02, aquí se utilizaron los mismos parámetros usados en el ensamblaje de las secuencias en contigs.

Aunque la astringencia en este caso especifica la función de puntuación usada para alinear las secuencias de entrada a la secuencia de referencia.

Utilizando la herramienta de detector de variantes, se realizó la detección de SNPs donde se tuvieron en cuenta diferentes parámetros como:

- Cubrimiento mínimo de las secuencias: **10**, que corresponde al número de lecturas que cubren la secuencia
- Frecuencia mínima de la variante: **35%**.

La identificación se realizó teniendo en cuenta el modelo de control standard por vecindad (Neighborhood quality Standard-NQS).

Se realizó la detección de variantes de baja frecuencia la cual es una herramienta dentro del CLC que se basa en un modelo estadístico para las muestras analizadas y un modelo para la secuenciación de errores. El parámetro que se debe establecer para el análisis es la significancia, el cual determina el valor cut-off para los test estadísticos para la variante que asume no es debida a errores de secuenciación, el valor de significancia que nosotros incluimos fue del 1%.

2.3 Equilibrio Hardy-Weinberg

Para comprobar si la población se encontraba en equilibrio Hardy-Weinberg, se compararon las frecuencias genéticas observadas con las esperadas empleando la prueba χ^2 de bondad de ajuste. Para el análisis se empleó un grado de libertad y un $p < 0,05$, se consideró significativo.

2.4 Análisis estadísticos

2.4.1 Asociación entre genotipos de TRIM62 con las variables epidemiológicas e histopatológicas

Para evaluar la asociación de las variantes genéticas en *DEAR1* con cáncer de seno se realizó un análisis de regresión logística.

Para el análisis de regresión logística condicional se tuvieron en cuenta dos modelos en el primer modelo se analizaron las variables de exposición (factores de riesgo epidemiológicos) más las variables control (genotipos de los polimorfismos en *DEAR1*) como variables independientes y la presencia de cáncer de seno como la variable dependiente. Las variables epidemiológicas de riesgo que se incluyeron en el análisis fueron menopausia, antecedentes familiares de cáncer de seno, el uso de anticonceptivos orales, uso de la terapia hormonal en periodo posmenopáusico, el índice de masa corporal, tabaquismo y la edad.

Para el segundo modelo se evaluaron las asociaciones entre genotipos de las variantes SNPs analizadas y 7 características clínicas e histopatológicas del tumor (histología, grado histológico, tamaño del tumor, estado de los nódulos linfoides, el estado de receptores de estrógenos (ER), progestágenos (PR) y de receptores celulares de membrana ErbB (HER2) de los casos de cáncer de seno, en este caso no se incluyeron covariables. Todas las pruebas se realizaron a 2 colas.

En el primer modelo se evaluaron las interacciones y confusiones de las covariables, así como la colinealidad con el fin de conocer cuales covariables incluidas en el análisis no eran independientes. Aquellas variables que no mostraron independencia fueron excluidas del modelo final de regresión. Para el análisis de colinealidad se tuvo en cuenta la opción VIF (factor de inflación de la varianza) dentro del análisis en SAS. Estos factores miden la inflación en las varianzas de los parámetros estimados debido a colinealidades que existen entre los regresores (variables independientes). Por lo cual un valor superior a 10 en el índice de condición se tomó como colineal y se retiraron los predictores correspondientes del modelo.

Para la identificación del modelo final se utilizó la aproximación “step-wise” que consiste en una selección de las variables independientes que muestran un efecto significativo sobre la variable respuesta.

La medida de asociación que se empleó para los dos modelos fue “odds ratios” (OR) y un intervalo de confianza (IC) del 95%. Para los análisis estadísticos se utilizó el paquete estadístico SAS versión 9.4 (SAS Institute Inc, Cary, NC, EE.UU.).

2.4.2 Análisis de supervivencia

Para los análisis de supervivencia se tuvieron en cuenta los siguientes puntos finales:

- 1.) Tiempo a la recurrencia: definido como el tiempo desde el diagnóstico o documentación al azar de un evento de seno, cualquier recurrencia local, regional o recurrencia distante de cáncer de seno o un cáncer de seno contralateral.
- 2.) Sobrevida específica de cáncer de seno (BCSS): se define como el tiempo desde el diagnóstico hasta la muerte por cáncer de seno.
- 3.) Sobrevida libre global (OSS): en general se calcula, como el tiempo desde la aparición de la enfermedad hasta la muerte por cualquier causa.

Los estimados paramétricos de la función de sobrevida se realizaron usando la distribución de Weibull para comparar las curvas de sobrevida con observaciones censuradas a la derecha (pacientes que se retiraron antes de conocer alguno de los eventos de desenlace) y para calcular las pruebas de asociación de la variable tiempo al evento con las covariables. Las tasas de sobrevida se muestran con sus intervalos de confianza del 95%. Un análisis multivariado de sobrevida con selección de variables pasó a paso, fue llevado a cabo con el fin de conocer la asociación de los genotipos en *DEAR1* con la tasa de sobrevida.

2.5 Consideraciones éticas

Los protocolos y procedimientos empleados en este trabajo fueron aprobados por el Comité de Ética de la Facultad de Medicina de la Pontificia Universidad Javeriana sede-Bogotá, así como de la Universidad de la Sabana.

El estudio además fue realizado de acuerdo a las normas y objetivos de la Declaración de Helsinki, adoptada en la 18ª Asamblea General de la Asociación Médica Mundial (Helsinki, Finlandia, Junio de 1964), revisada por la 29ª Asamblea Médica Mundial (Tokio, Japón, Octubre de 1975) y por la 35ª Asamblea Médica Mundial (Venecia, Italia, Octubre de 1983) (World Medical Association).

I

3. Resultados y discusión

3.1 Descripción de la población de estudio

3.1.1 Casos y controles

Se incluyeron 1022 casos y 1023 controles, uno de los casos fue excluido debido a que presentaba como diagnóstico primario cáncer de ovario. Los casos y controles fueron emparejados por edad en ± 2 años. Los casos provenían principalmente de las ciudades de Bogotá, Neiva y Villavicencio. Los controles fueron mujeres saludables no relacionadas con los casos quienes reportaron no tener historia familiar de cáncer de seno ni de ningún otro tipo de cáncer.

Información sobre etnicidad, historia familiar y personal de cáncer de seno fue obtenida de todos los participantes del estudio por medio del uso de un cuestionario. Las variables sobre factores de riesgo epidemiológicos incluyeron: uso de anticonceptivos orales, uso de terapia hormonal de reemplazo, período de transición entre la etapa reproductiva y no reproductiva (menopausia); Hábito de fumar e historia familiar de cáncer.

La estadística descriptiva sobre la información incluida en el estudio acerca de los factores de riesgo epidemiológicos para casos y controles, así como la información histopatológica de los casos es mostrada en las tablas 3-1 y 3-2.

Las variables epidemiológicas incluidas en el análisis se definieron de la siguiente manera:

- Hábito de fumar: variable categórica 0=nunca, 1=pasado, 2=Actualmente, (en el último año antes de la fecha de ingreso al estudio), para los casos se tuvo en

cuenta el año del diagnóstico de cáncer de seno y para los controles el año de ingreso al estudio. 888=desconocido.

- Uso de anticonceptivos orales: variable categórica. 0=nunca, 1=Si/actualmente, 888=desconocido. La categoría 1 se define como al menos 4 meses de uso.
- Uso de terapia de reemplazo hormonal: 0=nunca, 1=Si/actualmente, 888=desconocido. La categoría 1 se define como al menos 3 meses de uso.
- Menopausia: 1: pre/peri menopáusica 2: post menopáusica (definida como última menstruación hace más de doce meses antes de la fecha de ingreso).
- Historia Familiar: 0: No 1: Si definido como historia familiar de cáncer de seno en un pariente de primer grado.

Tabla 3-1: Estadística descriptiva variables categóricas de riesgo epidemiológico incluidas en el análisis

Factores de riesgo epidemiológicos- Variables Categóricas			
		Frecuencia	%
Fumar	Nunca	1459	72.30
	En el pasado	437	21.66
	Actualmente	122	6.05
	# Total de observaciones	2018	100
Uso de anticonceptivos orales	Nunca o < 3 meses	1429	71.13
	Si y/o actualmente	580	28.87
	# Total de observaciones	2009	100
Menopausia	pre-menopáusica	758	37.54
	pos-menopáusica	1261	62.46
	# Total de observaciones	2019	100

Factores de riesgo epidemiológicos- Variables Categóricas			
		Frecuencia	%
Uso de HRT	Nunca 0 < 3 meses	1860	92.40
	Si y/o > 3 meses	153	7.60
	# Total de observaciones	2013	100
Historia Familiar	No	1689	83.78
	Si	327	16.22
	# Total de observaciones	2016	100

En la tabla 3-2 se muestran aquellas variables de tipo continuo incluidas en el análisis.

Tabla 3-2: Estadística descriptiva variables continuas incluidas en el estudio

Variables continuas				
	Media	Error estándar	IC [95%]	# Total de Observaciones
BMI	25.49	0.091	25.31-25.66	1932
Edad de diagnóstico para los casos	49.75	0.367	49.03-50.47	1020
Edad de Ingreso	50.58	0.259	50.07-51.09	2044

La mayoría de las pacientes incluidas en el estudio, eran pacientes posmenopáusicas, en la muestra se observó que las pacientes presentaban sobrepeso y que la edad de diagnóstico para los casos, así como la edad de ingreso para los controles estuvo en el rango de 49 a 51 años.

Las variables histopatológicas incluidas se definieron como:

- Patología (Histología): 1: Cáncer invasivo incluyendo carcinoma ductal infiltrante o carcinoma lobular infiltrante o cualquier otro tipo de cáncer de tipo invasivo según la clasificación WHO (Sinn & Kreipe, 2013) 2:Cancer *In situ* esta categoría incluye carcinoma ductal *in situ* y lobular *in situ* .

- Receptor de estrógenos, progesterona y HER2: 0: Negativo y 1: Positivo.
- Estado de los nódulos: linfoides 0: negativo y 1: positivo.
- Tamaño del tumor: en esta variable se tuvo en cuenta el tamaño del tumor en el momento de la cirugía (mastectomía, lumpectomía o cuadrantectomía) en milímetros.

La estadística descriptiva para las variables histopatológicas es mostrada en la tabla 3-3 y 3-4.

Tabla 3-3: Estadística descriptiva variables histopatológicas categóricas

Variables histopatológicas			
		Frecuencia	%
Patología	Invasivo	800	84.39
	In -situ	148	15.61
	# Total de observaciones	948	100
Receptores de estrógeno	Negativo	170	22.31
	Positivo	592	77.69
	# Total de observaciones	762	100
Receptores de progesterona	Negativo	210	27.74
	Positivo	547	72.26
	# Total de observaciones	757	100
Receptor HER2	Negativo	455	75.58
	Positivo	147	24.42
	# Total de observaciones	602	100

Tabla 3-4: Estadística descriptiva variables histopatológicas continuas

Variables histopatológicas continuas				
Variabes	Media	Error estándar	IC [95%]	# Total de Observaciones
Tamaño del tumor	23.30	0.655	22.01 24.59	585
Numero de nódulos positivos	1.95	0.172	1.61 2.28	661

El comportamiento tumoral fue principalmente invasivo, con tumores respondedores a hormonoterapia y con un tamaño tumoral >10mm.

En el cuestionario sobre factores de riesgo epidemiológicos que completaron todas las pacientes del estudio también se incluyó información acerca de la paridad, nivel educativo, consumo de alcohol, uso de medicamentos entre otras variables de interés, sin embargo, para el desarrollo de los análisis estadísticos en este estudio solo se tuvieron en cuenta las variables mencionadas anteriormente ya que se conoce que tienen un riesgo importante asociado al desarrollo de la enfermedad.

En cuanto a las variables histopatológicas se tuvieron en cuenta aquellas variables que mostraron asociación con la pérdida de expresión de *DEAR1* en el estudio realizado por Lott (2009) (Lott et al., 2009a) estas variables fueron estado de los receptores de estrógeno, progesterona y HER2, además de las variables que se sabe están asociadas con el pronóstico de la enfermedad.

Las muestras de ADN de casos y controles se encuentran en el banco de muestras de la Pontificia Universidad Javeriana y se encuentran disponibles todas las muestras para estudios posteriores, en términos generales el método de extracción salina permitió purificar ADN de buena calidad (1,8 - 2,0) y en un amplio rango de concentraciones (25 ng/μL - 600 ng/μL).

El seguimiento de las pacientes se realizó cinco años después de la fecha de ingreso, con el fin de conocer la información con respecto a recaídas, recaídas contralaterales, muerte por cáncer de seno o muerte por cualquier otra causa, información indispensable para ser incluida en los análisis de sobrevida. El seguimiento de las pacientes se realizó por medio de llamadas telefónicas a las pacientes y/o familiares consignados dentro de la historia clínica, en algunos casos se contactó directamente a los médicos tratantes. La estadística descriptiva sobre la información recolectada para ser incluida en los análisis de sobrevida es mostrada en la tabla 3-5.

Tabla 3-5: Estadística descriptiva variables seguimiento

Datos supervivencia			
DESENLACES		Frecuencia	%
Recaída local-regional	Si	27	3
	Desconocido	327	32
	No	668	65
	Total	1022	
Metástasis Distantes	Si	109	11
	Desconocido	324	32
	No	589	58
	Total	1022	
Estado Vital	Viva	627	61
	Muerta	77	8
	Desconocido	318	31
	Total	1022	
Causa de Muerte	Cáncer de seno	25	32
	Otra causa	20	26
	Desconocido	32	42
	Total	77	

3.1.2 Tejidos embebidos en parafina

Las muestras de tejidos tumorales y tejidos normales embebidos en parafina que fueron incluidas en el estudio se seleccionaron a partir de la base de datos de patología del Hospital Universitario San Ignacio, según los criterios de selección expuestos en la metodología.

Los datos acerca de las características histopatológicas del tumor fueron recolectados desde los participantes del estudio. La recolección de la información fue hecha a través de la consulta en las bases de datos del HUSI. La información sobre datos histopatológicos fue recogida por medio de los reportes de patología de cada paciente y los datos de seguimiento se recolectaron teniendo en cuenta la fecha de la última consulta realizada por el paciente en el HUSI. La información clínica e histopatológica incluyó: edad del diagnóstico de cáncer de seno, histología, grado histológico, tamaño del tumor, estado de los nódulos linfoides, estado del receptor de estrógenos, estado del receptor de progesterona y estado del receptor de factor de crecimiento epidérmico.

La descripción de la información recolectada para las muestras de tejido tumoral embebido en parafina es mostrada en la tabla 3-6.

Tabla 3-6: Variables histopatológicas de los tejidos tumorales

Variables Histopatológicas tejidos tumorales			
Parámetro	Categorías	Frecuencia	%
Patología	Invasivo	25	49
	In -situ	19	37,3
	Otros	3	5,9
	desconocido	4	7,8
	# Total de observaciones	51	100
Receptor de estrógeno	Negativo	37	72,5
	Positivo	14	27,5

	# Total de observaciones	51	100
Receptores de progesterona	Negativo	37	72,5
	Positivo	14	27,5
	# Total de observaciones	51	100
Receptor HER2	Negativo	39	76,5
	Positivo	3	5,9
	Desconocido	9	17,6
	# Total de observaciones	51	100
Grado del tumor	I	3	5,9
	II	17	33,3
	III	17	33,3
	Desconocido	14	27,5
	# Total de observaciones	51	100,0
Estado tumoral	I	2	3,9
	II	6	11,8
	III	7	13,7
	Desconocido	35	68,6
	# Total de observaciones	51	100,0

Tabla 3-7: Variables histopatológicas continuas tejidos tumorales

Variables histopatológicas continuas				
Variables	Media	Error estándar	IC [95%]	# Total de Observaciones
Tamaño del tumor	1.85	0.0875	1.68 2.02	32
Número de nódulos positivos	1.14	0.34	0.47 1.81	35

Cabe resaltar que el volumen de la muestra, la pureza y la concentración fueron unas limitantes para este estudio, pues para la mayoría las concentraciones fueron bajas (1,5 –

51,2 ng/ μ L), los volúmenes reducidos (< 20 μ L) y la pureza en términos de relación A260/A280 estuvo por debajo de los valores de referencia. Cuatro muestras mostraron amplificación de los 10 exones y se obtuvieron secuencias de buena calidad. En los exones 2, 3^a, 5c, ninguna muestra mostro amplificación al visualizar los geles de agarosa, por ello para estas muestras se realizó de nuevo la extracción de ADN, usando el método de Fenol-Cloroformo-Alcohol - isoamílico (25:24:1). Nuevamente se hizo la amplificación de las muestras, donde del total de 72 muestras incluidas en el estudio solo 60 muestras mostraron amplificación.

3.2 Selección de los “Tag SNPs”

Para la selección de las etiquetas de SNP se utilizó el software Tagger, y se realizó la selección basados en las poblaciones CEU y CLM del proyecto HapMap y 1000 genomas respectivamente. Se realizaron 3 aproximaciones para la selección de las etiquetas de SNPs debido primero a que se quería hacer una comparación entre las dos poblaciones (CEU y CLM), y segundo a que se tomó la última actualización de los proyectos HapMap y 1000 genomas para la selección de los “TagSNPs”.

3.2.1 Primera aproximación “tagging SNPs” comparando dos poblaciones

En esta primera aproximación se realizó la comparación de dos poblaciones desde el proyecto de los 1000 genomas y el proyecto HapMap.

El objetivo del proyecto de los 1000 genomas es construir un recurso que ayude a entender la contribución genética a la enfermedad, este recurso provee el genoma de 1,092 individuos desde 14 poblaciones, estas poblaciones son: población China en Beijing, China (CHB), población del sur de china (CHS), población Japonesa en Tokio, Japón (JPT) , población Luhya en Webuye, Kenia (LWK), población Yoruba en Ibadán, Nigeria (YRI), población con ancestría Africana desde el sureste de los Estados Unidos (ASW) , población de ancestría Mexicana en los Ángeles, CA (MXL), población de puerto rico (PUR), población Finlandesa desde Finlandia (FIN), Toscana en Italia (TSI), residentes de Utah con ancestría Europea del norte y este (CEU), Británica desde Inglaterra y Escocia (GBR), poblaciones Ibéricas en España (IBS) y Colombia en Medellín, Colombia (CLM).

El proyecto HapMap tiene como objetivo principal determinar patrones comunes de variación en la secuencia de ADN en el genoma humano, por caracterizar variantes de la secuencia, sus frecuencias y correlaciones entre estas, en muestras de ADN desde poblaciones con ancestría Africana, Asiática y Europea y puede ser un recurso clave para encontrar genes implicados en salud, enfermedad y respuestas a fármacos y factores ambientales. Las poblaciones incluidas fueron: 30 tríos desde la población Yoruba de Ibadán-Nigeria (YRI), 30 tríos de residentes de Utah con ancestría del norte y este de Europa (CEU), población con ancestría africana en el sureste de Estados Unidos (ASW), 44 individuos desde Japón- Tokio no relacionados (JPT), 44 individuos no relacionados desde Beijing China (CHB), Indios Gujarati en Houston- Texas (GIH), población Luhya en Webuye , Kenia (LWK), población con ancestría Mexicana en los Ángeles-California (MEX), población Maasai en Kinyawa, Kenia (MKK) y población desde Toscana en Italia (TSI).

De acuerdo a HapMap (fase 3, publicación 27) el gen *DEAR1* ubicado en el cromosoma 1 se extiende desde la posición 33383591 a 33419854. Basándonos sobre datos de genotipos desde 116 individuos de la población CEU. El gen *DEAR1* más 5Kb de la región promotora (5,638 bp corriente arriba del codón de inicio ATG) contiene 94 SNPs (en los 4 transcritos). Fuera de estos 94 SNPs, 91 son intrónicos y tres están localizados en la región promotora.

De acuerdo al proyecto de los 1000 genomas (fase 1, publicación 20101123), el gen *DEAR1* se extiende desde la posición 33611003 a 33647671. Basados sobre datos de genotipos de 60 individuos desde la población CLM, el gen *DEAR1* mas la región promotora contiene 257 SNPs (en los 4 transcritos). En los tres primeros transcritos (ENST00000291416, ENST00000373432 y ENST00000373430) se encontraron 11 SNPs en la región 3'UTR, 1 en la región 5'UTR, 201 intrónicos, 2 no sinónimos codificantes, 8 sinónimos codificantes y 33 en la región promotora. En el cuarto transcrito (ENST00000543586) se encontraron 3 SNPs en la región 3' UTR ,1 en la región 5'UTR, 169 intrónicos, 2 codificantes no sinónimos, 7 sinónimos codificantes y 33 en la región promotora.

La primera aproximación "Tag SNPs" ($r^2 \geq 0.8$ y frecuencia alélica menor (MAF) $\geq 5\%$) reveló que los patrones de ligamiento de *DEAR1* difieren entre la población CEU (tres Tag SNPs capturan 26 alelos) y la población CLM desde el proyecto de los 1000 genomas (14 Tag SNPs capturan 64 alelos). Debido a esto, se realizó el análisis con los datos del

proyecto de los 1000 genomas, ya que se asumió que la población CLM podría estar genéticamente más relacionada con la población de estudio.

3.2.2 Segunda aproximación “Tagging SNPs” sobre la población CLM

De acuerdo al proyecto de los 1000 genomas (**fase 1**) el gen *DEAR1* se extiende desde la posición 33611003 a 33647671 sobre el cromosoma 1 (ncbi 37). Basados sobre datos de los genotipos de la población CLM la región del gen *DEAR1* mas la región promotora de 5kb contiene 170 SNPs (4 transcritos). Adicionalmente, 65 SNPs presentaban un MAF $\geq 10\%$, incluyendo 9 SNPs localizados en la región promotora y un SNP sinónimo codificante. Doce “tag SNPs” capturaban el 100% de los alelos de estos 65 SNPs con un coeficiente de correlación de $r^2 \geq 0.8$. Los SNP con alto valor de MAF entre un grupo de SNP etiquetados fueron seleccionados como “tag SNP”. Los tag SNPs seleccionados en la segunda aproximación son mostrados en la tabla 3-8.

