



VNIVERSITAT E VALÈNCIA

Identificación de variantes exómicas en diabetes tipo 2

Tesis doctoral presentada por:

Vanesa Martínez Barquero

Dirigida por:

Dr. Felipe Javier Chaves

Dr. Juan Francisco Ascaso

Dra. M^a Antonia Noguera

DOCTORADO EN BIOMEDICINA Y FARMACIA

Valencia, Mayo de 2015



VNIVERSITAT
ID VALÈNCIA



Fundación Investigación Clínica de Valencia
Instituto de Investigación Sanitaria – INCLIVA

El Dr. **Juan Francisco Ascaso**, catedrático del Departamento de Medicina de la Universidad de Valencia.

El Dr. **Felipe Javier Chaves**, Investigador y director de la Unidad de Genotipado y Diagnóstico Genético. Fundación Hospital Clínico de Valencia-INCLIVA.

La Dra. **M^a Antonia Noguera**, Profesora titular del Departamento de Farmacología de la Universidad de Valencia.

CERTIFICAN:

Que la memoria de Tesis Doctoral realizada por D^a Vanesa Martínez Barquero, titulada “Identificación de variantes exómicas en diabetes tipo 2”, ha sido realizada bajo su dirección y reúne todos los requisitos necesarios para su juicio y calificación.

En Valencia, marzo de 2015.

JF. Ascaso

FJ. Chaves

MA. Noguera

“La alegría está en la lucha, en el esfuerzo, en el sufrimiento que supone la lucha, y no en la victoria misma. Nuestra recompensa se encuentra en el esfuerzo y no en el resultado. Un esfuerzo total es una victoria completa.”

Gandhi.

A MIS ABUELOS, Luis, Matilde y Ana.

Agradecimientos

La realización de una tesis siempre es un camino largo y complicado, pero también hay momentos buenos y personas que están ahí, haciendo el camino más fácil.

Empezando por lo profesional, quiero agradecer a mis directores, el Dr. Ascaso y Javier, su apoyo en estos años, sus consejos y su sabiduría. A Javier, también su paciencia infinita en muchos momentos y el intentar encontrar siempre soluciones a todo.

A Marian, por ser codirectora de esta tesis, pero también, por confiar en mí desde el principio, y haberme dado la oportunidad de empezar de cero en la investigación.

No puedo olvidarme de Consuelo Guerri, por estar en un momento muy necesario para mí y haberme ayudado de corazón.

También recordar a todos mis compañeros del Laboratorio de Patología Autoinmune del CIPF, Carol, Sandra, Héctor, Marcos, Natalia, Jesús, Ernesto, M^a Carmen, Filo, Fernando, Paco, Zahara, Raul y Carmen. Aprendí mucho de vosotros tanto en lo profesional como en lo personal y viví muchos momentos que nunca olvidaré, gracias por todos esos buenos ratos!

Tras salir del CIPF, llegué a INCLIVA, a la UGDG, un mundo nuevo para mí, allí encontré otros compañeros fantásticos, de los que también he aprendido mucho y con los que he compartido estos años de tesis en que ha habido grandes momentos: Tani, Griselda, Pilar, Ana Bárbara, Verónica Navarro, Laura, Kike, Julia, Elena, Veronik, M^a Luisa, YoVero, Raquel. Mención especial para mis mamis: Cris y Victoria, para mis exomitas: Javi e Inma y para mi peque, Lucía. También agradecer a la unidad de Bioinformática, Dani, Pablo y Alba, su ayuda en la realización de esta tesis. Gracias a todos por vuestra ayuda y apoyo en el día a día durante estos años.

Llegaron más tarde INCLIVA, pero no por ello son menos importantes, María y como no, mis chicos, Alex y Miguel, que siempre están ahí para sacarme una sonrisa y ayudarme en lo que haga falta.

En este tiempo, también he tenido la oportunidad de dar clases en el grado de Nutrición Humana y Dietética, colaborando con Carlos Sánchez. Gracias Carlos por confiar en mí y hacerme ver que podía hacerlo.

En lo personal, agradecer a mis padres su apoyo. En especial, gracias a mi hermana Cintia, por estar siempre ahí para todo y por ser mi mejor amiga. También a mi tío Jose Luis, por ser el mejor tío del mundo. Os quiero mucho.

A mis amigas con mayúsculas, las que nunca fallan, Alba y Sandra. A mi French Club: Sara, Carmen y Carol. Y también a las últimas incorporaciones, Cris y Toni. Todos me habéis hecho pasar muy buenos ratos y también sé que os tengo para lo que necesite.

Para mí poder hacer la tesis era uno de mis sueños, pero lo mejor ha sido poder compartirlo con todos vosotros. GRACIAS!

Abreviaturas

ADN: Ácido desoxirribonucleico.

ARN: Ácido ribonucleico.

ATP: Adenosín trifosfato.

BWA: Alineador Burrows-Wheeler.

CNV: Variación en el número de copias.

ddNTP: didesoxinucleótido.

DM1: Diabetes mellitus tipo 1.

DM2: Diabetes mellitus tipo 2.

DMG: Diabetes mellitus gestacional.

DMN: Diabetes mellitus neonatal.

DMNP: Diabetes mellitus neonatal permanente.

DMNT: Diabetes mellitus neonatal transitoria.

FID: Federación Internacional de Diabetes.

GABA: Ácido gamma-aminobutírico

GAD: Glutamato descarboxilasa.

GWAS: Estudio de asociación de genoma completo.

IFN: Interferón.

IMC: Índice de masa corporal.

IR: Receptor de insulina.

IRS: Sustrato del receptor de insulina.

LADA: Latent autoimmune diabetes in adults.

MAF: Frecuencia del alelo minoritario.

MODY: Maturity-onset diabetes of the young.

NEFA: Ácidos grasos libres no esterificados.

NFkB: Factor nuclear kappa B.

NGS: Secuenciación de nueva generación.

OMS: Organización Mundial de la Salud.

OR: Odds ratio.

PCR: Reacción en cadena de la polimerasa.

PF: Polimorfismo frecuente.

PPF: Polimorfismo poco frecuente.

RI: Resistencia a la insulina.

SBS: Secuenciación por síntesis.

SEEDO: Sociedad Española para el Estudio de la Obesidad.

SNP: Polimorfismo de un solo nucleótido.

SNV: Variación de un solo nucleótido.

SOG: Sobrecarga oral de glucosa.

STR: Repeticiones cortas en tándem.

VCF: Archivo de llamado de variantes.

VEP: Predictor del efecto de las variantes.

Índice

INTRODUCCIÓN	3
<i>1-TIPOS DE DIABETES</i>	<i>4</i>
1.1- PRIMARIAS	5
1.1.1- DIABETES MONOGÉNICAS.....	5
Diabetes mellitus neonatal (DMN).....	5
MODY (“Maturity-onset diabetes of the young”).....	6
Diabetes mitocondrial (MIDD: Maternally Inherited Diabetes and Deafness).	7
1.1.2-DIABETES SINDRÓMICAS.....	7
1.1.3-DIABETES MULTIFACTORIALES:.....	9
DIABETES MELLITUS TIPO 1 (DM1).....	9
LADA (“Latent autoimmune diabetes in adults”).....	11
DIABETES GESTACIONAL	13
1.2- SECUNDARIAS	15
<i>2-DIABETES MELLITUS TIPO 2.....</i>	<i>17</i>
2.1- ASPECTOS CLÍNICOS Y EVOLUTIVOS DE LA DIABETES TIPO 2.	20
2.2-DIAGNÓSTICO DE LA DIABETES TIPO 2	22
2.3- FACTORES DE RIESGO Y PREVENCIÓN DE LA DIABETES TIPO 2.	23
2.4- PATOGÉNESIS DE LA DIABETES TIPO 2.....	27
2.4.1- Disfunción de las células β	27
Secreción de insulina en condiciones normales.....	27
Fisiopatología de la hiperglucemia.....	29
Toxicidad de la glucosa.....	30
Lipotoxicidad	31
Depósitos de polipéptido amiloide	32
2.4.2- Resistencia a la insulina.....	33
Obesidad	34
Mecanismos celulares.....	35
La fosforilación y desfosforilación de proteínas IRS.....	36
Papel de los productos de los adipocitos y la inflamación.....	37
Actividad NF κ B e IKK	38
Metabolismo mitocondrial.....	39

2.5- BASES GENÉTICAS DE LA DM2	41
3- SECUENCIACIÓN DE NUEVA GENERACIÓN.....	48
3.1- ROCHE 454	49
3.2- ILLUMINA	51
3.3- SOLID.....	54
3.4- ION TORRENT	56
3.5- TECNOLOGÍAS EMERGENTES	57
3.6-COMPARACIÓN DE LAS DIFERENTES PLATAFORMAS DE SECUENCIACIÓN.	62
4- NUEVAS ESTRATEGIAS DE ESTUDIO DE LA DIABETES MELLITUS TIPO 2	64
4.1- IDENTIFICACIÓN DE VARIANTES GENÉTICAS EN EL EXOMA	64
4.2- ESTUDIO DE POLIMORFISMOS DE BAJA FRECUENCIA Y VARIANTES RARAS	65
4.3- SELECCIÓN DE CASOS Y CONTROLES MEDIANTE NUEVOS CRITERIOS.....	67
HIPÓTESIS Y OBJETIVOS	71
MATERIAL Y MÉTODOS	75
1- SELECCIÓN DE MUESTRAS.....	75
2- PROCESADO DE LAS MUESTRAS.....	77
2.1-PREPARACION DE LIBRERIAS DE ADN	77
2.1.1-Cuantificacion de ADN	79
2.1.2-Fragmentacion de ADN.....	79
2.1.3-Reparación de los extremos.....	80
2.1.4-Adenilación	80
2.1.5-Ligación de los adaptadores.....	81
2.1.6-PCR de enriquecimiento	82
2.1.7- Puntos de control.....	82
2.2- ENRIQUECIMIENTO: EXOMA.....	84
2.3-VALIDACIÓN DE LOS POOLS	86
2.3.1- Visualización de los pools.....	86
2.3.2- Selección por tamaños.....	86
2.3.3- Cuantificación de los pools mediante PCR cuantitativa.....	87
2.4- AMPLIFICACIÓN CLONAL DE LOS POOLS: cBOT	88
2.4.1- Inmovilización de las moléculas de ADN.....	89
2.4.2- Amplificación isotérmica en puente.	89
2.4.3- Linealización, bloqueo e hibridación con los primers.	89
2.5- SECUENCIACIÓN: HiScanSQ y HiSeq 2000.	90
2.5.1- Secuenciación por síntesis.	90
2.5.2- HiScanSQ y HiSeq 2000.	94

3- ANÁLISIS BIOINFORMÁTICO	97
3.1- Alineamiento y mapeo de las secuencias: Paso de FastQ a SAM (Sequence Alignment/Map format).....	98
3.1.1- Análisis de la calidad de los archivos FastQ: FastQC.....	98
3.1.2 - Procesado de los archivos fastQ	98
3.1.3- Mapeo	99
3.2- Llamado de variantes.....	100
3.2.1- Obtención del archivo VCF a través de SAMtools y BCFtools.....	100
3.2.2- Tratamiento y comparación de VCFs	100
3.3- Anotación funcional de los SNPs detectados.....	102
3.4- Genes analizados.....	105
4- VALIDACIÓN DE LOS RESULTADOS OBTENIDOS POR SECUENCIACIÓN DE SANGER	106
RESULTADOS	111
1- SELECCIÓN DE GENES A ANALIZAR.....	111
1.1- Listado de genes seleccionados.....	112
2- RESULTADOS RELATIVOS A LA COBERTURA.....	121
2.1- Cobertura por exón.....	121
2.2- Porcentaje de bases con una cobertura mínima	124
3- POSIBLES CONSECUENCIAS FUNCIONALES DE LAS VARIACIONES OBTENIDAS.....	127
3.1- Consecuencias más importantes	129
4- EFECTOS FUNCIONALES DE LAS VARIACIONES.....	131
4.1- Según SIFT.	131
4.2- Según PolyPhen.....	132
5- SELECCIÓN DE VARIANTES MÁS RELEVANTES	133
5.1- Variaciones seleccionadas en controles	134
5.2- Variaciones seleccionadas en casos.....	136
6- VALIDACIÓN DE RESULTADOS POR SECUENCIACIÓN CAPILAR. COMPROBACIÓN DE GENOTIPOS SANGER-ILLUMINA.	138
6.1- Validación variaciones seleccionadas en los controles.....	139
6.2- Validación variaciones seleccionadas en los casos.	141
DISCUSIÓN	145
1-CARACTERÍSTICAS DE LA SELECCIÓN DE MUESTRAS.....	147
2-ESTUDIO DE VARIANTES RARAS	147
3-SECUENCIACIÓN DE NUEVA GENERACIÓN Y ESTUDIO DEL EXOMA	149
4-ANÁLISIS BIOINFORMÁTICO Y COBERTURA EMPLEADA	151
5-LIMITACIONES DEL ESTUDIO	152

6-TIPOS DE VARIANTES IDENTIFICADAS Y VERIFICADAS	154
7-VARIACIONES IDENTIFICADAS EN LOS GENES ESTUDIADOS	155
7.1-VARIACIONES IDENTIFICADAS Y VALIDADAS EN CONTROLES	155
Gen GCKR (“glucokinase (hexokinase 4) regulator”)	156
Gen THADA (“thyroid adenoma associated”)	159
Gen ITGB6 (“integrin, beta 6”)	161
Gen IFIH1 (“interferon induced with helicase C domain 1”).....	162
Gen TCF7L2 (“Transcription factor 7-like 2”)	164
Gen ABCC8 (“ATP-binding cassette, sub-family C (CFTR/MRP), member 8”)	167
Gen ERBB3 (“v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3”).....	169
Gen HMGA2 (“High mobility group AT-hook 2”)	171
7.2-VARIACIONES IDENTIFICADAS Y VALIDADAS EN CASOS	173
Gen NOTCH2 (“notch 2”).....	173
Gen THADA (“thyroid adenoma associated”)	175
Gen IFIH1 (“interferon induced with helicase C domain 1”).....	176
Gen AGMO (“alkylglycerolmonooxygenase”)	177
Gen CLEC7A (“C-type lectin domain family 7, member A”)	179
Gen LGR5 (“leucine-rich repeat containing G protein-coupled receptor 5”)	182
Gen OASL (“2'-5'-oligoadenylate synthetase-like”).....	183
Gen PAX4 (“paired box 4”).....	186
CONCLUSIONES	191
BIBLIOGRAFÍA	197
ANEXO 1	247

Introducción

INTRODUCCIÓN

La diabetes mellitus es un grupo de enfermedades metabólicas caracterizadas por una hiperglucemia debida a anomalías en la secreción de insulina, en su acción o en ambas. La hiperglucemia crónica de la diabetes está asociada a lesiones, disfunción y fallo de varios órganos, especialmente de ojos, riñón, nervios, corazón y vasos sanguíneos. En el desarrollo de la diabetes están implicados varios procesos patogénicos, que van desde la destrucción de las células β con déficit de insulina a aquellos que ocasionan resistencia a la insulina y disfunción de las células β pancreáticas, con la consiguiente deficiencia de acción de la insulina.

La diabetes mellitus es una de las principales causas de morbilidad y mortalidad en el mundo debido a las complicaciones cardiovasculares que produce. Entre éstas se encuentran: enfermedad coronaria (infarto, arritmias, muerte súbita), enfermedad isquémica cerebral, arteriosclerosis periférica de miembros inferiores y aneurisma abdominal, entre las principales complicaciones. Otras complicaciones crónicas de la diabetes están relacionadas con la microangiopatía (retinopatía y nefropatía) y con la neuropatía diabética (periférica y autonómica) (Marcovecchio et al, 2005). Los cálculos más recientes de la Federación Internacional de Diabetes (FID) se muestran en la figura 1 e indican que el 8,3% de adultos (382 millones de personas) padecen diabetes, y que el número de personas con esta enfermedad se incrementará en más de 592 millones en menos de 25 años. Además, los 175 millones de casos no diagnosticados actualmente suponen que un gran número de personas con diabetes van a desarrollar progresivamente complicaciones de las que no son conscientes. Por otra parte, el 80% de los afectados vive en países de ingresos medios y bajos, donde el aumento de casos se está acelerando a un ritmo alarmante. En otros países como España, el problema podría ser incluso más grave ya que se calcula que la prevalencia de la enfermedad es aproximadamente del 14% (Soriguer et al, 2012).

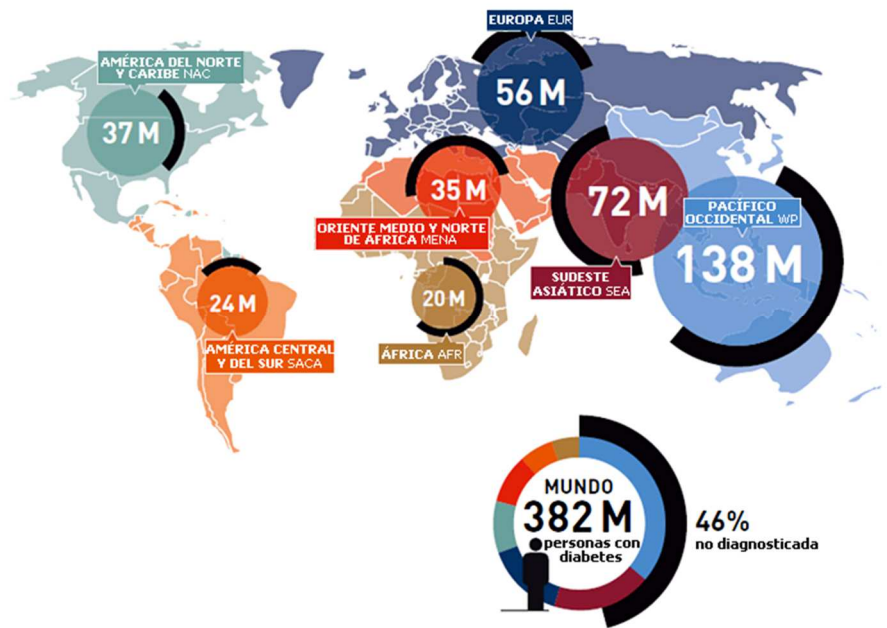


Figura 1: Número de personas con diabetes según la Federación Internacional de Diabetes (FID). *International Diabetes Federation, 2013.*

En los últimos años existe un problema creciente, causado por cambios en el estilo de vida (principalmente relacionados con la alimentación y la actividad física) que conducen a un mayor índice de masa corporal (IMC) y, por tanto a un aumento de la obesidad de la población (Kopelman et al, 2000; Lyssenko et al, 2008). Este incremento de la obesidad, fundamentalmente de la obesidad abdominal, es el principal causante del aumento de la incidencia de la diabetes mellitus tipo 2 (DM2).

1-TIPOS DE DIABETES

Existen diferentes tipos de diabetes y, en ocasiones, los mismos genes pueden estar implicados en varios de ellos. Los principales tipos se recogen en la figura 2 siendo la más frecuente la DM2, que se detallará en el apartado 2:

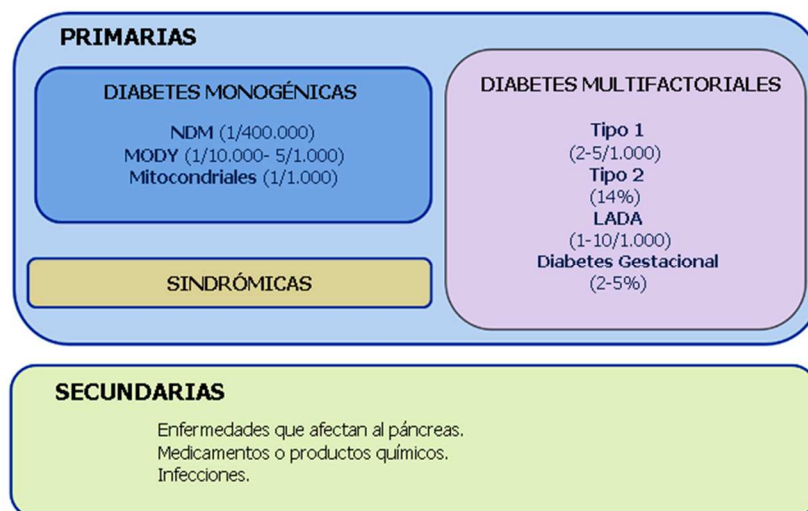


Figura 2: Diferentes tipos de diabetes.

Además de los diferentes tipos de diabetes recogidos en el esquema anterior, es importante definir también el concepto de prediabetes. La prediabetes es un trastorno en que el nivel de glucosa en sangre es mayor de lo normal, pero no lo suficientemente alto como para diagnosticar diabetes. La presencia de prediabetes supone un mayor riesgo de padecer diabetes tipo 2 con el tiempo.

1.1- PRIMARIAS

1.1.1- DIABETES MONOGÉNICAS

Diabetes mellitus neonatal (DMN)

La diabetes mellitus neonatal (DMN) es una forma de diabetes monogénica no autoinmune, poco frecuente, que afecta a 1 de cada ~400.000 nacimientos, y se diagnostica antes de los seis meses de edad. Existen dos tipos: transitoria (DMNT) y permanente (DMNP) según sea o no temporal, respectivamente. La etiología genética de este tipo de diabetes es bastante heterogénea. La mayoría de los casos de DMNT presentan una anomalía en el cromosoma 6q24 (Flanagan et al, 2007). Otras causas frecuentes de DMN son mutaciones sin sentido en los genes de los canales de potasio (KATP) de las células β pancreáticas, “potassium channel, inwardly rectifying subfamily J, member 11” (KCNJ11) y “ATP-binding cassette, sub-family C (CFTR/MRP), member 8”

Introducción

(ABCC8) y en el gen de la preproinsulina. La DMN se ha relacionado también con mutaciones puntuales en los genes: “glucokinase (hexokinase 4)” (GCK), “GLIS family zinc finger 3” (GLIS3), “eukaryotic translation initiation factor 2-alpha kinase 3” (EIF2AK3), “pancreatic and duodenal homeobox 1” (PDX1), “pancreas specific transcription factor, 1a” (PTF1A), “solute carrier family 2 (facilitated glucose transporter), member 2” (SLC2A2), “HNF1 homeobox B” (HNF1B) o “forkhead box P3” (FOXP3) (Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009).

Aunque ciertas características clínicas específicas (agenesia de páncreas, retraso en el desarrollo, insuficiencia renal, anemia, enfermedad de la glándula tiroides, trastornos cardíacos, etc) o una historia familiar de diabetes, pueden sugerir una etiología molecular para la DMN, el diagnóstico genético es crucial, ya que puede predecir el tratamiento más adecuado y mejorar la calidad de vida de estos pacientes. El ejemplo más notable se observa en los pacientes DMN con una mutación en los genes de los canales de potasio, ya que pueden ser tratados eficazmente con sulfonilureas orales, debido a que éstas se unen directamente a la subunidad reguladora SUR1 del canal, en lugar de requerir terapia de insulina de por vida (Gloyn et al, 2004; Babenko et al, 2006; Pearson et al, 2006).

MODY (“Maturity-onset diabetes of the young”)

MODY es una forma no autoinmune de diabetes de aparición temprana, que se transmite de forma autosómica dominante. MODY representa menos del 1-2% de casos de diabetes no autoinmunes y suele diagnosticarse en la infancia o en la juventud. Este trastorno monogénico se debe a la disfunción primaria de las células β pancreáticas y no suele estar asociado a la obesidad, a diferencia de lo que ocurre en la diabetes tipo 2 (Bonfond et al, 2010).

Se conocen al menos doce subtipos de MODY con etiologías genéticas distintas. Todos ellos han sido descritos en la literatura (Molven et al, 2011): MODY1-HNF4A, MODY2-GCK, MODY3-HNF1A, MODY4-PDX1, MODY5-HNF1B, MODY6-NeuroD1, MODY7-KLF11, MODY8-CEL, MODY9-PAX4, MODY10-INS, MODY11-BLK y, muy recientemente,

MODY12-ABCC8 (Bonfond et al, 2010; Fajans et al, 2011; Bowman et al, 2012). Estas diferentes etiologías genéticas se asocian con diferencias sustanciales en el curso clínico de la enfermedad, principalmente en términos de edad de inicio y niveles de hiperglucemia, explicando así diferentes respuestas al tratamiento (Vaxillaire et al, 2006; Fajans et al, 2011). Por lo tanto, el diagnóstico molecular temprano es crucial, ya que conduce al tratamiento adecuado y a la estimación del riesgo para los familiares asintomáticos (Vaxillaire et al, 2006).

Diabetes mitocondrial (MIDD: Maternally Inherited Diabetes and Deafness).

La diabetes mitocondrial es una forma “rara” o poco común de diabetes que representa hasta el 1% de todos los casos de diabetes en europeos y que se debe a defectos en el ADN mitocondrial (ADNmt). Además de la transmisión materna, sus características clínicas son principalmente la sordera neurosensorial, seguida de otros trastornos mitocondriales, miopatías y distrofia macular. A menudo MIDD se diagnostica como diabetes tipo 1, diabetes tipo 2 u otro tipo de diabetes monogénica (Maassen et al, 2006; Murphy et al, 2008). La ausencia de autoinmunidad y obesidad y la presencia de heredabilidad materna permiten distinguir este tipo de diabetes del resto (Murphy et al, 2008).

Además de la mutación del ADNmt 3243A>G, cuya importancia funcional ha sido ya evaluada (Maassen et al, 2004), existen otras variantes del ADNmt que se han asociado con un fenotipo propio de MIDD (Maechler et al, 2001; Crispim et al, 2008). Sin embargo, pocos estudios han explorado los cambios que las variantes detectadas en el ADNmt pueden producir en la eficiencia mitocondrial (Mariotti et al, 1994; Malfatti et al, 2007). En consecuencia, el significado patogénico de muchas variantes recientemente identificadas no está todavía establecido.

1.1.2-DIABETES SINDRÓMICAS

Se denominan diabetes sindrómicas, aquellas que se encuentran asociadas a algún tipo de síndrome. Son ejemplos los que se muestran en la tabla 1.

Síndrome	Herencia	Cromosoma	Gen	Características clínicas	Tipo de diabetes
Wolfram	Autosómica recesiva	4p16	WFS1	Diabetes con comienzo en la infancia, atrofia óptica, sordera, diabetes insípida	Insulino-deficiente
Wolcott-Rallison	Autosómica recesiva	2p12	EIF2AK3	Baja estatura, displasia epifisaria, complicaciones renales y hepáticas	Insulino-dependiente
Rabson-Mendenhall	Autosómica recesiva	19p13.2	INSR	Acantosis, alargamiento fállico, hirsutismo, hiperpigmentación	Insulino-resistente
Bardet-Biedl	Autosómica recesiva	16q,11q13,20q12, 3p13,15q22,2q31	-	Retraso en el desarrollo, retinopatía pigmentaria, polidactilia, obesidad, hipogonadismo	Insulino-resistente
Laurence-Moon	Autosómica recesiva	-	-	Retraso en el desarrollo, retinopatía pigmentaria, hipogonadismo, paraplejía espástica	Insulino-resistente
Alstrom	Autosómica recesiva	2p13	-	Retinopatía pigmentaria, sordera, obesidad, cardiomiopatía dilatada	Insulino-resistente
Prader-Willi	Autosómica dominante	15q12	-	Retraso en el desarrollo, hipotonía, hipogonadismo, baja estatura, obesidad, manos y pies pequeños	Insulino-resistente

Tabla 1: Desórdenes hereditarios que cursan con diabetes. Adaptado de Barrett TG. (2001). *Best Pract Res Clin Endocrinol Metab.* 15, 325-343.

1.1.3-DIABETES MULTIFACTORIALES:

DIABETES MELLITUS TIPO 1 (DM1)

La diabetes mellitus tipo 1 representa el 5-10% de todos los casos de diabetes mellitus. Se caracteriza por una severa deficiencia de insulina debida a la destrucción autoinmune de las células β pancreáticas. En consecuencia, los pacientes con DM1 dependen durante toda su vida, de la administración de insulina (Pirrot et al, 2008; Eizirik et al, 2009). Los marcadores de la destrucción autoinmune de las células β incluyen autoanticuerpos contra la insulina, los islotes de Langerhans, la descarboxilasa del ácido glutámico o glutamato descarboxilasa (GAD-65), el transportador de zinc 8 (ZnT8 o SLC30A8), y las tirosina fosfatasas (IA-2 y IA-2 β). Al menos uno (pero por lo general más) de estos anticuerpos están presentes inicialmente en el 85-90% de los pacientes cuando se detecta la hiperglucemia en ayunas.

La glutamato descarboxilasa es la enzima responsable de la síntesis del ácido gamma-aminobutírico (GABA), un importante neurotransmisor inhibitorio que se encuentra en las neuronas y en las células β pancreáticas. Las dos isoformas de la glutamato descarboxilasa son conocidas como GAD-65 y GAD-67. Ambas isoformas están presentes tanto en el cerebro como en el páncreas. El transportador de zinc 8, es un transportador implicado en la acumulación de zinc en vesículas intracelulares. Se expresa fuertemente en el páncreas, en particular en los islotes de Langerhans y colocaliza con la insulina en los gránulos de las células secretoras de insulina, INS-1. Similar a los GAD, IA-2 se expresa en tejido nervioso y otros tejidos endocrinos y también, en las células α y β de los islotes pancreáticos. En los individuos con diabetes tipo 1 es más frecuente encontrar anticuerpos contra IA-2 que contra IA-2 β .

La inflamación de los islotes de Langerhans (insulitis) se desarrolla probablemente en el contexto de un "diálogo" entre las células inmunes y las células β . Este cuadro de diálogo está mediado en parte por citocinas y quimiocinas, que se liberan tanto por células inmunitarias y por las células β , como por otras señales inmunogénicas liberadas al morir las células β . Esto puede conducir al comienzo y amplificación del proceso inflamatorio, pero en algunos casos, conduce a la resolución de la insulitis (Eizirik et al, 2009). El curso de la inflamación de las células β y su progresión a la DM1 clínica depende de la compleja

Introducción

interacción entre un fuerte componente genético y gran variedad de factores ambientales desencadenantes (Kim et al, 2005; Jahromi et al, 2007).

Entre los diversos loci asociados con DM1, determinados alelos del complejo mayor de histocompatibilidad de clase II (moléculas implicadas en la presentación de antígenos a los linfocitos T y en la diferenciación entre lo propio y lo ajeno por parte del sistema inmunitario) localizado en el cromosoma 6p21 son, con diferencia, el mayor factor de riesgo genético que conduce a la DM1, representando así el 30-50% de riesgo genético para la enfermedad (Steck et al, 2011). Otras variantes genéticas están asociadas con el riesgo de padecer DM1, pero con efectos relativamente menores, por ejemplo: “insulin gene” (INS), “cytotoxic T-lymphocyte associated protein 4 gene” (CTLA4), “protein tyrosine phosphatase nonreceptor type 22 gene” (PTPN22), “interleukin 2 receptor alpha gene” (IL2RA), “interferon induced with helicase C domain 1 gene” (IFIH1) y otros genes recientemente descubiertos por medio de GWAS (Genome-Wide Association Studies) o estudios de asociación de genoma completo. A continuación, en la figura 3 se reflejan algunos de estos genes.

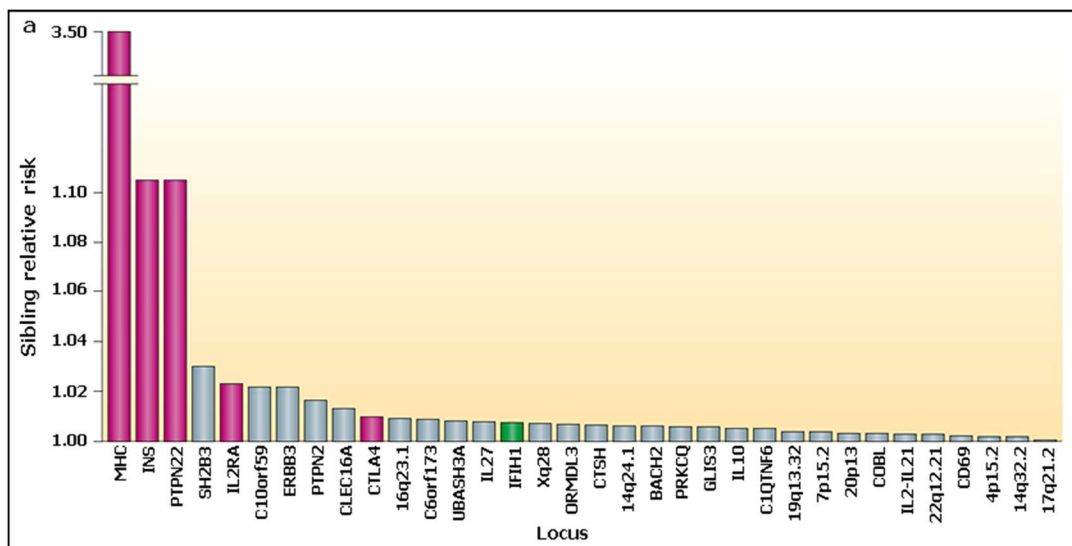


Figura 3: Principales genes implicados en DM1. Polychronakos y Li. (2011). *Nat Rev Genet.* 18, 781-792.

En cuanto a los desencadenantes ambientales que influyen en la destrucción autoinmune de las células β , éstos son entre otros, las infecciones virales, algunas vacunas, y ciertas toxinas (Knip et al, 2005; Pearl-Yafe, 2007). Hay evidencias de que algunos patógenos virales: enterovirus, virus de la rubéola, virus de las paperas, rotavirus, parvovirus y citomegalovirus, juegan un papel importante en el desencadenamiento de la destrucción autoinmune de las células β pancreáticas (Filippi et al, 2008; Hober et al, 2010). Entre estos virus, destacan los enterovirus, ya que muestran tropismo específico por el páncreas y han sido asociados con el desarrollo de la DM1 en humanos (Jaidane et al, 2008; Tauriainen et al, 2011).

LADA (“Latent autoimmune diabetes in adults”)

LADA es un tipo de diabetes autoinmune que se asemeja a la diabetes tipo 1 pero que parece tener componentes típicos de la diabetes tipo 2. En los estadíos iniciales se puede confundir con diabetes tipo 2 (Tuomi et al, 1993; Gale, 2005; Rolandsson et al, 2010). Clínicamente, se definen los pacientes con LADA, como el subgrupo de pacientes adultos con un fenotipo de diabetes tipo 2 y que dan positivo para el anticuerpo GAD o antiglutamato descarboxilasa (Zimmet, 1995). Debido a que las manifestaciones clínicas de LADA incluyen rasgos de las diabetes mellitus tipo 1 y tipo 2, se ha descrito esta condición como diabetes tipo 1.5 (Scherthaner et al, 2001). LADA representa el 2-12% de todos los casos de diabetes (Stenstrom et al, 2005).

Los pacientes con LADA presentan autoinmunidad y disfunción de las células β mediada por el sistema inmune. La progresión a la dependencia de la insulina en pacientes LADA, es más rápida que para los pacientes con diabetes mellitus tipo 2 clásica, negativos para autoanticuerpos y que no presentan daño progresivo en las células β (Ekholm et al, 2012). Sin embargo, la patogénesis de LADA no está claramente definida, y los criterios para el diagnóstico de la enfermedad varían entre los diferentes estudios. Por ello, la prevalencia de pacientes de LADA varía desde el 2,8% hasta el 22,3% en diferentes estudios publicados (Turner et al, 1997), y el 8-10% de los pacientes diagnosticados de diabetes tipo 2 están, de hecho, mal diagnosticados. Por esta razón, los esfuerzos para

Introducción

establecer una estrategia de tratamiento específico y una detección temprana son objeto de numerosas investigaciones.

En cuanto al componente genético de esta enfermedad, el locus en que se encuentra el complejo mayor de histocompatibilidad (brazo corto del cromosoma 6) que confiere la mayor parte de la susceptibilidad genética a la diabetes tipo 1 (Kantarova et al, 2007), ha mostrado asociaciones similares con LADA (Horton et al, 1999; Desai et al, 2007), pero también claras diferencias, por ejemplo, DQB1 * 0201 / * 0302 es un genotipo más común en la diabetes tipo 1 que en LADA, mientras que los genotipos de protección * 0602 / X y 0603 / X, son más comunes en LADA que en diabetes tipo 1 (Tuomi et al, 1999). El número variable de repeticiones en tándem en el gen de la insulina (INS VNTR) situado en el cromosoma 11 se divide en dos clases: clase I (26-63 repeticiones) y clase III (141-209 repeticiones) y la forma más corta ha mostrado una fuerte susceptibilidad a padecer diabetes tipo 1 (Bennett et al, 1996), y también LADA, tal y como se muestra en el "U.K. Prospective Diabetes Study" (Desai et al, 2006). Sin embargo, no se observó asociación en el estudio finlandés Botnia (Tuomi et al, 1999).

En un estudio realizado por Cervin et al (2008) se buscaba comparar las características genéticas de DM1, DM2 y LADA. Este estudio se realizó en 361 pacientes de LADA, 718 diabéticos tipo 1 y 1676 diabéticos tipo 2, así como en 1.704 controles sanos procedentes de Suecia y Finlandia. Se estudiaron: el locus HLA, INS VNTR y el gen PTPN22, para comparar con DM1 y, el gen TCF7L2 como referencia para comparar con DM2. La figura 4 muestra una gráfica sobre los resultados obtenidos. Estos resultados indican que LADA comparte características genéticas con ambos tipos de diabetes: tipo 1 (HLA, INS VNTR y PTPN22) y tipo 2 (TCF7L2). Lo que justifica la consideración de LADA como una mezcla de los dos tipos principales de diabetes.

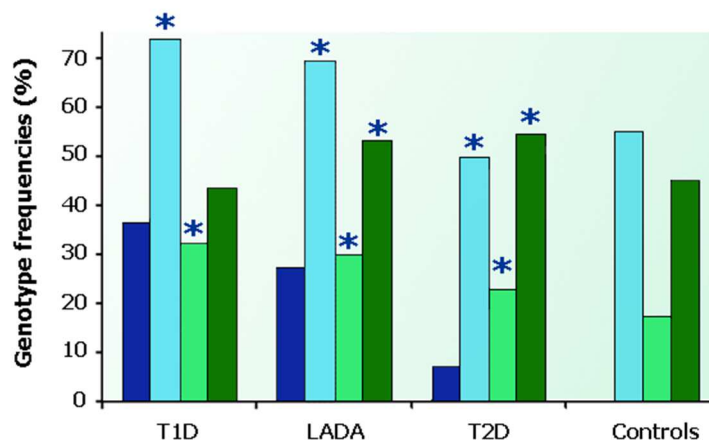


Figura 4: Frecuencias de los genotipos de riesgo para los 4 loci de susceptibilidad en DM1, LADA y DM2. *Asociación significativa con diabetes ($P < 0.05$). En azul oscuro, HLA-DQB1 (*0201/*0302); en azul claro, INS VNTR-rs689 (genotipo AA); en verde claro, PTPN22-rs2476601 (genotipos CT/TT); en verde oscuro, TCF7L2-rs7903146 (genotipos CT/TT). Cervin et al. (2008). *Diabetes*, 57, 1433-1437.

DIABETES GESTACIONAL

La diabetes mellitus gestacional (DMG) se define como una alteración de la tolerancia a la glucosa que se inicia o se detecta por primera vez durante el embarazo (Gabbe, 1986). Es la complicación metabólica más frecuente del embarazo (Virjee et al, 2001), con una prevalencia creciente en relación directa con el aumento de la obesidad (Dabelea et al, 2005). Durante el embarazo, la mujer se enfrenta a un aumento de la adiposidad y de la resistencia a insulina. La resistencia a la insulina que se desarrolla durante el embarazo se explica en parte, por el aumento en la producción del lactógeno placentario humano, los estrógenos y la prolactina (Di Cianni et al, 2003; Karnik et al, 2007; Kim et al, 2010). Por lo que, aquellas mujeres con una capacidad limitada para la compensación de la resistencia a la insulina por parte de las células β , son más susceptibles a desarrollar DMG (Buchanan et al, 2005).

Este tipo de diabetes es un factor de riesgo para el desarrollo posterior de diabetes tipo 2 (Buchanan et al, 2005), de hecho, hasta un 70% de las mujeres con DMG desarrollará DM2 en los primeros 5 años después del embarazo (Kim et al, 2002). Por otra parte, los hijos de mujeres con DMG tienen un mayor riesgo de desarrollar obesidad y diabetes tipo 2 a lo largo de su vida (Gillman et al, 2003). Basándose en estos hallazgos, se puede

Introducción

suponer que la DMG y la diabetes tipo 2 poseen un fondo genético común. Algunas de las variantes genéticas fuertemente asociadas con la diabetes tipo 2 también lo están con el riesgo de padecer DMG (Cho et al, 2009). Sin embargo, el conocimiento genético de la DMG es aún limitado.

Entre los loci asociados con DM2, que se han asociado también con el riesgo de padecer DMG están los siguientes: KCNJ11, TCF7L2, CDKAL1, KCNQ1, CDKN2A/CDKN2B, HHEX/IDE, IGF2BP2, SLC30A8, TCF2, FTO, INS, IRS1, ADRB3, MBL2, CAPN10, SERPINE1, INSR, GCK, HNF1A (Shaath et al, 2005; Shaath et al, 2007; Watanabe et al, 2007; Cho et al, 2009; Lauenborg et al, 2009; Shin et al, 2009; Zhou et al, 2009; Liton et al, 2007; Fallucca et al, 2006; Festa et al, 1999; Megia et al, 2004; Leipold et al, 2004; Leipold et al, 2006; Ober et al, 1983; Shaath et al, 2006; Shaath et al, 2006).

Sin embargo, los estudios sobre DMG, se basan a menudo en un tamaño de muestra relativamente pequeño, con respecto al utilizado para descubrir nuevos marcadores de DM2. Estos resultados no suelen estar confirmados en otras poblaciones y, por tanto no se consideran tan robustos. Al igual que para DM1 y DM2, el riesgo genético global de DMG se compone probablemente de la suma de las contribuciones de diferentes genes (Lauenborg et al, 2009).

Dado que la DMG en mujeres embarazadas parece revelar una predisposición genética a padecer otros tipos de diabetes, y al relacionarse con la resistencia a insulina, no es sorprendente que la más frecuente sea la predisposición a padecer DM2, especialmente si otros factores de riesgo como la obesidad están también presentes. Sin embargo, los estudios de agregación familiar sobre diabetes, muestran también un aumento de la prevalencia de la diabetes tipo 1 en hijos de madres con DMG (Dörner et al, 1987). Esto sugiere que la DMG también puede ser causada por variaciones genéticas que predisponen a una mujer a padecer diabetes tipo 1 o LADA. De hecho, alrededor del 10% de las mujeres con DMG puede padecer una forma autoinmune de diabetes (Lapolla et al, 2009).

1.2- SECUNDARIAS

Son las diabetes inducidas por:

Enfermedades del páncreas exocrino

Cualquier proceso que pueda lesionar de forma difusa el páncreas puede causar diabetes. Algunos de estos procesos son: pancreatitis, traumas, infecciones, pancreatectomías y carcinoma pancreático (Schwartz et al, 1978; Larsen et al, 1987; Cersosimo et al, 1991). Con la excepción del cáncer, para que se produzca diabetes, la lesión del páncreas debe ser muy extensa. Por el contrario, adenocarcinomas pancreáticos que sólo afectan una pequeña porción del páncreas están asociados a diabetes. Esto implica que, además de la reducción en el número de células β , debe haber algún otro mecanismo implicado. La fibrosis quística y la hemocromatosis, cuando son muy extensas, también dañan las células β y pueden alterar la secreción de insulina (Handwerger et al, 1969; Phelps et al, 1989).

Endocrinopatías

Algunas hormonas, como la hormona de crecimiento, el cortisol, el glucagón y la adrenalina antagonizan la acción de la insulina. Por lo tanto, cantidades excesivas de estas hormonas (en la acromegalia, enfermedad de Cushing, glucagonoma, feocromocitoma, etc.) pueden ocasionar diabetes (Soffer et al, 1961; Jadresic et al, 1982; Stenstrom et al, 1988; Berelowitz et al, 1996). Esto ocurre generalmente en individuos con defectos preexistentes en la secreción de insulina y, la hiperglucemia suele desaparecer cuando se controla el exceso de estas hormonas.

La hipokalemia inducida por el somatostatina y por la aldosterona puede ocasionar diabetes al menos parcialmente, al inhibir la secreción de la insulina (Conn, 1965; Berelowitz et al, 1996). La hiperglucemia se resuelve, en general, cuando se elimina el tumor.

Introducción

Fármacos

Muchos fármacos pueden alterar la secreción de insulina. Estos fármacos no causan diabetes por sí mismos, pero pueden precipitar la aparición de diabetes en individuos con cierta resistencia a la insulina (O'Byrne et al, 1990; Pandit et al, 1998). Algunas toxinas como el vacor (un raticida) y la pentamidina intravenosa pueden destruir permanentemente las células β del páncreas (Gallanosa et al, 1981; Bouchard et al, 1982; Assan et al, 1995; Esposti et al, 1996). Sin embargo, tales reacciones a fármacos son relativamente raras.

También existen numerosos fármacos y hormonas que pueden afectar a la acción de la insulina. Algunos ejemplos son el ácido nicotínico y los glucocorticoides (O'Byrne et al, 1990; Pandit et al, 1998). Se ha comunicado el desarrollo de diabetes asociada con anticuerpos contra las células β y, en algunos casos, severa deficiencia en insulina en pacientes tratados con interferón (Fabris et al, 1992; Shiba et al, 1996).

Infecciones

Algunos virus han sido asociados a la destrucción de las células β . Se produce diabetes en pacientes con rubeola congénita (Forrest et al, 1971), aunque muchos de estos pacientes poseen marcadores como HLA y otros, característicos de la diabetes de tipo 1. También los coxsackievirus B, citomegalovirus, adenovirus, y virus de las paperas han sido identificados como inductores de algunos casos de diabetes (King et al, 1983; Karjalainen et al, 1988; Pak et al, 1992).

2-DIABETES MELLITUS TIPO 2.

La diabetes mellitus tipo 2 (DM2) se ha convertido en una epidemia mundial que contribuye de forma importante al incremento de la morbi-mortalidad prematura en la población, actualmente afecta a cientos de millones de personas en el mundo y hasta a un 14% de la población adulta española (Soriguer et al, 2012); y se prevé que este porcentaje aumente en los próximos años. Este incremento se debe a cambios importantes en nuestra forma de vida que suponen un aumento importante de la obesidad y de sus consecuencias, siendo una de las más importantes el aumento de la DM2. En la obesidad parece claro que, principalmente la grasa abdominal genera, en muchos casos, un incremento de la resistencia a la insulina (RI), la cual puede conducir al desarrollo de DM2 (Kopelman et al, 2000; Kahn et al, 2006). El incremento del índice de masa corporal (IMC), del perímetro de cintura o de otros marcadores de obesidad, supone un incremento importante del riesgo de desarrollar DM2 (Kopelman et al, 2000; Lyssenko et al, 2008).

La DM2 resulta de la interacción de factores genéticos de predisposición y factores ambientales y adquiridos, principalmente la obesidad. El componente genético de la DM2, frente a similares factores ambientales y adquiridos (igual obesidad, nutrición, nivel de ejercicio, etc.), supone que unos pacientes desarrollen RI y algunos de ellos DM2, mientras que otros no desarrollan RI ni DM2. Diversos estudios han mostrado la importante heredabilidad de la DM2 y de parámetros relacionados (superior al 70%). Así, existe una importante concordancia entre gemelos idénticos en la presencia de DM2 o de intolerancia a la glucosa, que es superior al 70% en estudios transversales; pudiendo llegar al 76% y al 96% en estudios de seguimiento (Medici et al, 1999; Poulsen et al, 1999); además, la presencia de historia familiar de DM2 aumenta el riesgo de forma considerable (Lyssenko et al, 2008). Por otro lado, el incremento del IMC aumenta el efecto de polimorfismos asociados con DM2 y el riesgo de desarrollar DM2, tras un seguimiento de más de 20 años (Lyssenko et al, 2008).

Hasta el momento se han realizado numerosos estudios genéticos para conocer los genes asociados a la DM2 utilizando diferentes estrategias: estudios de genes

Introducción

candidatos, estudios de ligamiento en familias, asociación de polimorfismos a lo largo de todo el genoma o GWAS (Florez et al, 2008; Bonnefond et al, 2010). En los últimos años se han realizado diferentes estudios GWAS que han permitido la identificación de varios loci de susceptibilidad a la DM2 (Frayling et al 2007; Scott et al, 2007; Sladek et al 2007; Zeggini et al, 2007 y 2008; Unoki et al, 2008; Rung et al, 2009; Oi et al, 2010; Tsai et al, 2010; Voight et al, 2010; Saxena et al, 2013; Hara et al, 2014). Si bien, solo una parte de los resultados han sido validados en estudios de replicación en otras poblaciones (Palmer et al, 2008; Tan et al 2010).

Es importante destacar que pese a todo el trabajo realizado, el conjunto de loci y de variantes genéticas identificadas explican entre el 10 y el 15% de la heredabilidad de la DM2 (Zeggini et al, 2008; Billings et al, 2010), al igual que sucede con otras enfermedades complejas (So et al, 2011). Por otro lado, ninguno de los estudios poblacionales más importantes ha incluido población española. Pese a que el conocimiento actual de los genes implicados en el desarrollo de DM2 es escaso, la información obtenida está permitiendo que se abran nuevas posibilidades para entender esta enfermedad, su prevención, el diseño de nuevos tratamientos y el desarrollo de nuevas área de investigación (Prokopenko et al, 2008).

Los estudios GWAS realizados hasta el momento incluyen normalmente miles de muestras en estudios individuales y unas 95.000 en un metaanálisis (Voight et al, 2010). Estos estudios permiten identificar solamente la asociación de polimorfismos frecuentes o PFs (frecuencia del alelo menor o $MAF > 0.05$), sin embargo estos PFs en general parecen tener un efecto pequeño a nivel funcional y sobre el fenotipo (OR entre 1.1 y 1.5) (Bodmer y Bonilla, 2008; So et al, 2011).

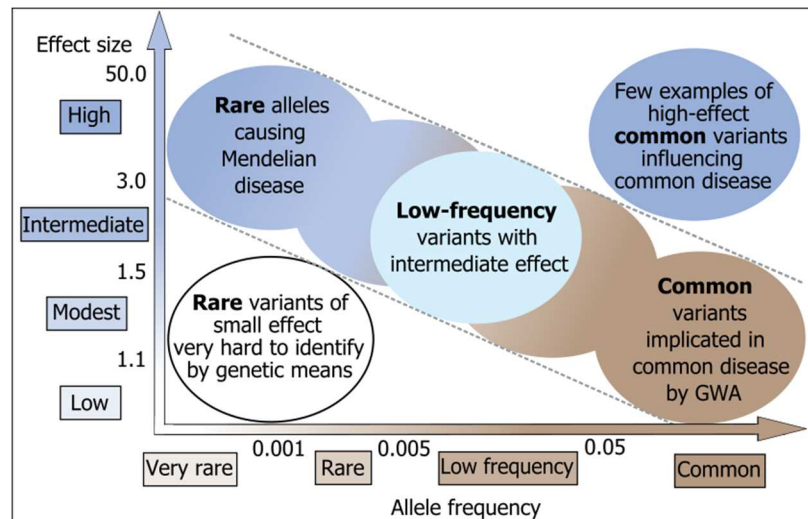


Figura 5: Efecto vs frecuencia de las variantes genéticas. Manolio et al. (2009). *Nature*, 461, 747-753.

Otra limitación de los GWAS anteriores es que los chips utilizados hasta el momento dejan muchas regiones genómicas sin una buena cobertura. Si bien, en este momento ya hay chips con muchos más polimorfismos (más de 2 millones de SNPs por muestra) que también incluyen PPFs o polimorfismos poco frecuentes ($MAF < 0.05$) y presentan una mayor cobertura. El aumento en el número de polimorfismos puede tener diferentes limitaciones: la cobertura podría seguir siendo insuficiente en zonas del genoma donde se necesitaría una densidad mucho mayor de la alcanzada para poder asegurar que se analiza toda su variabilidad, los tamaños muestrales necesarios son excesivamente grandes para poder detectar muchas de las asociaciones esperables, principalmente en los PPFs, y el coste es muy elevado.

Todo lo anterior supone que muchos PFs y PPFs, y muchos de los loci que los contienen no hayan sido identificados hasta el momento. En el caso de la diabetes tipo 2 se han identificado 62 loci asociados (Mahajan et al, 2014) y se espera que haya otros muchos que tengan un papel relevante.

Por tanto, actualmente se sabe que el componente genético de la DM2 es del 70%, pero con los GWAS llevados a cabo sólo se ha identificado un 10-15% de éste. Por tanto, el gran parte de la heredabilidad de la enfermedad todavía no se ha encontrado. Uno de los métodos planteados para encontrar esta heredabilidad perdida, se basa en la teoría

Introducción

“enfermedades comunes-polimorfismos raros” versus “enfermedades comunes-polimorfismos comunes”. El estudio de los polimorfismos comunes ha sido muy amplio hasta el momento debido al uso de los GWAS. Sin embargo, la teoría “enfermedades comunes-polimorfismos raros”, toma cada vez más fuerza y hace imprescindible el estudio de los polimorfismos poco frecuentes, los cuales podrían tener un mayor efecto en el fenotipo tal y como indica la figura 5.

2.1- ASPECTOS CLÍNICOS Y EVOLUTIVOS DE LA DIABETES TIPO 2.

En muchos casos, la DM2 tarda en ser diagnosticada porque muchos de sus síntomas pueden parecer inofensivos. Sin embargo, la detección temprana y el tratamiento de los síntomas pueden disminuir la posibilidad de padecer las complicaciones que conlleva la enfermedad.

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

Las complicaciones microvasculares son las siguientes:

Retinopatía: Se producen daños en la retina con aparición de manchas pequeñas y hemorragias microscópicas, que pueden ocasionar pérdida de visión.

Nefropatía: Se producen daños en el riñón. Cursa primero con proteinuria y luego con incapacidad progresiva para eliminar toxinas, además del aumento de creatinina y urea en la sangre, lo que puede derivar a diálisis e incluso causar la muerte.

Neuropatía: Se producen daños en los nervios periféricos, encargados de la transmisión de los impulsos nerviosos y de las sensaciones del tacto, calor, frío, dolor, etc. El compromiso de las fibras nerviosas es responsable de sensaciones alteradas como calambres, hormigueos, sensación de pies y plantas quemadas y/o de frío, dolores

nocturnos de piernas. Pueden aparecer lesiones en los pies, situación conocida como "Pie Diabético".

Y las complicaciones macrovasculares son las que se explican a continuación:

Arteriopatía de arterias carótidas y cerebrales: Se pueden producir accidentes vasculares cerebrales graves, que pueden causar la parálisis de un lado del cuerpo o pequeños infartos cerebrales que irán causando trastornos progresivos de memoria.

Arteriopatía de las coronarias: Se produce la obstrucción de las arterias responsables del flujo sanguíneo al corazón, lo que puede generar un Infarto de Miocardio.

Arteriopatía de extremidades inferiores: Disminuye el flujo sanguíneo a las piernas, apareciendo un intenso dolor al caminar. Puede causar gangrena en los pies.

La glicación de las proteínas tisulares y otras macromoléculas y la excesiva producción de polioles a partir de la glucosa son dos de los mecanismos que se han propuesto para explicar el daño tisular resultante de la hiperglucemia crónica.

La figura 6 muestra la prevalencia de las complicaciones de la diabetes más comunes entre las personas con diabetes tipo 2. Se puede observar que la prevalencia de las complicaciones crónicas microvasculares es mucho más alta que la de las complicaciones macrovasculares.

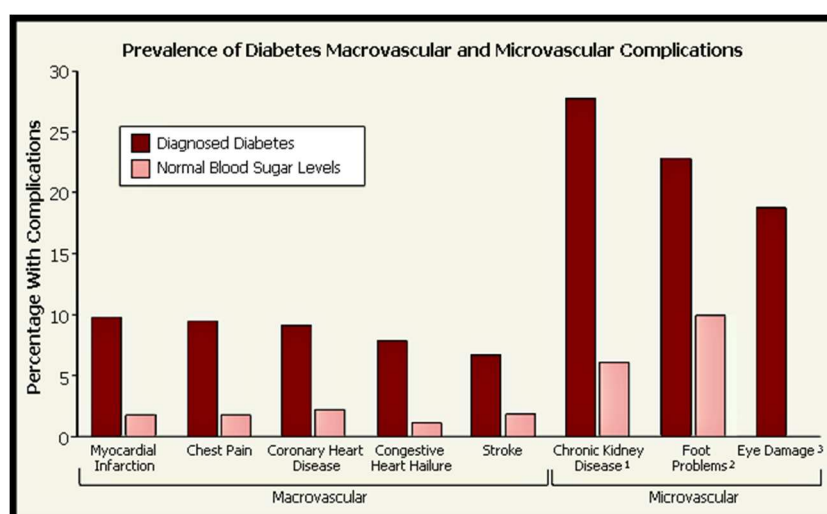


Figura 6: Prevalencia de las complicaciones producidas por la diabetes según los niveles de glucosa en sangre. Mitka. (2007). JAMA. 297, 2337-2338.

Introducción

Otras complicaciones incluyen problemas dentales, menor resistencia a las infecciones (como gripe y neumonía), macrosomía y otras complicaciones en el parto en mujeres diabéticas y embarazadas.

2.2-DIAGNÓSTICO DE LA DIABETES TIPO 2

Existen varias pruebas para diagnosticar la diabetes. Las más comunes se exponen a continuación, estableciéndose los valores para el diagnóstico en la tabla 3:

Medición de la hemoglobina glicosilada (HbA1C). Mide el nivel promedio de glucosa en la sangre durante los últimos 2 o 3 meses. El paciente no tiene que ayunar ni beber nada para hacerse esta prueba.

Medición de glucosa plasmática en ayunas. Esta prueba generalmente se realiza a primera hora de la mañana ya que su objetivo es medir los niveles de glucosa en sangre en ayunas.

Prueba de tolerancia a la glucosa oral. Es una prueba que mide los niveles de glucosa en la sangre antes de tomar una bebida azucarada (para el test Standard la cantidad utilizada es de 75-100 gr.) y 2 horas después de tomarla. Indica cómo el cuerpo procesa la glucosa.

Esta sobrecarga oral de glucosa (SOG) debe realizarse después de 8-12 h de ayuno, manteniendo una dieta con un contenido mínimo de 250 g diarios de hidratos de carbono, al menos durante los tres días previos. La glucosa administrada se diluye al 25% y debe ser ingerida en 5 minutos.

Prueba aleatoria (o casual) de glucosa plasmática. Esta prueba es un análisis de sangre en cualquier momento del día cuando se tienen síntomas de diabetes severa.

En la tabla 2 se muestran los valores que permiten el diagnóstico de la diabetes según la American Diabetes Association:

	Normal	Prediabetes	Diabetes
HbA1c (%)	<5.7	5.7-6.4	≥6.5
Glucemia plasmática basal (mg/dl)	<100	100-125	≥126
Glucemia 2h SOG con 75 g (mg/dl)	<140	140-199	≥200
Glucemia al azar con clínica hiperglucémica (mg/dl)	-	-	≥200

Tabla 2. Parámetros para el diagnóstico de la DM2. *American Diabetes Association, 2013.*

2.3- FACTORES DE RIESGO Y PREVENCIÓN DE LA DIABETES TIPO 2.

Cada año se producen cuatro millones de muertes atribuibles a la diabetes, lo que constituye el 6.8% del total de la mortalidad mundial (para todas las edades y de todas las causas) (Roglic et al, 2006; Roglic et al, 2009). La diabetes es la cuarta causa principal de muerte relacionada con una enfermedad y casi el 80% de las muertes relacionadas con esta enfermedad se dan en países en vías de desarrollo.

Según la Organización Mundial de la Salud (OMS), las muertes relacionadas con la diabetes aumentarán en un 50% en los próximos 10 años si no se toman medidas. En la figura 7, aparece el número de personas que padecen diabetes tipo 2 en las diferentes regiones del mundo, en la actualidad y también, las que se calcula que la padecerán en 2035. Esto permite ver el aumento de casos que se va a producir en este período de tiempo.



REGIÓN DE LA FID	2013 MILLONES	2035 MILLONES	AUMENTO %
● África	19,8	41,4	109%
● Oriente Medio y Norte de África	34,6	67,9	96%
● Sudeste Asiático	72,1	123	71%
● América Central y del Sur	24,1	38,5	60%
● Pacífico Occidental	138,2	201,8	46%
● América del Norte y Caribe	36,7	50,4	37%
● Europa	56,3	68,9	22%
Mundo	381,8	591,9	55%

Figura 7. Proyecciones por región del número de personas con diabetes entre 20 y 79 años, para 2013-2035. International Diabetes Federation, 2013.

La DM2 supone una notable carga económica para los pacientes, las familias, los sistemas de salud y los diferentes países, en particular para los países con pocos recursos (Mbanya et al, 2010). A pesar de ello, más del 80% de los gastos relacionados con la diabetes en el mundo se produce en los países desarrollados (Brown et al, 2006). Dados los costes asociados a la diabetes en los países en vías de desarrollo, donde los recursos para hacer frente a la enfermedad son muy limitados, es necesario el desarrollo, implementación y evaluación de los programas para prevenir esta enfermedad y sus complicaciones (Alberti et al, 2007). Sin embargo, la mayoría de los estudios de prevención llevados a cabo hasta el momento, se han realizado en países desarrollados.

Estos ensayos incluyen el “Estudio de viabilidad de Malmo”, en Suecia (Eriksson et al, 1991), el “Estudio de la Prevención de la Diabetes de Finlandia” (DPS) (Uusitupa et al, 2000; Tuomilehto et al, 2001; Lindstrom et al, 2003) y el “Programa de Prevención de Diabetes de EE.UU” (DPP) (Knowler et al, 2002). Todos ellos han demostrado ser

eficaces, ya que se ha producido una reducción del 63% en el riesgo de desarrollar diabetes tipo 2, en los grupos en que se ha dado intervención en el estilo de vida en comparación con los grupos control. El metaanálisis realizado en 2007 por Gillies et al, ha confirmado estos hallazgos, poniendo de manifiesto que las intervenciones no farmacológicas son eficaces en la reducción del riesgo de diabetes tipo 2 en personas con intolerancia a la glucosa.

Los ensayos centrados en la prevención de complicaciones de la DM2 realizados en países desarrollados no han evaluado directamente los efectos de las intervenciones no farmacológicas en la incidencia de complicaciones. No obstante, los estudios realizados han demostrado que las intervenciones no farmacológicas son eficaces para reducir los niveles de glucosa en personas con diabetes tipo 2 (Eriksson et al, 1991; Tuomilehto et al, 2001) y estudios realizados en Reino Unido y Japón indican que esta disminución de los niveles de glucosa en personas con diabetes tipo 2 recién diagnosticada puede reducir el desarrollo de complicaciones (Agurs-Collins et al, 1997; Davies et al, 2008).

A pesar de que la patogénesis de la DM2 es compleja, se han identificado una serie de factores de riesgo modificables y no modificables que predisponen a padecerla:

- a) Los factores de riesgo modificables suelen estar relacionados con el estilo de vida y, por tanto puede incidirse sobre ellos para hacer cambios que conduzcan a un estado de vida más saludable. Entre ellos están un mayor índice de masa corporal (IMC), la inactividad física, la mala alimentación, la hipertensión, el tabaquismo y el consumo de alcohol (Zimmet et al, 2001; Egede et al, 2005). El aumento del IMC y de la cintura son, posiblemente, los mayores factores de riesgo (Hu et al, 2001; Rewers et al, 1995; Kaye et al, 1991). En cuanto a la inactividad física, diferentes estudios muestran que menores niveles de actividad física aumentan el riesgo de padecer diabetes. Una revisión de 10 estudios prospectivos de cohortes que estudian la relación entre la diabetes y una actividad física de intensidad moderada, revela que las personas con niveles de actividad física medios suelen presentar alrededor del 30% menos de probabilidad de desarrollar diabetes que las personas que no practican ninguna

Introducción

actividad física (Jeon et al, 2007). La ingesta calórica total, así como la de carbohidratos y grasas saturadas, se han relacionado también con el desarrollo de diabetes. El consumo moderado de alcohol puede reducir el riesgo de desarrollar diabetes (Narayan et al, 2006), pero se ha demostrado que el tabaquismo sí que constituye un factor de riesgo para la diabetes (Will et al, 2001). Por último, los factores psicosociales como la depresión, el estrés, los problemas sociales, y un mal estado de salud mental también se asocian con un mayor riesgo de desarrollar diabetes (Eaton et al, 1996; Grandinetti et al, 2000; Diez Roux et al, 2002; Arroyo et al, 2004; Strodl et al, 2006).

- b) En cuanto a los factores de riesgo no modificables, como su nombre indica, no pueden variar. Éstos incluyen la edad, la raza u origen étnico, la historia familiar (predisposición genética), antecedentes de diabetes gestacional y bajo peso al nacer. La incidencia y prevalencia de la diabetes aumentan con la edad. Por otra parte, en cuanto a las diferentes razas, los afroamericanos son más propensos a desarrollar diabetes que los individuos de raza blanca (Egede et al, 2005). Y para los nativos americanos, las tasas de diagnóstico de la diabetes presentan diferencias de entre el 5% al 50% en diferentes tribus y grupos de población.

