



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MÉDICAS,
ODONTOLÓGICAS Y DE LA SALUD

INSTITUTO MEXICANO DEL SEGURO SOCIAL
CENTRO MÉDICO NACIONAL SIGLO XXI

“ESTUDIO DE VARIANTES DEL NÚMERO DE COPIAS (CNVS) EN EL GENOMA DE PACIENTES CON DIABETES TIPO 2”

TESIS QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS MÉDICAS

PRESENTA:

IVÁN DE JESÚS ASCENCIO MONTIEL

TUTOR PRINCIPAL:

DR. MIGUEL CRUZ LÓPEZ

ENTIDAD DE ADSCRIPCIÓN:

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MÉDICAS,
ODONTOLÓGICAS Y DE LA SALUD
INSTITUTO MEXICANO DEL SEGURO SOCIAL
CENTRO MÉDICO NACIONAL SIGLO XXI

U.M.A.E HOSPITAL DE ESPECIALIDADES “DR. BERNARDO SEPÚLVEDA”
UNIDAD DE INVESTIGACIÓN MÉDICA EN BIOQUÍMICA.

COLABORADOR:

DR. ADÁN VALLADARES SALGADO

ENTIDAD DE ADSCRIPCIÓN:

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MÉDICAS,
ODONTOLÓGICAS Y DE LA SALUD
INSTITUTO MEXICANO DEL SEGURO SOCIAL
CENTRO MÉDICO NACIONAL SIGLO XXI

U.M.A.E HOSPITAL DE ESPECIALIDADES “DR. BERNARDO SEPÚLVEDA”
UNIDAD DE INVESTIGACIÓN MÉDICA EN BIOQUÍMICA.

MÉXICO, D.F., ABRIL DE 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

A Dios

Por el don de la vida

A mis Padres

Por su apoyo incondicional durante todo este tiempo

A mis hermanos

Por su compañía y sus ánimos

A mis profesores

Quienes alentaron mi desarrollo profesional

A mis asesores de tesis

Por sus valiosa colaboración en este proyecto

A mis amigos

Por su ayuda y alegría.

Dr. Miguel Cruz López

Jefe de la Unidad de Investigación Médica en Bioquímica.
Hospital de Especialidades “Dr. Bernardo Sepúlveda”
U.M.A.E Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

Dr. Adán Valladares Salgado

Unidad de Investigación Médica en Bioquímica.
Hospital de Especialidades “Dr. Bernardo Sepúlveda”
U.M.A.E Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

Dr. Fabio Abdel Salamanca Gómez

Director de la Entidad Académica
Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

Índice General

Índice General.....	4
Índice de Tablas, figuras y anexos	5
Resumen.....	8
1. Antecedentes científicos	9
2. Planteamiento del problema	20
3. Justificación	20
4. Pregunta de Investigación	21
5. Hipótesis	21
6. Objetivos	22
7. Variables	23
8. Material y Método	26
8.1. Diseño de estudio.....	26
8.2. Universo de estudio.....	26
8.3. Criterios de selección	26
8.4. Tamaño de muestra	27
8.5 Poder del estudio	29
9. Plan general: procedimientos y métodos.....	30
10. Análisis estadístico	32
11. Consideraciones éticas.....	33
12. Recursos humanos y materiales.....	35
13. Cronograma de actividades	36
14. Resultados	37
15. Discusión	45
16. Conclusiones	48
17. Anexos	49
18. Referencias bibliográficas.....	61

Índice de tablas, figuras y anexos

Tablas

Tabla 1. <i>Loci</i> asociados a la DT2	10
Tabla 2. <i>Loci</i> asociados a la DT2 en población Mexicana	12
Tabla 3. Características de las diferentes plataformas de microarreglos y las herramientas de detección de CNVs	15
Tabla 4. Antecedentes científicos de estudios de CNVs y DT2	17
Tabla 5. Fases del proyecto de investigación “Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”	18
Tabla 6. Características de los sujetos participantes del “Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes” genotipificados con el microarreglo <i>Affymetrix Genome-Wide Human SNP array 5.0</i>	19
Tabla 7. Características de los CNVs en los grupos de estudio	37
Tabla 8. Número de muestras con CNVs en los grupos de estudio	40
Tabla 9. Diferencia de proporciones de muestras con CNVs entre los grupos de estudio	41
Tabla 10. Regiones génicas con diferencia de proporciones cruda y ajustada significativas, en los grupos de estudio	43

Tablas Suplementarias

Tabla Suplementaria 1. Resumen del análisis de control de calidad (CC)	56
Tabla Suplementaria 2. Características de los sujetos participantes del “Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes” genotipificados con el microarreglo <i>Affymetrix Genome-Wide Human SNP array 5.0</i> , después del análisis de Control de Calidad	56
Tabla Suplementaria 3. Características de los sujetos participantes del “Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes” genotipificados con el microarreglo <i>Affymetrix Genome-Wide Human SNP array 5.0</i> , después del análisis de Control de Calidad. Comparativo entre muestras retenidas y eliminadas por grupo de estudio	57
Tabla Suplementaria 4. Número de CNVs obtenidos a través de los algoritmos de detección, en 964 muestras (753 sujetos DT2 y 211 donadores)	59

Tabla Suplementaria 5. Pasos para la selección final de los CNVs, en la base de datos común	60
--	----

Figuras

Figura 1. Fases para la caracterización de CNVs	31
Figura 2. Cronograma de actividades.....	37
Figura 3. Conteo de CNVs totales y génicos en los grupos de estudio.....	36
Figura 4. Porcentaje de CNVs génicos en duplicaciones y deleciones en los grupos de estudio	38
Figura 5. Porcentaje de duplicaciones y deleciones en los CNVs totales y génicos por los grupos de estudio	39
Figura 6. Mediana del tamaño de los CNVs totales y génicos por los grupos de estudio	39
Figura 7. Frecuencia de CNVs en los grupos de comparación.	40
Figura 8. Diferencia en la frecuencia CNVs totales y génicos por tipo de CNV (duplicación, deleción o ambos) en los grupos de comparación	41
Figura 9. Comparativo del tamaño de los CNVs de acuerdo a su longitud.	42
Figura 10. Diferencia de proporciones de CNVs en regiones génicas, en los grupos de estudio.	43
Figura 11. Regiones con diferencias de proporciones cruda y ajustada significativas entre los grupos de estudio	44

Figuras Suplementarias

Figura Suplementaria 1. Flujograma del proceso de análisis de control de calidad para la selección de las muestras y CNVs	54
Figura Suplementaria 2. Pasos de eliminación de muestras en la primera etapa del análisis de control de calidad.....	55
Figura Suplementaria 3. Correlación entre el número de CNVs ≥ 5 sondas ≥ 5 kb entre los algoritmos iPattern, PennCNV y Birdsuite	60

Anexos

Anexo 1. Procedimiento de selección de los sujetos y del procesamiento de las muestras de los participantes del protocolo “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”	49
---	----

Anexo 2. Estimación de la Ancestría en 967 sujetos con DT2 y 343 donadores, participantes del estudio y poblaciones ancestrales	50
Anexo 3. Carta de Aprobación Comisión Nacional de Investigación Científica.....	51
Anexo 4. Licencia Sanitaria 06 T 09 006 0074	52
Anexo 5. Comisión de autorización sanitaria de entrada-salida 04930200377	53
Anexo 6. Fases de pre-procesamiento, análisis de control de calidad y detección de CNVs	54

Resumen

Ascencio Montiel IJ, Valladares-Salgado A, Cruz-López M. **Estudio de Variantes del Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2.**

En la actualidad, la diabetes tipo 2 (DT2) es considerada un problema de salud mundial. Existe evidencia que muestra que la enfermedad tiene un fuerte componente hereditario, lo cual han generado esfuerzos importantes para descifrar los factores genéticos que contribuyen en la génesis de la enfermedad. Los estudios de asociación se han enfocado hasta el momento al estudio de SNPs, sin embargo las Variantes en el Número de Copias (CNVs) permanecen poco estudiados.

Objetivo: Medir y comparar la frecuencia de las CNVs en sujetos con DT2 y sujetos no diabéticos (donadores del Banco Central de Sangre), residentes de la Ciudad de México, genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*.

Material y método: Estudio comparativo de 2 grupos realizado utilizando la información de las muestras de sujetos mexicanos, participantes del proyecto de investigación "Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes", genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*

Previo al análisis estadístico, se realizó un análisis de control de calidad al conjunto inicial de 1310 muestras con el fin de garantizar que la detección de CNVs fuera consistente entre los grupos de comparación. La detección de CNVs se realizó mediante los algoritmos iPattern, Birdsuite y PennCNV. Se seleccionaron los CNVs con tamaño $\geq 100\text{kb}$.

Se calcularon las frecuencias absolutas y relativas de dichos CNVs en los grupos de estudio y las diferencias fueron analizadas con las pruebas χ^2 o Prueba exacta de Fisher. Se calcularon las diferencias de proporciones crudas (DP) y ajustadas (DPa) por sexo, edad, índice de masa corporal y ancestría, utilizando un modelo de regresión logística no condicional. Un valor de $p < 0.05$ fue considerado como estadísticamente significativo.

Resultados: Un total de 880 muestras (686 DT2 y 194 donadores) y 4,023 CNVs finalizaron la fase de control de calidad y detección de CNVs. Del total de 4,023 CNVs que finalizaron el control de calidad, 3,182 fueron génicos. Se encontró una distribución similar de duplicaciones y deleciones entre los grupos de comparación. Las duplicaciones presentaron un tamaño mayor en comparación con las deleciones (223kb vs 129kb).

La frecuencia de duplicaciones fue de 89.4% en el grupo de DT2 y de 90.2% en el grupo donador, mientras que las frecuencias de deleciones fueron de 91.5% y 96.4% respectivamente. Se observó una menor proporción de muestras con deleciones en el grupo de DT2 (DP -4.85%, $p=0.020$), dicha diferencia permaneció significativa al ajustarse por las variables edad, sexo, índice de masa corporal y ancestría (DPa -5.16%, $p=0.003$). Se encontraron 3 regiones génicas con diferencia en las proporciones de CNVs entre los grupos de estudio (una región duplicada en el gen *NPIP* y dos regiones deletadas en los genes *ANKRD20A2* y *DUSP22*)

Conclusiones: La frecuencia de CNVs en los sujetos con DT2 y los sujetos no diabéticos fue mayor al 84%. Se observó una menor frecuencia de deleciones en el grupo de DT2 en comparación con el grupo de donadores. Las regiones génicas *DUSP22*, *NPIP* y *ANKRD20A2* presentaron frecuencias diferentes en los grupos de comparación.

1. ANTECEDENTES CIENTÍFICOS

La Diabetes Mellitus es un desorden metabólico caracterizado por hiperglucemia crónica con alteraciones en el metabolismo de carbohidratos, lípidos y proteínas debido a un defecto en la secreción y/o acción de la insulina (1).

La epidemia de la Diabetes Mellitus es reconocida por la Organización Mundial de la Salud, como una amenaza mundial. Se calcula que en el planeta existen más de 346 millones de personas con diabetes y 3.4 millones de muertes, y se prevé que ésta última cifra aumente a más del doble para el año 2030 (2).

La DT2 es un problema de Salud Pública en México. Se ha descrito una mayor letalidad de la DT2 en mexicanos (3); y la prevalencia del padecimiento ha sido estimada en 14.42% por la Encuesta Nacional de Salud Nutrición (ENSANUT) 2006, lo que representa que 7.3 millones de mexicanos, mayores de 20 años padecen la enfermedad (4). La información nacional más reciente sobre la prevalencia de DT2 se encuentra en la ENSANUT 2012 quien la halló en 9.2% por diagnóstico médico previo (mayor con respecto al 7% de prevalencia de diabetes por diagnóstico médico previo en la ENSANUT 2006) (5)

Existe evidencia clara de la importancia del ambiente en la ocurrencia de la DT2. Los factores que se han asociado a la presencia de la DT2 en población mexicana han sido la edad, la escolaridad baja, el antecedente familiar de diabetes mellitus, la coexistencia de enfermedad renal o hipercolesterolemia y la obesidad abdominal en mujeres (6,7).

La genética de la DT2

La DT2 es una enfermedad compleja, cuya patogénesis esta influenciada por factores genéticos y ambientales; y que se explica mediante dos mecanismos fisiopatológicos principales: la reducción de la secreción de insulina (ya sea debida a disfunción de las células beta pancreáticas o por reducción de la masa celular) y la resistencia a la insulina (8).

La alta concordancia entre gemelos y la diferencia en las tasas de desarrollo de DT2 entre hijos de padres con y sin diabetes (70% si ambos padres son diabéticos y 40% si uno de ellos lo es), han hecho sospechar la influencia de factores genéticos en la ocurrencia de la enfermedad (9).

Los genes pueden influenciar el desarrollo de la DT2 en dos vertientes: alterando directamente la secreción o acción de la insulina, o bien modificando la interacción del individuo con el ambiente; de igual forma, se ha documentado la modulación de la expresión de genes por factores ambientales y por la obesidad (10,11). Así, se ha demostrado un mayor riesgo para el desarrollo de diabetes en portadores de variantes de nucleótido único (SNP, *Single nucleotide polymorphism*) en los genes *FTO* y *MC4R* con una dieta inadecuada (12), así como disminución del riesgo para DT2 en portadores de variantes *TCF7L2* sometidos a intervenciones en el estilo de vida (13).

El descubrimiento de genes asociados a la diabetes ha seguido tres grandes metodologías: los análisis de ligamiento basados en familias (que permitieron el descubrimiento de genes asociados a formas mendelianas de diabetes), los estudios

de asociación (dirigidos a genes candidatos) y los estudios de asociación en todo el genoma (*Genome Wide-Association Studies, GWAS*) (14)

La disponibilidad de cantidades considerables de muestras de DNA de pacientes, así como la accesibilidad a tecnologías de genotipado de alta precisión (como los microarreglos), han consolidado a los GWAS como la metodología más exitosa para el descubrimiento de variantes genéticas y enfermedades. Por ejemplo, los primeros GWAS permitieron identificar al gen *TCF7L2* como asociado a DT2, cuya función en la modulación de la función pancreática fue demostrada con posterioridad (15).

Los GWAS han permitido hasta el momento, describir alrededor de 76 *loci* de susceptibilidad para la DT2 entre los que se encuentran los genes *CKAL1*, *CKN2A*, *CDKN2B*, *MTNR1B*, *TCF7L2*, *KCNJ11* (relacionados con la reducción en la secreción de la insulina) *FTO*, *IRS1* y *PPARG* (asociados a la resistencia a la insulina) (16,17, 18).

Del total de *loci* asociados a la DT2, 31 se reportan en la base de datos *Online Mendelian Inheritance in Man* (asociados a la DT2, número de fenotipo MIM 125853) [Tabla 1] (19). Los GWAS se han enfocado en el estudio de SNPs, sin embargo los datos de genotipificación pueden permitir el análisis de CNVs (20)

Tabla 1. Loci asociados a DT2

	Gene/Locus	Nombre del gen	Localización	Gene/Locus
1	<i>GPD2</i>	<i>Glycerol-3-phosphate dehydrogenase 2</i>	2q24.1	138430
2	<i>NEUROD1</i>	<i>Neuronal differentiation 1</i>	2q31.3	601724
3	<i>IRS1</i>	<i>Insulin receptor substrate 1</i>	2q36.3	147545
4	<i>PPARG</i>	<i>Peroxisome proliferator-activated receptor gamma</i>	3p25.2	601487
5	<i>IGF2BP2</i>	<i>Insulin-like growth factor 2 mRNA binding protein 2</i>	3q27.2	608289
6	<i>WFS1</i>	<i>Wolfram syndrome 1 (wolframin)</i>	4p16.1	606201
7	<i>NIDDM4</i>	<i>Diabetes mellitus, noninsulin-dependent</i>	5q34-q35.2	608036
8	<i>CDKAL1</i>	<i>CDK5 regulatory subunit associated protein 1-like 1</i>	6p22.3	611259
9	<i>HMGA1</i>	<i>High mobility group AT-hook 1</i>	6p21.31	600701
10	<i>ENPP1</i>	<i>Ectonucleotide pyrophosphatase/phosphodiesterase 1</i>	6q23.2	173335
11	<i>GCK</i>	<i>Glucokinase (hexokinase 4)</i>	7p13	138079
12	<i>PAX4</i>	<i>Paired box 4</i>	7q32.1	167413
13	<i>SLC30A8</i>	<i>Solute carrier family 30 (zinc transporter), member 8</i>	8q24.11	611145
14	<i>TCF7L2</i>	<i>Transcription factor 7-like 2 (T-cell specific, HMG-box)</i>	10q25.2-q25.3	602228
15	<i>KCNJ11</i>	<i>Potassium inwardly-rectifying channel, subfamily J, member 11</i>	11p15.1	600937
16	<i>ABCC8</i>	<i>ATP-binding cassette, sub-family C (CFTR/MRP), member 8</i>	11p15.1	600509
17	<i>MAPK8IP1</i>	<i>Mitogen-activated protein kinase 8 interacting protein 1</i>	11p11.2	604641
18	<i>MTNR1B</i>	<i>Melatonin receptor 1B</i>	11q14.3	600804
19	<i>HNF1A</i>	<i>HNF1 homeobox A</i>	12q24.31	142410
20	<i>IPF1</i>	<i>Insulin promoter factor</i>	13q12.2	600733
21	<i>IRS2</i>	<i>Insulin receptor substrate 2</i>	13q34	600797
22	<i>LIPC</i>	<i>Lipase, hepatic</i>	15q21.3	151670
23	<i>SLC2A4</i>	<i>Solute carrier family 2 (facilitated glucose transporter), member 4</i>	17p13.1	138190
24	<i>HNF1B</i>	<i>HNF1 homeobox B</i>	17q12	189907
25	<i>GCGR</i>	<i>Glucagon receptor</i>	17q25.3	138033
26	<i>RETN</i>	<i>Resistin</i>	19p13.2	605565
27	<i>AKT2</i>	<i>V-AKT murine thymoma viral oncogene homolog 2</i>	19q13.2	164731
28	<i>NIDDM3</i>	<i>Noninsulin-dependent diabetes mellitus 3</i>	20q12-q13.1	603694
29	<i>HNF4A</i>	<i>Hepatocyte nuclear factor 4, alpha</i>	20q13.12	600281
30	<i>PTPN1</i>	<i>Protein tyrosine phosphatase, non-receptor type 1</i>	20q13.13	176885

La diversidad genética en la población mexicana

La población mexicana se caracteriza por alto grado de diversidad genética. La mayoría de población mexicana es mestiza, es decir, se ha formado de una mezcla de nativos de México, migrantes europeos (principalmente de España, que arribaron durante la conquista de México en el Siglo XVI) y Africanos del Este (traídos a México durante los siglos XVI y XVIII).

Las contribuciones de ancestralidad que han sido estimados para el componente americano nativo oscilan entre 49% y 65%, para el europeo entre 30% y 45.8% y para el africano entre 1.8 y 6%, evidenciándose diferencias en la estructura ancestral de acuerdo a la región geográfica del país estudiada (21-25).

Así, ante las diferencias de los componentes ancestrales en la población mexicana, se ha propuesto una definición de "mexicano" que considera el haber nacido en México, el hablar el idioma Español y el poseer antecedente de 3 generaciones de haberse instalado en el país (25).

La medición de la ancestría en estudios genéticos en poblaciones con mezclas étnicas es necesaria, dado que la estratificación constituye un factor de confusión para las asociaciones genéticas (26)

Los métodos para estimar la distribución de los valores de ancestría en una población con mezclas raciales, utilizan modelos estadísticos, basados en marcadores informativos de ancestría (MIAs), los cuales presentan diferencias extremas en las frecuencia alélicas de dos o más poblaciones ancestrales de referencia de las cuales proviene la población bajo estudio (27).

El programa ADMIXMAP se basa en el método de Admixture Mapping el cual utiliza dos modelos (frecuentista y bayesiano) para la estimación individual de ancestría (27). A diferencia del programa ANCESTRYMAP (28), que únicamente se restringe a *loci* dialélicos, y del programa STRUCTURE (29) que requiere incluir correlaciones entre alelos en subpoblaciones cuyos marcadores no son informativos, el programa ADMIXMAP permite el análisis de *loci* multialélicos sin asumir correlaciones entre frecuencias alélicas en las subpoblaciones de estudio. La estructura de la población se expresa en diferentes vectores mediante un escalamiento multidimensional (MDS *multidimensional scaling*) tomando como referencia poblaciones ancestrales de comparación (30).

La genética de la DT2 en México

Las diferencias étnicas entre la población mexicana y las poblaciones de origen europeo, en quienes se ha realizado la mayor parte de los GWAS, han evidenciado diferencias en las señales genéticas asociadas a DT2.

El primer GWAS para DT2 realizado en Mexicanos, demostró replicación de únicamente 4 regiones génicas del total de las regiones reportadas en la literatura (*HNF1A*, *KCNQ1*, *CDKN2A*, *CDKN2B*) (31).

