



UNIVERSIDAD  
**MAYOR**

para espíritus emprendedores

Facultad de Ciencias

**ESCUELA DE  
BIOTECNOLOGÍA**

**Establecimiento de un Puntaje de Riesgo Poligénico para estimación de riesgo de enfermedad de Hígado Graso No Alcohólico en una población mestiza chilena**

Stefania Andrea Rosso Villanelo

Proyecto de Tesis para optar al título de Biotecnóloga

Santiago - Chile  
2021



UNIVERSIDAD  
**MAYOR**

para espíritus emprendedores

Facultad de Ciencias

---

**ESCUELA DE  
BIOTECNOLOGÍA**

---

**Establecimiento de un Puntaje de Riesgo Poligénico para estimación de riesgo de enfermedad de Hígado Graso No Alcohólico en una población mestiza chilena**

Stefania Andrea Rosso Villanelo

Proyecto de Tesis para optar al título de Biotecnóloga

Tutor: Ricardo Alejandro Verdugo Salgado, PhD.

Co-tutor: Alberto Jesús Martín Martín, PhD.

Santiago - Chile

2021

*A Laura Kanthak, Antonia Cardemil y Francesca Rosso.*

*Nunca dejen de cuestionar y explorar su mundo.*

SOLO USO ACADÉMICO

## AGRADECIMIENTOS

Quiero agradecer a mi tutor, doctor Ricardo Verdugo, y a la doctora Catterina Ferreccio por haberme dado la oportunidad de hacer este trabajo de tesis y guiado en todo el proceso. Todo lo que pude aprender fue más de lo que esperaba y estoy segura de que será fundamental de hoy en adelante en cualquier proyecto en que participe.

A mi co-tutor, doctor Alberto Martin, por todos sus consejos e invaluable apoyo moral, por toda la amabilidad y disposición para ayudarme en todo momento, por sus palabras de ánimo que me mantuvieron con la motivación que a veces perdía, por darme la seguridad que muchas veces no tuve y que necesitaba.

A Cristian Yañez y Rocío Retamales por siempre estar pacientemente resolviendo mis dudas cuando me vi atrapada, desde el día en que me instalé con mi computador en el laboratorio.

A todos los profesores que se involucraron en mi formación profesional desde lo académico hasta lo valórico durante el programa de pregrado.

Y, por supuesto, a mis padres Carolina Villanelo y Alejandro Rosso, mis hermanos Paulina, Angelo y Lorena, mis amigos, y a mi pareja, Daniel Rivera, quienes me acompañaron e impulsaron a siempre seguir adelante en todos los momentos en que quise rendirme.

## INFORMACIÓN DEL CENTRO DE INVESTIGACIÓN

### **Programa de Genética Humana - Instituto de Ciencias Biomédicas (ICBM)**

El Laboratorio de Genómica Biomédica (GENOMED) tiene por objetivo la comprensión de los mecanismos del genoma para formar características complejas en el ser humano, utilizando métodos computacionales y tecnología de alto rendimiento. El Laboratorio ChileGenómico cuenta con equipamiento y personal especializados en la generación de datos genómicos masivos, con especial énfasis en estudios poblacionales humanos. GENOMED y ChileGenómico son dirigidos por el doctor Ricardo Alejandro Verdugo y forman parte del Programa de Genética Humana del ICBM, de la Facultad de Medicina de la Universidad de Chile. Desde 2017, GENOMED colabora con el Centro Avanzado de Enfermedades Crónicas (ACCDiS) y en particular con la Dra. Catterina Ferreccio, para la caracterización genética de los participantes del estudio MAUCO. El presente trabajo de tesis se realizó en GENOMED con datos fenotípicos recolectados por ACCDiS y datos genéticos generados por ChileGenómico.

### **Fuentes de Financiamiento**

FONDEF Continuación D10E1007 (ChileGenómico)

FONDAP Nº 15130011 (ACCDiS)

## ABREVIATURAS

**AIC:** Criterio de Información de Aikaike

**ALT:** Alanina Aminotransferasa

**AST:** Aspartato Aminotransferasa

**AYM:** Aymara

**CI:** Intervalo de Confianza

**CL:** Chilena

**DBP:** Presión Sanguínea Diastólica

**EUR:** Europea

**FLD:** Enfermedad de Hígado Graso

**HDL:** Lipoproteínas de Alta Densidad

**IMC:** Índice de Masa Corporal

**LD:** Desequilibrio de Ligamiento

**LDL:** Lipoproteínas de Baja Densidad

**MAF:** Frecuencia de Alelo Menos Común

**MAP:** Mapuche

**NAFLD:** Enfermedad de Hígado Graso No Alcohólico

**NASH:** Esteatohepatitis No Alcohólica

**OR:** Odd Ratio

**PRL:** Nivel de Riesgo Poligénico

**PRS:** Puntaje de Riesgo Poligénico

**SBP:** Presión Sanguínea Sistólica

**SNP:** Polimorfismo de Un Nucleótido

**TG:** Triglicéridos

## **i. ÍNDICE DE CONTENIDOS**

<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. Variables demográficas y perfil bioquímico	1
1.2. Condiciones clínicas	2
1.3. Variantes genéticas asociadas a riesgo	2
1.4. Puntaje de Riesgo Poligénico	5
1.5. Aplicaciones Biotecnológicas	5
1.6. Contexto nacional	7
<b>2. HIPÓTESIS Y OBJETIVOS</b>	<b>9</b>
<b>3. MATERIALES Y MÉTODOS</b>	<b>10</b>
<b>4. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS</b>	<b>13</b>
4.1. Características de la población	13
4.1.1. Perfil bioquímico y condiciones clínicas	13
4.2. Genotipo	14
4.2.1. Variantes génicas	14
4.2.2. Componente ancestral	20
<b>5. DISCUSIÓN DE RESULTADOS</b>	<b>25</b>

5.1. Variables demográficas	25
5.2. Perfil bioquímico y condiciones clínicas	25
5.3. Genotipo	26
5.3.1. Riesgo Poligénico	26
5.3.2. Componente ancestral	27
5.4. Aplicaciones Biotecnológicas	27
<b>6. CONCLUSIONES</b>	<b>29</b>
<b>7. IMPLICANCIAS, RECOMENDACIONES Y PROYECCIONES FUTURAS</b>	<b>30</b>
<b>8. REFERENCIAS</b>	<b>31</b>
<b>9. MATERIAL SUPLEMENTARIO</b>	<b>38</b>



## ii. ÍNDICE DE TABLAS

<b>Tabla 1.</b> <i>Loci de variación genética asociados a riesgo de NAFLD reportados en la literatura</i>	<b>3</b>
<b>Tabla 2.</b> <i>Prevalencia a nivel nacional y regional de condiciones de riesgo de enfermedad de hígado graso</i>	<b>7</b>
<b>Tabla 3.</b> <i>Prevalencia de condiciones clínicas de riesgo en la muestra</i>	<b>13</b>
<b>Tabla 4:</b> <i>Asociación entre NAFLD y variables demográficas, clínicas y bioquímicas</i>	<b>14</b>
<b>Tabla 5.</b> <i>Frecuencias y asociación de variantes de riesgo con FLD</i>	<b>16</b>
<b>Tabla 6.</b> <i>Asociación entre FLD y PRL</i>	<b>17</b>
<b>Tabla 7.</b> <i>Asociación entre condiciones clínicas de riesgo y PRL</i>	<b>18</b>
<b>Tabla 8.</b> <i>Asociación entre alteración de niveles plasmáticos de transaminasas y PRL CL</i>	<b>19</b>
<b>Tabla 9.</b> <i>Asociación entre componente ancestral y FLD</i>	<b>21</b>
<b>Tabla 10.</b> <i>Asociación entre componente ancestral, PRS y transaminasas</i>	<b>21</b>
<b>Tabla 11.</b> <i>Asociación entre NAFLD y PRL por tercil de ancestría amerindia</i>	<b>22</b>
<b>Tabla 12.</b> <i>Asociación entre NAFLD y PRL por tercil de ancestría europea</i>	<b>23</b>
<b>Tabla suplementaria 1.</b> <i>Frecuencias alélicas en variantes de riesgo reportadas a nivel global y en poblaciones europea, chilena, mapuche y aymara</i>	<b>38</b>
<b>Tabla suplementaria 2.</b> <i>Asociación entre variantes de riesgo y condiciones clínicas</i>	<b>38</b>
<b>Tabla suplementaria 3.</b> <i>Asociación entre transaminasas hepáticas y ancestrías</i>	<b>39</b>

### iii. ÍNDICE DE FIGURAS

<b>Figura 1.</b> <i>Secuencia de procedimientos realizados en esta investigación</i>	10
<b>Figura 2.</b> <i>Proporción y número de individuos FLD- y FLD+ en presencia de transaminasas alteradas, obesidad, y colesterol HDL disminuido</i>	15
<b>Figura 3.</b> <i>Proporción y número de individuos FLD- y FLD+ por PRL y PRL CL</i>	17
<b>Figura 4.</b> <i>Proporción y número de individuos con niveles alterados y normales de transaminasas ALT y AST en función de PRL y PRL CL</i>	20
<b>Figura 5.</b> <i>Proporción y número de individuos FLD- y FLD+ en función de PRL y PRL CL por tercil de ancestría amerindia y europea</i>	24
<b>Figura suplementaria 1.</b> <i>Tendencia de niveles de transaminasas hepáticas ALT y AST en función de ancestrías amerindia y europea</i>	40

SOLO USO ACADÉMICO

#### iv. RESUMEN

La enfermedad de hígado graso no alcohólico es la más prevalente de las enfermedades al hígado a nivel global, y su prevalencia en Chile se estima en un 22%. Presentan mayor riesgo las personas con obesidad, resistencia a la insulina, dislipidemias, y personas con predisposición genética. Se han reportado nueve variantes genéticas en diferentes poblaciones europeas, asiáticas, y estadounidenses, pero los estudios han cubierto escasamente poblaciones latinoamericanas. En este contexto, la cohorte Maule (MAUCO) desarrolló un protocolo de estudio para realizar seguimiento a enfermedades crónicas en pacientes provenientes de la región del Maule, Chile, donde se centran las mayores tasas de mortalidad por enfermedades crónicas en el país. En este estudio, se evaluó la replicación de asociación entre nueve variantes génicas con hígado graso no alcohólico reportadas en la literatura en 3966 participantes de MAUCO. Sólo dos de las nueve variantes de riesgo replicaron su asociación: rs2896019 en el gen *PNPLA3* y rs2073080 en el gen *SAMM50*. Un puntaje de riesgo poligénico a partir de estas dos variantes mostró asociación con niveles plasmáticos elevados de las transaminasas hepáticas ALT (OR=1,17,  $p < 0,001$ ) y AST (OR=1,231,  $p < 0,001$ ), y con hígado graso sólo en individuos con ancestría sobre 54.6% europea (OR=1,154,  $p = 0,016$ ), en una muestra donde el 52,13% de los pacientes padece la enfermedad. Este es el primer estudio sobre riesgo poligénico de padecer enfermedad de hígado graso no alcohólico en una población chilena. Nuestros resultados sugieren que los marcadores de riesgo reportados en la literatura para hígado graso no alcohólico tienen modesta replicación en la población chilena, especialmente en personas sin un alto componente de ancestría europea. Este trabajo indica además la necesidad de realizar estudios de asociación a nivel genómico en población chilena orientados a descubrir loci de riesgo que no hayan mostrado asociación con la enfermedad en otras poblaciones.

## v. SUMMARY

Non-Alcoholic Fatty Liver Disease is the most prevalent liver disease worldwide, and its prevalence in Chile is estimated at 22%. People with obesity, insulin resistance, dyslipidemia, and genetic predisposition are at more risk. There have been 9 gene variants frequently reported among different populations. However, these studies have been focused in European, Asian and American populations, and poorly in Latin Americans. In this context, Maule Cohort (MAUCO) developed a study protocol to track chronic diseases in patients from Maule region, Chile. This region has the highest national mortality rates for chronic disease. In the present study, it was evaluated the replication of association between nine previously reported gene variants with non-alcoholic fatty liver disease in 3966 participants from MAUCO. Only two variants replicated their association: rs2896019 in the *PNPLA3* gene, and rs2073080 in the *SAMM50* gene. A polygenic risk score calculated from these two variants was associated with altered levels of ALT (OR=1,173,  $p < 0,001$ ) and AST (OR = 1,231,  $p < 0,001$ ) and with fatty liver disease only in patients above 56.4% of European ancestry (OR=1,154,  $p = 0,016$ ) in a sample where the 52,13% of patients suffer from fatty liver disease. This is the first study on polygenic risk score to develop non-alcoholic fatty liver disease in an admixed Chilean population. These results suggest that previously reported biomarkers for non-alcoholic-fatty liver disease have a modest replication in the Chilean population, especially in patients with a lower level of European ancestry. These findings point to the need of developing genome-wide association studies in the Chilean population to discover risk loci that have not shown association with the disease in other populations.

