

UNIVERSIDAD POLITÉCNICA DE VALENCIA

DEPARTAMENTO DE BIOTECNOLOGÍA



ESTUDIO DE LAS BASES GENÉTICAS DE LA DIABETES TIPO 2: IDENTIFICACIÓN DE VARIANTES DE BAJA FRECUENCIA MEDIANTE SECUENCIACIÓN MASIVA

TRABAJO FIN DE MÁSTER EN BIOTECNOLOGÍA BIOMÉDICA

ALUMNO/A: **Francisco Lara Hernández**

COTUTOR/A EXTERNO: **Dr. D. Felipe Javier Chaves Martínez**

TUTOR/A ACADÉMICO: **Prof. D. José Gadea Vacas**

Curso académico: **2018/2019**

Valencia, 6 de septiembre de 2019

RESUMEN

La diabetes mellitus tipo 2 (DM2) es una de las enfermedades más frecuentes a nivel mundial, contando con más de 3.5 millones de afectados en la población española. Se trata de una patología metabólica de origen multifactorial, desencadenada por la combinación de factores ambientales y genéticos. Mientras que la obesidad representa el mayor factor de riesgo conocido, el componente genético continúa siendo desconocido en gran medida. Numerosos estudios de asociación de genoma completo (GWAS) y, posteriormente, de exoma y genoma mediante secuenciación han identificado una gran cantidad de *loci* potencialmente implicados en esta patología. Sin embargo, los polimorfismos encontrados únicamente explican el 20% de su heredabilidad. El objetivo del presente proyecto fue validar e identificar nuevas variantes, especialmente poco frecuentes y raras, mediante la secuenciación masiva de amplicones de un total de 4784 individuos pertenecientes al estudio Di@bet.es. Se seleccionaron un total de 6 genes potencialmente relacionados con la DM2 a partir de estudios previos de nuestro grupo. Tras el diseño de cebadores específicos y las pruebas pertinentes de optimización, se procedió a preparar y secuenciar las librerías mediante un sistema NovaSeq 6000 de Illumina. El posterior análisis bioinformático permitió identificar 654 variantes. Veinticuatro variantes de cambio de sentido fueron asociadas de forma significativa con alguna variable clínico-bioquímica, mientras que los SNPs rs56343424 y rs77673441 fueron asociados con la DM2 y obesidad, respectivamente. A pesar de la cantidad de variantes encontradas, es necesario realizar más estudios poblacionales que validen estos hallazgos, así como estudios funcionales que permitan dilucidar los mecanismos moleculares implicados en el desarrollo o protección frente a la DM2.

PALABRAS CLAVE

Diabetes mellitus tipo 2, estudio Di@bet.es, secuenciación de nueva generación, amplicones, polimorfismo de un solo nucleótido.

ABSTRACT

Type 2 diabetes mellitus (T2DM) is one of the most common diseases worldwide, with more than 3.5 million people affected in the Spanish population. It is a metabolic pathology of multifactorial origin, triggered by the combination of environmental and genetic factors. While obesity represents the greatest known risk factor, the genetic factor remains largely unknown. Numerous genome-wide association studies (GWAS) and, subsequently, exome and genome sequencing have identified a large amount of *loci* potentially involved in this disease. However, the polymorphisms found only explain 20% of their heritability. The aim of this project was to validate and identify new variants, especially of low frequency and rare, through targeted sequencing from 4784 individuals belonging to the Di@bet.es study. A total of 6 T2DM-related genes were selected from our previous studies. Following the design of specific primers and the relevant optimization tests, libraries were prepared and sequenced using an Illumina NovaSeq 6000 system. The subsequent bioinformatic analysis allowed the identification of 654 variants. Twenty-four missense variants were significantly associated with some clinical-biochemical variable, while SNPs rs56343424 and rs77673441 were associated with DM2 and obesity, respectively. Despite the number of variants found, it is necessary to conduct more population-based studies that can validate these findings, as well as functional studies to elucidate the molecular mechanisms involved in the development or protection against T2DM.

KEYWORDS

Type 2 diabetes mellitus, Di@bet.es study, next generation sequencing, amplicons, single nucleotide polymorphism.

AGRADECIMIENTOS

A Javier por haber confiado en mí y darme la oportunidad de formar parte de este gran equipo.

A Tani, Kike, Ana B. y Veronik por haberme ayudado y orientado durante todo el proyecto.

A Azahara por sus momentos de locura y aguantar mis incontables y repetidas preguntas.

A Claudia y Soraya por haberme hecho la estancia mucho más amena.

A Lorena, Ana, Mapi y Ángel por su apoyo incondicional.

Estén o no estén, a toda UGD.

ÍNDICE

1	INTRODUCCIÓN.....	10
1.1	ETIOLOGÍA Y EPIDEMIOLOGÍA DE LA DIABETES.....	10
1.2	DIAGNÓSTICO Y TIPOS DE DIABETES.....	11
1.3	DIABETES MELLITUS TIPO 2.....	13
1.4	ASPECTOS MOLECULARES DE LA DM2. FISIOPATOLOGÍA.....	14
1.5	ESTUDIOS GENÉTICOS.....	16
1.6	TECNOLOGÍA NGS.....	18
2	ANTECEDENTES Y OBJETIVOS.....	20
3	MATERIAL Y MÉTODOS.....	22
3.1	POBLACIÓN DE ESTUDIO.....	22
3.2	DISEÑO EXPERIMENTAL.....	22
3.3	OBTENCIÓN DE MUESTRAS.....	23
3.4	PREPARACIÓN DE LIBRERÍAS.....	23
3.4.1	DISEÑO DE PRIMERS.....	23
3.4.2	PCR 1 MULTIPLEX.....	24
3.4.3	PCR 2 ANIDADA.....	25
3.4.4	ELECTROFORESIS CAPILAR Y GENERACIÓN DEL <i>POOL</i>	26
3.4.5	PURIFICACIÓN.....	26
3.4.6	CUANTIFICACIÓN.....	27
3.4.7	SECUENCIACIÓN.....	28
3.4.8	ANÁLISIS BIOINFORMÁTICO.....	29
3.5	ANÁLISIS ESTADÍSTICO.....	30
4	RESULTADOS Y DISCUSIÓN.....	32
4.1	CARACTERÍSTICAS DE LA POBLACIÓN DI@BET.ES.....	32
4.2	DISEÑO IN SILICO DE CEBADORES.....	32
4.3	OPTIMIZACIÓN DE LAS CONDICIONES DE PCR 1.....	33

4.4	VALIDACIÓN DE LAS LIBRERÍAS DE LA POBLACIÓN DI@BET.ES	36
4.5	RESULTADOS DEL ANÁLISIS BIOINFORMÁTICO.....	36
4.5.1	CALIDAD DE LA SECUENCIACIÓN.....	36
4.5.2	COBERTURA DE LA SECUENCIACIÓN	38
4.6	ANÁLISIS DE VARIANTES.....	40
4.6.1	VALIDACIÓN DE VARIANTES CON ESTUDIOS PREVIOS BASADOS EN EXOMAS...	42
4.7	ANÁLISIS POR GEN.....	42
4.7.1	ACAD10.....	42
4.7.2	CYP1A1	44
4.7.3	GCKR.....	45
4.7.4	MANBA	47
4.7.5	SREBF2.....	49
4.7.6	TCF7L2	51
4.8	LIMITACIONES DEL ESTUDIO	53
5	CONCLUSIONES	55
6	BIBLIOGRAFÍA.....	56

ÍNDICE DE FIGURAS

Figura 1. Estimación del número de individuos adultos (20-79 años) afectados de DM por regiones y a nivel mundial en 2017 y 2045.....	11
Figura 2. Mecanismos moleculares implicados en la RI.	16
Figura 3. Representación de los posibles efectos de las variantes genéticas en función de su frecuencia.	18
Figura 4. Esquema del diseño experimental.	23
Figura 5. Representación de la separación de multiplexes..	24
Figura 6. Esquemización sobre la generación de librerías..	25
Figura 7. Fases del proceso de purificación de librerías para secuenciación.	27
Figura 8. Ejemplo de recta patrón para la cuantificación de librerías.	27
Figura 9. Resumen del proceso de secuenciación de los sistemas Illumina.....	29
Figura 10. Esquema de la pipeline bioinformática.	29
Figura 11. Resultados de electroforesis capilar para la evaluación de las 8 condiciones a testar en la PCR 1.....	34
Figura 12. Resultados de electroforesis capilar sobre las tres medidas de concentración del aditivo 1. .	34
Figura 13. Resultados de electroforesis capilar tras el reajuste de la concentración de cada pareja de cebadores y pruebas a 0.1X y 0.05X.....	35
Figura 14. Cobertura mínima antes y después del reajuste de concentración de cebadores para el gen GCKR.....	35
Figura 15. Validación del proceso de amplificación de las muestras bajo estudio..	36
Figura 16. Calidad de la secuenciación por base e individuo..	37
Figura 17. Distribución de la calidad de todas las lecturas por individuo.	37
Figura 18. Contenido de bases indeterminadas (N) y distribución del tamaño de las lecturas..	38
Figura 19. Distribución de la cobertura media del gen MANBA.....	39
Figura 20. Distribución de la cobertura media del gen SREBF2.....	39
Figura 21. Relación entre la cobertura media y la amplitud de cobertura por amplicón de todas las muestras analizadas.	40
Figura 22. Porcentaje de variantes encontradas según el tipo de mutación.	41
Figura 23. Representación del efecto de la variante P446L en el metabolismo de la glucosa.....	47
Figura 24. Representación de la activación proteolítica de las proteínas SREBP.....	49

ÍNDICE DE TABLAS

Tabla 1. Diagnóstico de la DM y prediabetes según la AAD.	12
Tabla 2. Factores de riesgo para la DM2.	13
Tabla 3. Características genómicas del panel de genes seleccionado.	20
Tabla 4. Variantes encontradas en estudios previos basados en exoma por la Unidad de Genómica y Diabetes.....	21
Tabla 5. Condiciones utilizadas en la PCR 1 o multiplex.	25
Tabla 6. Condiciones utilizadas en la PCR 2 o anidada.	26
Tabla 7. Características clínico-bioquímicas de la población Di@bet.es.	32
Tabla 8. Condiciones a testar en el proceso de optimización de PCR 1.	33
Tabla 9. Cobertura media de cada gen estudiado.	38
Tabla 10. Variantes seleccionadas tras aplicar los filtros de calidad.	41
Tabla 11. Variantes validadas con los estudios previos de caso-control en exomas.	42
Tabla 12. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen ACAD10.	43
Tabla 13. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen CYP1A1.	44
Tabla 14. Variante asociada a la DM2 mediante la prueba exacta de Fisher.	45
Tabla 15. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen GCKR.	46
Tabla 16. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen MANBA.	48
Tabla 17. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen SREBF2.	49
Tabla 18. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen TCF7L2.	52
Tabla 19. Variante asociada a la obesidad mediante la prueba exacta de Fisher.	52

ABREVIATURAS

ACAD10: Acil-CoA deshidrogenasa 10

ADN: Ácido desoxirribonucleico

CT: Colesterol total

CYP1A1: Citocromo P450 1A1

DM: Diabetes mellitus

DM1: Diabetes mellitus tipo 1

DM2: Diabetes mellitus tipo 2

GCKR: Regulador de la glucoquinasa

GWAS: Estudio de Asociación de Genoma Completo

HDL: Lipoproteína de alta densidad

HOMA-IR: Modelo Homeostático de Resistencia a la Insulina

IMC: Índice Masa Corporal

LDL: Lipoproteína de baja densidad

MAF: Frecuencia Alélica Mínima

MANBA: β -manosidasa

NGS: Secuenciación de nueva generación

PCR: Reacción en cadena de la polimerasa

RI: Resistencia a la insulina

SNP: Polimorfismo de un solo nucleótido

SREBF2: Factor de transcripción de unión al elemento regulador de esterol 2

TCF7L2: Factor de transcripción 7 tipo 2

TG: Triglicéridos

WES: Secuenciación de exoma completo

WGS: Secuenciación de genoma completo

1 INTRODUCCIÓN

1.1 ETIOLOGÍA Y EPIDEMIOLOGÍA DE LA DIABETES

La diabetes mellitus (DM) es un conjunto de enfermedades metabólicas cuyo rasgo común es una hiperglucemia crónica, resultado de alteraciones en la secreción de insulina, en el mecanismo de acción de esta hormona o la combinación de ambos factores. La desregulación de la glucemia presenta efectos a nivel celular y tisular en múltiples órganos, conllevando al desarrollo de complicaciones a largo plazo. Dichas complicaciones se clasifican como microvasculares (retinopatías, nefropatías y neuropatías) y macrovasculares (relacionadas con enfermedades cardiovasculares), siendo éstas últimas la causa más frecuente de mortalidad en diabéticos (Papatheodorou, Banach, Bekiari, Rizzo, & Edmonds, 2018; Punthakee, Goldenberg, & Katz, 2018).

La DM representa uno de los desafíos sanitarios más importantes del último siglo. Los últimos datos proporcionados por la Federación Internacional de Diabetes (FID) (Cho et al., 2017) señalan un incremento de casos de DM en adultos (20-79 años) entre 2013 y 2017, pasando de 385 a 425 millones de afectados, respectivamente. Además, las últimas estimas prevén que la cifra siga aumentando y alcance los 629 millones de diabéticos en 2045 (*Figura 1*). No obstante, esta cifra podría ser todavía mayor, puesto que alrededor del 50% de personas adultas (212 millones) desconocen que padecen la enfermedad. La DM afecta principalmente a personas adultas (20-64 años) residentes en países de ingresos bajos y medios, y muestra una prevalencia ligeramente superior en hombres (9.1%) respecto a las mujeres (8.4%). No obstante, cada vez se diagnostican más casos de DM en niños y adolescentes, principalmente por los cambios en el estilo de vida de la sociedad actual (Chatterjee, Khunti, & Davies, 2017). Esta enfermedad constituye un grave problema de salud pública, ya que se encuentra entre las 10 causas principales de muerte y supone un elevado coste económico a nivel social.

Las estimaciones realizadas en 2017 por la FID (Cho et al., 2017) indican que las regiones del Pacífico Occidental y Sudeste Asiático son las zonas con mayor número de personas diagnosticadas de DM (158 y 82 millones respectivamente). En el caso de Europa, los datos marcan una prevalencia del 8.8% (58 millones de europeos) en personas adultas, siendo la Federación Rusa y Turquía los países más afectados. En la población española se ha estimado una prevalencia del 13,8% (Soriguer et al., 2012), representando así más de 3.5 millones de españoles afectos. España se ha situado entre los 10 países dentro de Europa con el mayor número de diabéticos en edad adulta (18-99 años).

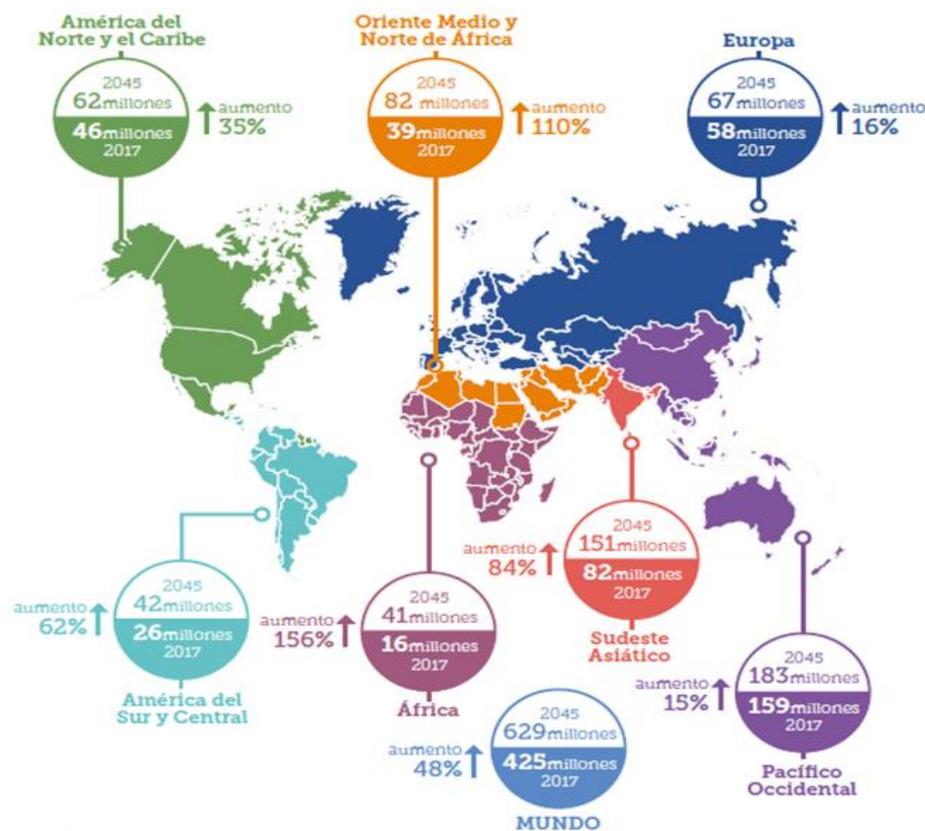


Figura 1. Estimación del número de individuos adultos (20-79 años) afectados de DM por regiones y a nivel mundial en 2017 y 2045. Extraída de (Cho et al., 2017).

1.2 DIAGNÓSTICO Y TIPOS DE DIABETES

Actualmente existen diferentes pruebas clínicas para el diagnóstico de la DM. Éstas se basan en la estimación de la glucemia del paciente, ya sea midiendo los niveles de glucosa plasmática tras 8 horas de ayuno, a través de la prueba de tolerancia a la glucosa oral (medición de la glucosa plasmática 2 horas después de la ingesta de 75 g de glucosa) o bien mediante una prueba aleatoria de sangre con niveles glucémicos superiores a 200 mg/dL. Por otro lado, se utiliza también como criterio diagnóstico la medida de los niveles de hemoglobina glicosilada (HbA1c), la cual permite analizar la glucemia a largo plazo (últimos 2-3 meses). En base a esto, los criterios vigentes para el diagnóstico de la DM establecen valores para el riesgo de sufrir complicaciones microvasculares, especialmente la retinopatía (American Diabetes Association, 2019; Nathan, 2015).

El término prediabetes hace referencia a un estado intermedio de hiperglucemia, donde los niveles de glucosa en plasma están por encima de los valores normales, pero se encuentran por debajo del límite de la DM. La prediabetes se asocia con muchas de las complicaciones microvasculares y macrovasculares que se encuentran en la DM, por lo que predispone al desarrollo de la enfermedad a largo plazo (Bansal, 2015).

La *Tabla 1* muestra los criterios diagnósticos para la DM y prediabetes según la Asociación Americana de Diabetes (AAD):

Tabla 1. Diagnóstico de la DM y prediabetes según la AAD. Adaptada de (Nathan, 2015).

	Normal	Prediabetes	Diabetes
Glucemia plasmática en ayunas	≤ 99 mg/dL	100 – 125 mg/dL	≥ 126 mg/dL
Prueba de tolerancia a la glucosa oral	≤ 139 mg/dL	140 – 199 mg/dL	≥ 200 mg/dL
Glucemia aleatoria + sintomatología hiperglucémica	-	-	≥ 200 mg/dL
HbA1c	≤ 5.6%	5.7 – 6.4%	≥ 6.5%

Tras el diagnóstico, establecer una clasificación adecuada de la DM es fundamental para determinar el tratamiento correcto. Sin embargo, no siempre es una tarea sencilla debido a la heterogeneidad de los síntomas, por lo que muchos pacientes son sometidos a una revisión. En el año 1997, la AAD estableció diferentes grupos atendiendo a su etiología, clasificación que sigue vigente en la actualidad (American Diabetes Association, 2019; Kharroubi & Darwish, 2015; Yeung, Hannah-Shmouni, Niederhoffer, & Walker, 2018):

1. Diabetes mellitus tipo 1 (DM1). Representa el 5-10% de los casos, afectando en gran medida a niños y adolescentes (80-90%). Su origen radica en la destrucción autoinmunitaria de las células β pancreáticas y, por consiguiente, una carencia en la secreción de insulina. Los afectados de DM1 son insulino dependientes.
2. Diabetes mellitus tipo 2 (DM2). Es el tipo más frecuente, constituyendo el 90-95% de los casos. Afecta principalmente a personas en edad adulta y su etiología está asociada con la resistencia a la insulina (RI) y un descenso progresivo en la secreción de ésta.
3. Diabetes tipo LADA (del inglés *Latent Autoimmune Diabetes in Adults*). Comparte la etiología de carácter autoinmunitario con la DM1. No obstante, se presenta en el grupo de edad propio de la DM2 con algunos factores similares a ésta. Por esta razón, es considerada como un tipo intermedio entre ambas, llevando en muchas ocasiones a un diagnóstico erróneo.
4. Diabetes gestacional. Se define como un estado de intolerancia a la glucosa cuyo inicio o primer diagnóstico tiene lugar durante el embarazo. A menudo desaparece una vez ha finalizado el mismo, pero haber padecido esta diabetes supone un riesgo elevado de desarrollo de DM2 en el futuro.

