



VNIVERSITAT
DE VALÈNCIA

**IDENTIFICACIÓN DE VARIANTES GENÉTICAS
POCO FRECUENTES Y RARAS
EN DIABETES MELLITUS TIPO 2
MEDIANTE SECUENCIACIÓN DE EXOMA**

Tesis Doctoral

Presentada por:

INMACULADA GALÁN CHILET

Dirigida por:

Felipe Javier Chaves Martínez

PROGRAMA DE DOCTORADO EN BIOQUÍMICA Y BIOMEDICINA

Valencia, 2015

UNIVERSITAT DE VALÈNCIA
DOCTORADO EN BIOQUÍMICA Y BIOMEDICINA
Departamento de Bioquímica y Biología Molecular

FUNDACIÓN INVESTIGACIÓN CLÍNICO DE VALENCIA-INCLIVA
Unidad de Genotipado y Diagnóstico Genético

**IDENTIFICACIÓN DE VARIANTES GENÉTICAS POCO
FRECUENTES Y RARAS EN DIABETES MELLITUS TIPO 2
MEDIANTE SECUENCIACION DE EXOMA**

Tesis doctoral

Presentada por:
Inmaculada Galán Chilet

Dirigida por:
Felipe Javier Chaves Martínez

Valencia, 2015



VNIVERSITAT
DE VALÈNCIA



D. **FELIPE JAVIER CHAVES MARTÍNEZ**, doctor en Ciencias Biológicas, investigador y director de la Unidad de Genotipado y Diagnóstico Genético en la Fundación Hospital Clínico de Valencia-INCLIVA.

INFORMA:

Que Dña. **INMACULADA GALÁN CHILET**, Licenciada en Ciencias Biológicas por la Universitat de València, ha realizado bajo mi dirección en la Fundación de Investigación Clínico de Valencia-INCLIVA, en la Unidad de Genotipado y Diagnóstico Genético, el trabajo experimental conducente a la elaboración de la presente memoria de investigación destinada a Tesis Doctoral, que se presenta bajo el título de: **“IDENTIFICACIÓN DE VARIANTES GENÉTICAS POCO FRECUENTES Y RARAS EN DIABETES MELLITUS TIPO 2 MEDIANTE SECUENCIACIÓN DE EXOMA”**.

Y para que así conste, expido el presente en Valencia, 5 de Noviembre de 2015.

Fdo. Dr. F. Javier Chaves Martínez

Este trabajo de tesis doctoral ha sido financiado por el proyecto coordinado "Identificación de variantes genéticas en el exoma asociadas con DM2" (PI11/00726), Instituto de Salud Carlos III, Fondo de Investigaciones Sanitarias, cuyo investigador principal es el **Dr. F. Javier Chaves Martínez**.

Dña. **Inmaculada Galán Chilet**, ha recibido una ayuda predoctoral de formación en investigación del Instituto de Salud Carlos III, Fondo de Investigaciones Sanitarias (F112/00247).

If I have seen further, it's by standing upon the shoulders of giants

Isaac Newton - Carta a Robert Hooke (1675)

*Als meus pares, Juanjo i Conchin,
per l'esforç en donar-me una educació,
pel seu suport en els moments difícils
i per ser el meu exemple de vida.*

Agraïments

L'elaboració d'una tesi doctoral no sols és producte de l'esforç individual sinó també del col·lectiu. S'ha de reconèixer el mèrit propi per aconseguir els objectius proposats, encara que sense oblidar que mai es parteix de zero, i que només es pot arribar més lluny perquè uns altres abans han acumulat el coneixement necessari per a fer progressar el nostre.

Des d'ací he d'expressar el meu agraïment a una gran quantitat de persones; els meus particulars gegants. Alguns m'han permès veure més enllà amb les seues orientacions i consells, uns altres m'han portat més lluny amb els seus ensenyaments i amb el seu suport, i uns altres m'han ajudat amb el seu afecte. A tots els dec la meua gratitud per la seua paciència i consideració.

Per ixe motiu, m'agradaria en primer lloc agrair al meu director, Javier Chaves, l'oportunitat que m'ha brindat i la confiança que ha dipositat en mi en tot moment. Gràcies per la teua motivació, pel teu suport, i per ser tot un exemple professional i personal.

Seguidament, vull donar gràcies a tots els meus companys, als presents i als que han passat a llarg d'aquests anys, perquè res seria el mateix sense esta xicoteta (o no tan xicoteta...) família! Gràcies per eixos dinars de Nadal, per eixos discursos "improvisats", per eixe "paridero", per les llistes de spotify dels divendres, per les risses continues, els beneits croissants i per totes ixes senyes d'entitat que fan més gran si cap este equip. Gràcies: Griselda, Pilar, Veronik, Tani, Kike, Victoria, Cris, Raquel, Javi, Loles, Ana Barbara, Charo, Laura i Vero. En particular, gràcies Griselda per la teua ajuda al lab 14 i per aprendre valencià involuntàriament; gràcies a les meues mamis, Cris i Victòria, per cuidar-me i mimar-me tant, conclusió: sou genials!; gràcies al meu pare telefònic per ser un exemple d'humilitat i superació, Tani; gràcies Raquel per la teua ajuda amb la tesi, disposició i consells; gràcies Vero per estar sempre ací per a tot des de el primer dia, i gràcies Javi, amic i company, pel teu suport incondicional. També vull recordar les que han passat al llarg d'este temps pel meu camí i que estan també molt present: Carmen, ML, Verónica, Julia, Pilar, Elena, Lucía i Vane. I en general, gràcies Unitat de Genotipazo per formar part d'esta etapa de la meua vida i fer este període més fàcil.

Agrair a l'Unitat de Bioinformàtica, en especial a Dani, la seua ajuda en l'anàlisi dels resultats, les seues explicacions i el seu temps. També a Miguel, pel suport informàtic i per eixe bon rotllo. Al Morenazo, Alex, per ser el millor veí i motivar-me diàriament, i a Maria, gràcies per les teues ensenyances a l'hora d'escriure articles, per les col·laboracions i per l'ajuda.

No podria oblidar-me tampoc de l'altra gran família que tinc en EEUU, Gràcies a Carlos Fernández Hernando, per confiar en mi i donar-me l'oportunitat de fer l'estada al seu laboratori. Gràcies a la meua postdoc Cristina, per ensenyar-me tant en tan poc de temps, per la seua ajuda continua i per ser millor persona. I com no a la resta de "americans": Lee, Nate, Binhod, Paramita, Noemi, Jose i Alberto. També a tots els amics que es van quedar allí, vos duc sempre en la ment! Moltes gràcies per fer que l'experiència fora inoblidable.

Gràcies també a la gent que va confiar en mi abans d'arribar a l'Unitat, a M^aJesús Pujalte i Esperanza Garay per donar-me l'oportunitat de començar els meus primers passos en el món de l'investigació. A Virginia, per ser una companya extraordinària i millor amiga. I a José Antonio López Guerrero i el seu equip, per l'oportunitat de treballar amb ells i donar-me les primeres lliçons en diagnòstic genètic.

En relació a allò personal, també vull recordar-me de les meues ganxilleres, les meues amigues de tota la vida, ixes que sempre han estat i estan en lo bo i lo roïn: Lidi, Yoli, Laura, Gio, Nuri, Eva, Lidon, M^aJosé i Melvi. Sou les millors!

Gràcies a la meua família. Gràcies a tots, als presents i als que segueixen estant present cada dia de la meua vida, gràcies iaaios, Manolo i Conchin, per cuidar-me tant! Però sobretot agraïda al meu germà Juanjo, per ser una extraordinària persona i el millor germà del món, i als meus pares, per ser un exemple a seguir i un espill on mirar-se en tots els aspectes de la vida. Gràcies perquè sense el vostre suport res seria el mateix!

I per descomptant, agrair especialment al meu company de vida tot el seu suport, comprensió, paciència i confiança depositada en mi. Molt orgullosa de tenir-te al meu costat, te vuic moltíssim, Pepe!

Abreviaturas

ACV,	Accidente cerebrovascular
ADA,	Asociación americana de diabetes
ADN,	Ácido desoxirribonucleico
ADP,	Adenosin difosfato
AGA,	Alteración de la glucosa en ayunas
AIRg,	Respuesta aguda de insulina a la glucosa
ARN,	Ácido ribonucleico
ATP,	Adenosin trifosfato
cAMP,	Adenosin monofosfato cíclico
CI,	Cardiopatía isquémica
COSM	COSMIC, Catálogo de mutaciones somáticas en cáncer
CS, CR, CD	HGMD, Base de datos de mutaciones génicas humanas
DM,	Diabetes mellitus
DM1,	Diabetes mellitus tipo 1
DM2,	Diabetes mellitus tipo 2
DMG,	Diabetes mellitus gestacional
DMN,	Diabetes mellitus neonatal
DMNP,	Diabetes mellitus neonatal permanente
DMNT,	Diabetes mellitus neonatal transitoria
dNTPs,	Dinucleótidos trifosfato
FANCE	GeneCards, Base de datos genética humana
FFAs,	Ácidos grasos libres
FID,	Federación internacional de diabetes
g, mg, µg, ng, kg	Gramo, miligramo, microgramo, nanogramo, kilogramo
GBA,	Glucosa basal alterada
GH,	Hormona del crecimiento
GLUT2,	Transportador de glucosa tipo 2
GSIS,	Secreción de insulina estimulada por glucosa
GWAS,	Estudios de asociación de genoma completo
HDL,	Lipoproteína de alta densidad
HLA	Antígeno leucocitario humano
HOMA	Del inglés, Homeostasis model assessment
IAPP,	Polipéptido amiloide del islote
IC,	Intervalo de confianza
IGV,	Del inglés, Integrative genome viewer
IMC,	Índice de masa corporal
IPE,	Insuficiencia pancreática exocrina
IRS,	Sustratos proteicos del receptor de insulina
ITG,	Intolerancia a la glucosa
LDL,	Lipoproteína de baja densidad
M, µM, nM, pM	Molar, micromolar, nanomolar, picomolar
MAF,	Frecuencia del alelo minoritario
MODY,	Del inglés, Maturity-onset diabetes of the young

ND,	Nefropatía diabética
NGS,	Secuenciación de nueva generación
NRD,	Neuropatía diabética
OR,	Del inglés, Odds ratio
pb, Mpb,	Pares de bases, megapares de bases
PCR,	Reacción en cadena de la polimerasa
PF,	Polimorfismo frecuente
PGE,	Producción endógena de glucosa
PPF,	Polimorfismo poco frecuente
RD,	Retinopatía diabética
RE,	Retículo endoplasmático
RI,	Resistencia a la insulina
RISN	Base de datos del boletín internacional de mutaciones de retina
ROS,	Especies reactivas de oxígeno
SBDSBase	Registro de variantes para el síndrome Shwachman-Diamond
SOG,	Sobrecarga oral de glucose
TAG,	Tolerancia anormal a la glucosa
VCF,	Del inglés, Variant call format
VEP,	Del inglés, Variant effect predictor
WHO,	Organización mundial de la salud

Índice

RESUMEN/ABSTRACT	1
INTRODUCCIÓN	3
1. EPIDEMIOLOGÍA DE LA DIABETES MELLITUS	10
2. CRITERIOS DIAGNÓSTICOS	12
3. TIPOS DE DIABETES	13
3.1. Diabetes mellitus tipo 1	13
3.2. Diabetes mellitus tipo LADA (<i>Latent Autoimmune Diabetes of Adults</i>)	14
3.3. Diabetes mellitus gestacional	15
3.4. Otros tipos de diabetes mellitus	16
3.4.1. Diabetes mellitus monogénicas	16
3.4.2. Diabetes sindrómicas	18
3.4.3. Enfermedades del páncreas exocrino	18
3.4.4. Endocrinopatías	18
3.4.5. Diabetes mellitus asociada a fármacos	19
3.4.6. Diabetes mellitus asociada a infecciones	19
3.5. Diabetes mellitus tipo 2	19
3.5.1. Prevalencia de la diabetes mellitus tipo 2 en población española	20
3.5.2. Aspectos clínicos y evolución de la diabetes mellitus tipo 2	20
3.5.3. Prevención y diagnóstico de la diabetes mellitus tipo 2	21
3.5.4. Factores de riesgo de la diabetes mellitus tipo 2	23
3.5.5. Mecanismos moleculares y fisiopatología de la diabetes mellitus tipo 2	26
3.5.5.1. Resistencia a insulina en la diabetes mellitus tipo 2	30
3.5.5.2. Disfunción de la célula β en la diabetes mellitus tipo 2	32
3.5.6. Mortalidad y coste sanitario de la diabetes mellitus tipo 2	37
4. GENÉTICA DE LA DIABETES MELLITUS TIPO 2	39
4.1. El factor ambiental en la diabetes mellitus tipo 2	39
4.2. Heredabilidad de la diabetes mellitus tipo 2	39
4.3. Análisis de ligamiento genético en familias	40
4.4. Estudio de genes candidatos	41
4.5. Estudios de asociación de genoma completo (GWAS)	42
4.6. Estudios de meta-análisis a gran escala (<i>Large-scale meta-analysis of genome-wide association studies</i>)	43
5. NUEVAS ESTRATEGIAS PARA EL ESTUDIO GENÉTICO DE LA DIABETES MELLITUS TIPO 2	46
5.1. Identificación de variantes genéticas en el exoma	47
5.2. Estudio de polimorfismos de baja frecuencia y variantes raras	47
5.3. Selección de casos y controles mediante criterios específicos	49

6. SECUENCIACION DE NUEVA GENERACIÓN: PLATAFORMA ILLUMINA	51
HIPÓTESIS Y OBJETIVOS	55
MATERIAL Y MÉTODOS	59
1. SELECCIÓN DE LA POBLACIÓN DE ESTUDIO	61
2. SECUENCIACIÓN DEL EXOMA COMPLETO	61
2.1. Preparación de las librerías de ADN	62
2.1.1. Cuantificación del ADN	63
2.1.2. Fragmentación del ADN	63
2.1.3. Reparación de los extremos	64
2.1.4. Adenilación	64
2.1.5. Ligación de los adaptadores	64
2.1.6. PCR de enriquecimiento	66
2.2. Enriquecimiento del exoma	66
2.3. Validación de los pools	69
2.3.1. Visualización y selección de los pools	69
2.3.2. Cuantificación de los pools	70
2.4. Amplificación clonal de los pools: cBOT	71
2.5. Secuenciación por síntesis: HiScanSQ (Illumina)	72
3. ANÁLISIS BIOINFORMÁTICO	74
3.1. Control de calidad y mapeo de las secuencias	75
3.1.1. Control de calidad: FastQC	76
3.1.2. Limpieza de secuencias: Seqtk	77
3.1.3. Mapeo: alineamiento de las lecturas contra el genoma de referencia	78
3.2. Llamado de variantes genéticas	79
3.3. Anotación funcional de variantes genéticas	81
4. VERIFICACIÓN DE LAS VARIANTES GENÉTICAS MEDIANTE SECUENCIACIÓN CLÁSICA SANGER	83
RESULTADOS	85
1. CARACTERÍSTICAS GENERALES DE LA POBLACIÓN DE ESTUDIO	87
2. CONTROL DE CALIDAD DE LA SECUENCIACIÓN: FASTQC	89
3. COBERTURA DE LOS DATOS SECUENCIADOS	92
4. VARIANTES GENÉTICAS IDENTIFICADAS TRAS LA SECUENCIACIÓN DEL EXOMA	95
5. SELECCIÓN DE VARIANTES GENÉTICAS CON POSIBLE EFECTO FUNCIONAL	97
6. VARIANTES GENÉTICAS DE <i>SPLICING</i> , <i>MISSENSE</i> Y EN SECUENCIAS CODIFICANTES PARA microARNs	99
6.1. Estudio de variantes genéticas de <i>splicing</i> en el exoma	99
6.1.1. Variantes genéticas de <i>splicing</i> en el grupo control	100
6.1.2. Variantes genéticas de <i>splicing</i> en el grupo casos	108
6.2. Estudio de variantes genéticas <i>missense</i> en el exoma	115
6.2.1. Variantes genéticas <i>missense</i> en el grupo control	115

6.2.2. Variantes genéticas <i>missense</i> en el grupo de casos	117
6.3. Estudio de variantes genéticas en secuencias codificantes para microARNs en el exoma.....	119
6.3.1. Variantes genéticas en secuencias codificantes para microARNs en el grupo control	119
6.3.2. Variantes genéticas en secuencias codificantes para microARNs en el grupo de casos	120
7. ESTUDIO DE VARIANTES GENÉTICAS DE CODÓN DE STOP EN EL EXOMA.....	120
7.1. Estudio de variantes genéticas de codón de stop en el grupo control.....	121
7.2. Estudio de variantes genéticas de codón de stop en el grupo de casos	128
DISCUSIÓN	135
1. CRITERIOS DE SELECCIÓN DE LA POBLACIÓN DE ESTUDIO	138
2. ESTUDIO DE VARIANTES GENÉTICAS DE BAJA FRECUENCIA O RARAS	139
3. SECUENCIACIÓN DE NUEVA GENERACIÓN: ESTUDIO DEL EXOMA.....	142
4. COBERTURA DE SECUENCIACIÓN Y ANÁLISIS BIOINFORMÁTICO	144
5. VARIANTES GENÉTICAS IDENTIFICADAS CON POSIBLE EFECTO FUNCIONAL	145
5.1. Variantes genéticas en regiones de <i>splicing</i>	145
5.2. Variantes genéticas de cambio de sentido o <i>missense</i>	148
5.3. Variantes genéticas en regiones codificantes para microARNs maduros	151
6. VARIANTES GENÉTICAS IDENTIFICADAS Y VERIFICADAS CON EFECTO FUNCIONAL: VARIANTES DE CODÓN DE STOP	154
7. INNOVACIÓN, LIMITACIONES Y PERSPECTIVAS FUTURAS DEL ESTUDIO	157
CONCLUSIONES/CONCLUSIONS	159
BIBLIOGRAFÍA	165
ANEXO	167

Identificación de variantes genéticas poco frecuentes y raras en diabetes mellitus tipo 2 mediante secuenciación de exoma

La diabetes mellitus tipo 2 (DM2) se ha convertido en una epidemia global que contribuye significativamente a la morbi-mortalidad prematura. La DM2 es resultado de múltiples factores, incluyendo ambientales, genéticos y adquiridos. Se ha demostrado que la heredabilidad de la DM2 se encuentra alrededor del 70%. No obstante, los estudios realizados sólo han podido identificar una pequeña proporción del componente genético de la enfermedad. El resto del componente podría residir, en parte, en variantes genéticas poco frecuentes (MAF <0,05) o variantes raras (MAF <0,01). Sin embargo, la implicación de estas últimas en la enfermedad no ha sido determinada por las limitaciones de los estudios genéticos realizados hasta el momento. Además, más del 90% de las mutaciones y polimorfismos asociados a enfermedades se sitúan en las regiones codificantes de proteínas, y por tanto, en el exoma. Por ello, una fracción importante de las variantes de baja frecuencia y raras puede localizarse en el exoma. Finalmente, este proyecto representa un trabajo pionero en cuanto a la técnica empleada y población a la que se dirige.

Objetivos: Identificar variantes genéticas presentes en el exoma, incluyendo variantes poco frecuentes y raras, en relación a la DM2 en población española.

Metodología: Secuenciación del exoma (cobertura: 20x) en 200 pacientes con DM2 y 200 controles sanos de población española; todos con un índice de masa corporal de 25 a 34,9 kg/m² y una edad comprendida entre 40 y 65 años. Las regiones exónicas fueron capturadas y secuenciadas empleando el sistema de secuenciación de nueva generación HiScanSQ de Illumina, generando lecturas de 2x100 pares de bases en cada dirección (*paired-end*). A continuación, mediante el análisis bioinformático se realizaron los controles de calidad, el alineamiento de las lecturas contra el genoma de referencia y la identificación de las variantes genéticas. Se seleccionaron aquellas variantes con un MAF <20% presentes en controles o en casos, una calidad de genotipado (Q) >20 y con valores predictivos de alteración de la funcionalidad, SIFT y PolyPhen, significativos. Posteriormente, se estudiaron las variantes genéticas de mayor impacto funcional que estuvieran presentes en controles o en casos; centrándonos en las variantes de codón

de stop o parada, variantes de *splicing*, variantes de cambio de sentido o *missense*, y variantes en regiones codificantes para microARNs. Finalmente, las variantes de codón de stop o parada fueron verificadas mediante el método de secuenciación de Sanger.

Resultados: La secuenciación del exoma generó aproximadamente 1.000 GB de datos iniciales que tras el análisis bioinformático se convirtieron en unos 2.000 GB en total. Se identificaron 21.822 SNPs en controles y 17.238 SNPs en casos con efecto funcional, presentes únicamente en uno de los grupos y cumpliendo con los criterios de filtrado establecidos. En concreto, 160 y 132 SNPs fueron identificados como variantes de *splicing*, en controles y casos, respectivamente. Mientras, 1.817 y 1.614 SNPs fueron identificados como variantes *missense* para el grupo control y diabéticos, respectivamente. Además, 23 SNPs fueron identificados en secuencias codificantes para microARNs en controles y 11 en casos. Finalmente, 102 variantes de codón de stop fueron identificadas en controles y 50 en casos.

Conclusiones: Se han identificado un gran número de variantes genéticas que pueden estar implicadas en el desarrollo de DM2 o en la protección frente a la misma, incluyendo nuevas variantes genéticas de baja frecuencia y raras. Con la finalidad de identificar variantes genéticas válidas será necesario validar los resultados obtenidos mediante otras estrategias, su replicación en un amplio número de muestras de pacientes control y diabéticos así como el diseño de experimentos funcionales.

Identification of low-frequency and rare genetic variants related to type 2 diabetes mellitus by exome sequencing

Type 2 diabetes mellitus (T2DM) has become a global epidemic contributing significantly to morbidity and premature mortality. T2DM is the result from the interaction of different factors, including environmental, genetic and acquired. It has shown that the heredability of T2DM is around 70%. Thus far genetic studies can explain only a fraction of the estimated genetic component of the disease. The rest of it could be explained in low-frequency variants (MAF <0,05) and rare variants (MAF <0,01). However, the implications of these variants in the disease have not been determined yet due to the limitations of genetic studies conducted. Moreover, more than 90% of mutations and polymorphisms associated with diseases are in the protein coding region, the exome. Therefore, an important fraction of low-frequency and rare variants could be found in the exome. Finally, this project represents a novel study because of the approach and population used.

Objectives: To identify genetic variants in the exome, including low-frequency and rare variants, in relation to T2DM in a Spanish population.

Methodology: Exome sequencing (coverage: 20x) in 200 patients with T2DM and 200 Spanish healthy controls; all subjects had a body mass index between 25-34.9 kg/m² and were 40 to 65 years old. Exome regions were captured and sequenced by next-generation sequencing technology using Illumina HiScanSQ system to generate 2x100 bp paired end reads. A bioinformatic analysis pipeline was used to perform quality controls, to align the reads to a reference genome and identify genetic variants. We selected variants with MAF <20%, present in controls or cases, with genotyping quality (Q) >20 and significant effect predictors, SIFT and PolyPhen. We analysed genetic variants with an important functional consequences which were present in controls or cases. Specifically, we focused in the analysis of splicing variants, missense variants, variants in mature microRNAs coding sequences and stop variants. Finally, we verified the stop variants by Sanger sequencing.

Results: Exome sequencing approximately generated 1,000 GB data which after the bioinformatics analysis became around 2,000 GB in total. It was identified 21,822 SNPs in controls and 17,238 in cases with a functional effect, present only in controls or cases that meet quality criteria. In particular, 160 and 132 SNPs were splicing variants which were identified in controls and cases, respectively. While 1,817 and 1,614 SNPs were missense variants in controls and cases, respectively. Furthermore, 23 SNPs were identified as mature microRNA variants in controls and 11 SNPs in cases. Finally, we identified 102 SNPs as stop variants in controls and 50 SNPs in cases.

Conclusions: We have identified a large number of genetic variants, including low-frequency and rare variants, which may be involved in the development of T2DM or in the protection from it. In order to establish the true genetic variants involved in the disease we will need to validate them by different strategies, replication in a large sample of controls and diabetics as well as carrying out functional studies.

La diabetes mellitus (DM) es un síndrome o conjunto de síndromes que se caracterizan por una alteración metabólica cuyo marcador es la hiperglucemia crónica, resultante de defectos en la secreción de insulina, de la acción de la insulina o de ambas. La hiperglucemia produce daño a nivel celular que finalmente ocasiona una alteración, disfunción e insuficiencia de diversos órganos, especialmente de los ojos, riñones, nervios, corazón y vasos sanguíneos.

La DM es una enfermedad compleja resultado de la interacción de muchos factores, incluyendo ambientales, adquiridos y genéticos. Además, es una de las principales enfermedades crónicas en el momento actual y constituye un problema sanitario importante al haberse convertido en una epidemia mundial; contribuyendo de forma importante a la morbi-mortalidad prematura. Los nuevos hábitos de vida, fundamentalmente el exceso de ingesta calórica y el sedentarismo (involucrados en un aumento de la obesidad), junto con el envejecimiento de la población, permiten predecir un gran incremento de esta enfermedad en las próximas décadas.

El desorden de la regulación metabólica que acompaña a la DM provoca alteraciones fisiopatológicas secundarias en muchos sistemas orgánicos. La morbilidad viene dada en primer lugar, por las lesiones microvasculares (retinopatía diabética (RD), nefropatía diabética (ND) y neuropatía diabética (NRD)), relacionadas íntimamente con los años de evolución de la enfermedad y control de glucemia. En segundo lugar, por las lesiones macrovasculares; los pacientes diabéticos tienen un riesgo más elevado de cardiopatía isquémica (CI), accidente cerebrovascular (ACV) y enfermedad vascular periférica. Todas relacionadas y/o causantes de una mayor mortalidad de los pacientes (1).

La presencia de complicaciones asociadas a la DM depende en buena medida del grado de control metabólico, entendido éste no sólo por la glucemia, sino también por el resto de los factores de riesgo asociados a la enfermedad. Para ello es fundamental una asistencia sanitaria multidisciplinar de calidad, junto con la implicación del propio paciente en todo este proceso. Diversos consensos y guías de práctica clínica así lo recomiendan (2, 3).

1. EPIDEMIOLOGÍA DE LA DIABETES MELLITUS

Mientras la prevalencia de la DM varía con la edad, sexo y población, la Federación Internacional de Diabetes (FID) estima que la prevalencia mundial es actualmente mayor del 8,3% y se estima que más de 387 millones de adultos tienen diabetes. Además, más del 77% de las personas con diabetes viven en países de ingresos medianos o bajos y el incremento previsto es de más de 205 millones de nuevos casos para el año 2035 (Figura 1) (4).

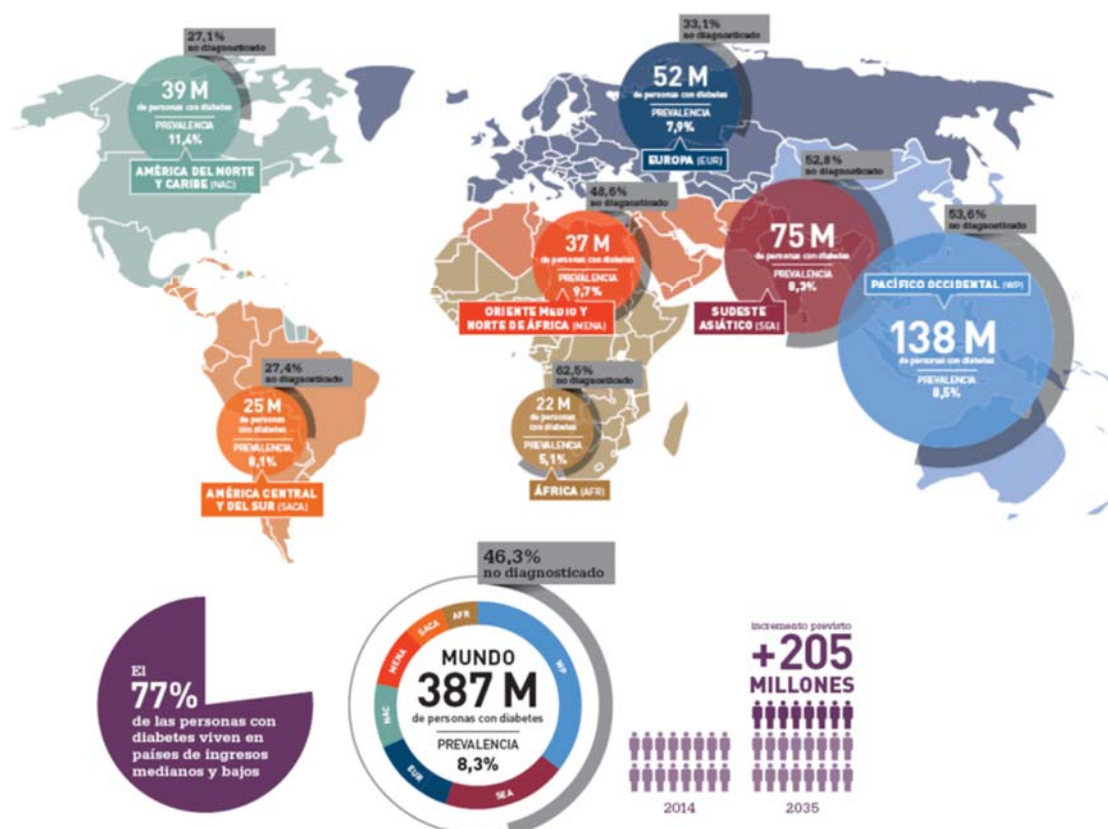


Figura 1. Indicadores sobre diabetes por países de acuerdo a la Federación Internacional de Diabetes (FID), 2014.

En Europa, se estima que el número de personas con diabetes es de 56,3 millones, el 8,5% de la población adulta. Los países con el mayor número de personas con diabetes están en su mayor parte en Europa Occidental, como Alemania, España, Italia, Francia y el Reino Unido (Figura 2).

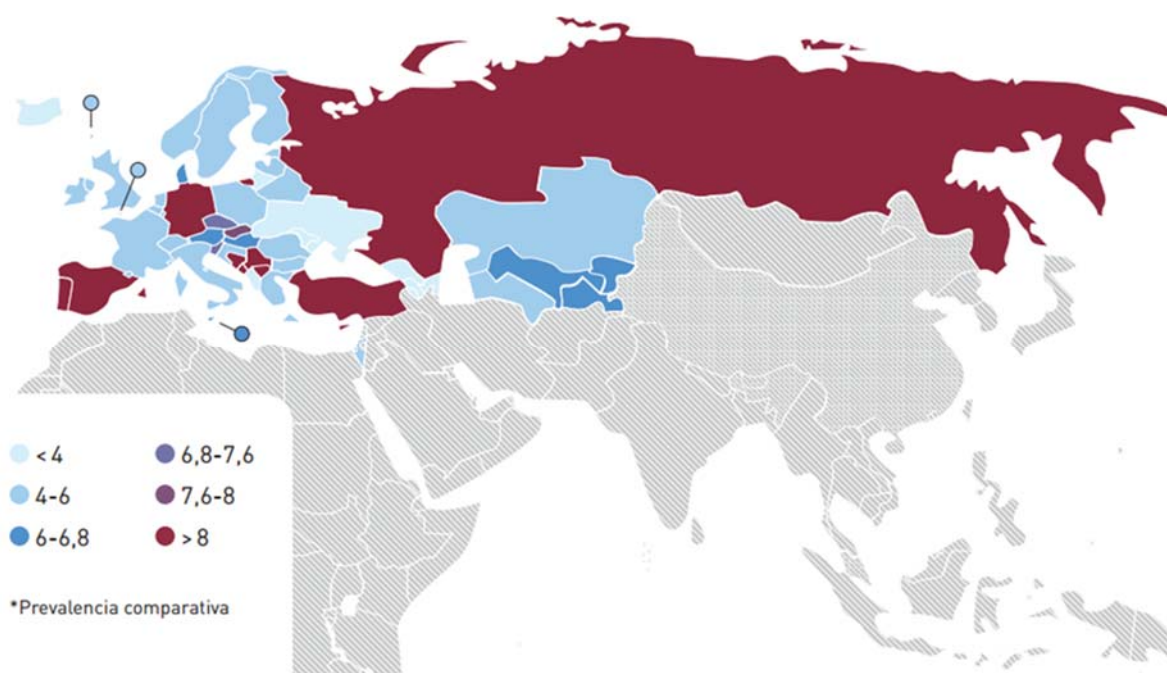


Figura 2. Estimaciones de prevalencia comparativa (%) de diabetes (20-79 años) en Europa según la Federación Internacional de Diabetes (FID), 2013.

La edad es un factor de riesgo importante para la diabetes mellitus tipo 2 (DM2). En Europa, el 37% de la población tiene más de 50 años y se espera que aumente a más del 44% en 2035. En gran medida, la alta prevalencia de la DM2 y la intolerancia a la glucosa son una consecuencia del envejecimiento de la población.

Además, Europa tiene el mayor número de niños con DM tipo 1 (DM1) en comparación con otras regiones, aproximadamente 129.300 casos. También tiene una de las tasas más altas de incidencia de la DM1 en niños, con 20.000 nuevos casos por año. Los países con las cifras globales más elevadas de DM1 en jóvenes son el Reino Unido, Rusia y Alemania.

En el caso concreto de España, los últimos resultados de la FID estiman que la prevalencia nacional está entorno al 11% y el número de casos de diabetes en población adulta (20-79 años) es de alrededor de 3.700.000 personas (Tabla 1) (5). Sin embargo, los resultados procedentes del estudio Di@bet.es, primer estudio español que tiene como objetivo conocer con mayor precisión la prevalencia de DM en España, estima que entorno al 13,8% (IC 95%: 12,8, 14,7%) mayores de 18 años tienen DM (6).

Tabla 1. Indicadores de prevalencia de diabetes para España de acuerdo con la Federación Internacional de Diabetes (FID) en el año 2014.

Indicadores Prevalencia	
Casos de diabetes (20-79), en miles	3.704
Casos de diabetes no diagnosticada (20-79), en miles	1.260
Prevalencia nacional de diabetes (%)	10,58
Prevalencia comparativa de diabetes (%)	7,85
Muertes relacionadas con la diabetes (20-79), en miles	24.428

Por todo ello, la DM es una enfermedad que aumenta progresivamente año tras año, siendo las principales causas el envejecimiento de la población y el incremento de la obesidad, causado por profundos cambios en el estilo de vida (sedentarismo, alimentación con alto contenido calórico y grasas saturadas).

2. CRITERIOS DIAGNÓSTICOS

Los criterios de diagnóstico para la DM según la ADA (*American Diabetes Association*) se basan en la presencia de uno de los siguientes criterios:

- Glucemia al azar ≥ 200 mg/dl ($\geq 11,1$ mmol/l) en presencia de síntomas de diabetes (poliuria, polidipsia o pérdida de peso inexplicada) o,
- Glucemia en ayunas (al menos durante 8 horas) ≥ 126 mg/dl (7,0 mmol/L) o,
- Glucemia ≥ 200 mg/dl ($\geq 11,1$ mmol/l) a las 2 horas tras la sobrecarga oral con 75 gr de glucosa (SOG) o,
- Hemoglobina glicada (HbA1c) $\geq 6,5\%$.

La glucemia se determina en plasma venoso por métodos enzimáticos y la HbA1c siguiendo un método similar al de la *International Federation of Clinical Chemistry* (ADA, (2); WHO, (7); *International Expert Committee*, (8); Jeppsson JO, (9)). Sin embargo, no se puede utilizar la HbA1c como test diagnóstico en pacientes con anemia o hemoglobinopatía (10, 11).

Cuando los niveles de glucemia de un paciente se encuentran elevados pero no alcanzan las cifras diagnósticas de diabetes, se clasifican como:

- Glucemia basal alterada (GBA) o alteración de la glucosa en ayunas (AGA): paciente con niveles de glucemia en ayunas entre 100-125 mg/dl, según la ADA (2); y entre 110-125 mg/dl según la Organización Mundial de la Salud (7).
- Intolerancia a la glucosa (ITG) o tolerancia anormal a la glucosa (TAG): pacientes con niveles a las 2 horas de la SOG entre 140-199 mg/dl (ADA; WHO) (2, 7).
- Riesgo elevado de desarrollar diabetes: pacientes con HbA1c entre 5,7-6,4% (ADA) (2).

También se utiliza el término “prediabetes” para describir la condición de las personas con un estado intermedio, donde los niveles de glucosa en sangre están por encima de los valores normales pero no son lo suficientemente altos para causar diabetes. La TAG se refiere a altos niveles de glucosa en sangre después de comer; mientras que la AGA se define como alta concentración de glucosa en sangre tras un período de ayuno (12).

3. TIPOS DE DIABETES

Existen varias clasificaciones útiles, sin embargo, tres parámetros pueden tenerse en cuenta para clasificar la enfermedad (2): 1) basándose en la patofisiología, 2) basándose en un defecto genético específico o 3) basándose en otro fenotipo común. En el caso de la diabetes todavía no hay una clasificación ideal y va evolucionando de acuerdo al conocimiento que se va obteniendo en relación a su etiología y patogénesis (13, 14).

3.1. Diabetes mellitus tipo 1

La DM1, anteriormente denominada diabetes mellitus insulino dependiente o juvenil, se debe a la destrucción de las células β del islote de Langerhans pancreático que conduce a una deficiencia de insulina. La DM1 representa entre el 5 y 10% de todos los casos de DM en el mundo. La presentación de este tipo de diabetes tiene una importante variabilidad global en relación con su prevalencia e incidencia. Un estudio reciente en España estima que la incidencia media de la DM1 en menores de 15 años sería de 17,69 casos/100.000 habitantes-año (15). Aunque la DM1 normalmente representa tan sólo

una minoría de la carga total de la diabetes en la población, es la forma predominante en los grupos de edad más jóvenes en la mayoría de los países desarrollados.

La DM1 se clasifica en los siguientes subtipos:

- **DM1 A o autoinmune:** enfermedad autoinmune en la que existe una destrucción selectiva de las células β del páncreas mediada por linfocitos T activados en sujetos con haplotipos HLA de predisposición. La respuesta autoinmune se realiza contra determinadas moléculas del islote: insulina, glutamato-descarboxilasa, tirosina-fosfatasa (IA-2); carboxipeptidasa H; ICA69, etc. Los autoanticuerpos contra estos antígenos pueden ser detectados en el suero de pacientes con DM1, y esto ha sido utilizado como una ayuda en el diagnóstico, clasificación y predicción de la enfermedad (16).

- **DM1 B o idiopática:** como entidad de reciente descripción se conoce poco de su etiología, evolución y pronóstico. Como contraposición a la DM1 A, describe a aquellos pacientes con deficiencia de insulina inicial, tendencia a la cetosis o cetoacidosis, en los que no se encuentran datos de autoinmunidad ni haplotipos HLA de predisposición (17). Cabe señalar que la deficiencia de insulina puede ser fluctuante a lo largo de la enfermedad, pero en algunas poblaciones (japonesa) puede tener un carácter fulminante (18). Inicialmente, y con un fuerte componente familiar, se ha descrito como más frecuente en la población afroamericana, asiática o hispana estadounidense (19). Hoy día existen pocos datos sobre su existencia y características en población española.

3.2. Diabetes mellitus tipo LADA (*Latent Autoimmune Diabetes of Adults*)

La diabetes mellitus tipo LADA tiene una base autoinmune y se caracteriza por mantener suficiente secreción insulínica para evitar la cetoacidosis durante períodos de tiempo prolongados. Sin embargo, y a pesar de tener una patogenia similar a la de la DM1, se presenta clínicamente en los adultos, pudiéndose confundir en muchos casos como DM2. Se calcula que alrededor del 10% de los diabéticos adultos diagnosticados como tipo 2 padecen, en realidad, esta forma de diabetes autoinmune, tipo 1, de evolución muy lenta (20). El perfil clínico de estos pacientes se caracteriza por ser sujetos más jóvenes que los tipo 2 clásicos en el momento del diagnóstico, habitualmente no son obesos, y muestran pocas o ninguna de las alteraciones que componen el síndrome

metabólico. El principal marcador de autoinmunidad en la diabetes LADA son los anticuerpos contra la descarboxilasa del ácido glutámico (anti-GAD), a veces asociados a anticuerpos contra antígenos citoplasmáticos de las células insulares (21) . LADA representa el 2-12% de todos los casos de diabetes (22).

3.3. Diabetes mellitus gestacional

La diabetes mellitus gestacional (DMG) se define como una alteración de la tolerancia a la glucosa que se inicia o se reconoce durante el embarazo (23). Generalmente, la DMG suele aparecer en estadios tardíos del embarazo, alrededor de la semana 24. La condición se produce debido a que la acción de la insulina es bloqueada, en parte, por las hormonas generadas por la placenta, generando resistencia a la insulina (24-26).

La DMG es la complicación metabólica más frecuente del embarazo (27), con una prevalencia creciente en relación directa con el aumento de la obesidad (28). La prevalencia varía ampliamente entre las diferentes poblaciones de todo el mundo. Gran parte de la variabilidad es debida a las diferencias en los criterios de diagnóstico y las poblaciones de estudio. En España, la prevalencia de la DMG se estima entre el 2,5 y el 7,3% en población obstétrica general, lo que representa un 90% de todos los embarazos de riesgo (29).

La DMG normalmente desaparece después del parto. Sin embargo, las mujeres que han tenido DMG tienen un mayor riesgo de desarrollar diabetes gestacional en embarazos posteriores y de desarrollar DM2 más adelante en la vida (30, 31). Los bebés que nacen de madres con DMG también tienen un mayor riesgo de obesidad y DM2 en la adolescencia o en la edad adulta temprana (32). Las mujeres con DMG tienen que vigilar y controlar sus niveles de glucosa en sangre para reducir al mínimo los riesgos para el bebé. Normalmente, esto se puede hacer mediante la adopción de una dieta sana y ejercicio moderado, aunque en algunos casos puede ser necesario también administrar insulina o medicación oral.

3.4. Otros tipos de diabetes mellitus

3.4.1. Diabetes mellitus monogénicas

La diabetes monogénica es aquella diabetes causada por mutaciones producidas en un solo gen y que presentan un patrón de herencia autosómica.

- **Diabetes mellitus neonatal**

La diabetes mellitus neonatal (DMN) es una diabetes que se diagnostica en los primeros 6 meses de vida y no tiene causa autoinmune. Se debe a una disfunción de las células β originada por alteraciones en genes que regulan la síntesis o secreción de insulina. Se han descrito numerosos genes en la etiopatogenia de la DMN pero el 80-90% de los casos se originan por alteraciones localizadas a 3 niveles: genes situados en el brazo largo del cromosoma 6 (6q24), genes que codifican las subunidades del canal de K^+ sensible al ATP (subunidad SUR1 y subunidad Kir 6.2), y en el gen de la insulina (33-36). Clásicamente se han descrito dos formas clínicas de presentación de la enfermedad: transitoria (DMNT), con resolución en los primeros 18 meses de vida y predisposición a diabetes posterior, y permanente (DMNP). Desde el punto de vista clínico, el 50-60% de los casos es DMNT (37).

- **Diabetes mellitus tipo MODY (*Maturity-onset diabetes of the young*)**

Es una forma de diabetes familiar de herencia autosómica dominante, caracterizada por una disfunción de la célula β que, en general, se manifiesta durante la infancia o la juventud. Se debe a defectos genéticos en factores que intervienen en la formación y función de dicha célula β , pero no suele asociarse a obesidad como la DM2 (38).

Su etiología genética es diferente, pues se han descrito hasta la actualidad doce subtipos (39, 40). Las mutaciones del gen de la glucoquinasa (*GCK*) y del factor heptonuclear 1A (*HNF1A*) son las más frecuentes; las menos comunes se localizan en el factor heptonuclear 4A (*HNF4A*), en el factor heptonuclear 1B (*HNF1B*) y en el gen de la insulina (*INS*) (41). El tipo MODY supone entre el 1 y el 5% de todos los tipos de diabetes.

Dos de los MODY están causados por mutaciones en genes que codifican enzimas: el MODY 2 (consecuencia de alteraciones en el gen que codifica la GCK) (42) y el MODY 7 (por mutaciones en el gen de la carboxil éster lipasa [*CEL*]) (43).

Otros cinco genes implicados codifican factores de transcripción de la célula β (43): *IPF-1* (MODY 4), *HNF1A* (MODY 3), *4A* (MODY 1) y *1B* (MODY 5), que se expresan en el hígado e islotes pancreáticos y desempeñan un papel fundamental en la normalidad del desarrollo y la función de las células β (44). También se ha identificado el *NeuroD1/BETA2* (MODY 6), que interviene en el desarrollo pancreático y en la transcripción del gen de la insulina.

Recientemente, se han descrito nuevas formas de MODY en un número limitado de familias causadas por mutaciones en genes que codifican para una glicoproteína secretada por el páncreas (MODY8-*CEL*), un factor de transcripción (MODY9-*PAX4*), insulina (MODY10-*INS*), una tirosina quinasa (MODY11-*BLK*) y un transportador (MODY12-*ABCC8*) (39, 40, 45).

Los sujetos con distintos tipos de MODY presentan características clínicas diferentes que se expresan a distintas edades, aunque los defectos estén presentes desde el nacimiento. Existen estudios que relacionan la edad de comienzo de la diabetes con la presencia de la misma durante el embarazo (46). Parece que la hiperglucemia durante la vida fetal puede acelerar el comienzo de la diabetes interfiriendo en la secreción de insulina en fases posteriores de la vida, ya que los islotes fetales son muy sensibles al efecto tóxico de la glucosa. La medida de los niveles de glucosa en ayunas así como la prueba de SOG son claves para el diagnóstico de la enfermedad, reflejando no sólo el grado de hiperglucemia sino también la causa subyacente (47).

• **Diabetes mellitus mitocondrial**

La diabetes mitocondrial se debe a mutaciones en el ADN mitocondrial (la mejor conocida es la causada por la mutación A3243G) que impiden la entrada de leucina en la mitocondria, afectando a la cadena respiratoria (48, 49). Aunque la mutación está presente en todas las células, la expresión clínica afecta solamente a las células β y a las del oído interno. En las primeras, se ha observado una menor producción de ATP y, como consecuencia, de insulina. Clínicamente se caracteriza por herencia materna, diabetes de aparición en adultos jóvenes (sobre los 40 años) generalmente leve, sordera

neurosensorial y antecedentes maternos de diabetes. No se encuentra insulino-resistencia pero la evolución natural es hacia la deficiencia de insulina (50). Más rara vez, la diabetes se asocia a miopatía, encefalopatía, acidosis láctica y accidentes isquémicos cerebrales. Algunos casos del llamado síndrome de Wolfram se han descrito asociados a mutaciones en el ADN mitocondrial (51).

3.4.2. Diabetes sindrómicas

Las formas sindrómicas hereditarias más comunes de la DM incluyen síndromes con degeneración pancreática, desórdenes endocrinológicos, errores innatos de la homeostasis de carbohidratos, alteraciones neurológicas, desórdenes citogenéticos, etc. Entre ellos, encontramos los síndromes genéticos de Wolfram, Wolcott-Rallison, Rabson-Mendenhall, Bardet-Biedl, Laurence-Moon, Alstrom o Prader-Willi (52).

3.4.3. Enfermedades del páncreas exocrino

La insuficiencia pancreática exocrina (IPE) es un trastorno grave que se produce cuando el páncreas no produce enzimas digestivas o las produce en cantidad insuficiente, lo que significa que el organismo no puede digerir los alimentos, causando una absorción inadecuada de nutrientes (53). La IPE es una consecuencia importante de un amplio rango de enfermedades graves que alteran la función pancreática normal. Algunas de las enfermedades comúnmente relacionadas con la IPE son: pancreatitis crónica, pancreatectomía, hemocromatosis, neoplasias o la fibrosis quística. Se estima que de promedio un 51% de los pacientes con DM1 padecen una IPE, mientras que la presentan aproximadamente un 32% de los pacientes con DM2 (54).

3.4.4. Endocrinopatías

La compleja regulación de la homeostasis metabólica requiere la participación de muchas hormonas además de la insulina. Por ejemplo, la hormona de crecimiento (GH), que a través del factor de crecimiento insulino similar tipo 1 (IGF-1), controla diferentes aspectos del transporte de aminoácidos, la síntesis de proteínas y el metabolismo lipídico. El glucagón y las catecolaminas son las principales hormonas contra-regulatorias que protegen al organismo contra la hipoglucemia. El cortisol y la GH contribuyen en un paso posterior sólo si la hipoglucemia es severa y prolongada. Asimismo, el cortisol, por ser protector ante el estrés y las agresiones del medio, ejerce

por un lado un rol permisivo en la regulación fisiológica de la gluconeogénesis y, por el otro, un papel fundamental para proveer de una mayor disponibilidad de glucosa durante estas situaciones. Tanto la GH como las catecolaminas, el glucagón y el cortisol son hormonas contra-reguladoras, antagonistas de la acción de la insulina. Por consiguiente, la hipersecreción autónoma de cualquiera de las hormonas (en la acromegalia, enfermedad de Cushing, glucagonoma, etc.), puede ocasionar diabetes (55, 56).

3.4.5. Diabetes mellitus asociada a fármacos

Compuestos como vacor (raticida), pentamidina, ácido nicotínico, glucocorticoides, interferón- α , hormona tiroidea, agonistas β -adrenérgicos o tiazidas pueden inducir diabetes por su toxicidad directa sobre las células β , pero también induciendo resistencia insulínica periférica. Un ejemplo es la acción de la simvastatina y atorvastatina las cuales disminuyen la secreción de insulina por las células β (57, 58).

3.4.6. Diabetes mellitus asociada a infecciones

Algunos virus como el de la rubeola, coxsackievirus B, citomegalovirus, adenovirus y virus de las paperas, se asocian con la destrucción de las células β y promueven diabetes (59-62). Independientemente de los efectos secundarios de los medicamentos, la historia natural del virus VIH (Virus de la Inmunodeficiencia Humana) también se asocia al desarrollo de la diabetes (63).

3.5. Diabetes mellitus tipo 2

La DM2, anteriormente conocida como diabetes independiente de insulina o diabetes del adulto, es la forma mayoritaria explicando entre el 90-95% de casos de diabetes. La DM2 engloba individuos que presentan resistencia a la insulina o una deficiencia relativa de insulina, generando así una acumulación de glucosa en sangre.

Muchas personas con DM2 no son conscientes de su enfermedad durante mucho tiempo, ya que los síntomas pueden tardar años en aparecer o en ser reconocidos, pero durante este tiempo el organismo está siendo dañado por el exceso de glucosa en sangre. Estas personas suelen ser diagnosticadas sólo cuando las complicaciones de la diabetes ya se han desarrollado. A diferencia de las personas con DM1, la mayoría de las personas con DM2 no requieren, por lo general, dosis diarias de insulina para

sobrevivir, aunque pueden necesitarla en fases avanzadas de la enfermedad. Muchas personas pueden controlar su enfermedad a través de una dieta sana, una mayor actividad física y otra medicación. Sin embargo, si no son capaces de regular sus niveles de glucosa en sangre, puede que tengan que tomar insulina.

Las causas que desencadenan la DM2 se desconocen en el 70-80% de los casos; al parecer, influyen diversos factores genéticos (concordancia en gemelos monocigóticos superior al 90% y herencia de carácter poligénico), obesidad (85%), dislipidemia, hipertensión arterial, dieta rica en carbohidratos, factores hormonales y resistencia a la acción de la insulina en los tejidos periféricos, entre otros.

3.5.1. Prevalencia de la diabetes mellitus tipo 2 en población española

La DM2 ha sido catalogada como la epidemia del siglo XXI tanto por su creciente magnitud como por su impacto en la enfermedad cardiovascular (primera causa de mortalidad en las sociedades desarrolladas). El estudio Di@bet.es, el más reciente a nivel nacional incluye, de manera aleatoria, el estudio de 5.419 personas mayores de 18 años, y estima que entorno al 14% de la población adulta española tiene diabetes (6). Este resultado es ligeramente más elevado que los estudios descritos anteriormente (64-66). Varios factores, como el cambio de criterios diagnósticos, el envejecimiento de la población, la menor mortalidad de los diabéticos o un verdadero aumento de la incidencia, han podido influir en este fenómeno (67).

Los estudios realizados para estimar la incidencia de DM2 son menos numerosos que en el caso de la DM1 y, además, no siguen una metodología estandarizada, con lo que es más complicado establecer comparaciones entre ellos (68). Como ejemplo, un estudio reciente en población adulta de Madrid ha determinado que la incidencia de DM2 es de 3,5 casos/1.000 persona-año (69).

3.5.2. Aspectos clínicos y evolución de la diabetes mellitus tipo 2

Con el tiempo, la diabetes puede afectar diferentes órganos y sistemas del organismo, causando así graves complicaciones e incluso la muerte. Estas complicaciones pueden ser microvasculares (lesiones que afectan a los vasos sanguíneos de menor calibre) y macrovasculares (lesiones que afectan a vasos sanguíneos más grandes).

Las principales manifestaciones de las alteraciones microvasculares relacionadas con la DM son la retinopatía (RD), la nefropatía (NFD) y la neuropatía (NRD).

- **Retinopatía (RD):** puede estar presente en el momento del diagnóstico de la DM2, tiene una prevalencia que aumenta con el tiempo de evolución de la enfermedad (alcanzando el 60% tras 20 años de evolución) y es la complicación microvascular más frecuente (70, 71). Es una causa importante de ceguera y discapacidad visual. Está causada por el daño de los vasos sanguíneos de la retina, lo que ocasiona una pérdida progresiva de la vista, que a veces llega a ser ceguera. Se distinguen dos formas de RD: no proliferativa y proliferativa, según la ausencia o presencia de neovascularización en la retina, respectivamente.

- **Nefropatía diabética (NFD):** se manifiesta con un aumento de la excreción urinaria de albúmina y alteraciones en la filtración glomerular. La NFD está causada por lesiones en los capilares glomerulares. Ello puede causar insuficiencia renal y al final la muerte. En los países desarrollados, esta es una causa importante de diálisis y trasplante renal. Como la RD, puede estar presente en el momento de realizar el diagnóstico de DM2. Su prevalencia aumenta con el tiempo de evolución de la enfermedad, y se estima que a los 10 años oscila entre el 25 y el 40% (72, 73).

- **Neuropatía diabética (NRD):** es un grupo heterogéneo de alteraciones del sistema nervioso periférico que puede expresarse en cualquier localización corporal con alteraciones sensitivas, motoras, autónomas o mixtas. Pueden aparecer lesiones en los pies, por alteraciones de los vasos sanguíneos y nervios, dando lugar al conocido "Pie Diabético" (74).

Las enfermedades cardiovasculares o complicaciones macrovasculares suponen la principal causa de morbilidad y mortalidad entre las personas con DM (75, 76). Estas incluyen:

- **Arteriopatía de arterias carótidas y cerebrales:** se produce una disminución del flujo sanguíneo en el cerebro pudiendo producir ACVs graves, que pueden causar la parálisis de un lado del cuerpo o pequeños infartos cerebrales que irán causando trastornos progresivos de memoria.

• **Arteriopatía de las coronarias:** se produce la obstrucción de las arterias responsables del flujo sanguíneo al corazón, lo que puede generar un infarto de miocardio.

• **Arteriopatía periférica:** disminuye el flujo sanguíneo a las extremidades inferiores apareciendo un intenso dolor al caminar, pudiendo causar gangrena en los pies.

3.5.3. Prevención y diagnóstico de la diabetes mellitus tipo 2

Los datos actuales sugieren que 1 de cada 4 personas con DM2 es consciente de su enfermedad. La fase asintomática de la diabetes tiende a ser bastante larga y sólo algunos pacientes desarrollan complicaciones en estadio temprano, tales como la RD o NFD (77).

El diagnóstico precoz de la DM2 permite evitar las complicaciones agudas de la enfermedad, como, la cetoacidosis o el síndrome hiperosmolar hiperglucémico, cuya mortalidad no es baja (2 y 20-40% respectivamente). La ADA propone el cribado a partir de los 45 años mediante glucemia basal (en ayunas) en plasma (Tabla 2) (2).

Tabla 2. Detección precoz de la diabetes mellitus tipo 2 según la *American Diabetes Association (ADA)*.

DETECCIÓN PRECOZ DE LA DIABETES MELLITUS TIPO 2
Personas mayores de 45 años (repetir cada 3 años)
Anualmente, y a cualquier edad, en personas con algún factor de riesgo de diabetes:
- Índice de masa corporal >27 Kg/m ²
- Antecedentes de diabetes en familiares de primer grado
- Etnias de alto riesgo
- Macrosomía y/o diabetes gestacional previa
- HDL <35 mg/dl y/o triglicéridos >250 mg/dl
- Hipertensión arterial
- Diagnóstico previo de intolerancia a la glucosa o glucemia basal alterada
Embarazadas entre la 24-28 semanas

Varios ensayos han mostrado que los cambios en el estilo de vida, dieta y ejercicio pueden reducir de una manera significativa la incidencia de DM2 sobre todo en personas con “prediabetes” (78-81). Además, algunos medicamentos, como la metformina, puede prevenir la enfermedad en pacientes con estadios intermedios de la enfermedad (78, 79). Sin embargo, aunque la dieta y el ejercicio son beneficiosos, el *screening* para prevenir la enfermedad debería realizarse en personas con riesgo elevado, de forma que

pudiesen beneficiarse a tiempo de la medicación, evitando el desarrollo de la enfermedad. Los criterios específicos de diagnóstico para la prediabetes y la DM2, se muestran en la Tabla 3.

Table 3. Criterios diagnósticos de la diabetes mellitus tipo 2 según la *American Diabetes Association (ADA)*.

Diagnóstico	Hemoglobina A1c , nivel, %	Niveles de glucosa en ayunas mmol/L	mg/dL
Prediabetes	5,7–6,4	5,55–6,94	100–125
Diabetes	≥6,5	≥7,0	≥126

3.5.4. Factores de riesgo de la diabetes mellitus tipo 2

Los factores de riesgo para la DM2 son de origen multifactorial, siendo la obesidad y los cambios en el estilo de vida (sedentarismo y dieta) factores clave. Además, los factores de base genética, como la historia familiar de diabetes, participan también en el riesgo de desarrollar DM2 (Tabla 4).

Dentro de la complejidad de la DM2 podemos distinguir entre factores de riesgo reversibles o irreversibles (no modificables). El primer grupo incluye factores como la obesidad, la dieta o el sedentarismo; el segundo: la edad, raza u origen étnico, historia familiar (predisposición genética), antecedentes de DMG o bajo peso al nacer.

Tabla 4. Factores de riesgo de la diabetes mellitus de tipo 2 según la *American Diabetes Association (ADA)*.

FACTORES DE RIESGO DE LA DIABETES MELLITUS TIPO 2
Antecedente familiar de primer grado con DM2
Edad (>45 años)
Obesidad (IMC ≥ 27 kg/m ²)
Raza/Etnia
Glucosa basal anómala o alteración de la tolerancia a la glucosa identificadas previamente
Antecedentes de DMG
Hipertensión (tensión arterial $\geq 140/90$ mm Hg)
HDL <35 mg/dl o nivel de triglicéridos >250 mg/dl
Síndrome del ovario poliquístico

• **Obesidad:** es el factor más estrechamente asociado con el desarrollo de diabetes en el adulto, ya que se ha estimado que el riesgo atribuible a la obesidad es alrededor del 75%. La distribución de la grasa corporal tiene una gran importancia a la hora de definir a los individuos con mayor riesgo cardiovascular (82). La obesidad abdominal (por encima de la cintura), más frecuente en los varones, condiciona una mayor morbi-mortalidad cardiovascular, probablemente porque conlleva un incremento de la grasa abdominal visceral (83, 84). Sin embargo, el efecto del exceso de peso es reversible, y pérdidas del 5% lo disminuyen; así en la cohorte de Framingham se observó una reducción importante en la incidencia de DM2, incluso en personas obesas con intolerancia a la glucosa (81, 85).

• **Dieta:** los países en los que se ha producido una occidentalización de los hábitos dietéticos (alto consumo de hidratos de carbono simples, grasas saturadas y proteína con disminución de fibra), junto con un mayor sedentarismo, muestran un aumento en la prevalencia de obesidad y diabetes. Parece que el contenido de grasas de la dieta puede ser un factor de riesgo importante, se ha comprobado que su reducción disminuye el riesgo independiente de la pérdida de peso (79). Por otro lado, el consumo moderado de alcohol puede reducir el riesgo de desarrollar diabetes (86), pero se ha demostrado que el tabaquismo constituye un factor de riesgo para la enfermedad (87).

