

Universidad de Costa Rica  
Facultad de Ciencias Básicas  
Escuela de Biología

**Caracterización de variantes genéticas y frecuencias alélicas vinculadas a  
hiperlipidemias a partir de genomas costarricenses**

Trabajo final de graduación modalidad “Tesis de Graduación” para optar por el grado de  
Licenciatura en Biología con énfasis en Biología Molecular y Biotecnología

Juan Carlos Valverde-Hernández

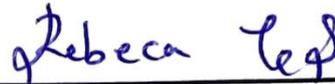
B37179

Ciudad Universitaria Rodrigo Facio  
2022



## Miembros del tribunal

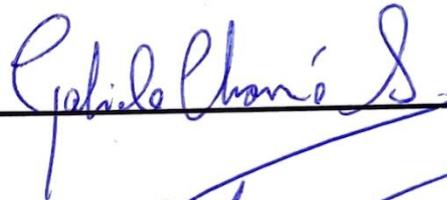
PhD. Rebeca Campos Sánchez  
*rebeca.campos@ucr.ac.cr*  
**Directora de Trabajo Final de Graduación**

  
\_\_\_\_\_

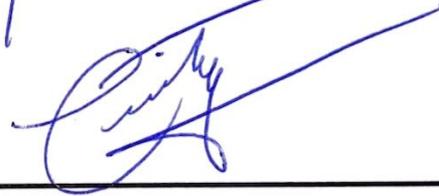
Dra. Henriette Raventós Vorst  
**Lectora de Trabajo Final de Graduación**

  
\_\_\_\_\_

PhD. Gabriela Chavarría Soley  
**Lectora de Trabajo Final de Graduación**

  
\_\_\_\_\_

PhD. Cindy Fernández García  
**Presidente del tribunal**

  
\_\_\_\_\_

PhD. Alejandro Leal Esquivel  
**Miembro del tribunal**

  
\_\_\_\_\_

B.Sc. Juan Carlos Valverde-Hernández  
*jvalverdehernandez@gmail.com*  
**Postulante**

  
\_\_\_\_\_

Ciudad Universitaria Rodrigo Facio, junio 2022



## Agradecimientos

Infinitas gracias a mi familia por las herramientas, el apoyo incondicional y el amor que me han dado. Gracias por sumar a mi vida.

Agradezco enormemente la PhD. Rebeca Campos Sánchez por brindarme la oportunidad de desarrollar este proyecto y por la guía invaluable que me brindó en el proceso.

Gracias a las lectoras Dra. Henriette Raventós Vorst, PhD. Gabriela Chavarría Soley por su disposición de utilizar su tiempo para revisar el trabajo. Estimo su criterio muchísimo. Muchas gracias también a ustedes y el resto del grupo de Psiquiatría Genética del CIBCM por brindarme acceso a los datos genómicos analizados en el estudio.

Mis agradecimientos para la M.Sc. Sandra Silva de la Fuente, quien fue lectora del anteproyecto y ha sido quien ha impulsado el estudio de la genética de la dislipidemia en Costa Rica en los últimos años.

Agradezco al Centro de Investigación en Ciencia e Ingeniería de Materiales (UCR, CICIMA) por permitirme hacer uso de su clúster computacional, sin el cual hubiera sido imposible llevar a cabo las operaciones y los análisis de datos para este estudio. Muchas gracias al PhD. Federico Muñoz Rojas del CICIMA por siempre atender mis consultas relacionadas con el uso del clúster.

Un agradecimiento al Colaboratorio Nacional de Computación Avanzada (CENAT, CNCA) por brindarme acceso al clúster computacional Kabré, el cual fue necesario para algunas etapas del análisis.

Eternas gracias al país por priorizar la inversión en la educación pública y permitir que la Universidad de Costa Rica exista. Gracias por darme la oportunidad de tener educación de alta calidad y permitirme utilizar mis energías y conocimientos para completar este trabajo.



## Tabla de contenidos

Índice de cuadros.....	ix
Índice de figuras .....	xi
Abreviaciones.....	xiii
Resumen.....	1
Abstract .....	2
Introducción.....	3
Metabolismo de lípidos .....	5
Dislipidemias.....	9
Hipertrigliceridemia (HTG) .....	10
Hipercolesterolemia (HC) .....	11
Enfoque genómico en la investigación de dislipidemias.....	13
Justificación.....	16
Objetivos .....	18
Objetivo general .....	18
Objetivos específicos.....	18
Materiales y métodos.....	19
Sujetos y datos genómicos .....	19
Descubrimiento de variantes y determinación de genotipos .....	21
Evaluación del procesamiento bioinformático .....	24
Análisis de ancestría genética.....	24
Anotación de variantes .....	26
Identificación y caracterización de variantes de interés .....	27
Código para análisis bioinformático.....	28

Resultados .....	29
Métricas del llamado de variantes cumplen con los estándares de calidad .....	29
Ancestría de genomas costarricenses es concordante con estudios previos .....	31
Sitios polimórficos identificados en genes y promotores de interés.....	33
Discusión.....	48
Evaluación de flujo de análisis bioinformático .....	48
Concordancia con ancestría de costarricenses del Valle Central .....	49
Variantes de interés farmacogenómico .....	50
Variantes categorizadas como de riesgo.....	51
Variantes previamente reportadas en la región Latinoamericana.....	53
Conclusiones y recomendaciones.....	56
Referencias .....	60
Anexos.....	74

## Índice de cuadros

Cuadro 1. Características principales de las seis clases de lipoproteínas. ....	5
Cuadro 2. Categorías de Fredrickson. ....	10
Cuadro 3. Coordinadas genómicas caracterizadas en la etapa de llamado de variantes. ....	20
Cuadro 4. Estadísticas del llamado de variantes para el panel de genes y promotores involucrados en el metabolismo de lípidos mencionados en el Cuadro 3. ....	33
Cuadro 5. Variantes encontradas en genes de interés asociadas fenotípicamente a rasgos con significancia clínica farmacogenómica o protectora ante enfermedades. ....	40
Cuadro 6. Variantes que producen alteraciones en el marco abierto de lectura encontradas en genes involucrados en el metabolismo de lípidos categorizadas como riesgosas por más de dos fuentes. ....	41
Cuadro 7. Variantes identificadas en promotores de genes involucrados en el metabolismo de lípidos categorizadas como riesgosas por más de dos fuentes. ....	43
Cuadro 8. Variantes previamente reportadas en costarricenses y latinoamericanos en genes involucrados en el metabolismo de lípidos y sus promotores. ....	44
Anexo 1. Variantes que producen alteraciones en el marco abierto de lectura encontradas en genes involucrados en el metabolismo de lípidos categorizadas como riesgosas por una fuente como mínimo. ....	74



## Índice de figuras

Figura 1. Ruta endógena y exógena del transporte de lípidos. ....	8
Figura 2. Flujo de análisis bioinformático basado en las buenas prácticas de llamado de variantes propuestas por GATK. ....	22
Figura 3. Métricas de calidad para el proceso de llamado de variantes efectuado en los grupos CR-WGS y AMR-80. ....	30
Figura 4. Relación lineal entre 15 frecuencias alélicas obtenidas en costarricenses del Valle Central de la cohorte CR-WGS y costarricenses sanos de la misma región elegidos aleatoriamente para el <i>Costa Rica Heart Study</i> . ....	31
Figura 5. Similitud genética entre individuos americanos con base en genotipos de 250 AIMs reportados por Wang <i>et al.</i> (2019).....	32
Figura 6. Mezcla genética observada en los individuos costarricenses calculada con base en genotipos de 78 AIMs reportados por Campos-Sánchez <i>et al.</i> (2013). ....	34
Figura 7. Conteo de variantes detectadas en genes involucrados en el metabolismo de lípidos categorizados por su impacto sobre el transcrito.....	36
Figura 8. Significancia clínica de acuerdo con las asociaciones fenotípicas reportadas en ClinVar para las variantes identificadas en los genes de interés. ....	37
Figura 9. Diferencias observadas entre frecuencias alélicas en genes asociados a metabolismo de lípidos en costarricenses frente a las reportadas 1KGP. ....	38
Figura 10. Concordancia entre fuentes utilizadas para identificar variantes de interés de acuerdo con su patogenicidad o su asociación con alteraciones en el perfil lipídico.....	39
Figura 11. Expresión de genes afectados por variantes de interés en tejidos protagónicos en el metabolismo de lípidos y su vínculo con procesos biológicos. ....	47



## Abreviaciones

<b>Abreviación</b>	<b>Término</b>
1KGP	- <i>1000 Genomes Project</i>
AFR	- Africanos del Este incluidos en <i>1000 Genomes Project</i>
AIM	- <i>Ancestry Informative Markers</i> o Marcadores informativos de ancestría
ALL	- Sujetos incluidos en <i>1000 Genomes Project</i>
AMR	- Latinoamericanos incluidos en <i>1000 Genomes Project</i>
ANGPLT3	- Proteína semejante a angiopoyetina 3
apo(a)	- Apolipoproteína (a)
apoA-I	- Apolipoproteína AI
apoA-II	- Apolipoproteína AII
apoA-IV	- Apolipoproteína AIV
apoB-100	- Apolipoproteína B-100
apoB-48	- Apolipoproteína B-48
apoC-I	- Apolipoproteína CI
apoC-III	- Apolipoproteína CII
apoE	- Apolipoproteína E
BAM	- Mapa de alineamiento de secuencias en compresión binaria
BQSR	- <i>Base quality score recalibration</i>
BWA	- Burrows-Wheeler Aligner
C.C.S.S	Caja Costarricense del Seguro Social
CENAT	- Centro Nacional de Alta Tecnología
CETP	- Colesteryl ester transferasa
CICIMA	- Centro de Investigación en Ciencia e Ingeniería de Materiales
CREB	- Proteína de adhesión a CRE
CRTC2	- Coactivador de transcripción 2 regulado por CREB
CR-RC	- Costarricenses de la Región Caribe reportados por Campos-Sánchez et al. 2013
CR-VC	- Costarricenses del Valle Central reportados por Campos-Sánchez et al. 2013
CR-WGS	- Cohorte de costarricenses analizados en el presente estudio
CR-ZN	- Costarricenses de la Zona Norte reportados por Campos-Sánchez et al. 2013
CR-ZS	- Costarricenses de la Zona Sur reportados por Campos-Sánchez et al. 2013
ddNTPs	- Dideoxinucleótido trifosfatos
EAS	- Asiáticos del Este incluidos en <i>1000 Genomes Project</i>
EUR	- Europeos del Este incluidos en <i>1000 Genomes Project</i>
FAT4	- Ácido graso translocasa 4
FATHM	- <i>Functional Analysis through Hidden Markov Models</i>
GATK	- <i>Genome Analysis Toolkit</i>

<b>Abreviación</b>	<b>Término</b>
gnomAD	- <i>Genome Aggregation Database</i>
GPD1	- Glicerol-3-fosfato deshidrogenasa
GPIHBP1	- Proteína 1 de unión a lipoproteína de alta densidad
GVCF	- Formato variant call genómico
GWAS	- <i>Whole Genome Association Study</i>
GWAVA	- <i>Genome Wide Annotation of VARIants</i>
HC	- Hipercolesterolemia
HDL	- Lipoproteína de alta densidad
HMGCR	- 3-hidroxi-3-metil-glutaril-CoA reductasa
HTG	- Hipertrigliceridemia
HTGL	- Triglicérido lipasa hepática
ICD	- Clasificación Internacional de Enfermedades
IDL	- Lipoproteína de densidad intermedia
LCAT	- Lecitina-colesterol aciltransferasa
LDLR	- Receptor de lipoproteínas de baja densidad
LMF1	- Factor 1 de maduración de la lipasa
LPL	- Lipoproteína lipasa
LRP	- Proteína relacionada con el receptor LDL
miRNA	- Micro ARN
MYLIP/IDOL	- degradador inducible de receptor de LDL
NIH	- <i>National Institute of Health</i>
NPC1L1	- Transportador intestinal de colesterol
OMS	- Organización Mundial de la Salud
PacBio	- <i>Pacific Biosciences</i>
PCSK9	- Proproteína convertasa subtilisina/kexina tipo 9
REST	- <i>Representational State Transfer</i> o Transferencia de Estado Representacional
REVEL	- <i>Rare Exome Variant Ensemble Learner</i>
SIFT	- <i>Structural Interaction Fingerprint</i>
SR-BI	- Receptor carroñero clase B
tGRCh38/hg38	- <i>Genome Reference Consortium Human Build 38</i>
VCF	- Formato variant call
VEP	- <i>Variant Effect Predictor</i>
VLDL	- Lipoproteína de muy baja densidad
VQSR	- <i>Variant quality score recalibration</i>

## Resumen

Las hiperlipidemias son factores de riesgo en enfermedades de gran importancia para la salud pública tales como la pancreatitis aguda y la aterosclerosis, condiciones que contribuyen al desarrollo de cáncer pancreático y enfermedades cardiovasculares, respectivamente. Estilos de vida poco saludables, la preexistencia de enfermedades y la acumulación de variantes genéticas en algunos *loci* son factores que participan en el desarrollo de las hiperlipidemias. Si bien hay estudios abundantes que analizan la causalidad genética detrás de esta enfermedad, su mayoría se han enfocado en estudiar individuos en países europeos y norteamericanos. En Costa Rica hay escasos estudios de este tipo y ninguno de estos se ha enfocado en identificar variantes con potencial de afectar los niveles de lípidos en sangre y determinar su frecuencia en el país. Por consiguiente, este estudio se focalizó en identificar variantes ubicadas en genes involucrados en el metabolismo de lípidos a partir de genomas pertenecientes a costarricense del Valle Central, contrastar su frecuencia alélica con las de grupos reportados en *1000 Genomes Project* e identificar potenciales variantes que podrían incidir en el desarrollo de dislipidemias. Fueron identificadas aproximadamente 17500 variantes en los genes y promotores evaluados, de las cuales la amplia mayoría no habían sido reportadas en costarricenses. Utilizando herramientas bioinformáticas predictivas se identificaron de 29 variantes con potencial de afectar el desempeño de estos genes en regiones codificantes y promotoras. Algunas de estas variantes han sido vinculadas directamente con alteraciones en los niveles de lípidos en sangre en el pasado. Ciertas variantes reportadas previamente en estudios realizados localmente no fueron identificadas probablemente debido a que estos estudios incluyeron únicamente sujetos con dislipidemias severas, enriqueciendo la muestra de variantes potencialmente raras en la población general.

**Palabras clave:** dislipidemia, hiperlipidemia, hipertrigliceridemia, hipercolesterolemia, descubrimiento de variantes, variantes de riesgo, secuenciación de nueva generación, bioinformática, farmacogenómica, América Latina, América Central, Costa Rica.

## Abstract

Hyperlipidemias are risk factors in diseases of significant importance to public health such as acute pancreatitis and atherosclerosis, conditions that contribute to the development of pancreatic cancer and cardiovascular diseases, respectively. Unhealthy lifestyles, the pre-existence of diseases and the accumulation of genetic variants in some *loci* participate in the development of hyperlipidemias. The genetic causality behind this disease has been mostly in populations with a large European ancestry. Few studies have explored this topic in Costa Rica and none of them have focused on identifying variants with the potential to alter blood lipid levels and quantifying their frequency. To fill this gap, this study focused on identifying variants located in genes and promoters involved in lipid metabolism from genomes belonging to Costa Ricans of the Central Valley, contrasting their allelic frequency with those of groups reported in *the 1000 Genomes Project* and identifying potential variants that could affect the development of dyslipidemias. Approximately 17500 variants were identified in the evaluated regions, of which the vast majority were not reported in Costa Ricans. A set of 29 variants with the potential to alter the performance of these genes were detected on their coding and promoter regions using predictive bioinformatics tools. Some of them have been linked to changes in blood lipid levels in other studies. Some of the variants previously reported in locally conducted studies were not identified most likely because those studies only included subjects with severe dyslipidemias, which could have enriched the sample with variants that are probably rare in the general population due to their pathogenicity.

**Key words:** dyslipidemia, hyperlipidemia, hypertriglyceridemia, hypercholesterolemia, variant discovery, risk variant, NGS, bioinformatics, pharmacogenomics, Latin America, Central America, Costa Rica.

## Introducción

Una importante parte de la población mundial padece de alguna clase de dislipidemia. En ciudades de América Latina tales como Barquisimeto, Lima y Bogotá, por ejemplo, se han registrado incidencias para esta condición mayores al 70% en hombres y superiores al 50% en mujeres (Vinueza et al., 2010). Costa Rica no es la excepción. En un estudio realizado en la década del 2000 que involucró 107.000 habitantes de San José, se reportó que 36% de hombres y 22% de las mujeres presentaron hipercolesterolemia, mientras que un 48% de los hombres y un 52% de las mujeres reportaron hipertrigliceridemia (Gutiérrez et al., 2010). Esto, sumado a que tales condiciones se han vinculado estrechamente con el desarrollo de padecimientos complejos entre los cuales se destacan las enfermedades cardiovasculares y la pancreatitis aguda (Bruikman et al., 2017; Paredes et al., 2019, de Pretis et al., 2018), hacen que las hiperlipidemias sean un problema de salud pública en el siglo XXI.

El sedentarismo y malos hábitos alimenticios pueden tener un impacto profundo sobre el desarrollo de estas enfermedades (Brahm & Hegele, 2013), el abordaje clínico de estos casos suele incluir la implementación de regímenes de ejercicio y de restricción calórica. Adicionalmente, una cantidad robusta de evidencia ha demostrado que las características genéticas de un individuo tienen un papel protagónico en el desarrollo de las hiperlipidemias (Brahm & Hegele, 2013; Johansen et al., 2011; Wierzbicki & Reynolds, 2019). Actualmente la enfermedad se considera mayoritariamente poligénica, sin embargo, se ha demostrado que alteraciones de genes como el gen de lipoproteína lipasa (*LPL*), el receptor de lipoproteínas de baja densidad (*LDLR*) y la apolipoproteína B (*APOB*) suelen tener efectos más marcados que otros genes involucrados en el metabolismo lipídico (Dron et al., 2019, 2020; Johansen et al., 2011, 2014; Lewis et al., 2015).

La mayoría de los estudios dirigidos a identificar el efecto que tiene el componente genético sobre la presencia de alteraciones en el metabolismo de lípidos y el desarrollo de dislipidemias han sido realizados principalmente en países anglosajones y europeos. Dicho esto, el

estudio de Andaleon et al. (2019) ha sido de los más exhaustivos realizados a la fecha de esta índole para América Latina, centroamericanos incluidos. No obstante, actualmente se conoce poco sobre la arquitectura genética de los genes que han sido vinculados con estos padecimientos tanto a nivel latinoamericano como a nivel nacional.

En Costa Rica a la fecha se han publicado pocos estudios sobre esta materia. Por un lado, algunas investigaciones derivadas del proyecto *Dietary Fat and Heart Disease in Costa Rica* también conocido como *Costa Rica Heart Study* financiado por el *National Institute of Health* (NIH) han cuantificado las frecuencias alélicas de ciertas variantes en los genes *APOC*, *LPL*, *APOE*, *PCSK9*, *FADS1-2-3* y *USF1* a partir de 4000 individuos del Valle Central costarricense y han vinculado la presencia de algunas de estas variantes con un incremento en el riesgo de padecer enfermedades coronarias e hiperlipidemias (Airewele et al., 2001; Aslibekyan et al., 2012; Brown et al., 2003; Gong et al., 2011; Ruiz-Narváez et al., 2005, 2008; Yang et al., 2004; Yu et al., 2017). Por otro lado, dos investigaciones se han enfocado en identificar variantes genéticas en un grupo de 38 costarricenses con hipertrigliceridemia ubicadas en regiones de interés tales como el gen *LPL* y la región promotora de *APOCII* (González Cordero, 2018; Gutiérrez Ávila, 2019).

Esta investigación propone usar datos de genomas completos costarricenses del Valle Central para identificar sitios polimórficos en genes vinculados con la incidencia de dislipidemias, cuantificar las frecuencias alélicas en las variantes identificadas y caracterizar funcionalmente las variantes *in silico*. Esto con el fin de conocer si las herramientas bioinformáticas actuales las categorizan como variantes de interés debido a su patogenicidad o posible relevancia farmacogenómica. Esto permitirá tener el primer retrato nacional de la frecuencia de variantes de riesgo previamente reportadas en genes vinculados con este grupo de enfermedades obtenido a partir de datos genómicos. Adicionalmente, se reportan las frecuencias alélicas encontradas para las variantes ubicadas en genes de interés (i.e. *LDLR* y *APOCII*) previamente reportadas en costarricenses. El impacto de este estudio se podría observar en el abordaje clínico de pacientes con hiperlipidemias desde la perspectiva de la medicina personalizada.

## Metabolismo de lípidos

Los triglicéridos y esteroides forman parte del grupo de biomoléculas hidrofóbicas denominadas lípidos, vitales en procesos biológicos tales como la señalización celular, almacenamiento energético, como componente estructural en membranas y transporte de nutrientes liposolubles (Cox & García-Palmieri, 1990; Di Angelantonio et al., 2009). Debido a la naturaleza hidrofóbica de los mismos, su transporte y eventual metabolización es posible puesto que dichas moléculas se ensamblan en partículas denominadas lipoproteínas (Cox & García-Palmieri, 1990; Tso et al., 2004). Estas son globulares con un núcleo lipofílico mayormente constituido por triglicéridos y colesterol, encapsulado por una monocapa constituida por fosfolípidos, colesterol esterificado y apolipoproteínas, las cuales están implicadas en su reconocimiento molecular, procesamiento y estabilización (Mahmood, 2014; Torres-Romero et al., 2020; L. Zhou et al., 2015). Las lipoproteínas usualmente se dividen en seis clases principales de acuerdo con su densidad, la cual está asociada negativamente con la relación entre su contenido lipídico frente al proteico (Cuadro 1).

**Cuadro 1.** Características principales de las seis clases de lipoproteínas.

(Feingold & Grunfeld, 2018; McDonald et al., 2016).

Clase	Quilomacrón	Lipoproteína de muy baja densidad (VLDL)	Lipoproteína de densidad intermedia (IDL)	Lipoproteína de baja densidad (LDL)	Lipoproteína (a)	Lipoproteína de alta densidad (HDL)
Tamaño (nm)	80-500	30-80	25-35	18-25	~30	5-12
Densidad (g/mL)	<0.930	0.930 - 1.006	1.006 - 1.019	1.019 - 1.063	1.055 - 1.085	1.063 - 1.210
Lípidos totales (%)	98-99	90-92	-	75-80	-	40-48
Glicerolípidos (%)	81-89	7	-	10	-	5
Colesterol esterificado (%)	2 - 4	15 - 23	-	47 - 51	-	24 - 45
Colesterol no esterificado (%)	1 - 3	4 - 9	-	10 - 12	-	6 - 8
Fosfolípidos (%)	7 - 9	19 - 12	-	28 - 30	-	42 - 51
Apolipoproteínas	apoB-48, apoC, apoE, apoA-I, A-II, A-IV	apoB-100, apoE, apoC	apoB-100, apoE, apoC	apoB-100	apoB-100, apo(a)	apoA-I, apoA-II, apoC, apoE

Tanto los lípidos como las lipoproteínas se pueden originar tanto por vía exógena como endógena (Cox & García-Palmieri, 1990) (Figura 1). La primera involucra la formación de estas a partir de lípidos absorbidos intestinalmente mediante la dieta. Una vez ingeridos los alimentos, la masticación y la liberación de lipasa salival inician el proceso de digestión y absorción (Mahmood, 2014). En la cavidad intestinal entran en contacto con lipasas pancreáticas y sales biliares. Estas últimas emulsifican los lípidos y facilitan su hidrólisis mediada por las lipasas, convirtiendo así una porción de los triglicéridos en ácidos grasos libres y monoacilglicerol, y el colesterol esterificado en colesterol libre, lo que reduce su hidrofobicidad y a su vez facilita su absorción por las células intestinales denominadas enterocitos (Xiao et al., 2011).

El influjo de lípidos a los enterocitos sucede tanto por transporte pasivo como mediante transporte activo, facilitado por proteínas como el ácido graso translocasa CD36 (FAT/CD36), el ácido graso translocasa (FAT4) y el receptor carroñero clase B (SR-BI), involucrados en la captación de colesterol (Cox & García-Palmieri, 1990). Muchos de los lípidos absorbidos son esterificados una vez que ingresan a las células. Seguidamente, el complejo de Golgi y el retículo endoplásmico median el ensamblaje de las lipoproteínas de baja densidad denominadas quilomicrones al añadir progresivamente a la apolipoproteína B-48 (apoB-48) lípidos obtenidos de la dieta. Estas apolipoproteínas son secretadas a vasos linfáticos y eventualmente alcanzan el torrente sanguíneo.

La vía endógena de producción de lípidos se da en el hígado, donde se forman lipoproteínas mayormente absorbidas a partir de las lipoproteínas circulantes, principalmente remanentes de quilomicrones (Cox & García-Palmieri, 1990; Xiao et al., 2011). Los hepatocitos ensamblan las lipoproteínas de muy baja densidad (VLDL) a partir de lípidos y la apolipoproteína B-100 (apoB-100). A diferencia de los quilomicrones, los VLDL son formados en luego de comer, también conocida como etapa postprandial (Tiwari & Siddiqi, 2012).

Las VLDLs y los quilomicrones una vez en el torrente sanguíneo se metabolizan por medio de lipólisis vascular y experimentan cambios en su contenido lipídico y proteico. La lipoproteína

lipasa (LPL) tiene un rol protagónico tanto en la lipólisis gradual de las lipoproteínas como en la distribución de los ácidos grasos libres generados en dicho proceso. Por un lado, LPL al entrar en contacto con los quilomicrones hidroliza los triglicéridos presentes en el núcleo de estas partículas, generando ácidos grasos libres y monoacilglicerol, y transformando los *VLDLs* en lipoproteínas de densidad intermedia (*IDL*), y los quilomicrones en remanentes de quilomicrones. Por lo general, la LPL se localiza en el endotelio capilar próximo al músculo esquelético, cardíaco y del tejido adiposo. Otras enzimas relevantes en la lipólisis de las lipoproteínas son la triglicérido lipasa hepática (HTGL), la lecitina-colesterol aciltransferasa (LCAT) y la proteína colesterol éster transferasa (CETP) (Dominiczak & Caslake, 2011).

