



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MÉDICAS,

ODONTOLÓGICAS Y DE LA SALUD

INSTITUTO NACIONAL DE CIENCIAS MÉDICAS Y NUTRICIÓN SALVADOR ZUBIRÁN

DESARROLLO DE UN MODELO PREDICTIVO DE LA INCIDENCIA DE LA DIABETES TIPO 2 EN ADULTOS

MEXICANOS USANDO VARIABLES CLÍNICAS Y METABOLITOS CUANTIFICADOS POR RESONANCIA

MAGNETICA NUCLEAR

TESIS QUE PARA OPTAR POR EL GRADO DE MAESTRIA EN CIENCIAS MÉDICAS PRESENTA:

DRA. CRISTINA MARTÍNEZ BERDEJA

TUTOR: DR. CARLOS ALBERTO AGUILAR SALINAS, INCMNSZ

MIEMBROS DEL COMITÉ TUTOR: DRA. PALOMA ALMEDA VALDÉS, INCMNSZ, DR. MIGUEL
KLÜNDER KLÜNDER, HIM, DRA. HORTENSIA MORENO MACÍAS, FACULTAD DE MEDICINA, DR.

JORGE ROJAS SERRANO, INER

CIUDAD DE MÉXICO, JUNIO 2016.



Universidad Nacional
Autónoma de México



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 está 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.

2. AGRADECIMIENTOS

A mi hija Isabel, quien le da un nuevo sentido a mi vida.

A mis padres, Cristina y Raúl por su apoyo incondicional.

A mi hermana Alejandra, por su motivación y por servir de ejemplo en el campo de la investigación.

Al Dr. Carlos A. Aguilar, por ser un gran asesor académico y un pilar fundamental en mi formación profesional.

A la Dra. Hortensia Moreno, por su ayuda y gran disposición para realizar el análisis estadístico.

3. ÍNDICE

1.	Portada de tesis	Pág. 1
2.	Agradecimientos	Pág. 2
3.	Índice.	Pág. 3
4.	Título	Pág. 5
5.	Resumen	Pág.5
6.	Introducción	Pág. 7
7.	Antecedentes	Pág. 8
8.	Objetivos	Pág. 10
9.	Hipótesis	Pág. 11
10.	Métodos	Pág. 11
10.0	Métodos: Diseño del estudio	Pág. 11
10.1	Métodos: Criterios de selección	Pág. 11
10.2	Métodos: Procedimientos	Pág. 12
10.3	Métodos: Desenlaces y variables	Pág. 13
10.4	Métodos: Análisis estadístico	Pág. 14
11.	Resultados	Pág. 15
11.0	Resultados: Participantes.	Pág. 15

11.1	Resultados: Análisis descriptivo de los metabolitos.	Pág. 16
11.2	Resultados: Modelos de predicción	Pág. 18
12.	Discusión	Pág. 24
13.	Fortalezas y debilidades del estudio	Pág. 28
14.	Conclusión	Pág. 29
15.	Bibliografía	Pág. 29

4. TITULO

Desarrollo de un modelo predictivo de la incidencia de la diabetes tipo 2 en adultos mexicanos usando variables clínicas y metabolitos cuantificados por resonancia magnética nuclear

5. RESUMEN

La prevención de la diabetes depende de la identificación y el tratamiento de los individuos en riesgo. Existe la necesidad de crear un modelo predictivo de diabetes incidente para mexicanos, que incorpore a los factores de riesgo conocidos y a la metabolómica como nueva estrategia. Se plantea la hipótesis de que la combinación de varios indicadores clínicos y de metabolómica permitirá predecir la incidencia de diabetes con mayor precisión que lo alcanzado con indicadores clínicos convencionales. Métodos: se trata de un estudio de casos y controles anidado en una cohorte. La cohorte de adultos incluyó individuos sin diabetes de 25 a 75 años de edad residentes del centro del país. Se realizó una evaluación basal y otra a los 3 años para ver qué sujetos desarrollaban diabetes. Se midieron parámetros bioquímicos y antropométricos: peso, talla, perímetro de cintura y tensión arterial. Se obtuvo información de los antecedentes médicos, de la dieta y del ejercicio mediante cuestionarios estandarizados. Los criterios de selección para los casos fueron los pacientes que tuvieran diabetes incidente a los 3 años de seguimiento, definido como un valor de glucosa de ayuno ≥ 126 mg/dl. Los controles fueron todos los participantes sin diabetes a los 3 años de seguimiento. Se excluyeron a todos los individuos que no tuvieran las mediciones completas. Análisis estadístico: se realizó el análisis descriptivo para describir las características de la cohorte, del grupo de casos y de controles y de los metabolitos. Se utilizó regresión logística para buscar las variables asociadas a la diabetes incidente. Se crearon modelos con regresión logística de solo los metabolitos, solo los factores clínicos y la combinación de

ambos. La capacidad de predicción se estimó con el área bajo la curva. Resultados: la cohorte final fue de 6144 adultos, de los cuales 345 tuvieron diabetes incidente a los 3 años. Se seleccionó un grupo de casos y de controles pareado por edad, sexo y puntaje de propensión (propensity score). El grupo analizado incluyó 957 controles sin diabetes y 123 casos con diabetes. Se encontraron 26 metabolitos con diferencia estadística significativa ($p < 0.05$) entre los grupos. El modelo de solo metabolitos alcanzó un AUC = 0.815, el modelo de solo factores clínicos un AUC = 0.771 y el modelo completo un AUC = 0.817. El modelo completo predice mejor para los sujetos \geq a 50 años y con un IMC ≥ 30 kg/m². Conclusión: el adicionar los metabolitos al modelo clínico convencional mejora la capacidad de predicción de diabetes incidente de manera modesta.

6. INTRODUCCION

Actualmente existen 383 millones de personas con diabetes. Se estima que la diabetes tipo 2 afectará a más de 550 millones de personas para el 2030 (1). En la Encuesta Nacional de Salud y Nutrición (ENSANUT 2006) se encontró una prevalencia de la diabetes del 14.4%, de los cuales el 7.3% corresponde a casos previamente diagnosticados y el 7.1% a casos diagnosticados durante la encuesta. En 2012 el porcentaje de casos diagnosticados se incrementó a 9.2%, lo que se traduce en 6.4 millones de adultos afectados (2). Debido a los altos costos que genera la enfermedad, tanto personales como para los sistemas de salud, resulta esencial identificar a los sujetos en alto riesgo de diabetes, como primer paso para proveer intervenciones preventivas en diabetes (3). La historia médica y las pruebas de laboratorio proveen información sobre el riesgo futuro de presentar diabetes (4), sin embargo muchos de estos predictores surgen después de muchos años de disfunción metabólica subclínica (5). Los marcadores nuevos, como son los metabolitos, pueden ayudar a elucidar aspectos de disfunción metabólica que contribuyan al riesgo de diabetes y a la identificación temprana de los individuos que se benefician de las terapias preventivas en diabetes (6).

