



VNIVERSITAT
DE VALÈNCIA

FACULTAD DE CIENCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR
**PROGRAMA DE DOCTORADO EN BIOMEDICINA Y
BIOTECNOLOGÍA**

**IDENTIFICACIÓN DE VARIANTES GENÉTICAS
RELACIONADAS CON DIABETES MELLITUS
TIPO 2 EN POBLACIÓN ESPAÑOLA**

Tesis Doctoral presentada por:
VERÓNICA LENDÍNEZ TORTAJADA

Dirigida por:
**DR. FELIPE JAVIER CHAVES MARTÍNEZ
DR. JOSÉ TOMÁS REAL COLLADO**

VALENCIA, DICIEMBRE 2019



UNIVERSITAT DE VALÈNCIA

Programa de Doctorado en Biomedicina y Biotecnología

INCLIVA | VLC
Instituto de Investigación Sanitaria

INSTITUTO DE INVESTIGACIÓN SANITARIA-INCLIVA

Unidad de Genómica y Diagnóstico Genético/ Unidad de Genómica y
Diabetes

**IDENTIFICACIÓN DE VARIANTES GENÉTICAS
RELACIONADAS CON DIABETES MELLITUS
TIPO 2 EN POBLACIÓN ESPAÑOLA**

Tesis Doctoral presentada por:
VERÓNICA LENDÍNEZ TORTAJADA

Dirigida por:
DR. FELIPE JAVIER CHAVES MARTÍNEZ
DR. JOSÉ TOMÁS REAL COLLADO

VALENCIA, DICIEMBRE 2019



VNIVERSITAT
D VALÈNCIA



INCLIVA | VLC
Instituto de Investigación Sanitaria

D. FELIPE JAVIER CHAVES MARTÍNEZ, Doctor en Ciencias Biológicas, Licenciado en Biología y especialista en Genética e investigador y director de la Unidad de Genómica y Diagnóstico Genético/Unidad de Genómica y Diabetes del Instituto de Investigación Sanitaria-INCLIVA

Y

D. JOSÉ TOMÁS REAL COLLADO, Doctor en Medicina y Cirugía, Catedrático del Departamento de Medicina de la Facultad de Medicina y Odontología de la Universitat de València, y Médico especialista en Endocrinología y Nutrición del Hospital Clínico Universitario de Valencia.

CERTIFICAN:

Que la presente memoria, titulada **“IDENTIFICACIÓN DE VARIANTES GENÉTICAS RELACIONADAS CON DIABETES MELLITUS TIPO 2 EN POBLACIÓN ESPAÑOLA”**, corresponde al trabajo realizado en la Unidad de Genómica y Diagnóstico Genético/Unidad de Genómica y Diabetes del Instituto de Investigación Sanitaria-INCLIVA bajo nuestra dirección por Dña. **Verónica Lendínez Tortajada**, Licenciada en Bioquímica, reúne los requisitos necesarios para su presentación y lectura como Tesis Doctoral en el Programa de Doctorado en Biomedicina y Biotecnología con el fin de optar al grado de Doctora por la Universitat de València.

Y para que conste donde proceda firman el presente certificado en Valencia, a 4 de diciembre de 2019.

Firmado:

Firmado:

Dr. F. Javier Chaves Martínez

Dr. J. Tomás Real Collado

Este trabajo de tesis doctoral ha sido financiado por el proyecto coordinado "Identificación de variaciones de secuencia y de metilación e hidroximetilación en el exoma asociadas al desarrollo de diabetes tipo 2" (P114/00874), del Instituto de Salud Carlos III, Fondo de Investigaciones Sanitarias, cuyo investigador principal es el Dr. Felipe Javier Chaves Martínez.

"Investigar es ver lo que todo el mundo ha visto y pensar lo que nadie más ha pensado"

Albert Szent-Györgyi

"Lo bueno de la ciencia es que es cierta, creas o no en ella"

Neil deGrasse Tyson

*En memòria dels meus iaïos
Francisca, Antonio i Dolores,
i de la tia Amparo.*

AGRADECIMIENTOS:

*"Yo soy una parte de todo aquello
que he encontrado en el camino"
Alfred Tennyson*

Durante el desarrollo de una tesis doctoral existen muchos altibajos, mucho desconocimiento al principio, muchos experimentos fallidos y pocos exitosos pero revitalizantes que te ayudan a seguir un poco más, ganas de aprender conocimientos, conceptos, técnicas nuevas, esfuerzo personal, sacrificios sociales... todo ello para llegar a un fin y poder continuar con la carrera profesional.

Sin embargo, hay un aspecto del mismo que hace que todo el proceso sea más llevadero e inolvidable: las personas que conoces en el camino.

Personas con más experiencia, con menos, personas presentes durante un corto período o que aún sigues teniendo a tu lado, que te enseñan cuando tienen oportunidad, que te hacen ver que puedes llegar a donde te propongas, que te dan otra visión de la vida, que te animan en tus días malos y celebran contigo tus triunfos, que te hacen sentir parte de un todo y que no estás sola en el camino. Es gracias a todas estas personas que he podido llegar a donde estoy ahora, gracias a ellas conozco un poco más donde están mis límites y mis posibilidades.

Me gustaría empezar dando las gracias a mis dos directores por haber confiado en mí para la realización de este trabajo.

Al Dr. José T. Real por abrirme una puerta hace 6 años y aceptarme como estudiante de máster, por ser uno de mis directores antes y ahora, por ayudarme en todo lo que he necesitado. Pero sobre todo por presentarme al Dr. Javier Chaves, porque conocerlo a él y a la gran familia que es para mí y siempre será la UGDG, ha sido uno de las mejores experiencias de vida que puedo recordar.

A Javier, gracias por darme la oportunidad de trabajar en tu grupo, por creer en mí sin apenas conocerme, por ayudarme en todos los momentos de dificultad, que no han sido pocos, por proponerme hacer esta tesis y sacar adelante éste y otros proyectos, por contar conmigo y tener en cuenta mi opinión, por enseñarme algo nuevo cada día, por ayudarme a aprender y mejorar, por no ser solo mi jefe, sino también un apoyo tanto profesional como personal, gracias.

Porque no todas las familias están unidas por lazos de sangre o apellidos, a veces ocurre que un grupo de personas, al principio desconocidas, pueden

volverse con el tiempo un gran apoyo y formar parte de ti y sin que te des cuenta, pasas a considerarlas familia.

Primero quería agradecer infinitamente a una compañera, a una amiga que ha estado conmigo en todo momento desde que entró hace 4 años, y que sin ella este trabajo no hubiera sido posible; Gracias de todo corazón, Azahara. Por tu inmensa ayuda, por tu paciencia, por tus explicaciones, por permitirme aprender contigo el "maravilloso mundo de la bioinformática"... Gracias también por haber respondido todas mis dudas, mis preguntas repetitivas, por aconsejarme y dejarme trabajar contigo, para mí ha sido un verdadero placer. Eres una persona con un gran potencial y corazón, una buena amiga, divertida y una gran artista. Siempre estaré disponible para todo lo que necesites, no dejes de ser como eres, aún tienes mucho que ofrecer y espero estar ahí para poder ver como consigues todo lo que te propongas.

A Tani, maestro del conocimiento infinito, consejero de vida y amante de las bicis, gracias por acogerme a tu manera, por ayudarme y enseñarme los misterios del laboratorio, por tus charlas y compartir conmigo tus puntos de vista sobre aspectos de la vida, por tus ánimos y apoyo, gracias.

A Victoria, una de les meues mamis del laboratorí. Gràcies per tota la teua ajuda al laboratorí i fora d'ell, al principi sobretot, i durant tota la teua estada. Gràcies pels teus bons consells, xarrades i carinyo, per haver-me deixat formar part d'eixa família tan bonica que tens, per haver-te preocupat per mi sempre, gràcies de tot cor.

A Cristina, mi otra mami, que siempre has estado ahí para cualquier duda o pregunta, tanto laboral como personal, gracias por tus consejos, por tu sinceridad, por todas las risas compartidas, por tratarme también como alguien más de tu familia, gracias Cris.

A Kike, mi compañero de penurias y alegrías. Por ese mítico verano junioreando, por dejarme aprender contigo, por ser nuestra propia calculadora humana, por tus consejos infinitos, por cuidarme y estar siempre ahí, gracias.

A Griselda, mi otra mami desde el laboratorio 14, gracias por enseñarme y ayudarme en todo lo que he necesitado. Por compartir tu experiencia y tus consejos y también por ser parte de tu familia, gracias.

A las chicas de endocrino, Vero y Charo, por esos comienzos experimentando con técnicas "raras", por todas esas risas y lloros que hemos compartido. Por ser siempre un remanso de paz y ánimo cuando las cosas se ponían tensas, por vuestro alegría y apoyo en todo momento, gracias chicas.

A Laura, mi otra chica de endocrino, aunque entramos a sitios diferentes al principio, poco a poco te convertiste también en otro apoyo incondicional. Gracias por toda tu paciencia, por escucharme y ayudarme, por todos tus consejos claros, por dejarme formar parte de tu familia también, por contar conmigo y por saber que puedo contar contigo, espero que por mucho tiempo, gracias.

A la gente del lab 2, a Rakel, Javi y Loles, porque, aunque no hayamos trabajado en el mismo campo siempre habéis estado ahí para cualquier cosa que he necesitado, para resolver cualquier duda o simplemente compartir momentos de distensión. Gracias.

A Ana O y Olga, aunque hayamos compartido menos tiempo, también habéis sido un gran apoyo en mis últimos meses en el grupo y habéis hecho que la despedida haya sido menos amarga. Ha sido un placer haber coincidido también con vosotras, gracias.

A mis predecesoras Inma y Vanesa, por haberme ayudado y enseñado en mis inicios, en todo lo que he necesitado, por haberme dejado el "futuro" de vuestras investigaciones en mis manos. Gracias por haber compartido conmigo un poquito de vuestro tiempo, gracias.

A Ana B, por tu ayuda cuando la he necesitado y por todo lo que has podido enseñarme, gracias.

A todas esas compañeras que, aunque, discontinuamente, también han dejado su huella en mi camino y me han aportado, ayudado y enseñado. Gracias a Verónica N, a Irene y a las FIR/MIR que han rotado una temporada: Lidia, Blanca F, Ester, Ana R, entre otras personas, que siempre han aportado un poquito de ellas mismas.

A Pablo, Dani y Alba, por esos comienzos en la bioinformática, por mostrarme un mundo nuevo y lleno de posibilidades. Por vuestras enseñanzas básicas y útiles que después he podido ir desarrollando, gracias.

A Miguel, Álex y a los de Big Data, porque sin vosotros, esos días interminables en bioinfo, no habrían sido lo mismo. Por hacer que el ambiente fuera bueno, lleno de risas, consejos y buen humor. Por haber estado pendientes y ayudándonos en todo momento, gracias.

A la gente de prácticas, TFM's TFG's, FPs y proyectos varios que han pasado por el laboratorio y que, con sus ganas de aprender y motivación, me han enseñado y ayudado a ver las cosas de diferentes maneras. A Pilar, Bárbara, Cristina, Tania, Noemí, Dora, Eugenia, Fran, Soraya, Claudia, Cristinass y todos/as los demás que me faltan, gracias.

A Carol, la incorporació, encara que breu, més revolucionària que he conegut. Crec que has canviat la vida de tot el món, jo mateixa inclosa, amb la teua energia, el teu somriure i bon humor, però sobretot, per les ganes que li poses a tot allò que et proposes. Gràcies per haver-me ajudat en tot, amb bons consells i crítiques constructives. Eres una inspiració i una persona que difícilment oblidaré. Gràcies per estar ací i no canviés mai.

A Alicia, nuestra "infiltrada" de hemato favorita. Gracias por alegrar cada día con tus visitas, por tu preocupación y tu ayuda en momentos de crisis, por aconsejarme y compartir buenos momentos que espero que no se acaben nunca, gracias.

Quería agradecer en especial a dos personas que han sido para mí el mejor apoyo que he podido tener durante esta larga etapa y que sin ellas, no hubiera sido lo mismo. Toda buena amistad empieza por casualidad y no por ello tiene menos valor.

Pilar, ¿Quién iba a decir que conversaciones sobre metilación con un océano de distancia iban a conducir a todo lo demás? Has sido para mí, aparte de mi mentora en el laboratorio, una buena amiga que ha estado en todo momento que he necesitado, enseñándome, ayudándome, compartiendo viajes, congresos, actividades varias, alegrías, penas y amigas. Como M^a José, gracias por "dejarme compartir" a Pilar contigo, por tu cariño, por tus consejos, por dejarme ser tu amiga. A las dos, por haberme dejado formar parte de vuestras vidas y vuestras familias y que siga siendo así, gracias.

A veces una conversación medio en broma puede conducir a que sucedan cosas inesperables. Quien me iba a decir que hace más de un año un simple "va sí tú te apuntas yo me apunto" me iba a permitir conocer a personas con una afición común y convertirse en grandes amigos/as con ritmo y diversión.

Gracias a toda mi gente de baile, que habéis llenado de buenos momentos, risas, algún que otro pisotón... pero sobre todo de buena salsa y bachata este último año largo.

A Vicenta, porque si ti esto no hubiera ocurrido, nos conocemos de hace tiempo, pero nunca hubiéramos imaginado que encontraríamos esta gente maravillosa en una academia de baile.

A mis primeros profes Maribel y Alberto, por el buen rollo de las clases y de la música con ¡sabor!

A mis compis, ahora amigos, Javi Núñez y Javi Jr, dos torrentinos con mucho ritmo y ganas de pasarlo bien. A Vicente, a María y a todos los compañeros/as

de antes y ahora por estar siempre dispuestos a echar algunos pasos en la pista.

A Pedro, por llevarme a bailar, por esas charlas y por estar ahí, gracias.

A Marisa, por aparecer en mi vida, por tu cariño, por tus palabras, por tu ejemplo y tu apoyo, por poder contar contigo ahora y espero que siempre, por todo, gracias de corazón.

A Juanlu, por tu motivación, por tu estilo de enseñar y transmitir, porque me ha ayudado últimamente aparte de en el baile. "...hay que saber esperar..." Gracias por confiar en mí y sacarme una sonrisa siempre.

A Ana, porque todo es buena excusa para reencontrarse, salir a bailar y divertirse.

A mis compis de penas y alegrías del master Elisa, Isa y Ester, porque, aunque nuestras quedadas sean difusas en el tiempo, siempre he sabido que estáis ahí para cualquier cosa. Gracias por vuestro apoyo y ánimos sobretodo en estos últimos meses, gracias.

A esas amigas que aun en la distancia, sabes que están ahí, y que un mensaje o llamada bastan para saber que todo sigue igual.

A Lourdes, ambas iniciamos nuestras aventuras hace 6 años y aunque estás a muchos kilómetros, saber de ti y seguir contando contigo, tus consejos y ánimos me ha ayudado siempre. Porque nunca cambie, gracias.

A Eva, amiga de penas y alegrías tésínicas, gracias por esas quedadas de desahogo, risas y socorro. Por preocuparte por mí y porque sigamos en contacto, gracias.

A Blanca, ¿Qué voy a decir de mi "lab partner"? Hemos recorrido parte del camino juntas desde hace mucho y, aunque se haya desviado por senderos diferentes, siempre has estado presente de la manera que fuera y siempre he sabido que puedo contar contigo, gracias.

Y, por último, pero no menos importante, a mi familia porque ellos son el principal apoyo en todo lo que haga, y que sé que siempre tendré.

Gracias a mis padres, Andrés y Paqui, porque ellos son mi modelo a seguir, por todo lo que habéis hecho y hacéis por mí, gracias de corazón.

A mi hermana Débora, mi otra versión, mi apoyo incondicional en todo, te quiero hermanita, gracias.

*A mi tía Mari, por tu cariño, por siempre estar ahí y por tu apoyo, gracias tía.
A mis tíos Rosa y Toni, porque siempre me han ayudado y animado en todo,
gracias.*

A mi iaia Juli y a mi iaio Paco, por cuidarme y quererme tanto, gracias.

*Y al resto de mi familia porque sé que estarán ahí para cualquier cosa que
necesite.*

A todos/as, gracias infinitas.

ÍNDICE

ABREVIATURAS	VII
ÍNDICE DE FIGURAS	XIII
ÍNDICE DE TABLAS	XVIII
RESUMEN.....	1
INTRODUCCIÓN	5
1.- DIABETES MELLITUS.....	7
1.1.- Definición	7
1.2.- Epidemiología.....	7
1.3.- Diagnóstico.....	9
1.4.- Clasificación y tipos de diabetes	12
1.4.1.- Diabetes Mellitus tipo 1 (DM1)	12
1.4.2.- Diabetes Mellitus tipo 2.....	13
1.4.3.- Diabetes latente autoinmune en el adulto (<i>LADA</i>).....	13
1.4.4.- Diabetes Mellitus Gestacional (<i>GDM</i>).....	13
1.4.5.- Otros tipos específicos de diabetes:.....	14
1.4.5.1.- Defectos genéticos de la función de las células β -pancreáticas	15
1.4.5.2.- Diabetes Mellitus Neonatal (DMN)	15
1.4.5.3.- Diabetes Mellitus Mitocondrial (DMM)	16
1.4.5.4.- Diabetes Mellitus de la juventud de aparición en la madurez (MODY)	16
1.4.5.5.- Defectos genéticos en la acción de la insulina.....	17
1.4.5.6.- Enfermedades del páncreas exocrino	17
1.4.5.7.- Endocrinopatías	17
1.4.5.8.- Diabetes asociada a fármacos o químicos.....	18
1.4.5.9.- Diabetes asociada a infecciones	18
1.4.5.10.- Otros síndromes genéticos asociados con diabetes	18
1.5.- Diabetes Mellitus tipo 2 (DM2).....	19
1.5.1.- Prevalencia de la DM2.....	19
1.5.2.- Factores de riesgo ambientales evolución y prevención de la DM2	21
1.5.2.1.- Factores de riesgo ambientales	21
1.5.2.2.- Complicaciones a largo plazo de la DM2	24
1.5.2.3.- Prevención de la DM2.....	25
1.5.3.- Fisiopatología de la DM2	26
1.5.3.1.- Alteración de las células β -pancreáticas.....	28
1.5.3.2.- Resistencia a la insulina	29
1.5.4.- Factores genéticos de la DM2	29
1.5.4.1.- Estudios de ligamiento en familias y genes candidatos	31
1.5.4.2.- Estudios de asociación de genoma completo (GWAS)	32
1.5.4.3.- Estudios de variantes raras	35

1.5.4.3.1.- Estudios de variantes raras en exoma	37
HIPÓTESIS y OBJETIVOS.....	42
MATERIAL Y MÉTODOS	48
1.- Validación de los genes seleccionados en el estudio previo de exoma .	50
1.1.- Antecedentes y resultados previos	50
1.2.- Diseño del estudio	52
1.2.1.- Población de estudio.....	52
1.2.2.- Extracción y procesado del ADN.....	53
1.2.3.- Diseño por amplicones para multiplex.....	53
1.2.4.- Puesta a punto y PCRs	57
1.2.4.1.- PCR1: Amplificación en multiplex	57
1.2.4.2.- PCR2: Incorporación de identificadores específicos.....	58
1.2.5.- Tratamiento post-PCR.....	59
1.2.6.- Secuenciación con tecnología de nueva generación (NGS)	60
1.2.6.1.- Preparación de la muestra	61
1.2.6.2.- Tecnología NGS Illumina, Inc.....	61
2.- Identificación de variantes genéticas de exoma implicadas en el desarrollo de la DM2 mediada o no por resistencia a insulina (RI).....	64
2.1.- Antecedentes previos	64
2.2.- Diseño del estudio	64
2.2.1.- Poblaciones de estudio	65
2.2.1.1.- Estudio Egabro-Pizarra.....	65
2.2.1.2.- Estudio caso/control	66
2.2.2.- Obtención y preparación de las muestras del Estudio Egabro-Pizarra. 67	
2.2.3.- Proceso de generación de las genotecas de exoma.....	67
Estudio caso/control.....	67
Estudio Egabro-Pizarra.....	67
2.2.3.1.- Fragmentación e incorporación de adaptadores	69
2.2.3.2.- Hibridación y captura	70
2.2.3.3.- Indexación y Secuenciación.....	71
3.- Análisis bioinformático	72
3.1.- Flujo de análisis bioinformático general o <i>pipeline</i>	72
3.1.1.- Demultiplexado y generación de archivos <i>FastQ</i>	74
3.1.2.- Control de calidad	76
3.1.3.- <i>Trimming</i>	78
3.1.4.- Alineamiento	78
3.1.5.- Coberturas.....	80

3.1.6.- Identificación de variantes.....	81
3.1.7.- Normalización de las variantes.....	82
3.1.8.- Anotación de variantes.....	83
3.1.9.- Filtrado de variantes.....	83
3.1.9.1.- Priorización de variantes.....	83
3.1.10.- Validación del flujo bioinformático.....	87
4.- Análisis estadístico.....	88
4.1.- Estadística general.....	88
4.2.- Estudios de asociación de variantes.....	88
4.2.1.- Análisis básicos de asociación.....	89
RESULTADOS Y DISCUSIÓN.....	93
1.- Validación de los genes seleccionados en el estudio previo de exoma..	95
1.1.- Características de la población de estudio.....	95
1.2.- Diseño de genes seleccionados.....	96
1.3.- Control de calidad: Programa <i>FastQC</i>	96
1.4.- Análisis de Coberturas.....	97
1.5.- Selección de las variantes genéticas.....	100
1.6.- Descripción de resultados por gen.....	101
1.6.1.- GEN ACAD10.....	101
1.6.1.1.- Proceso de secuenciación.....	102
1.6.1.2.- Variantes identificadas.....	102
1.6.2.- GEN ACPL2/PXYLP1.....	104
1.6.2.1.- Proceso de secuenciación.....	104
1.6.2.2.- Variantes identificadas.....	105
1.6.3.- GEN ACSL3.....	106
1.6.3.1.- Proceso de secuenciación.....	107
1.6.3.2.- Variantes identificadas.....	107
1.6.4.- GEN ACSM3.....	109
1.6.4.1.- Proceso de secuenciación.....	110
1.6.4.2.- Variantes identificadas.....	110
1.6.5.- GEN AFMID.....	112
1.6.5.1.- Proceso de secuenciación.....	113
1.6.5.2.- Variantes identificadas.....	114
1.6.6.- GEN ARHGEF38.....	115
1.6.6.1.- Proceso de secuenciación.....	116
1.6.6.2.- Variantes identificadas.....	116
1.6.7.- GEN CACNG5.....	118
1.6.7.1.- Proceso de secuenciación.....	118
1.6.7.2.- Variantes identificadas.....	119
1.6.8.- GEN CCNDBP1.....	120

1.6.8.1.- Proceso de secuenciación	121
1.6.8.2.- Variantes encontradas	122
1.6.9.- GEN CCT8L2.....	123
1.6.9.1.- Proceso de secuenciación	124
1.6.9.2.- Variantes identificadas	125
1.6.10.- GEN CIAPIN1.....	125
1.6.10.1.- Proceso de secuenciación.....	125
1.6.10.2.- Variantes identificadas.....	126
1.6.11.- GEN CYPIA1.....	127
1.6.11.1.- Proceso de secuenciación.....	128
1.6.11.2.- Variantes identificadas.....	128
1.6.12.- GEN <i>DDO</i>	130
1.6.12.1.- Proceso de secuenciación.....	130
1.6.12.2.- Variantes identificadas.....	131
1.6.13.- GEN EBLN2.....	132
1.6.13.1.- Proceso de secuenciación.....	133
1.6.13.2.- Variantes identificadas.....	133
1.6.14.- GEN <i>GCKR</i>	135
1.6.14.1.- Proceso de secuenciación.....	136
1.6.14.2.- Variantes identificadas.....	136
1.6.15.- GEN HABP2.....	139
1.6.15.1.- Proceso de secuenciación.....	139
1.6.15.2.- Variantes identificadas.....	140
1.6.16.- GEN OLFML1.....	141
1.6.16.1.- Proceso de secuenciación.....	142
1.6.16.2.- Variantes identificadas.....	143
1.6.17.- GEN SLC12A4.....	144
1.6.17.1.- Proceso de secuenciación.....	¡Error! Marcador no definido.
1.6.17.2.- Variantes identificadas.....	145
1.6.18.- GEN SLC15A3.....	147
1.6.18.1.- Proceso de secuenciación.....	147
1.6.18.2.- Variantes identificadas.....	148
1.6.19.- GEN SREBF2.....	149
1.6.19.1.- Proceso de secuenciación.....	150
1.6.19.2.- Variantes identificadas.....	151
2.- Identificación de variantes genéticas de exoma implicadas en el desarrollo de la DM2 mediada o no por RI	154
2.1.- Diseño del estudio	154
2.2.- Características de la población de estudio	154
2.2.1.- Estudio Egabro-Pizarra	154
2.2.2.- Estudio caso/control.....	156

2.2.3.- Estudio caso/control conjunto.....	156
2.3.- Control de calidad: Programa <i>FastQC</i>	157
2.4.- Análisis de Coberturas.....	160
2.5.- Validación de <i>miracleV2</i>	163
2.6.- Variantes encontradas en el análisis de exoma por estudio	163
2.6.1.- Estudio Egabro-Pizarra	164
2.6.1.1.- Estudio de evolución 1	165
2.6.1.2.- Estudio de evolución 2	169
2.6.2.- Estudio caso/control.....	173
2.6.3.- Estudio caso/control conjunto.....	181
3.- Limitaciones y perspectivas futuras	189
CONCLUSIONES.....	192
REFERENCIAS	197

ABREVIATURAS

A	Adenina
ABCC8	Del inglés <i>ATP binding cassette subfamily C member 8 gene</i>
ACAD10	Del inglés <i>Acyl-CoA deshydrogenase 10 gene</i>
ACPL2/PXYLP1	Del inglés <i>2-phosphoxylose phosphatase 1 gene</i>
ACSL3	Del inglés <i>Acyl-CoA Synthetase Long Chain Family member 3 gene</i>
ACSM3	Del inglés <i>Acyl-CoA Synthetase Medium Chain family member 3 gene</i>
ADA	Del inglés <i>American Diabetes Association</i>
	Del inglés <i>Adiponectin, C1Q and collagen domain containing gene</i>
ADIPOQ	<i>gene</i>
ADN	Ácido desoxirribonucleico
AFMID	Del inglés <i>Arylformamidase gene</i>
ARHGEF38	Del inglés <i>Rho guanine nucleotide exchange factor 38 gene</i>
ARN	Ácido ribonucleico
ATP	Adenosin trifosfato
BCL	Del inglés <i>Base Call files</i>
BWA	Del inglés <i>Burrows-Wheeler Aligner program</i>
C	Citosina
	Del inglés <i>Calcium voltage-gated channel auxiliary subunit gamma 5 gene</i>
CACNG5	<i>gene</i>
CADD	Del inglés <i>Combined Annotation Dependent Depletion</i>
CAPN10	Del inglés <i>Calpain 10 gene</i>
CBPs	Células Beta Pancreáticas
CCNDBP1	Del inglés <i>Cyclin D1 binding protein 1 gene</i>
CCR	Cáncer Colonorectal
CCT8L2	Del inglés <i>Chaperonin containing TCP1 subunit 8 like 2 gene</i>
CDH18	Del inglés <i>Cadherin 18 gene</i>
CIAPIN1	Del inglés <i>Cytokine induced apoptosis inhibitor 1 gene</i>
Cr	Cromosoma
CYP1A1	Del inglés <i>Cytochrome P450 family 1 subfamily A member 1 gene</i>
DAG	Diacilglicerol
DCCT	Del inglés <i>Diabetes Control and Complications Trial</i>
DDO	Del inglés <i>D-aspartate oxidase gene</i>
DEP	Diabetes del páncreas exocrino

ABREVIATURAS

DGCR2	Del inglés <i>DiGeorge syndrome critical region gene 2 gene</i>
DM1	Diabetes Mellitus Tipo 1
DM2	Diabetes Mellitus tipo 2
DMG	Diabetes Mellitus Gestacional
DMM	Diabetes Mellitus Mitocondrial
EBLN2	Del inglés <i>Endogenous Bornavirus like nucleoprotein 2 gene</i>
EHW	Equilibrio Hardy-Weinberg
	Del inglés <i>Ectonucleotide pyrophosphatase/phosphodiesterase 1 gene</i>
ENPP1	<i>gene</i>
EO	Estrés oxidativo
FID	Federación internacional de Diabetes
FOXX1	Del inglés <i>Forkhead box K1 gene</i>
G	Guanina
GAD65	Ácido Glutámico Descarboxilasa
GAGS	Glucosaminglicanos
GBA	Glucosa Basal Alterada
GCK	Glucocinasa
GCKR	Del inglés <i>Glucokinase receptor gene</i>
GEF	Del inglés <i>Guanine Exchange Factor</i>
GIAB	Del inglés <i>The Genome in a Bottle Consortium</i>
GPa	Niveles de glucosa en plasma al azar
	Del inglés <i>Genome Reference Consortium Human genome build 37</i>
GrCh37	<i>37</i>
GWAS	Del inglés <i>Genome-Wide Association Studies</i>
HABP2	Del inglés <i>Hyaluronan binding protein 2 gene</i>
HbA1c	Hemoglobina glicosilada
HDL	Del inglés <i>High Density Lipoprotein Cholesterol</i>
HGNC	Del inglés <i>HUGO Gene Nomenclature Committee</i>
HLA	Del inglés <i>Human Leucocyte Antigen System</i>
	Del inglés <i>Major histocompatibility complex, class II, DQ alpha 1 gene</i>
HLA-DQA1	<i>gene</i>
	Del inglés <i>Major histocompatibility complex, class II, DQ beta 1 gene</i>
HLA-DQB1	<i>gene</i>
HNF1A	Del inglés <i>Hepatocyte nuclear factor 1 alpha gene</i>
HNF1B	Del inglés <i>Hepatocyte nuclear factor 1 beta gene</i>
HNF4A	Del inglés <i>Hepatocyte nuclear factor 4 alpha gene</i>

ABREVIATURAS

HOMA	Del inglés <i>Homeostatic Model Assessment test</i>
HRC	Del inglés <i>Haplotype Reference Consortium</i>
HT1	Tampón de Hibridación 1 Asociación Internacional de los grupos de estudio de diabetes y embarazo
IADPSG	
IMC	Índice de Masa Corporal
ITG	Intolerancia a la Glucosa Del inglés <i>Potassium inwardly rectifying channel subfamily J member 11</i>
KCNJ11	
kg/m ²	Kilogramos/metro cuadrado
LADA	Del inglés <i>Latent Autoimmune Diabetes in Adults</i>
LDL	Del inglés <i>Low Density Lipoprotein</i>
MAF	Del inglés <i>Minority Allele Frequency</i> Miopatía mitocondrial, encefalopatía, acidosis láctica y episodios tipo ictus
MELAS	
mg/dL	Miligramos/decilitro
MIDD	Del inglés <i>Maternity inherited diabetes and deadness</i>
mmHg	Milímetros de Mercurio
mmol/L	Milimol/litro
MODY	Del inglés <i>Maturity onset Diabetes of the Young</i>
MTMR7	Del inglés <i>Myotubularin related protein 7 gene</i>
MYO10	Del inglés <i>Myosin X gene</i>
NaOH	Hidróxido Sódico
NCBI	Del inglés <i>National Centre for Biotechnology Information</i>
ng/uL	Nanogramos/microlitro
NGS	Del inglés <i>New Generation Sequencing</i>
NGSP	Programa Nacional de estandarización de la glicohemoglobina
NIH	Del inglés <i>National Institute of Health</i>
NIST	Del inglés <i>National Institute of Standards and Technology</i>
nM	NanoMolar
OLFML1	Del inglés <i>Olfactomedin like 1 gene</i>
OMS	Organización Mundial de la Salud
OR	Del inglés <i>Odds Ratio</i>
ORF	Del inglés <i>Open Reading Frame</i>
p	p-valor
PARP2	Del inglés <i>Poly(ADP-ribose) polymerase 2 gene</i>

ABREVIATURAS

pb	pares de bases
PCR	Del inglés <i>Polymerase Chain Reaction</i>
ped	Del inglés <i>Pedigree File</i>
pM	picoMolar Del inglés <i>Peroxisome proliferator activated receptor gamma gene</i>
PPARG	<i>gene</i>
PPFs	Polimorfismos poco frecuentes
PREDIMED	Estudio de Prevención con Dieta Mediterránea
RE	Retículo Endoplasmático
RI	Resistencia a Insulina Del inglés
ROS	Del inglés <i>Reactive Oxygen Species</i>
SBS	Del inglés <i>Sequencing by Synthesis</i>
SLC12A4	Del inglés <i>Solute carrier family 12 member 4 gene</i>
SLC15A3	Del inglés <i>Solute carrier family 15 member 3 gene</i>
SNP	Del inglés <i>Single Nucleotide Polimorfism</i>
SQL	Del inglés <i>Strutured Query Language</i> Del inglés <i>Sterol regulatory element binding transcription factor 2 gene</i>
SREBF2	<i>gene</i>
TCF7L2	Del inglés <i>Transcription Factor 7 like 2 gene</i>
TG	Triglicéridos
Tm	Temperatura de hibridación
TRIS	Del inglés (hydroxymethyl)aminomethane
TSOG	Test de sobrecarga oral de glucosa
UGDG	Unidad de Genómica y Diagnóstico Genético
UTR	Del inglés <i>untranslated region</i>
VCF	Del inglés <i>Variant Calling Format File</i>
VLDL	Del inglés <i>Very Low Density Lipoprotein</i> Del inglés ZAC: <i>Zinc finger protein associated with apoptosis and cell cycle arrest</i> and HYMAI: <i>Imprinted in hydatidiform mole</i>
ZAC/HYAMI	

ÍNDICE DE FIGURAS

Figura 1: Datos en millones de personas por continentes y por edad en el año 2017 y su estimación para el año 2045 según la Federación Internacional de Diabetes. (FID).	8
Figura 2: Factores de modulación de la patogénesis de la DM2. Adaptación de Farooqui <i>et al.</i> 2018 (47).	21
Figura 3: Proceso de regulación por insulina de la homeostasis de la glucosa. Adaptación de Sayem A, <i>et al.</i> (68).	27
Figura 4: Diagrama de Venn con los <i>loci</i> más significativos ($P > 5 \times 10^{-8}$) asociados con DM2 publicados antes del 2016.	34
Figura 5: Viabilidad de identificar variantes genéticas por la frecuencia de alelos de riesgo y la fuerza del efecto genético (<i>Odds Ratio</i> (OR)). Adaptación de Manolio <i>et al.</i> , 2009 (109).	36
Figura 6: Ejemplo de archivo para el diseño de amplicones solapados.	54
Figura 7: Esquema explicativo de ambas opciones de diseño de amplicones.	56
Figura 8: Secuencia universal añadida a cada par de oligonucleótidos específicos diseñados.	56
Figura 9: Condiciones de temperaturas y fases de amplificación utilizadas en la PCR1.	58
Figura 10: Secuencias universales para los <i>barcodes</i> de la casa Illumina, Inc.	58
Figura 11: Condiciones de temperatura de la segunda PCR de identificación de muestras.	59
Figura 12: Esquema de las PCRs necesarias para amplificar la zona de interés y adición de identificadores.	59
Figura 13: Proceso de amplificación clonal en forma de puente. Adaptación de Tecnología Illumina, Inc.	62
Figura 14: Proceso de secuenciación por síntesis (SBS). Adaptación de Tecnología Illumina, Inc (129).	62

Figura 15: Flujo de trabajo para la generación y enriquecimiento en exoma de librerías de ADN genómico.....	68
Figura 16: Esquema del proceso de generación de librerías de ADN y enriquecimiento en exoma para secuenciación por NGS.....	70
Figura 17: Esquema del flujo de análisis bioinformático utilizado en este trabajo.....	74
Figura 18: Ejemplo de archivo <i>FastQ</i> . Muestra las 4 líneas representativas de este tipo de archivo de texto.....	76
Figura 19: Visión de las secuencias alineadas y toda la información del proceso de alineamiento en un archivo BAM.....	79
Figura 20: Extracto de un archivo VCF donde se puede ver por cada línea la variante llamada y la información correspondiente.....	82
Figura 21: Gráfica que representa la cobertura media de todas las muestras de la población y el porcentaje de bases cubiertas a esa profundidad de cobertura.....	98
Figura 22: Distribución de la cobertura en la población secuenciada para el gen <i>ACAD10</i>	102
Figura 23: Distribución de la cobertura en la población secuenciada para el gen <i>ACPL2</i>	105
Figura 24: Representación gráfica del gen <i>ACSL3</i> y su cobertura en la población.....	107
Figura 25: Distribución de la cobertura de la población para el gen <i>ACSM3</i>	110
Figura 26: Distribución cobertura para el gen <i>AFMID</i>	113
Figura 27: Distribución de la cobertura media en la población del gen <i>ARHGEF38</i>	116
Figura 28: Distribución de la cobertura en la población secuenciada del gen <i>CACNG5</i>	119
Figura 29: Distribución de la cobertura en la población del gen <i>CCNDBP1</i>	121

Figura 30: Distribución de la cobertura en la población secuenciada del gen *CCT8L2*..... 124

Figura 31: Distribución de la cobertura en la población secuenciada del gen *CIAPIN1*..... 126

Figura 32: Distribución de la cobertura en la población secuenciada del gen *CYP1A1*..... 128

Figura 33: Distribución de la cobertura en la población secuenciada del gen *DDO*..... 131

Figura 34: Distribución de la cobertura en la población secuenciada del gen *EBLN2*..... 133

Figura 35: Distribución de la cobertura en la población secuenciada del gen *GCKR*..... 136

Figura 36: Distribución de la cobertura en la población secuenciada del gen *HABP2*..... 140

Figura 37: Distribución de la cobertura en la población secuenciada del gen *OLFML1*..... 143

Figura 38: Distribución de la cobertura en la población secuenciada del gen *SLC12A4*..... 145

Figura 39: Distribución de la cobertura en la población secuenciada del gen *SLC15A3*..... 148

Figura 40: Distribución de la cobertura en la población secuenciada del gen *SREBF2*..... 151

Figura 41: Representación de la densidad frente a la cobertura media en exones, sondas y la suma de ambos..... 161

Figura 42: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos A y B mediante el programa *PANTHER Gene List Analysis*..... 167

Figura 43: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos A y B mediante el programa *PANTHER Gene List Analysis*..... 168

Figura 44: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos C y D mediante el programa *PANTHER Gene List Analysis*..... 171

Figura 45: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos C y D mediante el programa *PANTHER Gene List Analysis*..... 172

Figura 46: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos E y F mediante el programa *PANTHER Gene List Analysis*..... 179

Figura 47: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos E y F mediante el programa *PANTHER Gene List Analysis*..... 180

Figura 48: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos AE (controles) y BDF (casos) mediante el programa *PANTHER Gene List Analysis*. 186

Figura 49: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos AE (controles) y BDF (casos) mediante el programa *PANTHER Gene List Analysis*. 187

ÍNDICE DE TABLAS

Tabla 1: Descripción de las variantes seleccionadas de los estudios anteriores en exomas de DM2. Genoma de referencia GrCh37.	51
Tabla 2: Información del gen <i>SERBF2</i> . Genoma de referencia GrCh37.....	51
Tabla 3: Distribución según diagnóstico de RI y/o DM2 de los pacientes de la población Egabro-Pizarra (Málaga).	66
Tabla 4: Distribución de los pacientes del estudio caso/control de estudios previos realizados por la UGDG.....	66
Tabla 5: Características generales del kit <i>Surseselect All human exon v5 + UTRs</i> (Adaptación Agilent Technologies, Inc., Santa Clara, EE.UU.).....	71
Tabla 6: Características clínicas de la población de estudio.....	95
Tabla 7: Información general de los genes secuenciados.....	96
Tabla 8: Valores de los parámetros del control de calidad de los archivos <i>FastQ</i> de las muestras del estudio antes y después del proceso de <i>trimming</i>	97
Tabla 9: Resultados de cobertura media por gen, número de exones y número de amplicones diseñados.....	99
Tabla 10: Clasificación por impacto de las variantes del gen <i>ACAD10</i>	103
Tabla 11: Características y descripción de las variantes encontradas para el gen <i>ACAD10</i>	103
Tabla 12: Clasificación por impacto de las variantes del gen <i>ACPL2</i>	105
Tabla 13: Características y descripción de las variantes encontradas para el gen <i>ACPL2</i>	106
Tabla 14: Clasificación por impacto de las variantes del gen <i>ACSL3</i>	108
Tabla 15: Características y descripción de las variantes encontradas para el gen <i>ACSL3</i>	108
Tabla 16: Clasificación por impacto de las variantes del gen <i>ACSM3</i>	110
Tabla 17: Características y descripción de las variantes encontradas para el gen <i>ACSM3</i>	111

Tabla 18: Características de las variantes seleccionadas del gen *AFMID* y su distribución en la población..... 114

Tabla 19: Características y descripción de las variantes encontradas para el gen *AFMID*..... 114

Tabla 20: Clasificación por impacto de las variantes del gen *ARHGEF38*..... 117

Tabla 21: Características y descripción de la variante encontradas para el gen *ARHGEF38*..... 117

Tabla 22: Clasificación por impacto de las variantes del gen *CACNG5*..... 119

Tabla 23: Características y descripción de las variantes encontradas para el gen *CACNG5*..... 120

Tabla 24: Clasificación por impacto de las variantes del gen *CCNDBP1*..... 122

Tabla 25: Características y descripción de las variantes encontradas para el gen *CCNDBP1*..... 122

Tabla 26: Clasificación por impacto de las variantes del gen *CIAPIN1*..... 126

Tabla 27: Características y descripción de la variante encontrada para el gen *CIAPIN1*..... 127

Tabla 28: Clasificación por impacto de las variantes del gen *CYP1A1*..... 129

Tabla 29: Características y descripción de las variantes encontradas para el gen *CYP1A1*..... 129

Tabla 30: Clasificación por impacto de las variantes del gen *DDO*..... 131

Tabla 31: Características y descripción de las variantes encontradas para el gen *DDO*..... 132

Tabla 32: Clasificación por impacto de las variantes del gen *EBLN2*..... 133

Tabla 33: Características y descripción de las variantes encontradas para el gen *EBLN2*..... 134

Tabla 34: Clasificación por impacto de las variantes del gen *GCKR*..... 136

Tabla 35: Características y descripción de las variantes encontradas para el gen *GCKR*..... 137

Tabla 36: Descripción de la variante del estudio previo en la población caso/control BN-BCN.....	138
Tabla 37: Clasificación por impacto de las variantes del gen <i>HABP2</i>	140
Tabla 38: Características y descripción de las variantes encontradas para el gen <i>HABP2</i>	141
Tabla 39: Clasificación por impacto de las variantes del gen <i>OLFML1</i>	143
Tabla 40: Características y descripción de las variantes encontradas para el gen <i>OLFML1</i>	143
Tabla 41: Clasificación por impacto de las variantes del gen <i>SLC12A4</i>	145
Tabla 42: Características y descripción de las variantes encontradas para el gen <i>SLC12A4</i>	146
Tabla 43: Clasificación por impacto de las variantes del gen <i>SLC15A3</i>	148
Tabla 44 Características y descripción de las variantes encontradas para el gen <i>SLC15A3</i>	149
Tabla 45: Clasificación por impacto de las variantes del gen <i>SREBF2</i>	151
Tabla 46: Características y descripción de las variantes encontradas para el gen <i>SREBF2</i>	152
Tabla 47: Relación de asociación con DM2 de los genes estudiados.....	153
Tabla 48: Características clínicas de los 119 individuos de la población Egabro-Pizarra para el estudio de exoma.....	155
Tabla 49: Características clínicas de los 369 individuos de la población caso/control para el estudio de exoma. Hay que considerar que los diagnosticados como DM2 están bajo tratamiento.....	156
Tabla 50: Características clínicas de toda la población conjunta.....	157
Tabla 51: Valores de los parámetros del control de calidad de los archivos <i>FastQ</i> de las muestras del Estudio Egabro-Pizarra antes y después del proceso de <i>trimming</i>	158
Tabla 52: Valores de los parámetros del control de calidad de los archivos <i>FastQ</i> de las muestras del estudio caso/control antes y después del proceso de <i>trimming</i>	159

Tabla 53: Descripción de las coberturas en las diferentes regiones y por subgrupos de ambas poblaciones. A) Estudio Egabro-Pizarra con 119 exomas y B) Estudio Caso/Control con 369 exomas.....	162
Tabla 54: Clasificación por impacto y consecuencia de las variantes entre el grupo A y el grupo B del estudio Egabro-Pizarra.	166
Tabla 55: Clasificación por impacto y consecuencia de las variantes entre el grupo C y el grupo D del estudio Egabro-Pizarra.	169
Tabla 56: Clasificación por impacto y consecuencia de las variantes de exoma del estudio caso/control.....	174
Tabla 57: Información descriptiva de las variantes significativas encontradas solo en uno de los grupos del análisis de asociación con el Test de Fisher entre el grupo E y el grupo F del estudio caso/control.	175
Tabla 58: Información descriptiva de las variantes significativas del análisis de asociación con el Test de Fisher entre el grupo E y el grupo F del estudio caso/control.....	177
Tabla 59: Clasificación por impacto y consecuencia de las variantes de exoma del estudio caso/control conjunto.....	181
Tabla 60: Información descriptiva de las variantes significativas encontradas solo en uno de los grupos del análisis de asociación con el Test de Fisher entre el grupo control (AE) y el grupo de casos (BDF) del estudio caso/control conjunto.....	182
Tabla 61: Información descriptiva de las variantes significativas del análisis de asociación con el Test de Fisher entre el grupo AE y el grupo BDF del estudio caso/control conjunto.	183
Tabla 62: Listado de genes con asociación estadísticamente significativa con DM2 del estudio caso/control (E-F) y del estudio caso/control conjunto.....	188

RESUMEN

IDENTIFICACIÓN DE VARIANTES GENÉTICAS RELACIONADAS CON
DIABETES MELLITUS TIPO 2 EN POBLACIÓN ESPAÑOLA.

INTRODUCCIÓN: La DM2 es una enfermedad metabólica multifactorial compleja en la que participan tanto factores ambientales como genéticos. Numerosos estudios han demostrado que la influencia de la genética en el riesgo de padecer DM2 es significativa. La heredabilidad estimada puede variar entre un 25% al 80% dependiendo de la población de estudio y de la herencia familiar. Sin embargo, las variantes encontradas hasta el momento, que tienen valores de Frecuencia del Alelo Minoritario (MAF) > 0.05 , solo explican una pequeña parte del riesgo de padecer DM2, sobre un 15%, el resto aún está por identificar. La búsqueda de variantes poco comunes o con valores de MAF < 0.05 pueden tener un efecto funcional relevante y afectar al desarrollo de la enfermedad. Actualmente, gracias al rápido desarrollo de la tecnología, se pueden realizar estudios y análisis de secuenciación de exoma completo que permiten encontrar variantes poco frecuentes con mayor fiabilidad.

OBJETIVOS: Identificar variantes genéticas en genes seleccionados de estudios previos, así como en muestras de exoma que puedan estar asociadas a DM2 en población española. Idear un flujo de análisis bioinformático fiable y automático para poder utilizar analizar los datos y que pueda ser utilizado en futuros proyectos.

METODOLOGÍA: Parte 1: Estudio de 19 genes en una población compuesta por 588 casos de entre 30 y 65 años y 832 controles de los 50 y los 75 años de edad. Secuenciación de multiplexes de amplicones por técnicas NGS, Desarrollo de un flujo de análisis bioinformático y análisis estadístico de asociación posterior.

Parte 2: Preparación de 119 genotecas de exoma caso/control, secuenciación por NGS y análisis bioinformático conjunto con 369 exomas caso/control extra disponibles, análisis descriptivo y estadístico de las variantes encontradas.