Hasta el momento, se han descrito 18 genes *loci* asociados a DT2 en población mexicana obtenidos en su mayoría en estudios comparativos y de casos y controles (32-51) [Tabla 2].

Tabla 2. *Loci* asociados a DT2 en población Mexicana

No	Gen/locus	Nombre del gen	Localización	Asociación	Ref.
1	<i>ACE</i>	Enzima convertidora de angiotensina I	17q23.3	Nefropatía diabética	32
2	<i>APOE</i>	Apolipoproteína E	19q13.2	Edema Macular	33
3	<i>CAPN10</i>	Calpaina 10	2q37.3	DT2	34
4	<i>CDKN2AIPNL</i>	Proteína N-terminal de interacción CDKN2A	5q31.1	DT2	31
5	<i>CDKN2A</i>	Inhibidor de ciclina dependiente de cinasa 2A	9p21	DT2	31
6	<i>HLADR7</i>	HLA-DR7	6p21.3	Retinopatía diabética	35
7	<i>HLADB1</i>	HLADB1	6p21.3	DT2 y enfermedad renal	36,37
8	<i>HNF1A</i>	HNF1 homeobox A	12q24.2	DT2 y niveles de lípidos	31, 38,39
9	<i>IRS1</i>	Sustrato 1 de receptor de insulina	2q36	DT2	40,41
10	<i>IGF2BP2</i>	Proteína de unión a mRNA del factor de crecimiento similar a insulina	3q27.2	DT2	31
11	<i>LTBP1</i>	Proteína de unión al factor de crecimiento transformante beta 1 latente	2p22-p21	Nefropatía diabética	42
12	<i>LOC387761</i>	LOC387761	11p12	DT2	43
13	<i>MGEA5</i>	Antígeno 5 de expresión en meningioma	10q24.1-q24.3	DT2	44
14	<i>KCNQ1</i>	Miembro 1 de la subfamilia KQT de canales de K voltaje-dependiente	11p15.5	DT2	31,37
15	<i>SLC16A11</i>	Miembro 11 de la familia de acarreadores de solutos 16	17p13.1	DT2	51
16	<i>TCF7L2</i>	Factor de transcripción 7 similar a 2	10q25.3	DT2 y triglicéridos	39,45,46,47
17	<i>TNF</i>	Factor de necrosis tumoral	6p21.3	DT2	48,49
18	10q	-	10q	DT2	50

Recientemente un meta-análisis de 26,488 casos de DT2 y 83,964 controles de origen Europeo, de Asia del Este, de Asia del Sur, Mexicano y México- Americano demostró 7 *loci* asociados a Diabetes independientemente de su ancestría (*TMEM154*, *SSR1-RREB1*, *FAF1*, *POU5F1-TCF19*, *LPP*, *ARL15*, *MPHOSPH9*) (18)

Las Variantes del Número de Copias

Las variantes del Número de Copias (*Copy Number Variation, CNVs*) se definen como segmentos de DNA de al menos 1kb de tamaño, cuya diferencia en número de copias se evidencia al comparar con dos o más genomas, y pueden describirse como duplicaciones (ganancia de número de copias) o deleciones (pérdida de número de copias) (52)

Hasta el momento, se ha estimado que el 71.48% del genoma presenta algún tipo de Variación en el número de copias (53).

Los CNVs pueden generar enfermedad por efectos cuantitativos o efectos disruptivos. Las deleciones pueden causar hemicigocidad o haploinsuficiencia (en genes dosis-sensibles); y las duplicaciones pueden generar un aumento en la producción de una proteína (efecto de dosis). Los efectos disruptivos se ocasionan por la desregulación de la expresión de un gen (por ejemplo la falta de expresión de una proteína por una deleción en regiones promotoras) (54,55).

El mayor efecto del descubrimiento de los CNVs en la medicina se ha dado en el laboratorio, optimizando la precisión del diagnóstico del cariotipo en enfermedades cromosómicas, y mejorando el diagnóstico de enfermedades previamente tamizadas mediante SNPs (por ejemplo, en la detección de fibrosis quística, los casos negativos a las pruebas habituales han mostrado deleciones largas en el gen *CFTR*) (54,56). La aplicación clínica de los CNVs actualmente constituye un reto clínico; bajo ciertas circunstancias los CNVs que se han asociados a enfermedades complejas, tiene potencial aplicación en la confirmación de un diagnóstico clínico existente, o bien en el tamizaje de población abierta que permita el diagnóstico oportuno de casos con antecedentes familiares del padecimiento (57).

Los microarreglos en la detección Variantes del Número de Copias.

La detección de CNVs se ha convertido en una disciplina por sí misma y una parte importante de los estudios genéticos de susceptibilidad a diferentes enfermedades tales como autismo, esquizofrenia, neuroblastoma, psoriasis, cáncer, entre otras.

La identificación de CNVs a través de todo el genoma puede realizarse mediante secuenciación o bien mediante microarreglos. El uso de los microarreglos ha sido el instrumento más utilizado y que ha permitido avanzar en el conocimiento de los CNVs.

Los microarreglos consisten en superficies sólidas que contienen fragmentos de DNA en parrillas microscópicas en las cuales se realiza la hibridación entre la sonda del arreglo y el DNA en estudio, generando una fluorescencia que indica el nivel de expresión del gen.

Existen dos tipos de microarreglos para la detección de CNVs: los microarreglos de hibridación genómica comparativa (*comparative genomic hybridization arrays, CGH*) y los microarreglos de SNPs (*SNP arrays*). Los arreglos CGH pueden detectar CNVs mediante sondas de oligonucleótidos; en tanto los arreglos de SNPs tienen la ventaja de permitir la detección de CNVs junto con los datos de genotipificación de SNPs (54). Sin embargo, los arreglos más actuales contienen tanto sondas para la detección de SNPs como sondas para detección de CNVs

El microarreglo *Affymetrix Genome-Wide Human SNP array 5.0* contiene un total de 1,340,037 sondas: 500,568 para SNPs (Polimorfismos de un solo nucleótido), 340,742 para CNVs y 6,113 sondas de control, que permiten la detección de SNPs y CNVs a lo largo de los 22 pares de cromosomas y los cromosomas sexuales (58). Aunque se trató de una versión intermedia entre el microarreglo *Affymetrix 6.0* y el 500K, tiene la ventaja, sobre éste último, de tener cobertura con sondas para detección de CNVs.

Proceso de detección de CNVs

El procedimiento de detección de CNVs incluye diferentes fases.

Fase de pre-procesamiento

En la fase de pre-procesamiento, la información de la intensidad de señal contenida en el microarreglo es transformada en valores del Log R ratio (*LRR*) y de la frecuencia del alelo B (*BAF*, *B Allele Frequency*). El *LRR* corresponde a la medida normalizada de la totalidad de la intensidad de señal para dos alelos de un SNP; mientras que la *BAF* es la medida normalizada de la relación de intensidad alélica de dos alelos. Los valores de *LRR* y *BAF* se usan para inferir la presencia de variaciones en el número de copias (59). Una tasa de genotipificación <95% o de control de intensidad <86% indican una mala calidad del genotipado del microarreglo. En el caso del microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*, la fase de pre-procesamiento se realiza mediante el uso de las herramientas de *Affymetrix (Affymetrix Power Tools)* (60).

Fase de análisis de control de calidad

La fase de análisis de control de calidad se realiza con la finalidad de reducir la heterogeneidad debido al tratamiento de muestras y de disminuir el número de detecciones de CNVs falsas positivas, debido a variaciones en la intensidad de la señal. Se utilizan los valores de la desviación estandar (DE) de *BAF* y de *LRR* así como el factor de variación (*waviness factor*) y la variación del *BAF (BAF drift)*. Los puntos de corte para eliminar muestras varían entre los tipos de microarreglos, los diferentes estudios, y deben individualizarse en lote de muestras. Para el caso del microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*, los puntos de corte descritos han sido <0.20 para la DE de *LRR*, *WF* entre -0.05 y +0.05, y *BAF drift* de <0.01 (61).

Adicional a estas mediciones, un número alto de CNVs por muestra puede indicar heterogeneidad y posiblemente mala calidad de la muestra; también los CNVs con tamaño bajo (<5kb) y detectados por pocas sondas (<5 sondas) pueden ser falsas detecciones (62).

Fase de detección de CNVs

Una vez seleccionadas las muestras con calidad adecuada, se realiza la detección de CNVs mediante programas estadísticos. Las herramientas estadísticas para la detección de CNVs y las características de los diferentes microarreglos se enlistan en la Tabla 3 (63).

Tabla 3. Características de las diferentes plataformas de microarreglos y las herramientas de detección de CNVs

Tipo de arreglo	Plataforma	Número de Sondas	Distancia promedio entre sondas	Herramienta de detección de CNVs
Hibridación genómica comparativa (CGH)	Sanger WGTP	29,043	170,000 pb	CNV finder, Nexus Copy Number
	Agilent 244K	236,381	60 pb	ADM-2, Nexus Copy Number
	Agilent 2x244K	462,609	60 pb	ADM-2, Nexus Copy Number
	NimbleGen 720K	720,412	60 pb	Nexus Copy Number
	NimbleGen2.1M	2,161,679	60 pb	Nexus Copy Number
SNP	Affymetrix 500K	500,568	25 pb	dCHIP, Partek Genomics Suite, GTC
	Illumina 650Y	660,918	50 pb	cnv Partition, PennCNV, QuantiSNP
SNPs + sondas para CNV	Affymetrix 5.0	1,340,037	25 pb	PennCNV, Birdsuite, iPattern
	Affymetrix 6.0	1,852,600	25 pb	dCHIP, Partek Genomics Suite, GTC, Birdsuite, iPattern
	Illumina 1M	1,072,820	50 pb	cnv Partition, PennCNV, QuantiSNP, iPattern
	Illumina 660W	657,366	50 pb	cnv Partition, PennCNV, QuantiSNP, iPattern
	Illumina Omni	1,140,419	50 pb	cnv Partition, PennCNV, QuantiSNP, iPattern

GTC: genotyping console.

Modificado de Pinto, et al. Nat Biotechnol. 2011

Las herramientas de detección de CNVs, específicas para el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0* son PennCNV, Birdsuite y iPattern.

El programa PennCNV utiliza un modelo de cadenas ocultas de Markov que utiliza información sobre la intensidad de las sondas, para la detección de CNVs (64). En tanto, Birdsuite es un paquete compuesto por cuatro programas (*Canary*, *Birdseed*, *Birdseye* y *Fawkes*) para la detección de CNVs utilizando información de genotipificación de SNPs como de CNVs (65). El programa iPattern utiliza algoritmos similares al programa Birdsuite (62).

A pesar que existen otras herramientas para la detección de CNVs, tal como CNVTools (66) QuantiSNP (67) o dChip (68); se ha evidenciado que la detección de CNVs se comporta mejor en herramientas desarrolladas específicamente para ciertas plataformas, que aquellas independientes de plataformas (63). De esta manera, con motivo del protocolo se utilizarán los programas PennCNV, Birdsuite y iPattern.

Dada la variabilidad en ciertas regiones del genoma, se excluyen a los CNVs localizados en regiones pseudoautosómicas, de inmunoglobulinas y localizadas en regiones gap. El tamaño de los CNVs que se reporta habitualmente es el de 100Kb o más (62,69). El presente protocolo se enfocará en este tipo de CNVs.

Fase de análisis de los CNVs y validación.

Una vez obtenidos los CNVs, el programa PLINK permite calcular las frecuencias de los CNVs y las diferencias en las proporciones de dichos CNVs. Adicionalmente puede realizarse la verificación de la presencia de los CNVs mediante visualización directa o mediante validación con PCR (*Polymerase Chain Reaction*) (62). El programa PennCNV permite la visualización de regiones mediante imágenes y utilizando la intensidad de señal (64).

Estudios de Variantes del Número de Copias y DT2

Se cuentan con escasos estudios que evalúan la asociación entre los CNVs y la DT2 [Tabla 4].

En 2009, Shtir buscó asociación entre la DT2 y CNVs en una submuestra de 194 pacientes con glucosa >126mg/dL y 213 controles del *Framingham Heart Study*, genotipificados con el microarreglo *Affymetrix Mapping 500K*. No se encontró asociación entre regiones de CNVs y una correlación pobre entre los algoritmos de detección (70).

En otro estudio, Wang y colaboradores no hallaron asociación entre 5 CNVs de la región 1q, utilizando el ensayo de PCR tiempo real cuantitativa, en 128 casos diabéticos y 128 controles de origen Americano-Europeo y Americano-Africano (71).

En 2010, Jeon publicó asociación entre DT2 y el número de copias <2 en el gen *LEPR* (*Odds Ratio*, OR 1.92), en un estudio que incluyó 137 casos y 258 controles de origen coreano genotipificados con el microarreglo *Affymetrix Mapping 50K* (72).

En estudios conducidos en población asiática y publicados en 2011, Kudo reportó la asociación entre delección de la región 4p16.3 y la DT2 de inicio temprano en 100 casos y 100 controles de origen japonés (73); en tanto Bae, en un estudio de 275 casos y 496 controles de origen coreano, reportó 3 regiones asociadas (15q21.1, 22q11.22 y 22q11.22) a la DT2 (74). Los estudios utilizaron microarreglos *Illumina (deCODE-Illumina CNV 370K Bead Chip y Illumina HumanHap 300 BeadChip* respectivamente).

En un estudio publicado en *Nature* en 2010, se encontró asociación entre la región de número variable CNVR5583.1 (rs1798090) localizada en el gen de tetraspanina 8 (*TSPAN8*) 12q14.1 y la DT2. El estudio se llevó a cabo utilizando aproximadamente 2000 casos de DT2 del *Wellcome Trust Case Control Consortium* y aproximadamente 3000 controles provenientes de la Cohorte Británica de 1958 y del Servicio de Sangre del Reino Unido (75). De manera similar, utilizando las mismas muestras de pacientes genotipificadas con el microarreglo *Affymetrix Mapping 500K*, Chen halló tres regiones asociadas a DT2 (1p34.3, rs16824514; 1q41, rs337147; y 19q13.2, rs2016070) utilizando una nueva estrategia de análisis de asociaciones de CNVs en GWAS (76).

En un estudio publicado en 2011 por Grassi, se encontró asociación entre la retinopatía diabética severa y la región CNVR6685.1 de 28.5 Mb, localizada en el intrón del gen *CCDC101* (valor de p 3.4×10^{-6}) en el cromosoma 16p11.2. El estudio incluyó 281 sujetos con retinopatía diabética y 1856 controles (77). Los participantes fueron de origen caucásico y presentaban diabetes tipo 1; se obtuvieron de los estudios EDIC (Epidemiology of Diabetes Intervention and Control Trial) y GoKinD (Genetics of Kidney in Diabetes). Se utilizó la plataforma *Affymetrix 5.0* e *Illumina HumanHap550*.

En 2011 Irvin publicó un estudio realizado en 1040 afroamericanos sin diabetes para buscar la asociación entre los niveles de insulina y la resistencia a la insulina (medida por el HOMA-IR) con los CNVs. Se realizó la genotipificación con el microarreglo *Affymetrix 6.0*. Se obtuvieron los CNVs mediante el programa Birdsuite. Para los niveles de insulina se asociaron dos regiones (rs10277702 y rs361367 del gen *TCRVB*; rs12552047 en cromosoma 9); mientras que para la resistencia a la insulina los 4 marcadores (rs10277702 y rs361367 del gen *TCRVB*; rs13003829 del

gen *ARHGEF4*; y rs12509348 del gen *DCK*). Así, las asociaciones más significativas se dieron con el gen *TCRVB* (78)

En 2012, Plengvidhya condujo un estudio de casos y controles en 305 pacientes con DT2 y 250 controles del Hospital Siriraj Hospital, de Bangkok, Tailandia para buscar CNVs en la región del gen *CAPN10*, utilizando PCR múltiple y cromatografía líquida de alta resolución. No se encontraron diferencias en la presencia de duplicaciones o deleciones entre casos y controles (3.9 vs 2.9, 1.3 vs. 0.3 $p = 0.692$); sin embargo el indel19 se encontró asociado con la DT2 en el modelo recesivo (OR: 0.66; 95%CI: 0.47-0.94, $p = 0.022$) (79)

Finalmente, en un estudio realizado en 1677 sujetos Mexicanos Americanos en San Antonio, Texas genotificados con el microarreglo Illumina Infinium Beadchip, se estudió la heredabilidad de 2937 CNVs detectados mediante los algoritmos PennCNV y QuantiSNP. Se encontraron regiones con bajas frecuencias (0.2%) y 765 previamente no reportadas (80).

Tabla 4. Antecedentes científicos de estudios de CNVs y DT2

Autor, año	n (casos controles)	Población	Hallazgos	Ref
Shtir, 2009	407 (194 213)	Caucásica.	No asociación.	70
Wang, 2009	256 (128 128)	Africano/Caucásico-Americano	No asociación con región 1q	71
Jeon, 2010	395 (137 258)	Coreano	Delección Gen <i>LEPR</i>	72
Kudo, 2011	200 (100 100)	Japonés	Región 4p16.3	73
Bae, 2011	771 (275 496)	Coreano	Regiones 15q21.1, 22q11.22 y 22q11.22	74
WTCCC, 2010	5,000 (2000 3000)	Caucásica	Región 12q14.1, CNVR5583.1 (rs1798090) <i>TSPAN8</i>	75
Chen, 2010	5,000 (2000 3000)	Caucásica	1p34.3, (gen <i>INPP5B</i>), 1q41, gen <i>MOSC2</i>) y 19q13.2	76
Grassi, 2011	2137 (281 1856)	Caucásica	Retinopatía diabética y CNVR6685.1 gen <i>CCDC101</i>	77
Irvin, 2011	1,040	Africana-Americana	Asociación gen <i>TCRVB</i> y Resistencia a la insulina	78
Plengvidhya, 2012	555 (305 250)	Tailandesa	No asociación CNVs - DT2 en gen <i>CAPN10</i>	79
Blackburn, 2013	1,677	Mexicana-Americana	765 CNVs no reportados previamente	80

El proyecto de investigación “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”.

El presente protocolo de Maestría se encuentra inmerso en la primera fase del proyecto de investigación titulado “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con DT2 y su implicación en las complicaciones más frecuentes” el cual, se encuentra en curso en la Unidad de Investigación Médica en Bioquímica del Hospital de Especialidades “Dr. Bernardo Sepúlveda” de Centro Médico Nacional Siglo XXI.

Las 3 fases de dicho proyecto son: 1) caracterización de CNVs, 2) validación de regiones de CNVs mediante PCR tiempo real cuantitativa y 3) búsqueda de asociaciones de CNVs y las complicaciones de la diabetes [Tabla 5].

Tabla 5. Fases del proyecto de investigación “Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”

Fase	Objetivo de la fase
I	Caracterizar CNVs
II	Validar regiones con CNVs
III	Buscar asociaciones entre CNVs y complicaciones de la Diabetes

El objetivo de la fase I y del presente proyecto, es la caracterización de CNVs mediante algoritmos de detección, en sujetos con DT2 y sujetos no diabéticos (donadores del Banco Central de Sangre de Centro México Nacional Siglo XXI), que han sido genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*. En las fases subsecuente se procederá a la validación de las regiones mediante PCR tiempo real cuantitativa y finalmente en la fase III se buscarán CNVs en grupos de sujetos diabéticos con complicaciones.

El proyecto general ha sido aprobado por la Comisión Nacional de Investigación Científica del Instituto Mexicano del Seguro Social y cuenta con el apoyo del Fondo Sectorial de Investigación en Salud y Seguridad Social del CONACYT.

Se disponen de 967 muestras de sujetos con DT2 y muestras de 343 donadores del Banco Central de Sangre genotipificadas con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*. [Para las características de obtención de los sujetos y procesamiento de las muestras, refiérase al Anexo 1]

Se dispone también de la información de ancestralidad en vectores MDS, obtenidos mediante el programa ADMIXMAP y basado en poblaciones europeas, Africanas y Amerindias; así como de información sobre sexo, edad, índice de masa corporal y edad al diagnóstico en cada uno de los sujetos participantes del estudio [Anexo 2].

Al analizar las características de la población, la mediana de edad de la población total es de 50 años siendo igual entre los sujetos con DT2 y los donadores. El 63.6% de los participantes fueron mujeres, siendo mayor la proporción de éstas en el grupo de DT2 (68.7% vs 49.3%) [Tabla 6]

Tabla 6. Características de los sujetos participantes del “Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes” genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*

Característica	Grupo DT2	Grupo Donador	Total	p
n	967	343	1310	
Sexo masculino	303 (31.3%)	174 (50.3%)	477 (36.4%)	
femenino	664 (68.7%)	169 (49.3%)	833 (63.6%)	<0.001
Edad, años	50 (45, 57)	50 (47, 54)	50 (46, 56)	0.784
Edad al diagnóstico, años	45 (40, 49)	ND	ND	ND
Tiempo de evolución, años	4 (1,10)	ND	ND	ND
Índice de masa corporal (kg/m ²)	29.1 (26.3, 32.5)	27.2 (25, 29.4)	28.5 (25.9, 31.7)	< 0.001

Los datos se presentan como número (%) ó mediana (percentil 25, percentil 75).