Se han utilizado distintas herramientas que ayuden a mejorar la predicción en diabetes. Una de ellas es la información genética, la cual incluye a los polimorfismos de nucleótido único, agrupados en un puntaje de riesgo genético. La información genética tiene una baja capacidad para predecir diabetes incidente y solo mejora de manera moderada la predicción de los factores de riesgo clínicos tradicionales (7). Dentro de las nuevas herramientas para mejorar la predicción en diabetes está la metabolómica.

La metabolómica es un conjunto de mediciones globales realizadas en muestras biológicas, que tiene como meta cuantificar tantos metabolitos como sea posible y evaluar cambios en los niveles

de metabolitos bajo algún tipo de estrés o reto. Existen 2 plataformas analíticas para el estudio de los metabolitos: la espectroscopía de masas acoplada a cromatografía líquida y la espectroscopía basada en resonancia magnética nuclear (RMN) (8, 9).

Las aplicaciones de la metabolómica en investigación en diabetes incluyen el descubrimiento de biomarcadores diagnósticos y pronósticos, el descubrimiento de vías metabólicas nuevas, de biomarcadores funcionales en la acción de nuevos fármacos y la identificación de efectos adversos de los fármacos (8, 9).

El panel de metabolómica utilizado por los investigadores Kettunen J, Ala-Korpela M et al. incluye la medición de las subclases de lipoproteínas, los extractos de lípidos en suero, los aminoácidos y otros metabolitos de bajo peso molecular y otras medidas que derivan de la medición de los metabolitos como son los metabolitos de la glicólisis, los cuerpos cetónicos, las glucoproteínas y la albúmina. Dicho panel se obtiene mediante la espectroscopía acoplada a RMN (6, 10).

Los metabolitos descritos en la literatura asociados a la predicción del riesgo de diabetes son los aminoácidos ramificados y aromáticos (11), los triacilgliceroles de pocos carbonos y doble enlace (12), la relación glutamina glutamato (13), el ácido 2 aminoadípico (14) y las acilcarnitinas (15). En otro estudio se encontró que un cetoácido ramificado llamado 3-metil-2-oxovalerato es un biomarcador predictivo para la glucosa anormal de ayuno (16).

7. ANTECEDENTES

En el estudio de TJ Wang et al. se encontraron cinco metabolitos que predicen diabetes a futuro, tres de ellos son aminoácidos ramificados: leucina, isoleucina y valina y dos más son aminoácidos aromáticos: fenilalanina y tirosina. Se crearon modelos de regresión logística para evaluar la

asociación entre el nivel basal de los metabolitos y la diabetes a futuro, se ajustó para la edad, el sexo, el IMC y la glucosa de ayuno. Para los 5 aminoácidos de interés, por cada incremento en la desviación estándar en el marcador log (log marker), se asoció a un incremento en los momios del 57 al 102% para presentar diabetes en el futuro. La mejor combinación de aminoácidos para predecir diabetes a futuro fue la isoleucina, la fenilalanina y la tirosina. Los aminoácidos pueden predecir diabetes hasta 12 años antes del diagnóstico clínico (11).

En otro estudio de TJ Wang et al. en el cual se utilizó a la metabolómica, se encontró que el ácido 2 aminoadípico estaba fuertemente asociado con el riesgo de desarrollar diabetes. En estudios experimentales se encontró que el ácido 2 aminoadípico, era capaz de bajar los niveles de la glucosa de ayuno en ratones, de potenciar la secreción de insulina de una línea de células β , así como en islotes humanos y murinos. Además de encontrarse elevado hasta 12 años antes de la aparición de diabetes manifiesta. Se sugiere que el ácido 2 aminoadípico es un marcador y un potencial modulador de la homeostasis de glucosa en humanos (14).

Se crearon modelos de regresión logística ajustados por sexo y por edad para probar la asociación de los metabolitos con la diabetes incidente. La capacidad de predicción de cada modelo se evaluó mediante el área bajo la curva. Se incluyeron 1622 pacientes no diabéticos del Framingham Offspring Study (FOS). Se observaron 206 casos de diabetes incidente durante 13.5 años de seguimiento. El modelo que incluyó solo a los metabolitos alcanzó un área bajo la curva de 0.803. El modelo con mayor área bajo la curva (AUC = 0.880) incluyó el puntaje de riesgo genético (GRS), los metabolitos (9) y los factores clínicos. El modelo que incluyó a los metabolitos más los factores clínicos alcanzó un área bajo la curva de 0.874. Los metabolitos asociados con diabetes incidente fueron 3 aminoácidos (isoleucina, tirosina y fenilalanina) y 6 metabolitos de lípidos (C18:2 lisofosfatidilcolina LPC, C38:6 fosfatidilcolina PC, triacilglicerolos: C44:1, C48:0, C52:1 y C56:9 TAG)

(17). Se concluye que tanto los metabolitos como la información genética proveen información complementaria para la predicción a futuro de la diabetes tipo 2. Los nuevos marcadores del riesgo de diabetes parecen mejorar la capacidad de predicción de la diabetes tipo 2 de manera modesta (17).

El primer modelo de predicción de la incidencia de diabetes en mexicanos fue creado en esta misma cohorte utilizando parámetros clínicos y de laboratorio. Se alcanzó un área bajo la curva de 0.764 en población general. La mayoría de los estudios sobre modelos de predicción de la incidencia de diabetes como de metabolómica han sido realizados en población caucásica. Resulta importante conocer cuáles son los metabolitos y parámetros que mejor representan los defectos metabólicos en los mexicanos. Por otro lado, resulta esencial conocer si el agregar los metabolitos a los parámetros clínicos convencionales aumenta la capacidad de predicción de diabetes incidente en mexicanos.

8. OBJETIVOS

Crear un modelo predictivo de la incidencia de la diabetes tipo 2 compuesto por variables clínicas y por la concentración de metabolitos cuantificados por resonancia magnética nuclear que supere la capacidad predictiva de los modelos actuales en adultos mexicanos.

Comparar la capacidad predictiva de dos modelos predictivos de la incidencia de diabetes: uno basado en los parámetros de metabolómica y otro basado en las variables clínicas convencionales.

9. HIPOTESIS

La combinación de varios indicadores clínicos y de metabólica permitirá predecir la incidencia de diabetes con mayor precisión que lo alcanzado con los indicadores clínicos convencionales.

