



**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL**

UNIDAD IRAPUATO

**“Análisis de la incidencia de mutaciones somáticas *de novo* y
del perfil transcripcional de genes asociados con la
aterosclerosis en un modelo murino”**

Tesis que presenta

M.C. Maria del Pilar Valencia Morales

Para obtener el grado de

Doctora en Ciencias

En la especialidad de

Biotecnología de Plantas

Directores de tesis: Dra. Gertrud Lund

Dr. Silvio Zaina

Irapuato, Guanajuato

febrero 2017

El presente trabajo titulado “Análisis de la incidencia de mutaciones somáticas *de novo* y del perfil transcripcional de genes asociados con la aterosclerosis murina”, se realizó en el Laboratorio de Epigenética del Departamento de Ingeniería Genética del Centro de Investigación y de Estudios Avanzados del IPN, Unidad Irapuato; bajo la dirección de la Dra. Gertrud Lund y en el Departamento de Ciencias Médicas de la División de Ciencias de la Salud, Campus León de la Universidad de Guanajuato, bajo la codirección del Dr. Silvio Zaina, y con el soporte de una beca CONACyT #234871.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT), por la beca de Doctorado otorgada durante el periodo comprendido entre septiembre de 2011 y agosto de 2015.

Al Centro de Investigación y de Estudios Avanzados del IPN de la Unidad Irapuato (CINVESTAV-Irapuato), por permitirme realizar mis estudios de Doctorado y formar parte de este reconocido centro de investigación.

A la Dra. Gertrud Lund y el Dr. Silvio Zaina; les agradezco infinitamente toda su asesoría, comentarios y valiosos aportes al proyecto. Porque siempre me han brindado su apoyo.

A mis sinodales, los Doctores: Stefan De Folter, Stewart Gillmor, Alfredo Cruz y Maria Luisa Lazo, por aceptar ser parte de mi comité tutorial y enriquecer con sus acertadas observaciones y sugerencias el presente trabajo.

Al Dr. Luis Delaye Arredondo; quien no pudo estar presente en el examen de titulación, pero fungió como miembro de mi comité tutorial durante la totalidad del proyecto y de igual forma realizó aportes valiosos al mismo.

A la Lic. BQ Dalia Rodríguez Ríos, por la asesoría otorgada durante el desarrollo experimental de la tesis, específicamente durante la extracción de ácidos nucleicos de tejidos de ratón.

A todo el personal académico, administrativo y de mantenimiento del CINVESTAV-Irapuato por todas las facilidades prestadas. En especial quiero agradecer a Dora Elia Anguiano porque amablemente me orientó, incluso desde antes de llegar a México.

Al Departamento de Ciencias Médicas de la universidad de Guanajuato, sede Campus León; por el apoyo otorgado para la realización de diferentes experimentos. En especial, a la Dra. Yolanda Caudillo y al Dr. Rubén Rangel Salazar por su amable asesoría experimental. También quiero agradecer al Dr. Joel Ramírez Emiliano, quien amablemente me facilitó el microscopio estereoscópico para todas y cada una de las disecciones que realicé.

Al Laboratorio de Epigenética y todos sus integrantes por su apoyo y colaboración.

A Magda Ramírez-Nava y a Gladys López-González, quienes fueron de gran ayuda durante el trabajo de bioterio previo al inicio de este proyecto (mantenimiento de las cepas de ratones); por ayudarme a desarrollar habilidades en el bioterio.

A todos mis amigos y compañeros durante la maestría y el doctorado, los presentes y ausentes, todos y cada uno de ellos han sido primordiales para sentirme como en casa.

Por último y no menos importante, a mi familia, quienes son el motor que me impulsa a seguir adelante cada día.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	VI
ÍNDICE DE TABLAS	VII
ABREVIATURAS	VIII
RESUMEN	IX
ABSTRACT	X
CAPITULO I	11
Introducción	11
1.1. Mutaciones en el DNA	12
1.2. Variantes germinales y somáticas	14
1.3. El mosaicismo somático en humanos	15
1.4. Nuevas asociaciones genéticas en la era de la secuenciación masiva	17
1.5. Variación genética en el genoma de ratón	19
1.6. Aterosclerosis	19
1.7. ApoE knock-out: modelo de aterosclerosis	21
Justificación	23
Objetivos	24
CAPITULO II	25
Materiales y métodos	25
2.1. Muestras biológicas	25
2.2. Secuenciación de RNA (RNA-Seq)	26
2.3. Secuenciación de exoma completo (WES)	26
2.4. Llamado de variantes	27
2.5. Anotación funcional de las variantes detectadas	27
2.6. Análisis de enriquecimiento de vías de señalización	28
CAPITULO III	29
Resultados	29
3.1. Variantes somáticas de novo en ratones con aterosclerosis	29

3.1.2.	Obtención de ratones ateroscleróticos y normales.....	29
3.1.3.	Determinación de variantes genéticas por medio de secuenciación del exoma (WES) de ratones consanguíneos.....	30
3.1.3.1.	Distribución cromosómica diferencial de las variantes.....	32
3.1.3.2.	Variantes somáticas de novo de ratones ApoE ^{-/-} y silvestres.....	35
3.1.3.3.	Análisis de las variantes compartidas entre individuos ApoE ^{-/-}	38
3.1.4.	Identificación de variantes por secuenciación de RNA total (RNA seq) de ratones consanguíneos.....	45
3.1.4.1.	Variantes específicas de ratones ApoE ^{-/-} y silvestres.....	46
3.1.4.2.	v-WES y v-RNAseq específicas de ApoE ^{-/-}	47
3.1.4.3.	Relevancia de los genes con variantes comunes a ApoE ^{-/-}	48
3.2.	Expresión diferencial de genes y su influencia con la determinación de variantes.....	61
3.3.	Asociación del perfil de expresión de genes en aortas de humanos y ratones que sufren aterosclerosis.....	63
CAPITULO IV	69
	Discusión.....	69
	Conclusiones.....	75
	Perspectivas.....	76
Apéndice	77
Bibliografía	90

ÍNDICE DE FIGURAS

Figura 1. Mosaicismo genético (modificado de Miliaras 2015)	15
Figura 2. Factores que afectan el genoma de un individuo desde la concepción hasta la muerte (Modificado de Gottlieb et al., 2014)	16
Figura 3. Progresión de la lesión aterosclerótica en humanos (Modificado de Moore & Tabas, 2011)	20
Figura 4. Placa aterosclerótica en la aorta de un ratón ApoE ^{-/-}	30
Figura 5. Cuantificación de variantes totales en el DNA por cromosoma	32
Figura 6. Visualización general de variantes en los cromosomas	33
Figura 7. Visualización de las vWES del cromosoma 1 y 6	34
Figura 8. Visualización de v-WES y lecturas totales del cromosoma 6	35
Figura 9. Porcentaje de genes por cromosoma.	38
Figura 10. v-COMP de aorta e hígado	39
Figura 11. Distribución cromosómica de variantes del RNA	45
Figura 12. Distribución cromosómica de genes con variantes	47
Figura 13. Ubicación del SNP rs30693033 de Immt en ratón	60
Figura 14. Distribución cromosómica de v-WES, v-RNAseq y de genes expresados diferencialmente	62

ÍNDICE DE TABLAS

Tabla 1. Niveles de colesterol y triglicéridos en plasma de ratones ApoE ^{-/-} alimentados con una dieta Chow por 2-3 semanas. (Tomado de Jawień et al, 2004).	22
Tabla 2. Datos generales de la WES	31
Tabla 3. Número de variantes y su respectiva cobertura por tejido	32
Tabla 4. Generalidades de las v-WES/ApoE ^{-/-} y v-WES/WT	36
Tabla 5. Análisis de enriquecimiento GO y KEGG de las v-COMP específicas de aorta	40
Tabla 6. Análisis de enriquecimiento GO y KEGG de las v-COMP específicas de hígado	42
Tabla 7. Datos generales de las v-RNAseq/ApoE ^{-/-} y v-RNAseq/WT	46
Tabla 8. Variantes putativas de aorta	49
Tabla 9. Variantes putativas de hígado	52
Tabla 10. Análisis de Gene Ontology (GO) y KEGG de aorta	57
Tabla 11. Análisis de Gene Ontology y KEGG de hígado	58
Tabla 12. Número de genes únicos expresados diferencialmente en ApoE ^{-/-} por tejido	61
Tabla 13. Análisis de correlación de grado-CpGs y expresión diferencial de genes en ratón	65

ABREVIATURAS

ApoB-LPs	apolipoprotein B-containing lipoproteins.
BP	Biological Process.
CC	Cellular Component.
DAVID	Database for Annotation, Visualization and Integrated Discovery.
ECVs	Enfermedades Cardiovasculares.
GO	Gene Ontology.
GWAS	Genome-wide association study.
IMC	Índice de Masa Corporal.
INEGI	Instituto Nacional de Estadística y Geografía.
INTERPRO	Protein Sequence Analysis & Classification.
LDL	Low Density Lipoprotein.
MAF	Minor Allele Frequency.
Mb	Millones de bases.
mDNA	Metilación del DNA
ME	Músculo esquelético.
MF	Molecular Function.
MH	Missing heritability.
NK	Natural Killer.
OPS	Organización Panamericana de la Salud.
Pb	pares de bases.
RNA-Seq	RNA-Sequencing.
SNP	Single Nucleotide Polymorphism.
SNV	Single Nucleotide Variant.
Ti	Transversiones.
Tv	Transiciones.
v-COMP	Variantes compartidas.
VEP	Variant Effect Predictor.
V(D)J	Variable, Diversity, Joining.
VLDL	Very Low Density lipoprotein.
VSDNs	Variantes somáticas <i>de novo</i> .
VSN	Variante de un solo nucleótido.
v-WES	Variantes obtenidas por WES.
WES	Whole Exome Sequencing.
WGS	Whole Genome Sequencing.
WHO	World Health Organization.

RESUMEN

La aterosclerosis es una enfermedad inflamatoria que se caracteriza por la presencia lesiones fibrocelulares en la pared vascular de las arterias de mediano y grueso calibre. Las complicaciones clínicas que conlleva este padecimiento hacen parte de las enfermedades cardiovasculares (ECVs) y representan la mayor causa de muerte a nivel mundial. Los factores de riesgo incluyen tanto componentes ambientales como genéticos. Los polimorfismos de un solo nucleótido (Single Nucleotide Polymorphism o SNP) son un factor genético que explica menos del 30% del riesgo de esta enfermedad. La falta de contundencia a la hora de encontrar asociaciones genéticas, sugiere la participación de otros mecanismos como las variantes somáticas *de novo*, cuyo papel primordial ya ha sido determinado en las enfermedades neurológicas y en el cáncer.

El objetivo de este trabajo fue analizar la incidencia de las variantes somáticas *de novo* en ratones deficientes en la apolipoproteína E ($ApoE^{-/-}$), un modelo de aterosclerosis inducida por hiperlipidemia. Se realizó secuenciación total del exoma (WES) de la aorta y el hígado de dos parejas de hermanos consanguíneos C57BL/6J (silvestres y homocigotos para la mutación $ApoE^{-/-}$), en una de ellas se analizó también el músculo. Además, en esta última, se analizaron los tres tejidos por secuenciación de RNA (RNA-Seq).

Se identificó un mayor número de variantes somáticas *de novo* por WES y RNA-seq en ratones $ApoE^{-/-}$ comparado con los silvestres. De manera interesante, la mayoría de estas variantes mapearon diferencialmente entre individuos $ApoE^{-/-}$ y silvestres en los cromosomas 6, 7 y 8. De igual forma, hubo variantes comunes entre diferentes ratones $ApoE^{-/-}$ (24 y 31 % en aorta e hígado, respectivamente), en donde la mayoría fueron tejido específicas. Adicionalmente, encontramos que solo el ~3.5 % de las variantes específicas de $ApoE^{-/-}$ fueron conservadas a nivel de RNA. Una de las explicaciones a este resultado, podría estar relacionada con expresión diferencial, ya que observamos que los cromosomas con mayor número de variantes estaban asociados con mayor número de genes expresados diferencialmente. De este modo, obtuvimos dos grupos de variantes (16 y 37 en aorta e hígado, respectivamente), todas ellas reportadas como SNP y cuyas implicaciones funcionales sugieren su participación en procesos relacionados con la aterosclerosis como: la respuesta inmune, la proliferación celular, la regulación de la transcripción; así como la regulación de las vías de señalización de insulina y glucagón.

ABSTRACT

Atherosclerosis is an inflammatory disease characterized by fibrocellular lesions in the vascular wall of large and medium sized arteries. Clinical complications associated with this condition are part of cardiovascular disease (CVD) and represent the leading cause of death worldwide. Risk factors include both environmental and genetic components. The single nucleotide polymorphisms (SNPs) explain only ~ 30% of the disease risk. The lack of forcefulness in finding genetic associations suggests the involvement of other mechanisms such as *de novo* somatic variants, whose primary role has already been determined in cancer and neurological diseases.

The aim of this study was to analyze the incidence of *de novo* somatic variants in apolipoprotein E mice deficient (*ApoE*^{-/-}), a model of atherosclerosis induced by hyperlipidemia. The aorta, liver and skeletal muscle of C57BL/6J consanguineous male ApoE-null/WT sibling pairs were analyzed by whole exome sequencing (WES) and RNA-seq (one of the pairs subjected to whole-exome sequencing).

We identified a high number of *de novo* somatic variants in *ApoE*^{-/-} compared to WT mice. Interestingly, most of these variants mapped differentially between *ApoE*^{-/-} and WT mice on chromosomes 6, 7 and 8. Similarly, there were common variants between different *ApoE*^{-/-} individuals (24 and 31% in aorta and liver, respectively), where most variants were tissue-specific. Additionally, we found that only 3.5%~ of *ApoE*^{-/-} specific variants were preserved at the RNA level. One explanation for this result could be related to differential expression, as we observed that chromosomes with more variants were associated with a higher number of differentially expressed genes. Thus, we obtained two groups of variants (16 and 37 in aorta and liver, respectively), all reported as SNPs (database SNP). The functional implications suggest their involvement atherosclerosis related processes as immune response, cell proliferation and regulation of transcription and insulin/glucagon signaling pathways.

CAPITULO I

Introducción

La población mexicana está viviendo una transición epidemiológica en la cual la mayoría de los fallecimientos son causados por enfermedades no transmisibles. El sobrepeso, la obesidad, el consumo de azúcares y de alcohol, son responsables de la mayoría de estas enfermedades (Stevens *et al.*, 2008). Se estima que el 71.2% de los adultos en México tiene sobrepeso y obesidad (Gutiérrez *et al.*, 2012) y que las dos primeras causas de muerte son las enfermedades cardiovasculares y la diabetes tipo II (INEGI, 2014). La dieta determina en gran medida la incidencia de este tipo de padecimientos. En un informe de la Organización Panamericana de la Salud (OPS), se describe cómo en 13 países de Latinoamérica, el consumo de alimentos y bebidas ultra-procesados ha aumentado entre el año 2000 y 2013. Dicho estudio puntualiza, entre otras cosas, que el aumento en la venta de estos productos, esta correlacionado positiva y significativamente con un incremento en el índice de masa corporal (IMC>25) en la población adulta, en donde México ocupa uno de los primeros lugares (WHO, 2015).

La aterosclerosis, a menudo ocasionada por factores de riesgo como la obesidad, es un origen patológico importante de las enfermedades cardiovasculares (ECVs) y que a su vez son una de las principales causas de la morbilidad y mortalidad en el mundo. En Estados Unidos, por ejemplo, en el 2013 hubo una tasa de 222.9 muertes por cada cien mil habitantes, en promedio una muerte cada 40 segundos, lo que representa más fallecimientos por año en comparación con el cáncer. A nivel mundial, se estima que fallecen 17.3 millones de personas cada año a causa de las ECVs (30% del total de muertes en el 2013), y se estima que esta cifra crezca a 23.6 millones para el 2030. Los países de ingreso económico medio y bajo son los que padecen la mayoría de dichas muertes (80%) (revisado en Mozaffarian *et al.*, 2015).

Al igual que en otras enfermedades comunes, en la aterosclerosis existen interacciones complejas entre el genoma y múltiples factores ambientales, en donde la epigenética es un modulador importante (revisado en Zhong *et al.*, 2016). Tal es así, que algunos factores de la dieta como por ejemplo los ácidos grasos, pueden inducir cambios en la expresión de genes que están asociados con alteraciones en marcas epigenéticas

como son la metilación de DNA y la modificación pos-transcripcional de las histonas (revisado en Lund & Zaina, 2011).

Aunque la genética también cumple un papel importante en las ECVs, éste no es claro ya que se han identificado numerosos SNP - la gran mayoría son estudios en sangre periférica - que en conjunto explican solo un pequeño porcentaje de la heredabilidad del riesgo cardiovascular (Stylianou *et al.*, 2012). Este fenómeno es conocido como “missing heritability” (MH) (Maher, 2008; Manolio *et al.*, 2009) y aunque hay diversas hipótesis para explicarlo, las causas definitivas siguen siendo desconocidas. Las posibilidades abarcan desde la falta de detección de los componentes genéticos de fenotipos intermedios que influyen el fenotipo principal (ej: la susceptibilidad a la enfermedad isquémica del corazón es influenciada por fenotipos intermedios como presión sanguínea e hipercolesterolemia, entre otros) (Blanco-Gómez *et al.*, 2016); hasta la incidencia de variantes genéticas con baja frecuencia (minor allele frequency o MAF entre 1–5%,) o variantes raras (MAF < 1%), tal como se ha determinado en algunos rasgos lipídicos (Kanoni & Masca, 2016).

Un papel poco estudiado, pero en principio no menos crucial, es la contribución de las mutaciones post-cigóticas somáticas. Estas pueden dar lugar a un genotipo diferencial de SNPs entre el tejido vascular y las células de sangre periférica. Las mutaciones somáticas post-cigóticas no son heredadas y por lo tanto no explican la MH, pero si existen pueden afectar las conclusiones derivadas de los análisis de asociación genética.

En esta sección de Introducción, se presentará primero la relevancia de las mutaciones somáticas en el contexto de las variantes genéticas. Posteriormente, se revisarán las características más importantes del proceso aterosclerótico y del modelo animal utilizado en el presente trabajo experimental.

1.1. Mutaciones en el DNA

El término “mutación” ha sido comúnmente acuñado para describir cambios en la secuencia nucleotídica del DNA que derivan o no en una enfermedad (Condit *et al.*, 2002). A su vez, una mutación es llamada “polimorfismo” cuando su frecuencia poblacional es del 1% o más (Brookes, 1999). Sin embargo, en la era de la secuenciación genómica personal, se ha propuesto una revisión a estos términos con miras a definir las mutaciones con relación al genoma germinal de cada individuo. De esta forma, una mutación podría ser una variante

en el DNA, adquirida a lo largo de la vida de un individuo (“mutación somática”, presente en todas o algunas células somáticas) o ser transmitida por la línea germinal y estar presente en todas las células de la progenie (“mutaciones germinales”) (Karki *et al.*, 2015). Por lo anterior, y aunado a la connotación negativa del término “mutación” generalmente adoptada (asociadas a enfermedades, a exposición a diferentes tipos de radiación, entre otras), en adelante utilizaremos el término “variante” al igual que otros han sugerido (Condit *et al.*, 2002; MacArthur *et al.*, 2014; Richards *et al.*, 2015).

De acuerdo a la porción del genoma que involucran, las variantes pueden clasificarse como cromosómicas o génicas. Las primeras implican la ganancia o pérdida ya sea de segmentos cromosómicos, hasta cromosomas individuales o juegos completos de los mismos. Por su parte, las variantes génicas o puntuales (mapean en un solo locus cromosomal), afectan secuencias pequeñas de DNA (inserciones y deleciones, o INDELS) o bien sitios específicos como las sustituciones de un solo nucleótido o VSNs (variantes de un solo nucleótido) (Vissers *et al.*, 2003). Las VSNs se dividen en transiciones, cuando una purina es reemplazada por otra purina, o una pirimidina por otra pirimidina (C-T); y en transversiones cuando la base sustituida es contraria a la base inicial (una purina por una pirimidina y viceversa) (Griffiths *et al.*, 2000). Otro tipo de variante genética es definida como variación en el número de copias o CNV (por su nombre en inglés Copy Number Variation); en este caso, la expresión de un gen es amplificada o reducida por el incremento (o disminución) en el número de copias de un alelo (Sebat *et al.*, 2004).

A nivel de proteína, las VSNs pueden ser silenciosas, si codifican para el mismo aminoácido; sinónimas cuando el codón involucrado es diferente, pero codifica para un aminoácido equivalente (Lys por Arg, ambos son aminoácidos básicos). También hay variantes de pérdida de sentido o sin sentido cuando el codón codifica un aminoácido diferente sin funcionalidad o el nuevo codón es de terminación, respectivamente. En cuanto a las inserciones o deleciones de uno o varios nucleótidos, éstos pueden generar un desplazamiento del marco de lectura, así que cuando la adición o deleción no es múltiplo de 3, la proteína codificada será diferente a partir del punto donde ocurrió la variante o incluso tendrá un codón de paro prematuro (Griffiths *et al.*, 2000).

Como se había mencionado antes, los SNPs son VSNs que tienen un contexto poblacional, de tal manera que se encuentran en una frecuencia de al menos 1% y constituyen la variante genética más común en el genoma humano (1 en 1000 pb). El patrón de SNPs sobre un mismo cromosoma puede ser heredado en bloques (haplotipos) y brinda

información sustancial en cuanto a la diversidad evolutiva y asociaciones genéticas en humanos (Frazer *et al.*, 2007a). A nivel poblacional, el consorcio del proyecto de 1000 genomas reportó en el 2015, 88 millones de variantes en 26 poblaciones, de las cuales el 96% son SNPs; de las variantes restantes, 3.6 millones son INDEls y 60.000 son variantes estructurales. De manera interesante, encontraron también que un genoma típico de un individuo, difiere del genoma de referencia en 4.1 a 5 millones de sitios (0.1% del genoma) en donde los individuos con ancestría africana son los más diferentes (Auton *et al.*, 2015). Esta pequeña variación podría explicar las diferencias individuales en cuanto a susceptibilidad a enfermedades, respuesta a medicamentos o a factores ambientales (Frazer *et al.*, 2007a).

En el contexto de enfermedades humanas, las variantes o SNPs de transición son particularmente importantes ya que se ha reportado que el dinucleótido CpG es blanco de variantes patológicas en el genoma humano. El 35 % de las variantes genéticas que causan enfermedades en humanos caen en dinucleótidos CpG y de estas, el 90 % son transiciones C-T o G-A (Cooper & Youssoufian, 1988; Cooper *et al.*, 2010).

1.2. Variantes germinales y somáticas

Una variación genética en la línea germinal es cualquier cambio detectable y heredable en el linaje de las células germinales, teniendo como referencia al genoma del linaje germinal parental. Este tipo de variantes se dan predominantemente durante la gametogénesis y como consecuencia, están presentes en todos los tejidos somáticos de la progenie del individuo (Griffiths *et al.*, 2000). Ahora se sabe que muchas variantes que se creía que ocurrían en la línea germinal, son en realidad variantes somáticas *de novo*, o en algunos casos fueron heredadas de uno de los padres, pero con un nivel de mosaicismo muy bajo (Acuna-Hidalgo *et al.*, 2015; Campbell *et al.*, 2014; Scally, 2016). Por otro lado, las variantes somáticas pueden surgir durante el desarrollo temprano del cigoto y dan lugar a un mosaicismo embrionario, en cuyo caso coexistirían dos o más poblaciones celulares distintas en un mismo individuo (**Figura 1**). A su vez, un organismo mosaico, aparentemente sano, podría llevar una variante germinal en sus células germinales y transmitirla a la progenie (Huang *et al.*, 2014).

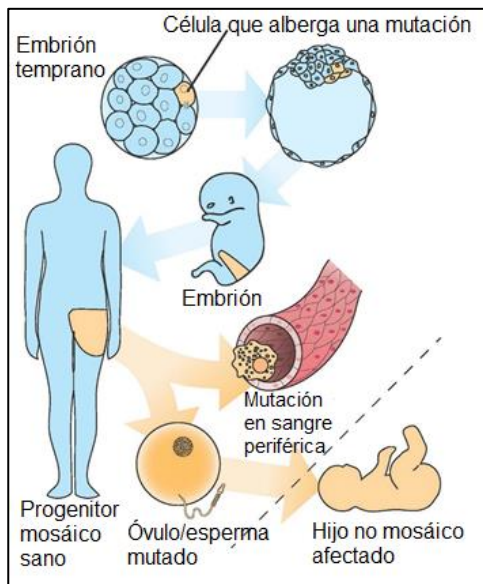


Figura 1. Mosaicismo genético (modificado de Miliaras 2015)

Numerosas líneas de evidencia, muchas de ellas basadas en secuenciación completa del genoma de tríos (padres e hijo), muestran que la tasa de sustitución de bases en humanos esta entre 1.1 y 1.7×10^{-8} variantes por nucleótido por generación (Besenbacher *et al.*, 2015; Campbell *et al.*, 2012; Kong *et al.*, 2012; Lynch, 2010; Séguirel *et al.*, 2014). Esto corresponde a ~ 74 VSN por genoma por generación. Por su parte, la tasa de INDELS es $\sim 8\%$ de la tasa de sustitución de bases, dando alrededor 3 nuevos INDELS por generación (Besenbacher *et al.*, 2015; Kloosterman *et al.*, 2015; O'Roak *et al.*, 2012). Por otro lado, la tasa de variación somática por sustitución de bases es de 4-25 veces más grande que la correspondiente tasa para la línea germinal (revisado en Lynch, 2010).

1.3. El mosaicismo somático en humanos

El término "mosaicismo somático" fue acuñado a mediados de los 1950s por Cotterman en su trabajo sobre la variación antigénica (Cotterman, 1956) y se refiere a la coexistencia de poblaciones celulares somáticas con distintos genotipos en un mismo individuo. Con frecuencia, el mosaicismo somático es enmascarado, pero también puede dar lugar a grandes cambios fenotípicos y revelar la expresión de variantes genéticas de otro modo letales para el organismo.

En la actualidad, se conoce que entre los tejidos somáticos sanos, así como al interior de ellos, hay una cuantiosa variabilidad genética. Los grandes proyectos poblacionales como el “1000 Genomes Project Consortium” (Consortium *et al.*, 2010) y el “The Cancer Genome Atlas”, así como investigaciones hechas en menor número de individuos, han mostrado una gran heterogeneidad inter e intra-individual (Gerlinger *et al.*, 2012; Russnes *et al.*, 2011).

Aunque las variaciones somáticas *de novo* son una fuente importante de variación genética poblacional, en ocasiones también son la causa del desarrollo de enfermedades, ya sea causando directamente las condiciones de la enfermedad, o en otros casos modulando el fenotipo ya existente (Poduri *et al.*, 2013). En efecto, su relevancia ha sido encontrada en enfermedades como el cáncer (Berenblum and Shubik, 1949; Rozhok and DeGregori, 2015), en donde incluso se ha visto que la carga de variantes somáticas puede reducir la adecuación de las células sanas en el tejido benigno circundante, incrementando la selección de los eventos oncogénicos. En el caso del cáncer de ovario, hay variantes somáticas que se generan *de novo* y que tienen gran heterogeneidad genética al interior del tumor mamario (Bashashati *et al.*, 2013), además de ello, algunos de los genes expresan preferencialmente la variante alélica (Halabi *et al.*, 2016).

Las variantes genéticas que se generan durante el desarrollo del individuo y que posteriormente se van seleccionando a lo largo de su vida gracias a cambios en el microambiente de los tejidos, dan lugar a una heterogeneidad fenotípica en las células y tejidos y pueden tener implicaciones importantes para el desarrollo de cáncer y de ECVs (**Figura 2**) (Gottlieb, 2010; Gottlieb, 2014).

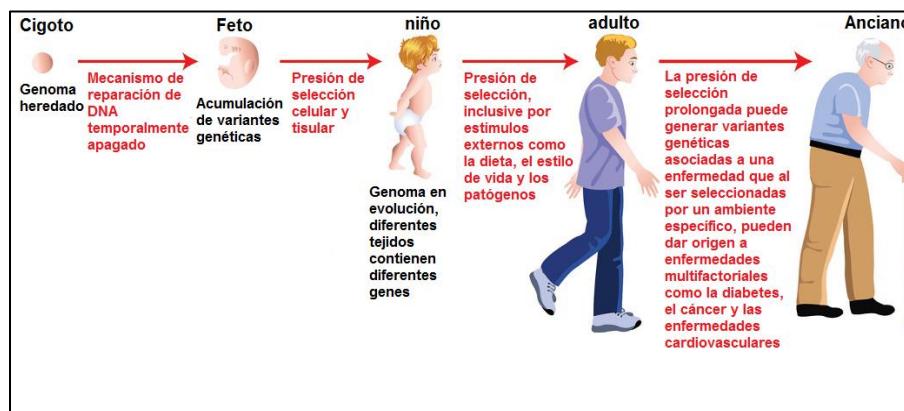


Figura 2. Factores que afectan el genoma de un individuo desde la concepción hasta la muerte (Modificado de Gottlieb *et al.*, 2014)

Debido al avance en el campo de la secuenciación masiva de DNA, las variantes somáticas *de novo* han sido relacionadas con otro tipo de patologías diferentes al cáncer. Tal es el caso de algunas formas raras o comunes de enfermedades del desarrollo neurológico como la discapacidad cognitiva, autismo, esquizofrenia y los desórdenes del espectro de autismo (ASD), en donde las variantes *de novo* juegan un papel prominente (revisado en Freed and Pevsner, 2016; Veltman and Brunner, 2012). Existen otros casos como la Hemimegalencefalia (HMG), en donde la presencia de una variante, incluso en el 8% de las células, es suficiente para cambiar el tamaño y la arquitectura de un hemisferio cerebral completo (Evrony *et al.*, 2012; Lee *et al.*, 2012; Poduri *et al.*, 2012).