Tabla3-8 Alelos capturados por los "Tag SNPs" segunda aproximación

Tag SNP	MAF	Alelos capturados
rs784769	0,142	rs149228047,rs627558,rs6679508,rs622407,rs12037147,rs3766820,rs784769,rs683849,rs679602,rs643402,rs634307,rs3766815,rs680960,rs614039,rs784770,rs636561,rs603186,rs635677,rs587910
rs584298	0,217	rs654336,rs685235,rs621539,rs584298,rs599167,rs689187,rs641657,rs607008,rs784773,rs631646
rs59983645	0,15	rs35111341,rs12734992,rs72656238,rs59983645,rs623596,rs3912145,rs12563712,rs12046456,rs4481817,rs57784925,rs3863726,rs3766813,rs35643025,rs71647911,rs79987911,rs2306257,rs17465232,rs12026515,rs7541761,rs75976605,rs35164421,rs72885925
rs784772	0,108	rs784772
rs2927970	0,367	rs2927970
rs3862970	0,225	rs3862970
rs11803151	0,475	rs11803151
rs645998	0,183	rs645998
rs628466	0,192	rs590844,rs635614,rs628466,rs72656240,rs646065,rs3863727

rs10798929	0,433	rs10798929
rs14147677 8	0,125	rs141476778
rs2927966	0,133	rs2927966

En la tabla 3-8 se muestran los SNPs que cubren los “tag-SNPs” seleccionados en el software tagger. Los SNPs marcados con azul son los SNPs que se encontraban en la región promotora y los SNPs marcados con verde son los SNPs sinónimos en la región codificante. A pesar de encontrar los SNPs a genotipificar se decidió realizar una tercera aproximación debido a la liberación de la fase 3 del proyecto 1000 genomas el cual incluía 94 individuos.

3.2.3 Tercera aproximación tag SNP sobre el poder para población CLM

De acuerdo al proyecto de los 1000 genomas (**fase 3**) el gen *DEAR1* se extiende desde la posición 33611003 a 33647660 sobre el cromosoma 1 (ncbi 37). Basados sobre los datos de genotipos desde la población CLM (94 individuos) la región génica de *DEAR1* más la región promotora de 5kb (posición 33611003 a 33652660) comprende 230 SNPs (4 transcritos). Fuera de estos 230 SNPs, 65 SNPs tenían un MAF $\geq 5\%$. Ocho tag SNPs capturaron el 100% de los alelos de estos 65 SNPs con un coeficiente de correlación de $r^2 \geq 0.8$. Los tag SNPs seleccionados son mostrados en la tabla 3-9. Desde la segunda aproximación “tag SNP” los genotipos fueron obtenidos para 1 SNP en el grupo tagging 1 (rs59983645), para 2 SNPs en el grupo tagging 2 (rs584298) y para rs2927970 y rs599167 en el grupo tagging 3. Cuatro SNPs (rs673894, rs35622844, rs10798929 y rs2927966) no fueron genotipados. En total cuatro fuera de los 65 SNPs capturan 93.8% (61/65) de la variación genética en la región investigada considerando un $r^2 \geq 0.8$. A pesar que algunos de los tagSNPs capturaban SNPs que se encontraban en la región promotora del gen, así como SNPs que se encontraban en la región codificante sus valores de MAF eran bajos, por tal motivo no fueron incluidos dentro de la genotipificación. El uso del MAF como criterio de selección de tag SNPs se basa en que más alelos comunes confieren gran poder estadístico, por lo tanto, haciendo el enriquecimiento para alelos comunes se incrementa el poder para detectar asociaciones.

Tabla 3-9 Alelos capturados por los TagSNPs tercera aproximación

Test	Alelos capturados
rs584298	rs654336, rs685235, rs621539, rs584298, rs599167, rs689187, rs641657, rs607008, rs784773, rs631646
rs59983645	rs35111341,rs12734992,rs72656238,rs59983645,rs623596,rs3912145,rs12563712,rs12046456,rs4481817,rs57784925,rs3863726,rs3766813,rs35643025,rs71647911,rs79987911,rs2306257,rs17465232,rs12026515,rs7541761, rs75976605, rs35164421, rs72885925
rs599167	rs599161,rs631646,rs654336,rs623596,,rs685235,rs641657,rs607008,rs784773,rs621539
rs2927970	rs2927970

Luego de seleccionar los tagSNPs para ser genotipados en el total de casos y controles se procedió a realizar el análisis *in silico* tanto de los tagSNPs como de los SNPs que se encontraron en LD con estos.

Los SNPs fueron analizados y catalogados en el software FuncPred (<https://snpinf.niehs.nih.gov/snpinf/snpfunc.html>) esto con el fin de conocer si se encontraban en alguna región regulatoria como por ejemplo en sitios de unión a factores de transcripción, sitios de splicing, y sitios de unión a miRNA. Al caracterizar los tagSNPs los resultados no mostraron ningún resultado en la predicción de la función, sin embargo, al catalogar los SNPs que capturan cada uno de los tagSNP se encontró que 2 SNPs están ubicados en la región 3'UTR (rs689187, rs3766813) y en el análisis de predicción se encontraron en sitios de unión a microRNA. Al obtener este resultado se procedió a evaluar los miRNAs asociados a estos sitios de unión predichos para esto se utilizó el software Polymirts 3.0 (<http://compbio.uthsc.edu/miRSNP/>), los resultados arrojaron una lista de 12 miRNA que se unen en alguna de estas dos regiones genómicas en el 3'UTR de *DEAR1*, sin embargo hasta ahora solo se tienen resultados predictivos de la posible unión *in silico* ya que no hay evidencia funcional. En la tabla 3-10 se muestran los miRNAs predichos *in silico*.(Sun et al., 2015; Yoon *et al.*, 2011).

Tabla 3-10 miRNAs predichos como sitios de unión a los SNPs rs689187 y rs3766813. (D: el alelo derivado interrumpe un sitio miRNA conservado; C: el alelo derivado crea un nuevo sitio de unión a miRNA)

SNP	Alelo	miRNA	Función
rs689187	G	<u>hsa-miR-615-5p</u>	D
	A	<u>hsa-miR-4662a-5p</u> <u>hsa-miR-601</u>	C C
<u>rs3766813</u>	C	<u>hsa-miR-1207-5p</u>	D
		<u>hsa-miR-1909-3p</u>	D
		<u>hsa-miR-3147</u>	D
		<u>hsa-miR-4763-3p</u>	D
		<u>hsa-miR-6721-5p</u>	D
		<u>hsa-miR-6722-3p</u>	D
		<u>hsa-miR-92a-1-5p</u>	D
	T	<u>hsa-miR-4520a-3p</u> <u>hsa-miR-4520b-3p</u>	C C

Debido a que no hay análisis funcionales que puedan soportar la evidencia *in silico* sería interesante realizar la validación de estos SNPs, ya que al encontrarse en regiones regulatorias podrían estar actuando como las variantes causales asociadas con el riesgo a cáncer de seno.

3.3 Genotipificación Tag SNPs en casos y controles

3.3.1 Descripción genotipos y frecuencias alélicas

La genotipificación de los 4 Tag SNPs se realizó Empleando las condiciones de amplificación e iniciadores descritos en la metodología (2.1.5).

Se utilizaron dos estrategias de genotipificación distintas debido a la disponibilidad de los reactivos para llevar a cabo el proceso de genotipificación.

Los 4 tag-SNPs fueron analizados en 1022 casos y 1023 controles. El SNP rs59983645 no mostró amplificación en 31 muestras, el SNP rs584298 no mostro amplificación en 27 muestras, el SNP rs628466 no mostró amplificación en 35 muestras y el SNP s2927970

no mostró amplificación en 89 muestras. Los datos sobre la frecuencia de los genotipos y alelos son mostrados en la tabla 3-11.

Tabla 3-11 Frecuencias de genotipos de los 4 Tag SNPs en casos y controles

rs584298			
Casos	Frecuencias	Lab. result	1000 Genomas_fase3 (94 individuos CLM)
AA	533	53,1	
AG	396	39,5	
GG	74	7,4	
A	1462	72,9	
G	544	27,1	
Controles			
AA	565	55,9	65.0
AG	380	37,6	26.7
GG	66	6,5	8.3
A	1510	74,7	78.3
G	512	25,3	21.7
rs2927970			
Casos			
TT	389	38,5	
TC	457	45,2	
CC	164	16,2	
T	1235	61,1	
C	785	38,9	
rs2927970			
	Frecuencias	Lab. result	1000 Genomas_fase3 (94 individuos CLM)
Controles			
TT	410	40,7	37.2
TC	456	45,2	45.7
CC	142	14,1	17.0
T	1276	63,3	60.1
C	740	36,7	39.9
rs59983645			
Casos			
CC	596	59,6	
TC	350	35,0	
TT	54	5,4	
C	1542	77,1	
T	458	22,9	
Controles			
CC	630	62,4	66.0
TC	325	32,2	29.8
TT	55	5,4	4.3
C	1585	78,5	80.9

T	435	21,5	19.1
rs599167			
Casos			
TT	571	58,4	
TA	342	35,0	
AA	64	6,6	
T	1484	75,9	
A	470	24,1	
Controles			
TT	583	59,6	59.6
TA	338	34,5	33.0
AA	58	5,9	7.4
T	1504	76,8	76.1
A	454	23,2	23.9

En el anexo G se muestra la representación gráfica del ensayo KASP e iPLEX para los SNPs genotificados. Debido a que los SNPs analizados mostraron una buena calidad en la genotificación, es decir no se presentaron muchos datos faltantes, los 4 Tag SNPs fueron incluidos dentro de los análisis estadísticos.

3.3.2 Equilibrio de Hardy Weinberg

Para establecer si los genotipos presentes en la población se ajustaban al equilibrio Hardy-Weinberg, las frecuencias genotípicas observadas se compararon con las esperadas usando una prueba X^2 . Además, se realizaron los respectivos test de asociación donde se obtuvo el OR (odds ratio) para la diferencia en las frecuencias alélicas, χ^2 , OR para homocigotos y heterocigotos y un OR común. Se realizó la evaluación del coeficiente de consanguinidad (estadístico F) y el coeficiente de bondad de ajuste de Pearson. Los resultados se muestran en la tabla, 3-12, 3-13,3-14 y 3-15 y sugieren que la población se encuentra en equilibrio de Hardy Weinberg

Por lo cual este resultado apoya la homogeneidad de la población examinada, de acuerdo a los postulados de Hardy-Weinberg

Tabla 3-12 Prueba para la desviación del equilibrio de Hardy Weinberg rs584298. Si $P > 0,05$ se acepta la H_0 , concluyendo que las frecuencias observadas se ajustan a las esperadas y por tanto se encuentran en equilibrio Hardy-Weinberg.

	Análisis para la desviación del equilibrio de HW		Análisis de asociación (IC: 95% intervalo de confianza)				
	Controles	Casos	Diferencias en la frecuencia alélica	Heterocigotos	Homocigotos	Positividad del alelo	Prueba Armitage's
rs584298	n11=565 (563.82) n12=380 (382.35) n22=66 (64.82) f_a1=0.75 +/-0.010 F=0.00616 p=0.844794 (Pearson) p=0.844994 (Llr) p=0.867873 (Exact)	n11=533 (532.76) n12=396 (396.47) n22=74 (73.76) f_a1=0.73 +/-0.010 F=0.00120 p=0.969760 (Pearson) p=0.969767 (Llr) p=1.000000 (Exact)	Riesgo del alelo 2				
			[1]<->[2]	[11]<->[12]	[11+]<->[22]	[11]<->[12+22]	OR común
			Odds_ratio=1.097 C.I.=[0.954-1.263] chi2=1.68 p=0.19474 (P)	Odds_ratio=1.105 C.I.=[0.919-1.328] chi2=1.13 p=0.28863	Odds_ratio=1.189 C.I.=[0.836-1.690] chi2=0.92 p=0.33621	Odds_ratio=1.117 C.I.=[0.937-1.331] chi2=1.53 p=0.21616	Odds_ratio=1.095 chi2=1.67 p=0.19564
			Riesgo del alelo 1				
		[2]<->[1]	[22]<->[12]	[22]<->[11]	[11+12]<->[22]	OR común	
		Odds_ratio=0.911 C.I.=[0.792-1.049] chi2=1.68 p=0.19474 (P)	Odds_ratio=0.929 C.I.=[0.648-1.333] chi2=0.16 p=0.69070	Odds_ratio=0.841 C.I.=[0.592-1.197] chi2=0.92 p=0.33621	Odds_ratio=0.877 C.I.=[0.622-1.237] chi2=0.56 p=0.45346	Odds_ratio=0.913 chi2=1.67 p=0.19564	

Tabla 3-13 Prueba para la desviación del equilibrio de Hardy Weinberg rs2927970. Si $P > 0,05$ se acepta la H_0 , concluyendo que las frecuencias observadas se ajustan a las esperadas y por tanto se encuentran en equilibrio Hardy-Weinberg.

	Análisis para la desviación del equilibrio de HW		Análisis de asociación (IC: 95% intervalo de confianza)				
	Controles	Casos	Diferencias en la frecuencia alélica	Heterocigotos	Homocigotos	Positividad del alelo	Prueba Armitage's
rs2927970	n11=410 (403.81)	n11=389 (377.53)	Riesgo del alelo 2				
	n12=456 (468.37)	n12=457 (479.94)					
	n22=142 (135.81)	n22=164 (152.53)					
	f_a1=0.63 +/-0.011	f_a1=0.61 +/-0.011	[1]<->[2]	[11]<->[12]	[11+]<->[22]	[11]<->[12+22]	OR común
	F=0.02642	F=0.04779	Odds_ratio=1.096	Odds_ratio=1.056	Odds_ratio=1.217	Odds_ratio=1.095	Odds_ratio=1.097
	p=0.401630 (Pearson)	p=0.128784 (Pearson)	C.I.=[0.965-1.245]	C.I.=[0.874-1.277]	C.I.=[0.935-1.585]	C.I.=[0.916-1.308]	
p=0.402215 (Llr)	p=0.129317 (Llr)	chi2=1.99	chi2=0.32	chi2=2.13	chi2=0.98	chi2=1.92	
p=0.415962 (Exact)	p=0.128133 (Exact)	p=0.15799 (P)	p=0.57195	p=0.14418	p=0.32123	p=0.16576	
			Riesgo del alelo 1				
			[2]<->[1]	[22]<->[12]	[22]<->[11]	[11+12]<->[22]	OR común
			Odds_ratio=0.912	Odds_ratio=0.868	Odds_ratio=0.822	Odds_ratio=0.846	Odds_ratio=0.912
			C.I.=[0.803-1.036]	C.I.=[0.669-1.125]	C.I.=[0.631-1.070]	C.I.=[0.663-1.079]	
			chi2=1.99	chi2=1.15	chi2=2.13	chi2=1.81	chi2=1.92
			p=0.15799 (P)	p=0.28371	p=0.14418	p=0.17811	p=0.16576

Tabla 3-14 Prueba para la desviación del equilibrio de Hardy Weinberg rs59983645. Si $P > 0,05$ se acepta la H_0 , concluyendo que las frecuencias observadas se ajustan a las esperadas y por tanto se encuentran en equilibrio Hardy-Weinberg.

	Análisis para la desviación del equilibrio de HW		Análisis de asociación (IC: 95% intervalo de confianza)				
	Controles	Casos	Diferencias en la frecuencia alélica	Heterocigotos	Homocigotos	Positividad del alelo	Prueba Armitage's
rs59983645	n11=630 (621.84)	n11=596 (594.44)	Riesgo del alelo 2				
	n12=325 (341.32)	n12=350 (353.12)	[1]<->[2]	[11]<->[12]	[11+]<->[22]	[11]<->[12+22]	OR común
	n22=55 (46.84)	n22=54 (52.44)	Odds_ratio=1.082	Odds_ratio=1.138	Odds_ratio=1.038	Odds_ratio=1.124	Odds_ratio=1.061
	f_a1=0.78 +/-0.009	f_a1=0.77 +/-0.009	C.I.=[0.933-1.256]	C.I.=[0.943-1.374]	C.I.=[0.701-1.536]	C.I.=[0.939-1.345]	chi2=1.05
F=0.04783	F=0.00883	chi2=1.08	chi2=1.83	chi2=0.03	chi2=1.63	chi2=1.05	
p=0.128526 (Pearson)	p=0.780071 (Pearson)	p=0.29776 (P)	p=0.17657	p=0.85265	p=0.20200	p=0.30449	
p=0.133925 (Llr)	p=0.780587 (Llr)	Riesgo del alelo 1					
p=0.135954 (Exact)	p=0.788221 (Exact)	[2]<->[1]	[22]<->[12]	[22]<->[11]	[11+12]<->[22]	OR común	
		Odds_ratio=0.924	Odds_ratio=1.097	Odds_ratio=0.964	Odds_ratio=1.009	Odds_ratio=0.942	
		C.I.=[0.796-1.072]	C.I.=[0.732-1.644]	C.I.=[0.651-1.426]	C.I.=[0.686-1.484]	chi2=1.05	
		chi2=1.08	chi2=0.20	chi2=0.03	chi2=0.00	chi2=1.05	
		p=0.29776 (P)	p=0.65424	p=0.85265	p=0.96404	p=0.30449	

Tabla 3-15 Prueba para la desviación del equilibrio de Hardy Weinberg rs599167. Si $P > 0,05$ se acepta la H_0 , concluyendo que las frecuencias observadas se ajustan a las esperadas y por tanto se encuentran en equilibrio Hardy-Weinberg.

	Análisis para la desviación del equilibrio de HW		Análisis de asociación (IC: 95% intervalo de confianza)				
	Controles	Casos	Diferencias en la frecuencia alélica	Heterocigotos	Homocigotos	Positividad del alelo	Prueba Armitage's
rs599167	n11=583 (577.63)	n11=571 (563.53)	Riesgo del alelo 2				
	n12=338 (348.73)	n12=342 (356.95)					
	n22=58 (52.63)	n22=64 (56.53)					
	f_a1=0.77 +/-0.010	f_a1=0.76 +/-0.010	[1]<->[2]	[11]<->[12]	[11+]<->[22]	[11]<->[12+22]	OR común
	F=0.03077	F=0.04188	Odds_ratio=1.049	Odds_ratio=1.033	Odds_ratio=1.127	Odds_ratio=1.047	Odds_ratio=1.052
	p=0.335627 (Pearson)	p=0.190496 (Pearson)	C.I.=[0.905- 1.216]	C.I.=[0.855- 1.249]	C.I.=[0.775- 1.637]	C.I.=[0.874- 1.254]	
	p=0.339410 (Llr)	p=0.194743 (Llr)	chi2=0.41	chi2=0.11	chi2=0.39	chi2=0.25	chi2=0.39
	p=0.324788 (Exact)	p=0.189480 (Exact)	p=0.52358 (P)	p=0.73629	p=0.53143	p=0.61889	p=0.53098
	Riesgo del alelo 1						
			[2]<->[1]	[22]<->[12]	[22]<->[11]	[11+12]<->[22]	OR común
		Odds_ratio=0.953	Odds_ratio=0.917	Odds_ratio=0.888	Odds_ratio=0.898	Odds_ratio=0.951	
		C.I.=[0.822- 1.105]	C.I.=[0.623- 1.349]	C.I.=[0.611- 1.290]	C.I.=[0.622- 1.297]		
		chi2=0.41	chi2=0.19	chi2=0.39	chi2=0.33	chi2=0.39	
		p=0.52358 (P)	p=0.65965	p=0.53143	p=0.56688	p=0.53098	

3.3.3 Análisis de asociación de los genotipos y las variables epidemiológicas con cáncer de seno

Los análisis de asociación se realizaron de acuerdo a los parámetros establecidos en la metodología. Antes de realizar el análisis por regresión logística se realizaron los análisis de interacción y confusión para cada una de las variables, así como el análisis de colinealidad.

El análisis de multicolinealidad se realizó incluyendo todas las covariables, así como los 12 genotipos (3 genotipos para cada tagSNP), en este análisis las covariables que presentaron multicolinealidad fueron aquellas que presentaron un índice mayor a 10 entre estas se encontraron:

- IMC: mostro colinealidad con la edad y la covariable menopausia
- Edad: mostro colinealidad con menopausia y Terapia de reemplazo hormonal
- Los genotipos en los tag SNPs 2927970,599167 así como el genotipo TT y TC del tagSNP rs59983645 mostraron colinealidad entre sí.

Las covariables que mostraron colinealidad fueron retiradas del modelo de regresión. Los datos sobre el análisis de colinealidad son mostrados en el anexo H.

Se realizó la selección de variables para el modelo final usando la aproximación "Stepwise", el modelo final incluyo historia familiar, menopausia, terapia de reemplazo hormonal, amamantar y el SNP rs584298 (genotipos AG y GG), los análisis incluyeron 1963 observaciones. Los valores de OR, así como los valores de probabilidad de las variables incluidas en el modelo final son mostrados en la tabla 3-16.

Tabla 3-16 Estadísticos del análisis Stepwise para el modelo final de regresión

Variables	OR	IC- [intervalo de confianza 95%]	Valor de probabilidad
Historia Familiar	4.54	[2,84-7,26]	P<0.0001
Menopausia	2,28	[1,39-3,74]	P< 0.0001
BMI	0,99	[0.997-0,999]	P=0.004
rs584298 AG	1.41	[1,00-2,00]	P=0.004
rs584298 GG	2,57	[1,35-4,90]	P=0.048

En el análisis del primer modelo se retiraron covariables que podrían aportar peso al análisis de regresión debido a la colinealidad. En el análisis de colinealidad se pueden

realizar ciertas modificaciones con el fin de no excluir covariables importantes estas modificaciones pueden ser: intentar eliminar la variable menos necesaria implicada en la colinealidad, a riesgo de obtener un modelo menos válido. Sin embargo, un problema común es no saber qué variable debemos omitir. Cualquiera de las variables problemáticas puede ser omitida, no hay fundamentos estadísticos para suprimir una variable en vez de otra.

Se recomienda que, si eliminamos una variable predictora, ésta se reemplace por otra igualmente importante que no tenga una colinealidad tan fuerte.

Se puede intentar cambiar la escala de medida de la variable en conflicto (es decir, transformarla). Sin embargo, estas transformaciones hacen al modelo muy dependiente de los datos actuales, invalidando su capacidad predictiva.

También se puede recurrir a aumentar la muestra para así aumentar la información en el modelo y ver si la multicolinealidad puede disminuir, aunque no siempre será posible.

Debido a que ninguna de las posibles soluciones aplicaba para nuestro caso, se decidió retirar estas covariables del modelo.