RESULTADOS: Parte 1: En total se han encontrado 539 variantes en los 19 genes analizados de las cuales 11 de ellas, han sido estadísticamente significativas en la asociación con la DM2 en el estudio caso/control realizado.

Parte 2: Se han encontrado en conjunto 12708 variantes genéticas en los 488 exomas analizados. De entre ellas, 49 variantes han resultado significativas en la asociación con DM2 dentro de la comparación caso/control analizada.

CONCLUSIONES: En este trabajo se ha llevado a cabo la identificación de variantes genéticas relacionadas con la DM2 mediante estudios de exoma y validación de genes previamente identificados. Se ha diseñado y automatizado un flujo de análisis bioinformático para ser utilizado en futuros proyectos. Aun así, son necesarios otros estudios que puedan replicar estos u otros resultados en poblaciones de mayor tamaño, así como estudios funcionales.

INTRODUCCIÓN

1.- DIABETES MELLITUS

1.1.- Definición

Se conoce como Diabetes Mellitus al conjunto de desórdenes metabólicos caracterizados, principalmente, por un aumento crónico de los niveles de glucosa en sangre. Diversos procesos patogénicos están implicados en su desarrollo, entre ellos, la alteración de la secreción de insulina, la destrucción de las células β -pancreáticas y a la alteración de la actividad de la insulina como reguladora del metabolismo.

Los síntomas más comunes son la poliuria, polidipsia, polifagia, visión borrosa y pérdida de peso. En casos más severos se puede producir cetoacidosis, el llamado síndrome no-cetónico hiperosmolar, que sin tratamiento efectivo puede conducir al coma y/o la muerte.

El progresivo desarrollo de la enfermedad contribuye, a largo plazo, a la aparición de disfunciones y daño orgánico causando retinopatías y ceguera, nefropatías que pueden conducir a un fallo renal, úlceras en los pies, aumento del riesgo cardiovascular y enfermedades cerebrovasculares (1-3).

1.2.- Epidemiología

Las diferencias en la prevalencia de la Diabetes Mellitus a nivel mundial dependen de varios factores como el género, la edad, la región geográfica, la situación económica del país, etcétera. Datos analizados en la última versión del *Diabetes Atlas* de la Federación internacional de Diabetes (FID) representados en la figura 1 estimaron

que en 2017 había 425 millones de personas de entre 20 a 79 años de edad con diabetes. Si se aumentaba el rango de edad de 18 a 99 años el número de personas subía a 451 millones. Esto significa que un 8.8% de la población mundial padece diabetes. Se calcula que estos números van en aumento y en 2045 se estima que habrá un 9.9% de la población mundial, 693 millones de personas enfermas de diabetes.

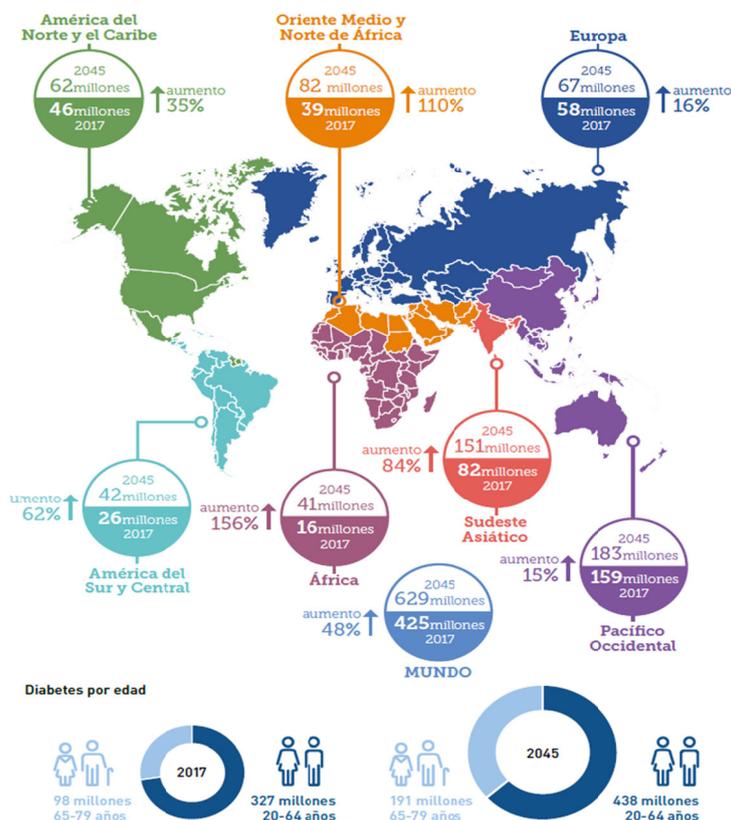


Figura 1: Datos en millones de personas por continentes y por edad en el año 2017 y su estimación para el año 2045 según la Federación Internacional de Diabetes. (FID).

Comparando con los datos analizados en el año 2000, se ha incrementado el número de personas diabéticas del rango de edad entre los 20 y los 79 años en un 28%. Las mujeres tenían una

prevalencia menor a los hombres de 8.4% y 8.9% respectivamente, aunque para 2045 se estima que para ambos sexos se incremente hasta un 9.9%.

Los países con ingresos más altos poseen una prevalencia 3 veces mayor en el rango de edad de entre 65-69 años comparado con los de menores ingresos. Esto es debido a los cambios socio-económicos que llevan a tener vidas más sedentarias y altos ratios de urbanización, aunque también mejoras en la salud y en la esperanza de vida de las personas con diabetes (4). Por otro lado, en los países con menores ingresos y/o que están en vías de desarrollo, donde la edad de debut de la enfermedad es menor, también hay un gran número de casos de diabetes, debido al bajo nivel educacional, bajo nivel sociocultural y falta de recursos sanitarios para su diagnóstico (5,6).

En cuanto a los datos de DM en España, los resultados del estudio Di@bet.es realizado en 2012 y que se va actualizando para conocer mejor la prevalencia de la DM en este país, indicaba un valor de 13.8% en individuos mayores a los 18 años (7). Datos más recientes de la FID de 2017 indican una prevalencia, en el rango de edad de 20 a 79 años, del 10.4% un poco menor que la encontrada en el estudio español, con un 28.4% de casos no diagnosticados.

1.3.- Diagnóstico

Según la Organización Mundial de la Salud (OMS) (8), junto con la FID (9) y la Asociación Americana de Diabetes (ADA) (2), las pruebas clínicas de diagnóstico para diabetes están basadas en la medición de los niveles de glucosa en plasma:

- Medición al azar de niveles de glucosa en plasma (GP) con clásicos síntomas de hiperglucemia o hipoglucemia: Valores \geq 200 mg/dL (11.1 mmol/L).
- Medición de glucosa en plasma en ayunas, al menos 8 horas sin comer (GPa): \geq 126 mg/dL (7 mmol/L).
- Glucemia plasmática a las 2 horas del test de sobrecarga oral con 75 gramos de anhídrido de glucosa disuelta en agua (TSOG): Valores \geq 200 mg/dL (11.1 mmol/L).

También es válida la medición de los niveles de hemoglobina glicosilada (HbA1c) siempre que el paciente no tenga ninguna hemoglobinopatía y/o anemia, no sea una mujer embarazada, ni niños, ni con sospecha de Diabetes tipo 1 y teniendo en cuenta la predisposición a ello de su etnia. Este método debe estar certificado por el Programa Nacional de estandarización de la glicohemoglobina (NGSP) y llevado a cabo según el ensayo *Trial Diabetes Control and Complications Trial* (DCCT). Se consideran valores \geq 6.5% (48 mmol/mol) como diagnósticos de diabetes.

Por otro lado, cuando se realizan estos test y no se sobrepasan los umbrales recomendados, pero los niveles siguen siendo altos para considerarse normales, se define como prediabetes. Los criterios que definen la prediabetes son: (10)

- Glucosa Basal Alterada (GBA): Valores alterados después de hacer el test GPA: De 100 mg/dL a 125 mg/dL (5.6–6.9 mmol/L). Según la OMS se establece el umbral en 110 mg/dL (6.1 mmol/L).

- Intolerancia a la Glucosa (ITG): Valores alterados después de realizar el test TSOg: De 140 mg/dL (7.8 mmol/L) a 199 mg/dL (11.0 mmol/L).
- Valores de HbA1c alterados: De 5.7–6.5% (39–47 mmol/mol).

También se deben tener en cuenta los siguientes criterios para comprobar la existencia de prediabetes en adultos asintomáticos.

- Personas con sobrepeso u obesidad (Índice de Masa Corporal (IMC) $\geq 25 \text{ kg/m}^2$ o $\geq 23 \text{ kg/m}^2$ en asiáticos americanos) y que cumplan uno o más de los siguientes criterios:
 - Ser de etnia con alto riesgo (americanos africanos, latinos, nativos americanos, americanos asiáticos, isleños del pacífico).
 - Historial de enfermedad cardiovascular.
 - Hipertensión ($\geq 140/90$ mmHg o tratados).
 - Niveles de colesterol HDL > 35 mg/dL (0.90 mmol/L) y/o niveles de triglicéridos (TG) > 250 mg/dL (2.82 mmol/L).
 - Mujeres con síndrome del ovario poliquístico.
 - Inactividad física.
 - Otras condiciones clínicas asociadas con la resistencia a insulina (obesidad severa, acantosis pigmentaria...).
- Pacientes con prediabetes (HbA1c $\geq 5.7\%$ (39 mmol/mol), GBA, o ITG).

- Mujeres que fueron diagnosticadas con Diabetes Mellitus Gestacional (DMG), deben hacerse pruebas de por vida al menos cada 3 años.
- Edad mayor a 45 años.
- Si los resultados son normales se deben repetir al menos en 3 intervalos al año considerando las pruebas más frecuentes y los resultados iniciales de riesgo.

1.4.- Clasificación y tipos de diabetes

Según los criterios de la Asociación Americana de Diabetes (ADA), se establecen diversos tipos de diabetes dependiendo de sus características fisiopatológicas, defectos a nivel genético conocidos, y otros síndromes asociados que cursan con síntomas asociados a la diabetes.

1.4.1.- Diabetes Mellitus tipo 1 (DM1)

La DM1 es una enfermedad autoinmune que destruye las células β -pancreáticas causando una deficiencia de insulina, seguida de una hiperglucemia en sangre conduciendo a una cetoacidosis. Constituye del 5-10% de los casos totales de diabetes, se presenta en la infancia y adolescencia. Se conocen marcadores diana para el diagnóstico como son: los autoanticuerpos de las células beta, de la insulina, de la ácido glutámico descarboxilasa (GAD65), y los de las tirosina-fosfatasa IA-2 y IA-2 β . Estos autoanticuerpos suelen estar presentes en el plasma en un 85-90% en estado de hiperglucemia y son una forma rápida de diagnóstico y clasificación de la enfermedad. Normalmente está fuertemente asociada a tipos específicos de HLA, y con los genes *HLA-DQA1* y *HLA-DQB1* relacionados.

Algunos tipos de diabetes tipo 1 pueden ser idiopáticos ya que se desconocen las causas verdaderas, pero sí que cursan con insulinopenia en diferentes grados y cetoacidosis, aunque no tienen evidencias de autoinmunidad ni tampoco relación con el HLA. Normalmente se da en personas descendientes de africanos o asiáticos (11,12).

1.4.2.- Diabetes Mellitus tipo 2

Descrita en el apartado 1.5 de este trabajo.

1.4.3.- Diabetes latente autoinmune en el adulto (LADA)

Sobre este tipo de diabetes no hay un entendimiento uniforme en cuanto a su clasificación, ya que posee signos y síntomas tanto de la DM1 como de la Diabetes Mellitus tipo 2 (DM2). Se clasifica también como Diabetes tipo 1.5, Diabetes mellitus de progresión lenta de insulino-dependencia o doble diabetes (13).

Aparece en edad adulta normalmente sobre los 30-35 años, hay existencia de autoanticuerpos de los islotes β -pancreáticos y de la GAD65 como en la DM1, pero no insuficiencia total de insulina. Ciertos haplotipos de HLA y otros genes asociados que son típicos de la DM1 también pueden existir. También se relaciona con la DM2 por presentar una asociación típica del gen *TCF7L2*, además de la edad de aparición y de otros factores ambientales relacionados como el sobrepeso, la inactividad física y la resistencia a insulina (14,15).

1.4.4.- Diabetes Mellitus Gestacional (GDM)

La diabetes gestacional es una intolerancia a carbohidratos resultante en una hiperglucemia de severidad variada diagnosticada durante el

primer o segundo trimestre de embarazo. Mujeres con riesgo de padecer este tipo de diabetes son aquellas con edad avanzada, con historial previo de intolerancia a la glucosa, un primer embarazo con diabetes, que sean de ciertos grupos étnicos de riesgo o que presenten elevación de los niveles de glucosa en sangre en cualquier momento del embarazo. En 2017 se estimaron que de 21.3 millones de nacimientos en el mundo, el 16.2% estuvo afectado por hiperglucemia en el embarazo, y al menos el 84.6% se le diagnosticó a la madre este tipo de diabetes (16).

Como criterio sistemático, en la semana 24 y 28 de gestación se realiza el test *one-step* establecido por el consenso de la Asociación Internacional de los grupos de estudio de diabetes y embarazo (IADPSG) que conlleva a la medición a 1 hora y a 2 horas después de la ingesta de 75 gramos de glucosa disuelta en agua, con un ayuno de al menos 8 horas. Los criterios diagnósticos son:

- Medición en ayunas: Niveles \geq 92 mg/dL (5.1 mmol/L).
- 1 hora después: Niveles \geq 180 mg/dL (10.0 mmol/L).
- 2 horas después: Niveles \geq 153 mg/dL (8.5 mmol/L).

Existe otro test denominado *two-steps*, recientemente adoptado por la OMS que tiene ciertas variaciones en los criterios y valores, y que, en algunos estudios, se ha visto una mejoría en cuanto a la precisión del diagnóstico.

Después de 6 semanas o más del alumbramiento la mujer que ha sido diagnosticada con GDM debe volver a ser reclasificada mediante los test ITG y o TSOG para saber el posible efecto a largo plazo (3,12).

1.4.5.- Otros tipos específicos de diabetes:

Seguendo la clasificación de la ADA existen otros tipos de diabetes menos comunes, pero que se puede diagnosticar como tal en el transcurso de la enfermedad.

1.4.5.1.- Defectos genéticos de la función de las células β -pancreáticas

Diversas formas de diabetes están asociadas a defectos monogénicos, es decir, que se conocen los genes que las causan, relacionados con la pérdida de función de las células beta. Se caracterizan por la aparición de una hiperglucemia temprana, pero sin autoinmunidad, por lo que no se clasifican como DM1. Normalmente siguen un patrón de herencia autosómico dominante.

1.4.5.2.- Diabetes Mellitus Neonatal (DMN)

Se diagnostica en los primeros 6 meses de edad y hay dos tipos, la transitoria, la cual puede ser debida a varios mecanismos genéticos interrelacionados: por la herencia de ambos cromosomas 6 paternos, conocida también como isodisomía uniparental paterna, por una duplicación en la región 6q24 o por una sobreexpresión de los genes *ZAC/HYAMI* por defecto en la metilación. Estos genes están asociados con el proceso de apoptosis al final del ciclo celular, su sobreexpresión disminuye el número de células β -pancreáticas durante el desarrollo del individuo haciendo que sean insuficientes para la regulación glucídica. Normalmente hay remisión de los efectos a los 18 meses, aunque puede aparecer más tardíamente como DM2 (17). El segundo tipo es permanente ya que afecta a los genes *KCNJ11* y *ABCC8* que codifican las subunidades que conforman de los canales de potasio de las células beta. Estos defectos permiten una configuración abierta permanente de dichos canales, causando una

hiperpolarización de la membrana plasmática inhibiendo a su vez la secreción de insulina. El uso como tratamiento de sulfonilureas, bloqueantes de los canales de potasio que permiten su cierre, mejora la calidad de vida del paciente (18,19).

1.4.5.3.- Diabetes Mellitus Mitocondrial (DMM)

Este tipo de diabetes ocurre por una mutación en el ARN de transferencia de la leucina, una transición de A a G en la posición 3243 del ADN mitocondrial (A3243G) produciendo la llamada diabetes hereditaria mitocondrial y sordera (*MIDD*). Esta mutación se localiza en el complejo I de la cadena respiratoria mitocondrial afectando a la producción de ATP necesario para el correcto funcionamiento celular. Esto afecta a la viabilidad de las células beta y a la producción de insulina(20,21). La variante A3243G también se ha relacionado con el Síndrome *MELAS* aunque este no cursa con diabetes, lo cual indica una diferente expresión fenotípica de la mutación (22).

1.4.5.4.- Diabetes Mellitus de la juventud de aparición en la madurez (MODY)

Este tipo de diabetes engloba un grupo de desórdenes monogénicos autosómicos dominantes sin dependencia de insulina que se suelen presentar antes de los 25 años. Se ha comprobado que es la forma monogénica más común de diabetes y afecta al 1-2% de todos los diabéticos en Europa, sin contar con el porcentaje de individuos a los que se les diagnostica erróneamente con DM1 o DM2 (23). Dependiendo de las localizaciones de las mutaciones, se pueden clasificar hasta en 14 tipos diferentes. Las más frecuentes conocidas son las variantes en los genes *HNF4A* (MODY 1), *GCK* (MODY 2), *HNF1A* (MODY 3), *HNF1B* (MODY 5). Los efectos clínicos difieren

dependiendo de la edad de expresión, localización de la mutación y la afectación del correcto funcionamiento de las células beta, pero todas cursan características clínicas parecidas que hacen posible su diagnóstico como puede ser: la ausencia de autoanticuerpos de los islotes pancreáticos, acción de la insulina intacta aunque de baja producción, ausencia de relación con obesidad, niveles lipídicos normales o historial familiar de diabetes entre otras características (24,25).

1.4.5.5.- Defectos genéticos en la acción de la insulina

Otras enfermedades menos comunes relacionadas con diabetes son debidas a defectos genéticos en el receptor celular de la insulina, afectando a su unión y produciendo desde una hiperinsulinemia y modesta hiperglucemia hasta una diabetes severa. Normalmente cursan con acantosis pigmentaria. Entre las más conocidas están: la resistencia a insulina tipo A, el leprechaunismo, el Síndrome de Rabson-Mendenhall, Diabetes Lipoatrófica y otros síndromes pediátricos (26).

1.4.5.6.- Enfermedades del páncreas exocrino

La diabetes del páncreas exocrino (DEP) también conocida como Diabetes tipo 3c, se debe a la destrucción o daño en el páncreas y depende de varias condiciones, entre ellas, la pancreatitis, fibrosis cística, neoplasias pancreáticas, hemocromatosis y otras dolencias que puedan afectar al correcto funcionamiento pancreático (27,28).

1.4.5.7.- Endocrinopatías

Alteraciones en los niveles de diversas hormonas que actúan en la regulación metabólica como: la hormona del crecimiento, la

aldosterona, el glucagón, el cortisol, la epinefrina, etc. Antagonizan la acción de la insulina. Desórdenes asociados al exceso de estas hormonas como acromegalia, el Síndrome de Cushing, hipertiroidismo, hipercotisolismo, glucagonoma, feocromocitomas, somatostatinoma, aldosteronoma, pueden cursar con hiperglicemias que suelen resolverse al eliminar el exceso de hormona o extirpar el tumor creado (29,30).

1.4.5.8.- Diabetes asociada a fármacos o químicos

Muchos fármacos pueden afectar a la secreción insulínica del organismo y por ello producir hiperglicemia y/o resistencia a insulina. Algunos de ellos son las estatinas (31), ácido nicotínico (32), glucocorticoides (33). Otros afectan a la integridad de las células beta como pueden ser los tóxicos tipo tiazidas, pentaminas, isonizidas, el veneno de rata Vacor entre otros (34).

1.4.5.9.- Diabetes asociada a infecciones

Algunos virus han sido asociados a la destrucción de las células beta induciendo diabetes. Ejemplos de ellos son la rubeola congénita, citomegalovirus, rotavirus, entre otros (35,36).

1.4.5.10.- Otros síndromes genéticos asociados con diabetes

Dentro de los síndromes más comunes producidos con aberraciones cromosómicas como el Síndrome de Klinefelter, el Síndrome de Down, el Síndrome de Prader-Willi, el Síndrome de Wolfram, entre otros, cursan con manifestaciones relacionadas con la insuficiencia de insulina, diabetes insípida o deficiencia de células beta-pancreáticas (37).

1.5.- Diabetes Mellitus tipo 2 (DM2)

1.5.1.- Prevalencia de la DM2

La Diabetes Mellitus tipo 2 (DM2) se caracteriza por una deficiencia de insulina y/o su acción relativa causada por la disfunción de las células β -pancreáticas y/o resistencia a insulina en sus órganos diana (38). Es una de las causas de muerte más frecuentes, ya que está asociada con un amplio rango de enfermedades cardiovasculares, desórdenes del sistema nervioso, y problemas hepáticos (39). Según los datos más recientes de la OMS, 422 millones de personas sufren diabetes en el mundo y más del 90% son DM2. Se ha visto una relación inversa entre la existencia de DM2 en la población y su posición socioeconómica. Varios estudios han analizado por separado los países con alto, medio y bajo nivel de ingresos y se ha comprobado que un 40-60% de mayor riesgo relativo se da en países con ingresos más bajos (40). Países dentro de este rango como India, China, Brasil, Pakistán, Indonesia, tienen los niveles más altos de personas con diabetes. Globalmente la prevalencia de la diabetes es similar entre hombres y mujeres (41). Según la OMS, solo en la región europea hay 60 millones de personas (10.3% de hombres y 9.6% de mujeres) mayores de 25 años que han sido diagnosticadas de diabetes.

Según la FID, en España se estimó, en 2017, que había unos 3.5 millones de personas con diabetes con una prevalencia del 10.4%. El estudio Di@bet.es de 2012 fue el primero en realizarse en territorio español, recogió 5072 personas al azar a los cuales se les midieron entre otros, niveles de glucosa en ayunas y tras realizar el TSO, lípidos, hábitos de vida, edad, sexo, medidas corporales, etc. Se concluyó que la prevalencia general en mayores de 18 años fue de un

13.8%. Los valores alterados de GBA y ITG hacen que los trastornos relacionados con el metabolismo de carbohidratos estén sobre un 28% (7). En otros estudios realizados en diferentes regiones de la geografía española, cuyos habitantes comparten un entorno parecido y están expuestos a los mismos factores ambientales, se han encontrado valores de prevalencia diversos. Según la Encuesta Europea de Salud en España realizada en 2014, se observó que las comunidades con valores más altos de prevalencia eran las ciudades de Ceuta y Melilla, Región de Murcia y Canarias. Siendo las comunidades del norte las que poseían valores más bajos. En esta encuesta el valor medio de prevalencia se situaba en 6.84% (42). En el estudio prospectivo Pizarra realizado en la región de Andalucía, vieron que la prevalencia aumentaba desde un 13.4 hasta un 20.3% a los 6 años (43). Por otro lado, en el reciente estudio de la Generalitat Valenciana sobre diabetes de 2017, se estimó que la prevalencia fue del 7.8%. Algo mayor a la obtenida en el estudio anterior de 2009 la cual fue de 6.74%. Esto supone un incremento de casos de DM2 en la comunidad significativo (44). El último estudio de la Comunidad de Madrid prueba que, en hombres la prevalencia era de un 12.3%, casi el doble que en mujeres (45).

En un reciente estudio realizado, teniendo en cuenta todos los datos disponibles hasta la fecha, de todos los estudios realizados con población española, en el que infirieron 3 casos sobre la evolución de la prevalencia de la DM, vieron que, los tres escenarios mostraban una tendencia a la baja o estabilización, en cuanto al progreso de la prevalencia, sobre un 8.6% para el 2025 (46).

1.5.2.- Factores de riesgo ambientales evolución y prevención de la DM2

1.5.2.1.- Factores de riesgo ambientales

Se considera la DM2 una enfermedad multifactorial ya que influyen tanto factores genéticos como ambientales. En la figura 2 se puede ver como interaccionan dichos factores entre sí y como favorecen la aparición de la DM2.

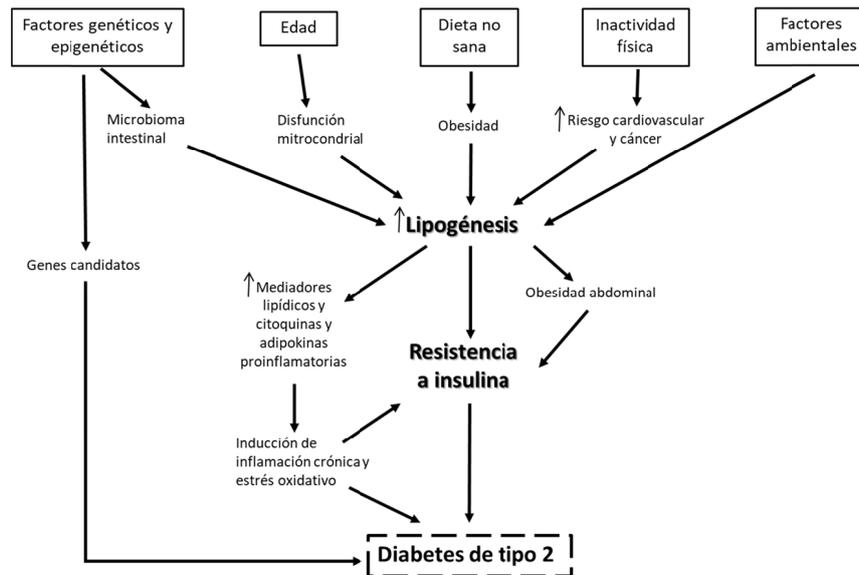


Figura 2: Factores de modulación de la patogénesis de la DM2. Adaptación de Farooqui *et al.* 2018 (47).

Según la ADA se consideran como factores de riesgo de la DM2 los siguientes parámetros:

- No modificables, propios del individuo:
 - Etnia
 - Edad > 45 años

- Antecedentes familiares de diabetes
- Diabetes gestacional en embarazo previo
- Dieta materna pobre durante el embarazo
- Modificables, adquiridos con el estilo de vida:
 - Valores de IMC ≥ 25 kg/m² (sobrepeso), ≥ 27 kg/m² (obesidad)
 - Hipertensión arterial valores $\geq 140/90$ mmHg
 - Valores HDL > 35 mg/dL y/o valores TG > 250 mg/dL
 - Test GBA alterado
 - Sedentarismo
 - Dieta poco sana

Dentro de los no modificables son relevantes: la etnia a la que pertenece el individuo, se tiene mayor o menor predisposición a desarrollar DM2. Así, los descendientes de americanos africanos, nativos americanos, asiáticos o nativos del pacífico poseen una mayor susceptibilidad y riesgo a padecer diabetes (48); se conoce que en la DM2 influyen los antecedentes familiares y que existe un componente genético asociado, por lo que se han realizado una gran diversidad estudios en busca de los posibles genes implicados (ver apartado 1.5.4). Además, la influencia del ambiente intrauterino y problemas asociados al embarazo pueden añadir cambios epigenéticos que predispongan al individuo a un mayor riesgo de padecer no solo DM2 sino otras enfermedades (49). Uno de estos problemas es que la madre haya sufrido DMG (50). Los bebés nacidos de madres que han sufrido DMG también tienen un riesgo a largo plazo de desarrollar obesidad, DM2 u otros tipos de diabetes (51). Por otro lado, las guías para la prevención de la DM2 del Instituto Nacional de Salud

Americano (NIH) aconsejan realizar actividad física, cuidar la dieta durante este periodo con el fin de mejorar el control glucémico, el peso y la estabilización del IMC, ya que se ha documentado que, mujeres que han sufrido DMG tienen predisposición a desarrollar DM2, ocurre entre un 15 y un 50% a los cinco años (52).

En cuanto a los factores de riesgo que se asocian al estilo de vida, estos están muy interrelacionados entre sí, aun considerando el amplio rango de tipos de dietas en todo el mundo, se ha visto que la ingesta de carbohidratos refinados, grasas saturadas en exceso y carne roja contribuye al riesgo diabético (53). La globalización y el desarrollo de la tecnología han supuesto un aumento del sedentarismo y la falta de ejercicio físico lo que favorece el riesgo cardiovascular y de diabetes (54).

Por otro lado, la obesidad es una enfermedad crónica muy asociada a la DM2 y se la considera en sí misma un factor de riesgo importante. Según datos de la OMS, los casos de obesidad en el mundo se han triplicado desde 1975. Se ha calculado que en 2016 sobre un 13% de la población mundial adulta eran obesos. Actualmente no solo afecta a la población adulta, alarmantemente 41 millones de niños por debajo de los 5 años en todo el mundo, se consideran obesos o con sobrepeso. Considerado un problema de los países más desarrollados, se ha visto que, actualmente, los casos de obesidad se dan en aquellos con bajos ingresos. Esto es debido a la falta de información médica, la falta de actividad física, la inadecuada nutrición durante las etapas de vida más tempranas, la exposición a alimentos altos en grasas, azúcar, sal, y energéticos.

Las serias complicaciones cardiovasculares que conllevan la obesidad y la diabetes ponen en riesgo la vida del individuo ya que se consideran la causa más frecuente de muerte mundialmente (55).

1.5.2.2.- Complicaciones a largo plazo de la DM2

Los daños en el organismo producidos a largo plazo por la diabetes, afectan a la mayoría de los órganos y sistemas del organismo, algunos pueden tratarse y disminuir su progresión, aunque pueden llegar a ser irreversibles pudiendo causar más complicaciones o la muerte.

Retinopatía diabética: Es una complicación microvascular frecuente en pacientes con diabetes y una de las causas más frecuentes de ceguera en países en desarrollo. Se estima que una tercera parte de los diabéticos tienen problemas graves de visión (56). La retinopatía diabética tiene componentes tanto vasculares como neurológicos. Por lo que aquellas personas que la desarrollan tienen alto riesgo de padecer otras complicaciones (57).

Nefropatía diabética: Se caracteriza por fallo funcional y remodelado estructural del riñón debido principalmente a la hiperglucemia asociada a la diabetes. La permeabilidad glomerular se ve afectada produciendo proteinuria con altos niveles de microalbúmina (58). Es también una de las complicaciones microvasculares más comunes en pacientes diabéticos y una de las causas principales de la prescripción de diálisis crónica (59).

Neuropatía Diabética: Es la tercera consecuencia microvascular producida por la DM2 y se define como los signos y síntomas de disfunción de los nervios periféricos en pacientes diabéticos cuando

otras causas han sido excluidas. Se clasifican en simétricas y asimétricas dependiendo de la localización de los nervios afectados y producen problemas a nivel motor y sensorial (60). También es la causa de úlceras en los pies, conocido como pie diabético, amputaciones, heridas que no sanan, que suelen llevar a infecciones y gangrena y disfunción sexual debida al estrés oxidativo en los tejidos cavernosos (61).

Como complicaciones macrovasculares existen diferentes tipos de arteriopatías según el tipo de vaso sanguíneo afectado y su localización. Las obstrucciones de dichos vasos se producen por calcificaciones de hidroxapatita por el exceso de colesterol o por glicosilación de las proteínas vasculares debido al estado de hiperglucemia (62). Principalmente las complicaciones cerebrales y cardíacas son las más importantes, dadas sus consecuencias.

1.5.2.3.- Prevención de la DM2

A pesar de que existan varios factores de riesgo que favorezcan la aparición de la DM2 que no se pueden modificar, sí que hay otros sobre los que se puede actuar con el fin de retrasar o prevenir su aparición. Los cambios en los hábitos de alimentación, por ejemplo, pueden favorecer una disminución en el riesgo de aparición de la DM2 (63). Un estudio realizado siguiendo la dieta mediterránea (PREDIMED) comprobó que, una dieta rica en grasas insaturadas puede ser útil para disminuir el riesgo de padecer resistencia a insulina y/o reducir los niveles de biomarcadores de inflamación asociados (64). Otro estudio realizado en población china concluyó que, el consumo de frutas y vegetales, los cuales son ricos en fibra y antioxidantes, pueden considerarse una barrera protectora contra la

DM2 y otros problemas de salud (65). Diferentes estudios sobre el aumento de la actividad física diaria, desde caminar a realizar ejercicios de alta intensidad, han demostrado que se puede reducir el riesgo asociado a la DM2. Siempre dependiendo de las condiciones físicas de cada persona, realizar un poco de ejercicio ayuda a mejorar la calidad de vida y reducir también la aparición de otras enfermedades relacionadas (66).

1.5.3.- Fisiopatología de la DM2

En condiciones normales las células beta del páncreas producen la insulina necesaria para controlar los niveles de glucosa en plasma. Además, en situación postprandial, actúa como una hormona antilipolítica evitando la liberación de lípidos del tejido adiposo, favoreciendo el proceso de lipogénesis en ellos y la gluconeogénesis en el músculo y en el hígado (67) (Figura 3).

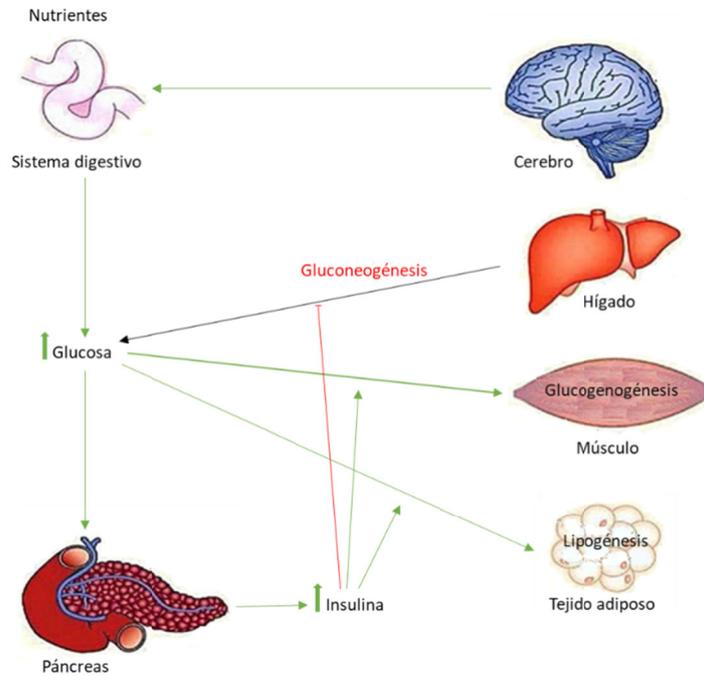


Figura 3: Proceso de regulación por insulina de la homeostasis de la glucosa. Adaptación de Sayem A, *et al.* (68).

Sin embargo, cualquier alteración a lo largo de este proceso propicia la aparición de los síntomas y procesos fisiopatológicos propios de la diabetes.

Defectos en la secreción y sensibilidad de la insulina ante la presencia de elevados niveles de glucosa, la aparición de la resistencia a insulina (RI) y la disfunción y/o destrucción de las células β -pancreáticas, teniendo en cuenta la predisposición genética y las variantes patogénicas de los genes implicados, son los mecanismos principales de la desregulación metabólica característica de la DM2 (69).

1.5.3.1.- Alteración de las células β -pancreáticas

El fallo de las células beta del páncreas (CBPs) ocurre por una serie de procesos como: la presencia de hiperglucemia y metabolismo de lípidos alterado, aumento del estrés oxidativo, procesos inflamatorios y autoinmunes, incremento de apoptosis y desdiferenciación celular combinados con la predisposición genética del individuo (70).

La resistencia a insulina (RI) producida en los tejidos periféricos, hace que las CBPs aumenten la producción de insulina para controlar los niveles de glucosa, lo cual genera un importante estrés oxidativo (EO) en estas células y más concretamente en el retículo endoplásmico (RE) generándose a su vez, cambios metabólicos internos. Además, de la hiperglucemia crónica existente, que genera un estado de glucotoxicidad en las CBPs, en conjunto, todo afecta a su capacidad secretora de insulina pudiendo desencadenar el proceso de muerte celular programada (71,72).

Un efecto parecido produce la alta presencia de ácidos grasos liberados a la sangre debido a la RI y la inhibición de su oxidación. Esto supone un grado de toxicidad lipídica grave que afecta a las CBPs inhibiendo la producción de insulina y favoreciendo su apoptosis (73,74).

Recientemente se ha comprobado que la pérdida de funcionalidad de las CBPs no solo se debe a su apoptosis sino también, al proceso de desdiferenciación que siguen con el fin de preservar su identidad y volver a diferenciarse en presencia de condiciones metabólicas más favorables (75,76).

1.5.3.2.- Resistencia a la insulina

Es producida por diferentes alteraciones metabólicas incluyendo la exposición a niveles elevados de insulina (77). El aumento excesivo principalmente, en situaciones de obesidad, de las moléculas provenientes del tejido adiposo como ácidos grasos, hormonas esteroideas y mediadores proinflamatorios, estimulan la gluconeogénesis hepática y favorecen el estado resistente a insulina (78). El estado de inflamación derivado, provoca estrés en el RE que perjudica al correcto plegamiento de las proteínas y puede llegar a iniciar el proceso de muerte celular programada al no poder restaurar su homeostasis (79). La disfunción mitocondrial debido a alteraciones metabólicas derivadas de la obesidad, por ejemplo, promueve en DM2 la liberación de especies reactivas del oxígeno (ROS) afectando entre otros procesos celulares a la cascada de señalización de la insulina fomentando la RI (80).

El proceso patológico de RI es una condición característica de los tejidos insulino-dependientes, como las células musculares o adiposas, cuando no responden correctamente a los niveles normales de insulina (81).

La RI es una característica común a muchos tipos de desórdenes metabólicos y está muy ligada a la obesidad. Condición que a menudo se refleja también en la mayoría de los pacientes con DM2.

1.5.4.- Factores genéticos de la DM2

La DM2 es una enfermedad metabólica multifactorial compleja en la que participan tanto factores ambientales como genéticos. Numerosos estudios han demostrado que la influencia de la genética

en el riesgo de padecer DM2 es significativa. La heredabilidad estimada puede variar entre un 25% al 80% dependiendo de la población de estudio y de la herencia familiar (82). Se ha visto que hay un riesgo aditivo en el modelo de transmisión de la DM2 si los padres, sobretodo la madre, han sufrido algún tipo de diabetes o alteraciones en los niveles glucídicos (83). Por otro lado, muchos trabajos, con el fin de dilucidar el componente genético de la DM2, se han centrado en seguimiento y estudio de parejas de gemelos registrados y se ha calculado una heredabilidad sobre el 72%, cambiante dependiendo del tiempo de duración de dichos estudios (84,85).

Otros en cambio, se han centrado en entender la influencia de los orígenes étnicos en el desarrollo de DM2. Por ejemplo, ciertos grupos indígenas, como los indios Pima o tribus de islas del Pacífico tienen un riesgo muy alto (86,87). En la población caucásica se ha visto aproximadamente un 40% de riesgo si uno de los padres tiene DM2, incrementándose a un 70% si los dos padres están afectados. Sin embargo, las variantes encontradas hasta el momento solo explican una pequeña parte del riesgo de padecer DM2, sobre un 15%, es necesario realizar más estudios interrelacionando tanto datos de variantes genéticas, de herencia familiar, como influencia ambiental y epigenética para poder llegar a conocer en su totalidad el papel de la genética en la DM2 (88).

Han sido muchos los métodos y aproximaciones de estudio que se han llevado a cabo en estos 30 años o más con el interés de revelar la predisposición genética de la DM2. Por ejemplo, estudios de ligamiento en familias sobre los años 90, estudios de genes candidatos elegidos por su importancia biológica, estudios más

recientes de asociación de genoma completo (GWAS), los cuales buscan marcadores genéticos con MAF mayores a 0.05 a lo largo del genoma. La evolución de la tecnología y la aparición de la secuenciación de nueva generación (NGS) han permitido estudios a mucha mayor escala poblacional, análisis de exomas y genomas completos y búsqueda de variantes con valores de MAF menores a 0.05, que están ayudando a entender mejor la parte genética de la DM2.

1.5.4.1.- Estudios de ligamiento en familias y genes candidatos

A finales del siglo pasado, se empezaron a utilizar los análisis de ligamiento en familias como método de búsqueda de posibles regiones genómicas relacionadas con un fenotipo de interés respecto de marcadores conocidos en la secuencia de ADN. Este método solo servía para identificar *loci* de alta penetrancia y con un modo de herencia conocido, como en la mayoría de enfermedades monogénicas. En el campo de la diabetes han servido para encontrar genes relacionados con la diabetes tipo *MODY* o diabetes neonatal. Sin embargo, al ser la DM2 un desorden multifactorial y poligenético, estos estudios no eran los más indicados. Aun así, sí que se han realizado varios análisis sobre diversos genes que, posteriormente, se han podido corroborar. Entre ellos está el gen *CAPN10*, encontrado en el cromosoma 2 en un estudio en indios México-americanos (89), el gen *ENPP1* asociado con la obesidad en la infancia y madurez (90) o el gen *ADIPOQ* en caucásicos franceses (91). El gen *TCF7L2* estudiado en varias poblaciones diferentes (92,93),

posee una gran importancia en la actualidad siendo uno de genes con mayor asociación conocida.

Por otro lado, teniendo en cuenta los mecanismos biológicos relacionados con la patogénesis de la DM2, se realizaron diversos estudios sobre genes candidatos cuyas variantes pudieran predisponer o estar relacionadas con el fenotipo de DM2. EL primer gen reproducible fue *PPARG* codificante del receptor nuclear PPAR- γ , estudios lo asocian con el incremento de la sensibilidad a insulina y, por lo tanto, con un potencial efecto de protección frente a DM2 (94,95). Aunque otros estudios, que incluyen el gen *KCNJ11* (96), indicarían que determinadas variantes de estos genes actuarían de manera aditiva en el incremento del riesgo de padecer DM2 (97).

1.5.4.2.- Estudios de asociación de genoma completo (GWAS)

Los estudios de GWAS han marcado un antes y un después en la búsqueda e identificación de variantes en muchas enfermedades conocidas. Son un diseño experimental usado para detectar asociaciones entre variantes genéticas y rasgos clínicos en poblaciones grandes. Se basan en la asociación directa de marcadores o en el principio del desequilibrio de ligamiento a nivel poblacional. En el segundo caso, podemos detectar asociaciones de polimorfismos no directamente implicados en la enfermedad dado que los alelos de diferentes *loci* o genes cercanos forman haplotipos que se heredan juntos con mayor probabilidad que al azar. Por otro lado, también se rigen por la teoría de que una enfermedad común debe producirse por el efecto de variantes comunes o con valores de MAF > 0.05 (98).

Desde la puesta en marcha del proyecto *HapMap* en 2003, después de la secuenciación del primer genoma humano, con la finalidad de construir una base de datos de más de 1 millón de variaciones genéticas representativas poblacionales (99) hasta la actualidad, se han realizado diversos estudios de GWAS para muchas enfermedades. En referencia a la DM2 se han podido encontrar más de 100 genes potencialmente asociados en estudios multi-étnicos y meta-análisis relacionando las variantes encontradas con rasgos clínicos como: niveles de glucosa, de HbA1c, presencia de resistencia a insulina, valores de IMC, etc. (100). Gracias a estos estudios a mayor escala, se han podido identificar muchos genes y definir con mayor precisión, las rutas metabólicas implicadas en DM2 (101).

En la figura 4 se pueden ver todos los genes con variantes principalmente identificados por GWAS, resaltándose aquellos que se han encontrado significativamente asociados con variables clínicas relacionadas con la glucemia.

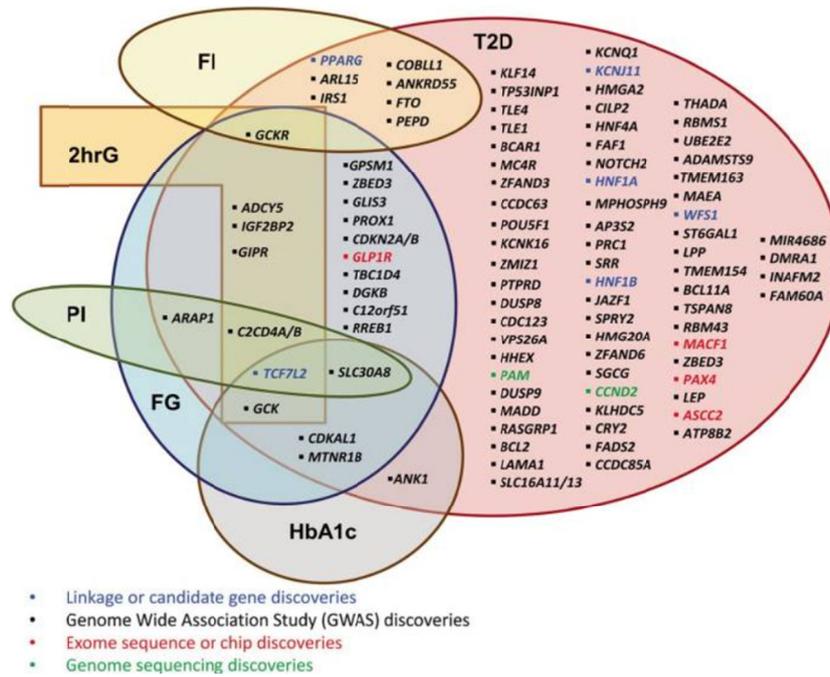


Figura 4: Diagrama de Venn con los loci más significativos ($P > 5 \times 10^{-6}$) asociados con DM2 publicados antes del 2016. FI: *Fasting insulin*, 2hrG: *2-hour glucose following an oral glucose tolerance test*. PI: *Fasting proinsulin*. FG: *fasting glucose*, HbA1c: *Hemoglobin A1c*, T2D: *Type 2 Diabetes* (102).

A pesar de que ha sido un gran avance a la hora de identificar variantes potencialmente asociadas a diversas características de diferentes enfermedades, incluida la DM2, ha permitido identificar una parte pequeña del componente genético. La alta frecuencia de falsos positivos ($p < 10^{-2}$), los valores de OR bajos o la alta frecuencia de aparición en la población ($MAF > 0.05$) hace que estas variantes solo representen una pequeña proporción de la heredabilidad estimada de una enfermedad, en el caso de la DM2, no supera el 15% (103,104).

1.5.4.3.- Estudios de variantes raras

Dado que las numerosas variantes comunes encontradas a lo largo de los estudios GWAS no pueden explicar en su totalidad las bases genéticas de las enfermedades, se empezó a estudiar también la importancia de las variantes con menor frecuencia poblacional ($MAF < 0.05$) o variantes raras ($MAF < 0.01$). Estas variantes son más numerosas en el genoma y algunas de ellas podrían tener un efecto mucho más relevante en el riesgo de desarrollar determinadas enfermedades que las variantes comunes. Numerosos estudios han visto que más del 60% de las variantes del genoma son polimorfismos con valores de $MAF < 0.05$.

Al tener una frecuencia mucho menor, es posible que no sean detectadas en estudios pequeños o que no se hayan incluido en muchos chips. Causas importantes de que esto sea así puede ser, la selección natural o los cambios que se han producido en nuestra forma de vida en las últimas décadas o siglos. La selección natural previene la aparición de las variantes de alto efecto negativo en la población, lo que podría explicar el efecto más modesto (*Odds Ratio* (OR) < 1.1) de las variantes comunes con mayor incidencia poblacional (105,106). En este sentido, la selección natural explicaría, en parte, que sean las variantes raras o de baja frecuencia aquellas que expliquen las enfermedades comunes (figura 5). Otras cuestiones relevantes serían los posibles efectos de variantes que se seleccionaron en contra durante milenios e incluso aquellas que eran beneficiosas pero que bajo las condiciones de vida actuales pueden tener efectos totalmente diferentes al haber cambiado el ambiente.

Las variantes raras se distribuyen en todo el genoma, si bien, es más probable que tengan un efecto funcional aquellas que se localizan en las regiones codificantes (incluyendo regiones contiguas) y otras regiones funcionales del genoma (107). Estas variantes pueden incrementar el riesgo de la enfermedad de dos a tres veces sin llegar a demostrar una clara segregación mendeliana y aun así contribuir substancialmente a esa heredabilidad perdida (108).