10. METODOS

El estudio forma parte del proyecto “Consolidación de una cohorte de adultos residentes en el Centro de México diseñada para medir la incidencia de la diabetes y de los componentes del síndrome metabólico”. La cohorte analizada incluyó 6144 adultos normoglucémicos, residentes de las zonas urbanas del Centro de México. El tiempo de seguimiento fue de 3 años.

10.0 Diseño del estudio

Se diseñó un estudio de casos y controles anidado en una cohorte. Se seleccionaron ocho controles por cada caso. Los pacientes seleccionados fueron pareados por edad, sexo y puntaje de propensión (propensity score), además debían tener una medición completa de los 233 metabolitos.

El pareamiento por puntaje de propensión (propensity score) es un método utilizado para reducir los efectos de la confusión en estudios observacionales. Es un puntaje de balanceo o equilibrio que hace que las covariables basales sean similares entre los tratados y los no tratados o entre los casos y los controles. Se utilizaron los siguientes parámetros para el puntaje de propensión: el IMC ± 2 kg/m², la edad ± 5 años y el sexo (29).

10.1 Criterios de selección

Los criterios de inclusión para los casos fueron sujetos con diabetes incidente y para los controles fueron sujetos sin diabetes. Se definió como diabetes incidente a los participantes que tuvieran glucosa de ayuno mayor o igual a 126 mg/dl (≥ 7 mmol/l) o que estuvieran con hipoglucemiantes orales o insulina indicada por el médico tratante en la segunda visita. Además debían tener las dos valoraciones completas: la visita inicial y la segunda visita a los 2 a 3 años. Se tomó como criterio de eliminación aquellos casos y controles que no tuvieran completas las dos valoraciones.

10.2 Procedimientos

En la valoración inicial, se les realizó a los participantes del estudio, una historia médica mediante un cuestionario estandarizado, se les midió el peso, la talla, el perímetro de la cintura y la cadera, la presión arterial, mediante métodos de medición estandarizados. Se aplicó un cuestionario validado sobre la actividad física y la conducta alimentaria y un registro retrospectivo del consumo en la dieta. Se les pidió un ayuno de 9 a 12 horas para la toma de una muestra de sangre para la medición de la glucosa, el colesterol, los triglicéridos, el colesterol HDL, el colesterol LDL, la apolipoproteína B, la proteína C reactiva, la insulina, la tirotrópina, la tiroxina libre y los 233 metabolitos. Todos los participantes firmaron un consentimiento informado antes de su participación en el estudio.

En la segunda evaluación se realizó el mismo procedimiento antes mencionado a los 2 a 3 años de la valoración inicial.

Las muestras fueron procesadas en el laboratorio central del Departamento de Endocrinología y Metabolismo del Instituto Nacional de Ciencias Médicas y Nutrición, usando técnicas estandarizadas con reactivos comerciales. Para la determinación de la glucemia y los lípidos en sangre se usaron métodos enzimáticos colorimétricos (Synchron CX5, Beckman Coulter), la determinación de la apolipoproteína B se realizó por el método de nefelometría (IMAGE,

Beckman Coulter). La insulina se determinó mediante tecnología de enzimoanálisis de micropartículas (MEIA con AxSYM, Abbott). La proteína C reactiva se determinó por inmunofelometría (BN ProSpec, Siemens). Las muestras se mantuvieron en congelación (-80°C) hasta su medición.

Los 233 metabolitos fueron medidos de alícuotas almacenadas de plasma, de suero y de sangre total. Para la medición de los metabolitos se utilizó espectroscopía acoplada a resonancia magnética nuclear (espectrómetro: Bruker AVANCE III 500 MHz, Bruker AVANCE III HD 600 MHz). Los metabolitos fueron medidos en Finlandia por el grupo de investigadores del Profesor Ala-Korpela.

10.3 Desenlaces y variables

El desenlace primario fue la incidencia de diabetes.

Los parámetros clínicos y bioquímicos se catalogaron como variables dimensionales, continuas, excepto el sexo que fue clasificada como una variable dicotómica.

Dentro de los parámetros clínicos se incluyó a la edad (años), el IMC (kg/m²), la talla (mt), el peso (kg), el perímetro de la cintura (cm), el perímetro de la cadera (cm) y la presión arterial (mmHg).

Los parámetros bioquímicos incluyeron a la glucosa (mg/dl), el colesterol total (mg/dl), el colesterol HDL (mg/dl), el colesterol LDL (mg/dl), los triglicéridos (mg/dl), la insulina (μU/l), la proteína C reactiva (mg/l), la apolipoproteína B (mg/dl), la TSH (μU/ml) y la tiroxina libre (μg/dl).

Los parámetros de metabolómica también se clasificaron como variables dimensionales, continuas.

La variable dependiente de diabetes incidente, al ser categórica, se codificó como 1 para los pacientes con diabetes incidente y como 0 para los pacientes sin diabetes. La variable independiente fue cada uno de los 233 metabolitos (se detallan en la página 15) y cada uno de los parámetros clínicos o bioquímicos a los cuales nos referiremos como factores clínicos.

10.4 Análisis estadístico

Se realizó el análisis estadístico descriptivo para describir las características generales de la cohorte y del grupo de los casos y los controles, así como las características de los metabolitos, para lo cual se utilizó mediana e intervalo intercuartilar para las variables no paramétricas. Se utilizó la prueba de U de Mann Whitney para la comparación de medianas de dos grupos.

Se realizó un análisis multivariado con regresión logística para ver la asociación y la significancia estadística de cada uno de los metabolitos con la variable de diabetes incidente, ajustando por edad y por sexo. Se utilizó regresión por pasos hacia atrás (backward stepwise regression) para seleccionar las covariables para los modelos. Se consideró una $p \leq 0.05$ para la selección de las covariables. Además se consideraron aquellos metabolitos, previamente descritos en la literatura que predicen diabetes incidente. Se crearon modelos de predicción para diabetes incidente por medio de regresión logística, uno solo con metabolitos, otro solo con factores clínicos y un tercero incluyendo tanto metabolitos como factores clínicos. Se calculó el área bajo la curva para evaluar la capacidad de predicción de diabetes incidente de cada modelo. Se utilizó la prueba de χ^2 para comparar las dos curvas ROC. El análisis estadístico se realizó con el programa STATA versión 12 para Windows.

11. RESULTADOS

11.0 Participantes

La cohorte final analizada incluyó a 6144 participantes normoglucémicos. Participaron más mujeres en el estudio (n= 4091). Se observaron 345 casos de diabetes incidente en un periodo de seguimiento de 3 años. Se seleccionó un grupo de casos y de controles pareado por edad, sexo y propensity score. El grupo analizado incluyó 957 controles sin diabetes y 123 casos con diabetes. Todos los pacientes debían tener completa la medición de los 233 metabolitos. Se encontraron diferencias al comparar ambos grupos en su valoración basal en todas las características, excepto para el colesterol (total, HDL, LDL), la apolipoproteína B, los triglicéridos, la cadera, la talla y la presión arterial. Las características basales de la cohorte final y del grupo de casos y controles seleccionado se muestran en la tabla 1.