En el modelo final se puede observar como las covariables incluidas tienen una asociación significativa con la variable dependiente (cáncer de seno). El genotipo GG y AG en el tagSNP rs584298 mostro una asociación estadísticamente significativa con un $OR > 1$ por lo cual se puede inferir que este genotipo podría ser un factor de riesgo en cáncer de seno ajustado por las covariables de riesgo epidemiológico. El modelo final con sus correspondientes estadísticos es mostrado en el anexo J

3.3.4 Análisis de asociación de los genotipos con las variables histopatológicas del tumor

Para evaluar la asociación entre los genotipos de *DEAR1* con algunas características específicas tumorales nosotros realizamos un análisis de regresión logística condicional entre los genotipos de *DEAR1* y 5 características histopatológicas del tumor (histología, grado histológico, receptores de estrógeno, receptores de progesterona, y receptor HER2). Los resultados no mostraron ningún tipo de asociación entre los genotipos de *DEAR1* con: Histología (868 observaciones incluidas), grado histológico, ER (694 observaciones incluidas), PR (689 observaciones incluidas) y HER2 (552 observaciones incluidas) $P \geq 0.05$. Este resultado sugiere que en principio la muestra con la que se realizó el análisis de regresión no es suficiente para detectar asociaciones positivas y por otro lado sugiere que variantes en otros genes así como la evaluación de variantes funcionales en *DEAR1* es necesaria para la asociación con parámetros histológicos de la enfermedad ya que hay evidencia que en el desarrollo tumoral desde DCIS hasta carcinoma invasivo se deben presentar diferentes mutaciones a lo largo de genes conductores e incluso dentro del mismo gen *DEAR1* (Chen *et al.*,2013).

Para evaluar si los genotipos en el gen *DEAR1* se asociaron con algunas características específicas del tumor, se realizó una regresión logística multinomial entre los genotipos *DEAR1* y cinco características clínicas e histopatológicas del tumor (histología, grado histológico, estado de los receptores de estrógenos (ER), receptores de progesterona estado (PR) y el estado del receptor 2 del factor de crecimiento epidérmico humano (HER2). Los resultados muestran la asociación entre el tagSNP rs584298 genotipo AA con la expresión del receptor HER2 ($p = 0.048$) y el genotipo AG con expresión del receptor PR ($p = 0.015$) por lo tanto, si un paciente tiene expresión de HER2, el riesgo relativo de tener el genotipo AA sobre GG (grupo de referencia) aumentaría el riesgo en un factor de 3.48 veces cuando las otras variables en el modelo se mantienen constantes. La asociación con el estado de PR también aumentará, si un paciente tiene una expresión de PR, el riesgo relativo de tener el genotipo AG sobre GG (grupo de referencia) se espera que aumente en un factor de 2.78 veces cuando las otras variables en el modelo se mantienen constantes.

Tabla 3-17 Asociación rs584298 y el estado PR y HER2

TagSNP	RRR	Std. Err.	z	P> z	[95% CI]	
rs584298						
A						
PR Status	1.758	0.706	1.40	0.161	0.79	3.86
HER2 status	3.48	2.20	1.98	0.048	1.00	12.01
AG						
PR status	2.78	1.17	2.43	0.015	1.22	6.36
HER2 status	3.10	2.00	1.76	0.078	0.879	10.99
G	Base outcome					

Ajustado por tamaño tumoral y estado de los nodulos linfoides

Las Variables histopatológicas tamaño tumoral y estado de los nódulos linfoides, son consideradas en la práctica clínica como variables de pronóstico para la enfermedad.

Se ha descrito una relación directa entre el estado de los ganglios linfáticos axilares y el riesgo de recurrencia a distancia. La sobrevida a 5 años en pacientes sin compromiso ganglionar es del 82,8% comparado al 73% para pacientes con 1 a 3 ganglios positivos, 45,7% para 4 a 12 ganglios positivos y del 28,4% para pacientes con más de 13 ganglios positivos (González Ortega, Morales Wong, López Cuevas, & Díaz Valdéz, 2011).

3.3.5 Análisis de supervivencia

Para el análisis de sobrevida los tres puntos finales definidos fueron: Tiempo a la recurrencia, Sobrevida específica de cáncer de seno:(SS) y Sobrevida libre total (OSS) como se mencionó en la metodología (2.4)

Los múltiples análisis de Sobrevida con selección de variables paso a paso fueron llevados a cabo usando la distribución Weibull.

En el anexo I se muestran los análisis de sobrevida para cada uno de los puntos finales.

Los resultados revelaron que ninguno de los genotipos de *DEAR1* tiene asociación significativa con el tiempo de recurrencia de cáncer de seno, muerte por cáncer de seno o muerte por cualquier otra causa.

Este resultado sugiere que las observaciones incluidas para el análisis no son significativas para encontrar cualquier tipo de asociación, principalmente por el número de datos censurados y el número de fallas (muerte o recurrencia) que se presentaron dentro del tiempo de seguimiento, además de los datos que fueron retirados debido a fallas en el tiempo de censura.

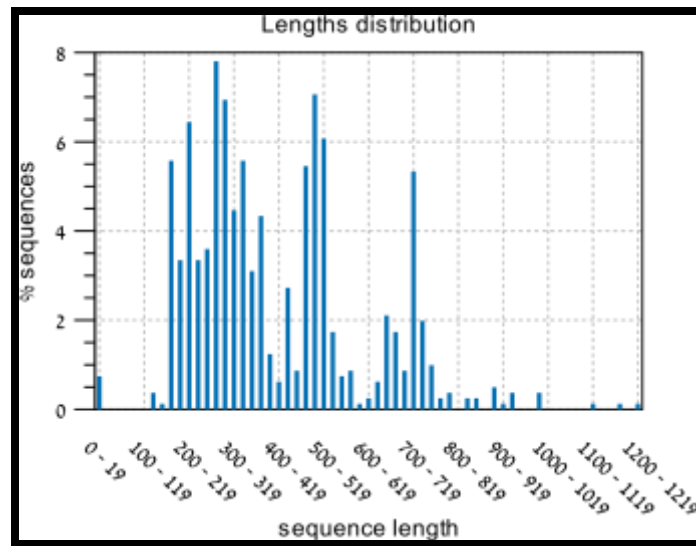
3.4 Secuenciación de los tejidos embebidos en parafina

3.4.1 Análisis de las secuencias obtenidas

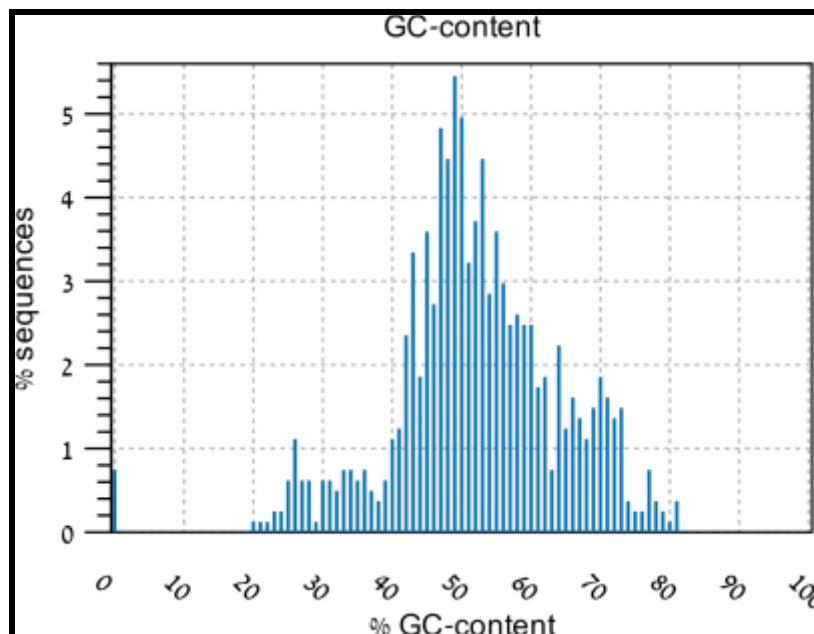
Al servicio de secuenciación de MacroGen se enviaron 60 productos de PCR correspondientes a los tres grupos de pacientes incluidos en el estudio, la secuenciación se realizó en sentido bidireccional por duplicado para cada una de las muestras incluidas.

Se obtuvieron como resultado 1200 lecturas de secuencias de las cuales 393 no fueron ensambladas de acuerdo a los parámetros establecidos, por lo cual en el análisis solo se incluyeron 807 lecturas de secuencias.

Los resultados de control de calidad de las lecturas son mostrados en las figuras 3-1 a 3-5.

Figura 3-1 Longitud de distribución de las secuencias

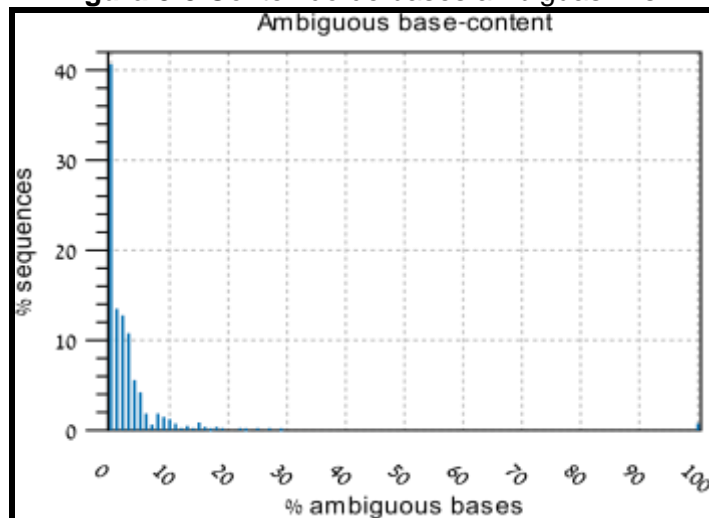
X: longitud de la secuencia en pares de bases Y: número de secuencias con una característica de longitud particular normalizado a el total de secuencias. Aquí podemos observar como la mayoría de las secuencias analizadas solo lograron tener una longitud de distribución de hasta 1000 pares de bases.

Figura 3-2 Porcentaje de GC

En el eje de las X se muestra el contenido relativo de GC de una secuencia y en el eje Y se muestra el número de secuencias con la característica en particular. Los porcentajes

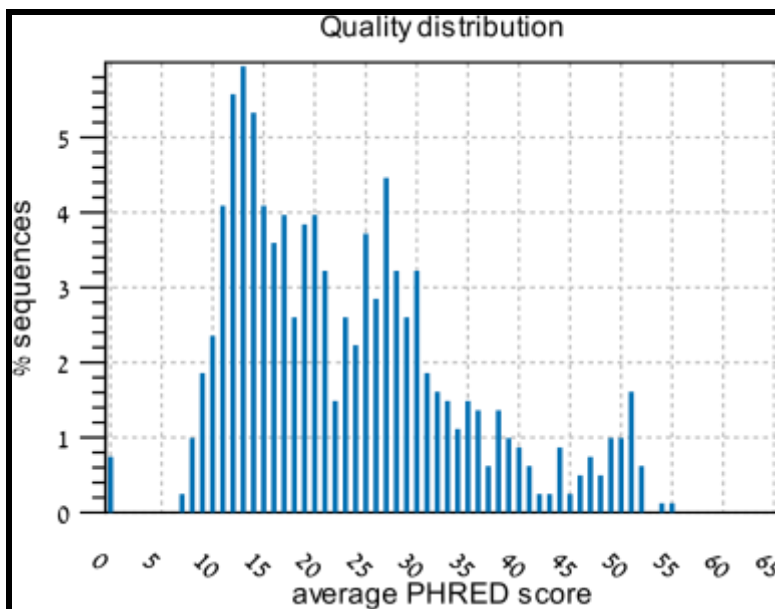
de GC son normalizados al total del número de secuencias. Este resultado muestra como el contenido de GC fue de 40 al 60% en el 5% de las secuencias.

Figura 3-3 Contenido de bases ambiguas "Ns"



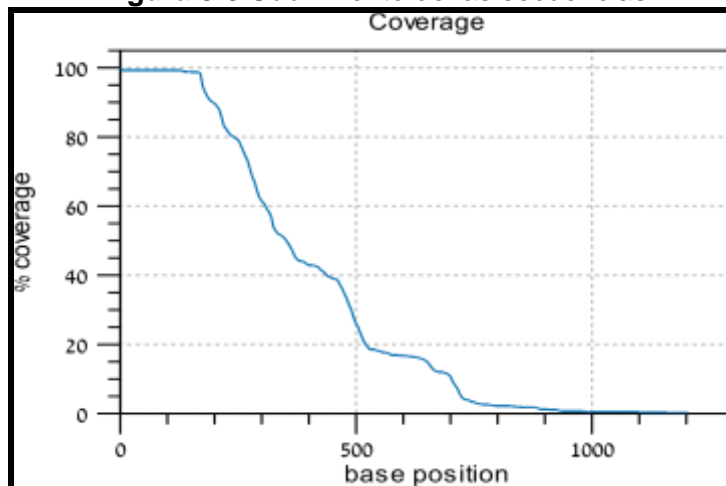
En el eje de las X se muestra el contenido relativo de Ns de una secuencia y en el eje Y se muestra el número de secuencias con nucleótidos ambiguos. Aquí podemos observar que el contenido de bases ambiguas en nuestras secuencias es mínimo entre un 0-10% en el 40% de secuencias.

Figura 3-4 Distribución de la calidad PHRED score



En el eje X se muestra el puntaje PHRED y en el eje Y se muestra el número de secuencias observadas, el puntaje es normalizado a el número total de las secuencias. Este resultado muestra como la calidad de las secuencias tiene un puntaje PHRED entre 10 y 50 lo que significa que las bases muestran una precisión entre el 90% y 99.99% de ser correctas en al menos 5% de las secuencias.

Figura 3-5 Cubrimiento de las secuencias



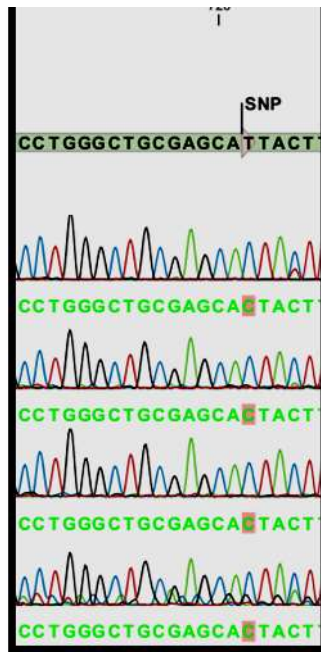
El cubrimiento muestra el número de secuencias que cubren la posición individual de una base, en el eje X está la posición de la base y en el eje Y el número de secuencias cubriendo las posiciones individuales de cada base normalizadas a el número total de secuencias, Aquí podemos observar una total correlación entre el resultado de longitud de la distribución de secuencias ya que la longitud fue solo de hasta 1000 pares de bases. Podemos observar que el cubrimiento es mayor en las primeras 500 pares de bases.

El análisis de variantes mostró la presencia de una variante tipo SNP en la posición 722 de la secuencia de referencia NM018207.2 la cual se encuentra anotada en la base de datos del NCBI rs622407 esta variante es una variante sinónima CAT>CAC codifica para el aminoácido histidina, por lo cual no se presenta ningún cambio estructural ni funcional a nivel de la proteína. Los resultados del SNP identificado sobre la secuencia son mostrados en la figura 3-6.

El análisis de variantes de baja frecuencia mostro 35 variantes de diversos tipos (SNV y/o deleción y/o inserción), según los parámetros mencionados en la metodología. Sin embargo, se aplicó un filtro adicional (filtro de llamada de variantes marginales) con el fin de remover variantes de baja calidad.

El polimorfismo identificado se encuentra en el dominio RING del gen *TRIM62*, dominio que está involucrado directamente en la transferencia de ubiquitin desde la E2 a la E3 en la cascada de ubiquitinación.

Figura 3-2 SNP en la posición 722



La frecuencia genotípica para el SNP encontrado en la población de estudio es mostrada en la tabla 3-18.

Tabla 3-18: Frecuencias genotípicas rs622407 en muestras de tejido tumoral

Muestras		Frecuencias Genotípicas			Total
		CC	TC	TT	
Casos	Receptor Triple negativo	24	3	4	31
	Receptor triple positivo	4	0	1	5
Controles		15	6	0	21
Total de observaciones		43	9	5	57

Según este resultado en la muestra evaluada hay una disminución en la presencia del alelo silvestre (T) tanto en los casos como en los controles, lo cual se puede deber en parte al tamaño de la muestra. De igual manera al realizar el análisis de equilibrio de Hardy- Weinberg se observa que la población de estudio (casos) no se encuentra en equilibrio de Hardy- Weinberg lo cual es de esperarse debido a que son muestras tumorales. Los resultados del análisis de Hardy Weinberg son mostrados en la tabla 3-19.

Como se puede observar en la tabla 3-19, el alelo 2 (C) está en alta frecuencia tanto en los casos como en los controles. Sin embargo, La N puede ser un factor que influye en la desviación de Hardy Weinberg. También se observó una deficiencia de heterocigotos en los casos, por lo cual esto también puede influir en la desviación significativa del equilibrio de Hardy-Weinberg.

Al revisar la literatura no se encontró un registro en que el que se describa la distribución de este polimorfismo en pacientes con cáncer de seno, este sería el primer reporte.

Los resultados de este estudio sugieren realizar un análisis en muestras pareadas (normal y tumoral) para determinar si la condición tumoral influye sobre los cambios en la distribución de este polimorfismo detectados en los pacientes con cáncer de seno.

Si bien como parte de nuestro análisis no encontramos alguna variante patogénica, debido en parte a la baja calidad del ADN extraído desde los bloques tumorales, al intentar realizar una asociación de la variante identificada con los parámetros histopatológicos tumorales encontramos que los casos no se encontraban en equilibrio lo cual es interesante ya que este polimorfismo podría estar asociado con la enfermedad.

Los análisis de asociación del polimorfismo con diferentes parámetros histológicos no mostraron ningún resultado estadísticamente significativo (datos no mostrados), esto debido al tamaño de la muestra , así como al procedimiento de inclusión en su mayoría de casos triple negativos por lo cual no hubo homogeneidad en la selección de los mismos.

Tabla 3-19 Prueba de desviación del equilibrio de Hardy-Weinberg para los tejidos embebidos en parafina

SNP	Análisis para la desviación del equilibrio de Hardy Weinberg		Análisis de asociación (IC: 95% intervalo de confianza)				
	Controles	Casos	Diferencias en la frecuencia alélica	Heterocigotos	Homocigotos	Positividad del alelo	Prueba Armitage's
Snp1722	n11=0 (0.43) n12=6 (5.14) n22=15 (15.43) f_a1=0.14 +/- 0.049 F=-0.16667 p=0.445009 (Pearson) p=0.316180 (Llr) p=1.000000 (Exact)	n11=5 (1.21) n12=3 (10.59) n22=27 (23.21) f_a1=0.19 +/- 0.061 F=0.71660 p=0.000022 (Pearson) p=0.000118 (Llr) p=0.000224 (Exact)	Riesgo del alelo 2				
			[1]<->[2]	[11]<->[12]	[11+]<->[22]	[11]<->[12+22]	OR común
			Odds ratio=0.731 C.I.=[0.255-2.096] chi2=0.34 p=0.55852 (P)	Odds ratio=0.049 C.I.=[0.002-1.169] chi2=5.83 p=0.01573	Odds ratio=0.161 C.I.=[0.008-3.116] chi2=2.62 p=0.10534	Odds ratio=0.129 C.I.=[0.007-2.457] chi2=3.29 p=0.06953	Odds ratio=0.761 chi2=0.24 p=0.62462
			Riesgo del alelo 1				
			[2]<->[1]	[22]<->[12]	[22]<->[11]	[11+12]<->[22]	OR común
			Odds ratio=1.368 C.I.=[0.477-3.924] chi2=0.34 p=0.55852 (P)	Odds ratio=0.278 C.I.=[0.061-1.274] chi2=2.93 p=0.08686	Odds ratio=6.200 C.I.=[0.321-119.795] chi2=2.62 p=0.10534	Odds ratio=0.741 C.I.=[0.216-2.540] chi2=0.23 p=0.63259	Odds ratio=2.962 chi2=0.24 p=0.62462

4. Consideraciones finales

La pregunta de investigación que se planteó en este trabajo estuvo dirigida a conocer si las variantes genéticas en *DEAR1* estaban asociadas con cáncer de seno en una muestra de población Colombiana y podrían ser útiles como marcadores pronósticos de la enfermedad. Para dar respuesta a esta pregunta utilizamos una metodología de Tag SNPs, con el fin de seleccionar aquellos polimorfismos que podrían ser causales o que podrían estar en desequilibrio de ligamiento con los SNP causales involucrados en el desenlace clínico que en este caso fue cáncer de seno.

La muestra de población Colombiana provino de una cohorte de pacientes con cáncer de seno no seleccionados, que incluyó 1022 casos y 1023 controles. Información sobre factores de riesgo epidemiológicos para cáncer de seno, características histopatológicas del tumor, así como información de seguimiento y sobrevida fue incluida en el estudio. La información de riesgo epidemiológico incluyo variables que se sabe están asociadas con riesgo de cáncer de seno para estas variables la información que se encontró disponible para los análisis estadísticos fue de un 90 a 98%. La información sobre características histopatológicas del tumor fue recolectada a través de los reportes de patología de las pacientes sin embargo muchas de las pacientes que ingresaron al estudio eran pacientes recién diagnosticadas razón por la cual mucha de la información acerca de la histopatología no se encontró disponible para los análisis estadísticos en este estudio. En las pacientes que no tenían su reporte de patología y que no se acordaban de la misma se les solicito a sus médicos tratantes copia de la historia clínica, la información disponible sobre la histopatología tumoral fue de un 30% a un 47%. Por último, el seguimiento de las pacientes se realizó a los 5 años desde la fecha de ingreso al estudio, se logró recolectar un 70% de la información requerida para llevar a cabo los análisis de sobrevida.

Para la selección de los tagSNPs se tuvieron en cuenta los datos del genoma desde el proyecto HapMap población CEU y del proyecto de los 1000 genomas población CLM. Al

realizar la primera aproximación tagSNP entre las dos poblaciones se observó que los patrones de ligamiento de *DEAR1* diferían entre las dos poblaciones (3 Tag SNPs capturan 26 alelos) y la población CLM desde el proyecto de los 1000 genomas (14 Tag SNPs capturan 64 alelos). Teniendo esto en cuenta se evaluaron solo tag SNPs desde la población CLM del proyecto de los 1000 genomas ya que se asumió que esta población podría ser más similar a la población de estudio.

En el proyecto de los 1000 genomas la población CLM reveló que el número de variantes por genoma es de 4 a 4.6 millones de variantes siendo en su mayoría variantes compartidas a través de todos los continentes y sólo una pequeña proporción de variantes representativas únicas de la población, que en su mayoría corresponden a variantes raras ($MAF < 0.5\%$) (Auton et al., 2015). Además, este estudio reveló similitudes genéticas con los genomas europeos, africanos y nativos americanos. Los genomas desde Medellín mostraron predominantemente ascendencia europea 74.6%, 18.1% Asia (nativo-americanos) y 7.3% ascendencia africana. Sin embargo, los individuos de la población CLM varían ampliamente con respecto a sus proporciones ancestrales desde estos tres grupos ancestrales.