## 1.INTRODUCCIÓN

Se define hígado graso no alcohólico (NAFLD) como la acumulación de grasa en el hígado mayor al 5% de su peso total, no asociado al consumo de alcohol. A nivel global se trata de la enfermedad al hígado con mayor prevalencia, con una incidencia promedio cercana al 20%<sup>1</sup>. En algunos individuos, esta enfermedad genera una respuesta inflamatoria en el hígado, condición conocida como esteatohepatitis no alcohólica (NASH), que a su vez puede generar tejido fibroso y progresar a cirrosis e incluso a carcinoma hepatocelular en los casos más graves<sup>2</sup>, y en conjunto, estas enfermedades representan la principal causa de trasplantes de hígado<sup>3</sup>. Los pacientes con condiciones metabólicas tales como obesidad, resistencia a la insulina y/o dislipidemias son quienes se encuentran en mayor riesgo de enfermedad<sup>2</sup>. Sin embargo, la predisposición a la enfermedad varía entre individuos debido a variaciones en secuencias de genes relacionados con el metabolismo y la etnicidad: de acuerdo a lo reportado en EEUU, NAFLD afecta a un 45% de población predominantemente hispanica, 33% europea y 34% afroamericana<sup>4</sup>.

### 1.1. Variables demográficas y perfil bioquímico

Si bien NAFLD puede afectar a cualquier grupo etario, suele afectar con mayor frecuencia a personas entre 40 y 50 años<sup>5</sup>. Con relación al sexo, no hay claridad ya que se ha descrito que el paciente promedio con NAFLD es una mujer de mediana edad, mientras que otros estudios han demostrado que el hombre es quien tiene mayor predisposición<sup>6,7</sup>

NAFLD es la principal causa de niveles plasmáticos elevados (>35 [U/L]) de las enzimas hepáticas alanina aminotransferasa (ALT) y aspartato aminotransferasa (AST). Estas enzimas se encuentran normalmente al interior de hepatocitos y por lo tanto su presencia libre en la sangre es indicativa de lesión hepática<sup>8,9</sup>. Las dislipidemias, por otra parte, son factores que permiten determinar si el individuo se ve propenso a desarrollar FLD ya que se relacionan con la concentración de grasa en el hígado. Los niveles plasmáticos aumentados de triglicéridos (>150 [mg/dL]) y lipoproteínas de baja densidad (LDL) (>100 [mg/dL]) predisponen a generación de tejido graso, mientras que los niveles disminuidos de lipoproteínas de alta densidad (HDL) (<40 [mg/dL] en varones o <50 [mg/dL] en mujeres) implican menor catabolismo de los niveles excesivos de colesterol en sangre. Adicionalmente, se ha descrito que un nivel elevado de glicemia (>105 [mg/dL]) indica posible resistencia a insulina que está especialmente asociada al padecimiento de FLD y a su progreso a fibrosis hepática<sup>1</sup>.

## 1.2. Condiciones clínicas

La prevalencia de esta enfermedad aumenta en pacientes obesos hasta un 74%<sup>10</sup>. Kim y cols.<sup>11</sup> reportaron una fuerte relación positiva entre hígado graso y síndrome metabólico (MS), el cual comprende obesidad abdominal y alteraciones en glicemia. Incluso, se ha descrito MS como un factor predictivo para FLD<sup>12</sup>. Sin embargo, un 30% de los pacientes con FLD no presenta anormalidades metabólicas en US<sup>11</sup>. En todos los casos puede verse involucrado el factor genético<sup>13</sup>.

## 1.3. Variantes genéticas asociadas a riesgo

A partir de estudios de asociación de genoma completo (GWAS) ha sido posible identificar polimorfismos de un solo nucleótido (SNPs) que han mostrado tener asociación a NAFLD. La selección de variantes de riesgo para una determinada población puede ser basada en los filtros de frecuencia de alelo menos común (MAF) y de desequilibrio de ligamiento (LD). MAF es una técnica aplicada en estudios de asociación que consiste en seleccionar variantes en que el alelo con menor frecuencia es común y es mayor al 5%, indicando que se tratan de polimorfismos asociados a enfermedades comunes<sup>14</sup>, y LD consiste en la selección de variantes que se encuentren en loci cercanos, por lo que la frecuencia de recombinación es baja, por lo que la probabilidad de heredar esas variantes es similar<sup>15</sup>. Variantes en *PNPLA3*, *SAMM50*, *GCKR*, *PPARγ*, *IL-6*, *TCF7L2*, *PPP1R3B* y *GC* se han asociado a riesgo de NAFLD en distintas poblaciones<sup>16-21</sup>. La frecuencia alélica de SNPs en estos genes difiere entre poblaciones, indicando que el componente ancestral se encuentra involucrado en la predisposición a la enfermedad<sup>22</sup> (Tabla 1, Tabla suplementaria 1). A continuación, se resume el estado del arte actual de cada gen.

### ***PNPLA3 (Gene ID 80339)***

El gen *PNPLA3* codifica para la proteína adiponutrina, que se encuentra al interior de adipocitos y hepatocitos. Esta proteína está involucrada en adipogénesis<sup>23</sup>. Variantes en este gen generan una ganancia de función, promoviendo la acumulación anormal de lípidos en el tejido hepático<sup>24, 25</sup>. La variante rs2896019 se ha relacionado al desarrollo de esta enfermedad y a su progreso a esteatohepatitis no alcohólica (NASH), así como a alteraciones en los niveles plasmáticos de ALT y AST, en poblaciones asiáticas y latinoamericanas<sup>21, 26</sup>.

### **SAMM50 (Gene ID 25813)**

El gen *SAMM50* (homólogo del componente 50 de la maquinaria de clasificación y ensamblaje) está involucrado en el ensamblaje de proteínas barril beta en la membrana externa de la mitocondria<sup>27</sup>. Se ha descrito que la variante rs2073080 de este gen conduce a una disfunción en la mitocondria, alterando su capacidad para remover especies reactivas de oxígeno y así promoviendo el desarrollo de NAFLD<sup>21,26</sup>.

**Tabla 1.** Loci de variación genética asociados a riesgo de NAFLD reportados en la literatura.

Gen	SNP	Chr.	Posición (Mb)	Alelos		Frecuencia alélica
				N	R	Global <sup>(60)</sup> (n=5.008)
<i>PNPLA3</i>	rs2896019 <sup>21,24,25,26</sup>	22	43937814	T	G	0,738
<i>SAMM50</i>	rs2073080 <sup>21,25,26,27</sup>	22	43998522	T	C	0,744
<i>GCKR</i>	rs780094 <sup>28,29</sup>	2	27518370	T	C	0,302
	rs1260326 <sup>28,29</sup>	2	27508073	T	C	0,404
<i>PPARγ</i>	rs1801282 <sup>32</sup>	3	12351626	C	G	0,901
<i>IL-6</i>	rs1800795 <sup>34</sup>	3	12351626	C	G	0,901
<i>TCF7L2</i>	rs7903146 <sup>35</sup>	10	112998590	C	T	0,708
<i>PP1R3B</i>	rs4240624 <sup>37,38</sup>	8	9326721	G	A	0,114
<i>GC</i>	rs222054 <sup>40</sup>	4	71738582	C	G	0,709

### **GCKR (Gene ID 2646)**

El gen *GCKR* (regulador de glucoquinasa) codifica para una proteína que regula la actividad de la glucoquinasa en hepatocitos y en el páncreas, en respuesta a los niveles de glucosa en la sangre, por lo que juega un rol central en el metabolismo de carbohidratos y en la activación de lipogénesis

hepática<sup>28</sup>. Las variantes rs780094 y rs1260326 se han asociado a hipertrigliceridemia<sup>29</sup> e hiperglicemia en diferentes poblaciones, aumentando el riesgo de desarrollar NAFLD<sup>29, 30</sup>.

#### ***PPAR $\gamma$* (Gene ID 5468)**

PPAR corresponde a una subfamilia de receptores nucleares activados por proliferador de peroxisomas. *PPAR $\gamma$*  es un subtipo de receptores PPAR que funciona como regulador en la diferenciación de adipocitos<sup>31</sup>. La variante rs1801282 en este gen se ha visto implicada en la patología de enfermedades metabólicas tales como obesidad y resistencia a la insulina<sup>32</sup>.

#### ***IL-6* (Gene ID 3569)**

La citoquina codificada por este gen participa en la respuesta inflamatoria, así como en la maduración de linfocitos B. Este gen está implicado en las etapas inflamatorias de múltiples enfermedades<sup>33</sup>. La variante rs1800795 en este gen se ha asociado a diabetes mellitus y NASH<sup>34</sup>.

#### ***TCF7L2* (Gene ID 6934)**

Este gen codifica para un factor de transcripción con un rol clave en la ruta de señalización de Wnt, glicoproteínas cruciales en regulación y función de señalizaciones. Esta proteína se encuentra involucrada en la homeostasis de glucosa en sangre. La variante rs7903146 en *TCF7L2* se ha asociado con riesgo de diabetes mellitus tipo II (DM2), enfermedad coronaria y NAFLD<sup>35</sup>.

#### ***PPP1R3B* (Gene ID 79660)**

La proteína codificada por este gen es la subunidad reguladora 3B de fosfatasa 1, cuya función está involucrada en la regulación de síntesis de glucógeno en tejido esquelético y hepático<sup>36</sup>. La variante rs4240624 en este gen se ha asociado a acumulación de triglicéridos y glucógeno en el hígado<sup>37, 38</sup>.

#### ***GC* (Gene ID 2638)**

Este gen codifica para la proteína de unión a vitamina D. Se encuentra en muchos tipos celulares, y se encuentra involucrada en el transporte de metabolitos, incluyendo ácidos grasos<sup>39</sup>. La variante rs222054 se ha asociado a alteraciones en rutas metabólicas involucradas en adipogénesis y resistencia a la insulina<sup>40</sup>.



#### **1.4. Puntaje de Riesgo Poligénico**

Los puntajes de riesgo poligénico (PRS) son expresiones matemáticas construidas a partir de los alelos que porta una persona en múltiples loci asociados a riesgo. Existen distintos tipos de PRS, desde la simple multiplicatoria de los riesgos asociados a cada alelo que porta una persona a estimaciones bayesianas que consideran las frecuencias alélicas poblacionales como probabilidades *a priori* de portar un alelo. El valor del PRS para enfermedades de herencia compleja, como la de hígado graso no alcohólico, es que, si bien cada locus aporta sólo una pequeña fracción de riesgo, al agrupar múltiples loci el riesgo acumulado puede ser significativo y por lo tanto el PRS puede ser predictivo de la probabilidad que tiene una persona de desarrollar una enfermedad<sup>41</sup>. Al basarse en el genotipo inferido a partir de un ensayo de ADN, la predicción puede realizarse incluso antes del nacimiento, mediante un test genético *in vitro*. Sin embargo, el desarrollo y aplicación PRS está aún en etapas iniciales. La gran mayoría de los estudios GWAS de los cuales derivan las asociaciones utilizadas en el cálculo de los PRS se han realizado en poblaciones de ancestría Europea<sup>42,43</sup>, y las asociaciones reportadas no han sido evaluadas en otras poblaciones. Por lo tanto, el valor predictivo del PRS que de ellos deriva es desconocido en poblaciones no europeas, tal como en Latinoamérica<sup>44</sup>. Recientemente se describió por primera vez que PRS funciona como predictor de riesgo de transaminasas elevadas en una población mexicana, pero no su asociación directa con NAFLD<sup>21</sup>.

Múltiples estudios han establecido diferencias de riesgo de desarrollar NAFLD en distintas etnias, pero se han centrado mayoritariamente en poblaciones de Asia y Europa<sup>45</sup>. Dichos trabajos determinaron que NAFLD es más prevalente en individuos de ancestría hispánica en estudios que contemplan también alimentación y estilo de vida, que expone a las personas a desarrollar NAFLD por factores ambientales<sup>45</sup>. Por lo tanto, es posible hipotetizar que existen diferencias en el riesgo poligénico para NAFLD entre hispanos y poblaciones de otros orígenes.

#### **1.5. Aplicaciones biotecnológicas**

Existe interés en el mercado en el desarrollo de pruebas diagnósticas que permitan un diagnóstico temprano de la enfermedad, antes de que progrese a una cirrosis, momento en el cual la condición se vuelve irreversible. La detección de NAFLD se realiza por pruebas de ultrasonido abdominal y tomografía computarizada, y su diagnóstico definitivo sólo se logra mediante exámenes de biopsia de hígado, un proceso invasivo<sup>46,47</sup>. En grupos de alto riesgo, particularmente personas obesas, se recomienda realizar pruebas imagenológicas como ecografías para detectar signos de la enfermedad. Sin embargo, el NAFLD puede pasar desapercibido a estas pruebas y además también

puede presentarse en personas no obesas<sup>47</sup>. Por lo tanto, existe demanda por pruebas que permitan identificar grupos de alto riesgo en los cuales la realización de ensayos diagnósticos permita identificar pacientes de forma temprana y eficiente.

La empresa BioPredictive ha desarrollado pruebas diagnósticas de enfermedades al hígado mediante métodos no invasivos, basándose en niveles plasmáticos de transaminasas hepáticas, proteínas plasmáticas, colesterol y glicemia como biomarcadores asociados a daño hepático<sup>48</sup>. Es un producto económico y preciso, sin embargo, se trata de información sobre el estado de salud actual de la persona, por lo que no permite tomar medidas preventivas. Otros estudios han logrado utilizar biomarcadores para predecir de forma precisa el progreso de NAFLD a fibrosis o carcinoma<sup>49-51</sup>, pero nuevamente se trata de un proceso que funciona sólo cuando ya se ha desarrollado NAFLD en algún grado.