Tabla 1. Características de la cohorte y del grupo de casos y controles analizado

Factores clínicos	Cohorte n = 6144	Sin diabetes incidente n = 957	Con diabetes incidente n = 123	p
sexo M/H (n)	4091/2053	675/87	282/36	-
edad (años)	42 (35 – 49)	46 (39 – 52)	46 (38 – 52)	0.9518
tiempo seguimiento (meses)	27 (24 – 31)	27 (24 – 37)	31 (24 – 43)	0.0023
IMC (kg/m ²)	28.04 (25.59 – 31.05)	28.02 (25.65 – 31.06)	29.34 (26.76 – 33.36)	0.0006
peso (kg)	71.7 (63.7 – 81.8)	70.5 (62.3 – 80.5)	75.5 (65.5 – 84.3)	0.0028
talla (mt)	1.59 (1.53 – 1.66)	1.57 (1.53 – 1.65)	1.57 (1.52 – 1.65)	0.960
cintura (cm)	92 (85 – 99)	92 (86 – 99)	95 (87 – 105.5)	0.0034
cadera (cm)	105.1 (103.7 – 105.1)	105.1 (103.7 – 105.1)	105.1 (103.7 – 108)	0.1057
TAS (mmHg)	114 (110 – 120)	115 (110 – 120)	115 (110 – 125)	0.264
TAD (mmHg)	79 (70 – 80)	79 (70 – 80)	79 (70 – 80)	0.548
glucosa (mg/dl)	85 (79 – 93)	86 (80 – 93)	99 (89 – 107)	<0.001
insulina (μU/ml)	10.15 (7.0 – 14.7)	10.2 (7.0 – 14.6)	12.0 (8.8 – 19.5)	0.0001
HOMA-IR	2.14 (1.42 – 3.23)	2.18 (1.44 – 3.20)	2.8 (2.08 – 4.73)	<0.001
c-HDL (mg/dl)	43 (36 – 51)	43 (36 – 51)	39 (33 – 46)	0.0008
colesterol (mg/dl)	202 (177 – 231)	203 (178 – 229)	202 (177 – 227)	0.972
triglicéridos (mg/dl)	158 (110 – 227)	158 (111 – 228)	180 (127 – 258)	0.101
c-LDL (mg/dl)	125 (109 – 140)	124 (106 – 139)	124 (112 – 138)	0.632
apo B (mg/dl)	106 (89.6 – 125)	107 (90 – 125)	108 (94 – 128)	0.2907
PCR (mg/l)	2.02 (1.01 – 3.84)	1.91 (1.01 – 3.53)	2.93 (1.45 – 4.39)	0.0011

Los valores de los factores clínicos están expresados en mediana (Md) e intervalo intercuartil (IIC). Se utilizó U Mann Whitney para la diferencia de Md. La selección de los pacientes se realizó pareando por edad, sexo y propensity score.

11.1 Análisis descriptivo de los metabolitos

Los metabolitos fueron cuantificados por espectroscopía acoplada a RMN incluyeron a los cuerpos cetónicos (acetato, acetoacetato, 3 β hidroxibutirato), a los metabolitos relacionados a la glucólisis (glucosa, lactato, piruvato, citrato, glicerol), a los metabolitos relacionados a la inflamación (α 1 ácido glucoproteína), a los ácidos grasos y la saturación (ácidos grasos totales, ácidos grasos de cierta longitud de cadena y cierto grado de saturación), a los ácidos grasos (FA) (omega 3, omega 6 (FAw6), ácidos grasos poliinsaturados (PUFA), ácidos grasos monoinsaturados (MUFA), ácidos grasos saturados, ácido docosahexanoico 22:6, ácido linoleico 18:2 (LA), ácido linoleico conjugado), a las subclases de lipoproteínas (VLDL, IDL, LDL, HDL), a las mediciones de lípidos para cada subclase (colesterol esterificado, colesterol libre, triglicéridos, fosfolípidos, colesterol total, lípidos totales, concentración de partículas), a los aminoácidos (alanina, glutamina, glicina, histidina), a los aminoácidos ramificados (isoleucina, leucina, valina), a los aminoácidos aromáticos (fenilalanina, tirosina), al colesterol (colesterol VLDL, LDL, HDL, HDL2 HDL3, colesterol, colesterol libre, colesterol esterificado, colesterol remanente), a las apolipoproteínas (Apo A1, Apo B, Apo B/Apo A1), a los metabolitos del balance de fluidos (creatinina, albúmina), los glicéridos y los fosfolípidos (VLDL triglicéridos, LDL triglicéridos, HDL triglicéridos, triglicéridos, diglicéridos, fosfoglicéridos, relación diglicérido/triglicérido, relación triglicérido/fosfoglicérido, fosfatidilcolina y otras colinas, esfingomiélinas, colinas totales) y al tamaño de la partícula de la lipoproteína (diámetro medio de partículas VLDL, LDL y HDL).

Se realizó una comparación de los metabolitos entre los casos y los controles. Se encontraron 26 metabolitos con diferencia estadística significativa ($p < 0.05$) entre los grupos (tabla 2). Los

metabolitos encontrados fueron los siguientes: lipoproteína de densidad intermedia en su medición de triglicéridos (IDL TG), lipoproteína de alta densidad de tamaño extra grande en su medición de fosfolípidos (XL HDL PL), lipoproteína de alta densidad grande en la medición de la concentración de dichas lipoproteínas, los lípidos totales, los fosfolípidos, el colesterol, los ésteres de colesterol, el colesterol libre [L HDL (P, L, PL, C, CE, FC)], la lipoproteína de alta densidad de tamaño mediano en la medición de los componentes enlistados en el paréntesis [M HDL (P, L, C, CE, FC)], la lipoproteína de alta densidad de tamaño pequeño [S HDL (C, CE)], el diámetro medio de las lipoproteínas de alta densidad (HDL D), el colesterol de la lipoproteína de alta densidad (HDL C), la descripción estimada del largo de las cadenas de los ácidos grasos (FALen), las relaciones de los siguientes ácidos grasos medidos: LA/FA, FAw6/FA, PUFA/FA, MUFA/FA, la glucosa (Glc), la isoleucina (Ile), la leucina (Leu) y la tirosina (Tyr).