5. Existen otros tipos de DM de etiología variada. Entre ellas podemos destacar las diabetes monogénicas, las cuales representan una fracción inferior al 5% de los casos diagnosticados. Se deben principalmente a mutaciones puntuales que resultan en la destrucción de las células β pancreáticas o un descenso en el número de éstas, como son la diabetes neonatal y MODY (del inglés *Maturity Onset Diabetes of the Young*); mutaciones en genes mitocondriales (diabetes mitocondrial) o incluso mutaciones cuyo resultado es una insulina truncada. Por último, se ha observado la asociación entre DM y ciertos síndromes como Prader-Willi, Cohen, Wolfram, etc.

1.3 DIABETES MELLITUS TIPO 2

La diabetes mellitus tipo 2 (DM2) ha alcanzado dimensiones epidémicas en los últimos años, contribuyendo en gran medida a la mortalidad y discapacidad mundial (Zheng, Ley, & Hu, 2018). En España se ha estimado la incidencia de la DM2 en 11.58 casos por cada 1000 personas al año (2016-2017), gracias al estudio di@bet.es llevado a cabo por el CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM) y presentado en el XXIX Congreso Nacional de la Sociedad Española de Diabetes (CIBERDEM, 2018).

La DM2 es una patología multifactorial originada por la combinación de factores ambientales y genéticos. Se han descrito multitud de factores relacionados con el estilo de vida que guardan una asociación directa con el desarrollo de la enfermedad (*Tabla 2*). No obstante, una dieta saludable acompañada de ejercicio físico y evitando el consumo de ciertas sustancias ayuda a disminuir el riesgo de padecer DM2 (Kolb & Martin, 2017; Olokoba, Obateru, & Olokoba, 2012).

Tabla 2. Factores de riesgo para la DM2. Adaptada de (Zheng et al., 2018).

Edad avanzada	Consumo de alcohol y/o tabaco
Antecedentes familiares de DM2	Estrés o depresión
Ascendencia caucásica	Presión sanguínea alta
Sobrepeso u obesidad	Niveles altos de triglicéridos
Dieta poco saludable	Niveles bajos de colesterol HDL
Vida sedentaria	Medicamentos como estatinas o tiazidas
Antecedentes de diabetes gestacional	Estatus socioeconómico bajo

La clínica y progresión de la DM2 varían de forma considerable entre individuos, pudiendo ser asintomáticos durante las primeras etapas, hecho que dificulta el diagnóstico. La DM2 se caracteriza por una alteración en el metabolismo de carbohidratos, lípidos y proteínas como consecuencia del incremento de la glucemia y/o de la RI. Entre los síntomas encontramos micción frecuente y excesiva (poliuria), aumento de la sed (polidipsia), pérdida de peso, visión borrosa, cansancio, etc. La progresión de la enfermedad, sin prevención y/o tratamiento, conlleva al desarrollo de complicaciones microvasculares y macrovasculares mencionadas con anterioridad (DeFronzo et al., 2015; Olokoba et al., 2012).

En cuanto al tratamiento, por lo general los afectos de DM2 no son dependientes de insulina, pero tienen acceso a diferentes fármacos capaces de regular algunos efectos de esta patología. Cabe destacar el antidiabético oral metformina, ya que es la terapia de primera línea. Éste actúa disminuyendo la salida de glucosa hepática y, por tanto, los niveles de este azúcar en sangre. Sin embargo, en muchas ocasiones está contraindicada y se recurre a otros fármacos como las sulfonilureas o tiazolidinedionas (DeFronzo et al., 2015).

1.4 ASPECTOS MOLECULARES DE LA DM2. FISIOPATOLOGÍA

El páncreas endocrino está constituido por los islotes de Langerhans, un conjunto de células encargadas de la producción de diversas hormonas. Entre las células residentes encontramos las células α y β , las cuales sintetizan y liberan glucagón e insulina, respectivamente. Estas dos hormonas presentan efectos antagónicos: la insulina actúa regulando la entrada de glucosa en las células, mientras que el glucagón estimula la salida de este azúcar al torrente sanguíneo, consiguiendo una regulación precisa de la glucemia en todo el organismo (Rojas et al., 2018).

En el caso de la DM2, la RI juega un papel fundamental en el desarrollo de la enfermedad, produciéndose por la combinación de factores ambientales (Ej. Obesidad visceral) y genéticos (Ej. Variantes genéticas que modulan el receptor de insulina y/u otros elementos de la cascada de señalización). La RI se caracteriza por una ausencia o baja respuesta a la insulina, principalmente por el tejido hepático, músculo y tejido adiposo, que confluente en un incremento de la demanda de esta hormona (DeFronzo et al., 2015).

En condiciones normales, las células β pancreáticas integran señales (Ej. Sustratos como glucosa, ácidos grasos o aminoácidos) que les permiten liberar la cantidad de insulina necesaria para suplir la demanda de cada momento (ciclos ayuno-ingesta, ejercicio, estrés) (DeFronzo et al., 2015). La insulina ejerce su acción biológica al unirse a su receptor de membrana, induciendo de esta forma la autofosforilación de residuos de tirosina localizados en la región citosólica del

receptor, los cuales son reconocidos por sustratos del receptor de insulina (*IRS-1 e IRS-2*). Finalmente, los IRS actúan como adaptadores moleculares que inducen la activación de la vía de señalización PI3K/Akt, responsable de acciones metabólicas como la movilización de transportadores de glucosa a la membrana, glucogénesis, adipogénesis y síntesis de proteínas. La señalización desencadenada por la insulina, mediada por otros adaptadores moleculares, también activa la vía MAPK/Ras, encargada de activar la expresión de genes relacionados con el crecimiento, proliferación y diferenciación celular (Gutiérrez-Rodelo, Roura-Guiberna, & Olivares-Reyes, 2017).

En el contexto de la DM2, la RI provoca un aumento de la secreción de insulina por parte de las células β , en un proceso denominado hiperinsulinemia compensatoria. Actualmente se conocen varios mecanismos extrínsecos e intrínsecos implicados en el desarrollo de RI (*Figura 2*). El principal factor extrínseco es la obesidad, la cual conlleva una expansión del tejido adiposo y la secreción de citoquinas proinflamatorias (TNF- α , IL-6, IL-18 e IL-1B) y ácidos grasos libres (acumulación de diacilglicerol y ceramidas). Las citoquinas proinflamatorias inducen RI debido a que influyen de forma indirecta en la fosforilación de residuos de serina en las IRS, llevando a su inactivación e impidiendo la acción de la insulina. La acumulación de diacilglicerol implica la fosforilación de la proteína quinasa C (PKC) y, por tanto, la inhibición del receptor de insulina; mientras que las ceramidas activan a los receptores TLR-4, los cuales terminan regulando de forma negativa la vía PI3K/Akt. En cuanto a los mecanismos intrínsecos encontramos la disfunción mitocondrial (bien por el aumento de especies reactivas del oxígeno (EROs) o por descenso de la oxidación metabólica de ácidos grasos) y el estrés en el retículo endoplasmático (RE, como alteraciones en el plegamiento de proteínas o en la vía mTOR). En ambos casos se induce la fosforilación de residuos de serina en las IRS, produciendo su inactivación e inhibiendo la señalización de la insulina (DeFronzo et al., 2015; Gutiérrez-Rodelo et al., 2017).

Finalmente, algunos de los mecanismos implicados en la RI (Ej. Estrés oxidativo, estrés en el RE, entre otros) inducen la muerte, ya sea por apoptosis o autofagia, de las células β pancreáticas. De hecho, en pacientes con DM2 se han observado reducciones superiores al 30% de estas células, y su progresiva desaparición o inactivación lo que conduce a una progresiva menor secreción de insulina (DeFronzo et al., 2015; Rojas et al., 2018).

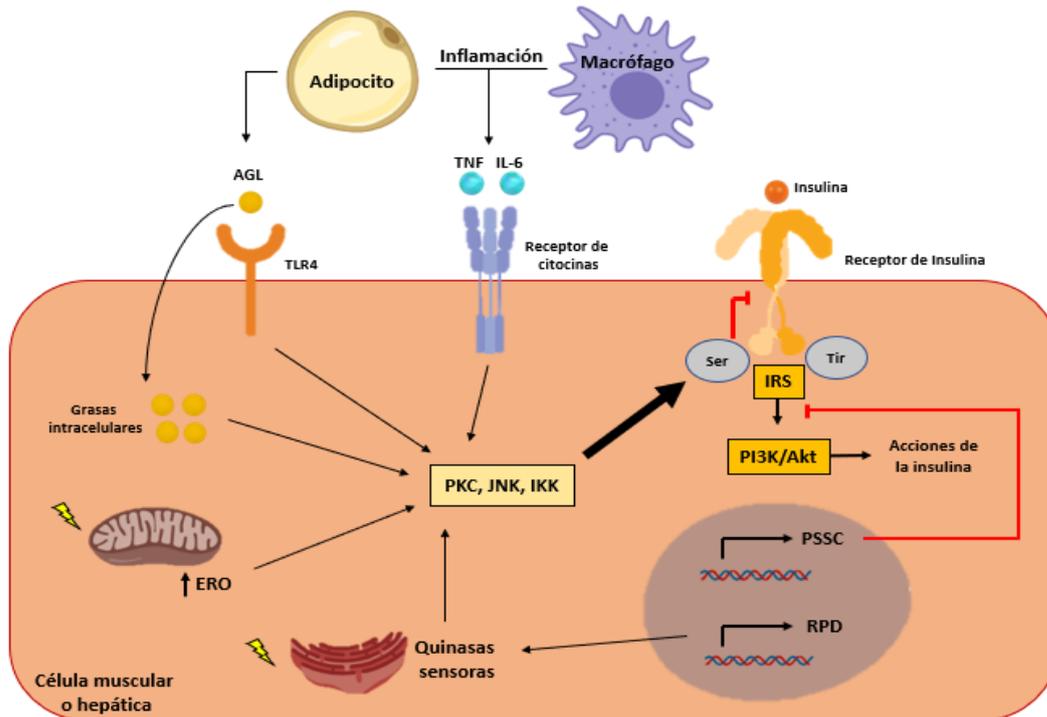


Figura 2. Mecanismos moleculares implicados en la RI. El incremento de adipocitos a causa de la obesidad, y la infiltración de macrófagos en el tejido adiposo conducen al desarrollo de RI e inflamación por la mayor secreción de AGL y citocinas. La señalización por parte de estas moléculas promueve la activación de diferentes quinasas, cuya acción culmina con la fosforilación de residuos de serina en los IRS e impidiendo la acción de la insulina. De la misma forma, la presencia de ERO por alteraciones en las mitocondrias o estrés en el RE también conducen de forma indirecta a la inhibición de los IRS y por tanto al desarrollo de RI. AGL: ácidos grasos libres; ERO: especies reactivas del oxígeno; PSSC: proteínas supresoras de la señalización de citocinas; RPD: respuesta a proteínas desplegadas. Adaptada de (DeFronzo et al., 2015).

1.5 ESTUDIOS GENÉTICOS

La obesidad o falta de actividad física son algunos factores asociados al desarrollo de DM2. Sin embargo, los factores ambientales no afectan del mismo modo a toda la población debido a que el riesgo de susceptibilidad parece ser heredado (Ali, 2013). Varios estudios poblacionales, en familias y gemelos han permitido establecer una estimación de la heredabilidad de la DM2, siendo de hasta el 80%. Se ha estimado que el riesgo de desarrollar DM2 cuando uno de los padres se encuentra afectado es del 40%, mientras que cuando ambos padres lo están este porcentaje alcanza el 70%. Estudios en gemelos muestran una tasa de concordancia en gemelos dicigóticos del 20-30%, mientras que en gemelos monocigóticos aumenta hasta el 70%; llegando al 100% cuando se realizan seguimientos a largo plazo (Kaprio et al., 1992; Tillil & Kobberling, 1987). En todos los casos, se pone de manifiesto la importancia de la genética en el desarrollo de la enfermedad.

El carácter poligénico de la enfermedad hace que la búsqueda de genes de susceptibilidad para DM2 no sea una tarea sencilla. El análisis de ligamiento fue la primera estrategia empleada, permitiendo la identificación de dos genes relacionados con la DM2: calpaína 10 (*CAPN10*) y

factor de transcripción 7 tipo 2 (*TCF7L2*). Mientras que la asociación del primero no es consistente debido a los resultados negativos obtenidos en diferentes estudios de asociación de genoma completo (GWAS, del inglés *Genome-Wide Association Study*), el gen *TCF7L2* es hasta la fecha el que mayor asociación presenta, mostrando un riesgo relativo de 1.4 y confirmado en diferentes grupos étnicos. Más tarde, el estudio de genes candidatos, relacionados con la enfermedad por su función biológica, permitió hallar nuevos *loci* relacionados con la DM2 como son el receptor γ activado por el proliferador de peroxisoma (*PPAR- γ*), un canal rectificador de potasio (*KCNJ11*), *IRS1*, entre otros (Ahlqvist, Ahluwalia, & Groop, 2011; Prasad & Groop, 2015).

El análisis de ligamiento y el estudio de genes candidatos han permitido identificar varios genes asociados a la DM2, aun así, su contribución a la heredabilidad de la enfermedad sigue siendo pequeña. La revolución en este campo llegó con el desarrollo de las tecnologías de genotipado de alto rendimiento que, junto con herramientas como HapMap o el proyecto 1000 genomas, lograron genotipar miles de polimorfismos de un solo nucleótido (SNPs, del inglés *Single Nucleotide Polymorphism*) de forma simultánea. Hasta el momento, los estudios GWAS y los posteriores metaanálisis reúnen más de 100 regiones genómicas de riesgo para la DM2 (Ali, 2013; Langenberg & Lotta, 2018).

Hasta hace algunos años, la mayoría de los estudios de asociación han operado bajo la premisa “enfermedad común – variante común”, en la que la predisposición a una determinada enfermedad compleja, en este caso la DM2, deriva de un número moderado de variantes comunes (frecuencia del alelo menos común, “minor allele frequency” o $MAF \geq 0.05$), aun presentando efectos modestos. No obstante, las variantes significativas asociadas a la DM2 son capaces de explicar hasta el 20% de su heredabilidad, dejando un 80-90% de heredabilidad perdida, es decir, la parte de variación genética que no puede ser explicada por estos polimorfismos. Este hecho ha impulsado grandes esfuerzos para resolver dicho problema, surgiendo diversas teorías que podrían dar una explicación: presencia de variaciones estructurales, interacciones gen-gen y gen-ambiente, efectos epigenéticos o variantes de baja frecuencia con fuerte efecto fenotípico (Dziewulska, Dobosz, & Dobrzyn, 2018; Prasad & Groop, 2015; Stančáková & Laakso, 2016).

Entre las teorías propuestas, una de las más plausibles en la DM2 sería la de “variante rara – enfermedad común”. Las variantes raras ($MAF < 0.01$) o poco frecuentes ($0.01 < MAF < 0.05$) suelen ser sucesos más recientes en la evolución, además de estar mayormente representadas dentro del genoma humano ya que la selección natural se encargaría de eliminar aquellos alelos deletéreos antes de que alcancen una frecuencia mayor. En este sentido, esta premisa apunta a

que parte de la etiología de las enfermedades multifactoriales puede ser debida a las variantes raras, cuyo efecto conjunto compensaría su baja frecuencia (*Figura 3*). La búsqueda de variantes raras a través de estudios de asociación ha sido posible gracias al desarrollo de las tecnologías de secuenciación, especialmente de la secuenciación de genoma y exoma completos (WGS y WES, del inglés *Whole Genome Sequencing* y *Whole Exome Sequencing*, respectivamente). Incluso se han realizado chips que incluyen parte de estas variantes. Estas nuevas estrategias permiten introducir una mayor cobertura y resolución a un coste más asequible, además de aportar un mayor poder estadístico ya que permiten la secuenciación de miles de individuos paralelamente (Dziewulska et al., 2018; Prasad & Groop, 2015; J. Wang et al., 2016).

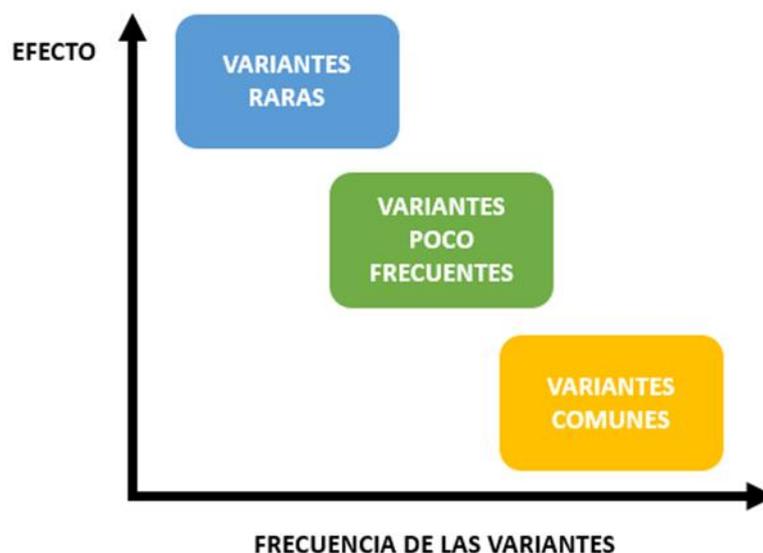


Figura 3. Representación de los posibles efectos de las variantes genéticas en función de su frecuencia. Adaptada de (Bonnetfond & Froguel, 2015).

1.6 TECNOLOGÍA NGS

Las tecnologías de secuenciación representan uno de los grandes avances en el mundo de la genómica, ya que no solo utilizan la composición del ADN para describir el código genético de todo organismo vivo, sino que también tienen aplicaciones prácticas como por ejemplo estudiar la etiología y tratamiento de enfermedades genéticas. Las primeras tecnologías de secuenciación surgieron en 1977, siendo la técnica de secuenciación Sanger la más utilizada durante muchos años gracias a su elevada eficiencia y baja radioactividad. Está basada en el uso de dideoxinucleótidos (ddNTPs) marcados fluorescentemente, los cuales se van añadiendo a un molde de ADN durante un primer paso de amplificación. Posteriormente se termina generando una escalera de fragmentos génicos que se distinguen entre sí por una base de longitud y que pueden ordenarse a través de una electroforesis capilar. La secuenciación por Sanger fue la

técnica de referencia durante más de 3 décadas. No obstante, su rendimiento es limitado cuando se trata de estudios a gran escala debido a que únicamente trabaja con amplicones individuales, lo que incrementa el coste y el tiempo empleado (Kchouk, Gibrat, & Elloumi, 2017; Slatko, Gardner, & Ausubel, 2018).

