

# UNIVERSIDAD POLITÉCNICA DE VALENCIA

DEPARTAMENTO DE BIOTECNOLOGÍA



## IDENTIFICACIÓN DE VARIANTES GENÉTICAS POCO FRECUENTES EN DIABETES MELLITUS TIPO 2 MEDIANTE SECUENCIACIÓN MASIVA DE AMPLICONES

TRABAJO FIN DE MÁSTER EN BIOTECNOLOGÍA BIOMÉDICA

ALUMNA: Tania Otero Rodríguez

TUTOR: Felipe Javier Chaves Martínez

*Curso Académico: 2017/2018*

VALENCIA, 07/09/2018

# RESUMEN

## IDENTIFICACIÓN DE VARIANTES GENÉTICAS POCO FRECUENTES EN DIABETES MELLITUS TIPO 2 MEDIANTE SECUENCIACIÓN MASIVA DE AMPLICONES

La *diabetes mellitus* tipo 2 (DM2) es una epidemia global que afecta a más de 3,6 millones de adultos en España. Se trata de una enfermedad compleja en cuyo desarrollo están implicados tanto factores ambientales, como genéticos. Mediante diferentes estudios a nivel genómico se han identificado más de 120 *loci* relacionados con DM2, sin embargo se estima que el conjunto de las variantes identificadas explica aproximadamente un 10-15% de la heredabilidad. Hasta ahora, los estudios se han centrado en la búsqueda de variantes comunes, pero parte del componente genético podría deberse a variantes génicas funcionales poco frecuentes o raras. En el presente estudio se realizó la búsqueda de variantes poco frecuentes en una población de 1415 individuos (395 diabéticos y 829 controles) mediante secuenciación dirigida. Se secuenció un panel de 10 genes previamente relacionados con DM2 mediante secuenciación de exoma. Para ello se diseñaron los cebadores específicos y se pusieron a punto las condiciones de amplificación para la generación de librerías que permitieran la secuenciación de prácticamente todas las regiones de interés. Las muestras fueron secuenciadas mediante el sistema HiSeq de la plataforma Illumina y se realizó el análisis bioinformático y la anotación para la identificación de las variantes. Fueron genotipadas 372 variantes, 153 de las cuales superaron el filtro del 70% de individuos genotipados. Se estudió la posible relación de las variantes génicas analizadas con la DM2, así como los posibles mecanismos moleculares implicados. Se identificaron un gran número de variantes genéticas raras y poco frecuentes que podrían estar implicadas en riesgo o protección a DM2, siendo necesaria una validación posterior de estos resultados, así como experimentos funcionales que los confirme.

## PALABRAS CLAVE

*Diabetes mellitus* tipo 2, NGS, SNP, variantes raras, heredabilidad, enfermedad compleja, variantes funcionales.

# ABSTRACT

## LOW-FREQUENCY GENETIC VARIANTS IDENTIFICATION IN TYPE 2 *DIABETES MELLITUS* THROUGH AMPLICON MASSIVE SEQUENCING

Type 2 *diabetes mellitus* (DM2) is a global epidemic affecting more than 3.6 million adults in Spain. It is a complex disease whose development involves both environmental factors such as genetic. Through different studies at the genomic level more than 120 *loci* related to DM2 have been identified, however it is estimated that the set of identified variants explains approximately only a 10%-15 of the heritability. So far, studies have focused on the search for common variants, but part of genetic component may be due to rare or low frequency functional variants. In the present study, low frequency variants were studied in 1415 individuals (395 diabetics and 829 controls) through targeted sequencing. A panel of 10 genes previously related to DM2 was sequenced by exome sequencing. To this end, the specific primers were designed and the amplification conditions were set up for the generation of libraries that allowed the sequencing of practically all the regions of interest. Samples were sequenced using the Illumina HiSeq system and bioinformatic analysis and annotation were performed to identify the variants. 372 variants were genotyped, surpassing 153 variants the filter of 70% of genotyped individuals. The relationship between identified gene variants with DM2 and the possible molecular mechanisms involved were studied. It was identified a large number of rare and low-frequency genetic variants that could be involved in risk or protection to DM2, requiring further validation of these results, as well as functional experiments that confirm them.

### KEYWORDS

*Diabetes mellitus* tipo 2, NGS, SNP, rare variants, heredability, complex disease, functional variants.

Alumna: Tania Otero Rodríguez

Valencia, septiembre de 2018

Prof. D. Felipe Javier Chaves Martínez



## AGRADECIMIENTOS

Me gustaría agradecer a Javier Chaves el apoyo y el haberme dejado formar parte de su equipo, un grupo de grandes profesionales que son a su vez una pequeña familia.

Gracias a todos con los que compartí mis horas en el laboratorio por estar siempre dispuestos a enseñarme, en especial a Veronik por su paciencia y todas esas explicaciones que hicieron posible llevar a cabo este trabajo. Del resto, Kike, Carol, Tani, Pilar, Azahara, Loles... me llevo un trocito de aprendizaje de cada uno, a nivel científico, pero sobretodo personal. Os deseo lo mejor.

No me olvido tampoco de esa funjifamilia valenciana que me ha aportado tantos momentos absurdos y me harán recordar esta etapa con una sonrisa.

Por último, gracias a la familia de verdad, la incondicional, la de ADN, porque nadie me dará nunca consejos tan sabios: "Envíalo como esté, si total".

# ÍNDICE

1.	INTRODUCCIÓN.....	1
1.1.	Epidemiología y etiología de la diabetes.....	1
1.2.	Diagnóstico y pronóstico de la diabetes .....	2
1.3.	Clasificación y síntomas de la diabetes .....	3
1.4.	<i>Diabetes Mellitus</i> tipo 2 .....	4
1.5.	Fisiopatología de la DM2.....	5
1.6.	Tecnología NGS .....	7
1.7.	Estudios genéticos en DM2 .....	8
2.	ANTECEDENTES Y OBJETIVOS.....	11
3.	MATERIALES Y MÉTODOS .....	13
3.1.	Selección de la población a estudiar .....	13
3.2.	Diseño del estudio.....	13
3.3.	Obtención de muestras .....	15
3.4.	Preparación de la librería .....	15
3.4.1.	Diseño de <i>primers</i> .....	16
3.4.2.	PCR1 multiplex-generación de amplicones.....	16
3.4.3.	PCR2- Amplificación con códigos de barras- <i>barcodes</i> .....	17
3.4.4.	Electroforesis capilar .....	18
3.4.5.	Generación del <i>pool</i> .....	18
3.4.6.	Purificación.....	19
3.4.7.	Cuantificación.....	19
3.4.8.	Secuenciación.....	20
3.5.	Análisis bioinformático.....	22
3.6.	Análisis estadístico .....	25
3.7.	Análisis funcional.....	25
4.	RESULTADOS Y DISCUSIÓN.....	26
4.1.	Diseño de cebadores .....	26
4.2.	Puesta a punto de la PCR1-multiplex.....	26
4.2.1.	Ajuste de condiciones de la PCR .....	26
4.2.2.	Reajuste en la concentración y rediseño de <i>primers</i> .....	29
4.3.	Electroforesis capilar de las muestras de la población a estudio .....	30
4.4.	Análisis bioinformático.....	30
4.5.	Análisis de variantes por gen. ....	34
4.5.1.	<i>ACPL2</i> .....	34

4.5.2.	<i>AFMID</i> .....	34
4.5.3.	<i>ARHGEF38</i> .....	36
4.5.4.	<i>CACNG5</i> .....	38
4.5.5.	<i>CCT8L2</i> .....	39
4.5.6.	<i>CIAPIN 1</i> .....	40
4.5.7.	<i>DDO</i> .....	41
4.5.8.	<i>OLFM1</i> .....	42
4.5.9.	<i>SLC12A4</i> .....	43
4.5.10.	<i>SLC15A3</i> .....	44
4.6.	Limitaciones .....	45
5.	CONCLUSIONES.....	46
6.	REFERENCIAS BIBLIOGRÁFICAS .....	48

## ÍNDICE DE FIGURAS

Figura 1. Estimación de la prevalencia de diabetes en 2017 en adultos (20 a 79 años).....	2
Figura 2. Interacción entre los genes y el medio ambiente en individuos que mantienen una tolerancia normal a la glucosa y aquellos que desarrollan diabetes tipo 2.....	5
Figura 3. Resumen del diseño del estudio. ....	14
Figura 4. Flujo de trabajo para la preparación de la librería a secuenciar.....	15
Figura 5. Preparación de la librería.....	15
Figura 6. Representación de las interacciones entre las dos multiplexes de cebadores.....	16
Figura 7. Ejemplo de electroforesis capilar utilizando el sistema Qiaxcel.....	18
Figura 8. Esquema del proceso de purificación del ADN amplificado mediante el kit Magsi-NGS PREP®.....	19
Figura 9. Recta de cuantificación del pool de ADN a secuenciar.....	20
Figura 10. Secuenciación Illumina.....	21
Figura 11. Esquema de los formatos (arriba) y software y herramientas (abajo) utilizados en el análisis informático.....	22
Figura 12. Representación de una lectura del secuenciador Miseq en formato FastQ.....	23
Figura 13. Formato de archivo SAM/BAM.....	23
Figura 14. Formato de archivo VCF.....	24
Figura 15. Resultado de la electroforesis capilar mediante el sistema Qiaxcel de las muestras amplificadas sin añadir los reactivos R2 y R3 ya unas concentraciones de R1 de 0, 1,5 y 3 mM .	26
Figura 16. Resultado de la electroforesis capilar mediante el sistema Qiaxcel de las muestras amplificadas añadiendo los reactivos R2 (5 y 10%) y R3 (0,5 y 1%) y a unas concentraciones de R1 de 0, 1,5 y 3 mM.....	27
Figura 17. Visualización por electroforesis capilar de la librería amplificada con la multiplex A (izquierda) y multiplex B (derecha) de cebadores a diferentes concentraciones (0,007, 0,01, 0,013, 0,017).....	28
Figura 18. Representación general de la distribución de las lecturas por amplicón en una de las pruebas de validación.....	28
Figura 19. Representación general de la distribución de las lecturas por amplicón en la prueba de ajuste de concentración de cebadores.....	29
Figura 20. Representación general de las lecturas por amplicón en la prueba de ajuste de concentración de cebadores.....	29
Figura 21. Resultado de la electroforesis capilar mediante el sistema Qiaxcel de las muestras elegidas al azar en una de las placas control.....	30
Figura 22. Porcentaje de variantes por efecto en el total secuenciadas (izquierda) y en las variantes filtradas por 70% de la población genotipada (derecha).....	32
Figura 23. Mecanismo de acción de la enzima arilformamidasa.....	35
Figura 25. Representación de los plots de cobertura de los controles de validación y de las muestras de la población.....	46



## ÍNDICE DE TABLAS

Tabla 1. Valores para el diagnóstico de diabetes según la American Diabetes Association.....	3
Tabla 2. Factores de riesgo a DM2 .....	4
Tabla 3. Algunos genes relacionados con diabetes tipo 2 seleccionadas en Estudios de Asociación de Genoma Completo. ....	9
Tabla 4. Variantes encontradas en los genes seleccionados a estudio en el presente trabajo ..	11
Tabla 5. Características de los genes a estudio en el presente trabajo .....	12
Tabla 6. Valores promedio de edad, IMC y obesidad de los individuos seleccionados a estudio .....	13
Tabla 7. Fases y condiciones utilizadas en la PCR1. ....	17
Tabla 8. Fases y condiciones utilizadas en la PCR2. ....	17
Tabla 9. Clasificación de las variantes secuenciadas antes del filtro del 70% de individuos genotipados.....	31
Tabla 10. Clasificación de las variantes secuenciadas tras el filtro del 70% de individuos genotipados.....	31
Tabla 11 . Variantes encontradas que coincidieron con el estudio previo .....	32
Tabla 12. Variantes estadísticamente significativas (p-valor <0,05).....	33
Tabla 13. Clasificación de variantes missense en el gen AFMID genotipadas en más del 70% de individuos. ....	35
Tabla 14. Clasificación de variantes missense en el gen ARHGEF38 genotipadas en más del 70% de individuos. ....	37
Tabla 15. Variantes 5'UTR genotipadas en >70% en el gen CACNG5.....	39
Tabla 16. Variantes missense genotipadas en >70% en el gen CACNG5. ....	39
Tabla 17. Variantes missense genotipadas en >70% en el gen CIAPIN1.....	41
Tabla 18. Variantes missense genotipadas en >70% en el gen OLFML1.....	43
Tabla 19. Variantes missense genotipadas en >70% en el gen SLC12A4.....	44
Tabla 20. Variantes missense genotipadas en >70% en el gen SLC15A3 .....	44

## ABREVIATURAS

**ACPL2:** Proteína 2 similar a fosfatasa ácida

**AFMID:** Arilformamidasas

**ADA:** Asociación Americana de la Diabetes

**ADN:** Ácido Desoxirribonucleico

**ADNdc:** ADN de doble cadena

**BAM:** Alineamiento/Mapa binario

**CACNG5:** Subunidad Auxiliar Gamma 5 del Canal de Calcio dependiente de Voltaje

**CCT8L2:** Proteína chaperonina que contiene la subunidad 8 tipo 2 de TCP1

**CIAPIN 1:** Inhibidor de la apoptosis inducido por citoquinas 1

**GTPasa:** Trifosfatasas de guanosina

**DDO:** D-aspartato oxidasa

**DM:** *Diabetes Mellitus*

**DM2:** *Diabetes Mellitus* tipo 2

**dNTP:** Desoxirribonucleótido trifosfato

**FOXO:** Familia Forkhead de factores de transcripción.

**GAG:** glucosaminoglucano

**GLUT:** Transportador de glucosa

**GWAS:** Estudios de Asociación de Genoma Completo

**MAF:** Frecuencia del alelo menos común

**NMDA:** N-metil-D-Asp

**NGS:** Secuenciación de Nueva Generación

**OLFM1:** Olfactomedina-1

**OMIM:** *Online Mendelian Inheritance in Man*

**PCR:** Reacción en Cadena de la Polimerasa

**pb:** Pares de bases

**SAM:** Alineamiento/Mapa de Secuencia

**SNP:** Polimorfismo de un solo nucleótido

**SLC:** Familia de transportadores de soluto

**VCF:** *Variant Call Format*



# 1. INTRODUCCIÓN

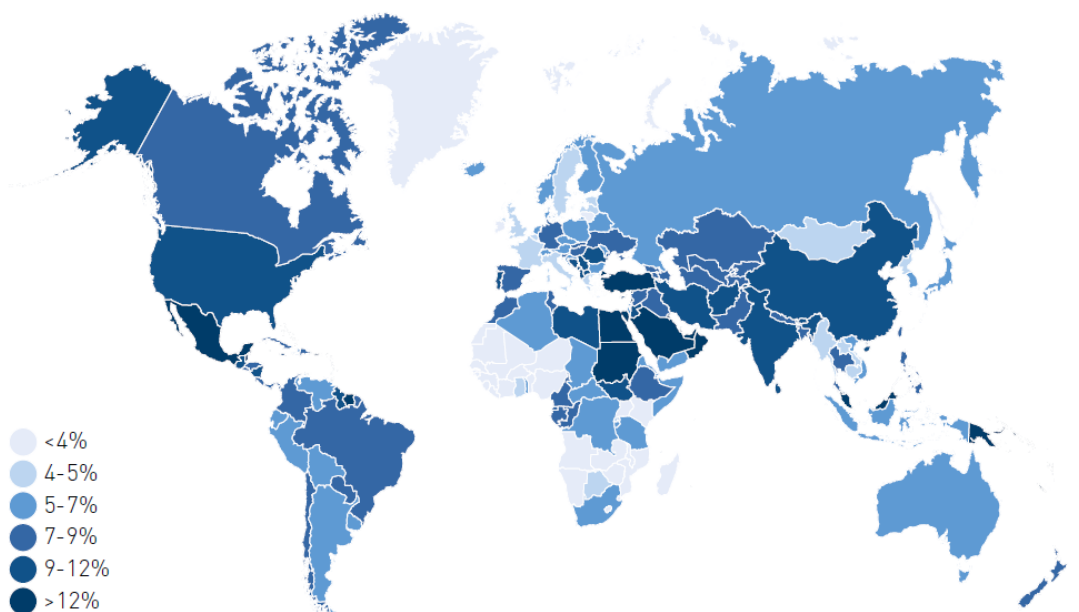
## 1.1. Epidemiología y etiología de la diabetes

La *diabetes mellitus* (DM) es un grupo de desórdenes metabólicos caracterizados por deficiencias en la secreción de insulina, problemas para que esta hormona ejerza su acción o una combinación de ambas, que resultan en hiperglucemia, ya sea en ayuno o estado postprandial. Se trata de una enfermedad multisistémica que conlleva una alta morbilidad, dado que se asocia con daño, disfunción y fallo en órganos y tejidos que incluyen la retina, riñones, nervios, corazón y vasos sanguíneos. También conlleva una alta mortalidad ya que se ve incrementada la propensión a enfermedades cardiovasculares, infarto de miocardio, accidentes cerebrovasculares y fallo renal (Crawford *et al.*, 2017, Alam *et al.*, 2014).

Según la *International Diabetes Federation* (IDF), (8ª edición del Diabetes Atlas de la FID, 2017), se estima una prevalencia de la diabetes de más de 425 millones de personas en todo el mundo (más de 58 millones en Europa), afectando a una tercera parte de personas mayores de 65 años. En las últimas décadas, se ha producido un aumento importante en la prevalencia de DM2 en la mayoría de las regiones del mundo, habiéndose duplicado entre 1980 y 2014, de forma más pronunciada en hombres que mujeres (Kolb *et al.*, 2017). Se calcula una prevalencia de la diabetes del 8,8% en adultos de entre 20 a 79 años. Alrededor del 79% vive en países de ingresos bajos y medios. Las regiones más afectadas por la enfermedad se encuentran en el Sudeste Asiático (con 121 millones de personas afectas en China) y en el Pacífico Occidental (Fig. 1). La incidencia ha aumentado en los últimos años y se espera que para el año 2045 alcance 693 millones de personas afectas, presentando un mayor incremento las regiones de África, Oriente Medio, Norte de África y el Sudeste Asiático. En este incremento general están implicados factores como el envejecimiento de la población y el aumento de obesidad y estilo de vida sedentario. Sin embargo, la diabetes ha comenzado a disminuir en algunos países desarrollados. El presupuesto sanitario empleado en diabetes en todo el mundo es elevado, significando aproximadamente un 9% del total en Europa y un 17% en la región del Medio Oriente y Norte de África (Fig. 1).

En España, la *International Diabetes Federation* estimó que en 2017 unos 3,6 millones de adultos entre 20 y 79 años padecían diabetes. Además, se situó como el décimo país con mayor número de personas mayores de 65 años diabéticas (2,2 millones). El estudio Di@bet.es es el primero realizado en España destinado a evaluar la prevalencia de *diabetes mellitus* tipo 2 (DM2) a nivel nacional. Se llevó a cabo entre 2008 y 2010 en una muestra de 5.072 adultos

españoles mayores de 18 años, seleccionados de forma aleatoria del Sistema Nacional de Salud, de 110 Centros de Atención Primaria. En 2011 se publicaron los resultados, que indicaban que el 13,8% de los españoles mayores de 18 años padece diabetes tipo 2. El 43% del total de diagnosticados desconocían que padecían la enfermedad (Soriguer *et al.*, 2012). Los últimos datos presentados en 2018 en el marco del XXIX Congreso Nacional de la Sociedad Española de Diabetes indican que cada año se diagnostican 386.003 nuevos casos de diabetes en España (11,58 por cada 1.000 personas).



**Figura 1.** Estimación de la prevalencia de diabetes en 2017 en adultos (20 a 79 años) (Diabetes Atlas de la FID - 8ª edición).