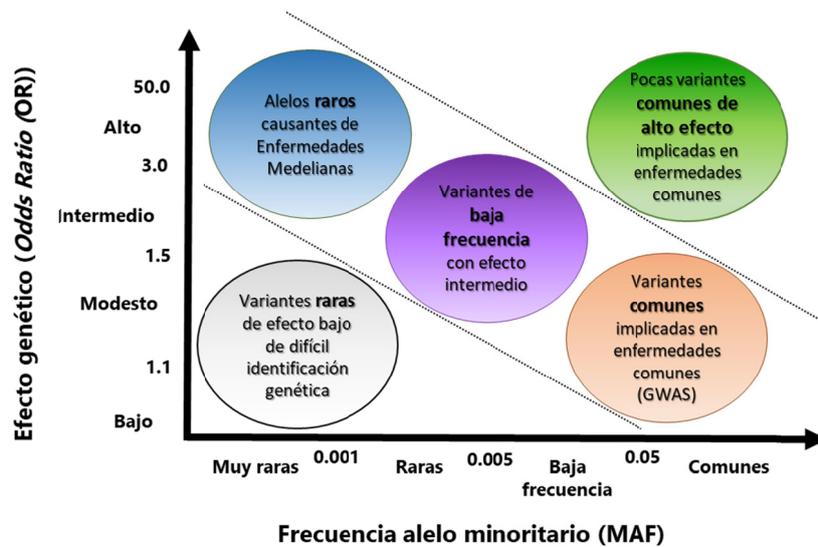


Figura 5: Viabilidad de identificar variantes genéticas por la frecuencia de alelos de riesgo y la fuerza del efecto genético (*Odds Ratio* (OR)). Adaptación de Manolio *et al.* 2009 (109).

Para poder estudiar e identificar variantes de baja frecuencia, solo puede ser posible a través de la secuenciación, necesitando tecnología mucho más avanzada que la secuenciación tradicional por Sanger. El desarrollo tecnológico ha permitido disponer de los sistemas de secuenciación masiva en paralelo que, con mucho menor coste, posibilitan obtener muchos más datos (110).

1.5.4.3.1.- Estudios de variantes raras en exoma

El desarrollo de la tecnología de secuenciación junto con eficientes procesos de captura de ADN, hace que el estudio del exoma sea un nuevo enfoque para estudiar las bases genéticas de las enfermedades. En aquellas más complejas, como la DM2, puede revelar variantes poco comunes que no han sido detectadas en los GWAS. Es un enfoque complementario que permite evaluar el papel de todas las variantes de secuencia sean comunes o de baja frecuencia. Para lograr un poder estadístico mayor es necesario disponer de un alto número de muestras en estudios caso/control o poblacionales (111). Debido a la compleja arquitectura genética que subyace en la base genética de las enfermedades, es difícil saber cuál de las muchas señales de asociación de un mismo *locus* es la causante real (112).

El genoma humano contiene 3 billones de pares de bases de ADN, pero solo el 2% de ellas son codificantes para proteínas. La inmensa mayoría de las mutaciones causantes de los trastornos monogénicos se encuentran en las zonas codificantes exónicas del genoma (113). También hay que considerar que solo una pequeña parte de las variantes genéticas son causales de enfermedad. Por otro lado, los estudios de exoma permiten identificar variantes en zonas intrónicas adyacentes, zonas de *splicing*, zonas promotoras y regiones no traducidas (UTR). Variantes encontradas en estas regiones pueden ser potencialmente importantes al afectar a la regulación génica y ser ellas, y no las variantes de dichos genes, las que posean la asociación al riesgo estudiada (114).

Además de ello, en el ámbito de la DM2, así como en otras enfermedades complejas, se ha intentado desentrañar su complejidad genética e intentar identificar mejor las causas de la misma. Numerosos estudios se han realizado en los últimos años y sobretodo, comparaciones entre estudios para poder aumentar el poder estadístico mediante el incremento muestral y conceder mayor credibilidad a las asociaciones encontradas. Se ha visto que este aumento es necesario para poder encontrar un efecto calculable de las variantes raras o de baja frecuencia.

Algunos ejemplos de trabajos que han llegado a estas conclusiones son, por ejemplo, la investigación publicada en 2016 por el consorcio *GoT2D* y *T2D-GENES* que comparó datos genómicos de tres fuentes diferentes, unos 12940 exomas multiétnicos, 2657 muestras europeas de genoma completo y 111548 con genotipos imputados de exoma y genoma completo. Vieron que la inmensa mayoría de las variantes encontradas eran de frecuencia común y estaban en regiones previamente identificadas por los estudios anteriores de *GWAS* y que solo habían surgido unas pocas variantes raras nuevas. Aunque a pesar de sus resultados que empíricamente indican la mayor contribución de las variantes comunes a la heredabilidad de la DM2 que las raras, los autores defienden que incrementando el tamaño muestral de casos de DM2 se podrá discernir, sin ninguna duda, el valor clínico y la explicación biológica de las variantes de baja frecuencia en el riesgo de la DM2 (115).

En el estudio genotipado mediante *arrays* de exoma en 81412 casos de DM2 y 370832 controles de diversas procedencias y su comparación con datos de *GWAS* en 50160 casos de DM2 y 465572

controles de ascendencia europea, se identificaron 40 asociaciones de las cuales solo 5 tenían valores de MAF <0.05 y todas con OR < 1.4 . Al realizar la comparación con los datos de *GWAS*, menos del 50% de las variantes codificantes encontradas tenían relación con el riesgo en la DM2. El término *smoking gun* explica bien el hecho de que no por ser variante codificante sea la "culpable" del fenotipo expresado, ya que la mayoría de ellas se encuentran en desequilibrio de ligamiento y su presencia no es casual. Este estudio no es muy informativo en cuanto a variantes raras, ya que en los *array* de exoma están poco representadas y también defienden el hecho de que se necesita aumentar el muestreo poblacional para poder identificarlas (112).

Recientemente se han encontrado 56 asociaciones de variantes de baja frecuencia ($0.5\% < \text{MAF} < 5\%$) y 14 raras ($\text{MAF} < 0.5\%$) en 60 *loci*, en 900000 muestras europeas (75000 de ellas con DM2), ha sido utilizando datos imputados del *Haplotype Reference Consortium (HRC)* covariados y no por IMC y sexo (116). El estudio demuestra como el incremento del tamaño muestral combinado con buena y precisa comprensión de la imputación de los datos, se puede aumentar la caracterización de la contribución de la genética al riesgo de la DM2. Aun así, la explicación estimada por las variantes raras a la heredabilidad de la DM2, sigue siendo inferior a la otorgada por las variantes comunes encontradas.

Hace unos meses se publicó el mayor análisis de secuenciación de exoma de un estudio caso/control de DM2. Se secuenciaron de 20791 pacientes con Diabetes Tipo 2 y 24440 controles en 5 grupos de ascendencia. Calcularon que alrededor de 75000 a 185000 muestras pueden ser necesarias para conseguir significación en la asociación de

variantes raras en estudios de exoma completo tanto para DM2 como para otras enfermedades. Los resultados encontrados enuncian que, la proporción encontrada de variantes raras a nivel de gen, parece estar distribuida entre numerosos genes, es decir, su nivel de aportación al riesgo de DM2 se desvanece en presencia de otras variantes comunes más frecuentes, lo cual no significa que dejen de contribuir a la heredabilidad de la DM2. Es más, su efecto puede ser mayor en el sentido de que, si son variantes codificantes de cambio proteico, al estar menos correlacionadas con el desequilibrio de ligamiento que las variantes más comunes, su aparición pueden implicarse directamente en la patogénesis de la enfermedad. La dificultad empírica de encontrarlas limita la habilidad para identificarlas de *novo*.

Sin embargo, cuando en un gen se establece una implicación con una enfermedad, la búsqueda de variantes raras que puedan influir permitirá explicar mejor la relación entre su función y la enfermedad. Además, se justifica la necesaria combinación de datos de exoma y *GWAS* para conocer los rasgos biológicos de las enfermedades complejas. La futura adición de datos de secuenciación de genoma completo cuando sean suficientemente efectivos a nivel de coste, incrementará el poder de resolución y la imputación a paneles de genes referencia para muchas enfermedades (117,118).

HIPÓTESIS y OBJETIVOS

HIPÓTESIS

Una gran parte del componente genético de la DM2 no se ha identificado. Hasta el momento se han analizado principalmente variantes frecuentes. Por otro lado, una parte de las variantes raras puede tener un efecto funcional relevante y afectar al fenotipo de forma importante. Además, la DM2 se puede desarrollar de forma dependiente e independiente de la resistencia a insulina.

En base a los datos anteriores, nuestras hipótesis principales son:

- Parte del componente genético que puede explicar la DM2 se encuentra en variantes o polimorfismos poco frecuentes (PPFs) ($MAF \leq 0.05$), en las regiones codificantes del genoma las cuales, pueden interferir en la funcionalidad génica.
- El desarrollo de DM2 dependiente o independiente de la resistencia a insulina se produce por mecanismos diferentes y, por lo tanto, puede estar provocado por variantes genéticas que afectan a genes o rutas metabólicas diferentes.

OBJETIVOS

El objetivo principal de este proyecto es la identificación y/o validación de variantes genéticas de exoma relacionadas con la DM2 en población española.

Este trabajo se ha dividido en dos objetivos principales con sus respectivos objetivos secundarios:

1) Validación de los genes seleccionados en el estudio previo de exoma mediante la secuenciación del gen completo por tecnología de secuenciación de nueva generación (NGS).

- a) Puesta a punto de la preparación de las librerías de amplicones de los genes seleccionados.
- b) Diseño y automatización de un flujo de análisis bioinformático adecuado para el análisis de las variantes genéticas encontradas.
- c) Análisis de asociación estadística y descripción de los diversos tipos de variantes genéticas encontradas diferenciando entre casos y controles.

2) Identificación de variantes genéticas que puedan estar implicadas en el desarrollo de la DM2 (mediada o no por RI) mediante secuenciación de exoma.

- a) Realización y preparación de las librerías de 119 exomas para secuenciar por tecnología de nueva generación (NGS).
- b) Puesta a punto y establecimiento del flujo de análisis bioinformático automatizado para el análisis conjunto de los

exomas disponibles (488 en total) para identificación de variantes genéticas raras relacionadas con DM2.

- c) Selección y análisis descriptivo y de asociación estadística de las variantes genéticas encontradas en los exomas secuenciados.

MATERIAL Y MÉTODOS

1.- Validación de los genes seleccionados en el estudio previo de exoma

1.1.- Antecedentes y resultados previos

En el desarrollo de la línea de investigación "Estudio de las bases genéticas de la DM2" en la Unidad de Genómica y Diagnóstico Genético (UGDG) se realizó la identificación y análisis de variantes genéticas de baja frecuencia en exoma, en controles sanos y pacientes con DM2 de población española, con el fin de identificar aquellas que pudieran tener una posible implicación en el desarrollo de diabetes o en su protección frente a ella. Teniendo en cuenta estos resultados, se seleccionaron 18 variantes a analizar (Tabla 1), dado que todas ellas se encontraban solo en uno de los grupos estudiados (casos o controles) y tienen un efecto funcional importante. Además, se decidió analizar las regiones codificantes e intrónicas contiguas de los genes donde se localizan dado que, en un mismo gen pueden darse diferentes variantes raras que puedan tener un efecto similar, e incluso el efecto contrario.

Tabla 1: Descripción de las variantes seleccionadas de los estudios anteriores en exomas de DM2. Genoma de referencia GrCh37.

Cr	Posición	SNP	Cambio	Gen	Ensembl ID	Consecuencia
2	27745372	rs146053779	C/T	GCKR	ENSG00000087734	Ganancia codón STOP
2	2233785788	rs139220113	T/C	ACSL3	ENSG00000123983	Region donadora de splicing
3	73111724	rs2231925	T/C	EBLN2	ENSG00000255423	Ganancia codón STOP
3	140993422	-	T/G	ACPL2	ENSG00000155893	Ganancia codón STOP
4	106474097	rs148016102	G/T	ARHGGEF38	ENSG00000236699	Ganancia codón STOP
6	110736730	rs141023778	C/T	DDO	ENSG00000203797	Ganancia codón STOP
10	115338424	rs41292628	C/T	HABP2	ENSG00000148702	Ganancia codón STOP
11	7509386	rs141351486	C/T	OLFML1	ENSG00000183801	Ganancia codón STOP
11	60714017	rs116597710	G/A	SLC15A3	ENSG00000110446	Ganancia codón STOP
12	112167654	rs146335507	C/T	ACAD10	ENSG00000111271	Ganancia codón STOP
15	43483698	-	C/T	CCNDBP1	ENSG00000166946	Ganancia codón STOP
15	750015234	-	G/A	CYP1A1	ENSG00000140465	Ganancia codón STOP
16	20802007	rs34381224	C/G	ACSM3	ENSG00000005187	Ganancia codón STOP
16	57464236	rs170358	G/A	CIAPIN1	ENSG00000005194	Ganancia codón STOP
16	67980235	-	G/T	SLC12A4	ENSG00000124067	Ganancia codón STOP
17	64881037	rs62621240	C/T	CACNG5	ENSG00000075429	Ganancia codón STOP
17	76200907	rs138105144	A/G	AFMID	ENSG00000183077	Región aceptora de splicing
22	17072347	rs139948519	C/T	CCTBL2	ENSG00000198445	Ganancia codón STOP

Cr: Cromosoma. SNP: Single Nucleotide Polymorphism.

Al presente estudio se añadió el gen *SREBF2* para su completo análisis por ser un gen candidato que se está analizando en la UGDG (Tabla2) y al haberse encontrado una variante genética rara causante de DM2 mediante otros estudios llevados a cabo en la Unidad.

Tabla 2: Información del gen *SREBF2*. Genoma de referencia GrCh37.

Cr	Posición	Gen	Ensembl ID	Ensembl Transcrito
22	42229109-42303312	<i>SREBF2</i>	ENSG00000198911	ENST00000361204

Cr: Cromosoma. SNP: Single Nucleotide Polymorphism.

1.2.- Diseño del estudio

En esta primera parte del trabajo se planteó realizar la validación de las variantes halladas en el estudio previo de exoma, además de la secuenciación y análisis de las posibles nuevas variantes que se pudieran encontrar a lo largo de los genes a los que pertenecen, utilizando un estudio caso/control de población española. Se realizó el diseño por amplicones solapados, de las regiones exónicas y zonas de *splicing*, amplificándolos en diversas multiplexes. Se llevó a cabo la secuenciación con tecnología de nueva generación (NGS) utilizando los sistemas Miseq y HiSeq (Illumina, Inc, San Diego, EE.UU.). A raíz de la necesidad de analizar las variantes encontradas, se desarrolló un flujo de análisis bioinformático eficaz, universal y automático para que pudiera ser utilizado en éste y en futuros análisis de la UGDG. La parte estadística y de asociación se llevó a cabo usando el programa *PLINK v1.07*(119).

1.2.1.- Población de estudio

Las muestras utilizadas en esta primera validación de los polimorfismos seleccionados, pertenecen al Biobanco del CIBERDEM y al Biobanco Nacional de ADN (Salamanca, España). Se seleccionaron para este estudio un total de 1419 individuos teniendo en cuenta la edad. Para los pacientes diagnosticados con DM2 se estableció un rango de edad de 30 a 65 años. En cambio, para los controles, se eligieron edades comprendidas entre los 50 a 75 años ya que la posible aparición de la enfermedad en estos individuos es menos probable.

Se contó con la aprobación de los correspondientes Comité de Ética y Comité Científico (INCLIVA y Biobancos) y todos los participantes dieron su Consentimiento Informado para donar sus muestras a los mencionados biobancos.

1.2.2.- Extracción y procesado del ADN

Se realizó la extracción del ADN genómico de las muestras mediante el uso del sistema automático Chemagic Magnetic Separator Module 1 (Chemagen, Baesweiler, Alemania). Se cuantificó por fluorescencia el ADN obtenido con el kit QuantiFluor® dsDNA (Promega, Madison, EE.UU.) (120), fluoróforo que emite en un rango de 510-570nm. Se utilizó el Sistema de Detección GLOMAX Multi+ Detection System (Promega, Madison, EE.UU.) siguiendo las especificaciones pertinentes de la casa comercial. Se diluyeron las muestras con tris 10nM pH8.0 hasta tener un rango entre 10 y 20 ng/uL para homogeneizar las concentraciones de todas las muestras.

1.2.3.- Diseño por amplicones para multiplex

El primer paso para la amplificación de las regiones de interés de los genes seleccionados fue el diseño de los amplicones necesarios. Para ello, se seleccionó el transcrito consenso de la base de datos COSMIC (121) y se utilizaron las secuencias de los genes del genoma humano de referencia GrCh37 de las bases de datos públicas NCBI (122) y Ensembl (123). Una vez seleccionadas las regiones de interés en un documento de secuencia, se elaboró manualmente el diseño inicial de cada fragmento según las especificaciones de los programas a usar, teniendo en cuenta que había que cubrir toda la secuencia de una

fragmentos, más grandes o más pequeños a los esperados, aumentando la inespecificidad y disminuyendo la efectividad de la amplificación.

Para evitar este problema, la solución más sencilla fue separar todos los amplicones cuyos diseños solaparan, en como mínimo, dos agrupaciones de multiplexes. De esta manera, solo habría que tener en cuenta las interacciones entre los oligonucleótidos de los amplicones vecinos (Figura 7).

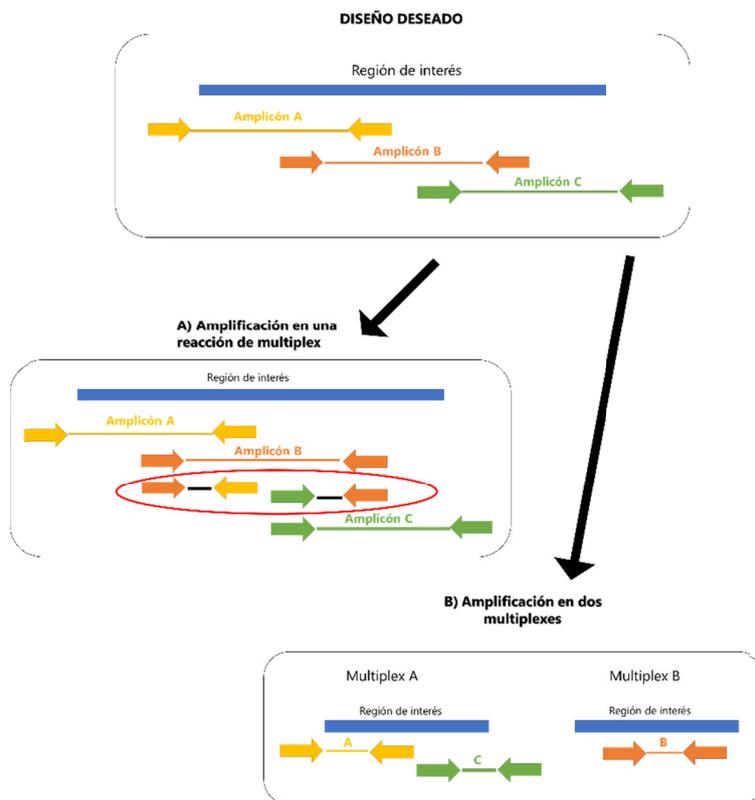


Figura 7: Esquema explicativo de ambas opciones de diseño de amplicones. Según el tamaño de fragmento deseado, se pueden diseñar amplicones sin tener en cuenta el solapamiento de la secuencia, permitiendo así la aparición de fragmentos inespecíficos (rodeados en rojo) debido a la interacción de los cebadores de diferentes diseños contiguos al utilizar una sola multiplex como en el ejemplo A. Otra forma es diseñar amplicones solapados, pero separar aquellos que contengan la misma secuencia parcial en dos multiplexes, para evitar así inespecificidades como en el ejemplo B.

Respecto a este hecho, el mismo programa poseía un módulo de comprobación de la estabilidad, calidad e interacciones intra e interespecíficas de dimerización que pudieran ocurrir entre los cebadores de la misma multiplex. Un apunte importante para realizar correctamente este análisis es que, a los cebadores diseñados se les debía de añadir una secuencia de nucleótidos específica, también llamada cola universal necesaria en pasos posteriores de amplificación (Figura 8). Por lo tanto, el análisis de estabilidad e interacciones se debía de realizar una vez añadida dicha secuencia extra.

Adaptadores universales:

Oligo 5': TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

Oligo 3': GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Figura 8: Secuencia universal añadida a cada par de oligonucleótidos específicos diseñados.

Para aquellos amplicones que dieron problemas al integrarse en las multiplexes establecidas, se optó por utilizar el programa libre *Primer3* (125,126) para su diseño individual. Se añadieron a la multiplex de origen y se verificaron de nuevo las interacciones entre cebadores.

Las condiciones generales de diseño de los amplicones fueron: Tamaño de los cebadores de 20 a 30 pb, siendo 22-25 pb el tamaño óptimo. Temperatura de hibridación (T_m) entre 58-64°C, óptimo de 60°C con una diferencia máxima de 2 grados dentro de la misma

multiplex. El rango de tamaños debía oscilar entre 150-250 pb máximo.

1.2.4.- Puesta a punto y PCRs

1.2.4.1.- PCR1: Amplificación en multiplex

Para llevar a cabo tanto la primera PCR como la segunda se utilizaron los protocolos establecidos en ese momento en el laboratorio, usando Multiplex PCR Taq Polymerase (127) con pequeñas variaciones, tanto en reactivos como en la concentración de oligonucleótidos, para poner a punto las condiciones más óptimas para cada multiplex. Esta parte del trabajo fue realizada en conjunto con dos estudiantes de Máster que participaron en la realización de la parte de laboratorio como parte de su Trabajo Final de Máster.

Al trabajar con gran cantidad de muestras se utilizaron placas de PCR de 384 y las amplificaciones se llevaron a cabo en termocicladores Veriti 384 (Applied Biosystem, Foster City, EE.UU.) siguiendo el programa de temperaturas descrito en la figura 9.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	20 min	
Desnaturalización	98 °C	20 seg	30x
Hibridación	55°C	30 seg	
	60°C	30 seg	
Extensión	72 °C	60 seg	
Extensión final	72 °C	5 min	
Espera	15 °C	∞	

Figura 9: Condiciones de temperaturas y fases de amplificación utilizadas en la PCR1.

1.2.4.2.- PCR2: Incorporación de identificadores específicos

El paso siguiente a la primera amplificación de las regiones de interés en todas las muestras es identificar cada muestra. Para ello fue necesario añadir a cada muestra un par específico de índices o *barcodes* con una secuencia conocida de 8 letras compatibles con la tecnología Illumina, Inc (Figura 10).

Estructura <i>Barcodes</i>:
Índice en 5': AATGATACGGCGACCACCGAGATCTACAC [i5] TCGTCGGCAGCGTC
Índice en 3': CAAGCAGAAGACGGCATAACGAGAT [i7] GTCTCGTGGGCTCGG

Figura 10: Secuencias universales para los *barcodes* de la casa Illumina, Inc.

Las condiciones de temperatura para esta segunda PCR fueron las mostradas en la figura 11.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	98°C	15 min	
Desnaturalización	98 °C	20 seg	40x
Hibridación	60°C	30 seg	
Extensión	72 °C	1 min	
Extensión final	72 °C	5 min	
Espera	15 °C	∞	

Figura 11: Condiciones de temperatura de la segunda PCR de identificación de muestras.

En la figura 12 se describe el proceso de amplificación final mediante dos PCRs anidadas que se lleva a cabo para poder obtener las genotecas necesarias para ser secuenciadas con técnicas de nueva generación (NGS).

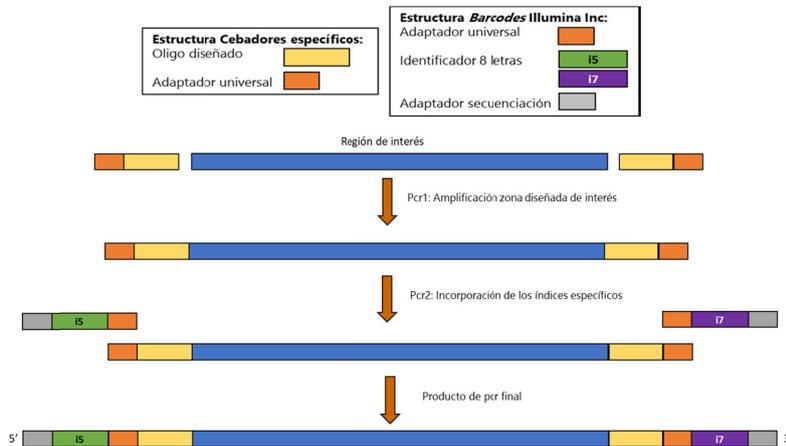


Figura 12: Esquema de las PCRs necesarias para amplificar la zona de interés y adición de identificadores. Primera PCR de amplificación de la región de interés. Segunda PCR de identificación de las muestras y amplicones mediante la adición de los *barcodes*.

1.2.5.- Tratamiento post-PCR

Al tener una gran cantidad de muestras, se prefirió trabajar con ellas asumiendo que todas tuvieron el mismo rendimiento en la amplificación y se juntaron todas en un mismo tubo o mismo *pool*.

Mediante el sistema *QiAxcel* modelo SK-QiAxcel-Slider (Qiagen, Hilden, Alemania), se verificó por electroforesis capilar, la integridad y el correcto tamaño de las diferentes multiplexes.

Seguidamente, se utilizó el kit de purificación por bolas magnéticas *MagSi-NGS PREP Plus Kit* (AMS Biotechnology (Europe) Limited, Adbington, Reino Unido) (128) a una concentración de 0.6x y siguiendo las especificaciones de la casa comercial para eliminar posibles productos inespecíficos y restos de oligonucleótidos que estuvieran en exceso en la reacción.

Se comprobó por electroforesis capilar que se habían eliminado los posibles fragmentos inespecíficos y se cuantificó el *pool* purificado de nuevo con el kit *QuantiFluor® dsDNA* (Promega, Madison, EE.UU.) (120), siguiendo las especificaciones de la casa comercial.

1.2.6.- Secuenciación con tecnología de nueva generación (NGS)

En esta parte del trabajo se utilizaron, el sistema *Miseq* (Illumina, Inc, San Diego, EE.UU.) para realizar las pruebas de puesta a punto de las condiciones óptimas de PCR y el sistema *HiSeq2000* (Illumina, Inc, San Diego, EE.UU.) para la secuenciación de toda la población con todos los amplicones seleccionados para así obtener la cantidad de lecturas necesarias para identificar las posibles variantes que se pudieran encontrar.

Ambos sistemas utilizan la tecnología típica de Illumina, que se basa en dos partes, primero una amplificación clonal y segundo un paso de secuenciación por síntesis.

1.2.6.1.- Preparación de la muestra

Para llevar a cabo la secuenciación, siguiendo el protocolo de la casa comercial (129), se diluyó el *pool* purificado hasta una concentración de 4 nM. Se desnaturalizó con NaOH y se diluyó con el tampón de hibridación (HT1) a conseguir una concentración final de 10 pM. Como control del proceso se utilizó el PhiX Control v3 (Illumina, Inc, San Diego, EE.UU.) diluido a 12.5 pM y en una proporción adecuada a la variabilidad de la muestra. Siguiendo el protocolo de la casa comercial.

1.2.6.2.- Tecnología NGS Illumina, Inc.

El tipo de secuenciación utilizada por la Tecnología Illumina consta de dos fases. Primero, sobre una superficie sólida llamada celda de flujo o *flow-cell*, en la que hay anclados pequeños oligonucleótidos complementarios a los adaptadores de secuenciación, hibridan las moléculas desnaturalizadas de ADN de la librería y se produce una primera amplificación mediante una polimerasa de alta fidelidad por extensión del extremo 3'. Seguidamente, se produce la hibridación del extremo libre con otros adaptadores de la celda de flujo en una amplificación isotérmica en forma de puente, produciéndose nuevas cadenas de ADN. Este proceso se repite muchas veces formándose agrupaciones de moléculas idénticas o *clústeres* clonales. (Figura 13).

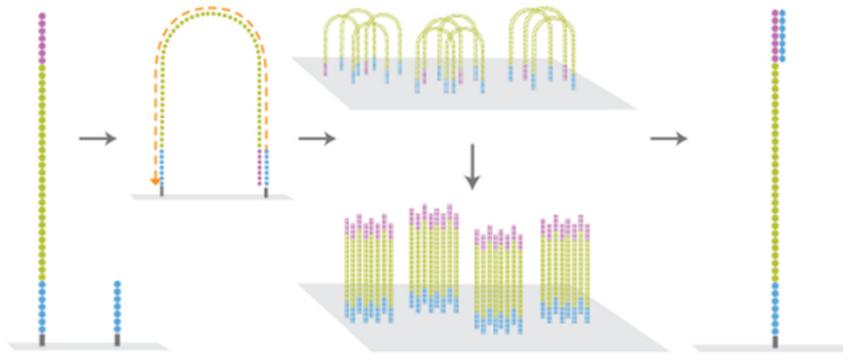


Figura 11: Proceso de amplificación clonal en forma de puente. Adaptación de Tecnología Illumina, Inc (129).

La segunda parte del proceso es la secuenciación por síntesis (SBS) característica. Se utilizan nucleótidos terminadores reversibles, fluorescentes, con un color para cada base. Se van añadiendo uno por cada ciclo de síntesis para unirse complementariamente a las cadenas de ADN hibridadas. Se sucede un paso de lavado de nucleótidos no necesarios y en cada ciclo se obtiene la imagen de señal fluorescente emitida, al terminar se agrupan todas las imágenes y se compone la secuencia completa (Figura 14).

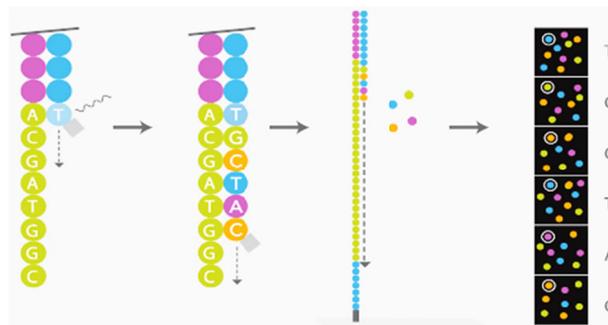


Figura 12: Proceso de secuenciación por síntesis (SBS). Adaptación de Tecnología Illumina, Inc (129).

Para la puesta a punto de las condiciones óptimas en el sistema *Miseq* se utilizaron los reactivos tipo *Miseq Reagents kit v2 (300cycles)* (Illumina, Inc, San Diego, EE.UU.) los cuales permiten una secuenciación en ambas direcciones o tipo *pair-end* de 150 pb, suficientes para cubrir al menos en una dirección la región de interés amplificada. Para la secuenciación de todas las muestras se utilizó el sistema *Hiseq2000* (Illumina, Inc) en la empresa MacroGen, Inc (Korea del sur) con una secuenciación *pair-end* de tipo 100x2.

2.- Identificación de variantes genéticas de exoma implicadas en el desarrollo de la DM2 mediada o no por resistencia a insulina (RI)

2.1.- Antecedentes previos

Como se ha comentado en el apartado 1.1 se realizaron trabajos previos de secuenciación de exoma de pacientes con DM2 y controles sanos. Los archivos de secuenciación iniciales llamados *FastQs* se analizaron de nuevo en conjunto con la nueva población para identificar posibles variantes genéticas implicadas en el desarrollo de la DM2 y así aumentar el tamaño poblacional. Como el proceso bioinformático fue diseñado de nuevo para este trabajo, se debía reanalizar todas las muestras juntas para evitar falsos negativos que podrían haber surgido de comparar resultados provenientes de dos análisis bioinformáticos diferentes.

2.2.- Diseño del estudio

En esta parte del trabajo se llevó a cabo el procesado de 119 muestras para secuenciar y analizar su exoma mediante un protocolo de captura y enriquecimiento utilizando la tecnología de secuenciación de nueva generación (NGS) en el sistema *Hiseq2000* (Illumina, Inc, San Diego, EE.UU.). Junto con 369 exomas ya procesados en proyectos anteriores (130,131) se estableció desde el principio también, un flujo de análisis bioinformático automático y se utilizaron programas de uso libre para filtrar y seleccionar las variantes de interés. Se

analizaron los resultados teniendo en cuenta el origen y tipo de muestras y se realizaron análisis estadísticos y de asociación de las variantes con parámetros clínicos utilizando el programa *PLINK v1.07* (119).

2.2.1.- Poblaciones de estudio

2.2.1.1.- Estudio Egabro-Pizarra

En esta segunda parte del trabajo se seleccionaron 119 muestras de población española del estudio Egabro-Pizarra procedentes del Hospital Regional Universitario de Málaga (Málaga, España). Esta población fue concebida como un estudio prospectivo con un seguimiento de evolución de los pacientes de 5 años. Se recogieron datos clínicos al inicio del estudio y al final. Los parámetros de inclusión fueron:

- Índice de Masa Corporal (IMC) entre 20-35%
- Resistencia a Insulina (RI): HOMA > 3.45, no RI: HOMA < 3.00
- Edad comprendida entre: 35-60 años.
- Diagnóstico de DM2 según criterios de la OMS.

Se clasificaron en 4 grupos dependiendo de la existencia o no de RI al inicio del estudio y del desarrollo o no de DM2 al final del estudio (Tabla 3).

Tabla 3: Distribución según diagnóstico de RI y/o DM2 de los pacientes de la población Egabro-Pizarra (Málaga).

	Grupo A (N = 42)	Grupo B (N = 19)	Grupo C (N = 42)	Grupo D (N = 16)
RI (SI/NO)	NO	NO	SI	SI
DM2_final (SI/NO)	NO	SI	NO	SI

RI: resistencia a insulina. DM2: Diabetes Mellitus tipo 2.

2.2.1.2.- Estudio caso/control

Esta población española fue la utilizada en las tesis doctorales previas de la UGDG (130,131). Proceden del Biobanco del CIBERDEM y del Biobanco Nacional de ADN. Los criterios de selección para las muestras fueron:

- Edad: entre 40 y 65 años
- IMC: 25 y 34.9 kg/m²
- Diagnóstico de DM2 según criterios de la OMS

Para seguir la codificación de la otra población se consideró a los controles como grupo E y a los casos como grupo F (Tabla 4).

Tabla 4: Distribución de los pacientes del estudio caso/control de estudios previos realizados por la UGDG.

	Grupo E (N = 184)	Grupo F (N = 185)
DM2 (SI/NO)	NO	SI

DM2: Diabetes Mellitus tipo 2.

2.2.2.- Obtención y preparación de las muestras del Estudio Egabro-Pizarra.

Las muestras utilizadas, procedentes del Hospital Regional Universitario de Málaga, fueron procesadas para la extracción del ADN desde sangre total usando el kit *FlexiGene DNA* (Qiagen, Hilden, Alemania) (132) siguiendo las especificaciones de la casa comercial. Se cuantificaron mediante fluorescencia con el kit *QuantiFluor® dsDNA* (Promega, Madison, EE.UU.) (120), y se diluyeron hasta 25 ng/uL por necesidad del protocolo de captura.

2.2.3.- Proceso de generación de las genotecas de exoma.

Estudio caso/control

Como se ha comentado, estas muestras fueron procesadas en trabajos previos de la unidad, en aquel momento se utilizó el kit *TruSeq DNA Sample Preparation v2* (Illumina, Inc, San Diego, EE.UU.) para generar las genotecas y el kit *Truseq Exome Enrichment* (Illumina, Inc, San Diego, EE.UU.) para el enriquecimiento del exoma. Se utilizaron los reactivos necesarios para el uso del equipo *HiScanSQ* (Illumina, Inc, San Diego, EE.UU.) perteneciente a la UGDG. Esta preparación fue algo diferente a la utilizada en la siguiente población y puede encontrarse en los trabajos ya mencionados.

Estudio Egabro-Pizarra

Para poder obtener las genotecas de exoma se utilizó el kit *QXT SureSelect Target Enrichment for Illumina Multiplexed Sequencing. Featuring Transposase-Based Library Prep Technology* (Agilent Technologies, Santa Clara, EE.UU.) (133) y para el proceso de

enriquecimiento en exoma se utilizó el kit *SureSelect Target Enrichment System Human all exon V5 + UTRs* (Agilent Technologies, Santa Clara, EE.UU.) (133). El flujo de trabajo general utilizado se puede ver en la figura 15 y se describe en los apartados posteriores.

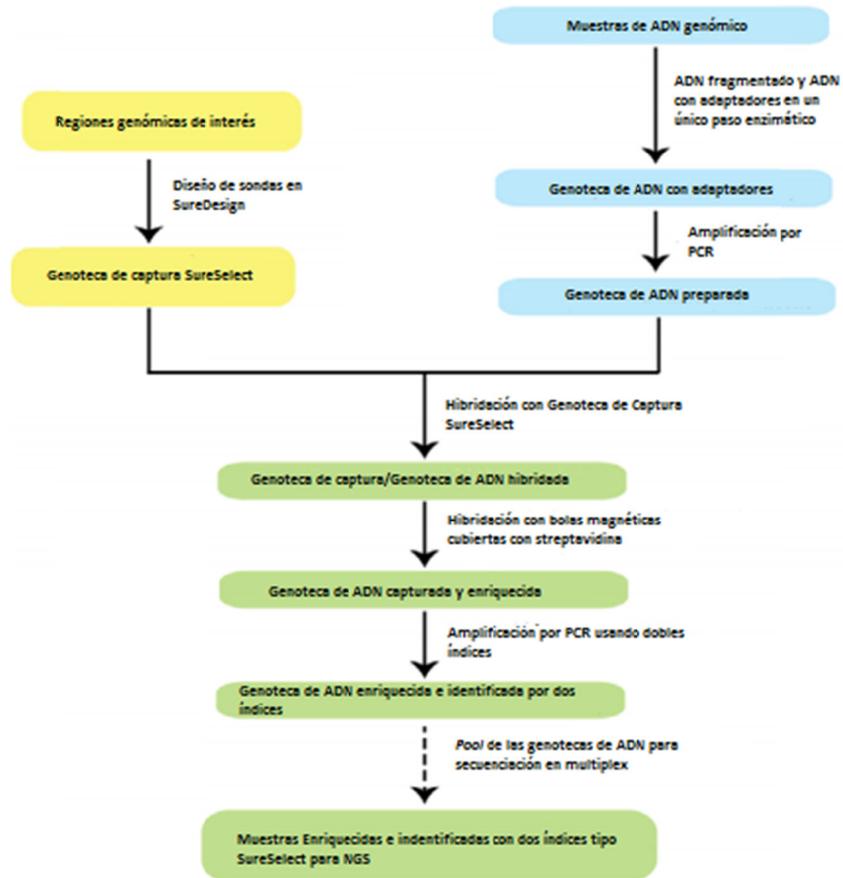


Figura 13: Flujo de trabajo para la generación y enriquecimiento en exoma de librerías de ADN genómico. (Adaptación del protocolo de *QXT SureSelect Target Enrichment for Illumina Multiplexed Sequencing, Featuring Transposase-Based Library Prep Technology, version C1, june 2015*) (133).

2.2.3.1.- Fragmentación e incorporación de adaptadores

El primer paso del protocolo fue la fragmentación enzimática del ADN para generar moléculas más pequeñas mediante enzimas transposasas. Esta técnica permite utilizar poca cantidad de material de partida y procesar muchas muestras a la vez. Después de una purificación con bolas magnéticas, para eliminar restos de enzima y otras moléculas no deseadas, se llevó a cabo la reparación de los extremos de los fragmentos generados y la adición de adaptadores necesarios para el proceso de captura, mediante amplificación por PCR de alta fidelidad. Después de este proceso se purificó con bolas magnéticas y se visualizaron las genotecas en el Sistema *Qiaxcel* modelo SK-QIAxcel-Slider (Qiagen, Hilden, Alemania) (Figura 16A).

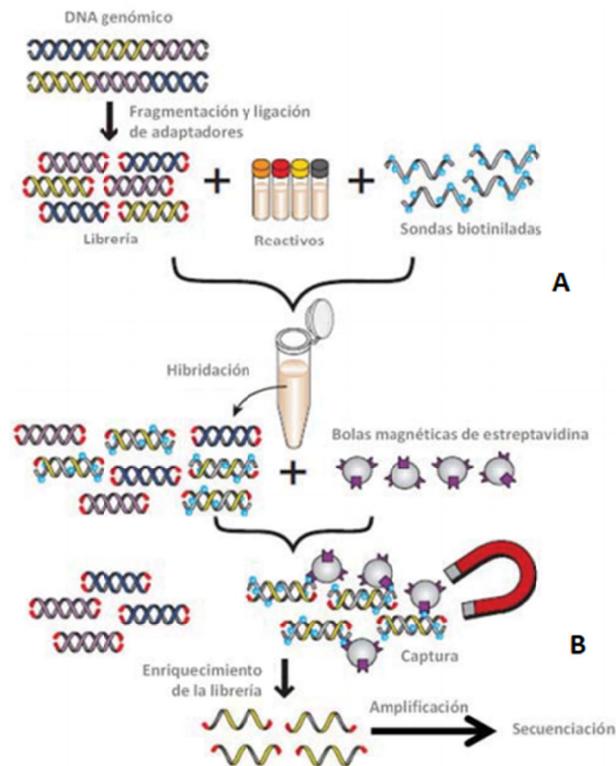


Figura 14: Esquema del proceso de generación de librerías de ADN y enriquecimiento en exoma para secuenciación por NGS. A) Fragmentación y adición de adaptadores. B) Hibridación y captura de enriquecimiento en exoma mediante sistema biotina-estreptavidina y bolas magnéticas, identificación de las muestras y finalización del proceso para secuenciación. (Adaptación Agilent Technologies, Santa Clara, Inc).

2.2.3.2.- Hibridación y captura

El siguiente paso fue añadir las sondas comerciales *All human exón v5 + UTRs*, (Agilent Technologies, Santa Clara, EE.UU.) que cubren todo el exoma y extremos UTRs de todos los genes. Las características se muestran en la tabla 5. Estas sondas están marcadas con biotina y al ser complementarias a las regiones genómicas de interés, mediante

una hibridación se seleccionaron dichas zonas de forma bastante específica.

Tabla 5: Características generales del kit *Surselect All human exon v5 + UTRs* (Adaptación Agilent Technologies, Inc., Santa Clara, EE.UU.).

	Parámetros	All exon V5 + UTRs
Detalles del diseño	Regiones de interés	75 Mb
	# Genes	21522
	# Exones diana	359555
Detalles del flujo de trabajo	Cantidad de producto total	6 Gb
	Tiempo procesado	1.5 días

A continuación, se añadieron bolas magnéticas recubiertas con estreptavidina que permitieron seleccionar los fragmentos hibridados y así capturar solo las regiones de interés. Se hicieron varios lavados para eliminar restos de sondas, bolas y fragmentos inespecíficos (Figura 16B).

2.2.3.3.- Indexación y Secuenciación

Una vez capturadas las regiones de interés se pasó a identificar cada muestra de forma única, utilizando dos índices o *barcodes* compatibles con la tecnología Illumina. Siguiendo las especificaciones del protocolo de Agilent, se combinaron dichos índices de forma que se pudieran identificar el mayor número de muestras posibles. Por último, se purificó con bolas magnéticas, se visualizó en el sistema *Qiaxcel* modelo SK-QIAxcel-Slider (Qiagen, Hilden, Alemania) y se cuantificó con el kit *QuantiFluor® dsDNA* (Promega, Madison, EE.UU.)

(120) para poder diluir las muestras y mezclarlas en un solo *pool* para enviar a secuenciar con un sistema *Hiseq2000* (Illumina, Inc, San Diego, EE.UU.) con una cobertura de 50x en una secuenciación tipo *pair-end* de 100x2 en el Centro CRG-CNAG de Barcelona. El proceso de secuenciación es el mismo descrito en el punto 1.2.6.2.

3.- Análisis bioinformático

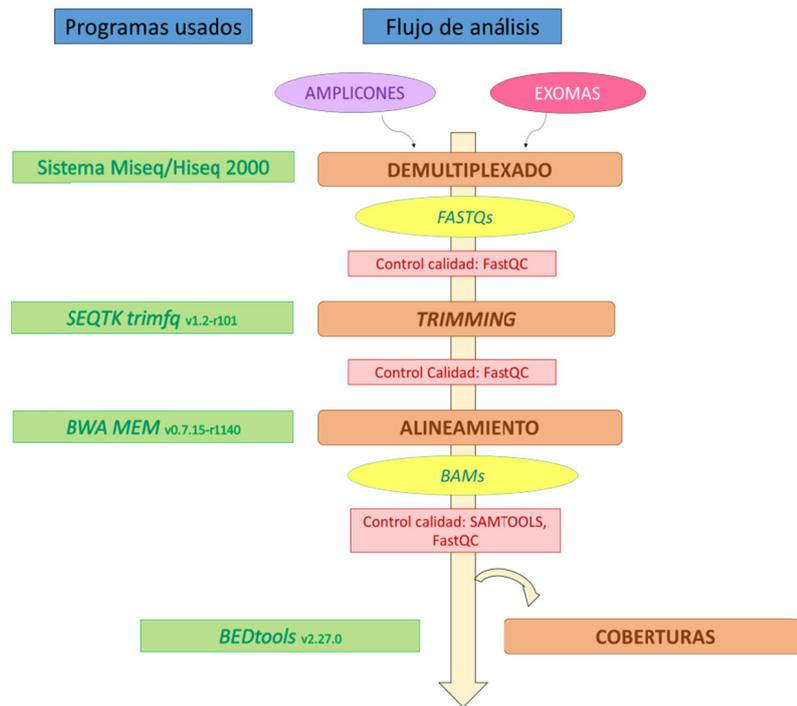
Tanto en el apartado 1 (Proyecto de validación de variantes en amplicones) como en el apartado 2 (Identificación de variantes en muestras de exoma) el proceso de secuenciación es el mismo. Por lo que se han desarrollado dos flujos de trabajo bioinformático o *pipelines* parecidas, con una parte común, mediante el uso de aplicaciones de software libre, y otra más específica, mediante programas o *scripts* manuales dependiendo de la finalidad del análisis.

Se ha llevado a cabo con la principal ayuda de Azahara Fuentes Trillo y de Carolina Monzó Cataluña, bioinformáticas en la UGDG de INCLIVA y para la gestión de recursos se ha contado con la Unidad de Bioinformática de INCLIVA.

3.1.- Flujo de análisis bioinformático general o *pipeline*

En la figura 17 se muestra la *pipeline* integrada, llamada *miracleV2*, utilizada para llevar a cabo el análisis bioinformático de los datos obtenidos en este trabajo. En ella se detallan los nombres de los procesos llevados a cabo, los programas informáticos de acceso libre

usados, los puntos de control de calidad y los archivos utilizados y generados en cada etapa según el análisis. Cada parte de la misma se probó y se puso a punto por separada antes de integrarla en el programa o *script* automático final. En todos los pasos que lo permitieron, se utilizó el programa *GNU parallel* (134) para la paralelización de los mismos y aprovechamiento de los recursos computacionales.



