

**UNIVERSIDAD DE CHILE  
FACULTAD DE MEDICINA  
ESCUELA DE POSTGRADO**



**ANÁLISIS DE LIGAMIENTO Y ASOCIACIÓN PARA IDENTIFICAR  
MARCADORES GENÉTICOS DE RIESGO DE COLELITIASIS Y CÁNCER DE VESÍCULA BILIAR EN  
CHILE**

**PABLO FELIPE BÁEZ BENAVIDES**

**TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS BIOMEDICAS**

**Directores de Tesis: Prof. Dra. Katherine Marcelain Cubillos  
Prof. Dr. Justo Lorenzo Bermejo**

**2018**

**UNIVERSIDAD DE CHILE  
FACULTAD DE MEDICINA  
ESCUELA DE POSTGRADO**

**INFORME DE APROBACION TESIS DE  
DOCTORADO EN CIENCIAS BIOMÉDICAS**

**Se informa a la Comisión de Grados Académicos de la Facultad de Medicina, que la Tesis de Doctorado en Ciencias Biomédicas presentada por el candidato**

**PABLO FELIPE BAEZ BENAVIDES**

**ha sido aprobada por la Comisión Informante de Tesis como requisito para optar al Grado de Doctor en Ciencias Biomédicas en Examen de Defensa de Tesis rendido el día 4 de mayo de 2018.**

**Prof. Dra. Katherine Marcelain Cubillos  
Dpto. de Oncología Básico Clínica, Facultad de Medicina, Universidad de Chile**

**Prof. Dr. Justo Lorenzo Bermejo  
Statistical Genetics Group, Institute of Medical Biometry and Informatics,  
Heidelberg University, Germany.**

**Directores de Tesis**

**COMISION INFORMANTE DE TESIS**

**PROF. DRA. GABRIELA REPETTO**

**PROF. DR. PATRICIO GONZÁLEZ**

**PROF. DR. RICARDO VERDUGO**

**PROF. DRA. LUCIA CIFUENTES  
Presidente Comisión de Examen**

## FINANCIAMIENTO

Esta tesis de doctorado fue realizada en el laboratorio de Genómica del Cáncer, perteneciente al Departamento de Oncología Básico-Clínico de la Facultad de Medicina, Universidad de Chile. Y contó con el siguiente financiamiento:

1. Ministerio Alemán de Educación e Investigación (BMBF), *grant* 01DN15021 (Justo Lorenzo).
2. FONDEF IT16I10051. “TumorSec: Optimization, validation and widespread use of a gene sequencing test to improve the accuracy of cancer therapies” (Katherine Marcelain).
3. FONDEF D11I1029 “Incorporation of Next Generation Sequencing on the cancer patient’s care” (Katherine Marcelain).
4. Beca de Doctorado para Extranjeros en Chile Año Académico 2013, CONICYT 63130089 (Pablo Báez).
5. Beca CONICYT Nacional 63130089, Beneficio de gastos operacionales 2015 (Pablo Báez).
6. Beca CONICYT Nacional 63130089, Beneficio complementario de pasantía doctoral 2015 (Pablo Báez).
7. Becas de pasantías en el extranjero Programa MECESUP Uch1304 2015 (Pablo Báez).
8. Beca CONICYT Nacional 63130089, Beneficio de gastos operacionales 2016 (Pablo Báez).

## ÍNDICE

ÍNDICE.....	IV
ÍNDICE DE FIGURAS .....	VII
ÍNDICE DE TABLAS .....	IX
LISTA DE SÍMBOLOS Y ABREVIATURAS .....	XI
RESUMEN.....	XIV
ABSTRACT .....	XVI
<b>I. INTRODUCCIÓN .....</b>	<b>18</b>
<b>1. Cáncer: Antecedentes Generales.....</b>	<b>18</b>
<b>2. Cáncer de Vesícula Biliar: Generalidades y Factores de Riesgo .....</b>	<b>20</b>
2.1. Definición del Cáncer de Vesícula Biliar .....	20
2.2. Epidemiología del Cáncer de Vesícula Biliar.....	20
2.3. Diagnóstico, Pronóstico y Tratamiento del Cáncer de Vesícula Biliar.....	21
2.4. Etiología del Cáncer de Vesícula Biliar .....	23
2.4.1. Edad y sexo como factores de riesgo de CVB. ....	23
2.4.2. Colelitiasis y riesgo de CVB. ....	24
2.4.3. Susceptibilidad genética al CVB. ....	26
<b>3. Fundamentos del Análisis de Ligamiento Genético .....</b>	<b>29</b>
3.1. Enfoques para el Análisis de Ligamiento.....	29
3.1.1. Análisis de ligamiento paramétrico. ....	30
3.1.2. Análisis de ligamiento no paramétrico. ....	31
3.2. Herramientas Informáticas para la Estimación del Ligamiento .....	32
<b>4. Priorización de Variantes .....</b>	<b>33</b>
<b>5. Mapeo Fino por Asociación .....</b>	<b>33</b>
<b>II. Hipótesis y Objetivos .....</b>	<b>35</b>
<b>1. Hipótesis .....</b>	<b>35</b>
<b>2. Objetivos.....</b>	<b>35</b>
2.1. Objetivo General .....	35
2.2. Objetivos Específicos.....	35
<b>III. Individuos y Métodos .....</b>	<b>36</b>
<b>1. Diseño Experimental.....</b>	<b>36</b>
<b>2. Familias con Múltiples Casos de Cáncer de Vesícula Biliar .....</b>	<b>38</b>
<b>3. Pacientes con Cáncer de Vesícula Biliar e Individuos Control.....</b>	<b>49</b>
<b>4. Métodos .....</b>	<b>52</b>
4.1. Aprobación Ética .....	52
4.2. Procesamiento de las Muestras .....	52
4.3. Genotipificación, Control de Calidad y Estructura de los Datos .....	53
4.4. Análisis de Ligamiento Genético .....	54



4.4.1.	Métodos de ligamiento genético .....	54
4.4.2.	Análisis de sensibilidad .....	56
4.4.2.1.	Contribuciones por edad de diagnóstico, colecistectomía y fenotipo .....	56
4.4.2.2.	Contribución de cada familia a los resultados de ligamiento .....	57
4.4.2.3.	Variación de los parámetros en el ligamiento paramétrico .....	57
4.5.	Búsqueda de Genes Candidatos por Anotación Funcional .....	58
4.6.	Métodos de Mapeo Fino por Asociación Genética .....	58
<b>IV.</b>	<b>Resultados .....</b>	<b>59</b>
<b>1.</b>	<b>Análisis Descriptivo de las Familias y los Individuos .....</b>	<b>59</b>
<b>2.</b>	<b>Estimación del riesgo de recurrencia y riesgo relativo en las familias.....</b>	<b>63</b>
<b>3.</b>	<b>Análisis de Ligamiento Genético.....</b>	<b>67</b>
3.1.	Ligamiento Paramétrico .....	67
3.2.	Ligamiento No Paramétrico. ....	70
3.3.	Análisis de Sensibilidad .....	74
3.3.1.	Modelos con restricción por edad de diagnóstico o colecistectomía, y considerando dos definiciones de individuos afectados (CVB y CVB + CCA). ....	74
3.3.2.	Contribuciones de cada familia a los resultados del análisis de ligamiento. ....	79
3.3.3.	Variación de los parámetros en el ligamiento paramétrico.....	82
<b>4.</b>	<b>Búsqueda de Genes Candidatos mediante Anotación Funcional.....</b>	<b>82</b>
4.1.	Análisis de Clasificación y Enriquecimiento Funcional.....	82
4.2.	Análisis de las Categorías de Anotación de DAVID Individualmente .....	83
4.2.1.	Anotación por enfermedades .....	83
4.2.2.	Anotación en Categorías funcionales basadas en palabras clave.....	86
<b>5.</b>	<b>Mapeo Fino por Asociación Genética .....</b>	<b>89</b>
<b>V.</b>	<b>Discusión .....</b>	<b>96</b>
<b>VI.</b>	<b>Conclusiones .....</b>	<b>107</b>
<b>VII.</b>	<b>Material Científico Generado .....</b>	<b>108</b>
<b>1.</b>	<b>Publicaciones .....</b>	<b>108</b>
<b>2.</b>	<b>Presentaciones .....</b>	<b>108</b>
<b>VIII.</b>	<b>Material Suplementario .....</b>	<b>109</b>
<b>1.</b>	<b>Tablas .....</b>	<b>109</b>
<b>2.</b>	<b>Figuras .....</b>	<b>111</b>
<b>3.</b>	<b>Encuesta.....</b>	<b>130</b>
<b>4.</b>	<b>Certificado de Aprobación Ética .....</b>	<b>137</b>
<b>5.</b>	<b>Consentimiento Informado .....</b>	<b>139</b>
<b>6.</b>	<b>Estructura de los datos .....</b>	<b>143</b>
<b>7.</b>	<b>Código Fuente (Mendel) .....</b>	<b>145</b>
7.1.	Código Fuente 1. Análisis de Ligamiento Paramétrico en 7 familias Chilenas con Agregación del CVB .....	145

7.2.	Código Fuente 2. Análisis de Ligamiento No Paramétrico en 7 familias Chilenas con Agregación del CVB .....	146
7.3.	Código Fuente 3. Análisis de Ligamiento Paramétrico Multipuntos en 7 familias Chilenas con Agregación del CVB .....	148
<b>8.</b>	<b>Código Fuente (R) .....</b>	<b>151</b>
8.1.	Código Fuente 4. Preparación de Archivos de Entrada Necesarios Para el Análisis de Ligamiento Genético en el Programa Mendel .....	151
8.2.	Código Fuente 5. Control de Calidad de 1.161 SNPs Genotipificados en Pacientes con CVB e Individuos Sanos.....	154
8.3.	Código fuente 6. Asociación Genética Mediante Regresión Logística Binaria en Pacientes con CVB e Individuos Sanos.....	162
<b>IX.</b>	<b>Referencias .....</b>	<b>169</b>

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Visión global de la contribución genética al cáncer. ....	19
<b>Figura 2.</b> Modelo etiológico del Cáncer de Vesícula Biliar. ....	23
<b>Figura 3.</b> Diseño del estudio. ....	37
<b>Figura 4.</b> Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar. ....	40
<b>Figura 5.</b> Mapa de Chile representando los lugares de nacimiento de todos los individuos del estudio. ....	50
<b>Figura 6.</b> Evaluación de la integridad del ADN genómico a través de electroforesis en gel de agarosa. ....	52
<b>Figura 7.</b> Curvas de puntuación LOD del análisis de ligamiento del genoma completo para el modelo I (alelo de susceptibilidad al CVB) en 7 familias chilenas con múltiples casos de CVB. ....	68
<b>Figura 8.</b> Curvas de puntuación LOD del análisis de ligamiento del genoma completo para el modelo II (alelo de susceptibilidad al CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados. ....	69
<b>Figura 9.</b> Test de ligamiento no paramétrico del genoma completo para el modelo I (alelo de susceptibilidad al CVB) en 7 familias chilenas con múltiples casos de CVB. ....	71
<b>Figura 10.</b> Test de ligamiento no paramétrico del genoma completo para el modelo II (alelo de susceptibilidad a CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados. ....	72
<b>Figura 11.</b> Resultados del test de ligamiento no paramétrico sobre el cromosoma 9 para el modelo II (alelo de susceptibilidad a CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados. ....	73
<b>Figura 12.</b> Resultados del test de ligamiento no paramétrico sobre el cromosoma 16 para el modelo II (alelo de susceptibilidad a CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados. ....	73
<b>Figura 13.</b> Efecto de la restricción por edad e inclusión de individuos colecistectomizados, sobre las máximas puntuaciones LOD por cromosoma obtenidas a través de análisis de ligamiento paramétrico. ....	76
<b>Figura 14.</b> Efecto de la restricción por edad e inclusión de individuos colecistectomizados, sobre los máximos valores p ( $-\log_{10}$ ) por cromosoma obtenidos a través de análisis de ligamiento no paramétrico. ....	78
<b>Figura 15.</b> Contribuciones acumuladas de las familias a las señales de ligamiento. ....	81
<b>Figura 16.</b> Redes de interacción molecular para 25 genes que podrían albergar variantes de riesgo para el desarrollo de CVB. ....	88

<b>Figura 17.</b> Manhattan plot del mapeo fino por asociación sobre el cromosoma 9 suponiendo un modelo genético dominante, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.....	90
<b>Figura 18.</b> Manhattan plot del mapeo fino por asociación sobre el cromosoma 16 suponiendo un modelo genético dominante, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.....	90
<b>Figura 19.</b> Manhattan plot del mapeo fino por asociación sobre el cromosoma 9 suponiendo un modelo genético aditivo, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.....	91
<b>Figura 20.</b> Manhattan plot del mapeo fino por asociación sobre el cromosoma 16 suponiendo un modelo genético aditivo, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.....	91
<b>Figura 21.</b> Análisis de asociación en los genes ABCA1, SEC14L5 y PTPN3, en 183 casos con CVB y 1.962 controles de población chilena.....	94

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Genes candidatos analizados en estudios de asociación en pacientes con CVB provenientes de diversas poblaciones. ....	27
<b>Tabla 2.</b> Variantes de susceptibilidad al CVB identificadas mediante estudios de asociación del genoma completo (GWAS). ....	28
<b>Tabla 3.</b> Razón de incidencia estandarizada (RIS) para el CVB en descendientes de padres diagnosticados con CVB entre 1961-2012. ....	29
<b>Tabla 4.</b> Características generales de los individuos incluidos en el análisis de genealogías y ligamiento. ....	39
<b>Tabla 5.</b> Riesgos relativos (RR) identificables para diferentes frecuencias alélicas en una muestra de 183 casos con CVB y 1.962 controles poblacionales. ....	51
<b>Tabla 6.</b> Estadística descriptiva para 12 variables con información completa, analizadas en individuos con CVB, CCA y NAF pertenecientes a 7 familias. ....	59
<b>Tabla 7.</b> Promedio estimado de las proporciones de ascendencia europea, africana, mapuche y aymara en 7 familias chilenas con múltiples casos de CVB. ....	60
<b>Tabla 8.</b> Distribución de los individuos por sexo, rango de edad, nivel de educación y componentes de ascendencia, incluidos en el análisis de mapeo fino por asociación en dos regiones candidatas de ligamiento. ....	61
<b>Tabla 9.</b> Distribución de los casos con CVB y controles por sexo, edad y proporción de ascendencia mapuche. ....	63
<b>Tabla 10.</b> Riesgo relativo de <u>CVB</u> para individuos con familiares afectados del primer al tercer grado de consanguinidad en 16 familias chilenas. ....	64
<b>Tabla 11.</b> Riesgo relativo de <u>colecistitis</u> para individuos con familiares afectados del primer al tercer grado de consanguinidad en 16 familias chilenas. ....	65
<b>Tabla 12.</b> Riesgo relativo de <u>CVB</u> para individuos con colecistitis y antecedentes familiares de CVB en 16 familias chilenas. ....	66
<b>Tabla 13.</b> Mejores resultados por cromosoma (Cr), obtenidos mediante análisis de ligamiento del genoma completo en 7 familias con múltiples casos de CVB y colecistectomía. ....	67
<b>Tabla 14.</b> Máximas puntuaciones LOD por cromosoma obtenidas en 12 modelos definidos para el análisis de sensibilidad del ligamiento genético. ....	74
<b>Tabla 15.</b> Puntuaciones LOD sugerentes de ligamiento en cuatro cromosomas obtenidas mediante análisis de ligamiento paramétrico en dos modelos con restricción de edad en 7 familias chilenas con múltiples casos de CVB. ....	75
<b>Tabla 16.</b> Máximos valores p ( $-\log_{10}$ ) por cromosoma obtenidas en 12 modelos definidos para el análisis de sensibilidad del ligamiento genético. ....	77
<b>Tabla 17.</b> Contribución individual de cada genealogía a las señales de ligamiento no paramétrico en 7 familias chilenas con múltiples casos de CVB. ....	79

<b>Tabla 18.</b> Contribuciones individuales y acumuladas de las familias a las señales de ligamiento no paramétrico en los cromosomas 9 y 16. ....	79
<b>Tabla 19.</b> Lista de 95 genes contenidos en la región de ligamiento chr9:105016960-115983216 identificada al analizar 7 familias chilenas con múltiples casos de CVB y cálculos biliares. ....	82
<b>Tabla 20.</b> Lista de 33 genes contenidos en la región de ligamiento chr16:105444-5995388 identificada al analizar 7 familias chilenas con múltiples casos de CVB y cálculos biliares. ....	83
<b>Tabla 21.</b> Lista de genes contenidos en las regiones de ligamiento identificadas y que están relacionados con enfermedades humanas registradas en la base de datos OMIM. ....	84
<b>Tabla 22.</b> Genes anotados en GAD de acuerdo a 11 fenotipos/enfermedades asociados con CVB y cálculos biliares. ....	85
<b>Tabla 23.</b> Genes anotados en categorías funcionales basadas en palabras claves. ....	86
<b>Tabla 24.</b> Genes potencialmente relevantes dentro de la etiopatogenia del CVB y cálculos biliares. ....	87
<b>Tabla 25.</b> Resumen de variantes asociadas con el riesgo de CVB y sus odds ratios estimados identificadas en 4 estudios de asociación. ....	93

## LISTA DE SÍMBOLOS Y ABREVIATURAS

$\lambda$ : Lambda (Riesgo relativo) .....	38
$\alpha$ [Alpha]: Proporción de marcadores ligados .....	55
$\theta$ [Theta]: Fracción de Recombinación .....	30
<i>ABCA1</i> : ATP Binding Cassette Subfamily A Member 1 .....	83
<i>ABCB1</i> : ATP binding cassette subfamily B member 1 .....	28
<i>ABCB4</i> : ATP binding cassette subfamily B member 4 .....	28
ALNP: Análisis de ligamiento no paramétrico .....	67
ALP: Análisis de ligamiento paramétrico.....	67
ASP: Pares de hermanos afectados.....	31
<i>BAAT</i> : Ácido biliar-CoA; aminoácido N-aciltransferasa .....	83
CCA: Colectomía .....	56
<i>CDKN2A</i> : Gen del inhibidor 2A de quinasa dependiente de ciclina .....	20
COX2: Ciclooxygenasa 2 .....	20
<i>CREBBP</i> : Gen de la proteína de unión a CREB.....	87
CVB: Cáncer de vesícula biliar .....	20
DAVID: Database for Annotation, Visualization and Integrated Discovery .....	33
<i>DCC</i> : Gen Delecionado en Cáncer de Colon .....	27
DE: Desviación estándar.....	47
<i>DEC1</i> : Gen delecionado en cáncer esofágico 1 .....	83
DL: Desequilibrio de ligamiento .....	33
<i>FHIT</i> : Gen frágil de la tríada de la histidina .....	20
GAD: Genetic Association Database.....	83
GCTA: Acrónimo en inglés de Genome-Wide Complex Trait Analysis tool.....	28
GES: Garantías Explícitas en Salud .....	22

GWAS: Acrónimo en inglés de Genome-wide association study .....	27
IBD: Idéntico por descendencia .....	31
IBS: Idéntico por estado .....	31
ICD-O: Acrónimo en inglés de International Classification of Diseases for Oncology.....	20
<i>KMT2C</i> : Gen de la lisina metiltransferasa 2C .....	28
<i>KRAS</i> : Gen homólogo del oncogen viral del sarcoma de la rata Kirsten .....	20
MAF: Frecuencia del alelo menor .....	53, 112
Mb: Megabase, 1 millón de bases.....	36
NAF: No afectado .....	59
NGS: Acónimo en inglés de Next Generation Sequencing .....	28
OMIM: Online Mendelian Inheritance in Man.....	83
PANTHER: Protein ANalysis THrough Evolutionary Relationships .....	33
<i>PTPN3</i> : Protein Tyrosine Phosphatase, Non-Receptor Type 3.....	92
$\phi$ : Phi, Configuración fenotípica de una genealogía especial en la función de Score .....	32
R0: Cirugía que permite alcanzar la remisión completa .....	21
$r^2$ : Coeficiente de correlación al cuadrado, medida del desequilibrio de ligamiento.....	53
RIS: Razón de incidencia estandarizada .....	29
RME: Razón de mortalidad estandarizada .....	21
RRF: Riesgo relativo familiar .....	28
S: Función de Score .....	32
<i>SEC14L5</i> : SEC14 Like Lipid Binding 5 .....	92
<i>SMAD4</i> : Gen del Miembro 4 de la Familia SMAD .....	28
SNP: Acrónimo en inglés de Single Nucleotide Polymorphism .....	27
SR-BI: Receptor scavenger clase B tipo I .....	100
<i>TAL2</i> : Factor de transcripción TAL BHLH 2 .....	83
TCGA: Acrónimo en inglés de The Cancer Genome Atlas .....	100



<i>TLR4</i> : Gen del Receptor tipo Toll 4.....	83
TNFSF15: Miembro de la superfamilia del factor de necrosis tumoral 15.....	86
TNFSF8: Miembro de la superfamilia del factor de necrosis tumoral 8.....	86
<i>TP53</i> : Gen de la Proteina Tumoral P53 .....	20
UPDB: Acrónimo en inglés de Utah Population Database .....	29
v: Vector de herencia de la función de Score.....	32

## RESUMEN

El cáncer de vesícula biliar (CVB) es un tumor maligno con una sobrevida de cinco años inferior al 15%. Mientras que en el mundo se presenta en menos de 2/100.000 individuos, en Chile las tasas de incidencia son superiores a 25 casos por cada 100.000 personas-años en mujeres y 9 casos por cada 100.000 personas-años en hombres. El CVB fue responsable de 1.316 muertes en 2015 (931 mujeres y 385 hombres). El conocimiento actual sobre los desencadenantes genéticos y ambientales de este tumor es muy limitado. La coleditiasis es considerada el principal factor de riesgo para el CVB. La existencia de un componente genético que predispone al desarrollo de estas patologías se sustenta en dos observaciones principales: (1) la amplia asociación de la etnia con la prevalencia de coleditiasis y CVB y (2) la agregación familiar observada para ambas enfermedades. A pesar del gran impacto sanitario de esta neoplasia en Chile y de la probable existencia de variantes genéticas de susceptibilidad, se han llevado a cabo pocos estudios y aún no se han identificado marcadores moleculares que permitan predecir el riesgo individual de desarrollar CVB y prevenir esta agresiva enfermedad en Chile. En este sentido, es necesario identificar marcadores de susceptibilidad del desarrollo de coleditiasis y CVB en población chilena. El objetivo final de esta tesis fue identificar variantes genéticas que se asocien con la coleditiasis y cáncer de vesícula biliar con agregación familiar, y con el riesgo de cáncer de vesícula biliar esporádico, aplicando una estrategia con tres pasos: (1) análisis de ligamiento genético en familias, (2) búsqueda de genes candidatos en las regiones de ligamiento utilizando anotaciones funcionales y (3) mapeo fino por asociación investigando los genotipos de casos y controles poblacionales. En total se investigaron los datos demográficos y genotípicos de 2.193 individuos: 48 individuos (6 casos de CVB, 15 pacientes colecistectomizados y 27 individuos no afectados) pertenecientes a 7 familias con múltiples casos de CVB; 183 pacientes con CVB esporádico y 1.962 controles poblacionales. El análisis de ligamiento no paramétrico identificó dos regiones candidatas que contenían un total de 128 genes: Chr9:105016960-115983216 con 95 genes, y Chr16:105444-5995388 con 33 genes. Mediante un análisis integrativo por anotación funcional se redujo la lista total a 30 genes potencialmente relevantes dentro de la etiopatogenia del CVB. El análisis de asociación en

los genes candidatos reveló dos variantes asociadas con el riesgo de CVB esporádico: rs363717 (OR=1,84 95% IC 1,26-2,66) del gen *ABCA1* y rs1558562 (OR=2,02 95% IC 1,4-2,95) del gen *SEC14L5*. Los resultados obtenidos sugieren que tanto la agregación familiar del CVB y la colelitiasis, como la susceptibilidad al CVB esporádico en chilenos está asociada al menos en parte, a variantes genéticas localizadas en *ABCA1* y *SEC14L5*.

## ABSTRACT

Gallbladder cancer (GBC) is a very aggressive tumor with a five-year survival rate under 15%. Worldwide GBC is considered a rare disease, but in Chile incidence rates amount to 25 cases per 100,000 person-years in women and 9 cases per 100,000 person-years in men. GBC caused 1.316 deaths in Chile in 2015 (931 women and 385 men). Current knowledge about the genetic and environmental triggers of this tumor is sparse. Gallstones disease is considered the main risk factor for CVB. Two observations support the important role of a genetic component that predisposes to these pathologies: (1) the broad association between ancestry and gallstones/GBC prevalence, and (2) the familial clustering observed for both diseases. However, genetic variants associated to GBC susceptibility in Chile have not yet been identified. Despite the large health impact of this cancer in Chile, few studies have been carried out in its population. In this sense, it is necessary to identify gallstones and GBC susceptibility markers in the Chilean population. The final objective of this thesis was to identify genetic variants associated with gallstones and GBC with familial clustering, and with the risk of sporadic GBC, through a three-step analysis: (1) genetic linkage analysis using parametric and non-parametric linkage methods in families with multiple GBC cases; (2) search for candidate genes through an integrative functional annotation analysis of the harbored genes in the identified linkage regions; and (3) fine-mapping by genetic association based on cases and population controls and using 1,028 variants distributed throughout the candidate regions. A total of 2,193 individuals were analyzed: 48 individuals (six cases of GBC, 15 cholecystectomized, 27 unaffected) belonging to 7 families with multiple GBC cases, 183 sporadic GBC patients and 1,962 population controls. Two candidate linkage regions were identified using non-parametric linkage analysis: Chr9: 105016960-115983216 and Chr16: 105444-5995388, which harbored 128 coding genes: 95 on chromosome 9 and 33 on chromosome 16. By functional annotation, the list was reduced to 30 genes potentially relevant for the disease etiopathogenesis. Candidate gene association analysis revealed two variants associated with sporadic GBC risk: *ABCA1* rs363717 (OR = 1.84 95% CI 1.26-2.66) and *SEC14L5* rs1558562 (OR = 2.02 95% CI 1.4-2.95). These results suggest that both GBC and

gallstones with familial clustering, as well as sporadic GBC susceptibility in Chile, is associated at least in part with *ABCA1* and *SEC14L5* genetic variants.

## I. INTRODUCCIÓN

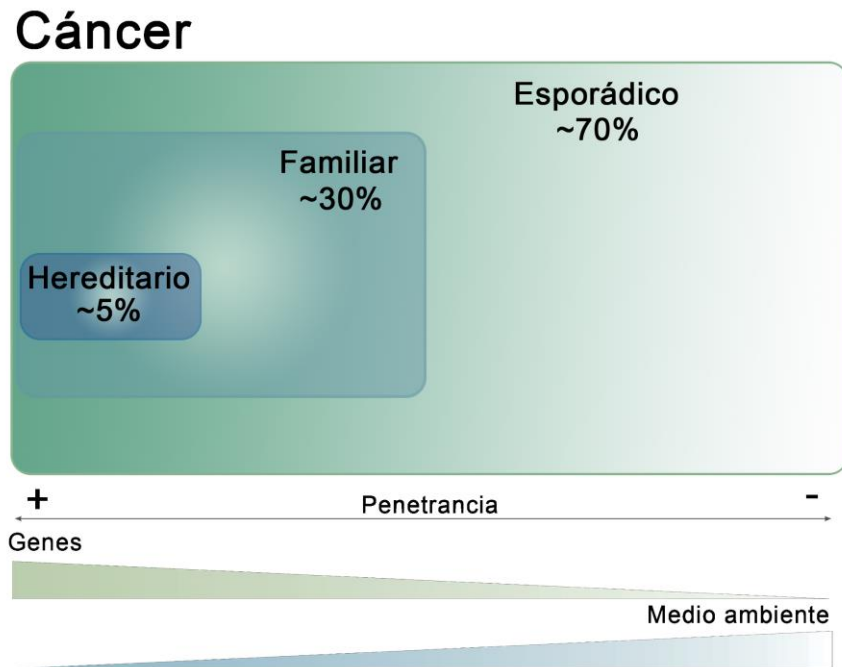
### 1. Cáncer: Antecedentes Generales

El cáncer es una de las principales causas de morbilidad y mortalidad en todo el mundo, con aproximadamente 14,1 millones de casos nuevos en 2012. Se estima que este número aumentará a 24 millones en 2035 [1]. En Chile esta enfermedad se presenta como la segunda causa de muerte, cuya tasa de mortalidad alcanza 139 por 100.000 habitantes [2].

Los datos epidemiológicos y genéticos actuales sugieren que el cáncer se presenta en tres formas: esporádica, familiar y hereditaria (**Figura 1**). El cáncer esporádico se define como aquel que ocurre en personas sin historial familiar de ese cáncer y sin cambios heredados en su ADN que aumentarían su riesgo de padecer ese cáncer en particular [3]. En el cáncer familiar por otro lado, los pacientes tienen uno o más familiares con el mismo tipo de cáncer, lo que se define como agregación familiar, sin embargo, no presentan un patrón específico de herencia. La evidencia de agregación familiar es importante, pues implica que los parientes cercanos de las personas con cáncer tienen un mayor riesgo de cáncer en ese mismo sitio. Finalmente, el cáncer hereditario o sindrómico, se caracteriza por la presencia de un gen alterado (asociado con el desarrollo del cáncer) que se transmite de padres a hijos en la familia, por lo que los pacientes suelen tener familiares con el mismo tipo o un tipo relacionado de cáncer, pueden desarrollar más de un cáncer y a menudo se presentan a edades tempranas. [4, 5].

Los cánceres esporádico y familiar caen dentro de la categoría de enfermedades complejas, las cuales son consecuencia de la combinación de factores de riesgo ambientales y genéticos, que en la mayoría de los casos son desconocidos [5]. Además, el grado de susceptibilidad y la variación fenotípica en la presentación de estas enfermedades, puede depender de un conjunto de enfermedades complejas fisiopatológicamente relacionadas, denominadas fenotipos intermedios o de segundo orden, y estos a su vez pueden depender de otros fenotipos intermedios denominados de tercer orden, lo cual conduce a un incremento en el nivel de complejidad del principal rasgo complejo (el cáncer en este caso) [6].

Es por esto que establecer las principales causas genéticas subyacentes a la susceptibilidad al cáncer, se ha convertido en uno de los retos más importantes de la actualidad. En este sentido, el análisis de familias con agregación del cáncer mediante enfoques de ligamiento o asociación genética contemplando el análisis de fenotipos relacionados, podría orientar la identificación de nuevas variantes de riesgo que también podrían ser responsables de un subconjunto de casos de cáncer esporádico, lo cual contribuiría a la comprensión de los mecanismos de susceptibilidad [7], ayudaría a identificar a las personas con mayor riesgo de desarrollar cáncer, facilitaría el desarrollo de paneles genéticos para la evaluación del riesgo y predisposición genética a ciertos tipos de cáncer y contribuiría con la oncología de precisión para un mejor tratamiento antineoplásico [8].



**Figura 1.** Visión global de la contribución genética al cáncer.

En los casos de cáncer hereditario las mutaciones causales son altamente penetrantes, tienen un fuerte componente hereditario y poca influencia ambiental. Sin embargo, en la mayoría de los casos familiares hay varias mutaciones de baja penetrancia que contribuyen a la susceptibilidad al cáncer de una manera aditiva, que implica interacciones entre genes y con factores ambientales. Estas mutaciones también pueden contribuir a los casos de cáncer que se clasifican como “esporádicos”. Por lo tanto, la distinción entre los casos “esporádicos” y “familiares” y entre los factores predisponentes “genéticos” y “ambientales” se ilustró de forma difuminada, pues podría considerarse mejor como un continuo de riesgos que contribuyen al desarrollo del cáncer. Adaptada de [9].

## **2. Cáncer de Vesícula Biliar: Generalidades y Factores de Riesgo**

### **2.1. Definición del Cáncer de Vesícula Biliar**

El Cáncer de Vesícula Biliar (CVB) (ICD-O C23) es uno de los tumores malignos más comunes y agresivos del tracto gastrointestinal [10]. Más del 85% de los tumores de la vesícula biliar son de tipo adenocarcinoma, distribuidos de la siguiente manera: 75% adenocarcinoma no papilar, 6% adenocarcinoma papilar, que se desarrolla en el tejido conectivo y tiene mejor pronóstico debido a su menor probabilidad de diseminarse, y 2% adenocarcinoma mucinoso. El restante 17% se distribuyen entre tumores de células escamosas, adenoescamosos, de células pequeñas, sarcomas y neuroendocrinos [11].

La iniciación y progresión del CVB se presume que ocurre bajo la secuencia displasia-carcinoma, en donde los cambios histológicos se acompañan de diversas anomalías moleculares progresivas [12]. En las etapas tempranas se observan mutaciones del gen de la Proteína Tumoral P53 (*TP53*), sobreexpresión de la Ciclooxygenasa 2 (*COX2*), mutaciones del ADN mitocondrial e hipermetilación de promotores en genes supresores tumorales. Durante el desarrollo de la displasia se produce la pérdida alélica en varios sitios cromosómicos (en particular en los brazos p de los cromosomas 3 y 8). En el estadio *in situ* del carcinoma ocurre la inactivación de genes supresores tumorales como el gen frágil de la tríada de la histidina (*FHIT*) y el gen del inhibidor 2A de quinasa dependiente de ciclina (*CDKN2A*) y pérdidas adicionales en otras regiones cromosómicas, particularmente en los brazos q de los cromosomas 9, 18 y 22. Finalmente, se producen cambios específicos, que se detectan sólo en tumores invasivos, tales como mutaciones en el gen homólogo del oncogen viral del sarcoma de la rata Kirsten (*KRAS*) [13].

### **2.2. Epidemiología del Cáncer de Vesícula Biliar**

A nivel mundial, el CVB tiene una incidencia baja ( $\approx 2,2$  por 100.000) [14], sin embargo, la incidencia de esta enfermedad es ampliamente heterogénea; en EE.UU. se presentan menos de 5.000 casos por año, con tasas de incidencia de 0,9 y 0,5 por cada 100.000 mujeres y hombres, respectivamente [15], mientras que en Europa central y septentrional, India, Chile, Bolivia y México se encuentran las poblaciones con mayores incidencias. En Chile, las



tasas son superiores a 25 casos por cada 100.000 mujeres y 9 casos por cada 100.000 hombres [16]. La provincia de Valdivia en Chile, tiene la incidencia de CVB más alta del mundo tanto en mujeres como en hombres con 27 y 12 casos por cada 100.000 personas por año, respectivamente. De igual forma, Chile presenta una de las tasas de mortalidad por CVB más altas en el mundo, afectando en mayor proporción a mujeres que a hombres. En Chile, el CVB representa la quinta causa de muerte por cáncer en mujeres, detrás de los cánceres de mama, pulmón, estómago y colon. En 2015, 1.316 personas (931 mujeres y 385 hombres), murieron por CVB en Chile ([www.deis.cl](http://www.deis.cl)). No obstante, se presentan variaciones regionales significativas; las tasas más altas se observan en la XIV, X y IX región, mientras que las más bajas en las regiones I, II y VI. La región De los Ríos presenta la mayor mortalidad en Chile con una razón de mortalidad estandarizada (RME) de 2,14 (95% CI: 1,57-2,69) [17, 18].

### **2.3. Diagnóstico, Pronóstico y Tratamiento del Cáncer de Vesícula Biliar**

El hallazgo de CVB suele realizarse bajo dos escenarios: (1) de forma incidental (cerca de la mitad de los casos), tras la colecistectomía en pacientes con sospecha de enfermedad benigna, fundamentalmente de colelitiasis, principal factor de riesgo para el CVB [19], y (2) de forma preoperatoria, en pacientes con síntomas inicialmente vagos e inespecíficos como dolor abdominal en el cuadrante superior derecho/cólico biliar, náuseas y/o vómitos, pérdida del apetito, fiebre e ictericia. En cualquiera de los casos, el diagnóstico tiende a realizarse en etapas avanzadas de la enfermedad [20], a menudo con focos diseminados localmente al hígado y otros órganos adyacentes, y por lo general con un mal pronóstico. De hecho, el CVB exhibe el peor pronóstico de los cánceres gastrointestinales y hepatobiliares, con una supervivencia global a los cinco años no mayor al 15% y que rara vez supera los 24 meses después del diagnóstico [17]. Actualmente, el diagnóstico de CVB depende principalmente de las imágenes auxiliares no invasivas y los exámenes invasivos como la laparoscopia y la biopsia, no existiendo un marcador tumoral único ideal para el diagnóstico y el pronóstico de CVB [21-25].

En cuanto al tratamiento, no existe una terapia curativa, excepto en los casos en que se puede realizar una colecistectomía con éxito (resección R0 entre el 10-20% de los casos de CVB no incidentales y 85% entre los incidentales) [19], lo cual depende principalmente de la

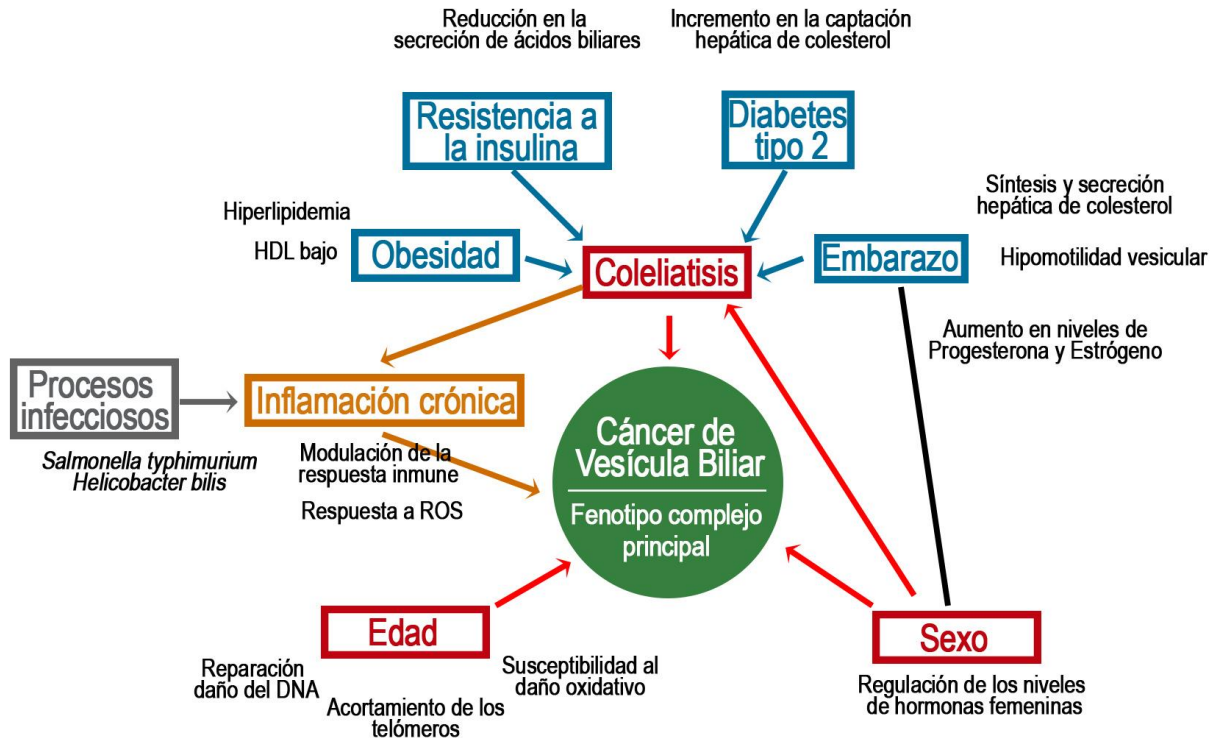
estadificación del tumor (**Tabla Suplementaria 1**). Para los casos descubiertos incidentalmente suele indicarse una segunda intervención quirúrgica. Para los tumores T1b, T2 y T3, la operación puede incluir la resección no anatómica del lecho de la vesícula biliar (segmentos IVB y V del hígado) o una resección anatómica tipo hepatectomía derecha [20]. En los pacientes con enfermedad T1a marginal negativa (sin invasión de la lámina propia), enfermedad localmente irresecable o metástasis a distancia, la intervención quirúrgica adicional no está indicada [26]. En aquellos casos de pronóstico pobre (60% del total, que corresponden a enfermedad avanzada) como CVB no resecables, recurrentes o metastásicos, se ha establecido como régimen de referencia para el tratamiento de primera línea con quimioterapia, la combinación cisplatino/gemcitabina. Mientras que el papel de la terapia de segunda línea aún no está claro, y no ha surgido un régimen único [27, 28].

En Chile, dado que la mortalidad por CVB es elevada, se incluyó desde el año 2006 la ecotomografía abdominal y la colecistectomía preventiva dentro de las Garantías Explícitas en Salud (GES), lo cual en teoría garantiza la realización de dichos procedimientos a los pacientes portadores o con sospecha de colelitiasis (sintomáticos o no, a partir de 2010) entre 35 y 49 años de edad [29]. El objetivo de esta estrategia preventiva es ampliar la cobertura de colecistectomía a edades tempranas, previo al desarrollo y mortalidad por CVB. Sin embargo, se ha descrito que la cobertura diagnóstica y terapéutica en Chile es insuficiente, especialmente la cobertura ecográfica, que adicionalmente muestra una inequidad regional y socioeconómica que se traduciría en una demanda potencial no cubierta de 176 mil colecistectomías [29], sin contar el número de pacientes sin síntomas previos, que también quedan por fuera de la cobertura diagnóstica.

En este sentido, se hace necesaria una mejor estrategia preventiva, basada en la priorización de las intervenciones en función de la identificación de los individuos susceptibles y los factores de riesgo asociados al desarrollo de CVB.

## 2.4. Etiología del Cáncer de Vesícula Biliar

El CVB es el resultado de una intrincada red de interacciones funcionales entre múltiples condiciones fisiológicas y fisiopatológicas junto con exposiciones medioambientales (**Figura 2**). La edad, el sexo femenino, la colelitiasis y la etnia se han observado fuertemente asociados con el desarrollo del CVB.



**Figura 2.** Modelo etiológico del Cáncer de Vesícula Biliar.

Diversos factores de riesgo crean una intrincada red de interacciones complejas que, junto con exposiciones medioambientales, conducen al desarrollo de CVB. Esta serie de factores de riesgo corresponden a condiciones fisiológicas y fisiopatológicas que en mayor o en menor medida influyen en la patogénesis y aportan los determinantes moleculares y genéticos de la susceptibilidad al CVB. Modelo adaptado al CVB con base en la propuesta de Blanco-Gómez *et al.* [6].

### 2.4.1. Edad y sexo como factores de riesgo de CVB.

Como con la mayoría de las neoplasias, la incidencia del CVB incrementa con la edad. En Chile por ejemplo, en el periodo 2011-2015 fallecieron 28, 236, 804, 1211 y 2113 mujeres en los grupos de <40, 40-50, 51-60, 61-70 y >70 años, respectivamente. El número de hombres que fallecieron en el mismo periodo fue de 13, 80, 270, 444 y 844, en los mismos

grupos de edad, respectivamente ([www.deis.cl](http://www.deis.cl)). Este comportamiento podría ser consecuencia de diferentes grados de susceptibilidad relacionados con mecanismos biológicos tales como la capacidad de reparar el daño en el ADN, el acortamiento de los telómeros o la susceptibilidad al daño oxidativo, lo cual conduciría a la acumulación de las alteraciones moleculares necesarias para el desarrollo del cáncer [30].

En cuanto al sexo; las mujeres presentan a nivel mundial, aproximadamente el doble de riesgo respecto a los hombres, mientras que en Chile la razón de incidencia entre mujeres y hombres es de 2,7 [20]. Los procesos regulatorios de los niveles de hormonas como el estrógeno y la progesterona podrían estar relacionados con esta mayor susceptibilidad, pues exposiciones más prolongadas a estas hormonas en las mujeres (menarquia temprana, menopausia tardía, embarazos y/o partos múltiples), se asocia con un mayor riesgo de CVB [31, 32]. Los estrógenos aumentan la formación de cálculos biliares mediante la elevación del colesterol biliar, estimulan la angiogénesis y promueven la proliferación celular anormal [33]. El mecanismo no es completamente conocido, pero se ha sugerido que el riesgo de colestitis y CVB podría estar mediado por polimorfismos en los receptores de estrógeno y progesterona [34].

#### **2.4.2. Colelitiasis y riesgo de CVB.**

La evidencia existente sugiere que el desarrollo de CVB en el escenario de la colestitis se produce al menos en parte, a través de la inflamación crónica, la producción local de agentes carcinógenos tales como ácidos biliares secundarios y la acumulación de toxinas que se excretan en la bilis [35]. Este proceso de inflamación crónica secundaria a cálculos biliares daría lugar a displasia epitelial y finalmente a carcinoma, mediante la presunta secuencia metaplasia-displasia-carcinoma, que se propone en el denominado “modelo de carcinogénesis y diseminación de la vesícula biliar” de Barreto *et al.* [36].

En este modelo, la inflamación crónica y el daño tisular generado por los cálculos biliares conducirían a un proceso adaptativo de reprogramación que daría lugar a la metaplasia epitelial, lo que se traduce en la transformación de un epitelio diferenciado en otro epitelio diferenciado [37]. La metaplasia es un hallazgo común en los tejidos de la

vesícula biliar expuestos a cálculos biliares [38, 39] y se encuentra en el 66% de las vesículas con carcinoma infiltrante [40]. Posteriormente se desarrollarían la displasia y el carcinoma *in situ* que también son relativamente frecuentes en las muestras de colecistectomía de rutina y se observan en la mucosa adyacente en más del 90% de los casos de CVB [41]. Estos cambios histológicos vendrían acompañados de diversas alteraciones moleculares progresivas (**Tabla Suplementaria 2**) [42-55].

La colelitiasis también es una enfermedad compleja, con múltiples factores de riesgo genéticos y medioambientales, algunos de los cuales han mostrado asociación con el riesgo de CVB [56, 57]. Es por esto que muchos de los factores de riesgo son compartidos entre ambas patologías, incluyendo dieta, estilo de vida sedentario, pérdida rápida de peso, diabetes, obesidad y sexo femenino [56]. En este sentido la colelitiasis podría considerarse como uno de los ejes principales dentro de una red jerárquica de factores de riesgo que determinan la susceptibilidad al CVB.

Al respecto, se ha observado que las frecuencias de cálculos biliares y CVB son paralelas en poblaciones específicas, como los indígenas de Chile y Norteamérica [58-61], de las que se desprende un gradiente coherente de prevalencias desde las poblaciones más mezcladas hacia las menos mezcladas. Además de la coexistencia en las mismas poblaciones, la asociación entre ambas patologías está determinada por el hallazgo de cálculos biliares en la mayoría de los pacientes con CVB (79-100%) [62, 63]. No obstante, sólo un 1-3% de los individuos con cálculos biliares desarrollan CVB [64, 65], y un porcentaje de los pacientes con CVB no muestran evidencia de colelitiasis. Por lo tanto, factores genéticos u otros factores ambientales podrían contribuir a la carcinogénesis en la vesícula biliar [47].

Los estudios familiares por su parte, han demostrado que las personas con parientes de primer grado con cálculos biliares tienen de 2 a 4 veces más probabilidades de desarrollar cálculos en comparación con los controles [66, 67]. Además, se ha estimado que la colelitiasis incrementa entre 2,3 y 34 veces el riesgo de CVB [68], y que las personas con antecedentes familiares de cálculos biliares que también tienen cálculos biliares tienen 57 veces mayor riesgo (IC 95% 32,0-110,5) de desarrollar CVB [69].

Sin embargo, los estudios en familias con colelitiasis y CVB son escasos, probablemente debido a la baja frecuencia global y a la edad de inicio tardía del CVB. De manera que los reportes de la co-segregación de la colelitiasis y el CVB en genealogías son prácticamente inexistentes. En cuanto a la frecuencia de colelitiasis dentro de las familias con CVB, no existen antecedentes que sugieran la existencia una tasa más alta que en la población general. No obstante, como se podrá observar en la sección de resultados de esta tesis, el número de individuos no afectados con CVB que presentaron colelitiasis en las familias estudiadas fue superior al esperado de acuerdo con la incidencia de la enfermedad en la población.

#### **2.4.3. Susceptibilidad genética al CVB.**

La búsqueda de variantes que otorguen susceptibilidad al CVB se ha abordado principalmente mediante estudios de genes candidatos y de asociación de genoma completo en casos esporádicos de CVB, mientras que los estudios en familias se han limitado al enfoque epidemiológico.

El estudio de genes candidatos de susceptibilidad al CVB se ha focalizado principalmente en la búsqueda de SNPs en genes litogénicos (afectan la homeostasis del colesterol y probablemente promueven la formación y el crecimiento de los cálculos de colesterol) y en genes asociados al proceso de carcinogénesis. Algunos de los genes estudiados en CVB, durante la última década se listan en la **Tabla 1**. Es interesante notar la asociación encontrada entre CVB y ciertas variantes en genes litogénicos, que también se relacionan de forma independiente con la colelitiasis, pues dicha observación soporta la idea de construir un modelo de redes jerárquicas basado en fenotipos intermedios que aportan diversos grados de susceptibilidad genética y que en conjunto conducen al desarrollo del CVB.

No obstante, y aunque algunos de estos genes se relacionarían con la fisiopatogenia molecular del CVB, este enfoque de estudio sobre genes candidatos no ha sido suficiente para establecer una asociación firme con la susceptibilidad al CVB. Adicionalmente, la gran mayoría de los estudios se han llevado a cabo en poblaciones chinas e indias, mientras que

los estudios en pacientes chilenos son escasos.

**Tabla 1.** Genes candidatos analizados en estudios de asociación en pacientes con CVB provenientes de diversas poblaciones.

En azul se resaltan los genes que también han mostrado asociación con coleditiasis. [41, 70, 71].

Población	Gen	OR
China	ABCG8, ALOX5, APOB, APOE, CETP, COMT, CYP17, CYP19A1, CYP1A1, ESR1, GSTM1, HSD3B2, IL1B, LDLR, LRPAP1, OGG1, PTGS2, RAD23B, RXR-B, SHBG, TGFB1, TNFa, XPC	1,0 - 3,4
India	ABCG8, ADRB3, ACE I/D, APOB, CCKAR, CYP1A1, CYP7A1, DNMT3B, EGF, ERCC2, GSTM1, GSTP1, IL1B, KRAS, LRPAP1, MSH2, MTHFR, NAT2, OGG1, PTGS2, TGFB1, TLR2, TLR4, TNFa	1,05 - 7,57
Japonesa	CYP1A1, CYP2C19, GSTM1, TP53	1,08 - 1,28
Chilena	APOB, APOE, CETP, CYP2C9, CYP2E1, CYP3A4, CYP7A1	1,06 - 1,28
	CYP1A1, GSTM1, TP53	NS

Se ha sugerido que la susceptibilidad genética a una enfermedad compleja como el CVB, es resultado de la acción conjunta de diversas variantes genéticas comunes, que aportan cada una un riesgo relativo moderado a la enfermedad. Por tanto, las investigaciones que se enfocan en los polimorfismos génicos de penetrancia moderada serían más acertadas [70]. No obstante, los estudios moleculares con este diseño recién se están empezando a usar en CVB.

Hasta la fecha se han publicado sólo dos estudios de asociación del genoma completo (GWAS) en CVB (Tabla 2) [72, 73]. El primero de ellos incluyó a 41 pacientes con CVB y 866 controles de población japonesa. El polimorfismo de un solo nucleótido (SNP) rs7504990 del gen Delecionado en Cáncer de Colon (*DCC*) (18q21.3), que codifica para el receptor de Netrina 1, mostró un OR estimado de 6,95 (IC del 95%: 3,43 a 14,08). El análisis de imputación subsiguiente identificó de manera similar, múltiples SNPs con asociaciones fuertes en una región genómica adyacente, en la cual se ha informado pérdida de heterocigosidad en CVB y otros cánceres. La reducción en la expresión de *DCC* se asocia con el tipo histológico pobremente diferenciado, aumento en la proliferación y metástasis a través de la pérdida de adhesividad [72].

**Tabla 2.** Variantes de susceptibilidad al CVB identificadas mediante estudios de asociación del genoma completo (GWAS).

Población	Casos	Controles	Gen	SNP	OR	IC 95%	Ref.
Japonesa	41	866	DCC	rs7504990	6,95	3,43-14,08	[72]
India	1042	1709	ABCB1	rs17209837	1,61	1,38-1,89	[73]
			ABCB4	rs1558375	1,47	1,30-1,66	
				rs4148808	1,57	1,35-1,82	

El segundo estudio analizó 1.042 casos de CVB y 1.709 controles, junto con una cohorte de replicación con 428 casos de CVB y 420 controles de población india. Se observaron asociaciones significativas en 3 SNPs con OR por alelo de 1,61 (1,38-1,89;  $p=2,26 \times 10^{-9}$ ) para el marcador rs17209837 del gen *ABCB1* y 1,47 (IC 95% 1,30-1,66;  $p=2,31 \times 10^{-1}$ ) y 1,57 (1,35-1,82;  $p=2,71 \times 10^{-9}$ ) para los marcadores rs1558375 y rs4148808 del gen *ABCB4*, respectivamente [73]. De estos tres genes, sólo para algunas variantes del gen *ABCB4* se ha reportado también asociación con el riesgo de colelitiasis [74-76]. Los autores también estimaron la heredabilidad combinada de los datos de GWAS utilizando la herramienta GCTA (Genome-Wide Complex Trait Analysis tool) e informaron que los SNPs comunes incluidos en su análisis explican el 23% de la variación del riesgo de CVB, lo que corresponde a un riesgo relativo familiar (RRF) de 3,15 [73].

Otro enfoque que se ha utilizado para la identificación de genes candidatos de susceptibilidad a CVB, es la secuenciación de nueva generación (NGS) de 409 genes relacionados con cáncer. Mediante esta aproximación, Yadav *et al.* [77], identificaron, en 11 pacientes con CVB, la presencia de 60 mutaciones germinales raras en 46 genes incluyendo controladores (*drivers*) de cáncer como el Miembro 4 de la Familia SMAD (*SMAD4*), la lisina metiltransferasa 2C (*KMT2C*) y *TP53*. Adicionalmente todos los casos de inicio temprano o casos hipermutados presentaban mutaciones germinales en genes críticos de la reparación del ADN, lo cual podría estar altamente relacionado con la susceptibilidad al cáncer.

Finalmente, los análisis en familias con múltiples casos afectados con CVB no ha sido empleado con frecuencia, debido a las dificultades mencionadas previamente, como alta letalidad y remoción temprana de la vesícula biliar. No obstante, la agregación de el CVB se ha documentado desde hace décadas [59, 78, 79]. Hemminki *et al.* [80, 81] por ejemplo,



publicaron un estudio utilizando la base de datos Familia-Cáncer sueca, que incluye 10,2 millones de individuos seguidos entre 1961 y 2012. Durante ese periodo se diagnosticaron más de 11.000 CVB, de los cuales 25 correspondían a casos con antecedentes familiares (**Tabla 3**). Esta cifra representa un riesgo 2,47 veces mayor de CVB para los hijos de pacientes con CVB respecto a la población general. También se observó una asociación significativa para los descendientes de madres diagnosticadas con CVB, lo que sugiere una transmisión materna preferencial del riesgo de la enfermedad (2,83).