## 1.2. Diagnóstico y pronóstico de la diabetes

El diagnóstico de la diabetes se lleva a cabo mediante tres pruebas principales que miden la glucosa plasmática tras 8 h en ayunas, la glucosa en plasma 2h después de la ingestión de 75g de glucosa (glucosa postprandial) y la hemoglobina glicosilada (HbA1c). Los criterios para el diagnóstico de la diabetes son los siguientes (American Diabetes Association, 2018, Crawford *et al.*, 2017) (Tabla 1):

**Tabla 1. Valores para el diagnóstico de diabetes según la American Diabetes Association**

Prueba	Valores normales	Prediabetes	Diabetes
Glucosa en ayunas	<99 mg/dL	100-125 mg/dL	≥126 mg/dL
Glucosa postprandial	≤139 mg/dL	140-199 mg/dL	≥200 mg/dL
HbA1C	≤5,6 %	5,7-6,4 %	≥6,5 %
Aleatorio	-	-	≥200 mg/dL + síntomas

La prediabetes conlleva un aumento de riesgo de desarrollar diabetes o enfermedades cardiovasculares (CVD). Está asociada a obesidad, dislipemia con triglicéridos elevados, colesterol HDL bajo e hipertensión.

### **1.3. Clasificación y síntomas de la diabetes**

Según la American Diabetes Association (ADA), los pacientes afectados de diabetes se pueden clasificar en los siguientes grupos principales:

- Diabetes tipo 1 (DM1): Tiene origen autoinmune y conlleva deficiencia de insulina debido a la destrucción de células  $\beta$ .
- Diabetes tipo 2 (DM2): Está normalmente ligada a resistencia a insulina, produciéndose la pérdida progresiva de secreción de insulina de las células  $\beta$ . Representa un 90-95% de la población diabética. Se conoce también como diabetes no insulino dependiente o diabetes del adulto.
- Diabetes gestacional: Aparece en el segundo o tercer trimestre de embarazo e implica riesgo de desarrollar DM2, al darse niveles altos de glucosa en sangre en embarazadas que no han padecido diabetes previamente.

Existen otros tipos de diabetes, entre los que se incluyen las diabetes monogénicas. Entre ellas se encuentra la diabetes neonatal, que aparece en el primer mes de vida y puede ser transitoria o permanente; la DM tipo MODY (diabetes de la edad madura que se presenta en el joven), o la diabetes tipo LADA (diabetes autoinmune latente en adultos), que comprende a los pacientes diabéticos que no suelen requerir tratamiento con insulina durante al menos 6 meses después del diagnóstico, y que presentan marcadores inmunes séricos detectables de DM1. Las diabetes monogénicas representan menos del 5% de los casos y conllevan una disfunción de las células  $\beta$ . También se puede desarrollar DM tras un trasplante o puede estar asociada a la enfermedad de fibrosis quística (ADA, Buzzetti *et al.*, 2017).

## 1.4. Diabetes Mellitus tipo 2

La *diabetes mellitus* tipo 2 es una epidemia global relacionada con cambios ambientales y el estilo de vida. Entre estos factores ambientales se incluyen dietas occidentales densas en energía, actividad física disminuida, obesidad (principalmente abdominal), exposición al ruido o polvo fino, sueño corto o perturbado, tabaquismo, estrés, depresión y un bajo nivel socioeconómico (Kolb *et al.*, 2017)

Los principales factores de riesgo de desarrollar DM2 según el Instituto Nacional de Diabetes y Enfermedades Digestivas y Renales (NIDDK), son los siguientes (Tabla 2):

Tabla 2. Factores de riesgo a DM2

Factores de riesgo de DM2	
Sobrepeso u obesidad	Antecedentes de infarto o ictus
≥ 45 años	Antecedentes de diabetes gestacional
Familiar de primer grado con DM2	Sedentarismo
Raza étnica de alto riesgo	Depresión
Presión sanguínea alta	Síndrome de Ovario Poliquístico
↓ Colesterol HDL o ↑ triglicéridos	<i>Acanthosis nigricans</i>

Los pacientes de DM2 a menudo presentan síndrome metabólico durante muchos años antes del inicio de la hiperglucemia y el diagnóstico. En este período, la obesidad y la resistencia a la insulina a menudo están presentes, asociadas con la hipertensión y la dislipidemia, existiendo un mayor riesgo cardiovascular (Schofield *et al.*, 2016).

Los estudios epidemiológicos indican que los principales factores protectores de la diabetes comprenden las dietas basadas en alimentos vegetales y el trabajo muscular de intensidad moderada a alta. A pesar de que es más frecuente en adultos, la incidencia de DM2 está aumentando en jóvenes y niños (Kolb *et al.*, 2017, Dendup *et al.*, 2018).

La evolución de la DM2 suele ser lenta y normalmente se presenta sin los trastornos metabólicos agudos que aparecen en la diabetes tipo 1, por lo que resulta muy difícil determinar el momento exacto de su aparición. Los pacientes con DM2 suelen presentar resistencia periférica a la insulina y producción deficiente relativa de insulina; y los síntomas comprenden visión borrosa, boca seca y aumento de la sed, micción frecuente y abundante, entumecimiento y hormigueo de manos y pies, cansancio, infecciones fúngicas recurrentes en la piel y lentitud en la curación de heridas (8ª edición del Diabetes Atlas de la FID 2017).

Actualmente, el tratamiento de primera línea para tratar la DM2 es la metformina (biguanida) debido a su eficacia y seguridad elevada. Este fármaco actúa reduciendo la glucosa en plasma postprandial y basal, pero existen otros fármacos cuando está contraindicada o no es tolerada. Si no se producen mejoras se administrará doble, triple terapia, o terapia combinada inyectable (Upadhyay *et al.*, 2018; Trhasher *et al.*, 2017).

### 1.5. Fisiopatología de la DM2

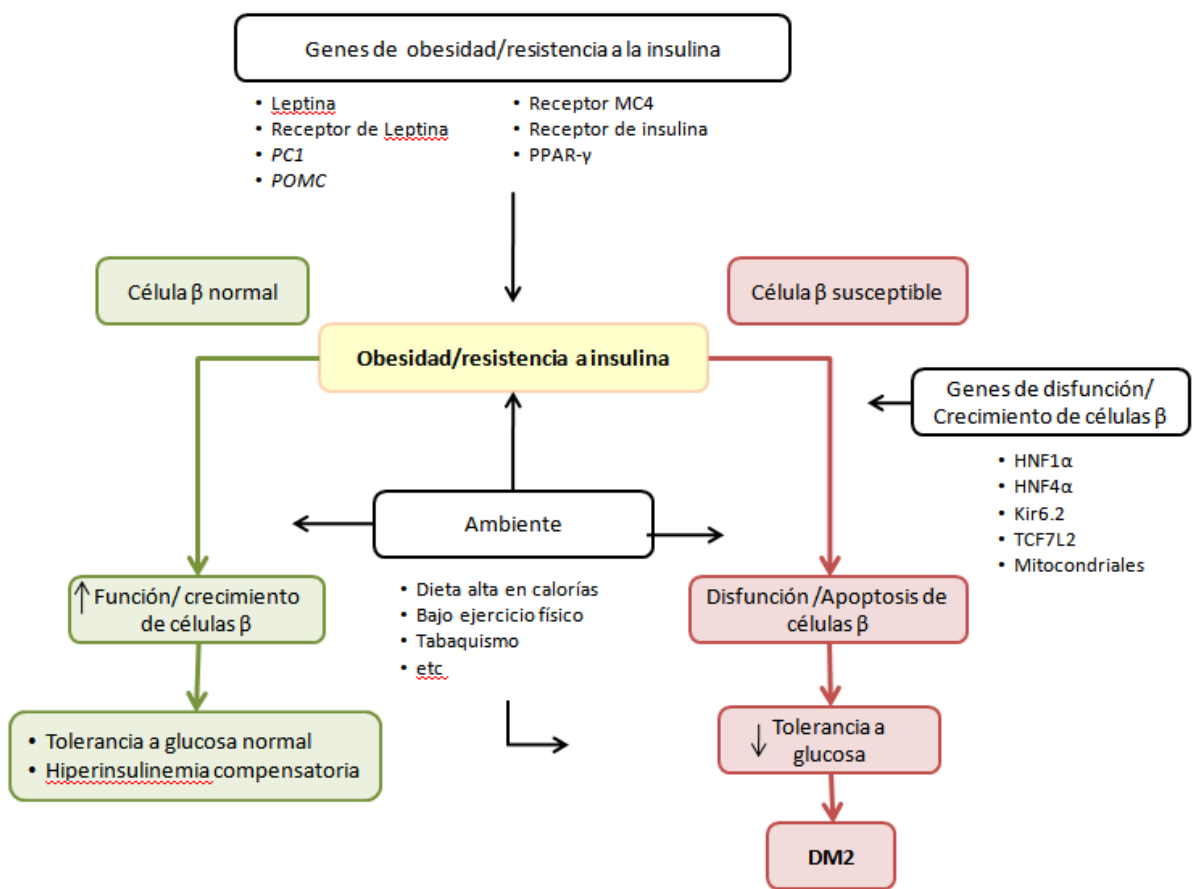


Figura 2. Interacción entre los genes y el medio ambiente en individuos que mantienen una tolerancia normal a la glucosa y aquellos que desarrollan diabetes tipo 2. Figura adaptada a Kahn *et al.*, 2006.



En individuos sanos la insulina y el glucagón ayudan a mantener constantes los niveles de glucemia en el cuerpo. La insulina es una hormona producida por las células  $\beta$  pancreáticas que tiene como función la internalización celular de la glucosa, mientras que el glucagón es producido por las células  $\alpha$  del páncreas y desencadena la liberación de glucosa al torrente sanguíneo (Kchouk *et al.*, 2017).

La obesidad y la resistencia a insulina son los principales factores implicados en el desarrollo de DM2 y se producen por la acción de ciertos genes implicados (*PPAR- $\gamma$* , *leptin...*) y factores ambientales (dieta alta en calorías, ejercicio físico disminuido...). Este estado de insulinoresistencia y obesidad conlleva un el aumento en la demanda de secreción en las células  $\beta$  (Fig.2).

En condiciones fisiológicas normales, las células  $\beta$  son capaces de producir una respuesta adaptativa al aumento de la demanda de insulina mediante cambios en su función y masa, manteniendo la tolerancia a la glucosa normal mediante una hiperinsulinemia compensatoria. Sin embargo, ya sea como consecuencia de ese aumento en la secreción o por factores genéticos, las células  $\beta$  pueden estar alteradas y presentar disfunción y disminución de su masa. Esto resulta en una alteración de la tolerancia a glucosa, con estados de hiperglucemia en ayunas, y el desarrollo de DM2. En el aumento de los niveles de glucosa en sangre postprandial y en ayunas también están implicadas la supresión incompleta de la producción de glucosa hepática y la internalización deficiente por parte de las células de hígado y músculo. El incremento de los niveles de glucosa, junto con la exposición crónica a niveles elevados de ácidos grasos libres (NEFAs), a su vez podría contribuir al avance de la enfermedad al afectar a las células  $\beta$  y a la sensibilidad a la insulina. De hecho, aunque en pacientes diabéticos las células  $\beta$  funcionan aproximadamente a un 25% o menos de su capacidad en el momento del diagnóstico, la pérdida de función es continua, lo que hace que la sintomatología de la enfermedad sea progresiva. El número de células secretoras también se ve reducido, aproximadamente a la mitad (Kahn *et al.*, 2006, Czech *et al.*, 2017). Sin embargo, existen grupos con mayor riesgo de desarrollar diabetes en los que se produce una disfunción de las células  $\beta$  mucho antes de que se considere que tienen una tolerancia reducida a la glucosa. Entre estos grupos de riesgo encontramos mujeres con antecedentes de diabetes gestacional, síndrome de ovario poliquístico, personas con intolerancia a la glucosa o parientes de primer grado de individuos con diabetes tipo 2 (Kahn *et al.*, 2006).

Entre los mecanismos de inicio de resistencia a insulina se han propuesto recientemente la activación del factor de transcripción FOXO1 en el hígado y la alteración de la translocación del

transportador de glucosa GLUT4 a la membrana superficial en el músculo esquelético. FOXO1 aumenta la expresión de enzimas de la gluconeogénesis; por lo tanto, el aumento de su actividad resulta en una producción incrementada de glucosa en el hígado. En condiciones normales, la insulina activa la fosforilación de este factor de transcripción a través de la proteína quinasa Akt, reteniéndolo inactivo en el citoplasma. Por otro lado, la disminución en los niveles de GLUT4 en la membrana superficial del músculo, conlleva una menor absorción de glucosa circulante (Czech *et al.*, 2017).

En individuos obesos, se produce la inflamación persistente de bajo grado que podría contribuir a la progresión a la resistencia a la insulina y la diabetes tipo 2. Las citoquinas proinflamatorias que se han visto más relacionadas con resistencia a insulina son IL-1 $\beta$ , TNF $\alpha$  y IL-6 (Pirola *et al.*, 2017).

## 1.6. Tecnología NGS

En 1977, fueron publicados los primeros métodos de secuenciación de ADN, describiéndose las tecnologías de secuenciación de Primera Generación, Sanger y Maxam-Gilbert. En concreto, el método desarrollado por Frederick Sanger ha sido el más utilizado durante aproximadamente 40 años, hasta la llegada de los métodos de Secuenciación de Segunda Generación (NGS) en 2005. Estas nuevas tecnologías han supuesto una revolución en el ámbito de la secuenciación al solucionar los inconvenientes de la primera generación, dado que permiten generar millones de lecturas cortas de forma paralela, así como la detección sin el uso de electroforesis. Esto conlleva un gran aumento de velocidad y un abaratamiento del proceso (Kchouk *et al.*, 2017).

Los métodos NGS de segunda generación se dividen en dos enfoques principales: secuenciación por ligamiento (SBL) y secuenciación por síntesis (SBS), y se clasifican en cuatro plataformas principales de secuenciación: Roche/454 (2005), Illumina/Solexa (2006), ABI/SOLiD (2007) e Ion Torrent (2010). Estas tecnologías generalmente requieren una amplificación previa por Reacción en Cadena de la Polimerasa (PCR), que resulta en un incremento del tiempo y los costes del proceso.

Las tecnologías de Secuenciación de Tercera Generación permiten la secuenciación de moléculas sin proceso de amplificación (Heather *et al.*, 2016), abaratando los costes; y la obtención de lecturas largas, que facilitan el posterior ensamblado. Los principales métodos son la secuenciación a tiempo real de una sola molécula o SMRT (*Single Molecule Real Time*) mediante fluorescencia, utilizado por las plataformas Pacific Biosciences y Oxford Nanopore

Technologies, y la secuenciación por síntesis, utilizada por las plataformas Illumina y10x Genomics (Goodwin *et al.*, 2016).

### 1.7. Estudios genéticos en DM2

En el desarrollo de la *diabetes mellitus* tipo 2 están implicados tanto factores ambientales como genéticos, tratándose de una enfermedad multifactorial. Actualmente, la genética implicada en DM2 es poco conocida. De hecho, los últimos estudios de la heredabilidad de la DM2, estiman que puede variar desde un 25% a un 80% en diferentes poblaciones. En la población caucásica se ha estimado un riesgo aproximado del 40% de heredar DM2 siendo uno de los padres afecto (3 veces más que en la población general), y aproximadamente del 70% si se trata de ambos progenitores. Sin embargo, en la población indígena de indios Pima se observó una heredabilidad cercana al 100%. Por otra parte, en un estudio se comparó la heredabilidad entre gemelos monocigóticos y dicigóticos, resultando ser de aproximadamente un 70% y 25%, respectivamente (Poulsen *et al.*, 1999; Meigs *et al.*, 2000, Prasad *et al.*, 2015). En el caso de gemelos monocigóticos se ha llegado a estimar una heredabilidad del 96% para DM2 o tolerancia alterada a la glucosa a los 15 años de seguimiento (Medici *et al.*, 1999). Estos estudios muestran la importancia del componente genético en el desarrollo de la enfermedad.

Las primeras estrategias utilizadas en el mapeo genético de la DM2 incluían estudios de ligamiento o estudios de asociación de genes candidatos, que presentaban inconvenientes relacionados con el diseño de experimentos y tamaño de la muestra. En 2006 se identificó una variante de riesgo en el locus *TCF7L2* mediante un estudio de ligamiento en población islandesa, que sigue considerándose una de las más relacionadas con DM2. Desde 2007, la aparición de nuevas técnicas, como el genotipado basado en *array*, junto con un incremento del tamaño de muestras por un aumento de la colaboración internacional, permitió avanzar en la asociación de *loci* genómicos con enfermedades complejas como la DM2 y desórdenes asociados (Grarup *et al.*, 2014).

Con el desarrollo de las nuevas tecnologías se han encontrado principalmente variantes comunes (frecuencia del alelo menos común o MAF >5%) asociadas a la enfermedad, que presentan un efecto pequeño sobre el fenotipo. Los estudios de Asociación de Genoma Completo (GWAS) han identificado más de 90 *loci* de susceptibilidad para DM2 y más de 50 *loci* para parámetros relacionados (Tabla 3). Para la mayoría de los *loci*, la variante causal y el gen son desconocidos. Entre los *loci* identificados se encuentran *PPARG* y *KCNJ11*, que codifican

proteínas que son diana de los fármacos antidiabéticos tiazolidinedionas y sulfonilureas. Por otra parte, aunque la DM2 y la obesidad están muy relacionadas, solo se han identificado unos 5 *loci* compartidos (Stančáková *et al.*, 2016).

**Tabla 3. Algunos genes relacionados con diabetes tipo 2 seleccionadas en Estudios de Asociación de Genoma Completo.**

<b>Locus</b>	<b>Implicación en DM2</b>
<i>PPARG</i>	Desregulación del metabolismo de grasas
<i>KCNJ11</i>	Asociado a hiperinsulinemia congénita
<i>TCF7L2</i>	Disminución de la sensibilidad de las células $\beta$ , llevando a procesamiento alterado y secreción disminuida de insulina
<i>MTNR1B, FADS1, DGKB, GCK</i>	Disminución de la liberación temprana de insulina estimulada por glucosa
<i>FSADS1</i>	Metabolismo alterado de ácidos grasos insaturados
<i>FTO, IGF2BP2</i>	Aumento de adiposidad y resistencia a insulina
<i>HHEX</i>	Alteración en el control del desarrollo de las estructuras pancreáticas, incluidas las células del islote beta
<i>SLC30A8</i>	Transporte alterado de zinc a las células del islote beta (producción y secreción de insulina)
<i>WFS1</i>	Supervivencia y función alteradas de las células del islote beta

Datos obtenidos de Khardori *et al.*, 2018

A pesar del alto número de *loci* identificados, el impacto individual de cada variante estudiada es bajo y los *loci* de susceptibilidad a DM2 identificados explicarían aproximadamente un 10% de la heredabilidad de la DM2. A esta discrepancia entre la heredabilidad preestimada en estudios familiares y los resultados observados en los GWAS, se le ha denominado el problema de la “heredabilidad perdida”. Algunas de las hipótesis que podrían explicar este fenómeno son las interacciones genético-ambientales y gen-gen, una heredabilidad sobreestimada en estudios previos al no tener en cuenta efectos epigenéticos e interacciones genéticas complejas, o las variantes de baja frecuencia o raras con efecto relevante sobre el fenotipo (Lohmueller *et al.*, 2013, Stančáková *et al.*, 2016).

Hasta la actualidad, en los estudios de asociación de casos y controles en enfermedades complejas como la DM2 ha predominado la hipótesis enfermedad común-variante común. La mayoría de estos estudios se centran en la búsqueda de polimorfismos de un solo nucleótido (SNP) frecuentes en la población (MAF  $\geq 5\%$ ), al considerar que pueden contribuir

significativamente a la prevalencia de las enfermedades complejas, incluso si su efecto sobre el riesgo de la enfermedad es modesto.