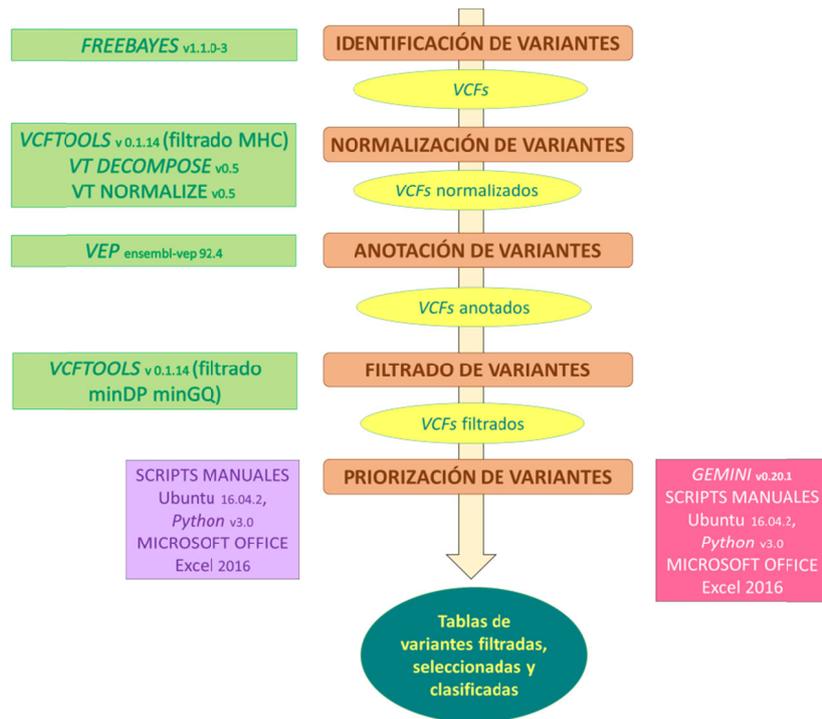


Figura 15: Esquema del flujo de análisis bioinformático utilizado en este trabajo. Los programas utilizados se detallan a la izquierda de la figura en fondo verde. Los procesos realizados se identifican sobre fondo anaranjado en la parte central. Los archivos producidos en cada proceso se resaltan en fondo amarillo. Los controles de calidad se visualizan en rojo. Todo el proceso ha sido el mismo salvo las modificaciones pertinentes dependiendo del experimento a analizar, se ha señalado en morado para el flujo de análisis de los genes del apartado 1 y en rosa para el análisis de exoma del apartado 2 de este trabajo.

3.1.1.- Demultiplexado y generación de archivos *FastQ*

Durante cada ciclo de secuenciación, el programa informático propio del aparato mide la intensidad de las señales fluorescentes emitidas por cada base marcada, además de la información de la confianza del

identificación de cada base como una nota de calidad y los almacena en archivos binarios llamados BCL (en inglés *Base Call*).

Con la información recogida en un archivo llamado *samplesheet.csv*, que enlaza el nombre de la muestra y la combinación de índices específicos que la identifican, se facilita el reconocimiento de cada lectura con cada muestra llevando a cabo el demultiplexado y la unificación de toda la información en un archivo llamado *FastQ*.

Como se ha realizado una secuenciación *pair-end* o bidireccional, cada muestra codificada tendrá dos archivos *FastQ* correspondientes a cada tipo de lectura, R1 o *forward* y R2 o *reverse*.

Estos archivos *FastQ* son archivos iniciales que se usaron como datos de entrada de las *pipelines* desarrolladas. Tienen la estructura característica representada en la figura 18. Se componen de grupos de 4 líneas que siempre siguen el mismo patrón, su cantidad depende del número de lecturas identificadas por cada muestra. El patrón que siguen es el siguiente:

- La primera línea es el identificador de la lectura, donde se recoge la información básica del aparato que la ha generado, las coordenadas del clúster donde se ha generado, si es la lectura R1 o R2 y los índices que la han identificado.
- La segunda línea es la lectura secuenciada.
- La tercera es el indicador de la calidad representado con un símbolo "+".
- La cuarta línea son los valores de calidad de la secuencia de la línea 2 codificados en escala logarítmica *Phred Quality Score* (Q) y expresados en código ASCII. El valor Q está relacionado

- **Valores de calidad por secuencia:** Representa la relación entre los valores de calidad Q y el número de lecturas que poseen dicho valor.
- **Contenido de bases por secuencia:** Se ve representada la proporción de cada una de las 4 bases del ADN por cada posición en la lectura.
- **Contenido de pares GC en la secuencia:** Se compara el contenido de pares GC en toda la longitud de cada secuencia respecto a una distribución normal consensuada.
- **Contenido de N por secuencia:** Se representa en una gráfica el porcentaje de bases no identificadas en una secuencia. Al no haber una suficiente confianza en la identificación, se sustituye por N la base llamada.
- **Distribución de la longitud de la secuencia:** Permite visualizar la longitud y tamaño de las lecturas generadas.
- **Secuencias duplicadas:** En esta gráfica se representa el grado de duplicación por cada secuencia analizada.
- **Secuencias sobrerrepresentadas:** Este módulo dará algún dato si hay alguna secuencia repetida a partir de un 0.1% del total.
- **Contenido de adaptadores:** Aquí se reflejarán cualquier sobrerrepresentación en las secuencias de cualquiera de los diferentes adaptadores conocidos a partir de un 5% sobre la secuencia total.
- **Contenido en K-mer:** Este parámetro mostrará cualquier secuencia pequeña que al analizar el total de secuencias con

un test binomial tenga una desviación significativa de sobrerrepresentación.

El comando genérico utilizado fue:

```
fastqc {input}.FastQ -o {output_name} --extract
```

El segundo paso después de conocer el estado de las secuencias, es el recorte de las mismas teniendo en cuenta su calidad. Normalmente, los extremos de las lecturas suelen tener más errores en el proceso de secuenciación y la calidad de las bases es más baja, por ello se realiza un paso de limpieza de extremos o *trimming* para homogeneizar la calidad de las lecturas recortando aquellas cuyas bases posean un valor de calidad (Q) menor al umbral especificado. Para ello se ha utilizado la herramienta *seqtk trimfq* (136) en su versión *v1.2*. El comando genérico fue el siguiente:

```
seqtk trimfq {input}.FastQ | gzip >  
{output_name}.trimmed.FastQ"
```

3.1.4.- Alineamiento

Después de seleccionar las lecturas de mejor calidad es necesario realizar un alineamiento con un genoma de referencia para saber exactamente qué región se ha secuenciado. En este caso se ha utilizado en genoma de referencia humano *hs37d5* y se ha llevado a cabo el alineamiento con el software BWA (*Burrows-Wheeler-Aligner*) *v0.7.15* (137). Este programa tiene varios algoritmos; BWA-backtrack, BWA-SW y BWA-MEM. En este caso se ha utilizado el BWA-MEM por el ser más reciente y el que mejor se ajustaba a las características de las secuencias. Este algoritmo indexa el genoma de referencia para

Para poder continuar con el siguiente proceso, es necesario tener ordenados los archivos BAM por cromosoma y posición, para ello se ha utilizado el software *SAMtools v1.7* (139) con las opciones *sort* e *index* para llevar a cabo la ordenación e indexación de los archivos. Los comandos usados fueron:

```
samtools sort -O bam -o {sample_name}-sorted.bam  
samtools index {sample_name}-sorted.bam"
```

Otro proceso necesario es la eliminación de duplicados de las lecturas, que se crean fácilmente en cualquier amplificación por PCR. Para ello se utilizó la opción de *SAMtools rmdup*.

```
samtools rmdup -S {sample_name}-sorted.bam {sample_name}-  
rmdup.bam
```

3.1.5.- Coberturas

Una vez procesados los archivos *BAM* se pudo realizar el análisis de coberturas para visualizar a modo de tabla y/o gráfica el estado de las muestras. Para ello se utilizó el programa *BEDtools v2.27.0* (140) y los archivos *bed* con las localizaciones de las zonas secuenciadas para crear un histograma que mostrara la distribución general de la cobertura en las diferentes poblaciones analizadas. Se hicieron varios sub-análisis a partir de estos datos para conocer otros parámetros interesantes. El comando general utilizado fue:

```
bedtools intersect -bed -g ../bed_files/chrom_complete.sizes -
a {bam_file} -b {bed_file} | bedtools coverage -hist -b - -a
{bed_file} | grep ^all > {output_dir}
/all_{sample_name}.hist.txt
```

3.1.6.- Identificación de variantes

El siguiente paso es el proceso de identificación de variantes o *calling variant*, es decir, a partir de haber alineado las secuencias contra el genoma de referencia, ver si había sido una alineación exacta o ha habido algún cambio con respecto a la referencia. Para ello se utilizó el programa *Freebayes v1.1.10* (141). Este es el comando:

```
freebayes -f {reference_genome}.fa -t {input}.bed --genotype-
qualities --strict-vcf -C 1 {input}.bam | bgzip -c >
{output name}.vcf
```

En este paso se procesan los archivos BAM y se crean los archivos *VCF* (Variant Calling Format) (142) que contienen la información de las variantes identificadas. Es un archivo que también se divide en dos partes, la cabecera que almacena toda la información del proceso, las características del programa usado, detalla el significado de los campos que se codifican las características de las variantes, el comando usado, el genoma de referencia y cualquier otro apunte que sea necesario para explicar con claridad cómo se ha creado el archivo. A continuación, está el campo de las variantes encontradas (Figura 20). En cada línea del archivo se detalla una variante, se especifica el cromosoma, posición, el alelo de referencia y el/los alternativos, se muestra la calidad del proceso de identificación de cada una, la

profundidad de lectura que ha permitido llamarla, la frecuencia de cada alelo encontrado y el genotipo.

```
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT AG-417-MIX-A-MIX-8-08-Peque_SS2
chr1 115252168 . A 290.777 . DP=541;MQSB=0.362529;MQ0F=0;AF1=0;AC1=0;
DP4=264,5,0,0;MQ=49;FQ=-289.528 GT:PL:DP:ADF:ADR:AD:GQ 0/0:0:269:264,0:5,0:269,0:99
chr1 115252169 . A 290.777 . DP=543;MQSB=0.831993;MQ0F=0;AF1=0;AC1=0;
DP4=267,6,0,0;MQ=49;FQ=-289.528 GT:PL:DP:ADF:ADR:AD:GQ 0/0:0:273:267,0:6,0:273,0:99
chr1 115252170 . T 290.777 . DP=544;MQSB=0.361511;MQ0F=0;AF1=0;AC1=0;
DP4=270,3,0,0;MQ=49;FQ=-289.528 GT:PL:DP:ADF:ADR:AD:GQ 0/0:0:273:270,0:3,0:273,0:99
```

Figura 18: Extracto de un archivo VCF donde se puede ver por cada línea la variante llamada y la información correspondiente.

3.1.7.- Normalización de las variantes

Sucesivamente a la identificación de variantes se utilizó el programa *Vt v0.5* (Variant Tool) (143) para descomponer y normalizar las variantes identificadas que presentaban un cambio multialélico, es decir, que tenían más de un alelo alternativo, se debían separar, una por línea, ya que cada cambio podría tener un efecto diferente, convirtiéndose en otra variante. Para ello se usaron dos opciones de este programa; *decompose* y *normalize*. Los comandos usados fueron los siguientes:

```
vt decompose -s {input}.vcf |
vt normalize -r {reference_genome}.fa -n - > {sample_name}-
dec-norm.vcf
```

Con el fin de evitar variantes situadas en zonas hipervariables del genoma, como el Complejo Mayor de Histocompatibilidad (MHC) codificado en el cromosoma 6, esta parte se eliminó mediante el uso de la herramienta *VCftools v0.1.14* (144).

```
vcftools --gzvcf {input_chr6}.vcf.gz --recode --recode-INFO-
all --exclude-bed {MHC_regions}.bed --out {output}-
noMHC.vcf.gz
```

3.1.8.- Anotación de variantes

A continuación, se llevó a cabo el proceso de anotación de variantes. Para ello se utilizó el programa *VEP* (145) con la versión de variantes de la base de datos *Ensembl 92*. La cual es una amplia fuente de información de funcionalidad de variantes, su impacto, localización génica, frecuencias alélicas, valores de patogenicidad y otros detalles importantes para conocer la variante analizada. Este fue el comando usado:

```
vep --cache --dir $vep_db --assembly GRCh37 --offline --
force_overwrite --sift b --polyphen b --check_existing --
allele_number --variant_class --canonical -i {input}.vcf.gz --
format vcf --compress_output bgzip --vcf -o {output_name}-dec-
norm-VEP.vcf.gz --af_1kg --af_esp --af_gnomad
```

3.1.9.- Filtrado de variantes

Una vez anotadas las variantes con la información disponible se hizo un filtrado de las mismas usando el programa *VCFTools v0.1.14* (144) usando los parámetros *minDP*, valor mínimo de cobertura para considerar una variante creíble y *minGQ*, valor mínimo de calidad del genotipado para considerar que una variante no es un error mendeliano. Este es el comando usado:

```
vcftools --gzvcf {input}-dec-norm-VEP.vcf.gz --recode --recode-
INFO-all --minDP 10 --minGQ 20 --out {output_name}-
minDP10minGQ20.vcf.gz"
```

3.1.9.1.- Priorización de variantes

En referencia a lo comentado al principio de este capítulo, hasta este punto, todo el flujo de análisis que se llevó a cabo tanto para las

muestras de amplicones como las de exoma, fue prácticamente el mismo, adecuando los programas al tipo de datos analizados. Dada la diferente complejidad y cantidad del contenido de los archivos finales, se decidieron realizar dos tipos de análisis para la priorización de variantes.

Para el apartado 1 de resultados este trabajo, la validación de genes por amplicones, se desarrollaron una serie de pasos de análisis para estructurar y organizar de una forma más eficiente y visual los datos de las variantes, con el fin de favorecer su interpretación posterior. Estos pasos se automatizaron una vez puestos a punto, para así tener un *script* de análisis sencillo y disponible para procesar cualquier dato de esta naturaleza en proyectos futuros. Se utilizó el lenguaje de programación *Python v2.7*

Con esta *pipeline* automatizada se generaron archivos por gen de todas las variantes encontradas, así como los análisis de dichas variantes con el programa estadístico *PLINK*.

A partir de ahí se organizaron, filtraron y adecuaron los resultados para poder presentarlos con la mayor claridad posible.

Para el apartado 2 de resultados de exoma, se decidió utilizar *GEMINI v0.20.1* (GEnome MINIng) (146). Es una herramienta que trabaja sobre bases de datos relacionales estructuradas tipo SQL (*Structured Query Language*) es decir, son tablas relacionadas entre sí por una columna común o *primary key* y que permite hacer varias acciones, filtrados y llamamientos sobre diferentes características a la vez de forma rápida y eficaz.

Este programa precisa de *vcfs* anotados con *VEP* o *SNPeffect* y procesados de una manera especial para poder ser utilizado, como la decomposición y normalización de variantes, también necesita que los archivos se compriman en forma *bgzip* y se cree un índice de las posiciones genómicas tipo *tabix* para facilitar la búsqueda a la hora de procesar los archivos. Para estas últimas acciones se utilizaron los programas *bgzip v1.7(147)* y *tabix v1.7(148)*. Los comandos fueron:

```
bgzip -c {input}.vcf > {output_name}.vcf.gz
tabix -p vcf {input}.vcf.gz
```

Otro requisito necesario, es el uso de un archivo de relación de las muestras utilizadas con su fenotipo o si es el caso, entre ellas mismas para análisis de parentesco. Se conoce como archivo *ped* (pedigree file) y se compone de 6 columnas obligatorias separadas por tabulación y otras adicionales además de que es necesario respetar su codificación de las variables. La cabecera del archivo debe tener los siguientes campos:

#family_id, name, paternal_id, maternal_id, sex, phenotype

Family_id: relación de parentesco en caso de este tipo de análisis, sino debe ser igual a *names*.

Names: nombre de la muestra tal cual aparece en el archivo *VCF*.

Paternal_id y *Maternal_id*: nombre la de muestra que tiene este grado de parentesco. Si no existe se codifica con 0 o -9.

Sex: Sexo del individuo codificado como 1=hombre, 2=mujer otro=0 o -9

Phenotype: Diferencia entre las muestras. En estudios caso/control, control=1, caso=2. Si hay más de un fenotipo se puede seguir codificando en números.

GEMINI además de usar la anotación de la base de datos *VEP* permite incorporar otras bases de información interesante para completar el estudio, en este caso se añadieron los datos procedentes de la base de datos *CADD* (*Combined Annotation Dependent Depletion*) (149) la cual le da un valor de patogenicidad a la variante teniendo en cuenta otros parámetros externos. A partir de un valor 20 se considera una variante patogénica.

Para poder trabajar con *GEMINI*, primero es necesario crear la base de datos tipo *SQL*; para ello se utilizó el siguiente comando:

```
gemin_i load -v {input}-dec-norm-VEP.vcf.gz -t VEP -p {pedigree}.ped {output_database_name}.db
```

Una vez creada, se pueden realizar infinidad de análisis y obtener diversos datos. En este trabajo se establecieron ciertas pautas:

- Extraer variantes genotipadas en un 60% o más de las muestras secuenciadas como filtro general.
- Utilizar el transcrito considerado canónico del gen.
- Considerar solo aquellas variantes en las que, al menos, haya una muestra con genotipo heterocigoto (0/1) y/o homocigoto recesivo (1/1).
- Extraer las variantes de cada grupo de muestras, así como por agrupación por fenotipo o tipo de estudio.

A continuación, se muestra un comando ejemplo que cumple todas estas condiciones para obtener las variantes del grupo A codificado con el fenotipo 1 en el archivo *ped*.

```
gemini query --header -q "select chrom, start, end, ref, alt,
rs_ids, type, sub_type, gene, transcript, aa_change,aa_length,
impact_severity, impact_so, aaf, aaf_gnomad_nfe, cadd_scaled,
hwe, call_rate, (gts).(phenotype==1) from variants where
vep_canonical = 'YES' and call_rate >= 0.6" --gt-filter
"(gt_types).(phenotype==1).(==HET).(any) or
(gt_types).(phenotype==1).(==HOM_ALT).(any)"
{input_database}.db > {output_name}-groupA-basal.txt"
```

Una vez generados todos los archivos necesarios, se trabajó su organización, filtrado y presentación final de manera clara y eficiente.

Por otro lado, se quiso averiguar las diferentes funciones moleculares, biológicas y rutas metabólicas donde los genes encontrados en el análisis de exoma (apartado 2). Para ello se utilizó por simplicidad de ejecución y resultados, el programa libre *PANTHER Gene List Analysis* (150,151).

3.1.10.- Validación del flujo bioinformático

Con el fin de validar el flujo de análisis *miracleV2*, se utilizaron datos de secuenciación tipo *Gold Standard*, es decir, que han sido verificados por diferentes técnicas de secuenciación y análisis. Se utilizó la NIST7035 de exoma completo, del GIAB (*The Genome in a Bottle Consortium*), el cual es una infraestructura del NIST (*National Institute of Standards and Technology*) que tiene el objetivo de

caracterizar genomas humanos para el análisis, validación y desarrollo de tecnología para la práctica clínica.

Los datos brutos de esta muestra, fueron analizados con *miracleV2* hasta generar el archivo *VCF* y posteriormente comparados frente al *VCF* original con el fin de conocer la sensibilidad y especificidad del flujo bioinformático utilizado. Para ello se utilizó la herramienta *vcfeval* de RTGtools *v3.10.1* (152). El comando utilizado fue el siguiente:

```
vcfeval -b {input}NIST.vcf.gz -c all-bg.vcf.gz --evaluation-  
regions intersect_exome_NIST.bed -o {output_name}.vcf.gz -t  
{reference_genome}.fa --decompose --sample {input_NIST}.fastq  
--squash-ploidy
```

4.- Análisis estadístico

4.1.- Estadística general

Las características de las poblaciones se expresaron aquellas cuantitativas como media \pm desviación estándar y las cualitativas como porcentaje para recoger de forma visual la información.

4.2.- Estudios de asociación de variantes

Con el fin de conocer si la presencia de las variantes encontradas en un grupo control o caso determinaba un posible efecto protector o de riesgo, o si tenían alguna relación con las variables clínicas de la población, se utilizó el programa *PLINK v1.07* (119) para llevar a cabo estos estudios de asociación utilizando diversos test estadísticos

conocidos ya que esta herramienta permite utilizarlos de forma sencilla.

Para el uso de este programa se deben estructurar los datos en varios archivos de texto específicos:

- Archivo *ped*. Posee una estructura parecida al usado para *GEMINI*. La cabecera debe tener los siguientes campos:
sample_id, sample_name, father, mother, sex, group, variant 1, variant 2...

Las variantes se deben codificar con letra y dejando un espacio entre los alelos. Ej. A C.

Para su uso la cabecera debe ser eliminada.

- Archivo *map*. Este archivo de texto almacena el nombre de la variante añadida en el archivo *ped*, en el mismo orden, el cromosoma y la posición de la misma en un formato de 4 columnas también sin cabeceras.

Tanto el archivo *ped* como *map* deben de llamarse de la misma manera para que el sistema pueda reconocer que archivos están relacionados.

4.2.1.- Análisis básicos de asociación

En este apartado se utilizó el test de asociación alélica básica, Test de Fisher, en vez del Test Chi-cuadrado, al no haber al menos 5 elementos mínimos en cada variable. Además, se tenía la opción de filtrar aquellas variantes que no cumplieran el Equilibrio de Hardy-Weingberg (EHW) y calcular el intervalo de confianza el 95%. El

comando general utilizado fue el siguiente utilizando como ejemplo los archivos *analisis1.ped* y *analisis1.map*

```
Test de Fisher:  
plink --file analisis1 --fisher --ci 0.95 --hardy -out  
{output_name}
```

Como resultado, se obtuvieron dos archivos, uno con el análisis estadístico por variante con los valores del intervalo de confianza y el p-valor asociado para las variantes que cumplían el EHW y otro con los valores de este equilibrio de todas las variantes. Además, se tuvo en cuenta no solo el p-valor significativo del test Fisher, sino que también se valoraron aquellas variantes que no siendo significativas o si, tenían un valor de OR (*Odds Ratio*) menores a 0.5 y mayores a 2 ya que permite identificar aquellas variantes con un posible efecto protector o un posible efecto de riesgo respectivamente.

Este programa también permite analizar qué tipo de modelo genético podía seguir la variante analizada. Tiene en cuenta 5 tipos de modelos teniendo en cuenta que *a* es el alelo referencia y *A* el alelo alternativo. Dependerá su codificación del método de asociación utilizado: (153).

- Test de Tendencia de Cochran-Armitage (Modelo Aditivo) (TREND): Test estadístico para analizar datos ordenados en categorías. Considera una tabla de contingencia de 2x3 teniendo en cuenta los 3 genotipos, AA, A/a, a/a de forma aditiva con 1 grado de libertad.

- Test Alélico (ALL): A vs a. Solo tiene en cuenta la presencia de cada alelo en una tabla de contingencia 2x2 con 1 grado de libertad.
- Test genotípico (Modelo codominante) (GENO): Es el test general de asociación con 2 grados de libertad para la distribución de tabla 2x3, el cual compara todos los genotipos contra todos con la misma importancia; AA vs Aa vs aa.
- Test de acción dominante (DOM): Este test tiene en cuenta la dominancia del alelo A frente al a. En el cual el incremento de copias del alelo A incrementa el riesgo de la enfermedad. La tabla de contingencia es de 2x2 con un 1 grado de libertad del genotipo A/A frente a la suma de los genotipos A/a y a/a.
- Test de acción recesiva (REC): Este test contempla la recesividad del alelo A minoritario sobre el alelo a, en el cual son necesarias las 2 copias de A para incrementar el riesgo de la enfermedad. Se describe con una tabla de contingencia de 2x2 de 1 grado de libertad, en la que se cuenta el genotipo a/a vs la suma de AA y Aa.

Con el fin de saber qué modelo genético sigue la variante, al comando anterior se le debía de añadir la opción `--model`. Esto generaba otro archivo con los p-valores asociados a cada tipo de modelo para cada una de las variantes analizadas. Comando usado:

```
Test de Fisher y modelo genético:  
plink --file analisis1 -fisher --model --ci 0.95 --hardy -out  
{output_name}
```


RESULTADOS Y DISCUSIÓN

Debido a la naturaleza de los resultados de este trabajo de tesis doctoral, los directores y yo hemos convenido, con el fin de facilitar su comprensión, realizar la exposición de los resultados en conjunto con la discusión de los mismos.

1.- Validación de los genes seleccionados en el estudio previo de exoma.

1.1.- Características de la población de estudio

La descripción de las características generales de la población utilizada en esta parte del trabajo se puede ver en la tabla 6. La selección final por edad incluyó a 832 controles (462 hombres, 369 mujeres y 1 sin datos) con glucemia normal, por debajo de 110 mg/dL y 395 casos de DM2 (258 hombres, 278 mujeres y 52 sin datos), el 81.06% de los pacientes de DM2 estaban tratados.

Tabla 6: Características clínicas de la población de estudio.

<i>Características</i>	Controles (N = 832)	Casos (N = 588)
Sexo:		
Hombre (%)	462 (55.53)	258 (43.88)
Mujer (%)	369 (44.35)	278 (47.28)
NA (%)	1 (0.12)	52 (8.84)
Edad (años)	55.92 ± 4.56	53.78 ± 18.02
IMC (Kg/m²)	27.26 ± 3.81	27.91 ± 11.11
Glucosa basal (mg/dL)	< 110	139.58 ± 74.16

Valores expresados como media ± desviación estándar. IMC: Índice de masa corporal.

1.2.- Diseño de genes seleccionados

Basándose en datos previos de la UGDG, se seleccionaron 19 genes para amplificar las regiones exónicas y las zonas de *splicing* contiguas con el fin de analizar las posibles variantes genéticas que pudieran aparecer y poder asociarlas o no con el desarrollo de la DM2. En la tabla 7 se puede ver la información de los genes seleccionados.

Tabla 7: Información general de los genes secuenciados.

Chr	Posición	Gen	Ensembl ID	Ensembl Transcrito	Nº exones
2	223725652-223809357	<i>ACSL3</i>	ENSG00000123983	ENST00000357430	17
2	27719709-27746554	<i>GCKR</i>	ENSG00000087734	ENST00000264717	19
3	140950670-141013748	<i>ACPL2</i>	ENSG00000155893	ENST00000286353	6
3	73110810-73112488	<i>EBLN2</i>	ENSG00000255423	ENST00000533473	1
4	106473777-106629250	<i>ARHGEF38</i>	ENSG00000236699	ENST0000042047.2	14
6	110712974-110736765	<i>DDO</i>	ENSG00000203797	ENST00000368624	5
10	115310596-115349361	<i>HABP2</i>	ENSG00000148702	ENST00000351270	13
11	60704556-60720002	<i>SLC15A3</i>	ENSG00000110446	ENST00000227880	8
11	7506619-7532608	<i>OLFML1</i>	ENSG00000183801	ENST00000329293.3	3
12	112123857-112194903	<i>ACAD10</i>	ENSG00000111271	ENST00000013698	21
15	43477455-43487396	<i>CCNDBP1</i>	ENSG00000166946	ENST00000300213	11
15	75011883-75017951	<i>CYP1A1</i>	ENSG00000140465	ENST00000395048	7
16	20775024-20808903	<i>ACSM3</i>	ENSG00000005187	ENST00000289416	14
16	57462081-57481440	<i>CIAPIN1</i>	ENSG00000005194	ENST0000039439.8	9
16	67977377-68003504	<i>SLC12A4</i>	ENSG00000124067	ENST00000316341	24
17	64831235-64881603	<i>CACNG5</i>	ENSG00000075429	ENST00000533854	6
17	76183443-76203782	<i>AFMID</i>	ENSG00000183077	ENST00000327898	11
22	17071667-17073700	<i>CCT8L2</i>	ENSG00000198445	ENST00000359963	1
22	42229109-42303312	<i>SREBF2</i>	ENSG00000198911	ENST00000361204	19

Cr: cromosoma.

1.3.- Control de calidad: Programa *FastQC*

El primer punto de control del proceso es analizar la calidad de los archivos *FastQ* que se generan en el secuenciador y comprobar la calidad de los mismos después de ser procesados en el paso de *trimming*. Para ello se utilizó el programa *FastQC v0.11.5* (135) y se evaluaron los parámetros de calidad resultantes. En la tabla 8 se

pueden ver los valores expresados en porcentaje de los 4 parámetros de calidad más interesantes.

Tabla 8: Valores de los parámetros del control de calidad de los archivos *FastQ* de las muestras del estudio antes y después del proceso de *trimming*.

PARÁMETROS	<i>FastQ</i> preprocesados			<i>FastQ</i> postprocesados		
	<i>PASS</i>	<i>WARN</i>	<i>FAIL</i>	<i>PASS</i>	<i>WARN</i>	<i>FAIL</i>
Estadísticos básicos	100.00	0.00	0.00	100.00	0.00	0.00
Calidad de secuencia por base	46.58	11.48	41.95	77.03	11.23	11.75
Valores de calidad por secuencia	99.29	0.16	0.55	99.64	0.04	0.32
Contenido de N por secuencia	35.57	35.19	29.25	39.92	46.81	13.27
Duplicación de secuencias	0.92	1.09	97.99	0.92	1.14	97.94

Valores expresados en porcentaje.

Como se puede observar, la calidad de las secuencias mejora considerablemente después de llevar a cabo el proceso de *trimming* eliminando las bases con menor valor de calidad y aumentando la calidad global de la muestra. No obstante, los niveles de duplicados de este tipo de muestras son muy altos debido a que no es posible distinguir la cobertura real de las secuencias de los duplicados generados por PCR en los procesos previos a la secuenciación, ya que se parte de una amplificación por PCR. En este caso, este parámetro no es tenido en cuenta a la hora de seguir con el análisis.

1.4.- Análisis de Coberturas

Una vez secuenciados todos los amplicones, se llevó a cabo el análisis de la cobertura de los mismos, independientemente para cada gen,

teniendo en cuenta todos los individuos de la población. Para ello se contó con la ayuda, de nuevo, de Azahara Fuentes. En la figura 21 se puede ver la representación para cada amplicón (líneas grises), la media de cobertura de todas las muestras secuenciadas.

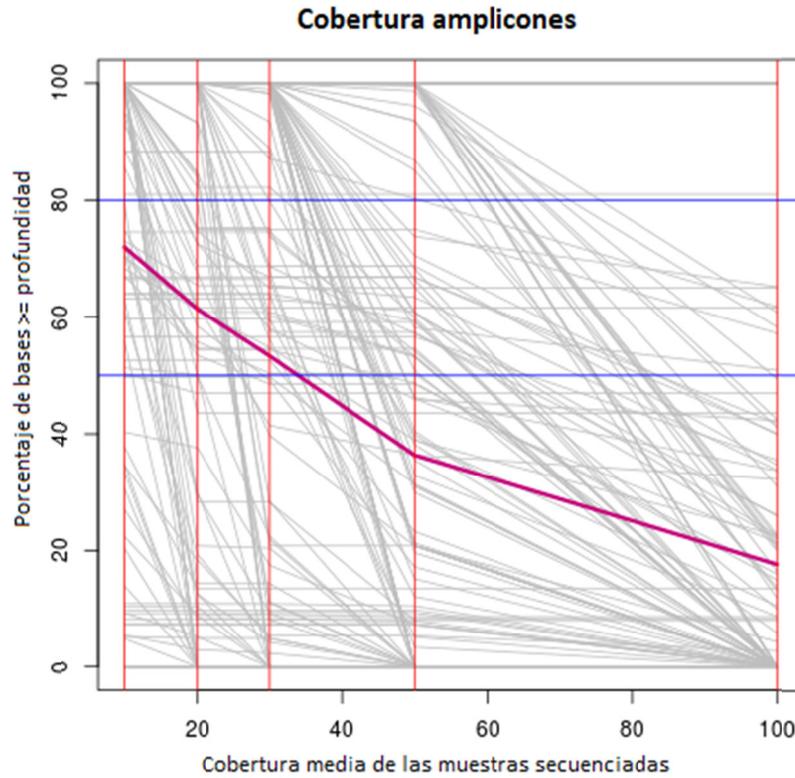


Figura 19: Gráfica que representa la cobertura media de todas las muestras de la población y el porcentaje de bases cubiertas a esa profundidad de cobertura. La línea rosada indica la media de la relación profundidad/amplitud de cobertura en todos los amplicones.

Teniendo en cuenta la media total se puede ver, por ejemplo, que más del 60% de las regiones de interés estarían cubiertas a una profundidad de cobertura de 20 lecturas.

En la tabla 9 se recogen los valores de cobertura media por gen, así como el número de exones que poseen según sus transcritos canónicos, y el número de amplicones que ha sido necesario diseñar para cubrir toda la secuencia codificante y las regiones intrónicas adyacentes.

Tabla 9: Resultados de cobertura media por gen, número de exones y número de amplicones diseñados.

GEN	Nº EXONES	Nº AMPLICONES	COBERTURA MEDIA
<i>ACAD10</i>	21	29	35.28
<i>ACPL2</i>	6	9	31.20
<i>ACSL3</i>	17	18	115.18
<i>ACSM3</i>	14	30	25.13
<i>AFMID</i>	11	11	996.87
<i>ARHGEF38</i>	14	20	136.35
<i>CACNG5</i>	6	7	1227.91
<i>CCNDBP1</i>	11	12	77.11
<i>CCT8L2</i>	1	10	110.82
<i>CIAPIN1</i>	9	9	199.27
<i>CYP1A1</i>	7	12	32.74
<i>DDO</i>	5	9	201.93
<i>EBLN2</i>	1	9	46.25
<i>GCKR</i>	19	19	18.40
<i>HABP2</i>	13	23	65.09
<i>OLFML1</i>	3	9	25.06
<i>SLC12A4</i>	24	31	83.74
<i>SLC15A3</i>	8	13	264.44
<i>SREBF2</i>	19	38	76.04

El gen con mayor cobertura media ha sido *CACNG5* con un valor de 1227.91X, seguido del gen *AFMID* con un valor de 996.87X. Por otro lado, el gen *GCKR* ha obtenido el valor más bajo de cobertura 18.40X.

La diferencia de valores en la relación profundidad/amplitud de cobertura entre unos genes y otros, se debe tanto a la diferente eficiencia de amplificación de los diversos amplicones, como a la limitación en la cantidad de lecturas disponibles, y a la diversidad de la calidad del ADN utilizado. Todo esto ha determinado que unas muestras se hayan amplificado mejor que otras y que la distribución de lecturas sea dispareja entre ellas.

1.5.- Selección de las variantes genéticas

Durante el proceso bioinformático se establecieron una serie de filtros para priorizar las variantes finales. Estos fueron:

- Uso de un valor mínimo de cobertura de 10 lecturas. Es decir, fueron necesarias 10 lecturas mínimo para considerar una variante candidata.
- Uso de un valor mínimo de calidad de genotipado de 20 para poder considerar una variante apta y no un error de secuenciación.
- Genotipado de al menos el 60% de las muestras en ambos grupos.
- Selección del transcrito canónico de cada gen.
- Clasificación según su grado de impacto por el valor de *VEP_Impact* que utiliza la anotación utilizada, en este caso, la *Variant Effect Predictor (VEP)*, de Ensembl (145). Cuatro categorías: alto, moderado, modificador y bajo impacto.
- Distinción de las variantes específicas y únicas de los diferentes grupos, es decir, que solo ha aparecido en dichos grupos y de aquellas comunes a ambos grupos comparados.

En cuanto a los resultados estadísticos de las variantes encontradas, se ha optado por mostrar todas las variantes con un p-valor menor al 0.05 mediante el Test de asociación Fisher para estudios de caso/control, así como el test genético con más probabilidad de asociación. También se ha decidido mostrar aquellas variantes con valores de OR (*Odds Ratio*) menores a 0.5 por indicar un cierto efecto protector y aquellas variantes con valores de OR mayores a 2 como ejemplo de alteraciones con efecto de riesgo que, aunque no hayan tenido un p-valor significativo pueden ser interesantes en estudios futuros.

En los resultados hay que considerar que no se ha incluido una corrección por múltiples análisis dado que, estos estudios ya son la validación de resultados previos con los mismos genes. Si se quisiera considerar el valor p corregido se debería considerar un valor umbral de $9.28E-5$ para considerar que un polimorfismo es estadísticamente significativo.

1.6.- Descripción de resultados por gen

1.6.1.- GEN ACAD10

El gen *ACAD10* (HGNC:21597) codifica la enzima (Acil-CoA deshidrogenasa 10) perteneciente a la familia de las Acil-CoA deshidrogenasas, las cuales participan en el proceso de α y β -oxidación de ácidos grasos en la mitocondria. Este gen posee 21 exones y 20 intrones (154). Se localiza mayoritariamente en tejido muscular y renal. Una desregulación de la oxidación de ácidos grasos y el aumento de los niveles plásmaticos de acilcarnitinas se ha relacionado con el desarrollo de resistencia a la insulina periférica, con

elevados niveles de insulina, tolerancia anormal a la glucosa y ganancia de peso, procesos que favorecen el desarrollo de DM2 (155).

1.6.1.1.- Proceso de secuenciación

Para amplificar este gen se diseñaron 29 fragmentos que cubrían los 21 exones que lo componen y las regiones intrónicas adyacentes. En la figura 22 se puede observar la cobertura de todas las muestras secuenciadas cuyo patrón coincide con las regiones exónicas del transcrito utilizado (ENST00000013698). La media de cobertura para este gen fue de 35.28X.

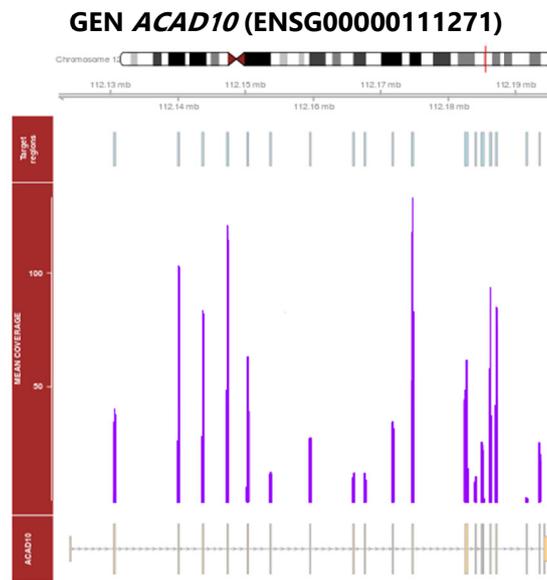


Figura 20: Distribución de la cobertura en la población secuenciada para el gen *ACAD10*.

1.6.1.2.- Variantes identificadas

Para el gen *ACAD10* se han encontrado las variantes resumidas en la tabla 10 clasificadas según su impacto. Se han encontrado un total de 71 variantes una vez aplicados los filtros previamente descritos.

Tabla 10: Clasificación por impacto de las variantes del gen *ACAD10*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL
			casos	controles	comunes	
ACAD10	ALTO	V. pauta de lectura	0	1	0	1
		V. donadora de <i>splicing</i>	1	0	0	1
	MODERADO	V. <i>missense</i>	11	18	5	34
		V. <i>missense</i> y <i>splicing</i>	0	2	0	2
	MODIFICADOR	V. intrónica	7	10	2	19
	BAJO	V. <i>splicing</i> e intrónica	0	2	1	3
		V. sinónima	2	9	0	11
				21	42	8
						TOTAL

En el grupo de casos se han visto 21 variantes y 42 variantes únicas en el grupo de controles. Las 8 variantes restantes, se han encontrado en ambos grupos en proporciones similares. Casi un 78% de las variantes corresponden a un impacto intermedio y sobretodo son mutaciones *missense* o mutaciones con cambio de sentido.

En la tabla 11 se recogen las variantes del estudio estadístico realizado.

Tabla 11: Características y descripción de las variantes encontradas para el gen *ACAD10*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
12:112130708	A/G	rs7132509	V. <i>splicing</i> e intrónica	g.112130708A>G	2.86	ALLELIC	4.025(1.258-12.880)	0.023
12:112186202	T/G		V. <i>missense</i>	p.Val856Gly	23.00	TREND	3.030(0.274-33.460)	0.341
12:112186347	C/T		V. intrónica	g.112186347C>T	7.65	TREND	0.376(0.108-1.313)	0.109

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
12:112130708	0.0039	0.0127	0.0032	386	6	2	394	622	4	0	626
12:112186202		0.0020	0.0007	508	2	0	510	761	1	0	762
12:112186347		0.0042	0.0112	350	3	0	353	609	14	0	623

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Solamente el polimorfismo rs7132509 ha obtenido un resultado significativo y una OR superior a 4.00. Su localización intrónica próximo al exón (+7bp) hace necesario analizar su efecto en el procesado del ARN de *ACAD10*. En un estudio hecho con indios Pima se ha visto que este polimorfismo junto con otros, pueden estar asociados con el riesgo de padecer DM2 (156).

1.6.2.- GEN ACPL2/PXYLP1

El producto del gen *ACPL2* o recientemente conocido como *PXYLP1* (HGNC:26303) codifica una fosfatasa de xilosa en su extremo 2-O-fosfato y está implicada en la regulación de la síntesis de glucosaminglucanos (GAGs) como el condrotin-sulfato y el heparan-sulfato. La desfosforilación de este residuo de xilosa es necesaria para que se prosiga con normalidad la elongación de las cadenas de GAGs. (157).

1.6.2.1.- Proceso de secuenciación

El gen *ACPL2/PXYLP1* está compuesto de 6 exones según el transcrito ENST00000286353. Se diseñaron 9 amplicones para completar su secuenciación. La cobertura media de la población fue de 31.20X. En la figura 23 se puede ver su representación.

GEN *ACPL2*/*PXYLP1* (ENSG00000155893)

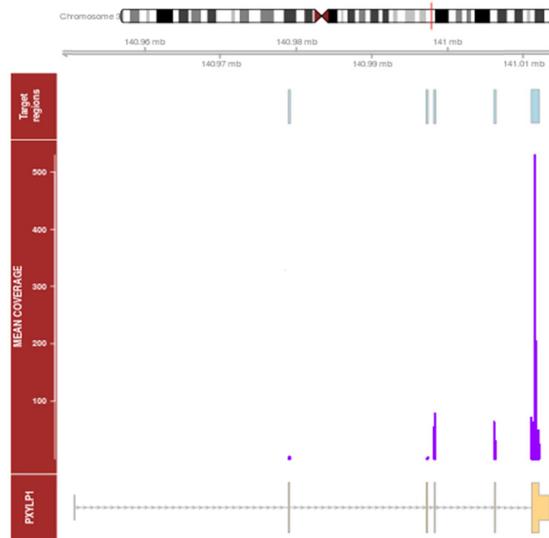


Figura 21: Distribución de la cobertura en la población secuenciada para el gen *ACPL2*.

1.6.2.2.- Variantes identificadas

En el análisis del gen *ACPL2* se han encontrado las variantes de la tabla 12. En total han sido 4 variantes sinónimas.

Tabla 12: Clasificación por impacto de las variantes del gen *ACPL2*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL
			casos	controles	comunes	
<i>ACPL2</i>	BAJO	V. sinónima	2	1	1	4

El análisis de asociación mediante el Test de Fisher, muestra que ningún polimorfismo tiene una asociación significativa. Si bien, en la tabla 13 se muestra el único que tiene un valor de OR mayor a 2.

Tabla 13: Características y descripción de las variantes encontradas para el gen *ACPL2*.

A) Información de las variantes y resultados estadísticos.

cr.posición	Alelos	SNP	Consecuencia	Cambio	TEST	OR(IC 95%)	P
3:141011642	C/T	rs145988315	V. sinónima	p.Leu346=	TREND	2.579 (0.233-28.480)	0.422

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de la variante en la población estudiada.

cr.posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
3:141011642	2.98E-04	0.0018	0.0007	561	2	0	563	714	1	0	715

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

No se han encontrado datos bibliográficos que así lo demuestren y se deberían de hacer otros estudios para poder llegar a alguna conclusión más certera.

1.6.3.- GEN ACSL3

La enzima codificada por el gen *ACSL3* (HGNC:3570) forma parte de la familia de las sintetasas de acyl-CoA de cadena larga. Intracelularmente, se localiza en la membrana del retículo endoplasmático celular (RE) y forma parte en la creación de los almacenes de lípidos o gotas lipídicas. Además, favorece la entrada de lípidos extracelulares (158,159). También participa en la secreción de las lipoproteínas de muy baja densidad (VLDL) debido a su influencia en la síntesis de lecitina, el lípido más abundante en la membrana de las VLDL, a través del proceso de síntesis del diacilglicerol (DAG) (160). Otros estudios han mostrado que actúa favoreciendo la actividad de diversos factores de transcripción asociados a la síntesis lipídica, como puede ser el receptor de peroxisoma-proliferador-activado-gamma (PPAR γ) o el elemento de unión a proteína regulador del estero

(SREBP-1c) (161). Un estudio reciente demuestra su presencia en los gránulos secretores de insulina tanto de células β humanas como en islotes pancreáticos murinos aunque su papel no está claro todavía (162).

1.6.3.1.- Proceso de secuenciación

Para la amplificación de este gen, el cual posee 17 exones, se diseñaron 18 amplicones para cubrir toda la secuencia del transcrito canónico elegido ENST00000357430. La cobertura media de este gen fue de 115.18x, siendo un buen resultado global. En la figura 24 se puede ver la distribución de la cobertura en la población.

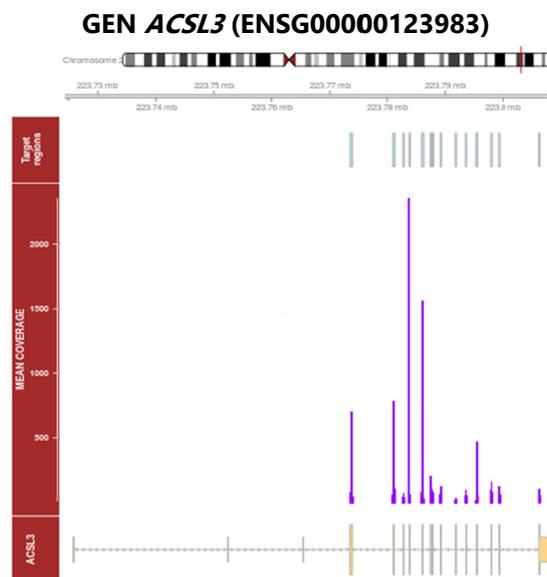


Figura 22: Representación gráfica del gen *ACSL3* y su cobertura en la población.

1.6.3.2.- Variantes identificadas

Se han identificado 26 variantes en total en el gen *ACSL3*, que se pueden ver en la tabla 14. Se han clasificado según su impacto.

Tabla 14: Clasificación por impacto de las variantes del gen *ACSL3*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL
			casos	controles	comunes	
<i>ACSL3</i>	ALTO	V. ganancia de <i>STOP</i>	0	1	0	1
	MODERADO	V. <i>missense</i>	4	7	4	15
	MODIFICADOR	V. intrónica	3	1	2	6
	BAJO	V. sinónima	1	3	0	4
			8	12	6	26

Se han visto 8 variantes, en total, en el grupo de casos y 12 variantes únicas en el grupo de controles. El resto pertenece a ambos grupos siendo en más de un 50% de impacto intermedio, sobretodo, variantes *missense*.

En la tabla 15 se pueden ver los resultados del análisis estadístico y características de los polimorfismos.

Tabla 15: Características y descripción de las variantes encontradas para el gen *ACSL3*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
2:223786084	A/G	rs749148769	V. <i>missense</i>	p.Met298Val	22.30	ALLELIC	0.278(0.180-0.430)	2.32E-10
2:223787470	T/C		V. <i>missense</i>	p.Leu334Pro	31.00	GENO/DOM	0.300(0.131-0.684)	2.21E-03
2:223797861	C/T		V. intrónica	g.223797861C>T	9.26	TREND	NA	1.12E-03

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(casos)	MAF(controles)	Casos				Controles			
			0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
2:223786084	0.0319	0.1050	383	24	1	408	444	97	10	555
2:223787470	0.0084	0.0278	411	7	0	418	527	31	0	558
2:223797861	0.0081	0	487	8	0	495	643	0	0	643

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Si bien estas tres variantes tienen un p-valor significativo en las muestras estudiadas, el rs749148769 tiene un valor de p mucho mayor

que supone una importante asociación con DM2, con un efecto protector importante (OR=0.27). No se han encontrado referencias en la bibliografía. Viendo su valor de OR se podría decir que estas variantes tendrían un cierto factor de protección ya que podrían favorecer, en este caso una disminución de la actividad del gen *ACSL3* y una mejor regulación metabólica. En algunos estudios se le ha atribuido un papel de regulador del mecanismo de lipogénesis de *ново* hepática. Un aumento relevante de la síntesis lipídica puede contribuir al desarrollo de esteatosis hepática, resistencia a la insulina y dislipemias (163). Por esto, se le puede considerar una potencial diana terapéutica en algunas enfermedades metabólicas. En cualquier caso, nuevos estudios a nivel funcional y poblacional deber realizarse para conocer el impacto de estas variantes.