**Tabla 3.** Razón de incidencia estandarizada (RIS) para el CVB en descendientes de padres diagnosticados con CVB entre 1961-2012.

El número observado corresponde al número de descendientes con CVB con antecedentes familiares.  
\* IC del 95% no incluye al 1,00.

	Número observado	RIS (95% IC)
<b>Padre</b>	3	1,22 (0,23–3,61)
<b>Madre</b>	22	2,83 (1,77–4,29)*
<b>Total</b>	25	2,47 (1,59-3,65)*

Adaptada de [81].

En un estudio similar, realizado por Jackson *et al.* [82], se analizó la Base de Datos de la Población de Utah (UPDB) que relaciona la información genealógica con el Registro de Cáncer, y se estimó un riesgo relativo de CVB de 2,28 (IC 95%: 1,35-3,60) para los individuos con parientes de tercer grado con CVB. Con base en la incidencia de genealogías con múltiples casos de CVB, se estimó que alrededor del 26% de los casos son causados por factores genéticos (94 de 364 casos que presentaban al menos un familiar afectado). Esta agregación familiar del CVB sugiere un papel crítico de la genética en el desarrollo de esta enfermedad, aunque los factores medioambientales también deben considerarse [83].

### 3. Fundamentos del Análisis de Ligamiento Genético

#### 3.1. Enfoques para el Análisis de Ligamiento

La pregunta principal en el análisis de ligamiento genético es si existe o no cosegregación de cualquier segmento de un cromosoma con el fenotipo, con una frecuencia mayor a la esperada por azar. Esta probabilidad se calcula mediante un elaborado análisis matemático [84], basado en test estadísticos que se han subdividido en dos categorías

principales: paramétricos y no paramétricos.

### **3.1.1. Análisis de ligamiento paramétrico.**

El análisis de ligamiento paramétrico se basa en la cálculo de la fracción de recombinación ( $\theta$  [Theta]) (la probabilidad de recombinación entre un locus marcador y el supuesto locus causal), para estimar la puntuación del logaritmo en base diez del *odds ratio* o verosimilitud (LOD), en donde la hipótesis alternativa, que supone que el ligamiento existe ( $\theta < 0,5$ ), se fija en el numerador, y está determinada por la verosimilitud de que el locus de enfermedad putativo se encuentre en una posición  $x$ , dentro del mapa de marcadores (ligamiento), mientras que la hipótesis nula, que representa el no ligamiento ente los loci ( $\theta = 0,5$ ), se ubica en el denominador, y está determinada por la verosimilitud de que el locus de la enfermedad se encuentre fuera del mapa, es decir, a una gran distancia de los marcadores. [85, 86]

La función de la puntuación LOD se define como:

$$LOD(\theta) = \log_{10} \left[ \frac{Like(\theta)}{Like(\theta = \frac{1}{2})} \right]$$

El análisis de ligamiento paramétrico se usa tradicionalmente para enfermedades que involucran un solo gen o que siguen un modo de herencia Mendeliana. De manera que se debe especificar un modelo genético para la enfermedad, es decir: (1) el número de loci génicos, (2) el modo de herencia en cada locus (dominante, recesivo, codominante), (3) la frecuencia del alelo causante de enfermedad, (4) la penetrancia de la enfermedad para cada genotipo, (5) la tasa de fenocopias y (6) las frecuencias alélicas de los marcadores. Las enfermedades mendelianas simples, típicamente presentan una alta penetrancia (llegando a 100%); sin embargo, para enfermedades de herencia compleja, no todas las personas que heredan el genotipo predisponente de la enfermedad desarrollan la enfermedad, y además hay individuos que no portan el genotipo de riesgo pero que desarrollan la enfermedad (fenocopias) [87]. La principal debilidad del análisis de ligamiento paramétrico es justamente la necesidad de especificar un modelo genético subyacente para la enfermedad. En general, los métodos paramétricos son bastante eficientes cuando el modelo se especifica

correctamente, siendo más potentes estadísticamente que los métodos no paramétricos. Sin embargo, si los modelos especificados son incorrectos, entonces el poder estadístico de los métodos de ligamiento paramétrico disminuye, la tasa de errores tipo 1 se incrementa considerablemente, y el poder del test puede ser excedido por el de los métodos no paramétricos [88].

Aunque el CVB carece de un modo de herencia genético claro, hay cierta evidencia hacia un patrón autosómico dominante. En el trabajo de Hemminki *et al.*, todos los pares identificados fueron de padres e hijos y ninguno de hermanos [80]. Sin embargo, en las familias identificadas en el presente estudio, 2 pares de afectados eran madre-hija, 3 pares eran hermanos y dos tía-sobrino(o). Para la colestiasis se ha encontrado evidencia de herencia tanto autosómica dominante como autosómica recesiva [75, 89, 90]. Se ha descrito que 51% de los casos familiares de colestiasis, presentan antecedente familiares parentales, 35% sólo entre hermanos y 14% tienen antecedentes en padres y hermanos [91].

### **3.1.2. Análisis de ligamiento no paramétrico.**

El análisis de ligamiento no paramétrico (o libre de modelo), por otra parte, es comúnmente usado con enfermedades multifactoriales, donde múltiples variantes genéticas, frecuentemente en conjunto con factores medioambientales, pueden contribuir al desarrollo de enfermedades y por ende no hay un modo de herencia claro. El análisis de ligamiento no paramétrico se basa en la estimación del número de alelos compartidos entre pares de parientes. El primer método fue desarrollado por Penrose [92] y analiza los alelos compartidos entre pares de hermanos que son idénticos por estado (IBS), es decir, llevan el mismo alelo en el mismo locus. Posteriormente, estos métodos fueron extendidos por Haseman y Elston [93] al análisis de alelos compartidos que son idénticos por descendencia (IBD), o sea, aquellos alelos que se han transmitido a cada individuo desde un ancestro común [85]. El tipo más simple de análisis de alelos compartidos se basa en estudiar pares de hermanos afectados (ASP), pero se han desarrollado enfoques más sofisticados aplicados a diversos tipos de familiares [94].

Con esta información, se realiza una comparación entre la puntuación de alelos IBD

estimado y la puntuación esperado de acuerdo a la hipótesis nula. Los nuevos enfoques hacen uso de genealogías grandes [95], tienen en cuenta la configuración fenotípica (sanos-afectados) y el patrón de herencia de un locus en particular dentro de la genealogía, definido bajo la siguiente función de puntuación:

$$S(v, \phi)$$

donde  $\phi$  representa la configuración fenotípica de una genealogía especial y  $v$ , el vector de herencia. Dado que normalmente el patrón de herencia real es desconocido, el modelo calcula la media de todos los posibles vectores de herencia ponderadas por la probabilidad del vector [96].

### **3.2. Herramientas Informáticas para la Estimación del Ligamiento**

Existen diversos programas informáticos que implementan los algoritmos de estimación del ligamiento (Elston-Stewart: LINKAGE [97], FASTLINK [98]. Lander-Green: GeneHunter [99], MERLIN [100]). También existen paquetes integrales, como por ejemplo Mendel [101], que permiten el análisis estadístico de rasgos genéticos cualitativos y cuantitativos. Si bien Mendel está orientado al análisis de genealogías, contiene varias opciones de análisis que emplean datos de casos/controles o muestras aleatorias de individuos. Una de sus ventajas para el análisis de datos de genealogías, es que incorpora los algoritmos de Elston-Stewart y de Lander-Green-Kruglyak, y en algunas de sus aplicaciones (como las utilizadas para el análisis de ligamiento), elige genealogía por genealogía, el algoritmo que sea más rápido [101].

Si bien el objetivo final de los estudios de ligamiento es identificar genes o variantes relacionadas con la enfermedad, a menudo sólo se consiguen identificar regiones genómicas que podrían albergarlos. Estas regiones son mucho más grandes que las identificadas en los estudios de asociación y pueden llegar a abarcar incluso 40 Mb [102]. Aunque la resolución suele ser bastante pobre (la región de ligamiento contiene una gran cantidad de genes), una señal de ligamiento estadísticamente significativa limita la búsqueda del gen predisponente a esa región candidata, y por tanto reduce los costos financieros y temporales de estudios posteriores. El análisis de ligamiento con una mayor densidad de marcadores (ligamiento

multipunto), la priorización de variantes por anotación funcional, el mapeo fino por asociación y la secuenciación genética son métodos para complementar la identificación de la variante genética predisponente [103].

#### **4. Priorización de Variantes**

La priorización de variantes dentro de las regiones de ligamiento es importante para permitir la conversión de las señales estadísticas en la identificación de los genes directamente relacionados con la biología de la enfermedad. El objetivo de este proceso es tratar de conectar estas variantes a genes probables cuya alteración conduciría al riesgo de enfermedad, mediante anotación funcional.

Aquí se aprovechan las herramientas disponibles para anotar variantes genéticas por ubicación genómica (secuencias intrónicas, exónicas, reguladoras), consecuencias funcionales, y en categorías - para aquellas variantes que caen en regiones codificantes para genes - como por ejemplo por función molecular o proceso biológico, entre otros. Este enfoque ha ayudado en la inferencia causal de las variantes identificadas en estudios de ligamiento y asociación, y aporta fuentes de información independientes sobre lo que cada variante probablemente haga biológicamente [104][105, 106].

#### **5. Mapeo Fino por Asociación**

El mapeo fino se define como el proceso por el cual se asignan probabilidades de causalidad bien ajustadas a las variantes dentro de una región candidata previamente identificada [103]. Al igual que en los GWAS, en el mapeo fino se aprovecha el principio del desequilibrio de ligamiento (DL) a nivel poblacional, entre marcadores (SNPs) con una frecuencia alélica moderada-alta que son genotificados, y variantes desconocidas que se supone causan directamente la enfermedad compleja. El DL entre los marcadores genéticos y las variantes causales es creado por fuerzas evolutivas como la mutación, la deriva genética y la selección, y se descompone a través de la recombinación [107]. En la práctica, para estimar dichas probabilidades, se usan modelos de asociación de regresión logística en donde el fenotipo (enfermo/sano) es la variable de respuesta, con combinaciones lineales (aditivas) de las variables explicatorias que entran en el modelo como predictores, por ejemplo variables

de genotipo o covariables como factores de riesgo epidemiológico (fumar, sexo), variables clínicas (gravedad de la enfermedad y edad de inicio) y estratificación poblacional (componentes principales que capturan la variación debido a la ascendencia diferencial de los individuos) [108]. En esta tesis, el objetivo final del mapeo fino fue establecer si variantes asociadas con el riesgo de colelitiasis y CVB en familias también se asocian con la susceptibilidad al CVB en los casos esporádicos.

A pesar del gran impacto sanitario de esta neoplasia en Chile (el riesgo acumulado de ser diagnosticado con CVB para un chileno antes de los 75 años es de un 1,2%) y de la probable existencia de variantes genéticas de susceptibilidad, se han llevado a cabo pocos estudios y aún no se han identificado marcadores moleculares que permitan predecir el riesgo individual de desarrollar CVB y prevenir esta agresiva enfermedad. En este sentido, es necesario identificar marcadores de susceptibilidad del desarrollo de CVB en población chilena. Por ello, en este trabajo de tesis se propuso identificar regiones cromosómicas candidatas de ligamiento en familias con múltiples casos de CVB y colelitiasis, para posteriormente, estudiar dichas regiones empleando métodos de mapeo fino, que permitan identificar variantes genéticas asociadas con la susceptibilidad al CVB dentro de estas familias y en casos esporádicos.

## **II. Hipótesis y Objetivos**

### **1. Hipótesis**

La agregación familiar de cáncer de vesícula biliar y colelitiasis se asocia con variantes genéticas de riesgo para cáncer de vesícula biliar esporádico.

### **2. Objetivos**

#### **2.1. Objetivo General**

Identificar variantes genéticas que se asocien con el riesgo de colelitiasis y cáncer de vesícula biliar con agregación familiar y esporádico.

#### **2.2. Objetivos Específicos**

1. Identificar regiones genómicas asociadas a la colelitiasis y el CVB con agregación familiar mediante análisis de ligamiento.
2. Establecer la relevancia fisiopatológica de los genes contenidos en las regiones genómicas candidatas mediante análisis de anotación funcional.
3. Identificar variantes genéticas asociadas con el riesgo de CVB esporádico mediante un mapeo fino por asociación en las regiones cromosómicas candidatas.

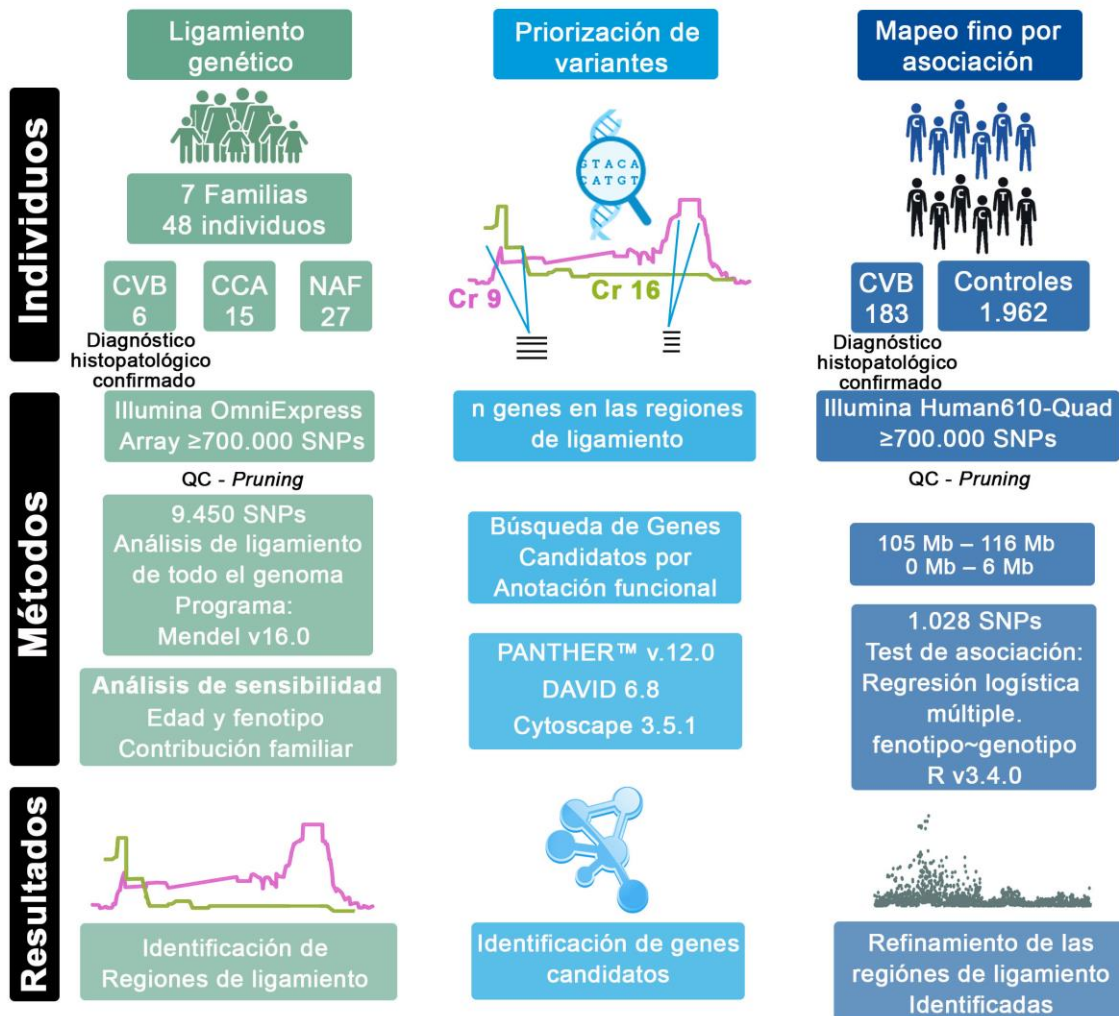
### III. Individuos y Métodos

#### 1. Diseño Experimental

Una estrategia comúnmente aplicada en el mapeo de variantes genéticas que predisponen al desarrollo de enfermedades complejas consiste en realizar un análisis de tres etapas [109-116]: (1) análisis de ligamiento de genoma completo en familias, (2) priorización de variantes usando métodos de anotación funcional, y (3) mapeo fino por asociación genética (basada en familias o en casos y controles) con marcadores y/o ligamiento multipunto en las regiones candidatas identificadas tras el análisis de ligamiento.

La **Figura 3** ilustra la aplicación del diseño de tres etapas en esta tesis. La elección del diseño se basó fundamentalmente en la complementariedad de los enfoques empleados. Los estudios de ligamiento permiten reducir la región de búsqueda de las variantes genéticas asociadas, del genoma completo a un número variable de regiones candidatas de diversos tamaños entre 1 y 40 Mb [117, 118]. Las herramientas para clasificar y anotar funcionalmente facilitan la identificación de genes candidatos entre los contenidos en regiones identificadas en estudios de ligamiento y/o asociación [103]. Estos enfoques complementan el refinamiento de las regiones a estudiar. Finalmente, con este conocimiento *a priori*, se puede aumentar la densidad de los marcadores genéticos en las regiones candidatas obtenidas del análisis de ligamiento, y realizar un mapeo fino en casos y controles, aumentando la potencia para detectar y localizar con mayor precisión las regiones o variantes asociadas con el riesgo de desarrollar la enfermedad [87]. Tal como se mencionó anteriormente, el objetivo final del mapeo fino en esta tesis, fue establecer si existen variantes asociadas al riesgo de CVB en las regiones candidatas, tanto en casos esporádicos como familiares.





**Figura 3.** Diseño del estudio.

Se emplearon tres enfoques complementarios: ligamiento genético en familias, priorización de variantes en las regiones de ligamiento utilizando anotaciones funcionales, y mapeo fino por asociación estadística basada en casos y controles poblacionales de conveniencia. Se muestran las diferentes etapas del estudio según su orden cronológico de izquierda a derecha. Todos los individuos con CVB contaban con el diagnóstico histopatológico. Hacia abajo se describen las muestras/regiones cromosómicas, los pasos y técnicas utilizadas y los resultados a conseguir en cada una de las etapas. Se describen los conjuntos de marcadores, los límites de los intervalos cubiertos por estos marcadores en las regiones de ligamiento, las técnicas utilizadas para la genotipificación inicial de los SNPs, el tipo de análisis genético realizado y el programa utilizado en cada caso. CVB: individuos con cáncer de vesícula biliar. CCA: individuos con colecistectomía. NAF: individuos no afectados.

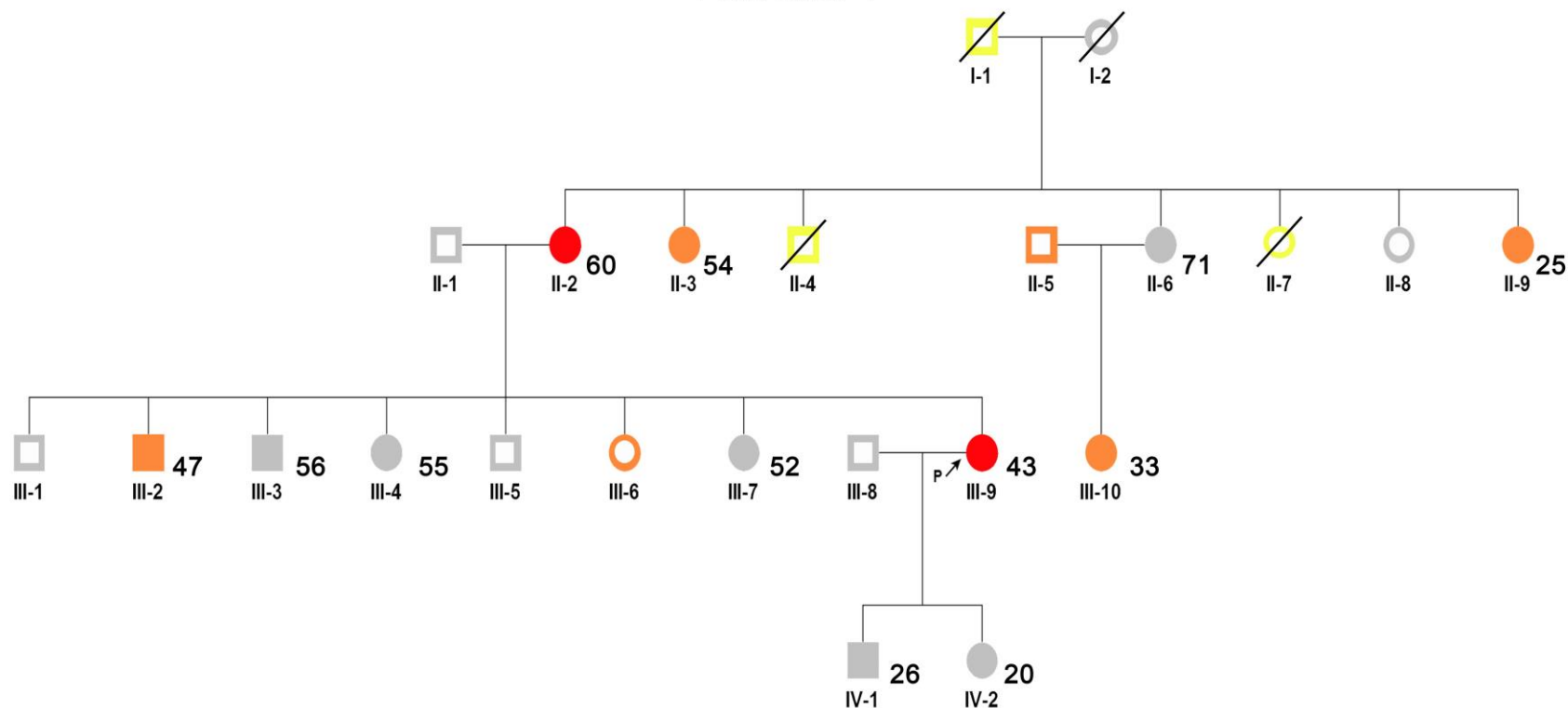
## 2. Familias con Múltiples Casos de Cáncer de Vesícula Biliar

Fueron considerados casos familiares, aquellos en los cuales el caso índice presentaba al menos un familiar directo – entre el primer y tercer grado de consanguinidad – diagnosticado con CVB. En total se identificaron 16 familias con múltiples miembros afectados por CVB. Las genealogías de estas familias fueron utilizadas para la estimación de los riesgos de recurrencia empíricos y los riesgos relativos ( $\lambda$ ) para los parientes del primer al tercer grado de consanguinidad de los probandos. Del total de familias, siete accedieron a participar y fueron incluidas en el análisis de ligamiento. Las familias restantes rechazaron su participación por diversas razones, por ejemplo, motivos personales como el fallecimiento reciente de familiares cercanos (por otras causas) o ausencia de familiares por viajes al extranjero, mientras que otras familias prefirieron no expresar sus motivos. Estos individuos no mostraron características distintivas en cuanto a sexo o edad de diagnóstico respecto a los que aceptaron participar. En las genealogías de las 7 familias participantes se representaron 179 individuos con los siguientes fenotipos: CVB=14, colecistectomía=22, cáncer no vesicular=5, no afectado=138 (**Figura 4**). Un total de 48 individuos (CVB=6, colecistectomía=15, no afectado=27) fueron entrevistados (**Material Suplementario 3**) y se les tomó una muestra de sangre para extraer ADN genómico. Todos los probandos y sus familiares afectados con CVB contaban con un diagnóstico histológico (individuos genotipificados; n=6) y/o el certificado de defunción correspondiente (n=8). Las características y distribuciones de los individuos incluidos en el análisis de genealogías y ligamiento se detallan en la **Tabla 4**.

**Tabla 4.** Características generales de los individuos incluidos en el análisis de genealogías y ligamiento.

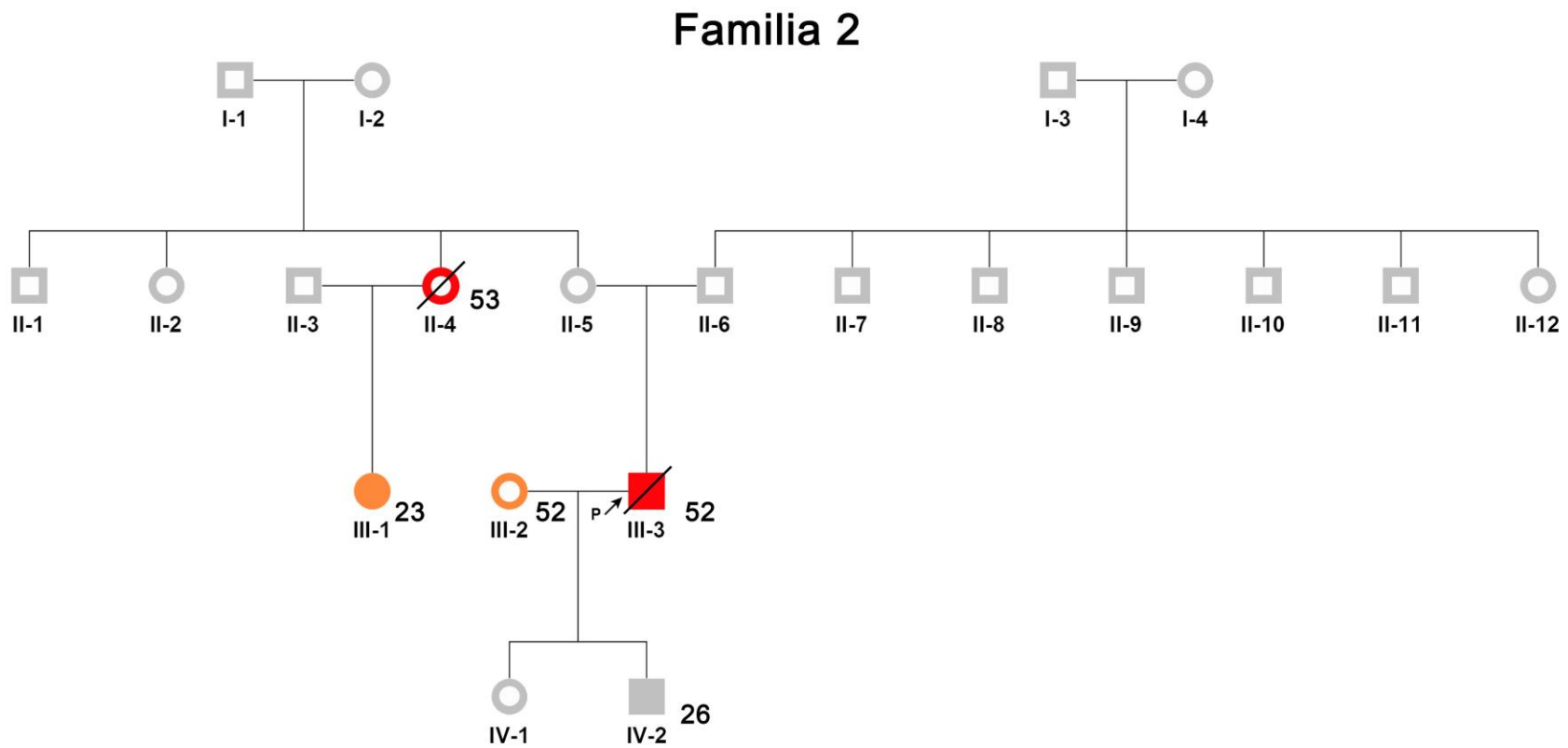
	<b>Datos de segregación</b>			<b>Datos de ligamiento</b>			
<b>Número de genealogías</b>	16			7			
<b>3 generaciones</b>	8			2			
<b>4 generaciones</b>	6			5			
<b>5 generaciones</b>	1			0			
<b>Número promedio de individuos por genealogía ± DS</b>	22,7 ± 6,7			25,3 ± 3,6 (Genotipificados: 6,9 ± 3,9)			
<b>Número de individuos</b>	<b>Total</b>	<b>Hombres</b>	<b>Mujeres</b>	<b>Total</b>	<b>Genotipificados</b>	<b>Hombres</b>	<b>Mujeres</b>
<b>No afectados</b>	282	170	112	138	27	8	19
<b>Cáncer no vesicular</b>	7	3	4	5	0	0	0
<b>Afectados</b>	76	23	53	36	21	3	18
<b>CVB</b>	34	7	27	14	6	1	5
<b>Colecistectomías</b>	42	16	26	22	15	2	13
<b>Total</b>	365	196	169	179	48	11	37
<b>Proporción de afectados</b>	0,21	0,12	0,31	0,20	0,44	0,27	0,49

## Familia 1



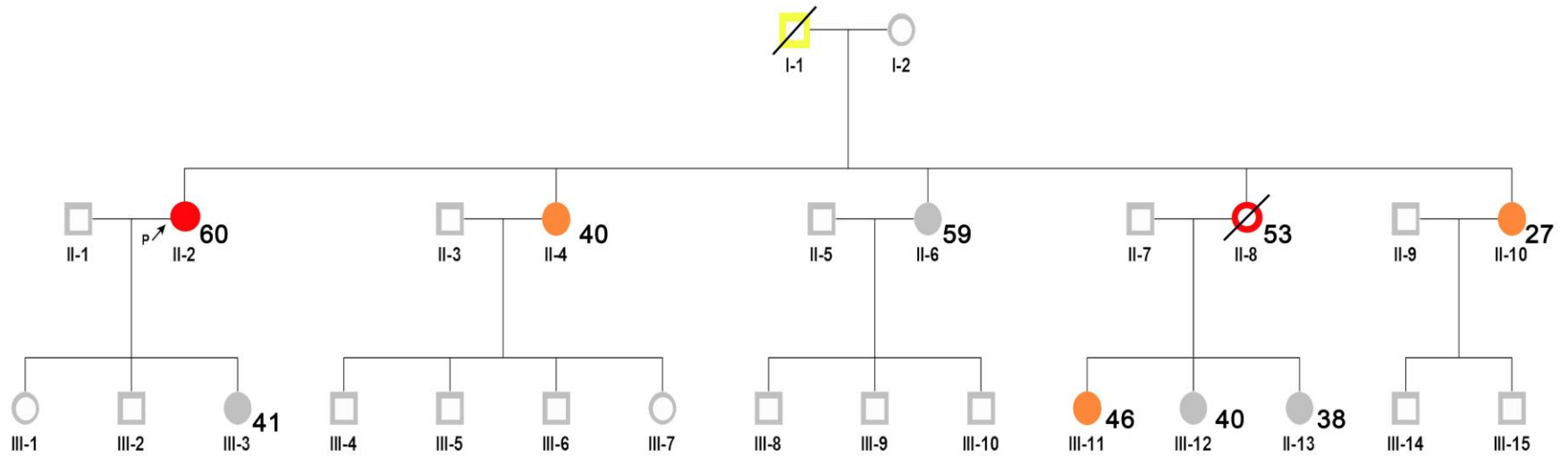
**Figura 4.** Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar.

Los colores representan la disponibilidad tanto del genotipo (vacío = sin datos de genotipo, relleno = genotipo disponible) como del fenotipo (rojo = cáncer de vesícula biliar (CVB), naranja = colecistectomía, amarillo = Cáncer No Vesicular, gris = no afectado). Los números indican la información sobre edad más relevante (Edad de diagnóstico de CVB o de colecistectomía o de realización de la entrevista). Continúa.



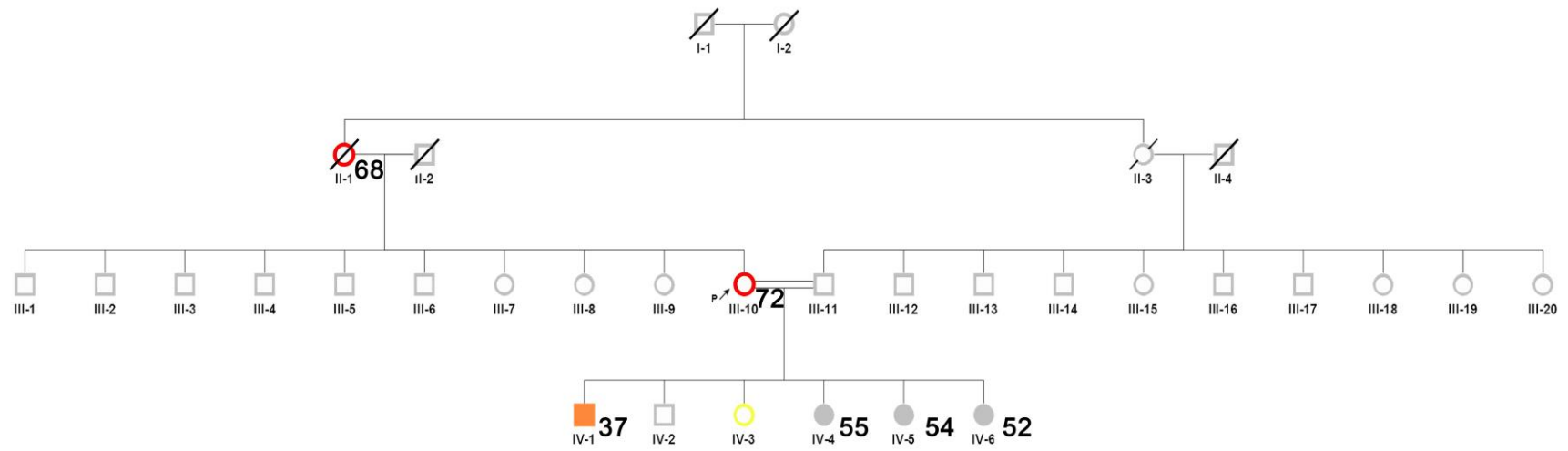
**Figura 4.** Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar. Continuación 2/7.

### Familia 3



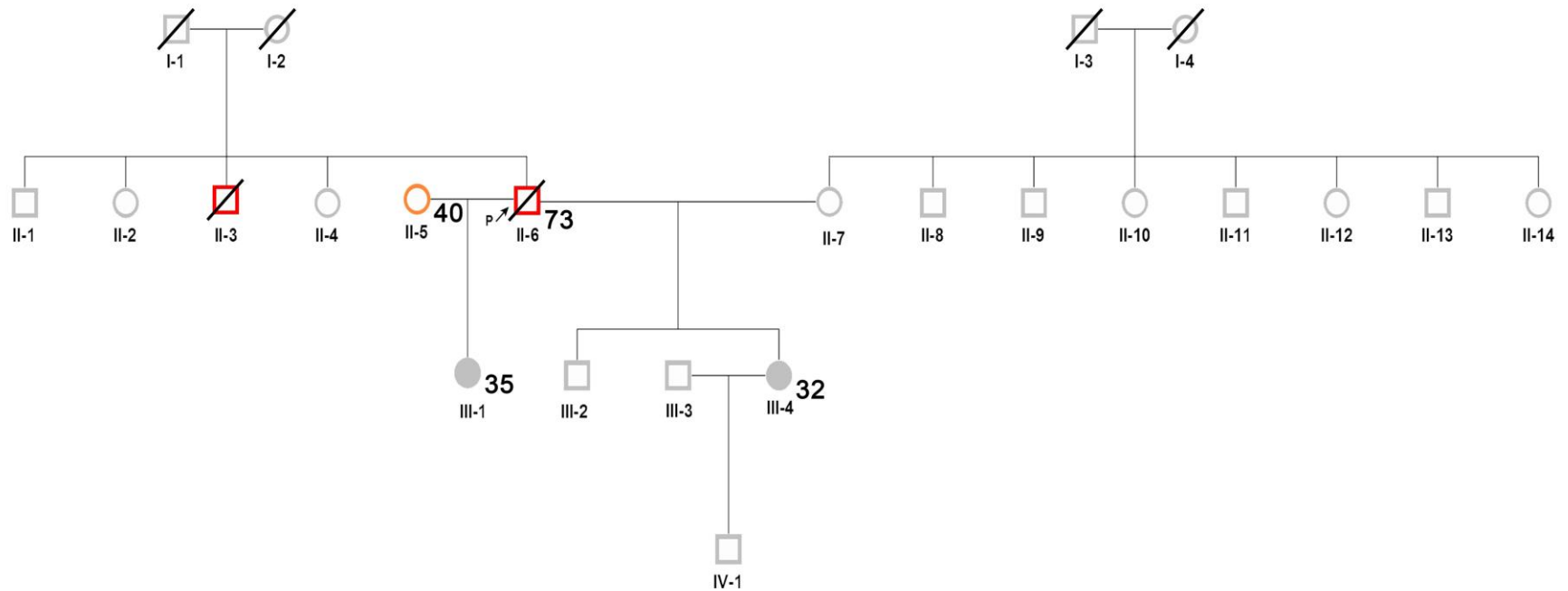
**Figura 4.** Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar. Continuación 3/7.

## Familia 4



**Figura 4.** Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar. Continuación 4/7.

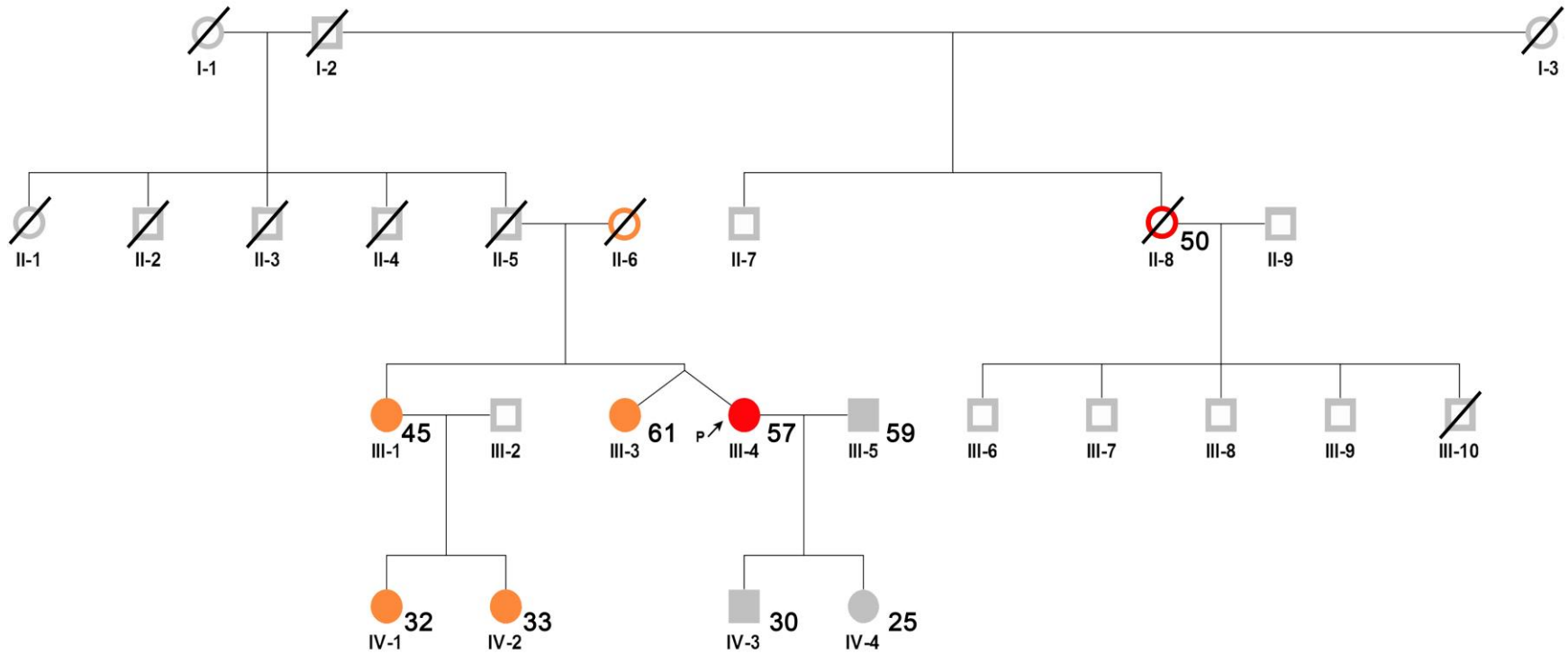
## Familia 5



**Figura 4.** Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar. Continuación 5/7.

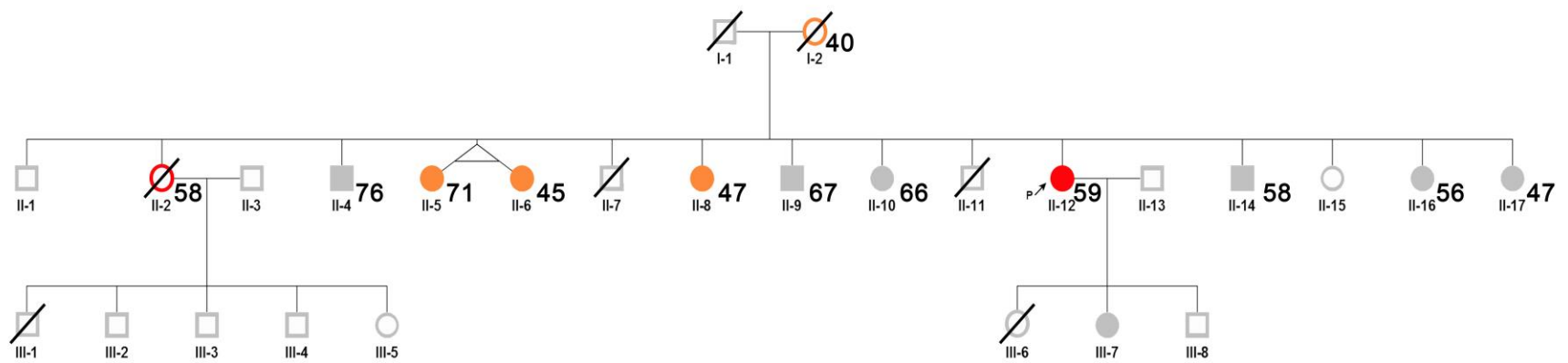


## Familia 6



**Figura 4.** Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar. Continuación 6/7.

### Familia 7



**Figura 4.** Genealogías de 7 familias chilenas con varios miembros diagnosticados con cáncer de vesícula biliar. Continuación 7/7.

A continuación, se describen detalladamente tres de las familias más relevantes para el estudio. La familia 1 contaba con un total de 23 individuos distribuidos en 4 generaciones, de los cuales 12 pudieron ser contactados y accedieron a participar en el estudio. El caso índice (III-9) de esta familia corresponde a una mujer con diagnóstico de CVB a los 43 años, cuya madre (II-2) presentó CVB a una edad de 60 años. Además, un hermano, dos tías y una prima del caso índice reportaron haber sido colecistectomizados con una edad promedio de  $40 \pm 13$  años (DE). Esta familia es un buen ejemplo de la problemática encontrada al estudiar el CVB familiar: tan pronto se diagnostica un caso en la familia, y con mayor razón si se diagnostican varios casos, los familiares de los pacientes son colecistectomizados. Esto impide observar si otros individuos en esas familias desarrollan CVB, motivando los análisis de sensibilidad llevados a cabo en la presente tesis. Al momento de la encuesta, el caso índice reportó tres presuntos casos de cáncer en la familia: El abuelo (I-1) con cáncer de páncreas, un tío materno (II-4) con cáncer de vías biliares, y una tía materna (II-7) del caso índice con cáncer de mama. No obstante, estos diagnósticos no pudieron ser corroborados mediante reportes de anatomía patológica, y las causas de muerte certificadas por el Registro Civil e Identificación de Chile correspondieron a hemorragia cerebral (I-1), paro cardiorrespiratorio / cáncer gástrico (II-4) y cáncer de mediastino / secuela pulmonar TBC / insuficiencia cardiaca congestiva (II-7).

La genealogía de la familia 3 incluía un total de 27 individuos distribuidos en 3 generaciones. Ocho familiares accedieron a participar. El caso índice (II-2) corresponde a una mujer con diagnóstico de CVB a los 60 años, cuya hermana (II-8) falleció por CVB terminal a una edad de 53 años. Además, dos de sus hermanas, y una sobrina reportaron haber sido colecistectomizadas a una edad promedio de  $35 \pm 7$  años (DE). El padre del caso índice (I-1) tuvo un presunto cáncer esofágico.

Finalmente, la familia 7 contaba con 3 generaciones y un total de 27 individuos, de los cuales 11 accedieron a participar. El caso índice (II-12) corresponde a una mujer con diagnóstico de CVB a los 59 años, cuya hermana (II-2) presentó un CVB terminal a una edad de 58 años. Además, se reportaron colecistectomías en la madre (I-2) y en tres hermanas del caso índice (II-5, II-6 y II-8), quienes fueron operadas a una edad promedio de  $51 \pm 14$  años.

Como hallazgo relevante en esta familia, se destaca el par de gemelas monocigotas (II-5 y II-6) concordantes para los cálculos biliares.

Esta información descriptiva de las familias, junto con la inspección visual de las genealogías fue fundamental para el diseño de análisis de sensibilidad considerando los fenotipos: estrictamente CVB o, alternativamente, CVB junto con colecistectomía. En concreto, en el segundo caso se investigaron los resultados de los análisis de ligamiento suponiendo que los pacientes colecistectomizados hubieran desarrollado CVB de no haber sido removida su vesícula. Esto debido a que, en términos del riesgo de desarrollar CVB, un paciente esporádico colecistectomizado, y un paciente colecistectomizado con dos familiares diagnosticados con cáncer de vesícula biliar son casos completamente diferentes [69]. Los familiares con cálculos biliares que no se habían sometido a colecistectomía fueron considerados como no afectados en los análisis de sensibilidad, debido a que aún conservaban su vesícula y no habían desarrollado CVB. Por el contrario, en el análisis de ligamiento, fueron considerados junto con los individuos sanos con el fenotipo “*unknown*”.

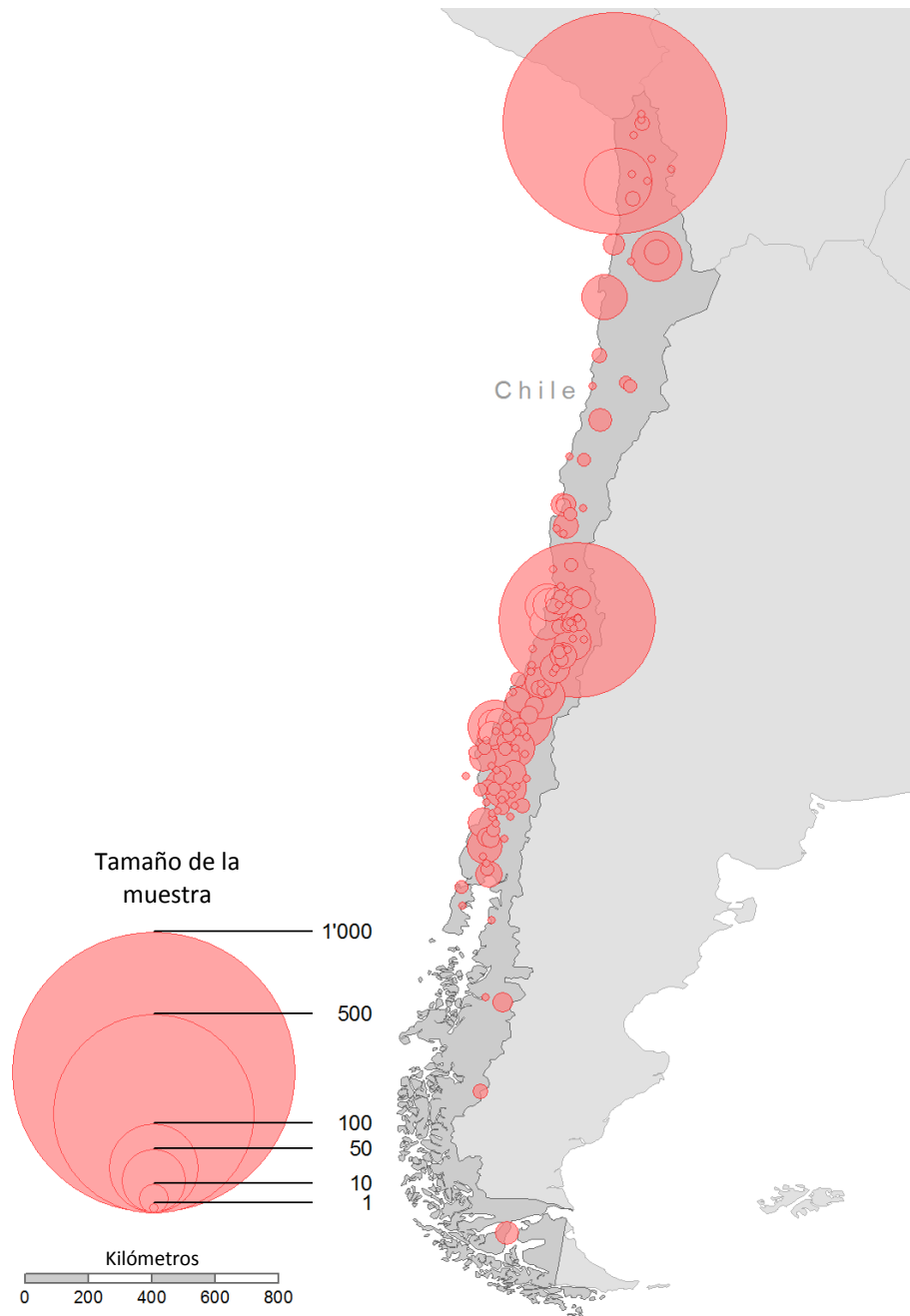
En las formas familiar y hereditaria de diversos tipos de cáncer, los familiares de pacientes jóvenes tienen mayor riesgo de desarrollar el cáncer que aquellos con parientes afectados de mayor edad. La agregación familiar suele ser más evidente en las familias con diagnóstico temprano del cáncer, y tanto el número de miembros afectados en la familia como una temprana edad de diagnóstico son los determinantes más importantes del riesgo familiar de la enfermedad [119-122]. En el caso del CVB esporádico, la edad promedio de las personas al momento del diagnóstico es de 72 años, siendo más de dos tercios de los pacientes mayores de 65 años [123]. La edad promedio del diagnóstico de CVB en las familias identificadas fue de 58 años, de los cuales tan solo una quinta parte fue diagnosticada sobre los 65 años de edad, lo que sugiere una tendencia al desarrollo del CVB a edades tempranas. Adicionalmente, la edad promedio de colecistectomía en los familiares fue de 41 años, mientras que el 55% de las colelitiasis en Chile se diagnostican en personas de 45 a >80 años [124]. Sin embargo, la ventana de tiempo promedio entre estas dos edades (edad de colecistectomía en individuos sin CVB y edad de diagnóstico de CVB en individuos con el cáncer, dentro de las familias), fue de 17 años. En este sentido, la remoción de la vesícula a

edad temprana tiene implicaciones obvias sobre el riesgo de desarrollar CVB, especialmente considerando que la coleditiasis incrementa entre 2,3 y 34 veces el riesgo de CVB [68], y que las personas con antecedentes familiares de cálculos biliares que también tienen cálculos biliares tienen 57 veces mayor riesgo (IC 95% 32,0-110,5) de desarrollar CVB [69]. Si además se contempla el riesgo adicional que otorgan los antecedentes familiares de CVB (RR de 2,1 a 13,9 en familiares de primer grado, dependiendo de la población) [125], la probabilidad de desarrollar CVB que tendrían los pacientes colecistectomizados con dos familiares diagnosticados con cáncer de vesícula biliar sería considerablemente alta.

Teniendo en cuenta estas consideraciones se implementaron varios modelos de análisis integrando dos definiciones de fenotipo y seis restricciones por edad que se describen más adelante en la sección de métodos, análisis de sensibilidad.

### **3. Pacientes con Cáncer de Vesícula Biliar e Individuos Control**

Para el mapeo fino por asociación genética se utilizó la información de 2.145 individuos provenientes de tres estudios independientes: los datos chilenos del “Consortio para el Análisis de la Diversidad y Evolución de Latinoamérica” (Candela), liderado por el Dr. Francisco Rothhammer, el “Estudio Piloto sobre Cáncer de Vesícula en Chile” liderado por la Dra. Jill Koshiol, y el proyecto de investigación “Identificación de Biomarcadores para una Prevención y Tratamiento Personalizados del Cáncer de Vesícula Biliar” liderado por el Dr. Justo Lorenzo Bermejo. En estos tres estudios se reclutaron individuos chilenos de ambos sexos, mayores de edad. Del total de individuos, 183 eran pacientes con CVB y 1.962 eran controles poblacionales. Todos los casos contaban con un diagnóstico confirmado de CVB. Los individuos control fueron reclutados mediante un muestreo por conveniencia, invitándolos a participar a través de conferencias públicas y presentaciones en medios de comunicación. Alrededor de dos tercios de los individuos control reclutados eran soldados profesionales, con una proporción relativamente grande de hombres nacidos en el sur de Chile, pero en general representando bien a la población chilena [18, 126]. El mapa de la **Figura 5** muestra el lugar de nacimiento de los participantes en los tres estudios.



**Figura 5.** Mapa de Chile representando los lugares de nacimiento de todos los individuos del estudio. La ubicación de los círculos corresponde a los lugares de nacimiento. Los tamaños de los círculos son proporcionales a la cantidad de voluntarios nacidos en ese lugar. Adaptada de [126].

Teniendo en cuenta el riesgo acumulado de CVB para un chileno (1,2%), aproximadamente 24 individuos dentro del grupo control probablemente desarrollen cáncer de vesícula biliar antes de los 75 años, resultando en estimaciones algo conservadoras de los efectos genéticos asociados – los efectos genéticos serían mayores si el grupo de referencia incluyera solamente individuos sin cáncer de vesícula biliar.

Con el tamaño de muestra alcanzado (183 casos con CVB y 1.962 controles poblacionales), se podrían identificar variantes que otorguen un alto RR de CVB (**Tabla 5**) [127]. Por ejemplo, suponiendo una frecuencia alélica del 10% en la población se podrían identificar variantes dominantes que otorguen un RR = 2,7 (potencia estadística 80%, nivel de significación  $\alpha = 4,9 \times 10^{-5}$ ) [128, 129] .

**Tabla 5.** Riesgos relativos (RR) identificables para diferentes frecuencias alélicas en una muestra de 183 casos con CVB y 1.962 controles poblacionales.

Las estimaciones fueron realizadas con la herramienta *Genetic Association Study (GAS) Power Calculator* [129] fijando la potencia estadística en 80% y el nivel de significación en  $\alpha = 4,9 \times 10^{-5}$ , usando una prevalencia del CVB de 3% en la población general [1], bajo 3 modelos genéticos suponiendo diferentes frecuencias alélicas.