Valor de p entre los grupos mediante la prueba, χ^2 o U-Mann-Whitney. ND, no disponible

Con los datos del protocolo se ha generado GWAS, el cual ha sido publicado en la revista *Diabetología* (31).

Con fines de hipótesis, y utilizando el estudio de Jeon (72) que compara las frecuencias de CNVs entre sujetos con diabetes y sin diabetes, y dadas las similitudes en el microarreglo utilizado y entre las poblaciones mexicana y asiática, se propone la región del gen *LEPR* para el ensayo de hipótesis.

Sin embargo, cabe aclarar que el presente protocolo pretende realizar una búsqueda de CNVs en los 22 pares de cromosomas y el cromosoma X, de acuerdo a la cobertura de sondas del microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*.

De igual forma, y a pesar de la gran relevancia de los factores ambientales en la génesis de la DT2 y dado el avance actual del conocimiento respecto a los CNVs, el presente protocolo no contempla la inclusión de interacciones gen-ambiente, sin embargo incluye el ajuste de las variables confusoras disponibles en el GWAS: edad, sexo, índice de masa corporal y ancestría.

2. PLANTEAMIENTO DEL PROBLEMA

México es uno de los países con más alta prevalencia de DT2. La población Mexicana caracterizada por su alta proporción de mestizaje, constituye un grupo de estudio para marcadores genéticos de alto interés.

Los CNVs constituyen marcadores relativamente inexplorados para la DT2: se han realizado 8 estudios, todos ellos en poblaciones de origen europeo, asiático o africano y ninguno en población de origen latinoamericano.

Los CNVs localizados en regiones génicas pueden generar enfermedad por efectos cuantitativos o efectos disruptivos.

Si bien existe un reporte de la extensión, localización y distribución de los CNVs en población Mexicana Americana, no se ha explorado su impacto en la presencia de la DT2.

3. JUSTIFICACIÓN

En la actualidad, la DT2 es considerada un problema de salud mundial. Existe evidencia que muestra que la enfermedad tiene un fuerte componente hereditario, lo cual han generado esfuerzos importantes para descifrar los factores genéticos que contribuyen en la génesis de la enfermedad.

Los estudios de asociación se han enfocado hasta el momento al estudio de SNPs, sin embargo los CNVs permanecen poco estudiados. Se dispone de información de genotipificación de sujetos con DT2 y donadores del Banco Central de Sangre en el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*.

El estudio de las variantes en el número de copias en sujetos con DT2 y donadores del banco central de Sangre, permitirá conocer por primera vez las regiones del genoma que se encuentran en número variable en una muestra de sujetos mexicanos de la ciudad de México, además de dar indicios sobre genes posiblemente relacionados con fisiopatología de la enfermedad en nuestra población.

4. PREGUNTA DE INVESTIGACIÓN

En las muestras de sujetos mexicanos, residentes de la Ciudad de México, participantes del proyecto de investigación “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”, genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0* y posterior a un análisis de control de calidad:

- ¿Cuál es la frecuencia de las Variantes en el Número de Copias $\geq 100\text{kb}$, en los sujetos con DT2 y cuál en los sujetos no diabéticos (donadores del Banco Central de Sangre)?
- ¿Cuál es la diferencia en la frecuencia de las Variantes en el Número de Copias $\geq 100\text{kb}$, en los sujetos con DT2 y los sujetos no diabéticos (donadores del Banco Central de Sangre), ajustada por edad, sexo, índice de masa corporal y ancestría?

5. HIPÓTESIS

En las muestras de sujetos mexicanos, residentes de la Ciudad de México, participantes del proyecto de investigación “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”, genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0* y posterior a un análisis de control de calidad:

- La frecuencia de las Variantes en el Número de Copias $\geq 100\text{kb}$ para la región del gen *LEPR**, será de 63.7% en los sujetos con DT2, y de 47.7% en los sujetos no diabéticos (donadores del Banco Central de Sangre) (72)
- Existe una diferencia de 16% en la frecuencia de las Variantes en el Número de Copias $\geq 100\text{kb}$ para la región del gen *LEPR**, entre los sujetos con DT2 y los sujetos no diabéticos (donadores del Banco Central de Sangre, ajustada por edad, sexo, índice de masa corporal y ancestría) (72)

** Si bien el estudio realizará una búsqueda de CNVs en los 22 pares de cromosomas y el cromosomas X (de acuerdo a la cobertura de sondas del microarreglo Affymetrix Genome-Wide Human SNP array 5.0), con fines de prueba de hipótesis, se propone la región del gen LEPR publicada por Jeon en población coreana (69), por presentar la mayor similitud de los estudios de CNVs y DT2 a la población mexicana.*

6. OBJETIVOS

En las muestras de sujetos mexicanos, residentes de la Ciudad de México, participantes del proyecto de investigación “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”, genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0* y posterior a un análisis de control de calidad:

- Medir la frecuencia de las Variantes en el Número de Copias $\geq 100\text{kb}$ (detectados mediante los algoritmos iPattern, Birdsuite y PennCNV y posterior a un análisis de control de calidad), en los sujetos con DT2 y los sujetos no diabéticos (donadores del Banco Central de Sangre).
- Medir la diferencia en la frecuencia de las Variantes en el Número de Copias $\geq 100\text{kb}$ (detectados mediante los algoritmos iPattern, Birdsuite y PennCNV y posterior a un análisis de control de calidad) en los sujetos con DT2 y los sujetos no diabéticos (donadores del Banco Central de Sangre), ajustada por edad, sexo, índice de masa corporal y ancestría.

7. VARIABLES DEL ESTUDIO

La variable independiente del estudio correspondió a la presencia de DT2, mientras que la variable dependiente correspondió a las variantes del número de copias totales y génicas. Las variables sexo, edad, índice de masa corporal, ancestría fueron consideradas variables confusoras, mientras que las variables edad al diagnóstico de la DT2 y tamaño del CNV fueron consideradas variables descriptoras.

6.1 Variabes de estudio

a) Variable independiente: Diabetes tipo 2.

Definición conceptual: Tipo de diabetes en la que hay capacidad residual de secreción de la insulina, pero sus niveles no superan la resistencia a la insulina concomitante, insuficiencia relativa de secreción de insulina o cuando coexisten ambas posibilidades y aparece la hiperglucemia (81). Se definen por cualquiera de los 4 criterios establecidos por la Asociación Americana de Diabetes: 1) glucosa plasmática en ayunas $\geq 126\text{mg/dL}$, 2) glucosa plasmática a las 2h $\geq 200\text{mg/dL}$ en la curva de tolerancia a la glucosa, 3) glucosa plasmática casual $\geq 200\text{mg/dL}$ en presencia de síntomas clásicos de hiperglucemia o crisis hiperglucémica o 4) hemoglobina glicosilada $\geq 6.5\%$ (82)

Definición operacional: Presencia de diagnóstico de DT2 de acuerdo a los criterios de la Asociación Americana de Diabetes.

Tipo de variable: Cualitativa

Escala de medición: Dicotómica

Indicador: 1=grupo DT2, 0=grupo donador

b) Variable dependiente: Variante del número de copias (CNVs) $\geq 100\text{kb}$

Definición conceptual: Segmentos de DNA cuya diferencia en número de copias se evidencia al comparar con dos o más genomas (54)

Definición operacional: Región del genoma explorada en el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*, detectada por los dos algoritmos PennCNV (64) iPattern (65) y Birdsuite (66) posterior al análisis de control de calidad. Se obtendrá el número de copias (NC) del CNV como 0,1,2 ó 3. Se considera como número de copias normal NC=2 en autosomas y cromosoma X en mujeres y NC=1 en cromosoma X en hombres. La delección será definida si NC<2 en autosomas y cromosoma X en mujeres y NC=0 en cromosoma X en hombres; mientras en la duplicación si NC>2 en autosomas y cromosoma X en mujeres y NC>1 en cromosoma X en hombres, posteriores al análisis de control de calidad y con tamaño $\geq 100\text{kb}$.

Tipo de variable: Cualitativa

Escala de medición: Nominal

Indicador: 0=número de copias normal , 1=delección, 2=duplicación

6.2 Variables confusoras

a) Sexo

Definición conceptual: Condición orgánica que distingue mujer y hombre y les asigna una función específica en la generación (83).

Definición operacional: Sexo referido por el paciente durante su inclusión al estudio.

Tipo de variable: Cualitativa

Escala de medición: Dicotómica

Indicador: 1=hombre, 2=mujer

b) Edad

Definición conceptual: Cantidad de tiempo que ha pasado desde el nacimiento de una persona hasta el presente (84)

Definición operacional: Diferencia entre la fecha de inclusión al estudio y la fecha de nacimiento del paciente.

Tipo de variable: Cuantitativa

Escala de medición: Discreta

Indicador: Años

c) Ancestría

Definición conceptual: Conjunto de rasgos genéticos que se heredan a los descendientes y determinan la herencia étnica de una persona (84).

Definición operacional: Valor del vector 1 del escalamiento multidimensional (MDS), determinado por el método Admixture Mapping utilizando un panel de 3,500 marcadores informativos de ancestría (27) basados en información de estudios de genoma amplio de 6 poblaciones ancestrales (Nahua, Maya, Andino, Españoles, Europeos CEPH y Yoruba) (85) [Anexo 2]

Tipo de variable: Cuantitativa

Escala de medición: Continua

Indicador: valor del vector 1 del escalamiento multidimensional (MDS vector 1).

d) Índice de Masa Corporal

Definición conceptual: Cociente entre la peso y talla elevada al cuadrado, utilizado para identificar el sobrepeso y la obesidad en los adultos (86)

Definición operacional: Cociente entre el peso y la talla medida en el paciente en el momento su inclusión al estudio.

Tipo de variable: Cuantitativa

Escala de medición: Continua

Indicador: kg/m²

6.3 Variables descriptoras

a) Edad al diagnóstico de la DT2 (únicamente para los sujetos con DT2)

Definición conceptual: Cantidad de tiempo que ha pasado desde un momento dado en una persona hasta el presente (84)

Definición operacional: Diferencia entre la fecha del diagnóstico de la DT2 y la fecha de nacimiento del paciente.

Tipo de variable: Cuantitativa

Escala de medición: Discreta

Indicador: Años

b) Tamaño del CNV

Definición conceptual: Dimensión lineal o distancia comprendida entre dos puntos (84)

Definición operacional: Dimensión lineal, en pares de bases (kilobases o megabases), comprendido entre el inicio de un CNV y su final.

Tipo de variable: Cuantitativa

Escala de medición: Discreta

Indicador: Kilobases (kb) ó Megabases (Mb).

8. MATERIAL Y MÉTODO

8.1 Diseño de estudio

Se realizó un estudio comparativo de dos grupos.

Por la intervención del investigador: *observacional*

Por el número de grupos a estudiar: *comparativo*

Por el número de mediciones: *transversal*

Por la forma de recolección de la información: *retrolectivo*

Por la dirección: *retrospectivo*

8.2 Universo de estudio

El universo de estudio estuvo formado por los datos obtenidos de las muestras de sujetos mexicanos residentes de la Ciudad de México, participantes del proyecto de investigación “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes” y genotificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0* (31). Se dispuso de 967 muestras de sujetos con DT2 y 343 sujetos no diabéticos donadores del Banco Central de Sangre que cumplieron los criterios de selección.

8.3 Criterios de Selección

Se enlistan los criterios de selección de los sujetos incluidos en el proyecto de investigación principal; los criterios de eliminación fueron aplicados en el desarrollo del presente protocolo.

Criterios de inclusión

a) Para los sujetos con DT2

- Derechohabientes del Instituto Mexicano del Seguro Social con edad igual o mayor a 18 años.
- Con diagnóstico de DT2 de acuerdo a los criterios de Asociación Americana de Diabetes (*ver definición operacional de la variable independiente*)
- Usuarios de la Unidad de Medicina Familiar No. 4 de la Ciudad de México.
- Que acepten participar en el estudio.
- Con consentimiento informado por escrito con autorización para análisis genético y su empleo para estudios futuros.
- Mexicanos por nacimiento, que hablen el idioma Español y antecedente de 3 generaciones de haberse instalado en la ciudad de México.

b) Para los sujetos no diabéticos (donadores del Banco Central de Sangre)

- Donadores del Banco Central de Sangre de Centro Médico Nacional Siglo XXI
- Edad igual o mayor a 40 años
- Con glucosa plasmática en ayuno <100mg/dL.
- Sin antecedentes familiares en primer grado de DT2.
- Que acepten participar en el estudio.
- Con consentimiento informado por escrito con autorización para análisis genético y su empleo para estudios futuros
- Mexicanos por nacimiento, que hablen el idioma Español y antecedente de 3 generaciones de haberse instalado en la ciudad de México.

Criterios de no inclusión

a) Para ambos grupos

- Tasa de genotipificación < 95%
- Tasa de control de intensidad <86%

Criterios de eliminación

a) Para ambos grupos

- Que no cumplan los criterios de control de calidad:
 - Desviación estándar de la intensidad de señal >0.3
 - Desviación estándar del Log R ratio (LRR) ≥ 0.3
 - Desviación estándar de la frecuencia del Alelo B (BAF) ≥ 0.1
 - Variación de la frecuencia del alelo B > 0.01
 - Factor de variación (WF) ≤ -0.035 y $\geq + 0.035$
 - Numero de CNVs (≥ 5 kb y con ≥ 5 sondas) obtenidos mediante cada uno de los algoritmos por encima al percentil 95
 - Ausencia de CNVs con tamaño de 100kb o más

8.4 Tamaño de muestra

a) Tamaño de muestra para una proporción

Tomando en cuenta la región *LEPR* (72), utilizando la fórmula para el cálculo de tamaño mínimo de muestra en un estudio descriptivo de una variable dicotómica, un valor alfa de 0.05 (1.96), una amplitud total (W) de 0.15 y asumiendo una proporción esperada (P) de 0.637 para los casos y una (P) de 0.477 de , tenemos:

$$n = 4z_{\alpha}^2 P(1-P)/W^2$$

Para el grupo de DT2, desarrollando, tenemos:

$$n = [4(1.96)^2 \times 0.637(1-0.637)] / (0.15)^2 = [4 \times 3.8416 \times 0.637 \times 0.363] / 0.0225 = 3.5531 / 0.0225 = 158 \text{ sujetos}$$

Para el grupo de sujetos no diabéticos, desarrollando, tenemos:

$$n = [4(1.96)^2 \times 0.477(1-0.477)] / (0.15)^2 = [4 \times 3.8416 \times 0.477 \times 0.523] / 0.0225 = 3.8334 / 0.0225 = 170 \text{ sujetos}$$

b) Tamaño de muestra para una diferencia de proporciones

Tomando en cuenta la región *LEPR* (72), utilizando la fórmula para el cálculo de tamaño mínimo de muestra en para una diferencia de proporciones; y asumiendo una proporción de casos expuestos (P_1) 0.637, una proporción de controles expuestos (P_2) 0.477, y una proporción casos (q_1) de 0,783, un valor alfa de 0.05 ($z_\alpha = 1.96$) y un poder de 0.8 ($z_\beta = 0.84$), tenemos:

$$n = \{ [z_\alpha \sqrt{P(1-P)(1/q_1 + 1/q_2)}] + [z_\beta \sqrt{P_1(1-P_1)(1/q_1) + P_2(1-P_2)(1/q_2)}] \}^2 / (P_1 - P_2)^2$$

donde:

$$q_2 = 1 - q_1$$

$$P = q_1 P_1 + q_2 P_2$$

Desarrollando, tenemos:

$$q_2 = 1 - 0.738 = 0.262$$

$$P = (0.738 \times 0.637) + (0.262 \times 0.477) = 0.4702 + 0.1249 = 0.5951$$

$$n = \{ [1.96 \sqrt{0.5951(1-0.5951)(1/0.347 + 1/0.653)}] + [0.84 \sqrt{0.477(1-0.477)(1/0.347 + 0.637(1-0.637)(1/0.653))}] \}^2 / (0.477 - 0.637)^2$$

$$= \{ [1.96 \sqrt{1.246}] + [0.84 \sqrt{1.266}] \}^2 / (0.160)^2$$

$$= \{ [1.96 \times 1.1165] + [0.84 \times 1.1251] \}^2 / 0.0256$$

$$= \{ 2.188 + 0.9451 \}^2 / 0.0256 = \{ 3.134 \}^2 / 0.0256 = 9.8194 / 0.0256$$

$$= 383.6$$

Calculando el número de sujetos por grupo:

$$\text{Para el grupo de DT} = n \times q_1 = 383.6 \times 0.738 = 283.1$$

$$\text{Para el grupo de sujetos no diabéticos} = n \times q_2 = 383.6 \times 0.262 = 100.4$$

Así los tamaños de muestra estimados fueron: para una proporción de 158 sujetos con DT y 170 sujetos sin diabetes (total 328); mientras que para una diferencia de proporciones de 283 y 101 respectivamente (total 384).

Considerando un porcentaje de pérdidas de 10%, el tamaño de muestra final estimado para una proporción fue de 174 sujetos con DT2 y 187 sujetos sin diabetes (total 361); mientras que para la diferencia de proporciones fue de 311 y 111 respectivamente (total 422).

Técnica de muestreo: No probabilístico consecutivo.

8.5 Poder del tamaño de muestra

b) Poder para una diferencia de proporciones

Tomando en cuenta la región *LEPR* (72), utilizando la fórmula para el cálculo del poder para una diferencia de proporciones; y asumiendo una proporción de casos expuestos (P_1) 0.637, una proporción de controles expuestos (P_2) 0.477, un valor alfa de 0.05 ($z_\alpha = 1.96$) y un tamaño de muestra para casos (n_1) de 967 y para controles (n_2) de 343, tenemos:

$$Poder = \Phi\{\Delta / \sqrt{[(p_1q_1/n_1) + (p_2q_2/n_2)]}\} - z_\alpha\{\sqrt{p'q'}(1/n_1 + 1/n_2) / \sqrt{[(p_1q_1/n_1) + (p_2q_2/n_2)]}\}$$

donde:

$$q_1 = 1 - p_1$$

$$q_2 = 1 - p_2$$

$$\Delta = |p_2 - p_1|$$

$$p' = (n_1p_1 + n_2p_2) / (n_1 + n_2)$$

$$q' = 1 - p'$$

Desarrollando, tenemos:

$$q_1 = 1 - 0.637 = 0.363$$

$$q_2 = 1 - 0.477 = 0.523$$

$$\Delta = |0.637 - 0.477| = 0.16$$

$$p' = [(967 \times 0.637) + (343 \times 0.477)] / (967 + 343) = (615.97 + 163.61) / 331682 \\ = 779.59 / 331682 = 0.595$$

$$q' = 1 - 0.595 = 0.405$$

$$Poder = \Phi\{0.16 / \sqrt{[(0.637 \times 0.363 / 967) + (0.477 \times 0.523 / 343)]}\} - \\ 1.96\{\sqrt{0.595 \times 0.405} (1/967 + 1/343) / \sqrt{[(0.637 \times 0.363 / 967) + (0.477 \times 0.523 / 343)]}\}$$

$$Poder = \Phi\{0.16 / \sqrt{(0.0002391 + 0.0007273)}\} - 1.96\{\sqrt{0.24095 \times (0.0010 + 0.0029)} / \\ \sqrt{(0.0002391 + 0.0007273)}\}$$

$$Poder = \Phi\{0.16 / \sqrt{(0.0096)}\} - 1.96\{\sqrt{0.24095 \times 0.003949} / \sqrt{(0.0096)}\}$$

$$Poder = \Phi\{0.16 / 0.0311\} - 1.96\{\sqrt{0.0308} / 0.0311\} = 5.1467 - (-1.96 \times 0.9923)$$

$$Poder = \Phi(5.1467 - 1.9449) = 3.2017$$

$$Poder = \Phi(3.2017) = 0.9993$$

El poder estimado de la muestra para una diferencia de proporciones fue de 0.9993.

9. PLAN GENERAL: PROCEDIMIENTOS Y MÉTODOS

Al inicio del estudio se dispuso de 967 sujetos con DT2 y 343 sujetos donadores, genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0* el cual contiene 500,568 sondas para SNPs (Polimorfismos de un solo nucleótido) accesibles y 1,340,037 sondas totales (58).

El plan general de análisis de las muestras genotipificadas se llevó a cabo en las siguientes etapas [Figura 1]:

a) Fase de pre-procesamiento [Anexo 6].

Con el paquete estadístico *Affymetrix Power Tools* (60), se obtuvieron los valores de Log R ratio y frecuencia del alelo B por cada muestra, así como los valores de las desviaciones estándar, el WF y el BAF drif.