Es claro que diferentes regiones de Colombia muestran diferentes patrones demográficos, sin embargo de acuerdo al estudio llevado a cabo por Ruiz-Linares (2014) (Ruiz-Linares et al., 2014) se evidencia como en la muestra de población Colombiana la ascendencia africana es elevada en las regiones costeras particularmente sobre el pacífico y alta ascendencia Europea en las áreas centrales, mientras la ascendencia Amerindia es alta en el sureste y este del país (amazonia) (Ruiz-Linares et al., 2014). Por lo cual primero existe una alta correlación entre los resultados mostrados por el proyecto de los 1000 genomas y los resultados mostrados por Ruiz-Linares donde se observa claramente el mestizaje desde los tres cursos de poblaciones ancestrales y segundo la selección de los tagSNP desde la población CLM debería cubrir las variantes de la población colombiana en estudio, ya que esta provenía principalmente desde las regiones centrales del país (Bogotá, Neiva y Villavicencio). Además de esto según el estudio del proyecto mil genomas se observa como la diversidad en la muestra CLM es baja comparada con otras poblaciones incluidas en el proyecto de los 1000 genomas (Auton et al., 2015).

En el análisis de equilibrio de Hardy-Weinberg se rechazó la hipótesis nula por lo cual se encontró que la población de estudio tanto los casos como controles se encontraban en

equilibrio de Hardy- Weinberg es decir que no se evidenciaron procesos de migraciones recientes, selección, mutación y/o subestructura en la población. Lo cual apoya la homogeneidad de las poblaciones examinadas.

La asociación encontrada entre el genotipo homocigoto para el alelo G y el heterocigoto AG del SNP rs584298 con cáncer de seno ajustada para las otras covariables de riesgo mostro una probabilidad de asociación con cáncer de seno. En los estudios de asociación genética es importante genotipificar aquellos SNPs analizados en la fase de identificación que muestren asociación con la enfermedad en una cohorte de estudio diferente con el fin de realizar la validación del mismo. Otra razón por la cual se pudo obtener esta asociación modesta es que el verdadero SNP causal no fue genotipado por lo cual en estos casos donde el SNP genotipado está en desequilibrio de ligamiento con el SNP causal se mostraran solo efectos moderados ya que cada SNP genotipado sirve como un sustituto imperfecto para el SNP causal (Wu *et al.*, 2010).

Otros tagSNP no mostraron asociación debido a que mostraron colinealidad entre sí en el primer modelo de regresión evaluado por lo cual tuvieron que ser retiradas del modelo final de regresión.

En el segundo modelo de regresión donde se evaluó la asociación entre genotipos de *DEAR1* y variables histopatológicas, la asociación encontrada en el SNP rs584298 para homocigotos del alelo A y heterocigotos AG mostró una asociación positiva. Sin embargo, como se explicó anteriormente para asegurar que este SNP genotipificado tiene una asociación real con la expresión de los marcadores “HER2 y PR” es necesario realizar un estudio de replicación.

Tras distintos intentos de descifrar la base genética de enfermedades complejas, actualmente la investigación en este campo se centra en dos hipótesis de trabajo:

-Hipótesis de enfermedad común -variaciones comunes. Postula que las enfermedades complejas estarían producidas por el efecto acumulativo de múltiples variantes genéticas comunes en la población, cada una de ellas con un efecto pequeño.

- Hipótesis de enfermedad común - múltiples variantes raras. Defiende que estas enfermedades estarían causadas por variaciones genéticas poco frecuentes, pero con

mucha penetrancia. Estas variaciones podrían ser mutaciones puntuales, variaciones en el número de copia (copy number variants o CNVs), o inserciones o deleciones de mayor longitud. Es probable que ambas hipótesis sean ciertas, contribuyendo las dos en cierta medida a constituir la realidad genética de las enfermedades complejas. (Frazer, Murray, Schork, & Topol, 2009).

En este estudio y como parte del consorcio BCAC decidimos probar la primera hipótesis donde se seleccionaron aquellas variantes que tuvieran un $MAF > 0.05$ y que pudieran actuar como un factor de riesgo de baja penetrancia en cáncer de seno. Se seleccionaron 4 TagSNPs que mostraron altos valores MAF en dos aproximaciones diferentes y que cubrían más del 90% de los SNPs para la región analizada. Este se escogió como criterio de inclusión para la selección de los tagSNPs ya que la selección de alelos comunes ($MAF > 0.05$) confieren gran poder estadístico incrementando de esta manera el poder para detectar asociaciones. En algunos estudios es realmente arriesgado tomar como único criterio de selección de los tagSNP el MAF ya que las frecuencias alélicas pueden no estar bien establecidas o pueden estar basadas sobre una población de referencia que difiere desde la población de estudio.

La utilidad de las aproximaciones tagSNP ha sido de gran ayuda para la búsqueda de variantes asociadas con enfermedades complejas, sin embargo, este tipo de aproximaciones presentan algunas desventajas, una de ellas es que se requiere una gran cantidad de muestras para encontrar algún tipo de asociación que en la mayoría de los casos es modesta y por lo tanto requerirá una replicación en otra cohorte de estudio para asegurar que si es la variante real asociada a la enfermedad. Segundo no se puede conocer la variante causal ya que en la mayoría de los casos se evalúan variantes en desequilibrio de ligamiento con la variante causal.

Ya que las enfermedades complejas como el cáncer son enfermedades multifactoriales, se podrían encontrar asociaciones robustas si se incluyen aquellos SNPs que ya se encuentren validados en otras poblaciones, que se encuentren en regiones regulatorias del gen que puedan afectar la estabilidad de la proteína. Además de evaluar varios SNPs en diferentes genes candidatos ya que las variantes en un solo gen pueden ser un factor

débil de asociación con la enfermedad pero que sumados a otras variantes de baja penetrancia pueden generar un efecto sumativo de riesgo.

En cuanto al uso de tejidos embebidos en parafina para análisis genómicos es claro que son una herramienta útil para la descripción de variantes genéticas asociadas con enfermedad, la utilidad de los datos obtenidos a partir del análisis de tejidos está directamente relacionada con la calidad de la muestra, la cual se puede ver afectada por las condiciones de manejo y almacenamiento que pueden contribuir a su deterioro o degradación. Existen algunos factores, incluyendo el tipo de tejido, la hipoxia antes de la escisión, el método de preservación, el tipo y tiempo de almacenamiento, la congelación y descongelación y los métodos de extracción, que afectan la calidad de las biomoléculas (Berger *et al.*, 2001). En nuestro estudio utilizamos muestras incluidas en parafina, según los protocolos por el laboratorio de patología el fijador utilizado fue formalina al 10% el papel de la formalina es producir entrecruzamiento con las proteínas, principalmente histonas que ayudan a mantener el empaquetamiento del DNA en los cromosomas, de esta forma previene que la doble hélice de DNA se libere del complejo. La formalina penetra el tejido en una proporción de 0,5 mm/h, por lo tanto, entre mayor tiempo permanezca el tejido en el fijador, mayor será el entrecruzamiento y el DNA se puede fragmentar ([REN Z.P., 2000](#)), o el entrecruzamiento tan fuerte impide su extracción. El tamaño de los fragmentos producidos por el efecto de la fijación depende del tipo de tejido y las condiciones del proceso, la mayoría de los estudios reportan la generación de pequeños fragmentos de entre 200 a 500 bp ([Coombs, Gough, & Primrose, 1999](#)). En nuestro caso la amplificación se pudo hacer en fragmentos de hasta 300 bp, la calidad del ADN obtenido no fue muy bueno en parte porque fueron muestras con más de 6 a 8 años de almacenamiento y la cantidad de tejido tumoral en las muestras fue mínimo. Además, debido a la manipulación de los mismos en formalina y luego incluidos en parafina nos hace suponer que el DNA en la mayoría de las muestras se encontraba fragmentado.

A pesar de que las muestras patológicas de archivo son de incalculable valor para hacer estudios moleculares retrospectivos, para este tipo de estudios donde se requiere una excelente calidad del ADN para obtener secuencias limpias que puedan proporcionar información sobre nuevas variantes genómicas, sería mejor incluir tejidos frescos donde no solo se puedan realizar análisis genómicos sino que también se puedan hacer comparaciones con datos proteómicos que permitan abordar la enfermedad como un todo.

5. Conclusiones y recomendaciones

5.1 Conclusiones

- Se analizaron 4 SNPs que abarcan la región 1p35.1 en la que se ha informado LOH y asociación con el desarrollo de cáncer de mama y la invasión. No se observaron diferencias de frecuencia significativa en 4 SNPs entre los casos de cáncer de mama y los controles. Este estudio sobre variantes genéticas en el gen *DEAR1* y su asociación con el cáncer de mama en la población colombiana a través de un enfoque de Tag-SNP es el primer estudio informado
- Se encontró una asociación marginal entre el genotipo AA y AG del rs584298 mas las covariables de menopausia, terapia de reemplazo hormonal e historia familiar con cáncer de seno.
- Se encontró una asociación significativa entre las variables histopatológicas HER 2 y PR con el TagSNP rs584298. Este SNP se encuentra en una región intrónica en *DEAR1* y hasta ahora no tenía una importancia clínica.
- Dos SNPs en DL con los tagSNP rs584298 y rs59983645, por análisis de predicción se encontraron en la región 3'UTR de *DEAR1* como sitios de unión a miRNA.
- El efecto conocido de este gen sobre el cáncer de mama en el análisis funcional llevado a cabo por (Lott et al., 2009b) aumenta la probabilidad de que una o muchas variantes puedan estar asociadas con el riesgo de cáncer de mama.
- La falta de asociaciones robustas con cáncer de mama puede atribuirse a varios factores. Primero, los SNPs evaluados no son las variantes causales más bien éstos podrían estar en desequilibrio del ligamiento (LD) con otras variantes

asociadas a riesgo; Debido a la heterogeneidad de la población, la naturaleza de LD puede variar de población a población y los SNP seleccionados en este estudio representan sólo una población en Colombia (CLM) y esto puede influir en los resultados. En segundo lugar, el tamaño de la muestra de nuestro estudio puede carecer de poder estadístico para capturar la asociación significativa en algunos marcadores.

- En el análisis de tejidos embebidos en parafina se logró identificar un polimorfismo en el dominio RING del gen TRIM62, dominio que está involucrado directamente en la transferencia de ubiquitin desde la E2 a la E3 en la cascada de ubiquitinación, dicho polimorfismo ya se encuentra anotado y es un polimorfismo sinónimo, por lo cual no afecta la función de la proteína.

5.2 Recomendaciones

- No Incluir sólo tagSNPs basados sobre el MAF, tratar de evaluar aquellos SNPs que ya se encuentren validados.
- Evaluar SNPs en regiones regulatorias del gen que afecten la estabilidad de la proteína.
- Realizar la validación de los SNPs que mostraron asociación con cáncer de seno y con uno de los parámetros pronósticos de la enfermedad.
- Incluir dentro de los análisis de asociación gen candidato no sólo un gen sino varios genes que se hayan asociado con cáncer de seno con el fin de obtener asociaciones más robustas.
- Realizar la validación de los SNPs, que mostraron en el análisis de predicción sitios de unión a miRNA.
- Emplear otro tipo de metodologías para el análisis de variantes patógenicas en tejidos tumorales embebidos en parafina tales como, secuenciación de última

generación donde se pueda hacer la detección de indels, variantes en el número de copias entre otras.

- Emplear tejidos frescos para la búsqueda de nuevas variantes.
- Establecer protocolos en las instituciones hospitalarias en los departamentos de patología para la preservación de los tejidos en parafina que mejoren la preservación del ADN para ser utilizados en análisis retrospectivos.
- Crear una base de datos centralizada donde se pueda tener acceso a la información sobre factores de riesgo, datos histopatológicos, información sobre el tratamiento y demás información que sea relevante para el desarrollo de este tipo de estudios.

Referencias

- A map of human genome variation from population-scale sequencing. (2010). *Nature*, 467(7319),1061-1073.
- Allred, D. C., Mohsin, S. K., & Fuqua, S. A. (2001). Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer*, 8(1), 47-61.
- Anderson, E. (2002). The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis. *Breast Cancer Res*, 4(5), 197-201.
- Antoniou, A. C., & Easton, D. F. (2006). Models of genetic susceptibility to breast cancer. *Oncogene*, 25(43), 5898-5905. doi:10.1038/sj.onc.1209879
- Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68-74. doi:10.1038/nature15393
- Bayés, M., & Gut, I. G. (2011). Overview of Genotyping *Molecular Analysis and Genome Discovery* (pp. 1-23): John Wiley & Sons, Ltd.
- Beral, V., Bull, D., Doll, R., Peto, R., Reeves, G., & Cancer, C. G. o. H. F. i. B. (2001). 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*, 358(9291), 1389-1399.
- Berliner, J. L., & Fay, A. M. (2007). Risk assessment and genetic counseling for hereditary breast and ovarian cancer: recommendations of the National Society of Genetic Counselors. *J Genet Couns*, 16(3), 241-260. doi:10.1007/s10897-007-9090-7
- Berndsen, C. E., & Wolberger, C. (2014). New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol*, 21(4), 301-307. doi:10.1038/nsmb.2780
- Birch, J. M., Alston, R. D., McNally, R. J., Evans, D. G., Kelsey, A. M., Harris, M., . . . Varley, J. M. (2001). Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene*, 20(34), 4621-4628. doi:10.1038/sj.onc.1204621
- Bloom HJG, R. W. (1957). Histologic grading and prognosis in breast cancer A study of 1409 cases of which 359 have been followed for 15 years. *Breast Journal Cancer*, 11(3), 359-377.

- Boardman, L. A., Thibodeau, S. N., Schaid, D. J., Lindor, N. M., McDonnell, S. K., Burgart, L. J., Hartmann, L. C. (1998). Increased risk for cancer in patients with the Peutz-Jeghers syndrome. *Ann Intern Med*, 128(11), 896-899.
- Chan, M., Ji, S. M., Liaw, C. S., Yap, Y. S., Law, H. Y., Yoon, C. S., Lee, A. S. (2012). Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population. *Breast Cancer Res Treat*, 136(1), 209-220. doi:10.1007/s10549-012-2234-y
- Chen, C., Seth, A. K., & Aplin, A. E. (2006). Genetic and Expression Aberrations of E3 Ubiquitin Ligases in Human Breast Cancer. *Molecular Cancer Research*, 4(10), 695-707. doi:10.1158/1541-7786.mcr-06-0182
- Chen, N., Balasenthil, S., Reuther, J., Frayna, A., Wang, Y., Chandler, D. S., Killary, A. M. (2013). DEAR1 is a Chromosome 1p35 Tumor Suppressor and Master Regulator of TGF β -Driven Epithelial-Mesenchymal Transition. *Cancer discovery*, 3(10), 1172-1189. doi:10.1158/2159-8290.cd-12-0499
- Chen, Z. J. (2005). Ubiquitin signalling in the NF-[kappa]B pathway. *Nat Cell Biol*, 7(8), 758-765. doi:http://www.nature.com/ncb/journal/v7/n8/supinfo/ncb0805-758_S1.html
- Ciechanover, A. (2005). Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol*, 6(1), 79-87.
- Cox, A., Dunning, A. M., Garcia-Closas, M., Balasubramanian, S., Reed, M. W. R., Pooley, K. A., Consortium, B. C. A.C (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics*, 39(3), 352-358. doi:10.1038/ng1981
- de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nat Genet*, 37(11), 1217-1223. doi:10.1038/ng1669
- Deshaies, R. J., & Joazeiro, C. A. (2009). RING domain E3 ubiquitin ligases. *Annu Rev Biochem*, 78, 399-434. doi:10.1146/annurev.biochem.78.101807.093809
- Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D., Thompson, D., Ballinger, D. G., Ponder, B. A. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447(7148), 1087-1093. doi:nature05887 [pii]10.1038/nature05887
- Elston, C. W., & Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19(5), 403-410.
- Engle, L. J., Simpson, C. L., & Landers, J. E. (2006). Using high-throughput SNP technologies to study cancer. *Oncogene*, 25(11), 1594-1601. doi:10.1038/sj.onc.1209368

- Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., . . . Iggo, R. (2005). Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene*, *24*(29), 4660-4671. doi:1208561 [pii]10.1038/sj.onc.1208561
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., . . . Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, *136*(5), E359-386. doi:10.1002/ijc.29210
- Fish, E. M., & Molitoris, B. A. (1994). Alterations in epithelial polarity and the pathogenesis of disease states. *N Engl J Med*, *330*(22), 1580-1588. doi:10.1056/NEJM199406023302207
- FitzGerald, M. G., Marsh, D. J., Wahrer, D., Bell, D., Caron, S., Shannon, K. E., . . . Haber, D. A. (1998). Germline mutations in PTEN are an infrequent cause of genetic predisposition to breast cancer. *Oncogene*, *17*(6), 727-731. doi:10.1038/sj.onc.1201984
- Frazer, K. A., Murray, S. S., Schork, N. J., & Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, *10*(4), 241-251. doi:10.1038/nrg2554
- Freemont, P. S. (1993). The RING finger. A novel protein sequence motif related to the zinc finger. *Ann N Y Acad Sci*, *684*, 174-192.
- Gago-Dominguez, M., Castelao, E. J., Gude, F., Fernandez, M. P., Aguado-Barrera, M. E., Ponte, S. M., . . . Martínez, M. E. (2016). Alcohol and breast cancer tumor subtypes in a Spanish Cohort. *SpringerPlus*, *5*(1), 1-9. doi:10.1186/s40064-015-1630-2
- Garber, J. E., Goldstein, A. M., Kantor, A. F., Dreyfus, M. G., Fraumeni, J. F., Jr., & Li, F. P. (1991). Follow-up study of twenty-four families with Li-Fraumeni syndrome. *Cancer Res*, *51*(22), 6094-6097.
- Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., Richesson, D. A., . . . Pharoah, P. D. (2008). Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet*, *4*(4), e1000054. doi:10.1371/journal.pgen.1000054
- Ghousaini, M., & Pharoah, P. D. (2009). Polygenic susceptibility to breast cancer: current state-of-the-art. *Future Oncol*, *5*(5), 689-701.
- Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., . . . Nussbaum, R. L. (2003). The International HapMap Project. *Nature*, *426*(6968), 789-796. doi:doi:10.1038/nature02168
- Gjorevski, N., & Nelson, C. M. (2011). Integrated morphodynamic signalling of the mammary gland. *Nature Reviews Molecular Cell Biology*, *12*(9), 581-593. doi:10.1038/nrm3168

- Gomez. (2006). Evaluación cuantitativa del riesgo de cáncer de mama. *Rev. Med. Clin. Condes*, 17(4), 149-163.
- González Ortega, J. M., Morales Wong, M. M., López Cuevas, Z., & Díaz Valdéz, M. (2011). Factores pronósticos del cáncer de mama. *Revista Cubana de Cirugía*, 50, 130-138.
- Gupta, S., Jaworska-Bieniek, K., Lubinski, J., & Jakubowska, A. (2013). Can selenium be a modifier of cancer risk in CHEK2 mutation carriers? *Mutagenesis*, 28(6), 625-629. doi:10.1093/mutage/get050
- Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., & King, M. C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, 250(4988), 1684-1689.
- Hatakeyama, S. (2011). TRIM proteins and cancer. *Nat Rev Cancer*, 11(11), 792-804. doi:10.1038/nrc3139nrc3139 [pii]
- Hennesy, B. T., Gonzalez-Angulo, A. M., Stemke-Hale, K., Gilcrease, M. Z., Krishnamurthy, S., Lee, J. S., . . . Mills, G. B. (2009). Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res*, 69(10), 4116-4124. doi:10.1158/0008-5472.CAN-08-34410008-5472.CAN-08-3441 [pii]
- Hofmann, W., & Schlag, P. M. (2000). BRCA1 and BRCA2--breast cancer susceptibility genes. *J Cancer Res Clin Oncol*, 126(9), 487-496.
- Howard, B. A., & Gusterson, B. A. (2000). Human breast development. *J Mammary Gland Biol Neoplasia*, 5(2), 119-137.
- Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., . . . Perou, C. M. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, 7, 96. doi:1471-2164-7-96 [pii]10.1186/1471-2164-7-96
- Huibregtse, J. M., Scheffner, M., Beaudenon, S., & Howley, P. M. (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci U S A*, 92(7), 2563-2567.
- Hulka, B. S., & Moorman, P. G. (2008). Breast cancer: hormones and other risk factors. *Maturitas*, 61(1-2), 203-213; discussion 213.
- Jarvius, J., Nilsson, M., & Landegren, U. (2003). Oligonucleotide Ligation Assay. In P.-Y. Kwok (Ed.), *Single Nucleotide Polymorphisms: Methods and Protocols* (pp. 215-228). Totowa, NJ: Springer New York.
- Kenemans, P., Verstraeten, R. A., & Verheijen, R. H. (2004). Oncogenic pathways in hereditary and sporadic breast cancer. *Maturitas*, 49(1), 34-43. doi:10.1016/j.maturitas.2004.06.005S0378512204002075 [pii]
- Kitagawa, K., Kotake, Y., & Kitagawa, M. (2009). Ubiquitin-mediated control of oncogene and tumor suppressor gene products. *Cancer Science*, 100(8), 1374-1381. doi:10.1111/j.1349-7006.2009.01196.x

- Kwok, P. Y. (2000). High-throughput genotyping assay approaches. *Pharmacogenomics*, 1(1), 95-100. doi:10.1517/14622416.1.1.95
- Kwok, P. Y. (2001). Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet*, 2, 235-258. doi:10.1146/annurev.genom.2.1.235
- Lahmann, P. H., Hoffmann, K., Allen, N., Van Gils, C. H., Khaw, K. T., Tehard, B., . . . Riboli, E. (2004). Body size and breast cancer risk: Findings from the european prospective investigation into cancer and nutrition (EPIC). *International Journal of Cancer*, 111(5), 762-771. doi:10.1002/ijc.20315
- Lambertini, M., Santoro, L., Del Mastro, L., Nguyen, B., Livraghi, L., Ugolini, D., . . . Azim, H. A., Jr. (2016). Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. *Cancer Treat Rev*, 49, 65-76. doi:S0305-7372(16)30062-7 [pii]10.1016/j.ctrv.2016.07.006
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921.
- Lim, W., Olschwang, S., Keller, J. J., Westerman, A. M., Menko, F. H., Boardman, L. A., . . . Houlston, R. S. (2004). Relative frequency and morphology of cancers in STK11 mutation carriers. *Gastroenterology*, 126(7), 1788-1794.
- Livak, K. J. (2003). SNP genotyping by the 5'-nuclease reaction. *Methods Mol Biol*, 212, 129-147.
- Lott, S. T., Chen, N., Chandler, D. S., Yang, Q., Wang, L., Rodriguez, M., . . . Killary, A. M. (2009a). DEAR1 is a dominant regulator of acinar morphogenesis and an independent predictor of local recurrence-free survival in early-onset breast cancer. *PLoS Med*, 6(5), e1000068. doi:10.1371/journal.pmed.1000068
- Lott, S. T., Chen, N. Y., Chandler, D. S., Yang, Q. F., Wang, L., Rodriguez, M., . . . Killary, A. M. (2009b). DEAR1 Is a Dominant Regulator of Acinar Morphogenesis and an Independent Predictor of Local Recurrence-Free Survival in Early-Onset Breast Cancer. *Plos Medicine*, 6(5). doi:ARTN e100006810.1371/journal.pmed.1000068
- Marblestone, J. G., LaRocque, J. P., Mattern, M. R., & Leach, C. A. (2012). Analysis of ubiquitin E3 ligase activity using selective polyubiquitin binding proteins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(11), 2094-2097. doi:<http://dx.doi.org/10.1016/j.bbamcr.2012.06.013>
- Mavaddat, N., Antoniou, A. C., Easton, D. F., & Garcia-Closas, M. (2010). Genetic susceptibility to breast cancer. *Molecular Oncology*, 4(3), 174-191. doi:<http://dx.doi.org/10.1016/j.molonc.2010.04.011>
- McGuire WL, C. P., Sears ME, Escher GC. (1975). Estrogen receptors in human breast cancer: an overview. . *Raven Press, New York*, 1-8.