Frecuencia alélica	Modelo genético		
	Multiplicativo	Aditivo	Dominante
<b>0,15</b>	2,0	2,2	2,5
<b>0,10</b>	2,3	2,5	2,7
<b>0,05</b>	3,0	3,3	3,4

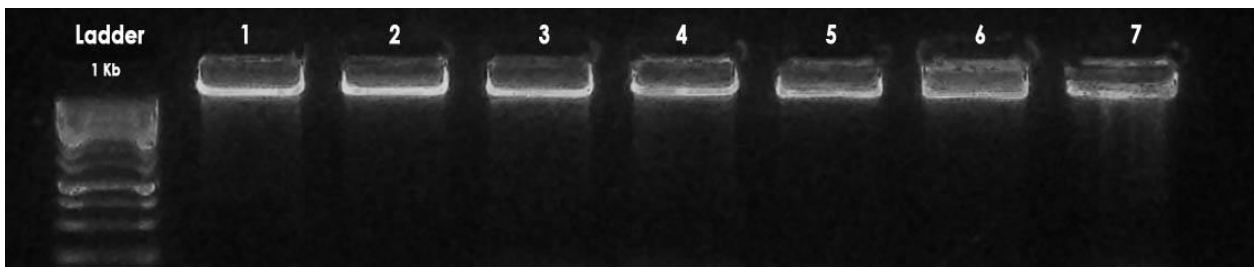
## 4. Métodos

### 4.1. Aprobación Ética

El estudio contó con la aprobación ética de los Servicios de Salud Metropolitano Central, Metropolitano Sur Oriente, Metropolitano Oriente, Metropolitano Occidente, del Maule, de Concepción y de Valdivia, del Comité Ético Científico de la IV Región, de las Facultades de Medicina de la Universidad de Chile y de la Pontificia Universidad Católica de Chile, y de la Universidad de Tarapacá. A modo de ejemplo, se adjunta Acta de aprobación y consentimiento informado del Comité de ética en investigación en seres humanos de la Facultad de Medicina de la Universidad de Chile (**Material Suplementario 4**). Todos los participantes proporcionaron su consentimiento informado por escrito (**Material Suplementario 5**).

### 4.2. Procesamiento de las Muestras

Las células mononucleares de las muestras de sangre periférica se aislaron usando un sistema de separación por centrifugación en gradiente de densidad (Leucosep Tubes®). El pellet celular se procesó inmediatamente, usando el kit de purificación de ADN genómico Wizard® (Promega). El ADN se cuantificó usando el reactivo PicoGreen (Invitrogen, Darmstadt, Alemania), realizando mediciones por triplicado a fin de establecer la variabilidad en la cuantificación del ADN. La integridad y calidad del ADN se verificó mediante electroforesis, cargando 100 ng de cada muestra en un gel de agarosa al 1% teñido con GelRed™ (Biotium Inc.) (un ejemplo se muestra en la **Figura 6**).



**Figura 6.** Evaluación de la integridad del ADN genómico a través de electroforesis en gel de agarosa. Los carriles marcados 1-7 contienen ADN genómico de 7 miembros de una familia con agregación familiar de CVB.

Finalmente, se prepararon alícuotas de 1µg que fueron almacenadas a -80 °C y



posteriormente enviadas a una compañía especializada en Alemania (LGC<sup>®</sup> Limited), para la genotipificación.

### 4.3. Genotipificación, Control de Calidad y Estructura de los Datos

Los familiares y los casos esporádicos con CVB fueron genotipificados mediante un chip Illumina OmniExpress, mientras que los participantes controles, con un *beadchip* Illumina Human610-Quad. Los dos chips incluyen al menos 700.000 SNPs distribuidos en todo el genoma. Como parte de los procedimientos del control de calidad en estudios de asociación, se excluyeron los duplicados intencionales, las muestras con más del 5% de genotipos faltantes y los polimorfismos no autosómicos [130]. Los datos fueron proporcionados por el Grupo de Estadística Genética - IMBI, Universidad de Heidelberg – Alemania, en formato plink (.map y .ped) para el análisis de ligamiento y en formato de texto estándar (.txt) para el análisis de asociación, de forma codificada tal como se detalla en el **Material suplementario 6**.

De acuerdo con la descripción de los datos proporcionada por el Grupo de Estadística Genética, se seleccionaron sólo variantes que no estaban en desequilibrio de ligamiento ( $r^2 < 0,1$ ) con otras variantes y se efectuó un análisis de componentes principales utilizando la función *eigenstrat* disponible en [www.popgen.dk/software/index.php/Rscripts](http://www.popgen.dk/software/index.php/Rscripts) [131] para descartar cualquier subestructura en la población que pudiera conducir a falsas asociaciones (**Figura Suplementaria 1**). Además, los datos de los casos y los controles fueron estructurados y agrupados conforme a 4 u 8 variables: Rango de edad, sexo, fenotipo individual y grupo de ascendencia mapuche o grupo según los 5 primeros componentes principales. La estimación de los componentes individuales de ascendencia europea, mapuche, aymara y africana se realizó con el software ADMIXTURE, de forma supervisada empleando referencias descritas anteriormente [18, 132].

Una vez recibidos los datos, se identificaron y eliminaron los SNPs con un porcentaje de genotipos faltantes superior al 5%. Se excluyeron las posiciones que mostraron diferentes tasas de genotipos faltantes en casos y controles ( $p < 0,01$ ). Se estimó la frecuencia del alelo menor (MAF) de cada marcador y se eliminaron los SNPs con un MAF  $< 0,05$ . Se removieron

los SNPs con un  $p < 0,001$  en el test de equilibrio de Hardy–Weinberg entre los controles (**Figura Suplementaria 2**) [133, 134]. Tras el control de calidad, permanecieron 9.450 variantes distribuidas a lo largo de todo el genoma para el análisis de ligamiento y 1.161 posiciones localizadas en el cromosoma 9: 105-116 Mb y cromosoma 16: 0-6 Mb para el análisis de asociación genética.

#### **4.4. Análisis de Ligamiento Genético**

##### **4.4.1. Métodos de ligamiento genético.**

En esta tesis se llevaron a cabo análisis de ligamiento paramétrico y no paramétrico, utilizando el programa Mendel versión 16.0 [101]. Los umbrales sugerentes y de significancia se fijaron con base en lo reportado en la literatura [135, 136]. Los parámetros generales del análisis (**Material Suplementario 7**) se ajustaron en los archivos control siguiendo el manual de referencia, de la siguiente manera [136-138]:

##### *1. Análisis de ligamiento paramétrico*

- Nivel de significancia para el genoma completo:  $LOD \geq 3,3$ .
- Umbral sugerente de ligamiento:  $LOD \geq 1,86$ .
- Se utilizó un modelo de penetrancia simple, en donde la función de penetrancia es uniforme para todos los individuos. Por ejemplo, los siguientes comandos establecen las penetrancias del fenotipo *AFFECTED* en el locus hipotético “CVB” para cada uno de los tres genotipos:
  - Penetrancia = 0,05 :: A/A
  - Penetrancia = 0,95 :: A/B
  - Penetrancia = 0,95 :: B/B

En este modelo, la penetrancia del fenotipo “NORMAL” para el genotipo A/A sería de 0,95. Sin embargo, el modelo no supone la existencia de más de un locus predisponente que explique el 0,05 de penetrancia del fenotipo *AFFECTED* para el genotipo A/A.

- Modelos de herencia autosómico dominante, recesivo y aditivo.
- Frecuencia del alelo de susceptibilidad: 0,0001.

- Modelo I: Alelo/s de susceptibilidad al CVB con agregación familiar (Afectados = 6 individuos con CVB).
- Modelo II: Alelo/s de susceptibilidad al CVB y cálculos biliares con agregación familiar (Afectados = 6 individuos con CVB + 15 individuos colecistectomizados, ver página 49 donde se justifica este modelo).

Los parámetros del modelo genético utilizado se fijaron considerando: (1) Los riesgos relativos estimados para individuos con familiares de primer a tercer grado con CVB estimados por Hemminki *et al.* [81] y Jackson *et al.* [82], (2,3 - 2,5) que equivalen a un tamaño de efecto genético intermedio, para el cual se espera (2) una baja frecuencia alélica (menor de 0,05) de la variante causal [139]. (3) La heterogeneidad alélica se evaluó usando el modelo de mezcla habitual [140, 141] en el que se estima el parámetro  $\alpha$  [Alpha] (la probabilidad de que en cualquier genealogía se esté segregando el alelo de enfermedad ligado a la posición actual  $\theta$ ), y la puntuación LOD se maximiza tanto en la distancia genética entre el locus marcador y la enfermedad ( $\theta$ ), así como también sobre la proporción de familias ligadas ( $\alpha$ ). Siguiendo estos criterios, los mejores resultados se obtuvieron utilizando una penetrancia del 95% y una frecuencia del alelo de susceptibilidad de 0,0001 (ver más adelante la descripción de los análisis de sensibilidad). Dado que en las regiones candidatas encontradas mediante el análisis de ligamiento paramétrico  $\alpha$  siempre fue igual a 1 (todas las genealogías presentan un locus de la enfermedad cercano los marcadores identificados), no fue necesario estimar la probabilidad posterior de cada genealogía dado el ligamiento. (4) En cuanto a la heterogeneidad de la enfermedad, que es significativamente más difícil de considerar que la heterogeneidad genética, es ampliamente aceptado que las neoplasias del tracto biliar no son una enfermedad única, sino un grupo de diferentes enfermedades con características demográficas, moleculares y opciones de tratamiento distintas. Teniendo en cuenta esta heterogeneidad del CVB, los diagnósticos histopatológicos confirmados de CVB fueron considerados como un requisito indispensable para los casos índice, lo que permitió que se descartaran posibles casos de cáncer de las vías biliares y hepáticos reportados por los familiares como casos de CVB.

## II. *Análisis de ligamiento no paramétrico*

- Nivel de significancia para el genoma completo:  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ).
- Umbral sugerente de ligamiento:  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ).
- Modelo I: Alelo/s de susceptibilidad al CVB con agregación familiar (Afectados = 6 individuos con CVB).
- Modelo II: Alelo/s de susceptibilidad al CVB y cálculos biliares con agregación familiar (Afectados = 6 individuos con CVB + 15 individuos colecistectomizados).

A los individuos que informaron estar sanos o tener colelitiasis, pero no haber sido colecistectomizados ( $n=3$ ), así como también aquellos sin información, se les asignó, tanto en el análisis paramétrico como no paramétrico, el fenotipo “*unknown*”.

Mediante la estimación del coeficiente de correlación de Spearman se evaluó la posible asociación o interdependencia entre las puntuaciones (LOD o valores  $p$ ) obtenidas entre los modelos I y II.

### **4.4.2. *Análisis de sensibilidad***

El impacto de la edad al diagnóstico y colecistectomía de los individuos, así como también la contribución de cada familia en los resultados del ligamiento genético fueron cuantificados a través de un análisis de sensibilidad como se indica a continuación:

#### *4.4.2.1. Contribuciones por edad de diagnóstico, colecistectomía y fenotipo*

Se definieron 12 modelos de análisis con base en la definición de “individuo afectado” (CVB o CVB + CCA) y en seis edades límite de diagnóstico:

#### *I. Afectados; todos los individuos diagnosticados con CVB:*

- a. A cualquier edad.
- b. Antes de los 71 años.
- c. Antes de los 70 años.
- d. Antes de los 65 años.
- e. Antes de los 60 años.
- f. Antes de los 55 años.

#### *II. Afectados; todos los individuos diagnosticados con CVB, más los colecistectomizados:*

- a. A cualquier edad.
- b. Antes de los 71 años.
- c. Antes de los 70 años.
- d. Antes de los 65 años.
- e. Antes de los 60 años.
- f. Antes de los 55 años.

Por ejemplo, en el modelo II-f, abreviado como “CVB + CCA<55”, se consideraron afectados aquellos individuos con un diagnóstico de CVB y aquellos con un historial de colecistectomía antes de cumplir los 55 años. En el modelo I-c denominado “CVB < 70”, los individuos se consideraron afectados si tenían diagnóstico de CVB antes de cumplir 70 años. Los individuos colecistectomizados a edades por sobre los límites establecidos, fueron considerados en los análisis como “sanos”.

#### *4.4.2.2. Contribución de cada familia a los resultados de ligamiento*

Se realizaron análisis de ligamiento no paramétrico para estimar las contribuciones individuales de cada familia, con los parámetros generales anteriormente descritos y definiendo los afectados como aquellos individuos diagnosticados con CVB o con CVB más los colecistectomizados a cualquier edad. Adicionalmente se calcularon y representaron gráficamente las contribuciones acumuladas de acuerdo a la contribución de cada familia en orden descendente, explícitamente:

- a. Familia 1
- b. Familia 1+3
- c. Familia 1+3+8
- d. Familia 1+3+8+...

#### *4.4.2.3. Variación de los parámetros en el ligamiento paramétrico*

El efecto de la variación en los parámetros del modelo genético, se registraron en un análisis de sensibilidad para el ligamiento paramétrico en los cromosomas 7, 9 y 16 utilizando los parámetros que se detallan a continuación:

Penetrancia: 70%, 80%, 90%, 95%, 99%.

Tasa de fenocopias: 30%, 20%, 10%, 5%, 1%.

Frecuencia del alelo de susceptibilidad: 0,0001; 0,001; 0,01; 0,05.

#### **4.5. Búsqueda de Genes Candidatos por Anotación Funcional**

Con el fin de identificar potenciales genes candidatos, se extrajo desde Ensembl la lista de genes codificantes contenidos en las regiones identificadas mediante ligamiento no paramétrico: Chr9:105339690-115983216 y Chr16:3707747-5878184 (GRCh37). Posteriormente se realizó un análisis de clasificación y anotación funcional mediante las herramientas PANTHER™ Versión 12.0 [104] y DAVID 6.8 [105, 106], respectivamente. Usando Cytoscape 3.5.1 [142, 143] se generó un gráfico de redes de interacción molecular para los genes que podrían albergar variantes causales, identificados en las regiones candidatas de ligamiento para el CVB y los cálculos biliares.

#### **4.6. Métodos de Mapeo Fino por Asociación Genética**

Para cada variante seleccionada, se realizó un test de asociación mediante regresión logística múltiple, considerando el fenotipo como variable de respuesta y el genotipo como variable explicatoria bajo los modelos de herencia dominante y aditivo por separado. Para definir las covariables a incluir en el modelo, se realizó un análisis exploratorio de los datos comparando la distribución de los casos y los controles por sexo, nivel de edad y proporción de ascendencia mapuche mediante una prueba  $\chi^2$  de Pearson. Así, se incluyó en el modelo de regresión las covariables edad, sexo y proporción de ascendencia mapuche, o alternativamente edad, sexo y los 5 primeros componentes principales agrupados en cuartiles (**Material Suplementario 6**). El umbral de significancia a nivel de genoma completo se fijó en  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ), mientras que el umbral de significancia para las variantes genéticas seleccionadas se fijó mediante la corrección de Bonferroni en  $p < 0,05/1.028$  SNPs =  $4,9e^{-5}$  ( $-\log_{10}(p) > 4,3$ ). Los ORs se calcularon mediante la exponenciación de los coeficientes de regresión estimados ( $e^{\beta_1}$ ). Todos los análisis se realizaron en R versión 3.4.0 (para mayor detalle de las librerías usadas refiérase a la sección Anexos).

## IV. Resultados

### 1. Análisis Descriptivo de las Familias y los Individuos

La **Tabla 6** compara la distribución de 12 variables en individuos afectados por CVB, colecistectomizados (CCA) por cálculos biliares e individuos no afectados (NAF) pertenecientes a 7 familias. Los individuos con CVB presentaron una edad de cirugía mayor que los colecistectomizados ( $p = 0,01$ ). Las 11 variables restantes no mostraron diferencias significativas entre los tres grupos de individuos. Al analizar los individuos afectados en conjunto (CVB + CCA) frente a los no afectados, no hubo variación en los resultados antes mencionados.

**Tabla 6.** Estadística descriptiva para 12 variables con información completa, analizadas en individuos con CVB, CCA y NAF pertenecientes a 7 familias.

CVB: Cáncer de Vesícula Biliar, CCA: Colecistectomizados, NAF: No afectados.

Variable	Niveles	CVB (n=6)		CCA (n=15)		NAF (n=27)		p-val
		n	%	n	%	n	%	
Sexo	F	5	83,3	13	86,7	19	70,4	0,40 <sup>§</sup>
	M	1	16,7	2	13,3	8	29,6	
Edad de cirugía	<34	0	0,0	6	40,0	NA	-	0,01 <sup>†</sup>
	34-44	0	0,0	3	20,0	NA	-	
	45-55	2	33,3	5	33,3	NA	-	
	>55	4	66,7	1	6,7	NA	-	
Otro cáncer	No	5	83,3	14	93,3	26	96,3	0,69 <sup>§</sup>
	Si	1	16,7	1	6,7	1	3,7	
Diabetes	No	4	80,0	8	53,3	22	81,5	0,13 <sup>§</sup>
	Si	1	20,0	7	46,7	5	18,5	
Hipertensión	No	3	60,0	6	40,0	17	63,0	0,18 <sup>§</sup>
	Si	2	40,0	9	60,0	10	37,0	
Colesterol alto	No	2	40,0	9	60,0	18	66,7	0,48 <sup>§</sup>
	Si	3	60,0	6	40,0	9	33,3	
Peso (Kg)	<60	1	20,0	3	42,9	3	20,0	0,31 <sup>†</sup>
	60-65	0	0,0	1	14,3	3	20,0	
	66-74	4	80,0	3	42,9	2	13,3	
	>75	0	0,0	0	0,0	7	46,7	
IMC	<18,5	0	0,0	0	0,0	0	0,0	0,59 <sup>†</sup>
	18,5-24,99	1	20,0	4	57,1	5	33,3	
	25-29,99	4	80,0	3	42,9	6	40,0	
	>30	0	0,0	0	0,0	4	26,7	

<b>Actividad física</b>	No	5	83,3	13	86,7	20	74,1	0,45 <sup>§</sup>
	Si	1	16,7	2	13,3	7	25,9	
<b>Uso de anticonceptivos</b>	Nunca he usado	1	20,0	4	33,3	6	31,6	0,63 <sup>§</sup>
	Usaba	4	80,0	8	66,7	10	52,6	
	Uso actualmente	0	0,0	0	0,0	3	15,8	
<b>Número de embarazos</b>	<1	0	0,0	3	21,4	6	35,3	0,12 <sup>†</sup>
	1-2	1	25,0	2	14,3	4	23,5	
	3-4	2	50,0	3	21,4	4	23,5	
	>4	1	25,0	6	42,9	3	17,6	
<b>Terapia de reemplazo hormonal</b>	Nunca he usado	5	100,0	11	84,6	14	82,4	0,90 <sup>§</sup>
	Usaba	0	0,0	2	15,4	3	17,6	
	Uso actualmente	0	0,0	0	0,0	0	0,0	

§ Test Chi cuadrado de Pearson con  $p$  valor simulado (10.000 réplicas)

† Test Jonckheere-Terpstra (10.000 réplicas).

Respecto a la ancestría, las familias investigadas presentaron niveles similares en sus porcentajes de ascendencia genética, con promedios de 49% europeo, 39% mapuche, 9% aymara y 3% africano (**Tabla 7**).

**Tabla 7.** Promedio estimado de las proporciones de ascendencia europea, africana, mapuche y aymara en 7 familias chilenas con múltiples casos de CVB.

<b>Familia</b>	<b>Ascendencia</b>			
	Europea	Mapuche	Aymara	Africana
1	0,51	0,38	0,09	0,03
2	0,51	0,38	0,09	0,02
3	0,44	0,35	0,16	0,04
4	0,47	0,40	0,09	0,04
5	0,51	0,43	0,05	0,01
6	0,51	0,38	0,09	0,03
7	0,50	0,38	0,09	0,02
Promedio ± DE	0,49 ± 0,03	0,39 ± 0,02	0,09±0,03	0,03 ± 0,01

La **Tabla 8** muestra las características demográficas de los 2.145 individuos incluidos en el mapeo fino. La mediana de edad de los participantes en el estudio fue de 27 años y 44% eran mujeres. Las regiones chilenas con mayor número de participantes fueron De Arica y Parinacota (34%), Metropolitana de Santiago (18%) y Del Biobío (10%).



**Tabla 8.** Distribución de los individuos por sexo, rango de edad, nivel de educación y componentes de ascendencia, incluidos en el análisis de mapeo fino por asociación en dos regiones candidatas de ligamiento.

Variable Demográfica	Total (n=2145)		Casos CVB (n=183)		Controles (n=1962)	
	n	%	n	%	n	%
Sexo						
Hombres	1181	56	42	24	1139	58
Mujeres	948	44	138	76	810	42
Edad						
0-19	17	0	1	0	16	0
20-29	1074	50	3	2	1071	54
30-39	474	22	3	2	471	24
40-49	174	8	20	12	154	8
50-59	134	6	42	24	92	4
>=60	250	12	106	58	144	8
Faltante	6	0	5	2	1	0
Nivel de educación						
Escuela primaria	135	6	95	52	40	2
Escuela secundaria	1330	62	61	34	1269	66
Educación técnica	71	4	6	4	65	4
Universidad	400	18	11	6	389	20
Postgrado	55	2	1	0	54	2
Faltante	138	6	6	4	132	6
Región						
Arica	707	34	3	2	704	36
Tarapacá	64	4	0	0	64	4
Antofagasta	89	4	2	2	87	4
Atacama	27	2	1	0	26	2
Coquimbo	29	2	4	2	25	2
Valparaíso	92	4	1	0	91	4
Metropolitana	389	18	43	24	346	18
O'Higgins	51	2	10	6	41	2
Maule	165	8	24	14	141	8
Biobío	206	10	26	14	180	10
Araucanía	122	6	13	8	109	6
Ríos	55	2	26	14	29	2
Lagos	47	2	14	8	33	2
Aisén	8	0	1	0	7	0
Magallanes	5	0	0	0	5	0
Otro país	5	0	0	0	5	0
Faltante	68	4	12	6	56	2
Ascendencia Europea (%)						
0-9	25	2	5	2	20	2
10-19	26	2	5	2	21	2
20-30	81	4	4	2	77	4
30-39	161	8	7	4	154	8

40-49	493	24	39	22	454	24
50-59	963	46	83	46	880	46
60-69	294	14	27	16	267	14
70-79	47	2	6	4	41	2
80-89	23	2	3	2	20	2
90-100	16	0	1	0	15	0
Ascendencia Mapuche (%)						
0-9	27	2	2	2	25	2
10-19	59	2	3	2	56	2
20-30	243	12	5	2	238	12
30-39	495	24	24	14	471	24
40-49	844	40	73	40	771	40
50-59	329	16	49	28	280	14
60-69	76	4	11	6	65	4
70-79	32	2	4	2	28	2
80-89	6	0	2	2	4	0
90-100	18	0	7	4	11	0
Ascendencia Aymara (%)						
0-9	669	32	87	48	582	30
10-19	894	42	84	46	810	42
20-30	258	12	5	2	253	12
30-39	108	6	1	0	107	6
40-49	60	2	0	0	60	4
50-59	58	2	0	0	58	2
60-69	43	2	1	0	42	2
70-79	12	0	1	0	11	0
80-89	10	0	0	0	10	0
90-100	17	0	1	0	16	0
Ascendencia Africana (%)						
0-9	1943	92	175	98	1768	90
10-19	176	8	5	2	171	8
20-30	10	0	0	0	10	0

La **Tabla 9** compara la distribución según sexo, edad y proporción de ascendencia mapuche de los casos y controles incluidos en el estudio. Como es de esperar, dada la diferente incidencia del CVB en mujeres y hombres, la proporción de mujeres fue mayor en los casos con CVB (77%) que en los controles (23%) ( $p < 0,0001$ ). Asimismo, los individuos mayores de 50 años fueron más frecuentes entre los casos con CVB que entre los controles ( $p < 0,0001$ ). Finalmente, un 42,6% de los casos con CVB presentaron una proporción de ascendencia mapuche en el cuarto cuartil, mientras que los controles mostraron una distribución uniforme en cuanto a las proporciones mapuche ( $p < 0,0001$ ). Dadas las

diferencias entre los casos y controles, se consideró un análisis pareado empleando una submuestra de controles que fuese comparable con los casos. Sin embargo, el número de controles no fue suficiente para aparear a los individuos por sexo, edad y ascendencia mapuche.

**Tabla 9.** Distribución de los casos con CVB y controles por sexo, edad y proporción de ascendencia mapuche.

Variable	Niveles	Controles (n=1962)		CVB (n=183)		p <sup>§</sup>
		n	%	n	%	
Sexo	Hombre	1139	58,1	42	23,0	<0,0001
	Mujer	823	41,9	141	77,0	
Edad	≤ 50	1729	88,1	34	18,6	<0,0001
	>50	223	11,4	149	81,4	
Ascendencia Mapuche	1 (≤ 30%)	515	26,2	16	8,7	<0,0001
	2 (> 30% y ≤ 38%)	497	25,3	39	21,3	
	3 (> 38% y < 44%)	496	25,3	50	27,3	
	4 (≥ 44%)	454	23,1	78	42,6	

<sup>§</sup> Test Chi cuadrado de Pearson.

## 2. Estimación del riesgo de recurrencia y riesgo relativo en las familias

Las **Tablas 10 - 12** muestran los riesgos relativos estimados a partir de la información registrada en las entrevistas a los probandos con CVB de las 16 familias con agregación de ese fenotipo. Los individuos con familiares de primer grado afectados con CVB presentan un riesgo 6,2 veces de desarrollar CVB mayor que un individuo de la población general, mientras que los antecedentes de segundo y tercer grado proporcionan un riesgo 3,6 veces mayor, cada uno. El riesgo que proporciona tener un padre con CVB resultó ser sorprendentemente alto; casi 21 veces mayor que el de un individuo de la población general (**Tabla 10**).

Por otro lado, el riesgo de coleditiasis para individuos con familiares de primer grado afectados con coleditiasis presentan el mismo riesgo que individuos de la población general (**Tabla 11**). Mientras que el riesgo de CVB para individuos con coleditiasis que además presentan parientes de primer grado (padres o hermanos) con CVB es 2,7 veces mayor que un individuo con coleditiasis de la población general (**Tabla 12**).

**Tabla 10.** Riesgo relativo de CVB para individuos con familiares afectados del primer al tercer grado de consanguinidad en 16 familias chilenas.

Grado de parentesco de familiares del <b>probando con CVB</b>		<b>Nº de familiares</b>	<b>Nº de afectados con CVB</b>	<b>Riesgo de recurrencia empírico</b>	<b>Riesgo relativo (<math>\lambda</math>)*</b>
<b>Parientes de 1<sup>er</sup> grado</b>	Padres	32	8	0,25	20,8
	Hijos	45	0	0,00	0,0
	Hermanos	98	5	0,05	4,3
	<b>Total</b>	<b>175</b>	<b>13</b>	<b>0,07</b>	<b>6,2</b>
<b>Parientes de 2<sup>o</sup> grado</b>	Medios hermanos	5	0	0,00	0,0
	Abuelos	16	0	0,00	0,0
	Sobrinos	41	1	0,02	2,0
	Tíos	31	3	0,10	8,1
	<b>Total</b>	<b>93</b>	<b>4</b>	<b>0,04</b>	<b>3,6</b>
<b>Parientes de 3<sup>er</sup> grado</b>	Primos	21	0	0,00	0,0
	Medios tíos	2	1	0,50	41,7
	<b>Total</b>	<b>23</b>	<b>1</b>	<b>0,04</b>	<b>3,6</b>

\* Estimado con base en el riesgo acumulado de CVB en la población general (1,2%).

**Tabla 11.** Riesgo relativo de colecistitis para individuos con familiares afectados del primer al tercer grado de consanguinidad en 16 familias chilenas.

Grado de parentesco de familiares del <b>probando con CVB</b>		<b>Nº de familiares</b>	<b>Nº de afectados con CVB o colecistitis</b>	<b>Riesgo de recurrencia empírico</b>	<b>Riesgo relativo (<math>\lambda</math>)*</b>
<b>Parientes de 1<sup>er</sup> grado</b>	Padres	32	11	0,34	1,3
	Hijos	45	4	0,10	0,3
	Hermanos	98	30	0,30	1,1
	<b>Total</b>	<b>175</b>	<b>45</b>	<b>0,26</b>	<b>1,0</b>
<b>Parientes de 2º grado</b>	Medios hermanos	5	1	0,20	0,7
	Abuelos	16	0	0,00	0,0
	Sobrinos	41	3	0,07	0,3
	Tíos	31	5	0,16	0,6
	<b>Total</b>	<b>93</b>	<b>9</b>	<b>0,10</b>	<b>0,4</b>
<b>Parientes de 3<sup>er</sup> grado</b>	Primos	21	2	0,09	0,4
	Medios tíos	2	1	0,50	1,9
	<b>Total</b>	<b>23</b>	<b>3</b>	<b>0,13</b>	<b>0,5</b>

\* Estimado con base en la prevalencia de colecistitis en la población general (27%).

**Tabla 12.** Riesgo relativo de CVB para individuos con colestiasis y antecedentes familiares de CVB en 16 familias chilenas.

Grado de parentesco de familiares del <b>probando con colestiasis</b>		<b>Nº de familiares</b>	<b>Nº de afectados con CVB</b>	<b>Riesgo de recurrencia empírico</b>	<b>Riesgo relativo (<math>\lambda</math>)*</b>
<b>Parientes de 1<sup>er</sup> grado</b>	Padres	24	7	0,29	3,5
	Hermanos	60	12	0,20	2,4
	<b>Total</b>	<b>84</b>	<b>19</b>	<b>0,23</b>	<b>2,7</b>
<b>Parientes de 2º grado</b>	Medios hermanos	4	1	0,25	3,0
	Abuelos	22	1	0,05	0,5
	Tíos	41	3	0,07	0,9
	<b>Total</b>	<b>67</b>	<b>5</b>	<b>0,07</b>	<b>0,9</b>
<b>Parientes de 3<sup>er</sup> grado</b>	Primos	15	2	0,13	1,6
	Medios tíos	8	1	0,12	1,5
	<b>Total</b>	<b>23</b>	<b>3</b>	<b>0,13</b>	<b>1,6</b>

\* Estimado con base en el riesgo de CVB para individuos con colestiasis de la población general (8,4%) [144].

### 3. Análisis de Ligamiento Genético

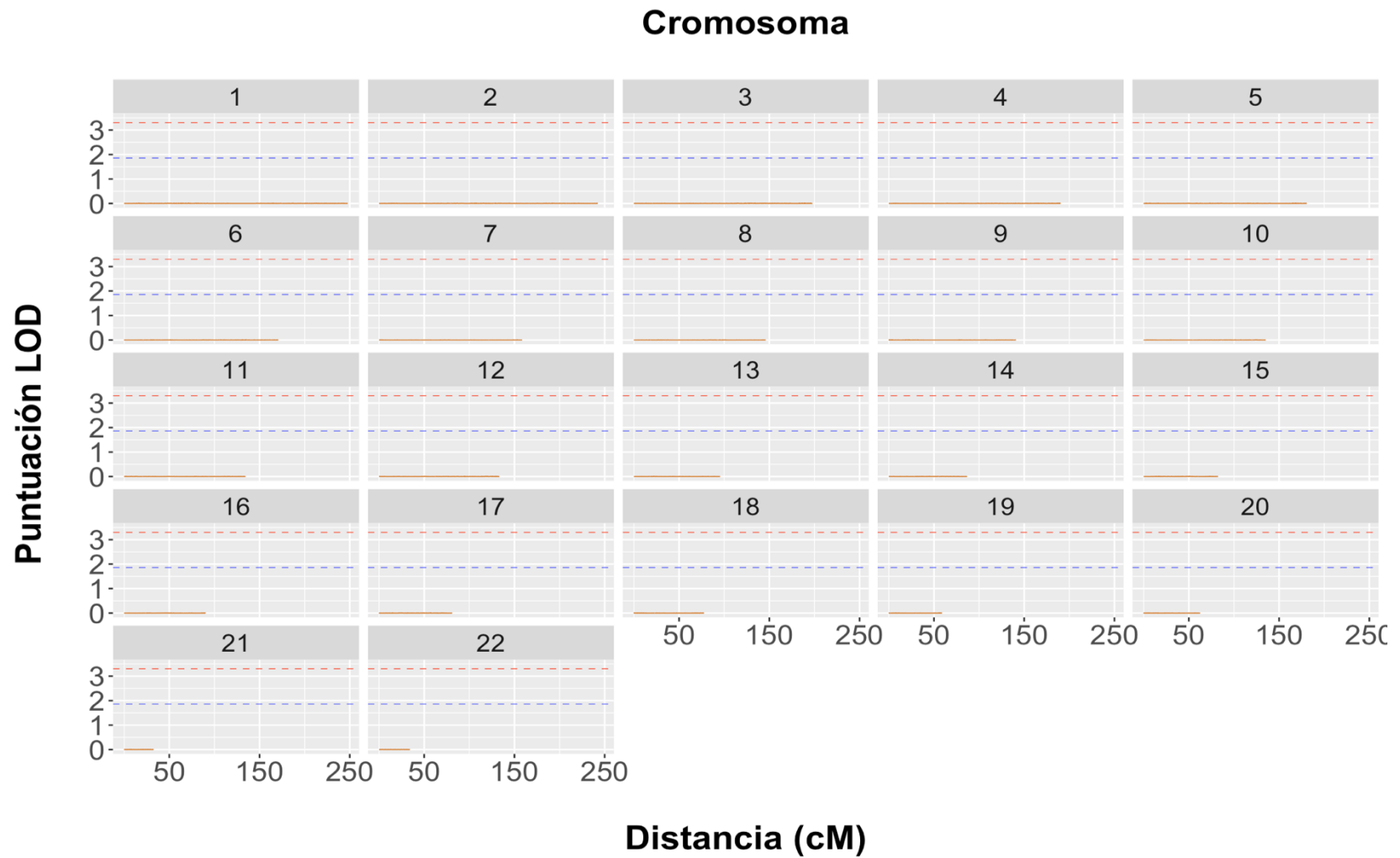
#### 3.1. Ligamiento Paramétrico

El análisis de ligamiento paramétrico arrojó los siguientes resultados, de acuerdo al modelo utilizado. Para el *Modelo I* de alelo de susceptibilidad al CVB (Afectados = 6 individuos con CVB), la máxima puntuación LOD observada fue de 0,012 en 5 cromosomas (**Tabla 13, Figura 7**). Para el *Modelo II* de alelo de susceptibilidad a CVB y cálculos biliares (Afectados = 6 individuos con CVB + 15 con colecistectomía), se observó una máxima puntuación LOD de 1,98 en el marcador rs2437817 del cromosoma 9 (**Tabla 13, Figura 8**).

**Tabla 13.** Mejores resultados por cromosoma (Cr), obtenidos mediante análisis de ligamiento del genoma completo en 7 familias con múltiples casos de CVB y colecistectomía.

Izquierda: Máximas puntuaciones LOD obtenidas mediante un análisis de ligamiento paramétrico para los modelos I (susceptibilidad al CVB) y II (susceptibilidad al CVB y a los cálculos biliares) sin restricción de edad. Derecha: Se muestran los valores  $p$  en escala logarítmica, obtenidos mediante un análisis de ligamiento no paramétrico para los modelos I y II. En negrita se resaltan las puntuaciones con valores más altos. El nivel de significancia para el genoma completo y el umbral sugerente de ligamiento en el análisis paramétrico (ALP) son  $LOD \geq 3,3$  y  $LOD \geq 1,86$ , respectivamente. El nivel de significancia para el genoma completo y el umbral sugerente de ligamiento para el análisis no paramétrico (ALNP) son  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ), respectivamente.

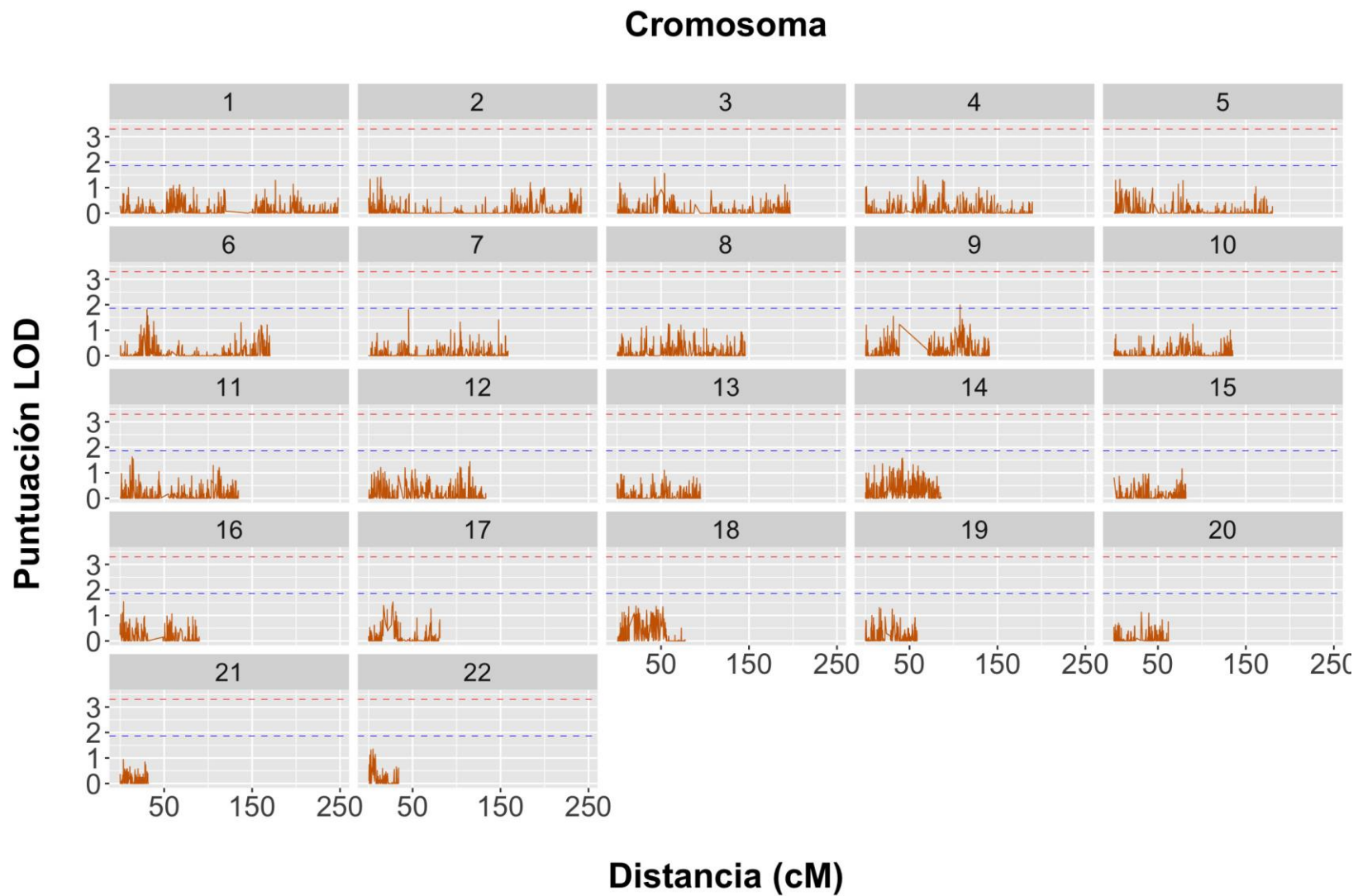
Cr	ALP		ALNP	
	CVB	CVB y CCA	CVB	CVB y CCA
	Max puntuación LOD		$-\log_{10}(p)$	
<b>1</b>	0,012	1,04	0,78	1,54
<b>2</b>	0,011	1,39	0,87	1,17
<b>3</b>	0,012	1,36	0,87	1,66
<b>4</b>	0,012	1,25	0,87	0,61
<b>5</b>	0,011	1,12	0,87	1,26
<b>6</b>	0,011	1,58	0,87	1,95
<b>7</b>	0,008	<b>1,81</b>	0,50	1,49
<b>8</b>	0,007	1,00	0,50	1,53
<b>9</b>	0,012	<b>1,98</b>	0,50	<b>2,88</b>
<b>10</b>	0,007	1,24	0,87	1,12
<b>11</b>	0,007	1,44	0,87	1,23
<b>12</b>	0,008	1,14	0,87	2,03
<b>13</b>	0,008	0,87	0,49	1,51
<b>14</b>	0,007	1,36	0,87	1,87
<b>15</b>	0,012	0,98	0,50	1,91
<b>16</b>	0,007	1,24	0,50	<b>2,42</b>
<b>17</b>	0,011	1,33	0,87	1,78
<b>18</b>	0,011	1,13	0,78	1,71
<b>19</b>	0,007	1,24	0,50	1,65
<b>20</b>	0,007	0,95	0,87	1,62
<b>21</b>	0,007	0,94	0,50	1,66
<b>22</b>	0,006	1,10	0,87	1,67



**Figura 7.** Curvas de puntuación LOD del análisis de ligamiento del genoma completo para el modelo I (alelo de susceptibilidad al CVB) en 7 familias chilenas con múltiples casos de CVB.

Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo ( $LOD \geq 3,3$ ) y el umbral sugerente de ligamiento ( $LOD \geq 1,86$ ), respectivamente.



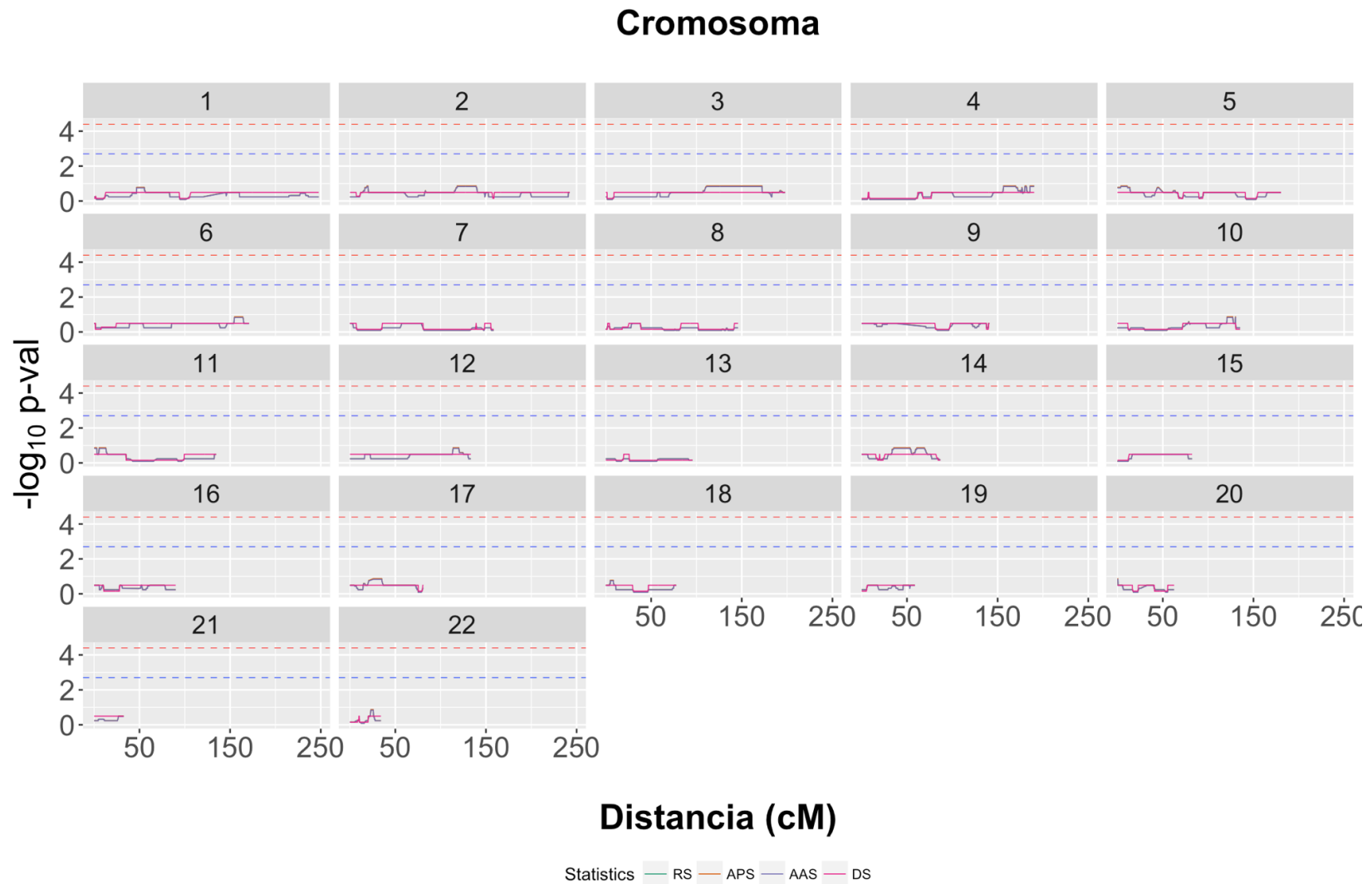


**Figura 8.** Curvas de puntuación LOD del análisis de ligamiento del genoma completo para el modelo II (alelo de susceptibilidad al CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados.

Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo ( $LOD \geq 3,3$ ) y el umbral sugerente de ligamiento ( $LOD \geq 1,86$ ), respectivamente.

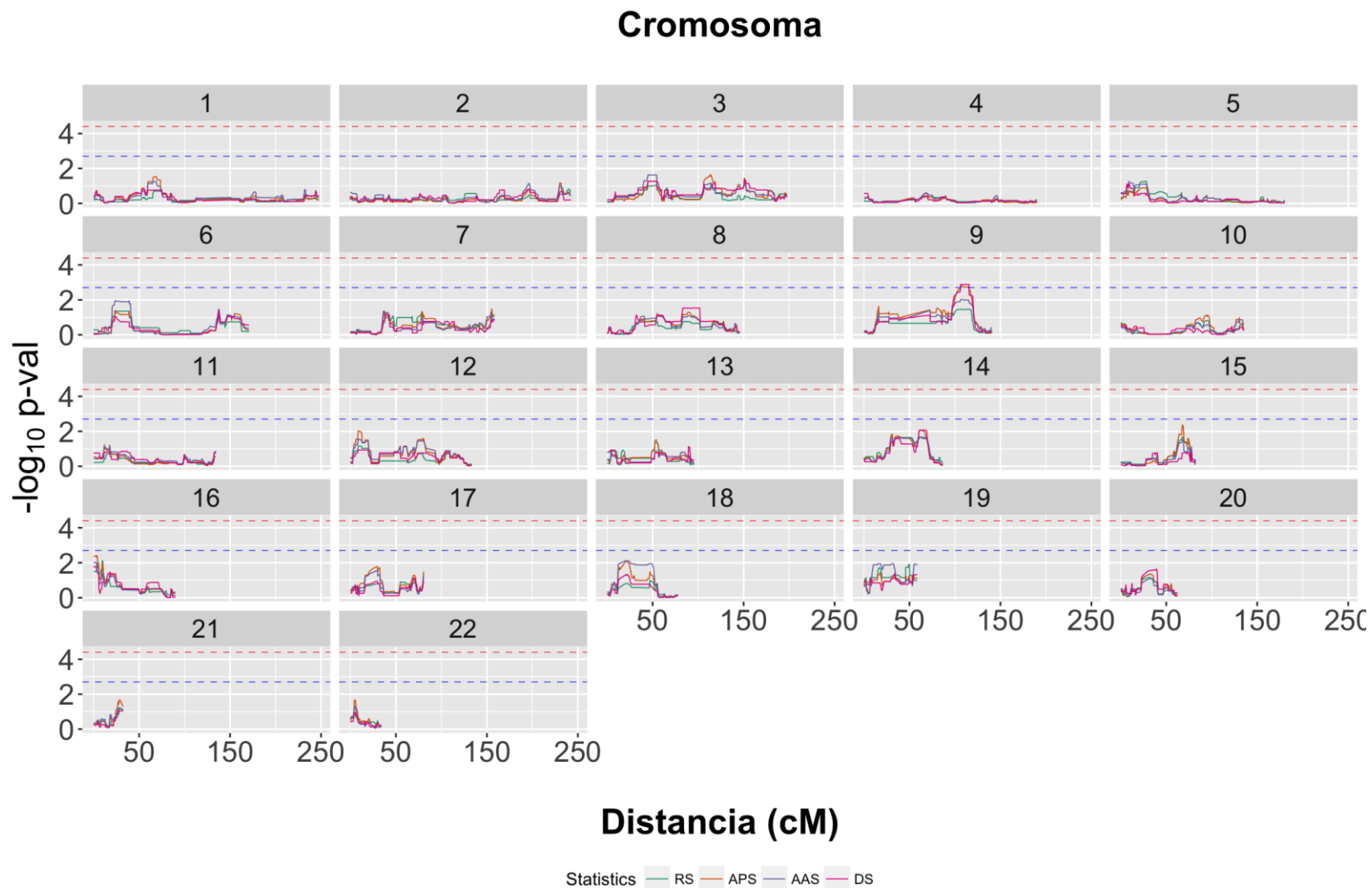
### 3.2. Ligamiento No Paramétrico.

Por su parte, el análisis de ligamiento no paramétrico arrojó los siguientes resultados, de acuerdo al modelo utilizado. Para el *Modelo I*: Alelo de susceptibilidad al CVB (Afectados = 6 individuos con CVB), el máximo valor  $p$  ( $-\log_{10}$ ) observado fue de 0,87 en los cromosomas 2, 3, 4, 5, 6, 10, 11, 12, 14, 17, 20 y 22 (**Tabla 13, Figura 9**). Para el *Modelo II*: Alelo de susceptibilidad a CVB y cálculos biliares (Afectados = 6 individuos con CVB + 15 con colecistectomía), el máximo valor  $p$  ( $-\log_{10}$ ) observado fue = 2,88 entre los marcadores rs4742892 y rs10116897 del cromosoma 9. También se observó una segunda región candidata en el Cromosoma 16 con un valor  $p$  ( $-\log_{10}$ ) de 2,42 en el marcador rs1053874 (**Tabla 13, Figuras 10-12**). La correlación de LOD entre los modelos I y II fue de 0,09 para el cromosoma 7, -0,02 para el cromosoma 9 y 0,06 para el cromosoma 16.



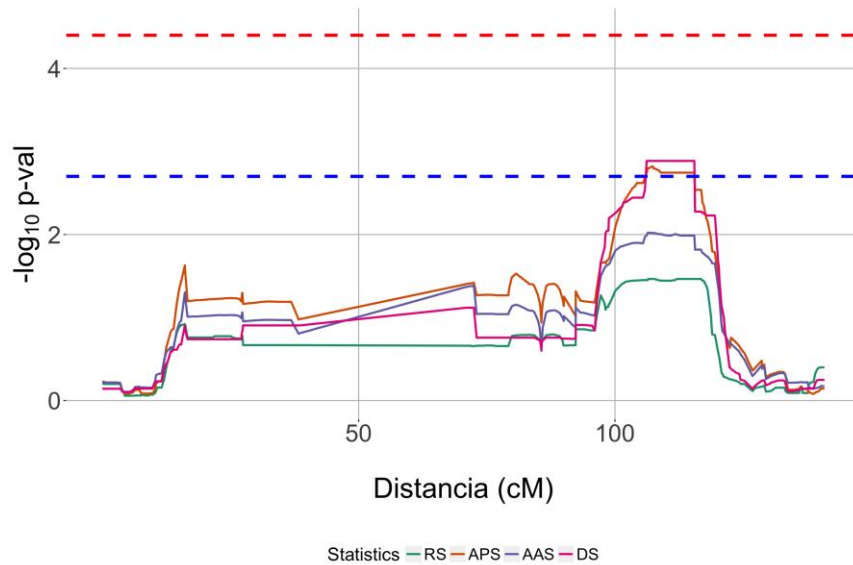
**Figura 9.** Test de ligamiento no paramétrico del genoma completo para el modelo I (alelo de susceptibilidad al CVB) en 7 familias chilenas con múltiples casos de CVB.

Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y el umbral sugerente de ligamiento  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ), respectivamente. RS: *Recessive Blocks Statistic*, APS: *Additive Pairs Statistic*, AAS: *Additive All Statistic*, DS: *Dominant Blocks Statistic*.



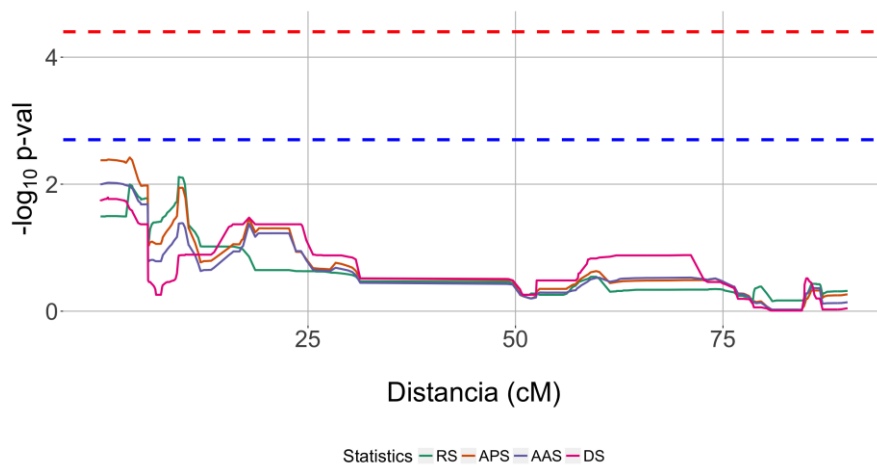
**Figura 10.** Test de ligamiento no paramétrico del genoma completo para el modelo II (alelo de susceptibilidad a CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados.

Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y el umbral sugerente de ligamiento  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ), respectivamente. RS: *Recessive Blocks Statistic*, APS: *Additive Pairs Statistic*, AAS: *Additive All Statistic*, DS: *Dominant Blocks Statistic*.



**Figura 11.** Resultados del test de ligamiento no paramétrico sobre el cromosoma 9 para el modelo II (alelo de susceptibilidad a CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados.

Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y el umbral sugerente de ligamiento  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ), respectivamente. RS: *Recessive Blocks Statistic*, APS: *Additive Pairs Statistic*, AAS: *Additive All Statistic*, DS: *Dominant Blocks Statistic*.



**Figura 12.** Resultados del test de ligamiento no paramétrico sobre el cromosoma 16 para el modelo II (alelo de susceptibilidad a CVB y cálculos biliares) en 7 familias chilenas con múltiples casos de CVB e individuos colecistectomizados.

Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y el umbral sugerente de ligamiento  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ), respectivamente. RS: *Recessive Blocks Statistic*, APS: *Additive Pairs Statistic*, AAS: *Additive All Statistic*, DS: *Dominant Blocks Statistic*.

### 3.3. Análisis de Sensibilidad

#### 3.3.1. Modelos con restricción por edad de diagnóstico o colecistectomía, y considerando dos definiciones de individuos afectados (CVB y CVB + CCA).