1.4. Nuevas asociaciones genéticas en la era de la secuenciación masiva

En los últimos años, se han identificado factores genéticos de riesgo ya sea a través del estudio de genes candidatos, rasgos mendelianos, o de los estudios de asociación del genoma completo (GWAS). A nivel poblacional, los GWAS tienen el propósito de encontrar variantes comunes asociadas a enfermedades comunes (Reich & Lander, 2001). Aunque estos análisis incluyen miles de muestras en estudios individuales y metaanálisis, solo permiten identificar la asociación con SNPs frecuentes (frecuencia del alelo menor o $MAF > 0.05$), que comúnmente tienen un efecto pequeño a nivel funcional y sobre el fenotipo. Otras limitaciones de los GWAS incluyen la poca cobertura de las micromatrices usadas (2.5 millones de SNPs en humanos), así como el error inherente que conllevan dado su alto número de hipótesis (cada uno de los SNPs), por lo que exigen para ser significativa una $p < 5 \times 10^{-8}$ y su replicación en una población independiente (Munroe and Tinker, 2015).

Periódicamente, el catálogo de GWAS del NHGRI-EBI adiciona cientos de variantes genéticas asociadas a enfermedades (Welter *et al.*, 2014). Los GWAS, han revelado asociaciones estadísticamente robustas entre determinadas variantes y el riesgo de padecer una enfermedad. La genética de la enfermedad de arterias coronarias ha permitido la identificación por GWAS de 51 loci de asociaciones (Ozaki and Tanaka, 2016). Sin embargo, en el caso de las ECVs, estas variaciones explican en conjunto solo un ~30% o menos de la heredabilidad (McPherson and Tybjaerg-Hansen, 2016). Además, existen

variantes con baja frecuencia (alelos infrecuentes), cuya determinación es difícil por este método y que junto a otros factores también influyen el desarrollo de enfermedades. Por esto, se ha sugerido la creación de plataformas de bases de datos genéticos de nueva generación (NGDB), que puedan ser capaces de incorporar y analizar los factores que contribuyen a la disociación del genotipo y fenotipo. Dichas plataformas deberían incorporar nuevas herramientas bioinformáticas basadas en algoritmos que pueden evaluar heterogeneidad genética, así como herramientas poderosas de análisis de biología de sistemas, que puedan usarse activamente para procesar y evaluar la vasta suma tanto de información genética y epigenética, requerida para revelar las relaciones verdaderas entre genotipo y fenotipo (Gottlieb *et al.*, 2014).

Las técnicas de secuenciación modernas, que son más cuantitativas y profundas, permiten averiguar si por ejemplo los alelos infrecuentes existen en una enfermedad y que impacto biológico tienen. Estas técnicas incluyen la secuenciación completa de un genoma (WGS) o la secuenciación de la porción codificante del mismo (conocido como “exome sequencing o WES”). Aunque los costos de la WGS han disminuido abismalmente con relación a sus inicios, sigue siendo una metodología costosa y computacionalmente engorrosa. Por su parte, la WES provee la oportunidad de secuenciar el DNA codificante de un individuo con un mínimo de mapeo y con relativamente bajo costo. El estudio del exoma, resulta especialmente interesante dado que más del 90-95% de las variantes o polimorfismos asociados a enfermedades se sitúan en las regiones codificantes de proteínas y, por consiguiente, en el exoma (Botstein & Risch, 2003; Stenson *et al.*, 2003).

Ahora se conoce que los alelos menores existen y están cobrando cada vez más importancia. Nelson *et al.* (2012), secuenciaron 202 genes que son blanco de fármacos en 14,002 personas y encontraron que los alelos menores probablemente asociados con enfermedades eran abundantes (Nelson *et al.*, 2012). Recientemente, el consorcio de agregación de exoma (Exome Aggregation Consortium, ExAC) analizó el exoma de al menos 60 mil individuos. Ellos observaron que en promedio había una variante cada ocho bases del genoma, lo que indica una alta recurrencia. Identificaron 3200 genes con variantes que aunque truncaban las proteínas predichas, el 72 % de dichos genes no tienen un fenotipo establecido con alguna enfermedad (Lek *et al.*, 2016).

Estos datos podrían indicar que existen variaciones que los métodos de genotipificación clásica no habían permitido identificar eficientemente. Por eso se plantea por ejemplo que la tasa de variación somática y/o germinal es más alta y relevante de lo

que anteriormente se consideraba. Así, la genotipificación de células de sangre periférica en aterosclerosis no sería una representación fiel del tejido vascular. (Gottlieb *et al.*, 2010).

1.5. Variación genética en el genoma de ratón

El ratón de laboratorio, el organismo modelo más popular en la genética de mamíferos, fue derivado de cruces endogámicos entre ratones silvestres de diferentes subespecies de *Mus musculus* (Beck *et al.*, 2000; Wade *et al.*, 2002). En consecuencia, la variación genética entre las diferentes cepas de laboratorio no es al azar y a menudo viene dada por la herencia de las cepas silvestres. Se ha determinado que la sub especie *M. m. domesticus* contribuyó en mayor medida a dicha variación con un 68%, seguida del híbrido *M. m. molossinu* con un 10%, *M. m. musculus* con un 6% y por último de *M. m. castaneus* con un 3% (Frazer *et al.*, 2007b).

La primera determinación de variación genética entre cepas fue realizada con relación al borrador del genoma de ratón (MGSCv3) realizado con la cepa C57BL/6J. La colección inicial incluyó la determinación de SNPs de las cepas 129S1/SvImJ, C3H/HeJ y BALB/cByJ, en donde encontraron alrededor de 79.000 SNPs (Wade *et al.*, 2002). Posteriormente, se originó el proyecto del genoma de ratón (Mouse Genome Project - MGP) con el objetivo de secuenciar los genomas de las diferentes cepas de ratón de laboratorio y crear un catálogo genómico de las mismas. En una primera fase, se secuenciaron 17 cepas de laboratorio en las cuales se identificaron 56.7 M de SNPs únicos y 8.8 M de INDELS. (Keane *et al.*, 2011). Recientemente, este catálogo ha sido ampliado a 71 M de SNPs y 12 M de INDELS por Doran *et al.* (2016), al secuenciar el genoma completo de 13 cepas endogámicas de ratón de laboratorio; en las cuales se incluyen modelos que son importantes para el estudio de la respuesta inmune, leucemia, pérdida de oído relacionada a la edad y la artritis reumatoide (Doran *et al.*, 2016).

1.6. Aterosclerosis

La aterosclerosis es una enfermedad de las arterias de mediano y grueso calibre que consiste en la formación de una lesión fibrocelular (placa aterosclerótica) en la pared

vascular. La lesión contiene una población celular diversa, incluyendo células inflamatorias, endoteliales y de músculo liso y su formación involucra décadas de cambios estructurales de la capa íntima del tejido vascular. Dichos cambios son inicialmente provocados por la infiltración y acumulación de lipoproteínas como la de baja densidad (LDL), y consisten en la migración de células inflamatorias desde la sangre hacia la pared vascular acompañada por la migración y proliferación de las células del músculo liso vascular y la producción de matriz extracelular (**Figura 3**).

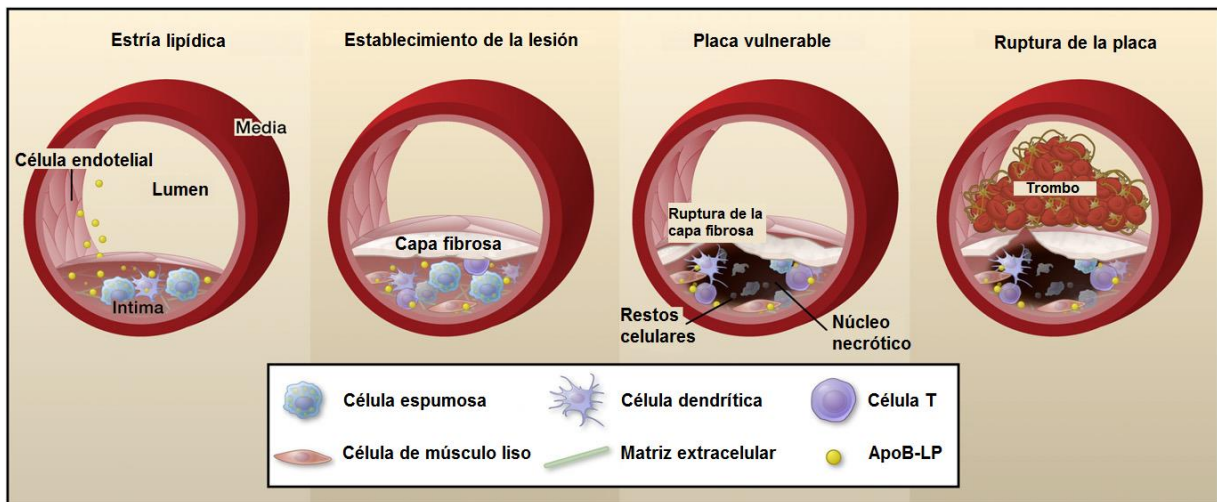


Figura 3. Progresión de la lesión aterosclerótica en humanos (Modificado de Moore & Tabas, 2011)

Esta condición conocida como aterosclerosis subclínica suele ser asintomática por años. Sin embargo, eventualmente con el paso del tiempo, la acumulación de células apoptóticas y de sus restos celulares, junto a la producción de proteasas por los macrófagos, debilita la lesión vascular y desencadena un evento trombótico (revisado en Moore & Tabas, 2011). Estos eventos conllevan a las complicaciones clínicas de la aterosclerosis: infartos al miocardio, ictus isquémico, lesiones isquémicas en riñones e intestino, enfermedad vascular periférica entre otras manifestaciones (Libby *et al*, 2013).

Un factor predominante en el proceso aterosclerótico y sus complicaciones es la inflamación. Los monocitos y macrófagos constituyen la mayoría de las células inmunes en la lesión y están involucrados con el tamaño, composición y vulnerabilidad de la placa aterosclerótica (Swirski *et al*, 2016). También hay marcadores de inflamación cuyas

variantes alélicas están asociadas con ECVs. Algunos comprenden proteínas de fase aguda como la proteína C reactiva (CRP) y fibrinógeno, citoquinas como como la interleucina 6 (IL-6) y moléculas de adhesión celular como la molécula de adhesión intercelular (ICAM) (Raman *et al.*, 2013). Desde el punto de vista terapéutico, algunas vías de señalización inflamatorias han sido objeto de estrategias que buscan disminuir la incidencia de la aterosclerosis, incluso algunas han probado la inmunización contra los autoantígenos de la LDL oxidada y Hsp60 en modelos animales (Nilsson *et al.*, 1997; Palinski *et al.*, 1995; Xu *et al.*, 1992).

1.7. ApoE knock-out: modelo de aterosclerosis

Una limitación del ratón como modelo de estudio de la aterosclerosis ha sido su resistencia metabólica a la dieta alta en grasa, que se traduce en la falta de inducción de hiperlipidemia (niveles altos de lípidos circulantes) aún utilizando dietas con altos porcentaje de lípidos. La modificación genética del gen *ApoE* ha solucionado ese problema. La glicoproteína APOE es sintetizada en el hígado, cerebro y otros tejidos tanto en humanos como en ratones y su peso molecular es de ~34 kD. Esta glicoproteína, es un componente estructural de todas las partículas de lipoproteínas (excepto LDL) y participa en la internalización de dichas lipoproteínas por el hígado (Mahley *et al.*, 1984). Una de sus funciones, es ser ligando del receptor de APOB y de APOE (LDL) y del receptor de remanentes de quilomicrones, de este modo es posible la absorción específica de partículas que contienen APOE por parte del hígado y por lo tanto el mantenimiento de niveles normales de lipoproteinemia (Zhang *et al.*, 1992).

En humanos, *APOE* es polimórfico y consta de tres alelos: *APOE-ε2*, *APOE-ε3* y *APOE-ε4*, de los cuales *APOE-ε3* es el más común. La variante *APOE-ε2* difiere de *APOE-ε3* porque tiene una cisteína en la posición 158 en lugar de una arginina. Esta sustitución de aminoácido que se da en el sitio de unión al receptor LDL, disminuye la capacidad de unión a menos del 2% comparado a *APOE-ε3*. La homocigocidad de *APOE-ε2* está asociada con la hiperlipoproteinemia tipo III, que se caracteriza por incremento de los niveles de triglicéridos y colesterol en plasma, además del desarrollo de aterosclerosis y nódulos amarillos cutáneos xantomatosos cargados de lípidos (revisado en Jawien *et al.*, 2004). En el cerebro, la homocigocidad en *APOE-ε4*, se ha asociado con el riesgo de

padecer Alzheimer (Corder *et al.*, 1993).

La función de la APOE en el recambio de las lipoproteínas séricas ha sugerido que su inactivación podría causar hiperlipidemia en ratones, logrando así un modelo de aterosclerosis. Confirmando esta hipótesis, en 1992, dos grupos de investigación generaron ratones deficientes en la producción de APOE a través de la inactivación del gen *ApoE* por recombinación homóloga (Piedrahita *et al.*, 1992; Plump *et al.*, 1992). Fue un gran avance, pues se pudo observar que, aunque la deficiencia en APOE no interfiere con el normal desarrollo de los individuos, éstos presentan hasta cinco veces más colesterol en plasma que sus contrapartes normales (**Tabla 1**). Además, se identificaron depósitos ricos en células espumosas en la aorta proximal, así como el desarrollo y progresión de lesiones ateroscleróticas desde edad temprana (10 semanas) (Plump *et al.*, 1992; Zhang, 1992).

Tabla 1. Niveles de colesterol y triglicéridos en plasma de ratones *ApoE*^{-/-} alimentados con una dieta Chow por 2-3 semanas. (Tomado de Jawieñ *et al.*, 2004)

Animales	Colesterol total en mg/dl ±SD	HDL en mg/dl ±SD	Triglicéridos en mg/dl ±SD
Normales	86 ± 20	73 ± 28	73 ± 36
Heterocigotos	88 ± 22	75 ± 18	102 ± 40
Homocigotos	434 ± 129	33 ± 15	123 ± 51

Dichas lesiones son espontáneas y se dan en todo el árbol arterial, incluso bajo una dieta normal aparecen estrías lipídicas que con el tiempo progresan hasta la capa fibrosa (Nakashim *et al.*, 1994). Actualmente, este modelo de ratón es considerado como uno de los mejores para el estudio de desórdenes de lipoproteínas y aterosclerosis (Jawien *et al.*, 2004; Meir and Leitersdorf, 2004). El modelo de ratón *ApoE*^{-/-} ha sido estudiado desde diferentes enfoques, todos ellos han ayudado a entender la aterosclerosis como un padecimiento multifactorial. De particular importancia para este trabajo, se sabe que durante las primeras semanas de vida de estos individuos, incluso antes de la aparición de lesiones ateroscleróticas visibles, ocurren cambios como la metilación global del DNA (Lund *et al.*, 2004).

Justificación

Es un hecho que el desarrollo de las ECVs ha aumentado durante las últimas décadas en la población humana. En su mayoría, el estudio de dichas enfermedades se concentra en el análisis genético, ambiental y la convergencia de ambos. Sin lugar a dudas, esto ha permitido la implementación de planes de prevención en cuanto al estilo de vida, así como de tratamientos médicos más específicos. Sin embargo, en la aterosclerosis, una de las principales causas de las ECVs, se conoce poco acerca de la incidencia de las VSDNs y sus implicaciones en la predicción del riesgo genético que conllevan. Aun sin ser heredables, las VSDN pueden ser importantes en el desarrollo de enfermedades no transmisibles y tener implicaciones acerca la causa, el mecanismo y las estrategias preventivas para la ECVs.

Objetivos

General

Determinar y cuantificar las variantes somáticas *de novo* en diferentes tejidos de ratones hiperlipidémicos *ApoE*^{-/-}, y evaluar la incidencia de las mismas en el perfil transcripcional de genes.

Específicos

- Determinar la frecuencia de variantes somáticas *de novo* en tejido de ratones hiperlipidémicos *ApoE*^{-/-} y silvestres.
- Evaluar la especificidad de las variantes somáticas *de novo* a nivel inter e intra-individual (mosaicismo).
- Evaluar el impacto de dichas variantes sobre la expresión génica por medio de secuenciación masiva de RNA.
- Buscar una relación transcripcional entre el modelo murino *ApoE*^{-/-} y aortas ateroscleróticas de humanos.

CAPITULO II

Materiales y métodos

2.1. Muestras biológicas

El manejo de los ratones utilizados en este estudio, se realizó en todo momento según los estándares de la norma oficial mexicana NOM-062-ZOO-1999. Cada grupo de individuos fue alimentado permanentemente (comida y agua *ad libitum*) con la dieta control Rodent Laboratory Chow 5001* (Purina) y mantenidos por cruces endogámicos. Todos los especímenes usados pertenecían a la cepa C57BL/6 ya sea silvestres o deficientes en la producción de APOE (Piedrahita *et al.*, 1992). Para generar hermanos *ApoE^{-/-}* y silvestres, se realizaron cruces entre heterocigotos y una vez genotipificada la progenie, se eligieron dos parejas de hermanos machos. Las dos parejas fueron designadas como: hermanos A y hermanos B con edades de 14 y 16 meses, respectivamente. Las disecciones y obtención de tejidos se llevaron a cabo teniendo en cuenta el mapa anatómico del ratón y las muestras colectadas fueron mantenidas en RNAlater (Ambion) a -80°C hasta su utilización. Antes de la extracción de ácidos nucleicos, cada tejido fue macerado con mortero y pistilo en presencia de nitrógeno líquido. Posteriormente, el tejido macerado fue separado en porciones equiparables tanto para la extracción de DNA como de RNA (**Figuras suplementarias 1 y 2**).

Los tejidos elegidos para el análisis fueron aorta, hígado y músculo esquelético. Esto se debió principalmente a las diferencias en cuanto a la acumulación de lípidos y al origen embrionario de cada uno. En la aorta aterosclerótica, la acumulación de lípidos es crónica y se da a lo largo de todo el proceso, mientras que en el hígado y en músculo es alta y baja respectivamente, esto último debido al dinamismo con que se procesan. De esta forma, si los lípidos al causar metilación del DNA inciden en las transiciones C a T, se podría detectar una tasa de variación genética diferencial. En cuanto al origen embrionario, los tres son de linajes embrionarios diferentes: la aorta del mesodermo de la placa lateral, el hígado del endodermo y el músculo esquelético del mesodermo paraxial. Si algunas de las variaciones son post-cigóticas tardías, será posible detectar mosaicismos.

2.2. Secuenciación de RNA (RNA-Seq)

La secuenciación de RNA se llevó a cabo en la Unidad Universitaria de secuenciación masiva de DNA del Instituto de Biotecnología de la UNAM. Este experimento se realizó con los tejidos de aorta, hígado y músculo de la pareja de hermanos A. El RNA extraído con TRIzol® (Thermo) fue analizado en cuanto a su integridad y concentración con el sistema Bioanalyzer 2100 (Agilent). Después de constatar que cada muestra contara con un número RIN >7, se elaboraron bibliotecas pareadas de Illumina (Kit TruSeq) que posteriormente fueron secuenciadas con la plataforma de Illumina Genome Analyzer II con una configuración pareada de 144 ciclos (lecturas de 72 pares de bases).

Las lecturas obtenidas mediante la secuenciación de RNA fueron alineadas al genoma de referencia C57BL/6J GRCm38.p2 con el alineador BWA. Los duplicados fueron eliminados usando la herramienta Picard MarkDuplicate (<http://picard.sourceforge.net>) y una vez filtrados, los alineamientos se usaron para hacer el conteo de lecturas por cada gen, así como el análisis de expresión diferencial (paquete NOIseq R con normalización TMM). Este programa determina que un gen está diferencialmente expresado si tiene una probabilidad arriba de 0.8, en nuestro caso, tomamos aquellos genes con valor $p \geq 0.9$. Los datos de secuenciación de RNA de aorta están disponibles en la base de datos BioProject del NCBI (Project accession PRJNA262445). Para la identificación de variantes, se utilizó el programa SAMtools.

2.3. Secuenciación de exoma completo (WES)

Se realizó secuenciación exónica de las parejas de hermanos A y B, cabe resaltar que el DNA de la pareja A fue obtenido de la misma porción de tejido analizado en el RNA-Seq. El DNA extraído de cada uno de los tejidos fue cuantificado por un método fluorométrico (Qubit®) y posteriormente se fragmentaron por ultrasonificación (COVARIS M220 Focused-ultrasonicator™) 200 ng de cada muestra. Para corroborar tanto la fragmentación, así como la concentración del DNA, se corrió un en gel en el sistema Agilent 2100 Bioanalyzer. La construcción de las bibliotecas y la captura de exoma se llevaron a cabo con el sistema Agilent SureSelect Mouse All Exon bajo las indicaciones del fabricante. Cada biblioteca fue hibridada al exoma de ratón durante 48 horas y posteriormente la fracción exónica aislada

(sistema estreptavidina-biotina) se sometió una ronda de amplificación por PCR. La plataforma de secuenciación usada fue Illumina HiSeq2000 y se llevó a cabo en la Unidad Universitaria de Secuenciación Masiva de DNA del Instituto de Biotecnología de la UNAM. Las lecturas pareadas obtenidas de la secuenciación de exoma fueron alineadas al genoma de referencia C57BL/6J GRCm38.p4, por medio del algoritmo BWA (Burrows-Wheeler Aligner). Los alineamientos generados en formato BAM fueron sometidos a las herramientas de Picard MarkDuplicate (<http://picard.sourceforge.net>) para eliminar duplicados ópticos y de PCR.

2.4. Llamado de variantes

El llamado de variantes en el RNA-seq fue realizado con el programa SAMtools (Li *et al.*, 2009) y en el WES con FreeBayes (versión v0.9.20-16-g3e35e72) (Garrison and Marth, 2012). En ambos casos, los archivos VCF obtenidos luego del llamado de variantes, fueron procesados con el paquete VCFtools con el objetivo de determinar parámetros estadísticos, de cobertura, así como de realizar comparaciones entre muestras. Para la visualización de las variantes, así como de las lecturas de cada muestra, se usó el programa IGV (Integrative Genomics Viewer).

2.5. Anotación funcional de las variantes detectadas

Para filtrar y anotar las variantes se usó la herramienta Variant Effect Prediction (VEP) (McLaren *et al.*, 2010), esta sirve para anotar variantes de un solo nucleótido, así como inserciones/deleciones. VEP determina el efecto de las variantes (SNPs, inserciones, deleciones, CNVs, o variantes estructurales) sobre los genes, los transcritos, y la secuencia de la proteína, así como sobre regiones reguladoras. El funcionamiento de este programa se basa dos herramientas: SIFT y PolyPhen, para dar información sobre los cambios a nivel de proteína y por lo tanto, sobre su posible efecto. SIFT predice si la sustitución de un aminoácido afecta a la función de la proteína basándose en la homología de secuencia y las propiedades físicas de los aminoácidos. PolyPhen es una herramienta que predice el posible impacto de una sustitución de aminoácidos en la estructura y función de una proteína humana utilizando consideraciones físicas y comparativas directas.

2.6. Análisis de enriquecimiento de vías de señalización

Para la agrupación funcional de genes ya sean expresados diferencialmente o con variantes genéticas, se usó el programa DAVID (Database for Annotation, Visualization and Integrated Discovery) Bioinformatic Resources 6.7(Huang *et al.*, 2009a, 2009b). Esta herramienta *on line*, permite interpretar y relacionar de manera funcional listas de genes, que en nuestro caso corresponden a genes diferencialmente expresados y a genes con variantes genéticas. Este programa hace uso de diferentes herramientas de anotación como el *Gene Functional Classification*, el Functional Annotation, el Gene ID Conversion, el Gene Name Batch Viewer y el NIAID Pathogen Genome Browser para predecir el significado biológico de las listas sometidas. El programa permite identificar términos del GO (Gene Ontology) como: función molecular (Molecular Function o MF), proceso biológico (Biological Process o BP) y componente celular (compound cellular o CC). De igual forma, se pueden visualizar vías de señalización en los mapas de vías de BioCarta y del KEGG (Kyoto Encyclopedia of Genes and Genomes), así como predecir dominios y proteínas por medio del INTERPRO (protein sequence analysis & classification).

CAPITULO III

Resultados

A continuación, los resultados descritos se muestran en tres bloques principales: los dos primeros corresponden al análisis del perfil de mutaciones genéticas y transcripcional en ratones *ApoE*^{-/-} y el último a la correlación entre la metilación del DNA y la expresión diferencial de genes en aortas de humanos y ratones con aterosclerosis.

3.1. Variantes somáticas *de novo* en ratones con aterosclerosis

3.1.2. Obtención de ratones ateroscleróticos y normales

Para determinar si la aterosclerosis estaba asociada con un incremento en variaciones somáticas, se analizaron dos parejas de hermanos de la cepa C57BL/6 ya sea silvestres o knock-out para ApoE (Apolipoproteína E). Las parejas de hermanos –denominadas parejas A y B– fueron sacrificadas a los 14 y 16 meses, respectivamente, y se colectaron los tejidos de aorta, hígado y músculo esquelético. En esta etapa del desarrollo, todos los individuos *ApoE*^{-/-}, a diferencia de los silvestres, exhibieron un ateroma avanzado (**Figura 4**). La placa aterosclerótica se vio extendía a lo largo de diferentes puntos de la aorta, tales como la raíz aortica, la aorta torácica, la aorta abdominal e incluso hasta la bifurcación aórtica. Cabe subrayar que la detección visual del fenotipo aterosclerótico es una importante validación de la asignación del genotipo realizada previamente por genotipación molecular.

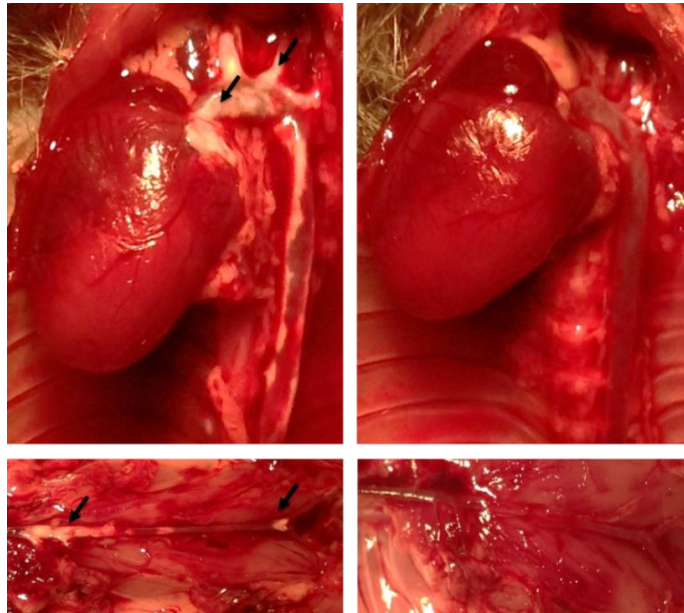


Figura 4. Placa aterosclerótica en la aorta de un ratón $ApoE^{-/-}$

Se muestran el corazón y la aorta torácica de una pareja de hermanos machos de 16 meses. Nótese la lesión prominente de la raíz aórtica y aorta ascendente (indicada por flechas en los paneles superior e inferior, respectivamente) en el ratón $ApoE^{-/-}$ (izquierda) comparado con su ausencia en el hermano silvestre (derecha).

3.1.3. Determinación de variantes genéticas por medio de secuenciación del exoma (WES) de ratones consanguíneos

Se realizó WES (Whole Exome Sequencing) de un total de 10 muestras; de las cuales seis fueron de la pareja A (aorta, hígado y músculo) y cuatro de la pareja B (aorta e hígado) (**Tabla 2**). Aunque los datos obtenidos con relación al número de lecturas totales fueron similares en cada muestra, hubo mayor número de éstas en la pareja A. En promedio el 91 % de las lecturas tenían un valor de Phred arriba de 30 un. El mapeo contra el genoma de referencia fue en promedio del 99.8 %, con valores muy similares entre hermanos y con una cobertura promedio de 76X. Este último valor es considerablemente bueno si se tiene en cuenta que incluso con coberturas de 20X se han podido determinar variantes en ratón con gran fiabilidad (Fairfield *et al.*, 2011).