Una de las causas por las que la hipótesis opuesta, enfermedad común-variante rara, no ha sido muy explorada en los últimos años es por la dificultad a la hora de explicar la alta prevalencia de enfermedades complejas siendo la frecuencia de cada SNP tan baja (MAF  $\leq$ 5%). Una posible explicación es el gran número de SNPs raros en conjunto en el genoma, por lo que muchos de ellos pueden tener un efecto en estas enfermedades y compensar su MAF reducido. En este sentido, la mayoría de SNPs del genoma humano son de frecuencia menor al 5%. Otro de los aspectos a considerar es el efecto funcional de estas variantes. El efecto de los SNPs poco frecuentes puede ser deletéreo, afectando la función de un gen y favoreciendo la selección poblacional. Sin embargo, el efecto también puede resultar beneficioso en relación a algunas enfermedades o darse el efecto perjudicial tras la edad de reproducción, por lo que se favorecería la permanencia de esas SNPs en la población. Asimismo, también podría tratarse de variantes recientes en la población (Gorlov *et al.*, 2011).

Uno de los factores que han dificultado los estudios de variantes raras en enfermedades complejas es el tamaño de muestra. Debido a las limitaciones tecnológicas muchos de los estudios se llevaban a cabo con un tamaño reducido de individuos ( $\leq$ 1000), por lo que el poder estadístico no era suficiente para la búsqueda de SNP raros. Gracias a la aparición de las tecnologías de nueva generación (NGS) y la asequibilidad de la secuenciación de exoma, se pueden diseñar estudios de SNPs poco frecuentes con un tamaño muestral suficiente, aunque hasta ahora, se han llevado a cabo pocos estudios basados en la búsqueda de variantes raras en DM2. En una investigación de 2008 se compararon las MAF de SNPs sinónimas con la de SNPs no sinónimas clasificadas previamente como probablemente dañinas, posiblemente dañinas y benignas, concluyendo que en estudios con grandes tamaños de muestra es más adecuada la búsqueda de SNPs de baja frecuencia (Gorlov *et al.*, 2008).

## 2. ANTECEDENTES Y OBJETIVOS

Previamente a este estudio se llevó a cabo la búsqueda de variantes genéticas poco frecuentes y raras en DM2 mediante secuenciación de exoma en 200 pacientes con DM2 y 200 controles sanos de población española (Galán, 2015), dentro de un proyecto coordinado por varios grupos de investigación españoles. Se identificó un gran número de variantes genéticas con posible implicación en el desarrollo de DM2 o en la protección frente a la misma. En base a estas variantes y a la información de bases de datos, se seleccionaron 10 genes para secuenciación dirigida en el presente estudio (Tablas 4 y 5). Las variantes seleccionadas tenían un alto efecto funcional en el fenotipo, siendo 9 de ellas variantes de ganancia de codón de *STOP* y la restante una variante en una región aceptora de *splicing*. 6 de estas variantes fueron clasificadas como protectoras (genes *ACPL2*, *ARHGEF38*, *CCT8L2*, *CIAPIN1*, *DDO*, *OLFML1*), dado que estaban presentes solo en controles sanos; y 4 serían de riesgo (genes *AFMID*, *CACNG5*, *SLC12A4*, *SLC15A3*), estando sobrerrepresentadas en individuos diabéticos.

**Tabla 4. Variantes encontradas en los genes seleccionados a estudio en el presente trabajo**

Posición Cr.	Cambio	Gen	Tipo de Variación	SNP	Efecto
3:140993422	T/G	ACPL2	Ganancia codón STOP	-	Protector
17:76200907	A/G	AFMID	Región aceptora de <i>splicing</i>	rs138105144	Riesgo
4:106474097	G/T	ARHGEF38	Ganancia codón STOP	rs148016102, COSM1671189	Protector
17:64881037	C/T	CACNG5	Ganancia codón STOP	rs62621240	Riesgo
22:17072347	C/T	CCT8L2	Ganancia codón STOP	rs139948519	Protector
16:57464236	G/A	CIAPIN1	Ganancia codón STOP	rs170358	Protector
6:110736730	C/T	DDO	Ganancia codón STOP	rs141023778	Protector
11:7509386	C/T	OLFML1	Ganancia codón STOP	rs141351486, COSM1322159	Protector
16:67980235	G/T	SLC12A4	Ganancia codón STOP	-	Riesgo
11:60714017	G/A	SLC15A3	Ganancia codón STOP	rs116597710	Riesgo

Los genes a estudio están relacionados con diferentes mecanismos moleculares y codifican para proteínas con funciones como la desfosforilación y desaminación, el transporte, la regulación de receptores, función de chaperona o efecto antiapoptótico (Tabla 5).

**Tabla 5. Características de los genes a estudio en el presente trabajo**

<b>Gen</b>	<b>Función</b>	<b>Posible relación con diabetes</b>
<i>ACPL2</i> ( <i>PXYLP1</i> )	2-O-desfosforilación de la xilosa en la unión glucosaminoglucano-proteína de los proteoglicanos	Alteración estructural y de concentración de proteoglicanos en condiciones de diabetes.
<i>AFMID</i>	Cataliza el segundo paso de la conversión de Trp a ácido nicotínico, NAD(H) y NADP(H),	El catabolismo anormal de triptófano se ha relacionado en estudios observacionales con DM2. Ratones <i>knockout</i> para <i>AFMID</i> y <i>TK</i> muestran tolerancia a la glucosa alterada
<i>ARHGEF38</i>	Activación de GTPasas y regulación de la transducción de señal de proteína Rho.	Implicado en la señalización del receptor de neurotrofina p75, que regula la homeostasis de la glucosa y la sensibilidad a la insulina
<i>CACNG5</i>	Proteína reguladora del receptor transmembrana de glutamato AMPA tipo II (TARP).	Varios estudios han indicado que la diabetes induce cambios en los receptores de glutamato.
<i>CCT8L-2</i>	Posible chaperona molecular; forma parte del complejo antiapoptótico TRiC/CCT.	El complejo TRiC/CCT defiende contra el estrés y apoptosis causado por agregados de polipéptidos. Posible implicación en la supervivencia de células $\beta$
<i>CIAPIN1</i>	Efectos antiapoptóticos en la célula inducidos por citoquinas.	Las citoquinas pueden inducir apoptosis en células $\beta$ . CIAPIN1 está altamente expresado en páncreas, pudiendo estar implicado en la supervivencia de células $\beta$
<i>DDO</i>	Flavoproteína peroxisomal. Cataliza la desaminación oxidativa de D-aspartato y D-aspartato de N-metilo.	Los receptores de glutamato N-metil-D-aspartato (NMDA) se expresan en todo el riñón, y su abundancia aumenta en la diabetes
<i>OLFML1</i>	Regulación de la proliferación celular al aumentar el porcentaje de células en fase S.	Su expresión se ve aumentada al inducir un ambiente tolerogénico en el páncreas
<i>SLC12A4</i> ( <i>KCC1</i> )	Cotransportador Potasio/Cloro.	Los islotes pancreáticos de rata expresan varios extrusores Cl <sup>-</sup> incluyendo <i>SLC12A4</i> ( <i>KCC1</i> ), <i>SLC12A6</i> ( <i>KCC3</i> ) y <i>SLC12A7</i> ( <i>KCC4</i> ).
<i>SLC15A3</i> ( <i>PHT2</i> )	Transporte de oligopéptidos.	<i>SLC15A3</i> podría estar relacionado con el sistema inmunitario.

El objetivo de este estudio es el estudio de estos genes, así como la identificación de nuevas variantes y la validación de las variantes previamente encontradas mediante el estudio de los genes indicados por secuenciación masiva de amplicones.

### 3. MATERIALES Y MÉTODOS

#### 3.1. Selección de la población a estudiar

En el estudio se incluyeron 1415 muestras de una población procedente del Biobanco de Barcelona (CIBERDEM). Se seleccionó un grupo de casos compuesto por 586 individuos que padecían DM2, y otro de controles, que comprendía 829 individuos no afectados de DM2. Los controles presentaban glucemia normal, mientras que los casos diabéticos presentaban diagnóstico para esta enfermedad.

Para los casos se utilizó como criterio de inclusión una edad comprendida entre 30 y 65 años. Para los controles, se seleccionaron pacientes dentro de un rango de edad de 50 a 73 años, dado que en este rango es más probable que no hubiera un posterior desarrollo de la enfermedad a lo largo de la vida de los individuos que en personas más jóvenes. La edad media de ambos grupos fue similar y el grupo de diabéticos presentó un índice de masa corporal (IMC) mayor y un mayor porcentaje de individuos obesos, tal y como era esperable (Tabla 6).

**Tabla 6. Valores promedio de edad, IMC y obesidad de los individuos seleccionados a estudio**

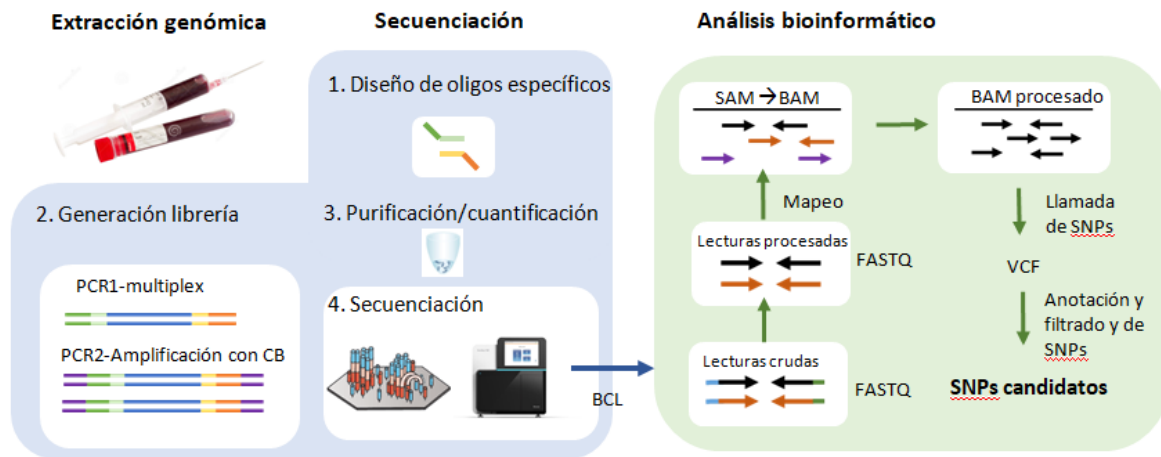
	Casos	Controles
Edad (años) <sup>a</sup>	55,28±7,91	55,91±4,55
IMC (kg/m <sup>2</sup> ) <sup>a</sup>	31,46±6,05	26,64±3,68
Obesidad (%)	55,05	20,6

IMC: Índice de Masa Corporal. <sup>a</sup>Datos expresados como media ± desviación estándar

#### 3.2. Diseño del estudio

En este proyecto se estudiaron variantes génicas raras en exoma. Dado que el exoma representa únicamente un 2% del ADN, este tipo de estudios permite acelerar, simplificar y abaratar el proceso de búsqueda de variantes. Las variantes no sinónimas en estas zonas suelen tener efecto sobre el fenotipo, a diferencia de las SNPs en regiones no codificantes, que suelen presentar efectos más neutros sobre este (Gaulton *et al.*, 2017). Utilizando como base la hipótesis enfermedad común-variante rara, se filtraron las SNPs con una MAF≤5%.





**Figura 3. Resumen del diseño del estudio.** Consta de 3 fases principales: extracción genómica, secuenciación y análisis bioinformático. CB: Código de Barras

El estudio de las variantes se llevó a cabo mediante secuenciación por amplicones (*target sequencing*), que permite el genotipado de cientos de muestras en paralelo. Se trata de una secuenciación dirigida, en la que las regiones génicas de interés fueron secuenciadas mediante el diseño de cebadores específicos. Una de las ventajas de este tipo de secuenciación es que permite obtener información más fácil de analizar, además de conseguir coberturas mayores y abaratar costes con respecto a otros tipos de estudio. Las principales limitaciones de esta técnica son el diseño de las sondas u oligos, que no es escalable a muchos genes y que no permite la identificación de nuevos genes.

El estudio constó de 3 fases principales (Fig. 3). En primer lugar, se llevó a cabo la extracción del ADN de los pacientes y controles a partir de sangre en EDTA. A continuación se generó la librería mediante las PCRs diseñadas: el primer paso fue la amplificación con oligos específicos de las secuencias genómicas de interés y las colas para la segunda amplificación; el segundo consistió en incorporar mediante estas colas los códigos de barras o *barcodes* y el resto de secuencias necesarias para las reacciones de NGS. Tras la purificación y cuantificación del *pool* de muestras amplificadas se llevó a cabo la secuenciación. Los datos obtenidos se analizaron mediante *softwares* y *scripts* bioinformáticos generados en la Unidad de Genómica y Diagnóstico Genético (UGDG) de INCLIVA. Las lecturas crudas fueron procesadas y mapeadas y se llevó a cabo la búsqueda de variantes, el filtrado, el análisis estadístico y la anotación para seleccionar los SNPs candidatos.

### 3.3. Obtención de muestras

El estudio fue aprobado previamente por el comité ético del Hospital Clínico de Valencia y de su Fundación. El ADN para las pruebas de validación fue extraído mediante el Kit de purificación de ADN Maxwell® (Promega®, USA) en el Hospital Clínico de Valencia. El ADN de las muestras se obtuvo del Biobanco de Barcelona.

### 3.4. Preparación de la librería

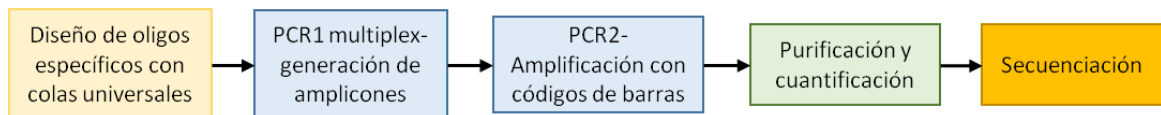


Figura 4. Flujo de trabajo para la preparación de la librería a secuenciar.

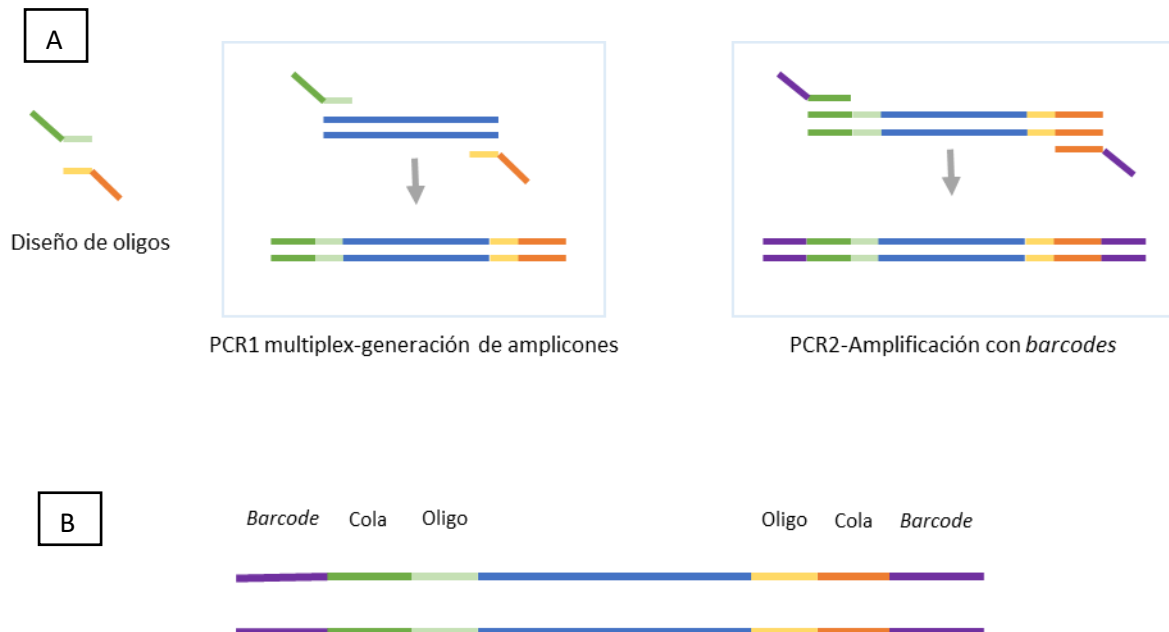


Figura 5. Preparación de la librería. A. Esquema de las amplificaciones; B. Estructura general de los amplicones

En primer lugar, se llevó a cabo el diseño *in silico* de los oligos o cebadores específicos 5' y 3' de las secuencias genómicas de interés con las colas universales añadidas para permitir la posterior unión de los *barcodes*. A continuación, se generaron los amplicones mediante PCR (PCR1) y se añadieron los *barcodes* para la identificación de las muestras mediante una segunda PCR anidada (PCR2) (Figs 4 y 5).

### 3.4.1. Diseño de *primers*

Se utilizó el genoma de referencia hg19/GRCh37 y las secuencias génicas y transcritos de las bases de datos COSMIC y Ensembl. El diseño de cebadores se llevó a cabo mediante el programa JPCR y el programa *online* Primer3.

Los cebadores se diseñaron en dos multiplexes (multiplex A y B), es decir, para dos reacciones de PCR independientes, evitando así fragmentos cortos resultado de la interacción entre oligos 5' y 3' de fragmentos solapantes (Fig. 6). Una vez realizada la primera PCR se juntaron los dos *pools* y se realizó la segunda PCR.

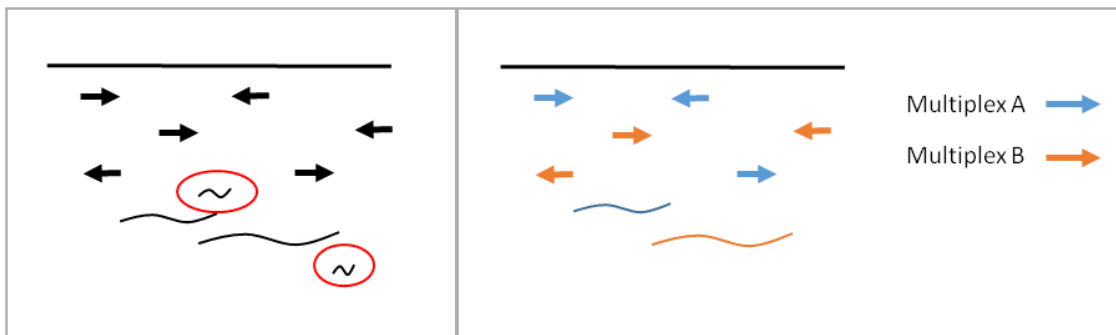


Figura 6. Representación de las interacciones entre los dos multiplexes de cebadores

### 3.4.2. PCR1 multiplex-generación de amplicones

Tras el diseño de los cebadores en dos multiplexes diferentes (multiplex A y B), se llevaron a cabo dos reacciones de PCR1 paralelas utilizando las mismas condiciones.

Las reacciones se realizaron mediante enzimas comerciales y en la puesta a punto de la reacción de amplificación se añadieron diferentes reactivos a distintas concentraciones y se llevaron a cabo pruebas variando la concentración de los *pools*.

Para el proceso de amplificación se utilizaron termocicladores de placas 384 Veriti® (AppliedBiosystems®, California). La PCR se inició a una temperatura de 95°C durante 20 minutos para permitir la activación de la enzima polimerasa. Posteriormente se desnaturalizaron las cadenas de ADN a 98°C, permitiendo así la unión de los cebadores. Para este alineamiento se utilizó una temperatura de 55°C, seguida de 60°C, evitando así problemas relacionados con las diferencias en la temperatura de hibridación de los cebadores; y la temperatura seleccionada para la extensión de la polimerasa fue de 72°C. Los procesos de

desnaturalización, alineamiento y extensión se repitieron 25 veces. La extensión final fue de 5 min a 72°C (Tabla 7).

**Tabla 7. Fases y condiciones utilizadas en la PCR1.**

Programa de PCR	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	20 min	
Desnaturalización	98 °C	20 seg	25x
Hibridación	55°C	30 seg	
	60°C	30 seg	
Extensión	72 °C	60 seg	
Extensión final	72 °C	5 min	
Conservación	15 °C	∞	

PCR: Reacción en Cadena de la Polimerasa

### 3.4.3. PCR2- Amplificación con códigos de barras-*barcodes*

En la PCR anidada se añadieron los *barcodes* mediante la hibridación con las colas universales a las muestras amplificadas, juntando en una misma reacción las amplificaciones con multiplex A y B. Estos *barcodes* son secuencias cortas que permiten la identificación de cada muestra en el proceso de análisis de datos tras la secuenciación. Mediante este método, se puede secuenciar de forma simultánea un alto número de muestras, disminuyendo los costes y simplificando el análisis. Los oligos de la PCR2 incluyeron además otras secuencias necesarias para el proceso de NGS.