Se eliminaron las muestras con tasa de genotipificación <95% y control de intensidad <86%

b) Análisis de control de calidad [Anexo 6].

Se utilizaron los siguientes criterios para la eliminación de las muestras de mala calidad (59):

- Desviación estándar de la intensidad de señal >0.3
- Desviación estándar del Log R ratio (LRR) ≥ 0.3
- Desviación estándar de la frecuencia del Alelo B (BAF) ≥ 0.1
- Variación de la frecuencia del alelo B > 0.01
- Factor de variación (WF) ≤ -0.035 y $\geq + 0.035$

Posterior a la detección mediante los tres algoritmos, se eliminaron las muestras con

- Numero de CNVs (≥ 5 kb y con ≥ 5 sondas) obtenidos mediante cada uno de los algoritmos por encima al percentil 95
- Sin CNVs con tamaño de 100kb o más

c) Detección de CNVs [Anexo 6]

Con las muestras restantes se realizó la detección de CNVs utilizando los programas iPattern (62), PennCNV (64) y Birdsuite (65). Se obtuvo el número total de CNVs obtenidos, así como la mediana de CNVs por muestra, tanto para el total de CNVs detectados como para los CNVs con ≥ 5 kb y con ≥ 5 sondas.

Se observó la correlación entre los tres algoritmos utilizando la prueba de Spearman. Se generó una base de datos común integrando aquellos CNVs que hallan sido detectados por los algoritmos mejor relacionados.

Se eliminaron los CNVs localizados en regiones con alta variabilidad (regiones pseudoautosómicas, regiones de inmunoglobulinas, regiones gap, (de acuerdo a UCSC Genome Bioinformatics) y aquellos localizados en ambos brazos del cromosoma o catalogados como duplicaciones y deleciones al mismo tiempo por los algoritmos de detección.

Finalmente se seleccionaron los CNVs con tamaño mayor a 100kb.

d) Análisis de CNVs

El análisis de CNVs se llevó a cabo con el programa PLINK (87) calculando la proporción de muestras con CNVs (porcentaje de muestras con uno o más CNVs).

e) Visualización de regiones

Las regiones que presenten CNVs fueron visualizadas en todas las muestras, utilizando el programa PennCNV (64).

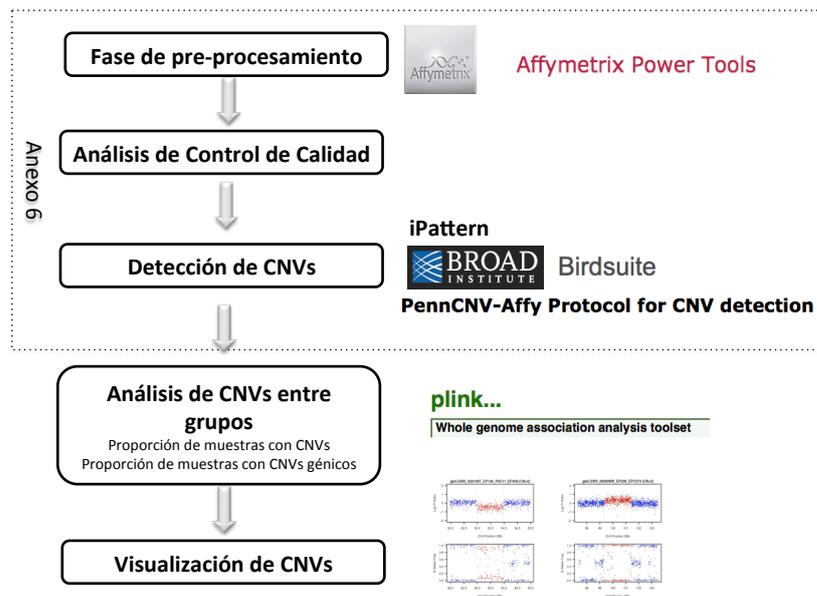


Figura 1: Fases para la caracterización de CNVs.

10. ANÁLISIS ESTADÍSTICO

Se realizó un análisis descriptivo de los datos. Las variables cualitativas se expresaron como número y porcentaje. Para las variables cuantitativas, se utilizó promedio \pm desviación estándar, o mediana (percentiles 25, 75), en caso que la distribución de las variables fuese normal o no. Se consideró como distribución normal, si la prueba Shapiro-Wilk presentó un valor de $p \geq 0.05$.

Para comparar la distribución de las variables descriptoras entre los grupos, se utilizaron las pruebas χ^2 o Prueba exacta de Fisher (para variables cualitativas) y t de Student o U de Mann-Whitney (para variables continuas), de acuerdo a su distribución.

Los CNVs fueron analizados en 2 grupos (totales y génicos) y tomando en cuenta tres categorías (duplicaciones, deleciones y duplicaciones + deleciones).

Un CNV fue considerado génico si su ubicación se intersectó con la localización de alguno de los genes según las anotaciones RefSeq versión Marzo 2006, NCBI v36, hg18 (disponible en <http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/>)

Se calcularon las frecuencias absolutas y relativas de dichos CNVs en los grupos de estudio.

Las diferencias entre los sujetos con DT2 y los sujetos sin diabetes con respecto los CNVs, fueron analizadas con las pruebas χ^2 o Prueba exacta de Fisher (para variables cualitativas). Se calcularon las diferencias de proporciones con Intervalos de Confianza al 95% (IC95%) crudas (DP) y ajustadas (DPa) por sexo, edad, índice de masa corporal y ancestría, utilizando un modelo de regresión logística no condicional.

Un valor de $p < 0.05$ fue considerado como estadísticamente significativo.

11. CONSIDERACIONES ÉTICAS

El protocolo de estudio fue aprobado por la Comisión Nacional de Investigación Científica del Instituto Mexicano del Seguro Social con el número de registro 2008-785-073 (Anexo 3).

El desarrollo del protocolo estuvo regido por los principios especificados en la Declaración de Helsinki y ha cumplido con los apartados de la normatividad nacional en Materia de Investigación para la Salud vigente.

Riesgos de la investigación: El protocolo constituyó una “Investigación de riesgo mínimo”, según los criterios del Reglamento de la Ley General de Salud en Materia de Investigación para la Salud, Título Segundo, Capítulo I, Artículo 17, Inciso II; dado que requirió la punción venosa para la extracción de sangre periférica.

Las disposición de las muestras de DNA por parte de la Unidad de Investigación Médica en Bioquímica fue autorizada por la Licencia Sanitaria 06 T 09 006 0074 (Anexo 4); en tanto el envío de muestras de ADN no representó riesgo sanitario, tal y como se explica en el oficio de comisión de autorización sanitaria de entrada-salida 04930200377 (Anexo 5), documentos emitidos por la Comisión Federal para la Protección contra Riesgos Sanitarios.

Contribuciones y beneficios para los participantes y la sociedad en su conjunto: La presente investigación no brindó ningún beneficio directo al participante, sin embargo el beneficio se da a la sociedad en su conjunto, al generar conocimiento sobre factores genéticos posiblemente involucrados en la DT2.

Confidencialidad: Los datos obtenidos del análisis genético fueron codificados con claves numéricas y protegidos por contraseñas, con la finalidad de evitar la identificación del sujeto, y asegurar el anonimato y confidencialidad de la información. Los sujetos recibieron los resultados de los estudios bioquímicos realizados como parte de la investigación y los participantes con resultados anormales fueron canalizados al servicio de Medicina Familiar. No se brindaron los resultados de los estudios genéticos.

Las bases de datos se encuentran bajo resguardo del Dr. Miguel Cruz López, responsable de la Unidad de Investigación Médica en Bioquímica, quien es el único que cuenta con las claves de identificación.

Condiciones en las que se solicitó el consentimiento: A todos los sujetos se les solicitó consentimiento informado por escrito, en cumplimiento con el Reglamento de la Ley General de Salud en Materia de Investigación para la Salud, Título Segundo “De los Aspectos Éticos de la Investigación en Seres Humanos”, Capítulo I, Artículo 14. Dicho consentimiento informado incluyó la autorización para la extracción de sangre, análisis genético, almacenamiento de muestras y empleo de la información para estudios futuros. El consentimiento informado fue solicitado por los investigadores de la Unidad de Investigación Médica en Bioquímica. Para los sujetos con DT2, el consentimiento se solicitó durante la cita de la entrevista; mientras que para el grupo donadores fue al momento de ingresar al Banco Central de Sangre.

No se contempló el envío de las muestras de DNA fuera del país en la carta de consentimiento informado, debido a la novedad del tema y que el avance del conocimiento impedía vislumbrar el mejor método para la detección de CNVs, en el momento que los sujetos firmaron dicha carta.

Desde el año 2009, las muestras de DNA almacenadas de los sujetos participantes fueron agotadas en su totalidad, con motivo de la genotipificación. No se recibieron solicitudes para la eliminación de muestras, ni se han documentado solicitudes de eliminación de los datos por parte de los sujetos participantes.

Forma de selección de los participantes: Los sujetos con DT2 fueron invitados a participar por el personal de la Unidad de Investigación Médica en Bioquímica en forma consecutiva, en las salas de espera de la Unidad de Medicina Familiar No. 4 de la Delegación Sur del Distrito Federal del Instituto Mexicano del Seguro Social e invitados a una cita en la Unidad de Investigación Médica en Bioquímica; mientras que los donadores fueron invitados a participar en forma consecutiva en el momento de ingresar al Banco Central de Sangre de Centro Médico Nacional Siglo XXI.

Esta investigación genética se realizó como parte de un proyecto de investigación, tal como lo establece la Ley General de Salud, en su Título Quinto, Capítulo Único, Artículo 100, inciso V.

Los resultados del análisis genético han sido manejados tomando en consideración lo establecido en el Acta de No Discriminación por Información Genética de 2008 (88).

12. RECURSOS HUMANOS Y MATERIALES

La Unidad de Investigación Médica en Bioquímica contó con los recursos humanos necesarios para la realización del proyecto

a) Recursos humanos

Se contó con la participación del alumno y tutor, así como de la colaboración de un doctor en ciencias. Las actividades que realizaron cada uno de los participantes del protocolo fueron:

- El alumno llevó a cabo el pre-procesamiento de información, el análisis de control de calidad, la detección de CNVs, el análisis de CNVs y de la visualización de regiones.
 - Cabe mencionar que la información fenotípica de los sujetos del presente estudio, así como la extracción y almacenamiento de las muestras de DNA, fue realizada durante los años 2008 y 2009 por personal de la Unidad de investigación Médica en Bioquímica, por lo que el alumno de maestría no participó en dichos procedimientos con motivo del presente protocolo.
- El Dr. Adán Valladares Salgado, responsable del banco de DNA de la Unidad de investigación Médica en Bioquímica, colaboró con la edición de datos del microarreglo y análisis de control de calidad.
- El Dr. Miguel Cruz López, jefe de la Unidad de Investigación Médica en Bioquímica vigiló el desarrollo y los avances del protocolo.

b) Recursos físicos

Se contó de un área física con computadoras para el análisis de los datos con la paquetería básica, así como los datos en código binario.

c) Recursos financieros

El protocolo contó con apoyo del Fondo Sectorial de Investigación en Salud y Seguridad Social del CONACYT (clave del proyecto: 2008-785-073; convocatoria:S0008-2008-1).

d) Experiencia del grupo de investigación.

El grupo tiene experiencia reconocida en el área de la genética de la DT2, posee colaboraciones internacionales, y publicaciones en revistas de alto impacto tales como Diabetologia, Diabetes Care, Diabetes, Human Genetics, Clinical Genetics, American Human Biology, Biochemical and Biophysical Research Communications, BMC Genetics, Journal of Endocrinological Investigation, Archives of Medical Research, European Journal of Pharmacology, Human Immunology, Molecular and Cellular Biochemistry, Nature Genetics, entre otras.

13. CRONOGRAMA DE ACTIVIDADES

El proyecto se desarrolló siguiendo el siguiente cronograma de actividades, que estuvo formado por un total de 24 meses (2 años), divididos en 4 semestres [Figura 2].

El protocolo fue presentado en cinco ocasiones en el Seminario de Investigación del Programa de Maestría (en marzo, junio y septiembre de 2012, y en mayo y septiembre del 2013) y en 3 ocasiones en los Coloquios de Investigación (agosto 2012 y febrero y julio de 2013); mostrando avances del mismo a partir del año 2013.

Se incrementó el tiempo del análisis de control de calidad debido a que sus fases coincidieron con el proceso de detección de CNVs, a que se consideró la inclusión de la detección de CNVs con el programa iPattern así como la generación de la base de datos común.



Figura 2: Cronograma de actividades.

14. RESULTADOS

Un total de 880 muestras (686 DT2 y 194 donadores) y 4,023 CNVs finalizaron las fases de pre-procesamiento, análisis de control de calidad y detección de CNVs [Anexo 6].

14.1 Análisis de CNVs entre grupos de comparación

Características de los CNVs entre los grupos de comparación

Del total de 4,023 CNVs que finalizaron el control de calidad, 3,182 fueron génicos. El grupo de DT presentó 3,103 CNVs totales y 2,458 CNVs génicos; mientras que el grupo donador tuvo 920 y 724 CNVs respectivamente [Tabla 7, Figura 3].

Tabla 7. Características de los CNVs en los grupos de estudio

Característica de los CNVs	Grupo DT2 (n=686)	Grupo donador (n=194)	Total (n=880)	p
Total de CNVs				
Número de CNVs	3,103	920	4,023	
Duplicaciones	1502 (48.4)	449 (48.8)	1951 (48.5)	
Deleciones	1601 (51.6)	471 (51.2)	2072 (51.5)	0.831
Tamaño de CNVs en kb (mediana)	169 (124-264)	165 (124-261)	166 (124-263)	0.856
Duplicaciones	223 (189-393)	223 (198-385)	223 (192-389)	0.300
Deleciones	129 (111-156)	127 (111-151)	128 (111-155)	0.210
CNVs génicos				
Número de CNVs	2,458	724	3,182	
Duplicaciones	1272 (51.7)	374 (51.7)	1646 (51.7)	
Deleciones	1186 (48.3)	350 (48.3)	1536 (48.3)	0.965
Tamaño de CNVs en kb (mediana)	160 (125-255)	164 (125-255)	161 (125-255)	0.671
Duplicaciones	223 (180-291)	223 (198-329)	223 (189-298)	0.037*
Deleciones	127 (111-151)	125 (111-151)	127 (111-151)	0.623

Los datos se presentan como número (%) ó mediana (percentil 25, percentil 75). Valor de p mediante prueba X² o prueba exacta de Fisher o U de Mann-Whitney *p<0.05

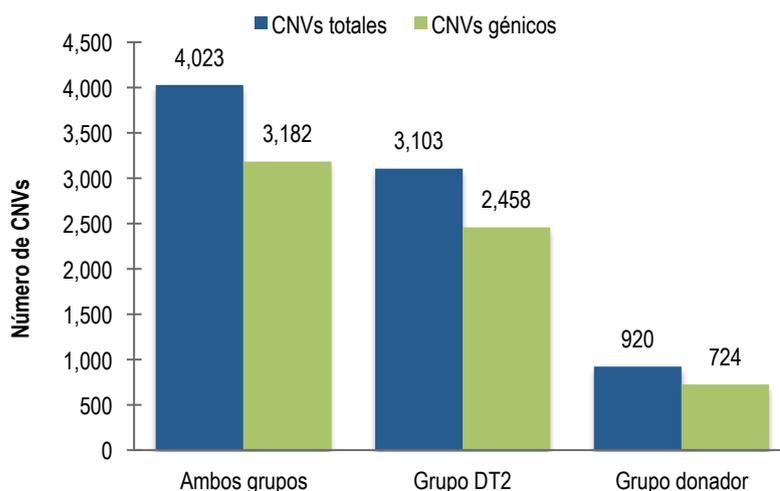


Figura 3. Conteo de CNVs totales y génicos en los grupos de estudio. DT2: Diabetes tipo 2

El porcentaje de CNVs génicos fue de 79.1%. El 84.4% de las duplicaciones fueron génicas, mientras que el 74.1% de las deleciones tuvieron esta característica. La distribución de los porcentajes de CNVs génicos fueron similares entre los grupos de comparación (Figura 4).

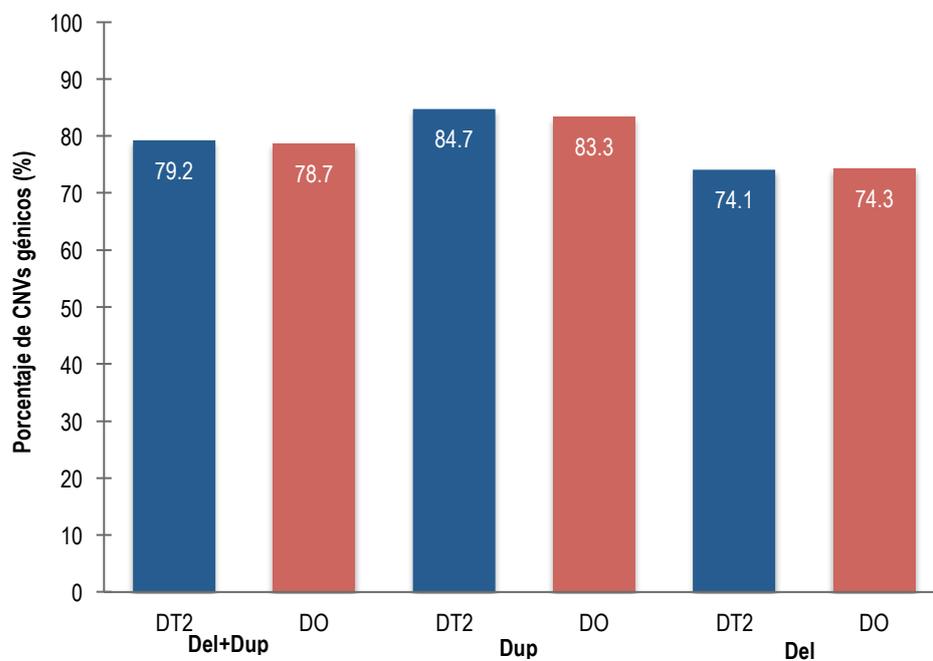


Figura 4. Porcentaje de CNVs génicos en duplicaciones y deleciones, en los grupos de estudio. Dup: Duplicaciones, Del: Deleciones, DT2: grupo diabetes tipo 2, DO: grupo donador.

El 48.5% de los CNVs totales fueron duplicaciones y el 51.5% deleciones; mientras que para los CNVs génicos el 48.3% correspondieron a deleciones y 51.7% a duplicaciones.

Se encontró una distribución similar de duplicaciones y deleciones entre los grupos de comparación (duplicaciones de CNVs totales: 48.4% en el grupo DT2 y 48.8% en el grupo donador, $p=0.831$; duplicaciones en CNVs génicos 51.7% en ambos grupos, $p=0.965$) [Figura 5].

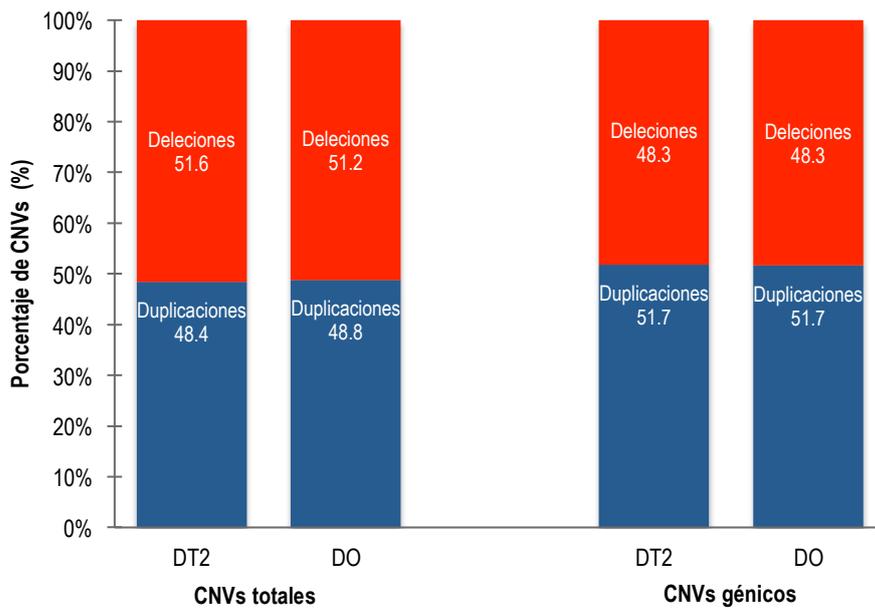


Figura 5. Porcentaje de duplicaciones y deleciones en los CNVs totales y génicos por los grupos de estudio. DT2: grupo diabetes tipo 2, DO: grupo donador.