Tabla 2. Datos generales de la WES

Pareja de hermanos	Tejido	Individuo	Lecturas totales	Lecturas alineadas	% lecturas alineadas	pares de bases secuenciadas	Cobertura del exoma (X)
A	Aorta	<i>ApoE</i> ^{-/-}	39576371	39492189	99.79	3997213471	80
		Silvestre	39835320	39734449	99.75	4023367320	80
	Hígado	<i>ApoE</i> ^{-/-}	41027456	40937205	99.78	4143773056	83
		Silvestre	39131822	39053187	99.80	3952314022	79
	Músculo	<i>ApoE</i> ^{-/-}	37819057	37739732	99.79	3819724757	76
		Silvestre	35848781	35771641	99.78	3620726881	72
B	Aorta	<i>ApoE</i> ^{-/-}	36944946	36849035	99.74	3731439546	75
		Silvestre	37599617	37507547	99.76	3797561317	76
	Hígado	<i>ApoE</i> ^{-/-}	36838401	36734655	99.72	3720678501	74
		Silvestre	33843416	33743731	99.71	3418185016	68

Como producto del alineamiento de secuencias se obtuvieron archivos tipo “.bam”, los cuales fueron usados para realizar el llamado de variantes de WES (v-WES) mediante el algoritmo FreeBayes. Una vez obtenidas las v-WES, éstas fueron filtradas con la herramienta VCFtools de acuerdo a aquellas que presentaban 10 o más lecturas. En promedio, se encontraron ~7000 v-WES por tejido con un valor mínimo de 5381 para el hígado del hermano silvestre B y un máximo de 8436 para el mismo tejido del hermano *ApoE*^{-/-} A (**Tabla 3**).

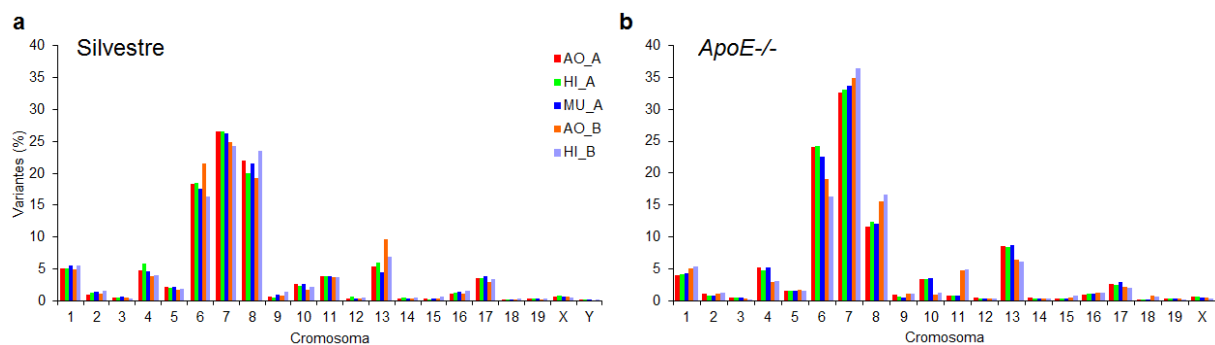
Cabe resaltar que la profundidad promedio, es decir, el número de lecturas en cada posición en donde se halló una variante, estuvo por arriba de 100X en todas las muestras. De igual forma, se realizó un segundo filtrado más estricto de las v-WES específicas de los individuos *ApoE*^{-/-} incluyendo solo aquellas que presentaban una cobertura mínima de 60 %, una calidad de base de 30 y de mapeo de 20 en la escala de Phred. Dado que bajo este filtro se conservaron más del 60 % de las variantes iniciales, se decidió trabajar con el total de las v-WES halladas bajo el primer filtro.

Tabla 3. Número de variantes y su respectiva cobertura por tejido

Pareja de hermanos	Tejido	Individuo	Variantes totales	Profundidad promedio
A	Aorta	<i>ApoE</i> ^{-/-}	8177	125.13
		Silvestre	7110	116.43
	Hígado	<i>ApoE</i> ^{-/-}	8436	115.74
		Silvestre	6319	127.98
	Músculo	<i>ApoE</i> ^{-/-}	7859	112.76
		Silvestre	5947	107.46
B	Aorta	<i>ApoE</i> ^{-/-}	7003	110.48
		Silvestre	7207	112.18
	Hígado	<i>ApoE</i> ^{-/-}	6766	109.44
		Silvestre	5381	113.28

3.1.3.1. Distribución cromosómica diferencial de las variantes

Al cuantificar el número de v-WES por cromosoma, se observó que la mayoría de ellas mapeaban preferentemente en los cromosomas 6, 7 y 8, tanto en los tejidos de *ApoE*^{-/-} como silvestres (χ^2 , $p=6 \times 10^{-177}$ y $p=1.3 \times 10^{-113}$, respectivamente; en comparación con una distribución al azar a lo largo del genoma; **Figura 5**).

**Figura 5.** Cuantificación de variantes totales en el DNA por cromosoma

Con un análisis visual de los datos generados con el programa IGV, se pudo ver la distribución de las v-WES a lo largo los cromosomas en cada muestra (**Figura 6**). Tanto en tejidos silvestres como en *ApoE*^{-/-}, se pudo observar que dicha distribución no estaba

relacionada ni con el tamaño de los cromosomas (el 1 es el más grande y el Y el más pequeño), ni con la densidad génica del genoma de referencia (**Figura 6**).

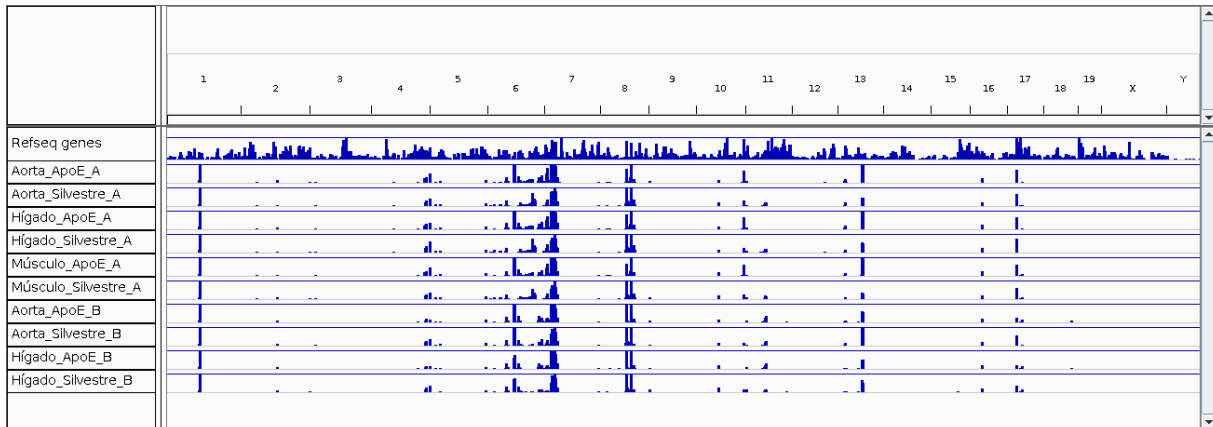


Figura 6. Visualización general de variantes en los cromosomas

Se observan las variantes de cada muestra con respecto a cada cromosoma. La primera fila muestra la distribución de los genes en el genoma de referencia.

Se hizo también una comparación entre la distribución de v-WES de los cromosomas 1 (el más grande) y 6 (uno de los que tiene mayor número de v-WES). En el cromosoma 1 se observó una distribución más homogénea de variantes comparado con el cromosoma 6 (**Figura 7a**). En este último, gran cantidad de v-WES, se concentraron en sitios específicos, como por ejemplo en la región incluida entre las bandas qC3 y qF1. Además, se observaron regiones - como la comprendida entre las bandas qB3 y qC3 - en donde la distribución de v-WES fue diferencial entre tejidos *ApoE*^{-/-} y silvestres (**Figura 7b**, nótese como en estos últimos dicha región está desprovista de variantes).

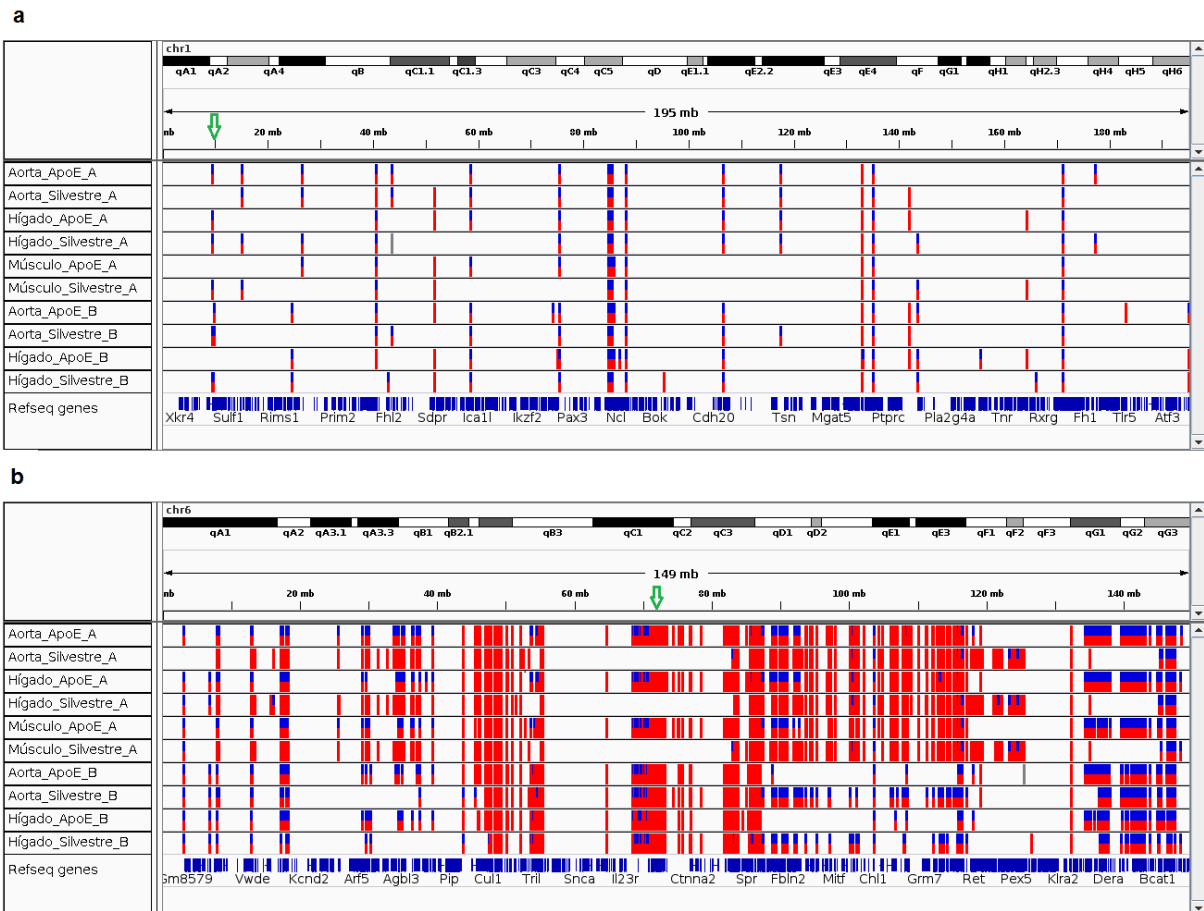


Figura 7. Visualización de las vWES del cromosoma 1 y 6

Se muestran las v-WES encontradas en los cromosomas 1 y 6 de cada muestra. La proporción de variantes difiere en algunos puntos de los cromosomas. El color rojo describe variantes homocigotas, mientras que la combinación de rojo y azul se refiere a variantes heterocigotas. En la última fila, en color azul, se muestran los genes del genoma de referencia. Las flechas verdes indican ejemplos de variantes diferenciales entre tejidos *ApoE*^{-/-} y silvestres.

Sin embargo, la falta de v-WES en esta última región no se debía a una ausencia de lecturas. Por ejemplo, en la aorta de la pareja A se observó que la cobertura – pero no el llamado de v-WES – fue similar en el individuo *ApoE*^{-/-} y el silvestre (**Figure 8a**); un acercamiento a la región qB3-qC3 confirmó la presencia de lecturas en dicha región en ratones silvestres y *ApoE*^{-/-} (**Figura 8b**).

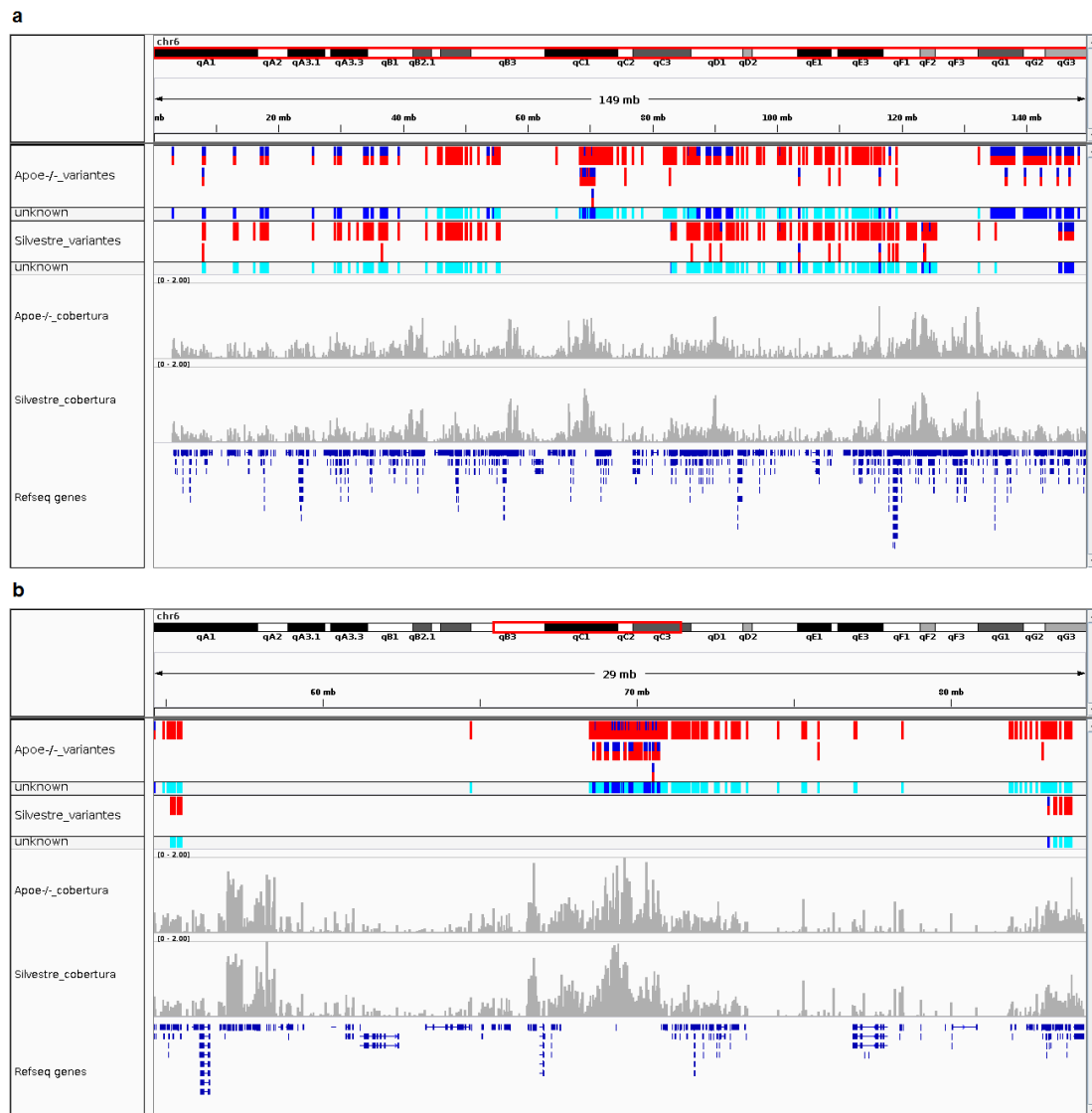


Figura 8. Visualización de v-WES y lecturas totales del cromosoma 6

Cromosoma 6 completo (a), así como la región de qB3-qC3 del mismo (b). El color rojo describe variantes homocigotas, mientras que la combinación de rojo y azul se refiere a variantes heterocigotas. Las barras en gris indican la cobertura. En azul están los genes de referencia. Nótese las regiones genómicas desprovistas de variantes, pero con lecturas.

3.1.3.2. Variantes somáticas *de novo* de ratones *ApoE*^{-/-} y silvestres

En conjunto, el análisis anterior sugería la existencia de variantes comunes y específicas entre los ratones silvestres y *ApoE*^{-/-}, que además se distribuían de manera diferencial entre

los cromosomas. Debido al hecho de que el DNA de los ratones progenitores no fue secuenciado, no podremos excluir que las variantes comunes entre hermanos hayan sido transmitidas a través de la línea germinal. Sin embargo, las v-WES únicas a cada individuo *ApoE*^{-/-} y silvestre (v-WES/*ApoE*^{-/-} y v-WES/WT, respectivamente), podrían representar variantes somáticas *de novo*.

Para identificar las v-WES/*ApoE*^{-/-} y v-WES/WT, se compararon las variantes entre cada pareja de hermanos mediante el uso del programa VCFtools (**Tabla 4**). En ambas parejas, la mayoría de las v-WES fueron compartidas entre los ratones silvestres y *ApoE*^{-/-} (65 y 69 % en A y B, respectivamente), pero sin diferencias significativas entre ellos (χ^2 , $p=0.67$). Con relación a las variantes únicas, se observó un porcentaje mayor en los ratones *ApoE*^{-/-} comparado con los silvestres en todos los tejidos (χ^2 , $p=0.04$), con la excepción de la aorta de la pareja B (χ^2 , $p=0.15$).

Tabla 4. Generalidades de las v-WES/*ApoE*^{-/-} y v-WES/WT

Pareja de hermanos	Tejido	Individuo	% del total	Genes	Ti/Tv	% dbSNPs
A	Aorta	<i>ApoE</i> ^{-/-}	39	502	1.8	73
		Silvestre	30	377	2.1	80
	Hígado	<i>ApoE</i> ^{-/-}	45	606	1.9	78
		Silvestre	27	300	2.1	73
	Músculo	<i>ApoE</i> ^{-/-}	44	549	1.7	75
		Silvestre	26	284	2.4	79
B	Aorta	<i>ApoE</i> ^{-/-}	29	462	1.9	73
		Silvestre	31	362	1.8	70
	Hígado	<i>ApoE</i> ^{-/-}	40	504	2.0	77
		Silvestre	24	279	1.6	66

Las v-WES/*ApoE*^{-/-} y v-WES/WT representaban en promedio 525 y 322 genes con un promedio de 3 y 4 variantes por gen, respectivamente (**Tabla 4**). De hecho, más de la mitad de los genes tuvieron una sola variante (52 % para tejidos *ApoE*^{-/-} y 60 % para silvestres) y solo el 2% tenían 10 variantes o más. El número de variantes por gen no estuvo correlacionado con el tamaño de los genes (ej. aorta *ApoE*^{-/-}, $r=0.05$ $p=0.2$). Un ejemplo de ello son las inmunoglobulinas kappa variable que siendo genes con un promedio de 500pb, estuvieron en el grupo de los genes con más v-WES (**Tablas suplementarias 1 y 2**). En

cada tejido más del 85% de las v-WES/*ApoE*^{-/-} y v-WES/WT correspondían a VSN y la mayoría de ellas fueron variaciones de transición (C-T o A-G), lo cual se refleja en el radio de transiciones entre transversiones (Ti/TV). De manera sorprendente, la mayoría de los v-WES/*ApoE*^{-/-} (75%) y v-WES/WT (76%) representaban VSN ya reportadas (dbSNP = Single Nucleotide Polymorphism Database) (**Tabla 4**).

Al clasificar las v-WES/*ApoE*^{-/-} y v-WES/WT de acuerdo a su consecuencia biológica, se encontró que los cambios más prevalentes fueron sinónimos y de pérdida de función, mientras que en menor medida estaban los desplazamientos del marco de lectura, entre otros (**Figura suplementaria 3**). Aunque hubo más prevalencia de cambios sinónimos en los tejidos silvestres, y de cambios de pérdida de sentido en los tejidos *ApoE*^{-/-}, esta no fue significativa (χ^2 , $p=0.4$).

Las v-WES/*ApoE*^{-/-} y v-WES/WT de cada pareja se localizaron principalmente en los cromosomas 6, 7, 8, 11 y 13 (**Figura suplementaria 4, 4a-4b y 4c-4d**, respectivamente), y de manera diferencial en los cromosomas 7, 8 y 11. Esta distribución cromosómica no tuvo ninguna relación con el tamaño de los mismos en las v-WES/*ApoE*^{-/-} (χ^2 , $p=2 \times 10^{-4}$), mientras que en las v-WES/WT no hubo diferencias significativas (χ^2 , $p=0.998$). Lo anterior también se evidenció en una distribución cromosómica casi idéntica de los genes que fueron mapeados por las v-WES/*ApoE*^{-/-} y v-WES/WT (**Figura 9a-9b y 9c-9d, respectivamente**).

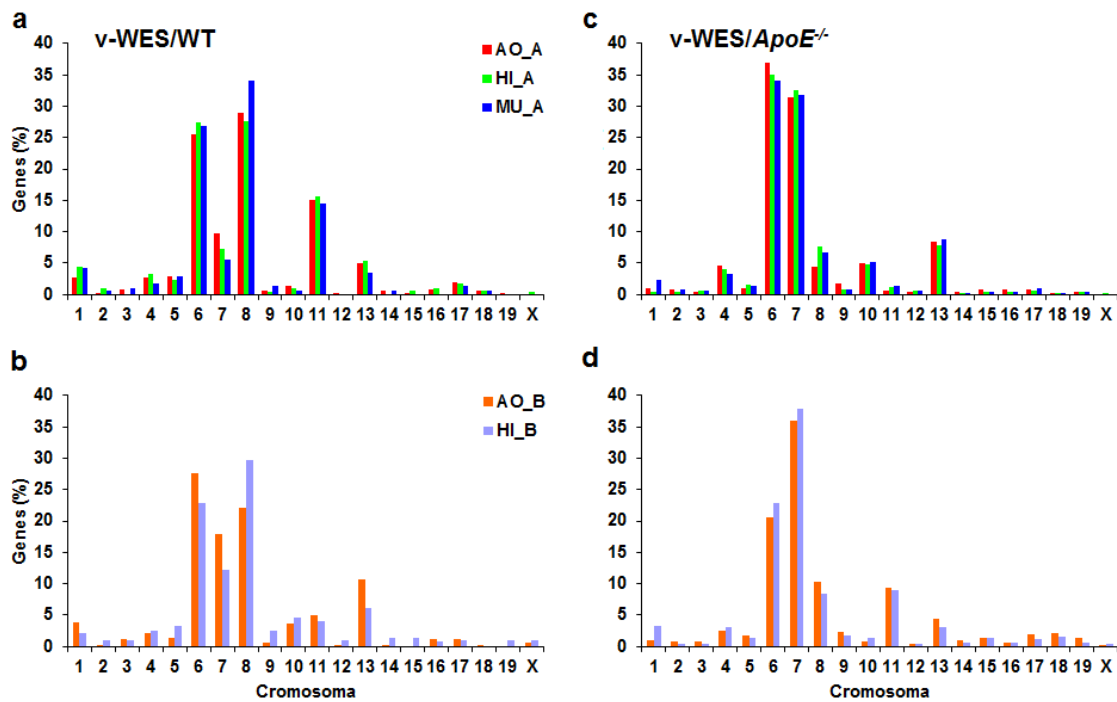


Figura 9. Porcentaje de genes por cromosoma.

Se observan el porcentaje de genes mapeados por las v-WES/WT (a, b) y las v-WES/*ApoE*^{-/-} (c, d) en cada cromosoma, en las parejas A (a, c) y B (b, d).

Hasta aquí, los datos indicaban que los dos individuos de *ApoE*^{-/-} tenían un mayor número de variantes somáticas con respecto a los ratones silvestres en la mayoría de los tejidos. Aunque se observaron diferencias entre los individuos *ApoE*^{-/-} con relación al número de variantes, ambos mostraron un enriquecimiento de variantes en cromosomas específicos.

3.1.3.3. Análisis de las variantes compartidas entre individuos *ApoE*^{-/-}

Se procedió a identificar las v-WES/*ApoE*^{-/-} que estaban conservadas entre las parejas A y B en la aorta e hígado. No se realizó este análisis en músculo ya que se contaba únicamente con WES de la pareja A. Se encontraron un total de 618 y 999 variantes compartidas (v-COMP/*ApoE*^{-/-}) en la aorta e hígado, respectivamente (**Figura 10**). En la aorta esto corresponde al 19 y 30 % de v-WES/*ApoE*^{-/-} del hermano A y B, respectivamente; en el hígado los porcentajes de los respectivos hermanos fueron 26 y 37%.

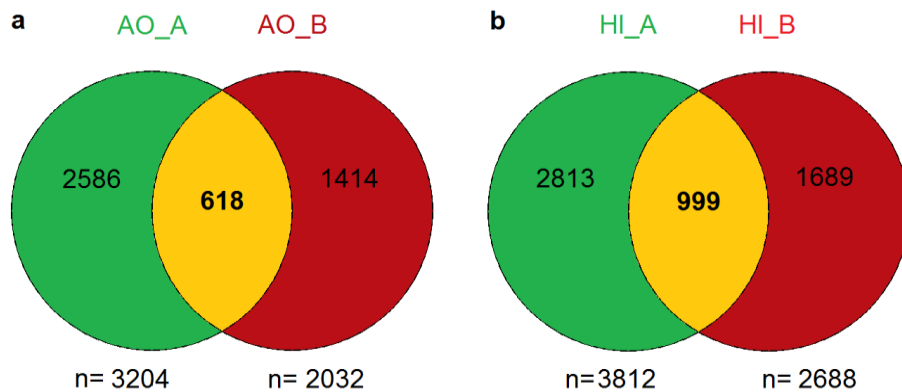


Figura 10. v-COMP de aorta e hígado

Número total de variantes específicas de *ApoE*^{-/-} en aorta (verde) e hígado (rojo) del individuo A (AO_A y HI_A, respectivamente) y del B (AO_B y HI_B, respectivamente).

La mayoría de las v-COMP/*ApoE*^{-/-} eran específicas de aorta o de hígado (55 y 72%, respectivamente) y en ambos casos se localizaban preferentemente en los cromosomas 7 (62 y 45 %) y 6 (18 y 28 %, respectivamente). En cuanto a las v-COMP/*ApoE*^{-/-} comunes entre tejidos, el 81 % estaban en el cromosoma 7.

Las v-COMP específicas de tejido estaban localizadas en 270 y 516 genes en aorta e hígado, respectivamente. Como se observó para las v-WES/*ApoE*^{-/-}, la mayoría de las v-COMP/*ApoE*^{-/-} fueron dbSNPs. Respecto a sus consecuencias biológicas, la mayoría representaban cambios sinónimos (63 y 56 %, en aorta e hígado, respectivamente) y cambios sin sentido (31 y 38 %, respectivamente).

Para conocer si las v-COMP/*ApoE*^{-/-} conllevaban a un impacto funcional, se usó la herramienta DAVID Bioinformatic usando como input las v-COMP/*ApoE*^{-/-} tejido específicos o comunes. Para las variantes tejido-específicas, se encontró un enriquecimiento en términos del GO en aorta y en hígado relacionadas con la especificidad funcional del tejido (**Tablas 5 y 6, respectivamente**). En la aorta y el hígado, los términos del GO con mayor significancia en la categoría BP (= Biological Process) y CC (=celular function) fueron idénticos; es decir embarazo ($p=2.8 \times 10^{-10}$ y $p=3.7 \times 10^{-9}$, respectivamente) y las partículas de lipoproteínas ricas en triglicéridos ($p=0.004$ y $p=0.004$, respectivamente) y VLDL ($p=0.004$ y 0.004 , respectivamente), (**Tablas 5 y 6, respectivamente**). En cuanto a los términos de MF (=molecular function) se encontró mayor significancia en unión a ATP ($p=0.0026$) y a metales de transición ($p=0.001$), en la aorta y hígado, respectivamente.