Se utilizaron las siguientes condiciones (Tabla 8):

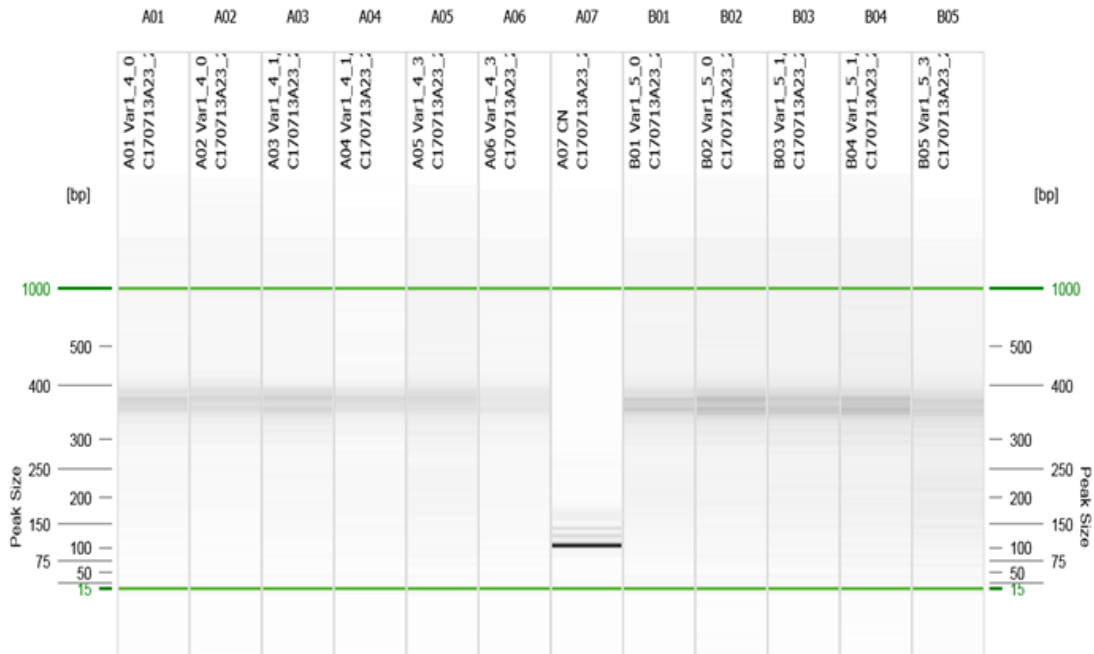
**Tabla 8. Fases y condiciones utilizadas en la PCR2.**

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95 °C	15 min	
Desnaturalización	98 °C	20 seg	35x
Hibridación	60°C	20 seg	
Extensión	72 °C	2 min	
Extensión final	72 °C	5 min	
Conservación	15 °C	∞	

PCR: Reacción en Cadena de la Polimerasa

### 3.4.4. Electroforesis capilar

Para la visualización de los fragmentos amplificados se utilizó el sistema de electroforesis capilar avanzado QIAxcel (QIAxcel DNA Screening Kit, Qiagen®, Alemania), que permite la automatización de la electroforesis capilar de forma sensible y con una capacidad de hasta 96 muestras por ciclo (Fig. 7).



**Figura 7. Ejemplo de electroforesis capilar utilizando el sistema QIAxcel.** El sistema QIAxcel se basa en una matriz de diodos emisores de luz y colectores microópticos adheridos a los capilares dentro de cartuchos de gel por los que migran los fragmentos de ADN por tamaño al aplicar corriente. El software QIAxcelScreenGel, permite el análisis de resultados.

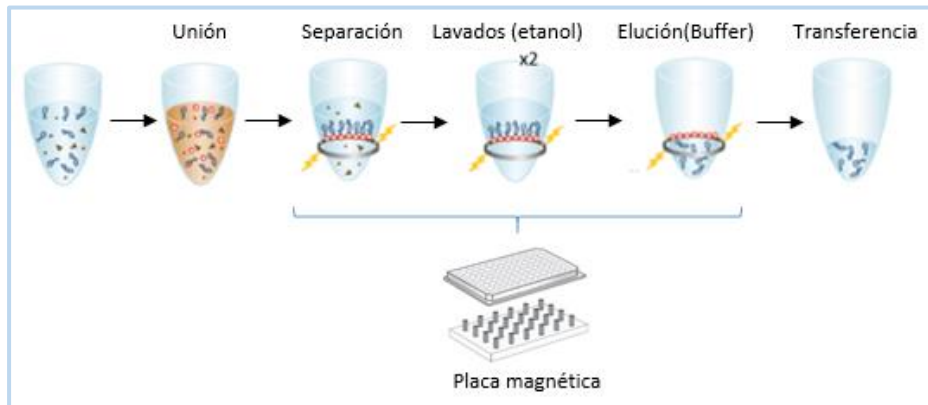
### 3.4.5. Generación del *pool*

Una vez amplificado el ADN y añadidos los códigos de barras en la PCR2, se procedió a la obtención del *pool* a secuenciar, añadiendo 3  $\mu$ l de cada muestra amplificada en un tubo Falcon. Este paso sólo es posible habiendo identificado cada muestra con su *barcode* para poder realizar el posterior análisis de la secuenciación.

### 3.4.6. Purificación

En el proceso de purificación del ADN amplificado se utilizó el kit Magsi-NGS PREP® (MagnaMedics Diagnostics B.V., Países Bajos). En este proceso se eliminaron productos inespecíficos, así como restos de cebadores y dímeros de estos.

Para la eliminación de tamaños inferiores a 100 pb se utilizó una concentración de muestra de 0.6x y se añadieron bolas magnéticas. Tras centrifugar a 1000 rpm, en la etapa de separación se utilizó una placa magnética y se eliminó el sobrenadante, obteniendo la muestra unida a las bolas magnéticas. A continuación se realizaron dos lavados con etanol al 70% en la placa magnética y se dejó evaporar. En la etapa de elución se añadió agua ultrapura (Gibco) para obtener el ADN purificado en el sobrenadante, y se centrifugó a 1000 rpm. Por último, sobre la placa magnética, se transfirió el sobrenadante que contenía la muestra purificada (Fig. 8).



**Figura 8.** Esquema del proceso de purificación del ADN amplificado mediante el kit Magsi-NGS PREP®. Imagen adaptada de BECKMAN COULTER®.

### 3.4.7. Cuantificación

La cuantificación de las muestras y los controles de las pruebas de la puesta a punto se llevó a cabo mediante el sistema QuantiFluor® dsDNA (Promega, USA), basado en la unión al ADN de un colorante fluorescente que presenta unos picos máximos de excitación y emisión de 504nm y 531nm, respectivamente. Este sistema permite la cuantificación de forma selectiva de pequeñas cantidades de ADN de doble cadena (ADNdc) en solución frente a otros tipos de ácidos nucleicos, y es lineal en un rango de 0.05-200 ng/ $\mu$ l de muestra de ADNdc.

Para preparar la curva estándar se utilizó como ADNdc el ADN estándar del fago lambda. Para cuantificar las muestras se añadió el colorante Quantifluor. Como tampón y blanco se utilizó TE. La fluorescencia utilizando el Sistema de Detección GloMax®-Multi+ (Promega®, USA).

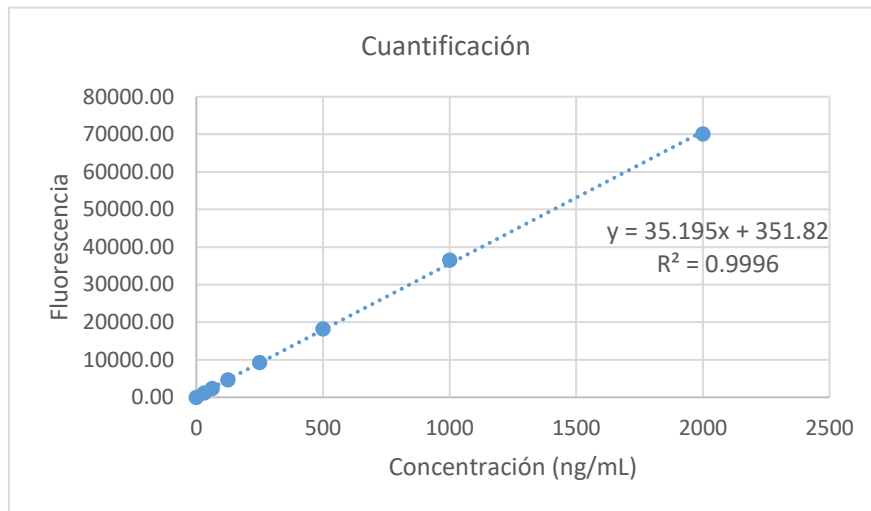


Figura 9. Recta de cuantificación del pool de ADN a secuenciar

Se obtuvo una fluorescencia final del *pool* de 23752,1 UFR (Unidades de Fluorescencia Relativa). La concentración se calculó extrapolando en la recta estándar (Fig. 9) y fue de 66,49 ng/ $\mu$ L. Finalmente, la muestra fue diluida en Tris/Tween a 20 nM, concentración previa requerida para enviar a secuenciar. Para ello se calculó el peso molecular promedio, teniendo en cuenta el peso molecular promedio por par de bases (660 g/mol) y el número promedio de bases por amplicón (396 pb).

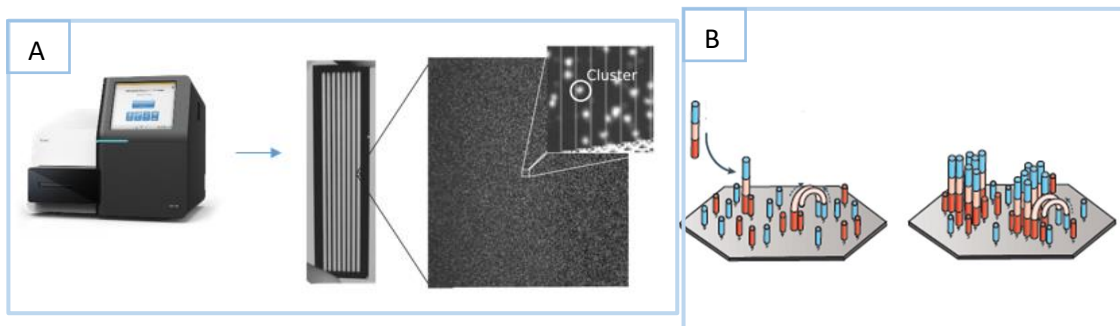
### 3.4.8. Secuenciación

Para las pruebas de validación se utilizó el secuenciador Miseq® (Illumina®, California) de la Unidad de Genómica Diagnóstico Genético de INCLIVA, mientras que las muestras fueron secuenciadas mediante el sistema HiSeq® (Illumina®, California) por la empresa MacroGen® (Corea del Sur), dado el elevado número de individuos a secuenciar y la gran cantidad de lecturas necesarias para obtener una cobertura por amplicón suficiente para identificar el genotipo de los pacientes. Ambos secuenciadores de Illumina se basan en la secuenciación por síntesis mediante amplificación en puente en fase sólida.

En el primer paso de la amplificación en puente en fase sólida las cadenas de ADN desnaturalizadas a secuenciar se ligan por su extremo 5' a cebadores complementarios a los

adaptadores, que se encuentran inmovilizados sobre una célula de flujo (Fig. 10A). Tras la amplificación, el extremo 3' libre puede interactuar con otros cebadores cercanos, formando una estructura de puente que permite la amplificación en el sentido inverso (Fig. 10B). En el proceso de generación de *cluster*, se llevan a cabo varias rondas de amplificación, formando unos 100-200 *clusters* clonales a partir de los cebadores inmovilizados. Por último, las cadenas reversas se eliminan, se hace un lavado del ADN libre y se bloquean los cebadores 3'.

En la secuenciación por síntesis, se utilizan dNTPs terminadores marcados con fluoróforos, que se añadirán en los millones de *clusters* en paralelo a la célula de flujo. La célula de flujo es un portaobjetos de vidrio que contiene 8 pequeños canales a través de los cuales se pueden bombear polimerasas, dNTP y tampones. Estos dNTPs presentan el grupo ribosa 3'-OH bloqueado de forma reversible, por lo que en cada ronda de detección se unirá un solo nucleótido a las cadenas de ADN. Para eliminar los nucleótidos no unidos y las polimerasas, se hace un lavado y se añade un tampón o *buffer* que permite la detección de la fluorescencia correspondiente a cada base. Los grupos bloqueantes de los dNTPs-OH se eliminan añadiendo un agente químico y se vuelve a realizar el mismo proceso hasta secuenciar toda la molécula. La señal fluorescente se analiza mediante un *software* informático. Para que las librerías hibriden de forma que haya separación física suficiente entre las cadenas de ADN, se utiliza una concentración muy baja, de 6 a 20 pM.



**Figura 10. Secuenciación Illumina.**A. Secuenciador Miseq y célula de flujo, en la que hibridarán las librerías y se formarán los *clusters*; B. Formación de puentes para la amplificación en puente en fase sólida previa a la secuenciación por síntesis. Imagen adaptada de Whiteford *et al.*, 2009 y Goodwin *et al.*, 2016

En el caso de las muestras utilizadas para la validación, secuenciadas mediante la plataforma Miseq, se utilizó el Kit MiseqReagent v2 (Illumina®, California), para el que se recomienda cargar en el secuenciador 600  $\mu$ L de la librería desnaturalizada a una concentración entre 6-20 pM, partiendo de una concentración mínima de librería de 4nM. En primer lugar, la librería fue



diluida en Tris (0.1%)/tween20 a 4 nM y a continuación fue desnaturalizada añadiendo NaOH y Tampón de Hibridación (HT1) hasta obtener 1 mL a 20 pM. Teniendo en cuenta la variabilidad, fue diluida de nuevo en HT1 a una concentración de 12 pM. La concentración de NaOH no superó el 1 mM en la solución final después de diluir con HT1, dado que concentraciones más altas de NaOH en la librería inhiben la hibridación a la celda de flujo y disminuyen la densidad del *cluster*. Como control de secuencia se utilizó el fago PhiX de Illumina. El fago fue desnaturalizado y diluido a 12,5pM del mismo modo que la librería de interés y se utilizó a una concentración baja del 1%, teniendo en cuenta la diversidad de la librería.

Mediante este sistema se obtuvieron lecturas de 2x150 pares de bases en cada dirección (*paired-end*) y el tiempo aproximado de secuenciación fue de 24 h. El secuenciador Miseq permite obtener un archivo de salida u *output* de 4,5-5,1 Gb.

### 3.5. Análisis bioinformático

El proceso de análisis bioinformático fue llevado a cabo por personal de la UGDG (INCLIVA) mediante *softwares* y *script* bioinformáticos generados en esta unidad (Fig. 11).

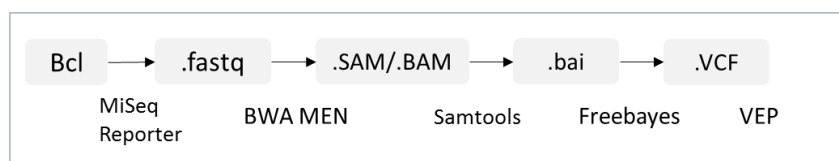


Figura 11. Esquema de los formatos (arriba) y *software* y herramientas (abajo) utilizados en el análisis informático

Los secuenciadores tipo Illumina generan archivos en formato .bcl. Una vez obtenido el archivo *output*, se lleva a cabo el demultiplexado utilizando el *software* "MiSeqReporter", en el que el archivo .bcl se procesa a .fastq. Para ello se utiliza un archivo "sample\_sheet" que indica cada muestra y su *barcode*, y se generan ficheros de lecturas para cada individuo comparando las lecturas de índices con los del archivo. Para cada muestra secuenciada se obtendrán dos archivos .fastq, al realizarse la secuenciación desde los dos extremos (*paired-endreads*).

FastQC permite hacer un control de calidad de las secuencias brutas y obtener una visión rápida de la calidad de la secuenciación. Además de permitir la detección de áreas con posibles problemas de secuenciación, FastQC proporciona gráficos y tablas resumen acerca de los datos obtenidos. Los archivos .fastq se muestran como información codificada y cada símbolo aporta datos acerca de cada lectura (instrumento, carrera, *flowcell*, índice...), así como información de su calidad (Phred Score) escrita en formato ASCII (Fig. 12).



Mediante la herramienta samtoolsview (Li et al., 2009) se obtuvo el formato .bam (Alineamiento/Mapa Binario), que es la forma binaria del formato .sam. El comando -samtoolsort se utiliza para ordenar el archivo BAM según su posición de alineación en la referencia y -samtoolsindex para crear un índice que permite realizar búsquedas rápidas en un BAM, generando un archivo .bam.bai.

El llamado de variantes se llevó a cabo con FreeBayes (Garrison *et al.*, 2012), un detector bayesiano de variantes génicas que permite la detección de pequeños polimorfismos, principalmente SNPs (polimorfismos de un solo nucleótido), indels (inserciones y deleciones), MNPs (polimorfismos multi-nucleótido), y variaciones complejas (eventos compuestos de inserción y sustitución). Usa alineamientos de lecturas cortas (archivos BAM) para cada individuo de la población, determinando la combinación más probable de genotipos para la población en cada posición en un genoma de referencia (FASTA). El archivo obtenido proporcionó las posiciones con más probabilidad de ser polimórficas en formato VCF (*Variant Call Format*).