En la **Tabla 14** se resumen las máximas puntuaciones LOD por cromosoma obtenidas mediante el análisis de ligamiento paramétrico en 12 modelos diferentes con dos definiciones de los individuos afectados de acuerdo al fenotipo (CVB, o bien CVB + colecistectomía) y con 6 restricciones por edad de diagnóstico o cirugía.

**Tabla 14.** Máximas puntuaciones LOD por cromosoma obtenidas en 12 modelos definidos para el análisis de sensibilidad del ligamiento genético.

En la tabla se muestra un resumen de las máximas puntuaciones LOD por cromosoma (Cr) obtenidas mediante análisis de ligamiento paramétrico en 7 familias chilenas con múltiples casos de CVB. Se estudiaron 12 modelos diferentes con dos definiciones de los casos de acuerdo al fenotipo (CVB o CVB+CCA) y con 6 restricciones de los mismos por edad de diagnóstico o cirugía. CVB: Cáncer de Vesícula Biliar, CCA: Colecistectomía, Edades: Todas las edades (TE), 71, 70, 65, 60 y 55 años. Nivel de significancia para el genoma completo y umbral sugerente de ligamiento para el análisis de ligamiento paramétrico:  $LOD \geq 3,3$ ;  $LOD \geq 1,86$ , respectivamente. En negrita se resaltan las puntuaciones con valores más altos.

Cr	TE	CVB					CVB+CCA					
		< 71	< 70	< 65	< 60	< 55	TE	< 71	< 70	< 65	< 60	< 55
1	0,012	0,02	0,02	0,02	0,02	0,004	1,04	0,88	0,88	0,88	1,02	1,35
2	0,011	0,02	0,02	0,02	0,02	0,003	1,39	1,21	1,21	1,21	1,36	<b>2,14</b>
3	0,012	0,02	0,02	0,02	0,02	0,003	1,36	1,25	1,25	1,25	1,39	1,74
4	0,012	0,02	0,02	0,02	0,02	0,003	1,25	0,98	0,98	0,98	0,87	1,6
5	0,011	0,02	0,02	0,02	0,02	0,002	1,12	1,24	1,24	1,24	1,21	1,78
6	0,011	0,02	0,02	0,02	0,02	0,002	1,58	1,00	1,00	1,00	1,25	1,19
7	0,008	0,02	0,02	0,02	0,01	0,003	<b>1,81</b>	0,87	0,87	0,87	0,74	1,16
8	0,007	0,02	0,02	0,02	0,02	0,003	1,00	1,37	1,37	1,37	1,62	<b>1,97</b>
9	0,012	0,02	0,02	0,02	0,02	0,002	<b>1,98</b>	1,37	1,37	1,37	1,22	1,45
10	0,007	0,02	0,02	0,02	0,02	0,002	1,24	0,74	0,74	0,74	0,96	1,4
11	0,007	0,02	0,02	0,02	0,02	0,005	1,44	1,15	1,15	1,15	1,23	1,73
12	0,008	0,02	0,02	0,02	0,01	0,002	1,14	1,21	1,21	1,21	1,19	1,75
13	0,008	0,02	0,02	0,02	0,02	0,003	0,87	1,00	1,00	1,00	1,15	1,82
14	0,007	0,01	0,01	0,01	0,01	0,002	1,36	1,23	1,23	1,23	1,48	1,72
15	0,012	0,02	0,02	0,02	0,02	0,004	0,98	1,07	1,07	1,07	1,32	1,46
16	0,007	0,02	0,02	0,02	0,02	0,003	1,24	1,18	1,18	1,18	1,09	1,37
17	0,011	0,01	0,01	0,01	0,01	0,002	1,33	1,21	1,21	1,21	1,46	1,68
18	0,011	0,02	0,02	0,02	0,02	0,002	1,13	0,86	0,86	0,86	1,08	1,12
19	0,007	0,01	0,01	0,01	0,01	0,003	1,24	0,91	0,91	0,91	1,05	1,22
20	0,007	0,01	0,01	0,01	0,02	0,002	0,95	1,36	1,36	1,36	1,5	1,73
21	0,007	0,02	0,02	0,02	0,02	0,002	0,94	0,8	0,8	0,8	0,89	0,76
22	0,006	0,01	0,01	0,01	0,02	0,002	1,10	0,44	0,44	0,44	0,32	0,51

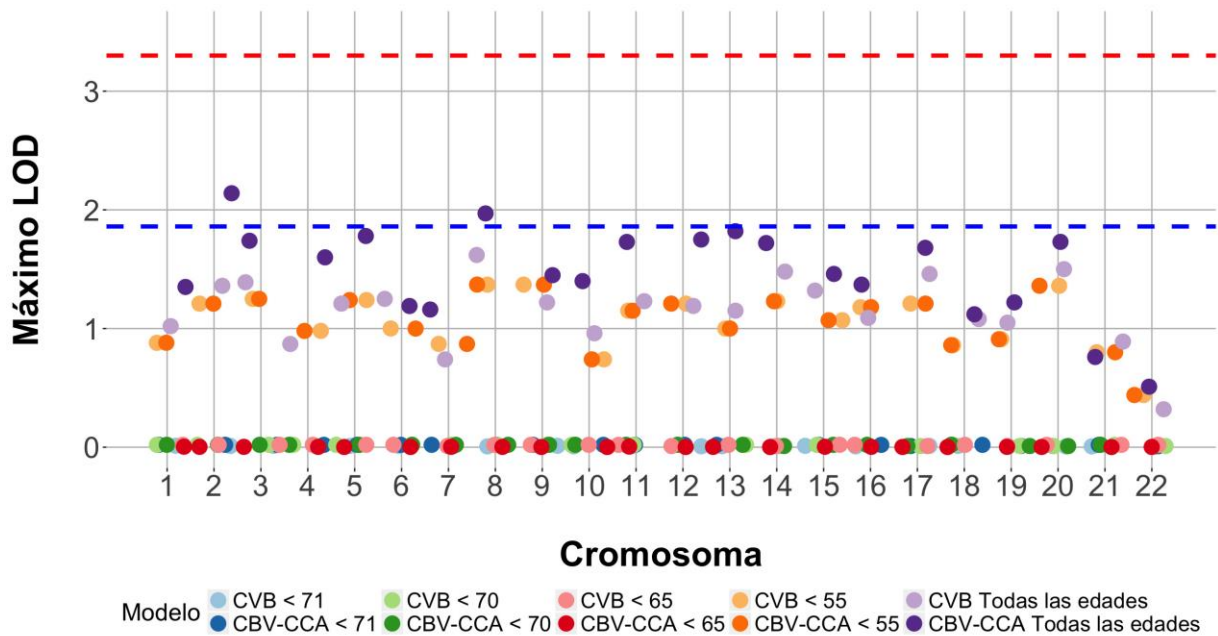
Al analizar los 12 modelos en conjunto se observó el importante efecto que tiene la definición del fenotipo afectado en las puntuaciones LOD (**Figura 13**). En tres modelos (“CVB+CCA < 55”, “CVB+CCA < 65” y “CVB+CCA Todas las edades”) se observaron puntuaciones LOD superiores a 1 en la mayoría de los cromosomas, mientras que los nueve modelos restantes arrojaron puntuaciones LOD máximas cercanas a 0.

En dos modelos (“CVB+CCA < 55” y “CVB+CCA Todas las edades”), se observaron cuatro SNPs (rs1876110, rs7006918, rs1525655 y rs2437817) con puntuaciones LOD que sugieren ligamiento (**Tabla 15**). Al inspeccionar visualmente los resultados del test de ligamiento paramétrico en los individuos con CVB, se observó un efecto mínimo de la edad sobre las estimaciones de la puntuación LOD. Por otro lado, cuando se consideran como afectados los individuos con CVB y los colecistectomizados, se observa un claro efecto de la edad con las máximas puntuaciones LOD observadas en el grupo de pacientes diagnosticados/colecistectomizados con menos de 55 años y en el grupo que incluye pacientes afectados a cualquier edad.

**Tabla 15.** Puntuaciones LOD sugerentes de ligamiento en cuatro cromosomas obtenidas mediante análisis de ligamiento paramétrico en dos modelos con restricción de edad en 7 familias chilenas con múltiples casos de CVB.

CVB: Cáncer de Vesícula Biliar, CCA: Colecistectomía, Edades: 55 años, Todas las edades (TE). “CVB+CCA<55”: los individuos se consideraron como afectados si fueron diagnosticados con CVB o colecistectomizados antes de cumplir los 55 años. “CVB+CCA TE”: los individuos se consideraron como afectados si fueron diagnosticados con CVB o colecistectomizados a cualquier edad. Nivel de significancia para el genoma completo y umbral sugerente de ligamiento para el análisis de ligamiento paramétrico:  $LOD \geq 3,3$ ;  $LOD \geq 1,86$ , respectivamente.

Modelo	Cromosoma	SNP Id	Max LOD	Gen
CVB+CCA < 55	2	rs1876110	2,14	LOC105373700
	8	rs7006918	1,97	LINC00535
CVB+CCA TE	7	rs1525655	1,81	-
	9	rs2437817	1,98	ABCA1



**Figura 13.** Efecto de la restricción por edad e inclusión de individuos colecistectomizados, sobre las máximas puntuaciones LOD por cromosoma obtenidas a través de análisis de ligamiento paramétrico.

Se estudiaron 12 modelos diferentes combinando 6 restricciones de los casos por edad de diagnóstico o cirugía así: <71, <70, <65, <60, <55 años, y todos los individuos independientemente de su edad, con dos definiciones de los casos de acuerdo al fenotipo: CVB y CVB+CCA. En la figura se muestran los resultados de los 12 modelos. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $LOD \geq 3,3$  y el umbral sugerente  $LOD \geq 1,86$  de ligamiento, respectivamente. CVB: Cáncer de Vesícula Biliar, CCA: Colecistectomía.

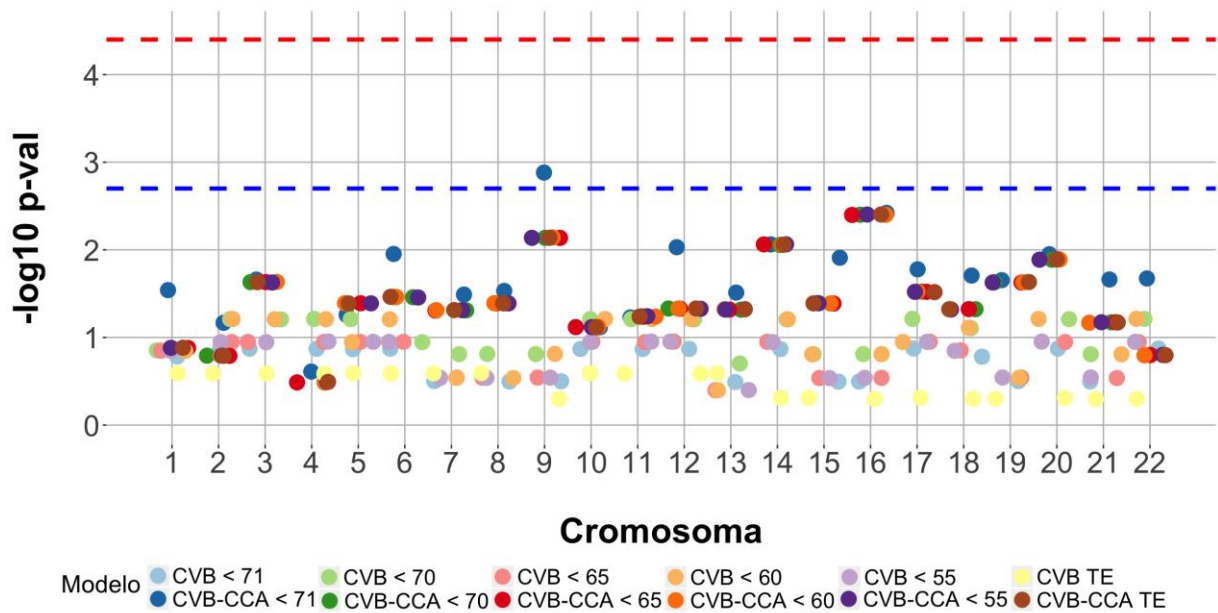
Por otro lado, en los resultados del análisis de ligamiento no paramétrico que se resumen en la **Tabla 16** se observaron las mismas 2 regiones cromosómicas candidatas de ligamiento previamente identificadas (Cromosomas 9 y 16). Mientras que el efecto de la edad sobre la estimación de los valores  $p$  siguió un patrón similar al descrito anteriormente, con pocas variaciones al analizar tanto a los pacientes con CVB como a los individuos con CVB más los colecistectomizados. No obstante, en este último grupo es bastante llamativo que el máximo valor  $p$  ( $-\log_{10}$ ) para el modelo que incluye a los individuos de todas las edades, siempre se aleja de los valores  $p$  ( $-\log_{10}$ ) para los modelos con restricción por edad, los cuales presentan valores robustos al límite de edad (**Figura 14**).



**Tabla 16.** Máximos valores  $p$  ( $-\log_{10}$ ) por cromosoma obtenidas en 12 modelos definidos para el análisis de sensibilidad del ligamiento genético.

En la tabla se muestran los máximos valores  $p$  ( $-\log_{10}$ ) por cromosoma de un análisis de ligamiento no paramétrico de 7 familias chilenas con múltiples casos de CVB. Se consideraron 12 modelos diferentes con dos definiciones de los individuos afectados conforme al fenotipo (CVB o CVB + CCA) y con 6 restricciones de los mismos conforme a la edad de diagnóstico o cirugía. CVB: Cáncer de Vesícula Biliar, CCA: Colectomía, Edades: Todas las edades (TE), 71, 70, 65, 60 y 55 años. El nivel de significancia para el genoma completo y el umbral sugerente de ligamiento para el análisis no paramétrico (ALNP) es de  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ), respectivamente. En negrita se resaltan las puntuaciones con valores más altos.

Cr	CVB						CVB+CCA					
	TE	< 71	< 70	< 65	< 60	< 55	TE	< 71	< 70	< 65	< 60	< 55
<b>1</b>	0,78	0,85	0,85	0,85	0,85	0,59	1,54	0,88	0,88	0,88	0,88	0,88
<b>2</b>	0,87	1,21	0,95	1,21	0,95	0,59	1,17	0,79	0,79	0,79	0,79	0,79
<b>3</b>	0,87	1,21	0,95	1,21	0,95	0,59	1,66	1,63	1,63	1,63	1,63	1,63
<b>4</b>	0,87	1,21	0,95	1,21	0,95	0,59	0,61	0,49	0,49	0,49	0,49	0,49
<b>5</b>	0,87	1,21	0,95	0,95	0,95	0,59	1,26	1,39	1,39	1,39	1,39	1,39
<b>6</b>	0,87	0,95	0,95	1,21	0,95	0,59	1,95	1,46	1,46	1,46	1,46	1,46
<b>7</b>	0,50	0,81	0,54	0,54	0,54	0,59	1,49	1,31	1,31	1,31	1,31	1,31
<b>8</b>	0,50	0,81	0,54	0,54	0,54	0,59	1,53	1,39	1,39	1,39	1,39	1,39
<b>9</b>	0,50	0,81	0,54	0,81	0,54	0,30	<b>2,88</b>	2,14	2,14	2,14	2,14	2,14
<b>10</b>	0,87	1,21	0,95	1,21	0,95	0,59	1,12	1,12	1,12	1,12	1,12	1,12
<b>11</b>	0,87	1,21	0,95	1,21	0,95	0,59	1,23	1,24	1,24	1,24	1,24	1,24
<b>12</b>	0,87	1,21	0,95	1,21	0,95	0,59	2,03	1,33	1,33	1,33	1,33	1,33
<b>13</b>	0,49	0,70	0,40	0,40	0,40	0,59	1,51	1,32	1,32	1,32	1,32	1,32
<b>14</b>	0,87	1,21	0,95	1,21	0,95	0,31	2,06	2,06	2,06	2,06	2,06	2,06
<b>15</b>	0,50	0,81	0,54	0,81	0,54	0,31	1,91	1,39	1,39	1,39	1,39	1,39
<b>16</b>	0,50	0,81	0,54	0,81	0,54	0,30	<b>2,42</b>	2,40	2,40	2,40	2,40	2,40
<b>17</b>	0,87	1,21	0,95	0,95	0,95	0,31	1,78	1,52	1,52	1,52	1,52	1,52
<b>18</b>	0,78	1,11	0,85	1,11	0,85	0,30	1,71	1,32	1,32	1,32	1,32	1,32
<b>19</b>	0,50	0,54	0,54	0,54	0,54	0,30	1,65	1,63	1,63	1,63	1,63	1,63
<b>20</b>	0,87	1,21	0,95	1,21	0,95	0,31	1,95	1,89	1,89	1,89	1,89	1,89
<b>21</b>	0,50	0,81	0,54	0,81	0,54	0,30	1,66	1,17	1,17	1,17	1,17	1,17
<b>22</b>	0,87	1,21	0,95	1,21	0,95	0,30	1,67	0,80	0,80	0,80	0,80	0,80



**Figura 14.** Efecto de la restricción por edad e inclusión de individuos colecistectomizados, sobre los máximos valores  $p$  ( $-\log_{10}$ ) por cromosoma obtenidos a través de análisis de ligamiento no paramétrico.

Se estudiaron 12 modelos diferentes combinando 6 restricciones de los casos por edad de diagnóstico o cirugía así: <71, <70, <65, <60, <55 años, y todos los individuos independiente de su edad con dos definiciones de los casos de acuerdo al fenotipo: CVB y CVB+CCA. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y el umbral sugerente  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ) de ligamiento, respectivamente. CVB: Cáncer de Vesícula Biliar, CCA: Colecistectomía.

### 3.3.2. Contribuciones de cada familia a los resultados del análisis de ligamiento.

Las contribuciones individuales y acumuladas de las familias a las señales de ligamiento no paramétrico se muestran en las **Tablas 17 y 18**. Las familias 1, 3, 7, 6 y 4 arrojaron los mayores aportes individuales a la señal del cromosoma 9 ( $-\log_{10}(p \text{ total})=2,88$ ) con valores de 1,14; 1,08; 0,89; 0,87 y 0,71 respectivamente, mientras que las familias 2 y 5 contribuyeron 0,3 cada una. Por otro lado, en el cromosoma 16 ( $-\log_{10}(p \text{ total}) = 2,42$ ), se observaron los aportes más importantes para las familias 3, 4 y 1 con señales de ligamiento de 1,79; 1,20 y 0,99 respectivamente. Las familias restantes aportaron 1,11 en conjunto.

**Tabla 17.** Contribución individual de cada genealogía a las señales de ligamiento no paramétrico en 7 familias chilenas con múltiples casos de CVB.

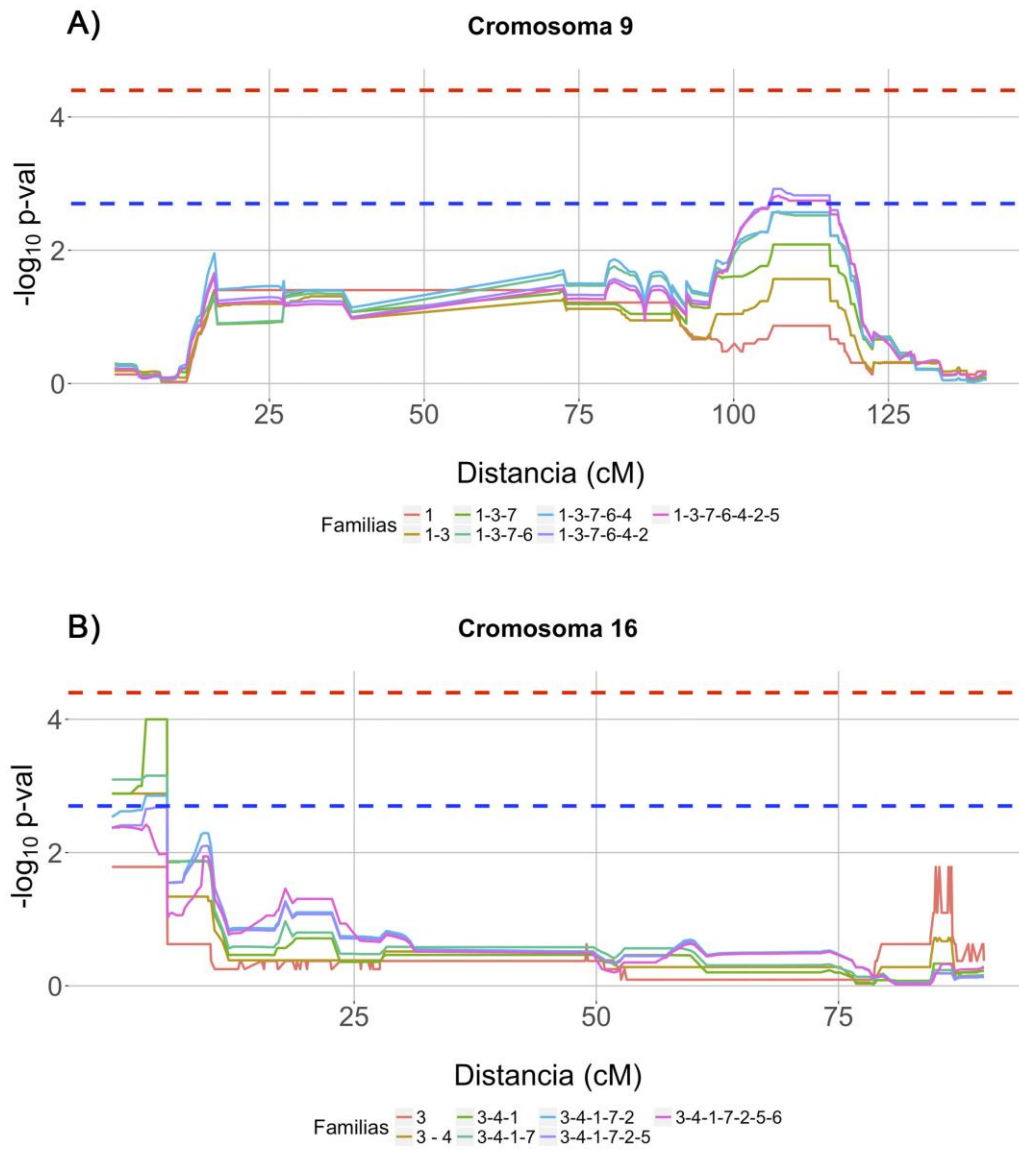
Familia	$-\log_{10}(p)$	
	Cr9	Cr16
1	1,14	0,99
2	0,30	0,30
3	1,08	1,79
4	0,71	1,20
5	0,30	0,13
6	0,87	0,05
7	0,89	0,63
1-7	2,88	2,42

**Tabla 18.** Contribuciones individuales y acumuladas de las familias a las señales de ligamiento no paramétrico en los cromosomas 9 y 16.

Familias	$p$	$-\log_{10}(p)$
<b>Cr9</b>		
1	0,0715	1,14
1-3	0,0270	1,57
1-3-7	0,0082	2,09
1-3-7-6	0,0027	2,57
1-3-7-6-4	0,0026	2,59
1-3-7-6-4-2	0,0012	2,92
1-3-7-6-4-2-5	0,0015	2,88
<b>Cr16</b>		
3	0,0164	1,79
3-4	0,0013	2,89
3-4-1	<0,0001	>4,00
3-4-1-7	0,0007	3,15
3-4-1-7-2	0,0014	2,85
3-4-1-7-2-5	0,0021	2,68
3-4-1-7-2-5-6	0,0038	2,42

Tal como se mencionó en el apartado de resultados de ligamiento no paramétrico, con las 7 familias disponibles para el análisis se identificaron dos regiones; una sugerente en el cromosoma 9 ( $-\log_{10}(p)=2,88$ ) y otra con niveles cercanos al umbral sugerente de ligamiento en el cromosoma 16 ( $-\log_{10}(p)=2,42$ ). Al analizar las contribuciones acumuladas de las familias a la señal de ligamiento del cromosoma 9 (**Tabla 18**), se observó que en general cada familia hizo una contribución a favor del ligamiento sobre la región indicada, en donde la inclusión de cada nueva familia (salvo la numero 5) en el análisis resultó en un aumento de la señal global. La máxima señal obtenida fue  $-\log_{10}(p)=2,92$  y se alcanzó al incluir 6 familias (**Figura 15A**). Los resultados en el cromosoma 16, por otro lado, identificaron algunas familias con datos a favor y otras con datos en contra de la evidencia de ligamiento. Los análisis incluyendo las 7 familias arrojaron una señal cercana al nivel sugerente de ligamiento ( $-\log_{10}(p)=2,42$ ). Al incluir únicamente las familias 3, 4 y 1, la señal aumentó a  $-\log_{10}(p)>4,0$  (**Figura 15B**).

Al aplicar simulación de Monte Carlo para calcular los valores  $p$ , en el análisis de ligamiento no paramétrico entre los SNPs rs10760688 y rs10739300 del cromosoma 9, la señal basada en las 1, 3, 7, 6, 4 y 2 disminuyó a  $-\log_{10}(p)=2,38$ . Mientras que entre los marcadores rs2562188 y rs9934064 del cromosómo 16, la señal basada en las familias 3, 4 y 1 disminuyó a  $-\log_{10}(p)=3,1$ ; estando localizada en el marcador rs1659495.



**Figura 15.** Contribuciones acumuladas de las familias a las señales de ligamiento.

Se presentan los valores  $p$  ( $-\log_{10}$ ) obtenidos a través de un análisis de ligamiento no paramétrico de 7 familias chilenas con múltiples casos de Cáncer de Vesícula Biliar, así como individuos colecistectomizados. Se estudiaron 7 grupos diferentes combinando las familias en orden secuencial descendente de acuerdo a su contribución individual a la señal de ligamiento. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 2,2e^{-5}$  ( $-\log_{10}(p) > 4,4$ ) y el umbral sugerente  $p < 7,4e^{-4}$  ( $-\log_{10}(p) > 2,7$ ) de ligamiento. **(A)** Se observa la contribución de cada familia en favor del ligamiento sobre el cromosoma 9, en donde la inclusión de una nueva familia en el análisis conduce al incremento en la señal. La máxima señal obtenida fue  $-\log_{10}(p) = 2,92$  y se alcanzó al incluir 6 familias. **(B)** Los resultados en el cromosoma 16 muestran la presencia de algunas familias que contribuyen en favor y otras en contra del ligamiento. El análisis inicial con las 7 familias arrojó una señal sugerente de ligamiento ( $-\log_{10}(p) = 2,42$ ), mientras que al incluir únicamente las familias 3, 4 y 1 se obtuvo una señal de  $-\log_{10}(p) > 4,0$ .

### 3.3.3. Variación de los parámetros en el ligamiento paramétrico

Los análisis de sensibilidad para el ligamiento paramétrico en los cromosomas 7, 9 y 16 no mostraron cambios significativos al variar los valores de la penetrancia (70%, 80%, 90%, 95%, 99%), tasa de fenocopias (30%, 20%, 10%, 5%, 1%) y frecuencia del alelo de susceptibilidad (0,0001; 0,001; 0,01; 0,05). Los gráficos correspondientes se muestran en las Figuras Suplementarias 3-5.

## 4. Búsqueda de Genes Candidatos mediante Anotación Funcional

### 4.1. Análisis de Clasificación y Enriquecimiento Funcional

En las dos regiones candidatas se identificaron 128 genes: 95 del cromosoma 9 (**Tabla 19**) y 33 del cromosoma 16 (**Tabla 20**). La clasificación funcional por vías de señalización celular realizada en PANTHER™ [104], permitió establecer que los 128 genes estaban relacionados con 58 vías, siendo la vía de señalización WNT (P00057) y las vías de señalización de proteína G heterotrimérica mediadas por Gi alfa y Gs alfa (P00026), las que incluían un mayor número de genes (4 cada una). A fin de refinar y esclarecer los resultados, se realizó un análisis de enriquecimiento funcional en DAVID 6.8 [105, 106].

**Tabla 19.** Lista de 95 genes contenidos en la región de ligamiento chr9:105016960-115983216 identificada al analizar 7 familias chilenas con múltiples casos de CVB y cálculos biliares.

ABCA1	DEC1	LPAR1	PPP3R2	TMEM246
ACTL7A	DFNB31	MRPL50	PRPF4	TMEM38B
ACTL7B	DNAJC25	MUSK	PTBP3	TNC
AKAP2	DNAJC25-GNG10	NIPSNAP3A	PTGR1	TNFSF15
AKNA	EPB41L4B	NIPSNAP3B	PTPN3	TNFSF8
ALAD	FAM206A	OR13C2	RAD23B	TRIM32
ALDOB	FKBP15	OR13C3	RGS3	TXN
AMBP	FKTN	OR13C4	RNF183	TXNDC8
ASTN2	FRRS1L	OR13C5	RNF20	UGCG
ATP6V1G1	FSD1L	OR13C8	SLC31A1	WDR31
BAAT	GNG10	OR13C9	SLC31A2	ZFP37
BSPRY	GRIN3A	OR13D1	SLC44A1	ZNF189
C9orf152	HDHD3	OR13F1	SLC46A2	ZNF462
C9orf43	HSDL2	OR2K2	SMC2	ZNF483
C9orf84	IKBKAP	ORM1	SNX30	ZNF618
C9orf91	INIP	ORM2	SUSD1	
CDC26	KIAA0368	PALM2	SVEP1	
COL27A1	KIAA1958	PALM2-AKAP2	TAL2	
CTNNAL1	KIF12	PAPPA	TLR4	
CYLC2	KLF4	POLE3	TMEM245	

**Tabla 20.** Lista de 33 genes contenidos en la región de ligamiento chr16:105444-5995388 identificada al analizar 7 familias chilenas con múltiples casos de CVB y cálculos biliares.

A2BP1	CDIP1	GLYR1	PPL	TRAP1
ADCY9	CORO7	HMOX2	ROGDI	UBALD1
ALG1	CORO7-PAM16	MGRN1	SEC14L5	UBN1
ANKS3	CREBBP	NAGPA	SEPT12	VASN
C16orf71	DNAJA3	NMRAL1	SMIM22	ZNF500
C16orf89	EEF2KMT	NUDT16L1	SRL	
C16orf96	GLIS2	PAM16	TFAP4	

Del total de genes, 34 fueron clasificados funcionalmente, de acuerdo al proceso biológico relacionado, en cuatro grupos. La mayoría de genes en el grupo 1 corresponde a proteínas receptoras olfatorias y pertenecen a la familia 1 de receptores acoplados a proteína G. En el grupo 2 se ubicaron mayoritariamente genes de proteínas transmembrana. En el grupo 3 estaban proteínas que contenían repeticiones WD. Y finalmente, las proteínas del grupo 4, probablemente involucradas en la regulación transcripcional, pertenecen a la familia de proteínas con dedos de zinc. No obstante, 95 genes quedaron por fuera de la clasificación.

## **4.2. Análisis de las Categorías de Anotación de DAVID Individualmente**

### **4.2.1. Anotación por enfermedades**

Se identificaron 22 genes relacionados con enfermedades humanas, con un componente genético conocido (registradas en la base de datos Online Mendelian Inheritance in Man (OMIM)) (**Tabla 21**). De estos genes, dos están relacionados con el metabolismo de lípidos y ácidos biliares (*ABCA1*, *BAAT* o ácido biliar-CoA: aminoácido N-aciltransferasa) y otros tres están asociados a cáncer (*TAL2* (Factor de transcripción TAL BHLH 2), *DEC1* (Gen deletado en cáncer esofágico 1), *TLR4* (Receptor tipo Toll 4)). De forma similar, la anotación de los genes de acuerdo a enfermedades registradas en la Base de datos de Asociación Genética (GAD), permitió identificar 25 genes relacionados con 11 fenotipos/enfermedades considerados como factores de riesgo de CVB y cálculos biliares (**Tabla 22**).

**Tabla 21.** Lista de genes contenidos en las regiones de ligamiento identificadas y que están relacionados con enfermedades humanas registradas en la base de datos OMIM.

Se resaltan los genes directamente relacionados con el metabolismo de lípidos y con cáncer en negrita y azul respectivamente.

<b>Gen</b>	<b>Enfermedad</b>
Aldolase, fructose-bisphosphate B (ALDOB)	Intolerancia a la fructosa
ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase (ALG1)	Trastorno congénito de la glicosilación tipo Ik
Aminolevulinate dehydratase (ALAD)	Susceptibilidad al envenenamiento por plomo. Porfiria hepática aguda.
<b>ATP binding cassette subfamily A member 1 (ABCA1)</b>	Enfermedad de la arteria coronaria en la hipercolesterolemia familiar, protección contra. Enfermedad de Tángen. Deficiencia de HDL tipo 2
<b>Bile acid-CoA:amino acid N-acyltransferase (BAAT)</b>	Hipercolanemia familiar
Collagen type XXVII alpha 1 chain (COL27A1)	Síndrome de Steel
CREB binding protein (CREBBP)	Síndrome de Rubinstein-Taybi
<b>Deleted in esophageal cancer 1 (DEC1)</b>	Carcinoma de células escamosas esofágico
Fukutin (FKTN)	Distrofia muscular-Distroglicanopatía tipo A, tipo C, tipo B
GLIS family zinc finger 2 (GLIS2)	Nefronoptosis
Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein (IKBKAP)	Disautonomía familiar
Muscle associated receptor tyrosine kinase (MUSK)	Secuencia de deformación de aquinesia fetal, síndrome miasténico, congénito, asociado con deficiencia del receptor de acetilcolina
Pre-mRNA processing factor 4 (PRPF4)	Retinitis pigmentosa
Presequence translocase associated motor 16 homolog (PAM16)	Displasia espondilometafisaria tipo Megarbane-Dagher-Melike
Rogdi homolog (ROGDI)	Síndrome de Kohlschütter-Tonz
Septin 12 (SEPT12)	Fallo espermatogénico
<b>TAL bHLH transcription factor 2 (TAL2)</b>	Leucemia, células T linfocítica aguda, somática
Tenascin C (TNC)	Sordera, autosómica dominante
<b>Toll like receptor 4 (TLR4)</b>	Cáncer colorrectal, susceptibilidad a. Degeneración macular relacionada con la edad, hiporreactividad a endotoxinas.
Transmembrane protein 38B (TMEM38B)	Osteogénesis imperfecta tipo XIV
Tripartite motif containing 32 (TRIM32)	Distrofia muscular, cintura-cadera, tipo 2H, síndrome de Bardet-Biedl
Whirlin (WHRN)	Sordera, autosómica recesiva, síndrome de Usher, tipo 2D



**Tabla 22.** Genes anotados en GAD de acuerdo a 11 fenotipos/enfermedades asociados con CVB y cálculos biliares.

<b>Fenotipo</b>	<b>Gen</b>
Aumento de peso	Glutamate ionotropic receptor NMDA type subunit 3A (GRIN3A)
	Protein phosphatase 3 regulatory subunit B, beta (PPP3R2)
Peso corporal	Chromosome 9 open reading frame 84 (c9orf84)
	Cylicin 2 (CYLC2)
	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein (IKBKAP)
BMI	Structural maintenance of chromosomes 2 (SMC2)
	RAD23 homolog B, nucleotide excision repair protein (RAD23B)
	Adenylate cyclase 9 (ADCY9)
	Alpha-1-microglobulin/bikunin precursor (AMBP)
	Astrotactin 2 (ASTN2)
	Protein phosphatase 3 regulatory subunit B, beta (PPP3R2)
	Structural maintenance of chromosomes 2 (SMC2)
Hipertensión	Toll like receptor 4 (TLR4)
	ATP binding cassette subfamily A member 1 (ABCA1)
	Adenylate cyclase 9 (ADCY9)
	Aminolevulinatase dehydratase (ALAD)
	Astrotactin 2 (ASTN2)
	Heme oxygenase 2 (HMOX2)
	Whirlin (WHRN)
Presión sanguínea	Zinc finger protein 618 (ZNF618)
	Toll like receptor 4 (TLR4)
	Tumor necrosis factor superfamily member 8 (TNFSF8)
Colesterol	Zinc finger protein 618 (ZNF618)
	ATP binding cassette subfamily A member 1 (ABCA1)
	KIAA0368
	Astrotactin 2 (ASTN2)
Niveles plasmáticos de colesterol HDL (HDL-C)	Structural maintenance of chromosomes 2 (SMC2)
	ATP binding cassette subfamily A member 1 (ABCA1)
	CREB binding protein (CREBBP)
Lipoproteínas HDL	ATP binding cassette subfamily A member 1 (ABCA1)
	Collagen type XXVII alpha 1 chain (COL27A1)
	Cyclin 2 (CYLC2)
Colesterol LDL	ATP binding cassette subfamily A member 1 (ABCA1)
	Astrotactin 2 (ASTN2)
	Muscle associated receptor tyrosine kinase (MUSK)
	Protein phosphatase 3 regulatory subunit B, beta (PPP3R2)
Triglicéridos	Toll like receptor 4 (TLR4)
	ATP binding cassette subfamily A member 1 (ABCA1)
	Collagen type XXVII alpha 1 chain (COL27A1)
	Muscle associated receptor tyrosine kinase (MUSK)
	Toll like receptor 4 (TLR4)
DMT2	Tumor necrosis factor superfamily member 15 (TNFSF15)
	ATP binding cassette subfamily A member 1 (ABCA1)
	Kruppel like factor 4 (KLF4)
	Aldolase, fructose-bisphosphate B (ALDOB)
Colangiocarcinoma	Toll like receptor 4 (TLR4)
	Protein Tyrosine Phosphatase, Non-Receptor Type 3 (PTPN3)

#### 4.2.2. Anotación en Categorías funcionales basadas en palabras clave

Fueron identificados 15 genes por su clasificación mediante la búsqueda de la palabra “lipid” (**Tabla 23**), de los cuales 7 están directamente relacionados con el metabolismo de lípidos, mientras que los 8 restantes se identificaron debido a que contienen una región de unión a residuos lipídicos. Al realizar una búsqueda usando las palabras “cancer”, “tumor” o “neoplasm” dentro de esta categoría, se identificaron los genes *DEC1*, *TNFSF8* y *TNFSF15*.

**Tabla 23.** Genes anotados en categorías funcionales basadas en palabras claves.

En negrita se resaltan los genes directamente relacionados con el metabolismo de lípidos.

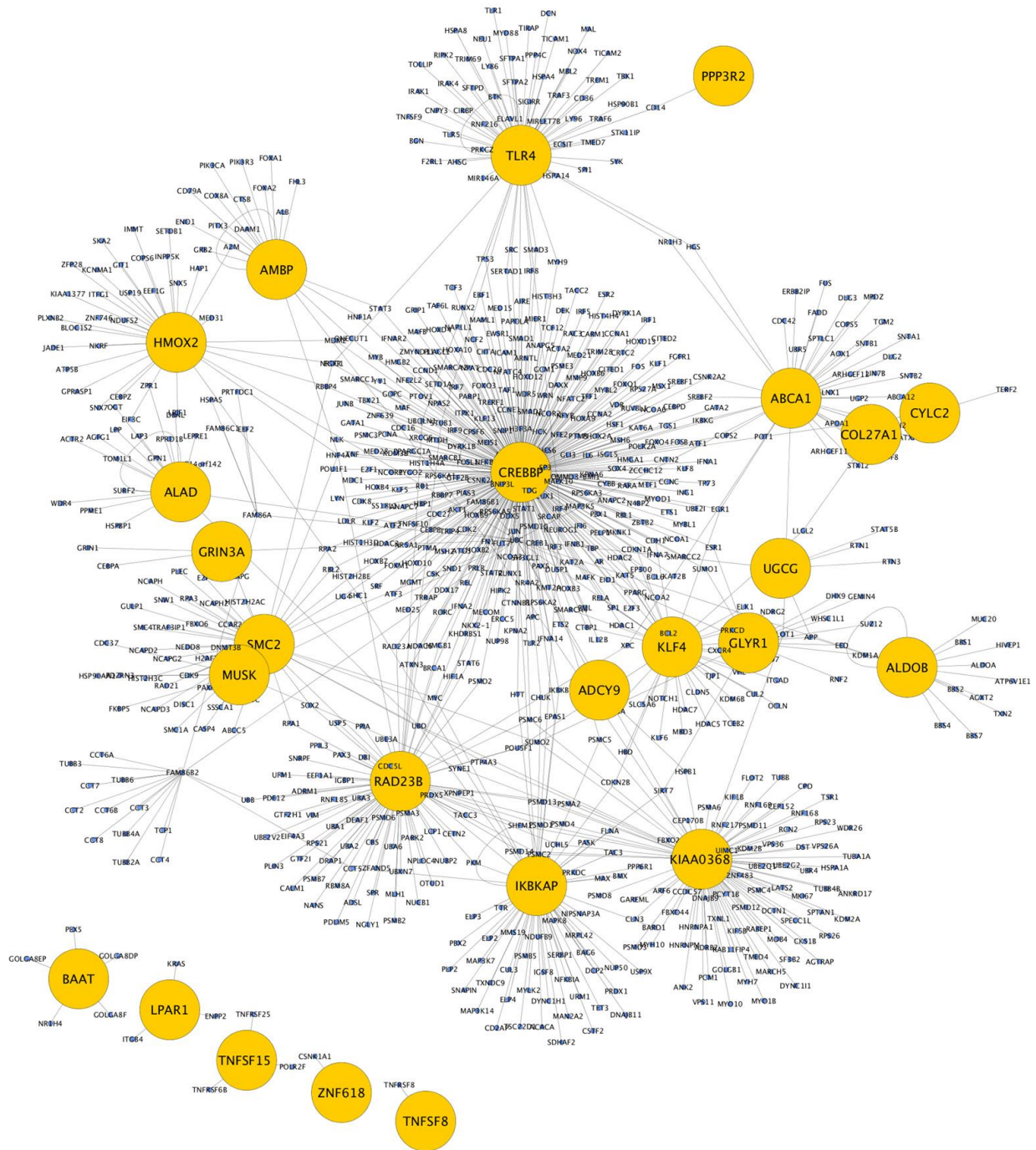
Gen	Relación funcional
A-kinase anchoring protein 2 (AKAP2)	Lipid moiety-binding region
<b>ATP binding cassette subfamily A member 1 (ABCA1)</b>	Nuclear Receptors in Lipid Metabolism and Toxicity
<b>CREB binding protein(CREBBP)</b>	Cellular lipid metabolic process
DNAJC25-GNG10 readthrough (DNAJC25-GNG10)	Lipid moiety-binding region
DnaJ heat shock protein family (Hsp40) member C25(DNAJC25)	Lipid moiety-binding region
G protein subunit gamma 10 (GNG10)	Lipid moiety-binding region
PALM2-AKAP2 readthrough (PALM2-AKAP2)	Lipid moiety-binding region
TNF receptor associated protein 1 (TRAP1)	Lipid particle
<b>UDP-glucose ceramide glucosyltransferase (UGCG)</b>	Glycosphingolipid biosynthetic process
<b>Bile acid-CoA:amino acid N-acyltransferase (BAAT)</b>	Fatty acid metabolism, Lipid metabolism
<b>Glyoxylate reductase 1 homolog (GLYR1)</b>	Lipid metabolism
<b>Lysophosphatidic acid receptor 1 (LPAR1)</b>	Phospholipid binding
Paralemmin 2 (PALM2)	Lipid moiety-binding region
Protein phosphatase 3 regulatory subunit B, beta (PPP3R2)	Lipid moiety-binding region
<b>SEC14 like lipid binding 5 (SEC14L5)</b>	Regulate both lipid metabolism and vesicular transport

Mediante un análisis integrativo, estos resultados (anotación funcional por procesos biológicos, por enfermedades (OMIM y GAD), por palabras clave y por vías celulares), permitieron una reducción del total, a una lista de 30 genes (**Tabla 24**). De estos genes, *ABCA1* fue el gen que más se repitió, con 9 apariciones dentro de las categorías de anotación, seguido por *TLR4* y *PPP3R2* con 6 y 4 apariciones respectivamente. Las variantes en estos 30 genes podrían ser relevantes en la etiopatogenia del CVB y cálculos biliares y, por ende, ser responsables de las señales de ligamiento identificadas en las 7 familias chilenas analizadas. Mediante la generación de un gráfico de redes de interacción molecular, se pudo observar

que 20 de estos genes están interconectados, siendo el gen de la proteína de unión a CREB (*CREBBP*) la molécula que, por sus funciones, tiene mayor cantidad de vínculos moleculares (Figura 16).

**Tabla 24.** Genes potencialmente relevantes dentro de la etiopatogenia del CVB y cálculos biliares.

Gen	Nombre
ABCA1	ATP binding cassette subfamily A member 1
ADCY9	Adenylate cyclase 9
ALAD	Aminolevulinatase dehidratasa
ALDOB	Aldolasa, fructosa-bisfosfato B
AMBP	Alfa-1-microglobulina/bikunina precursora
ASTN2	Astrotactina 2
BAAT	Aciltransferasa de ácido biliar-CoA:ácido amino N-acil
COL27A1	Cadena alfa 1 del tipo XXVII de colágeno
CREBBP	Proteína de unión a CREB
CYLC2	Ciclina 2
DEC1	1 Deleted In Esophageal Cancer
GLYR1	Homólogo 1 de reductasa de glicolato
GRIN3A	Subunidad 3A del receptor ionotrópico de glutamato tipo NMDA
HMOX2	2 Oxigenasa de hemo
IKBKAP	Proteína asociada al complejo de potenciación del gen de la proteína ligandera kappa en células B, inhibidor
KIAA0368	KIAA0368
KLF4	Factor 4 como Kruppel
LPAR1	1 Receptor de ácido fosfatídico lisofosfatídico
MUSK	Quinasa de tirosina asociada al músculo receptor
PPP3R2	Beta subunidad reguladora 3 de fosfatasa de proteínas
PTPN3	3 Fosfatasa de tirosina, tipo no receptor
RAD23B	Proteína de reparación de excisión de nucleótidos B, homólogo de RAD23
SEC14L5	5 Unión a lípidos como SEC14
SMC2	2 Mantenimiento estructural de los cromosomas
TLR4	4 Receptor como Toll
TNFSF8	8 Miembro de la superfamilia de factores de necrosis tumoral
TNFSF15	15 Miembro de la superfamilia de factores de necrosis tumoral
UGCG	Transferasa de glucosilación de ceramida de glucosa-UDP
WHRN	Whirlin
ZNF618	618 Proteína de dedo de zinc



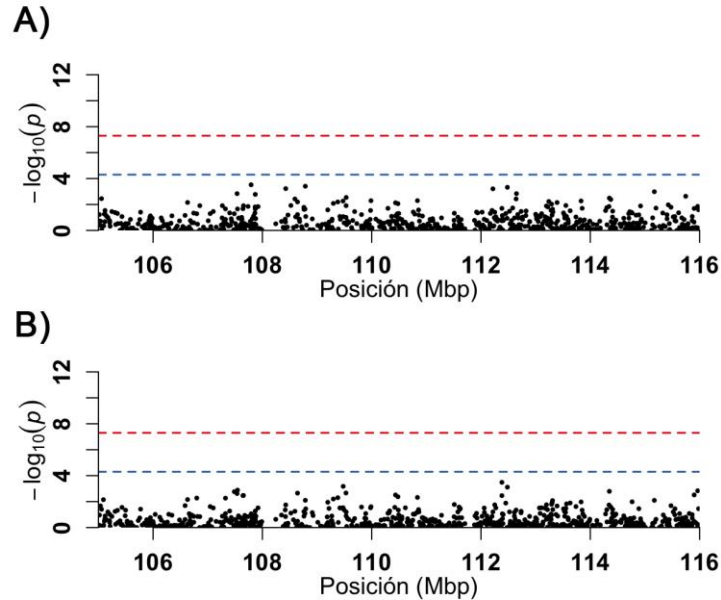
**Figura 16.** Redes de interacción molecular para 25 genes que podrían albergar variantes de riesgo para el desarrollo de CVB.

Los genes que se muestran fueron identificados en las regiones de ligamiento para el CVB y cálculos biliares Chr9:105016960-115983216 y chr16:105444-5995388 en 7 familias chilenas con múltiples casos.

## 5. Mapeo Fino por Asociación Genética

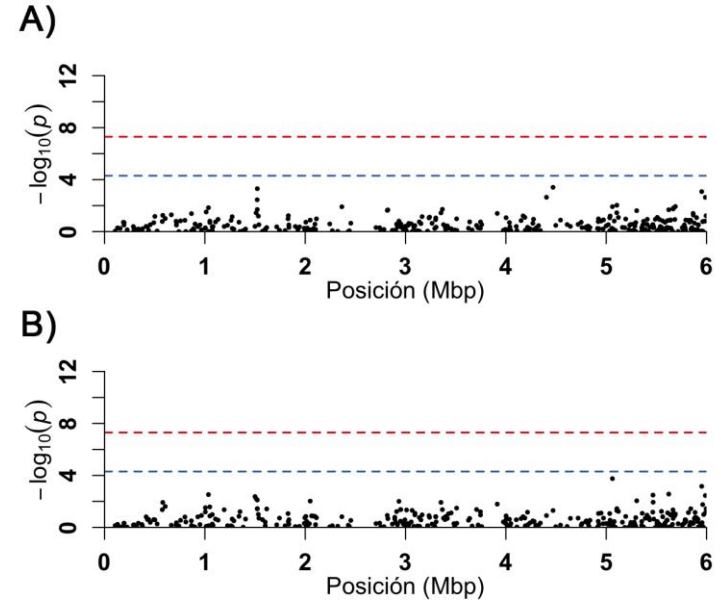
Finalmente, se evaluó si las variantes contenidas en las regiones candidatas identificadas mediante los análisis de ligamiento en familias, se asocian con un mayor riesgo de desarrollar CVB en la población chilena. Para esto, se realizó un mapeo fino por asociación genética.

Tras realizar el control de calidad (sección III-4.3), 1.028 marcadores en las regiones chr9:105016960-115983216 y chr16:105444-5995388 fueron utilizados para el mapeo fino por asociación. Al analizar las regiones de los cromosomas 9 y 16 bajo el modelo dominante con y sin covariables no se observaron variantes genéticas significativamente asociadas al riesgo de CVB (**Figuras 17 y 18**). Bajo un modelo genético aditivo sin covariables, el marcador rs9409373 en el cromosoma 9 presentó un  $-\log_{10}(p)=9,9$  (OR=0,35 IC del 95% [0,25-0,47]). Sin embargo, al incluir las covariables sexo, edad y proporción de ascendencia mapuche en el modelo, la asociación entre esta variante y el riesgo del CVB no alcanzó significancia estadística (**Figuras 19 y 20**). La proporción de la varianza en el fenotipo explicada en el modelo de regresión por el sexo, la edad y la ascendencia mapuche fue 0,38; 0,09 y 0,05 respectivamente ( $R^2$  de Nagelkerke).



**Figura 17.** Manhattan plot del mapeo fino por asociación sobre el cromosoma 9 suponiendo un modelo genético dominante, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.

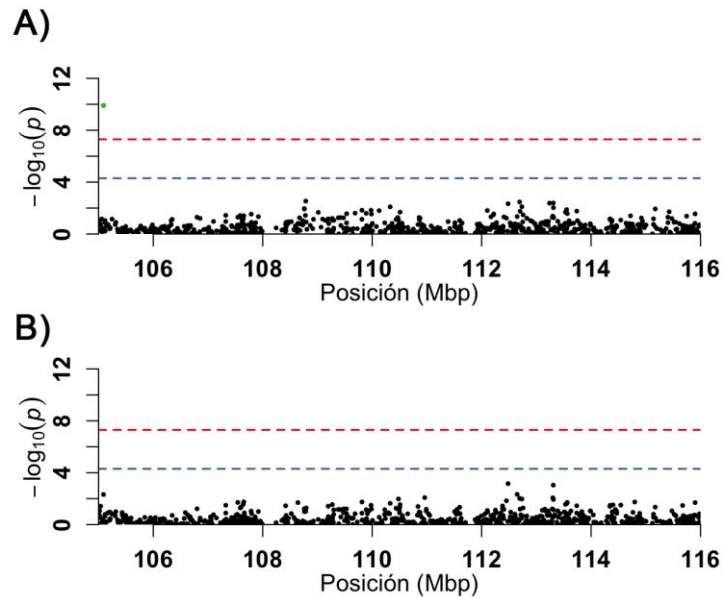
No se observaron variantes significativamente asociadas al CVB usando un modelo de regresión logística y suponiendo un modelo genético dominante **(A)** sin covariables ni **(B)** con las covariables sexo, edad y proporción de ascendencia mapuche. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ) y el umbral de significancia con corrección de Bonferroni para el dataset  $p < 4,9e^{-5}$  ( $-\log_{10}(p) > 4,3$ ), respectivamente.



**Figura 18.** Manhattan plot del mapeo fino por asociación sobre el cromosoma 16 suponiendo un modelo genético dominante, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.

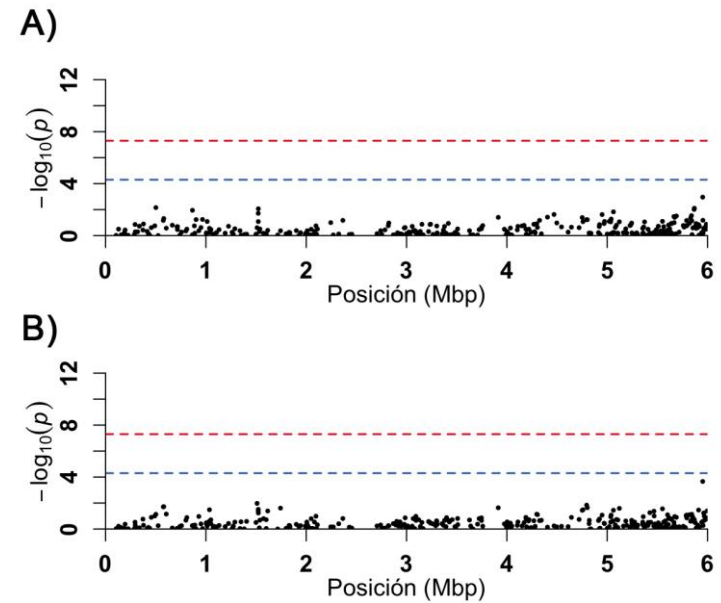
No se observaron variantes significativamente asociadas al CVB usando un modelo de regresión logística y suponiendo un modelo genético dominante **(A)** sin covariables ni **(B)** con las covariables sexo, edad y proporción de ascendencia mapuche. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ) y el umbral de significancia con corrección de Bonferroni para el dataset  $p < 4,9e^{-5}$  ( $-\log_{10}(p) > 4,3$ ), respectivamente.





**Figura 19.** Manhattan plot del mapeo fino por asociación sobre el cromosoma 9 suponiendo un modelo genético aditivo, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.

**(A)** Se destaca en verde el marcador rs9409373 que presentó un  $-\log_{10}(p)=9,9$  usando un modelo de regresión logística sin covariables y suponiendo un modelo genético aditivo. **(B)** Al incluir las covariables sexo, edad y proporción de ascendencia mapuche en el modelo de regresión, la variante rs9409373 perdió la significancia. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ) y el umbral de significancia con corrección de Bonferroni para el dataset  $p < 4,9e^{-5}$  ( $-\log_{10}(p) > 4,3$ ), respectivamente.