1.6.4.- GEN ACSM3

El gen *ACSM3* (HGNC:10522) codifica una enzima perteneciente a la familia de las sintetasas de Acil-CoA de cadena media. Se expresa en los riñones (164) e hígado. Hay estudios en carcinoma hepatocelular que demuestran que el aumento característico de los niveles de lípidos en estos tumores tiene una relación directa con la disminución de la expresión del gen *ACSM3*, colocándolo como una posible diana terapéutica. También se ha visto que los factores de transcripción como PPAR γ y el Factor de crecimiento transformante-1 beta 1(TGFB1) actuarían de forma negativa sobre el gen, disminuyendo su actividad (165). La desregulación del metabolismo de ácidos grasos, está relacionada con alteraciones en los niveles de glucosa y aumenta el riesgo de padecer obesidad y DM2 (166,167).

1.6.4.1.- Proceso de secuenciación

El gen *ACSM3* posee 14 exones y han sido necesarios diseñar 30 amplicones para cubrir toda la secuencia del transcrito ENST00000289416. La cobertura media de este proceso fue de 25.13X. En la figura 25 se puede la cobertura media de todas las muestras y su distribución a lo largo del gen.

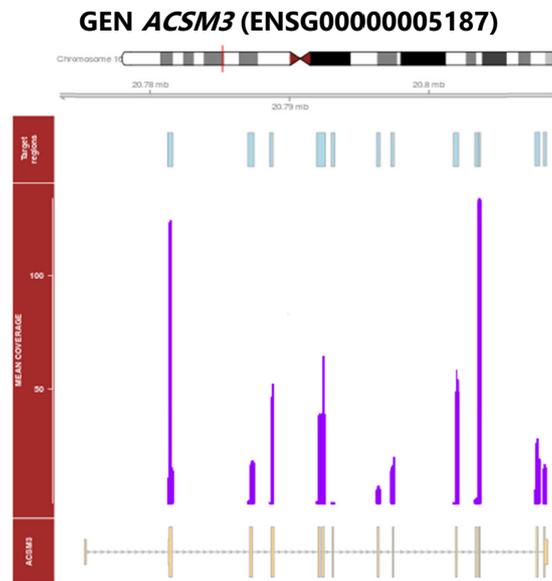


Figura 23: Distribución de la cobertura de la población para el gen *ACSM3*.

1.6.4.2.- Variantes identificadas

Para el gen *ACSM3* se han encontrado 13 variantes, recogidas en la tabla 16.

Tabla 16: Clasificación por impacto de las variantes del gen *ACSM3*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL
			casos	controles	comunes	
ACSM3	ALTO	V. aceptora de <i>splicing</i>	1	0	0	1
		V. de ganancia de <i>STOP</i>	1	1	1	3
	MODERADO	V. de <i>missense</i>	1	3	1	5
		V. de <i>missense</i> e <i>splicing</i>	1	0	0	1
	BAJO	V. sinónima	1	2	0	3
			5	6	2	13
						TOTAL

Para el grupo de casos se han visto 5 variantes únicas, mientras que para el grupo de controles han aparecido 6 variantes.

De entre ellas, no se han encontrado variantes estadísticamente significativas. En la tabla 17 se puede ver las características de las 13 variantes encontradas.

Tabla 17: Características y descripción de las variantes encontradas para el gen *ACSM3*.

A) Información de las variantes.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD
16:20781381	A/G		V. <i>missense</i>	p.Met9Val	3.26
16:20788810	C/T	rs147662918	V. sinónima	p.Asp182=	9.72
16:20792328	T/C		V. <i>missense</i>	p.Met272Thr	22.70
16:20792340	C/A		V. de <i>STOP</i>	p.Ser276Ter	40.00
16:20792387	T/C		V. <i>missense</i>	p.Trp292Arg	26.20
16:20792404	T/C		V. sinónima	p.Cys297=	16.87
16:20792417	C/T	rs117563656	V. <i>missense</i>	p.His302Tyr	15.11
16:20792419	T/C		V. sinónima	p.His302=	16.78
16:20792435	C/A	rs7196188	V. <i>missense</i>	p.Pro308Thr	16.72
16:20801981	C/T	rs757648352	V. de <i>STOP</i>	p.Arg433Ter	38.00
16:20802007	C/G	rs34381224	V. de <i>STOP</i>	p.Tyr441Ter	35.00
16:20802008	G/A	rs146721389	V. <i>missense</i> y <i>splicing</i>	p.Val442Ile	11.10
16:20803556	A/T		V. <i>splicing</i>	g.20803556A>T	28.40

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
16:20781381		0	0.0007	491	0	0	491	740	1	0	741
16:20788810	1.80E-05	0	0.0007	456	0	0	456	715	1	0	716
16:20792328		0.0013	0	391	1	0	392	653	0	0	653
16:20792340		0.0026	0	387	2	0	389	663	0	0	663
16:20792387		0	0.0008	412	0	0	412	665	1	0	666
16:20792404		0.0012	0	415	1	0	416	667	0	0	667
16:20792417	2.14E-03	0.0012	0.0008	407	1	0	408	661	1	0	662
16:20792419		0	0.0008	409	0	0	409	662	1	0	663
16:20792435	2.25E-04	0	0.0015	407	0	0	407	660	2	0	662
16:20801981	8.96E-06	0	0.0007	438	0	0	438	701	1	0	702
16:20802007	9.75E-03	0.0126	0.0128	426	11	0	437	684	18	0	702
16:20802008	8.97E-06	0.0011	0	436	1	0	437	703	0	0	703
16:20803556		0.0010	0	502	1	0	503	751	0	0	751

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

No se han encontrado polimorfismos con asociaciones significativas, posiblemente porque todos los identificados tienen un MAF muy bajo. Al ser un gen relacionado también con el metabolismo de lípidos, sería interesante realizar estudios con un número mayor de individuos y analizar parámetros lipídicos para conocer su posible efecto en la población y su relación con la DM2. No hemos encontrado bibliografía en la que se observe una asociación de este gen con DM2.

1.6.5.- GEN AFMID

El gen *AFMID* (HGNC:20910) codifica para la enzima arilformidasa, o quirunenina formidasa. Es la segunda enzima en la vía de degradación del aminoácido triptófano para generar los cofactores NAD⁺ y NAD(P)H. Se expresa mayoritariamente en el hígado y en riñones. La desregulación de esta vía puede generar neuro e inmunotoxicidad, así como degradación gradual del tejido renal a nivel glomerular, afectando a su funcionalidad (168,169). Se ha sugerido en un estudio con ratones que la alteración de esta ruta y la acumulación de derivados de la quinurenina puede favorecer la aparición de

resistencia a la insulina, así como una disminución de la tolerancia a la glucosa (170). Estudios más recientes han visto que *AFMID* genera 4 isoformas diferentes de la enzima, la expresión de unas u otras añade un nuevo mecanismo de regulación al metabolismo del triptófano produciendo una alteración en los niveles de NAD⁺, afectando a los mecanismos de reparación del ADN, sobretodo en hepatocitos, favoreciendo los primeros estadios del carcinoma hepatocelular (CHC) (171).

1.6.5.1.- Proceso de secuenciación

Para la secuenciación del gen *AFMID*, el cual tiene 11 exones se han diseñado 12 amplicones para cubrir los exones y zonas de *splicing* adyacentes del transcrito ENST00000327898. En la figura 26 se puede ver la cobertura de la población distribuida a lo largo del gen. La cobertura media fue bastante alta, de 996.87X.

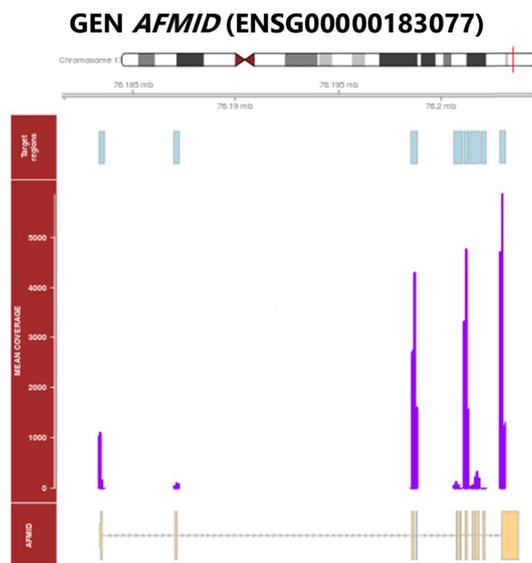


Figura 24: Distribución cobertura para el gen *AFMID*.

1.6.5.2.- Variantes identificadas

En el análisis del gen *AFMID* se han encontrado un total de 45 variantes. Se clasifican según su impacto en la tabla 18.

Tabla 18: Características de las variantes seleccionadas del gen *AFMID* y su distribución en la población.

GEN	IMPACTO	CONSECUENCIA	N° DE VARIANTES			TOTAL	
			casos	controles	comunes		
AFMID	ALTO	V. aceptora de <i>splicing</i> codificante	0	1	0	1	
	MODERADO	V. <i>missense</i>	2	6	2	10	
		V. en 3' UTR	6	2	0	8	
	MODIFICADOR	V. intrónica	7	4	9	20	
		V. <i>upstream</i>	0	2	0	2	
		V. <i>splicing</i> e intrónica	1	0	0	1	
	BAJO	V. <i>splicing</i> e intrónica	1	0	0	1	
		V. sinónima	2	1	0	3	
			18	16	11	45	TOTAL

Se han visto en el grupo de casos 18 variantes únicas, mientras que en el grupo de controles 16 variantes. Los 11 polimorfismos restantes se han encontrado en ambos grupos. Más de un 88% de ellas se definen con un impacto intermedio, sobretodo son variantes de tipo *missense* e intrónicas.

Después de realizar en análisis estadístico mediante el Test de Fisher, se han encontrado dos variantes significativas. En la tabla 19 se observan sus características.

Tabla 19: Características y descripción de las variantes encontradas para el gen *AFMID*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
17:76187160	C/T	rs17640750	V. intrónica	g.76187160C>T	1.53	ALLELIC	0.269(0.032-2.276)	0.276
17:76201330	T/C		V. intrónica	g.76201330T>C	1.32	DOM	6.291(0.034-2.500)	1.463E-18
17:76203128	A/G	rs547095488	V. 3' UTR	g.76203128A>G	0.86	TREND	NA	3.15E-03

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
17:76187160	2.54E-01	0.0024	0.0090	420	2	0	421	606	5	3	610
17:76201330		0.0826	0.0148	467	88	2	557	788	22	1	811
17:76203128	7.46E-04	0.0052	0	576	6	0	582	829	0	0	829

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Para conocer el efecto (aumento o disminución de actividad final del gen o de la proteína) de las variantes con valores significativos se deberían hacer estudios funcionales. Si bien, los datos previos en relación a la pérdida de actividad y RI podrían indicar que pueden suponer una pérdida de función, especialmente la variante en 17:76201330. Por otro lado, podrían estar en desequilibrio de ligamiento con otras variantes funcionales que no se han detectado en el presente estudio o que están en las regiones no analizadas. En cualquier caso, estos datos indican una importante asociación con el riesgo de desarrollar DM2 que serán validados en breve mediante el análisis de estudios casos/control y poblacionales más amplios.

1.6.6.- GEN ARHGEF38

La proteína codificada por el gen *ARHGEF38* (HGNC:25968) o Factor 38 de intercambio de nucleótidos de guanina asociado a proteínas de la familia de GTPasas tipo Rho, estimula la disociación del GDP y favoreciendo la unión del GTP activando la GTPasa. Estas proteínas de intercambio también son conocidas como *GEF* (Guanine Exchange Factor) (172).

Las proteínas Rho, así como otros tipos de GTPasas, están implicadas en una gran variedad de procesos, siendo normalmente de las primeras proteínas en casi todas las vías de señalización celular más importantes (173).

1.6.6.1.- Proceso de secuenciación

El gen *ARHGEF38*, según el transcrito ENST0000042047.2 tiene 14 exones, los cuales han sido diseñados en 20 amplicones para cubrir todas las regiones exónicas y adyacentes. La secuenciación de este gen en la población ha resultado con una cobertura media de 136.35X. En este caso, si se observa la figura 27, el dato de cobertura media está un poco desviado, ya que ha habido un amplicón que se amplifica mucho mejor que el resto, que tienen unos datos de cobertura parecida.

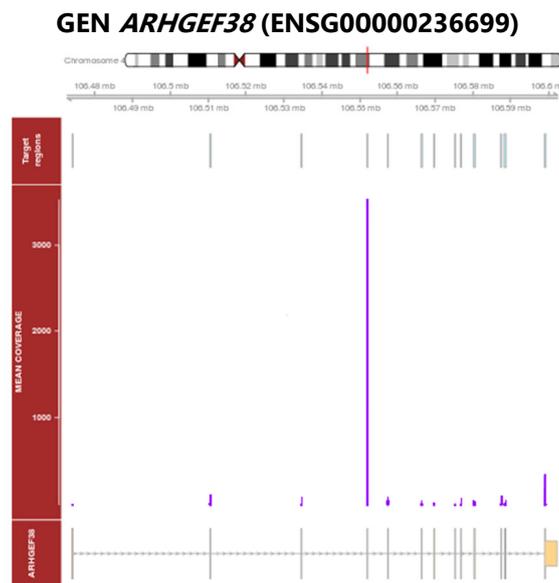


Figura 25: Distribución de la cobertura media en la población del gen *ARHGEF38*.

1.6.6.2.- Variantes identificadas

En el gen *ARHGEF38* se han encontrado 10 variantes totales. En la tabla 20 están resumidas según su impacto.

Tabla 20: Clasificación por impacto de las variantes del gen *ARHGEF38*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
<i>ARHGEF38</i>	MODERADO	V. <i>missense</i>	3	4	0	7	
	MODIFICADOR	V. intrónica	0	0	1	1	
	BAJO	V. sinónima	0	2	0	2	
			3	6	1	10	TOTAL

En el grupo de casos se han encontrado 3 variantes únicas y en el de controles 6 variantes. Solo hay una variante intrónica común a ambos grupos.

Teniendo en cuenta el análisis de asociación estadístico usando el Test de Fisher, se ha encontrado una variante con valor de OR menor a 0.5. En la tabla 21 se pueden ver sus características.

Tabla 21: Características y descripción de la variante encontradas para el gen *ARHGEF38*.

A) Información de la variante y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
4:106552022	G/T	rs189615529	V. intrónica	g.106552022G>T	2.32	TREND	0.301(0.035-2.575)	0.244

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de la variante en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
4:106552022	7.97E-05	0.0009	0.0030	557	1	0	558	823	5	0	828

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

En algunos análisis a gran escala se han identificado otros polimorfismos, como el rs7684945 (174) no analizados en el presente estudio, que poseen cierta relación con parámetros relacionados con la DM2 por lo que un análisis más exhaustivo de este gen en otras

poblaciones es necesario para conocer su implicación en el desarrollo de la enfermedad.

1.6.7.- GEN CACNG5

El gen *CACNG5* (HGNC:1409) codifica la subunidad auxiliar *gamma 5* del canal de Ca^{2+} dependiente de voltaje. Por sus diferencias con otras proteínas de la familia, también se le considera de la clase II de las Proteínas transmembrana reguladoras del receptor AMPA (TARPs). Están implicadas en tráfico y plasticidad neuronal de los receptores AMPA. Los cuales son receptores ionotrópicos de glutamato y son los principales receptores del sistema nervioso (175).

1.6.7.1.- Proceso de secuenciación

Para el gen *CACNG5* se diseñaron 7 amplicones para cubrir los 6 exones y sus regiones contiguas según el transcrito ENST00000533854. La cobertura media del proceso de secuenciación fue de 1227.91X, siendo un buen resultado. En la figura 28 se puede observar la cobertura de la población según los amplicones diseñados.

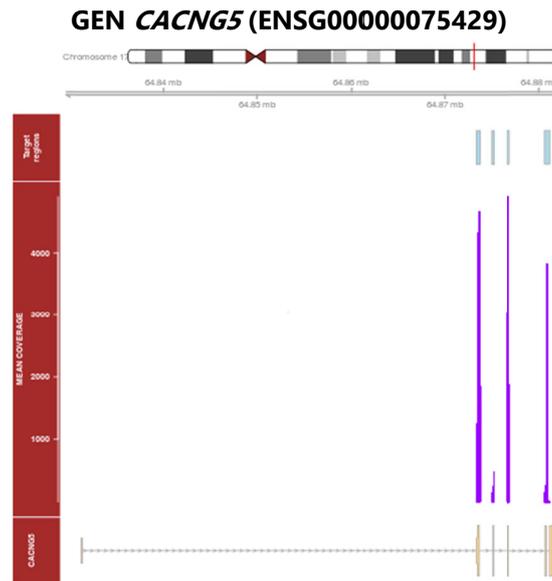


Figura 26: Distribución de la cobertura en la población secuenciada del gen *CACNG5*.

1.6.7.2.- Variantes identificadas

En el estudio del gen *CACNG5* han aparecido 22 variantes totales. Se pueden ver en la tabla 23.

Tabla 22: Clasificación por impacto de las variantes del gen *CACNG5*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
<i>CACNG5</i>	MODERADO	V. <i>missense</i>	2	2	3	7	
	MODIFICADOR	V. 5' UTR	2	0	2	4	
		V. intrónica	2	0	2	4	
	BAJO	V. sinónima	1	1	4	6	
			7	3	11	21	TOTAL

Se han encontrado 7 variantes únicas en casos y 3 variantes únicas en controles. El resto, están distribuidas en ambos grupos. Más de un 77% tiene un impacto intermedio.

Teniendo en cuenta el análisis de asociación estadístico, realizado con el Test de Fisher, no se han encontrado variantes significativas. En la tabla 23 están reflejadas sus características.

Tabla 23: Características y descripción de las variantes encontradas para el gen *CACNG5*.

A) Información de las variantes y resultados estadísticos.

cr.posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
17:64873387	C/T	rs186268864	V. 5'UTR	g.64873387C>T	4.12	TREND	0.224(0.029-1.753)	0.150
17:64873528	C/T	rs769615801	V. sinónima	p.Ile26=	6.11	TREND	2.966(0.269-32.750)	0.351
17:64876773	C/T	rs149159754	V. <i>missense</i>	p.Thr128Met	23.90	TREND	3.809(0.345-42.070)	0.240

Cr: cromosoma. CADD: Combined Annotation Dependent Depletion value. SNP: Single Nucleotide Polymorphism. OR: Odds Ratio. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr.posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
17:64873387		0.001	0.006	374	1	0	375	817	8	1	826
17:64873528	8.13E-06	0.002	0.001	566	2	0	568	827	1	0	828
17:64876773	9.31E-04	0.002	0.001	442	2	0	444	828	1	0	829

Cr: cromosoma. MAF: Minority Allele Frequency. EUR: Europa.

Nuestros resultados no muestran una asociación significativa y no hay datos bibliográficos descritos sobre la asociación de este gen con DM2.

1.6.8.- GEN CCNDBP1

El gen *CCNDBP1* (HGNC:1587) codifica para la proteína de unión a la ciclina D1. También se conoce como GCIP o de unión a ciclina D1 (CD1) y GRAP2. La primera es una proteína que actúa en las rutas de proliferación y mitogénesis celular y la segunda posee un papel de conector central en la señalización y activación inmune celular (176). GCIP o CCNDBP1 tiene localización nuclear mayoritariamente para actuar en los procesos de división celular, es considerado un gen

supresor de tumores ya que actúa inhibiendo a la CD1, evitando así la fosforilación de la proteína retinoblastoma (Rb), otro gen supresor de tumores, y paralizando el proceso de proliferación celular. Se han estudiado varias proteínas que actúan sobre GCIP regulando su activación y la progresión celular, como por ejemplo la proteína Rad, la cual, traslocaría GCIP del núcleo al citoplasma evitando su acción sobre CD1 promoviendo así la senescencia celular (177). El oncogén miR-9 también actúa inhibiendo a GCIP en osteosarcomas (178).

1.6.8.1.- Proceso de secuenciación

El gen *CCNDBP1* (HGNC:1587) tiene 11 exones según su transcrito ENST00000300213. Fueron necesarios 12 amplicones para cubrir toda su secuencia. La cobertura media de este gen fue de 77.11X. En la figura 29 se puede ver su distribución en toda la población secuenciada.

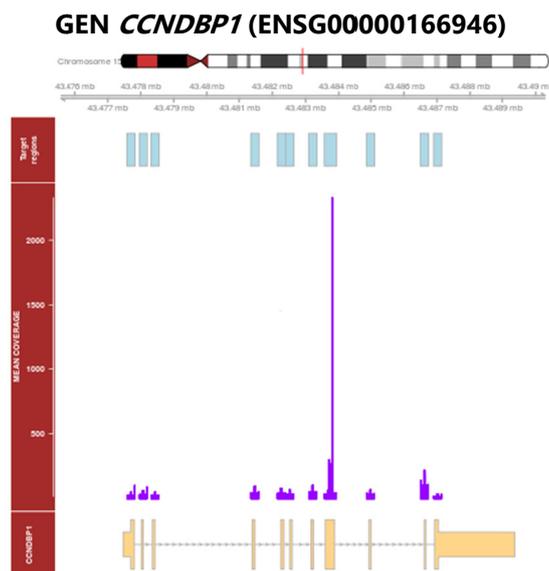


Figura 27: Distribución de la cobertura en la población del gen *CCNDBP1*.

1.6.8.2.- Variantes encontradas

Tras el análisis del gen *CCNDBP1* se han identificado un total de 22 variantes. Se recogen clasificadas por su grado de impacto en la tabla 24.

Tabla 24: Clasificación por impacto de las variantes del gen *CCNDBP1*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL
			casos	controles	comunes	
<i>CCNDBP1</i>	ALTO	V. ganancia de <i>STOP</i> y <i>splicing</i>	0	1	0	1
	MODERADO	V. <i>missense</i>	2	6	1	9
	MODIFICADOR	V. intrónica	1	6	1	8
	BAJO	V. <i>splicing</i> e intrónica	0	1	0	1
		V. sinónima	0	2	1	3
			3	16	3	22
						TOTAL

En el grupo de casos se han encontrado 3 variantes únicas mientras que en el grupo de controles 16 variantes. Son más comunes, las variantes *missense* y las intrónicas.

Tras realizar el Test de Fisher, solo una variante ha resultado significativa. En la tabla 25 están sus características.

Tabla 25: Características y descripción de las variantes encontradas para el gen *CCNDBP1*.

A) Información de las variantes y resultados estadísticos.

cr.posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
15:43482282	G/A	rs139518864	V. <i>missense</i>	p.Ala121Thr	10.07	TREND	2.581(0.234-28.520)	0.4216
15:43483811	G/C	COSM3501355	V. sinónima	p.Gly266=	5.16	ALLELIC	0.256(0.154-0.425)	6.48E-09

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr.posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
15:43482282	2.06E-04	0.0025	0.0010	392	2	0	394	499	1	0	500
15:43483811		0.0244	0.0863	390	18	1	409	417	76	5	498

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

El polimorfismo encontrado COSM3501355 codifica para una variante sinónima, lo cual no significa que no pueda ser funcional; si bien, sería necesario determinar el efecto que pueda tener por su localización dentro de la proteína o si podría afectar por ejemplo a sitios de *splicing* cercanos. No hemos encontrado estudios que indiquen que variantes del gen *CCNDBP1* se asocien con DM2 u otras enfermedades relacionadas, pero dado su papel en la regulación de la ciclina D1 dentro del ciclo celular, sí que podría tener cierta relación, ya que esta ciclina sí que se ha asociado en ciertos estudios, con el metabolismo glucídico. Se ha visto que su sobre-expresión en células β -pancreáticas induce a un aumento de este tipo celular, pero sin llegar a formar un cáncer ni influir en la homeostasis de la glucosa. Se la podría considerar una posible diana terapéutica para aumentar las células beta y así el aumento de la producción de insulina (179). También se ha estudiado su papel en la regulación de la gluconeogénesis hepática. Su presencia junto con la ciclina CDK4 en condiciones normales e inducidas por insulina, inhiben a PGC-1 α por acetilación y aumentan los niveles de GCN5 favoreciendo la inhibición del proceso gluconeogénico. Su disminución de concentración o desregulación en condiciones de DM2 puede favorecer la hiperglucemia y la resistencia a la insulina (180).

1.6.9.- GEN CCT8L2

El gen *CCT8L2* (HGNC:15553) codifica la subunidad 8 tipo 2 de la familia de carabinas moleculares TCP-1. Filogenéticamente se ha estudiado su aparición, los genes *CCT8L* (1 y 2) tienen un origen reciente, debido a una tardía duplicación en el linaje de *CCT8*, al principio de la aparición de los mamíferos. Conservan la típica

estructura de carabina molecular pero no forman el complejo oligomérico típico de las proteínas CCT, e interactúan con sus sustratos en regiones catalíticas diferentes, aun así, se consideran proteínas funcionales del grupo II de carabinas moleculares eucariotas (181).

1.6.9.1.- Proceso de secuenciación

Para secuenciar el gen *CCT8L2* el cual según su transcrito ENST00000359963, está compuesto por un exón de 2034 pb, para el cual han sido necesarios diseñar 10 amplicones para cubrir la totalidad de su secuencia y regiones adyacentes. La cobertura media de este proceso ha sido de 110.82X. En la figura 30 se puede ver la distribución de la misma de toda la población secuenciada.

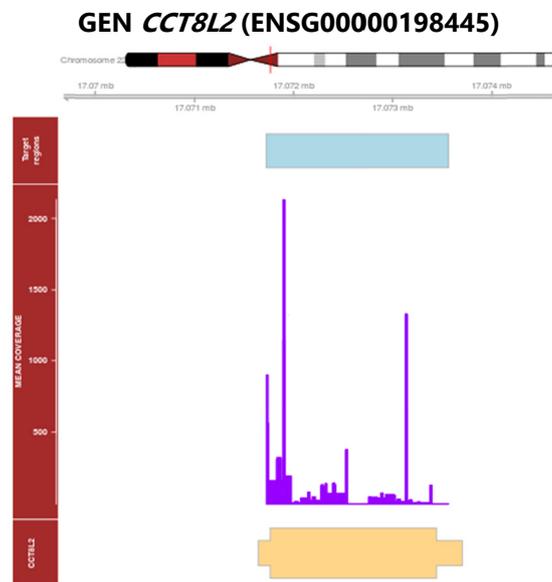


Figura 28: Distribución de la cobertura en la población secuenciada del gen *CCT8L2*.

1.6.9.2.- Variantes identificadas

No se han identificado variantes que superen los filtros de genotipado en al menos un 60% de las muestras.

1.6.10.- GEN CIAPIN1

El gen *CIAPIN1* (HGNC:28050) codifica una proteína antiapoptótica inducida por citoquinas también conocida como AMORISINA. Es parte de la maquinaria de ensamblaje proteico FE-S citosólica (*CIA*) necesaria para la maduración extramitocondrial de las proteínas FE-S (182). No tiene ninguna homología con otras proteínas reguladoras de la apoptosis como la familia BCL-2, la familia de las caspasas u otras moléculas implicadas en la transducción de señales. Su expresión es totalmente dependiente de la estimulación de los factores de crecimiento de los tejidos donde se expresa. Se ha visto que participa en los procesos de proliferación y tumorigénesis de diversos tipos de cánceres (183,184).

1.6.10.1.- Proceso de secuenciación

Para poder secuenciar el gen *CIAPIN1* de 9 exones según su transcrito ENST0000039439.8, se diseñaron 9 amplicones que cubrieron toda su secuencia y regiones adyacentes. La cobertura media del proceso fue de 199.27X. En la figura 31 se puede ver su distribución en la población a lo largo del gen.

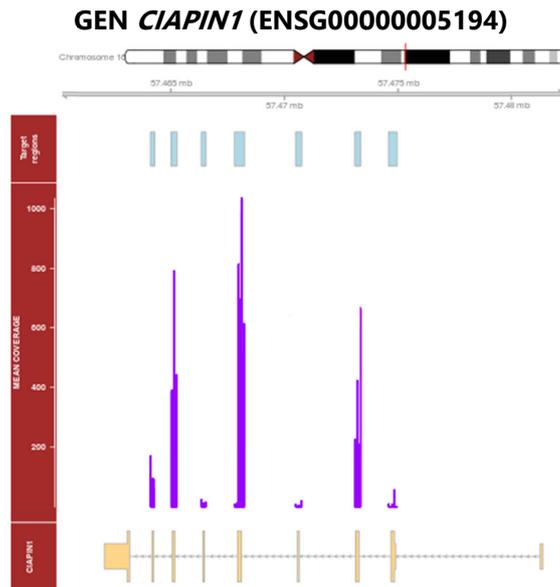


Figura 29: Distribución de la cobertura en la población secuenciada del gen *CIAPIN1*.

1.6.10.2.- Variantes identificadas

En el análisis del gen *CIAPIN1* se han encontrado 3 variantes en total. En la tabla 26 se pueden ver clasificadas por impacto.

Tabla 26: Clasificación por impacto de las variantes del gen *CIAPIN1*.

GEN	IMPACTO	CONSECUENCIA	N° DE VARIANTES			TOTAL
			casos	controles	comunes	
<i>CIAPIN1</i>	MODIFICADOR	V. intrónica	0	0	1	1
	BAJO	V. sinónima	1	1	0	2
			1	1	1	3
						TOTAL

Se han encontrado una variante intrónica común a ambos grupos y dos variantes sinónimas, una en casos y otra en controles.

Después de realizar el Test estadístico de Fisher no se han encontrado variantes significativas, aunque si una variante intrónica con valor de OR menos a 0.5. En la tabla 27 se pueden ver las características.

Tabla 27: Características y descripción de la variante encontrada para el gen *CIAPIN1*.

A) Información de la variante y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
16:57468150	T/G	rs372235600	V. intrónica	g.57468150T>G	6.67	TREND	0.397(0.088-1.794)	0.246

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de la variante en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(contróles)	Casos				Contróles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
16:57468150	1.61E-04	0.0026	0.0067	383	2	0	385	816	9	1	826

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

No se han encontrado datos de asociación de variantes de este gen con la DM2.

1.6.11.- GEN *CYP1A1*

El gen *CYP1A1* (HGNC:2595) codifica el miembro 1 de la subfamilia A de la familia 1 de los citocromos P450. Estas proteínas están implicadas en mecanismos de biotransformación de químicos endógenos o xenobióticos, así como medicamentos o drogas. También están implicados en la síntesis de colesterol, esteroides y otros lípidos. Se localizan en el retículo endoplasmático y son inducidos por factores externos e internos. Se ha visto que *CYP1A1* está implicado en aclaramiento metabólico de drogas y es responsable de la variabilidad interindividual de su farmacogenética (185). Se ha relacionado con enfermedades vasculares como resistencia a la insulina y DM2, ya que tanto la familia CYP1 y la insulina tienen papeles importantes en el metabolismo oxidativo relevante para la aterosclerosis (186). También se ha comprobado que *CYP1A1* puede estar implicado en la regulación de la proliferación de

las células cancerosas y su supervivencia por inhibir la quinasa AMPK además de participar en el proceso de metástasis por actuar sobre la señal de la β -catenina (187).

1.6.11.1.- Proceso de secuenciación

EL gen *CYP1A1* tiene 7 exones según el transcrito ENST00000395048 utilizado. Se han necesitado diseñar 12 amplicones para cubrir toda la secuencia codificante y sus adyacentes. En el proceso de secuenciación se ha conseguido una cobertura media de 32.74X. En la figura 32 se puede ver el patrón de distribución de la cobertura de la población a lo largo de la secuencia del gen.

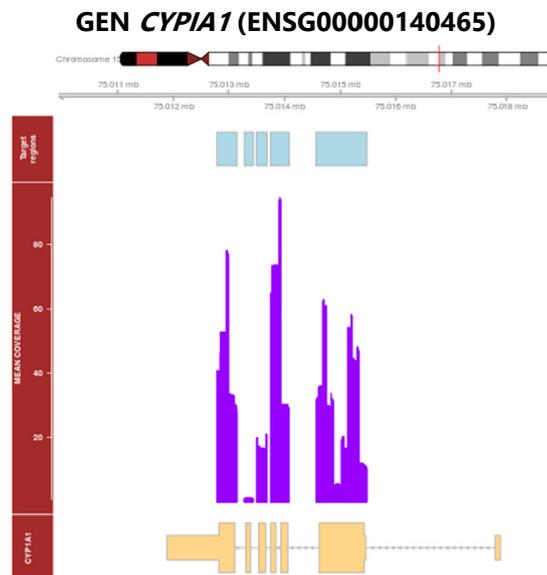


Figura 30: Distribución de la cobertura en la población secuenciada del gen *CYP1A1*.

1.6.11.2.- Variantes identificadas

En el gen *CYP1A1* se han encontrado 32 variantes totales. En la tabla 28 se pueden ver.

Tabla 28: Clasificación por impacto de las variantes del gen *CYP1A1*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
<i>CYP1A1</i>	ALTO	V. aceptora de <i>splicing</i>	0	1	0	1	
		V. ganancia de <i>STOP</i>	1	0	1	2	
	MODERADO	V. <i>missense</i>	3	11	6	20	
	MODIFICADOR	V. 3' UTR	0	1	1	2	
		V. intrónica	0	1	0	1	
	BAJO	V. <i>splicing</i> e intrónica	0	1	0	1	
		V. sinónima	1	2	2	5	
				5	17	10	32

Un total de 5 variantes se han encontrado en el grupo de casos y 17 en el grupo de controles. 10 variantes se han contado en ambos grupos. Más de un 74% de las variantes son de impacto intermedio, siendo las más frecuentes las *missense*.

En el análisis de asociación estadístico mediante el Test de Fisher, no se han encontrado variantes significativas, aunque si algunas interesantes respecto a su valor de OR. Sus características se recogen en la tabla 29.

Tabla 29: Características y descripción de las variantes encontradas para el gen *CYP1A1*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
15:75012827	C/T		V. 3'UTR	g.75012827C>T	1.79	TREND	2.012(0.539-7.515)	0.288
15:75014697	G/A		V. sinónima	p.Leu248=	5.78	TREND	2.054(0.459-9.196)	0.336
15:75014727	G/A	rs61747605	V. <i>missense</i>	p.Pro238Ser	26.10	TREND	3.240(0.293-35.780)	0.310
15:75015234	G/A	rs572138243	V. de <i>STOP</i>	p.Gln69Ter	29.80	TREND	0.478(0.050-4.600)	0.513

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	CADD	Casos				Controles			
					0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
15:75012827		0.0058	0.0029	1.79	425	5	0	430	684	4	0	688
15:75014697		0.0041	0.0020	5.78	489	4	0	493	745	3	0	748
15:75014727	5.46E-04	0.0021	0.0007	26.10	467	2	0	470	749	1	0	750
15:75015234	1.79E-05	0.0013	0.0027	29.80	385	1	0	386	549	3	0	552

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

El MAF de los polimorfismos más interesantes es muy bajo por lo cual con una muestra limitada, no podemos saber realmente su importancia. No hemos encontrado otros datos que indiquen una asociación de este gen con DM2. Serían necesarios más trabajos sobre este gen para conocer su posible papel en el desarrollo de la DM2.

1.6.12.- GEN *DDO*

El gen *DDO* (HGNC:2722) codifica para la proteína D-aspartato oxidasa, es una flavoproteína esteroespecífica que actúa en la vía de degradación de los D-aminoácidos generando 2-oxo ácidos, peróxido de hidrógeno y amonio. Es altamente específica para los D-aminoácidos ácidos como el D-aspártico y el D-glutámico. Tiene localización peroxisomal y alta actividad en el riñón, hígado y cerebro. El D-aspártico es un neurotransmisor que actúa en diferentes procesos del organismo como en la reproducción, a nivel endocrino y nervioso. Al ser la *DDO* la única enzima conocida que cataliza su degradación, variantes en esta enzima, pueden reducir su concentración de forma anómala, desregulando la homeostasis del sistema glutaminérgico, pudiéndose producir procesos neurodegenerativos como la esquizofrenia. Este hecho, convierte a la *DDO* en una diana plausible de medicamentos antipsicóticos (188,189).

1.6.12.1.- Proceso de secuenciación

Con el fin de secuenciar el gen *DDO*, de 5 exones según su transcrito ENST00000368624, fueron necesarios 9 amplicones para cubrir la totalidad de la secuencia. La cobertura media de este gen fue de 201.93X. En la figura 33 se puede ver su distribución en la población secuenciada.

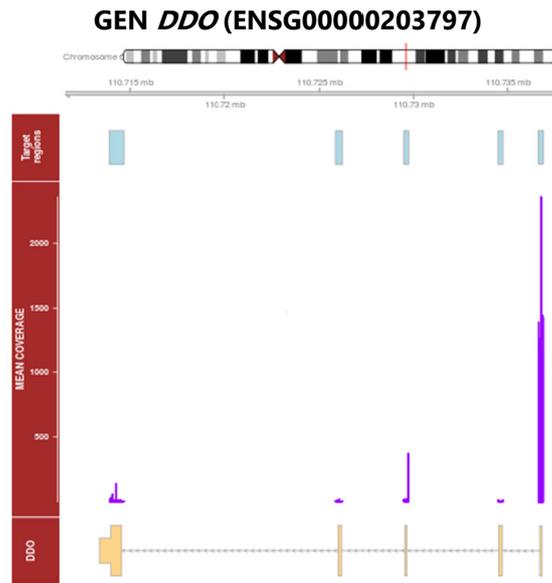


Figura 31: Distribución de la cobertura en la población secuenciada del gen *DDO*.

1.6.12.2.- Variantes identificadas

En el estudio del gen *DDO* se han encontrado 2 variantes totales. En la tabla 30 se pueden ver agrupadas por impacto.

Tabla 30: Clasificación por impacto de las variantes del gen *DDO*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
<i>DDO</i>	MODIFICADOR	V. intrónica	0	0	1	1	
	BAJO	V. sinónima	0	1	0	1	
			0	1	1	2	TOTAL

La variante intrónica se ha visto en ambos grupos y la variante sinónima en el grupo de controles.

No se han encontrado variantes estadísticamente significativas en el análisis realizado. En la tabla 31 se puede ver la distribución de genotipos y características de ambas variantes.

Tabla 31: Características y descripción de las variantes encontradas para el gen *DDO*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD
6:110736610	T/C	rs9384742	V. intrónica	g.110736610T>C	8.64
6:110736696	G/T	rs184476725	V. sinónima	p.Gly17=	9.44

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
6:110736610		0.4251	0.4613	149	239	79	467	226	425	163	814
6:110736696	7.16E-05	0	0.0006	500	0	0	500	824	1	0	825

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

No hay información descrita que relacione este gen con la DM2.

1.6.13.- GEN EBLN2

El gen *EBLN2* (HGNC:25493) codifica una proteína similar a la nucleoproteína endógena 2 del Bornavirus. Estos elementos homólogos existen en muchos genomas de mamíferos, contienen un ORF (*Open Reading Frame*) intacto que permite su expresión como ARN mensajero. Se les considera inserciones antiguas del virus que pueden ser procesados por transcriptasas reversas de retrotransposones tipo LINE-1s. Por ahora no se conoce la función exacta de este gen en los humanos, pero sí que se ha visto que su expresión se ve disminuida en el cáncer cervical, debido a una variante en su secuencia que trunca su expresión (190,191):

1.6.13.1.- Proceso de secuenciación

Para la secuenciación del gen *EBLN2*, el cual según el transcrito ENST00000533473 tiene 1 exón de 1679 pb, se han necesitado diseñar 9 amplicones para cubrir toda su secuencia codificante. La cobertura media alcanzada fue de 46.25X. En la figura 34 se puede ver la distribución de la misma en la población.

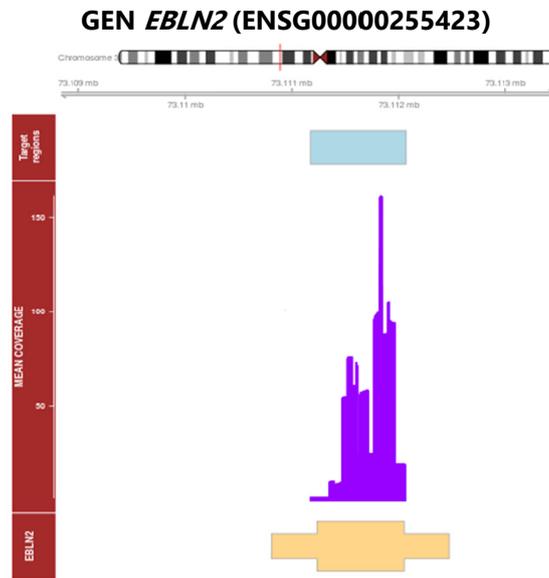


Figura 32: Distribución de la cobertura en la población secuenciada del gen *EBLN2*.

1.6.13.2.- Variantes identificadas

En el análisis del gen *EBLN2* se han encontrado 12 variantes totales. En la tabla 32 se pueden ver organizadas por impacto.

Tabla 32: Clasificación por impacto de las variantes del gen *EBLN2*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
EBLN2	ALTO	V. pauta de lectura	0	1	1	2	
	MODERADO	V. inserción sin cambio*	0	0	1	1	
		V. <i>missense</i>	1	5	2	8	
	BAJO	V. sinónima	0	1	0	1	
			1	7	4	12	TOTAL

*Original: Variante de inserción sin cambio en pauta de lectura.

Se ha encontrado una variante en el grupo de casos y 7 variantes en el grupo de controles. Más del 73% de las variantes tienen un impacto intermedio, siendo las más frecuentes las *missense*.

No se han encontrado variantes significativas al hacer el Test de asociación de Fisher. En la tabla 33 se pueden ver la distribución de los genotipos y las características de las 12 variantes encontradas.

Tabla 33: Características y descripción de las variantes encontradas para el gen *EBLN2*.

A) Información de las variantes.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD
3:73111481	C/CA	rs3832186	V. pauta de lectura	p.Arg84fs	0
3:73111503	T/TTGG	rs143235716	V. inserción sin cambio*	p.Leu91Glu92insVal	0
3:73111521	A/G		V. <i>missense</i>	p.Ile97Val	7.06
3:73111541	TA/T		V. pauta de lectura	p.Lys105fs	0
3:73111558	C/G	COSM260633	V. <i>missense</i>	p.Ser109Cys	13.65
3:73111624	T/A	rs200840536	V. <i>missense</i>	p.Phe131Tyr	9.75
3:73111807-9	GCA/ACG		V. <i>missense</i>	p.Gly192_Ile193delinsAspVal	0
3:73111809	A/G	rs2231926	V. <i>missense</i>	p.Ile193Val	1.66
3:73111828	A/G	rs201170532	V. <i>missense</i>	p.Asp199Gly	11.07
3:73111879	T/G	rs2231927	V. <i>missense</i>	p.Met216Arg	16.04
3:73111886	T/C	rs377295977	V. sinónima	p.Ser2180=	3.13
3:73111917	G/T		V. <i>missense</i>	p.Asp229Tyr	15.30

*Original: Variante de inserción sin cambio en pauta de lectura. Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*.

B) Distribución de las variantes en la población estudiada.

crposición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
3:73111481	0.105	0.1210	0.1317	340	90	8	438	478	145	11	634
3:73111503	0.028	0.0405	0.0379	332	23	3	358	475	39	0	514
3:73111521		0.00	0.0007	511	0	0	511	763	1	0	764
3:73111541		0.00	0.0007	511	0	0	511	763	1	0	764
3:73111558		0.00	0.0007	511	0	0	511	763	1	0	764
3:73111624	1.90E-04	0.00	0.0013	385	0	0	385	755	2	0	757
3:73111807-9		0.00	0.0006	522	0	0	522	781	1	0	782
3:73111809	0.511	0.5069	0.5207	118	265	125	508	166	409	198	773
3:73111828	0.001	0.0010	0.0013	522	1	0	523	780	2	0	782
3:73111879	1.08E-04	0.00	0.0007	502	0	0	502	752	1	0	753
3:73111886	2.70E-05	0.00	0.0013	502	0	0	502	749	2	0	751
3:73111917		0.0010	0.00	517	1	0	518	776	0	0	776

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Tampoco se ha encontrado información en bibliografía que relacione este gen con DM2.

1.6.14.- GEN *GCKR*

El gen *GCKR* (HGNC:4196) codifica para la proteína reguladora de la glucoquinasa, (GCKP). Tiene localización nuclear y es exclusiva de los hepatocitos. Su función principal es modular la concentración de la glucoquinasa (GK), primera enzima de la glicólisis, que convierte la glucosa en glucosa-6-fosfato (G6P) (192). De forma natural, la GCKP se encuentra formando un complejo estable con GK en el núcleo, cuando la concentración de glucosa en sangre aumenta y entra en la célula, la fructosa-1-fosfato, la cual también se ve aumentada, es la encargada de romper la interacción entre ambas y la GK migra al citoplasma donde puede ejercer su función. El aumento de G6P, favorece de forma alostérica la unión de ambas proteínas en el núcleo, disminuyendo a su vez su propia síntesis. Diversas variantes que disminuyen la actividad de *GCKP* se han visto asociadas a DM2, Diabetes tipo MODY 2, hipertrigliceridemia o el síndrome del hígado graso no alcohólico (193,194).

1.6.14.1.- Proceso de secuenciación

Con el fin de secuenciar el gen *GCKR*, el cual tiene 19 exones, según el transcrito ENST00000264717, se diseñaron 19 amplicones para cubrir todos ellos y las secuencias flanqueantes. El proceso de secuenciación tuvo una cobertura media de 18.40X. En la figura 35 se puede ver su perfil en la población secuenciada.

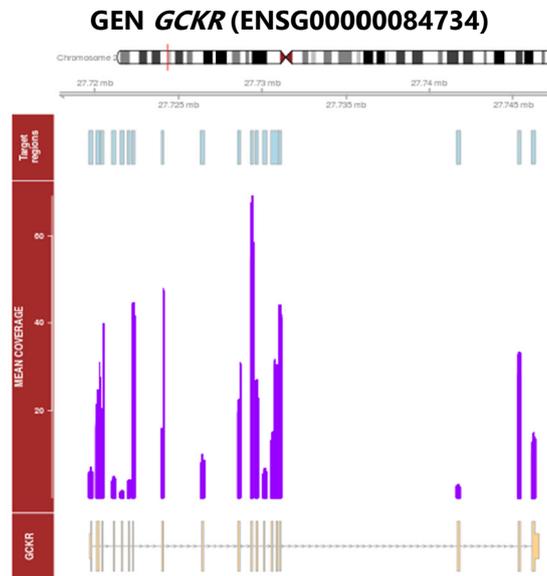


Figura 33: Distribución de la cobertura en la población secuenciada del gen *GCKR*.

1.6.14.2.- Variantes identificadas

En el análisis del gen *GCKR* se han encontrado 25 variantes en total. Están resumidas en la tabla 34.

Tabla 34: Clasificación por impacto de las variantes del gen *GCKR*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
GCKR	ALTO	V. ganancia de STOP	0	0	1	1	
	MODERADO	V. missense	3	12	1	16	
	MODIFICADOR	V. intrónica	1	1	3	5	
	BAJO	V. splicing e intrónica	0	1	1	2	
		V. sinónima	0	1	0	1	
			4	15	6	25	TOTAL

En el grupo de casos han aparecido 4 variantes únicas y en el grupo de controles 15 variantes. Son de impacto intermedio más de un 84%, siendo las más comunes las *missense* y las intrónicas.