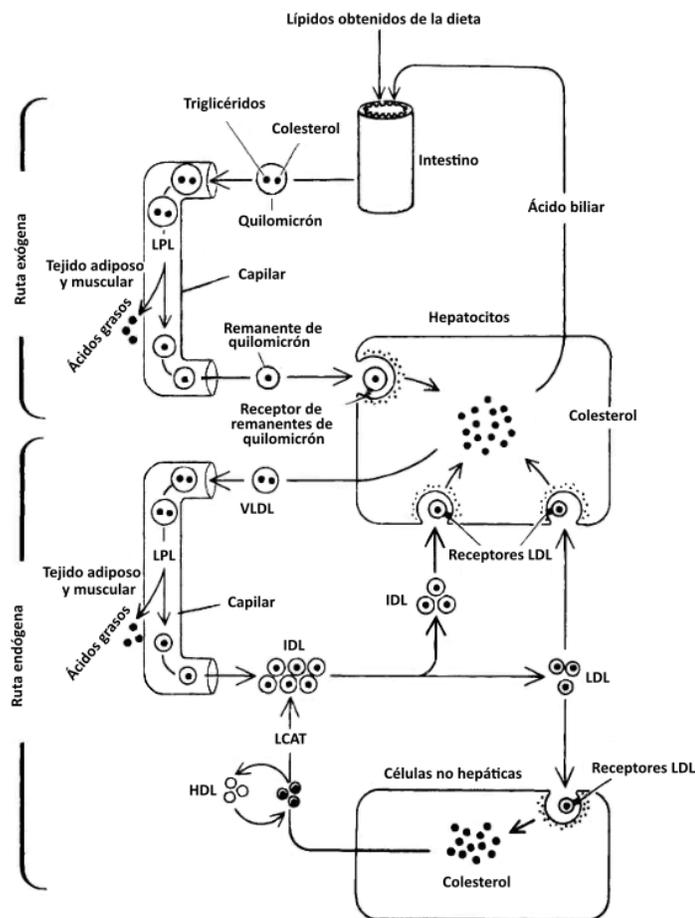
Las lipoproteínas de densidad intermedia (*IDL*), lipoproteínas de densidad baja (*LDL*) y los remanentes de quilomicrones se originan de lipoproteínas que han sido sujetas a lipólisis por parte de la LPL (Bhatnagar et al., 2008), debido a esto dichas lipoproteínas tienen porcentualmente una menor cantidad de triglicéridos y una mayor cantidad de colesterol. Estas tienen un rol importante en la distribución del colesterol a lo largo del cuerpo, principalmente hacia tejidos glandulares (Berglund et al., 2012; Brahm & Hegele, 2013).

Las lipoproteínas de alta densidad (*HDL*) son sintetizadas mayormente en el hígado y el intestino delgado. Poseen un alto contenido proteico, teniendo la apolipoproteína apoA-1 un rol protagónico en su ensamblaje. Los *HDL* se destacan por su capacidad de absorber lípidos de células y trasladarlos al hígado, promoviendo la redistribución de los lípidos en el cuerpo (L. Zhou et al., 2015).

Los procesos de distribución y metabolización de las lipoproteínas están altamente regulados. El conjunto de apolipoproteínas que contiene una lipoproteína es determinante para su reconocimiento celular puesto que pueden ser son identificados como ligandos por receptores celulares (Dominiczak & Caslake, 2011) La apoB-100 y la apolipoproteína E (apoE) son reconocidos por el receptor LDL (*LDL-R*). Del mismo modo, apoE es reconocido por la proteína relacionada con el receptor LDL (*LRP*) y las apolipoproteínas A con el receptor carroñero BI

(Dominiczak & Caslake, 2011). Las apolipoproteínas también tienen la capacidad de modular la actividad enzimática de proteínas que actúan sobre las lipoproteínas.

ApoC-I y apoC-III por ejemplo, tienen la capacidad de inhibir la acción lipasa de enzimas como la LPL y la HTGL, mientras que apoC-II promueve su actividad (Dominiczak & Caslake, 2011). La regulación de estos procesos es altamente compleja, el humano codifica 21 apolipoproteínas y su mayoría pueden ser transferidas dinámicamente entre lipoproteínas s (Su & Peng, 2020; von Zychlinski & Kleffmann, 2015; Y. Zhou et al., 2018).



**Figura 1.** Ruta endógena y exógena del transporte de lípidos.

Modificada de Cox & García-Palmeri (1980).

Por otro lado, la actividad de actores importantes en el metabolismo de lípidos, como el LPL, es regulada en múltiples niveles. Actores de transcripción tales como CREB y CRT2 dictan en cierta medida la expresión de LPL. Los microARNs miR-29 y el miR-410 por ejemplo, se ha demostrado que tienen la capacidad de regular a la baja los transcritos del gen *LPL* (Caussy et al., 2016). El funcionamiento pleno de LPL depende también en gran medida de sufrir modificaciones postraduccionales mediadas por el factor 1 de maduración de la lipasa (LMF1). Existen mecanismos que están involucrados en la movilización de la proteína LPL madura hacia la membrana endotelial cuyo funcionamiento es vital para que LPL mantenga su bioactividad. Adicionalmente, a la hora de que LPL actúa sobre las lipoproteínas, una importante cantidad de especies participan en la modulación de este proceso, GPIHBP1 por ejemplo promueve la lipólisis de LPL hacia las lipoproteínas mientras que ANGPT3 la inhiben (Lewis et al., 2015).

### *Dislipidemias*

Concentraciones lipídicas en sangre distantes a la media pueden ser ocasionadas tanto por estilos de vida poco saludables, condiciones médicas y ambientales preexistentes, como por anomalías en la síntesis, degradación y transporte de lípidos ocasionadas por a las características genómicas de los individuos (Cox & García-Palmieri, 1990; Lewis et al., 2015). Las dislipidemias son un grupo de condiciones cuyo sello es la presencia de niveles anormales en los lípidos. La incidencia de estas condiciones se ha vinculado estrechamente con el desarrollo y prevalencia de padecimientos cardiovasculares, pancreatitis, problemas de crecimiento, obesidad, entre otras (Bartłomiejczyk et al., 2019; Brahm & Hegele, 2013; de Pretis et al., 2018; Vekic et al., 2019).

Las condiciones que involucran niveles lipídicos anormalmente altos son denominadas hiperlipidemias o hiperlipoproteinemias. En la clínica los casos suelen ser categorizados de acuerdo con los parámetros como niveles de colesterol, triglicéridos y lipoproteínas específicas en subclases definidas por sistemas como el propuesto por Fredrickson y por la Clasificación Internacional de Enfermedades (ICD) de la Organización Mundial de la Salud (OMS) (Fredrickson & Lees, 1965;

Lee & Hegele, 2017). La clasificación de fenotipos de Fredrickson está basada en los niveles de lipoproteínas en plasma (Cuadro 2). Anteriormente la clasificación de las hiperlipidemias de la ICD clasificaba estas enfermedades en subtipos semejantes a la clasificación de Fredrickson, no obstante, recientemente la OMS sustituyó en la ICD-11 tales categorías por hipertrigliceridemia (HTG), hipercolesterolemia (HC), hiperlipidemia mixta e hipoalfalipoproteinemia (World Health Organization, 2020).

**Cuadro 2.** Categorías de Fredrickson.

(Fredrickson & Lees, 1965; Stroes et al., 2017).

Fenotipo	Lipoproteínas incrementadas	Lípidos incrementados
I	LDL	Colesterol
IIa	LDL-VLDL	Triglicéridos y colesterol
IIb	$\beta$ -lipoproteínas	Triglicéridos y colesterol
III	VLDL	Triglicéridos y colesterol en menor medida
IV	Quilomicrones	Triglicéridos y colesterol

*Hipertrigliceridemia (HTG)*

Existen diversos criterios para definir la HTG, no obstante, la *Endocrine Society* indica que niveles de triglicéridos en plasma durante ayuno mayores a 150mg/dL desvían de lo normal, de 200mg/dL a 1000mg/dL se clasifica como hipertrigliceridemia moderada, de 1000mg/dl a 2000mg/dL como severa, y muy severa de 2000mg/dL en adelante (Berglund et al., 2012; Brahm & Hegele, 2013). En la mayoría de los casos estos desniveles van de la mano con la acumulación patológica de quilomicrones y VLDLs circulantes (Dron et al., 2019; Lee & Hegele, 2017). Algunas manifestaciones clínicas de esta enfermedad son hepatoesplenomegalia, xantomas cutáneos eruptivos, *lipidemia retinalis* y pancreatitis aguda (Brahm & Hegele, 2013; Chyzyk et al., 2019; Hegele, 2014; Johansen et al., 2011). Se ha demostrado que altos niveles de triglicéridos en plasma aumentan el riesgo de enfermedades cardiovasculares (Lewis et al., 2015).

Múltiples factores influyen en el desarrollo e incidencia de HTG. Factores asociados al estilo de vida del individuo tales como el consumo de alcohol y malos hábitos alimenticios tienen un rol importante en el desarrollo de la enfermedad (Lewis et al., 2015). De igual modo, la prevalencia de condiciones preexistentes entre ellas la insuficiencia renal, la obesidad, diabetes tipo 2 y el embarazo aumentan el riesgo a desarrollar HTG (Brahm & Hegele, 2013). Las características genéticas de los actores moleculares involucrados en el ciclo de vida de las lipoproteínas ricas en triglicéridos también tienen un papel en la predisposición de un individuo a desarrollar esta enfermedad.

Alteraciones en las proteínas involucradas en la regulación de la actividad de la LPL tales como apoA-5, apoC-2, GPIHBP1 y LMF1 pueden tener efectos profundos sobre la capacidad de un individuo de procesar los triglicéridos (Lewis et al., 2015; Surendran et al., 2012). Del mismo modo, mutaciones vinculadas a genes involucrados en el metabolismo de ácidos grasos libres, entre ellos la glicerol-3-fosfato deshidrogenasa (*GPDI*) pueden tener efectos significativos sobre los triglicéridos en plasma de un individuo. Si bien ciertas mutaciones en los genes *APOA5*, *APOC2*, *GPDI*, *GPIHBP1*, *LMF1* y *LPL* son causantes del desarrollo de dislipidemia monogénica, una creciente cantidad de estudios indica que en la mayoría de los casos la incidencia de HTG es poligénica (Brahm & Hegele, 2013; Dron et al., 2019; Johansen et al., 2011; Reiner, 2017).

### *Hipercolesterolemia (HC)*

Generalmente se considera que un individuo posee HC cuando presenta niveles plasmáticos de colesterol total superiores a 190 mg/dL (5.0 mmol/L) (Ahangari et al., 2018). No obstante, el colesterol presente en las LDL (C-LDL), en los remanentes de lipoproteínas y en los HDL (C-HDL) son parámetros claves en el manejo de esta enfermedad debido a que explican en cierta medida las dinámicas de acumulación y redistribución de colesterol en el cuerpo (Bhatnagar et al., 2008; Henderson et al., 2016; Nordestgaard, 2016).

Elevados niveles de colesterol total, C-LDL y colesterol de lipoproteínas remanentes propician la acumulación de este compuesto en las paredes endoteliales (Bhatnagar et al., 2008; Henderson et al., 2016; Nordestgaard, 2016). La endocitosis de LDLs y remanentes de lipoproteínas por macrófagos presentes en las paredes arteriales es un paso trascendental en el desarrollo de ateromas, disfunción endotelial y procesos inflamatorios (Bhatnagar et al., 2008; Henderson et al., 2016; Nordestgaard, 2016). Por esta razón se considera que la HC es el principal factor de riesgo para el desarrollo de enfermedades arteriales coronarias, una de las principales causas de muerte a nivel mundial (Bhatnagar et al., 2008). Del mismo modo que la HTG, el desarrollo de hipercolesterolemia está vinculado con el estilo de vida del individuo, factores ambientales, condiciones preexistentes y su perfil genético. Malos hábitos alimenticios, obesidad, diabetes, el fumado y la edad repercuten sobre la incidencia de esta enfermedad (Langlois & Nordestgaard, 2018; Paredes et al., 2019).

Se han identificado una amplia cantidad de variantes genéticas asociadas a genes involucrados en la biogénesis, transporte y catabolismo del colesterol en el cuerpo. Si bien esta enfermedad tiende a ser de naturaleza poligénica (Bianconi et al., 2020; Henderson et al., 2016), históricamente el enfoque investigativo ha sido dirigido al estudio de los genes cuyas mutaciones tienen un impacto más profundo, siendo estos *LDL-R*, *APOB* y *PCSK9* (Bhatnagar et al., 2008; Bianconi et al., 2020; Bruikman et al., 2017; Henderson et al., 2016). Afectaciones en la capacidad de LDL-R de reconocer las LDL dificulta la distribución del colesterol a tejidos periféricos, se reporta que en algunas poblaciones un 80% de los casos de HC monogénica se atribuye a mutaciones en este gen (Bhatnagar et al., 2008). Mutaciones en el gen *APOB* son la segunda causa de HC monogénica, cambios en la proteína pueden afectar negativamente el reconocimiento celular de las lipoproteínas ricas en colesterol, resultando en acumulación de colesterol en plasma (Henderson et al., 2016). Por otro lado, se ha reportado que alrededor de un 1% de los casos de HC monogénica presentan mutaciones en la proproteína convertasa subtilisina/kexina tipo 9 (*PCSK9*)

al ser un inhibidor de la LDL-R, se ha observado que incrementos en la actividad de la misma se asocia con el desarrollo de HC (Henderson et al., 2016).

Si bien se ha observado que algunas variantes en *LDL-R*, *APOB* y *PCSK9* pueden tener predisponer en gran medida a un individuo de padecer HC, hay una amplia gama de genes los cuales también pueden ser afectados por variantes que aportan a la predisposición genética de un individuo. Algunos de los más estudiados son el transportador intestinal de colesterol *NPC1L1*, el degradador inducible de receptor de LDL *MYLIP/IDOL*, el regulador positivo de secreción y actividad de PCSK9 denominado *SORT1*; *APOE* y 3-hidroxi-3-metil-glutaril-CoA reductasa (*HMGCR*) involucrada en la biogénesis de colesterol (Ahangari et al., 2018; Bianconi et al., 2020; Ference et al., 2016; Gustafsen et al., 2014; Huff et al., 2014; Tada et al., 2017).

### *Enfoque genómico en la investigación de dislipidemias*

En la década de 1970, Frederick Sanger desarrolló la metodología de secuenciación de ADN mediada por el uso de inhibidores de amplificación. Otros refinamientos de la técnica permitieron que la técnica fuera adoptada extensivamente por la comunidad científica, entre ellos la suplantación del gel de electroforesis por electroforesis capilar como método de visualización y la sustitución de los inhibidores tipo didesoxinucleótido trifosfatos (ddNTPs) radiomarcados por ddNTPs fluorocromados (França et al., 2002; Sanger et al., 1977). Si bien los resultados de la secuenciación Sanger son considerados el estándar de oro a la fecha, tienen la limitación de que la longitud máxima de los fragmentos secuenciados (~1kb) hace inviable la secuenciación de genomas por este método. Esta limitación condujo a que a mediados de la década del 2000 se desarrollara la primera tecnología de secuenciación masiva (Liu et al., 2012). Si bien inicialmente los costos elevados de secuenciar un genoma humano completo eran una barrera para que laboratorios de investigación hicieran uso de esta tecnología, la reducción exponencial de su costo en los recientes años, el desarrollo de programas bioinformáticos y el encadenamiento de ellos para el análisis de estos datos han permitido que la implementación de secuenciación de genomas

completos, exomas y transcriptomas sea cada vez más común. Actualmente las tecnologías de secuenciación desarrolladas por Illumina, PacBio y Nanopore lideran este creciente mercado (Kumar et al., 2019).

La secuenciación de genomas y exomas ha permitido caracterizar miles de variantes génicas que previamente no habían sido identificadas y han ayudado a elucidar el efecto de estas sobre una inmensa cantidad de fenotipos (Carrasco-Ramiro et al., 2017). Si bien el manejo y análisis de datos generados por medio de estas técnicas ha sido un reto, actualmente existen flujos de trabajo los cuales permiten la identificación de variantes genéticas, entre los cuales se destacan las buenas prácticas de llamado de variantes a través del uso del *Genome Analysis Toolkit* o GATK desarrollado por el *Broad Institute* (Poplin et al., 2017).

Por su parte, la anotación funcional es facilitada por herramientas como *Variant Effect Predictor (VEP)*, ANNOVAR y snpEff que integran la información de las variantes de diferentes fuentes y simplifican el proceso de anotar las variantes con dicha información (Cingolani et al., 2012; Cunningham et al., 2021; K. Wang et al., 2010). Estas pueden anotar las variantes con datos provenientes o generados con:

- Modelos predictivos tales como SIFT, PolyPhen2, ClinPred y REVEL que ayudan a determinar *in silico* si las variantes pueden ser patogénicas (Adzhubei et al., 2013; Flanagan et al., 2010; Ng & Henikoff, 2003; Alirezaie et al., 2018; Gunning et al., 2021; Ioannidis et al., 2016).
- Bases de datos como ClinVar o *NHGRI-EBI GWAS catalog* que vinculan sitios polimórficos y sus variantes con efectos fenotípicos previamente reportados en la literatura (Buniello et al., 2019; Landrum et al., 2018).
- Proyectos enfocados en caracterizar la diversidad genética humana como *1000 Genomes Project* (1KGP) que contienen frecuencias alélicas para una importante cantidad de poblaciones (Auton et al., 2015; Sudmant et al., 2015).

La implementación de los estudios de asociación en todo el genoma (GWAS) con características tales como el índice de masa corporal o los niveles tanto de colesterol y triglicéridos en plasma han permitido demostrar que las enfermedades dislipidémicas son de índole poligénico y permiten la identificación de posibles dianas moleculares para el desarrollo de potenciales terapéuticos (Berberich & Hegele, 2019; Dron et al., 2020; Johansen et al., 2011, 2014; Luo et al., 2020). Una aplicación destacada del uso de tecnologías de secuenciación masiva para el estudio de dislipidemias es el panel de secuenciación denominado “*LipidSeq*” (Dron et al., 2020). Este panel ha permitido generar índices de susceptibilidad a trastornos en el metabolismo de lípidos en individuos, ayudando a diagnosticar casos complejos, afinando el abordaje clínico de los pacientes y, de paso, identificando variantes que previamente no habían sido reportadas (Dron et al., 2020).

## Justificación

Las enfermedades no transmisibles son responsables del 71% de las muertes a nivel mundial, siendo una de estas las enfermedades cardiovasculares (World Health Organization, 2018). En Costa Rica alrededor del 83% de las muertes son vinculadas a enfermedades no transmisibles y alrededor de un 29% del total de muertes se atribuye a enfermedades cardiovasculares, esto las convierte en la principal causa de muerte en el país (Instituto Nacional de Estadística y Censos, 2019; World Health Organization, 2018). Si bien múltiples factores participan en el desarrollo de estas enfermedades, se sabe que las hiperlipidemias, la hipercolesterolemia en particular, tienen un papel protagónico en su prevalencia (Bruikman et al., 2017; Paredes et al., 2019; World Health Organization, 2009). Esto se refleja en estadísticas de la Organización Mundial de la Salud (OMS) que posicionan la incidencia de colesterol alto entre los 10 principales factores de riesgo de mortalidad (World Health Organization, 2009).

El estudio de las hiperlipidemias es relevante para la salud pública no solo por el impacto que tienen en las enfermedades vasculares sino también por el papel que tienen en el desarrollo de otros padecimientos, entre ellos la pancreatitis aguda. Se estima que la hipertrigliceridemia es uno de los principales causantes del 10% de los casos de pancreatitis (de Pretis et al., 2018). Esta enfermedad puede generar secundariamente condiciones como problemas cardiacos, trombosis, encefalopatía y se ha observado que puede incrementar el riesgo de desarrollar cáncer de páncreas, uno de los tipos de cáncer con menor tasa de supervivencia (Kirkegård et al., 2018; Muniraj et al., 2012; Stroes et al., 2017).

La incidencia de dislipidemias en algunas ciudades latinoamericanas alcanza valores tan altos como el 75% en hombres y 62% en mujeres (Vinueza et al., 2010). En Costa Rica, la publicación más reciente sobre el tema encontró, basado en 107.000 residentes de la provincia de San José, que un 36% de los hombres y un 22% de las mujeres presentaban hipertrigliceridemia,

mientras que los hombres presentaron hipercolesterolemia en un 48% y las mujeres un 52% (Gutiérrez et al., 2010). Si bien los estudios epidemiológicos sobre el tema no son abundantes o actuales, existe un mayor vacío en lo que respecta al papel que tiene el componente genético sobre las dislipidemias en la región latinoamericana y en menor medida en Costa Rica. En la Universidad de Costa Rica, se desarrollaron dos trabajos de investigación, uno dirigido a caracterizar variantes genéticas asociadas al gen *LPL* y otro enfocado en la región promotora y la región codificante del gen *APOCII* presentes en un grupo de 38 individuos con hipertrigliceridemia (González Cordero, 2018; Gutiérrez Ávila, 2019). De los 38 individuos evaluados, se detectaron polimorfismos genéticos en los exones del gen *LPL* en 18 individuos, a partir de los cuales se identificaron en total 11 variantes potencialmente patogénicas (González Cordero, 2018). Por otro lado, se identificaron cuatro variantes puntuales en la región promotora del gen *APOCII*. Tres de estas variantes fueron previamente identificadas. Particularmente, la variante rs12721063 se observó que putativamente podía afectar el sitio de unión del factor de transcripción CIIC, la misma se observó en uno de los 38 individuos (Gutiérrez Ávila, 2019).

El presente estudio pretende caracterizar en costarricenses por primera vez la variación genética albergada en genes cuya afectación ha sido vinculada con el desarrollo de hiperlipidemias. Para ello, se analizaron genomas pertenecientes a un grupo de costarricenses del Valle Central no emparentados y sin enfermedad diagnosticada. En primera instancia, se caracterizan las frecuencias de variantes previamente identificadas en genes involucrados en el desarrollo de dislipidemia en costarricenses y en poblaciones mesoamericanas. Asimismo, se identifican variantes de riesgo las cuales no habían sido identificadas costarricenses a la fecha. La información generada a partir de este estudio pretende ayudar a orientar y contextualizar futuros estudios sobre dislipidemias en Costa Rica y la región. A mediano y largo plazo, estudios como este pueden ayudar en el abordaje clínico de pacientes con cuadros complejos de hiperlipidemia desde la perspectiva de la medicina de precisión.

## Objetivos

### *Objetivo general*

Identificar variantes genéticas vinculadas a genes asociados a hiperlipidemias en los genomas de un grupo del Valle Central costarricense para estimar las frecuencias alélicas.

### *Objetivos específicos*

1. Establecer un flujo de análisis bioinformático para realizar la identificación de variantes genómicas y la cuantificación de las respectivas frecuencias alélicas usando paquetes y algoritmos preexistentes.
2. Identificar variantes genéticas en genes vinculados con el desarrollo de hiperlipidemias citados en el Cuadro 3 para cuantificar su frecuencia alélica en el grupo costarricense.
3. Comparar las frecuencias alélicas poblacionales obtenidas en el grupo costarricense con las reportadas en poblaciones descritas en *1000 Genomes Project*.
4. Cuantificar las frecuencias alélicas de variantes genéticas en los genes de interés previamente descritas en costarricenses para complementar estudios previos.

## Materiales y métodos

### *Sujetos y datos genómicos*

Se utilizaron datos genómicos provenientes del estudio inscrito ante la Vicerrectoría de Investigación de la Universidad de Costa Rica bajo el identificador B5331 denominado “Secuenciación del genoma completo en familias con trastornos psicóticos y afectivos” (Chavarria-Soley et al., 2021). Este incluye 214 genomas completos de costarricenses del Valle Central, 99 de ellos forman parte de tríos y 115 son parte de una familia extendida. El comité Ético Científico de la Universidad de Costa Rica aprobó el uso de estos datos en esta investigación.

La secuenciación se realizó utilizando el equipo Illumina HiSeq 2000 con extremos pareados. Los datos cuentan con una cobertura mínima de 30x y sus lecturas tienen una extensión mínima de 100pb. Los datos fueron previamente alineados con la herramienta BWA-MEM del paquete BWA v0.7.15 utilizando como genoma de referencia GRCh38 y almacenados en formato CRAM.

Con la finalidad de tener estimaciones de frecuencias alélicas las cuales reflejen una muestra aleatoria de costarricenses del Valle Central, se incluyeron al estudio únicamente individuos no emparentados y sin enfermedad diagnosticada. De los 214 genomas disponibles, 72 cumplieron con dicho criterio, por tanto, fueron seleccionados para ser incluidos en el estudio. Se hará referencia a este grupo de acá en adelante como CR-WGS.

**Cuadro 3.** Coordenadas genómicas caracterizadas en la etapa de llamado de variantes.

Rol	Cohorte	Identificador	Fuente de coordenadas	Fuente de identificadores
<b>Análisis de control de calidad</b>	CR-WGS AMR-80	Coordenadas genómicas descritas en el <i>Twist Human Comprehensive Exome Panel</i> para GRCh38 <a href="https://www.twistbioscience.com/sites/default/files/resources/2020-09/Twist_ComprehensiveExome_targets_hg38.bed">https://www.twistbioscience.com/sites/default/files/resources/2020-09/Twist_ComprehensiveExome_targets_hg38.bed</a>		
<b>Proveer una amplia cantidad de variantes para entrenar la herramienta de GATK, 'Variant Quality Score Recalibration' (VQSR)</b>	CR-WGS AMR-80	Coordenadas genómicas descritas en el <i>Twist Human Comprehensive Exome Panel</i> para GRCh38 <a href="https://www.twistbioscience.com/sites/default/files/resources/2020-09/Twist_ComprehensiveExome_targets_hg38.bed">https://www.twistbioscience.com/sites/default/files/resources/2020-09/Twist_ComprehensiveExome_targets_hg38.bed</a>		
<b>Contrastar ancestría con grupos costarricenses previamente reportados (variantes de dbSNP)</b>	CR-WGS	78 variantes registradas en dbSNP	<b>Variantes registradas en dbSNP:</b> Base de datos <i>Ensembl Genes 105</i> , genoma GRCh38.p13. Coordenadas extraídas con <i>BioMart</i>	Campos-Sánchez et al. 2013
<b>Contrastar ancestría con grupos de América reportados en <i>IKGP phase 3</i> (variantes de dbSNP)</b>	CR-WGS	250 variantes registradas en dbSNP	<b>Variantes registradas en dbSNP:</b> Base de datos <i>Ensembl Genes 105</i> , genoma GRCh38.p13. Coordenadas extraídas con <i>BioMart</i>	Wang et al., 2019
<b>Identificar variantes en genes y sus promotores involucrados en el metabolismo de lípidos y el desarrollo de dislipidemias (genes)</b>	CR-WGS	<b>Símbolos genéticos:</b> ABCA1, ABCG1, ABCG4, ABCG5, ABCG8, ABHD5, ANGPTL3, APOA1, APOA2, APOA4, APOA5, APOB, APOC1, APOC2, APOC3, APOC4, APOD, APOE, APOF, APOH, APOL1, APOL2, APOL3, APOL4, APOL5, APOL6, APOM, APOO, CD36, CELSR2, CETP, CILP2, CREB3L3, CYP26A1, FADS1, FADS2, FADS3, GALNT2, GCKR, GPD1, GPIHBP1, HMGCR, KLHL8, LCAT, LDLR, LDLRAP1, LIPA, LIPC, LIPE, LIPG, LMF1, LPL, LRP1, MLXIPL, MTPP, MYLIP, NCAN, NPC1L1, PCSK9, PLA2G7, PLIN1, PLTP, PNPLA2, PPARA, SCARB1, SORT1, STAP1, TRIB1, USF1	<b>Símbolos genéticos:</b> Base de datos <i>Ensembl Genes 105</i> , genoma GRCh38.p13. Coordenadas extraídas con <i>BioMart</i> <b>Promotores:</b> Obtenidos de <i>Eukaryotic Promoter Database</i> , -500pb y +100pb del sitio de inicio de transcripción más probable	Dron et al., 2019, 2020; Johansen et al., 2011, 2014; Johansen & Hegele, 2011; Lewis et al., 2015; Nakayama et al., 2010; Plaisier et al., 2009; Sarraju & Knowles, 2019; Vazquez Vidal, 2014
<b>Cuantificar frecuencias alélicas de variantes previamente reportadas en costarricenses y latinoamericanos en genes involucrados en el metabolismo de lípidos y el desarrollo de dislipidemias (genes)</b>	CR-WGS	<b>Variantes registradas en dbSNP:</b> rs12740374, rs11206510, rs4245791, rs6882842, rs1231383321, rs118204057, rs268, rs316, rs328, rs9282541, rs174570, rs2070895, rs6511720, rs12721063, rs2288912, rs919639084, rs2288911 <b>Variantes no reportadas en dbSNP, identificador HGMD:</b> LPL(NM_000237.3):c.655C>T, LPL(NM_000237.3):c.680T>G, LPL(NM_000237.3):c.1420_1421insG, LPL(NM_000237.3):c.123_124insA, LPL(NM_000237.3):c.1163_1164insA, LPL(NM_000237.3):c.1164_1165insA, LPL(NM_000237.3):c.1277_1278insA, LPL(NM_000237.3):c.1304_1305insG, LPL(NM_000237.3):c.1317G>C	<b>Variantes registradas en dbSNP:</b> Base de datos <i>Ensembl Genes 105</i> , genoma GRCh38.p13. Coordenadas extraídas con <i>BioMart</i>	Andaleon, Mogil & Wheeler 2019; González Cordero 2018; Gutierrez Ávila 2019; Villareal-Molina et al. 2012; Yang et al. 2004; Yu et al. 2017

Por otro lado, con la finalidad de evaluar el desempeño de la etapa de llamado de variantes y determinación de genotipos, se incluyeron exomas de individuos provenientes de individuos latinoamericanos (AMR) secuenciados y genotipados dentro del marco del *1000 Genomes Project* (1KGP), fase 3 (Auton et al., 2015). Se seleccionaron 80 individuos aleatoriamente a partir del grupo de 536 sujetos latinoamericanos enlistados, constituido por sujetos puertorriqueños (PUR), colombianos (CLM), mexicanos (MXL) y peruanos (PEL). Estos exomas fueron secuenciados utilizando el equipo Illumina HiSeq 2000 y fueron alineados contra el genoma GRCh38 utilizando BWA-ALN, parte del paquete BWA v0.5.9. Los exomas se descargaron en formato CRAM de [http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/1000\\_genomes\\_project/data/](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/data/) en diciembre 2022. Se eligieron individuos latinoamericanos debido a que se han observado diferencias en ciertas estadísticas utilizadas para evaluar la calidad del llamado de variantes entre poblaciones humanas (Wang et al., 2015). Este grupo será referido como AMR-80.