Con respecto al tamaño, las medianas de los CNVs totales y génicos fueron de 166kb y 161kb respectivamente. Las duplicaciones presentaron un tamaño mayor que las deleciones (223kb vs 129kb). No se observaron diferencias de las medianas de tamaño de CNVs entre los grupos de comparación, a excepción de las duplicaciones génicas las cuales fueron ligeramente mayores en el grupo donador, en comparación con el grupo DT2 [Figura 6].

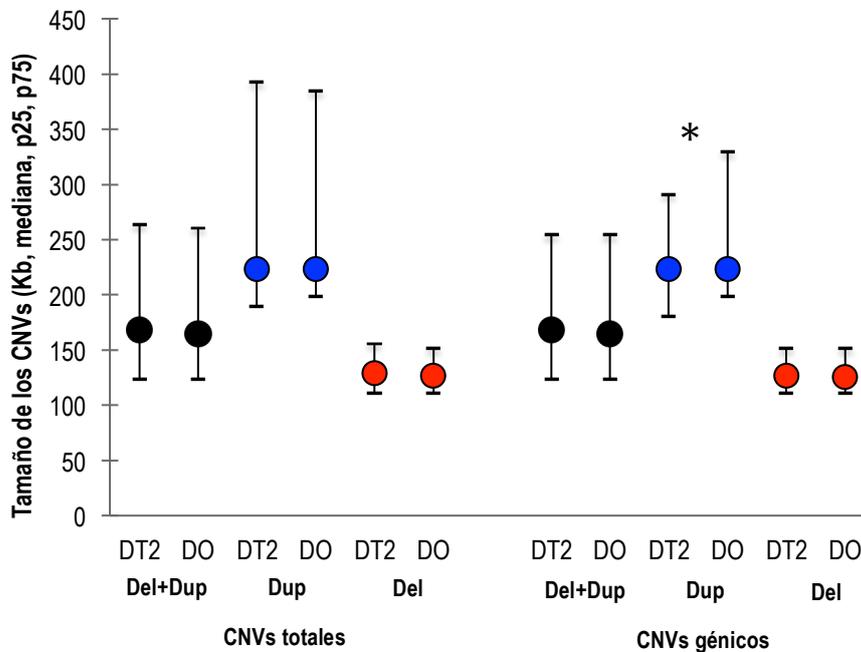


Figura 6. Mediana del tamaño de los CNVs totales y génicos por los grupos de estudio (mediana, percentil 25, percentil 75). En pares se presenta la categoría de CNVs; en negro los CNVs totales, en azul las duplicaciones, en rojo las deleciones. DT2: grupo diabetes tipo 2, DO: grupo donador, Del: deleciones, Dup: Duplicaciones. *Valor de $p < 0.05$ entre los grupos DT2 y DO.

Frecuencia de CNVs en los grupos de comparación

La frecuencia de las duplicaciones fueron de 89.4% en el grupo de DT2 y de 90.2% en el grupo donador, mientras que las frecuencias de deleciones fueron de 91.5% y 96.4% respectivamente.

Con respecto a los CNVs génicos, la frecuencia fue de 98.4% y 99.5% en los grupos de DT2 y donador. Para las duplicaciones génicas, las frecuencias fueron de 85.3% y 87.6% respectivamente y para las deleciones de 84.3% y 88.1%.

Al analizar la frecuencia en los grupos de comparación se observó que el grupo de donadores presentó una mayor frecuencia de deleciones que el grupo de DT2 (96.4% vs 91.5%, $p=0.020$) [Tabla 8, Figura 7]

Tabla 8. Número de muestras con CNVs en los grupos de estudio

Número de muestras con CNVs	Grupo DT2 (n=686)	Grupo donador (n=194)	p
Todos los CNVs			
Duplicaciones+Deleciones	686 (100)	194 (100)	1.000
Duplicaciones	613 (89.4)	175 (90.2)	0.733
Deleciones	628 (91.5)	187 (96.4)	0.020*
CNVs génicos			
Duplicaciones+Deleciones	675 (98.4)	193 (99.5)	0.481
Duplicaciones	585 (85.3)	170 (87.6)	0.407
Deleciones	578 (84.3)	171 (88.1)	0.179

Los datos se presentan como número (%). Valor de p entre grupo DT2 y grupo donador mediante prueba X2 o exacta de Fisher * $p<0.05$

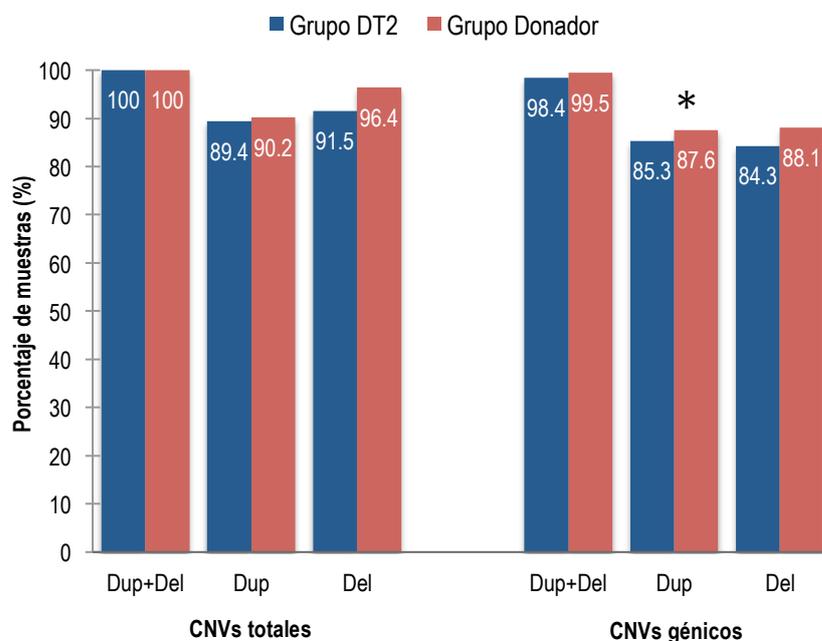


Figura 7. Frecuencia de CNVs en los grupos de comparación. Se muestra las duplicaciones y deleciones en los CNVs totales y génicos en el grupo de DT2 (barra azul) y donadores (barra roja), Del: deleciones, Dup: Duplicaciones. *Valor de $p<0.05$ entre los grupos DT2 y DO.

Diferencia en la frecuencia de CNVs entre los grupos de comparación

Con respecto a la diferencia en la frecuencia de CNVs en los grupos de comparación, se observó una diferencia en las deleciones (grupo DT2 -4.85%, $p=0.020$), la cual permaneció significativa al ajustarse por las variables edad, sexo, índice de masa corporal y ancestría (-5.16%, $p=0.003$) [Tabla 9, Figura 8].

Tabla 9. Diferencia de proporciones de muestras con CNVs entre los grupos de estudio

Característica	DP (IC95%)	p	DPa (IC95%)	p
Todos los CNVs				
Duplicaciones+Deleciones	0 (0 - 0)	1.000	0 (0 - 0)	1.000
Duplicaciones	-0.85 (-5.62 - +3.90)	0.733	+0.16 (-5.06 - +5.40)	0.953
Deleciones	-4.85 (-8.20 - -1.50)	0.020*	-5.16 (-8.54 - -1.80)	0.003*
CNVs génicos				
Duplicaciones+Deleciones	-1.09 (-2.47 - +0.30)	0.481	-0.86 (-2.46 - +0.70)	0.289
Duplicaciones	-2.35 (-7.69 - +3.00)	0.407	+0.02 (-5.97 - +6.00)	0.995
Deleciones	-3.89 (-9.19 - +1.40)	0.179	-4.97 (-10.3 - +0.30)	0.064

DP: diferencia de proporciones, DPa: diferencia de proporciones ajustada por sexo, edad, índice de masa corporal y ancestría (MDS vector 1). *Valor de $p < 0.05$

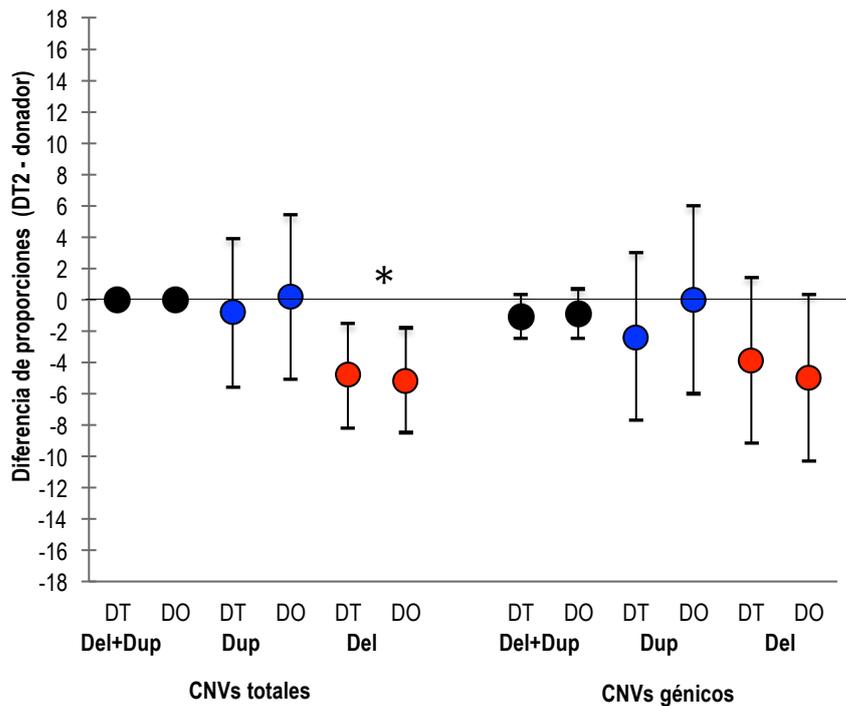


Figura 8. Diferencia en la frecuencia CNVs totales y génicos (A) por tipo de CNV (duplicación, deleción o ambos) en los grupos de comparación. DT2: grupo diabetes tipo 2, DO: grupo donador, Del: deleciones, Dup: Duplicaciones. *Valor de $p < 0.05$ entre los grupos DT2 y DO.

14.2 Visualización de regiones.

Se realizó la visualización de los diferentes CNVs. Con longitudes de 100Kb o mayores, los CNVs presentaron una mejor aspecto (Figura 9).

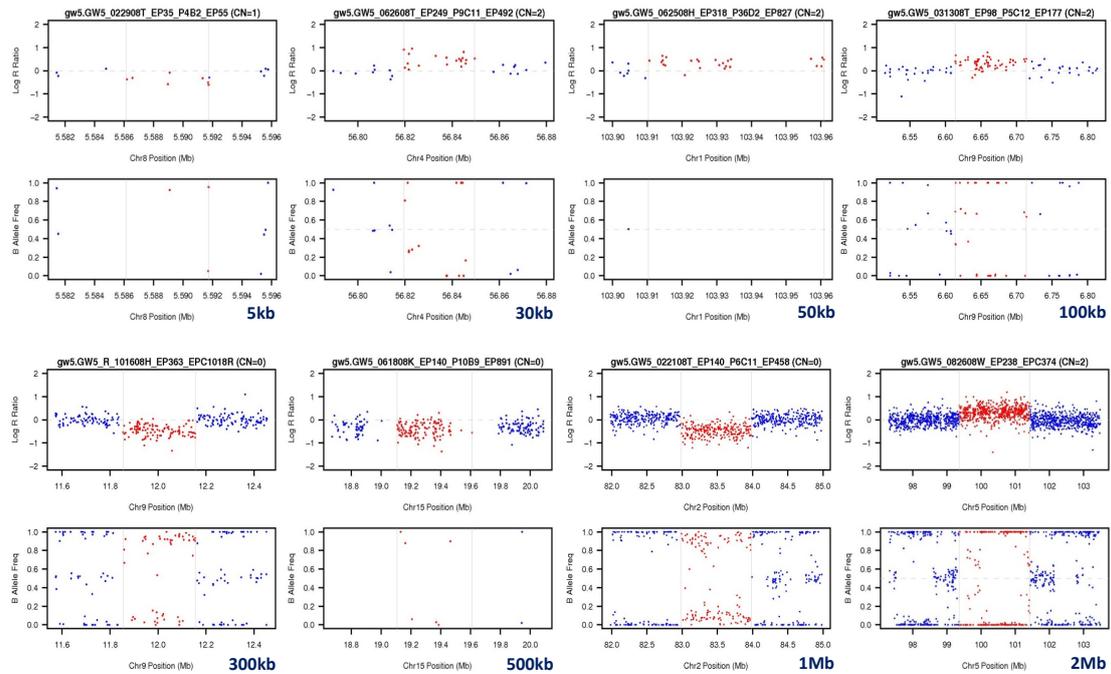


Figura 9. Comparativo del tamaño de los CNVs de acuerdo a su longitud: 5kb, 30kb, 50kb, 100kb, 300kb, 500kb, 1Mb y 2Mb.

Al realizar el análisis de segmentos génicos, se encontraron 3 regiones génicas con diferencia en las proporciones de CNVs entre los grupos de estudio, una región duplicada en el gen *NPIP* y dos regiones deletadas en los genes *ANKRD20A2* y *DUSP22* [Tabla 10, Figuras 10 y 11].

Tabla 10. Regiones génicas con diferencia de proporciones cruda y ajustada significativas, en los grupos de estudio

Región génica	Coordenadas	Tipo	Grupo DT2 (n=686)	Grupo donador (n=194)	DP (IC95%)	p	DPa (IC95%)	p
<i>ANKRD20A2</i>	chr9:43080420-43123540	Delección	31 (4.5)	2 (1)	3.5 (1.4 - 5.6)	0.024*	3.4 (1.2 - 5.6)	0.0026*
<i>NPIP</i>	chr16:1493880-14953432	Duplicación	28 (4.1)	1 (0.5)	3.6 (1.8 - 5.4)	0.014*	3.4 (1.6 - 5.3)	0.0003*
<i>DUSP22</i>	chr6:237100-296355	Delección	28 (4.1)	15 (7.7)	-3.7 (-7.7 - 0.4)	0.037*	-5 (-9.8 - -0.2)	0.0403*

DP: diferencia de proporciones, DPa: diferencia de proporciones ajustada por sexo, edad, índice de masa corporal y ancestría (MDS vector 1). *Valor de p <0.05

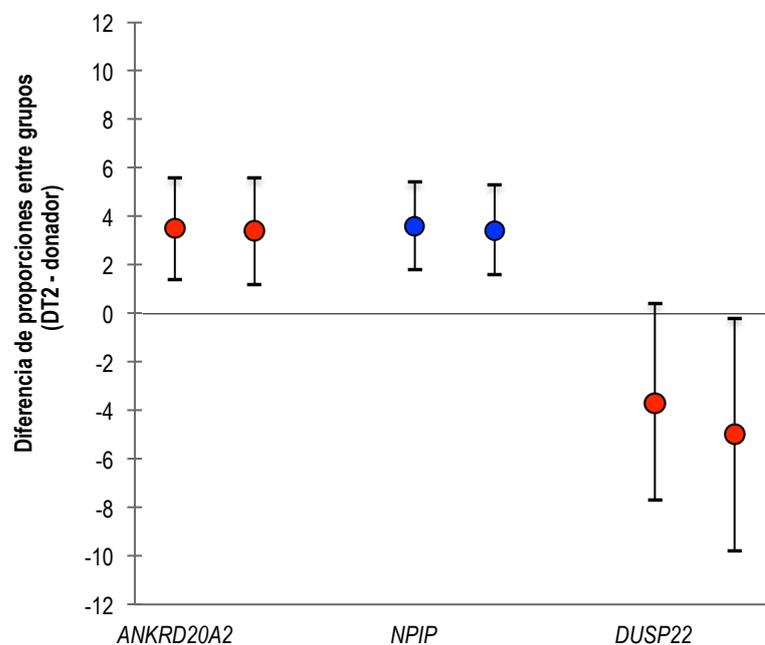


Figura 10. Diferencia de proporciones de CNVs en regiones génicas, en los grupos de estudio. Se muestra como un punto la estimación puntual de la diferencia de proporciones con su intervalo de confianza al 95% a la derecha cruda y a la izquierda ajustada. Los círculos azules corresponden a duplicaciones, en tanto los círculos rojos corresponden a delecciones.

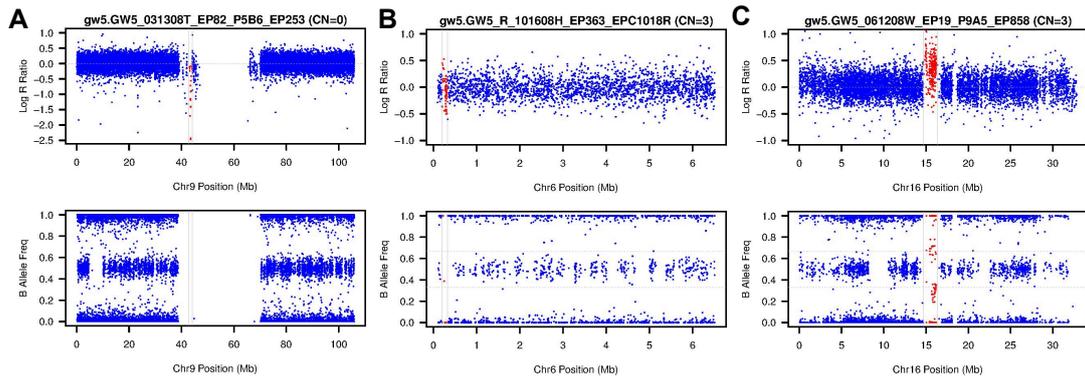


Figura 11. Regiones con diferencias de proporciones cruda y ajustada significativas entre los grupos de estudio. A: Delección de 1.2Mb en la región ANKRD20A2 en un caso de DT2; B: Delección de 123.9kb en la región génica DUSP22 en un caso de DT2; C: Duplicación de 1.6Mb en la región NPIP en un sujeto del grupo donador.

15. DISCUSIÓN

La Diabetes es un problema importante de Salud Pública en México. Se estima que el 9.2% de los Mexicanos mayores a 20 años padecen esta enfermedad (5); y que el costo de la Diabetes en 2011 fue de \$7.7 billones de dólares americanos (89).

La DT2 es una enfermedad compleja causada por factores ambientales y genéticos. La alta concordancia entre gemelos y las disparidades en las tasas de desarrollo del padecimiento entre hijos de padres con y sin diabetes muestran la influencia de factores genéticos en la ocurrencia de la enfermedad (9). La evidencia del fuerte componente hereditario de la DT2, han generado esfuerzos importantes para descifrar los factores genéticos que contribuyen en la génesis de la enfermedad

La disponibilidad de cantidades considerables de muestras de DNA de pacientes, así como la accesibilidad a tecnologías de genotipado de alta precisión (como los microarreglos), han consolidado a los GWAS como la metodología más exitosa para el descubrimiento de variantes genéticas asociadas a la DT2. Por ejemplo, los primeros GWAS permitieron identificar al gen *TCF7L2* como asociado a DT2, cuya función en la modulación de la función pancreática fue demostrada con posterioridad. Hasta el momento, se han descrito un total de 76 *loci* asociados a la DT2, la mayoría de ellos, detectados mediante esta metodología (18).

En forma adicional al análisis de SNPs, las plataformas de microarreglos permiten la caracterización de CNVs. Estas variantes de número de copias se han consolidado como una nueva línea de investigación debido al efecto débil de los SNPs en la contribución genética de enfermedades complejas (90). La detección de CNVs se ha convertido en una disciplina por sí misma y una parte importante de los estudios genéticos de susceptibilidad a diferentes enfermedades.

Se estima que aproximadamente el 71.48% del genoma contiene CNVs (53). Es posible que este tipo de variación genómica, puede ocasionar efectos fenotípicos debido a efectos cuantitativos o disruptivos; por ejemplo, las duplicaciones pueden crear un desbalance debido al exceso de productos de genes duplicados (54).

Los CNVs constituyen marcadores relativamente inexplorados para la DT2: se han realizado 8 estudios, todos ellos en poblaciones de origen europeo, asiático o africano y ninguno en población de origen latinoamericano.

La actual población mexicana es el resultado de un proceso de mestizaje que involucró a las poblaciones Nativas Americanas, Europeas y Africanas durante los siglos XVI y XVIII. Las contribuciones ancestrales que han sido estimados para el componente americano nativo oscilan entre 49% y 65%, para el europeo entre 30% y 45.8% y para el africano entre 1.8 y 6% (21-25).

Es bien conocido que la presencia de estratificación poblacional puede incrementar la tasa de falsos positivos en estudios de asociación con poblaciones mestizas. Una forma de controlar este efecto es el desarrollar un análisis multivariado como el realizado en este estudio.

El presente estudio comparativo describió la frecuencia de las CNVs en 686 muestras de sujetos con DT2 y 194 muestras de sujetos no diabéticos (donadores del Banco Central de Sangre), residentes de la Ciudad de México, genotipificados con el *microarreglo Affymetrix Genome-Wide Human SNP array 5.0*.