- McGuire WL, C. P., Sears ME, Escher GC. (1980). An update on estrogen and progesterone receptors in prognosis for primary and advanced breast cancer. . *Progr Cancer Res Ther*, 14, 337-343.
- McPherson, K., Steel, C. M., & Dixon, J. M. (2000). ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ*, 321(7261), 624-628.
- Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., . . . Stratton, M. R. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet*, 31(1), 55-59.
- Meroni, G., & Diez-Roux, G. (2005). TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. *Bioessays*, 27(11), 1147-1157. doi:10.1002/bies.20304
- Metzger Filho, O., Ignatiadis, M., & Sotiriou, C. (2011). Genomic Grade Index: An important tool for assessing breast cancer tumor grade and prognosis. *Crit Rev Oncol Hematol*, 77(1), 20-29. doi:10.1016/j.critrevonc.2010.01.011S1040-8428(10)00012-0 [pii]
- Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., . . . Easton, D. F. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*, 45(4), 353-361. doi:<http://www.nature.com/ng/journal/v45/n4/abs/ng.2563.html#supplementary-information>
- Michelle Alizart, J. S., Margaret Cummings, Sunil R. Lakhani. (2012). Molecular classification of breast carcinoma. *Diagnostic Histopathology*, 18(3), 97-103. doi: <http://dx.doi.org/10.1016/j.mpdhp.2011.12.003>
- Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, 16(3), 1215.
- Muthuswamy, S. K. (2009). A new tumor suppressor that regulates tissue architecture. *PLoS Med*, 6(5), e1000073. doi:10.1371/journal.pmed.1000073
- Oeth, P., del Mistro, G., Marnellos, G., Shi, T., & van den Boom, D. (2009). Qualitative and quantitative genotyping using single base primer extension coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MassARRAY). *Methods Mol Biol*, 578, 307-343. doi:10.1007/978-1-60327-411-1_20
- Olayioye, M. A., Neve, R. M., Lane, H. A., & Hynes, N. E. (2000). The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J*, 19(13), 3159-3167. doi:10.1093/emboj/19.13.3159
- Olefsky, J. M. (2001). Nuclear receptor minireview series. *J Biol Chem*, 276(40), 36863-36864. doi:10.1074/jbc.R100047200
- Patnala, R., Clements, J., & Batra, J. (2013). Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC Genetics*, 14(1), 39. doi:10.1186/1471-2156-14-39

- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., . . . Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*(6797), 747-752. doi:10.1038/35021093
- Pharoah, P. D., Guilford, P., & Caldas, C. (2001). Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology*, *121*(6), 1348-1353.
- Pérez Sánchez VM, V. C. T., Mora Tiscareño A. (2008). Diagnóstico histopatológico y factores pronóstico en cáncer infiltrante de glándula mamaria. *Cancerología*, *3*, 7-17.
- Quintas-Cardama, A., Post, S. M., Solis, L. M., Xiong, S., Yang, P., Chen, N., . . . Lozano, G. (2014). Loss of the novel tumour suppressor and polarity gene Trim62 (Dear1) synergizes with oncogenic Ras in invasive lung cancer. *J Pathol*, *234*(1), 108-119. doi:10.1002/path.4385
- Quintas-Cardama, A., Zhang, N., Qiu, Y. H., Post, S., Creighton, C. J., Cortes, J., . . . Kornblau, S. M. (2015). Loss of TRIM62 expression is an independent adverse prognostic factor in acute myeloid leukemia. *Clin Lymphoma Myeloma Leuk*, *15*(2), 115-127.e115. doi:10.1016/j.clml.2014.07.011
- Rahman, N., Seal, S., Thompson, D., Kelly, P., Renwick, A., Elliott, A., . . . Stratton, M. R. (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet*, *39*(2), 165-167. doi:10.1038/ng1959
- Renwick, A., Thompson, D., Seal, S., Kelly, P., Chagtai, T., Ahmed, M., . . . Rahman, N. (2006). ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet*, *38*(8), 873-875. doi:10.1038/ng1837
- Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., . . . Ballabio, A. (2001). The tripartite motif family identifies cell compartments. *EMBO J*, *20*(9), 2140-2151. doi:10.1093/emboj/20.9.2140
- Reynolds, P., Hurley, S., Goldberg, D. E., Anton-Culver, H., Bernstein, L., Deapen, D., . . . Ziogas, A. (2004). Active smoking, household passive smoking, and breast cancer: evidence from the California Teachers Study. *J Natl Cancer Inst*, *96*(1), 29-37.
- Ripperger, T., Gadzicki, D., Meindl, A., & Schlegelberger, B. (2009). Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet*, *17*(6), 722-731. doi:10.1038/ejhg.2008.212
- Roberts, J. M., Eide, E., & Faltermier, C. (2009). Inhibition of trim62 activity reduces cancer cell proliferation: Google Patents.

- Ruiz-Linares, A., Adhikari, K., Acuna-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., . . . Gonzalez-Jose, R. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet*, *10*(9), e1004572. doi:10.1371/journal.pgen.1004572
- Savage, S. A., & Chanock, S. J. (2006). Genetic association studies in cancer: good, bad or no longer ugly? *Hum Genomics*, *2*(6), 415-421.
- Seal, S., Thompson, D., Renwick, A., Elliott, A., Kelly, P., Barfoot, R., . . . Rahman, N. (2006). Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet*, *38*(11), 1239-1241.
- Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I. A., Avril, S., . . . Stingl, J. (2012). Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res*, *14*(5), R134. doi:10.1186/bcr3334 [pii]
- Shen, R., Fan, J. B., Campbell, D., Chang, W., Chen, J., Doucet, D., . . . Oliphant, A. (2005). High-throughput SNP genotyping on universal bead arrays. *Mutat Res*, *573*(1-2), 70-82. doi:10.1016/j.mrfmmm.2004.07.022
- Sinn, H. P., & Kreipe, H. (2013). A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care (Basel)*, *8*(2), 149-154. doi:10.1159/000350774
- Smid, M., Wang, Y., Klijn, J. G., Sieuwerts, A. M., Zhang, Y., Atkins, D., . . . Foekens, J. A. (2006). Genes associated with breast cancer metastatic to bone. *J Clin Oncol*, *24*(15), 2261-2267. doi:JCO.2005.03.8802 [pii]10.1200/JCO.2005.03.8802
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., . . . Borresen-Dale, A. L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, *98*(19), 10869-10874. doi:10.1073/pnas.19136709898/19/10869 [pii]
- Spratt, Donald E., Walden, H., & Shaw, Gary S. (2014). RBR E3 ubiquitin ligases: new structures, new insights, new questions. *Biochemical Journal*, *458*(Pt 3), 421-437. doi:10.1042/bj20140006
- Sun, Y., Zhang, T., Wang, C., Jin, X., Jia, C., Yu, S., & Chen, J. (2015). MiRNA-615-5p functions as a tumor suppressor in pancreatic ductal adenocarcinoma by targeting AKT2. *PLoS One*, *10*(4), e0119783. doi:10.1371/journal.pone.0119783
- Syvanen, A. C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet*, *2*(12), 930-942. doi:10.1038/35103535
- Tan, M.-H., Mester, J. L., Ngeow, J., Rybicki, L. A., Orloff, M. S., & Eng, C. (2012). Lifetime Cancer Risks in Individuals with Germline PTEN Mutations. *Clinical Cancer Research*, *18*(2), 400-407. doi:10.1158/1078-0432.ccr-11-2283
- The International HapMap Project. (2003a). *Nature*, *426*(6968), 789-796. doi:10.1038/nature02168 nature02168 [pii]

- The International HapMap Project. (2003b). *Nature*, 426(6968), 789-796. doi:http://www.nature.com/nature/journal/v426/n6968/suppinfo/nature02168_S1.html
- Thomas, G., Jacobs, K. B., Kraft, P., Yeager, M., Wacholder, S., Cox, D. G., . . . Hunter, D. J. (2009). A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nature Genetics*, 41(5), 579-584. doi:10.1038/ng.353
- Thompson, D., & Easton, D. (2004). The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia*, 9(3), 221-236. doi:494761 [pii] 10.1023/B:JOMG.0000048770.90334.3b
- Torres, D., Rashid, M. U., Gil, F., Umana, A., Ramelli, G., Robledo, J. F., Hamann, U. (2007). High proportion of BRCA1/2 founder mutations in Hispanic breast/ovarian cancer families from Colombia. *Breast Cancer Res Treat*, 103(2), 225-232. doi:10.1007/s10549-006-9370-1
- Turnbull, C., & Rahman, N. (2008). Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet*, 9, 321-345. doi:10.1146/annurev.genom.9.081307.164339
- Uchil, P. D., Quinlan, B. D., Chan, W. T., Luna, J. M., & Mothes, W. (2008). TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathog*, 4(2), e16. doi:10.1371/journal.ppat.004001607-PLPA-RA-0403 [pii]
- Vega, A., Salas, A., Milne, R. L., Carracedo, B., Ribas, G., Ruibal, A., . . . Carracedo, A. (2009). Evaluating new candidate SNPs as low penetrance risk factors in sporadic breast cancer: a two-stage Spanish case-control study. *Gynecol Oncol*, 112(1), 210-214. doi:10.1016/j.ygyno.2008.09.012
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., . . . Zhu, X. (2001). The sequence of the human genome. *Science*, 291(5507), 1304-1351.
- Veronesi, U., Viale, G., Rotmensz, N., & Goldhirsch, A. (2006). Rethinking TNM: breast cancer TNM classification for treatment decision-making and research. *Breast*, 15(1), 3-8. doi:S0960-9776(05)00297-3 [pii]10.1016/j.breast.2005.11.011
- Walsh, T., Casadei, S., Coats, K. H., Swisher, E., Stray, S. M., Higgins, J., . . . King, M. C. (2006). Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA*, 295(12), 1379-1388.
- Washbrook, E. (2006). Breast disorders Risk factors and epidemiology of breast cancer. *Women's Health Medicine*, 3(1), 8-14. doi:<http://dx.doi.org/10.1383/wohm.2006.3.1.8>
- Weigelt B, G. F., Reis-Filho JS. (2010). Histological types of breast cancer: how special are they? *Molecular Oncology*, 4(3), 192-208. doi:10.1016/j.molonc.2010.04.004
- Winston, J. S., Ramanaryanan, J., & Levine, E. (2004). HER-2/neu evaluation in breast cancer are we there yet? *Am J Clin Pathol*, 121 Suppl, S33-49.

- Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., . . . Micklem, G. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature*, 378(6559), 789-792. doi:10.1038/378789a0
- Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., & Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *Am J Hum Genet*, 86(6), 929-942. doi:10.1016/j.ajhg.2010.05.002
- Yoon, A. R., Gao, R., Kaul, Z., Choi, I. K., Ryu, J., Noble, J. R., . . . Wadhwa, R. (2011). MicroRNA-296 is enriched in cancer cells and downregulates p21WAF1 mRNA expression via interaction with its 3' untranslated region. *Nucleic Acids Res*, 39(18), 8078-8091. doi:10.1093/nar/gkr492

A. Anexo: Consentimiento informado

CONSENTIMIENTO INFORMADO PARA LA PARTICIPACIÓN EN UN PROYECTO DE INVESTIGACIÓN

Título del Proyecto de Investigación:

“ESTUDIO DE FACTORES GENÉTICOS Y NO-GENÉTICOS PARA CÁNCER DE SENO Y DE OVARIO EN COLOMBIA”.

Usted (o su pariente) está invitado a participar en un estudio de investigación propuesto por el Instituto de Genética Humana de la Pontificia Universidad Javeriana, con la participación del Deutschen Krebsforschungszentrum:

Dr. Ignacio Briceño Balcázar, Instituto de Genética Humana, PUJ (Bogotá - Colombia).
Dra. Diana Torres López, Instituto de Genética Humana, PUJ (Bogotá - Colombia).
Dra. Ute Hamann, Deutschen Krebsforschungszentrum (Heidelberg - Germany).

Es muy importante que usted lea y entienda ciertos puntos importantes en la realización de este estudio: (a) Su participación en este estudio es totalmente voluntaria. (b) El participar en este estudio puede no beneficiarlo a usted directamente, pero esta investigación nos permite clarificar muchos conceptos sobre la enfermedad y su mecanismo de herencia, de manera que los beneficios posteriores sean para usted, su familia u otros individuos afectados. (c) Usted puede retirarse del estudio cuando lo desee. La revocación de este consentimiento no tendrá perjuicio alguno sobre la relación médico-paciente. (d) Ninguna persona involucrada en este estudio recibirá beneficios económicos como pago por su participación. (e) Este estudio no tiene ningún interés económico por parte nuestra o de las instituciones colaboradoras, así como tampoco el desarrollo de patentes con base en el material genético. (f) **CONFIDENCIALIDAD:** Los registros con la información de cada individuo permanecerán archivados en el Instituto de Genética Humana. Las historias médicas, los resultados de exámenes y la información que usted nos ha dado son de carácter absolutamente confidencial, de manera que, solamente usted y el equipo de investigación tendrá acceso a estos datos. Por ningún motivo se divulgará esta información sin su consentimiento. Cuando los resultados de este estudio sean reportados en revistas médicas científicas o en congresos científicos, los nombres de todos aquellos que tomaron parte en el estudio serán omitidos. (g) La naturaleza de este estudio, sus riesgos, sus inconvenientes, incomodidades y cualquier información importante está resumida a continuación y será explicada por el grupo investigador. (h) Los resultados de las pruebas genéticas podrían beneficiar sus parientes consanguíneos, mediante asesoría genética usted podrá decidir si desea compartirlos con parientes que pudieran estar en riesgo. (i) Si tiene algún interrogante sobre el estudio por favor no dude en manifestarlo a alguno de los investigadores, quien, con mucho gusto, le contestará sus preguntas.

Cualquier información adicional usted puede obtenerla de los investigadores, o directamente con:

Dr. Ignacio Briceño Balcázar, Instituto de Genética Humana, PUJ: (571) 3208320, Ext: 2792.
Dra. Diana Torres López, Instituto de Genética Humana, PUJ: (571) 3208320, Ext: 2792.
Dra. Ute Hamann, Deutschen Krebsforschungszentrum (Heidelberg-Germany), (49) 6221422347.

EXPLICACIÓN DEL PROYECTO DE INVESTIGACIÓN

JUSTIFICACIÓN Y OBJETIVO:

Cáncer de seno es el cáncer más común entre las mujeres a nivel mundial. Se ha observado que la incidencia de cáncer de seno es mayor en Europa y Norte América aun cuando las tasas en países en vías de desarrollo incluido Sur América se han venido incrementando. En Colombia, la incidencia edad-estandarizada para cáncer de seno (Registro de Cali) fue estimada en 38.8 pacientes por 100.000 casos, por año.

El objetivo principal de esta investigación es estudiar la prevalencia y penetrancia de las mutaciones fundadoras encontradas en los genes *BRCA1* y *BRCA2*, que fueron identificadas en la fase inicial del proyecto de investigación y evaluar, en relación a la etiología del cáncer de seno, el papel de la susceptibilidad genética, mediante el estudio de factores genéticos e interacciones genético-ambientales en población Colombiana.

PROCEDIMIENTO:

Se realizará una entrevista clínica con usted y se tomará una muestra de aproximadamente 8 ml de sangre mediante punción en vena periférica respectivamente. En caso de que sea necesario repetir los exámenes, usted será notificado para tomar las muestras nuevamente. Estas muestras serán manejadas y analizadas únicamente por personas involucradas directamente en este proyecto y almacenadas en nuestro laboratorio en el Instituto de Genética Humana.

RIESGOS E INCOMODIDADES La participación en este estudio representa riesgo mínimo para su salud e integridad y las molestias y efectos adversos estarán representadas exclusivamente por la toma de muestra referida en el procedimiento.

BENEFICIOS ADICIONALES:

El conocimiento generado de esta investigación será de gran utilidad en la identificación de mujeres con alto riesgo de desarrollar cáncer de seno para dirigir los esfuerzos a estrategias de prevención. Los resultados de este estudio proporcionarán información sobre su riesgo a desarrollar cáncer de seno u ovario contra-lateral y el riesgo de sus familiares a desarrollar cáncer de seno y/o ovario. La detección temprana de mutaciones en los genes *BRCA1* y *BRCA2* permitirá tomar medidas preventivas como controles frecuentes, mastectomía y/o ooforectomía profiláctica y/o quimioprevención.

RESPONSABILIDAD DEL PACIENTE Y PRECAUCIONES

Al tomar parte de este estudio es importante que usted contemple las siguientes responsabilidades y precauciones:

Comprende que el riesgo existente en una toma de muestra de sangre en vena periférica es muy bajo y por lo tanto no reviste riesgo en la salud del paciente.

Comprende que no hay indemnización para su participación en este estudio.

Comprende que la participación en este proyecto de investigación no incluye el tratamiento requerido.

MANEJO DE RESULTADOS:

Los resultados de los estudios genéticos solo le serán revelados, **si usted lo desea**, cuando representen una ayuda diagnóstica para la entidad objeto de este estudio; estos resultados le serán entregados durante consulta clínica en presencia de un MÉDICO GENETISTA. Los resultados de pruebas diagnósticas no genéticas empleadas en el estudio serán remitidos en sobre cerrado directamente al médico tratante que lo remitió.

OTRA INFORMACIÓN PERTINENTE: El riesgo a desarrollar cáncer de seno para población general es del 10-13% y para cáncer de ovario del 1-2%. Portadores de mutación en el gen *BRCA1* tienen un incremento en el riesgo a desarrollar cáncer de seno entre 36-87% y para *BRCA2* entre 36-71%. El correspondiente riesgo a cáncer de ovario para portadores en *BRCA1* es de 16-63% y en portadores en *BRCA2* es de 16-27%.

Adicionalmente, portadores de mutación en *BRCA1* o *BRCA2* tiene un riesgo a desarrollar cáncer de seno contra-lateral del 64% y 52% respectivamente y cáncer en el ovario contrario de 44% y 16%, respectivamente. Hombres portadores de mutación en el gen *BRCA2* tienen un riesgo del 6% de desarrollar cáncer de seno durante su vida.

“ESTUDIO DE FACTORES GENÉTICOS Y NO-GENÉTICOS PARA CÁNCER DE SENO Y OVARIO EN COLOMBIA”.

AUTORIZACIÓN:

La utilización de la muestra en estudios posteriores nos podría ayudar en el futuro a entender las causas y/o el comportamiento de la(s) entidad(es) anteriormente mencionada(s). Se puede dar el caso en donde usted y su familia no se beneficien directamente de estos estudios, pero tanto su familia como otros individuos afectados podrían beneficiarse. Por lo tanto, por favor marque su decisión con respecto al almacenamiento de la muestra y su utilización en estudios de investigación posteriores.

Deseo que la muestra que me fue extraída sea DESECHADA una vez completado el estudio de investigación.

Autorizo conservar la muestra que me fue extraída con la posibilidad de emplearla en las situaciones señaladas a continuación:

- En estudios de investigación colaborativos con otras instituciones Nacionales y/o Internacionales, enviando la muestra al exterior a el(los) laboratorio(s) del(los) Sí No instituto(s) antes mencionado(s).
- En estudios complementarios de diagnóstico para mí o algún miembro de mi familia : Sí No
- En estudios de investigación específicos para la(s) entidad(es), objeto de esta toma de muestra, siempre y cuando se conserve en anonimato mis datos de identificación: Sí No
- En estudios de investigación de entidades distintas a la(s) entidad(es) objeto de esta toma de muestra, siempre y cuando se conserve en anonimato mis datos de identificación: Sí No

- En estudios de investigación colaborativos con otras instituciones Nacionales y/o Internacionales, siempre y cuando exista acuerdo interinstitucional previo, aprobación del comité de ética y se conserve en anonimato mis datos de Sí No identificación.

AUTORIZACIÓN PARA LA TOMA DE MUESTRAS E INCLUSIÓN VOLUNTARIA EN EL PROYECTO DE INVESTIGACIÓN:

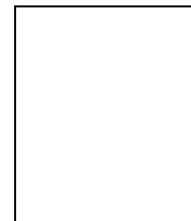
“ESTUDIO DE FACTORES GENÉTICOS Y NO-GENÉTICOS PARA CÁNCER DE SENO Y OVARIO EN COLOMBIA”.

Yo, _____ con documento de identificación No. _____ de _____, acepto voluntariamente que se me tome una muestra de sangre con el fin de realizar el análisis: “Estudio de factores genéticos y no-genéticos para cáncer de seno y ovario en Colombia”. Así mismo, declaro que se me ha explicado la presencia de los riesgos y el manejo que se le dará a la muestra.

Fecha: _____

Firma del Paciente

Testigo



Huella Digital dedo índice derecho

B. Anexo: Cuestionario sobre factores potenciales de riesgo para pacientes con cáncer de seno

“ESTUDIO DE CÁNCER DE SENO Y OVARIO EN COLOMBIA”.

El propósito de este proyecto es conocer acerca del desarrollo, tratamiento y prevención de cáncer de seno y ovario.

Nombre: _____

Fecha de Nacimiento: _____ **Edad:** ____

Fecha de hoy: _____

SECCION I – FERTILIDAD (Por favor describa la historia de sus embarazos, se ha encontrado que esta afecta el riesgo de cáncer en algunas mujeres).