**Figura 20.** Manhattan plot del mapeo fino por asociación sobre el cromosoma 16 suponiendo un modelo genético aditivo, en un análisis de 183 casos con CVB y 1.962 controles de población chilena.

No se observaron variantes significativamente asociadas al CVB usando un modelo de regresión logística suponiendo un modelo genético aditivo **(A)** sin covariables ni **(B)** con las covariables sexo, edad y proporción de ascendencia mapuche. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ) y el umbral de significancia con corrección de Bonferroni para el dataset  $p < 4,9e^{-5}$  ( $-\log_{10}(p) > 4,3$ ), respectivamente.

Al ajustar los modelos de regresión logística incluyendo las cinco primeras componentes principales además del sexo y la edad como covariables, no se observaron diferencias respecto a los resultados anteriores. En las **Figuras Suplementarias 6 y 7** se muestran los ORs y los  $-\log_{10}(p)$  para las 1.028 variantes en los modelos aditivo y dominante ajustados por sexo, edad y proporción de ascendencia mapuche o sólo sexo y edad, respectivamente.

Finalmente se analizaron los genotipos de las variantes contenidas en 13 de los 30 genes candidatos identificados mediante anotación funcional (para los 17 restantes no se contaba con información genotípica). Los genes *ABCA1* y *SEC14L5* mostraron resultados significativos bajo un modelo genético dominante ajustado por las covariables sexo, edad y proporción de ascendencia mapuche, mientras que el gen *PTPN3* mostró resultados significativos en un modelo no ajustado. El marcador rs363717 del gen *ABCA1* presentó un OR=1,84 (IC del 95% 1,26-2,66). En el gen *SEC14L5*, el marcador rs1558562 presentó un OR=2,02 (IC del 95% 1,4-2,95). Mientras que en el gen *PTPN3*, el marcador rs7041037 presentó un OR=0,62 (IC del 95% 0,47-0,81) (**Tabla 25 y Figura 21**). Sin embargo, al ajustar el modelo por sexo, edad y proporción de ascendencia mapuche este marcador perdió la significancia.

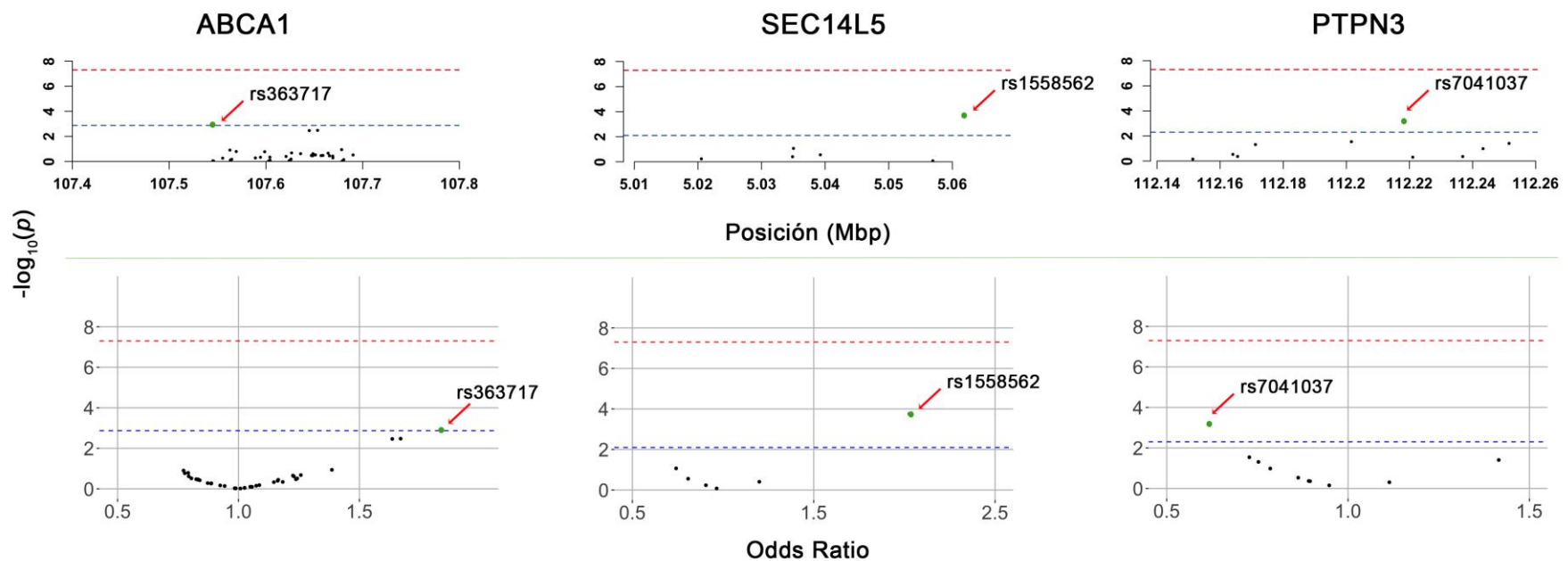


**Tabla 25.** Resumen de variantes asociadas con el riesgo de CVB y sus *odds ratios* estimados identificadas en 4 estudios de asociación.

Los primeros cinco loci fueron identificados en esta tesis mediante un análisis de regresión logística en 183 casos con CVB y 1.962 controles.

\*Homocigoto del alelo alternativo como referencia. MA: Modelo ajustado por covariables sexo, edad y proporción de ascendencia mapuche. US: Umbral de significancia con corrección de Bonferroni.

Locus	Cromosoma	SNP ID	OR*	95% IC	p	Ajuste	Estudio	Diseño/Población
<b>ABCA1</b>	9q31.1	rs363717	1.84	1.26-2.66	0.001	Sexo, edad y ascendencia MAP	Presente	CVB vs Controles / Chile
<b>SEC14L5</b>	16p13.3	rs1558562	2.02	1.4-2.95	0.0002			
<b>PTPN3</b>	9q31.3	rs7041037	0.62	0.47-0.81	0.0006	Ninguno		
-	9q31.1	rs9409373	0.35	0.25-0.48	1.24E-10	Ninguno		
<b>PALM2</b>	9q31.3	rs10283686	2.02	1.46-2.81	2.46E-05	Sexo y edad		
<b>ABCG8</b>	2p21	rs11887534	1.74	1.5-2.0	3.24E-08	Sexo, edad, IMC, DM2 y ascendencia	[145]	Colelitiasis-CVB vs controles /Chile
<b>TRAF3</b>	14q32.32	rs12882491	1,40	1.2-1.6	1.11E-07			
<b>CETP</b>	16q13	rs708272	5.04	1.43-17.8	0.01	Edad	[146]	Colelitiasis vs CVB / Chile
<b>APOB</b>	2p21.1	rs693	0.14	0.03-0.63	0.01			
<b>ABCB1</b>	7q21.12	rs17209837	1,61	1,38-1,89	2,26E-9	Sexo, edad, 5 eigenvectores y antecedentes de colelitiasis	[73]	CVB vs Controles / India
<b>ABCB4</b>	7q21.12	rs1558375	1,47	1.30-1.66	2,31E-10			
		rs4148808	1,57	1,35-1,82	2,71E-9			



**Figura 21.** Análisis de asociación en los genes ABCA1, SEC14L5 y PTPN3, en 183 casos con CVB y 1.962 controles de población chilena.

Panel superior: Los marcadores rs363717, rs1558562 y rs7041037 de los genes ABCA1, SEC14L5 y PTPN3 respectivamente, se observaron significativamente asociados al CVB. Panel inferior: Plots de los valores  $p$  para cada variante y su *odds ratio* (OR) asociado. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ) y el umbral de significancia con corrección de Bonferroni para cada gen.

A modo de resumen; se identificaron dos regiones de ligamiento candidatas (chr9:105016960-115983216 y chr16:105444-5995388) asociadas con la colelitiasis y el CVB con agregación familiar, 30 genes potencialmente relevantes dentro de la etiopatogenia de estas enfermedades y dos variantes (rs363717 del gen *ABCA1* (OR=1,84 95% IC 1,26-2,66) y rs1558562 del gen *SEC14L5* (OR=2,02 95% IC 1,4-2,95)) asociadas con el riesgo de CVB esporádico.

## V. Discusión

Los estudios epidemiológicos señalan a la colelitiasis como el principal factor de riesgo de CVB. Sin embargo, entre ambas patologías existen muchos factores de riesgo compartidos [56], y su comportamiento epidemiológico es muy similar. Las frecuencias de colelitiasis y CVB son directamente proporcionales en poblaciones específicas, como los indígenas de Chile y Norteamérica [58-61], ambas enfermedades presentan agregación familiar [82, 91] y las familias con agregación de CVB suelen presentar también múltiples casos de colelitiasis. Sumado a esto, se ha observado que ciertas variantes en genes litogénicos asociadas con el riesgo de colelitiasis también se asocian de forma independiente con el riesgo de CVB. Por ejemplo, Bustos *et al.* reportaron recientemente que los genes *ABCG8* (rs11887534;  $p=3.24 \times 10^{-8}$ , OR=1.74) y *TRAF3* (rs12882491;  $p=1.11 \times 10^{-7}$ , OR=1.40) se asocian con el riesgo de colelitiasis, así como también de CVB en chilenos [145].

Como resultado de la fuerte asociación entre ambas patologías se genera una gran dificultad al tratar de identificar factores de riesgo exclusivos para el CVB. No obstante, en el primer objetivo específico de esta tesis, se aprovechó esa característica a fin de incrementar la potencia del análisis de ligamiento. Esto se logró definiendo y estudiando dos fenotipos; en el primero de ellos se consideraron afectados sólo a los individuos con CVB, mientras que en el segundo se consideraron como afectados a los individuos con CVB y a los individuos colecistectomizados (CVB+CCA), suponiendo que éstos hubiesen desarrollado el cáncer de no haber sido colecistectomizados.

Esta propuesta está apoyada en tres observaciones (las dos primeras descritas en la literatura y mencionadas anteriormente y la última producto del análisis de genealogías en esta tesis): (1) los diversos factores de riesgo que comparten el CVB y la colelitiasis, (2) las asociaciones observadas entre genes litogénicos y el riesgo de colelitiasis y CVB, y (3) el riesgo diferencial que presenta un paciente colecistectomizado con dos familiares de primer grado diagnosticados con cáncer de vesícula biliar, respecto a un chileno de la población general con colelitiasis, (2,7 veces mayor; ver **Tabla 12**).

Este último es uno de los hallazgos más relevantes de esta tesis, pues de nuestro conocimiento, en la literatura científica chilena no existen antecedentes de la estimación de los riesgos de recurrencia empíricos ni los riesgos relativos para los individuos con antecedentes de CVB. Además, estos resultados sugieren que los familiares con antecedentes tienen seis veces mayor riesgo de CVB que individuos de la población general, y que la probabilidad de desarrollar colelitiasis para los individuos de las familias es la misma que la de la población general.

Si bien no se puede conocer con exactitud la probabilidad de que los familiares colecistectomizados desarrollen CVB, su riesgo *per se*, determinado por el  $\lambda=2,7$  hace suponer que esa probabilidad sea más cercana a 1 que a 0. Por esto, el modelo implementado sería particularmente informativo, sólo mediante el análisis de familias que presenten una alta frecuencia de colelitiasis y múltiples casos de CVB, de lo contrario se estarían estudiando aquellos casos comunes de colelitiasis que no conducen al desarrollo de CVB.

#### *Regiones candidatas de ligamiento en familias con agregación del CVB*

Mediante análisis de ligamiento no paramétrico se identificaron dos señales sugerentes de ligamiento para el fenotipo CVB+CCA, la primera con un  $-\log_{10}(p)>2,88$  entre los marcadores rs4742892 y rs7043293 del cromosoma 9, y la segunda con un  $-\log_{10}(p)=3,1$  entre los marcadores rs12709237 y rs1529917 del cromosoma 16.

Los análisis de sensibilidad en las dos regiones candidatas permitieron establecer la robustez de los resultados y dejaron en evidencia características muy importantes de las familias analizadas. En primer lugar, las contribuciones acumuladas de las familias a la señal de ligamiento del cromosoma 9 condujeron a un incremento global en la señal, mientras que en el cromosoma 16 mostraron contribuciones al ligamiento en sentidos opuestos, sugiriendo que las familias probablemente presentan una reducida heterogeneidad en el locus del cromosoma 9 y una evidente heterogeneidad en el locus del cromosoma 16.

Al excluir los casos con CVB de mayor edad, los cuales pueden haber desarrollado el CVB de forma esporádica por el efecto de la edad, y al excluir los casos con CCA de mayor

edad (individuos libres de CVB a edades avanzadas), los cuales podrían afectar negativamente las señales de ligamiento, se observó que el grupo de familiares afectados era bastante homogéneo y que la mayoría de ellos tuvieron el cáncer o la colecistectomía a edades tempranas. Esto se traduce en escasas variaciones en los resultados de ligamiento al eliminarlos del análisis.

Además, se establecieron rigurosas definiciones para asignar los fenotipos; en el análisis de ligamiento, los individuos con colelitiasis no colecistectomizados, al igual que los que reportaron ser sanos, fueron designados con el fenotipo “*unknown*”, de manera que los posibles efectos sobre los puntajes del ligamiento por errores en la especificación del fenotipo serían iguales entre ellos, pero menores a los efectos negativos que se generan al asignar un fenotipo afectado a un individuo sano. Incluir estos individuos como afectados en los análisis tampoco era conveniente porque no se tenía la certeza del diagnóstico de colelitiasis mediante pruebas imagenológicas. Por el contrario, en el análisis de sensibilidad con restricciones por edad, los individuos con colelitiasis no colecistectomizados fueron considerados como “sanos” (cuando correspondiera), generando pequeñas reducciones en las señales de ligamiento, probablemente atribuibles a los efectos descritos anteriormente.

#### *Variantes asociadas con el riesgo de CVB*

En cuanto al mapeo fino por asociación genética, los determinantes de potencia estadística son similares a los de los estudios de ligamiento; el tamaño del efecto y la frecuencia del alelo de riesgo, y el tamaño de muestra disponible son los principales. En términos generales, con el tamaño de muestra utilizado en el mapeo fino se podrían haber identificado variantes que otorguen un alto riesgo de CVB (**Tabla 5**). No obstante, el análisis de regresión logística con covariables no mostró variantes asociadas al CVB suponiendo los modelos genéticos dominante y aditivo. La ausencia de marcadores significativamente asociados con el CVB era un resultado muy probable teniendo en cuenta que las variantes de riesgo suelen mostrar efectos modestos [139]. Los estudios de asociación son especialmente útiles en la identificación de ese tipo de variantes comunes pero sus moderados efectos suelen requerir muestras de gran tamaño [128].

Si bien los resultados iniciales del mapeo fino no fueron satisfactorios, la reducción en el número de variantes, al restringir el análisis a 13 de los 30 genes candidatos que podrían ser altamente relevantes dentro de la etiopatogenia del CVB, permitió identificar la asociación de los genes *ABCA1* (rs363717; OR=1,84 95% IC 1,26-2,66) y *SEC14L5* (rs1558562; OR=2,02 95% IC 1,4-2,95) con el riesgo de CVB esporádico. Este resultado es altamente relevante, pues sugiere que variantes albergadas en las regiones candidatas identificadas en los cromosomas 9 y 16 estarían asociadas no solo con el riesgo de colestiasis y CVB con agregación familiar sino también con los casos esporádicos de este cáncer.

Considerando su función, es factible que estos genes sean biológicamente relevantes en el desarrollo del CVB y la colestiasis. La asociación con el gen *ABCA1* es particularmente relevante, teniendo en cuenta las recientes asociaciones identificadas para 3 variantes de los genes *ABCB1* y *ABCB4* con CVB en población india y una del gen *ABCG8* en chilenos, con el riesgo de CVB [73, 145] (**Tabla 25**), por lo que variantes que afecten el funcionamiento de *ABCA1* podrían jugar un papel importante en la susceptibilidad al CVB y la colestiasis en Chile.

#### *ABCA1: papel en la homeostasis de lípidos y en cáncer*

El gen *ABCA1* codifica para un exportador de colesterol unido a la membrana celular, que forma parte de un grupo de genes que controla la homeostasis del colesterol en las células hepáticas y epiteliales de la vesícula biliar. Entre ellos, *ABCB11*, *ABCB4* y el heterodímero *ABCG5/ABCG8* que transportan hacia la bilis: fosfatidilcolina, colesterol y sales biliares, respectivamente [147].

El transportador *ABCA1* hepático e intestinal contribuyen significativamente a los niveles plasmáticos de colesterol HDL [148]. Se ha demostrado que la sobreexpresión del gen *ABCA1* en macrófagos e hígado de ratones transgénicos (h*ABCA1*-Tg) altera el perfil lipídico del plasma y modula el catabolismo de HDL plasmático, así como el transporte hepático inverso de colesterol y la excreción de colesterol biliar y por lo tanto puede alterar marcadamente el riesgo litiasico [149]. Estudios en epitelio vesicular de humanos han

demostrado una expresión diferencial de los transportadores de tipo ABC entre individuos sanos y con cálculos biliares [150, 151].

La colelitiasis generalmente presenta un perfil con una mayor expresión de transportadores de salida de colesterol (ABCA1 y ABCG5/G8) y una disminución de la expresión del receptor *scavenger* clase B tipo I (SR-BI), receptor capaz de facilitar la captación selectiva del colesterol HDL [152]. Este hallazgo es sorprendente dado que, en la vesícula biliar, ABCA1 se expresa en la membrana basolateral epitelial, y por tanto el aumento de la expresión de ABCA1 podría traducirse en un aumento del flujo de salida de colesterol de las células epiteliales de la vesícula biliar como un mecanismo para prevenir la sobrecarga de colesterol celular y la toxicidad relacionada [153]. En este sentido, ciertas variantes que afecten la función de ABCA1 podrían ser responsables del deterioro del transporte de colesterol desde el interior de las células hacia las apolipoproteínas, lo que podría facilitar la precipitación del colesterol al aumentar el ISC. Esto podría conducir a un incremento en la expresión del transportador como mecanismo compensatorio.

El análisis de la información (estado mutacional y niveles de expresión) obtenida mediante NGS por el proyecto del Atlas de Genoma del Cáncer (TCGA) de todos los genes en diversos cánceres, incluidos los implicados en el metabolismo del colesterol, ha proporcionado un soporte correlativo del papel de la vía del colesterol y las consecuencias de su desregulación en el desarrollo del cáncer [154].

Chou *et al.* analizaron el estado de metilación del promotor de ABCA1 en líneas celulares de cáncer de ovario y muestras de pacientes. Sus resultados demostraron que pacientes con cáncer de ovario con mayor metilación y menor expresión de ABCA1 presentaron una supervivencia más corta [155]. De forma similar, se ha observado que la hipermetilación del promotor ABCA1 es un hallazgo común en tumores de próstata de grado intermedio a alto, mientras que no es detectable en próstatas sanas. Además, los niveles de expresión se correlacionan inversamente con el grado de Gleason, de manera que los altos niveles de colesterol intracelular contribuirían con un entorno propicio para la progresión tumoral [156].



Las mutaciones de pérdida de función o el silenciamiento del gen *ABCA1*, generan un aumento de los niveles de colesterol intracelular, que es una característica que se observa con frecuencia en células neoplásicas. Estudiando líneas celulares de cáncer de colon, Smith y Land [157] establecieron que conjuntamente con la síntesis elevada de colesterol en las células cancerosas, la deficiencia de *ABCA1* incrementa el colesterol mitocondrial, inhibe la liberación de moléculas promotoras de la muerte celular mitocondrial y facilita la supervivencia de las células, lo que sugiere que el colesterol mitocondrial elevado es esencial para el fenotipo del cáncer. Por esto se ha postulado una actividad supresora tumoral de la función de eflujo de *ABCA1*, que se ve comprometida después de la inhibición en la expresión del gen por mutaciones oncogénicas o por mutaciones de pérdida de función de la proteína específicas de cáncer.

Toda esta evidencia resalta la importancia que tiene *ABCA1* en el metabolismo celular; su participación en el metabolismo del colesterol, su posible efecto sobre la formación de cálculos y las consecuencias de estos sobre el epitelio de la vesícula biliar, así como también su aparente contribución al desarrollo y progresión de otros tipos de cáncer, sugieren que *ABCA1* podría tener un rol relevante en la susceptibilidad al CVB.

#### *SEC14L5 y otros genes a estudiar*

A diferencia de *ABCA1*, *SEC14L5* es un gen funcionalmente poco descrito en el humano. Pertenece al subgrupo de proteínas similares a *SEC14* localizadas en los cromosomas 16 y 17, y se ha asociado con el Trastorno de Estrés Posttraumático [158]. En *Saccharomyces cerevisiae* se ha establecido que *Sec14* es una proteína de transferencia de fosfatidilcolina/fosfatidilinositol y regula tanto el metabolismo lipídico como el transporte vesicular.

Si bien las variantes rs363717 y rs1558562 de los genes *ABCA1* y *SEC14L5* respectivamente, mostraron asociación con el riesgo de CVB, es necesario explorar otros genes que por sus funciones biológicas asociadas podrían influir en el desarrollo de la coleditiasis y el CVB. Especialmente bajo la idea de que variantes en múltiples genes, junto con ciertas exposiciones medioambientales, actúan en conjunto para dar lugar al CVB.

Es importante además, tener en cuenta que usualmente las variantes asociadas a patologías identificadas en los GWAS no son las causales, pero sirven para marcar una región de DL que contiene la o las variantes causales [159]. Los métodos estadísticos actuales para identificar variantes causales en loci de riesgo usan la fuerza de la señal de asociación en un marco iterativo condicional o estiman las probabilidades de que las variantes sean las causales [160]. Los estudios futuros deberán complementarse además, con ensayos funcionales de las variantes identificadas que permitan establecer el efecto biológico de las mismas y su impacto a nivel celular y fisiológico.

Los tres genes relacionados con cáncer (*DEC1*, *TAL2* y *TLR4*), identificados en el análisis de anotación funcional, también deben contemplarse en futuros estudios. Srivastava *et al.* por ejemplo, establecieron la asociación entre el polimorfismo g.14143C>T; rs4986791 del gen *TLR4* y el riesgo de CVB en mujeres y en casos con colestiasis (OR = 2,85 y 2,22 respectivamente) [161]. Sin embargo, los genotipos de estos genes para los individuos incluidos en este estudio no estuvieron disponibles.

El análisis de este tipo de genes también es importante, pues dentro de la sumatoria de susceptibilidades que conducirían al desarrollo del fenotipo final podrían estar involucradas variantes relacionadas con el desarrollo del cáncer y no solamente las relacionadas con el metabolismo de lípidos. En este sentido, la asociación del gen *DCC* (supresor tumoral; impide la proliferación y el crecimiento celular) con CVB en japoneses [72], puede considerarse un hallazgo coherente, pues la neoplasia vesicular en Japón tiene una etiología distinta, se relacionan con anomalías en los conductos pancreatobiliares y los cálculos de colesterol se observan casi con la misma frecuencia que los pigmentados [162]. En occidente, en donde la colestiasis parece tener un rol más preponderante, se han identificado variantes asociadas con el riesgo de CVB, en genes involucrados tanto en el metabolismo lipídico y/o biliar y en genes regulatorios de la proliferación muerte celular, reparación del daño de ADN y vías inflamatorias [70].

### *Limitaciones y fortalezas del estudio*

Entre las debilidades de esta tesis pueden mencionarse: (1) el número reducido de individuos afectados con CVB dentro de las familias (6 en total), (2) la ausencia de exámenes imagenológicos para confirmar el diagnóstico de coledocitis en todos los familiares, (3) los estudios de riesgo de recurrencia, al igual que otros tipos de estudio observacional, son propensos a confusión o sesgo, (4) el tamaño de muestra alcanzado de casos con CVB esporádico (183 individuos) y (5) las características del grupo control utilizado en el mapeo fino.

En cuanto al número de individuos en las familias, este es un aspecto de gran relevancia, no sólo porque determina la agregación familiar del rasgo, sino también porque es la principal fuente de información en los análisis de ligamiento genético. Aunque se estima que el 26% de los casos de CVB tienen antecedentes familiares [82], identificar y reclutar familias extensas con varios individuos diagnosticados con CVB es una tarea difícil, probablemente por características inherentes a la enfermedad, especialmente la alta letalidad.

Para abordar esta dificultad, se incluyeron en los análisis las dos definiciones de fenotipo (CVB, CVB+CCA), las cuales dejaron en evidencia el impacto del número de individuos incluidos en el análisis sobre la señal de ligamiento. Esta estrategia para incrementar el poder estadístico puede considerarse evidentemente una debilidad del estudio, sin embargo, las estimaciones del riesgo relativo en 16 familias chilenas sugieren, que entre dos individuos de la población que desarrollan coledocitis, existe una susceptibilidad casi 3 veces mayor de CVB en aquel que presente más de 2 antecedentes familiares del mismo cáncer. Adicionalmente, la confirmación diagnóstica de los individuos fue rigurosa, entre los individuos con coledocitis sólo se consideró a aquellos que hubiesen sido colecistectomizados, lo ofrece una garantía del diagnóstico en este grupo de individuos. Esto hace más precisa y específica la definición de individuo afectado.

No obstante, debe notarse la imposibilidad en el presente estudio, de estimar cual es la probabilidad de que los familiares colecistectomizados desarrollaran CVB en el futuro. Para

tratar de cuantificar el ruido generado desde diversas fuentes, se llevaron a cabo amplios análisis de sensibilidad considerando las dos definiciones de fenotipo afectado, diferentes edades al diagnóstico y las contribuciones individuales de cada familia reclutada. Aunque estos análisis arrojaron resultados consistentes con el modelo propuesto, una alternativa a futuro sería contrastar los hallazgos con estudios en familias con agregación familiar de colelitiasis sin cáncer.

Debido a que no se pudo obtener exámenes imagenológicos que permitan verificar el fenotipo de todos los individuos, se utilizaron algunas estrategias para reducir los sesgos que pudiesen generarse por esta causa; fundamentalmente, considerando sólo a los individuos colecistectomizados, y mediante la asignación del fenotipo “*unknown*” en el análisis de ligamiento genético.

Las estimaciones del riesgo de recurrencia y de riesgo relativo realizadas en esta tesis no son de base poblacional, tampoco se pudo ajustar las estimaciones por variables de confusión como edad, y las exposiciones medioambientales diferenciales como es usual en este tipo de estudios epidemiológicos, no fueron controladas. No obstante, los resultados fueron útiles para la comprensión del comportamiento de las dos enfermedades (CVB y colelitiasis) en las familias estudiadas.

En cuanto al número pequeño de casos esporádicos de CVB, se utilizó como estrategia para aumentar el poder estadístico el uso de un grupo control al menos cuatro veces más grande que el de casos [163], sin embargo la gran cantidad de controles empleados (casi 10 veces mayor) no fue suficiente para compensar un número tan reducido de casos. En este mismo sentido, también se buscó controlar la heterogeneidad de los individuos reclutados (la heterogeneidad de la enfermedad puede provocar una pérdida adicional de la potencia estadística) verificando su diagnóstico histopatológico de CVB, estableciendo que efectivamente se tratasen de casos esporádicos (sin antecedentes familiares), y entrevistándolos a fin de identificar posibles covariables que debiesen incluirse en los análisis.

Si bien es probable que el tamaño del efecto de las variantes involucradas no haya sido lo suficientemente grande como para ser captado con el tamaño de muestra utilizado, también es importante tener en consideración que el nivel de significancia tiende a hacerse más exigente a medida que se aumenta el número de marcadores analizados, debido a las comparaciones o pruebas múltiples que se realizan [164]. Aunque se acepta la aplicación de un ajuste de Bonferroni al nivel de significancia, basado en el número de comparaciones realizadas en el estudio, la prevención de errores de tipo 1 con este método puede llevar a incrementar los errores de tipo 2. Es por esto que la búsqueda y selección de 30 genes candidatos resultó ser una estrategia acertada, la cual favoreció la identificación de dos variantes asociadas con el riesgo de CVB (rs363717 del gen *ABCA1* y rs1558562 del gen *SEC14L5*).

Otro gran desafío en los estudios de asociación (incluyendo el mapeo fino) es generar la colección de controles. En esta investigación fue necesario usar un grupo control que, si bien representa en general a la población chilena, no necesariamente es el grupo ideal para el mapeo fino realizado en pacientes con CVB. A pesar de ello, las posibles fuentes de confusión (sexo, edad y proporción de ascendencia mapuche global) se incluyeron como covariables y mediante un análisis de componentes principales se descartaron posibles subestructuras derivadas de las diferencias genéticas entre los dos grupos.

En términos generales, las debilidades del estudio se abordaron con estrategias que, en mayor o en menor medida, redujeron los posibles efectos negativos sobre los resultados. La idea de utilizar el modelo II (CVB+CCA) en familias con agregación de este cáncer tiene un buen soporte. Dado el considerable riesgo de desarrollar CVB dentro de las familias estudiadas, se podría suponer que las regiones de ligamiento identificadas estarían efectivamente asociadas con la susceptibilidad al CVB, lo que implicaría suponer además que los individuos colecistectomizados hubieran desarrollado CVB en algún momento de su vida. Sin embargo, los resultados del estudio son susceptibles a múltiples interpretaciones, especialmente porque representan tan sólo una parte del panorama, el cual deberá complementarse con estudios en individuos y familias con diagnósticos de cálculos biliares, pero sin antecedentes de CVB.

Es probable que las familias estudiadas en esta tesis presenten agregación del CVB y cálculos biliares como consecuencia del efecto de múltiples variantes de susceptibilidad, de las cuales unas otorgan susceptibilidad a los cálculos, y otras son responsables de la susceptibilidad al CVB, la interacción de ambas en algunos familiares conduciría al desarrollo del CVB. La misma explicación podría considerarse para los casos esporádicos de CVB, quienes también tenían antecedentes de colestiasis. La comprensión de la etiología del CVB dependerá de la identificación de esa serie de variantes e interacciones. Se necesitan más estudios para identificar y caracterizar las variantes causales involucradas en el desarrollo del CVB en Chile.

De momento, los hallazgos que aquí se presentan sugieren que tanto la agregación familiar del CVB y la colestiasis, como la susceptibilidad al CVB esporádico en chilenos está asociada, a variantes genéticas localizadas en las dos regiones candidatas identificadas, una en el cromosoma 9 y la otra en el cromosoma 16. Es destacable además, lo novedoso de esta investigación, pues a la fecha no existen antecedentes de estudios de ligamiento genético en familias con agregación del CVB, y es además el primer análisis de este tipo en población chilena.

## VI. Conclusiones

1. Los individuos con antecedentes de CVB en familiares de primer grado podrían tener un riesgo seis veces mayor de desarrollar CVB que un individuo de la población general. Estudios de base poblacional en Chile podrían confirmar estos hallazgos.
2. Las regiones genómicas chr9:105016960-115983216 y chr16:105444-5995388 se asocian con la colestiasis y el CVB con agregación familiar.
3. Las dos regiones genómicas candidatas identificadas contienen 30 genes potencialmente relevantes dentro de la etiopatogenia del CVB y la colestiasis.
4. Las variantes rs363717 del gen *ABCA1* (OR=1,84 95% IC 1,26-2,66) y rs1558562 del gen *SEC14L5* (OR=2,02 95% IC 1,4-2,95) se asocian con el riesgo de CVB esporádico.
5. La agregación familiar de cáncer de vesícula biliar y colestiasis se asocia con variantes genéticas de riesgo para cáncer de vesícula biliar esporádico.
6. De nuestro conocimiento, este es el primer análisis de ligamiento genético en familias con agregación del CVB y su realización en población chilena destaca aún más la trascendencia de nuestros resultados.

## **VII. Material Científico Generado**

### **1. Publicaciones**

- Lorenzo Bermejo J, Boekstegers F, González Silos R, Marcelain K, **Báez Benavides P**, Barahona Ponce C, *et al.* (2017) Subtypes of Native American ancestry and leading causes of death: Mapuche ancestry-specific associations with gallbladder cancer risk in Chile. *PLoS Genet* 13(5): e1006756. <https://doi.org/10.1371/journal.pgen.1006756>

### **2. Presentaciones**

- Baez Benavides P. “Update on Environmental and Genetic Characteristics of Chilean Gallbladder Cancer. Patients and Families”. Symposium Towards Personalized Prevention and Treatment of Gallbladder Cancer 2017.
- Baez Benavides P. “Environmental and Genetic Characteristics of Chilean Gallbladder Cancer Families”. Symposium Towards Personalized Prevention and Treatment of Gallbladder Cancer 2016.
- Baez Benavides P. “Genetic Linkage Analysis in Chilean families with gallbladder cancer”. Journal Club Presentation at the Statistical Genetics Group, IMBI, Heidelberg University 2016.



## VIII. Material Suplementario

### 1. Tablas

**Tabla Suplementaria 1.** Estadificación del CVB.

T = Tamaño del tumor (0-4), N = Número de ganglios afectados/linfáticos regionales (x-0-3). M = Metástasis a distancia. (0-1).

Etapas	Grupo T	Grupo N	Grupo M	Descripción
0	Tis	N0	M0	Existe solo un pequeño cáncer en la capa epitelial de la vesícula biliar (Tis). No se ha propagado fuera de la vesícula biliar.
I	T1(a o b)	N0	M0	El tumor crece hacia la lámina propia (T1a) o la capa muscular (T1b). No se ha propagado fuera de la vesícula biliar.
II	T2	N0	M0	El tumor creció hacia el tejido fibroso perimuscular (T2). No se ha propagado fuera de la vesícula biliar.
IIIA	T3	N0	M0	El tumor se extiende a través de la serosa (la capa exterior de la vesícula biliar) y/o crece directamente hacia el hígado y/o a otra estructura cercana (T3). No se ha propagado a los ganglios linfáticos cercanos (N0) ni a tejidos u órganos distantes a la vesícula biliar (M0).
IIIB	T1 a T3	N1	M0	El tumor no ha crecido hacia los vasos sanguíneos principales que van hacia el hígado ni ha alcanzado más de un órgano cercano aparte del hígado (T1 a T3), pero ha alcanzado los ganglios linfáticos cercanos (N1). El cáncer no se ha propagado a los tejidos ni a órganos distantes a la vesícula biliar (M0).
IVA	T4	N0 o N1	M0	El tumor crece hacia los vasos sanguíneos principales que van hacia el hígado o ha alcanzado más de un órgano cercano además del hígado (T4). Puede o no haberse propagado a los ganglios linfáticos adyacentes (N0 o N1). El cáncer no se ha propagado a los tejidos ni a órganos distantes a la vesícula biliar (M0).
IVB	Cualquier T	N2	M0	El tumor principal puede o no haberse propagado fuera de la vesícula biliar. Se ha propagado a los ganglios linfáticos lejanos a la vesícula biliar (N2). El cáncer no se ha propagado a los tejidos ni a órganos distantes a la vesícula biliar (M0).
	Cualquier T	Cualquier N	M1	El tumor principal podría o no haberse propagado fuera de la vesícula biliar. Podría o no haberse propagado a los ganglios linfáticos. El tumor se ha propagado a los tejidos u órganos distantes de la vesícula biliar (M1).

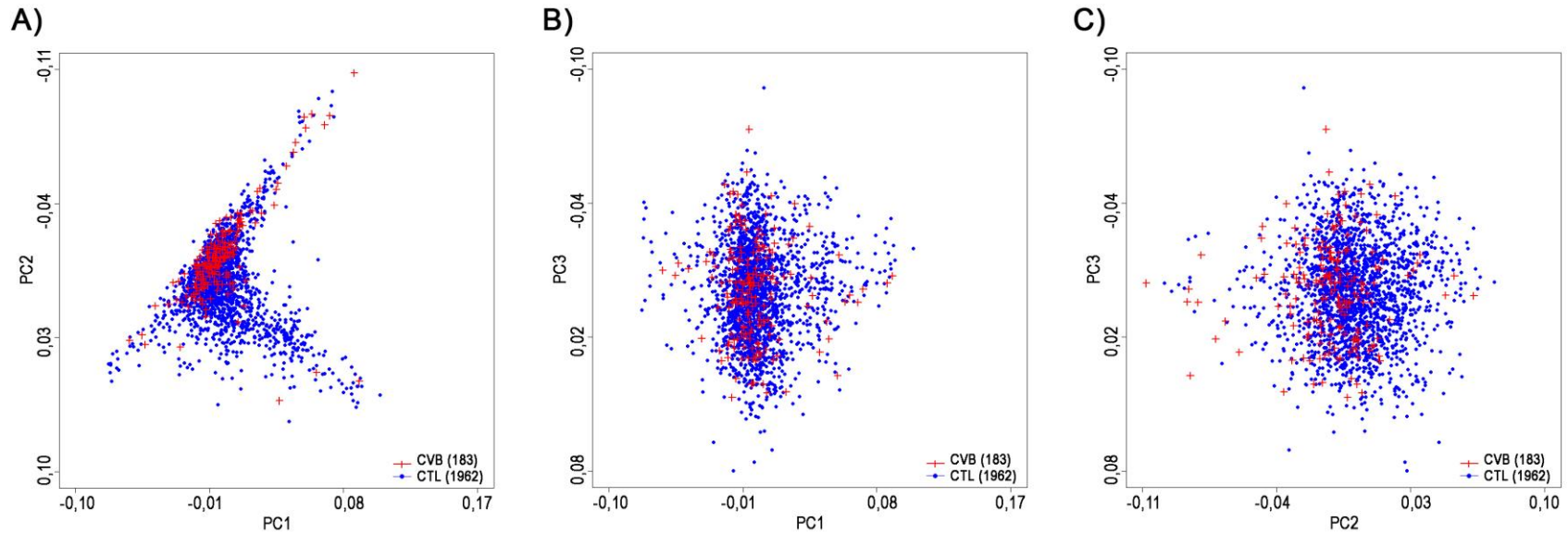
Adaptada de [26].

**Tabla Suplementaria 2.** Anormalidades genéticas frecuentes en tumores de la vesícula biliar.

<b>Cromosomales [51-53, 165, 166]</b>	
Bandas preferentemente implicadas	1p32, 1p36, 1q32, 3p21, 6p21, 7p13, 7q11, 7q32, 19p13, 19q13, 22q13
Aberraciones estructurales	del(3)(p13), i(5)(p10), del(6)(q13), del(9)(p13), del(16)(q22), del(17)(p11), i(17)(q10), del(19)(p13), i(21)(q10)
Ganancias	3q, 5p, 7p, 7q, 8q, 11q, 13q, 17q
Pérdidas	3p, 4q, 5q, 9p, 10p, 10q, 11p, 14p, 14q, 15p, 17p, 19p, 21p, 21q, Xp
<b>Moleculares</b>	
Oncogenes	ERBB, KRAS [42-48, 167-169]
Genes Supresores Tumorales	TP53, p16, FHIT, EGFR, PIK3CA [48, 50, 168, 170-172]
Firmas mutacionales	firma A (TC(A/C/T)>TG(A/C/T) y TCN>TTN)* firma B ((A/C/G)CG>(A/C/G)TG) [48, 173]

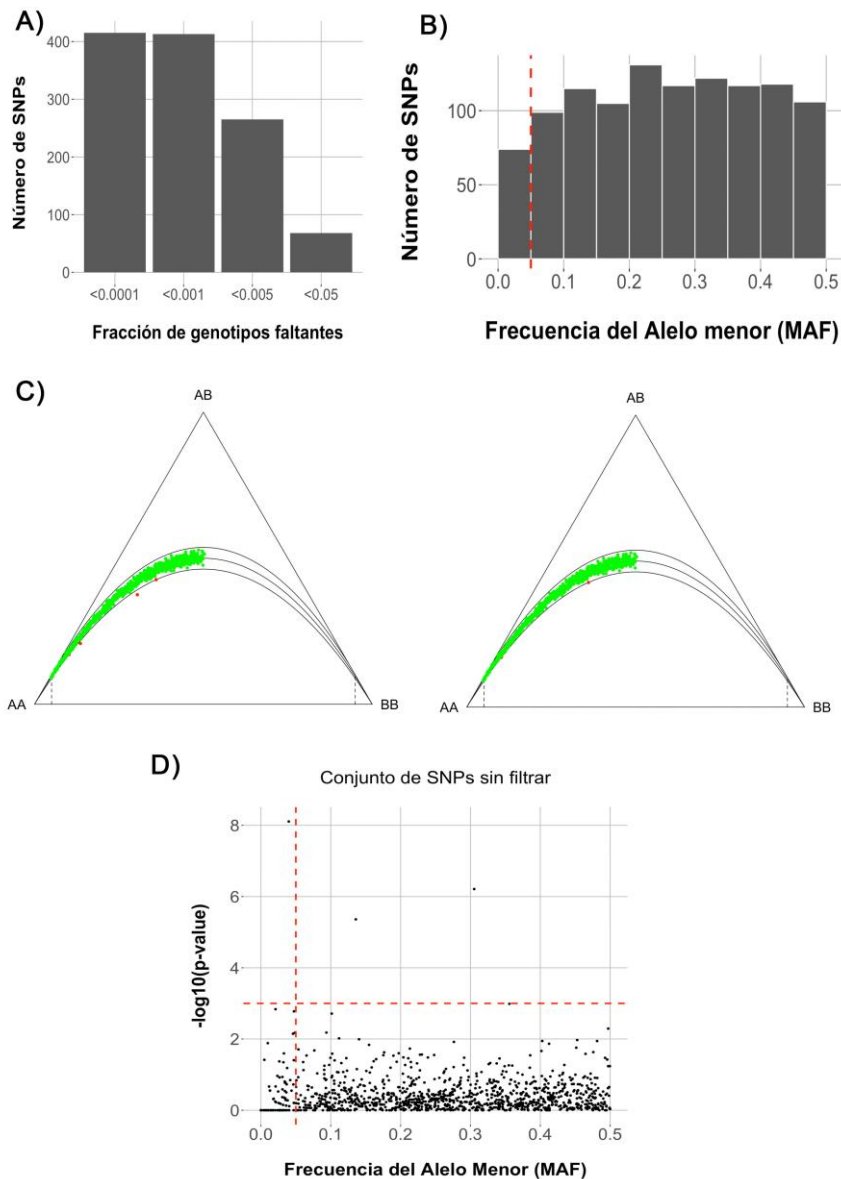
\*N: Cualquier base

## 2. Figuras



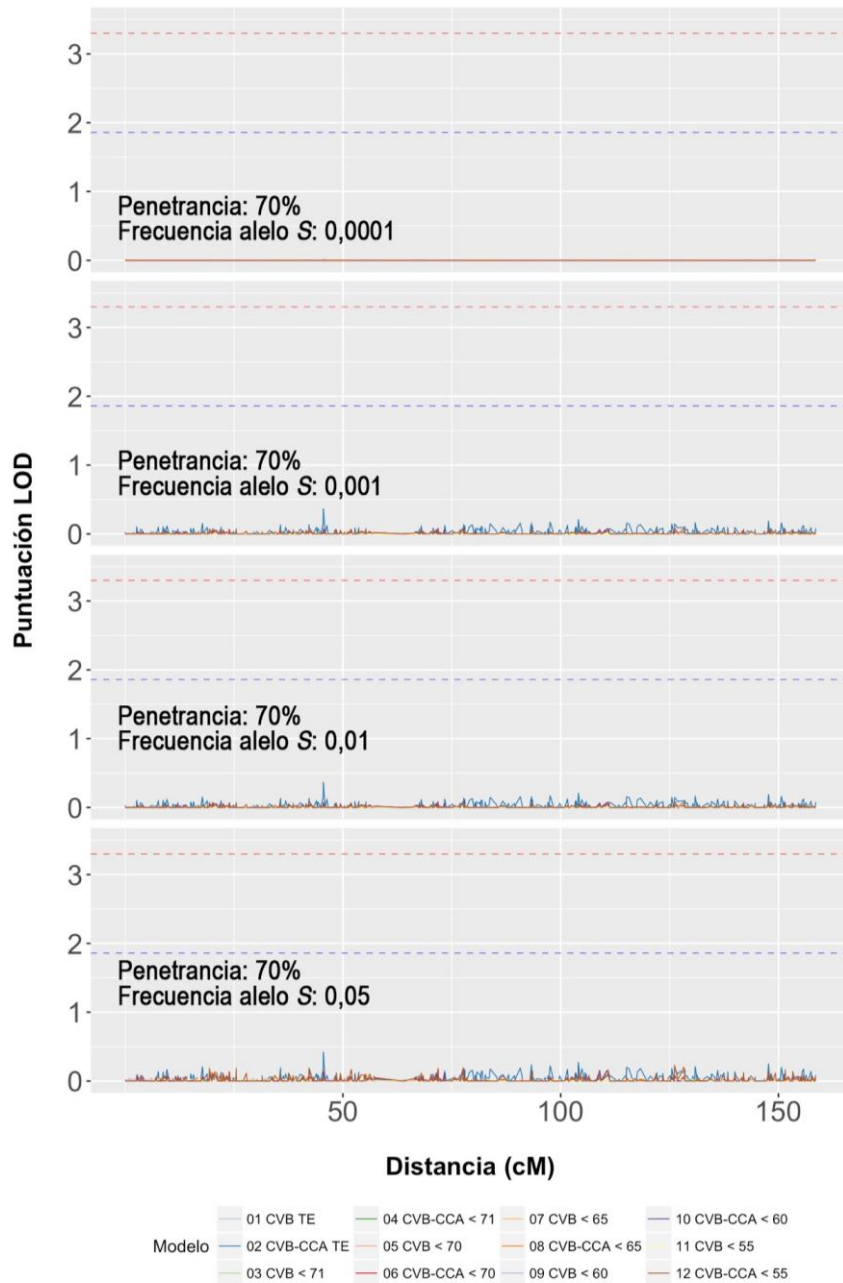
**Figura Suplementaria 1.** Análisis de componentes principales (ACP) de 2.145 individuos chilenos: 183 casos de CVB y 1.962 controles.

Se muestran los resultados de un análisis de componentes principales usando 1.161 SNPs localizados en las regiones de ligamiento candidatas de los cromosomas 9: 105-116 Mb y 16: 0-6 Mb, y que se emplearon en el análisis de asociación genética. En la figura se observa la ausencia de subestructuras dentro de los individuos con CVB (rojo) y los controles poblacionales (azul), para las componentes: **A)** 1 y 2, **B)** 1 y 3, **C)** 2 y 3.



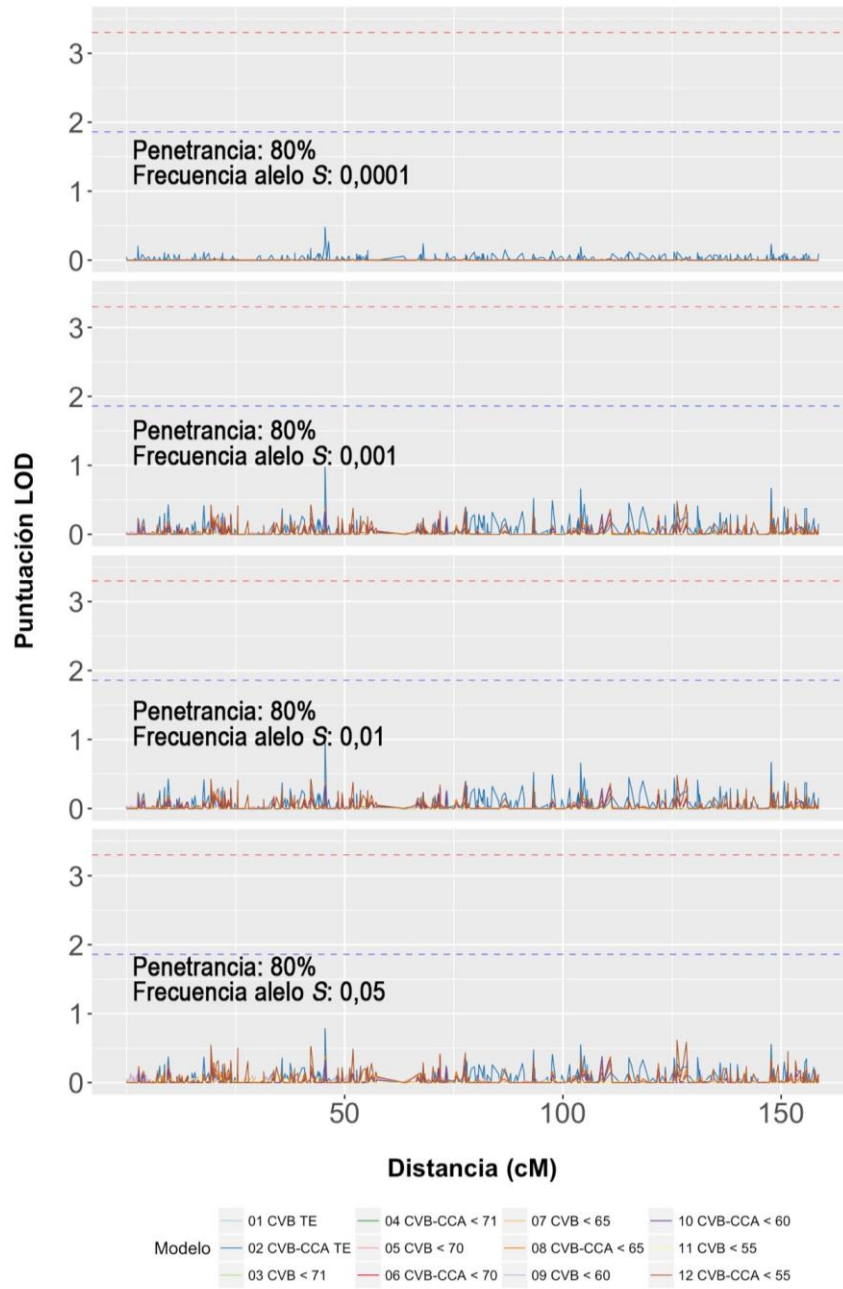
**Figura Suplementaria 2.** Control de calidad para un conjunto de 1.161 variantes utilizadas en un mapeo fino por asociación en casos con CVB y controles.

**(A)** En ningún SNP se observó un porcentaje de genotipos faltantes superior al 5%, por lo que no se eliminaron variantes en este paso. **(B)** Se estimó la frecuencia del alelo menor (MAF) de cada marcador y se eliminaron 79 SNPs con un MAF <math><0,05</math>, por eso el dominio es de 0 a 0,5. **(C)** Se removieron los SNPs con un Ternary Plot para 1.030 variantes en 1.962 individuos control. Izquierda: *Ternary plot* en donde se removieron 2 SNPs. **(D)** *Scatter plot* para la frecuencia del alelo menor (MAF) y el valor

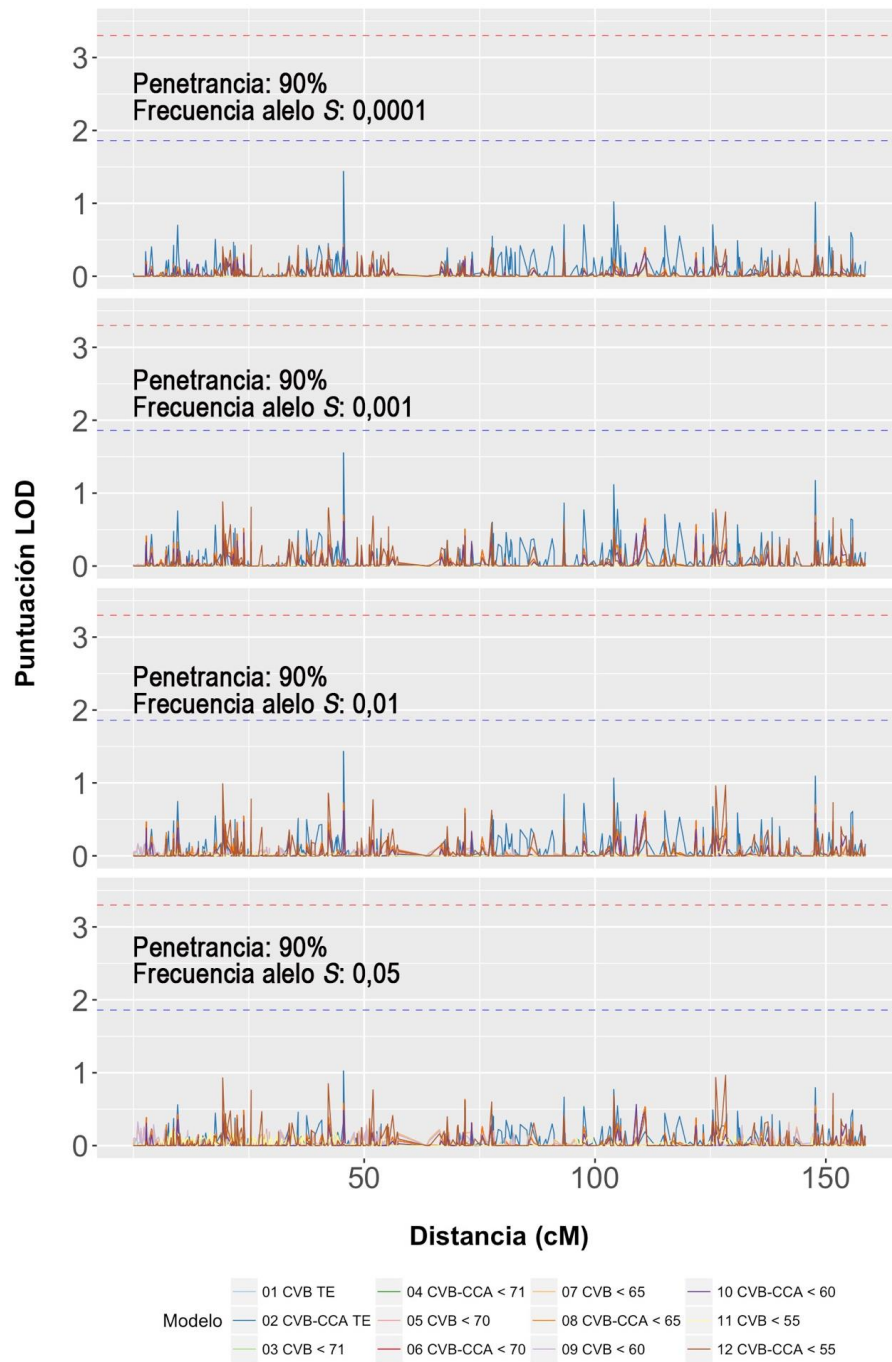


**Figura Suplementaria 3.** Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 7. (Páginas 114-118).

Se realizaron análisis de ligamiento paramétrico variando los valores de la penetrancia (70%, 80%, 90%, 95%, 99%), tasa de fenocopias (30%, 20%, 10%, 5%, 1%) y frecuencia del alelo de susceptibilidad (0,0001; 0,001; 0,01; 0,05). Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $LOD \geq 3,3$  y el umbral sugerente  $LOD \geq 1,86$  de ligamiento, respectivamente. CVB: Cáncer de Vesícula Biliar, CCA: Colectectomía, TE: Todas las edades.