Los PRS tienen la potencialidad de convertirse en biomarcadores de riesgo útiles para la clínica. Si bien el uso de PRS se ha restringido principalmente al ámbito de la investigación, recientemente se está evaluando su utilidad para predecir riesgo en aplicaciones clínicas y de salud pública<sup>52-54</sup>. A diferencia de otro tipo de biomarcadores, los PRS pueden ser calculados mucho antes del inicio de un proceso patológico, incluso antes del momento de nacimiento. En teoría, los PRS podrían ser usados para discriminar grupos de individuos con distinto nivel de riesgo a sufrir una enfermedad y de esa manera mejorar la dirección de estrategias preventivas, como imagenología, o de tratamiento.

Además de la aplicación en la clínica, los marcadores genéticos pueden ser útiles para particulares interesados en conocer su composición genética. Conociendo su riesgo genético, es posible que los individuos adopten cambios en su estilo de vida tendientes a disminuir las probabilidades de enfermarse. Esto motiva a empresas de test genéticos de acceso directo al consumidor tales como 23andMe que, por ejemplo, incluyó en 2019 un reporte de riesgo a desarrollar diabetes tipo II en su plataforma<sup>43</sup>.

Por lo tanto, los PRS podrían derivar en el desarrollo de ensayos aplicables tanto para la medicina de precisión como para la orientación de políticas públicas por parte de las autoridades sanitarias. Sin embargo, antes de poder ser aplicados en alguna población, los PRS deben ser validados en cuanto a su nivel predictivo en la población de interés y se debe evaluar si su poder predictivo depende de la ancestría de las personas.

## 1.6. Contexto nacional

La población chilena se ha enfrentado a grandes variaciones socioeconómicas y ambientales en los últimos años. Estos cambios han significado alteraciones en el panorama de la salud: por un lado, ha aumentado considerablemente la esperanza de vida<sup>55</sup>, y en paralelo también ha crecido la proporción de chilenos afectados por condiciones de riesgo de enfermedades crónicas<sup>56</sup>, que representan el 85% de las causas de muerte en el país<sup>57</sup>. El año 2018, el Departamento de Estadísticas e Información de Salud (DEIS)<sup>58</sup> reportó que las enfermedades al hígado representaron la séptima causa de muerte a nivel nacional y cerca del 51% de los fallecimientos por enfermedades del sistema digestivo. En los resultados de la Encuesta Nacional de Salud 2016-2017 se reportó que el 18,9% de los chilenos presenta niveles de transaminasas asociados a daño hepático, más del 36,6% presenta alguna dislipidemia, el 41,2% padece síndrome metabólico y el 10,2% reportó tener diabetes tipo II (Tabla 2).

**Tabla 2.** Prevalencia a nivel nacional y regional de condiciones de riesgo de enfermedad de hígado graso. Fuente: Encuesta Nacional de Salud, 2016-2017, Ministerio de Salud<sup>58</sup>. 'n' indica el número de individuos y, entre paréntesis, la proporción respecto de la población expresada como porcentaje.

	Chile (n)	Maule (n)	Hombres (n)	Mujeres (n)
<b>Hígado graso, cirrosis, o daño hepático crónico (autorreporte)</b>	323 (5,2%)	14 (8,3%)	75 (3,3%)	248 (7,0%)
<b>Transaminasas elevadas</b>	778 (18,9%)	29 (10,6%)	280 (18,3%)	498 (19,4%)
<b>Sobrepeso u obesidad</b>	4.170 (75,0%)	264 (76,2%)	1.489 (74,5%)	2.681 (75,6%)
<b>HDL &lt; 40 mg/dL H, &lt;50 mg/dL M</b>	1.662 (44,9%)	110 (47,7%)	501 (40,7%)	1.161 (48,9%)
<b>LDL &gt; 100 mg/dL</b>	1.858 (52,2%)	111 (52,9%)	678 (52,9%)	1.180 (51,6%)
<b>TG ≥ 100 [mg/dL]</b>	1.294 (36,6%)	71 (33,5%)	526 (42,3%)	768 (31,0%)
<b>Síndrome Metabólico</b>	1.574 (41,2%)	98 (38,7%)	578 (44,0%)	996 (38,5%)
<b>Diabetes tipo 2 (autorreporte)</b>	738 (10,2%)	58 (11,1%)	246 (8,6%)	568 (11,8%)

En este contexto, se creó la cohorte de Maule (MAUCO), con el objetivo de analizar la situación y progreso de enfermedades crónicas en la comuna de Molina, Chile, a través del seguimiento a datos de hábitos de riesgo, condiciones clínicas y muestras biológicas de 10.000 individuos<sup>59</sup>. El trabajo realizado por MAUCO se centró en cuantificar la prevalencia de un grupo de enfermedades crónicas, incluyendo hígado graso. Para ello colaboraron con el laboratorio Genomed, perteneciente a la Universidad de Chile, con el fin de determinar si las prevalencias en esta

población en particular podrían estar relacionadas a un factor genético y a la ancestría de las personas. Según un estudio realizado por Riquelme y col. (2009)<sup>61</sup>, NAFLD se manifiesta en un 23% de chilenos de ancestría hispánica. Adicionalmente, Riquelme sugiere que la ancestría amerindia puede ser un antecedente importante en la predisposición al desarrollo de NAFLD. Sumado a lo anterior, hoy sabemos que, en promedio, el componente ancestral de la población chilena es 53% europeo, 42% amerindio y 4% africano en una muestra de 2.843 individuos en diferentes zonas del país<sup>61</sup>.

En el presente trabajo, se genotipificaron 3.966 muestras de la cohorte MAUCO mediante secuenciación dirigida, que incluyó 150 SNPs informativos de ancestría y 9 SNPs asociados a riesgo de NAFLD, con la finalidad de estimar la influencia del componente genético y ancestral en el desarrollo de la enfermedad de hígado graso en la población chilena.

El objetivo de esta investigación es evaluar el grado de replicación de las asociaciones reportadas para nueve variantes de riesgo en los genes *PNPLA3*, *SAMM50*, *GCKR*, *PPP1R3B*, *PPAR $\gamma$* , *IL-6*, *GC* y *TCF7L2* con hígado graso no alcohólico en participantes de MAUCO en la zona central de Chile, con la finalidad de determinar los marcadores de riesgo más relevantes de esta enfermedad, y así abrir la posibilidad de generar instrumentos que utilicen la tecnología de PRS para estimar riesgo de enfermedad de hígado graso de forma dirigida a población chilena.

## 2.HIPÓTESIS Y OBJETIVOS

Existen diferencias en las asociaciones entre enfermedad de hígado graso no alcohólico y variantes genéticas de riesgo entre la población chilena y las reportadas en otras poblaciones producto de la ancestría mestiza de los chilenos.

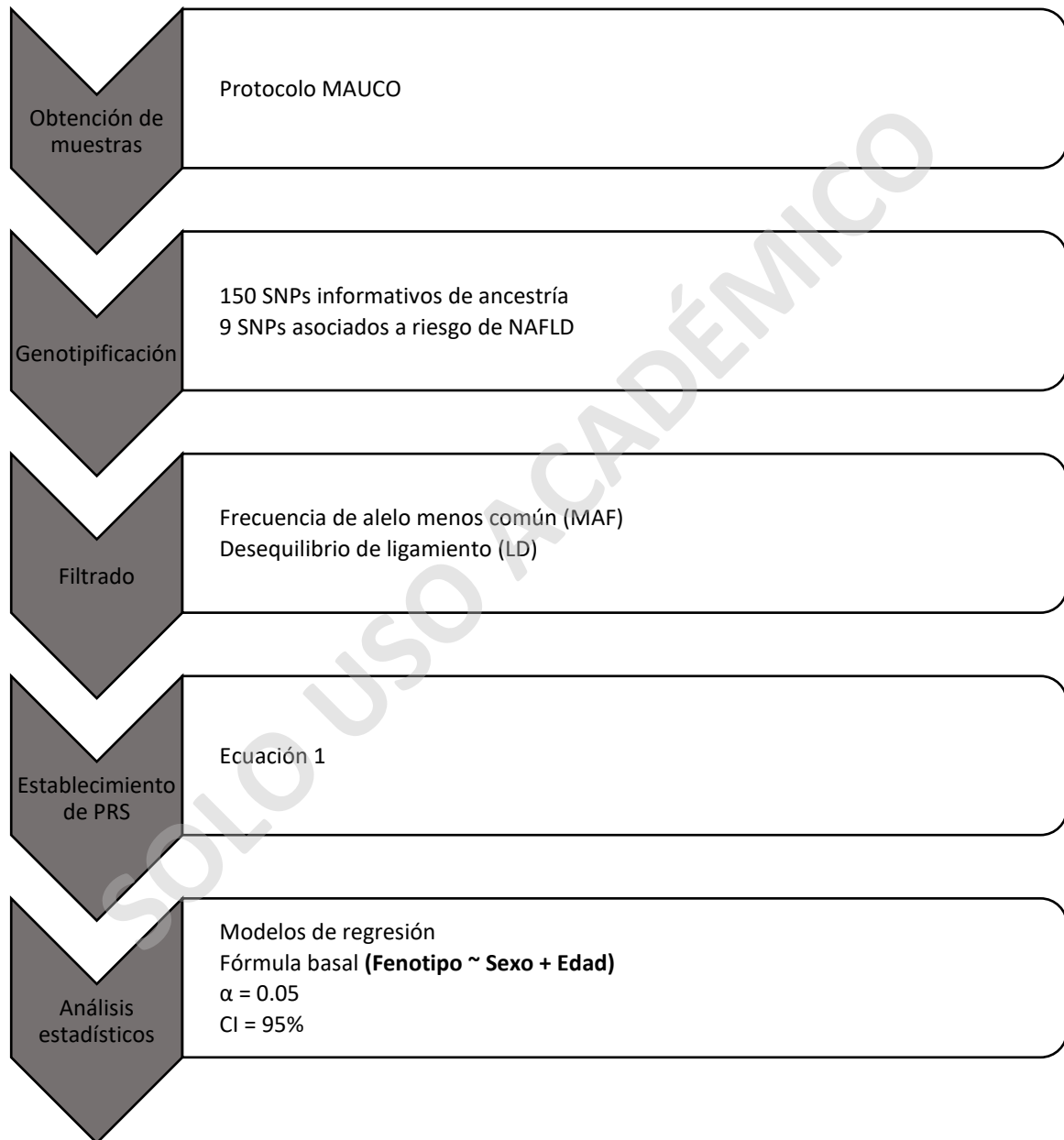
**Objetivo general:** Evaluar la replicación de asociación entre enfermedad de hígado graso no alcohólico y variantes genéticas previamente identificadas, en una población mestiza chilena.

1. **Objetivo específico 1.** Determinar las variantes genéticas asociadas a enfermedad de hígado graso no alcohólico.
2. **Objetivo específico 2.** Determinar la asociación entre el componente ancestral y la presencia de enfermedad de hígado graso.
3. **Objetivo específico 3.** Evaluar si existe interacción entre componente ancestral y las variables de riesgo asociadas a la enfermedad.

SOLO USO ACADÉMICO

### 3. MATERIALES Y MÉTODOS

El estudio fue diseñado para analizar relaciones entre la enfermedad de hígado graso no alcohólico y el genotipo de individuos pertenecientes al grupo de estudio de la cohorte MAUCO. La metodología utilizada para realizar los análisis se rigió por una secuencia lógica de pasos, detallada en el siguiente esquema (Figura 1):



**Figura 1.** Secuencia de procedimientos realizados en esta investigación

Los datos demográficos, bioquímicos y clínicos fueron obtenidos según el protocolo de estudio del Maule Cohort (MAUCO) de enfermedades crónicas. La presencia de hígado graso se determinó a través de imagenología, específicamente mediante un examen ecográfico, el que permite mediante ultrasonido detectar infiltración de tejido graso.

El perfil bioquímico de los individuos fue determinado a partir de muestras de sangre, que fueron analizadas con los reactivos de los kits de detección de Wiener Lab, de acuerdo con las instrucciones del fabricante. En los análisis que requerían la variable en forma categórica, los niveles fueron expresados de forma dicotómica como 'normal' o 'alterado', considerando este último de acuerdo con los criterios recomendados por los fabricantes de los kits de detección : ALT y AST > 35 U/L, glicemia > 105 mg/dL, triglicéridos > 150 mg/dL, HDL < 40 mg/dL en mujeres y < 50 mg/dL en varones, LDL > 160 mg/dL y presión sanguínea > 135/85 mmHg. Se consideró obeso al individuo con índice de masa corporal (IMC) igual o superior a 30 [kg/m<sup>2</sup>]. La presencia de síndrome metabólico se determinó de acuerdo con los criterios de diagnóstico de síndrome metabólico del National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III)<sup>62</sup>.

Todos los análisis estadísticos fueron ejecutados en Python 3.7.6. Para determinar la significancia estadística, se estableció  $\alpha=0,05$ . El nivel de confianza para los intervalos de confianza se determinó a 95%. Las librerías usadas fueron Pandas<sup>63</sup>, Numpy<sup>64</sup> y Scipy<sup>65</sup> en sus versiones 1.0.1, 1.18.1 y 1.4.1, respectivamente.