En el análisis de asociación estadístico con el Test de Fisher se ha encontrado una variante intrónica significativa y una *missense* con alto valor de OR. En la tabla 35 están sus características.

Tabla 35: Características y descripción de las variantes encontradas para el gen *GCKR*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
2:27728723	G/A		V. intrónica	g.27728723G>A	1.67	GENO/DOM	0.257(0.074-0.886)	0.019
2:27745373	G/A	rs8179249	V. missense	p.Arg540Gln	8.92	TREND	6.073(0.678-54.430)	0.066

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

chr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
2:27728723		0.0043	0.0167	344	3	0	347	462	16	0	478
2:27745373	0.0070	0.0048	0.0008	412	4	0	416	626	1	0	627

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Estos dos polimorfismos no han sido descritos en la bibliografía, pero si se han estudiado otros de este mismo gen asociados a alteraciones metabólicas relacionadas con diabetes.

Por ejemplo, las variantes más comunes descritas de este gen, los polimorfismos rs780094 y rs1260326 se encuentran en desequilibrio de ligamiento entre ellos, ($r^2 = 0.96$) según datos del HapMap II se ha visto en varios estudios, que reducen los niveles de glucosa e insulina protegiendo en este caso contra el desarrollo de resistencia a insulina y diabetes (195). Aunque se han asociado negativamente con el riesgo de dislipemia al favorecer el aumento de los niveles de triglicéridos (196,197).

La variante de ganancia de STOP que aparece en ambos grupos se ha identificado como la misma variante que apareció en el estudio previo, el polimorfismo rs146053779. En esta población no ha resultado significativa. Pero es dato importante a tener en cuenta en otros estudios posteriores. En la tabla 36 se pueden ver sus características.

Tabla 36: Descripción de la variante del estudio previo en la población caso/control BN-BCN.

A) Información de la variante y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
2:27745372	C/T	rs146053779	V. pérdida STOP	R540*	35	TREND	0.602(0.117-3.112)	0.71

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
2:27745372	0.0014	0.0024	0.0040	415	2	0	417	622	5	0	627

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

1.6.15.- GEN HABP2

El gen *HABP2* (HGNC:4798) codifica para la proteína 2 de unión al ácido hialurónico. Pertenece a la familia de peptidasas S1 serina proteasas. También es conocida como proteasa activadora del Factor VII (FSAP), la cual actúa en el proceso de coagulación sanguínea activando el Factor VII y en la fibrinólisis regulando el activador de la pro-uroquinasa tipo plasminógeno (pro-uPA). Además, se ha relacionado con otros procesos como la trombosis venosa profunda, la permeabilidad endotelial, regulación del factor de crecimiento, supervivencia celular, así como se ha visto implicada en el riesgo de cáncer de tiroides y algunos tipos de cáncer de pulmón de células no pequeñas (198,199). Se conoce una variante, llamada polimorfismo de Marburgo I (p.Gly534Glu, rs080536), la cual está presente en un 5% de la población caucásica, que produce un descenso en la actividad proteolítica de la proteína sobre varios de sus sustratos. Siendo asociado este hecho a un incremento en el riesgo de tromboembolismos y estenosis carotídeas (200,201).

1.6.15.1.- Proceso de secuenciación

El gen *HABP2*, según el transcrito ENST00000533473 tiene 13 exones, para los cuales se han diseñado 23 amplicones con el fin de secuenciar todas las zonas codificantes y las regiones intrónicas adyacentes. El proceso de secuenciación tuvo una cobertura media de 46.25X. Su distribución en la población se puede ver en la figura 36.

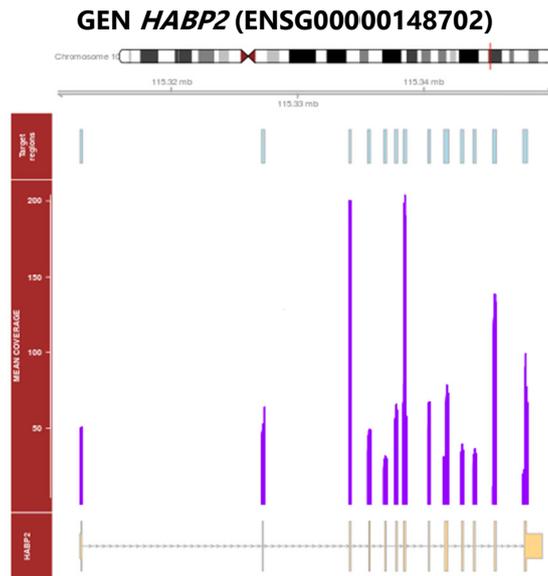


Figura 34: Distribución de la cobertura en la población secuenciada del gen *HABP2*.

1.6.15.2.- Variantes identificadas

Se han encontrado 80 variantes totales en el análisis del gen *HABP2*. En la tabla 37 se puede ver su clasificación por impacto.

Tabla 37: Clasificación por impacto de las variantes del gen *HABP2*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
<i>HABP2</i>	ALTO	V. donadora de <i>splicing</i>	1	0	0	1	
		V. ganancia de <i>STOP</i>	1	0	1	2	
	MODERADO	V. <i>missense</i>	7	18	6	31	
		V. <i>missense</i> y de <i>splicing</i>	1	1	0	2	
	MODIFICADOR	V. 3' UTR	1	1	0	2	
		V. 5' UTR	1	0	0	1	
		V. intrónica	6	10	7	23	
	BAJO	V. <i>splicing</i> e intrónica	1	1	0	2	
		V. sinónima	1	10	5	16	
				20	41	19	80

De entre ellas, 20 variantes se han encontrado solo en caso y 41 solo en controles. Hay 19 variantes que se han encontrado en ambos

grupos. Más del 73% de las variantes se clasifican con un impacto intermedio, siendo las más comunes las *missense* y las intrónicas.

En la tabla 38 se pueden ver los datos de las variantes que poseen valores de OR menores a 0.5 o mayores a 2 en el análisis de asociación, ninguna de ellas ha resultado significativa.

Tabla 38: Características y descripción de las variantes encontradas para el gen *HABP2*.

A) Información de las variantes y resultados estadísticos.

cr.posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
10:115337833	C/T	rs142936446	V. <i>missense</i>	p.Ser497Phe	8.93	TREND	0.553(0.057-5.327)	0.603
10:115342945	C/T		V. intrónica	g.115342945C>T	1.83	TREND	0.554(0.200-1.530)	0.245
10:115345563	C/T	rs141433300	V. <i>missense</i>	p.Arg1384Cys	33.00	TREND	3.029(0.274-33.450)	0.341

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr.posición	MAF(EUR)	MAF(casos)	MAF(contróles)	Casos				Contróles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
10:115337833	0.0015	0.0011	0.0020	448	1	0	449	734	3	0	737
10:115342945		0.0058	0.0105	427	5	0	432	699	15	0	714
10:115345563		0.0019	0.0006	519	2	0	521	776	1	0	777

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

No existen datos de la posible relación de este gen con la DM2. Sería necesario realizar otros estudios para corroborarlo.

1.6.16.- GEN OLFML1

El gen *OLFML1* (HGNC:24473) se expresa en un producto parecido a la proteína 1 olfactomedina. Pertenece a la subfamilia VII de la familia de glicoproteínas OLF, ya que posee un dominio OLF en su parte C-terminal muy semejante (202). Se expresa mayormente en las microvellosidades intestinales, pulmón, hígado, bazo y corazón. Su caracterización ha confirmado que es una proteína secretada al exterior celular, la cual sufre modificaciones postranscripcionales

como la glicosilación en su parte N-terminal. Trabajos realizados *in vitro* ha demostrado que su presencia, favorece la proliferación celular al acelerar la entrada en la fase S (203). También se ha encontrado una mutación de *nov* Met250Leu que podría tener un efecto inhibitor sobre la ruta de señalización Hippo, afectando negativamente al proceso de mineralización de los osteoblastos (204). Más recientemente, se ha realizado un análisis de rutas metabólicas y perfiles de interacción entre genes resultando el gen *OLFML1*, entre otros, poseer una actividad aumentada asociada a la presencia de resistencia a insulina en pacientes con diabetes no dependiente de insulina u obesos (205).

1.6.16.1.- Proceso de secuenciación

El gen *OLFML1* tiene 3 exones según el transcrito ENST0000329293.3. Se han diseñado 9 amplicones en total para cubrir su secuencia exónica y regiones adyacentes para llevar a cabo el proceso de secuenciación. La cobertura media del mismo ha sido de 25.06X. En la figura 37 se puede ver su distribución en la población.

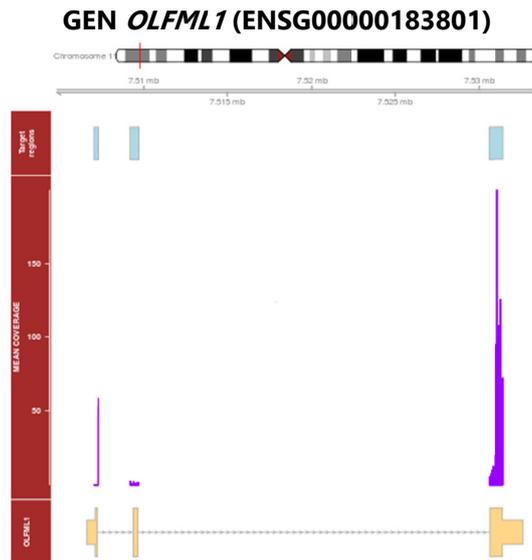


Figura 35: Distribución de la cobertura en la población secuenciada del gen *OLFML1*.

1.6.16.2.- Variantes identificadas

En el proceso de análisis del gen *OLFML1* se han encontrado solamente 5 variantes *missense*, están recogidas en la tabla 39.

Tabla 39: Clasificación por impacto de las variantes del gen *OLFML1*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL
			casos	controles	comunes	
<i>OLFML1</i>	MODERADO	V. <i>missense</i>	1	1	3	5

Tras el análisis estadístico con el Test de Fisher solo se ha encontrado una variante significativa En la tabla 40 se pueden ver sus características.

Tabla 40: Características y descripción de las variantes encontradas para el gen *OLFML1*.

A) Información de las variantes y resultados estadísticos.

cr.posición	Alelos	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
11:7531084	T/G	V. <i>missense</i>	p.Leu292Val	23.10	TREND	0	0.018

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr.posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
11:7531084	0	0	0.005	571	0	0	571	701	7	0	708

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Esta variante no está descrita en la bibliografía ni se ha podido calcular un valor OR, siendo muy interesante su análisis en otras muestras caso/control y poblaciones para saber su papel y su relación con la DM2.

1.6.17.- GEN SLC12A4

El gen *SLC12A4* (HGNC:10913) codifica un co-transportador de K^+ / Cl^- independiente de Na^+ también conocido como KCC1. Pertenece a una de las ramas de la familia SLC12, cuyos miembros se regulan por modificaciones de fosforilación/desfosforilación en residuos diana serina/treonina localizados en los dominios N-terminal y C-terminal. En este caso, KCC1 es inhibida por la quinasa WNK4 al ser fosforilada (206,207). Estos tipos de cotransportadores tienen papeles importantes en varios procesos fisiológicos como la regulación del volumen celular, la modulación de la concentración intracelular del cloruro y el transporte iónico transepitelial. KCC1 se expresa mayoritariamente en el riñón, más específicamente en algunos segmentos de la nefrona como los túbulos proximal y distal y la médula renal (207,208). La actividad de las proteínas KCC es osmóticamente sensible a los cambios de volumen celular, en algunos estudios con líneas celulares cancerosas, se ha visto que el Factor 1 de crecimiento parecido a la insulina (IGF-1), el cual está en concentraciones superiores a las normales en este tipo de células, estimula la actividad de las proteínas KCC, su sobreexpresión

permitiría a estas células, una ventaja selectiva para su estabilidad y proliferación (209).

1.6.17.1.-

El gen *SLC12A4*, según su transcrito ENST00000316341, tiene 24 exones. Se han necesitado 31 amplicones para cubrirlos enteros. La cobertura media del proceso de secuenciación ha sido de 83.74X. En la figura 38 se puede ver su perfil en la población secuenciada.

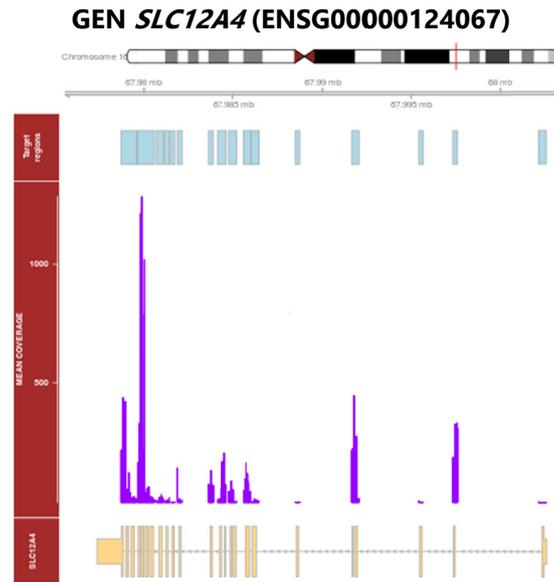


Figura 36: Distribución de la cobertura en la población secuenciada del gen SLC12A4.

1.6.17.2.- Variantes identificadas

Para el gen *SLC12A4* se han encontrado 17 variantes totales, las cuales se recogen en la tabla 41.

Tabla 41: Clasificación por impacto de las variantes del gen *SLC12A4*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL
			casos	controles	comunes	
SLC12A4	MODERADO	V. <i>missense</i>	2	1	2	5
	MODIFICADOR	V. intrónica	2	3	3	8
	BAJO	V. <i>splicing</i> e intrónica	2	0	0	2
		V. sinónima	1	1	0	2
			7	5	5	17
						TOTAL

Han salido 7 variantes en el grupo de casos y 5 variantes únicas en el grupo de controles. Se han encontrado 5 variantes comunes a ambos grupos. De impacto intermedio hay más de un 76% de las variantes totales.

En cuanto al análisis estadístico mediante el Test de Fisher se ha encontrado una variante intrónica significativa y una variante *missense* con un valor de OR mayor a 2. En la tabla 42 se pueden ver sus características.

Tabla 42: Características y descripción de las variantes encontradas para el gen *SLC12A4*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
16:67983745	A/G	rs62059925	V. <i>missense</i>	p.Ile569Thr	26.8	TREND	3.232(1.220-0.293)	0.311
16:67984546	A/G	rs189512607	V. intrónica	g.16:67984546A>G	13.61	TREND	0.468(0.330-0.245)	0.020

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
16:67983745	3.017E-03	2.02E-03	6.34E-04	494	2	0	496	788	1	0	789
16:67984546	7.227E-03	0.012	0.026	478	12	0	491	765	40	1	806

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Para poder conocer la posible relación de este gen con la DM2 es necesario realizar otros estudios poblacionales y analizar sus resultados.

1.6.18.- GEN SLC15A3

El gen *SLC15A3* (HGNC:18068) expresa una proteína perteneciente a la familia de transportadores electrogénicos SLC15. Más concretamente codifica el cotransportador 2 de péptidos/histidina, también conocido como PHT2 o hPTR3. Se ha visto expresado en pulmón, bazo, y timo principalmente (210). A nivel intracelular, se ha visto localizado en lisosomas, endosomas tardíos y peroxisomas. Además, se ha estudiado su participación en la respuesta inmune innata a virus, favoreciendo la producción de las citoquinas IFN (211). En otros trabajos, en líneas celulares de macrófagos y dendríticas de ratón se observó su activación por la acción de los ligandos de los receptores TLR mediante la señalización de las vías NF- κ B, MAPK and TRIF (212).

1.6.18.1.- Proceso de secuenciación

Con el fin de secuenciar el gen *SLC15A3*, el cual tiene 8 exones según su transcrito ENST00000227880, se han diseñado 13 amplicones. La cobertura media del proceso ha sido de 264.44X. En la figura 39 se puede ver su distribución en la población.

GEN *SLC15A3* (ENSG00000110446)

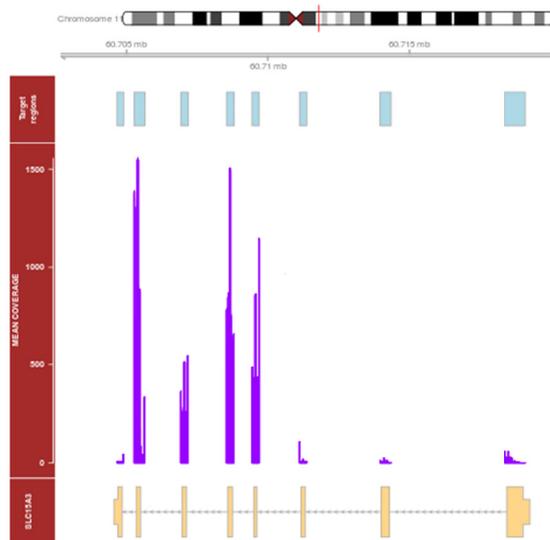


Figura 37: Distribución de la cobertura en la población secuenciada del gen *SLC15A3*.

1.6.18.2.- Variantes identificadas

Tras el análisis del gen *SLC15A3* se han encontrado 20 variantes totales. En la tabla 43 se clasifican según su impacto.

Tabla 43: Clasificación por impacto de las variantes del gen *SLC15A3*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
<i>SLC15A3</i>	ALTO	V. pauta de lectura	0	1	0	1	
		V. ganancia de <i>STOP</i>	1	0	0	1	
	MODERADO	V. <i>missense</i>	3	2	3	8	
	MODIFICADOR	V. intrónica	2	2	0	4	
	BAJO	V. <i>splicing</i> e intrónica	0	0	1	1	
		V. sinónima	4	1	0	5	
			10	6	4	20	TOTAL

En el grupo de casos se han contado 10 variantes únicas y en el grupo de controles 6 variantes. Se han visto 4 variantes que aparecen en ambos grupos. En más del 60% las variantes se clasifican con un

impacto intermedio, siendo las más frecuentes las *missense* y las intrónicas.

Al realizar el Test de Fisher se ha encontrado una variante significativa. En la tabla 44 se pueden ver sus características.

Tabla 44 Características y descripción de las variantes encontradas para el gen *SLC15A3*.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
11:60705454	A/G	rs2905692	V. sinónima	p.Gly493=	11,51	ALLELIC	NA	0,024
11:60705468	G/A	rs36792477	V. <i>missense</i>	p.Arg489Cys	32,00	TREND	131(0.283-34.57)	0,326

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
11:60705454	0.9981	0.0074	0	539	0	4	543	827	0	0	827
11:60705468	1.28E-04	0.0019	0.0006	535	2	0	537	826	1	0	827

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Ambos polimorfismos serían interesantes para su análisis en muestras caso/control y estudios poblacionales de mayor tamaño para poder conocer si se repite este resultado y se pudiera sacar alguna conclusión relacionada con la DM2.

1.6.19.- GEN SREBF2

El gen *SREBF2* (HGNC:11290) codifica para uno de los tres elementos reguladores SREBP, la proteína de unión al elemento regulador de esteroles, SREBP2. Se expresa en forma ubiquitinada y es uno de los elementos más importantes, junto a SREBP1 y SREBP1c en la regulación del metabolismo de colesterol y los ácidos grasos. Se localiza en la membrana nuclear y del retículo endoplasmático ya que posee un dominio de cremallera de leucina hélice-bucle-hélice. Está regulado por los niveles de colesterol de la célula, así como por

fosforilación. Modificaciones postraduccionales son necesarias para generar la forma madura de la proteína y traslocarla al núcleo, con el fin de que active la transcripción de sus genes diana (213,214). Uno de ellos es el promotor del receptor de LDLs, donde se une al elemento clásico regulador de esterol *cis* (sre-1/ATCACCCCAC). Este elemento se encuentra en una gran cantidad de genes que codifican enzimas relacionadas con el metabolismo de lípidos. También se ha encontrado este elemento de unión al gen de la insulina, en el del factor de crecimiento parecido a la insulina (IGF-1), factor de crecimiento derivado de plaquetas (PDGF) así como del factor de crecimiento epidérmico (EGF). Se ha comprobado que la cascada de señalización de las MAP quinasas, en especial las ERK quinasas, las cuales controlan la regulación de la insulina y del PDGF, también actuarían fosforilando a los elementos SREBPs y activándolos. Este hecho hace que, los elementos SREBPs tengan un papel de convergencia entre varias rutas de señalización tanto metabólicas como endocrinas (215). Variantes encontradas en este gen, por lo tanto, pueden producir una desregulación de las vías en las que está implicado aumentando el riesgo de sufrir diversas enfermedades como aterosclerosis (216) y síndromes relacionados con el sistema inmune, ya que los SREBP controlan el metabolismo glicolítico de las células NK y los linfocitos T activos (217). También se han hechos estudios de asociación de variantes en el gen con DM2 (218).

1.6.19.1.- Proceso de secuenciación

El gen *SREBF2* tiene 19 exones según el transcrito ENST00000361204, se diseñaron 38 amplicones incluyendo además 600 pb de la región

promotora. La cobertura media del proceso fue de 76.06X. En la figura 40 se puede ver su distribución en la población.

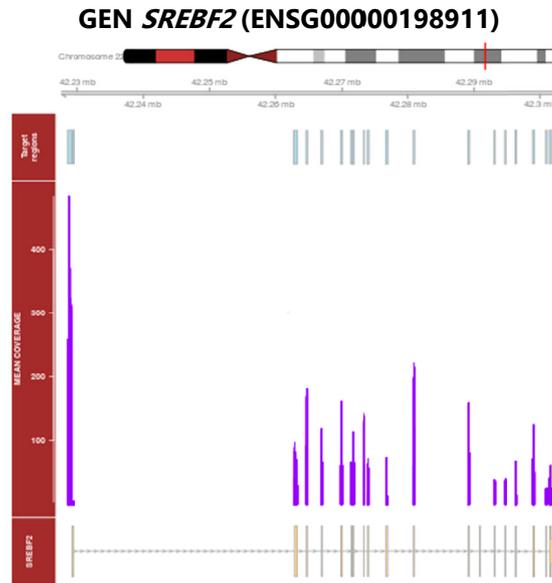


Figura 38: Distribución de la cobertura en la población secuenciada del gen *SREBF2*.

1.6.19.2.- Variantes identificadas

En el análisis del gen *SREBF2* se han contado un total de 123 variantes. Se pueden ver clasificadas por impacto en la tabla 45.

Tabla 45: Clasificación por impacto de las variantes del gen *SREBF2*.

GEN	IMPACTO	CONSECUENCIA	Nº DE VARIANTES			TOTAL	
			casos	controles	comunes		
<i>SREBF2</i>	ALTO	V. pauta de lectura	2	0	0	2	
		V. donadora de <i>splicing</i>	0	1	0	1	
	MODERADO	V. <i>missense</i>	8	20	7	35	
	MODIFICADOR	V. intrónica	3	7	7	17	
		V. <i>upstream</i>	25	18	2	45	
	BAJO	V. de <i>splicing</i> e intrónica	1	1	0	2	
		V. de <i>splicing</i> y sinónima	1	0	0	1	
V. sinónima		7	5	8	20		
			47	52	24	123	TOTAL

Separándolas por grupos, se han contado 47 variantes en los casos y 52 variantes en los controles. Las 24 variantes restantes se han visto en ambos grupos. Más de un 78% de ellas se clasifican con un impacto intermedio, siendo las más frecuentes las de tipo *upstream*, pero porque en este caso se ha analizado también la región promotora del gen.

En el análisis de asociación estadística, no se han encontrado variantes significativas pero sí 4 con valores de OR relevantes. En la tabla 46 están sus características.

Tabla 46: Características y descripción de las variantes encontradas para el gen SREBF2.

A) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	Consecuencia	Cambio	CADD	TEST	OR(IC 95%)	P
22:42264651	C/T		V. <i>missense</i>	p.Ser192Phe	24.70	TREND	0.379(0.042-3.396)	0.367
22:42276857	G/A	rs142713090	V. sinónima	p.Leu633=	9.78	TREND	2.591(0.618-10.870)	0.176
22:42280967	T/C	rs374234598	V. sinónima	p.Ala720=	9.17	TREND	4.453(0.462-42.870)	0.157
22:42280996	G/T	rs146030234	V. <i>missense</i>	p.Gly730Val	22.90	TREND	0.502(0.052-4.832)	0.542

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

B) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(casos)	MAF(controles)	Casos				Controles			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
22:42264651		0.0010	0.0026	516	1	0	517	770	4	0	774
22:42276857	3.39E-03	0.0059	0.0023	418	5	0	423	647	3	0	650
22:42280967	4.30E-04	0.0028	0.0006	535	3	0	538	785	1	0	786
22:42280996	4.48E-04	0.0010	0.0020	494	1	0	495	738	3	0	741

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

En este estudio no se han encontrado variantes significativas, pero al tratarse de un gen clave implicado en diversos procesos metabólicos, se deberían hacer más estudios para corroborar su posible efecto en relación a la DM2. Se ha visto que una variante de este gen, no encontrada en el presente estudio, el polimorfismo rs133291, se ha

asociado con la desregulación de la homeostasis del colesterol controlada por SREBF2 en el hígado, lo cual provoca el aumento de los niveles internos de colesterol, favoreciendo la esteatosis hepática no alcohólica, la resistencia a insulina, inflamación y apoptosis celular (219).

1.6.20.- Resumen resultados

Se recoge la relación de los genes analizados en los que se han encontrado variantes con asociación estadísticamente significativa con DM2 en la tabla 47.

Tabla 47: Relación de asociación con DM2 de los genes estudiados.

GENES	ASOCIACIÓN
ACAD10	SI
ACPL2/PXYLP1	NO
ACSL3	SI
ACSM3	NO
AFMID	SI
ARHGEF38	NO
CACNG5	NO
CCNDBP1	SI
CCT8L2	NO
CIAPIN1	NO
CYP1A1	NO
DDO	NO
EBLN2	NO
GCKR	SI
HABP2	NO
OLFML1	SI
SLC12A4	SI
SLC15A3	SI
SRBEF2	NO

2.- Identificación de variantes genéticas de exoma implicadas en el desarrollo de la DM2 mediada o no por RI

2.1.- Diseño del estudio

Los resultados de este apartado de la tesis se estructuraron de la siguiente manera:

1. Se realizó un estudio de evolución de la DM2 y RI con la población Egabro-Pizarra de 119 individuos.
2. Se analizó de nuevo la población caso/control utilizada en los estudios previos.
3. Se agruparon los casos y los controles de ambas poblaciones para obtener una nueva y analizarlos conjuntamente.

En cada uno de los tres sub-análisis se llevaron a cabo las siguientes acciones:

- a) Se contabilizaron las variantes por impacto y por consecuencia de cada estudio, de forma general, por subgrupo y por procedencia de grupo caso o grupo control.
- b) Se llevó a cabo el análisis estadístico de asociación utilizando el Test de Fisher para estudios caso/control y conocer las variantes con significatividad estadística.

2.2.- Características de la población de estudio

2.2.1.- Estudio Egabro-Pizarra

Esta población constó de 119 individuos separados en 4 grupos según su diagnóstico de RI y DM2. Las características más representativas se

pueden ver en la tabla 47. Hay que tener en cuenta que sobre un 80% de los pacientes de los grupos B, C, D tenían algún tipo de tratamiento.

Durante 5 años se hizo el seguimiento de los participantes de los Estudio Egabro y Estudio Pizarra (43,220), de ellos se seleccionaron 4 grupos de individuos que conformaron la nueva muestra:

GRUPO A: no presentaban ni IR ni DM2 en todo el estudio.

GRUPO B: no presentaban ni RI ni DM2 al principio del estudio no presentaban RI al final, pero si DM2.

GRUPO C: presentaban IR al principio, DM2 no, y al final del periodo no presentaban RI y si DM2.

GRUPO D: presentaban RI al principio, DM2 no, y al final del periodo presentaban RI y DM2.

Tabla 48: Características clínicas de los 119 individuos de la población Egabro-Pizarra para el estudio de exoma.

Características	Grupo A (N = 42)	Grupo B (N = 19)	Grupo C (N = 42)	Grupo D (N = 16)
Sexo:				
Hombre (%)	18 (42.85)	7 (36.84)	17 (40.48)	6 (37.50)
Mujer (%)	24 (57.14)	12 (63.16)	25 (59.52)	10 (62.50)
Edad (años)	56.02 ± 7.63*	51.97 ± 5.48*	46.32 ± 13.40	47.89 ± 16.15
IMC_basal (Kg/m²)	29.69 ± 4.78	28.36 ± 10.28	31.57 ± 4.79**	35.60 ± 5.20**
IMC_final (Kg/m²)	29.41 ± 5.21	28.09 ± 8.30	30.62 ± 4.18**	34.96 ± 5.40**
Glucosa_basal (mg/dL)	103.98 ± 12.41*	113.63 ± 14.88*	110.81 ± 16.89**	121.38 ± 19.68**
Glucosa_final (mg/dL)	102.88 ± 10.07*	129.58 ± 53.99*	103.07 ± 12.16**¥	136.69 ± 36.74**
HOMAIR_basal	1.79 ± 0.69	1.83 ± 0.74	5.24 ± 4.12**	7.21 ± 5.64**
HOMAIR_final	3.51 ± 5.29¥	3.09 ± 4.05	3.11 ± 2.82**¥	7.16 ± 6.91**
RI (SI/NO)	NO	NO	SI	SI
DM2_final (SI/NO)	NO	SI	NO	SI

Valores expresados como media ± desviación estándar. IMC: Índice de masa corporal, HOMAIR: Índice de Resistencia a la Insulina. RI: Resistencia a la Insulina. DM2: Diabetes Mellitus tipo 2. *p-valor < 0.05 entre grupos A y B. **p-valor < 0.05 entre grupos C y D. ¥ p-valor < 0.05 comparando valores iniciales y finales dentro del mismo grupo.

2.2.2.- Estudio caso/control

En la tabla 48 se muestran las características clínicas generales finales de la población seleccionada para el análisis de su exoma. Estos datos provienen de la línea de trabajo de la unidad en la que se incluye el presente estudio.

Tabla 49: Características clínicas de los 369 individuos de la población caso/control para el estudio de exoma. Hay que considerar que los diagnosticados como DM2 están bajo tratamiento.

<i>Características</i>	Controles Grupo E (N =184)	Casos Grupo F (N = 185)
Sexo:		
Hombres (%)	112 (60.86)	117 (63.24)
Mujeres (%)	71 (38.59)	68 (36.75)
NA (%)	1 (0.54)	0
Edad (años)	52.72 ± 8.77*	56.42 ± 6.54*
IMC (Kg/m2)	29.16 ± 3.78	29.71 ± 2.83
Glucosa_basal (mg/dL)	81.52 ± 9.78*	150.91 ± 71.86*
DM2 (SI/NO)	NO	SI

Valores expresados como media ± desviación estándar. IMC: Índice de masa corporal. DM2: Diabetes Mellitus tipo 2. *p-valor > 0.05 entre grupo E y F.

2.2.3.- Estudio caso/control conjunto

Esta nueva población proviene de unir los individuos de los grupos A y E como nuevo grupo de controles y los grupos B, D y F como nuevo grupo de casos. El grupo C queda excluido de esta nueva selección por estar compuesto por individuos sin DM2, pero con RI (característica que los diferencia con el grupo de controles iniciales). Por lo que este nuevo subconjunto se compone de 446 individuos, 226 controles y 220 casos.

Las características clínicas de la nueva población serían las indicadas en la tabla 49.

Tabla 50: Características clínicas de toda la población conjunta.

<i>Características</i>	Controles (N = 226)	Casos (N = 220)
Sexo:		
Hombres (%)	130 (57.52)	130 (59.09)
Mujeres (%)	95 (42.03)	90 (40.90)
NA (%)	1 (0.44)	0
Edad (años)	53.33 ± 8.65	55.41 ± 7.89
IMC (Kg/m2)	29.26 ± 3.98*	30.01 ± 4.43*
Glucosa_basal (mg/dL)	85.69 ± 13.51*	145.54 ± 67.36*
DM2_final (SI/NO)	NO	SI

Valores expresados como media ± desviación estándar. IMC: Índice de masa corporal. DM2: Diabetes Mellitus tipo 2. *p-valor > 0.05 entre grupo controles y casos.

2.3.- Control de calidad: Programa *FastQC*

Para poder conocer la calidad de las secuencias de exoma obtenidas directamente del secuenciador y comprobar si tras el procesado de *trimming* la calidad de éstas mejora, se utilizó de nuevo el programa *FastQC v0.11.5* y se compararon los resultados de ambos tipos de archivos. En este caso como se parte de dos estudios diferentes, el estudio Egabro-Pizarra y el estudio caso/control, que han sido secuenciados en dos tecnologías diferentes, los datos deben presentarse por separado, dada la posible variabilidad que ello añade al análisis de la calidad.

En la tabla 50 se pueden ver los valores de calidad para el estudio Egabro-Pizarra realizado en el Sistema Hiseq 2000 (Illumina, Inc, San Diego, EE.UU.).

Tabla 51: Valores de los parámetros del control de calidad de los archivos *FastQ* de las muestras del Estudio Egabro-Pizarra antes y después del proceso de *trimming*.

PARÁMETROS	<i>FastQ</i> preprocesados			<i>FastQ</i> postprocesados		
	<i>PASS</i>	<i>WARN</i>	<i>FAIL</i>	<i>PASS</i>	<i>WARN</i>	<i>FAIL</i>
Estadísticos básicos	100	0	0	100	0	0
Calidad de secuencia por base	100	0	0	100	0	0
Valores de calidad por secuencia	100	0	0	100	0	0
Contenido de N por secuencia	100	0	0	100	0	0
Duplicados de secuencia	31.09	53.36	15.55	31.93	52.94	15.13

Valores expresados en porcentaje.

En este caso se puede ver que los valores de calidad de las secuencias son muy buenos desde el principio, y permiten continuar el análisis con más seguridad. En cuanto al número de duplicados, más del 50% de las muestras se encuentra en la categoría intermedia, esto quiere decir que la existencia de dichos duplicados puede ser debidos, tanto a las amplificaciones por PCR del proceso de realización de las librerías, como a posibles duplicados biológicos de secuencia donde diferentes copias de la misma secuencia son, al azar, seleccionadas y marcadas como duplicados. Estos dos hechos son imposibles de distinguir entre si y hacen que el programa los considere duplicados dentro de la categoría de *Warning*. También influye el nivel conjunto de diversidad de la librería en el proceso de secuenciación. El 15% de las muestras de la categoría *Fail*, tienen niveles de duplicados mayores al 50% de las secuencias totales por muestra, en estos casos se podría sobreestimar sus valores de cobertura, pero en pasos posteriores del

flujo bioinformático se eliminan estos duplicados corrigiendo los valores y permitir una mejor interpretación de los resultados.

En la tabla 51 se pueden ver los parámetros de calidad del estudio caso/control realizado en el Sistema HiScanSQ (Illumina, Inc, San Diego, EE.UU.).

Tabla 52: Valores de los parámetros del control de calidad de los archivos *FastQ* de las muestras del estudio caso/control antes y después del proceso de *trimming*.

PARÁMETROS	<i>FastQ</i> preprocesados			<i>FastQ</i> postprocesados		
	<i>PASS</i>	<i>WARN</i>	<i>FAIL</i>	<i>PASS</i>	<i>WARN</i>	<i>FAIL</i>
Estadísticos básicos	100	0	0	100	0	0
Calidad de secuencia por base	71.79	0.56	27.65	100	0	0
Valores de calidad por secuencia	100	0	0	100	0	0
Contenido de N por secuencia	100	0	0	100	0	0
Duplicados de secuencia	89.94	7.82	2.23	90.22	8.10	1.68

Valores expresados en porcentaje.

Estos datos también indican que la secuenciación de esta población ha sido correcta y que el proceso de *trimming* ha corregido cualquier alteración en la calidad de las secuencias. En cuanto al nivel de duplicados, en este análisis, el 89.94% de las muestras pasa correctamente el umbral de duplicados de secuencia y solo un 2.23% de muestras falla. De todas formas, en el paso posterior de eliminación de duplicados se corrigen los valores llegando a más de un 90.22% de muestras sin duplicados de secuencia, lo cual, favorece su posterior análisis.

2.4.- Análisis de Coberturas

Para el análisis de coberturas de las muestras se contó de nuevo con la ayuda de Carolina Monzó y Azahara Fuentes a la hora de utilizar los programas pertinentes y la automatización de los resultados.

Se llevó a cabo el análisis de coberturas obtenidas de las regiones de interés para ser estudiadas, en este caso todos los exones y las zonas UTRs de los genes, y se comparó con las regiones cubiertas por las sondas del kit comercial utilizado. Según el diseño de cada empresa, las sondas diseñadas no cubren completamente las regiones de interés esto se debe a que en el proceso de enriquecimiento se pueden incluir regiones contiguas a las que hibridan con las sondas creando el llamado cono de secuenciación. Este cono ayudará a secuenciar estas regiones y cubrir todo el exón y las regiones contiguas. Para ello hay que tener en cuenta no solo las coordenadas de las sondas comerciales sino también las de los exones y regiones contiguas basadas, en este caso en el genoma de referencia *GrCh37* de *Ensembl*.

En la figura 41 se puede ver la densidad de cobertura de las regiones cubiertas por las sondas comerciales, la de las regiones exónicas no cubiertas y la combinación de las zonas exónicas y las sondas.

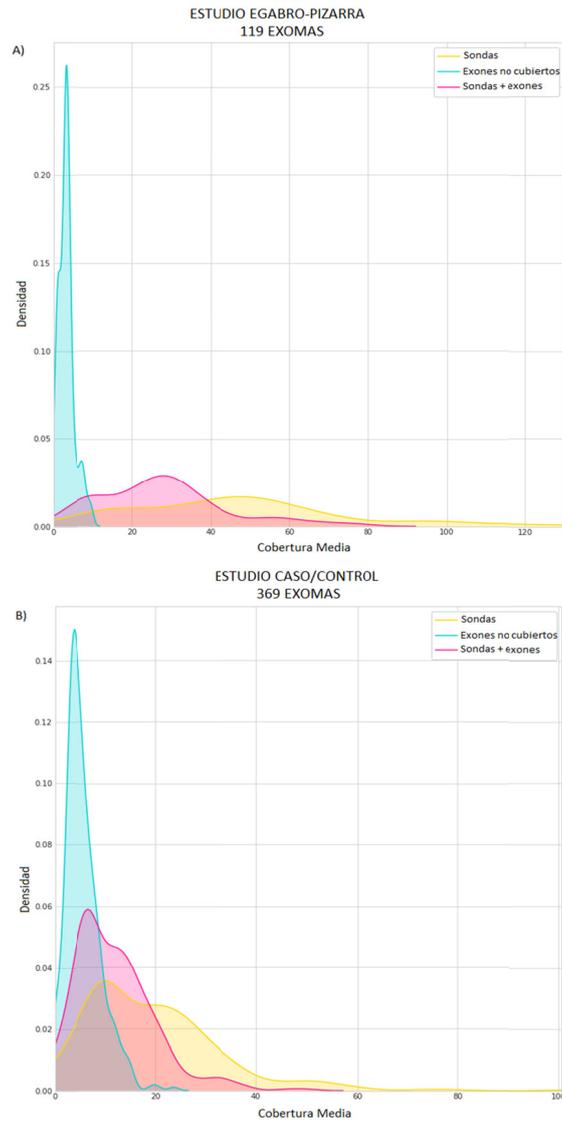


Figura 39: Representación de la densidad frente a la cobertura media en exones, sondas y la suma de ambos. A) Estudio Egabro-Pizarra con 119 exomas secuenciado en el Sistema Hiseq2000 (Illumina, Inc) y Estudio caso/control con 369 exomas secuenciado en el Sistema HiScanSQ (Illumina, Inc).

Comparando por separado las coberturas de las diferentes zonas entre los grupos, en la tabla 52, se puede comprobar que, de acuerdo a lo esperado, la cobertura media es mayor en las zonas cubiertas por las sondas que en la combinación de las sondas con las regiones exónicas teóricas. Aun así, la cobertura alcanzada total es suficiente para continuar con la identificación de variantes genéticas de estos exomas.

Tabla 53: Descripción de las coberturas en las diferentes regiones y por subgrupos de ambas poblaciones. A) Estudio Egabro-Pizarra con 119 exomas y B) Estudio Caso/Control con 369 exomas.

Estudio Egabro-Pizarra 119 exomas		Cobertura media	Desviación estándar	Estudio Caso/Control 369 exomas		Cobertura media	Desviación estándar
A) Sondas	Total	46.737X	± 24.202	B) Sondas	Total	19.245X	± 13.506
	Grupo A	50.579X	± 24.847		Casos	16.823X	± 9.980
	Grupo B	50.836X	± 41.551		Controles	21.628X	± 15.916
	Grupo C	43.295X	± 18.895	Exones	Total	5.587X	± 3.505
	Grupo D	40.816X	± 30.866		Casos	4.956X	± 2.751
Exones	Total	3.338X	± 2.013	Controles	6.212X	± 4.028	
	Grupo A	3.712X	± 1.891	Sondas + Exones	Total	11.881X	± 7.871
	Grupo B	3.797X	± 2.952		Casos	10.589X	± 6.109
	Grupo C	2.910X	± 1.275		Controles	13.175X	± 9.140
	Grupo D	2.936X	± 2.442				
Sondas + Exones	Total	27.523X	± 16.043				
	Grupo A	29.839X	± 14.682				
	Grupo B	30.053X	± 24.455				
	Grupo C	25.386X	± 11.043				
	Grupo D	24.047X	± 18.321				

En el estudio Egabro-Pizarra, la cobertura media en las regiones cubiertas solo por las sondas fue de 46.74X mientras que en la del estudio caso/control fue de 19.24X. También hay diferencia en la cobertura si se unen las coordenadas de las sondas y los exones, siendo de 27.52X en el primer estudio y de 11.88X en el segundo. Esto se debe a la importante diferencia en los costes de NGS de estos momentos en relación a cuando se realizaron los primeros estudios y la importante limitación en los presupuestos de investigación en los proyectos que se han realizado.

2.5.- Validación de *miracleV2*

Una vez analizada la muestra NIST7035 del Consorcio GIAB con el flujo de análisis *miracleV2* se compararon los resultados de las variantes con los propios *gold standard* de la muestra original y se calculó la precisión (P) y sensibilidad (S) para evaluar la calidad de la *pipepline* desarrollada en este trabajo.

Con el análisis general se obtuvieron valores de P: 80% y S: 94%. Esto es debido a que en *miracleV2*, en el paso de identificación de variantes con el programa *Freebayes* se llama a cualquier variante con una cobertura de al menos una lectura (parámetro -C 1), ya que el objetivo es no perder ninguna posible variante por falta de cobertura. Pero, si se continúa con flujo de análisis llevado a cabo en este trabajo, se produce un filtrado de variantes mediante el programa *VCFTools* con los parámetros *minDP* de 10, el cual acepta una variante como válida si está cubierta por 10 lecturas, y el parámetro *minGQ* de 20, que establece un mínimo de calidad de 20 de genotipado para afirmar que esa variante es apta para su consideración. Con este posterior filtrado, el valor P aumentó al 97% a expensas de la sensibilidad, la cual disminuyó al 79% al descartar variantes por calidad de genotipado en la muestra analizada en comparación con la original que no es filtrada por estos parámetros.

2.6.- Variantes encontradas en el análisis de exoma por estudio

El proceso de filtrado bioinformático inicial de las variantes en este estudio ha sido el mismo que en el apartado anterior:

- Filtrado por un valor mínimo de cobertura de 10 lecturas, para tomar una variante por válida.
- Filtrado por un valor mínimo de calidad de genotipado de 20 para evitar variaciones por error de secuenciación.

Los siguientes filtros se han realizado, esta vez, con el programa GEMINI v0.20.1 (146) considerando también un genotipado mínimo en el 60% de las muestras estudiadas y la clasificación por impacto tomada de la anotación por VEP, pero en este caso simplificada a tres categorías: Alto, Medio, Bajo.

En todos los resultados de enumeración de variantes se han considerado solo aquellas variantes únicas por grupo y que, por tanto, no se han encontrado en ningún otro grupo analizado.

El análisis estadístico de asociación en población caso/control utilizando el Test de Fisher, se ha realizado sobre todas las variantes del sub-estudio en cuestión, pero solo se muestran los resultados de aquellas variantes específicas solo presentes en los grupos implicados en cada sub-estudio. Se muestran aquellas variantes con p-valores menores de 0.05 en todos los casos, y con valores de OR > 2 o valores de OR < 0.5.

En este caso también hay que tener en cuenta que no se ha utilizado ningún test de corrección por múltiples análisis. Un valor de p corregido igual a 3.55E-6 debería considerarse para decir que una variación es estadísticamente significativa.

2.6.1.- Estudio Egabro-Pizarra

Se ha llevado a cabo el análisis bioinformático de los exomas pertenecientes al estudio Egabro-Pizarra de un total de 119

individuos, clasificado en 4 grupos según la existencia o no de RI y DM2. Se han realizado dos sub-estudios caso/control según la naturaleza de los grupos enfrentados:

Estudio de evolución 1: Análisis de la aparición de DM2 sin antecedentes de RI utilizando el grupo A como controles y el grupo B como casos.

Estudio de evolución 2: Análisis de la aparición de DM2 con antecedentes de RI utilizando el grupo C como controles de DM2, pero con RI y el grupo D como casos, al haber desarrollado DM2 después del periodo de seguimiento teniendo RI en el inicio del mismo.

2.6.1.1.- Estudio de evolución 1

En este primer análisis se han tenido en cuenta los grupos A y B. Las muestras del grupo A se han tomado como controles sanos y las muestras del grupo B como casos ya que han sido diagnosticadas con DM2 al final del estudio. En la tabla 53 se muestran clasificadas por impacto y consecuencia las variantes únicas del grupo A, las únicas del grupo B y las comunes entre ambos grupos.

Tabla 54: Clasificación por impacto y consecuencia de las variantes entre el grupo A y el grupo B del estudio Egabro-Pizarra.

IMPACTO	CONSECUENCIA	Nº VARIANTES			TOTAL
		GRUPO A	GRUPO B	COMUNES	
ALTO	V. pauta de lectura	2	0	0	2
	V. aceptora de <i>splicing</i>	1	0	0	1
	V. pérdida de <i>START</i>	1	0	0	1
	V. ganancia de <i>STOP</i>	3	1	0	4
MEDIO	V. <i>missense</i>	253	73	15	341
	V. <i>splicing</i>	20	7	0	27
BAJO	V. 3' UTR	206	92	8	306
	V. 5' UTR	18	12	2	32
	V. <i>downstream</i>	1	2	0	3
	V. intrónica	92	31	3	126
	V. no codificante	17	5	1	23
	V. sinónima	147	56	9	212
	V. <i>upstream</i>	3	1	0	4
		764	280	38	1082

En total han sido 1082 variantes, de las cuales más de un 65% se clasifican con un impacto bajo, siendo las más comunes las variantes en el extremo 3'UTR y las sinónimas. Más de un 31% de las variantes totales son de tipo *missense*.

Dado el escaso número de individuos era poco probable identificar variantes con suficiente significación estadística ya que el poder estadístico no es suficiente como para dar un resultado positivo.

Si bien, algunas de las variantes identificadas podrían ser relevantes para el desarrollo de DM2. Habría que confirmarlas en estudios con mayor número de muestras.

Por otro lado, se han explorado las posibles rutas metabólicas y funciones biológicas de las que forman parte los genes donde se localizan las variantes identificadas. En total han sido 620 genes

analizados, de los cuales más de un 57% tenían, al menos, dos variantes analizadas en este estudio. Se ha utilizado la aplicación de acceso libre *PANTHER Gene List Analysis*. (150,151) para conocer los procesos biológicos (figura 42) y rutas metabólicas (figura 43) en las que están implicados los genes de cada grupo de este sub-estudio.