### *Descubrimiento de variantes y determinación de genotipos*

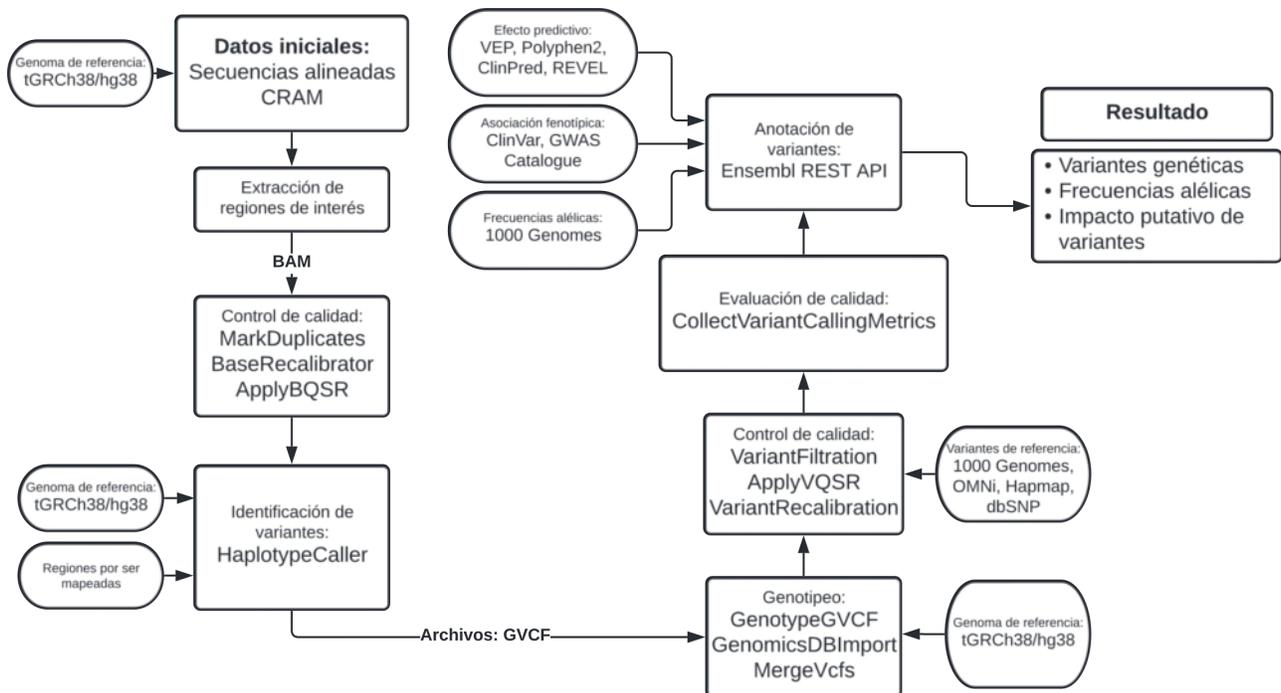
Se efectuó el proceso de llamado de variantes y determinación de genotipos para ambas cohortes de manera independiente. El procesamiento bioinformático aplicado en esta etapa a los datos CR-WGS se diferenció únicamente al efectuado para los de AMR-80 por las coordenadas a las que se limitó el análisis. Las regiones analizadas por cohorte se encuentran detalladas en el Cuadro 3. Se extendió cada intervalo de coordenadas 500pb corriente arriba y abajo ya que se ha observado que la ausencia de esta extensión o *padding* puede afectar negativamente el procesamiento bioinformático (Zanti et al., 2021).

El flujo de análisis bioinformático implementado se encuentra resumido en la Figura 2. Este análisis se implementó en el clúster computacional Kabré (CENAT) y el clúster del CICIMA (UCR). El preprocesamiento de los archivos CRAM y el descubrimiento de variantes se implementó basado

en las mejores prácticas para el llamado de variantes *Germline short variant discovery (SNPs + Indels)* con GATK v4.2.1.0 descritas por el *Broad Institute* (Poplin et al., 2017).

Como medida de control de calidad sobre las lecturas, primeramente, se eliminaron las lecturas duplicadas utilizando la herramienta MarkDuplicates, la cual forma parte del paquete GATK. Seguidamente, para ajustar los errores sistemáticos observados ocasionados por el secuenciador, se implementó el modelo de aprendizaje de máquinas de GATK denominado *Base Quality Score Recalibrator* utilizando los comandos BaseRecalibrator y ApplyBQSR.

Para el llamado de variantes tipo indel o SNV, se utilizó HaplotypeCaller, GenomicsDBImport, GenotypeGVCF, y MergeVcfs. Durante este proceso, se seleccionó tGRCh38/hg38 como el genoma de referencia y la base de datos de variantes dbSNP Build 151, se utilizó como fuente de referencia para las variantes.



**Figura 2.** Flujo de análisis bioinformático basado en las buenas prácticas de llamado de variantes propuestas por GATK.

Como control de calidad sobre las variantes identificadas, se calculó un puntaje de error referido como VQSLOD para las variantes identificadas utilizando el modelo de aprendizaje de máquinas de GATK, *Variant Quality Score Recalibrator* (VQSR). Para ello, métricas obtenidas para cada variante son alimentadas al modelo VQSR, entre ellas la profundidad de variante, el sesgo de hebra, y la calidad de la variante asignada en la etapa anterior, junto a listas de variantes con diferentes grados de confianza (Depristo et al., 2011). La evaluación de los errores de llamado de variantes se realizó para las indels y SNVs por separado.

Las bases de datos suministradas al modelo VQSR son las sugeridas por GATK y se encuentran almacenadas en su repositorio “*Resource bundle*” “*genomics-public-data*”, a excepción de la base de datos *dbSNP v151*, la cual fue extraída del sitio FTP del *National Center for Biotechnology Information* de los Estados Unidos o NCBI. Para calcular el puntaje de error en los indels, se consideraron como variantes verdaderas las altamente validadas en el set de datos *Mills and 1000 genomes gold standard* (Mills et al., 2006). Se utilizaron como datos de entrenamiento los genotipos de los sujetos provenientes de la primera fase del estudio 1KGP obtenidos con el chip *Axiom Exome Plus*. La base de datos *dbSNP v151* también fue suministrada al modelo, pero se consideró como una base de datos con menor grado de validación.

Para calcular los puntajes de error para los SNVs, se consideraron como variantes verdaderas las que se encontraban en la base de datos *HapMap phase 3 release 3*, parte del *International HapMap Project* (Altshuler et al., 2010). Se definieron como bases de datos de entrenamiento el panel de la fase 3 de 1KGP genotipados con el chip *OMNI 2.5* junto a la base de datos constituida por genotipos con alto nivel de confianza provenientes de la fase 1 de 1KGP. Finalmente, la base de datos de *dbSNP* se utilizó como fuente de referencia para las variantes conocidas, pero con escasa validación. Utilizando *ApplyVQSR* se excluyeron de análisis posteriores las variantes que tuvieran un VQSLOD menor al 97.5% de las variantes tipo SNVs y 95% de las variantes tipo indel.

### *Evaluación del procesamiento bioinformático*

Utilizando la herramienta de GATK CollectVariantCallingMetrics se calculó la relación transición vs transversión ( $T_i/T_v$ ) y la relación heterocigota vs homocigoto de alelo alternativo (Het/non-ref Hom), métricas utilizadas comúnmente para describir la calidad del proceso de llamado de variantes. El análisis se limitó a las coordenadas exónicas mencionadas en el Cuadro 3 y efectuó tanto para la cohorte CR-WGS como para la cohorte AMR-80. Únicamente se incluyeron estas regiones genómicas al análisis debido a que las métricas  $T_i/T_v$  y Het/non-ref Hom para permitir una evaluación comparativa o *benchmarking* veraz, puesto que únicamente existen valores consenso de estas métricas para el exoma. Se obtuvieron estas métricas tanto para cada cromosoma por separado como a nivel de exoma. Los valores obtenidos se compararon entre cohortes y con los valores reportados por Wang et al. (2015).

Con la finalidad de evaluar la concordancia entre los genotipos identificados para AMR-80 con los obtenidos de manera independiente para los mismos sujetos por el consorcio de 1KGP, se seleccionaron aleatoriamente 300 variantes por cromosoma y se descargó el genotipo reportado por 1KGP para las mismas. Se excluyeron de la comparación los genotipos obtenidos para AMR-80 donde no fue posible identificar los 2 alelos. Se calculó la concordancia de los genotipos obtenidos para AMR-80 y los reportados para los mismos sujetos. Adicionalmente, con la intención de evaluar la concordancia entre las frecuencias alélicas obtenidas en estudios que involucran costarricenses del Valle Central elegidos de manera aleatoria, se generó un modelo lineal para contrastar las frecuencias previamente reportadas en publicaciones del *Costa Rica Heart Study* y las obtenidas para CR-WGS (Brown et al., 2003; Ruiz-Narváez et al., 2005, 2010).

### *Análisis de ancestría genética*

Para determinar si los sujetos incluidos en la cohorte CR-WGS presentan un perfil de ancestría el cual calza dentro del patrón observado en otras poblaciones latinoamericanas se

contrastaron los genotipos de 250 AIMs (*Ancestry Informative Markers*) descritos por Wang *et al.* (2019) con los de varias poblaciones americanas y ancestrales reportadas en 1KGP. Se usaron como grupos ancestrales las poblaciones Europa (EUR), África (AFR) y Asia del Este (EAS) descritas en el panel 1KGP (Auton *et al.*, 2015; Sudmant *et al.*, 2015). Debido a que la mayor parte de la ascendencia de los nativos americanos proviene de la población del este asiático (Wang *et al.*, 2019), y dado a la escasez de datos genómicos para este grupo poblacional, en este estudio se utiliza el grupo EAS como grupo ancestral alternativo. No se consideraron los sujetos del grupo de Barbados (ACB) como integrantes de AFR ni los habitantes de Utah (CEU) como parte del grupo EUR puesto que son habitantes de América. Fueron considerados como americanos los grupos ACB, CEU, PUR, CLM, MXL y PEL.

Se descargaron los genotipos de los 250 AIMs para 150 individuos seleccionados aleatoriamente por cada grupo ancestral (AFR, EUR, EAS) y para la totalidad de individuos ACB, CEU, CLM, MXL y PEL. Se extrajeron genotipos para la cohorte CR-WGS y se integraron estos genotipos con los de los sujetos provenientes de 1KGP. Se efectuó un análisis de componentes principales (PCA) utilizando la cantidad de alelos alternativos por AIM. Únicamente se incluyeron los AIMs sin genotipos faltantes. Por otra parte, se obtuvieron frecuencias alélicas de las poblaciones de América previamente mencionadas y las frecuencias de las subpoblaciones que forman parte de AFR, EUR y EAS. A partir de estas frecuencias alélicas se estimaron las relaciones de similitud entre dichos estos grupos utilizando TreeMix v1.13 (Pickrell & Pritchard, 2012).

Con el objetivo de evaluar si es concordante la ancestría observada en CR-WGS con el perfil reportado previamente para sujetos del Valle Central costarricense se utilizaron 78 AIMs descritos por Campos-Sánchez *et al.* (2013) para realizar un análisis de mezcla genética utilizando el programa STRUCTURE v2.3.4 (Hubisz *et al.*, 2009). Se descargaron los genotipos para los 78 marcadores de los sujetos seleccionados para los grupos ancestrales en la etapa anterior (AFR, EUR, EAS). Estos fueron integrados con los genotipos para tales AIMs obtenidos en CR-WGS y los reportados para los grupos costarricenses provenientes de la Zona Norte (CR-ZN), Zona Sur (CR-ZS), la región Caribe

(CR-RC) y del Valle Ventral (CR-VC) (Campos-Sánchez et al., 2013). Los genotipos de los grupos CR-ZN, CR-ZS, CR-RC y CR-VC fueron determinados por medio de la metodología Golden Gate de Illumina.

La base de datos integrada utilizada para los análisis en STRUCTURE fue conformada por 782 individuos (Hubisz et al., 2009). Los parámetros de corrida que se utilizaron fueron: '*Length of Burnin Period*' o la cantidad de iteraciones para reducir el efecto de la configuración inicial como 50000, el '*Number of MCMC Reps after Burnin*' o la cantidad de iteraciones del modelo para obtener estimativas certeras como 100000, se asumió que existía mezcla genética en los individuos, se especificó además que los grupos podían tener frecuencias alélicas correlacionadas y se definieron los grupos *EUR*, *AFR* y *EAS* como los grupos ancestrales. Con estos parámetros se realizaron 10 simulaciones asumiendo que existían 3 grupos ancestrales en la población. Estos resultados fueron unificados haciendo uso de CLUMPP y DISTRUCT por medio de la herramienta CLUMPAK (Jakobsson & Rosenberg, 2007; Kopelman et al., 2015; Rosenberg, 2004). Se generaron tres tipos de gráficos, uno que representa la mezcla genética, un diagrama ternario a partir de las proporciones unificadas, y un análisis de componentes principales (PCA) utilizando la cantidad de alelos alternativos por variante. Únicamente se incluyeron los AIMS con genotipos completos.

#### *Anotación de variantes*

Las variantes encontradas en las regiones de interés fueron anotadas con información albergada en *Ensembl release 104* utilizando su *API REST v14.1* (Cunningham et al., 2021). Se extrajeron predicciones de patogenicidad, asociaciones fenotípicas e información de genética poblacional para cada variante.

El tipo de variante fue determinada utilizando *Variant Effect Predictor* (VEP) v7 (Cunningham et al., 2021). Se generaron predicciones *in silico* de patogenicidad para variantes con cambio de sentido utilizando la herramienta tradicional PolyPhen2 (Flanagan et al., 2010), y dos

herramientas más recientemente desarrolladas, ClinPred y REVEL (Alirezaie et al., 2018; Gunning et al., 2021; Ioannidis et al., 2016). Para la anotación con asociaciones fenotípicas, *Ensembl API REST* hizo recurso de las bases de datos ClinVar y *NHGRI-EBI GWAS catalog* (Buniello et al., 2019; Landrum et al., 2018). Por otra parte, las variantes identificadas en los promotores fueron anotadas funcionalmente con GWAVA y FATHM-XF (Ritchie et al., 2014; Rogers et al., 2018).

Para contrastar las frecuencias poblacionales encontradas en el grupo CR-WGS con las reportadas en poblaciones caracterizadas extensivamente, fueron anotadas las frecuencias de los grupos de 1KGP, EAS, EUR, AFR, AMR y la totalidad de 1KGP (ALL). Se realizaron pruebas exactas de Fisher con la finalidad de determinar cuáles de las variantes encontradas presentan una frecuencia alélica distinta en el grupo de genomas costarricenses respecto a las poblaciones de 1KGP. Se utilizó un nivel de significancia del 0.05 ajustado con la corrección Bonferroni como umbral para determinar si la frecuencia entre dos poblaciones es distinta.

### *Identificación y caracterización de variantes de interés*

El estudio consideró un sitio polimórfico como variante de interés si (1) es una variante de riesgo de acuerdo con dos o más fuentes de anotación funcional o si (2) la variante fue previamente reportada en Costa Rica o Latinoamérica dentro del contexto de metabolismo de lípidos y dislipidemias. Dicho esto, se generaron tres listas de variantes de interés, una constituida por variantes consideradas como riesgosas por predicciones bioinformáticas ubicadas dentro de los genes reportados en el Cuadro 3, otra contiene variantes de riesgo encontradas en los promotores de dichos genes y la última incluye las variantes que se han reportado en costarricenses y latinoamericanos en los genes y promotores de interés.

La lista de variantes de riesgo según predicciones bioinformáticas localizadas en las regiones codificantes de los genes está conformada por variantes que cumplen al menos dos de los siguientes criterios: (1) ser categorizadas por PolyPhen2 como variantes posiblemente dañinas (P) o

probablemente dañinas (D), (2) tener un índice calculado por REVEL mayor a 0.5, (3) que el puntaje de ClinPred fuera mayor a 0.5 o (4) que tuviera un fenotipo reportado por ClinVar o *NHGRI-EBI GWAS catalog* el cual tuviera relación con el metabolismo de lípidos o un riesgo aumentado a desarrollar y padecer dislipidemias. Por otro lado, se incluyeron a la lista de variantes de riesgo localizadas en los promotores de los genes, las variantes con un puntaje mayor a 0.5 en los tres puntajes GWAVA y que estuvieran categorizadas como variantes deletéreas por FATHM-XF. Para visualizar el consenso entre las diferentes fuentes en la determinación si una variante es de riesgo o no, se generaron diagramas de Venn utilizando la herramienta jVenn (Bardou et al., 2014). Las variantes de la tercera lista reportadas en estudios enfocados en analizar el rol del componente genético sobre el metabolismo de lípidos o el desarrollo de dislipidemias en Costa Rica y Latinoamérica. Finalmente, con el objetivo de determinar los procesos biológicos en los que participan los genes afectados por al menos una variante de interés, se efectuó un enriquecimiento de términos de ontología genética (GO) con una significancia  $p \leq 1.94e^{-18}$  utilizando la herramienta GOnet (Pomaznoy et al., 2018).

### *Código para análisis bioinformático*

Además de las herramientas mencionadas anteriormente, para la elaboración del estudio se utilizaron los lenguajes de programación libres Python 3.7 y R 4.1.2. Durante la etapa de llamado de variantes se utilizó Python para administrar el flujo de trabajo. Este también se utilizó para anotar las variantes, manipular los datos y generar visualizaciones de estos. R se utilizó para generar las visualizaciones producidas a partir de los resultados de TreeMix. Información referente al código puede encontrarse en repositorio de GitHub [https://github.com/jcvalverdehernandez/cr\\_dislipidemia\\_2022](https://github.com/jcvalverdehernandez/cr_dislipidemia_2022).

## Resultados

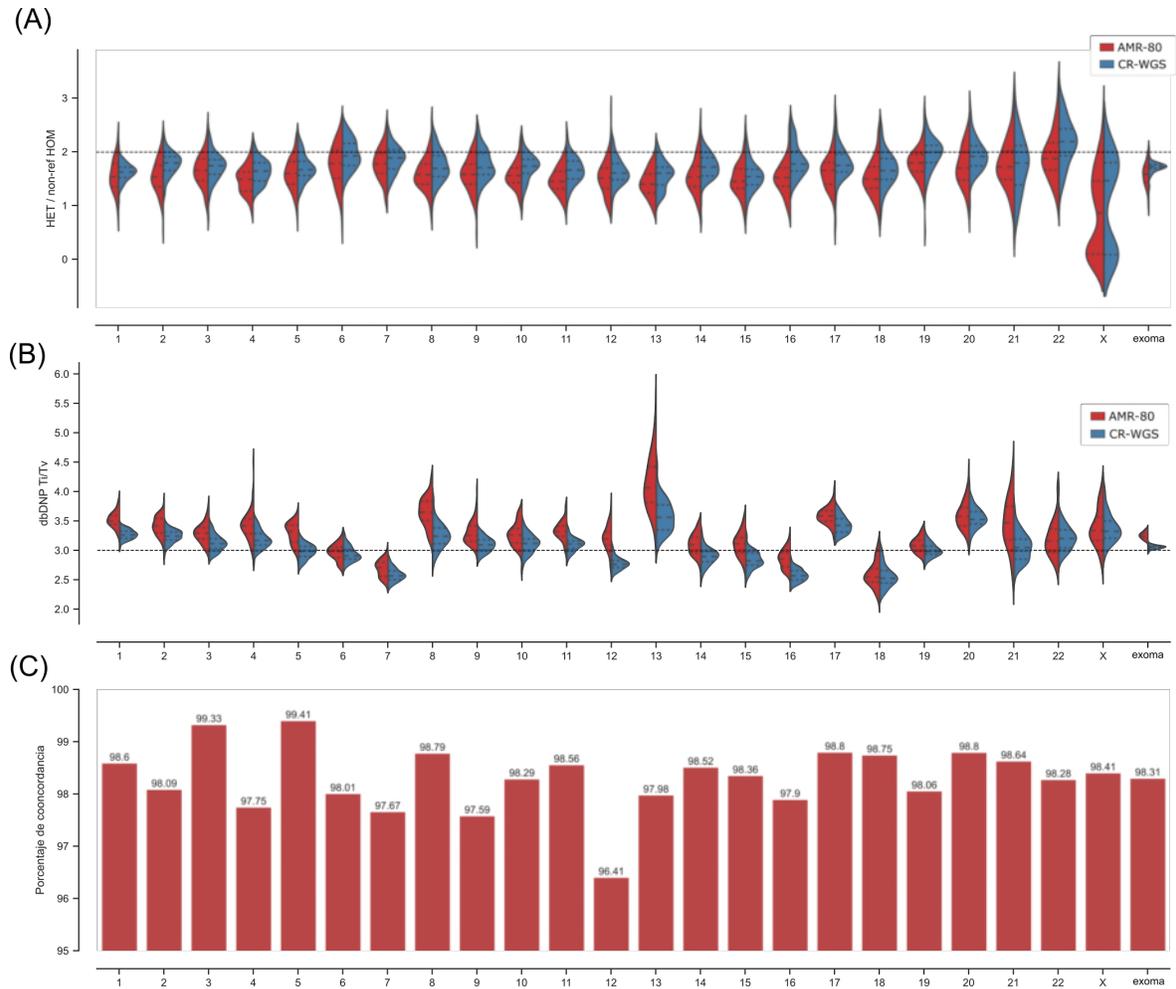
### *Métricas del llamado de variantes cumplen con los estándares de calidad*

La cantidad de variantes promedio por exoma, la relación  $T_i/T_v$  y la relación HET/non-ref HOM promedio obtenida para AMR-80 fueron consistentes con los valores considerados estándar en exomas de alta calidad (Wang et al., 2015). Por su parte, las métricas obtenidas para CR-WGS también fueron consistentes con los valores estándar a excepción de la cantidad de SNVs detectados por exoma en promedio, se encontraron más SNVs en el grupo CR-WGS que en el grupo AMR-80. Los genotipos encontrados para AMR-80 con el flujo de trabajo utilizado en este estudio fueron altamente consistentes a los reportados originalmente por 1KGP, lo cual sugiere que el llamado de variantes es consistente con las buenas practicas de llamado de variantes de GATK.

Se identificaron 18461 SNVs y 258 indels en promedio por muestra dentro de las regiones exónicas de AMR-80, mientras que en el grupo CR-WGS se encontró un promedio 22468 SNVs y 610 indels por muestra. Por otro lado, se encontró que, de manera generalizada, los sujetos que conforman AMR-80 presentaron un índice  $T_i/T_v$  de 3.25 que es mayor al observado en los sujetos de CR-WGS, 3.05 (Figura 3A). Dicho esto, el índice obtenido para ambas cohortes calza dentro de lo esperado según la literatura, se reporta que valores  $T_i/T_v$  que rondan 3.0 suelen indicar que los datos presentan una calidad adecuada (Wang et al., 2015). Por otro lado, la relación HET/non-ref HOM promedio observada para el exoma de AMR-80 fue de 1.57 y para CR-WGS fue de 1.72 (Figura 3B). El estándar general para este índice es de 2.0. No obstante, Wang et al. (2015) reportaron que el grupo AMR tiene un valor promedio cercano a 1.75, semejante al observado en el presente análisis. La distribución de datos observada para el índice HET/non-ref HOM es más consistente entre cohortes que la observada para el índice  $T_i/T_v$ .

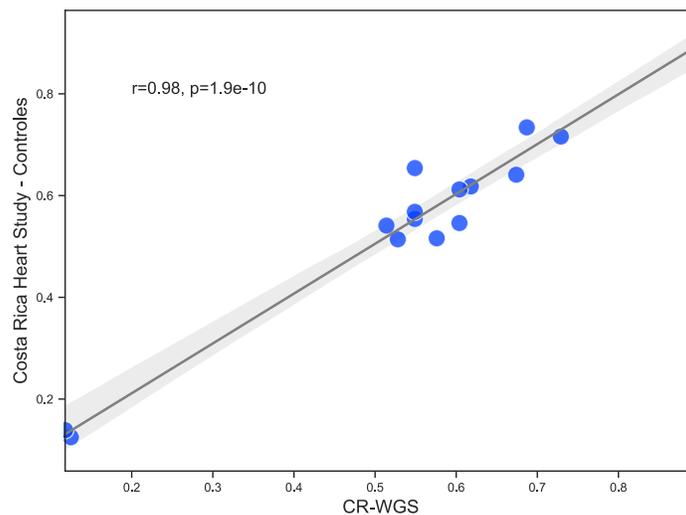
Se encontró una concordancia del 98.31% entre los genotipos obtenidos para la cohorte AMR-80 y los reportados por 1KGP utilizando el procesamiento bioinformático descrito en la Figura

2 , esto sugiere que el procesamiento bioinformático utilizado en este proyecto genera genotipos altamente fidedignos (Figura 3C). Finalmente, se encontró que las frecuencias alélicas previamente reportadas en diversos sitios polimórficos en estudios de *Costa Rica Heart Study* están correlacionados significativamente con los observados CR-WGS, lo que sugiere que existe una alta similitud entre estas cohortes (Figura 4).