### **Modelos de regresión**

El modelo de regresión logística permite modelar la probabilidad  $p$  de que ocurra un evento binario en función de una o más variables explicativas. Para esto, el modelo genera una función lineal entre las variables explicativas  $x$  y el log-odd (*logarithm of odds*) de una variable de respuesta. El *odd* es la razón entre las probabilidades de un evento y su complemento (Ecuación 1). La regresión logística modela el logaritmo natural (ln) de esta expresión como una función lineal de un intercepto y una o más variables independientes (Ecuación 2). Los estimadores  $b_1$ , para el genotipo de cada SNP reportado en la literatura, fueron utilizados para estimar un *Odds Ratio* (OR) asociado a portar un alelo del SNPs (Ecuación 3). Las regresiones logísticas para las variables fenotípicas binarias fueron ajustadas en Python 3.0 con la librería StatsModel v0.11.0 con la función GLM (family = binomial) (Generalized Linear Models), que utiliza el método de los mínimos cuadrados ponderados iterativamente (IRLS por su sigla en inglés). Para las variables fenotípicas continuas, se ajustó una regresión lineal múltiple con la función OLS (*Ordinary Least Squares*) de StatsModel<sup>66</sup>. La bondad del ajuste de los modelos de regresión se determinó con el Criterio de Información de Akaike (AIC).

Los gráficos correspondientes a cada análisis se generaron con las librerías Matplotlib y Seaborn<sup>67</sup> en sus versiones 3.1.3 y 0.10.0 respectivamente.

**Ecuación 1** 
$$\text{odd} = \frac{p}{1-p}$$
  $p$  = probabilidad de ocurrencia de un evento

**Ecuación 2** 
$$\text{log-odd} = \ln\left(\frac{p}{1-p}\right) = b_0 + b_1x$$
  $p$  = probabilidad de ocurrencia de un evento  
 $b$  = coeficientes

**Ecuación 3** 
$$OR = e^{b_1}$$
  $x$  = variable explicativa

### Puntaje de Riesgo Poligénico

3.966 muestras fueron genotipadas por secuenciación dirigida para 150 SNPs informativos de ancestría y 9 SNPs de riesgo de FLD seleccionados de la literatura ( $OR \geq 1,2$  y  $p < 3 \times 10^{-5}$ ) y filtrados por frecuencia de alelo menos común (MAF) y desequilibrio de ligamiento (LD). Con un análisis por regresión logística, ajustado a edad y sexo, se evaluó asociación entre cada SNP con FLD. Se generó un PRS a partir de los SNPs filtrados, y un PRS a partir de los SNPs que mostraron asociación con FLD en la muestra chilena (PRS CL), ambos mediante la Ecuación 4. Ambos puntajes de escala continua fueron categorizados en 4 niveles de riesgo poligénico (PRL y PRL CL respectivamente), correspondientes a cuartiles de cada PRS. Los componentes ancestrales de los individuos fueron categorizados en terciles.

**Ecuación 4** 
$$PRS = \ln\left(\prod_{i=1}^n OR_i^x\right)$$
  $OR$ : Odd Ratio  
 $n$ : Multiplicatoria  
 $n$ : Número de variantes génicas asociadas a la enfermedad  
 $x$ : Número de alelos de riesgo portados por el individuo



## 4. PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

### 4.1. Características de la población

Del total de 3.966 pacientes, el 51,4% padece FLD, de los que un 85,4% no consume niveles riesgosos de alcohol. De los pacientes enfermos, la edad promedio se acerca a los 55 años, y el 68,71% son mujeres. Se observaron diferencias significativas en edad ( $p = 0,035$ ) entre casos y controles. El sexo por sí solo no mostró asociación significativa (Tabla 3), pero sí como covariable en otros modelos.

#### 4.1.1. Perfil bioquímico y condiciones clínicas

La prevalencia observada de condiciones de riesgo se resume en Tabla 3.

**Tabla 3.** Prevalencia de condiciones clínicas de riesgo en la muestra. *n*: número de individuos

	<b>MAUCO (n=3.966)</b>	<b>Hombres (n=1.291)</b>	<b>Mujeres (n=2.675)</b>
<b>Transaminasas elevadas</b>	1243 (31,34%)	553 (42,84%)	690 (25,79%)
<b>Sobrepeso u obesidad</b>	3342 (84,27%)	1084 (83,97%)	2258 (84,41%)
<b>HDL &lt; 40 mg/dL H, &lt;50 mg/dL M</b>	2294 (57,84%)	598 (46,32%)	1696 (63,40%)
<b>LDL &gt; 100 mg/dL</b>	368 (9,28%)	105 (8,13%)	263 (9,83%)
<b>TG ≥ 100 mg/dL</b>	984 (24,81%)	384 (29,74%)	600 (22,43%)
<b>SM</b>	2028 (51,13%)	606 (46,94%)	1422 (53,16%)
<b>Diabetes (autorreporte)</b>	642 (16,19%)	200 (15,49%)	442 (16,52%)

Se observó asociación entre FLD y presión sistólica (SBP) y diastólica (DBP), transaminasas hepáticas (ALT y AST), índice de masa corporal (IMC), glicemia, triglicéridos y colesterol HDL, no así con el colesterol total. IMC y transaminasas mostraron ser los mayores estimadores de riesgo de FLD (Tabla 4).

El 16,31% de los pacientes presentó un nivel plasmático elevado de AST y un 29,77% de los pacientes presentó un nivel plasmático alterado de ALT, y un 14,73% presentó ambas transaminasas elevadas. Del total de la muestra, un 21,1% de los pacientes presenta FLD y niveles elevados de ALT, y un 11,8% presenta FLD y niveles alterados de AST (Figura 2); mientras que un 31,4% presenta FLD y obesidad. Una cuarta variable que mostró una alta asociación ( $OR = 0,962$ ,  $p$

< 0,001) con FLD fue el colesterol HDL, ya que el 66,4% de los pacientes con FLD tenía niveles de HDL por debajo de lo deseable (Figura 1).

**Tabla 4:** Asociación entre NAFLD y variables demográficas, clínicas y bioquímicas. Odd ratio estimados en regresión logística, ajustada a edad y sexo.

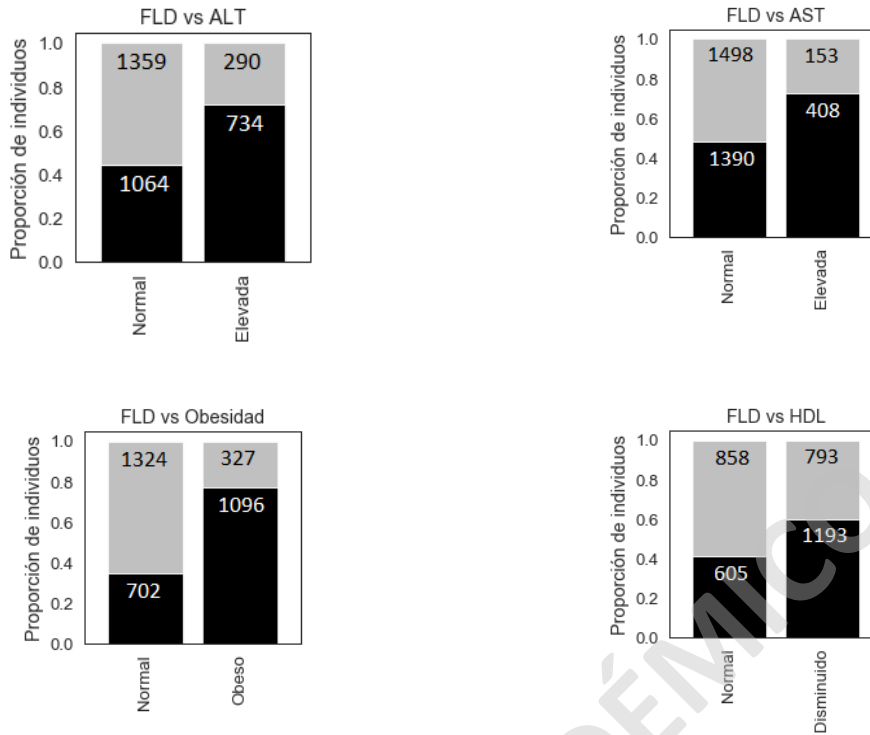
	Todos	FLD -	FLD+	OR (CI)	p
<b>Hombres (n)</b>	1.291 (-6)	647	638		
<b>Mujeres (n)</b>	2.675 (-3)	1.271	1.401	0,894 (0,782-1,022)	0,101
<b>Edad</b>	54,5 ± 0,2	54,2±0,2	54,8±0,2	<b>1,007 (1,001-1,013)</b>	<b>0,035</b>
<b>Presión sist* [mmHg]</b>	132,0 ± 0,3	130,1 ± 0,5	133,7 ± 0,4	<b>1,010 (0,993-1,007)</b>	<b>0,000</b>
<b>Presión diast* [mmHg]</b>	76,0 ± 0,2	75,0 ± 0,2	76,9 ± 0,2	<b>1,018 (1,012-1,024)</b>	<b>0,000</b>
<b>IMC [kg/m<sup>2</sup>]</b>	29,6 ± 4,9	27,2 ± 3,7	31,8 ± 4,8	<b>1,316 (1,289-1,343)</b>	<b>0,000</b>
<b>ALT [U/L]</b>	32,3 ± 0,3	26,8 ± 0,4	37,3 ± 0,5	<b>1,038 (1,033-1,042)</b>	<b>0,000</b>
<b>AST [U/L]</b>	28,3 ± 0,25	25,7 ± 0,3	30,6 ± 0,4	<b>1,031 (1,025-1,037)</b>	<b>0,000</b>
<b>Glucosa [mg/dL]</b>	103,4 ± 0,6	96,8 ± 0,7	109,4 ± 0,9	<b>1,014 (1,011-1,017)</b>	<b>0,000</b>
<b>Total col. [mg/dL]</b>	193,5 ± 0,6	192,7 ± 0,9	194,2 ± 0,9	1,000 (0,999-1,002)	0,579
<b>TG [mg/dL]</b>	166,5 ± 2,1	142,8 ± 2,5	188,3 ± 3,1	<b>1,005 (1,004-1,006)</b>	<b>0,000</b>
<b>LDL [mg/dL]</b>	115,7 ± 0,5	117,1 ± 0,8	114,3 ± 0,8	<b>0,996 (0,995-0,999)</b>	<b>0,001</b>
<b>HDL [mg/mL]</b>	45,3 ± 0,2	47,5 ± 0,3	43,4 ± 0,2	<b>0,962 (0,956-0,968)</b>	<b>0,000</b>

*Condiciones clínicas y variables bioquímicas expresados en media ± error estándar*

## 4.2. Genotipo

### 4.2.1. Variantes génicas

Cinco SNPs fueron seleccionados por su asociación a NAFLD de acuerdo con la literatura: rs2896019 (*PNPLA3*), rs2073080 (*SAMM50*), rs780094 (*GCKR*), rs4240624 (*PPP13RB*) y rs222054 (*GC*), con los que se generó un PRS. Dos de estas cinco variantes mostraron asociación a FLD en la muestra: rs2896019 y rs2073080, con una frecuencia alélica de 0,477 y 0,594 respectivamente, a partir de los que se obtuvo un PRS para la muestra (PRS CL) (Tabla 5). Los valores mínimos de PRS y PRS CL fueron de 0,0 para ambos, y los valores máximos fueron de 0,489 y 2,177 respectivamente, con medias de 0,289 y 1,044.



**Figura 2.** Proporción y número de individuos FLD- (gris) y FLD+ (negro) en presencia de transaminasas alteradas, obesidad, y HDL disminuido.

Mediante análisis de regresión logística, se obtuvo que cada PRL y PRL CL representa un aumento de un 8,38% y 11,68% de riesgo de desarrollar NAFLD respecto a los correspondientes PRL anteriores (Tabla 6). Se observa que a partir del tercer PRL, los individuos NAFLD+ superan el 50%, mientras que el mismo valor se obtiene a partir del segundo PRL CL (Figura 3).