La revolución en las técnicas de secuenciación de ADN surge con la aparición de las tecnologías de secuenciación de nueva generación (NGS, del inglés *Next Generation Sequencing*), divididas en segunda y tercera generación, y caracterizadas por ser capaces de procesar millones de lecturas en paralelo, presentar un coste más asequible y un mayor rendimiento. Las plataformas de secuenciación de segunda generación pueden clasificarse en función de la química que utilicen: Roche/454 (2005), Illumina/Solexa (2006) e Ion Torrent (2010) utilizan la secuenciación por síntesis, mientras que ABI/SOLiD se basa en la secuenciación por ligación (Kchouk et al., 2017; Singh, 2017; Slatko et al., 2018).

Las plataformas de segunda generación han supuesto una auténtica revolución en el análisis del ADN. No obstante, éstas precisan de un trabajo laborioso debido a que necesitan la generación de librerías, paso previo a la secuenciación y que en muchos casos supone un importante incremento del coste y el tiempo de ejecución. La secuenciación de tercera generación puede eliminar algunos problemas de las técnicas previas (generación de librerías, longitud de las lecturas, etc.). La generación de lecturas de mayor tamaño implica una mejora en la resolución del ensamblaje y una reducción del problema que representan las regiones repetitivas y variaciones estructurales en el genoma, una clara desventaja en la secuenciación de segunda generación. Existen diferentes estrategias de secuenciación de tercera generación: la secuenciación a tiempo real de una única molécula (SMRT, del inglés *Single Molecule Real Time Sequencing*) a través de nucleótidos marcados fluorescentemente y desarrollada por Pacific Biosciences, la secuenciación del ADN a través de un poro a nanoescala llevada a cabo por Oxford Nanopore y la secuenciación por síntesis de lecturas largas usadas por 10xGenomics (Kchouk et al., 2017; Slatko et al., 2018).

2 ANTECEDENTES Y OBJETIVOS

En el exoma, incluyendo las regiones intrónicas contiguas, se encuentra el 90-95% de las variantes genéticas conocidas responsables de las enfermedades hereditarias humanas. Esta región alberga una gran cantidad de polimorfismos con un posible efecto funcional, muchos de ellos desconocidos o no investigados con relación a la DM2. Más del 50% de éstos son raros o poco frecuentes, por lo que es fundamental su estudio en enfermedades complejas. Los estudios GWAS realizados hasta el momento no han permitido la identificación de variantes poco frecuentes, razón por la cual el estudio del exoma se ha convertido en una estrategia más sencilla y rentable, a diferencia del estudio del genoma completo (Bodmer & Bonilla, 2008).

Por lo anterior, el grupo de investigación Unidad de Genómica y Diabetes (UGD) de INCLIVA ha llevado a cabo diversos proyectos con el fin de describir las bases genéticas de la DM2. Se realizaron dos estudios basados en la secuenciación de exoma para la identificación de variantes genéticas poco frecuentes y raras en casi 600 exomas (289 casos de DM2 y 289 controles sanos) pertenecientes a la población española (Galán Chilet, 2016; Martínez, 2015). Entre ambos estudios se seleccionaron más de 150 variantes ($MAF < 0,01$) distribuidas en más de 140 genes, los cuales podrían estar implicados en el desarrollo o protección de la enfermedad debido a su función biológica: transcripción, metabolismo, exocitosis, etc. Estos estudios han permitido seleccionar un panel de genes (*Tabla 3*) con el fin de identificar nuevas variantes y validar las descritas anteriormente gracias a la tecnología de secuenciación masiva de amplicones.

Tabla 3. Características genómicas del panel de genes seleccionado.

GEN	TRANSCRITO	CROMOSOMA	COORDENADAS	N.º EXONES
ACAD10	ENST00000313698	12	112,123,902 - 112,194,901	21
CYP1A1	ENST00000395048	15	75,011,883 - 75,017,951	7
GCKR	ENST00000264717	2	27,719,709 - 27,746,551	19
MANBA	ENST00000226578	4	103,553,050 - 103,682,151	17
SREBF2	ENST00000361204	22	42,229,109 - 42,303,312	19
TCF7L2	ENST00000369397	10	114,710,211 - 114,927,437	14

En el presente trabajo también se incluyó el análisis de *SREBF2* como gen candidato, ya que a través de estudios paralelos llevados a cabo en la UGD se ha encontrado una variante genética rara que podría estar directamente relacionada con la DM2 y parámetros relacionados.

Para ello, se establecieron los siguientes objetivos parciales:

- Realizar un diseño de cebadores para estudiar la región exónica del panel de genes seleccionado, así como la posterior optimización de las librerías de amplicones a secuenciar.
- Secuenciar 4784 individuos pertenecientes al estudio Di@bet.es utilizando el panel previamente diseñado.
- Validar los polimorfismos previamente encontrados en estudios basados en exoma (*Tabla 4*) e identificar nuevas variantes no descritas hasta el momento.
- Analizar la posible relación entre los genotipos y ciertas variables bioquímicas (rasgos glucémicos, perfil lipídico, etc.), y su posible asociación con la DM2.

Tabla 4. Variantes encontradas en estudios previos basados en exoma por la Unidad de Genómica y Diabetes.

Gen	cr:posición	rs	Alelos	Consecuencia	MAF (EUR)
ACAD10	12:112167654	rs146335507	C/T	Ganancia codón de parada	0.0001
CYP1A1	15:75015234	-	G/A	Ganancia codón de parada	-
	15:75012979	rs41279188	G/T	Cambio de sentido	0.005
GCKR	2:27745372	rs146053779	C/T	Ganancia codón de parada	0.004
TCF7L2	10:114925406	rs77673441	C/G	Cambio de sentido	0.005
MANBA	4:103771478	rs12644381	A>G	Variante 3' UTR	0.493
	4:103771116	rs7690700	T/C	Variante 3' UTR	0.493
	4:103772591	rs735403	C/T	Variante intrónica	0.493
	4:103775364	rs909349	T/A-C	Variante intrónica	0.493
	4:103776125	rs2866413	G/A-C	Cambio de sentido	0.493
	4:103830959	rs227368	C/T-G-A	Cambio de sentido	0.497

MAF: Frecuencia Alélica Mínima

3 MATERIAL Y MÉTODOS

3.1 POBLACIÓN DE ESTUDIO

Las muestras seleccionadas provenían del estudio Di@bet.es, un estudio poblacional, nacional y transversal que fue impulsado entre 2009 y 2010 con el objetivo de examinar la prevalencia de la obesidad, DM y otros factores de riesgo cardiovasculares en la población española. La elección de los participantes se realizó a través de un muestreo aleatorio contando con más de 100 centros de salud repartidos por todo el territorio español. De los más de 10000 candidatos, solo el 55.8% fueron examinados y posteriormente, el 9.9% de ellos fueron excluidos por institucionalización, enfermedad grave, embarazo o parto reciente, obteniendo un muestreo final de 5072 individuos mayores de edad (Soriguer et al., 2012). El estudio fue aprobado por los correspondientes comités de investigación y éticos, y todos los participantes dieron su consentimiento informado.

3.2 DISEÑO EXPERIMENTAL

En el presente proyecto se analizaron varios polimorfismos mediante NGS dirigida, utilizando concretamente la tecnología de secuenciación por amplicones (del inglés *targeted sequencing*). Se trata de una técnica que permite la secuenciación de variantes exónicas, intrónicas e intergénicas en regiones seleccionadas del genoma. A diferencia de las estrategias WGS y WES, que son más costosas y requieren un amplio manejo de datos, la secuenciación dirigida requiere un análisis bioinformático más moderado y es una opción más rentable, sobre todo cuando se trata de secuenciar grandes cantidades de individuos y se han identificado las regiones de interés.

El diseño experimental se dividió en tres fases (*Figura 4*). En la primera fase se llevó a cabo el diseño de cebadores y la preparación del material genómico a la concentración deseada. En segundo lugar, se procedió a preparar las librerías de amplicones realizando dos amplificaciones por reacción en cadena de la polimerasa (PCR, del inglés *Polymerase Chain Reaction*): una primera PCR selectiva que permitió seleccionar la región genómica concreta a estudiar a través de cebadores específicos e incluir una cola de nucleótidos, y una segunda PCR nominativa cuyo objetivo era identificar a cada una de las muestras mediante etiquetas moleculares (del inglés *Barcodes*) e incluir todas las secuencias necesarias para la reacción de NGS a partir de las colas introducidas en la primera PCR. Posteriormente, tras la purificación y cuantificación de las librerías de amplicones, se procedió a su secuenciación. Por último, en la tercera fase se realizó el análisis bioinformático que permitió la identificación de polimorfismos candidatos.



Figura 4. Esquema del diseño experimental. El diseño constó de 3 fases: diseño de cebadores y preparación del material genómico, preparación de librerías y secuenciación, y análisis bioinformático.

3.3 OBTENCIÓN DE MUESTRAS

El ADN genómico de la población de estudio se extrajo previamente a partir de muestras de sangre, utilizando diferentes sistemas de extracción en el CIBERDEM (Madrid). El ADN para las pruebas de optimización fue extraído a través del kit ADN Maxwell® (Promega®, USA) en el Hospital Clínico de Valencia.

Dichas muestras se encontraban a diferentes concentraciones de partida, por lo que se realizaron diluciones de éstas en placas nuevas con el objetivo de que todas estuviesen a 5 ng/μL, siendo una concentración adecuada para proceder a la preparación de librerías.

3.4 PREPARACIÓN DE LIBRERÍAS

3.4.1 DISEÑO DE PRIMERS

La obtención de secuencias genómicas se llevó a cabo empleando las bases de datos COSMIC (v89, mayo 2019) y Ensembl (v97, julio 2019), utilizando la versión del genoma de referencia GRCh37/hg19 y aplicando el filtro MAF>0,01 (se muestran las variantes con un MAF superior al 1%). A continuación, para el diseño *in silico* de los cebadores se utilizaron los *softwares* online ExonPrimer (<https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html>) y Primer3 (v0.4.0., <http://bioinfo.ut.ee/primer3-0.4.0/>).

Las condiciones generales del diseño de cebadores fueron: tamaño del producto de alrededor de 150 pares de bases (pb), una longitud de cebadores de 18 a 22 bases, una temperatura de *melting* de 58-62°C con una diferencia máxima de un grado entre cebadores de cada pareja, y un porcentaje de GC del 20-80%.

Reunidas todas las secuencias de cebadores y tras añadir las colas universales, se analizaron las posibles interacciones que pudiesen darse entre ellos en la región terminal 3', lo que culminaría afectando a su actividad. Para este paso se empleó el *software* FastPCR (Kalendar, Lee, & Schulman, 2009).

Finalmente, los cebadores se separaron en dos multiplexes (multiplex A y B). El fundamento de este paso era impedir la amplificación de parejas de cebadores contiguas o solapantes en una misma reacción, ya que por la propia naturaleza inespecífica de la enzima polimerasa se formarán fragmentos que no son de interés, produciendo un sesgo posterior en los resultados (Figura 5).

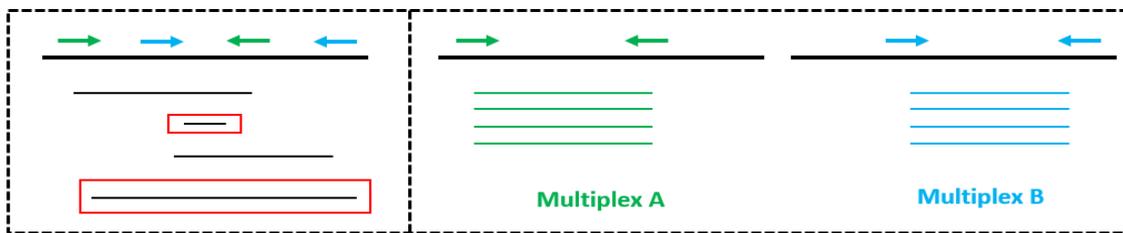


Figura 5. Representación de la separación de multiplexes. Izquierda: dos parejas de cebadores contiguas amplificadas en la misma reacción de PCR generan productos inespecíficos (recuadros rojos). Derecha: obtención de los productos de PCR deseados gracias a la división de las parejas de cebadores en dos multiplexes diferentes.

3.4.2 PCR 1 MULTIPLEX

Una vez clasificadas las parejas de cebadores en dos multiplexes se procedió a realizar la primera PCR selectiva. Para ello, las muestras fueron amplificadas en dos reacciones de PCR independientes, en una se utilizó la multiplex A de cebadores y en la otra la multiplex B, ambas bajo las mismas condiciones.

Las reacciones de PCR se realizaron empleando una enzima comercial y diferentes aditivos, y se utilizaron los termocicladores de placas 384 Veriti® (AppliedBiosystems®, California) para llevar a cabo la amplificación. Durante el proceso de puesta a punto se probaron diferentes concentraciones de cebadores y aditivos.

La PCR se inició a 95°C durante 20 minutos con objeto de activar la enzima polimerasa. A continuación, se produjo la desnaturalización de las cadenas de ADN a 98°C, seguido del paso de hibridación a 55°C y 60°C. Seguidamente, la extensión se produjo a 72°C durante 1 minuto. Los pasos de desnaturalización, hibridación y extensión se replicaron durante 38 ciclos. La extensión final se realizó a 72°C durante 5 minutos (Tabla 5).

Tabla 5. Condiciones utilizadas en la PCR 1 o multiplex.

PROGRAMA	TEMPERATURA (°C)	TIEMPO	CICLOS
Activación inicial	95	20 min	
Desnaturalización	98	30 seg	38x
Hibridación	55	1 min	
	60	1 min	
Extensión	72	1 min	
Extensión final	72	5 min	
Mantenimiento	15	∞	

3.4.3 PCR 2 ANIDADA

Para la PCR anidada se juntaron los dos productos amplificados (multiplex A + multiplex B) correspondientes a la misma muestra, y se llevó a cabo una segunda amplificación, cuyo objeto fue añadir los adaptadores, los cuales incluían la secuencia identificadora o *barcode* (permite analizar a cada sujeto por separado) y otras secuencias adaptadoras específicas para el proceso de secuenciación (Figura 6).

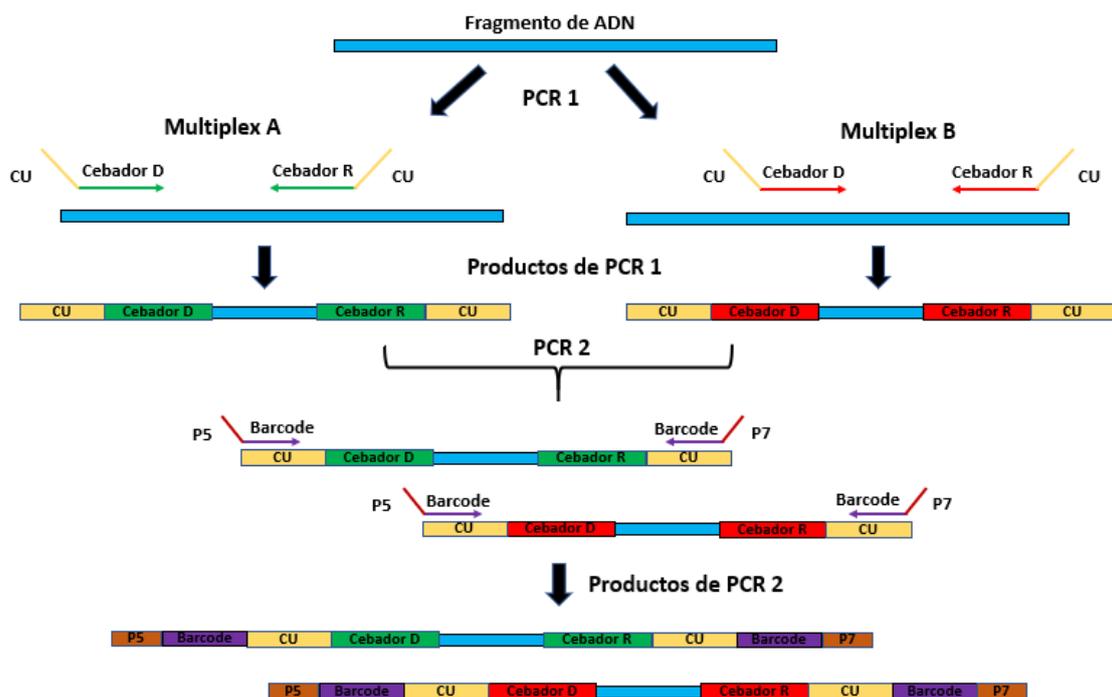


Figura 6. Esquematización sobre la generación de librerías. Durante la PCR 1 se consigue la amplificación de las regiones de interés gracias al uso de cebadores específicos (separados en dos multiplexes), los cuales llevan incorporados las CU. En la PCR2 se incorporan los barcodes y secuencias adaptadoras P5 y P7, que van a permitir el análisis de cada muestras de forma individual. Cebador D: Cebador directo; Cebador R: Cebador reverso; CU: Colas universales.

En la *Tabla 6* se muestran las distintas fases y condiciones utilizadas durante la amplificación en la PCR 2 o anidada:

Tabla 6. Condiciones utilizadas en la PCR 2 o anidada.

PROGRAMA	TEMPERATURA (°C)	TIEMPO	CICLOS
Activación inicial	95	15 min	
Desnaturalización	98	20 seg	35x
Hibridación	60	30 seg	
Extensión	72	1 min	
Extensión final	72	15 min	
Mantenimiento	15	∞	

3.4.4 ELECTROFORESIS CAPILAR Y GENERACIÓN DEL *POOL*

Finalizadas las dos PCRs, se procedió a comprobar la correcta amplificación de los fragmentos génicos empleando un paso de electroforesis capilar. Concretamente, se utilizó el sistema Qiaxcel (Qiaxcel DNA Screening Kit, Qiagen®, Alemania), el cual permite analizar hasta 96 reacciones de forma simultánea, en un periodo moderado de tiempo y con alta sensibilidad.

Tras la visualización de los productos de PCR con la electroforesis, se procedió a generar la mezcla de muestras o *pool* añadiendo 2 µL de cada muestra amplificada en un tubo Falcon. Gracias al uso de *barcodes*, las muestras podrán distinguirse en el posterior análisis.

3.4.5 PURIFICACIÓN

La purificación del *pool* tiene como objetivo la eliminación de restos de reactivos, productos de PCR inespecíficos y cebadores que no han hibridado o han dimerizado.

Para llevar a cabo la purificación se empleó el kit comercial Magsi-NGS PREP® (MagnaMedics Diagnostics B.V., Países Bajos), el cual se compone de bolas magnéticas que garantizan una elevada eficiencia a la hora de seleccionar tamaños y la limpieza de librerías destinadas a la secuenciación. El kit fue optimizado por el personal de UGD para la eliminación de productos de PCR menores de 200 pb. Primero, se añadieron las bolas magnéticas cargadas positivamente (en condiciones limitantes) sobre las muestras de ADN (cargado negativamente) provocando la unión de éstos por afinidad de cargas, por lo que aquellos fragmentos de mayor tamaño (y mayor número de cargas negativas) quedarán unidos a dichas bolas. Tras distintos lavados con etanol y sobre una placa magnética, se añadió agua, un potente nucleófilo que desplazará al ADN de las bolas haciendo que pase al sobrenadante y permitiendo su recuperación.

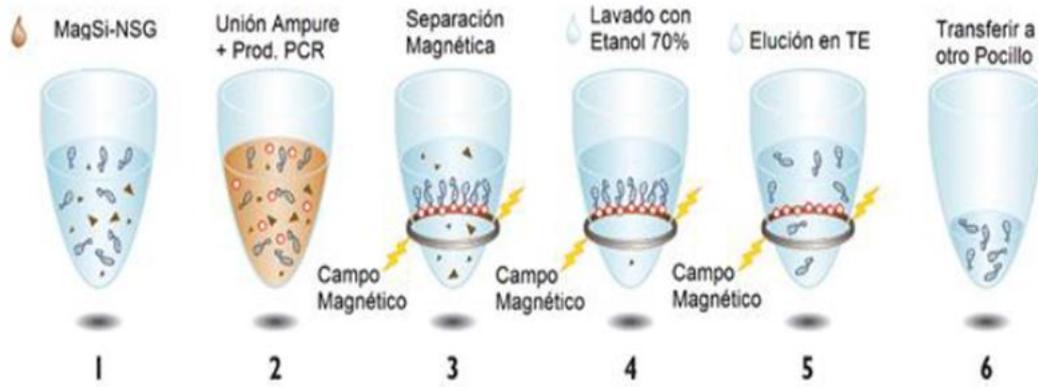


Figura 7. Fases del proceso de purificación de librerías para secuenciación. Extraída de (Seda García, 2018).