Tabla 2. Análisis descriptivo de los metabolitos

Metabolito	Sin diabetes incidente n = 957	Con diabetes incidente n = 123	p
IDL-TG (mmol/l)	0.125 (0.105 – 0.15)	0.131 (0.113 – 0.156)	0.041
XL-HDL-PL (mmol/l)	0.108 (0.0702 – 0.151)	0.0912 (0.055 – 0.127)	0.006
L-HDL-P (mol/l)	7.46e-07 (5.67e-07 – 9.64e-07)	6.24e-07 (4.59e-07 – 8.79e-07)	0.0007
L-HDL-L (mmol/l)	0.459 (0.345 – 0.6)	0.384 (0.279 – 0.543)	0.0007
L-HDL-PL (mmol/l)	0.24 (0.186 – 0.303)	0.209 (0.154 – 0.281)	0.0009
L-HDL-C (mmol/l)	0.192 (0.134 – 0.264)	0.157 (0.106 – 0.228)	0.0008
L-HDL-CE (mmol/l)	0.154 (0.11 – 0.211)	0.128 (0.0879 – 0.183)	0.0008
L-HDL-FC (mmol/l)	0.0372 (0.0237 – 0.0538)	0.0286 (0.0159 – 0.0473)	0.0008
M-HDL-P (mol/l)	1.78e-06 (1.55e-06 – 2.02e-06)	1.68e-06 (1.52e-06 – 1.93e-06)	0.023
M-HDL-L (mmol/l)	0.749 (0.653 – 0.854)	0.705 (0.639 – 0.816)	0.019
M-HDL-C (mmol/l)	0.341 (0.288 – 0.4)	0.318 (0.27 – 0.369)	0.002
M-HDL-CE (mmol/l)	0.273 (0.232 – 0.322)	0.253 (0.216 – 0.286)	0.001
M-HDL-FC (mmol/l)	0.0674 (0.0572 – 0.0794)	0.0629 (0.0545 – 0.0734)	0.032
S-HDL-C (mmol/l)	0.443 (0.412 – 0.478)	0.432 (0.402 – 0.463)	0.025
S-HDL-CE (mmol/l)	0.324 (0.294 – 0.355)	0.311 (0.286 – 0.346)	0.007
HDL-D (nm)	9.7 (9.6 – 9.8)	9.6 (9.5 – 9.8)	0.004
HDL-C (mmol/l)	1.0 (0.0 – 1.0)	1.0 (0.0 – 1.0)	0.0004
FALen	17.3 (17.1 – 17.5)	17.4 (17.2 – 17.5)	0.007

LA/FA (%)	28.9 (26.9 – 30.9)	27.8 (25.6 – 30.2)	0.001
FAw6/FA (%)	34.6 (32.5 – 36.6)	33.9 (31.7 – 35.7)	0.001
PUFA/FA (%)	37.6 (35.4 – 39.6)	36.7 (34.6 – 38.6)	0.002
MUFA/FA (%)	24.5 (22.7 – 26.6)	25.2 (23.5 – 26.9)	0.022
Glc (mmol/l)	3.8 (3.5 – 4.1)	4.4 (3.9 – 4.7)	<0.001
Ile (mmol/l)	5.0 (4.0 – 6.0)	5.0 (5.0 – 6.0)	0.0008
Leu (mmol/l)	7.0 (6.0 – 8.0)	8.0 (6.0 – 8.0)	0.0039
Tyr (mmol/l)	5.0 (4.0 – 6.0)	5.0 (5.0 – 6.0)	0.0116

Valores de metabolitos expresados en Md (IIC). Se utilizó U Mann Whitney para la diferencia de Md.

11.2 Modelos de predicción

Con el fin de desarrollar el mejor modelo de predicción de riesgo de diabetes incidente para población mexicana, se generaron diversos modelos de regresión incluyendo distintos componentes a los modelos: solo los metabolitos, solo los factores clínicos (parámetros clínicos y bioquímicos) y la combinación de ambos.

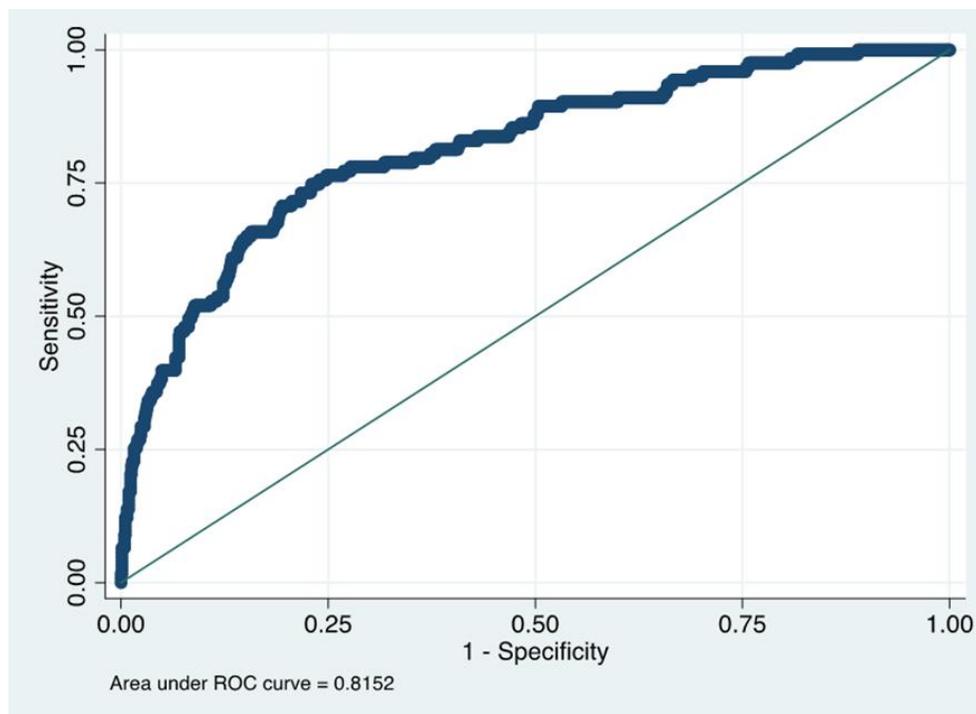
Para la selección de variables específicas, se realizó una regresión por pasos hacia atrás (backward stepwise regression) incluyendo a todos los metabolitos. Tomando en cuenta dicha selección de metabolitos, más los metabolitos ya descritos previamente que tenían asociación con diabetes incidente, se crearon modelos usando regresión logística. El primer modelo incluyó solo a los metabolitos, el cual incluyó 8 metabolitos y se obtuvo un área bajo la curva de 0.815 (tabla 3, figura 1). El mejor modelo de solo metabolitos incluyó a 2 aminoácidos ramificados, 3 lipoproteínas de alta densidad, los ésteres de colesterol, la glucosa y el ácido linoleico.