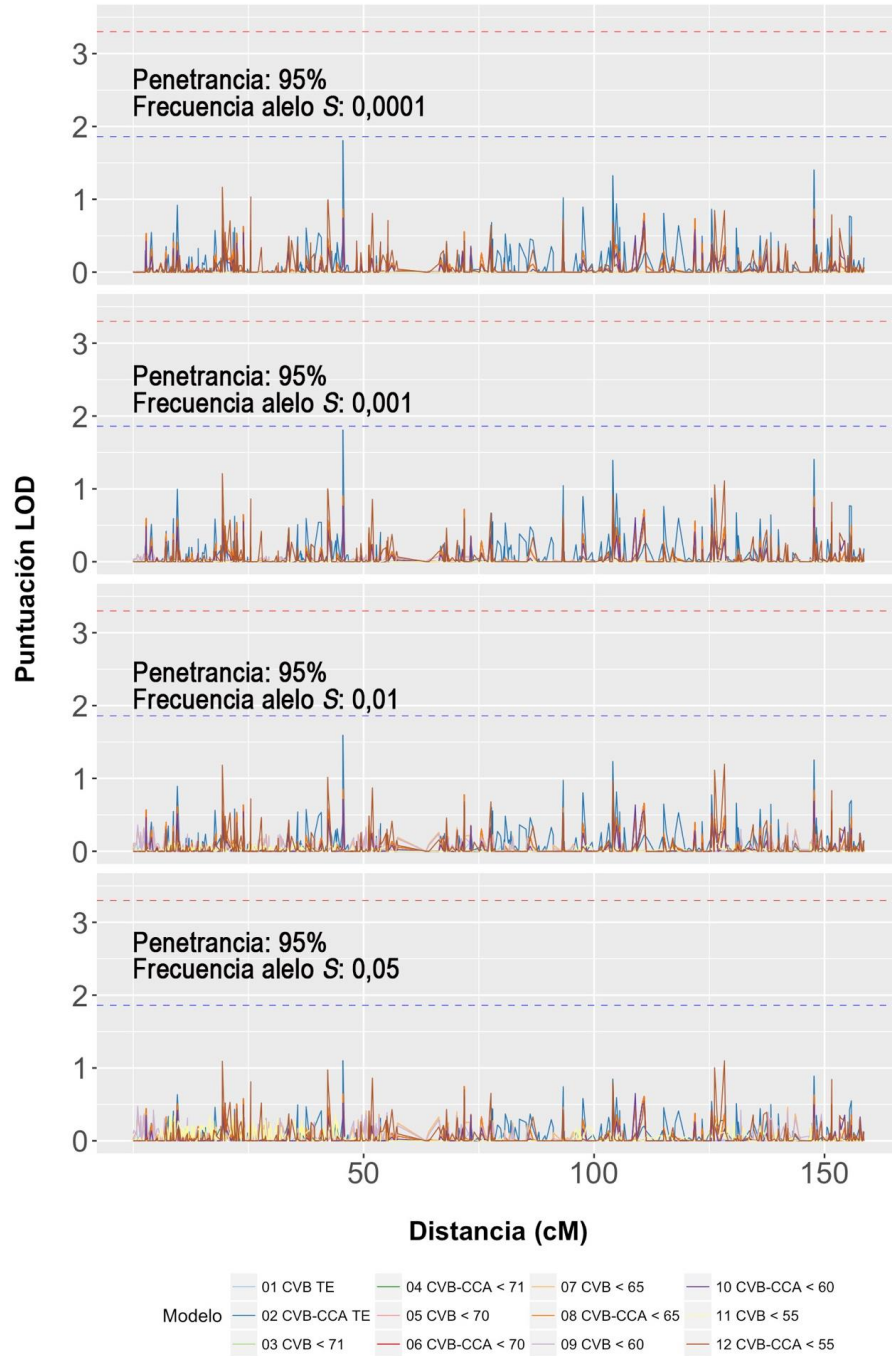


**Figura Suplementaria 3.** Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 7.



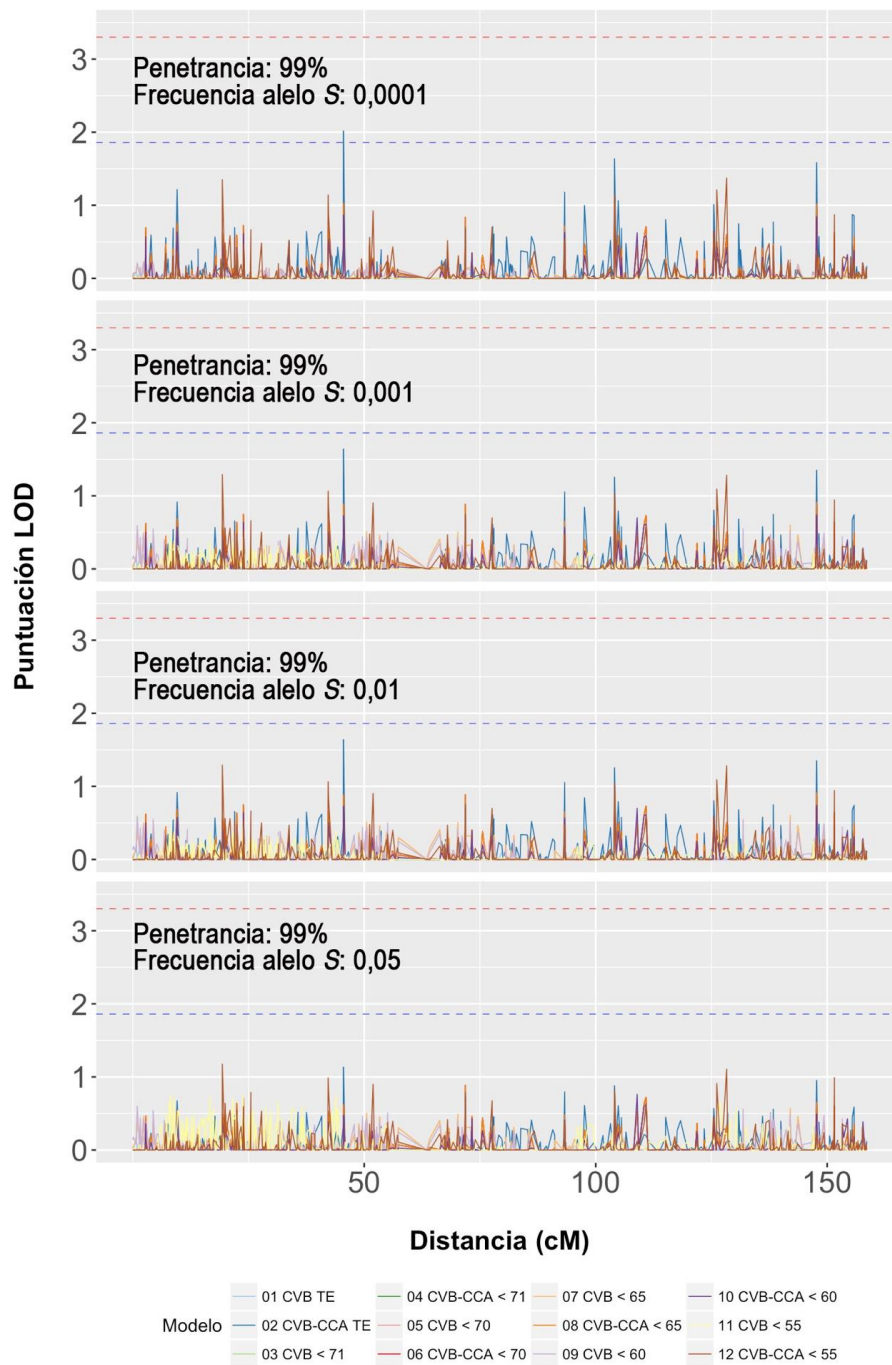
**Figura Suplementaria 3.** Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 7.



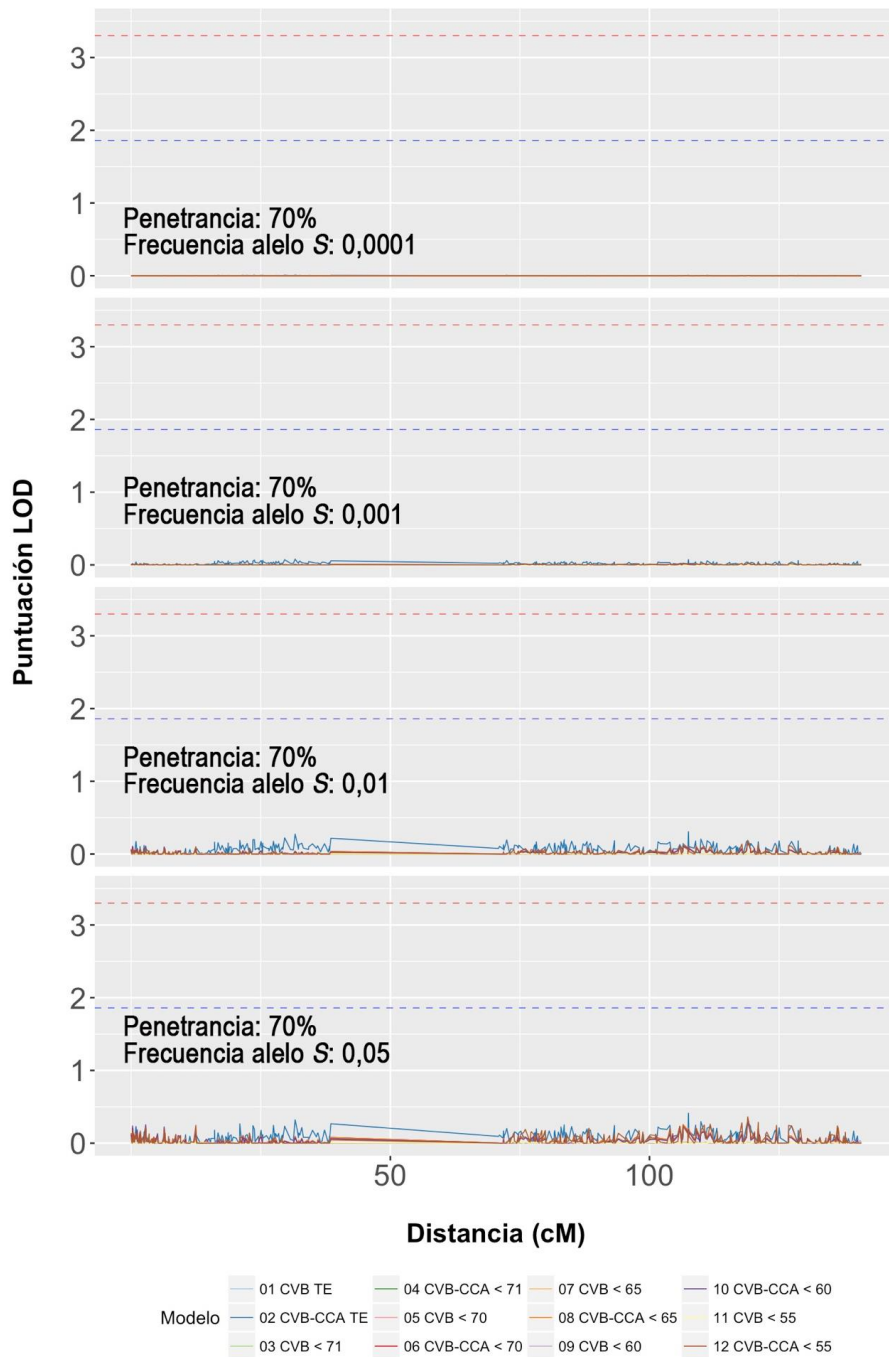


**Figura Suplementaria 3.** Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 7.





**Figura Suplementaria 3.** Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 7.



**Figura Suplementaria 4.** Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 9. (Páginas 119 - 123)

Se realizaron análisis de ligamiento paramétrico variando los valores de la penetrancia (70%, 80%, 90%, 95%, 99%), tasa de fenocopias (30%, 20%, 10%, 5%, 1%) y frecuencia del alelo de susceptibilidad (0,0001; 0,001; 0,01; 0,05). Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $\text{LOD} \geq 3,3$  y el umbral sugerente  $\text{LOD} \geq 1,86$  de ligamiento, respectivamente. CVB: Cáncer de Vesícula Biliar, CCA: Colectectomía, TE: Todas las edades.

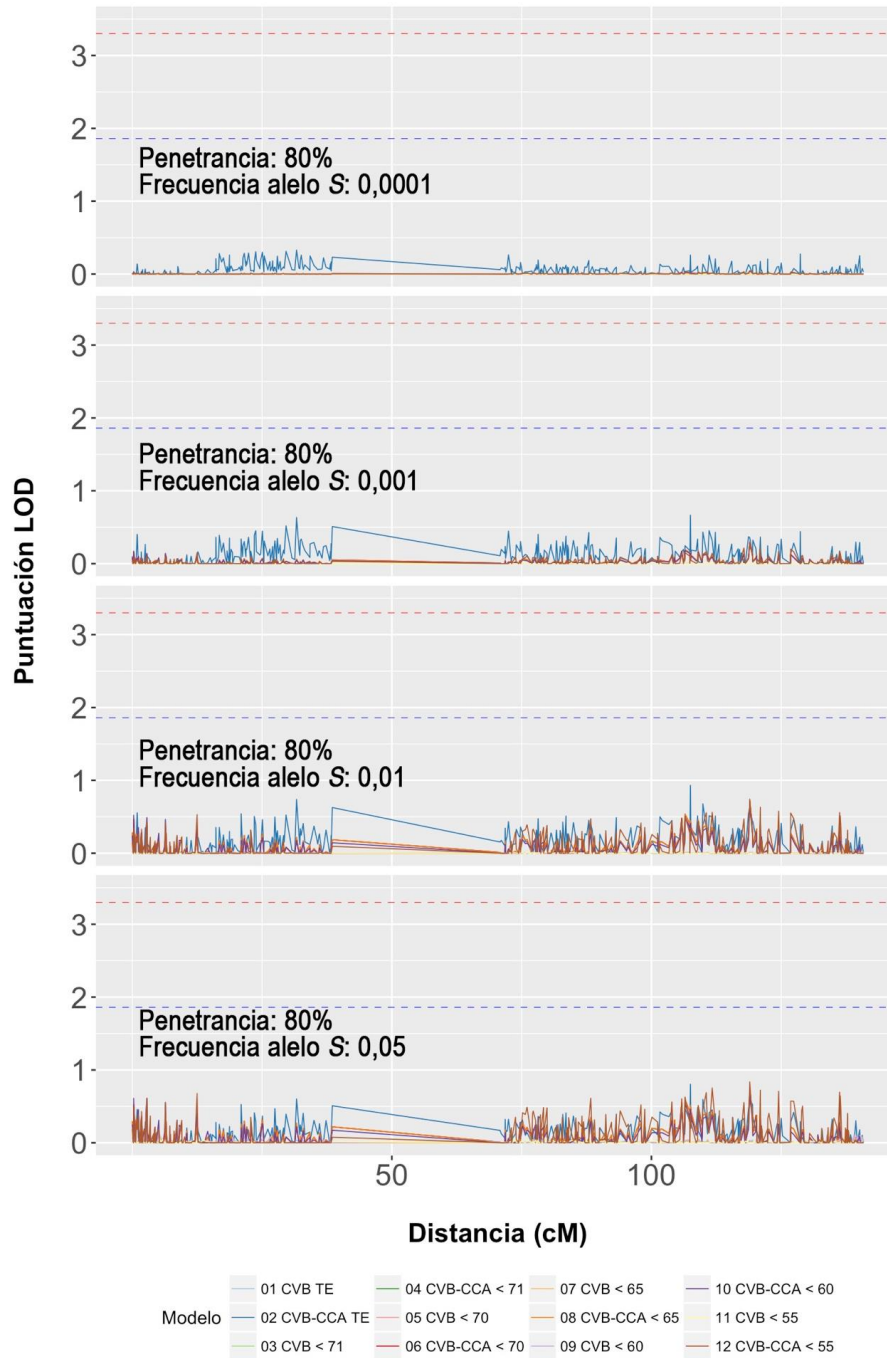


Figura Suplementaria 4. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 9.

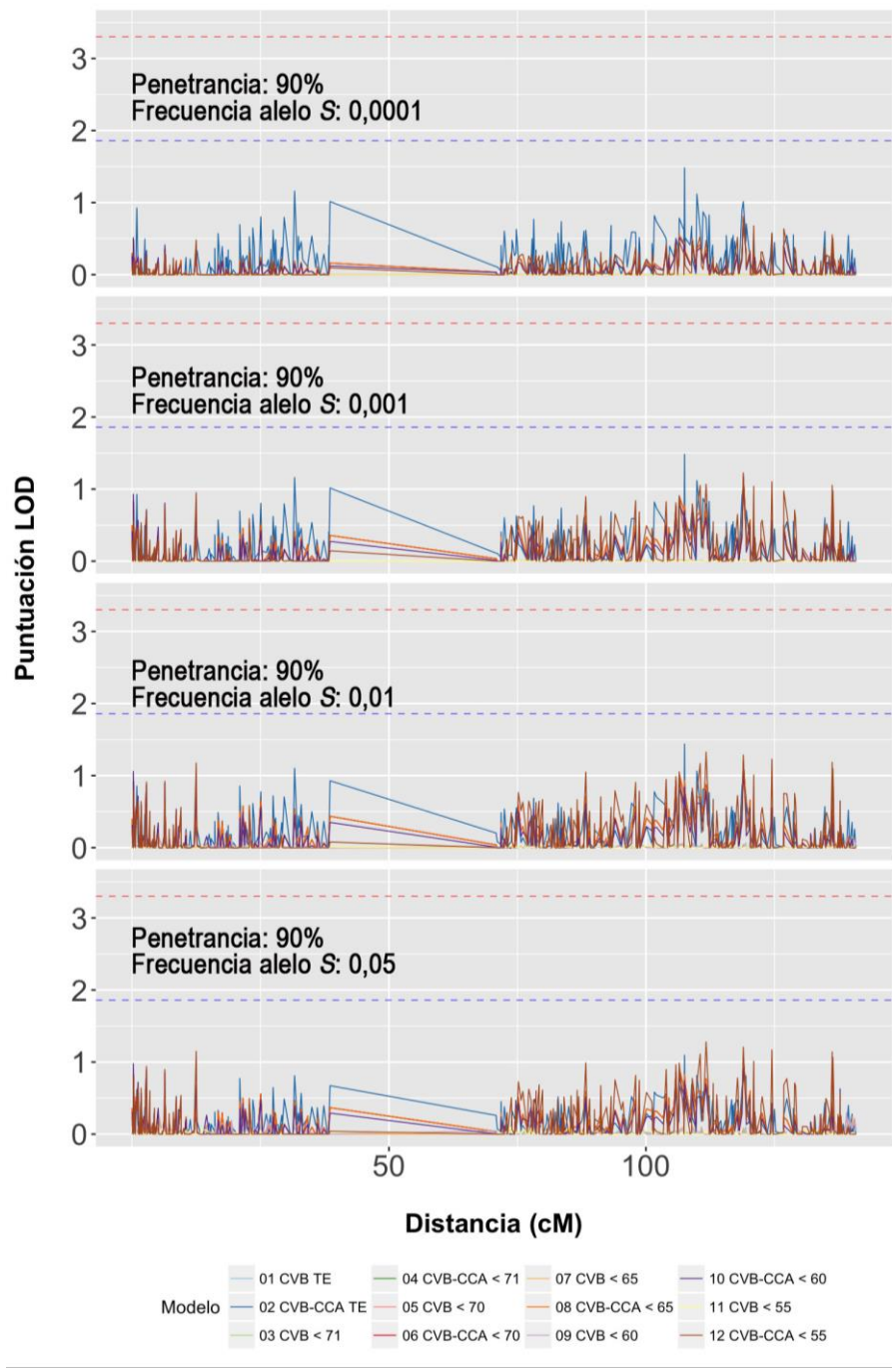


Figura Suplementaria 4. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 9.

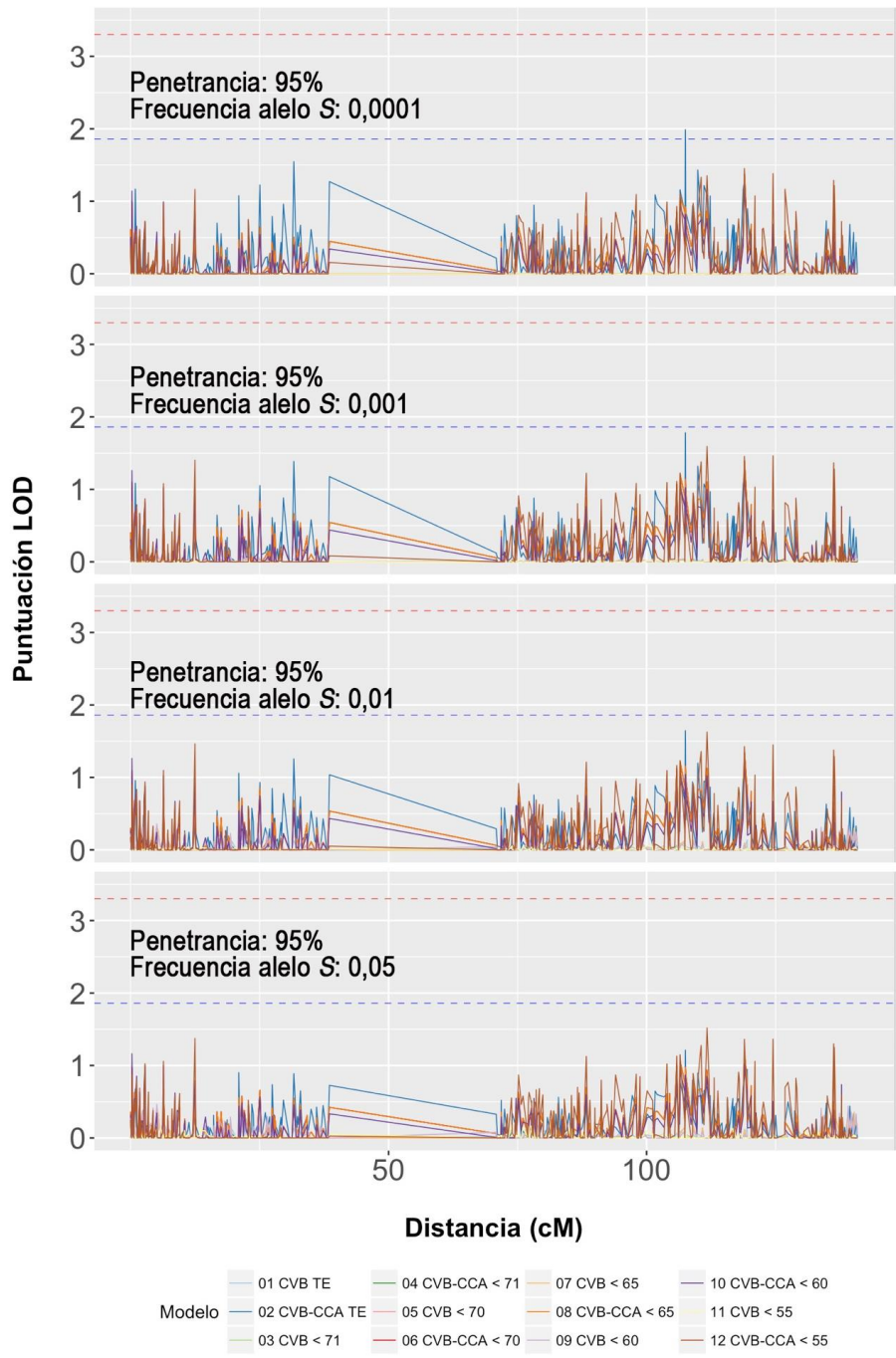


Figura Suplementaria 4. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 9.

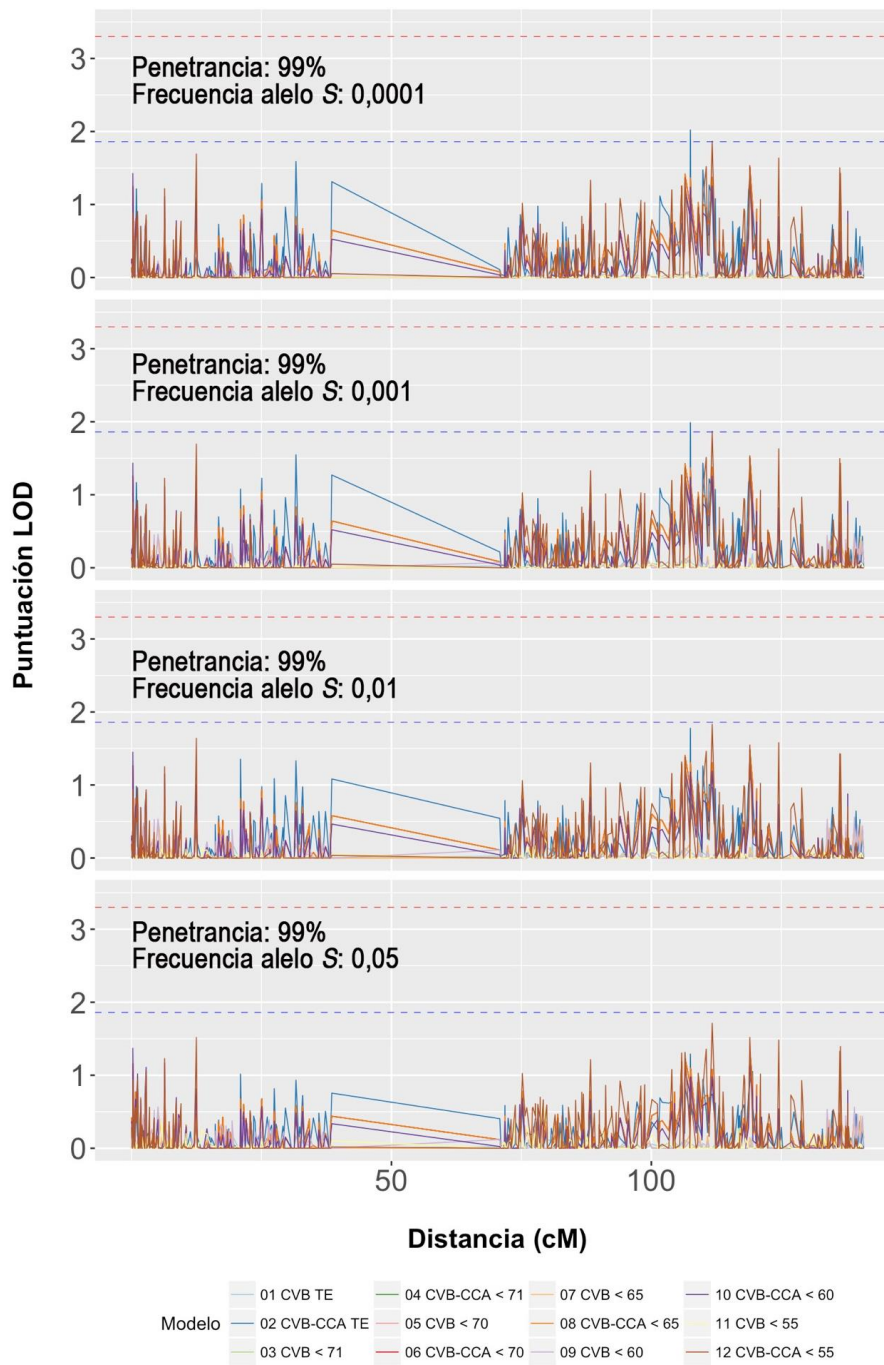
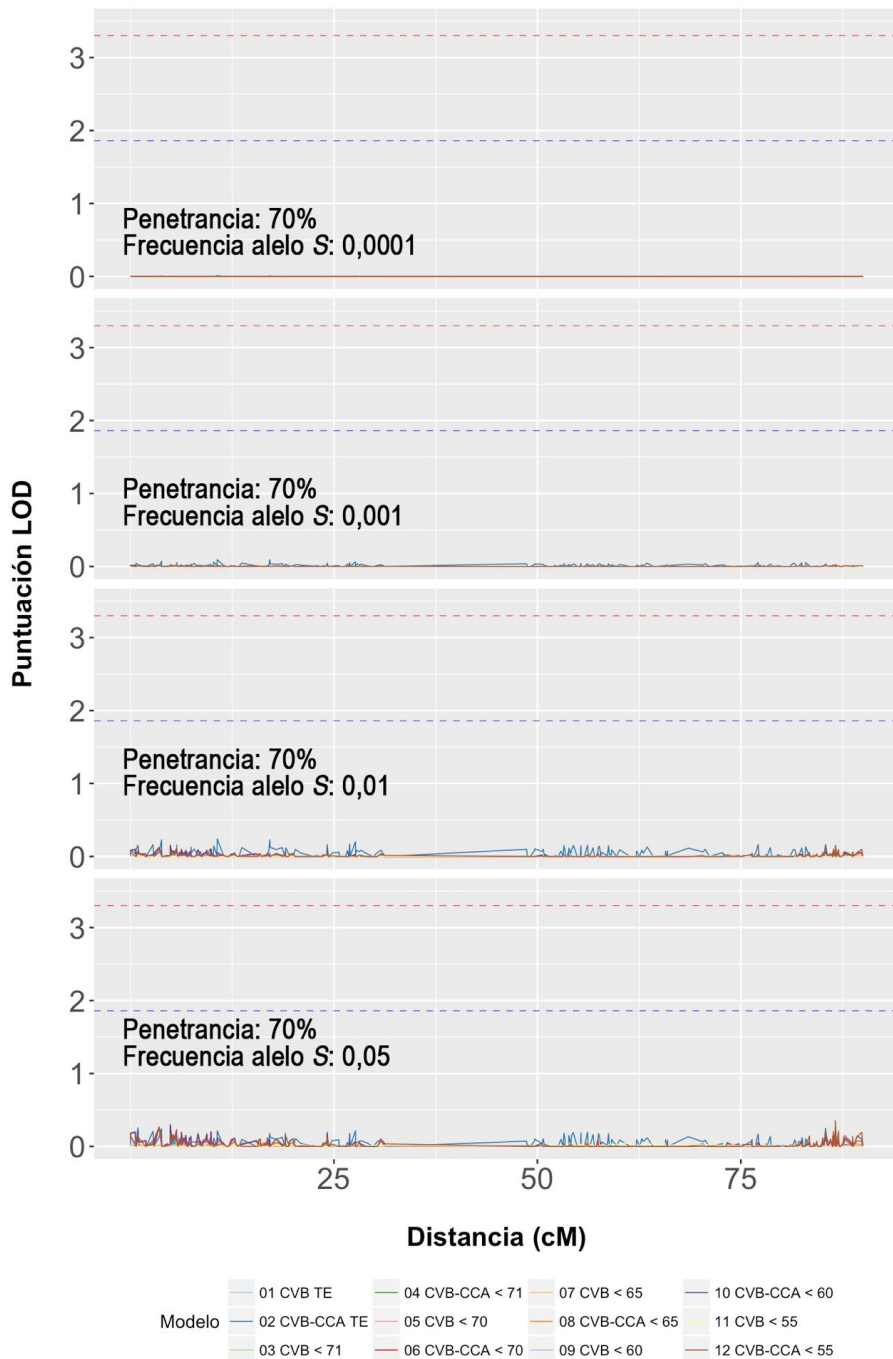


Figura Suplementaria 4. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 9.





**Figura Suplementaria 5.** Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 16. (Páginas 124-128)

Se realizaron análisis de ligamiento paramétrico variando los valores de la penetrancia (70%, 80%, 90%, 95%, 99%), tasa de fenocopias (30%, 20%, 10%, 5%, 1%) y frecuencia del alelo de susceptibilidad (0,0001; 0,001; 0,01; 0,05). Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $\text{LOD} \geq 3,3$  y el umbral sugerente  $\text{LOD} \geq 1,86$  de ligamiento, respectivamente. CVB: Cáncer de Vesícula Biliar, CCA: Colectectomía, TE: Todas las edades.

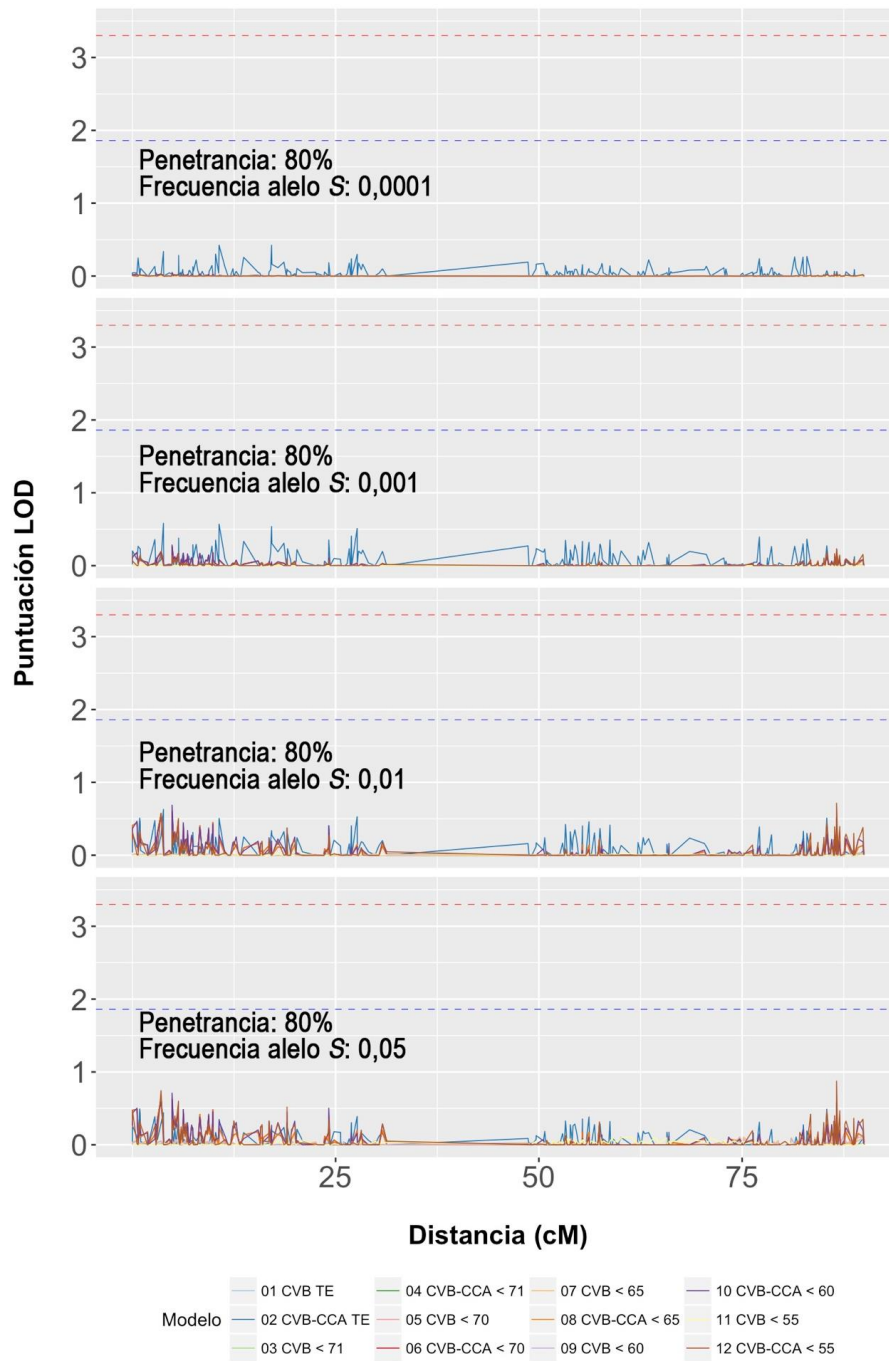


Figura Suplementaria 5. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 16.



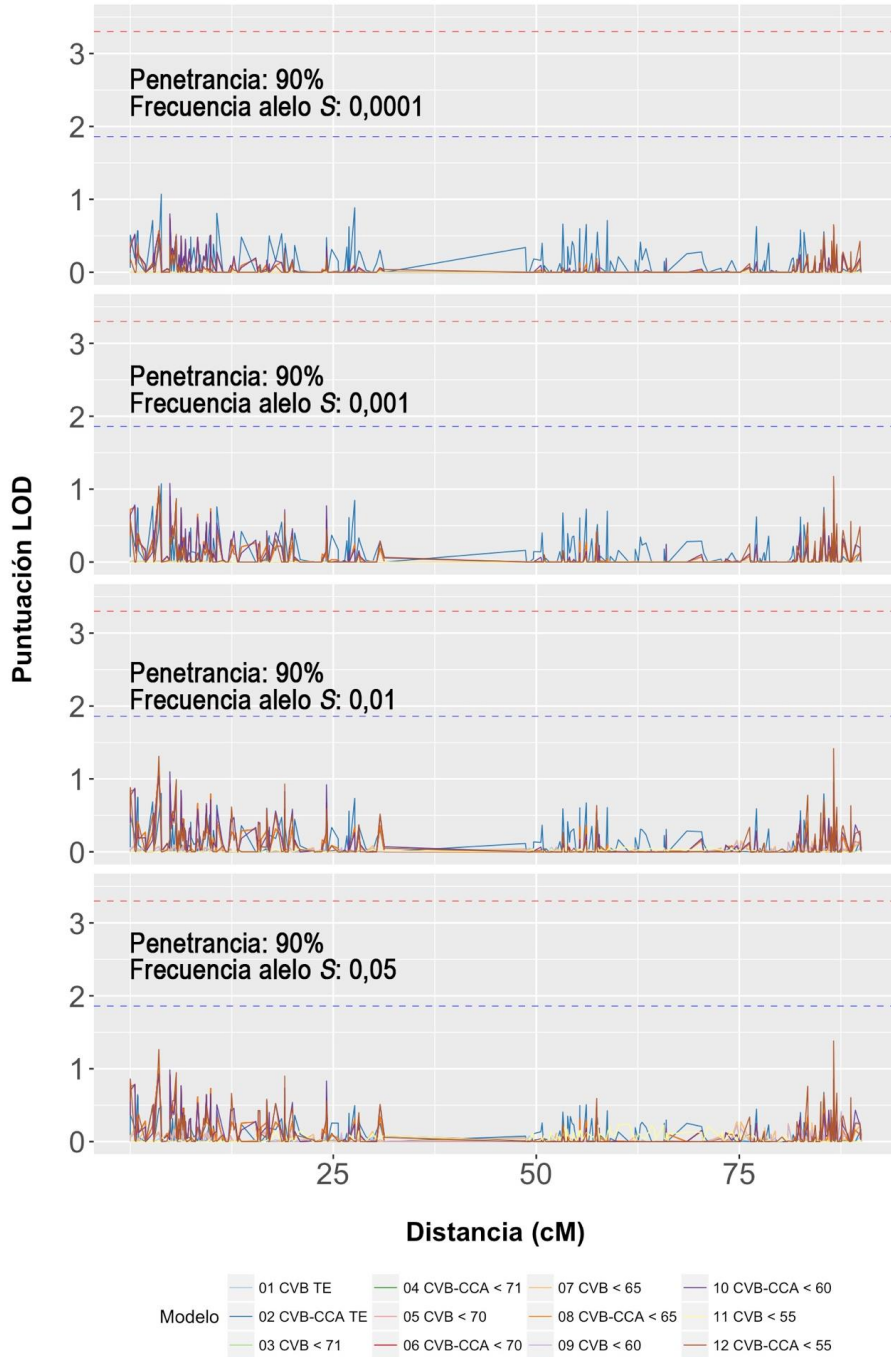


Figura Suplementaria 5. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 16.

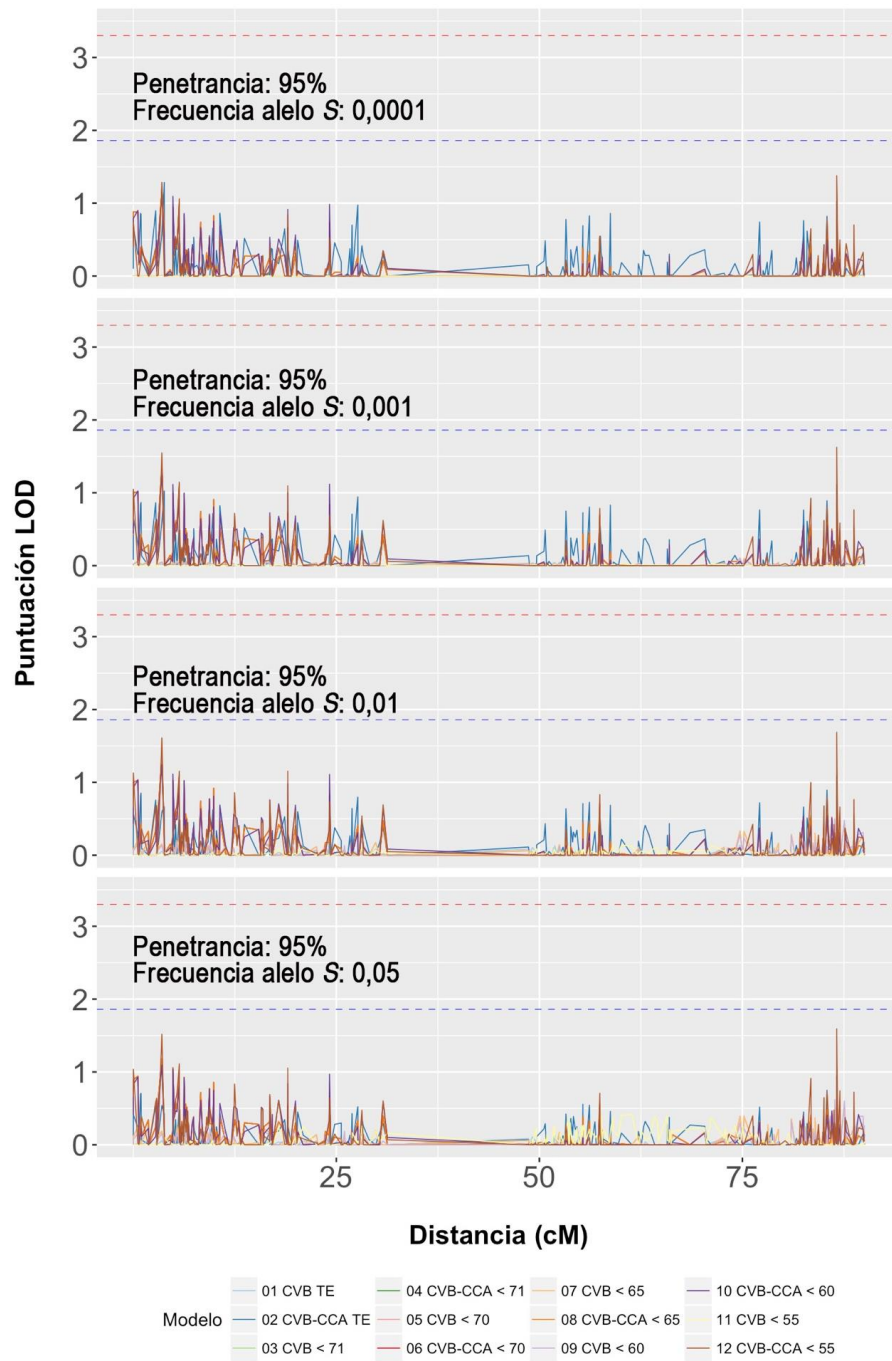


Figura Suplementaria 5. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 16.

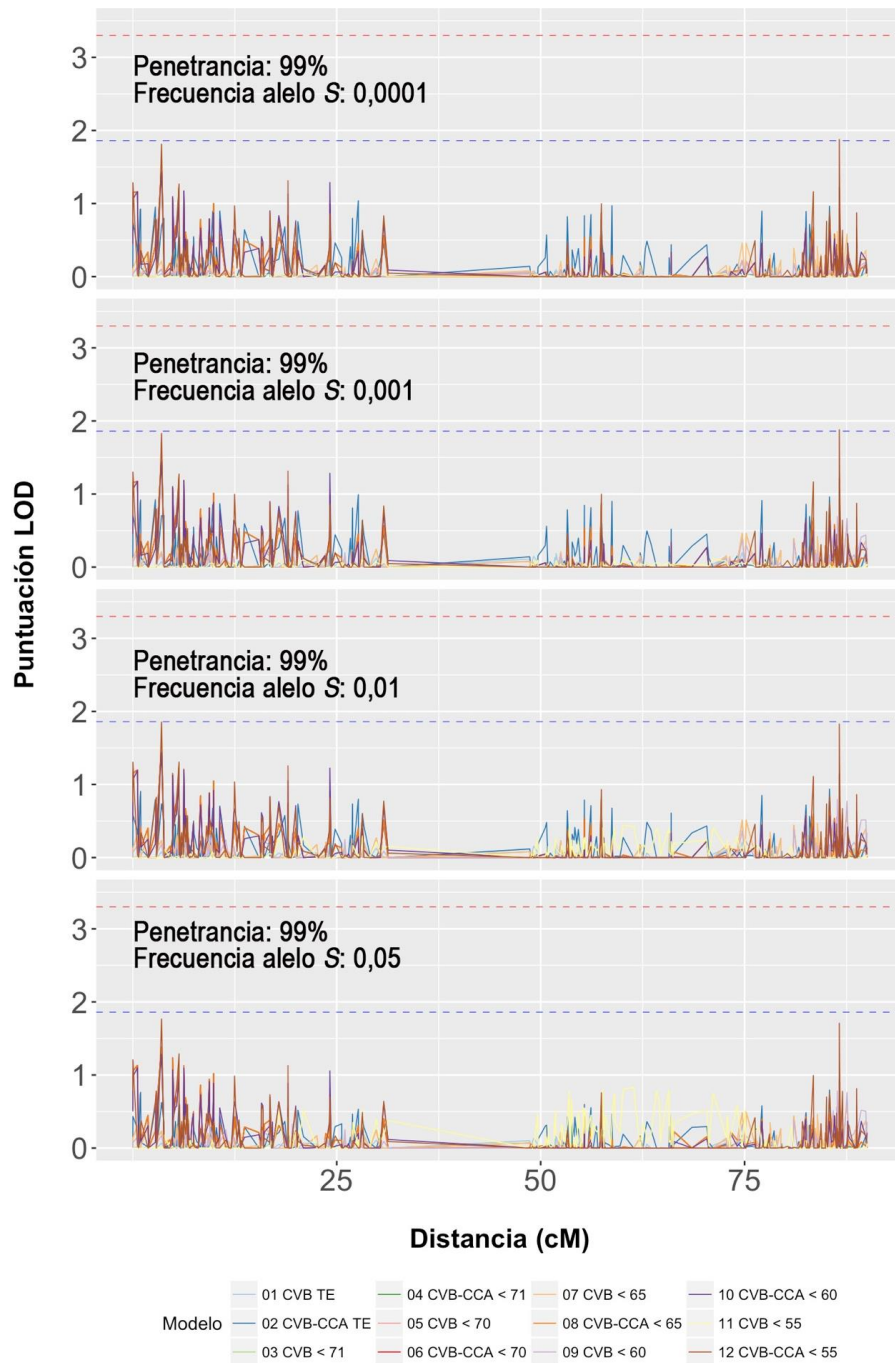
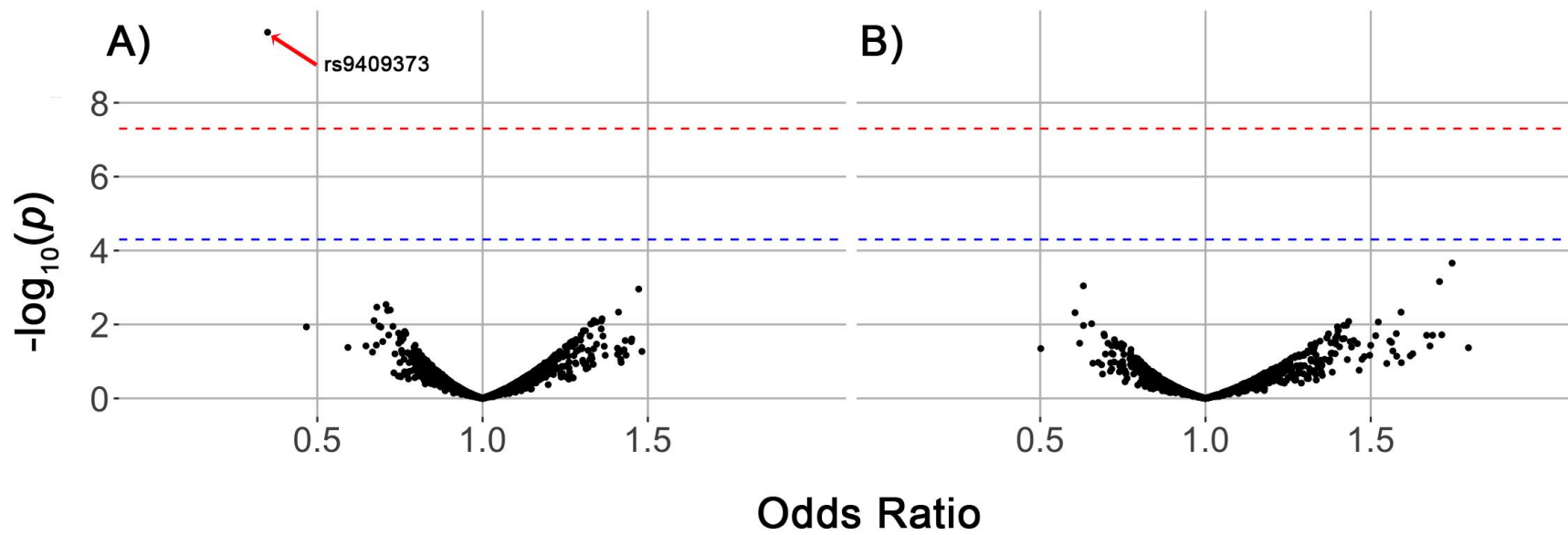
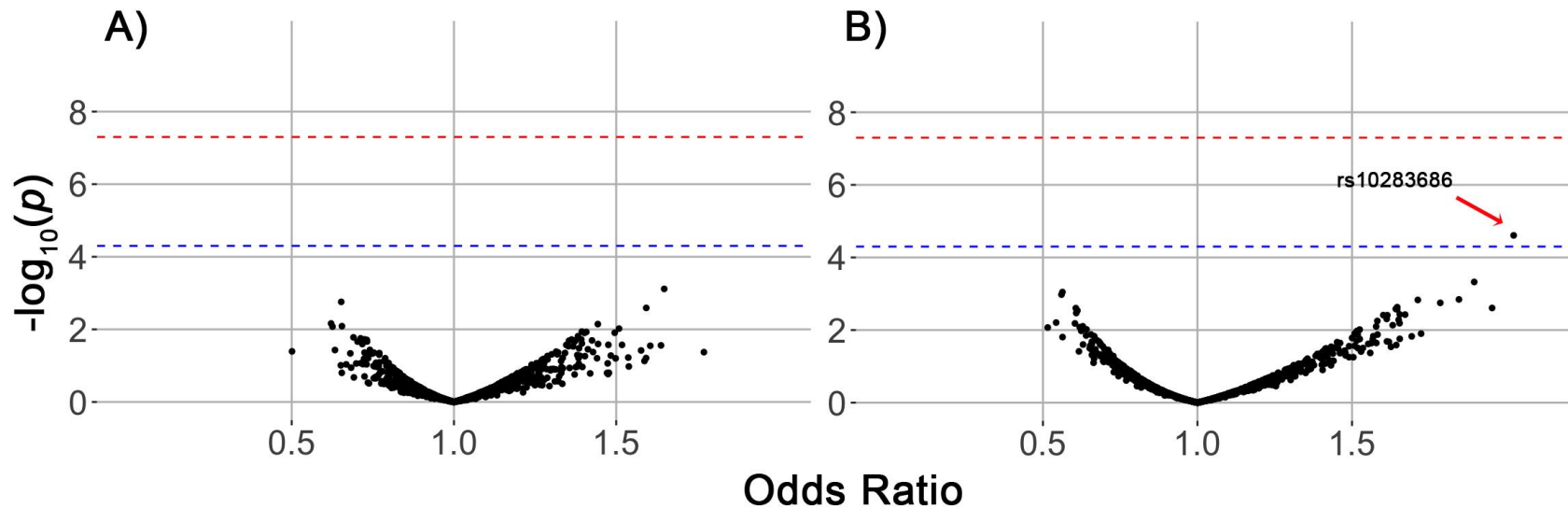


Figura Suplementaria 5. Análisis de sensibilidad para el ligamiento paramétrico en el cromosoma 16.



**Figura Suplementaria 6.** Odds ratios estimados a partir del análisis de regresión logística en 183 casos con CVB y 1.962 controles.

Se muestran los *odds ratios* y *p* valores ( $-\log_{10}(p\text{-val})$ ) de 1.028 variantes analizadas, en un modelo aditivo **A)** sin covariables y **B)** con covariables (Edad, sexo y porcentaje de ascendencia mapuche). Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ) y el umbral de significancia con corrección de Bonferroni para el *dataset*  $p < 4,9e^{-5}$  ( $-\log_{10}(p) > 4,3$ ), respectivamente.



**Figura Suplementaria 7.** Odds ratios estimados a partir del análisis de regresión logística en 183 casos con CVB y 1.962 controles.

Se muestran los *odds ratios* y *p* valores ( $-\log_{10}(p\text{-val})$ ) de 1.028 variantes analizadas, en un modelo genético **A)** aditivo y **B)** dominante, ajustados por las variables edad y sexo. Las líneas punteadas roja y azul indican el nivel de significancia para el genoma completo  $p < 5e^{-8}$  ( $-\log_{10}(p) > 7,3$ ) y el umbral de significancia con corrección de Bonferroni para el *dataset*  $p < 4,9e^{-5}$  ( $-\log_{10}(p) > 4,3$ ), respectivamente.

### 3. Encuesta

BIOBANCO DE TUMORES Y FLUIDOS  
DE LA UNIVERSIDAD DE CHILE

#### ENTREVISTA PRESENCIAL A DONANTES DE MUESTRA AL BIOBANCO PARA EL ESTUDIO SOBRE CÁLCULOS BILIARES Y CÁNCER DE VESÍCULA BILIAR

Código del Donante: \_\_\_\_\_

Entrevistador/a: \_\_\_\_\_

Se le ha leído y obtenido el consentimiento al entrevistado Si \_\_\_ No \_\_\_

Fecha de encuesta: \_\_/\_\_/\_\_\_\_

El objetivo de este cuestionario es conocer algunos factores de riesgo y determinantes que pudieron ocasionar su enfermedad.

Esta información contribuirá a conocer por qué la población Chilena se enferma de cálculos biliares y de cáncer de vesícula biliar, permitiendo en el futuro prevenir y tratar de mejor forma estas enfermedades.

Su colaboración es esencial e insustituible, por lo que le agradecería responder todo el cuestionario, siguiendo las instrucciones que le iré mencionando.

SU INFORMACIÓN SE TRATARÁ DE FORMA ANÓNIMA Y  
CONFIDENCIAL.