3.4.6 CUANTIFICACIÓN

Para la cuantificación se empleó el kit QuantiFluor® dsDNA System (Promega, USA). Este kit proporciona un colorante fluorescente capaz de unirse a ADN de doble cadena, presentando unos rangos de excitación y emisión de 504 nm y 531 nm, respectivamente. Permite la cuantificación de pequeñas cantidades de ADN en un rango entre 0.05 y 200 ng gracias a la interpolación con una recta patrón (Figura 8).

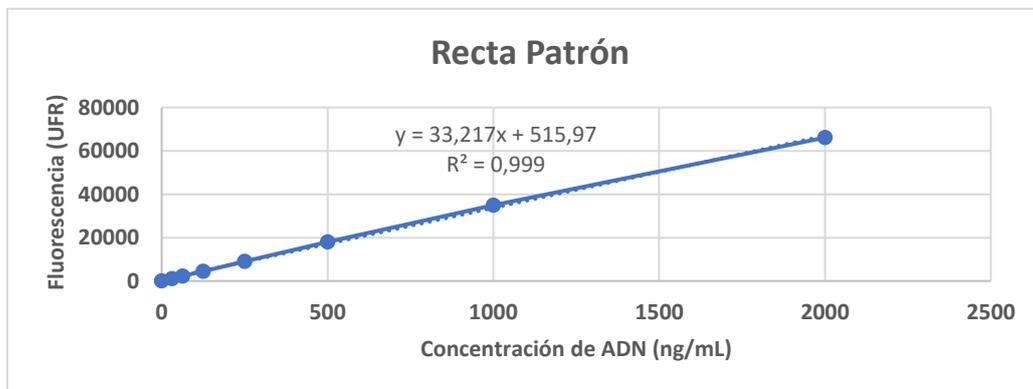


Figura 8. Ejemplo de recta patrón para la cuantificación de librerías.

La recta patrón se realizó a partir de disoluciones seriadas de una concentración conocida de ADN bicatenario correspondiente al fago Lambda. Se utilizó Tris-EDTA como tampón y blanco. Tras la adición del colorante QuantiFluor se analizó la fluorescencia a través del espectrofluorímetro GLOMAX Multi-Detection System (Promega®, USA).

La fluorescencia final del pool fue de 16362.05 UFR (Unidades de Fluorescencia Relativa), obteniendo una concentración final de 48.67 ng/μL. Finalmente se calcularon los nanomoles/L (nM) teniendo en cuenta la longitud promedio de todos los amplicones (312 pb) y su peso molecular, obteniendo un resultado de 236.35 nM.

3.4.7 SECUENCIACIÓN

El proceso de secuenciación se realizó empleando dos tipos de secuenciadores: el sistema de secuenciación MiSeq® (Illumina®, California) de INCLIVA para las pruebas de optimización, y el sistema NovaSeq® 6000 (Illumina®, California) del Centro Nacional de Análisis Genómico (CNAG, Barcelona) para el estudio poblacional, debido a la gran cantidad de muestras empleadas y el elevado número de lecturas necesarias para el posterior genotipado. Ambos sistemas emplean la secuenciación por síntesis mediante la amplificación en puente.

La primera fase del proceso se basa en la formación de grupos o *clusters*. En la celda de flujo del secuenciador hay disponibles una gran cantidad de secuencias adaptadoras, las cuales son complementarias a los adaptadores incorporados en los amplicones durante la PCR anidada, junto a los *barcodes*. Después de un primer proceso de desnaturalización, las hebras de cadena sencilla hibridan con dichas secuencias por su extremo 5'. Tras un proceso de amplificación isotérmica, el extremo 3' de cada cadena puede hibridar con adaptadores complementarios cercanos, formando una estructura en puente que permite la amplificación en el sentido inverso. Este paso se repite en varios ciclos llegando a formar hasta 200 *clusters* compuestos de moléculas idénticas. Finalmente, tras diversos lavados para eliminar las cadenas reversas y el ADN no unido, se produce el bloqueo final de los extremos 3' con objeto de evitar un cebado no deseado (Metzker, 2010; Shendure & Ji, 2008).

La secuenciación propiamente dicha de los sistemas Illumina se basa en la secuenciación por síntesis, donde se emplean desoxinucleótidos (dNTPs) marcados con un fluoróforo y con una modificación en el grupo -OH de su extremo 3', convirtiéndolos en terminadores reversibles. Por tanto, la secuenciación se inicia con la adición de un nuevo dNTP sobre la cadena molde, el cual impide la extensión de la cadena por su extremo bloqueante. A continuación, tras un lavado de los reactivos, se produce la identificación de dicho dNTP gracias a la excitación del fluoróforo que lleva incorporado. Finalmente, se añade un agente químico capaz de eliminar el grupo bloqueante, pudiendo repetir el ciclo hasta completar la secuenciación de cada molécula de ADN. La señal de fluorescencia será posteriormente analizada por un *software* informático (Metzker, 2010; Shendure & Ji, 2008).

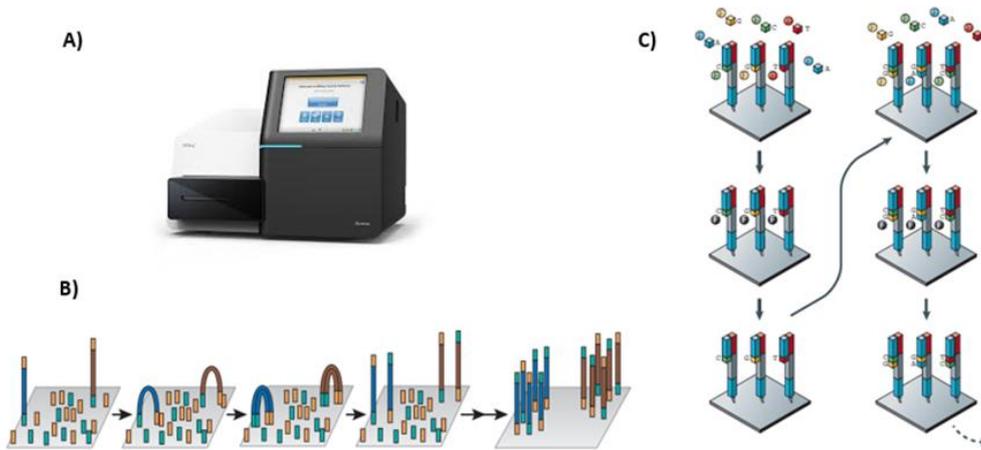


Figura 9. Resumen del proceso de secuenciación de los sistemas Illumina. A) Secuenciador MiSeq utilizado para las pruebas de optimización. B) Proceso de generación de clusters. C) Secuenciación de librerías por síntesis. Adaptada de (Metzker, 2010; Shendure & Ji, 2008).

Para el caso particular de las pruebas de optimización se utilizó el kit Miseq Reagent v2 (Illumina®, California) disponible para el sistema MiSeq. Siguiendo las instrucciones del fabricante, se cargaron 600 µL de librería con un control interno al 5%. Como control interno se empleó el fago Phix proporcionado por la casa comercial Illumina. Tras un primer paso de desnaturalización con hidróxido sódico (NaOH) de las librerías y fago, éstos fueron diluidos con tampón de hibridación hasta una concentración final de 17 pM y 12.5 pM, respectivamente. A través de este sistema pueden utilizarse hasta 22 millones de lecturas, obteniendo lecturas de 2x150 pb en ambos sentidos (*paired-end*) en un tiempo estimado de 24 horas.

3.4.8 ANÁLISIS BIOINFORMÁTICO

Finalizado el proceso de secuenciación, se procedió al estudio de los posibles polimorfismos presentes en los distintos amplicones. Para ello, el personal bioinformático de UGD llevó a cabo un análisis empleando diferentes *softwares* y *scripts*. A continuación, se describen los pasos generales realizados durante el análisis bioinformático (*Figura 10*):



Figura 10. Esquema de la pipeline bioinformática.

El primer paso fue el demultiplexado. Durante este proceso se utilizó la información del fichero *sample sheet* (informa de la combinación de *barcodes* que presenta cada individuo) para identificar que secuencias corresponden a cada muestra. De este modo, los archivos generados por los secuenciadores de Illumina en formato BCL son procesados a archivos en formato FASTQ, los cuales almacenan las secuencias biológicas de cada individuo, así como la calidad de cada base en escala Phred (representado por el valor Q, estima la probabilidad de que una base sea incorrecta). A partir de los FASTQ también se realizó un primer control de calidad empleando el programa FastQC (v0.11.5, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), el cual indica sí la secuenciación pasa los filtros de calidad deseados.

A continuación, se llevó a cabo el procesamiento de lecturas crudas mediante el proceso conocido como *trimming*. Éste consistió en la eliminación de los adaptadores y bases de baja calidad, especialmente de los extremos de las secuencias. Tras este paso, se volvió a realizar un control de calidad con el mismo programa para ver si se produjo una mejora con respecto al anterior, en aquellos casos mejorables.

Una vez obtenidas las lecturas curadas se procedió a su mapeo. Las lecturas fueron alineadas contra un genoma de referencia (versión GRCh37/hg19), y los datos fueron almacenados en archivos en formato BAM (del inglés *Binary Aligment Map*), que corresponde a la versión binaria del formato SAM (del inglés *Sequence Aligment Map*).

Posteriormente se realizó el llamado de variantes. En esta etapa, las lecturas alineadas fueron utilizadas para identificar las posiciones que diferían con el genoma de referencia, obteniendo de esta forma el genotipo de cada muestra. La información quedó recogida en los archivos con formato VCF (del inglés *Variant Call Format*).

Finalmente, los archivos VCF se anotaron empleando información contenida en la base de datos Ensembl (v97, julio 2019). Durante este paso, se predice el posible efecto funcional de cada uno de los polimorfismos encontrados y pueden ser filtrados posteriormente por consecuencia, efecto, impacto, entre otros.

3.5 ANÁLISIS ESTADÍSTICO

Los datos descriptivos de la población se expresaron como media \pm desviación estándar para las variables cuantitativas y como porcentaje para las variables cualitativas. Las medidas sobre las distintas variables clínico-bioquímicas de cada sujeto fueron proporcionadas por el CIBERDEM. Los niveles de triglicéridos se midieron en mg/dL y se realizó su transformación logarítmica.

El análisis estadístico fue realizado por el personal bioinformático de UGD y se empleó RStudio (versión 1.1.456). Una vez seleccionadas las variantes que fueron genotipadas en al menos el 70% de los individuos, se realizaron las pruebas de Shapiro-Wilk y de Levene para comprobar la normalidad y homocedasticidad de los datos, respectivamente. Esto llevó a emplear la prueba no paramétrica de Kruskal-Wallis para estudiar la posible asociación entre los genotipos de las variantes encontradas y las variables clínico-bioquímicas (cuantitativas). Para examinar la posible asociación entre los SNPs y las variables dicotómicas (DM2 y obesidad) se empleó la prueba de Chi-cuadrado o la prueba exacta de Fisher, en aquellos casos que al menos uno de los grupos presentaba menos de 5 observaciones. Se analizó si las frecuencias genotípicas cumplían el equilibrio de Hardy-Weinberg, permitiendo descartar aquellas variantes que no estuviesen en equilibrio. En todas las pruebas se asumió un nivel de significatividad ≤ 0.05 .

4 RESULTADOS Y DISCUSIÓN

4.1 CARACTERÍSTICAS DE LA POBLACIÓN DI@BET.ES

En el presente trabajo se incluyeron 4784 muestras pertenecientes al estudio Di@bet.es de edades comprendidas entre 18 y 93 años. Los niveles medios de glucemia basal fueron de 99.07 mg/dL, teniendo en cuenta que los pacientes diabéticos estaban tratados. El principal criterio de inclusión fue la disponibilidad de las muestras en el momento del estudio. En la *Tabla 7* se muestran las características clínico-bioquímicas de la población bajo estudio:

Tabla 7. Características clínico-bioquímicas de la población Di@bet.es.

ESTUDIO DI@BET.ES	
Edad (años)*	50.43 ± 16.94
Hombres (%)	42.75
Mujeres (%)	57.25
Glucosa basal (mg/dL)*	99.07 ± 29.73
Insulina basal (μU/mL)*	69.97 ± 128.22
HOMA-IR*	2.57 ± 6.86
Diabetes (%)	14.03
IMC (Kg/m ²)*	28.05 ± 5.17
Obesidad (IMC > 30) (%)	30.44
LDL (mg/dL)*	104.53 ± 29.64
HDL (mg/dL)*	51.56 ± 12.84
LogTG*	2.02 ± 0.22
Colesterol total (mg/dL)*	195.68 ± 39.75

HOMA-IR: Modelo homeostático de resistencia a la insulina; IMC: Índice Masa Corporal; LDL: Lipoproteínas de baja densidad; HDL: Lipoproteínas de alta densidad; TG: Triglicéridos. *Datos expresados como media ± desviación estándar.

4.2 DISEÑO IN SILICO DE CEBADORES

Con el objetivo de estudiar las regiones exónicas de los 6 genes seleccionados se diseñaron 312 cebadores utilizando los programas ExonPrimer y Primer3. A continuación, dichos cebadores se separaron en dos multiplexes atendiendo a las parejas solapantes o contiguas y/o a las posibles interacciones entre ellos, resultando en 80 parejas para la multiplex A y 76 parejas para la multiplex B. Las dos multiplexes en conjunto representaron 156 amplicones con una longitud promedio de 312 bases.

Inicialmente, las dos multiplexes de cebadores se realizaron de forma equimolar, es decir, se mezclaron 3 μL de cada pareja de cebadores a una concentración de 50 μM , obteniéndose concentraciones finales de 0.63 μM y 0.66 μM para las multiplexes A y B, respectivamente.

4.3 OPTIMIZACIÓN DE LAS CONDICIONES DE PCR 1

Obtenidas todas las parejas de cebadores y separadas en dos multiplexes, se realizaron diferentes pruebas de optimización utilizando muestras de ADN control con el fin de encontrar las mejores condiciones de PCR1 que permitiesen analizar el mayor número posible de amplicones.

Como punto de partida se tomaron las condiciones de PCR establecidas en un estudio previo de características similares. En este caso, se mantuvo la concentración de la enzima comercial y se probó el efecto de dos aditivos con el objetivo de ver si mejoraban la reacción. Además, dichos aditivos se combinaron con concentraciones diferentes de las multiplexes de cebadores, resultando en 8 condiciones a testar que se muestran en la *Tabla 8*:

Tabla 8. Condiciones a testar en el proceso de optimización de PCR 1.

CEBADORES 1X				CEBADORES 0.1X			
SIN AD.	AD. 1 [3 Mm]	AD. 2 [5% v/v]	AD. 1 [3 mM] + AD. 2 [5% v/v]	SIN AD.	AD. 1 [3 Mm]	AD. 2 [5% v/v]	AD. 1 [3 mM] + AD. 2 [5% v/v]

AD: Aditivo

En la *Figura 11* se observan los resultados de la electroforesis capilar tras haber realizado las dos PCRs pertinentes. Se seleccionó la condición “Cebadores 0.1X + Aditivo 1 [3 mM]” tanto por la presencia de bandas del tamaño esperado como por la menor suciedad de la reacción, a diferencia de las condiciones positivas de los cebadores a 1X, en las que se observó una mayor cantidad de cebadores dimerizados. Además, se descartó la condición “Cebadores 0.1X + Aditivo 1 [3 mM] + Aditivo 2 [5% v/v]” debido a que los resultados eran muy similares a la seleccionada, por lo que no había necesidad de añadir un segundo reactivo.

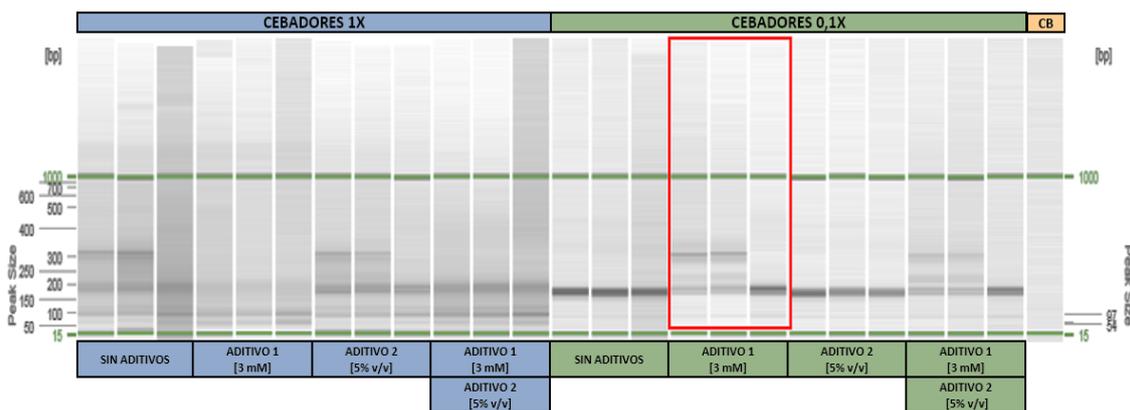


Figura 11. Resultados de electroforesis capilar para la evaluación de las 8 condiciones a testar en la PCR 1. Se observan dos muestras y un control negativo por condición, así como el control de barcode (CB). El recuadro rojo indica la condición seleccionada como la más favorable.

A continuación, debido a la mejora en la reacción de PCR producida por el aditivo 1, se realizó una segunda prueba variando la concentración del mismo: se mantuvo la concentración inicial como control [3 mM] y se testó a la mitad [1.5 mM] y al doble [6 mM] de concentración.

En este segundo caso se seleccionó la condición “Aditivo 1 [6mM]” como la más favorable, de igual modo que la anterior, debido a la mayor intensidad de las bandas como a la menor presencia de dímeros de cebadores, indicando una amplificación más específica (Figura 12).

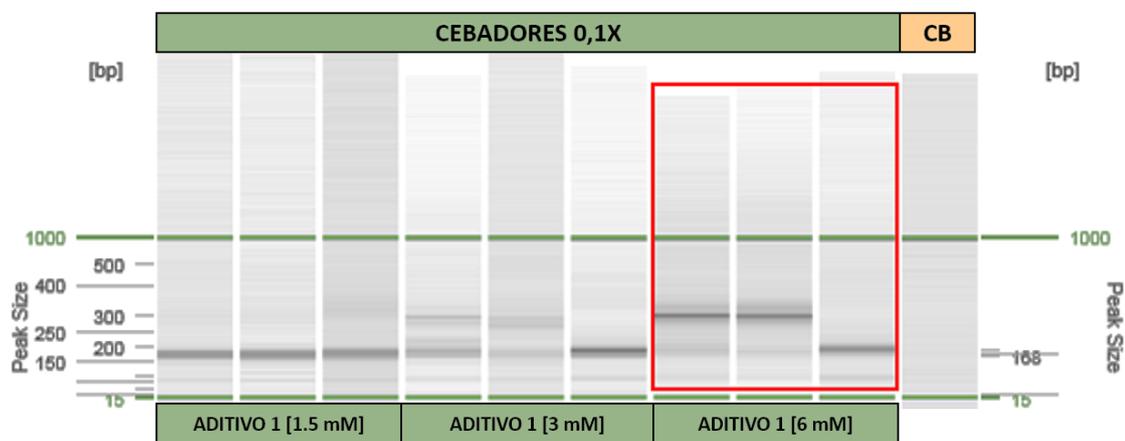


Figura 12. Resultados de electroforesis capilar sobre las tres medidas de concentración del aditivo 1. Se observan dos muestras y un control negativo por condición, así como el control de barcode (CB). El recuadro rojo indica la condición seleccionada como la más favorable

Atendiendo a estos resultados, se realizó un reajuste de la concentración de cada pareja de cebadores en función de la cobertura mínima obtenida en la prueba anterior, añadiendo la mitad, el doble, el triple de volumen o manteniéndose igual. Además, las dos multiplexes también se probaron a concentraciones 0.1X y 0.05X.