1. ¿Ha estado usted embarazada?

- No** → **Continué con le pregunta 6.**
 Si

2. Por favor describa todos sus embarazos en orden, desde el primero hasta el último. Especifique el año de nacimiento de cada uno. Marque con X la columna apropiada según el resultado de cada embarazo. Traté de recordar cuanto tiempo amamantó (le dio pecho) a cada recién nacido, (marque O a los recién nacidos que no amamantó) e indique en la última columna los meses que los amamantó.

		HISTORIAL DE EMBARAZOS						
#	Año de Embarazo	Embarazo perdido	Aborto terapéutico	Nacido Muerto	Nacido Vivo (Mes/Año de nacimiento)	Semanas de embarazo	Peso de Nacimiento (si sabe)	Meses que Amamanto (dio pecho)
1								
2								
3								
4								
5								
6								
7								
8								
9								

*(Para gemelos marque G a la columna correspondiente)

3. ¿Cuántos embarazos de éstos fueron cesáreas? _____

Encierre en un círculo cuántos fueron por cesáreas (1-9 del gráfico) **1 2 3 4 5**
6 7 8 9

4. ¿Tuvo alguna vez dificultad al amamantar?

- No**
 Si → Razón: ___ poca producción de leche ___ dolor
 ___ bebe prematuro ___ inflamación del seno
 ___ Otra razón: _____

Encierre en un círculo en qué embarazo (1-9 del gráfico) experimentó esto:

1 2 3 4 5 6 7 8

5. ¿Ha tomado alguna vez medicamentos para finalizar la producción de leche?
- No**
- Si** → Nombre de medicamento (si sabe): _____
- Método: ___ inyección o ___ pastillas
- Encierre en un círculo en qué embarazo (1-9 del gráfico) tomó ese medicamento:
- 1 2 3 4 5 6 7 8 9**
6. ¿Cuántos años tenía cuando tuvo su primer periodo menstrual? _____ años.
7. Describiría sus ciclos menstruales tal como: siempre regular, normalmente regular o nunca regular (regularidad significa: que el principio de su periodo era seguro dentro de los mismos 5 días del mes). (Marque uno de ellos)
- Mis periodos fueron siempre regulares.**
- Mis periodos fueron normalmente regulares.**
- Mis periodos fueron irregulares.**
8. ¿Cuántos días tardó/tarda su ciclo menstrual? (un ciclo es definido como el principio del primer día de su periodo y el último día antes de su próximo periodo.). (Marque uno)
- ___ **28 días** ___ **29 días** ___ **30 días** ___ **31 días** ___ **32 días** **Otro: ___ días.**
9. ¿Tiene usted todavía periodos menstruales? En otras palabras ¿Ha tenido usted algún periodo menstrual durante el año pasado?
- No**
- Si**
10. ¿Han terminado completamente sus periodos?
- No** → **Continué con la pregunta 13**
- Si**
11. ¿Qué edad tenía cuando su periodo terminó completamente? _____
12. ¿Cuál fue la razón? (Marque una razón)
- Menopausia natural**
- Histerectomía** (solo el útero, no los ovarios)
- Útero y ovarios extirpados**
- Ovarios extirpados** (no el útero)
- Medicación / Quimioterapia** tomó medicamento que le hizo finalizar el periodo
- Otra razón (por favor especifique)**

13. ¿Ha visitado al doctor por algún problema o dificultad en quedar embarazada o llevar el embarazo hasta su término completo? Tal como: ¿algunos abortos?

- No** → **Continué con la pregunta 16.**
 Si → ¿Qué razón le dio el doctor por la que no podía quedar embarazada? (*Por favor marque todas las que le aplican.*)

- | | |
|-----------------------------------|--|
| __ no se encontró ningún problema | __ problema con la cerviz |
| __ Problema con los ovarios | __ cónyuge tiene problemas fértiles |
| __ Problema con las trompas | __ otra razón (Por favor especifique): |
| __ Problema con endometriosis | _____ |

14. ¿Ha tomado algún medicamento para quedar embarazada?

- No**
 Si → Nombre de medicamento(s) _____
 ¿Por cuantos meses lo tomó? _____ meses
 ¿En qué año tomó ese medicamento? _____

15. ¿Ha recibido tratamiento para quedarse embarazada tal como: fertilización *in-vitro*/¿Embrión trasferido?

- No**
 Si → ¿Qué tipo de tratamiento recibió? _____

16a. ¿Ha usado alguna vez la píldora, Norplant (implantes) o Depo-Provera (inyecciones) para prevenir el embarazo ó por alguna otra razón?

- No** → **Continué con la pregunta 17.**
 Si → Por favor describa las ocasiones?

	Nombre de medicamento	Año que inició	Año que terminó	Lapso de tiempo que uso	Método (marque)
1					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones
2					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones
3					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones
4					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones

16b. ¿Está usted tomando la píldora actualmente? (Depo-Provera o Norplant)

- No**
 Si

17a. ¿Ha tomado usted terapia de hormonas para la menopausia?

- No** → **Continué con la pregunta 18.**
 Si → Complete la información:

	Nombre de las Hormonas	Año que inició	Año que terminó	Lapso de tiempo que usó (años)	Método (marque)
1					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales
2					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales
3					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales
4					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales

17b. ¿Está tomando terapia de hormonas actualmente?

No

Si → ¿Qué tipo? _____ ¿Dosis? _____

SECCION II – CÁNCER DE SENO

18. ¿Ha sido usted tratada alguna vez por cáncer del seno?

No → Continué con la pregunta 32 (Sección III).

Si → Año de diagnóstico: _____

Marque uno → ¿Ha tenido usted cáncer en un seno? un seno
o dos senos

Por favor explique su primer diagnóstico.

¿Qué año fue? _____

¿Qué mes fue (si sabes)? _____

¿Cuál Seno? Izquierdo

Derecho

Tipo de Cáncer (Marque uno):

Este fue un cáncer invasivo.

- Este no fue un cáncer invasivo (estado temprano del cáncer de seno).
→ Fue Carcinoma Ductal In-situ (DCIS)? **Sí** **No**
→ Fue Carcinoma Lobular In-situ (LCIS)? **Sí** **No**
- Fue ambos, invasivo y no-invasivo.
 No estoy segura.

Tratamiento

- ¿Fue usted operada con el **primer** cáncer de seno?
 No
 Si → Año de operación _____
¿Qué clase de operación fue?
Marque uno → ___ Cuadrantectomía
___ Lumpectomía (tumor removido, pecho no removido)
___ Mastectomía (un pecho removido)
___ Mastectomía **bilateral** (ambos pechos removidos)

Detección

- ¿Cómo se detectó el tumor?
Marque uno → ___ ¿Con mamografía?
___ ¿Palpó usted una masa?
___ ¿Palpó su doctor una masa?
___ ¿Con resonancia magnética - (MRI)?
___ Otro: _____

19. ¿Recibió usted tratamiento de Quimioterapia con su primer cáncer de seno?

- No**
 Si

20. ¿Recibió usted tratamiento de Radiación con su primer cáncer de seno?

- No**
 Si

21. ¿Recibió usted Tamoxifen con su primer cáncer del seno?

- No**
 Si → Cantidad de pastillas por día _____
Desde _____ (año) hasta _____ (año)

22. ¿Recibió usted otros medicamentos con su primer cáncer del seno?

- No**
 Si → Cuál? Femara (letrozole)
 Aromasin (exemstane)
 Arimidex (anastrozole)
 Otro: _

Cantidad de pastillas por día _____
Desde _____ (año) hasta _____ (año): _____

23. ¿Ha tenido una repetición del primer cáncer?

No

Si → Fecha: ___/___/____ (mes/día/año)

Lugar: _____

24. Ha usted tenido cirugías adicionales en el seno afectado (tales como mastectomía completa o reconstrucción)?

No

Si → Fecha: ___/___/____ (mes/día/año)

Tipo de Cirugía: _____

[Si usted no desarrolló cáncer en su Segundo Seno, CONTINÚE CON LA PREGUNTA 32]

25. Por favor explique su segundo cáncer. (No incluye una recurrencia, esto es, un cáncer en el mismo seno).

¿En qué año fue? _____ Qué mes (si recuerda)? _____ Fue →

Izquierdo

Derecho

Tipo de Cáncer (conteste cuando aplica):

Este fue un cáncer invasivo.

Este no fue un cáncer invasivo (estado temprano del cáncer de seno).

→ Fue Carcinoma Ductal In-situ (DCIS)? **Si** **No**

→ Fue Carcinoma Lobular In-situ (LCIS)? **Sí** **No**

Fue ambos, invasivo y no-invasivo.

No estoy segura.

Tratamiento

¿Fue usted operada para su segundo cáncer de seno?

No

Si → Año de operación _____.

¿Qué clase de operación fue?

Marque uno → ___ **Lumpectomía** (tumor removido, pecho no removido)

___ Mastectomía (un pecho removido)

___ Mastectomía **bilateral** (ambos pechos removido)

Detección

¿Cómo se detectó el tumor?

Marque uno → ___ ¿Con mamografía?

___ ¿Palpó usted una masa?

___ ¿Palpó su doctor una masa?

___ ¿Con resonancia magnética - (MRI)?

___ Otro: _____

26. ¿Recibió usted tratamiento de Quimioterapia con su primer cáncer de seno?

No

Si

27. ¿Recibió usted tratamiento de Radiación con su segundo cáncer de seno?

No

Si

28. ¿Recibió usted Tamoxifen con su segundo cáncer de seno?

No

Si → Cantidad de pastillas por día: _____

Desde: _____ (año) hasta _____ (año).

29. ¿Recibió usted otros medicamentos con su segundo cáncer de seno?

No

Si → Cual?

Femara (letrozole)

Aromasin (exemstane)

Arimidex (anastrozole)

Otros: _____

Cantidad de pastillas por día _____

Desde _____ (año) hasta _____ (año).

30. Tuvo recurrencia de su segundo cáncer de seno?

No

Si → Fecha: ___ / ___ / ___ (mes/día/año)

Sitio de recurrencia: _____

31. Tuvo una cirugía adicional en el seno afectado (como una mastectomía completa o reconstrucción)?

No

Si → Fecha: ___ / ___ / ___ (mes/día/año)

Tipo de Cirugía: _____

SECCION III – MÉTODOS PARA PREVENIR CÁNCER DE SENO

32. Ha tomado alguna droga para prevención de cáncer de seno (drogas tomadas antes de cualquier diagnóstico de cáncer de seno)?

- No**
- No estoy segura**
- Si** → Medicamento: ___ Tamoxifen (Nolvadex) ___ Raloxifene (Evista)
 ___ Aromasin (Exemestane) ___ Femara (Letrozole)
 ___ Arimidex (Anastrozole)
 ___ Otro (*explique*): _____

Cantidad de pastillas al día: _____

Desde: ___ / ___ / ___ (mes/año) hasta ___ / ___ / ___ (mes/año)

33. ¿Es ó fue usted participante en un estudio para prevención con Tamoxifen ó Raloxifen?

No

- Si** → Otro? ___ Ensayo con Tamoxifen (Tamoxifen vs. placebo)
 ___ Ensayo "STAR" (Tamoxifen vs. Raloxifen)

Raloxifen)

___ Otro: _____

___ Desconocido

¿Sabe usted que droga tomó? **Sí** **No** ¿Cuál? _____

34. Tiene o ha tenido implantes de seno o reconstrucción de seno?

No

- Si** → Qué tipo? ___ Salino ___ Silicona
 ___ TRAM-flap ___ Otro: _____

35. ¿Le han operado alguna vez para prevenir cáncer de seno? (i.e. remoción de ambos senos).

(NOTA: La cirugía es solo preventiva si usted nunca ha sido previamente diagnosticado con cáncer de seno).

No

- Si** → En qué año? _____

¿Cuál seno? ___ **Izquierdo** ___ **Derecho** ___ **Ambos**

Procedimiento: ___ **Mastectomía Total** (pezón y areola removidos)

___ **Mastectomía Subcutánea** (pezón y areola preservados)

36. Ha tenido alguna vez una biopsia de seno?

No

- Si** → Número de biopsias: _____

Mes/Año de la última biopsia: ___ / ___

Cuál fue el resultado de la biopsia? (*Marque todos los que aplique*)

___ normal ___ DCIS ___ Hiperplasia atípica

___ cáncer ___ LCIS ___ Fibroadenoma

___ otro (especifique): _____

37. ¿Aproximadamente cuántas mamografías ha tenido en toda su vida?

He tenido _____ mamografías.

¿A qué edad se hizo su primera mamografía? _____ años.

¿Edad que se hizo su última mamografía? _____ años

38. ¿Han detectado alguna anomalía con alguna de sus mamografías?

No

Si → Por favor explique la anomalía: _____

Mes/Año de anomalía: ___ / _____

39. Ha tenido alguna vez estudio por MRI (resonancia magnética) en sus senos?

No

Si → Qué año? _____

SECCION IV – CÁNCER DE OVARIO

40. ¿Ha tenido usted alguna vez cáncer de ovario?

No → **Continúe con la pregunta 45 (Sección V).**

Si → Año: _____

41. ¿Fue usted operada?

No

Si → Marque uno → ___ Un ovario fue removido
___ Ambos ovarios fueron removidos

42. ¿Recibió usted tratamientos de quimioterapia?

No

C. Anexo: Cuestionario sobre factores potenciales de riesgo para controles

ESTUDIO DE CÁNCER DE SENO Y OVARIO EN COLOMBIA”.

El propósito de este proyecto es conocer acerca del desarrollo, tratamiento y prevención de cáncer de seno y ovario.

CONTROL

Nombre: _____

Fecha de Nacimiento: _____ **Edad:** ____

Fecha de hoy: _____

SECCION I – FERTILIDAD (Por favor describa la historia de sus embarazos, se ha encontrado que esta afecta el riesgo de cáncer en algunas mujeres).

1. ¿Ha estado usted embarazada?

No → **Continué con le pregunta 6.**

Si

3. Por favor describa todos sus embarazos en orden, desde el primero hasta el último. Especifique el año de nacimiento de cada uno. Marque con X la columna apropiada según el resultado de cada embarazo. Trate de recordar cuanto tiempo amamantó (le dio pecho) a cada recién nacido, (marque O a los recién nacidos que no amamantó) e indique en la última columna los meses que los amamantó.

HISTORIAL DE EMBARAZOS								
#	Año de Embarazo	Embarazo perdido	Aborto terapéutico	Nacido Muerto	Nacido Vivo (Mes/Año de nacimiento)	Semanas de embarazo	Peso de Nacimiento (si sabe)	Meses que Amamanto (dio pecho)
1								
2								
3								
4								
5								
6								
7								
8								
9								

*(Para gemelos marque G a la columna correspondiente)

3. Cuántos embarazos de éstos fueron cesáreas? _____

Encierre en un círculo cuántos fueron por cesáreas (1-9 del gráfico)

1	2	3
4	5	6
7	8	9

5. ¿Tuvo alguna vez dificultad al amamantar?

No

Si → Razón: ___ poca producción de leche ___ dolor
___ bebe prematuro ___ inflamación del seno
___ Otra razón: _____

Encierre en un círculo en qué embarazo (1-9 del gráfico) experimentó esto:

1 2 3 4 5 6 7 8 9

5. ¿Ha tomado alguna vez medicamentos para finalizar la producción de leche?

No

Si → Nombre de medicamento (si sabe):

Método: ___ inyección o ___ pastillas

Encierre en un círculo en qué embarazo (1-9 del gráfico) tomó ese medicamento:

1 2 3 4 5 6 7 8 9

6. ¿Cuántos años tenía cuando tuvo su primer periodo menstrual? _____ años.

7. Describiría sus ciclos menstruales tal como: siempre regular, normalmente regular ó nunca regular (regularidad significa: que el principio de su periodo era seguro dentro de los mismos 5 días del mes). (*Marque uno de ellos*)

Mis periodos fueron siempre regulares.

Mis periodos fueron normalmente regulares.

Mis periodos fueron irregulares.

9. ¿Cuántos días tardó/tarda su ciclo menstrual? (un ciclo es definido como el principio del primer día de su periodo y el último día antes de su próximo periodo.). (*Marque uno*)

___ 28 días ___ 29 días ___ 30 días ___ 31 días ___ 32 días Otro: ___ días.

9. ¿Tiene usted todavía periodos menstruales? En otras palabras ¿Ha tenido usted algún periodo menstrual durante el año pasado?

No

Si

10. ¿Han terminado completamente sus periodos?

- No** → **Continué con la pregunta 13**
 Si

12. ¿Qué edad tenía cuando su periodo terminó completamente? _____

12. ¿Cuál fue la razón? (Marque una razón)

- Menopausia natural**
 Histerectomía (solo el útero, no los ovarios)
 Útero y ovarios extirpados
 Ovarios extirpados (no el útero)
 Medicación / Quimioterapia tomó medicamento que le hizo finalizar el periodo
 Otra razón (por favor especifique)

13. ¿Ha visitado al doctor por algún problema ó dificultad en quedar embarazada ó llevar el embarazo hasta su término completo? Tal como: ¿algunos abortos?

- No** → **Continué con la pregunta 16.**
 Si → ¿Qué razón le dio el doctor por la que no podía quedar embarazada? (*Por favor marque todas las que le aplican.*)

___ no se encontró ningún problema
___ problema con los ovarios fértiles
___ problema con las trompas
___ problema con endometriosis

___ problema con la cerviz
___ cónyuge tiene problemas
___ otra razón (Por favor especifique): _____

14. ¿Ha tomado algún medicamento para quedar embarazada?

- No**
 Si → Nombre de medicamento(s) _____
¿Por cuantos meses lo tomó? _____ meses
¿En qué año tomó ese medicamento? _____

16. ¿Ha recibido tratamiento para quedarse embarazada tal como: fertilización *in-vitro*/ Embrión trasferido?

- No**
 Si → ¿Qué tipo de tratamiento recibió?

16a. ¿Ha usado alguna vez la píldora, Norplant (implantes) o Depo-Provera (inyecciones) para prevenir el embarazo ó por alguna otra razón?

No → **Continué con la pregunta 17.**

Si → Por favor describa las ocasiones?

	Nombre de medicamento	Año que inició	Año que terminó	Lapso de tiempo que uso	Método (marque)
1					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones
2					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones
3					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones
4					<input type="checkbox"/> Píldoras <input type="checkbox"/> Implantes <input type="checkbox"/> Inyecciones

16b. ¿Está usted tomando la píldora actualmente? (Depo-Provera o Norplant)

No

Si

17a. ¿Ha tomado usted terapia de hormonas para la menopausia?

No → **Continué con la pregunta 18.**

Si → Complete la información:

	Nombre de las Hormonas	Año que inició	Año que terminó	Lapso de tiempo que usó (años)	Método (marque)
1					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales
2					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales
3					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales
4					<input type="checkbox"/> Píldoras <input type="checkbox"/> Parches <input type="checkbox"/> Supositorios vaginales

17b. ¿Está tomando terapia de hormonas actualmente?

No

Si → ¿Qué tipo? _____ ¿Dosis? _____

SECCION III – MÉTODOS PARA PREVENIR CÁNCER DE SENO

32. Ha tomado alguna droga para prevención de cáncer de seno (drogas tomadas antes de cualquier diagnóstico de cáncer de seno)?

No

No estoy segura

Si → Medicamento: ___ Tamoxifen (Nolvadex) ___ Raloxifene (Evista)

___ Aromasin (Exemestane) ___ Femara (Letrozole)

___ Arimidex (Anastrozole)

___ Otro (*explique*): _____

Cantidad de pastillas al día: _____

Desde: ___ / ___ / ___ (mes/año) hasta ___ / ___ / ___

(mes/año)

33. ¿Es ó fue usted participante en un estudio para prevención con Tamoxifen ó Raloxifen?

No

Si → Otro? ___ Ensayo con Tamoxifen (Tamoxifen vs.

placebo) ___ Ensayo “STAR” (Tamoxifen vs.

Raloxifen) ___ Otro: _____ Desconocido

Sabe usted que droga tomó? **Si** **No**

¿Cuál? _____

34. Tiene o ha tenido implantes de seno o reconstrucción de seno?

No

Si → Qué tipo? ___ Salino ___ Silicona

___ TRAM-flap ___ Otro: _____

35. ¿Le han operado alguna vez para prevenir cáncer de seno? (i.e. remoción de ambos senos).

(NOTA: La cirugía es solo preventiva si usted nunca ha sido previamente diagnosticado con cáncer de seno).

No

Si → En que año? _____

Cuál seno? ___ **Izquierdo** ___ **Derecho** ___ **Ambos**

Procedimiento: ___ **Mastectomía Total** (pezón y areola removidos)

___ **Mastectomía Subcutánea** (pezón y areola preservados)

36. Ha tenido alguna vez una biopsia de seno?

No

Si → Número de biopsias: _____

Mes/Año de la última biopsia: _____

___ / ___

Cuál fue el resultado de la

biopsia? (Marque todos los que aplique)

___ normal ___ DCIS ___ Hiperplasia atípica
___ cáncer ___ LCIS ___ Fibroadenoma
___ otro (especifique): _____

37. ¿Aproximadamente cuántas mamografías ha tenido en toda su vida?
He tenido _____ mamografías.
¿A qué edad se hizo su primera mamografía? _____ años.
¿Edad que se hizo su última mamografía? _____ años.

38. ¿Han detectado alguna anomalía con alguna de sus mamografías?

No

Si → Por favor explique la anomalía: _____

Mes/Año de anomalía: ___ / ___

39. Ha tenido alguna vez estudio por MRI (resonancia magnética) en sus senos?

No

Si → Qué año? _____

SECCION V – OPERACIONES REPRODUCTIVAS/ABDOMINALES

45. ¿Le han removido uno o ambos ovarios por otra razón, no relacionada con cáncer de ovarios?

No

Si → ¿En qué año fue la operación?: _____

¿Cuál fue la razón por la operación?: _____

¿Un ovario removido? _____ o ambos ovarios removidos? _____

46. ¿Le han hecho alguna vez ligamiento de trompas (salpingo)?

No

Si → ¿Qué año fue?: _____

47. ¿Ha tenido usted alguna otra operación en sus órganos reproductivos, incluyendo: ovarios, trompas, o útero (histerectomía)?

No

Si → ¿Qué año fue?: _____

¿Qué clase de operación? _____

¿Cuál fue la razón?: _____

48. ¿Ha tenido usted alguna otra operación en el abdomen? Tal como: vesícula, apéndice, etc.