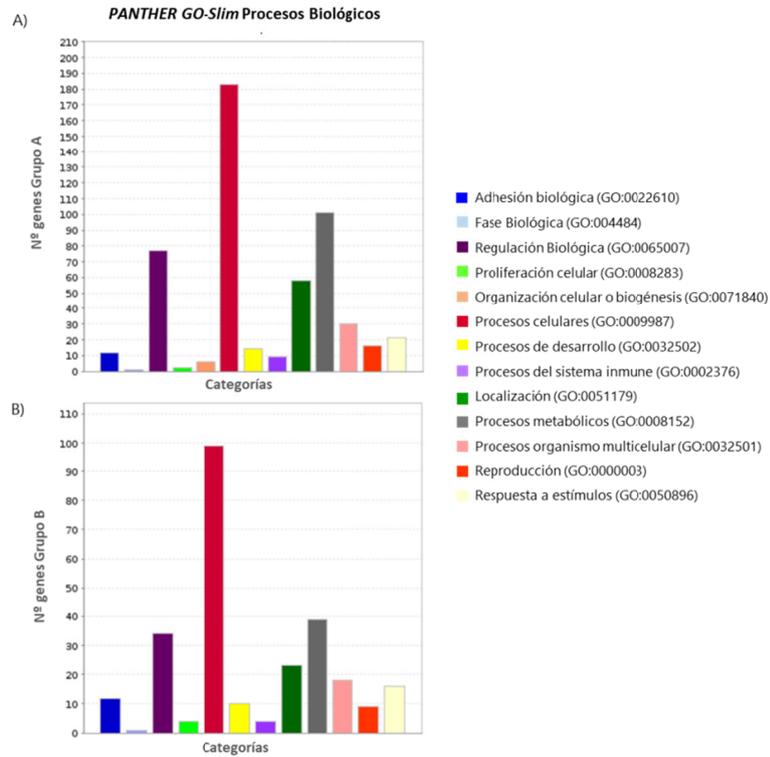


Figura 40: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos A y B mediante el programa *PANTHER Gene List Analysis*. (150,151)

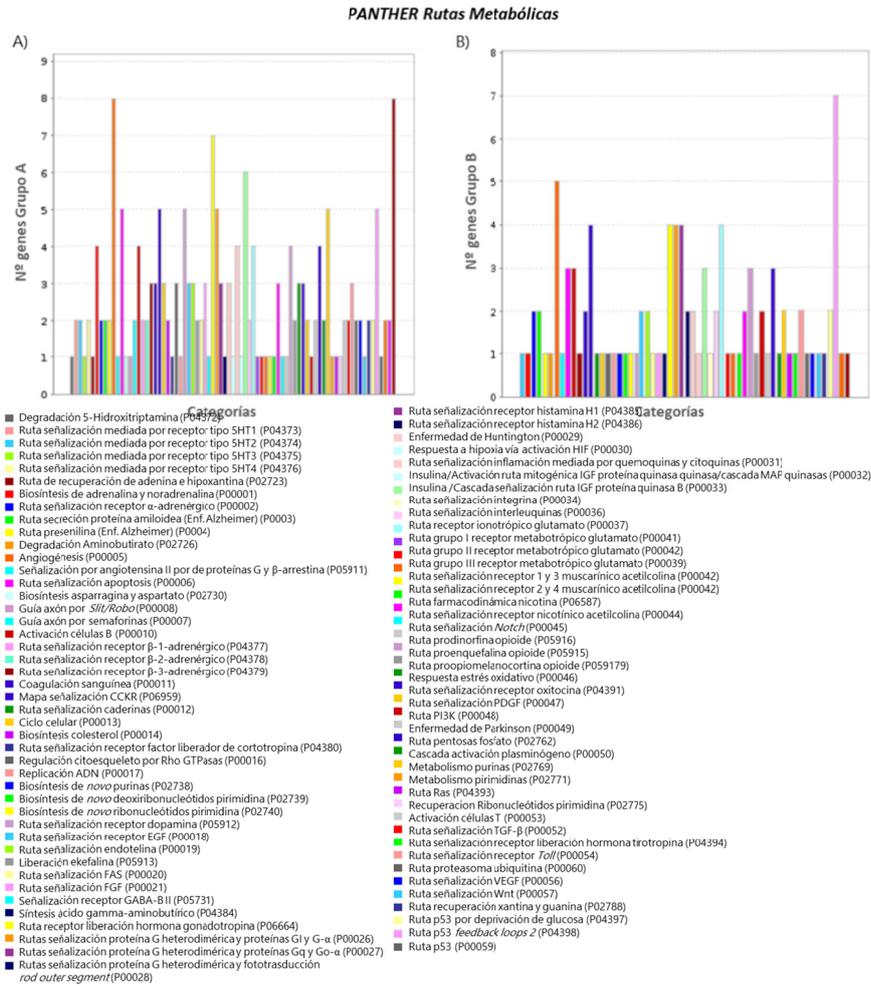


Figura 41: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos A y B mediante el programa *PANTHER Gene List Analysis*. (150,151).

Como se puede comprobar tanto en la figura 42 como en la figura 43, los genes analizados participan en multitud de rutas metabólicas de diferentes naturalezas, este hecho refuerza el carácter multifactorial característico de la DM2.

2.6.1.2.- Estudio de evolución 2

El segundo análisis en esta población ha sido utilizando los grupos C y D. El grupo C que tenía RI al inicio del estudio, pero no desarrolló DM2 al final se ha utilizado como grupo control, en cambio el grupo D si desarrolló DM2 al final del estudio estando también diagnosticados con RI al principio. En la tabla 54 se pueden ver, clasificadas por impacto y consecuencia las variantes únicas encontradas en el grupo C, las únicas del grupo D y las comunes entre ambos grupos.

Tabla 55: Clasificación por impacto y consecuencia de las variantes entre el grupo C y el grupo D del estudio Egabro-Pizarra.

IMPACTO	CONSECUENCIA	Nº VARIANTES			TOTAL
		GRUPO C	GRUPO D	COMUNES	
ALTO	V. pauta de lectura	2	0	6	8
	V. ganancia de <i>STOP</i>	2	2	0	4
	V. pérdida de <i>STOP</i>	0	1	0	1
MEDIO	V. <i>missense</i>	193	65	1	259
	V. <i>splicing</i>	12	5	4	21
BAJO	V. 3' UTR	159	48	10	217
	V. 5' UTR	15	6	0	21
	V. <i>downstream</i>	1	2	0	3
	V. intrónica	72	28	1	101
	V. no codificante	11	2	0	13
	V. sinónima	133	49	4	186
	V. <i>upstream</i>	2	2	0	4
		602	210	26	838

En total se han encontrado 838 variantes y de ellas, más de un 65% son de impacto bajo, siendo también las más frecuentes las sinónimas y las localizadas en el extremo 3'UTR. Las variantes *missense* esta vez representan un 33% del total.

Después de llevar a cabo el análisis de asociación usando el Test de Fisher, tampoco se ha encontrado ninguna variante significativa al analizar las variantes solo existentes en los grupos C y D. Debido al bajo número de participantes hace que no sea significativamente estadístico ningún resultado. Aun así, permite conocer variantes interesantes para analizar en futuros trabajos.

Sí que se ha investigado las funciones biológicas y las posibles rutas metabólicas en las que pueden estar implicados los genes cuyas variantes han sido encontradas en estos dos grupos. En total han sido 620 genes, algunos con más de dos variantes encontradas. Mediante el programa *PANTHER Gene List Analysis*. (150,151) se han podido conocer los procesos biológicos (figura 44) y las rutas (figura 45) a los que pertenecen dichos genes estudiados.

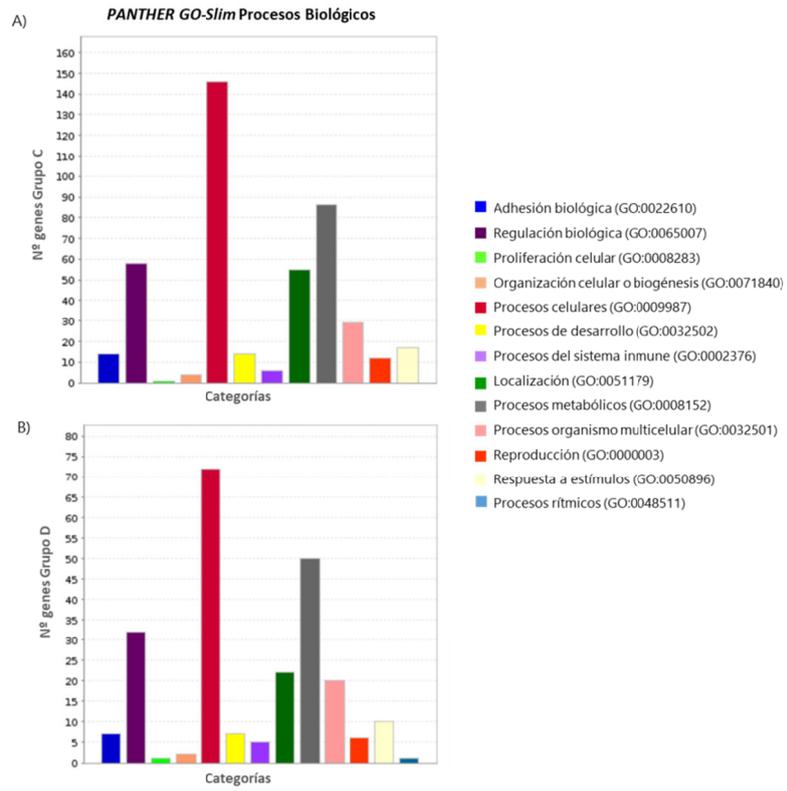


Figura 42: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos C y D mediante el programa *PANTHER Gene List Analysis*. (150,151).

PANTHER Rutas Metabólicas

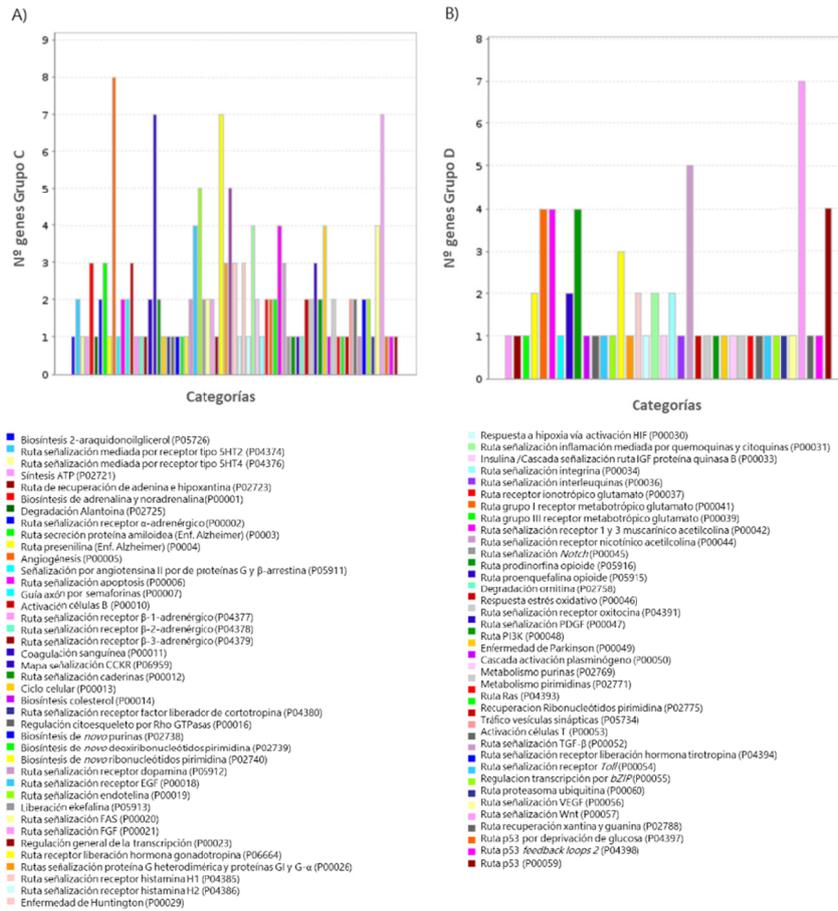


Figura 43: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos C y D mediante el programa *PANTHER Gene List Analysis*. (150,151).

2.6.2.- Estudio caso/control

Esta subpoblación se utilizó en estudios de exoma anteriores del grupo cuyos datos han servido para inicializar la validación de genes (apartado 1 de este trabajo). Hemos vuelto a reanalizar los datos ya que todo el flujo de análisis bioinformático, posteriores filtrados y análisis de variantes ha sido diferente y por lo tanto, los resultados pueden cambiar con respecto al análisis anterior.

En este subconjunto de datos, en el que hay 184 casos diagnosticados con DM2 (Grupo F) y 185 controles (Grupo E), en total 369 individuos, se han encontrado en este análisis 9333 variantes totales. En la tabla 55 se resumen clasificadas por impacto y consecuencia.

Tabla 56: Clasificación por impacto y consecuencia de las variantes de exoma del estudio caso/control.

IMPACTO	CONSECUENCIA	Nº VARIANTES			TOTAL
		GRUPO E	GRUPO F	COMUNES	
ALTO	V. pauta de lectura	4	4	1	9
	V. aceptora de <i>splicing</i>	4	4	1	9
	V. donadora de <i>splicing</i>	3	1	1	5
	V. pérdida de <i>START</i>	0	2	1	3
	V. ganancia de <i>STOP</i>	38	25	6	69
	V. pérdida de <i>STOP</i>	0	0	1	1
MEDIO	V. delección sin cambio*	0	1	2	3
	V. <i>missense</i>	1013	832	325	2170
	V. <i>splicing</i>	67	51	17	135
BAJO	V. 3' UTR	1114	878	442	2434
	V. 5' UTR	89	68	55	212
	V. <i>downstream</i>	72	55	117	244
	V. intrónica	394	349	251	994
	V. no codificante	473	355	630	1458
	V. sinónima	656	508	248	1412
	V. <i>upstream</i>	72	40	63	175
		3999	3173	2161	9333

*Original: Variante de delección sin cambio en pauta de lectura.

En este análisis, más del 74% de las variantes tienen un impacto bajo, siendo las más comunes las variantes en el extremo 3' UTR y las variantes no codificantes y sinónimas. En este caso también el tipo *missense* se encuentran en más de un 23% de la totalidad de variantes.

En el análisis de asociación de este estudio caso/control utilizando el Test de Fisher, se han obtenido en total 50 variantes significativas, 9 de ellas se han encontrado solo en uno de los dos grupos, 5 variantes en el grupo E control y 4 variantes en el grupo F con DM2. En la tabla 56 se pueden ver toda su información.

Tabla 57: Información descriptiva de las variantes significativas encontradas solo en uno de los grupos del análisis de asociación con el Test de Fisher entre el grupo E y el grupo F del estudio caso/control.

A) Clasificación por impacto y consecuencia.

IMPACTO	CONSECUENCIA	Nº VARIANTES
MEDIO	V. <i>missense</i>	1
BAJO	V. 3' UTR	1
	V. <i>downstream</i>	1
	V. intrónica	5
	V. sinónima	1

9

B) Información de las variantes y resultados estadísticos

cr:posición	Alelos	SNP	GEN	Cambio	Consecuencia	CADD	TEST	OR (IC 95%)	P
1:14677	G/A	rs201327123	<i>DDX11L1</i>	g.14677G>A	V. <i>downstream</i>	7.35	TREND	NA	0.019
2:8743427	C/T	rs116414452	<i>AC011747.6</i>	g.8743427C>T	V. intrónica	11.20	TREND	NA	0.020
8:17156603	G/A	rs117299491	<i>MTMR7</i>	g.17156603G>A	V. 3' UTR	5.80	TREND	NA	0.022
14:24709506	C/T	rs184422577	<i>TINF2</i>	p.Leu364=	V. sinónima	12.99	TREND	NA	0.030
18:2945860	C/A	rs73936873	<i>LPIN2</i>	g.2945860C>A	V. intrónica	8.71	TREND	NA	0.027
21:15436787	G/A	rs557353737	<i>AP001347.6</i>	g.15436787G>A	V. intrónica	1.22	TREND	NA	0.029
22:19034627	G/A	rs1051243	<i>DGCR2</i>	g.19034627G>A	V. intrónica	2.36	TREND	NA	0.013
X:1554024	G/A	rs45447192	<i>ASMTL</i>	g.1554024G>A	V. intrónica	8.88	TREND	NA	0.012
MT:7805	G/A	None	<i>MT-CO2</i>	p.Val74Ile	V. <i>missense</i>	0.01	ALLELIC	NA	0.037

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

C) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(E)	MAF(F)	GRUPO E				GRUPO F			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
1:14677	0.0096	0.0214	0	134	6	0	140	143	0	0	143
2:8743427	0.0084	0	0.0184	159	0	0	159	131	5	0	136
8:17156603	0.0095	0.0204	0	141	6	0	147	148	0	0	148
14:24709506	0.0057	0	0.0157	160	0	0	160	154	5	0	159
18:2945860	0.0074	0	0.0151	174	0	0	174	161	5	0	166
21:15436787	0.0085	0	0.0174	148	0	0	148	139	5	0	144
22:19034627	0.0102	0.0194	0	142	6	0	155	139	0	0	139
X:1554024	0.0080	0.0213	0	135	6	0	141	119	0	0	119
MT:7805	0.0102	0.0291	0	100	0	3	103	80	0	0	80

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

Dentro de este grupo, encontramos el gen *MTMR7*, miembro de la familia de miotubularina lípido fosfatasas, inhibidor de la señalización de insulina (221). Se ha visto que, en pacientes con cáncer colonorectal (CCR) y además DM2, la hiperinsulinemia establecida y el aumento de los niveles del factor de crecimiento, inhiben su acción permitiendo el crecimiento del tumor. Constituye un marcador potencial en este tipo de cáncer (222). Por otro lado, el gen *DGCR2* el cual es un receptor de adhesión que se expresa durante la embriogénesis (223) se ha descubierto recientemente en la CBPs secretoras de insulina como uno de sus biomarcadores exclusivos (224).

Los demás genes y/o variantes no se encuentran en la bibliografía relacionados con DM2.

El resto de variantes significativas son 41 en total. De entre ellas, 6 variantes tienen valores de OR superiores a 2 siendo consideradas con efecto de riesgo potencial para la DM2. Por otro lado 8 variantes poseen valores de OR por debajo de 0.5 y se les podría considerar un efecto más bien protector frente a la DM2. Las 27 variantes significativas restantes tienen valores intermedios de OR, por lo que su posible efecto es más ambiguo. En la tabla 57 se puede ver la información de las 14 variantes significativas encontradas en ambos grupos con valores de OR <0.5 y OR >2.

Tabla 58: Información descriptiva de las variantes significativas del análisis de asociación con el Test de Fisher entre el grupo E y el grupo F del estudio caso/control.

A) Clasificación por impacto y consecuencia.

IMPACTO	CONSECUENCIA	Nº VARIANTES
MEDIO	V. missense	2
	V. splicing	1
BAJO	V. downstream	3
	V. intrónica	2
	V. no codificante	6

14

B) Información de las variantes y resultados estadísticos

cr:posición	Alelos	SNP	Gen	Cambio	Consecuencia	CADD	TEST	OR (IC 95%)	P
3:659806	A/G	rs149448530	AC090044.1	g.659805A>G	V. intrónica	6.89	TREND	0.422(0.182-0.980)	4.84E-02
8:11990766	G/A	rs184094604	USP17L2	g.11990765G>A	V. downstream	2.78	TREND	2.454(1.187-5.073)	1.57E-02
8:11991210	C/T	rs117041743	USP17L2	g.11991209C>T	V. downstream	5.80	DOM	2.049(1.030-4.078)	4.46E-02
8:11991419	G/A	rs56195358	USP17L2	g.11991418G>A	V. downstream	0.32	TREND	2.442(1.181-5.051)	1.58E-02
10:131636	T/C	rs4508132	IL9RP2	g.131635T>C	V. splicing	0.46	DOM	0.454(0.276-0.746)	1.74E-03
10:126402	T/C	rs1984905	IL9RP2	g.126401T>C	V. no codificante	1.11	DOM	0.377(0.226-0.629)	1.14E-04
12:5141881	G/A	rs16932783	RP11-429A20.2	g.5141880G>A	V. no codificante	6.32	ALLELIC	3.131(1.125-8.714)	2.38E-02
14:20158095	G/T	rs764147054	ARHGAP42P4	g.20158094G>T	V. no codificante	1.90	TREND	0.197(0.043-0.896)	2.35E-02
15:20876452	G/A	rs6422239	NBEAP1	g.20876451G>A	V. no codificante	10.66	TREND	0.218(0.062-0.762)	1.41E-02
18:14537970	C/T	rs148283099	POTEC	p.Val214Ile	V. missense	11.01	ALLELIC	0.396(0.180-0.870)	1.86E-02
22:16365578	A/G	rs773311848	NEK2P2	g.16365577A>G	V. no codificante	1.70	TREND	7.255(0.888-59.290)	3.65E-02
X:6451932	C/G	rs200714581	VXC3A	p.Glu139Gln	V. missense	0.02	TREND	7.498(0.917-61.330)	3.39E-02
X:74965306	C/T	rs146563627	TTC3P1	g.74965305C>T	V. no codificante	0.00	ALLELIC	0.161(0.036-0.720)	7.40E-03
X:85416170	C/A	rs10126219	DACH2	g.85416169C>A	V. intrónica	0.00	ALLELIC	0.431(0.208-0.892)	2.47E-02

Cr: cromosoma. CADD: Combined Annotation Dependent Depletion value. SNP: Single Nucleotide Polymorphism. OR: Odds Ratio. IC: Intervalo de confianza. P: p-valor Test de Fisher.

C) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(E)	MAF(F)	GRUPO E				GRUPO F			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
3:659806	4.22E-02	0.0669	0.0310	144	19	0	142	149	8	0	129
8:11990766	5.39E-02	0.0357	0.0909	158	11	0	154	139	24	1	143
8:11991210	6.38E-02	0.0417	0.0884	158	11	1	156	141	24	1	147
8:11991419	6.22E-02	0.0377	0.0949	154	11	0	146	136	24	1	137
10:131636	8.72E-01	0.9962	1.1681	5	43	110	132	3	18	130	119
10:126402	8.81E-01	0.9211	0.9414	5	46	117	152	2	19	143	162
12:5141881	2.94E-02	0.0158	0.0490	166	5	0	158	155	13	1	153
14:20158095	1.99E-02	0.0357	0.0084	158	11	0	154	150	2	0	119
15:20876452	8.38E-03	0.0625	0.0185	137	15	0	120	130	3	0	81
18:14537970	7.97E-02	0.0737	0.0328	149	19	2	156	152	9	0	137
22:16365578	1.17E-02	0.0031	0.0229	171	1	0	160	162	7	0	153
X:6451932	1.43E-03	0.0042	0.0333	151	1	0	120	138	7	0	105
X:74965306	2.28E-02	0.0411	0.0076	162	5	4	158	157	0	1	131
X:85416170	6.09E-02	0.1048	0.0478	134	14	6	124	140	9	1	115

Cr: cromosoma. MAF: Minority Allele Frequency. EUR: Europa.

Estos genes y/o variantes no se han relacionado por ahora con DM2.

Se ha analizado, mediante el programa *PANTHER Gene List Analysis*. (150,151) las funciones biológicas y las posibles rutas metabólicas en las que pueden estar implicados los genes de este estudio. En total han sido 1693, algunos de ellos con dos o más variantes encontradas. En la figura 46 se muestran los procesos biológicos y en la figura 47 las rutas metabólicas.

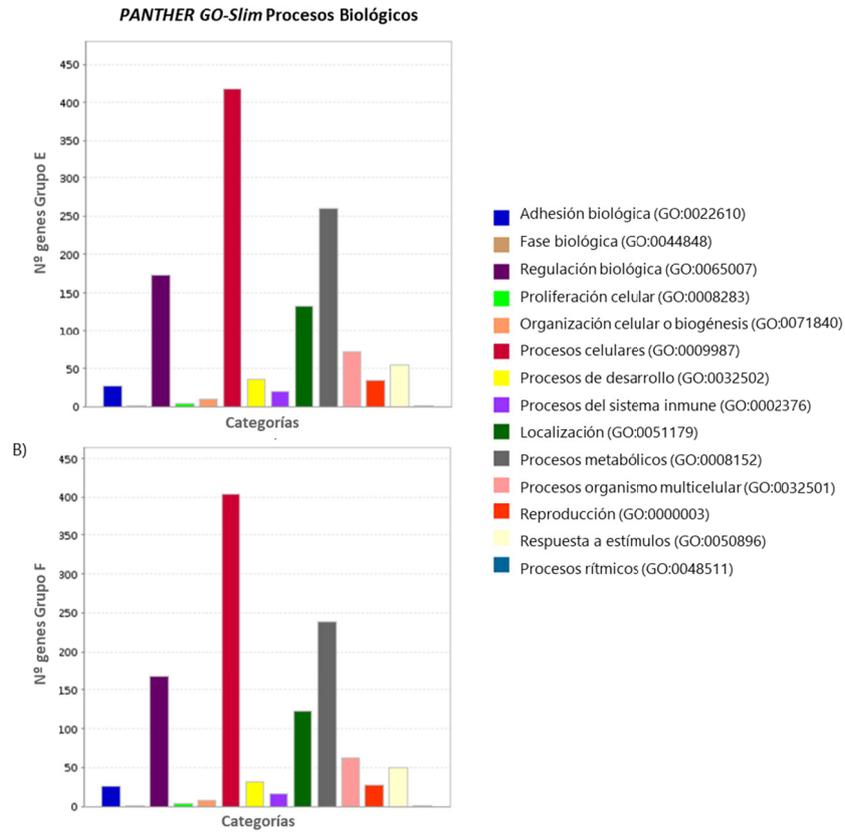


Figura 44: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos E y F mediante el programa PANTHER Gene List Analysis. (Mi et al. 2017, 2019).

PANTHER Rutas Metabólicas

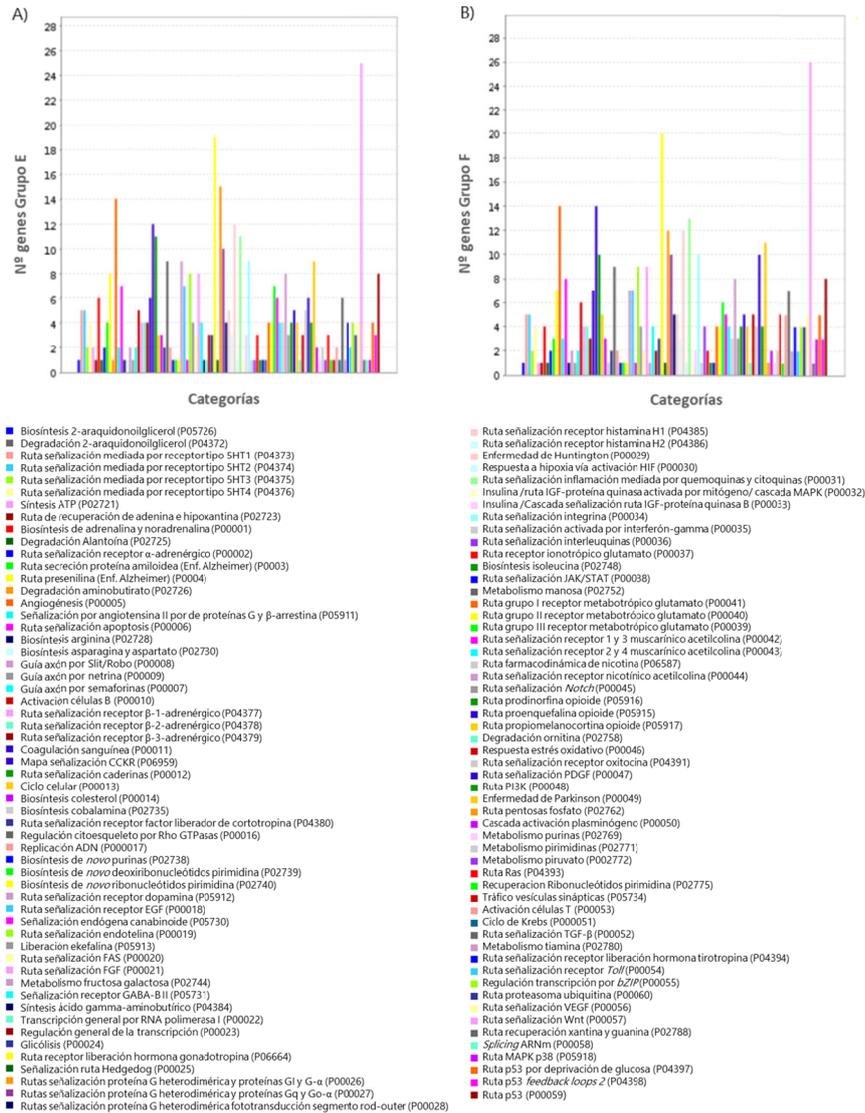


Figura 45: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos E y F mediante el programa PANTHER Gene List Analysis. (Mi et al. 2017, 2019).

2.6.3.- Estudio caso/control conjunto

En el análisis de la unión de todas las muestras de exoma disponibles, menos el grupo C, por tener RI, quitando las variantes encontradas en el análisis caso/control anterior, se han encontrado un total de 1544 variantes. Se clasifican en la tabla 58.

Tabla 59: Clasificación por impacto y consecuencia de las variantes de exoma del estudio caso/control conjunto.

IMPACTO	CONSECUENCIA	Nº VARIANTES			TOTAL
		CONTROLES	CASOS	COMUNES	
ALTO	V. pauta de lectura	0	1	1	2
	V. donadora de <i>splicing</i>	0	0	1	1
	V. ganancia de <i>STOP</i>	0	0	4	4
MEDIO	V. <i>missense</i>	64	34	278	376
	V. <i>splicing</i>	5	2	20	27
BAJO	V. 3' UTR	61	37	352	450
	V. 5'UTR	11	1	38	50
	V. <i>downstream</i>	3	0	17	20
	V. intrónica	38	11	138	187
	V. no codificante	8	4	106	118
	V. sinónima	51	29	207	287
	V. <i>upstream</i>	1	0	21	22
		242	119	1183	1544

Más de un 73% son variantes de impacto bajo, en su mayoría en la región 3' UTR, sinónimas, intrónicas y no codificantes. En más de un 24% están representadas las variantes *missense*.

Después de realizar el estudio de asociación con el Test de Fisher se han encontrado un total de 27 variantes significativas. Que solo hayan aparecido en uno de los grupos, han sido un total de 4 variantes, 1 en el grupo de controles (AE) y 3 en el grupo de casos (BDF). En la tabla 59 están los resultados.

Tabla 60: Información descriptiva de las variantes significativas encontradas solo en uno de los grupos del análisis de asociación con el Test de Fisher entre el grupo control (AE) y el grupo de casos (BDF) del estudio caso/control conjunto.

A) Clasificación por impacto y consecuencia.

IMPACTO	CONSECUENCIA	Nº VARIANTES
BAJO	V. 3' UTR	1
	V. intrónica	1
	V. no codificante	1
	V. sinónima	1

4

B) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	GEN	Cambio	Consecuencia	CADD	TEST	OR (IC 95%)	P
5:917697	T/G	rs115313592	TRIP13	g.917697T>G	V. 3' UTR	4.15	TREND	NA	0.021
6:12514623	T/C	rs148225566	RPL15P3	g.12514623T>C	V. no codificante	14.53	TREND	NA	0.009
14:20813642	G/A	rs3093889	PARP2	g.20813642G>A	V. intrónica	0.13	TREND	NA	0.016
14:20823095	G/A	rs3093927	PARP2	Pro310=	V. sinónima	12.92	TREND	NA	0.018

Cr: cromosoma. CADD: *Combined Annotation Dependent Depletion value*. SNP: *Single Nucleotide Polymorphism*. OR: *Odds Ratio*. IC: Intervalo de confianza. P: p-valor Test de Fisher.

C) Distribución de las variantes en la población estudiada.

cr:posición	MAF(EUR)	MAF(AE)	MAF(BDF)	GRUPO AE (controles)				GRUPO F (casos)			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
5:917697	0.006	0	0.0167	158	0	0	158	116	4	0	120
6:12514623	0.010	0.0190	0	177	7	0	184	174	0	0	174
14:20813642	0.008	0	0.0200	143	0	0	143	120	5	0	125
14:20823095	0.008	0	0.0177	156	0	0	156	136	5	0	141

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

El gen *PARP2* (Poli (ADP-ribosa) polimerasa), que está implicado en la señalización de los procesos de daño y reparación del ADN (225), se ha visto que algunas mutaciones en este gen son propias de los pacientes diabéticos con adenocarcinomas pancreáticos ductales (226).

Aparte, se han encontrado 22 variantes significativas que aparecen tanto en casos como en controles. 11 de ellas tienen valores de OR >2 indicando un posible efecto de riesgo, 8 variantes poseen un valor de OR < 0.5 lo cual podría conferirles un cierto efecto protector. Las 3

restantes tienen valores intermedios de OR. En la tabla 60 se pueden los resultados.

Tabla 61: Información descriptiva de las variantes significativas del análisis de asociación con el Test de Fisher entre el grupo AE y el grupo BDF del estudio caso/control conjunto.

A) Clasificación por impacto y consecuencia.

IMPACTO	CONSECUENCIA	Nº VARIANTES
MEDIO	V. missense	5
BAJO	V. splicing	1
	V. 3' UTR	4
	V. intrónica	2
	V. no codificante	5
	V. sinónima	2
	V. upstream	3

22

B) Información de las variantes y resultados estadísticos.

cr:posición	Alelos	SNP	GEN	Cambio	Consecuencia	CADD	TEST	OR (IC 95%)	P
5:436443	T/A	rs796509582	AHRR	g.436443T>A	V. 3' UTR	2.92	TREND	3.730(1.376-10.11)	4.87E-03
5:16754930	A/G	rs72732223	MYO10	g.16754930A>G	V. splicing	12.56	TREND	2.796(1.212-6.453)	1.02E-02
5:20305371	T/C	rs186763609	CDH18	g.20305371T>C	V. intrónica	0.04	TREND	3.657(0.999-13.390)	3.48E-02
6:7517239	G/A	rs6597269	RP11-288G3.4	g.7517239G>A	V. no codificante	2.37	TREND	0.578(0.390-0.857)	5.98E-03
6:16279097	T/C	rs739473	GMPR	p.His210=	V. sinónima	0.01	TREND	0.124(0.015-0.996)	1.89E-02
7:4804952	C/T	rs565940809	FOXP1	g.4804952C>T	V. 3' UTR	2.71	TREND	8.934(1.112-71.810)	1.24E-02
7:9767725	G/T	rs72583761	AC060834.2	g.9767725G>T	V. no codificante	4.37	ALLELIC	0.406(0.167-0.985)	4.29E-02
8:623906	C/T	rs118014343	ERICH1	p.Arg149Gln	V. missense	1.99	TREND	0.201(0.044-0.917)	2.06E-02
8:19221664	G/C	rs877386	SH2D4A	p.Gly263Ala	V. missense	0.00	TREND	3.226(1.149-9.054)	1.74E-02
9:990235	C/T	rs61737966	DMRT3	p.Arg217Cys	V. missense	14.61	TREND	6.074(1.318-27.990)	7.94E-03
10:94545	C/T	rs10904047	TUBB8	g.94545C>T	V. intrónica	0.65	ALLELIC	0.462(0.291-0.732)	8.22E-04
11:1028379	C/T	rs116256283	MUC6	p.Gly534Ser	V. missense	24.90	TREND	9.583(1.207-76.080)	8.55E-03
13:21353626	C/CA	rs140765614	XPO4	g.21353626C>A	V. 3' UTR	None	TREND	0.229(0.050-1.053)	3.72E-02
14:19411980	T/C	rs141267552	RP11-536C10.7	g.19411980T>C	V. upstream	0.82	TREND	0.207(0.045-0.943)	2.32E-02
15:22692646	G/A	rs551215980	ABC810P1	g.22692646G>A	V. no codificante	5.55	TREND	3.625(1.171-11.220)	1.59E-02
16:817407	C/T	rs72773454	MSLN	p.Leu519=	V. sinónima	10.67	TREND	3.086(1.098-8.676)	2.27E-02
16:1398934	G/A	rs116986044	BIAIP3	g.1398934G>A	V. 3' UTR	5.01	TREND	0.105(0.013-0.823)	8.31E-03
18:14478894	T/A	rs72883372	CXADR3	g.14478894T>A	V. no codificante	3.40	ALLELIC	0.392(0.192-0.799)	8.30E-03
22:21363341	T/C	rs2516541	AC002472.1	g.21363341T>C	V. upstream	0.01	TREND	1.772(1.031-3.046)	3.39E-02
22:21363505	C/T	rs2516542	AC002472.1	g.21363505C>T	V. upstream	4.75	TREND	1.884(1.084-3.272)	2.11E-02
X:3228966	C/G	rs148675322	MXRA5	p.Glu2426Asp	V. missense	20.30	ALLELIC	9.722(1.224-77.200)	9.46E-03
X:101034869	C/T	rs4145243	RP1-232L22 B.1	g.101034869C>T	V. no codificante	15.82	ALLELIC	2.734(1.353-5.526)	4.13E-03

Cr: cromosoma. CADD: Combined Annotation Dependent Depletion value. SNP: Single Nucleotide Polymorphism. OR: Odds Ratio. IC: Intervalo de confianza. P: p-valor Test de Fisher.

C) Distribución de las variantes en la población estudiada..

RESULTADOS Y DISCUSIÓN

cr.posición	MAF(EUR)	MAF(AE)	MAF(BDF)	GRUPO AE (controles)				GRUPO BDF (casos)			
				0/0	0/1	1/1	Total	0/0	0/1	1/1	Total
5:436443	0.0339	0.0215	0.0636	158	5	0	163	153	20	0	173
5:16754930	0.0418	0.0323	0.0759	147	8	0	155	125	20	0	145
5:20305371	0.0160	0.0125	0.0323	197	3	0	200	176	10	0	186
6:7517239	0.1976	0.2182	0.1531	95	60	10	165	113	44	3	160
6:16279097	0.0124	0.0303	0.0092	157	8	0	165	162	1	0	163
7:4804952	0.0112	0.0080	0.0291	187	1	0	188	164	8	0	172
7:9767725	0.0375	0.0545	0.0292	149	14	2	165	147	7	0	154
8:623906	0.0181	0.0371	0.0131	164	11	0	175	151	2	0	153
8:19221664	0.0244	0.0185	0.0473	184	5	0	189	155	14	0	169
9:990235	0.0189	0.0135	0.0472	146	2	0	148	117	10	0	127
10:94545	0.8589	0.4647	0.4816	6	40	116	170	2	20	135	163
11:1028379	0.0136	0.0090	0.0344	165	1	0	166	151	9	0	160
13:21353626	0.0181	0.0364	0.0143	155	10	0	165	138	2	0	140
14:19411980	0.0221	0.0411	0.0148	147	11	0	158	133	2	0	135
15:22692646	0.0201	0.0149	0.0405	197	4	0	201	172	13	0	185
16:817407	0.0271	0.0219	0.0537	155	5	0	160	135	14	0	149
16:1398934	0.0148	0.0351	0.0094	161	10	0	171	158	1	0	159
18:14478894	0.0519	0.0714	0.0359	164	22	3	189	170	11	0	181
22:21363341	0.0916	0.0744	0.1152	145	23	0	168	129	34	2	165
22:21363505	0.0905	0.0723	0.1178	144	22	0	166	122	33	2	157
X:3228966	0.0140	0.0092	0.0290	162	1	0	163	148	5	2	155
X:101034869	0.0651	0.0331	0.0786	157	6	3	166	120	12	8	140

Cr: cromosoma. MAF: *Minority Allele Frequency*. EUR: Europa.

El gen *MYO10* codifica para la miosina X la cual es una proteína motora basada en actina que tiene papel en la integración de la F-actina y los microtúbulos del citoesqueleto durante el proceso de meiosis. Se ha visto que aumenta su expresión en pacientes africano-americanos con obesidad mórbida y DM2 (227). Por otro lado, en el estudio de GWAS de cuatro poblaciones caso/control para la DM2 (*DiaGen Consortium*) se vio que la siguiente asociación más fuerte con la DM2, después del gen por excelencia *TCF7L2* fue el gen *MYO10*. Tres polimorfismos (rs31313, rs173728, y rs253336) fueron significativamente asociados (228). Esto lo convierte en un gen interesante para futuros análisis.

El gen *CDH18* codifica una caderina clásica tipo II que participa en el proceso de adhesión de caderinas célula-célula calcio dependiente. Ha sido asociado con GDM en un estudio GWAS en mujeres chinas

(229). Otro estudio de GWAS mostró que este gen, y *MYO10*, ambos relacionados entre sí por su papel en la adhesión de las células y movilidad celular, están asociados con el síndrome metabólico, obesidad y DM2 (230).

El gen *FOXK1* codifica un factor de transcripción que se trasloca al núcleo desde el citoplasma debido a la señalización de la insulina en la célula, juega por lo tanto un papel regulador en esa transducción de señal y está relacionado con el control de crecimiento celular, el proceso de β -oxidación mitocondrial y procesos metabólicos (231).

En este estudio también se han analizado las funciones biológicas (figura 48) y las rutas metabólicas (figura 49) en las que pueden estar implicados los genes analizados.

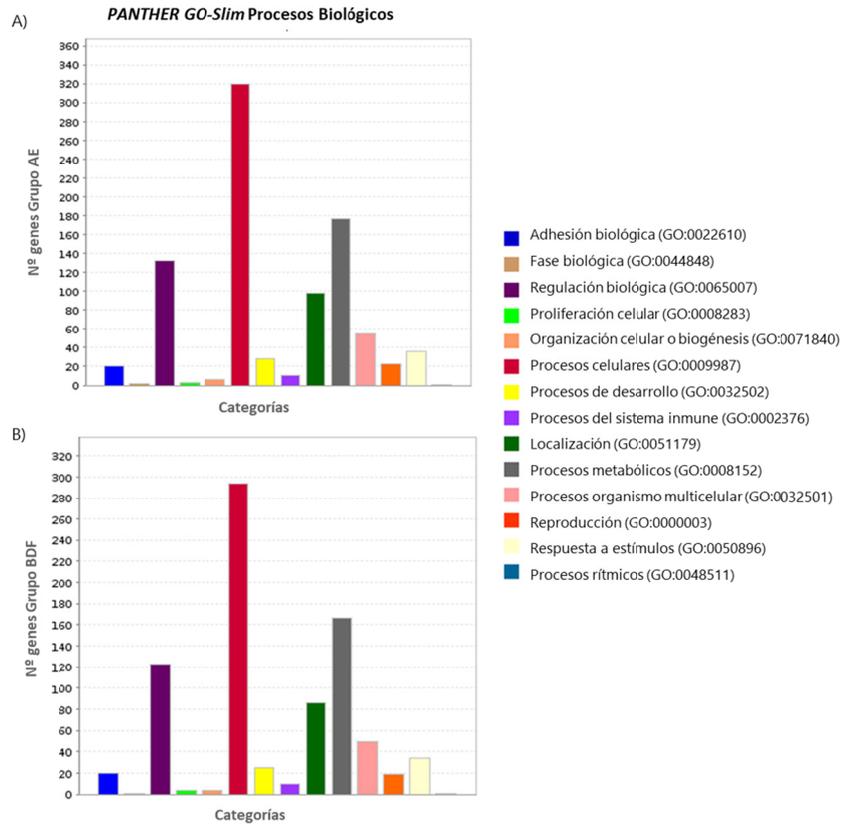


Figura 46: Resultados del análisis de procesos biológicos en los que están implicados los genes encontrados en los grupos AE (controles) y BDF (casos) mediante el programa PANTHER Gene List Analysis. (Mi et al. 2017, 2019).

PANTHER Rutas Metabólicas

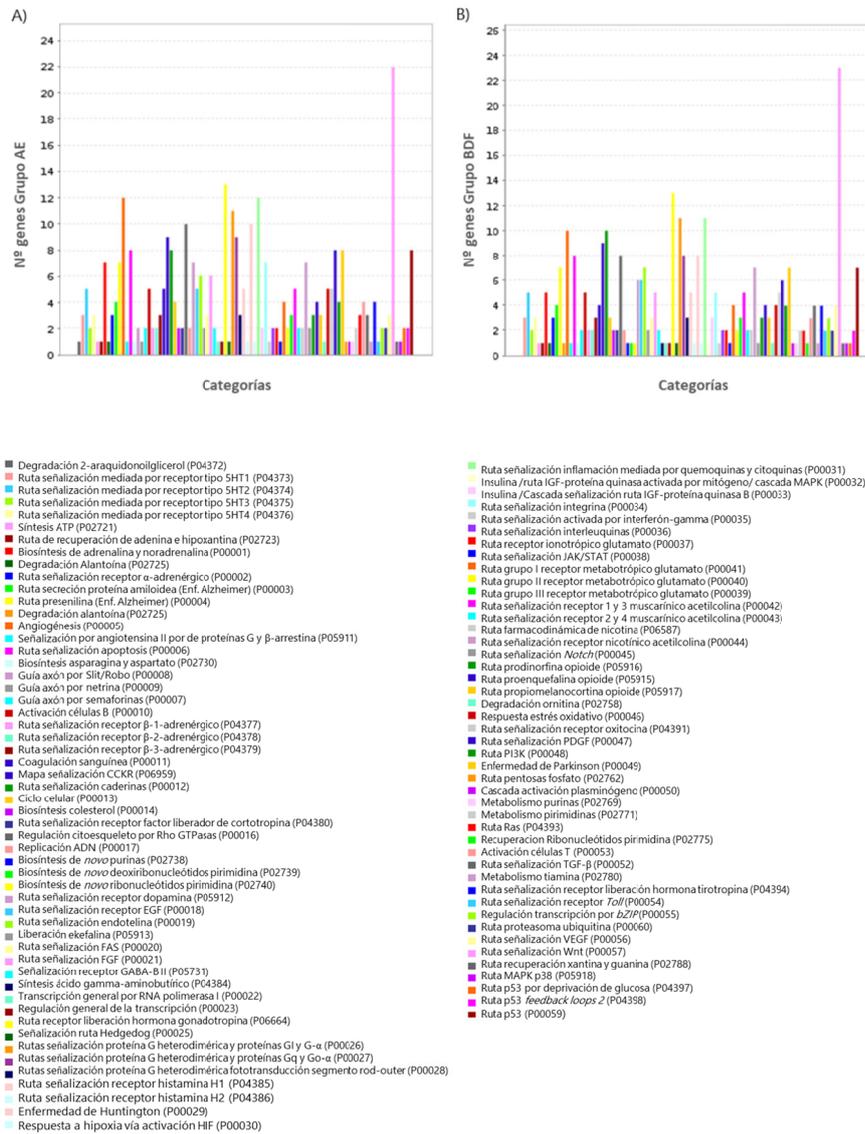


Figura 47: Resultados del análisis de las rutas metabólicas en las que están implicados los genes encontrados en los grupos AE (controles) y BDF (casos) mediante el programa PANTHER Gene List Analysis. (Mi et al. 2017, 2019).

2.6.4.- Resumen resultados

En la tabla c se pueden ver los genes del análisis de exoma del estudio caso/control (E-F) y del estudio caso/control conjunto en los que se han encontrado variantes significativas asociadas a DM2.

Tabla 62: Listado de genes con asociación estadísticamente significativa con DM2 del estudio caso/control (E-F) y del estudio caso/control conjunto.