• **Sedentarismo:** los hábitos sedentarios y el bajo nivel de ejercicio físico son un factor de riesgo para el desarrollo de la DM2. El ejercicio físico disminuye la resistencia a la insulina, contribuyendo a mejorar el metabolismo hidrocarbonado. La práctica de ejercicio de forma regular es una medida preventiva. Se ha demostrado capaz de disminuir de forma significativa, independientemente de la pérdida de peso, la aparición de diabetes en mujeres obesas y con antecedentes familiares (82) y también cuando ya existe una tolerancia alterada a la glucosa (79, 81).

• **Raza/Etnia:** existen grandes diferencias étnicas y geográficas en cuanto a la prevalencia de DM2. Mientras que unas poblaciones parecen muy castigadas (en los indios Pima o los micronesios de Nauru es superior al 30%), en otras zonas es muy poco frecuente (África Central, 0,3%) (75, 88).

• **Diabetes gestacional:** la DMG en las mujeres normalmente desaparece después del nacimiento, como anteriormente se ha señalado. Sin embargo, las mujeres que han tenido DMG tienen un mayor riesgo de desarrollar diabetes en embarazos posteriores y de DM2 más adelante en la vida. Los bebés que nacen de madres con DMG también tienen un mayor riesgo de obesidad y DM2 en la adolescencia o en la edad adulta temprana (89, 90).

• **Antecedentes familiares:** la presencia de antecedentes familiares y por tanto, predisposición genética, es un importante factor de riesgo. Estudios de seguimiento han estimado que la concordancia entre gemelos idénticos en la presencia de DM2 puede ser mayor del 76%, cuando se realiza un seguimiento de hasta 15 años (91).

Desde un punto de vista genético y epigenético, el fondo genético del individuo es uno de los principales factores de susceptibilidad que puede determinar la respuesta frente a un exceso energético y el desarrollo de la DM2 (92, 93). Individuos con sobrepeso u obesos no desarrollan diabetes o la desarrollarán muy tarde durante su vida, lo que se podría considerar como individuos resistentes a la DM2. Estos individuos presentan las siguientes características: células β pancreáticas funcionales, niveles de nutrientes en sangre correctos, predomina el tejido adiposo subcutáneo frente al visceral, ausencia de resistencia a insulina e incremento limitado de grasa hepática, etc (94, 95).

Sin embargo, los individuos con susceptibilidad genética al exceso calórico desarrollan DM2 debido a la insuficiencia de respuestas adaptativas para gestionar el exceso energético, presentando defectos metabólicos cruciales para el desarrollo de la DM2: resistencia periférica a la insulina, aumento de la producción de glucosa endógena, incapacidad de las células β pancreáticas para compensar el exceso energético, aumento en la secreción de glucagón y respuesta reducida de incretina, alteración del tejido adiposo subcutáneo, depósitos de tejido adiposo visceral en órganos e inflamación del tejido adiposo (Figura 3).

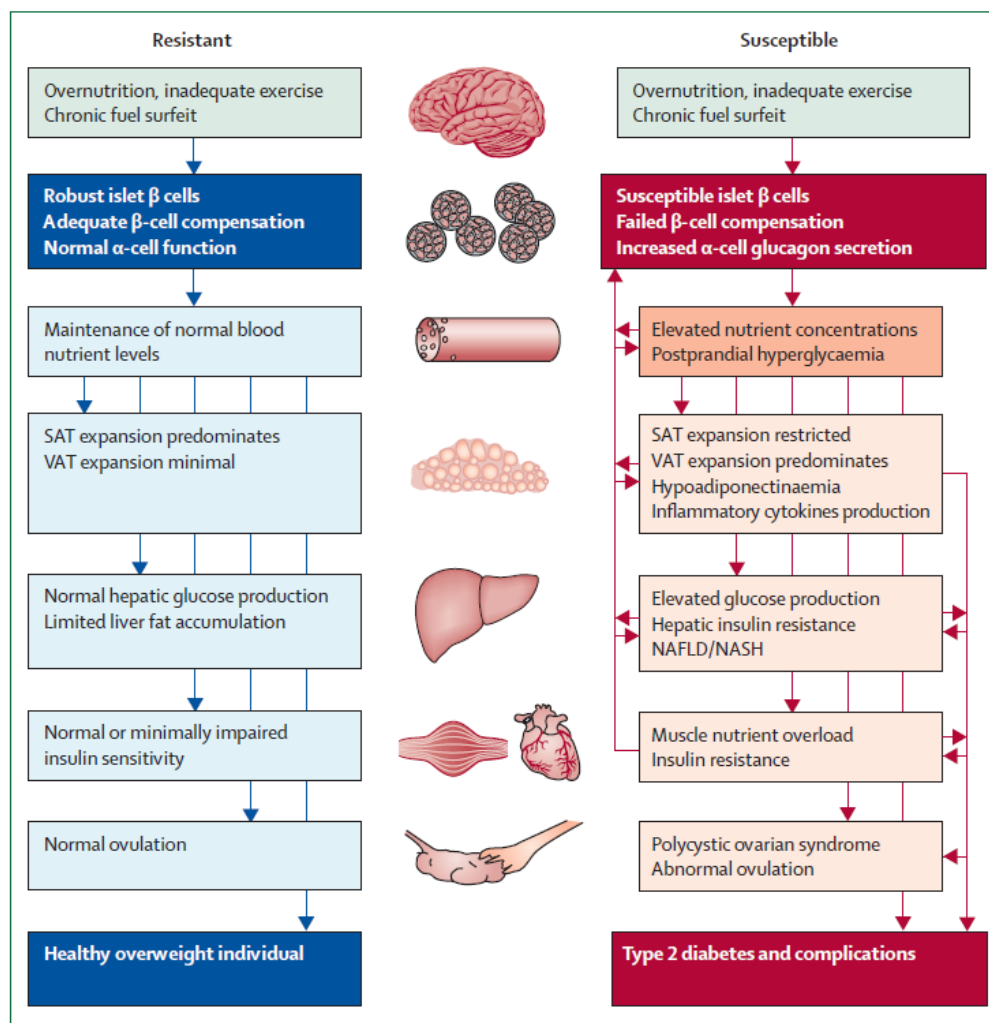


Figura 3. Diabetes mellitus tipo 2 y sus complicaciones de acuerdo a la susceptibilidad genética. Individuos genéticamente resistentes son capaces de compensar el exceso energético gracias a que sus células β pancreáticas son funcionales. De este modo los niveles de nutrientes en sangre se mantienen normales sin afectar a órganos como el hígado, músculo esquelético, corazón u ovarios. Por el contrario, los individuos con susceptibilidad a la diabetes presentan un exceso energético que no es capaz de ser compensado por las células β pancreáticas y que conducen a un estrés y producción de lesiones a nivel de órganos. SAT= Tejido adiposo subcutáneo (*subcutaneous adipose tissue*). VAT= tejido visceral adiposo (*visceral adipose tissue*). NAFLD= enfermedad del hígado graso no alcohólico (*non-alcoholic fatty liver disease*). NASH=esteatohepatitis no alcohólica (*non-alcoholic steatohepatitis*). Nolan et al (2011) *Lancet*; 378(9786):169-8.

3.5.5. Mecanismos moleculares y fisiopatología de la diabetes mellitus tipo 2

El desarrollo de la DM2 se puede producir mediante diferentes procesos (96-98). En general está mediado por la resistencia a la insulina (RI) asociada a la obesidad y el fallo de la célula β pancreática (96). Para entender en profundidad los mecanismos celulares y moleculares responsables de la DM2 es necesario entender los mecanismos implicados en la homeostasis de glucosa e insulina (Figura 4).

Liberación y acción de la insulina

La liberación de insulina es un proceso indispensable en la homeostasis del cuerpo como respuesta al aporte energético por el consumo de alimentos. Su liberación es inducida principalmente en respuesta al incremento de glucemia, pero al mismo tiempo es regulada por diversas sustancias (hormonas gastrointestinales y pancreáticas, neurotransmisores del sistema nervioso autónomo, etc). La glucosa, aminoácidos, ácidos grasos y cuerpos cetónicos favorecen la secreción de insulina, al igual que la activación del receptor β_2 -adrenérgico y la estimulación del nervio vago, mientras que los receptores α_2 -adrenérgicos inhiben la liberación de insulina (99-101).

La despolarización de la célula β provoca la liberación de insulina. El proceso se inicia con el aumento de la concentración plasmática de carbohidratos: la fructosa y la glucosa entran en la célula β a través del transportador de glucosa 2 (GLUT2) y el transportador de fructosa (GLUT5). El GLUT2 es un transportador de glucosa de baja afinidad, se expresa en el hígado, riñón, células β del páncreas y en la membrana basolateral de las células epiteliales del intestino delgado (102, 103). El GLUT-2 actúa como un regulador permitiendo únicamente la entrada de glucosa cuando está lo suficiente elevada en plasma (>70 mg/dL) como para requerir la liberación de una cantidad significativamente importante de insulina. Después de la ingesta de alimento, el hígado, por su parte, es capaz de incorporar la glucosa a través del GLUT2 para convertirla rápidamente en glucógeno. De forma inversa, durante el periodo postprandial tardío (periodo comprendido entre las 6 y 8 horas de ayuno), el glucógeno sufre una degradación para generar moléculas de glucosa, que salen de la célula hepática a la circulación sistémica, preservando de esta manera la glucemia en valores fisiológicos. Por lo anterior, el GLUT2 es un transportador bidireccional que puede transportar glucosa desde la sangre al tejido o desde el tejido hacia la sangre, según se requiera. Recientemente se ha descubierto que GLUT2 tiene también la capacidad de transportar fructosa, como el transportador clásico de fructosa GLUT5 que se expresa en las células de la mucosa intestinal, donde modula la absorción de fructosa desde el lumen a la célula epitelial intestinal (104).

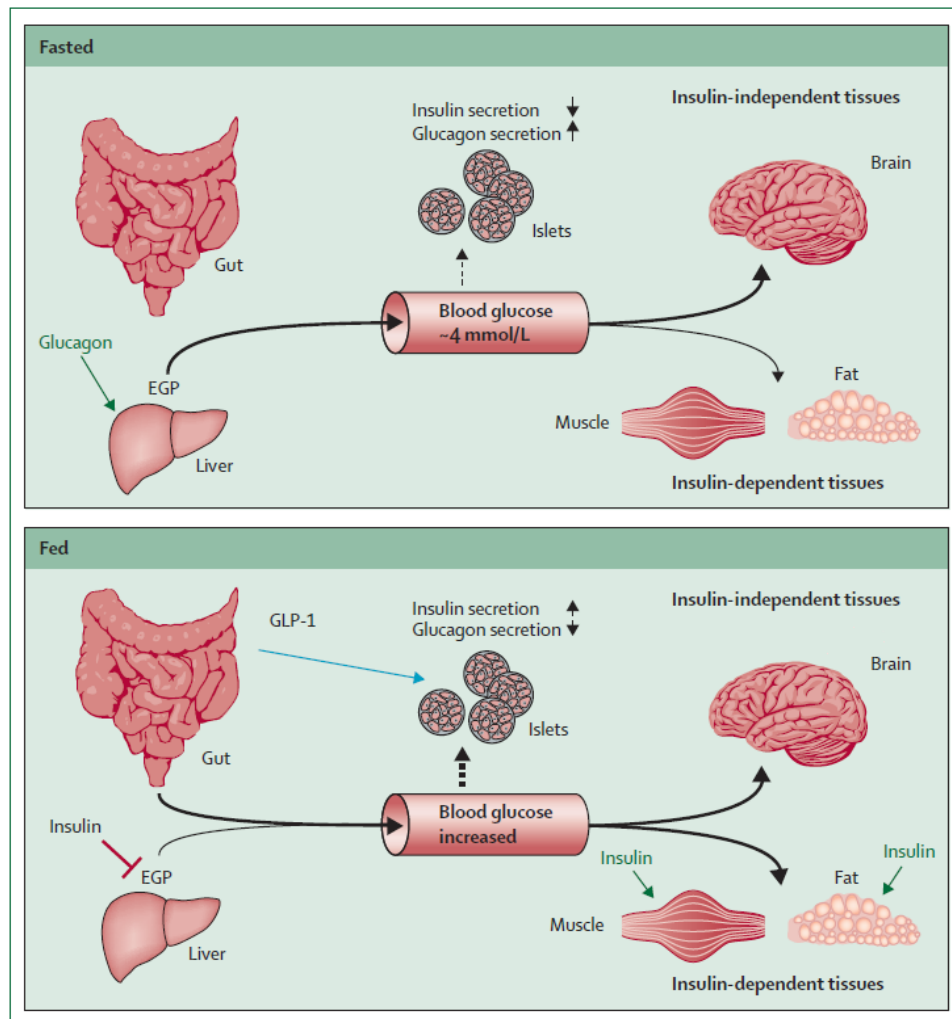


Figura 4. Visión general de la homeostasis normal de la glucosa. En ayunas, la concentración de glucosa en sangre está determinada por el equilibrio entre la producción endógena de glucosa (EPG); principalmente a través de la glucogenólisis y la gluconeogénesis hepática, y la utilización de tejidos independientes de insulina, como el cerebro. La EPG evita la hipoglucemia y es acompañada por un bajo ratio insulina/glucagón en plasma. Durante la ingesta de alimentos con hidratos de carbono, las concentraciones de glucosa en sangre se elevan, lo que estimula la secreción de insulina por las células β pancreáticas, suprimiendo la secreción de glucagón por las células α . La EPG se inhibe y la absorción de insulina en los tejidos periféricos como el corazón, músculo esquelético y tejido adiposo se activa. Procesos neurohormonales, como la secreción de incretinas tales como la GLP-1, aumenta la secreción de insulina y disminuye secreción de glucagón. *Adapt. Nolan et al (2011) Lancet; 378(9786):169-8.*

Tras el ingreso de la glucosa (o fructosa) al interior de la célula β mediante GLUT2, ésta es fosforilada a glucosa-6-fosfato (G-6-P) por la glucoquinasa. Este proceso determina la velocidad de glucólisis y de los subsecuentes procesos oxidativos que culminan con el incremento en la relación ATP/ADP del citosol. La despolarización de la célula ocurre a causa del cierre de los canales de K^+ sensibles a ATP (K_{ATP}), incrementando el potencial de membrana hasta alcanzar la apertura de canales de Ca^{2+}

citosólico. Finalmente, ello induce la fusión de la vesícula exocítica que contiene insulina con la membrana plasmática (99, 105).

El canal K_{ATP} es un octámero compuesto de cuatro subunidades Kir 6.2 y cuatro SUR1. La subunidad Kir 6.2 se encarga de la respuesta inhibitoria inducida por la unión con ATP. La subunidad SUR1 tiene sitios de unión para el ADP y el diazóxido (que favorecen la apertura del conducto), así como para las sulfonilureas y meglitinida (ambas inhiben la apertura del conducto). Mutaciones en estas subunidades alteran la liberación de insulina (106).

Transducción de señales del receptor de insulina

La insulina es un miembro de la familia de péptidos denominada “factores de crecimiento insulinoides”. El receptor de insulina está presente en todas las células de los mamíferos y está conformado por dos unidades α y dos subunidades β . Las subunidades α son extracelulares y tienen el sitio de unión a ligando, mientras que las subunidades β atraviesan la membrana, tienen un dominio con varios residuos de tirosina, un dominio tirosina quinasa y un sitio de unión de ATP. Cuando se une la insulina al receptor, la subunidad α influye en la β para accionar la tirosina quinasa, la cual se autofosforila en residuos de tirosina, iniciando la actividad quinasa contra otras proteínas, como pueden ser los sustratos proteicos del receptor de insulina (IRS-1 a 4) que junto con la proteína Shc participan como proteínas de andamiaje para otras.

El receptor de la insulina se internaliza inmediatamente tras la unión de la insulina, lo que puede llevar a su degradación o reciclaje (107). La actividad de la tirosina quinasa disminuye por el AMPc o la fosforilación de residuos de serina/treonina en la subunidad β . Además, las diversas quinasas serina/treonina también fosforilan los IRSs como mecanismo de retroalimentación negativa del receptor a insulina (108-110).

Existen dos isoformas del receptor para insulina, IR-A e IR-B. En el músculo y páncreas se expresa principalmente el receptor IR-A y es colocalizado con el IRS-1, mientras que en el hígado, el tejido adiposo y en los riñones se expresa el IR-B junto con el IRS-3. La vía transduccional de cada receptor lleva a dos señales distintas, el IR-A señala fundamentalmente vías antiapoptóticas, mientras que el IR-B señala la diferenciación celular (111). El efecto de la insulina sobre las células β promueve la supervivencia para mantener la función e integridad de las mismas.

Los IRS participan en la señalización de la insulina para el crecimiento, supervivencia y metabolismo. Los IRS pueden activar a la fosfatidilinositol-3-quinasa (PI3K), cuyo resultado es activar diferentes proteínas, entre ellas la proteína quinasa B (PKB/Akt), la cual, dentro de sus funciones, activa factores de transcripción y la síntesis de glucógeno. Variantes genéticas como el polimorfismo Gly972Arg presente en el IRS-1, determina un defecto en un sitio de unión de PI3K lo que ocasiona una disminución en la secreción de insulina por la célula β pancreática en respuesta a glucosa (112).

3.5.5.1. Resistencia a insulina en la diabetes mellitus tipo 2

La resistencia a la insulina (RI) es un estado patológico en el que las células que ordinariamente responden a la insulina dejan de hacerlo. Los individuos con RI están predispuestos al desarrollo de DM2, además de asociárseles frecuentemente con un número importante de desórdenes de salud entre los que se encuentran la obesidad, la hipertensión, dislipemia y la inflamación crónica.

De manera general, la RI se manifiesta por una disminución en el transporte de glucosa inducido por la insulina en adipocitos y músculo esquelético, un aumento de la producción de glucosa hepática y alteraciones en el metabolismo de lípidos en tejido adiposo y hepático. A nivel molecular, los mecanismos por los que se genera la RI pueden ser múltiples y variar de un individuo a otro.

La RI es la consecuencia de una deficiente señalización de la insulina causada por mutaciones o modificaciones post-traduccionales en el receptor de insulina o moléculas efectoras *downstream* del mismo. En algunos casos la RI se debe a un defecto en la unión de la insulina a su receptor, pero más a menudo se atribuye a alteraciones posteriores a la unión de la insulina, que alteran desde la funcionalidad de su receptor hasta la actividad de proteínas localizadas *downstream* del mismo y que desempeñan funciones importantes en la señalización de la insulina (113, 114).

Entre las alteraciones más comunes se encuentran la disminución en el número de receptores y de su actividad quinasa, un aumento en el estado de fosforilación en residuos de Ser/Thr de proteínas clave como el receptor y su sustrato, la disminución de la actividad de las quinasas PI3K y Akt, y defectos en la expresión y función del transportador GLUT4 (presente en adipocitos, músculo esquelético y miocardio) (112).

De estas alteraciones el aumento en la fosforilación en residuos de Ser/Thr a nivel del receptor de insulina y de IRS, ha sido considerado como uno de los mecanismos clave en el desarrollo de la RI. Un aumento en el estado de fosforilación de ambas proteínas puede alterar su asociación a otras proteínas, bloquear sitios de fosforilación en Tyr, disminuir su activación e inducir su degradación (113, 115). La importancia de un aumento en el estado de fosforilación en residuos de Ser/Thr de las proteínas IRS también ha sido documentado en estudios clínicos, en donde se ha demostrado que en hígado, músculo y tejido adiposo de pacientes obesos (tejidos que desempeñan un papel importante en el desarrollo de la RI), la expresión de las proteínas IRS-1 disminuye alrededor del 54%, y este aumento en la degradación de IRS puede estar dado por un aumento en la fosforilación de residuos de Ser/Thr en IRS (116).

Varios agentes y condiciones metabólicas se han descrito como inductores de la RI. Los más comunes son los ácidos grasos libres (FFAs) y sus metabolitos; el factor de necrosis tumoral- α (TNF- α) y otras citoquinas, hormonas catabólicas (como la epinefrina, el glucagón y la angiotensina II), y hormonas secretadas por el tejido adiposo como la resistina. De esta forma parece que la resistencia a la insulina es consecuencia de la acción de una multitud de diferentes inductores. Por ejemplo, el incremento en la concentración plasmática de FFAs se encuentra asociado con muchos estados de RI, incluyendo obesidad y DM2. Inicialmente, el incremento en la concentración plasmática de FFAs, induce RI por la inhibición del transporte de glucosa estimulado por la insulina, que es seguido por una reducción en la síntesis de glucógeno en músculo y la oxidación de la glucosa.

Estudios a nivel molecular han determinado que el incremento en la concentración de FFAs puede llevar a cambios en la expresión del receptor de insulina y alterar, tanto la unión de la insulina con el receptor como el estado de fosforilación de su dominio de quinasa (117). Así mismo, pueden inhibir la activación de la enzima PI3K dependiente de IRS-1. La inhibición de la PI3K por FFAs ha sido asociada a un aumento en la fosforilación en residuos de Ser/Thr del IRS-1. Recientemente se ha descrito que los FFAs también pueden alterar la activación de Akt debido a un aumento en la cantidad de ceramida y diacilglicerol en células musculares en cultivo (115).

Por otro lado, la acumulación de triglicéridos en depósitos viscerales de manera ectópica (principalmente en el hígado) ha sugerido la existencia de un defecto en la

oxidación lipídica mitocondrial en pacientes con DM2 a nivel del musculo esquelético (118). *PGC-1* (*PPAR-γ coactivator 1*), factor de transcripción para genes implicados en la oxidación mitocondrial de ácidos grasos y síntesis de ATP, presenta una disminución en su expresión en pacientes jóvenes, delgados y con antecedentes familiares con DM2 (119). Esto sugiere un posible defecto hereditario en la fosforilación oxidativa mitocondrial que explicaría la acumulación lipídica celular. Estudios de expresión génica han mostrado una disminución en la expresión de *PGC-1* y genes relacionados que podrían afectar a la función mitocondrial en pacientes con RI y DM2 (120, 121).

3.5.5.2. Disfunción de la célula β en la diabetes mellitus tipo 2

Independientemente de los factores causales y los mecanismos patogénicos implicados en la disfunción de la célula β , una vez tiene lugar la aparición clínica de la enfermedad, la disminución gradual de la función/masa de las células β durante la historia natural de la misma, contribuye al deterioro continuo del control metabólico. Varios factores pueden contribuir a la progresiva pérdida de función/masa de la célula β en pacientes con DM2 (122):

- **Hiper glucemia-Glucotoxicidad**

Las células β son extremadamente sensibles a las concentraciones de glucosa en sangre; por ello, cambios en la homeostasis de glucosa influyen en la función y dinámica de la célula β . Además, la exposición crónica a altas concentraciones anómalas de glucosa en sangre tiene efectos perjudiciales en la síntesis/secreción de insulina, en la supervivencia celular y la sensibilidad a la insulina (“glucotoxicidad”). Todo ello a su vez produce hiper glucemia y, finalmente, el círculo vicioso de deterioro continuo de la función de células β . La glucotoxicidad implica cambios irreversibles en los componentes celulares relacionados con la producción/secreción de insulina (123):

a) La exposición prolongada a altas concentraciones de glucosa causa la pérdida gradual de expresión en el gen de la insulina y principales reguladores de la actividad promotora de la misma. Estos procesos están mediados, al menos en parte, por el estrés oxidativo (124-126).

b) Las células β expuestas a un aumento en la secreción de insulina requieren una gran demanda de síntesis de proinsulina por parte del retículo endoplasmático (RE);

conduciendo a un estrés celular (93). El estrés del RE genera una acumulación de proteínas desplegadas y supone la pérdida de la funcionalidad del mismo. En casos de estrés crónico, se produce la muerte de las células β a través de mecanismos de apoptosis. Además de la glucosa, FFAs y el polipéptido amiloide de los islotes (IAPP) son factores desencadenantes de estrés en el RE (127).

c) La hiperglucemia a largo plazo aumenta el flujo metabólico hacia la mitocondria e induce la generación excesiva de especies reactivas de oxígeno (ROS), conduciendo a un estrés oxidativo crónico (126). Estudios clínicos apuntan que los pacientes diabéticos presentan mayor cantidad de pro-oxidantes y marcadores de daño oxidativo tisular (128, 129). Además, el estrés oxidativo también podría estar implicado en la RI inducida por la hiperglucemia, repercutiendo en la función de la célula β (130).

d) La hiperglucemia crónica conduce a la disminución en el número de mitocondrias y a cambios en su morfología (aumento de volumen, reducción de proteínas de membrana interna, etc) en las células β . Estos cambios se asocian con defectos en la secreción de insulina estimulada por glucosa (GSIS) así como una disminución en la capacidad de transporte de Ca^{2+} o generación de ATP (131, 132). Además, la alteración en la homeostasis de Ca^{2+} influye negativamente sobre el crecimiento/función de las células β y en las vías de secreción de insulina (132).

e) Existe evidencia que la hiperglucemia crónica induce una respuesta inflamatoria no inmune en los islotes pancreáticos (producción de interleucina (IL) - 1β , factor NF-Kb, etc.). Las señales pro-inflamatorias desencadenarían la apoptosis de las células β mediante mecanismos asociados a oxidación y el estrés en el RE (93, 133, 134).

• Lipotoxicidad

La diabetes se asocia a menudo con cambios en los perfiles de lipoproteínas y el aumento en las concentraciones de FFAs (122). La exposición prolongada a niveles elevados de FFAs (resultantes de la lipólisis del adipocito o hidrólisis de lipoproteínas) pueden tener efectos negativos sobre la célula β al promover la acumulación de estos metabolitos tóxicos en los islotes pancreáticos (“lipotoxicidad”) (92, 125, 134). Sin embargo, hay estudios (procedentes de experimentos *in vitro* e *in vivo*) que indican que los efectos negativos de los lípidos sobre la célula β se producen solo en presencia de hiperglucemia. La oxidación de los ácidos grasos en las células β es inhibida en la

mitocondria produciéndose la acumulación de *long-chain acyl coenzyme A (CoA)* en el citoplasma (“glucolipototoxicidad”) (93, 135). Varios son los mecanismos por los cuales los ácidos grasos pueden afectar a la función y supervivencia de las células β (125).

La acumulación de compuestos lipídicos puede afectar al metabolismo de las células β , directa o indirectamente, a través de señales lipídicas (134). La exposición prolongada a altas concentraciones de ácidos grasos bajo condiciones de hiperglucemia disminuye la expresión génica y la secreción de insulina (125). Aunque la mayor parte de los estudios se ha centrado en los efectos de los FFAs sobre la función de las células β , nuevos datos sugieren que el metabolismo del colesterol/fracciones lipoproteicas puede también desempeñar un papel en la insuficiencia de la célula β . Se ha demostrado que lipoproteínas de baja densidad (LDL) pueden inducir apoptosis, y la LDL-oxidada reduce los niveles de expresión de la pre-proinsulina en células β aisladas mientras que las lipoproteínas de alta densidad (HDL) son protectoras (136, 137). Así, nuevos estudios apuntan a que las lipoproteínas parecen ser importantes moduladores de la función y supervivencia de la célula β (138).

• Autoinmunidad e Inflamación

El papel de la autoinmunidad en la DM1 es conocido desde hace mucho tiempo. Sin embargo, algunos fenotipos de DM2 también presentan marcadores de autoinmunidad que se correlacionan con la función de las células β , sugiriendo un posible papel del sistema inmunológico en la patogénesis de la DM2 (139-141). Las células β de los islotes están metabólicamente e inmunológicamente sobre reguladas cuando están expuestas a hiperglucemia (142). Se ha indicado que aproximadamente el 10% de individuos con DM2 presenta autoanticuerpos específicos de la diabetes, aunque el porcentaje es mayor en individuos de edad joven y delgados (143). Datos recientes identificaron individuos DM2 con respuestas de célula T reactiva contra antígenos del islote que se correlacionaron con una función inferior de las células β (141).

Un mecanismo patogénico que induce una respuesta inflamatoria en los islotes pancreáticos en pacientes con DM2 es el estrés metabólico crónico, debido a los altos niveles de glucosa en sangre y FFAs (aumento de citoquinas y quimiocinas). La citoquina IL-1 β juega un papel crucial al regular muchas otras citoquinas proinflamatorias (TNF- α o IL-6) y quimiocinas (CXCL-1, MCP-1 o MIP1- α) (144). La producción incrementada de citoquinas y quimiocinas contribuye a la apoptosis y deterioro de la célula β . Por otra

parte, factores derivados del tejido adiposo (adipoquinas proinflamatorias) pueden tener efectos locales y sistémicos en el metabolismo y contribuir a la inflamación crónica (145). Esta situación induce la disfunción y muerte de las células β , siendo un factor de riesgo para la autoinmunidad. Así pues, en la DM2 los posibles procesos de autoinmunidad en los islotes pancreáticos serían el resultado de respuestas inflamatorias crónicas asociadas con la obesidad, factores de estrés metabólico y de estrés de las propias células β (140).

• Adipoquinas

Las adipoquinas incluyen una gran variedad de péptidos liberados por el tejido adiposo. Algunas actúan como citoquinas proinflamatorias y otras tienen efectos protectores sobre la función y supervivencia de la célula β (82). La leptina puede inducir inflamación y apoptosis, ejerciendo efectos directos sobre el páncreas a través de sus receptores. Ello conduce a la inhibición de la GSIS y la disminución de los niveles de pre-insulina. Otras citoquinas liberadas por los adipocitos (IL-6 y TNF- α) también pueden modular la supervivencia de la célula β . Por otro lado, la adiponectina está relacionada estructuralmente con TNF- α y se ha visto que presenta efectos beneficiosos: acciones sensibilizadoras a insulina (principalmente en hígado), promueve la función y supervivencia de las células β y tiene efectos anti-inflamatorios locales y sistémicos (146). A nivel de las células β , la adiponectina actúa a través de sus receptores y parece aumentar la GSIS (147). Su efecto principal parece ser anti-apoptótico a través de la activación de quinasas pro-supervivencia y la reducción de los niveles de ceramida intracelulares (148). Otra adipoquina, visfatina, originalmente conocida como un mimético de la insulina, se ha demostrado que mejora la secreción de insulina, la proliferación de las células β e inhibe la apoptosis, posiblemente a través de la regulación de la expresión de *PDX-1* (*pancreas duodenum homeobox-1*) (149).

• Depósito de sustancia amiloide

La hipersecreción de depósito intracelular de polipéptido amiloide de los islotes pancreáticos (*islet amyloid polypeptide*, IAPP) ha sido una característica patológica de la DM2, siendo considerablemente estudiado su papel en la muerte de la célula β . El IAPP es un producto de las células β que es almacenado con la insulina en gránulos secretorios. Su función no es del todo conocida pero parece que limita la GSIS y tiene un efecto inhibitor en el vaciado gástrico y apetito (150, 151). Para que IAPP forme

agregados, primero tiene que adquirir conformación de hoja β , posteriormente formar precipitados de IAPP con forma de agregados amorfos y finalmente, fibrillas (152). Aunque la agregación empieza intracelularmente, la mayoría de fibrillas de amiloide se encuentran en el espacio extracelular donde forma los típicos depósitos. Una alta demanda en la secreción de insulina puede dar lugar a una proteína mal plegada en el RE, alteraciones de su procesamiento (en los gránulos secretores) y su posterior tráfico en las células β . Las moléculas de insulina (pero no la proinsulina) parecen formar complejos heteromoleculares que estabilizan el IAPP e inhiben la formación de la fibrilla de amiloide. En pacientes diabéticos se ha observado un aumento en la relación proinsulina/insulina que afectaría a la formación de las fibrillas. Se ha sugerido que un aumento en las concentraciones de glucosa y lípidos favorece la formación de las fibrillas de amiloide. Los mecanismos por los que la formación del amiloide están relacionados con la disfunción celular y muerte de la célula β todavía son desconocidos. Algunos estudios han indicado que pequeños oligómeros y fibrillas más grandes son más citotóxicos. La generación de iones ROS podría perturbar la cinética en la secreción normal de insulina y producir la muerte de las células β (150). Recientes estudios apuntan a que la agregación de IAPP activa la producción de IL-1 β y a través del estrés del RE, induce apoptosis y muerte de las células β (151, 153). Así pues, una fuerte correlación existe entre el depósito de amiloide y la reducción de la superficie junto con el aumento de la apoptosis en las células β (154).

• Incretinas

Los efectos positivos de las incretinas sobre la función y masa celular de las células β son conocidos (155, 156). Mediante la unión a receptores específicos de las células β , aumenta el AMP cíclico intracelular (cAMP), el calcio y estimulan la GSIS (156-158). También regulan varios genes específicos incluyendo la proinsulina, la glucoquinasa y transportadores de glucosa. Además, se ha demostrado que GLP-1 (*“glucagon-like peptide-1”*) impide la glucolipotoxicidad a través de la activación de PKB (159). Así, las incretinas protegen y mejoran la función de la célula β . Los mecanismos moleculares que están detrás de los efectos positivos de las incretinas no son del todo conocidos pero los datos indican que cAMP/PKA, vías de señalización PI3K/PKB o factores de transcripción como el PDX-1, estarían implicados (160). En la DM2, se produce tanto un deterioro en la liberación de incretina como una resistencia a su acción, contribuyendo de manera significativa a la deficiencia de la insulina y disfunción de las células β (161). Además, la

hiperglucemia altera el efecto de la incretina debido a una bajada en la regulación y una desensibilización de sus receptores específicos (156).

3.5.6. Mortalidad y coste sanitario de la diabetes mellitus tipo 2

Una de cada diez muertes de adultos en Europa puede atribuirse a la diabetes (619.000 en 2013) según la FID. La gran mayoría (90%) de estas muertes se produjeron en personas mayores de 50 años, lo que refleja en parte la distribución por edades de la población, pero también puede estar relacionado con la mejora de las tasas de supervivencia debido a los sistemas de salud más eficaces. La mortalidad por diabetes en las mujeres es ligeramente superior que en los hombres (329.000 frente a 289.000, respectivamente) (Figura 5).

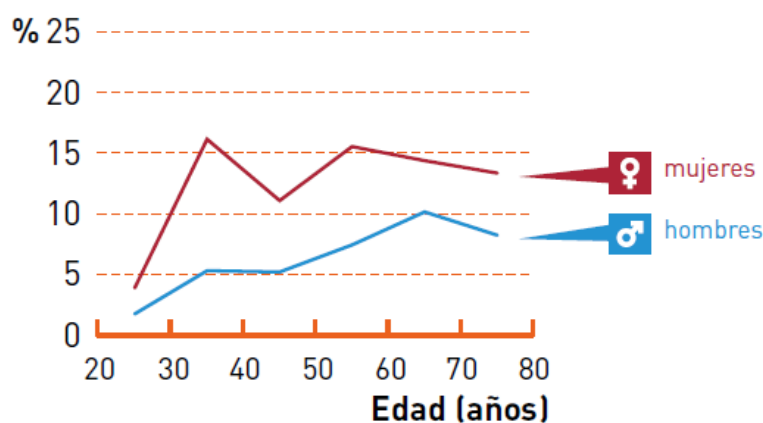


Figura 5. Muertes debidas a diabetes en Europa, 2013. Porcentaje de todas las causas de mortalidad debida a la diabetes por edad y sexo en Europa según la Federación Internacional de Diabetes (FID).

Además, existe un claro aumento de la mortalidad según los factores de riesgo, siendo mayor con la presencia de RI e hiperglucemia (Figura 6).

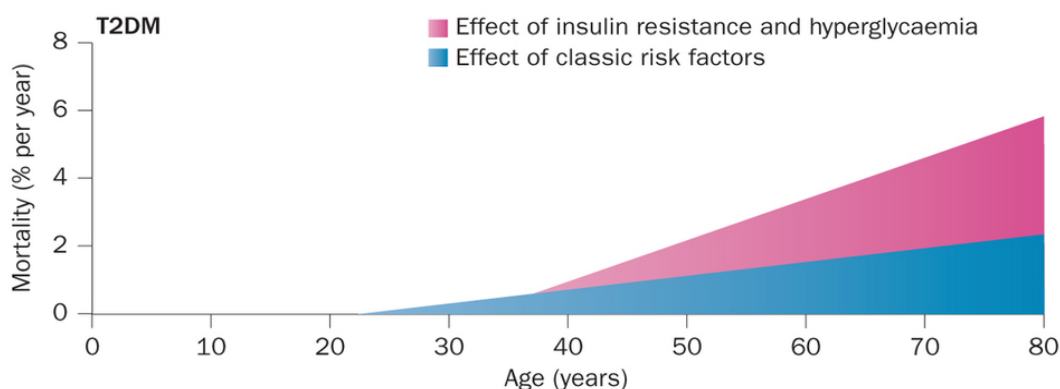


Figura 6. Efecto sobre la mortalidad de los factores de riesgo clásicos (azul) frente a los efectos de la resistencia a la insulina y la hiperglucemia (rosa). Adapt. *Nature Reviews Endocrinology* 10, 293-302 (2014).

Los resultados del estudio SECCA ("Spain estimated cost Ciberdem-Cabimer in Diabetes") estiman que el coste directo total anual de la DM ascendió a 5.809 millones de euros, representando el 8,2% del gasto sanitario total en 2012 (162); proporción muy elevada del total del gasto sanitario español. Esto hace necesario introducir estrategias y medidas para mejorar la eficiencia en el control y el tratamiento de la DM reduciendo así sus complicaciones y los enormes costes humanos y económicos asociados a la enfermedad (Figura 7). En el caso concreto de la DM2, un estudio observacional descriptivo en una muestra de población española estimó que el coste en salud anual se sitúa entre el 758 y 4.348 €/persona/año (68).

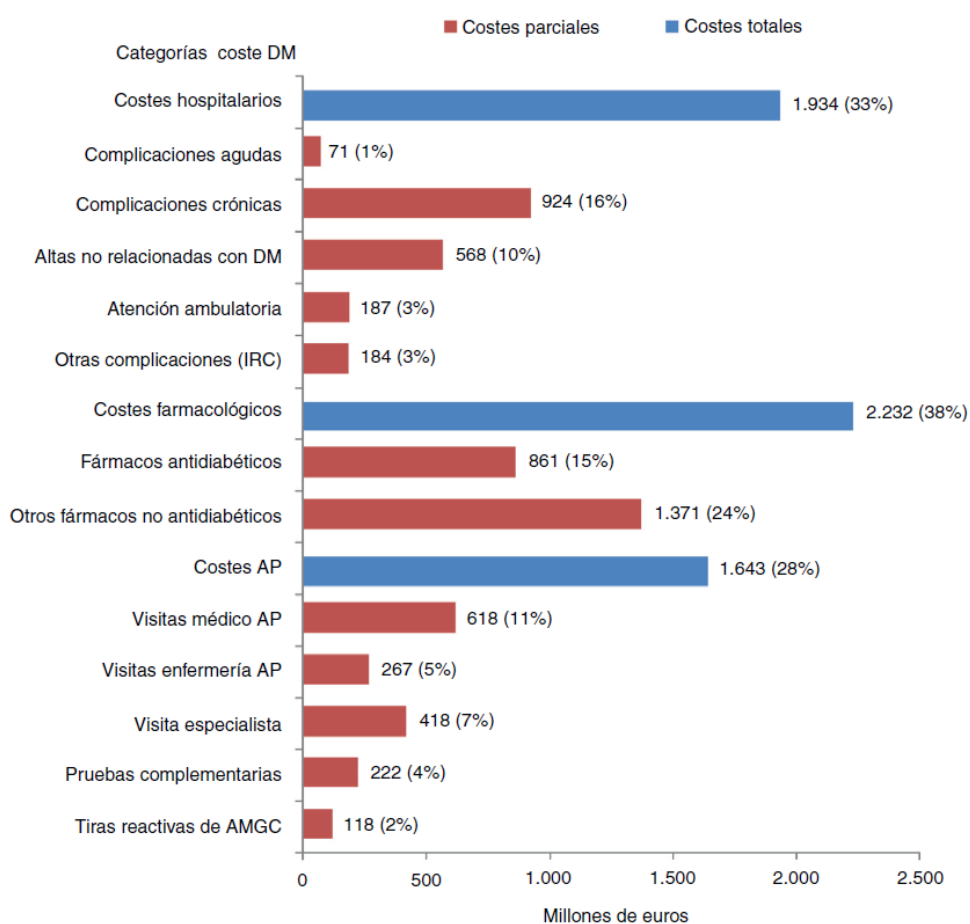


Figura 7. Costes directos relacionados con la diabetes mellitus y sus complicaciones (% sobre el total). AMGC: automonitorización de glucemia capilar; AP: atención primaria; DM: diabetes mellitus; IRC: insuficiencia renal crónica.

4. GENÉTICA DE LA DIABETES MELLITUS TIPO 2

La DM2 es una enfermedad compleja en la cual intervienen múltiples factores ambientales y genéticos. El conocimiento sobre los genes implicados en la patogénesis de la enfermedad ha incrementado sustancialmente en los últimos años. Aunque más de 100 *loci* asociados a la DM2 o parámetros glucémicos han sido descritos, sólo una pequeña parte del componente genético de la enfermedad es explicado, y las variantes causantes y genes afectados son desconocidos para la mayoría de los *loci* (163).

4.1. El factor ambiental en la diabetes mellitus tipo 2

Si bien los factores genéticos son importantes en la DM2, es obvio que los factores ambientales juegan un papel clave en el desarrollo de la enfermedad. El cambio hacia un estilo de vida occidental durante los últimos 50 años ha provocado una epidemia mundial incrementando la prevalencia de la DM2 y la obesidad (definida por el índice de masa corporal, IMC en kg/m^2). Los últimos estudios apuntan que en la Unión Europea entre el 30-70% de los adultos tiene sobrepeso (IMC $>25 \text{ kg}/\text{m}^2$) y entre el 10-30% obesidad (IMC $>30 \text{ kg}/\text{m}^2$) (164). Estos cambios han ocurrido en un periodo relativamente corto de tiempo, la mayoría debido a cambios en la dieta y la actividad física; por lo que nuestros genes no han cambiado. Sin embargo, se han descrito diferentes factores genéticos moduladores del riesgo de desarrollar DM2 que aumentan su efecto con el aumento del IMC. Esto no significa que el papel de la genética sea menor en el desarrollo de la enfermedad, pues nuestra respuesta a la nueva situación ambiental o a cada situación particular está determinada genéticamente. Una razón probable es que la selección haya favorecido genotipos conservadores de energía (*energy-preserving genotypes*), para poblaciones que viven en ambientes con un abastecimiento inestable y escaso de alimentos, maximizando el almacenaje energético y así, la probabilidad de supervivencia (165). Por lo que en el mundo occidental los individuos con estos genotipos tendrían un riesgo mayor de desarrollar enfermedades relacionadas con el metabolismo energético como obesidad o diabetes.

4.2. Heredabilidad de la diabetes mellitus tipo 2

Aunque factores ambientales, como el estilo de vida y la dieta, están implicados en la patogénesis de la enfermedad, los factores genéticos también están involucrados en

este proceso. Entre el 15-25% de los casos con antecedentes de primer grado de DM2, desarrollan intolerancia a la glucosa o diabetes (166).

Los estudios entre gemelos dicigóticos y monocigóticos han sido claves para estudiar el componente genético de la enfermedad. Diversos estudios han cuantificado la heredabilidad de la DM2 y de parámetros relacionados. La concordancia entre gemelos idénticos en la presencia de DM2 o intolerancia a la glucosa es superior al 60% en estudios transversales, pudiendo llegar al 76% y al 96%, respectivamente, si se realiza seguimiento hasta de 15 años (91, 167).

La genética de la DM2 ha ido avanzando de acuerdo a la tecnología. Ahora, con la secuenciación de nueva generación, se espera obtener información importante en variantes con fuerte efecto. No obstante, y hasta el momento, se han realizado diferentes estudios genéticos (estudios de ligamiento, genes candidatos, estudios de asociación de todo el genoma (GWAS), meta-análisis, etc.) con la finalidad de desgranar el componente genético de la enfermedad.

4.3. Análisis de ligamiento genético en familias

El objetivo del mapeo genético es la identificación de variantes en el genoma que incrementen la susceptibilidad a una enfermedad o que afecten a parámetros fisiológicos de ésta. Los métodos usados para mapear genes causantes de enfermedad han ido evolucionando en las últimas décadas. El método tradicional de mapeo genético es el denominado análisis de ligamiento. El objetivo es identificar la ubicación aproximada del gen con respecto a otra secuencia de ADN o marcador genético de posición conocida, basándose en el principio de desequilibrio de ligamiento en familias afectadas por la enfermedad.

Esta estrategia ha sido útil para mapear enfermedades genéticas con una fuerte penetrancia y con un modo de herencia conocido. Sin embargo, esta técnica es menos útil para identificar genes implicados en enfermedades complejas, como la DM2. Sólo algunos genes han sido identificados utilizando el análisis de ligamiento: *CAPN10* (*calpain 10*) y *TCF7L2* (*transcription factor 7-like 2*).

CAPN10, presente en el cromosoma 2, fue mapeado por análisis de ligamiento y descrito como gen de susceptibilidad para DM2 (168). Posteriormente, varios meta-análisis y GWAS han confirmado la asociación de *CAPN10* con DM2 (169, 170). Por otro lado un número de polimorfismos de un sólo nucleótido (SNPs) en *TCF7L2*, localizado en el cromosoma 10, han sido asociados con DM2 y confirmados en diferentes grupos étnicos (171-174).

4.4. Estudio de genes candidatos

La identificación de genes implicados en DM2 también se ha hecho utilizando pruebas de asociación en estudios casos-control y poblacionales, en lugar de con familias. Hasta hace un tiempo, la única estrategia viable ha sido estudiar los genes candidatos que tenían una alta probabilidad de afectar al rasgo de estudio en base a la función conocida del gen. Aunque se han realizado una gran cantidad de estudios de este tipo, sólo 6 genes han sido asociados con DM2: *PPARG* (*peroxisome proliferator-activated receptor gamma*) (175), *IRS1* (*insulin receptor substrate 1*) (176), *KCNJ11* (*potassium inwardly-rectifying channel, subfamily J, member 11*) (106), *WFS1* (*Wolfram syndrome 1*) (177), *HNF1A* (*HNF1 homeobox A*), y *HNF1B* (*HNF1 homeobox B*) (178).

El gen *PPARG*, codifica para el receptor nuclear PPAR- γ y está asociado con DM2. El receptor PPAR- γ es una molécula diana para compuestos de tiazolidinediona, una clase de drogas utilizadas en el tratamiento de la DM2. La presencia de un SNP en este gen produce el cambio de una prolina por una alanina en la posición 12 de la proteína y se calcula que está presente en el 15% de la población europea. Esta variación está asociada con un aumento de la actividad transcripcional, un incremento en la sensibilidad a la insulina, y una protección frente a la DM2 (175).

IRS1 codifica para una proteína la cual es fosforilada por el receptor tirosina quinasa de insulina y es esencial para la función de la insulina. La asociación entre un SNP (Gly972Arg) y la DM2 fue descrito por primera vez en 1993 y ha sido replicada en posteriores GWAS (176).

Además, se ha demostrado que variantes comunes presentes en genes causantes de enfermedades monogénicas incrementan el riesgo de DM2; es el caso de SNPs presentes en genes como *KCNJ11*, *WFS1*, *HNF1A* o *HNF1B*.

El gen *KCNFJ11* codifica una proteína que forma parte de un canal de potasio sensible a ATP (K^+ -ATP) el cual regula el potencial de membrana y por tanto, la secreción de insulina dependiente de glucosa en células β pancreáticas. Mutaciones en este gen causan graves diabetes neonatales (42). Por otro lado, el SNP Glu23Lys se ha asociado con deterioros en la secreción de insulina y con DM2 (106), siendo confirmada esta asociación por posteriores meta-análisis y GWAS (179, 180). *WFS1* codifica para Wolframin a, una proteína que es detectada en pacientes con el síndrome de Wolfram, siendo una de las características la diabetes juvenil. Posteriores estudios de genes candidatos, de asociación y meta-análisis han confirmado su asociación con DM2 (177, 181, 182). Mutaciones en *HNF1A* son causantes de la diabetes tipo MODY más frecuente en los países europeos, MODY3 (*mature onset diabetes of the young 3*). Estudios prospectivos y GWAS realizados en los últimos años han mostrado que variantes comunes en este gen disminuyen su actividad transcripcional e incrementan el riesgo de desarrollar DM2 (178, 183, 184). *HFN1B* es un factor de transcripción que puede heterodimerizar con HNF1A. Mutaciones en *HFN1B* son causantes de MODY5. Variantes en este gen se han asociado también a DM2 (185, 186).

4.5. Estudios de asociación de genoma completo (GWAS)

El rápido desarrollo de las plataformas de genotipado de SNPs ha permitido estudiar millones de SNPs simultáneamente, abriendo nuevas posibilidades para los estudios de asociación. El Proyecto HapMap ha proporcionado otra herramienta importante y ha demostrado que el genotipar alrededor de 500.000 SNPs es suficiente para cubrir entorno al 75% de las variantes comunes (frecuencia del alelo minoritario, MAF >5%) en el genoma (187).

Los primeros resultados del primer GWAS de DM2 fueron publicados en 2007 por Sladek et al (188). En este estudio se analizaron 661 casos y 614 controles, de población francesa, y se identificaron dos nuevos *loci*: *HHEX* (*hematopoietically expressed homeobox*) y *SLC30A8* (*solute carrier family 30, member 8*). A continuación, otros 4 estudios en población Europea fueron publicados incluyendo: *Welcome Trust Case Control Consortium* (189), *Diabetes Genetics Initiative Study* (190), *FUSION genetic study* (191) y otro por Steinthorsdottir et al. (192). Genes como *IGF2BP2* (*insulin-like growth factor 2 mRNA binding protein 2*), *CDKN2A/CDKN2B* (*cyclin-dependent kinase inhibitor 2A/cyclin-dependent kinase inhibitor 2B*) o *CDKAL1* (*CDK5 regulatory subunit*

protein 1-like 1) fueron identificados en estos GWAS. Además, numerosos *loci* descritos anteriormente fueron confirmados.

Posteriormente, también se han publicado otros estudios de GWAS en población no europea (193-196). SNPs en genes como *KCNQ1* (*KQT-like subfamily, member 1*), *PTPRD* (*protein tyrosine phosphatase, receptor type D*) o *YY1AP1* (*YY1 associated protein 1*), entre otros, se asociaron con DM2.

4.6. Estudios de meta-análisis a gran escala (*Large-scale meta-analysis of genome-wide association studies*)

Los estudios de meta-análisis combinan los resultados de diferentes GWAS, permitiendo aumentar así el número de individuos en el estudio. El primero de ellos fue *DIAGRAM consortium* donde se hizo un análisis combinado de 3 GWAS (*Diabetes Genetics Initiative, FUSION y Welcome Trust Case Control Consortium*). Posteriormente *DIAGRAM consortium* fue ampliado con otros GWAS y como resultado se descubrieron nuevos *loci* asociados a la enfermedad (196). A continuación, se muestran aquellas variantes comunes (MAF >0,05) asociadas a DM2 e identificadas mediante estudios de GWAS (Tabla 5).

En el *Meta-analysis of Glucose and Insulin-Related Traits (MAGIC) Consortium* se combinaron 21 GWAS procedentes de población europea y se descubrieron 9 *loci* nuevos asociados con glucosa o concentraciones de insulina (197). Entre estos genes 5 se asociaron también a DM2: *GCK* (*glucokinase hemokinase 4*), *DGKB-TMEM195* (*diacylglycerol kinase, beta 90KDa-transmembrane protein 195*), *GCKR* (*glucokinase hemokinase 4 regulator*), *ADCY5* (*adenylate cyclase 5*) y *PROX1* (*prospero homeobox 1*).

El último meta-análisis realizado ha incorporado otros grupos étnicos, además del europeo, y se han descubiertos otros nuevos *loci* de susceptibilidad a la DM2 (Figura 8) (198).

En resumen, y por todo lo expuesto anteriormente, la DM2 es una enfermedad compleja en la que interactúan tanto factores genéticos como ambientales determinando la resistencia a insulina y el fallo en la función de la célula β (Figura 9). Los genes implicados en el correcto funcionamiento del metabolismo glucídico y su regulación son muy numerosos. Así, la presencia de variantes genéticas en un gran número de genes implicados en ellos podrían conllevar alteraciones en los mismos y causar DM, pudiendo muchas de ellas incluirse en la DM2.

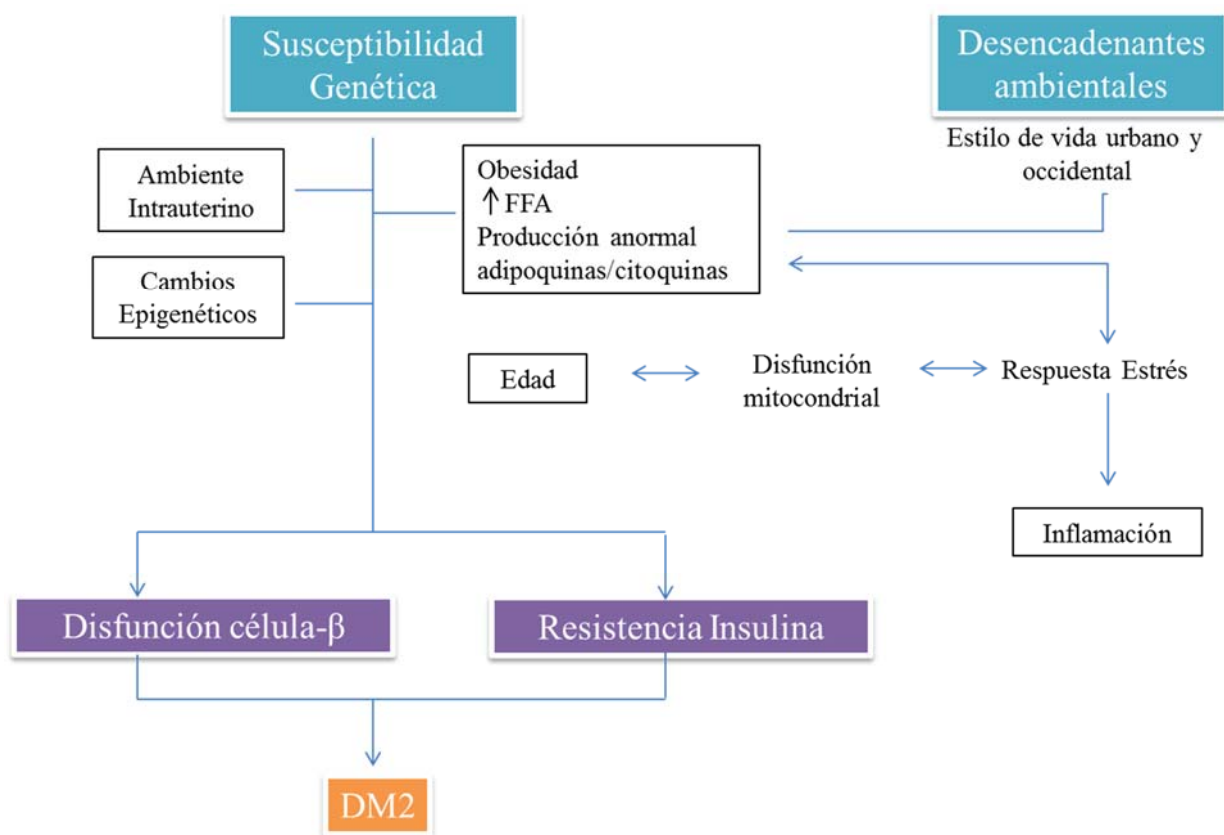


Figure 9. Esquema global de los factores patogénicos implicados en el desarrollo de diabetes mellitus tipo 2. Como enfermedad compleja, la DM2 es debida a una combinación de factores ambientales, genéticos y epigenéticos, que interactúan conjuntamente para producir la resistencia a insulina y la disfunción de la célula β . *Adapt. Brunetti et al (2014). World J Diabetes.*

5. NUEVAS ESTRATEGIAS PARA EL ESTUDIO GENÉTICO DE LA DIABETES MELLITUS TIPO 2

La información obtenida de los estudios genéticos realizados ha permitido que se abran nuevas posibilidades para entender la DM2, su prevención, tratamiento y el desarrollo de nuevas áreas de investigación (96). Los estudios GWAS realizados hasta el momento tienen varias limitaciones, pero la principal es que sólo permiten identificar la asociación de polimorfismos frecuentes o PFs (MAF >0,05). Estos polimorfismos, mayoritariamente, tienen un efecto pequeño sobre el fenotipo (odds ratio (OR) entre 1,1 y 1,5) (199, 200). Los polimorfismos poco frecuentes o PPFs (MAF <0,05) y con un efecto funcional fuerte podrían explicar parte del componente genético de la DM2 y presentar ORs mayores de 3 o 4 (201). La asociación de estos PPFs con DM2 y de muchos de los *loci* que los contienen no ha sido identificada hasta el momento.

Los PPFs son muy numerosos en nuestro genoma y pueden representar más del 50% de los polimorfismos. La mayoría de los polimorfismos funcionales son raros (95%) y normalmente proceden de eventos recientes en la historia humana; posiblemente por una fuerte expansión y falta de una selección de la población (202, 203). Un grupo de ellos podría tener un efecto funcional relevante en relación al desarrollo de la DM2 (204-207). Por tanto, es importante su estudio en cada población dado que en diferentes poblaciones existen diferentes variantes y el número de variantes en cada individuo es elevado (208-210).

Debido a los escasos resultados obtenidos hasta la actualidad se deben plantear nuevas estrategias para abordar el estudio del genoma y otras aproximaciones en el análisis de datos, que permitan sobrepasar las limitaciones de los estudios genéticos realizados hasta el momento en relación a la genética de las enfermedades complejas, entre las que se encuentra la DM2 (211, 212). En este sentido tres posibles estrategias relevantes emergen con el objetivo de hacer frente a estos estudios: a) la identificación de variantes genéticas en el exoma; b) estudio de polimorfismos PPFs y variantes raras; y c) selección de casos y controles mediante criterios específicos.

5.1. Identificación de variantes genéticas en el exoma

La tecnología de secuenciación de nueva generación ha tenido un gran desarrollo en los últimos años convirtiéndose en una herramienta de gran potencial para realizar estudios genéticos de secuenciación. Una aplicación que resulta de gran interés es la secuenciación completa del exoma.

El estudio del exoma resulta especialmente interesante dado que más del 90% de las mutaciones y polimorfismos asociados a enfermedades se sitúan en las regiones codificantes de proteínas, y por tanto, en el exoma (213, 214). Además, el análisis de polimorfismos distribuidos en todo el genoma muestra una mayor concentración de polimorfismos asociados con enfermedades complejas en el exoma frente al resto (209, 215). El estudio del exoma también permite analizar aquellas alteraciones que modifican el *splicing* del ARN y se encuentran en las regiones intrónicas contiguas a los exones (216, 217). A pesar de que existen otras regiones del genoma de gran interés cuyo análisis es muy importante, los resultados esperables son menos prometedores siendo el coste de estos estudios más elevado en la actualidad.

El número de polimorfismos en el exoma es muy elevado: en el estudio de 200 exomas, 120.000 polimorfismos de los que 53.000 eran SNPs codificantes no descritos previamente; sospechándose que hay muchos más por identificar, principalmente PPFs (209, 215, 218). Es importante remarcar el elevado número de variantes de pérdida de función presentes en cada persona, hasta 250-300 en todos los genes identificados y entre 50 y 100 en genes causantes de enfermedades hereditarias (208).

5.2. Estudio de polimorfismos de baja frecuencia y variantes raras

Los principales estudios realizados hasta el momento se basan en el análisis de PFs en la población, es decir el estudio de aquellos que tienen una MAF superior al 5 o al 10 % (178, 219-222). Estos estudios se basan en la hipótesis conocida como *Common Disease, Common Variant*, si bien, en la actualidad tiene cada vez más fuerza la hipótesis *Common Disease, Rare Variants* debido a la búsqueda de la “heredabilidad perdida” o no encontrada (199, 221).