Tal y como se observa en la *Figura 13*, se decidió aceptar como condición final “cebadores 0.05X + aditivo [6 mM]”. La dilución de los cebadores aumenta su eficiencia, reduciendo de esta forma la saturación de éstos y la menor suciedad de la reacción de PCR.

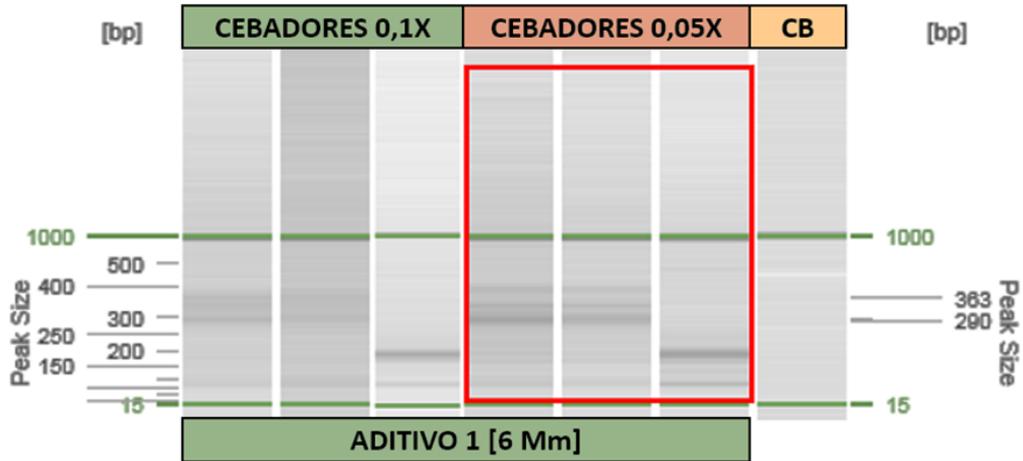


Figura 13. Resultados de electroforesis capilar tras el reajuste de la concentración de cada pareja de cebadores y pruebas a 0.1X y 0.05X. Se observan dos muestras y un control negativo por condición, así como el control de barcode (CB). El recuadro rojo indica la condición seleccionada como la más favorable.

En la *Figura 14* se observa la diferencia entre las coberturas mínimas antes y después de realizar el ajuste de concentraciones tomando como ejemplo el gen GCKR. No se encontró una homogeneidad en todos los amplicones, no obstante, se observó como la cobertura mínima aumentó en la mayor parte de ellos. El resto de los genes mostraron resultados similares tras el reajuste de la concentración de las parejas de cebadores.

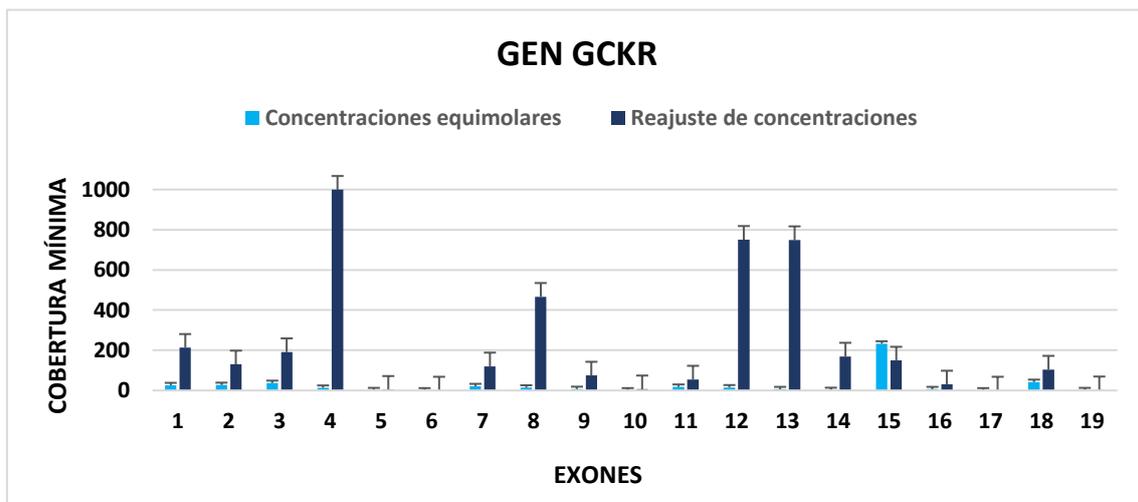


Figura 14. Cobertura mínima antes y después del reajuste de concentración de cebadores para el gen GCKR.

4.4 VALIDACIÓN DE LAS LIBRERÍAS DE LA POBLACIÓN DI@BET.ES

Una vez obtenidas las mejores condiciones de PCR se procedió a generar las librerías de cada muestra de la población Di@bet.es. Antes de enviar las muestras a secuenciar, se validó su amplificación mediante electroforesis capilar (*Figura 15 A*).

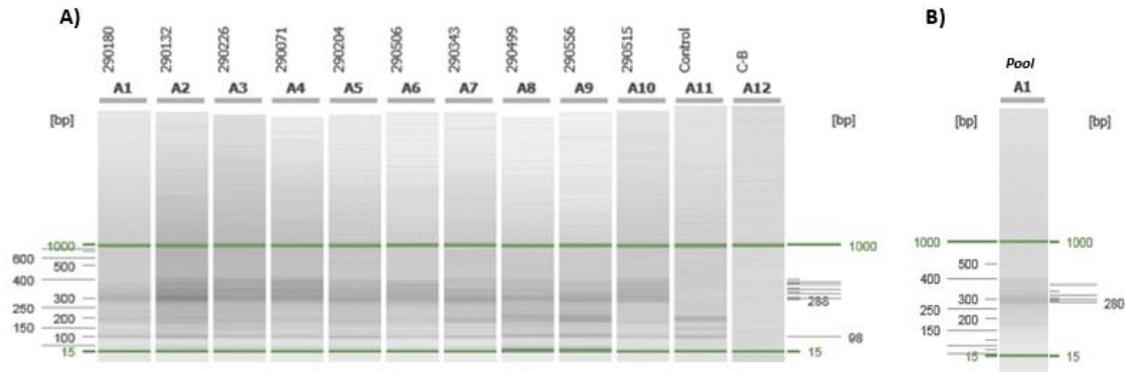


Figura 15. Validación del proceso de amplificación de las muestras bajo estudio. A) Se observa la amplificación de 10 muestras al azar representativas de la población Di@bet.es. En algunas de las muestras (A1, A7, A8 y A9) la amplificación es de menor calidad, probablemente debido a la peor calidad del ADN o menor concentración. B) Electroforesis capilar del pool final tras la purificación.

A continuación, se generó la librería de todas las muestras mezclando 2 μ L de cada una y se procedió a su purificación con el objetivo de eliminar restos de reactivos y dímeros de cebadores (*Figura 15 B*).

4.5 RESULTADOS DEL ANÁLISIS BIOINFORMÁTICO

4.5.1 CALIDAD DE LA SECUENCIACIÓN

El número medio de lecturas crudas/individuo fue de 195433.52 ± 117886.83 . A continuación, se llevó a cabo un procesado de dichas lecturas en el que se eliminaron los adaptadores y se aplicaron varios filtros de calidad con el objetivo de eliminar aquellas cuyo valor en escala Phred fuese de $Q < 20$ y/o presentasen una longitud ≤ 80 bases. Como resultado final se obtuvo un número medio de lecturas curadas/individuo de 131837.78 ± 78102.76 . Una vez completado el procesado se procedió a realizar un control de calidad de la secuenciación utilizando diversos parámetros que proporciona el programa FastQC, indicando de este modo el grado de fiabilidad de los datos.

Primeramente, se analizó la distribución de los valores de calidad por base y el valor medio de calidad por lectura. Respecto al primero, el valor medio de calidad por base fue bastante homogéneo antes y después del procesado, además presentó valores de $Q > 30$ a lo largo de todas las lecturas (Figura 16). De la misma forma, se observó que la gran mayoría de lecturas/individuo presentaban una calidad media de $Q > 30$ (Figura 17). Ambos parámetros indicaron una buena calidad de los datos obtenidos de la secuenciación.

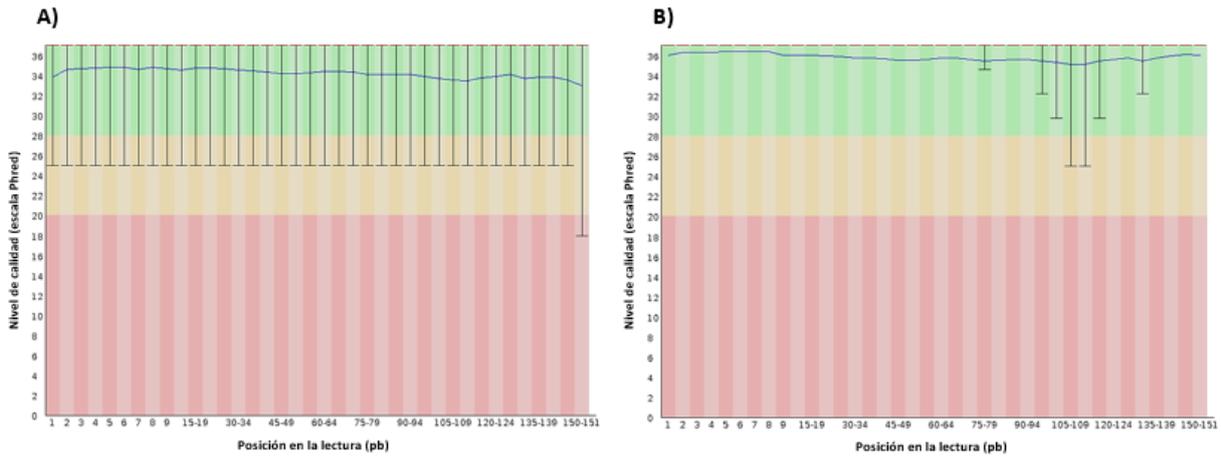


Figura 16. Calidad de la secuenciación por base e individuo. A) Calidad antes del procesado de lecturas. B) Calidad tras aplicar el recorte de adaptadores y filtros estándar de calidad. La línea roja representa la mediana y la línea azul la media.

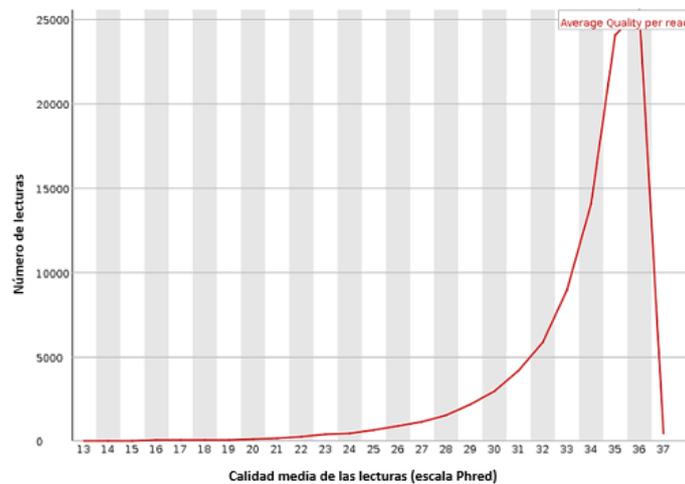


Figura 17. Distribución de la calidad de todas las lecturas por individuo. La mayor parte de lecturas presentan una calidad de $Q > 30$.

También se analizaron parámetros como el contenido de N por base y la distribución de la longitud de secuencias. En los dos casos se observó que la calidad de la secuenciación fue buena, tanto por la ausencia de indeterminaciones como por que la gran mayoría de lecturas se encontraban dentro del tamaño esperado (Figura 18).

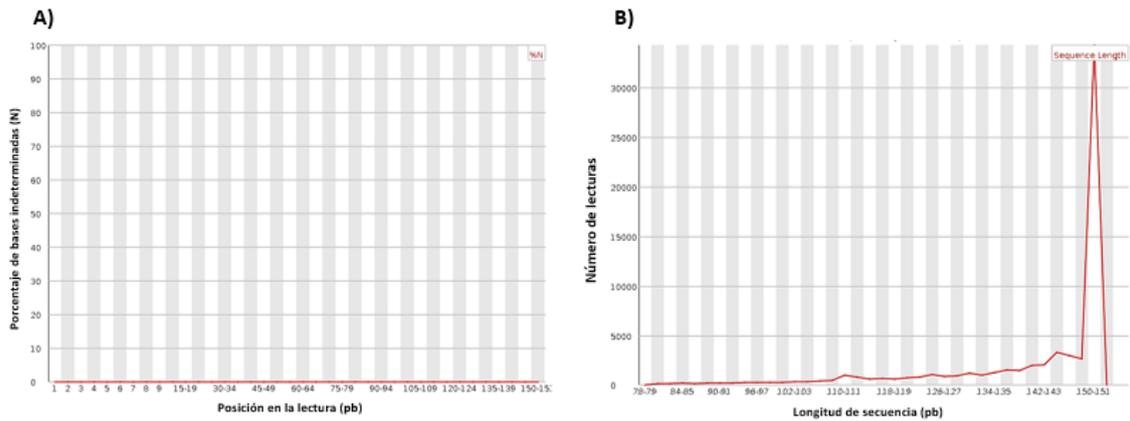


Figura 18. Contenido de bases indeterminadas (N) y distribución del tamaño de las lecturas. A) Se observa que el porcentaje de bases sin lectura es cero. B) La mayoría de las lecturas procesadas presenta un tamaño esperado de alrededor de 150 pb.

4.5.2 COBERTURA DE LA SECUENCIACIÓN

Pasado el control de calidad, se procedió al análisis de los datos de cobertura. Para ello, se estableció como umbral una cobertura mínima de 20X, la cual permite la identificación de variantes (incluyendo de baja y/o rara frecuencia) de una forma adecuada.

Por un lado, se examinó la cobertura media de cada gen estudiado (Tabla 9). Se observó que el gen *MANBA* obtuvo la mayor cobertura (840.17X), mientras que en *ACAD10* se encontró la menor (42.01X). También se examinó la distribución de cobertura para cada gen. Como ejemplos, se muestra la distribución de la cobertura media para los genes *MANBA* y *SREBF2* respecto a todas las muestras analizadas. Se observó que todas las regiones fueron cubiertas, aunque no de una forma homogénea (Figuras 19 y 20). El resto de los genes mostraron resultados similares.

Tabla 9. Cobertura media de cada gen estudiado.

GEN	COBERTURA MEDIA
<i>ACAD10</i>	42.01
<i>CYP1A1</i>	105.66
<i>GCKR</i>	59.95
<i>MANBA</i>	840.17
<i>SREBF2</i>	310.42
<i>TCF7L2</i>	588.74

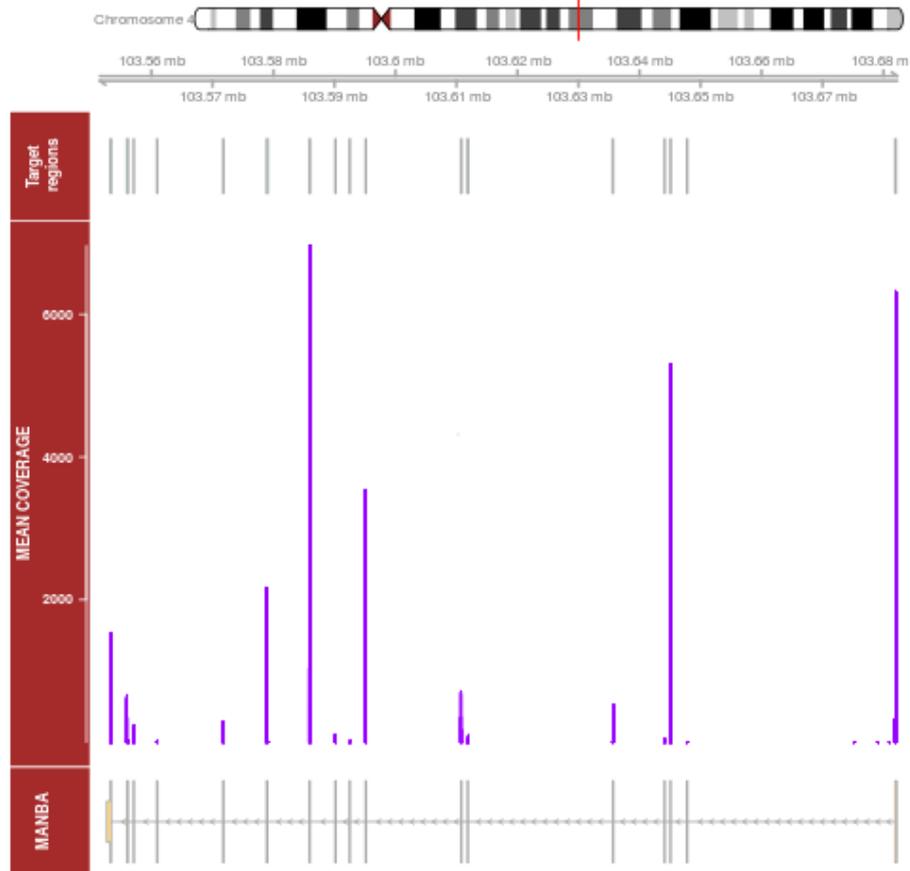


Figura 19. Distribución de la cobertura media del gen MANBA.

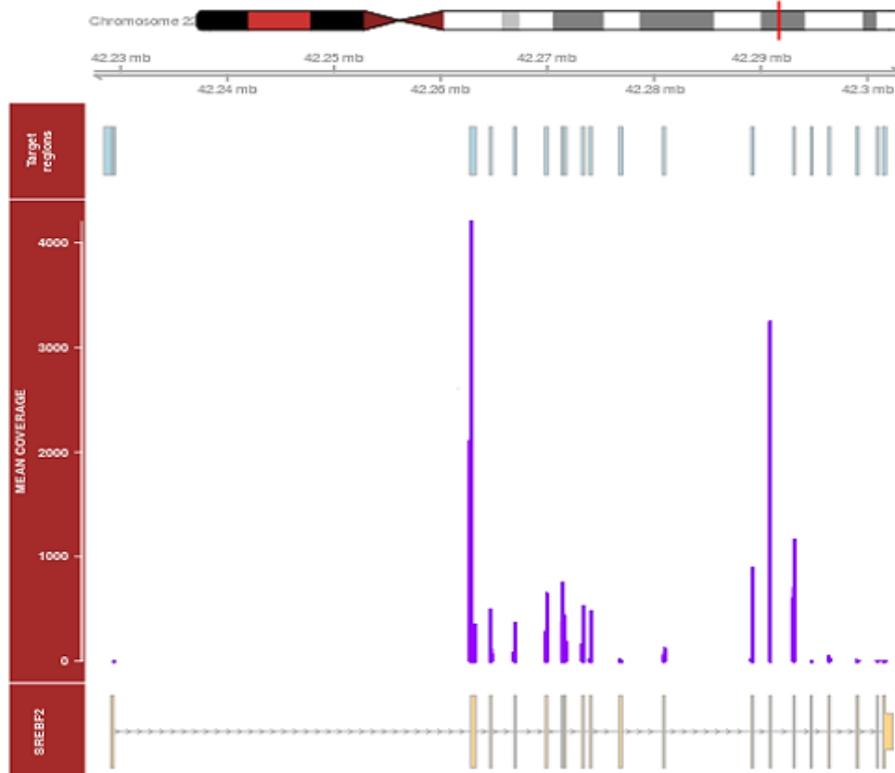


Figura 20. Distribución de la cobertura media del gen SREBF2.

La *Figura 21* muestra la relación entre la cobertura media de todas las muestras analizadas por amplicón (líneas grises) y la fracción de amplicón cubierta. La línea magenta representa la media de todos los amplicones, por lo que aproximadamente el 55% de las regiones de interés presentan una cobertura de 20X (umbral determinado anteriormente).