Tabla 3. Modelo multivariante con regresión logística solo metabolitos en grupo selecto de casos y de controles

Variable Independiente	OR	E.T.	Wald	sig	IC 95%	AUC
sexo	0.723	0.199	-1.17	0.241	0.421 – 1.241	0.815
edad	0.972	0.010	-2.48	0.013	0.951 – 0.994	
leu	1.143	0.043	3.54	<0.001	1.061 – 1.231	
ile	1.397	0.175	2.67	0.008	1.092 – 1.786	
estc	1.947	0.475	2.73	0.006	1.205 – 3.143	
glc	10.379	2.486	9.77	<0.001	6.489 – 16.600	
xl-hdl-tg	1.24e+36	3.84e+37	2.68	0.007	5.01e-09 – 3.06e+62	
hdl-c	0.434	0.110	-3.28	0.001	0.264 – 0.715	
s-hdl-tg	5.53e-19	8.84e-18	-2.63	0.009	1.38e-32 – 0.00002	
la	0.436	0.102	-3.52	<0.001	0.274 – 0.692	
cons	1.41e-11	9.94e-11	-3.55	<0.001	1.42e-17 – 0.00001	

OR = odds ratio, ET = error standard, z = prueba de Wald, sig = significancia estadística, IC = intervalo de confianza, AUC = área bajo la curva, cons = constante. leu = leucina, ile = isoleucina, estc = colesterol esterificado, glc = glucosa, xl-hdl-tg = lipoproteína de alta densidad tamaño extragrande componente triglicéridos, hdl-c = lipoproteína de alta densidad, s-hdl-tg = lipoproteína de alta densidad de tamaño pequeño componente triglicéridos, la = ácido linoleico.

Figura 1. Curva ROC del modelo solo metabolitos en grupo selecto de casos y de controles



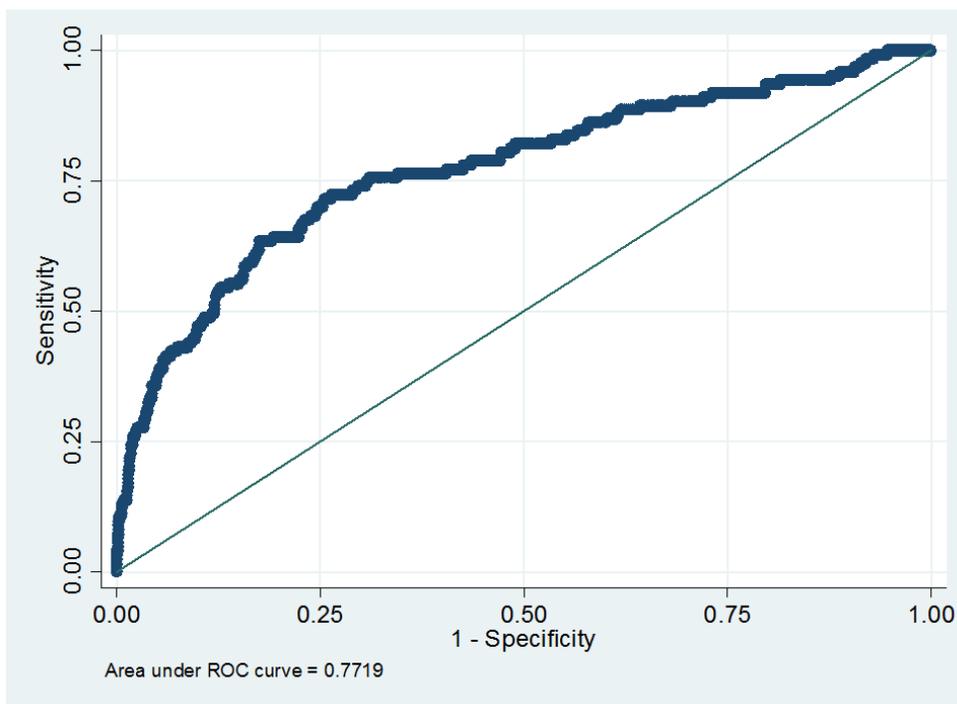
El segundo modelo multivariante con regresión logística incluyó solo a los factores clínicos. Se incluyeron 3 factores clínicos: el sexo, la edad y la glucosa y se obtuvo un área bajo la curva de 0.771 (tabla 4, figura 2).

Tabla 4. Modelo multivariante con regresión logística solo factores clínicos en grupo selecto de casos y de controles

Variable Independiente	OR	E.T.	Wald	sig	IC 95%	AUC
sexo	0.846	0.196	-1.72	0.474	0.536 – 1.335	0.771
edad	0.975	0.010	-2.36	0.018	0.955 – 0.995	
glucosa	1.109	0.011	10.44	<0.001	1.088 – 1.131	
cons	0.00002	0.00002	-10.89	<0.001	4.02e-06 – 0.0001	

OR = odds ratio, ES = error standard, z = prueba de Wald, Sig = significancia estadística, IC = intervalo de confianza, AUC = área bajo la curva

Figura 2. Curva ROC del modelo solo factores clínicos en grupo selecto de casos y de controles



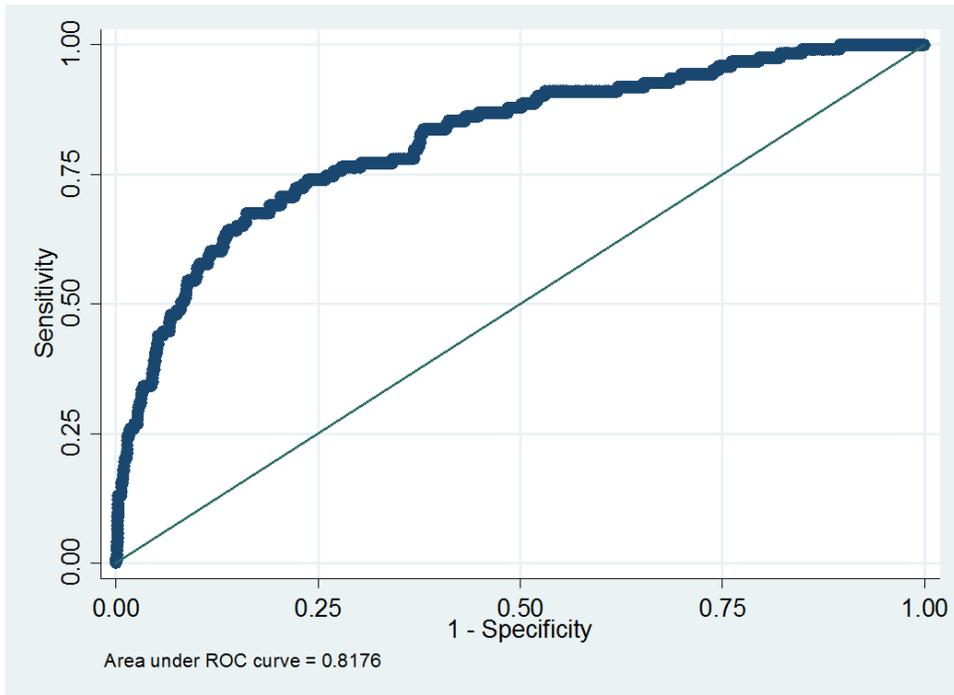
Por último se creó un modelo multivariante con regresión logística completo el cual incluyó tanto a los metabolitos como a los factores clínicos. Incluyó 11 variables, 3 de ellas son factores clínicos y 8 variables son metabolitos. Se obtuvo un área bajo la curva de 0.817 (tabla 5, figura 3).