En vista a los resultados obtenidos hasta el momento, cada vez parece más claro que cada enfermedad común puede ser causada por polimorfismos de muy variada

frecuencia en la población, integrando ambas teorías e incluyendo los diferentes tipos de polimorfismos: CNVs, SNPs, ins/dels, etc. y con diferente nivel de efecto funcional (211, 219, 221, 223, 224). Esta hipótesis también se ha propuesto en relación a la DM2 y tras analizar los datos de numerosos GWAS con esta enfermedad (178). Es decir, pueden existir numerosos PPFs (incluyendo las variantes raras; MAF <0,01) con un potencial impacto sobre el fenotipo (OR >3), por lo que en conjunto pueden estar detrás de un importante porcentaje del componente genético de la enfermedad (199, 221, 223, 225) (Figura 10). Si bien, existen algunos argumentos en contra (222).

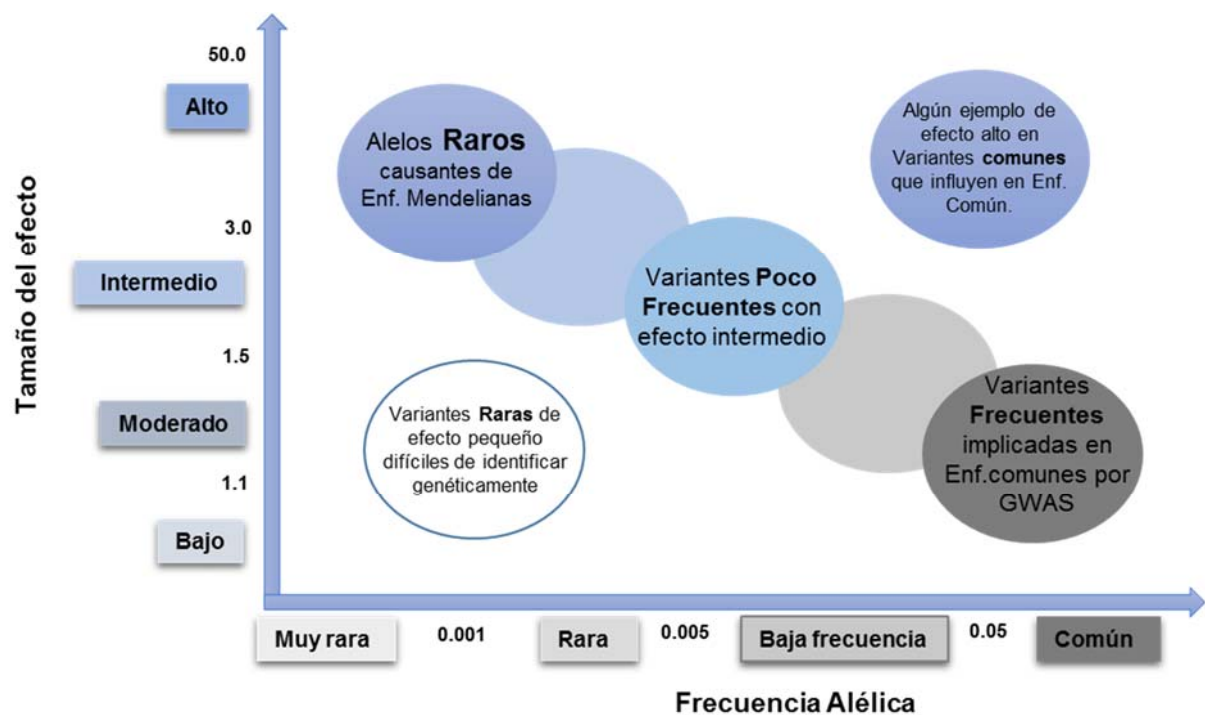


Figura 10. Efecto vs frecuencia alélica de las variantes genéticas. *Adapt. de Owens y cols., 2009 (Owen, Williams, & O'Donovan, 2009).*

Los PPFs actualmente son desconocidos en su inmensa mayoría (226). El proyecto *The 1000 Genomes Project Consortium* ha permitido la identificación de aproximadamente 15 millones de SNPs (208). Este trabajo supone un gran avance tanto en la identificación de nuevas variantes como en el desarrollo de herramientas bioinformáticas. Pero no garantiza la identificación de aquellos PPFs asociados a la DM2 y menos en población española (al incluirse sólo una pequeña muestra de población española en el proyecto). Esto se debe a que gran parte de los PPFs localizados en el exoma pueden proceder de eventos de mutación relativamente recientes y/o haber sufrido selección en base a su efecto (226), por lo que pueden existir diferencias importantes entre poblaciones.

Otro aspecto importante a tener en cuenta es que parte del componente genético calculado para la DM2 podría deberse a alteraciones epigenéticas (227). La epigenética se ha definido como cambios heredables en la función génica que ocurren sin producirse cambios en la secuencia de ADN. La epigenética es determinante en el desarrollo de enfermedades complejas: obesidad, resistencia a insulina y la propia DM2 (228-230). Estas alteraciones pueden transmitirse durante una o varias generaciones y pueden semejar una herencia genética convencional en algunos aspectos. Cambios epigenéticos tienen un efecto importante en la funcionalidad de genes implicados en el metabolismo hidrocarbonado, desarrollo de células β pancreáticas y en su funcionamiento (231, 232).

Dentro de las modificaciones epigenéticas se incluye: la metilación del ADN, la acetilación de las histonas o la interferencia postranscripcional por medio de ARN. En esta última el papel de los microARNs ha sido de gran estudio durante los últimos años. Los microARNs conforman una familia de ARNs no codificantes de pequeño tamaño que regulan negativamente la expresión de genes a nivel post-transcripcional. Además, los miARNs juegan un papel crucial como reguladores de la expresión génica mediante la unión a la región 3' UTR de sus mRNA diana; desestabilizando o inhibiendo su traducción. Los microARNs juegan un importante papel en el desarrollo del páncreas y en la función de la célula β (expresión y secreción de insulina) (233-235). Por ello, estas moléculas son dianas terapéuticas prometedoras y demuestran un gran potencial como biomarcadores de diagnóstico para la enfermedad cardiovascular, incluida la DM2 (236).

Así pues, el estudio de variantes genéticas en las secuencias codificantes para microARNs así como en sus sitios de unión a sus genes diana también puede ayudar a desengranar mejor la etiología y patofisiología de la enfermedad.

5.3. Selección de casos y controles mediante criterios específicos

Hasta la actualidad en los estudios de asociación de DM2 se han utilizado principalmente casos y controles (con similares características de edad, sexo y IMC) sin tener en cuenta otros factores, pudiendo ser muchas veces "supercasos vs supercontroles", que permiten solo identificar una parte del fondo genético en población general.

Para aumentar la probabilidad de identificar variantes genéticas en un grupo reducido de muestras que puedan estar asociadas a la DM2 y bien, aumenten el riesgo de desarrollar DM2 en una gran proporción de los casos o, lo disminuyan en los controles, debemos seleccionarlas correctamente. Como ya se ha indicado, el incremento en DM2 se produce por cambios que suponen un incremento de la obesidad. Por ello, es muy importante identificar qué factores genéticos se asocian con el desarrollo de la DM2 en pacientes con preobesidad u obesidad clase I (según la clasificación SEEDO-2007). En conjunto, este grupo representa al paciente "tipo" con un elevado riesgo de desarrollar DM2 y los casos más frecuentes en la población, principalmente de edad media (40-65 años), aunque la edad de diagnóstico se está reduciendo. Por el contrario, los pacientes con obesidad clase II o superior presentan una exposición a los efectos de la obesidad elevadísima por lo que el componente genético podría verse desvirtuado.

Así pues, dado que nuestro interés se centra en identificar aquellos genes implicados en el desarrollo o en la protección frente a la DM2 debida a un incremento ligero de IMC, hemos seleccionado pacientes con diagnóstico previo de DM2 o presencia de diabetes según la OMS (niveles de glucosa basal >126 mg/dl) y un IMC entre 28 y 35 kg/m² (237). Los individuos control presentan similar IMC pero niveles de glucosa en ayunas inferiores a 90 mg/dl (con el objetivo de evitar fenotipos intermedios que puedan provocar ruido y poder detectar genes y variantes que protejan frente a la DM2 o que ayuden a tener niveles bajos de glucosa).

Por otro lado, con el paso del tiempo el efecto de la exposición a un factor de riesgo aumenta y, por lo tanto, se precisa un menor componente genético para ocasionar las alteraciones metabólicas. Por lo tanto, se ha seleccionado una muestra de casos y controles de una edad moderada (40- 65 años) tanto para casos como para controles (en el caso de los controles cuanto mayor fuera su edad, se podría esperar una mayor carga genética protectora frente a un factor de riesgo, siendo esta una potencial limitación del estudio). Finalmente, y dado que el sexo puede ser un factor determinante, se han seleccionado tanto hombres como mujeres en ambos grupos.

Por otro lado, como se ha descrito anteriormente, los conocimientos acerca de la genética de la DM2 han ido avanzando con la tecnología, permitiendo superar las

limitaciones de los estudios anteriores y el desarrollo de los nuevos sistemas para el estudio genético de la enfermedad.

6. SECUENCIACION DE NUEVA GENERACIÓN: PLATAFORMA ILLUMINA

Desde la invención a finales de los setenta de los métodos de secuenciación de ADN, el método enzimático de Sanger ha sido el *gold standard* de la secuenciación. Este método automatizado permite la lectura de secuencias de ADN de una longitud de hasta 1000 pb, con una altísima precisión de lectura. Sin embargo, y a pesar de la innegable aplicabilidad de la técnica, sólo permite realizar 96 o 384 reacciones en paralelo; lo que conlleva mucho tiempo en la ejecución de los experimentos. Así, a partir del año 2003 empezaron a aparecer nuevas estrategias que combinaban la secuenciación de miles de pequeños fragmentos de ADN con un ensamblaje posterior para crear lecturas de varios Mpb. Estas nuevas técnicas de secuenciación fueron denominadas secuenciación de nueva generación o *next generation sequencing (NGS)*.

La principal ventaja de estas nuevas plataformas es la capacidad de producción de grandes volúmenes de datos de forma coste-eficiente (238). Existen diferentes plataformas en la actualidad: Roche (454 FLX Pyrosequencer), Illumina (HiSeq), Ion torrent (Proton / PGM sequencing) y Applied Biosystems (SOLiD), entre otros. A pesar de las diferencias que difieren en cuanto a la metodología química usada y a la generación del array de secuencias, el esquema de trabajo es similar: preparación de una librería de fragmentos de ADN, seguida de una ligación de adaptadores y amplificación, para pasar después a la secuenciación. El proceso de secuenciación alterna ciclos enzimáticos con un registro de fluorescencia que permite determinar la secuencia. Los avances en las tecnologías NGS son de especial utilidad para la resecuenciación de genomas, secuenciación de genomas de *novo*, secuenciación del exoma, determinación de regiones de interacción proteína-ADN, análisis de metilación, análisis de distintos tipos de ARN (RNA-Seq), etc. Esto permite detectar mutaciones, variantes genéticas, cambios en la metilación o en la expresión de genes, etc. De hecho, cabe destacar la importancia de estas tecnologías en proyectos como el proyecto 1000 genomas (239).

En el presente trabajo de investigación se ha utilizado, concretamente, el sistema HiScanSQ de la plataforma Illumina ya que era el sistema más adecuado para la finalidad del mismo. Así, la ADN polimerasa cataliza la incorporación de dNTPs marcados fluorescentemente a la cadena molde de ADN durante los ciclos de secuenciación. Durante cada ciclo, en el momento de incorporación, los nucleótidos son identificados por excitación del fluoróforo y posteriormente se elimina esta fluorescencia y el bloqueo de la síntesis de ADN, dejando la molécula de ADN lista para la incorporación del siguiente nucleótido. La diferencia crítica es que, en lugar de secuenciar un sólo fragmento de ADN, este proceso se extiende a través de millones de fragmentos de una manera masiva y paralelamente. La secuenciación química de Illumina es por síntesis permitiendo una gran precisión, alto rendimiento, tasas de error bajas y alto porcentaje de llamado de bases. El flujo de trabajo en Illumina consta de 4 pasos claves (Figura 11):

- **Preparación de la librería:** la librería de secuenciación es preparada por una fragmentación aleatoria del ADN seguida de una ligación de adaptadores específicos. Los fragmentos unidos al adaptador pueden ser amplificados por PCR y purificados posteriormente (Figura 11A).
- **Generación de *clusters*:** el conducto de microfluidos en Illumina se basa en una celda de flujo (*flow cell*) que consiste en un vidrio plano con 8 canales (*lanes*) donde se han unido covalentemente los adaptadores con secuencia complementaria a los adaptadores de la librería. Por estos canales pasan los reactivos para realizar los diferentes pasos necesarios. A continuación, los fragmentos de librería son amplificados *in situ* en la superficie de la celda de flujo mediante una etapa de amplificación “en puente” para finalmente producir los focos de secuenciación (*clusters*) (Figura 11B).
- **Secuenciación por síntesis:** la tecnología por síntesis de Illumina se basa en el método del terminador reversible el cual detecta bases individuales a medida que se incorporan a la hebra de ADN molde. El terminador reversible evita la extensión de la cadena y permite la detección mediante fluorescencia del nucleótido incorporado. Una vez se ha realizado este proceso, se elimina el marcaje fluorescente de la base y pasa a ser normal, con lo que en el siguiente ciclo se puede incorporar el nuevo terminador reversible correspondiente. Esta estrategia permite tener una baja tasa de error en las lecturas en comparación con otras tecnologías de NGS (240) (Figura 11C).

- **Análisis de datos:** la secuencia leída es alineada frente al genoma de referencia. A continuación tiene lugar la identificación de variantes (SNPs), inserciones/delecciones, etc. (Figura 11D).

Una posibilidad de la NGS es la secuenciación en *paired-end*. Esta permite secuenciar los dos extremos de los fragmentos de ADN de la librería de secuenciación y alinear las lecturas tanto en *forward* como en *reverse*. Esto genera el doble número de lecturas al mismo tiempo y aumenta la precisión en el alineamiento de las lecturas.

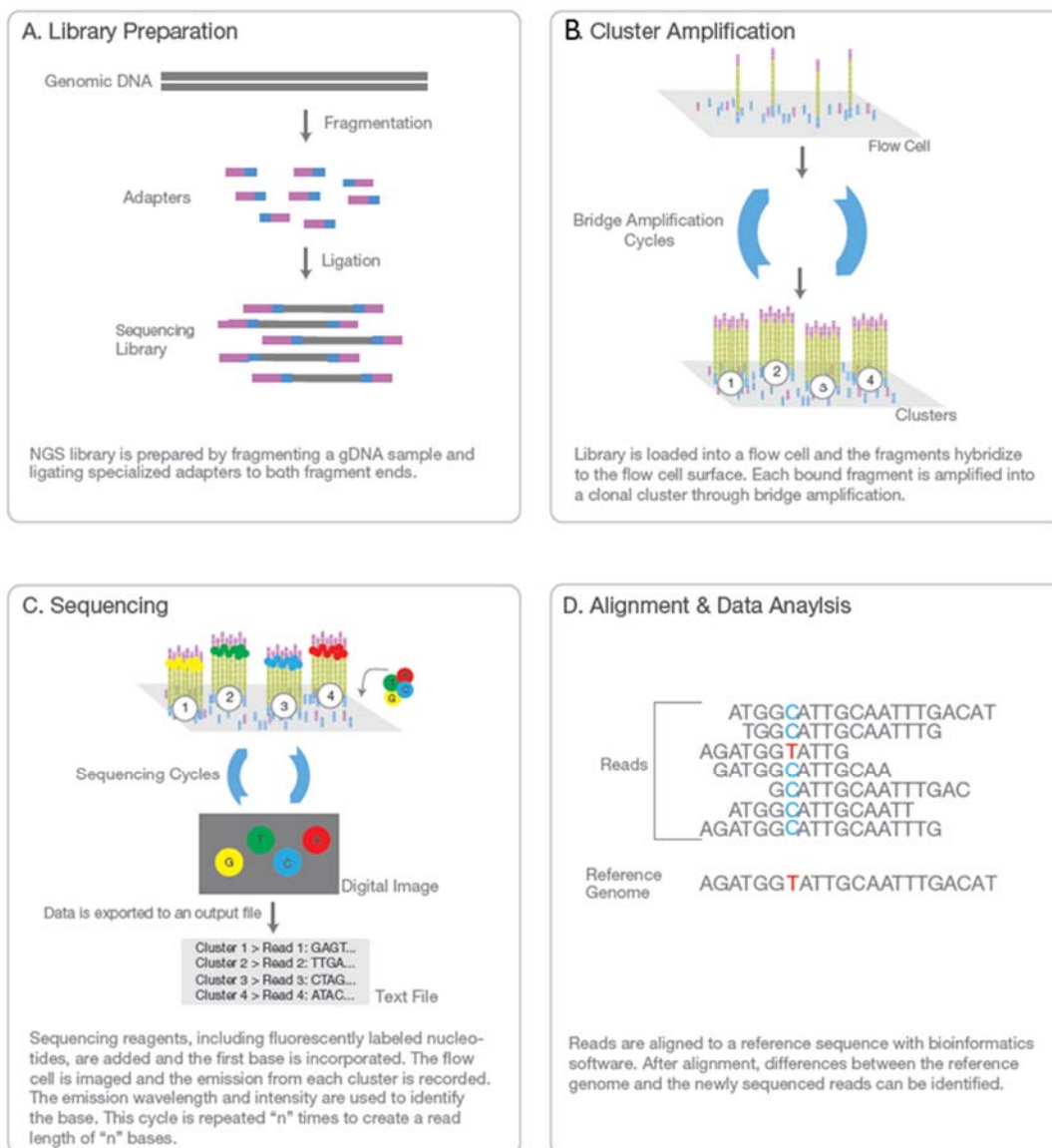


Figura 11. Flujo de trabajo para la secuenciación por Illumina. A) Construcción de la librería; B) Generación de *clusters* por amplificación clonal; C) Secuenciación por síntesis con terminadores reversibles; D) Análisis de los datos.

Hipótesis y Objetivos

La DM2 es una enfermedad compleja y heterogénea en muchos aspectos, principalmente en los mecanismos que conducen hasta ella y en los factores que la desencadenan. La DM2 surge de la interacción entre numerosos factores ambientales y genéticos, al igual que los niveles de diferentes parámetros relacionados con ella (niveles de glucosa, de insulina, resistencia a la insulina, etc.). En relación a los factores genéticos, tal y como se ha indicado en la introducción, existe un gran número de genes potencialmente implicados en el desarrollo de la DM2. No obstante, solamente se ha llegado a explicar entre el 5-10% del componente genético calculado mediante diferentes tipos de estudios, generalmente basados en el análisis de polimorfismos frecuentes en la población.

Más del 90% de las mutaciones causantes de enfermedades y polimorfismos asociados al desarrollo de enfermedades se localizan en el exoma. Las técnicas de secuenciación de nueva generación nos permiten identificar las variantes presentes en el exoma de una persona. Muchas de estas variantes genéticas son potencialmente funcionales al alterar la secuencia de las proteínas o el procesado de su ARN y, por lo tanto, pueden tener un efecto sobre el fenotipo.

Nuestra hipótesis es que numerosas variantes genéticas presentes en baja proporción en la población, pueden estar implicadas en el desarrollo de la DM2. Una parte de ellas están localizadas en el exoma y pueden explicar parte de la heredabilidad de la enfermedad que no ha podido ser encontrada por otros métodos. Además, para poder demostrar esta hipótesis general, nos basamos en que la secuenciación del exoma de casos y controles para la DM2 permitirá aumentar la proporción de variantes genéticas detectadas que estén implicadas en esta enfermedad.

Así, el presente trabajo surge dentro de un proyecto coordinado por varios grupos de investigación españoles con el objetivo principal de caracterizar la DM2 en España, siendo el objetivo de nuestro grupo la identificación de variantes genéticas presentes en el exoma en la DM2, incluyendo variantes poco frecuentes y raras, en relación a la DM2 en una muestra representativa de población española.

Para alcanzar este objetivo general nos planteamos en el presente trabajo los siguientes objetivos parciales específicos:

- Purificar y secuenciar el exoma de 400 muestras mediante secuenciación de nueva generación (200 casos y 200 controles).
- Identificar y analizar todos los polimorfismos presentes en los exomas secuenciados, centrándonos en variantes genéticas de baja frecuencia o raras.
- Seleccionar aquellos polimorfismos con una distribución diferente entre casos y controles, principalmente aquellos que puedan tener un mayor efecto funcional.
- Verificar los resultados obtenidos de la secuenciación de nueva generación mediante otros procedimientos.

Material y Métodos

1. SELECCIÓN DE LA POBLACIÓN DE ESTUDIO

Se seleccionaron 400 muestras para la secuenciación del exoma completo. Todos los individuos seleccionados tenían una edad comprendida entre 40 y 65 años y un IMC entre 25 y 34,9 kg/m²; procedentes del Biobanco del CIBERDEM y del Biobanco Nacional de ADN. Los criterios de inclusión fueron los siguientes:

- **Controles:** 200 sujetos con ausencia de DM2

Niveles de glucosa basal <90 mg/dl, insulina <12 µU/ml, HOMA <2,6, metabolismo hidrocarbonado no alterado tras sobrecarga oral de glucosa (glucosa <140 mg/dl e insulina <64 µU/ml).

- **Casos:** 200 pacientes diagnosticados de DM2

Diagnóstico previo de DM2 o presencia de diabetes según la organización mundial de la salud; niveles de glucosa basal >126 mg/dl, insulina >12 µU/ml y HOMA >3,2.

2. SECUENCIACIÓN DEL EXOMA COMPLETO

El ADN de todas las muestras del estudio fue extraído a partir de sangre periférica. Todos los participantes dieron su consentimiento para participar en el estudio habiendo sido aprobado previamente por el comité ético del Hospital Clínico de Valencia y de su Fundación.

Tras la extracción de ADN, las muestras fueron cuantificadas, purificadas, enriquecidas y secuenciadas mediante secuenciación de nueva generación empleando la tecnología TruSeq de Illumina (Figura 12).

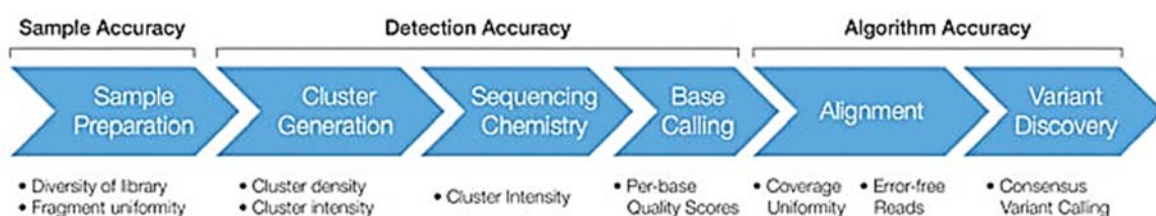


Figura 12. Esquema de trabajo para la secuenciación de exoma con la plataforma Illumina.

2.1. Preparación de las librerías de ADN

La generación de las librerías de ADN se llevó a cabo empleando el kit “TruSeq DNA Sample Preparation v2” (Illumina) siguiendo el esquema de trabajo representado en la Figura 13.

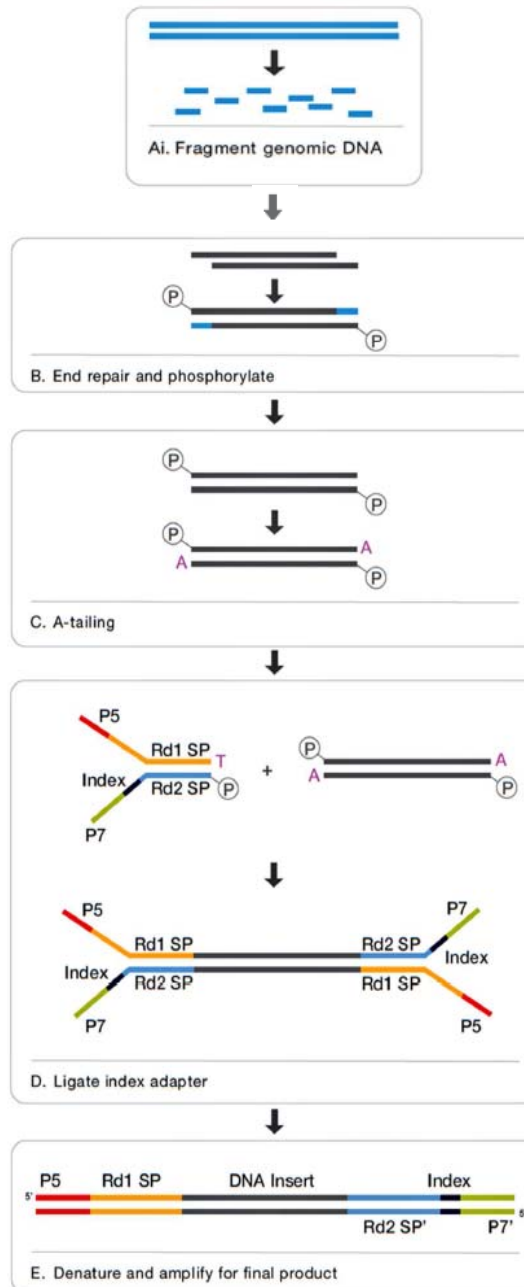


Figura 13. Esquema para la generación de la librería de ADN. El primer paso es fragmentar el ADN genómico (A). A continuación tiene lugar la generación de extremos romos (B) y la adición de una base de adenina (C), para seguir con la ligación de los adaptadores con el índice (D). De este modo, el producto final (E) queda listo para su amplificación en el cBot.

2.1.1. Cuantificación del ADN

Tras la extracción del ADN a partir de sangre periférica, el primer paso fue la cuantificación del ADN genómico utilizando el kit “PicoGreen® dsDNA quantitation assay” (Invitrogen) y el sistema “GloMax®-Multi Detection System” (Promega). Las ventajas de este método de cuantificación respecto de otros sistemas convencionales de cuantificación son: su alta sensibilidad, no requiere gran cantidad de muestra, es específico para ADN de doble cadena en presencia de cantidades equimolares de ARN, fácil de usar y apto para formatos de placas de 96 y 384 pocillos. La plataforma de secuenciación de Illumina recomienda partir de una concentración de 20 ng/μL en un volumen final de 50 μL (1 μg de ADN total).

2.1.2. Fragmentación del ADN

El siguiente paso en la preparación de las librerías es la fragmentación del ADN genómico. Este procedimiento se llevó a cabo con el sistema Covaris™ S220 (Life Technologies) el cual permite la obtención de fragmentos de ADN de una forma sencilla y precisa. El sistema implica un paso de sonicación (ultrasonidos y longitudes de onda muy cortas) donde se generan fragmentos un rango de tamaños entre 100-250 pb, si bien se necesita cierta optimización de parámetros. La principal ventaja de este método es que permite un control de los diferentes parámetros y una reproducibilidad en la fragmentación de las muestras para condiciones similares. A continuación, el ADN fragmentado se pueden visualizar en el sistema de QIAxcel (Qiagen). El sistema QIAxcel permite realizar una electroforesis capilar de forma automatizada de hasta 96 muestras en una carga, mediante electroforesis de 12 muestras simultáneamente. Es altamente sensible y resolutivo y reduce los errores de manipulación manual y tiempo (Figura 14).

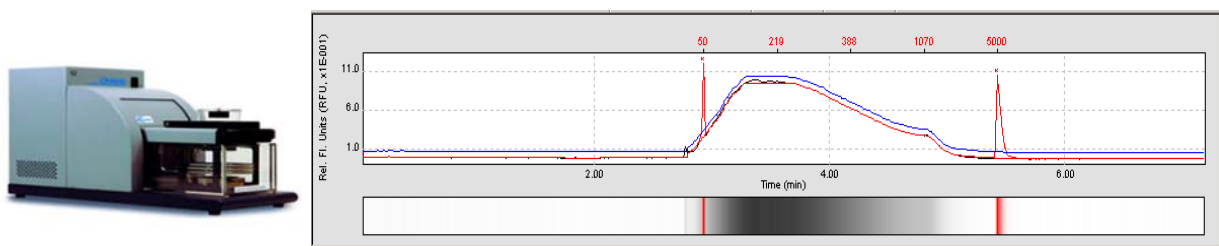


Figura 14. Fragmentación del ADN genómico utilizando el sistema Covaris y visualización en el sistema QIAxcel. Aparato Covaris de fragmentación de ADN (izquierda) y visualización de una muestra de ADN fragmentado en el Qiaxcel (derecha).

2.1.3. Reparación de los extremos

Los extremos salientes del fragmento de ADN generado son transformados mediante una exonucleasa 3'-5' en extremos romos. De este modo con la actividad exonucleasa se eliminan los extremos sobresalientes en 3' y con la polimerasa los extremos en 5' son completados.

2.1.4. Adenilación

La adición de una base de adenina (A) a los extremos 3' de las cadenas de ADN impide que los fragmentos de extremos romos se ligen entre sí durante la reacción de ligación de adaptadores. Además, este proceso reduce la formación de quimeras y la ligación entre oligonucleótidos, evitando uno de los principales problemas que es la formación de *primer-dimer* (dímeros de oligonucleótidos).

2.1.5. Ligación de los adaptadores

A continuación, se añaden a los extremos del ADN unas secuencias adaptadoras con la finalidad de identificar y amplificar posteriormente la muestra en una celda de flujo o *flow cell*. Este proceso es clave para identificar cada una de las muestras por su índice tras la secuenciación. Además durante este proceso tiene lugar la incorporación de todas las secuencias necesarias para realizar las reacciones de secuenciación del fragmento interno e índices (Figura 15).

- Rd1 SP: oligonucleótido para la secuenciación del read 1
5' AACTCTTTCCCTACACGACGCTCTTCCGATCT
- Rd2 SP: oligonucleótido para la secuenciación del read 2
5' CGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATCT
- P5 Y P7: oligonucleótidos que se unen a la celda de flujo.
5' P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
5' AACTCTTTCCCTACACGACGCTCTTCCGATCT
- Índice: secuencia específica que identifica cada muestra.

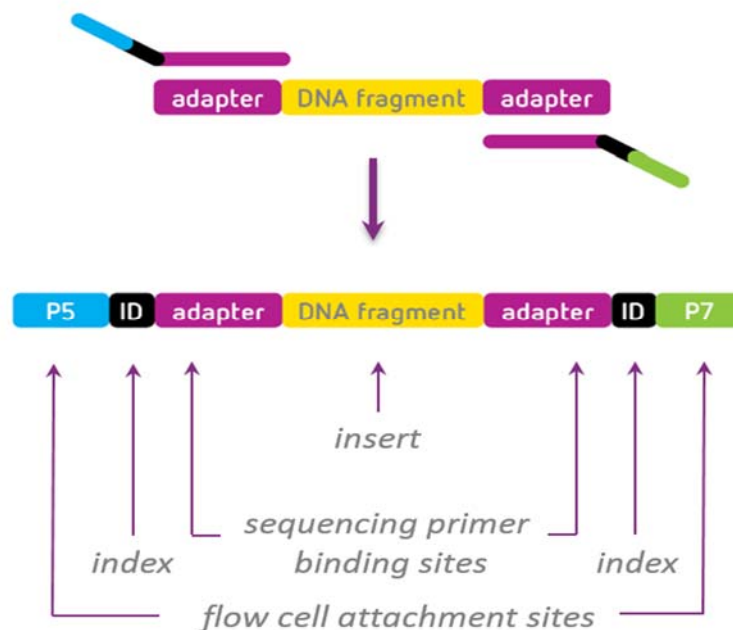


Figura 15. Esquema de los adaptadores de secuenciación.

Los índices y secuencias de los mismos que se emplearon para la identificación de las muestras tras la secuenciación se muestran en la Tabla 6.

Tabla 6. Secuencia de nucleótidos para los diferentes índices empleados.

SET A		SET B	
Índice	Secuencia	Índice	Secuencia
AD002	CGATGT(A)	AD001	ATCACG(A)
AD004	TGACCA(A)	AD003	TTAGGC(A)
AD005	ACAGTG(A)	AD008	ACTTGA(A)
AD006	GCCAAT(A)	AD009	GATCAG(A)
AD007	CAGATC(A)	AD010	TAGCTT(A)
AD012	CTTGTA(A)	AD011	GGCTAC(A)
AD013	AGTCAA(C)	AD020	GTGGCC(T)
AD014	AGTTCC(G)	AD021	GTTTCG(G)
AD015	ATGTCA(G)	AD022	CGTACG(T)
AD016	CCGTCC(C)	AD023	GAGTGG(A)
AD018	GTCCGC(A)	AD025	ACTGAT(A)
AD019	GTGAAA(C)	AD027	ATTCCT(T)

El conjunto de todos los fragmentos de ADN de una misma muestra unidos a los adaptadores tiene como resultado la generación de la librería.

2.1.6. PCR de enriquecimiento

Con la finalidad de favorecer el enriquecimiento de aquellos fragmentos de ADN con los adaptadores y por consiguiente, la cantidad de ADN en la librería, se lleva a cabo una amplificación selectiva utilizando un par de oligonucleótidos complementarios a los extremos de los adaptadores.

Después de cada etapa de amplificación tiene lugar un paso de purificación mediante esferas magnéticas Agencourt AMPure XP (Beckman Coulter, Inc). Ello permite la selección de fragmentos de ADN mayores a 150 pb.

2.2. Enriquecimiento del exoma

Seguidamente se procedió a capturar la zona exómica y circundante (“Truseq Exome Enrichment Kit”, Illumina), un 90% aproximadamente de la zona codificante conocida del genoma.

Este paso empieza con la mezcla equimolar de las librerías en pools de 6. Siguiendo las recomendaciones de Illumina, se emplea 500 ng de cada librería de ADN y se cuantifica empleando el sistema GloMaxR-Multi Detection (Promega).

El proceso de enriquecimiento cuenta con dos procesos de hibridación intercalados con sus correspondientes lavados. En el paso de hibridación, la librería de ADN desnaturalizada se mezcla con sondas específicas marcadas con biotina y que capturan la región exónica. A continuación, en los pasos de lavados se utilizan esferas de streptavidina para capturar las sondas unidas a las regiones de interés y eliminar el material genético que no haya hibridado de forma específica con las sondas (Figura 16).

El kit “TruSeq Exome Enrichment” incluye más de 340.0000 sondas, cada una diseñada contra el genoma de referencia humano hg19/GRCh37. El conjunto de las sondas ha sido diseñado para enriquecer más de 200.000 exones y cubre alrededor de 20.794 genes de interés y unos 62.000.000 pb. Las características generales de cobertura se muestran en la Tabla 7.

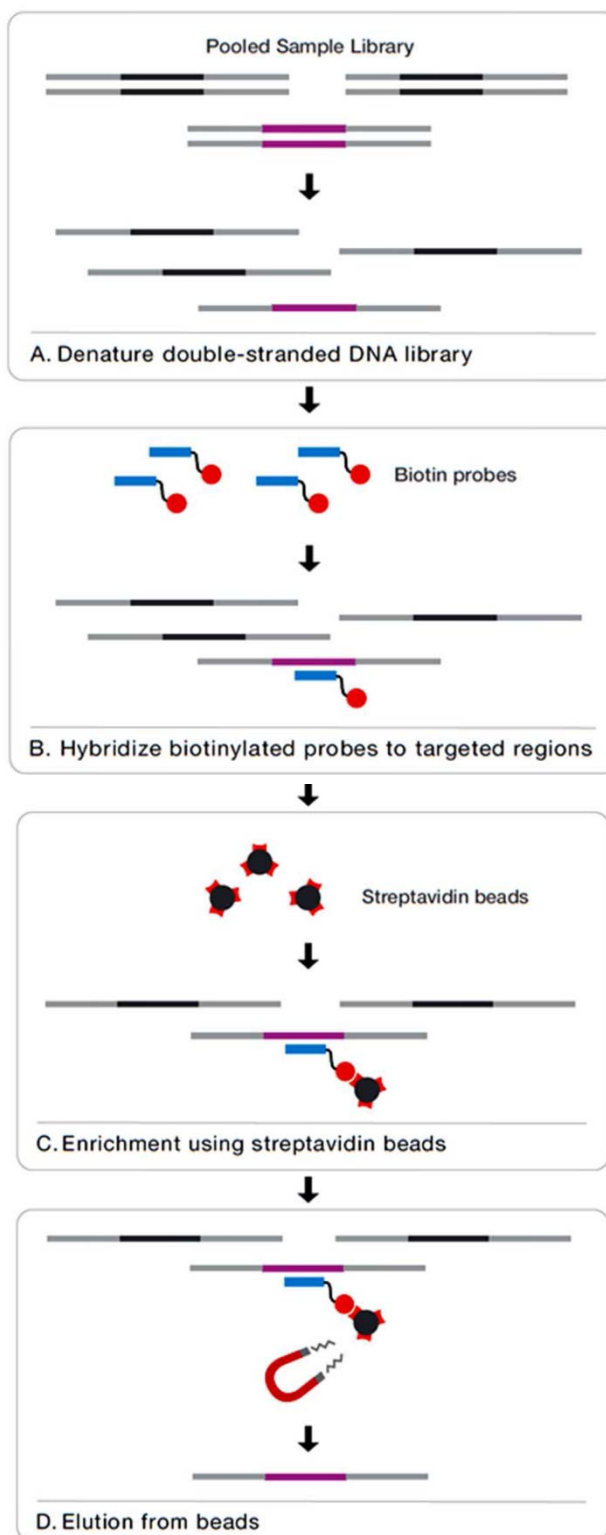


Figura 16. Proceso de enriquecimiento exómico.

2.3. Validación de los pools

2.3.1. Visualización y selección de los pools

A continuación tiene lugar la selección de los pools de acuerdo con su tamaño. Para ello se visualizan los pools (6 librerías de ADN) en un gel utilizando el sistema de QIAxcel (Figura 18). Luego se calcula el tamaño medio del pool para hacer su posterior normalización tras la cuantificación del mismo (qPCR).

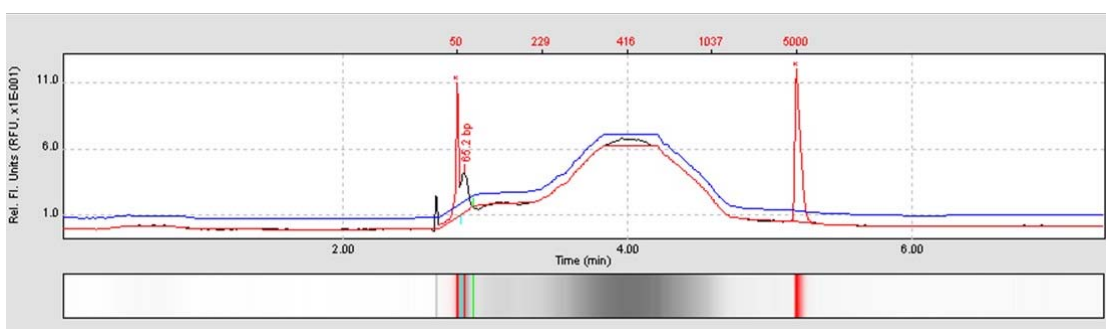


Figura 18. Visualización de un pool (6 librerías de ADN) en el sistema QIAxcel.

Si existe formación de dímeros, estos son eliminados mediante el sistema Pippin Prep (Sage Science). Este sistema consiste en una electroforesis capilar que permite obtener fragmentos de ADN del tamaño especificado de forma sencilla, rápida y sin posibilidad de contaminación cruzada (Figura 19).

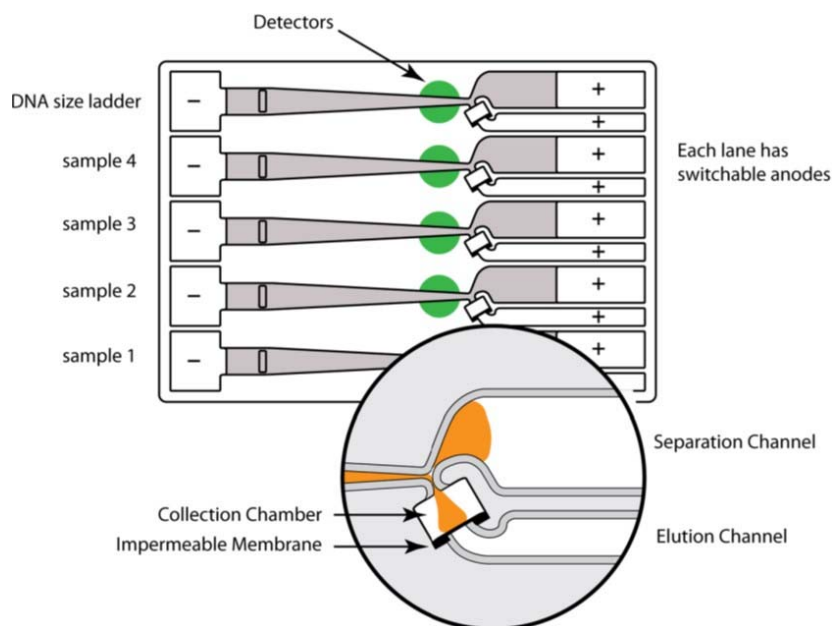


Figura 19. Sistema Pippin Prep

2.3.2. Cuantificación de los pools

La cuantificación de los pools se realizó mediante PCR cuantitativa usando el sistema LightCycler 480 (Roche). Concretamente se empleó un método no específico donde se utilizó el SYBR Green como fluorocromo. Este tiene la propiedad de generar una señal fluorescente al unirse al ADN de doble cadena que es proporcional a la cantidad de producto.

Para la cuantificación de las librerías se utilizó una curva estándar (Kappa BioSystems) con muestras de concentración conocida. A continuación, se determinó el C_p (*crossing point*), ciclo en el cual la fluorescencia de la amplificación supera la fluorescencia de fondo para las muestras estándar. Finalmente, se calculó la concentración de las muestras extrapolando el valor de C_p de la muestras en la curva estándar (Figura 20).

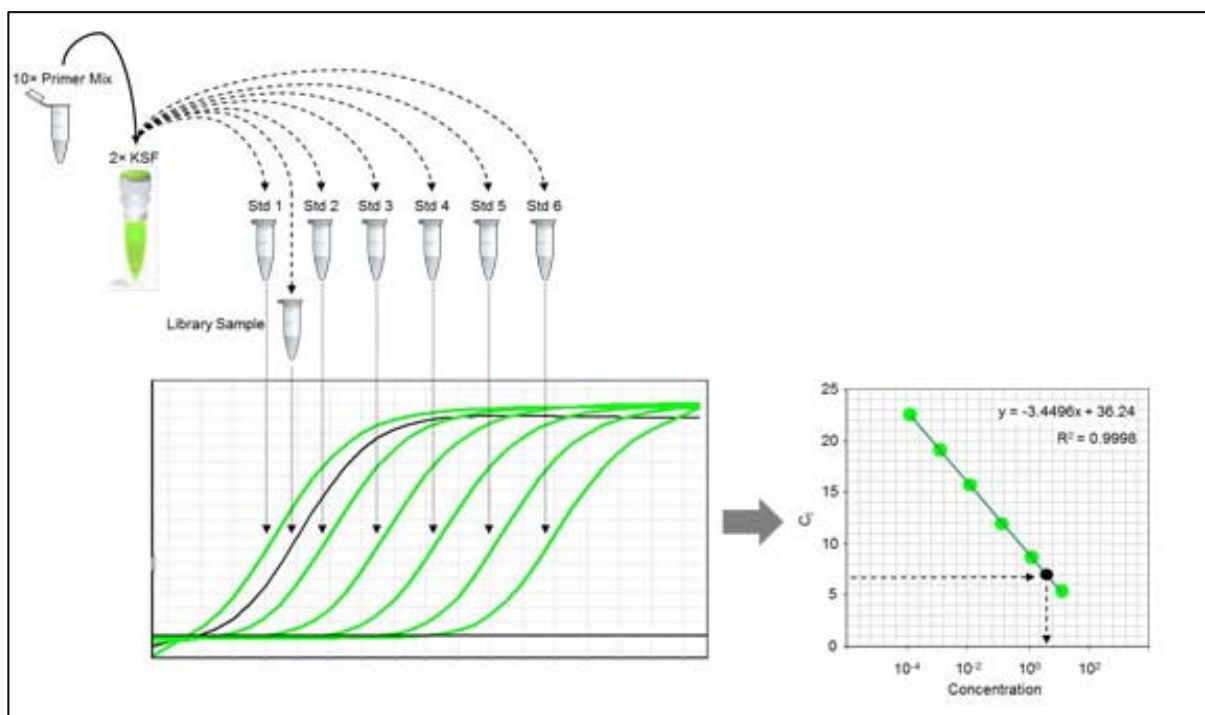


Figura 20. Esquema de cuantificación de las librerías de exoma mediante qPCR. La muestra a cuantificar de concentración desconocida es la señalada en color negro. Su concentración es obtenida al extrapolar su valor de C_p de la curva realizada a partir de las concentraciones de las muestras estándar.

2.4. Amplificación clonal de los pools: cBOT

Se realizó la amplificación clonal de los fragmentos de ADN empleando el sistema cBOT de Illumina con el kit “TruSeq PE Cluster v3-cBot-HS”. Antes de secuenciar las muestras, las moléculas de ADN son amplificadas “en puente” para formar los *clusters* clonales sobre una superficie sólida denominada celda de flujo o *flow cell*. Las muestras se preparan a una concentración de 2 nM por pool (definiendo un pool como la mezcla de 6 muestras diferentes con índices compatibles), posteriormente se desnaturalizan y se preparan a la concentración deseada para introducir las en el cBOT. Tras diferentes experimentos de secuenciación, concluimos que la concentración óptima para nuestras librerías y que permitía obtener, con nuestro equipamiento, *clusters* lo suficientemente separados unos de otros para garantizar unas secuenciaciones de calidad fue de 8 pM.

Dicha amplificación consta de tres fases:

- **Inmovilización de las moléculas de ADN:** cientos de millones de moléculas de ADN se hibridan con oligonucleótidos (directos e inversos) que se encuentran inmovilizados sobre la *flow cell*. A continuación, se produce una amplificación (mediante PCR) de las secuencias unidas donde se van generando copias por extensión en 3', empleando la ADN polimerasa de alta fidelidad para prevenir errores de incorporación. Las muestras originales son desnaturalizadas, dejando a su copia inmovilizada sobre la superficie sólida de la *flow cell* (Figura 21, A).
- **Amplificación isotérmica en puente:** las copias inmovilizadas giran para hibridarse con los oligonucleótidos que están libres sobre la *flow cell*. La ADN polimerasa copia la muestra formando puentes de ADN que se desnaturalizan para obtener 2 cadenas de ADN, que a su vez, giran y se hibridan con otros oligonucleótidos, extendiéndose para formar otras dos nuevas cadenas de ADN. La operación se repite en el área circundante hasta conseguir muchas localizaciones aisladas que contienen múltiples copias idénticas de un fragmento único, conocido como *cluster* (Figura 21, B).
- **Linealización, bloqueo e hibridación:** cada *cluster* de los puentes de ADN es desnaturalizado y la cadena en dirección 3' se elimina, dejando la cadena 5'. El extremo 3' de las cadenas de ADN y los oligonucleótidos presentes en la *flow cell* son bloqueados para que no interfieran en la reacción de secuenciación posterior. El oligonucleótido de secuenciación se hibrida con la secuencia complementaria del adaptador de Illumina en

los extremos sueltos de las secuencias de los *clusters*. La *flow cell* contendría en este paso alrededor de 750 millones de *clusters* con aproximadamente 1000 moléculas por *cluster*, estando lista para ser secuenciada (Figura 21, C).

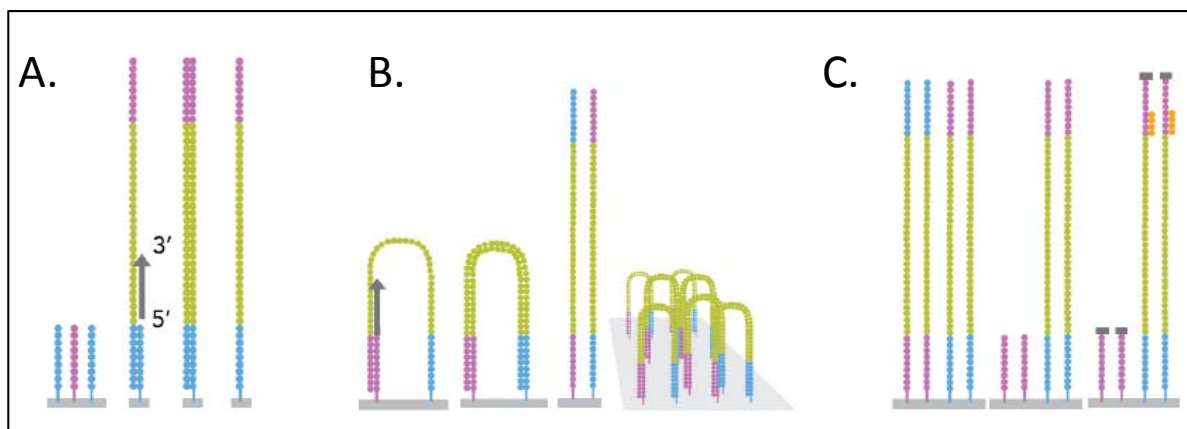


Figura 21. Generación de los *clusters* por amplificación en puente. A) Inmovilización de las moléculas de ADN de la librería generada, B) Amplificación isotérmica en puente y, C) Linealización, bloque e hibridación.

2.5. Secuenciación por síntesis: HiScanSQ (Illumina)

Este trabajo se realizó utilizando el sistema HiScanSQ (Illumina) y los kits de secuenciación “TruSeq SBS v3-HS”. En este paso, tiene lugar la secuenciación por síntesis de los *clusters* empleando para ello nucleótidos marcados fluorescentemente con 4 fluoróforos distintos y con terminadores reversibles. En cada ciclo de síntesis se incorporan los nucleótidos modificados cuando estos son complementarios a la cadena molde (solo 1 nucleótido en cada ciclo por molécula). Posteriormente, se eliminan los nucleótidos sobrantes y se recoge la imagen de fluorescencia mediante la excitación de los fluoróforos por los laser del equipo y la fluorescencia emitida es captada por una cámara CCD que se va desplazando a lo largo de la *flow cell*. Este proceso se va repitiendo de forma cíclica y finalmente las imágenes se analizan para identificar la secuencia completa (Figura 22).

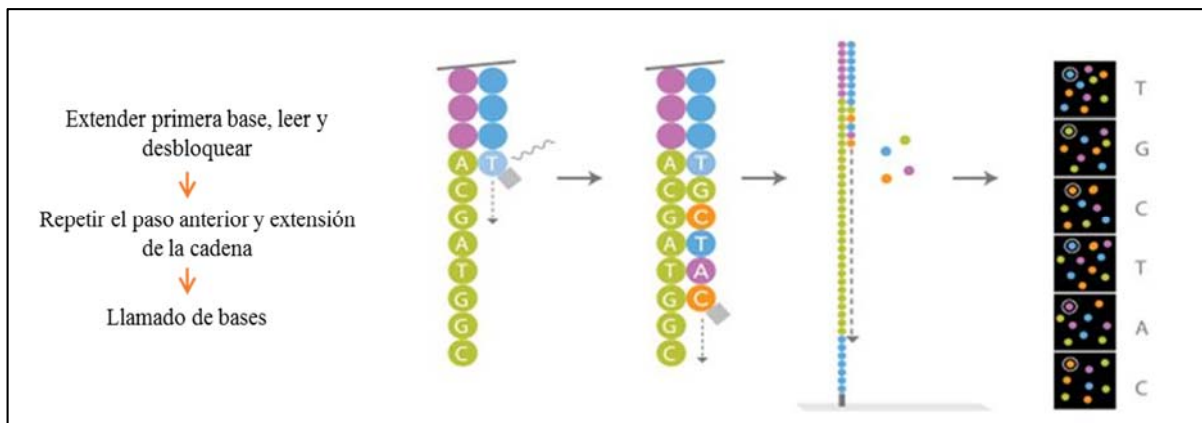


Figura 22. Esquema de flujo en la secuenciación por síntesis en la plataforma de Illumina.

La secuenciación realizada durante este proceso es de tipo *paired end*. Este tipo de secuenciación hace referencia a la generación de dos lecturas a partir de los oligonucleótidos de secuenciación correspondientes a los adaptadores presentes en cada extremo de cada fragmento de ADN de la librería. Teniendo en cuenta que la distancia entre ellas es conocida de manera aproximada, se generan secuencias con un alineamiento de alta precisión (Figura 23).

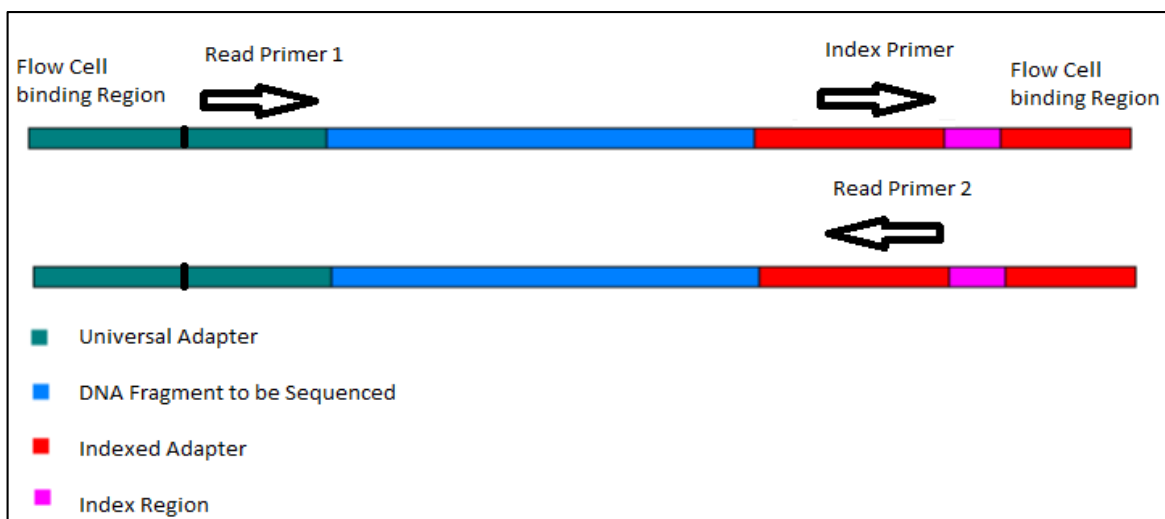


Figura 23. Secuenciación tipo *paired-end*.

3. ANÁLISIS BIOINFORMÁTICO

Tras el proceso de secuenciación, el siguiente paso fue la identificación de todas las variantes funcionales presentes en los exomas garantizando el mejor análisis posible mediante los controles de calidad adecuados y una priorización apropiada de dichas variantes.

La plataforma de secuenciación de Illumina incorpora un software que permite seguir el proceso de secuenciación a tiempo real y posibilita el análisis preliminar de los datos. Hay dos procesos principales (Figura 24):

- HiSeq Control Software: proporciona información en tiempo real del ciclo de secuenciación, notificaciones, *clusters* y lugar de lectura dentro de la *flow cell*.
- Sequencing Analysis Viewer (SAV): proporciona información de los *clusters* efectivos y la densidad de *clusters* en cada zona de la *flow cell*. También ofrece información de la lectura, de los índices y de cada muestra.

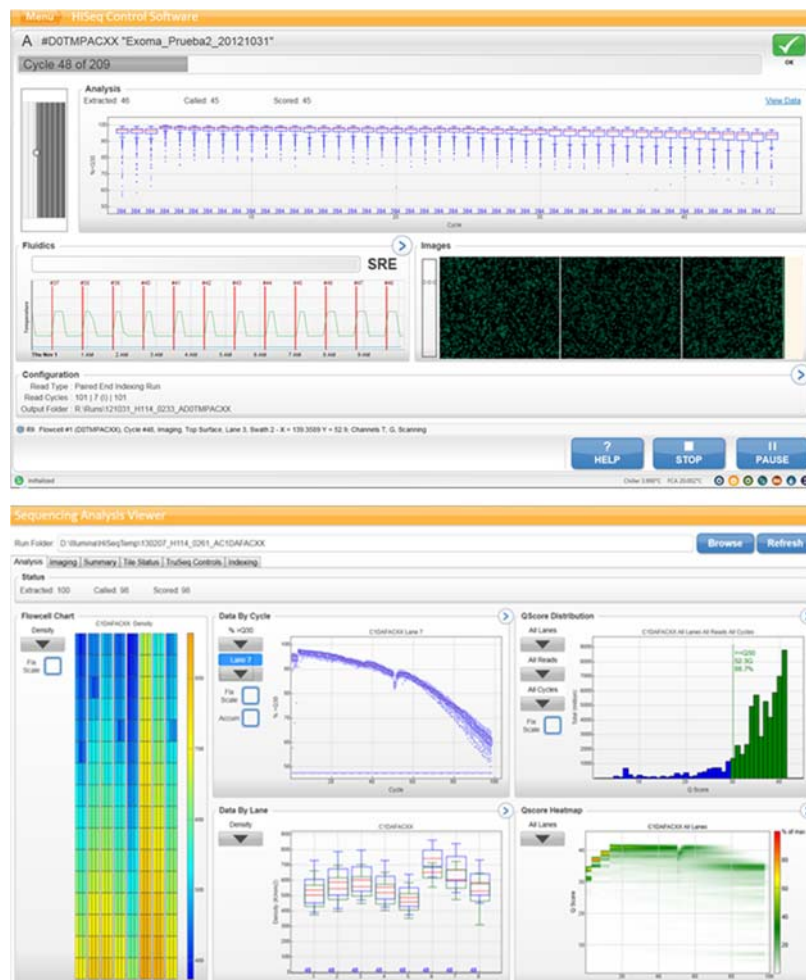


Figura 24. Visualización del proceso de secuenciación a tiempo real. En la parte superior se representa el *HiSeq Control Software* y en la parte inferior el *Sequencing Analysis Viewer*.

Tras la secuenciación, se procedió al análisis bioinformático de los resultados obtenidos de la misma (en el Anexo de la presente memoria se detalla todo el pipeline bioinformático empleado para el análisis de las secuencias). Un esquema general de los procesos seguidos en el análisis junto con las diferentes herramientas bioinformáticas necesarias en cada uno de los pasos se representa en la Figura 25.

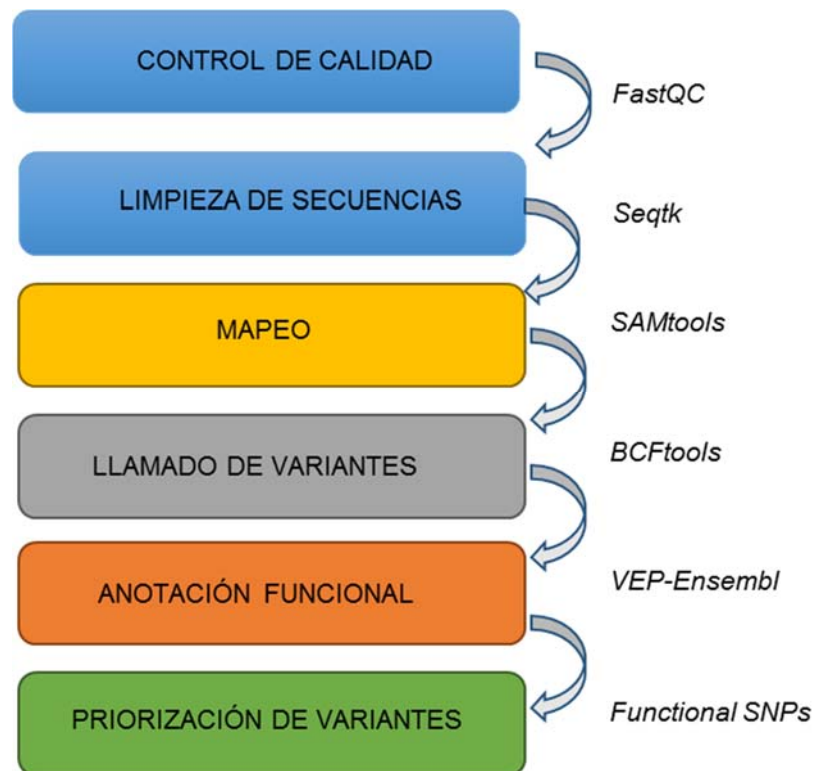


Figura 25. Esquema del análisis bioinformático de las secuencias y las diferentes herramientas de análisis utilizadas.