Tabla 5. Análisis de enriquecimiento GO y KEGG de las v-COMP específicas de aorta

Categoría	Término	Valor p	FDR	Genes
	bile acid metabolic process	4E-04	0.550	SLCO1C1, GM6614, SLCO1A5, SLC27A5
	positive regulation of cAMP biosynthetic process	0.006	8.554	PTGIR, ADCY7, GSK3A, GIPR
	antigen processing and presentation	0.007	10.538	MILL1, MILL2, RELB, HFE
	axon guidance	0.018	23.824	UNC5A, SEMA4F, DPYSL2, PTPRO, VASP
BP	sodium-independent organic anion transport	0.018	24.751	SLCO1C1, GM6614, SLCO1A5
	protein phosphorylation	0.031	37.877	PRKD2, STK32C, RET, GSK3A, MYLK3, GUCY2C, MARK4, CDK20, DMPK, ERCC2
	nucleotide-excision repair	0.031	38.298	ERCC1, FANCC, ERCC2
	RNA secondary structure unwinding	0.044	50.118	DDX39, DDX47, GM8994
	learning or memory	0.048	52.934	THRA, GRIN2B, MAN2B1
	neuron projection	0.009	10.996	ANKRD27, CTSL, GRIN2B, KLHL14, CALM3, KLC3, DPYSL2, MARK4, PPP5C
	cytosol	0.022	23.770	NLRP5, THRA, C5AR1, NFKBIB, MYLK3, RELB, HK2, DPYSL2, INMT, SIRT2, DMPK, PTGIR, HEBP1, GIPR, LRP6, GEMIN7, FANCC, PPP5C
CC	microtubule	0.026	27.698	EML2, SLC8A2, GSK3A, KLC3, DPYSL2, DCTN1, SIRT2
	mitotic spindle	0.040	39.748	EML2, CAPG, SIRT2
	apical part of cell	0.048	45.232	CTSL, C5AR2, C5AR1, HFE
	bile acid transmembrane transporter activity	0.002	3.257	SLCO1C1, GM6614, SLCO1A5
	microtubule binding	0.012	14.660	EML2, KLC3, DPYSL2, MARK4, CASC1, PPP5C
MF	ATP-dependent RNA helicase activity	0.012	15.109	DDX39, DDX47, GM8994, DHX34
	helicase activity	0.014	17.574	DDX39, RECQL, DDX47, DHX34, ERCC2
	C5a anaphylatoxin receptor activity	0.015	18.458	C5AR2, C5AR1
	protein kinase activity	0.016	19.014	PRKD2, STK32C, RET, GSK3A, MYLK3, GUCY2C, MARK4, CDK20, DMPK, ERCC2

	sodium-independent organic anion transmembrane transporter activity	0.018	21.281	SLCO1C1, GM6614, SLCO1A5
	protein domain specific binding	0.023	26.641	THRA, GUSB, CAPG, NPHS1, CALM3, HIST1H4C, ERCC1
	transcription factor activity, sequence-specific DNA binding	0.025	28.828	ZFP551, THRA, HOXA3, NR1D1, HOXA4, ZFP772, ZFP954, RELB, HIF3A, BCL3, FOSB, ZFP110, CREBL2
	protein binding	0.034	36.853	MILL1, PVR, THRA, MILL2, NFKBIB, NAPA, KLC3, ZFP110, DMPK, GRIN2B, RINL, NR1D1, ZFP296, CEACAM16, BCL3, NOVA2, GEMIN7, ERCC1, CREBL2, RET, NLRP5, IMMT, RELB, DPYSL2, PTPRO, PPP1R13L, SIRT2, VASP, DCTN1, ANKRD27, CD3EAP, DOK1, CTSL, CLGN, GSK3A, SEMA4F, KLHL14, NPHS1, LRP6, HIF3A, LRMP, IGFL3, CDK20, PPP5C
	thyroid hormone transmembrane transporter activity	0.037	39.962	SLCO1C1, SLCO1A5
	Vascular smooth muscle contraction	0.013	13.887	PTGIR, ADCY7, MYLK3, CALM3, PLA2G4C
KEGG	Platelet activation	0.013	14.587	PTGIR, ADCY7, MYLK3, PLA2G4C, VASP
	cGMP-PKG signaling pathway	0.031	30.972	SLC8A2, ADCY7, MYLK3, CALM3, VASP

Tabla 6. Análisis de enriquecimiento GO y KEGG de las v-COMP específicas de hígado

Categoría	Término	Valor P	FDR	Genes
	xenobiotic glucuronidation	2.5E-14	4E-11	UGT1A10, UGT1A9, UGT1A8, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, UGT1A1
	flavonoid glucuronidation	4.7E-10	8E-07	UGT1A10, UGT1A9, UGT1A8, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, UGT1A1
	flavonoid biosynthetic process	4.7E-10	8E-07	UGT1A10, UGT1A9, UGT1A8, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, UGT1A1
	regulation of transcription, DNA-templated	0.000	0.69	ZFP541, FOXA3, CNOT2, ZKSCAN3, ZFP786, ZFP93, ATOH8, ZFP273, ATF7IP, ZFP329, RELB, PKN1, SIX5, IRF2BP1, MXD3, JUNB, PCGF1, PAF1, ZFP369, ZSCAN12, NSD1, HOXA13, SOX5, ZFP112, ZFP457, DPF1, PRDM13, NPAS1, MEIS3, ZFP956, POU2F2, ZSCAN26, ETV2, KDM3A, THAP1, CC2D1A, ERCC2, ZFP58, TRIM28, ZFP775, ZFP708, SMYD1, MYPOP, PPP1R13L, SIRT2, SALL1, ZFP772, THRAP3, HIF3A, RFX1, SP6
	transcription, DNA-templated	0.001	1.36	ZFP541, HOXA13, FOXA3, SOX5, CNOT2, ZKSCAN3, ZFP112, ZFP786, DPF1, PRDM13, NPAS1, ZFP93, POU2F2, ZSCAN26, THAP1, KDM3A, CC2D1A, ATOH8, ERCC2, ATF7IP, ZFP329, ZFP58, TRIM28, RELB, POLR1A, ZFP775, PKN1, SMYD1, SIX5, IRF2BP1, MYPOP, PPP1R13L, SIRT2, PCGF1, MXD3, JUNB, SALL1, MAPK3, THRAP3, RFX1, HIF3A, SP6, PAF1, ZFP369, NSD1, ZSCAN12
BP	cellular glucuronidation	0.003	5.37	UGT1A2, UGT1A5, UGT1A1
	negative regulation of transcription from RNA polymerase II promoter	0.013	19.77	ATF7IP, RELB, TRIM28, CNOT2, IRF2BP1, MYPOP, ZKSCAN3, PPP1R13L, ZFP128, JUNB, SIRT2, PRDM13, NPAS1, ZFP568, RSL1, SALL1, ZFP296, CC2D1A, PAF1, NSD1
	peptidyl-serine phosphorylation	0.016	22.75	PRKD2, MAST1, MAPK3, PKN1, PRKACA, EIF2AK3, DMPK
	negative regulation of transcription, DNA-templated	0.017	23.83	ATF7IP, RELB, TRIM28, SOX5, SIX5, SMYD1, ZKSCAN3, ZFP457, MXD3, SIRT2, CDKN1B, RSL1, SALL1, LOXL3, ATOH8
	retina vasculature development in camera-type eye	0.027	36.45	MIR27A, MIR24-2, MIR23A
	bile acid metabolic process	0.034	42.91	GM6614, SLCO1A5, SLC27A5
	protein phosphorylation	0.036	44.66	MYLK3, TRIM28, AURKC, PKN1, GUCY2C, MARK4, DMPK, PRKD2, MAST1, AAK1, MAPK3, GRK6, PRKACA, EIF2AK3, CDK20, ERCC2
	behavioral response to pain	0.037	46.08	GRIN2B, TACR1, CACNA1A
	epithelial tube branching involved in lung morphogenesis	0.037	46.08	CTSL, KRAS, HHIP
CC	protein complex	0.004	5.11	NLRP5, SUCLG1, STRN4, PKN1, DCTN1, MUC4, PRELID1, NOD2, CDKN1B, UGT1A6B, UGT1A6A, MAPK3, NPHS1, RYR1, PRKACA, FABP1, CACNA1A, SLC27A5, PPP5C

	cytosol	0.017	20.10	CHMP2A, TACR1, NFKBIB, HK2, SNRPD2, GIPC1, PRDX2, DMPK, NOD2, RAC3, GYS2, PSMD8, ARHGDI, NLRP5, OSBPL3, CMAS, MYLK3, RELB, PDE3A, SIRT2, XPNPEP1, CDKN1B, PLEKHA5, MAPK3, GIPR, RYR1, GRK6, SP6, FABP1, CACNA1A, PPP5C, REPIN1
	endoplasmic reticulum	0.029	32.54	PDIA4, RTN2, DMPK, UGT1A9, RNF103, UGT1A7C, UGT1A2, APOE, UGT1A6B, UGT1A6A, UGT1A5, HSD17B3, PTDSS1, FKRP, AIFM3, ERP27, CYP2B10, GUCY2C, UGT1A1, RNF170, MUC4, UGT1A10, CLGN, TMEM8B, SEMA4F, KLHL14, LRMP, EIF2AK3, SLC27A5
	nuclear membrane	0.034	37.01	HPN, OSBPL3, CDC14B, SCRNI, TNPO2, DMPK, REPIN1
	membrane	0.048	47.64	CHMP3, ADCY7, QPCTL, CNOT2, DMPK, GNG8, NKPD1, GRIN2B, UNC5A, APOE, DHX34, RAB24, PRKACA, CMAS, LRP1B, STRN4, CYP2B10, PTPRO, DCTN1, RNF225, DDX39, UGT1A10, TELO2, RYR1, PAF1, EIF2AK3, PPP5C, CHMP2A, SCN1B, ST8SIA1, CCDC91, GIPC1, FXYD5, FXYD7, PRR7, KRAS, RAC3, PLXND1, FKRP, ARHGDI, OSBPL3, TOMM40, GUCY2C, RPS5, MUC4, ATP6V0E2, PLEKHA5, CLGN, EPS8, TMEM8B, CD207, GRK6, CLCN4, CACNA1A, TM7SF3, CACHD1, IL27RA, TACR1, AP2S1, BCAM, RTN2, ART4, SLC1A5, NOD2, UGT1A7C, RNF103, UGT1A6A, SMC03, CNTNAP2, HHIP, LOXL3, KIRREL2, PTDSS1, SYMPK, PTGER1, TRPM8, ADGRE5, PKN1, UGT1A1, VASP, RNF170, PRKD2, RECQL, MAST1, ARHGAP33, SEMA4F, KLHL14, GIPR, LRMP, SLC27A5, GRAMD1A, NKD1, HK2, GPRC5A, APLP1, UGT1A9, UGT1A2, RASGRP4, PALM3, ETNK1, KDM3A, CC2D1A, CATSPERG1, RNF121, ELMOD2, GPR107, HPN, SLC8A2, IMMT, PDE3A, GDPD3, SLC01A5, SIRT2, CLPTM1, ABCC9, NPHS1, ADGRL1, SPATA31
	glucuronosyltransferase activity	1.4E-08	2.1E-05	UGT1A10, UGT1A9, UGT1A8, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, UGT1A1
	transferase activity, transferring hexosyl groups	0.001	1.52	UGT1A9, UGT1A7C, UGT1A2, UGT1A6A, UGT1A1
	transcription factor activity, sequence-specific DNA binding	0.003	3.76	HOXA13, FOXA3, RELB, TRIM28, ZFP775, SOX5, ZFP708, MYPOP, ZKSCAN3, ZFP457, ZFP128, JUNB, ZFP568, ZSCAN22, ZFP956, ZFP772, POU2F2, ETV2, RFX1, ZSCAN26, HIF3A, KDM3A, ATOH8, ZFP273, ZFP369, ZSCAN12
MF	protein binding	0.011	14.77	PSG17, RARRES2, IL27RA, PRDX2, DMPK, HOOK2, NOD2, GRIN2B, APOE, CEACAM16, CNTNAP2, RAB24, PRKACA, HHIP, KIRREL2, TUBB3, NLRP5, TRPM8, CEP89, LRP1B, RELB, POLR1A, PKN1, KRT13, PTPRO, ZFP128, VASP, JUNB, DCTN1, MXD3, CD3EAP, CTSL, DOK1, MAST1, ARHGAP33, SEMA4F, DACT3, KLHL14, TELO2, MAPK3, RYR1, LRMP, PAF1, ZFP369, NSD1, EIF2AK3, PPP5C, MILL1, CHMP2A, NKD1, MILL2, SCN1B, HOXA13, NFKBIB, SOX5, WBP11, KLC3, GIPC1, GPRC5A, APLP1, PRDM13, ZFP568, KRAS, RAC3, ZFP296, KDM3A, PLXND1, TNPO2, ERCC1, GPR107, CYP2G1, LGALS4, IMMT, TRIM28, GARS, SMYD1, MYPOP, PPP1R13L, SIRT2, CDKN1B, CLGN, EPS8, SALL1, NPHS1, THRAP3, HIF3A, GRK6, FCGBP, CDK20, CACNA1A
	nucleic acid binding	0.015	20.02	ZFP541, ZFP329, ZFP58, ZFP775, ZFP112, ZKSCAN3, RNASEH2A, ZFP786, DPF1, PRDM13, DDX39, RECQL, ZFP93, SALL1, ZSCAN26, DHX34, SP6, THAP1, ZFP369, ZSCAN12, ERCC2, REPIN1
	DNA binding	0.019	24.10	ZFP541, HOXA13, FOXA3, SOX5, ZKSCAN3, ZFP112, ZFP786, PRDM13, NPAS1, MEIS3, ZFP93, HJURP, POU2F2, ETV2, THAP1, CC2D1A, ATOH8, ERCC1, ERCC2, ZFP329, ZFP58, TRIM28, RELB, ZFP775, POLR1A, SMYD1,

				SIX5, MYPOP, MXD3, JUNB, RECQL, STRA13, SALL1, HIF3A, RFX1, SP6, ZFP369, ZSCAN12, CACNA1A, REPIN1
	ATP binding	0.022	28.32	ADCY7, HK2, DMPK, NOD2, DHX34, PALM3, ETNK1, PRKACA, NLRP4E, ERCC2, NLRP5, MYLK3, GARS, PKN1, PBP2, GUCY2C, MARK4, SARS2, RECQL, PRKD2, DDX39, ABCC9, MAST1, CKM, MAPK3, THRAP3, GRK6, EIF2AK3, NLRP9B, CDK20, CLCN4, SLC27A5, PPP5C
	RNA polymerase II core promoter proximal region sequence-specific DNA binding	0.026	31.83	ZFP58, SALL1, POU2F2, RELB, ETV2, RFX1, SIX5, CC2D1A, ZKSCAN3, MYPOP, ZFP457, JUNB
	RNA polymerase II transcription factor activity, sequence-specific DNA binding	0.030	35.99	NPAS1, FOXA3, HIF3A, SOX5, ZSCAN18, ZFP369, ZFP446
	protein kinase activity	0.041	45.56	MYLK3, TRIM28, AURKC, PKN1, GUCY2C, MARK4, DMPK, PRKD2, MAST1, MAPK3, GRK6, PRKACA, EIF2AK3, CDK20, ERCC2
	Starch and sucrose metabolism	4.0E-08	5.0E-05	UGT1A10, UGT1A9, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, HK2, GYS2, UGT1A1
	Ascorbate and aldarate metabolism	4.0E-08	5.1E-05	UGT1A10, UGT1A9, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, UGT1A1
	Pentose and glucuronate interconversions	4.6E-07	5.7E-04	UGT1A10, UGT1A9, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, UGT1A1
	Porphyrin and chlorophyll metabolism	1.2E-06	0.001	UGT1A10, UGT1A9, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, UGT1A1
	Steroid hormone biosynthesis	2.6E-06	0.003	UGT1A10, UGT1A9, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, HSD17B3, CYP2B10, UGT1A1
KEGG	Retinol metabolism	2.8E-05	0.034	UGT1A10, UGT1A9, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, CYP2B10, UGT1A1
	Chemical carcinogenesis	3.5E-05	0.044	UGT1A10, UGT1A9, UGT1A7C, UGT1A2, UGT1A6B, UGT1A6A, UGT1A5, CYP2B10, UGT1A1
	Circadian entrainment	0.002	2.739	GNG8, ADCY7, GRIN2B, MAPK3, RYR1, CALM3, PRKACA
	Calcium signaling pathway	0.003	3.485	PTGER1, SLC8A2, ADCY7, TACR1, MYLK3, RYR1, CALM3, PRKACA, CACNA1A
	cAMP signaling pathway	0.005	6.593	ADCY7, GRIN2B, RAC3, GIPR, MAPK3, CALM3, PRKACA, PDE3A, HHIP
	Morphine addiction	0.009	10.603	GNG8, ADCY7, GRK6, PRKACA, PDE3A, CACNA1A

3.1.4. Identificación de variantes por secuenciación de RNA total (RNA seq) de ratones consanguíneos

Se realizó secuenciación de RNA total de los tejidos de aorta, hígado y músculo de la pareja A. Se obtuvieron aproximadamente quince millones de lecturas, de las cuales el 80 % fueron alineadas correctamente contra el genoma de referencia (C57BL/6 J GRCm38.p2). Se secuenciaron 864 Mb de DNA, lo que corresponde a 25 veces la porción codificante del genoma de ratón (sin considerar isoformas de cada gen). Después del llamado de variantes por RNA-Seq (v-RNAseq), realizado con el programa SAMtools, se descartaron aquellas que presentaban menos de 5 lecturas. En promedio se encontraron 9400 v-RNAseq por muestra, con un mayor número en la aorta del ratón *ApoE*^{-/-} (11739) y menor en el hígado del ratón silvestre (5483). El número total de las v-RNAseq fue en promedio un 29% más alto, comparado con las v-WES.

Las v-RNAseq se distribuyeron de manera más homogénea en comparación con las v-WES, sobre todo en los tejidos del ratón silvestre (**Figura 11 y Figura 5**). Sin embargo, al igual que las v-WES, se observó un enriquecimiento de v-RNAseq en los cromosomas 6,7,8,11,13 en ambos casos; pero más evidente en el ratón *ApoE*^{-/-}. (**Figura 18, a y b, respectivamente**). De manera interesante se observó también una alta frecuencia de v-RNAseq en el cromosoma X, pero no de los v-WES (**Figura 5**).

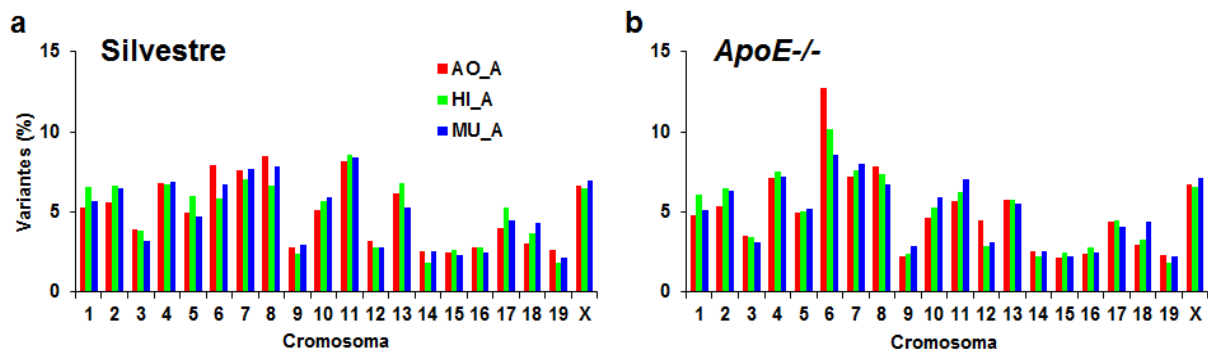


Figura 11. Distribución cromosómica de variantes del RNA

Porcentaje de variantes totales en las muestras silvestres (a) y *ApoE*^{-/-} (b).

3.1.4.1. Variantes específicas de ratones *ApoE*^{-/-} y silvestres

Como las v-WES, la mayoría de las v-RNAseq del hermano *ApoE*^{-/-} y del hermano silvestre eran compartidas entre ellos. El hígado fue la excepción, en donde el hermano silvestre tuvo más variantes comunes. De la misma forma, las v-RNAseq específicas de *ApoE*^{-/-} y ratones silvestres (v-RNAseq/*ApoE*^{-/-} y v-RNAseq/WT, respectivamente) fueron mayores en los tejidos del *ApoE*^{-/-}, excepto el músculo, (χ^2 , $p=3 \times 10^{-6}$). Las variantes identificadas correspondían en promedio a 160 genes con un promedio de 2 variantes por gen. Contrario a las v-WES, la mayoría de las v-RNAseq eran variantes nuevas (**Tabla 7**). El promedio de la relación Ti/Tv fue 2.2, lo cual indica un incremento en el número de transiciones en las v-RNAseq comparado con las v-WES (ver **Tabla 4**).

Tabla 7. Datos generales de las v-RNAseq/*ApoE*^{-/-} y v-RNAseq/WT

Tejido	Individuo	% del total	Genes	Ti/Tv	%dbSNPs
Aorta	<i>ApoE</i> ^{-/-}	31	268	2.2	49
	Silvestre	22	182	2.3	47
Hígado	<i>ApoE</i> ^{-/-}	52	247	2.4	45
	Silvestre	6	60	2.2	54
Músculo	<i>ApoE</i> ^{-/-}	17	82	2.2	47
	Silvestre	25	119	2.1	47

Al analizar las v-RNAseq/ *ApoE*^{-/-} y v-RNAseq/WT, se observó que las primeras estaban enriquecidas en el cromosoma 6, mientras que las segundas se concentraron principalmente en el cromosoma 11 y en menor medida en el cromosoma 6 (**Figura suplementaria 5**). A nivel de los genes en donde mapearon dichas variantes, también hubo una tendencia similar (**Figura 12**). De hecho, esta distribución fue similar la observada en los v-WES/*ApoE*^{-/-} y v-WES/WT (ver **Figura 9**), en especial si se comparan las variantes de los tejidos silvestres. El análisis de enriquecimiento mostró incidencia de las v-RNAseq/ *ApoE*^{-/-} en diversos procesos celulares en los tres tejidos (ver **Tablas suplementarias 3, 4 y 5**).

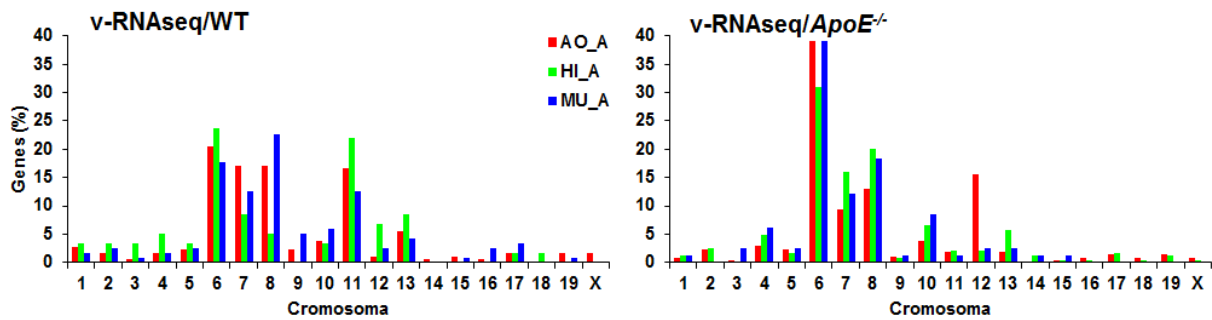


Figura 12. Distribución cromosómica de genes con variantes

De esta forma, los datos mostraron que las variantes identificadas por RNA-seq tenían una distribución cromosómica muy parecida a las v-WES, en particular a nivel de genes; es decir cuando se comparan la distribución cromosómica de los genes que contienen a las variantes encontradas a nivel de DNA y RNA. Lo anterior podría sugerir que las variantes somáticas están mapeando en determinados genes.

3.1.4.2. v-WES y v-RNAseq específicas de *ApoE*^{-/-}

Con el fin de validar algunas de las v-WES con las v-RNAseq, nos enfocamos en las 618 y 999 v-COMP/*ApoE*^{-/-} de aorta o el hígado, respectivamente (**ver Figura 10**). En la aorta solo el 3% (16/618) de estas variantes traslapaban con las v-RNAseq/*ApoE*^{-/-} (**Tabla 7**). Algo similar fue observado en hígado, en donde el 4% – 37/999 – fueron conservadas a nivel de DNA y de RNA.

Esta baja concordancia en ambos tejidos fue inesperada, ya que las v-WES y v-RNAseq tenían distribuciones cromosómicas muy parecidas, sobre todo tomando en cuenta los genes que tenían dichas variantes (**ver Figuras 9 y 12**). De hecho, al realizar la misma comparación, a nivel de genes, el porcentaje de solapamiento aumento al 14 % para la aorta y al 17 % para el hígado. Para saber si la expresión génica estaba implicada en lo anterior; se determinó el porcentaje de genes mapeados por variantes que estaban expresados. Se observó que en ambos tejidos, más del 70 % (72 y 70 % en aorta e hígado, respectivamente) de los genes estaban expresados. Sin embargo, esto no quiere decir que las lecturas del RNA-seq concuerden exactamente con las lecturas en el WES.

De las variantes confirmadas por ambas plataformas, todas (excepto una variante en el hígado que correspondía a un denominado cambio complejo SNV+INDEL), eran SNPs reportados (dbSNPs) en donde en promedio el 78 % correspondían a transiciones (**Tablas 8 y 9**). En ambos tejidos, la mayoría de estas variantes ocasionaban cambios sinónimos (91 y 76 %, en aorta e hígado, respectivamente).

Aunque a un nivel muy bajo, logramos a identificar variantes somáticas conservadas a nivel de DNA y RNA. La baja conservación puede incluir desde errores de la secuenciación, expresión diferencial de los alelos en *ApoE^{-/-}*, como la falta de transcripción en las regiones que contienen v-WES, o ediciones del RNA.

3.1.4.3. Relevancia de los genes con variantes comunes a *ApoE^{-/-}*

Encontramos 16 y 37 variantes específicas a *ApoE^{-/-}* en aorta e hígado, respectivamente. Algunos genes mapeados por estas variantes destacan por su relación con la aterosclerosis (indicados por un asterisco en las **Tablas 8 y 9**), en aorta suman el 75 % de las variantes (12/16), mientras que en hígado correspondieron al 27 % (10/37). Acorde a la tendencia general de los datos, la gran mayoría de las variantes mapeaban en los cromosomas 6 y 7, (en aorta por ejemplo todas estaban en dichos cromosomas).

Es de notar, que la mayoría de estas variantes fueron específicas de tejido y solo una estaba presente en los dos tejidos (chr6:135152868, *Hebp1*). Esta especificidad se fue reflejada en un análisis de GO y del KEGG, el cual mostró enriquecimiento en categorías como la respuesta inmune, la proliferación celular y la regulación de la transcripción en aorta; por su parte en hígado la regulación negativa de la transcripción y las vías de señalización de insulina y glucagón fueron las más relevantes (**Tablas 10 y 11**).