VCF es un formato de archivo de texto que contiene líneas de metainformación, una línea de encabezado y líneas de datos que contienen información sobre una posición en el genoma. El formato también tiene la capacidad de contener información de genotipo en muestras para cada posición (Fig. 14).

```
##fileformat=VCFv4.1
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36.md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

Figura 14. Formato de archivo VCF.

En el proceso de anotación de variantes se utilizó VEP (VariantEffect Predictor), un *script* que predice el efecto funcional de las variantes genómicas a partir de las coordenadas de las variantes y de los cambios de nucleótidos.

### **3.6. Análisis estadístico**

Para el análisis estadístico se utilizó el *software* PLINK (v1.07) (URL: <http://pngu.mgh.harvard.edu/purcell/plink/>), una herramienta estadística para estudios genómicos poblacionales.

Las variantes fueron filtradas por un 70% mínimo de individuos genotipados en esa posición. Como prueba estadística se realizó el test de Fisher para todas las variantes que superaron el filtro, y se seleccionaron las que presentaron un p-valor menor de 0,05. También se obtuvieron datos de frecuencia, el *odd ratio* y la desviación, el equilibrio de Hardy-Weinberg y los límites inferior y superior del intervalo de confianza al 95%. El modelo estadístico utilizado para la obtención del p-valor fue el test *allelic*, basado en las frecuencias alélicas de cada grupo.

Para descartar que las variantes fueran un artefacto se analizó si las proporciones genotípicas cumplían el equilibrio de Hardy-Weinberg.

### **3.7. Análisis funcional**

Para el estudio de la severidad del efecto de las mutaciones puntuales se utilizaron los predictores SIFT y PolyPhen, que permiten predecir el efecto funcional del cambio de aminoácido en una proteína causado por una mutación no sinónima.

## 4. RESULTADOS Y DISCUSIÓN

### 4.1. Diseño de cebadores

Se diseñaron cebadores para 258 amplicones, 132 para la multiplex A y 126 para la multiplex B. El número de amplicones final para la secuenciación de los 10 genes de interés fue de 128, con un tamaño medio de 391 bases.

### 4.2. Puesta a punto de la PCR1-multiplex

#### 4.2.1. Ajuste de condiciones de la PCR

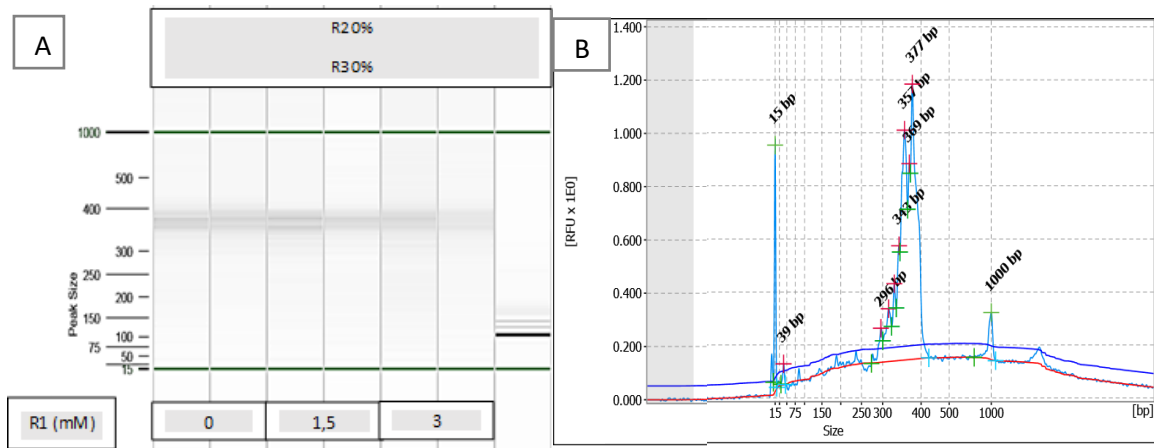
En la puesta a punto de la PCR multiplex se probaron diferentes condiciones en ADN control variando la concentración de distintos reactivos:

[multiplex] (0,007; 0,01; **0,013**; 0,017; 0,02; 0,027; 0,033; 0,04; 0,15  $\mu$ M)

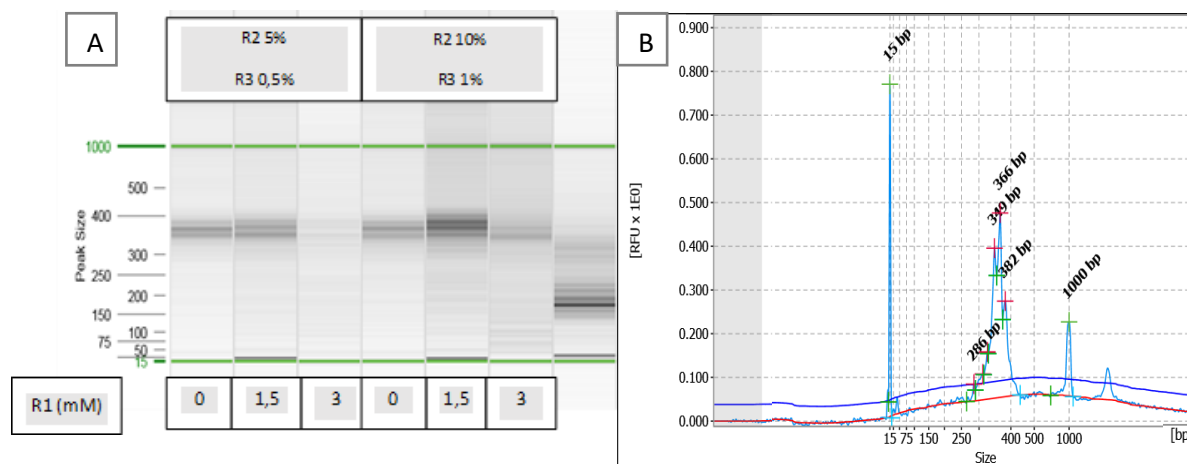
Reactivo 1 (**0**; 1,5; 3; 4; 5; 6 M)

Reactivo 2 (**0%**, 5%, 10%)

reactivo 3 (**0%**; 0,5%, 1%)



**Figura 15. Resultado de la electroforesis capilar mediante el sistema Qiaxcel de las muestras amplificadas sin añadir los reactivos R2 y R3 ya unas concentraciones de R1 de 0, 1,5 y 3 mM .A. Visualización de la banda aproximadamente al tamaño medio esperado de las muestras (391 pb); B. Visualización de los picos de fluorescencia de la electroforesis capilar, que presentan el tamaño esperado y evidencia que las bandas corresponden a los amplicones.**



**Figura 16. Resultado de la electroforesis capilar mediante el sistema Qiaxcel de las muestras amplificadas añadiendo los reactivos R2 (5 y 10%) y R3 (0,5 y 1%) y a unas concentraciones de R1 de 0, 1,5 y 3 mM .A. Visualización de la banda aproximadamente al tamaño medio esperado de las muestras (391 pb); B. Visualización de los picos de fluorescencia de la electroforesis capilar, que presentan el tamaño esperado y evidencia que las bandas corresponden a los amplicones.**

Se utilizaron condiciones modificadas de PCR con una mezcla de enzimas comerciales. Tras las diferentes pruebas se seleccionaron como condiciones específicas para la PCR1 las condiciones sin añadir los reactivos 1, 2 y 3, al obtener resultados similares en la secuenciación de las muestras control y simplificando así el posterior proceso de secuenciación de muestras. Además, en algunas pruebas se observó en la electroforesis capilar mayor suciedad al añadir estos reactivos, que podría deberse a una amplificación menos específica (Figs. 15 y 16).

Respecto a la concentración de mezcla de oligos, se utilizó una concentración de 0,013  $\mu\text{M}$  por pocillo en la PCR1 valorando obtener una intensidad suficiente de las bandas de la electroforesis capilar y la mínima obtención de fragmentos de ADN amplificados inespecíficos posible (Fig. 17).

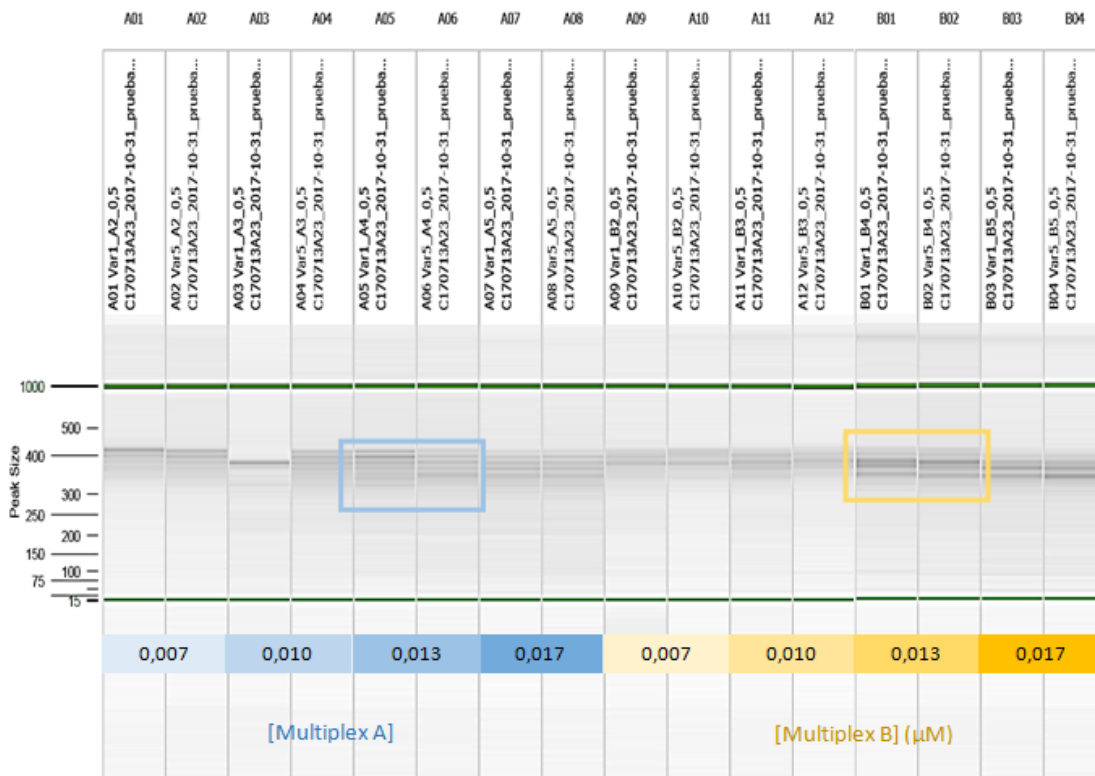


Figura 17. Visualización por electroforesis capilar de la librería amplificada con la multiplex A (izquierda) y multiplex B (derecha) de cebadores a diferentes concentraciones (0,007, 0,01, 0,013, 0,017). Para llevar a cabo la electroforesis se utilizó el sistema Qiaxcel.

Tras la secuenciación con el sistema Miseq se obtuvo la siguiente distribución de coberturas:

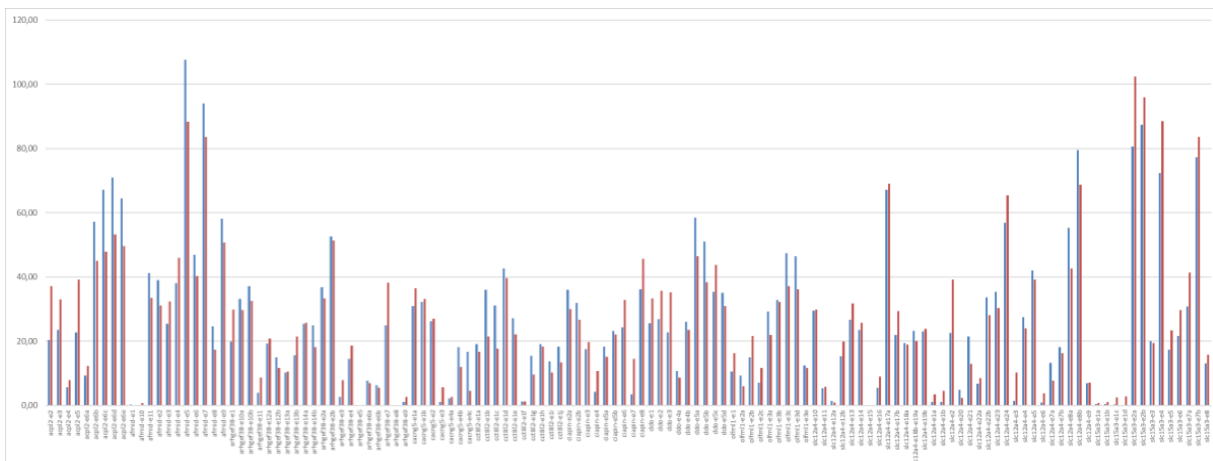


Figura 18. Representación general de la distribución de las lecturas por amplicón en una de las pruebas de validación. En esta prueba se utilizaron las condiciones elegidas para la secuenciación de la población (0,013 nM de Multiplex, sin añadir los reactivos R1, R2 y R3). En el eje y se muestra el número de lecturas y en el eje x el nombre del amplicón.

#### 4.2.2. Reajuste en la concentración y rediseño de primers

Después de ajustar la concentración de los oligos correspondientes a amplicones con demasiado o insuficiente número de lecturas, se obtuvo la siguiente distribución de coberturas:

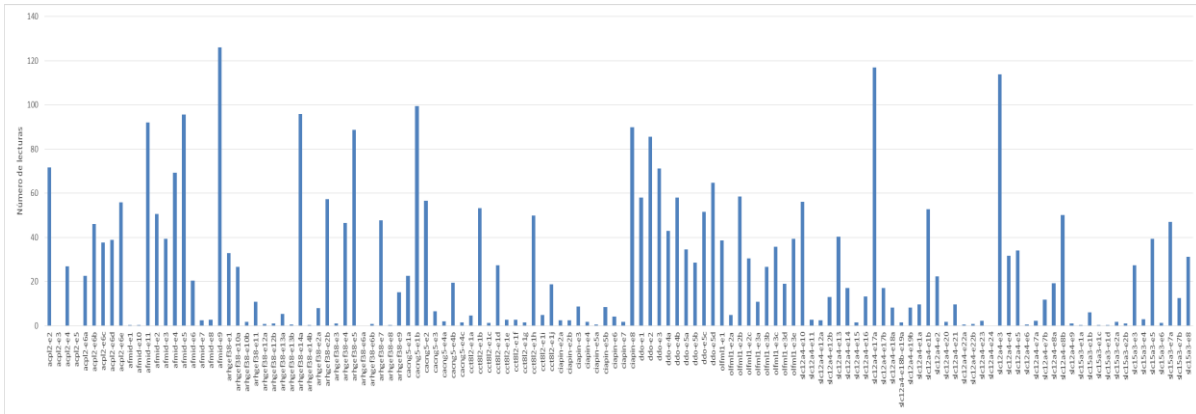


Figura 19. Representación general de la distribución de las lecturas por amplicón en la prueba de ajuste de concentración de cebadores. En el eje y se muestra el número de lecturas y en el eje x el nombre del amplicón.

Dado que se obtuvo una distribución menos uniforme y con un mayor número de amplicones que presentaban muy pocas lecturas (Fig. 19), se procedió al rediseño de los amplicones con bajo número de lecturas.

Tras rediseñar 38 cebadores, la distribución general del número de lecturas no presentó una distribución más uniforme (Fig. 20), aunque de este nuevo diseño se utilizaron 3 oligos en la multiplex (para los amplicones *afmid-e1*, *slc12a4-e20*, *slc12a4-e12a*), que mostraban una mejora significativa respecto al número de lecturas de la primera prueba.

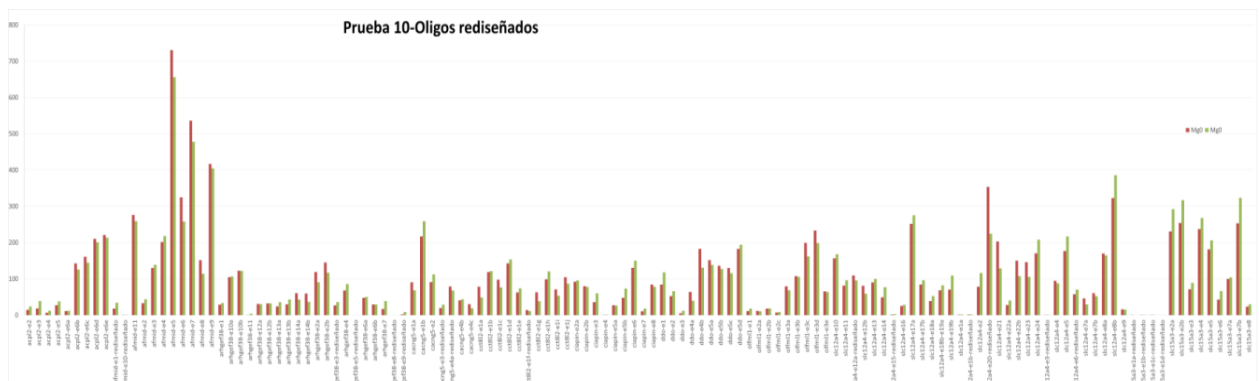


Figura 20. Representación general de las lecturas por amplicón en la prueba de ajuste de concentración de cebadores. En el eje y se muestra el número de lecturas y en el eje x el nombre del amplicón



### 4.3. Electroforesis capilar de las muestras de la población a estudio

Tras las amplificaciones de las librerías de la población, se comprobó que hubiera ADN amplificado previamente a la secuenciación mediante el sistema Qiaxcel (Fig. 21).

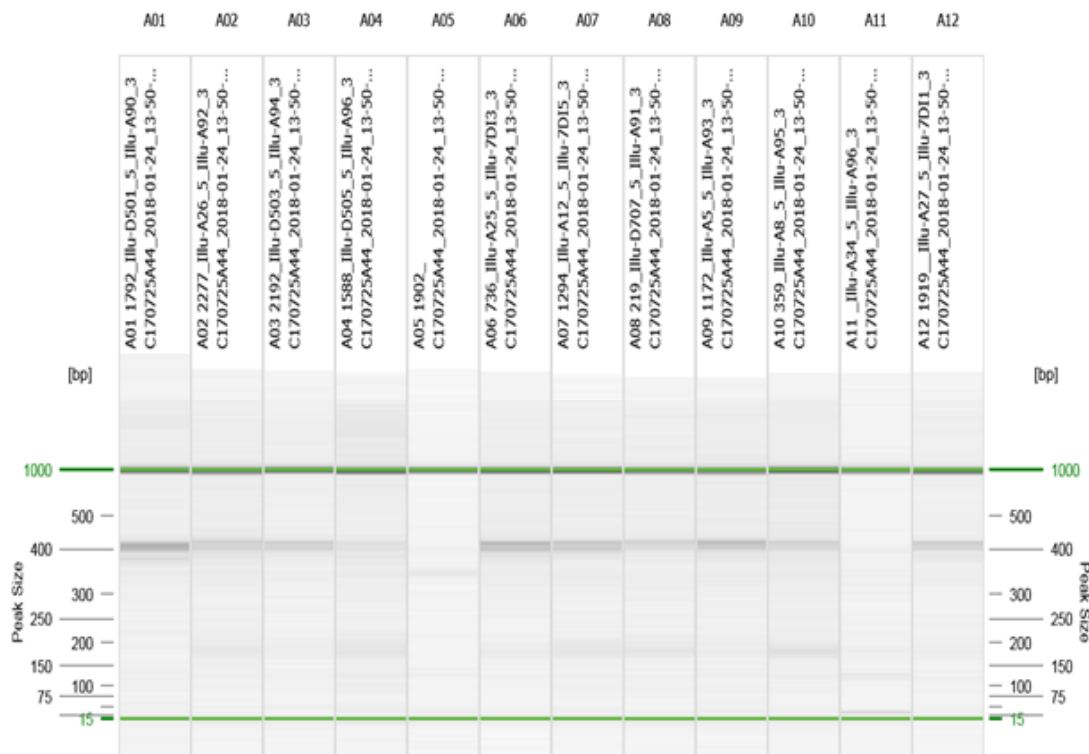


Figura 21. Resultado de la electroforesis capilar mediante el sistema Qiaxcel de las muestras elegidas al azar en una de las placas control.

### 4.4. Análisis bioinformático

El valor medio de Q30 (%) fue de 77,3, indicando que el 77,3 % de las lecturas tienen una probabilidad de lectura errónea de una base del 0,1% o menor. El conteo medio de lecturas fue de  $99121 \pm 47787$ .

Para el filtrado de variantes, se escogieron las que habían sido genotipadas en un 70% de los individuos de la población o superior, quedando seleccionadas 155 variantes de las 372 encontradas (Tablas 9 y 10).

**Tabla 9. Clasificación de las variantes secuenciadas antes del filtro del 70% de individuos genotipados.**

Tipo de variante	ACPL2	AFMID	ARHGEF38	CACNG5	CCT8L2	CIAPIN1	DDO	OLFML1	SLC12A4	SLC15A3	Total
sinónima	1	4	10	7	4	9	7	5	14	9	70
<i>splicing</i>	20	0	0	0	0	3	0	0	5	1	29
<i>missense</i>	0	13	25	7	10	4	6	14	19	15	113
donadora de <i>splicing</i>	1	0	1	0	0	0	0	0	0	0	2
aceptora de <i>splicing</i>	0	1	0	0	0	0	0	0	0	0	1
<i>frameshift</i>	25	0	0	0	0	0	0	0	0	2	27
pérdida de <i>START</i>	0	0	0	0	0	1	0	0	0	0	1
ganancia de <i>STOP</i>	0	0	1	0	1	1	0	1	0	2	6
<i>upstream</i>	0	2	0	0	0	0	0	0	0	0	2
intrónica	3	22	16	13	0	5	4	1	27	14	105
3' UTR	0	8	0	0	0	0	0	0	0	1	9
5' UTR	0	0	0	4	1	1	0	0	0	1	7
total	50	50	53	31	16	24	17	21	65	45	372

Verde: severidad del efecto probablemente más baja, amarillo: severidad moderada, naranja: severidad alta, azul: efecto modificador.

**Tabla 10. Clasificación de las variantes secuenciadas tras el filtro del 70% de individuos genotipados.**

Tipo de variante	ACPL2	AFMID	ARHGEF38	CACNG5	CCT8L2	CIAPIN1	DDO	OLFML1	SLC12A4	SLC15A3	Total
sinónima	0	3	2	7	0	6	1	1	5	6	31
<i>splicing</i>	0	0	0	0	0	2	0	0	3	1	6
<i>missense</i>	0	11	8	7	0	0	0	5	5	9	45
donadora de <i>splicing</i>	0	0	1	0	0	0	0	0	0	0	1
aceptora de <i>splicing</i>	0	1	0	0	0	0	0	0	0	0	1
<i>frameshift</i>	0	0	0	0	0	0	0	0	0	1	1
pérdida de <i>START</i>	0	0	0	0	0	0	0	0	0	0	0
ganancia de <i>STOP</i>	0	0	1	0	0	0	0	0	0	1	2
<i>upstream</i>	0	0	0	0	0	0	0	0	0	0	0
intrónica	0	18	5	10	0	1	1	0	9	10	54
3' UTR	0	8	0	0	0	0	0	0	0	0	8
5' UTR	0	0	0	4	0	0	0	0	0	0	4
total	0	41	17	28	0	9	2	6	22	28	153

Verde: severidad del efecto probablemente más baja, amarillo: severidad moderada, naranja: severidad alta, azul: efecto modificador.

La consecuencia más frecuente encontrada en las variantes secuenciadas fue el grupo de variantes de cambio de sentido, principalmente mutaciones no sinónimas de un único nucleótido que causan la aparición de un codón que codifica para un aminoácido diferente al original pudiendo tener un impacto sobre la funcionalidad de la proteína muy variado, desde no tener ningún efecto funcional a inactivar la proteína, aumentar o desregular su actividad, etc. También se secuenció un alto número de variantes intrónicas; sinónimas, consideradas de baja severidad de impacto al no producir un cambio de codón; y de *splicing*, que podrían interrumpir sitios de empalme existentes o secuencias reguladoras de corte y empalme, crear sitios nuevos o activar los crípticos. Se secuenciaron 6 variantes de ganancia de *STOP*, de las que dos variantes superaron el filtro del 70%, que presentan una severidad de impacto más alta al producirse una proteína truncada (Tabla 11, Fig. 22).

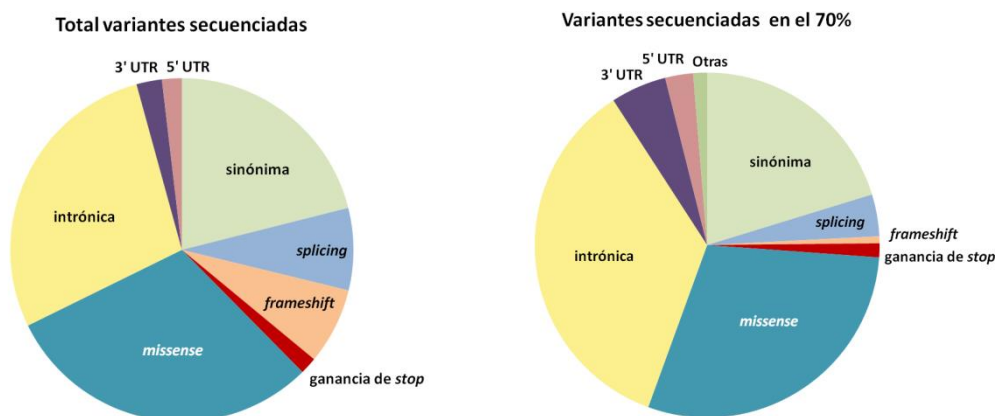


Figura 22. Porcentaje de variantes por efecto en el total secuenciadas (izquierda) y en las variantes filtradas por 70% de la población genotipada (derecha).

En la comprobación de las variantes ya descritas en el proyecto de secuenciación de exoma previo se encontraron tres posiciones en común (Tablas 4 y 11).

Tabla 11 . Variantes encontradas que coincidieron con el estudio previo

Posición Cr.	SNP	Cambio		Gen	Tipo de Variación	Efecto	% Genotipado	F casos	F ctroles
22:17072347	rs139948519	c.1094G>A	p.(W365*)	CCT8L2	Ganancia STOP	Protector <sup>a</sup>	≥ 70	0,002	0,0014
16:57464236	rs170358	c.723C>T	p.(R216*)	CIAPIN1	Ganancia STOP	Protector <sup>a</sup>	≥ 70	0,066	0,08
11:60714017	rs116597710	c.835C>T	p.(R279*)	SLC15A3	Ganancia STOP	Riesgo	≤ 70 <sup>b</sup>	0,011	0

<sup>a</sup> En el previo estudio se clasificó como variante protectora. En este estudio se observó una frecuencia similar en controles y casos.

<sup>b</sup> No identificada tras el filtro de ≤ 70% de individuos genotipados para esa posición. Genotipados 88 casos y 98 controles.

De los tres cambios nucleotídicos coincidentes con el anterior proyecto se secuenciaron los siguientes:

- En el gen *CCT8L2* se encontró la variante de ganancia de *STOP* rs139948519. Mientras que en el anterior estudio se clasificó como variante con efecto protector, en el presente trabajo se obtuvieron frecuencias similares en casos (0,002) y controles (0,0014).
- En el gen *CIAPIN1* se detectó la variante de ganancia de *STOP* rs170358 en una frecuencia similar en casos (0,066) y controles (0,08), a pesar de estar clasificada previamente como protectora.

- En el gen *SLC15A3* fue secuenciada la variante de ganancia de *STOP* rs116597710, previamente clasificada como variante de riesgo. Dado que en este estudio solo fue secuenciada en un paciente diabético, se estaría contribuyendo a la validación del resultado previo, aunque es necesario destacar que esta variante no fue secuenciada en más del 70% de casos y controles.

En el análisis de las variantes encontradas se priorizaron las de efecto moderado o alto, incluyendo las variantes de ganancia de *STOP*, *splicing* y *missense*. Se consideraron con efecto alto las variables de *splicing* presentes a  $\pm 3$  nucleótidos del transcrito, y dentro de las variantes *missense*, las que presentaban un valor de SIFT  $\leq 0,05$ .

Se obtuvieron 5 variantes estadísticamente significativas según el test de Fisher utilizando el modelo *allelic*, comparando su frecuencia en casos y controles. 4 fueron genotipadas en más del 70% de la población. Entre estas variantes encontramos dos variantes intrónicas: una en el gen *SLC15A3*, que tendría efecto protector, y otra en el gen *SLC12A4*, que sería clasificada como de riesgo según las frecuencias obtenidas. A pesar de que las variantes intrónicas no suelen tener un gran efecto en la proteína, pueden alterar los motivos reguladores de la transcripción, afectando a los niveles de expresión, la señalización o las modificaciones post-transcripcionales. Por otro lado, se obtuvo como significativa una variante clasificada como sinónima en el gen *CIAPIN1*. Dado que no se produciría un cambio aminoacídico en la proteína, podría tratarse de un falso positivo. Por último, se identificó una variante *missense* en el gen *OLFML1*, que tendría un efecto protector al ser más frecuente en controles y presentó efecto tolerado en la proteína según el predictor predictor SIFT. Entre las variantes genotipadas en menos del 70% de la población, se identificó una variante *missense* en el gen *ARHGEF38* que sería clasificada de riesgo en base a la frecuencia, y presentaría efecto de dañina según el predictor SIFT, pudiendo verse alterada la estructura y función de la proteína (Tabla 12).

**Tabla 12. Variantes estadísticamente significativas (p-valor <0,05).**

Posición	rs	Gen	Variante		Transcrito	MAF	Tipo	p-valor <sup>b</sup>	Efecto <sup>c</sup>	SIFT
11:60706937	-	<i>SLC15A3</i>	G>A	-	ENST00000227880	-	intrónica	0,0176	Protector	-
4:106580338	-	<i>ARHGEF38</i> <sup>a</sup>	c.1507A>G	Thr455Ala	ENST00000420470	-	<i>missense</i>	0,0275	Riesgo	Dañina (N/A)
16:57465116	-	<i>CIAPIN1</i>	c.335G>T	Ser231Ser	ENST00000394391	-	sinónima	0,0381	Riesgo	-
16:67984416	-	<i>SLC12A4</i>	G>T	-	ENST00000316341	-	intrónica	0,0454	Riesgo	-
11:7531229	rs76280703	<i>OLFML1</i>	c.1019G>A	p.Arg340His	ENST00000329293	< 0.01	<i>missense</i>	0,025	Protector	Tolerada (0,11 <sup>d</sup> )

<sup>a</sup> No identificada tras el filtro de  $\leq 70\%$  de individuos genotipados para esa posición. Genotipados 277 casos y 402 controles.

<sup>b</sup> p-valor utilizando el modelo "Allelic".

<sup>c</sup> Valor SIFT