3.1. Control de calidad y mapeo de las secuencias

Los secuenciadores de nueva generación, como el sistema HiScanSQ de la plataforma Illumina, generan cientos de millones de secuencias en una sola carrera de secuenciación. Antes de analizar los resultados de la secuenciación y extraer conclusiones, siempre es necesario realizar controles de calidad. Su finalidad es asegurar que los datos brutos son correctos y no hay sesgos en los datos que puedan afectar al análisis de resultados posteriores o a los propios resultados del experimento.

A continuación se describen una serie de controles de calidad de datos, filtros y parámetros bioinformáticos que se aplicaron con la finalidad de obtener unos resultados fiables y de calidad.

3.1.1. Control de calidad: FastQC

El tipo de formato de las lecturas de secuenciación obtenidas tras la secuenciación es FastQ. Para cada uno de los individuos secuenciados se obtienen dos archivos al realizarse la secuenciación desde los dos extremos (*paired-end reads*). En este tipo de archivos, la calidad de las lecturas se expresa en base logarítmica (*Phred Score*, $Q = -10 \log_{10} P$). Así, por ejemplo, una calidad Q20 significa una probabilidad de 0.01 de haber leído una base erróneamente (Figura 26).

```
@H114:263:C0K6MACXX:7:1101:1123:1976 1:N:0:CGATGT
GTGGATACCTGTTGCTGACTATAAGAGATGAGACTG
+
@@CFFFFFHJHHHJJJFHGEFHFEHHJJJJJJ
```

Figura 26. Representación de una lectura del secuenciador en formato FastQ. La primera línea muestra tras la “@” diversa información acerca de la máquina, la muestra, el índice, la carrera o la posición de la lectura. La segunda línea es la secuencia de bases leída. La última línea informa de la calidad de la lectura (*Phred Score*) en formato ASCII.

FastQC tiene como objetivo realizar un control de calidad de los datos de secuenciación de forma sencilla y rápida. Las principales funciones del FastQC son: a) importar archivos FastQ, b) identificar si hay algún problema ya sea en el secuenciador o en la librería del ADN de partida, c) proporcionar gráficos y tablas para evaluar los datos de forma rápida y d) generar informes de calidad de las secuencias obtenidas.

Los diferentes parámetros de control que se pueden evaluar mediante FastQC:

- Calidad de la secuencia por base: muestra la calidad de las bases según la posición de lectura. Los datos son buenos si la calidad de las lecturas en las distintas posiciones es alta y homogénea (las barras que muestran el rango intercuartílico 25-75% son pequeñas, de tamaño semejante en todas las posiciones de lectura y de alta puntuación). Son malos datos si la calidad de las lecturas largas es baja, observándose un aumento de variabilidad en las lecturas de las últimas posiciones.
- Puntuaciones de calidad de la secuencia: representa la distribución de la calidad de las lecturas. Permite ver si hay algún subconjunto de secuencias con valores bajos de calidad.

- Contenido de bases por secuencia: informa de la frecuencia de cada base en cada posición de lectura. Al emplear la plataforma Illumina, que utiliza librerías de fragmentos generados de forma aleatoria, no deberían aparecer diferencias significativas en las frecuencias de las distintas bases en cada posición de lectura. Se considera que hay un error cuando la diferencia entre A y T o G y C es mayor del 20% en alguna de las posiciones de lectura.
- Contenido de GC por base: indica la distribución de GC por posición de lectura. En una librería aleatoria debería ser homogéneo.
- Contenido de GC por secuencia: muestra el porcentaje de GC por secuencia.
- Contenido de Ns por base. Se utiliza N para marcar las bases de las que no se ha obtenido lectura.
- Distribución de la longitud de secuencias: representa las longitudes de las lecturas en pares de bases (pb).
- Duplicados de secuencias: indica el grado de duplicación de cada secuencia. Cuanto menor es este número, mejor es la calidad de los datos, y suele traducirse en un adecuado nivel de cobertura de las secuencias.
- Secuencias sobrerrepresentadas: informa de la presencia de secuencias que aparecen en mayor cantidad de la esperada. Generalmente suelen corresponder a las secuencias de los cebadores utilizados para la PCR.
- Contenido en K-mer: K-mer se refiere a oligómeros específicos (de longitud k) que se pueden utilizar para localizar regiones de interés. Muestra la cantidad de veces que esas secuencias aparecen en la muestra. Por defecto, muestra los pentámeros que aparecen en mayor frecuencia de la esperada. Puede identificar problemas, por ejemplo, secuencias que corresponden a los adaptadores localizadas en posiciones intermedias de las lecturas. Igualmente, puede mostrar si el enriquecimiento de todas las secuencias ha sido homogéneo o algunas están sobre-enriquecidas.

3.1.2. Limpieza de secuencias: Seqtk

El siguiente paso consiste en eliminar tanto los adaptadores utilizados durante el proceso de secuenciación como aquellas bases que tengan una calidad baja, como por

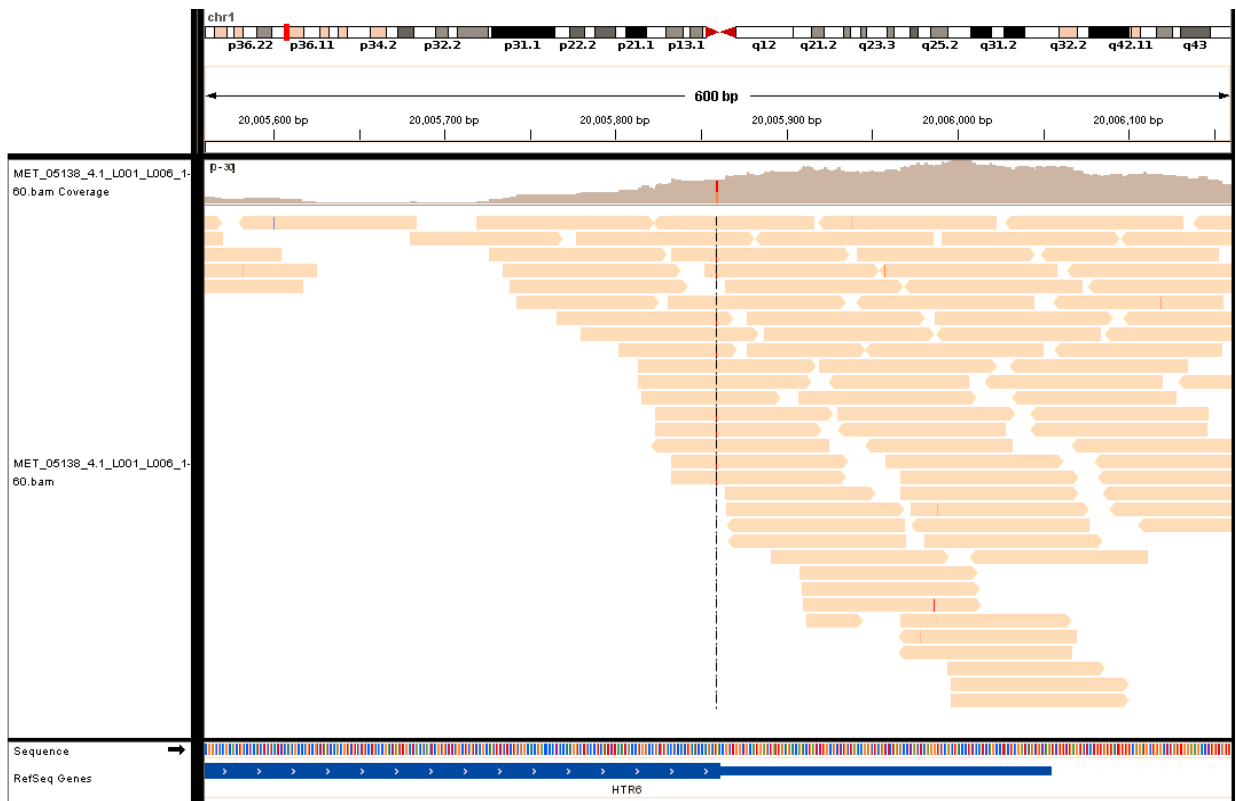


Figura 28. Ejemplo de visualización de una variante genética mediante el programa IGV.

3.2. Llamado de variantes genéticas

Para el llamado de las variantes empleamos el BCFtools que permite obtener dicha información en un archivo denominado VCF (*Variant Call Format*) a partir de la información contenida en los archivos BAM.

BCFtools forma parte del paquete SAMtools. Este último recopila la información del BAM y calcula probabilidades de los datos para cada posible genotipo, almacenando esta información en el formato BCF. A continuación, a través del BCFtools se hace el llamado de las variantes. El formato del archivo VCF resultante se muestra en la Figura 29.

```

##fileformat=VCFv4.1
##fileDate=20110413
##source=VCFtools
##reference=file:///refs/human_NCBI36.fasta
##contig=<ID=1,length=249250621,md5=1b22b98cdeb4a9304cb5d48026a85128,species="Homo Sapiens">
##contig=<ID=X,length=155270560,md5=7e0e2e580297b7764e31dbc80c2540dd,species="Homo Sapiens">
Header {
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##ALT=<ID=DEL,Description="Deletion">
##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE1 SAMPLE2
Body {
1 1 . ACG A,AT 40 PASS . GT:DP 1/1:13 2/2:29
1 2 . C T,CT . PASS H2;AA=T GT 0|1 2/2
1 5 rs12 A G 67 PASS . GT:DP 1|0:16 2/2:20
X 100 . T <DEL> . PASS SVTYPE=DEL;END=299 GT:GQ:DP 1:12:. 0/0:20:36

```

Figura 29. Formato del archivo VCF.

La cabecera de un archivo VCF se caracteriza por empezar por # o ## seguidas de diferentes datos entre ellos el programa, número de versión o genoma de referencia utilizado. Seguidamente, se muestra la información general de las variantes donde aparece la siguiente información:

- CHROM: Cromosoma
- POS: Posición en la secuencia de referencia (posición 1 es la primera base)
- ID: rs del SNP si se ha identificado previamente
- REF: base de referencia
- ALT: base alternativas
- QUAL: calidad. Cuanto mayor es la calidad más alta es esta variable.
- FILTER: PASS indica que esa posición ha pasado todos los filtros.
- INFO: información adicional (AA, alelo ancestral; DP, cobertura; etc)

La información específica sobre el genotipo se encuentra contenida en el apartado FORMAT y se compone de los siguientes parámetros:

- GT: Genotipo. Un valor de 0 indica el alelo de referencia, un valor de 1, el primer alelo alternativo en la lista, y un valor de 2, el segundo.
- DP: Cobertura en esa posición para esa muestra.
- GQ: Calidad del genotipo.

3.3. Anotación funcional de variantes genéticas

Para el llamado de variantes se empleó la herramienta VEP (*Variant Effect Predictor*) de Ensembl el cual determina el efecto de las variantes genéticas, SNPs, en los genes, transcritos y secuencias proteicas así como en regiones reguladoras. Este programa nos permite obtener información de:

- Genes y transcritos donde se encuentran nuestras variantes
- Localización de las variantes genéticas (por ejemplo, en secuencias codificantes)
- Consecuencia de la variante genética en la secuencia proteica (por ejemplo, pérdida de un codón stop)
- Frecuencia del alelo minoritario (MAF) de la variante genética en población europea (según El Proyecto 1000 Genomas)
- Test predictivos de funcionalidad: SIFT y PolyPhen. Se utilizan para la identificación de variantes con efecto negativo sobre el fenotipo. SIFT predice si la sustitución de un aminoácido por otro afecta a la función de la proteína basándose en la homología de la secuencia y las propiedades físicas de los aminoácidos. PolyPhen predice el posible efecto de la sustitución de aminoácidos según la estructura y función de la proteína utilizando consideraciones físicas y comparativas directas.

Las variantes filtradas y seleccionadas por su impacto funcional fueron: a) variantes en regiones de *splicing*, b) variantes con cambio de sentido (*missense*), c) variantes en regiones que codifican para microRNAs y, d) variantes de codón de stop. También se anotaron en un inicio, todas las variantes genéticas que están en regiones exónicas. A continuación, se definen las variantes más importantes que se identificaron empleando la herramienta bioinformática VEP:

- **Variante en región donante de *splicing* (splice donor variant):** variante que afecta a la región de *splicing* en el extremo 5' de un intrón.
- **Variante en región aceptora de *splicing* (splice acceptor variant):** variante que afecta a la región de *splicing* en el extremo 3' de un intrón.
- **Variante de ganancia de codón de parada (stop gained):** variante en la secuencia que ocasiona un cambio en un codón de tal forma que se origina un codón de stop prematuro y, por tanto, un péptido más corto.

- **Variante de pérdida de codón de parada (stop lost):** variante en la secuencia que ocasiona la desaparición de un codón stop, dando lugar a un péptido más largo.
- **Variante con cambio de sentido (*missense variant*):** variante en la secuencia que cambia una o más bases, resultando en una secuencia de aminoácidos diferente, pero con la misma longitud.
- **Variante en microARN maduro (microARN mature):** variante en la secuencia que codifica para un microARN.

Tras la obtención de todas las variantes funcionales obtenidas en los 400 individuos se priorizó su estudio en función de:

- a) Genotipos con una cobertura mínima de 4x y un valor de calidad de genotipado mayor de 20.
- b) Presencia de la variante genética funcional en casos o controles, no en ambos.
- c) MAF en la población europea inferior a 20%.
- d) Ambos predictores, SIFT y PolyPhen, con puntuaciones significativas.

A continuación, estudiamos mediante *VEP run statistics* las variantes genéticas con posible efecto funcional en el grupo de casos y controles. Finalmente, se analizaron las posibles funciones moleculares, procesos biológicos o rutas metabólicas de los genes donde se encontraron las diferentes variantes encontradas empleando herramientas informáticas (PANTHER, *Gene List Analysis*) (241).

4. VERIFICACIÓN DE LAS VARIANTES GENÉTICAS MEDIANTE SECUENCIACIÓN CLÁSICA SANGER

Las variantes genéticas encontradas en codones de parada o stop en las muestras de casos y controles fueron comprobadas mediante secuenciación Sanger y electroforesis capilar. El motivo de esta validación fue comprobar que los resultados encontrados mediante secuenciación de nueva generación se correspondían a los de secuenciación por Sanger y por tanto, eran reales. En ocasiones pueden producirse errores en la anotación de variantes debido a que hay errores en la secuencia, la calidad de los datos es baja, cobertura insuficiente, etc.

La secuenciación se ha realizado mediante “BigDye® Terminator v3.1 CycleSequencing kit” (Applied Biosystems) y la electroforesis capilar en un secuenciador ABI3730 (Applied Biosystems). Se emplearon los programas de Primer 3 (versión 4.0.0), BLAST (Basic Local Alignment Search Tool) y Ensembl Genome Browser para diseñar todos los oligonucleótidos con la finalidad de amplificar y secuenciar la región del genoma dónde se encontraba cada una de las variantes genéticas previamente identificadas. Mediante los programas de Sequencing Analysis y Sequencher se analizaron las secuencias permitiendo visualizar en un cromatograma si el individuo es homocigoto o heterocigoto para la variación encontrada (Figura 30).

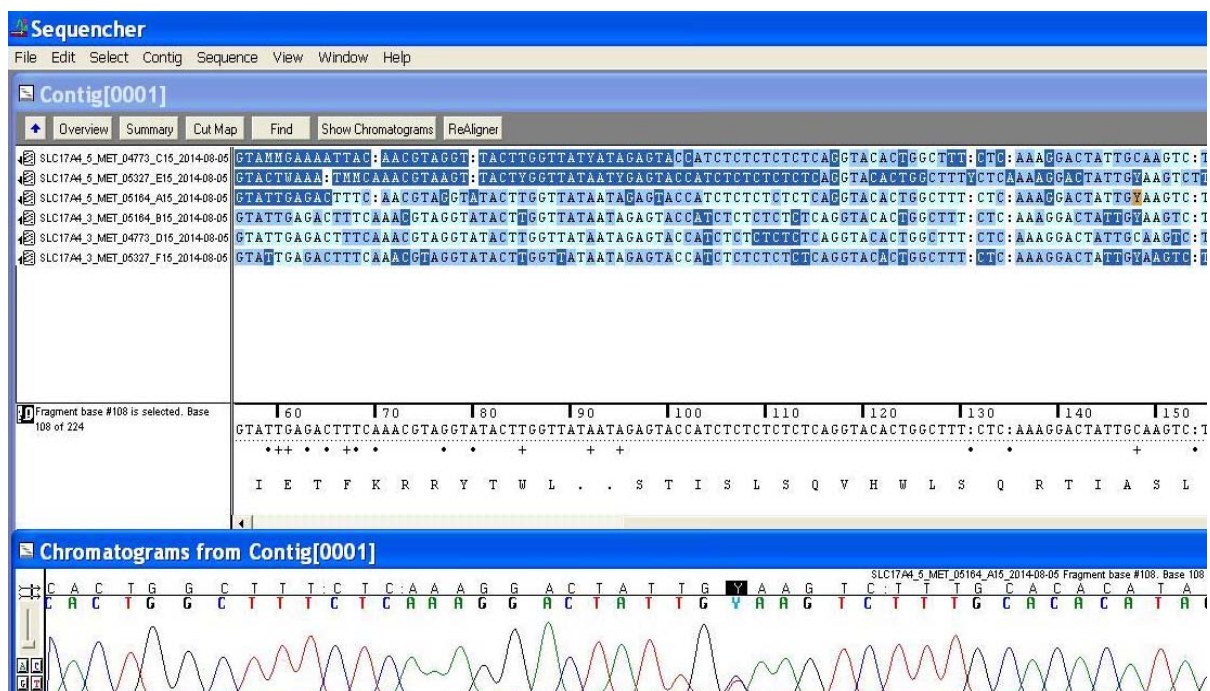


Figura 30. Visualización en un cromatograma una de las variantes genéticas de stop identificada y verificada tras secuenciación por Sanger.

Resultados

1. CARACTERÍSTICAS GENERALES DE LA POBLACIÓN DE ESTUDIO

Las características de la población de estudio tras la selección final de los individuos para ambos grupos, control y casos, se muestran en la Tabla 8. La edad media fue de 53 y 56 años y los niveles de glucosa fueron de 82,26 mg/dL y de 162,09 mg/dL para el grupo control y casos, respectivamente. Los datos de glucosa para el grupo de casos (n=200) incluye a individuos con y sin tratamiento para la DM2; no obstante, el 83,5% de los individuos estaban tratados.

Tabla 8. Características descriptivas de la población de estudio.

	Controles	Casos
Número de individuos	200	200
Sexo (hombres)	121	121
Edad	52,91 ± 0,53	56,17 ± 0,47
Glucosa (mg/dL)	82,26 ± 0,50	162,09 ± 4,31
IMC (kg/m ²)	29,38 ± 0,18	29,74 ± 0,20
Colesterol Total (mg/dL)	202,43 ± 2,70	184,48 ± 4,37
LDL (mg/dL)	112,31 ± 1,90	109,64 ± 4,49

Con la finalidad de obtener dos grupos, controles y casos, lo más similares posible, el primer paso fue la selección de las muestras a partir de las bases de datos disponibles. Tras la selección, las características generales para ambos grupos fueron homogéneas para los parámetros de edad, IMC, colesterol y LDL, exceptuando para los niveles de glucosa (Figura 31 y Figura 32).

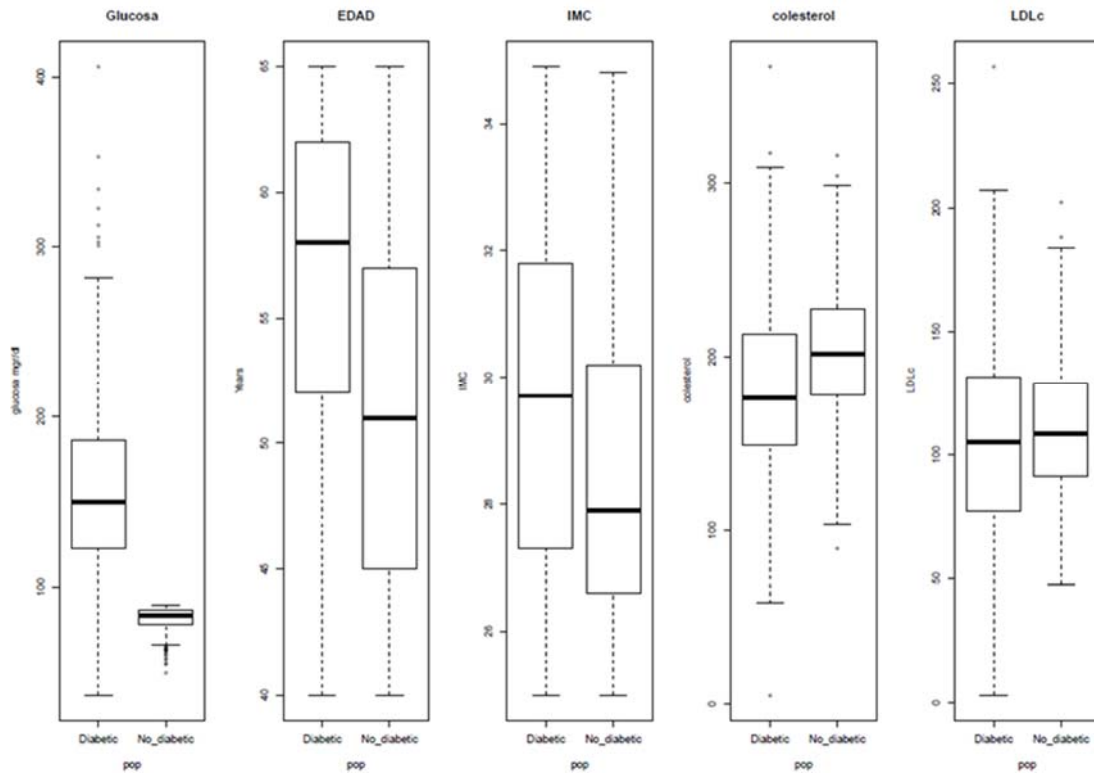


Figura 31. Boxplot de muestras disponibles para seleccionar el grupo de controles (*No Diabetic*) y casos (*Diabetic*). El diagrama de cajas representa los valores para los percentiles 25, 50 (mediana, línea) y 75, en cada uno de los parámetros de estudio.

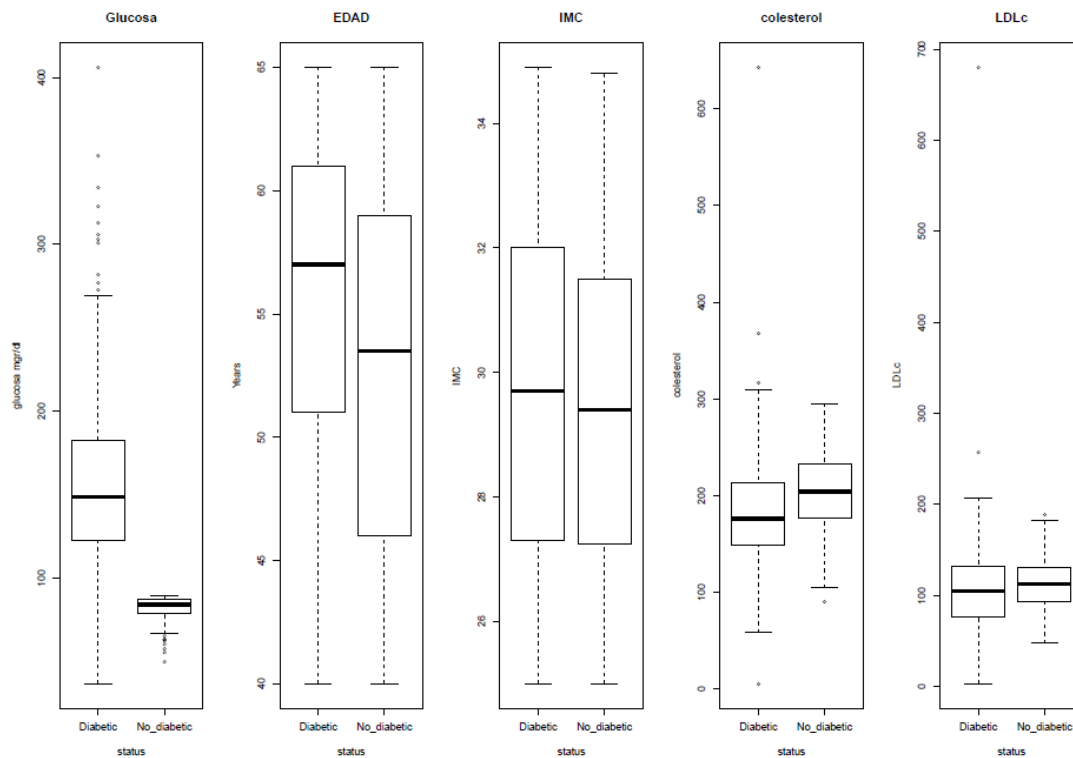


Figura 32. Boxplot de las muestras seleccionadas para el grupo de controles (*No Diabetic*) y casos (*Diabetic*). El diagrama de cajas representa los valores para los percentiles 25, 50 (mediana, línea) y 75, en cada uno de los parámetros de estudio.

En la selección se tuvo en cuenta la distribución de sexo e IMC con la finalidad de homogenizar ambos grupos. Tras la selección de las muestras tanto hombres como mujeres quedaron representados de manera uniforme (Figura 33).

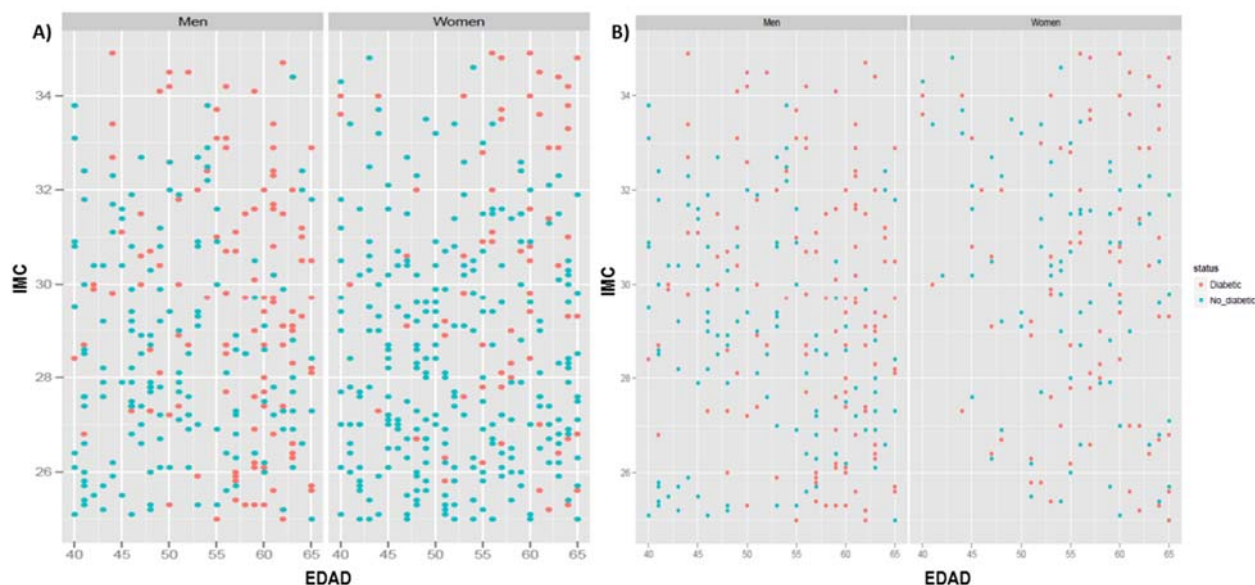


Figura 33. Representación de la distribución por sexo según el índice de masa corporal (IMC) en el grupo de controles (*No Diabetic*) y casos (*Diabetic*) antes (A) y después (B) de la selección.

2. CONTROL DE CALIDAD DE LA SECUENCIACIÓN: FASTQC

Tras realizar el control de las secuencias con el programa FastQC, se examinaron los resultados obtenidos con el fin de comprobar la calidad de la secuenciación realizada. Así pues, se analizaron diferentes parámetros de control de calidad de los datos obtenidos tras la secuenciación con Illumina.

En primer lugar, se analizó la calidad de la secuencia por base (Figura 34). La calidad media de secuenciación para cada base fue alta y homogénea (eje X). Se observó que conforme aumenta la longitud de la secuencia, baja la calidad de secuenciación o, lo que es lo mismo, el valor de Q (eje Y). Este resultado es normal por el propio ruido de fondo originado por el procedimiento de secuenciación, que genera una disminución de la calidad al aumentar los ciclos de secuenciación. Las secuencias presentaron un $Q \geq 30$ durante casi toda la reacción, indicando que los datos obtenidos tras la secuenciación eran fiables.

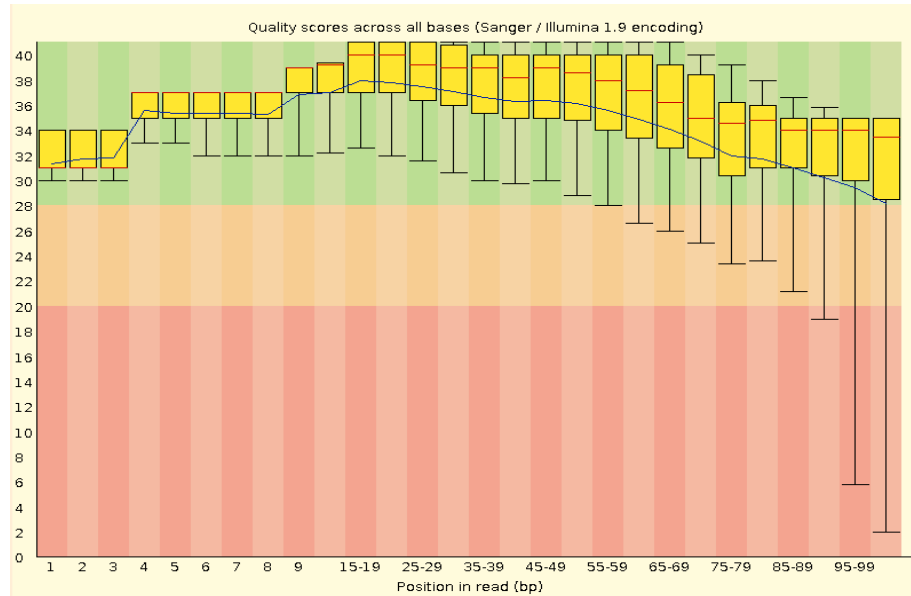


Figura 34. Calidad de la secuencia por base y por muestra. La línea roja representa la mediana y la línea azul la calidad media de secuenciación.

Otro parámetro de calidad analizado fue el contenido de bases en la secuencia (Figura 35). Los resultados mostraron que la secuenciación había sido correcta, mostrando un balance entre las bases complementarias y sin presentar diferencias significativas en sus frecuencias en cada posición de lectura.

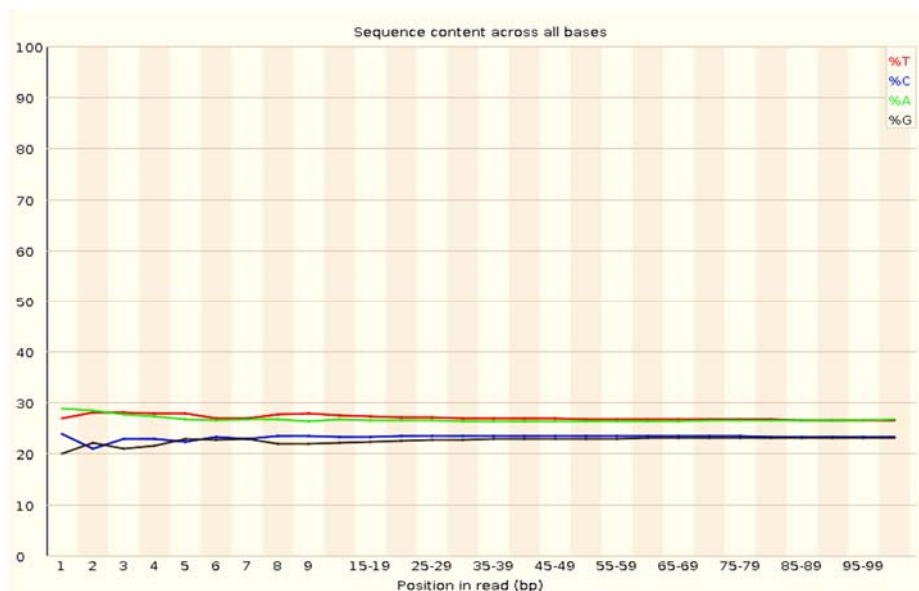


Figura 35. Contenido de las bases de la secuencia (A-T y G-C). Muestra la frecuencia de cada base en cada posición de lectura.

La mayoría de las secuencias coinciden en un mismo contenido medio de GC; característico de cada organismo. Así, tras analizar la distribución de GC de nuestras muestras, se observó que la distribución de GC real y teórica era muy parecida (Figura 36).

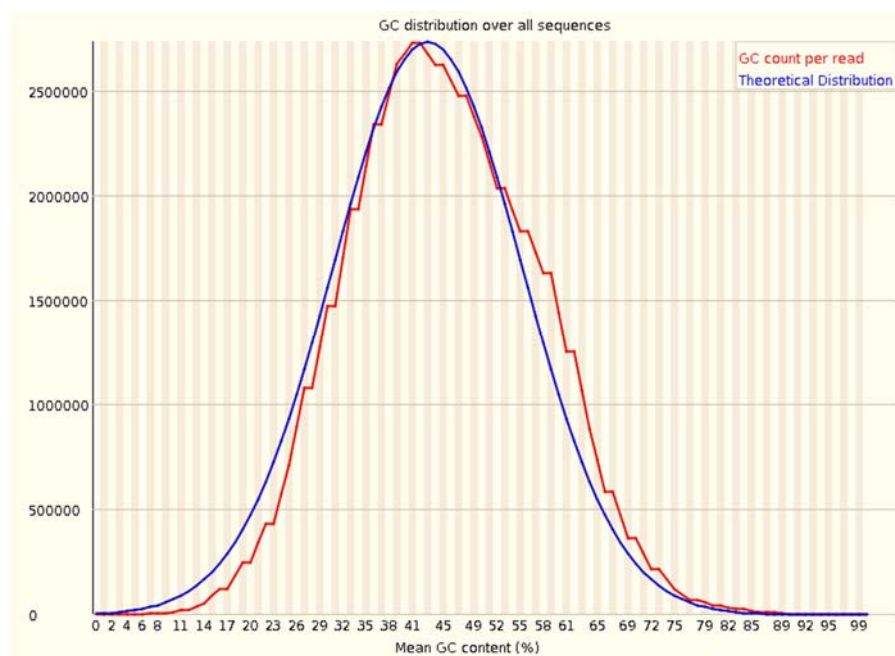


Figura 36. Contenido de GC por secuencia. Muestra la distribución de GC por secuencia. La línea azul muestra la distribución esperada y la roja la real.

Además, se comprobó que la secuenciación fue correcta y se realizó de manera adecuada, presentando un porcentaje de bases ambiguas (N) por base igual a cero (Figura 37).

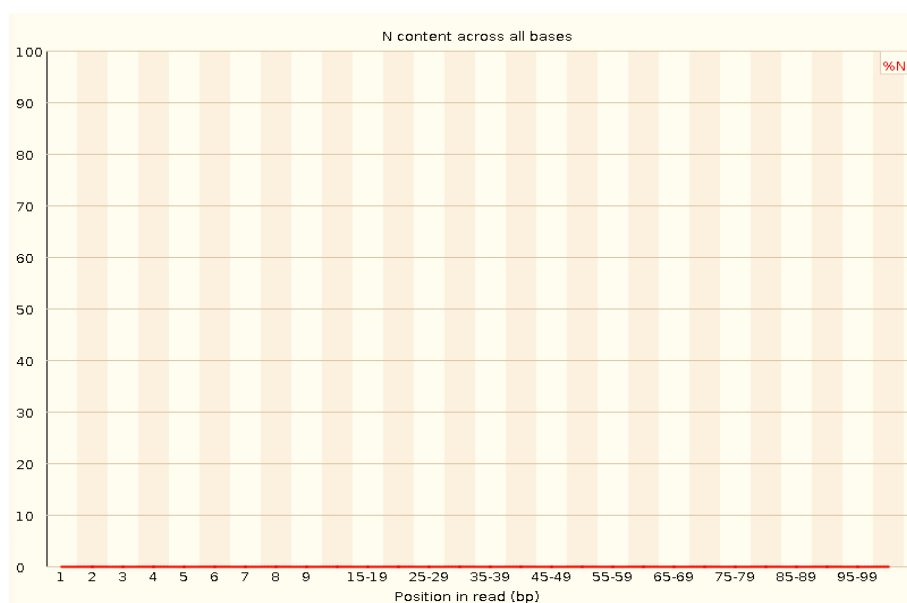


Figura 37. Contenido de bases sin lectura. Se observa como el porcentaje de nucleótidos no identificados en cada secuencia (N) es cero.

Por tanto, los resultados de los parámetros de control de calidad de la secuenciación permitieron concluir que las secuencias obtenidas eran correctas, presentando buenas calidades y un porcentaje de bases equilibrado.

3. COBERTURA DE LOS DATOS SECUENCIADOS

Después de realizar el control de calidad de la secuenciación, se analizaron los datos de cobertura de las secuencias obtenidas. La cobertura media de las muestras fue alrededor de 20x (Figura 38), la cual fue adecuada para el tipo de experimento diseñado. Esta cobertura fue suficiente para identificar un gran número de variantes genéticas, incluyendo variantes poco frecuentes y raras.

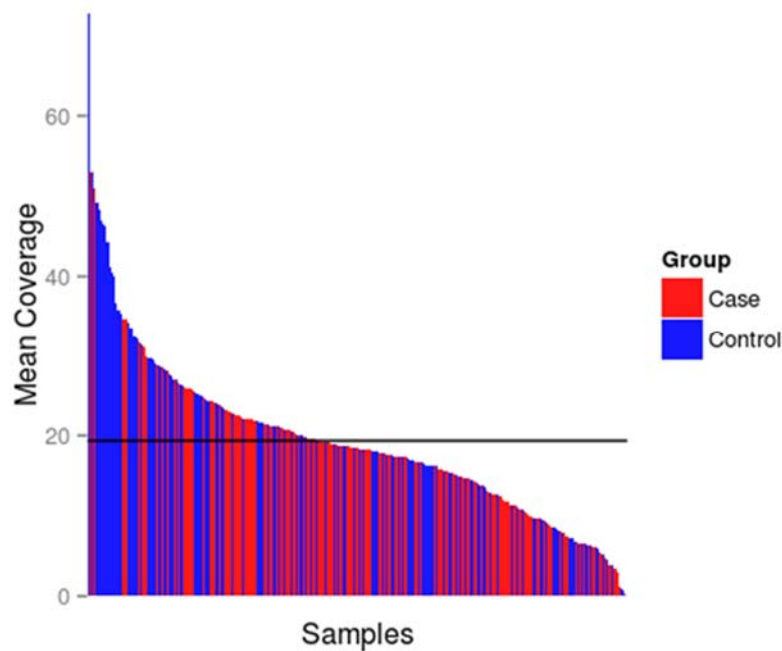


Figura 38. Cobertura media de secuenciación para el grupo de casos y de controles. Las muestras están coloreadas según el grupo al que pertenecen. La línea negra indica la media total.

Después de analizar la cobertura para cada una de las muestras secuenciadas en el grupo de casos y controles, se estudió el porcentaje de bases que tenían una cobertura mínima a 10x y 20x para cada uno de los grupos. Los resultados mostraron que alrededor del 60% de las secuencias tenían una cobertura mínima de 10x y alrededor del 40% tenían un 20x (Figura 39).

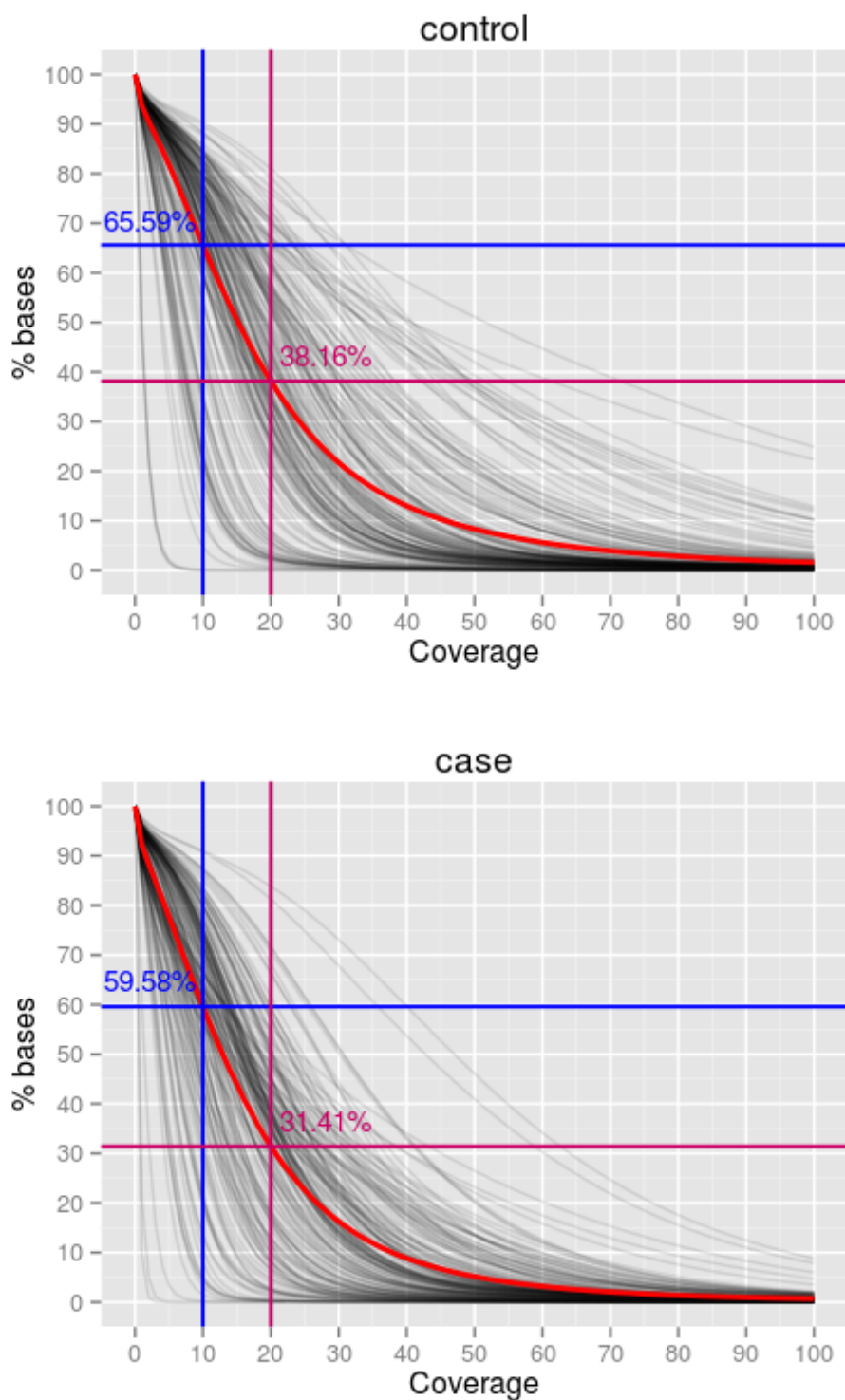


Figura 39. Porcentaje de bases que tienen una cobertura mínima concreta en los grupos control y casos. Se señala para cada muestra (líneas grises) el porcentaje de bases que tienen una cobertura mínima concreta a 10x y 20x para controles (parte superior) y para casos (parte inferior). La línea roja es la media del conjunto de muestras en cada gráfica.

Se analizó la cobertura para las diferentes regiones potencialmente analizables en base a si estaban incluidas dentro de las sondas o en regiones contiguas a éstas (región extendida). La región extendida en este caso se corresponde con unas 100 pb por cada extremo de la sonda.

Como puede observarse en la Figura 40, la cobertura media fue mayor en la región de las sondas y menor en la zona extendida contigua a las sondas. Estos resultados indican que el método de captura y alineamiento de las secuencias se realizó correctamente, aumentando la cobertura en aquellas zonas de interés donde va dirigida la sonda, es decir, las regiones exónicas.

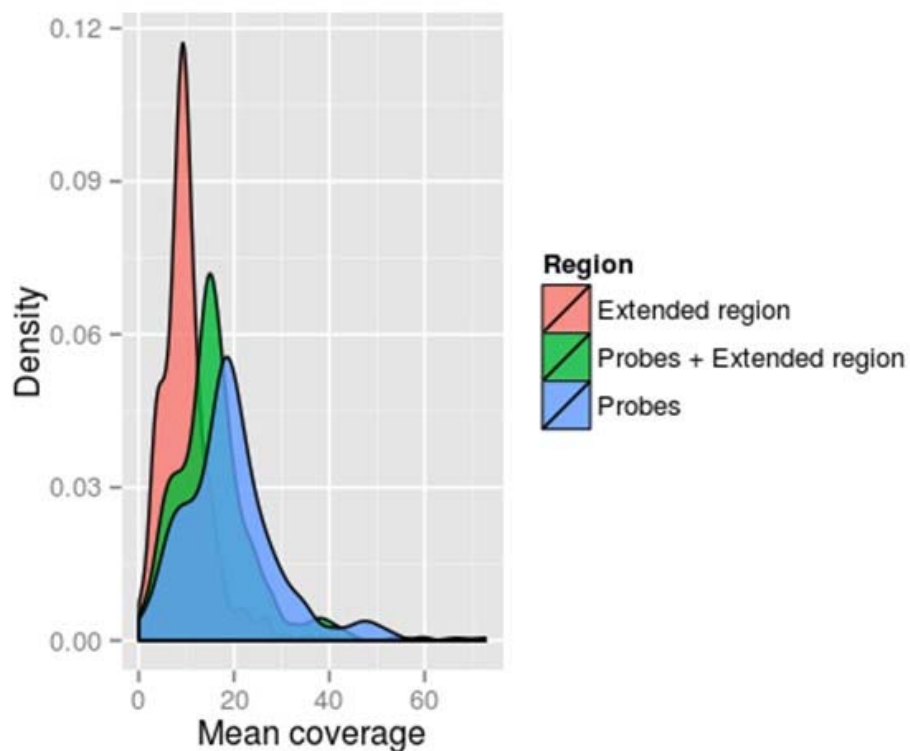


Figura 40. Densidad para la cobertura media en las diferentes regiones cubiertas por las sondas. La densidad de cobertura se analizó para cada una de las tres regiones: región extendida (rosa), dentro de las sondas (azul) y dentro de las sondas más la región extendida (verde).

La cobertura media por muestra en las diferentes regiones cubiertas y para los dos grupos, control y casos, indica que la cobertura media por muestra fue mayor en las zonas de las sondas, siendo de media de 20x para ambos grupos (Tabla 9).

Tabla 9. Cobertura media por muestra en las diferentes regiones cubiertas.

	Grupo	Cobertura media por muestra
Sonda	control	21,35
	caso	17,38
Sonda ± Región extendida	control	17,37
	caso	14,00
Región extendida	control	11,07
	caso	8,64

*Región extendida = 100 pb

Seguidamente, se analizó el porcentaje de bases con una cobertura determinada dentro de las sondas para los dos grupos (Tabla 10). Alrededor del 40% de las bases secuenciadas presentaron una cobertura de 20x. Ello resultó en una cobertura suficiente para llevar a cabo el objetivo de identificar el mayor número de variantes genéticas posible en una gran cantidad de muestras.

Tabla 10. Porcentaje de bases con una cobertura determinada dentro de las sondas para controles y casos.

	5x	10x	15x	20x	25x	30x
Controles	81,74	65,59	50,54	38,16	28,68	21,70
Casos	77,66	59,57	43,75	31,41	22,46	16,19

4. VARIANTES GENÉTICAS IDENTIFICADAS TRAS LA SECUENCIACIÓN DEL EXOMA

Tras el análisis bioinformático de las variantes genéticas encontradas en las 400 muestras, se identificaron todas aquellas variantes con una posible funcionalidad biológica tanto para casos como para controles (Figura 41 y Figura 42).

El número de variantes genéticas encontradas antes de aplicar ninguno de los filtros establecidos fue muy elevado; siendo alrededor de 1 millón el número de SNPs encontrados en ambos grupos. Las variantes en regiones intrónicas, variantes en genes que codifican para un ARN no codificante o variantes intergénicas fueron las más abundantes. Por otro lado, las variantes menos frecuentes fueron las presentes en regiones codificantes para microARNs, variantes de stop o de *splicing*.

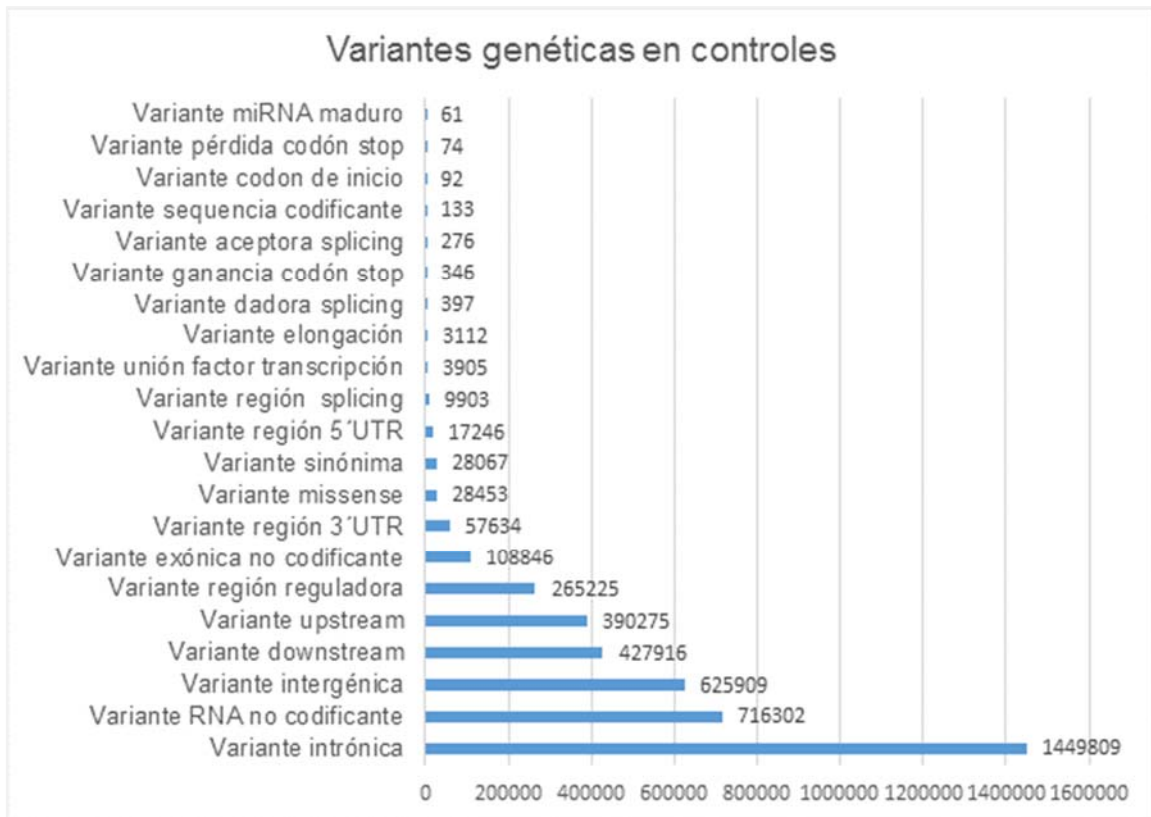


Figura 41. Representación del número de variantes genéticas encontradas para el grupo de controles clasificadas según su impacto funcional.



Figura 42. Representación del número de variantes genéticas encontradas para el grupo de casos clasificadas según su impacto funcional.

Tras observar el gran número de variantes genéticas identificadas y con posible funcionalidad biológica, decidimos establecer una serie de parámetros y filtros que permitieran identificar únicamente aquellas variantes genéticas de mayor importancia a nivel funcional.

5. SELECCIÓN DE VARIANTES GENÉTICAS CON POSIBLE EFECTO FUNCIONAL

Después de secuenciar, analizar e identificar las variantes genéticas de los 400 individuos, se seleccionaron las variantes de mayor relevancia y se clasificaron según la consecuencia funcional para ambos grupos (Tabla 11). Como se ha descrito previamente en metodología, las variantes fueron seleccionadas y priorizadas de acuerdo una serie de criterios:

- Cobertura igual o superior a 4 y una calidad de genotipado mayor de 20.
- Variante genética presente en casos o controles, no en ambos.
- MAF inferior al 20%
- Ambos predictores de funcionalidad, SIFT y PolyPhen, con puntuaciones significativas.

Tabla 11. Número de variantes genéticas encontradas mediante secuenciación de nueva generación según su funcionalidad biológica.

	Controles	Casos
Variantes de codón de parada	110	67
Variantes de <i>splicing</i>	160	132
Variantes <i>missense</i>	1.817	1.614
Variantes en región de <i>splicing</i>	2.316	1.806
Variantes en secuencia codificante	104	113
Variantes en región codificante para un microARN maduro	23	11
Variantes en 5' UTR	4.090	3.047
Variantes en 3' UTR	13.202	10.448

Además de estudiar las consecuencias funcionales de mayor severidad, tuvimos en cuenta los valores predictivos de SIFT y PolyPhen para las variantes que ocasionaban cambios aminoacídicos. Así pues, se priorizaron aquellas variantes cuyos valores predictivos indicaban, en ambas puntuaciones, una consecuencia deletérea o dañina. En la Figura 43 se representa un resumen de las consecuencias funcionales

para el Cromosoma 1 como ejemplo, incluyendo los valores de predicción SIFT y PolyPhen.

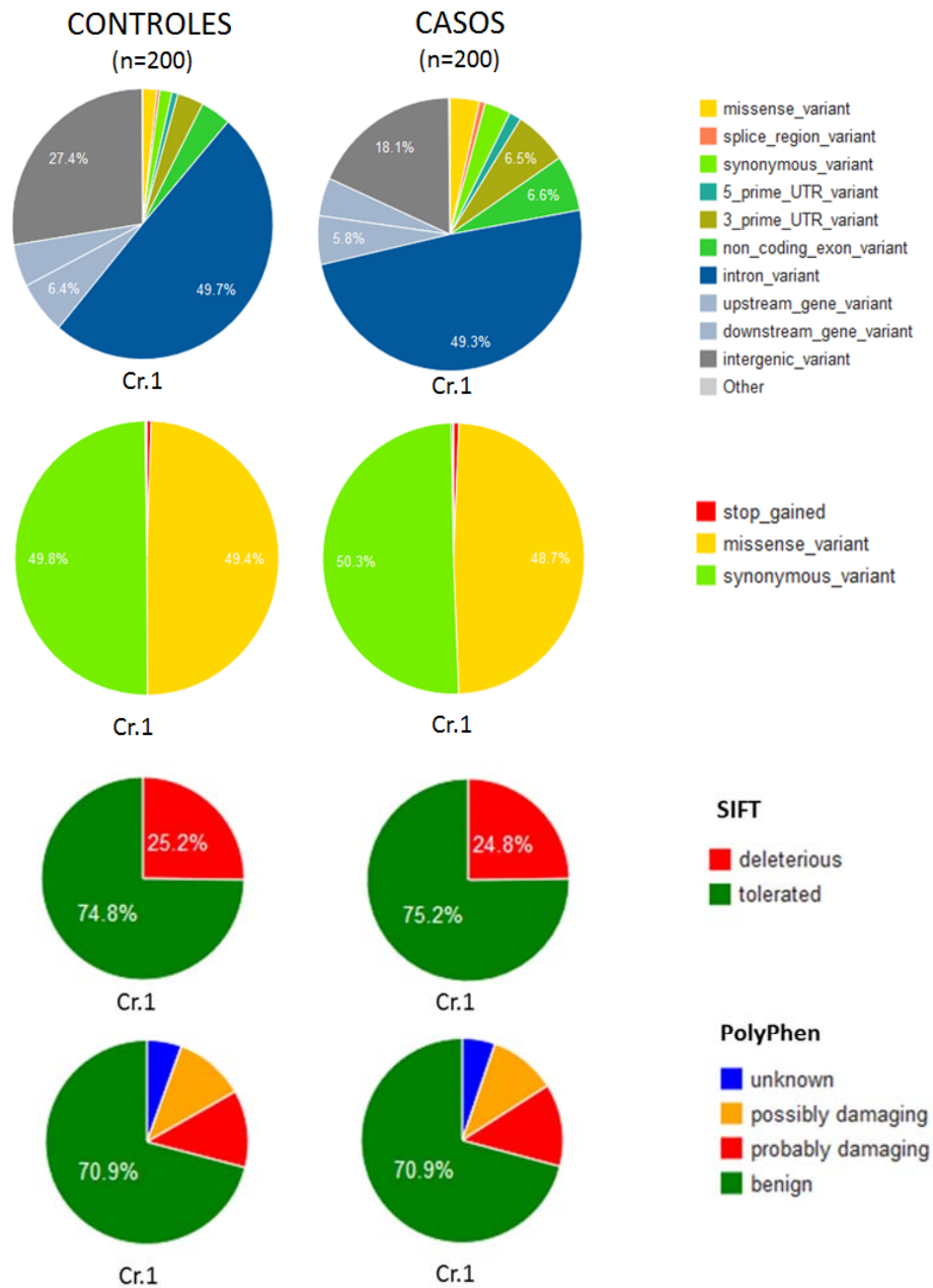


Figura 43. Representación gráfica de las variantes genéticas con posible consecuencia funcional para el cromosoma 1 en el grupo de controles y de casos.

Concretamente, este proyecto se centró en el estudio de variantes genéticas de un sólo nucleótido (SNP) con impacto funcional; en particular: variantes de *splicing* y *missense*, variantes presentes en secuencias codificantes de microARNs y variantes de stop o parada.

6. VARIANTES GENÉTICAS DE *SPLICING*, *MISSENSE* Y EN SECUENCIAS CODIFICANTES PARA microARNs

6.1. Estudio de variantes genéticas de *splicing* en el exoma

Las variantes genéticas de *splicing* afectan al procesamiento de los intrones. Las variantes genéticas de mayor impacto funcional son aquellas situadas en el sitio donante o aceptor de *splicing*. Tras el procesamiento y filtrado de los datos, se obtuvieron 160 SNPs en controles (Tabla 12) y 132 SNPs en casos (Tabla 13). Dentro de las variantes de *splicing* encontramos algunas variantes que a su vez también pueden ser variantes *missense*. Los criterios establecidos para definir una variante de *splicing*, según la herramienta de anotación funcional VEP, es toda variante que cambie +/- dos pares de bases con respecto a la región 3' UTR (aceptora) o 5' UTR (dadora) de un intron.

Algunas de las variantes genéticas identificadas fueron variantes poco frecuentes (MAF <0,05) e incluso raras (MAF <0,01); siendo alguno de ellos SNPs nuevos no identificados previamente.

A continuación, se analizaron las posibles funciones moleculares, procesos biológicos o rutas metabólicas de los genes donde se identificaron variantes de *splicing* tras la secuenciación del exoma. Para ello, se emplearon herramientas informáticas (PANTHER, *Gene List Analysis*) que permitieron estudiar que rutas metabólicas y funciones moleculares estaban representadas en mayor medida en el total de genes analizados.

6.1.1. Variantes genéticas de *splicing* en el grupo control

Las variantes genéticas de *splicing* presentes en el grupo control se detallan en la Tabla 12, indicando la posición cromosómica, el cambio de nucleótido para ese SNP, su identificador en base a los datos de *Ensembl*, el nombre del gen, el tipo de variación (donante o aceptora de *splicing*), el SNP y MAF europeo si se conoce.

Tabla 12. Variantes genéticas de *splicing* en el grupo de controles (n=200).