Tabla 8. Variantes putativas de aorta

Cr	Posición	Varian te	Gen	Tipo	dbSNP	Consecuencia
6	69197754	A/G	Igkv4-73	IG_V_pseudogene	rs58074438 3	non_coding_transcript_exon_variant, non_coding_transcript_variant
6	69197754	A/G	-	promoter_flanking_region	rs58074438 3	regulatory_region_variant
6	69197889	G/T	Igkv4-73	IG_V_pseudogene	rs57923784 4	splice_region_variant, non_coding_transcript_exon_variant, non_coding_transcript_variant
6	69197889	G/T	-	promoter_flanking_region	rs57923784 4	regulatory_region_variant
6	70074188	C/T	Igkv6-32*	IG_V_gene	rs24575345 8	synonymous_variant
6	70074188	C/T	-	CTCF_binding_site	rs24575345 8	regulatory_region_variant
6	71831432	G/A	Immt*	protein_coding	rs30693033	5_prime_UTR_variant
6	71831432	G/A	Immt*	retained_intron	rs30693033	non_coding_transcript_exon_variant, non_coding_transcript_variant
6	71831432	G/A	Immt*	nonsense_mediated_decay	rs30693033	5_prime_UTR_variant, NMD_transcript_variant
6	71831432	G/A	Mrpl35	processed_transcript	rs30693033	upstream_gene_variant
6	71831432	G/A	-	promoter	rs30693033	regulatory_region_variant
6	71831432	G/A	-	-	rs30693033	TF_binding_site_variant
6	72556204	T/C	Capg*	protein_coding	rs38237719	downstream_gene_variant
6	72556204	T/C	Capg*	retained_intron	rs38237719	downstream_gene_variant
6	72556204	T/C	Capg*	retained_intron	rs38237719	non_coding_transcript_exon_variant, non_coding_transcript_variant
6	72556204	T/C	Capg*	protein_coding	rs38237719	synonymous_variant
6	83195676	G/A	Dctn1*	retained_intron	rs13466206	upstream_gene_variant
6	83195676	G/A	Dctn1*	retained_intron	rs13466206	non_coding_transcript_exon_variant, non_coding_transcript_variant
6	83195676	G/A	Dctn1*	protein_coding	rs13466206	synonymous_variant
6	83195676	G/A	Dctn1*	retained_intron	rs13466206	downstream_gene_variant
6	134484297	C/T	Lrp6*	protein_coding	rs30946154	synonymous_variant
6	135152868	A/G	Gm22881	snoRNA	rs13467604	downstream_gene_variant
6	135152868	A/G	Hebp1*	protein_coding	rs13467604	synonymous_variant
6	135152868	A/G	Hebp1*	processed_transcript	rs13467604	downstream_gene_variant
6	135152868	A/G	Hebp1*	processed_transcript	rs13467604	non_coding_transcript_exon_variant, non_coding_transcript_variant

6	135155194	C/T	Hebp1*	protein_coding	rs51965372	synonymous_variant
6	135155194	C/T	Hebp1*	processed_transcript	rs51965372	downstream_gene_variant
6	135155194	C/T	Hebp1*	processed_transcript	rs51965372	non_coding_transcript_exon_variant, non_coding_transcript_variant
6	136609682	C/G	Plbd1	retained_intron	rs30126435	downstream_gene_variant
6	136609682	C/G	Atf7ip	protein_coding	rs30126435	3_prime_UTR_variant
6	136609682	C/G	Plbd1	protein_coding	rs30126435	downstream_gene_variant
7	16248935	A/G	C5ar1*	protein_coding	rs3023150	synonymous_variant
7	16248935	A/G	C5ar2*	protein_coding	rs3023150	upstream_gene_variant
7	16837727	A/C	Strn4*	protein_coding	rs31961396	missense_variant
7	16837727	A/C	Strn4*	retained_intron	rs31961396	downstream_gene_variant
7	16837727	A/C	Strn4*	retained_intron	rs31961396	upstream_gene_variant
7	16837727	A/C	Strn4*	nonsense_mediated_decay	rs31961396	3_prime_UTR_variant, NMD_transcript_variant
7	16837727	A/C	Strn4*	retained_intron	rs31961396	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	16883213	G/A	Dact3	protein_coding	rs25543877 2	synonymous_variant
7	19606691	T/C	Relb*	retained_intron	rs23996716 9	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19606691	T/C	Clasrp	nonsense_mediated_decay	rs23996716 9	upstream_gene_variant
7	19606691	T/C	Clasrp	protein_coding	rs23996716 9	upstream_gene_variant
7	19606691	T/C	Gm26890	lincRNA	rs23996716 9	upstream_gene_variant
7	19606691	T/C	Relb*	protein_coding	rs23996716 9	synonymous_variant
7	19606691	T/C	Relb*	nonsense_mediated_decay	rs23996716 9	3_prime_UTR_variant, NMD_transcript_variant
7	19606691	T/C	Clasrp	retained_intron	rs23996716 9	upstream_gene_variant
7	19619655	C/T	Relb*	retained_intron	rs21682703 1	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19619655	C/T	Relb*	protein_coding	rs21682703 1	synonymous_variant
7	19619655	C/T	Relb*	processed_transcript	rs21682703 1	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19619655	C/T	Relb*	nonsense_mediated_decay	rs21682703 1	synonymous_variant, NMD_transcript_variant
7	19619655	C/T	-	CTCF_binding_site	rs21682703 1	regulatory_region_variant

7	19619655	C/T	-	promoter_flanking_region	rs21682703 1	regulatory_region_variant
7	19809598	G/A	Bcl3*	protein_coding	rs49235820	synonymous_variant
7	19809598	G/A	Bcl3*	processed_transcript	rs49235820	downstream_gene_variant
7	19809598	G/A	Bcl3*	retained_intron	rs49235820	downstream_gene_variant
7	19809598	G/A	Gm16175	antisense	rs49235820	upstream_gene_variant
7	19809598	G/A	-	promoter	rs49235820	regulatory_region_variant

*Genes relevantes en la aterosclerosis

Tabla 9. Variantes putativas de hígado

Chr	Position	Variant	SYMBOL	BIOTYPE	Existing_variation	Consequence
1	88259678	A/G	Mroh2a	protein_coding	rs49400006	intron_variant
1	88259678	A/G	Hjurp	protein_coding	rs49400006	downstream_gene_variant
1	88259678	A/G	-	promoter_flanking_region	rs49400006	regulatory_region_variant
6	48598289	C/T	Repin1*	protein_coding	rs29924209	downstream_gene_variant
6	48598289	C/T	Zfp775	protein_coding	rs29924209	upstream_gene_variant
6	48598289	C/T	Repin1*	retained_intron	rs29924209	downstream_gene_variant
6	48598289	C/T	Repin1*	protein_coding	rs29924209	3_prime_UTR_variant
6	71199872	A/G	Fabp1*	protein_coding	rs257021964	5_prime_UTR_variant
6	71199872	A/G	-	CTCF_binding_site	rs257021964	regulatory_region_variant
6	71199872	A/G	-	promoter	rs257021964	regulatory_region_variant
6	71414791	A/G	Rmnd5a	protein_coding	rs30569747	synonymous_variant
6	71414866	G/C	Rmnd5a	protein_coding	rs30649753	missense_variant
6	71510450	G/T	Rnf103	protein_coding	rs30667120	3_prime_UTR_variant
6	71510450	G/T	Rnf103	retained_intron	rs30667120	downstream_gene_variant
6	73260489	T/C	Suclg1	nonsense_mediated_decay	rs13467077	3_prime_UTR_variant, NMD_transcript_variant
6	73260489	T/C	Suclg1	retained_intron	rs13467077	non_coding_transcript_exon_variant, non_coding_transcript_variant
6	73260489	T/C	Suclg1	protein_coding	rs13467077	synonymous_variant
6	73260489	T/C	Suclg1	processed_transcript	rs13467077	downstream_gene_variant
6	73260489	T/C	Suclg1	protein_coding	rs13467077	intron_variant
6	134920985	G/T	Cdkn1b	protein_coding	rs13467280	missense_variant
6	134920985	G/T	-	promoter	rs13467280	regulatory_region_variant
6	135152868	A/G	Gm22881	snoRNA	rs13467604	downstream_gene_variant
6	135152868	A/G	Hebp1*	protein_coding	rs13467604	synonymous_variant
6	135152868	A/G	Hebp1*	processed_transcript	rs13467604	downstream_gene_variant
6	135152868	A/G	Hebp1*	processed_transcript	rs13467604	non_coding_transcript_exon_variant, non_coding_transcript_variant
6	142456335	T/C	Gys2*	protein_coding	rs32172841	synonymous_variant
6	142456425	A/T	Gys2*	protein_coding	rs240772411	synonymous_variant

6	142772159	T/C	Cmas	retained_intron	rs46054869	downstream_gene_variant
6	142772159	T/C	Cmas	nonsense_mediated_decay	rs46054869	3_prime_UTR_variant, NMD_transcript_variant
6	142772159	T/C	Cmas	protein_coding	rs46054869	synonymous_variant
6	142772159	T/C	Cmas	nonsense_mediated_decay	rs46054869	downstream_gene_variant
6	142775178	C/G	Cmas	retained_intron	rs47460449	downstream_gene_variant
6	142775178	C/G	Cmas	nonsense_mediated_decay	rs47460449	downstream_gene_variant
6	142775178	C/G	Cmas	protein_coding	rs47460449	synonymous_variant
7	12988608	C/T	Slc27a5*	processed_transcript	rs6198049	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	12988608	C/T	Zfp446	retained_intron	rs6198049	downstream_gene_variant
7	12988608	C/T	Zfp446	protein_coding	rs6198049	downstream_gene_variant
7	12988608	C/T	Slc27a5	protein_coding	rs6198049	synonymous_variant
7	12988608	C/T	Slc27a5	protein_coding	rs6198049	intron_variant
7	12988608	C/T	Slc27a5	retained_intron	rs6198049	downstream_gene_variant
7	12988608	C/T	Slc27a5	protein_coding	rs6198049	downstream_gene_variant
7	13029558	C/T	Trim28	retained_intron	rs13479127	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	13029558	C/T	Trim28	retained_intron	rs13479127	upstream_gene_variant
7	13029558	C/T	Trim28	retained_intron	rs13479127	downstream_gene_variant
7	13029558	C/T	Chmp2a	protein_coding	rs13479127	downstream_gene_variant
7	13029558	C/T	Trim28	protein_coding	rs13479127	missense_variant
7	16497608	T/C	Gm17826	processed_pseudogene	rs31335150	upstream_gene_variant
7	16497608	T/C	Arhgap35	protein_coding	rs31335150	3_prime_UTR_variant
7	16561623	A/G	Arhgap35	protein_coding	rs31773949	synonymous_variant
7	16563519	A/G	Arhgap35	protein_coding	rs32014716	synonymous_variant
7	16810589	A/C	Fkrp	protein_coding	rs31769119	synonymous_variant
7	16810589	A/C	Fkrp	retained_intron	rs31769119	downstream_gene_variant
7	16810589	A/C	Fkrp	protein_coding	rs31769119	downstream_gene_variant
7	16810589	A/C	-	promoter_flanking_region	rs31769119	regulatory_region_variant
7	16917193	A/G	Calm3	processed_transcript	rs31735892	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	16917193	A/G	Calm3	protein_coding	rs31735892	synonymous_variant

7	16917193	A/G	Calm3	nonsense_mediated_decay	rs31735892	upstream_gene_variant
7	17006986	G/A	Ppp5c*	protein_coding	rs36342204	synonymous_variant
7	17006986	G/A	Ppp5c*	retained_intron	rs36342204	downstream_gene_variant
7	17006986	G/A	Ppp5c*	retained_intron	rs36342204	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	17006986	G/A	Ppp5c*	processed_transcript	rs36342204	downstream_gene_variant
7	19005657	G/A	Myppos	antisense	rs13461383	upstream_gene_variant
7	19005657	G/A	Mypop	protein_coding	rs13461383	downstream_gene_variant
7	19005657	G/A	lrf2bp1	protein_coding	rs13461383	synonymous_variant
7	19005676	T/A	Myppos	antisense	rs13461382	upstream_gene_variant
7	19005676	T/A	Mypop	protein_coding	rs13461382	downstream_gene_variant
7	19005676	T/A	lrf2bp1	protein_coding	rs13461382	missense_variant
7	19006050	G/A	Myppos	antisense	rs249023235	upstream_gene_variant
7	19006050	G/A	Mypop	protein_coding	rs249023235	downstream_gene_variant
7	19006050	G/A	lrf2bp1	protein_coding	rs249023235	synonymous_variant
7	19054461	GCC/ACT	Sympk	protein_coding	rs6211046	intron_variant
7	19054461	GCC/ACT	Sympk	retained_intron	rs6211046	downstream_gene_variant
7	19054461	GCC/ACT	Sympk	processed_transcript	rs6211046	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19054461	GCC/ACT	Rsph6a	protein_coding	rs6211046	upstream_gene_variant
7	19054461	GCC/ACT	Sympk	nonsense_mediated_decay	rs6211046	downstream_gene_variant
7	19054461	GCC/ACT	Sympk	protein_coding	rs6211046	frameshift_variant
7	19054461	GCC/ACT	Rsph6a	processed_transcript	rs6211046	upstream_gene_variant
7	19054461	GCC/ACT	Sympk	processed_transcript	rs6211046	downstream_gene_variant
7	19054461	GCC/ACT	-	CTCF_binding_site	rs6211046	regulatory_region_variant
7	19054461	GCC/ACT	-	TF_binding_site	rs6211046	regulatory_region_variant
7	19080149	T/C	Dmpk*	nonsense_mediated_decay	rs51876975	upstream_gene_variant
7	19080149	T/C	Dmwd	protein_coding	rs51876975	synonymous_variant
7	19080149	T/C	Dmpk*	protein_coding	rs51876975	upstream_gene_variant
7	19080149	T/C	Dmpk*	processed_transcript	rs51876975	upstream_gene_variant
7	19080149	T/C	Dmwd	processed_transcript	rs51876975	intron_variant, non_coding_transcript_variant

7	19081103	C/T	Dmpk*	nonsense_mediated_decay	rs32460915	upstream_gene_variant
7	19081103	C/T	Dmwd	protein_coding	rs32460915	synonymous_variant
7	19081103	C/T	Dmpk*	protein_coding	rs32460915	upstream_gene_variant
7	19081103	C/T	Dmpk*	processed_transcript	rs32460915	upstream_gene_variant
7	19081103	C/T	Dmwd	processed_transcript	rs32460915	intron_variant, non_coding_transcript_variant
7	19081103	C/T	Dmwd	processed_transcript	rs32460915	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19091818	C/T	Dmpk*	processed_transcript	rs8254017	downstream_gene_variant
7	19091818	C/T	Dmpk*	retained_intron	rs8254017	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19091818	C/T	Dmpk*	processed_transcript	rs8254017	upstream_gene_variant
7	19091818	C/T	Dmpk*	nonsense_mediated_decay	rs8254017	synonymous_variant, NMD_transcript_variant
7	19091818	C/T	Mir3100	miRNA	rs8254017	downstream_gene_variant
7	19091818	C/T	Six5	protein_coding	rs8254017	upstream_gene_variant
7	19091818	C/T	Dmpk*	processed_transcript	rs8254017	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19091818	C/T	Dmpk*	protein_coding	rs8254017	synonymous_variant
7	19091818	C/T	Dmpk*	retained_intron	rs8254017	upstream_gene_variant
7	19091818	C/T	Dmpk*	retained_intron	rs8254017	downstream_gene_variant
7	19091818	C/T	Dmpk*	protein_coding	rs8254017	downstream_gene_variant
7	19147080	G/C	Qpctl	protein_coding	rs256294294	synonymous_variant
7	19147080	G/C	Snrpd2	retained_intron	rs256294294	upstream_gene_variant
7	19147080	G/C	Snrpd2	protein_coding	rs256294294	upstream_gene_variant
7	19147080	G/C	Qpctl	retained_intron	rs256294294	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	19244821	A/G	Gm25134	snRNA	rs13460149	downstream_gene_variant
7	19244821	A/G	Opa3	protein_coding	rs13460149	intron_variant
7	19244821	A/G	-	promoter_flanking_region	rs13460149	regulatory_region_variant
7	28398962	T/C	Samd4b	nonsense_mediated_decay	rs3160012	downstream_gene_variant
7	28398962	T/C	Samd4b	protein_coding	rs3160012	downstream_gene_variant
7	28398962	T/C	Paf1	retained_intron	rs3160012	downstream_gene_variant
7	28398962	T/C	Paf1	retained_intron	rs3160012	non_coding_transcript_exon_variant, non_coding_transcript_variant
7	28398962	T/C	Paf1	protein_coding	rs3160012	synonymous_variant

7	31130430	A/G	Scn1b	protein_coding	rs260651625	upstream_gene_variant
7	31130430	A/G	Gramd1a	nonsense_mediated_decay	rs260651625	3_prime_UTR_variant, NMD_transcript_variant
7	31130430	A/G	Gramd1a	protein_coding	rs260651625	3_prime_UTR_variant
7	31130430	A/G	Gramd1a	retained_intron	rs260651625	non_coding_transcript_exon_variant, non_coding_transcript_variant
8	83994854	G/A	Prkaca	protein_coding	rs13460088	synonymous_variant
8	83994854	G/A	Samd1	protein_coding	rs13460088	upstream_gene_variant
8	89028672	A/G	Mir8110	miRNA	rs33111193	downstream_gene_variant
8	89028672	A/G	Sall1	protein_coding	rs33111193	synonymous_variant
13	65000298	T/C	-	-	rs29237663	intergenic_variant
13	66974514	G/T	Ptdss1	protein_coding	rs46681114	missense_variant
19	53006170	G/C	Mir6407	miRNA	rs223196633	downstream_gene_variant
19	53006170	G/C	Xpnpep1	processed_transcript	rs223196633	non_coding_transcript_exon_variant, non_coding_transcript_variant
19	53006170	G/C	Xpnpep1	nonsense_mediated_decay	rs223196633	3_prime_UTR_variant, NMD_transcript_variant
19	53006170	G/C	Xpnpep1	protein_coding	rs223196633	missense_variant

Tabla 10. Análisis de Gene Ontology (GO) y KEGG de aorta

Categoría	Término	Valor P	FDR	Genes
	T-helper 1 type immune response	5.95E-03	7.04	RELB, BCL3
	I-kappaB kinase/NF-kappaB signaling	2.03E-02	22.26	RELB, BCL3
GO_BP	negative regulation of Wnt signaling pathway	3.52E-02	35.52	DACT3, LRP6
	negative regulation of transcription, DNA-templated	3.86E-02	38.23	ATF7IP, RELB, BCL3
	positive regulation of epithelial cell proliferation	4.67E-02	44.32	C5AR2, C5AR1
GO_MF	C5a anaphylatoxin receptor activity	1.39E-03	1.27	C5AR2, C5AR1

Tabla 11. Análisis de Gene Ontology y KEGG de hígado

Categoría	Término	Valor P	FDR	Genes
a				
GO_BP	negative regulation of transcription from RNA polymerase II promoter	9.44E-03	11.63	SALL1, TRIM28, ARHGAP35, PAF1, IRF2BP1, MYPOP
	negative regulation of transcription, DNA-templated	1.17E-02	14.20	CDKN1B, SALL1, TRIM28, SIX5, ARHGAP35
	neural tube closure	1.32E-02	15.95	SALL1, PRKACA, ARHGAP35
	regulation of transcription, DNA-templated	2.88E-02	31.71	SALL1, PAK4, TRIM28, ZFP775, SIX5, ARHGAP35, PAF1, IRF2BP1, MYPOP
	regulation of sodium ion transport	3.85E-02	40.04	SCN1B, DMPK
	transcription, DNA-templated	4.67E-02	46.37	SALL1, TRIM28, ZFP775, SIX5, ARHGAP35, PAF1, IRF2BP1, MYPOP
	mRNA polyadenylation	4.74E-02	46.92	SYMPK, PAF1
GO_CC	cytosol	4.99E-05	0.05	CHMP2A, CDKN1B, CMAS, HEBP1, SNRPD2, GYS2, FABP1, ARHGAP35, XPNPEP1, DMPK, REPIN1, PPP5C
	nucleus	5.29E-03	5.64	SYMPK, CMAS, TRIM28, ZFP775, SNRPD2, SIX5, ARHGAP35, IRF2BP1, MYPOP, ZFP446, DMPK, CDKN1B, HJURP, SALL1, CALM3, PRKACA, FABP1, PAF1, PPP5C, REPIN1
	protein complex	6.12E-03	6.50	CDKN1B, SUCLG1, PRKACA, FABP1, SLC27A5, PPP5C
	cytoplasm	1.91E-02	19.08	CHMP2A, SYMPK, SUCLG1, SNRPD2, SIX5, ARHGAP35, SAMD1, XPNPEP1, DMPK, CDKN1B, HJURP, SALL1, PAK4, HEBP1, CALM3, GYS2, PRKACA, FABP1, PAF1, PPP5C
GO_MF	ligase activity	4.53E-03	5.11	RNF103, SUCLG1, TRIM28, IRF2BP1, SLC27A5
	protein complex binding	2.85E-02	28.36	CDKN1B, PRKACA, ARHGAP35, SLC27A5

	DNA binding	3.69E-02	35.22	HJURP, SALL1, TRIM28, ZFP775, SIX5, ARHGAP35, MYPOP, REPIN1
	transferase activity	4.89E-02	43.98	QPCTL, CMAS, GYS2, PRKACA, PTDSS1, FKRP, DMPK
KEGG	Glucagon signaling pathway	1.57E-02	16.03	GYS2, CALM3, PRKACA
	Insulin signaling pathway	2.95E-02	28.16	GYS2, CALM3, PRKACA
	Adrenergic signaling in cardiomyocytes	3.35E-02	31.36	SCN1B, CALM3, PRKACA

En aorta, uno de los genes relevantes fue *Immt* (**Figura 13**). La sustitución de G/A en la posición chr6:71831432 de este gen ya había sido reportada como un SNP (rs30693033). Se ubica hacia la región 5' UTR del gen, cerca del primer exón en una isla CpG (isla 17531). De manera interesante el SNP rs30693033 se encuentra en el inicio de una región regulatoria (chr6:71831433-71831443) en el sitio de unión a E2F3 y rio arriba de otro gen (*Mrpl35*).

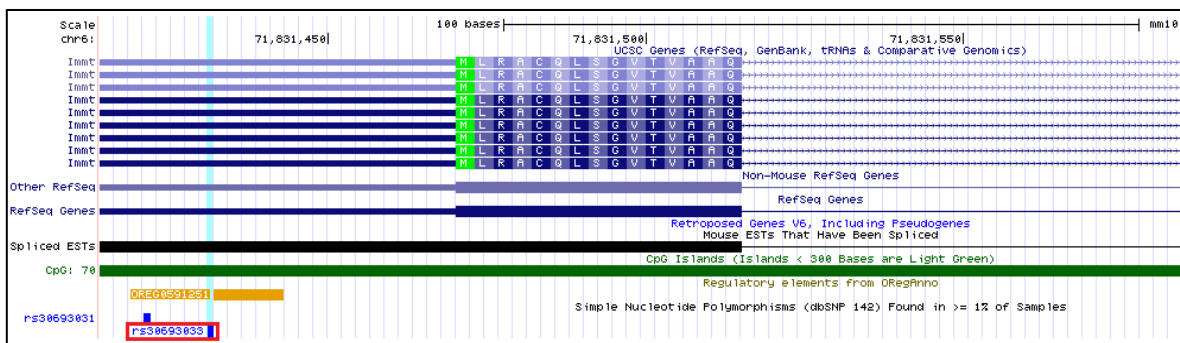


Figura 13. Ubicación del SNP rs30693033 de *Immt* en ratón

Se observan diferentes SNPs hacia la región hacia el 5' del gen *Immt* (en color azul), en una isla CpG (verde); el SNPs rs30693033 se marca con un recuadro rojo.

Immt codifica para una proteína mitocondrial con un 90 % de identidad a la proteína de humanos y es indispensable para la correcta morfología de las crestas mitocondriales (Liu *et al.*, 2013). Ciertamente, el gen *IMMT* es interesante dada su metilación diferencial en ateromas aórticos y sangre periférica en humanos, más exactamente, porque su región promotora se encuentra hipermetilada en aortas ateroscleróticas y podría otorgar información importante como marcador funcional (Sharma *et al.*, 2008; Zaina *et al.*, 2014).

Otros genes importantes son *Lrp6*, para el cual justamente se han detectado algunos alelos de riesgo asociados con el síndrome metabólico y la aterosclerosis en humanos (Go, 2015). *Hebp1* que promueve la movilización de calcio y quimiotaxis en monocitos y células dendríticas (Migeotte *et al.*, 2005). Por último, *Relb* un miembro de la familia de factores de transcripción NFκB, un marcador putativo asociado a la salud cardiovascular (Labonté *et al.*, 2014).

Por su parte en hígado, el gen que llamó la atención por codificar para una proteína de dedos de zinc, altamente expresada en hígado y tejido adiposo fue Repin1. Este gen mapea en un QTL de susceptibilidad a dislipidemia y desordenes metabólicos en ratas y su deficiencia ocasiona cambios en moléculas como Cd36, Ppar γ , Glut2, fosforilación de Akt y expresión del mRNA de lipocalin2, Vamp4, y Snap23. Al parecer este gen juega un papel importante en la sensibilidad a insulina y el metabolismo lipídico (Kern *et al.*, 2014). Otro gen importante es Fabp1, el cual es crítico en la absorción y transporte intracelular de ácidos grasos. Es un importante regulador del metabolismo de lípidos y de señalización celular (Wang *et al.*, 2015).

3.2. Expresión diferencial de genes y su influencia con la determinación de variantes

Como hemos observado que la mayoría de las v-COMP/*ApoE*^{-/-} eran tejido específicas (55 y 71% de la aorta y hígado, respectivamente) y la gran mayoría mapeaban en los cromosomas 6 y 7 (ver apartado 5.2.3), nos preguntamos si la falta de solapamiento entre las v-COMP con v-RNA-seq podría estar relacionada con expresión diferencial de genes. Para este fin, se determinaron los genes que estaban diferencialmente expresados en los tejidos del ratón *ApoE*^{-/-} con respecto al silvestre bajo un valor $p \geq 0.9$ (**Tabla 12**). La lista completa de los genes con sus respectivos valores de p y “fold-change” se encuentra en el disco adjunto a esta tesis. En todos los tejidos del ratón *ApoE*^{-/-}, hubo una tendencia hacia el aumento en la expresión de genes respecto al silvestre (**Tabla suplementaria 6**). El tejido con más genes con expresión diferencial fue la aorta, seguido por el hígado y músculo. Cabe resaltar que el hecho de que el 89% de los genes (1436/1615), se encontraban expresados en los tres tejidos, indicaba que la falta de expresión diferencial a algún tejido no fue debida a la ausencia de lectura.

Tabla 12. Número de genes únicos expresados diferencialmente en *ApoE*^{-/-} por tejido

Expresión diferencial	Aorta	Hígado	Músculo
Up	910	79	55
Down	577	54	26
TOTAL*	1487	133	81

*el total de genes incluye genes expresados diferencialmente en dos o más tejidos

De los 1615 genes que mostraron expresión diferencial, 95% (1532/1615) cambiaron su expresión solo en un tejido y solo 0.2% (3/1615) fueron diferenciales entre ratones *ApoE*^{-/-} y silvestres en todos los tres tejidos. Estos genes corresponden a *ApoE* (apolipoprotein E), *Pianp* (PILR alpha associated neural protein) y *Thns2* (threonine synthase-like 2 (bacterial)), cuya expresión estaba reprimida en todos los tres tejidos de ratones *ApoE*^{-/-}.

Con respecto a la distribución por cromosoma, la mayoría de los genes que mostraron expresión diferencial se distribuyeron de manera significativa en determinados cromosomas. En los tres tejidos, se observó un enriquecimiento de genes diferencialmente expresados en los cromosomas 6, 8 y 11. De manera interesante, dicha distribución estuvo correlacionada positivamente con las distribuciones de las v-WES y v-RNAseq en aorta y músculo (**Figura 14**). En aorta, el porcentaje de genes mapeados por las v-WES y v-RNAseq estuvo correlacionado significativamente con el porcentaje de expresión de genes (Pearson; $r=0.56$, $p=0.008$ y $r=0.52$, $p=0.016$; respectivamente) (**Figura suplementaria 6**).

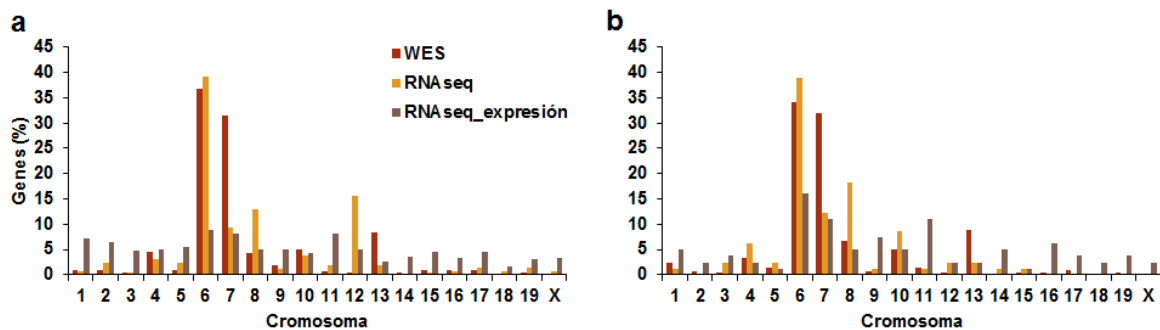


Figura 14. Distribución cromosómica de *v-WES*, *v-RNAseq* y de genes expresados diferencialmente

Se muestran el porcentaje de genes con variantes encontradas en WES, RNA-seq y de genes diferencialmente expresados en aorta (a) y músculo (c).

Para evaluar la posible implicación funcional de los genes diferencialmente expresados, se usó la herramienta DAVID bioinformatics, en especial para observar la presencia de vías de señalización enriquecidas descritas por el KEGG. En aorta, los genes sobreexpresados participan significativamente en procesos relacionados con respuesta inmune y adhesión celular. En el caso de los genes reprimidos, las vías de señalización enriquecidas tienen que ver con enfermedades neurológicas, cardíacas y cáncer. Por otra parte, en hígado las vías de señalización comprometidas tienen relación con cáncer, miocarditis y metabolismo

de glutatión. En cuanto al músculo, dado que hubo pocos genes expresados diferencialmente, las vías enriquecidas son la señalización por PPAR, el metabolismo de xenobióticos y de drogas para los genes sobreexpresados y de contracción de músculo cardiaco y miocarditis viral para los genes reprimidos (**Tabla suplementaria 6**).

De esta misma forma, se realizó un análisis de los genes que muestran un perfil de expresión tejido-específico y diferencial respecto al silvestre (**ver Tabla 6**). En conjunto, los 7 genes reprimidos afectan significativamente el metabolismo de tirosina (KEGG_PATHWAY, $p=6.8E-3$), mientras que los 11 genes sobre expresados, aparentemente no afectan ninguna vía de señalización aterosclerótica en la aorta de un ratón *ApoE^{-/-}*.

3.3. Asociación del perfil de expresión de genes en aortas de humanos y ratones que sufren aterosclerosis

La dinámica en la expresión de genes puede estar dirigida por alteraciones epigenéticas como la metilación del DNA. En humanos, los perfiles de metilación de arterias carótidas y aortas con aterosclerosis, han sido previamente documentados, así como su relevancia en cuanto a los genes y/o vías de señalización que son regulados. Estos análisis se han concentrado en la comparación de los metilomas de porciones sanas y ateroscleróticas inter e intra/individual. Ahora se conoce qué a nivel genómico, la hipermetilación global del DNA, involucra tanto regiones codificantes como no codificantes (elementos repetidos), además de que claramente se diferencia de regiones de tejido normal (hipometilación en la porción asintomática de un individuo con aterosclerosis). De mayor relevancia, ha sido la determinación de CpGs metilados que mapean en genes específicos de la aterosclerosis y cuya utilidad terapéutica podría explorarse (Zaina et al., 2014, 2015).

Aunque a nivel genómico la hipermetilación del DNA podría ser una firma molecular observada en la aterosclerosis, varias preguntas permanecen por ser abordadas. Dos de ellas están relacionadas con cambios en perfiles de metilación en loci específicos que pudieran estar siendo modulados a través del desarrollo del ateroma; de igual forma, resulta interesante observar si estos cambios están relacionados tanto con la expresión de los genes mapeados tanto en humanos como en ratón.

Como un acercamiento a los aspectos anteriores, se realizó un trabajo conjunto entre los datos (aortas de humanos) obtenidos por el Dr. Silvio Zaina de la universidad de Guanajuato en su estancia sabática en el IDIBELL, (Bellvitge Biomedical Research Institute, Barcelona, España) (Zaina et al., 2014) y el transcrito de aorta de ratón *ApoE*^{-/-}. Los datos obtenidos en humano comprendían el metiloma de regiones de aorta aterosclerótica y asintomática de 15 donadores *post mortem*. Los análisis de metilación se realizaron utilizando el microarreglo *HumanMethylation450* (450 k array) de Illumina que contiene 485,764 *loci* de citosinas distribuidas a lo largo del genoma humano que a su vez cubren cerca del 99% de los genes anotados y del 96% de las islas CpG (ICGs). La anotación de las CpG está entonces relacionada a las ICGs. De esta forma, se distinguen tres regiones: “playa”, 2kb flanqueando a la ICG; “arrecife”, 2kb inmediatamente adyacente a la playa; “mar abierto”, DNA por fuera de las ICG, playas o arrecifes. En cuanto a la anotación génica, TSS1500 o TSS200 indican la región -1500 pb o -200 pb del sitio de inicio de la transcripción, respectivamente (Sandoval et al., 2011).