SI TIENE PROBLEMAS O DIFICULTADES PARA COMPRENDER LAS  
PREGUNTAS, POR FAVOR CONSULTEME.

El entrevistador necesita el siguiente material:

- Huincha de medir (cinta métrica)
- Calculadora para el IMC
- Lápiz y papel para dibujar el árbol genealógico de familias con varios miembros afectados por cáncer de vesícula biliar

El encuestador lee todas las alternativas posibles antes de que el encuestado responda.

En caso de una respuesta múltiple del encuestado, el encuestador le ayuda a elegir una y sólo una respuesta. Por ejemplo, en la pregunta 8: Bebe usted MÁS BIEN 2 ó 3 vasos por vez?

BIOBANCO DEL HOSPITAL CLINICO UNIVERSIDAD DE CHILE  
Versión 05/02/2014

1/7

**ANTECEDENTES SOCIODEMOGRAFICOS**

A1 ¿Cuál es su nombre \_\_\_\_\_ y apellido(s) \_\_\_\_\_ ?

A2 Sexo: Hombre \_\_\_ Mujer \_\_\_

A3 Teléfono: \_\_\_\_\_

A4 ¿Cuál es su fecha de nacimiento? \_\_\_/\_\_\_/\_\_\_

A5 ¿Cuántos años tenía cuándo le operaron? \_\_\_\_\_ años

A6 ¿Dónde nació usted? \_\_\_\_\_

A7 (Si nació en Chile) ¿En qué ciudad o pueblo de Chile? \_\_\_\_\_

A8 ¿En qué ciudades o pueblos ha residido usted desde su nacimiento?

En \_\_\_\_\_ hasta los \_\_\_\_\_ años de edad

En \_\_\_\_\_ hasta los \_\_\_\_\_ años de edad

En \_\_\_\_\_ hasta los \_\_\_\_\_ años de edad

En \_\_\_\_\_ hasta los \_\_\_\_\_ años de edad

En \_\_\_\_\_ hasta los \_\_\_\_\_ años de edad

¿En qué ciudad o pueblo nació su ...	A9 Padre:	A10 Madre:
	A11 Abuelo paterno:	A13 Abuelo materno:
	A12 Abuela paterna:	A14 Abuela materna:
¿Qué apellidos tiene o tenía su...	A15 Madre:	A16 Padre:
	A17 Abuela materna:	A19 Abuelo materno:
	A18 Abuela paterna:	A20 Abuelo paterno:

B1 ¿Cuáles de los siguientes describe su origen étnico?

Chileno-Hispano	
Mapuche	
Aymara/Quechua	
Pascuense (Isla de Pascua)	
Europeo	
Otro (especifique)	

C1 ¿En qué ciudad o pueblo vive usted actualmente? \_\_\_\_\_

C2 ¿A qué comuna pertenece? \_\_\_\_\_



BIOBANCO DE TUMORES Y FLUIDOS  
DE LA UNIVERSIDAD DE CHILE

D1 Estado Civil

Soltero (a)	
Conviviente	
Casado(a)	
Separado (a)	
Divorciado (a)	
Viudo (a)	

E1 Nivel de educación

Sin escolarización formal	
Escuela primaria incompleta	
Escuela primaria completa	
Escuela secundaria incompleta	
Escuela secundaria completa	
Estudios Técnicos incompletos	
Estudios Técnicos completos	
Estudios Universitarios incompletos	
Estudios Universitarios completos	
Estudios de postgrado	
No sabe/no contesta	

F1 Ocupación

Profesionales, Técnicos y Afines	
Gerentes, Administradores y Directivos	
Empleados oficina y afines	
Vendedores y afines	
Agricultores, Ganaderos, Pescadores	
Conductores y afines	
Artesanos y Operarios	
Otros Artesanos y Operarios	
Obreros y Jornaleros	
Trabajadores en Servicios Personales	
Dueña de casa	
Jubilado	
Otros trabajadores	

¿Trabaja ó trabajó...

F2 en Industria metalúrgica? No \_\_\_  
Si \_\_\_ ¿Durante cuánto tiempo? \_\_\_\_\_ años

F3 en Industria de plásticos? No \_\_\_  
Si \_\_\_ ¿Durante cuánto tiempo? \_\_\_\_\_ años

F4 con pesticidas? No \_\_\_  
Si \_\_\_ ¿Durante cuánto tiempo? \_\_\_\_\_ años  
¿Cuál era/es el pesticida? \_\_\_\_\_



**PREGUNTAS ASOCIADAS AL ESTILO DE VIDA**

G1 ¿Consumes alcohol?

No \_\_\_\_\_

Si \_\_\_\_\_

¿Qué tipo de alcohol? Vino \_\_\_ Cerveza \_\_\_ Pisco \_\_\_ Ron \_\_\_ Otros (especifique) \_\_\_\_\_

Antes consumía, ya no consumo \_\_\_\_\_

¿Qué tipo de alcohol? Vino \_\_\_ Cerveza \_\_\_ Pisco \_\_\_ Ron \_\_\_ Otros (especifique) \_\_\_\_\_

¿De qué edad hasta que edad? Desde los \_\_\_\_\_ años hasta los \_\_\_\_\_ años

G2 ¿Cuántos días por semana consume o consumía, incluyendo días del fin de semana?

1 día \_\_\_\_\_ 2 días \_\_\_\_\_ 3 a 4 días \_\_\_\_\_ 5 días o más \_\_\_\_\_

G3 ¿Cuántos vasos cada día?

1 vaso \_\_\_\_\_ 2 vasos \_\_\_\_\_ 3 vasos \_\_\_\_\_ 4 vasos o más \_\_\_\_\_

H1 ¿Actualmente, fuma usted cigarrillos?

Sí, uno o más cigarrillos al día \_\_\_\_\_

¿A qué edad comenzó a fumar? A los \_\_\_\_\_ años

Como promedio, ¿cuántos cigarrillos fumó al día el último mes? \_\_\_\_\_ cigarrillos (1 paquete=20 cigarrillos)

Sí, ocasionalmente /menos de un cigarrillo al día \_\_\_\_\_

¿A qué edad comenzó a fumar? A los \_\_\_\_\_ años

Como promedio, ¿cuántos cigarrillos fumó durante el último mes? \_\_\_\_\_ cigarrillos

No, dejé de fumar \_\_\_\_\_

¿Qué edad tenía cuando empezó a fumar? \_\_\_\_\_ años

¿Qué edad tenía cuando dejó de fumar? \_\_\_\_\_ años

Como promedio, ¿cuántos cigarrillos fumaba al día? \_\_\_\_\_ cigarrillos

No, nunca he fumado \_\_\_\_\_

H2 ¿Alguna persona fuma habitualmente dentro de su vivienda?

Si \_\_\_\_\_ No \_\_\_\_\_

H3 ANTES DE LOS ULTIMOS TRES AÑOS, en el hogar, trabajo, escuela, restaurantes y cualquier otro sitio, ¿cuántas veces al día, semana o mes comió usted ?

Ají verdes frescos

Nunca \_\_\_\_\_ veces por \_\_\_\_\_ día \_\_\_\_\_ semana \_\_\_\_\_ mes \_\_\_\_\_

Ají rojos frescos

Nunca \_\_\_\_\_ veces por \_\_\_\_\_ día \_\_\_\_\_ semana \_\_\_\_\_ mes \_\_\_\_\_

Ají rojo en polvo

Nunca \_\_\_\_\_ veces por \_\_\_\_\_ día \_\_\_\_\_ semana \_\_\_\_\_ mes \_\_\_\_\_

Ají rojo en pasta

Nunca \_\_\_\_\_ veces por \_\_\_\_\_ día \_\_\_\_\_ semana \_\_\_\_\_ mes \_\_\_\_\_

**ANTECEDENTES DE SALUD Y SALUD FAMILIAR**

11 ¿Le han diagnosticado cálculos en la vesícula?

No

Si

¿A qué edad? Cuando tenía \_\_\_\_\_ años

¿Cómo le diagnosticaron los cálculos biliares? Una máquina (rayos X/eco) \_\_\_\_\_ Examen físico \_\_\_\_\_

¿Alguna vez ha sido hospitalizado por esta condición? Si  No

¿Cuántas veces estuvo hospitalizado por esta condición? \_\_\_\_\_ veces

12 ¿Ha tenido usted algún otro cáncer?

No

Si

¿Que tipo de cáncer? \_\_\_\_\_

¿A qué edad se lo diagnosticaron? Cuando tenía \_\_\_\_\_ años

13 ¿Ha sido diagnosticado con alguna de las siguientes condiciones?

	Si	¿Qué edad tenía cuando le diagnosticaron?	¿Cuántas veces fue hospitalizado?	¿Qué tipo de tratamiento recibió o recibe?
Pólipos en la vesícula biliar				
Otros problemas de la vesícula biliar (arenilla,...) Tipo: _____				
Hipertensión (presión alta)				
Diabetes (azúcar alta en la sangre)				
Colesterol alto				
Insuficiencia renal				
Epilepsia				
Cardiopatías				
Tuberculosis				
Gastritis				
Cirrosis hepática				
Fiebre tifoidea/paratifoidea				
Apendicitis				
Enfermedad autoinmune Tipo: _____				
Artrosis				
Síndrome de colon irritable				
Alérgias Desencadenante: _____ Desencadenante: _____ Desencadenante: _____				
Enfermedad a la tiroides Tipo: _____				
Otras enfermedades _____ _____ _____				

14 ¿A algún familiar le han diagnosticado problemas (cálculos o piedras, pólipos, arenilla) en la vesícula?

No \_\_\_\_\_

Si \_\_\_\_\_

¿Qué, a quién se lo diagnosticaron, y a qué edad?

Madre biológica \_\_\_\_\_ A los \_\_\_\_\_ años Cálculos\_Pólipos\_Arenilla\_Otros (especifique) \_\_\_\_\_

Padre biológico \_\_\_\_\_ A los \_\_\_\_\_ años Cálculos\_Pólipos\_Arenilla\_Otros (especifique) \_\_\_\_\_

Hermanos(as) \_\_\_\_\_ A los \_\_\_\_\_ años Cálculos\_Pólipos\_Arenilla\_Otros (especifique) \_\_\_\_\_

Hijos(as) biológicos(as) \_\_\_\_\_ A los \_\_\_\_\_ años Cálculos\_Pólipos\_Arenilla\_Otros (especifique) \_\_\_\_\_

Otros familiares (especifique) \_\_\_\_\_ A los \_\_\_\_\_ años Cálculos\_Pólipos\_Arenilla\_Otros (especifique) \_\_\_\_\_

15 ¿A algún familiar le han diagnosticado cáncer de vesícula?

No \_\_\_\_\_

Si \_\_\_\_\_

¿A quién o quienes se lo diagnosticaron y a qué edad?

Madre biológica \_\_\_\_\_ A los \_\_\_\_\_ años

Padre biológico \_\_\_\_\_ A los \_\_\_\_\_ años

Hermanos(as) \_\_\_\_\_ A los \_\_\_\_\_ años

Hijos(as) biológicos(as) \_\_\_\_\_ A los \_\_\_\_\_ años

Otros familiares (especifique) \_\_\_\_\_ A los \_\_\_\_\_ años

(En caso de familias con varios miembros con cáncer de vesícula, el encuestador dibuja el árbol genealógico del entrevistado con los individuos afectados, y solicita permiso para un contacto telefónico posterior)

El encuestado consiente un contacto telefónico posterior Si \_\_\_\_\_ No \_\_\_\_\_

16 ¿A algún familiar le han diagnosticado diabetes?

No \_\_\_\_\_

Si \_\_\_\_\_

¿A quién o quienes se lo diagnosticaron y a qué edad?

Madre biológica \_\_\_\_\_ A los \_\_\_\_\_ años

Padre biológico \_\_\_\_\_ A los \_\_\_\_\_ años

Hermanos(as) \_\_\_\_\_ A los \_\_\_\_\_ años

Hijos(as) biológicos(as) \_\_\_\_\_ A los \_\_\_\_\_ años

Otros familiares (especifique) \_\_\_\_\_ A los \_\_\_\_\_ años

17 ¿A algún familiar le han diagnosticado cáncer de hígado/hepático?

No \_\_\_\_\_

Si \_\_\_\_\_

¿A quién o quienes se lo diagnosticaron y a qué edad?

Madre biológica \_\_\_\_\_ A los \_\_\_\_\_ años

Padre biológico \_\_\_\_\_ A los \_\_\_\_\_ años

Hermanos(as) \_\_\_\_\_ A los \_\_\_\_\_ años

Hijos(as) biológicos(as) \_\_\_\_\_ A los \_\_\_\_\_ años

Otros familiares (especifique) \_\_\_\_\_ A los \_\_\_\_\_ años

18 ¿Han sido diagnosticados otros tipos de cáncer en su familia? ¿Qué tipos de cáncer y a qué familiares?

Madre biológica \_\_\_\_\_ Tipo(s) de cáncer \_\_\_\_\_ A los \_\_\_\_\_ años

Padre biológico \_\_\_\_\_ Tipo(s) de cáncer \_\_\_\_\_ A los \_\_\_\_\_ años

Hermanos(as) \_\_\_\_\_ Tipo(s) de cáncer \_\_\_\_\_ A los \_\_\_\_\_ años

Hijos(as) biológicos(as) \_\_\_\_\_ Tipo(s) de cáncer \_\_\_\_\_ A los \_\_\_\_\_ años

Otros familiares (especifique) \_\_\_\_\_ Tipo(s) de cáncer \_\_\_\_\_ A los \_\_\_\_\_ años

J1 ¿Cual era su peso cuando fué diagnosticada/o con cálculos biliares/cáncer de vesícula? \_\_\_\_\_ kgs

J2 ¿Ha experimentado un cambio de peso en los dos últimos años?

No \_\_\_\_\_

Si \_\_\_\_\_ ¿Cuánto ha cambiado su peso en los últimos dos años? \_\_\_\_\_ kgs Subida \_\_\_\_\_ Bajada \_\_\_\_\_

J3 Talla \_\_\_\_\_ cms

IMC \_\_\_\_\_ (Peso/Talla<sup>2</sup>)

(Si la paciente tiene un IMC mayor o igual a 25) ¿Recuerda desde qué edad tiene exceso de peso?

Desde siempre \_\_\_\_\_ Desde que tenía \_\_\_\_\_ años

(El entrevistador mide a la entrevistada)

J4 Circunferencia de cintura \_\_\_\_\_ cms

J5 Circunferencia de cadera \_\_\_\_\_ cms

J7 ¿Realiza algún tipo de actividad física?

No \_\_\_\_\_

Si \_\_\_\_\_

¿Qué tipo de actividad/es física/s? \_\_\_\_\_

¿Cuántas veces por semana y por cuánto tiempo cada vez? \_\_\_\_\_ días por semana \_\_\_\_\_ horas esos días

¿Desde hace cuánto tiempo realiza esta actividad física? Desde que tenía \_\_\_\_\_ años

#### **SOLO PARA MUJERES**

K1 ¿A qué edad tuvo su primera regla o menstruación? A los \_\_\_\_\_ años

K2 ¿Cual fue el día de inicio de su última regla?

No estoy mestruando \_\_\_\_\_

Mi último período empezó el \_\_\_\_/\_\_\_\_/\_\_\_\_

K3 ¿Usa usted o ha usado anticonceptivos ó píldoras anticonceptivas (píldora, inyección, parche,...)

Uso actualmente \_\_\_\_\_ Empecé a los \_\_\_\_\_ años Tipo de anticonceptivo \_\_\_\_\_

Usé, pero ya no \_\_\_\_\_ Empecé a los \_\_\_\_\_ años y dejé a los \_\_\_\_\_ años Tipo de anticonceptivo \_\_\_\_\_

Nunca he usado \_\_\_\_\_

K4 ¿Tuvo en alguno de sus embarazos cálculos biliares? En \_\_\_\_\_ embarazos

K5 ¿Usted usa o ha usado tratamiento de reemplazo hormonal (dar ejemplos)

Uso actualmente \_\_\_\_\_ Empecé a los \_\_\_\_\_ años Tipo de tratamiento \_\_\_\_\_

Usé, pero ya no \_\_\_\_\_ Empecé a los \_\_\_\_\_ años y dejé a los \_\_\_\_\_ años Tipo de tratamiento \_\_\_\_\_

Nunca he usado \_\_\_\_\_

Muchísimas gracias por su participación. ¿Tiene alguna pregunta? ¿Quiere que anote algún dato que usted considere importante?

#### 4. Certificado de Aprobación Ética



UNIVERSIDAD DE CHILE - FACULTAD DE MEDICINA  
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES HUMANOS

1/2



### ACTA DE APROBACIÓN DE PROYECTO

FECHA: 11 OCT. 2012

11 OCT. 2012

**PROYECTO: GENOME-WIDE INVESTIGATION OF THE INHERITED GENETIC AND EPIGENETIC DIVERSITY OF CHILEAN WOMEN AFFECTED BY GALLBLADDER CANCER AND GALLSTONE DISEASE**

**INVESTIGADOR RESPONSABLE: DRA. KATHERINE MARCELAIN**

**INSTITUCIÓN: PROGRAMA DE GENÉTICA HUMANA, FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE.**

Con fecha 09 de octubre de 2012, el proyecto ha sido analizado a la luz de los postulados de la Declaración de Helsinki, de la Guía Internacional de Ética para la Investigación Biomédica que involucra sujetos humanos CIOMS 1992, y de las Guías de Buena Práctica Clínica de ICH 1996.

Sobre la base de la información proporcionada en el texto del proyecto el Comité de Ética de la Investigación en Seres Humanos de la Facultad de Medicina de la Universidad de Chile, estima que el estudio propuesto está bien justificado y que no significa para los sujetos involucrados riesgos físicos, psíquicos o sociales mayores que mínimos.

Este comité también analizó y aprobó el correspondiente documento de Consentimiento Informado en su versión original con fecha 24 de septiembre de 2012. Se adjunta aprobación de los siguientes documentos: Consentimiento Informado.

En virtud de las consideraciones anteriores el Comité otorga la aprobación ética para la realización del estudio propuesto, dentro de las especificaciones del protocolo.

*Teléfono: 9786923 Fax: 9786189 Email: ceiha@med.uchile.cl*



**INTEGRANTES DEL COMITÉ DE ÉTICA DE LA INVESTIGACIÓN  
EN SERES HUMANOS**

11 OCT 2012

<b>NOMBRE</b>	<b>CARGO</b>	<b>RELACIÓN CON LA INSTITUCIÓN</b>
Dr. Manuel Oyarzún	Presidente	Sí
Sra. Marianne Gaudlitz	Vicepresidente	Sí
Dr. Hugo Amigo	Miembro	Sí
Dr. Leandro Biagini	Miembro	Sí
Dra. Lucia Cifuentes	Miembro	Sí
Sra. Nina Horwitz	Miembro	Sí
Dra. María Eugenia Pinto	Miembro	Sí
Sra. Claudia Marshall	Miembro	No

Santiago, 10 de octubre de 2012.

  
**Dr. Manuel Oyarzún G.**  
**Profesor Titular y Presidente CEISH**

MOG/mva  
c.c: Archivo Proy. N°123-2012.

## 5. Consentimiento Informado

*Consentimiento Informado – Hospital Clínico Universidad de Chile  
Versión 2.0- 8.02.16*



### CONSENTIMIENTO INFORMADO

#### TITULO DEL ESTUDIO

Identificación de biomarcadores para una prevención y tratamiento personalizados del Cáncer de Vesícula Biliar.

#### CONSENTIMIENTO INFORMADO

- Nombre de investigador responsable en Chile: Dra. Katherine Marcelain
- Institución: Departamento de Oncología Básico Clínica, Facultad de Medicina, Universidad de Chile
- Número de contacto: (2) 2 978 9562

Le estamos invitando a participar en el proyecto “Identificación de biomarcadores para una prevención y tratamiento personalizados del Cáncer de Vesícula Biliar”.

#### Antecedentes del Estudio

El cáncer de vesícula biliar es muy poco común en la mayoría de los países del mundo. Sin embargo, en Chile la frecuencia y la mortalidad de este cáncer son extraordinariamente altas. En la población mapuche, la mortalidad por cáncer de vesícula biliar es la más alta a nivel mundial.

Esta alta mortalidad afecta preferentemente a mujeres. En Chile el cáncer de vesícula biliar es el tumor que mata a más mujeres, después del cáncer de mama.

Dada la alta frecuencia de cálculos biliares y cáncer de vesícula en nuestro país, es posible que existan características genéticas propias de la población chilena que aumenten la probabilidad de desarrollar esta enfermedad. El fin último del estudio en que le invitamos a participar consiste en identificar los determinantes genéticos del cáncer de vesícula biliar para una mejor prevención y tratamiento de estas enfermedades. Esto lo realizaremos a través de proyectos de investigación, tanto el presente como otros futuros, empleando las entrevistas y las muestras que le están siendo solicitadas.

Junto con nuestros colaboradores internacionales, realizaremos estudios genéticos de las muestras de ADN recolectadas en este estudio, y compararemos los perfiles genéticos de nuestra población con los perfiles de otras poblaciones con una baja frecuencia de cáncer de vesícula biliar, por ejemplo Alemania. Para realizar estos estudios necesitamos obtener y estudiar muestras de 100 individuos diagnosticados con cáncer de vesícula biliar y con ellas identificar las características que su enfermedad presenta.

#### Procedimiento

Si usted está de acuerdo en participar en este estudio, se le solicitará donar 20cc (cuatro cucharaditas) de sangre. Esta cantidad de sangre es relativamente pequeña y será extraída de manera segura por personal calificado. También le solicitamos responder algunas preguntas que pudieran ser relevantes para el estudio presente, en una entrevista que será realizada por alguna persona de confianza que el investigador designe. El procedimiento completo durará aproximadamente 45 minutos y se podrá realizar, ya sea en las dependencias del Hospital o en su domicilio, dependiendo de cada caso. También le pedimos permiso para, en caso necesario, re-examinar su información médica y confirmar su diagnóstico y, si es posible, utilizar restos de tejido tumoral que se encuentren almacenados en el Servicio de Anatomía Patológica del Hospital en que se



realizó el diagnóstico. De las muestras que usted donó se obtendrá el material genético y se medirán diversos marcadores genéticos. Se evaluarán también sus antecedentes familiares de enfermedades de vesícula biliar y en el caso de reportarse algún caso de cáncer de vesícula biliar, se incorporará a todos los familiares de primer grado que deseen participar, independientemente si han presentado alguna enfermedad a la vesícula o no.

Dado que la investigación avanza rápidamente y se generan nuevas preguntas respecto al cáncer, los investigadores le solicitan (a través de este documento) su autorización para almacenar sus muestras de forma indefinida o hasta que sean utilizadas en su totalidad. Sus muestras serán empleadas sólo para investigación y no serán vendidas. Es substancial destacar que su participación es esencial para identificar los rasgos genéticos que contribuyen al desarrollo de cáncer de vesícula, y que el objetivo final de este estudio es prevenir y tratar mejor esta enfermedad.

Si su muestra fuese de utilidad para una investigación distinta de la descrita en este documento, para poder utilizarla los investigadores deberán obtener una aprobación del Comité de Ética de la Institución en que se realizase dicho estudio, previo a cualquier tipo de análisis.

#### **Confidencialidad**

Se respetará la confidencialidad de todos sus datos en todo momento. No se liberará ni publicará información alguna relacionada con su identidad. Su identidad tampoco será entregada en ningún informe sobre este estudio. Cualquier tipo de muestra o información que se facilite para investigación será codificada y anónima. La información se mantendrá asegurada en gabinetes especiales (de acceso restringido) y en archivos computacionales con claves de seguridad.

Cualquier dato o muestra que se envíe fuera de Chile no contendrá su nombre ni ningún tipo de información personal que permita identificarle, y se realizará cumpliendo todas las exigencias que la ley chilena establece. Al firmar este consentimiento, usted autoriza que se transfiera información derivada de su muestra hacia organizaciones que se encuentran fuera de Chile.

**Riesgos del Procedimiento** Los riesgos asociados con la extracción de sangre son leves, pero pueden incluir dolor, hinchazón, sangrado excesivo, desmayos o sensación de mareo, moratones, infección (un leve riesgo cada vez que se rompe la piel) y múltiples punciones para localizar las venas.

**Costos:** La realización de entrevista y toma de muestra no tienen ningún costo para usted. En caso de que usted incurra en gastos de desplazamiento para la realización del procedimiento, este le será reembolsado presentando el comprobante correspondiente.

**Beneficios:** Este estudio no le proporcionará ningún beneficio personal, pero podrá ser útil para otras personas en el diagnóstico, pronóstico y tratamiento del cáncer de vesícula biliar.

**Alternativas:** Usted tiene la alternativa de participar o no en este estudio. Si usted decide no participar no habrá consecuencias negativas para usted y recibirá el mismo trato que se aplica habitualmente a los pacientes que concurren a este Servicio de Salud.



**Compensación:** Usted no recibirá ninguna compensación económica por su participación en este estudio.

**Voluntariedad:** Su participación en esta investigación es VOLUNTARIA y se puede retirar en cualquier momento, comunicándole su decisión al Investigador Responsable. Si usted decide retirarse del estudio, las muestras que usted donó serán eliminadas siguiendo las normas de bioseguridad correspondientes.

**Derecho del Participante:** Como participante de este estudio, usted tiene derecho a conocer los detalles del proyecto de investigación y sus resultados globales.

**Contacto:** Si usted tiene cualquier consulta o desea más información respecto a este estudio, puede contactar a la Dra. Katherine Marcelain Cubillos (Teléfono: 2 2978 9562, [kmarcelain@med.uchile.cl](mailto:kmarcelain@med.uchile.cl)). Si usted considera amagados los derechos que le asisten como sujeto de investigación, puede contactar al comité de Ética de la Facultad de Medicina de la Universidad de Chile. Departamento de Bioética y Humanidades Médicas, Facultad de Medicina, Universidad de Chile (Teléfono: 2 2978 6189).

Después de haber recibido y comprendido la información de este documento, y haber podido aclarar mis dudas, otorgo mi consentimiento para participar en el proyecto **Identificación de biomarcadores para una prevención y tratamiento personalizados del Cáncer de Vesícula Biliar.**



## 6. Estructura de los datos

Para cada combinación posible, se calculó la frecuencia de individuos con el genotipo correspondiente (homocigotos para el alelo de referencia o el alternativo, heterocigotos y con genotipos faltantes). Por ejemplo: para el SNP rs10946430 y la combinación "No afectado \* Hombre \* ≤ 50 años \* Cluster 1 de ascendencia mapuche" se encontraron los genotipos de 216 individuos distribuidos así: 65 homocigotos al alelo de referencia, 110 heterocigotos, 41 homocigotos al alternativo y 0 genotipos faltantes [calculados como: 216 - (65 + 110 + 41) = 0 faltantes] (**Tabla Suplementaria 3**).

**Tabla Suplementaria 3.** Estructura y agrupación de la información genotípica de 1.161 SNPs en 2.145 individuos, con base en su fenotipo, edad, sexo y proporción de ascendencia mapuche.

Cr: Cromosoma, CVB: Cáncer de vesícula biliar, HLT: Sano, Young: ≤ 50 años, Old: >50 años, bp: pares de bases. REF: Alelo de referencia. ALT: Alelo alternativo. MAP: Grupo de ascendencia mapuche. homREF: Homocigoto al alelo de referencia. het: Heterocigoto. homALT: Homocigoto al alelo alternativo. Miss: Genotipo faltante.

Cr	SNP ID	bp	REF	ALT	Fenotipo	Sexo	Edad	MAP	homREF	het	homALT	miss
6	rs10946430	21124521	T	C	CVB	mujer	old	1	1	2	1	0
6	rs10946430	21124521	T	C	CVB	mujer	old	2	5	6	0	0
6	rs10946430	21124521	T	C	CVB	mujer	young	3	5	14	2	0
6	rs10946430	21124521	T	C	CVB	mujer	old	4	6	7	4	0
6	rs10946430	21124521	T	C	HLT	hombre	young	1	65	110	41	0
6	rs10946430	21124521	T	C	HLT	hombre	old	2	83	118	59	0
...	...	...	...	...	...	...	...	...	...	...	...	...

Una vez realizado el control de calidad, los datos fueron transformados nuevamente: los recuentos totales de individuos que compartían las mismas características, por cada marcador, se separaron en observaciones por individuo por marcador (**Tabla Suplementaria 4**).

**Tabla Suplementaria 4.** Datos transformados desde recuentos totales de individuos agrupados (Tabla 12) por SNP, edad, sexo, fenotipo y proporción de ascendencia mapuche, a observaciones separadas que representan a cada individuo con sus características y genotipo para cada marcador.

MAP: Grupo de ascendencia mapuche. \* 0: Homocigoto al alelo de referencia. 1: Heterocigoto. 2: Homocigoto al alelo alternativo.

Individuo	SNP ID	Fenotipo	Sexo	Edad	MAP	Genotipo *
1	rs912211	CVB	hombre	young	3	0
2	rs912211	CVB	hombre	young	3	0
3	rs912211	CVB	hombre	young	3	0
4	rs912211	CVB	hombre	old	2	0
5	rs912211	CVB	hombre	old	2	1
6	rs912211	CVB	hombre	old	2	1
7	rs912211	CVB	hombre	old	2	0
8	rs912211	CVB	hombre	old	3	2
...	...	...	...	...	...	...

## 7. Código Fuente (Mendel)

Con el software MENDEL se realizaron los análisis de ligamiento paramétrico, no paramétrico y multipuntos, así como también los análisis de sensibilidad en 7 familias chilenas con múltiples casos de CVB.

### 7.1. Código Fuente 1. Análisis de Ligamiento Paramétrico en 7 familias Chilenas con Agregación del CVB

La opción de análisis 2 del programa MENDEL se usó para buscar el locus hipotético del rasgo, el cual se desliza a lo largo del mapa de marcadores, buscando el punto que maximice la probabilidad conjunta del rasgo y los marcadores.

# Los siguientes archivos de entrada son requeridos para el análisis de LP:  
# - Archivo Control: determina los valores de los parámetros que Mendel ejecutará.  
# - Archivo de Definición: Define los loci genéticos, los rasgos cualitativos (factores) o variables  
# cuantitativas a analizar.  
# - Archivo Mapa: especifica los loci genéticos que se analizarán y las distancias entre ellos.  
# - Archivo Pedigrí: Datos específicos de cada individuo. (preparados con el Código Fuente 4 en R).  
# MENDEL arroja dos archivos de salida (output y summary); el primero describe y detalla cada paso del análisis, incluye un reporte de los archivos de entrada e incluye posibles errores de procesamiento, mientras que el segundo resume los hallazgos del análisis.

Archivo Control:

```
! LP_Control_Cr1.txt

! Input Files
DEFINITION_FILE = LP_DefinfileCr1.txt
MAP_FILE = LP_MapfileCr1.txt
PEDIGREE_FILE = LP_Cr1.txt

! Output Files
SUMMARY_FILE = LP_Cr1_Summary.txt
OUTPUT_FILE = LP_Cr1_Output.txt
!ECHO = Yes

! General Analysis Parameters
MAX_ADJUSTED_MEIOSES = 30
AFFECTED_LOCUS_OR_FACTOR = GBC
AFFECTED = 2

! Location Score Analysis Options
MAP_DISTANCE_UNITS = BP
```

```

ANALYSIS_OPTION = Location_Scores
PENETRANCE = 0.05 :: A/A
PENETRANCE = 0.95 :: A/B
PENETRANCE = 0.95 :: B/B
TRAVEL = GRID
STANDARD_GRID = TRUE
PROBABILITY_PEDIGREE_LINKED = True
DEVIANCES=TRUE

```

### Archivo de Definición:

```

! LP_DefinfileCr1.txt

GBC, AUTOSOME, 2, 2 ! Locus-name, Type, # of Alleles, # of Phenotypes
  A, 0.9999 ! Allele, Frequency
  B, 0.0001 ! Allele, Frequency
  1 ! Phenotype
  2 ! Phenotype
rs1240708
rs424079
rs6693447
rs10909845
rs2483278
rs1181883
rs7367066
rs4233258
rs12040805
rs966321
!1028 SNPs
!...

```

### Archivo Mapa:

```

! LP_MapfileCr1.txt

!1028 SNPs, posición mapa en pb

rs1240708,1335790
rs424079,2071340
rs6693447,2330190
rs10909845,2876525
rs2483278,3252104
rs1181883,3677933
!...

```

## 7.2. Código Fuente 2. Análisis de Ligamiento No Paramétrico en 7 familias Chilenas con Agregación del CVB

La opción de análisis 4 del programa MENDEL se usó para mapear el gen de susceptibilidad al CVB y los CB, identificando el exceso de alelos compartidos entre las afectadas en las genealogías.

# Los siguientes archivos de entrada son requeridos para el análisis de LP:  
# - Archivo Control: determina los valores de los parámetros que Mendel ejecutará.  
# - Archivo de Definición: Define los loci genéticos, los rasgos cualitativos (factores) o  
# variables  
# cuantitativas a analizar.  
# - Archivo Mapa: especifica los loci genéticos que se analizarán y las distancias entre ellos.  
# - Archivo Pedigrí: Datos específicos de cada individuo. (preparados con el Código Fuente 4  
# en R).  
# MENDEL arroja dos archivos de salida (output y summary); el primero describe y detalla  
cada paso del análisis, incluye un reporte de los archivos de entrada e incluye posibles  
errores de procesamiento, mientras que el segundo resume los hallazgos del análisis.

#### Archivo Control:

```
! LNP_Control_Cr1.txt

!Input Files
DEFINITION_FILE = LNP_DefinfileCr1.txt
MAP_FILE = LNP_MapfileCr1.txt
PEDIGREE_FILE = LNP_Cr1.txt

! Output Files
SUMMARY_FILE = LNP_Cr1_Summary.txt
OUTPUT_FILE = LNP_Cr1_Output.txt
!ECHO = Yes

! General Analysis Parameters
AFFECTED_LOCUS_OR_FACTOR = HEALTH

! Location Score Analysis Options
MAP_DISTANCE_UNITS = BP
ANALYSIS_OPTION = NPL
```

#### Archivo de Definición:

```
! LNP_DefinfileCr1.txt

!1028 SNPs, Cromosoma
!...
rs4658642,1
rs2653158,1
rs6671796,1
rs6692147,1
rs6428914,1
rs10754458,1
rs2153409,1
rs4445429,1
rs12062784,1
rs10924820,1
exm2265098,1
rs4925650,1
rs10802523,1
rs10888267,1
```

```
rs7418152,1
rs4926506,1
HEALTH,Factor,2
0
2
```

#### Archivo Mapa:

```
! LNP_MapfileCr1.txt

!1028 SNPs, posición mapa en pb
!...
rs4658642,244652199
rs2653158,244815439
rs6671796,245358619
rs6692147,245417562
rs6428914,245486835
rs10754458,245774993
rs2153409,245995920
rs4445429,246226566
rs12062784,246733870
rs10924820,246928540
exm2265098,247259684
rs4925650,247584075
rs10802523,247745313
rs10888267,248059423
rs7418152,248550894
rs4926506,249081330
```

### 7.3. Código Fuente 3. Análisis de Ligamiento Paramétrico Multipuntos en 7 familias Chilenas con Agregación del CVB

Mediante la opción de análisis 2 del programa MENDEL se realizó un análisis de puntuación de ubicación en el cromosoma 9 usando todos los marcadores simultáneamente.

```
# Los siguientes archivos de entrada son requeridos para el análisis de LP multipuntos:
# - Archivo Control: determina los valores de los parámetros que Mendel ejecutará.
# - Archivo de Definición: Define los loci genéticos, los rasgos cualitativos (factores) o
# variables
# cuantitativas a analizar.
# - Archivo Mapa: especifica los loci genéticos que se analizarán y las distancias entre ellos.
# - Archivo Pedigrí: Datos específicos de cada individuo. (preparados con el Código Fuente 4
# en R).
# MENDEL arroja dos archivos de salida (output y summary); el primero describe y detalla
cada paso del análisis, incluye un reporte de los archivos de entrada e incluye posibles
errores de procesamiento, mientras que el segundo resume los hallazgos del análisis.
```

#### Archivo Control:

```
! LP-mp_Control_Cr9.txt
```



```
! Input Files
DEFINITION_FILE = LP-mp_DefinfileCr9mp.txt
MAP_FILE = LP-mp_MapfileCr9mp.txt
PEDIGREE_FILE = LP-mp_Cr9.txt
```

```
! Output Files
SUMMARY_FILE = LP-mp_Cr9_Summary.txt
OUTPUT_FILE = LP-mp_Cr9_Output.txt
!ECHO = Yes
```

```
! General Analysis Parameters
MAX_ADJUSTED_MEIOSES = 30
AFFECTED_LOCUS_OR_FACTOR = GBC
AFFECTED = 2
```

```
! Location Score Analysis Options
MAP_DISTANCE_UNITS = BP
NUMBER_OF_MARKERS_INCLUDED=ALL
ANALYSIS_OPTION = Location_Scores
PENETRANCE = 0.05 :: A/A
PENETRANCE = 0.95 :: A/B
PENETRANCE = 0.95 :: B/B
TRAVEL = GRID
STANDARD_GRID = TRUE
```

### Archivo de Definición:

```
! LP-mp_DefinfileCr9mp.txt
```

```
GBC, AUTOSOME, 2, 2 ! Locus-name, Type, # of Alleles, # of Phenotypes
  A, 0.9999 ! Allele, Frequency
  B, 0.0001 ! Allele, Frequency
  1 ! Phenotype
  2 ! Phenotype
rs10760688,9
rs2211079,9
rs2417250,9
rs12342601,9
rs1002665,9
rs2995195,9
rs1352979,9
rs7865219,9
rs4540458,9
rs4742892,9
rs10820559,9
rs4149339,9
exm2262569,9
rs2437817,9
rs2417598,9
rs10820780,9
rs2417682,9
rs7043293,9
rs902143,9
rs784676,9
rs942292,9
rs498924,9
```

rs837999,9  
rs7030851,9  
rs2805888,9  
rs4978813,9  
rs10980011,9  
rs12115781,9  
rs10739300,9

#### Archivo Mapa:

! LP-mp\_MapfileCr9mp.txt

rs10760688,102247368  
rs2211079,103539701  
rs2417250,104108650  
rs12342601,104409985  
rs1002665,104630136  
rs2995195,104782096  
rs1352979,105632182  
rs7865219,105999196  
rs4540458,106344556  
rs4742892,106571202  
rs10820559,106637033  
rs4149339,107545156  
exm2262569,107571241  
rs2437817,107690057  
rs2417598,107782100  
rs10820780,107851745  
rs2417682,108779678  
rs7043293,109065718  
rs902143,109181911  
rs784676,109912752  
rs942292,110076032  
rs498924,110732838  
rs837999,110989319  
rs7030851,111281164  
rs2805888,111880736  
rs4978813,112251563  
rs10980011,112477012  
rs12115781,112672877  
rs10739300,113155541

## 8. Código Fuente (R)

La preparación y transformación de los datos genotípicos, el control de calidad por SNPs, el mapeo fino por asociación, la extracción de la lista de genes contenidos en las regiones de ligamiento identificadas, y las figuras de los análisis fueron realizados con el software R.

### 8.1. Código Fuente 4. Preparación de Archivos de Entrada Necesarios Para el Análisis de Ligamiento Genético en el Programa Mendel

```
#Titulo:      Generación de los archivos preliminares de entrada: "pedigrí", "mapa" y
#             "definición" necesarios en Mendel.
#Fecha de ultima actualización:  20.02.17
#Autor:      Pablo Báez
#Objetivo:   Generar la información base de los archivos de entrada necesarios
#            para un análisis de ligamiento genético en el programa Mendel.
#Descripción: Este Script permite preparar los archivos input necesarios para un análisis
#             de Ligamiento genético en Mendel, los cuales se describen originalmente así:
#             1. Definition file:
#             The definition file is a plain text file that describes the genetic loci,
#             qualitative factors, and quantitative variables in your data set.
#             These entries must appear in the same order in the definition file as their
#             corresponding phenotypes appear in the pedigree file.
#             2. Map file:
#             The map file is a plain text file indicating the genetic loci to be analyzed and
#             the distances between them. The loci in common between the map and definition
#             file are called the model loci and are the only loci used in analysis. The map
#             file should contain at least one locus in common with the definition file. The
#             map file should not contain any factors or quantitative variables.
#             There are two map file types, one listing marker positions and the other listing
#             intermarker distances. Position map files require physical distances.
#             Inter-marker distance map files require genetic distances or recombination
#             fractions. Your choice of distance units informs Mendel which type of map file
#             to expect.
#             3. Pedigree file:
#             The pedigree file is also a simple text file. It names individuals, describes
#             their relationships, and records their phenotypes at all genetic loci, factors,
#             and quantitative variables. Usually names in this file should be eight or
#             fewer characters. Technically they can be longer, but only the first eight
#             characters will be used, so they must be uniquely determined by the first eight.
#             Genotypes and phenotypes at loci and factors can be up to 10 characters long,
#             and variable values up to 64. The default maximum line length in the pedigree
```

# file is 262,144 = 2<sup>18</sup> characters. This value can be overridden using the keyword PEDIGREE\_MAX\_LINE\_LEN in the control file. Comment strings (which recall are initiated by an ! and continue to the end of the line) are allowed on any line after any data values.

# Para mayor detalle de los archivos y su funcionamiento consulte el manual del programa Mendel.

# Los archivos de trabajo fueron enviados por F\_Boekstegers (GEG-IMBI), con formatos .ped y .map, con los siguientes nombres: "26\_families\_pruned.ped " y "26\_families\_pruned.map".

# Descripción original de esta clase de archivos:

# **1. .map (PLINK text fileset variant information file):** Variant information file accompanying a .ped text pedigree + genotype table. Also generated by '--recode rlist'. A text file with no header file, and one line per variant with the following 3-4 fields:

- # -Chromosome code. PLINK 1.9 also permits contig names here, but most older programs do not.
- # -Variant identifier
- # -Position in morgans or centimorgans (optional; also safe to use dummy value of '0')
- # -Base-pair coordinate

# All lines must have the same number of columns (so either no lines contain the morgans/centimorgans column, or all of them do).

# **2. .ped (PLINK/MERLIN/Haploview text pedigree + genotype table)**

# Original standard text format for sample pedigree information and genotype calls. Normally must be accompanied by a .map file; Haploview requires an accompanying .info file instead.

# Loaded with --file, and produced by --recode. Contains no header line, and one line per sample with 2V+6 fields where V is the number of variants. The first six fields are the same as those in a .fam file, one line per sample with the following six fields:

- # 1. Family ID ('FID')
- # 2. Within-family ID ('IID'; cannot be '0')
- # 3. Within-family ID of father ('0' if father isn't in dataset)
- # 4. Within-family ID of mother ('0' if mother isn't in dataset)
- # 5. Sex code ('1' = male, '2' = female, '0' = unknown)
- # 6. Phenotype value ('1' = control, '2' = case, '-9'/'0'/non-numeric = missing data if case/control)

# The seventh and eighth fields are allele calls for the first variant in the .map file ('0' = no call); the 9th and 10th are allele calls for the second variant; and so on.

# If all alleles are single-character, PLINK 1.9 will correctly parse the more compact 'compound genotype' variant of this format, where each genotype call is

```
# represented as a single two-character string. This does not require the use of an
# additional loading flag. You can produce such a file with '--recode compound-
# genotypes'. It is also possible to load .ped files missing some initial fields.
#
# Bibliografía: 1. https://www.genetics.ucla.edu/software/Mendel\_current\_doc.pdf
#                 2. https://www.cog-genomics.org/plink2/formats
#                 3.
# NOTA IMPORTANTE: Lea, analice y trate de entender completamente este script
# antes de ejecutarlo.
```

```
#####
```

```
#####
#           Importar las librerías           #
#####
```

```
library(reshape)
library(readr)
library(data.table)
library(beepr)
library(dplyr)
```

```
#####
#           Directorio de trabajo           #
#####
```

```
# Ubicar el directorio de trabajo e importar los archivos ped y map:
```

```
setwd("~/")
fam <- read.table("26_families_pruned.ped", header = F, sep=" ", colClasses = "character")
map <- read.table("26_families_pruned.map", header = F, sep="\t")
```

```
#####
#           Preparar la tabla de genotipos           #
#####
```

```
# Fusionar los dos archivos en uno solo mediante los siguientes pasos y definiendo
# una estructura determinada:
# 1. Extraer la lista de marcadores (SNPs) en un vector denominado "mapm"
# 2. Extraer en un objeto denominado "gen" las 6 primeras columnas del archivo .ped
# 3. Combinar las columnas 7 y 8 correspondientes cada una al genotipo de la primera
# variante, separandolas mediante "/", y así sucesivamente con los demas marcadores.
# 4. Almacenar la tabla de genotipos como .csv.
```

```

mapm <- as.vector(t(map[,2]))
gen <- fam[,1:6]
for (i in 1 : length(mapm))
{
  gen$dummy = paste(fam[,6+2*i-1], fam[,6+2*i], sep="/")
  gen <- rename(gen, c(dummy = mapm[i]))
}
beep(8)
gen2<-setDT(as.data.frame(t(gen)), keep.rownames = TRUE)[]
colnames(map)<-c("chr", "snpid", "cM", "bp")
colnames(gen2)[1]<-"snpid"
map$snpid<-as.character(map$snpid)
gen2<-merge(map,gen2,by="snpid", all=T)
#cols=c(2:4)
#gen2[,cols]<-apply(gen2[,cols], 2, function(x) as.numeric(x))
gen2<-gen2[order(gen2[,2],gen2[,4]),]
gent<-as.data.frame(t(gen2[9451:9456,]))
gent[7:9456]<-t(gen2[1:9450,])
gent$`9452`<-toupper(as.character(gent$`9452`))

```

#Recodificar las muestras codificadas por el servicio de genotipado con los códigos #internos.

```

CodigosBBLGC <- read_csv("~/Clave Codigos Biobanco-LGC codes.csv", col_names = FALSE)
pb2 <- mutate_all(CodigosBBLGC, funs(toupper))
pb2$X1 == pb2$X2
names(pb2)[1]<-paste("LGCID")
names(pb2)[4]<-paste("ID")
names(pb2)[5]<-paste("FAMILIA")
pb2<-pb2[,-c(2,3)]
FFPED <- read_csv("FFPEDpart1.csv")
FFPED<-FFPED[-c(53:183),]
FFped<-merge(pb2, FFPED, by="ID", all= TRUE)
names(gent)[2]<-paste("LGCID")
genfam <- merge(FFped, gent, by="LGCID",all= TRUE)
write.table(genfam, "Genotypes.csv", sep=",")

```

# Fusione Genotypes.csv con el archivo FFPEDpart2 (que contiene los miembros de las familias que no fueron genotipados, pero que se requieren para formar el archivo pedigrí).

## 8.2. Código Fuente 5. Control de Calidad de 1.161 SNPs Genotificados en Pacientes con CVB e Individuos Sanos

```

#Titulo: Per-marker QC de 1161 SNPs genotipados en 2145 individuos:
# 183 con CVB y 1962 Sanos
#Fecha: 29.09.16
#Autor: Pablo Báez
#Objetivo: Remover los marcadores subóptimos del set de datos a fin de controlar
# falsos positivos y evitar la reducción en la habilidad para identificar
# asociaciones verdaderamente correlacionadas con el riesgo de enfermedad.
#Proyecto: Análisis de ligamiento y asociación para identificar marcadores
# genéticos de riesgo de cáncer de vesícula biliar en Chile.
#Descripción: Este Script permite realizar el control de calidad de los marcadores
# mediante cuatro pasos:
# 1. Se identificaron y removieron los SNPs con un exceso de genotipos
# faltantes (> 5%)
# 2. Se removieron los marcadores con diferente tasa de llamado de variantes
# entre casos y controles (p<0,01)
# 3. Se estimó la frecuencia del alelo menor (MAF) de cada marcador y se
# eliminaron los SNPs con un MAF <0,05
# 4. Se removieron los SNPs con un p <0,001 en el test de equilibrio de
# Hardy–Weinberg entre los controles.
# Para usar el script importe su dataset (en este caso
# baez_genotypes_updated.txt) y siga las instrucciones.
#
# Bibliografía:
# 1. https://www.ncbi.nlm.nih.gov/pubmed/21085122
# 2. https://www.ncbi.nlm.nih.gov/pubmed/21234875
# 3. https://www.ncbi.nlm.nih.gov/pubmed/20718045
# 4. https://cran.r-project.org/web/packages/HardyWeinberg/index.html
#
# NOTA IMPORTANTE: Lea, analice y trate de entender completamente este script
# antes de ejecutarlo.

```

```
#####
```

```
#####
# Descripción del dataset #
#####
```

```

# El dataset base (baez_genotypes_updated.txt) fue construido a partir de los genotipos
# de 2.145 individuos con 1.161 SNPs. Para ello se seleccionaron SNPs dentro de los #rangos:
# Cr9:105-116 Mb y Cr16:0-6 Mb que corresponden a las regiones obtenidas en el #análisis de
# ligamiento genético. (Nota: Los genomas utilizados anteriormente para los #análisis de
# ligamiento fueron excluidos).

```

```

#Se combinaron datos de tres estudios y sólo se mantuvieron los SNP presentes en todos
#los estudios.