Posición Cr.	Cambio	Ensembl ID	Gen	Tipo de Variación	SNP	MAF europeo
1:20640814	G/A	ENSG00000158816	VWA5B1	Variante en región aceptora de <i>splicing</i>	-	-
1:22453975	C/G	ENSG00000162552	WNT4	Variante en región aceptora de <i>splicing</i>	rs2235527	0,2
1:225328432	G/T	ENSG00000185842	DNAH14	Variante en región aceptora de <i>splicing</i>	rs192476647	0,02
1:172431386	G/A	ENSG00000180999	C1orf105	Variante en región donante de <i>splicing</i>	rs41310899	0,09
1:26644537	C/A	ENSG00000158062	UBXN11	Variante en región donante de <i>splicing</i>	rs77928789	0,0026
1:32230280	C/T	ENSG00000121753	BAI2	Variante en región donante de <i>splicing</i>	rs71516323	0,04
2:138030234	C/T	ENSG00000144229	THSD7B	Variante en región donante de <i>splicing</i>	rs12622896	0,02
2:1507598	G/T	ENSG00000115705	TPO	Variante en región donante de <i>splicing</i>	rs28912994	0,03
2:163136505	C/G	ENSG00000115267	IFIH1	Variante en región donante de <i>splicing</i>	rs35337543, CS0920222	0,01
2:47572363	G/C	ENSG00000119888	EPCAM	Variante en región donante de <i>splicing</i>	rs17036526	0,17
2:144903307	C/A	ENSG00000121964	GTDC1	Variante en región aceptora de <i>splicing</i>	-	-
2:209138744	G/T	ENSG00000115020	PIKFYE	Variante en región aceptora de <i>splicing</i>	-	-
3:129151450	C/A	ENSG00000129071	MBD4	Variante en región aceptora de <i>splicing</i>	-	-
3:167023739	C/T	ENSG00000169064	ZBBX	Variante en región aceptora de <i>splicing</i>	-	-
3:49723167	T/G	ENSG00000173531	MST1	Variante en región aceptora de <i>splicing</i>	-	-
3:132337477	C/A	ENSG00000240303	ACAD11	Variante en región donante de <i>splicing</i>	rs41272317, COSM149473	0,02
3:180379648	C/G	ENSG00000145075	CCDC39	Variante en región donante de <i>splicing</i>	rs397515392	-
4:170458959	C/A	ENSG00000137601	NEK1	Variante en región donante de <i>splicing</i>	-	-
4:69917321	G/T	ENSG00000171234	UGT2B7	Variante en región donante de <i>splicing</i>	rs11935951	0,2
4:39246987	A/C	ENSG00000157796	WDR19	Variante en región aceptora de <i>splicing</i>	-	-
5:148989259	G/A	ENSG00000183111	ARHGEF37	Variante en región donante de <i>splicing</i>	rs200166388	-
5:161274793	G/T	ENSG00000223355	GABRA1	Variante en región donante de <i>splicing</i>	rs191560793	0
5:169131080	G/A	ENSG00000134516	DOCK2	Variante en región donante de <i>splicing</i>	rs264843	0,16
5:54721690	C/A	ENSG00000067113	PPAP2A	Variante en región donante de <i>splicing</i>	-	-
6:161006077	C/T	ENSG00000198670	LPA	Variante en región donante de <i>splicing</i>	rs41272114, CS9922248	0,03

6:31378575	G/A	ENSG00000204520	MICA	Variante en región donante de <i>splicing</i>	rs181430930	0
7:101952374	A/G	ENSG00000160999	SH2B2	Variante en región aceptora de <i>splicing</i>	rs76100825	0,06
7:141897443	G/C	ENSG00000257743	RP11-1220K2.2	Variante en región aceptora de <i>splicing</i>	-	-
7:73452035	A/C	ENSG00000049540	ELN	Variante en región aceptora de <i>splicing</i>	-	-
7:117174420	G/T	ENSG0000001626	CFTR	Variante en región donante de <i>splicing</i>	rs77188391,CS001826,CS910420	-
7:129825030	A/G	ENSG00000146842	TMEM209	Variante en región donante de <i>splicing</i>	rs118139998	0,01
7:158662182	G/A	ENSG00000126870	WDR60	Variante en región donante de <i>splicing</i>	rs842695	0,19
7:43659323	G/T	ENSG00000164543	STK17A	Variante en región donante de <i>splicing</i>	-	-
7:66459197	A/G	ENSG00000126524	SBDS	Variante en región donante de <i>splicing</i>	rs113993993,SBDSbase_D0076.g.2392T>C,CS030125	-
8:144463516	T/G	ENSG00000158106	RHPN1	Variante en región donante de <i>splicing</i>	-	-
8:16043667	C/T	ENSG00000038945	MSR1	Variante en región donante de <i>splicing</i>	rs435815	0,12
8:27649570	G/T	ENSG00000171320	ESCO2	Variante en región donante de <i>splicing</i>	-	-
8:75537578	T/C	ENSG00000254349	RP11-758M4.1	Variante en región donante de <i>splicing</i>	rs10957734	0,05
8:113318422	C/A	ENSG00000164796	CSMD3	Variante en región aceptora de <i>splicing</i>	-	-
8:17749317	C/G	ENSG00000104760	FGL1	Variante en región aceptora de <i>splicing</i>	rs35522859	0,04
9:114520363	C/A	ENSG00000165181	C9orf84	Variante en región donante de <i>splicing</i>	rs200554409	-
9:130236212	T/G	ENSG00000148356	LRSAM1	Variante en región donante de <i>splicing</i>	-	-
9:95640171	C/T	ENSG00000127081	ZNF484	Variante en región donante de <i>splicing</i>	rs117360596	0,03
10:37478486	G/T	ENSG00000148513	ANKRD30A	Variante en región donante de <i>splicing</i>	-	-
10:99258150	C/T	ENSG00000155229	MMS19	Variante en región aceptora de <i>splicing</i>	-	-
11:111945610	G/A	ENSG00000150776	C11orf57	Variante en región aceptora de <i>splicing</i>	rs59038861	0,004
11:3010503	T/G	ENSG00000205531	NAP1L4	Variante en región aceptora de <i>splicing</i>	rs7109106	0,1
11:32657377	C/A	ENSG00000186714	CCDC73	Variante en región aceptora de <i>splicing</i>	-	-
11:71279907	A/G	ENSG00000204572	KRTAP5-10	Variante en región aceptora de <i>splicing</i>	rs111350381	0,06
11:111753928	T/C	ENSG00000137720	C11orf1	Variante en región donante de <i>splicing</i>	-	-
11:116701354	G/A	ENSG00000110245	APOC3	Variante en región donante de <i>splicing</i>	rs138326449	-
11:123612256	C/A	ENSG00000166261	ZNF202	Variante en región donante de <i>splicing</i>	-	-
11:34937486	G/C	ENSG00000110435	PDHX	Variante en región donante de <i>splicing</i>	rs3763930	0,12

11:46748178	T/G	ENSG00000180210	F2	Variante en región donante de <i>splicing</i>	rs111566191	-
12:132464340	T/G	ENSG00000183495	EP400	Variante en región donante de <i>splicing</i>	-	-
12:80855835	C/T	ENSG00000139304	PTPRQ	Variante en región donante de <i>splicing</i>	rs11524045	0,2
12:120884459	G/A	ENSG00000111780	AL021546,6	Variante en región aceptora de <i>splicing</i>	-	-
12:120884459	G/A	ENSG00000257218	GATC	Variante en región aceptora de <i>splicing</i>	-	-
12:124268493	G/A	ENSG00000197653	DNAH10	Variante en región aceptora de <i>splicing</i>	rs200903075	-
12:40931655	G/A	ENSG00000205592	MUC19	Variante en región aceptora de <i>splicing</i>	rs76684015	0,15
12:57108461	T/C	ENSG00000196531	NACA	Variante en región aceptora de <i>splicing</i>	rs148340957	-
13:100925449	G/C	ENSG00000175198	PCCA	Variante en región aceptora de <i>splicing</i>	rs367615795	-
14:69928413	G/A	ENSG0000029364	SLC39A9	Variante en región aceptora de <i>splicing</i>	rs111495442	0,02
14:37149866	C/A	ENSG00000183032	SLC25A21	Variante en región donante de <i>splicing</i>	rs17104991	0
15:65496921	C/A	ENSG00000138615	CILP	Variante en región aceptora de <i>splicing</i>	-	-
15:43635274	G/T	ENSG00000168803	ADAL	Variante en región donante de <i>splicing</i>	-	-
15:48444430	C/A	ENSG00000104177	MYEF2	Variante en región donante de <i>splicing</i>	rs117103804	-
15:50593417	C/A	ENSG00000104064	GABPB1	Variante en región donante de <i>splicing</i>	rs80042819	0,004
15:75641682	T/C	ENSG00000140398	NEIL1	Variante en región donante de <i>splicing</i>	rs5745908,CS088022	0,01
16:55853446	C/A	ENSG00000198848	CES1	Variante en región donante de <i>splicing</i>	rs4513095	0,01
16:56774787	C/T	ENSG00000102900	NUP93	Variante en región donante de <i>splicing</i>	rs11861901	-
16:75566985	A/T	ENSG00000135702	CHST5	Variante en región donante de <i>splicing</i>	-	-
17:1386153	A/C	ENSG00000197879	MYO1C	Variante en región donante de <i>splicing</i>	rs115519639	-
17:39596412	C/T	ENSG00000171360	KRT38	Variante en región donante de <i>splicing</i>	-	-
17:28887134	G/A	ENSG00000266733	TBC1D29	Variante en región aceptora de <i>splicing</i>	rs111780165	0,07
17:40985792	G/C	ENSG00000131467	PSME3	Variante en región aceptora de <i>splicing</i>	rs115525472	0,0013
17:45447802	G/A	ENSG00000178852	EFCAB13	Variante en región aceptora de <i>splicing</i>	rs76299620	0,03
18:54388902	T/G	ENSG00000091157	WDR7	Variante en región donante de <i>splicing</i>	rs2080681	0,12
18:63525189	G/A	ENSG00000081138	CDH7	Variante en región donante de <i>splicing</i>	-	-
19:11688889	A/T	ENSG00000102575	ACP5	Variante en región donante de <i>splicing</i>	rs76639504	0,12
19:42192030	C/A	ENSG00000007306	CEACAM7	Variante en región donante de <i>splicing</i>	rs141024482	-
19:4722206	C/T	ENSG00000142002	DPP9	Variante en región donante de <i>splicing</i>	rs138088554	0,01
19:51014372	A/C	ENSG00000161677	JOSD2	Variante en región donante de <i>splicing</i>	-	-

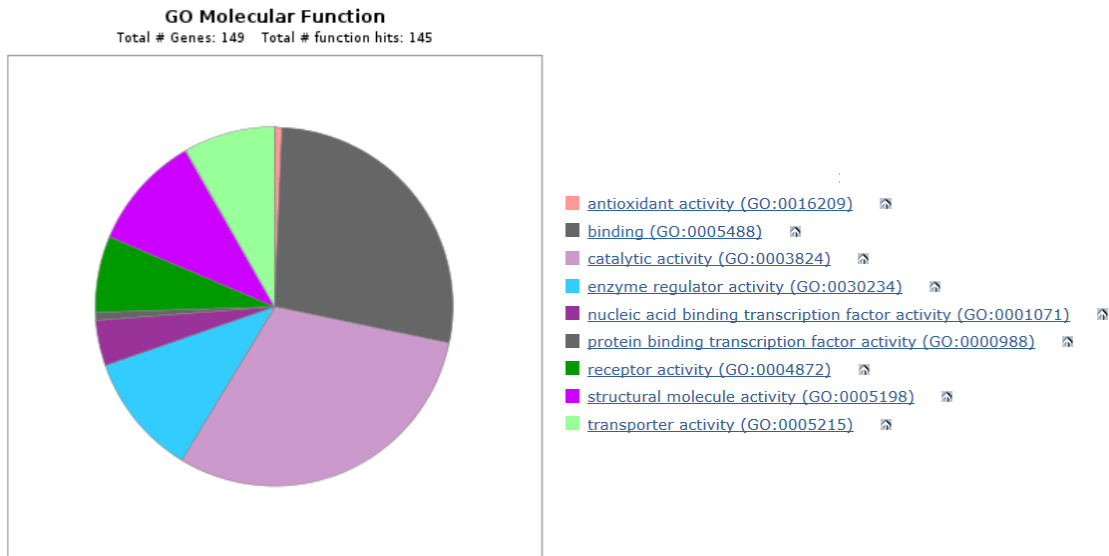
19:40554705	T/C	ENSG00000128000	ZNF780B	Variante en región aceptora de <i>splicing</i>	rs112293642	0,01
19:49336010	C/G	ENSG0000087076	HSD17B14	Variante en región aceptora de <i>splicing</i>	rs184605823	0,01
19:55098667	G/A	ENSG00000239988	LILRA2	Variante en región aceptora de <i>splicing</i>	rs2241524	0,01
20:31647696	G/T	ENSG00000186190	BPIFB3	Variante en región aceptora de <i>splicing</i>	rs11697151	0,02
20:44433784	G/T	ENSG00000101457	DNTTIP1	Variante en región aceptora de <i>splicing</i>	-	-
22:21965145	G/T	ENSG00000185651	UBE2L3	Variante en región aceptora de <i>splicing</i>	rs112359876	-
20:18119089	G/T	ENSG00000149474	CSRP2BP	Variante en región donante de <i>splicing</i>	rs116884076	0,06
20:258970	C/T	ENSG00000196476	C20orf96	Variante en región donante de <i>splicing</i>	rs149841409	0,004
20:31486430	G/A	ENSG00000215529	EFCAB8	Variante en región donante de <i>splicing</i>	rs375401148	-
21:28216059	A/G	ENSG00000154734	ADAMTS1	Variante en región donante de <i>splicing</i>	rs400852	0,2
22:36597744	A/C	ENSG00000100336	APOL4	Variante en región donante de <i>splicing</i>	rs9610445	0,14
22:51144595	T/G	ENSG00000251322	SHANK3	Variante en región donante de <i>splicing</i>	rs150909992	-
X:152939728	C/T	ENSG00000130822	PNCK	Variante en región donante de <i>splicing</i>	rs5987128	0,12
X:48434994	T/G	ENSG00000102317	RBM3	Variante en región donante de <i>splicing</i>	COSM69284	-
X:54838087	G/T	ENSG00000102316	MAGED2	Variante en región donante de <i>splicing</i>	-	-
X:6452998	C/T	ENSG00000169059	VCX3A	Variante en región donante de <i>splicing</i>	-	-
X:8139147	T/C	ENSG00000177504	VCX2	Variante en región donante de <i>splicing</i>	rs4066728	-
Variantes genéticas de <i>splicing</i> y <i>missense</i>						
1:100472710	C/A	ENSG00000117620	SLC35A3	Variante <i>splicing</i> y Variante <i>missense</i>	-	-
1:11807600	A/G	ENSG00000177674	AGTRAP	Variante <i>splicing</i> y Variante <i>missense</i>	rs144060195	-
1:24105970	C/G	ENSG00000057757	PITHD1	Variante <i>splicing</i> y Variante <i>missense</i>	-	-
1:28806965	G/C	ENSG00000204138	PHACTR4	Variante <i>splicing</i> y Variante <i>missense</i>	rs72661785	0,02
1:84876680	T/C	ENSG00000137976	DNASE2B	Variante <i>splicing</i> y Variante <i>missense</i>	rs149919241	0,01
1:92733663	A/G	ENSG00000174842	GLMN	Variante <i>splicing</i> y Variante <i>missense</i>	rs35258161	0
1:94467548	C/G	ENSG00000198691	ABCA4	Variante <i>splicing</i> y Variante <i>missense</i>	rs41292677,RISN_ABC R:c,6148G>C, CM970020	0,0026
2:233243586	C/T	ENSG00000163283	ALPP	Variante <i>splicing</i> y Variante <i>missense</i>	rs1130335	0,08
2:29497967	G/A	ENSG00000171094	ALK	Variante <i>splicing</i> y Variante <i>missense</i>	rs35228363	0,02
2:3482702	G/C	ENSG00000171853	TRAPPC12	Variante <i>splicing</i> y Variante <i>missense</i>	-	-
2:36737128	G/A	ENSG00000150938	CRIM1	Variante <i>splicing</i> y Variante <i>missense</i>	rs12997487	0,01

3:108780836	T/A	ENSG000000114487	MORC1	Variante splicing y Variante missense	rs17225637	0,004
3:42732351	G/T	ENSG000000157119	KLHL40	Variante splicing y Variante missense	-	-
3:49055300	A/C	ENSG000000178149	DALRD3	Variante splicing y Variante missense	-	-
4:100512392	G/A	ENSG000000138823	MTTP	Variante splicing y Variante missense	rs61750974	0,02
4:184622961	G/T	ENSG000000168538	TRAPPC11	Variante splicing y Variante missense	-	-
6:24780892	A/C	ENSG000000112312	GMNN	Variante splicing y Variante missense	rs1923185	0,04
6:36979483	T/G	ENSG000000146192	FGD2	Variante splicing y Variante missense	COSM304089	-
6:43100425	A/T	ENSG000000112655	PTK7	Variante splicing y Variante missense	rs34021075	0,02
7:105656471	G/A	ENSG000000128536	CDHR3	Variante splicing y Variante missense	rs143495908	0,01
7:151079066	C/A	ENSG000000187260	WDR86	Variante splicing y Variante missense	-	-
7:44098033	C/T	ENSG000000136279	DBNL	Variante splicing y Variante missense	rs147233974	0,0026
7:925694	T/G	ENSG000000239857	GET4	Variante splicing y Variante missense	rs200535612	-
8:120580415	G/A	ENSG000000136960	ENPP2	Variante splicing y Variante missense	rs149152218	-
8:22449181	C/T	ENSG000000248235	AC037459,4	Variante splicing y Variante missense	rs73229817	0,05
8:24349417	T/C	ENSG000000069206	ADAM7	Variante splicing y Variante missense	rs3736281	0,01
9:139934887	G/A	ENSG000000107281	NPDC1	Variante splicing y Variante missense	rs140988877	-
10:33113448	A/G	ENSG000000150076	C10ORF68	Variante splicing y Variante missense	rs35739764	0,01
10:81465798	C/T	ENSG000000188199	NUTM2B	Variante splicing y Variante missense	-	-
11:117054496	G/A	ENSG000000149577	SIDT2	Variante splicing y Variante missense	rs149441737, COSM1225698	0,0026
11:92569867	C/T	ENSG000000165323	FAT3	Variante splicing y Variante missense	rs200404766	-
11:93170710	C/T	ENSG000000165325	CCDC67	Variante splicing y Variante missense	rs190272170	0
12:52312771	A/T	ENSG000000139567	ACVRL1	Variante splicing y Variante missense	rs141653630	-
12:7985318	C/A	ENSG000000173262	SLC2A14	Variante splicing y Variante missense	rs10846018	0,04
12:88588615	G/A	ENSG000000139324	TMTC3	Variante splicing y Variante missense	-	-
13:52332329	C/T	ENSG000000139668	WDFY2	Variante splicing y Variante missense	rs35409673, COSM432485	0,01
14:64669681	G/T	ENSG000000054654	SYNE2	Variante splicing y Variante missense	-	-
14:92076825	G/C	ENSG000000133962	CATSPERB	Variante splicing y Variante missense	rs149037988	-
14:94731284	C/T	ENSG000000119698	PPP4R4	Variante splicing y Variante missense	rs140391785	-

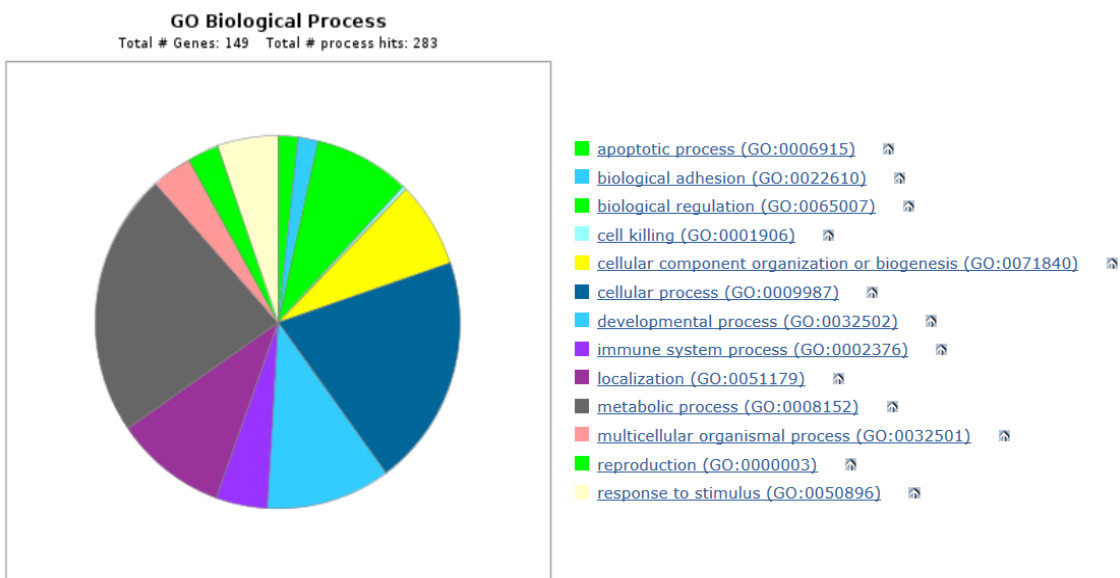
15:43020850	T/C	ENSG000000140326	CDAN1	Variante splicing y Variante missense	rs371548844	-
17:10356960	C/G	ENSG00000264424	MYH4	Variante splicing y Variante missense	-	-
17:37815304	G/C	ENSG00000131748	STARD3	Variante splicing y Variante missense	rs11556624	0,01
17:7674784	C/T	ENSG00000183914	DNAH2	Variante splicing y Variante missense	rs73232344	0,01
19:52505068	G/A	ENSG00000197619	ZNF615	Variante splicing y Variante missense	-	-
19:5694912	G/A	ENSG00000196365	LONP1	Variante splicing y Variante missense	-	-
21:44317156	A/C	ENSG00000160194	NDUFB3	Variante splicing y Variante missense	rs141922962	0,01
22:37273671	C/T	ENSG00000100365	NCF4	Variante splicing y Variante missense	rs200598824, TMP_ESP_22_3727367 1_37273671	-
22:38165269	G/A	ENSG00000100106	TRIOBP	Variante splicing y Variante missense	rs138139146	0,01
X:55103110	G/T	ENSG00000238269	PAGE2B	Variante splicing y Variante missense	-	-

A continuación, se estudiaron que funciones moleculares y metabólicas tenían los genes en los cuales encontramos las variantes de *splicing* en el grupo control. Los siguientes gráficos muestran los resultados obtenidos tras analizar los datos en el programa informático PANTHER- *Gene List Analysis*.

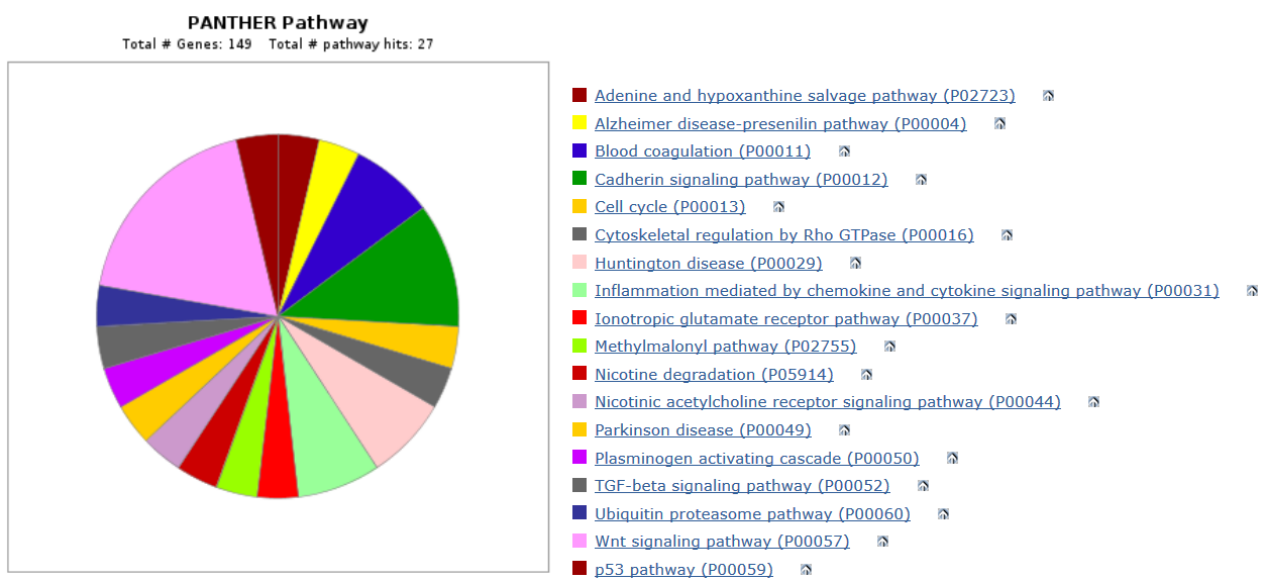
- Función Molecular:



- Proceso biológico:



- Rutas metabólicas implicadas:



6.1.2. Variantes genéticas de *splicing* en el grupo casos

Del mismo modo, se analizaron las variantes genéticas de *splicing* presentes únicamente en el grupo de diabéticos (Tabla 13).

Tabla 13. Variantes genéticas de *splicing* en el grupo de casos (n=200).

Posición Cr.	Cambio	Ensembl ID	Gen	Tipo de Variación	SNP	MAF europeo
1:212985592	G/A	ENSG00000203705	TATDN3	Variante en región aceptora de <i>splicing</i>	rs78095001	0,03
1:161565381	A/G	ENSG00000244682	FCGR2C	Variante en región donante de <i>splicing</i>	rs76277413	0,12
1:20650026	G/A	ENSG00000158816	VWA5B1	Variante en región donante de <i>splicing</i>	rs199957660	-
1:247030545	C/A	ENSG00000153207	AHCTF1	Variante en región donante de <i>splicing</i>	rs141409693	-
1:32154650	C/A	ENSG00000084636	COL16A1	Variante en región donante de <i>splicing</i>	rs72887331	0
1:43273064	A/T	ENSG00000177868	CCDC23	Variante en región donante de <i>splicing</i>	rs12402142	0,01
2:223785788	T/C	ENSG00000123983	ACSL3	Variante en región donante de <i>splicing</i>	rs139220113	0,01
2:43968183	G/C	ENSG00000152527	PLEKHH2	Variante en región donante de <i>splicing</i>	-	-
2:64211155	T/C	ENSG00000143952	VPS54	Variante en región aceptora de <i>splicing</i>	-	-
3:172674615	C/A	ENSG00000144962	SPATA16	Variante en región aceptora de <i>splicing</i>	-	-
3:183752964	A/C	ENSG00000186090	HTR3D	Variante en región aceptora de <i>splicing</i>	rs77099580	0,05
3:125856655	C/T	ENSG00000144908	ALDH1L1	Variante en región donante de <i>splicing</i>	rs151039331	-
3:186947546	A/G	ENSG00000127241	MASP1	Variante en región donante de <i>splicing</i>	-	-
3:52423604	T/G	ENSG00000114841	DNAH1	Variante en región donante de <i>splicing</i>	rs183004153	-
3:75834065	C/T	ENSG00000227124	ZNF717	Variante en región donante de <i>splicing</i>	rs58564752	0,03
4:142142171	G/A	ENSG00000109445	ZNF330	Variante en región donante de <i>splicing</i>	rs112082999	-
4:5975385	C/T	ENSG00000181215	C4orf50	Variante en región donante de <i>splicing</i>	rs138928790	0,02
4:160262715	G/T	ENSG00000109756	RAPGEF2	Variante en región aceptora de <i>splicing</i>	-	-
4:69687987	C/A	ENSG00000109181	UGT2B10	Variante en región aceptora de <i>splicing</i>	rs2942857	-
4:76841130	T/C	ENSG00000138744	NAAA	Variante en región aceptora de <i>splicing</i>	-	-
5:82648943	G/A	ENSG00000152422	XRCC4	Variante en región aceptora de <i>splicing</i>	rs1805377, CS075240	0,15
5:134077214	G/T	ENSG00000164615	CAMLG	Variante en región donante de <i>splicing</i>	-	-
5:150726857	C/T	ENSG00000186335	SLC36A2	Variante en región donante de <i>splicing</i>	rs371203963, CS086103	-
5:176314201	C/A	ENSG00000160883	HK3	Variante en región donante de <i>splicing</i>	-	-
5:35013058	C/T	ENSG00000113492	AGX72	Variante en región donante de <i>splicing</i>	rs114286107	0,02
6:117866762	T/G	ENSG00000164465	DCBLD1	Variante en región donante de <i>splicing</i>	-	-
6:170036466	C/A	ENSG00000184465	WDR27	Variante en región donante de <i>splicing</i>	rs111656694	0,004

6:26370833	T/G	ENSG00000186470	BTN3A2	Variante en región donante de <i>splicing</i>	rs58367598	0,04
6:27278875	C/T	ENSG00000158553	POM121L2	Variante en región donante de <i>splicing</i>	rs377639347	-
6:32548522	C/G	ENSG00000196126	HLA-DRB1	Variante en región donante de <i>splicing</i>	rs35121789	-
6:43267805	G/A	ENSG00000137204	SLC22A7	Variante en región donante de <i>splicing</i>	-	-
6:28195466	A/T	ENSG00000137185	ZSCAN9	Variante en región aceptora de <i>splicing</i>	-	-
7:108158746	C/A	ENSG00000135241	PNPLA8	Variante en región aceptora de <i>splicing</i>	-	-
7:105668925	T/C	ENSG00000128536	CDHR3	Variante en región donante de <i>splicing</i>	rs17152481	0,01
7:126890902	C/T	ENSG00000179603	GRM8	Variante en región donante de <i>splicing</i>	rs17866749	0,03
7:151791170	T/G	ENSG00000178234	GALNT11	Variante en región donante de <i>splicing</i>	rs17173334	0,01
8:12285023	C/G	ENSG00000145002	FAM86B2	Variante en región donante de <i>splicing</i>	-	-
8:92213022	T/A	ENSG00000214954	LRRRC69	Variante en región donante de <i>splicing</i>	rs72666050	0,02
9:114905750	C/T	ENSG00000106868	SUSD1	Variante en región donante de <i>splicing</i>	-	-
9:140348896	C/T	ENSG00000165802	NSMF	Variante en región aceptora de <i>splicing</i>	-	-
10:12148244	G/A	ENSG00000181192	DHTKD1	Variante en región aceptora de <i>splicing</i>	COSM367023	-
10:104154532	G/T	ENSG00000077150	NFKB2	Variante en región donante de <i>splicing</i>	-	-
10:50149895	T/G	ENSG00000128815	WDFY4	Variante en región donante de <i>splicing</i>	-	-
10:99218978	A/G	ENSG00000155229	MMS19	Variante en región donante de <i>splicing</i>	rs201440186	-
11:102668717	A/G	ENSG00000196611	MMP1	Variante en región donante de <i>splicing</i>	rs139018071	0,0026
11:377329	T/G	ENSG00000182272	B4GALNT4	Variante en región donante de <i>splicing</i>	rs373329286	-
11:58611906	C/T	ENSG00000156689	GLYATL2	Variante en región donante de <i>splicing</i>	rs117831599	0,01
11:102649482	T/C	ENSG00000166670	MMP10	Variante en región aceptora de <i>splicing</i>	rs17860955	0,02
11:65629060	C/T	ENSG00000172757	CFL1	Variante en región aceptora de <i>splicing</i>	rs575085	0,01
12:11420454	T/C	ENSG00000197870	PRB3	Variante en región aceptora de <i>splicing</i>	rs11054202	-
12:123344759	T/G	ENSG00000130787	HIP1R	Variante en región donante de <i>splicing</i>	-	-
12:55248425	G/T	ENSG00000172551	MUCL1	Variante en región donante de <i>splicing</i>	rs142903254	-
12:6807210	C/A	ENSG00000139200	PIANP	Variante en región donante de <i>splicing</i>	-	-
14:23623467	C/T	ENSG00000092068	SLC7A8	Variante en región donante de <i>splicing</i>	-	-
14:73706541	G/A	ENSG00000100767	PAPLN	Variante en región donante de <i>splicing</i>	rs202232594	-
15:41797504	C/T	ENSG00000062524	LTK	Variante en región donante de <i>splicing</i>	rs144415087	-
15:29368296	T/G	ENSG00000034053	APBA2	Variante en región donante de <i>splicing</i>	-	-
15:34825386	C/G	ENSG00000215252	GOLGA8B	Variante en región donante de <i>splicing</i>	-	-

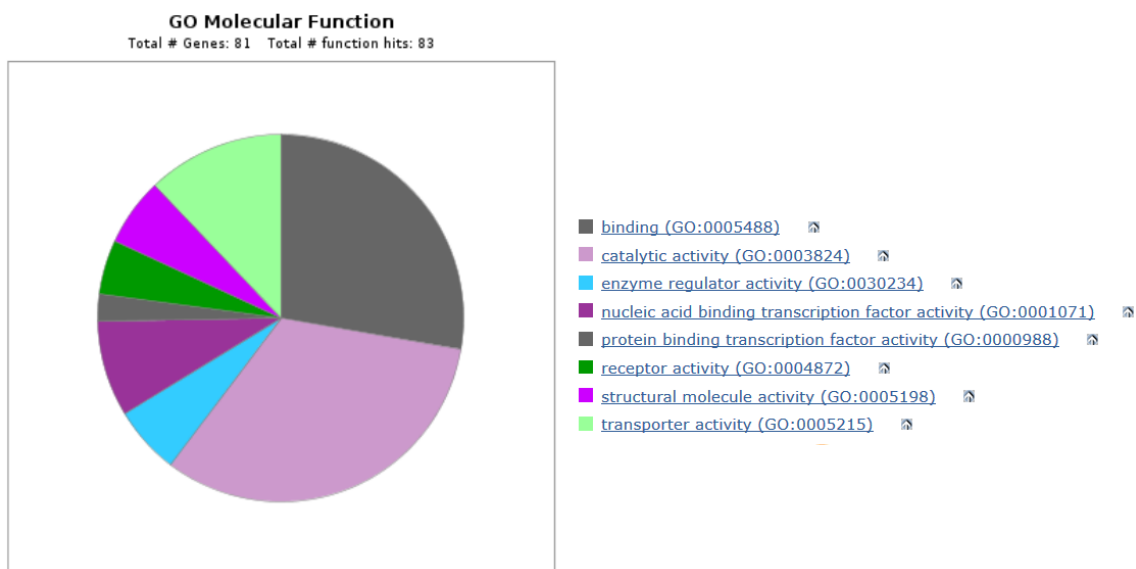
16:57984284	C/G	ENSG00000070729	CNGB1	Variante en región donante de <i>splicing</i>	-	-
16:707223	T/G	ENSG00000161996	WDR90	Variante en región donante de <i>splicing</i>	rs78538490	-
16:19584445	G/A	ENSG00000103544	C16orf62	Variante en región aceptora de <i>splicing</i>	rs139739334	-
16:84495318	A/C	ENSG0000064270	ATP2C2	Variante en región aceptora de <i>splicing</i>	rs4782970	0,01
17:5487821	C/G	ENSG0000091592	NLRP1	Variante en región aceptora de <i>splicing</i>	rs117569222	0,04
17:7290695	A/G	ENSG00000174292	TNK1	Variante en región aceptora de <i>splicing</i>	rs7220814	0,07
17:76200907	A/G	ENSG00000183077	AFMID	Variante en región aceptora de <i>splicing</i>	rs138105144	-
17:18863967	T/C	ENSG00000154025	SLC5A10	Variante en región donante de <i>splicing</i>	rs142768495	-
17:21199083	T/A	ENSG0000034152	MAP2K3	Variante en región donante de <i>splicing</i>	-	-
17:42786208	G/T	ENSG00000161692	DBF4B	Variante en región donante de <i>splicing</i>	rs34198215	0,01
17:80170660	C/G	ENSG00000176155	CCDC57	Variante en región donante de <i>splicing</i>	-	-
18:8248334	A/G	ENSG00000173482	PTPRM	Variante en región aceptora de <i>splicing</i>	rs192187739	0,05
18:8248335	G/T	ENSG00000173482	PTPRM	Variante en región aceptora de <i>splicing</i>	rs7236743	0,05
19:1041352	A/G	ENSG0000064687	ABCA7	Variante en región aceptora de <i>splicing</i>	rs3752229	0,07
19:1475258	C/G	ENSG00000119559	C19orf25	Variante en región aceptora de <i>splicing</i>	rs201353187	-
19:22846601	G/T	ENSG00000229676	ZNF492	Variante en región aceptora de <i>splicing</i>	rs71355985	0,04
19:2876090	G/A	ENSG00000172000	ZNF556	Variante en región aceptora de <i>splicing</i>	rs144545156, COSM994312	0,004
19:51768620	A/G	ENSG00000179213	SIGLECL1	Variante en región aceptora de <i>splicing</i>	rs140988748	-
19:58197847	A/C	ENSG00000204519	ZNF551	Variante en región aceptora de <i>splicing</i>	-	-
19:19144471	G/A	ENSG00000105676	ARMC6	Variante en región donante de <i>splicing</i>	-	-
21:10907041	C/A	ENSG00000166157	TPTTE	Variante en región aceptora de <i>splicing</i>	-	-
21:47635708	C/A	ENSG00000160285	LSS	Variante en región aceptora de <i>splicing</i>	-	-
21:34889295	C/T	ENSG00000159131	GART	Variante en región donante de <i>splicing</i>	COSM1616024	-
22:25124143	C/T	ENSG00000184571	PIWIL3	Variante en región donante de <i>splicing</i>	rs73152579	0,01
22:38042938	T/G	ENSG00000100092	SH3BP1	Variante en región donante de <i>splicing</i>	rs202091066	-
Variantes genéticas de <i>splicing</i> y missense						
1:113159464	T/A	ENSG00000007341	ST7L	Variante <i>splicing</i> y Variante missense	rs114199731	0,01
1:114372214	C/G	ENSG00000134242	PTPN22	Variante <i>splicing</i> y Variante missense	rs56048322	0,01
1:1147322	C/T	ENSG00000186827	TNFRSF4	Variante <i>splicing</i> y Variante missense	-	-
1:155002562	G/T	ENSG00000163354	DCST2	Variante <i>splicing</i> y Variante missense	-	-
1:222892605	G/A	ENSG00000162819	BROX	Variante <i>splicing</i> y Variante missense	-	-

1:228562288	G/A	ENSG00000154358	OBSCN	Variante splicing y Variante missense	rs377038292	-
1:230391090	G/A	ENSG00000143641	GALNT2	Variante splicing y Variante missense	-	-
1:28362189	G/T	ENSG00000158161	EYA3	Variante splicing y Variante missense	-	-
1:52759347	G/A	ENSG00000157077	ZFYVE9	Variante splicing y Variante missense	rs375050885	-
1:60503779	C/A	ENSG00000162598	C1orf87	Variante splicing y Variante missense	-	-
2:136396311	T/G	ENSG00000048991	R3HDM1	Variante splicing y Variante missense	-	-
2:136701048	A/G	ENSG00000115866	DARS	Variante splicing y Variante missense	-	-
2:160840584	C/A	ENSG00000153246	PLA2R1	Variante splicing y Variante missense	rs149133741	0,01
2:167085202	A/G	ENSG00000169432	SCN9A	Variante splicing y Variante missense	rs200266532	-
2:31605868	G/A	ENSG00000158125	XDH	Variante splicing y Variante missense	rs142388231	-
2:43625278	C/A	ENSG00000115970	THADA	Variante splicing y Variante missense	rs56269749	0,01
3:11667831	C/A	ENSG00000144560	VGLL4	Variante splicing y Variante missense	-	-
3:194387016	G/C	ENSG00000041802	LSG1	Variante splicing y Variante missense	rs146856473	0,0013
3:195591056	C/T	ENSG00000061938	TNK2	Variante splicing y Variante missense	rs13433937	0,04
3:197408001	C/T	ENSG00000145016	KIAA0226	Variante splicing y Variante missense	rs141198920	0
3:42734569	C/A	ENSG0000010282	HHATL	Variante splicing y Variante missense	-	-
3:52836533	G/A	ENSG00000162267	ITIH3	Variante splicing y Variante missense	rs74320783	0,01
5:112198206	T/G	ENSG00000153037	SRP19	Variante splicing y Variante missense	rs201250324,COSM1288199	-
5:160029708	G/T	ENSG00000118322	ATP10B	Variante splicing y Variante missense	-	-
5:81608669	A/G	ENSG00000205464	ATP6AP1L	Variante splicing y Variante missense	-	-
6:111587079	G/T	ENSG00000173214	KIAA1919	Variante splicing y Variante missense	rs200070624	-
6:31901558	G/A	ENSG00000166278	C2	Variante splicing y Variante missense	rs147186833	-
7:107822267	C/T	ENSG00000091129	NRCAM	Variante splicing y Variante missense	-	-
7:132481204	T/G	ENSG00000106554	CHCHD3	Variante splicing y Variante missense	-	-
7:21640629	T/C	ENSG00000105877	DNAH11	Variante splicing y Variante missense	-	-
8:17503646	T/C	ENSG00000129422	MTUS1	Variante splicing y Variante missense	-	-
8:30541607	G/A	ENSG00000104687	GSR	Variante splicing y Variante missense	-	-
8:59407065	C/T	ENSG00000167910	CYP7A1	Variante splicing y Variante missense	rs8192875	0
9:131151542	G/A	ENSG00000167118	URM1	Variante splicing y Variante missense	-	-
9:138670668	G/A	ENSG00000107147	KCNT1	Variante splicing y Variante missense	rs151272083	-
9:14842658	C/G	ENSG00000164946	FREM1	Variante splicing y Variante missense	rs41298151	0,0026

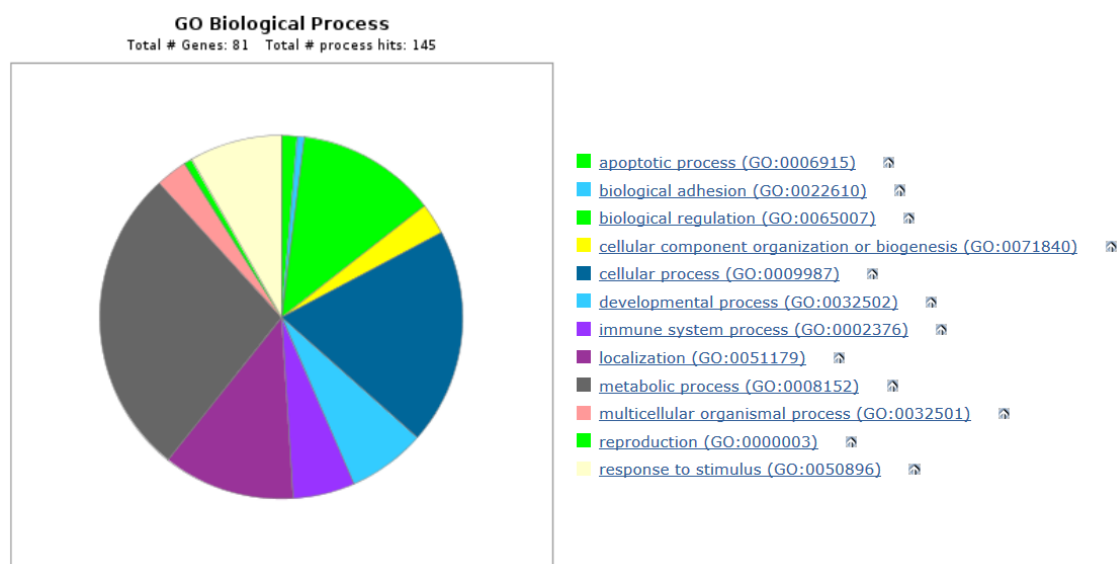
9:32481339	C/T	ENSG00000107201	DDX58	Variante splicing y Variante missense	rs61752945	0,01
10:116201483	G/A	ENSG00000099204	ABLIM1	Variante splicing y Variante missense	-	-
11:124967634	A/C	ENSG00000150433	TMEM218	Variante splicing y Variante missense	-	-
11:16824599	C/T	ENSG00000166689	PLEKHA7	Variante splicing y Variante missense	rs34556458	0,01
11:57089266	C/T	ENSG00000149115	TNKS1BP1	Variante splicing y Variante missense	rs148840217	-
11:65293512	G/A	ENSG00000142186	SCYL1	Variante splicing y Variante missense	rs78243061	0
11:85420374	G/A	ENSG00000137501	SYTL2	Variante splicing y Variante missense	rs199632665	-
12:7242740	C/G	ENSG00000159403	C1R	Variante splicing y Variante missense	rs139531404	0,01
12:9994550	G/C	ENSG00000150045	KLRF1	Variante splicing y Variante missense	-	-
13:31513915	G/A	ENSG00000175664	TEX26	Variante splicing y Variante missense	rs76662628	0
14:21570565	T/G	ENSG00000232070	TMEM253	Variante splicing y Variante missense	-	-
16:1308333	G/A	ENSG00000095917	TPSD1	Variante splicing y Variante missense	rs142998820	0,02
16:2260523	C/A	ENSG00000182685	BRICD5	Variante splicing y Variante missense	-	-
16:67236691	A/G	ENSG00000102890	ELMO3	Variante splicing y Variante missense	-	-
17:18862551	A/T	ENSG00000154025	SLC5A10	Variante splicing y Variante missense	rs148178887	-
17:33504629	G/A	ENSG00000141161	UNC45B	Variante splicing y Variante missense	rs139715157	0
17:6946390	G/A	ENSG00000174326	SLC16A11	Variante splicing y Variante missense	rs371508899	-
17:71468356	A/G	ENSG00000069188	SDK2	Variante splicing y Variante missense	rs117687984	0,01
19:12812267	T/G	ENSG00000105576	TNPO2	Variante splicing y Variante missense	-	-
19:37870148	G/T	ENSG00000189164	ZNF527	Variante splicing y Variante missense	-	-
19:49335923	G/A	ENSG00000087076	HSD17B14	Variante splicing y Variante missense	rs143254849	0
19:51361850	C/G	ENSG00000142515	KLK3	Variante splicing y Variante missense	rs61729813	0,01
19:52934741	A/C	ENSG00000198633	ZNF534	Variante splicing y Variante missense	rs118147154	0,01
19:53901281	C/G	ENSG00000196417	ZNF765	Variante splicing y Variante missense	rs141876832	0,02
20:36760744	G/A	ENSG00000198959	TGM2	Variante splicing y Variante missense	rs201666703	-
22:44324767	G/A	ENSG00000100344	PNPLA3	Variante splicing y Variante missense	rs141106484	-

Se analizaron también que posibles funciones metabólicas y en que procesos biológicos y rutas estaban implicados los genes en los cuales se habían identificado las variantes de *splicing* en los pacientes diabéticos.

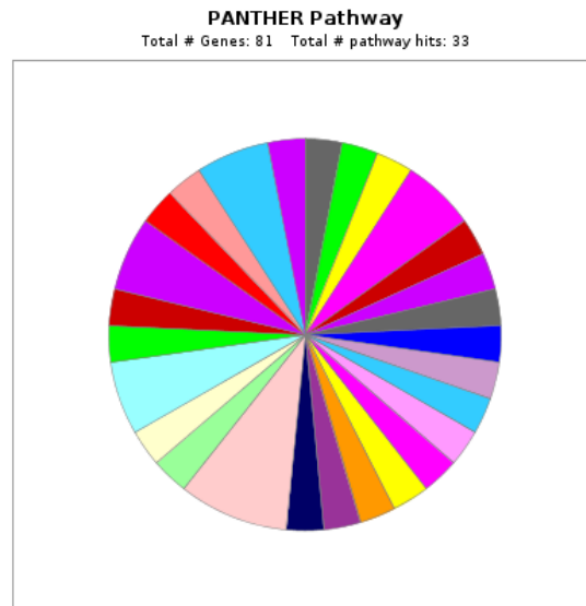
- Función Molecular:



- Proceso biológico:



- Rutas metabólicas implicadas:



- [5-Hydroxytryptamine degradation \(P04372\)](#) ⓘ
- [Alzheimer disease-amyloid secretase pathway \(P00003\)](#) ⓘ
- [Alzheimer disease-presenilin pathway \(P00004\)](#) ⓘ
- [Apoptosis signaling pathway \(P00006\)](#) ⓘ
- [B cell activation \(P00010\)](#) ⓘ
- [Cholesterol biosynthesis \(P00014\)](#) ⓘ
- [Cytoskeletal regulation by Rho GTPase \(P00016\)](#) ⓘ
- [De novo purine biosynthesis \(P02738\)](#) ⓘ
- [Dopamine receptor mediated signaling pathway \(P05912\)](#) ⓘ
- [EGF receptor signaling pathway \(P00018\)](#) ⓘ
- [FGF signaling pathway \(P00021\)](#) ⓘ
- [Folate biosynthesis \(P02742\)](#) ⓘ
- [Gonadotropin releasing hormone receptor pathway \(P06664\)](#) ⓘ
- [Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway \(P00026\)](#) ⓘ
- [Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway \(P00027\)](#) ⓘ
- [Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction \(P00028\)](#) ⓘ
- [Huntington disease \(P00029\)](#) ⓘ
- [Inflammation mediated by chemokine and cytokine signaling pathway \(P00031\)](#) ⓘ
- [Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade \(P00032\)](#) ⓘ
- [Integrin signalling pathway \(P00034\)](#) ⓘ
- [Metabotropic glutamate receptor group III pathway \(P00039\)](#) ⓘ
- [Oxidative stress response \(P00046\)](#) ⓘ
- [Plasminogen activating cascade \(P00050\)](#) ⓘ
- [Ras Pathway \(P04393\)](#) ⓘ
- [T cell activation \(P00053\)](#) ⓘ
- [Toll receptor signaling pathway \(P00054\)](#) ⓘ
- [p38 MAPK pathway \(P05918\)](#) ⓘ

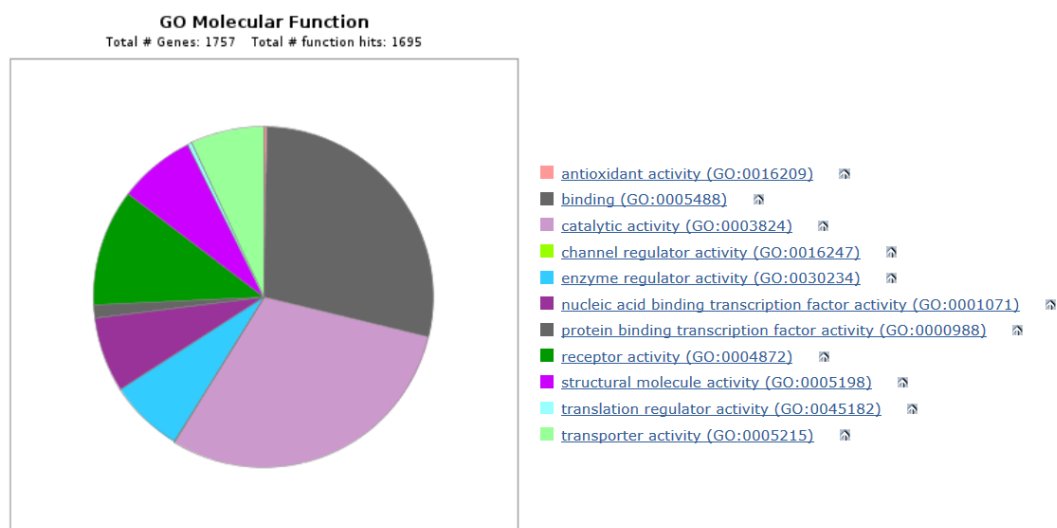
6.2. Estudio de variantes genéticas *missense* en el exoma

Las variantes de cambio de sentido o *missense* implican la aparición de un nuevo triplete que codifica para un aminoácido de distinto tipo. El número de SNPs encontrados de este tipo de variación puntual tras la secuenciación fue muy elevado; concretamente, 1817 SNPs en controles y 1614 SNPs en casos (datos no mostrados).

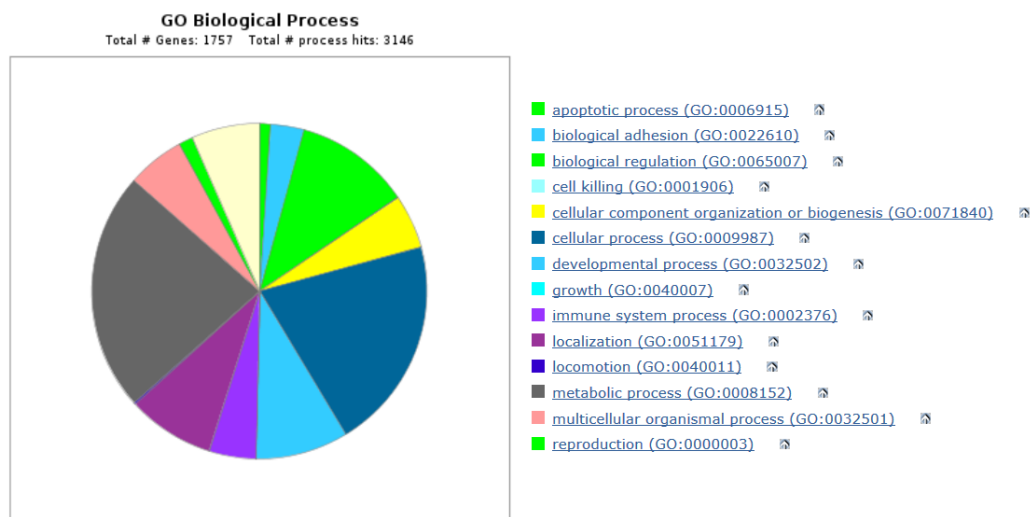
6.2.1. Variantes genéticas *missense* en el grupo control

También estudiamos que posibles funciones moleculares, procesos biológicos y rutas metabólicas estaban sobre-representados en los genes donde se encontraron las diferentes variantes genéticas *missense* en controles.

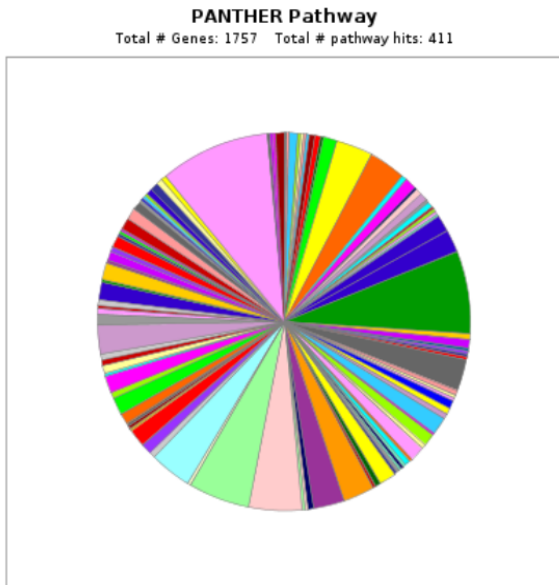
- Función Molecular:



- Proceso biológico:



• Rutas metabólicas implicadas:



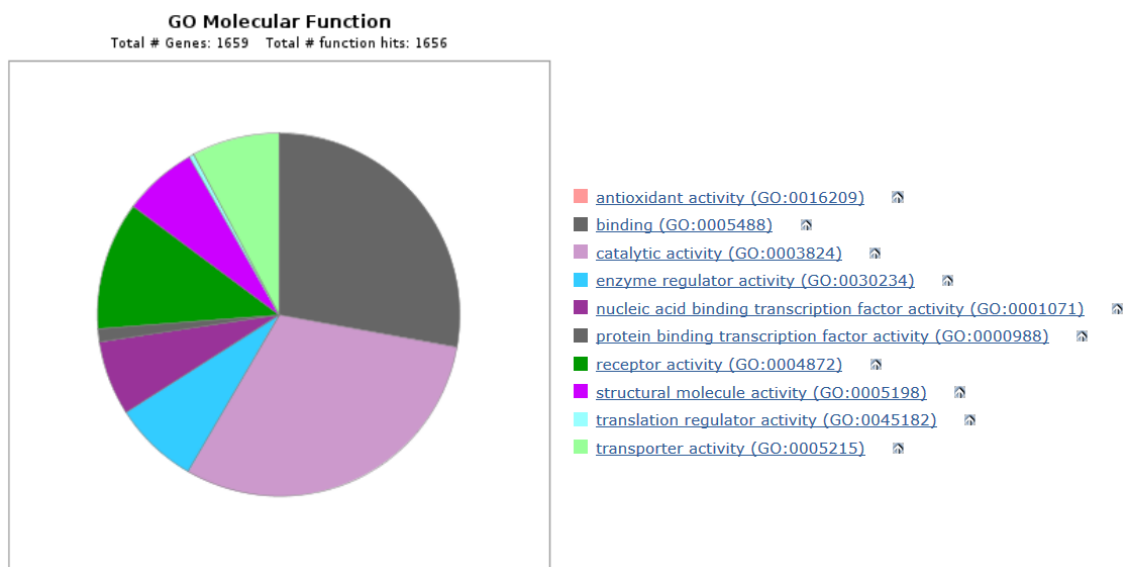
- 5-Hydroxytryptamine degradation (P04372) ⓘ
- SHT1 type receptor mediated signaling pathway (P04373) ⓘ
- SHT2 type receptor mediated signaling pathway (P04374) ⓘ
- SHT3 type receptor mediated signaling pathway (P04375) ⓘ
- SHT4 type receptor mediated signaling pathway (P04376) ⓘ
- ALP23B signaling pathway (P06209) ⓘ
- Activinbetasignaling pathway (P06210) ⓘ
- Adenine and hypoxanthine salvage pathway (P02723) ⓘ
- Adrenaline and noradrenaline biosynthesis (P00001) ⓘ
- Allantoin degradation (P02725) ⓘ
- Alzheimer disease-amyloid secretase pathway (P00003) ⓘ
- Alzheimer disease-presenilin pathway (P00004) ⓘ
- Angiogenesis (P00005) ⓘ
- Angiotensin II-stimulated signaling through G proteins and beta-arrestin (P05911) ⓘ
- Apoptosis signaling pathway (P00006) ⓘ
- Arginine biosynthesis (P02728) ⓘ
- Ascorbate degradation (P02729) ⓘ
- Axon guidance mediated by Slit/Robo (P00008) ⓘ
- Axon guidance mediated by netrin (P00009) ⓘ
- Axon guidance mediated by semaphorins (P00007) ⓘ
- B cell activation (P00010) ⓘ
- BMP signaling pathway-drosophila (P06211) ⓘ
- Beta1 adrenergic receptor signaling pathway (P04377) ⓘ
- Beta2 adrenergic receptor signaling pathway (P04378) ⓘ
- Blood coagulation (P00011) ⓘ
- CCKR signaling map (P06959) ⓘ
- Cadherin signaling pathway (P00012) ⓘ
- Cell cycle (P00013) ⓘ
- Cholesterol biosynthesis (P00014) ⓘ
- Circadian clock system (P00015) ⓘ
- Coenzyme A biosynthesis (P02736) ⓘ
- Corticotropin releasing factor receptor signaling pathway (P04380) ⓘ
- Cysteine biosynthesis (P02737) ⓘ
- Cytoskeletal regulation by Rho GTPase (P00016) ⓘ
- DNA replication (P00017) ⓘ
- DPP-SCW signaling pathway (P06212) ⓘ
- DPP signaling pathway (P06213) ⓘ
- De novo purine biosynthesis (P02738) ⓘ
- De novo pyrimidine ribonucleotides biosynthesis (P02740) ⓘ

- Dopamine receptor mediated signaling pathway (P05912) ⓘ
- EGF receptor signaling pathway (P00018) ⓘ
- Endogenous cannabinoid signaling (P05730) ⓘ
- Endothelin signaling pathway (P00019) ⓘ
- FAS signaling pathway (P00020) ⓘ
- FGF signaling pathway (P00021) ⓘ
- Flavin biosynthesis (P02741) ⓘ
- GABA-B receptor II signaling (P05731) ⓘ
- GRR signaling pathway (P06214) ⓘ
- Gamma-aminobutyric acid synthesis (P04384) ⓘ
- General transcription by RNA polymerase I (P00022) ⓘ
- Glutamine glutamate conversion (P02745) ⓘ
- Glycolysis (P00024) ⓘ
- Gonadotropin releasing hormone receptor pathway (P06664) ⓘ
- Hedgehog signaling pathway (P00025) ⓘ
- Heme biosynthesis (P02746) ⓘ
- Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026) ⓘ
- Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027) ⓘ
- Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028) ⓘ
- Histamine H1 receptor mediated signaling pathway (P04385) ⓘ
- Histamine synthesis (P04387) ⓘ
- Huntington disease (P00029) ⓘ
- Inflammation mediated by chemokine and cytokine signaling pathway (P00031) ⓘ
- Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032) ⓘ
- Integrin signalling pathway (P00034) ⓘ
- Interferon-gamma signaling pathway (P00035) ⓘ
- Interleukin signaling pathway (P00036) ⓘ
- Ionotropic glutamate receptor pathway (P00037) ⓘ
- Leucine biosynthesis (P02749) ⓘ
- MYO signaling pathway (P06215) ⓘ
- Mannose metabolism (P02752) ⓘ
- Metabotropic glutamate receptor group I pathway (P00041) ⓘ
- Metabotropic glutamate receptor group III pathway (P00039) ⓘ
- Methylmalonyl pathway (P02755) ⓘ
- Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042) ⓘ
- Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043) ⓘ
- N-acetylglucosamine metabolism (P02756) ⓘ
- Nicotine degradation (P05914) ⓘ
- Nicotine pharmacodynamics pathway (P06587) ⓘ
- Nicotinic acetylcholine receptor signaling pathway (P00044) ⓘ
- Notch signaling pathway (P00045) ⓘ
- O-antigen biosynthesis (P02757) ⓘ
- Oxidative stress response (P00046) ⓘ
- Oxytocin receptor mediated signaling pathway (P04391) ⓘ
- PDGF signaling pathway (P00047) ⓘ
- PI3 kinase pathway (P00048) ⓘ
- Parkinson disease (P00049) ⓘ
- Peptidoglycan biosynthesis (P02763) ⓘ
- Plasminogen activating cascade (P00050) ⓘ
- Pyruvate metabolism (P02772) ⓘ
- Ras Pathway (P04393) ⓘ
- SCW signaling pathway (P06216) ⓘ
- Salvage pyrimidine ribonucleotides (P02775) ⓘ
- Sulfate assimilation (P02778) ⓘ
- Synaptic vesicle trafficking (P05734) ⓘ
- T cell activation (P00053) ⓘ
- TGF-beta signaling pathway (P00052) ⓘ
- Threonine biosynthesis (P02781) ⓘ
- Thyrotropin-releasing hormone receptor signaling pathway (P04394) ⓘ
- Toll receptor signaling pathway (P00054) ⓘ
- Transcription regulation by bZIP transcription factor (P00055) ⓘ
- Tryptophan biosynthesis (P02783) ⓘ
- Ubiquitin proteasome pathway (P00060) ⓘ
- VEGF signaling pathway (P00056) ⓘ
- Vitamin D metabolism and pathway (P04396) ⓘ
- Wnt signaling pathway (P00057) ⓘ
- Xanthine and guanine salvage pathway (P02788) ⓘ
- p38 MAPK pathway (P05918) ⓘ
- p53 pathway feedback loops 2 (P04398) ⓘ
- p53 pathway (P00059) ⓘ

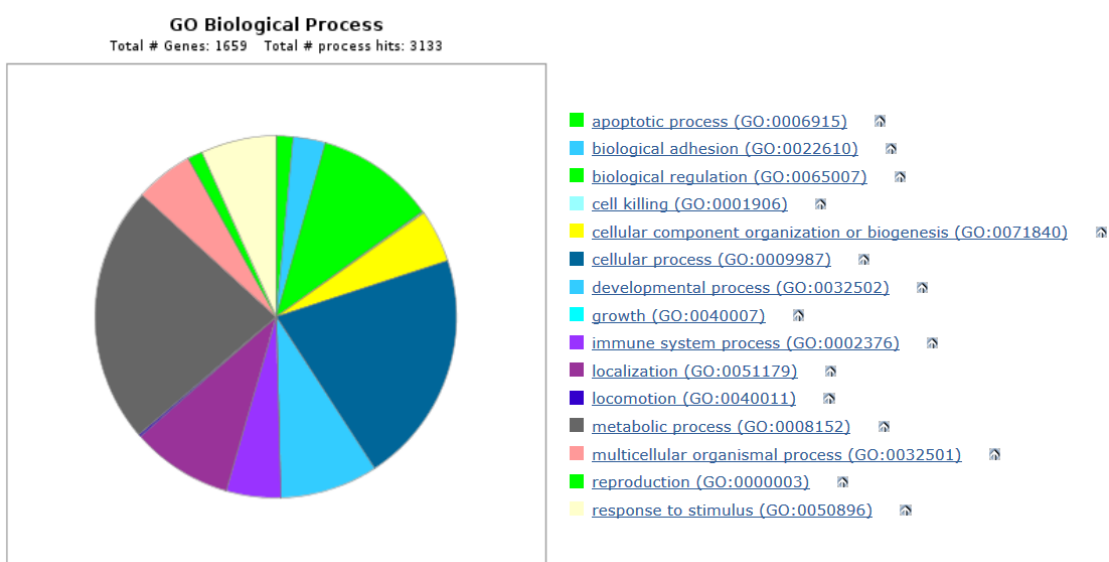
6.2.2. Variantes genéticas *missense* en el grupo de casos

Del mismo modo que en el grupo control, también estudiamos que funciones, procesos y rutas biológicas estaban enriquecidas en aquellos genes donde encontramos este tipo de variantes genéticas para el grupo de diabéticos.