El *microarreglo Affymetrix Genome-Wide Human SNP array 5.0* contiene un total de 1,340,037 sondas: 500,568 para SNPs, 340,742 para CNVs y 6,113 sondas de control, que permiten la detección de SNPs y CNVs a lo largo de los 22 pares de cromosomas y los cromosomas sexuales (58).

Nuestro estudio reportó cuatro hallazgos interesantes: 1) una mayor proporción de deleciones en comparación con las duplicaciones; 2) un mayor tamaño de las duplicaciones en comparación con las deleciones; 3) una menor frecuencia de deleciones en el grupo de DT2 en comparación con el grupo de donadores; y finalmente 4) tres regiones génicas con diferencia de en la frecuencia entre el grupo DT2 y los donadores.

El reporte de Blackburn (80), en el cual utilizaron diferentes tipos de *microarreglos Beadchip* y el programa PennCNV, también halló una mayor proporción de deleciones en comparación con las duplicaciones, obteniendo una razón deleción/duplicación de entre 1.56 y 2.75. En nuestro estudio dicha razón fue moderadamente inferior (1.06).

El tamaño de las duplicaciones también fue mayor que el de las deleciones en el reporte de Blackburn (con medianas de entre 2.6Kb y 20.4Kb para las deleciones y de entre 3.5kb a 31.2kb en las duplicaciones). Sin embargo dicho tamaño fue menor en comparación con nuestro estudio (223 kb en duplicaciones y 128 kb en deleciones), ya que reportamos los CNVs con tamaño mayor a 100kb.

Si bien, porcentaje de duplicaciones y deleciones del total de CNVs no presentó diferencias entre los grupos de comparación, se observó una menor frecuencia de muestras con deleciones en el grupo de Diabetes tipo 2 en comparación con el grupo de donadores (91.5% vs 96.4%, $p=0.020$, DPa -5.1%, $p=0.003$).

El presente estudio reveló también tres regiones génicas en las cuales se encontraron diferencias en la frecuencia de CNVs. El grupo de DT2 presentó una mayor frecuencia de deleciones en el gen *ANKRD20A2* y duplicaciones en el gen *NPIP*; mientras que el grupo donador presentó una frecuencia mayor de deleciones en el gen *DUSP22*.

El gen *NPIP* (OMIM#606406) se localiza en el cromosoma 16p13.11 y genera una proteína de interacción poro nuclear. Las duplicaciones en este gen se han relacionado con la divergencia entre humanos y simios africanos (91). Un incremento en el transcrito de este gen, se ha reportado en el perfil de expresión de la mácula humana (92).

Con respecto al gen *ANKRD20A2*, se han reportado CNVs de novo (duplicaciones) en familias con Esquizofrenia. Dicho gen se localiza en el cromosoma 9p12 (93).

Finalmente se han reportado CNVs raros (duplicaciones) en el gen *DUSP22* en sujetos con disección de la arteria cervical (94), además que se ha descrito la hipometilación del promotor de dicho gen en sangre periférica de bomberos en servicio (95).

En nuestro conocimiento, este es el primer estudio realizado en población mexicana que describe CNVs en un grupo de sujetos con DT2.

El diseño comparativo que limita la inferencia de los hallazgos, el porcentaje elevado de eliminación de muestras durante el control de calidad, y el tamaño de muestra limitado, son las principales debilidades del estudio.

Con respecto al porcentaje de muestras eliminadas, éste ha superior a reportes realizados con el mismo microarreglo. En el estudio de Grassi (77), el porcentaje de muestras eliminadas fue de 12.8%, dado que sus criterios de control de calidad fueron laxos, pues únicamente eliminaron muestras con incompatibilidades de sexo y con tasa de genotipificación <90%.

Sin embargo, la selección estricta de muestras y de CNVs, el uso de medidas de control de calidad rigurosas, la construcción de una base de datos común, y la novedad del tema son fortalezas dignas de mencionar en el presente estudio.

Nuestros resultados sugieren que los sujetos con DT2 poseen una menor frecuencia de deleciones en comparación con un grupo de sujetos no diabéticos y que las regiones génicas *DUSP22*, *NPIP* y *ANKRD20A2* presentan frecuencias diferentes en los grupos de comparación.

16. CONCLUSIONES

La realización del Estudio de Variantes del Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2, brinda las siguientes conclusiones:

- La frecuencia de las Variantes en el Número de Copias $\geq 100\text{kb}$, en los sujetos con Diabetes tipo 2 y los sujetos no diabéticos fue mayor al 84%.
 - Las deleciones fueron más frecuentes que las duplicaciones.
 - Las duplicaciones presentaron un mayor tamaño en comparación de con las deleciones.
- Se observó una menor frecuencia de deleciones en el grupo de Diabetes tipo 2 en comparación con el grupo de donadores.
- Las regiones génicas *DUSP22*, *NPIP* y *ANKRD20A2* presentaron frecuencias estadísticamente diferentes en los grupos de comparación.

17. ANEXOS

Anexo 1. Procedimiento de selección de los sujetos y del procesamiento de las muestras de los participantes del protocolo “Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes”.

Los participantes del estudio fueron sujetos habitantes de la ciudad de México.

El grupo de DT2 fue constituido por sujetos con diagnóstico conocido de DT2, edad al diagnóstico menor a 55 años y diagnosticados de acuerdo a los criterios de la Asociación Americana de Diabetes. Fueron reclutados en forma consecutiva de la Unidad de Medicina Familiar No. 4 de la Delegación Sur del Distrito Federal del Instituto Mexicano del Seguro Social entre 2008 y 2009.

El grupo de sujetos no diabéticos fue formado por donadores del banco central de sangre sin historia familiar de diabetes, edad mayor a 40 años y concentraciones de glucosa plasmática en ayunas <100mg/dL. La recolección de las muestras se realizó también entre los años 2008 y 2009.

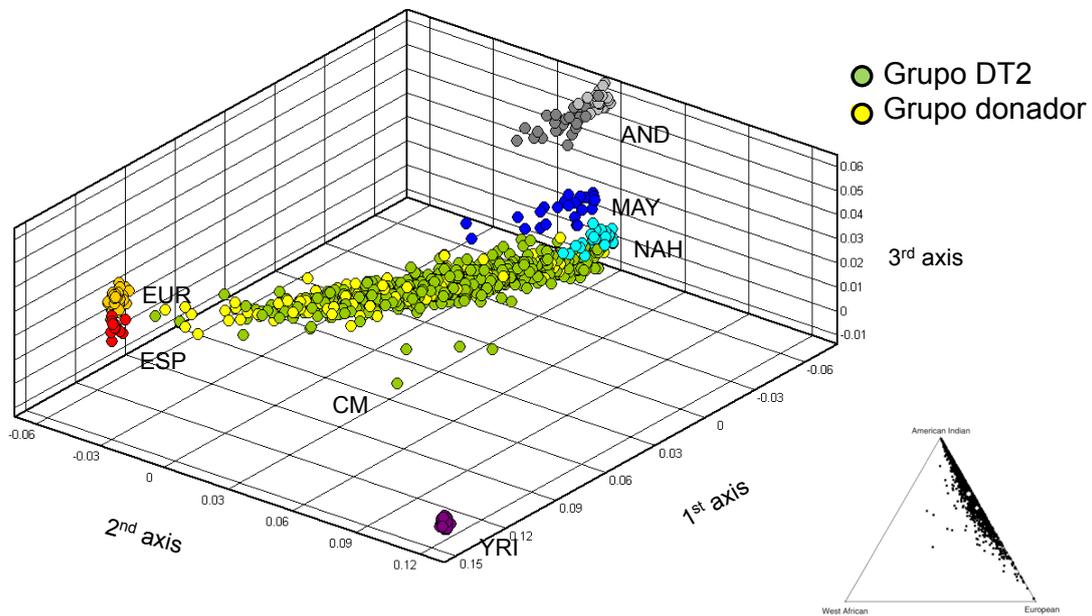
Los sujetos fueron invitados directamente a participar por los investigadores de la Unidad de Investigación en Bioquímica, en las salas de espera de la consulta externa de la Unidad de Medicina Familiar No. 4 (grupo de DT2) y requeridos para una cita en dicha Unidad; mientras que los donadores fueron invitados al momento de ingresar al Banco Central de Sangre de Centro Médico Nacional Siglo XXI. Todos firmaron la hoja de consentimiento informado. Se obtuvo de sangre venosa de cada participante.

La extracción de DNA de las muestras sanguíneas se realizó a partir de leucocitos por el método basado en la separación en columnas (QIAamp DNA Blood Midi/ Kit, Qiagen, Alemania). Se analizó la pureza y concentración mediante espectrofotometría a 260/280 nm. Para evaluar la integridad del DNA se realizó electroforesis en geles de agarosa al 0.8% teñidas con bromuro de etidio.

Las alícuotas de DNA fueron almacenadas en ultracongeladores a -70° C hasta el momento del envío; y posteriormente fueron transportadas por vía aérea en recipientes utilizando hielo seco como refrigerante, en un lapso de 24 horas hasta su destino final.

La genotipificación se realizó en las instalaciones de The Centre for Applied Genomics (Toronto, ON, Canada) usando el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*, en el año 2009.

Anexo 2. Estimación de la Ancestría en 967 sujetos con DT2 y 343 donadores, participantes del estudio y poblaciones ancestrales.



Gráfica Multidimensional de los sujetos del estudio (puntos verdes: 967 sujetos con DT2 y puntos amarillos: 343 donadores) y 6 poblaciones ancestrales (NAH (Nahua), MAY (Maya), AND (Andino), ESP (Españoles), EUR (Europeos CEPH), YRI (Yoruba), CM (Ciudad de México)) obtenidos con el programa ADMIXMAP utilizando un panel de 3,500 marcadores informativos de ancestría (24) basados en información de estudios de genoma amplio en poblaciones Europea, Africana del este y Amerindia.

Anexo 3. Carta de Aprobación Comisión Nacional de Investigación Científica.



INSTITUTO MEXICANO DEL SEGURO SOCIAL

DIRECCIÓN DE PRESTACIONES MÉDICAS
UNIDAD DE EDUCACIÓN, INVESTIGACIÓN Y POLÍTICAS DE SALUD
COORDINACIÓN DE INVESTIGACIÓN EN SALUD
"2008, Año de la Educación Física y el Deporte"

COMISIÓN NACIONAL DE INVESTIGACIÓN CIENTÍFICA

Ref. 09-B5-61-2800/200800

Diciembre 10, 2008

2110

DR. MIGUEL CRUZ LÓPEZ
Unidad de Investigación Médica en Bioquímica
UMAE, Hospital de Especialidades
Centro Médico Nacional SXXI
México, D. F.

Informo a usted que el protocolo titulado: "**Estudio de variantes de Número de Copia (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes**", fue sometido a la consideración de esta Comisión Nacional de Investigación Científica.

Los procedimientos propuestos en el protocolo cumplen con los requerimientos de las normas éticas vigentes y la carta de consentimiento informado es suficientemente explícita, por lo cual tengo el agrado de hacerle saber que con base en las opiniones de los vocales de esta Comisión, se ha emitido dictamen de **AUTORIZADO**, con número de registro: **2008-785-073**.

Atentamente

DOCTOR CÉSAR A. CRUZ SANTIAGO
Presidente
Comisión Nacional de Investigación Científica

Con copia:

- Dr. Mario Madrazo Navarro, Director de la UMAE, Hospital de Especialidades, C.M.N. Siglo XXI.
- Dr. Adolfo Chávez Negrete, Director de Educación e Investigación en Salud, UMAE Hospital de Especialidades, C.M.N. Siglo XXI.

DTP/iah
2694-2008-065

IMSS

SEGURIDAD Y SOLIDARIDAD SOCIAL

4º piso Bloque "B" de la Unidad de Congresos Av. Cuauhtémoc 330 Col. Doctores México 06720

Anexo 4. Licencia Sanitaria 06 T 09 006 0074

05330080010035		ESTADOS UNIDOS MEXICANOS SECRETARIA DE SALUD COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS SANITARIOS COMISIÓN DE AUTORIZACIÓN SANITARIA	
NOMBRE DEL PROPIETARIO O RAZÓN SOCIAL: DENOMINACIÓN DEL ESTABLECIMIENTO:	INSTITUTO MEXICANO DEL SEGURO SOCIAL U.M.A.E. HOSPITALES DE ESPECIALIDADES DEL CENTRO MEDICO NACIONAL	DOMICILIO: COLONIA Y/O LOCALIDAD: DELEGACIÓN POLÍTICA O MUNICIPIO: C.P.:	AV. CUAUHTEMOC No. 330 DOCTORES, CD. DE MÉXICO CUAUHTEMOC 06720
ENTIDAD FEDERATIVA:	DISTRITO FEDERAL		
LICENCIA SANITARIA No.			
06 T 09 006 0074			
AUTORIZADO PARA:			
DISPOSICIÓN DE ÓRGANOS, TEJIDOS Y CÉLULAS CON FINES TERAPEUTICOS, EN SU MODALIDAD DE PROCURACIÓN Y TRASPLANTE DE CórNEA, RIÑÓN, HIGADO, CORAZÓN, PANCREAS, CÉLULAS PROGENITORAS HEMATOPOYETICAS E INTESTINO.			
FECHA DE EXPEDICIÓN 10/02/2006	LA PRESENTE AUTORIZACIÓN ES VALIDA SIEMPRE Y CUANDO NO SEAN MODIFICADAS LAS CONDICIONES EN LAS QUE FUE EXPEDIDA		
POR TIEMPO INDETERMINADO DE CONFORMIDAD AL ARTICULO 370 DE LA LEY GENERAL DE SALUD			DR. ALVARO HERRERA HUERTA SUBDIRECTOR EJECUTIVO DE AUTORIZACIONES EN SERVICIOS DE SALUD
AUTOSUJETO DECIMO PRIMERO TRACCION Y VIGESIMO PRIMERO DEL ACUERDO POR EL QUE SE DELEGAN LAS FACULTADES QUE SE SEÑALAN EN LOS ORGANOS ADMINISTRATIVOS QUE SE INDICAN DE LA COMISION FEDERAL PARA LA PROTECCION CONTRA RIESGOS SANITARIOS			
ESTA LICENCIA DEBERA ESTAR EXHIBIDA EN LUGAR VISIBLE DEL ESTABLECIMIENTO			

Anexo 5. Comisión de autorización sanitaria de entrada-salida 04930200377



SECRETARIA
DE SALUD

DEPENDENCIA: COMISION FEDERAL PARA LA
PROTECCION CONTRA RIESGOS SANITARIOS.

COMISION DE AUTORIZACION SANITARIA

N° DE ENTRADA-SALIDA: 04930200377

MONTERREY N° 33, COL. ROMA, C.P. 06700,
MEXICO, D.F.

México DF., a 14 de Junio del 2004

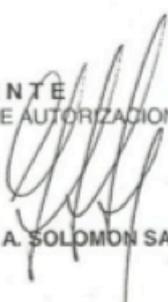
INSTITUTO MEXICANO DEL SEGURO SOCIAL
DR. MIGUEL CRUZ LOPEZ
AV. CUAUHTEMOC 330
COL. DOCTORES
MEXICO, DF
CP. 06720

Con fundamento en los artículos 4º, párrafo tercero, 14 y 16 de la Constitución Política de los Estados Unidos Mexicanos; 39 fracciones XIII, XXI y XXIV de la Ley Orgánica de la Administración Pública Federal; 3º Fracción XXII, XXVI y XXVIII, 4 fracción III, 17 bis fracción II, IV, VI, VIII, X y XIII, 17 bis 1, 17 bis 2, 313 fracción 1, 333, 340, 368, 371, 375 fracción VI de la Ley General de Salud, 2 inciso C fracción II, 34 y 35 del Reglamento Interior de la Secretaría de Salud, 14 del Reglamento de la Comisión Federal para la Protección Contra Riesgos Sanitarios, artículo Segundo, fracción X del Decreto de Creación del Centro Nacional de la Transfusión Sanguínea, publicado en el Diario Oficial de la Federación el 21 de enero de 1988, y en atención a su solicitud con número de entrada al rubro citado de fecha 14 de Junio, para el permiso de salida del territorio nacional de 100 muestras de DNA de pacientes con diabetes tipo 2, esta Comisión, le informa lo siguiente:

Las muestras antes citadas no requieren permiso de salida expedido por esta Comisión, ya que de acuerdo al DOF del 29 de marzo del 2002, la fracción arancelaria 3002.10.14 acota a paquete globular humano o concentrado de eritrocitos y no al producto a importar. Además que serán empleadas en la investigación titulada "Caracterización de pacientes con Diabetes tipo 2 y sujetos sanos, con marcadores genéticos por Admixture mapping", no representando riesgo sanitario.

Lo anterior sin menoscabo de las atribuciones que otras dependencias ejerzan en la materia

ATENTAMENTE
COMISIONADO DE AUTORIZACION SANITARIA


DR. GUILLERMO A. SOLOMON SANTIBAÑEZ
LGP/ZESA

Anexo 6. Fases de pre-procesamiento, análisis de control de calidad y detección de CNVs.

Al inicio de estas tres etapas, se contó con un total de 1310 muestras (967 sujetos con DT2 y 343 donadores) y se concluyó con 880 muestras (686 sujetos con DT2 y 194 donadores) y 4,023 CNVs [Figura Suplementaria 1]. El porcentaje de eliminación de muestras fue 38.2%

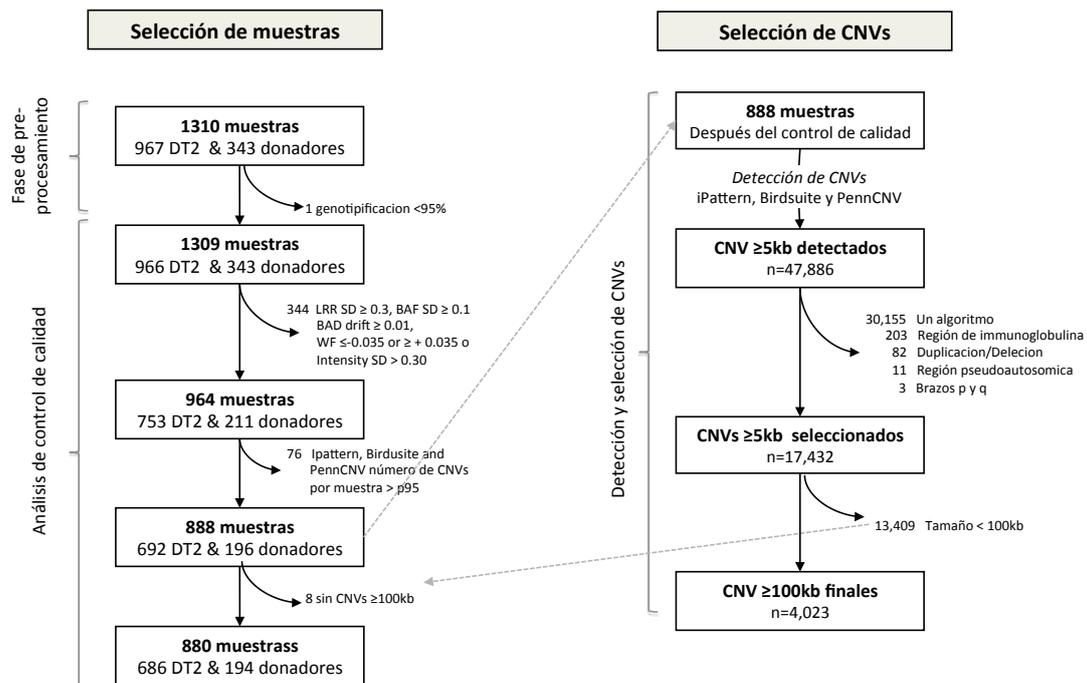


Figura Suplementaria 1: Flujograma del proceso de análisis de control de calidad para la selección de las muestras y CNVs.

6.1 Fase de pre-procesamiento

En esta etapa se obtuvo la información del LRR y BAF de las muestras y se eliminó una de ellas debido a una genotipificación menor al 95%.