Genes con asociación con DM2	
Estudio E-F	Estudio Conjunto
AC011747.6	ABCB10P1
AC090044.1	AC002472.1
AP001347.6	AC060834.2
ARHGAP42P4	AHRR
ASMTL	BAIAP3
DACH2	CDH18
DDX11L1	CXADRP3
DGCR2	DMRT3
IL9RP2	ERICH1
LPIN2	FO XK1
MT-CO2	GMPR
MTMR7	MSLN
NBEAP1	MUC6
NEK2P2	MXRA5
POTEC	MYO10
RP11-429A20.2	PARP2
TINF2	RP11-288G3.4
TTC3P1	RP11-536C10.7
USP17L2	RP1-232L22_B.1
VCX3A	RPL15P3
	SH2D4A
	TRIP13
	TUBB8
	XPO4

3.- Limitaciones y perspectivas futuras

Las principales limitaciones en el presente estudio han sido varias, se ha contado con un presupuesto limitado, el cual ha condicionado el número de individuos totales analizados, la cobertura media que se ha podido alcanzar, así como el tipo de protocolos comerciales que se han utilizado. Los recursos disponibles en el centro de trabajo han restringido en cierta medida las condiciones del flujo de análisis utilizado, lo cual también, limita la cantidad de variaciones genéticas finales y su cobertura en todas las muestras secuenciadas.

A pesar de ello, la finalidad de este trabajo de poder identificar variantes genéticas en cierto modo fiables que puedan relacionarse con DM2, se ha conseguido, y este hecho permitirá diseñar nuevos proyectos para seguir verificando resultados y conseguir un poder estadístico lo bastante elevado con el fin de confirmar o desmentir resultados previamente establecidos.

La necesidad de realizar este tipo de trabajos en poblaciones de mayor tamaño es sustancial debido a la baja frecuencia de las variantes genéticas, aunque factores económicos y temporales estén en contra, es la mejor vía para poder llevar a cabo su identificación y su asociación con la DM2.

Los siguientes trabajos para la verificación de variantes se esperan realizar en un gran número de muestras (hasta 15000) procedentes de varios estudios poblacionales a los que tiene acceso la UGDG. Son muestras con buena fenotipación clínica, entre ellos está el estudio Di@bet.es (232) o el Estudio Hortega (233). Se pretende disminuir la

limitación de resultados relacionada con el número de individuos analizados. Además, se están llevando a cabo mejoras en los procedimientos de generación de librerías de NGS para implementar la cobertura de las regiones y aumentar los datos a analizar.

CONCLUSIONES

1. Los estudios de exoma permiten seleccionar genes candidatos para conocer su posible relación con enfermedades, en este caso DM2, mediante estudios de asociación en poblaciones caso/control.
2. De 19 genes seleccionados en estudio previo de exoma, en 17 de ellos se han encontrado, en conjunto, 11 variantes significativas relacionadas con DM2 en la población caso/control estudiada, localizadas en los genes *ACAD10*, *ACSL3*, *AFMID*, *CCNDBP1*, *GCKR*, *OLFML1*, *SLC12A4*, *SLC15A3* y *SREBF2*.
3. Los pasos de diseño, puesta a punto y posterior automatización de flujos de análisis bioinformáticos, que se han llevado a cabo en este trabajo, permiten investigar y probar, durante su desarrollo, diferentes opciones de análisis hasta establecer uno fiable, rápido y duradero para futuros estudios relacionados.
4. El aumento de individuos en los estudios de exoma permiten una mayor confianza en la identificación de variantes genéticas asociadas, en este caso a DM2. Aún así, se necesita un número mayor para conseguir mejor poder estadístico.
5. Se han podido identificar 12708 variantes genéticas en los 488 exomas estudiados de las cuales 49 variantes han sido asociadas significativamente a DM2 en las comparaciones caso/control analizadas. Estos datos deben ser validados para demostrar su posible asociación e implicación en la DM2.

CONCLUSIONES

6. Todos los estudios realizados en el presente trabajo indican que hay numerosos genes y variantes raras de los mismos que pueden estar asociados a la DM2 en nuestra población y que pueden explicar parte del componente genético de la misma.

REFERENCIAS

REFERENCIAS

1. Alberti KGMM, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO Consultation. *Diabet Med.* 1998 Jul
2. American Diabetes Association AD. Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 2013 Jan 36(Supl1):S67-74.
3. Kharroubi AT. Diabetes mellitus: The epidemic of the century. *World J Diabetes.* 2015;6(6):850.
4. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract.* 2018;138:271-81.
5. Vijayakumar V, Mavathur R, Sharma MNK. Ethnic Disparity and Increased Prevalence of Type 2 Diabetes Among South Asians: Aetiology and Future Implications for Diabetes Prevention and Management. *Curr Diabetes Rev.* 2018 Oct 10;14(6):518-22.
6. Pradeepa R, Mohan V. Prevalence of type 2 diabetes and its complications in India and economic costs to the nation. *Eur J Clin Nutr.* 2017 Jul 19;71(7):816-24.
7. Soriguer F, Goday A, Bosch-Comas A, Bordiú E, Calle-Pascual A, Carmena R, et al. Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. *Diabetologia.* 2012 Jan;55(1):88-93.
8. World Health Organization. definition and diagnosis of diabetes mellitus and intermediate hyperglycemia Report of a WHO/IDF Consultation. 2006.
9. International Diabetes Foundation. WINGS GDM Model of Care. 2015;20.
10. Mata-Cases M, Artola S, Escalada J, Ezkurra-Loyola P, Ferrer-García JC, Fornos JA, et al. Consenso sobre la detección y el manejo de la prediabetes. Grupo de Trabajo de Consensos y Guías Clínicas de la Sociedad Española de Diabetes. *Rev Clínica Española.* 2015 Mar 1

- ;215(2):117–29.
11. Kahanovitz L, Sluss PM, Russell SJ. Type 1 Diabetes-A Clinical Perspective. *Point Care*. 2017 Mar;16(1):37-40.
 12. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes 2019. *Diabetes Care*. 2019;42(Suppl. 1).
 13. Laugesen E, Østergaard JA, Leslie RDG. Latent autoimmune diabetes of the adult: current knowledge and uncertainty. *Diabet Med*. 2015;32:843–52.
 14. Ramu D, Perumal V, Paul SFD. Association of common type 1 and type 2 diabetes gene variants with latent autoimmune diabetes in adults: A meta-analysis. *J Diabetes*. 2019 Jun 12;11(6):484–96.
 15. Carlsson S. Etiology and Pathogenesis of Latent Autoimmune Diabetes in Adults (LADA) Compared to Type 2 Diabetes. *Front Physiol*. 2019;10:320.
 16. Dickens LT, Thomas CC. Updates in Gestational Diabetes Prevalence, Treatment, and Health Policy. *Curr Diab Rep*. 2019 Jun 9;19(6):33.
 17. Temple IK, Shield JP. Transient neonatal diabetes, a disorder of imprinting. *J Med Genet*. 2002 Dec;39(12):872-5.
 18. Slingerland AS, Hattersley AT. Mutations in the Kir6.2 subunit of the K_{ATP} channel and permanent neonatal diabetes: New insights and new treatment. *Ann Med* . 2005 May 8;37(3):186–95.
 19. Greeley SAW, Letourneau LR, Philipson LH. Precision medicine in KCNJ11 permanent neonatal diabetes. *Lancet Diabetes Endocrinol*. 2018 Aug 1;6(8):594–5.
 20. Olsson C, Johnsen E, Nilsson M, Wilander E, Ènen A-CS, Lagerstro M, et al. The level of the mitochondrial mutation A3243G decreases upon ageing in epithelial cells from individuals with diabetes and deafness. *Eur J Hum Genet*. 2001 Dec;9(12):917-21.
 21. De Andrade PBM, Rubi B, Frigerio F, van den Ouweland JMW, Maassen JA, Maechler P. Diabetes-associated mitochondrial DNA mutation

REFERENCIAS

- A3243G impairs cellular metabolic pathways necessary for beta cell function. *Diabetologia*. 2006 Aug 31;49(8):1816–26.
22. Zhang J, Guo J, Fang W, Jun Q, Shi K. Clinical features of MELAS and its relation with A3243G gene point mutation. *Int J Clin Exp Pathol*. 2015;8(10):13411–5.
 23. McDonald TJ, Ellard S. Maturity onset diabetes of the young: identification and diagnosis. *Ann Clin Biochem An Int J Biochem Lab Med*. 2013 Sep;50(5):403–15.
 24. Anik A, Çatlı G, Abacı A, Böber E. Maturity-onset diabetes of the young (MODY): an update. *J Pediatr Endocrinol Metab*. 2015 Jan 1 ;28(3–4):251–63.
 25. Naylor R, Knight Johnson A, Del Gaudio D, Adam MP, Ardinger HH, Pagon RA. Maturity-Onset Diabetes of the Young Overview. *Gen Rev*. 2018 May 24.
 26. Challis BG, Semple RK. Genetic Disorders of Insulin Action: Far More than Diabetes. *Curr Obes Rep*. 2013 Dec 13;2(4):293–300.
 27. Hardt PD, Brendel MD, Kloer HU, Bretzel RG. Is pancreatic diabetes (type 3c diabetes) underdiagnosed and misdiagnosed? *Diabetes Care*. 2008 Feb 1;31 Suppl 2(Suppl 2):S165–9.
 28. Wynne K, Devereaux B, Dornhorst A. Diabetes of the exocrine pancreas. *J Gastroenterol Hepatol*. 2019 Feb 1;34(2):346–54.
 29. Rich P. Nail changes due to diabetes and other endocrinopathies. *Derm Ther*. 2002;15:107–110.
 30. Resmini E, Minuto F, Colao A, Ferone D. Secondary diabetes associated with principal endocrinopathies: the impact of new treatment modalities. *Acta Diabetol*. 2009 Jun 26;46(2):85–95.
 31. Ganda OP. Statin-induced diabetes: incidence, mechanisms, and implications. *F1000Research*. 2016;5.
 32. Al-Shaer MH, AbuSabha HS. Are the effects of nicotinic acid on insulin resistance precipitated by abnormal phosphorous metabolism? *Lipids*

- Health Dis. 2004 Oct 28;3:23.
33. Suh S, Park MK. Glucocorticoid-Induced Diabetes Mellitus: An Important but Overlooked Problem. *Endocrinol Metab* (Seoul, Korea). 2017 Jun;32(2):180–9.
 34. Ferner RE. 8 Drug-induced diabetes. *Baillieres Clin Endocrinol Metab*. 1992 Oct 1;6(4):849–66.
 35. Yeung W-CG, Rawlinson WD, Craig ME. Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. *BMJ*. 2011 Feb 3;342:d35.
 36. Allen DW, Kim KW, Rawlinson WD, Craig ME. Maternal virus infections in pregnancy and type 1 diabetes in their offspring: Systematic review and meta-analysis of observational studies. *Rev Med Virol*. 2018 May 1;28(3):e1974.
 37. Barrett TG. Mitochondrial diabetes, DIDMOAD and other inherited diabetes syndromes. *Best Pract Res Clin Endocrinol Metab*. 2001 Sep;15(3):325–43.
 38. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet*. 2017 Jun 3;389(10085):2239–51.
 39. Zaccardi F, Webb DR, Yates T, Davies MJ. Pathophysiology of type 1 and type 2 diabetes mellitus: A 90-year perspective. *Postgrad Med J*. 2016 Feb;92(1084):63-9.
 40. Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. *BMC Med*. 2017 Jul 19;15(1):131.
 41. Wild S, Roglic G, Green A, Sicree R, King H. Global Prevalence of Diabetes. *Diabetes Care*. 2004 May 1;27(5):1047–53.
 42. Ministerio de Sanidad SS e I. Encuesta Europea de Salud en España 2014. 2014.
 43. Soriguer F, Rojo-Martínez G, Almaraz MC, Esteva I, Ruiz de Adana MS, Morcillo S, et al. Incidence of type 2 diabetes in southern Spain (Pizarra Study). *Eur J Clin Invest*. 2008 Jan 22;38(2):126–33.

REFERENCIAS

44. Valenciana C. Estrategia de Diabetes Comunitat Valenciana 2017-2021. 2017.
45. Ana María Gandarillas Grande, Virginia del Pino Valero MOG. Estudio PREDIMERC2015 PREvalencia de Diabetes Mellitus y Riesgo Cardiovascular en población adulta de la Comunidad de Madrid. 2015.
46. Docampo García M, José López Rey M. Evolución futura de la diabetes mellitus. un análisis en el caso español. *Rev Esp Salud Pública*. 2018;92.
47. Effect of Mediterranean Diet Components on Type 2 Diabetes and Metabolic Syndrome. *Role Mediterr Diet Brain Neurodegener Dis*. 2018 Jan 1;209–22.
48. Spanakis EK, Golden SH. Race/ethnic difference in diabetes and diabetic complications. *Curr Diab Rep*. 2013 Dec;13(6):814–23.
49. Davegårdh C, García-Calzón S, Bacos K, Ling C. DNA methylation in the pathogenesis of type 2 diabetes in humans. *Mol Metab*. 2018;14:12.
50. Coustan DR, Lowe LP, Metzger BE, Dyer AR, International Association of Diabetes and Pregnancy Study Groups. The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study: paving the way for new diagnostic criteria for gestational diabetes mellitus. *Am J Obstet Gynecol*. 2010 Jun;202(6):654.e1-6.
51. Carolan M, Davey M-A, Biro MA, Kealy M. Maternal age, ethnicity and gestational diabetes mellitus. *Midwifery*. 2012 Dec 1;28(6):778–83.
52. Gilinsky AS, Kirk AF, Hughes AR, Lindsay RS. Lifestyle interventions for type 2 diabetes prevention in women with prior gestational diabetes: A systematic review and meta-analysis of behavioural, anthropometric and metabolic outcomes. *Prev Med Reports*. 2015 Jan 1;2:448–61.
53. Panagiotakos DB, Tzima N, Pitsavos C, Chrysohoou C, Papakonstantinou E, Zampelas A, et al. The Relationship between Dietary Habits, Blood Glucose and Insulin Levels among People without Cardiovascular Disease and Type 2 Diabetes; The ATTICA Study. *Rev Diabet Stud*. 2005;2(4):208–208.

54. Jenkins DW, Jenks A. Exercise and Diabetes: A Narrative Review. *J Foot Ankle Surg.* 2017 Sep 1;56(5):968–74.
55. Hossain P, Kawar B, El Nahas M. Obesity and Diabetes in the Developing World — A Growing Challenge. *N Engl J Med* . 2007 Jan 18;356(3):213–5.
56. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet.* 2010 Jul 10;376(9735):124–36.
57. Jenkins AJ, Joglekar M V, Hardikar AA, Keech AC, O’Neal DN, Januszewski AS. Biomarkers in Diabetic Retinopathy. *Rev Diabet Stud.* 2015;12(1–2):159–95.
58. Wada J, Makino H. Inflammation and the pathogenesis of diabetic nephropathy. *Clin Sci.* 2013;124:139–52.
59. Haneda M, Utsunomiya K, Koya D, Babazono T, Moriya T, Makino H, et al. A new classification of Diabetic Nephropathy 2014: a report from Joint Committee on Diabetic Nephropathy. *Clin Exp Nephrol.* 2015 Feb 20;19(1):1–5.
60. Bansal V, Kalita J, Misra UK. Diabetic neuropathy. *Postgrad Med J* . 2006 Feb;82(964):95–100.
61. Wu Y, Ding Y, Tanaka Y, Zhang W. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. *Int J Med Sci.* 2014;11(11):1185–200.
62. John A. Bittl M. Percutaneous Coronary Interventions in the Diabetic Patient Where Do We Stand? *Circ Cardiovasc Interv.* 2015;8.
63. Kalergis M, Leung Yinko SSL, Nedelcu R. Dairy products and prevention of type 2 diabetes: implications for research and practice. *Front Endocrinol (Lausanne).* 2013;4:90.
64. Salas-Salvadó J, Bulló M, Babio N, Martínez-González MÁ, Ibarrola-Jurado N, Basora J, et al. Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes Care.* 2011 Jan ;34(1):14–9.

65. Villegas R, Shu XO, Gao Y-T, Yang G, Elasy T, Li H, et al. Vegetable but not fruit consumption reduces the risk of type 2 diabetes in Chinese women. *J Nutr.* 2008 Mar;138(3):574–80.
66. Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. *Eur J Epidemiol.* 2015 Jul 20;30(7):529–42.
67. Curtis L. Triplitt, PharmD C. Examining the Mechanisms of Glucose Regulation. *Am J Manag Care.* 2012;18.
68. Sayem A, Arya A, Karimian H, Krishnasamy N, Ashok Hasamnis A, Hossain C, et al. Action of Phytochemicals on Insulin Signaling Pathways Accelerating Glucose Transporter (GLUT4) Protein Translocation. *Molecules.* 2018 Jan 28;23(2):258.
69. Skyler JS, Bakris GL, Bonifacio E, Darsow T, Eckel RH, Groop L, et al. Differentiation of Diabetes by Pathophysiology, Natural History, and Prognosis. *Diabetes.* 2017;66(2):241–55.
70. Muoio D NC. Molecular and metabolic mechanisms of insulin resistance and β -cell failure in type 2 diabetes. *Nat Rev Mol Cell Biol.* 2008;9:193–205.
71. LeRoith D. Beta-cell dysfunction and insulin resistance in type 2 diabetes: role of metabolic and genetic abnormalities. *Am J Med.* 2002;113 Suppl 6A:3S-11S.
72. Grill V, Björklund A. Impact of metabolic abnormalities for beta cell function: Clinical significance and underlying mechanisms. *Mol Cell Endocrinol.* 2009 Jan 15;297(1–2):86–92.
73. Poitout V, Robertson RP. Minireview: Secondary β -Cell Failure in Type 2 Diabetes—A Convergence of Glucotoxicity and Lipotoxicity. *Endocrinology.* 2002 Feb 1;143(2):339–42.
74. Chang-Chen KJ, Mullur R, Bernal-Mizrachi E. β -cell Failure as a Complication of Diabetes. *Rev Endocr Metab Disord.* 2008 Dec;9(4):329–43.
75. Talchai C, Xuan S, Lin H V, Sussel L, Accili D. Pancreatic β cell

- dedifferentiation as a mechanism of diabetic β cell failure. *Cell*. 2012 Sep 14;150(6):1223–34.
76. Bensellam M, Jonas JC LD. Mechanisms of β -cell dedifferentiation in diabetes: recent findings and future research directions. *J Endocrinol*. 2018;236(2):109–43.
 77. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*. 2006;444(7121):840–6.
 78. Kahn BB, Flier JS. Obesity and insulin resistance. *J Clin Invest*. 2000 Aug;106(4):473–81.
 79. Back SH, Kaufman RJ. Endoplasmic reticulum stress and type 2 diabetes. *Annu Rev Biochem*. 2012;81:767–93.
 80. Rains JL, Jain SK. Oxidative stress, insulin signaling, and diabetes. *Free Radic Biol Med*. 2011 Mar 1;50(5):567–75.
 81. Samuel VT, Shulman GI. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. *J Clin Invest* . 2016 Jan 4;126(1):12–22.
 82. Szabo M, Máté B, Csép K, Benedek T. Genetic Approaches to the Study of Gene Variants and Their Impact on the Pathophysiology of Type 2 Diabetes. *Biochem Genet*. 2018 Apr 16;56(1–2):22–55.
 83. Meigs JB, Cupples LA, Wilson PWF. Parental Transmission of Type 2 Diabetes The Framingham Offspring Study. *Diabetes*. 2000 Dec;49(12):2201–7.
 84. Lo SSS, Tun RYM, Hawa M, Leslie RDG. Studies of diabetic twins. *Diabetes / Metab Rev*. 1991 Dec 1;7(4):223–38.
 85. Willemsen G, Ward KJ, Bell CG, Christensen K, Bowden J, Dalgård C, et al. The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin Res Hum Genet*. 2015;18.
 86. Knowler WC, Bennett PH, Hamman RF, Miller M. Diabetes incidence and prevalence in pima indians: A 19-fold greater incidence than in

- Rochester, Minnesota. *Am J Epidemiol.* 1978 Dec 1;108(6):497–505.
87. Karter AJ, Schillinger D, Adams AS, Moffet HH, Liu J, Adler NE, et al. Elevated rates of diabetes in Pacific Islanders and Asian subgroups: The Diabetes Study of Northern California (DISTANCE). *Diabetes Care.* 2013 Mar 1;36(3):574–9.
 88. Prasad RB, Groop L. Genetics of type 2 diabetes—pitfalls and possibilities. *Genes (Basel).* 2015;6(1):87–123.
 89. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet.* 2000 Oct;26(2):163–75.
 90. Meyre D, Bouatia-Naji N, Tounian A, Samson C, Lecoeur C, Vatin V, et al. Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat Genet.* 2005 Aug;37(8):863–7.
 91. Vasseur F, Helbecque N, Dina C, Lobbens S, Delannoy V, Gaget S, et al. Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum Mol Genet.* 2002 Oct 2 Jul;11(21):2607–14.
 92. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, et al. Linkage of Type 2 Diabetes Mellitus and of Age at Onset to a Genetic Location on Chromosome 10q in Mexican Americans. *Am. J. Hum. Genet.* 1999;64.
 93. Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet.* 2006 Mar 15;38(3):320–3.
 94. Deeb SS, Fajas L, Nemoto M, Pihlajamäki J, Mykkänen L, Kuusisto J, et al. A Pro12Ala substitution in PPAR γ 2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet.* 1998 Nov;20(3):284–7.
 95. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl M-C,

- Nemesh J, et al. The common PPAR γ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet.* 2000 Sep;26(1):76–80.
96. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes.* 2003 Feb 1;52(2):568–72.
 97. Hansen SK, Nielsen E-MD, Ek J, Andersen G, Glümer C, Carstensen B, et al. Analysis of Separate and Combined Effects of Common Variation in KCNJ11 and PPARG on Risk of Type 2 Diabetes. *J Clin Endocrinol Metab.* 2005 Jun 1;90(6):3629–37.
 98. Kato N. Insights into the genetic basis of type 2 diabetes. *J Diabetes Investig.* 2013 May 6;4(3):233–44.
 99. Consortium TIH. The International HapMap Project. *Nature.* 2003 Dec;426(6968):789–96.
 100. Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. *Hum Mol Genet.* 2015 Oct 15;24(R1):R85–92.
 101. Tallapragada DSP, Bhaskar S, Chandak GR. New insights from monogenic diabetes for “common” type 2 diabetes. *Front Genet.* 2015;6:251.
 102. Merino J, Udler MS, Leong A, Meigs JB. A Decade of Genetic and Metabolomic Contributions to Type 2 Diabetes Risk Prediction. *Curr Diab Rep.* 2017 Nov 4;17(12):135.
 103. Ikegawa S. A short history of the genome-wide association study: where we were and where we are going. *Genomics Inform.* 2012 Dec;10(4):220–5.
 104. Zaitlen N, Kraft P. Heritability in the genome-wide association era. *Hum Genet.* 2012 Oct;131(10):1655–64.
 105. Kryukov G V, Pennacchio LA, Sunyaev SR. Most rare missense alleles

- are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet.* 2007 Apr;80(4):727–39.
106. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet.* 2010 Jun 1;11(6):415–25.
 107. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet.* 2008;82(1):100–12.
 108. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008 May;9(5):356–69.
 109. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009 Oct 8;461(7265):747–53.
 110. Metzker ML. Sequencing technologies — the next generation. *Nat Rev Genet.* 2010 Jan 8;11(1):31–46.
 111. Kiezun A, Garimella K, Do R, Stitzel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet.* 2013;44(6):623.
 112. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet.* 2018;50(4):559–71.
 113. Stranneheim H, Wedell A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *J Intern Med.* 2016;279:3–15.
 114. Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet.* 2015 Dec;47(12):1415–25.

REFERENCIAS

115. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. *Nature* . 2016 [cited 2017 Mar;536(7614):41–7.
116. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505.
117. Flannick J. The Contribution of Low-Frequency and Rare Coding Variation to Susceptibility to Type 2 Diabetes. *Curr Diab Rep* . 2019 Apr 8;19(5):25.
118. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature*. 2019 Jun 22;570(7759):71–6.
119. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MARR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. 2007 Sep 1;81(3):559–75.
120. Corporation P. QuantiFluor[®] dsDNA System Instructions for Use of Product E2670. Available from: www.promega.com/protocols
121. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*. 2004 Jul 8;91(2):355–8. Available from: <http://www.nature.com/articles/6601894>
122. NCBI . Available from: www.ncbi.nlm.nih.gov
123. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, et al. The Ensembl genome database project. *Nucleic Acids Research*. 2002;30. Available from: <http://www.bioperl.org/>
124. Kalendar R, Muterko A, Shamekova M, Zhambakin K. In Silico PCR Tools for a Fast Primer, Probe, and Advanced Searching. In: Domingues L, editor. *PCR: Methods and Protocols*. New York, NY: Springer New York; 2017:1–31. Available from: http://link.springer.com/10.1007/978-1-4939-7060-5_1

125. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007 May 15;23(10):1289–91. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btm091>
126. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res* . 2012 Aug 1;40(15):e115–e115. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks596>
127. Technologies A. QIAGEN Multiplex PCR Handbook For fast and efficient multiplex PCR without optimization. Qiagen. 2010 Oct. Available from: <https://www.qiagen.com/es/resources>
128. MagSi-NGS PREP Plus Product Manual MagSi-NGS PREP Plus Clean-up and size selection in sample preparation for NGS library construction. Available from: www.magnamedics.com
129. Illumina. Customize a short end-to-end workkow guide with the Custom Protocol Selector Nextera[®] DNA Library Prep Reference Guide. 2016 Available from: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplerepreps_nextera/nexteradna/nextera-dna-library-prep-reference-guide-15027987-01.pdf
130. Martínez-Barquero V. Identificación de variantes exómicas en diabetes tipo 2 Tesis doctoral presentada por : Vanesa Martínez Barquero. 2015.
131. Galan-Chilet I. Identificación de variantes genéticas poco frecuentes y raras en Diabetes mellitus tipo 2 mediante secuenciación de exoma. 2016.
132. Qiagen. FlexiGene[®] DNA Handbook Sample & Assay Technologies QIAGEN Sample and Assay Technologies. 2010 Apr. Available from: <https://www.qiagen.com/us/resources>
133. Agilent Technologies (California U. Enrichment for Illumina Multiplexed Sequencing Featuring Transposase-Based. *Tech Bull*. 2016 Dec:1–77.

REFERENCIAS

134. Tange Ole. GNU Parallel: The Command-Line Power Tool. Usenix. 2011 Feb;36(1). Available from: <https://www.usenix.org/system/files/login/articles/105438-Tange.pdf>
135. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
136. GitHub - lh3/seqtk: Toolkit for processing sequences in FASTA/Q formats. Available from: <https://github.com/lh3/seqtk>
137. Burrows-Wheeler Aligner download | SourceForge.net. Available from: <https://sourceforge.net/projects/bio-bwa/>
138. Sequence Alignment/Map Format Specification. Available from: <https://samtools.github.io/hts-specs/SAMv1.pdf>
139. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011 Nov 1;27(21):2987–93.
140. Bedtools: a powerful toolset for genome arithmetic — bedtools 2.29.0 documentation. Available from: <https://bedtools.readthedocs.io/en/latest/>
141. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012. Available from: <https://arxiv.org/pdf/1207.3907.pdf>
142. The Variant Call Format (VCF). Available from: <https://samtools.github.io/hts-specs/VCFv4.2.pdf>
143. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics*. 2015 Jul 1;31(13):2202–4.
144. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
145. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.

REFERENCIAS

146. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. Gardner PP, editor. *PLoS Comput Biol*. 2013 Jul 18;9(7):e1003153.
147. Bgzip program. Available from: <http://www.htslib.org/doc/bgzip.html>
148. Tabix program. Available from: <http://www.htslib.org/doc/tabix.html>
149. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014 Mar;46(3):310–5.
150. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017;45(D1):D183–9.
151. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc*. 2019 Mar 25;14(3):703–21.
152. RTG Tools Operations Manual v3.10. Available from: <https://cdn.rawgit.com/RealTimeGenomics/rtg-tools/master/installer/resources/tools/RTGOperationsManual/index.html>
153. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc*. 2011 Feb;6(2):121–33.
154. Ye X, Ji C, Zhou C, Zeng L, Gu S, Ying K, et al. Cloning and characterization of a human cDNA ACAD10 mapped to chromosome 12q24.1. *Mol Biol Rep*. 2004 Sep;31(3):191–5.
155. Bloom K, Mohsen AW, Karunanidhi A, El Demellawy D, Reyes-Múgica M, Wang Y, et al. Investigating the link of ACAD10 deficiency to type 2 diabetes mellitus. *J Inherit Metab Dis*. 2018;41(1):49–57.
156. Bian L, Hanson RL, Muller YL, Ma L, Kobes S, Knowler WC, et al. Variants in ACAD10 are associated with type 2 diabetes, insulin

- resistance and lipid oxidation in Pima Indians. *Diabetologia*. 2010 Jul 14;53(7):1349–53.
157. Koike T, Izumikawa T, Kitagawa H. Identification of Phosphatase That Dephosphorylates Xylose in the Glycosaminoglycan-Protein Linkage Region of Proteoglycans. *J Biol Chem*. 2014 Mar 7;289(10):6695-708.
 158. Chang YS, Tsai CT, Huangfu CA, Huang WY, Lei HY, Lin CF, et al. ACSL3 and GSK-3 β are essential for lipid upregulation induced by endoplasmic reticulum stress in liver cells. *J Cell Biochem*. 2011;112(3):881–93.
 159. Poppelreuther M, Rudolph B, Du C, Großmann R, Becker M, Thiele C, et al. The N-terminal region of acyl-CoA synthetase 3 is essential for both the localization on lipid droplets and the function in fatty acid uptake. *J Lipid Res*. 2012 May;53(5):888-900.
 160. Yao H, Ye J. Long chain Acyl-CoA synthetase 3-mediated phosphatidylcholine synthesis is required for assembly of very low density lipoproteins in human hepatoma Huh7 cells. *J Biol Chem*. 2008;283(2):849–54.
 161. Yan S, Yang X-F, Liu H-L, Fu N, Ouyang Y, Qing Sheng Yan K, et al. Long-chain acyl-CoA synthetase in fatty acid metabolism involved in liver and other diseases: An update. *World J Gastroenterol*. 2015;21(12):3492–8.
 162. Ansari I-UH, Longacre MJ, Stoker SW, Kendrick MA, O’neill LM, Zitur LJ, et al. Characterization of Acyl-CoA Synthetase Isoforms In Pancreatic Beta Cells: Gene Silencing Shows Participation of ACSL3 and ACSL4 In Insulin Secretion HHS Public Access. *Arch Biochem Biophys*. 2017;618:32–43.
 163. Bu SY, Mashek MT, Mashek DG. Suppression of long chain acyl-CoA synthetase 3 decreases hepatic de Novo fatty acid synthesis through decreased transcriptional activity. *J Biol Chem*. 2009;284(44):30474–83.
 164. Iwai N, Inagami T. Isolation of Preferentially Expressed Genes in the Kidneys of Hypertensive Rats. *Hypertension*. 1991 Feb;17(2):161-9.
 165. Gopal R, Selvarasu K, Pandian PP, Ganesan K. Integrative

- transcriptome analysis of liver cancer profiles identifies upstream regulators and clinical significance of ACSM3 gene expression. *Cell Oncol (Dordr)*. 2017 Jun;40(3):219-233.
166. Boomgaarden I, Vock C, Klapper M, Döring F. Comparative Analyses of Disease Risk Genes Belonging to the Acyl-CoA Synthetase Medium-Chain (ACSM) Family in Human Liver and Cell Lines. *Biochem Genet*. 2009 Oct;47(9-10):739-48.
 167. Currie E, Schulze A, Zechner R, Walther TC, Farese R V. Cellular Fatty Acid Metabolism and Cancer. *Cell Metab*. 2013 Aug 6;18(2):153-61.
 168. Pabarcus MK, Casida JE. Kynurenine formamidase: determination of primary structure and modeling-based prediction of tertiary structure and catalytic triad 1. *Biochim Biophys Acta*. 2002 Apr 29;1596(2):201-11.
 169. Dobrovolsky VN, Bowyer JF, Pabarcus MK, Heflich RH, Williams LD, Doerge D, et al. Effect of arylformamidase (kynurenine formamidase) gene inactivation in mice on enzymatic activity, kynurenine pathway metabolites and phenotype. *Biochim Biophys Acta*. 2005 Jun 20;1724(1-2):163-72.
 170. Hugill AJ, Stewart ME, Yon MA, Probert F, Cox IJ, Hough TA, et al. Loss of arylformamidase with reduced thymidine kinase expression leads to impaired glucose tolerance. *Biol Open*. 2015 Nov 15; 4(11): 1367–1375.
 171. Lin K-T, Ma WK, Scharner J, Liu Y-R, Krainer AR. A human-specific switch of alternatively spliced AFMID isoforms contributes to TP53 mutations and tumor recurrence in hepatocellular carcinoma. *Genome Res*. 2018 Mar; 28(3): 275–284.
 172. Bishop AL, Hall A. Rho GTPases and their effector proteins. *Biochem J*. 2000 Jun 1; 348(Pt 2): 241–255.
 173. Van Aelst L, D'Souza-Schorey C. Rho GTPases and signaling networks. *Genes Dev*. 1997;11(18):2295–322.
 174. Type 2 Diabetes Knowledge Portal: Available from: <http://www.type2diabetesgenetics.org/>

175. Kato AS, Siuda ER, Nisenbaum ES, Bredt DS. AMPA Receptor Subunit-Specific Regulation by a Distinct Family of Type II TARPs. *Neuron*. 2008 Sep 25;59(6):986–96.
176. Xia C, Bao Z, Tabassam F, Ma W, Qiu M, Hua SB, et al. GCIP, a novel human Grap2 and cyclin D interacting protein, regulates E2F-mediated transcriptional activity. *J Biol Chem*. 2000;275(27):20942–8.
177. Lee I, Yeom SY, Lee SJ, Kang WK, Park C. A novel senescence-evasion mechanism involving Grap2 and cyclin D interacting protein inactivation by Ras associated with diabetes in cancer cells under doxorubicin treatment. *Cancer Res*. 2010;70(11):4357–65.
178. Zhu SW, Li JP, Ma XL, Ma JX, Yang Y, Chen Y, et al. miR-9 modulates osteosarcoma cell growth by targeting the GCIP tumor suppressor. *Asian Pacific J Cancer Prev*. 2015;16(11):4509–13.
179. Zhang X, Gaspard JP, Mizukami Y, Li J, Graeme-Cook F, Chung DC. Overexpression of cyclin D1 in pancreatic beta-cells in vivo results in islet hyperplasia without hypoglycemia. *Diabetes*. 2005 Mar 1;54(3):712–9.
180. Lee Y, Dominy JE, Choi YJ, Jurczak M, Tolliday N, Camporez JP, et al. Cyclin D1-CDK4 Controls Glucose Metabolism Independently of Cell Cycle Progression. *Nature*. 2014 Jun 26;510(7506):547-51.
181. Mukherjee K, Conway De Macario E, Macario AJ, Brocchieri L. Chaperonin genes on the rise: new divergent classes and intense duplication in human and other vertebrate genomes. *BMC Evol Biol*. 2010 Mar 1;10:64.
182. Shibayama H, Takai E, Matsumura I, Kouno M, Morii E, Kitamura Y, et al. Identification of a Cytokine-induced Antiapoptotic Molecule Anamorsin Essential for Definitive Hematopoiesis. *J Exp Med*. 2004 Feb 16;199(4):581-92.
183. Chen X, Li X, Chen J, Zheng P, Huang S, Ouyang X. Overexpression of CIAPIN1 inhibited pancreatic cancer cell proliferation and was associated with good prognosis in pancreatic cancer. *Cancer Gene Ther*. 2012;19:538–44.

184. Wang X, Pan J, Lil J. Cytokine-induced apoptosis inhibitor 1 inhibits the growth and proliferation of multiple myeloma. *Mol Med Rep.* 2015 Aug 1;12(2):2056–62.
185. Lang D, Radtke M, Bairlein M. Highly Variable Expression of CYP1A1 in Human Liver and Impact on Pharmacokinetics of Riociguat and Granisetron in Humans. *Chem Res Toxicol.* 2019 Jun 17;32(6):1115-1122.
186. Wang XL, Greco M, Sim AS, Duarte N, Wang J, Wilcken DE. Effect of CYP1A1 MspI polymorphism on cigarette smoking related coronary artery disease and diabetes. *Atherosclerosis.* 2002 Jun;162(2):391–7.
187. Go R-E, Hwang K-A, Choi K-C. Cytochrome P450 1 family and cancers. *J Steroid Biochem Mol Biol.* 2015 Mar 1;147:24–30.
188. Katane M, Kanazawa R, Kobayashi R, Oishi M, Nakayama K, Saitoh Y, et al. Structure–function relationships in human D-aspartate oxidase: characterisation of variants corresponding to known single nucleotide polymorphisms. *Biochim Biophys Acta - Proteins Proteomics.* 2017 Sep;1865(9):1129–40.
189. Li Y, Han H, Yin J, Li T, Yin Y. Role of D-aspartate on biosynthesis, racemization, and potential functions: A mini-review. *Anim Nutr.* 2018 Sep;4(3):311-315.
190. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, et al. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature.* 2010 Jan 7;463(7277):84–7.
191. Renieri A, Mencarelli MA, Cetta F, Baldassarri M, Mari F, Furini S, et al. Oligogenic germline mutations identified in early non-smokers lung adenocarcinoma patients. *Lung Cancer.* 2014 Aug;85(2):168–74.
192. Li H, Xu R, Peng X, Wang Y, Wang T. Association of glucokinase regulatory protein polymorphism with type 2 diabetes and fasting plasma glucose: a meta-analysis. *Mol Biol Rep.* 2013 Jun;40(6):3935-42.
193. Irwin DM, Tan H. Evolution of glucose utilization: Glucokinase and glucokinase regulator protein. *Mol Phylogenet Evol.* 2014;70:195–203.

194. Brouwers MCGJ, Jacobs C, Bast A, Stehouwer CDA, Schaper NC. Modulation of Glucokinase Regulatory Protein: A Double-Edged Sword? *Trends Mol Med*. 2015 Oct;21(10):583–94.
195. Mohás M, Kisfali P, Járomi L, Maász A, Fehér E, Csöngéi V, et al. GCKR gene functional variants in type 2 diabetes and metabolic syndrome: do the rare variants associate with increased carotid intima-media thickness? *Cardiovasc Diabetol*. 2010 Nov 29;9:79.
196. Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C, et al. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes*. 2008 Nov;57(11):3112–21.
197. Vaxillaire M, Cavalcanti-Proença C, Dechaume A, Tichet J, Marre M, Balkau B, et al. The common P446L polymorphism in GCKR inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the DESIR prospective general French population. *Diabetes*. 2008 Aug;57(8):2253–7.
198. Passi A, Sardao VA, Singleton PA, Mirzapooiazova T, Mambetsariev N, Lennon FE, et al. HABP2 is a novel regulator of hyaluronan-mediated human lung cancer progression. *Front Oncol*. 2015;5:164.
199. Byskov K, Boettger T, Ruehle PF, Nielsen NV, Etscheid M, Kanse SM. Factor VII activating protease (FSAP) regulates the expression of inflammatory genes in vascular smooth muscle and endothelial cells. *Atherosclerosis*. 2017 Oct;265:133–9.
200. Kanse SM, Declerck PJ, Ruf W, Broze G, Etscheid M. Factor VII activating protease (FSAP) promotes the proteolysis and inhibition of tissue factor pathway inhibitor (TFPI). *Arter Thromb Vasc Biol*. 2012;32(2):427–33.
201. Kern B, Coppin L, Romanet P, Crépin M, Szuster I, Renaud F, et al. Multiple HABP2 variants in familial papillary thyroid carcinoma: Contribution of a group of “thyroid-checked” controls. *Eur J Med Genet*. 2017 Mar;60(3):178–84.
202. Tomarev SI, Nakaya N. Olfactomedin domain-containing proteins:

- possible mechanisms of action and functions in normal development and pathology. *Mol Neurobiol.* 2009 Oct;40(2):122–38.
203. Wan B, Zhou Y-B, Zhang X, Zhu H, Huo K, Han Z-G. hOLFML1, a novel secreted glycoprotein, enhances the proliferation of human cancer cell lines in vitro. *FEBS Lett.* 2008 Sep 22;582(21–22):3185–92.
204. Murakami K, Kikugawa S, Kobayashi Y, Uehara S, Suzuki T, Kato H, et al. Olfactomedin-like protein OLFML1 inhibits Hippo signaling and mineralization in osteoblasts. *Biochem Biophys Res Commun.* 2018 Oct;505(2):419–25.
205. Pujar MK, Vastrad B, Vastrad C. Integrative Analyses of Genes Associated with Subcutaneous Insulin Resistance. *Biomolecules.* 2019 Jan 22;9(2).
206. Adragna NC, Lauf PK. K-Cl cotransport function and its potential contribution to cardiovascular disease. *Pathophysiology.* 2007;14:135–46.
207. Bazúa-Valenti S, Castañeda-Bueno M, Gamba G. Physiological role of SLC12 family members in the kidney. *Am J Physiol Physiol.* 2016 Jul 1;311(1):F131–44.
208. Garneau AP, Marcoux AA, Slimani S, Tremblay LE, Frenette-Cotton R, Mac-Way F, et al. Physiological roles and molecular mechanisms of K^+ - Cl^- cotransport in the mammalian kidney and cardiovascular system: where are we? *J Physiol.* 2019 Mar 9;597(6):1451–65.
209. Shen M-R, Lin A-C, Hsu Y-M, Chang T-J, Tang M-J, Alper SL, et al. Insulin-like growth factor 1 stimulates KCl cotransport, which is necessary for invasion and proliferation of cervical cancer and ovarian cancer cells. *J Biol Chem.* 2004 Sep 17;279(38):40017–25.
210. Daniel H, Kottra G. The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Pflugers Arch-Eur J Physiol.* 2004 Feb;447:610–8.
211. He L, Wang B, Li Y, Zhu L, Li P, Zou F, et al. The Solute Carrier Transporter SLC15A3 Participates in Antiviral Innate Immune Responses against Herpes Simplex Virus-1. *J Immunol Res.* 2018.

212. Song F, Yi Y, Li C, Hu Y, Wang J, Smith DE, et al. Regulation and biological role of the peptide/histidine transporter SLC15A3 in Toll-like receptor-mediated inflammatory responses in macrophage. *Cel Death & Dis.* 2018 Jul;9(7):770.
213. Sato R, Inoue J, Kawabe Y, Kodama T, Takano T, Maeda M. Sterol-dependent transcriptional regulation of sterol regulatory element-binding protein-2. *J Biol Chem.* 1996 Oct 25;271(43):26461-4.
214. Miserez AR, Cao G, Probst LC, Hobbs HH. Structure of the Human Gene Encoding Sterol Regulatory Element Binding Protein 2 (SREBF2). *Genomics.* 1997 Feb 15;40(1):31-40.
215. Kotzka J, Müller-Wieland D, Roth G, Kremer L, Munck M, Schürmann S, et al. Sterol regulatory element binding proteins (SREBP)-1a and SREBP-2 are linked to the MAP-kinase cascade. *J Lipid Res.* 2000 Jan 1;41(1):99-108.
216. Robinet P, Védie B, Chironi G, Gariépy J, Simon A, Moatti N, et al. Characterization of polymorphic structure of SREBP-2 gene: role in atherosclerosis. *Atherosclerosis.* 2003 Jun 1;168(2):381-7.
217. Assmann N, O'Brien KL, Donnelly RP, Dyck L, Zaiatz-Bittencourt V, Loftus RM, et al. Srebp-controlled glucose metabolism is essential for NK cell functional responses. *Nat Immunol.* 2017 Nov 18;18(11):1197-206.
218. Galavi H, Noorzehi N, Saravani R, Sargazi S, Mollashahee-Kohkan F, Shahraki H. Association study of SREBF-2 gene polymorphisms and the risk of type 2 diabetes in a sample of Iranian population. *Gene .* 2018 Jun 20;660:145-50.
219. Musso G, Cassader M, Bo S, De Michieli F, Gambino R. Sterol regulatory element-binding factor 2 (SREBF-2) predicts 7-year NAFLD incidence and severity of liver disease and lipoprotein and glucose dysmetabolism. *Diabetes.* 2013;62(4):1109-20.
220. Rojo-Martínez G, Esteva I, Ruiz De Adana S, Catalá M, Merelo MJ, Tinahones F, et al. Patterns of insulin resistance in the general population of southeast Spain. *Diabetes Res Clin Pract.* 2004 Sep;65(3):247-56.

221. Nicot A-S, Laporte J. Endosomal Phosphoinositides and Human Diseases. *Traffic*. 2008 Aug;9(8):1240-9.
222. Weidner P, Söhn M, Gutting T, Friedrich T, Gaiser T, Magdeburg J, et al. Myotubularin-related protein 7 inhibits insulin signaling in colorectal cancer. *Oncotarget*. 2016 Aug 2;7(31):50490-506.
223. Taylor C, Wadey R, O'Donnell H, Roberts C, Mattei M-G, Kimber WL, et al. Cloning and mapping of murine Dgcr2 and its homology to the Sez-12 seizure-related protein. *Mamm Genome*. 1997 May;8(5):371-5.
224. Lindskog C, Korsgren O, Pontén F, Eriksson JW, Johansson L, Danielsson A. Novel pancreatic beta cell-specific proteins: Antibody-based proteomics for identification of new biomarker candidates. *J Proteomics*. 2012 May 17;75(9):2611-20.
225. Ghosh R, Roy S, Kamyab J, Danzter F, Franco S. Common and unique genetic interactions of the poly(ADP-ribose) polymerases PARP1 and PARP2 with DNA double-strand break repair pathways. *DNA Repair (Amst)*. 2016 Sep;45:56-62.
226. Wang K, Zhou W, Meng P, Wang P, Zhou C, Yao Y, et al. Immune-related somatic mutation genes are enriched in PDACs with diabetes. *Transl Oncol*. 2019 Jun 13;12(9):1147-54.
227. Doumatey AP, Xu H, Huang H, Trivedi NS, Lei L, Elkahloun A, et al. Global Gene Expression Profiling in Omental Adipose Tissue of Morbidly Obese Diabetic African Americans. *J Endocrinol Metab*. 2015 Jun;5(3):199-210.
228. Salonen JT, Uimari P, Aalto J-M, Pirskanen M, Kaikkonen J, Todorova B, et al. Type 2 Diabetes Whole-Genome Association Study in Four Populations: The DiaGen Consortium. *Am J Hum Genet*. 2007 Aug;81(2):338.
229. Wu N-N, Zhao D, Ma W, Lang J-N, Liu S-M, Fu Y, et al. A genome-wide association study of gestational diabetes mellitus in Chinese women. *J Matern Neonatal Med*. 2019 Jul 15;1-8.
230. Zhang Y, Kent JW, Olivier M, Ali O, Cerjak D, Broeckel U, et al. A comprehensive analysis of adiponectin QTLs using SNP association,

- SNP cis-effects on peripheral blood gene expression and gene expression correlation identified novel metabolic syndrome (MetS) genes with potential role in carcinogenesis and systemic inflammation. *BMC Med Genomics*. 2013 Apr 29;6:14.
231. Sakaguchi M, Cai W, Wang C-H, Cederquist CT, Damasio M, Homan EP, et al. FoxK1 and FoxK2 in insulin regulation of cellular and mitochondrial metabolism. *Nat Commun*. 2019;10(1):1582.
232. Soriguer F, Goday A, Bosch-Comas A, Bordi?? E, Calle-Pascual A, Carmena R, et al. Prevalence of diabetes mellitus and impaired glucose regulation in Spain: The Di@bet.es Study. *Diabetologia*. 2012 Jan;55(1):88-93
233. Mena-Martin FJ, Martin-Escudero JC, Simal-Blanco F, Carretero-Ares JL, Arzua-Mouronte D, Herreros-Fernandez V. Health-related quality of life of subjects with known and unknown hypertension: results from the population-based Horteiga study. *J Hypertens*. 2003 Jul;21(7):1283-9.

REFERENCIAS