**Tabla 5.** Frecuencias y asociación de variantes de riesgo con FLD. Odd Ratio estimados en regresión logística, ajustada a edad y sexo

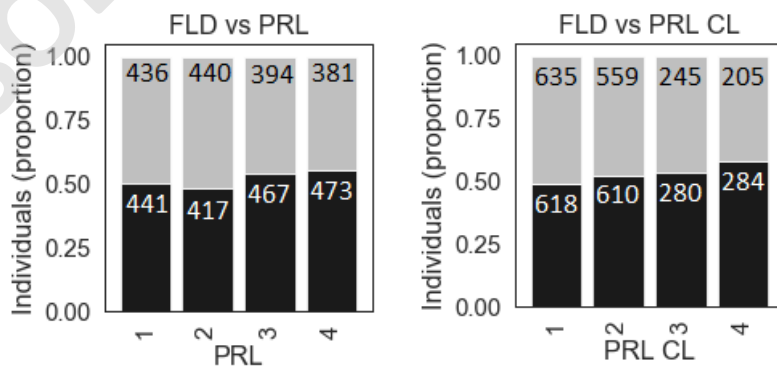
Gen	SNP	Alelos		Frecuencia alélica			OR (CI)	p
		Ref.	Ries.	MAUCO	FLD-	FLD+		
<b>PNPLA3</b>	<b>rs2896019</b>	T	G	0,477 (n=3.429)	0,500 (n=1.641)	0,457 (n=1.788)	<b>1,189</b> (1,082-1,308)	<b>0,000</b>
<b>SAMM50</b>	<b>rs2073080</b>	T	C	0,594 (n=3.414)	0,608 (n=1.631)	0,580 (n=1.783)	<b>1,126</b> (1,025-1,241)	<b>0,016</b>
<b>GCKR</b>	<b>rs780094</b>	T	C	0,667 (n=3.392)	0,674 (n=1.618)	0,661 (n=1.774)	1,060 (0,958-1,172)	0,257
<b>PP1R3B</b>	<b>rs4240624</b>	G	A	0,184 (n=3.350)	0,178 (n=1.611)	0,189 (n=1.739)	0,930 (0,822-1,052)	0,252
<b>GC</b>	<b>rs222054</b>	C	G	0,259 (n=3.431)	0,262 (n=1.638)	0,255 (n=1.793)	1,029 (0,924-1,145)	0,599

**Tabla 6.** Asociación entre FLD y PRL. Odd Ratio estimados en regresión logística, ajustada a edad y sexo.

PRL	FLD- (n)	FLD+ (n)	Estimate	OR (CI)	p	AIC
1	435 (49,61%)	442 (50,39%)	<b>0,081</b>	<b>1,083</b> <b>(1,021 – 1,150)</b>	<b>0,008</b>	<b>4.765,334</b>
2	440 (51,34%)	417 (48,66%)				
3	395 (45,88%)	466 (54,12%)				
4	381 (44,61%)	473 (55,39%)				

PRL CL	FLD- (n)	FLD+ (n)	Estimate	OR (CI)	p	AIC
1	636 (50,76%)	617 (49,24%)	<b>0,111</b>	<b>1,117</b> <b>(1,047 - 1,191)</b>	<b>0,001</b>	<b>4.742,421</b>
2	558 (47,73%)	611 (52,27%)				
3	245 (46,67%)	280 (53,33%)				
4	205 (41,92%)	284 (58,08%)				



**Figura 3.** Proporción y número de individuos FLD- (gris) y FLD+ (negro) por PRL y PRL CL

PRL y PRL CL se encontraron también asociados a niveles elevados de transaminasas hepáticas (Figura 4), pero no a las demás condiciones de riesgo asociadas a FLD (Tabla 7). Cada PRL y PRL CL mostró una probabilidad 8,2% y 17,2% mayor de presentar niveles elevados de ALT, y 14.2% y 23% mayor de presentar niveles alterados de AST respecto al correspondiente PRL anterior (Tabla 8, Figura 4).

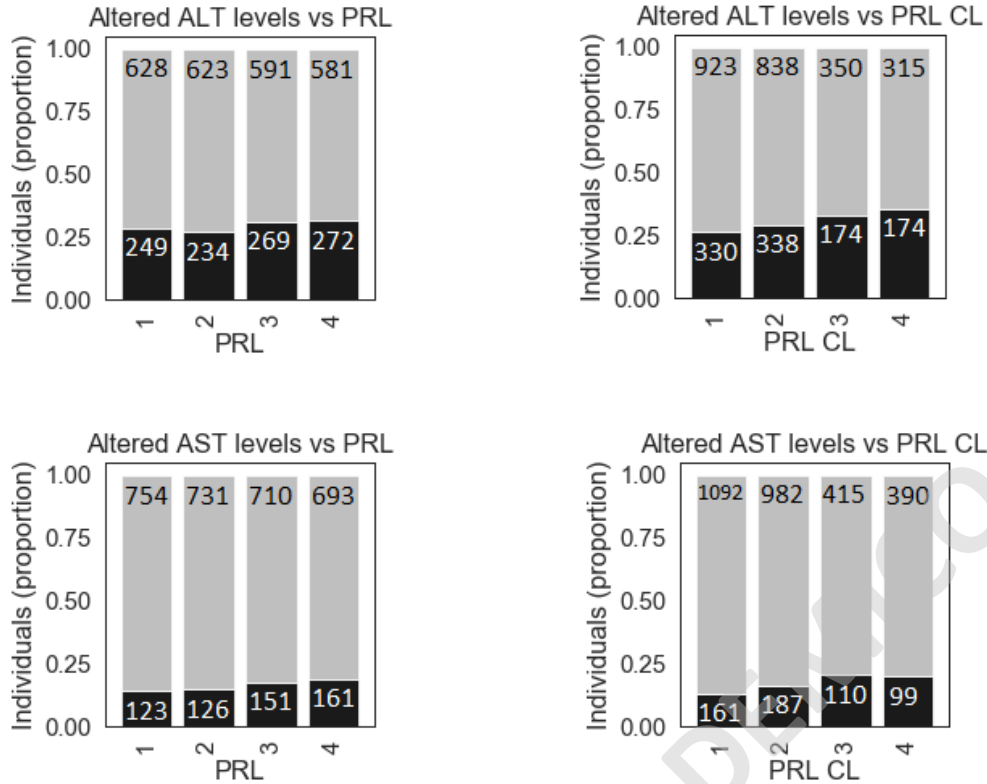
**Tabla 7.** Asociación entre condiciones clínicas de riesgo y PRL. Odd Ratio estimados en regresión logística, ajustada a edad y sexo.

	<b>Intercept</b>	<b>OR</b>	<b>CI</b>	<b>p</b>
<b>AST &gt; 35 U/L</b>	0,207	<b>1,231</b>	(1,130 - 1,340)	<b>0,000</b>
<b>ALT &gt; 35 U/L</b>	0,159	<b>1,173</b>	(1,092 - 1,259)	<b>0,000</b>
<b>Enfermedad cardiovascular</b>	-0,001	0,999	(0,882 - 1,132)	0,992
<b>Glicemia &gt; 105 mg/dL</b>	-0,004	0,996	(0,920 - 1,078)	0,919
<b>Triglicéridos</b>	-0,050	0,951	(0,882 - 1,025)	0,188
<b>HDL (&lt;40 H, &gt;50 M)</b>	0,023	1,023	(0,959 - 1,092)	0,491
<b>LDL</b>	0,052	1,053	(0,946 - 1,174)	0,345
<b>SM</b>	-0,010	0,990	(0,928 - 1,055)	0,751
<b>Obesidad</b>	-0,038	0,963	(0,901 - 1,028)	0,258

**Tabla 8.** Asociación entre alteración de niveles plasmáticos de transaminasas y PRL CL. Odd Ratio estimados en regresión logística, ajustada a edad y sexo.

ALT					
PRL CL	< 35[U/L]	> 35[U/L]	Estimate	OR (CI)	p
1	923 (73,66%)	330 (26,34%)	<b>0,159</b>	<b>1,172</b>	<b>0,000</b>
2	830 (71,06%)	338 (28,94%)		<b>(1,092-1,258)</b>	
3	350 (66,79%)	174 (33,21%)			
4	315 (64,42%)	174 (35,58%)			
AST					
1	1092 (87,15%)	161 (12,85%)	<b>0,207</b>	<b>1,230</b>	<b>0,000</b>
2	982 (84,00%)	187 (16,00%)		<b>(1,129-1,340)</b>	
3	415 (79,05%)	110 (20,95%)			
4	390 (79,75%)	99 (20,25%)			

SOLO USO ACADÉMICO



**Figura 4.** Proporción y número de individuos con niveles alterados (negro) y normales (gris) de transaminasas ALT y AST en función de PRL y PRL CL

#### 4.2.2. Componente ancestral

En la muestra, las ancestrías europea y amerindia son las más prevalentes, con una media cercana al 60% y al 35% respectivamente. Se observa que la media de ancestría europea es menor en pacientes FLD+ respecto a pacientes NAFLD- mientras que la media de ancestría amerindia en pacientes FLD+ es mayor que en pacientes NAFLD- (Tabla 9).

Ambas ancestrías mostraron estar asociadas a hígado graso, puntaje de riesgo poligénico, y transaminasas. Por cada % de ancestría amerindia, PRS aumenta 0,006, PRS CL aumenta 0,018, ALT aumenta 0,028 [U/L] y AST aumenta 0,009 [U/L]; y por cada % de ancestría europea, PRS aumenta 0,007, PRS CL aumenta 0,016, ALT aumenta 0,027 [U/L] y no se asocia de forma significativa con AST (Tabla 10).

Se evaluó la interacción entre riesgo poligénico y ancestría. Tanto PRL como PRL CL mostraron tener mayor asociación con FLD en individuos del segundo tercil de ancestría amerindia, mientras que PRL mostró asociación con FLD en segundo tercil de ancestría europea, y PRL CL en el tercero (Figura



5). En el segundo tercil de ancestría amerindia, cada PRL CL representa un riesgo de NAFLD 21,6% mayor respecto al anterior (Tabla 11). En el segundo tercil de ancestría europea, por cada PRL CL el riesgo de FLD es un 12% mayor, y en el tercer tercil, 15.4% mayor (Tabla 12).

**Tabla 9.** Asociación entre componente ancestral y FLD. Odd Ratio estimados en regresión logística, ajustada a edad y sexo.

	Min	Max	All	FLD-	FLD +	OR (CI)	p
<b>Europea</b>	0,14	0,99	0,59 ± 0,001	0,60 ± 0,002	0,59 ± 0,002	<b>0,99</b> <b>(0,98-0,99)</b>	<b>0,013</b>
<b>Amerindia</b>	0,00	0,84	0,35 ± 0,001	0,34 ± 0,002	0,36 ± 0,002	<b>1,01</b> <b>(1,00-1,02)</b>	<b>0,015</b>

**Tabla 10,** Asociación entre componente ancestral, PRS y transaminasas. Valores estimados en regresión lineal, ajustada a edad y sexo.

		r <sup>2</sup>	Coef	CI	p	AIC
<b>Amerindia</b>	<b>PRS</b>	0,006	<b>0,0098</b>	0,006 - 0,014	<b>0,000</b>	932
	<b>PRS CL</b>	0,018	<b>0,0113</b>	0,008 - 0,014	<b>0,000</b>	730
	<b>ALT</b>	0,028	<b>0,1066</b>	0,018 - 0,195	<b>0,018</b>	30.850
	<b>AST</b>	0,009	<b>0,0731</b>	0,013 - 0,134	<b>0,018</b>	28.240
<b>Europea</b>	<b>PRS</b>	0,007	<b>-0,0094</b>	-0,013, -0,006	<b>0,000</b>	932
	<b>PRS CL</b>	0,016	<b>-0,0098</b>	-0,013, -0,007	<b>0,000</b>	731
	<b>ALT</b>	0,027	<b>-0,0882</b>	-0,171, -0,005	<b>0,038</b>	30.850
	<b>AST</b>	0,009	-0,0509	-0,108, 0,006	0,079	28.240

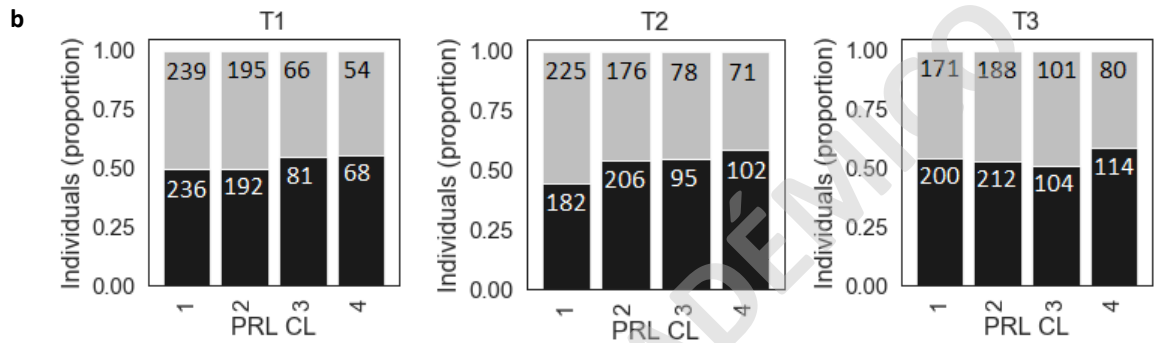
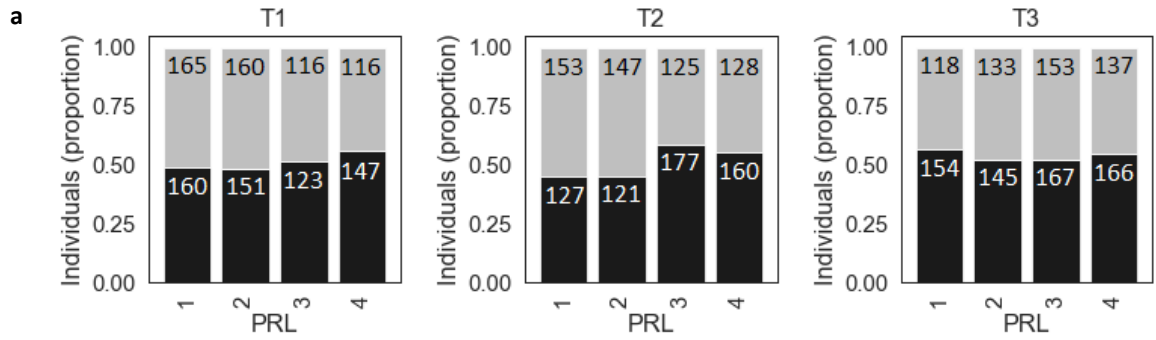
**Tabla 11.** Asociación entre NAFLD y PRL por tercil de ancestría amerindia. Odd Ratio estimados en regresión logística, ajustada a edad y sexo.