En relación al sexo, el estudio de Choi et al (2013) concluyó que en afroamericanos e hispanos, las mujeres presentaron una mayor prevalencia de DM2 que los hombres (Harris et al., 1998; Hertz et al., 2006). En asiáticos, por el contrario, los hombres presentaron mayor prevalencia de DM2 que las mujeres (Wang et al., 2011). Y, mientras que un estudio anterior (Cowie et al., 2010) se observó que la prevalencia de DM2 fue similar en ambos sexos en los caucásicos, este estudio mostró mayor prevalencia en hombres que en mujeres. No obstante, es importante señalar que la diferente acumulación de grasa en ambos géneros condiciona el riesgo de desarrollar la enfermedad, ya que los hombres acumulan mayor cantidad de grasa en torno a la región abdominal y las mujeres mayormente en la cadera y el muslo (Logue et al, 2011).

Como ya se ha indicado previamente, la heredabilidad de la DM2 es de un 70%, lo que indica que el componente genético es muy importante. Además, el incremento constante de los factores de riesgo modificables hace que los factores genéticos de riesgo sean cada vez más relevantes y tengan un efecto más potente (Quintanilla-García et al, 2010).

2.4- PATOGÉNESIS DE LA DIABETES TIPO 2

La patogénesis de esta enfermedad se basa en dos pilares fundamentales: la resistencia a la insulina y la disfunción de las células β . En la mayoría de los casos, se dan ambas en el desarrollo de la enfermedad y conducen a una marcada hiperglucemia. Para estos casos se propone como mecanismo para el desarrollo de la enfermedad, el incremento de la resistencia a insulina, lo cual supone un incremento del estrés de las células β (por una alta producción de insulina, la hiperglucemia, exceso de estrés oxidativo, etc.) que finalmente puede conducir a la falta de funcionalidad o a la muerte de estas células.

2.4.1- Disfunción de las células β

Secreción de insulina en condiciones normales

En condiciones normales, la glucosa es captada por la célula β pancreática a través del transportador de glucosa 2 (GLUT2). En este paso, la glucosa se fosforila mediante la enzima glucoquinasa, este es el paso limitante de la velocidad del metabolismo de la glucosa en las células β . Al metabolizarse, la glucosa conduce a la formación de piruvato, el cual se dirige posteriormente a la mitocondria y conduce a la formación de ATP. El ATP es necesario para la liberación de insulina, pero también está implicado en la despolarización de la membrana celular. La proporción ADP/ATP conduce a la activación del receptor de sulfonilurea 1 (SUR1), lo que dará lugar a cierre del canal de potasio adyacente. El cierre de los canales de potasio altera el potencial de membrana y abre los canales de calcio, lo que desencadena la liberación de gránulos preformados que contienen insulina. Un resumen de este proceso se muestra en la figura 8:

Introducción

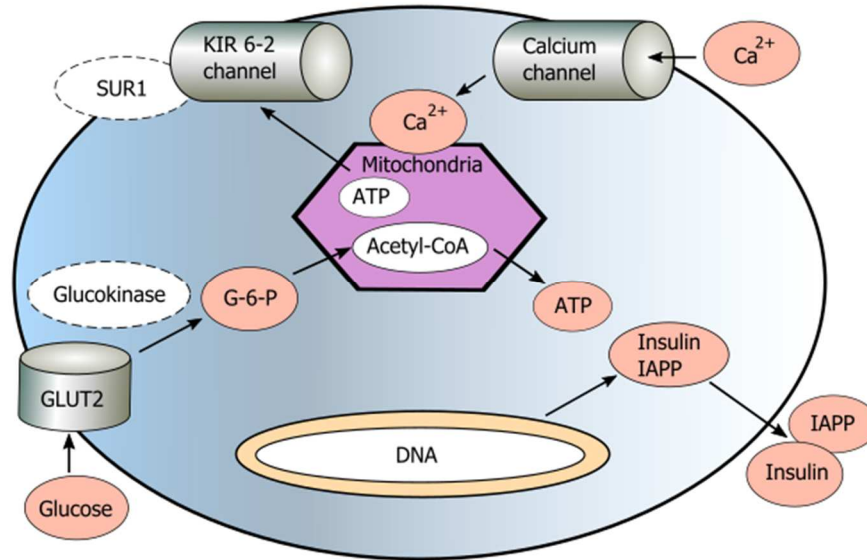


Figura 8: Secreción de insulina inducida por glucosa en condiciones normales. (IAPP = polipéptido amiloide de los islotes. G-6-P = glucosa-6-fosfato. CoA = coenzima A. GLUT2 = transportador de glucosa 2). Stumvoll et al. (2005). *Lancet*. 365, 1333-1346.

Los pacientes con diabetes tipo 2 padecen diferentes problemas en la regulación de la secreción de insulina. Como consecuencia, las concentraciones basales de ésta pueden llegar a elevarse a incluso más del doble del valor normal, sobre todo en pacientes hiperglucémicos obesos, pero esto se debe, presumiblemente, a un aumento de los niveles de glucosa en plasma. Del mismo modo, después de una comida, las concentraciones de insulina en el plasma pueden aparecer más elevadas de lo normal, también debido al aumento de la glucosa plasmática.

Pero incluso antes de la hiperglucemia manifiesta, que en sí misma afecta negativamente a la función de las células β , se han demostrado defectos en la secreción de insulina por ejemplo, en individuos prediabéticos (Van Haeften et al, 2002). Además, tanto en hijos no diabéticos de padres diabéticos, como en gemelos no diabéticos con padres diabéticos, se ha observado que existe una reducción en la secreción de insulina (Gerich, 1998). Por lo tanto, en individuos predispuestos existe un defecto, posiblemente de base genética, en la secreción de insulina (Bonnadonna et al, 2003). La obesidad, enfermedades agudas, o simplemente el envejecimiento pueden agravar el

defecto subyacente, conduciendo en última instancia a una diabetes manifiesta (Gerich, 1998).

En los indios Pima, conocidos por padecer resistencia a la insulina, los estudios prospectivos han demostrado que la baja liberación de insulina determina la transición desde una tolerancia normal a la glucosa hasta una intolerancia a ésta. Por el contrario, los indios Pima que no llegan a ser diabéticos suelen ser capaces de aumentar la secreción de insulina con el tiempo (Weyer et al, 1999).

Fisiopatología de la hiperglucemia

Para entender los mecanismos celulares y moleculares responsables de la diabetes tipo 2, es necesario conocer cómo se controla la glucemia. La insulina es la hormona clave para la regulación de los niveles de glucosa en sangre y, en general, la normoglucemia se mantiene por un equilibrio entre la acción y la secreción de insulina. Es importante destacar que, la célula β pancreática normal puede adaptarse a los cambios es decir, una disminución de la acción de la insulina está acompañada por la regulación positiva de la secreción de insulina (y viceversa). La figura 9 ilustra la relación curvilínea entre la función celular β normal y la sensibilidad a la insulina (Bergman, 1989). La desviación de esta hipérbola, ocurre por ejemplo en los pacientes con intolerancia a la glucosa o con diabetes tipo 2, y se produce cuando la función de las células β está disminuída y con ello aparece una menor sensibilidad a la insulina. Por lo tanto, la disfunción de las células β es un componente crítico en la patogénesis de la diabetes tipo 2.

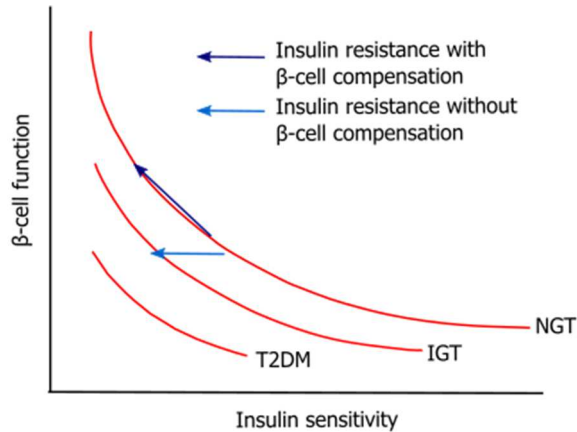


Figura 9: Relación hiperbólica entre la función de las células β y la sensibilidad a la insulina.

En las personas con tolerancia normal a la glucosa (NGT) existe una relación cuasi-hiperbólica entre la función de las células β y la sensibilidad a la insulina. Conforme se desvía esta hipérbola, se produce el deterioro en la tolerancia a la glucosa (tolerancia alterada a la glucosa [IGT], y diabetes tipo 2 [T2DM]). Stumvoll et al. (2005). *Lancet*. 365, 1333-1346.

Sin embargo, no sólo la desviación, sino también la progresión a lo largo de la hipérbola afecta a la glucemia. Cuando la acción de la insulina disminuye (como ocurre en la obesidad) hay una compensación mediante el aumento de la función de las células β . No obstante, a pesar de esta compensación, las concentraciones de glucosa en sangre en ayunas y 2 horas después de una sobrecarga oral de glucosa aumentarán levemente (Stumvoll et al, 2003). Este aumento puede ser leve, pero con el tiempo se vuelve perjudicial debido a la toxicidad de la glucosa, y es en sí mismo un motivo de disfunción de las células β . Por lo tanto, la resistencia a la insulina abre el camino para la hiperglucemia y la diabetes tipo 2, incluso con una ilimitada reserva de células β .

Toxicidad de la glucosa

La idea de que la hiperglucemia en sí puede disminuir la secreción de insulina ha llevado al concepto de toxicidad de la glucosa, lo que implica un daño irreversible con el paso del tiempo a los componentes celulares implicados en la producción de insulina (Robertson et al, 2003). En efecto, el deterioro de la secreción de insulina con el tiempo es lo habitual en la mayoría de los pacientes, de hecho muchos de ellos terminan con una deficiencia más o menos severa de insulina después de 10 años de diabetes (Wallace

et al, 2002). En las células β , el metabolismo oxidativo de la glucosa conduce a la producción de especies reactivas del oxígeno, normalmente eliminadas por la catalasa y la superóxido dismutasa. Las células β están equipadas con una baja cantidad de estas enzimas, y también de la enzima glutatión peroxidasa, reguladora del estado redox (Robertson et al, 2003). La hiperglucemia da lugar a la producción de gran cantidad de especies reactivas del oxígeno en estas células, con el consiguiente daño a los componentes celulares.

La pérdida del “pancreas duodenum homeobox 1” (PDX-1), el cual regula la actividad del promotor de insulina, también ha sido propuesta como un mecanismo que conduce a la disfunción de las células β (Robertson et al, 2003). Además, las especies reactivas de oxígeno son conocidas por aumentar la actividad del factor nuclear kB (NFkB), que induce potencialmente la apoptosis de las células β .

Lipotoxicidad

En general, en los pacientes obesos tanto diabéticos como no diabéticos, las concentraciones de ácidos grasos libres no esterificados (NEFA, por sus siglas en inglés “Non-Esterified Fatty Acids”) aparecen como resultado del incremento de la lipólisis de los adipocitos. Los ácidos grasos conducen a una mayor secreción de insulina en estudios de toxicidad aguda, pero después de 24 h se observa que inhiben la secreción de insulina. En presencia de glucosa, se inhibe la oxidación de ácidos grasos en las células β y se da la acumulación de acil coenzima A de cadena larga (Robertson et al, 2004). Este mecanismo ha sido propuesto como parte del proceso normal de secreción de insulina. Sin embargo, este acil coenzima A de cadena larga en sí, también puede disminuir el proceso de secreción de insulina mediante la apertura de canales de potasio en las células β (figura 10). Un segundo mecanismo podría ser el aumento de expresión de la “uncoupling protein-2” (UCP2), lo que llevaría a una reducción en la producción de ATP y, por tanto, a la disminución de la secreción de insulina. Un tercer mecanismo podría implicar la apoptosis de las células β , posiblemente a través la síntesis de ceramida inducida por triglicéridos, ácidos grasos o de la generación de óxido nítrico.

Depósitos de polipéptido amiloide

El polipéptido amiloide o amilina se localiza conjuntamente con la insulina en los gránulos secretores de las células β pancreáticas. Existen evidencias de que se secreta conjuntamente con la insulina en condiciones basales y en respuesta tanto a la glucosa como a diferentes estímulos secretagogos. Experimentos *in vitro* con islotes pancreáticos humanos aislados, han demostrado que la regulación del gen de amilina depende de las concentraciones de glucosa y de las señales derivadas de su metabolismo, de una forma similar a lo que ocurre con el gen de la insulina (Novials et al, 1993; Gasa et al, 1997). Entre las acciones biológicas de la amilina (para algunas de las cuales se han descrito receptores específicos) destacan: la inhibición de la secreción de glucagón, la estimulación del eje renina-angiotensina, el retraso del vaciamiento gástrico, y la inhibición de los centros hipotalámicos reguladores del apetito. Algunas de estas acciones se están estudiando como tratamiento coadyuvante de la insulina para ciertos casos de diabetes.

Se ha sugerido que los pequeños agregados amiloides son citotóxicos, (Janson et al, 1999) y están posiblemente relacionados con la producción de radicales. Los NEFA pueden aumentar la citotoxicidad de estos agregados (Hull et al, 2004). Los depósitos de amiloide se encuentran en la mayoría, pero no en todos los pacientes con diabetes tipo 2 (Knowles et al, 2002). Estos depósitos podrían ser una consecuencia de la sobreproducción de amilina que acompaña a la sobreproducción de insulina, y a continuación, contribuirían a la degeneración y pérdida de masa celular β , especialmente en presencia de concentraciones elevadas de NEFA. El hecho de que en los familiares de primer grado de pacientes con diabetes tipo 2, éste polipéptido se encuentre disminuido, se opone a esta hipótesis (Knowles et al, 2002). Además, el polipéptido amiloide no está presente en individuos resistentes a la insulina de mediana edad. Por lo tanto, todavía no se conoce el papel de los depósitos de amiloide en la fisiopatología de la diabetes tipo 2.

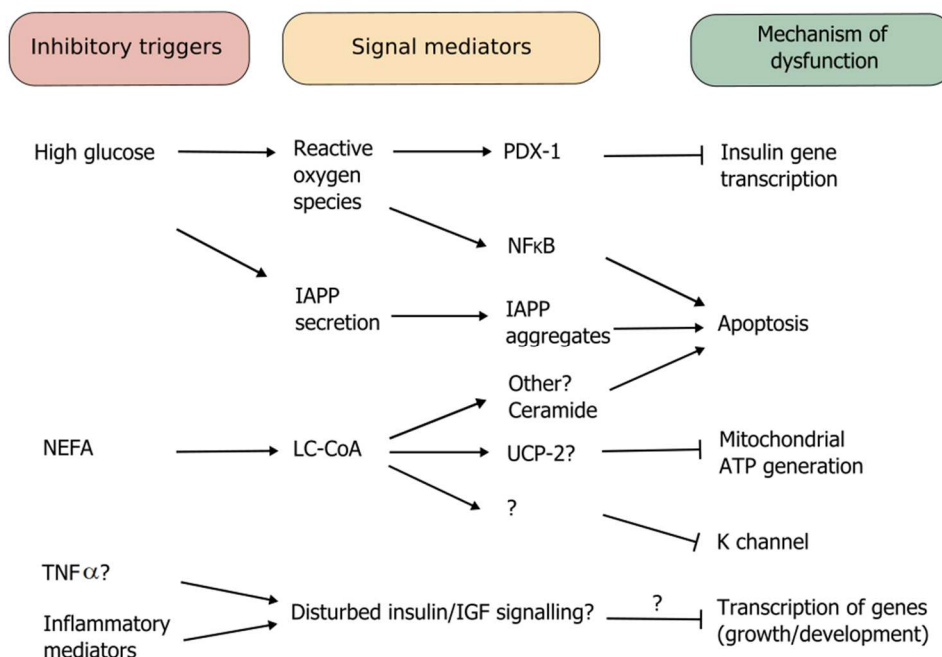


Figura 10: Los posibles efectos negativos de la hiperglucemia, aumento de NEFA, y varios moduladores implicados en la resistencia a la insulina en la disfunción de las células β
 IAPP = polipéptido amiloide de los islotes. LC-CoA de cadena larga = acetil coenzima A. PDX-1 = pancreas duodenum homeobox-1. UCP-2 = Uncoupled protein- 2. Stumvoll et al. (2005). *Lancet*. 365, 1333-1346.

2.4.2- Resistencia a la insulina

La resistencia a la insulina se produce cuando los efectos biológicos de ésta son menores de lo establecido, tanto para el consumo de glucosa por parte del músculo esquelético como para la supresión de la producción endógena de glucosa en el hígado (Dinneen et al, 1992). En condiciones de ayuno, el músculo disminuye el consumo de glucosa y el hígado por su parte, aumenta la producción de glucosa para mantener la glucemia dentro de los niveles adecuados.

La producción de glucosa endógena se encuentra aumentada en pacientes con diabetes tipo 2 (Meyer et al, 1998; Weyer et al, 1999). Sin embargo, este aumento se produce en presencia de hiperinsulinemia, pero, al menos en las etapas iniciales e intermedias de la enfermedad, la resistencia a la insulina hepática es la fuerza impulsora de la hiperglucemia en la diabetes tipo 2, como se muestra en la figura 11.

Introducción

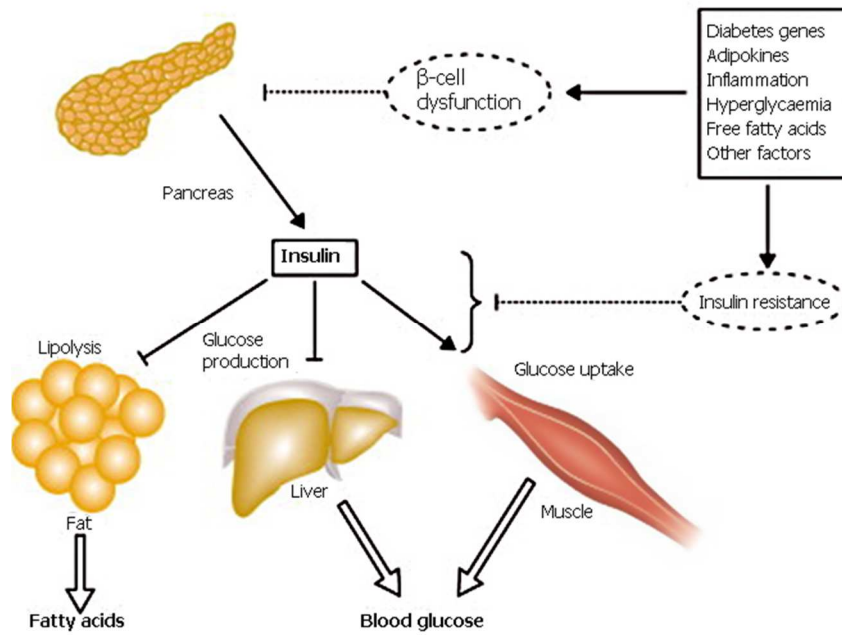


Figura 11: Fisiopatología de la hiperglucemia y el aumento de los ácidos grasos circulantes en la DM2. La secreción de insulina reduce la producción de glucosa por parte del hígado, aumenta la captación de glucosa por el músculo esquelético, y suprime la liberación de ácidos grasos del tejido adiposo. Los factores que contribuyen a la patogénesis de la diabetes tipo 2 afectan tanto a la secreción de insulina como a su acción. La resistencia a la insulina afecta a su acción en cada uno de sus tejidos diana, lo que lleva a un aumento de los niveles de ácidos grasos circulantes y a una hiperglucemia propia de la diabetes. A su vez, las concentraciones elevadas de glucosa y los ácidos grasos en el torrente sanguíneo conducirán tanto a una menor secreción de insulina como a una mayor resistencia a ésta. Stumvoll et al. (2005). *Lancet*. 365, 1333-1346.

Obesidad

La resistencia a la insulina se asocia con la obesidad y la inactividad física habiéndose identificado varios mecanismos que median esta asociación. Entre ellos están los mecanismos implicados por sustancias que tienen su origen en el adipocito. En este sentido, diferentes adipocinas (hormonas y citocinas producidas por los adipocitos) modulan la acción de la insulina. Un aumento de los triglicéridos, especialmente en los depósitos de tejido adiposo visceral, conduce a adipocitos de mayor tamaño y que son resistentes a la capacidad de la insulina para inhibir la lipólisis. Esta situación produce una mayor liberación de ácidos grasos libres no esterificados, aumentando sus niveles en sangre, lo cual agrava la resistencia a la insulina en el músculo esquelético y en el hígado, como se muestra en la figura 11 (Boden, 1997).

El almacenamiento de grasa excesiva, no sólo en los adipocitos, sino también de forma "ectópica" en células no adiposas tiene un papel importante (Danforth, 2000). Por ejemplo, un aumento de los lípidos intramiocelulares está asociado con la resistencia a la insulina del músculo esquelético bajo algunas circunstancias (Machann et al, 2004), y lo mismo ocurre en el hígado en relación a los lípidos intrahepáticos (Seppala-Lindros et al, 2002; Bajaj et al, 2003).

Mecanismos celulares

La insulina actúa mediante la unión a sus receptores específicos localizados en la membrana celular. El receptor de insulina (IR) es una glucoproteína que pertenece a la familia de receptores para factores de crecimiento con actividad intrínseca de quinasas de tirosina (RTK's), los cuales al ser estimulados por su ligando se autofosforilan en residuos de tirosina (Myers y White, 2002). Una vez que la insulina interacciona con su receptor y éste es activado, se inician las cascadas de señalización que dependen de un gran número de complejas interacciones proteicas. Dos vías principales de transducción son activadas por acción de la insulina: la vía de la fosfatidilinositol 3-quinasa (PI3K) y la vía de las quinasas activadas por mitógenos (MAP quinasas). Ambas vías regulan la mayoría de las acciones de la insulina asociadas a la regulación del metabolismo energético, de la expresión genética y de efectos mitogénicos (Avruch, 1998; Myers y White, 2002).

Los sustratos celulares del receptor de insulina (IRS), son proteínas que contienen dominios que permiten su unión al receptor de insulina activo. Adicionalmente, los IRS contienen entre 8 y 18 sitios potenciales de fosforilación, que al ser fosforilados por el IR se convierten en sitios de unión y activación de diferentes proteínas, de las cuales, muchas funcionan como proteínas adaptadoras (Virkamaki et al, 1999). Esto da lugar a la cascada de señalización que se muestra en la figura 12 (White, 2002). La insulina activa una serie de lípidos y quinasas vinculadas a la translocación de transportadores de glucosa a la superficie celular y se activa la síntesis de glucógeno, proteínas, ARNm, y ADN nuclear, lo cual afecta a la supervivencia y proliferación celular.

La fosforilación y desfosforilación de proteínas IRS

Después del estímulo con insulina, el IRS-1, que es el IRS que está implicado en el transporte de glucosa, se fosforila de manera notable en residuos de tirosina (Tyr) y de serina y treonina (Ser, Thr). Actualmente se sabe que, en la mayoría de los casos, la fosforilación de estos residuos está implicada en mecanismos de atenuación de la señal de insulina que desacopla la unión del IRS de proteínas efectoras de la vía de insulina como lo es la PI3K (Le Roith et al, 2003; Gual et al, 2005). La fosforilación en residuos de Ser/Thr de IRS puede llevarlo a: a) desacoplarse del IR lo que altera su capacidad de experimentar fosforilación en residuos de Tyr; b) su disociación de complejos intracelulares que lo mantienen en cercanía con el IR; c) su degradación o bien, d) convertirlo en proteínas inhibitoras de la actividad de quinasa del IR (Le Roith et al, 2003; Gual et al, 2005). Diferentes estudios han identificado varios residuos de Ser/Thr como blancos potenciales de fosforilación de quinasas de IRS que afectan a su activación, resultando en la disminución de la señalización (Paz y Zick, 2002; Le Roith et al, 2003; Gual et al, 2005; Engelman et al, 2006). Varias serina quinasas del IRS han sido identificadas, incluyendo varias proteínas quinasas activadas por mitógenos, “c-Jun NH2-terminal kinase”, “atypical protein kinase C”, y “phosphatidylinositol 3'-kinase”, entre otras (White, 2002).

Recientemente se ha demostrado que la familia de proteínas supresoras de señalización de citocinas (SOCS) juega un papel importante en regular negativamente la activación del IRS, ya sea por interacciones directas o indirectas. Se ha observado que su expresión es inducida por el tratamiento con la insulina en varios tejidos y líneas celulares. Cuando se induce la síntesis de las proteínas SOCS, éstas son capaces de asociarse con las proteínas IRS, alterando su estructura y su unión, tanto al IR, como a proteínas efectoras tales como PI3K. Además, se ha visto que la asociación de SOCS con IRS promueve su degradación y disminución en el número de células (Youngren, 2007). Por otra parte, la fosfotirosina fosfatasa 1B (PTP1B) se expresa ampliamente y tiene un papel importante en la regulación negativa de la señalización de insulina (Goldstein, 2003).

Papel de los productos de los adipocitos y la inflamación

El aumento de las concentraciones de NEFA y citoquinas inflamatorias (como el factor de necrosis tumoral α (TNF α) o la interleucina 6) liberados por el tejido adiposo visceral afectan negativamente a la cascada de señalización de la insulina (Ravussin, 2002; Rajala, 2003).

Los NEFA inhiben el metabolismo de la glucosa estimulado por insulina en el músculo esquelético y estimulan la gluconeogénesis en el hígado (Shulman, 2000; Boden et al, 2002). También se activan las quinasas celulares, incluyendo isoformas de la “atypical protein kinase C” mediante el aumento de los niveles de diacilglicerol celular, que pueden activar las quinasas inflamatorias: “inhibitor kappa B kinase” y “c-Jun NH2-terminal kinase”, el aumento de la fosforilación de serina / treonina de IRS-1 y la reducción de la señalización de IRS-1, como anteriormente se ha descrito. TNF α aumenta la lipólisis de los adipocitos, que aumenta aún más los NEFA, y también tiene efectos negativos directos sobre las vías de señalización de insulina (Hotamisligil, 2000). La neutralización de TNF invierte sustancialmente la resistencia a la insulina en roedores; sin embargo, su participación en la resistencia a la insulina humana no estaba del todo clara, pero hay evidencias de que en humanos, el TNF α disminuye la respuesta celular a la insulina, y tiene efectos sobre el metabolismo del colesterol y las lipoproteínas (Ramírez Alvarado et al, 2012). La interleucina 6 proinflamatoria inhibe la señal de la insulina mediante el aumento de la expresión de proteínas SOCS (Krebs et al, 2003; Senn et al, 2003).

La adiponectina, adipoquina que posee efectos antidiabéticos también tiene un papel importante ya que, mientras los NEFA circulantes y varias adipoquinas se encuentran aumentadas en la obesidad visceral, las concentraciones de adiponectina se reducen, lo que supone una disminución de sus efectos sensibilizantes a la insulina en el hígado y en el músculo (Goldstein et al, 2003; Rajala et al, 2003). Las señales de adiponectina a través de la vía quinasa AMP podría explicar sus beneficiosos efectos metabólicos dado que esta vía de señalización activada por estrés, está implicada en diversas respuestas metabólicas, incluyendo la supresión de la gluconeogénesis hepática, la captación de

Introducción

glucosa en el músculo esquelético durante el ejercicio, la oxidación de ácidos grasos, y la inhibición de la lipólisis (Goldstein et al, 2004; Yamauchi et al, 2002). La quinasa AMP también ha sido implicada en el mecanismo de acción de la metformina (Zhou et al, 2001; Zou et al, 2004) y, posiblemente, de las tiazolidindionas (Fryer et al, 2002), lo que sugiere que tiene un papel importante en las respuestas a los antidiabéticos orales.

Actividad NFκB e IKK

Existe también una estrecha relación entre la resistencia a la insulina y las vías de señalización inflamatorias clásicas. En condiciones de reposo, el factor nuclear kappa B (NFκB) se encuentra inactivo mediante la unión a un inhibidor, IκB (Karin et al, 2000). La fosforilación de IκB por su quinasa (IKK) conduce a su degradación, liberándolo de NFκB para que éste se transloque al núcleo, donde puede intervenir en la transcripción de diversos genes implicados en la respuesta inflamatoria. De hecho, las dosis altas de salicilatos, que bloquean la actividad de IKK, (Yin et al, 1998) pueden mejorar la hiperglucemia y la resistencia a la insulina en la diabetes y la obesidad (Kim et al, 2001; Yuan et al, 2001). La inactivación de IKKβ debida a alteraciones genéticas normaliza la resistencia a la insulina del músculo esquelético causada por los NEFA, a través de la fosforilación de IRS-1 y la posterior activación de su cascada de señal (Shoelson et al, 2003). En general, estos resultados sugieren que IKK podría ser un objetivo importante para el desarrollo de nuevas terapias en resistencia a la insulina.

Además de sus efectos sobre la señalización de la insulina, las sustancias liberadas por el tejido adiposo afectan en gran medida a la función endotelial vascular, lo que enlaza el aumento del riesgo vascular propio del síndrome metabólico con los mecanismos celulares de resistencia a la insulina (Wellen et al, 2003; Goldstein et al, 2004). Estas sustancias también reclutan y activan las células inflamatorias, lo que finalmente puede conducir al desarrollo de aterogénesis (Wellen et al, 2003).

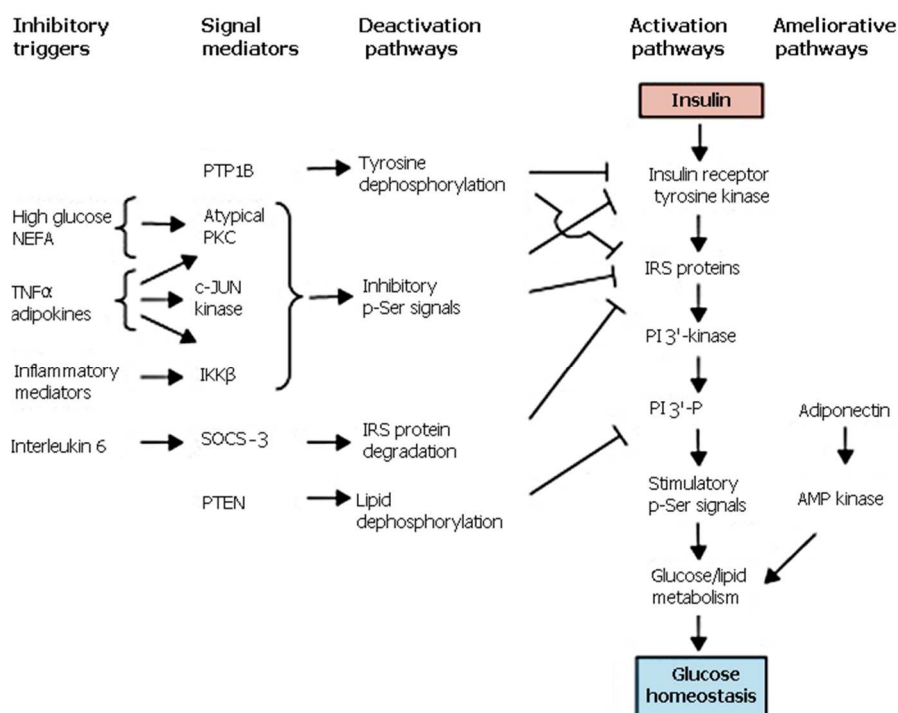


Figura 12: Señalización de la insulina y la resistencia a ésta. La señalización de la insulina implica la unión de ésta a su receptor seguida por una cascada de eventos intracelulares, representados como vías de activación. La modulación alterada de la acción de la insulina puede ser mediada a través de diversas vías, lo que conduciría a la resistencia a ésta. PKC = proteína quinasa; C PTEN = fosfatasa y homólogo de tensina; PI = fosfoinositol. Stumvoll et al. (2005). *Lancet*. 365, 1333-1346.

Metabolismo mitocondrial

La célula obtiene energía, principalmente en forma de ATP, a partir del poder reductor presente en las moléculas de glúcidos, lípidos y aminoácidos, entre otras. En la membrana interna mitocondrial se localiza la cadena respiratoria, que consta de una serie de transportadores de electrones. Su principal función es el transporte coordinado de protones y electrones, para producir un gradiente de protones que se utiliza para generar energía en forma de ATP, a partir de ADP y fosfato inorgánico.

La acumulación de triglicéridos en depósitos de grasa ectópicos (principalmente en el hígado) sugiere que existe un defecto en la oxidación lipídica mitocondrial en pacientes con diabetes tipo 2. En este sentido, estos pacientes padecen trastornos de la capacidad oxidativa y tienen mitocondrias más pequeñas en el músculo esquelético (Kelley et al, 2002). “PPAR γ co-activator 1” (PGC1, un factor de transcripción de genes implicados en

Introducción

la oxidación mitocondrial de ácidos grasos y la síntesis de ATP), se encuentra disminuido en sujetos con resistencia a la insulina, hijos de pacientes con diabetes tipo 2. Esto indica que un defecto hereditario en la fosforilación oxidativa mitocondrial podría conducir a la acumulación de lípidos a nivel celular (Petersen et al, 2004). Estudios de perfiles de expresión génica también han demostrado que la disminución de la expresión de PGC1 y productos de genes relacionados podría afectar la función mitocondrial en personas con resistencia a la insulina y diabetes tipo 2 (Mootha et al, 2003; Patti et al, 2003).

En el esquema de la figura 13 se resumen todos los mecanismos explicados que conducen a la patogénesis de la enfermedad:

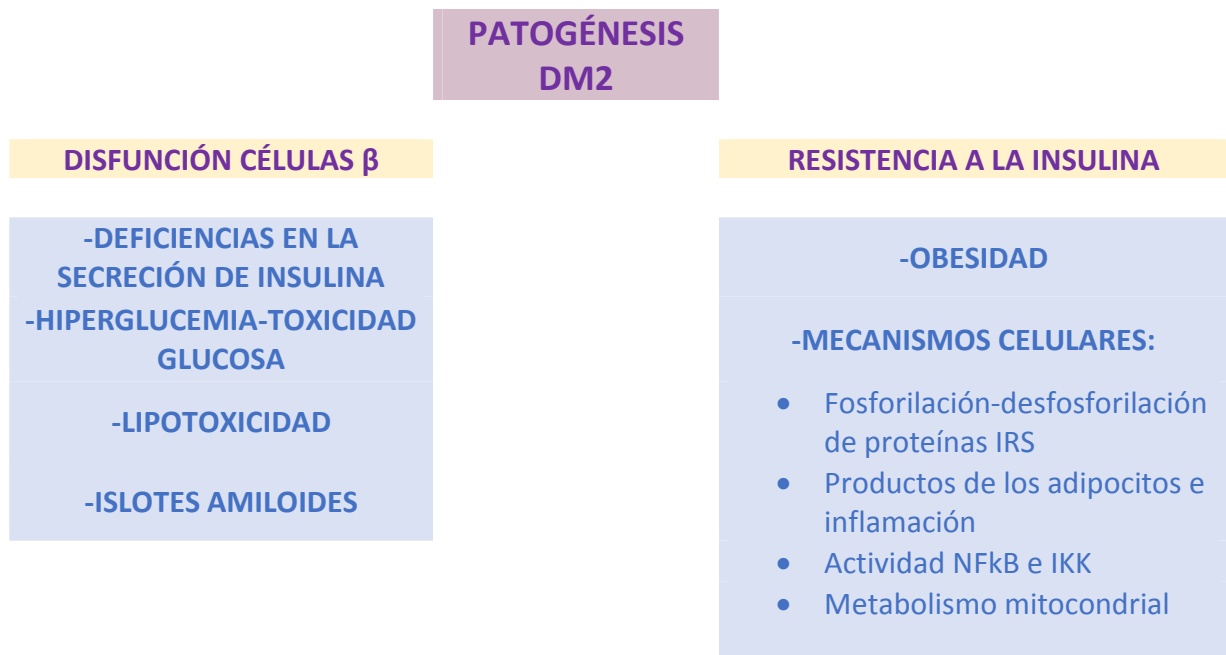


Figura 13: Resumen de los mecanismos implicados en la patogénesis de la DM2.

2.5- BASES GENÉTICAS DE LA DM2

La importancia de los factores genéticos en la DM2 comenzó a plantearse a partir de estudios realizados en gemelos, en los que se demostró una mayor incidencia de la enfermedad en gemelos homocigotos que en gemelos heterocigotos (Newman et al, 1987), así como en estudios que mostraban una mayor prevalencia de la enfermedad en ciertas poblaciones, como los indios Pima (Weir et al, 1994).

El estudio de los genes implicados en diabetes tipo 2 comenzó mediante el análisis de genes candidatos o de estudios de ligamiento en familias. La llegada de los GWAS a mediados de la década de 2000 permitió conocer muchos más genes asociados con la enfermedad, con decenas de loci identificados hasta la fecha. Simultáneamente comenzaron a desarrollarse técnicas de secuenciación de nueva generación, las cuales permiten obtener gran cantidad de información mediante la secuenciación de genomas o exomas.

Los análisis de genes candidatos se refieren normalmente al estudio de genes previamente relacionados con la enfermedad a estudiar. En los estudios de ligamiento, se realizan estudios de todo el genoma para identificar regiones asociadas a la enfermedad, principalmente mediante el estudio de microsatélites (o STRs, por sus acrónimos en inglés para “short tandem repeat”) en familias (Pulst, 1999; Feingold et al, 2001). Los marcadores que se segregan en los familiares que presentan la enfermedad, más frecuentemente de lo esperado, son utilizados para localizar el gen causante. Esta técnica ha sido muy útil, particularmente en los desórdenes monogénicos o de transmisión mendeliana. Sin embargo, estos estudios han sido menos útiles para encontrar genes asociados a enfermedades poligénicas y rasgos complejos. Esto es debido, quizás en parte, al poder limitado de la técnica para detectar herencias complejas.

Posteriormente llegaron los GWAS, los cuales aumentaron las posibilidades de identificar la influencia genética de los rasgos complejos y de las enfermedades comunes. Los estudios de GWAS utilizan cientos de miles o millones de SNPs como

Introducción

marcadores genéticos. Esta técnica se impuso debido a su gran capacidad para el análisis de gran número de polimorfismos a la vez.

El desarrollo técnico producido en los últimos años y la necesidad de técnicas de secuenciación de gran productividad y a un bajo coste, ha conducido al desarrollo de tecnologías de secuenciación que producen miles de millones de secuencias de forma simultánea (Schuster, 2008). La gran capacidad de estas técnicas ha supuesto también el desarrollo de herramientas informáticas para su análisis. La irrupción de las tecnologías de secuenciación de nueva generación en la genética molecular supone un gran avance en el estudio de ésta y permitirá superar muchas de las limitaciones de las estrategias utilizadas hasta el momento para determinar las causas genéticas de gran número de enfermedades.

Mediante técnicas clásicas, antes de la utilización de los GWAS, se identificaron algunos genes asociados con diabetes tipo 2, como son: “peroxisome proliferator-activated receptor gamma” (PPARG), “calpain 10” (CAPN10), y “potassium inwardly rectifying channel, subfamily J, member 11” (KCNJ11) (Altshuler et al, 2000; Horikawa, 2000; Gloyn et al, 2003). Los GWAS han permitido identificar numerosos genes asociados con la enfermedad, habiéndose identificado 62 loci hasta la fecha mediante estudios de este tipo (Mahajan et al, 2014). Varios de estos loci se han confirmado a través de varios GWAS posteriores u otros tipos de estudios de validación, funcionales, etc. Algunos de los genes inicialmente asociados a DM2 son: “CDK5 regulatory subunit associated protein 1-like 1” (CDKAL1), “solute carrier family 30 (zinc transporter) member 8” (SLC30A8), “homeobox hematopoietically expressed” (HHEX), “insulin-like growth factor 2 mRNA-binding protein 2” (IGF2BP2), y “cyclin-dependent kinase inhibitor 2A/2B” (CDKN2A/2B) (Burton et al, 2007; Saxena et al, 2007; Scott et al, 2007; Sladek et al, 2007; Zeggini et al, 2007).

El metaanálisis realizado por Zeggini y colaboradores en 2008, permitió identificar varios genes más: “juxtaposed with another zinc finger gene 1” (JAZF1), “cell division cycle 123, calcium/calmodulin-dependent protein kinase 1D” (CDC123, CAMK1D), “tetraspanin 8, leucine-rich repeat-containing G protein-coupled receptor 5” (TSPAN8, LGR5), “thyroid

adenoma associated" (THADA), "ADAM metallopeptidase with thrombospondin type 1 motif, 9" (ADAMTS9), y "notch gene homolog 2" (NOTCH2) (Zeggini et al, 2008).

Estudios posteriores permitieron identificar nuevas variantes asociadas a DM2 en genes tales como: "B-cell cl/lymphoma 11A" (BCL11A), "centaurin, delta-2" (CENTD2), "coiled-coil-helix-coiled-coil-helix domain 9" (CHCHD9), "dual-specificity phosphatase 9" (DUSP9), "high mobility group at-hook 2" (HMGA2), "hepatocyte nuclear factor-1-alpha" (HNF1A), "potassium voltage gated channel, KQT-like subfamily, member 1" (KCNQ1), "kruppel-like factor 14" (KLF14), "protein regulating cytokinesis 1" (PRC1), "tumor protein p53-inducible nuclear protein 1" (TP53INP1), "zinc finger BED domain-containing protein 3" (ZBED3), y "zinc finger, an1-type, domain-containing protein 6" (ZFAND6) (Voight et al, 2010).

GWAS en relación a variables relacionadas con DM2 han permitido identificar nuevos genes, principalmente los asociados a niveles de glucosa en ayunas. Por ejemplo, los GWAS de las concentraciones de glucosa en ayunas han identificado una variante del "melatonin receptor 1B" (MTNR1B) y posteriormente este gen también fue asociado con diabetes tipo 2 (Lyssenko et al, 2009; Prokopenko et al, 2009). Otro metaanálisis posterior sobre las concentraciones de glucosa en ayunas realizado por el "Meta-Analysis of Glucose and Insulin-related traits Consortium" (MAGIC) demostró asociación de los genes "adenylate cyclase-5" (ADCY5), "prospero-related homeobox-1" (PROX1), "glucokinase" (GCK), "glucokinase regulatory protein" (GCKR), y "diacylglycerol kinase (DGKB) transmembrane protein-195" (TMEM195) con DM2 (Dupuis et al, 2010).

GWAS sobre obesidad han demostrado claramente asociación del gen "fat mass and obesity-associated protein" (FTO) con obesidad (Frayling et al, 2007). Y, dado que la obesidad es un factor predisponente para el desarrollo de diabetes tipo 2, no es sorprendente que variantes en el gen FTO también se hayan encontrado asociadas en GWAS sobre DM2 (Burton et al, 2007).

Sin embargo, el gen que muestra la asociación más fuerte con diabetes tipo 2 en varios grupos étnicos (europeos, asiáticos y afrodescendientes) es el "transcription factor 7-

Introducción

like 2" (TCF7L2). La variante rs7903146 (siendo el alelo T el que confiere el riesgo de padecer la enfermedad) es común en la población, se encuentra en la región no codificante del gen y su presencia se traduce en un aumento, aproximadamente del 40% en el riesgo de padecer diabetes tipo 2 (Grant et al, 2006; Cauchi et al, 2007). No obstante, todavía no está claro cómo la variación en el locus TCF7L2 influye en la función del producto del gen ni cómo esta función alterada conduce a la patogénesis de la DM2.

En el último metaanálisis sobre DM2 (Mahajan et al, 2014), se han encontrado nuevos loci asociados y además, se ha intentado buscar representación de otros grupos étnicos, ya que la mayoría de GWAS siempre se habían llevado a cabo en población de origen europeo (Zeggini et al, 2008; Kong et al, 2009; Dupuis et al, 2010; Voight et al, 2010; Morris et al, 2012). A continuación, en la figura 14 (Mahajan et al, 2014) se muestra un Manhattan plot con los loci de susceptibilidad a DM2 previamente establecidos y los loci asociados en este metaanálisis. Además, en la tabla 3 (Mahajan et al, 2014) se explican de forma más detallada los nuevos loci encontrados en este estudio.

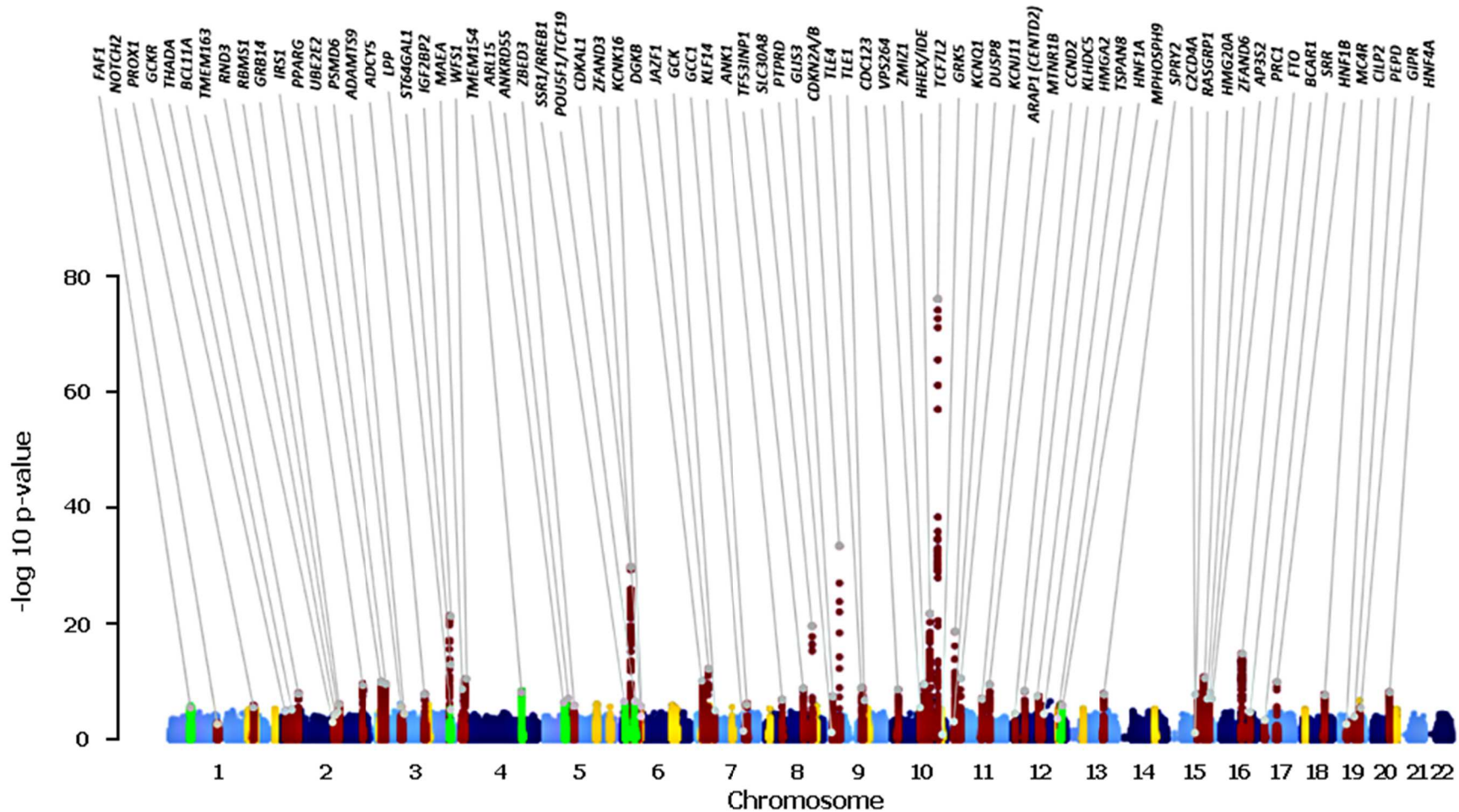


Figura 14: Manhattan plot de los resultados obtenidos en el metaanálisis. El metaanálisis recoge 26.488 casos y 83.964 controles DM2 de Europa, Asia Oriental, Asia del Sur, y de ascendencia mexicana y mexicoamericana. Se tomaron hasta 2,5 millones de SNPs autosómicos del proyecto HapMap. Los loci previamente establecidos como de susceptibilidad a DM2 se señalan en rojo. Los nuevos loci que lograron asociación nominal ($p < 5.10$) en la etapa 1 del meta-análisis, y son significativos en todo el genoma ($p < 5 \times 10^{-8}$) después de la validación con 21.491 casos y 55.647 controles de ascendencia europea, se representan en verde. Los loci que mostraron significación nominal en el metaanálisis, pero no alcanzaron significación en todo el genoma tras de la validación, están señalados en amarillo. Mahajan et al. Nat Genet. 2014 Mar;46(3):234-44.

Locus	Lead SNP	Chr	Build 36 position (bp)	Alleles ^a		Trans-ethnic discovery meta-analysis		European ancestry validation meta-analysis		Combined meta-analysis	
				Risk	Other	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P
TMEM154	rs6813195	4	153,739,925	C	T	1.08 (1.05–1.11)	4.2 × 10 ⁻⁹	1.08 (1.05–1.11)	2.0 × 10 ⁻⁶	1.08 (1.06–1.10)	4.1 × 10 ⁻¹⁴
SSR1-RREB1	rs9505118	6	7,235,436	A	G	1.06 (1.04–1.09)	1.9 × 10 ⁻⁶	1.06 (1.03–1.09)	1.7 × 10 ⁻⁴	1.06 (1.04–1.08)	1.4 × 10 ⁻⁹
FAF1	rs17106184	1	50,682,573	G	A	1.11 (1.07–1.16)	1.9 × 10 ⁻⁶	1.09 (1.04–1.15)	4.8 × 10 ⁻⁴	1.10 (1.07–1.14)	4.1 × 10 ⁻⁹
POU5F1-TCF19	rs3130501	6	31,244,432	G	A	1.07 (1.04–1.10)	1.5 × 10 ⁻⁶	1.06 (1.03–1.10)	7.0 × 10 ⁻⁴	1.07 (1.04–1.09)	4.2 × 10 ⁻⁹
LPP	rs6808574	3	189,223,217	C	T	1.08 (1.04–1.11)	4.3 × 10 ⁻⁶	1.06 (1.03–1.09)	2.6 × 10 ⁻⁴	1.07 (1.04–1.09)	5.8 × 10 ⁻⁹
ARL15	rs702634	5	53,307,177	A	G	1.08 (1.05–1.11)	3.4 × 10 ⁻⁷	1.05 (1.02–1.08)	2.1 × 10 ⁻³	1.06 (1.04–1.09)	6.9 × 10 ⁻⁹
MPHOSPH9	rs4275659	12	122,013,881	C	T	1.06 (1.03–1.09)	5.5 × 10 ⁻⁶	1.06 (1.02–1.09)	4.4 × 10 ⁻⁴	1.06 (1.04–1.08)	9.5 × 10 ⁻⁹

Tabla 3: Nuevos locus en que se encontró asociación con DM2 ($p < 5 \times 10^{-9}$).

Los locus de susceptibilidad a la DM2 mostrados han sido identificados a través de GWAS de 26,488 casos y 83,964 controles de Europeos, Asiáticos del este y del sur de Asia, Mexicanos y Mexicanos de ascendencia americana con el seguimiento en una fase de validación del metaanálisis con 21,491 casos y 55,647 controles de ascendencia europea genotipados con el MetaboChip. (Chr, cromosoma; OR, odds ratio; IC, intervalo de confianza. ^aLos alelos están alineados con la cadena forward del NCBI Build 36). Mahajan et al. (2014). *Nat Genet.* 46, 234-244.

Los nuevos loci incluyen SNPs que mapean cerca del gen POU5F1-TCF19 en el complejo mayor de histocompatibilidad, región del genoma esencial para la respuesta inmune. Esta región alberga los genes HLA de clase II, asociados con el riesgo de padecer diabetes mellitus tipo 1 (Bradfield et al, 2011). Además, como se ha mencionado previamente, se ha visto que variantes implicadas en DM1, están también asociadas con LADA (Cervin et al, 2008; Grant et al, 2010).

También se incluyen regiones que mapean en los genes ARL15 y SSR1-RREB1, que ya se habían asociado previamente con la regulación de los niveles de insulina en condiciones de ayuno y con los niveles de glucosa en ayuno, respectivamente (Scott et al, 2012). El locus ARL15 había sido asociado también con los niveles circulantes de adiponectina, proteína secretada por los adipocitos y que posee efectos antidiabéticos (Richards et al, 2009). Por otra parte, los alelos de riesgo para DM2 en SSR1-RREB1 y LPP indican un papel primordial en la susceptibilidad a la DM2 a través de la disfunción de las células β .

No obstante, es importante destacar que, a pesar de todos los estudios que se han hecho para establecer las bases genéticas de la DM2, la información que existe hasta el momento permite explicar solamente el 10-15% del componente genético de la enfermedad. Por ello es esencial seguir profundizando en la enfermedad y establecer nuevas estrategias de estudio, como por ejemplo el uso de las tecnologías de secuenciación de nueva generación.

3- SECUENCIACIÓN DE NUEVA GENERACIÓN

Tras completar la secuenciación del primer genoma humano en 2004, la creciente necesidad de secuenciar el genoma de mayor número de individuos y, ha inducido el cambio desde la secuenciación de Sanger hasta las tecnologías de alto rendimiento. En 2005, emergía la primera plataforma de secuenciación masiva de ADN, dando lugar al comienzo de la era de la secuenciación masiva (Margulies et al, 2005; Shendure et al, 2005).

Hasta la fecha, el desarrollo de las tecnologías de secuenciación ha permitido la generación de gran cantidad de datos. Estas nuevas tecnologías se han empezado a emplear para diferentes aplicaciones como: descubrimiento de mutaciones causantes de enfermedades hereditarias mediante secuenciación de genoma completo, identificación de marcadores epigenéticos, secuenciación de determinadas regiones, etc. Además, la simplificación de los procesos tanto de secuenciación como de análisis, está haciendo que la secuenciación de nueva generación (NGS, por sus siglas en inglés “Next Generation Sequencing”) sea cada vez más empleada, y, que esté desplazando a los GWAS en gran medida (Buermans et al, 2014).

Todas las plataformas de secuenciación de nueva generación requieren un riguroso proceso de preparación de librerías, que va desde la fragmentación del ADN hasta la ligación de adaptadores que identifican el ADN de cada muestra, la cuantificación de las librerías, la mezcla de muestras y, finalmente la secuenciación. Para cada una de las diferentes plataformas de secuenciación se han diseñado diferentes estrategias para la preparación de librerías, así como para la detección de la señal y en última instancia, la lectura de las secuencias de ADN:

- En los sistemas de segunda generación (Illumina, SOLID y GS454) se requiere una amplificación clonal de las moléculas iniciales en clusters (Mitra et al, 1999) para así aumentar el ratio señal-ruido de fondo, ya que los sistemas de secuenciación no son lo suficientemente sensibles como para detectar la extensión de una base en una sola molécula.

- Y en los de tercera generación, HeliScope y PacBio SMRT no necesitan ningún paso de pre-amplificación, ya que son lo suficientemente sensibles para detectar extensiones a nivel de molécula individual. Las diferentes estrategias para generar las secuencias también conducen a diferencias en la cantidad de datos generados.

3.1- ROCHE 454

El sistema de secuenciación 454 se basa en la combinación de la tecnología de PCR en emulsión y la pirosecuenciación en placa de nanopocillos (Margulies et al, 2005). En la PCR en emulsión, el molde de cadena sencilla, se encuentra unido a las denominadas “beads” y se limita a microesferas de emulsión individuales en las que se producen millones de copias de cada molde mediante amplificación por PCR (debería haber una sola molécula de ADN molde por bead). Las beads donde se produce amplificación son enriquecidas posteriormente y se depositan en pocillos individuales de una placa PicoTiter™ donde se lleva a cabo la pirosecuenciación. En este método de secuenciación por síntesis, la emisión de luminiscencia, procedente de la liberación de pirofosfato en la incorporación de nucleótidos dirigida por el molde, se controla en tiempo real y en cada nanopocillo de la placa.

El sistema 454 permite obtener lecturas largas de alta calidad. La última plataforma 454GSFLX puede producir aproximadamente un millón de lecturas con longitudes de hasta 1.000 pb por carrera. Debido a esta ventaja, a pesar de los costes más altos en comparación con otras plataformas de NGS, la plataforma 454 es la más apropiada para diversas aplicaciones, incluyendo el ensamblaje *de novo* (Gilles et al, 2011) y la metagenómica (Loman et al, 2012). Sin embargo, la tecnología de 454 tiene un problema inherente a la pirosecuenciación que ocasiona problemas en la detección de homopolímeros. Debido a la falta de un terminador en cada ciclo de incorporación de una base, se dan múltiples incorporaciones de nucleótidos idénticos en las regiones donde varias bases consecutivas son iguales. Esto puede conducir a la no linealidad entre la intensidad de la señal y la longitud de los tramos de homopolímero cuando cuatro o

Introducción

más nucleótidos se incorporan de forma consecutiva. En consecuencia, el sistema 454 tiene una tasa de error relativamente alta para detectar inserciones y deleciones (indels) en homopolímeros largos (Wommack et al, 2008).

El flujo de trabajo de secuenciación completa del sistema 454 tiene cuatro pasos principales:

- Generación de una librería de ADN molde de cadena sencilla
- Amplificación clonal a base de PCR en emulsión de la librería
- Generación de datos a través de la secuenciación por síntesis
- Análisis de datos utilizando diferentes herramientas bioinformáticas

En la figura 15 se muestra el esquema de trabajo que sigue este sistema:

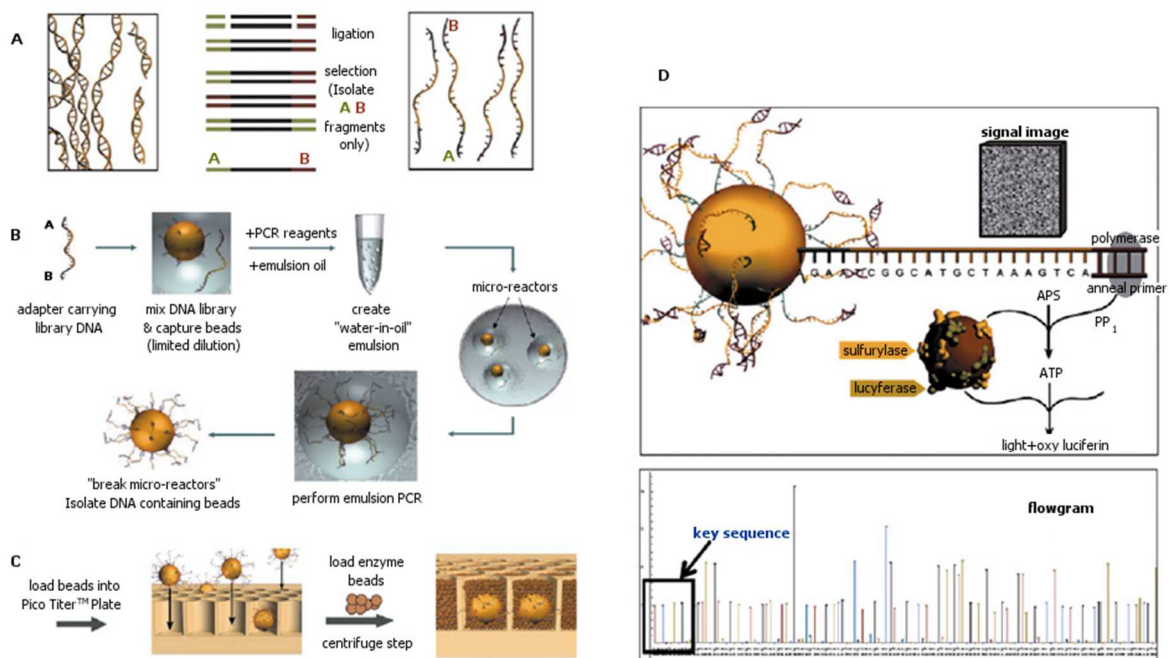


Figura 15: Esquema de trabajo del Sistema 454. A) Preparación de librería. B) Amplificación clonal. C) Carga de las microesferas de ADN en la placa PicoTiter™. D) Secuenciación. (<http://www.454.com>)

3.2- ILLUMINA

Todos los procesos enzimáticos y de creación de imágenes de la tecnología Illumina tienen lugar en una célula de flujo. Dependiendo de la plataforma, ésta célula de flujo está dividida en: 1 (MiSeq), 2 (HiSeq2500) u 8 (HiSeq2000, HiSeq2500) canales separados. Illumina utiliza la amplificación en puente para la generación de clusters y la posterior secuenciación por síntesis (SBS, por sus siglas del inglés, “sequencing by synthesis”) (Bentley et al, 2008).

El primer paso para cargar la librería en la célula de flujo, es la desnaturalización de los fragmentos de ADN de doble cadena en moléculas de ADN individuales. Cuando están en la superficie de la célula de flujo, estos se hibridan con los oligonucleótidos (complementarios a uno de los adaptadores presente en la librería) que se utilizan como cebadores para formar una copia inicial de la molécula de muestra individual a secuenciar, que queda unida de forma covalente a la superficie de la célula de flujo. Las moléculas de librería iniciales se eliminan y los fragmentos unidos a la célula de flujo se utilizan para generar un conjunto de copias mediante amplificación isotérmica. Cada molécula generada mediante el proceso anterior se hibrida con los oligos distribuidos en la misma superficie y que son complementarios al extremo 3' de los fragmentos (región donde se localiza el otro adaptador presente en la librería) formando la denominada “estructura de puente”. A partir de ahí se produce la extensión o copia del fragmento. Estos pasos se repiten de forma sucesiva mediante la alternancia cíclica de tres tampones específicos que median los pasos de: desnaturalización, anillamiento del oligo y extensión. El último paso consiste en eliminar una de las cadenas de los fragmentos de ADN de doble cadena, utilizando el sitio escindible en el oligo de la superficie y bloquear todos los extremos 3' con didesoxinucleótidos (ddNTPs) para evitar que por otra parte los extremos 3' abiertos actúen como cebadores de secuenciación en moléculas de librería adyacentes (Bentley et al, 2008).

El método de amplificación en puente no es un método muy eficiente para la amplificación clonal ya que, los 35 ciclos de amplificación isotérmica producen apenas ~1.000 copias de la molécula inicial. Por otra parte, se dará predominantemente un crecimiento de los clusters hacia el exterior y existe una alta probabilidad de que las

Introducción

hebras del molde se rehibriden entre ellas, en lugar de hibridarse a un nuevo oligo. Esto supone limitaciones en el tamaño de las moléculas de molde que se amplifican de manera realmente fiable.

Posteriormente, durante la secuenciación, los clusters en la célula de flujo se leen nucleótido por nucleótido en ciclos repetitivos. Durante estos ciclos, ddNTPs marcados con fluorescencia se incorporan en la cadena de ADN en crecimiento. Cada una de las cuatro especies de ddNTPs (A, C, T, G) tiene un fluoróforo diferente y único que sirve para identificar la base y actuar como terminador reversible evitando así múltiples eventos de extensión. Después de la obtención de imágenes, el fluoróforo es escindido, el terminador reversible se desactiva y las hebras del molde están listas para el siguiente ciclo de incorporación. La secuencia se lee siguiendo la señal fluorescente por cada paso de extensión para cada cluster. Bajo circunstancias ideales, todas las bases dentro de un cluster se extenderán en una fase. Sin embargo, una pequeña porción de las moléculas no se extienden adecuadamente. Durante muchos ciclos, estos errores se acumulan y disminuye la relación señal-ruido por cluster, causando una disminución en la calidad hacia los extremos de las lecturas.