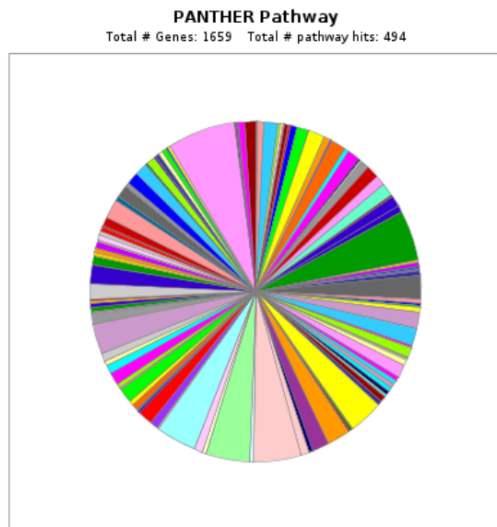
- Función molecular:



- Proceso biológico:



• Rutas metabólicas implicadas:



- 5-Hydroxytryptamine degradation (P04372) ⓘ
- 5HT1 type receptor mediated signaling pathway (P04373) ⓘ
- 5HT2 type receptor mediated signaling pathway (P04374) ⓘ
- 5HT3 type receptor mediated signaling pathway (P04375) ⓘ
- 5HT4 type receptor mediated signaling pathway (P04376) ⓘ
- ATP synthesis (P02721) ⓘ
- Adenine and hypoxanthine salvage pathway (P02723) ⓘ
- Adrenaline and noradrenaline biosynthesis (P00001) ⓘ
- Alpha adrenergic receptor signaling pathway (P00002) ⓘ
- Alzheimer disease-amyloid secretase pathway (P00003) ⓘ
- Alzheimer disease-presenilin pathway (P00004) ⓘ
- Aminobutyrate degradation (P02726) ⓘ
- Androgen/estrogene/progesterone biosynthesis (P02727) ⓘ
- Angiogenesis (P00005) ⓘ
- Angiotensin II-stimulated signaling through G proteins and beta-arrestin (P05911) ⓘ
- Apoptosis signaling pathway (P00006) ⓘ
- Arginine biosynthesis (P02728) ⓘ
- Ascorbate degradation (P02729) ⓘ
- Axon guidance mediated by netrin (P00009) ⓘ
- B cell activation (P00010) ⓘ
- Beta1 adrenergic receptor signaling pathway (P04377) ⓘ
- Beta2 adrenergic receptor signaling pathway (P04378) ⓘ
- Beta3 adrenergic receptor signaling pathway (P04379) ⓘ
- Blood coagulation (P00011) ⓘ
- CCKR signaling map (P06959) ⓘ
- Cadherin signaling pathway (P00012) ⓘ
- Cell cycle (P00013) ⓘ
- Cholesterol biosynthesis (P00014) ⓘ
- Circadian clock system (P00015) ⓘ
- Coenzyme A biosynthesis (P02736) ⓘ
- Corticotropin releasing factor receptor signaling pathway (P04380) ⓘ
- Cytoskeletal regulation by Rho GTPase (P00016) ⓘ
- DNA replication (P00017) ⓘ
- De novo purine biosynthesis (P02738) ⓘ
- De novo pyrimidine deoxyribonucleotide biosynthesis (P02739) ⓘ
- De novo pyrimidine ribonucleotides biosynthesis (P02740) ⓘ

- Dopamine receptor mediated signaling pathway (P05912) ⓘ
- EGF receptor signaling pathway (P00018) ⓘ
- Endogenous cannabinoid signaling (P05730) ⓘ
- Endothelin signaling pathway (P00019) ⓘ
- Enkephalin release (P05913) ⓘ
- FAS signaling pathway (P00020) ⓘ
- FGF signaling pathway (P00021) ⓘ
- Folate biosynthesis (P02742) ⓘ
- Formyltetrahydroformate biosynthesis (P02743) ⓘ
- Fructose galactose metabolism (P02744) ⓘ
- GABA-B receptor II signaling (P05731) ⓘ
- Gamma-aminobutyric acid synthesis (P04384) ⓘ
- General transcription by RNA polymerase I (P00022) ⓘ
- General transcription regulation (P00023) ⓘ
- Glycolysis (P00024) ⓘ
- Gonadotropin releasing hormone receptor pathway (P06664) ⓘ
- Hedgehog signaling pathway (P00025) ⓘ
- Heme biosynthesis (P02746) ⓘ
- Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026) ⓘ
- Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027) ⓘ
- Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028) ⓘ
- Histamine H1 receptor mediated signaling pathway (P04385) ⓘ
- Huntington disease (P00029) ⓘ
- Hypoxia response via HIF activation (P00030) ⓘ
- Inflammation mediated by chemokine and cytokine signaling pathway (P00031) ⓘ
- Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032) ⓘ
- Insulin/IGF pathway-protein kinase B signaling cascade (P00033) ⓘ
- Integrin signalling pathway (P00034) ⓘ
- Interferon-gamma signaling pathway (P00035) ⓘ
- Interleukin signaling pathway (P00036) ⓘ
- Ionotropic glutamate receptor pathway (P00037) ⓘ
- JAK/STAT signaling pathway (P00038) ⓘ
- Mannose metabolism (P02752) ⓘ
- Metabotropic glutamate receptor group I pathway (P00041) ⓘ
- Metabotropic glutamate receptor group II pathway (P00040) ⓘ
- Metabotropic glutamate receptor group III pathway (P00039) ⓘ
- Methionine biosynthesis (P02753) ⓘ
- Methylmalonyl pathway (P02755) ⓘ
- Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042) ⓘ
- Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043) ⓘ
- N-acetylglucosamine metabolism (P02756) ⓘ
- Nicotine pharmacodynamics pathway (P06587) ⓘ
- Nicotinic acetylcholine receptor signaling pathway (P00044) ⓘ
- Notch signaling pathway (P00045) ⓘ
- O-antigen biosynthesis (P02757) ⓘ
- Opioid prodynorphin pathway (P05916) ⓘ
- Opioid proenkephalin pathway (P05915) ⓘ
- Opioid proopiomelanocortin pathway (P05917) ⓘ
- Oxidative stress response (P00046) ⓘ
- Oxytocin receptor mediated signaling pathway (P04391) ⓘ
- PDGF signaling pathway (P00047) ⓘ
- PI3 kinase pathway (P00048) ⓘ
- Parkinson disease (P00049) ⓘ
- Pentose phosphate pathway (P02762) ⓘ
- Plasminogen activating cascade (P00050) ⓘ
- Proline biosynthesis (P02768) ⓘ
- Purine metabolism (P02769) ⓘ
- Pyrimidine Metabolism (P02771) ⓘ
- Ras Pathway (P04393) ⓘ
- S-adenosyl methionine biosynthesis (P02773) ⓘ
- Synaptic vesicle trafficking (P05734) ⓘ
- T cell activation (P00053) ⓘ
- TCA cycle (P00051) ⓘ
- TGF-beta signaling pathway (P00052) ⓘ
- Threonine biosynthesis (P02781) ⓘ
- Thyrotropin-releasing hormone receptor signaling pathway (P04394) ⓘ
- Toll receptor signaling pathway (P00054) ⓘ
- Toll pathway drosophila (P06217) ⓘ
- Transcription regulation by bZIP transcription factor (P00055) ⓘ
- Tryptophan biosynthesis (P02783) ⓘ
- Ubiquitin proteasome pathway (P00060) ⓘ
- VEGF signaling pathway (P00056) ⓘ
- Vasopressin synthesis (P04395) ⓘ
- Vitamin B6 metabolism (P02787) ⓘ
- Vitamin D metabolism and pathway (P04396) ⓘ
- Wnt signaling pathway (P00057) ⓘ
- Xanthine and guanine salvage pathway (P02788) ⓘ
- p38 MAPK pathway (P05918) ⓘ
- p53 pathway feedback loops 2 (P04398) ⓘ
- p53 pathway (P00059) ⓘ

6.3. Estudio de variantes genéticas en secuencias codificantes para microARNs en el exoma

Como se ha indicado anteriormente, el estudio de variantes genéticas en regiones codificantes para microARNs es de gran interés debido a la capacidad de estas moléculas de regular la expresión de otros genes. Concretamente, en el presente proyecto encontramos que este tipo de variantes en regiones codificantes para microARNs maduros eran minoritarias; sin embargo, debido a su posible efecto funcional, identificamos 23 SNPs en el grupo control y 11 SNPs en el grupo de diabéticos.

6.3.1. Variantes genéticas en secuencias codificantes para microARNs en el grupo control

A continuación, se señalan aquellas variantes presentes en zonas codificantes de microARNs identificadas en el grupo control (Tabla 14).

Tabla 14. Variantes genéticas en la secuencia codificante de un microRNA en el grupo control.

<i>Posición Cr.</i>	<i>Cambio</i>	<i>Ensembl ID</i>	<i>microRNA</i>	<i>SNP</i>	<i>MAF europeo</i>
3:194387926	C/T	ENSG00000221718	<i>AC046143.1</i>	rs143803572	0,01
3:195426305	T/C	ENSG00000207650	<i>MIR570</i>	rs9860655	0,11
3:41690767	G/A	ENSG00000221337	<i>AC099537.1</i>	rs141115358	0,01
4:49211441	C/G	ENSG00000266337	<i>AC118282.4</i>	-	-
5:53371399	C/T	ENSG00000265421	<i>MIR4459</i>	rs73112689	0,18
6:151361084	G/A	ENSG00000221469	<i>AL133260.1</i>	rs12215772	-
6:15150001	T/C	ENSG00000264687	<i>AL050335.1</i>	rs9296971	0,17
6:47430649	A/G,T	ENSG00000266330	<i>AL355353.1</i>	rs6920530	0
6:75429493	T/G	ENSG00000264884	<i>AL590684.1</i>	rs9343210	0,05
7:129625557	C/T	ENSG00000263557	<i>AC073320.1</i>	rs13228723	0,02
7:63387223	G/A	ENSG00000265017	<i>AC092634.2</i>	rs78348673	0,02
7:7107883	C/T	ENSG00000266409	<i>AC092104.5</i>	rs2965466	0,12
8:105767375	G/T	ENSG00000263732	<i>AC012564.1</i>	-	-
9:66861474	G/T	ENSG00000264898	<i>AL353626.2</i>	rs7468892	-
10:13747007	G/C	ENSG00000222235	<i>AL157392.1</i>	rs1409325	0,12
12:38715401	G/A	ENSG00000266333	<i>AC117372.1</i>	rs11183304	0
14:20685323	G/A	ENSG00000266336	<i>AL356019.1</i>	rs183826110	0,01
16:820249	G/A	ENSG00000207579	<i>MIR662</i>	rs9745376	0,004
17:63416750	C/A	ENSG00000265189	<i>AC006080.1</i>	rs9894176	0,06
17:77681036	G/C	ENSG00000266665	<i>MIR4739</i>	rs73410309	0,19
19:20473395	G/A	ENSG00000221181	<i>AC078899.5</i>	rs4809039	0,03
21:22770125	A/G	ENSG00000221210	<i>AP001137.1</i>	rs9636695	0,03
21:44371114	T/C	ENSG00000264580	<i>MIR5692B</i>	rs451887	0,13

6.3.2. Variantes genéticas en secuencias codificantes para microARNs en el grupo de casos

También se obtuvieron aquellas variantes presentes en zonas codificantes de microARNs para el grupo de casos (Tabla 15).

Tabla 15. Variantes genéticas en la secuencia codificante de un microRNA en el grupo de diabéticos.

<i>Posición Cr.</i>	<i>Cambio</i>	<i>Ensembl ID</i>	<i>microRNA</i>	<i>SNP</i>	<i>MAF europeo</i>
1:17185500	C/T	ENSG00000263811	<i>MIR3675</i>	rs202014433	-
2:176032376	G/A	ENSG00000215973	<i>MIR933</i>	rs79402775	0,02
5:56777677	G/A	ENSG00000264748	<i>AC025470.1</i>	rs16886992, COSM1437999	0,04
9:30774307	C/T	ENSG00000211510	<i>AL590726.1</i>	rs16916284	0
9:78338720	C/G	ENSG00000264310	<i>AL157883.1</i>	rs11144594	0,11
15:28947721	G/A	ENSG00000238340	<i>AC055876.2</i>	rs62009579	-
15:55665202	T/C	ENSG00000207964	<i>MIR628</i>	-	-
19:20473402	C/A	ENSG00000221181	<i>AC078899.5</i>	rs11666693	0,1
22:18246960	C/G	ENSG00000264757	<i>MIR3198-1</i>	rs56660112	0
X:1410673	A/T	ENSG00000264510	<i>BX649553.3</i>	-	-
X:1412530	G/A	ENSG00000263980	<i>BX649553.2</i>	-	-

7. ESTUDIO DE VARIANTES GENÉTICAS DE CODÓN DE STOP EN EL EXOMA

Tras el análisis bioinformático se obtuvieron las variantes de ganancia/pérdida de un codón de stop. Es decir, variantes que implican la adición (ganancia) o delección (pérdida) de un único par de nucleótidos o de varios pares de nucleótidos.

Por su relevancia funcional y para tener una aproximación del número de variantes reales, este tipo de variantes se verificaron en los dos grupos mediante secuenciación de Sanger. Esto permitió diferenciar que variantes genéticas eran válidas y reales, y cuales no eran ciertas: 102 fueron válidas en controles y 50 en casos. Así, del total de las 177 variantes de codón de stop identificadas inicialmente, el 85% fueron válidas.

Con el objetivo de encontrar parámetros diferenciales entre controles y diabéticos, también analizamos la posición genética de las variantes encontradas en ambos grupos. Algunas de las variantes genéticas identificadas se localizaron en la misma región cromosómica, pudiendo indicar lugares de mayor susceptibilidad genética y posible interés. Finalmente, se estudió también las funciones moleculares, procesos biológicos y rutas metabólicas de los genes donde identificamos y verificamos las variantes genéticas encontradas (PANTHER- *Gene List Analysis*).

7.1. Estudio de variantes genéticas de codón de stop en el grupo control

Para el grupo control, verificamos las 110 variantes de stop identificadas inicialmente y 102 resultaron ser válidas. Además, diferenciamos entre variantes de ganancia o pérdida de codón de stop, así como aquellas que también podían ser variantes de *splicing* (Tabla 16).

Tabla 16. Variantes genéticas identificadas y verificadas de codón de stop para el grupo de controles.

Posición Cr.	Cambio	Ensembl ID	Gen	Tipo de Variación	SNP	MAF europeo
1:145074975	G/A	ENSG00000178104	PDE4DIP	Variante ganancia codón stop	rs56128713	-
1:152058192	G/A	ENSG00000182898	TCHHL1	Variante ganancia codón stop	rs150014958	0
1:157805673	C/A	ENSG00000073754	CD5L	Variante ganancia codón stop	rs142870154, COSM897603	0,0013
1:248039789	T/C	ENSG00000162722	TRIM58	Variante pérdida codón stop	rs148369481	0,02
1:248722723	G/A	ENSG00000182783	OR2T29	Variante ganancia codón stop	rs2000037558	-
1:32280739	G/A	ENSG00000134668	SPOCD1	Variante ganancia codón stop	-	-
1:35223162	G/A	ENSG00000189280	GJB5	Variante ganancia codón stop	-	-
1:51584465	C/T	ENSG00000204006	C1orf185	Variante ganancia codón stop	rs184105748	0,0026
2:128522203	A/T	ENSG00000136709	WDR33	Variante ganancia codón stop	rs117753184, COSM148939	0
2:84934721	C/T	ENSG00000115423	DNAH6	Variante ganancia codón stop	COSM1023290	-
2:85576627	G/A	ENSG00000042445	RETSAT	Variante ganancia codón stop	rs143670128	0,0013
2:95539833	C/G	ENSG00000163060	TEKT4	Variante ganancia codón stop	rs111469930	0,01
3:108147680	G/A	ENSG00000144821	MYH15	Variante ganancia codón stop	rs61744539	0,004
3:108218327	G/A	ENSG00000144821	MYH15	Variante ganancia codón stop, Variante <i>splicing</i>	COSM1308363	-
3:126135332	G/A	ENSG00000163885	CCDC37	Variante ganancia codón stop	rs142615985	0,02
3:130159330	C/T	ENSG00000172752	COL6A5	Variante ganancia codón stop	rs2201717	0,01
3:140993422	T/G	ENSG00000155893	ACPL2	Variante ganancia codón stop	-	-
3:14180695	G/T	ENSG00000170876	TMEM43	Variante ganancia codón stop	-	-
3:157160693	T/A	ENSG00000163661	PTX3	Variante ganancia codón stop	-	-
3:159609989	G/A	ENSG00000250588	IQCJ- SCHIP1	Variante ganancia codón stop	rs74495285	0,03
3:161221491	C/T	ENSG00000182447	OTOL1	Variante ganancia codón stop	-	-
3:186461524	C/T	ENSG00000113889	KNG1	Variante ganancia codón stop	rs76438938	0,03
3:73111724	T/G	ENSG00000255423	EBLN2	Variante ganancia codón stop	rs2231925	0,01

4:106474097	G/T	ENSG000000236699	ARHGEF38	Variante ganancia codón stop	rs148016102, COSM1671188, COSM1671189	-
4:154702787	G/T	ENSG000000145423	SFRP2	Variante ganancia codón stop	COSM117303	-
4:15482360	C/T	ENSG00000048342	CC2D2A	Variante ganancia codón stop	rs1861050	0,04
4:155407595	C/A	ENSG000000197410	DCHS2	Variante pérdida codón stop	rs140554741, COSM1137151	0,03
4:15739428	A/T	ENSG000000109743	BST1	Variante pérdida codón stop	rs4320134	0,14
5:140049036	C/T	ENSG000000120314	WDR55	Variante ganancia codón stop	rs370437926	-
5:150859058	G/C	ENSG000000123643	SLC36A1	Variante pérdida codón stop	rs149620473	0,0026
5:180166545	G/A	ENSG000000174339	OR2Y1	Variante ganancia codón stop	rs372667309	-
5:95194593	G/T	ENSG000000236882	C5orf27	Variante ganancia codón stop	rs17085193	0,04
6:110736730	C/T	ENSG000000203797	DDO	Variante ganancia codón stop	rs141023778	0
6:139576544	G/A	ENSG000000164440	TXLNB	Variante ganancia codón stop	rs41289819	0,16
6:29407955	C/T	ENSG000000206474	OR10C1	Variante ganancia codón stop	rs17184009,CM035843	-
6:35423630	C/T	ENSG000000112039	FANCE	Variante ganancia codón stop	rs121434505, FANCE:c.355C>T, CM003675	-
6:41903798	C/A	ENSG000000112576	CCND3	Variante ganancia codón stop	rs33966734	0,01
6:42162372	G/A	ENSG000000112599	GUCA1B	Variante ganancia codón stop	rs200394928, COSM1209194	-
6:49459927	C/G	ENSG000000031691	CENPQ	Variante ganancia codón stop	rs144734447	0,0026
7:2749438	C/T	ENSG000000174945	AMZ1	Variante ganancia codón stop	-	-
7:44180639	G/A	ENSG000000106631	MYL7	Variante ganancia codón stop	-	-
7:87908735	C/T	ENSG000000127954	STEAP4	Variante ganancia codón stop	-	-
8:144732505	C/T	ENSG000000183309	ZNF623	Variante ganancia codón stop	rs200418048	-
8:18080115	C/T	ENSG000000171428	NAT1	Variante ganancia codón stop	rs5030839,CM981375	0,01
8:22053701	T/G	ENSG000000168487	BMP1	Variante ganancia codón stop	rs80055120	0,02
8:95384577	G/A	ENSG000000197275	RAD54B	Variante ganancia codón stop	-	-
9:100380062	G/C	ENSG000000136925	TSTD2	Variante ganancia codón stop	-	-
9:114804191	C/T	ENSG000000106868	SUSD1	Variante ganancia codón stop	rs146843448	0,0013
9:21481483	G/A	ENSG000000184995	IFNE	Variante ganancia codón stop	rs2039381	0,0026

9:95874220	G/T	ENSG000000165233	C9orf89	Variante ganancia codón stop, Variante <i>splicing</i>	rs141857568	-
11:4598956	C/T	ENSG000000171987	C11orf40	Variante ganancia codón stop	rs794448530	0,05
11:60184047	C/T	ENSG000000166928	MS4A14	Variante ganancia codón stop	-	-
11:7509386	C/T	ENSG000000183801	OLFML1	Variante ganancia codón stop	rs141351486, COSM1322159	0,01
11:8959327	G/A	ENSG000000176009	ASCL3	Variante ganancia codón stop	rs141568989	0,0026
12:1022569	A/C	ENSG00000002016	RAD52	Variante ganancia codón stop	rs4987208,CM051217	0,01
12:1023218	G/T	ENSG00000002016	RAD52	Variante ganancia codón stop	rs4987207,CM025484, COSM178088	0,02
12:111893803	T/C	ENSG000000204842	ATXN2	Variante pérdida codón stop	-	-
12:40873278	T/A	ENSG000000205592	MUC19	Variante ganancia codón stop	rs11176811	0,2
12:58019202	C/T	ENSG000000135502	SLC26A10	Variante ganancia codón stop	rs151077423	0,01
12:6666359	G/A	ENSG000000111641	NOP2	Variante ganancia codón stop	-	-
12:80169711	A/T	ENSG000000058272	PPP1R12A	Variante pérdida codón stop	-	-
13:114782808	G/A	ENSG000000185989	RASA3	Variante ganancia codón stop	COSM24409	-
13:22077082	T/A	ENSG000000165487	MICU2	Variante ganancia codón stop	rs145410850	0,0026
13:32972626	A/T	ENSG000000139618	BRCA2	Variante ganancia codón stop	rs11571833,CM993644	0,01
13:46287373	C/A	ENSG000000174015	SPERT	Variante ganancia codón stop	rs80072371, COSM277457	0,05
13:53617309	C/T	ENSG000000102837	OLFM4	Variante ganancia codón stop	rs34067666	0,03
14:100808596	G/A	ENSG000000140105	WARS	Variante ganancia codón stop	-	-
14:22690191	G/A	ENSG000000211814	TRAV35	Variante ganancia codón stop	rs17198047	0,05
16:33629700	G/A	ENSG000000271178	IGHV3OR1 6-13	Variante ganancia codón stop	rs2019670	-
16:57464236	G/A	ENSG000000005194	CIAPIN1	Variante ganancia codón stop	rs170358	0,04
16:81174978	A/G	ENSG000000166473	PKD1L2	Variante pérdida codón stop	rs117006360	0,06
16:88780062	C/A	ENSG000000174177	CTU2	Variante ganancia codón stop	rs201111272	-
16:89294865	C/A	ENSG000000170100	ZNF778	Variante ganancia codón stop	rs201425656	-
17:3195603	T/A	ENSG000000180090	OR3A1	Variante ganancia codón stop	rs7218125	0,0013
17:37898826	C/T	ENSG000000141738	GRB7	Variante ganancia codón stop	-	-

17:39595484	G/A	ENSG00000171360	KRT38	Variante ganancia codón stop	rs148768443, COSM162057	-
17:42429897	C/A	ENSG00000030582	GRN	Variante ganancia codón stop	-	-
17:72931912	C/T	ENSG00000182938	OTOP3	Variante ganancia codón stop	rs201177829	-
17:76475703	G/T	ENSG00000187775	DNAH17	Variante ganancia codón stop	-	-
17:7760574	C/T	ENSG00000183011	LSMD1	Variante ganancia codón stop	rs144163075	0,01
18:21399820	C/A	ENSG00000053747	LAMA3	Variante ganancia codón stop	-	-
19:11034179	C/T	ENSG00000130733	YIPF2	Variante ganancia codón stop	-	-
19:16268425	C/T	ENSG00000196684	HSH2D	Variante ganancia codón stop	rs113641420	0
19:1783027	T/C	ENSG00000130270	ATP8B3	Variante pérdida codón stop	rs60482625	0,05
19:35718891	C/T	ENSG00000177558	FAM187B	Variante ganancia codón stop	rs35001809	0,04
19:38875072	G/A	ENSG00000179168	GGN	Variante ganancia codón stop	rs62123481, COSM1680985, COSM1680986	0,02
19:43585262	C/T	ENSG00000242221	PSG2	Variante ganancia codón stop	-	-
19:4542967	G/A	ENSG00000167680	SEMA6B	Variante ganancia codón stop	-	-
19:46273841	A/T	ENSG00000104936	DMPK	Variante ganancia codón stop	-	-
19:53085847	C/T	ENSG00000167562	ZNF701	Variante ganancia codón stop	rs144556038	0
19:53626794	C/T	ENSG00000170954	ZNF415	Variante ganancia codón stop	rs35443760	0,09
19:54006913	T/A	ENSG00000198346	ZNF813	Variante ganancia codón stop	-	-
19:54870191	T/A	ENSG00000167613	LAIR1	Variante ganancia codón stop	rs111949142	0,16
19:55568144	C/A	ENSG00000160439	RDH13	Variante ganancia codón stop	-	-
19:57987104	C/T	ENSG00000197128	ZNF772	Variante ganancia codón stop	rs73067111	0,01
19:687035	T/A	ENSG00000185198	PRSS57	Variante ganancia codón stop	rs150190105	-
19:9804446	C/T	ENSG00000224689	ZNF812	Variante ganancia codón stop	rs142939425	0
20:44511257	G/A	ENSG00000168612	ZSWIM1	Variante ganancia codón stop	rs35972756	0,02
21:43913159	C/A	ENSG00000160188	RSPH1	Variante ganancia codón stop	rs138320978	-
22:17072347	C/T	ENSG00000198445	CCT8L2	Variante ganancia codón stop	rs139948519	0,0026
22:29656431	C/T	ENSG00000100263	RHBDD3	Variante ganancia codón stop	rs138870856	0,01
X:54777550	C/A	ENSG00000102313	ITIH6	Variante ganancia codón stop	rs139183502	-

A continuación, con la finalidad de encontrar diferencias entre ambos grupos de estudio (controles y diabéticos), se analizó la distribución y posición cromosómica de las diferentes variantes genéticas en el grupo control (Figura 43).

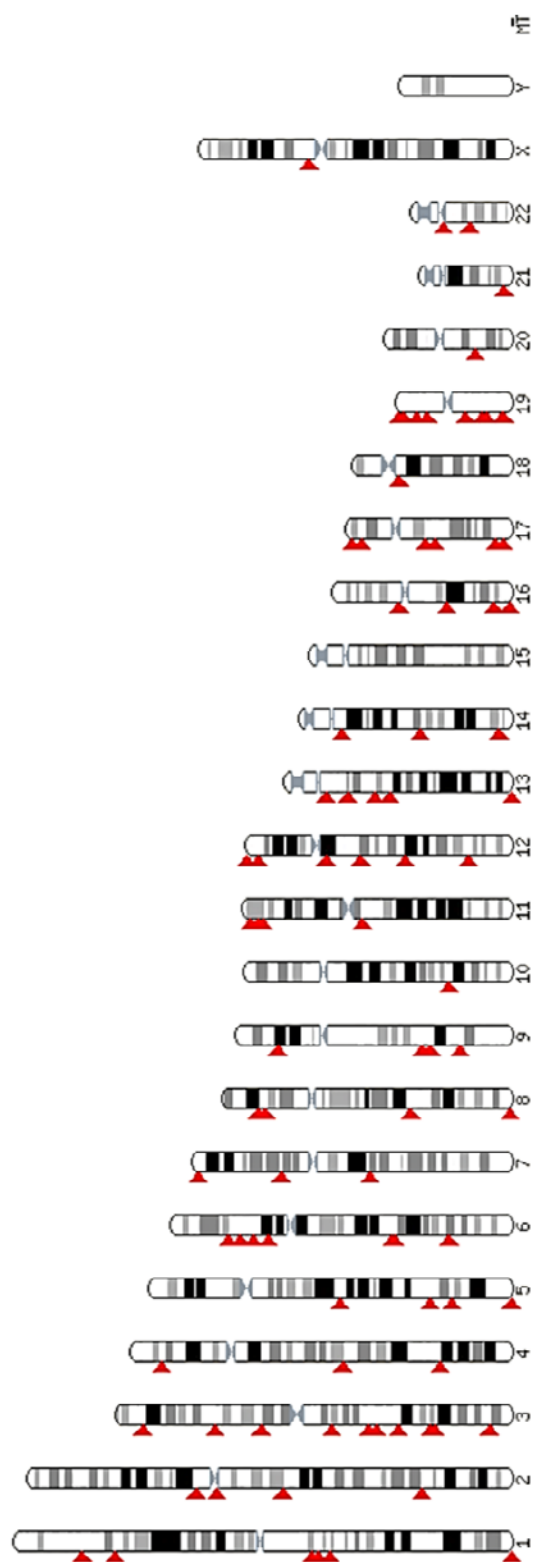
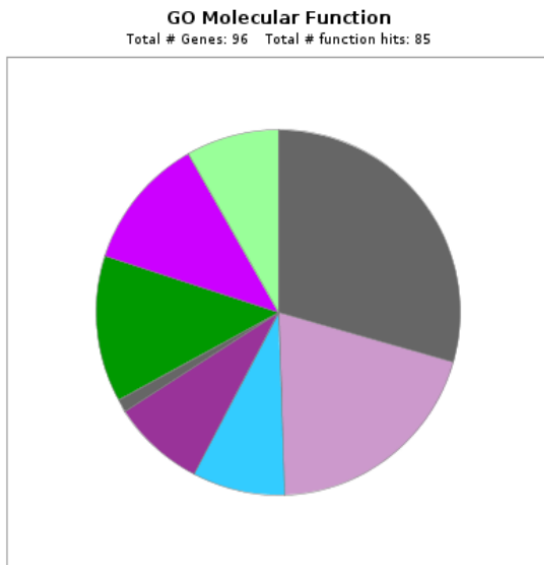


Figura 43. Cariograma donde se representa la posición cromosómica de las variantes genéticas de codón de stop verificadas en el grupo de controles.

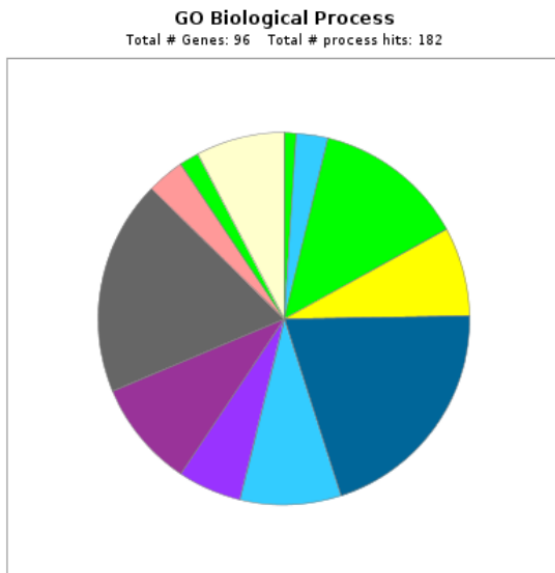
Tras comprobar que variantes genéticas eran ciertas en el grupo control, se realizó el análisis informático para ver que procesos biológicos estaban más enriquecidos para esos genes.

- **Función molecular:**



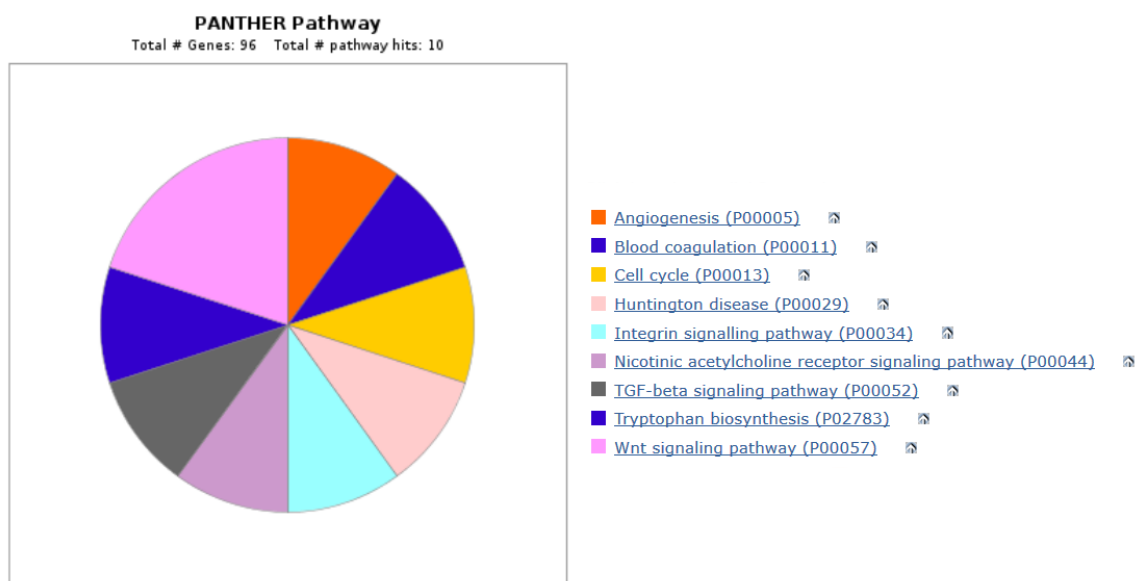
- [binding \(GO:0005488\)](#) ⓘ
- [catalytic activity \(GO:0003824\)](#) ⓘ
- [enzyme regulator activity \(GO:0030234\)](#) ⓘ
- [nucleic acid binding transcription factor activity \(GO:0001071\)](#) ⓘ
- [protein binding transcription factor activity \(GO:0000988\)](#) ⓘ
- [receptor activity \(GO:0004872\)](#) ⓘ
- [structural molecule activity \(GO:0005198\)](#) ⓘ
- [transporter activity \(GO:0005215\)](#) ⓘ

- **Proceso Biológico:**



- [apoptotic process \(GO:0006915\)](#) ⓘ
- [biological adhesion \(GO:0022610\)](#) ⓘ
- [biological regulation \(GO:0065007\)](#) ⓘ
- [cellular component organization or biogenesis \(GO:0071840\)](#) ⓘ
- [cellular process \(GO:0009987\)](#) ⓘ
- [developmental process \(GO:0032502\)](#) ⓘ
- [immune system process \(GO:0002376\)](#) ⓘ
- [localization \(GO:0051179\)](#) ⓘ
- [metabolic process \(GO:0008152\)](#) ⓘ
- [multicellular organismal process \(GO:0032501\)](#) ⓘ
- [reproduction \(GO:0000003\)](#) ⓘ
- [response to stimulus \(GO:0050896\)](#) ⓘ

- Rutas metabólicas implicadas:



7.2. Estudio de variantes genéticas de codón de stop en el grupo de casos

Seguidamente, se realizó el análisis y verificación de las variantes de stop para el grupo de casos; siendo 50 válidas de un total de 67 variantes identificadas inicialmente (Tabla 17).

Tabla 17. Variantes genéticas identificadas y verificadas de codón de stop para el grupo de casos.

Posición Cr.	Cambio	Ensembl ID	Gen	Tipo de Variación	SNP	MAF europeo
1:114267508	G/A	ENSG00000116793	<i>PHTF1</i>	Variante ganancia codón stop	rs200552382	-
1:145416808	C/T	ENSG00000168509	<i>HFE2</i>	Variante ganancia codón stop	CM041369	-
1:162829293	C/T	ENSG00000185860	<i>C1orf110</i>	Variante ganancia codón stop	-	-
1:20005860	G/T	ENSG00000158748	<i>HTR6</i>	Variante pérdida codón stop	rs369184848	-
1:201175258	C/T	ENSG00000163395	<i>IGFN1</i>	Variante ganancia codón stop	-	-
1:20517763	C/T	ENSG00000162543	<i>UBXN10</i>	Variante ganancia codón stop	rs142831657	0,0026
1:236706300	T/A	ENSG00000116977	<i>LGALS8</i>	Variante ganancia codón stop	rs2273865	0,03
1:85020683	C/G	ENSG00000117151	<i>CTBS</i>	Variante pérdida codón stop	rs138201181, COSM1249364	0
2:108863758	G/A	ENSG00000196228	<i>SULT1C3</i>	Variante ganancia codón stop	rs112050262	0,03
2:198593260	A/C	ENSG00000152430	<i>BOLL</i>	Variante ganancia codón stop	rs74375706	0
2:209025758	G/A	ENSG00000168582	<i>CRYGA</i>	Variante ganancia codón stop	rs116344874	0
4:111474532	C/T	ENSG00000138792	<i>ENPEP</i>	Variante ganancia codón stop	rs201624755, COSM265701	-
4:141310389	G/A	ENSG00000153132	<i>CLGN</i>	Variante ganancia codón stop	rs200583755	-
5:140229556	C/A	ENSG00000204961	<i>PCDHA9</i>	Variante ganancia codón stop	-	-
5:140588810	A/T	ENSG00000120328	<i>PCDHB12</i>	Variante ganancia codón stop	rs187438401	0
6:150390149	G/C	ENSG00000131019	<i>ULBP3</i>	Variante ganancia codón stop	rs72501734	0,01
6:25778182	C/T	ENSG00000146039	<i>SLC17A4</i>	Variante ganancia codón stop	rs2328894	0,01
6:31125257	C/A	ENSG00000204536	<i>CCHCR1</i>	Variante ganancia codón stop	rs72856718	-
6:47682599	G/T	ENSG00000153294	<i>GPR115</i>	Variante ganancia codón stop	-	-

7:100389677	C/T	ENSG000000146839	ZAN	Variante ganancia codón stop	rs149104440	0,01
7:141895976	C/T	ENSG000000257743	RP11-1220K2.2	Variante ganancia codón stop	-	-
7:45931540	C/T	ENSG000000146678	IGFBP1	Variante ganancia codón stop	-	-
8:16012594	G/A	ENSG000000038945	MSR1	Variante ganancia codón stop	rs41341748,CM023579	0,01
10:1065457	T/A	ENSG000000148377	IDJ2	Variante pérdida codón stop	rs41314629	0,01
10:115338424	C/T	ENSG000000148702	HABP2	Variante ganancia codón stop	rs41292628	0,0026
10:4889403	C/T	ENSG000000165568	AKR1E2	Variante ganancia codón stop	rs12240276	0,12
10:79784801	G/A	ENSG000000148606	POLR3A	Variante ganancia codón stop	rs267608678	-
10:85994115	C/T	ENSG000000148602	LRIT1	Variante ganancia codón stop	rs201482815	-
11:60714017	G/A	ENSG000000110446	SLC15A3	Variante ganancia codón stop	rs116597710	0
11:62850775	G/A	ENSG000000197658	SLC22A24	Variante ganancia codón stop	rs77002186	0,02
12:112167654	C/T	ENSG000000111271	ACAD10	Variante ganancia codón stop	rs146335507	-
12:13240162	G/A	ENSG000000111305	GSG1	Variante ganancia codón stop	rs141509089	0,0013
12:18800961	G/C	ENSG000000139144	PIK3C2G	Variante pérdida codón stop	rs61757718	0,02
12:6675094	G/A	ENSG000000139190	VAMP1	Variante ganancia codón stop	-	-
12:81064144	C/T	ENSG000000139304	PTPRQ	Variante ganancia codón stop	-	-
13:52605154	C/G	ENSG000000253797	UTP14C	Variante ganancia codón stop	rs141049018	-
15:43483698	C/T	ENSG000000166946	CCNDBP1	Variante ganancia codón stop	-	-
15:75015234	G/A	ENSG000000140465	CYP1A1	Variante ganancia codón stop	-	-
15:91474631	C/T	ENSG000000184508	HDDC3	Variante ganancia codón stop	rs117318472	0,01
16:16367613	G/T	ENSG000000103226	NOMO3	Variante ganancia codón stop	rs202070282	-
16:1825982	C/T	ENSG000000197774	EME2	Variante ganancia codón stop	rs61753375	0,02
16:20802007	C/G	ENSG000000005187	ACSM3	Variante ganancia codón stop	rs34381224,COSM702435	0,004
16:67220231	G/A	ENSG000000179044	EXOC3L1	Variante ganancia codón stop	rs141772271	0,01

16:67980235	G/T	ENSG000000124067	SLC12A4	Variante ganancia codón stop	-	-
16:77756501	G/T	ENSG000000140876	NUDT7	Variante ganancia codón stop	rs182579196	0,0013
17:64881037	C/T	ENSG000000075429	CACNG5	Variante ganancia codón stop	rs62621240	0
17:7077122	C/T	ENSG000000141505	ASGR1	Variante ganancia codón stop	rs200584799	-
17:78092070	C/T	ENSG000000171298	GAA	Variante ganancia codón stop	rs121907943,CM930288	-
19:55021777	T/A	ENSG000000167618	LAI/R2	Variante pérdida codón stop	rs62131570	0,06
19:57838058	G/A	ENSG000000178229	ZNF543	Variante ganancia codón stop	rs150392165	0,01

Del mismo modo, se analizó la posición cromosómica de las diferentes variantes genéticas en los casos (Figura 44).

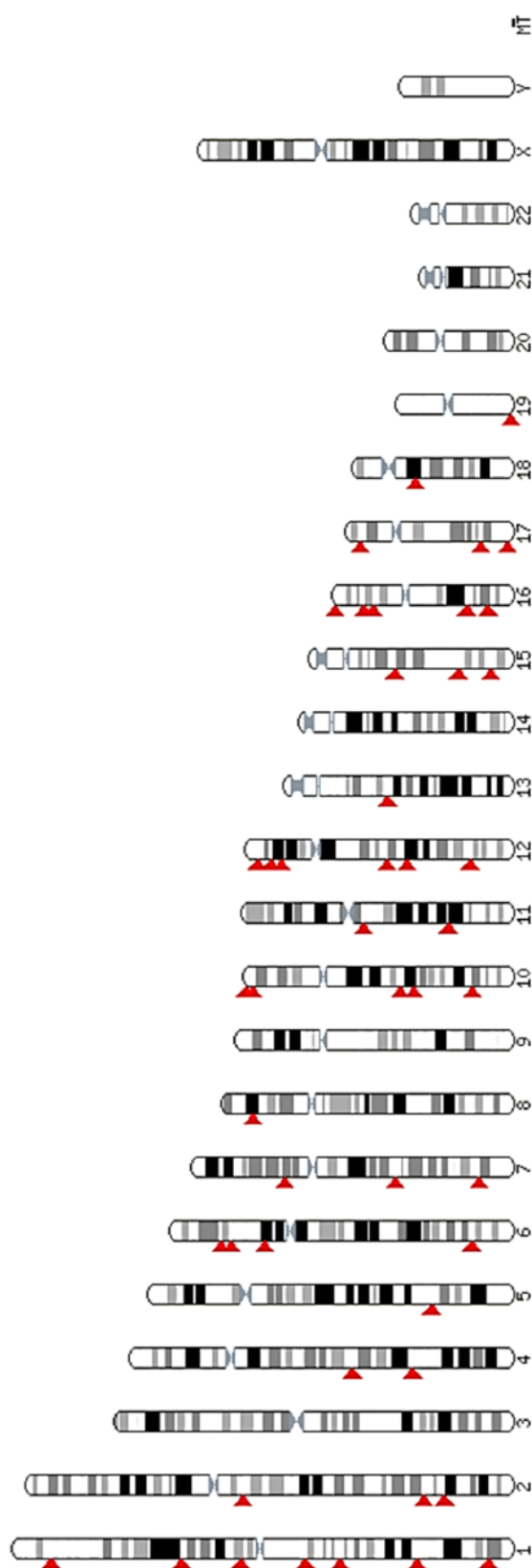
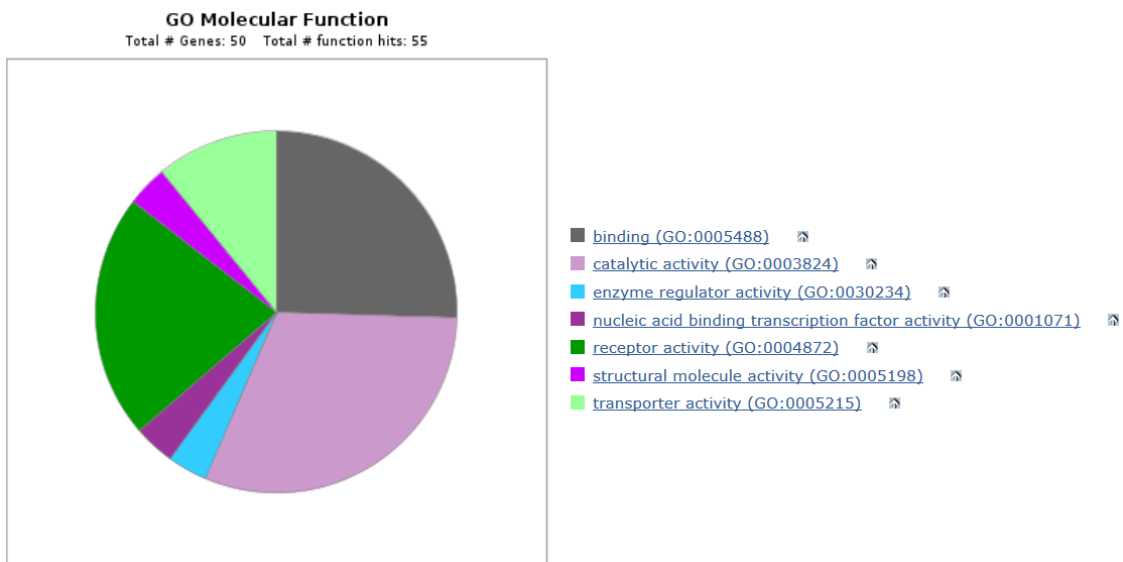


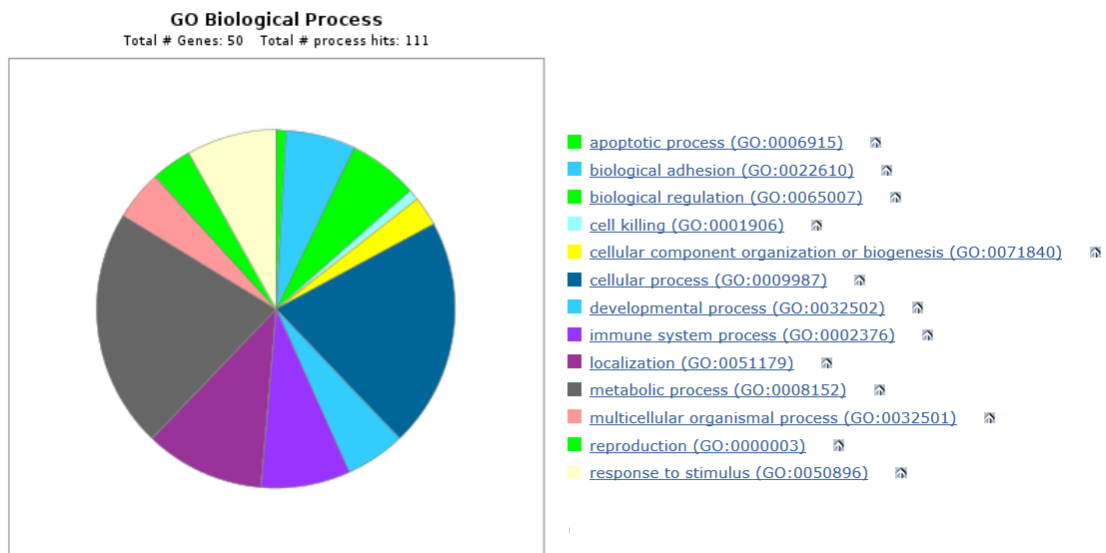
Figura 44. Cariograma donde se representa la posición cromosómica de las variantes genéticas de codón de stop verificadas en el grupo de casos.

Después de la verificación de los SNPs para el grupo control, se estudiaron que funciones moleculares y procesos biológicos así como rutas metabólicas estaban más representados en los pacientes con DM2.

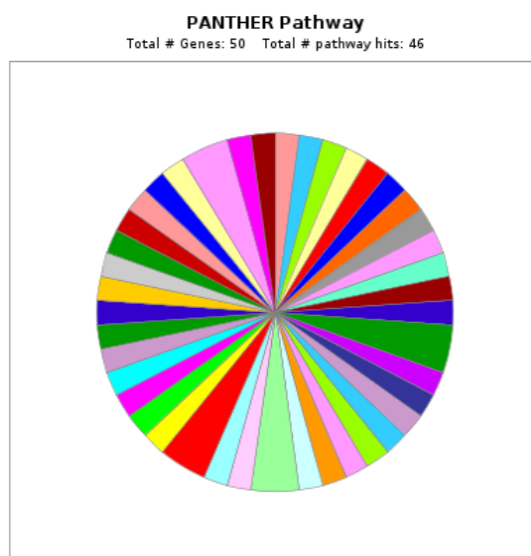
- **Función Molecular:**



- **Proceso biológico:**



- Rutas metabólicas implicadas:



- [SHT1 type receptor mediated signaling pathway \(P04373\)](#) ⓘ
- [SHT2 type receptor mediated signaling pathway \(P04374\)](#) ⓘ
- [SHT3 type receptor mediated signaling pathway \(P04375\)](#) ⓘ
- [SHT4 type receptor mediated signaling pathway \(P04376\)](#) ⓘ
- [Adrenaline and noradrenaline biosynthesis \(P00001\)](#) ⓘ
- [Alpha adrenergic receptor signaling pathway \(P00002\)](#) ⓘ
- [Angiogenesis \(P00005\)](#) ⓘ
- [Axon guidance mediated by netrin \(P00009\)](#) ⓘ
- [Beta1 adrenergic receptor signaling pathway \(P04377\)](#) ⓘ
- [Beta2 adrenergic receptor signaling pathway \(P04378\)](#) ⓘ
- [Beta3 adrenergic receptor signaling pathway \(P04379\)](#) ⓘ
- [Blood coagulation \(P00011\)](#) ⓘ
- [Cadherin signaling pathway \(P00012\)](#) ⓘ
- [Cholesterol biosynthesis \(P00014\)](#) ⓘ
- [Corticotropin releasing factor receptor signaling pathway \(P04380\)](#) ⓘ
- [Dopamine receptor mediated signaling pathway \(P05912\)](#) ⓘ
- [EGF receptor signaling pathway \(P00018\)](#) ⓘ
- [Endothelin signaling pathway \(P00019\)](#) ⓘ
- [FGF signaling pathway \(P00021\)](#) ⓘ
- [Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway \(P00026\)](#) ⓘ
- [Hypoxia response via HIF activation \(P00030\)](#) ⓘ
- [Inflammation mediated by chemokine and cytokine signaling pathway \(P00031\)](#) ⓘ
- [Insulin/IGF pathway-protein kinase B signaling cascade \(P00033\)](#) ⓘ
- [Integrin signalling pathway \(P00034\)](#) ⓘ
- [Ionotropic glutamate receptor pathway \(P00037\)](#) ⓘ
- [Metabotropic glutamate receptor group II pathway \(P00040\)](#) ⓘ
- [Metabotropic glutamate receptor group III pathway \(P00039\)](#) ⓘ
- [Muscarinic acetylcholine receptor 1 and 3 signaling pathway \(P00042\)](#) ⓘ
- [Muscarinic acetylcholine receptor 2 and 4 signaling pathway \(P00043\)](#) ⓘ
- [Nicotinic acetylcholine receptor signaling pathway \(P00044\)](#) ⓘ
- [Opioid prodynorphin pathway \(P05916\)](#) ⓘ
- [Opioid proenkephalin pathway \(P05915\)](#) ⓘ
- [Opioid proopiomelanocortin pathway \(P05917\)](#) ⓘ
- [Oxytocin receptor mediated signaling pathway \(P04391\)](#) ⓘ
- [PI3 kinase pathway \(P00048\)](#) ⓘ
- [Synaptic vesicle trafficking \(P05734\)](#) ⓘ
- [T cell activation \(P00053\)](#) ⓘ
- [Thyrotropin-releasing hormone receptor signaling pathway \(P04394\)](#) ⓘ
- [VEGF signaling pathway \(P00056\)](#) ⓘ
- [Wnt signaling pathway \(P00057\)](#) ⓘ
- [p53 pathway feedback loops 2 \(P04398\)](#) ⓘ
- [p53 pathway \(P00059\)](#) ⓘ

Discusión

La DM2 es una enfermedad compleja, heterogénea y multifactorial en la cual intervienen factores genéticos, ambientales y adquiridos. El conjunto de *loci* y variantes genéticas que intervienen en la enfermedad están relacionados principalmente con la disfunción de las células β pancreáticas y con alteraciones metabólicas como la resistencia a la insulina. El efecto de los diferentes genes implicados junto con factores ambientales, como la obesidad, modulan el umbral para que se desarrolle o no la enfermedad (242). Diversos estudios muestran la importante heredabilidad de la DM2, estableciendo que posiblemente es superior al 70% (91, 167).

Como se describió anteriormente, son varias las estrategias que se han seguido para conocer que genes están asociados a la DM2 (38, 243): estudio de genes candidatos, estudio de ligamiento de familias, GWAS, etc. Sin embargo, pese a todos los estudios realizados, el conjunto de *loci* y variantes genéticas identificadas explican únicamente entre el 10 y 15% de la heredabilidad de la enfermedad (196, 223), al igual que sucede en otras enfermedades complejas.

La hipótesis “enfermedades comunes, variantes raras” plantea la búsqueda de variantes poco frecuentes y raras que, en su conjunto, podrían explicar gran parte de la heredabilidad perdida de la enfermedad. Ello hace necesario la búsqueda de nuevas estrategias para abordar el estudio de la genética de enfermedades complejas mediante nuevas aproximaciones (211).

Por este motivo uno de los principales objetivos planteados en este estudio fue identificar nuevos *loci* y variantes genéticas que pudiesen contribuir a explicar parte del componente genético desconocido de la DM2, incluyendo variantes poco frecuentes y raras. En concreto, el abordaje del estudio fue la secuenciación de todo el exoma en 400 muestras (200 controles y 200 casos), permitiendo de esta manera hacer un *screening* genético global en población española.

Así, tras la secuenciación de exoma y posterior análisis bioinformático de las 400 muestras, se han identificado un gran número de variantes genéticas (Veáse apartado resultados). Después de filtrar los datos y priorizar las variantes según su posible efecto a nivel funcional (SIFT y PolyPhen), distribución diferencial entre casos y controles, y una frecuencia alélica inferior al 20%; más de 20.000 SNPs con importante efecto funcional fueron identificados en controles y casos, independientemente. En concreto, en este proyecto nos centramos en las variantes de mayor impacto funcional:

- Variantes que afectan a regiones de *splicing*
- Variantes de *missense* o de cambio de sentido
- Variantes en regiones que codifican para un microARN maduro
- Variantes genéticas de pérdida o ganancia de un codón stop o parada

1. CRITERIOS DE SELECCIÓN DE LA POBLACIÓN DE ESTUDIO

Nuestro estudio se centró en la identificación de variantes genéticas relacionadas con la DM2 en la situación de mayor riesgo y donde mayor cantidad de casos de DM2 se producen: personas con edad media y sobrepeso-obesidad ligera. Como se ha descrito anteriormente, la selección correcta de casos y controles fue un factor clave a la hora de diseñar el experimento. Así, con la finalidad de aumentar la probabilidad de encontrar variantes genéticas que puedan estar relacionadas con la enfermedad en un grupo reducido de muestras, 200 controles y 200 casos, nos centramos en seleccionar los individuos según dos características clave: el IMC y la edad.