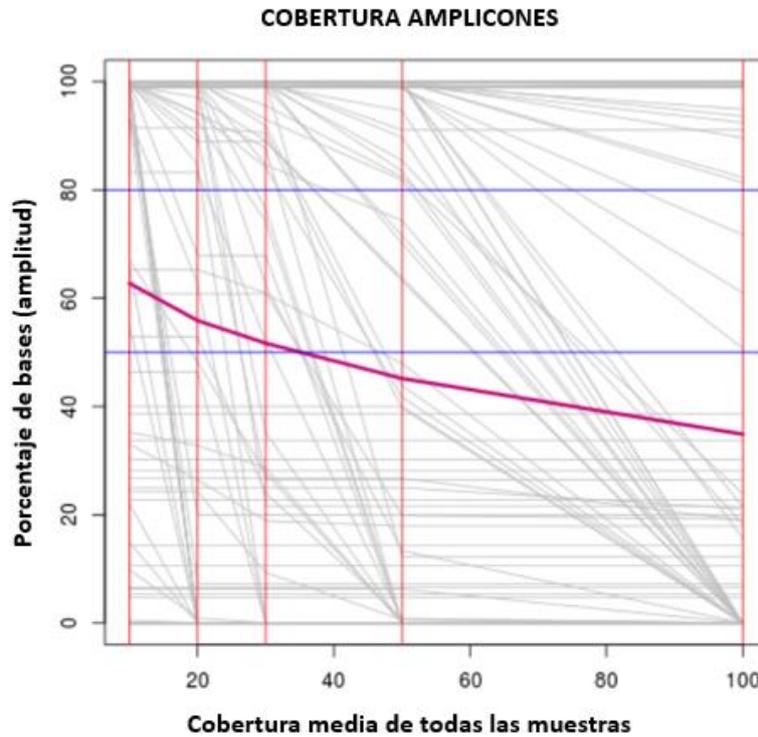


Figura 21. Relación entre la cobertura media y la amplitud de cobertura por amplicón de todas las muestras analizadas. Las líneas grises representan cada amplicón mientras que la de color magenta representa la relación cobertura media/porcentaje de la región cubierta de todos los amplicones.

La falta de homogeneidad en las coberturas entre los distintos amplicones y genes puede deberse a diferencias entre la eficacia de cada amplicón, la cantidad de lecturas disponibles y a la calidad del ADN de cada muestra.

4.6 ANÁLISIS DE VARIANTES

Para la selección final de variantes se aplicaron diferentes filtros. Por un lado, durante el análisis bioinformático se priorizó en aquellas variantes que cumplieren una cobertura mínima de 10X y una calidad de genotipado de 20, obteniéndose en este caso un total de 1123 variantes. Posteriormente, se seleccionaron solo aquellas variantes que se encontraron genotipadas en al menos el 70% de los individuos, asegurando así que no se trataba de un artefacto del proceso de secuenciación. El número final de variantes seleccionadas fue de 654 (*Tabla 10*).

Tabla 10. Variantes seleccionadas tras aplicar los filtros de calidad. Las variantes aparecen clasificadas según el gen al que pertenecen, tipo de mutación y efecto.

TIPO	ACAD10	CYP1A1	GCKR	MANBA	SREBF2	TCF7L2	TOTAL
Splicing	10	2	4	2	9	1	28
Sinónimas	19	11	17	24	47	34	152
3' UTR	0	0	0	0	0	6	6
5' UTR	0	3	3	1	0	2	9
Intrónicas	15	6	34	20	31	26	132
Upstream	0	0	0	0	0	0	0
Inframe	0	0	0	0	1	0	1
Cambio de sentido	53	54	42	53	56	44	302
Frameshift	0	2	0	1	1	0	4
Aceptora de splicing	0	0	1	2	1	0	4
Donadora de splicing	1	0	0	1	0	0	2
Ganancia codón de parada	0	1	8	2	1	2	14
Pérdida codón de parada	0	0	0	0	0	0	0
	98	79	109	106	147	115	654

Naranja: efecto bajo; Verde: efecto modificador; Azul: efecto moderado; Rojo: efecto alto.

Entre todas las variantes, la principal consecuencia encontrada fue la mutación con cambio de sentido o *missense mutation* (48.18%). Se trata de una mutación puntual en la que se produce un cambio en un solo nucleótido, provocando la aparición de un nuevo codón y a su vez la introducción de un nuevo aminoácido en la proteína. Las mutaciones con cambio de sentido pueden tener diversos efectos, pasando de no tener ninguna implicación en la proteína hasta alterar su estructura o función (alteraciones en el centro activo, plegamiento, estabilidad, localización final, interacción proteína-proteína, proteína-ADN, etc.). También se encontró un alto número de variantes sinónimas o de bajo efecto (23.24%) e intrónicas (20.18%). Se encontraron 24 variantes de efecto alto, siendo las principales las producidas por ganancia de un codón de parada.

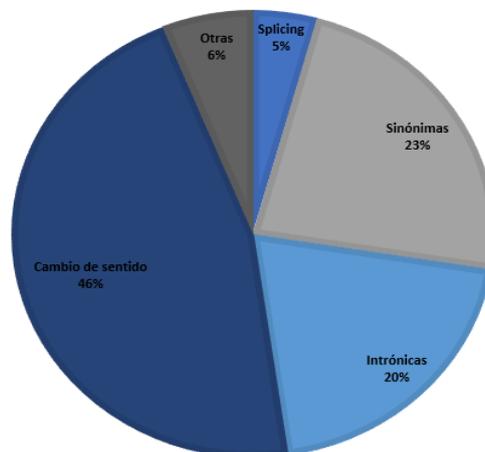


Figura 22. Porcentaje de variantes encontradas según el tipo de mutación.

4.6.1 VALIDACIÓN DE VARIANTES CON ESTUDIOS PREVIOS BASADOS EN EXOMAS

Con respecto a la validación de variantes, se encontraron dos posiciones coincidentes con los estudios previos de caso-control basados en exomas realizados por el grupo (*Tabla 11*):

Tabla 11. Variantes validadas con los estudios previos de caso-control en exomas.

cr:posición	rs	Alelos	Gen	Cambio	Tipo	Sift	Polyphen
15:75012979	rs41279188	G/T	CYP1A1	p.R464S	CS	Perjudicial	Perjudicial
10:114925406	rs77673441	C/G	TCF7L2	p.P489A	CS	Tolerado	Benigna

CS: Cambio de Sentido

La variante rs41279188 fue anotada como un SNP de cambio de sentido, y a través de los predictores de impacto *Sift* y *Polyphen* se observó que podía tener un efecto perjudicial. De hecho, en los estudios de caso-control únicamente se encontró genotipada en el grupo de diabéticos (Galán Chilet, 2016). Del mismo modo, la variante rs77673441 también fue anotada como cambio de sentido, sin embargo, fue clasificada como una variante protectora. En los estudios de exoma solamente se encontró en el grupo de controles, lo que indicaría su posible papel protector (Martínez, 2015). Ambas variantes fueron genotipadas en más del 70% de los individuos.

4.7 ANÁLISIS POR GEN

Para el presente trabajo se priorizaron las variantes de mayor efecto, siendo en este caso de cambio de sentido, que resultaron significativas con alguna variable propuesta. Se encontraron 24 variantes de alto impacto genotipadas en al menos el 70% de los individuos, no obstante, ninguna fue significativa con alguna de las variables.

4.7.1 ACAD10

El gen *ACAD10* se localiza en el cromosoma 12 y codifica para la enzima acil-CoA deshidrogenasa 10. Se trata de una flavoproteína mitocondrial que pertenece a la familia de las acil-CoA deshidrogenasas y que participa en el proceso de β -oxidación de los ácidos grasos. Más concretamente, actúa catalizando la oxidación de un ácido graso activado (unido a una molécula de acil-CoA) gracias al nucleótido FAD que tiene unido covalentemente (Ye et al., 2004).

Tras el análisis estadístico se encontraron cuatro SNPs de cambio de sentido significativos para los niveles de glucosa y TG (rs149500655), los niveles de insulina y el índice HOMA-IR (12-112174758-T-G) y los niveles de CT (rs545041174 y 12-112187101-G-A), tal y como se muestra en la *Tabla 12*. No se encontraron variantes asociadas a la DM2 u obesidad.

Tabla 12. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen ACAD10.

12-112187140-G-A (rs149500655)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
Glucosa (mg/dL)	99.15 ± 30.33	82.84 ± 16.63	-	0.026*	4083	11	-
LogTG	2.02 ± 0.22	1.88 ± 0.18	-	0.036*			
12-112174758-T-G					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
Insulina (µU/mL)	68.70 ± 79.23	50.13 ± 23.97	-	0.008*	3881	46	-
HOMA-IR	2.51 ± 4.89	1.62 ± 0.81	-	0.015*			
12-112184869-C-T (rs545041174)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
CT (mg/dL)	195.86 ± 39.63	132.81 ± 16.68	-	0.032*	3987	2	-
12-112187101-G-A					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
CT (mg/dL)	195.56 ± 39.80	261.18 ± 1.37	-	0.027*	4067	2	-

TG: Triglicéridos; HOMA-IR: Modelo Homeostático de Resistencia a la Insulina; CT: Colesterol total. Datos expresados como media ± desviación estándar. *Nivel de significatividad ≤ 0.05.

No se encontraron referencias bibliográficas sobre las cuatro variantes. No obstante, a pesar de que el gen *ACAD10* ha sido poco estudiado como gen candidato para la DM2, varias publicaciones relacionan alteraciones en este gen con el riesgo de desarrollar la enfermedad.

Un estudio publicado en el año 2010 demostró, en individuos de población Pima, la asociación existente entre dos variantes intrónicas (rs601663 y rs659964) y el aumento de susceptibilidad a la DM2. Además, estas dos variantes, junto a una tercera (rs7136874), también se asociaron con un descenso en las tasas de oxidación lipídica y lipólisis, un engrosamiento de adipocitos, así como una menor tasa en la eliminación de glucosa mediada por insulina (Bian et al., 2010). Resultados similares se obtuvieron más tarde a partir de un modelo de ratón *knockout* para el gen *ACAD10* (Bloom et al., 2018). Dichos ratones, a diferencia de los controles, exhibieron un notorio aumento de peso (especialmente por acumulación de tejido adiposo visceral y abdominal), un incremento de los niveles de TNF-α, así como características propias de la RI como mayores niveles de insulina plasmática y un menor grado de fosforilación de Akt.

Estos resultados ponen de manifiesto la importancia de la proteína acil-CoA deshidrogenasa 10 sobre la regulación del metabolismo energético. Un descenso de la oxidación lipídica en las mitocondrias junto con una menor tasa de lipólisis podría conllevar la acumulación de grasas en los adipocitos y, por tanto, una ganancia de peso y obesidad, factores clave en el desarrollo de la RI y aparición de DM2. Por esta razón, sería interesante llevar a cabo estudios funcionales para ver como el cambio aminoacídico producido en estas cuatro variantes podría afectar a la

función del gen *ACAD10* y la consiguiente asociación con alteraciones metabólicas y, posiblemente, con DM2.

4.7.2 CYP1A1

El citocromo P450 (CYP) representa una superfamilia de hemoproteínas con actividad monooxigenasa. Biológicamente, participan en el metabolismo de ciertos xenobióticos y componentes endógenos como hormonas, ácidos biliares y colesterol. El miembro *CYP1A1* codifica para enzima citocromo P450 1A1, la cual participa en la biotransformación de hidrocarburos aromáticos policíclicos para su posterior eliminación vía orina o bilis. Se trata de una enzima localizada en el mitocondria y RE, cuya expresión es inducible gracias al receptor de hidrocarburos arilo (RHA) y la presencia de sus ligandos, principalmente hidrocarburos aromáticos (Santes-Palacios et al., 2016). Estudios más recientes indican que *CYP1A1* también participa en el metabolismo de estrógenos y ácidos grasos poliinsaturados como el ácido araquidónico y eicosanoides (Mescher & Haarmann-Stemann, 2018).

Se encontraron cuatro variantes de cambio de sentido significativas para el perfil glucídico (niveles de glucosa, insulina y HOMA-IR) y para los niveles de HDL y TG, respectivamente (*Tabla 13*). Mientras que no se encontró ninguna variante asociada a la obesidad, el SNP rs56343424 también se encontró asociado a la DM2 tras aplicar la prueba de Fisher (*Tabla 14*).

Tabla 13. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen CYP1A1.

15-75012837-C-A (rs56343424)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
Glucosa (mg/dL)	99.18 ± 30.50	105.69 ± 22.02	-	0.036*	3994	10	-
Insulina (µU/mL)	69.78 ± 134.97	111.40 ± 105.10	-	0.037*			
HOMA-IR	2.57 ± 7.17	4.31 ± 4.98	-	0.032*			
HDL (mg/dL)	51.57 ± 12.87	41.00 ± 12.80	-	0.012*			
15-75013621-C-T					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
Glucosa (mg/dL)	98.67 ± 29.04	76.86 ± 1.15	-	0.045*	4121	2	-
HOMA-IR	2.47 ± 4.67	0.76 ± 0.23	-	0.040*			
15-75012987-G-T (rs1799814)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
HOMA-IR	2.57 ± 7.26	2.49 ± 3.33	3.60 ± 1.74	0.030*	4000	507	9
15-75012925-C-T (rs28399429)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
LogTG	2.02 ± 0.21	2.23 ± 0.11	-	0.050*	4273	3	-

HOMA-IR: Modelo Homeostático de Resistencia a la Insulina; HDL: Lipoproteínas de alta densidad; TG: Triglicéridos. Datos expresados como media ± desviación estándar. *Nivel de significatividad ≤ 0.05.

Tabla 14. Variante asociada a la DM2 mediante la prueba exacta de Fisher.

cr:posición	Alelos	rs	Tipo	MAF (EUR)	OR	p-valor	N.º Individuos		
							0/0	0/1	1/1
15-75012837	C/A	rs56343424	CS	-	4.20	0.038*	3994	10	-

CS: Cambio de Sentido; MAF: Frecuencia Alélica Mínima; OR: *ODDS Ratio*; *Nivel de significatividad ≤ 0.05 .

Respecto a la variante rs56343424 no se encontraron datos bibliográficos. Sin embargo, en los individuos heterocigotos (C/A) se observó un incremento en los niveles de glucosa, insulina y HOMA-IR, así como un descenso en los niveles de colesterol HDL, respecto a los homocigotos para el alelo de referencia (C/C). Además, la prueba exacta de Fisher para esta variante mostró un *odds ratio* superior a 4. El SNP rs56343424 podría suponer una fuerte asociación para el desarrollo de RI y DM2, por lo que es conveniente realizar estudios futuros que lo confirmen. Las variantes 15-75013621-C-T y rs28399429 tampoco resultaron estar descritas en la bibliografía.

Por otro lado, se encontró la variante común rs1799814. Se trata de un polimorfismo de cambio de sentido (p.T461N) que se ha relacionado con el riesgo de desarrollar cáncer de pulmón, dato no analizado en nuestro caso (Islam et al., 2013).

Los polimorfismos presentes en el gen *CYP1A1* han sido poco explorados en el ámbito de la DM2. Sin embargo, varios estudios indican una posible asociación entre el gen *CYP1A1* y el desarrollo de DM2 en pacientes con enfermedad de arteria coronaria (EAC) (X. L. Wang et al., 2002; Zou et al., 2015). Las hipótesis apuntan a que ciertos polimorfismos en el citocromo P450 1A1 y la consiguiente desregulación metabólica de sustancias presentes en el humo del tabaco o del ácido araquidónico podrían estar asociados con el desarrollo de RI y DM2 en pacientes con esta patología cardíaca. No obstante, los mecanismos moleculares implicados todavía están por dilucidar.

En base a todo lo anterior, el gen *CYP1A1* podría ser un buen candidato para futuras investigaciones con el objetivo de evaluar qué papel juega en la genética y fisiopatología de la DM2.

4.7.3 GCKR

El gen *GCKR* codifica para la proteína reguladora de la glucoquinasa (GKRP), una proteína clave en la homeostasis de la glucosa. Se localiza principalmente en los hepatocitos, donde se encuentra formando un complejo inhibitorio con la enzima glucoquinasa (GCK), la cual actúa en

el primer paso de la glucólisis. Esto es, en periodos de ayuno la presencia de fructosa-6-fosfato (un producto de la gluconeogénesis) promueve la formación del complejo inhibitorio GKR-P-GCK, el cual es retenido en el núcleo celular impidiendo la acción de la GCK y favoreciendo la síntesis *de novo* de glucosa. Por el contrario, tras la ingesta, la presencia de fructosa-1-fosfato promueve la disrupción del complejo y la posterior traslocación de la GCK al citoplasma, lugar donde actúa regulando la entrada de glucosa al interior celular (Raimondo, Rees, & Gloyn, 2015).

Tras aplicar los distintos filtros, se observó que tres variantes de cambio de sentido mostraron diferencias significativas con alguna de las variables clínicas estudiadas (Tabla 15). En concreto, las tres variantes resultaron determinantes para el perfil lipídico, ya fuese con el colesterol LDL o con los niveles de TG, respectivamente. No se encontraron asociaciones significativas entre las distintas variantes y la DM2 u obesidad.

Tabla 15. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen GCKR.

2-27731045-A-G					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
LDL (mg/dL)	104.27 ± 29.56	92.05 ± 25.02	-	0.042*	3575	24	-
LogTG	2.02 ± 0.21	1.90 ± 0.20	-	0.004*			
2-27745412-G-A (rs200860838)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
LDL (mg/dL)	104.24 ± 29.57	159.85 ± 42.04	-	0.048*	3492	2	-
2-27730940-T-C (rs1260326)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
LogTG	2.04 ± 0.23	2.02 ± 0.22	2.00 ± 0.21	0.001*	745	2018	1240

LDL: Lipoproteínas de baja densidad; TG: Triglicéridos. Datos expresados como media ± desviación estándar.

*Nivel de significatividad ≤ 0,05.

Mientras que los dos primeros SNPs no han sido descritos en la bibliografía, la variante rs1260326 es una de las más frecuentes de este gen. Ésta se encuentra en desequilibrio de ligamiento ($r^2 = 0.96$) con la variante rs780094 (no encontrada en este estudio, probablemente por localizarse en una región intrónica y no estar cubierta), y ambas se han asociado a diversos rasgos metabólicos en diferentes estudios de población caucásica (Mohás et al., 2010; Petit et al., 2016; Petta et al., 2014). Más concretamente, la presencia del alelo T conlleva un descenso de los niveles plasmáticos de glucosa junto con una mayor sensibilidad a la insulina. Por el contrario, se ha asociado con un aumento de los niveles de TG. El cambio aminoacídico de una prolina por una leucina (p.P446L) alteraría la estructura de la GKR-P, implicando una pérdida de afinidad por la fructosa-6-fosfato. Este hecho, impediría la formación del complejo inhibitorio GKR-P-GCK en el núcleo y como consecuencia, la GCK vería aumentada su función en el citoplasma promoviendo la entrada de glucosa, y procesos como la glucólisis y la síntesis de TG

(Figura 23). Por tanto, esta variante actuaría protegiendo contra la DM2, no obstante, podría asociarse con la presencia de ácido graso no alcohólico y obesidad (Raimondo et al., 2015).