A continuación, en la figura 16 se muestra un esquema del proceso completo:

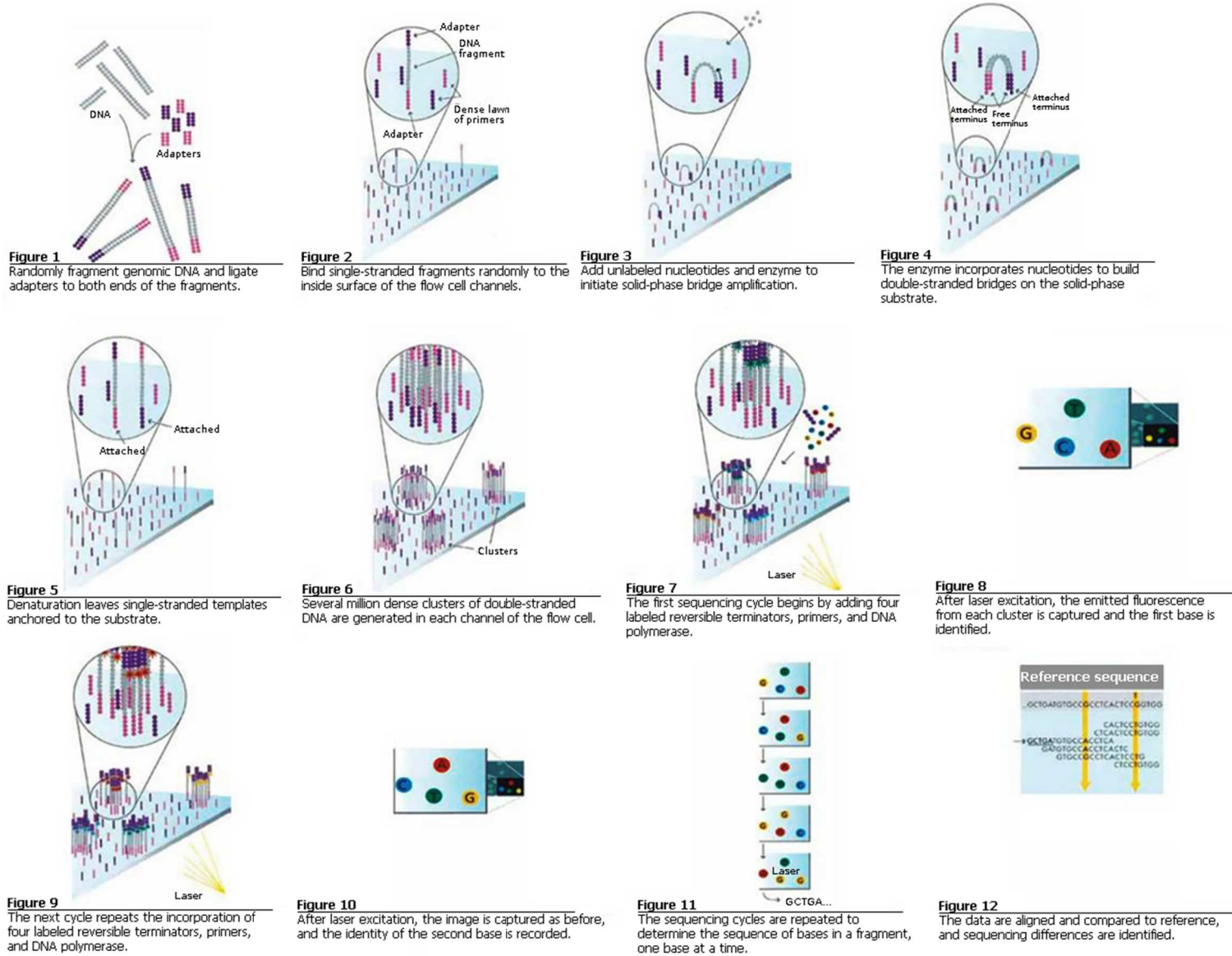


Figura 16: Esquema de trabajo de la plataforma Illumina. (<http://www.illumina.com>)

Introducción

La duración de un ciclo para HiSeq2000 es de aproximadamente 1h. Lo que más tiempo conlleva es la toma de imágenes de la célula de flujo, por lo que al reducir el tiempo de producción de imágenes, el proceso de secuenciación se puede acelerar considerablemente. Esto se ha implementado en las plataformas MiSeq, NexSeq500 y HiSeq2500 proporcionando la opción de reducir el área de superficie total a ser fotografiado. Así, el tiempo de ciclo se puede reducir a 5- 10 min en estos sistemas.

A principios de 2014, Illumina lanzó al mercado dos nuevos modelos: NextSeq500 y HiSeqXTen. El primero proporciona un rendimiento similar al HiSeq 2500, pero tiene un tamaño inferior, proporcionando un output medio (40 Gb) y alto (120 Gb) y con tiempos por carrera de menos de 30h. El HiSeqXTen, por su parte, permite la secuenciación del genoma humano con un coste menor. Esto se debe a la introducción de un nuevo tipo de células de flujo, las cuales contienen una rejilla con un formato de nano-pozos, en que cada uno puede producir un cluster. Esto permite obtener densidades de cluster óptimas y, en combinación con protocolos de exploración más rápidos, un software diseñado específicamente y una nueva química de secuenciación, este sistema puede producir ~600GB de datos en un solo día o ~1.8 terabases en un plazo de tres días, lo cual puede dar lugar a la secuenciación de 5 y 15 genomas humanos completos, respectivamente. Además de estos modelos, Illumina también ha introducido recientemente en el mercado los sistemas HiSeq 3000 y HiSeq 4000 los cuales proporcionan mayor rendimiento por carrera, a una velocidad mayor que sus antecesores.

3.3- SOLID

El sistema SOLID utiliza la tecnología de secuenciación por ligación. La librería se prepara mediante PCR en emulsión, de forma similar al protocolo del sistema 454 explicado previamente. La secuenciación se realiza a través de ciclos sucesivos de ligación, en los que cada cebador de secuenciación se liga a una sonda específica marcada con fluorescencia de acuerdo a la complementariedad entre las di-bases de la sonda y la muestra. Como cada cuatro di-bases (por ejemplo, AG, GA, TC, CT) están marcadas con

uno de los cuatro fluoróforos, los di-nucleótidos en las mismas posiciones de cada muestra se asocian con un fluoróforo único. A través de los ciclos de ligación, los di-nucleótidos se leen a intervalos de cinco bases, es decir, di-nucleótidos en las posiciones 4-5, 9-10, 14-15, 19-20 y así sucesivamente. Después de cinco rondas de ligación, cada nucleótido en la muestra se lee dos veces por dos señales fluorescentes, lo que aumenta la precisión de la llamada de bases. Entre las plataformas de secuenciación actuales, el sistema SOLiD es el que presenta la menor tasa de error. Su error más común es la sustitución. Además, también se ha demostrado que presenta una baja capacidad de representación de las regiones ricas en AT (Harismendy et al, 2009).

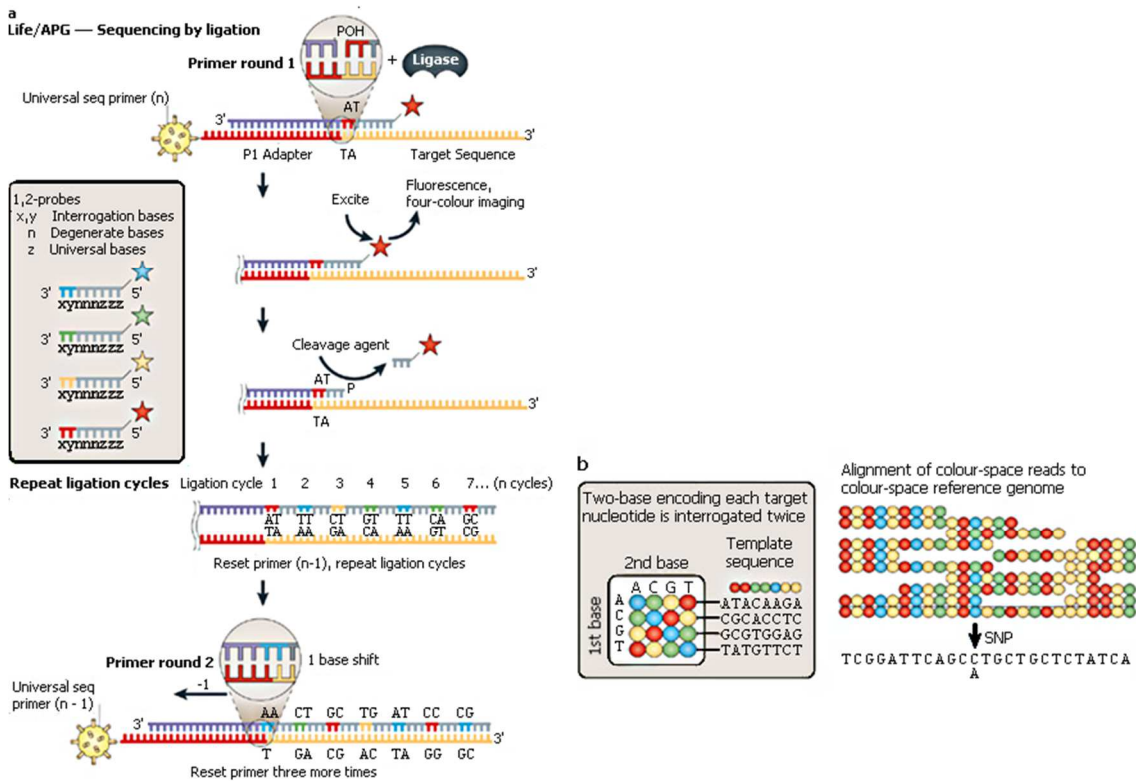


Figura 17: Esquema de funcionamiento del Sistema SOLiD. a) Secuenciación por ligación. b) Alineamiento de las secuencias. Metzker. (2010). *Nature Reviews Genetics*. 11, 31-46.

3.4- ION TORRENT

Por otro lado están las plataformas de Lifetech, Ion Torrent PGM e Ion Torrent Proton que son diferentes de otras plataformas NGS, en el sentido de que miden los cambios de pH, en lugar de la fluorescencia emitida durante la secuenciación.

Ion Torrent PGM posee un tiempo de respuesta más rápido que otros métodos actuales basados en nucleótidos marcados con fluorescencia y en la posterior formación de imágenes. El principio subyacente de Ion Torrent PGM es la detección de iones de hidrógeno liberados con la adición secuencial de desoxinucleótidos a una cadena creciente de ADN, lo que podría resultar en una alternativa algo más económica a las plataformas actuales de detección de fluorescencia (Merriman et al, 2012). Aunque Ion Torrent PGM produce megabases de secuencias de ADN en cuestión de horas, es insuficiente para generar datos de secuenciación de genoma o exoma completos con una cobertura adecuada por base. En 2012, Life Technologies lanzó el Ion Torrent Proton, el cual fue concebido para generar una cantidad mayor de datos, del orden de gigabases, lo que permite realizar estudios de alta calidad de secuenciación de exoma, transcriptoma y, finalmente, la secuenciación de todo el genoma de forma relativamente rápida (Boland et al, 2013).

En la figura 18 se muestra un esquema sobre el funcionamiento de estos sistemas:

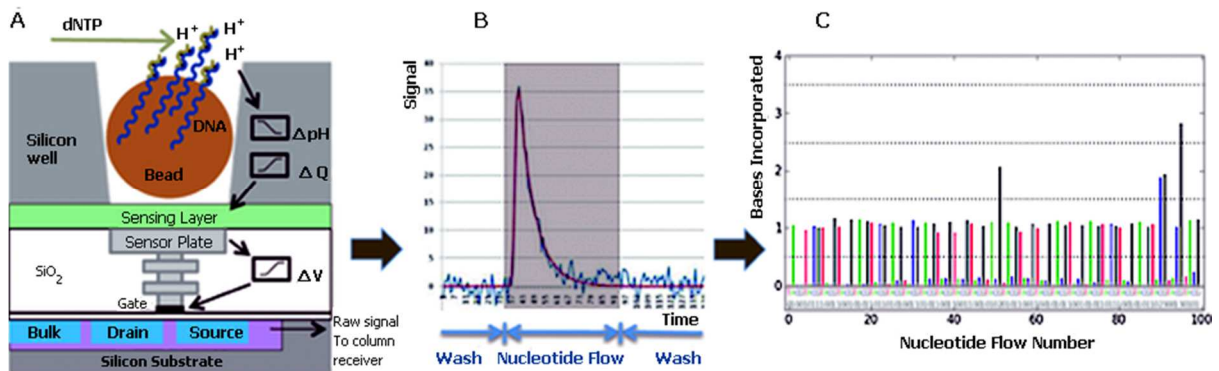


Figura 18: Descripción general de la secuenciación utilizando un sensor de pH semiconductor. (A) La muestra, en forma de esfera de ADN se encuentra en un pocillo muy por encima del sensor CMOS (complementary metal-oxide semiconductor) pHFET (pH sensitive field effect transistor). La muestra tiene una sola hebra, se le une el cebador para la extensión, y luego la polimerasa. Conforme se une un dNTP específico (N = A, C, G o T), se libera un protón si el nucleótido se incorpora como la siguiente base en esta reacción de síntesis. La liberación de protones resultante (Delta pH) se convierte en una acumulación de carga positiva (Q) en la capa de detección, que el conductor transmite como un cambio de voltaje (V) al transistor. La señal resultante se envía al chip (B) Estas señales de pH se toman del pocillo que contiene la muestra a alta frecuencia (curva azul), por ejemplo, 100 Hz, y luego se ajustan a un modelo que extrae la señal neta (curva roja). (C) Las señales de incorporación neta para cada flujo se convierten en el input primario, lo que refleja no incorporación, la incorporación de 1-mer, incorporación 2-mer, etc. Estos datos brutos son el input para el algoritmo que convierte estas señales de incorporación en las secuencias correspondientes. Merriman B et al. (2012). *Electrophoresis*, 33, 3397-3417.

3.5- TECNOLOGÍAS EMERGENTES

La aparición de la tecnología de secuenciación de molécula única (“single-molecule sequencing technology”) ha originado un gran salto en la evolución de la secuenciación de nueva generación. El hecho de que permita secuenciar moléculas sencillas de ADN o moléculas de ARN en muestras biológicas sin amplificación previa, la convierte en una tecnología realmente innovadora. Esta estrategia de secuenciación proporciona ventajas significativas frente a las tecnologías actuales de NGS ya que se minimiza la manipulación de la muestra, se reducen sus requisitos (cantidad de muestra...), se evitan los errores inducidos por la amplificación, aumenta la flexibilidad en la longitud de lectura y se permite la cuantificación exacta de moléculas de ácido nucleico. La simplicidad, sensibilidad y las capacidades cuantitativas de este tipo de secuenciación la hacen muy prometedora para el diagnóstico molecular (Milos, 2009). Además, esta

Introducción

tecnología evita sesgos en regiones, por ejemplo en zonas ricas en GCs, los cuales sí que se producen en todas las plataformas descritas anteriormente.

El Sistema de Análisis Genético Helicos (figura 18) fue la primera plataforma de secuenciación de este tipo disponible en el mercado (Harris et al, 2008). En este sistema, colas de poli (A) de la muestra de ADN de cadena sencilla son capturadas por oligonucleótidos poli (T) que actúan como cebadores y que se encuentran unidos a la superficie de una célula de flujo. La secuenciación se realiza a través de ciclos de extensión de una sola base mediados por la ADN polimerasa, usando una serie de cuatro nucleótidos terminadores fluorescentes virtuales. Cada uno de ellos representa un terminador reversible 3'-no bloqueado con un resto inhibidor marcado con un fluoróforo (Bowers et al, 2009). El secuenciador, con dos células de flujo de 25 canales, es capaz de capturar miles de millones de moléculas de ADN individuales y generar más de 21 a 35 Gb de datos de secuencias con una longitud media de lectura de 35 pb. Aunque el proceso de secuenciación es asíncrono en este sistema, no se dan efectos de desfase debidos a la amplificación, los cuales sí que existen en otras plataformas de secuenciación. Por otra parte, no existe un sesgo en el contenido de GC en las lecturas. Sin embargo, la tasa de error de lectura actual en Helicos es relativamente alta (~ 3-5%), y el tipo de error dominante es la eliminación, que resulta presumiblemente de la incorporación de nucleótidos no marcados y/o errores de detección. El uso de la química del Terminador Virtual resuelve el problema en la secuenciación de homopolímeros, y la incorporación base-por-base, resulta en tasas de error de sustitución muy bajas (generalmente 0.2%). Cuando se aplica una estrategia de dos pasos, en la que moléculas de molde individuales se secuencian dos veces, las tasas de error pueden reducirse todavía más. En la figura 18 se muestra un esquema del proceso:

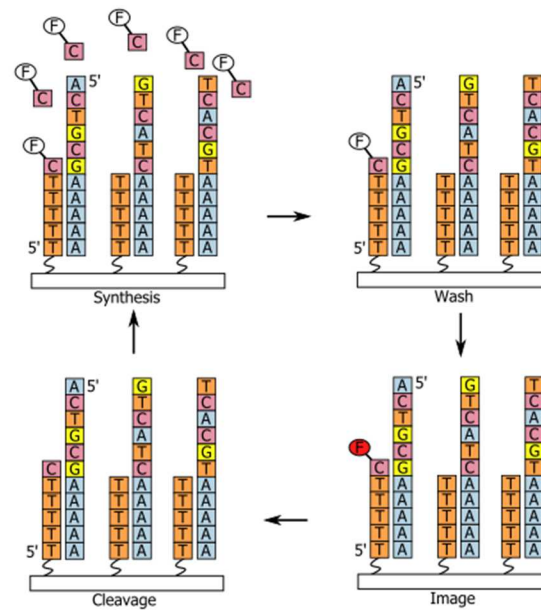


Figura 18: Esquema del funcionamiento de Helicos. Gupta PK. (2008). *Trends Biotechnol.* 26, 602-611.

También están surgiendo otras tecnologías de secuenciación de una sola molécula que dan lugar a mayores longitudes de lectura y a mayor velocidad de secuenciación. Un ejemplo es PacBio RS (figura 19), que es un sistema de secuenciación en tiempo real de una sola molécula desarrollado por Pacific Biosciences (Eid et al, 2009). En este sistema, una sola molécula de ADN polimerasa unida al molde es inmovilizada en la parte inferior de la llamada “zero-mode waveguide”, que funciona como una cámara de visualización nanofotónica para el seguimiento de la reacción de polimerización en un volumen de detección del orden de zeptolitros (10^{-21}). Durante la secuenciación, la incorporación de cuatro nucleótidos fluorescentes unidos a fosfato en la creciente cadena complementaria se registra ópticamente a tiempo real. El fluoróforo unido al resto fosfato terminal de cada nucleótido se elimina de forma natural por escisión enzimática después de la incorporación. Esto permite la síntesis de ADN mediante la polimerasa, dando lugar a secuencias de miles de bases. No obstante, el sistema PacBio ofrece actualmente un rendimiento de aproximadamente 50-100 Mb por carrera, es decir mucho más bajo que otras plataformas de NGS actuales. Además, la tasa de error de lectura es del 15%, superior a la que presentan otras tecnologías.

Introducción

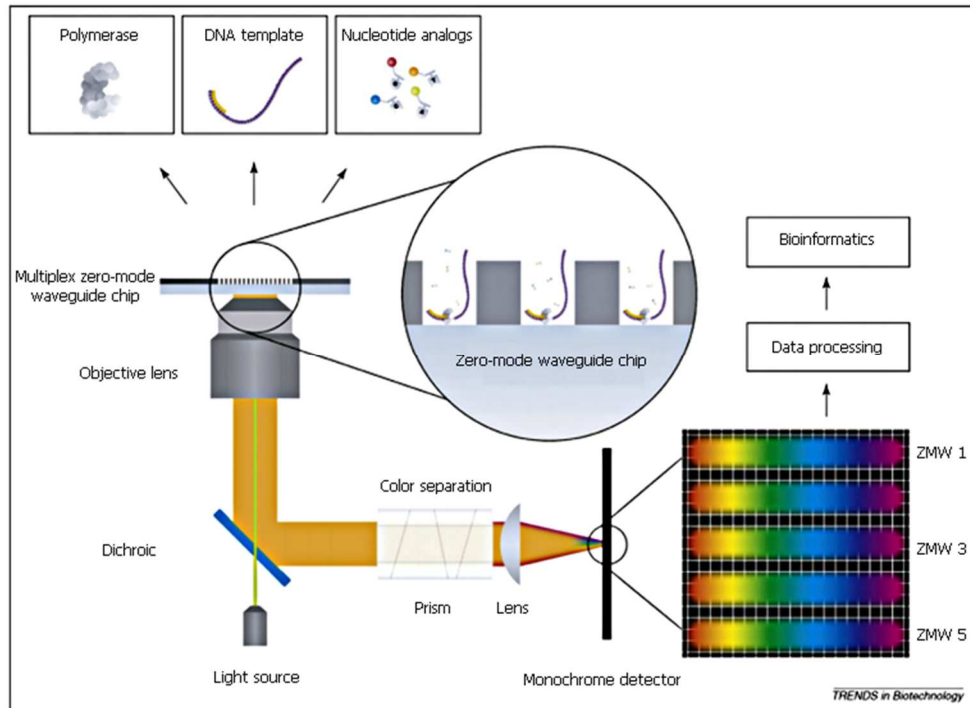


Figura 19: Representación esquemática del funcionamiento del sistema Pacific Bio. Gupta PK. (2008). *Trends Biotechnol.* 26, 602-611.

Otro ejemplo son las tecnologías de secuenciación por nanoporos, en el que las moléculas de ácido nucleico de una sola cadena son conducidas por electroforesis a través de un poro de tamaño nanométrico y se detectan por su efecto sobre una corriente iónica o señal óptica (Branton et al, 2008). Este tipo de secuenciación (figura 20) da lugar a lecturas de longitudes largas (de hasta decenas de kilobases), a requisitos mínimos de los reactivos y preparación de la muestra y, a una alta velocidad de secuenciación. Sin embargo, quedan varios problemas por resolver antes de que este tipo de secuenciación pueda aplicarse de forma más rutinaria. La alta velocidad de translocación del ADN a través de nanoporos hace que sea difícil distinguir las señales base de los ruidos de fondo. El movimiento aleatorio de las moléculas durante la translocación también se suma a la dificultad de alcanzar una resolución de una sola base.

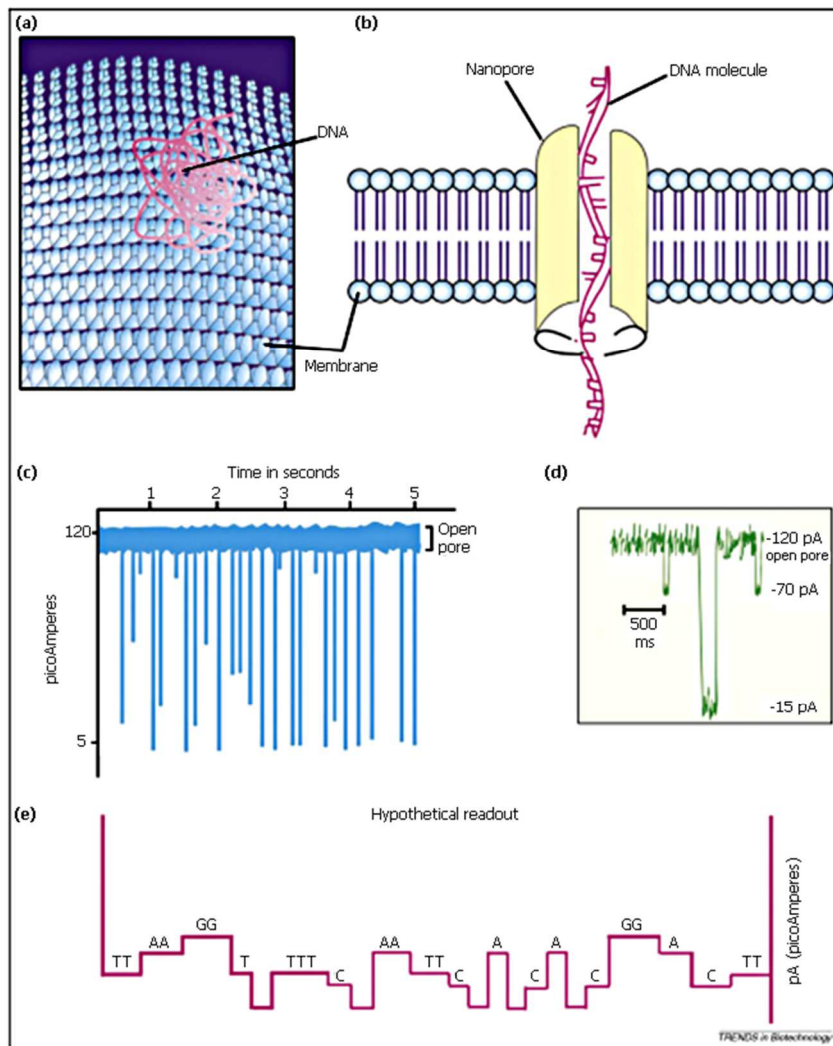


Figura 20: Representación esquemática de la secuenciación por nanoporos. (a) Vista de la superficie de una membrana con numerosos nanoporos. (b) Sección transversal de una parte de la membrana, con una molécula de ADN monocatenario atravesando un nanoporo. (c) Ejemplo de un perfil de pico obtenido como resultado de alteraciones que se producen cuando el ADN está pasando a través de un nanoporo. (d) Hipotético primer plano de picos que muestran los cambios esperados en la corriente que se derivarían de un ADN con segmentos de homopolímero, que se indican por la presencia de tres picos distintos. (e) Ejemplo hipotético de una secuencia de lectura que demuestra la resolución de una sola base de la técnica basada en diferencias menores pero detectables. Gupta PK. (2008). *Trends Biotechnol.* 26, 602-611.

Una limitación de todas estas plataformas, sería el uso de ADN polimerasas, a pesar de que, al poseer la función exonucleasa, éstas pueden corregir errores (si una base se incorpora mal, distorsiona la molécula, por lo que se detecta el error, retrocede y lo corrige). La tasa de error es del orden de 10^{-5} , pero con la corrección desciende hasta 10^{-8} . Aun así, el uso de ADN polimerasas para las amplificaciones supone que la muestra se puede sesgar en base a las especificidades de las mismas e introducirse errores en la secuencia.

3.6-COMPARACIÓN DE LAS DIFERENTES PLATAFORMAS DE SECUENCIACIÓN.

Los costes del proceso de secuenciación, son importantes para el usuario, pero a la vez difíciles de estimar. Los sistemas Illumina HiSeq 2000 o 2500 constituían la herramienta más rentable para la secuenciación de genoma hasta hace un tiempo, pero su mayor inconveniente era que cada carrera tenía una duración de aproximadamente 8 días. Actualmente existen los sistemas, HiSeq 3000 y HiSeq 4000, los cuales generan una mayor cantidad de datos por carrera, en un tiempo menor (unos 3 días y medio). En los últimos dos años, la esperada tasa de disminución en el coste unida a un aumento en el rendimiento del proceso de secuenciación, no se ha mantenido debido a la falta de competencia, por ello Illumina ha seguido dominando el mercado de la secuenciación. Plataformas de secuenciación más nuevas, como las tecnologías basadas en nanoporos, todavía tienen que demostrar su viabilidad en términos de rendimiento y precisión. En los próximos dos o tres años, es probable que Illumina siga dominando el mercado, de hecho, ya ha desarrollado sistemas como el HiSeq X Ten, el cual presenta la ventaja de que el coste por genoma a una cobertura del 30x está por debajo de 1000 dólares. Además, de todo el coste de reactivos, preparación y secuenciación, hay que tener en cuenta que el coste para el almacenamiento, análisis e interpretación de los datos de secuenciación es incluso más alto que el coste de generar los mismos.

Por tanto, la elección de un sistema u otro dependerá de la disponibilidad económica, pero también de las necesidades del proyecto, es decir, de la cantidad de datos que quieran generarse. Por ejemplo si se desean secuenciar genomas o exomas, el sistema más eficiente sería Illumina, pero en el caso de amplicones, sería suficiente con sistemas como el 454. A continuación, en la tabla 4 se comparan las diferentes características de las principales plataformas de secuenciación presentes en el mercado:

		Secuenciación por	Detección	Tipos de carrera	Tiempo de carrera	Longitud de lectura (pb)	Lecturas por carrera	Output por carrera
Roche	GS FLX Titanium XL+	Síntesis	Detección de pirofosfato	Single end	23 h	700	1 millones	700 Mb
	GS Junior System	Síntesis	Detección de pirofosfato	Single end	10 h	400	0.1 millones	40 Mb
LifeTechnologies	Ion torrent	Síntesis	Liberación de protones	Single end	4 h	200–400	4 millones	1.5–2 Gb
	Proton	Síntesis	Liberación de protones	Single end	4 h	125	60–80 millones	8–10 Gb
	Abi/solid	Ligación	Fluorescencia; detección de di-bases	Single & paired-end	10 días	75 + 35	2.7 billones	300 Gb
Illumina/solexa	HiSeq2000/2500	Síntesis	Fluorescencia; terminadores reversibles	Single & paired-end	12 días	2 × 100	3 billones	600 Gb
	MiSeq	Síntesis	Fluorescencia; terminadores reversibles	Single & paired-end	65 h	2 × 300	25 millones	15 Gb
Pacific biosciences	RSII	Síntesis de una sola molécula	Fluorescencia; terminadores unidos a fosfato	Single end	2 días	50% de lecturas > 10 kb	0.8 millones	5 Gb
Helicos	Heliscope	Síntesis de una sola molécula	Fluorescencia; terminadores virtuales	Single end	10 días	~ 30	500 millones	15 Gb

Tabla 4: Sumario de las diferentes plataformas de secuenciación. Adaptado de Buermans y den Dunnen. (2014). *Biochim Biophys Acta*. 1842, 1932-1941.

4- NUEVAS ESTRATEGIAS DE ESTUDIO DE LA DIABETES MELLITUS TIPO 2

El hecho de que los estudios GWAS realizados hasta el momento sólo permitan identificar la asociación de polimorfismos frecuentes y de que los chips utilizados hasta el momento dejen muchas regiones genómicas sin una buena cobertura, supone que la asociación con DM2 de muchos polimorfismos frecuentes y poco frecuentes (PFs y PPFs), y muchos de los loci que los contienen no hayan sido identificados hasta el momento.

Todo ello unido al desarrollo de las tecnologías de nueva generación, ha hecho que se busquen otras aproximaciones en el análisis de datos (Perry et al, 2009) y se intente abordar el estudio del genoma con nuevas aproximaciones que permitan sobrepasar las limitaciones de los estudios genéticos realizados hasta el momento en relación a la genética de enfermedades complejas, entre ellas la DM2 (Eichler et al, 2010). En este sentido emergen tres posibles estrategias que pueden ser relevantes: a) identificación de variantes funcionales en el exoma; b) estudio de polimorfismos poco frecuentes y raros (PPFs); c) selección de casos y controles mediante criterios específicos.

4.1- IDENTIFICACIÓN DE VARIANTES GENÉTICAS EN EL EXOMA

La tecnología de secuenciación de nueva generación ha tenido un gran desarrollo en los últimos años, permitiendo realizar estudios de secuenciación de numerosos exomas en una reacción y la capacidad de los equipos se incrementa constantemente. Por otro lado, también se ha mejorado enormemente la capacidad y fiabilidad de los programas de análisis y alineamiento de las secuencias obtenidas.

El estudio del exoma resulta especialmente interesante dado que más del 90-95% de las mutaciones y polimorfismos asociados a enfermedades se sitúan en las regiones codificantes de proteínas y, por tanto, en el exoma (Botstein et al, 2003; Stenson et al,

2003). Además, el análisis de polimorfismos distribuidos en todo el genoma muestra una mayor concentración de polimorfismos asociados con enfermedades complejas en el exoma frente al resto (Ng et al, 2009; Li B et al, 2010; Lehne et al. 2011). El estudio del exoma también permite analizar aquellas alteraciones que modifican el “splicing” del ARN y se encuentran en las regiones intrónicas contiguas a los exones (Mamanova et al, 2010; Fisher et al, 2011). Existen otras regiones del genoma de gran interés y cuyo análisis será muy importante; pero los resultados esperables son menos prometedores y el coste de su estudio será mucho más elevado. El estudio del exoma hace posible identificar gran número de variantes funcionales a un coste menor y generando una cantidad de datos más manejable en comparación con la secuenciación del genoma.

El número de polimorfismos en el exoma es muy elevado; en el estudio de 200 exomas se identificaron 120.000 polimorfismos de los que 53.000 no estaban descritos; sospechándose que hay muchos más por identificar, principalmente PPFs (Hedges et al, 2009; Ng et al, 2009; Li B et al, 2010). Es importante remarcar el elevado el número de variantes de pérdida de función presentes en cada persona, hasta 150 en todos los genes identificados y entre 10 y 20 en genes causantes de enfermedades hereditarias (1000 Genomes Project Consortium, 2012).

4.2- ESTUDIO DE POLIMORFISMOS DE BAJA FRECUENCIA Y VARIANTES RARAS

Los principales estudios realizados hasta el momento se basan en el análisis de polimorfismos frecuentes (PFs) en la población, es decir el estudio de aquellos que tienen una MAF superior al 5 o al 10 % (Pritchard et al, 2001; Yang et al, 2005; Panoussopoulou et al, 2009; Schork et al, 2009; Voight, et al, 2010). Estos estudios se basan en la hipótesis conocida como “Common Disease, Common Variant”, si bien, en la actualidad tiene cada vez más fuerza la hipótesis “Common Disease, Rare Variants” (Bodmer y Bonilla, 2008; Schork et al, 2009) debido a que cada vez son más los polimorfismos poco frecuentes que se asocian de forma importante a diferentes

Introducción

enfermedades y que pueden explicar parte de la "heredabilidad perdida" o no encontrada.

En vista a los resultados obtenidos hasta el momento, cada vez parece más claro que cada enfermedad común puede ser causada por polimorfismos de muy variada frecuencia en la población, integrando ambas teorías e incluyendo los diferentes tipos de polimorfismos: CNVs ("copy number variations"), SNPs, ins/dels, etc. y con diferente nivel de efecto funcional (Bodmer y Bonilla, 2008; Manolio et al, 2009; Panoustsopoulou et al, 2009; Schork et al, 2009; Billings et al, 2010; Eichler et al, 2010). Esta hipótesis también se ha propuesto en relación a la DM2 tras analizar los datos de numerosos GWAS con esta enfermedad (Voight et al, 2010). Es decir, pueden existir numerosos PPFs (incluyendo las variantes raras; $MAF < 0.001$) con un potencial impacto elevado sobre el fenotipo (muchos con $OR > 3$); por lo que en conjunto pueden estar detrás de un importante porcentaje del componente genético de la DM2 (Bodmer y Bonilla, 2008; Schork et al, 2009; Billings et al, 2010; Robinson, 2010). Si bien, existen algunos argumentos en contra (Yang et al, 2005).

Los PPFs actualmente son desconocidos en su inmensa mayoría (Gorlov et al, 2008; PC et al, 2008). El proyecto "1000 genomas" ha permitido la identificación de numerosos polimorfismos, 38 millones de SNPs exactamente (1000 Genomes Project Consortium, 2012). Este trabajo ha supuesto un gran avance tanto en la identificación de nuevas variantes como en el desarrollo de herramientas bioinformáticas. Pero no garantiza la identificación de aquellos PPFs asociados a la DM2 y menos en población española (al incluirse solo una pequeña muestra de población española en el proyecto). Esto se debe a que gran parte de los PPFs localizados en el exoma pueden proceder de eventos de mutación relativamente recientes y/o haber sufrido selección en base a su efecto (Gorlov et al, 2008); por lo que pueden existir diferencias importantes entre poblaciones. Un objetivo fundamental del Proyecto 1000 Genomas era identificar más del 95% de los SNPs al 1% de frecuencia en un amplio conjunto de poblaciones. Así, se incluyeron el 50%, 98% y 99,7% de los SNPs con frecuencias de, 0,1%, 1,0% y 5,0%, respectivamente, provenientes de muestras de 2500 genomas secuenciados de UK (proyecto UK10K, financiado por Wellcome Trust), cumpliendo así con este objetivo. Sin

embargo, la cobertura puede ser menor para las poblaciones que no están estrechamente relacionadas con las estudiadas. Por ejemplo, el proyecto incluye sólo el 23,7%, 76,9% y 99,3% de los SNPs con frecuencias de, 0,1%, 1,0% y 5,0%, respectivamente, en, 2000 genomas secuenciados en un estudio de la población Cerdeña (the SardiNIA study).

El análisis de los PPFs muestra, tanto los efectos generalizados de la selección purificadora o estabilizadora en sitios funcionalmente relevantes del genoma, como la forma en que se puede interactuar con la historia de la población para dar lugar a una diferenciación local sustancial. El efecto surge principalmente debido a que las variantes raras tienden a ser recientes y por lo tanto están geográficamente restringidas (Gravel et al, 2011; Mathieson y McVean, 2012; Nelson et al, 2012). Esto implica que la interpretación de las variantes raras en individuos con una enfermedad en particular debe hacerse dentro del contexto de los antecedentes genéticos locales. Por otra parte, aboga por el valor de continuar con la secuenciación de individuos de diversas poblaciones para caracterizar el espectro de la variación genética humana y para dar soporte a estudios sobre diferentes enfermedades en estas poblaciones.

En cuanto a la relación de los polimorfismos poco frecuentes con enfermedades complejas, tales como la DM2, ya hay estudios (Steinthorsdottir et al, 2014) que han demostrado la asociación de algunas variantes raras con diabetes tipo 2, poniendo así de manifiesto la implicación e importancia de este tipo de variantes en estas enfermedades.

4.3- SELECCIÓN DE CASOS Y CONTROLES MEDIANTE NUEVOS CRITERIOS.

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

Introducción

población general. Para aumentar la probabilidad de encontrar variantes genéticas en un grupo reducido de muestras, que puedan estar asociadas a la DM2 y que aumenten el riesgo de desarrollar DM2 en una gran proporción de los casos, deben seleccionarse correctamente. Como ya se ha indicado, el incremento en DM2 se está produciendo por cambios que suponen un incremento de la obesidad; por lo que es muy importante identificar qué factores genéticos se asocian con el desarrollo de la DM2 en pacientes con preobesidad u obesidad Tipo I (según la clasificación SEEDO-2007). En conjunto este grupo representa al paciente "tipo" con un elevado riesgo de desarrollar DM2 y los casos más frecuentes en la población, principalmente con una edad media (40-60 años), aunque la edad de diagnóstico se está reduciendo. Los pacientes con obesidad Clase II o superior presentan una exposición a los efectos de la obesidad muy elevada por lo que el componente genético podría verse desvirtuado.

Por ello, se deben seleccionar aquellos con la mayor carga genética posible dentro del objetivo planteado, es decir pacientes DM2 con IMC entre 28 y 35 (Lyssenko et al, 2010). La obesidad abdominal tiene un efecto similar al IMC, pero que no se ha analizado en muchos estudios; por lo que deben incluir casos y controles con ella. Por otro lado, con la edad la exposición a factores de riesgo se incrementa, esto supone que aquellos pacientes que desarrollan DM2 a mayores edades podrían tener un menor componente genético, aunque en el estudio de Voight et al (2010) la edad no parece alterar la asociación de los marcadores; posiblemente por la mezcla de fenotipos incluida en este estudio. Los antecedentes familiares suponen un aumento del riesgo, si bien la inclusión de este criterio podría aumentar la identificación de marcadores relacionados con factores de riesgo de DM2; por ejemplo, asociados a obesidad.

En resumen, en nuestro exoma existen numerosos polimorfismos con un efecto funcional y, por lo tanto, pueden tener un efecto sobre el fenotipo y algunos pueden estar asociados a DM2. Muchos de estos polimorfismos serán PPFs que pueden ser desconocidos hasta el momento en una gran parte y cuya asociación con DM2 no ha podido ser analizada mediante GWAS realizados. Estos polimorfismos estarán sobrerrepresentados en casos o en controles, según su efecto. Por lo tanto, resulta muy importante analizar las variantes genéticas presentes en el exoma en relación a la DM2.

Hipótesis y Objetivos

HIPÓTESIS Y OBJETIVOS

Por tanto, nuestra hipótesis es: “Diferentes variantes genéticas localizadas en el exoma pueden estar implicadas en el desarrollo de la DM2 y explicar parte de la heredabilidad que no ha podido ser encontrada por otros medios”. Además, para poder demostrar esta hipótesis, nos basamos en las siguientes:

- a) La secuenciación del exoma de casos y controles para esta enfermedad nos permitirá aumentar la proporción de polimorfismos detectados que estén implicados en esta enfermedad.
- b) La selección de polimorfismos con mayor efecto funcional, aumentará el porcentaje de polimorfismos implicados en la enfermedad.

Como consecuencia, los objetivos de esta tesis han sido los siguientes:

- Secuenciar el exoma de 178 individuos: 89 casos y 89 controles.
- Poner a punto toda la metodología de preparación de librerías de ADN, secuenciación masiva, y todos los métodos de análisis bioinformático para la obtención de resultados.
- Analizar los genes previamente relacionados con DM2 en los individuos secuenciados.
- Identificar variantes no conocidas hasta el momento y polimorfismos poco frecuentes que, pese a haber sido identificados previamente, no se habían relacionado con la DM2.
- Validar los resultados obtenidos mediante el método de referencia.

Material y Métodos

MATERIAL Y MÉTODOS

1- SELECCIÓN DE MUESTRAS:

Se seleccionaron 178 muestras para su posterior secuenciación de exoma, provenientes de la población de Valcar (Hospital Clínico de Valencia) y del biobanco del CIBERDEM, de acuerdo a los siguientes criterios:

- **Controles:** 89 sujetos sin DM2:

- Con IMC entre 26 y 32.
- De 45-65 años de edad.
- parámetros asociados con la diabetes lejos de los valores de DM2:
 - glucosa en ayunas <90 mg/dl.
 - metabolismo hidrocarbonado normal tras sobrecarga oral de glucosa (glucosa <140 mg/dl e insulina <64 μ U/ml).

- **Casos:** 89 pacientes de DM2 diagnosticados según los criterios de la organización mundial de la salud:

- Con IMC entre 26 y 32 (preobesidad y obesidad tipo I; según clasificación SEEDO 2007).
 - De 45-65 años de edad.
 - Diagnóstico de diabetes.
- (Todo ello en el momento del diagnóstico).

En la tabla 5 se muestran los parámetros generales de los 178 individuos estudiados:

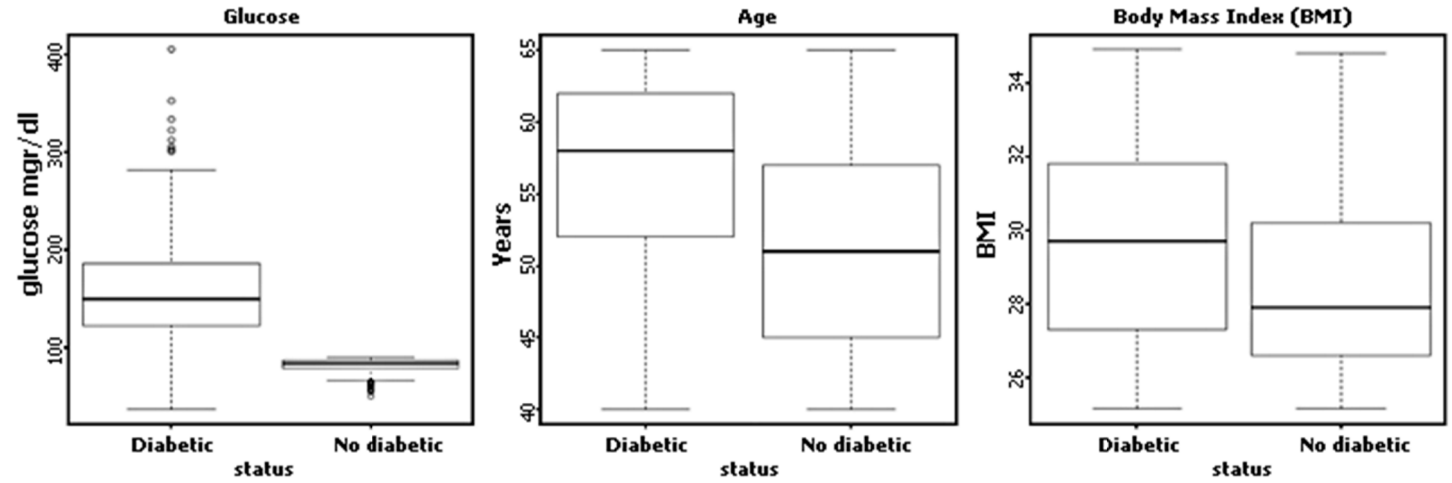
	CONTROLES	CASOS
Número de individuos	89	89
Edad	52.2 \pm 7.3	57 \pm 6.0
IMC (kg/m ²)	29.6 \pm 2.6	29.5 \pm 3.0
Glucosa (mg/dL)	81.7 \pm 7.3	154.9 \pm 62.4

Tabla 5: Características principales de las muestras empleadas.

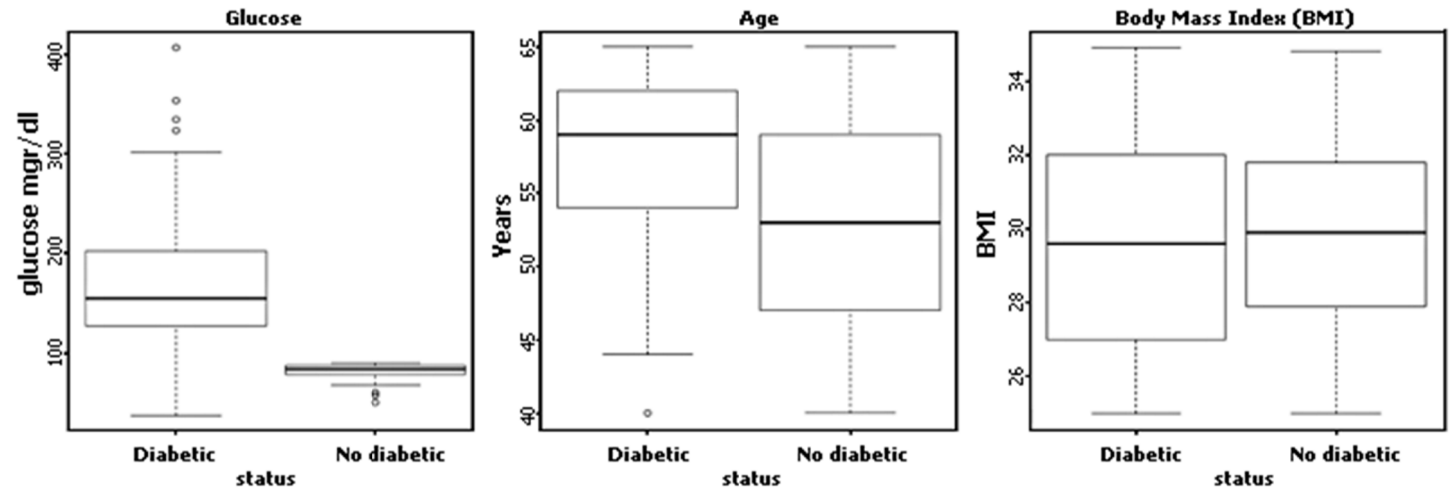
A continuación, en la figura 21 aparecen las gráficas que muestran los parámetros generales, antes y después de la selección. Se puede observar cómo se ha intentado obtener una muestra lo más homogénea posible con respecto a la población de partida:

Figura 21: Parámetros generales de las muestras

A) Previos a su selección.



B) Tras su selección. Características de los 178 individuos.



2- PROCESADO DE LAS MUESTRAS:

Los pasos que se han seguido, han sido los siguientes:

- Preparación de librerías.
- Enriquecimiento en exoma.
- Validación de los pools.
- Amplificación clonal.
- Secuenciación mediante la tecnología Illumina.
- Análisis bioinformático.
- Validación de resultados por secuenciación de Sanger.

2.1-PREPARACION DE LIBRERIAS DE ADN:

Para preparar las librerías, se ha empleado el kit “TruSeq DNA SamplePreparation v2”. Para ello previamente se debe cuantificar y fragmentar el ADN. En la figura 22 se representa un esquema de todo el proceso:

Material y Métodos

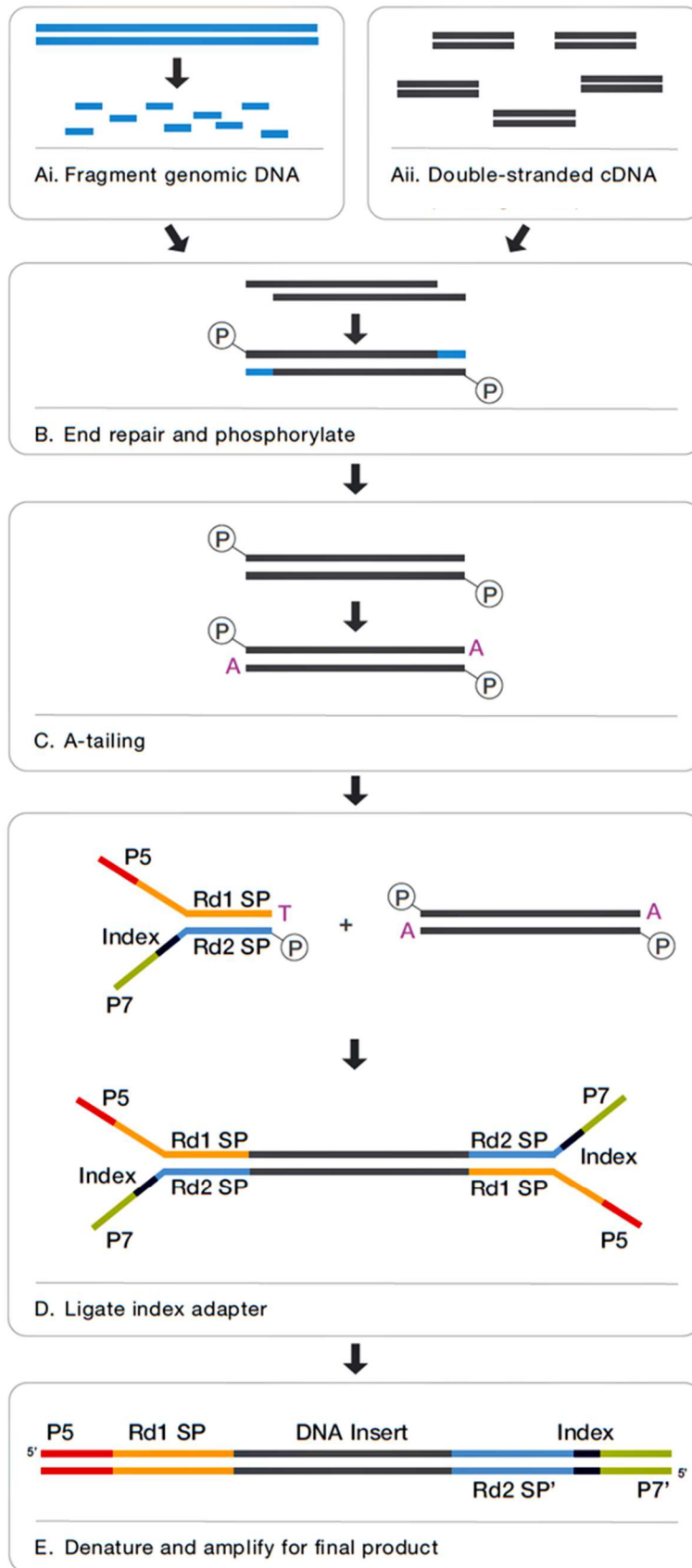


Figura 22: Esquema del proceso de preparación de librerías.

2.1.1-Cuantificación de ADN

El primer paso es la cuantificación del ADN de partida, Illumina recomienda partir de 50µL a una concentración de 20 ng/µL. Es decir, se parte siempre de 1 microgramo total de ADN. La cuantificación ha sido realizada con el kit “PicoGreen® dsDNA quantitation assay” (Invitrogen) y el sistema “GloMax®-MultiDetectionSystem” (Promega). La ventaja de este procedimiento es que se une a ADN de doble cadena únicamente, con lo cual aumenta la fiabilidad del método frente al tradicional (absorbancia a 260 y 280 nm). El principal inconveniente del método tradicional, era que podían existir interferencias en la medida como, por ejemplo, la contribución relativa de nucleótidos, ácidos nucleicos de cadena simple o proteínas. Con este reactivo, se asegura que se está cuantificando únicamente el ADN de doble cadena.

2.1.2-Fragmentación de ADN

El segundo paso es fragmentar el ADN genómico, para ello se empleó el sistema Covaris S220, veáse figura 23a y 23b (Unidad de Genética Molecular, Instituto de Biomedicina de Valencia-CSIC).



Figura 23: A) Covaris S220



B) Microtubos empleados

Funciona como un sonicador tradicional, pero presenta la ventaja de que el proceso puede controlarse de una forma muy precisa modificando los distintos parámetros: número de ciclos, intensidad, etc. Todo ello es llevado a cabo manteniendo constante la temperatura del baño y sin contacto con el ambiente (figura 24), lo cual hace que la

Material y Métodos

fragmentación sea totalmente reproducible en muestras que se encuentren en similares condiciones.

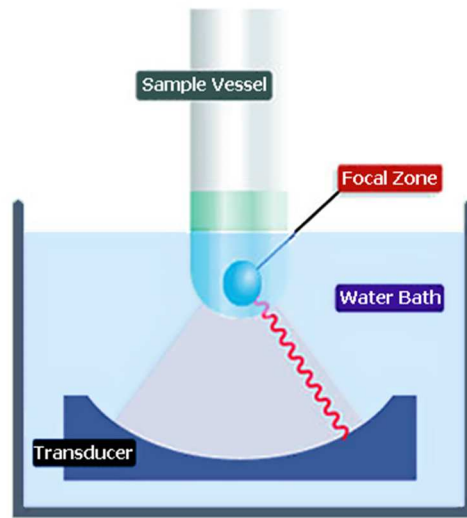


Figura 24: Esquema sobre el funcionamiento de Covaris.

Una vez el ADN está fragmentado, pasamos a emplear el kit TruSeq DNA Sample Preparation v2, que consta de los siguientes pasos:

2.1.3-Reparación de los extremos.

Este proceso convierte los extremos salientes resultantes de la fragmentación en extremos romos. Mediante la actividad exonucleasa 3'-5' de la reacción, sobre ADN de cadena sencilla, se eliminan los salientes en 3' y la actividad polimerasa rellena los salientes en 5' (veáse figura 22 A y 22 B).

2.1.4-Adenilación

Se añade una Adenina (A) en los extremos 3' para prevenir la ligación entre los diferentes fragmentos durante la ligación de los adaptadores. Esto asegura una baja tasa de formación de quimeras y además evita la ligación entre oligos, evitando así la formación de primer-dimer. La correspondiente Timina (T) en el extremo 3' del adaptador hará posible la ligación con el fragmento (veáse figura 22 C).

2.1.5-Ligación de los adaptadores

Este proceso permite adicionar unos índices a los fragmentos, de forma que cada muestra será marcada con un índice diferente. Esto prepara las muestras para su distribución en la célula de flujo o “flow-cell” es decir, permitirá identificar cada muestra en el proceso de secuenciación (véase figura 22 D).

Además, es en este paso cuando se añaden también todas las secuencias y oligos necesarios para el proceso de secuenciación:

P5 Y P7 son los adaptadores que hibridan con los que se encuentran unidos a la superficie de la célula de flujo del sistema HiScan SQ. Sus secuencias son:

5' P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

5' ACACTCTTCCCTACACGACGCTCTTCCGATCT

Rd1 SP: Oligo para secuenciación del read 1

5' ACACTCTTCCCTACACGACGCTCTTCCGATCT

Rd2 SP: Oligo para la secuenciación del read 2

5' CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

Index: índice que proporciona a cada muestra una secuencia específica para ser identificada. En la tabla 6 se muestran los índices empleados:

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

Tabla 6: Secuencia de los índices empleados.

2.1.6-PCR de enriquecimiento

Sirve para enriquecer selectivamente aquellos fragmentos de ADN que contienen moléculas de adaptador en sus extremos, y de esta forma, amplificar la cantidad de ADN en la librería. La PCR se lleva a cabo con un PCR primer cocktail que se une al final de los adaptadores. El número de ciclos debe ser el mínimo posible para evitar el sesgo de representación de cada librería (véase figura 22 E).

Tras cada paso se purifica con las bolas magnéticas AgencourtAMPure XP (BeckmanCoulter, Inc). Retiene ADN de cadena simple y doble con tamaños de 150pb-10kb.

2.1.7- Puntos de control

Durante el proceso hay varios puntos de control:

- Tras la fragmentación
- Tras la formación de la librería

Este control se ha realizado con el sistema Qiaxcel (Qiagen), empleando el “QIAxcel DNA Screening Kit”. Este sistema utiliza un gel de electroforesis capilar que permite una rápida separación de ácidos nucleicos por tamaños. A diferencia del gel de agarosa tradicional, la separación se lleva a cabo en un capilar con gel. Cada muestra se carga automáticamente en un capilar individual y se aplica voltaje. Las moléculas cargadas negativamente migran hacia el polo positivo. Las moléculas de bajo peso molecular migran más rápido que las de mayor peso.

Al mismo tiempo que las moléculas migran a través del capilar, pasan un detector que mide la señal fluorescente generada por bromuro de etidio. Un fotomultiplicador convierte la señal emitida en datos electrónicos que se transfieren al ordenador para ser procesados con el software BioCalculator. Tras el procesamiento, los datos se transforman en un electroferograma y en la imagen del gel. En la figura 25 se muestra un esquema del funcionamiento de este aparato.

Este sistema ofrece gran número de ventajas frente al gel de agarosa tradicional: Mayor sensibilidad de detección, menor pérdida muestral y un rápido análisis.

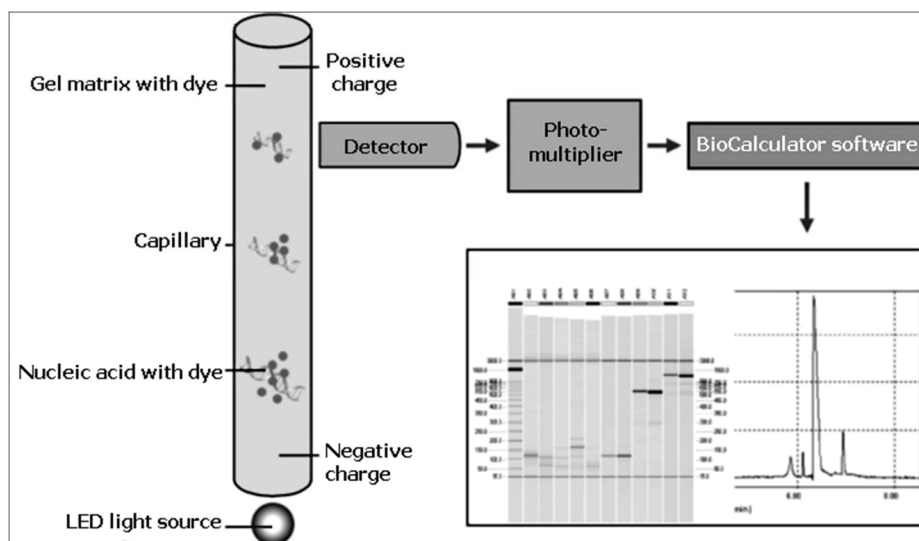
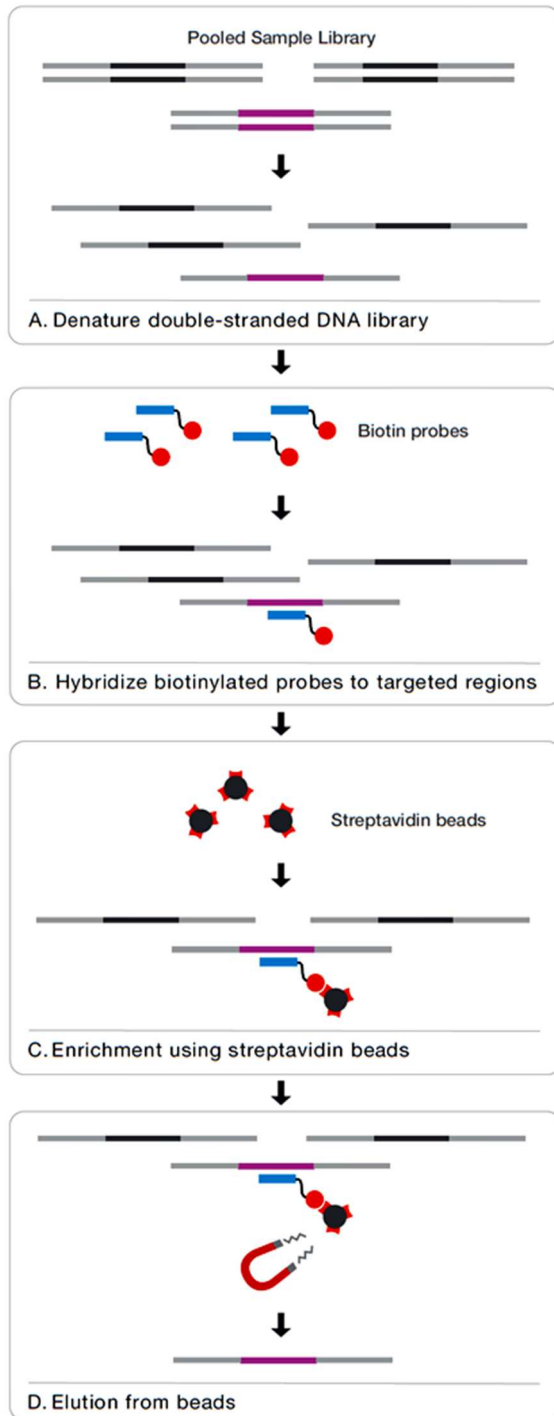


Figura 25: Funcionamiento del sistema QIAxcel.

2.2- ENRIQUECIMIENTO: EXOMA

El proceso empieza con la mezcla de las librerías, poniendo de cada una la misma cantidad, en este caso 500 ng por librería en pools de 6, tal y como indica Illumina para exoma. La cuantificación se hace con el aparato GloMax®-MultiDetection System (Promega) explicado previamente.



El enriquecimiento, consiste en 4 pasos principales:

Primera Hibridación:

Previa desnaturalización del ADN, la librería de ADN se mezcla con sondas que capturan las regiones deseadas. La hibridación, que dura 16-20 horas, asegura la unión de las sondas a las regiones correspondientes. Se pueden combinar diferentes librerías en un mismo pool siempre que tengan diferente índice.

Primer lavado:

Se emplean bolas de streptavidina para capturar las sondas, que estarán ya unidas a las regiones de interés. Hay hasta tres tipos de buffer de lavado para quitar, todo lo que no se haya unido a las sondas. Finalmente se eluyen las librerías y se preparan para una segunda hibridación de 16-20 horas.

Segunda hibridación:

La primera elución de las librerías se mezcla de nuevo, previa desnaturalización del ADN, con sondas que capturan las regiones de interés. Esta segunda hibridación asegura más a fondo la captura de las regiones de interés.

Segundo lavado:

Se vuelven a emplear las bolas de streptavidina para capturar las sondas que se encuentran unidas a las regiones de interés. Hay hasta tres tipos de buffer de lavado para quitar todo lo que no se haya unido a las sondas y tras ello, las librerías son eluidas.

Amplificación de los pools por PCR:

Se lleva a cabo con los mismos reactivos que la amplificación de la librería y tras ello se purifica con AMPure.

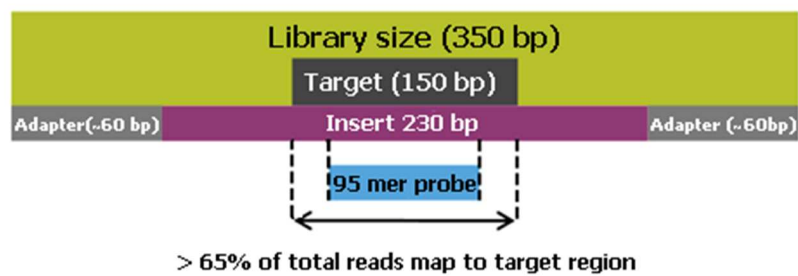
Figura 26: Proceso de enriquecimiento.

El kit TruSeq Exome Enrichment de Illumina incluye más de 340.000 sondas, cada una de ellas diseñadas contra el genoma de referencia humano NCBI37/hg19. Este kit enriquece más de 200.000 exones y, por tanto, más de 20.000 genes de interés. A continuación, en la tabla 7, se muestran detalles más concretos sobre el kit:

	TruSeq Exome Enrichment
Tamaño de la región amplificada	62 Mb
Número de genes	20.794
Número de exones	201.121
Tamaño de las sondas	95-mer
Número de sondas	340.427
Tamaño de librería recomendado	> 350 pb
% bases cubiertas a una cobertura media del 0.2x	> 80%

Tabla 7: Detalles de la cobertura proporcionada por el kit TruSeq Exome Enrichment.

Este kit cubre 62 Mb del genoma humano. Cada sonda tiene como diana, librerías de 300-400 pb (con insertos de un tamaño de 180-280 pb). Se enriquecen 265-465 bases situadas simétricamente en el punto medio de la sonda, tal y como muestra la figura 27. Esto supone que, además de cubrir los exones, también estén cubiertas las regiones adyacentes (promotores y UTRs).



With a 350 bp DNA library (mean insert size = 230 bp), the probe will enrich 365 bp ($2 \times \text{Insert} - \text{Probe}$) centered around its midpoint.

Figura 27: Esquema sobre el funcionamiento de las sondas.

2.3-VALIDACIÓN DE LOS POOLS:

2.3.1- Visualización de los pools

Para ello se emplea el sistema ya descrito Qiaxcel (Qiagen), de esta forma se ve si tiene el tamaño apropiado y la presencia o no de dímeros. Si está todo correcto, se calcula su tamaño medio, ya que en el paso posterior (cuantificación por qPCR) se efectúa la corrección por tamaños.

2.3.2- Selección por tamaños

Si la forma no es correcta y/o existe dímero, se selecciona el tamaño adecuado mediante el sistema Pippinprep, que consiste en una electroforesis que separa y extrae fragmentos de ADN. Se emplean cassettes de gel y se recoge el ADN en los pocillos de elución de acuerdo al rango de tamaño seleccionado (figura 28).

Las ventajas que presenta se encuentran principalmente a nivel de rapidez del proceso, ya que cada carrera dura 50-100 min y caben 4 muestras más el marcador. Además, no presenta riesgos tan graves como el método tradicional con bromuro de etidio.

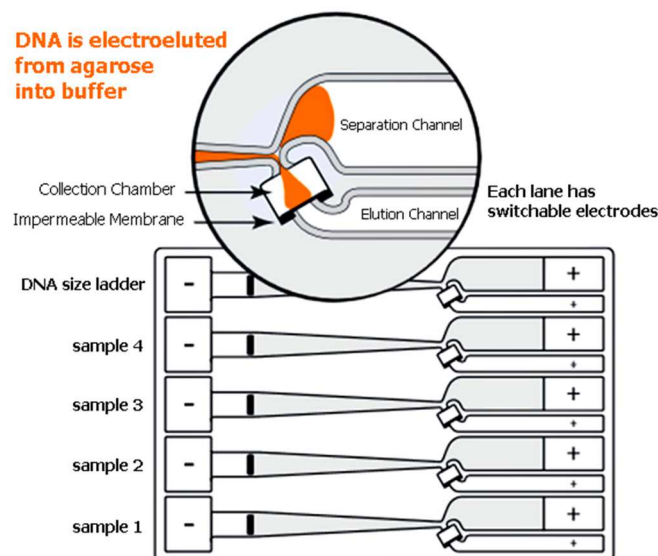


Figura 28: Estructura de un cassette de Pippinprep.