**Figura 3.** Métricas de calidad para el proceso de llamado de variantes efectuado en los grupos CR-WGS y AMR-80.

(A) Relación  $T_i/T_v$  por cromosoma calculado a partir de variantes también reportadas en dbSNP, (B) Relación HET/non-ref/HOM por cromosoma, (C) Porcentaje de concordancia de genotipos obtenidos para AMR-80 con los reportados por 1KGP calculado para cada cromosoma. Las líneas punteadas demarcan el valor consenso en el estado del arte para cada métrica a nivel de exoma.



**Figura 4.** Relación lineal entre 15 frecuencias alélicas obtenidas en costarricenses del Valle Central de la cohorte CR-WGS y costarricenses sanos de la misma región elegidos aleatoriamente para el *Costa Rica Heart Study*.

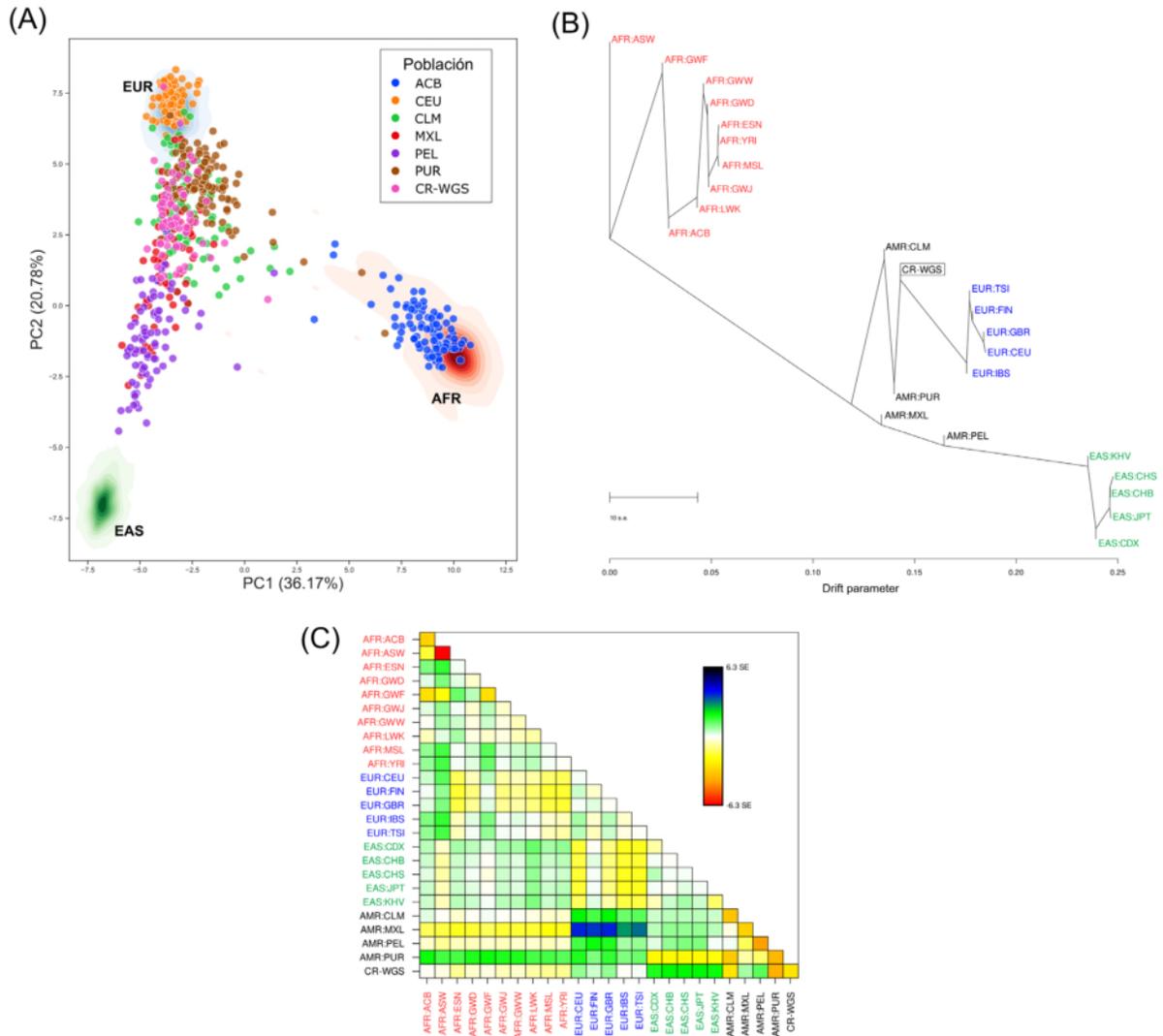
Las frecuencias alélicas del grupo *Costa Rica Heart Study* fueron extraídas de publicaciones asociadas (Brown et al., 2003; Ruiz-Narváez et al., 2005, 2010).

#### *Ancestría de genomas costarricenses es concordante con estudios previos*

Los análisis de ancestría realizados validan que la cohorte CR-WGS tiene un perfil genético consistente con el esperado para una muestra aleatoria de costarricenses del Valle Central. CR-WGS tiene un perfil consistente con los grupos latinoamericanos en 1KGP y es semejante a CR-VC, población aleatoria del Valle Central costarricense previamente descrita por Campos-Sánchez *et al.* (2013).

El análisis de componentes principales (PCA) logró captar alrededor de un 55% (entre el componente principal 1 y el 2) de la variación genética en el panel de 250 AIMs existente en los tres grupos ancestrales y los siete grupos americanos (Figura 5). El mismo revela claras diferencias entre los 250 AIMs entre los sujetos de los cinco grupos latinoamericanos (CR-WGS, CLM, PUR, MXL y PEL) y los de América Anglosajona (CEU, ACB). Los cinco grupos latinoamericanos presentan

un mayor grado de mezcla genética que los de América Anglosajona. Se observa que los sujetos costarricenses (CR-WGS) aparentan tener más similitud con los sujetos colombianos (CLM) y puertorriqueños (PUR) que con las muestras de sujetos mexicanos (MXL) y peruanos (PEL) presentes en 1KGP.



**Figura 5.** Similitud genética entre individuos americanos con base en genotipos de 250 AIMs reportados por Wang *et al.* (2019).

(A) Análisis de componentes principales. (B) Relaciones genéticas entre las poblaciones incluidas en el análisis de acuerdo con estimaciones de TreeMix. (C) Valores residuales del modelo representado en (B). Poblaciones de 1KGP destacadas - EUR: Europa del este, AFR: África, EAS: Asia del Este, ACB: Barbados, CEU: Utah, CLM: Colombia, MXL: México, PEL: Perú, PUR: Puerto Rico, CR-WGS: cohorte de genomas costarricenses del Valle Central analizados presente estudio.

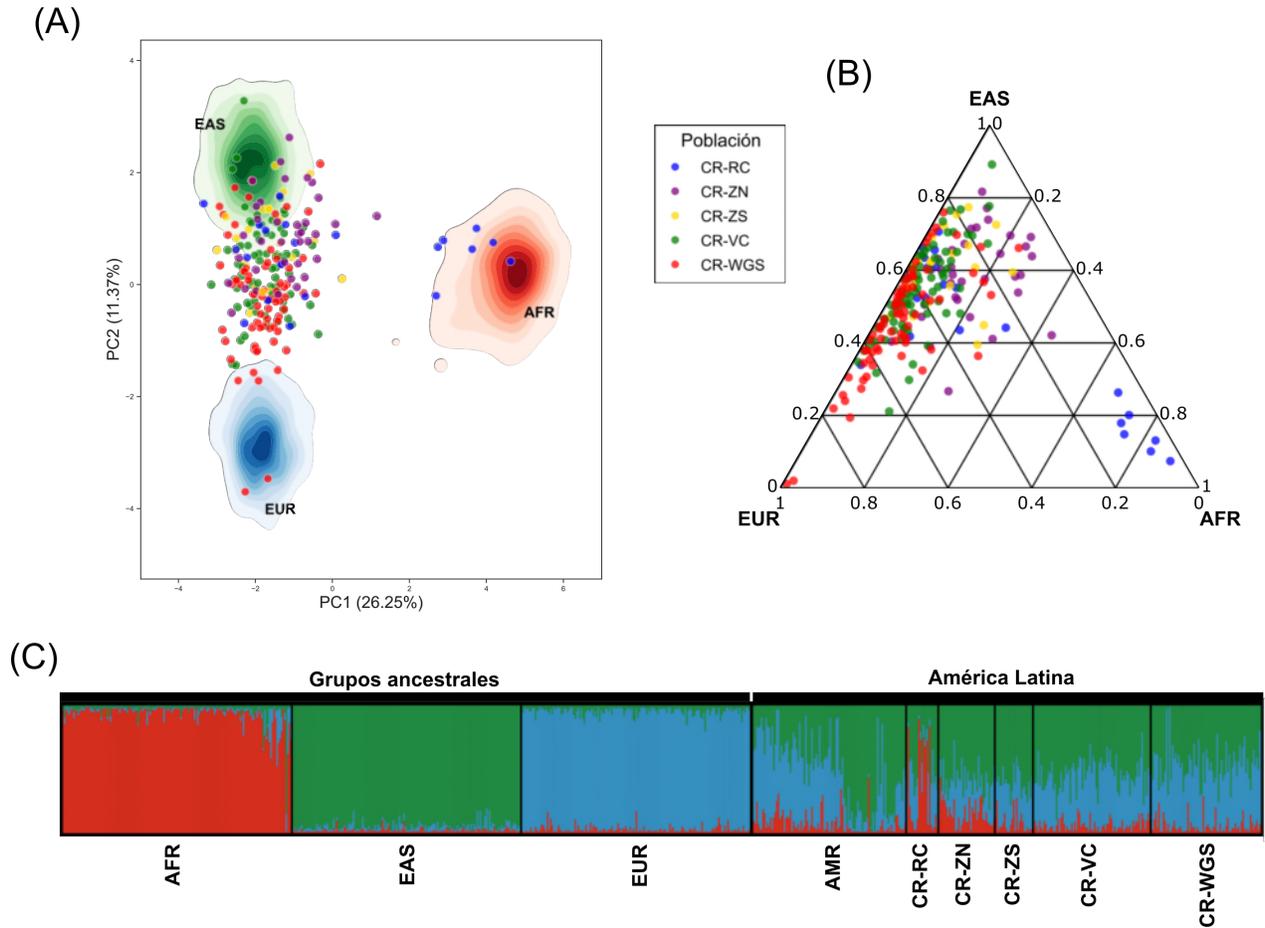
Al contrastar la mezcla genética de CR-WGS con la del resto de grupos costarricenses descritos por Campos-Sánchez *et al.* (2013), se observa que los grupos con los que CR-WGS tiene la mayor similitud son los costarricenses del Valle Central (CR-VC) y los de la Zona Sur (CR-ZN) (Figura 6). El PCA realizado en esta etapa logró captar aproximadamente un 35% de la variación genética encontrada entre el componente principal 1 y el 2.

#### *Sitios polimórficos identificados en genes y promotores de interés*

Se identificaron 17852 sitios polimórficos en el grupo CR-WGS ubicados en los genes de interés y sus promotores (Cuadro 3) constituidos por 15459 SNVs y 2568 indels (Cuadro 4). Se encontraron 290 variantes las cuales no habían sido previamente reportadas en dbSNP. Un 5.4% de las variantes detectadas presentaron más de un alelo alternativo. Los grupos de 1KGP que tuvieron las frecuencias alélicas más semejantes a las de CR-WGS fueron AMR, ALL y EUR. Se identificaron variantes de riesgo y variantes con relevancia farmacogenómica en los genes de interés que no habían sido reportadas previamente en Costa Rica. Por otra parte, algunas de las variantes previamente reportadas en costarricenses no fueron detectadas en la cohorte.

**Cuadro 4.** Estadísticas del llamado de variantes para el panel de genes y promotores involucrados en el metabolismo de lípidos mencionados en el Cuadro 3.

Métrica	Gen	Promotor	Total
<b>Variantes identificadas</b>	17593	232	17825
SNV	15052	205	15257
SNV dbSNP	14851	201	15052
SNV no reportado en dbSNP	201	4	205
SNV monoalélico	14769	201	14970
SNV multialélico	283	4	287
Indel	2541	27	2568
Indel dbSNP	2459	24	2483
Indel no reportado en dbSNP	82	3	85
Indel monoalélico	1870	23	1893
Indel multialélico	671	4	675



**Figura 6.** Mezcla genética observada en los individuos costarricenses calculada con base en genotipos de 78 AIMs reportados por Campos-Sánchez *et al.* (2013).

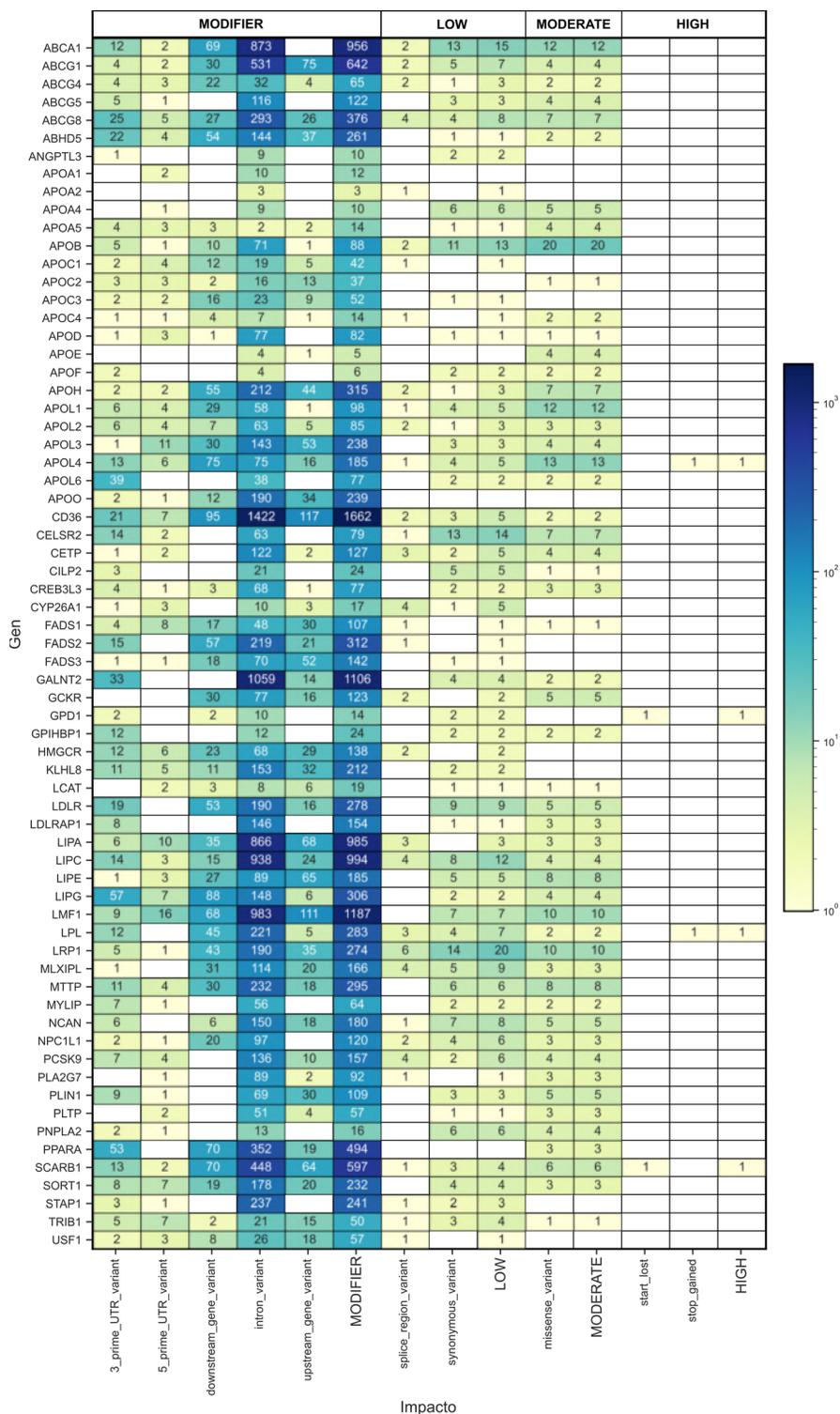
(A) Análisis de componentes principales, (B) Diagrama ternario de mezcla genética, (C) Diagrama de estructura genética. AFR: África, EAS: Asia del Este, EUR: Europa, AMR: América Latina, CR-RC: costarricenses de la Región Caribe, CR-ZN: costarricenses de la Zona Norte, CR-ZS: costarricenses de la Zona Sur, CR-VC: costarricenses de la Valle Central, CR-WGS: cohorte de genomas costarricenses del Valle Central analizados presente estudio.

De acuerdo con la anotación de impacto asignada por VEP basada en la consecuencia de la variante obtenida *in silico* según el término de *Sequence Ontology* (SO) asignado, 281 variantes tienen consecuencias esperadas de bajo riesgo de afectar transcritos del gen o *LOW*, 253 de riesgo medio o *MODERATE* y 4 de alto riesgo o *HIGH*. No fue posible asignarles un riesgo esperado a consecuencias asignadas a 14078 de las variantes utilizando VEP debido a la dificultad de predecir el impacto de las consecuencias, estas consecuencias son referidas como modificadoras o *MODIFIER* por VEP (Figura 7).

Las cuatro variantes asociadas a consecuencias de alto riesgo de acuerdo con VEP son: (1) rs328, causante de una mutación sin sentido en el gen LPL, (2) rs192225524, capaz de generar un cambio sin sentido en el gen APOL4, (3) rs144009925, capaz de suprimir el sitio de iniciación de GPD1 y (4) la variante rs749801989 ubicada en el gen SCARB1, también ubicada en el codón de iniciación. Por otro lado, ABCA1, APOL1, APOL4 y APOB fueron los genes con el mayor número de variantes con la capacidad de ocasionar modificaciones en el marco abierto de lectura.

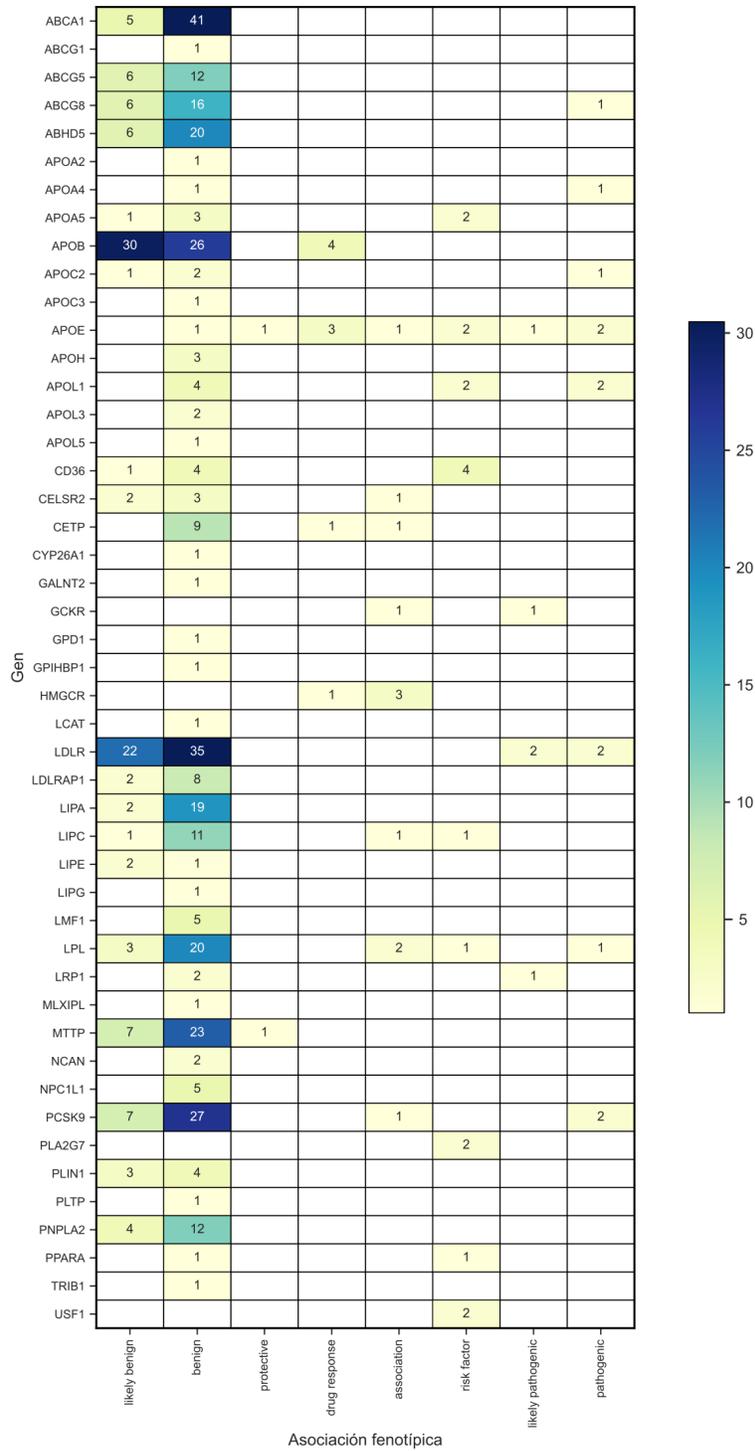
Producto de la anotación fenotípica, 115 variantes fueron asociadas con rasgos fenotípicos categorizados con una significancia clínica no benigna, es decir, términos distintos a *benign* o *likely benign* (Figura 8). Los genes con un mayor número de variantes con rasgos fenotípicos categorizados como factores de riesgo o patogénicos (*risk factor*, *pathogenic*, *likely pathogenic*) fueron APOE, APOL1 y CD36. Cabe destacar que nueve variantes fueron anotadas con rasgos asociados a farmacogenómica y dos con rasgos protectores (Cuadro 5).

De acuerdo con las pruebas exactas de Fisher implementadas para contrastar las frecuencias alélicas de CR-WGS y las de los grupos pertenecientes a 1KGP; los grupos AMR, ALL y EUR son los más semejantes a CR-WGS (Figura 9). Estos se diferenciaron individualmente de CR-WGS en menos de 700 variantes. Por otra parte, EAS y AFR presentaron diferencias estadísticamente significativas en la frecuencia de los alelos de más de 3000 sitios polimórficos con las encontradas en CR-WGS.



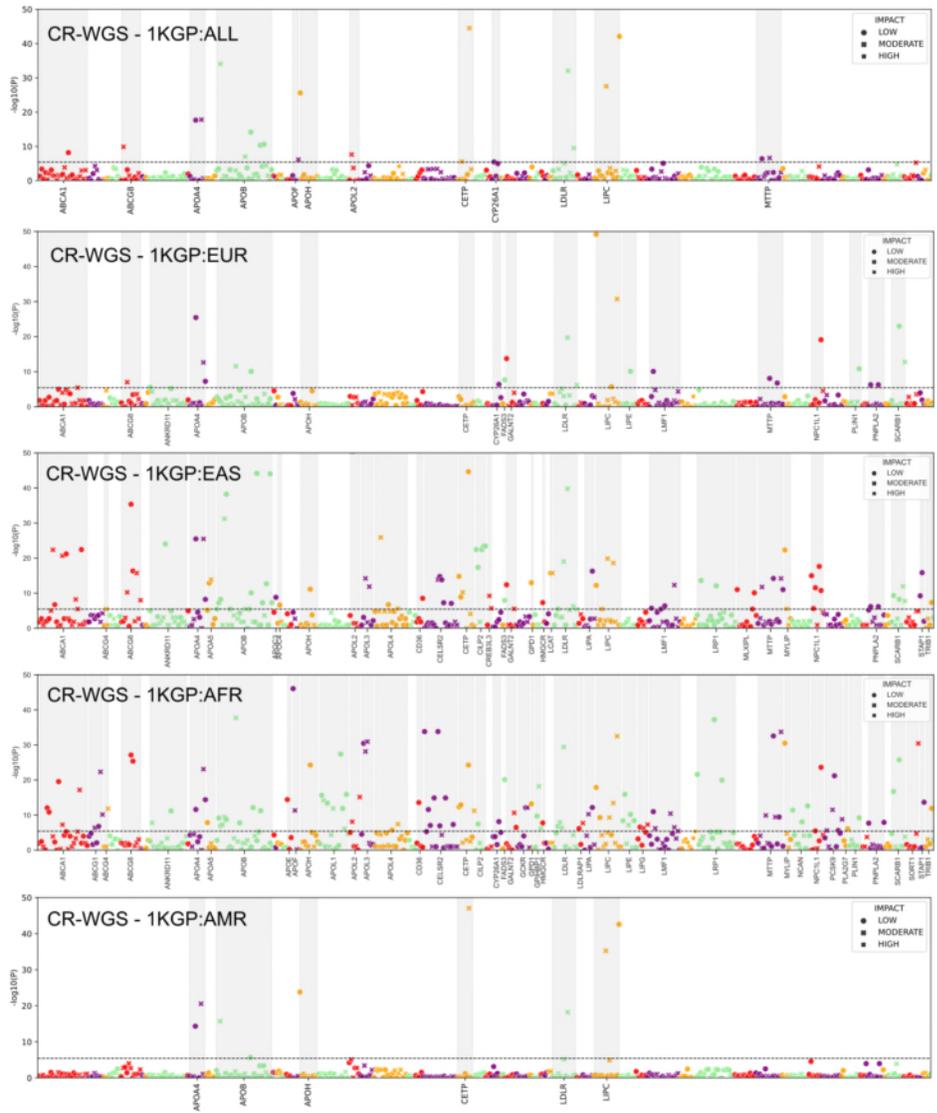
**Figura 7.** Conteo de variantes detectadas en genes involucrados en el metabolismo de lípidos categorizados por su impacto sobre el transcrito.

Categorización de acuerdo con Variant Effect Predictor (VEP).

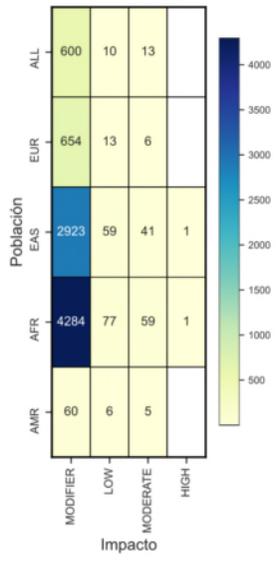


**Figura 8.** Significancia clínica de acuerdo con las asociaciones fenotípicas reportadas en ClinVar para las variantes identificadas en los genes de interés.