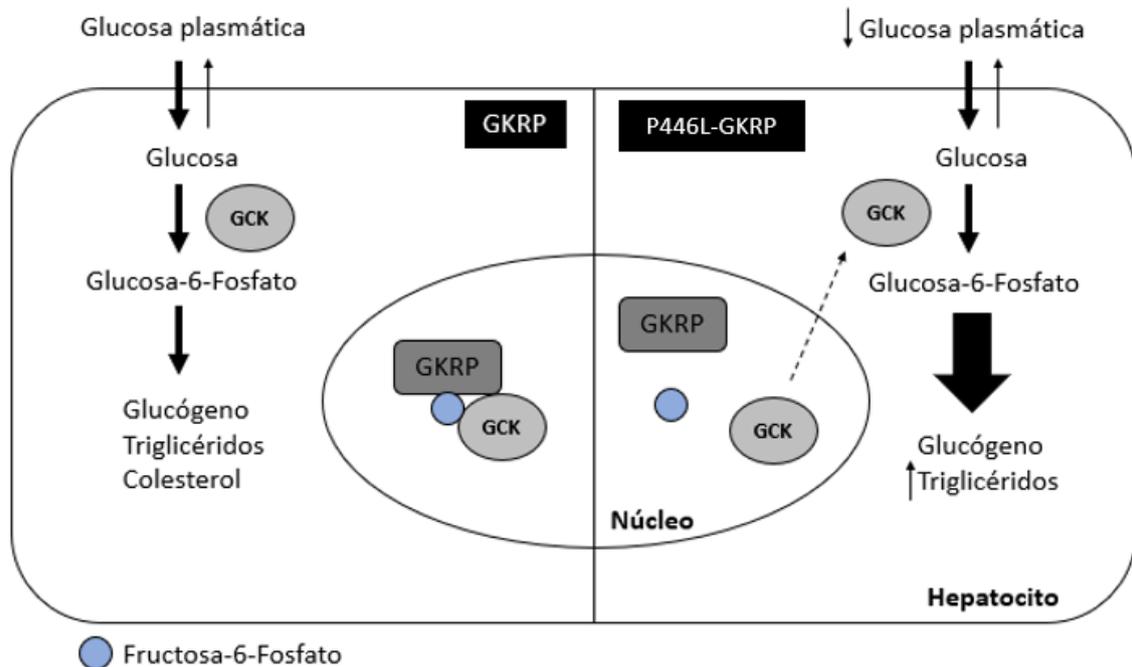


Figura 23. Representación del efecto de la variante P446L en el metabolismo de la glucosa. Izquierda: en estado de ayuno, GKRPs mantiene retenida a GCK en el núcleo gracias a la presencia de F6F. Cuando los niveles de glucosa en plasma se elevan, este complejo se disocia permitiendo la translocación de GCK al citoplasma para promover la entrada de glucosa y permitir su almacenamiento. Derecha: la variante P446L disminuiría la afinidad entre GKRPs y F6F, reduciendo los niveles plasmáticos de glucosa y aumentando la lipogénesis. GKRPs: proteína reguladora de la glucoquinasa; GCK: glucoquinasa; F6F: fructosa-6-fosfato. Adaptada de (Raimondo et al., 2015).

4.7.4 MANBA

El gen *MANBA*, localizado en el cromosoma 4, codifica para la enzima β -manosidasa. Se trata de una proteína de expresión ubicua y localizada en los lisosomas de las células. Es un tipo de exoglicosidasa que participa en el catabolismo de las N-glicoproteínas, concretamente actúa hidrolizando el residuo terminal de manosa unido por enlace $\beta(1-4)$ del extremo no reductor las mismas. Mutaciones en *MANBA* son causantes del desarrollo de β -manosidosis hereditaria, una enfermedad de almacenamiento lisosomal (Gytz et al., 2019; Suzuki, 2016).

En la *Tabla 16* se muestran las variantes estadísticamente significativas con un porcentaje de genotipado de al menos el 70%. En este caso, resultaron asociadas con los niveles de CT y colesterol HDL. No se encontró ninguna asociación significativa entre las variantes y la DM2 u obesidad.

Tabla 16. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen MANBA.

4-103555952-T-C					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
CT (mg/dL)	194.73 ± 39.57	204.56 ± 41.65	-	0.046*	4183	60	-
4-103611845-C-T (rs227368)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
CT (mg/dL)	194.62 ± 41.04	193.65 ± 40.36	198.46 ± 38.29	0.036*	858	1306	541
4-103681928-C-T (rs776864457)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
HDL (mg/dL)	51.45 ± 12.82	65.64 ± 8.36	-	0.039*	4302	3	-

CT: Colesterol total; HDL: Lipoproteínas de alta densidad. Datos expresados como media ± desviación estándar.

*Nivel de significatividad ≤ 0.05.

Las variantes 4-103555952-T-C y rs227368 resultaron asociadas con los niveles de CT. Mientras que el primer SNP no ha sido descrito en la bibliografía, un estudio pendiente de publicación por el grupo UGD ha visto que la variante rs227368 causa una reducción de la actividad enzimática en la β-manosidasa, asociándose con un incremento del IMC. Además, estudios en humanos y conejos han relacionado altos niveles plasmáticos de N-glicanos de alto contenido en manosa con un aumento del CT (Bai et al., 2016), lo que podría estar asociado con los resultados obtenido en este estudio.

Por otro lado, la variante rs776864457 ha resultado significativa con los niveles de colesterol HDL. No se han encontrado datos bibliográficos de esta variante, pero se observa que los homocigotos para el alelo de referencia (alelo C) presentan niveles más bajos de colesterol HDL, que los 3 heterocigotos genotipados. Debido a que se trata del tipo de colesterol que reduce la probabilidad de aterogenicidad podría ser una variante protectora, aunque habría que aumentar el poder estadístico genotipando a más individuos y realizando estudios funcionales que esclareciesen dicha hipótesis.

A pesar de no haber encontrado relación con la DM2, diferentes estudios, no relacionados directamente con polimorfismos en el gen MANBA, han observado una asociación entre niveles altos de manosa en plasma y la presencia de RI, aumento del IMC y DM2, aunque se desconocen los mecanismos moleculares implicados (Mardinoglu et al., 2017). La realización de más estudios sobre la genética de la enzima β-manosidasa podrían ayudar a esclarecer la posible asociación existente entre el metabolismo energético de la manosa y glicoproteínas asociadas, y la posible aparición de enfermedades metabólicas como DM2, RI o hipercolesterolemia.

4.7.5 SREBF2

El gen *SREBF2* codifica para la proteína de unión al elemento regulador de colesterol 2 (SREBP2), que junto a otros dos miembros (SREBP-1a y SREBP-1c), forma parte de una subfamilia de factores de transcripción con dominio de cremallera de leucina hélice-bucle-hélice. Las proteínas SREBPs interactúan directamente con el ADN promoviendo la expresión de genes relacionados con el metabolismo del colesterol, ácidos grasos y TG (DeBose-Boyd & Ye, 2018).

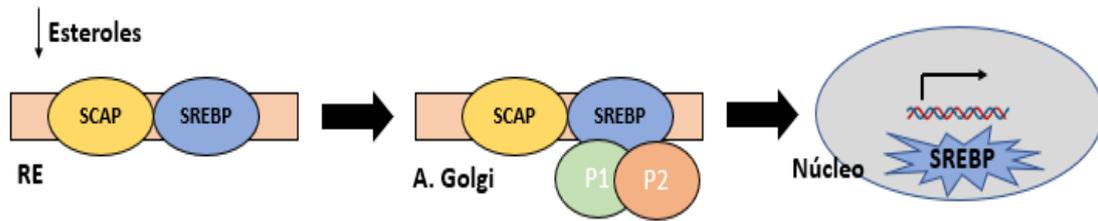


Figura 24. Representación de la activación proteolítica de las proteínas SREBP. Cuando los niveles de esteroides disminuyen y son reconocidos por SCAP, las proteínas SREBPs ancladas en la membrana del RE en su forma inactiva se transportan al aparato de Golgi, donde son procesadas por dos proteasas. Esto conduce a la liberación del dominio de unión al ADN de SREBP, el cual se transloca al núcleo e inicia la expresión de genes relacionados con el metabolismo de lípidos. Adaptada de (Bloise & Reinhart, 1981).

Se encontraron siete variantes de cambio de sentido asociadas a diferentes rasgos metabólicos. Cuatro de ellas (rs147075621, rs200675940, 22-42264791-G-T y rs765731064) se asociaron con el perfil lipídico, especialmente con niveles de CT, LDL y HDL. El SNP rs765731064, junto con el rs74792944, también se asoció con el IMC. Finalmente, las variantes 22-42271667-C-T y 22-42269857-A-C se asociaron con niveles de insulina y glucosa, respectivamente (Tabla 17).

Tabla 17. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen SREBF2.

22-42271667-C-T					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
Insulina (µU/mL)	70.22 ± 131.75	40.36 ± 12.76	-	0.016*	4332	9	-
22-42263057-C-T (rs147075621)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
LDL (mg/dL)	104.48 ± 29.55	138.61 ± 12.53	-	0.030*	4021	3	-
CT (mg/dL)	195.62 ± 39.70	250.55 ± 9.57	-	0.013*			
22-42273963-G-A (rs200675940)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
CT (mg/dL)	195.51 ± 39.74	172.70 ± 23.35	-	0.046*	3970	9	-

RESULTADOS Y DISCUSIÓN

22-42269857-A-C					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
Glucosa (mg/dL)	98.92 ± 29.95	103.78 ± 10.23	-	0.035*	4405	9	-
22-42289229-C-T (rs747929744)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
IMC (Kg/m ²)	28.04 ± 5.17	35.16 ± 3.14	-	0.049*	4556	2	-
22-42264791-G-T					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
LDL (mg/dL)	104.47 ± 29.40	97.58 ± 32.87	-	0.020*	3578	87	-
22-42294739-C-T (rs765731064)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
IMC (Kg/m ²)	28.08 ± 5.17	22.40 ± 0.86	-	0.022*	4291	3	-
HDL (mg/dL)	51.47 ± 12.84	73.10 ± 16.09	-	0.022*			
CT (mg/dL)	2.02 ± 0.22	1.74 ± 0.04	-	0.013*			

LDL: Lipoproteínas de baja densidad; CT: Colesterol total; IMC: Índice de Masa Corporal; HDL: Lipoproteínas de alta densidad. Datos expresados como media ± desviación estándar. *Nivel de significatividad ≤ 0.05.

Ninguno de los siete SNPs encontrados ha sido descrito previamente en la bibliografía. No obstante, atendiendo a los resultados obtenidos y la asociación con numerosas variables clínicas, este gen podría sugerir un riesgo tanto para DM2 como dislipemias.

Las proteínas SREBPs son factores de transcripción implicados en la regulación del metabolismo lipídico. Mientras que SREBF1 regula la expresión de genes relacionados con la síntesis de ácidos grasos y TG, SREBF2 regula la expresión de genes implicados en la síntesis e internalización de colesterol en las células, tales como HMG-CoA sintetasa, HMG-CoA reductasa o el receptor de LDL (DeBose-Boyd & Ye, 2018). Por tanto, variaciones genéticas en estos factores de transcripción conllevarían una desregulación del metabolismo lipídico. Diversas publicaciones no solo han contrastado esta hipótesis, sino que indican que SREBP2 parece causar alteraciones en la glucemia y secreción de insulina. Un estudio basado en ratones transgénicos demostró que la sobreexpresión de SREBF2 en islotes pancreáticos aumentaba tanto la expresión de los genes que regula para la síntesis de colesterol como los niveles del mismo. Además, también se observó que estos islotes presentaban un descenso en el número y masa celular, lo que se relacionó con la menor tasa de secreción de insulina, una hiperglucemia crónica y el desarrollo de DM2 (Ishikawa et al., 2008). Estos datos podrían indicar que la disfunción pancreática inducida por lipotoxicidad conduce al riesgo de DM2.

Estudios posteriores han relacionado la posible interacción existen entre SREBF2 y pacientes con DM2, si bien en el contexto de otras enfermedades. Un estudio de casos-controles para EAC y DM2 demostró que éstos presentaban una mayor tasa de expresión de SREBF2, acompañándose de hipertensión, hiperglucemia, un aumento de los niveles de CT, LDL y un descenso en los

niveles de HDL, en comparación con pacientes EAC no diabéticos y controles (Pérez-Belmonte et al., 2017). En otro estudio se analizó el grado de asociación de la variante rs133291-C/T en hígado graso no alcohólico y DM2. Los pacientes portadores del alelo T mostraron una mayor probabilidad de desarrollar DM2 en el tiempo, puesto que se observó una elevada RI en tejidos clave como hígado, músculo y adiposo, así como una disfunción de las células β pancreáticas (Musso, Cassader, Bo, De Michieli, & Gambino, 2013).

Algunas de estas observaciones podrían correlacionarse con los SNPs encontrados en este trabajo. Por un lado, en las variantes 22-42271667-C-T, rs147075621 y 22-42269857-A-C se observa que los individuos heterocigotos presentan niveles más bajos de insulina, y mayores de CT y glucosa, respectivamente. Esto podría indicar un factor de riesgo para la aparición de RI y DM2. Por el contrario, las variantes rs200675940, 22-42264791-G-T y rs765731064 podrían ser protectoras, ya que los portadores del alelo menos frecuente presentan niveles más bajos de CT, LDL y mayores de HDL. Cabe destacar que se trata de hipótesis, las cuales deben ser probadas con estudios adicionales de caso-control que permitiesen aumentar el poder estadístico, y/o funcionales.

4.7.6 TCF7L2

El gen TCF7L2 (Factor de transcripción 7 tipo 2) codifica para un factor de transcripción que media en la vía de señalización canónica de Wnt, gracias a su interacción con la molécula β -catenina. De este modo, la unión del ligando Wnt a su receptor desencadena una cascada de señalización que promueve la translocación de la β -catenina al núcleo celular, lugar donde interacciona con el receptor nuclear TCF7L2. A su vez, TCF7L2, gracias a la presencia de su dominio HMG-box (del inglés *High Mobility Group – box*), puede unirse al ADN induciendo la transcripción de genes implicados en el desarrollo embrionario, proliferación, migración y determinación celular (Smith, 2007).

Se encontraron tres variantes de cambio de sentido, 10-114710643-A-C, rs201257936 y rs77673441, que resultaron significativas para los niveles de CT, HOMA-IR e IMC, respectivamente (*Tabla 18*).

Tabla 18. Relación entre las variantes significativas y las variables clínico-bioquímicas para el gen TCF7L2.

10-114710643-A-C					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
CT (mg/dL)	195.59 ± 39.70	252.22 ± 45.13	-	0.033*	4381	3	-
10-114925405-C-G (rs201257936)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
HOMA-IR	2.57 ± 7.02	4.6 ± 7.02	-	0.041*	4493	2	-
10-114925406-C-A (rs77673441)					N.º Individuos		
	0/0	0/1	1/1	p-valor	0/0	0/1	1/1
IMC (Kg/m ²)	28.06 ± 5.16	31.06 ± 2.69	-	0.037*	4474	8	-

CT: Colesterol total; HOMA-IR: Modelo Homeostático de Resistencia a la Insulina; IMC: Índice de Masa Corporal. Datos expresados como media ± desviación estándar. *Nivel de significatividad ≤ 0.05.

No se encontraron datos bibliográficos sobre ninguno de los tres SNPs. No obstante, se observó que los heterocigotos de cada una de las variantes presentaron mayores valores medios de CT, HOMA-IR e IMC, respectivamente, en comparación con los individuos homocigotos para el alelo de referencia. Este hecho podría indicar *a priori* que representan variantes de riesgo para estas variables clínicas.

Por otro lado, se observó que el SNP rs77673441, relacionado con el IMC según los datos obtenidos, también supuso un factor de riesgo para el desarrollo de obesidad (Tabla 19). Este hallazgo es contrario a los resultados obtenidos en estudios previos basados en exoma por el grupo UGD (Martínez, 2015), en los que la variante rs77673441 encontrada únicamente en controles supondría un efecto protector contra la enfermedad. Por esta razón, es conveniente llevar a cabo estudios adicionales que permitan establecer el papel de estos polimorfismos raros en la DM2.

Tabla 19. Variante asociada a la obesidad mediante la prueba exacta de Fisher.

							N.º Individuos		
cr:posición	Alelos	rs	Tipo	MAF (EUR)	OR	p-valor	0/0	0/1	1/1
10-114925406	C/A	rs77673441	CS	0.001	13.55	0.004*	4474	8	-

CS: Cambio de Sentido; MAF: Frecuencia Alélica Mínima; OR: *ODDS Ratio*; *Nivel de significatividad ≤ 0.05.

Hasta la fecha, TCF7L2 es el gen estudiado con mayor grado de asociación a la DM2. El primer SNP relacionado con este desorden metabólico surgió de un estudio de casos-controles en población caucásica (Grant et al., 2006), en el que se observó que los portadores de la variante rs7903146 (IVS3C>T) presentaban un mayor riesgo de desarrollar dicha enfermedad. Con el tiempo, este hallazgo se replicó en diferentes grupos étnicos mediante estudios de metaanálisis (Ding et al., 2018).

Desde entonces, una gran variedad de SNPs comunes y poco frecuentes han sido objeto de estudio para ver su grado de implicación en la DM2. A pesar de ello, los mecanismos moleculares responsables que podrían conferir un riesgo o protección frente a la DM2 no están totalmente establecidos, pero la participación de este factor de transcripción en la vía de señalización Wnt ha permitido establecer diversas hipótesis.

Por un lado, se han realizado estudios que revelan el papel de este factor de transcripción en distintos órganos metabólicos. Por ejemplo, el silenciamiento del gen TCF7L2 en islotes pancreáticos humanos y murinos mostró una desregulación de las células β pancreáticas, conllevando al descenso de su proliferación y entrada en apoptosis (Shu, Sauter, Schulthess, & Matveyenko, 2008). Del mismo modo, TCF7L2 también se ha estudiado en el tejido adiposo, donde se ha visto que la depleción de esta proteína conduce a un engrosamiento de los adipocitos, inflamación y en última instancia RI (Chen et al., 2018).

Por otro lado, una de las hipótesis más plausibles hace referencia al papel clave que podría jugar TCF7L2 en la regulación de la homeostasis de la glucosa regulando la transcripción del gen proglucagón (*GCG*) en las células L intestinales y β pancreáticas (Shu et al., 2009). El gen *GCG* codifica para la hormona incretina péptido similar al glucagón tipo 1 (GLP-1), el cual está implicado directamente en la homeostasis de la glucosa mediante la estimulación de insulina e inhibición del glucagón. Por tanto, no resulta raro que SNPs en TCF7L2 puedan inducir DM2 mediante la regulación de este elemento génico.

Por tanto, dada la importancia que presenta este gen en el contexto de la DM2, es imperativo seguir realizando estudios funcionales que permitan descubrir los mecanismos moleculares subyacentes, ya que podrían descubrirse nuevas dianas terapéuticas para esta enfermedad.

4.8 LIMITACIONES DEL ESTUDIO

Primeramente, respecto al proceso de optimización y análisis de coberturas se obtuvo una profundidad y amplitud de cobertura poco homogénea entre los distintos fragmentos. Dado el número de cebadores implicados, las características fisicoquímicas de cada uno y el tiempo limitado, no se prosiguió con el proceso para hallar las condiciones óptimas de amplificación para todos ellos. Esto pone de manifiesto la necesidad de continuar trabajando en este paso, con el objetivo de conseguir cubrir el máximo número de regiones posibles que aporten más fiabilidad a los resultados. Para ello, sería conveniente llevar a cabo un rediseño de cebadores, un reajuste de concentraciones más preciso o incluso valorar la posibilidad de introducir una tercera multiplex.

Otra de las limitaciones ha sido el número de variantes a analizar. En el presente trabajo se han considerado las variantes de efecto moderado (siendo en este caso de cambio de sentido) genotipadas en más del 70% de la población, ya que las de efecto alto no resultaron significativas, quizá por su baja frecuencia en la población. A pesar de ello, sería conveniente estudiarlas y ver si han sido descritas en relación con la DM2. Del mismo modo, también habría que considerar el estudio de variantes con efecto más bajo como las intrónicas o de *splicing*, puesto que la DM2 se ha asociado a bastantes polimorfismos de este tipo. Por último, dada la cobertura obtenida en las muestras, alrededor del 40% de las variantes no fueron consideradas, pero una mejora en las condiciones de PCR y secuenciación que proporcionen una cobertura mayor podrían permitir analizarlas en trabajos posteriores y/o repetir aquellas muestras con menor cobertura. Por otro lado, la mejora en la calidad de las muestras posiblemente supondría un estudio mucho más homogéneo.