6.2 Análisis de control de Calidad

En la primera etapa del análisis de control de calidad se eliminaron un total de 345 muestras (26.3%) [Figura Suplementaria 1 y Tabla suplementaria 1]:

- 238 por presentar una desviación estándar de la intensidad de señal >0.3
- 105 por tener una desviación estándar del Log R ratio ≥ 0.3
- 1 por variación de la frecuencia del alelo B > 0.01
- 1 por presentar factor de variación < -0.035 y $> +0.035$

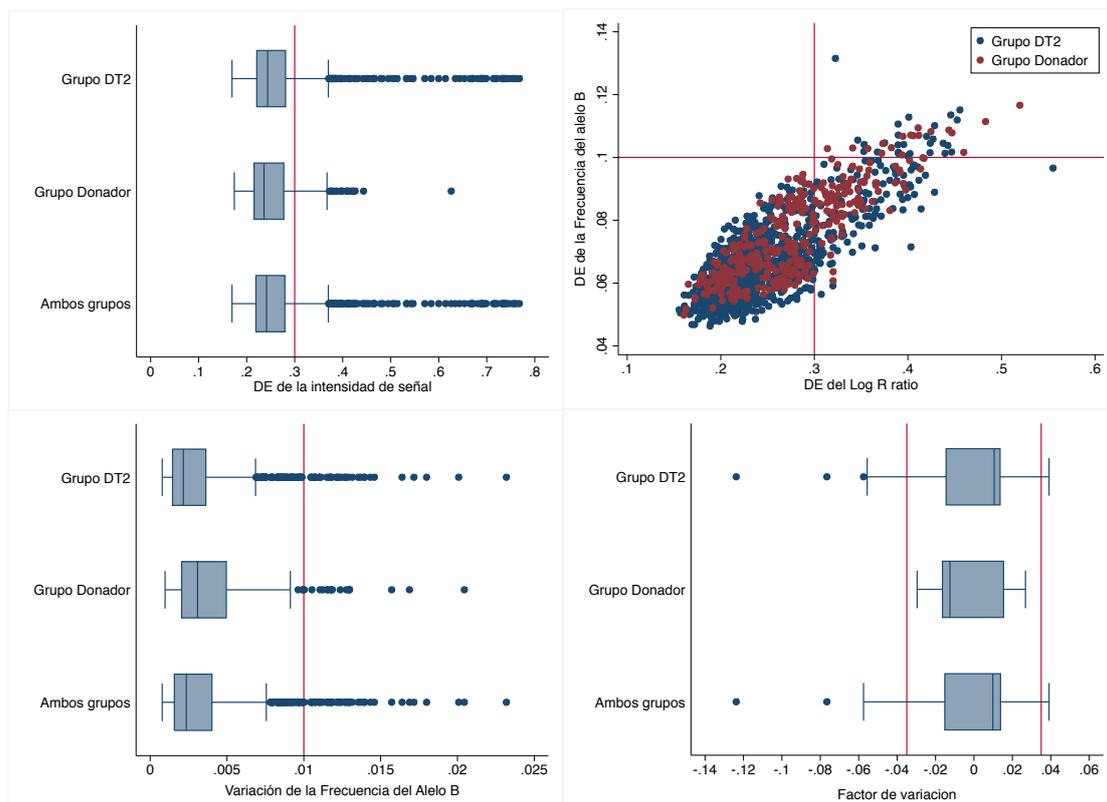


Figura Suplementaria 2: Pasos de eliminación de muestras en la primera etapa del análisis de control de calidad: a) Desviación estándar de la intensidad de señal, b) Desviación estándar de frecuencia del alelo B y desviación estándar de Log R ratio, c) Variación de la frecuencia del alelo B y d) factor de variación.

Durante la segunda etapa y superponiéndose a la fase de detección de CNVs (ver más adelante), se eliminaron 84 muestras adicionales

- 76 por presentar un número de CNVs obtenidos mediante cada uno de los tres algoritmos PennCNV, iPatern y Birdsuite por encima del percentil p95
- 8 por carecer de CNVs con tamaño de 100kb o más.

Tabla Suplementaria 1. Resumen del análisis de control de calidad (CC)

Paso de control de calidad de las muestras	Grupo DT2		Grupo donador		Total	
	# Muestras eliminadas	# Muestras después del CC	# Muestras eliminadas	# Muestras después del CC	# Muestras eliminadas	# Muestras después del CC
Criterios básicos de exclusión						
Número de muestras	0	967	0	343	0	1,310
Tasa de genotipificación <95%	1	966	0	343	1	1,309
Tasa de control de intensidad < 86%	0	966	0	343	0	1,309
Control de calidad a nivel de lotes						
Desviación estándar de la intensidad de señal >0.3	172	794	66	277	238	1,071
Control de calidad a nivel de muestras						
Desviación estándar del Log R ratio ≥ 0.3	39	755	66	211	105	966
Desviación estándar de la frecuencia del Alelo B ≥ 0.1	0	755	0	211	0	966
Variación de la frecuencia del alelo B > 0.01	1	754	0	211	1	965
Factor de variación < -0.035 y > +0.035	1	753	0	211	1	964
Control de calidad a nivel de detección de CNVs						
Muestras con número CNVs obtenidos >p95	61	692	15	196	76	888
Muestras sin CNVs con tamaño ≥ 100 kb	6	686	2	194	8	880
Resultado final del control de calidad						
Total de muestras	281	686	149	194	430	880

Las características de los sujetos de estudio después del análisis de control de calidad se muestran en la Tabla Suplementaria 2.

Tabla Suplementaria 2. Características de los sujetos participantes del "Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes" genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*, después del análisis de Control de Calidad

Característica	Grupo DT2	Grupo Donador	Total	p
n	686	194	880	
Sexo masculino	205(29.9%)	89(45.9%)	294(33.4%)	
femenino	481(70.1%)	105(54.1%)	586(66.6%)	<0.001
Edad, años	49 (44, 56)	50 (47, 54)	49 (45, 55)	0.086
Edad al diagnóstico, años	44 (39, 48)	ND	ND	ND
Tiempo de evolución, años	3 (1, 10)	ND	ND	ND
Índice de masa corporal (kg/m ²)	29.2 (26.2, 32.6)	27.4 (25.3, 29.6)	28.6 (26,32)	<0.001

Los datos se presentan como número(%) ó mediana (percentil 25, percentil 75). Valor de p entre los grupos mediante la prueba, χ^2 o U-Mann-Whitney. ND, no disponible

Al comparar las muestras retenidas con aquéllas eliminadas, se observó que las muestras eliminadas presentaron una mayor proporción de varones, así como mayor edad, edad al diagnóstico y tiempo de evolución, sin encontrar diferencias en el Índice de Masa Corporal [Tabla Suplementaria 3].

Tabla Suplementaria 3. Características de los sujetos participantes del "Estudio de variantes de Número de Copias (CNVs) en el Genoma de Pacientes con Diabetes Tipo 2 y su implicación en las complicaciones más frecuentes" genotipificados con el microarreglo *Affymetrix Genome-Wide Human SNP array 5.0*, después del análisis de Control de Calidad. Comparativo entre muestras retenidas y eliminadas por grupo de estudio.

Característica	Retenidas	Eliminadas	p
Grupo DT2			
n	686	281	
Sexo masculino	205 (29.9%)	98 (34.9%)	0.129
femenino	481 (70.1%)	183 (65.1%)	
Edad, años	49 (44,56)	53 (48,58)	<0.001
Edad al diagnóstico, años	44 (39,48)	47 (40,51)	<0.001
Tiempo de evolución, años	3 (1,10)	5 (2,12)	<0.001
Índice de masa corporal (kg/m ²)	29.2 (26.2,32.6)	29.1 (26.6,32.3)	0.999
Grupo Donador			
n	194	149	
Sexo masculino	89 (45.9%)	85 (57%)	0.040
femenino	105 (54.1%)	64 (43%)	
Edad, años	50 (47,54)	50 (47,54)	0.411
Edad al diagnóstico, años	ND	ND	ND
Tiempo de evolución, años	ND	ND	ND
Índice de masa corporal (kg/m ²)	27.4 (25.3,29.6)	27.1 (24.9,28.8)	0.304
Total			
n	880	430	
Sexo masculino	294 (33.4%)	183 (42.6%)	0.001
femenino	586 (66.6%)	247 (57.4%)	
Edad, años	49 (44,56)	53 (48,58)	<0.001
Edad al diagnóstico, años	44 (39,48)	47 (40,51)	<0.001
Tiempo de evolución, años	3 (1,10)	5 (2,12)	<0.001
Índice de masa corporal (kg/m ²)	28.6 (26,32)	28.2 (25.9,31.1)	0.1434

Los datos se presentan como número (%) ó mediana (percentil 25, percentil 75). Valor de p entre los grupos mediante la prueba, χ^2 o U-Mann-Whitney. ND, no disponible

El poder resultante para el nuevo tamaño de muestra fue de 0.9790.

Tomando en cuenta la región *LEPR* (60), utilizando la fórmula para el cálculo del poder para una diferencia de proporciones; y asumiendo una proporción de casos expuestos (P_1) 0.637, una proporción de controles expuestos (P_2) 0.477, un valor alfa de 0.05 ($z_\alpha = 1.96$) y un tamaño de muestra para casos (n_1) de 686 y para controles (n_2) de 194, tenemos:

$$\text{Poder} = \Phi\{\Delta / \sqrt{[(p_1q_1/n_1) + (p_2q_2/n_2)]}\} - z_\alpha\{\sqrt{p'q'}/\sqrt{[(p_1q_1/n_1) + (p_2q_2/n_2)]}\}$$

donde:

$$q_1 = 1 - p_1$$

$$q_2 = 1 - p_2$$

$$\Delta = |p_2 - p_1|$$

$$p' = (n_1p_1 + n_2p_2) / (n_1 + n_2)$$

$$q' = 1 - p'$$

Desarrollando, tenemos:

$$p' = [(686 \times 0.637) + (194 \times 0.477)] / (686 + 194) = (436.98 + 92.5380) / 133084 = 529.52 / 133084 = 0.602$$

$$q' = 1 - 0.602 = 0.398$$

$$\text{Poder} = \Phi \left\{ \frac{0.16}{\sqrt{[(0.637 \times 0.363 / 753) + (0.477 \times 0.523 / 211)]}} - 1.96 \left\{ \frac{\sqrt{0.601 \times 0.398} (1/686 + 1/194)}{\sqrt{[(0.637 \times 0.363 / 686) + (0.477 \times 0.523 / 194)]}} \right\} \right\}$$

$$\text{Poder} = \Phi \left\{ \frac{0.16}{\sqrt{(0.0003070 + 0.001182)}} - 1.96 \left\{ \frac{\sqrt{0.23960 \times (0.0014 + 0.0051)}}{\sqrt{(0.0003370 + 0.001285)}} \right\} \right\}$$

$$\text{Poder} = \Phi \left\{ \frac{0.16}{\sqrt{(0.0016)}} - 1.96 \left\{ \frac{\sqrt{0.2396 \times 0.00606}}{\sqrt{(0.0096)}} \right\} \right\}$$

$$\text{Poder} = \Phi \{ 0.16 / 0.0386 \} - 1.96 \{ [0.03980] / 0.0402 \} = 3.9715 \quad (-1.96 \times 0.9881)$$

$$\text{Poder} = \Phi(3.9715 - 1.9367) = 2.0348$$

$$\text{Poder} = \Phi(2.0348) = 0.9790$$

6.2 Detección de CNVs

En las 964 muestras restantes en la primera fase de eliminación (753 sujetos con DT2 y 211 donadores), se realizó la detección de CNVs mediante los tres algoritmos. El número total de CNVs fueron de 67,112, 55,311 y 50,647 para los algoritmos PennCNV, Birdsuite y iPattern respectivamente.

Al enfocarnos en los CNVs con tamaño de 5kb o más y generadas por 5 o más sondas, los números totales de CNVs fueron de 45,854 , 34,578 y 45,342 respectivamente para los algoritmos. La mediana de CNVs por sujeto fue de 41 para iPattern, 40 para PennCNV y 35 para Birdsuite. La dispersión en el número de CNVs por muestra fue mayor en PennCNV, seguido de iPattern y de Birdsuite (rango intercuartil 10, 9 y 7 respectivamente) [Tabla Suplementaria 4].

Tabla Suplementaria 4. Número de CNVs obtenidos a través de los algoritmos de detección, en 964 muestras (753 sujetos DT2 y 211 donadores)

Paso para la selección de los CNVs	Total de CNVs			CNVs ≥ 5 sondas ≥ 5 kb		
	Grupo DT2	Grupo donador	Total	Grupo DT2	Grupo donador	Total
Número total de CNVs, detectados por PennCNV	53,444	13,668	67,112	36,686	9,168	45,854
Número total de CNVs, detectados por Birdsuite	43,089	12,222	55,311	26,981	7,597	34,578
Número total de CNVs, detectados por iPattern	40,755	9,892	50,647	36,591	8,751	45,342
Mediana de número de CNVs por muestra, detectados por PennCNV	60 (52,68)	62 (54,70)	60 (52.5,69)	39 (35,46)	42 (35,47)	40 (35,46)
Mediana de número de CNVs por muestra, detectados por Birdsuite	55 (50,61)	57 (53,63)	56 (50,62)	34 (31,38)	36 (32,39)	35 (31,38)
Mediana de número de CNVs por muestra, detectados por iPattern	47 (41,53)	46 (41,51)	47 (41,52)	42 (36,46)	41 (36,45)	41 (36,46)

Los datos se presentan como números absolutos ó mediana (percentil 25, percentil 75).

Comportamiento de los algoritmos de detección y generación de la base de datos común

Dada la posibilidad de que las muestras con un alto número de CNVs mostraran detecciones falsas, se eliminaron las muestras que presentaron un número de CNVs por encima del percentil 95 para cada uno de los algoritmos (72, 45 y 53 para PennCNV, Birdsuite y iPattern respectivamente). De esta forma, 76 muestras fueron descartadas [Tabla Suplementaria 1].

Al analizar el comportamiento de los CNVs entre los diferentes algoritmos en las 888 muestras restantes, se encontró la mayor correlación entre los algoritmos iPattern y Birdsuite (valores de correlación de Spearman: iPattern y Birdsuite, $\rho=0.3936$; PennCNV y Birdsuite, $\rho=0.3161$; PennCNV y iPattern, $\rho=0.2029$) [Figura Suplementaria 3].

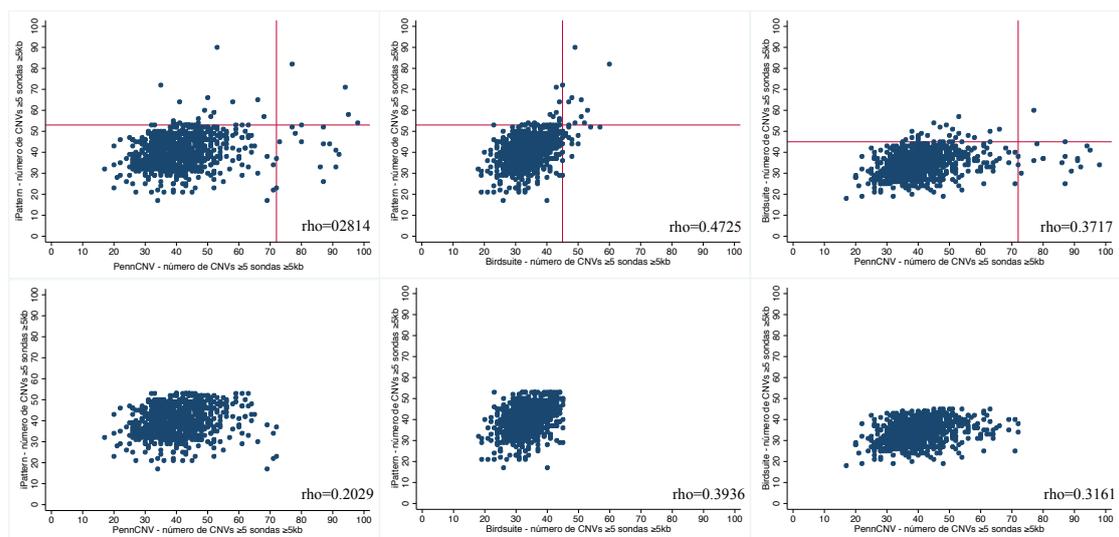


Figura Suplementaria 3: Correlación entre el número de CNVs ≥ 5 sondas ≥ 5 kb entre los algoritmos iPattern, PennCNV y Birdsuite antes y después de la selección de muestras por encima del percentil 95. Se muestra el valor del coeficiente de correlación de Spearman. Los algoritmos iPattern y PennCNV mostraron la mejor correlación.

De esta manera, los CNVs detectados por los algoritmos iPattern y Birdsuite (que presentaron la mejor correlación) fueron combinados e integrados dentro de una base de datos común, generando un total de 47,886 CNVs, de los cuales 30,454 (63.6%) fueron eliminados por ser detectados únicamente por un algoritmo, ser catalogadas como duplicaciones y deleciones al mismo tiempo ó por localizarse en regiones pseudo-autosómicas, gap, inmunoglobulinas o en ambos brazos del cromosoma.

Selección de CNVs con longitud ≥ 100 kb.

Finalmente, se eliminaron 13,409 CNVs con una longitud < 100 kb y 8 muestras que no presentaron CNVs con longitud ≥ 100 kb. De esta manera la base de datos común final esta formada por 4,023 CNVs en 880 muestras (686 sujetos con DT2 y 194 donadores) [Tablas Suplementarias 1 y 5, Figura Suplementaria 1].

Tabla Suplementaria 5. Pasos para la selección final de los CNVs, en la base de datos común

Paso para la selección de los CNVs	Autosomas		Cromosoma X		Total	
	# CNVs eliminados	# CNVs después del CC	# CNVs eliminados	# CNVs después del CC	# CNVs eliminados	# CNVs después del CC
Número de CNVs	0	46,908	0	978	0	47,886
CNVs detectados por un algoritmo (Birdsuite o iPattern)	29,309	17,599	846	132	30,155	17,731
CNVs en regiones pseudo-autosómicas	0	17,599	11	121	11	17,720
CNVs en regiones de inmunoglobulinas	203	17,396	0	121	203	17,517
CNVs localizados en brazos pq del cromosoma	3	17,393	0	121	3	17,514
CNVs con solapamiento $>50\%$ con regiones gap	0	17,393	0	121	0	17,514
CNVs catalogados como duplicación y deleción	82	17,311	0	121	82	17,432
CNVs con longitud < 100 kb	13,363	3,948	46	75	13,409	4,023
Resultado final del control de calidad de CNVs						
Número de CNVs	42,960	3,948	903	75	43,863	4,023

18. REFERENCIAS BIBLIOGRÁFICAS

1. WHO. Department of Noncommunicable Disease Surveillance Geneva. Definition, Diagnosis and classification of diabetes mellitus and its complications. Report of a WHO consultation. Part 1. Diagnosis and classification of diabetes mellitus. WHO/NCD/NCS/99.2 Disponible en: http://whqlibdoc.who.int/hq/1999/who_ncd_ncs_99.2.pdf. Último acceso 15 de marzo de 2014
2. Organización Mundial de la Salud. Diabetes. Datos fundamentales. Disponible en: <http://www.who.int/mediacentre/factsheets/fs312/es/index.html>. Último acceso 10 de marzo de 2014
3. Hunt KJ, Gonzalez ME, Lopez R, Haffner SM, Stern MP, Gonzalez-Villalpando C. Diabetes is more lethal in Mexicans and Mexican-Americans compared to Non-Hispanic whites. *Ann Epidemiol* 2011;21:899-906
4. Villalpando S, de la Cruz V, Rojas R, Shamah-Levy T, Avila MA, Gaona B, Rebollar R, Hernández L. Prevalence and distribution of type 2 diabetes mellitus in Mexican adult population: a probabilistic survey. *Salud Publica Mex* 2010;52 Suppl 1:S19-S26.
5. Gutiérrez JP, Rivera-Dommarco J, Shamah-Levy T, Villalpando-Hernández S, Franco A, Cuevas-Nasu L, Romero-Martínez M, Hernández-Ávila M. Encuesta Nacional de Salud y Nutrición 2012. Resultados Nacionales. Cuernavaca, México: Instituto Nacional de Salud Pública (MX), 2012.2 Disponible en: <http://ensanut.insp.mx/informes/ENSANUT2012ResultadosNacionales.pdf>. Último acceso 10 de marzo de 2014
6. Olaiz-Fernández G, Rojas R, Aguilar-Salinas CA, Rauda J, Villalpando S. Diabetes mellitus en adultos mexicanos. Resultados de la Encuesta Nacional de Salud 2000. *Salud Publica Mex* 2007;49 supl 3:S331-S337.
7. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Möhlig M, Pfeiffer AF, Spranger J, Thamer C, Häring HU, Fritsche A, Joost HG. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007;30:510-515.
8. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 2005;365:1333-1346.
9. Barroso I. Genetics of Type 2 diabetes. *Diabet Med* 2005;22:517-535.
10. Ali O. Genetics of type 2 diabetes. *World J Diabetes* 2013;4:114-123.
11. Murea M, Ma L, Freedman BI. Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications. *Rev Diabet Stud* 2012;9:6-22.
12. Ortega-Azorín C, Sorlí JV, Asensio EM, Coltell O, Martínez-González MÁ, Salas-Salvadó J, Covas MI, Arós F, Lapetra J, Serra-Majem L, Gómez-Gracia E, Fiol M, Sáez-Tormo G, Pintó X, Muñoz MA, Ros E, Ordovás JM, Estruch R, Corella D Associations of the FTO rs9939609 and the MC4R rs17782313

polymorphisms with type 2 diabetes are modulated by diet, being higher when adherence to the Mediterranean diet pattern is low. *Cardiovasc Diabetol* 2012;11:137.