(A)



(B)

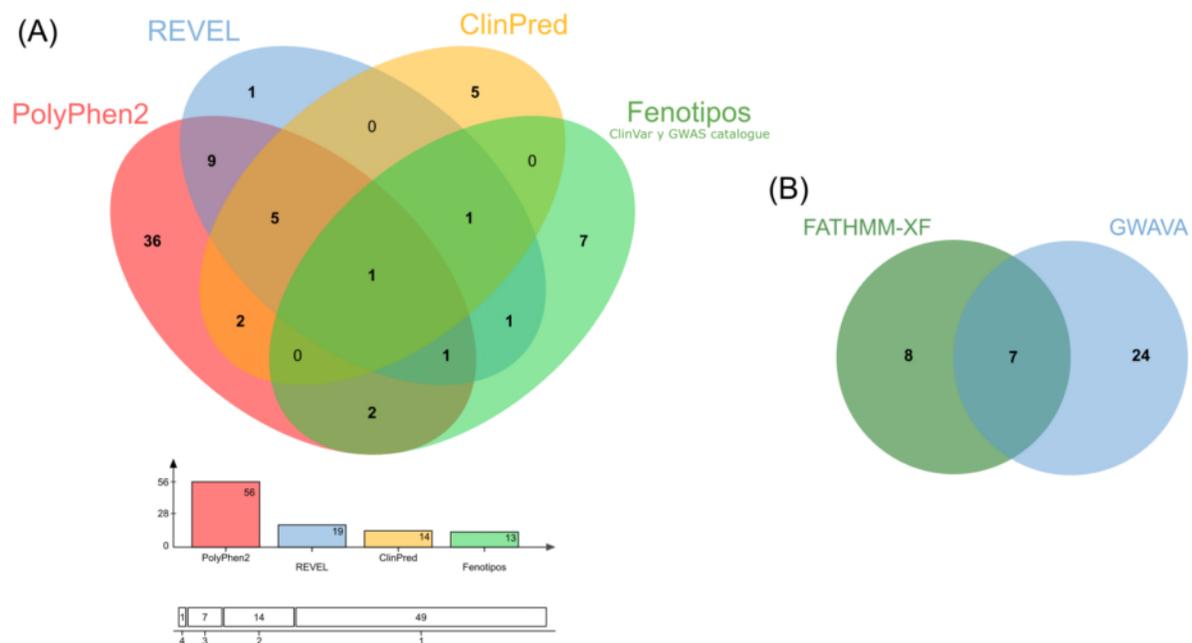


**Figura 9.** Diferencias observadas entre frecuencias alélicas en genes asociados a metabolismo de lípidos en costarricenses frente a las reportadas 1KGP.

(A) Probabilidad de acuerdo con pruebas de Fisher de que los sitios polimórficos tengan diferencias en sus frecuencias alélicas. La línea punteada demarca el umbral de significancia con el ajuste Bonferroni. Las variantes categorizadas como LOW, MODERATE y HIGH por VEP son representadas. (B) Cantidad de variantes con frecuencias alélicas significativamente distintas a las observadas en la cohorte costarricense estudiada (CR-WGAS). ALL: Todos los sujetos reportados en 1KGP, EUR: Europa, EAS: Asia del Este, AFR: África, AMR: América Latina.

De las variantes que cambian el sentido identificadas dentro de los genes de interés mencionados en el Cuadro 3, 22 fueron categorizadas como variantes de riesgo por más de dos de

las fuentes utilizadas para la anotación funcional (Figura 10A, Anexo 1). Las variantes fueron compiladas junto con su información funcional en el Cuadro 6. Del mismo modo, 7 variantes de riesgo fueron identificadas en los promotores de tales genes de interés (Figura 10B), estos se enlistan en el Cuadro 7. Finalmente, las variantes previamente vinculadas al metabolismo de lípidos o el desarrollo de dislipidemias en costarricenses y latinoamericanos se enlistan en el Cuadro 8. Algunas de las variantes del Cuadro 8 no fueron encontradas en CR-WGS.



**Figura 10.** Concordancia entre fuentes utilizadas para identificar variantes de interés de acuerdo con su patogenicidad o su asociación con alteraciones en el perfil lipídico.

(A) Diagrama de Venn con categorización de variantes cambio de sentido encontradas en los genes asociados al metabolismo de lípidos y el desarrollo de hiperlipidemias. (B) Diagrama de Venn con categorización de en promotores de genes de interés.

**Cuadro 5.** Variantes encontradas en genes de interés asociadas fenotípicamente a rasgos con significancia clínica farmacogenómica o protectora ante enfermedades.

CR: Cohorte nombrada CR-WGS, genomas costarricenses evaluados en el presente estudio. Grupos de 1KGP - ALL: todos los sujetos, EAS: Asia del Este, EUR: Europa, AFR: África, AMR: América Latina. \*Frecuencia alélica significativamente distinta ( $p < 0.05$ ) a la observada en costarricenses estudiados.

Gen	Variante (dbSNP)	Alelos (REF/ALT)	Frecuencia alélica de alelo alternativo						Rasgos protectores o farmacogenómicos asociados
			CR	ALL	EUR	EAS	AFR	AMR	
APOB	rs1367117	G/A	<b>0.29900</b>	0.19300	0.29821	<b>0.11507*</b>	0.07791	0.28674	<b>Alelo A según ClinVar:</b> Respuesta ante Warfarina ( <i>Warfarin response</i> )
CETP	rs1532624	C/T	<b>0.33300</b>	0.25427	0.42644	0.29067	<b>0.10438*</b>	0.34438	<b>Alelo A según ClinVar:</b> Respuesta ante inhibidores de HMG CoA reductasa ( <i>HMG CoA reductase inhibitors response – Efficacy</i> )
HMGCR	rs17244841	A/T	<b>0.00694</b>	0.07553	0.01789	0.00595	0.10741	0.02594	<b>Alelo T según ClinVar:</b> respuesta ante pravastatina ( <i>pravastatin response – Efficacy</i> ), respuesta ante simvastatina ( <i>simvastatin response – Efficacy</i> ), Respuesta ante inhibidores de HMG CoA reductasa ( <i>HMG CoA reductase inhibitors response – Efficacy</i> ), reducción atenuada de colesterol por estatinas ( <i>Statins - attenuated cholesterol lowering by</i> )
APOE	rs3816873	T/C	<b>0.19400</b>	0.25894	0.26044	0.13591	0.26097	0.17867	<b>Alelo C según ClinVar:</b> protección contra síndrome metabólico ( <i>Metabolic syndrome, protection against</i> )
APOE	rs429358	T/C	<b>0.13200</b>	0.17702	0.15507	0.08631	0.26778	0.10375	<b>Alelo C según ClinVar:</b> Respuesta ante Warfarina ( <i>Warfarin response</i> )
APOB	rs679899	G/A	<b>0.34000</b>	0.37588	0.47416	<b>0.86408*</b>	<b>0.13011*</b>	0.39193	<b>Alelo A según ClinVar:</b> Respuesta ante Warfarina ( <i>Warfarin response</i> )
APOE	rs7412	C/T	<b>0.08300</b>	0.08875	0.06262	0.10020	0.10287	0.04755	<b>Alelo T según ClinVar:</b> respuesta ante atorvastatina ( <i>atorvastatin response – Efficacy</i> ), Respuesta ante Warfarina ( <i>Warfarin response</i> ) <b>Alelo T según NHGRI-EBI GWAS catalog:</b> cambio de C-LDL como respuesta a niveles de estatinas ( <i>Response to statins (LDL cholesterol change)</i> )
APOE	rs769450	G/A	<b>0.31900</b>	0.35663	0.41153	0.21825	0.35023	0.29683	<b>Alelo A según ClinVar:</b> Respuesta ante Warfarina ( <i>Warfarin response</i> )

**Cuadro 6.** Variantes que producen alteraciones en el marco abierto de lectura encontradas en genes involucrados en el metabolismo de lípidos categorizadas como riesgosas por más de dos fuentes.

CR: Cohorte nombrada CR-WGS, genomas costarricenses evaluados en el presente estudio. Grupos de 1KGP - ALL: todos los sujetos, EAS: Asia del Este, EUR: Europa, AFR: África, AMR: América Latina. \* Frecuencia alélica significativamente distinta ( $p < 0.05$ ) a la observada en costarricenses estudiados.

Gen	Variante	Alelos (REF/ALT)	Frecuencia alélica del alelo alternativo						Modelos <i>in silico</i> categorizan variante como riesgosa	Anotación
			CR	ALL	EUR	EAS	AFR	AMR		
ABCA1	rs140365800	T/C	<b>0.00694</b>	0.00138	-	-	-	0.00144	PolyPhen2, REVEL	<b>Fuentes:</b> ClinVar <b>Términos:</b> Hipoalfalipoproteinemia familiar ( <i>Familial hypoalphalipoproteinemia</i> )
ABCG5	rs145164937	G/C	<b>0.02800</b>	0.00191	0.00199	-	-	0.00576	PolyPhen2, REVEL	-
ABCG8	rs11887534	G/C	<b>0.02800</b>	0.05798	0.07952	0.01389	0.07640	0.09654	PolyPhen2	<b>Fuentes:</b> ClinVar , NHGRI-EBI GWAS catalog <b>Términos:</b> niveles de proteína C reactiva y C-LDL ( <i>C-reactive protein levels or C-LDL cholesterol levels (pleiotropy)</i> ), <i>SITOSTEROLEMIA</i>
ABCG8	rs762225826	A/G	<b>0.00694</b>	0.00016	-	-	-	-	PolyPhen2, REVEL, ClinPred	-
APOA5	rs3135506	G/C	<b>0.09700</b>	0.06331	0.06759	-	0.06732	0.11671	PolyPhen2	<b>Fuentes:</b> NHGRI-EBI GWAS catalog <b>Términos:</b> niveles de colesterol total ( <i>Total cholesterol levels</i> ), susceptibilidad a hipertrigliceridemia ( <i>HYPERTRIGLYCERIDEMIA, SUSCEPTIBILITY TO</i> )
APOE	rs7412	C/T	<b>0.08300</b>	0.08875	0.06262	0.10020	0.10287	0.04755	PolyPhen2, REVEL	<b>Fuentes:</b> ClinVar , NHGRI-EBI GWAS catalog <b>Términos:</b> hipercolesterolemia ( <i>Hypercholesterolemia</i> ), Apolipoproteinemia E1, niveles de colesterol total, C-HDL y C-LDL ( <i>HDL cholesterol, LDL cholesterol levels, Total cholesterol levels</i> )
APOH	rs1801689	A/C	<b>0.01400</b>	0.01957	0.04076	0.00099	0.00151	0.03602	PolyPhen2, REVEL	-
CD36	rs1057086208	G/A	<b>0.01400</b>	0.00001	-	-	-	-	PolyPhen2, REVEL, ClinPred	-
CD36	rs146027667	G/T	<b>0.00694</b>	0.00058	-	-	-	-	PolyPhen2, REVEL	-
CELSR2	rs202022169	T/C	<b>0.00694</b>	0.00070	-	-	-	0.00432	PolyPhen2, REVEL	-
CREB3L3	rs779860332	C/A	<b>0.00694</b>	0.00024	-	-	-	-	PolyPhen2, REVEL, ClinPred	-

Gen	Variante	Alelos (REF/ALT)	Frecuencia alélica del alelo alternativo						Anotación	
			CR	ALL	EUR	EAS	AFR	AMR	Modelos <i>in silico</i> categorizan variante como riesgosa	Asociación fenotípica
GCKR	rs146175795	G/A	<b>0.02800</b>	0.00285	-	0.00694	-	0.02161	REVEL, ClinPred	<b>Fuentes:</b> ClinVar <b>Términos:</b> hipertrigliceridemia ( <i>Hypertriglyceridemia</i> )
LDLR	rs764042910	C/T	<b>0.01400</b>	0.00008	-	-	-	-	PolyPhen2, REVEL , ClinPred	<b>Fuentes:</b> ClinVar <b>Términos:</b> hipercolesterolemia familiar ( <i>Familial hypercholesterolemia</i> )
LIPE	rs138539064	G/A	<b>0.00694</b>	0.00242	-	-	-	-	PolyPhen2, REVEL	-
LIPE	rs750246327	G/A	<b>0.02100</b>	0.00003	-	-	-	-	PolyPhen2, ClinPred	-
LRP1	rs1185554986	C/T	<b>0.00694</b>	0.00001	-	-	-	-	PolyPhen2, ClinPred	-
LRP1	rs1377430371	C/T	<b>0.00694</b>	0.00002	-	-	-	-	PolyPhen2, REVEL , ClinPred	-
LRP1	rs1414626991	A/G	<b>0.01400</b>	0.00004	-	-	-	-	PolyPhen2, REVEL	-
LRP1	rs755040010	G/A	<b>0.00694</b>	0.00032	-	-	-	-	PolyPhen2, REVEL	-
NPC1L1	rs148405668	G/A	<b>0.00694</b>	0.00051	-	-	0.00227	-	PolyPhen2, REVEL	-
PPARA	rs1800206	C/G	<b>0.02800</b>	0.03620	0.05865	-	0.00530	0.03458	REVEL	<b>Fuentes:</b> ClinVar <b>Términos:</b> susceptibilidad a hiperapobetalipoproteinemia ( <i>HYPERAPOBETALIPOPROTEINEMIA, SUSCEPTIBILITY TO</i> )
PPARA	rs565902106	A/G	<b>0.00694</b>	0.00012	-	-	-	0.00144	PolyPhen2, REVEL , ClinPred	-

**Cuadro 7.** Variantes identificadas en promotores de genes involucrados en el metabolismo de lípidos categorizadas como riesgosas por más de dos fuentes.

CR: Cohorte nombrada CR-WGS, genomas costarricenses evaluados en el presente estudio. Grupos de 1KGP - ALL: todos los sujetos, EAS: Asia del Este, EUR: Europa, AFR: África, AMR: América Latina.

Gen	Elemento regulador	Variante	Alelos (REF/ALT)	Frecuencia de alelo alternativo						Antecedente científico
				CR	ALL	EUR	EAS	AFR	AMR	
<b>ABCA1</b>	ENSR00000889364	rs1800977	G/A	<b>0.38900</b>	0.318162	0.373757	0.343254	0.254917	0.43804	Niveles de triglicéridos y colesterol en HDLs asociados a prevalencia de alelo A en jóvenes (White et al., 2015). Susceptibilidad a diabetes tipo 2 también se ha asociado a esta variante (Du et al. 2020).
<b>MYLIP</b>	ENSR00000193925	rs182730226	A/T	<b>0.00694</b>	0.020004	0.006958	0.063492	-	0.004323	-
<b>APOC1</b>	ENSR00001221338	rs184802815	G/A	<b>0.03500</b>	0.010486	-	-	0.000756	0.090778	-
<b>PLTP</b>	ENSR00000137716	rs2294213	C/G	<b>0.06900</b>	0.111675	0.094433	0.317460	0.086989	0.060519	Colesterol aportado por HDLs en sangre asociado a prevalencia de alelo G (Engler et al. 2008).
<b>MTTP</b>	ENSR00000731288	rs41275707	G/A	<b>0.00694</b>	0.002127	0.005964	-	-	-	-
<b>SCARB1</b>	ENSR00000059010	rs41276688	C/T	<b>0.11100</b>	0.042531	0.047714	-	0.037821	0.051873	-
<b>APOH</b>	ENSR00001012430	rs8178822	G/T	<b>0.04200</b>	0.06425	0.069583	0.051587	0.088502	0.040346	Se ha confirmado que esta variante impacta la actividad del promotor de APOH (Qiu et al., 2020).

**Cuadro 8.** Variantes previamente reportadas en costarricenses y latinoamericanos en genes involucrados en el metabolismo de lípidos y sus promotores.

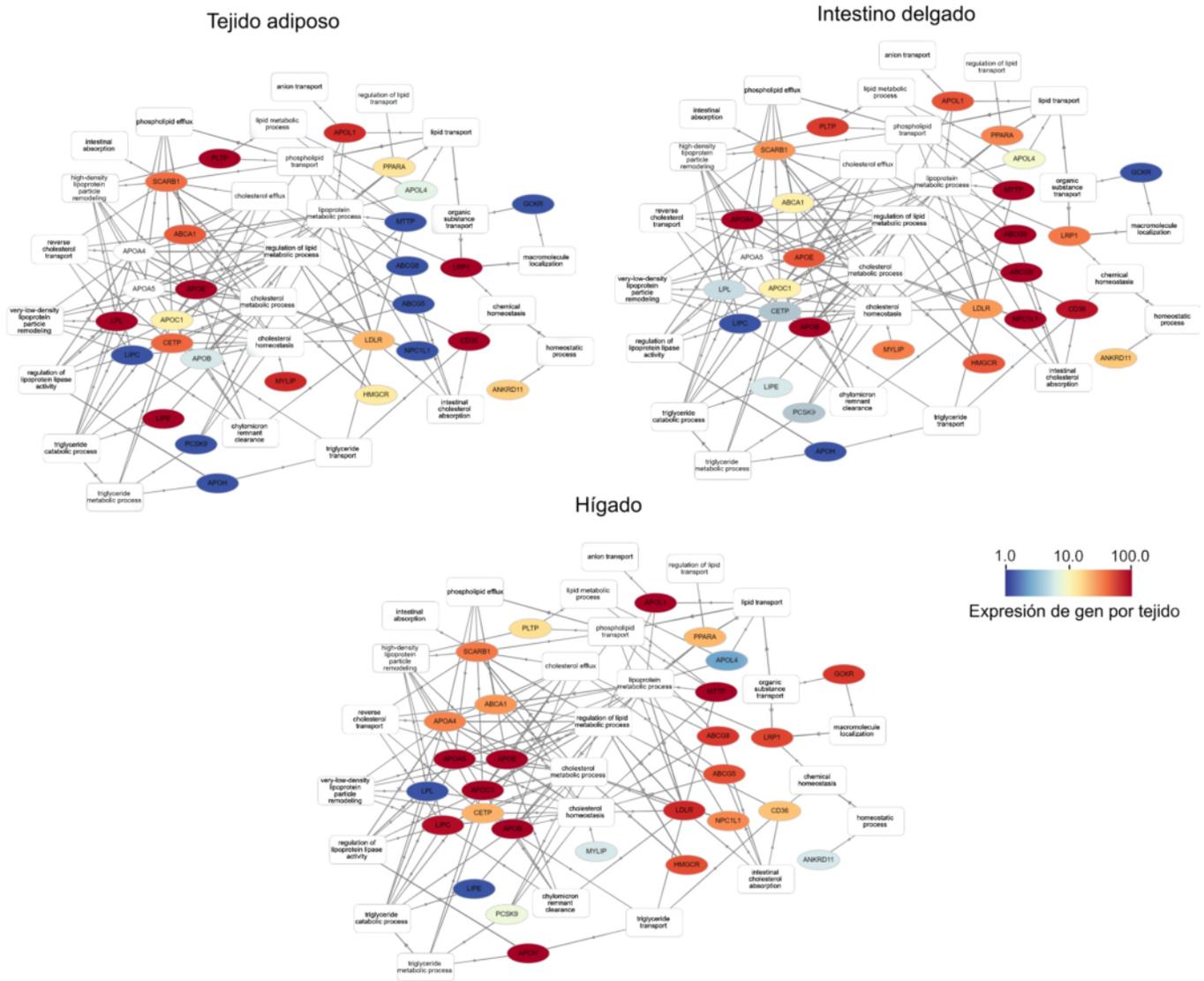
CR: Cohorte nombrada CR-WGS, genomas costarricenses evaluados en el presente estudio. Grupos de 1KGP - ALL: todos los sujetos, EAS: Asia del Este, EUR: Europa, AFR: África, AMR: América Latina. \*Frecuencia alélica de significativamente distinta ( $p < 0.05$ ) a la observada en costarricenses estudiados.

Coordenadas tGRCh38/hg38	Elemento genómico	Variante (dbSNP o HGMD)	Alelos (REF/ALT)	Frecuencia alélica de alelo alternativo						Asociación con metabolismo de lípidos en poblaciones latinoamericanas
				CR	ALL	EUR	EAS	AFR	AMR	
chr1: 109274968	CELSR2	rs12740374	G/T	<b>0.19400</b>	0.19549	0.21272	<b>0.04266*</b>	0.24735	0.20461	Estudio GWAS muestra una asociación entre el alelo menos abundante en EUR (T) con los niveles de LDLs y colesterol en latinoamericanos (Andaleon et al., 2019).
chr1: 55030366	Promotor PCSK9	rs11206510	T/C	<b>0.12500</b>	0.10184	0.17197	<b>0.04464*</b>	0.13464	0.11671	Ingesta de triglicéridos polinsaturados en costarricenses portadores de alelo C asociado a un menor riesgo de infarto del miocardio (Yu et al. 2017).
chr2: 43847292	ABCG8	rs4245791	C/T	<b>0.78500</b>	0.84105	0.68986	<b>0.99603*</b>	<b>0.89334*</b>	0.80259	Estudio GWAS muestra una asociación entre el alelo menos abundante en EUR (C) con los niveles de LDLs en latinoamericanos (Andaleon et al., 2019).
chr5: 75356084	HMGCR	rs6882842	G/A	<b>0.45100</b>	<b>0.61601*</b>	0.40855	0.50992	<b>0.94251*</b>	0.48559	Estudio GWAS muestra una asociación entre el alelo menos abundante en EUR (A) con los niveles de LDLs en latinoamericanos (Andaleon et al., 2019).
chr8: 19951902	LPL	rs1231383321	C/A	-	-	-	-	-	-	Alelo alternativo encontrado en grupo de costarricenses con hiperlipidemia severa (González Cordero 2018).
chr8: 19954222	LPL	rs118204057	G/A	-	0.00020	-	-	-	0.00144	
chr8: 19954233	LPL	LPL(NM_000237.3): c.655C>T	C/T	-	-	-	-	-	-	
chr8: 19954258	LPL	LPL(NM_000237.3): c.680T>G	T/G	-	-	-	-	-	-	
chr8: 19956018	LPL	rs268	A/G	<b>0.02100</b>	0.00519	0.01392	-	<b>0.00076*</b>	0.01153	
chr8: 19960925	LPL	rs316	C/A	<b>0.13900</b>	0.15256	0.12028	0.11210	0.23676	0.14553	

Coordenadas tGRCh38/hg38	Elemento genómico	Variante (dbSNP o HGMD)	Alelos (REF/ALT)	Frecuencia alélica de alelo alternativo						Asociación con metabolismo de lípidos en poblaciones latinoamericanas
				CR	ALL	EUR	EAS	AFR	AMR	
chr8: 19960925	LPL	rs328	C/G	<b>0.13900</b>	0.09245	0.13022	0.12202	0.06127	0.06340	Alelo alternativo asociado en costarricenses a un menor riesgo de infarto del miocardio (Yang et al. 2004).
chr8: 19960925	LPL	LPL(NM_000237.3): c.1420_1421insG	/G	-	-	-	-	-	-	Alelo alternativo encontrado en grupo de costarricenses con hiperlipidemia severa (González Cordero 2018).
chr8: 19948215	LPL	LPL(NM_000237.3): c.123_124insA	*/A	-	-	-	-	-	-	
chr8: 19960925	LPL	LPL(NM_000237.3): c.1163_1164insA	*/A	-	-	-	-	-	-	
chr8: 19960926	LPL	LPL(NM_000237.3): c.1164_1165insA	*/A	-	-	-	-	-	-	
chr8: 19961039	LPL	LPL(NM_000237.3): c.1277_1278insA	*/A	-	-	-	-	-	-	
chr8: 19961066	LPL	LPL(NM_000237.3): c.1304_1305insG	*/G	-	-	-	-	-	-	
chr8: 19961078	LPL	LPL(NM_000237.3): c.1317G>C	G/C	-	-	-	-	-	-	
chr9: 104858554	ABCA1	rs9282541	G/A	<b>0.04200</b>	<b>0.00599*</b>	-	-	<b>0.00076*</b>	0.04179	
chr11: 61829740	FADS2	rs174570	C/T	<b>0.31300</b>	0.22823	0.16004	0.55556	0.00908	0.48703	Estudio GWAS muestra una asociación entre el alelo menos abundante en EUR (T) con los niveles de colesterol en latinoamericanos (Andaleon et al., 2019).
chr15: 58431740	LIPC	rs2070895	G/A	<b>0.50700</b>	0.39437	<b>0.21073*</b>	0.41667	0.56430	0.42507	Estudio GWAS muestra una asociación entre el alelo menos abundante en EUR (A) con los niveles de HDLs en latinoamericanos (Andaleon et al., 2019).

Coordenadas tGRCh38/hg38	Elemento genómico	Variante (dbSNP o HGMD)	Alelos (REF/ALT)	Frecuencia alélica de alelo alternativo						Asociación con metabolismo de lípidos en poblaciones latinoamericanas
				CR	ALL	EUR	EAS	AFR	AMR	
chr19: 11091630	LDLR	rs6511720	G/T	<b>0.06900</b>	0.09165	0.11034	<b>0.01190*</b>	0.14675	0.09366	Estudio GWAS muestra una asociación entre el alelo menos abundante en EUR (T) con los niveles de LDLs y colesterol en latinoamericanos (Andaleon et al., 2019). Alelo alternativo encontrado en grupo de costarricenses con hiperlipidemia severa (Gutiérrez Ávila 2019).
chr19: 44945909	Promotor de APOC2	rs12721063	G/A	-	0.00240	0.00895	-	0.00151	0.00144	
chr19: 44945942	Promotor de APOC2	rs2288912	C/G	<b>0.64600</b>	0.66014	<b>0.48608*</b>	0.67460	<b>0.76324*</b>	0.72911	
chr19: 44946012	Promotor de APOC2	rs919639084	C/T	-	-	-	-	-	-	
chr19: 44946027	Promotor de APOC2	rs2288911	T/G	<b>0.66000</b>	0.66573	<b>0.49301*</b>	0.67361	<b>0.77307*</b>	0.73487	

De acuerdo con el análisis de enriquecimiento de términos GO efectuado, alteraciones en los genes portadores de las variantes de riesgo identificadas pueden generar cambios importantes en procesos biológicos efectuados en el hígado, intestino delgado y tejido adiposo vinculados con el metabolismo de lípidos, afectando tanto el anabolismo como el catabolismo de estos (Figura 11).



**Figura 11.** Expresión de genes afectados por variantes de interés en tejidos protagónicos en el metabolismo de lípidos y su vínculo con procesos biológicos.

Los óvalos son genes y cuadrados simbolizan procesos biológicos obtenidos del análisis de enriquecimiento de términos GO. Las flechas indican que un gen este asociado a determinado proceso biológico. La posición de los elementos es constante en las tres redes. La coloración de los óvalos representa los niveles de expresión del gen en el tejido mencionado. La expresión de gen por tejido fue extraída de *Human Protein Atlas*.