## **4.5. Análisis de variantes por gen.**

### **4.5.1. ACPL2**

El gen *ACPL2* (Proteína 2 similar a fosfatasa ácida) o *PXPYLP1* (2-Fosfoxilosa fosfatasa 1) está implicado en la biosíntesis de los proteoglicanos sulfato de condroitina y sulfato de heparán. La proteína *ACPL2* actúa a través de la 2-O-desfosforilación de la xilosa en la región de unión de glucosaminoglucano (GAG)-proteína de los proteoglicanos, regulando así la cantidad de cadenas de GAG maduros (*National Center for Biotechnology Information*).

En estudios previos se ha relacionado la diabetes con la alteración estructural del sulfato de condroitina en la orina de pacientes con DM (Zhao *et al.*, 2013), así como en riñón de ratas en condiciones de diabetes (Joladarashi *et al.*, 2011) y ratones db/db diabéticos (Reine *et al.*, 2013). También se observó una disminución del contenido de sulfato de heparán en la membrana basal glomerular en biopsias renales de pacientes con nefropatía diabética y en un experimento con células epiteliales viscerales glomerulares y mesangiales en condiciones de hiperglucemia, en el que se concluyó que concentraciones altas de glucosa producen la alteración de la producción de heparán sulfato de estas células (van Det *et al.*, 1996).

No se encontraron variantes genotipadas en más del 70% de la población, probablemente debido a la baja cobertura de este gen, dada las características de su secuencia.

Dentro de las variantes genotipadas en menos del 70% de la población se encontró una variante de cambio de pauta de lectura en un caso (de entre 99 casos secuenciados), en el que se producía una inserción de una C (c.255dup). También se genotiparon variantes de *splicing*, tanto en homocigosis, que al estar presente en los dos alelos podrían tener un efecto considerable en la proteína; como en heterocigosis, aunque no se observaron diferencias significativas entre casos y controles.

Las variantes perjudiciales en *ACPL2* en pacientes diabéticos podrían alterar la función de esta proteína y, por tanto, la biosíntesis de proteoglicanos, a su vez relacionada con diabetes.

### **4.5.2. AFMID**

La enzima arilformamidasa (AFMID), también conocida como quinureninaformamidasa, cataliza el segundo paso de la conversión de triptófano a ácido nicotínico, NAD (H) y NADP (H), hidrolizando N-formil-L-quinurenina a L-quinurenina (Fig. 23) (*National Center for Biotechnology Information*).

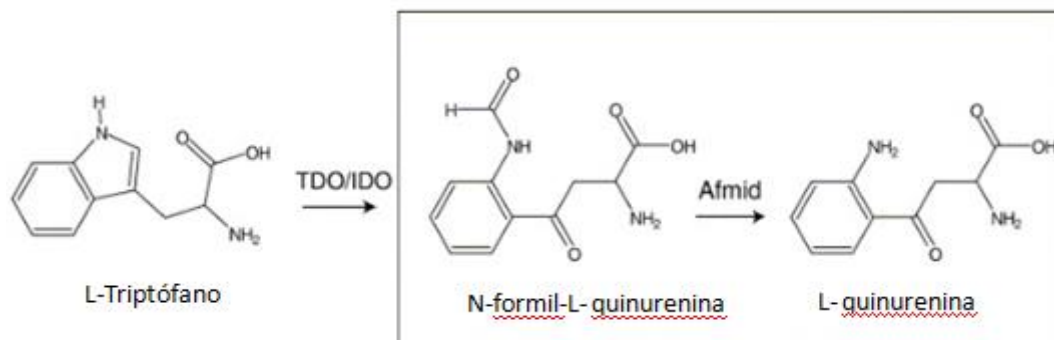


Figura 23. Mecanismo de acción de la enzima arilformamidasa.

El catabolismo anormal de triptófano se ha relacionado en estudios observacionales con DM2. En un estudio con ratones *knockout* para *AFMID* y *Tk*, estos presentaban tolerancia alterada a la glucosa debido a un defecto en la secreción de insulina estimulada por glucosa. Estos ratones mostraron una masa de islotes reducida con la edad y la sensibilidad a la insulina no se vio alterada. No se encontraron pruebas de un fenotipo renal, lo que sugiere que este fenotipo publicado resultó de la pérdida de la expresión de *Tk* en el doble *knockout*. Sin embargo, a pesar de eliminar específicamente solo el exón 2 de *AFMID*, también se observó cierta reducción de la expresión de *Tk*, posiblemente debido a un elemento regulador en esta región (Hugill *et al.*, 2015).

En otro estudio se estudió la activación de la vía de la quinurenina en individuos obesos, concluyendo que la presencia de macrófagos en el tejido adiposo podría contribuir a desviar la vía de la kururenina hacia la activación de la la monooxigenasa de quinurenina (Favennec *et al.*, 2015).

Dentro de las variantes *missense* filtradas encontradas, se estudió el efecto en la proteína de ciertas variantes *missense* que aparecían solo en el grupo de casos o solo en el de controles, dado que a pesar de estar presentes en un número insuficiente de pacientes para hacer el análisis estadístico, nos podrían aportar información observadas en conjunto (Tabla 13).

Tabla 13. Clasificación de variantes *missense* en el gen *AFMID* genotipadas en más del 70% de individuos.

chr:position	rs	Variante	MAF	SIFT	S.S.	POLYPHEN	Casos			Controles			
							0/0	0/1	1/1	0/0	0/1	1/1	
17:76187109	rs758457659	c.122C>A	p.Ala41Asp	< 0.01	deleterious	0,03	-	509	1	0	705	0	0
17:76198628	rs774921922	c.203A>G	p.Tyr68Cys	< 0.01	deleterious	0	damaging	577	1	0	829	0	0
17:76198636	rs144866146	c.211G>A	p.Gly71Ser	< 0.01	tolerated	0,42	benign	578	0	0	828	1	0
17:76198681	rs746267035	c.256G>A	p.Glu86Lys	< 0.01	tolerated	0,91	benign	578	0	0	828	1	0
17:76198828	rs143224831	c.304G>A	p.Gly102Arg	< 0.01	deleterious	0	damaging	577	1	0	829	0	0
17:76201182	rs61742958	c.476A>G	p.Tyr159Cys	0.02	deleterious	0,01	damaging	580	2	0	827	0	0
17:76201208	-	c.502C>T	p.His168Tyr	-	deleterious	0	damaging	584	1	0	828	0	0

Se utilizaron los transcritos ENST00000327898 y ENSP00000328938. S.S.: Sift Score.

Tal y como se observa en la tabla, 6 de las variantes *missense* fueron secuenciadas en heterocigosis solo en casos, presentando un efecto sobre la proteína según los predictores *SIFT/POLYPHEN* deletéreo/dañino; y dos de las variantes se secuenciaron solo en controles, con un efecto tolerado/benigno. Mediante el análisis estadístico se comparó el número de casos con variantes deletéreas respecto a controles sanos, obteniendo un p-valor de 0,005, por lo que se trataría de una diferencia considerablemente significativa. Respecto al número de mutaciones *missense* total, se obtuvo un p-valor <0,05, concretamente de 0,03, siendo mayor la frecuencia en casos.

Teniendo en cuenta estos resultados y los estudios previos, las variantes *missense* identificadas con mayor efecto deletéreo en la proteína, podrían alterar la estructura y función de la proteína AFMID y estar relacionadas con una mayor predisposición a alteraciones en la tolerancia a la glucosa y, por tanto con una mayor predisposición a DM2.

#### **4.5.3. ARHGEF38**

*ARHGEF38* (Rho Guanine Nucleotide Exchange Factor 38) codifica una proteína que activa las GTPasas monoméricas, estimulando la liberación de difosfato de guanosina (GDP) para permitir la unión del trifosfato de guanosina (GTP). Las GTPasas Rho desempeñan un papel fundamental en numerosos procesos iniciados en la célula mediante estímulos extracelulares que funcionan a través de receptores acoplados a proteínas G. La proteína codificada puede formar complejos con proteínas G y estimular señales dependientes de rho (*National Center for Biotechnology Information*).

Entre sus vías relacionadas están la señalización mediada por el receptor de neurotrofina p75 (p75 NTR) y la señalización por receptor acoplado a proteínas G (GPCR). Previamente se ha observado que el receptor p75 NTR regula la homeostasis de la glucosa y la sensibilidad a la insulina (Baeza-Raja *et al.*, 2012).

Variantes en el gen *ARHGEF11* (Rho guaninenucleotide exchange factor (GEF) 11), perteneciente a la misma familia génica que *ARHGEF38*, se han asociado a resistencia a insulina y DM2 en indios Pima (Ma *et al.*, 2007). En otro estudio (Zhang *et al.*, 2016), se relacionó la sobreexpresión de la enzima heparanasa con mejoras en la homeostasis de glucosa, y se midieron diferencias en la expresión génica en los islotes pancreáticos de ratones transgénicos que sobreexpresaban esta enzima. Uno de los genes sobreexpresados fue *ARHGEF38*. Esto sugiere que este gen podría estar regulando a la enzima heparanasa y por tanto regulando la concentración de heparán-sulfato, previamente relacionado con diabetes, tal y como se ha explicado anteriormente.

Entre las variantes genotipadas en menos del 70%, en el análisis estadístico de las variantes mediante *plink*, se obtuvo una variante *missense* (A>G) estadísticamente significativa según el test de Fisher en la posición 4:106580340, con un p-valor de 0,028 según el modelo *allelic* (<0,05). Esta variante fue secuenciada en homocigosis únicamente en 2 pacientes diabéticos. Cabe destacar, que esta variante fue genotipada en 277 casos y 402 controles, por lo que no se superó el filtro del 70% de individuos secuenciados. Aun así, fue clasificado por SIFT como dañina, produciendo un cambio de aminoácido de treonina a alanina. En este caso la variante *missense* se encontró en homocigosis, por lo que en los individuos que presentaron esta variante en los dos alelos la función proteica de ARHGEF38 podría estar alterada, al no presentar ninguna de las copias funcionales. Variantes con este tipo de efecto dañino en la proteína ARHGEF38 podrían estar relacionadas con DM2, pudiendo incrementar el riesgo de resistencia a insulina o alteraciones en la homeostasis de la glucosa. Estos datos se deberán verificar en sucesivos estudios ampliando las muestras analizadas.

Dentro del conjunto de variantes *missense* genotipadas en más del 70% de individuos, se comparó el total de las genotipadas en el grupo de casos y en el de controles teniendo en cuenta el efecto en la proteína según los predictores SIFT y POLYPHEN (Tabla 14).

**Tabla 14. Clasificación de variantes missense en el gen ARHGEF38 genotipadas en más del 70% de individuos.**

chr:position	rs	Variante		MAF	SIFT	S.S.	POLYPHEN	Casos			Controles		
								0/0	0/1	1/1	0/0	0/1	1/1
4:106510492	rs776205833	c.284G>A	p.Arg95Gln	< 0.01	deleterious	-	damaging	510	1	0	758	0	0
4:106510540	rs765574141	c.332A>G	p.Asn111Ser	< 0.01	tolerated	0,19	benign	517	0	0	721	1	0
4:106510558	-	c.350T>C	p.Val117Ala	-	deleterious	0,01	-	517	0	0	717	1	0
4:106510581	-	c.517A/G	p.Arg125Gly	-	deleterious	0,01	damaging	529	0	0	740	1	0
4:106510582	rs144590203	c.374G>C	p.Arg125Thr	< 0.01	deleterious	0,01	-	530	0	0	741	1	0
4:106552103	rs771256401	c.557T>G	p.Ile186Ser	< 0.01	deleterious	0	damaging	569	1	0	817	0	0
4:106552180	rs774704317	c.634A>T	p.Ser212Cys	< 0.01	tolerated	0,34	-	569	0	0	816	1	0
4:106599018	-	c.476A>G	p.Tyr159Cys	-	deleterious	0,01	damaging	465	0	0	670	1	0

Transcritos: ENST00000420470y ENSP00000265154.2. S.S.: Sift Score.

En este caso, analizando las variantes *missense* genotipadas en más del 70% de individuos, no se encontró una sobrerrepresentación en uno de los dos grupos (diabéticos/controles) de las variantes más deletéreas. Estas variantes se encontraron en heterocigosis, por lo que en caso de tener efecto funcional sobre ARHGEF38, supondría que con una copia funcional podría ser suficiente para no presentar un fenotipo alterado.



#### 4.5.4. CACNG5

El gen *CACNG5* (OMIM\*606405) codifica la Subunidad Auxiliar Gamma 5 del Canal de Calcio dependiente de Voltaje (*CACNG5*), también llamada Proteína Reguladora Transmembrana de AMPAR Gamma-5. Este tipo de proteínas regulan el tráfico y la canalización de los receptores de ácido  $\alpha$ -amino-3-hidroxi-5-metilo-4-isoxazolpropiónico o AMPAR, que son receptores transmembrana ionotrópicos para glutamato y median la transmisión sináptica rápida en el sistema nervioso central. Los receptores de glutamato se dividen en dos grupos principales: los canales ionotrópicos activados por ligando de acción rápida, entre los que se encuentran los receptores AMPA, kainato y NMDA (N-metil-D-aspartato) y los receptores metabotrópicos de acción más lenta. Los AMPAR se localizan en muchas regiones del cerebro y es el receptor más abundante en el sistema nervioso.

Varios estudios han relacionado la diabetes con alteraciones en los receptores de glutamato. Se observó que el ratón *ob/ob*, modelo de diabetes tipo 2, tenía sitios de unión aumentados para los receptores NMDA y AMPA en la sustancia gris de la médula espinal, así como un aumento de la afinidad del receptor NMDA, con respecto a ratones control. Por ello, ambos receptores podrían estar involucrados en las neuropatías periféricas que pueden acompañar a la DM2 (Li *et al.*, 1999). En otro estudio en ratas diabéticas inducidas por estreptozotocina, tanto las propiedades de unión de los receptores AMPA, como la expresión de la subunidad GluR1 se vieron disminuidas (Chabot *et al.*, 1997).

El glutamato también es el principal neurotransmisor excitador en la retina. En una investigación en ratas se concluyó que la diabetes afecta al contenido en la retina de las subunidades del receptor de glutamato ionotrópico a nivel proteico, por lo que el gen *CACNG5* podría estar implicado en la retinopatía diabética característica de la enfermedad, al regular los receptores AMPA (Santiago *et al.*, 2009).

De entre las variantes que superaron el filtro del 70% se secuenciaron 4 variantes en la región 5'UTR (Tabla 15). Este tipo de variantes génicas pueden estar implicadas en la regulación de la expresión génica y, por tanto, podrían tener un efecto sobre la proteína. Dentro de este grupo, destaca la variante c.174C>T, que fue genotipada con frecuencia mayor en controles, siendo uno de los individuos control, homocigoto para la variante. Por otra parte, la variante c.189C>T fue genotipada en 3 de los controles y en ningún paciente diabético. Estas variantes podrían estar relacionadas con un efecto protector, ya sea por una ganancia de función o pérdida de esta. Tal y como se ha comentado previamente, en ratones modelo de DM2 se observó un aumento en los sitios de unión de los AMPAR. Dado que las variantes 5'UTR podrían estar

relacionadas con la expresión de *CACNG5*, que a su vez regula a AMPAR, podrían producir cambios en la actividad de estos receptores.

**Tabla 15. Variantes 5'UTR genotipadas en >70% en el gen *CACNG5*.**

chr:position	rs	Variante	MAF	Casos			Controles		
				0/0	0/1	1/1	0/0	0/1	1/1
17:64873368	-	c.155G>A	-	482	0	0	718	1	0
17:64873374	-	c.161C>A	-	481	1	0	719	0	0
17:64873387	rs186268864	c.174C>T	0,01	479	3	0	712	6	1
17:64873402	rs189220145	189C>T	< 0,01	495	0	0	732	3	0

Transcritos: ENST00000533854 y ENSP00000169565.3.

Por otra parte, también se secuenciaron 7 variantes *missense*, genotipadas en más del 70% de la población (Tabla 16). Estas variantes no aparecieron sobrerrepresentadas de forma significativa en uno de los dos grupos (diabéticos/controles), aunque destacaría la variante c.383C>T, que fue genotipada en 3 controles y ningún caso, pudiendo ser considerada como protectora, posiblemente por procesos similares a los descritos anteriormente, aunque serían necesarios más estudios (el efecto en la proteína sería benigno según los predictores SIFT y POLYPHEN).

**Tabla 16. Variantes *missense* genotipadas en >70% en el gen *CACNG5*.**

chr:position	rs	Variante		MAF	SIFT	S.S.	POLYPHEN	Casos			Controles		
								0/0	0/1	1/1	0/0	0/1	1/1
17:64873529	rs993681024	c.79G>A	p.Ala27Thr	< 0.01	deleterious	0	damaging	582	0	0	825	1	0
17:64873550	-	-	-	-	deleterious	0	damaging	580	1	0	825	1	0
17:64873605	rs758545923	c.155T>C	p.Met52Thr	< 0.01	tolerated	0,59	benign	579	1	0	826	1	0
17:64875099	rs758742270	c.206G>A	p.Arg69Gln	< 0.01	tolerated	0,14	-	502	0	0	732	1	0
17:64876761	rs145214481	c.371G>A	p.Arg124His	< 0.01	tolerated	0,46	benign	516	1	0	751	1	0
17:64876770	rs142916987	c.380G>T	p.Arg127Leu	< 0.01	deleterious	0,01	damaging	519	1	0	752	0	0
17:64876773	rs149159754	c.383C>T	p.Thr128Met	< 0.01	tolerated	0,09	benign	520	0	0	750	3	0

Transcritos: ENST00000533854 y ENSP00000169565.3. S.S.: Sift Score.