```

```

#
# La información del dataset corresponde a:
#   - 1181 hombres, 964 mujeres
#   - 183 con CVB y 1962 Sanos
#   - 1763 menores (<= 50) y 382 mayores (> 50)
#   - La ascendencia Mapuche se agrupo igualmente:
#       1 = ascendencia Mapuche <= 30% (531 individuos)
#       2 = ascendencia Mapuche > 30% y <= 38% (536)
#       3 = ascendencia Mapuche > 38% y <= 44% (546)
#       4 = ascendencia Mapuche > 44% (532)
#   - Para cada SNP se tienen 32 combinaciones posibles:
#       Pheno (2) * sexo (2) * edad (2) * ascendencia Mapuche (4)
#   - Para cada combinación se cuentan los genotipos correspondientes (homocigotas
#     de referencia o alelo alternativo, heterocigotos, genotipos faltantes)
#   NOTA: Un genotipo nunca se cuenta dos veces

```

#Anotaciones:

```

# Chr = Cromosoma
# Snpid = SNP ID
# Cm = distancia genética (vacía)
# Bp = distancia física en pares de bases
# REF = referencia / alelo principal
# ALT = alelo alterno / menor
# Pheno = identificador de respuesta (GBC/HEALTHY)
# Sexo = male / female
# Edad = young / old (young: <= 50)
# MAP = 1/2/3/4 (definiciones ver arriba)
# Hom_REF = número de homocigotos del alelo de referencia
# Het = número de heterocigotos
# Hom_ALT = número de homocigotos de alelo alternativo
# Miss = número de genotipos faltantes

```

# Los genotipos faltantes se contaron de la siguiente manera:

```

#   - Genotipos faltantes = Número de genotipos disponibles para esta combinación de
#   - Identificador - (hom_ALT + het + hom_REF)
# p.ej. Para la combinación HEALTHY * mujer * joven * cluster de ascendencia Mapuche 2
# uno tiene genotipos de 197 individuos
# Entonces, para el SNP rs10946430 tendríamos 197 - (50 + 95 + 52) = 0 genotipos
# faltantes

```

```

#####
#           Importar las librerías           #
#####

```



```

library(plyr)
library(HardyWeinberg)
library(ggplot2)
library(data.table)

#####
#           Directorio de trabajo           #
#####
setwd("~/baez_genotypes")

#####
# Importar y preparar el archivo baez_genotypes #
#####

baez_genotypes <- read.delim("baez_genotypes_updated.txt")

#####
#           Verificar el # de SNPs y de individuos           #
#####

individuals<-baez_genotypes[,c(2,11:14)]
individuals<-ddply(individuals, "snpid", numcolwise(sum)) # 1161 Markers = número obs.
individuals$total<-rowSums(individuals[,2:5]) # 2145 Individuals.

#####
#           SNP QC           #
#####

# Identificar y remover los SNPs con un exceso de genotipos faltantes (> 5%) #

missvalues<-baez_genotypes[,c(2,14)]
summissval<-lapply(split(missvalues, (1:nrow(missvalues) - 1)%/%32),
                    function(data) sum(data$miss))
summissval<-as.data.frame(summissval)
summissval<-as.data.frame(t(summissval))
snpidcol<- as.vector(unique(baez_genotypes[,2]))
rownames(summissval) <- t(snpidcol)
summissval<-as.data.frame(summissval)
rm(snpidcol)
summissval$missrate<- summissval[,1]/2145*100
names(summissval)[1]<-paste("Number_of_Indiv")
missqcsnps<-summissval[which(summissval$Number_of_Indiv > 107),] # 107 = 2145 * 0.05
missqcsnpspercent<-summissval[which(summissval$missrate > 5),]
all(missqcsnps$V1 == missqcsnpspercent$V1)

```

```

all(missqcsnps$missrate == missqcsnpspercent$missrate)
all(missqcsnps == missqcsnpspercent) # 0 SNPs have a high missing call rate
  ## Plot
summissval$missrate<-summissval$Number_of_Indiv/2145
summissval$missrank <- cut(summissval$missrate,
                           breaks=c(0,0.0001, 0.0003, 0.001, 0.005, 0.05) ,
                           right=T,include.lowest = T)

tiff('MissingCallRate.tiff', units="cm", width=14, height=10, res=200)
qplot(x = missrank, data = summissval) +
  labs(title="") +
  theme(aspect.ratio = 1) +
  labs(x=" \nFracción de genotipos faltantes", y="Número de SNPs\n ") +
  scale_x_discrete(labels=c("<0.0001", "<0.001", "<0.005", "<0.05"))+
  theme(axis.text=element_text(size=12),
        axis.title=element_text(size=16,face="bold"))
dev.off()

# No se requiere remover marcadores subóptimos

# Remover marcadores con diferente tasa de llamado de variantes entre casos y controles
(p<0,01) #

Health<-baez_genotypes[which(baez_genotypes$pheno == "HEALTHY"),]
SummaryHealth<-ddply(Health, .(chr, bp, snpid), numcolwise(sum))
NonHealth<-baez_genotypes[which(baez_genotypes$pheno != "HEALTHY"),]
SummaryNonHealth<-ddply(NonHealth, .(chr, bp, snpid), numcolwise(sum))
Health$missSC<- NonHealth$miss
row.names(Health) <- paste(1:18576)
missingvalHvNH<- Health[,c(14:15)]
tttest<-lapply(split(missingvalHvNH,rep(1:1161,each=16)),
               function(data) t.test(data, correct = F, simulate.p.value = T, B = 10000)$p.value)
Health2<-Health[,c(2,10)]
Health2<-ddply(Health2, "snpid", numcolwise(sum))
Health2$MAP<-tttest
Health2<-as.data.frame(Health2)
names(Health2)[2]<- paste("p-val.t.test")
signdifHvsNH<-Health2[which(Health2$"p-val.t.test" <0.01),] # 57 SNPs
genoclean_dif<-baez_genotypes[!(baez_genotypes$snpid %in% signdifHvsNH$snpid),]

# Estimar la frecuencia del alelo menor (MAF) de cada marcador y eliminar los SNPs con un
MAF <0,05 #

MAFestim<-genoclean_dif[,c(2,11:13)]

```

```

MAFestim<-ddply(MAFestim, "snpid", numcolwise(sum))
MAFestim2 <- MAFestim[,-1]
rownames(MAFestim2) <- MAFestim[,1]
names(MAFestim2)[1]<-paste("MM")
names(MAFestim2)[2]<-paste("MN")
names(MAFestim2)[3]<-paste("NN")
MAFestim2<-as.matrix(MAFestim2)
MAFcalc<-maf(MAFestim2)
MAFcalc<-as.data.frame(MAFcalc)
setDT(MAFcalc, keep.rownames = TRUE)[[]]
names(MAFcalc)[1]<-paste("snpid")
#MAFcalc$MAFcalc <- as.numeric(as.character(MAFcalc$MAFcalc))
MAFcalc<-as.data.frame(MAFcalc)
MAFqcsnps<-MAFcalc[which(MAFcalc$MAFcalc < 0.05),] # 79 SNPs have a < 0.05 MAF
#Plot MAF SNPs
MAFplot<-genoclean_dif[,c(1:2,4,11:13)]
MAFplot<-ddply(MAFplot, .(chr, bp, snpid), numcolwise(sum))
MAFplot2<-merge(MAFplot, MAFcalc, by="snpid", sort = F)

tiff('MAF.tiff', units="cm", width=14, height=10, res=200)
ggplot(data = MAFplot2, aes(MAFplot2$MAFcalc)) +
  geom_histogram(col="white", breaks=seq(0,0.5, by=0.05))+
  labs(title="", x="\nFrecuencia del Alelo menor (MAF)", y="Número de SNPs\n")+
#stat_bin(bins = 10)+
  geom_vline(xintercept=0.05, linetype = "dashed", colour = "red", size=0.8) +
  theme(axis.text=element_text(size=16),
        axis.title=element_text(size=18,face="bold"))
dev.off()
# Remover los SNPs < 0.05 MAF:
genoclean_dif_MAF<-genoclean_dif[!(genoclean_dif$snpid
                                %in%MAFqcsnps$snpid),]
totalsnps<-ddply(genoclean_dif_MAF, .(chr, bp, snpid), numcolwise(sum))
# Quedan 1030 SNPs

# Remover los SNPs con un p <0,001 en el test de equilibrio de Hardy–Weinberg entre los
controles #

HWpheno<-genoclean_dif_MAF[,c(2,7,11:13)]
HWpheno<-HWpheno[which(HWpheno$pheno == "HEALTHY"),]
HWpheno<-ddply(HWpheno, "snpid", numcolwise(sum))
names(HWpheno)[2]<-paste("MM")
names(HWpheno)[3]<-paste("MN")
names(HWpheno)[4]<-paste("NN")
HWTTest<-HWExactMat(as.matrix(HWpheno[,2:4]), verbose=F)

```

```

HWTest<-as.data.frame(HWTest)
HWpheno$pvalvec<-HWTest$pvalvec
#HW - Ternary Plot
HWfrec<-HWpheno[,c(2:4)]
HWfrec<-as.matrix(HWfrec)
HWfrec[,3]<-HWpheno$NN/HWpheno$total

tiff('TernaryPlot.tiff', units="cm", width=20, height=20, res=200)
HWTernaryPlot(HWfrec, 1962, region=1,vertex.cex=1.5,signifcolour=TRUE, hwcurve=TRUE,
alpha = 0.001, pch=20,
      vertexlab = c("AA", "AB", "BB"), vbounds=T, cex = 1,axislab="\n\nTernary plot de
1030 SNPs \nen 1962 individuos control")
dev.off()
# Remover los SNPs HWE < 0.001:
HWexclude<-HWpheno[which(HWpheno$pvalvec < 0.001),] # 2 SNPs a remover
genoclean_dif_MAF_HW<-genoclean_dif_MAF[!(genoclean_dif_MAF$snpid
      %in%HWexclude$snpid),]
totalsnps<-ddply(genoclean_dif_MAF_HW, .(chr, bp, snpid), numcolwise(sum))
#HW - Ternary Plot clean data
HWphenoclean<-HWpheno[!(HWpheno$snpid %in% HWexclude$snpid),]
HWphenoclean2<-HWphenoclean[,c(2:4)]
HWphenoclean2<-as.matrix(HWphenoclean2)
tiff('TernaryPlotClean.tiff', units="cm", width=20, height=20, res=200)
HWTernaryPlot(HWphenoclean2,1962, region=1,vertex.cex=1.5,signifcolour=TRUE, alpha =
0.001,
      pch=20, vertexlab = c("AA", "AB", "BB"), vbounds=T, cex = 1,
      axislab="\n\nTernary plot de 1030 SNPs \nen 1962 individuos control")
dev.off()

write.table(genoclean_dif_MAF_HW, file = "GenoClean.txt", sep =";")

#####
#           Plots del dataset de SNPs sin QC           #
#####

## Missing call rate> 0.05
SummaryGeno<-baez_genotypes[,c(1:2,4,11:14)]
SummaryGeno<-ddply(SummaryGeno, .(chr, bp, snpid), numcolwise(sum))
SummaryGeno$missrate<- SummaryGeno[,7]/2145
missqcsnps<-SummaryGeno[which(SummaryGeno$missrate > 0.05),] #0 SNPs have a missing
call rate> 0.05
names(SummaryGeno)[4]<-paste("MM")
names(SummaryGeno)[5]<-paste("MN")
names(SummaryGeno)[6]<-paste("NN")

```

```

## MAF < 0.05
MAF<-as.matrix(SummaryGeno[,c(4:6)])
MAFc<-maf(MAF)
MAFc<-as.data.frame(MAFc)
SummaryGeno$MAFc<-MAFc$MAFc
MAFqcsnps<-SummaryGeno[which(SummaryGeno$MAFc < 0.05),] #79 SNPs have a < 0.05
MAF

## HWE <0.001
HWpheno<-baez_genotypes[,c(2,7,11:13)]
HWpheno<-HWpheno[which(HWpheno$pheno == "HEALTHY"),]
HWpheno<-ddply(HWpheno, "snpid", numcolwise(sum))
names(HWpheno)[2]<-paste("MM")
names(HWpheno)[3]<-paste("MN")
names(HWpheno)[4]<-paste("NN")
HWTTest<-HWExactMat(as.matrix(HWpheno[,2:4]), verbose=F)
HWTTest<-as.data.frame(HWTTest)
HWTTest$snpid<-HWpheno$snpid
SummaryGeno<-merge(SummaryGeno, HWTTest, by="snpid", sort = F)
SummaryGeno$Log_Pval= -log10(SummaryGeno[,10])
rm(HWpheno, HWTTest, MAF, MAFc, MAFqcsnps, missqcsnps)
beep(2)
ElimSNPsHWpval<-SummaryGeno[which(SummaryGeno$pvalvec <0.001),] # 3 SNPs
ElimSNPsMAF<-SummaryGeno[which(SummaryGeno$MAFc <0.05),] # 78 SNPs

SummaryGenoplt<-SummaryGeno[which(SummaryGeno$Log_Pval != Inf),]
tiff('plotPval_MAF.tiff', units="cm", width=20, height=20, res=200)
qplot(x = MAFc, y = Log_Pval, data = SummaryGenoplt, size=l(0.4)) +
  labs(title="Conjunto de SNPs sin filtrar\n", x="\nFrecuencia del Alelo Menor (MAF)", y="-
log10(p-value)\n") +
  geom_hline(yintercept=3, linetype = "dashed", colour = "red", size=0.8) +
  geom_vline(xintercept=0.05, linetype = "dashed", colour = "red", size=0.8) +
  theme(axis.text=element_text(size=16),
        axis.title=element_text(size=18,face="bold"))+
  theme(plot.title = element_text(size=18, hjust = 0.5))
dev.off()

tiff('plotPval_MAFZoom.tiff', units="cm", width=20, height=20, res=200)
qplot(x = MAFc, y = Log_Pval, data = SummaryGenoplt, size=l(0.4)) +
  labs(title="", x="\nFrecuencia del Alelo Menor (MAF)", y="-log10(p-value)\n") +
  scale_y_continuous(limits = c(0, 5)) + geom_hline(yintercept=3, linetype = "dashed", colour =
"red", size=0.8) +
  scale_x_continuous(limits = c(0, 0.1)) +

```

```
geom_vline(xintercept=0.05, linetype = "dashed", colour = "red", size=0.7)+
theme(axis.text=element_text(size=16),
      axis.title=element_text(size=18,face="bold"))
dev.off()
```

### 8.3. Código fuente 6. Asociación Genética Mediante Regresión Logística Binaria en Pacientes con CVB e Individuos Sanos

```
#Titulo:      Análisis de asociación genética, mediante regresión logística binaria en
#             pacientes con CVB e individuos sanos.
#Fecha:      29.03.17
#Autor:      Pablo Báez
#Objetivo:   Predecir el resultado de una variable categórica (Fenotipo) en función de
#             las variables independientes o predictoras (Ancestría, Genero y Edad)
#Proyecto:   Análisis de ligamiento y asociación para identificar marcadores
#             genéticos de riesgo de cáncer de vesícula biliar en Chile.
#Descripción: Este Script contiene las ordenes para preparar y ejecutar una
#             logística binaria en R (library(MASS)).
#             Para usar el script importe su dataset (GenoClean.txt, o el archivo bruto,
#             en este caso baez_genotypes.txt) y siga las instrucciones.
#
#Nociones básicas:
#   Dichotomous case/control traits are generally analyzed using either contingency
#   table methods or logistic regression. Contingency table tests examine and
#   measure the deviation from independence that is expected under the null
#   hypothesis that there is no association between the phenotype and genotype
#   classes. The most ubiquitous form of this test is the popular chi-square test
#   (and the related Fisher's exact test).
#   Logistic regression is an extension of linear regression where the outcome of
#   a linear model is transformed using a logistic function that predicts the
#   probability of having case status given a genotype class. Logistic regression
#   is often the preferred approach because it allows for adjustment for clinical
#   covariates (and other factors), and can provide adjusted odds ratios as a
#   measure of effect size. Logistic regression has been extensively developed,
#   and numerous diagnostic procedures are available to aid interpretation of the
#   model.
#
#Bibliografía: 1. https://es.wikipedia.org/wiki/Regresión\_logística
#              2. http://rpubs.com/sinhrks/plot\_lm
#              3. http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/Categor/Tema3Cate.pdf
#              4. http://www.uv.es/lejarza/mcaf/glm.pdf
#              5. http://www.dm.uba.ar/materias/modelos\_lineales\_generalizados\_Mae/2002/1/GLM\_parte1.pdf
```

```

#           6. http://analisisydecision.es/interpretacion-de-los-parametros-de-un-modelo-glm/
#           7. http://data.princeton.edu/R/glms.html
#
#           NOTA IMPORTANTE: Lea, analice y trate de entender completamente este
#           script antes de ejecutarlo.
#####

```

```

#####
#           Descripción del dataset           #
#####

```

```

# El dataset base (baez_genotypes_updated.txt) fue construido a partir de los genotipos
# de 2.145 individuos con 1.161 SNPs. Para ello se seleccionaron SNPs dentro de los rangos:
# Cr9:105-116 Mb y Cr16:0-6 Mb que corresponden a las regiones obtenidas en el análisis
# de ligamiento genético. (Nota: Los genomas utilizados anteriormente para los análisis de
# ligamiento fueron excluidos).
# Se combinaron datos de tres estudios y sólo se mantuvieron los SNP presentes en todos
# los estudios.
# El archivo GenoClean.txt , con el que se trabaja en este script tiene el mismo formato del
# archivo original (baez_genotypes_updated.txt) pero viene con el SNP QC que se realizó
# en el programa de Código Fuente 5 en R.

```

```

#####
#           Importar las librerías           #
#####

```

```

library(readr)
library(MASS)
library(ggplot2)
library(plyr)
library(ggfortify)
library(beepr)
library(qvalue)
library(qqman)
library(RColorBrewer)
library(data.table)

```

```

#####
#           Directorio de trabajo           #
#####

```

```

setwd("~/baez_genotypes")

```

```

#####
#   Importar y preparar el archivo GenoClean   #
#####

GenoClean <- read_csv("GenoClean.txt",
                     col_types = cols(MAP = col_factor(levels = c("1",
                                                                "2", "3", "4"))))
#####
#   Verificar el # de SNPs y de individuos     #
#####

#GenoClean<-GenoClean[,-1]
individuals<-GenoClean[,c(2,11:14)]
individuals<-ddply(individuals, "snpid", numcolwise(sum)) # 1028 Markers = obs. number
individuals$total<-rowSums(individuals[,2:5]) # 2145 Individuals.
rm(individuals)

#####
#           Preparación de los datos           #
#####

# El objetivo principal en este punto es preparar un df que contenga individuos con
# cada una de sus características (genero, ancestría, edad), en lugar del conteo de
# individuos pertenecientes a cada grupo. Para mayor detalle vea el programa Código Fuente
# 5 en R.
GenoClean1 <-GenoClean
# Asignar a cada genotipo un valor de acuerdo a un modelo genético dominante, de la
# siguiente forma:
# Homocigotos al alelo de referencia = 0
# Heterocigotos = 1
# Homocigotos al alternativo = 1
# Missing = 9

names(GenoClean1)[11]<-paste("0")
names(GenoClean1)[12]<-paste("1")
names(GenoClean1)[13]<-paste("1")
names(GenoClean1)[14]<-paste("9")
GenoClean1 <-GenoClean1[, -c(1,3:6)]

# Obtener el vector con los nombres que se usaran para repetir (0, 1, 1, 9):
GenoCats <- tail(names(GenoClean1), 4)
# Construir un data.frame donde cada fila represente un individuos:

```



```

temp <- Map(function(x) GenoClean1[rep(seq_len(nrow(GenoClean1)), x), 1:5],
GenoClean1[GenoCats])
dfNew <- cbind(do.call(rbind, temp),
              "GenoCats"=rep(GenoCats, colSums(GenoClean1[GenoCats])))
beep(2)

#####
#                               Análisis glm                               #
#####

# GBC vs Health #
# Análisis sin covariables #

GBCvsHLT<-dfNew
#GBCvsHLT$pheno<-as.character(GBCvsHLT$pheno)
GBCvsHLT$pheno[GBCvsHLT$pheno == "GBC"] <- 1
GBCvsHLT$pheno[GBCvsHLT$pheno == "HEALTHY"] <- 0
GBCvsHLT<-subset(GBCvsHLT, GenoCats != 9)
GBCvsHLT$GenoCats<-as.numeric(as.character(GBCvsHLT$GenoCats))
GBCvsHLT$pheno<-as.numeric(GBCvsHLT$pheno)
GBCvsHLT$MAP<-as.factor(GBCvsHLT$MAP)
AssoGBC <- dplyr::dply(GBCvsHLT, "snpid", function(df)
  glm(pheno~GenoCats, family = binomial(link=logit), data=df))
beep(2)
summary(AssoGBC[[1]])
anova(AssoGBC[[1]])
Assocpval<-as.data.frame(sapply(AssoGBC, function(f) summary(f)$coefficients[,4]))
beep(2)
Assocpval<-as.data.frame(t(Assocpval))
setDT(Assocpval, keep.rownames = TRUE)[,]
setnames(Assocpval, 1, "snpid")
SNPsInfo<-GenoClean[,c(1:4,6)]
SNPsInfo<-ddply(SNPsInfo, .(chr, bp, snpid), numcolwise(sum))
SNPsInfo<-SNPsInfo[,-4]
glm1Results<-merge(SNPsInfo, Assocpval, by="snpid")
glm1Results<-glm1Results[order(glm1Results$bp),]
glm1Results<-glm1Results[ order(glm1Results[,2], glm1Results[,3]), ]
names(glm1Results)[5]<-paste("P")
glm1Results$logP<- -log(glm1Results$P,10)
#write.csv(glm1Results, file="GBCvsHealth-covALL.csv")
GBCvsHealthnocov<-subset(glm1Results, logP >= 4.3)
write.csv(GBCvsHealthnocov, file="GBCvsHealth-cov.csv")
snps_to_highlight16<-subset(glm1Results, logP >= 4.3 & chr == 16)
snps_to_highlight16<-as.vector(snps_to_highlight16$snpid)

```

```
snps_to_highlight9<-subset(glm1Results, logP >= 4.3 & chr == 9)
snps_to_highlight9<-as.vector(snps_to_highlight9$snpid)
```

```
#Info plots:
```

```
#Suggestive line:
```

```
##- Where to draw a "suggestive" line. Default -log10(1e-5) = 5. Set to FALSE to disable.
```

```
#Genomewide line:
```

```
##- Where to draw a "genome-wide significant" line. Default -log10(5e-8) = 7.30103.
```

```
#Set to FALSE to disable.
```

```
#Suggestive line for this study (established by bonferroni post-test):
```

```
##-log10(4.863813e-05)=4.3
```

```
Azul <- brewer.pal(8, "Accent")
```

```
Red <- brewer.pal(9, "Set1")
```

```
Dots<-brewer.pal(9, "Greys")
```

```
signiflines<-c(Azul[5],Red[1])
```

```
# Plot Cr 9 #
```

```
tiff('Cr9_No-Cov.tiff', units="cm", width=16, height=8, res=200)
manhattan(subset(glm1Results, chr == 9), chr = "chr", bp = "bp", p = "P",
  snp = "snpid", col=Dots[9], cex=0.6, main="Cromosoma 9\n Modelo sin covariables",
  suggestiveline = F, genomewideline = F,
  xlim = c(105000000, 116000000), ylim = c(0, 12), yaxt="n",xaxt="n",
highlight=snps_to_highlight9,
  cex.lab=1.2, mgp=c(2,1,.5), las=0.1, xlab="Posición (Mbp)")
abline(h=c(4.3,7.3), col=signiflines, lwd=1.5, lty=2)
my.axis <-paste(axTicks(1)/1000000)
axis(1,at=axTicks(1), labels=my.axis, xlim = c(105000000, 116000000), cex.axis = 1.2,
font.axis=2)
axis(2,at=axTicks(2), cex.axis = 1.2, ylim = c(0, 12), font.axis=2)
dev.off()
```

```
# Plot Cr 16 #
```

```
tiff('Cr16_No-Cov.tiff', units="cm", width=16, height=8, res=200)
manhattan(subset(glm1Results, chr == 16), chr = "chr", bp = "bp", p = "P",
  snp = "snpid", col=Dots[9], cex=0.6, main="Cromosoma 16\n Modelo sin covariables",
  suggestiveline = F, genomewideline = F,
  xlim = c(0, 60000000), ylim = c(0, 8), yaxt="n",xaxt="n",highlight=snps_to_highlight16,
  cex.lab=1.2, mgp=c(2,1,.5), las=0.1, xlab="Posición (Mbp)")
abline(h=c(4.3,7.3), col=signiflines, lwd=1.5, lty=2)
my.axis <-paste(axTicks(1)/1000000)
axis(1,at=axTicks(1), labels=my.axis, xlim = c(0, 60000000), cex.axis = 1.2, font.axis=2)
```

```
axis(2,at=axTicks(2), cex.axis = 1.2, ylim = c(0, 8),font.axis=2)
dev.off()
```

### # Análisis con covariables #

```
AssoGBCallcov <- dply(GBCvsHLT, "snpid", function(df)
  glm(pheno~age+gender+MAP+GenoCats, family = binomial(link=logit), data=df))
beep(2)
summary(AssoGBCallcov[[1]])
Assocpval2<-as.data.frame(sapply(AssoGBCallcov, function(f) summary(f)$coefficients[,4]))
beep(2)
Assocpval2<-as.data.frame(t(Assocpval2))
setDT(Assocpval2, keep.rownames = TRUE)[[]]
setnames(Assocpval2, 1, "snpid")
glm2Results<-merge(SNPsInfo, Assocpval2, by="snpid")
glm2Results<-glm2Results[order(glm2Results$bp),]
glm2Results<-glm2Results[ order(glm2Results[,2], glm2Results[,3]), ]
names(glm2Results)[10]<-paste("P")
glm2Results$logP<- -log10(glm2Results$P)
#write.csv(glm2Results, file="GBCvsHealth+covALL.csv")
GBCvsHealthsicov<-subset(glm2Results, logP >= 4.3)
write.csv(GBCvsHealthsicov, file="GBCvsHealth+cov.csv")
snps_to_highlight16<-subset(glm2Results, logP >= 4.3 & chr == 16)
snps_to_highlight16<-as.vector(snps_to_highlight16$snpid)
snps_to_highlight9<-subset(glm2Results, logP >= 4.3 & chr == 9)
snps_to_highlight9<-as.vector(snps_to_highlight9$snpid)
```

### # Plot Cr 9 #

```
tiff('Cr9_Plus_Cov.tiff', units="cm", width=16, height=8, res=200)
manhattan(subset(glm2Results, chr == 9), chr = "chr", bp = "bp", p = "P",
  snp = "snpid", col=Dots[9], cex=0.6, main="Cromosoma 9\n Modelo con covariables",
  suggestiveline = F, genomewideline = F,
  xlim = c(105000000, 116000000), ylim = c(0, 12), yaxt="n",xaxt="n",
highlight=snps_to_highlight9,
  cex.lab=1.2, mgp=c(2,1,.5), las=0.1, xlab="Posición (Mbp)")
abline(h=c(4.3,7.3), col=signiflines, lwd=1.5, lty=2)
my.axis <-paste(axTicks(1)/1000000)
axis(1,at=axTicks(1), labels=my.axis, xlim = c(105000000, 116000000), cex.axis = 1.2,
font.axis=2)
axis(2,at=axTicks(2), cex.axis = 1.2, ylim = c(0, 12), font.axis=2)
dev.off()
```

### # Plot Cr 16 #

```
tiff('Cr16_Plus_Cov.tiff', units="cm", width=16, height=8, res=200)
manhattan(subset(glm2Results, chr == 16), chr = "chr", bp = "bp", p = "P",
  snp = "snpid", col=Dots[9] , cex=0.6, main="Cromosoma 16\n Modelo con covariables",
  suggestiveline = F, genomewideline = F,
  xlim = c(0, 6000000), ylim = c(0, 8), yaxt="n",xaxt="n",highlight=snps_to_highlight16,
  cex.lab=1.2, mgp=c(2,1,.5), las=0.1, xlab="Posición (Mbp)")
```

```
abline(h=c(4.3,7.3), col=signiflines, lwd=1.5, lty=2)
my.axis <-paste(axTicks(1)/1000000)
axis(1,at=axTicks(1), labels=my.axis, xlim = c(0, 6000000), cex.axis = 1.2, font.axis=2)
axis(2,at=axTicks(2), cex.axis = 1.2, ylim = c(0, 8),font.axis=2)
dev.off()
```

```
#####
#           Estimación ORs           #
#####
```

```
# GBC vs Health #
# Análisis sin covariables #
```

```
glm1Results$ORsI<-exp(glm1Results$(Intercept))
glm1Results$ORsGc<-exp(glm1Results$P)
intervalos<-as.data.frame(sapply(AssoGBC, function(f) exp(confint(f))))
beep(2)
exp(confint(AssoGBC[[1]], level = 0.95))
exp(confint(AssoGBC[[2]]))
intervalosInterGen<-cbind(as.data.frame(t(intervalos[c(1,3),])),
as.data.frame(t(intervalos[c(2,4),])))
setDT(intervalosInterGen, keep.rownames = TRUE)[]
setnames(intervalosInterGen, 1:5, c("snpid", "I 2.5 %", "I 97.5 %", "G 2.5 %", "G 97.5 %"))
glm1Results<-merge(glm1Results, intervalosInterGen, by="snpid")
glm1ORs<-glm1Results[,c(1:4,7,9,10,5,8,11,12,6)]
glm1ORs[,c(2:12)]<-as.numeric(glm1ORs[,c(2:12)])
glm1ORs$chr<-as.numeric(glm1ORs$chr)
write.csv(glm1ORs, file="OddsRatioGBCvsHealth-cov.csv")
```

## IX. Referencias

1. Ferlay J, S.I., Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. 2012 [cited 2017 December 11]; Available from: <http://globocan.iarc.fr/>.
2. *Departamento de Estadística e Información en Salud. Ministerio de Salud de Chile. Tasa de Tumores Malignos según Sexo, Chile 1997-2013*. 2013. 2013 [cited 2017 December 11]; Available from: <http://www.deis.cl/series-y-graficos-de-mortalidad/>.
3. Tempfer, C.B. and M.A. Reymond, *Sporadic Cancers*, in *The Molecular Basis of Human Cancer*. 2017, Humana Press: New York, NY.
4. Hodgson, S., *Mechanisms of inherited cancer susceptibility*. J Zhejiang Univ Sci B, 2008. **9**(1): p. 1-4.
5. Albright, F., et al., *Significant evidence for a heritable contribution to cancer predisposition: a review of cancer familiarity by site*. BMC Cancer, 2012. **12**: p. 138.
6. Blanco-Gomez, A., et al., *Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes*. Bioessays, 2016. **38**(7): p. 664-73.
7. Sverdlov, E., *Missing heritability: Illegitimate quantitative comparison of holistic and reductionist data (response to DOI 10.1002/bies.201600084)*. Bioessays, 2016. **38**(12): p. 1195-1196.
8. Petersen, G.M., *Familial pancreatic cancer*. Semin Oncol, 2016. **43**(5): p. 548-553.
9. de la Chapelle, A., *Genetic predisposition to colorectal cancer*. Nat Rev Cancer, 2004. **4**(10): p. 769-80.
10. Hundal, R. and E.A. Shaffer, *Gallbladder cancer: epidemiology and outcome*. Clin Epidemiol, 2014. **6**: p. 99-109.
11. Edge, S.B. and American Joint Committee on Cancer., *AJCC cancer staging manual*. 7th ed. 2010, New York: Springer. xiv, 648 p.
12. Silva, V.W., et al., *Biliary carcinomas: pathology and the role of DNA mismatch repair deficiency*. Chin Clin Oncol, 2016. **5**(5): p. 62.
13. Roa, I. and X. de Aretxabala, *Gallbladder cancer in Chile: what have we learned?* Curr Opin Gastroenterol, 2015. **31**(3): p. 269-75.
14. Ferlay, J., et al., *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]*. Lyon, France: International Agency for Research on Cancer; 2013. Available from: <http://globocan.iarc.fr/>, accessed on 03/03/2014., 2012.
15. Castro, F.A., et al., *Biliary tract cancer incidence in the United States-Demographic and temporal variations by anatomic site*. Int J Cancer, 2013. **133**(7): p. 1664-71.
16. Randi, G., et al., *Epidemiology of biliary tract cancers: an update*. Ann Oncol, 2009. **20**(1): p. 146-59.
17. MINISTERIO\_DE\_SALUD, *Guía Clínica COLECISTECTOMÍA PREVENTIVA EN ADULTOS DE 35 A 49 AÑOS*. 2010, Minsal: Santiago.
18. Lorenzo Bermejo, J., et al., *Subtypes of Native American ancestry and leading causes of death: Mapuche ancestry-specific associations with gallbladder cancer risk in Chile*. PLoS Genet, 2017. **13**(5): p. e1006756.

19. Shih, S.P., et al., *Gallbladder cancer: the role of laparoscopy and radical resection*. Ann Surg, 2007. **245**(6): p. 893-901.
20. Wernberg, J.A. and D.D. Lucarelli, *Gallbladder cancer*. Surg Clin North Am, 2014. **94**(2): p. 343-60.
21. Jiao, X., et al., *Upregulated plasma and urinary levels of nucleosides as biological markers in the diagnosis of primary gallbladder cancer*. J Sep Sci, 2014. **37**(21): p. 3033-44.
22. Wang, Y.F., et al., *Combined detection tumor markers for diagnosis and prognosis of gallbladder cancer*. World J Gastroenterol, 2014. **20**(14): p. 4085-92.
23. Rana, S., et al., *Evaluation of CA 242 as a tumor marker in gallbladder cancer*. J Gastrointest Cancer, 2012. **43**(2): p. 267-71.
24. Sharma, P., et al., *Global methylation profiling to identify epigenetic signature of gallbladder cancer and gallstone disease*. Tumour Biol, 2016. **37**(11): p. 14687-14699.
25. Khandelwal, A., et al., *The emerging role of long non-coding RNA in gallbladder cancer pathogenesis*. Biochimie, 2017. **132**: p. 152-160.
26. *AJCC cancer staging manual*. 2016, New York, NY: Springer Science+Business Media. pages cm.
27. Valle, J.W., et al., *New Horizons for Precision Medicine in Biliary Tract Cancers*. Cancer Discov, 2017. **7**(9): p. 943-962.
28. Muller, B.G., X. De Aretxabala, and M. Gonzalez Domingo, *A review of recent data in the treatment of gallbladder cancer: what we know, what we do, and what should be done*. Am Soc Clin Oncol Educ Book, 2014: p. e165-70.
29. Latorre, S.G., et al., *[Coverage of the gallbladder cancer prevention strategy in Chile: results from the 2009-2010 National Health Survey]*. Rev Med Chil, 2015. **143**(2): p. 158-67.
30. Arroyo, G.F., A. Gentile, and L.A. Parada, *Gallbladder cancer: South American experience*. Chin Clin Oncol, 2016. **5**(5): p. 67.
31. Shukla, V.K., et al., *Lifestyle, reproductive factors and risk of gallbladder cancer*. Singapore Med J, 2008. **49**(11): p. 912-5.
32. Pandey, M. and V.K. Shukla, *Lifestyle, parity, menstrual and reproductive factors and risk of gallbladder cancer*. Eur J Cancer Prev, 2003. **12**(4): p. 269-72.
33. Zhang, L.Q., et al., *Potential therapeutic targets for the primary gallbladder carcinoma: estrogen receptors*. Asian Pac J Cancer Prev, 2013. **14**(4): p. 2185-90.
34. Srivastava, A., et al., *Significant role of estrogen and progesterone receptor sequence variants in gallbladder cancer predisposition: a multi-analytical strategy*. PLoS One, 2012. **7**(7): p. e40162.
35. Tazuma, S. and G. Kajiyama, *Carcinogenesis of malignant lesions of the gall bladder. The impact of chronic inflammation and gallstones*. Langenbecks Arch Surg, 2001. **386**(3): p. 224-9.
36. Barreto, S.G., A. Dutt, and A. Chaudhary, *A genetic model for gallbladder carcinogenesis and its dissemination*. Ann Oncol, 2014. **25**(6): p. 1086-97.
37. Mills, J.C. and O.J. Sansom, *Reserve stem cells: Differentiated cells reprogram to fuel repair, metaplasia, and neoplasia in the adult gastrointestinal tract*. Sci Signal, 2015. **8**(385): p. re8.

38. Seretis, C., et al., *Metaplastic changes in chronic cholecystitis: implications for early diagnosis and surgical intervention to prevent the gallbladder metaplasia-dysplasia-carcinoma sequence*. J Clin Med Res, 2014. **6**(1): p. 26-9.
39. Duarte, I., et al., *Metaplasia and precursor lesions of gallbladder carcinoma. Frequency, distribution, and probability of detection in routine histologic samples*. Cancer, 1993. **72**(6): p. 1878-84.
40. Roa, I., et al., *Preneoplastic lesions in gallbladder cancer*. J Surg Oncol, 2006. **93**(8): p. 615-23.
41. Sharma, A., et al., *Gallbladder cancer epidemiology, pathogenesis and molecular genetics: Recent update*. World J Gastroenterol, 2017. **23**(22): p. 3978-3998.
42. Zhu, A.X. and A.F. Hezel, *Development of molecularly targeted therapies in biliary tract cancers: reassessing the challenges and opportunities*. Hepatology, 2011. **53**(2): p. 695-704.
43. Rashid, A., *Cellular and molecular biology of biliary tract cancers*. Surg Oncol Clin N Am, 2002. **11**(4): p. 995-1009.
44. Roa E, I., et al., *Mutación del gen p53 en el cáncer de la vesícula biliar*. Revista médica de Chile, 2000. **128**: p. 251-258.
45. Roa S, J.C., et al., *Mutación del gen K-ras en el cáncer de la vesícula biliar*. Revista médica de Chile, 2004. **132**: p. 955-960.
46. CASTILLO A, J., P. GARCÍA M, and J.C. ROA S, *Alteraciones genéticas en lesiones preneoplásicas y neoplásicas de la vesícula biliar*. Revista médica de Chile, 2010. **138**: p. 595-604.
47. Wistuba, II and A.F. Gazdar, *Gallbladder cancer: lessons from a rare tumour*. Nat Rev Cancer, 2004. **4**(9): p. 695-706.
48. Li, M., et al., *Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway*. Nat Genet, 2014. **46**(8): p. 872-6.
49. Kawamoto, T., et al., *HER Receptor Family: Novel Candidate for Targeted Therapy for Gallbladder and Extrahepatic Bile Duct Cancer*. Gastrointest Cancer Res, 2007. **1**(6): p. 221-7.
50. Roa, J.C., et al., *Promoter methylation profile in gallbladder cancer*. J Gastroenterol, 2006. **41**(3): p. 269-75.
51. Hecht, F., et al., *Adenocarcinoma of the gallbladder: chromosome abnormalities in a genetic form of cancer*. Cancer Genet Cytogenet, 1983. **8**(3): p. 185-90.
52. Roa, I., et al., *DNA content in gallbladder carcinoma: a flow cytometric study of 96 cases*. Histopathology, 1993. **23**(5): p. 459-64.
53. Gorunova, L., et al., *Nonrandom chromosomal aberrations and cytogenetic heterogeneity in gallbladder carcinomas*. Genes Chromosomes Cancer, 1999. **26**(4): p. 312-21.
54. Wistuba, II, et al., *Genome-wide allelotyping analysis reveals multiple sites of allelic loss in gallbladder carcinoma*. Cancer Res, 2001. **61**(9): p. 3795-800.
55. Castillo, J., P. Garcia, and J.C. Roa, *[Genetic alterations in preneoplastic and neoplastic injuries of the gallbladder]*. Rev Med Chil, 2010. **138**(5): p. 595-604.
56. Lowenfels, A.B., et al., *Gallstone growth, size, and risk of gallbladder cancer: an interracial study*. Int J Epidemiol, 1989. **18**(1): p. 50-4.

57. Henson, D.E., J. Albores-Saavedra, and D. Corle, *Carcinoma of the gallbladder. Histologic types, stage of disease, grade, and survival rates*. *Cancer*, 1992. **70**(6): p. 1493-7.
58. Moore, S.P., et al., *Cancer in indigenous people in Latin America and the Caribbean: a review*. *Cancer Med*, 2014. **3**(1): p. 70-80.
59. Devor, E.J. and R.W. Buechley, *Gallbladder cancer in Hispanic new mexicans. II. Familial occurrence in two northern New Mexico kindreds*. *Cancer Genetics and Cytogenetics*, 1979. **1**(2): p. 139-145.
60. Nir, I., et al., *Diversification and trends in biliary tree cancer among the three major ethnic groups in the state of New Mexico*. *Am J Surg*, 2012. **203**(3): p. 361-5; discussion 365.
61. Lanier, A.P., et al., *Disparities in cancer mortality among Alaska Native people, 1994-2003*. *Alaska Med*, 2008. **49**(4): p. 120-5.
62. Hamdani, N.H., et al., *Clinicopathological study of gall bladder carcinoma with special reference to gallstones: our 8-year experience from eastern India*. *Asian Pac J Cancer Prev*, 2012. **13**(11): p. 5613-7.
63. Hsing, A.W., et al., *Gallstones and the risk of biliary tract cancer: a population-based study in China*. *Br J Cancer*, 2007. **97**(11): p. 1577-82.
64. Dorobisz, T., et al., *Incidental gallbladder cancer after cholecystectomy: 1990 to 2014*. *Onco Targets Ther*, 2016. **9**: p. 4913-6.
65. Lazcano-Ponce, E.C., et al., *Epidemiology and molecular pathology of gallbladder cancer*. *CA Cancer J Clin*, 2001. **51**(6): p. 349-64.
66. Gilat, T., et al., *An increased familial frequency of gallstones*. *Gastroenterology*, 1983. **84**(2): p. 242-6.
67. Sarin, S.K., et al., *High familial prevalence of gallstones in the first-degree relatives of gallstone patients*. *Hepatology*, 1995. **22**(1): p. 138-41.
68. Shaffer, E.A., *Gallbladder cancer: the basics*. *Gastroenterol Hepatol (N Y)*, 2008. **4**(10): p. 737-41.
69. Hsing, A.W., et al., *Family history of gallstones and the risk of biliary tract cancer and gallstones: a population-based study in Shanghai, China*. *Int J Cancer*, 2007. **121**(4): p. 832-8.
70. Srivastava, K., et al., *Candidate gene studies in gallbladder cancer: a systematic review and meta-analysis*. *Mutat Res*, 2011. **728**(1-2): p. 67-79.
71. Rai, R., et al., *A Multiple Interaction Analysis Reveals ADRB3 as a Potential Candidate for Gallbladder Cancer Predisposition via a Complex Interaction with Other Candidate Gene Variations*. *Int J Mol Sci*, 2015. **16**(12): p. 28038-49.
72. Cha, P.C., et al., *A genome-wide association study identifies SNP in DCC is associated with gallbladder cancer in the Japanese population*. *J Hum Genet*, 2012. **57**(4): p. 235-7.
73. Mhatre, S., et al., *Common genetic variation and risk of gallbladder cancer in India: a case-control genome-wide association study*. *Lancet Oncol*, 2017. **18**(4): p. 535-544.
74. Rosmorduc, O., et al., *ABCB4 gene mutation-associated cholelithiasis in adults*. *Gastroenterology*, 2003. **125**(2): p. 452-9.



75. Rosmorduc, O., B. Hermelin, and R. Poupon, *MDR3 gene defect in adults with symptomatic intrahepatic and gallbladder cholesterol cholelithiasis*. *Gastroenterology*, 2001. **120**(6): p. 1459-67.
76. Lucena, J.F., et al., *A multidrug resistance 3 gene mutation causing cholelithiasis, cholestasis of pregnancy, and adulthood biliary cirrhosis*. *Gastroenterology*, 2003. **124**(4): p. 1037-42.
77. Yadav, S., et al., *Targeted Gene Sequencing of Gallbladder Carcinoma Identifies High-impact Somatic and Rare Germline Mutations*. *Cancer Genomics Proteomics*, 2017. **14**(6): p. 495-506.
78. Trajber, H.J., et al., *Adenocarcinoma of the gallbladder in two siblings*. *Cancer*, 1982. **50**(6): p. 1200-3.
79. Garber, J.E. and W. Shipley, *Carcinoma of the gall bladder in three members of a family*. *Cancer Genet Cytogenet*, 1989. **39**(1): p. 141-2.
80. Hemminki, K. and X. Li, *Familial liver and gall bladder cancer: a nationwide epidemiological study from Sweden*. *Gut*, 2003. **52**(4): p. 592-6.
81. Hemminki, K., et al., *Genetics of gallbladder cancer*. *Lancet Oncol*, 2017. **18**(6): p. e296.
82. Jackson, H.H., et al., *Familial risk in gallbladder cancer*. *Journal of the American College of Surgeons*. **205**(3): p. S38.
83. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland*. *N Engl J Med*, 2000. **343**(2): p. 78-85.
84. Ziegler, A. and I.R. König, *A statistical approach to genetic epidemiology*. 2nd ed. 2010, Weinheim: Wiley-VCH. xxiv, 497 p.
85. Dawn Teare, M. and J.H. Barrett, *Genetic linkage studies*. *Lancet*, 2005. **366**(9490): p. 1036-44.
86. Ott, J., J. Wang, and S.M. Leal, *Genetic linkage analysis in the age of whole-genome sequencing*. *Nat Rev Genet*, 2015. **16**(5): p. 275-84.
87. Klein, A.P., *Overview of linkage analysis: application to pancreatic cancer*. *Methods Mol Med*, 2005. **103**: p. 329-41.
88. Bailey-Wilson, J.E., *Parametric and Nonparametric Linkage Analysis*, in *eLS*. 2001, John Wiley & Sons, Ltd.
89. Kosters, A., M. Jirsa, and A.K. Groen, *Genetic background of cholesterol gallstone disease*. *Biochim Biophys Acta*, 2003. **1637**(1): p. 1-19.
90. Ziol, M., et al., *ABCB4 heterozygous gene mutations associated with fibrosing cholestatic liver disease in adults*. *Gastroenterology*, 2008. **135**(1): p. 131-41.
91. Hemminki, K., et al., *Familial risks for gallstones in the population of Sweden*. *BMJ Open Gastroenterol*, 2017. **4**(1): p. e000188.
92. Penrose, L.S., *THE DETECTION OF AUTOSOMAL LINKAGE IN DATA WHICH CONSIST OF PAIRS OF BROTHERS AND SISTERS OF UNSPECIFIED PARENTAGE*. *Annals of Eugenics*, 1935. **6**(2): p. 133-138.
93. Haseman, J.K. and R.C. Elston, *The investigation of linkage between a quantitative trait and a marker locus*. *Behav Genet*, 1972. **2**(1): p. 3-19.

94. McPeck, M.S., *Optimal allele-sharing statistics for genetic mapping using affected relatives*. *Genet Epidemiol*, 1999. **16**(3): p. 225-49.
95. Basu, S., et al., *A likelihood-based trait-model-free approach for linkage detection of binary trait*. *Biometrics*, 2010. **66**(1): p. 205-13.
96. Bailey-Wilson, J.E., *Parametric versus nonparametric and two-point versus multipoint: controversies in gene mapping*, in *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*. 2004, John Wiley & Sons, Ltd.
97. Lathrop, G.M., et al., *Strategies for multilocus linkage analysis in humans*. *Proc Natl Acad Sci U S A*, 1984. **81**(11): p. 3443-6.
98. Cottingham, R.W., Jr., R.M. Idury, and A.A. Schaffer, *Faster sequential genetic linkage computations*. *Am J Hum Genet*, 1993. **53**(1): p. 252-63.
99. Kruglyak, L., et al., *Parametric and nonparametric linkage analysis: a unified multipoint approach*. *Am J Hum Genet*, 1996. **58**(6): p. 1347-63.
100. Abecasis, G.R., et al., *Merlin--rapid analysis of dense genetic maps using sparse gene flow trees*. *Nat Genet*, 2002. **30**(1): p. 97-101.
101. Lange, K., et al., *Mendel: the Swiss army knife of genetic analysis programs*. *Bioinformatics*, 2013. **29**(12): p. 1568-70.
102. Bailey-Wilson, J.E. and A.F. Wilson, *Linkage analysis in the next-generation sequencing era*. *Hum Hered*, 2011. **72**(4): p. 228-36.
103. Spain, S.L. and J.C. Barrett, *Strategies for fine-mapping complex traits*. *Hum Mol Genet*, 2015. **24**(R1): p. R111-9.
104. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements*. *Nucleic Acids Res*, 2017. **45**(D1): p. D183-D189.
105. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists*. *Nucleic Acids Res*, 2009. **37**(1): p. 1-13.
106. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. *Nat Protoc*, 2009. **4**(1): p. 44-57.
107. Visscher, P.M., et al., *Five years of GWAS discovery*. *Am J Hum Genet*, 2012. **90**(1): p. 7-24.
108. Clarke, G.M., et al., *Basic statistical analysis in genetic case-control studies*. *Nat Protoc*, 2011. **6**(2): p. 121-33.
109. Bull, S.B., S. John, and L. Briollais, *Fine mapping by linkage and association in nuclear family and case-control designs*. *Genetic Epidemiology*, 2005. **29**(S1): p. S48-S58.
110. Chang, Y.C., et al., *Genome-wide linkage analysis and regional fine mapping identified variants in the RYR3 gene as a novel quantitative trait locus for circulating adiponectin in Chinese population*. *Medicine (Baltimore)*, 2016. **95**(44): p. e5174.
111. Hsueh, W.C., et al., *Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24*. *Diabetes*, 2003. **52**(2): p. 550-7.

112. Zinovieva, E., et al., *Comprehensive linkage and association analyses identify haplotype, near to the TNFSF15 gene, significantly associated with spondyloarthritis*. PLoS Genet, 2009. **5**(6): p. e1000528.
113. Herr, M., et al., *Evaluation of fine mapping strategies for a multifactorial disease locus: systematic linkage and association analysis of IDDM1 in the HLA region on chromosome 6p21*. Hum Mol Genet, 2000. **9**(9): p. 1291-301.
114. Painter, J.N., et al., *High-density fine-mapping of a chromosome 10q26 linkage peak suggests association between endometriosis and variants close to CYP2C19*. Fertil Steril, 2011. **95**(7): p. 2236-40.
115. Dastani, Z., et al., *Fine mapping and association studies of a high-density lipoprotein cholesterol linkage region on chromosome 16 in French-Canadian subjects*. Eur J Hum Genet, 2010. **18**(3): p. 342-7.
116. Park, H., et al., *A family-based association study after genome-wide linkage analysis identified two genetic loci for renal function in a Mongolian population*. Kidney Int, 2013. **83**(2): p. 285-92.
117. Devlin, B. and N. Risch, *A comparison of linkage disequilibrium measures for fine-scale mapping*. Genomics, 1995. **29**(2): p. 311-22.
118. Kruglyak, L. and E.S. Lander, *High-resolution genetic mapping of complex traits*. Am J Hum Genet, 1995. **56**(5): p. 1212-23.
119. Klein, E.A., P.A. Kupelian, and J.S. Witte, *Does a family history of prostate cancer result in more aggressive disease?* Prostate Cancer Prostatic Dis, 1998. **1**(6): p. 297-300.
120. Anders, C.K., et al., *Breast cancer before age 40 years*. Semin Oncol, 2009. **36**(3): p. 237-49.
121. Brenner, H., et al., *Individual and joint contribution of family history and Helicobacter pylori infection to the risk of gastric carcinoma*. Cancer, 2000. **88**(2): p. 274-9.
122. Fuchs, C.S., et al., *A prospective study of family history and the risk of colorectal cancer*. N Engl J Med, 1994. **331**(25): p. 1669-74.
123. Acevedo, A., et al., *Morbimortalidad y estadía hospitalaria en los egresados por patología biliar del Servicio de Cirugía del Hospital del Salvador*. Rev. chil. cir, 2005. **57**(5): p. 412-416.
124. Ministerio de Salud, *Guía Clínica Colectomía Preventiva en Adultos de 35 a 49 años*. 2014, MINSAL: Santiago.
125. Randi, G., S. Franceschi, and C. La Vecchia, *Gallbladder cancer worldwide: geographical distribution and risk factors*. Int J Cancer, 2006. **118**(7): p. 1591-602.
126. Ruiz-Linares, A., et al., *Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals*. PLoS Genet, 2014. **10**(9): p. e1004572.
127. Gershon, E.S. and K.S. Grennan, *Genetic and genomic analyses as a basis for new diagnostic nosologies*. Dialogues Clin Neurosci, 2015. **17**(1): p. 69-78.
128. Purcell, S., S.S. Cherny, and P.C. Sham, *Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits*. Bioinformatics, 2003. **19**(1): p. 149-50.

129. Johnson, J.L., *Genetic Association Study (GAS) Power Calculator*. 2017, Department of Biostatistics Center for Statistical Genetics: University of Michigan School of Public Health. p. [http://csg.sph.umich.edu/abecasis/cats/gas\\_power\\_calculator/index.html](http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html).
130. Wise, A.L., L. Gyi, and T.A. Manolio, *eXclusion: toward integrating the X chromosome in genome-wide association analyses*. *Am J Hum Genet*, 2013. **92**(5): p. 643-7.
131. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies*. *Nat Genet*, 2006. **38**(8): p. 904-9.
132. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals*. *Genome Res*, 2009. **19**(9): p. 1655-64.
133. Anderson, C.A., et al., *Data quality control in genetic case-control association studies*. *Nat Protoc*, 2010. **5**(9): p. 1564-73.
134. Turner, S., et al., *Quality control procedures for genome-wide association studies*. *Curr Protoc Hum Genet*, 2011. **Chapter 1**: p. Unit1 19.
135. Nyholt, D.R., *All LODs are not created equal*. *Am J Hum Genet*, 2000. **67**(2): p. 282-8.
136. Lange, E.M. and K. Lange, *Powerful allele sharing statistics for nonparametric linkage analysis*. *Hum Hered*, 2004. **57**(1): p. 49-58.
137. Sobel, E. and K. Lange, *Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics*. *Am J Hum Genet*, 1996. **58**(6): p. 1323-37.
138. Day-Williams, A.G., et al., *Unifying ideas for non-parametric linkage analysis*. *Hum Hered*, 2011. **71**(4): p. 267-80.
139. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges*. *Nat Rev Genet*, 2008. **9**(5): p. 356-69.
140. Ott, J., *Linkage analysis and family classification under heterogeneity*. *Ann Hum Genet*, 1983. **47**(Pt 4): p. 311-20.
141. Smith, C.A., *Testing for Heterogeneity of Recombination Fraction Values in Human Genetics*. *Ann Hum Genet*, 1963. **27**: p. 175-82.
142. Assenov, Y., et al., *Computing topological parameters of biological networks*. *Bioinformatics*, 2008. **24**(2): p. 282-4.
143. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. *Genome Res*, 2003. **13**(11): p. 2498-504.
144. Nervi, F., et al., *Frequency of gallbladder cancer in Chile, a high-risk area*. *Int J Cancer*, 1988. **41**(5): p. 657-60.
145. Bustos, B.I., et al., *Common variants in ABCG8 and TRAF3 genes confer risk for gallstone disease and gallbladder cancer in admixed Latinos with Mapuche Native American ancestry*. *bioRxiv*, 2018.
146. Baez, S., et al., *Genetic variants involved in gallstone formation and capsaicin metabolism, and the risk of gallbladder cancer in Chilean women*. *World J Gastroenterol*, 2010. **16**(3): p. 372-8.
147. Grunhage, F. and F. Lammert, *Gallstone disease. Pathogenesis of gallstones: A genetic perspective*. *Best Pract Res Clin Gastroenterol*, 2006. **20**(6): p. 997-1015.
148. Dikkers, A. and U.J. Tietge, *The neglected cousin of the hepatocyte: how gallbladder epithelial cells might contribute to cholesterol gallstone formation*. *Dig Dis Sci*, 2013. **58**(2): p. 296-8.

149. Vaisman, B.L., et al., *ABCA1 overexpression leads to hyperalphalipoproteinemia and increased biliary cholesterol excretion in transgenic mice*. J Clin Invest, 2001. **108**(2): p. 303-9.
150. Yoon, J.H., et al., *ATP-binding cassette sterol transporters are differentially expressed in normal and diseased human gallbladder*. Dig Dis Sci, 2013. **58**(2): p. 431-9.
151. Tsaroucha, A.K., et al., *Megalin and cubilin in the human gallbladder epithelium*. Clin Exp Med, 2008. **8**(3): p. 165-70.
152. Chen, Y., J. Kong, and S. Wu, *Cholesterol gallstone disease: focusing on the role of gallbladder*. Lab Invest, 2015. **95**(2): p. 124-31.
153. Lee, J., et al., *Polarized cholesterol and phospholipid efflux in cultured gall-bladder epithelial cells: evidence for an ABCA1-mediated pathway*. Biochem J, 2002. **364**(Pt 2): p. 475-84.
154. Kuzu, O.F., M.A. Noory, and G.P. Robertson, *The Role of Cholesterol in Cancer*. Cancer Res, 2016. **76**(8): p. 2063-70.
155. Chou, J.L., et al., *Hypermethylation of the TGF-beta target, ABCA1 is associated with poor prognosis in ovarian cancer patients*. Clin Epigenetics, 2015. **7**: p. 1.
156. Solomon, K.R., et al., *Words of wisdom. Re: Dysregulation of cholesterol homeostasis in human prostate cancer through loss of ABCA1*. Eur Urol, 2013. **63**(6): p. 1128-9.
157. Smith, B. and H. Land, *Anticancer activity of the cholesterol exporter ABCA1 gene*. Cell Rep, 2012. **2**(3): p. 580-90.
158. Chitrala, K.N., P. Nagarkatti, and M. Nagarkatti, *Prediction of Possible Biomarkers and Novel Pathways Conferring Risk to Post-Traumatic Stress Disorder*. PLoS One, 2016. **11**(12): p. e0168404.
159. Balding, D.J., *A tutorial on statistical methods for population association studies*. Nat Rev Genet, 2006. **7**(10): p. 781-91.
160. Hormozdiari, F., et al., *Identifying causal variants at loci with multiple signals of association*. Genetics, 2014. **198**(2): p. 497-508.
161. Srivastava, K., et al., *Significant association between toll-like receptor gene polymorphisms and gallbladder cancer*. Liver Int, 2010. **30**(7): p. 1067-72.
162. Tazuma, S., et al., *Report on the 2013 national cholelithiasis survey in Japan*. J Hepatobiliary Pancreat Sci, 2015. **22**(5): p. 392-5.
163. Hong, E.P. and J.W. Park, *Sample size and statistical power calculation in genetic association studies*. Genomics Inform, 2012. **10**(2): p. 117-22.
164. Yi, N., et al., *Multiple comparisons in genetic association studies: a hierarchical modeling approach*. Stat Appl Genet Mol Biol, 2014. **13**(1): p. 35-48.
165. Bardi, G., et al., *Abnormal karyotypes in three carcinomas of the gallbladder*. Cancer Genet Cytogenet, 1994. **76**(1): p. 15-8.
166. Roa, I., et al., *DNA ploidy pattern and tumor suppressor gene p53 expression in gallbladder carcinoma*. Cancer Epidemiol Biomarkers Prev, 1997. **6**(7): p. 547-50.
167. Pai, R.K., K. Mojtahed, and R.K. Pai, *Mutations in the RAS/RAF/MAP kinase pathway commonly occur in gallbladder adenomas but are uncommon in gallbladder adenocarcinomas*. Appl Immunohistochem Mol Morphol, 2011. **19**(2): p. 133-40.
168. Asai, T., et al., *High frequency of TP53 but not K-ras gene mutations in Bolivian patients with gallbladder cancer*. Asian Pac J Cancer Prev, 2014. **15**(13): p. 5449-54.

169. Wistuba, IJ, et al., *Allele-specific mutations involved in the pathogenesis of endemic gallbladder carcinoma in Chile*. *Cancer Res*, 1995. **55**(12): p. 2511-5.
170. Kim, Y.T., et al., *Genetic alterations in gallbladder adenoma, dysplasia and carcinoma*. *Cancer Lett*, 2001. **169**(1): p. 59-68.
171. Chan, E. and J. Berlin, *Biliary tract cancers: understudied and poorly understood*. *J Clin Oncol*, 2015. **33**(16): p. 1845-8.
172. Yokoyama, N., et al., *Mutations of p53 in gallbladder carcinomas in high-incidence areas of Japan and Chile*. *Cancer Epidemiol Biomarkers Prev*, 1998. **7**(4): p. 297-301.
173. Nakamura, H., et al., *Genomic spectra of biliary tract cancer*. *Nat Genet*, 2015. **47**(9): p. 1003-10.