Los datos relativos a la información de los donadores, tratamiento de las muestras y análisis de los microarreglos de metilación se describen en Zaina *et al* 2015.

Después de comparar los perfiles de metilación entre muestras con diferentes grados histológicos de aterosclerosis (desde III hasta VII), se identificaron 1,985 sitios CpG (grado-GpG, correspondientes a 1,206 genes) cuyo agrupamiento tendía hacia la hipermetilación en muestras con grado histológico alto, en donde el 82.2% de ellos mostró correlación positiva con el grado de la lesión.

Dado que parte de la relevancia de la metilación diferencial de los grado-CpG es su impacto transcripcional, se procedió a correlacionar los perfiles de metilación con datos de expresión de genes. Algunos de los genes encontrados comprenden a *PDGFA* y *C9orf3/MIR23b* cuya metilación diferencial se ha visto relacionada con cambios de expresión génica en aorta (Zaina *et al.*, 2014). Adicionalmente, 105 genes que contienen grado-CpGs han sido relacionados con genes firma de placa aterosclerótica (Puig et al., 2011) y 87 con genes reportados como relacionados a la aterosclerosis en la base de datos HuGENet™.

De especial importancia para el presente trabajo es conocer si la relación metilación-transcripción observada en humanos, podría replicarse un modelo animal de aterosclerosis. Para esto, se compararon el estado de metilación de las grado-CpG y el perfil de expresión diferencial génica de la aorta del ratón *ApoE*^{-/-}. Se encontraron en total 99 grado-CpG que

estaban correlacionadas con 93 genes expresados diferencialmente en ratón (**Tabla 13**). En especial, hubo correlación inversa los grado-CpG de la región promotora y la expresión diferencial de sus correspondientes genes en ratón (correlación de Pearson $r=-0.71$, $p=1.5 \times 10^{-4}$). Así mismo, la región de mar abierto mostró una correlación similar ($r=-0.384$, $p=0.009$, $n=46$), en donde las regiones de 5'UTR y de primer exón fueron más representadas (21 de 46, 45.6%).

Un grupo de genes mapeados por grado-CpGs (29) sobreexpresados en ratón, son blanco del regulador transcripcional NFKB (Nuclear Factor Kappa, $p=1.8 \times 10^{-4}$) (Shimizu et al., 1990). Esto se ve reflejado en el enriquecimiento de genes CARM1 (PRKAR1B, RARA; $p=0.006$ y $p=0.05$, respectivamente) que regulan la respuesta inflamatoria y la transcripción. En cuanto a los genes reprimidos, 32 eran regulados por MEF2A (myocyte enhancer factor 2A), un factor de transcripción implicado en arterias coronarias ($p=3.5 \times 10^{-7}$).

Tabla 13. Análisis de correlación de grado-CpGs y expresión diferencial de genes en ratón

Gen (m)	Radio de expresión ApoE ^{-/-} /wt (m)	CpG_ID (h)	Región génica (h)	Posición relativa a CGIs (h)*	Correlación de Spearman rho metilación/grado (h)
Abr	2.14	cg10024094	Body	Island	0.59
Adap2	5.29	cg06957310	Body		0.55
Adcy3	3.49	cg25562925	Body		0.57
Arhgap25	11.07	cg15477363	5'UTR		-0.55
Arhgap27	4.72	cg11189268	TSS200	Island	0.52
Baiap2	2.64	cg23261327	Body	N_Shore	0.52
C1qc	7.00	cg05308979	Body	Island	0.66
C3	4.88	cg04601383	Body	S_Shelf	0.65
Cacna1a	2.87	cg22084460	Body	S_Shore	0.59
Cacna1c	0.47	cg17401938	Body	S_Shelf	0.62
Casp4	4.74	cg15665081	5'UTR		0.66
Ccr5	15.59	cg00803692	TSS200		-0.55
Cdh11	2.20	cg04977618	5'UTR		0.56
Ckm	0.17	cg20444256	TSS1500	N_Shore	0.53
Clip2	2.22	cg04130408	TSS1500	N_Shore	-0.55
Comp	3.38	cg15784332	TSS200	Island	0.56
Csgalnact1	0.35	cg27510066	Body		0.53
Ctnna3	0.36	cg04030146	Body		0.54
Cyp2e1	0.39	cg00436603	TSS200	N_Shore	0.61

Ddc	0.06	cg11111332	5'UTR		0.62
Dmd	0.42	cg03252296	Body	N_Shore	0.52
Dnajc8	0.36	cg07424927	TSS200	Island	0.52
Dnmt3a	5.39	cg20948740	Body	Island	0.52
Dtx4	3.84	cg07745373	Body	S_Shore	0.62
Dusp7	3.57	cg07615364	Body	N_Shore	0.56
Dynll1	0.45	cg01168339	Body	S_Shore	0.60
Efhd1	0.34	cg22193436	Body	S_Shore	0.52
Eno3	0.44	cg12331425	1stExon	S_Shore	0.66
F10	8.81	cg17835016	Body	Island	0.62
Fam110b	0.47	cg09379188	5'UTR		0.69
Fam134b	5.14	cg00819078	Body	N_Shore	-0.53
Fcho1	135.97	cg15822394	Body	Island	0.59
Fnip2	4.12	cg22441846	Body	S_Shelf	0.58
Frzb	7.19	cg26700716	1stExon	Island	0.63
Gadl1	0.39	cg04556868	TSS200	S_Shore	0.69
Galns	5.51	cg09308897	Body	S_Shore	0.57
Gng2	3.48	cg08001559	TSS200		0.54
Gpnmb	83.20	cg14444376	5'UTR	N_Shore	0.56
Gpr176	18.10	cg20588193	Body		0.60
Gpr65	7.10	cg19755435	5'UTR		-0.52
Gsta3	0.32	cg12445586	1stExon		0.66
Hrc	0.20	cg21697944	TSS200	N_Shelf	0.58
Hspb8	0.47	cg11187110	Body		0.52
Ikake	3.15	cg11739399	3'UTR		0.58
Inpp5a	0.49	cg09415366	Body		0.63
Irs2	0.43	cg00191629	TSS200	Island	0.56
Itgb2	18.47	cg18012089	Body	S_Shelf	0.57
Kcnma1	0.39	cg18507365	Body		-0.52
Ldhd	0.35	cg02258387	Body	N_Shelf	0.58
Lmod1	0.46	cg09404633	1stExon		0.71
Lpl	5.28	cg07263235	TSS200	N_Shore	0.56
Lrp1	2.08	cg01515074	Body		0.57
Ly86	7.35	cg03349397	TSS1500		-0.54
Lyn	6.68	cg03388786	5'UTR		0.68
Mbnl1	0.49	cg18861231	Body		0.58
Mid1	2.55	cg06814121	TSS1500		0.56
Mpeg1	81.06	cg11179518	1stExon		-0.52
Mrvi1	0.48	cg26262049	Body		0.54

Msra	0.42	cg10300895	Body		-0.54
Myo10	2.65	cg26422488	1stExon	Island	0.57
Myo7a	5.53	cg08036798	TSS200		0.63
Nov	0.41	cg05020510	TSS1500	N_Shore	0.64
Nr4a2	0.29	cg00426720	5'UTR	N_Shore	-0.54
Ntf3	0.26	cg17775989	Body		0.52
Osbpl8	6.31	cg07815177	5'UTR		0.60
Pdpn	3.21	cg04886857	5'UTR	S_Shelf	0.56
Pik3cd	5.36	cg20994801	TSS1500	N_Shore	-0.52
Ppef1	0.26	cg18811184	5'UTR		0.64
Prdm16	0.38	cg04852443	Body	Island	0.61
Prkar1b	3.91	cg24986615	Body	N_Shore	0.53
Pstpip1	6.68	cg11829253	Body	S_Shore	0.54
Ptpro	5.46	cg21027526	Body		0.57
Rai14	2.80	cg22429911	TSS1500		0.58
Rara	2.50	cg25362050	5'UTR	N_Shelf	0.52
Rarres1	3.40	cg16708174	Body		-0.52
Rbpms2	0.35	cg13474370	TSS1500	S_Shore	0.56
Rcan2	0.42	cg26677394	Body		0.53
Sbno2	4.88	cg09506661	Body	Island	0.76
Slc10a3	2.34	cg03025340	Body	N_Shore	0.67
Slc16a3	2.52	cg23664708	3'UTR	S_Shore	0.52
Slc39a11	3.64	cg09696465	Body		0.64
Slc6a8	2.71	cg03025466	3'UTR		0.64
Snrnp27	0.39	cg14608581	Body		0.66
Snrpn	0.29	cg22491305	TSS1500		0.77
St5	0.48	cg08710564	5'UTR		-0.52
Synpo2	0.31	cg08384155	Body		0.53
Thsd4	0.28	cg24753094	Body		0.73
Tlcd1	0.37	cg09479241	Body	N_Shore	0.56
Tmem176b	2.68	cg05017199	TSS1500	Island	0.55
Tnik	0.40	cg15237829	Body		0.52
Tns3	2.68	cg18335223	Body		0.52
Top2a	5.13	cg18314885	Body		0.54
Tspan2	0.40	cg19763774	Body		0.53
Tspan6	0.39	cg19041765	Body	Island	0.55
Vcam1	11.01	cg25763716	TSS1500		-0.54
Wfdc1	0.28	cg09904140	TSS200	N_Shore	0.60
Wif1	0.20	cg15862358	TSS1500	S_Shore	0.52

Zbtb20	0.46	cg17875356	5'UTR	0.74
Zmynd15	4.76	cg17900689	Body	0.63

Genes diferencialmente expresados en ratón (m) y metilación de CpGs en humanos (h)

*la ausencia de término indica región de mar abierto.

CAPITULO IV

Discusión

El presente trabajo ha sido un acercamiento hacia la determinación de las variantes somáticas *de novo* en la aterosclerosis murina. Su importancia radica en la necesidad de encontrar nuevos mecanismos que ayuden a predecir eficientemente el riesgo cardiovascular. En este sentido, el modelo de ratón *ApoE*^{-/-}, cuyo fenotipo se da en ausencia de estímulos externos, fue idóneo para retratar la variación genética intrínseca (con poca influencia del ambiente) que conlleva al desarrollo de la aterosclerosis. Precisamente, una de las primeras observaciones en los individuos analizados, fue la presencia de una placa aterosclerótica extendida por diferentes puntos de la arteria aorta (**Figura 4**). Dicha observación no incluyó a los hermanos silvestres y es una característica anatómica típica del fenotipo aterosclerótico, similar a la presentada en humanos (Nakashima *et al.*, 1994).

En promedio, encontramos 7600 variantes (SNVs + INDELS) en los individuos *ApoE*^{-/-} y 6300 en los silvestres (**Tabla 3**). Este valor está acorde con datos publicados en donde el rango de variantes exónicas de bazo se encuentra entre 8000 para cepas que tienen un fondo C57BL6/J y 200 mil, para cepas más divergentes (Fairfield *et al.*, 2011). A nivel de genoma completo, el número total de los SNPs también varía en un rango amplio (entre 350 mil y 19.3 Mb) a lo largo de diferentes cepas (Doran *et al.*, 2016). Por ejemplo la cepa C57BL/10J, genéticamente cercana a C57BL6/J (Beck *et al.*, 2000), tiene 349 mil SNPs en el genoma, lo que correspondería a un promedio de 4900 SNPs en el exoma, si se tiene en cuenta que éste es el 1.4 del genoma (Doran *et al.*, 2016). En nuestro caso, el número de SNVs estuvo en promedio en 6700 para los tejidos de ratones *ApoE*^{-/-} y 6300 para sus hermanos silvestres.

La distribución de las variantes llamó la atención por su marcada tendencia hacia determinados cromosomas tanto en el DNA (**Figuras 5 y 9**) como el RNA (**Figuras 11 y 12**). También hubo variación a nivel regional; el cromosoma 6 por ejemplo, mostró regiones de alta densidad de variantes en tejidos de ratones *ApoE*^{-/-} comparado con los ratones silvestres (**Figuras 7 y 8**). Interesantemente, esta distribución se asemeja a la observada en C58/J, una cepa homocigota para *Cdh23ahl* que desarrolla leucemia y que tiene tendencia a la aterosclerosis (Kane *et al.*, 2012). Concretamente, la región comprendida entre las bandas qB3 y qC3 (**Figura suplementaria 7**), alberga variantes en las muestras

ApoE^{-/-} y en la cepa C58/J, pero no así en las muestras de ratones silvestres. Esta región contiene diferentes genes, entre ellos un grupo considerable de genes de inmunoglobulinas (Igs). Es de notar que la mayoría de las variantes que mapearon en Igs se presentaron en tejidos de ratones *ApoE*^{-/-} (**Tablas suplementarias 1 y 2**) y que además algunas de ellas fueron específicas de tejido, tanto en DNA como en RNA. Uno de los mecanismos que podría estar dirigiendo este fenómeno podría ser la hipermetilación somática de loci de inmunoglobulinas kappa variable (Igkv) en donde están involucradas enzimas citidina desaminasa como AID (activation-induced deaminase) (Jacob *et al.*, 1991; Muramatsu *et al.*, 1999).

Fue notable también, la acumulación diferencial de variantes en los cromosomas 7, 8 y 11. El primero tuvo un número considerable de variantes en los ratones *ApoE*^{-/-} comparado con los silvestres. Una posible explicación a este resultado, podría ser la inestabilidad genómica a menudo observada en placas ateroscleróticas inestables, que contienen anomalías cromosómicas como trisomía/tetrasomía del cromosoma 7 y monosomía del cromosoma 11. Al contrario de las placas estables, en donde no es común encontrar dichas alteraciones cromosómicas (Maturri *et al.*, 2001).

Nuestro mayor interés estuvo centrado en la determinación de SNVs, especialmente en las variantes de transición toda vez que éstas conllevan un gran impacto funcional (Mort *et al.*, 2008), además de ser las más abundantes tanto en el genoma de humanos como de ratones (Steeland *et al.*, 2016; Wade *et al.*, 2002). Tal y como ya se ha reportado, las transiciones fueron más prevalentes en todas las muestras (**Tablas 4 y 7**); en la aorta, por ejemplo, más del 80% de las variantes confirmadas en el RNA eran transiciones (**Tablas 7 y 8**).

En genes humanos, los cambios de C-T son los más frecuentes ya que corresponden a 2/3 del total de las transiciones. De hecho, en el contexto de variantes genéticas que causan enfermedades en humanos, éstas son desde cinco veces hasta un orden de magnitud más comunes que la tasa de mutación basal (Koeberl *et al.*, 1990; Krawczak *et al.*, 1998; Nachman & Crowell, 2000). Uno de los mecanismos implicados en esta incidencia parece ser la metilación de la citosina (Cooper *et al.*, 2010).

La 5mdC es un blanco de variación debido a que puede desaminarse a timina generando un mal apareamiento G-T (Duncan & Miller, 1980). En eucariotes, hay un mecanismo de reparación que reconoce los malos apareamientos G-T y reemplaza la timina con una citosina. Sin embargo, esta reparación es poco eficiente, como se ha demostrado

por el hecho de que las transiciones 5mdC>T son 10 veces más frecuentes que otras sustituciones (Cooper *et al.*, 2010). Shen *et al.* (1994), determinaron las constantes de desaminación hidrolítica espontánea para 5mdC y C ($5.8 \times 10^{-13} \text{ s}^{-1}$ y $2.6 \times 10^{-13} \text{ s}^{-1}$, respectivamente). Interesantemente, estos valores están por encima de los necesarios para explicar la generación de “hot spots” de variación genética que causan enfermedades transmitidas por línea germinal y cáncer en tejidos somáticos (Shen *et al.*, 1994).

Las transiciones en sitios CpG no solo se dan de manera espontánea, también pueden ser consecuencia de otros mecanismos. Entre ellos, la acción de la familia de proteínas desaminasas AID/APOBEC (activation-induced deaminase/apolipoproteinB mRNA editing enzyme, catalytic polypeptide-like) (Chiu & Greene, 2008). Recientemente, se encontró que la proteína APOBEC3G de humano reconoce DNA de cadena sencilla y podría ser la responsable de las variantes de transición que ocurren en la cadena rezagada durante la replicación del DNA (Bhagwat *et al.*, 2016).

Es conveniente señalar que la hipermetilación del DNA, particularmente del cuerpo del gen (intrones y exones), es una firma epigenética de la formación y progresión de la lesión aterosclerótica (Silva-Martinez *et al.*, 2016; Zaina *et al.*, 2014). Durante este proceso, uno de los factores clave es la dieta, cuyos componentes han sido asociados con la modulación de la mDNA. Tal es el caso de los ácidos grasos insaturados araquidónico y el oléico, los cuales exhiben diferentes efectos diferenciales sobre la mDNA (hipermetilación e hipometilación, respectivamente) (Silva-Martínez *et al.*, 2016). A su vez, en sistemas *in vitro*, las lipoproteínas aterogénicas VLDL (Very Low Density lipoprotein), ricas en triglicéridos, y LDL (Low Density lipoprotein) inducen mDNA *de novo* de macrófagos humanos THP-1. Dicha metilación *de novo* esta mediada por la metiltransferasa DNMT1 y está asociada con la represión de varios genes críticos para la función de macrófagos, incluyendo genes pro-inflamatorios (Lund *et al.*, 2004; Rangel-Salazar *et al.*, 2011).

Evidentemente la mDNA es sensible a la dieta y por lo tanto su potencial farmacológico-nutricional puede ser relevante. Un ejemplo ilustrativo sobre como la dieta puede directamente influir en un fenotipo debido a cambios epigenéticos, proviene de estudios realizados en la abeja *Apis mellifera*. A pesar de la naturaleza clonal de su genoma, la abeja reina y sus obreras difieren con respecto a su anatomía, fisiología y longevidad como resultado del consumo diferencial de jalea real (Kucharski *et al.*, 2008). A nivel molecular, aproximadamente 550 genes muestran diferencias en la metilación del DNA entre la reina y sus obreras. Interesantemente, dichas diferencias de metilación se

detectaron dentro de regiones exónicas y fueron asociadas con procesamiento alternativo de mRNA en la reina y obreras (Lyko *et al.*, 2010).

También en mamíferos, la exposición prenatal a diferentes factores nutricionales y ambientales es crucial para el establecimiento y mantenimiento de marcas epigenéticas. Un claro ejemplo de ello, son los estudios hechos en el modelo de ratón Avy (viable yellow Agouti). Un suplemento de bisfenol A (BPA, un monómero estrogénico utilizado en la producción de policarbonato de plástico), en la dieta materna disminuye significativamente la metilación en el promotor del gen Agouti de la descendencia, lo que ocasiona diferencias fenotípicas en el color de la piel. La suplementación nutricional de la madre con donadores de grupos metilo contrarrestan los efectos del BPA, lo cual demuestra que los cambios dietéticos pueden proteger contra efectos epigenéticos dañinos causados por toxinas ambientales (Dolinoy *et al.*, 2006; Waterland *et al.*, 2007). También se han documentado casos de herencia transgeneracional de las marcas epigenéticas, aunque los mecanismos subyacentes no están claros. Carone *et al.* (2010), encontraron que los perfiles del epigenoma de ratones cuyo padre había sido alimentado con una dieta baja en proteína, tenían cambios modestos (~20%) pero numerosos en la metilación de citosinas comparados con los de su padre. De igual forma, se elevaron los niveles de expresión de genes involucrados en biosíntesis de colesterol y lípidos (Carone *et al.*, 2010).

En el presente estudio, además de identificar variantes en el DNA y RNA, encontramos particularmente importante indagar un poco más en los cambios transcripcionales que conlleva el modelo *ApoE*^{-/-} y más aún, la heterogeneidad de dichos cambios en diferentes tejidos (aorta, hígado y músculo). En el conjunto de genes diferencialmente expresados algunos se destacan por su expresión diferencial en al menos dos tejidos, tal es el caso de *Plin1*, *Adipoq* y *Retn*. La inactivación homocigota de éstos se ha asociado con alteraciones en la lipólisis, beta-oxidación, producción de leptina, susceptibilidad a la obesidad, problemas con el reciclado del colesterol y resistencia a la insulina, entre otros (Ikeda *et al.*, 2013; Wang *et al.*, 2016; Zhao *et al.*, 2015). Por su parte el gen *Rps27l*, que está reprimido en aorta y sobreexpresado en hígado, es un supresor de la inestabilidad genómica y tumorigénesis asociada a p53 (Xiong *et al.*, 2014).

En conjunto, los genes diferencialmente expresados en los tres tejidos; están significativamente involucrados en procesos relacionados a cada uno de ellos. De esta forma, en aorta la respuesta pro-inflamatoria, la migración y adhesión celular se encuentran aumentadas, mientras que la contracción de musculo vascular liso y cardiaco están

disminuidas. En el caso de hígado y músculo se destacan diferentes vías metabólicas como el metabolismo de glutatión, de PPAR y de xenobióticos (**Tabla suplementaria 6**).

La mayoría de las variantes encontradas en este trabajo eran comunes entre los individuos *ApoE*^{-/-} y silvestres; esto es algo que se esperaba dada la consanguinidad de los mismos. Tampoco se aleja de lo encontrado en humanos, pues la determinación de variantes genéticas en gemelos monocigotos mediante secuenciación masiva de DNA y exoma, ha determinado que éstos son genéticamente idénticos (Petersen *et al.*, 2014). Aquí es indispensable tener en cuenta la variación entre tejidos dentro de un mismo individuo; una sola muestra de sangre no provee una información genómica exhaustiva que se pueda extrapolar a todo el organismo. Precisamente, al interior del tejido sanguíneo puede haber divergencias, por ejemplo, Forsberg *et al.* (2012) mostraron que el genoma post-cigoto de las células sanguíneas es dinámico con múltiples expansiones y contracciones de células clonales a través del tiempo (Forsberg *et al.*, 2012). Así mismo, a través de la vida de un organismo, la tasa de variación somática se ha visto implicada en la estabilidad genómica de células somáticas *in vivo*, particularmente durante el envejecimiento (Siudeja *et al.*, 2015). En cuanto a patologías como el cáncer, la variabilidad intra e inter-tumoral ha sido descrita como un fenómeno recurrente (Nik-Zainal *et al.*, 2016).

Contrariamente a lo que se esperaba, la concordancia entre las variantes encontradas en DNA y RNA fue baja. Algunas explicaciones podrían ser errores en los protocolos de captura de RNA poli(A), baja cobertura, transcripción alélica diferencial y ediciones de RNA. En primer lugar, los estándares de cobertura y filtros de calidad aplicados a la captura de exoma sugieren poca injerencia. Por su parte, aunque la mayoría de los genes con variantes estaban expresados, es posible que las lecturas que avalaron tanto la variante como la expresión no fueran las mismas. Por último, a nivel de exones, se ha reportado que casi en el 50% no hay coincidencia entre la secuencia de RNA y DNA en un mismo individuo (Li *et al.*, 2011).

Por último, si bien es cierto que las cepas de ratones de laboratorio conllevan gran variabilidad genética (Doran *et al.*, 2016), el hecho de que el modelo *ApoE*^{-/-} pertenezca a la misma cepa secuenciada para crear el genoma de referencia (C57BL/6J GRCm38.p4); reduce en gran medida la variabilidad dada por el fondo genético (Wade *et al.*, 2002). De igual forma, la selección de hermanos machos consanguíneos, hace que tanto las condiciones hormonales, como la variabilidad genética otorgada por vía parental sea más homogénea. A su vez, el uso de una estrategia integral como la WES junto al RNA-Seq,

predice el hallazgo de variantes con más certeza y aporta una buena relación funcional de las mismas (Codina-Solà *et al.*, 2015).

Las VSDNs constituyen un campo poco explorado en las ECVs y al igual que las variaciones epigenéticas y perturbaciones ambientales, pueden dirigir la heterogeneidad fenotípica entre las células de un tejido, así como tener implicaciones importantes para el desarrollo normal o de enfermedades. Por tal razón, trabajos como este, evidencian variantes cuyo papel a nivel de blancos moleculares podría identificarse y posiblemente puedan transferirse a la práctica clínica.

Conclusiones

- El modelo de ratón *ApoE*^{-/-} ha sido explorado desde diferentes puntos de vista; por primera vez se ha podido determinar a nivel global, en el genoma codificante (exoma), el conjunto de variantes genéticas posiblemente de origen somático que conllevan al progreso de la enfermedad.
- La distribución cromosómica de las variantes, no es aleatoria y se da preferencialmente en determinados cromosomas. Lo anterior podría implicar que puede haber regiones de susceptibilidad que son blanco de mutaciones asociadas con la aterosclerosis.
- Tanto los genes mapeados diferencialmente por las variantes, así como los diferencialmente expresados; contribuyen al enriquecimiento de procesos y vías de señalización implicadas directa o indirectamente con la aterosclerosis.
- Tanto en humanos como en ratones, los perfiles de expresión de genes en aortas con aterosclerosis pueden estar correlacionados con cambios en la metilación del DNA.

Perspectivas

- Analizar la expresión alélica diferencial con miras a esclarecer la baja concordancia entre WES y RNA-seq.
- Determinar las variantes del RNA-seq que son producto de ediciones del RNA.
- Este trabajo expone la incidencia de las mutaciones somáticas *de novo* en la aterosclerosis murina y desde este punto de vista es importante indagar en un mayor número de muestras su relevancia.

Apéndice

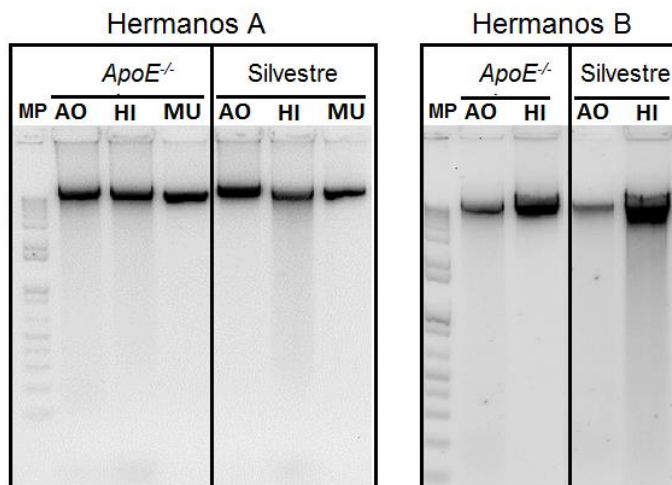


Figura suplementaria 1. DNA de ratones *ApoE*^{-/-} y silvestres utilizado para la WES

Electroforesis de DNA en gel de agarosa al 1 % de las muestras de aorta (AO), hígado (HI) y músculo (MU) de las parejas de hermanos A y B.

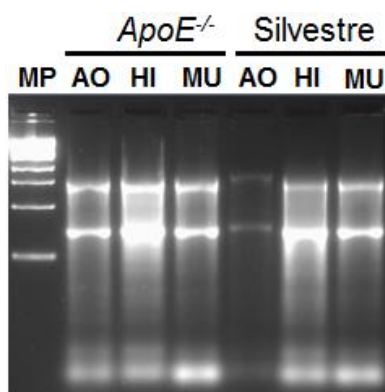


Figura suplementaria 2. RNA de ratones *ApoE*^{-/-} y silvestres utilizado para el RNA-seq

Electroforesis de DNA en gel de agarosa al 1.2 % de las muestras de aorta (AO), hígado (HI) y músculo (MU) de la pareja de hermanos A.