#### 4.5.5. *CCT8L2*

La proteína chaperonina que contiene la subunidad 8 tipo 2 de TCP1 (*CCT8L2*) forma parte del complejo chaperonina TRiC/CCT, que es dependiente de ATP y tiene como función el plegamiento de proteínas citosólicas. TRiC/CCT regula a la proteína HSF1, un factor de transcripción que protege a las células de apoptosis y estrés inducido por proteínas mal plegadas. La expresión de proteínas chaperonas y la actividad de HSF1 se ha visto disminuida en tejidos resistentes a la insulina en DM2, pudiendo ser consecuencia de alteraciones en las subunidades que forman el complejo TRiC/CCT (Neef *et al.*, 2014).

En DM2 se ha observado la formación de agregados del polipéptido amiloide de los islotes o amilina, que están implicados en la muerte celular por apoptosis en las células beta. Una posible hipótesis sería que el incorrecto funcionamiento de TRiC/CCT conllevara la disminución del efecto antiapoptótico en células beta y del efecto protector ante la amilina (Chiti *et al.*, 2006).

No se encontraron variantes genotipadas en más del 70% de la población, probablemente debido a la baja cobertura de este gen.

Dentro de las variantes genotipadas en menos del 70% de la población se encontró la variante 5'UTR c.-35G>A en 2 casos de pacientes diabéticos de 28 secuenciados, y ninguno en 36 controles, pudiendo estar relacionada con riesgo. Además, de entre todas las variantes *missense*, dos fueron secuenciadas en 2 casos diabéticos (c.1425C>A) y en 1 caso diabético (c.1406G>A) sin aparecer en ningún control, y con predicción SIFT de dañinas, pudiendo indicar una relación entre alteraciones en la proteína con riesgo a DM2. Dado el bajo número de muestras genotipadas, estos datos se deberán verificar en sucesivos estudios ampliando las muestras analizadas.

#### **4.5.6. CIAPIN1**

CIAPIN1 (OMIM\*608943) es un inhibidor de la apoptosis inducido por citoquinas, que promueve el desarrollo de células hematopoyéticas (Chen *et al.*, 2012).

El gen *CIAPIN1* está altamente expresado en páncreas, pudiendo estar implicado en la supervivencia de células  $\beta$ . De hecho, se ha estudiado su implicación en la inhibición de la proliferación de células de cáncer de páncreas, al retrasar la progresión del ciclo celular e inducir la apoptosis celular. Por tanto, mutaciones en *CIAPIN1* podrían estar relacionadas con el incremento en la apoptosis de células  $\beta$  en pacientes diabéticos

Por otra parte, este gen también está implicado en la maduración de proteínas Fe-S extramitocondriales, al formar parte, junto con NADPH y la proteína diflavina NDOR1, de la cadena de transporte de electrones, necesarios para la reacción de ensamblaje necesaria para la biogénesis de Fe-S citosólica (Paul *et al.*, 2015). El dominio CDGSH es parte de un motivo de unión hierro-azufre (2Fe-2S), que aparece en varias proteínas humanas, como mitoNEET, un homodímero que está anclado a la membrana mitocondrial externa y se ha visto involucrado en la diabetes, la obesidad y las enfermedades cardiovasculares, entre otros (Sengupta *et al.*, 2018).

**Tabla 17. Variantes *missense* genotipadas en >70% en el gen *CIAPIN1*.**

chr:position	rs	Variante		MAF	SIFT	S.S.	Casos			Controles		
							0/0	0/1	1/1	0/0	0/1	1/1
17:57464175	-	c.823G>A	p.Gly275Arg	-	deleterious	0	445	1	0	645	0	0
17:57464219	-	c.799A>G	p.Glu260Gly	-	tolerated	0,13	445	1	0	647	0	0

Transcritos: ENST00000394391.4y ENSP00000377914.4. S.S.: Sift Score.

De las dos variantes *missense* encontradas tras el filtro del 70% de individuos genotipados (Tabla 17), la variante c.823G>A podría ser de interés al tener una predicción SIFT dañina y estar presente únicamente en un caso. Este cambio podría estar relacionado con riesgo a DM2 al verse alterada la estructura proteica por la sustitución de glicina por arginina. El aminoácido glicina está implicado en la conformación estructural de la proteína y suele estar presente en zonas de giro u “horquilla” al no tener cadena lateral.

De entre las variantes genotipadas en <70% destaca la variante de pérdida de *START* c.1A>G, que fue genotipada en heterocigosis en uno de los controles y está clasificada como deletérea por los predictores SIFT y POLYPHEN. También se secuenció una variante de ganancia de *STOP* (c.350C>A), en este caso en uno de los pacientes diabéticos. Ambas variantes tenían una MAF<0,01. En ambas se produciría una proteína más corta, pero en el primer caso faltaría secuencia proteica al inicio hasta el siguiente codón *START*, mientras que la variante de *STOP* originaría una proteína truncada en el aminoácido 116 de 312, es decir, tendría un tamaño menor a la mitad del normal, por lo que se trataría de una variante probablemente patogénica que afectaría a la función.

#### **4.5.7. DDO**

La D-aspartato-oxidasa (DDO) (OMIM\*124450) es una flavoproteína que cataliza la desaminación oxidativa de los D-aminoácidos dicarboxílicos a  $\alpha$ -cetoácidos y amoníaco. DDO es inactivo con D-aminoácidos que se oxidan por D-aminoácido oxidasa (DAO), que cataliza la misma reacción con D-aminoácidos neutros y básicos como sustrato. DDO es la única enzima catabólica que actúa sobre D-Asp en mamíferos. Se trata de un D-aminoácido libre que tiene funciones cruciales en los sistemas neuroendocrino, endocrino y del sistema nervioso central (Setoyama *et al.*, 1997).

Estudios recientes han implicado a D-Asp en la fisiopatología de la infertilidad y las enfermedades relacionadas con el receptor de esta molécula, N-metil-D-Asp (NMDA), un subtipo del receptor de L-glutamato. Los NMDAr se expresan en todo el riñón, y la abundancia de estos receptores y algunos de sus agonistas endógenos aumenta en condiciones de

diabetes. Actualmente, la proteína DDO humana se considera un objetivo terapéutico, y los inhibidores de DDO pueden tener potencial como nuevos fármacos contra las enfermedades relacionadas con los sistemas mencionados anteriormente. Sin embargo, esta proteína no se ha caracterizado en detalle en humanos (Katane *et al.*, 2011; Roshanravan *et al.*, 2016).

Se genotiparon únicamente dos variantes en más del 70% de la población para este gen, siendo una sinónima y otra intrónica, consideradas variantes con bajo efecto en la proteína.

Dentro de las variantes secuenciadas en menos del 70% se encontraron 6 tipo *missense*, pero no se pudo establecer una relación entre estas mutaciones en el gen *DDO* y el riesgo o protección a DM2.

#### **4.5.8. OLFM1**

La proteína OLFM1 (olfactomedina-1) (OMIM\*605366) comparte una gran similitud de secuencia (9%) con la proteína localizada en el retículo endoplasmático relacionada con la olfactomedina neuronal de rata. Si bien no se conoce la función exacta de la proteína codificada, su abundante expresión en el cerebro sugiere que puede tener un papel esencial en el tejido nervioso. Se han encontrado transcritos alternativos que codifican diferentes isoformas.

En un estudio muy reciente (Cobo *et al.*, 2018), se utilizó la molécula BL001 para provocar un ambiente tolerogénico dentro del páncreas. En consecuencia, el perfil de transcriptoma reveló que BL001 aumentó significativamente la expresión de genes clave asociados con la firma genética de macrófagos M2, que incluyen a *OLFM1* (Noelin-1 o Pancortin), entre otros.

Se genotiparon 6 variantes *missense* en más del 70% de la población, dos de ellas con efecto deletéreo según los predictores SIFT y POLYPHEN y 3 con efecto tolerado (Tabla 18). Se secuenciaron más variantes con efecto deletéreo en controles que en casos. Respecto a la comparación del total de variantes en casos y controles independientemente del efecto, se obtuvo un p-valor de 0,04 (<0,5), siendo mayor la frecuencia en controles. Esto podría indicar que variantes *missense* en el gen *OLFM1* podrían tener un efecto protector a padecer DM2, posiblemente por cambios en la actividad proteica.

**Tabla 18. Variantes missense genotipadas en >70% en el gen *OLFML1*.**

chr:position	rs	Variante		MAF	SIFT	S.S.	POLYPHEN	Casos			Controles		
								0/0	0/1	1/1	0/0	0/1	1/1
11:7531076	-	c.866A>C	p.His289Pro	-	tolerated	0,24	benign	508	1	0	700	4	1
11:7531084	-	c.874T>G	p.Leu292Val	-	deleterious	0,05	damaging	518	2	0	754	5	0
11:7531132	-	c.926A>G	p.Asp309Gly	-	deleterious	N/A	-	482	0	0	667	1	0
11:7531135	rs370996733	c.925G>A	p.Asp309Asn	< 0.01	tolerated	0,14	-	479	3	0	674	0	0
11:7531229	rs76280703	c.1019G>A	p.Arg340His	< 0.01	tolerated	0,12	benign	532	3	1	728	16	2

Transcritos: ENST00000394391.4 y ENST00000329293.3. S.S.: Sift Score.

Entre las variantes genotipadas en menos del 70% de la población se secuenció la variante de *STOP* c.745C>T (rs759190256) en heterocigosis en uno de los casos de pacientes con DM2. Esta variante causaría una proteína truncada en el aminoácido 249, siendo el tamaño normal de la proteína *OLFML1* de 402 aminoácidos. Esta reducción a casi la mitad del tamaño podría causar una pérdida de función total o parcial de la proteína, que podría estar compensada por el otro alelo funcional del gen.

#### 4.5.9. *SLC12A4*

El gen *SLC12A4* (miembro 3 de la familia 15 de transportadores de solutos) o *KCC1* (OMIM \*604119), es un cotransportador de cloruro que media el movimiento acoplado de iones potasio y cloruro a través de la membrana plasmática.

La concentración de cloruro intracelular en las células  $\beta$  pancreáticas se mantiene por encima del equilibrio electroquímico debido a la predominancia funcional de cargadores de  $\text{Cl}^-$ , como el  $\text{Na}^+ \text{K}^+ 2\text{Cl}^-$  cotransportador 1 (*SLC12A2*) sobre los extrusores  $\text{Cl}^-$ . Este mecanismo contribuye a la fuerza de despolarización de  $\text{Cl}^-$  ligada a la secreción de insulina. Los islotes pancreáticos de rata también expresan varios extrusores de  $\text{Cl}^-$ , incluyendo *SLC12A4* (*KCC1*), *SLC12A6* (*KCC3*) y *SLC12A7* (*KCC4*). Estos transportadores parecen estar enriquecidos en células  $\alpha$  secretoras de glucagón (Kursan *et al.*, 2017). Teniendo en cuenta su función, mutaciones en el gen *SLC12A4* podrían estar implicadas en la alteración del equilibrio electroquímico en células  $\alpha$  y  $\beta$  pancreáticas.

Entre las variantes genotipadas en más del 70% se encontraron tres variantes de *splicing*, que no fueron consideradas con efecto importante sobre la proteína debido a que se encontraban a más de 3 nucleótidos del exón codificante. Además, se secuenciaron 5 variantes *missense*, de las cuales 4 de ellas aparecieron en controles, pudiendo estar relacionadas con un efecto protector. 2 de estas variantes eran deletéreas y 2 toleradas según los predictores SIFT y POLYPHEN (Tabla 19).

**Tabla 19. Variantes *missense* genotipadas en >70% en el gen *SLC12A4*.**

chr:position	rs	Variante		MAF	SIFT	S.S.	POLYPHEN	Casos			Controles		
								0/0	0/1	1/1	0/0	0/1	1/1
16:67978795	rs146713633	c.3206G>A	p.Arg1069Gln	< 0.01	deleterious	0	-	549	2	0	771	2	0
16:67979874	rs373734460	c.2719G>A	p.Glu907Lys	< 0.01	tolerated	0,12	benign	580	0	0	822	1	0
16:67983745	rs62059925	c.1706T>C	p.Ile569Thr	< 0.01	deleterious	0	-	534	0	0	748	3	0
16:67985884	rs763679917	c.974A>G	p.Lys325Arg	< 0.01	tolerated	0,08	-	422	0	0	580	1	0
16:67997459	-	c.119T>C	p.His40Arg	-	deleterious	0,02	benign	565	0	0	805	1	0

Transcritos: ENST00000316341.3 y ENSP00000318557.3. S.S.: Sift Score.

#### 4.5.10. *SLC15A3*

El miembro 3 de la familia 15 de transportadores de solutos (*SLC15A3*) o PHT2 (OMIM \*610408) es un transportador presente en lisosomas y vesículas que se encarga del transporte de oligopéptidos e histidina al citosol de forma dependiente de protones. A pesar de no conocerse exactamente su función, esta proteína estaría relacionada con el sistema inmunitario, dado que está implicada en la regulación fagocítica en macrófagos y se expresa principalmente en el sistema linfático (Bissa *et al.*, 2016). Además se ha visto expresada en células dendríticas tras la estimulación de los TLR (*Toll-Like Receptors*), que median la salida de componentes derivados de bacterias, como NOD2 (Song *et al.*, 2018).

**Tabla 20. Variantes *missense* genotipadas en >70% en el gen *SLC15A3***

chr:position	rs	Variante		MAF	SIFT	S.S.	POLYPHEN	Casos			Controles		
								0/0	0/1	1/1	0/0	0/1	1/1
11:60705355	-	c.1571T>C	p.Leu524Pro		-	-	-	571	1	0	820	0	0
11:60705468	rs367924777	c.1465C>T	p.Arg489Cys	< 0.01	-	-	-	560	1	0	801	2	0
11:60705473	-	c.1460C>T	p.Ala487Val		Deleterious	-	Damaging	564	0	0	801	2	0
11:60706993	rs751726944	c.1394A>G	p.Tyr465Cys	< 0.01	Deleterious	-	Damaging	468	1	0	691	0	0
11:60708657	rs199509067	c.1213C>T	p.Pro405Ser	< 0.01	Deleterious	0,03	-	541	0	0	779	1	0
11:60708672	rs140945002	c.1198C>T	p.Arg400Trp	< 0.01	Deleterious	0	benign	545	2	0	786	1	0
11:60708696	rs140596697	c.1174C>T	p.Arg392Cys	< 0.01	Deleterious	0	-	567	0	0	808	2	0
11:60709520	rs78193365	c.1094G>T	p.Gly365Val	0.01	Tolerated	-	benign	460	14	0	683	28	0
11:60709584	-	c.1030C>T	p.Leu344Phe		Deleterious	0,01	-	512	1	0	739	0	0

Transcritos: ENST00000227880 y ENSP00000227880. S.S.: Sift Score.

Las variantes *missense* con efecto más deletéreo genotipadas en más del 70% de individuos no se encontraron sobrerrepresentadas significativamente en uno de los dos grupos (diabéticos/controles), por lo que no se pudo establecer una relación entre estas mutaciones en el gen *SLC15A3* y el riesgo o protección a padecer DM2. En el análisis estadístico por tipo de mutación se obtuvo un p-valor de 0,04 (<0,05) para el conjunto de variantes de *splicing*, dado que la frecuencia fue mayor en controles. Sin embargo, el efecto de estas variantes probablemente sería bajo, al localizarse a más de 3 nucleótidos del transcrito.

De entre las variantes genotipadas en <70% destacó la variante de ganancia de *STOP* rs116597710, previamente clasificada como variante de riesgo en el anterior estudio. Fue secuenciada en uno de los pacientes diabéticos y en ningún control. Esta variante presenta una

MAF en la población menor al 0,01, por lo que sería considerada una variante rara. El cambio nucleotídico C>T en la posición 835 del cDNA en el cromosoma 11 (11q12.2), origina un cambio de aminoácido de arginina a codón terminador en la posición 279 en la proteína (NM\_016582). La proteína SLC15A3 tiene 581 aminoácidos por lo que se produciría una proteína truncada aproximadamente a la mitad de su longitud, que podría significar una pérdida total o parcial de la función proteica. Este resultado podría contribuir a la validación de la variante, a pesar de que serían necesarios más estudios.

#### **4.6. Limitaciones**

Dada la elevada cantidad de información obtenida en este trabajo y el tiempo limitado, es necesario seguir analizando estos datos para completar el análisis, por lo que este proyecto se podría considerar un estudio preliminar. Un aspecto importante a posterior estudio son las variantes en el número de copias o CNV.

Otra de las limitaciones encontradas fue la baja calidad de las muestras, que implicó una cobertura poco uniforme y algunas regiones de interés en el ADN sin cubrir. En la puesta a punto se utilizaron muestras de una mayor calidad y se obtuvo una mejor cobertura (Fig. 24).

Por este mismo motivo, tras aplicar el filtro del 70% de individuos genotipados para cada variante, se descartaron 219 variantes (más de la mitad), que no habían sido secuenciadas en un número suficiente de muestras. Esto supone que las muestras volverán a ser secuenciadas para mejorar la cobertura y mejorar los resultados.