Tabla 5. Modelo multivariante completo con regresión logística en grupo selecto de casos y de controles

Variable Independiente	OR	E.T.	Wald	sig	IC 95%	AUC
sexo	0.743	0.206	-1.07	0.287	0.431 – 1.282	0.817
edad	0.976	0.010	-2.08	0.038	0.955 – 0.998	
glucosa	1.112	0.011	9.91	<0.001	1.089 – 1.136	
leu	1.134	0.043	3.30	0.001	1.052 – 1.222	
ile	1.411	0.174	2.78	0.005	1.107 – 1.799	
estc	1.854	0.457	2.51	0.012	1.144 – 3.007	
FAlen	2.200	0.865	2.01	0.045	1.018 – 4.756	
xl-hdl-tg	1.47e+36	4.52e+37	2.71	0.007	1.03e+10 – 2.11e+62	
hdl-c	0.457	0.116	-3.06	0.002	0.277 – 0.755	
s-hdl-tg	2.20e-18	3.52e-17	-2.55	0.011	5.65e-32 – 0.00008	
la	0.464	0.109	-3.25	0.001	0.292 – 0.738	
cons	1.69e-11	1.17e-10	-3.59	<0.001	2.23e-17 – 0.00001	

OR = odds ratio, ES = error standard, z = prueba de Wald, sig = significancia estadística, IC 95% = intervalo de confianza, AUC = área bajo la curva, leu = leucina, ile = isoleucina, estc = colesterol esterificado, FAlen = largo o longitud de los ácidos grasos, xl-hdl-tg = lipoproteína de alta densidad de tamaño extra grande en su medición de triglicéridos, la = ácido linoleico.

Figura 3. Curva ROC en modelo completo en grupo selecto de casos y de controles



El OR de cada covariable en el modelo de solo metabolitos no fue atenuado en el modelo completo al adicionar los factores clínicos.

Se analizó el grupo de los casos y de los controles dividiéndolos por edad: menor a 50 años y mayor o igual a 50 años y por índice de masa corporal: IMC menor a 30 kg/m² y mayor o igual a 30 kg/m². Se encontró que el modelo completo predice mejor para los sujetos con edad mayor o igual a 50 años (AUC 0.814 vs 0.857) y para los sujetos con un IMC mayor o igual a 30 kg/m² (AUC 0.826 vs 0.854).

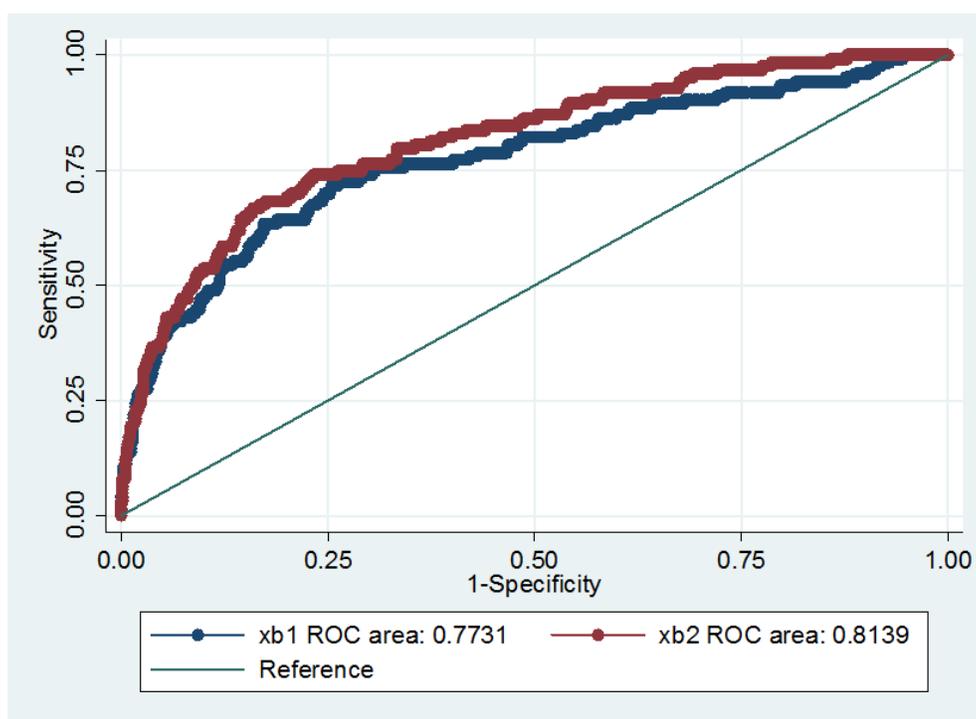
Se realizó la comparación del área bajo la curva del modelo de sólo metabolitos (AUC = 0.813) con el modelo clínico (AUC = 0.773) y se encontró una diferencia estadística significativa ($p = 0.0092$).

Tabla 6. Comparación estadística de dos áreas bajo la curva de los modelos de sólo metabolitos y clínico

Modelo	AUC	Std. Err.	IC 95%	p
Metabolitos	0.813	0.0214	0.772 – 0.855	0.0092
Clínico	0.773	0.0253	0.723 – 0.822	

Se utilizó la prueba de χ^2 para comparar las dos áreas bajo la curva. AUC = área bajo la curva, Std. Err. = error standard, IC 95% = intervalo de confianza.

Figura 4. Comparación de dos áreas bajo la curva del modelo de sólo metabolitos y del modelo clínico



12. DISCUSION

Este es el primer estudio en medir metabolitos a través de la metabolómica en población mexicana. Es un estudio novedoso, al crear modelos de predicción de diabetes incidente que incluye a los parámetros de la metabolómica. Es uno de los primeros estudios en considerar múltiples metabolitos relacionados con los lípidos y las lipoproteínas. Todos los metabolitos y los factores clínicos fueron medidos de manera basal y el tiempo de desarrollo de diabetes incidente o el tiempo libre de diabetes, se derivó de la misma evaluación y seguimiento. El método de área bajo la curva se utilizó para evaluar la capacidad de predicción de los modelos.