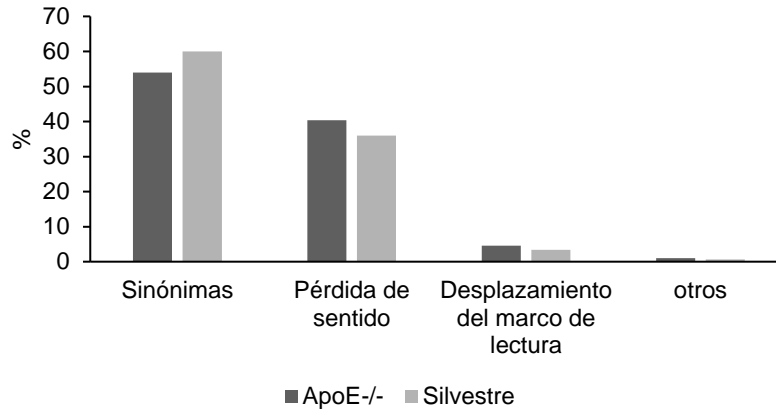


Figura suplementaria 3. Porcentaje de v-WES/ApoE^{-/-} y v-WES/WT con diferentes consecuencias funcionales

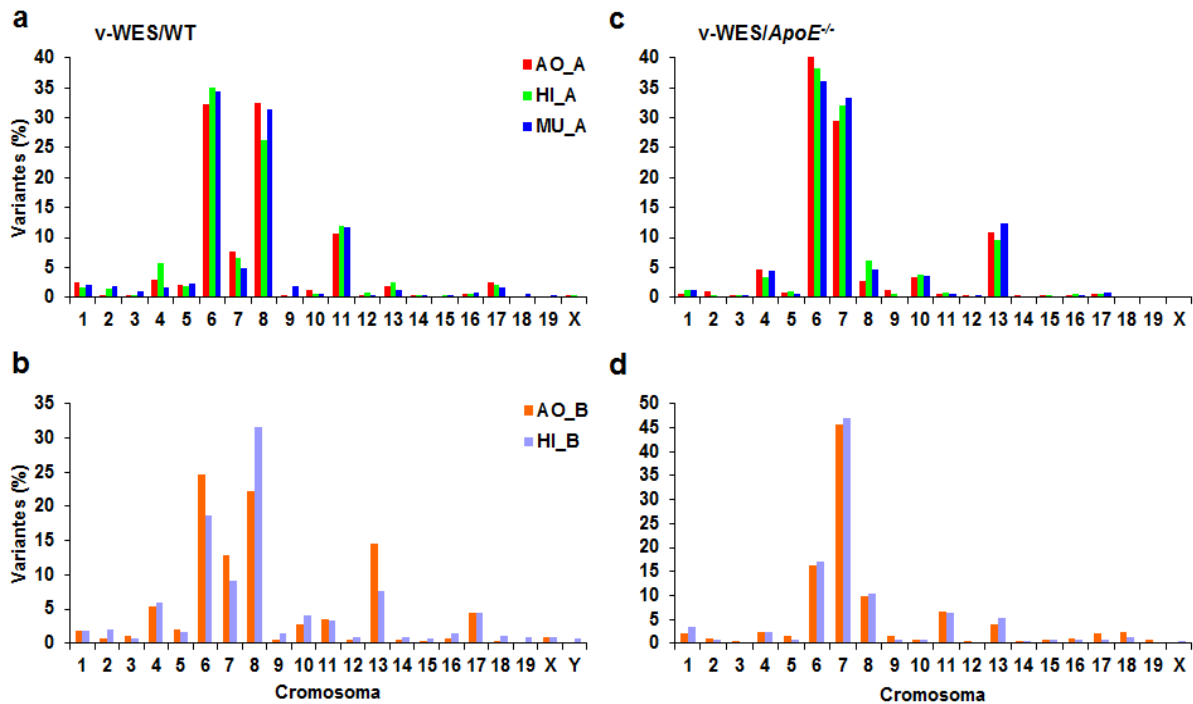


Figura suplementaria 4. Distribución de las v-WES/ApoE^{-/-} y v-WES/WT por cromosoma

Se observan el porcentaje de las v-WES/WT (a, b) y de las v-WES/ApoE^{-/-} (c, d) por cromosoma, en las parejas A (a, c) y B (b, d).

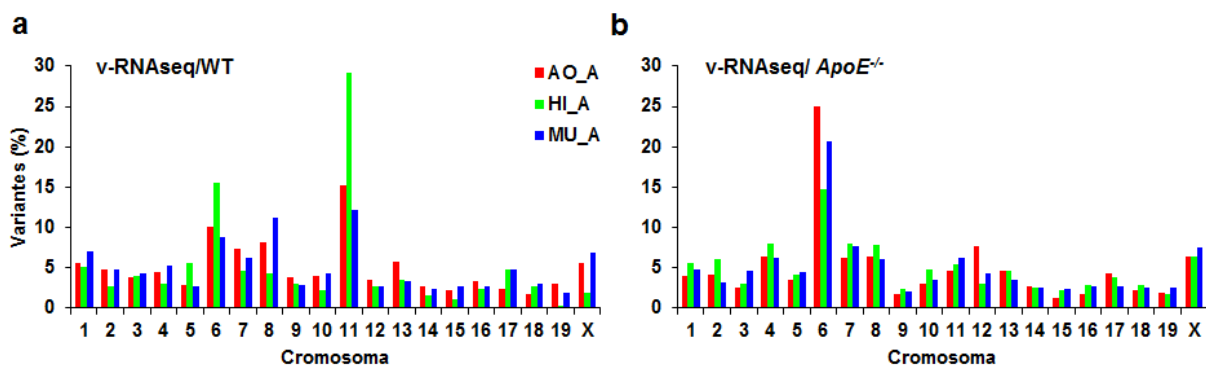


Figura suplementaria 5. Distribución de variantes y genes en el RNA por cromosoma

Se observan, el porcentaje de variantes (a, b) y de genes (c, d) por cromosoma, específicos de los tejidos de ratones *ApoE*^{-/-} y silvestres.

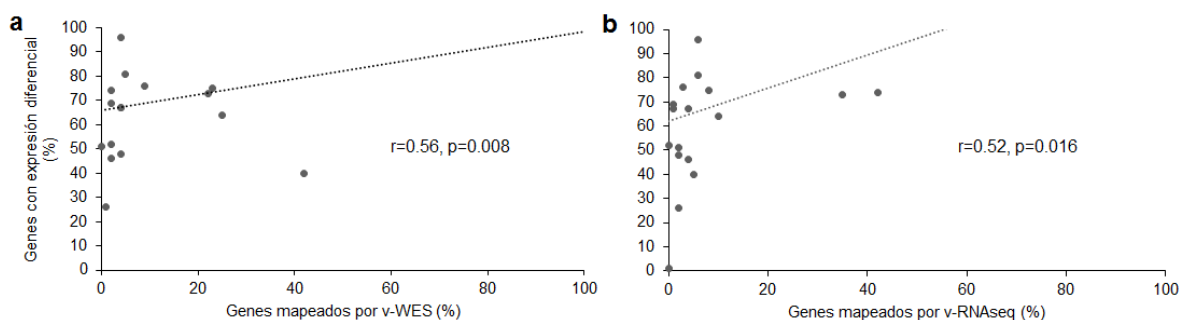


Figura suplementaria 6. Correlación entre el porcentaje de genes mapeados por variantes y la expresión diferencial.

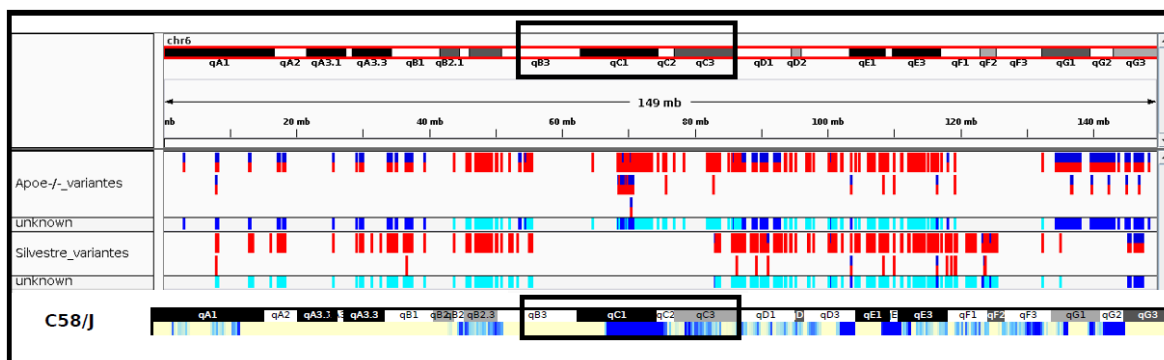


Figura suplementaria 7. Distribución de SNPs a lo largo del cromosoma 6 de aorta de individuos *ApoE*^{-/-} y silvestre (C57BL/6) comparados con la cepa C58/J. En el recuadro negro se identifica la región de comparación.

Tabla suplementaria 1. Número de v-WES/*ApoE*^{-/-} por gen

Cr	Gen	AO_A	HI_A	MU_A	AO_B	HI_B
4	Fhad1	10	<	<	ND	ND
4	Spn	11	13	14	ND	ND
4	Thrap3	10	16	12	15	16
6	Gm6614	10	<	<	<	<
6	Igkv12-41	<	10	11	ND	<
6	Igkv3-12	21	18	20	<	<
6	Igkv3-4	17	15	14	ND	ND
6	Igkv4-50	10	10	10	ND	ND
6	Igkv4-57-1	11	10	11	<	<
6	Igkv4-59	10	<	<	<	<
6	Igkv4-81	13	11	10	ND	<
6	Igkv4-91	11	11	11	ND	<
6	Igkv6-23	11	11	11	<	<
7	Ceacam20	<	<	<	10	10
7	Ceacam3	<	<	11	<	10
7	Fbxo46	<	10	10	<	<
7	Nlrp4e	10	10	13	<	10
7	Psg20	<	<	<	<	18
7	Psg22	<	<	<	<	14
7	Psg26	<	12	11	<	16
7	Psg28	12	10	<	<	<
7	Vmn2r44	<	<	11	<	<
8	Mylk3	<	<	<	ND	10
8	Thap1	<	15	<	<	17
11	Fat2	ND	ND	ND	<	11
13	Cntnap3	11	15	11	<	<
16	Muc4	<	<	10	<	13

Conteo de v-WES por gen en tejidos de ratones *ApoE*^{-/-}. Las v-WES fueron anotadas tanto en regiones codificantes como no codificantes. ND = variante no determinada, < = menos de 10 variantes.

Tabla suplementaria 2. Número de v-WES/WT por gen

Cr	Gen	AO_A	HI_A	MU_A	AO_B	HI_B
6	A2m	13	11	11	ND	ND
6	Ankrd26	21	21	18	ND	ND
6	Bms1	11	10	<	ND	ND
6	Cacna1c	12	11	<	ND	ND
8	Chd9	11	11	<	ND	ND
8	Cilp2	<	ND	<	ND	12
11	Fat2	10	<	<	<	<
11	Hmmr	<	11	<	<	ND
8	Ints10	<	<	<	10	<
16	Muc4	<	<	<	10	<
6	Olf212	ND	ND	ND	14	<
6	Pdzrn3	<	ND	<	<	10
8	Tbc1d9	<	<	<	10	<
6	Vmn2r22	11	10	10	ND	ND
6	Vmn2r24	15	14	14	ND	ND
6	Vmn2r27	16	20	17	ND	ND
17	Vmn2r114	<	<	<	26	<
17	Vmn2r115	<	<	<	24	<
17	Vmn2r117	<	<	<	18	19
13	Zfp457	ND	ND	ND	11	ND

Conteo de v-WES por gen en tejidos de ratones silvestres. Las v-WES fueron anotadas tanto en regiones codificantes como no codificantes. ND = variante no determinada, < = menos de 10 variantes.

Tabla suplementaria 3. Análisis de enriquecimiento GO y KEGG de las v-RNAseq/ApoE^{-/-} de aorta

Category	Term	PValue	FDR	Genes
GO_BP	apoptotic process	0.0066	9.807575	DDX47, C5AR1, HTRA2, MAP1S, MFSD10, ZFP369, PPP1R13L, PDCD4, DEDD2, ELMO1, BLCAP, AKT2
	negative regulation of transcription, DNA-templated	0.0197	26.51589	CDKN1B, MDFIC, RELB, GATAD2A, BCL3, ARHGAP35, SPEN, LOXL3, ATOH8, PDCD4
	B cell differentiation	0.0259	33.42626	LYL1, JAK3, IGKC, GIMAP1
	dichotomous subdivision of terminal units involved in salivary gland branching	0.0412	47.90543	PLXNA1, PLXND1
	phagocytosis, engulfment	0.0461	51.93062	IGHV1-84, IGKC, ELMO1
	cellular calcium ion homeostasis	0.0479	53.3039	ATP13A1, CHERP, APOE, DMPK
	apoptotic nuclear changes	0.0492	54.27637	DEDD2, BLCAP
	cytosol	0.0011	1.338466	C5AR1, MAT2A, STRAP, CMAS, NFKBIB, RELB, ARHGAP35, PDE3A, ATP6V1B2, PDCD4, ARFGEF1, XPNPEP1, DMPK, CDKN1B, HTRA2, MAP1S, SYDE1, VAMP8, HK3, HEBP1, LRP6, VPS35, JAK3, AKT2
	mitochondrion	0.0059	7.186988	LDHB, OPA3, APOOL, SLC27A1, GCDH, SHMT2, GTPBP3, IMMT, LYRM5, GARS, PISD, DMPK, ABCC9, GADD45GIP1, HTRA2, KCNJ8, VAMP8, ACOT10, HK3, HEBP1, MRPL19, YME1L1, CSL, HSPA5, MGST1
	cytoplasm	0.0074	8.854041	SLC27A1, LDHB, CHMP3, CHERP, PLXNA1, PLEKHM2, PDCD4, DMPK, TMF1, DDI2, APOE, MDFIC, ATOH8, AKT2, RBL2, ANKS1, RELB, STRN4, FBLIM1, RNASEH2A, DCTN1, ELMO1, WDR83, DOK1, VAMP8, DACT3, FRMD4B, D930015E06RIK, UCHL4, FARSA, ZFP369, EIF2AK3, CCDC8, ELL, STRAP, NFKBIB, MYO9B, KLC3, ARHGAP35, WBP11, ATP6V1B2, ARFGEF1, MTM1, CHD9, HJURP, AHNAK2, BCL3, VPS35, HSPA5, CC2D1A, IQSEC1, SNX20, SHMT2, GARS, THUMPD3, SFI1, PPP1R13L, XPNPEP1, STAT2, SAFB2, CDKN1B, EPS8, MAP1S, HEBP1, CAPG, JAK3, MERTK, CDK20
GO_CC	membrane	0.0077	9.222774	SLC27A1, LDHB, CHMP3, CHERP, PLXNA1, BCAM, VPS37C, DMPK, TMF1, EFHD2, ART4, HTRA2, ST3GAL5, RNF103, APOE, LOXL3, AKT2, C5AR1, CMAS, STRN4, PISD, MOGS, DCTN1, LRRC25, ELMO1, VAMP8, D930015E06RIK, VAMP5, MFSD10, YME1L1, FARSA, EIF2AK3, GRAMD1A, EMP1, BLCAP, GGCX, APOOL,

				AP1M1, TPRA1, HFE, MYO9B, ATP6V1B2, EDEM1, ARFGEF1, MTM1, DDX47, SLC39A6, VPS35, CC2D1A, HSPA5, PLXND1, FKRP, IQSEC1, SNX20, B4GALNT1, IMMT, PDE3A, TMEM51, DAGLB, GIMAP1, CLPTM1, KCNN4, ABCC9, ATP13A1, EPS8, FAM151A, KCNJ8, LRP6, JAK3, MERTK, NOMO1, MGST1, RSC1A1
	early endosome	0.0170	19.36943	CHMP3, VAMP8, APOE, LRP6, HFE, VPS35
	late endosome	0.0175	19.84998	MTM1, CHMP3, APOE, VAMP5, VPS35
	cytoplasmic vesicle membrane	0.0182	20.58505	VAMP5, MFSD10, AHNAK2
	ruffle	0.0238	26.09077	MTM1, FRMD4B, MYO9B, ATP6V1B2
	ATP-sensitive potassium channel complex	0.0248	26.99437	ABCC9, KCNJ8
	endoplasmic reticulum	0.0370	37.7333	GGCX, SLC27A1, CHERP, MOGS, EDEM1, DMPK, GIMAP1, TMF1, ATP13A1, RNF103, APOE, D930015E06RIK, LRP6, SLC39A6, HSPA5, EIF2AK3, FKRP, MGST1
	mitochondrial inner membrane	0.0389	39.19991	GCDH, SLC27A1, APOOL, SHMT2, IMMT, PISD, YME1L1, MGST1
	MICOS complex	0.0489	46.70466	APOOL, IMMT
	protein phosphatase binding	0.0054	7.237754	KCNN4, CDKN1B, JAK3, EIF2AK3, PIK3R2
	identical protein binding	0.0065	8.579853	LDHB, SHMT2, PPP1R13L, STAT2, SAFB2, ABCC9, HJURP, DACT3, APOE, MAP1S, LRP6, EIF2AK3, MGST1
	antigen binding	0.0147	18.51966	IGHV5-17, HFE, IGHV1-84, IGKC
	ATP binding	0.0314	35.60072	MAT2A, GARS, MYO9B, ATP6V1B2, DMPK, CHD9, ABCC9, DDX47, ATP13A1, KCNJ8, HK3, YME1L1, HSPA5, JAK3, FARSA, MERTK, EIF2AK3, CDK20, AKT2
GO_MF	protein binding	0.0407	43.68933	AP1M1, CHERP, PLXNA1, STRAP, NFKBIB, WBP11, KLC3, EDEM1, PDCD4, DMPK, TMF1, EFHD2, MTM1, GADD45GIP1, HTRA2, MDFIC, APOE, ACAN, BCL3, VPS35, HSPA5, PLXND1, AKT2, PIK3R2, RBL2, IMMT, RELB, GARS, SPEN, PPP1R13L, DCTN1, STAT2, ELMO1, DOK1, CDKN1B, EPS8, SYDE1, VAMP8, DACT3, FRMD4B, LRP6, JAK3, ZFP369, TBX18, CDK20, EIF2AK3, RSC1A1
	Measles	0.0001	0.163745	DOK1, CDKN1B, NFKBIB, JAK3, EIF2AK3, STAT2, AKT2, PIK3R2
	Epstein-Barr virus infection	0.0097	11.08445	CDKN1B, NFKBIB, RELB, JAK3, EIF2AK3, AKT2, PIK3R2
KEGG	Central carbon metabolism in cancer	0.0172	18.8971	HK3, G6PD2, AKT2, PIK3R2
	Chemokine signaling pathway	0.0248	26.14955	NFKBIB, JAK3, STAT2, ELMO1, AKT2, PIK3R2
	Biosynthesis of antibiotics	0.0377	37.06188	LDHB, GCDH, SHMT2, HK3, CSL, G6PD2
	Carbohydrate digestion and absorption	0.0499	46.04962	HK3, AKT2, PIK3R2

Tabla suplementaria 4. Análisis de enriquecimiento GO y KEGG de las v-RNAseq/ApoE^{-/-} de hígado

Categoría	Term	PValue	FDR	Genes
	bile acid metabolic process	0.0122	11.26	SLCO1B2, SLCO1A4, SLCO1A1, SLC27A5
	protein transport	0.0701	36.40	CHMP3, AP2S1, TOMM40, CCDC91, NAPA, VPS41, MON2, HOOK2, AKTIP, VAMP8, ATG7, VPS35, TNPO2, COPE, PIK3R2
	drug transport	0.0011	1.69	SLCO1B2, SLC17A3, ABCC3
	transport	0.0050	7.57	SCN1B, CHMP3, AP2S1, HFE, CCDC91, NAPA, HOOK2, SLCO1A4, SLCO1A1, APOE, ATG7, SLC35D2, VPS35, DYNC1H1, TNPO2, PIK3R2, TOMM40, VPS41, MON2, ITPR1, DCTN1, SLCO1B2, ATP6V0E2, ATP13A1, AKTIP, KCNJ8, VAMP8, SLC17A1, ABCC3, FABP1, CLCN4, COPE
	tRNA aminoacylation for protein translation	0.0050	7.62	GARS, FARSA, IARS2, MARS
GO_BP	protein deubiquitination	0.0058	8.81	CYLD, USP39, USP38, UCHL4, USP15
	endosome to lysosome transport	0.0064	9.69	HOOK2, CHMP3, AKTIP, VPS41
	RNA secondary structure unwinding	0.0075	11.20	DDX39, DDX47, DDX49, DDX21
	drug transmembrane transport	0.0081	12.04	SLCO1B2, SLC17A3, ABCC3
	vesicle-mediated transport	0.0129	18.51	KXD1, VAMP8, AP2S1, VAMP5, NAPA, VPS41, COPE
	protein homotetramerization	0.0179	24.73	HSD17B10, ALDH5A1, DHPS, VASP
	sodium-independent organic anion transport	0.0229	30.61	SLCO1B2, SLCO1A4, SLCO1A1
	negative regulation of transcription, DNA-templated	0.0367	44.49	NACC1, PSMC5, CDKN1B, MDFIC, SALL1, TRIM28, GATAD2A, BRD7, BCL3, ARHGAP35, ATOH8
	epoxygenase P450 pathway	0.0371	44.85	CYP2A22, CYP2B10, CYP2A4
	cytoplasm	0.0378	45.46	IER2, MRPS35, CHMP3, PLEKHM2, IARS2, DMPK, DDI2, HOOK2, MKLN1, MDFIC, APOE, ATG7, FGFR1OP2, NUDT16L1, BRD7, PRKACA, ATOH8, DYNC1H1, USP15, SDR9C7, SYMPK, RBM42, RBL2, SUCLG1, POLR1A, PKN1, RNASEH2A, VASP, DCTN1, PRKD2, DDX39, PGLS, VAMP8, FRMD4B, UCHL4, PAF1, FARSA, ZFP369, AKR1D1, COPE, PPP5C, FGFR4, NKD1, ASUN, ELL, STRAP, GK2, NFKBIB, MYO9B, GIPC1, ARHGAP35, WBP11, CYLD, ISYNA1, HJURP, GYS2, BCL3, VPS35, HSPA5, CC2D1A, TNPO2, IQSEC1, MARS, INF2, SERGEF, NACC1, UPF1, GARS, TOMM40,
GO_CC				

extracellular exosome	0.0421	49.19	THUMPD3, SFI1, ANXA4, SAMD1, XPNPEP1, ITPR1, STAT2, CDKN1B, PSMC5, AKTIP, SLC17A3, MAP1S, IRF7, SALL1, HEBP1, CALM3, CHN2, HABP4, JAK1, RAP1B, FABP1, ACADSB, TM7SF3, CHMP3, ABHD8, GK2, LTBP4, GIPC1, NAPA, AGMAT, ECE1, APOE, PVRL2, PRKACA, VPS35, CC2D1A, SEPP1, HSPA5, MAN2B1, DYNC1H1, MARS, LPL, RBL2, SUCLG1, ADGRE5, GARS, TOMM40, MGP, MOGS, H2AFJ, H2-Q6, ANXA4, MON2, VASP, XPNPEP1, PGLS, PSMC5, VAMP8, HEBP1, VAMP5, FABP1, RAP1B, PCYOX1, GDF15, AKR1D1
HOPS complex	0.0013	1.67	HOOK2, AKTIP, VPS41
mitochondrion	0.0026	3.34	MRPS35, HSD17B10, ACADSB, GK2, PRDX6B, TRMT1, IARS2, AGMAT, DMPK, GADD45GIP1, HTRA2, PRKACA, HSPA5, CSL, MARS, OPA3, RPUSD3, ALDH5A1, IMMT, SUCLG1, LYRM5, GARS, TOMM40, PISD, KCNJ8, PTCDC3, VAMP8, COA7, HEBP1, MGST1
nucleolus	0.0061	7.61	RBL2, TRIM28, POLR1A, WBP11, THUMPD3, ITPR1, DDX47, AKTIP, MAP1S, MDFIC, HJURP, DDX21, ZFP330, DHX40, SDR9C7, MARS
early endosome	0.0084	10.34	CHMP3, ECE1, VAMP8, APOE, HFE, VPS35, VPS41
very-low-density lipoprotein particle	0.0091	11.24	LPL, APOE, PCYOX1
endoplasmic reticulum	0.0137	16.42	GGCX, EVA1A, HSD17B10, FGFR4, MGP, MOGS, CYP2B10, H2-Q6, EDEM1, COLGALT1, ITPR1, TMCC1, DMPK, DOLK, ATP13A1, RNF103, APOE, HSPA5, PTDSS1, FKRP, MGST1, SLC27A5, TM6SF2
cytosol	0.0161	19.00	STRAP, CMAS, NFKBIB, VPS41, ARHGAP35, GIPC1, XPNPEP1, ITPR1, DMPK, CYLD, MKLN1, CDKN1B, HTRA2, MAP1S, VAMP8, ATG7, HEBP1, GYS2, FABP1, RAP1B, VPS35, AKR1D1, PPP5C, REPIN1
midbody	0.0220	25.13	CYLD, CHMP3, PKN1, ZFP330, HSPA5
endoplasmic reticulum membrane	0.0276	30.47	SLC17A3, MOGS, HSPA5, PTDSS1, TM6SF2, ITPR1, DMPK
integral component of endoplasmic reticulum membrane	0.0278	30.66	HSPA5, EDEM1, SLC27A5, DOLK
membrane	0.0361	38.01	CHMP3, TM7SF3, QPCTL, AP2S1, TMEM161A, KLHDC7A, DMPK, EFHD2, HTRA2, ST3GAL5, RNF103, APOE, SLC35D2, DDX21, PRKACA, DYNC1H1, PTDSS1, SYMPK, EVA1A, CMAS, ADGRE5, PKN1, PISD, VPS41, MOGS, CYP2B10, VASP, DCTN1, TMEM150A, PRKD2, DDX39, VAMP8, VAMP5, PAF1, FARSA, TM6SF2, SLC27A5, COPE, PPP5C, GGCX, FGFR4, NKD1, SCN1B, GK2, TPRA1, TMEM82, HFE, NAPA,

			MYO9B, CCDC91, GIPC1, EDEM1, COLGALT1, TMCC1, CYLD, SLCO1A4, SLCO1A1, DDX47, ECE1, PVRL2, VPS35, CC2D1A, HSPA5, FKRP, IQSEC1, MARS, LPL, IMMT, TOMM40, H2-Q6, ITPR1, DOLK, CLPTM1, SLCO1B2, PSMC5, ATP6V0E2, ATP13A1, AKTIP, KCNJ8, SLC17A1, ABCC3, FAM234B, CHN2, JAK1, RAP1B, RSC1A1, CLCN4, MGST1	
	late endosome	0.0394	40.69	CHMP3, APOE, VAMP5, VPS35, VPS41
	nucleoplasm	0.0406	41.66	IER2, FGFR4, WBP11, CCDC91, DDI2, MDFIC, GATAD2A, BCL3, SYMPK, UPF1, EIF1AD, TRIM28, GARS, TOMM40, IRF2BP1, RNASEH2A, JUNB, PRKD2, IRF7, RFX1, HABP4, FABP1, PAF1, RSC1A1, COPE, PPP5C
	poly(A) RNA binding	0.0411	42.05	MRPS35, HSD17B10, ZC3H4, STRAP, TRMT1, WBP11, LSM14B, DDX47, DDX49, NUDT16L1, DDX21, DYNC1H1, RBM28, RBM42, UPF1, IMMT, RPU3D3, SUCLG1, TRIM28, H2-Q6, DDX39, R3HDM2, PTC3, FARSA, DHX40, REPIN1
	ATP-dependent RNA helicase activity	0.0496	48.35	DDX39, DDX47, UPF1, DDX49, DDX21, DHX40
	organic anion transmembrane transporter activity	0.0005	0.75	SLCO1B2, SLCO1A4, SLCO1A1, SLC17A3
	bile acid transmembrane transporter activity	0.0008	1.11	SLCO1B2, SLCO1A4, SLCO1A1
	ATP binding	0.0017	2.42	FGFR4, GK2, MYO9B, IARS2, DMPK, DDX47, DDX49, PRKACA, DDX21, HSPA5, DYNC1H1, MARS, UPF1, GARS, PKN1, DDX39, PRKD2, PSMC5, ATP13A1, KCNJ8, ABCC3, JAK1, DHX40, FARSA, CLCN4, SLC27A5, PPP5C
GO_MF	ubiquitinyl hydrolase activity	0.0056	7.68	CYLD, USP39, USP38, UCHL4, USP15
	aminoacyl-tRNA ligase activity	0.0070	9.50	GARS, FARSA, IARS2, MARS
	drug transmembrane transporter activity	0.0091	12.19	SLCO1B2, SLC17A3, ABCC3
	helicase activity	0.0094	12.64	DDX39, DDX47, UPF1, DDX49, DDX21, DHX40
	hydrolase activity	0.0109	14.55	ABHD8, PLBD1, AGMAT, CYLD, DDX47, ECE1, HTRA2, DDX49, USP38, NUDT16L1, DDX21, MAN2B1, USP15, LPL, UPF1, MOGS, RNASEH2A, XPNPEP1, XRCC6BP1, DDX39, PGLS, ATP13A1, ATP6V0E2, PSMC5, UCHL4, DHX40, PPP5C
	ligase activity	0.0139	18.10	RNF103, SUCLG1, TRIM28, GARS, FARSA, IARS2, IRF2BP1, RNF181, SLC27A5, MARS
	RNA binding	0.0215	26.74	RBM42, UPF1, EIF1AD, RPU3D3, TRMT1, THUMP3, RNASEH2A, DDX47, DDX49, PTC3, NUDT16L1, DDX21, RBM28, MARS, PPP5C
	transcription factor binding	0.0237	29.03	PSMC5, MDFIC, ATG7, BRD7, BCL3, ATOH8, IRF2BP1, JUNB, PIK3R2