Otra de las limitaciones a la hora del análisis estadístico fue el trabajar con variantes raras, que presentan una frecuencia muy baja en la población. Se utilizó el test de Fisher al existir en algunos casos valores de 0 en uno de los dos grupos. Además, se tuvieron en cuenta en conjunto las variantes no clasificadas previamente como estadísticamente significativas que se presentaban solo en casos o controles, dado que al tratarse de variantes raras podrían tener efecto sobre el fenotipo.





**Figura 24. Representación de los *plots* de cobertura de los controles de validación (A) y de las muestras de la población (B).** A pesar de que las muestras de la población fueron secuenciadas a una profundidad mucho mayor de lecturas, cabría esperar una distribución general similar a la de las pruebas de validación. Sin embargo, debido a la mala calidad y la fragmentación del ADN de las muestras de la población, estas presentaron menos regiones de interés cubiertas.

## 5. CONCLUSIONES

1. En una población de 1415 individuos, 586 afectados de *diabetes mellitus* tipo 2 (DM2) y 829 controles, se identificaron mediante secuenciación por amplicones 372 variantes raras o poco frecuentes en 10 genes seleccionados de un estudio previo. 152 de estas variantes fueron genotipadas en más del 70% de la población, siendo la mayor parte variantes de tipo *missense*.
2. En el gen *SLC15A3* los resultados obtenidos en el presente estudio han contribuido a la validación de la variante de ganancia de *STOP* rs116597710, previamente clasificada como variante de riesgo a DM2, dado que fue secuenciada en un paciente diabético.
3. Entre las variantes estadísticamente significativas, se identificó una variante *missense* en el gen *OLFML1* que podría tener efecto protector a DM2 y una variante *missense* en el gen *ARHGEF38*, detectada en homocigosis en dos casos y con efecto dañino según los predictores, que podría suponer riesgo a DM2.
4. En el gen *AFMID* se relacionaron variantes *missense* deletéreas con un aumento del riesgo a DM2, mientras que en *SLC12A4* y *OLFML1* este tipo de variantes podría estar relacionado con un efecto protector.
5. El elevado número de nuevas variantes genéticas encontradas podría ayudar a explicar parte del componente genético de la DM2. Los resultados aportan nuevas variantes a estudio, que podrían estar relacionadas con la enfermedad.
6. Es necesaria la confirmación de estos hallazgos con la validación y replicación de los resultados encontrados en otras poblaciones, así como el diseño de experimentos funcionales.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- Alam, U., Asghar, O., Azmi, S., & Malik, R. A. (2014). *General aspects of diabetes mellitus. Handbook of Clinical Neurology* (1st ed., Vol. 126). Elsevier B.V. <http://doi.org/10.1016/B978-0-444-53480-4.00015-1>
- Baeza-rajá, B., Li, P., Le, N., Sachs, B. D., & Schachtrup, C. (2012). P75 Neurotrophin Receptor Regulates Glucose Homeostasis and Insulin Sensitivity. *Pnas*, *109*(15), 5838–5843. <http://doi.org/10.1073/pnas.1103638109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1103638109>
- Bissa, B., Beedle, A. M., & Govindarajan, R. (2016). Lysosomal solute carrier transporters gain momentum in research. *Clinical Pharmacology and Therapeutics*, *100*(5), 431–436. <https://doi.org/10.1002/cpt.450>
- Brylinski, M., & Skolnick, J. (2010). Structural and Functional Characterization of the Human Kinome, *271*(13), 1839–1854.
- Buzzetti, R., Zampetti, S., & Maddaloni, E. (2017). Adult-onset autoimmune diabetes: Current knowledge and implications for management. *Nature Reviews Endocrinology*, *13*(11), 674–686. <http://doi.org/10.1038/nrendo.2017.99>
- Care, D., & Suppl, S. S. (2018). Classification and diagnosis of diabetes: Standards of medical care in Diabetes2018. *Diabetes Care*, *41*(January), S13–S27. <http://doi.org/10.2337/dc18-S002>
- Center for Disease Control and Prevention; National Center for Health. (2017). National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States. Atlanta, GA: Centers for Disease Control and Prevention; 2014. *US Department of Health and Human Services*, (Cdc), 2009–2012. <http://doi.org/10.1177/1527154408322560>
- Chabot, C., Massicotte, G., Milot, M., Trudeau, F., & Gagné, J. (1997). Impaired modulation of AMPA receptors by calcium-dependent processes in streptozotocin-induced diabetic rats. *Brain Research*, *768*(1–2), 249–256. [http://doi.org/10.1016/S0006-8993\(97\)00648-3](http://doi.org/10.1016/S0006-8993(97)00648-3)
- Chen, X., Li, X., Chen, J., Zheng, P., Huang, S., & Ouyang, X. (2012). Overexpression of CIAPIN1 inhibited pancreatic cancer cell proliferation and was associated with good prognosis in pancreatic cancer. *Cancer Gene Therapy*, *19*(8), 538–544. <http://doi.org/10.1038/cgt.2012.28>
- Chiti, F., & Dobson, C. M. (2006). Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry*, *75*(1), 333–366. <http://doi.org/10.1146/annurev.biochem.75.101304.123901>
- Cobo-Vuilleumier, N., Lorenzo, P. I., Rodríguez, N. G., Herrera Gómez, I. D. G., Fuente-Martin, E., López-Noriega, L., ... Gauthier, B. R. (2018). LRH-1 agonism favours an immune-islet dialogue which protects against diabetes mellitus. *Nature Communications*, *9*(1), 1–15. <http://doi.org/10.1038/s41467-018-03943-0>

- Comité de la 8va edición de Diabetes Atlas. (2017). *Diabetes Atlas. Federación Internacional de Diabetes* (Vol. 8). <http://doi.org/2-930229-80-2>
- Crawford, K. (2017). Review of 2017 Diabetes Standards of Care. *Nursing Clinics of North America*, 52(4), 621–663. <http://doi.org/10.1016/j.cnur.2017.07.010>
- Czech, M. P. (2017). Insulin action and resistance in obesity and type 2 diabetes. *Nature Medicine*, 23(7), 804–814. <http://doi.org/10.1038/nm.4350>
- Dendup, T., Feng, X., Clingan, S., & Astell-Burt, T. (2018). Environmental risk factors for developing type 2 diabetes mellitus: A systematic review. *International Journal of Environmental Research and Public Health*, 15(1). <http://doi.org/10.3390/ijerph15010078>
- Favennec, M., Hennart, B., Caiazzo, R., Leloire, A., Yengo, L., Verbanck, M., ... Poulain-Godefroy, O. (2015). The kynurenine pathway is activated in human obesity and shifted toward kynurenine monooxygenase activation. *Obesity*, 23(10), 2066–2074. <http://doi.org/10.1002/oby.21199>
- Galan I (2015). Identificación de variantes genéticas poco frecuentes y raras en diabetes mellitus tipo 2 mediante secuenciación de exoma (Tesis doctoral). Fundación de Investigación Clínico de Valencia-INCLIVA, Valencia.
- Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at <http://arxiv.org/abs/1207.3907> (2012)
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <http://doi.org/10.1038/nrg.2016.49>
- Grarup, N., Sandholt, C. H., Hansen, T., & Pedersen, O. (2014). Genetic susceptibility to type 2 diabetes and obesity: From genome-wide association studies to rare variants and beyond. *Diabetologia*, 57(8), 1528–1541. <http://doi.org/10.1007/s00125-014-3270-4>
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1–8. <http://doi.org/10.1016/j.ygeno.2015.11.003>
- Hugill, A. J., Stewart, M. E., Yon, M. A., Probert, F., Cox, I. J., Hough, T. A., ... Cox, R. D. (2015). Loss of arylformamidase with reduced thymidine kinase expression leads to impaired glucose tolerance. *Biology Open*, 4(11), 1367–75. <http://doi.org/10.1242/bio.013342>
- Joladarashi, D., Salimath, P. V., & Chilkunda, N. D. (2011). Diabetes results in structural alteration of chondroitin sulfate/dermatan sulfate in the rat kidney: Effects on the binding to extracellular matrix components. *Glycobiology*, 21(7), 960–972. <http://doi.org/10.1093/glycob/cwr029>
- Kahn, S. E., Hull, R. L., & Utzschneider, K. M. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*, 444(7121), 840–846. <http://doi.org/10.1038/nature05482>

- Khadori *et al.* (2018). Type 2 Diabetes Mellitus: Practice Essentials, Background, Pathophysiology. Retrieved from <https://emedicine.medscape.com/article/117853-overview>
- Katane M., Saitoh K., Maeda K., Hanai T., Sekine M., Furuchi T., Homma H. (2011). Role of the active site residues arginine-216 and arginine-237 in the substrate specificity of mammalian D-aspartate oxidase. *Amino Acids* 40:467–476. <http://doi.org/10.1007/s00726-010-0658-4>.
- Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine*, 09(03). <http://doi.org/10.4172/0974-8369.1000395>
- Kolb, H., & Martin, S. (2017). Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. *BMC Medicine*, 15(1), 1–11. <http://doi.org/10.1186/s12916-017-0901-x>
- Kursan, S., McMillen, T. S., Beesetty, P., Dias-Junior, E., Almutairi, M. M., Sajib, A. A., ... Di Fulvio, M. (2017). The neuronal K+Cl-co-transporter 2 (Slc12a5) modulates insulin secretion. *Scientific Reports*, 7(1), 1–14. <http://doi.org/10.1038/s41598-017-01814-0>
- Li H.\*, Handsaker B.\*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078-9. [PMID: 19505943]
- Li, N., Young, M. M., Bailey, C. J., & Smith, M. E. (1999). NMDA and AMPA glutamate receptor subtypes in the thoracic spinal cord in lean and obese-diabetic ob/ob mice. *Brain Research*, 849(1–2), 34–44. [http://doi.org/S0006-8993\(99\)02070-3](http://doi.org/S0006-8993(99)02070-3) [pii]
- Lohmueller, K. E., Sparsø, T., Li, Q., Andersson, E., Korneliussen, T., Albrechtsen, A., ... Pedersen, O. (2013). Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *American Journal of Human Genetics*, 93(6), 1072–1086. <http://doi.org/10.1016/j.ajhg.2013.11.005>
- Ma, L., Hanson, R., Que, L. N., Cali, A. M. G., Fu, M., Mack, J. L., ... Baier, L. J. (2007). Variants in ARHGE11, a Candidate Gene for the Linkage to Type 2 Diabetes on Chromosome 1q, Are Nominally Associated With Insulin Resistance and Type 2 Diabetes in Pima Indians. *Diabetes*, 56, 1454–1459. <http://doi.org/10.2337/db06-0640>.Additional
- Medici, F., Hawa, M., Ianari, A., Pyke, D. A., & Leslie, R. D. G. (1999). Concordance rate for type II diabetes mellitus in monozygotic twins: Actuarial analysis. *Diabetologia*, 42(2), 146–150. <http://doi.org/10.1007/s001250051132>
- Meigs, J. B., Cupples, L. A., & Wilson, P. W. F. (2000). The Framingham Offspring Study. *World Health*, 57(23), 2201–2207. <http://doi.org/10.2337/db08-0700>.J.C.F.
- Mitchell, M. J., & King, M. R. (2014). NIH Public Access, 79(1), 1–23. <http://doi.org/10.1088/1367-2630/15/1/015008>.Fluid

- National Center for Biotechnology Information. Home - Gene - NCBI. (2018). Retrieved from <https://www.ncbi.nlm.nih.gov/gene/>
- Neef, D. W., Jaeger, A. M., Gomez-Pastor, R., Willmund, F., Frydman, J., & Thiele, D. J. (2014). A direct regulatory interaction between chaperonin TRiC and stress-responsive transcription factor HSF1. *Cell Reports*, 9(3), 955–966. <http://doi.org/10.1016/j.celrep.2014.09.056>
- Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), 29/08/2018. World Wide Web URL: <https://omim.org/>
- Paul, V. D., & Lill, R. (2015). Biogenesis of cytosolic and nuclear iron-sulfur proteins and their role in genome stability. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1853(6), 1528–1539. <http://doi.org/10.1016/j.bbamcr.2014.12.018>
- Pirola, L., & Ferraz, J. C. (2017). Role of pro- and anti-inflammatory phenomena in the physiopathology of type 2 diabetes and obesity. *World Journal of Biological Chemistry*, 8(2), 120. <http://doi.org/10.4331/wjbc.v8.i2.120>
- Poulsen, P., Ohm Kyvik, K., Vaag, A., & Beck-Nielsen, H. (1999). Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance - A population-based twin study. *Diabetologia*, 42(2), 139–145. <http://doi.org/10.1007/s001250051131>
- Prasad, R. B., & Groop, L. (2015). Genetics of type 2 diabetes—pitfalls and possibilities. *Genes*, 6(1), 87–123. <http://doi.org/10.3390/genes6010087>
- Reine, T. M., Grøndahl, F., Jenssen, T. G., Hadler-Olsen, E., Prydz, K., & Kolset, S. O. (2013). Reduced Sulfation of Chondroitin Sulfate but Not Heparan Sulfate in Kidneys of Diabetic db/db Mice. *Journal of Histochemistry and Cytochemistry*, 61(8), 606–616. <http://doi.org/10.1369/0022155413494392>
- Roshanravan, H., Kim, E. Y., & Dryer, S. E. (2016). NMDA receptors as potential therapeutic targets in diabetic nephropathy: Increased renal NMDA receptor subunit expression in akita mice and reduced nephropathy following sustained treatment with memantine or MK-801. *Diabetes*, 65(10), 3139–3150. <http://doi.org/10.2337/db16-0209>
- Santiago, a R., Gaspar, J. M., Baptista, F. I., Cristovao, a J., Santos, P. F., Kamphuis, W., & Ambrosio, a F. (2009). Diabetes changes the levels of ionotropic glutamate receptors in the rat retina. *Mol Vis*, 15, 1620–1630. Retrieved from [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=19693289](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19693289)
- Schofield, J. D., Liu, Y., Rao-Balakrishna, P., Malik, R. A., & Soran, H. (2016). Diabetes Dyslipidemia. *Diabetes Therapy*, 7(2), 203–219. <http://doi.org/10.1007/s13300-016-0167-x>
- Sengupta, S., Nechushtai, R., Jennings, P. A., Onuchic, J. N., Padilla, P. A., Azad, R. K., & Mittler, R. (2018). Phylogenetic analysis of the CDGSH iron-sulfur binding domain reveals its ancient origin. *Scientific Reports*, 8(1), 1–12. <http://doi.org/10.1038/s41598-018-23305-6>

- Song, F., Yi, Y., Li, C., Hu, Y., Wang, J., Smith, D. E., & Jiang, H. (2018). Regulation and biological role of the peptide/histidine transporter SLC15A3 in Toll-like receptor-mediated inflammatory responses in macrophage. *Cell Death and Disease*, 9(7), 1–15. <https://doi.org/10.1038/s41419-018-0809-1>
- 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. <http://doi.org/10.1007/s00125-011-2336-9>
- Stančáková, A., & Laakso, M. (2016). Genetics of type 2 diabetes. *Endocrine Development*, 31, 203–220. <http://doi.org/10.1159/000439418>
- Thrasher, J. (2017). Pharmacologic Management of Type 2 Diabetes Mellitus: Available Therapies. *American Journal of Medicine*, 130(6), S4–S17. <http://doi.org/10.1016/j.amjmed.2017.04.004>
- Upadhyay, J., Polyzos, S. A., Perakakis, N., Thakkar, B., Paschou, S. A., Katsiki, N., ... Mantzoros, C. S. (2018). Pharmacotherapy of type 2 diabetes: An update. *Metabolism: Clinical and Experimental*, 78, 13–42. <http://doi.org/10.1016/j.metabol.2017.08.010>
- van Det, N. F., van den Born, J., Tamsma, J. T., Verhagen, N. a, Berden, J. H., Bruijn, J. a, ... van der Woude, F. J. (1996). Effects of high glucose on the production of heparan sulfate proteoglycan by mesangial and epithelial cells. *Kidney International*, 49(4), 1079–89. <http://doi.org/10.1038/ki.1996.157>
- Zhang, D., Wang, F., Lal, N., Chiu, A. P. L., Wan, A., Jia, J., ... Rodrigues, B. (2017). Heparanase overexpression induces glucagon resistance and protects animals from chemically induced diabetes. *Diabetes*, 66(1), 45–57. <http://doi.org/10.2337/db16-0761>
- Zhao, T., Lu, X., Davies, N. M., Gong, Y., Guo, J., Zhang, H., ... Li, P. (2013). Diabetes results in structural alteration of chondroitin sulfate in the urine. *Journal of Pharmacy and Pharmaceutical Sciences*, 16(3), 486–493.