## Discusión

### *Evaluación de flujo de análisis bioinformático*

El flujo de trabajo bioinformático utilizado para realizar el llamado de variantes sobre las cohortes AMR-80 y CR-WGS generó métricas que se encuentran dentro de los valores esperados. Sin embargo, estas estadísticas se desviaron de los valores estándar (Wang et al., 2015) y también se encontraron diferencias entre ambas cohortes. Se identificaron aproximadamente 18461 SNVs por exoma en AMR-80 y 22468 en CR-WGS. De acuerdo con la literatura el conteo esperado de SNVs por exoma oscila entre 15000 y 20000, siendo el factor determinante de esta variación las coordenadas utilizadas para definir el exoma (S. B. Ng et al., 2009; Stitzel et al., 2011). No se encontró reportado en la literatura un valor esperado específico para el panel *Twist Human Comprehensive Exome Panel* utilizado para definir las regiones exónicas. Así mismo, si bien AMR-80 y CR-WGS presentaron valores de  $T_i/T_v$  y HET/ non-ref HOM cercanos a los reportados para exomas de latinoamericanos (J. Wang et al., 2015), AMR-80 tendió a presentar valores de  $T_i/T_v$  más altos que los observados en CR-WGS (Figura 3).

Las diferencias entre AMR-80 y CR-WGS mencionadas fueron probablemente ocasionadas en cierta medida debido a que el proceso de llamado de variantes aplicado a AMR-80 únicamente evaluó las regiones exónicas mientras que para CR-WGS, se evaluaron además algunas regiones no codificantes, incluidas promotores e intrones (Cuadro 3). Debido a que se evaluaron más regiones genómicas por sujeto en CR-WGS se pudo detectar una mayor cantidad de variantes por sujeto. Esto pudo haber sido determinante a la hora de excluir el 2.5% de las variantes de acuerdo con su puntaje de calidad VQSLOD. Esto implica que en AMR-80 se descartaron el 2.5% de las variantes exónicas con menor puntaje, mientras que en CR-WGS además incluyó secuencias no codificantes. Los valores menores a 3 en  $T_i/T_v$  en exomas suelen reflejar una menor calidad en

el llamado de variantes (J. Wang et al., 2015), por lo que el valor de 3.0 para CR-WGS puede considerarse que cumple con los estándares de calidad reportados para el llamado de variantes.

### *Concordancia con ancestría de costarricenses del Valle Central*

Los resultados obtenidos a partir de los análisis de ancestría muestran que los genomas del Valle Central de Costa Rica evaluados tienen una mezcla genética concordante con lo esperado dentro del contexto latinoamericano, así como con lo reportado en estudios de ancestría efectuados en costarricenses del Valle Central elegidos aleatoriamente. También existe una alta concordancia entre las frecuencias alélicas reportadas para CR-WGS con respecto a la muestra de costarricenses del Valle Central sin enfermedad diagnosticada estudiada en el *Costa Rica Heart Study*. Todo esto sugiere que las frecuencias obtenidas a partir de CR-WGS son representativas de la población general del Valle Central de Costa Rica.

Al comparar la ancestría de los costarricenses con datos internacionales, los genomas de los costarricenses del Valle Central analizados (CR-WGS) presentaron un perfil de ancestría semejante al de los grupos latinoamericanos reportados en 1KGP. De los cuatro grupos hispanos incluidos en 1KGP, los análisis efectuados indican que el grupo costarricense se asemeja más a los colombianos y puertorriqueños que a los mexicanos y los peruanos. Esto se debe a que la relación entre la ancestría europea frente a la aportada por Asia del Este es más alta en los costarricenses, colombianos y puertorriqueños que la observada en los mexicanos y peruanos (Adhikari et al., 2017; Wang et al., 2019).

Al contrastar la mezcla genética de CR-WGS con la de costarricenses genotipados por Campos-Sánchez et al. (2013) pertenecientes a la Zona Norte, Zona Sur, Región Caribe y Valle Central, se encontró que los sujetos más semejantes a los de CR-WGS fueron los del Valle Central, lo que concuerda con el origen de las muestras de CR-WGS. Ruiz-Narváez et al., (2010) determinaron la ancestría a partir de 1998 sujetos del Valle Central sin patología reportada elegidos

aleatoriamente; este estudio mostró una ancestría 57.8% europea, 38.3% amerindia y 3.8% de África del Este, un balance semejante al observado para CR-WGS en este estudio. Dicho esto, cabe mencionar que el análisis de mezcla genética realizado a los cinco grupos costarricenses tendió a mostrar balances de ancestría europea menores que los obtenidos para cuatro de estos grupos en los análisis reportados por Campos-Sánchez et al. (2013). Esto se puede deber a que en el presente estudio se utilizó la población de Asia del Este (EAS) reportada en 1KGP como grupo ancestral en vez de un grupo de amerindios como en el estudio de Campos-Sánchez et al. (2013). Si bien EAS ha sido utilizado en estudios previos de ancestría como grupo análogo a los nativos americanos debido al origen histórico de los mismos y debido a que EAS es un grupo amplio y estandarizado (Wang et al., 2019), se recomienda en estudios futuros utilizar información genómica estandarizada de nativos americanos para esta clase de análisis.

#### *Variantes de interés farmacogenómico*

De acuerdo con la anotación funcional extraída de ClinVar y *GWAS Catalogue*, se ha reportado que al menos ocho de las variantes identificadas tienen un impacto ya sea sobre la eficacia, seguridad o metabolismo de agentes terapéuticos. Se ha observado en trabajos previos que las variantes rs1532624 en CETP, rs17244841 en HMGCR, y rs429358, rs7412 junto con rs769450 en APOE, pueden alterar la eficacia de fármacos tipo estatinas tales como lovastatina, atorvastatina o pravastatina para reducir los niveles de colesterol en sangre (Ciuculete et al., 2017; De Keyser et al., 2011; Guan et al., 2019; Mega et al., 2009; Thompson et al., 2009). Las frecuencias alélicas obtenidas para estas variantes en el grupo de costarricenses estudiados fueron semejantes a las de los europeos y latinoamericanos reportados en 1KGP, únicamente se encontraron diferencias significativas para rs1532624 con el grupo AFR y para rs679899 con AFR y EAS.

### *Variantes categorizadas como de riesgo*

Se ha observado que alteraciones en los niveles de expresión o en el funcionamiento de los genes involucrados en el metabolismo de lípidos evaluados en este estudio pueden ocasionar desequilibrios en el perfil lipídico de las personas y dar paso al desarrollo de dislipidemias. Se encontraron variantes en 20 de los 69 genes evaluados que fueron categorizadas como de riesgo por más de dos herramientas bioinformáticas (Cuadro 6 y Cuadro 7). De estas variantes, cinco están asociadas a genes de la familia de las apolipoproteínas, ocho están presentes en genes involucrados la movilización de lípidos, seis variantes en genes que tienen un rol regulatorio, 8 en genes estrechamente relacionados con las dinámicas de reconocimiento de lipoproteínas y la detección de lípidos y 2 en el gen de la lipasa E (*LIPE*).

Los genes pertenecientes a la familia de las apolipoproteínas afectados por variantes de riesgo mencionadas previamente son, *APOA5*, *APOC1*, *APOE* y *APOH*. De acuerdo con Su & Peng (2020), *APOA5* y *APOE* participan en el ensamblaje de las VLDLs. En un estudio realizado por Y. Zhou et al., (2018) se encontró que variantes en *APOA* suelen impactar más los niveles de triglicéridos en plasma que los de colesterol. Varios estudios han vinculado la presencia del alelo C en el SNV rs3135506 con niveles elevados de triglicéridos en plasma (Li et al., 2014; Ruiz-Narváez et al., 2005). Surendran et al. (2012) encontraron una frecuencia alélica del 21% en un grupo de pacientes con hipertrigliceridemia severa mientras que el grupo control presentó una frecuencia del 9%. En el grupo costarricense esta variante alcanzó una frecuencia alélica del 9.7% y no mostró diferencias significativas con los otros grupos de 1KGP. Por otro lado, varios estudios han asociado la presencia del alelo T de la variante rs7412 perteneciente a *APOE* con niveles elevados de colesterol en sangre aportado principalmente por LDLs y con índices de masa corporal elevados (Tejedor et al., 2014; Thompson et al., 2009). A pesar de que la frecuencia de esta variante en los costarricenses es de 8.30% mientras que la de los latinoamericanos registrados en 1KGP es de 4.75%, no se encontraron diferencias estadísticamente significativas entre ellas, evaluar esto en

otras partes del país o incrementar el tamaño de la muestra puede ayudar a esclarecer si esta tendencia se disipa o torna más robusta.

Alteraciones en las apolipoproteínas APOC1 y APOH también han sido asociadas con cambios en el perfil lipídico. APOC1 inhibe la unión mediada por APOE de lipoproteínas con sus receptores, inhibe la actividad enzimática de la lipasa hepática (LIPC) y promueve la actividad de la LCAT, enzima encargada de esterificar colesterol, modificación que posibilita su movilización de células a lipoproteínas (Paththinige et al., 2017). Pocos estudios han asociado la presencia de variantes en el gen con alteraciones en el perfil lipídico, no obstante, sí hay evidencia de esta asociación (R. Zhang et al., 2018). No se encontró literatura que explore el efecto de la variante rs184802815 identificada en la región promotora de este gen. Por otro lado, si bien se conoce poco sobre el papel molecular de APOH en el metabolismo de lípidos, se ha observado en varias poblaciones que la presencia de algunas variantes asociadas al funcionamiento de esta apolipoproteína afectan los niveles de colesterol LDL (Willer et al., 2013). Las dos variantes de riesgo identificadas en este gen han sido asociadas con efectos fenotípicos. Un estudio GWAS asoció el alelo T de la variante rs8178822 ubicado en la región promotora con los niveles de APOH en plasma; mientras que el alelo C de la variante rs1801689 ha sido vinculado con modificaciones en los niveles de LDL en sangre, esta variación altera la afinidad de APOH con fosfolípidos (Mather et al., 2016).

Se identificaron variantes de riesgo en seis genes involucrados en el transporte de lípidos. De las ocho variantes en este grupo de genes, cinco se encontraron en los genes de la familia de transportadores ABC estrechamente asociados en el metabolismo de lípidos, *ABCA1*, *ABCG5* y *ABCG8*. *ABCA1* participa en la formación de HDLs al trasladar colesterol y fosfolípidos del interior de la célula hacia HDLs nacientes. Por otro lado *ABCG5* forma un heterodímero con *ABCG8* que median en múltiples niveles la absorción y excreción de esteroides (Feingold & Grunfeld, 2018; Kerr et al., 2011). De las variantes de riesgo identificadas, únicamente rs1800977 en *ABCA1* y rs11887534 en *ABCG8* han sido asociadas a cambios en los niveles de HDLs en sangre

(Porchay et al., 2006; Sařacka et al., 2021). Por otro lado, la proteina transportadora de fosfolipidos o PLTP fundamental en el mecanismo molecular para transferir fosfolipidos de LDLs y VLDLs hacia HDLs alberga la variante de riesgo identificada rs2294213. Se ha visto que el alelo G de esta variante está más presente en sujetos con hiperalfalipoproteinemia o altos niveles de colesterol HDL en sangre y se observa el efecto inverso en grupos con mayor abundancia del alelo C (Chen et al., 2009; Engler, 2009; Engler et al., 2008; Paththinige et al., 2017). Las dos variantes de riesgo restantes identificadas en genes involucrados en el transporte de lípidos se encontraron en MTTP y NPC1L1, no obstante, no se encontró ningún vínculo entre estas variantes y algún fenotipo de interés.

Se encontraron variantes de riesgo en cinco genes con un papel regulatorio o señalizador en el metabolismo de lípidos. No se encontraron investigaciones previas que asocien la presencia de las variantes de riesgo encontradas en CELSR2, CREB3L3, GCKR, MYLIP y PPARA con alteraciones en el perfil lipídico o riesgo a padecer dislipidemias.

Se encontraron variantes de riesgo putativo en los genes *CD36*, *LDLR*, *LRP1* y *SCARB1*, involucrados en la detección de lípidos y lipoproteínas. Sin embargo, ninguna de dichas variantes ha sido vinculada con alteraciones en el perfil lipídico. Del mismo modo, se encontraron dos variantes de riesgo en la lipasa E sensible a hormonas, pero tampoco se ha reportado vínculo entre estas variantes y alteraciones en el perfil lipídico.

#### *Variantes previamente reportadas en la región Latinoamericana*

Las variantes rs2288911 y rs2288912 en el promotor de APOC2, activadora de la lipoproteinlipasa (LPL), fueron las únicas variantes reportadas por Gutiérrez Ávila (2019) encontradas en los genomas costarricenses evaluados. La variante rs2288912 ha sido asociada con los niveles de triglicéridos y colesterol en sangre por varios estudios (Pirim et al., 2019; X. Zhang et al., 2014), mientras que para la variante rs2288911 no hay evidencia robusta que la asocie con alteraciones en el perfil lipídico. No obstante, en un reporte de caso clínico, un paciente con

hipertrigliceridemia severa era heterocigoto para una variante en LPL y además contaba con las variantes rs2288911 y rs2288912 (Nocoń-Bohusz et al., 2015). Cabe mencionar que estas dos variantes se encuentran a 15pb de distancia y de acuerdo con Schuster et al. (2011), se encuentran asociadas por desequilibrio de ligamiento. Evidencia indirecta de esto es que los alelos alternativos en ambos sitios polimórficos tienen frecuencias alélicas altamente semejantes, siendo estas 0.660 para rs2288911 y 0.646 para rs2288912 en la población costarricense, el mismo fenómeno se observa en las poblaciones de 1KGP reportadas.

De las 15 variantes descritas por González Cordero (2018) en *LPL*, únicamente las variantes rs268 y rs316 fueron encontradas en los genomas costarricenses analizados. De acuerdo con un meta-análisis, el alelo G en la variante rs268 está asociado a menores niveles de colesterol HDL en plasma (Boes et al., 2009). Rs316 es una variante intrónica la cual, de acuerdo con Pirim et al. (2014), es posible que se encuentre junto a un sitio regulador. La presencia del alelo A en dicha variante ha sido asociada con un incremento de colesterol HDL en repetidas ocasiones (Pirim et al., 2014, 2015; Schuster et al., 2011). Por otro lado, la variante asociada previamente en una publicación del *Costa Rica Heart Study* con una reducción en el riesgo ante infarto de miocardio en costarricenses también fue identificada en la cohorte estudiada (Yang et al., 2004). El alelo G en el sitio polimórfico rs328 suprime la codificación de los últimos dos aminoácidos de LPL lo cual incrementa su actividad lipasa, notablemente esto está asociado a niveles bajos de triglicéridos en plasma e incrementos en colesterol HDL en sujetos sanos, pero en sujetos con obesidad este alelo en vez se asocia con niveles elevados de triglicéridos en plasma (Palacio Rojas et al., 2017). No se encontraron diferencias estadísticamente significativas para las frecuencias alélicas encontradas en estas tres variantes entre costarricenses y los grupos 1KGP, sin embargo, sí se observaron diferencias aritméticas entre los costarricenses y AMR principalmente en rs268 y rs328 que podrían ser exploradas en un futuro.

Algunas de las variantes reportadas en el estudio de Andaleon et al., (2019), enfocado en identificar variantes asociadas a cambios en el perfil lipídico latinoamericanos habitantes de

Estados Unidos, fueron encontradas en la cohorte costarricense analizada. De acuerdo con la investigación, las variantes rs12740374 en CELSR2, rs4245791 en ABCG8, rs6882842 en HMGCR y rs6511720 en LDLR influyen en los niveles de colesterol LDL en hispanos. Por otro lado, rs12740374, rs174570 localizadas en FADS2 y rs6511720 en LDLR, afectan el colesterol total y rs2070895 en LIPC este asociado con los niveles de HDLs en este grupo. La variante rs9282541 de ABCA1, considerada como variante privada en nativos americanos y sus descendientes por otro lado (Du et al., 2020; Villarreal-Molina et al., 2012), se encontró también en el grupo costarricense con una frecuencia alélica semejante a la observada en los latinoamericanos reportados en 1KGP. Villarreal-Molina et al. (2012) encontraron en sujetos mexicanos que la presencia de esta variante estaba asociada a niveles más bajos de colesterol total y colesterol HDL en plasma y asociaron que el efecto observado depende del sexo del sujeto, por esta línea se argumenta que este último hecho demuestra que el efecto de esta variante interactúa con otros factores.

## Conclusiones y recomendaciones

De los análisis de calidad efectuados al flujo de trabajo implementado para llamado de variantes y el análisis de ancestría realizado se concluyen dos puntos. En primera instancia se determinó que los genotipos obtenidos del procesamiento bioinformático son fidedignos. En segunda instancia, se concluyó que el grupo de genomas es consistente con el perfil genético de los latinoamericanos y los costarricenses del Valle Central. Además de ello, las frecuencias alélicas obtenidas en la población de estudio son altamente semejantes a las reportadas al subgrupo de sujetos seleccionados del Valle Central del estudio *Human Heart Study*, señalan que las frecuencias obtenidas pueden ser extrapolables al Valle Central. Dicho esto, genomas provenientes de bases de datos conformadas por sujetos seleccionados de manera no aleatoria pueden ser reanalizados con el fin de estudiar frecuencias alélicas poblacionales luego de seleccionar únicamente los sujetos no emparentados y de aplicar medidas de control de calidad como las implementadas en este estudio.

El estudio identificó en genomas de personas del Valle Central de Costa Rica alrededor de 17500 variantes localizadas en 69 genes y promotores involucrados en el metabolismo de lípidos. Entre estas, más de 30 variantes con potencial de afectar el funcionamiento de estos genes, y algunas han sido vinculadas directamente con el desarrollo de hiperlipidemias. Otras variantes podrían afectar el desempeño de proteínas involucradas en el metabolismo de lípidos según análisis bioinformáticos, aunque no han sido asociadas directamente con el desarrollo de tales condiciones en la literatura. Por otra parte, se encontraron 8 variantes con relevancia farmacogenómica, varias de estas pueden modular la respuesta del sujeto ante la aplicación de fármacos tipo estatinas, terapias comúnmente utilizadas para tratar casos de hiperlipidemia severa.

Algunas variantes previamente descritas en costarricenses localizadas en los genes y promotores de interés fueron encontradas en el grupo estudiado. No obstante, ciertas variantes reportadas por González Cordero (2018) en el gen *LPL* no fueron encontradas. Esto puede deberse

a que González Cordero (2018) estudió sujetos con hipertrigliceridemia severa, debido a esto las variantes al ser patogénicas son raras en la población general. Además, se encontraron variantes las cuales de acuerdo con Andaleon et al. (2019), tienen un impacto significativo en los niveles de lípidos en sangre de personas latinoamericanas que viven en Estados Unidos de América.

Es importante indagar si las variantes reportadas previamente en latinoamericanos pueden producir alteraciones en los niveles de lípidos en costarricenses. Por un lado, las investigaciones de González Cordero (2018) y Gutiérrez Ávila (2019) reportan la existencia de ciertas variantes en personas con hiperlipidemia severa pero no establecen un vínculo causal o cuantitativo entre la presencia de estas variantes en los sujetos y los niveles de triglicéridos en sangre. Por otro lado, si bien la investigación de Andaleon et al. (2019) logra vincular la presencia de algunas variantes con un impacto cuantitativo sobre los niveles de lípidos en plasma para latinoamericanos, es importante mencionar que las personas incluidas en dicho estudio residen en Estados Unidos, lo cual puede significar que estos estén expuestos a un estilo de vida y condiciones ambientales distintas a las que estarían sujetas si vivieran en su país de origen. Se ha visto que el ambiente por sí solo puede afectar un 21% en la variación de los niveles de colesterol total en plasma y un 29% en los niveles de triglicéridos en plasma y que aproximadamente un 6% de la variación se atribuye a la interacción entre ambiente y genética (Elder et al., 2009). Considerando esto, en un futuro es importante realizar estudios que capturen datos ambientales, genotípicos y fenotípicos de costarricenses habitantes de Costa Rica para entender con mayor claridad los factores e interacciones que participan en la incidencia de hiperlipidemias.

Para responder con mayor amplitud y precisión a los objetivos planteados en estudio se hacen dos recomendaciones. En primer lugar, se sugiere utilizar como grupo ancestral genomas de nativos americanos en vez del grupo de Asia del Este reportado en 1KGP para establecer con mayor claridad la ancestría del grupo estudiado. En segundo lugar, con la finalidad de poder ampliar la cantidad de variantes identificadas y estimar frecuencias alélicas con mayor precisión, se sugiere incluir genomas costarricenses provenientes de otras investigaciones.

La información generada a partir de este estudio ayudará a orientar y contextualizar futuros estudios sobre dislipidemias en Costa Rica y la región. Debido a que la investigación compila la frecuencia alélica de una serie de variantes identificadas en costarricenses con el potencial de impactar el metabolismo de lípidos, el diseño experimental de futuras investigaciones a nivel nacional se podrá realizar de una manera más informada. Por un lado, las frecuencias alélicas obtenidas para las diferentes variantes pueden ayudar a obtener mejores estimados del tamaño de muestra requerido para detectar ciertas variantes en la población general. Por otro lado, debido a que el estudio identificó alrededor de 30 variantes de interés a partir de las aproximadamente 17500 identificadas en los 69 genes, futuras investigaciones podrían focalizarse en estas variantes. Limitar el análisis genético a ciertas variantes puntuales puede generar una reducción en los costos de ejecución de los proyectos y abre la posibilidad de utilizar metodologías de genotipado tradicionales tales como secuenciación Sanger, RFLPs o qPCR con sondas.

Para entender con mayor claridad las dinámicas que participan en la incidencia de hiperlipidemias en costarricenses y latinoamericanos en general, grupos subrepresentados en el estado del arte, se recomienda realizar estudios de cohorte los cuales recolecten información multidimensional de los participantes. Considerando que la hiperlipidemia es una enfermedad compleja resulta vital compilar información sobre el ambiente en el que se desenvuelven los individuos, su estilo de vida y su genética para vincularla con rasgos fenotípicos de interés como el perfil lipídico. Debido a que la Caja Costarricense del Seguro Social (C.C.S.S.) alberga una amplia cantidad de información médica de costarricenses, el acceso a estos datos podría hacer posible la realización de un estudio de cohorte retrospectivo en el cual se incluyan costarricenses de todas las regiones del país.

El análisis genético podría limitarse a un grupo de variantes de interés o alternativamente, se podría realizar secuenciación del genoma completo, el exoma o un panel de genes involucrados en el metabolismo de lípidos como el panel *LipidSeq* descrito por Johansen et al. (2014). Una estrategia similar ya ha sido utilizada en América Latina en un estudio realizado en Argentina por

Corral et al., (2018), en dicha investigación la secuenciación se limitó a genes vinculados al metabolismo de lípidos.

La obtención de datos genotípicos pareados con datos fenotípicos haría posible efectuar estudios GWAS, desarrollar índices de riesgo poligénicos específicos para la población costarricense, validar si índices desarrollados para otras poblaciones son extrapolables para la población costarricense o incluso identificar maneras más efectivas para tratar estas condiciones a nivel nacional. Adicionalmente, se podría entender con mayor claridad si existe variación en el rol que tiene la genética sobre el desarrollo de estas enfermedades entre poblaciones latinoamericanas, pregunta que ha sido planteada y explorada anteriormente por Andaleon et al. (2019) la cual podría esclarecer si aprendizajes en otras poblaciones de la región son extrapolables para la costarricense y en qué medida.

## Referencias

- Adhikari, K., Chacón-Duque, J. C., Mendoza-Revilla, J., Fuentes-Guajardo, M., & Ruiz-Linares, A. (2017). The Genetic Diversity of the Americas. *Annual Review of Genomics and Human Genetics*, 18, 277–296. <https://doi.org/10.1146/annurev-genom-083115-022331>
- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. In *Current Protocols in Human Genetics* (Issue SUPPL.76). <https://doi.org/10.1002/0471142905.hg0720s76>
- Ahangari, N., Ghayour Mobarhan, M., Sahebkar, A., & Pasdar, A. (2018). Molecular aspects of hypercholesterolemia treatment: current perspectives and hopes. *Annals of Medicine*, 50(4), 303–311. <https://doi.org/10.1080/07853890.2018.1457795>
- Airewele, G. E., Sigurdson, A. J., Wiley, K. J., Frieden, B. E., Calderera, L. W., Riccardi, V. M., Lewis, R. A., Chintagumpala, M. M., Ater, J. L., Plon, S. E., & Bondy, M. L. (2001). Gene-diet interactions and plasma lipoproteins: Role of apolipoprotein E and habitual saturated fat intake. *Genetic Epidemiology*, 20(1), 117–128. [https://doi.org/10.1002/1098-2272\(200101\)20:1<117::AID-GEPI10>3.0.CO;2-C](https://doi.org/10.1002/1098-2272(200101)20:1<117::AID-GEPI10>3.0.CO;2-C)
- Alirezaie, N., Kernohan, K. D., Hartley, T., Majewski, J., & Hocking, T. D. (2018). ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. *American Journal of Human Genetics*, 103(4), 474–483. <https://doi.org/10.1016/j.ajhg.2018.08.005>
- Altshuler, D. M., Gibbs, R. A., Peltonen, L., Schaffner, S. F., Yu, F., Dermitzakis, E., Bonnen, P. E., De Bakker, P. I. W., Deloukas, P., Gabriel, S. B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Aarno Palotie, Parkin, M., Whittaker, P., Chang, K., Hawes, A., ... McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311), 52–58. <https://doi.org/10.1038/nature09298>
- Andaleon, A., Mogil, L. S., & Wheeler, H. E. (2019). Genetically regulated gene expression underlies lipid traits in Hispanic cohorts. *PLoS ONE*, 14(8), 1–21. <https://doi.org/10.1371/journal.pone.0220827>
- Aslibekyan, S., Jensen, M. K., Campos, H., Linkletter, C. D., Loucks, E. B., Ordovas, J. M., Deka, R., Rimm, E. B., & Baylin, A. (2012). Fatty acid desaturase gene variants, cardiovascular risk factors, and myocardial infarction in the Costa Rica Study. *Frontiers in Genetics*, 3(MAY), 1–11. <https://doi.org/10.3389/fgene.2012.00072>
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurler, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., ... Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>

- Bardou, P., Mariette, J., Escudié, F., Djemiel, C., & Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*, *15*(293), 1–7. <http://www.biomedcentral.com/1471-2105/15/293>
- Bartłomiejczyk, M. A., Penson, P., & Banach, M. (2019). Worldwide Dyslipidemia Guidelines. *Current Cardiovascular Risk Reports*, *13*(2), 1–7. <https://doi.org/10.1007/s12170-019-0597-x>
- Berberich, A. J., & Hegele, R. A. (2019). The role of genetic testing in dyslipidaemia. *Pathology*, *51*(2), 184–192. <https://doi.org/10.1016/j.pathol.2018.10.014>
- Berglund, L., Brunzell, J. D., Goldberg, A. C., Goldberg, I. J., Sacks, F., Murad, M. H., & Stalenhoef, A. F. H. (2012). Evaluation and treatment of hypertriglyceridemia: An endocrine society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism*, *97*(9), 2969–2989. <https://doi.org/10.1210/jc.2011-3213>
- Bhatnagar, D., Soran, H., & Durrington, P. N. (2008). Hypercholesterolaemia and its management. *Bmj*, *337*(7668), 503–508. <https://doi.org/10.1136/bmj.a993>
- Bianconi, V., Banach, M., & Pirro, M. (2020). Why patients with familial hypercholesterolemia are at high cardiovascular risk? Beyond LDL-C levels. *Trends in Cardiovascular Medicine*, *xxxx*. <https://doi.org/10.1016/j.tcm.2020.03.004>
- Boes, E., Coassin, S., Kollerits, B., Heid, I. M., & Kronenberg, F. (2009). Genetic-epidemiological evidence on genes associated with HDL cholesterol levels: A systematic in-depth review. *Experimental Gerontology*, *44*(3), 136–160. <https://doi.org/10.1016/j.exger.2008.11.003>
- Brahm, A., & Hegele, R. A. (2013). Hypertriglyceridemia. *Nutrients*, *5*(3), 981–1001. <https://doi.org/10.3390/nu5030981>
- Brown, S., Ordovás, J. M., & Campos, H. (2003). Interaction between the APOC3 gene promoter polymorphisms, saturated fat intake and plasma lipoproteins. *Atherosclerosis*, *170*(2), 307–313. [https://doi.org/10.1016/S0021-9150\(03\)00293-4](https://doi.org/10.1016/S0021-9150(03)00293-4)
- Bruikman, C. S., Hovingh, G. K., & Kastelein, J. J. P. (2017). Molecular basis of familial hypercholesterolemia. *Current Opinion in Cardiology*, *32*(3), 262–266. <https://doi.org/10.1097/HCO.0000000000000385>
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., ... Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005–D1012. <https://doi.org/10.1093/nar/gky1120>
- Campos-Sánchez, R., Raventós, H., & Barrantes, R. (2013). Ancestry informative markers clarify the regional

admixture variation in the Costa Rican population. *Human Biology*, 85(5), 721–740.  
<https://doi.org/10.3378/027.085.0505>