Y si está todo correcto, se pasa a cuantificar los pools mediante qPCR.

2.3.3- Cuantificación de los pools mediante PCR cuantitativa

Para ello se ha empleado el sistema LightCycler 480 (Roche). La PCR cuantitativa es una variación de la técnica PCR estándar que se emplea para la cuantificación tanto de DNA como de mRNA en una muestra. Utilizando primers o cebadores específicos puede medirse el número de copias relativo de ADN que contiene una muestra.

Para realizar esta técnica se ha empleado el reactivo de unión al ADN de doble cadena (no específico de secuencia) SYBR Green. Este reactivo emite poca fluorescencia cuando no está unido al ADN y aumenta la fluorescencia que emite cuando se intercala en el ADN. Por tanto, la fluorescencia es proporcional a la cantidad de producto y es la medida de esta fluorescencia lo que permite la cuantificación del ADN.

Teóricamente en cada ciclo se duplica la cantidad de producto: éste aumenta exponencialmente con el número de ciclos de PCR y depende del número inicial de copias del molde de ADN. Los pasos a seguir para hacer una qPCR son los siguientes:

- Se emplean muestras estándar, muestras con concentración conocida. En expresión génica se emplean diluciones seriadas de un cDNA de concentración conocida.
- Se determina el Cp ("Crossing point", punto en el que la fluorescencia de la muestra supera el umbral o fluorescencia de fondo) de las muestras estándar y de las muestras desconocidas.
- Se realiza una curva estándar a partir de los datos de Cp y concentración de las muestras estándar.
- Se calcula la concentración de las muestras desconocidas extrapolando el valor de Cp de dichas muestras en la curva estándar (figura 29).

Material y Métodos

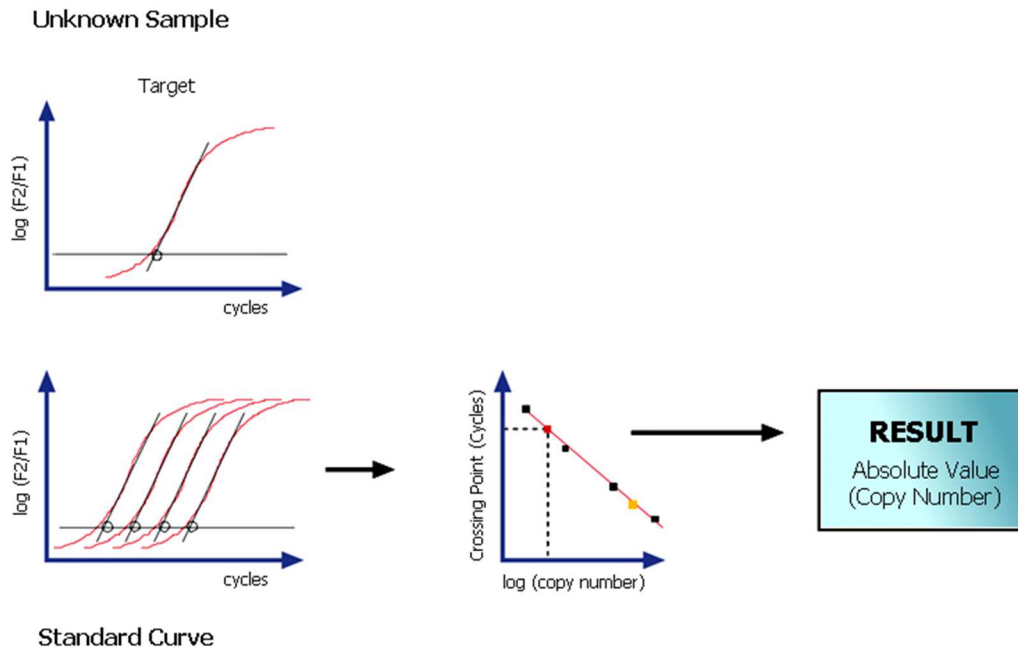


Figura 29: Imágenes obtenidas al realizar la cuantificación de ADN mediante qPCR. En este caso la muestra roja es la de concentración desconocida. Su concentración se obtiene al extrapolar su valor de Cp de la curva patrón realizada a partir de las concentraciones de las muestras estándar.

2.4- AMPLIFICACIÓN CLONAL DE LOS POOLS: cBOT

cBOT es el sistema de amplificación clonal de Illumina. Antes de ser secuenciadas, las moléculas de ADN son amplificadas “en puente” para formar clusters clonales en la célula de flujo. Las muestras se preparan a una concentración de 2nM por pool (definiendo un pool como la mezcla de 6 muestras con índices compatibles), posteriormente se desnaturalizan y se preparan a la concentración deseada para introducirlas en el cBOT. Tras diferentes pruebas, la concentración que dió mejores resultados fue la de 8pM.

Ya en el cBOT, el proceso consta de varias fases:

2.4.1- Inmovilización de las moléculas de ADN.

Cientos de millones de moléculas de ADN se hibridan con oligonucleótidos que se encuentran inmovilizados en la superficie de la célula de flujo. Se van generando copias por extensión en 3', empleando la ADN polimerasa de alta fidelidad para prevenir errores de incorporación. Las muestras originales son desnaturalizadas, dejando a su copia inmovilizada en la superficie de la célula de flujo (figura 30a).

2.4.2- Amplificación isotérmica en puente.

Las copias inmovilizadas giran para hibridarse con los oligonucleótidos que están libres sobre la célula de flujo. La ADN polimerasa copia la muestra formando puentes de ADN que se desnaturalizan para obtener 2 cadenas de ADN, que a su vez giran y se hibridan con otros oligonucleótidos y se extienden otra vez para formar dos nuevas cadenas de ADN. Este proceso se repite en cada muestra para crear millones de clusters clonales individuales que contendrán unas 2000 moléculas (figura 30b).

2.4.3- Linealización, bloqueo e hibridación con los primers.

Cada cluster de los puentes de DNA es desnaturalizado y la cadena en dirección 3' se elimina, dejando la cadena en 5'. El extremo 3' de las cadenas de ADN y los oligonucleótidos presentes en la célula de flujo son bloqueados para que no interfieran con la reacción de secuenciación. El oligo de secuenciación se hibrida con la secuencia complementaria del adaptador de Illumina en los extremos sueltos de las muestras presentes en los clusters. La célula de flujo contiene ahora cerca de 750 millones de clusters con aproximadamente 1000 moléculas por cluster, de esta forma está lista para ser secuenciada (figura 30c).

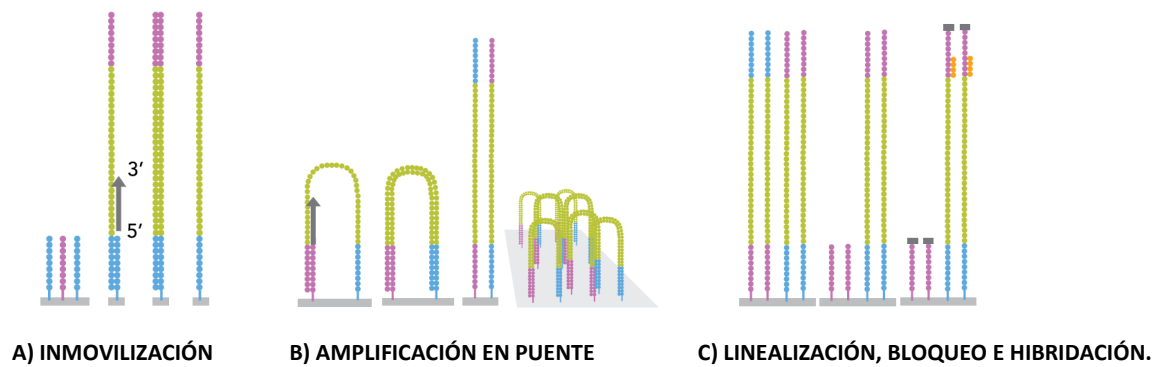


Figura 30: Generación de clusters por amplificación isotérmica en puente.

2.5- SECUENCIACIÓN: HiScanSQ y HiSeq 2000.

2.5.1- Secuenciación por síntesis.

La tecnología Illumina es ampliamente utilizada en el campo de la secuenciación masiva. Emplea el método denominado secuenciación por síntesis, que consiste en el uso de terminadores reversibles en un método cíclico que incluye la incorporación de nucleótidos, la captura de fluorescencia y la ruptura:

- La ADN polimerasa unida a la muestra con el primer, incorpora (por complementariedad de bases), un solo nucleótido marcado con fluorescencia.
- Después de la incorporación, se eliminan los nucleótidos que no han sido incorporados y se lee la fluorescencia mediante láseres para saber qué nucleótido se ha incorporado.
- Sigue un paso de ruptura para eliminar la fluorescencia y la modificación que impide la elongación de la cadena.

En la figura 31 se muestra un ejemplo del proceso de secuenciación:

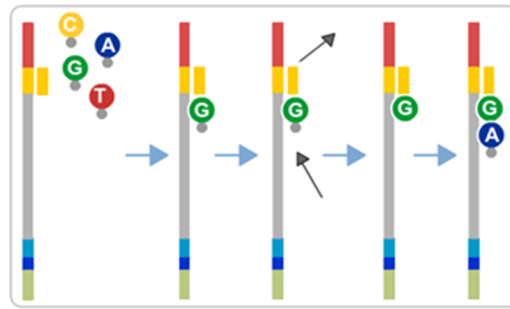


Figura 31: Proceso de secuenciación por síntesis. 1. Los ddNTP marcados con fluorescencia se añaden junto con la polimerasa. Los ddNTP son bases modificadas ligeramente que evitan que la polimerasa se añada a otros nucleótidos de la cadena en crecimiento después de haber añadido la primera. Los nucleótidos sin incorporar se eliminan. 2. Un láser activa el fluoróforo verde. 3. La cámara captura el color de emisión. 4. El fluoróforo y el terminador se eliminan. El nucleótido ha vuelto a su forma natural y la polimerasa se puede así añadir a otra base de la cadena en crecimiento, en el siguiente ciclo de extensión. 5. El siguiente ciclo comienza añadiendo un nuevo grupo de ddNTPs marcados con fluorescencia.

Las lecturas que se han hecho en este estudio han sido del tipo paired end, es decir, en una misma carrera, se producen 3 reads o lecturas: uno en un sentido de la hebra de ADN, otro para la lectura de los índices y el restante en el otro sentido, tal y como se muestra en la figura 31b:

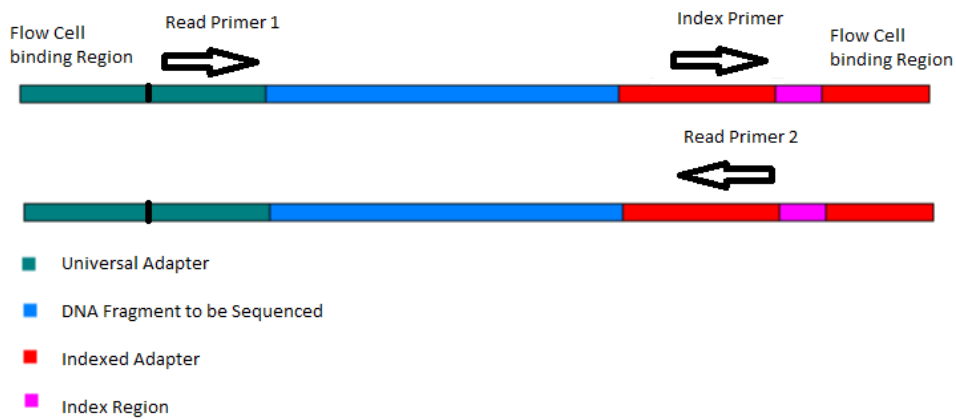


Figura 31b: Esquema de las lecturas del tipo paired-end.

Para tener un análisis muy preliminar a tiempo real de la secuenciación, Illumina proporciona unas herramientas de software para ver cómo está funcionando el proceso (figuras 32 y 33). Son las siguientes:

- HiSeq Control Software:

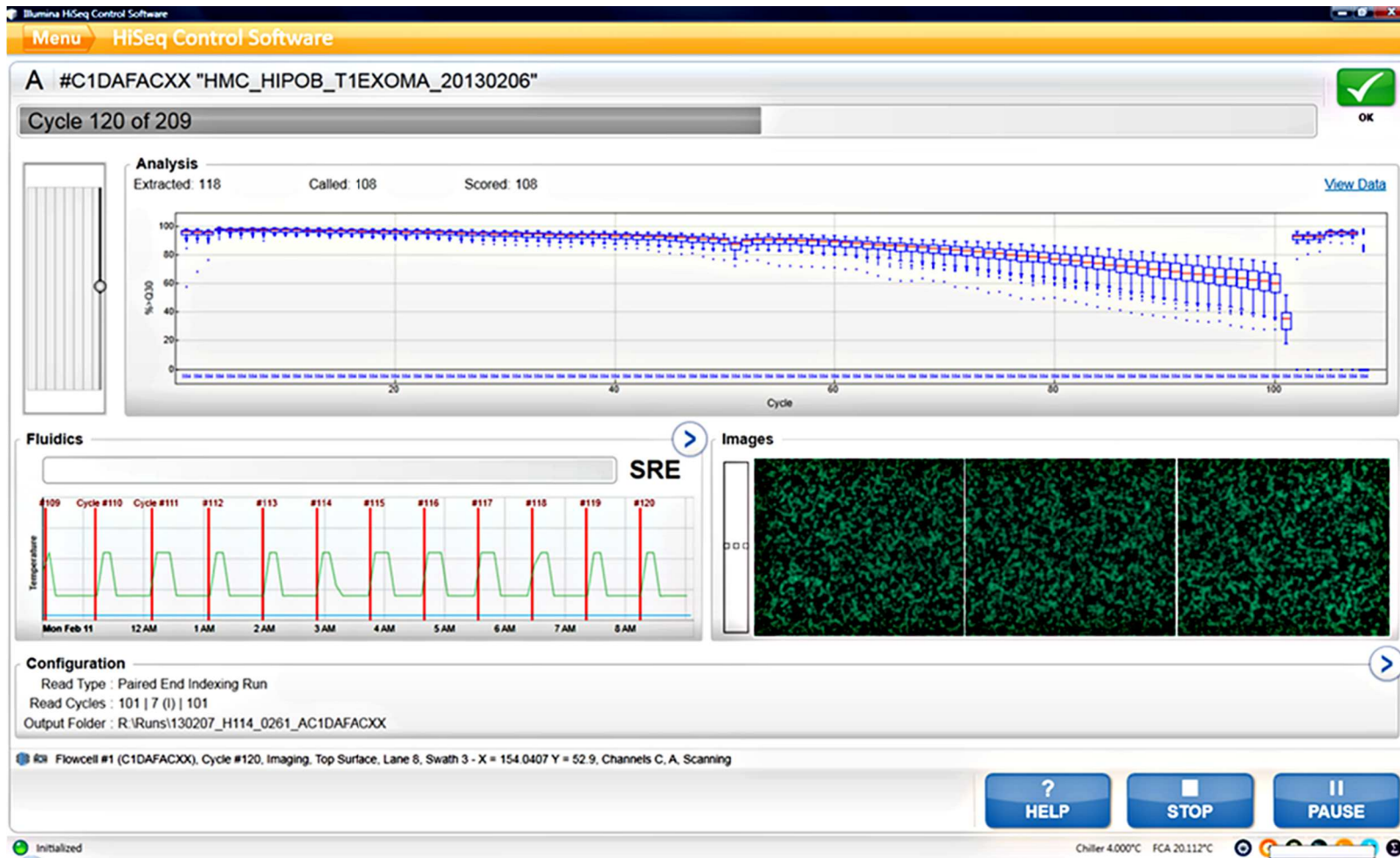


Figura 32: HiSeq control software. Proporciona información a tiempo real sobre el ciclo en que se encuentra la secuenciación y también es el software en que aparecen las notificaciones si hay alguna incidencia. Sirve para controlar el run, pudiendo pausarlo o pararlo si fuera necesario. Además, ofrece la posibilidad de ver los clusters que se están leyendo en ese momento y en qué parte de la célula de flujo se está produciendo la lectura.

- Sequencing Analysis Viewer (SAV):



Figura 33: Sequencing Analysis Viewer. Proporciona información sobre los clusters efectivos, densidad de clusters en cada zona de la célula de flujo, representación de los diferentes índices y, por tanto de cada muestra en cada línea de la célula de flujo...

2.5.2- HiScanSQ y HiSeq 2000.

El sistema HiScan integra la secuenciación de nueva generación con la capacidad de realizar análisis de genotipado, expresión génica y arrays de metilación. Esto se debe a que se compone de dos módulos, como se muestra en la figura 34: módulo HiScan, para los análisis de microarrays; y módulo SQ, para llevar a cabo la secuenciación. Ambos módulos son independientes. Esta combinación le convierte en una herramienta muy útil para diversos tipos de estudios genéticos.



Figura 34: Características del HiScanSQ.

En la tabla 8 se muestran los distintos tipos de lecturas que pueden llevarse a cabo y la información relativa al funcionamiento de este aparato:

Longitud de lectura	Duración de la carrera	Output
1 x 35 pb	1.5 días	23-26 Gb
2 x 50 pb	4.5 días	67-75 Gb
2 x 100 pb	8.5 días	135-150 Gb
Lecturas	Más de 750 millones de clusters que pasan los filtros y más de 1.5 billones de lecturas paired-end.	
Rendimiento	Más de 17.5 Gb por día para una carrera de 2 x 100 pb	
	Más de 85 % de bases con calidad mayor de 30 en 2 x 50 pb	
	Más de 80 % de bases con calidad mayor de 30 en 2 x 100 pb	

Tabla 8: Parámetros de secuenciación de HiScanSQ

En cuanto al HiSeq 2000 (figura 35), el método de secuenciación es el mismo, sólo que con una generación de resultados mayor y más rápida. Con este sistema, los clusters de ambas caras de la célula de flujo pueden ser secuenciados, aumentando así el número de lecturas y por tanto, la generación de resultados.

Material y Métodos

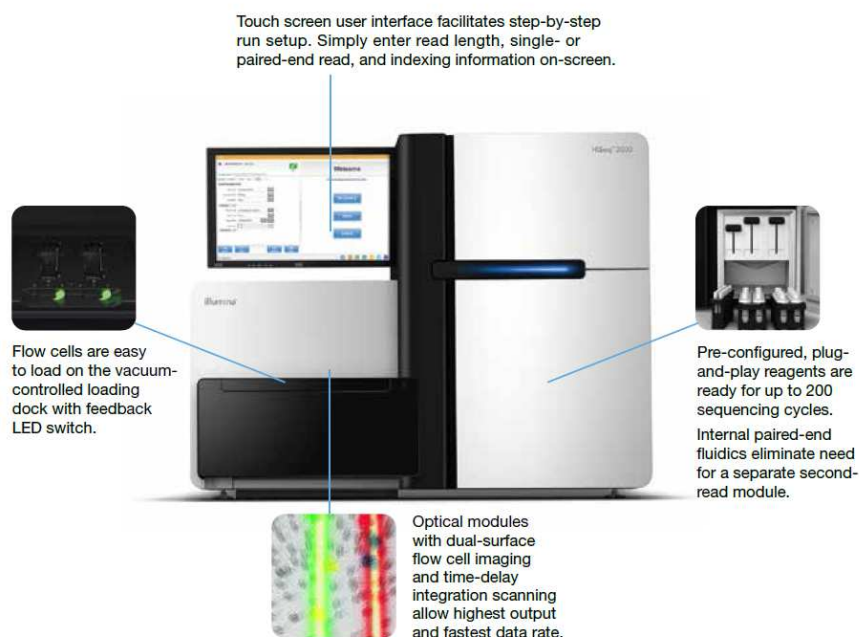


Figura 35: Características de HiSeq 2000

Los parámetros que emplea se representan en la tabla 9:

Longitud de lectura	Duración de la carrera	Output
1 x 35 pb	1.5 días	26-35 Gb
2 x 50 pb	4 días	75-100 Gb
2 x 100 pb	8 días	150-200 Gb
Lecturas	Más de 1 billón de clusters que pasan los filtros y más de 2 billones de lecturas paired-end.	
Rendimiento	Más de 25 Gb por día para una carrera de 2 x 100 pb	
	Más de 90 % de bases con calidad mayor de 30 en 2 x 50 pb	
	Más de 85 % de bases con calidad mayor de 30 en 2 x 100 pb	

Tabla 9: Parámetros de secuenciación de HiSeq 2000

3- ANÁLISIS BIOINFORMÁTICO

En el procesado de las muestras son importantes tanto la preparación y secuenciación de librerías como la generación y procesado de los datos, que se resumen en la figura 36:



Figura 36: Pasos a seguir en el análisis bioinformático de las secuencias.

Para seguir este esquema, es necesario el uso de diferentes programas bioinformáticos y obtener así diferentes archivos, que finalmente conducirán a la obtención de las variantes deseadas. Un esquema muy general de los pasos a seguir sería el que se representa en la figura 37:

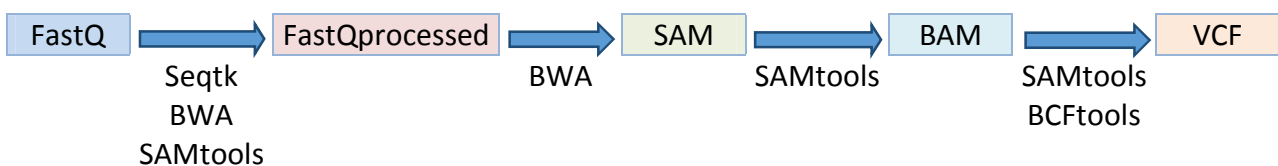


Figura 37: Programas empleados y archivos generados durante el proceso de análisis.

3.1- Alineamiento y mapeo de las secuencias: Paso de FastQ a SAM (Sequence Alignment/Map format).

3.1.1- Análisis de la calidad de los archivos FastQ: FastQC

El formato de un archivo fastQ es el siguiente (figura 38):

```
@H114:263:C0K6MACXX:7:1101:1123:1976 1:N:0:CGATGT
GTGGATACCTGTTGCTGACTATAAGAGATGAGACTGCACATGGAGATGTTCTGTCCCCCTCAGTGTCTGAGCCCTCCGACTTCTTTCTTCCCCATCAA
+
@@@CFFFFFFHHHHJJJFHGEFHHFEHHJJJJJJJJJJJJIFHHGIIIEHIIJGIIJGIIJHFDFDGGGIIHHEE8ADDDDDDDDDDDDCDCDDCCC
```

Figura 38: Formato de un archivo fastQ.

Estos archivos constan de subunidades de 4 líneas:

Línea 1: Información sobre la muestra, es decir, lane, barcode, nombre de la muestra...

Línea 2: Secuencia

Línea 3: Separación entre secuencia y calidades. Suele ser “+”.

Línea 4: Calidades en Phred Score codificado en formato ASCII.

El programa FastQC proporciona información sobre la calidad de los archivos fastQ obtenidos.

3.1.2 - Procesado de los archivos fastQ

Se basa en eliminar las bases de baja calidad presentes en las lecturas. Se eliminan de los extremos de cada lectura, ya que al final de cada secuenciación se va perdiendo calidad. Este paso se realiza con el programa Seqtk, que es una herramienta para el procesado de secuencias en formato FASTA/Q (véase figura 37).

3.1.3- Mapeo

Consiste en el alineamiento de las secuencias obtenidas frente a un genoma de referencia. Se puede hacer ejecutando primero el programa Burrows-Wheeler Aligner (BWA) (Li et al, 2009a; Li H et al, 2010) y luego el programa SAMtools (Li et al, 2009b) o ejecutando ambos a la vez. BWA es un paquete de software para mapear secuencias contra un genoma de referencia. Consiste en tres algoritmos: BWA-backtrack, BWA-SW and BWA-MEM. El primer algoritmo está diseñado para secuencias mayores de 100 pb en Illumina, mientras que los otros dos se emplean para secuencias más largas, 70 pb a 1Mpb. BWA-MEM and BWA-SW comparten ciertas características, pero BWA-MEM, es más novedoso, y permite un análisis más rápido y preciso. Además también se emplea para secuencias de 70 a 100pb provenientes de Illumina. Es el que se ha empleado en el presente análisis.

El formato SAM (SequenceAlignment/Map) es un formato genérico para almacenar secuencias alineadas en un tamaño menor. SAMtools posee varias utilidades para manipular alineamientos en formato SAM, incluyendo, ordenar, indexar y mezclar estos archivos. Las líneas de su cabecera empiezan por '@'. Esta cabecera contiene información sobre cómo se ha generado el archivo, los programas empleados para procesar las lecturas, la referencia utilizada, etc.

Un ejemplo de la cabecera de un archivo SAM sería el que se muestra en la figura 39:

```
@HD      VN:1.0  SO:coordinate
@SQ      SN:1    LN:249250621  AS:NCBI37
         UR:file:/data/local/ref/GATK/human_g1k_v37.fasta
         M5:1b22b98cdeb4a9304cb5d48026a85128
@SQ      SN:2    LN:243199373  AS:NCBI37
         UR:file:/data/local/ref/GATK/human_g1k_v37.fasta
         M5:a0d9851da00400dec1098a9255ac712e
@SQ      SN:3    LN:198022430  AS:NCBI37
         UR:file:/data/local/ref/GATK/human_g1k_v37.fasta
         M5:fdfd811849cc2fadebc929bb925902e5
@RG      ID:UM0098:1    PL:ILLUMINA    PU:HWUSI-EAS1707-615LHAAXX-
L001
LB:80   DT:2010-05-05T20:00:00-0400    SM:SD37743    CN:UMCORE
@RG      ID:UM0098:2    PL:ILLUMINA    PU:HWUSI-EAS1707-615LHAAXX-
L002
LB:80   DT:2010-05-05T20:00:00-0400    SM:SD37743    CN:UMCORE
@PG      ID:bwa  VN:0.5.4
```

Figura 39: Cabecera de un archivo SAM.

Material y Métodos

El siguiente paso es el de SAM a BAM, también realizado con SAMtools.

Para llegar a obtener el archivo VCF es necesario, ordenar, indexar y eliminar duplicados en el archivo BAM. A veces es necesario fusionar varios archivos BAM, por ejemplo cuando hay un mismo individuo que se ha procesado varias veces, es decir en canales o lanes diferentes de la misma célula de flujo.

3.2- Llamado de variantes

3.2.1- Obtención del archivo VCF a través de SAMtools y BCFtools

BCFtools es parte del paquete SAMtools. SAMtools recoge información del archivo BAM, computa probabilidades de los datos dado cada posible genotipo y almacena las probabilidades en el formato BCF. No llama variantes. BCFtools aplica lo anterior y hace el llamado real. Puede también concatenar archivos BCF, indexarlos y convertirlos en VCF. Además, BCFtools puede trabajar sobre algunos VCFs, pero no sobre todos.

3.2.2- Tratamiento y comparación de VCFs

Para ello tienen que ser comprimidos e indexados. El manejo se lleva a cabo con VCFtools, que es un paquete de programas diseñados para trabajar con archivos VCF. El objetivo de VCFtools es proporcionar métodos para trabajar con archivos VCF: validación, fusión, comparación y cálculo de algunas estadísticas básicas de genética de poblaciones, etc.

El formato de un VCF es el que se muestra a continuación, en la figura 40:

```

##fileformat=VCFv4.0

##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##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=.,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
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
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
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

```

Figura 40: Formato de un VCF

Los VCF tienen una cabecera cuyas líneas comienzan por # o ##, y líneas con datos, que no comienzan por #. La cabecera contiene información como por ejemplo, el programa empleado para generar el archivo, el número de versión del formato VCF o el nombre del genoma de referencia tal y cómo se observa en la figura 40.

Cada línea de datos debe tener 8 columnas:

1. CHROM: Cromosoma
2. POS: Posición en la referencia, con la primera base en posición 1
3. ID: Si el SNP está identificado previamente, aparecerá su rs.
4. REF: Base de referencia.
5. ALT: Bases alternativas que aparecen y que no son el alelo de referencia.
6. QUAL: Informa de la calidad. Altos valores indicarán que la llamada es fiable.
7. FILTER: PASS indica que esa posición ha pasado todos los filtros. Cuando no se aplican filtros, no aparece ningún valor.

Material y Métodos

8. INFO: Información adicional. Por ejemplo: DP indica la cobertura; DB si el SNP está en dbSNP; AA, el alelo ancestral, etc

Si la información sobre el genotipo está presente, aparece un campo adicional llamado FORMAT, dentro del campo INFO, mostrando los siguientes parámetros:

- GT: Genotipo. Un valor de 0 indica el alelo de referencia, un valor de 1, el primer alelo alternativo en la lista, y un valor de 2, el segundo.
- DP: Cobertura en esa posición para esa muestra.
- FT: Filtro de genotipo por muestra. Indica si ese genotipo fue “llamado”.
- GL: Probabilidad de que el genotipo indicado sea el correcto.
- GQ: Calidad del genotipo.
- HQ: Calidad de haplotipo.

3.3- Anotación funcional de los SNPs detectados.

Las grandes plataformas de secuenciación de alto rendimiento generan cantidades masivas información, pero sigue siendo un desafío identificar las variantes funcionalmente importantes. Entre las herramientas que existen para filtrar y anotar estas variantes están: ANNOVAR (Wang et al, 2010) y VEP (McLaren et al, 2010), ambas sirven para anotar variantes de nucleótido único (SNVs) e inserciones / deleciones, y su consecuencia funcional.

En este análisis se ha empleado la herramienta VEP (Variant Effect Predictor). VEP determina el efecto de las variantes (SNP, inserciones, deleciones, CNVs, o variantes estructurales) sobre los genes, transcritos, y secuencia de la proteína, así como sobre regiones reguladoras. Basta con introducir las coordenadas de las variantes, y los cambios de nucleótidos para conocer la siguiente información:

- Genes y transcritos afectados por las variantes.
- Ubicación de las variantes (por ejemplo, en la secuencia de codificación, en el ARN no codificante, en regiones reguladoras).
- Consecuencia a nivel de la secuencia de la proteína (por ejemplo, con cambio

de sentido, sin sentido...).

- Variantes conocidas que responden a las encontradas, y las MAF asociadas según el Proyecto 1000 Genomas.

VEP emplea dos herramientas: SIFT y PolyPhen, para dar información sobre los cambios a nivel de proteína y por tanto, sobre su posible efecto. SIFT predice si la sustitución de un aminoácido afecta a la función de la proteína basándose en la homología de secuencia y las propiedades físicas de los aminoácidos. PolyPhen es una herramienta que predice el posible impacto de una sustitución de aminoácidos en la estructura y función de una proteína humana utilizando consideraciones físicas y comparativas directas.

Las variaciones filtradas y seleccionadas por su importancia funcional han sido las de stop, cambio de sentido y de *splicing*, habiéndose anotado también en un inicio, todas las que están en regiones exónicas. A continuación se definen las variantes más importantes que puede detectar VEP:

- **Variación donante de *splicing* (Splice donor variant):** variante de *splicing* que cambia la región base 2 en el extremo 5' de un intrón.
- **Variación aceptora de *splicing* (Splice acceptor variant):** variante de *splicing* que cambia la región base 2 en el extremo 3' de un intrón.
- **Variación de ganancia de codón de parada (stop gained):** variante en la secuencia que hace que al menos una base de un codón cambie, dando lugar a un codón de stop prematuro y, por tanto a un péptido más corto.
- **Variación de pérdida de codón de parada (stop lost):** variante en la secuencia que hace que al menos una base de un codón cambie, dando lugar a un péptido más largo.
- **Variación con cambio de sentido (Missense variant):** variante en la secuencia que cambia una o más bases, resultando en una secuencia de aminoácidos diferente, pero con la misma longitud.

Material y Métodos

- **Variación en la región de *splicing* (Splice region variant):** variante en la secuencia en la que se ha producido un cambio dentro de la región del sitio de *splicing*, ya sea dentro de 13 bases del exón o 38 bases del intrón.
- **Variación sinónima (synonimus variant):** variante en la secuencia en que cambia una base, pero no cambia el aminoácido al que da lugar.
- **Variación en la zona 5' UTR (5 prime UTR variant):** variante en la zona 5' UTR.
- **Variación en la zona 3' UTR (3 prime UTR variant):** variante en la zona 3' UTR.
- **Variación en la secuencia no codificante exónica (non coding exon variant):** variante que cambia la secuencia no codificante de un exón.
- **Variación intrónica (Intron variant):** variante en un transcrito que se produce dentro de un intrón.
- **Variación en transcrito diana de NMD (NMD transcript variant):** variante en un transcrito que es diana de NMD (vía cuya función principal es la de reducir los errores en la expresión génica mediante la eliminación de los transcritos de ARNm que contienen codones de parada prematuros).
- **Variación en transcrito de ARN no codificante (nc transcript variant):** variante en un transcrito en un ARN no codificante.
- **Variación en la región 5' anterior a un gen (upstream gene variant):** variante en la secuencia de un gen que se localiza en 5'.
- **Variación en la región 3' posterior a un gen (downstream gene variant):** variante en la secuencia de un gen que se localiza en 3'.
- **Variación en sitio de unión de un factor de transcripción (TF binding site variant):** variante en la secuencia localizada en un sitio de unión de un factor de transcripción.
- **Elongación (Feature elongation):** variante de secuencia que causa la extensión de una característica genómica, con respecto a la secuencia de referencia.
- **Variación en una región reguladora (Regulatory region variant):** variante en la secuencia localizada en una región reguladora.

En la figura 41 aparece un esquema que refleja las variantes que puede detectar VEP.

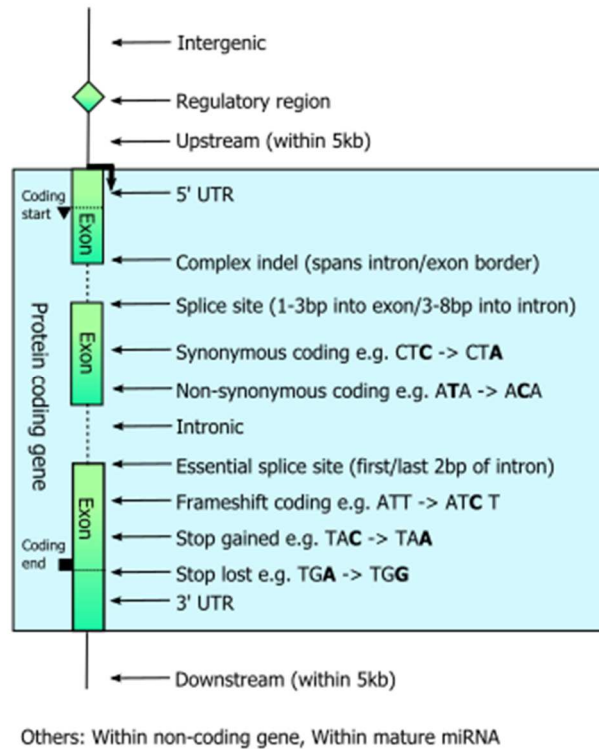


Figura 41: Esquema en que se representan los diferentes tipos de variaciones anotadas por VEP. McLaren et al. (2010). *Bioinformatics*. 26, 2069-2070.

En el anexo 1 (página 247) se muestran todos los comandos generales que se han empleado para llevar a cabo el análisis bioinformático completo de las muestras.

3.4- Genes analizados.

Tras una exhaustiva revisión bibliográfica, se seleccionaron 121 genes que se habían relacionado con diferentes tipos de diabetes hasta el momento de comenzar el análisis. Estos genes fueron identificados a través del análisis de genes candidatos y estudios de ligamiento en familias, diferentes GWAS (tanto sobre susceptibilidad a DM2 como GWAS sobre variables relacionadas) y posteriores metaanálisis.

Además, al tratarse de un proyecto coordinado, también se incluyeron genes relacionados con diabetes, los cuales se encuentran en investigación por los otros grupos. Los genes finalmente seleccionados se exponen en el apartado de resultados.

4- VALIDACIÓN DE LOS RESULTADOS OBTENIDOS POR SECUENCIACIÓN DE SANGER

La validación de los resultados obtenidos por NGS puede realizarse mediante secuenciación estándar o Sanger. Este estudio se realiza dado que puede existir una proporción de variantes que no sean reales debido a errores en la secuenciación, baja cobertura de algunas regiones, calidad de los datos, etc. Esta secuenciación se ha realizado mediante el sistema BigDye® Terminator v3.1 CycleSequencing kit (Applied Biosystems) y electroforesis capilar en un secuenciador ABI3730 (Applied Biosystems).

Esta secuenciación se basa en el método enzimático de Sanger, en el cual se emplean análogos dideoxirribonucleótidos (que carecen de grupo 3'OH) que se incorporan a la cadena creciente de ADN compitiendo con los nucleótidos normales y actuando como terminadores dando lugar a un bloqueo en la síntesis de la cadena. Al incorporarse de manera aleatoria se crea un conjunto de fragmentos de ADN de diferentes longitudes complementarios al molde de ADN que está siendo secuenciado. En la secuenciación automática basada en este método, en lugar de isótopos radiactivos para marcar los fragmentos de ADN, se utilizan terminadores marcados con diferentes fluoróforos.

El secuenciador o "Genetic Analyzer ABI3730" (Applied Biosystems), que aparece en la figura 42, funciona de manera totalmente automática inyectando las muestras en un capilar previamente cargado con un polímero que funciona como si fuera un gel de secuenciación, permitiendo separar fragmentos de ADN de cadena sencilla que se diferencian en una única base al tratarse de una electroforesis desnaturizante. A una altura determinada, el láser del dispositivo detecta la fluorescencia emitida por cada cadena sencilla de ADN marcado con fluorescencia y traduce dicha emisión en el fluorograma correspondiente y este en la secuencia o en los datos correspondientes.



Figura 42: Secuenciador ABI 3730.

Mediante programas de análisis tales como Sequencing analysis o SeqScape, se pueden analizar los datos que proporciona el secuenciador en forma de cromatogramas (figura 43), en los cuales se puede visualizar cada base y así observar si un individuo es homocigoto o heterocigoto para una determinada posición o si hay alguna variación con respecto a la secuencia de referencia.

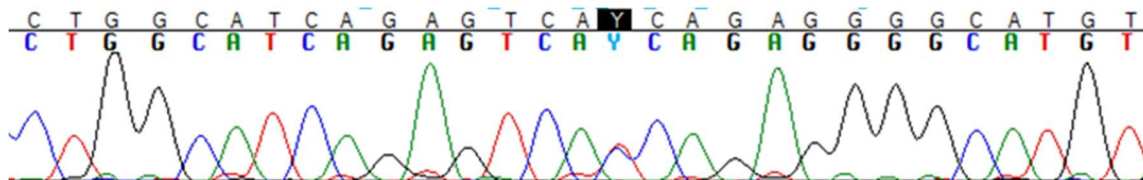


Figura 43: Imagen de un cromatograma en que se muestra un polimorfismo para el que el individuo es heterocigoto.

Resultados

RESULTADOS

Durante la realización de esta tesis, se han analizado 178 muestras, 89 procedentes de individuos afectados con DM2 y 89 controles, en los cuales se han obtenido las variaciones expuestas más adelante. En estos individuos, los genes analizados han sido elegidos mediante una exhaustiva revisión bibliográfica sobre los diferentes tipos de diabetes y los genes implicados en ellas.

Para la obtención de las variaciones, se ha aplicado un primer filtro de los datos basado en dos criterios: una profundidad de cobertura igual o superior a 4 y una calidad de genotipado igual o superior a 20. Nos hemos centrado en el estudio de las variaciones de nucleótido único (SNVs o SNPs) dejando para próximos estudios, variaciones tales como inserciones, deleciones, etc.

1- SELECCIÓN DE GENES A ANALIZAR

Tras la correspondiente revisión bibliográfica, se seleccionaron 121 genes previamente relacionados con diferentes tipos de diabetes. Estos genes fueron identificados a través del análisis de genes candidatos y estudios de ligamiento en familias, diferentes GWAS (tanto sobre susceptibilidad a DM2 como sobre variables relacionadas) y posteriores metaanálisis. Las palabras clave de la búsqueda fueron: diabetes, glucose, SNP, polymorphism, glucose metabolism alteration, obesity.

Además, también se incluyeron genes potencialmente relacionados con diabetes y cuyo efecto está en estudio por diferentes grupos con los que colabora nuestro grupo, pero se han obtenido resultados en alguno de ellos, como es el caso de CLEC7A.

Por otra parte, se han dejado para estudios posteriores los genes del cromosoma 6, debido a que en él se sitúa el complejo mayor de histocompatibilidad, por lo que se requiere un análisis bioinformático específico dada la alta variabilidad genética de esta región.

1.1- Listado de genes seleccionados

A continuación, en la tabla 10 se muestran todos los genes seleccionados para ser analizados y estudiados y también las referencias bibliográficas que los relacionan con los diferentes tipos de diabetes o con otras variables relacionadas: IMC, metabolismo lipídico, niveles de glucosa, etc. Además, en la figura 44, se muestra un cariograma en que aparecen todos los genes estudiados.

GEN	NOMBRE COMPLETO DEL GEN	OTRAS DENOMINACIONES	CROMOSOMA	REFERENCIAS BIBLIOGRÁFICAS
DIABETES SINDROMICAS				
WFS1	Wolfram syndrome 1 (wolframin)	CTRCT41, WFRS, WFS, WFSL	4	Barret TG, 2001
SLC19A2	solute carrier family 19 (thiamine transporter), member 2	RP1-206D15.4, TC1, THMD1, THT1, THTR1, TRMA	1	Barret TG, 2001
INSR	insulin receptor	CD220, HHF5	19	Barret TG, 2001
KCNJ11	potassium inwardly-rectifying channel, subfamily J, member 11	BIR, HHF2, IKATP, KIR6.2, PPH1, TNDM3	11	Barret TG, 2001
DIABETES NEONATAL				
ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	ABC36, HHF1, HI, HRINS, MRP8, PPH1, SUR, SUR1, SUR1delta2, TNDM2	11	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
GCK	glucokinase (hexokinase 4)	FGQTL3, GK, GLK, HHF3, HK4, HKIV, HXKP, LGLK, MODY2	7	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
IPF1	pancreatic and duodenal homeobox 1	GSF, IDX-1, IPF1, IUF1, MODY4, PDX-1, STF-1	13	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
PTFLA	pancreas specific transcription factor, 1a	PTF1-p48, bHLHa29	10	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
FOXP3	forkhead box P3	JM2, AIID, DIETER, IPEX, PIDX, XPID	X	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
EIF2AK3	eukaryotic translation initiation factor 2-alpha kinase 3	PEK, PERK, WRS	2	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009

Resultados

ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	ABC36, HNF1, H1, HRINS, MRP8, PHLI, SUR, SUR1, SUR1delta2, TNDM2	11	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
KCNJ11	potassium inwardly-rectifying channel, subfamily J, member 11	BIR, HNF2, IKATP, KIR6.2, PHLI, TNDM3	11	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
HNF1B	HNF1 homeobox B	FJHN, HNF-1B, HNF1beta, HNF2, HPC11, LF-83, LFB3, MODY5, TCF-2, TCF2, VHNF1	17	Aguilar-Bryan et al, 2008; Murphy et al, 2008; Vaxillaire et al, 2009
MODY				
MODY 1: HNF4A	hepatocyte nuclear factor 4, alpha	RP5-1013A22.1, HNF4, HNF4a7, HNF4a8, HNF4a9, HNF4alpha, MODY, MODY1, NR2A1, NR2A21, TCF, TCF14	20	Molven et al, 2011
MODY 2: GCK	glucokinase (hexokinase 4)	FGOTL3, GK, GLK, HNF3, HK4, HKIV, HXKP, LGLK, MODY2	7	Molven et al, 2011
MODY 3: HNF1A	HNF1 homeobox A	HNF-1A, HNF1, IDDM20, LFB1, MODY3, TCF-1, TCF1	12	Molven et al, 2011
MODY 4: IPF1	pancreatic and duodenal homeobox 1	GSF, IDX-1, IPF1, IUF1, MODY4, PDX-1, STF-1	13	Molven et al, 2011
MODY 5: HNF1B	HNF1 homeobox B	FJHN, HNF-1B, HNF1beta, HNF2, HPC11, LF-83, LFB3, MODY5, TCF-2, TCF2, VHNF1	17	Molven et al, 2011
MODY 6: NEUROD 1	neuronal differentiation 1	BETA2, BHF-1, MODY6, NEUROD, bHLHa3	2	Molven et al, 2011
MODY 7: KLF11	Kruppel-like factor 11	FKLF, FKLFI, MODY7, TIEG2, Tieg3	2	Molven et al, 2011
MODY 8: CEL	carboxyl ester lipase	BAL, BSD1, BSSL, CELL, CEase, FAP, FAPP, LIPA, MODY8	9	Molven et al, 2011
MODY 9: PAX4	paired box 4	KPD, MODY9	7	Molven et al, 2011

MODY 10: INSR	insulin receptor	CD220, HNF5	19	Molven et al, 2011
MODY 11: BLK	B lymphoid tyrosine kinase	MODY11	8	Molven et al, 2011
DIABETES MELLITUS TIPO 1				
INS	insulin	IDDM1, IDDM2, ILPR, IRDN, MODY10	11	Polychronakos y Li, 2011
PTPN22	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	LYP, LYP1, LYP2, PEP, PTPN8	1	Polychronakos y Li, 2011
ERBB3	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3	ErbB-3, HER3, LCCS2, MDA-BF-1, c-erbB-3, c-erbB3, erbB3-5, p180-ErbB3, p45-sErbB3, p85-sErbB3	12	Polychronakos y Li, 2011
CTLA4	cytotoxic T-lymphocyte-associated protein 4	CD, CD152, CELIAC3, CTLA-4, GRD4, GSE, IDDM12	2	Polychronakos y Li, 2011
IFIH1	interferon induced with helicase C domain 1	HLcd, IDDM19, MDA-5, MDA5, RLR-2	2	Polychronakos y Li, 2011
IL2RA	interleukin 2 receptor, alpha	RP11-536K7.1, CD25, IDDM10, IL2R, TCGFR	10	Polychronakos y Li, 2011
CLECT16A	C-type lectin domain family 16, member A	Gop-1, KIAA0350	16	Polychronakos y Li, 2011
PTPN2	protein tyrosine phosphatase, non-receptor type 2	PTN2, PTP, TC-PTP, TCELLPTP, TCPPTP	18	Polychronakos y Li, 2011
INSR	insulin receptor	CD220, HNF5	19	Polychronakos y Li, 2011
DIABETES MELLITUS TIPO 2				
IRS1	insulin receptor substrate 1	HIRS-1	2	Voight et al, 2010
IRS2	insulin receptor substrate 2	IRS-2	13	Vilà et al, 2011
KCNJ11	potassium inwardly-rectifying channel, subfamily J, member 11	BIR, HNF2, IKATP, KIR6.2, PHLI, TNDM3	11	Horikawa, 2000
HNF1A	HNF1 homeobox A	HNF-1A, HNF1, IDDM20, LFB1, MODY3, TCF-1, TCF1	12	Voight et al, 2010
HNF1B	HNF1 homeobox B	FIHN, HNF-1B, HNF1beta, HNF2, HPC11, LF-B3, LFB3, MODY5, TCF-2, TCF2, VHNF1	17	Marselli et al, 2010

Resultados

HNMF4A	hepatocyte nuclear factor 4, alpha	RP5-1013A22.1, HNMF4, HNMF4a7, HNMF4a8, HNMF4a9, HNMF4alpha, MODY, MODY1, NR2A1, NR2A21, TCF, TCF14	20	Kooner et al, 2011
IPF1	pancreatic and duodenal homeobox 1	GSF, IDX-1, IPF1, IUF1, MODY4, PDX-1, STF-1	13	Pinney y Simmons, 2010
IB1	mitogen-activated protein kinase 8 interacting protein 1	JIP-1, JIP1, PRKM8IP, MAPK8IP1	11	Milkossy et al, 2010
NEUROD1	neuronal differentiation 1	BETA2, BHF-1, MODY6, NEUROD, bHLHa3	2	Lefebvre et al, 2010
WFS1	Wolfram syndrome 1 (wolframin)	CTRCT41, WFRS, WFS, WFSL	4	Voight et al, 2010
ADIPOQ	adiponectin, C1Q and collagen domain containing	ACDC, ACRP30, ADIPQTL1, ADPN, APM-1, APM1, GBP28	3	Gong et al, 2010
PPARA	peroxisome proliferator-activated receptor alpha	SC22CB-5E3.5, NR1C1, PPAR, PPARalpha, hPPAR	22	Maida et al, 2011
TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)	RP11-357H24.1, TCF-4, TCF4	10	Grant et al, 2006; Cauchi et al, 2007
CAPN10	calpain 10	CANP10, NIDDM1	2	Gloyn et al, 2003
APOA2	apolipoprotein A-II	Apo-AII, ApoA-II, apoAII	1	Hasstedt et al, 2008; Fryirs et al, 2010
PKLR	pyruvate kinase, liver and RBC	PK1, PKL, PKR, PKRL, RPK	1	Hasstedt et al, 2008
LMX1A	LIM homeobox transcription factor 1, alpha	LMX1, LMX1.1	1	Wongseree et al, 2009
LMX1B	LIM homeobox transcription factor 1, beta	RP11-489N22.3, LMX1.2, NP51	9	Speliotes et al, 2010
KCNQ1	potassium voltage-gated channel, KQT-like subfamily, member 1	ATFB1, ATFB3, JLNS1, KCNA8, KCNA9, KVLQT1, KV1.9, KV7.1, LOT, LQT1, RWS, SQT2, WRS	11	Voight et al, 2010
DUSP9	dual specificity phosphatase 9	MKP-4, MKP4	X	Voight et al, 2010
CDKN2A	cyclin-dependent kinase inhibitor 2A	ARF, CDK4I, CDKN2, CMM2, INK4, INK4A, MLM, MTS-1, MTS1, P14, P14ARF, P16, P16-INK4A, P16INK4, P16INK4A, P19, P19ARF, TP16	9	Zeggini et al, 2007

CDKN2B	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	RP11-149I2.1, CDK4I, INK4B, MTS2, P15, TP15, p15INK4b	9	Zeggini et al, 2007
IGF2BP2	insulin-like growth factor 2 mRNA binding protein 2	IMP-2, IMP2, VICKZ2	3	Burton et al, 2007
FTO	fat mass and obesity associated	ALKBH9	16	Burton et al, 2007
THADA	thyroid adenoma associated	GITA	2	Zeggini et al, 2008
ARAP1	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1	CENTD2	11	Voight et al, 2010
HHEX	hematopoietically expressed homeobox	HEX, HMPH, HOX11L-PEN, PRH, PRHX	10	Sladek et al, 2007
NOTCH2	notch 2	AGS2, HJCYS, hn2	1	Zeggini et al, 2008
ADCY5	adenylate cyclase 5	ACS, FDFM	3	Dupuis et al, 2010
SLC30A8	solute carrier family 30 (zinc transporter), member 8	ZNT8, Znt-8	8	Scott et al, 2007
CAMK1D	calcium/calmodulin-dependent protein kinase ID	RP11-462F15.1, CKUK, CaM-K1, CaMKID	10	Zeggini et al, 2008
CDCC123	cell division cycle 123	RP11-186N15.4, C10orf7, D123	10	Zeggini et al, 2008
RBM51	RNA binding motif, single stranded interacting protein 1	C2orf12, HCC-4, MSSP, MSSP-1, MSSP-2, MSSP-3, SCR2, YC1	2	Qi et al, 2010
ITGB6	integrin, beta 6	ITGB6	2	Qi et al, 2010
TLE4	transducin-like enhancer of split 4	RP11-79D8.3, BCE-1, BCE1, E(spl), ESG, ESG4, GRG4	9	Vassy y Meigs, 2012
HMGGA2	high mobility group AT-hook 2	BABL, HMG1-C, HMG1C, LIPO, STQTL9	12	Voight et al, 2010
JAZF1	JAZF zinc finger 1	tcag7.981, TIP27, ZNF802	7	Zeggini et al, 2008
ADAMTS9	ADAM metalloproteinase with thrombospondin type 1 motif, 9	ADAMTS9	3	Zeggini et al, 2008
MTNR1B	melatonin receptor 1B	FGQTL2, MEL-1B-R, MT2	11	Lyssenko et al, 2009; Prokopenko et al, 2009
TSPAN8	tetraspanin 8	CO-029, TM4SF3	12	Zeggini et al, 2008

Resultados

LGR5	leucine-rich repeat containing G protein-coupled receptor 5	FEX, GPR49, GPR67, GRP49, HG38	12	Zeggini et al, 2008
BCL11A	B-cell CLL/lymphoma 11A (zinc finger protein)	BCL11A-L, BCL11A-S, BCL11A-XL, BCL11a-M, CTIP1, EVI9, HBFQTL5, ZNF856	2	Voight et al, 2010
ZBED3	zinc finger, BED-type containing 3	ZBED3	5	Voight et al, 2010
GCK	glucokinase (hexokinase 4)	FGQTL3, GK, GLK, HHF3, HK4, HKIV, HXKP, LGLK, MODY2	7	Dupuis et al, 2010
KLF14	Kruppel-like factor 14	BTEB5	7	Voight et al, 2010
OASL	2'-5'-oligoadenylate synthetase-like	OASLd, TRIP-14, TRIP14, p59 OASL, p59-OASL, p59OASL	12	Voight et al, 2010
PRC1	protein regulator of cytokinesis 1	ASE1	15	Voight et al, 2010
PROX1	prospero homeobox 1	PROX1	1	Dupuis et al, 2010
DGKB	diacylglycerol kinase, beta 90kDa	DAGK2, DGK, DGK-BETA	7	Dupuis et al, 2010
TMEM195	alkylglycerol monoxygenase	tcag7.1136, AGMO	7	Dupuis et al, 2010
GCKR	glucokinase (hexokinase 4) regulator	FGQTL5, GKR	2	Dupuis et al, 2010
TP53INP1	tumor protein p53 inducible nuclear protein 1	SIP, TP53DINP1, TP53INP1A, TP53INP1B, Teap, p53DINP1	8	Voight et al, 2010
ZFAND6	zinc finger, AN1-type domain 6	HT032, AWP1, ZA20D3, ZFAND5B	15	Voight et al, 2010
SREBF1	sterol regulatory element binding transcription factor 1	SREBP-1c, SREBP1, bHLHD1	17	Shao et al, 2010
SREBF2	sterol regulatory element binding transcription factor 2	CTA-250D10.14-005, SREBP-2, SREBP2, bHLHD2	22	Langhi y Carrou, 2010
SCAP	SREBF chaperone	PSEC0227	3	Sato, 2010
INSIG1	insulin induced gene 1	CL-6, CL6	7	Sato, 2010
S1P	membrane-bound transcription factor peptidase, site 1	PCKS8, SKI-1, MBTPS1	16	Cantrell Stanford et al, 2012

S2P	membrane-bound transcription factor peptidase, site 2	BRESEK, IFAP, KFSD, KFSDX, OLMSX, S2P, MBTPS2	X	Sato, 2010
FABP1	fatty acid binding protein 1, liver	FABPL, L-FABP	2	Mansego et al, 2012
FABP2	fatty acid binding protein 2, intestinal	FABPI, I-FABP	4	Albala et al, 2007
FABP3*	fatty acid binding protein 3, muscle and heart (mammary-derived growth inhibitor)	FABP11, H-FABP, M-FABP, MDGI, O-FABP	1	Shin et al, 2003
FABP4	fatty acid binding protein 4, adipocyte	A-FABP, AFABP, ALBP, HEL-S-104, aP2	8	Chan et al, 2010
FABP5*	fatty acid binding protein 5 (psoriasis-associated)	E-FABP, EFABP, KFABP, PA-FABP, PAFABP	8	Bu et al, 2011
FABP6	fatty acid binding protein 6, ileal	I-15P, I-BABP, I-BALB, I-BAP, ILBP, ILBP3, ILLBP	5	Fisher et al, 2009
FABP9	fatty acid binding protein 9, testis	PERF, PERF15, T-FABP	8	Smathers y Petersen, 2011
FABP12	fatty acid binding protein 12	-	8	Smathers y Petersen, 2011
PMP2	peripheral myelin protein 2	FABP8, M-FABP, MP2, P2	8	Smathers y Petersen, 2011
GCG	glucagon	GLP1, GLP2, GRPP	2	Hayashi, 2011
GCGR	glucagon receptor	GGR, GL-R	17	Hayashi, 2011
GENES IDENTIFICADOS POR GRUPOS CON LOS QUE COLABORAMOS				
CD163	CD163 molecule	M130, MM130	12	Møller, 2012
CD36	CD36 molecule (thrombospondin receptor)	BDPLT10, CHDS7, FAT, GP3B, GP4, GPIV, PASIV, SCARB3	7	Tan et al, 2005
CLEC7A	C-type lectin domain family 7, member A	UNQ539/PRO1082, BGR, CANDF4, CLECSF12, DECTIN1	12	Polychronakos y Li, 2011

* Se han estudiado algunos pseudogenes de estos genes:

FABP3P2, FABP5P1, FABP5P2, FABP5P3, FABP5P4, FABP5P5, FABP5P6, FABP5P7, FABP5P10, FABP5P11, FABP5P12, FABP5P13, FABP5P14, FABP5P15.

Tabla 10: Listado de genes según enfermedad

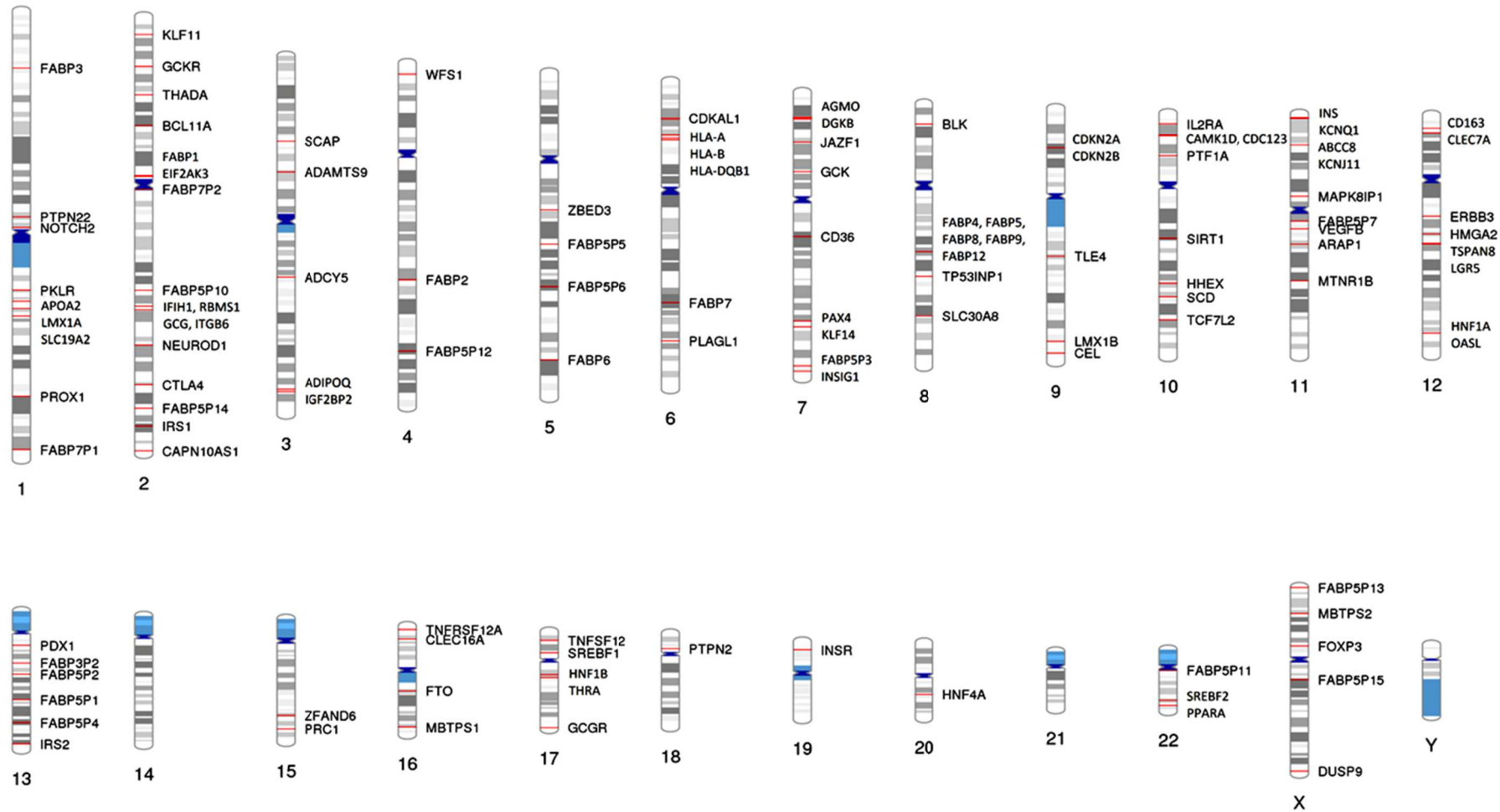


Figura 44: Cariograma en el que se representan todos los genes analizados en el estudio

2- RESULTADOS RELATIVOS A LA COBERTURA

Como se ha indicado previamente, para la obtención de las variaciones, se ha aplicado un primer filtro de los datos basado en dos criterios: una profundidad de cobertura igual o superior a 4 y una calidad de genotipado igual o superior a 20. La principal limitación para no obtener una mayor cobertura es económica, dado que el coste aumenta proporcionalmente a su incremento. La cobertura obtenida para cada una de las regiones no es homogénea, tal y como sucede en este tipo de estudios. En este apartado se han seleccionado dos genes para ser mostrados como ejemplo por las diferencias que presentan en cuanto a éste parámetro: ABCC8 e INSIG1.

2.1- Cobertura por exón.

Se ha obtenido una cobertura media aproximada de 20x de todas las regiones analizadas (veáse Tabla 11), que era la cobertura deseada para maximizar el número de muestras con los recursos disponibles.

Cobertura media (X)			
	Controles	Casos	Total
Dentro de los exones*	18.07	13.69	15.70
Dentro de las sondas	22.41	16.71	19.33

Tabla 11: Cobertura media por exón en casos y controles. Los datos relativos a los exones se refieren a cómo éstos están cubiertos, ya que las sondas están diseñadas para cubrir la mayoría de exones, pero no cubren estrictamente cada exón, aunque esto depende del kit empleado.
*(Se incluyen todos los exones del gen, incluso los no cubiertos por las sondas.)

Las siguientes figuras (45 y 46) muestran la cobertura en cada posición de las regiones transcritas, es decir, la concatenación de cada exón de los genes usados como modelo. Los límites de los exones se representan como líneas verticales y las sondas de captura se dibujan como cintas grises horizontales. En la figura 45 se muestra el ejemplo del gen ABCC8, del cual se ha obtenido una buena cobertura media en todos los exones, tanto en casos como en controles. Por otro lado, en la figura 46 se muestra el ejemplo del gen INSIG1, del cual se observa claramente que la cobertura media obtenida ha sido menor que la observada en otros genes.