		Estimate	OR	CI	p
<b>T1</b>	<b>PRL</b>	0,078	1,081	(0,975-1,199)	0,138
	<b>AIC: 1589.259</b>				
≥ 0,00					
≤ 0,321					
	<b>PRL CL</b>	0,086	1,089	(0,968-1,227)	0,154
	<b>AIC: 1579.366</b>				
<b>T2</b>	<b>PRL</b>	0,194	<b>1,214</b>	(1,094-1,349)	<b>0,000</b>
	<b>AIC: 1585.778</b>				
> 0,321					
≤ 0,385					
	<b>PRL CL</b>	0,202	<b>1,223</b>	(1,094-1,368)	<b>0,000</b>
	<b>AIC: 1582.204</b>				
<b>T3</b>	<b>PRL</b>	-0,041	0,960	(0,864-1,066)	0,446
	<b>AIC: 1591,060</b>				
> 0,385					
≤ 0,845					
	<b>PRL CL</b>	0,029	1,029	(0,922-1,149)	0,603
	<b>AIC: 1587,152</b>				

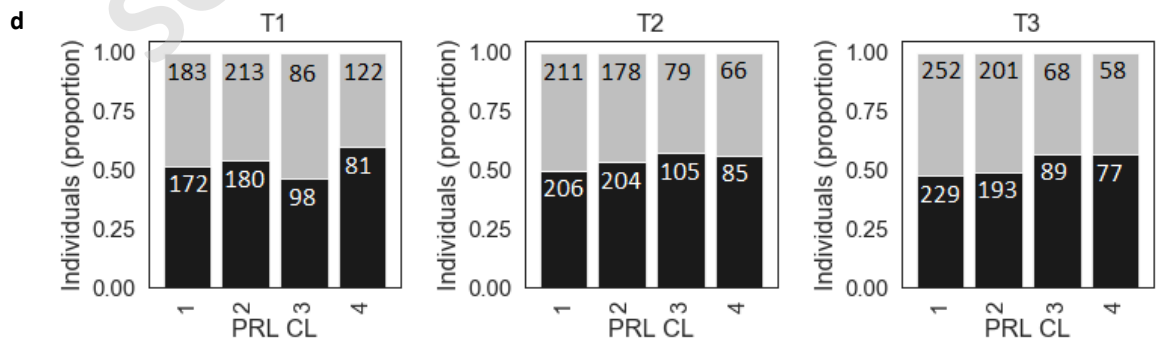
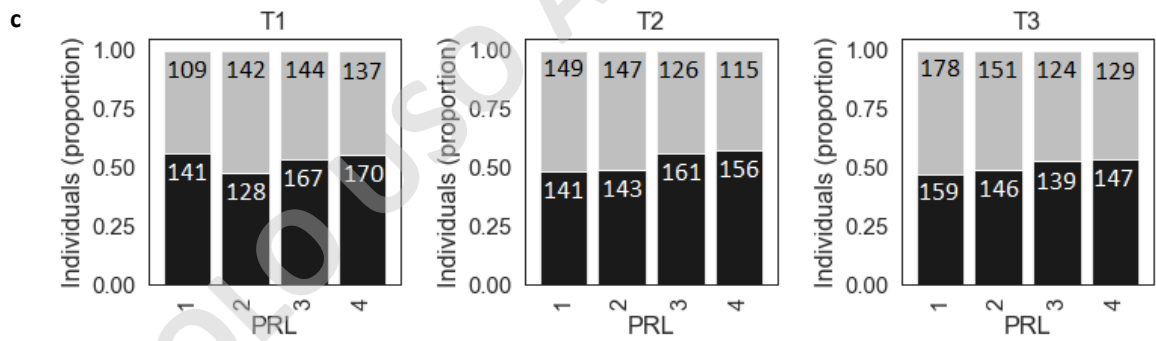
**Tabla 12.** Asociación entre NAFLD y PRL por tercil de ancestría europea. Odd Ratio estimados en regresión logística, ajustada a edad y sexo.

		Estimate	OR	CI	p
<b>T1</b>	<b>PRL</b>	0,008	1,008	(0,908-1,120)	0,886
	AIC: 1.596,330				
	<b>PRL CL</b>	0,070	1,072	(0,961-1,196)	0,207
		AIC: 1.590,575			
<b>T2</b>	<b>PRL</b>	0,148	<b>1,159</b>	(1,044-1,288)	<b>0,006</b>
	AIC: 1.585,122				
	<b>PRL CL</b>	0,114	<b>1,120</b>	(1,001-1,254)	<b>0,049</b>
		AIC: 1.583,063			
<b>T3</b>	<b>PRL</b>	0,080	1,084	(0,978-1,201)	0,124
	AIC: 1.592,831				
	<b>PRL CL</b>	0,144	<b>1,154</b>	(1,027-1,298)	<b>0,016</b>
		AIC: 1.580,870			

### AMERINDIA



### EUROPEA



**Figura 5:** Proporción y número de casos (negro) y controles (gris) de hígado graso en función de PRL (**a** y **c**) y PRL CL (**b** y **d**) por tercil de ancestría amerindia y europea

## 5. DISCUSIÓN DE RESULTADOS

### 5.1. Variables demográficas

La mayor prevalencia de la enfermedad de hígado graso se observó en mujeres y en pacientes con una edad promedio de 55 años. Lonardo y col.<sup>7</sup> reportaron resultados similares, sugiriendo que el estrógeno puede ser un factor protector, ya que en su estudio observaron que la prevalencia de la enfermedad en edad reproductiva es mayor en hombres que en mujeres. Una forma de confirmar si esto es lo que ocurre en la población estudiada de MAUCO sería dirigir estudios similares en pacientes menores de 47,5 años, que corresponde a la edad promedio en que se manifiesta la menopausia en Chile<sup>68</sup>.

La prevalencia de la enfermedad de hígado graso en esta muestra (51,4%) fue más del doble de lo reportado por Riquelme y cols<sup>60</sup> (23%) y por el Ministerio de Salud<sup>58</sup>(5,2%), aunque es importante tener en cuenta que, tanto los datos recopilados de MAUCO como los del MINSAL no tienen una diferenciación entre individuos con hígado graso causado por consumo de alcohol y los individuos con hígado graso no alcohólico, mientras que Riquelme se enfocó exclusivamente en hígado graso no alcohólico. Por otro lado, el Ministerio de Salud reportó el número de casos de acuerdo con el auto reporte de la enfermedad, sin diferenciar enfermedad de hígado graso de cirrosis o daño hepático crónico, mientras que los encontrados en MAUCO fueron contabilizados de acuerdo con un examen ecográfico. La diferencia entre estos parámetros de medición puede implicar que los resultados no sean comparables entre sí.

Un resultado sorprendente fue el alto número de pacientes con daño hepático según niveles de transaminasas: un 31,34% de los individuos tenía al menos una de las transaminasas hepáticas en niveles elevados, en comparación al 18,9% reportado en Chile y un 10,6% reportado en la región del Maule, según la información del Ministerio de Salud<sup>58</sup>. Sin embargo, es importante considerar que los resultados entregados por el MINSAL fueron obtenidos a partir de estudios en muestras poblacionales muy pequeñas, no sólo a nivel regional (n=29) sino que también a nivel nacional (n=778), en comparación a los 3.966 individuos de este estudio (Tabla 1 y 2).

### 5.2. Perfil bioquímico y condiciones clínicas.

La forma más certera de diagnosticar lesión hepática es a través de la medición de niveles sanguíneos de alanina aminotransferasa<sup>10</sup>, lo cual se condice con la asociación entre las transaminasas ALT y AST con hígado graso, tanto en la diferencia de sus medias como con los coeficientes de regresión logística obtenidos en este estudio (Tabla 2 y 3). La presencia de obesidad

determinada por el índice de masa corporal también mostró asociación significativa con la presencia de hígado graso, tal como ha sido reportado en estudios previos<sup>8</sup>.

### 5.3. Genotipo

#### 5.3.1. Riesgo poligénico

En el presente estudio, se consideraron nueve SNPs cuya asociación a hígado graso no alcohólico se ha reportado frecuentemente, de los que cinco fueron seleccionados tras pasar por filtro de frecuencia de alelo mínimo y desequilibrio de ligamiento para evaluar la replicación de asociación a la enfermedad. Sólo dos SNPs rs2896019 (*PNPLA3*) y rs2073080 (*SAMM50*) replicaron asociación con hígado graso en la muestra chilena, que coinciden con variantes reportadas por Larrieta-Carrasco<sup>21</sup> (Tabla 4). De acuerdo con estudios previos, la frecuencia alélica de la variante rs2896019 es similar a la reportada en la población chilena, y la frecuencia alélica de rs2073080 se acerca a la reportada para la población mapuche<sup>61</sup>. Las frecuencias de las otras tres variantes difieren a nivel global y respecto a la población europea, lo que puede explicar la ausencia de asociación entre ellas y la presencia de hígado graso en esta población (Tabla suplementaria 1). El nivel de riesgo poligénico obtenido a partir de estos dos SNPs mostró tener mayor asociación con hígado graso (OR = 1,117,  $p=0,001$ ) al compararlo con el nivel de riesgo poligénico obtenido a partir de los 5 SNPs en la muestra de población chilena (OR = 1,083,  $p = 0,008$ ), pero estos resultados se observaron sólo en individuos de ancestría europea mayor a 56,4% (Tabla 5). Esto también se condice con los estudios que, en su conjunto, han demostrado que el riesgo poligénico de desarrollar la enfermedad de hígado graso no alcohólico se manifiesta de forma diferencial entre grupos étnicos distintos<sup>22</sup>.

Las variantes utilizadas para calcular el puntaje de riesgo poligénico en la muestra se encuentran asociados a niveles alterados de transaminasas ALT y AST<sup>21</sup>, lo que fue corroborado en el presente estudio (rs2896019:  $\chi^2 = 22,87$ ,  $p<0,001$ ; rs2073080:  $\chi^2 = 9,94$ ,  $p=0,007$ ) y AST (rs2896019:  $\chi^2 = 28,59$ ,  $p<0,001$ ; rs2073080:  $\chi^2 = 13,24$ ,  $p=0,001$ ) (Tabla suplementaria 3). De forma concordante, el nivel de riesgo poligénico obtenido a partir de estas dos variantes mostró estar asociado de forma significativa tanto a niveles alterados de ALT (OR = 1,172,  $p<0,001$ ) como de AST (OR = 1,230,  $p<0,001$ ) (Tabla 4 y Figura 4) y no a la presencia de obesidad ni otros factores de riesgo, que sí mostraron asociación con las otras variantes (Tabla suplementaria 3). Estos resultados son comparables a los obtenidos por Larrieta-Carrasco<sup>21</sup> en el estudio de riesgo poligénico para estimar riesgo de hígado graso en población mexicana. Estas condiciones de riesgo no relacionadas a PRL, pero sí a la enfermedad de hígado graso, se deben a que el desarrollo de la enfermedad se debió

probablemente a hábitos de riesgo y no a un componente genético, acorde a lo expuesto por Kim y col<sup>11</sup>.

### 5.3.2. Componente ancestral

La media del componente tanto amerindio como europeo se acerca a la descrita por Verdugo y col.<sup>61</sup>. Ambas ancestrías mostraron estar asociadas a hígado graso, puntaje de riesgo poligénico, y transaminasas, siendo una correlación positiva con ancestría amerindia y correlación negativa con ancestría europea, salvo con AST con la que el componente europeo no mostró asociación significativa. Los niveles de riesgo poligénico obtenidos a partir de los cuartiles de puntajes mostraron estar asociados con padecimiento de hígado graso en los individuos con un 32% a 38.5% de ancestría amerindia (PRL: OR = 1,19,  $p=0,001$ ; PRL CL: OR = 1,216,  $p=0,001$ ), mientras que sólo PRL CL en individuos sobre el 56.4% de ancestría europea (OR = 1,148,  $p=0,019$ ), lo que sugiere que la asociación entre PRL CL y la presencia de hígado graso se relaciona al perfil ancestral promedio de la población, donde la ancestría es predominantemente europea. Estos resultados sugieren que los marcadores de riesgo reportados en la literatura para hígado graso no alcohólico tienen un sesgo por ancestría europea. Estas diferencias de asociación entre puntaje de riesgo, transaminasas y ancestrías puede explicarse por las diferencias en la frecuencia de las variantes de riesgo. La frecuencia de la variante rs2896019 en *PNPLA3* en la muestra (0,477) es significativamente distinta a la reportada para población europea (0,801) como mapuche (0,907), por lo que el impacto que esta variante pueda tener difiere considerablemente en este grupo, que contiene el perfil ancestral promedio de la muestra. Por otro lado, la frecuencia de la variante rs2073080 en *SAMM50* (0,594) se acerca más a la reportada para la población europea (0,789) y mapuche (0,611), por lo que su impacto puede ser comparable (Tabla 4 y Tabla suplementaria 1).

### 5.4. Aplicaciones biotecnológicas

Profundizar en esta área permitiría desarrollar kits de ensayos clínicos para estimación de riesgo de daño hepático asociado a alteraciones en los niveles de transaminasas hepáticas, así como de desarrollar la enfermedad de hígado graso no alcohólico, y complementarlo con factores ambientales para pronosticar el estado de salud del hígado en los pacientes. La información recopilada a largo plazo sobre el puntaje de riesgo poligénico en una población puede permitir a los servicios de salud implementar tecnologías de análisis de datos donde el puntaje de riesgo poligénico sea una herramienta tanto de pronóstico, así como de estratificación clínica; o bien generar productos comercializables de forma directa al consumidor para conocer sus probabilidades de desarrollar la enfermedad de hígado graso no alcohólico, lo que significa una ventaja respecto a los métodos que reportan la situación de salud actual. Sin embargo, los hallazgos

de este trabajo indican que, para encontrar las variantes genéticas más relevantes asociadas a la enfermedad de hígado graso no alcohólico, es necesario realizar estudios de asociación de genoma completo en la población chilena.