Carrasco-Ramiro, F., Peiró-Pastor, R., & Aguado, B. (2017). Human genomics projects and precision medicine. *Gene Therapy*, 24(9), 551–561. <https://doi.org/10.1038/gt.2017.77>

Caussy, C., Charrière, S., Meirhaeghe, A., Dallongeville, J., Lefai, E., Rome, S., Cuerq, C., Euthine, V., Delay, M., Marmontel, O., Di Filippo, M., Lagarde, M., Moulin, P., & Marçais, C. (2016). Multiple microRNA regulation of lipoprotein lipase gene abolished by 3'UTR polymorphisms in a triglyceride-lowering haplotype harboring p.Ser474Ter. *Atherosclerosis*, 246, 280–286.  
<https://doi.org/10.1016/j.atherosclerosis.2016.01.010>

Chavarria-Soley, G., Francis-Cartin, F., Jimenez-Gonzalez, F., Peralta, J. M., Blangero, J., Gur, R. E., Almas, L., Glahn, D., & Raventos, H. (2021). Identification of Genetic Risk Variants for Major Psychiatric Disorders in Costa Rican Families Using Wgs. *European Neuropsychopharmacology*, 51, e16–e17.  
<https://doi.org/10.1016/j.euroneuro.2021.07.042>

Chen, X., Sun, A., Mansoor, A., Zou, Y., Ge, J., Lazar, J. M., & Jiang, X. C. (2009). Plasma PLTP activity is inversely associated with HDL-C levels. *Nutrition and Metabolism*, 6, 1–6. <https://doi.org/10.1186/1743-7075-6-49>

Chyzyk, V., Kozmic, S., Brown, A. S., Hudgins, L. C., Starc, T. J., Davila, A. D., Blevins, T. C., Diffenderfer, M. R., He, L., Geller, A. S., Rush, C., Hegele, R. A., & Schaefer, E. J. (2019). Extreme hypertriglyceridemia: Genetic diversity, pancreatitis, pregnancy, and prevalence. *Journal of Clinical Lipidology*, 13(1), 89–99. <https://doi.org/10.1016/j.jacl.2018.09.007>

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>

Ciuculete, D. M., Bandstein, M., Benedict, C., Waeber, G., Vollenweider, P., Lind, L., Schiöth, H. B., & Mwinyi, J. (2017). A genetic risk score is significantly associated with statin therapy response in the elderly population. *Clinical Genetics*, 91(3), 379–385. <https://doi.org/10.1111/cge.12890>

Corral, P., Geller, A. S., Polisecki, E. Y., Lopez, G. I., Bañares, V. G., Cacciagu, L., Berg, G., Hegele, R. A., Schaefer, E. J., & Schreier, L. E. (2018). Unusual genetic variants associated with hypercholesterolemia in Argentina. *Atherosclerosis*, 277, 256–261. <https://doi.org/10.1016/j.atherosclerosis.2018.06.009>

Cox, R. A., & García-Palmieri, M. R. (1990). Cholesterol, Triglycerides, and Associated Lipoproteins. In *Clinical Methods: The History, Physical, and Laboratory Examinations*.  
<http://www.ncbi.nlm.nih.gov/pubmed/21250192>

- Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., ... Flicek, P. (2021). Ensembl 2022. *Nucleic Acids Research*, 1–8. <https://doi.org/10.1093/nar/gkab1049>
- De Keyser, C. E., Eijgelsheim, M., Hofman, A., Sijbrands, E. J. G., Maitland-Van Der Zee, A. H., Van Duijn, C. M., Uitterlinden, A. G., Witteman, J. C. M., & Ch Stricker, B. H. (2011). Single nucleotide polymorphisms in genes that are associated with a modified response to statin therapy: The Rotterdam Study. *Pharmacogenomics Journal*, 11(1), 72–80. <https://doi.org/10.1038/tpj.2010.11>
- de Pretis, N., Amodio, A., & Frulloni, L. (2018). Hypertriglyceridemic pancreatitis: Epidemiology, pathophysiology and clinical management. *United European Gastroenterology Journal*, 6(5), 649–655. <https://doi.org/10.1177/2050640618755002>
- Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–501. <https://doi.org/10.1038/ng.806>
- Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K. K., Thompson, A., Wood, A. M., Lewington, S., Sattar, N., Packard, C. J., Collins, R., Thompson, S. G., & Danesh, J. (2009). Major lipids, apolipoproteins, and risk of vascular disease. *JAMA - Journal of the American Medical Association*, 302(18), 1993–2000. <https://doi.org/10.1001/jama.2009.1619>
- Dominiczak, M. H., & Caslake, M. J. (2011). Apolipoproteins: Metabolic role and clinical biochemistry applications. *Annals of Clinical Biochemistry*, 48(6), 498–515. <https://doi.org/10.1258/acb.2011.011111>
- Dron, J. S., Wang, J., Cao, H., McIntyre, A. D., Iacocca, M. A., Menard, J. R., Movsesyan, I., Malloy, M. J., Pullinger, C. R., Kane, J. P., & Hegele, R. A. (2019). Severe hypertriglyceridemia is primarily polygenic. *Journal of Clinical Lipidology*, 13(1), 80–88. <https://doi.org/10.1016/j.jacl.2018.10.006>
- Dron, J. S., Wang, J., McIntyre, A. D., Iacocca, M. A., Robinson, J. F., Ban, M. R., Cao, H., & Hegele, R. A. (2020). Six years' experience with LipidSeq: Clinical and research learnings from a hybrid, targeted sequencing panel for dyslipidemias. *BMC Medical Genomics*, 13(1), 1–15. <https://doi.org/10.1186/s12920-020-0669-2>
- Du, W., Hu, Z., Wang, L., Li, M., Zhao, D., Li, H., Wei, J., & Zhang, R. (2020). ABCA1 Variants rs1800977 (C69T) and rs9282541 (R230C) Are Associated with Susceptibility to Type 2 Diabetes. *Public Health Genomics*, 23(1–2), 20–25. <https://doi.org/10.1159/000505344>
- Elder, S. J., Lichtenstein, A. H., Pittas, A. G., Roberts, S. B., Fuss, P. J., Greenberg, A. S., McCrory, M. A.,

- Bouchard, T. J., Saltzman, E., & Neale, M. C. (2009). Genetic and environmental influences on factors associated with cardiovascular disease and the metabolic syndrome. *Journal of Lipid Research*, *50*(9), 1917–1926. <https://doi.org/10.1194/jlr.P900033-JLR200>
- Engler, M. B. (2009). Nutrigenomics in cardiovascular disease: Implications for the future. *Progress in Cardiovascular Nursing*, *24*(4), 190–195. <https://doi.org/10.1111/j.1751-7117.2009.00058.x>
- Engler, M. B., Pullinger, C. R., Malloy, M. J., Natanzon, Y., Kulkarni, M. V., Song, J., Eng, C., Huuskonen, J., Rivera, C., Poon, A., Bensley, M., Sehnert, A., Zellner, C., Kane, J., & Aouizerat, B. E. (2008). Genetic variation in phospholipid transfer protein modulates lipoprotein profiles in hyperalphalipoproteinemia. *Metabolism: Clinical and Experimental*, *57*(12), 1719–1724. <https://doi.org/10.1016/j.metabol.2008.07.031>
- Feingold, K. R., & Grunfeld, C. (2018). Introduction to Lipids and Lipoproteins. In *Endotext* (Issue 1, pp. 1–20). MDText.com, Inc. <https://www.ncbi.nlm.nih.gov/books/NBK305896/>
- Ference, B. A., Robinson, J. G., Brook, R. D., Catapano, A. L., Chapman, M. J., Neff, D. R., Voros, S., Giugliano, R. P., Davey Smith, G., Fazio, S., & Sabatine, M. S. (2016). Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes . *New England Journal of Medicine*, *375*(22), 2144–2153. <https://doi.org/10.1056/nejmoa1604304>
- Flanagan, S. E., Patch, A. M., & Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic Testing and Molecular Biomarkers*, *14*(4), 533–537. <https://doi.org/10.1089/gtmb.2010.0036>
- França, L. T. C., Carrilho, E., & Kist, T. B. L. (2002). A review of DNA sequencing technologies. *Quarterly Reviews of Biophysics*, *35*(2), 169–200.
- Fredrickson, D. S., & Lees, R. S. (1965). A system for phenotyping hyperlipoproteinemia. *Circulation*, *31*(3), 321–327. <https://doi.org/10.1161/01.CIR.31.3.321>
- Gong, J., Campos, H., McGarvey, S., Wu, Z., Goldberg, R., & Baylin, A. (2011). Genetic Variation in Stearoyl-coa Desaturase 1 is associated with metabolic syndrome prevalence in costa rican adults. *Journal of Nutrition*, *141*(12), 2211–2218. <https://doi.org/10.3945/jn.111.143503>
- González Cordero, M. E. (2018). Mutaciones en la región codificante del gen de la lipoproteína lipasa (LPL), en una muestra de pacientes con hipertrigliceridemia residentes en Costa Rica. In *Mutaciones en la región codificante del gen de la lipoproteína lipasa (LPL), en una muestra de pacientes con hipertrigliceridemia residentes en Costa Rica*. Universidad de Costa Rica.
- Guan, Z. W., Wu, K. R., Li, R., Yin, Y., Li, X. L., Zhang, S. F., & Li, Y. (2019). Pharmacogenetics of statins treatment: Efficacy and safety. *Journal of Clinical Pharmacy and Therapeutics*, *44*(6), 858–867.

<https://doi.org/10.1111/jcpt.13025>

- Gunning, A. C., Fryer, V., Fasham, J., Crosby, A. H., Ellard, S., Baple, E. L., & Wright, C. F. (2021). Assessing performance of pathogenicity predictors using clinically relevant variant datasets. *Journal of Medical Genetics*, 58(8), 547–555. <https://doi.org/10.1136/jmedgenet-2020-107003>
- Gustafsen, C., Kjolby, M., Nyegaard, M., Mattheisen, M., Lundhede, J., Buttenschøn, H., Mors, O., Bentzon, J. F., Madsen, P., Nykjaer, A., & Glerup, S. (2014). The hypercholesterolemia-risk gene SORT1 facilitates PCSK9 secretion. *Cell Metabolism*, 19(2), 310–318. <https://doi.org/10.1016/j.cmet.2013.12.006>
- Gutiérrez Ávila, J. D. (2019). *Caracterización de la Región Promotora del Gen de la Apolipoproteína CII (APO CII), cofactor de la Lipoproteína Lipasa (LPL)*. Universidad de Costa Rica.
- Gutiérrez, E. G., José, J., Zúñiga, R., Atención, P. De, Salud, I. De, & Rica, U. D. C. (2010). *Dislipidemia y niveles de lípidos Sanguíneos en pacientes tratados en centros de atención primaria de la zona este de San José, Costa Rica, año 2006*. 7(2), 1–13.
- Hegele, R. A. (2014). Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases. *Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases, January 2014*. <https://doi.org/10.1007/978-3-642-40337-8>
- Henderson, R., O'Kane, M., McGilligan, V., & Watterson, S. (2016). The genetics and screening of familial hypercholesterolaemia. *Journal of Biomedical Science*, 23(1), 1–12. <https://doi.org/10.1186/s12929-016-0256-1>
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322–1332. <https://doi.org/10.1111/j.1755-0998.2009.02591.x>
- Huff, M. W., Assini, J. M., & Hegele, R. A. (2014). Gene therapy for hypercholesterolemia: Sweet dreams and flying machines. *Circulation Research*, 115(6), 542–545. <https://doi.org/10.1161/CIRCRESAHA.114.304800>
- Instituto Nacional de Estadística y Censos. (2019). *Panorama demográfico 2018*.
- Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L. A., Teerlink, C. C., Stanford, J. L., Isaacs, W. B., Xu, J., Cooney, K. A., Lange, E. M., Schleutker, J., Carpten, J. D., ... Sieh, W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of Human Genetics*, 99(4), 877–885. <https://doi.org/10.1016/j.ajhg.2016.08.016>
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing

with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>

Johansen, C. T., Dubé, J. B., Loyzer, M. N., MacDonald, A., Carter, D. E., McIntyre, A. D., Cao, H., Wang, J., Robinson, J. F., & Hegele, R. A. (2014). LipidSeq: A next-generation clinical resequencing panel for monogenic dyslipidemias. *Journal of Lipid Research*, 55(4), 765–772. <https://doi.org/10.1194/jlr.D045963>

Johansen, C. T., & Hegele, R. A. (2011). Genetic bases of hypertriglyceridemic phenotypes. *Current Opinion in Lipidology*, 22(4), 247–253. <https://doi.org/10.1097/MOL.0b013e3283471972>

Johansen, C. T., Kathiresan, S., & Hegele, R. A. (2011). Genetic determinants of plasma triglycerides. *Journal of Lipid Research*, 52(2), 189–206. <https://doi.org/10.1194/jlr.R009720>

Kerr, I. D., Haider, A. J., & Gelissen, I. C. (2011). The ABCG family of membrane-associated transporters: You don't have to be big to be mighty. *British Journal of Pharmacology*, 164(7), 1767–1779. <https://doi.org/10.1111/j.1476-5381.2010.01177.x>

Kirkegård, J., Cronin-Fenton, D., Heide-Jørgensen, U., & Mortensen, F. V. (2018). Acute Pancreatitis and Pancreatic Cancer Risk: A Nationwide Matched-Cohort Study in Denmark. *Gastroenterology*, 154(6), 1729–1736. <https://doi.org/10.1053/j.gastro.2018.02.011>

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. <https://doi.org/10.1111/1755-0998.12387>

Kumar, K. R., Cowley, M. J., & Davis, R. L. (2019). Next-Generation Sequencing and Emerging Technologies. *Seminars in Thrombosis and Hemostasis*, 45(7), 661–673. <https://doi.org/10.1055/s-0039-1688446>

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., ... Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1), D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>

Langlois, M. R., & Nordestgaard, B. G. (2018). Which Lipids Should Be Analyzed for Diagnostic Workup and Follow-up of Patients with Hyperlipidemias? *Current Cardiology Reports*, 20(10). <https://doi.org/10.1007/s11886-018-1036-1>

Lee, J., & Hegele, R. A. (2017). Investigated treatments for lipoprotein lipase deficiency and related metabolic disorders. *Expert Opinion on Orphan Drugs*, 5(5), 411–420. <https://doi.org/10.1080/21678707.2017.1311784>

Lewis, G. F., Xiao, C., & Hegele, R. A. (2015). Hypertriglyceridemia in the genomic era: A new paradigm.

*Endocrine Reviews*, 36(1), 131–147. <https://doi.org/10.1210/er.2014-1062>

Li, Y., Maciel, D., Shi, X., & Toma, H. (2014). *Biodegradable Polymer Nanogels for Drug / Nucleic Acid Delivery*. <https://doi.org/10.1021/cr500131f>

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012. <https://doi.org/10.1155/2012/251364>

Luo, Y., Eran, A., Palmer, N., Avillach, P., Levy-Moonshine, A., Szolovits, P., & Kohane, I. S. (2020). A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia. *Nature Medicine*, 26(9), 1375–1379. <https://doi.org/10.1038/s41591-020-1007-0>

Mahmood, H. (2014). Intestinal Lipid Absorption and Lipoprotein Formation. *Current Opinion in Lipidology*, 25(3), 200–206. <https://doi.org/10.1097/MOL.000000000000084>. Intestinal

Mather, K. A., Thalamuthu, A., Oldmeadow, C., Song, F., Armstrong, N. J., Poljak, A., Holliday, E. G., McEvoy, M., Kwok, J. B., Assareh, A. A., Reppermund, S., Kochan, N. A., Lee, T., Ames, D., Wright, M. J., Trollor, J. N., Schofield, P. W., Brodaty, H., Scott, R. J., ... Sachdev, P. S. (2016). Genome-wide significant results identified for plasma apolipoprotein H levels in middle-aged and older adults. *Scientific Reports*, 6(November 2015), 1–9. <https://doi.org/10.1038/srep23675>

McDonald, J. G., Ivanova, P. T., & Brown, H. A. (2016). Approaches to Lipid Analysis. In *Biochemistry of Lipids, Lipoproteins and Membranes* (pp. 41–72). Elsevier. <https://doi.org/10.1016/B978-0-444-63438-2.00002-X>

Mega, J. L., Morrow, D. A., Brown, A., Cannon, C. P., & Sabatine, M. S. (2009). Identification of genetic variants associated with response to statin therapy. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 29(9), 1310–1315. <https://doi.org/10.1161/ATVBAHA.109.188474>

Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., & Devine, S. E. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research*, 16(9), 1182–1190. <https://doi.org/10.1101/gr.4565806>

Muniraj, T., Gajendran, M., Thiruvengadam, S., Raghuram, K., Rao, S., & Devaraj, P. (2012). Acute Pancreatitis. *Disease-a-Month*, 58(3), 98–144. <https://doi.org/10.1016/j.disamonth.2012.01.005>

Nakayama, K., Bayasgalan, T., Tazoe, F., Yanagisawa, Y., Gotoh, T., Yamanaka, K., Ogawa, A., Munkhtulga, L., Chimedregze, U., Kagawa, Y., Ishibashi, S., & Iwamoto, S. (2010). A single nucleotide polymorphism in the FADS1/FADS2 gene is associated with plasma lipid profiles in two genetically similar Asian ethnic groups with distinctive differences in lifestyle. *Human Genetics*, 127(6), 685–690. <https://doi.org/10.1007/s00439-010-0815-6>

- Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. <https://doi.org/10.1093/nar/gkg509>
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., & Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, *461*(7261), 272–276. <https://doi.org/10.1038/nature08250>
- Nocoń-Bohusz, J., Wikiera, B., Basiak, A., Śmigiel, R., & Noczyńska, A. (2015). LPL gene mutation as the cause of severe hypertriglyceridemia in the course of ketoacidosis in a patient with newly diagnosed type 1 diabetes mellitus. *Pediatric Endocrinology, Diabetes and Metabolism*, *21*(2), 89–92. <https://doi.org/10.18544/PEDM-21.02.0029>
- Nordestgaard, B. G. (2016). Triglyceride-Rich Lipoproteins and Atherosclerotic Cardiovascular Disease: New Insights from Epidemiology, Genetics, and Biology. *Circulation Research*, *118*(4), 547–563. <https://doi.org/10.1161/CIRCRESAHA.115.306249>
- Palacio Rojas, M., Prieto, C., Bermúdez, V., Garicano, C., Núñez Nava, T., Martínez, M. S., Salazar, J., Rojas, E., Pérez, A., Marca Vicuña, P., González Martínez, N., Maldonado Parra, S., Hoedebecke, K., D’Addosio, R., Cano, C., & Rojas, J. (2017). Dyslipidemia: Genetics, lipoprotein lipase and HindIII polymorphism. *F1000Research*, *6*(May), 2073. <https://doi.org/10.12688/f1000research.12938.1>
- Paredes, S., Fonseca, L., Ribeiro, L., Ramos, H., Oliveira, J. C., & Palma, I. (2019). Novel and traditional lipid profiles in Metabolic Syndrome reveal a high atherogenicity. *Scientific Reports*, *9*(1), 1–7. <https://doi.org/10.1038/s41598-019-48120-5>
- Paththinige, C. S., Sirisena, N. D., & Dissanayake, V. H. W. (2017). Genetic determinants of inherited susceptibility to hypercholesterolemia - a comprehensive literature review. *Lipids in Health and Disease*, *16*(1), 1–22. <https://doi.org/10.1186/s12944-017-0488-4>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11). <https://doi.org/10.1371/journal.pgen.1002967>
- Pirim, D., Radwan, Z. H., Wang, X., Niemsiri, V., Hokanson, J. E., Hamman, R. F., Feingold, E., Bunker, C. H., Yesim Demirci, F., & Ilyas Kamboh, M. (2019). Apolipoprotein E-C1-C4-C2 gene cluster region and inter-individual variation in plasma lipoprotein levels: A comprehensive genetic association study in two ethnic groups. *PLoS ONE*, *14*(3), 1–29. <https://doi.org/10.1371/journal.pone.0214060>
- Pirim, D., Wang, X., Radwan, Z. H., Niemsiri, V., Bunker, C. H., Barmada, M. M., Kamboh, M. I., & Demirci, F. Y. (2015). Resequencing of LPL in African Blacks and associations with lipoprotein-lipid levels. *European Journal of Human Genetics*, *23*(9), 1244–1253. <https://doi.org/10.1038/ejhg.2014.268>

- Pirim, D., Wang, X., Radwan, Z. H., Niemsiri, V., Hokanson, J. E., Hamman, R. F., Barmada, M. M., Demirci, F. Y., & Kamboh, M. I. (2014). Lipoprotein lipase gene sequencing and plasma lipid profile. *Journal of Lipid Research*, *55*(1), 85–93. <https://doi.org/10.1194/jlr.M043265>
- Plaisier, C. L., Horvath, S., Huertas-Vazquez, A., Cruz-Bautista, I., Herrera, M. F., Tusie-Luna, T., Aguilar-Salinas, C., & Pajukanta, P. (2009). A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genetics*, *5*(9). <https://doi.org/10.1371/journal.pgen.1000642>
- Pomaznoy, M., Ha, B., & Peters, B. (2018). GOnet: A tool for interactive Gene Ontology analysis. *BMC Bioinformatics*, *19*(1), 1–8. <https://doi.org/10.1186/s12859-018-2533-3>
- Poplin, R., Ruano-Rubio, V., DePristo, M., Fennell, T., Carneiro, M., Van der Auwera, G., Kling, D., Gauthier, L., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M., Neale, B., MacArthur, D., & Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178. <https://doi.org/10.1101/201178>
- Porchay, I., Péan, F., Bellili, N., Royer, B., Cogneau, J., Chesnier, M.-C., Caradec, A., Tichet, J., Balkau, B., Marre, M., & Fumeron, F. (2006). ABCA1 Single Nucleotide Polymorphisms on High-Density Lipoprotein-Cholesterol and Overweight: the D.E.S.I.R. Study\*. *Obesity*, *14*(11), 1874–1879. <https://doi.org/10.1038/oby.2006.217>
- Qiu, J., Qin, X., Wen, J., Wu, L., Kong, L., Ou, Y., & Yang, S. (2020). Single-nucleotide polymorphisms influence the promoter activities of systemic lupus erythematosus-associated genes. *Biotechnology Letters*, *42*(10), 1887–1896. <https://doi.org/10.1007/s10529-020-02916-y>
- Reiner, Ž. (2017). Hypertriglyceridaemia and risk of coronary artery disease. *Nature Reviews Cardiology*, *14*(7), 401–411. <https://doi.org/10.1038/nrcardio.2017.31>
- Ritchie, G. R. S., Dunham, I., Zeggini, E., & Flicek, P. (2014). Functional annotation of noncoding sequence variants. *Nature Methods*, *11*(3), 294–296. <https://doi.org/10.1038/nmeth.2832>
- Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., & Campbell, C. (2018). FATHMM-XF: Accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*, *34*(3), 511–513. <https://doi.org/10.1093/bioinformatics/btx536>
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Ruiz-Narváez, E. A., Bare, L., Arellano, A., Catanese, J., & Campos, H. (2010). West african and amerindian ancestry and risk of myocardial infarction and metabolic syndrome in the central valley population of costa rica. *Human Genetics*, *127*(6), 629–638. <https://doi.org/10.1007/s00439-010-0803-x>

- Ruiz-Narváez, E. A., Sacks, F. M., & Campos, H. (2008). Abdominal obesity and hyperglycemia mask the effect of a common APOC3 haplotype on the risk of myocardial infarction. *American Journal of Clinical Nutrition*, 87(6), 1932–1938. <https://doi.org/10.1093/ajcn/87.6.1932>
- Ruiz-Narváez, E. A., Yang, Y., Nakanishi, Y., Kirchdorfer, J., & Campos, H. (2005). APOC3/A5 haplotypes, lipid levels, and risk of myocardial infarction in the Central Valley of Costa Rica. *Journal of Lipid Research*, 46(12), 2605–2613. <https://doi.org/10.1194/jlr.M500040-JLR200>
- Sałaacka, A., Boroń, A., Gorący, I., Hornowska, I., Safranow, K., & Ciechanowicz, A. (2021). An association of ABCG8: rs11887534 polymorphism and HDL-cholesterol response to statin treatment in the Polish population. *Pharmacological Reports*, 73(6), 1781–1786. <https://doi.org/10.1007/s43440-021-00302-7>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sarraj, A., & Knowles, J. W. (2019). Genetic Testing and Risk Scores: Impact on Familial Hypercholesterolemia. *Frontiers in Cardiovascular Medicine*, 6. <https://doi.org/10.3389/fcvm.2019.00005>
- Schuster, K. B., Wilfert, W., Evans, D., Thiery, J., & Teupser, D. (2011). Identification of mutations in the lipoprotein lipase (LPL) and apolipoprotein C-II (APOC2) genes using denaturing high performance liquid chromatography (DHPLC). *Clinica Chimica Acta*, 412(3–4), 240–244. <https://doi.org/10.1016/j.cca.2010.10.006>
- Stitzel, N. O., Kiezun, A., & Sunyaev, S. (2011). Computational and statistical approaches to analyzing variants identified by exome sequencing. *Genome Biology*, 12(9). <https://doi.org/10.1186/gb-2011-12-9-227>
- Stroes, E., Moulin, P., Parhofer, K. G., Rebours, V., Löhr, J. M., & Averna, M. (2017). Diagnostic algorithm for familial chylomicronemia syndrome. *Atherosclerosis Supplements*, 23, 1–7. <https://doi.org/10.1016/j.atherosclerosissup.2016.10.002>
- Su, X., & Peng, D. (2020). The exchangeable apolipoproteins in lipid metabolism and obesity. *Clinica Chimica Acta*, 503(139), 128–135. <https://doi.org/10.1016/j.cca.2020.01.015>
- Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Casale, F. P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., ... Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571), 75–81. <https://doi.org/10.1038/nature15394>
- Surendran, R. P., Visser, M. E., Heemelaar, S., Wang, J., Peter, J., Defesche, J. C., Kuivenhoven, J. A., Hosseini, M., Péterfy, M., Kastelein, J. J. P., Johansen, C. T., Hegele, R. A., Stroes, E. S. G., & Dallinga-Thie, G. M. (2012). Mutations in LPL, APOC2, APOA5, GPIHBP1 and LMF1 in patients with severe

hypertriglyceridaemia. *Journal of Internal Medicine*, 272(2), 185–196. <https://doi.org/10.1111/j.1365-2796.2012.02516.x>