Resultados

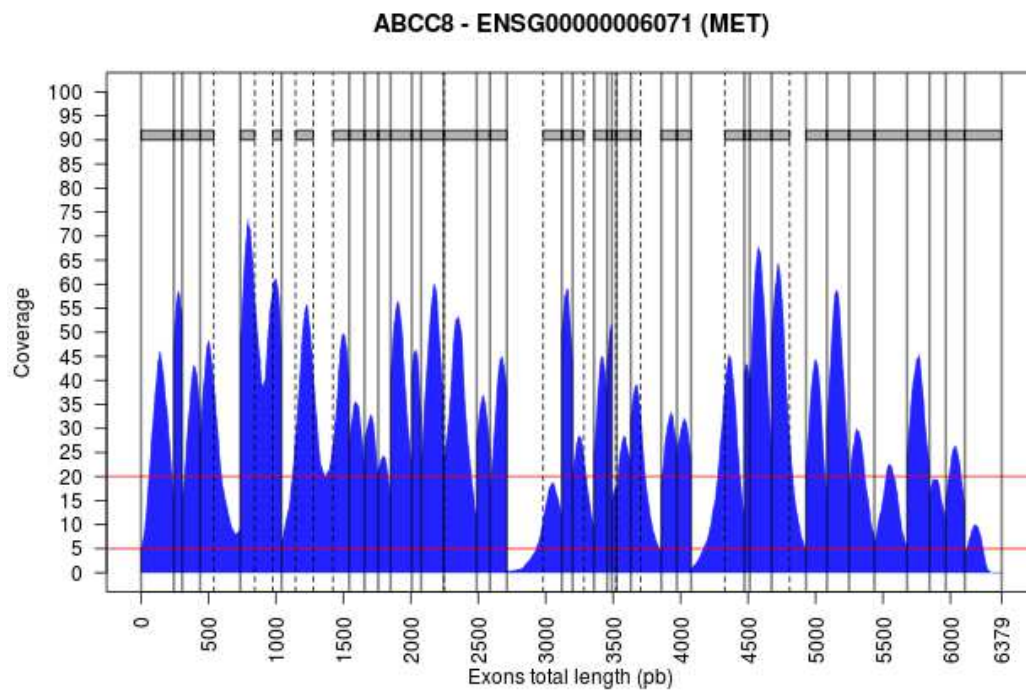
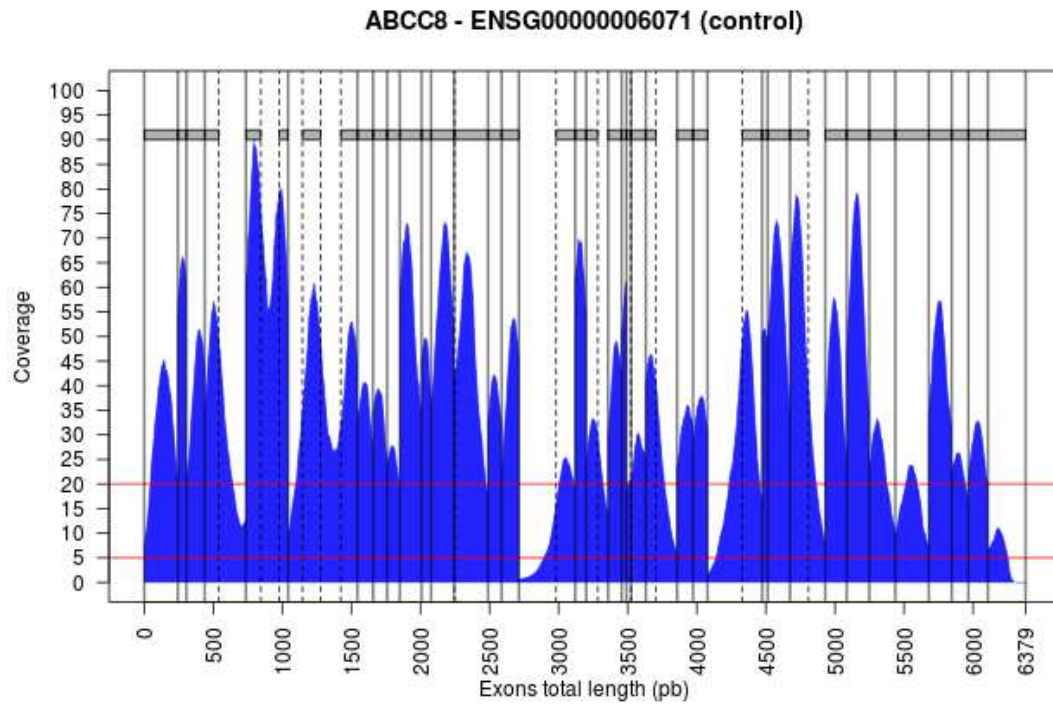


Figura 45: Cobertura media obtenida para el gen ABCC8 (superior: en controles, inferior: en casos)

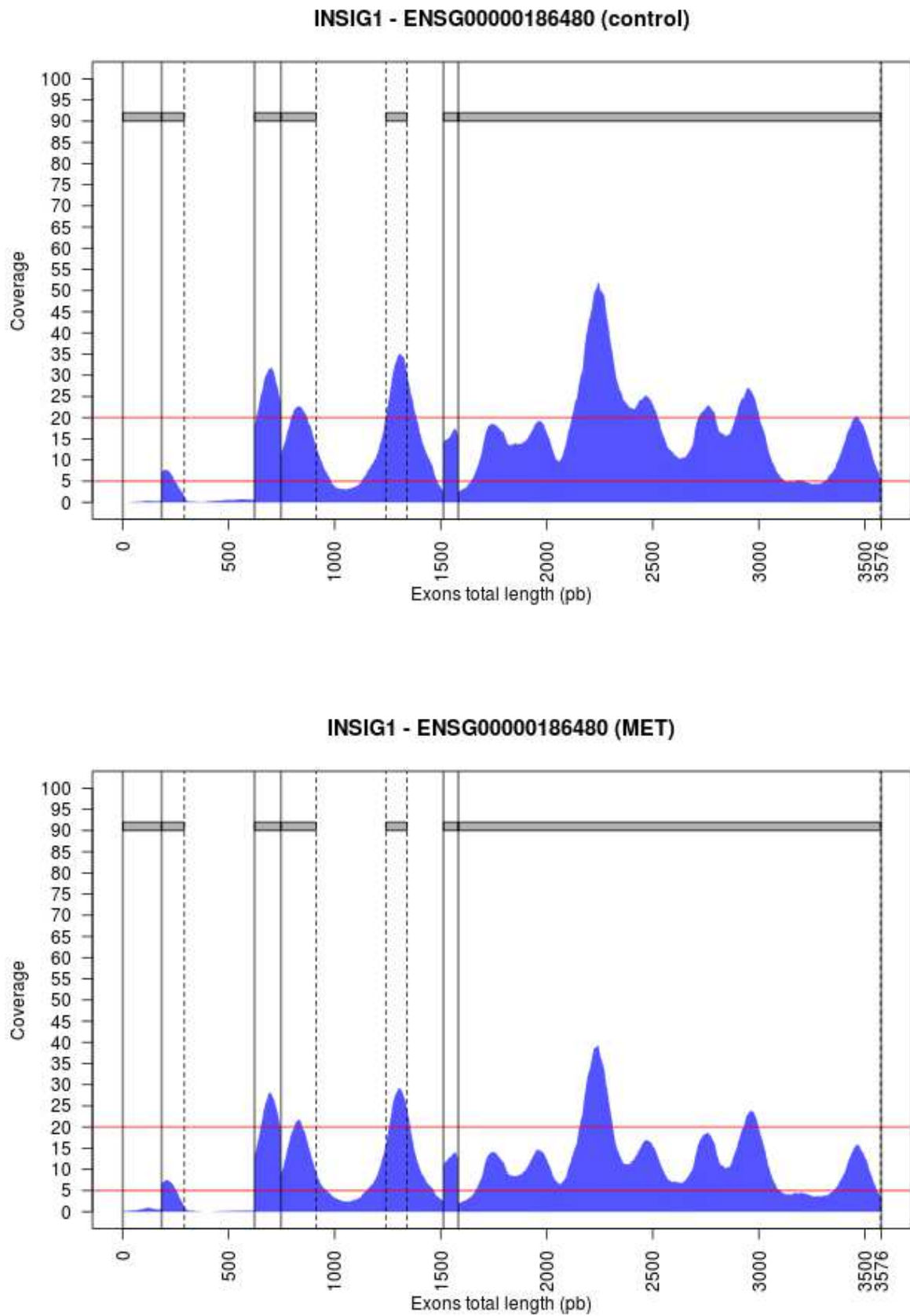


Figura 46: Cobertura media obtenida para el gen INSIG1 (superior: en controles, inferior: en casos)

2.2- Porcentaje de bases con una cobertura mínima

En cuanto a la cobertura por base, los resultados se muestran en la tabla 12 y han sido los siguientes:

	% de bases con una cobertura mínima de						
	1X	5X	10X	15X	20X	25X	30X
Dentro de los exones*	72.19	56.03	46.08	37.27	29.62	23.33	18.32
Dentro de las sondas	90.47	73.85	57.59	44.50	34.28	26.43	20.47

Tabla 12: Resultados del porcentaje de bases con una cobertura determinada. Los datos relativos a los exones se refieren a cómo éstos están cubiertos, ya que las sondas están diseñadas para cubrir la mayoría de exones, pero no cubren estrictamente cada exón, aunque esto depende del kit empleado. *(Se incluyen todos los exones del gen, incluso los no cubiertos por las sondas).

Los siguientes gráficos muestran el porcentaje de bases frente a la cobertura mínima que tienen. Cada muestra se representa como una línea gris y la media, como una línea roja.

A continuación, en las figuras 47 y 48, se muestran los dos genes del ejemplo anterior: ABCC8 e INSIG1.

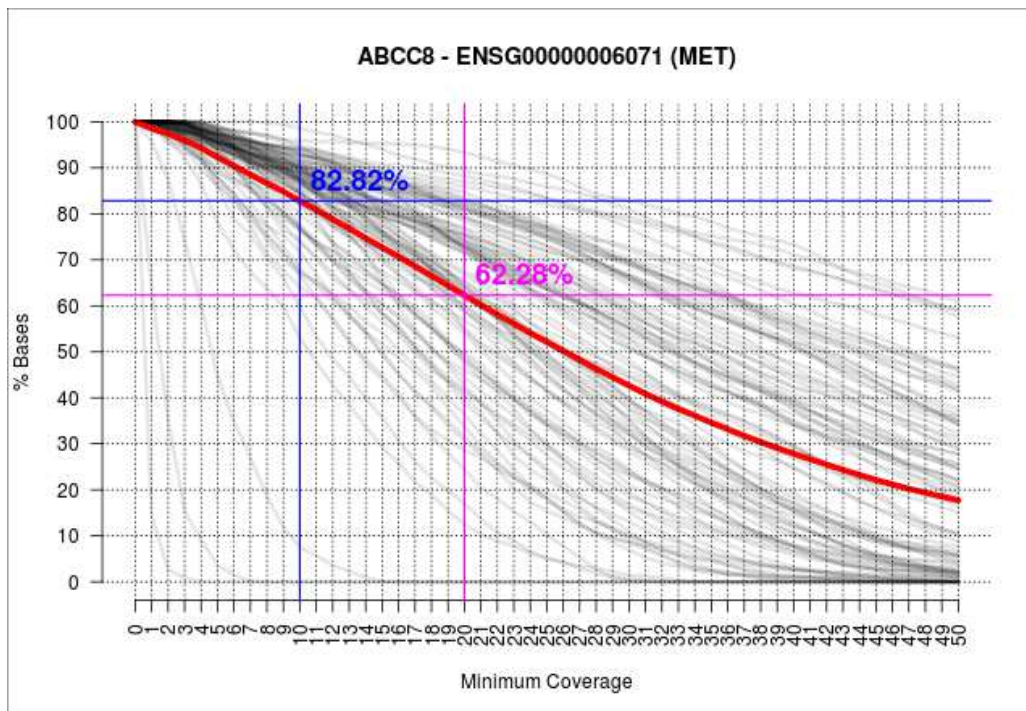
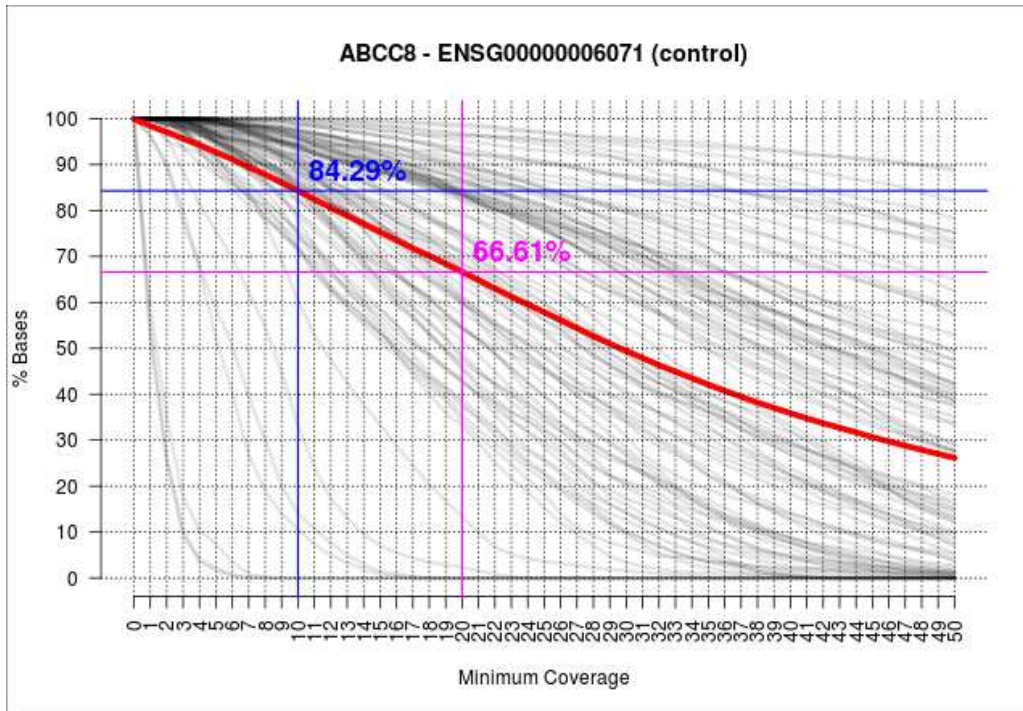


Figura 47: Cobertura por base obtenida para el gen ABCC8 (superior: en controles, inferior: en casos). Cada muestra se representa como una línea gris y la media, como una línea roja. Se puede observar que más del 80% de bases presentan una cobertura mínima de 10x y más del 60%, de 20x.

Resultados

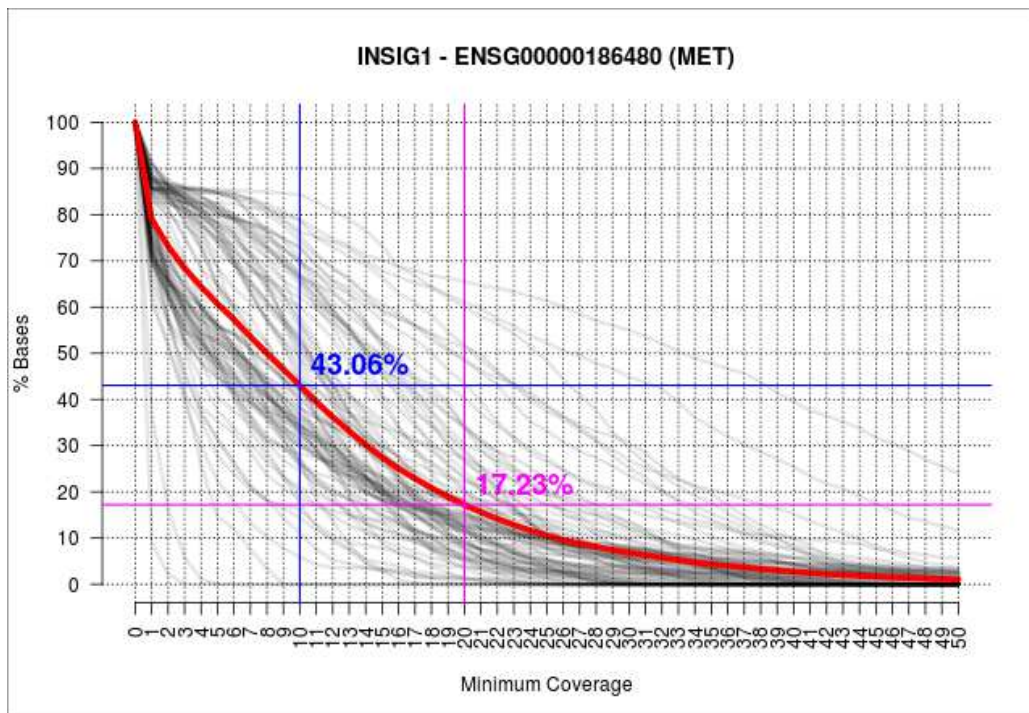
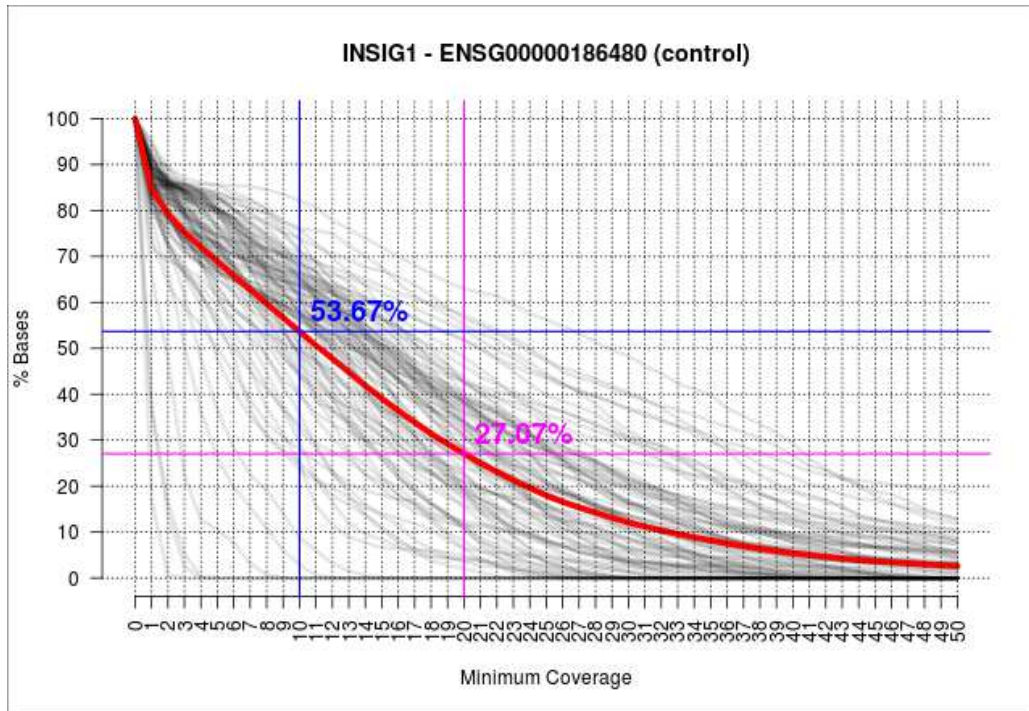
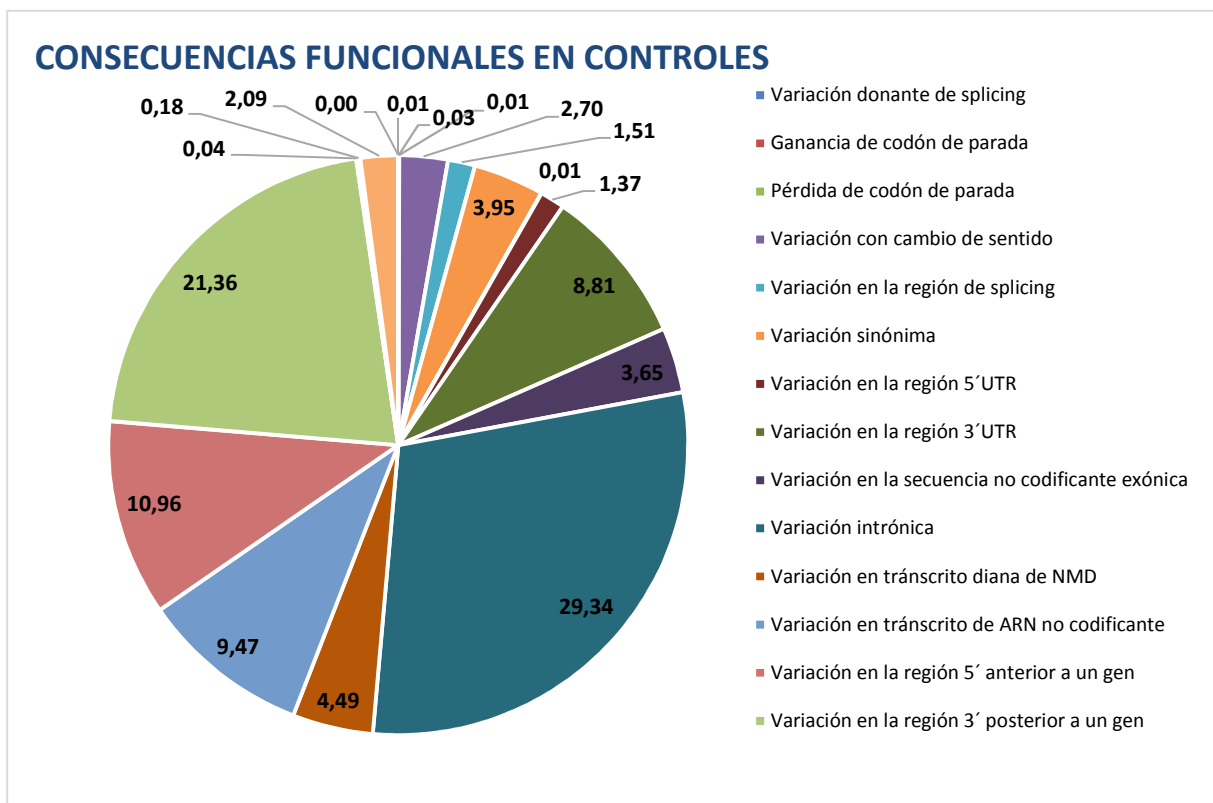


Figura 48: Cobertura por base obtenida para el gen INSIG1 (superior: en controles, inferior: en casos). Cada muestra se representa como una línea gris y la media, como una línea roja. Se puede observar que más del 50% (40% en casos) de bases presentan una cobertura mínima de 10x y más del 20% (17% en casos), de 20x.

3- POSIBLES CONSECUENCIAS FUNCIONALES DE LAS VARIACIONES OBTENIDAS:

Tras la secuenciación de los 178 exomas (89 controles y 89 casos) y el posterior análisis de los genes previamente relacionados con diferentes tipos de diabetes, se obtuvo una gran cantidad de variaciones, las cuales se muestran a continuación (véase tabla 13).

Aplicando solamente los 2 filtros iniciales: profundidad de cobertura ≥ 4 y calidad de genotipado ≥ 20 , se han obtenido 9838 variaciones en controles, de las cuales el 70.7% son exónicas y el 29.3 % son intrónicas. En cuanto a los casos, se han obtenido un total de 8863 variaciones, de las que el 72.8% se encuentran en el exoma y el 27.1% son intrónicas. Los diferentes tipos de variaciones obtenidas se representan a continuación, en la gráfica 49.



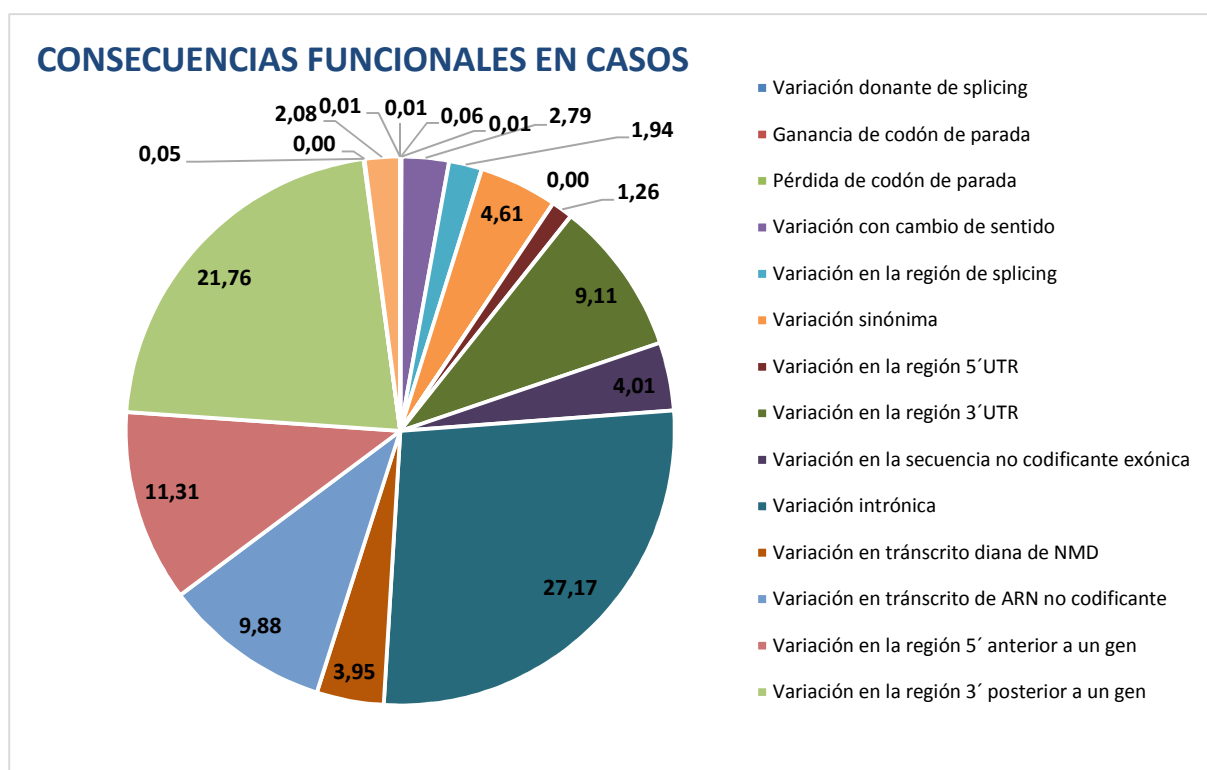


Figura 49: Posibles consecuencias funcionales de las variaciones encontradas (superior: en controles, inferior: en casos). Los números representan el % de cada tipo de variación.

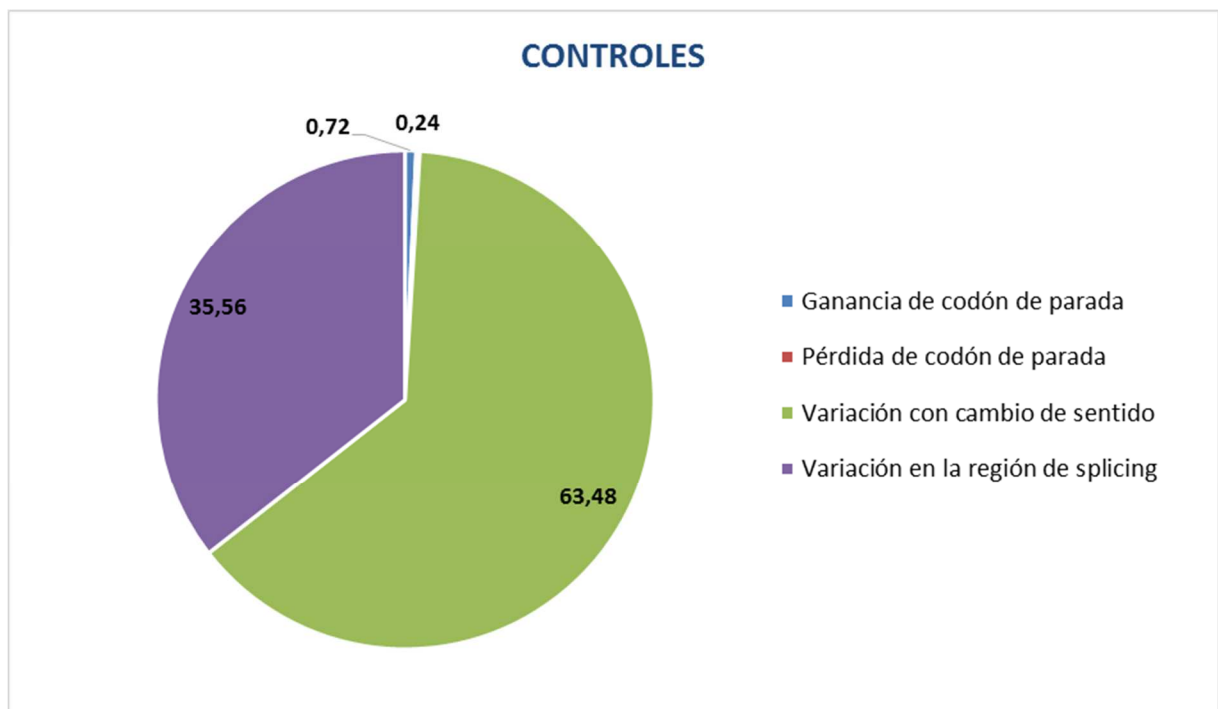
	CONTROLES	CASOS
Variación donante de <i>splicing</i>	1	1
Ganancia de codón de parada	3	5
Pérdida de codón de parada	1	1
Variación con cambio de sentido	266	247
Variación en la región de <i>splicing</i>	149	172
Variación sinónima	389	409
Variación en la región 5' UTR	135	112
Variación en la región 3' UTR	867	807
Variación en la secuencia no codificante exónica	359	355
Variación intrónica	2886	2408
Variación en transcrito diana de NMD	442	350
Variación en transcrito de ARN no codificante	932	876
Variación en la región 5' anterior a un gen	1078	1002
Variación en la región 3' posterior a un gen	2101	1929

Variación en sitio de unión de un factor de transcripción	4	4
Elongación	18	0
Variación en una región reguladora	206	184
Variación aceptora de <i>splicing</i>	0	1
TOTAL	9838	8863

Tabla 13: Posibles consecuencias funcionales de las variaciones encontradas

3.1- Consecuencias más importantes:

Se han destacado como más importantes, las variaciones con mayor efecto funcional: variaciones de pérdida o ganancia de codón de parada, con cambio de sentido y las que afectan a las regiones de *splicing*. Los resultados obtenidos se reflejan en la figura 50 y en la tabla 14.



Resultados

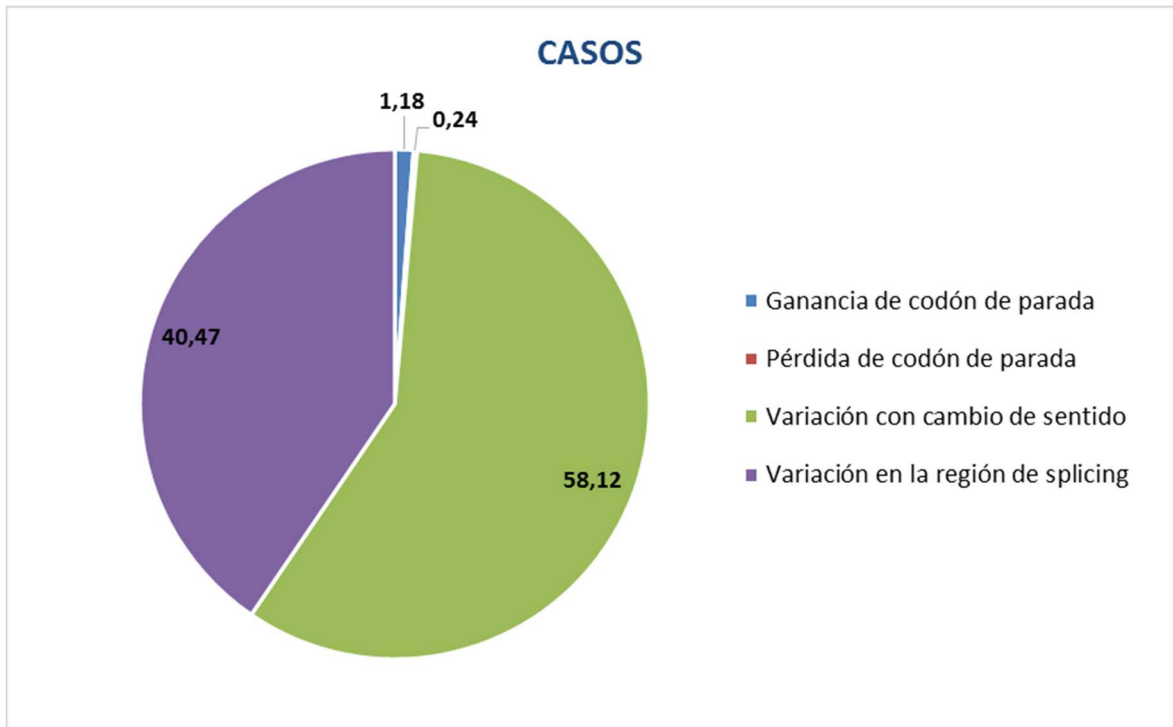


Figura 50: Consecuencias funcionales más importantes de las variaciones encontradas (superior: en controles, inferior: en casos). Los números representan el % de cada tipo de variación.

	CONTROLES	CASOS
Ganancia de codón de parada	3	5
Pérdida de codón de parada	1	1
Variación con cambio de sentido	266	247
Variación en la región de <i>splicing</i>	149	172

Tabla 14: Consecuencias funcionales más importantes de las variaciones encontradas en casos y controles.

4- EFECTOS FUNCIONALES DE LAS VARIACIONES.

Para determinar las posibles consecuencias funcionales que puede tener una variación se emplean principalmente 2 herramientas: SIFT y PolyPhen. SIFT predice si la sustitución de un aminoácido afecta a la función de la proteína basándose en la homología de secuencia y las propiedades físicas de los aminoácidos. PolyPhen es una herramienta que predice el posible impacto de una sustitución de aminoácidos en la estructura y función de una proteína humana utilizando consideraciones físicas y comparativas directas.

Los resultados obtenidos con SIFT se muestran en la figura 51 y la tabla 15; y los obtenidos con PolyPhen, en la figura 52 y la tabla 16.

4.1- Según SIFT.

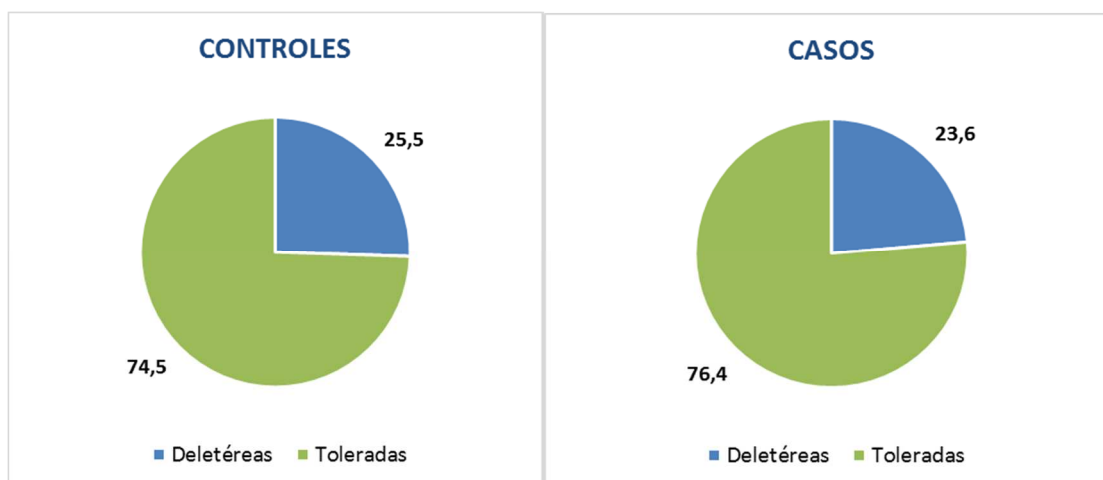


Figura 51: Consecuencias funcionales de las variaciones encontradas según SIFT (izquierda: en controles, derecha: en casos). Los números representan el % de cada tipo de variación.

	CONTROLES	CASOS
Deletéreas	67	58
Toleradas	196	188

Tabla 15: Consecuencias funcionales de las variaciones encontradas en casos y controles, según SIFT.

4.2- Según PolyPhen.

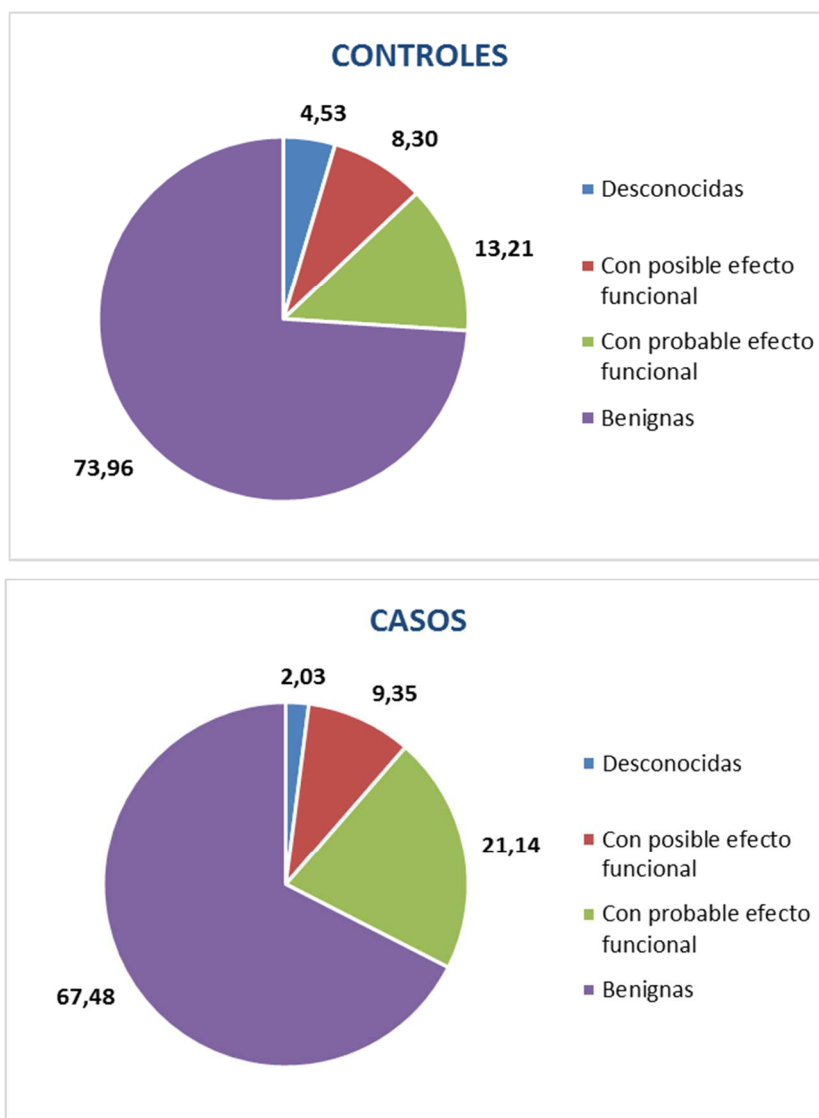


Figura 52: Consecuencias funcionales de las variaciones encontradas según PolyPhen (superior: en controles, inferior: en casos). Los números representan el % de cada tipo de variación.

	CONTROLES	CASOS
Desconocidas	12	5
Con posible efecto negativo	22	23
Con probable efecto negativo	35	52
Benignas	196	166

Tabla 16: Consecuencias funcionales de las variaciones encontradas en casos y controles, según PolyPhen.

5- SELECCIÓN DE VARIANTES MÁS RELEVANTES

De entre todas las variaciones encontradas (véase tabla 13), las elegidas para este estudio han sido las consideradas con mayores consecuencias funcionales: variaciones con cambio de sentido, de pérdida o ganancia de codón de parada y que afectan a las regiones de *splicing*. Y de entre todas ellas, se seleccionaron las que afectan a la funcionalidad de la proteína, según PolyPhen, siendo este resultado corroborado por SIFT, que lo expresa como que la variación es deletérea. Se obtuvieron aproximadamente 80 variaciones de este tipo. Posteriormente se seleccionaron para ser estudiadas y validadas las que tenían una MAF<0.2 en la población europea del proyecto 1000 Genomas y además, se excluyeron las que estaban presentes tanto en casos como en controles de manera coincidente.

Finalmente 20 variaciones se seleccionaron para ser validadas por secuenciación capilar. De este total, 16 resultaron ser verdaderos positivos. En las tablas 17 y 18 se muestran las variaciones seleccionadas. Y en la figura 53 se muestra un esquema que representa cómo se hizo la selección de variantes:

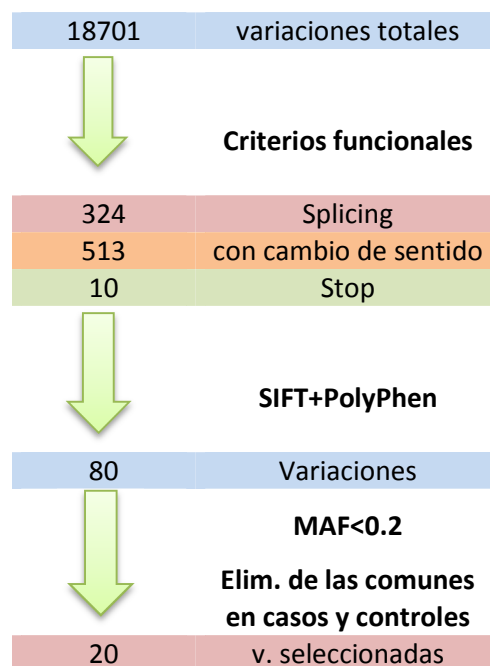


Figura 53: Representación de los criterios de selección y del número de variantes obtenidas durante el proceso de análisis.

Resultados

5.1- Variaciones seleccionadas en controles

Se seleccionaron 8 variaciones con una MAF <0.2 en la población europea del proyecto 1000 Genomas y que alteraban la funcionalidad de la proteína tanto para SIFT como para PolyPhen. Se muestran en la tabla 17:

Posición cromosómica	Gen	Consecuencia funcional	rs	MAF europea	Muestra	Genotipo	Cambio de Aa	Codón	Puntuación POLYPHEN	Puntuación SIFT
2:27745372	GCKR	Ganancia de stop	rs146053779	0.0040	170056	CT	Arg/*	Cga/Tga	-	-
2:43818077	THADA	Variación con cambio de sentido	rs10210191	-	170073	AG	Glu/Gly	gAg/gGg	Posible efecto funcional (0.804)	Deletérea (0.01)
2:160994197	ITGB6	Variación con cambio de sentido	rs55841905	0.01	280152	AG	Gly/Arg	Ggg/Agg	Posible efecto funcional (0.489)	Deletérea (0)
2:163144780	IFIH1	Variación con cambio de sentido	-	-	170056	CG	Leu/Phe	ttG/ttC	Probable efecto funcional (0.982)	Deletérea (0)
10:114925406	TCF7L2	Variación con cambio de sentido	rs77673441	-	170067	CG	Pro/Arg	cCc/cGc	Posible efecto funcional (0.501)	Deletérea (0.05)
11:17436078	ABCC8	Variación con cambio de sentido	-	-	170055	AT	Ser/Cys	Agt/Tgt	Probable efecto funcional (0.964)	Deletérea (0.02)
12:56494932	ERBB3	Variación con cambio de sentido	rs77822103	-	170056	CT	Ser/Pro	Tca/Cca	Probable efecto funcional (0.999)	Deletérea (0)
12:66308656	HMG2	Variación con cambio de sentido	rs139459932	0.04	290780	CT	Cys/Arg	Tgt/Cgt	Probable efecto funcional (0.983)	Deletérea (0.01)

Tabla 17: Variaciones seleccionadas en controles

5.2- Variaciones seleccionadas en casos

Se seleccionaron 12 variaciones con una MAF <0.2 en la población europea del proyecto 1000 Genomas y que alteraban la funcionalidad de la proteína tanto para SIFT como para PolyPhen. Se muestran en la tabla 18:

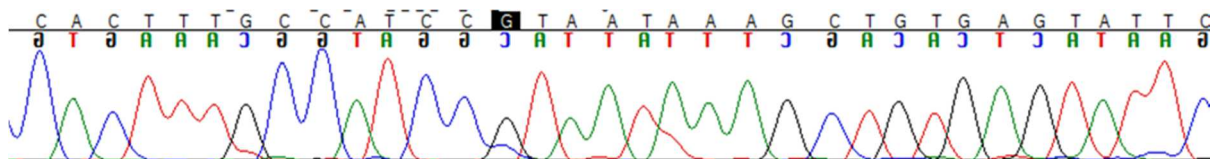
Posición cromosómica	Gen	Consecuencia funcional	rs	MAF europea	Muestra	Genotipo	Cambio de Aa	Codón	Puntuación POLYPHEN	Puntuación SIFT
1:114397536	PTPN22	Variación con cambio de sentido	-	-	MET_00035	AC	His/Asn	Cac/Aac	Probable efecto funcional (0.976)	Deletérea (0.03)
1:120458122	NOTCH2	Variación con cambio de sentido	rs35586704	-	MET_03928	AT	Leu/His	cTc/cAc	Probable efecto funcional (0.968)	Deletérea (0.01)
2:43547672	THADA	Variación con cambio de sentido	rs17334247	0.01	MET_02054	CT	Pro/Ser	Cca/Tca	Posible efecto funcional (0.842)	Deletérea (0.01)
2:163139085	IFIH1	Variación con cambio de sentido/ Variación en la región de splicing	rs140977021	-	MET_04575	AT	Val/Glu	gTa/gAa	Probable efecto funcional (1)	Deletérea (0)
7:15599863	AGMO	Variación con cambio de sentido	rs146766067	0.0013	MET_00054	CG	Glu/Gln	Gaa/Caa	Probable efecto funcional (1)	Deletérea(0)
7:127253572	PAX4	Variación con cambio de sentido	-	-	MET_00032	CG	Gln/Glu	Cag/Gag	Probable efecto funcional (0.985)	Deletérea (0.01)
10:6067847	IL2RA	Variación con cambio de sentido	-	-	MET_00323	CG	Gly/Ala	gGa/gCa	Posible efecto funcional (0.897)	Deletérea (0.01)
10:12811686	CAMK1D	Variación con cambio de sentido	-	-	MET_00138	GT	Leu/Phe	ttG/ttT	Probable efecto funcional (0.999)	Deletérea (0)
12:10271087	CLEC7A	Ganancia de stop	rs16910526	0.03	MET_02702 MET_00341 MET_00159	GT GT GT	Tyr/*	taT/taG	-	-
12:71971766	LGR5	Variación con cambio de sentido	rs138215571	0.004	MET_04575	CG	Leu/Val	Cta/Gta	Probable efecto funcional (0.999)	Deletérea (0)
12:121458600	OASL	Variación con cambio de sentido	-	-	MET_00032	GT	Phe/Val	Ttc/Gtc	Probable efecto funcional (0.971)	Deletérea (0.02)
19:7119593	INSR	Variación con cambio de sentido/ Variación en la región de splicing	rs201979105	-	MET_00072 MET_03924	GT GT	Ser/Ala	Tcc/Gcc	Probable efecto funcional (0.999)	Deletérea (0)

Tabla 18: Variaciones seleccionadas en casos

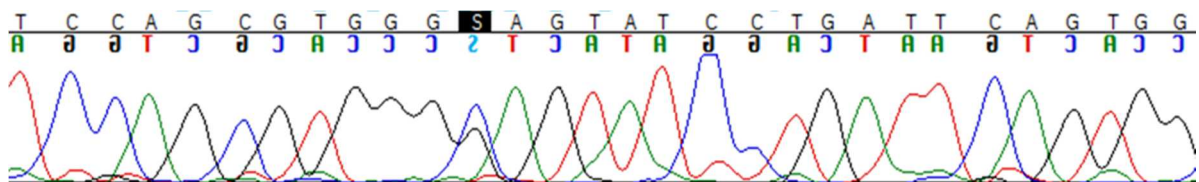
6- VALIDACIÓN DE RESULTADOS POR SECUENCIACIÓN CAPILAR. COMPROBACIÓN DE GENOTIPOS SANGER-ILLUMINA.

De un total de 20 variaciones seleccionadas para ser validadas, se han encontrado 4 falsos positivos, lo cual representa un error del 20% en Illumina, bajo las condiciones de este estudio y debido al reducido número de variaciones a validar. Por tanto definiendo precisión como (verdaderos positivos/(verdaderos positivos + falsos positivos)), la precisión del método ha sido del 80%.

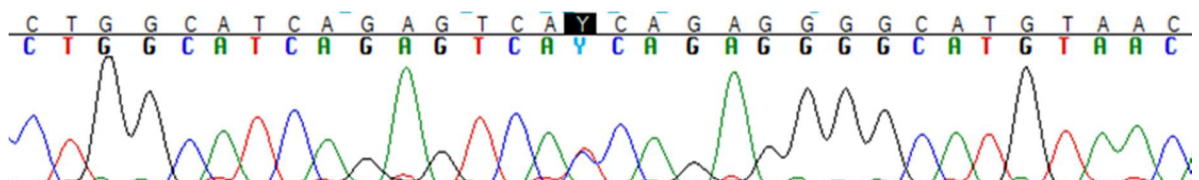
A continuación se muestran algunas de las imágenes de los cromatogramas obtenidos por secuenciación capilar: en la figura 54 se muestran 3 variantes que han resultado ser verdaderos positivos y se han podido verificar y, a continuación, en la figura 55, 2 variantes que han resultado ser falsos positivos.



A: LGR5_rs138215571 en la muestra MET_04575.

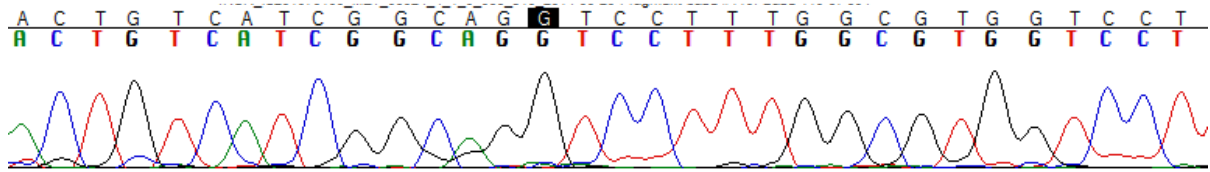


B: PAX4_7:127253572 en la muestra MET_00032.

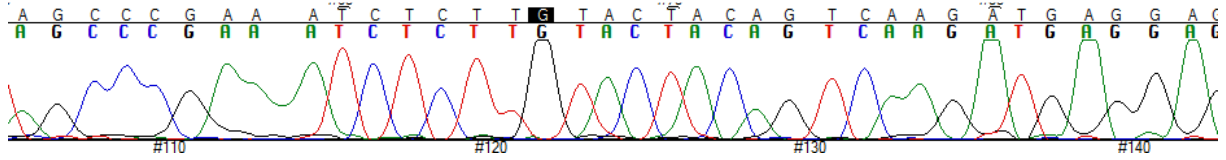


C: ERBB3_rs77822103 en la muestra 170056.

Figura 54: Ejemplos de variantes positivas.



A: INSR_rs201979105 en la muestra MET_03924.



B: CAMK1D_10:12811686 en la muestra MET_00138.

Figura 55: Ejemplos de falsos positivos.

6.1- Validación variaciones seleccionadas en los controles.

Tras la validación mediante secuenciación capilar en los controles, se pudo verificar que todas las variaciones detectadas eran verdaderos positivos, tal y como se muestra en la tabla 19:

Posición cromosómica	Gen	Consecuencia funcional	rs	MAF europea	Muestra	Genotipo Illumina	Genotipo Sanger	Cambio de Aa	Codón
2:27745372	GCKR	Ganancia de stop	rs146053779	0.0040	170056	CT	CT	Arg/*	Cga/Tga
2:43818077	THADA	Variación con cambio de sentido	rs10210191	-	170073	AG	AG	Glu/Gly	gAg/gGg
2:160994197	ITGB6	Variación con cambio de sentido	rs55841905	0.01	280152	AG	AG	Gly/Arg	Ggg/Agg
2:163144780	IFIH1	Variación con cambio de sentido	-	-	170056	CG	CG	Leu/Phe	ttG/ttC
10:114925406	TCF7L2	Variación con cambio de sentido	rs77673441	-	170067	CG	CG	Pro/Arg	cCc/cGc
11:17436078	ABCC8	Variación con cambio de sentido	-	-	170055	AT	AT	Ser/Cys	Agt/Tgt
12:56494932	ERBB3	Variación con cambio de sentido	rs77822103	-	170056	CT	CT	Ser/Pro	Tca/Cca
12:66308656	HMGA2	Variación con cambio de sentido	rs139459932	0.04	290780	CT	CT	Cys/Arg	Tgt/Cgt

Tabla 19: Validación variaciones seleccionadas en los controles.

6.2- Validación variaciones seleccionadas en los casos.

Tras la validación mediante secuenciación capilar en los casos, se detectaron 4 falsos positivos, mientras que las 8 variaciones restantes sí que pudieron ser confirmadas, como se puede observar en la tabla 20:

Posición cromosómica	Gen	Consecuencia funcional	rs	MAF europea	Muestra	Genotipo Illumina	Genotipo Sanger	Cambio de Aa	Codón
1:114397536	PTPN22	Variación con cambio de sentido	-	-	MET_00035	AC	CC	His/Asn	Cac/Aac
1:120458122	NOTCH2	Variación con cambio de sentido	rs35586704	-	MET_03928	AT	AT	Leu/His	cTc/cAc
2:43547672	THADA	Variación con cambio de sentido	rs17334247	0.01	MET_02054	CT	CT	Pro/Ser	Cca/Tca
2:163139085	IFIH1	Variación con cambio de sentido/ Variación en la región de splicing	rs140977021	-	MET_04575	AT	AT	Val/Glu	gTa/gAa
7:15599863	AGMO	Variación con cambio de sentido	rs146766067	0.0013	MET_00054	CG	CG	Glu/Gln	Gaa/Caa
7:127253572	PAX4	Variación con cambio de sentido	-	-	MET_00032	CG	CG	Gln/Glu	Cag/Gag
10:6067847	IL2RA	Variación con cambio de sentido	-	-	MET_00323	CG	GG	Gly/Ala	gGa/gCa
10:12811686	CAMK1D	Variación con cambio de sentido	-	-	MET_00138	GT	GG	Leu/Phe	ttG/ttT
12:10271087	CLEC7A	Ganancia de stop	rs16910526	0.03	MET_02702 MET_00341 MET_00159	GT GT GT	CT CT CT	Tyr/*	taT/taG
12:71971766	LGR5	Variación con cambio de sentido	rs138215571	0.004	MET_04575	CG	CG	Leu/Val	Cta/Gta
12:121458600	OASL	Variación con cambio de sentido	-	-	MET_00032	GT	GT	Phe/Val	Ttc/Gtc
19:7119593	INSR	Variación con cambio de sentido/ Variación en la región de splicing	rs201979105	-	MET_00072 MET_03924	GT GT	GG GG	Ser/Ala	Tcc/Gcc

Tabla 20: Validación variaciones seleccionadas en los casos.

Discusión

DISCUSIÓN

La DM2 es una enfermedad compleja que resulta de la interacción de factores genéticos de predisposición y factores ambientales y adquiridos. Los genes involucrados suelen estar relacionados con alteraciones metabólicas (como la resistencia a la insulina) o con la disfunción de las células β principalmente. El efecto aditivo de los diferentes loci genéticos implicados y los factores ambientales, tales como la obesidad, modulan el umbral para que se desarrolle o no la enfermedad (Quintanilla-García et al, 2010). Diversos estudios han mostrado la importante heredabilidad de la DM2, estableciendo que ésta es posiblemente superior al 70% (Medici et al, 1999; Poulsen et al, 1999).

Como ya se ha indicado previamente, se han realizado numerosos estudios genéticos para conocer los genes asociados a la DM2 utilizando diferentes estrategias: estudios de genes candidatos, ligamiento en familias, GWAS. Sin embargo, el conjunto de loci y de variantes genéticas identificadas hasta el momento, explican entre el 10 y el 15% de la heredabilidad de la enfermedad (Billings et al, 2010; Zeggini et al, 2008). Por lo que, con el objetivo de encontrar esta heredabilidad perdida, se desarrolla la hipótesis “enfermedades comunes, variantes raras” que plantea la búsqueda de estas variantes en el exoma. Para identificar estas variantes y parte del componente genético de la diabetes no identificado hasta el momento, en este estudio, se han secuenciado 178 exomas y se han analizado los genes que previamente se habían relacionado con diferentes tipos de diabetes.

Tras la secuenciación del exoma y posterior análisis bioinformático de las 178 muestras de casos y controles, se han identificado un gran número de variaciones (Veáse la tabla 13 del apartado resultados). Entre éstas se han seleccionado para este estudio las consideradas con mayor efecto funcional:

- variaciones con cambio de sentido.
- variaciones de pérdida o ganancia de codón de parada.
- variaciones que afectan a las regiones de *splicing*.

Discusión

Continuando con la selección de las variaciones con un posible efecto y con la idea de identificar aquellas que puedan estar implicadas en la DM2, se han seleccionado:

- Las variantes que afectan a la funcionalidad de la proteína, según PolyPhen, siendo este resultado corroborado por SIFT, que lo expresa como que la variación es deletérea. Se obtuvieron aproximadamente 80 variaciones de este tipo y de entre éstas, se han seleccionado para ser estudiadas y validadas, las que tienen un MAF<20% en la población europea del proyecto 1000 Genomas.
- Por otro lado, se han excluido las que estaban presentes en casos y controles simultáneamente.

Finalmente, 16 variaciones han sido identificadas y validadas por su posible relación con el desarrollo de la DM2. Se recogen en la tabla 21:

VARIACIONES QUE SE ENCUENTRAN SOLO EN CONTROLES		
rs	Posición cromosómica	GEN
rs146053779	2:27745372	GCKR
rs10210191	2:43818077	THADA
rs55841905	2:160994197	ITGB6
-	2:163144780	IFIH1
rs77673441	10:114925406	TCF7L2
-	11:17436078	ABCC8
rs77822103	12:56494932	ERBB3
rs139459932	12:66308656	HMGA2

VARIACIONES QUE SE ENCUENTRAN SOLO EN CASOS		
rs	Posición cromosómica	GEN
rs35586704	1:120458122	NOTCH2
rs17334247	2:43547672	THADA
rs140977021	2:163139085	IFIH1
rs146766067	7:15599863	AGMO
-	7:127253572	PAX4
rs16910526	12:10271087	CLEC7A
rs138215571	12:71971766	LGR5
-	12:121458600	OASL

Tabla 21: Variaciones identificadas y validadas en casos y controles

1-CARACTERÍSTICAS DE LA SELECCIÓN DE MUESTRAS

Dado que el objetivo de este estudio era encontrar variaciones poco frecuentes en un número de muestras limitado, la selección de muestras se realizó buscando en ellas unas características determinadas para poder identificar el mayor número de variaciones posible. Se centró en dos características, IMC y edad:

- El desarrollo de obesidad aumenta fuertemente el riesgo de padecer DM2. Los pacientes con obesidad Clase II o superior (según la clasificación SEEDO-2007) presentan una exposición a los efectos de la obesidad muy elevada, por lo que el componente genético podría verse desvirtuado. La inclusión de pacientes con un IMC bajo también podría conducir a la detección de pacientes con algún tipo de alteración, lo cual nos alejaría del objetivo planteado. Por ello, se deben seleccionar aquellos con la mayor carga genética posible dentro del objetivo planteado, es decir pacientes DM2 con IMC entre 26 y 32 (Lyssenko et al, 2010).
- Otro factor muy importante es la edad, ya que con ella, la exposición a factores de riesgo se incrementa, esto supone que aquellos pacientes que desarrollan DM2 a mayores edades podrían tener un menor componente genético.

Por tanto, el hecho de seleccionar pacientes medianamente jóvenes (45-65 años) y con preobesidad u obesidad tipo I, ha permitido tener una muestra con un componente genético predominante y poco desvirtuado, en la que se ha podido encontrar mayor número de variaciones.

2-ESTUDIO DE VARIANTES RARAS

Tanto estudios de segregación como estudios en gemelos han sugerido que las enfermedades complejas, se determinan en gran parte, por factores genéticos (King et al, 2002). Una de las enfermedades comunes que ha sido sometida a un mayor estudio

Discusión

genético es la diabetes tipo 2 (Permutt et al, 2005). La heredabilidad de ésta se ha estimado en alrededor del 70% (Kaprio et al, 1992; Poulsen et al, 1999; Almgren et al, 2011). A través de los GWAS llevados a cabo, 62 loci han sido asociados con esta enfermedad. Sin embargo, los SNPs asociados sólo pueden dar cuenta de menos del 10-15 % de la heredabilidad estimada a partir de estudios familiares (Morris et al, 2012).

En cuanto a la relación de los polimorfismos poco frecuentes con enfermedades complejas, tales como la DM2, un estudio reciente (Steinthorsdottir et al, 2014) ha demostrado la asociación de algunas variantes raras con diabetes tipo 2, poniendo así de manifiesto la implicación e importancia de este tipo de variantes en estas enfermedades. Se descubrieron 4 nuevas variantes que afectan el riesgo de diabetes tipo 2. Una variante de baja frecuencia (1.47%) en el intrón 1 del gen CCND2, rs76895963 [G], reduce el riesgo de diabetes tipo 2 a la mitad (odds ratio (OR) = 0.53, $P = 5.0 \times 10^{-21}$). Además, dos variantes de cambio de sentido en el gen PAM, p.Asp563Gly (frecuencia de 4.98%) y p.Ser539Trp (frecuencia de 0.65%), confieren un riesgo mayor de padecer la enfermedad (OR = 1.23, $P = 3.9 \times 10^{-10}$) y OR = 1.47, $P = 1.7 \times 10^{-5}$, respectivamente). Y, una variante en el gen PDX1 (0.20%) p.Gly218Ala, se asocia con un alto riesgo de diabetes tipo 2 (OR = 2.27, $P = 7.3 \times 10^{-7}$).

Por tanto, las razones principales por las que en este proyecto se decidió estudiar el papel de las variantes raras o polimorfismos poco frecuentes en la diabetes tipo 2 fueron:

- Los estudios GWAS realizados hasta hace unos años, solo permiten identificar la asociación de polimorfismos frecuentes ($MAF > 0.05$). Sin embargo, estos pueden tener un efecto pequeño a nivel funcional y sobre el fenotipo (OR entre 1.1 y 1.5) (Bodmer y Bonilla, 2008; So et al, 2011).
- La base genética de una gran parte de la heredabilidad calculada no se ha identificado.
- Los GWAS utilizados hasta el momento se han realizado con chips que dejan muchas regiones genómicas sin una buena cobertura y se centran en polimorfismos frecuentes.

- Además, el desarrollo de las tecnologías de nueva generación, está haciendo posible llevar a cabo proyectos para secuenciar mayor número de muestras y analizar así variantes tanto frecuentes como poco frecuentes.
- Los polimorfismos poco frecuentes y con efecto funcional podrían explicar parte de la heredabilidad no encontrada.

Uno de los principales objetivos de esta tesis era analizar el papel de las variantes raras en una enfermedad compleja como la diabetes mellitus tipo 2. Finalmente se han verificado 16 variantes raras que tienen, *a priori* un efecto funcional según los programas de predicción empleados.

3-SECUENCIACIÓN DE NUEVA GENERACIÓN Y ESTUDIO DEL EXOMA

El método seleccionado para realizar el estudio ha sido la secuenciación de exoma mediante la tecnología Illumina. Se ha empleado el kit de enriquecimiento en exoma TruSeq de Illumina. En un estudio muy reciente, (Chilamakuri et al, 2014) se evaluó el rendimiento de cuatro tecnologías de captura de exoma, y se observó que todas las tecnologías de captura del exoma funcionan bien, pero existen algunas diferencias consistentes entre ellas. Illumina cubre una mayor parte de las bases del exón según las bases de datos, seguida por NimbleGen y Agilent. Todas las tecnologías presentan una alta cobertura de las diferentes regiones, con la tecnología Agilent dando lugar a la cobertura más alta (99,8%) seguido de Nextera (98,2%), Truseq (96,9%), y NimbleGen (96,5%).

El genoma humano se compone de alrededor de 180.000 exones, lo que se traduce en 60 megabases (Mb) de longitud que representan el 1-2% del genoma humano. Se estima que las regiones codificantes del genoma o exoma, albergan aproximadamente el 90% del efecto de las mutaciones causantes de enfermedad (Choi et al., 2009), lo que convierte el estudio del exoma en una herramienta práctica y prometedora para el estudio de enfermedades genéticas.

Discusión

La secuenciación del exoma ya ha demostrado su utilidad en la investigación de las causas de diferentes enfermedades genéticas. En el estudio de Choi et al., (2009) se muestra por primera vez el valor de la secuenciación del exoma en este sentido, concretamente en el diagnóstico genético de la diarrea de cloruro congénita en un pequeño grupo de pacientes con sospecha de este síndrome. Se llevó a cabo la secuenciación del exoma en 6 pacientes que carecían de mutaciones en los genes conocidos para el síndrome de Bartter analizados mediante secuenciación estándar. Los resultados revelaron mutaciones homocigotas y deletéreas en el locus SLC26A3 de los 6 pacientes, lo que permitió realizar un diagnóstico molecular de la enfermedad que fue confirmado posteriormente en la evaluación clínica.

La secuenciación de exoma también está ayudando a mejorar la comprensión de las bases genéticas de otros tipos de enfermedades tales como el cáncer. En el estudio de Jiao y colaboradores (2011) se utilizó la secuenciación del exoma para explorar las bases genéticas de 10 tumores neuroendocrinos pancreáticos. Los resultados identificaron mutaciones frecuentes en MEN1 (44% tumores), DAX / ATRX (43%), y genes de la vía mTOR (14%). Curiosamente, mutaciones en los genes MEN1 y el DAX / ATRX mostraron potencial clínico, ya que su presencia se asocia con un mejor pronóstico. Además, la implicación de mutaciones en la vía de mTOR también es de interés clínico, dada la disponibilidad de agentes terapéuticos que inhiben esta vía.

En cuanto a la secuenciación de exoma en relación con la diabetes tipo 2, existen multitud de estudios en que se han encontrado asociaciones. En el estudio de Huang et al (2014) se secuenció el exoma de 177 indios Pima y, tras genotipar las variantes seleccionadas en un mayor número de individuos, se observó que variaciones en los genes CYB5A y RNF10, aumentaban el riesgo de DM2 en la población estudiada. En otro estudio muy reciente (Moltke et al, 2014), utilizando el genotipado mediante arrays y la secuenciación de exoma, se identificó una variante sin sentido (p.Arg684Ter) en el gen TBC1D4. Homocigotos portadores de esta variante tienen mayores concentraciones de glucosa en plasma y de insulina en suero, 2 horas después de una sobrecarga oral de glucosa, en comparación con individuos con otros genotipos.