No

Si → 1. ¿Cuál fue la primera operación?:

_____ Año: _____

2. ¿Cuál fue la segunda operación?:

3. _____
Año: _____

4. ¿Cuál fue la tercera operación?:

Año: _____

SECCION VI – INFORMACIÓN PERSONAL

Fecha de hoy: __ __ / __ __ / __ __ __
 Mes Día Año

Fecha de Nacimiento: __ __ / __ __ / __ __ __
 Mes Día Año

Lugar de Nacimiento: _____
 Ciudad Provincia/Estado

Residencia Actual: _____
 Ciudad Provincia/Estado

49a. ¿De qué nacionalidad son los antepasados de su padre? (*Por favor marque una opción.*)

Hispanico (País de Origen): _____	Canadiense francés: _____
Alemán: _____	Judío Ashkenazi: _____
Inglés: _____	Judío Sephardi: _____
Italiano: _____	Americano Nativo (Indio-Americano): _____
Irlandés: _____	Escocés: _____
Polaco/Eslovaco Oriental: _____	Africano o Africano-Americano: _____

Países del bloque Europeo (País de Origen): _____
Ruso: _____ Asiático/de Islas Pacíficas: _____
Escandinavo (Sueco/Finlandés/ Noruego/Danés) (País de Origen): _____
Holandés: _____
Otro (especifique): _____
No sabe: _____

49b. ¿De qué nacionalidad son los antepasados de su madre? (*Por favor marque una opción.*)

Hispanico (País de Origen): _____ Canadiense francés: _____
Alemán: _____ Judío Ashkenazi: _____
Inglés: _____ Judío Sephardi: _____
Italiano: _____ Americano Nativo (Indio-Americano): _____
Irlandés: _____ Escocés: _____
Polaco/Eslovaco Oriental: _____ Africano o Africano-Americano: _____

Países del bloque Europeo (País de Origen): _____
Ruso: _____ Asiático/de Islas Pacíficas: _____
Escandinavo (Sueco/Finlandés/ Noruego/Danés) (País de Origen): _____
Holandés: _____
Otro (especifique): _____
No sabe: _____

50. ¿Cuánto pesa actualmente? _____ kilogramos
¿Cuánto mide actualmente? _____ metros _____ centímetros
¿Recuerda su peso a los 18 años de edad? _____ kilogramos
¿Cuánto pesaba cuando tenía 30 años? _____ kilogramos
¿Cuánto pesaba cuando tenía 40 años? _____ kilogramos
¿Peso máximo alcanzado (sin considerar embarazos)? _____ kilogramos _____ años

51. ¿Recuerda usted cuanto pesó cuando nació?
 No
 Si → _____ kilogramos _____ onzas, o _____ gramos
52. ¿En qué año nació su madre? _____
53. ¿En qué orden nació usted? (i.e. primera, segunda, etc.) _____
54. ¿Es usted gemela?
 No
 Si → (*Marque uno*)
_____ ¿Es usted gemela idéntica? **Si** **No**
_____ Trillizo **Si** **No**
_____ Otro (por favor explique): _____

SECCION VII – ESTILO DE VIDA

55. ¿Ha fumado usted regularmente?
 No
 Si → ¿A qué edad empezó a fumar? _____ años.

¿Fuma usted todavía?

Si **No**

¿A qué edad dejó usted de fumar? _____ años
¿Aproximadamente, cuántos paquetes fumó/fuma a la semana?

_____ Paquetes

56. ¿Bebe ó bebió café regularmente?

No

Si →

¿A que edad empezó a beber café regularmente? _____ años.

¿Todavía bebe usted café? _____ SI _____ NO

¿A qué edad dejó de beber café? _____ años.

¿Aproximadamente: durante este periodo, cuántas tazas de café bebe cada día?

Cafeinado _____ tazas, Descafeinado _____ tazas,

Total _____
_tazas

57. ¿Toma usted bebidas alcohólicas?

No

Si → ¿Aproximadamente, cuántas bebidas alcohólicas consume durante una _____ semana?

_____ **0-3** _____ **4-9** _____ **10-20** _____ **20 ó más**

→ ¿A qué edad empezó usted a consumir bebidas alcohólicas? _____ años.

58. ¿Ha usado usted regularmente polvos de talco?

No

Si → ¿Sé lo aplicó directamente a la vagina? **Si**

No

¿Sé lo aplicó a las toallas sanitarias? **Si**

No

¿Otro uso de talco? (explique) _____

¿A qué edad empezó a usar talco? _____ años.

SECCION VIII – OTRAS ENFERMEDADES

59. ¿Está ó ha estado tomando algún medicamento regularmente?

No

Si → **Complete**

1. Medicamento _____ Desde _____ (año) hasta _____ (año).
Razón _____

2. Medicamento _____ Desde _____ (año) hasta _____ (año)
Razón _____

3. Medicamento _____ Desde _____ (año) hasta _____ (año)
Razón _____

60. ¿Padece o padeció usted de lo siguiente?

Fiebre de Heno **Si** **No**

Asma **Si** **No**

Eczema **Si** **No**

61. ¿Ha sido usted diagnosticada con otro tipo de cáncer?

No

Si → Qué tipo? _____ Año de diagnóstico?

62. Por favor describa cualquier problema médico que haya tenido en el pasado.
Especialmente los que requirieron hospitalización

SECCION IX – HISTORIA FAMILIAR

63. Por favor describa otros casos de cáncer en sus parientes, incluyendo: nombre, parentesco, tipo de cáncer (reporte patológico), edad de nacimiento, edad de diagnóstico, tipo de tratamiento (si lo sabe), edad actual o edad de fallecimiento. Un árbol genealógico puede ser dibujado.

Gracias por su colaboración con éste cuestionario. En caso de necesitarla nuevamente, podemos contactar con usted?

No

Si → Teléfono _____

En el caso que no te podemos contactar a este numero, por favor nos proporciona el teléfono de un familiar para que te podamos contactar:

Nombre: _____ **Parentesco:** _____

Teléfono: _____

PARA USO PRIVADO

Centro Participante: _____

Persona Contacto: _____

Entrevistador (por teléfono): _____

Familia Número: _____

Identificación Individual No: _____

Datos del cuestionario enviados: ____/____/____
Mes Día Año

Datos del cuestionario recibidos: ____/____/____
Mes Día Año

D. Anexo: Cuestionario sobre características clínicas e histopatológicas

ESTUDIO DE CANCER DE SENO BRCA1 Y BRCA2				
DATOS PERSONALES				
Nombre:		Fecha de nacimiento:		
Edad:		Fecha recoleccion de datos:	Telefono:	
Historia N°:		Cedula N°:		
TRATAMIENTO				
Quimioterapia Adyuvante		Si:	no:	
Tipo	AC CMF	OTRA	CUAL:	
Numero de ciclos				
Fecha de inicio				
Fecha de finalizacion				
Recibio antraciclina				
Recibio Taxanos				
Duracion del tratamiento adyuvante en meses				
Terapia adyuvante anti-hormona				
recibio terapia adyuvante anti-hormona				
Nombre del medicamento				
fecha de inicio				
fecha de finalizacion				
Duracion de la terapia en meses				
Terapia adyuvante con Trastuzumab				
Recibio terapia adyuvante con trastuzumab				
Fecha de inicio				
fecha de finalizacion				
Duracion de la terapia en semanas				
Recibio quimioterapia concomitantemente con trastuzumab				
Tipo de cirugía de seno				
Mastectomia				
cuadrantectomia				
Otra:	Especifique:			
Radioterapia				
Recibio radioterapia adyuvante				
Dosis				
Campos (seno, seno y nodulos linfoides, pared de la axilla y region supraclavicular. I				
Quimioterapia Neoadyuvante				
Recibio quimioterapia Neoadyuvante				
Nombre del medicamento				

Numero de ciclos					
Fecha de inicio					
Fecha de finalizacion					
Recibio terapia con antraciclinas					
Recibio Taxanos					
Recibio tratamiento CMF					
Estado antes del tratamiento neoadyuvante					
Estado de los nodulos linfoides antes del tratamiento neoadyuvante					
Tamaño del tumor antes del tratamiento neoadyuvante					
Estado despues del tratamiento neoadyuvante					
Estado de los nodulos linfoides despues del tratamiento neoadyuvante					
Tamaño del tumor despues del tratamiento neoadyuvante					
Hubo respuesta completa o parcial a la terapia en el seno					
Hubo respuesta completa o parcial a la terapia en la axila					
Respuesta clinica por mamografia					

SOBREVIDA

Ha tenido cancer de seno bilateral				Fecha Dx Primer cancer de seno	
				Fecha Dx. Segundo cancer de seno:	
				Edad de Dx. Primer cancer de seno:	
				Edad de Dx. segundo cancer de seno:	
Fecha de Muerte :					
Tiempo en años desde el Dx. Hasta el ingreso al estudio:					
Estado vital:					
Causa de muerte:					
Murio por cancer de seno:					
Ha tenido recaida locorregionales:					
Fecha de la recaida					
Ha tenido metastasis distantes					
Fecha de las metastasis:					
Tiempo en años desde el diagnostico hasta la primera recaida:					

CARACTERISTICAS DEL TUMOR

Tipo 1:				Tipo 2:			
Grado 1:				Grado 2:			
Estado TNM1 :				Estado TNM 2:			
Numero de nodulos afectados1:				Numero de nodulos afectados 2:			
Tamaño del tumor 1 :				Tamaño del tumor 2 :			
RE 1:		%:		RE 2:		%:	
RP 1:		%:		RP 2:		%:	
HER2 1:		%:		HER2 2:		%:	

E. Anexo: Protocolo de extracción de ADN desde sangre periférica

- **Buffer de lisis de glóbulos rojos (100 mL):**

0.83 g de Cloruro de Amonio
0.1 g de Bicarbonato de Potasio
0.2 mL de EDTA 0.5 M pH 8.0
80 mL de agua destilada estéril

Mezclar en un Erlenmeyer con magneto hasta homogenización completa. Completar a 100 ml con agua destilada estéril. Autoclavar. Almacenar a 4°C.

- **Buffer de lisis celular (100 ml):**

5 ml de EDTA 0.5 M pH 8.0
20 ml de SDS 10%
60 ml de agua destilada estéril

Mezclar cuidadosamente (evitando la formación de espuma) en un erlenmeyer hasta homogenización completa. Completar a 100 ml en una probeta con agua destilada estéril. No Autoclavar. Almacenar a temperatura ambiente

- **Solución precipitante de proteínas (50 ml):**

38,54 g de Acetato de Amonio
10 ml de agua destilada estéril

Mezclar en un Erlenmeyer usando magneto hasta completa homogenización. Puede requerir calentamiento ligero. Completar a 50 ml con agua destilada estéril. Autoclavar. Almacenar a 4°C.

- **Solución de rehidratación 10 ml**

TE (Tris-HCL 10 mM EDTA 1mM)
Preparar TRIS 1 M pH 8.0 100ml
PM 121.1g
EDTA 0.5 M pH 8.0 100ml

Y a partir de estos hacer la mezcla TE y dilución correspondientes

• PROCEDIMIENTO

Extracción de ADN

1. En un tubo de 15 mL, medir 8 mL de Buffer de lisis de glóbulos rojos y sobre este adicionar con una pipeta de Pasteur 2 mL de sangre total. Con la misma pipeta mezclar muy bien y con mucha suavidad hasta lograr una completa homogenización.
 2. Por inversión agitar muy suavemente durante 7 minutos a temperatura ambiente.
 3. Centrifugar a 3000 R.P.M. durante 8 minutos.
 4. Descartar el sobrenadante y conservar el pellet.
 5. Mezclar con una pipeta Pasteur muy suavemente el pellet en el líquido residual.
 6. Adicionar 2.5 ml de Buffer de lisis celular y mezclar varias veces, hasta homogenizar. Dejar al menos 20 minutos a temperatura ambiente. (En este paso se puede suspender el procedimiento, incluso hasta el día siguiente, dejando los tubos a 4°C).
 7. Adicionar 800 µl de solución precipitante de proteínas. Mezclar fuerte, hasta observar grumos de las proteínas precipitadas.
 8. Centrifugar a 4000 R.P.M. durante 10 minutos.
 9. Verter todo el sobrenadante sobre un tubo de 15 mL que contenga 2,4ml de isopropanol frío (alcohol isopropílico), dejar en reposo al menos unos 5 minutos.
 10. Mezclar cuidadosamente por inversión hasta precipitar el ADN.
 11. Envolver el ADN en una pipeta y pasarlo rápidamente por etanol al 70% frío, dejar secar a temperatura ambiente y resuspender finalmente en buffer TE (Tris-HCL 10 mM EDTA 1mM) de 50-500 µl, dependiendo de la cantidad de ADN recuperado.
- Nota: El ADN se puede recuperar tomándolo con pipeta Pasteur y centrifugando a 10.000 rpm por 5 min y lavar con 200µl etanol al 70%.

F. Anexo: Extracción de ADN desde tejidos embebidos en parafina.

Protocolo extracción de ADN de tejido normal embebido en parafina:

Desparafinación

1. A 2-3 cortes de tejido de 10 um de tamaño de un bloque de tejido embebido en parafina adicionar 300 ul de xilol a los cortes, mezclar suavemente por inversión durante 5 min
2. Colocar en una plataforma shaker a 37°C por 5 min
3. Centrifugar a 13000 rpm por 3 min
4. Descartar el sobrenadante con pipeta
5. Repetir estos pasos 2 veces
6. Adicionar 300 ul de etanol 100%
7. Colocar en una plataforma shaker a 37°C por 5 min
8. Centrifugar a 14000 rpm por 3 min
9. Descartar el sobrenadante
10. Repetir estos pasos 2 veces
11. Dejar secar el etanol por 1 hora

Extracción

12. Adicionar 500 ul de buffer de lisis celular (extracción de sangre, guardar a T° ambiente) y 16 ul de proteinasa K (25 mg/ml)
13. Incubar a 65°C por 48 horas
14. Adicionar 200 ul de solución precipitante de proteínas (extracción de sangre, guardar a 4°C)
15. Vortex por 20 seg
16. Incubar en hielo por 5 min
17. Centrifugar a 13000 g por 3 min
18. Transferir el sobrenadante a otro tubo
19. Adicionar 600 ul de isopropanol
20. Centrifugar a 16000 g por 5 min

21. Adicionar 300 ul de etanol ala 70%
22. Centrifugar 16000 g por 1 min
23. Resuspender en 30 ul de buffer TE (extracción de sangre, almacenar a T° ambiente)
24. Almacenar a -20°C

Protocolo extracción de ADN de tejido tumoral embebido en parafina ZYMO

Deparaffin

Preparation of paraffin sections:

1. Mount a tissue section with a 10 µm thickness onto a glass slide and dry it at 60°C for 30 minutes.
2. Drop the slide with paraffin embedded tissue in Xylene at room temperature for 30 minutes, change Xylene and incubate another 30 minutes to remove the paraffin.
3. Hydrate slide through 100%, 70%, 50% ethanol, and sterile water for 2 minutes each.
4. Air-dry the slide. It now can be used for DNA isolation using the Pinpoint Slide DNA Isolation System.

Pinpoint Fractionation

The purpose of this procedure is to remove the selected tissue area from the slide.

1. Apply the Pinpoint Solution over the area you want to recover DNA on the prepared slide. Use a pipette tip or a syringe to gently spread a small amount of Pinpoint Solution over the selected region. Use about 0.5 µl of Pinpoint Solution per mm² of tissue area or cover the selected area with a 0.5 mm layer of Pinpoint Solution. It is OK to spread the Pinpoint Solution beyond area that you had selected. It is actually easier to define the exact tissue area for recovery in step 3, see below.

2. Dry the Pinpoint Solution completely at room temperature.

It usually takes about 30 to 45 minutes. The Pinpoint Solution will change into a thin blue film after completely drying; the underlying tissue cells are embedded in the film at this stage.

3. Remove the embedded tissue from the slide.

Use a clean and sharp knife to cut the exact area you want; then peel the area off the slide. Transfer the film to a 0.5 ml tube. Generally, a minimal 1 mm² tissue area with 10 µm thickness (about 200-400 cells depending on different tissue and cell density) is needed for successful gene amplification. The size of each sample spot can vary from 1 to over 20 mm² according to your needs.

1. Spin the tubes briefly.

DNA Extraction

This one buffer extraction procedure will extract DNA from the recovered tissue.

1. Add 50 µl of Extraction Buffer and 5 µl Proteinase K to the tube containing the recovered tissue. Mix gently.

For multiple samples, Extraction Buffer and Proteinase K can be premixed, add 55 µl of this mixture to each sample.

2. Incubate the tubes at 55°C for 4 hours.

Heat the tubes at 95°C-98°C for 10 minutes then immediately put them on ice. This can be easily accomplished on a PCR machine.

Make sure that the temperature is above 95°C. Incomplete inactivation of Proteinase K can cause problems in the PCR reaction.

- 3- Vortex hard for 10-15 seconds. Centrifuge briefly. The DNA is ready for PCR analysis.

DNA Purification

Note: Add 14 ml 95-100% ethanol to the PP Wash Buffer Concentrate to make final PP Wash Buffer.

1. Add 100 μ l Pinpoint Binding Buffer to the above Proteinase K treated DNA sample. Mix briefly.
2. Transfer the mixture to a Zymo-Spin I Column and place the column into a 2 ml Collection Tube.
3. Spin the column and tube at full speed in a microfuge for 10 seconds.
4. Add 150 μ l PP Wash Buffer to the Zymo-Spin I Column and centrifuge at full speed for 10 seconds to wash.

Add 150 μ l PP Wash Buffer and centrifuge at full speed for one minute.

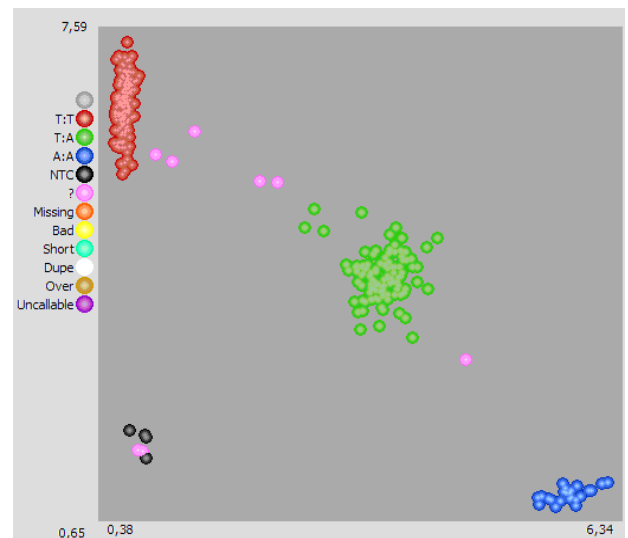
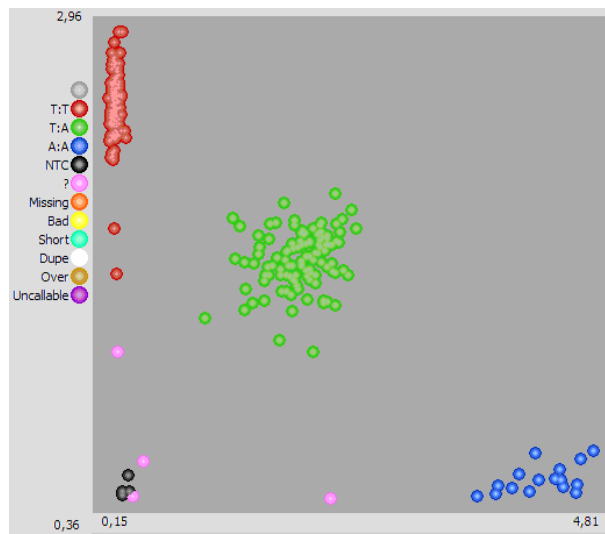
5. Transfer the column into a new 1.5 ml tube.
6. Add 10 μ l water or TE Buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) directly to the membrane of Zymo

Spin I Column. Wait for 1 minute. Spin briefly for 10 seconds to elute the DNA.

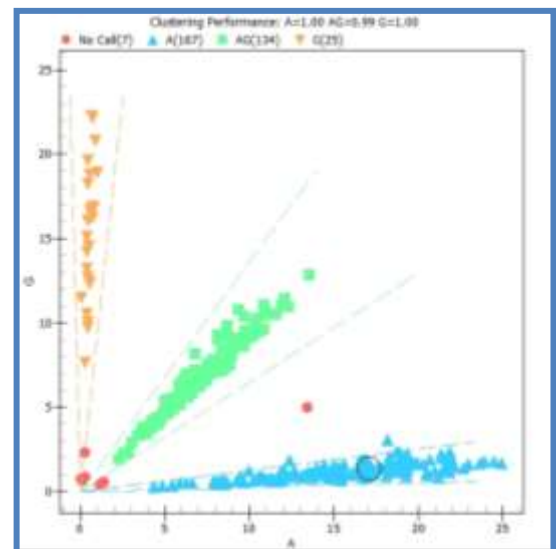
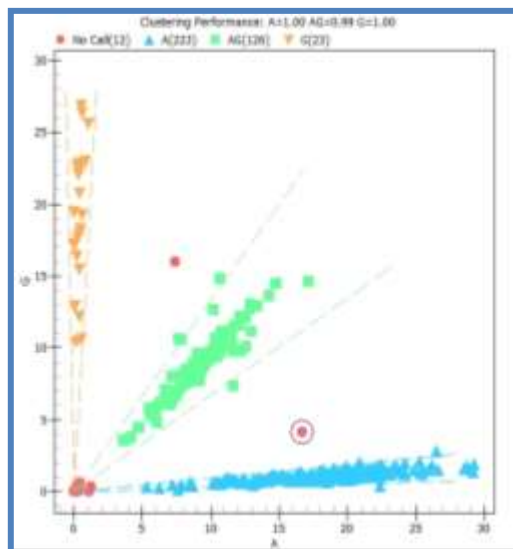
The eluted DNA now can be used for PCR amplification or can be stored at -20°C for future uses. Use 2-4 μ l of the purified DNA for each PCR reaction in 25-50 μ l reaction volume.