- Tada, H., Kawashiri, M. A., & Yamagishi, M. (2017). Comprehensive genotyping in dyslipidemia: Mendelian dyslipidemias caused by rare variants and Mendelian randomization studies using common variants. *Journal of Human Genetics*, 62(4), 453–458. <https://doi.org/10.1038/jhg.2016.159>
- Tejedor, M. T., Garcia-Sobreviela, M. P., Ledesma, M., & Arbones-Mainar, J. M. (2014). The apolipoprotein e polymorphism rs7412 associates with body fatness independently of plasma lipids in middle aged men. *PLoS ONE*, 9(9), 1–5. <https://doi.org/10.1371/journal.pone.0108605>
- Thompson, J. F., Hyde, C. L., Wood, L. S., Paciga, S. A., Hinds, D. A., Cox, D. R., Hovingh, G. K., & Kastelein, J. J. P. (2009). Comprehensive whole-genome and candidate gene analysis for response to statin therapy in the treating to new targets (TNT) cohort. *Circulation: Cardiovascular Genetics*, 2(2), 173–181. <https://doi.org/10.1161/CIRCGENETICS.108.818062>
- Tiwari, S., & Siddiqi, S. A. (2012). Intracellular trafficking and secretion of VLDL. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 32(5), 1079–1086. <https://doi.org/10.1161/ATVBAHA.111.241471>
- Torres-Romero, J., César Lara-Riegos, J., Antonio Estrella Parra, E., Fernández Sánchez, V., E. Arana-Argáez, V., Uc-Colli, S., Ángel Peña-Rico, M., Alberto Ramírez-Camacho, M., Dolores Ponce Regalado, M., & Elizabeth Alvarez-Sánchez, M. (2020). Lipoproteomics: Methodologies and Analysis of Lipoprotein-Associated Proteins along with the Drug Intervention. In *Drug Design - Novel Advances in the Omics Field and Applications [Working Title]: Vol. i* (Issue tourism, p. 13). IntechOpen. <https://doi.org/10.5772/intechopen.93634>
- Tso, P., Nauli, A., & Lo, C. M. (2004). Enterocyte fatty acid uptake and intestinal fatty acid-binding protein. *Biochemical Society Transactions*, 32(1), 75–78. <https://doi.org/10.1042/BST0320075>
- Vazquez Vidal, I. (2014). *Impact of Fatty Acid Desaturase (FADS) genotypes on the relationship between serum lipids and dietary fat intake in young mexican college students*. University of Illinois at Urbana-Champaign.
- Vekic, J., Zeljkovic, A., Stefanovic, A., Jelic-Ivanovic, Z., & Spasojevic-Kalimanovska, V. (2019). Obesity and dyslipidemia. *Metabolism: Clinical and Experimental*, 92, 71–81. <https://doi.org/10.1016/j.metabol.2018.11.005>
- Villarreal-Molina, T., Posadas-Romero, C., Romero-Hidalgo, S., Antúnez-Argüelles, E., Bautista-Grande, A., Vargas-Alarcón, G., Kimura-Hayama, E., Canizales-Quinteros, S., Juárez-Rojas, J. G., Posadas-Sánchez, R., Cardoso-Saldaña, G., Medina-Urrutia, A., González-Salazar, M. del C., Martínez-Alvarado, R., Jorge-Galarza, E., & Carnevale, A. (2012). The ABCA1 Gene R230C Variant Is Associated with Decreased Risk of Premature Coronary Artery Disease: The Genetics of Atherosclerotic Disease (GEA) Study. *PLoS ONE*,

7(11), 1–9. <https://doi.org/10.1371/journal.pone.0049285>

- Vinueza, R., Boissonnet, C. P., Acevedo, M., Uriza, F., Benitez, F. J., Silva, H., Schargrotsky, H., Champagne, B., & Wilson, E. (2010). Dyslipidemia in seven Latin American cities: CARMELA study. *Preventive Medicine, 50*(3), 106–111. <https://doi.org/10.1016/j.ypmed.2009.12.011>
- von Zychlinski, A., & Kleffmann, T. (2015). Dissecting the proteome of lipoproteins: New biomarkers for cardiovascular diseases? *Translational Proteomics, 7*, 30–39. <https://doi.org/10.1016/j.trprot.2014.12.001>
- Wang, J., Raskin, L., Samuels, D. C., Shyr, Y., & Guo, Y. (2015). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics, 31*(3), 318–323. <https://doi.org/10.1093/bioinformatics/btu668>
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research, 38*(16), 1–7. <https://doi.org/10.1093/nar/gkq603>
- Wang, L. J., Zhang, C. W., Su, S. C., Chen, H. I. H., Chiu, Y. C., Lai, Z., Bouamar, H., Ramirez, A. G., Cigarroa, F. G., Sun, L. Z., & Chen, Y. (2019). An ancestry informative marker panel design for individual ancestry estimation of Hispanic population using whole exome sequencing data. *BMC Genomics, 20*(Suppl 12), 1–14. <https://doi.org/10.1186/s12864-019-6333-6>
- White, M. J., Eren, F., Aիրbaşlı, D., Chen, J., Hu, T., Moore, J. H., Williams, S. M., & Aիրbaşlı, M. (2015). A systems genetics approach to dyslipidemia in children and adolescents. *OMICS A Journal of Integrative Biology, 19*(4), 248–259. <https://doi.org/10.1089/omi.2014.0140>
- Wierzbicki, A. S., & Reynolds, T. M. (2019). Genetic risk scores in lipid disorders. *Current Opinion in Cardiology, 34*(4), 406–412. <https://doi.org/10.1097/HCO.0000000000000623>
- Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., Beckmann, J. S., Bragg-Gresham, J. L., Chang, H. Y., Demirkan, A., Den Hertog, H. M., Do, R., Donnelly, L. A., Ehret, G. B., Esko, T., ... Abecasis, G. R. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics, 45*(11), 1274–1285. <https://doi.org/10.1038/ng.2797>
- World Health Organization. (2009). *Global Health Risks*. [http://www.who.int/healthinfo/global\\_burden\\_disease/GlobalHealthRisks\\_report\\_full.pdf](http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf)
- World Health Organization. (2018). Noncommunicable diseases country profiles 2018. In *Heart of Africa: Clinical Profile of an Evolving Burden of Heart Disease in Africa*.
- World Health Organization. (2020). *International Statistical Classification of Diseases and Related Health Problems*. (11th Ed.; ICD-11).

- Xiao, C., Hsieh, J., Adeli, K., & Lewis, G. F. (2011). Gut-liver interaction in triglyceride-rich lipoprotein metabolism. *American Journal of Physiology - Endocrinology and Metabolism*, 301(3).  
<https://doi.org/10.1152/ajpendo.00178.2011>
- Yang, Y., Ruiz-Narvaez, E., Niu, T., Xu, X., & Campos, H. (2004). Genetic variants of the lipoprotein lipase gene and myocardial infarction in the Central Valley of Costa Rica. *Journal of Lipid Research*, 45(11), 2106–2109. <https://doi.org/10.1194/jlr.M400202-JLR200>
- Yu, Z., Huang, T., Zheng, Y., Wang, T., Heianza, Y., Sun, D., Campos, H., & Qi, L. (2017). PCSK9 variant, long-chain n<sup>3</sup> PUFAs, and risk of nonfatal myocardial infarction in Costa Rican Hispanics. *American Journal of Clinical Nutrition*, 105(5), 1198–1203. <https://doi.org/10.3945/ajcn.116.148106>
- Zanti, M., Michailidou, K., Loizidou, M. A., Machattou, C., Pirpa, P., Christodoulou, K., Spyrou, G. M., Kyriacou, K., & Hadjisavvas, A. (2021). Performance evaluation of pipelines for mapping, variant calling and interval padding, for the analysis of NGS germline panels. *BMC Bioinformatics*, 22(1), 1–21.  
<https://doi.org/10.1186/s12859-021-04144-1>
- Zhang, R., Liu, Q., Liu, H., Bai, H., Zhang, Y., Guan, L., & Fan, P. (2018). Effects of apoC1 genotypes on the hormonal levels, metabolic profile and PAF-AH activity in Chinese women with polycystic ovary syndrome. *Lipids in Health and Disease*, 17(1), 1–10. <https://doi.org/10.1186/s12944-018-0725-5>
- Zhang, X., Johnson, A. D., Hendricks, A. E., Hwang, S. J., Tanriverdi, K., Ganesh, S. K., Smith, N. L., Peyser, P. A., Freedman, J. E., & O'Donnell, C. J. (2014). Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes. *Human Molecular Genetics*, 23(3), 795–782. <https://doi.org/10.1093/hmg/ddt461>
- Zhou, L., Li, C., Gao, L., & Wang, A. (2015). High-density lipoprotein synthesis and metabolism (Review). *Molecular Medicine Reports*, 12(3), 4015–4021. <https://doi.org/10.3892/mmr.2015.3930>
- Zhou, Y., Mägi, R., Milani, L., & Lauschke, V. M. (2018). Global genetic diversity of human apolipoproteins and effects on cardiovascular disease risk. *Journal of Lipid Research*, 59(10), 1987–2000.  
<https://doi.org/10.1194/jlr.P086710>

## Anexos

### Anexo 1. Variantes que producen alteraciones en el marco abierto de lectura encontradas en genes involucrados en el metabolismo de lípidos categorizadas como riesgosas por una fuente como mínimo.

Los resultados de las predicciones *in silico* en **negrita** superan el umbral determinado en la metodología para categorizar un alelo como variante de riesgo. Categorías PolyPhen2: B – variante benigna, P – variante posiblemente dañina, D – variante probablemente dañina.

Coordenadas	Símbolo genético	Identificador genético de Ensembl	Variante	Alelos (REF/ALT)	Frecuencia de alelo alternativo		Alteración en marco abierto de lectura		Predicción <i>in silico</i>			Asociaciones fenotípicas vinculadas con las condiciones estudiadas
					CR-WGS	IKGP - ALL	Codón	Aminoácido	Categoría PolyPhen2	Índice REVEL	Índice ClinPred	
chr1:25563756	LDLRAP1	ENSG00000157978	rs41291058	C/T	0.0210	0.0234	Cgg/Tgg	R/W	<b>P</b>	0.107	0.031	-
chr1:55039974	PCSK9	ENSG00000169174	rs11591147	G/T	0.0069	0.0103	cGt/cTt	R/L	B	0.028	0.013	<b>Fuentes:</b> ClinVar, NHGRI-EBI GWAS catalog <b>Términos:</b> colesterol total (Cholesterol, total), hipercolesterolemia familiar 1 y 3 ( <i>Familial hypercholesterolemia 1, Familial hypercholesterolemia 3</i> ), hipobetalipoproteinemia ( <i>Hypobetalipoproteinemia</i> ), hipocolesterolemia ( <i>Hypocholesterolemia</i> ), C-LDL ( <i>LDL cholesterol</i> )
chr1:55039995	PCSK9	ENSG00000169174	rs11583680	C/T	0.0970	0.0753	gCc/gTc	A/V	B	0.065	0.009	<b>Fuentes:</b> ClinVar <b>Términos:</b> hipercolesterolemia familiar 1 y 3 ( <i>Familial hypercholesterolemia 1, Familial hypercholesterolemia 3</i> ), hipobetalipoproteinemia ( <i>Hypobetalipoproteinemia</i> ), C-LDL ( <i>Low-density lipoprotein cholesterol</i> )
chr1:109253120	CELSR2	ENSG00000143126	rs115856488	T/A	0.0069	0.0030	cTt/cAt	L/H	<b>D</b>	0.303	0.045	-
chr1:109262946	CELSR2	ENSG00000143126	rs202022169	T/C	0.0069	0.0007	aTt/aCt	I/T	<b>D</b>	<b>0.710</b>	0.094	-
chr1:109272311	CELSR2	ENSG00000143126	rs375735096	C/T	0.0069	0.0001	Cgg/Tgg	R/W	<b>P</b>	0.151	0.356	-
chr2:21008823	APOB	ENSG00000084674	rs375053331	C/A	0.0069	3.7E-05	aGt/aTt	S/I	-	0.137	<b>0.704</b>	-
chr2:27498276	GCKR	ENSG00000084734	rs146175795	G/A	0.0280	0.0029	Gtg/Atg	V/M	-	<b>0.689</b>	<b>0.830</b>	<b>Fuentes:</b> ClinVar <b>Términos:</b> hipertrigliceridemia ( <i>Hypertriglyceridemia</i> )
chr2:43832056	ABCG5	ENSG00000138075	rs145164937	G/C	0.0280	0.0019	gCc/gGc	A/G	<b>P</b>	<b>0.871</b>	0.065	-
chr2:43837951	ABCG5	ENSG00000138075	rs6756629	G/A	0.0280	0.0666	Cgc/Tgc	R/C	<b>P</b>	0.332	0.025	-

Coordenadas	Símbolo genético	Identificador genético de Ensembl	Variante	Alelos (REF/ALT)	Frecuencia de alelo alternativo		Alteración en marco abierto de lectura		Predicción <i>in silico</i>			Asociaciones fenotípicas vinculadas con las condiciones estudiadas
					CR-WGS	IKGP - ALL	Codón	Aminoácido	Categoría PolyPhen2	Índice REVEL	Índice ClinPred	
chr2:43839108	ABCG8	ENSG00000143921	rs11887534	G/C	0.0280	0.0580	Gat/Cat	D/H	P	0.359	0.035	Fuentes: ClinVar , NHGRI-EBI GWAS catalog Términos: niveles de proteína C reactiva y -C-LDL (C-reactive protein levels or C-LDLcholesterol levels (pleiotropy))
chr2:43872126	ABCG8	ENSG00000143921	rs756606650	C/A	0.0140	0.0001	aCc/aAc	T/N	B	0.121	<b>0.920</b>	-
chr3:195573869	APOD	ENSG00000189058	rs76929107	C/T	0.0069	0.0075	Gtg/Atg	V/M	D	0.386	0.075	-
chr6:46705206	PLA2G7	ENSG00000146070	rs1051931	A/G	0.8400	0.8012	gTa/gCa	V/A	P	0.198	0.043	-
chr7:44536297	NPC1L1	ENSG00000015520	rs202127286	G/A	0.0069	0.0002	Cgt/Tgt	R/C	P	0.273	0.107	-
chr7:44539580	NPC1L1	ENSG00000015520	rs138140250	C/A	0.0350	0.0048	Gac/Tac	D/Y	P	0.421	0.043	-
chr7:44540069	NPC1L1	ENSG00000015520	rs148405668	G/A	0.0069	0.0005	Ctc/Ttc	L/F	D	<b>0.790</b>	0.048	-
chr7:80663155	CD36	ENSG00000135218	rs1057086208	G/A	0.0140	9.0E-06	Ggt/Agt	G/S	D	<b>0.712</b>	<b>0.996</b>	-
chr7:80672777	CD36	ENSG00000135218	rs146027667	G/T	0.0069	0.0006	gGa/gTa	G/V	D	<b>0.890</b>	0.415	-
chr8:19956018	LPL	ENSG00000175445	rs268	A/G	0.0210	0.0103	aAt/aGt	N/S	B	0.312	0.033	Fuentes: ClinVar Términos: susceptibilidad ante hiperlipidemia familiar combinada (HYPERLIPIDEMIA, FAMILIAL COMBINED, SUSCEPTIBILITY TO, Hyperlipidemia, familial combined, LPL related), hiperlipoproteinemia tipo 1 (Hyperlipoproteinemia, type 1)
chr8:125431236	TRIB1	ENSG00000173334	rs200584467	C/T	0.0210	0.0008	Cac/Tac	H/Y	B	0.208	<b>0.803</b>	-
chr9:104819977	ABCA1	ENSG00000165029	rs140365800	T/C	0.0069	0.0014	gAt/gGt	D/G	D	<b>0.841</b>	0.055	Fuentes: ClinVar Términos: Hipoalfalipoproteinemia familiar (Familial hypoalphalipoproteinemia)
chr10:89412847	LIPA	ENSG00000107798	rs115595484	G/A	0.0210	0.0301	gCg/gTg	A/V	P	0.121	0.033	-
chr11:824789	PNPLA2	ENSG00000177666	rs1138693	T/C	0.7430	0.6683	cTg/cCg	L/P	P	0.127	0.017	-
chr11:61802814	FADS1	ENSG00000149485	rs776632991	C/T	0.0069	1.4E-05	Gcc/Acc	A/T	-	0.226	<b>0.659</b>	-
chr11:116791691	APOA5	ENSG00000110243	rs3135506	G/C	0.0970	0.0633	tCg/tGg	S/W	P	0.223	0.017	Fuentes: NHGRI-EBI GWAS catalog Términos: niveles de colesterol total (Total cholesterol levels), susceptibilidad a hipertrigliceridemia (HYPERTRIGLYCERIDEMIA, SUSCEPTIBILITY TO)
chr11:116820918	APOA4	ENSG00000110244	rs5110	C/A	0.0760	0.0429	caG/caT	Q/H	-	0.047	0.001	Fuentes: ClinVar Términos: polimorfismo en apolipoproteína A (APOLIPOPROTEIN A-IV POLYMORPHISM, APOA4*1/APOA4*2), colesterol total (Cholesterol, total)

Coordenadas	Símbolo genético	Identificador genético de Ensembl	Variante	Alelos (REF/ALT)	Frecuencia de alelo alternativo		Alteración en marco abierto de lectura		Predicción <i>in silico</i>			Asociaciones fenotípicas vinculadas con las condiciones estudiadas
					CR-WGS	IKGP - ALL	Codón	Aminoácido	Categoría PolyPhen2	Índice REVEL	Índice ClinPred	
chr11:116821244	APOA4	ENSG00000110244	rs1366912241	C/T	0.0069	3.7E-05	Gag/Aag	E/K	-	0.386	<b>0.652</b>	-
chr12:57156140	LRP1	ENSG00000123384	rs1377430371	C/T	0.0069	2.0E-05	gCc/gTc	A/V	<b>D</b>	<b>0.611</b>	<b>0.550</b>	-
chr12:57173171	LRP1	ENSG00000123384	rs1185554986	C/T	0.0069	7.0E-06	aCg/aTg	T/M	<b>D</b>	0.418	<b>0.966</b>	-
chr12:57180338	LRP1	ENSG00000123384	rs1414626991	A/G	0.0140	4.3E-05	Att/Gtt	I/V	<b>P</b>	<b>0.570</b>	0.337	-
chr12:57195877	LRP1	ENSG00000123384	rs755040010	G/A	0.0069	0.0003	Ggc/Agc	G/S	<b>D</b>	<b>0.540</b>	0.074	-
chr12:124863717	SCARB1	ENSG00000073060	rs4238001	C/T	0.1110	0.0841	Ggc/Agc	G/S	<b>D</b>	0.178	0.029	-
chr15:58560876	LIPC	ENSG00000166035	rs140272400	A/G	0.0069	0.0026	cAg/cGg	Q/R	<b>D</b>	0.446	0.068	-
chr15:89667102	PLIN1	ENSG00000166819	rs8179071	G/A	0.0069	0.0101	tCg/tTg	S/L	<b>D</b>	0.224	0.020	-
chr16:870870	LMF1	ENSG00000103227	rs35168378	C/T	0.0420	0.0362	cGg/cAg	R/Q	<b>P</b>	0.179	0.034	-
chr16:56981179	CETP	ENSG00000087237	rs5880	G/C	0.1110	0.0436	Gcc/Ccc	A/P	<b>D</b>	0.136	0.019	-
chr16:56982180	CETP	ENSG00000087237	rs5882	G/A	0.7360	0.5377	Gtc/Atc	V/I	B	0.013	0.009	Fuentes: ClinVar Términos: hiperalfalipoproteinemia (HYPERALPHALIPOPROTEINEMIA 1), C-HDL (HIGH DENSITY LIPOPROTEIN CHOLESTEROL LEVEL QUANTITATIVE TRAIT LOCUS 10)
chr17:66212167	APOH	ENSG00000091583	rs1801690	C/G	0.0420	0.0414	tGg/tCg	W/S	<b>D</b>	0.175	0.016	-
chr17:66214462	APOH	ENSG00000091583	rs1801689	A/C	0.0140	0.0196	Tgc/Ggc	C/G	<b>D</b>	<b>0.757</b>	0.042	-
chr18:49581514	LIPG	ENSG00000101670	rs61729805	C/G	0.0140	0.0094	aCt/aGt	T/S	<b>D</b>	0.445	0.042	-
chr18:49583585	LIPG	ENSG00000101670	rs77960347	A/G	0.0350	0.0071	aAc/aGc	N/S	<b>D</b>	0.342	0.051	-
chr19:4157200	CREB3L3	ENSG00000060566	rs149917178	G/A	0.0069	0.0001	gGg/gAg	G/E	<b>D</b>	0.295	0.080	-
chr19:4164630	CREB3L3	ENSG00000060566	rs779860332	C/A	0.0069	0.0002	cCc/cAc	P/H	<b>D</b>	<b>0.749</b>	<b>0.963</b>	-
chr19:11089606	LDLR	ENSG00000130164	rs147509697	G/A	0.0069	0.0006	Ggg/Agg	G/R	B	<b>0.558</b>	0.025	-
chr19:11105549	LDLR	ENSG00000130164	rs764042910	C/T	0.0140	0.0001	Cgc/Tgc	R/C	<b>D</b>	<b>0.598</b>	<b>0.990</b>	Fuentes: ClinVar Términos: hipercolesterolemia familiar (Familial hypercholesterolemia)
chr19:19219115	NCAN	ENSG00000130287	rs2228603	C/T	0.0280	0.0387	Ccc/Tcc	P/S	<b>P</b>	0.098	0.035	-
chr19:42407617	LIPE	ENSG00000079435	rs34052647	G/A	0.0760	0.0139	Cgt/Tgt	R/C	<b>D</b>	0.389	0.048	-

Coordenadas	Símbolo genético	Identificador genético de Ensembl	Variante	Alelos (REF/ALT)	Frecuencia de alelo alternativo		Alteración en marco abierto de lectura		Predicción <i>in silico</i>			Asociaciones fenotípicas vinculadas con las condiciones estudiadas
					CR-WGS	IKGP - ALL	Codón	Aminoácido	Categoría PolyPhen2	Índice REVEL	Índice ClinPred	
chr19:42410620	LIPE	ENSG00000079435	rs138539064	G/A	0.0069	0.0024	cCg/cTg	P/L	D	0.564	0.060	-
chr19:42410677	LIPE	ENSG00000079435	rs201395817	G/A	0.0069	0.0004	cCg/cTg	P/L	P	0.391	0.213	-
chr19:42410729	LIPE	ENSG00000079435	rs750246327	G/A	0.0210	0.0000	Cgg/Tgg	R/W	D	0.380	0.920	-
chr19:44908684	APOE	ENSG00000130203	rs429358	T/C	0.1320	0.1770	Tgc/Cgc	C/R	B	0.229	0.000	Fuentes: ClinVar Términos: hipercolesterolemia familiar tipo 3 (familial hypercholesterolemia, Familial type 3 hyperlipoproteinemia)
chr19:44908822	APOE	ENSG00000130203	rs7412	C/T	0.0830	0.0888	Cgc/Tgc	R/C	D	0.588	0.031	-
chr19:44908915	APOE	ENSG00000130203	rs749750245	C/T	0.0069	0.0004	Cgc/Tgc	R/C	D	0.481	0.123	-
chr19:44949172	APOC2	ENSG00000234906	rs5126	A/C	0.0000	0.0093	Aaa/Caa	K/Q	B	0.337	0.073	Fuentes: ClinVar Términos: deficiencia de apolipoproteína C2 (Apolipoprotein C2 deficiency)
chr20:45905002	PLTP	ENSG00000100979	rs542300038	G/C	0.0069	0.0002	gaC/gaG	D/E	P	0.278	0.151	-
chr20:45909538	PLTP	ENSG00000100979	rs144016582	C/T	0.0140	0.0003	Gcg/Acg	A/T	P	0.091	0.050	-
chr21:42276899	ABCG1	ENSG00000160179	rs138056067	C/T	0.0069	0.0008	tCg/tTg	S/L	D	0.214	0.288	-
chr22:36141715	APOL3	ENSG00000128284	rs61741884	C/T	0.0210	0.0329	Gtg/Atg	V/M	P	0.094	0.045	-
chr22:36160775	APOL3	ENSG00000128284	rs132653	G/T	0.9170	0.6979	agC/agA	S/R	P	0.187	0.022	-
chr22:36191175	APOL4	ENSG00000100336	rs6000173	G/T	0.6600	0.6506	gCa/gAa	A/E	D	0.039	0.028	-
chr22:36191263	APOL4	ENSG00000100336	rs150226555	G/A	0.0069	0.0016	Cgg/Tgg	R/W	P	0.117	0.049	-
chr22:36199341	APOL4	ENSG00000100336	rs200978577	C/T	0.0069	0.0006	gGa/gAa	G/E	P	0.073	0.030	-
chr22:36265284	APOL1	ENSG00000100342	rs2239785	G/A	0.7780	0.5980	Gag/Aag	E/K	P	0.031	0.028	-
chr22:36265363	APOL1	ENSG00000100342	rs116136671	A/G	0.0069	0.0181	aAt/aGt	N/S	D	0.063	0.044	-
chr22:36265603	APOL1	ENSG00000100342	rs554687721	A/G	0.0069	0.0006	gAg/gGg	E/G	D	0.044	0.254	-
chr22:36265687	APOL1	ENSG00000100342	rs150685787	G/A	0.0069	0.0026	cGt/cAt	R/H	D	0.092	0.060	-
chr22:36265934	APOL1	ENSG00000100342	rs370010415	G/C	0.0069	0.0001	gaG/gaC	E/D	P	0.049	0.214	-
chr22:46218290	PPARA	ENSG00000186951	rs565902106	A/G	0.0069	0.0001	Aag/Gag	K/E	P	0.907	0.579	-

Coordenadas	Símbolo genético	Identificador genético de Ensembl	Variante	Alelos (REF/ALT)	Frecuencia de alelo alternativo		Alteración en marco abierto de lectura		Predicción <i>in silico</i>			Asociaciones fenotípicas vinculadas con las condiciones estudiadas
					CR-WGS	1KGP - ALL	Codón	Aminoácido	Categoría PolyPhen2	Índice REVEL	Índice ClinPred	
chr22:46218377	PPARA	ENSG00000186951	rs1800206	C/G	0.0280	0.0362	Ctt/Gtt	L/V	B	<b>0.668</b>	0.021	<b>Fuentes:</b> ClinVar <b>Términos:</b> susceptibilidad a hiperapobetalipoproteinemia (HYPERAPOBETALIPOPROTEINEMIA, SUSCEPTIBILITY TO)