Sin embargo, existen estudios de secuenciación de exoma en que no se han encontrado asociaciones con diabetes tipo 2. En el estudio de Lohmueller et al (2013) se planteaba la hipótesis de que las variantes raras en las regiones codificantes explican una parte sustancial de la heredabilidad de las enfermedades comunes. Se secuenciaron los exomas de 1.000 casos con diabetes tipo 2 y 1.000 controles sanos y se aplicó una serie de pruebas basadas para detectar los genes de susceptibilidad a la DM2. Sin embargo, ningún gen mostró una asociación significativa con el riesgo de la enfermedad. Otro ejemplo es el estudio de Li et al (2014), en el cual se obtuvo la secuencia del ADN mitocondrial a partir de los datos provenientes de un conjunto de 2000 exomas (1.000 individuos con diabetes tipo 2 y 1.000 controles). No se encontraron asociaciones significativas con DM2.

4-ANÁLISIS BIOINFORMÁTICO Y COBERTURA EMPLEADA

En el desarrollo de las tecnologías de secuenciación de nueva generación, el análisis bioinformático de los datos adquiere una gran importancia debido a la gran cantidad de datos generados y los pasos a seguir para llegar a seleccionar y obtener los resultados deseados. El objetivo de esta tesis era identificar nuevos polimorfismos y estudiar los polimorfismos poco frecuentes, sin embargo no era una prioridad el llevar a cabo el genotipado de las muestras. Por ello se ha optado por secuenciar un mayor número de muestras con una cobertura menor, dado que esta estrategia nos proporcionaría resultados de mayor interés (Kim et al, 2011). Es decir, con el fin de secuenciar más muestras, suele secuenciarse cada muestra con una cobertura menor. Por lo tanto, cada base sólo será cubierta por unas pocas lecturas, por lo que es más difícil inferir con precisión el genotipo de un individuo en un sitio en particular (Harismendy et al, 2009), sin embargo esto ha permitido encontrar un mayor número de variantes. De esta forma, la tasa de error puede verse aumentada, pero por otros métodos puede verificarse fácilmente si los polimorfismos son reales o no.

En el estudio de Clark et al (2011) se describe que, partiendo de la secuenciación de 80 millones de lecturas por individuo a una cobertura media de 48x con Illumina, el 97,1% de las bases estaban cubiertas por lo menos una vez, y el 90,0% en $\geq 10 \times$. Y un 35,6% de

Discusión

las lecturas se encuentran en regiones no exónicas, por tanto la máxima cobertura que se podría obtener sería de un 64,4%.

Estudios recientes muestran resultados similares a los que se han obtenido, por ejemplo el de Lohmueller et al, (2013), en que se secuenció el exoma de 2000 individuos. La cobertura media fue 56.3x. La cobertura media del exoma (porcentaje de bases que fueron cubiertos por al menos una lectura) fue de 97,27% por muestra. Ninguna muestra tenía menos del 90% de su exoma cubierto.

En nuestro estudio, como se ha mostrado en el apartado de resultados, se ha obtenido una cobertura media aproximada de 20x, que era la cobertura deseada para maximizar la diversidad muestral con los recursos disponibles. Además se obtuvo aproximadamente un 60% de bases cubiertas como mínimo a 10x y, un 90% a 1x.

5-LIMITACIONES DEL ESTUDIO

Como en todos los estudios, en el desarrollo de éste, hay ciertas limitaciones. No obstante, a pesar de éstas, se han alcanzado los objetivos planteados.

En cuanto al estudio de polimorfismos poco frecuentes:

- 1) la identificación de los PPFs en la población no ha sido completa principalmente debido al número limitado de muestras secuenciadas. Si bien, el número de muestras incluido en el proyecto permitiría identificar la mayoría de los PFs y más del 80% PPFs presentes en el exoma (con MAF entre 0.05 y 0.01) y asociados a la DM2. En el caso de las variantes raras (MAF< 0.001) la probabilidad es más baja (Li B et al, 2009).
- 2) El poder estadístico para detectar asociaciones viene limitado por la baja frecuencia de los PPFs. Los cálculos teóricos utilizados hasta el momento predicen que se podrán encontrar asociaciones con tamaños de muestra accesibles (2000 casos/controles) cuando la OR sea superior a valores de 2.0

(Gorlov et al, 2008). Por otro lado, se están desarrollando métodos estadísticos para mejorar su análisis individual o su interacción (Li B et al, 2008; Sul et al, 2011; Neale et al, 2011).

- 3) El número de PPFs encontrados ha sido muy elevado y no es posible abarcar un estudio completo de los mismos (Ng et al, 2008 y 2009) en estos momentos por cuestiones económicas. Por esta razón, se ha hecho una selección de variantes; aunque esto podría ocasionar que hayan quedado fuera polimorfismos realmente asociados a DM2.

En la continuación de este estudio se contará con más de 10.500 muestras de población general, 1000 casos y 2000 controles. Por lo que, si consideramos que muchos PPFs tendrán ORs superiores a 3 y $MAF > 0.01$, el poder estadístico para el análisis de la mayoría de los polimorfismos está garantizado.

Otra limitación es que el estudio se ha centrado en el análisis de las regiones codificantes del genoma. Sin embargo, existen otras regiones del genoma de gran interés y cuyo análisis sería importante, pero al ser los resultados esperables menos prometedores y el coste de su estudio mucho más elevado, el método utilizado ha sido la secuenciación del exoma frente a la del genoma completo.

En cuanto a las limitaciones del kit empleado para el enriquecimiento, éste posee una eficiencia de captura de las regiones exónicas de alrededor del 65%, pero se trata de limitaciones técnicas y por tanto, inevitables. Además, debido a los filtros aplicados en el análisis bioinformático de las secuencias, algunas variaciones no han sido detectadas porque no se han obtenido suficientes lecturas durante el proceso de secuenciación o por la baja calidad las secuencias. Por otra parte, también es posible que algunas variaciones no hayan sido detectadas (falsos negativos), debido a los filtros utilizados para reducir al máximo los falsos positivos.

Las variantes raras encontradas no han podido ser replicadas en poblaciones más grandes ni se han llevado a cabo estudios funcionales, si bien estos estudios se realizarán en un futuro próximo. Otra limitación sería la imposibilidad de calcular correctamente

Discusión

las MAF en nuestros genotipos (aunque sí que podría hacerse una aproximación), ya que el estudio trata de descubrir nuevos SNPs y no de genotiparlos, por lo que se ha buscado tener menor cobertura por muestra a cambio de secuenciar un mayor número de muestras o determinar su presencia en muestras más amplia mediante otras técnicas.

6-TIPOS DE VARIANTES IDENTIFICADAS Y VERIFICADAS

Los SNPs se encuentran a lo largo de todo el genoma, en los exones, intrones, regiones intergénicas, promotores, etc. Por lo tanto, en este sentido, es de gran interés el efecto que una simple sustitución de pares de bases puede tener en un rasgo o enfermedad, ya que muchas veces esto va a depender de la posición en que se encuentre el SNP. Un SNP en la región codificante de un gen puede afectar directamente a la proteína codificada, la cual quizá posea una función relevante; un SNP también puede influir en una región exónica o intrónica que afecte al *splicing* (Krawezak et al, 1992); un SNP en un promotor puede influir en la expresión génica (Drazen et al, 1999), etc. El grado en que cada SNP influye en el fenotipo está recibiendo una gran cantidad de atención a medida que se va identificando un mayor número de SNPs.

Esta tesis se ha centrado en 3 tipos principales de variaciones por su importancia funcional, aunque en un futuro se pretende estudiar también el resto:

- 1- Variaciones que afectan al proceso de *splicing*:
 - **Variación donante de *splicing***: variante de *splicing* que cambia la región del extremo 5' de un intrón.
 - **Variación aceptora de *splicing***: variante de *splicing* que cambia la región del extremo 3' de un intrón.
 - **Variación en la región de *splicing***: variante en la secuencia en la que se ha producido un cambio dentro de la región del sitio de *splicing*, ya sea dentro de 13 bases del exón o 38 bases del intrón.
- 2- Variaciones de stop:

- **Ganancia de codón de parada:** variante en la secuencia que hace que al menos una base de un codón cambie, dando lugar a un codón de stop prematuro y, por tanto a un péptido más corto.
- **Pérdida de codón de parada:** variante en la secuencia que hace que al menos una base de un codón cambie, dando lugar a un péptido más largo.

3- Variaciones con cambio de sentido:

Variantes en la secuencia que dan lugar al cambio de una o más bases, resultando en una secuencia de aminoácidos diferente, pero con la misma longitud.

7-VARIACIONES IDENTIFICADAS EN LOS GENES ESTUDIADOS

Los genes estudiados han sido genes implicados en diferentes tipos de diabetes o en alteraciones del metabolismo glucídico. Su selección se hizo mediante revisión bibliográfica. En ellos se han encontrado variantes no descritas previamente y variantes que, pese a haber sido ya estudiadas, no se habían relacionado con diabetes tipo 2. En todas las variantes encontradas, pese a que, según SIFT y PolyPhen, afectan a la funcionalidad de la proteína, sería necesario realizar estudios funcionales.

7.1-VARIACIONES IDENTIFICADAS Y VALIDADAS EN CONTROLES

Los alelos salvajes o wild type codifican un producto necesario para una función biológica específica. Por tanto, si se produce una mutación en un alelo, la función se altera: esto puede conducir a una pérdida o una ganancia de la función. Si bien, por el tipo de mutaciones seleccionadas en el presente estudio, la mayoría serán de pérdida de función.

Cuando se da una pérdida, el término general para denominar estas mutaciones es el de mutaciones de pérdida de función (“loss of function mutations”). El grado en que la función se pierde puede variar. Si ésta se ha perdido por completo, la mutación se llama una mutación nula. También es posible que parte de ésta permanezca, pero no al nivel del alelo salvaje.

Discusión

Por tanto, las variaciones que se han encontrado en controles, serán variaciones que alteran la funcionalidad de la proteína codificada por un gen determinado, confiriendo así una función protectora frente a la enfermedad. Un ejemplo concreto de este tipo de variaciones se encuentra en el estudio de Flannick et al (2014), en que se describen mutaciones de pérdida de función en el gen SLC30A8 que protegen contra la diabetes tipo 2.

Gen GCKR (“glucokinase (hexokinase 4) regulator”):

Según la secuencia de su ARNm NM_001486.3 se ha identificado la mutación c.1618C>T que en la secuencia de la proteína (NP_001477.2) corresponde con la mutación p.Arg540Ter. Es decir, se ha encontrado un cambio de citosina por timina en la posición 1618 del ARN mensajero que supone un cambio de la arginina 540 por un codón de parada de la traducción de un total de 625 aminoácidos. Se representa en la figura 56. No existen estudios sobre esta variación.

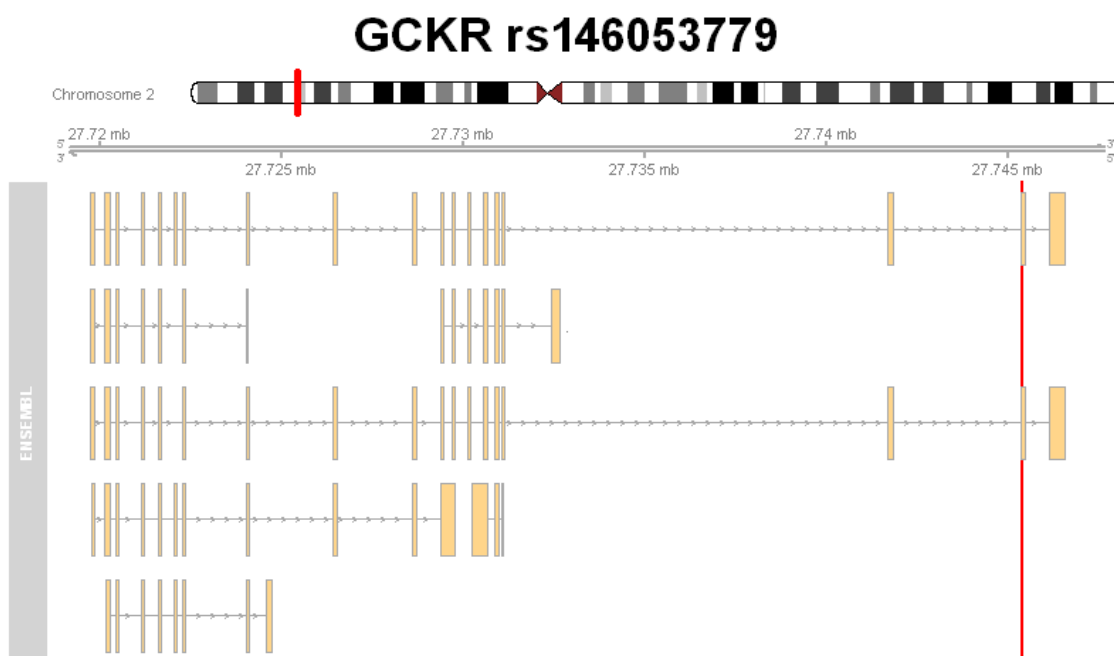


Figura 56: Representación de la variación rs146053779, marcada con la línea vertical roja, en los diferentes transcritos del gen GCKR.

Patogenicidad de la mutación.

En este caso el resultado de la mutación del gen es una proteína truncada, por lo que no se puede calcular la puntuación SIFT ni PolyPhen. Sin embargo, dada la importante alteración que esto supone en la proteína, será muy probablemente patogénica. No obstante, la proteína podría ser parcialmente funcional debido a que la mutación se encuentra en el penúltimo exón del gen, pero serían necesarios estudios funcionales para confirmarlo.

Función del gen.

El gen GCKR está situado en el cromosoma 2p23.2-3 (Hayward et al., 1996), una región genómica vinculada a rasgos metabólicos como la grasa corporal o las concentraciones de lípidos circulantes (Comuzzie et al., 1997). El ADNc de GCKR codifica para una proteína de 625 aminoácidos. En cuanto a su estructura, este gen contiene 19 exones y abarca 27 kb.

La glucoquinasa (GCK) es uno de los cuatro isoenzimas de la hexoquinasa que catalizan la fosforilación de la glucosa. Esta enzima está sujeta a inhibición, tanto en las células β pancreáticas como en las hepáticas, por medio de la “glucokinase regulatory protein” (GKRP), codificada por el gen GCKR. Según diferentes estudios, la proteína GKRP se encuentra fuertemente asociada con la hipertrigliceridemia (Saxena et al, 2007; Scott et al, 2007).

La unión de estas proteínas da como resultado la inhibición competitiva con respecto a la glucosa (Vandercammen et al, 1990). La constante de afinidad para GKRP de hígado de rata se reduce aproximadamente 15 veces con la presencia de fructosa 6-fosfato, y sus efectos inhibidores pueden ser antagonizados por la presencia de fructosa 1-fosfato (F1P). Hepatocitos de rata tratados con precursores de F1P conducen a aumento de la actividad GCK, presumiblemente por la unión de F1P a GKRP, lo que induce un cambio conformacional en GKRP que conduce a ésta a tener una menor afinidad por GCK (Vandercammen et al, 1989; Van Schaftingen, 1989; Van Schaftingen et al, 1994).

Discusión

Se ha demostrado que GKR_P es necesaria para la translocación de GCK al núcleo de los hepatocitos (De la Iglesia et al, 1999; Shiota et al, 1999). Después de unirse a GKR_P, GCK es retenida en el núcleo hasta que es liberada debido a cambios en el medio metabólico. Esta localización celular de GCK parece jugar un papel importante en la activación de la síntesis de glucógeno, que se produce tras la translocación de GCK al citosol (Agius et al, 1996; Agius, 1998), pero no se sabe si el ciclo de translocación de GCK hepática en el que está implicada GKR_P tiene otros papeles fisiológicos. Además, estudios con transgénicos han demostrado que la sobreexpresión de GCK en el hígado está asociada con un aumento concomitante de la utilización de glucosa y con una reducción de los niveles de glucosa en sangre (Niswender et al, 1997), mientras que los animales *knockout* de GCK hepática son ligeramente hiperglucémicos (Postic et al, 1999).

En el páncreas, GCK actúa como un sensor molecular para la homeostasis de la glucosa mediante el control del nivel umbral de glucosa estimulado por la liberación de insulina en las células β (Matschinsky et al, 1998). La unión de la glucosa a la GCK en los islotes se correlaciona positivamente con la liberación de insulina estimulada por glucosa (Liang et al, 1992). Ratones (Bali et al, 1995; Grupe et al, 1995; Postic et al, 1999; Terauchi et al, 1995) y humanos (Bell et al, 1996) con deficiencia de GCK poseen concentraciones elevadas de glucosa en la sangre. La enzima ha sido identificada en el hígado, en β -células pancreáticas (Iynedijan et al, 1986), en áreas discretas del sistema nervioso central, y en los enterocitos (Jetton et al, 1994).

Entre varios polimorfismos identificados en GCKR, el polimorfismo rs780094 ha sido ampliamente estudiado por su asociación con la susceptibilidad a la DM2. El alelo G de este polimorfismo también se ha asociado con mayores niveles de glucosa plasmática en ayunas. Mientras que el alelo T se asoció con menores niveles de glucosa, menor resistencia a la insulina y menor riesgo de DM2, pero mayores niveles de triglicéridos. Esto normalmente ocurre al contrario en humanos, pero coincide con lo que ocurre en estudios de sobreexpresión de GCKR y GCK en modelos de ratón (Saxena et al, 2007; Scott et al, 2007). Recientemente se publicó un metaanálisis (Li et al, 2013), en que se incluían 19 estudios (99.702 casos y 199.275 controles) con poblaciones de diferentes

etnias. Se demostró que el alelo G del polimorfismo rs780094 en GCKR es un factor de riesgo para desarrollar diabetes tipo 2.

En el estudio de Beer y colaboradores (2009), una variante de GCKR (P446L) se asocia con los niveles plasmáticos de triglicéridos y glucosa en ayunas en población general. El objetivo de este estudio fue determinar el mecanismo responsable de esta asociación genética. La alteración de la regulación de GCK en el hígado parece promover el flujo glucolítico, el metabolismo hepático de la glucosa y la elevación de las concentraciones de malonil-CoA, un sustrato para la lipogénesis de novo, proporcionando así un mecanismo mutacional de la asociación de esta variante con niveles elevados de triglicéridos y con niveles menores de glucosa.

En otro estudio (Tanaka et al, 2011) se reclutaron familias japonesas con una historia de 3 generaciones de diabetes. De todos los genes que se incluyeron en el estudio, la mutación GCKR g.6859C>G fue asociada. Se concluyó que GCKR es un gen implicado en la susceptibilidad a la diabetes en las familias japonesas.

La variación encontrada en este estudio, rs146053779 (gen GCKR), había sido descrita, pero no existen estudios sobre ella. Al encontrarse en controles, podría tener un papel protector, debido a que daría lugar a una proteína GGRP menos activa y esto conduciría a un aumento moderado en la actividad de GCK impidiendo así alcanzar niveles demasiado elevados de glucosa e insulina en sangre, los cuales podrían tener efectos negativos con el paso del tiempo.

Gen THADA (“thyroid adenoma associated”):

Según la secuencia de su ARNm NM_001083953.1 se ha identificado la mutación c.188A>G que en la secuencia de la proteína (NP_001077422.1) corresponde con la mutación p.Glu63Gly. Se representa en la figura 57.

No existen estudios sobre esta variación.

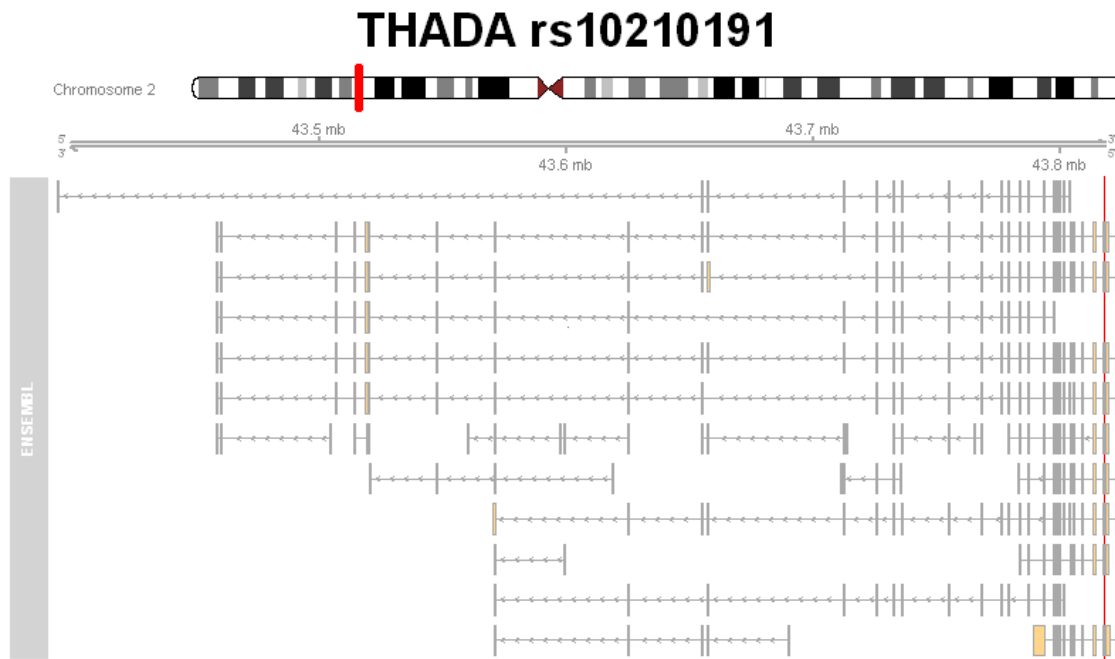


Figura 57: Representación de la variación rs10210191, marcada con la línea vertical roja, en los diferentes transcritos del gen THADA.

Patogenicidad de la mutación.

Según PolyPhen, esta mutación afectará posiblemente a la funcionalidad de la proteína resultante, con una puntuación de 0.804. Este resultado es corroborado por SIFT, que le da una puntuación de 0.01, lo cual la califica como deletérea.

Función del gen.

El gen THADA se encuentra en el cromosoma 2p21, tiene 38 exones y se extiende unas 365 kb. Se asocia a adenomas tiroideos, ya que en aproximadamente el 10% de los tumores benignos de la tiroides se encuentran reordenamientos en el cromosoma 2p21.

SNPs cercanos al gen THADA se han asociado con la diabetes tipo 2, y se ha observado que la expresión de este gen es diferente en los individuos con diabetes y los controles sanos (Parikh et al, 2009). Además, en el estudio de Simonis et al (2010) se estudia el papel de este gen en la DM2, concretamente mediante el estudio del polimorfismo rs7578597 (variación con cambio de sentido que provoca un cambio de treonina a alanina, T1187A). Se demuestra que los portadores homocigotos de los alelos de riesgo, tienen células β menos funcionales. Esto no se había observado en estudios previos,

aunque el estudio de Stancakova et al (2009) sí que mostró evidencias de una liberación de insulina reducida. Se ha sugerido que THADA, participa en el proceso de apoptosis celular (Rippe et al, 2003). Esto podría implicar que los sujetos con la variación rs7578597 tienen una masa de células β reducida debido a un aumento de la apoptosis. Pero serían necesarios más estudios para confirmar esta hipótesis.

Por tanto, si polimorfismos en el gen THADA pueden conducir a una menor o mayor funcionalidad de las células β y al aumento o disminución de su apoptosis, la variación encontrada en el presente estudio, rs10210191 (gen THADA), podría tener un papel protector mediante el aumento de actividad y la supervivencia de las células β .

Gen ITGB6 (“integrin, beta 6”):

Según la secuencia de su ARNm NM_000888.4 se ha identificado la mutación c.1408G>A que en la secuencia de la proteína (NP_000879.2) corresponde con la mutación p.Gly470Arg. Se representa en la figura 58.

No existen estudios sobre esta variación.



Figura 58: Representación de la variación rs55841905, marcada con la línea vertical roja, en los diferentes transcritos del gen ITGB6.

Discusión

Patogenicidad de la mutación.

Según PolyPhen, la mutación identificada afectará posiblemente a la funcionalidad de la proteína resultante, con una puntuación de 0.489. SIFT confirma este resultado dándole una puntuación de 0, lo cual la califica como deletérea.

Función del gen.

El gen ITGB6 codifica para la integrina beta 6. Las integrinas son receptores del tipo glicoproteína transmembrana, que median vínculos dinámicos entre el citoesqueleto de actina y la matriz extracelular y, también transducen señales hacia y desde el interior de la célula (Hynes, 1992). La expresión de la isoforma beta 6 puede afectar a la secreción de metaloproteinasas de la matriz y a la progresión de tumores (Niu et al, 2002; Li et al, 2003).

En la literatura existe poco apoyo para el papel de estas proteínas codificadas por el locus 2q24 en la etiología de la DM2. No obstante, en el estudio llevado a cabo por Qi et al (2010), este locus fue asociado con la enfermedad. El locus 2q24 incluye dos genes: RBMS1 e ITGB6. Variaciones en este locus, las cuales se asociaron con un menor riesgo diabetes tipo 2, también se relacionan con niveles más bajos de glucosa en ayunas. Estos datos sugieren que el locus 2q24 podría contener variantes que afectan al metabolismo de la glucosa y a la resistencia a insulina, pero es necesaria una mayor investigación al respecto.

Por ello, es posible que la variación encontrada, rs55841905 (gen ITGB6) también tenga un papel protector frente a la DM2 conduciendo a menores niveles de glucosa y a una menor resistencia a la insulina.

Gen IFIH1 (“interferon induced with helicase C domain 1”):

Según la secuencia de su ARNm, NM_022168.3 se ha identificado la mutación c.960G>C que en la secuencia de la proteína (NP_071451.2) corresponde con la mutación p.Leu320Phe. Se representa en la figura 59.

Esta variación no ha sido descrita previamente y por tanto no existen estudios sobre ella.

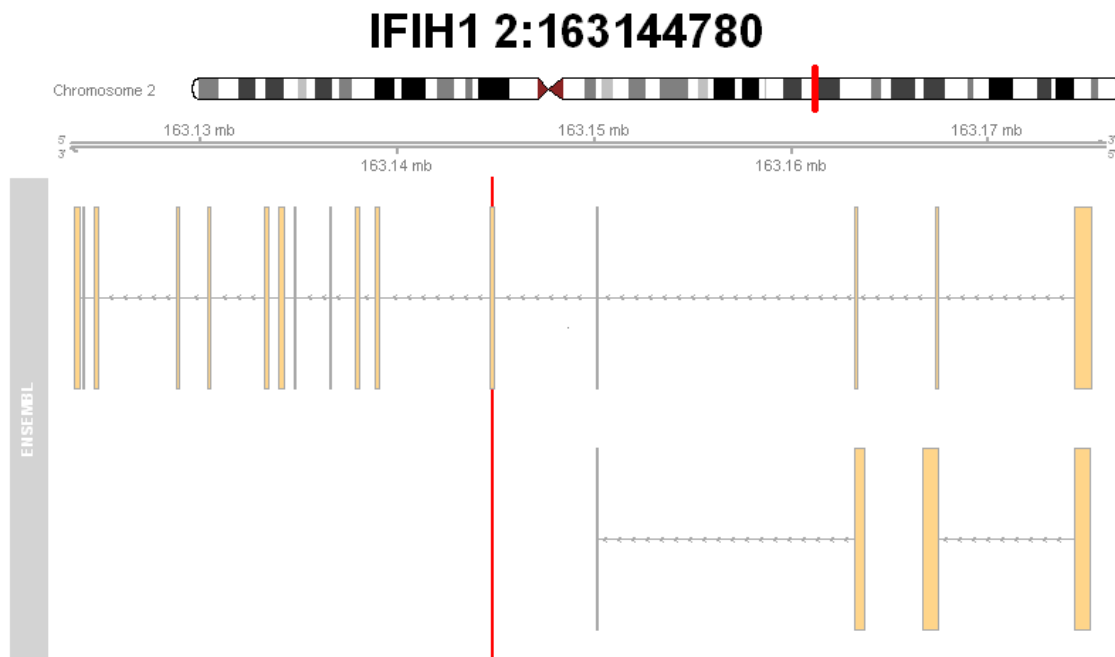


Figura 59: Representación de la variación 2:163144780, marcada con la línea vertical roja, en los diferentes transcritos del gen IFIH1.

Patogenicidad de la mutación.

Según PolyPhen, esta mutación tendrá probablemente un efecto importante sobre la actividad de la proteína, con una puntuación de 0.982. Este resultado es corroborado por SIFT, que le da una puntuación de 0, lo cual la califica como deletérea.

Función del gen.

El gen IFIH1 codifica para un receptor citoplasmático de la familia de receptores de reconocimiento de patrones (PRR) que reconoce ARN de doble cadena (dsRNA), jugando así un papel en la respuesta inmune innata desencadenada por la infección viral. La unión de dsRNA a este PRR desencadena la liberación de citoquinas proinflamatorias, tales como interferones (IFN), que poseen una potente actividad antiviral, protegiendo así las células no infectadas e induciendo la apoptosis de las células infectadas. Sin embargo, en algunos individuos susceptibles genéticamente, este sistema de defensa no funciona correctamente, induciendo una inflamación excesiva y la muerte

Discusión

prolongada de las células β , lo que predispone al desarrollo de DM1. De hecho, estudios llevados a cabo en diferentes poblaciones sugieren que variantes del gen IFIH1 se asocian con DM1 (Bouças et al, 2013). La variación no sinónima rs1990760 (Ala946Thr, exon 15, gen IFIH1) además de asociarse con DM1, también se ha asociado con otras enfermedades autoinmunes, como la enfermedad de Graves y el lupus eritematoso sistémico. No obstante, se requieren estudios adicionales para dilucidar los mecanismos moleculares que subyacen a la asociación entre estos polimorfismos y la DM1 y otras enfermedades autoinmunes.

A pesar de la gran multitud de estudios que asocian este gen con la DM1 y otras enfermedades autoinmunes, no existen evidencias de la asociación entre este gen y la diabetes tipo 2, por tanto se desconoce el mecanismo por el cual variaciones en este gen pueden estar relacionadas con la DM2.

Gen TCF7L2 (“Transcription factor 7-like 2”):

Según la secuencia de su ARNm NM_001146274.1 se ha identificado la mutación c.1484C>A que en la secuencia de la proteína (NP_001139746.1) corresponde con la mutación p.Pro495Arg. Se representa en la figura 60. No existen estudios sobre esta variación

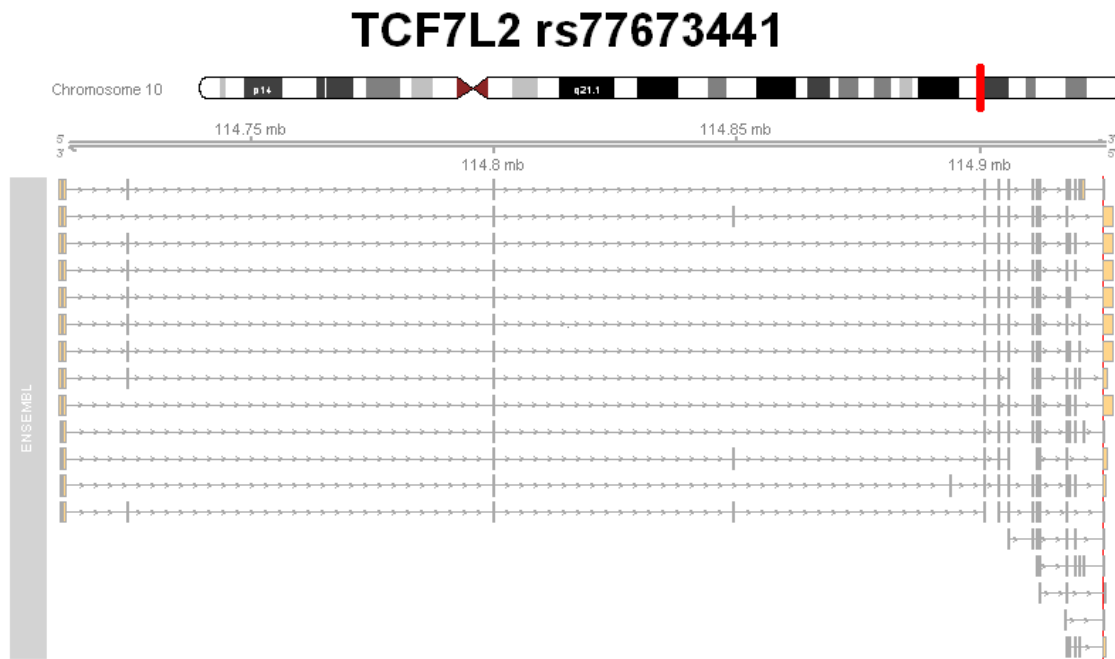


Figura 60: Representación de la variación rs77673441, marcada con la línea vertical roja, en los diferentes transcritos del gen TCF7L2.

Patogenicidad de la mutación.

Según PolyPhen, posiblemente esta mutación tendrá un efecto determinado sobre la funcionalidad de la proteína, con una puntuación de 0.501. Este resultado es corroborado por SIFT, que le da una puntuación de 0.05, lo cual la califica como deletérea.

Función del gen.

El locus que ha mostrado la asociación más fuerte con DM2 en diferentes etnias es el gen “Transcription factor 7-like 2” (TCF7L2), un miembro de la vía de señalización Wnt. La variante rs7903146 de este locus es común en la población, se localiza en la región no codificante del gen y da como resultado un aumento/disminución del riesgo de DM2 de aproximadamente un 40% (Grant et al, 2006). Sin embargo, sigue siendo desconocido el mecanismo por el cual el alelo T aumenta el riesgo de diabetes tipo 2 y la presencia del alelo C lo disminuye.

Existen diferentes estudios relacionados con el papel de TCF7L2 en DM2. Estos estudios han tratado de aclarar el papel de TCF7L2 en distintos órganos y tejidos incluyendo los

Discusión

islotos pancreáticos, intestino, tejido adiposo, o hígado. La depleción de TCF7L2 en islotos humanos resulta en la disminución de la proliferación y el aumento de la apoptosis de las células β . Además, la secreción de insulina estimulada por la glucosa se ve afectada por la pérdida de TCF7L2 tanto en islotos de ratón como humanos, mientras que la sobreexpresión de TCF7L2 en islotos no se ve influenciada por la glucosa ni por la apoptosis inducida por citoquinas (Shu et al, 2008). Se ha comprobado que mutantes TCF7L2 dominantes negativos para este gen reprimen la proliferación de células INS-1, lo que sugiere la importancia del papel de TCF7L2 en el mantenimiento de la masa de células β (Liu y Habener, 2008).

TCF7L2 también podría jugar un papel crítico en la homeostasis de la glucosa a través de la hormona “glucagon-like peptide 1” (GLP-1) producida en las células enteroendocrinas L del intestino delgado (Hansson et al, 2010). TCF7L2 se une a la región promotora del gen que codifica para el proglucagón y controla su actividad transcripcional en la línea celular GLUTag intestinal, donde el mutante dominante negativo de TCF7L2 suprime los niveles de ARNm de proglucagón (Yi et al, 2005).

Estudios recientes han revelado que TCF7L2 juega también un papel importante en el metabolismo, utilizando un modelo de ratón knockout condicional (Savic et al, 2011; Boj et al, 2012; Da Silva et al, 2012). La supresión de TCF7L2 en células β no revela ningún efecto en el desarrollo embrionario del páncreas endocrino, la proliferación de células β , ni la expresión de genes de células β . Sin embargo, el metabolismo lipídico hepático se ve afectado por la eliminación específica de TCF7L2 en el hígado. Esto sugiere que TCF7L2 puede actuar en la adaptación metabólica después del nacimiento, pero no durante el desarrollo embrionario. Otros estudios revelan que la expresión de varios genes clave implicados en el metabolismo del glucógeno, tales como glucógeno sintasa 2 (GYS2), y la gluconeogénesis, tales como fosfoenolpiruvatocarboxiquinasa 1 (PCK1) y glucosa-6-fosfato, subunidad catalítica (G6PC), se reduce significativamente en hígados de ratones knockout recién nacidos, en comparación con los de tipo salvaje (Boj et al, 2012).

Otros estudios muestran que la expresión de TCF7L2 se reduce en el tejido adiposo subcutáneo de pacientes con diabetes tipo 2, lo que sugiere además que TCF7L2 juega un papel en diversos tejidos metabólicos (Cauchi et al, 2006).

Por tanto dada la importante implicación de este gen en la diabetes tipo 2, la variación encontrada, rs77673441 (gen TCF7L2) podría tener un papel protector, antagonizando los efectos causados por la pérdida de función o la depleción de los niveles de TCF7L2 (menor masa de células β , menor liberación de insulina y mayor liberación de glucagón) aunque, al no haber estudios al respecto, se deberían realizar para tener un mayor conocimiento.

Gen ABCC8 (“ATP-binding cassette, sub-family C (CFTR/MRP), member 8”):

Según la secuencia de su ARNm, NM_001287174.1 se ha identificado la mutación c.2374A>T que en la secuencia de la proteína (NP_001274103.1) corresponde con la mutación p.Ser792Cys. Se representa en la figura 61.

Esta variación no ha sido descrita previamente y por tanto no existen estudios sobre ella.

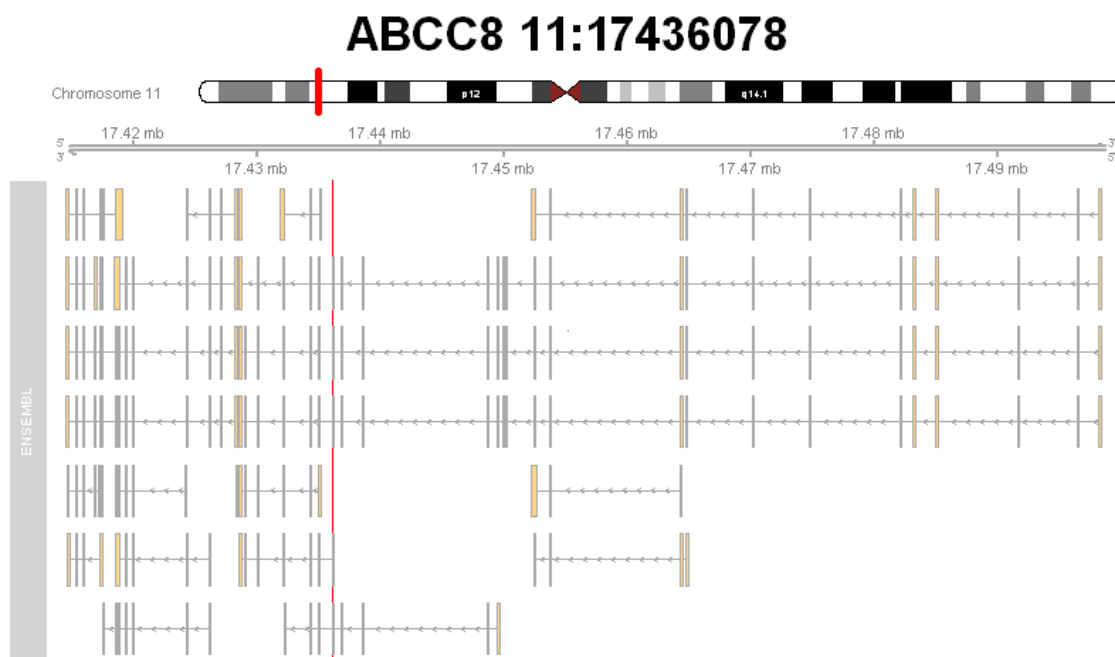


Figura 61: Representación de la variación 11:17436078, marcada con la línea vertical roja, en los diferentes transcritos del gen ABCC8.

Discusión

Patogenicidad de la mutación.

Según PolyPhen, probablemente esta mutación tendrá un efecto relevante en la funcionalidad de la proteína, con una puntuación de 0.964. Este resultado es corroborado por SIFT, que le da una puntuación de 0.02, lo cual la califica como deletérea.

Función del gen.

El gen ABCC8 está localizado en la región cromosómica 11p15.1 y codifica para el receptor de sulfonilurea 1 (SUR1) (Kapoor, 2010). SUR1 contiene tres dominios transmembrana, dos dominios citoplasmáticos de unión a nucleótidos y dos bucles citoplasmáticos. SUR1 puede acoplar la hidrólisis de ATP en los dominios de unión a nucleótidos para suministrar energía y desplazar a sustratos tales como azúcares, aminoácidos y lípidos (MaaBared, 2005; Pohl et al., 2005; Biemans-Oldehinkel et al, 2006; Kapoor, 2010).

ABCC8 tiene un papel esencial en la secreción de insulina: las células β pancreáticas regulan la producción y secreción de insulina a través del metabolismo de glucosa por el canal de potasio sensible a ATP (KATP) situado en la membrana plasmática. El metabolismo de la glucosa en las células β regula la secreción de insulina a través de la excitación de la célula por el canal de KATP. En condiciones normales, los iones de potasio fluyen hacia fuera de la célula a través de los canales KATP. Un aumento del metabolismo de la glucosa eleva los niveles de ATP y posteriormente inhibe SUR1, lo que conduce al cierre del canal KATP. Así, la membrana celular se despolariza y el calcio (Ca^{2+}) extracelular entra rápidamente al citosol, a través de canales de calcio dependientes de voltaje.

Los canales de calcio también activan los canales de potasio dependientes de voltaje, lo que repolariza la membrana celular y, en consecuencia se cierran los canales de calcio. El ión Ca^{2+} es un segundo mensajero intracelular ubicuo, crítico para el funcionamiento celular. El aumento de Ca^{2+} intracelular hace que otros componentes de la vía de secreción de insulina liberen gránulos en o cerca de la membrana plasmática. Una función anormal de ciertos componentes implicados en esta vía puede dar lugar a la

aparición de diabetes. Mutaciones patogénicas en el gen SUR1 pueden causar diferentes trastornos tales como la hipoglucemia hiperinsulinémica de la infancia y la DM2. La hipoglucemia hiperinsulinémica de la infancia es una enfermedad autosómica recesiva, resultado de una excesiva y mal regulada secreción de insulina (Ahlqvist et al, 2011; Lang et al, 2011).

Estudios en diferentes poblaciones han demostrado que SNPs en ABCC8 se asocian con el riesgo de padecer DM2 (Gloyn et al, 2001, 2003; Frazer et al., 2004; Hansen et al., 2005; Sakamoto et al., 2007; Alsmadi et al, 2008; Yu, 2009; Tang, 2009).

Por tanto, debido al papel tan importante del gen ABCC8 en el mantenimiento de la homeostasis de los niveles de glucosa, la variación encontrada en este estudio, al encontrarse en controles, podría tener un papel protector frente a la DM2 actuando en el proceso de liberación de insulina promovido por los propios niveles de glucosa.

Gen ERBB3 (“v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3”):

Según la secuencia de su ARNm NM_001982.3 se ha identificado la mutación c.3289T>C que en la secuencia de la proteína (NP_001973.2) corresponde con la mutación p.Ser1097Pro. Se representa en la figura 62.

No existen estudios sobre esta variación.

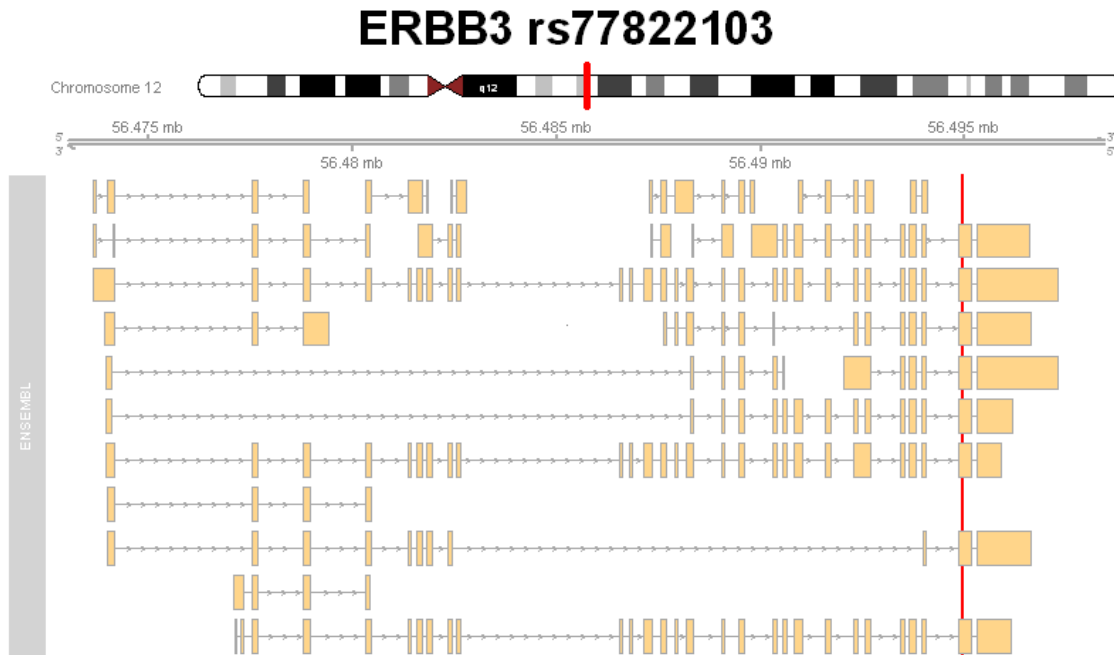


Figura 62: Representación de la variación rs77822103, marcada con la línea vertical roja, en los diferentes transcritos del gen ERBB3.

Patogenicidad de la mutación.

Según PolyPhen, probablemente esta mutación afectará la funcionalidad de la proteína resultante, con una puntuación de 0.999. SIFT, que le da una puntuación de 0, lo cual la califica como deletérea.

Función del gen.

Este gen codifica para un miembro del receptor del factor de crecimiento epidérmico (EGFR) de la familia de receptores tirosina quinasa. Forma heterodímeros con otros miembros de la familia del receptor EGF que tienen actividad quinasa. La heterodimerización conduce a la activación de las vías que conducen a la proliferación o diferenciación celular. El aumento en la expresión de este gen, se ha reportado en numerosos tipos de cáncer, incluyendo los de próstata, vejiga y mama. Se han caracterizado diferentes isoformas. Existe una isoforma que carece de la región intermembrana y se secreta fuera de la célula, la cual actúa modulando la actividad de la isoforma unida a la membrana.

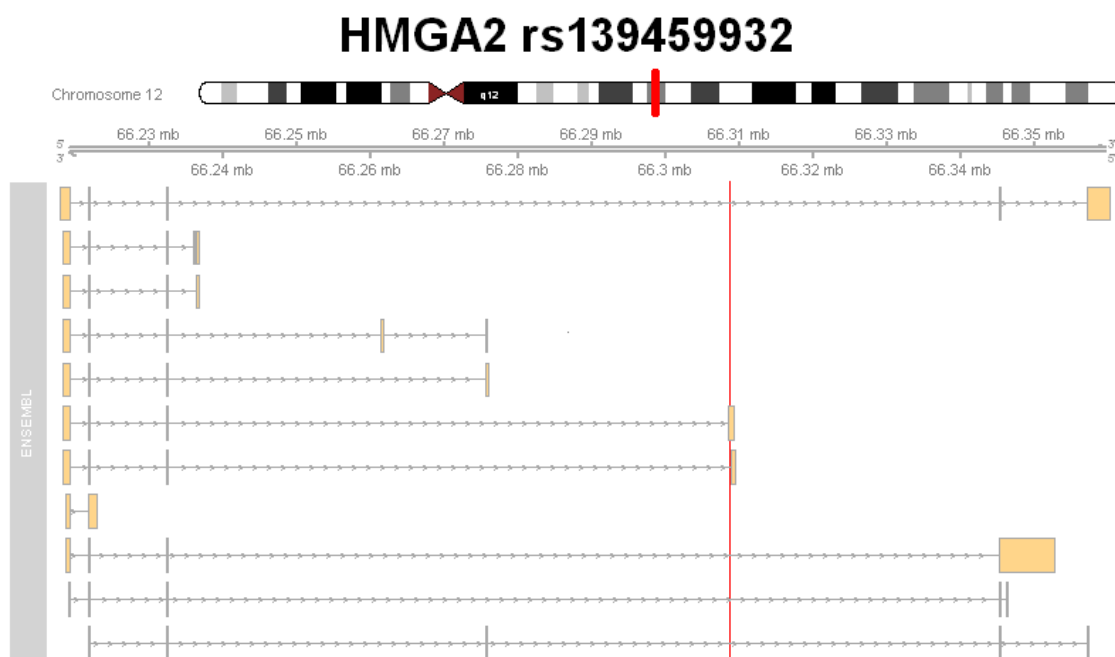
Se ha demostrado ampliamente que este gen se encuentra implicado en diabetes tipo 1 (Todd et al, 2007), sin embargo existen menos estudios que lo relacionen con diabetes tipo 2. Existe un estudio (Hemi et al, 2002) en que se relaciona ERBB3 con la fosforilación de las proteínas IRS, en él se concluye que la transactivación de los receptores ErbB2/ErbB3 desencadena la cascada PI3K, lo cual induce la fosforilación de las proteínas IRS ocasionando así una resistencia a la insulina.

Por tanto, la variación encontrada es una evidencia de que este gen se encuentra relacionado con la diabetes tipo 2, y el mecanismo por el cual se da esta relación podría ser la pérdida de actividad de ERBB3 que a su vez, reduciría la fosforilación de las proteínas IRS y de esta forma protegería frente a la resistencia a insulina y por tanto, frente a la diabetes.

Gen HMGA2 (“High mobility group AT-hook 2”):

Según la secuencia de su ARNm NM_001300919 se ha identificado la mutación c.406T>C que en la secuencia de la proteína (NP_001287848) corresponde con la mutación p.Cys136Arg. Se representa en la figura 63.

No existen estudios sobre esta variación.



Discusión

Figura 63: Representación de la variación rs139459932, marcada con la línea vertical roja, en los diferentes transcritos del gen HMGA2.

Patogenicidad de la mutación.

PolyPhen le otorga una puntuación de 0.983, lo cual confirma que esta mutación afecta a la funcionalidad de la proteína resultante. Este resultado es corroborado por SIFT, que le da una puntuación de 0.01, lo cual la califica como deletérea.

Función del gen.

Este gen codifica para una proteína que pertenece a la familia de proteínas no histonas de alta movilidad cromosómica (HMG). Estas funcionan como factores de soporte o estructurales y son componentes esenciales del enhanceosoma (complejo proteico que se une a la región promotora de un gen para activar su expresión). Esta proteína contiene dominios de unión al ADN y puede actuar como factor de regulación transcripcional. La identificación de deleciones, amplificaciones, y reordenamientos en este gen, se asocian con el liposarcoma mixoide, lo que sugiere un papel de éste en la adipogénesis y en la diferenciación mesenquimal.

En el metaanálisis de Voight et al (2010) se asoció este gen con diabetes tipo 2. El posible mecanismo que se plantea con respecto a su asociación con esta enfermedad es que parece tener un efecto primario en acción de la insulina, pero sin ninguna relación con la obesidad. Además, HMGA2 es un regulador transcripcional clave de IGF2BP2 (La proteína codificada por IGF2BP2 actúa mediante unión a la zona 5'UTR del ARNm del factor de crecimiento similar a la insulina 2 (IGF2) y regulando su traducción) (Brants et al, 2004; Cleynen et al, 2007). Sin embargo, debido a que los ratones knockout HMGA / Hmg1c son deficientes en la diferenciación de adipocitos (Anand y Chada, 2000), y el alelo de riesgo de IGF2BP2 se asocia con una reducción de la función de las células β (Groenewoud et al, 2008), es necesario seguir trabajando para establecer la relevancia de esta interacción en la patogénesis de la enfermedad.

Por tanto la variación encontrada en este estudio, rs139459932 (gen HMGA2), podría encontrarse protegiendo frente a la DM2, por un lado modulando la liberación de

insulina y por el otro, regulando la transcripción del gen IGF2BP2 y con ello, la función de las células β e incluso mediante el desarrollo del tejido adiposo.

7.2-VARIACIONES IDENTIFICADAS Y VALIDADAS EN CASOS

Estas variaciones se encuentran en individuos que padecen la enfermedad por tanto, podrían ser mutaciones causantes, por sí solas o en conjunto con otras.

Gen NOTCH2 (“notch 2”):

Según la secuencia de su ARNm, NM_024408.3 se ha identificado la mutación c.7223T>A que en la secuencia de la proteína (NP_077719.2) corresponde con la mutación p.Leu2408His. Se representa en la figura 64. No existen estudios sobre esta variación.

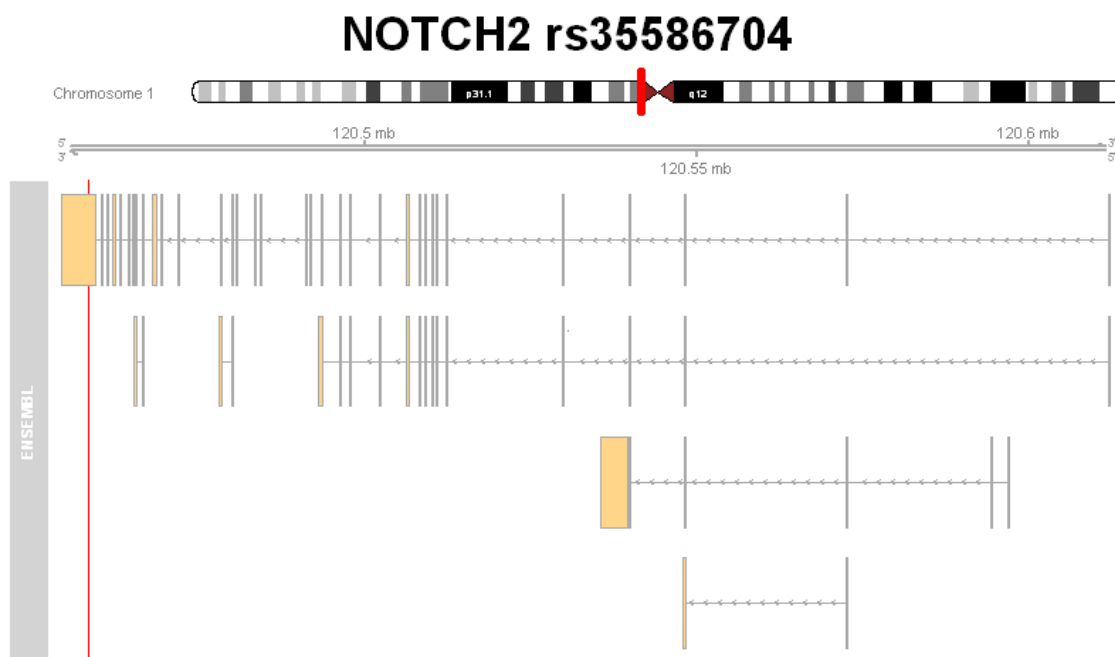


Figura 64: Representación de la variación rs35586704, marcada con la línea vertical roja, en los diferentes transcritos del gen NOTCH2.

Discusión

Patogenicidad de la mutación.

Según PolyPhen, esta mutación probablemente afectará la funcionalidad de la proteína, con una puntuación de 0.968. Este resultado es corroborado por SIFT, que le da una puntuación de 0.01, lo cual la califica como deletérea.

Función del gen.

Este gen codifica para un miembro de la familia Notch. Los miembros de esta familia de proteínas transmembrana tipo 1, comparten características estructurales tales como: un dominio extracelular que consiste en múltiples repeticiones del factor de crecimiento epidérmico (EGF), y un dominio intracelular que consta de múltiples y tipos de dominios. La vía Notch es una vía de señalización intercelular conservada evolutivamente que regula las interacciones entre células adyacentes. Es una de las vías de señalización básicas más implicadas en el desarrollo, y que está involucrada tanto en la especificación del tipo celular como en la organogénesis. Esta proteína funciona como un receptor para ligandos unidos a membrana, y puede desempeñar un papel en el desarrollo vascular, renal y hepático.

El papel de esta vía ha sido ampliamente estudiado y fue reportado por primera vez en 1914, en el desarrollo de *Drosophila*, pero su papel en el desarrollo humano y la enfermedad comenzó a estudiarse más tarde, cuando mutaciones en NOTCH1 se asociaron con una forma de leucemia linfoblástica aguda de células T (LLA-T) (Ellisen et al, 2001). En 1996, se encontraron mutaciones en NOTCH3 causantes de CADASIL, un trastorno caracterizado por el accidente cerebrovascular y demencia, con una edad de aparición en la tercera o cuarta década de la vida del paciente (Joutel et al, 1996).

Este gen fue asociado con diabetes tipo 2 en el metaanálisis de Zeggini et al (2008). Se encontró asociación con rs10923931, que se encuentra en el intrón 5 del gen. En ratones, Notch2 se expresa en las células ductales embrionarias (fuente de las células madre endocrinas y exocrinas) durante la organogénesis pancreática (Lammert et al, 2000). Por otra parte, existe un estudio (Jonsson et al, 2013) en que se ha confirmado que variantes en el gen NOTCH2 afectan a la secreción de glucagón *in vivo* e *in vitro*. De

hecho individuos con variaciones en este gen muestran concentraciones más elevadas de glucagón en ayunas.

Por tanto, variaciones en este gen, tales como la que se ha encontrado en este estudio (rs35586704) podrían causar algún problema durante la organogénesis pancreática y esto conduciría a anomalías en las células β , lo que aumentaría el riesgo de desarrollar DM2. Por otra parte, estas variaciones también pueden afectar a la función de las células α pancreáticas, aumentando la secreción de glucagón, lo que podría derivar en una hiperglucemia que conduciría también al desarrollo de diabetes.

Gen THADA (“thyroid adenoma associated”):

Según la secuencia de su ARNm NM_001083953.1 se ha identificado la mutación c.4351C>T que en la secuencia de la proteína (NP_001077422.1) corresponde con la mutación p.Pro1451Ser. Se representa en la figura 65.

No existen estudios sobre esta variación.

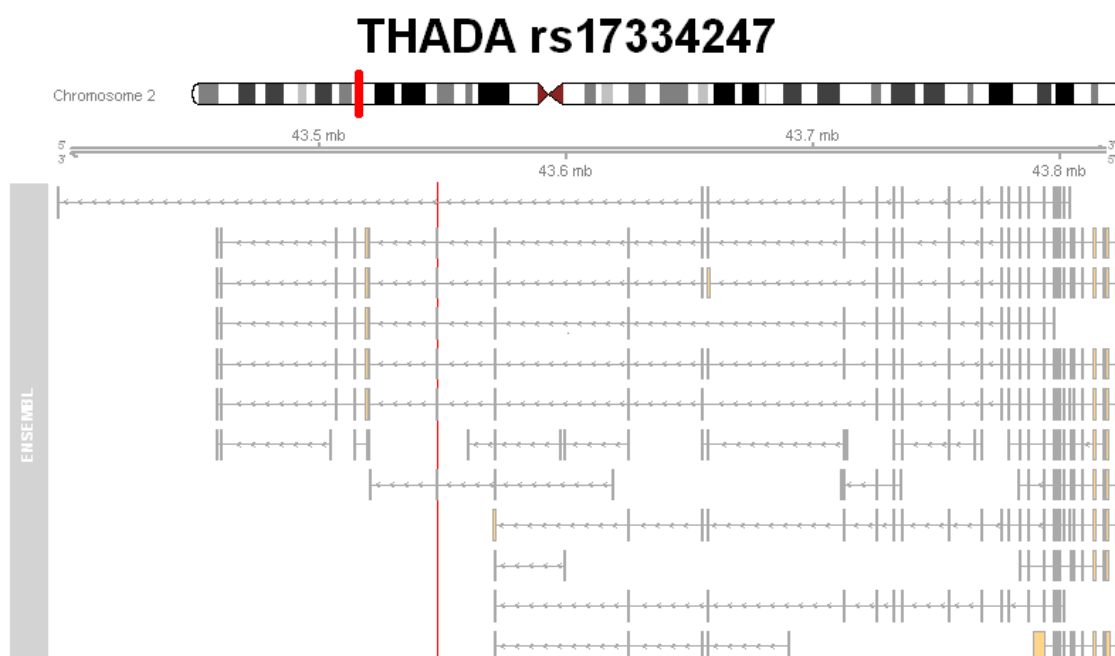


Figura 65: Representación de la variación rs17334247, marcada con la línea vertical roja, en los diferentes transcritos del gen THADA.

Discusión

Patogenicidad de la mutación.

PolyPhen, le otorga una puntuación de 0.842, con lo cual posiblemente esta mutación afectará a la funcionalidad de la proteína resultante. SIFT, corrobora este resultado dándole una puntuación de 0.01, lo cual la califica como deletérea.

Función del gen.

La función de este gen se ha detallado en el apartado anterior (variaciones identificadas y validadas en controles, gen THADA: página 159). Puesto que la variación identificada, se ha encontrado en los casos, su mecanismo de acción podría ser el descrito para otros polimorfismos del gen THADA, los cuales pueden conducir a una menor funcionalidad de las células β y al aumento de su apoptosis.

Gen IFIH1 (“interferon induced with helicase C domain 1”):

Según la secuencia del ARNm NM_022168.3 del gen IFIH1, se ha identificado la mutación c.1097T>A que, en la secuencia de la proteína (NP_071451.2) corresponde con la mutación p.Val366Glu. Se representa en la figura 66.

No existen estudios sobre esta variación.

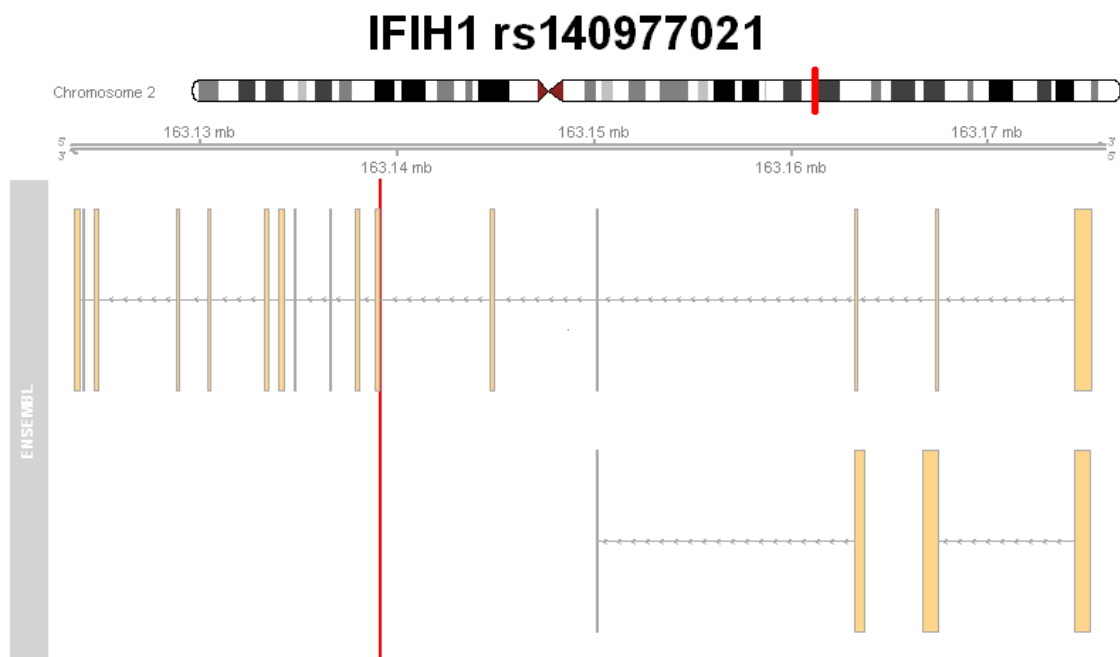


Figura 66: Representación de la variación rs140977021, marcada con la línea vertical roja, en los diferentes transcritos del gen IFIH1.

Patogenicidad de la mutación.

Según PolyPhen, probablemente esta mutación afectará la funcionalidad de la proteína resultante, con una puntuación de 1. SIFT confirma este resultado dándole una puntuación de 0, lo cual la califica como deletérea.

Función del gen.

La función de este gen se ha detallado en el apartado anterior (variaciones identificadas y validadas en controles, gen IFIH1: página 163). A pesar de la gran multitud de estudios que asocian este gen con la DM1 y otras enfermedades autoinmunes, no existen evidencias de la asociación entre este gen y la diabetes tipo 2, por tanto se desconoce el mecanismo por el cual variaciones en este gen pueden estar relacionadas con la DM2. Un segundo problema es que ésta u otras alteraciones en este gen puedan producir una diabetes similar a la tipo 1, pero cuyo desarrollo sea más tardío y/o más lento y, presentando un fenotipo similar al de la DM2.

Gen AGMO (“alkylglycerolmonooxygenase”):

Según la secuencia del ARNm NM_001004320.1 del gen AGMO, se ha identificado la mutación c.160G>C que en la secuencia de la proteína (NP_001004320.1) corresponde con la mutación p.Glu54Gln. Se representa en la figura 67.

No existen estudios sobre esta variación.



Figura 67: Representación de la variación rs146766067, marcada con la línea vertical roja, en los diferentes transcritos del gen AGMO.

Patogenicidad de la mutación.

Según PolyPhen, probablemente esta mutación afectará a la funcionalidad de la proteína, con una puntuación de 1. Este resultado es corroborado por SIFT, que le da una puntuación de 0, lo cual la califica como deletérea.

Función del gen.

AGMO codifica para el isotipo β del dominio catalítico de la diacilglicerol quinasa. Esta enzima cataliza la conversión de diacilglicerol (DAG) a ácido fosfatídico (PA) utilizando ATP como fuente de fosfato. En los islotes pancreáticos de rata, la glucosa aumenta los niveles de diacilglicerol (Peter-Riesch et al, 1998), lo que activa la proteína quinasa C (PKC) potenciando así la secreción de insulina (Prentki y Matschinsky, 1987).