G. Anexo: Representación gráfica de los resultados de genotipificación

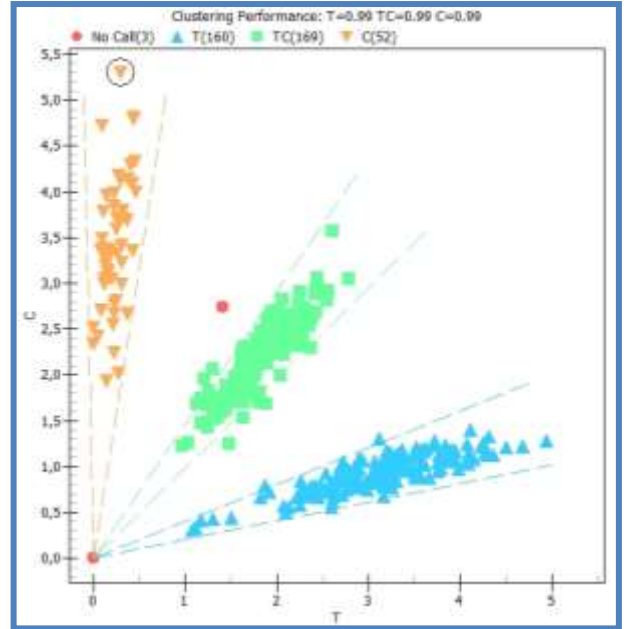
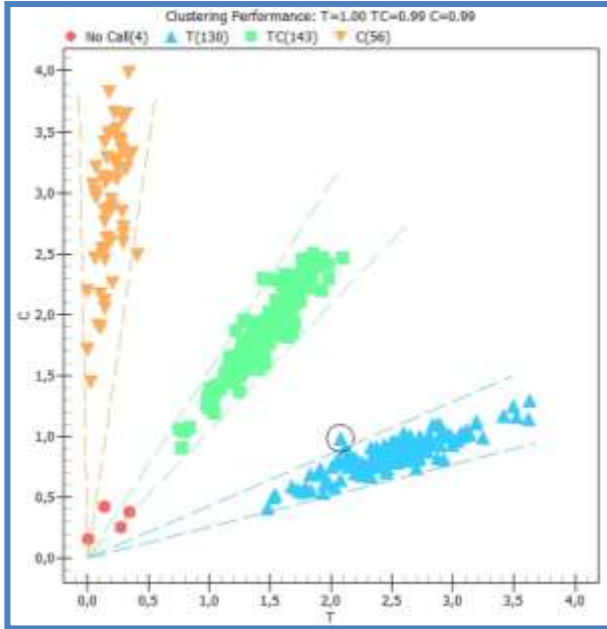
rs599167 genotyping using KASP assay



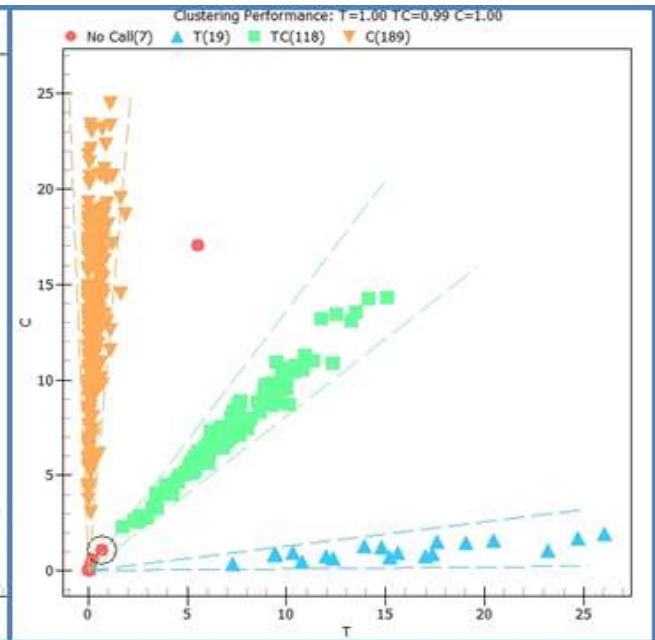
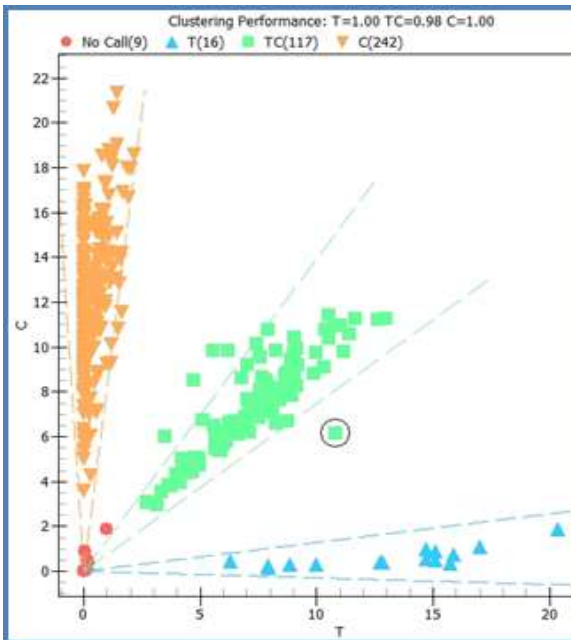
rs584298 genotyping by MALDI-TOF MS



rs2729970 genotyping by MALDI-TOF MS



rs59983645 genotyping by MALDI-TOF MS



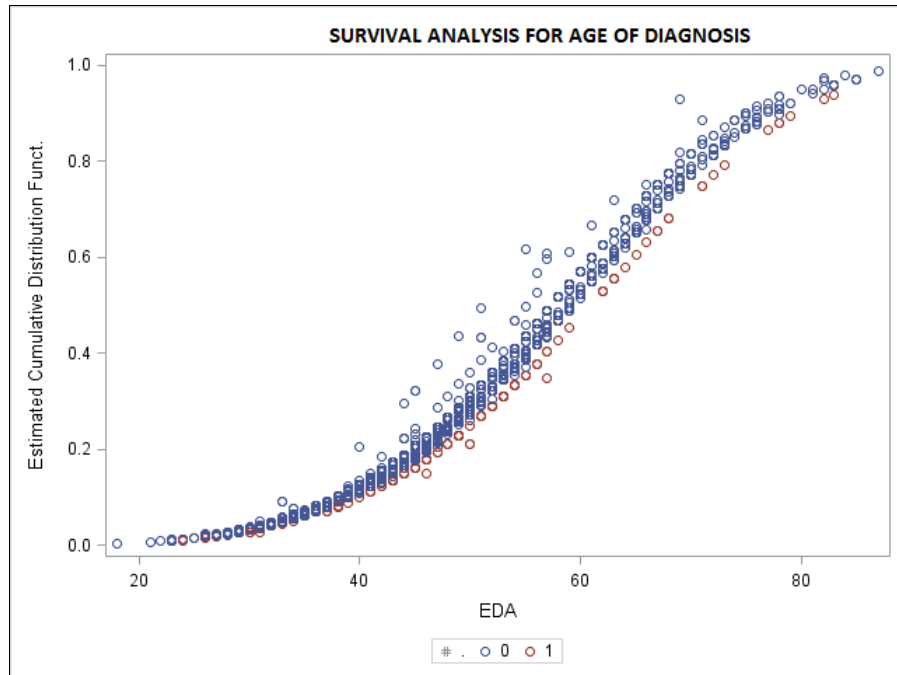
H. Anexo: Análisis de colinealidad

Colinearity Diagnostics

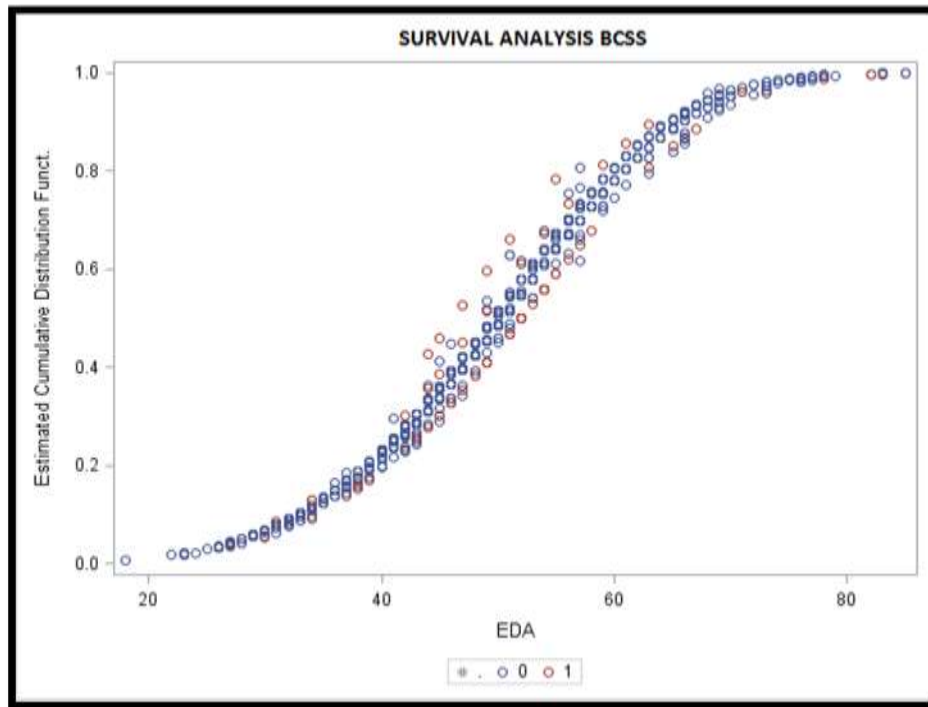
Number	Auto value	Index of the condition	Variation of proportion																			
			Intercept	Smoking	Family history	Age	IMC	Oral contraceptive	hormonal replacement therapy	Menopausal status	rs584298A	rs584298G	rs584298A G	rs2927970C	rs2927970T C	rs2927970T	rs59983645T	rs59983645T C	rs59983645C	rs599167A	rs599167T A	rs599167T
1	866.967	1.00	0.0007177	0.00294	0.00222	0.00039173	0.00029893	0.000319	0.00123	0.00060652	0.00013411	0.00009451	0.00012942	0.00014107	0.00015504	0.00015787	0.00006369	0.0002917	0.00014835	0.00009643	0.00037880	0.00045755
2	330.541	1.61953	6,49E-03	0.00009390	0.00003182	1,41E-03	1,05E-02	0.00003835	0.00004020	4,32E-03	0.0005515	0.00085	0.00038321	0.00203	0.00009212	0.00084030	0.00014	0.00061172	0.00047463	0.00441	0.00152	0.00163
3	296.425	1.71019	5,20E-04	0.00003765	0.00006552	6,44E-02	3,33E-02	0.00003470	0.00010428	1,97E-01	0.00015692	0.00441	0.00106	0.00127	0.00076498	0.00028096	0.00449	0.00113	0.00011145	0.00584	0.00356	0.00041465
4	0.91973	3.07024	0.00002664	0.00682	0.05510	0.00000950	0.00013036	0.00256	0.84214	0.00000153	0.00004593	0.00001664	0.00007934	0.00018046	0.00000927	0.00012866	0.00000279	0.0001354	0.00004655	0.00000134	0.00034531	0.00011018
5	0.81388	3.26378	0.00000450	0.02185	0.91408	0.00002995	0.00002125	0.00486	0.06495	0.00004788	0.00002632	0.00001404	5,97E-02	0.00023684	0.00002941	0.00005709	0.00011540	0.0000398	0.00000921	0.00024026	0.00000775	0.00001585
6	0.72897	3.44863	0.00003965	0.84149	0.01491	0.00043894	0.00022343	0.05183	0.02624	0.00061057	0.00007215	0.00002955	0.00010558	0.00042172	0.00011322	0.00077797	0.00006715	0.0006144	0.00003012	0.00000189	0.00047599	0.00016648
7	0.67836	3.57496	0.00002274	0.06649	0.00112	0.00028204	0.00008798	0.89197	0.00425	0.00028834	0.00000720	7,21E-06	0.00017862	0.00011950	0.00063805	0.00005504	0.00089616	0.00007310	0.00016078	0.00048157	0.00000776	0.00028173
8	0.64618	3.66289	2,23E-01	0.03982	0.00115	4,52E-02	6,30E-03	0.01514	0.00124	0.00000707	1,06E-03	0.00000916	0.00001306	0.03753	0.01370	0.00271	0.01085	0.00176	0.00014262	0.01096	0.00445	0.00055270
9	0.46245	4.32979	0.00000186	0.00849	0.00693	1,84E-03	0.00000878	0.00289	0.00043786	0.00000319	0.00007642	0.00012	0.00058068	0.04739	0.00278	0.02165	0.00504	0.000396	0.00226	0.00700	0.01044	0.00747
10	0.25764	5.80093	6,60E-01	0.00034082	0.00012853	0.00001304	1,93E-01	0.00639	0.00007633	0.00005808	0.00633	0.12050	0.03129	0.00069326	0.00597	0.00550	0.03119	0.00473	0.00016213	0.01752	0.00788	0.00017333
11	0.15034	7.59380	0.00000145	0.00166	0.00012002	0.00020586	7,76E-01	0.00028965	0.00021266	0.00017161	0.02948	0.22421	0.02028	0.01658	0.00054423	0.00597	0.06357	0.00019314	0.00149	0.00977	0.04789	0.01454
12	0.13021	8.15988	5,47E-01	0.00382	0.00013046	0.00002093	6,01E-01	0.00272	0.00049240	8,77E-01	0.00132	0.02659	0.00037192	0.00072920	0.00007755	0.00026899	0.03538	0.06819	0.01829	0.11817	0.17242	0.04220
13	0.08736	9.96177	0.00090855	0.00084751	0.00100121	0.07037	0.00050665	0.00719	0.04772	0.28635	0.00265	0.00546	0.00724	0.00480	0.00410	0.00896	0.00210	0.00395	0.00606	0.01062	0.04214	0.06079
14	0.08022	10.39589	0.00000218	0.00000485	0.00060793	0.00000905	8,74E-01	0.00005871	0.00003220	0.00023272	0.00065912	0.00441	0.00021575	0.00000203	0.00001059	0.00007266	0.36332	0.00768	0.02620	0.59081	0.01392	0.10226

15	0.03519	15.69565	0.00322	0.00023864	0.00053759	0.00472	0.09330	0.00010267	0.00537	0.11156	0.01195	0.00277	0.00616	0.00440	0.00527	0.00447	0.02192	0.01540	0.00721	0.19141	0.59253	0.64093
16	0.02377	19.09768	0.00019094	0.00119	0.00001204	0.39424	0.21864	0.00260	0.00039438	0.37245	0.04098	0.02562	0.04043	0.01906	0.01901	0.01816	0.03075	0.04491	0.04417	0.02815	0.08209	0.07763
17	0.02191	19.89048	0.00006693	0.00000220	0.00020134	0.50492	0.48995	0.00624	0.00481	0.22423	0.01574	0.00548	0.01291	0.00928	0.00624	0.00621	0.01286	0.01767	0.01359	0.00410	0.00544	0.00309
18	0.01243	26.40611	0.01147	0.00018222	0.00121	0.00144	0.01669	0.00091031	0.00000553	0.00050346	0.09531	0.06559	0.10025	0.56372	0.63259	0.63096	0.11198	0.16639	0.15513	0.00000209	0.00790	0.03166
19	0.00750	33.99057	0.00051416	0.00001582	0.00001343	0.00006744	0.00000761	0.00001300	0.00003637	0.00031705	0.73362	0.47439	0.72021	0.00364	0.00637	0.00550	0.28192	0.63529	0.69716	0.00002189	0.00008466	0.00064433
20	0.00452	43.80390	0.98346	0.00367	0.00023043	0.02284	0.18014	0.00095485	0.00021267	0.00256	0.06088	0.03443	0.05810	0.28777	0.30152	0.28727	0.02035	0.02745	0.02715	0.00038317	0.00653	0.01500

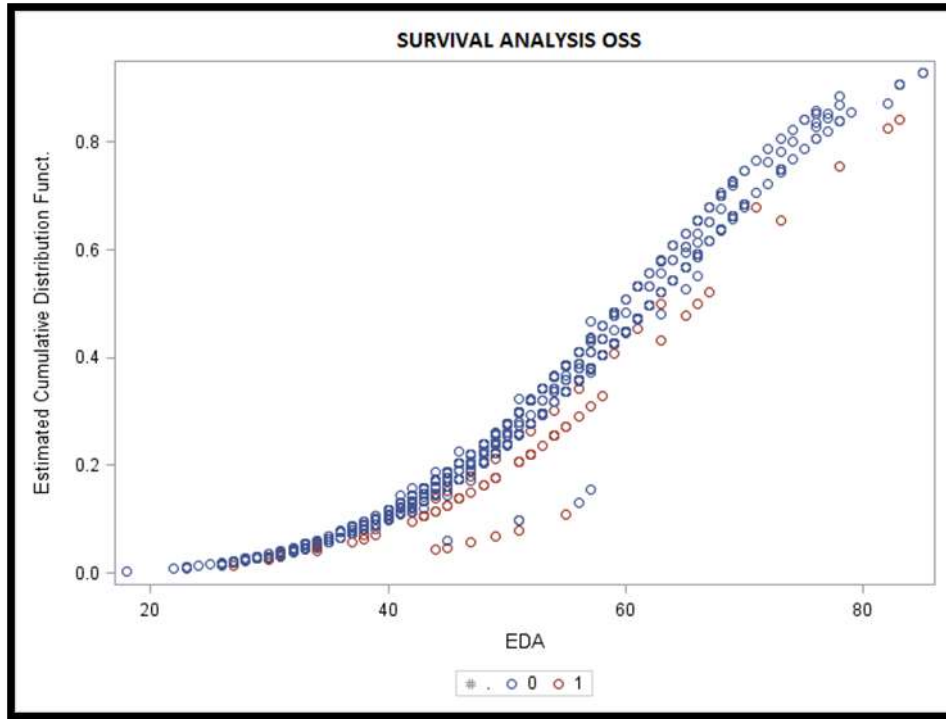
I. Anexo: Análisis de supervivencia



analysis of the estimator of maximum likelihood parameter age of diagnosis								
Parámetro		DF	Estimator	Standard error	IC		Chi-Square	Pr > ChiSq
					95%			
Intercept		1	41.599	0.0357	40.898	42.299	13547.0	<.0001
rs584298G	0	1	0.0516	0.0599	-0.0658	0.1691	0.74	0.3888
rs584298G	1	0	0.0000
rs584298AG	0	1	0.0036	0.0303	-0.0558	0.0629	0.01	0.9058
rs584298AG	1	0	0.0000
rs2927970TC	0	1	0.0071	0.0194	-0.0309	0.0452	0.14	0.7131
rs2927970TC	1	0	0.0000
rs59983645T	0	1	-0.1393	0.0846	-0.3051	0.0265	2.71	0.0997
rs59983645T	1	0	0.0000
rs59983645TC	0	1	0.0349	0.0382	-0.0400	0.1098	0.83	0.3613
rs59983645TC	1	0	0.0000
rs599167A	0	1	0.0590	0.0851	-0.1079	0.2258	0.48	0.4884
rs599167A	1	0	0.0000
rs599167TA	0	1	-0.0152	0.0402	-0.0941	0.0636	0.14	0.7054
rs599167TA	1	0	0.0000
Escal		1	0.2219	0.0054	0.2116	0.2327		
Forma Weibull		1	45.061	0.1095	42.966	4		



Analysis of the estimator of maximum likelihood parameter BCSS								
Parámetro	DF	Estimator	Standard error	IC		Chi-Square	Pr > ChiSq	
				95%				
Intercept	1	40.135	0.0363	39.424	40.845	12254.1	<.0001	
rs584298G	0	-0.0030	0.0574	-0.1156	0.1096	0.00	0.9586	
rs584298G	1	0.0000	
rs584298AG	0	-0.0395	0.0302	-0.0987	0.0198	1.70	0.1917	
rs584298AG	1	0.0000	
rs2927970TC	0	0.0167	0.0200	-0.0225	0.0560	0.70	0.4028	
rs2927970TC	1	0.0000	
rs59983645T	0	-0.0826	0.0837	-0.2465	0.0814	0.97	0.3236	
rs59983645T	1	0.0000	
rs59983645TC	0	0.0361	0.0335	-0.0295	0.1018	1.16	0.2809	
rs59983645TC	1	0.0000	
rs599167A	0	0.0558	0.0791	-0.0992	0.2107	0.50	0.4807	
rs599167A	1	0.0000	
rs599167TA	0	0.0056	0.0347	-0.0624	0.0737	0.03	0.8711	
rs599167TA	1	0.0000	
Escal	1	0.2207	0.0054	0.2103	0.2316			
Forma Weibull	1	45.306	0.1113	43.176	47.541			



Analysis of the estimator of maximum likelihood parameter OSS								
Parámetro	DF	Estimator	Standard error	IC		Chi-Square	Pr > ChiSq	
				95%				
Intercept	1	42.359	0.0642	41.100	43.617	4350.91	<.0001	
rs584298G	0	-0.0571	0.1082	-0.2692	0.1550	0.28	0.5978	
rs584298G	1	0.0000	
rs584298AG	0	-0.0262	0.0498	-0.1238	0.0714	0.28	0.5988	
rs584298AG	1	0.0000	
rs2927970TC	0	0.0234	0.0334	-0.0420	0.0888	0.49	0.4830	
rs2927970TC	1	0.0000	
rs59983645T	0	0.2416	0.2524	-0.2531	0.7363	0.92	0.3385	
rs59983645T	1	0.0000	
rs59983645TC	0	0.0029	0.0537	-0.1024	0.1082	0.00	0.9563	
rs59983645TC	1	0.0000	
rs599167A	0	-0.2454	0.2450	-0.7257	0.2348	1.00	0.3165	
rs599167A	1	0.0000	
rs599167TA	0	0.0402	0.0570	-0.0716	0.1519	0.50	0.4814	
rs599167TA	1	0.0000	
Escal	1	0.2343	0.0093	0.2169	0.2532			
Forma Weibull	1	42.672	0.1685	39.494	46.106			

Anexo J: Modelo final con selección de variables usando la aproximación Stepwise

Tag SNPs	Genotypes	frequencies Patients		FULL MODEL			REDUCED MODEL		
		Cases	Controls	OR	95% CI	p-Value	OR*	95% CI*	p-Value*
rs584298	AA	498	541						
	GG	74	63	2,24	0,88-5,70	0,09	2,57	1,35-4,90	0,004
	AG	371	360	1,44	0,86-2,41	0,164	1,41	1,00-2,00	0,048
rs2927970	CC	149	132						
	TT	366	392	0,9	0,44-1,88	0,798			
	TC	425	440	0,85	0,44-1,62	0,624			
	Age			7,15	5,21-9,82	0,000	7,28	5,32-9,97	0,000
	Menopausal Status			2,25	1,36-3,73	0,002	2,28	1,39-3,74	0,001
	BMI			0,99	0,997-0,999	0,004	0,99	0,997-0,999	0,004
Risk Variables	Family History			4,31	2,68-6,92	0,000	4,54	2,84-7,26	0,000
	Parous			0,65	0,28-1,50	0,323			
	Smoking			0,98	0,48-2,00	0,967			
	Breast feeding			1,07	0,52-2,20	0,847	1,41	0,93-2,15	0,100
	OC ever			1	0,701-1,43	0,993			
	HRTE ever			2,27	1,16-4,43	0,150	2,4	1,25-4,58	0,008