- Índice de Masa Corporal: se seleccionaron aquellos individuos con mayor carga genética posible, es decir, pacientes diabéticos con un IMC no excesivamente elevado (sobrepeso a obesidad grado I, 25-34.9 kg/m² según la clasificación SEEDO, 2007). Estudios han mostrado que la importancia del fondo genético es mayor en las personas con normopeso o sobrepeso ligero (IMC <24 y 24-27.9) que en obesos (IMC≥30) (244).
- Edad: la exposición a factores de riesgo se incrementa con la edad, esto supone que aquellos pacientes que desarrollan DM2 a edades mayores, podrían tener un menor componente genético. En nuestro estudio, seleccionamos tanto para controles como casos, una edad comprendida entre 40 y 65 años, siendo la media alrededor de 55 años. Sin embargo, en el estudio de Voight et al. la edad no parece alterar la asociación de los marcadores (178), posiblemente por la mezcla de fenotipos incluida en este estudio.

Todo ello implica que la selección de controles y casos tuvo que hacerse de forma cuidadosa e intentando seguir la información obtenida en base a estudios anteriores, con la finalidad de obtener mejores resultados. Así, la selección de pacientes de DM2 con

antecedentes familiares, con sobrepeso u obesidad tipo I y medianamente jóvenes (40-65 años) permitió seleccionar un grupo de casos con un componente genético predominante y poco desvirtuado, en el que se han encontrado un gran número de variantes genéticas con posible efecto funcional.

2. ESTUDIO DE VARIANTES GENÉTICAS DE BAJA FRECUENCIA O RARAS

Diversos estudios han cuantificado la heredabilidad de la DM2 y de parámetros relacionados. La concordancia entre gemelos idénticos en la presencia de DM2 o intolerancia a la glucosa es superior al 60% en estudios transversales, pudiendo llegar al 76% y al 96%, respectivamente, si se realizan seguimientos de hasta 15 años (91, 167).

Muchos son los estudios que se han realizado con el objetivo de identificar que genes están implicados en la enfermedad. No obstante, los estudios que mayor cantidad de resultados han proporcionado hasta el momento y cuyos resultados han podido ser validados en muchos casos, son los estudios de GWAS.

Los GWAS han permitido identificar variantes genéticas comunes que incrementan el riesgo de enfermedades complejas como la DM2, tal y como se ha indicado en la introducción. Sin embargo, la mayoría de *loci* identificados tienen un efecto pequeño a nivel funcional sobre el fenotipo o de asociación con el fenotipo (OR entre 1,1 y 1,5) (245). Otra limitación de los GWAS es que los chips utilizados hasta el momento dejan muchas regiones genómicas sin una cobertura buena. Si bien actualmente se están comercializando chips con muchos más polimorfismos (2 millones de SNPs por muestra) que también incluyen PPFs (MAF <0,05) y presentan una cobertura mayor. Sin embargo, el aumento en el número de polimorfismos puede ocasionar diferentes limitaciones. La cobertura podría ser insuficiente en muchas zonas del genoma dado que se necesitaría una densidad mucho mayor de la alcanzada para poder asegurar que se analiza toda la variabilidad de cada región. Además, los tamaños muestrales necesarios son excesivamente grandes para poder detectar muchas de las asociaciones esperables, principalmente en los PPFs, lo que supone un coste elevadísimo.

Las variantes comunes sólo han permitido explicar una pequeña parte del componente genético calculado en estudios con gemelos y de ligamiento de familias para la DM2 (224). Esta diferencia entre la heredabilidad explicada y la calculada con los estudios genéticos realizados hasta el momento se ha denominado “heredabilidad perdida”.

Una parte importante de los esfuerzos científicos de los últimos años se han centrado en intentar explicar parte de la “heredabilidad perdida” de enfermedades complejas (211, 224, 246). Así, la teoría genética poblacional sugiere que si variantes causantes de enfermedad son afectadas por selección natural que conducen a una ligera disminución en la capacidad reproductiva de los individuos portadores, entonces alteraciones que afecten al desarrollo de la enfermedad podrían estar en baja proporción en la población. Por ello, una proporción de la heredabilidad podría ser explicada por PPFs y variantes raras más que por comunes (220, 247). Además, el elevado número de PPFs y su potencial impacto (posiblemente con OR >3), en su conjunto, pueden estar detrás de un importante porcentaje del componente genético de la DM2 (199, 221, 225). De este modo, estudios recientes han empezado a plantear el estudio de las variantes de baja frecuencia o raras (frecuencia <1%) en relación a la DM2. (221, 248). No obstante, este tipo de variantes no han podido ser detectadas en su inmensa mayoría por los estudios realizados hasta el momento. En primer lugar, los chips de genotipado que se han utilizado están sesgados en contra de la inclusión de variantes de baja frecuencia, con lo que nunca se han asociado directamente con un parámetro concreto de enfermedad. Y en segundo lugar, las variantes de baja frecuencia no están en desequilibrio de ligamiento con las variantes comunes genotipadas en los GWAS, por lo que no han sido detectadas al no estar correlacionadas con un SNP común concreto.

Los nuevos avances tecnológicos y la aparición de la NGS, así como la relativa asequibilidad económica de la secuenciación de exoma, está permitiendo descubrir nuevas asociaciones entre variantes de baja frecuencia y enfermedades complejas. Recientemente, un estudio de secuenciación de genoma completo ha demostrado la asociación de algunas variantes raras con DM2, poniendo de manifiesto la implicación e importancia de este tipo de variantes en enfermedades complejas. Se descubrieron 4 nuevas variantes que afectan al riesgo de DM2 (207). Una variante de baja frecuencia (1,47%) en el intron 1 del gen *CCND2*, rs76895963 [G], reduce el riesgo de DM2 a la mitad (OR = 0,53, P = $5,0 \times 10^{-21}$). Además, dos variantes de cambio de sentido en el

gen PAM, p.Asp563Gly (frecuencia de 4,98%) y p.Ser539Trp (frecuencia de 0,65%), confieren un riesgo mayor de padecer la enfermedad (OR = 1,23, P = $3,9 \times 10^{-10}$) y OR = 1,47, P = $1,7 \times 10^{-5}$, respectivamente). Y, una variante en el gen PDX-1 (0,20%) p.Gly218Ala, se asocia con un alto riesgo de DM2 (OR = 2,27, P = $7,3 \times 10^{-7}$).

Por todo ello, y bajo la hipótesis de que, en conjunto, variantes de baja frecuencia pueden explicar una fracción sustancial de la heredabilidad de enfermedades comunes como la DM2, el principal objetivo planteado en este trabajo fue la identificación de variantes genéticas poco frecuentes (MAF <0,05) y raras (MAF <0,01) presentes en el exoma. Las principales razones por las cuales este proyecto se ha centrado en el estudio de este tipo de variantes pueden resumirse del siguiente modo:

- El componente genético de una gran parte de la heredabilidad calculada para la DM2 todavía no se ha identificado.
- Los estudios genéticos realizados anteriormente, como GWAS, solo permiten identificar la asociación de PFs (MAF >0,05). Sin embargo, estos tienen, en general, un efecto pequeño a nivel funcional y sobre el fenotipo. Además, los GWAS utilizados hasta el momento se han realizado con chips que dejan muchas regiones genómicas sin una buena cobertura.
- El avance tecnológico y los sistemas de secuenciación de nueva generación, están haciendo posible llevar a cabo proyectos para secuenciar un mayor número de muestras y analizar así variantes tanto frecuentes como poco frecuentes.
- Dado el elevado número de PPFs y su potencial impacto, en su conjunto, pueden estar detrás de un importante porcentaje del componente genético de la DM2.
- Los PPFs y variantes raras con efecto funcional podrían explicar gran parte de la heredabilidad no encontrada.

Así pues, en este trabajo todas las variantes genéticas encontradas fueron filtradas por una frecuencia alélica europea menor del 20%; lo que implica la inclusión de PPFs y variantes raras.

3. SECUENCIACIÓN DE NUEVA GENERACIÓN: ESTUDIO DEL EXOMA

La tecnología de NGS ha tenido un gran desarrollo en los últimos años, permitiendo realizar estudios de secuenciación de numerosos exomas en una reacción e incrementándose constantemente la capacidad de los equipos. Además, todo ello ha ido acompañado por una mejora en la capacidad y fiabilidad de los sistemas de análisis y alineamiento de las secuencias obtenidas.

El método de secuenciación seleccionado en este estudio ha sido la secuenciación de exoma (*Whole exome sequencing*). Más del 90% de mutaciones y polimorfismos asociados a enfermedades se sitúan en las regiones codificantes de proteínas y, por tanto, en el exoma (213, 214). Sin embargo, dado que el exoma representa alrededor del 1-2% de todo el genoma (≈ 180.000 exones, 60 Mpb), el análisis para la identificación de variantes resulta mucho más rentable económicamente comparado con la secuenciación del genoma completo (249). En este estudio, tal y como se indicó en la metodología, la tecnología de secuenciación empleada fue la TruSeq de Illumina. Comparado con otros sistemas de captura de exoma, este cubre una mayor parte de bases exónicas en todas las bases de datos; pudiendo obtener una cobertura máxima del 96,9% (250).

Asimismo, el análisis de polimorfismos en todo el genoma muestra una concentración mayor de polimorfismos asociados con enfermedades complejas en el exoma frente al resto del genoma (215, 251, 252). Incluso aquellas alteraciones que modifican el *splicing* del ARN y se encuentren en las regiones intrónicas contiguas a los exones podrán ser identificadas mediante la captura del exoma (216, 217).

Una de las principales características de la secuenciación de exoma es que permite identificar todos los polimorfismos presentes, incluyendo PFs, PPFs o variantes raras. Muchas de las variantes genéticas identificadas son potencialmente funcionales al alterar la secuencia de las proteínas o el procesado de su ARN y, por lo tanto, pueden tener un efecto sobre el fenotipo.

Así, en nuestro estudio un elevado número de SNPs codificantes fueron identificados. Concretamente, del total de SNPs identificados inicialmente (Veáse resultados), 17.238 polimorfismos en casos y 21.822 SNPs en controles cumplían con los criterios de selección y presentaban un impacto funcional sobre el fenotipo del individuo.

En el estudio de 200 exomas se identificaron 120.000 SNPs de los cuales 53.000 eran SNPs codificantes con posibles efectos funcionales sobre la proteína; apuntando que hay muchos más por identificar, principalmente PPFs (209, 215). Es importante remarcar el elevado el número de variantes de pérdida de función presentes en cada persona, hasta 250-300 en todos los genes identificados y entre 50 y 100 en genes causantes de enfermedades hereditarias (208, 209). Estos estudios apoyan nuestros resultados y corroboran el elevado número de variantes genéticas con un posible efecto funcional encontradas.

Un trabajo reciente secuenció el exoma de 177 indios Pima y, tras genotipar las variantes seleccionadas en un mayor número de individuos (n=7667), se observó que variantes en los genes *CYB5A* y *RNF10*, aumentaban el riesgo de DM2 en la población estudiada (253). En otro estudio, empleando chips de genotipado y secuenciación de exoma, se identificó una variante sin sentido (p.Arg684Ter) en el gen *TBC1D4*. Homocigotos portadores de esta variante tendrían mayores concentraciones de glucosa en plasma y de insulina en suero tras 2h de sobrecarga oral de glucosa, en comparación con individuos con otros genotipos (254).

Por otro lado, en el estudio de Lohmueller et al (2013) se secuenciaron los exomas de 1.000 casos con DM2 y 1.000 controles sanos y se aplicó una serie de simulaciones para detectar los genes de susceptibilidad a la DM2. Sin embargo, no se encontró ningún gen que mostrara una asociación significativa con el riesgo de desarrollar la enfermedad concluyendo que variantes raras responsables del riesgo de DM2 se distribuirían en un elevado número de genes y no únicamente en unos pocos *loci* (206). Estos resultados muestran la necesidad de un elevado poder estadístico para establecer asociación entre variante rara y enfermedad si se pretende realizar su estudio mediante la secuenciación del exoma únicamente. En nuestro caso, primero pretendemos hacer una criba de potenciales genes candidatos mediante su identificación en el exoma y en futuros estudios, verificar su posible asociación mediante su análisis en muestras de población amplias y estudios funcionales.

4. COBERTURA DE SECUENCIACIÓN Y ANÁLISIS BIOINFORMÁTICO

La secuenciación de nueva generación implica el manejo de una gran cantidad de datos que requiere de una infraestructura bioinformática compleja. El análisis bioinformático no sólo tiene que facilitar el preprocesado de la información obtenida, sino también el alineamiento, filtrado de los datos y controles de calidad necesarios para la detección, anotación de las variantes genéticas y estimación del efecto de cada alteración.

Uno de los factores que hay que tener en cuenta a la hora de diseñar un experimento de acuerdo a los objetivos planteados, es el compromiso que debe existir entre número de muestras y cobertura (profundidad de lectura) para un presupuesto determinado. Así, en el caso de muestras individuales el procedimiento correcto sería realizar una secuenciación profunda por individuo con una mayor sensibilidad para encontrar variantes en la muestra y un genotipado exacto. Por el contrario, en el caso de tener muchas muestras, la secuenciación multi-muestra se debería realizar a baja profundidad (baja cobertura) donde la sensibilidad depende de la frecuencia de la variación, el genotipado es menos exacto y hay mayor número de variantes genéticas encontradas (255).

Uno de los objetivos claves del proyecto ha sido identificar variantes genéticas incluyendo SNPs poco frecuentes o raros. Por ello, no fue prioritario el genotipado de las muestras pero si determinar el mayor número de variantes genéticas posible. Así, teniendo en cuenta este factor, se realizó la secuenciación exónica de un número elevado de muestras con una cobertura menor, dado que esta estrategia nos proporcionaría resultados de mayor interés (255). Cada base sólo es cubierta por unas pocas lecturas, por lo que es más difícil inferir con precisión el genotipo de un individuo en un sitio en particular (256). Sin embargo, esta aproximación ha permitido encontrar un mayor número de variantes genéticas de interés. De esta forma, la tasa de error puede verse aumentada, pero por otros métodos puede verificarse si los polimorfismos son reales o no, por ejemplo, secuenciación por Sanger.

No obstante, el tipo de secuenciación realizado en este proyecto es de tipo *paired-end* (2x100pb), como se indicó anteriormente. Se trata de un tipo de secuenciación en el que cada segmento corto de ADN obtenido tras la purificación de los exones es leído una vez por cada extremo en una longitud de 100 pb. De esta forma, para cada fragmento

se confirma su posición en el genoma dos veces. Esto ofrece más calidad y precisión en el alineamiento de la secuencia, reduciendo así la tasa de error.

Además, como se ha mostrado en el apartado de resultados, la cobertura media aproximada en nuestro estudio fue de 20x, siendo la cobertura deseada para maximizar la diversidad muestral con los recursos disponibles.

En relación al análisis bioinformático, se ha puesto a punto todo el protocolo o *pipeline* informático necesario para llevar a término el análisis completo del exoma, desde los controles de calidad de las secuencias hasta el llamado de variantes genéticas y la anotación funcional de las mismas. Para ello se han empleado las últimas herramientas bioinformáticas en el campo, descritas en el apartado de material y métodos (Anexo).

5. VARIANTES GENÉTICAS IDENTIFICADAS CON POSIBLE EFECTO FUNCIONAL

El número de variantes genéticas encontradas tras el análisis bioinformático reveló un gran número de variantes genéticas (Veáse Figura 42 y Figura 43). El elevado número de variantes encontrado obligó a priorizar y seleccionar aquellos SNPs que se ajustaban mejor a nuestros objetivos de estudio. Así, tras el filtrado y tratamiento bioinformático de los datos, el estudio se centró en las variantes genéticas con un efecto funcional importante sobre el fenotipo y que cumplían con los criterios establecidos previamente descritos.

Hay que resaltar que el estudio diferencial en los dos grupos de estudio, casos y controles, se basó en el fundamento de que variantes genéticas en los controles podrían ser “protectoras” frente al desarrollo de DM2. Por el contrario, variantes genéticas en casos podrían ser perjudiciales y favorecer el desarrollo de la misma.

5.1. Variantes genéticas en regiones de *splicing*

El proceso de *splicing* es un proceso post-transcripcional de maduración del ARN en el cual como resultado se eliminan ciertos fragmentos de secuencia intrónica. Variantes genéticas en los sitios de *splicing*, pueden influir sobre la síntesis proteica de distintas formas:

- Pérdida del sitio de *splicing*: puede originar la aparición prematura de un codón de stop, la pérdida de un exón o la inclusión de un intrón.
- Reducir la especificidad: puede variar la localización del sitio de *splicing*, lo que origina la inserción o deleción de aminoácidos o la pérdida de la pauta de lectura.
- Transposición del sitio de *splicing*: origina la inserción o deleción de ARN, generando cadenas de ARN más cortas o largas.

El número de variantes de *splicing* identificadas inicialmente en el estudio fue elevado (ver Figura 41 y Figura 42). Sin embargo, nos centramos en estudiar el posible efecto funcional de aquellos genes donde se encontraban variantes situadas en la región donante u aceptora de *splicing* presentes únicamente en alguno de los grupos (Tabla 12 y Tabla 13).

Tras agrupar los genes donde se encontraban las variantes genéticas según diversos criterios (función molecular, proceso biológico y ruta metabólica) se observa globalmente que la mayor parte de ellos participan en procesos como la vía de señalización por Wnt, la cascada de activación del plasminógeno o el estrés oxidativo, entre otros.

La mayor parte de genes donde se han encontrado nuevas variantes genéticas tanto en el grupo de casos como en el de controles no se han descrito previamente relacionados con DM2. Si bien, hay una proporción de genes en los que se ha descrito previamente una asociación con la enfermedad.

Con respecto a las variantes genéticas identificadas en este estudio en el grupo control, una variante genética en un sitio dador de *splicing*, rs41272114, presente en el gen *LPA* (*lipoprotein Lp(a)*) se asoció previamente con una disminución de los niveles de lipoproteína (a) circulante y una disminución del riesgo de enfermedad cardiovascular (257). Elevados niveles de Lp (a) constituyen un factor de riesgo independiente para enfermedad cardiovascular (258). El SNP rs41272114 es un polimorfismo funcional resultado de un *splicing* alternativo que da lugar a una proteína prematura que carece del dominio que normalmente va unido covalentemente a la apolipoproteína B y, por tanto, no es posible formar una Lp (a) madura. El efecto funcional de este SNP, generalmente referido como alelo nulo, ha resultado significativa para los niveles de Lp (a) circulante (259). En relación a la DM2, varios son los estudios que han examinado la

posible relación entre la concentración de Lp (a) y DM2. No obstante, los resultados de diferentes estudios casos-control son controvertidos. En sujetos con diabetes se han encontrado mayores (260, 261), similares (262) o inferiores niveles de Lp (a) que en los individuos control sin diabetes (263). Recientes estudios epidemiológicos en población europea sugieren que los niveles de Lp (a) tiene efectos opuestos en el riesgo de desarrollar enfermedad cardiovascular y diabetes (264, 265).

Una de las variantes genéticas no descritas anteriormente en el grupo control se localiza en el gen *MST1* (*macrophage Stimulating 1*) (ENSG00000173531, 3:49723167). Un estudio reciente publicado en *Nature Medicine* ha propuesto que *MST1* es un regulador crítico de la muerte por apoptosis de las células β y de su función. *MST1* induce específicamente la vía de apoptosis mitocondrial en las células β a través de la regulación de la proteína proapoptótica Bim. Así, *MST1* está potencialmente activada en células β bajo condiciones diabetogénicas; no obstante, la deficiencia de *MST1* completa restaura la normoglucemia, la función de la célula β y la supervivencia *in vitro* e *in vivo*. Por tanto, se ha propuesto *MST1* como diana para el desarrollo de nuevas terapias para la DM2 (266). De este modo, el SNP codificante identificado en este gen podría favorecer a los individuos control (no diabéticos) “protegiéndolos” frente al desarrollo de la enfermedad al generar una proteína *MST1* no activa.

Por otro lado, el *cluster* génico de apolipoproteínas (APOA1/C3/A4/A5) se ha asociado recientemente con triglicéridos y HDL en población no diabética; proponiéndose que variantes genéticas en estos genes podría afectar a la homeostasis lipídica en pacientes con DM2 (267). En nuestro estudio, encontramos un SNP (rs138326449) en un sitio dador de *splicing* presente en el gen *APOC3* (*apolipoprotein C-III*). Recientemente este SNP se ha descrito como una variante rara, asociada a niveles de triglicéridos (P-value=8.0 \times 10⁽⁻⁸⁾) y VLDL en población europea (rs138326449-A, MAF ~0.25%). Este análisis sugirió una relación independiente entre el SNP y los niveles de HDL; estando la presencia del alelo A asociada con una disminución de los niveles de VLDL (268). La presencia de esta variante en nuestra población sana (no diabética) corrobora los resultados de este estudio e indica que la presencia de esta variante rara estaría protegiendo a los individuos del desarrollo de la enfermedad.

En relación al grupo de casos, se encontró una mayor diversidad en cuanto a rutas metabólicas donde se localizaron las variantes genéticas identificadas. En este grupo encontramos una variante genética no descrita previamente localizada en el gen *MASP1*

(*mannan-binding lectin serine peptidase 1*) (ENSG00000127241, 3:186947546). Recientemente, un estudio ha demostrado que los niveles de *MASP1* son significativamente elevados en niños y adultos con diabetes. Además, los niveles de *MASP1* se correlacionan con HbA1c y la disminución de sus niveles mejora el control glucémico (269). Por otro lado, otro estudio mostró que el gen *BTN3A2* (*butyrophilin, subfamily 3, member A2*), perteneciente al complejo HMC-1, juega un papel clave en la regulación inmune y es un potencial gen de susceptibilidad a DM1 (270). En nuestro estudio, encontramos una variante (rs58367598) en el gen *BTN3A2* presente únicamente en individuos diabéticos la cual podría ser una variante de susceptibilidad para el desarrollo de diabetes.

5.2. Variantes genéticas de cambio de sentido o *missense*

Las variantes genéticas de cambio de sentido o *missense* son mutaciones puntuales en las cuales se produce un cambio en la secuencia de ADN que provoca el cambio de significado del codón donde se sitúa la variación, generando así la incorporación a la proteína de un aminoácido diferente al original. Esto puede provocar diferentes cambios a nivel traduccional: una proteína incapaz de cumplir su función, una proteína con una función aumentada, cambio de regulación, modificación de estabilidad, alteración de la degradación, etc. Como ejemplo, mutaciones descritas en un mismo gen, como el *PCSK9* (*proprotein convertase subtilisin/kexin type 9*) pueden causar tanto hipercolesterolemia como hipobetalipoproteinemia. Las mutaciones que producen una ganancia de función cursan con hipercolesterolemias autosómicas dominantes dando lugar a un aumento de la actividad de la proteína, desencadenando un incremento de las concentraciones de colesterol en sangre y un mayor riesgo de enfermedad cardiovascular (271, 272). Sin embargo, mutaciones de pérdida de función en este mismo gen, se asocian a reducciones significativas del colesterol total y de LDL (273).

El número de variantes genéticas de tipo *missense* encontradas es el más elevado entre las variantes con posible efecto funcional estudiadas en este trabajo; siendo de 1.817 SNPs en controles y de 1.614 SNPs en casos. Además, la cantidad de sistemas biológicos en los cuales están localizadas las variantes genéticas y sus genes es muy elevado y, en ambos grupos, existen un gran número de nuevas variantes genéticas localizadas en genes donde no se ha descrito una asociación previa con la enfermedad, haciendo que su análisis sea complejo. Por ello, y teniendo en cuenta la complejidad de

los datos, nos centraremos en aquellos procesos metabólicos más enriquecidos (mayor número de genes implicados) como son aquellos relacionados en el metabolismo proteico, lipídico o de carbohidratos.

En los individuos control, y tras revisión bibliográfica, algunos de los genes identificados se relacionaron previamente con DM2 o parámetros asociados con esta. Como se ha indicado anteriormente, el *cluster* génico de apolipoproteínas (APOA1/C3/A4/A5) se ha asociado con triglicéridos y HDL en población no diabética (267). Así, además de encontrar la variante de *splicing* (rs138326449) descrita anteriormente, también identificamos otra variante *missense* en APOA4 (*apolipoprotein A-IV*) en el grupo control (rs12721043).

Por otro lado, variantes genéticas en genes transportadores de la familia ABC, como el gen transportador de esteroides ABCG8 (*ATP-binding cassette, sub-family G (WHITE), member 8*), se han asociado con factores de riesgo para la DM2 (274, 275). Nosotros encontramos una variante *missense*, rs201817190, en pacientes control la cual podría estar implicada en DM2 al modular factores de susceptibilidad a la enfermedad.

Experimentos *in vivo* también han mostrado asociación entre genes, donde encontramos variantes *missense*, y DM2. PCTP (*phosphatidylcholine transfer protein*) es una proteína de unión a fosfolípidos la cual se expresa en hígado. Los ratones que carecen de esta proteína tienen mejorada la homeostasis de glucosa hepática y son resistentes a la diabetes inducida por dieta (276). Una variante *missense* en este gen, rs147596112, fue encontrada en nuestro estudio y podría dar lugar a una proteína inactiva que favoreciera el metabolismo glucémico en los individuos sanos. Además, se ha demostrado que el gen SLIT1 (*slit guidance ligand 1*) presenta un importante papel en la protección de la célula β frente a procesos de estrés oxidativo y muerte celular así como potenciando la segregación de insulina por estimulación de glucosa *in vivo* (277). En concreto, nuestros resultados identificaron una variante genética *missense* (rs200878365) presente en el gen SLIT1 la cual podría tener una implicación funcional sobre el fenotipo del individuo sano.

Respecto al grupo de pacientes diabéticos, también encontramos una gran cantidad de variantes genéticas que podrían estar asociadas con la enfermedad. Desde un punto de vista generalizado, muchas variantes *missense* se encuentran localizadas en genes que codifican para algún miembro de la superfamilia del citocromo p450, los cuales están implicados en una gran variedad de reacciones, entre ellas, el metabolismo de drogas,

la síntesis de colesterol, esteroides o de otros lípidos. Un ejemplo es el gen *CYP1A1* (*cytochrome P450, family 1, subfamily A, polypeptide 1*), el cual se ha identificado como posible gen candidato de diabetes gestacional (278, 279). Se localiza en el retículo endoplasmático y su expresión se ve inducida por hidrocarburos aromáticos policíclicos algunos de los cuales se encuentran en el humo del tabaco (280). En nuestro estudio identificamos el rs41279188 como variante *missense* en *CYP1A1*. Por otro lado, polimorfismos en el gen *PNPLA3* (*patatin-like phospholipase domain containing 3*) se han asociado también a diabetes gestacional (281); en él identificamos el rs141106484.

Otro de los genes donde detectamos una variación genética *missense* (rs760202) fue en el gen *HADH* (*hydroxyacyl-CoA dehydrogenase*). Anomalías en este gen se ha visto que pueden estar implicadas en el hiperinsulinismo congénito, enfermedad compleja en la que la secreción de insulina por parte de la célula β es inapropiada para los niveles de glucosa en sangre (282, 283). Por otro lado, polimorfismos en el gen *ADAR* (*adenosine deaminase, RNA-specific*) se han asociado a longevidad (284), la cual puede ser modulada por factores metabólicos como la obesidad abdominal o perfil lipídico (285). Nosotros identificamos una variante genética rara (MAF=0,0013), concretamente el rs145588689 en el gen *ADAR*, presente únicamente en diabéticos.

Un reciente trabajo de GWAS, se centró en el estudio de la disfunción del islote pancreático como posible factor de riesgo de DM2. Se realizó un chip con islotes pancreáticos de ratón y se examinó las asociaciones entre SNPs y respuesta aguda de insulina a la glucosa (*acute insulin response to glucose*, AIRg) (286). Se observó asociación entre AIRg y varios SNPs en diversos genes, entre ellos *PAMR1* (*peptidase domain containing associated with muscle regeneration 1*). *PAMR1* codifica para una proteína implicada en la regeneración muscular, en la cual encontramos una variación genética *missense* (rs143549343).

Otra observación interesante en los resultados obtenidos es que hubo diferencias en cuanto al número de variantes genéticas con posible efecto funcional por gen. En el grupo control encontramos múltiples variantes genéticas *missense* presentes en un mismo gen; dos variantes (*TARS2*, *ABCA4*, *FOXD4L1*, *TTN*, etc), tres (*PDE4DIP*, *ARHGEF11*, *OR10T2*, *INADL*, etc), seis (*DNAH7*), siete (*SYNE2*) e incluso nueve (*DST*). De similar modo en el grupo de diabéticos encontramos dos variantes presentes en un gen (*PTPN22*, *RPTN*, *SPTA1*, *CR1L*, etc.), tres (*GJB4*, *GPR113*, *TET2*, *FAT*, etc.) o seis (*DNAH3*).

Así, variantes en el gen *ARHGEF11* (*Rho guanine nucleotide exchange factor (GEF) 11*) se han asociado a resistencia a insulina y DM2 en indios Pima (287). Variantes genéticas en el gen *PTPN22* (*protein tyrosine phosphatase, non-receptor type 22 (lymphoid)*) se han asociado con DM1 y vitiligo (288, 289). Éste último es un desorden autoinmune que se puede encontrar aumentado en pacientes con DM, específicamente, tipo 1, y en forma menos frecuente, en tipo 2 (290).

Por otro lado, tanto en casos como en controles, se observó un gran número de variantes *missense* (diferentes entre grupos) que se concentran dentro de familias génicas. Un ejemplo es la protocadherina, la cual incluye una gran variedad de genes donde encontramos este tipo de variantes (*PCDHA4, PCDHA5, PCDHA7-9, PCDHB2, PCDHB8, PCDHB10-16*, etc). Otro caso es el gen *OBSCN* (*obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF*) el cual contiene 80 exones y codifica para una proteína de aproximadamente 780 KDa. En este gen encontramos cuatro variantes genéticas *missense* en controles (rs150312546, rs188302055, rs145597580 y rs35186354) y cinco en casos (rs191837710, rs200638282, rs56218706, rs72762066 y rs377038292).

Así pues, la mayoría de los genes identificados, tanto en controles como en diabéticos, no se han asociado previamente a DM2 ni a parámetros relacionados con esta. Sin embargo, la presencia de un número elevado de variantes genéticas con posible efecto funcional en un mismo gen hace plantearnos la importancia del posible estudio completo de estos genes. Ello nos ayudaría a entender mejor el componente genético de la DM2, pudiendo establecer nuevas dianas terapéuticas de la enfermedad.

5.3. Variantes genéticas en regiones codificantes para microARNs maduros

Como se ha remarcado en la introducción, la secreción adecuada de insulina por parte de las células β es necesaria para mantener los niveles de glucosa en sangre dentro del rango fisiológico. Durante la DM2, la secreción de insulina es insuficiente para normalizar los niveles de glucosa y eventualmente la masa de células β se reduce y la célula β muere, tanto en modelos animales como en humanos (291, 292).

Los microARNs son ARNs de aproximadamente 22 nucleótidos de longitud que regulan la expresión de genes al inducir la degradación y/o inhibición translacional

pudiendo cada microARN modular cientos de mARNs diana. Principalmente, los microARNs regulan la expresión de sus genes diana al unirse a su región 3' UTR a través de una pequeña secuencia de unos 7 nucleótidos (293). Así, los microARNs son fundamentales para la supervivencia de la célula dado que participan en procesos biológicos, estando su desregulación o expresión anormal relacionado con numerosas enfermedades como el cáncer, enfermedades neurológicas o diabetes (294). Existen estudios realizados *in vivo* e *in vitro* que muestran que los microARNs intervienen en procesos como el desarrollo del páncreas, la regulación de la glucosa, el control de las células β pancreáticas maduras y la diferenciación de islotes pancreáticos. Además, participan en la regulación de la producción, secreción y acción de la insulina (295-297).

Un ejemplo reciente son los resultados del grupo de investigación del Dr. Stoffel donde se ha demostrado que la familia de microARN-200 ejerce un papel crucial *in vivo* en la regulación endógena de la apoptosis de la célula β asociada a la DM2, proponiendo la inhibición de la actividad de los microARN-200 como una posible estrategia farmacológica para promover la supervivencia de la célula β (298).

Por todo ello, debido a la posible importancia de los mecanismos epigenéticos en el desarrollo de enfermedades complejas como la DM2, en este trabajo también nos centramos en analizar las variantes genéticas presentes en la secuencia codificadora de microARNs. En comparación con otros tipos de variantes genéticas con posible efecto funcional, el número de variantes encontradas fue mucho más bajo; siendo de 23 y 11 SNPs en secuencias codificantes para microARNs en controles y casos, respectivamente. Desafortunadamente, en este estudio no pudimos abarcar el análisis de variantes genéticas en las regiones 3' UTR (regiones de unión de los microARNs) dado que el número de variantes genéticas identificadas inicialmente fue demasiado elevado (57.634 SNP en controles y 52.920 en casos). En relación a los microARNs donde hemos localizado las variantes genéticas, la mayor parte de ellos no se han descrito previamente en la bibliografía y su función reguladora es desconocida.

En los pacientes diabéticos, encontramos diferentes variantes genéticas en microARNs, entre ellos el microARN-3675, microARN-933 y microARN-628. Un estudio de *deep sequencing* y bioinformático identificó el microARN-3675 como un microARN nuevo y funcional. Ensayos funcionales de 3' UTR luciferasa sugirieron además que este microARN regulaba negativamente genes como el CNTN3 (*contactin 3*) (299). Así, usando herramientas bioinformáticas (*TargetScan* y *miRanda*) encontramos que el

microARN-3675 tiene genes diana implicados en obesidad y diabetes como son el *HMGCS2* (*3-hydroxy-3-methylglutaryl-CoA synthase 2*) o el *INSIG1* (*insulin induced gene 1*) (300, 301).

Además, un trabajo reciente ha descrito que el rs79402775 presente en el microARN-933 maduro se asocia con el riesgo de desarrollar tumor tiroideo en población china (AA versus GG/GA, OR = 1.76, 1.00-3.12) (302). Este es un ejemplo de cómo variantes en los microARNs pueden contribuir a la susceptibilidad genética de enfermedades.

Otros microARNs como el microARN-628 se ha descrito asociado a respuesta inmune (303), como posible antiviral terapéutico de enfermedades respiratorias (304) o como biomarcador de infección vírica (305). En nuestro estudio, encontramos una variante genética no descrita anteriormente en la secuencia del microARN-628 maduro presente en diabéticos y no controles. Mediante la combinación de herramientas de predicción de genes diana para microARNs (*TargetScan* y *miRanda*) podemos concluir que los sitios de unión (3' UTR) del microARN-628 se encuentran en varios genes implicados en la homeostasis de glucosa y el metabolismo lipídico, incluyendo: *PFKFB2* (*6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2*), *GSK3B* (*glycogen synthase kinase 3 beta*) o *PPP1R3D* (*protein phosphatase 1, regulatory subunit 3D*).

Por todo ello, la alteración de la secuencia de los microARNs, la cual puede ser debida a la presencia de una determinada variante genética, puede afectar al desarrollo de una enfermedad, como la DM2. Ello convierte a estas moléculas reguladoras en potenciales biomarcadores para el diagnóstico clínico y posibles objetivos terapéuticos (306, 307).

No obstante, experimentos de validación y funcionales serían necesarios para comprobar la veracidad de los resultados de estos recientes estudios. Así pues, una estrategia de trabajo futura sería el diseño de experimentos funcionales que permitiesen identificar los mecanismos moleculares por los cuales estos microARNs regulan estos genes y su papel en la regulación de genes diana implicados en enfermedades cardiovasculares como la DM2 (308-310).

6. VARIANTES GENÉTICAS IDENTIFICADAS Y VERIFICADAS CON EFECTO FUNCIONAL: VARIANTES DE GANANCIA O PÉRDIDA DE UN CODÓN DE STOP

La variante genética de codón de stop, parada o sin sentido es un tipo de mutación puntual en la secuencia de ADN que provoca la aparición de un codón de terminación prematuro que conduce a su vez a la generación de un producto proteico truncado, incompleto y por lo general no funcional. Por ello, este tipo de variantes suelen tener un efecto funcional más drástico sobre el fenotipo y de ahí, su importancia en este estudio. El código genético cuenta únicamente con 3 codones de parada o stop (UAA, UAG y UGA), mientras que 61 son codones con sentido.

Como se mostró en el apartado de resultados, tras el análisis de los datos se obtuvieron las variantes genéticas de ganancia y pérdida de codón de stop, siendo las del primer tipo las más abundantes. A continuación, debido a su relevancia funcional, se verificaron mediante secuenciación por Sanger todas las variantes encontradas tanto en casos como en controles. El motivo de esta comprobación tuvo como objetivo principal confirmar qué las variantes genéticas encontradas eran ciertas y qué porcentaje de error teníamos mediante NGS. Este proceso implicó el diseño manual de los oligonucleótidos para cada una de las variantes genéticas encontradas, su amplificación mediante PCR, su posterior secuenciación y finalmente su verificación.

Tras verificar las mutaciones, concluimos que el 85% del total de variantes identificadas en casos y controles, eran válidas. El porcentaje de error restante pudo ser debido a varios factores, entre ellos, la calidad de las secuencias, la baja cobertura o errores de la propia técnica. Finalmente, el número de variantes genéticas de codón de parada reales fueron 102 SNPs en controles y 50 en diabéticos.

La mayor parte de las variantes genéticas encontradas y verificadas para cada uno de los genes no se han descrito en la bibliografía científica previamente. Además, ninguno de los polimorfismos poco frecuentes y variantes raras encontradas se han asociado previamente a la DM2 y/o parámetros asociados con esta. No obstante, algunas de estas variantes raras se han asociado a otras patologías como cáncer, trastornos psiquiátricos o enfermedades degenerativas.

En el grupo control, el rs11571833 en *BRCA2* (*breast cancer 2*) se asoció con cáncer de mama y pulmón (311-313); el rs5030839 en *NAT1* (*N-acetyltransferase 1*) se

relacionó con pólipos colorrectales (314), el rs4987208 en *RAD52* (*RAD52 homolog, DNA repair protein*) se asoció previamente a cáncer ovárico (315, 316), el rs203981 en *IFNE* (*interferon, epsilon*) con hemorragia intracerebral y vitíligo (317, 318), y el rs1861050 en *CC2D2A* (*coiled-coil and C2 domain containing 2*) con enfermedad psiquiátrica (319). En el caso de las variantes genéticas identificadas en los pacientes diabéticos, encontramos dos variantes raras previamente descritas. El rs12240276 en *AKR1E2* (*aldo-keto reductase family 1, member E2*) asociada a Alzheimer (320) y el rs61757718 en *PIK3C2G* (*phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 gamma*) asociado con colon irritable (321).

Por otro lado, cabe señalar que los genes donde identificamos estas variantes genéticas de codón de stop están implicados en una gran diversidad de rutas metabólicas potencialmente relacionadas con la DM2, entre ellas: la vía de señalización de Wnt, la cascada de señalización de la insulina, la vía de señalización PI3K o la vía de síntesis de colesterol.

En el grupo control encontramos una variación de stop en el gen *PPP1R12A* (*protein phosphatase 1, regulatory subunit 12*) no descrita previamente (ENSG00000058272, 12:80169711). Este gen está implicado en la señalización de insulina en el músculo esquelético (322). La proteína Rho, a través de la acción de la Rho-quinasa, parece inhibir esta proteína codificada por el gen *PPP1R12A*. Se ha descrito que la actividad Rho-quinasa está implicada en la disfunción endotelial en DM1 al promover el estrés oxidativo. Además, nuevos experimentos sugieren a los inhibidores de Rho quinasa como nuevos agentes terapéuticos para el tratamiento de la nefropatía diabética (323).

En los pacientes diabéticos encontramos también diferentes variantes genéticas en genes implicados en la enfermedad o en parámetros relacionados, entre ellos: *ACAD10*, *PIK3C2G*, *CYP1A1* o *PHTF1*.

En nuestro estudio localizamos el rs146335507 en el gen *ACAD10* (*acyl-CoA dehydrogenase family, member 10*) presente en el grupo de diabéticos. *ACAD10* codifica para una proteína implicada en la oxidación de ácidos grasos a nivel mitocondrial. No obstante, problemas en la oxidación de los ácidos grasos induce resistencia a insulina en hígado y músculo esquelético al aumentar los niveles de lípidos a nivel intracelular (324). Estudios previos mostraron que SNPs comunes en este gen se han asociado a DM2, resistencia a insulina y oxidación lipídica. Un estudio de GWAS

describió una asociación entre este gen y DM2 en indios americanos (325). Concretamente, encontraron un SNP común, rs632650, el cual se localizaba en el intrón 2 de *ACAD10* y presentaba asociación con DM2. Posteriormente, 23 SNPs fueron genotipados en 1.500 indios americanos y se mostró que SNPs, como el rs601663 o el rs659964 en este gen, se asociaban también con DM2 (326).

Otro de los genes donde encontramos una variante de stop poco frecuente (rs61757718) fue en *PIK3C2G* (*phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 gamma*). La proteína codificada por este gen forma parte de la familia PI3K, implicada en numerosas vías de señalización como la proliferación celular, transformación oncogénica, supervivencia celular, migración o tráfico proteico intracelular. Este gen tiene un importante papel en diversas enfermedades, entre ellas, la DM2. Un estudio previo identificó varios polimorfismos comunes en este gen asociados con DM2 en población japonesa (327).

En el gen *CYP1A1*, además de encontrar una variación *missense* descrita anteriormente, también encontramos una variante rara de stop no descrita anteriormente (ENSG00000140465, 15:75015234). Por otro lado, la presencia de diferentes variantes genéticas con posible efecto funcional para un mismo gen en los pacientes diabéticos, sugiere que estudios futuros de estos genes podrían ser clave en la búsqueda del componente genético de la enfermedad.

Finalmente, en el gen *PHTF1* (*putative homeodomain transcription factor 1*) se identificó el rs200552382. Polimorfismos comunes en este gen se han asociado con DM1 pero no hay estudios de asociación para DM2 (328, 329).

Por todo lo expuesto, la presencia de variantes genéticas poco frecuentes y raras con un impacto funcional fuerte como son las variantes de codón de stop, pueden estar detrás de parte de la heredabilidad no descrita hasta el momento de la DM2. La validación de estas variantes, su replicación en una población mayor y el estudio completo de los genes, pueden ser claves para desgranar parte del componente hereditario de la DM2.

7. INNOVACIÓN, LIMITACIONES Y PERSPECTIVAS FUTURAS DEL ESTUDIO

Hasta la fecha no se han realizado estudios cuyo objetivo principal sea la identificación de PPFs y variantes raras presentes en el exoma asociadas a la DM2 incluyendo población española. Por ello, nuestro trabajo representa un estudio novedoso en cuanto a técnica empleada y población a la que se dirige.

En relación a las variantes genéticas identificadas en el estudio, los PPFs y variantes raras en la población no han sido identificados en su totalidad debido principalmente, al número de muestras que se pueden abordar con la secuenciación de exoma y a la cobertura limitada que se ha podido alcanzar. No obstante, el número de muestras incluido ha sido suficiente para identificar prácticamente todos los PFs y más del 90% de PPFs presentes en el exoma (330). En el caso de la identificación de variantes raras (MAF <0,01) la probabilidad es más baja aunque los resultados muestran que hemos sido capaces de identificar un gran número de ellas.

El elevado número de variantes genéticas encontradas nos ha conducido a seleccionar aquellas con mayor impacto funcional. Debido al coste y tiempo que supone la verificación de las variantes, sólo han podido ser verificadas las variantes de codón de stop. Sin embargo, la validación de las variantes funcionales encontradas en las muestras analizadas se llevará acabo próximamente con otra de las plataformas de secuenciación de Illumina (MiSeq). Tras la validación, el diseño de experimentos funcionales futuros serán claves para determinar la funcionalidad biológica de estas nuevas variantes genéticas encontradas en el contexto de la DM2.

El poder estadístico para detectar asociaciones es limitado debido a la baja frecuencia de las variantes encontradas. No obstante, la posibilidad de replicación futura de los resultados obtenidos en una población de gran tamaño muestral (>10.000 muestras de población general, y 1.000 casos y 2.000 controles) permitirá confirmar nuestros resultados y establecer asociaciones estadísticas entre variantes y enfermedad.

La NGS está en continuo avance y mejora, lo que nos puede ayudar en el futuro a realizar nuevos estudios que nos permitan confirmar estos datos e identificar nuevas variantes implicadas en el desarrollo de diabetes o en la protección frente a esta. En el caso de la secuenciación de exoma los procedimientos son cada vez más eficaces y

económicos. El kit de enriquecimiento empleado posee una eficiencia de captura de regiones exónicas de alrededor del 65%; limitación técnica e inevitable. No obstante, este kit abarca la captura no sólo de la parte exónica sino de los extremos intrónicos que la rodean, permitiendo así también encontrar variantes de posible interés en estas zonas intergénicas.

La gran cantidad de datos generados ha hecho necesario el desarrollo de una plataforma bioinformática de análisis y el diseño de un protocolo o *pipeline* específico para este proyecto. Algunas de las variantes puede que no se hayan detectado al no superar los filtros de análisis establecidos, principalmente debido a que el número de lecturas y/o la calidad de la secuencia fueran bajos. Además, debido a la baja cobertura de secuenciación, no ha sido posible calcular correctamente el MAF de los genotipos pero si hacer una aproximación de ellos. Sin embargo, y como ya se indicó, este estudio busca identificar nuevas variantes genéticas implicadas en DM2 y no el genotipado de las muestras. Por ello, se decidió establecer una cobertura menor a cambio de secuenciar más número de muestras.

Finalmente, el número de PPFs y variantes raras es desconocido en su inmensa mayoría, aunque esta cifra va incrementándose con los resultados de los grandes proyectos de secuenciación total del genoma. Este trabajo supone un gran avance en la identificación de nuevas variantes genéticas en relación a la DM2, así como en el desarrollo de herramientas bioinformáticas necesarias para su análisis. Si bien hay que tener en cuenta que el efecto y la presencia de estos SNPs pueden estar modulados por factores ambientales específicos de nuestra población. La información obtenida abre nuevas posibilidades para entender esta enfermedad compleja, su prevención, el diseño de nuevos tratamientos y el desarrollo de nuevas áreas de investigación (331).

Conclusiones/Conclusions

1. La secuenciación del exoma es una herramienta útil para el estudio de enfermedades de herencia compleja como la diabetes mellitus de tipo 2, permitiendo la identificación de numerosas variantes genéticas potencialmente implicadas en el desarrollo de la enfermedad o en la protección frente a la misma.
2. La secuenciación de exoma ha permitido identificar tanto variantes genéticas previamente descritas, como nuevas; especialmente variantes poco frecuentes y raras, muchas de ellas potencialmente relacionadas con la diabetes mellitus de tipo 2.
3. El elevado número de nuevas variantes genéticas encontradas podría explicar parte del componente genético de la diabetes mellitus de tipo 2. Este hecho se apoya en que una parte relevante de las variantes encontradas están en genes previamente relacionados con diabetes y/o el metabolismo implicado en su desarrollo o en aspectos protectores frente a esta enfermedad.
4. Las nuevas variantes genéticas encontradas en este estudio abren nuevas líneas de investigación en el estudio de la diabetes mellitus de tipo 2 a diferentes niveles, contemplando rutas metabólicas y vías de señalización, entre otras.
5. Es necesario la confirmación de estos hallazgos con la validación y replicación de los resultados encontrados en una población mayor, así como el diseño de experimentos funcionales.

1. Exome sequencing is a useful tool for the study of complex inheritable diseases as type 2 diabetes mellitus, allowing the identification of potential genetic variants involved in the development of the disease or in the protection from it.
2. Exome sequencing has identified either genetic variants previously described as new, especially low-frequency and rare variants, most of them potentially related with type 2 diabetes mellitus.
3. The high amount of new genetic variants that we found could explain a part of the genetic component of type 2 diabetes mellitus. This fact is supported by a significant part of variants have been found in genes previously associated with diabetes and/or metabolism, being involved in its development or protective aspects to this disease.
4. The new genetic variants found in this study open new research lines into the genetic study of type 2 diabetes mellitus at different levels, as metabolic and signalling pathways, among others.
5. Confirmation of these findings with validation and replication in a larger population is needed, as well as the design of functional experiments.

Bibliografía

1. Leon BM, Maddox TM. Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research. *World journal of diabetes*. 2015;6(13):1246-58.
2. American Diabetes Association. Standards of medical care in diabetes--2014. *Diabetes care*. 2014;37 Suppl 1:S14-80.
3. Force IDFCGT. Global Guideline for Type 2 Diabetes: recommendations for standard, comprehensive, and minimal care. *Diabetic medicine : a journal of the British Diabetic Association*. 2006;23(6):579-93.
4. Forouhi NG, Wareham NJ. Epidemiology of diabetes. *Medicine*. 2014;42(12):698-702.
5. Diabetes FID. *Diabetes Atlas*. 2013;Sexta Edición.
6. Soriguer F, Goday A, Bosch-Comas A, Bordiu 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;55(1):88-93.
7. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation. WHO Guidelines Approved by the Guidelines Review Committee. Geneva2011.
8. Gillett MJ. International Expert Committee report on the role of the A1c assay in the diagnosis of diabetes: *Diabetes Care* 2009; 32(7): 1327-1334. *The Clinical biochemist Reviews / Australian Association of Clinical Biochemists*. 2009;30(4):197-200.
9. Jeppsson JO, Kobold U, Barr J, Finke A, Hoelzel W, Hoshino T, et al. Approved IFCC reference method for the measurement of HbA1c in human blood. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 2002;40(1):78-89.
10. Cowie CC, Rust KF, Byrd-Holt DD, Gregg EW, Ford ES, Geiss LS, et al. Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988-2006. *Diabetes care*. 2010;33(3):562-8.
11. Ziemer DC, Kolm P, Weintraub WS, Vaccarino V, Rhee MK, Twombly JG, et al. Glucose-independent, black-white differences in hemoglobin A1c levels: a cross-sectional analysis of 2 studies. *Annals of internal medicine*. 2010;152(12):770-7.
12. American Diabetes Association. Standards of medical care in diabetes-2013. *Diabetes care*. 2013;36 Suppl 1:S11-66.
13. American Diabetes Association. Classification and diagnosis of diabetes. *Diabetes care*. 2015;38 Suppl:S8-S16.

14. Thomas CC, Philipson LH. Update on diabetes classification. *The Medical clinics of North America*. 2015;99(1):1-16.
15. Conde Barreiro S, Rodriguez Rigual M, Bueno Lozano G, Lopez Siguero JP, Gonzalez Pelegrin B, Rodrigo Val MP, et al. [Epidemiology of type 1 diabetes mellitus in children in Spain]. *Anales de pediatria*. 2014;81(3):189 e1- e12.
16. Schmidt KD, Valeri C, Leslie RD. Autoantibodies in Type 1 diabetes. *Clinica chimica acta; international journal of clinical chemistry*. 2005;354(1-2):35-40.
17. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes care*. 1997;20(7):1183-97.
18. Imagawa A, Hanafusa T, Miyagawa J, Matsuzawa Y. A novel subtype of type 1 diabetes mellitus characterized by a rapid onset and an absence of diabetes-related antibodies. Osaka IDDM Study Group. *The New England journal of medicine*. 2000;342(5):301-7.
19. Pinero-Pilona A, Litonjua P, Aviles-Santa L, Raskin P. Idiopathic type 1 diabetes in Dallas, Texas: a 5-year experience. *Diabetes care*. 2001;24(6):1014-8.
20. Leslie RD, Pozzilli P. Type I diabetes masquerading as type II diabetes. Possible implications for prevention and treatment. *Diabetes care*. 1994;17(10):1214-9.
21. Tuomi T, Carlsson A, Li H, Isomaa B, Miettinen A, Nilsson A, et al. Clinical and genetic characteristics of type 2 diabetes with and without GAD antibodies. *Diabetes*. 1999;48(1):150-7.
22. Stenstrom G, Gottsater A, Bakhtadze E, Berger B, Sundkvist G. Latent autoimmune diabetes in adults: definition, prevalence, beta-cell function, and treatment. *Diabetes*. 2005;54 Suppl 2:S68-72.
23. Gabbe SG. Definition, detection, and management of gestational diabetes. *Obstetrics and gynecology*. 1986;67(1):121-5.
24. Di Cianni G, Miccoli R, Volpe L, Lencioni C, Del Prato S. Intermediate metabolism in normal pregnancy and in gestational diabetes. *Diabetes/metabolism research and reviews*. 2003;19(4):259-70.
25. Karnik SK, Chen H, McLean GW, Heit JJ, Gu X, Zhang AY, et al. Menin controls growth of pancreatic beta-cells in pregnant mice and promotes gestational diabetes mellitus. *Science*. 2007;318(5851):806-9.
26. Kim H, Toyofuku Y, Lynn FC, Chak E, Uchida T, Mizukami H, et al. Serotonin regulates pancreatic beta cell mass during pregnancy. *Nature medicine*. 2010;16(7):804-8.

27. Virjee S, Robinson S, Johnston DG. Screening for diabetes in pregnancy. *Journal of the Royal Society of Medicine*. 2001;94(10):502-9.
28. Dabelea D, Snell-Bergeon JK, Hartsfield CL, Bischoff KJ, Hamman RF, McDuffie RS, et al. Increasing prevalence of gestational diabetes mellitus (GDM) over time and by birth cohort: Kaiser Permanente of Colorado GDM Screening Program. *Diabetes care*. 2005;28(3):579-84.
29. Gorgojo Martínez JJ ARF, López Hernández E, Donnay Candil S. . Incidencia de la diabetes mellitus gestacional según distintos criterios diagnósticos en la zona suroeste de Madrid. Influencia del diagnóstico sobre los parámetros materno-fetales. *Rev Clin Esp*. 2002;202:136-41.
30. Barquiel B, Herranz L, Hillman N, Burgos MA, Pallardo LF. Prepregnancy body mass index and prenatal fasting glucose are effective predictors of early postpartum metabolic syndrome in Spanish mothers with gestational diabetes. *Metabolic syndrome and related disorders*. 2014;12(9):457-63.
31. Gillman MW, Rifas-Shiman S, Berkey CS, Field AE, Colditz GA. Maternal gestational diabetes, birth weight, and adolescent obesity. *Pediatrics*. 2003;111(3):e221-6.
32. Retnakaran R, Shah BR. Sex of the baby and future maternal risk of Type 2 diabetes in women who had gestational diabetes. *Diabetic medicine: a journal of the British Diabetic Association*. 2015.
33. Babenko AP, Polak M, Cave H, Busiah K, Czernichow P, Scharfmann R, et al. Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *The New England journal of medicine*. 2006;355(5):456-66.
34. Flanagan SE, Patch AM, Mackay DJ, Edghill EL, Gloyn AL, Robinson D, et al. Mutations in ATP-sensitive K⁺ channel genes cause transient neonatal diabetes and permanent diabetes in childhood or adulthood. *Diabetes*. 2007;56(7):1930-7.
35. Pearson ER, Flechtner I, Njolstad PR, Malecki MT, Flanagan SE, Larkin B, et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *The New England journal of medicine*. 2006;355(5):467-77.
36. De Leon DD, Stanley CA. Permanent Neonatal Diabetes Mellitus. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al., editors. *GeneReviews(R)*. Seattle (WA)1993.

37. Armentrout D. Neonatal diabetes mellitus. *Journal of pediatric health care : official publication of National Association of Pediatric Nurse Associates & Practitioners.* 1995;9(2):75-8.
38. Bonnefond A, Froguel P, Vaxillaire M. The emerging genetics of type 2 diabetes. *Trends in molecular medicine.* 2010;16(9):407-16.
39. Fajans SS, Bell GI. MODY: history, genetics, pathophysiology, and clinical decision making. *Diabetes care.* 2011;34(8):1878-84.
40. Molven A, Njolstad PR. Role of molecular genetics in transforming diagnosis of diabetes mellitus. *Expert review of molecular diagnostics.* 2011;11(3):313-20.
41. Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia.* 2010;53(12):2504-8.
42. Gloyn AL. Glucokinase (GCK) mutations in hyper- and hypoglycemia: maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemia of infancy. *Human mutation.* 2003;22(5):353-62.
43. Raeder H, Johansson S, Holm PI, Haldorsen IS, Mas E, Sbarra V, et al. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. *Nature genetics.* 2006;38(1):54-62.
44. Frayling TM, Evans JC, Bulman MP, Pearson E, Allen L, Owen K, et al. beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. *Diabetes.* 2001;50 Suppl 1:S94-100.
45. Bowman P, Flanagan SE, Edghill EL, Damhuis A, Shepherd MH, Paisey R, et al. Heterozygous ABCC8 mutations are a cause of MODY. *Diabetologia.* 2012;55(1):123-7.
46. Stride A, Shepherd M, Frayling TM, Bulman MP, Ellard S, Hattersley AT. Intrauterine hyperglycemia is associated with an earlier diagnosis of diabetes in HNF-1alpha gene mutation carriers. *Diabetes care.* 2002;25(12):2287-91.
47. Stride A, Vaxillaire M, Tuomi T, Barbetti F, Njolstad PR, Hansen T, et al. The genetic abnormality in the beta cell determines the response to an oral glucose load. *Diabetologia.* 2002;45(3):427-35.
48. Hoptasz M, Szczucinski A, Losy J. Heterogeneous phenotypic manifestations of maternally inherited deafness associated with the mitochondrial A3243G mutation. Case report. *Neurologia i neurochirurgia polska.* 2014;48(2):150-3.

49. Maassen JA, LM TH, Van Essen E, Heine RJ, Nijpels G, Jahangir Tafrechi RS, et al. Mitochondrial diabetes: molecular mechanisms and clinical presentation. *Diabetes*. 2004;53 Suppl 1:S103-9.
50. Fischel-Ghodsian N. Mitochondrial DNA mutations and diabetes: another step toward individualized medicine. *Annals of internal medicine*. 2001;134(9 Pt 1):777-9.
51. Pilz D, Quarrell OW, Jones EW. Mitochondrial mutation commonly associated with Leber's hereditary optic neuropathy observed in a patient with Wolfram syndrome (DIDMOAD). *Journal of medical genetics*. 1994;31(4):328-30.
52. Barrett TG. Mitochondrial diabetes, DIDMOAD and other inherited diabetes syndromes. *Best practice & research Clinical endocrinology & metabolism*. 2001;15(3):325-43.
53. Toouli J, Biankin AV, Oliver MR, Pearce CB, Wilson JS, Wray NH, et al. Management of pancreatic exocrine insufficiency: Australasian Pancreatic Club recommendations. *The Medical journal of Australia*. 2010;193(8):461-7.
54. Hardt PD, Ewald N. Exocrine pancreatic insufficiency in diabetes mellitus: a complication of diabetic neuropathy or a different type of diabetes? *Experimental diabetes research*. 2011;2011:761950.
55. Lombardi G, Galdiero M, Auriemma RS, Pivonello R, Colao A. Acromegaly and the cardiovascular system. *Neuroendocrinology*. 2006;83(3-4):211-7.
56. Prague JK, May S, Whitelaw BC. Cushing's syndrome. *Bmj*. 2013;346:f945.
57. Corrao G, Ibrahim B, Nicotra F, Soranna D, Merlino L, Catapano AL, et al. Statins and the risk of diabetes: evidence from a large population-based cohort study. *Diabetes care*. 2014;37(8):2225-32.
58. Ishikawa M, Okajima F, Inoue N, Motomura K, Kato T, Takahashi A, et al. Distinct effects of pravastatin, atorvastatin, and simvastatin on insulin secretion from a beta-cell line, MIN6 cells. *Journal of atherosclerosis and thrombosis*. 2006;13(6):329-35.
59. Gale EA. Congenital rubella: citation virus or viral cause of type 1 diabetes? *Diabetologia*. 2008;51(9):1559-66.
60. Karjalainen J, Knip M, Hyoty H, Leinikki P, Ilonen J, Kaar ML, et al. Relationship between serum insulin autoantibodies, islet cell antibodies and Coxsackie-B4 and mumps virus-specific antibodies at the clinical manifestation of type 1 (insulin-dependent) diabetes. *Diabetologia*. 1988;31(3):146-52.