El metaanálisis de Dupuis et al (2010), encontró asociación entre este gen y DM2. En otro estudio del mismo año, Boesgaard et al afirmaron que variantes en el locus AGMO se asocian con una disminución de la liberación de insulina estimulada por glucosa.

Según el estudio de Goodarzi et al (2013), el locus AGMO modula el metabolismo lipídico y puede contribuir a la resistencia hepática a la insulina. AGMO participa en la regulación de los niveles celulares de diacilglicerol. La acumulación de diacilglicerol en el hígado parece afectar a la señalización de insulina debido a la activación de la proteína quinasa C epsilon (Kumashiro et al, 2011). La actividad de esta quinasa también puede regular la cantidad de diacilglicerol disponible para su conversión a triacilglicerol. Las observaciones anteriores plantean la posibilidad de que AGMO afecte al efecto o la actividad de la insulina sobre diferentes tejidos.

En conclusión, la variación encontrada (rs146766067) podría asociarse a la alteración de la liberación de insulina estimulada por glucosa, al aumentar ésta los niveles de diacilglicerol en diferentes tejidos, lo que podría facilitar el desarrollo de DM2. De confirmarse esta actividad, apoyaría los hallazgos previos sobre la importancia del locus AGMO en la DM2.

Gen CLEC7A (“C-type lectin domain family 7, member A”):

Según la secuencia del ARNm NM_197947 del gen CLEC7A, se ha identificado la mutación c.714T>G que en la secuencia de la proteína (NP_922938) corresponde con la mutación p.Tyr238Ter. Se representa en la figura 68.

Solamente existe un estudio sobre esta variación (Kutikhin et al, 2012), en el que se señala este polimorfismo y algunos otros como objetivos prometedores para la oncogenómica y se sugiere que podrían ser utilizados en programas de prevención y en el diagnóstico precoz del cáncer en un futuro.

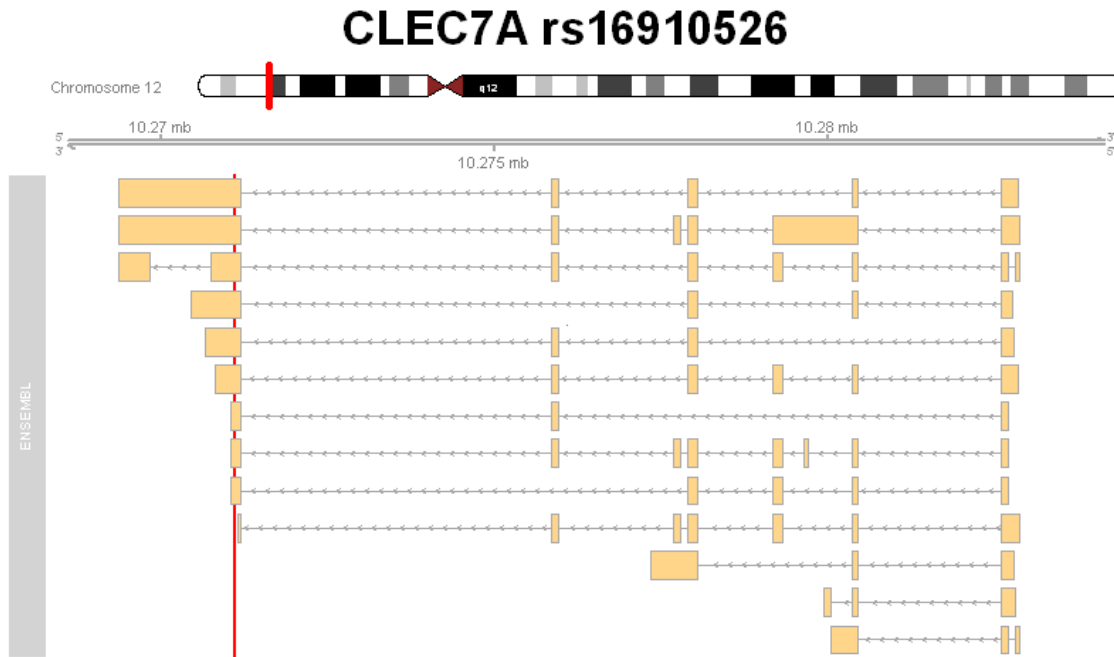


Figura 68: Representación de la variación rs16910526, marcada con la línea vertical roja, en los diferentes transcritos del gen CLEC7A.

Patogenicidad de la mutación.

En este caso el resultado de la mutación del gen es una proteína truncada, por lo que no se puede calcular la puntuación SIFT ni PolyPhen. Dada la importante alteración que esto supone en la proteína, será muy probablemente patogénica. Si bien, la mutación se encuentra en el último exón (penúltimo en uno de los transcritos), lo que podría originar una proteína parcialmente funcional, aunque serían necesarios ensayos para confirmarlo.

Función del gen.

Este gen codifica un miembro de la superfamilia “C-type lectin/C-type lectin-like domain” (CTL/CTLD). Se trata de un receptor de membrana tipo II con un dominio extracelular de tipo C-lectina y un dominio citoplasmático. Es un receptor de reconocimiento de patrones que reconoce gran variedad de glucanos presentes en hongos y plantas y, por tanto, desempeña un papel importante en la respuesta inmune innata. Este gen está estrechamente vinculado a otros miembros de la superfamilia

CTL/CTLD en el cromosoma 12p13, concretamente en la región complejo génico “natural killer”.

La proteína codificada por este gen está presente en células mieloides tales como macrófagos, células dendríticas y neutrófilos (Kerrigan et al, 2010). También se expresa en otros tipos de células incluyendo los eosinófilos y células B en humanos, aunque en ratones este receptor no se ha encontrado en estas poblaciones celulares (Brown, 2006). CLEC7A es un receptor de reconocimiento de patrones implicado en la respuesta inmune innata y por tanto, en el reconocimiento de algunos patógenos (Drummond et al, 2011).

Se ha demostrado que participa en el inicio de la fagocitosis y la producción de especies reactivas de oxígeno (ROS) en respuesta a zymosan (compuesto empleado para inducir inflamación) en los macrófagos, los neutrófilos y células dendríticas (Brown et al, 2001; Gantner et al, 2003; Kennedy et al, 2007; Bauer et al, 2008). Se sabe que es un mediador importante de estas funciones en ratones, sin embargo su papel en los fagocitos humanos es aún incierto. Otra función bien caracterizada de CLEC7A es la producción de mediadores lipídicos (Suram et al, 2006; Alvarez et al, 2010) y citoquinas (por ejemplo, TNF, IL-1 β , IL-10, IL-6, IL-23, CCL2, CCL3) (Dennehy et al, 2009; Reid et al, 2009). Esta producción está controlada por varias vías de señalización y es esencial para la inmunidad antifúngica (Netea et al, 2006).

Este gen no había sido asociado previamente con diabetes tipo 2, sin embargo ha sido incluido en este estudio por tratarse de un proyecto coordinado en que se estudian genes implicados o potencialmente implicados en diferentes tipos de diabetes. Por tanto se desconoce el mecanismo por el cual variaciones en él podrían estar relacionadas con DM2. Sí que existen estudios que lo relacionan con dolor neuropático (Rojewska et al, 2014), pero la mayor parte de estudios se centran en su papel en procesos inmunitarios (Duluc et al, 2014; Smith et al, 2014). Sin embargo, otro gen de la misma familia, CLEC16A, sí que se ha asociado con diabetes mellitus tipo 1 (Tomlinson et al, 2014; Sang et al, 2012) además de esclerosis múltiple, y disfunción adrenal (Hakonarson et al, 2007; Skinningsrud et al., 2008).

Discusión

Además, se han observado variaciones en este gen en una enfermedad similar a IPEX. IPEX es un trastorno inmunológico recesivo ligado al cromosoma X, que se caracteriza por diarrea severa debido a la enteropatía existente, diabetes mellitus insulino-dependiente, y dermatitis. Los pacientes también pueden padecer hipotiroidismo, anemia hemolítica autoinmune, trombocitopenia, linfadenopatía, hepatitis y nefritis (d'Hennezel et al, 2012). La importancia de este hallazgo radica en que la diabetes presente en el síndrome similar a IPEX quizá podría tener relación con este gen, pero todavía no existen estudios al respecto.

Gen LGR5 ("leucine-rich repeat containing G protein-coupled receptor 5"):

Según la secuencia del ARNm NM_003667 del gen LGR5, se ha identificado la mutación c.1270C>G que, en la secuencia de la proteína (NP_003658) corresponde con la mutación p.Leu424Val. Se representa en la figura 69.

No existen estudios sobre esta variación.

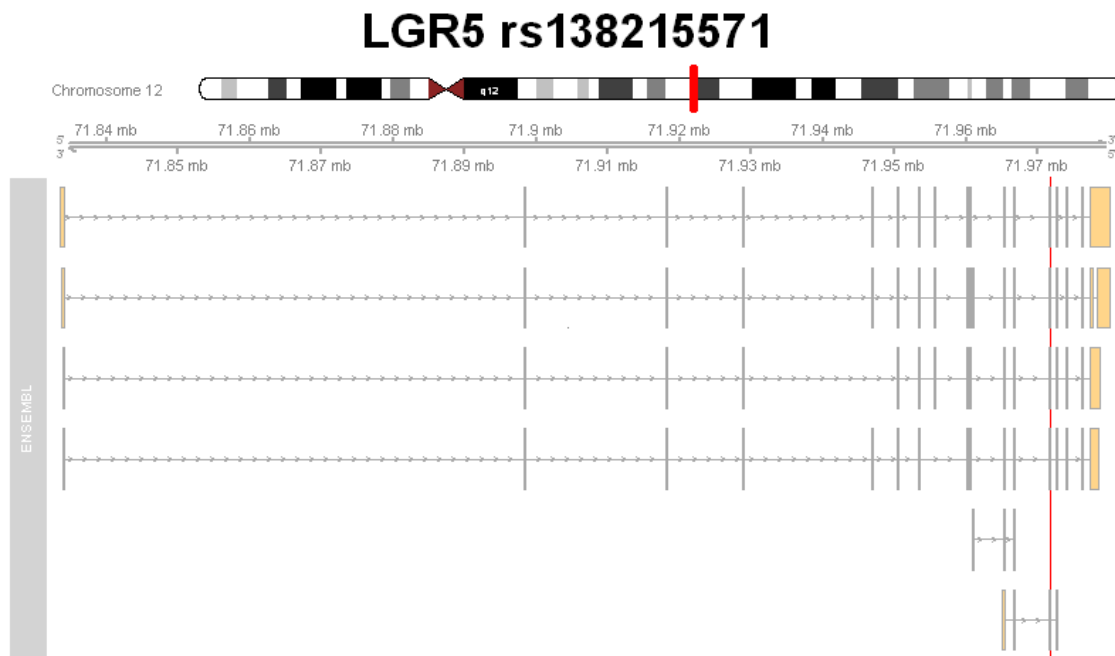


Figura 69: Representación de la variación rs138215571, marcada con la línea vertical roja, en los diferentes transcritos del gen LGR5.

Patogenicidad de la mutación.

Según PolyPhen, probablemente esta mutación tendrá algún efecto sobre la funcionalidad de la proteína resultante, con una puntuación de 0.999. Este resultado es corroborado por SIFT, que le da una puntuación de 0, calificándola así como deletérea.

Función del gen.

LGR5 es un miembro de la familia de receptores de proteína G, que comprende proteínas con siete dominios transmembrana. Ha sido identificado como un nuevo marcador de células madre adultas en el intestino delgado y los folículos pilosos (Jaks et al, 2008; Barker et al, 2010) y se encuentra ampliamente expresado en la médula espinal, mama, folículos pilosos, y tejidos cerebrales (Hsu et al, 1998). LGR5 también juega un papel importante durante la embriogénesis.

En los últimos años, muchos estudios han revelado que LGR5 se sobreexpresa en varios tipos de tumores (Yamamoto et al, 2003; Colon et al, 2006; Tanese et al, 2008; Becker et al, 2010; Al-Kharusi et al, 2013). A pesar de que la mayoría de estudios que existen sobre este gen se encuentran relacionados con cáncer, éste se ha asociado también con diabetes tipo 2 (Zeggini et al, 2008; Zhou et al, 2010). Sin embargo, no hay información acerca del mecanismo por el cual este gen puede estar implicado en la enfermedad.

Gen OASL (“2'-5'-oligoadenylate synthetase-like”):

Según la secuencia del ARNm NM_003733.3 del gen OASL, se ha identificado la mutación c.1309T>G que en la secuencia de la proteína (NP_003724.1) corresponde con la mutación p.Phe437Val. Se representa en la figura 70.

Esta variación no ha sido descrita previamente y por tanto no existen estudios sobre ella.

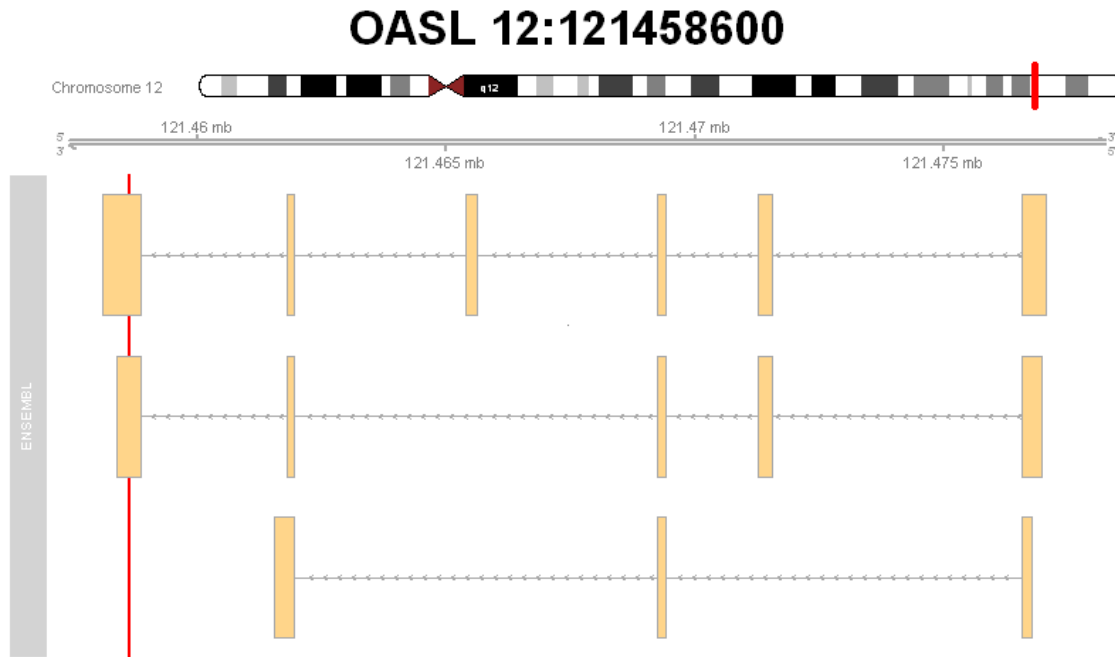


Figura 70: Representación de la variación 12:121458600, marcada con la línea vertical roja, en los diferentes transcritos del gen OASL.

Patogenicidad de la mutación.

Según PolyPhen, probablemente esta mutación tendrá un efecto sobre la funcionalidad de la proteína resultante, con una puntuación de 0.971. SIFT corrobora este resultado dándole una puntuación de 0.02, lo cual la califica como deletérea.

Función del gen.

Las 2', 5'- oligoadenilato sintetasas (OASs) se encuentran entre las llamadas proteínas antivirales inducidas por interferón (Sadler y Williams, 2008; Kristiansen et al, 2011). En humanos, la familia de genes de la OAS está compuesta de 4 genes: OAS1, OAS2, OAS3 y OASL.

El gen OASL codifica una proteína de dos dominios compuesta por una unidad de la OAS unida a un dominio C-terminal, que es homólogo a una repetición en tándem de ubiquitina (Hartmann et al., 1998). La función de OASL ha sido poco investigada, pero estudios recientes indican la importancia de este gen en la respuesta inmune relacionada con la actividad antiviral del interferón (Schoggins et al., 2011). El gen OASL está altamente inducido por infecciones víricas y bacterianas (Cheng et al., 2006; Zhou

y Amar, 2007), por la estimulación con LPS (Smith et al., 2002), y por algunas enfermedades como el lupus (Zhao et al., 2010).

La expresión ectópica de OASL en células Vero suprime la replicación de los virus de ARN de cadena sencilla (Marques et al., 2008). Además, se ha demostrado que OASL inhibe de la replicación viral del Virus de la Hepatitis C (Ishibashi et al., 2010 y Schoggins et al., 2011). Los mecanismos que conducen a la actividad antiviral de OASL siguen siendo desconocidos, sin embargo, esta actividad parece requerir el dominio semejante a la ubiquitina C-terminal para llevarse a cabo (Marques et al., 2008).

La mayoría de estudios sobre OASL se centran en su papel en la respuesta a diferentes tratamientos antivirales, ya que suele aumentar su expresión con la toma de los mismos, por lo que es importante por ejemplo, en pacientes con hepatitis C crónica (Jiménez-Sousa et al, 2011). Otros muchos estudios se centran en su papel en respuesta a las infecciones virales como el dengue (Lin et al, 2009).

Sin embargo, existe un estudio (Middelberg et al, 2011) en que se relaciona este gen con diferentes variables cardiovasculares. Se encontró asociación con los niveles de colesterol LDL, gamma glutamiltransferasa y proteína C reactiva. La conexión funcional entre el gen OASL y estos fenotipos no está clara. Por otra parte, entre los genes cercanos y en desequilibrio de ligamiento con el SNP principal de OASL, está HNF1A. HNF1A se expresa en hígado, riñón y páncreas endocrino y regula una serie de genes implicados en el metabolismo de las lipoproteínas, incluyendo las apolipoproteínas, enzimas de síntesis de colesterol y transportadores de ácidos biliares (Odom et al, 2004). HNF1A también tiene asociaciones alélicas con diabetes tipo 2 (Voight et al, 2010), con la PCR (Elliott et al, 2009; Dehghan et al, 2011; Okada et al, 2011) y con enfermedad coronaria (Erdmann et al, 2009).

En el metaanálisis de Voight y colaboradores (2010), se relaciona este gen con diabetes tipo 2. Se señalan algunos locus cercanos a ciertos genes relacionados con la enfermedad en los que, la biología existente, junto con datos fenotípicos y de expresión, señalan esos locus como candidatos para la susceptibilidad frente a la enfermedad.

Discusión

Estudios sobre esta alteración podrían ayudar a discernir la importancia de este gen en DM2.

Gen PAX4 (“paired box 4”):

Según la secuencia de su ARNm, NM_006193.2 se ha identificado la mutación c.553C>G que en la secuencia de la proteína (NP_006184.2) corresponde con la mutación p.Gln185Glu. Se representa en la figura 71.

Esta variación no ha sido descrita previamente y por tanto no existen estudios sobre ella.

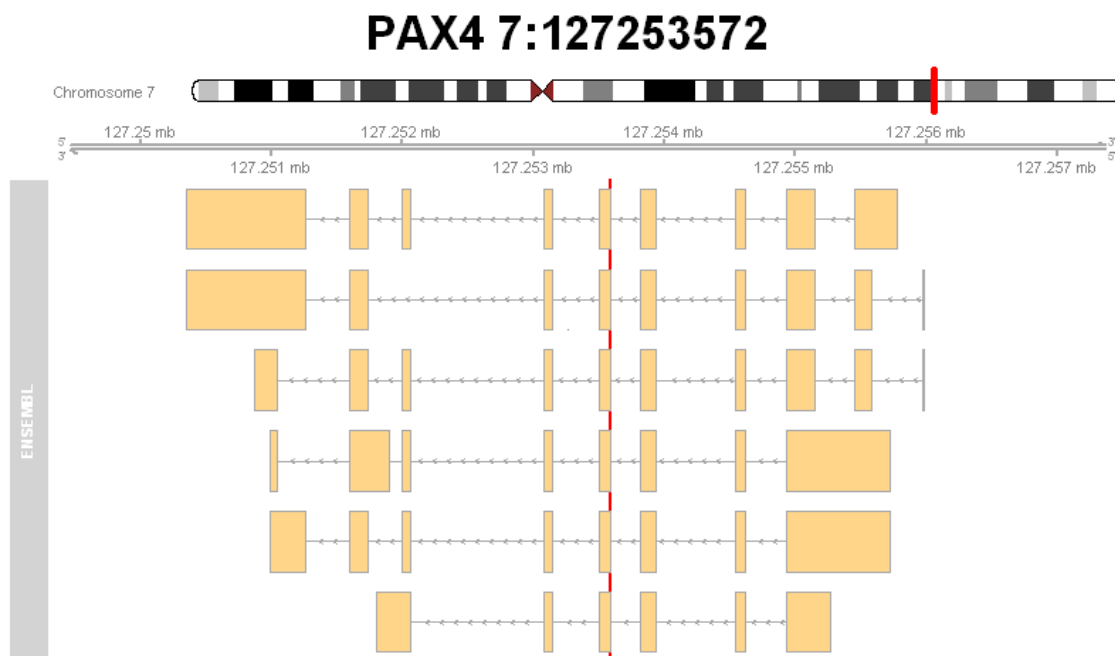


Figura 71: Representación de la variación 7:127253572, marcada con la línea vertical roja, en los diferentes transcritos del gen PAX4.

Patogenicidad de la mutación.

PolyPhen le otorga una puntuación de 0.985, por lo que probablemente esta mutación afectará a la funcionalidad de la proteína resultante. Por su parte, SIFT le da una puntuación de 0.01, lo cual la califica como deletérea.

Función del gen.

Este gen es un miembro de la familia de factores de transcripción “Paired box” (PAX). Estos genes juegan un papel crítico durante el desarrollo fetal y también en el desarrollo de cáncer. PAX4 está involucrado en el desarrollo de los islotes pancreáticos, de hecho estudios en ratones han demostrado la implicación de este gen en la diferenciación de las células β (Sosa-Pineda, 2004).

Se han caracterizado nueve genes Pax en mamíferos. Son reguladores de la expresión génica, y poseen un papel muy importante en la formación del sistema nervioso central, la cresta neural y la musculatura. Además, estos genes confieren a las células la capacidad para proliferar, asumir un destino celular y diferenciarse. Se expresan principalmente en ojos (Marquardt et al, 2001; Farhy et al, 2013), páncreas (Brun y Gauthier, 2008; Gosmain et al, 2011; Hu He et al, 2011; Kooptiwut et al, 2012; Pfeifer et al, 2013) y riñón (Dressler, 2011).

Mutaciones en PAX4 se han asociado con diabetes tipo 2. El estudio de Sosa-Pineda (2004) detalla estas asociaciones: en la población japonesa con diabetes tipo 2, existe una mutación con cambio de sentido en el gen PAX4 (R121W). Los pacientes con la mutación PAX4 R121W en homocigosis mostraron defectos graves en la secreción de insulina, lo que sugiere que esta alteración genética da lugar a la disfunción de las células β (Kanatsuka et al., 2002). En las personas de ascendencia africana (Mauvais-Jarvis et al, 2004) se identificaron otras dos mutaciones en este gen (R133W o R37W). Individuos homocigotos R133W o heterocigotos R37W mostraron alteraciones en los niveles de secreción de insulina durante una prueba de tolerancia a la glucosa. En general, estos estudios sugieren que mutaciones específicas modifican la función normal de PAX4 en combinación con factores ambientales y otros factores genéticos, dando lugar a una menor masa de células β o a una menor capacidad de respuesta de estas células al aumentar la demanda de insulina. Así, en algunos grupos étnicos, mutaciones en este gen aumentan su predisposición a la diabetes tipo 2. Por tanto, los resultados de estos estudios determinan que PAX4 es un componente clave de la maquinaria molecular responsable de iniciar la diferenciación celular β -pancreática.

Discusión

En base a estos datos, los mecanismos que podrían explicar la implicación de la variación identificada serían: generación de una menor masa de células β y/o la menor capacidad de respuesta de éstas para la liberación de insulina.

Conclusiones

CONCLUSIONES

La secuenciación de exoma completo ha demostrado ser una herramienta útil y eficiente en el descubrimiento de los genes responsables de enfermedades con herencia mendeliana. Sin embargo, hasta ahora no se había utilizado de forma generalizada para el estudio de enfermedades de herencia compleja.

Los objetivos planteados al inicio del estudio han sido alcanzados, habiéndose identificado diferentes variantes en genes asociados a la DM que podrían tener un papel relevante en el desarrollo de esta enfermedad. Se han identificado variaciones raras que, o no se habían descrito previamente o bien, se habían descrito pero no se habían relacionado con la DM2. Estas variaciones parecen tener un efecto relevante en la funcionalidad de las proteínas, según los programas de predicción empleados, pero son necesarios estudios funcionales para comprobar su efecto real.

Los genes analizados en este trabajo han sido previamente asociados con DM (excepto el gen CLEC7A), aunque en varios de ellos no se sabe realmente cual es el mecanismo por el que pueden estar implicados en la enfermedad. A través de revisión bibliográfica, se ha hecho un estudio de las funciones de los diferentes genes, aunque es necesario realizar diferentes estudios a diferentes niveles para poder conocer el mecanismo por el cual están relacionados con la enfermedad.

Es necesario analizar la presencia de estas variaciones tanto en estudios caso-control como en estudios poblacionales, con mayor número de muestras y de diferente procedencia, para poder conocer la importancia de estas alteraciones en nuestra población y en otras. Estos estudios están previstos en la siguiente fase del estudio. Además, otro objetivo es identificar también las variantes que puedan localizarse en el resto del exoma y que puedan estar implicadas en el desarrollo de la DM2. Este trabajo permitirá identificar genes cuyo papel o asociación con la DM2 no se ha descrito hasta el momento.

Conclusiones

Por tanto las conclusiones que se pueden obtener son:

- La secuenciación de exoma completo se dibuja como una herramienta útil para el estudio de las enfermedades de herencia compleja, lo que abre una nueva vía de investigación para aclarar las bases genéticas de estas enfermedades.
- Bajo las condiciones de este estudio, la tecnología Illumina y los filtros usados presentan un 20% de error en la identificación de mutaciones, tasa que es asumible para su validación por otros medios y que nos permite identificar un mayor número de alteraciones potencialmente implicadas en DM2.
- Los resultados de este estudio apoyan la hipótesis de que una parte de la heredabilidad de la DM2 se puede explicar a través de polimorfismos poco frecuentes.
- Se han identificado 16 variaciones relacionadas con DM2, algunas desconocidas y otras que no se habían relacionado previamente con ella a pesar de haber sido descritas.
- Alteraciones en los genes: GCKR, THADA, ITGB6, IFIH1, TCF7L2, ABCC8, ERBB3, HMGA2, NOTCH2, PAX4, CLEC7A, LGR5 y OASL pueden estar relacionados con DM2, lo que confirmaría los hallazgos de los GWAS realizados previamente.

Bibliografía

BIBLIOGRAFÍA

A:

-Agius L, Peak M, Newgard CB, Gomez-Foix AM, Guinovart JJ. (1996). Evidence for a role of glucose-induced translocation of glucokinase in the control of hepatic glycogen synthesis. *J. Biol. Chem.* 271, 30479–30486.

-Agius L. (1998). The physiological role of glucokinase binding and translocation in hepatocytes. *Adv. Enzyme Regul.* 38, 303–331.

-Aguilar-Bryan L, Bryan J. (2008). Neonatal diabetes mellitus. *Endocr Rev*, 29, 265–291.

-Agurs-Collins TD, Kumanyika SK, Have TR, Adams-Campbell LL. (1997). A randomized controlled trial of weight reduction and exercise for diabetes management in older African-American subjects. *Diabetes Care*, 20, 1503–1511.

-Ahlqvist, E., Ahluwalia, T.S., Groop, L. (2011). Genetics of type 2 diabetes. *Clinical Chemistry*, 57, 241–254.

-Albala C, Villarroel A, Santos JL, Angel B, Lera L, Liberman C, Sanchez H, Pérez-Bravo F. (2007). FABP2 Ala54Thr polymorphism and diabetes in Chilean elders. *Diabetes Res Clin Pract*, 77, 245-250.

-Alberti KG, Zimmet P, Shaw J. (2007). International Diabetes Federation: a consensus on type 2 diabetes prevention. *Diabet Med*, 24, 451–463.

-Al-Kharusi MR, Smartt HJ, Greenhough A, Collard TJ, Emery ED, Williams AC, Paraskeva C. (2013). LGR5 promotes survival in human colorectal adenoma cells and is upregulated by PGE2: implications for targeting adenoma stem cells with NSAIDs. *Carcinogenesis*, 34, 1150- 1157.

-Almgren P, Lehtovirta M, Isomaa B, Sarelin L, Taskinen MR, Lyssenko V, Tuomi T, Groop, L; Botnia Study Group. (2011). Heritability and familiarity of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia*, 54, 2811–2819.

-Alsmadi O, Al-Rubeaan K, Wakil SM, Imtiaz F, Mohamed G, Al-Saud H, Al-Saud NA, Aldaghri N, Mohammad S, Meyer BF. (2008). Genetic study of Saudi diabetes (GSSD):

Bibliografía

significant association of the KCNJ11 E23K polymorphism with type 2 diabetes. *Diabetes Metab. Res. Rev.*, 24, 137-140.

- Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES. (2000). The common PPAR γ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genetics*, 26, 76–80.

- Alvarez Y, Valera I, Municio C, Hugo E, Padron F, Blanco L, Rodriguez M, Fernandez N, Crespo AS. (2010). Eicosanoids in the innate immune response: TLR and non-TLR routes. *Mediators Inflamm*, pii: 201929.

- American Diabetes Association (2006). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 29, S43–S48.

- Anand A, Chada K. (2000). *In vivo* modulation of *Hmgic* reduces obesity. *Nat. Genet.*, 24, 377–380.

- Assan R, Perronne C, Assan D, Chotard L, Mayaud C, Matheron S, Zucman D. (1995). Pentamidine-induced derangements of glucose homeostasis. *Diabetes Care*, 18, 47–55.

- Arroyo C, Hu FB, Ryan LM, Kawachi I, Colditz GA, Speizer FE, Manson J. (2004). Depressive symptoms and risk of type 2 diabetes in women. *Diabetes Care*, 27, 129–133.

- Avruch J. (1998). Insulin signal transduction through protein kinase cascades. *Mol Cell Biochem*, 182, 31-48.

B:

- Babenko AP, Polak M, Cavé H, Busiah K, Czernichow P, Scharfmann R, Bryan J, Aguilar-Bryan L, Vaxillaire M, Froguel P. (2006). Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *N Engl J Med*, 355, 456–466.

- Bajaj M, Suraamornkul S, Pratipanawatr T, Hardies LJ, Pratipanawatr W, Glass L, Cersosimo E, Miyazaki Y, DeFronzo RA. (2003). Pioglitazone reduces hepatic fat content and augments splanchnic glucose uptake in patients with type 2 diabetes. *Diabetes*, 52, 1364–1370.

- Bali D, Svetlanov A, Lee HW, Fusco-DeMane D, Leiser M, Li B, Barzilai N, Surana M, Hou H, Fleischer N, DePinho R, Rossetti L, Efrat S. (1995). Animal model for maturity-onset diabetes of the young generated by disruption of the mouse glucokinase gene. *J. Biol. Chem.* 270, 21464–21467.
- Barker N, Clevers H. (2010). Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. *Gastroenterology*, 138, 1681-1696.
- Barrett TG. (2001). Mitochondrial diabetes, DIDMOAD and other inherited diabetes syndromes. *Best Pract Res Clin Endocrinol Metab*, 15, 325-43.
- Bauer S, Hartmann G, Goodridge HS, Underhill DM. Fungal recognition by TLR2 and Dectin-1. In Toll-Like Receptors (TLRs) and Innate Immunity. Vol 1, 1st edition. Berlin, Heidelberg: Springer, 2008.
- Becker L, Huang Q, Mashimo H. (2010). Lgr5, an intestinal stem cell marker, is abnormally expressed in Barrett's esophagus and esophageal adenocarcinoma. *Dis Esophagus*, 23, 168-174.
- Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, Orho-Melander M, Gloyn AL. The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. (2009). *Hum Mol Genet*, 18, 4081-4088.
- Bell GI, Pilkis SJ, Weber IT, Polonsky KS. (1996). Glucokinase mutations, insulin secretion, and diabetes mellitus. *Annu. Rev. Physiol.* 58, 171–186.
- Bennett ST, Todd JA. (1996). Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu Rev Genet*, 30, 343–370.
- Bennett K, James C, Mutair A, Al-Shaikh H, Sinani A, Hussain K. (2011). Four novel cases of permanent neonatal diabetes mellitus caused by homozygous mutations in the glucokinase gene. *Pediatr Diabetes*, 12, 192-196.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53–59.

Bibliografía

- Berelowitz M & Eugene HG. Non-insulin dependent diabetes mellitus secondary to other endocrine disorders. In Diabetes Mellitus. LeRoith D, Taylor SI, Olefsky JM, Eds. New York, Lippincott-Raven, 1996, p. 496–502.
- Bergman RN. (1989). Toward physiological understanding of glucose tolerance. Minimal-model approach. *Diabetes*, 38, 1512–1527.
- Biemans-Oldehinkel E, Doeven MK, Poolman B. (2006). ABC transporter architecture and regulatory roles of accessory domains. *FEBS Letters*, 580, 1023–1035.
- Billings LK, Florez JC. (2010). The genetics of type 2 diabetes: what have we learned from GWAS? *Ann N Y Acad Sci*, 1212, 59-77.
- Boden G. (1997). Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Diabetes*, 46, 3–10.
- Boden G, Shulman GI. (2002). Free fatty acids in obesity and type 2 diabetes: defining their role in the development of insulin resistance and b-cell dysfunction. *Eur J Clin Invest*, 32, 14–23.
- Bodmer W, Bonilla C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*, 40, 695-701.
- Boesgaard TW, Grarup N, Jorgensen T, Borch-Johnsen K, Hansen T, Pedersen O. (2010). Variants at DGKB/TMEM195, ADRA2A, GLIS3 and C2CD4B loci are associated with reduced glucosestimulated beta cell function in middle-aged Danish people. *Diabetologia*, 53, 1647–1655.
- Boj SF, van Es JH, Huch M, Li VS, José A, Hatzis P, Mokry M, Haegebarth A, van den Born M, Chambon P, Voshol P, Dor Y, Cuppen E, Fillat C, Clevers H. (2012). Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell*, 151, 1595–1607.
- Boland JF, Chung CC, Roberson D, Mitchell J, Zhang X, Im KM, He J, Chanock SJ, Yeager M, Dean M. (2013). The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Hum Genet*, 132, 1153-1163.
- Bonadonna RC, Stumvoll M, Fritsche A, Muggeo M, Häring H, Bonora E, van Haefen TW. (2003). Altered homeostatic adaptation of first- and second-phase beta-cell

secretion in the offspring of patients with type 2 diabetes: studies with a minimal model to assess beta-cell function. *Diabetes*, 52, 470–480.

- Bonnefond A, Froguel P, Vaxillaire M. (2010). The emerging genetics of type 2 diabetes. *Trends Mol Med*, 16, 407-416.

- Botstein D, Risch N. (2003). Discovering genotypes underlying human phenotypes. *Nat Genet*, 33, 228-237.

- Bouças AP, Oliveira Fdos S, Canani LH, Crispim D. (2013). The role of interferon induced with helicase C domain 1 (IFIH1) in the development of type 1 diabetes mellitus. *Arq Bras Endocrinol Metabol*, 57, 667-676.

- Bouchard P, Sai P, Reach G, Caubarrere I, Ganeval D, Assan R. (1982). Diabetes mellitus following pentamidine-induced hypoglycemia in humans. *Diabetes*, 31, 40–45.

- Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW, Jarosz M, Krzymanska-Olejniak E, Kung L, Lipson D, Lowman GM, Marappan S, McInerney P, Platt A, Roy A, Siddiqi SM, Steinmann K, Thompson JF. (2009). Virtual terminator nucleotides for next-generation DNA sequencing. *Nat Methods*, 6, 593–595.

- Bowman P, Flanagan SE, Edghill EL, Damhuis A, Shepherd MH, Paisey R, Hattersley AT, Ellard S. (2012). Heterozygous ABCC8 mutations are a cause of MODY. *Diabetologia*, 55, 123–127.

- Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, Frackelton EC, Chiavacci RM, Imielinski M, Monos DS, Pandey R, Bakay M, Grant SF, Polychronakos C, Hakonarson H. (2011). A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet*, 7, e1002293.

- Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wigginton M, Schloss JA. (2008). The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 26, 1146–1153.

- Brants JR, Ayoubi TA, Chada K, Marchal K, Van de Ven WJ, Petit MM. (2004). Differential regulation of the insulin-like growth factor II mRNA binding protein genes by architectural transcription factor HMGA2. *FEBS Lett*, 569, 277–283.

Bibliografía

-Brown JB, Vistisen D, Sicree R, Shaw J, Nichols G, Zhang P. The economic impacts of diabetes. In: Gan D, editor. *Diabetes Atlas*. 3. Brussels: International Diabetes Federation; 2006.

-Brown GD, Gordon S. (2001). Immune recognition: a new receptor for [beta]-glucans. *Nature*, 413, 36-37.

-Brown GD. (2006). Dectin-1: a signalling non-TLR pattern-recognition receptor. *Nat Rev Immunol*, 6, 33–43.

-Brun T, Gauthier BR. (2008). A focus on the role of Pax4 in mature pancreatic islet beta-cell expansion and survival in health and disease. *J. Mol. Endocrinol.* 40, 37-45.

-Bu L, Salto LM, De Leon KJ, De Leon M. (2011). Polymorphisms in fatty acid binding protein 5 show association with type 2 diabetes. *Diabetes Res Clin Pract*, 92, 82-91.

-Buchanan TA, Xiang AH. (2005) Gestational diabetes mellitus. *The Journal of clinical investigation*, 115, 485–491.

-Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, 661–678.

-Cantrell Stanford J , Morris AJ, Sunkara M, Popa GJ, Larson KL, Özcan S. (2012). Sphingosine 1-phosphate (S1P) regulates glucose-stimulated insulin secretion in pancreatic beta cells. *J Biol Chem*, 287, 13457-13464.

C:

-Cauchi S, Meyre D, Dina C, Choquet H, Samson C, Gallina S, Balkau B, Charpentier G, Pattou F, Stetsyuk V, Scharfmann R, Staels B, Frühbeck G, Froguel P. (2006). Transcription factor TCF7L2 genetic study in the French population: expression in human β -cells and adipose tissue and strong association with type 2 diabetes. *Diabetes*, 55, 2903–2908.

-Cauchi S, El Achhab Y, Choquet H, Dina C, Krempler F, Weitgasser R, Nejjari C, Patsch W, Chikri M, Meyre D, Froguel P. (2007). TCF7L2 is reproducibly associated with type 2

diabetes in various ethnic groups: a global meta-analysis. *Journal of Molecular Medicine*, 85, 777–782.

-Cersosimo E, Pister PWT, Pesola G, McDermott K, Bajorunas D, Brennan MF. (1991). Insulin secretion and action in patients with pancreatic cancer. *Cancer*, 67, 486–493.

-Cervin C, Lyssenko V, Bakhtadze E, Lindholm E, Nilsson P, Tuomi T, Cilio CM, Groop L. (2008). Genetic similarities between latent autoimmune diabetes in adults, type 1 diabetes, and type 2 diabetes. *Diabetes*, 57, 1433-1437.

-Chan KH, Song Y, Hsu YH, You NC, F Tinker L, Liu S. (2010). Common genetic variants in fatty acid-binding protein-4 (FABP4) and clinical diabetes risk in the Women's Health Initiative Observational Study. *Obesity (Silver Spring)*, 18, 1812-1820.

-Cheng G, Zhong J, Chisari FV. (2006). Inhibition of dsRNA-induced signaling in hepatitis C virus-infected cells by NS3 protease-dependent and -independent mechanisms. *Proc. Natl. Acad. Sci. U. S. A*, 103, 8499–8504.

-Chilamakuri CS, Lorenz S, Madoui MA, Vodák D, Sun J, Hovig E, Myklebost O, Meza-Zepeda LA. (2014). Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*, 15, 449.

-Cho YM, Kim TH, Lim S, Choi SH, Shin HD, Lee HK, Park KS, Jang HC. (2009). Type 2 diabetes-associated genetic variants discovered in the recent genome-wide association studies are related to gestational diabetes mellitus in the Korean population. *Diabetologia*, 52, 253–261.

-Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U.S.A*, 106, 19096–19101.

-Choi SE, Liu M, Palaniappan LP, Wang EJ, Wong ND. (2013). Gender and ethnic differences in the prevalence of type 2 diabetes among Asian subgroups in California. *J Diabetes Complications*, 27, 429-435.

-Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M. (2011). Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol*, 29, 908-914.

Bibliografía

-Cleyneen I, Brants JR, Peeters K, Deckers R, Debiec-Rychter M, Sciote R, Van de Ven WJ, Petit MM. (2007). HMGA2 regulates transcription of the *Imp2* gene via an intronic regulatory element in cooperation with nuclear factor-kappaB. *Mol. Cancer Res*, 5, 363–372.

-Colon H, Tumors OP, Mcclanahan T, Koseoglu S, Smith K, Grein J, Gustafson E, Black S, Kirschmeier P, Samatar AA. (2006). Identification of Overexpression of Orphan G Protein-Coupled Receptor GPR49 in Human Colon and Ovarian Primary Tumors. *Cancer biology & therapy*, 5, 419-426.

-Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, MacCluer JW, Blangero J. (1997). A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat Genet*, 15, 273–276.

-Conn JW. (1965). Hypertension, the potassium ion and impaired carbohydrate tolerance. *N Engl J Med*, 273, 1135–1143.

-Cowie CC, Rust KF, Byrd-Holt DD, Gregg EW, Ford ES, Geiss LS, Bainbridge KE, Fradkin JE. (2010). Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988–2006. *Diabetes Care*, 33, 562–568.

-Crispim D, Estivalet AA, Roisenberg I, Gross JL, Canani LH. (2008). Prevalence of 15 mitochondrial DNA mutations among type 2 diabetic patients with or without clinical characteristics of maternally inherited diabetes and deafness. *Arq Bras Endocrinol Metab*, 52, 1228–1235.

D:

-Dabelea D, Snell-Bergeon JK, Hartsfield CL, Bischoff KJ, Hamman RF, McDuffie RS; Kaiser Permanente of Colorado GDM Screening Program. (2005). Increasing prevalence of gestational diabetes mellitus (GDM) over time and by birth cohort: Kaiser Permanente of Colorado GDM Screening Program. *Diabetes care*, 28, 579–584.

-Danforth E Jr. (2000). Failure of adipocyte differentiation causes type II diabetes mellitus? *Nat Genet*, 26, 13.

-Da Silva Xavier G, Mondragon A, Sun G, Chen L, McGinty JA, French PM, Rutter GA. (2012). Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice. *Diabetologia*, 55, 2667–2676.

- Davies MJ, Heller S, Skinner TC, Campbell MJ, Carey ME, Cradock S, Dallosso HM, Daly H, Doherty Y, Eaton S, Fox C, Oliver L, Rantell K, Rayman G, Khunti K; Diabetes Education and Self Management for Ongoing and Newly Diagnosed Collaborative. (2008). Effectiveness of the diabetes education and self management for ongoing and newly diagnosed (DESMOND) programme for people with newly diagnosed type 2 diabetes: cluster randomised controlled trial. *BMJ*, 336, 491–495.
- Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, Pellikka N, Wallaschofski H, Kettunen J, Henneman P, et al. (2011). Meta-Analysis of Genome- Wide Association Studies in > 80 000 Subjects Identifies Multiple Loci for C-Reactive Protein Levels. *Circulation*, 123, 731-738.
- De la Iglesia N, Veiga-da-Cunha M, Van Schaftingen E, Guinovart JJ, Ferrer JC. (1999). Glucokinase regulatory protein is essential for the proper subcellular localisation of liver glucokinase. *FEBS Lett*, 456, 332–338.
- Dennehy KM, Willment JA, Williams DL, Brown GD. (2009). Reciprocal regulation of IL-23 and IL-12 following co-activation of Dectin-1 and TLR signaling pathways. *Eur J Immunol*, 39, 1379-1386.
- Desai M, Zeggini E, Horton VA, Owen KR, Hattersley AT, Levy JC, Hitman GA, Walker M, Holman RR, McCarthy MI, Clark A. (2006). The variable number of tandem repeats upstream of the insulin gene is a susceptibility locus for latent autoimmune diabetes in adults. *Diabetes*, 55, 1890 –1894.
- Desai M, Zeggini E, Horton VA, Owen KR, Hattersley AT, Levy JC, Walker M, Gillespie KM, Bingley PJ, Hitman GA, Holman RR, McCarthy MI, Clark A. (2007). An association analysis of the HLA gene region in latent autoimmune diabetes in adults. *Diabetologia*, 50, 68 –73.
- d'Hennezel E, Bin Dhuban K, Torgerson T, Piccirillo CA. (2012). The immunogenetics of immune dysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome. *J Med Genet*, 49, 291-302.
- Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, Roix JJ et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316, 1331–1336.
- Di Cianni G, Miccoli R, Volpe L, Lencioni C, Del Prato S. (2003). Intermediate metabolism in normal pregnancy and in gestational diabetes. *Diabetes Metab Res Rev*, 19, 259–270.

Bibliografía

-Diez Roux AV, Jacobs DR, Kiefe CI. (2002). Neighborhood characteristics and components of the insulin resistance syndrome in young adults: the coronary artery risk development in young adults (CARDIA) study. *Diabetes Care*, 25, 1976–1982.

-Dinneen S, Gerich J, Rizza R. (1992). Carbohydrate metabolism in noninsulin-dependent diabetes mellitus. *N Engl J Med*, 327, 707–713.

-Dörner G, Plagemann A, Reinagel H. (1987). Familial diabetes aggregation in type I diabetics: gestational diabetes an apparent risk factor for increased diabetes susceptibility in the offspring. *Exp Clin Endocrinol*, 89, 84–90.

-Drazen JM, Yandava CN, Dubé L, Szczerback N, Hippensteel R, Pillari A, Israel E, Schork N, Silverman ES, Katz DA, Drajesk J. (1999). Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat Genet*, 22, 168-170.

-Dressler, GR. (2011). Patterning and early cell lineage decisions in the developing kidney: the role of Pax genes. *Pediatr. Nephrol*, 26, 1387-1394.

-Drummond RA, Brown GD. (2011). The role of Dectin-1 in the host defence against fungal infections. *Curr Opin Microbiol*, 14, 392–399.

-Duluc D, Joo H, Ni L, Yin W, Upchurch K, Li D, Xue Y, Klucar P, Zurawski S, Zurawski G, Oh S. (2014). Induction and activation of human Th17 by targeting antigens to dendritic cells via dectin-1. *J Immunol*, 192, 5776-88.

-Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics*, 42, 105–116.

E:

-Eaton WW, Armenian H, Gallo J, Pratt L, Ford DE. (1996). Depression and risk for onset of type II diabetes: a prospective population-based study. *Diabetes Care*, 19, 1097–1102.

-Egede LE, Dagogo-Jack S. (2005). Epidemiology of type 2 diabetes: focus on ethnic minorities. *Med Clin North Am*, 89, 949–975.

-Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*, 11, 446-450.

-Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133–138.

-Eizirik DL, Colli ML, Ortis F. (2009). The role of inflammation in insulinitis and beta-cell loss in type 1 diabetes. *Nat Rev Endocrinol*, 5, 219-226.

-Ekholm E, Gottsater A, Dahlin LB, Sundkvist G. (2012). No signs of progressive beta cell damage during 20 years of prospective follow-up of autoantibody-negative diabetes. *Acta Diabetol*, 49, 57–62.

-Ellisen LW, Bird J, West DC, Soreng AL, Reynolds TC, Smith SD, Sklar J. (1991). TAN-1, the human homolog of the *Drosophila* notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell*, 66, 649–661.

-Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, Erdmann J, Braund P, Engert JC, Bennett D, Coin L, Ashby D, Tzoulaki I, Brown IJ, Mt-Isa S, McCarthy MI, Peltonen L, Freimer NB, Farrall M, Ruukonen A, Hamsten A, Lim N, Froguel P, Waterworth DM, Vollenweider P, Waeber G, Jarvelin MR, Mooser V, Scott J, Hall AS, Schunkert H, Anand SS, Collins R, Samani NJ, Watkins H, Kooner JS. (2009). Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA*, 302, 37-48.

-Engelman JA, Luo J, Cantley LC. (2006). The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat Rev Genet*, 7, 606-619.

-Erdmann J, Grosshennig A, Braund PS, König IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Trégouët DA, et al. (2009). New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet*, 41, 280-282.

Bibliografía

-Eriksson KF, Lindgärde F. (1991). Prevention of type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise, the 6-year Malmö feasibility study. *Diabetologia*, 34, 891–898.

-Esposti MD, Ngo A, Myers MA. (1996). Inhibition of mitochondrial complex I may account for IDDM induced by intoxication with rodenticide Vacor. *Diabetes*, 45, 1531–1534.

F:

-Fabris P, Betterle C, Floreani A, Greggio NA, de Lazzari F, Naccarato R, Chiaramonte M. (1992). Development of type 1 diabetes mellitus during interferon alpha therapy for chronic HCV hepatitis. *Lancet*, 340, 548.

-Fajans SS, Bell GI. (2011). MODY: history, genetics, pathophysiology, and clinical decision making. *Diabetes Care*, 34, 1878–1884.

-Fallucca F, Dalfrà MG, Sciuillo E, Masin M, Buongiorno AM, Napoli A, Fedele D, Lapolla A. (2006). Polymorphisms of insulin receptor substrate 1 and b3-adrenergic receptor genes in gestational diabetes and normal pregnancy. *Metabolism*, 55, 1451–1456.

-Farhy C, Elgart M, Shapira Z, Oron-Karni V, Yaron O, Menuchin Y, Rechavi G, Ashery-Padan R. (2013). Pax6 is required for normal cell-cycle exit and the differentiation kinetics of retinal progenitor cells. *PLoS ONE*, 8, e76489.

-Feingold E. (2001). Methods for Linkage Analysis of Quantitative Trait Loci in Humans. *Theoretical Population Biology*, 60, 167–180.

-Festa A, Krugluger W, Shnawa N, Hopmeier P, Haffner SM, Schernthaner G. (1999). Trp64Arg polymorphism of the b3-adrenergic receptor gene in pregnancy: association with mild gestational diabetes mellitus. *J Clin Endocrinol Metab*, 84, 1695–1699.

-Filippi CM, von Herrath MG. (2008). Viral trigger for type 1 diabetes: pros and cons. *Diabetes*, 57, 2863-2871.

-Fisher E, Grallert H, Klapper M, Pfäfflin A, Schrezenmeir J, Illig T, Boeing H, Döring F. (2009). Evidence for the Thr79Met polymorphism of the ileal fatty acid binding protein (FABP6) to be associated with type 2 diabetes in obese individuals. *Mol Genet Metab*, 98, 400-405.

- Flanagan SE, Patch AM, Mackay DJ, Edghill EL, Gloyn AL, Robinson D, Shield JP, Temple K, Ellard S, Hattersley AT. (2007). Mutations in ATP-sensitive K⁺ channel genes cause transient neonatal diabetes and permanent diabetes in childhood or adulthood. *Diabetes*, 56, 1930–1937.
- Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burt NP, Mahajan A, Fuchsberger C, Atzmon G, Benediktsson R, et al. (2014). Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet*, 46, 357-463.
- Florez JC. (2008). Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008. *J Clin Endocrinol Metab*, 93, 4633-4642.
- Forrest, JA, Menser MA, Burgess JA. (1971). High frequency of diabetes mellitus in young patients with congenital rubella. *Lancet*, 2, 332-334.
- Frayling TM. (2007). Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet*, 8, 657-662.
- Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, 316, 889–894.
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res*, 32, W273-W279.
- Fryer LG, Parbu-Patel A, Carling D. (2002). The anti-diabetic drugs rosiglitazone and metformin stimulate AMP-activated protein kinase through distinct signaling pathways. *J Biol Chem*, 277, 25226–25232.
- Fryirs MA, Barter PJ, Appavoo M, Tuch BE, Tabet F, Heather AK, Rye KA. (2010). Effects of high-density lipoproteins on pancreatic beta-cell insulin secretion. *Arterioscler Thromb Vasc Biol*, 30, 1642-1648.

G:

-Gabbe SG. (1986). Definition, detection, and management of gestational diabetes. *Obstetrics and gynecology*, 67, 121–125.

-Gale EA. (2005). Latent autoimmune diabetes in adults: a guide for the perplexed. *Diabetologia*, 48, 2195–2199.

-Gallanosa AG, Spyker DA, Curnow RT. (1981). Diabetes mellitus associated with autonomic and peripheral neuropathy after Vacor poisoning: a review. *Clin Toxicol*, 18, 441–449.

-Gantner BN, Simmons RM, Canavera SJ, Akira S, Underhill DM. (2003). Collaborative induction of inflammatory responses by Dectin- 1 and toll-like receptor 2. *J Exp Med*, 197, 1107-1117.

-Gasa R, Gomis R, Casamitjana R, Rivera F, Novials A. (1997). Glucose regulation of islet amyloid polypeptide gene expression in rat pancreatic islets. *Am J Physiol*, 272, E543-E549.

-Gerich JE. (1998). The genetic basis of type 2 diabetes mellitus: impaired insulin secretion versus impaired insulin sensitivity. *Endocr Rev*, 19, 491–503.

-Gilles A, Megléc E, Pech N, Ferreira S, Malausa T, Martin JF. (2011). Accuracy and quality assessment of 454 GS-FLX titanium pyrosequencing. *BMC Genomics*, 12, 245.

-Gillies CL, Abrams KR, Lambert PC, Cooper NJ, Sutton AJ, Hsu RT, Khunti K. (2007). Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *BMJ*, 334, 299–302.

-Gillman MW, Rifas-Shiman S, Berkey CS, Field AE, Colditz GA. (2003). Maternal gestational diabetes, birth weight, and adolescent obesity. *Pediatrics*, 111, e221–226.

-Gloyn AL, Hashim Y, Ashcroft SJ, Ashfield R, Wiltshire S, Turner RC; UK Prospective Diabetes Study (UKPDS 53). (2001). Association studies of variants in promoter and coding regions of beta-cell ATP-sensitive K-channel genes SUR1 and Kir6.2 with Type 2 diabetes mellitus (UKPDS 53). *Diabetic Med*, 18, 206-212.

-Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S, McCarthy MI, Hattersley AT, Frayling TM. (2003). Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes*, 52, 568-572.

-Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, Howard N, Srinivasan S, Silva JM, Molnes J, Edghill EL, Frayling TM, Temple IK, Mackay D, Shield JP, Sumnik Z, van Rhijn A, Wales JK, Clark P, Gorman S, Aisenberg J, Ellard S, Njølstad PR, Ashcroft FM, Hattersley AT. (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med*, 350, 1838–1849.

-Goldstein BJ. Protein-tyrosine phosphatases and the regulation of insulin action. In: LeRoith D, Taylor SI, Olefsky JM, eds. *Diabetes mellitus: a fundamental and clinical text*. Philadelphia: Lippincott, 2003: 255–68.

-Goldstein BJ, Scalia R. (2004). Adiponectin: a novel adipokine linking adipocytes and vascular function. *J Clin Endocrinol Metab*, 89, 2563–68.

-Gong M, Long J, Liu Q, Deng HC. (2010). Association of the ADIPOQ rs17360539 and rs266729 polymorphisms with type 2 diabetes: a meta-analysis. *Mol Cell Endocrinol*, 325,78-83.

-Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. (2008). Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet*, 82, 100-112.

-Gosmain Y, Cheyssac C, Heddad Masson M, Dibner C, Philippe, J. (2011). Glucagon gene expression in the endocrine pancreas: the role of the transcription factor Pax6 in α -cell differentiation, glucagon biosynthesis and secretion. *Diabetes Obes. Metab*, 13, 31-38.

-Grandinetti A, Kaholokula JK, Chang HK. (2000). Delineating the relationship between stress, depressive symptoms, and glucose intolerance. *Diabetes Care*, 23, 1443–1444.

-Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadóttir A, Styrkarsdóttir U, Magnusson KP, Walters GB, Palsdóttir E, Jonsdóttir T, Gudmundsdóttir T, Gylfason A, Saemundsdóttir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdóttir U, Gulcher JR, Kong A, Stefansson K. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature Genetics*, 38, 320–323.

Bibliografía

-Grant SF, Hakonarson H, Schwartz S. (2010). Can the genetics of type 1 and type 2 diabetes shed light on the genetics of latent autoimmune diabetes in adults? *Endocr. Rev.*, 31, 183–193.

-Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA; 1000 Genomes Project, Bustamante CD. (2011). Demographic history and rare allele sharing among human populations. *Proc. Natl Acad. Sci. USA*, 108, 11983–11988.

-Groenewoud MJ, Dekker JM, Fritsche A, Reiling E, Nijpels G, Heine RJ, Maassen JA, Machicao F, Schäfer SA, Häring HU, 't Hart LM, van Haefen TW. (2008). Variants of *CDKAL1* and *IGF2BP2* affect first-phase insulin secretion during hyperglycaemic clamps. *Diabetologia*, 51, 1659–1663.

-Grupe A, Hultgren B, Ryan A, Ma YH, Bauer M, Stewart TA. (1995). Transgenic knockouts reveal a critical requirement for pancreatic beta cell glucokinase in maintaining glucose homeostasis. *Cell*, 83, 69-78.

-Gual P, Marchand-Brustel YL, Tanti JF. (2005). Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. *Biochimie*, 87, 99-109.

-Gupta PK. (2008). Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol*, 26, 602-611.

H:

-Handwerger S, Roth J, Gorden P, Di Sant' Agnese P, Carpenter DF, Peter G. (1969). Glucose intolerance in cystic fibrosis. *N Engl J Med*, 281, 451–461.

-Hansson O, Zhou Y, Renström E, Osmark P. (2010). Molecular function of TCF7L2: Consequences of TCF7L2 splicing for molecular function and risk for type 2 diabetes. *Current Diabetes Reports*, 10, 444–451.

-Hansen SK, Nielsen EM, Ek J, Andersen G, Glümer C, Carstensen B, Mouritzen P, Drivsholm T, Borch-Johnsen K, Jørgensen T, Hansen T, Pedersen O. (2005). Analysis of separate and combined effects of common variation in *KCNJ11* and *PPARG* on risk of type 2 diabetes. *J. Clin. Endocrinol. Metab*, 90, 3629-3637.

-Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu

Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu HQ, Polychronakos C. (2007). A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature*, 448, 591-594.

-Hara K, Fujita H, Johnson TA, Yamauchi T, Yasuda K, Horikoshi M, Peng C, Hu C, Ma RC, Imamura M, Iwata M, Tsunoda T, Morizono T, Shojima N, So WY, Leung TF, Kwan P, Zhang R, Wang J, Yu W, Maegawa H, Hirose H; DIAGRAM consortium, Kaku K, Ito C, Watada H, Tanaka Y, Tobe K, Kashiwagi A, Kawamori R, Jia W, Chan JC, Teo YY, Shyong TE, Kamatani N, Kubo M, Maeda S, Kadowaki T. (2014). Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet*, 23, 239-246.

-Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol*, 10, R32.

-Harris MI, Flegal KM, Cowie CC, Eberhardt MS, Goldstein DE, Little RR, Wiedmeyer HM, Byrd-Holt DD. (1998). Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988–1994. *Diabetes Care*, 21, 518–524.

-Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z. (2008). Single-molecule DNA sequencing of a viral genome. *Science*, 320, 106–109.

-Hartmann R, Olsen HS, Widder S, rgensen J, Justesen RJ. (1998). p59OASL, a 2′–5′ oligoadenylate synthetase like protein: a novel human gene related to the 2′–5′ oligoadenylate synthetase family. *Nucleic Acids Research*, 26, 4121.

-Hasstedt SJ, Chu WS, Das SK, Wang H, Elbein SC. (2008). Type 2 diabetes susceptibility genes on chromosome 1q21-24. *Ann Hum Genet*, 72, 163-169.

-Hayashi Y. Metabolic impact of glucagon deficiency. (2011). *Diabetes Obes Metab*, 13, 151-157.

-Hayward BE, Fantès JA, Warner JP. (1996). Co-localization of the ketohexokinase and glucokinase regulator genes to a 500-kb region of chromosome 2p23. *Mamm Genome*, 7, 454–458.

-Hedges DJ, Burges D, Powell E, Almonte C, Huang J, Young S, Boese B, Schmidt M, Pericak-Vance MA, Martin E, Zhang X, Harkins TT, Züchner S. (2009). Exome sequencing of a multigenerational human pedigree. *PLoS One*, 4, e8232.

Bibliografía

-Hemi R, Paz K, Wertheim N, Karasik A, Zick Y, Kanety H. (2002). Transactivation of ErbB2 and ErbB3 by tumor necrosis factor-alpha and anisomycin leads to impaired insulin signaling through serine/threonine phosphorylation of IRS proteins. *J Biol Chem*, 277, 8961-8969.

-Hertz RP, Unger AN, Ferrario CM. (2006). Diabetes, hypertension, and dyslipidemia in Mexican Americans and non-Hispanic whites. *American Journal of Preventive Medicine*, 30, 103–110.

-Hober D, Sauter P. (2010). Pathogenesis of type 1 diabetes mellitus: interplay between enterovirus and host. *Nat Rev Endocrinol*, 6, 279-289.

-Horikawa Y. (2000). Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genetics*, 26, 163–175.

-Hotamisligil GS. (2000). Molecular mechanisms of insulin resistance and the role of the adipocyte. *Int J Obes Relat Metab Disord*, 24, S23-7

-Horton V, Stratton I, Bottazzo GF, Shattock M, Mackay I, Zimmet P, Manley S, Holman R, Turner R. (1999). Genetic heterogeneity of autoimmune diabetes: age of presentation in adults is influenced by HLA DRB1 and DQB1 genotypes (UKPDS 43): UK Prospective Diabetes Study (UKPDS) Group. *Diabetologia*, 42, 608–616.

-Huang K, Nair AK, Muller YL, Piaggi P, Bian L, Del Rosario M, Knowler WC, Kobes S, Hanson RL, Bogardus C, Baier LJ. (2014). Whole exome sequencing identifies variation in CYB5A and RNF10 associated with adiposity and type 2 diabetes. *Obesity (Silver Spring)*. 22, 984-988.

-Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, Willett WC. (2001). Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med*, 345, 790–797.

-Hu He KH, Lorenzo PI, Brun T, Jimenez Moreno CM, Aeberhard D, Vallejo Ortega J, Cornu M, Thorel F, Gjinovci A, Thorens B, Herrera PL, Meda P, Wollheim CB, Gauthier BR. (2011). In vivo conditional Pax4 overexpression in mature islet β -cells prevents stress-induced hyperglycemia in mice. *Diabetes*, 60, 1705-1715.

-Hull RL, Westermarck GT, Westermarck P, Kahn SE. (2004). Islet amyloid: a critical entity in the pathogenesis of type 2 diabetes. *J Clin Endocrinol Metab*, 89, 3629–3643.

-Hynes RO. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. *Cell*, 69, 11-25.

I:

-Ishibashi M, Wakita T, Esumi M. (2010). 2',5'-Oligoadenylate synthetase-like gene highly induced by hepatitis C virus infection in human liver is inhibitory to viral replication in vitro. *Biochemical and Biophysical Research Communications*, 392, 397–402.

-lynedjian PB, Mobius G, Seitz HJ, Wollheim CB, Renold, AE. (1986). Tissue-specific expression of glucokinase: identification of the gene product in liver and pancreatic islets. *Proc. Natl. Acad. Sci. U. S. A.* 83, 1998–2001.

J:

-Jadresic A, Banks LM, Child DF, Diamant L, Doyle FH, Fraser TR, Joplin GF. (1982). The acromegaly syndrome. *Q J Med*, 202, 189–204.

-Jahromi MM, Eisenbarth GS. (2007). Cellular and molecular pathogenesis of type 1A diabetes. *Cell Mol Life Sci*, 64, 865-872.

-Jaidane H, Hober D. (2008). Role of coxsackievirus B4 in the pathogenesis of type 1 diabetes. *Diabetes Metab*, 34, 537-548.

-Jaks V, Barker N, Kasper M, van Es JH, Snippert HJ, Clevers H and Toftgård R. (2008). Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nature genetics*, 40, 1291-1299.

-Janson J, Ashley RH, Harrison D, McIntyre S, Butler PC. (1999). The mechanism of islet amyloid polypeptide toxicity is membrane disruption by intermediate-sized toxic amyloid particles. *Diabetes*, 48, 491–498.

-Jeon CY, Lokken RP, Hu FB, van Dam RM. (2007). Physical activity of moderate intensity and risk of type 2 diabetes: a systematic review. *Diabetes Care*, 30, 744–752.

-Jetton TL, Liang Y, Pettepher CC, Zimmerman EC, Cox FG, Horvath K, Matschinsky FM, Magnuson MA. (1994). Analysis of upstream glucokinase promoter activity in transgenic mice and identification of glucokinase in rare neuroendocrine cells in the brain and gut. *J. Biol. Chem.* 269, 3641–3654.

-Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, Schulick RD, Tang LH, Wolfgang CL, Choti MA, Velculescu VE, Diaz LA Jr, Vogelstein B, Kinzler KW, Hruban RH,

Bibliografía

Papadopoulos N. (2011). DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*, 331, 1199–1203.