KEGG	sodium-independent organic anion	0.0354	40.25	SLCO1B2, SLCO1A4, SLCO1A1
	transmembrane transporter activity			
	ubiquitin-specific protease activity	0.0354	40.30	CYLD, USP38, UCHL4, USP15
	Bile secretion	0.0383	42.77	SLCO1B2, SLCO1A4, SLCO1A1, ABCC3, PRKACA, SLC27A5
	Biosynthesis of antibiotics	0.0468	49.62	ALDOART1, HSD17B10, PGLS, ISYNA1, SUCLG1, PRPS1L3, PCYOX1, CSL, G6PD2
	Pentose phosphate pathway	0.0015	1.88	ALDOART1, PGLS, PRPS1L3, G6PD2
	Carbon metabolism	0.0042	5.18	ALDOART1, PGLS, SUCLG1, PRPS1L3, CSL, G6PD2
	Platelet activation	0.0053	6.42	VAMP8, PRKACA, ARHGAP35, RAP1B, VASP, PIK3R2
	Measles	0.0123	14.36	CDKN1B, NFKBIB, IRF7, JAK1, STAT2, PIK3R2
	Metabolic pathways	0.0188	21.08	ALDOART1, HSD17B10, ACADSB, ALDH5A1, CMAS, GK2, SUCLG1, POLR1A, PISD, MOGS, AGMAT, CYP2B10, DOLK, ISYNA1, PGLS, ATP6V0E2, ST3GAL5, PRPS1L3, CSL, PTDSS1, AKR1D1, G6PD2, SLC27A5, CYP2A4
	Epstein-Barr virus infection	0.0217	24.02	PSMC5, CDKN1B, NFKBIB, JAK1, PRKACA, H2-Q6, PIK3R2
	Long-term potentiation	0.0227	24.94	CALM3, PRKACA, RAP1B, ITPR1
Aminoacyl-tRNA biosynthesis	0.0423	41.76	GARS, FARSA, IARS2, MARS	

Tabla suplementaria 5. Análisis de enriquecimiento GO y KEGG de las v-RNAseq/ApoE^{-/-} de músculo

Category	Term	PValue	FDR	Genes
GO_BP	vesicle fusion	0.0029	4.03	CAV2, VAMP8, VAMP5
	regulation of cholesterol transport	0.0110	14.38	LRP1, APOE
	positive regulation by host of viral process	0.0254	30.40	CAV2, APOE
	regulation of glucose import	0.0254	30.40	HK2, RTN2
	rRNA transcription	0.0432	46.27	CD3EAP, MARS
	negative regulation of transcription from RNA polymerase II promoter	0.0452	47.81	AEBP2, IFI27, STRAP, SPEN, CC2D1A, IRF2BP1, JUNB
	mitotic cell cycle arrest	0.0467	48.99	CDKN1B, GADD45GIP1
	mitochondrion	0.0000	0.02	LDHB, OPA3, IMMT, SUCLG1, LYRM5, GARS, HK2, PISD, REEP1, CD3EAP, ABCC9, GADD45GIP1, KCNJ8, VAMP8, ACOT10, HEBP1, CSL, FAM32A, MARS
GO_CC	early endosome	0.0041	4.68	CHMP3, VAMP8, APOE, HFE, VPS35
	late endosome	0.0094	10.34	CHMP3, APOE, VAMP5, VPS35
	ATP-sensitive potassium channel complex	0.0109	11.92	ABCC9, KCNJ8
	extracellular exosome	0.0213	22.05	LDHB, CHMP3, SUCLG1, ADGRE5, GARS, MGP, PLVAP, H2AFJ, SPEN, VAMP8, APOE, HEBP1, VAMP5, VPS35, CC2D1A, PRSS23, MARS
	cytosol	0.0274	27.52	CAV2, MKLN1, CDKN1B, GIMAP6, MAT2A, STRAP, VAMP8, HEBP1, HK2, VPS35, REPIN1
	endosome	0.0438	40.47	LRP1, CHMP3, CDKN1B, VAMP8, APOE, VPS35
	poly(A) RNA binding	0.0149	17.00	CD3EAP, LRP1, STRAP, IMMT, SUCLG1, WBP11, SPEN, FAM32A, RBM28, REPIN1
GO_MF	transcription corepressor activity	0.0164	18.57	AEBP2, SMYD1, SPEN, IRF2BP1
	SNARE binding	0.0185	20.72	CAV2, VAMP8, VAMP5
	syntaxin binding	0.0317	32.92	CAV2, ABCC9, VAMP8
	Biosynthesis of antibiotics	0.0014	1.40	LDHB, SUCLG1, HK2, PRPS1L3, CSL, AMPD1
KEGG	Carbon metabolism	0.0100	9.99	SUCLG1, HK2, PRPS1L3, CSL
	Biosynthesis of amino acids	0.0375	32.89	MAT2A, PRPS1L3, CSL

Tabla suplementaria 6. Vías de señalización enriquecidas (KEGG_PATHWAY) por genes diferencialmente expresados

Tejido/ expresión	Vía de señalización	Genes	Valor p	FDR
Aorta/ aumenta	Lysosome	41	7.40E-20	8.81E-17
	Chemokine signaling pathway	38	7.90E-11	9.40E-08
	B cell receptor signaling pathway	24	2.77E-10	3.30E-07
	Natural killer cell mediated cytotoxicity	28	4.70E-09	5.60E-06
	Cytokine-cytokine receptor interaction	41	1.09E-08	1.29E-05
	Fc gamma R-mediated phagocytosis	23	1.01E-07	1.20E-04
	ECM-receptor interaction	21	1.12E-07	1.33E-04
	Hematopoietic cell lineage	20	6.75E-07	8.04E-04
	Leukocyte transendothelial migration	24	9.07E-07	0.00108
	Cell adhesion molecules (CAMs)	27	2.64E-06	0.003139
	Other glycan degradation	8	2.52E-05	0.029988
Aorta/ disminuye	Ribosome	33	7.07E-24	8.00E-21
	Oxidative phosphorylation	29	2.03E-14	2.30E-11
	Parkinson's disease	29	3.78E-14	4.27E-11
	Alzheimer's disease	33	1.05E-13	1.19E-10
	Huntington's disease	30	2.48E-11	2.81E-08
	Cardiac muscle contraction	17	2.16E-08	2.45E-05
	Vascular smooth muscle contraction	18	2.11E-06	0.002392
	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	12	1.02E-04	0.115277
	Dilated cardiomyopathy	13	1.58E-04	0.178526
	Hypertrophic cardiomyopathy (HCM)	12	2.87E-04	0.323804
Hígado/ aumenta	Viral myocarditis	4	0.013736	13.02717
	Glutathione metabolism	3	0.031682	27.73929
	Prostate cancer	3	0.084398	58.92646
	Pathways in cancer	5	0.093778	62.98049
Hígado/ disminuye	Alzheimer's disease	6	3.20E-04	0.267077
	Oxidative phosphorylation	5	9.20E-04	0.766683
	Parkinson's disease	5	0.001002	0.834928
	Huntington's disease	5	0.003249	2.68547
Músculo/ aumenta	PPAR signaling pathway	6	1.0E-05	0.046227
	Metabolism of xenobiotics by cytochrome P450	4	0.001	4.518155
	Drug metabolism	4	0.001	6.40278
Músculo/ disminuye	Cardiac muscle contraction	3	0.006174	4.01717
	Viral myocarditis	3	0.00887	5.727879

Bibliografía

Acuna-Hidalgo, R., Bo, T., Kwint, M.P., Van De Vorst, M., Pinelli, M., Veltman, J.A., Hoischen, A., Vissers, L.E.L.M., and Gilissen, C. (2015). Post-zygotic Point Mutations Are an Underrecognized Source of de Novo Genomic Variation. *Am. J. Hum. Genet.* *97*, 67–74.

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., et al. (2015). A global reference for human genetic variation. *Nature* *526*, 68–74.

Bashashati, A., Ha, G., Tone, A., Ding, J., Prentice, L.M., Roth, A., Rosner, J., Shumansky, K., Kalloger, S., Senz, J., et al. (2013). Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.* *231*, 21–34.

Beck, J.A., Lloyd, S., Hafezparast, M., Lennon-Pierce, M., Eppig, J.T., Festing, M.F.W., and Fisher, E.M.C. (2000). Genealogies of mouse inbred strains. *Nat. Genet.* *24*, 23–25.

Berenblum, I., and Shubik, P. (1949). An experimental study of the initiating state of carcinogenesis, and a re-examination of the somatic cell mutation theory of cancer. *Br. J. Cancer* *3*, 109–118.

Besenbacher, S., Liu, S., Izarzugaza, J.M.G., Grove, J., Belling, K., Bork-Jensen, J., Huang, S., Als, T.D., Li, S., Yadav, R., et al. (2015). Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat. Commun.* *6*, 5969.

Blanco-Gómez, A., Castillo-Lluva, S., del Mar Sáez-Freire, M., Hontecillas-Prieto, L., Mao, J.H., Castellanos-Martín, A., and Pérez-Losada, J. (2016). Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes. *BioEssays* *664–673*.

Botstein, D., and Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* *33 Suppl*, 228–237.

Brookes, A.J. (1999). The essence of SNPs. *Gene* *234*, 177–186.

Campbell, C.D., Chong, J.X., Malig, M., Ko, A., Dumont, B.L., Han, L., Vives, L., O’Roak, B.J., Sudmant, P.H., Shendure, J., et al. (2012). Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* *44*, 1277–1281.

Campbell, I.M., Yuan, B., Robberecht, C., Pfundt, R., Szafranski, P., McEntagart, M.E., Nagamani, S.C.S., Erez, A., Bartnik, M., Wiśniowiecka-Kowalik, B., et al. (2014). Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. *Am. J. Hum. Genet.* *95*, 173–182.

Chen, J.M., Stenson, P.D., Cooper, D.N., and Férec, C. (2005). A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Hum. Genet.* *117*, 411–427.

Codina-Solà, M., Rodríguez-Santiago, B., Homs, A., Santoyo, J., Rigau, M., Aznar-Laín, G., Del Campo, M., Gener, B., Gabau, E., Botella, M.P., et al. (2015). Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum

disorders. *Mol. Autism* 6, 21.

Condit, C.M., Achter, P.J., Lauer, I., and Sefcovic, E. (2002). The changing meanings of “mutation:” A contextualized study of public discourse. *Hum. Mutat.* 19, 69–75.

Consortium, 1000 Genomes Project, others, and Africa, W. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.

Cooper, D.N., and Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. *Hum. Genet.* 78, 151–155.

Cooper, D.N., Mort, M., Stenson, P.D., Ball, E. V., and Chuzhanova, N. a (2010). Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum. Genomics* 4, 406–410.

Corder, E.H., Saunders, a M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, a D., Haines, J.L., and Pericak-Vance, M. a (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer’s disease in late onset families. *Science* 261, 921–923.

Cotterman, C.W. (1956). Somatic mosaicism for antigen A2. *Acta Genet. Stat. Med.* 6, 520–521.

Dolinoy, D.C., Weidman, J.R., Waterland, R.A., and Jirtle, R.L. (2006). Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ. Health Perspect.* 114, 567–572.

Doran, A., Adams, D., Hunter, K., Flint, J., Wong, K., and Keane, T. (2016). Deep genome sequencing and variation analysis of 13 inbred mouse strains defines candidate phenotypic alleles, private variation, and homozygous truncating mutations. *bioRxiv* 39131.

Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay, K.D., Gilmore, E.C., Poduri, A., et al. (2012). Single-neuron sequencing analysis of I1 retrotransposition and somatic mutation in the human brain. *Cell* 151, 483–496.

Fairfield, H., Gilbert, G.J., Barter, M., Corrigan, R.R., Curtain, M., Ding, Y., D’Ascenzo, M., Gerhardt, D.J., He, C., Huang, W., et al. (2011). Mutation discovery in mice by whole exome sequencing. *Genome Biol.* 12, R86.

Forsberg, L.A., Rasi, C., Razzaghian, H.R., Pakalapati, G., Waite, L., Thilbeault, K.S., Ronowicz, A., Wineinger, N.E., Tiwari, H.K., Boomsma, D., et al. (2012). Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet.* 90, 217–228.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al. (2007a). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861.

Frazer, K.A., Eskin, E., Kang, H.M., Bogue, M.A., Hinds, D.A., Beilharz, E.J., Gupta, R. V., Montgomery, J., Morenzoni, M.M., Nilsen, G.B., et al. (2007b). A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448, 1050–1053.

Freed, D., and Pevsner, J. (2016). The Contribution of Mosaic Variants to Autism Spectrum

Disorder. *PLOS Genet.* 12, e1006245.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv Prepr. arXiv1207.3907* 9.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* 366, 883–892.

Go, G.-W. (2015). Low-Density Lipoprotein Receptor-Related Protein 6 (LRP6) Is a Novel Nutritional Therapeutic Target for Hyperlipidemia, Non-Alcoholic Fatty Liver Disease, and Atherosclerosis. *Nutrients* 7, 4453–4464.

Gottlieb, B., Beitel, L.K., Alvarado, C., and Trifiro, M.A. (2010). Selection and mutation in the “new” genetics: An emerging hypothesis. *Hum. Genet.* 127, 491–501.

Gottlieb, B., Beitel, L.K., and Trifiro, M. (2014). Changing genetic paradigms: creating next-generation genetic databases as tools to understand the emerging complexities of genotype/phenotype relationships. *Hum. Genomics* 8, 9.

Griffiths, A.J., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. (2000). *An Introduction to Genetic Analysis*, 7th edition.

Gutiérrez, J.P., Rivera Dommarco, J., Shamah Levy, T., Villalpando Hernández, S., Franco, A., Cuevas Nasu, L., Romero Martínez, M., and Ávila, M.H. (2012). Encuesta Nacional de Salud y Nutrición 2012. Resultados Nacionales. (Cuernavaca), p.

Halabi, N.M., Martinez, A., Al-Farsi, H., Mery, E., Puydenus, L., Pujol, P., Khalak, H.G., McLurcan, C., Ferron, G., Querleu, D., et al. (2016). Preferential Allele Expression Analysis Identifies Shared Germline and Somatic Driver Genes in Advanced Ovarian Cancer. *PLoS Genet.* 12.

Huang, A.Y., Xu, X., Ye, A.Y., Wu, Q., Yan, L., Zhao, B., Yang, X., He, Y., Wang, S., Zhang, Z., et al. (2014). Postzygotic single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. *Cell Res.* 24, 1311–1327.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.

Huang, D.W., Lempicki, R. a, and Sherman, B.T. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.

Ikeda, Y., Tsuchiya, H., Hama, S., Kajimoto, K., and Kogure, K. (2013). Resistin affects lipid metabolism during adipocyte maturation of 3T3-L1 cells. *FEBS J.* 280, 5884–5895.

INEGI (2014). *Mujeres y Hombres en México 2013.* 117.

Jacob, J., Kelsoe, G., Rajewsky, K., and Weiss, U. (1991). Intracloonal generation of antibody mutants in germinal centres. *Nature* 354, 389–392.

Jawien, J., Nastalek, P., and Korbut, R. (2004). Mouse models of experimental atherosclerosis. *J. Physiol. Pharmacol.* 55, 503–517.

Kane, K.L., Longo-Guess, C.M., Gagnon, L.H., Ding, D., Salvi, R.J., and Johnson, K.R.

- (2012). Genetic background effects on age-related hearing loss associated with *Cdh23* variants in mice. *Hear. Res.* 283, 80–88.
- Kanoni, S., and Masca, N.G.D. (2016). Analysis with the exome array identifies multiple new independent variants in lipid loci. *0*, 1–13.
- Karki, R., Pandya, D., Elston, R.C., and Ferlini, C. (2015). Defining “mutation” and “polymorphism” in the era of personal genomics. *BMC Med. Genomics* 8, 37.
- Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289–294.
- Kern, M., Kosacka, J., Hesselbarth, N., Brückner, J., Heiker, J.T., Flehmig, G., Klötting, I., Kovacs, P., Matz-Soja, M., Gebhardt, R., et al. (2014). Liver-restricted *Repin1* deficiency improves whole-body insulin sensitivity, alters lipid metabolism, and causes secondary changes in adipose tissue in mice. *Diabetes* 63, 3295–3309.
- Kloosterman, W.P., Francioli, L.C., Hormozdiari, F., Marschall, T., Hehir-Kwa, J.Y., Abdellaoui, A., Lameijer, E.W., Moed, M.H., Koval, V., Renkens, I., et al. (2015). Characteristics of de novo structural changes in the human genome. *Genome Res.* 25, 792–801.
- Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father’s age to disease risk. *Nature* 488, 471–475.
- Labonté, L., Coulombe, P., Zago, M., Bourbeau, J., and Baglole, C.J. (2014). Alterations in the expression of the NF- κ B family member RelB as a novel marker of cardiovascular outcomes during acute exacerbations of chronic obstructive pulmonary disease. *PLoS One* 9, e112965.
- Lee, J.H., Huynh, M., Silhavy, J.L., Kim, S., Dixon-Salazar, T., Heiberg, A., Scott, E., Bafna, V., Hill, K.J., Collazo, A., et al. (2012). De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat. Genet.* 44, 941–945.
- Lek, M., Karczewski, K.J., Samocha, K.E., Banks, E., Fennell, T., O, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Birnbaum, D.P., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* 536, 30338.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., and Cheung, V.G. (2011). Widespread RNA and DNA Sequence Differences in the Human Transcriptome. *Science* (80-.). 333, 53–58.
- Libby, P., Lichtman, A., and Hansson, G. (2013). Immune Effector Mechanisms Implicated in Atherosclerosis: From Mice to Humans. *Immunity* 38, 1092–1104.
- Liu, W., Long, Q., Chen, K., Li, S., Xiang, G., Chen, S., Liu, X., Li, Y., Yang, L., Dong, D., et al. (2013). Mitochondrial metabolism transition cooperates with nuclear reprogramming

during induced pluripotent stem cell generation. *Biochem. Biophys. Res. Commun.* **431**, 767–771.

Lund, G., and Zaina, S. (2011). Atherosclerosis: an epigenetic balancing act that goes wrong. *Curr. Atheroscler. Rep.* **13**, 208–214.

Lund, G., Andersson, L., Lauria, M., Lindholm, M., Fraga, M.F., Villar-Garea, A., Ballestar, E., Esteller, M., and Zaina, S. (2004). DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. *J. Biol. Chem.* **279**, 29147–29154.

Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 961–968.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., et al. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476.

Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature* **456**, 18–21.

Mahley, R.W., Innerarity, T.L., Rall, S.C., and Weisgraber, K.H. (1984). Plasma lipoproteins: apolipoprotein structure and function. **25**.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorf, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.

Matturri, L., Cazzullo, A., Turconi, P., Lavezzi, A.M., Vandone, P.L., Gabrielli, L., Fernández Alonso, G., Grana, D., and Milei, J. (2001). Chromosomal alterations in atherosclerotic plaques. *Atherosclerosis* **154**, 755–761.

McPherson, R., and Tybjaerg-Hansen, A. (2016). Genetics of Coronary Artery Disease. *Circ. Res.* **118**, 564–578.

Meir, K.S., and Leitersdorf, E. (2004). Atherosclerosis in the apolipoprotein E-deficient mouse: A decade of progress. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1006–1014.

Migeotte, I., Riboldi, E., Franssen, J.-D., Grégoire, F., Loison, C., Wittamer, V., Detheux, M., Robberecht, P., Costagliola, S., Vassart, G., et al. (2005). Identification and characterization of an endogenous chemotactic ligand specific for FPRL2. *J. Exp. Med.* **201**, 83–93.

Moore, K.J., and Tabas, I. (2011). Macrophages in the pathogenesis of atherosclerosis. *Cell* **145**, 341–355.

Mozaffarian, D., Benjamin, E.J., Go, A.S., Arnett, D.K., Blaha, M.J., Cushman, M., Das, S.R., de Ferranti, S., Despr??s, J.P., Fullerton, H.J., et al. (2015). Heart Disease and Stroke Statistics???2016 Update: A Report From the American Heart Association.

Munroe, P.B., and Tinker, A. (2015). Genome-wide association studies and contribution to cardiovascular physiology. *Physiol. Genomics* **47**, 365–375.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a

novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476.

Nakashima, Y., Plump, A.S., Raines, E.W., Breslow, J.L., and Ross, R. (1994). ApoE-Deficient Mice Develop Lesions of All Phases of Atherosclerosis Throughout the Arterial Tree. *Arteriosclerosis* **133**–140.

Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St. Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S. -a., Fraser, D., et al. (2012). An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. *Science* (80-). **337**, 100–104.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 1–20.

Nilsson, J., Calara, F., Regnstrom, J., Hultgardh-Nilsson, A., Ameli, S., Cercek, B., and Shah, P.K. (1997). Immunization with homologous oxidized low density lipoprotein reduces neointimal formation after balloon injury in hypercholesterolemic rabbits. *J. Am. Coll. Cardiol.* **30**, 1886–1891.

O’Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250.

Ozaki, K., and Tanaka, T. (2016). Molecular genetics of coronary artery disease. *J. Hum. Genet.* **61**, 71–77.

Palinski, W., Miller, E., and Witztum, J.L. (1995). Immunization of low density lipoprotein (LDL) receptor-deficient rabbits with homologous malondialdehyde-modified LDL reduces atherogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 821–825.

Petersen, B.-S., Spehlmann, M.E., Raedler, A., Stade, B., Thomsen, I., Rabionet, R., Rosenstiel, P., Schreiber, S., and Franke, A. (2014). Whole genome and exome sequencing of monozygotic twins discordant for Crohn’s disease. *BMC Genomics* **15**, 564.

Piedrahita, J.A., Zhang, S.H., Hageman, J.R., Oliver, P.M., and Maeda, N. (1992). Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4471–4475.

Plump, A.S., Smith, J.D., Hayek, T., Aalto-Setälä, K., Walsh, A., Verstuyft, J.G., Rubin, E.M., and Breslow, J.L. (1992). Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell* **71**, 343–353.

Poduri, A., Evrony, G.D., Cai, X., Elhosary, P.C., Beroukhi, R., Lehtinen, M.K., Hills, B.L., Heinzen, E.L., Hill, A., Hill, S.R., et al. (2012). Somatic Activation of AKT3 Causes Hemispheric Developmental Brain Malformations. *Neuron* **74**, 41–48.

Poduri, A., Evrony, G.D., Cai, X., and Walsh, C.A. (2013). Somatic Mutation, Genomic Variation, and Neurological Disease. *Science* (80-). **341**, 1237758.

Puig, O., Yuan, J., Stepaniants, S., Zieba, R., Zycband, E., Morris, M., Coulter, S., Yu, X., Menke, J., Woods, J., et al. (2011). A gene expression signature that classifies human

- atherosclerotic plaque by relative inflammation status. *Circ. Cardiovasc. Genet.* *4*, 595–604.
- Raman, K., Chong, M., Akhtar-Danesh, G.-G., D’Mello, M., Hasso, R., Ross, S., Xu, F., and Paré, G. (2013). Genetic markers of inflammation and their role in cardiovascular disease. *Can. J. Cardiol.* *29*, 67–74.
- Reich, D.E., and Lander, E.S. (2001). On the allelic spectrum of human disease. *Trends Genet.* *17*, 502–510.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* *17*, 405–423.
- Rozhok, A.I., and DeGregori, J. (2015). Toward an evolutionary model of cancer: Considering the mechanisms that govern the fate of somatic mutations. *Proc. Natl. Acad. Sci.* *112*, 201501713.
- Russnes, H.G., Navin, N., Hicks, J., and Borresen-Dale, A.L. (2011). Insight into the heterogeneity of breast cancer through next-generation sequencing. *J. Clin. Invest.* *121*, 3810–3818.
- Sandoval, J., Heyn, H.A., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., and Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* *6*, 692–702.
- Scally, A. (2016). Mutation rates and the evolution of germline structure. *Phil. Trans. R. Soc. Lond. B* *371*, 20150137.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science* (80-.). *305*, 525–528.
- Ségurel, L., Wyman, M.J., and Przeworski, M. (2014). Determinants of Mutation Rate Variation in the Human Germline. *Annu. Rev. Genomics Hum. Genet.* *15*, 47–70.
- Sharma, P., Kumar, J., Garg, G., Kumar, A., Patowary, A., Karthikeyan, G., Ramakrishnan, L., Brahmachari, V., and Sengupta, S. (2008). Detection of Altered Global DNA Methylation in Coronary Artery Disease Patients. *27*, 357–365.
- Shimizu, H., Mitomo, K., Watanabe, T., Okamoto, S., and Yamamoto, K. (1990). Involvement of a NF-kappa B-like transcription factor in the activation of the interleukin-6 gene by inflammatory lymphokines. *Mol. Cell. Biol.* *10*, 561–568.
- Silva-Martinez, G.A., Rodriguez-Rios, D., Alvarado-Caudillo, Y., Vaquero, A., Esteller, M., Carmona, F.J., Moran, S., Nielsen, F.C., Wickstrom-Lindholm, M., Wrobel, K., et al. (2016). Arachidonic and oleic acid exert distinct effects on the DNA methylome. *Epigenetics* *2294*, 1–14.
- Siudeja, K., Nassari, S., Gervais, L., Skorski, P., Lameiras, S., Stolfa, D., Zande, M., Bernard, V., Frio, T.R., and Bardin, A.J. (2015). Frequent Somatic Mutation in Adult Intestinal Stem Cells Drives Neoplasia and Genetic Mosaicism during Aging. *Cell Stem Cell* *17*, 663–

674.

Steeland, S., Timmermans, S., Van Ryckeghem, S., Hulpliau, P., Saeys, Y., Van Montagu, M., Vandenbroucke, R.E., and Libert, C. (2016). Efficient analysis of mouse genome sequences reveal many nonsense variants. *Proc. Natl. Acad. Sci.* *113*, 5670–5675.

Stenson, P.D., Ball, E. V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S.T., Abeyasinghe, S., Krawczak, M., and Cooper, D.N. (2003). Human Gene Mutation Database (HGMD??): 2003 Update. *Hum. Mutat.* *21*, 577–581.

Stevens, G., Dias, R.H., Thomas, K.J.A., Rivera, J.A., Carvalho, N., Barquera, S., Hill, K., and Ezzati, M. (2008). Characterizing the epidemiological transition in Mexico: National and subnational burden of diseases, injuries, and risk factors. *PLoS Med.* *5*, 0900–0910.

Stylianou, I.M., Bauer, R.C., Reilly, M.P., and Rader, D.J. (2012). Genetic basis of atherosclerosis: Insights from mice and humans. *Circ. Res.* *110*, 337–355.

Swirski, F.K., Robbins, C.S., and Nahrendorf, M. (2016). Development and Function of Arterial and Cardiac Macrophages. *Trends Immunol.* *37*, 32–40.

Veltman, J. a, and Brunner, H.G. (2012). De novo mutations in human genetic disease. *Nat. Rev. Genet.* *13*, 565–575.

Vissers, L.E.L.M., De Vries, B.B.A., Osoegawa, K., Janssen, I.M., Feuth, T., Choy, C.O., Straatman, H., Van Der Vliet, W., Huys, E.H.L.P.G., Van Rijk, A., et al. (2003). Array-Based Comparative Genomic Hybridization for the Genomewide Detection of Submicroscopic Chromosomal Abnormalities. *Am. J. Hum. Genet.* *73*, 1261–1270.

Wade, C.M., Kulbokas, E.J., Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K., and Daly, M.J. (2002). The mosaic structure of variation in the laboratory mouse genome. *Nature* *420*, 574–578.

Wang, G., Bonkovsky, H.L., de Lemos, A., and Burczynski, F.J. (2015). Recent insights into the biological functions of liver fatty acid binding protein 1. *J. Lipid Res.* *56*, 2238–2247.

Wang, X., Chen, Q., Pu, H., Wei, Q., Duan, M., Zhang, C., Jiang, T., Shou, X., Zhang, J., and Yang, Y. (2016). Adiponectin improves NF-kappaB-mediated inflammation and abates atherosclerosis progression in apolipoprotein E-deficient mice. *Lipids Health Dis.* *15*, 33.

Waterland, R. a, Travisano, M., and Tahiliani, K.G. (2007). Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *FASEB J.* *21*, 3380–3385.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* *42*.

WHO (2015). Alimentos y bebidas ultraprocesados en América Latina: tendencias, efecto sobre la obesidad e implicaciones para las políticas públicas.

Xiong, X., Zhao, Y., Tang, F., Wei, D., Thomas, D., Wang, X., Liu, Y., Zheng, P., and Sun, Y. (2014). Ribosomal protein S27-like is a physiological regulator of p53 that suppresses genomic instability and tumorigenesis. *Elife* *3*, e02236.

Xu, Q., Dietrich, H., Steiner, H.J., Gown, a M., Schoel, B., Mikuz, G., Kaufmann, S.H., and Wick, G. (1992). Induction of arteriosclerosis in normocholesterolemic rabbits by immunization with heat shock protein 65. *Arterioscler. Thromb.* 12, 789–799.

Zaina, S., Heyn, H., Carmona, F.J., Varol, N., Sayols, S., Condom, E., Ramírez-Ruz, J., Gomez, A., Gonçalves, I., Moran, S., et al. (2014). DNA methylation map of human atherosclerosis. *Circ. Cardiovasc. Genet.* 7, 692–700.

Zaina, S., Gonçalves, I., Carmona, F., Gomez, A., Heyn, H., Mollet, I., Moran, S., Varol, N., and Esteller, M. (2015). DNA methylation dynamics in human carotid plaques after cerebrovascular events. *Arterioscler. Thromb. Vasc. Biol.* 35, 1835–1842.

Zhang, S.H., Reddick, R.L., Piedrahita, J.A., and Maeda, N. (1992). Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *Science* (80-.). 258, 468–471.

Zhao, X., Gao, M., He, J., Zou, L., Lyu, Y., Zhang, L., Geng, B., Liu, G., and Xu, G. (2015). Perilipin1 deficiency in whole body or bone marrow-derived cells attenuates lesions in atherosclerosis-prone mice. *PLoS One* 10, 1–16

Zhong, J., Agha, G., and Baccarelli, A.A. (2016). The Role of DNA Methylation in Cardiovascular Risk and Disease: Methodological Aspects, Study Design, and Data Analysis for Epidemiological Studies. *Circ. Res.* 118, 119–131.