SOLO USO ACADÉMICO



## 6.CONCLUSIONES

Sólo dos variantes de las cinco reportadas replicaron su asociación con enfermedad de hígado graso: rs2896019, en el gen *PNPLA3*, y rs2073080 en el gen *SAMM50*, indicando que existen diferencias de asociación entre las variantes reportadas para NAFLD en la literatura y las mismas variantes para la población chilena. Un puntaje de riesgo poligénico a partir de estas dos variantes entrega información sobre el riesgo de desarrollar enfermedad de hígado graso sólo en individuos con ancestría europea mayor al 56.4%, pero la asociación con el nivel de aspartato aminotransferasa ocurrió en todos los individuos. Esta expresión de riesgo, complementada con otros factores de riesgo tales como dislipidemias y obesidad, permite hacer una predicción sobre el riesgo de desarrollar hígado graso no alcohólico. Los resultados obtenidos con un PRS en base a las cinco variantes no mostraron mayor asociación a riesgo de enfermedad de hígado graso ni de transaminasas hepáticas elevadas respecto al PRS con sólo las dos variantes, que replicaron su asociación. La asociación del PRS con el riesgo de NAFLD fue mayor en individuos con alta ancestría europea, dando cuenta del sesgo prevalente que existe en los estudios de asociación de genoma completo en poblaciones con ese origen. Este es el primer estudio del riesgo genético de enfermedad de hígado graso no alcohólico en una población chilena mixta europea amerindia. Futuros estudios serán necesarios para replicar la asociación de otras variantes no evaluadas en el presente trabajo. Estos hallazgos indican la necesidad de realizar estudios de asociación de genoma completo en población chilena para poder detectar la variación genética asociada a riesgo con un origen distinto al europeo y que pueda ser más prevalente en nuestra población mestiza.

## 7. IMPLICANCIAS, RECOMENDACIONES Y PROYECCIONES FUTURAS

Las observaciones realizadas en este trabajo hacen posible la creación de kits de ensayos clínicos que permitan estimar riesgo de lesión hepática antes de su manifestación, ya sea estimando el riesgo de alteración en los niveles plasmáticos de transaminasas hepáticas o el riesgo de desarrollar la enfermedad de hígado graso no alcohólico. Los resultados de ensayos clínicos de este tipo permitirían al paciente tomar conocimiento y así adoptar medidas preventivas tanto en estilo de vida como alimentación, así como en controles médicos periódicos. A largo plazo, este tipo de medidas podría reducir la prevalencia de enfermedad de hígado graso no alcohólico y de lesión hepática causadas por riesgo genético, reduciendo costos en salud y significando una mejora en la calidad de vida de la población chilena.

Un resultado que debe ser estudiado a mayor profundidad es la gran prevalencia que mostró tener la presencia de daño hepático, tanto por la enfermedad de hígado graso como por alteraciones en los niveles de transaminasas, que además mostró ser significativamente mayor que en el resto del país. Esto sugiere que los estudios realizados hasta ahora por organismos públicos han sido en tamaños muestrales muy pequeños, dejando a la población subrepresentada. La situación debiera ser evaluada por un equipo multidisciplinario para poder abarcar la situación de salud, social, económica y cultural de cada región para estimar la realidad de una forma realmente representativa.

Los resultados de este trabajo abren oportunidades de investigación sobre el uso de polimorfismos en genes clave para estimar riesgo, de forma cuantitativa, para el desarrollo enfermedades frecuentes en población chilena, ya que deja en evidencia que las variantes génicas se distribuyen en distintas frecuencias, manifestando diferentes efectos entre grupos étnicos como los niveles de riesgo de desarrollo de la enfermedad. Lo anterior indica la necesidad de profundizar en el conocimiento y entendimiento de este campo mediante estudios de asociación de genoma completo en población chilena, y así poder identificar las variantes de riesgo más relevantes en el grupo étnico que representa nuestro país. El almacenamiento de datos obtenidos sobre marcadores de riesgo asociados a enfermedad de hígado graso en la población chilena, así como otras enfermedades, puede ser un recurso valioso para la creación de programas de salud enfocados a pronosticar y estratificar individuos y así tomar medidas preventivas en aquellas personas que se encuentren con mayor riesgo de desarrollar la enfermedad.

## 8. REFERENCIAS

1. Chalasani, N., Younossi, Z., Lavine, J., Diehl, A., Brunt, E., Cusi, K., Charlton, M. and Sanyal, A., 2012. The Diagnosis and Management of Non-alcoholic Fatty Liver Disease: Practice Guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology*, 142(7), pp.1592-1609.
2. American College of Gastroenterology. 2021. *Liver Transplantation - American College of Gastroenterology*. [online] Available at: <<https://gi.org/topics/liver-transplantation/>> [Accessed 31 January 2021].
3. Del Campo, J., Gallego-Durán, R., Gallego, P. and Grande, L., 2018. Genetic and Epigenetic Regulation in Nonalcoholic Fatty Liver Disease (NAFLD). *International Journal of Molecular Sciences*, 19(3), p.911.
4. ANGULO, P., 2007. GI Epidemiology: nonalcoholic fatty liver disease. *Alimentary Pharmacology & Therapeutics*, 25(8), pp.883-889
5. Summart, U., Thinkhamrop, B., Chamadol, N., Khuntikeo, N., Songthamwat, M. and Kim, C., 2017. Gender differences in the prevalence of nonalcoholic fatty liver disease in the Northeast of Thailand: A population-based cross-sectional study. *F1000Research*, 6, p.1630.
6. Lonardo, A., Nascimbeni, F., Ballestri, S., Fairweather, D., Win, S., Than, T., Abdelmalek, M. and Suzuki, A., 2019. Sex Differences in Nonalcoholic Fatty Liver Disease: State of the Art and Identification of Research Gaps. *Hepatology*, 70(4), pp.1457-1469.
7. Oh RC, Husted TR, Ali SM, Pantsari MW. Mildly Elevated Liver Transaminase Levels: Causes and Evaluation. 2017;96(11):7.
8. Kang K-S. Abnormality on Liver Function Test. *Pediatr Gastroenterol Hepatol Nutr*. 2013;16(4):225.
9. Luyckx, F., Desai, C., Thiry, A., Dewé, W., Scheen, A., Gielen, J. and Lefèbvre, P., 1998. Liver abnormalities in severely obese subjects: Effect of drastic weight loss after gastroplasty. *International Journal of Obesity*, 22(3), pp.222-226.
10. Kim, D., Touros, A. and Kim, W., 2018. Nonalcoholic Fatty Liver Disease and Metabolic Syndrome. *Clinics in Liver Disease*, 22(1), pp.133-140.
11. Hamaguchi, M., Takeda, N. and Kojima, T., 2006. The Metabolic Syndrome as a Predictor of Nonalcoholic Fatty Liver Disease. *Annals of Internal Medicine*, 144(5), p.380.

12. Chalasani, N., Guo, X., Loomba, R., Goodarzi, M., Haritunians, T., Kwon, S., Cui, J., Taylor, K., Wilson, L., Cummings, O., Chen, Y. and Rotter, J., 2010. Genome-Wide Association Study Identifies Variants Associated With Histologic Features of Nonalcoholic Fatty Liver Disease. *Gastroenterology*, 139(5), pp.1567-1576.e6.
13. Lasky-Su, J., 2017. Statistical Techniques for Genetic Analysis. *Clinical and Translational Science*, pp.347-362
14. Calabrese, B., 2019. Linkage Disequilibrium. *Encyclopedia of Bioinformatics and Computational Biology*, pp.763-765
15. Dong, X., 2019. PNPLA3—A Potential Therapeutic Target for Personalized Treatment of Chronic Liver Disease. *Frontiers in Medicine*, 6
16. Chung, G., Lee, Y., Yim, J., Choe, E., Kwak, M., Yang, J., Park, B., Lee, J., Kim, J. and Kim, J., 2018. Genetic Polymorphisms of PNPLA3 and SAMM50 Are Associated with Nonalcoholic Fatty Liver Disease in a Korean Population. *Gut and Liver*, 12(3), pp.316-323.
17. Valenti L, Baselli G. Genetics of Nonalcoholic Fatty Liver Disease: A 2018 Update. *Current Pharmaceutical Design*. 2019;24(38):4566-4573.
18. Seko Y, Yamaguchi K, Itoh Y. The genetic backgrounds in nonalcoholic fatty liver disease. *Clinical Journal of Gastroenterology*. 2018;11(2):97-102.
19. Sookoian S, Pirola C. Genetic Basis of Alcoholic and Nonalcoholic Fatty Liver Disease. *Alcoholic and Non-Alcoholic Fatty Liver Disease*. 2016;:147-163.
20. Alcoholic and Non-Alcoholic Fatty Liver Disease: Bench to Bedside [Internet]. Cham: Springer International Publishing; 2016 [citado 16 de noviembre de 2020]. p. 147-63.
21. Larrieta-Carrasco E, Flores Y, Macías-Kauffer L, Ramírez-Palacios P, Quiterio M, Ramírez-Salazar E et al. Genetic variants in COL13A1, ADIPOQ and SAMM50 , in addition to the PNPLA3 gene, confer susceptibility to elevated transaminase levels in an admixed Mexican population. *Experimental and Molecular Pathology*. 2018;104(1):50-58.
22. Szanto K, Li J, Cordero P, Oben J. <p>Ethnic differences and heterogeneity in genetic and metabolic makeup contributing to nonalcoholic fatty liver disease</p>. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 2019;Volume 12:357-367.
23. PNPLA3 patatin like phospholipase domain containing 3 [Homo sapiens (human)] - Gene - NCBI [Internet]. Ncbi.nlm.nih.gov. 2021 [cited 7 February 2021]. Available from: <https://www.ncbi.nlm.nih.gov/gene/80339>

24. Islek E, Sazci A, Ozel M, Aygun C. Genetic Variants in the PNPLA3 Gene Are Associated with Nonalcoholic Steatohepatitis. *Genetic Testing and Molecular Biomarkers*. 2014;18(7):489-496.
25. Delik A, Akkız H, Dinçer S. The effect of PNPLA3 polymorphism as gain in function mutation in the pathogenesis of non-alcoholic fatty liver disease. *Indian Journal of Gastroenterology*. 2020;39(1):84-91.
26. Kitamoto T, Kitamoto A, Yoneda M, Hyogo H, Ochi H, Nakamura T et al. Genome-wide scan revealed that polymorphisms in the PNPLA3, SAMM50, and PARVB genes are associated with development and progression of nonalcoholic fatty liver disease in Japan. *Human Genetics*. 2013;132(7):783-792.
27. SAMM50 sorting and assembly machinery component [Homo sapiens (human)] - Gene - NCBI [Internet]. Ncbi.nlm.nih.gov. 2021 [cited 7 February 2021]. Available from: <https://www.ncbi.nlm.nih.gov/gene/25813>
28. Ram R, Wakil S, Muiya N, Andres E, Mazhar N, Hagos S et al. A common variant association study in ethnic Saudi Arabs reveals novel susceptibility loci for hypertriglyceridemia. *Clinical Genetics*. 2017;91(3):371-378
29. Yang Z, Wen J, Tao X, Lu B, Du Y, Wang M et al. Genetic variation in the GCKR gene is associated with non-alcoholic fatty liver disease in Chinese people. *Molecular Biology Reports*. 2010;38(2):1145-1150.
30. Mohammadi S, Farajnia S, Shadmand M, Mohseni F, Baghban R. Association of rs780094 polymorphism of glucokinase regulatory protein with non-alcoholic fatty liver disease. *BMC Research Notes*. 2020;13(1)
31. PPARG peroxisome proliferator activated receptor gamma [Homo sapiens (human)] - Gene - NCBI [Internet]. Ncbi.nlm.nih.gov. 2021 [cited 7 February 2021]. Available from: <https://www.ncbi.nlm.nih.gov/gene/5468>
32. Tellechea M, Aranguren F, Pérez M, Cerrone G, Frechtel G, Taverna M. Pro12Ala Polymorphism of the Peroxisome ProliferatorActivated Receptor- $\gamma$  Gene is Associated With Metabolic Syndrome and Surrogate Measures of Insulin Resistance in Healthy Men. *Circulation Journal*. 2009;73(11):2118-2124.
33. IL6 interleukin 6 [Homo sapiens (human)] - Gene - NCBI [Internet]. Ncbi.nlm.nih.gov. 2021 [cited 7 February 2021]. Available from: <https://www.ncbi.nlm.nih.gov/gene/3569>
34. Wieckowska A, Papouchado B, Li Z, Lopez R, Zein N, Feldstein A. Increased Hepatic and Circulating Interleukin-6 Levels in Human Nonalcoholic Steatohepatitis. *The American Journal of Gastroenterology*. 2008;103(6):1372-1379.