-Jiménez-Sousa MÁ, Almansa R, de la Fuente C, Caro-Patón A, Ruiz L, Sánchez-Antolín G, González JM, Aller R, Alcaide N, Largo P, Ortiz de Lejarazu R, Bermejo-Martín JF. (2011). Gene expression profiling in the first twelve weeks of treatment in chronic hepatitis C patients. *Enferm Infecc Microbiol Clin*, 29, 573-580.

-Jonsson A, Ladenvall C, Ahluwalia TS, Kravic J, Krus U, Taneera J, Isomaa B, Tuomi T, Renström E, Groop L, Lyssenko V. (2013). Effects of common genetic variants associated with type 2 diabetes and glycemic traits on α - and β -cell function and insulin action in humans. *Diabetes*, 62, 2978-2983.

-Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, Alamowitch S, Domenga V, Cécillion M, Marechal E, Maciazek J, Vayssiere C, Cruaud C, Cabanis EA, Ruchoux MM, Weissenbach J, Bach JF, Bousser MG, Tournier-Lasserre E. (1996). Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature*, 383, 707–710.

K:

-Kahn SE, Hull RL, Utzschneider KM. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*, 444, 840-846.

-Kanatsuka A, Tokuyama Y, Nozaki O, Matsui K, Egashira T. (2002). Beta-cell dysfunction in late-onset diabetic subjects carrying homozygous mutation in transcription factors NeuroD1 and Pax4. *Metabolism*, 51, 1161–1165.

-Kantarova D, Buc M. (2007). Genetic susceptibility to type 1 diabetes mellitus in humans. *Physiol Res*, 56, 255–266.

-Kapoor R. (2010). Defining Genotype–Phenotype Correlations in Children With Congenital Hyperinsulinism. University College London (Doctoral dissertation).

-Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, Stengar J, Kesaniemi YA. (1992). Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia*, 35, 1060–1067.

- Karin M, Delhase M. (2000). The I kappa B kinase (IKK) and NF-kappa B: key elements of proinflammatory signalling. *Semin Immunol*, 12, 85–98.
- Karjalainen J, Knip M, Hyoty H, Linikki P, Ilonen J, Kaar M-L, Akerblom HK. (1988). Relationship between serum insulin antibodies, islet cell antibodies and Coxsackie-B4 and mumps virus-specific antibodies at the clinical manifestation of type I (insulin-dependent) diabetes. *Diabetologia*, 31, 146–152.
- Karnik SK, Chen H, McLean GW, Heit JJ, Gu X, Zhang AY, Fontaine M, Yen MH, Kim SK. (2007). Menin controls growth of pancreatic beta-cells in pregnant mice and promotes gestational diabetes mellitus. *Science*, 318, 806–809.
- Kaye SA, Folsom AR, Sprafka JM, Prineas RJ, Wallace RB. (1991). Increased incidence of diabetes mellitus in relation to abdominal adiposity in older women. *J Clin Epidemiol*, 44, 329–334.
- Kelley DE, He J, Menshikova EV, Ritov VB. (2002). Dysfunction of mitochondria in human skeletal muscle in type 2 diabetes. *Diabetes*, 51, 2944–2950.
- Kennedy AD, Willment JA, Dorward DW, Williams DL, Brown GD, Deleo FR. (2007). Dectin-1 promotes fungicidal activity of human neutrophils. *Eur J Immunol*, 37, 467-478.
- Kerrigan AM, Brown GD. (2010). Syk-coupled C-type lectin receptors that mediate cellular activation via single tyrosine based activation motifs. *Immunol Rev*, 234, 335–352.
- Kim JK, Kim YJ, Fillmore JJ, Chen Y, Moore I, Lee J, Yuan M, Li ZW, Karin M, Perret P, Shoelson SE, Shulman GI. (2001). Prevention of fat induced insulin resistance by salicylate. *J Clin Invest*, 108, 437–446.
- Kim C, Newton KM, Knopp RH. (2002). Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes care*, 25, 1862–1868.
- Kim H, Toyofuku Y, Lynn FC, Chak E, Uchida T, Mizukami H, Fujitani Y, Kawamori R, Miyatsuka T, Kosaka Y, Yang K, Honig G, van der Hart M, Kishimoto N, Wang J, Yagihashi S, Tecott LH, Watada H, German MS. (2010). Serotonin regulates pancreatic beta cell mass during pregnancy. *Nat Med*, 16, 804–808.
- Kim MS, Polychronakos C. (2005). Immunogenetics of type 1 diabetes. *Horm Res*, 64, 180-188.
- Kim SY, Lohmueller KE, Albrechtsen A, Li Y, Korneliussen T, Tian G, Grarup N, Jiang T, Andersen G, Witte D, Jorgensen T, Hansen T, Pedersen O, Wang J, Nielsen R. (2011).

Bibliografía

Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics*, 12, 231.

-King ML, Bidwell D, Shaikh A, Voller A, Banatvala JE. (1983). Coxsackie-B-virus-specific IgM responses in children with insulin-dependent (juvenile-onset; type I) diabetes mellitus. *Lancet*, 1, 1397-1399.

-King RA, Rotter JI, Motulsky AG. The Genetic Basis of Common Diseases. Vol1. 1st Ed. New York, NY: Oxford University Press; 2002.

-Knip M, Veijola R, Virtanen SM, Hyöty H, Vaarala O, Akerblom HK. (2005). Environmental triggers and determinants of type 1 diabetes. *Diabetes*, 54, S125-36.

-Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM; Diabetes Prevention Program Research Group. (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*, 346, 393–403.

-Knowles NG, Landchild MA, Fujimoto WY, Kahn SE. (2002). Insulin and amylin release are both diminished in first-degree relatives of subjects with type 2 diabetes. *Diabetes Care*, 25, 292–297.

-Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Jonasdottir A, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediksdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G; DIAGRAM Consortium, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. (2009). Parental origin of sequence variants associated with complex diseases. *Nature*, 462, 868–874.

-Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, Been LF, Chia KS, Dimas AS, Hassanali N, et al. (2011). Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet*, 43, 984-989.

-Kooptiwut S, Plengvidhya N, Chukijrungsroat T, Sujitjoo J, Semprasert N, Furuta H, Yenchitsomanus PT. (2012). Defective PAX4 R192H transcriptional repressor activities associated with maturity onset diabetes of the young and early onset-age of type 2 diabetes. *J. Diabetes Complications*, 26, 343-347.

-Kopelman PG. (2000). Obesity as a medical problem. *Nature*, 404, 635-643.

-Krawezak M, Reiss J, Cooper DN. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet*, 90, 41-54.

-Krebs DL, Hilton DJ. (2003). A new role for SOCS in insulin action. Suppressor of cytokine signaling. *Sci STKE*, 2003,PE6.

-Kristiansen H, Gad HH, Eskildsen-Larsen S, Despres P, Hartmann R. (2011). The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities. *Journal of Interferon and Cytokine Research*, 31, 41–47.

-Kumashiro N, Erion DM, Zhang D, Kahn M, Beddow SA, Chu X, Still CD, Gerhard GS, Han X, Dziura J, Petersen KF, Samuel VT, Shulman GI. (2011). Cellular mechanism of insulin resistance in nonalcoholic fatty liver disease. *Proc Natl Acad Sci U S A*, 108, 16381–16385.

-Kutikhin AG, Yuzhalin AE. (2012). C-type lectin receptors and RIG-I-like receptors: new points on the oncogenomics map. *Cancer Manag Res*, 4, 39-53.

L:

-Lammert E, Brown J, Melton DA. (2000). Notch gene expression during pancreatic organogenesis. *Mech. Dev*, 94, 199–203.

-Lang V, Youssef N, Light PE. (2011). The molecular genetics of sulfonyleurea receptors in the pathogenesis and treatment of insulin secretory disorders and type 2 diabetes. *Current Diabetes Reports*, 11, 543–551.

-Langhi C, Cariou B. (2010). Cholesterol metabolism and beta-cell function. *Med Sci (Paris)*, 26, 385-390.

-Lapolla A, Dalfrà MG, Fedele D. (2009). Diabetes related autoimmunity in gestational diabetes mellitus: is it important?. *Nutr Metab Cardiovasc Dis*, 19, 674–682.

-Larsen S, Hilsted J, Tronier B, Worning H. (1987). Metabolic control and B cell function in patients with insulin-dependent diabetes mellitus secondary to chronic pancreatitis. *Metabolism*, 36, 964–967.

Bibliografía

-Lauenborg J, Grarup N, Damm P, et al. (2009). Common type 2 diabetes risk gene variants associate with gestational diabetes. *J Clin Endocrinol Metab*, 94, 145–150.

-Lefebvre T, Dehennaut V, Guinez C, Olivier S, Drougat L, Mir AM, Mortuaire M, Vercoutter-Edouart AS, Michalski JC. (2010). Dysregulation of the nutrient/stress sensor O-GlcNAcylation is involved in the etiology of cardiovascular disorders, type-2 diabetes and Alzheimer's disease. *Biochim Biophys Acta*, 1800, 67-79.

-Lehne B, Lewis CM, Schlitt T. (2011). Exome localization of complex disease association signals. *BMC Genomics*, 12, 92.

-Leipold H, Knöfler M, Gruber C, Haslinger P, Bancher-Todesca D, Worda C. (2004). Calpain-10 haplotype combination and association with gestational diabetes mellitus. *Obstet Gynecol*, 103, 1235–1240.

-Leipold H, Knoefler M, Gruber C, Klein K, Haslinger P, Worda C. (2006). Plasminogen activator inhibitor 1 gene polymorphism and gestational diabetes mellitus. *Obstet Gynecol*, 107, 651–656.

-Le Roith D, Quon MJ, Zick Y. Molecular and cellular aspects of insulin resistance: Implications for diabetes. In: Signal Transduction and Human Disease. Editor: Finkel T, Gutkind JS. Hoboken, New Jersey Wiley-Interscience; 2003: pp 171-200.

-Li B, Leal SM. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*, 83, 311-321.

-Li H, Ruan J, Durbin R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*, 18, 1851-1858.

-Li B, Leal SM. (2009). Discovery of rare variants via sequencing: implications for the design of complex trait association studies. *PLoS Genet*, 5, e1000481.

-Li H, Durbin R. (2009a). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-1760.

-Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. (2009b). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-2079.

-Li H, Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26, 589-595.

- Li H, Xu R, Peng X, Wang Y, Wang T. (2013). Association of glucokinase regulatory protein polymorphism with type 2 diabetes and fasting plasma glucose: a meta-analysis. *Mol Biol Rep*, 40, 3935-3942.
- Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25, 1966-1967.
- Li S, Besenbacher S, Li Y, Kristiansen K, Grarup N, Albrechtsen A, Sparsø T, Korneliussen T, Hansen T, Wang J, Nielsen R, Pedersen O, Bolund L, Schierup MH. (2014). Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. *Eur J Hum Genet*, 22, 1040-1045.
- Li X, Yang Y, Hu Y, Dang D, Regezi J, Schmidt BL, Atakilit A, Chen B, EllisD, Ramos DM. (2003). Alphavbeta6-Fyn signaling promotes oral cancer progression. *J. Biol. Chem*, 278, 41646-41653.
- Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, Grarup N, Guo Y, Hellman I, Jin X, Li Q, Liu J, Liu X, Sparsø T, Tang M, Wu H, Wu R, Yu C, Zheng H, Astrup A, Bolund L, Holmkvist J, Jørgensen T, Kristiansen K, Schmitz O, Schwartz TW, Zhang X, Li R, Yang H, Wang J, Hansen T, Pedersen O, Nielsen R, Wang J. (2010). Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet*, 42, 969-972.
- Liang Y, Najafi H, Smith RM, Zimmerman EC, Magnuson, MA, Tal, M, Matschinsky, FM. (1992). Concordant glucose induction of glucokinase, glucose usage, and glucose-stimulated insulin release in pancreatic islets maintained in organ culture. *Diabetes*, 41, 792-806.
- Lin RJ, Yu HP, Chang BL, Tang WC, Liao CL, Lin YL. (2009). Distinct antiviral roles for human 2',5'-oligoadenylate synthetase family members against dengue virus infection. *J Immunol*, 183, 8035-8043.
- Lindström J, Louheranta A, Mannelin M, Rastas M, Salminen V, Eriksson J, Uusitupa M, Tuomilehto J; Finnish Diabetes Prevention Study Group. (2003). The Finnish diabetes prevention study (DPS), lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care*, 26, 3230-3236.
- Litou H, Anastasiou E, Thalassinou L, Sarika HL, Philippou G, Alevizaki M. (2007). Increased prevalence of VNTR III of the insulin gene in women with gestational diabetes mellitus (GDM). *Diabetes Res Clin Pract*, 76, 223-228.

Bibliografía

-Liu Z, Habener JF. (2008). Glucagon-like peptide-1 activation of TCF7L2-dependent Wnt signaling enhances pancreatic beta cell proliferation. *Journal of Biological Chemistry*, 283, 8723–8735.

-Logue J, Walker JJ, Colhoun HM, Leese GP, Lindsay RS, McKnight JA, Morris AD, Pearson DW, Petrie JR, Philip S, Wild SH, Sattar N; Scottish Diabetes Research Network Epidemiology Group. (2011). Do men develop type 2 diabetes at lower body mass indices than women?. *Diabetologia*, 54, 3003-3006.

-Lohmueller KE, Sparsø T, Li Q, Andersson E, Korneliussen T, Albrechtsen A, Banasik K, Grarup N, Hallgrimsdottir I, Kiil K, Kilpeläinen TO, Krarup NT, Pers TH, Sanchez G, Hu Y, Degiorgio M, Jørgensen T, Sandbæk A, Lauritzen T, Brunak S, Kristiansen K, Li Y, Hansen T, Wang J, Nielsen R, Pedersen O. (2013). Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *Am J Hum Genet*, 93, 1072-1086.

-Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol*, 30, 434–439.

-Lysenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P, Groop L. (2008). Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med*, 359, 2220-2232.

-Lysenko V, Nagorny CL, Erdos MR, Wierup N, Jonsson A, Spéjel P, Bugliani M, Saxena R, Fex M, Pulizzi N, Isomaa B, Tuomi T, Nilsson P, Kuusisto J, Tuomilehto J, Boehnke M, Altshuler D, Sundler F, Eriksson JG, Jackson AU, Laakso M, Marchetti P, Watanabe RM, Mulder H, Groop L. (2009). Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nature Genetics*, 41, 82–88.

M:

-Maa Bared S. (2005). Identification and Characterization of ABCA1-interactive Proteins and Their Relevance to Atherosclerosis (Doctoral dissertation).

-Maassen JA, 'T Hart LM, Van Essen E, Heine RJ, Nijpels G, Jahangir Tafrechi RS, Raap AK, Janssen GM, Lemkes HH. (2004). Mitochondrial diabetes: molecular mechanisms and clinical presentation. *Diabetes*, 53, S103–S109.

- Maassen JA, Jahangir Tafrechi RS, Janssen GM, Raap AK, Lemkes HH, 't Hart LM. (2006). New insights in the molecular pathogenesis of the maternally inherited diabetes and deafness syndrome. *Endocrinol Metab Clin North Am*, 35, 385–396.
- Maechler P, Wollheim CB. (2001). Mitochondrial function in normal and diabetic beta-cells. *Nature*, 414, 807–812.
- Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, et al. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*, 46, 234-244.
- Maida A, Lamont BJ, Cao X, Drucker DJ. (2011). Metformin regulates the incretin receptor axis via a pathway dependent on peroxisome proliferator-activated receptor- α in mice. *Diabetologia*, 54, 339-349.
- Malfatti E, Bugiani M, Invernizzi F, de Souza CF, Farina L, Carrara F, Lamantea E, Antozzi C, Confalonieri P, Sanseverino MT, Giugliani R, Uziel G, Zeviani M. (2007). Novel mutations of ND genes in complex I deficiency associated with mitochondrial encephalopathy. *Brain*, 130, 1894–1904.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. (2009). Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.
- Mansego ML, Martínez F, Martínez-Larrad MT, Zabena C, Rojo G, Morcillo S, Soriguer F, Martín-Escudero JC, Serrano-Ríos M, Redon J, Chaves FJ. (2012). Common variants of the liver fatty acid binding protein gene influence the risk of type 2 diabetes and insulin resistance in Spanish population. *PLoS One*, 7, e31853.
- Marcovecchio M, Mohn A, Chiarelli F. (2005). Type 2 diabetes mellitus in children and adolescents. *Journal of Endocrinological Investigation*, 28, 853–863.

Bibliografía

-Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376–380.

-Mariotti C, Tiranti V, Carrara F, Dallapiccola B, DiDonato S, Zeviani M. (1994). Defective respiratory capacity and mitochondrial protein synthesis in transformant cybrids harboring the tRNA(Leu(UUR)) mutation associated with maternally inherited myopathy and cardiomyopathy. *J Clin Invest*, 93, 1102–1107.

-Marquardt T, Ashery-Padan R, Andrejewski N, Scardigli R, Guillemot F, Gruss P. (2001). Pax6 is required for the multipotent state of retinal progenitor cells. *Cell*, 105, 43-55.

-Marques J, Anwar J, Eskildsen-Larsen S, Rebouillat D, Paludan SR, Sen G, Williams BR, Hartmann R. (2008). The p59 oligoadenylate synthetase-like protein possesses antiviral activity that requires the C-terminal ubiquitin-like domain. *Journal of General Virology*, 89, 2767–2772.

-Marselli L, Thorne J, Dahiya S, Sgroi DC, Sharma A, Bonner-Weir S, Marchetti P, Weir GC. (2010). Gene expression profiles of Beta-cell enriched tissue obtained by laser capture microdissection from subjects with type 2 diabetes. *PLoS One*, 5, e11499.

-Mathieson I, McVean G. (2012). Differential confounding of rare and common variants in spatially structured populations. *Nature Genet*, 44, 243–246.

-Matschinsky FM, Glaser B, Magnuson MA. (1998). Pancreatic beta-cell glucokinase: closing the gap between theoretical concepts and experimental realities. *Diabetes*, 47, 307–315.

-Matsuda A, Kuzuya T. (1994). Relationship between obesity and concordance rate for type 2 diabetes mellitus among twins. *Diabetes Res Clin Pract*, 26, 137-143.

-Mauvais-Jarvis F, Smith SB, Le May C, Leal SM, Gautier JF, Molokhia M, Riveline JP, Rajan AS, Kevorkian JP, Zhang S, Vexiau P, German MS, Vaisse C. (2004). PAX4 gene variations predispose to ketosis-prone diabetes. *Hum. Mol. Genet*, 13, 3151-3159.

- McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, 26, 2069-2070.
- Medici F, Hawa M, Ianari A, Pyke DA, Leslie RD. (1999). Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia*, 42, 146-150.
- Megia A, Gallart L, Fernández-Real JM, Vendrell J, Simón I, Gutierrez C, Richart C. (2004). Mannose-binding lectin gene polymorphisms are associated with gestational diabetes mellitus. *J Clin Endocrinol Metab*, 89, 5081–5087.
- Merriman B, Ion Torrent R&D Team, Rothberg JM. (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33, 3397-3417.
- Meyer C, Stumvoll M, Nadkarni V, Dostou J, Mitrakou A, Gerich J. (1998). Abnormal renal and hepatic glucose metabolism in type 2 diabetes mellitus. *J Clin Invest*, 102, 619–624.
- Mbanya JCN, Motala AA, Sobngwi E, Assah FK, Enoru ST. (2010). Diabetes in Sub-Saharan Africa. *Lancet*, 375, 2254–2266.
- Middelberg RP, Ferreira MA, Henders AK, Heath AC, Madden PA, Montgomery GW, Martin NG, Whitfield JB. (2011). Genetic variants in LPL, OASL and TOMM40/APOE-C1-C2-C4 genes are associated with multiple cardiovascular-related traits. *BMC Med Genet*, 12, 123.
- Miklossy J, Qing H, Radenovic A, Kis A, Vilenó B, László F, Miller L, Martins RN, Waeber G, Mooser V, Bosman F, Khalili K, Darbinian N, McGeer PL. (2010). Beta amyloid and hyperphosphorylated tau deposits in the pancreas in type 2 diabetes. *Neurobiol Aging*, 31, 1503-1515.
- Milos PM. (2009). Emergence of single-molecule sequencing and potential for molecular diagnostic applications. *Expert Rev. Mol. Diagn*, 9, 659–666.
- Mitra RD, Church GM. (1999). In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res*, 27, e34–e39.
- Møller HJ. (2012). Soluble CD163. *Scand J Clin Lab Invest*, 72, 1-13.
- Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliusen TS, Andersen MA, Nielsen TS, Krarup NT, Gjesing AP, Zierath JR, Linneberg A, Wu X, Sun G, Jin X, Al-Aama J, Wang J, Borch-Johnsen K, Pedersen O, Nielsen R, Albrechtsen A,

Bibliografía

Hansen T. (2014). A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*. 512, 190-193.

-Molven A, Njølstad PR. (2011). Role of molecular genetics in transforming diagnosis of diabetes mellitus. *Expert Rev Mol Diagn*, 11, 313-320.

-Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. (2003). PGC-1 alpha responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, 34, 267–273.

-Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet*, 44, 981–990.

-Murphy R, Turnbull DM, Walker M, Hattersley AT. (2008). Clinical features, diagnosis and management of maternally inherited diabetes and deafness (MIDD) associated with the 3243A>G mitochondrial point mutation. *Diabet Med*, 25, 383–399.

-Murphy R, Ellard S, Hattersley AT. (2008). Clinical implications of a molecular genetic classification of monogenic β -cell diabetes. *Nat Clin Pract Endocrinol Metab*, 4, 200–213.

-Myers MG Jr, White MF. The Molecular Basis of Insulin Action. In: *Insulin Signaling: From cultured cells to animal models*. Editor: Gruenberg G, Zick Y. Taylor and Francis, New York; 2002; pp 55-87.

N:

-Narayan KM, Boyle JP, Geiss LS, Saaddine JB, Thompson TJ. (2006). Impact of recent increase in incidence on future diabetes burden: U.S., 2005–2050. *Diabetes Care*, 29, 2114–2116.

-Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. (2011). Testing for an Unusual Distribution of Rare Variants. *PLoS Genet*, 7, e1001322.

-Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu SA, Fraser D, Warren L, Aponte J, Zawistowski M, Liu X, Zhang H, Zhang Y, Li J, Li Y, Li L, Woollard P, Topp S, Hall MD, Nangle K, Wang J, Abecasis G, Cardon LR, Zöllner S, Whittaker JC, Chisoe SL, Novembre J, Mooser V. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*, 337, 100–104.

-Netea MG, Gow NA, Munro CA, Bates S, Collins C, Ferwerda G, Hobson RP, Bertram G, Hughes HB, Jansen T, Jacobs L, Buurman ET, Gijzen K, Williams DL, Torensma R, McKinnon A, MacCallum DM, Odds FC, Van der Meer JW, Brown AJ, Kullberg BJ. (2006). Immune sensing of *Candida albicans* requires cooperative recognition of mannans and glucans by lectin and Toll-like receptors. *J Clin Invest*, 116, 1642-1650.

-Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD. (1987). Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia*, 30, 763-768.

-Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA, Strausberg RL, Venter JC. (2008). Genetic variation in an individual human exome. *PLoS Genet*, 4, e1000160.

-Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461, 272-276.

-Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 42, 30-35.

-Niswender KD, Shiota M, Postic C, Cherrington AD, Magnuson MA. (1997). Effects of increased glucokinase gene copy number on glucose homeostasis and hepatic glucose metabolism. *J. Biol. Chem.* 272, 22570–22575.

Bibliografía

-Niu J, Dorahy DJ, Gu X, Scott RJ, Draganic B, Ahmed N, Agrez MV. (2002). Integrin expression in colon cancer cells is regulated by the cytoplasmic domain of the beta6 integrin subunit. *Int. J. Cancer*, 99, 529-537.

-Novials A, Sarri Y, Casamitjana R, Rivera R, Gomis R. (1993). Regulation of islet amyloid polypeptide in human pancreatic islets. *Diabetes*, 42, 1514-1519.

O:

- Ober C, Xiang KS, Thisted RA, Indovina KA, Wason CJ, Dooley S. (1989). Increased risk for gestational diabetes mellitus associated with insulin receptor and insulin-like growth factor II restriction fragment length polymorphisms. *Genet Epidemiol*, 6, 559–569.

-O'Byrne S, Feely J. (1990). Effects of drugs on glucose tolerance in non-insulin-dependent diabetes (parts I and II). *Drugs*, 40, 203– 219.

-Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK, Fraenkel E, Bell GI, Young RA. (2004). Control of pancreas and liver gene expression by HNF transcription factors. *Science*, 303, 1378-1381.

-Okada Y, Takahashi A, Ohmiya H, Kumasaka N, Kamatani Y, Hosono N, Tsunoda T, Matsuda K, Tanaka T, Kubo M, Nakamura Y, Yamamoto K, Kamatani N. (2011). Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Hum Mol Genet*, 20, 1224-1231.

P:

-Pak CY, Eun H, McArthur RG, Yoon J. (1988). Association of cytomegalovirus infection with autoimmune type 1 diabetes. *Lancet*, 2, 1-4.

-Palmer ND, Goodarzi MO, Langefeld CD, Ziegler J, Norris JM, Haffner SM, Bryer-Ash M, Bergman RN, Wagenknecht LE, Taylor KD, Rotter JI, Bowden DW. (2008). Quantitative trait analysis of type 2 diabetes susceptibility loci identified from whole genome association studies in the Insulin Resistance Atherosclerosis Family Study. *Diabetes*, 57, 1093-1100.

- Pandit MK, Burke J, Gustafson AB, Minocha A, Peiris AN. (1993). Drug-induced disorders of glucose tolerance. *Ann Int Med*, 118, 529–540.
- Panoutsopoulou K, Zeggini E. (2009). Finding common susceptibility variants for complex disease: past, present and future. *Brief Funct Genomic Proteomic*, 8, 345-352.
- Parikh H, Lyssenko V, Groop LC. (2009). Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. *BMC Med Genomics*, 2, 72.
- Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, Miyazaki Y, Kohane I, Costello M, Saccone R, Landaker EJ, Goldfine AB, Mun E, DeFronzo R, Finlayson J, Kahn CR, Mandarino LJ. (2003). Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc Natl Acad Sci USA*, 100, 8466–8471.
- Paz K, Zick Y. Ser/Thr phosphorylation of insulin receptor signaling molecules and insulin resistance. In: *Insulin Signaling: From cultured cells to animal models*. Editor: Gruenberg G, Zick Y. Taylor & Francis, New York, 2002; pp 259-280.
- Pearl-Yafe M, Kaminitz A, Yolcu ES, Yaniv I, Stein J, Askenasy N. (2007). Pancreatic islets under attack: cellular and molecular effectors. *Curr Pharm Des*, 13, 749-760.
- Pearson ER, Flechtner I, Njølstad PR, Malecki MT, Flanagan SE, Larkin B, Ashcroft FM, Klimes I, Codner E, Iotova V, Slingerland AS, Shield J, Robert JJ, Holst JJ, Clark PM, Ellard S, Søvik O, Polak M, Hattersley AT; Neonatal Diabetes International Collaborative Group. (2006). Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med*, 355, 467–477.
- Permutt, MA, Wasson J, Cox N. (2005). Genetic epidemiology of diabetes. *J. Clin. Invest.* 115, 1431–1439.
- Perry JR, McCarthy MI, Hattersley AT, Zeggini E; Wellcome Trust Case Control Consortium, Weedon MN, Frayling TM. (2009). Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. *Diabetes*, 58, 1463-1467.
- Peter-Riesch B, Fathi M, Schlegel W, Wollheim CB. (1998). Glucose and carbachol generate 1,2-diacylglycerols by different mechanisms in pancreatic islets. *J Clin Invest*, 81, 1154–1161.

Bibliografía

- Petersen KF, Dufour S, Befroy D, Garcia R, Shulman GI. (2004). Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N Engl J Med*, 350, 664–671.
- Pfeifer A, Courtney M, Ben-Othman N, Al-Hasani K, Gjernes E, Vieira A, Druelle N, Avolio F, Faurite B, Mansouri A, Collombat P. (2013). Induction of multiple cycles of pancreatic β -cell replacement. *Cell Cycle*, 12, 3243-3244.
- Phelps G, Chapman I, Hall P, Braund W, Mackinnon M. (1989). Prevalence of genetic haemochromatosis among diabetic patients. *Lancet*, 2, 233-234.
- Pinney SE, Simmons RA. (2010). Epigenetic mechanisms in the development of type 2 diabetes. *Trends Endocrinol Metab*, 21, 223-229.
- Pirrot P, Cardozo AK, Eizirik DL. (2008). Mediators and mechanisms of pancreatic beta-cell death in type 1 diabetes. *Arq Bras Endocrinol Metabol*, 52, 156-165.
- Pohl A, Devaux PF, Herrmann A. (2005). Function of prokaryotic and eukaryotic ABC proteins in lipid transport. *Biochimica et Biophysica Acta*, 1733, 29–52.
- Polychronakos C, Li Q. (2011). Understanding type 1 diabetes through genetics: advances and prospects. *Nat Rev Genet*, 12, 781-792.
- Postic C, Shiota M, Niswender KD, Jetton TL, Chen Y, Moates JM, Shelton KD, Lindner J, Cherrington AD, Magnuson MA. (1999). Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic beta cell-specific gene knock-outs using Cre recombinase. *J. Biol. Chem.* 274, 305–315.
- Poulsen P, Kyvik KO, 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, 139-145.
- Prentki M, Matschinsky FM. (1987). Ca²⁺, cAMP, and phospholipid derived messengers in coupling mechanisms of insulin secretion. *Physiol Rev*, 67, 1185–1248.
- Pritchard JK. (2001). Are rare variants responsible for susceptibility to complex diseases?. *Am J Hum Genet*, 69, 124-137.
- Prokopenko I, McCarthy MI, Lindgren CM. (2008). Type 2 diabetes: new genes, new understanding. *Trends Genet*, 24, 613-621.

-Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, Loos RJ, Manning AK, Jackson AU, Aulchenko Y, et al. (2009). Variants in MTNR1B influence fasting glucose levels. *Nature Genetics*, 41, 77–81.

-Pulst SM. (1999). Genetic Linkage Analysis. *Arch Neurol*, 56, 667-672.

Q:

-Qi L, Cornelis MC, Kraft P, Stanya KJ, Linda Kao WH, Pankow JS, Dupuis J, Florez JC, Fox CS, Paré G, Sun Q, Girman CJ, Laurie CC, Mirel DB, Manolio TA, Chasman DI, Boerwinkle E, Ridker PM, Hunter DJ, Meigs JB, Lee CH, Hu FB, van Dam RM; Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC); Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium. (2010). Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet*, 19, 2706 -2715.

-Quintañilla-García C, Zuñiga-Guajardo S. (2010). El efecto incretina y su participación en la diabetes mellitus tipo 2. *Rev Med Inst Mex Seguro Soc*, 48, 509-520.

R:

-Rajala MW, Scherer PE. (2003). Minireview: the adipocyte-at the crossroads of energy homeostasis, inflammation, and atherosclerosis. *Endocrinology*, 144, 3765–3773.

-Ramírez Alvarado MM., Sánchez Roitz C. (2012). El factor de necrosis tumoral- α , la resistencia a la insulina, el metabolismo de lipoproteínas y la obesidad en humanos. *Nutr. Hosp*, 27, 1751-1757.

-Ravussin E, Smith SR. (2002). Increased fat intake, impaired fat oxidation, and failure of fat cell proliferation result in ectopic fat storage, insulin resistance, and type 2 diabetes mellitus. *Ann N Y Acad Sci*, 967, 363–378.

-Reid DM, Gow NAR, Brown GD. (2009). Pattern recognition: recent insights from Dectin-1. *Curr Opin Immunol*, 21, 30-37.

-Rewers M, Hamman RF. Risk factors for non-insulin-dependent diabetes. In: Harris MI, Cowie CC, Stern MP, et al, eds. *Diabetes in America*. 2nd ed. Bethesda, MD: National

Bibliografía

Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 1995:179–220.

-Richards JB, Waterworth D, O'Rahilly S, Hivert MF, Loos RJ, Perry JR, Tanaka T, Timpson NJ, Semple RK, Soranzo N, et al. (2009). A genome-wide association study reveals variants in *ARL15* that influence adiponectin levels. *PLoS Genet*, 5, e1000768.

-Rippe V, Drieschner N, Meiboom M, Murua Escobar H, Bonk U, Belge G, Bullerdiel J. (2003). Identification of a gene rearranged by 2p21 aberrations in thyroid adenomas. *Oncogene*, 22, 6111–6114.

-Robertson RP, Harmon J, Tran PO, Tanaka Y, Takahashi H. (2003). Glucose toxicity in beta-cells: type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes*, 52, 581–87.

-Robertson RP, Harmon J, Tran PO, Poitout V. (2004). Beta-cell glucose toxicity, lipotoxicity, and chronic oxidative stress in type 2 diabetes. *Diabetes*, 53, S119–24.

-Robinson R. (2010). Common disease, multiple rare (and distant) variants. *PLoS Biol*, 8, e1000293.

-Roglic G. Diabetes mortality. In: Gan D, editor. *Diabetes Atlas*. 3. Brussels: International Diabetes Federation; 2006.

-Roglic G, Unwin N, et al. Mortality attributable to diabetes: estimates for the year 2010. In: Unwin N, et al., editors. *Diabetes Atlas*. 4. Brussels: International Diabetes Federation; 2009.

-Rolandsson O, Palmer JP. (2010). Latent autoimmune diabetes in adults (LADA) is dead: long live autoimmune diabetes. *Diabetologia*, 53, 1250–1253.

-Rojewska E, Korostynski M, Przewlocki R, Przewlocka B, Mika J. (2014). Expression profiling of genes modulated by minocycline in a rat model of neuropathic pain. *Mol Pain*, 10, 47.

-Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proença C, Bacot F, Balkau B, Belisle A, Borch-Johnsen K, Charpentier G, Dina C, Durand E, Elliott P, Hadjadj S, Järvelin MR, Laitinen J, Lauritzen T, Marre M, Mazur A, Meyre D, Montpetit A, Pisinger C, Posner B, Poulsen P, Pouta A, Prentki M, Ribel-Madsen R, Ruukonen A, Sandbaek A, Serre D, Tichet J, Vaxillaire M, Wojtaszewski JF, Vaag A, Hansen T, Polychronakos C, Pedersen O, Froguel P, Sladek R. (2009). Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet*, 41, 1110–1115.

S:

-Sadler AJ, Williams BR. (2008). Interferon-inducible antiviral effectors. *Nature Reviews Immunology*, 8, 559–68.

-Sakamoto Y, Inoue H, Keshavarz P, Miyawaki K, Yamaguchi Y, Moritani M, Kunika K, Nakamura N, Yoshikawa T, Yasui N, Shiota H, Tanahashi T, Itakura M. (2007). SNPs in the KCNJ11-ABCC8 gene locus are associated with type 2 diabetes and blood pressure levels in the Japanese population. *J. Hum. Genet*, 52, 781-793.

-Sang Y, Zong W, Yan J, Liu M. (2012). The Correlation between the CLEC16A Gene and Genetic Susceptibility to Type 1 Diabetes in Chinese Children. *Int J Endocrinol*, 2012, 245384.

-Sato R. (2010). Sterol metabolism and SREBP activation. *Arch Biochem Biophys*, 501, 177-181.

-Savic D, Ye H, Aneas I, Park S-Y, Bell GI, Nobrega MA. (2011). Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism. *Genome Research*, 21, 1417–1425.

-Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316, 1331–1336.

-Saxena R, Saleheen D, Been LF, Garavito ML, Braun T, Bjornes A, Young R, Ho WK, Rasheed A, Frossard P, Sim X, Hassanali N, Radha V, Chidambaram M, Liju S, Rees SD, Ng DP, Wong TY, Yamauchi T, Hara K, Tanaka Y, Hirose H, McCarthy MI, Morris AP; DIAGRAM; MuTHER; AGEN, Basit A, Barnett AH, Katulanda P, Matthews D, Mohan V, Wander GS, Singh JR, Mehra NK, Ralhan S, Kamboh MI, Mulvihill JJ, Maegawa H, Tobe K, Maeda S, Cho YS, Tai ES, Kelly MA, Chambers JC, Kooner JS, Kadowaki T, Deloukas P, Rader DJ, Danesh J, Sanghera DK. (2013). Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes*, 62, 1746-55.

-Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007). A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*, 316, 1341–1345.

Bibliografía

-Schernthaner G, Hink S, Kopp HP, Muzyka B, Streit G, Kroiss A. (2001). Progress in the characterization of slowly progressive autoimmune diabetes in adult patients (LADA or type 1.5 diabetes). *Exp Clin Endocrinol Diabetes*, 109, S94–S108.

-Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, Rice CM. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, 472, 481–485.

- Schork NJ, Murray SS, Frazer KA, Topol EJ. (2009). Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*, 19, 212-219.

-Schuster SC. (2008). Next-generation sequencing transforms today's biology. *Nat Methods*, 5, 16–18.

-Schwartz SS, Zeidler A, Moossa AR, Kuku SF, Rubenstein AH. (1978). A prospective study of glucose tolerance, insulin, C-peptide, and glucagon responses in patients with pancreatic carcinoma. *Digestive Dis*, 23, 1107– 1114.

-Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. (2007). A genome wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, 316, 1341 -1345.

-Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Mägi R, Strawbridge RJ, Rehnberg E, Gustafsson S, et al. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet*, 44, 991–1005.

-Senn JJ, Klover PJ, Nowak IA, Zimmers TA, Koniaris LG, Furlanetto RW, Mooney RA. (2003). Suppressor of cytokine signaling-3 (SOCS-3), a potential mediator of interleukin-6- dependent insulin resistance in hepatocytes. *J Biol Chem*, 278, 13740–13746.

-Seppälä-Lindroos A, Vehkavaara S, Häkkinen AM, Goto T, Westerbacka J, Sovijärvi A, Halavaara J, Yki-Järvinen H. (2002). Fat accumulation in the liver is associated with defects in insulin suppression of glucose production and serum free fatty acids independent of obesity in normal men. *J Clin Endocrinol Metab*, 87, 3023–3028.

-Shaat N, Ekelund M, Lernmark A, Ivarsson S, Almgren P, Berntorp K, Groop L. (2005). Association of the E23K polymorphism in the KCNJ11 gene with gestational diabetes mellitus. *Diabetologia*, 48, 2544–2551.

-Shaat N, Karlsson E, Lernmark A, Ivarsson S, Lynch K, Parikh H, Almgren P, Berntorp K, Groop L. (2006). Common variants in MODY genes increase the risk of gestational diabetes mellitus. *Diabetologia*, 49, 1545–1551.

-Shaat N, Lernmark A, Karlsson E, Ivarsson S, Parikh H, Berntorp K, Groop L. (2007). A variant in the transcription factor 7-like 2 (TCF7L2) gene is associated with an increased risk of gestational diabetes mellitus. *Diabetologia*, 50, 972–979.

-Shao S, Liu Z, Yang Y, Zhang M, Yu X. (2010). SREBP-1c, Pdx-1, and GLP-1R involved in palmitate-EPA regulated glucose-stimulated insulin secretion in INS-1 cells. *J Cell Biochem*, 111, 634-642.

-Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309, 1728–1732.

-Shiba T, Morino Y, Tagawa K, Fujino H, Unuma T. (1996). Onset of diabetes with high titer anti-GAD antibody after IFN therapy for chronic hepatitis. *Diabetes Res Clin Pract*, 30, 237–241.

-Shin HD, Kim LH, Park BL, Jung HS, Cho YM, Moon MK, Lee HK, Park KS. (2003). Polymorphisms in fatty acid-binding protein-3 (FABP3) - putative association with type 2 diabetes mellitus. *Hum Mutat*, 22, 180.

-Shin HD, Park BL, Shin HJ, Kim JY, Park S, Kim B, Kim SH. (2010). Association of KCNQ1 polymorphisms with the gestational diabetes mellitus in Korean women. *J Clin Endocrinol Metab*, 95, 445–449.

-Shiota C, Coffey J, Grimsby J, Grippo JF, Magnuson MA. (1999). Nuclear import of hepatic glucokinase depends upon glucokinase regulatory protein, whereas export is due to a nuclear export signal sequence in glucokinase. *J. Biol. Chem.* 274, 37125–37130.

-Shoelson SE, Lee J, Yuan M. (2003). Inflammation and the IKK beta/Ikappa B/NF-kappa B axis in obesity- and diet-induced insulin resistance. *Int J Obes Relat Metab Disord*, 27, S49–52.

-Shu L, Sauter NS, Schulthess FT, Matveyenko AV, Oberholzer J, Maedler K. (2008). Transcription factor 7-like 2 regulates β -cell survival and function in human pancreatic islets. *Diabetes*, 57, 645–653.

Bibliografía

-Shu XO, Long J, Cai Q, Qi L, Xiang YB, Cho YS, Tai ES, Li X, Lin X, Chow WH, Go MJ, Seielstad M, Bao W, Li H, Cornelis MC, Yu K, Wen W, Shi J, Han BG, Sim XL, Liu L, Qi Q, Kim HL, Ng DP, Lee JY, Kim YJ, Li C, Gao YT, Zheng W, Hu FB. (2010). Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet*, 6, e1001127.

-Shulman GI. (2000). Cellular mechanisms of insulin resistance. *J Clin Invest*, 106, 171–176.

-Skingningsrud B, Husebye ES, Pearce SH, McDonald DO, Brandal K, Wolff AB, Løvås K, Egeland T, Undlien DE. (2008). Polymorphisms in CLEC16A and CIITA at 16p13 are associated with primary adrenal insufficiency. *Clin Endocrinol Metab*, 93, 3310-3317.

-Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445, 881 -885.

-Smathers RL, Petersen DR. (2011). The human fatty acid-binding protein family: evolutionary divergences and functions. *Hum Genomics*, 5, 170-191.

-Smith NL, Hankinson J, Simpson A, Denning DW, Bowyer P. (2014). Reduced expression of TLR3, TLR10 and TREM1 by human macrophages in CCPA, and novel associations of VEGFA, DENND1B and PLAT. *Clin Microbiol Infect*, 20, O960-8.

-So HC, Gui AH, Cherny SS, Sham PC. (2011). Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol*, 35, 310-317.

-Soffer LJ, Iannaccone A, Gabilove JL. (1961). Cushing's syndrome. *Am J Med*, 30, 129–146.

-Solimena M, Folli, Aparisi R, Pozza G, De Camilli P. (1992). Autoantibodies to GABA-nergic neurons and pancreatic beta cells in stiff-man syndrome. *N Engl J Med*, 41, 347–353.

-Soriguer F, Goday A, Bosch-Comas A, Bordiú E, Calle-Pascual A, Carmena R, Casamitjana R, Castaño L, Castell C, Catalá M, Delgado E, Franch J, Gaztambide S, Girbés J, Gomis R, Gutiérrez G, López-Alba A, Martínez-Larrad MT, Menéndez E, Mora-Peces I, Ortega E, Pascual-Manich G, Rojo-Martínez G, Serrano-Rios M, Valdés S, Vázquez JA, Vendrell J. (2012). Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Di@bet.es Study. *Diabetologia*, 55, 88-93.

- Sosa-Pineda B. (2004). The gene Pax4 is an essential regulator of pancreatic beta-cell development. *Mol Cells*, 18, 289-294.
- Stancakova A, Kuulasmaa T, Paananen J, Jackson AU, Bonnycastle LL, Collins FS, Boehnke M, Kuusisto J, Laakso M. (2009). Association of 18 confirmed susceptibility loci for type 2 diabetes with indices of insulin release, proinsulin conversion, and insulin sensitivity in 5,327 nondiabetic Finnish men. *Diabetes*, 58, 2129–2136.
- Steck AK, Rewers MJ. (2011). Genetics of type 1 diabetes. *Clin Chem*, 57, 176-85.
- Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, Helgadottir HT, Johannsdottir H, Magnusson OT, Gudjonsson SA, Justesen JM, Harder MN, Jørgensen ME, Christensen C, Brandslund I, Sandbæk A, Lauritzen T, Vestergaard H, Linneberg A, Jørgensen T, Hansen T, Daneshpour MS, Fallah MS, Hreidarsson AB, Sigurdsson G, Azizi F, Benediktsson R, Masson G, Helgason A, Kong A, Gudbjartsson DF, Pedersen O, Thorsteinsdottir U, Stefansson K. (2014). Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet*, 46, 294-298.
- Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeyasinghe S, Krawczak M, Cooper DN. (2003). Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*, 21, 577-581.
- Stenstrom G, Ernest I, Tisell L. (1988). Long-term results in 64 patients operated upon for pheochromocytoma. *Acta Med Scan*, 223, 345–352.
- Stenstrom G, Gottsater A, Bakhtadze E, Berger B, Sundkvist G. (2005). Latent autoimmune diabetes in adults: definition, prevalence, beta-cell function, and treatment. *Diabetes*, 54, S68–S72.
- Strodl E, Kenardy J. (2006). Psychosocial and non-psychosocial risk factors for the new diagnosis of diabetes in elderly women. *Diabetes Res Clin Pract*, 74, 57–65.
- Stumvoll M, Tataranni PA, Stefan N, Vozarova B, Bogardus C. (2003). Glucose allostasis. *Diabetes*, 52, 903–909.
- Sul JH, Han B, He D, Eskin E. (2011). An Optimal Weighted Aggregated Association Test for Identification of Rare Variants Involved in Common Diseases. *Genetics*, 188, 181-188.
- Suram S, Brown GD, Ghosh M, Gordon S, Loper R, Taylor PR, Akira S, Uematsu S, Williams DL, Leslie CC. (2006). Regulation of cytosolic phospholipase A2 activation and

Bibliografía

Cyclooxygenase 2 expression in macrophages by the beta-glucan receptor. *J Biol Chem*, 281, 5506-5514.

T:

-Tan JT, Ng DP, Nurbaya S, Ye S, Lim XL, Leong H, Seet LT, Siew WF, Kon W, Wong TY, Saw SM, Aung T, Chia KS, Lee J, Chew SK, Seielstad M, Tai ES. (2010). Polymorphisms identified through genome-wide association studies and their associations with type 2 diabetes in Chinese, Malays, and Asian-Indians in Singapore. *J Clin Endocrinol Metab*, 95, 390-397.

-Tan KT, Tayebjee MH, Lim HS, Lip GY. (2005). Clinically apparent atherosclerotic disease in diabetes is associated with an increase in platelet microparticle levels. *Diabet Med*, 22, 1657-1662.

-Tanaka D, Nagashima K, Sasaki M, Yamada C, Funakoshi S, Akitomo K, Takenaka K, Harada K, Koizumi A, Inagaki N. (2011). GCKR mutations in Japanese families with clustered type 2 diabetes. *Mol Genet Metab*, 102, 453-460.

-Tanese K, Fukuma M, Yamada T, Mori T, Yoshikawa T, Watanabe W, Ishiko A, Amagai M, Nishikawa T, Sakamoto M. (2008). G-protein-coupled receptor GPR49 is up-regulated in basal cell carcinoma and promotes cell proliferation and tumor formation. *The American journal of pathology*, 173, 835-843.

-Tang X. (2009). The Study of Single Nucleotide Polymorphism of Susceptibility Genes of Type Two Diabetes in Chinese Han Population. Master's thesis, CQMU, Chong Qing.

-Tauriainen S, Oikarinen S, Oikarinen M, Hyoty H. (2011). Enteroviruses in the pathogenesis of type 1 diabetes. *Semin Immunopathol*, 33, 45-55.

-Terauchi Y, Sakura H, Yasuda K, Iwamoto K, Takahashi N, Ito K, Kasai H, Suzuki H, Ueda O, Kamada N, Jishage K, Komeda K, Noda M, Kanazawa Y, Taniguchi S, Miwa I, Akanuma Y, Kodama T, Yazaki Y, Kadowaki T. (1995). Pancreatic beta-cell-specific targeted disruption of glucokinase gene. Diabetes mellitus due to defective insulin secretion to glucose. *J. Biol. Chem.* 270, 30253–30256.

-Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszek JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S,

Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tîrgoviște C; Genetics of Type 1 Diabetes in Finland, Simmonds MJ, Heward JM, Gough SC; Wellcome Trust Case Control Consortium, Dunger DB, Wicker LS, Clayton DG. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*, 39, 857–864.

-Tomlinson MJ 4th, Pitsillides A, Pickin R, Mika M, Keene K, Hou X, Mychaleckyj J, Chen WM, Concannon P, Onengut-Gumuscu S. (2014). Fine Mapping and Functional Studies of Risk Variants for Type 1 Diabetes at Chromosome 16p13.13. *Diabetes*, 63, 4360-8.

-Tsai FJ, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, Wang TY, Chen RH, Shiu CF, Liu YM, Chang CC, Chen P, Chen CH, Fann CS, Chen YT, Wu JY. (2010). A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet*, 6, e1000847.

-Tuomi T, Groop LC, Zimmet PZ, Rowley MJ, Knowles W, Mackay IR. (1993). Antibodies to glutamic acid decarboxylase reveal latent autoimmune diabetes mellitus in adults with a non-insulin-dependent onset of disease. *Diabetes*, 42, 359–362.

-Tuomi T, Carlsson A, Li H, Isomaa B, Miettinen A, Nilsson A, Nissen M, Ehrnstrom BO, Forsen B, Snickars B, Lahti K, Forsblom C, Saloranta C, Taskinen MR, Groop LC. (1999). Clinical and genetic characteristics of type 2 diabetes with and without GAD antibodies. *Diabetes*, 48, 150–157.

-Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P. (2001). Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*, 344, 1343–1350.

-Turner R, Stratton I, Horton V, Manley S, Zimmet P, Mackay IR, Shattock M, Bottazzo GF, Holman R. (1997). UKPDS 25: autoantibodies to islet-cell cytoplasm and glutamic acid decarboxylase for prediction of insulin requirement in type 2 diabetes UK Prospective Diabetes Study Group. *Lancet*, 350, 1288–1293.

U:

-Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, Ng DP, Holmkvist J, Borch-Johnsen K, Jørgensen T, Sandbaek A, Lauritzen T, Hansen T, Nurbaya S, Tsunoda T, Kubo M, Babazono T, Hirose H, Hayashi M, Iwamoto Y, Kashiwagi A, Kaku K, Kawamori

Bibliografía

R, Tai ES, Pedersen O, Kamatani N, Kadowaki T, Kikkawa R, Nakamura Y, Maeda S. (2008). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet*, 40, 1098 -1102.

-Uusitupa M, Louheranta A, Lindström J, Valle T, Sundvall J, Eriksson J, Tuomilehto J. (2000). The Finnish Diabetes Prevention Study. *Br J Nutr*, 83, S137–42.

V:

-Vandercammen A, Van Schaftingen E. (1990). The mechanism by which rat liver glucokinase is inhibited by the regulatory protein. *Eur. J. Biochem*. 191, 483–489.

-Vandercammen, A, Van Schaftingen E. (1990). Competitive inhibition of liver glucokinase by its regulatory protein. *Eur. J. Biochem*. 200, 545–551.

-Vandercammen, A, Van Schaftingen E. (1989). Stimulation of glucose phosphorylation by fructose in isolated rat hepatocytes. *Eur. J. Biochem*. 179, 173–177.

-van Haefen TW, Pimenta W, Mitrakou A, Korytkowski M, Jenssen T, Yki-Jarvinen H, Gerich JE. (2002). Disturbances in beta-cell function in impaired fasting glycemia. *Diabetes*, 51, S265–70.

-Van Schaftingen E. (1989). A protein from rat liver confers to glucokinase the property of being antagonistically regulated by fructose 6-phosphate and fructose 1-phosphate. *Eur. J. Biochem*. 179, 179–184.

-Van Schaftingen E, Detheux M, Veiga Da Cunha M. (1994). Short-term control of glucokinase activity: role of a regulatory protein. *FASEB J*. 8, 414–419.

-Vassy JL, Meigs JB. (2012). Is genetic testing useful to predict type 2 diabetes?. *Best Pract Res Clin Endocrinol Metab*, 26, 189-201.

-Vaxillaire M, Froguel P. (2006). Genetic basis of maturity-onset diabetes of the young. *Endocrinol Metab Clin North Am*, 35, 371–384.

-Vaxillaire M, DP, Bonnefond A, Froguel P. (2009). Breakthroughs in monogenic diabetes genetics: from pediatric forms to young adulthood diabetes. *Pediatr Endocrinol Rev*, 6, 405–417.

-Vilà L, Roglans N, Perna V, Sánchez RM, Vázquez-Carrera M, Alegret M, Laguna JC. (2011). Liver AMP/ATP ratio and fructokinase expression are related to gender differences in AMPK activity and glucose intolerance in rats ingesting liquid fructose. *J Nutr Biochem*, 22, 741-51.

-Virjee S, Robinson S, Johnston DG. (2001). Screening for diabetes in pregnancy. *Journal of the Royal Society of Medicine*, 94, 502–509.

-Virkamaki A, Ueki K, Kahn CR. (1999). Protein-protein interaction in insulin signaling and the molecular mechanisms of insulin resistance. *J Clin Invest*, 103, 931-943.

-Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G et al. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*, 42, 579-589.

W:

-Wallace TM, Matthews DR. (2002). Coefficient of failure: a methodology for examining longitudinal beta-cell function in Type 2 diabetes. *Diabet Med*, 19, 465–469.

-Wang K, Li M, Hakonarson H. (2010). ANNOVAR: functional annotation of genetic variants from high throughput sequencing data. *Nucleic Acids Res*, 38, e164.

-Wang EJ, Wong EC, Dixit AA, Fortmann SP, Linde RB, Palaniappan LP. (2011). Type 2 diabetes: identifying high risk Asian American subgroups in a clinical population. *Diabetes Research and Clinical Practice Journal*, 93, 248–254.

- Watanabe RM, Allayee H, Xiang AH, Trigo E, Hartiala J, Lawrence JM, Buchanan TA. (2007). Transcription factor 7-like 2 (TCF7L2) is associated with gestational diabetes mellitus and interacts with adiposity to alter insulin secretion in Mexican Americans. *Diabetes*, 56, 1481–1485.

-Weir GC, Leahy JL. Pathogenesis of non-insulin-dependant (type II) diabetes mellitus. In: Kahn CR, Weir GC, Pennsylvania, eds. *Joslin's diabetes mellitus*. 13th ed. Malvern: Lea and Febiger; 1994.

-Wellen KE, Hotamisligil GS. (2003). Obesity-induced inflammatory changes in adipose tissue. *J Clin Invest*, 112, 1785–1788.

Bibliografía

-Weyer C, Bogardus C, Mott DM, Pratley RE. (1999). The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest*, 104, 787–794.

-Weyer C, Bogardus C, Pratley RE. (1999). Metabolic characteristics of individuals with impaired fasting glucose and/or impaired glucose tolerance. *Diabetes*, 48, 2197–2203.

-White MF. (2002). IRS proteins and the common path to diabetes. *Am J Physiol Endocrinol Metab*, 283, E413–E422.

-Will JC, Galuska DA, Ford ES, Mokdad A, Calle EE. (2001). Cigarette smoking and diabetes mellitus: evidence of a positive association from a large prospective cohort study. *Int J Epidemiol*, 30, 540–546.

-Wommack KE, J. Bhavsar, J. Ravel. (2008). Metagenomics: read length matters. *Appl. Environ. Microbiol*, 74, 1453–1463.

-Wongseree W, Assawamakin A, Piroonratana T, Sinsomros S, Limwongse C, Chaiyaratana N. (2009). Detecting purely epistatic multi-locus interactions by an omnibus permutation test on ensembles of two-locus analyses. *BMC Bioinformatics*, 10, 294.

Y:

-Yamamoto Y, Sakamoto M, Fujii G, Tsuiji H, Kenetaka K, Asaka M and Hirohashi S. (2003). Overexpression of orphan G-protein-coupled receptor, Gpr49, in human hepatocellular carcinomas with beta-catenin mutations. *Hepatology*, 37, 528-533.

-Yamauchi T, Kamon J, Minokoshi Y, Ito Y, Waki H, Uchida S, Yamashita S, Noda M, Kita S, Ueki K, Eto K, Akanuma Y, Froguel P, Foufelle F, Ferre P, Carling D, Kimura S, Nagai R, Kahn BB, Kadowaki T. (2002). Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP activated protein kinase. *Nat Med*, 8, 1288–1295.

-Yang Q, Khoury MJ, Friedman J, Little J, Flanders WD. (2005). How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol*, 34, 1129-1137.

- Yang W, Kelly T, He J. (2007). Genetic epidemiology of obesity. *Epidemiol Rev*, 29, 49-61.
- Yi F, Brubaker PL, Jin T. (2005). TCF-4 mediates cell type-specific regulation of proglucagon gene expression by β -catenin and glycogen synthase kinase-3 β . *Journal of Biological Chemistry*, 280, 1457–1464.
- Yin MJ, Yamamoto Y, Gaynor RB. (1998). The anti-inflammatory agents aspirin and salicylate inhibit the activity of I(kappa)B kinase-beta. *Nature*, 396, 77–80.
- Yki-Järvinen H. (1992). Glucose toxicity. *Endocr Rev*, 13, 415–431.
- Youngren J. (2007). Regulation of insulin receptor function. *Cellular and Molecular Life Sciences*, 64, 873-891.
- Yu M. (2009). Association Study of the Genetic Polymorphisms of KCNJ11 and TCF7L2 Genes and Repaglinide Response in Type 2 Diabetes Patients. Master's thesis, CSU, Hunan.
- Yuan M, Konstantopoulos N, Lee J, Hansen L, Li ZW, Karin M, Shoelson SE. (2001). Reversal of obesity- and diet-induced insulin resistance with salicylates or targeted disruption of Ikkbeta. *Science*, 293, 1673–1677.

Z:

- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS; Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316, 1336 -1341.
- Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*, 40, 638-645.

Bibliografía

-Zhao ZY, Yang NP, Huang XY, Min L, Liu XX, Zhang J, et al. (2010). Interferon-inducible genes 2',5'-oligoadenylate synthetase-like and tetratricopeptide repeats 2 are correlated with clinical features of patients with systemic lupus erythematosus. *Sichuan Da Xue Xue Bao Yi Xue Ban*, 41, 247–251.

-Zhou G, Myers R, Li Y, Chen Y, Shen X, Fenyk-Melody J, Wu M, Ventre J, Doebber T, Fujii N, Musi N, Hirshman MF, Goodyear LJ, Moller DE. (2001). Role of AMP-activated protein kinase in mechanism of metformin action. *J Clin Invest*, 108, 1167–1174.

-Zhou Q, Amar S. (2007). Identification of signaling pathways in macrophage exposed to *Porphyromonas gingivalis* or to its purified cell wall components. *Journal of Immunology*, 179, 7777–7790.

-Zhou Q, Zhang K, Li W, Liu JT, Hong J, Qin SW, Ping F, Sun ML, Nie M. (2009). Association of KCNQ1 gene polymorphism with gestational diabetes mellitus in a Chinese population. *Diabetologia*, 52, 2466–2468.

-Zimmet PZ. (1995). The pathogenesis and prevention of diabetes in adults. Genes, autoimmunity, and demography. *Diabetes Care*, 18, 1050–1064

-Zimmet P, Alberti KG, Shaw J. (2001). Global and societal implications of the diabetes epidemic. *Nature*, 414, 782–787.

-Zou MH, Kirkpatrick SS, Davis BJ, Nelson JS, Wiles WG 4th, Schlattner U, Neumann D, Brownlee M, Freeman MB, Goldman MH. (2004). Activation of the AMP activated protein kinase by the anti-diabetic drug metformin in vivo: Role of mitochondrial reactive nitrogen species. *J Biol Chem*, 279, 43940–43951.

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

-1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56-65.

Anexo I

ANEXO 1

En este anexo se encuentran todos los comandos empleados en el proceso de análisis bioinformático:

Obtención de estadísticas de archivos FASTQ

Ejecutar FastQC

- Comando genérico:

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

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

Depuración de los archivos FASTQ

Recortar las bases de baja calidad de las lecturas obtenidas.

Ejecutar seqtk

- Comando genérico:

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

Mapeo

1.1. Ejecutar BWA

- Comando genérico:

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

- Opciones empleadas BWA:
 - **-t (INT)**: Número de hilos de ejecución [1]
 - **-C**: Anexar comentario FASTA /Q al archivo SAM resultante. Esta opción se puede utilizar para transferir información de lectura (por ejemplo, código de barras) al archivo resultante.
 - **-M**: Indica los alineamientos divididos como secundarios (para la compatibilidad con Picard).
 - **-R (STR)**: Completar la línea del read group de la cabecera. '\t' puede ser utilizado en STR y se convertirá a TAB en el archivo SAM resultante. El read group ID irá unido a cada lectura en el output. Un ejemplo es: '@RG\tID:foo\tSM:bar'. [null]
- Formato de la cabecera:
 - "@RG\tID:id\tPL:platform\tSM:sample\tDS:description\tCN:center\tDT:date\tPU:flowcell-barcode.lane"
- Ejemplo de la cabecera:
 - "@RG\tID:280397_CGATGT_L005_R1_001\tPL:illumina\tSM:280397\tDS:ref=1KG;pfx=/media/sf_ugdg_data/ref_data/human/1kg/hs37d5.fa.gz\tCN:INCLIVA\tDT:2013-09-19\tPU:130327_H114_0265_AD1LLUACXX-.2"

1.2. Conversión de SAM a BAM

- Comando genérico:

```
$ samtools view -b -S <input.sam> > <output.bam>
```
- Opciones empleadas SAMtools:
 - **-b**: El archivo de salida se crea en formato BAM.
 - **-S**: El archivo de entrada está en formato SAM. Si las líneas de la cabecera @SQ no están, se requiere la opción '-t'.

2. Ejecutar BWA y SAMtools

- Comando genérico:

```
$ bwa mem -t[INT] -C -M -R[STR] <reference.fa.gz> <input_R1.fastq>  
<input_R2.fastq> 2> <output_R1-R2.sam.log> | samtools view -b -S - >  
<output_R1-R2.bam>
```
 - Opciones empleadas SAMtools:
 - **-:** El archivo de entrada proviene de stdout
-

Visualizar SAM

Ejecutar SAMtools view

- Comando genérico:

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

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

Visualizar BAM

Ejecutar SAMtools view

- Comando genérico:

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

Inspeccionar archivos BAM

Ejecutar SAMtools tview

- Comando genérico:

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

Ordenar archivos BAM

Ejecutar SAMtools sort

- Comando genérico:

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

- Opciones empleadas SAMtools:
 - **-m (INT)**: Máxima memoria requerida aproximadamente. [500000000]
-

Indexar archivos BAM

Ejecutar SAMtools index

- Comando genérico:

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

Eliminar duplicados

Ejecutar SAMtools rmdup

- Comando genérico:

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

Unir varios archivos BAM

Unir cabeceras de archivos BAM

- Comando genérico:

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

- Opciones empleadas SAMtools:
 - **-H**: Devolver solamente la cabecera.

Ejecutar SAMtools merge

- Comando genérico:

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

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

Obtención estadísticas archivos BAM

Ejecutar SAMtools flagstat

- Comando genérico:

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

Seleccionar subconjuntos en archivos BAM

Ejecutar SAMtools view

- Comando genérico:

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

Obtención de cobertura

Ejecutar BEDtools para obtener un histograma de cobertura

- Comando genérico:

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

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

Ejecutar BEDtools para obtener la cobertura por base

- Comando genérico:

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

- Opciones empleadas BEDtools:
 - **-d:** Devuelve la profundidad de cobertura para cada posición en B. Las posiciones devueltas empiezan en 1.
 - **-abam:** El archivo de entrada A está en formato BAM.

- **-b:** El archivo de entrada b está dentro.

Obtención de variaciones

Ejecutar SAMtools y BCFtools para obtener archivos VCF

- Comando genérico:

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

- Opciones empleadas SAMtools:
 - **-d (INT):** En una posición determinada, leer un máximo de INT lecturas por BAM. [250]
 - **-D:** Salida de la cobertura por muestra.
 - **-C:** Coeficiente empleado para reducir la calidad de mapeo de las lecturas que contienen excesivos errores. Un valor de 0, deshabilita esta función. Si se emplea, se recomienda un valor de 50 para BWA.
 - **-E:** Computación BAQ extendida.
 - **-f (FILE):** La secuencia de referencia indexada se encuentra en formato FASTA.
 - **-l (FILE):** archivo o lista de posiciones que contiene una lista de las regiones o de los sitios donde se deben generar pileup o BCF.
 - **-u:** Similar a -g exceptuando que el output es un BCF sin comprimir.
- Opciones empleadas BCFtools:
 - **-:** El archivo de entrada viene de stdout
 - **-v:** Output con sólo posiciones que contengan variantes. (force -c)
 - **-g:** Llamada de genotipos por base en los sitios que contienen variantes. (force -c)
 - **-c:** Llamar variantes usando inferencia Bayesiana. Esta opción invoca la opción -e.
 - **-e:** Realiza la estimación por máxima verosimilitud, incluyendo la frecuencia alélica, el equilibrio Hardy-Weinberg y test de asociación con razones de verosimilitud.**[NOTA]:** Si se omite la opción "-v", se imprimirán todas las posiciones, no solo las que contengan variaciones.

Comparación VCFs

Comprimir VCF con tabix

- Comando genérico:

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

Indexar VCF con tabix

- Comando genérico:

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

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

Comparar archivos VCF con VCFtools

- Comando genérico:

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

Anotación funcional

Preparación del archivo de entrada para VEP

- Comando genérico:

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

Ejecutar VEP

- Comando genérico:

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

VERSIONES EMPLEADAS

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

-Seqtk Versión: 1.0-r31

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

Anexo I

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

-VCFtools (v0.1.11)

-BEDtools (2.17.0)