13. Florez JC, Jablonski KA, Bayley N, Pollin TI, de Bakker PI, Shuldiner AR, Knowler WC, Nathan DM, Altshuler D. TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 2006;355:241–250.

14. McCarthy MI, Zeggini E. Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 2009;9:164-171.

15. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006;38:320-323.

16. McCarthy MI. Genomics, type 2 diabetes, and obesity. *N Engl J Med* 2010;363:2339-2350.

17. Ntzani EE, Kavvoura FK. Genetic risk factors for type 2 diabetes: insights from the emerging genomic evidence. *Curr Vasc Pharmacol* 2012;10:147-155.

18. DIABetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014;46:234-244

19. Online Mendelian Inheritance in Man. Disponible en: <http://www.ncbi.nlm.nih.gov/omim>. Último acceso 10 de marzo de 2014

20. Xu L, Hou Y, Bickhart DM, Song J, Liu GE. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. *Microarrays* 2013;2:171-185

21. Martinez-Marignac VL, Valladares A, Cameron E, Chan A, Perera A, Globus-Goldberg R, Wachter N, Kumate J, McKeigue P, O'Donnell D, Shriver MD, Cruz M, Parra EJ. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum Genet* 2007;120:807-819.

22. Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, Hidalgo-Miranda A, Contreras AV, Figueroa LU, Raska P, Jimenez-Sanchez G, Zolezzi IS, Torres M, Ponte CR, Ruiz Y, Salas A, Nguyen E, Eng C, Borjas L, Zabala W, Barreto G, González FR, Ibarra A, Taboada P, Porras L, Moreno F, Bigham A, Gutierrez G, Brutsaert T, León-Velarde F, Moore LG, Vargas E, Cruz M, Escobedo J, Rodriguez-Santana J, Rodriguez-Cintrón W, Chapela R, Ford JG, Bustamante C, Seminara D, Shriver M, Ziv E, Burchard EG, Haile R, Parra E, Carracedo A; LACE Consortium. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet* 2012;8:e1002554.

23. Martinez-Fierro ML, Beuten J, Leach RJ, Parra EJ, Cruz-Lopez M, Rangel-Villalobos H, Riego-Ruiz LR, Ortiz-Lopez R, Martinez-Rodriguez HG, Rojas-

Martinez A. Ancestry informative markers and admixture proportions in northeastern Mexico. *J Hum Genet* 2009;54:504-509.

24. Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil J, Fernandez-Lopez JC, Uribe-Figueroa L, Contreras A, Balam-Ortiz E, del Bosque-Plata L, Velazquez-Fernandez D, Lara C, Goya R, Hernandez-Lemus E, Davila C, Barrientos E, March S, Jimenez-Sanchez G. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc Natl Acad Sci U S A* 2009;106:8611-8616.

25. Salazar-Flores J, Dondiego-Aldape R, Rubi-Castellanos R, Anaya-Palafox M, Nuño-Arana I, Canseco-Avila LM, Flores-Flores G, Morales-Vallejo ME, Barojas-Pérez N, Muñoz-Valle JF, Campos-Gutiérrez R, Rangel-Villalobos H. Population structure and paternal admixture landscape on present-day Mexican-Mestizos revealed by Y-STR haplotypes. *Am J Hum Biol* 2010;22:401-409.

26. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev* 2002;11:505-512.

27. McKeigue, P.M., Carpenter, J., Parra, E.J., Shriver, M.D. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Annals of Human Genetics* 2000;64:171-186.

28. Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, Reich D. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004;74:979-1000.

29. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003;164:1567-1587.

30. Liu Y, Nyunoya T, Leng S, Belinsky SA, Tesfaigzi Y, Bruse S. Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics* 2013;7:1

31. Parra EJ, Below JE, Krithika S, Valladares A, Barta JL, Cox NJ, Hanis CL, Wachter N, Garcia-Mena J, Hu P, Shriver MD; Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, Kumate J, McKeigue PM, Escobedo J, Cruz M. Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. *Diabetologia* 2011;54:2038-2046.

32. Ortega-Pierres LE, Gómez García A, Rodríguez-Ayala E, Figueroa-Núñez B, Farias-Rodríguez VM, Higareda-Mendoza AE, Pardo-Galván MA, Cortés-García JC, López-Meza JE, Alvarez-Aguilar C. Angiotensin-1 converting enzyme insertion/deletion gene polymorphism in a Mexican population with diabetic nephropathy. *Med Clin (Barc)* 2007;129:6-10.

33. Santos A, Salguero ML, Gurrola C, Muñoz F, Roig-Melo E, Panduro A. The epsilon4 allele of apolipoprotein E gene is a potential risk factor for the severity of

macular edema in type 2 diabetic Mexican patients. *Ophthalmic Genet* 2002;23:13-19.

34. del Bosque-Plata L, Aguilar-Salinas CA, Tusié-Luna MT, Ramírez-Jiménez S, Rodríguez-Torres M, Aurón-Gómez M, Ramírez E, Velasco-Pérez ML, Ramírez-Silva A, Gómez-Pérez F, Hanis CL, Tsuchiya T, Yoshiuchi I, Cox NJ, Bell GI. Association of the calpain-10 gene with type 2 diabetes mellitus in a Mexican population. *Mol Genet Metab* 2004;81:122-126.

35. Quiroz-Mercado H, Suárez-Licona A, Fromow-Guerra J, López-Carasa G, Cárdenas-Hernández R, Ruiz-Morales JA, Portal-Celhay C, Granados J. Human lymphocyte antigen DR7 protects against proliferative retinopathy with type II diabetes mellitus. *Arch Med Res* 2002;33:123-127.

36. Perez-Luque E, Alaez C, Malacara JM, Garay ME, Fajardo ME, Nava LE, Gorodezky C. Protective effect of DRB1 locus against type 2 diabetes mellitus in Mexican Mestizos. *Hum Immunol* 2003;64:110-118.

37. Pérez-Luque E, Malacara JM, Olivo-Díaz A, Aláez C, Debaz H, Vázquez-García M, Garay ME, Nava LE, Burguete A, Gorodezky C. Contribution of HLA class II genes to end stage renal disease in mexican patients with type 2 diabetes mellitus. *Hum Immunol* 2000;61:1031-1038.

38. Below JE, Gamazon ER, Morrison JV, Konkashbaev A, Pluzhnikov A, McKeigue PM, Parra EJ, Elbein SC, Hallman DM, Nicolae DL, Bell GI, Cruz M, Cox NJ, Hanis CL. Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. *Diabetologia* 2011 ;54:2047-2055.

39. Weissglas-Volkov D, Huertas-Vazquez A, Suviolahti E, Lee J, Plaisier C, Canizales-Quinteros S, Tusie-Luna T, Aguilar-Salinas C, Taskinen MR, Pajukanta P. Common hepatic nuclear factor-4alpha variants are associated with high serum lipid levels and the metabolic syndrome. *Diabetes* 2006;55:1970-1977.

40. Martínez-Gómez LE, Cruz M, Martínez-Nava GA, Madrid-Marina V, Parra E, García-Mena J, Espinoza-Rojo M, Estrada-Velasco BI, Piza-Roman LF, Aguilera P, Burguete-García AI. A replication study of the IRS1, CAPN10, TCF7L2, and PPARG gene polymorphisms associated with type 2 diabetes in two different populations of Mexico. *Ann Hum Genet* 2011;75:612-620

41. Burguete-García AI, Cruz-Lopez M, Madrid-Marina V, Lopez-Ridaura R, Hernández-Avila M, Cortina B, Gómez RE, Velasco-Mondragón E. Association of Gly972Arg polymorphism of IRS1 gene with type 2 diabetes mellitus in lean participants of a national health survey in Mexico: a candidate gene study. *Metabolism* 2010;59:38-45.

42. Valladares-Salgado A, Angeles-Martínez J, Rosas M, García-Mena J, Utrera-Barillas D, Gómez-Díaz R, Escobedo-de la Peña J, Parra EJ, Cruz M. Association of polymorphisms within the transforming growth factor- β 1 gene with diabetic nephropathy and serum cholesterol and triglyceride concentrations. *Nephrology (Carlton)* 2010;15:644-648.

43. Gutiérrez-Vidal R, Rodríguez-Trejo A, Canizales-Quinteros S, Herrera-Cornejo M, Granados-Silvestre MA, Montúfar-Robles I, Ortiz-López MG, Menjívar M. LOC387761 polymorphism is associated with type 2 diabetes in the Mexican population. *Genet Test Mol Biomarkers* 2011;15:79-83.
44. Cameron EA, Martinez-Marignac VL, Chan A, Valladares A, Simmonds LV, Wacher N, Kumate J, McKeigue P, Shriver MD, Kittles R, Cruz M, Parra EJ. MGEA5-14 polymorphism and type 2 diabetes in Mexico City. *Am J Hum Biol* 2007;19:593-596.
45. Cruz M, Valladares-Salgado A, Garcia-Mena J, Ross K, Edwards M, Angeles-Martinez J, Ortega-Camarillo C, de la Peña JE, Burguete-Garcia AI, Wacher-Rodarte N, Ambriz R, Rivera R, D'artote AL, Peralta J, Parra EJ, Kumate J. Candidate gene association study conditioning on individual ancestry in patients with type 2 diabetes and metabolic syndrome from Mexico City. *Diabetes Metab Res Rev* 2010;26:261-270.
46. Huertas-Vazquez A, Plaisier C, Weissglas-Volkov D, Sinsheimer J, Canizales-Quinteros S, Cruz-Bautista I, Nikkola E, Herrera-Hernandez M, Davila-Cervantes A, Tusie-Luna T, Taskinen MR, Aguilar-Salinas C, Pajukanta P. TCF7L2 is associated with high serum triacylglycerol and differentially expressed in adipose tissue in families with familial combined hyperlipidaemia. *Diabetologia* 2008;51:62-69.
47. Parra EJ, Cameron E, Simmonds L, Valladares A, McKeigue P, Shriver M, Wacher N, Kumate J, Kittles R, Cruz M. Association of TCF7L2 polymorphisms with type 2 diabetes in Mexico City. *Clin Genet* 2007;71:359-366.
48. Perez-Luque E, Malacara JM, Garay-Sevilla ME, Fajardo ME. Association of the TNF- α -308G/A polymorphism with family history of type 2 diabetes mellitus in a Mexican population. *Clin Biochem* 2012;45:12-15.
49. Guzmán-Flores JM, Muñoz-Valle JF, Sánchez-Corona J, Cobián JG, Medina-Carrillo L, García-Zapién AG, Cruz-Quevedo EG, Flores-Martínez SE. Tumor necrosis factor-alpha gene promoter -308G/A and -238G/A polymorphisms in Mexican patients with type 2 diabetes mellitus. *Dis Markers* 2011;30:19-24.
50. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am J Hum Genet* 1999;64:1127-1140.
51. Vázquez MJ, Burt NP, Aguilar-Salinas CA, González-Villalpando C, Florez JC, Orozco L, Haiman CA, Tusié-Luna T, Altshuler D. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 2014;506:97-101
52. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME, Feuk L. Challenges and standards in integrating surveys of structural variation. *Nat Genet* 2007;39(7 Suppl):S7-S15.
53. Database of Genomic Variants DGV. Disponible en <http://dgvbeta.tcag.ca/dgv/app/home>. Último acceso 15 de marzo de 2014

54. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med* 2010;12:e8
55. Henrichsen CN, Chagnat E, Reymond A. Copy number variants, diseases and gene expression. *Hum Mol Genet* 2009;18:R1-R8.
56. Tomaiuolo R, Sangiuolo F, Bombieri C, Bonizzato A, Cardillo G, Raia V, D'Apice MR, Bettin MD, Pignatti PF, Castaldo G, Novelli G. Epidemiology and a novel procedure for large scale analysis of CFTR rearrangements in classic and atypical CF patients: a multicentric Italian study. *J Cyst Fibros* 2008;7:347-351.
57. Buchanan JA, Scherer SW. Contemplating effects of genomic structural variation. *Genet Med* 2008;10:639-647.
58. Data Sheet Affymetrix® Genome-Wide Human SNP Array 5.0. Disponible en http://media.affymetrix.com/support/technical/datasheets/genomewide_snp5_data_sheet.pdf. Último acceso 15 de marzo de 2014
59. Wang K, Bucan M. Copy Number Variation Detection via High-Density SNP Genotyping. *Cold Spring Harb Protoc*; 2008; doi:10.1101/pdb.top46
60. Affymetrix Power Tools (APT) Software Package. Disponible en: http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx#1_1. Último acceso 15 de marzo de 2014
61. Kim SY, Kim JH, Chung YJ. Effect of Combining Multiple CNV Defining Algorithms on the Reliability of CNV Calls from SNP Genotyping Data. *Genomics Inform* 2012;10:194-199
62. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Iglizoi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansy K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI Jr, Paterson AD, Pericak-Vance

MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijnsman EM, Scherer SW, Sutcliffe JS, Betancur C. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;466:368-372.

63. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, Macdonald JR, Mills R, Prasad A, Noonan K, Gribble S, Prigmore E, Donahoe PK, Smith RS, Park JH, Hurles ME, Carter NP, Lee C, Scherer SW, Feuk L. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 2011;29:512-520.

64. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;17:1665-1674.

65. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008;40:1253-1260.

66. Barnes C, Plagnol V, Fitzgerald T, Redon R, Marchini J, Clayton D, Hurles ME. A robust statistical method for case-control association testing with copy number variation. *Nat Genet* 2008;40:1245-1252.

67. QuantiSNP. Disponible en: <https://sites.google.com/site/quantisnp/home> Último acceso 15 de marzo de 2014

68. dChip Software. Disponible en: <http://www.hsph.harvard.edu/cli/complab/dchip/> Último acceso 15 de marzo de 2014

69. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008;455:237-241.

70. Shtir C, Pique-Regi R, Siegmund K, Morrison J, Schumacher F, Marjoram P. Copy number variation in the Framingham Heart Study. *BMC Proc* 2009;3 Suppl 7:S133.

71. Wang H, Hays NP, Das SK, Craig RL, Chu WS, Sharma N, Elbein SC. Phenotypic and molecular evaluation of a chromosome 1q region with linkage and association to type 2 diabetes in humans. *J Clin Endocrinol Metab* 2009;94:1401-1408.

72. Jeon JP, Shim SM, Nam HY, Ryu GM, Hong EJ, Kim HL, Han BG. Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. *BMC Genomics* 2010;11:426.

73. Kudo H, Emi M, Ishigaki Y, Tsunoda U, Hinokio Y, Ishii M, Sato H, Yamada T, Katagiri H, Oka Y. Frequent loss of genome gap region in 4p16.3 subtelomere in early-onset type 2 diabetes mellitus. *Exp Diabetes Res* 2011;2011:498460.

74. Bae JS, Cheong HS, Kim JH, Park BL, Kim JH, Park TJ, Kim JY, Pasaje CF, Lee JS, Park YJ, Park M, Park C, Koh I, Chung YJ, Lee JY, Shin HD. The genetic

effect of copy number variations on the risk of type 2 diabetes in a Korean population. *PLoS One* 2011;6:e19091

75. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K, Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D, Edkins S, Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D, Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P, Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, Kirov G, Langford CF, Lango-Allen H, Lathrop GM, Lee J, Lee KL, Lees C, Lewis K, Lindgren CM, Maisuria-Armer M, Maller J, Mansfield J, Martin P, Massey DC, McArdle WL, McGuffin P, McLay KE, Mentzer A, Mimmack ML, Morgan AE, Morris AP, Mowat C, Myers S, Newman W, Nimmo ER, O'Donovan MC, Onipinla A, Onyiah I, Ovington NR, Owen MJ, Palin K, Parnell K, Pernet D, Perry JR, Phillips A, Pinto D, Prescott NJ, Prokopenko I, Quail MA, Rafelt S, Rayner NW, Redon R, Reid DM, Renwick, Ring SM, Robertson N, Russell E, St Clair D, Sambrook JG, Sanderson JD, Schuilenburg H, Scott CE, Scott R, Seal S, Shaw-Hawkins S, Shields BM, Simmonds MJ, Smyth DJ, Somaskantharajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Su Z, Symmons DP, Thompson JR, Thomson W, Travers ME, Turnbull C, Valsesia A, Walker M, Walker NM, Wallace C, Warren-Perry M, Watkins NA, Webster J, Weedon MN, Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling TM, Lee C, McVean G, Munroe PB, Palotie A, Sawcer SJ, Scherer SW, Strachan DP, Tyler-Smith C, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Gough SC, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR, Worthington J, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand W, Parkes M, Rahman N, Todd JA, Samani NJ, Donnelly P. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010;464:713-720.

76. Chen X, Li X, Wang P, Liu Y, Zhang Z, Zhao G, Xu H, Zhu J, Qin X, Chen S, Hu L, Kong X. Novel association strategy with copy number variation for identifying new risk loci of human diseases. *PLoS One* 2010;5:e12185.

77. Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL. Genome-wide meta-analysis for severe diabetic retinopathy. *Hum Mol Genet* 2011;20:2472-2481.

78. Irvin MR, Wineinger NE, Rice TK, Pajewski NM, Kabagambe EK, Gu CC, Pankow J, North KE, Wilk JB, Freedman BI, Franceschini N, Broeckel U, Tiwari HK, Arnett DK. Genome-wide detection of allele specific copy number variation associated with insulin resistance in African Americans from the HyperGEN study. *PLoS One* 2011;6:e24052

79. Plengvidhya N, Chanprasert K, Tangjittipokin W, Thongnoppakhun W, Yenchitsomanus PT. Identification of copy number variation of CAPN10 in Thais with type 2 diabetes by multiplex PCR and denaturing high performance liquid chromatography (DHPLC). *Gene* 2012;506:383-386.
80. Blackburn A, Göring HH, Dean A, Carless MA, Dyer T, Kumar S, Fowler S, Curran JE, Almasy L, Mahaney M, Comuzzie A, Duggirala R, Blangero J, Lehman DM. Utilizing extended pedigree information for discovery and confirmation of copy number variable regions among Mexican Americans. *Eur J Hum Genet* 2013;21:404-409.
81. MODIFICACION a la Norma Oficial Mexicana NOM-015-SSA2-1994, Para la prevención, tratamiento y control de la diabetes mellitus en la atención primaria para quedar como Norma Oficial Mexicana NOM-015-SSA2-1994, Para la prevención, tratamiento y control de la diabetes.
82. American Diabetes Association. Standards of medical care in diabetes--2014. *Diabetes Care* 2014;37 Suppl 1:S14-S80
83. Universidad de Salamanca. Diccionario médico-biológico, histórico y etimológico. Disponible en: dicciomed.eusal.es/ Último acceso 15 de marzo de 2014
84. Diccionario de referencia. Disponible en <http://dictionary.reference.com>. Último acceso 15 de marzo de 2014
85. Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, Leon-Velarde F, Moore LG, Vargas E, McKeigue PM, Shriver MD, Parra EJ. A genomewide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet* 2007;80:1171-1178.
86. Organización Mundial de la Salud. Obesidad y sobrepeso, nota descriptiva No 311. Disponible en <http://www.who.int/mediacentre/factsheets/fs311/es/>. Último acceso 15 de marzo de 2014
87. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575.
88. Genetic Information Nondiscrimination Act (GINA) of 2008, disponible en: <http://www.genome.gov/24519851>. Último acceso 15 de marzo de 2014
89. Arredondo A, Reyes G. Health disparities from economic burden of diabetes in middle-income countries: evidence from México. *PLoS One* 2013;8:e68443
90. Frazer KA1, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 2009;10:241-251
91. Johnson ME1, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE. Positive selection of a gene family during the emergence of humans and African apes. *Nature* 20014;413:514-519.

92. Hornan DM, Peirson SN, Hardcastle AJ, Molday RS, Cheetham ME, Webster AR. Novel retinal and cone photoreceptor transcripts revealed by human macular expression profiling. *Invest Ophthalmol Vis Sci* 2007;48:5388-5396.
93. Sujit Maiti, Kiran Halagur Bhoge Gowda Kumar, Christina A. Castellani, Richard O'Reilly, Shiva M. Singh. Ontogenetic De Novo Copy Number Variations (CNVs) as a Source of Genetic Individuality: Studies on Two Families with MZD Twins for Schizophrenia. *PLoS One* 2011;6:e17125
94. Grond-Ginsbach C, Chen B, Pjontek R, Wiest T, Jiang Y, Burwinkel B, Tchatchou S, Krawczak M, Schreiber S, Brandt T, Kloss M, Arnold ML, Hemminki K, Lichy C, Lyrer PA, Hausser I, Engelter ST. Copy number variation in patients with cervical artery dissection. *Eur J Hum Genet* 2012;20:1295-1299
95. Ouyang B1, Baxter CS, Lam HM, Yeramaneni S, Levin L, Haynes E, Ho SM. Hypomethylation of dual specificity phosphatase 22 promoter correlates with duration of service in firefighters and is inducible by low-dose benzo[a]pyrene. *J Occup Environ Med* 2012;54:774-780.