35. TCF7L2 transcription factor 7 like 2 [Homo sapiens (human)] - Gene - NCBI [Internet]. Ncbi.nlm.nih.gov. 2021 [cited 7 February 2021]. Available from: <https://www.ncbi.nlm.nih.gov/gene/6934>
36. PPP1R3B protein phosphatase 1 regulatory subunit 3B [Homo sapiens (human)] - Gene - NCBI [Internet]. Ncbi.nlm.nih.gov. 2021 [cited 7 February 2021]. Available from: <https://www.ncbi.nlm.nih.gov/gene/79660>
37. Stender S, Smagris E, Lauridsen B, Kofoed K, Nordestgaard B, Tybjaerg-Hansen A et al. Relationship between genetic variation at PPP1R3B and levels of liver glycogen and triglyceride. *Hepatology*. 2018;67(6):2182-2195.
38. Palmer N, Musani S, Yerges-Armstrong L, Feitosa M, Bielak L, Hernaez R et al. Characterization of european ancestry nonalcoholic fatty liver disease-associated variants in individuals of african and hispanic descent. *Hepatology*. 2013;58(3):966-975.
39. GC vitamin D binding protein [Homo sapiens (human)] - Gene - NCBI [Internet]. Ncbi.nlm.nih.gov. 2021 [cited 7 February 2021]. Available from: <https://www.ncbi.nlm.nih.gov/gene/2638>
40. Pan J. Gender and racial differences in nonalcoholic fatty liver disease. *World Journal of Hepatology*. 2014;6(5):274.
41. Shieh Y, Hu D, Ma L, Huntsman S, Gard C, Leung J et al. Breast cancer risk prediction using a clinical risk model and polygenic risk score. *Breast Cancer Research and Treatment*. 2016;159(3):513-525.
42. How it Works - 23andMe UK [Internet]. 23andme.com. 2021 [cited 7 February 2021]. Available from: <https://www.23andme.com/howitworks/>
43. 23andMe thinks polygenic risk scores are ready for the masses, but experts aren't so sure [Internet]. MIT Technology Review. 2021 [cited 7 February 2021]. Available from: <https://www.technologyreview.com/2019/03/08/136730/23andme-thinks-polygenic-risk-scores-are-ready-for-the-masses-but-experts-arent-so-sure/>
44. Lewis C, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*. 2020;12(1).
45. Iqbal U, Perumpail B, Akhtar D, Kim D, Ahmed A. The Epidemiology, Risk Profiling and Diagnostic Challenges of Nonalcoholic Fatty Liver Disease. *Medicines*. 2019;6(1):41
46. Bril F, McPhaul M, Caulfield M, Castille J, Poynard T, Soldevila-Pico C et al. Performance of the SteatoTest, ActiTest, NashTest and FibroTest in a multiethnic cohort of patients with type 2 diabetes mellitus. *Journal of Investigative Medicine*. 2018;67(2):303-311.

47. Non-alcoholic Fatty Liver Disease (NAFLD) - American College of Gastroenterology [Internet]. American College of Gastroenterology. 2021 [cited 7 February 2021]. Available from: <https://gi.org/topics/fatty-liver-disease-nafld/>
48. RATZIU V, GIRAL P, MUNTEANU M, MESSOUS D, MERCADIER A, BERNARD M et al. Screening for liver disease using non-invasive biomarkers (FibroTest, SteatoTest and NashTest) in patients with hyperlipidaemia. *Alimentary Pharmacology & Therapeutics*. 2006;25(2):207-218
49. Lydatakis H, Hager I, Kostadelou E, Mpousmpoulas S, Pappas S, Diamantis I. Non-invasive markers to predict the liver fibrosis in non-alcoholic fatty liver disease. *Liver International*. 2006;26(7):864-871.
50. Angulo P, Bugianesi E, Bjornsson E, Charatcharoenwitthaya P, Mills P, Barrera F et al. Simple Noninvasive Systems Predict Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2013;145(4):782-789.e4.
51. Ratziu V, Massard J, Charlotte F, Messous D, Imbert-Bismut F, Bonyhay L et al. Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterology*. 2006;6(1).
52. Lewis C, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome Medicine*. 2017;9(1).
53. Martin A, Kanai M, Kamatani Y, Okada Y, Neale B, Daly M. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*. 2019;51(4):584-591.
54. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*. 2019;10(1).
55. World Development Indicators | DataBank [Internet]. [Databank.worldbank.org](https://databank.worldbank.org/reports.aspx?source=2). 2021 [cited 7 February 2021]. Available from: <https://databank.worldbank.org/reports.aspx?source=2>
56. Informe Anual 2019 - Centro Avanzado de Enfermedades Crónicas [Internet]. Centro Avanzado de Enfermedades Crónicas. 2021 [cited 7 February 2021]. Available from: <https://www.accdis.cl/informe-anual-2019/>
57. Encuesta Nacional de Salud 2016-2017 [Internet]. EPI – Departamento de Epidemiología. 2021 [cited 7 February 2021]. Available from: <http://epi.minsal.cl/encuesta-nacional-de-salud-2015-2016/>
58. Departamento de Estadísticas e Información de Salud [Internet]. [Deis.minsal.cl](https://deis.minsal.cl/). 2021 [cited 7 February 2021]. Available from: <https://deis.minsal.cl/>
59. Ferreccio C, Roa J, Bambs C, Vives A, Corvalán A, Cortés S et al. Study protocol for the Maule Cohort (MAUCO) of chronic diseases, Chile 2014–2024. *BMC Public Health*. 2015;16(1).

60. Riquelme A, Arrese M, Soza A, Morales A, Baudrand R, Pérez-Ayuso R et al. Non-alcoholic fatty liver disease and its association with obesity, insulin resistance and increased serum levels of C-reactive protein in Hispanics. *Liver International*. 2009;29(1):82-88.
61. Verdugo R, Di Genova A, Herrera L, Moraga M, Acuña M, Berríos S et al. Development of a small panel of SNPs to infer ancestry in Chileans that distinguishes Aymara and Mapuche components. *Biological Research*. 2020;53(1).
62. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA*. 16 de mayo de 2001;285(19):2486-97.
63. McKinney W. *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference. 2010;.
64. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng*. marzo de 2011;13(2):22-30.
65. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. marzo de 2020;17(3):261-72.
66. Seabold S, Perktold J. *Statsmodels: Econometric and Statistical Modeling with Python*. Proceedings of the 9th Python in Science Conference. 2010
67. Waskom M, Botvinnik O, Kane D, Hobson P, Lukauskas S, Gemperline D et al. *mwaskom/seaborn: v0.8.1 (September 2017)* [Internet]. Zenodo. 2021 [cited 7 February 2021]. Available from: <https://zenodo.org/record/883859#.YCANtOhKhPY>
68. Capote Bueno MI, Segredo Pérez AM, Gómez Zayas O. Climaterio y menopausia. *Revista Cubana de Medicina General Integral*. diciembre de 2011;27(4):543-57.
69. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015 Oct;526(7571):68-74.
70. Kawaguchi T, Shima T, Mizuno M, Mitsumoto Y, Umemura A, Kanbara Y, et al. Risk estimation model for nonalcoholic fatty liver disease in the Japanese using multiple genetic markers. *PLoS One*. 2018;13(1):e0185490.
71. Kawaguchi T, Sumida Y, Umemura A, Matsuo K, Takahashi M, Takamura T, et al. Genetic polymorphisms of the human PNPLA3 gene are strongly associated with severity of non-alcoholic fatty liver disease in Japanese. *PLoS One*. 2012;7(6):e38322.



72. Adams LA, White SW, Marsh JA, Lye SJ, Connor KL, Maganga R, et al. Association between liver-specific gene polymorphisms and their expression levels with nonalcoholic fatty liver disease. *Hepatology*. febrero de 2013;57(2):590-600.

SOLO USO ACADÉMICO

## 9.MATERIAL SUPLEMENTARIO

**Tabla suplementaria 1.** Frecuencias alélicas en variantes de riesgo reportadas a nivel global y en poblaciones europea, chilena, mapuche y aymara

Gen	SNP	Global <sup>(69)</sup> (n=5008)	EUR <sup>(69)</sup> (n=1006)	CL <sup>(61)</sup>	MAP <sup>61</sup> (n=54)	AYM <sup>61</sup> (n=85)	OR
<b>PNPLA3</b>	<b>rs2896019</b>	0,738	0,801	0,507 (n=418)	0,907	0,159	2.02 <sup>(26)</sup>
<b>SAMM50</b>	<b>rs2073080</b>	0,744	0,789	0,388 (n=419)	0,611	0,588	1,47 <sup>(71)</sup>
<b>GCKR</b>	<b>rs780094</b>	0,302	0,411	0,329 (n=419)	0,213	0,206	1,22 <sup>(28)</sup>
<b>PP1R3B</b>	<b>rs4240624</b>	0,114	0,075	0,770 (n=417)	0,732	0,600	1,25 <sup>(38)</sup>
<b>GC</b>	<b>rs222054</b>	0,709	0,670	0,783 (n=390)	0,774	0,849	2.54 <sup>(72)</sup>

**Tabla suplementaria 2:** Asociación entre SNPs de riesgo y condiciones clínicas. Valores obtenidos con test  $\chi^2$

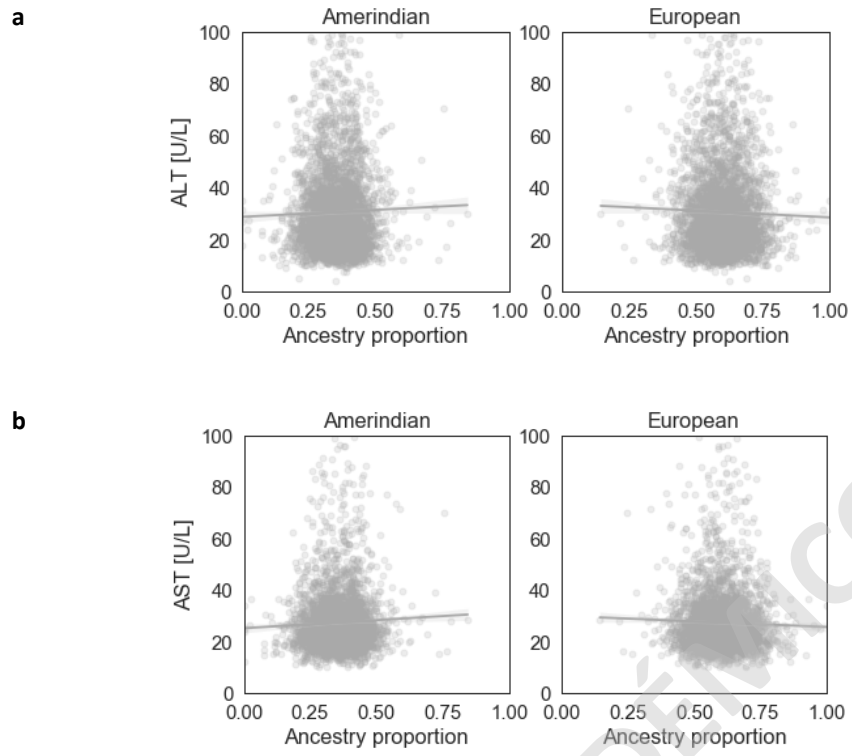
Gen	SNP	HDL		TG		Obesidad		SM	
		$\chi^2$	p	$\chi^2$	p	$\chi^2$	p	$\chi^2$	p
<b>PNPLA3</b>	<b>rs2896019</b>	0,21	0,90	2.81	0,25	1,16	0,56	2.50	0,29
<b>SAMM50</b>	<b>rs2073080</b>	0,31	0,86	5.37	0,07	1,43	0,49	1,05	0,59
<b>GCKR</b>	<b>rs780094</b>	2.03	0,36	<b>22.63</b>	<b>1,22E-5</b>	0,79	0,68	1,42	0,49
<b>PP1R3B</b>	<b>rs4240624</b>	<b>14.07</b>	<b>9.0E-5</b>	5.07	0,08	0,98	0,61	<b>28.00</b>	<b>8.33E-7</b>
<b>GC</b>	<b>rs222054</b>	2.41	0,30	1,40	0,50	1,26	0,53	3.31	0,53

**Tabla suplementaria 2** (continuación)

Gen	SNP	Glicemia		ALT		AST	
		$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>
<b>PNPLA3</b>	<b>rs2896019</b>	2.04	0,36	<b>22.87</b>	<b>1,08E-5</b>	<b>28.59</b>	<b>6.17E-7</b>
<b>SAMM50</b>	<b>rs2073080</b>	0,92	0,63	<b>9.94</b>	<b>0,007</b>	<b>13,24</b>	<b>0,001</b>
<b>GCKR</b>	<b>rs780094</b>	3.68	0,16	2.59	0,27	0,17	0,92
<b>PP1R3B</b>	<b>rs4240624</b>	4.49	0,11	2.51	0,28	3.73	0,15
<b>GC</b>	<b>rs222054</b>	3.31	0,19	3.08	0,21	4.71	0,10

**Tabla suplementaria 3.** Asociación entre transaminasas hepáticas y ancestrías. Odd Ratio estimados en regresión logística, ajustada a edad y sexo.

		Estimate	OR	<i>p</i>
ALT	<b>Sexo (M)</b>	0,797	2,218	0,000
	<b>Edad</b>	-0,019	0,980	0,000
	<b>Amerindio</b>	0,071	1,073	0,127
ALT	<b>Sexo (M)</b>	0,797	2,219	0,000
	<b>Edad</b>	-0,019	0,980	0,000
	<b>Europeo</b>	-0,063	0,938	0,170
AST	<b>Sexo (M)</b>	0,498	1,646	0,000
	<b>Edad</b>	-0,012	0,987	0,009
	<b>Amerindio</b>	0,096	1,100	0,091
AST	<b>Sexo (M)</b>	0,496	1,642	0,000
	<b>Edad</b>	-0,012	0,987	0,010
	<b>Europeo</b>	-0,050	0,950	0,377



**Figura suplementaria 1.** Tendencia de niveles de transaminasas hepáticas a) ALT y b) AST en función de ancestrías amerindia y europea.