61. King ML, Shaikh A, Bidwell D, Voller A, Banatvala JE. Coxsackie-B-virus-specific IgM responses in children with insulin-dependent (juvenile-onset; type I) diabetes mellitus. *Lancet*. 1983;1(8339):1397-9.
62. Pak CY, Eun HM, McArthur RG, Yoon JW. Association of cytomegalovirus infection with autoimmune type 1 diabetes. *Lancet*. 1988;2(8601):1-4.
63. Kalra S, Kalra B, Agrawal N, Unnikrishnan A. Understanding diabetes in patients with HIV/AIDS. *Diabetology & metabolic syndrome*. 2011;3(1):2.
64. Boronat M, Varillas VF, Saavedra P, Suarez V, Bosch E, Carrillo A, et al. Diabetes mellitus and impaired glucose regulation in the Canary Islands (Spain): prevalence and associated factors in the adult population of Telde, Gran Canaria. *Diabetic medicine : a journal of the British Diabetic Association*. 2006;23(2):148-55.
65. Botas P, Delgado E, Castano G, Diaz de Grenu C, Prieto J, Diaz-Cadorniga FJ. Comparison of the diagnostic criteria for diabetes mellitus, WHO-1985, ADA-1997 and WHO-1999 in the adult population of Asturias (Spain). *Diabetic medicine : a journal of the British Diabetic Association*. 2003;20(11):904-8.
66. Lorenzo C, Serrano-Rios M, Martinez-Larrad MT, Gabriel R, Williams K, Gonzalez-Villalpando C, et al. Was the historic contribution of Spain to the Mexican gene pool partially responsible for the higher prevalence of type 2 diabetes in mexican-origin populations? The Spanish Insulin Resistance Study Group, the San Antonio Heart Study, and the Mexico City Diabetes Study. *Diabetes care*. 2001;24(12):2059-64.
67. Sergio Valdés GR-MyFS. Evolución de la prevalencia de la diabetes tipo 2 en población adulta española. *Med Clin (Barc)*. 2007;129(15):599. .
68. Ruiz-Ramos M, Escolar-Pujolar A, Mayoral-Sanchez E, Corral-San Laureano F, Fernandez-Fernandez I. [Diabetes mellitus in Spain: death rates, prevalence, impact, costs and inequalities]. *Gaceta sanitaria / SESPAS*. 2006;20 Suppl 1:15-24.
69. Gil-Montalban E, Martin-Rios MD, Ortiz-Marron H, Zorrilla-Torras B, Martinez-Cortes M, Esteban-Vasallo MD, et al. Incidence of type 2 diabetes and associated factors in the adult population of the Community of Madrid. PREDIMERC cohort. *Rev Clin Esp*. 2015.
70. Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98(5 Suppl):786-806.

71. Mohamed Q, Gillies MC, Wong TY. Management of diabetic retinopathy: a systematic review. *Jama*. 2007;298(8):902-16.
72. Adler AI, Stevens RJ, Manley SE, Bilous RW, Cull CA, Holman RR, et al. Development and progression of nephropathy in type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS 64). *Kidney international*. 2003;63(1):225-32.
73. Parving HH, Lewis JB, Ravid M, Remuzzi G, Hunsicker LG, investigators D. Prevalence and risk factors for microalbuminuria in a referred cohort of type II diabetic patients: a global perspective. *Kidney international*. 2006;69(11):2057-63.
74. Boulton AJ, Vinik AI, Arezzo JC, Bril V, Feldman EL, Freeman R, et al. Diabetic neuropathies: a statement by the American Diabetes Association. *Diabetes care*. 2005;28(4):956-62.
75. O'Keefe JH, Jr., Miles JM, Harris WH, Moe RM, McCallister BD. Improving the adverse cardiovascular prognosis of type 2 diabetes. *Mayo Clinic proceedings*. 1999;74(2):171-80.
76. Yudkin JS, Blauth C, Drury P, Fuller J, Henley J, Lancaster T, et al. Prevention and management of cardiovascular disease in patients with diabetes mellitus: an evidence base. *Diabetic medicine : a journal of the British Diabetic Association*. 1996;13(9 Suppl 4):S101-21.
77. Vijan S. In the clinic. Type 2 diabetes. *Annals of internal medicine*. 2015;162(5):ITC1-16.
78. Diabetes Prevention Program Research G, Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet*. 2009;374(9702):1677-86.
79. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England journal of medicine*. 2002;346(6):393-403.
80. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes care*. 1997;20(4):537-44.
81. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *The New England journal of medicine*. 2001;344(18):1343-50.

82. Carey VJ, Walters EE, Colditz GA, Solomon CG, Willett WC, Rosner BA, et al. Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women. The Nurses' Health Study. *American journal of epidemiology*. 1997;145(7):614-9.
83. DeFronzo RA, Ferrannini E. Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. *Diabetes care*. 1991;14(3):173-94.
84. Maggio CA, Pi-Sunyer FX. The prevention and treatment of obesity. Application to type 2 diabetes. *Diabetes care*. 1997;20(11):1744-66.
85. Harati H, Hadaegh F, Momenan AA, Ghanei L, Bozorgmanesh MR, Ghanbarian A, et al. Reduction in incidence of type 2 diabetes by lifestyle intervention in a middle eastern community. *American journal of preventive medicine*. 2010;38(6):628-36 e1.
86. Fenwick EK, Xie J, Man RE, Lim LL, Flood VM, Finger RP, et al. Moderate consumption of white and fortified wine is associated with reduced odds of diabetic retinopathy. *Journal of diabetes and its complications*. 2015.
87. Will JC, Galuska DA, Ford ES, Mokdad A, Calle EE. Cigarette smoking and diabetes mellitus: evidence of a positive association from a large prospective cohort study. *International journal of epidemiology*. 2001;30(3):540-6.
88. Amos AF, McCarty DJ, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabetic medicine: a journal of the British Diabetic Association*. 1997;14 Suppl 5:S1-85.
89. Barker DJ, Hales CN, Fall CH, Osmond C, Phipps K, Clark PM. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia*. 1993;36(1):62-7.
90. Poulsen P, Grunnet LG, Pilgaard K, Storgaard H, Alibegovic A, Sonne MP, et al. Increased risk of type 2 diabetes in elderly twins. *Diabetes*. 2009;58(6):1350-5.
91. Medici F, Hawa M, Ianari A, Pyke DA, Leslie RD. Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia*. 1999;42(2):146-50.
92. DeFronzo RA. Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes*. 2009;58(4):773-95.
93. Prentki M, Nolan CJ. Islet beta cell failure in type 2 diabetes. *The Journal of clinical investigation*. 2006;116(7):1802-12.

94. Stefan N, Kantartzis K, Machann J, Schick F, Thamer C, Rittig K, et al. Identification and characterization of metabolically benign obesity in humans. *Archives of internal medicine*. 2008;168(15):1609-16.
95. Wajchenberg BL. Subcutaneous and visceral adipose tissue: their relation to the metabolic syndrome. *Endocrine reviews*. 2000;21(6):697-738.
96. Brunetti A, Chiefari E, Foti D. Recent advances in the molecular genetics of type 2 diabetes mellitus. *World journal of diabetes*. 2014;5(2):128-40.
97. Groop L. Pathogenesis of type 2 diabetes: the relative contribution of insulin resistance and impaired insulin secretion. *International journal of clinical practice Supplement*. 2000(113):3-13.
98. Matveyenko AV, Gurlo T, Daval M, Butler AE, Butler PC. Successful versus failed adaptation to high-fat diet-induced insulin resistance: the role of IAPP-induced beta-cell endoplasmic reticulum stress. *Diabetes*. 2009;58(4):906-16.
99. Doyle ME, Egan JM. Pharmacological agents that directly modulate insulin secretion. *Pharmacological reviews*. 2003;55(1):105-31.
100. Santulli G, Lombardi A, Sorriento D, Anastasio A, Del Giudice C, Formisano P, et al. Age-related impairment in insulin release: the essential role of beta(2)-adrenergic receptor. *Diabetes*. 2012;61(3):692-701.
101. Zhao A, Ohara-Imaizumi M, Brissova M, Benninger RK, Xu Y, Hao Y, et al. Galphao represses insulin secretion by reducing vesicular docking in pancreatic beta-cells. *Diabetes*. 2010;59(10):2522-9.
102. Mardones L, Ormazabal V, Romo X, Jana C, Binder P, Pena E, et al. The glucose transporter-2 (GLUT2) is a low affinity dehydroascorbic acid transporter. *Biochemical and biophysical research communications*. 2011;410(1):7-12.
103. Zheng Y, Scow JS, Duenes JA, Sarr MG. Mechanisms of glucose uptake in intestinal cell lines: role of GLUT2. *Surgery*. 2012;151(1):13-25.
104. Wright EM, Loo DD, Hirayama BA. Biology of human sodium glucose transporters. *Physiological reviews*. 2011;91(2):733-94.
105. Ashcroft FM. ATP-sensitive potassium channelopathies: focus on insulin secretion. *The Journal of clinical investigation*. 2005;115(8):2047-58.
106. Hani EH, Boutin P, Durand E, Inoue H, Permutt MA, Velho G, et al. Missense mutations in the pancreatic islet beta cell inwardly rectifying K⁺ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia*. 1998;41(12):1511-5.

107. Taylor R, Agius L. The biochemistry of diabetes. *The Biochemical journal*. 1988;250(3):625-40.
108. Belfiore A, Frasca F, Pandini G, Sciacca L, Vigneri R. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocrine reviews*. 2009;30(6):586-623.
109. Copps KD, White MF. Regulation of insulin sensitivity by serine/threonine phosphorylation of insulin receptor substrate proteins IRS1 and IRS2. *Diabetologia*. 2012;55(10):2565-82.
110. Ozaki K. Insulin receptor-related receptor in rat islets of Langerhans. *European journal of endocrinology / European Federation of Endocrine Societies*. 1998;139(2):244-7.
111. Serrano R, Villar M, Martinez C, Carrascosa JM, Gallardo N, Andres A. Differential gene expression of insulin receptor isoforms A and B and insulin receptor substrates 1, 2 and 3 in rat tissues: modulation by aging and differentiation in rat adipose tissue. *Journal of molecular endocrinology*. 2005;34(1):153-61.
112. Sesti G, Federici M, Hribal ML, Lauro D, Sbraccia P, Lauro R. Defects of the insulin receptor substrate (IRS) system in human metabolic disorders. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2001;15(12):2099-111.
113. Boura-Halfon S, Zick Y. Phosphorylation of IRS proteins, insulin action, and insulin resistance. *American journal of physiology Endocrinology and metabolism*. 2009;296(4):E581-91.
114. Saltiel AR, Kahn CR. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*. 2001;414(6865):799-806.
115. Bhattacharya S, Dey D, Roy SS. Molecular mechanism of insulin resistance. *Journal of biosciences*. 2007;32(2):405-13.
116. Pratipanawatr W, Pratipanawatr T, Cusi K, Berria R, Adams JM, Jenkinson CP, et al. Skeletal muscle insulin resistance in normoglycemic subjects with a strong family history of type 2 diabetes is associated with decreased insulin-stimulated insulin receptor substrate-1 tyrosine phosphorylation. *Diabetes*. 2001;50(11):2572-8.
117. Boden G, Cheung P, Stein TP, Kresge K, Mozzoli M. FFA cause hepatic insulin resistance by inhibiting insulin suppression of glycogenolysis. *American journal of physiology Endocrinology and metabolism*. 2002;283(1):E12-9.

118. Kelley DE, He J, Menshikova EV, Ritov VB. Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes. *Diabetes*. 2002;51(10):2944-50.
119. Razak F, Anand SS. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. Petersen KF, Dufour S, Befroy D, Garcia R, Shulman GI. *N Engl J Med* 2004; 350: 664-71. *Vascular medicine*. 2004;9(3):223-4.
120. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*. 2003;34(3):267-73.
121. Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, et al. Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(14):8466-71.
122. Wajchenberg BL. beta-cell failure in diabetes and preservation by clinical treatment. *Endocrine reviews*. 2007;28(2):187-218.
123. Robertson RP, Harmon J, Tran PO, Poitout V. Beta-cell glucose toxicity, lipotoxicity, and chronic oxidative stress in type 2 diabetes. *Diabetes*. 2004;53 Suppl 1:S119-24.
124. Kim JW, Yoon KH. Glucolipotoxicity in Pancreatic beta-Cells. *Diabetes & metabolism journal*. 2011;35(5):444-50.
125. Poitout V, Robertson RP. Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocrine reviews*. 2008;29(3):351-66.
126. Robertson RP, Harmon J, Tran PO, Tanaka Y, Takahashi H. Glucose toxicity in beta-cells: type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes*. 2003;52(3):581-7.
127. Eizirik DL, Cardozo AK, Cnop M. The role for endoplasmic reticulum stress in diabetes mellitus. *Endocrine reviews*. 2008;29(1):42-61.
128. Sakuraba H, Mizukami H, Yagihashi N, Wada R, Hanyu C, Yagihashi S. Reduced beta-cell mass and expression of oxidative stress-related DNA damage in the islet of Japanese Type II diabetic patients. *Diabetologia*. 2002;45(1):85-96.
129. Shin CS, Moon BS, Park KS, Kim SY, Park SJ, Chung MH, et al. Serum 8-hydroxy-guanine levels are increased in diabetic patients. *Diabetes care*. 2001;24(4):733-7.
130. Eriksson JW. Metabolic stress in insulin's target cells leads to ROS accumulation - a hypothetical common pathway causing insulin resistance. *FEBS letters*. 2007;581(19):3734-42.

131. Lu H, Koshkin V, Allister EM, Gyulkhandanyan AV, Wheeler MB. Molecular and metabolic evidence for mitochondrial defects associated with beta-cell dysfunction in a mouse model of type 2 diabetes. *Diabetes*. 2010;59(2):448-59.
132. Ma Z, Wirstrom T, Borg LA, Larsson-Nyren G, Hals I, Bondo-Hansen J, et al. Diabetes reduces beta-cell mitochondria and induces distinct morphological abnormalities, which are reproducible by high glucose in vitro with attendant dysfunction. *Islets*. 2012;4(3):233-42.
133. Maedler K, Sergeev P, Ris F, Oberholzer J, Joller-Jemelka HI, Spinas GA, et al. Glucose-induced beta cell production of IL-1beta contributes to glucotoxicity in human pancreatic islets. *The Journal of clinical investigation*. 2002;110(6):851-60.
134. van Raalte DH, Diamant M. Glucolipotoxicity and beta cells in type 2 diabetes mellitus: target for durable therapy? *Diabetes research and clinical practice*. 2011;93 Suppl 1:S37-46.
135. Yaney GC, Corkey BE. Fatty acid metabolism and insulin secretion in pancreatic beta cells. *Diabetologia*. 2003;46(10):1297-312.
136. Abderrahmani A, Niederhauser G, Favre D, Abdelli S, Ferdaoussi M, Yang JY, et al. Human high-density lipoprotein particles prevent activation of the JNK pathway induced by human oxidised low-density lipoprotein particles in pancreatic beta cells. *Diabetologia*. 2007;50(6):1304-14.
137. Cnop M, Hannaert JC, Gruppig AY, Pipeleers DG. Low density lipoprotein can cause death of islet beta-cells by its cellular uptake and oxidative modification. *Endocrinology*. 2002;143(9):3449-53.
138. Rutti S, Ehses JA, Sibling RA, Prazak R, Rohrer L, Georgopoulos S, et al. Low- and high-density lipoproteins modulate function, apoptosis, and proliferation of primary human and murine pancreatic beta-cells. *Endocrinology*. 2009;150(10):4521-30.
139. Brooks-Worrell B, Palmer JP. Is diabetes mellitus a continuous spectrum? *Clinical chemistry*. 2011;57(2):158-61.
140. Brooks-Worrell B, Palmer JP. Immunology in the Clinic Review Series; focus on metabolic diseases: development of islet autoimmune disease in type 2 diabetes patients: potential sequelae of chronic inflammation. *Clinical and experimental immunology*. 2012;167(1):40-6.
141. Brooks-Worrell BM, Reichow JL, Goel A, Ismail H, Palmer JP. Identification of autoantibody-negative autoimmune type 2 diabetic patients. *Diabetes care*. 2011;34(1):168-73.

142. Wilkin TJ. The accelerator hypothesis: weight gain as the missing link between Type I and Type II diabetes. *Diabetologia*. 2001;44(7):914-22.
143. Genovese S, Bazzigaluppi E, Goncalves D, Ciucci A, Cavallo MG, Purrello F, et al. Clinical phenotype and beta-cell autoimmunity in Italian patients with adult-onset diabetes. *European journal of endocrinology / European Federation of Endocrine Societies*. 2006;154(3):441-7.
144. Donath MY, Boni-Schnetzler M, Ellingsgaard H, Ehse JA. Islet inflammation impairs the pancreatic beta-cell in type 2 diabetes. *Physiology*. 2009;24:325-31.
145. Goldberg RB. Cytokine and cytokine-like inflammation markers, endothelial dysfunction, and imbalanced coagulation in development of diabetes and its complications. *The Journal of clinical endocrinology and metabolism*. 2009;94(9):3171-82.
146. Turer AT, Scherer PE. Adiponectin: mechanistic insights and clinical implications. *Diabetologia*. 2012;55(9):2319-26.
147. Wijesekara N, Krishnamurthy M, Bhattacharjee A, Suhail A, Sweeney G, Wheeler MB. Adiponectin-induced ERK and Akt phosphorylation protects against pancreatic beta cell apoptosis and increases insulin gene expression and secretion. *The Journal of biological chemistry*. 2010;285(44):33623-31.
148. Holland WL, Miller RA, Wang ZV, Sun K, Barth BM, Bui HH, et al. Receptor-mediated activation of ceramidase activity initiates the pleiotropic actions of adiponectin. *Nature medicine*. 2011;17(1):55-63.
149. Dunmore SJ, Brown JE. The role of adipokines in beta-cell failure of type 2 diabetes. *The Journal of endocrinology*. 2013;216(1):T37-45.
150. Hull RL, Westermark GT, Westermark P, Kahn SE. Islet amyloid: a critical entity in the pathogenesis of type 2 diabetes. *The Journal of clinical endocrinology and metabolism*. 2004;89(8):3629-43.
151. Westermark P, Andersson A, Westermark GT. Islet amyloid polypeptide, islet amyloid, and diabetes mellitus. *Physiological reviews*. 2011;91(3):795-826.
152. Clark A, Nilsson MR. Islet amyloid: a complication of islet dysfunction or an aetiological factor in Type 2 diabetes? *Diabetologia*. 2004;47(2):157-69.
153. Masters SL, Dunne A, Subramanian SL, Hull RL, Tannahill GM, Sharp FA, et al. Activation of the NLRP3 inflammasome by islet amyloid polypeptide provides a mechanism for enhanced IL-1beta in type 2 diabetes. *Nature immunology*. 2010;11(10):897-904.

154. Jurgens CA, Toukatly MN, Fligner CL, Udayasankar J, Subramanian SL, Zraika S, et al. beta-cell loss and beta-cell apoptosis in human type 2 diabetes are related to islet amyloid deposition. *The American journal of pathology*. 2011;178(6):2632-40.
155. Cernea S, Raz I. Therapy in the early stage: incretins. *Diabetes care*. 2011;34 Suppl 2:S264-71.
156. Vilsboll T. The effects of glucagon-like peptide-1 on the beta cell. *Diabetes, obesity & metabolism*. 2009;11 Suppl 3:11-8.
157. Fehse F, Trautmann M, Holst JJ, Halseth AE, Nanayakkara N, Nielsen LL, et al. Exenatide augments first- and second-phase insulin secretion in response to intravenous glucose in subjects with type 2 diabetes. *The Journal of clinical endocrinology and metabolism*. 2005;90(11):5991-7.
158. Quddusi S, Vahl TP, Hanson K, Prigeon RL, D'Alessio DA. Differential effects of acute and extended infusions of glucagon-like peptide-1 on first- and second-phase insulin secretion in diabetic and nondiabetic humans. *Diabetes care*. 2003;26(3):791-8.
159. Buteau J, El-Assaad W, Rhodes CJ, Rosenberg L, Joly E, Prentki M. Glucagon-like peptide-1 prevents beta cell glucolipotoxicity. *Diabetologia*. 2004;47(5):806-15.
160. Portha B, Tourrel-Cuzin C, Movassat J. Activation of the GLP-1 receptor signalling pathway: a relevant strategy to repair a deficient beta-cell mass. *Experimental diabetes research*. 2011;2011:376509.
161. Meier JJ, Nauck MA. Is the diminished incretin effect in type 2 diabetes just an epiphenomenon of impaired beta-cell function? *Diabetes*. 2010;59(5):1117-25.
162. Carlos Crespoa MB, Aitana Soria-Juan, Alfonso Lopez-Alba,, Soria NL-MyB. Costes directos de la diabetes mellitus y de sus complicaciones en España (Estudio SECCAID: Spain estimated cost Ciberdem-Cabimer in Diabetes). *Av Diabetol*. 2013;29:182-9.
163. Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. *Human molecular genetics*. 2015;24(R1):R85-92.
164. WHO. Obesity. <http://www.euro.who.int/en/health-topics/noncommunicable-diseases/obesity/data-and-statistics/>. (Accessed August 2015).
165. Neel JV. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? 1962. *Bulletin of the World Health Organization*. 1999;77(8):694-703; discussion 692-3.

166. Pierce M, Keen H, Bradley C. Risk of diabetes in offspring of parents with non-insulin-dependent diabetes. *Diabetic medicine : a journal of the British Diabetic Association*. 1995;12(1):6-13.
167. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia*. 1999;42(2):139-45.
168. 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. *Nature genetics*. 2000;26(2):163-75.
169. Song Y, Niu T, Manson JE, Kwiatkowski DJ, Liu S. Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *American journal of human genetics*. 2004;74(2):208-22.
170. Weedon MN, Schwarz PE, Horikawa Y, Iwasaki N, Illig T, Holle R, et al. Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *American journal of human genetics*. 2003;73(5):1208-12.
171. 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. *American journal of human genetics*. 1999;64(4):1127-40.
172. Grant SF, 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. *Nature genetics*. 2006;38(3):320-3.
173. Reynisdottir I, Thorleifsson G, Benediktsson R, Sigurdsson G, Emilsson V, Einarsdottir AS, et al. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *American journal of human genetics*. 2003;73(2):323-35.
174. Tong Y, Lin Y, Zhang Y, Yang J, Zhang Y, Liu H, et al. Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. *BMC medical genetics*. 2009;10:15.
175. Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, et al. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nature genetics*. 1998;20(3):284-7.

176. Almind K, Bjorbaek C, Vestergaard H, Hansen T, Echwald S, Pedersen O. Aminoacid polymorphisms of insulin receptor substrate-1 in non-insulin-dependent diabetes mellitus. *Lancet*. 1993;342(8875):828-32.
177. Minton JA, Hattersley AT, Owen K, McCarthy MI, Walker M, Latif F, et al. Association studies of genetic variation in the WFS1 gene and type 2 diabetes in U.K. populations. *Diabetes*. 2002;51(4):1287-90.
178. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature genetics*. 2010;42(7):579-89.
179. van Dam RM, Hoebee B, Seidell JC, Schaap MM, de Bruin TW, Feskens EJ. Common variants in the ATP-sensitive K⁺ channel genes KCNJ11 (Kir6.2) and ABCC8 (SUR1) in relation to glucose intolerance: population-based studies and meta-analyses. *Diabetic medicine : a journal of the British Diabetic Association*. 2005;22(5):590-8.
180. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316(5829):1336-41.
181. Franks PW, Rolandsson O, Debenham SL, Fawcett KA, Payne F, Dina C, et al. Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. *Diabetologia*. 2008;51(3):458-63.
182. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nature genetics*. 2007;39(8):951-3.
183. Holmkvist J, Almgren P, Lyssenko V, Lindgren CM, Eriksson KF, Isomaa B, et al. Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes. *Diabetes*. 2008;57(6):1738-44.
184. Holmkvist J, Cervin C, Lyssenko V, Winckler W, Anevski D, Cilio C, et al. Common variants in HNF-1 alpha and risk of type 2 diabetes. *Diabetologia*. 2006;49(12):2882-91.
185. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nature genetics*. 2007;39(8):977-83.
186. Winckler W, Weedon MN, Graham RR, McCarroll SA, Purcell S, Almgren P, et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes*. 2007;56(3):685-93.

187. International HapMap C. The International HapMap Project. *Nature*. 2003;426(6968):789-96.
188. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881-5.
189. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-78.
190. Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of BioMedical R, Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316(5829):1331-6.
191. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007;316(5829):1341-5.
192. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature genetics*. 2007;39(6):770-5.
193. Jonsson A, Isomaa B, Tuomi T, Taneera J, Salehi A, Nilsson P, et al. A variant in the KCNQ1 gene predicts future type 2 diabetes and mediates impaired insulin secretion. *Diabetes*. 2009;58(10):2409-13.
194. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. *Nature*. 2009;462(7275):868-74.
195. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nature genetics*. 2008;40(9):1098-102.
196. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature genetics*. 2008;40(5):638-45.
197. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature genetics*. 2010;42(2):105-16.
198. Replication DIG, Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes C, South Asian Type 2 Diabetes C, Mexican American Type 2 Diabetes C, Type

- 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples C, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics*. 2014;46(3):234-44.
199. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nature genetics*. 2008;40(6):695-701.
200. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genetic epidemiology*. 2011;35(5):310-7.
201. Grarup N, Sandholt CH, Hansen T, Pedersen O. Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia*. 2014;57(8):1528-41.
202. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*. 2013;493(7431):216-20.
203. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337(6090):64-9.
204. Consortium STD, Estrada K, Aukrust I, Bjorkhaug L, Burt NP, Mercader JM, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *Jama*. 2014;311(22):2305-14.
205. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nature reviews Genetics*. 2007;8(9):657-62.
206. Lohmueller KE, Sparso T, Li Q, Andersson E, Korneliussen T, Albrechtsen A, et al. Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *American journal of human genetics*. 2013;93(6):1072-86.
207. Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nature genetics*. 2014;46(3):294-8.
208. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-73.
209. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461(7261):272-6.

-
210. Service SK, Teslovich TM, Fuchsberger C, Ramensky V, Yajnik P, Koboldt DC, et al. Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. *PLoS genetics*. 2014;10(1):e1004147.
211. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews Genetics*. 2010;11(6):446-50.
212. Perry JR, McCarthy MI, Hattersley AT, Zeggini E, Wellcome Trust Case Control C, Weedon MN, et al. Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. *Diabetes*. 2009;58(6):1463-7.
213. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature genetics*. 2003;33 Suppl:228-37.
214. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, et al. Human Gene Mutation Database (HGMD): 2003 update. *Human mutation*. 2003;21(6):577-81.
215. Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nature genetics*. 2010;42(11):969-72.
216. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome biology*. 2011;12(1):R1.
217. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. *Nature methods*. 2010;7(2):111-8.
218. Hedges DJ, Burges D, Powell E, Almonte C, Huang J, Young S, et al. Exome sequencing of a multigenerational human pedigree. *PloS one*. 2009;4(12):e8232.
219. Panoutsopoulou K, Zeggini E. Finding common susceptibility variants for complex disease: past, present and future. *Briefings in functional genomics & proteomics*. 2009;8(5):345-52.
220. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *American journal of human genetics*. 2001;69(1):124-37.
221. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics & development*. 2009;19(3):212-9.
222. Yang Q, Khoury MJ, Friedman J, Little J, Flanders WD. How many genes underlie the occurrence of common complex diseases in the population? *International journal of epidemiology*. 2005;34(5):1129-37.

223. Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? *Annals of the New York Academy of Sciences*. 2010;1212:59-77.
224. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53.
225. Robinson R. Common disease, multiple rare (and distant) variants. *PLoS biology*. 2010;8(1):e1000293.
226. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *American journal of human genetics*. 2008;82(1):100-12.
227. Haumaitre C. Epigenetic regulation of pancreatic islets. *Current diabetes reports*. 2013;13(5):624-32.
228. Drong AW, Lindgren CM, McCarthy MI. The genetic and epigenetic basis of type 2 diabetes and obesity. *Clinical pharmacology and therapeutics*. 2012;92(6):707-15.
229. Raciti GA, Nigro C, Longo M, Parrillo L, Miele C, Formisano P, et al. Personalized medicine and type 2 diabetes: lesson from epigenetics. *Epigenomics*. 2014;6(2):229-38.
230. Ruchat SM, Hivert MF, Bouchard L. Epigenetic programming of obesity and diabetes by in utero exposure to gestational diabetes mellitus. *Nutrition reviews*. 2013;71 Suppl 1:S88-94.
231. Avrahami D, Kaestner KH. Epigenetic regulation of pancreas development and function. *Seminars in cell & developmental biology*. 2012;23(6):693-700.
232. Waki H, Yamauchi T, Kadowaki T. The epigenome and its role in diabetes. *Current diabetes reports*. 2012;12(6):673-85.
233. Kong L, Zhu J, Han W, Jiang X, Xu M, Zhao Y, et al. Significance of serum microRNAs in pre-diabetes and newly diagnosed type 2 diabetes: a clinical study. *Acta diabetologica*. 2011;48(1):61-9.
234. Li Y, Xu X, Liang Y, Liu S, Xiao H, Li F, et al. miR-375 enhances palmitate-induced lipoapoptosis in insulin-secreting NIT-1 cells by repressing myotrophin (V1) protein expression. *International journal of clinical and experimental pathology*. 2010;3(3):254-64.
235. Rottiers V, Naar AM. MicroRNAs in metabolism and metabolic disorders. *Nature reviews Molecular cell biology*. 2012;13(4):239-50.
236. Deiuliis JA. MicroRNAs as regulators of metabolic disease: Pathophysiologic significance and emerging role as biomarkers and therapeutics. *International journal of obesity*. 2015.

-
237. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *The New England journal of medicine*. 2008;359(21):2220-32.
238. Shendure J, Ji H. Next-generation DNA sequencing. *Nature biotechnology*. 2008;26(10):1135-45.
239. Kuehn BM. 1000 Genomes Project promises closer look at variation in human genome. *Jama*. 2008;300(23):2715.
240. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. *Genome biology*. 2013;14(5):R51.
241. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nature protocols*. 2013;8(8):1551-66.
242. Langenberg C, Sharp SJ, Franks PW, Scott RA, Deloukas P, Forouhi NG, et al. Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. *PLoS medicine*. 2014;11(5):e1001647.
243. Florez JC. Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. *The Journal of clinical endocrinology and metabolism*. 2008;93(12):4633-42.
244. Matsuda A, Kuzuya T. Relationship between obesity and concordance rate for type 2 (non-insulin-dependent) diabetes mellitus among twins. *Diabetes research and clinical practice*. 1994;26(2):137-43.
245. Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature genetics*. 2010;42(7):570-5.
246. Gibson G. Rare and common variants: twenty arguments. *Nature reviews Genetics*. 2011;13(2):135-45.
247. Eyre-Walker A. Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107 Suppl 1:1752-6.
248. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature reviews Genetics*. 2010;11(6):415-25.
249. Biesecker LG, Shianna KV, Mullikin JC. Exome sequencing: the expert view. *Genome biology*. 2011;12(9):128.

250. Chilamakuri CS, Lorenz S, Madoui MA, Vodak D, Sun J, Hovig E, et al. Performance comparison of four exome capture systems for deep sequencing. *BMC genomics*. 2014;15:449.
251. Lehne B, Lewis CM, Schlitt T. Exome localization of complex disease association signals. *BMC genomics*. 2011;12:92.
252. Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, et al. Genetic variation in an individual human exome. *PLoS genetics*. 2008;4(8):e1000160.
253. Huang K, Nair AK, Muller YL, Piaggi P, Bian L, Del Rosario M, et al. Whole exome sequencing identifies variation in CYB5A and RNF10 associated with adiposity and type 2 diabetes. *Obesity*. 2014;22(4):984-8.
254. Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*. 2014;512(7513):190-3.
255. Kim SY, Lohmueller KE, Albrechtsen A, Li Y, Korneliusen T, Tian G, et al. Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC bioinformatics*. 2011;12:231.
256. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome biology*. 2009;10(3):R32.
257. Kyriakou T, Seedorf U, Goel A, Hopewell JC, Clarke R, Watkins H, et al. A common LPA null allele associates with lower lipoprotein(a) levels and coronary artery disease risk. *Arteriosclerosis, thrombosis, and vascular biology*. 2014;34(9):2095-9.
258. Nordestgaard BG, Chapman MJ, Ray K, Boren J, Andreotti F, Watts GF, et al. Lipoprotein(a) as a cardiovascular risk factor: current status. *European heart journal*. 2010;31(23):2844-53.
259. Ogorelkova M, Gruber A, Utermann G. Molecular basis of congenital lp(a) deficiency: a frequent apo(a) 'null' mutation in caucasians. *Human molecular genetics*. 1999;8(11):2087-96.
260. Habib SS, Aslam M, Shah SF, Naveed AK. Lipoprotein (a) is associated with basal insulin levels in patients with type 2 Diabetes Mellitus. *Arquivos brasileiros de cardiologia*. 2009;93(1):28-33.
261. Jenkins AJ, Steele JS, Janus ED, Santamaria JD, Best JD. Plasma apolipoprotein (a) is increased in type 2 (non-insulin-dependent) diabetic patients with microalbuminuria. *Diabetologia*. 1992;35(11):1055-9.

262. Haffner SM, Morales PA, Stern MP, Gruber MK. Lp(a) concentrations in NIDDM. *Diabetes*. 1992;41(10):1267-72.
263. Rainwater DL, MacCluer JW, Stern MP, VandeBerg JL, Haffner SM. Effects of NIDDM on lipoprotein(a) concentration and apolipoprotein(a) size. *Diabetes*. 1994;43(7):942-6.
264. Boronat M, Saavedra P, Perez-Martin N, Lopez-Madrazo MJ, Rodriguez-Perez C, Novoa FJ. High levels of lipoprotein(a) are associated with a lower prevalence of diabetes with advancing age: results of a cross-sectional epidemiological survey in Gran Canaria, Spain. *Cardiovascular diabetology*. 2012;11:81.
265. Mora S, Kamstrup PR, Rifai N, Nordestgaard BG, Buring JE, Ridker PM. Lipoprotein(a) and risk of type 2 diabetes. *Clinical chemistry*. 2010;56(8):1252-60.
266. Ardestani A, Paroni F, Azizi Z, Kaur S, Khobragade V, Yuan T, et al. MST1 is a key regulator of beta cell apoptosis and dysfunction in diabetes. *Nature medicine*. 2014;20(4):385-97.
267. Qi L, Liu S, Rifai N, Hunter D, Hu FB. Associations of the apolipoprotein A1/C3/A4/A5 gene cluster with triglyceride and HDL cholesterol levels in women with type 2 diabetes. *Atherosclerosis*. 2007;192(1):204-10.
268. Timpson NJ, Walter K, Min JL, Tachmazidou I, Malerba G, Shin SY, et al. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nature communications*. 2014;5:4871.
269. Jenny L, Ajjan R, King R, Thiel S, Schroeder V. Plasma levels of mannan-binding lectin-associated serine proteases MASP-1 and MASP-2 are elevated in type 1 diabetes and correlate with glycaemic control. *Clinical and experimental immunology*. 2015;180(2):227-32.
270. Viken MK, Blomhoff A, Olsson M, Akselsen HE, Pociot F, Nerup J, et al. Reproducible association with type 1 diabetes in the extended class I region of the major histocompatibility complex. *Genes and immunity*. 2009;10(4):323-33.
271. Cariou B, Ouguerram K, Zair Y, Guerois R, Langhi C, Kourimate S, et al. PCSK9 dominant negative mutant results in increased LDL catabolic rate and familial hypobetalipoproteinemia. *Arteriosclerosis, thrombosis, and vascular biology*. 2009;29(12):2191-7.
272. Blesa S, Vernia S, Garcia-Garcia AB, Martinez-Hervas S, Ivorra C, Gonzalez-Albert V, et al. A new PCSK9 gene promoter variant affects gene expression and causes

autosomal dominant hypercholesterolemia. *The Journal of clinical endocrinology and metabolism*. 2008;93(9):3577-83.

273. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *The New England journal of medicine*. 2006;354(12):1264-72.

274. Gok O, Karaali ZE, Acar L, Kilic U, Ergen A. ABCG5 and ABCG8 Gene Polymorphisms in Type 2 Diabetes Mellitus in the Turkish Population. *Canadian journal of diabetes*. 2015.

275. Nicolas A, Fatima S, Lamri A, Bellili-Munoz N, Halimi JM, Saulnier PJ, et al. ABCG8 polymorphisms and renal disease in type 2 diabetic patients. *Metabolism: clinical and experimental*. 2015;64(6):713-9.

276. Shishova EY, Stoll JM, Ersoy BA, Shrestha S, Scapa EF, Li Y, et al. Genetic ablation or chemical inhibition of phosphatidylcholine transfer protein attenuates diet-induced hepatic glucose production. *Hepatology*. 2011;54(2):664-74.

277. Yang YH, Manning Fox JE, Zhang KL, MacDonald PE, Johnson JD. Intra-islet SLIT-ROBO signaling is required for beta-cell survival and potentiates insulin secretion. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(41):16480-5.

278. McRobie DJ, Glover DD, Tracy TS. Effects of gestational and overt diabetes on human placental cytochromes P450 and glutathione S-transferase. *Drug metabolism and disposition: the biological fate of chemicals*. 1998;26(4):367-71.

279. Zhang Q, He M, Wang J, Liu S, Cheng H, Cheng Y. Predicting of disease genes for gestational diabetes mellitus based on network and functional consistency. *European journal of obstetrics, gynecology, and reproductive biology*. 2015;186:91-6.

280. Merhi M, Demirdjian S, Hariri E, Sabbah N, Youhanna S, Ghassibe-Sabbagh M, et al. Impact of inflammation, gene variants, and cigarette smoking on coronary artery disease risk. *Inflammation research : official journal of the European Histamine Research Society [et al]*. 2015;64(6):415-22.

281. Bo S, Gambino R, Menato G, Canil S, Ponzio V, Pinach S, et al. Isoleucine-to-methionine substitution at residue 148 variant of PNPLA3 gene and metabolic outcomes in gestational diabetes. *The American journal of clinical nutrition*. 2015;101(2):310-8.

282. Rahman SA, Nessa A, Hussain K. Molecular mechanisms of congenital hyperinsulinism. *Journal of molecular endocrinology*. 2015;54(2):R119-29.

283. Senniappan S, Shanti B, James C, Hussain K. Hyperinsulinaemic hypoglycaemia: genetic mechanisms, diagnosis and management. *Journal of inherited metabolic disease*. 2012;35(4):589-601.
284. Sebastiani P, Montano M, Puca A, Solovieff N, Kojima T, Wang MC, et al. RNA editing genes associated with extreme old age in humans and with lifespan in *C. elegans*. *PloS one*. 2009;4(12):e8210.
285. Oguro R, Kamide K, Katsuya T, Akasaka H, Sugimoto K, Congrains A, et al. A single nucleotide polymorphism of the adenosine deaminase, RNA-specific gene is associated with the serum triglyceride level, abdominal circumference, and serum adiponectin concentration. *Experimental gerontology*. 2012;47(2):183-7.
286. Sharma PR, Mackey AJ, Dejene EA, Ramadan JW, Langefeld CD, Palmer ND, et al. An Islet-Targeted Genome-Wide Association Scan Identifies Novel Genes Implicated in Cytokine-Mediated Islet Stress in Type 2 Diabetes. *Endocrinology*. 2015;156(9):3147-56.
287. Ma L, Hanson RL, Que LN, Cali AM, Fu M, Mack JL, et al. Variants in ARHGEF11, a candidate gene for the linkage to type 2 diabetes on chromosome 1q, are nominally associated with insulin resistance and type 2 diabetes in Pima Indians. *Diabetes*. 2007;56(5):1454-9.
288. Torn C, Hadley D, Lee HS, Hagopian W, Lernmark A, Simell O, et al. Role of Type 1 Diabetes-Associated SNPs on Risk of Autoantibody Positivity in the TEDDY Study. *Diabetes*. 2015;64(5):1818-29.
289. Garcia-Melendez ME, Salinas-Santander M, Sanchez-Dominguez C, Gonzalez-Cardenas H, Cerda-Flores RM, Ocampo-Candiani J, et al. Protein tyrosine phosphatase PTPN22 +1858C/T polymorphism is associated with active vitiligo. *Experimental and therapeutic medicine*. 2014;8(5):1433-7.
290. Shemer A, Bergman R, Linn S, Kantor Y, Friedman-Birnbaum R. Diabetic dermopathy and internal complications in diabetes mellitus. *International journal of dermatology*. 1998;37(2):113-5.
291. Butler AE, Janson J, Bonner-Weir S, Ritzel R, Rizza RA, Butler PC. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes*. 2003;52(1):102-10.
292. Song B, Scheuner D, Ron D, Pennathur S, Kaufman RJ. Chop deletion reduces oxidative stress, improves beta cell function, and promotes cell survival in multiple mouse models of diabetes. *The Journal of clinical investigation*. 2008;118(10):3378-89.

293. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009;136(2):215-33.
294. Chang TC, Mendell JT. microRNAs in vertebrate physiology and human disease. *Annual review of genomics and human genetics*. 2007;8:215-39.
295. Guay C, Roggli E, Nesca V, Jacovetti C, Regazzi R. Diabetes mellitus, a microRNA-related disease? *Translational research : the journal of laboratory and clinical medicine*. 2011;157(4):253-64.
296. Kolfschoten IG, Roggli E, Nesca V, Regazzi R. Role and therapeutic potential of microRNAs in diabetes. *Diabetes, obesity & metabolism*. 2009;11 Suppl 4:118-29.
297. Osmai M, Osmai Y, Bang-Berthelsen CH, Pallesen EM, Vestergaard AL, Novotny GW, et al. microRNAs as regulators of beta-cell function and dysfunction. *Diabetes/metabolism research and reviews*. 2015.
298. Belgardt BF, Ahmed K, Spranger M, Latreille M, Denzler R, Kondratiuk N, et al. The microRNA-200 family regulates pancreatic beta cell survival in type 2 diabetes. *Nature medicine*. 2015;21(6):619-27.
299. Wu Q, Wang C, Guo L, Ge Q, Lu Z. Identification and characterization of novel microRNA candidates from deep sequencing. *Clinica chimica acta; international journal of clinical chemistry*. 2013;415:239-44.
300. Carobbio S, Hagen RM, Lelliott CJ, Slawik M, Medina-Gomez G, Tan CY, et al. Adaptive changes of the Insig1/SREBP1/SCD1 set point help adipose tissue to cope with increased storage demands of obesity. *Diabetes*. 2013;62(11):3697-708.
301. Zhang D, Yang H, Kong X, Wang K, Mao X, Yan X, et al. Proteomics analysis reveals diabetic kidney as a ketogenic organ in type 2 diabetes. *American journal of physiology Endocrinology and metabolism*. 2011;300(2):E287-95.
302. Lui WO, Pourmand N, Patterson BK, Fire A. Patterns of known and novel small RNAs in human cervical cancer. *Cancer research*. 2007;67(13):6031-43.
303. Jun H, Ying H, Daiwen C, Bing Y, Xiangbing M, Ping Z, et al. miR-628, a microRNA that is induced by Toll-like receptor stimulation, regulates porcine innate immune responses. *Scientific reports*. 2015;5:12226.
304. Hasan MM, Akter R, Ullah MS, Abedin MJ, Ullah GM, Hossain MZ. A Computational Approach for Predicting Role of Human MicroRNAs in MERS-CoV Genome. *Advances in bioinformatics*. 2014;2014:967946.

-
305. Tambyah PA, Sepramaniam S, Mohamed Ali J, Chai SC, Swaminathan P, Armugam A, et al. microRNAs in circulation are altered in response to influenza A virus infection in humans. *PloS one*. 2013;8(10):e76811.
306. Faruq O, Vecchione A. microRNA: Diagnostic Perspective. *Frontiers in medicine*. 2015;2:51.
307. Kim YK. Extracellular microRNAs as Biomarkers in Human Disease. *Chonnam medical journal*. 2015;51(2):51-7.
308. Leng S, Zhang W, Zheng Y, Liberman Z, Rhodes CJ, Eldar-Finkelman H, et al. Glycogen synthase kinase 3 beta mediates high glucose-induced ubiquitination and proteasome degradation of insulin receptor substrate 1. *The Journal of endocrinology*. 2010;206(2):171-81.
309. Luo X, Zhang Y, Ruan X, Jiang X, Zhu L, Wang X, et al. Fasting-induced protein phosphatase 1 regulatory subunit contributes to postprandial blood glucose homeostasis via regulation of hepatic glycogenesis. *Diabetes*. 2011;60(5):1435-45.
310. Rider MH, Bertrand L, Vertommen D, Michels PA, Rousseau GG, Hue L. 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase: head-to-head with a bifunctional enzyme that controls glycolysis. *The Biochemical journal*. 2004;381(Pt 3):561-79.
311. Delahaye-Sourdeix M, Anantharaman D, Timofeeva MN, Gaborieau V, Chabrier A, Vallee MP, et al. A rare truncating BRCA2 variant and genetic susceptibility to upper aerodigestive tract cancer. *Journal of the National Cancer Institute*. 2015;107(5).
312. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature genetics*. 2013;45(4):353-61, 61e1-2.
313. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nature genetics*. 2014;46(7):736-41.
314. Fu Z, Shrubsole MJ, Li G, Smalley WE, Hein DW, Cai Q, et al. Interaction of cigarette smoking and carcinogen-metabolizing polymorphisms in the risk of colorectal polyps. *Carcinogenesis*. 2013;34(4):779-86.
315. Beesley J, Jordan SJ, Spurdle AB, Song H, Ramus SJ, Kjaer SK, et al. Association between single-nucleotide polymorphisms in hormone metabolism and DNA repair genes and epithelial ovarian cancer: results from two Australian studies and an additional validation set. *Cancer epidemiology, biomarkers & prevention: a publication of the*

American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2007;16(12):2557-65.

316. Schildkraut JM, Iversen ES, Wilson MA, Clyde MA, Moorman PG, Palmieri RT, et al. Association between DNA damage response and repair genes and risk of invasive serous ovarian cancer. *PLoS one*. 2010;5(4):e10061.

317. Cho HR, Kim SK, Lim HK, Jeong Park H, Chung JH, Lee MH. Association study between nonsense polymorphism (rs2039381, Gln71Stop) of interferon-epsilon and susceptibility to vitiligo in Korean population. *Immunological investigations*. 2013;42(5):423-30.

318. Kim SK, Park HJ, Kim JW, Chung JH, Yoo SD, Kim DH, et al. T Allele of nonsense polymorphism (rs2039381, Gln71Stop) of interferon-epsilon is a risk factor for the development of intracerebral hemorrhage. *Human immunology*. 2014;75(1):88-90.

319. Dick DM, Aliev F, Krueger RF, Edwards A, Agrawal A, Lynskey M, et al. Genome-wide association study of conduct disorder symptomatology. *Molecular psychiatry*. 2011;16(8):800-8.

320. Grupe A, Li Y, Rowland C, Nowotny P, Hinrichs AL, Smemo S, et al. A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease. *American journal of human genetics*. 2006;78(1):78-88.

321. Camilleri M, Klee EW, Shin A, Carlson P, Li Y, Grover M, et al. Irritable bowel syndrome-diarrhea: characterization of genotype by exome sequencing, and phenotypes of bile acid synthesis and colonic transit. *American journal of physiology Gastrointestinal and liver physiology*. 2014;306(1):G13-26.

322. Zhang X, Ma D, Caruso M, Lewis M, Qi Y, Yi Z. Quantitative phosphoproteomics reveals novel phosphorylation events in insulin signaling regulated by protein phosphatase 1 regulatory subunit 12A. *Journal of proteomics*. 2014;109:63-75.

323. Yin H, Ru H, Yu L, Kang Y, Lin G, Liu C, et al. Targeting of Rho kinase ameliorates impairment of diabetic endothelial function in intrarenal artery. *International journal of molecular sciences*. 2013;14(10):20282-98.

324. Shulman GI. Cellular mechanisms of insulin resistance. *The Journal of clinical investigation*. 2000;106(2):171-6.

325. Hanson RL, Bogardus C, Duggan D, Kobes S, Knowlton M, Infante AM, et al. A search for variants associated with young-onset type 2 diabetes in American Indians in a 100K genotyping array. *Diabetes*. 2007;56(12):3045-52.

-
326. Bian L, Hanson RL, Muller YL, Ma L, Investigators M, Kobes S, et al. Variants in ACAD10 are associated with type 2 diabetes, insulin resistance and lipid oxidation in Pima Indians. *Diabetologia*. 2010;53(7):1349-53.
327. Daimon M, Sato H, Oizumi T, Toriyama S, Saito T, Karasawa S, et al. Association of the PIK3C2G gene polymorphisms with type 2 DM in a Japanese population. *Biochemical and biophysical research communications*. 2008;365(3):466-71.
328. Douroudis K, Kisand K, Nemvalts V, Rajasalu T, Uibo R. Allelic variants in the PHTF1-PTPN22, C12orf30 and CD226 regions as candidate susceptibility factors for the type 1 diabetes in the Estonian population. *BMC medical genetics*. 2010;11:11.
329. Smyth DJ, Cooper JD, Howson JM, Walker NM, Plagnol V, Stevens H, et al. PTPN22 Trp620 explains the association of chromosome 1p13 with type 1 diabetes and shows a statistical interaction with HLA class II genotypes. *Diabetes*. 2008;57(6):1730-7.
330. Li B, Leal SM. Discovery of rare variants via sequencing: implications for the design of complex trait association studies. *PLoS genetics*. 2009;5(5):e1000481.
331. Prokopenko I, McCarthy MI, Lindgren CM. Type 2 diabetes: new genes, new understanding. *Trends in genetics : TIG*. 2008;24(12):613-21.

Anexo

En este anexo se encuentran descritos todos los comandos empleados durante el proceso de análisis bioinformático de las secuencias obtenidas:

Obtención de estadísticas de archivos FASTQ

Ejecutar FastQC

- Comando generic:

```
$ fastqc -o /path/to/stats_folder input.fastq.gz
```

- Opciones empleadas FastQC:

-o: Crear todos los archivos de salida en el directorio especificado. Este directorio debe existir ya que el programa no lo creara. Si esta opción no se establece, se crea el archivo de salida para cada archivo de secuencias en el mismo directorio que el archivo de secuencia que se procesó.

Depuración de los archivos FASTQ

Recortar las bases de baja calidad de las lecturas obtenidas.

Ejecutar seqtk

- Comando genérico:

```
$ seqtk trimfq input.fastq.gz > output_trimmed.fastq
```

Mapeo

1.1. Ejecutar BWA

- Comando genérico:

```
$ bwa mem -t[INT] -C -M -R[STR] <reference.fa.gz> <input_R1.fastq>  
<input_R2.fastq> > <output_R1-R2.sam> 2> <output_R1-R2.sam.log>
```

- Opciones empleadas BWA:

- t (INT): Numero de hilos de ejecución [1]
- C: Anexar comentario FASTA /Q al archivo SAM resultante. Esta opción se puede utilizar para transferir información de lectura (por ejemplo, código de barras) al archivo resultante.
- M: Indica los alineamientos divididos como secundarios (para la compatibilidad con Picard).
- R (STR): Completar la línea del read group de la cabecera. '\t' puede ser utilizado en STR y se convertirá a TAB en el archivo SAM resultante. El read group ID ira unido a cada lectura en el output. Un ejemplo es: '@RG\tID:foo\tSM:bar'.[null]

- Formato de la cabecera:

```
@RG\tID:id\tPL:platform\tSM:sample\tDS:description\tCN:center\tDT:date\tPU:flow cell-
barcode.lane"
```

- Ejemplo de la cabecera:

```
"@RG\tID:280397_CGATGT_L005_R1_001\tPL:Illumina\tSM:280397\tD
S:ref=1KG;pfx=/media/sf_ugdg_data/ref_data/human/1kg/hs37d5.fa.gz
\tCN:INCLIVA\tDT:2013-09-19\tPU:130327_H114_0265_AD1LLUACXX-2"
```

1.2. Conversión de SAM a BAM

- Comando genérico:

```
$ samtools view -b -S <input.sam> > <output.bam>
```

- Opciones empleadas SAMtools:

-b: El archivo de salida se crea en formato BAM.

-S: El archivo de entrada esta en formato SAM. Si las líneas de la cabecera @SQ no están, se requiere la opción '-t'.

2. Ejecutar BWA y SAMtools

- Comando genérico:

```
$ bwa mem -t[INT] -C -M -R[STR] <reference.fa.gz> <input_R1.fastq> <input_R2.fastq> 2>
<output_R1-R2.sam.log> | samtools view -b -S - >
<output_R1-R2.bam>
```

- Opciones empleadas SAMtools:

-: El archivo de entrada proviene de stdout

Visualizar SAM

Ejecutar SAMtools view

- Comando genérico:

```
$ samtools view -S <input.sam>
```

- Opciones empleadas SAMtools:

-S: El archivo de entrada esta en formato SAM. Si las líneas de la cabecera @SQ no están, se requiere la opción '-t'.

Visualizar BAM

Ejecutar SAMtools view

- Comando genérico:

```
$ samtools view <input.bam>
```

Inspeccionar archivos BAM

Ejecutar SAMtools tview

- Comando genérico:

```
$ samtools tview <input.bam>
```

Ordenar archivos BAM

Ejecutar SAMtools sort

- Comando genérico:

```
$ samtools sort <input.bam> <output.sorted> -m[int]
```

- Opciones empleadas SAMtools:

-m (INT): Máxima memoria requerida aproximadamente. [500000000]

Indexar archivos BAM

Ejecutar SAMtools index

- Comando genérico:

```
$ samtools index <input.sorted.bam>
```

Eliminar duplicados

Ejecutar SAMtools rmdup

- Comando genérico:

```
$ samtools rmdup <input.sorted.bam> <output.sorted.rmdup.bam> 2>  
<output.bam.sorted.rmdup.log>
```

Unir varios archivos BAM

Unir cabeceras de archivos BAM

- Comando genérico:

```
$ cat <(samtools view -H <input_L1.sorted.rmdup.bam> | grep -v  
'^@RG') <(for file in <input_L1.sorted.rmdup.bam>  
<input_L2.sorted.rmdup.bam>; do samtools view -H $file | grep  
'^@RG'; done) > <newheader.txt>
```

- Opciones empleadas SAMtools:

-H: Devolver solamente la cabecera.

Ejecutar SAMtools merge

- Comando genérico:

```
$ samtools merge -h [<newheader.txt>] <output_L1-L2.bam>  
<input_L1.sorted.rmdup.bam> <input_L2.sorted.rmdup.bam>
```

- Opciones empleadas SAMtools:

-h (FILE): utilizar el contenido de FILE como cabecera para el fichero de salida, reemplazando así cualquier línea de cabecera que de otro modo sería copiada desde in1.bam. (FILE es en realidad formato SAM, aunque cualquier alineación que pueda contener será ignorada).

Obtención estadísticas archivos BAM

Ejecutar SAMtools flagstat

- Comando genérico:

```
$ samtools flagstat <input.bam> > <output.bas>
```

Seleccionar subconjuntos en archivos BAM

Ejecutar SAMtools view

- Comando genérico:

```
$ samtools view <input.bam> "chr:start-end" > <output.subset.bam>
```

Obtención de cobertura

Ejecutar BEDtools para obtener un histograma de cobertura

- Comando genérico:

```
$ coverageBed -hist -abam <input.sorted.rmdup.bam> -b <regions.bed> >  
<output.sorted.rmdup.hist.cov>
```

- Opciones empleadas BEDtools:

-hist: Dar lugar a un histograma de cobertura para cada función en B, así como un histograma resumido de todos los rasgos en B.

Ejecutar BEDtools para obtener la cobertura por base

- Comando genérico:

```
$ coverageBed -d -abam <input.sorted.rmdup.bam> -b <regions.bed>  
> <output.sorted.rmdup.pb.cov>
```

- Opciones empleadas BEDtools:

-d: Devuelve la profundidad de cobertura para cada posición en B. Las posiciones devueltas empiezan en 1.

-abam: El archivo de entrada A esta en formato BAM.

-b: El archivo de entrada b está dentro.

Obtención de variantes

Ejecutar SAMtools y BCFtools para obtener archivos VCF

- Comando genérico:

```
$ samtools mpileup -d[INT] -D -C50 -E -f <reference.fa.gz> -l  
<regions.bed> -u <input.sorted.rmdup.bam> | bcftools view - -v -c -g >  
<output.vcf> 2> <output.vcf.log>
```

- Opciones empleadas SAMtools:

-d (INT): En una posición determinada, leer un máximo de INT lecturas por BAM. [250]

-D: Salida de la cobertura por muestra.

-C: Coeficiente empleado para reducir la calidad de mapeo de las lecturas que contienen excesivos errores. Un valor de 0, deshabilita esta función. Si se emplea, se recomienda un valor de 50 para BWA.

-E: Computación BAQ extendida.

-f (FILE): La secuencia de referencia indexada se encuentra en formato FASTA.

-l (FILE): archivo o lista de posiciones que contiene una lista de las regiones o de los sitios donde se deben generar pileup o BCF.

-u: Similar a -g exceptuando que el output es un BCF sin comprimir.

- Opciones empleadas BCFtools:

-: El archivo de entrada viene de stdout

-v: Output con solo posiciones que contengan variantes. (force -c)

-g: Llamada de genotipos por base en los sitios que contienen variantes. (force -c)

-
- c: Llamar variantes usando inferencia Bayesiana. Esta opción invoca la opción -e.
 - e: Realiza la estimación por máxima verosimilitud, incluyendo la frecuencia alélica, el equilibrio Hardy-Weinberg y test de asociación con razones de verosimilitud.**[NOTA]**: Si se omite la opción "-v", se imprimirán todas las posiciones, no solo las que contengan variantes.
-
-

Comparación VCFs

Comprimir VCF con tabix

- Comando genérico:

```
$ bgzip <input.vcf>
```

Indexar VCF con tabix

- Comando genérico:

```
$ tabix -p [STR] <input.vcf>
```

- Opciones empleadas tabix:

-p (STR): Formato de entrada para indexar. Input format for indexing. Son valores validos: gff, bed, sam, vcf y pshtab. Esta opción no se debe aplicar junto con cualquiera de las siguientes: -s, -b, -e, -c y -0; no se utiliza para la recuperación de datos, ya que esta configuración se almacena en el índice. [gff]

Comparar archivos VCF con VCFtools

- Comando genérico:

```
$ vcf-compare <input_1.vcf> <input_2.vcf> <input_1-2.vcf> > <output.comp.txt>
```

Filtrado VCFs

- Exclusión de indels**

```
$ vcfutils --gzvcf <input.vcf> --remove-indels --recode --recode-INFO-all --out <output.vcf>
```

- Filtrado por coverage y calidad**

```
$ vcfutils --gzvcf <input.vcf> --minDP 4 --minQ 20 --recode --recode-INFO-all --out <output.vcf>
```

- **Exclusion MHC**

```
$ vcftools --vcf <input.vcf> --exclude-bed <MHC.bed> --recode --recode-INFO-all --out  
<output.vcf>
```

Anotación funcional

Preparación del archivo de entrada para VEP

- Comando genérico:

```
$ cat <input.vcf> | grep -e "1/1\|0/1" | perl -lane 'next if /^#/; print join("\t", $F[0], $F[1], $F[1],  
"$F[3]/$F[4]", "+")' > <input.vep>
```

Ejecutar VEP

- Comando genérico:

```
$ perl variant_effect_predictor.pl -i <input.vep> --cache --force_overwrite --everything --offline -o  
<output.vep>
```

VERSIONES EMPLEADAS

A continuación se exponen las versiones de los programas empleados en el análisis:

-Seqtk Version: 1.0-r31

-BWA (alignment via Burrows-Wheeler transformation) Version: 0.7.5a-r405

-Samtools (Tools for alignments in the SAM format) Version: 0.1.19-44428cd

-VCFtools (v0.1.11); BEDtools (2.17.0).