Por último, el poder estadístico está limitado para establecer asociaciones significativas fiables, dada la baja frecuencia de la mayor parte de los polimorfismos. Este hecho puede dar lugar a considerar con mayor probabilidad más falsos positivos, por lo que las variantes obtenidas deben ser validadas en estudios venideros, que permitan aumentar tanto el número muestral como la mayor fiabilidad de los datos. Además, hay que tener en cuenta que no se ha aplicado una corrección por el número de comparaciones dado que el presente estudio es, en sí mismo, la replicación de datos previos.

Por tanto, debido a la gran cantidad de datos obtenidos y el tiempo limitado, el presente trabajo debe considerarse como un estudio inicial en el que se necesita seguir trabajando con el análisis de variantes, refinar estudios estadísticos y aumentar el número de muestras.

5 CONCLUSIONES

1. La secuenciación masiva de amplicones es una buena herramienta para la identificación y validación de variantes en regiones genómicas de interés. No obstante, es necesario implementar un buen protocolo de optimización para conseguir una cobertura media mayor y más homogénea.
2. La secuenciación de 4784 individuos pertenecientes al estudio Di@bet.es ha permitido obtener 654 variantes genotipadas en más del 70% de la población, siendo en su mayoría SNPs de cambio de sentido.
3. Con relación a los estudios previos basados en exoma, en el presente proyecto se han validado los polimorfismos rs41279188 y rs77673441 pertenecientes a los genes CYP1A1 y TCF7L2, respectivamente. Sin embargo, es necesario realizar futuros estudios funcionales que permitan analizar su papel en el contexto de la DM2.
4. Respecto al panel de genes estudiado, se han identificado 24 variantes de cambio de sentido asociadas con alguna variable correspondiente al perfil glucídico, perfil lipídico o IMC. Además, los SNPs rs56343424 y rs77673441 correspondientes a los genes CYP1A1 y TCF7L2 fueron asociados a la DM2 y obesidad, respectivamente.
5. El número de variantes encontradas en los genes ACAD10, CYP1A1, GCKR, MANBA, SREBF2 y TCF7L2, especialmente de baja frecuencia, podría explicar parte del componente genético de la DM2, contribuyendo a la hipótesis de ‘variante rara – enfermedad común’. Sin embargo, se necesita realizar un estudio más exhaustivo de todas las variantes para valorar su grado de implicación en la enfermedad.
6. La confirmación de estos resultados precisa de su validación en otras poblaciones que permitan aumentar el poder estadístico, así como la realización de futuros estudios funcionales que permitan dilucidar los mecanismos moleculares implicados y la identificación de potenciales dianas terapéuticas.

6 BIBLIOGRAFÍA

- Ahlqvist, E., Ahluwalia, T. S., & Groop, L. (2011). Genetics of type 2 diabetes. *Clinical Chemistry*, 57(2), 241–254. <https://doi.org/10.1373/clinchem.2010.157016>.
- Ali, O. (2013). Genetics of type 2 diabetes. *World Journal of Diabetes*, 4(4), 114–123. <https://doi.org/10.4239/wjd.v4.i4.114>
- American Diabetes Association (2019). Classification and diagnosis of diabetes: Standards of medical care in diabetesd-2019-. *Diabetes Care*, 42, S13–S28. <https://doi.org/10.2337/dc19-S002>
- Bai, L., Li, Q., Li, L., Lin, Y., Zhao, S., Wang, W., ... Liu, E. (2016). Plasma High-mannose and complex/Hybrid N-Glycans are associated with hypercholesterolemia in humans and rabbits. *PLoS ONE*, 11(3), 1–16. <https://doi.org/10.1371/journal.pone.0146982>
- Bansal, N. (2015). Prediabetes diagnosis and treatment: A review. *World Journal of Diabetes*, 6(2), 296–303. <https://doi.org/10.4239/wjd.v6.i2.296>
- Bian, L., Hanson, R. L., Muller, Y. L., Ma, L., Kobes, S., Knowler, W. C., ... Baier, L. J. (2010). Variants in ACAD10 are associated with type 2 diabetes, insulin resistance and lipid oxidation in Pima Indians. *Diabetologia*, 53(7), 1349–1353. <https://doi.org/10.1007/s00125-010-1695-y>
- Bloise, R., & Reinhart, C. (1981). SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *Industrie Minerale, Les Techniques*, 109(6), 437–441. <https://doi.org/10.1172/JCI200215593>.Lipid
- Bloom, K., Mohsen, A. W., Karunanidhi, A., El Demellawy, D., Reyes-Múgica, M., Wang, Y., ... Vockley, J. (2018). Investigating the link of ACAD10 deficiency to type 2 diabetes mellitus. *Journal of Inherited Metabolic Disease*, 41(1), 49–57. <https://doi.org/10.1007/s10545-017-0013-y>
- Bodmer, W., & Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics*, 40(6), 695–701. <https://doi.org/10.1038/ng.f.136>
- Bonnefond, A., & Froguel, P. (2015). Rare and common genetic events in type 2 diabetes: What should biologists know? *Cell Metabolism*, 21(3), 357–368. <https://doi.org/10.1016/j.cmet.2014.12.020>
- Chatterjee, S., Khunti, K., & Davies, M. J. (2017). Type 2 diabetes. *The Lancet*, 389(10085), 2239–2251. [https://doi.org/10.1016/S0140-6736\(17\)30058-2](https://doi.org/10.1016/S0140-6736(17)30058-2)
- Chen, X., Ayala, I., Shannon, C., Fourcaudot, M., Acharya, N. K., Jenkinson, C. P., ... Norton, L. (2018). The diabetes gene and wnt pathway effector TCF7L2 regulates adipocyte development and function. *Diabetes*, 67(4), 554–568. <https://doi.org/10.2337/db17-0318>
- Cho, N. H., Kirigia, J., Mbanaya, J. C., Ogurstova, K., Guariguata, L., Rathmann, W., ... Reja, A. (2017). Diabetes Atlas de la FID. In *Diabetes India and Jaslok Hospital President-Elect International Diabetes Federation* (Vol. 8). <https://doi.org/10.1016/j.diabres.2017.09.002>

- CIBERDEM. (2018). Estudio di@bet.es: Cerca de 400.000 personas desarrollan diabetes cada año en España Centro. *Centro de Investigación Biomédica En Red. Diabetes y Enfermedades Metabólicas Asociadas*. <https://doi.org/https://www.ciberdem.org/noticias/estudio-diabetes-cerca-de-400000-personas-desarrollan-diabetes-cada-ano-en-espana>
- DeBose-Boyd, R. A., & Ye, J. (2018). SREBPs in Lipid Metabolism, Insulin Signaling, and Beyond. *Trends in Biochemical Sciences*, 43(5), 358–368. <https://doi.org/10.1016/j.tibs.2018.01.005>
- DeFronzo, R. A., Ferrannini, E., Groop, L., Henry, R. R., Herman, W. H., Holst, J. J., ... Weiss, R. (2015). Type 2 Diabetes Mellitus. *Nature Reviews Disease Primers*, 1(15019), 1–22. <https://doi.org/doi.org/10.1038/nrdp.2015.19>
- Ding, W., Xu, L., Zhang, L., Han, Z., Jiang, Q., Wang, Z., & Jin, S. (2018). Meta-analysis of association between TCF7L2 polymorphism rs7903146 and type 2 diabetes mellitus. *BMC Medical Genetics*, 19(1), 1–12. <https://doi.org/10.1186/s12881-018-0553-5>
- Dziewulska, A., Dobosz, A. M., & Dobrzyn, A. (2018). High-throughput approaches onto uncover (Epi)genomic architecture of type 2 diabetes. *Genes*, 9(8). <https://doi.org/10.3390/genes9080374>
- Galán Chilet, I. (2016). *Identificación de variantes genéticas poco frecuentes y raras en Diabetes mellitus tipo 2 mediante secuenciación de exoma. Tesis doctoral presentada por: Inmaculada Galán Chilet.*
- Grant, S. F. A., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., ... Stefansson, K. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature Genetics*, 38(3), 320–323. <https://doi.org/10.1038/ng1732>
- Gutiérrez-Rodelo, C., Roura-Guiberna, A., & Olivares-Reyes, J. A. (2017). Molecular Mechanisms of Insulin Resistance: An Update. *Gaceta Medica de Mexico*, 153(2), 197–209.
- Gytz, H., Liang, J., Liang, Y., Gorelik, A., Illes, K., & Nagar, B. (2019). The structure of mammalian β -mannosidase provides insight into β -mannosidosis and nystagmus. *FEBS Journal*, 286(7), 1319–1331. <https://doi.org/10.1111/febs.14731>
- Ishikawa, M., Iwasaki, Y., Yatoh, S., Kato, T., Kumadaki, S., Inoue, N., ... Shimano, H. (2008). Cholesterol accumulation and diabetes in pancreatic β -cell-specific SREBP-2 transgenic mice: A new model for lipotoxicity. *Journal of Lipid Research*, 49(12), 2524–2534. <https://doi.org/10.1194/jlr.M800238-JLR200>
- Islam, M. S., Ahmed, M. U., Bin Sayeed, M. S., Al Maruf, A., Mostofa, A. G. M., Akram Hussain, S. M., ... Hasnat, A. (2013). Lung cancer risk in relation to nicotinic acetylcholine receptor, CYP2A6 and CYP1A1 genotypes in the Bangladeshi population. *Clinica Chimica Acta*, 416, 11–19. <https://doi.org/10.1016/j.cca.2012.11.011>
- Kalendar, R., Lee, D., & Schulman, A. H. (2009). FastPCR software for PCR primer and probe design and repeat search. *Genes, Genomes and Genomics*, 3(1), 1–14.
- Kaprio, J., Tuomilehto, J., Koskenvuo, M., Romanov, K., Reunanen, A., Eriksson, J., ... Kesäniemi, Y. A. (1992). Concordance for Type 1 (insulin-dependent) and Type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia*, 35(11), 1060–1067. <https://doi.org/10.1007/BF02221682>

- Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine*, *09*(03). <https://doi.org/10.4172/0974-8369.1000395>
- Kharroubi, A. T., & Darwish, H. M. (2015). Diabetes mellitus: The epidemic of the century. *World Journal of Diabetes*, *6*(6), 850–867. <https://doi.org/10.4239/wjd.v6.i6.850>
- Kolb, H., & Martin, S. (2017). Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. *BMC Medicine*, *15*(1), 1–11. <https://doi.org/10.1186/s12916-017-0901-x>
- Langenberg, C., & Lotta, L. A. (2018). Genomic insights into the causes of type 2 diabetes. *The Lancet*, *391*(10138), 2463–2474. [https://doi.org/10.1016/S0140-6736\(18\)31132-2](https://doi.org/10.1016/S0140-6736(18)31132-2)
- Mardinoglu, A., Stančáková, A., Lotta, L. A., Kuusisto, J., Boren, J., Blüher, M., ... Smith, U. (2017). Plasma Mannose Levels Are Associated with Incident Type 2 Diabetes and Cardiovascular Disease. *Cell Metabolism*, *26*(2), 281–283. <https://doi.org/10.1016/j.cmet.2017.07.006>
- Martínez, V. (2015). *Identificación de variantes exómicas en diabetes tipo 2. Tesis doctoral presentada por: Vanesa Martínez Barquero*. Retrieved from <https://core.ac.uk/download/pdf/71040565.pdf>
- Mescher, M., & Haarmann-Stemmann, T. (2018). Modulation of CYP1A1 metabolism: From adverse health effects to chemoprevention and therapeutic options. In *Pharmacology and Therapeutics* (Vol. 187). <https://doi.org/10.1016/j.pharmthera.2018.02.012>
- Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, *11*(1), 31–46. <https://doi.org/10.1038/nrg2626>
- Mohás, M., Kísfali, P., Járomi, L., Maász, A., Fehér, E., Csöngéi, V., ... Melegh, B. (2010). GCKR gene functional variants in type 2 diabetes and metabolic syndrome: Do the rare variants associate with increased carotid intima-media thickness? *Cardiovascular Diabetology*, *9*, 1–7. <https://doi.org/10.1186/1475-2840-9-79>
- Musso, G., Cassader, M., Bo, S., De Michieli, F., & Gambino, R. (2013). Sterol regulatory element-binding factor 2 (SREBF-2) predicts 7-year NAFLD incidence and severity of liver disease and lipoprotein and glucose dysmetabolism. *Diabetes*, *62*(4), 1109–1120. <https://doi.org/10.2337/db12-0858>
- Nathan, D. M. (2015). Diabetes: Advances in diagnosis and treatment. *JAMA - Journal of the American Medical Association*, *314*(10), 1052–1062. <https://doi.org/10.1001/jama.2015.9536>
- Olokoba, A. B., Obateru, O. A., & Olokoba, L. B. (2012). Type 2 diabetes mellitus: a review of current trends. *Oman Medical Journal*, *27*(4), 269–273. <https://doi.org/10.5001/omj.2012.68>
- Papatheodorou, K., Banach, M., Bekiari, E., Rizzo, M., & Edmonds, M. (2018). Complications of Diabetes 2017. *Journal of Diabetes Research*, *2018*, 1–4. <https://doi.org/10.1155/2018/3086167>

- Pérez-Belmonte, L. M., Moreno-Santos, I., Cabrera-Bueno, F., Sánchez-Espín, G., Castellano, D., Such, M., ... Jiménez-Navarro, M. (2017). Expression of sterol regulatory element-binding proteins in epicardial adipose tissue in patients with coronary artery disease and diabetes mellitus: Preliminary study. *International Journal of Medical Sciences*, 14(3), 268–274. <https://doi.org/10.7150/ijms.17821>
- Petit, J. M., Masson, D., Guiu, B., Rollot, F., Duvillard, L., Bouillet, B., ... Verges, B. (2016). GCKR polymorphism influences liver fat content in patients with type 2 diabetes. *Acta Diabetologica*, 53(2), 237–242. <https://doi.org/10.1007/s00592-015-0766-4>
- Petta, S., Miele, L., Bugianesi, E., Cammà, C., Rosso, C., Boccia, S., ... Craxì, A. (2014). Glucokinase regulatory protein gene polymorphism affects liver fibrosis in non-alcoholic fatty liver disease. *PLoS ONE*, 9(2), 1–7. <https://doi.org/10.1371/journal.pone.0087523>
- Prasad, R. B., & Groop, L. (2015). Genetics of type 2 diabetes—pitfalls and possibilities. *Genes*, 6(1), 87–123. <https://doi.org/10.3390/genes6010087>
- Punthakee, Z., Goldenberg, R., & Katz, P. (2018). Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. *Canadian Journal of Diabetes*, 42, 10–15. <https://doi.org/10.1016/j.cjcd.2017.10.003>
- Raimondo, A., Rees, M. G., & Gloyn, A. L. (2015). Glucokinase regulatory protein: Complexity at the crossroads of triglyceride and glucose metabolism. *Current Opinion in Lipidology*, 26(2), 88–95. <https://doi.org/10.1097/MOL.0000000000000155>
- Rojas, J., Velasco, M., Tomey, D., Olivar, L. C., Bermudez, V., Palmar, J., ... Nava, M. (2018). Pancreatic Beta Cell Death: Novel Potential Mechanisms in Diabetes Therapy. *Journal of Diabetes Research*, 2018, 1–19. <https://doi.org/10.1155/2018/9601801>
- Santes-Palacios, R., Ornelas-Ayala, D., Cabañas, N., Marroquín-Pérez, A., Hernández-Magaña, A., Del Rosario Olgún-Reyes, S., ... Espinosa-Aguirre, J. J. (2016). Regulation of Human Cytochrome P4501A1 (hCYP1A1): A Plausible Target for Chemoprevention? *BioMed Research International*, 2016. <https://doi.org/10.1155/2016/5341081>
- Seda García, E. (2018). *DESARROLLO DE UN SISTEMA DE DETECCIÓN DE MUTACIONES SOMÁTICAS DE INTERÉS CLÍNICO MEDIANTE SECUENCIACIÓN MASIVA EN MUESTRAS ONCOLÓGICAS. Tesis doctoral presentada por: Enrique Seda García.*
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–1145. <https://doi.org/10.1038/nbt1486>
- Shu, L., Matveyenko, A. V., Kerr-Conte, J., Cho, J. H., McIntosh, C. H. S., & Maedler, K. (2009). Decreased TCF7L2 protein levels in type 2 diabetes mellitus correlate with downregulation of GIP- and GLP-1 receptors and impaired beta-cell function. *Human Molecular Genetics*, 18(13), 2388–2399. <https://doi.org/10.1093/hmg/ddp178>
- Shu, L., Sauter, N. S., Schulthess, F. T., & Matveyenko, A. V. (2008). *Transcription Factor 7-Like 2 Regulates β-Cell Survival and Function in Human Pancreatic Islets*. 57(March), 645–653. <https://doi.org/10.2337/db07-0847.N.S.S>
- Singh, R. R. (2017). Next generation sequencing technologies. *Comprehensive Medicinal Chemistry III*, 2–8, 354–361. <https://doi.org/10.1016/B978-0-12-409547-2.12327-3>

- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, 122(1), e59. <https://doi.org/10.1002/cpmb.59>
- Smith, U. (2007). TCF7L2 and type 2 diabetes - We WNT to know. *Diabetologia*, 50(1), 5–7. <https://doi.org/10.1007/s00125-006-0521-z>
- Soriguer, F., Goday, A., Bosch-Comas, A., Bordiú, E., Calle-Pascual, A., Carmena, R., ... Vendrell, J. (2012). Prevalence of diabetes mellitus and impaired glucose regulation in Spain: The Di@bet.es Study. *Diabetologia*, 55(1), 88–93. <https://doi.org/10.1007/s00125-011-2336-9>
- Stančáková, A., & Laakso, M. (2016). Genetics of type 2 diabetes. *Endocrine Development*, 31, 203–220. <https://doi.org/10.1159/000439418>
- Suzuki, T. (2016). Catabolism of N-glycoproteins in mammalian cells: Molecular mechanisms and genetic disorders related to the processes. *Molecular Aspects of Medicine*, 51, 89–103. <https://doi.org/10.1016/j.mam.2016.05.004>
- Tillil, H., & Kobberling, J. (1987). Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients. *Diabetes*, 36(1), 93–99. <https://doi.org/10.2337/diab.36.1.93>
- Wang, J., Skoog, T., Einarsdottir, E., Kaartokallio, T., Laivuori, H., Grauers, A., ... Jiao, H. (2016). Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. *Scientific Reports*, 6(August), 1–11. <https://doi.org/10.1038/srep33256>
- Wang, X. L., Greco, M., Sim, A. S., Duarte, N., Wang, J., & Wilcken, D. E. L. (2002). Effect of CYP1A1 MspI polymorphism on cigarette smoking related coronary artery disease and diabetes. *Atherosclerosis*, 162(2), 391–397. [https://doi.org/10.1016/S0021-9150\(01\)00723-7](https://doi.org/10.1016/S0021-9150(01)00723-7)
- Ye, X., Ji, C., Zhou, C., Zeng, L., Gu, S., Ying, K., ... Mao, Y. (2004). Cloning and characterization of a human cDNA ACAD10 mapped to chromosome 12q24.1. *Molecular Biology Reports*, 31(3), 191–195. <https://doi.org/10.1023/B:MOLE.0000043622.57408.6b>
- Yeung, R., Hannah-Shmouni, F., Niederhoffer, K., & Walker, M. (2018). Not quite type 1 or type 2, what now? Review of monogenic, mitochondrial, and syndromic diabetes. *Reviews in Endocrine and Metabolic Disorders*, 19, 35–52. <https://doi.org/10.1007/s11154-018-9446-3>
- Zheng, Y., Ley, S. H., & Hu, F. B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature Reviews. Endocrinology*, 14(2), 88–98. <https://doi.org/10.1038/nrendo.2017.151>
- Zou, J. G., Ma, Y. T., Xie, X., Yang, Y. N., Pan, S., Adi, D., ... Chen, B. D. (2015). The association between CYP1A1 genetic polymorphisms and coronary artery disease in the Uygur and Han of China. *Lipids in Health and Disease*, 14(1), 1–12. <https://doi.org/10.1186/s12944-015-0121-3>