Los resultados muestran la capacidad de los metabolitos de predecir el riesgo de la diabetes tipo 2. Los metabolitos aumentan la capacidad de predicción al adicionarlos a los factores clínicos convencionales, de una discriminación regular a una discriminación buena. La capacidad de predicción del modelo completo en sujetos menores de 50 años y sujetos con un IMC <30 kg/m² fue menor que en aquellos sujetos mayores o iguales de 50 años o con un IMC mayor o igual a 30 kg/m². En el estudio de Walford se encontró que el modelo completo predice mejor en aquellos sujetos menores de 50 años y sin obesidad en población caucásica (17).

La mayor área bajo la curva se obtuvo en el modelo completo el cual tomó en cuenta a los metabolitos y a los factores clínicos. Sin embargo, el área bajo la curva obtenida en dicho modelo (AUC = 0.817) fue menor a la encontrada en el estudio de Walford et al. (AUC = 0.874) (17) y similar a la encontrada en población amerindia (AUC = 0.822, p = 0.006) (21). Lo anterior podría explicarse por la menor área bajo la curva en el modelo de solo factores clínicos de nuestra cohorte. El modelo de solo metabolitos alcanzó un área bajo la curva similar (AUC = 0.815) a lo encontrado en el estudio de Walford (AUC = 0.803) (17). El modelo de solo factores clínicos alcanzó un área bajo la curva menor (AUC = 0.771) a lo obtenida en la cohorte de Framingham

Offspring Study (FOS) (0.856) (17), sin embargo dicha área bajo la curva es similar a la obtenida en el estudio AUSDRISK de Australia y Cambridge Risk Score del Reino Unido (AUC 0.78 y 0.74 respectivamente) (18, 19). Pareciera que el modelo de metabolitos predice mejor que el modelo de factores clínicos en población mexicana.

Los aminoácidos ramificados, leucina e isoleucina y el aminoácido aromático tirosina, se asocian a diabetes incidente en nuestra cohorte. Dichos metabolitos eran de particular interés, por ser metabolitos previamente descritos relacionados a diabetes incidente. Además de que dichos aminoácidos se han encontrado elevados hasta 12 años antes de la aparición de diabetes manifiesta en población caucásica (11). Los aminoácidos ramificados y aromáticos pudieran ser tanto marcadores como efectores de la resistencia a la insulina. Los aminoácidos circulantes pudieran promover la resistencia a la insulina, posiblemente generando una disrupción en las vías de señalización de la insulina en el músculo esquelético con la activación de mTOR, JUN e IRS-1 (28). La secreción anormal de insulina, es el otro mecanismo en la patogenia de la diabetes tipo 2. Múltiples aminoácidos, en particular los ramificados, son moduladores en la secreción de la insulina. Así, la hiperaminoacidemia pudiera promover la diabetes al generar hiperinsulinemia llevando a falla en la célula beta (27).

En nuestro estudio encontramos varios metabolitos del grupo de los ácidos grasos con asociación con la diabetes incidente como fueron la longitud de los ácidos grasos, la relación del ácido linoleico con los ácidos grasos totales (FA), la relación del omega 6 con los FA, la relación de los ácidos grasos poliinsaturados con los FA y la relación de los ácidos grasos monoinsaturados con los FA. Los ácidos grasos también tienen efectos importantes en la sensibilidad y la secreción de la insulina, los dos mecanismos que están alterados en la diabetes tipo 2. La insulina regula los niveles de glicerol y de ácidos grasos en el suero al inhibir a la lipólisis. En los estados de

resistencia a la insulina, la lipólisis incrementada lleva a la sobreproducción de glicerol y de ácidos grasos libres a partir de los triglicéridos (24). Cuando existe resistencia de insulina en el músculo esquelético, el exceso de glucosa llega al hígado resultando en lipogénesis de novo e hiperlipidemia. La infiltración de macrófagos en el tejido adiposo, lleva a un incremento en la lipólisis, lo que genera una síntesis incrementada de triglicéridos e hiperlipidemia por esterificación incrementada de ácidos grasos. La lipólisis también estimula la gluconeogénesis hepática, promoviendo la hiperglucemia a partir de los ácidos grasos hasta su conversión en glucosa (26).

En población caucásica los metabolitos asociados con diabetes incidente fueron 3 aminoácidos: la isoleucina, la tirosina y la fenilalanina y 6 metabolitos de lípidos: la lisofosfatidilcolina LPC C18:2, la fosfatidilcolina PC C38:6, los triacilglicerolos: C44:1, C48:0, C52:1 y C56:9 TAG (17). En nuestra cohorte también se encontró asociación de la isoleucina y la tirosina con la diabetes. En población amerindia se encontraron 7 metabolitos que predicen el riesgo de diabetes de manera significativa. Dos metabolitos el 2-hidroxi-bifenil y un químico X-1178 se asociaron con un riesgo incrementado de diabetes, mientras que cinco metabolitos, uno relacionado a la fosfatidilcolina (PC 22:6/20:4), un flavonoide (3S)-7hidroxi-2',3',4',5',8-pentametoxiflavan, dos tetrapéptidos (MEIR, LDYR) y un metabolito desconocido X-490, se asociaron de manera significativa con un riesgo disminuido de diabetes (21). En población asiática con obesidad y diabetes se encontró mayor concentración de ácidos grasos saturados, algunos aminoácidos incluidos la leucina, la isoleucina, la fenilalanina, el ácido láctico, el ácido 3 hidroxibutírico, la colina, el ácido 3,7-dimetilúrico, el ácido pantoténico, el mioinositol, el sorbitol, el glicerol y la glucosa (22). En la cohorte de mexicanos se encontró asociación de los aminoácidos, los ácidos grasos y la glucosa con la diabetes. Por otro lado, en población finlandesa se encontró que el glicerol, los ácidos grasos: libres, monoinsaturados, saturados, n-7, n-9 son biomarcadores de un riesgo elevado de

diabetes tipo 2, mientras que la elevación del ácido graso n-6 (principalmente ácido linoleico) se asoció a un riesgo disminuido de diabetes tipo 2 (24). En nuestra cohorte también se asoció al ácido linoleico con menor riesgo de diabetes. Por último, en el estudio de población alemana se encontró a la hexosa, la fenilalanina y las diacil-fosfatidilcolinas C32:1, C36:1, C38:3 y C40:5 con asociación positiva significativa para el riesgo de diabetes. En este mismo estudio se encontró a la glicina, la esfingomielina C16:1, la lisofosfatidilcolina C18:2 y las acil-alquil-fosfatidilcolinas C34:3, C40:6, C42:5, C44:4 y C:45:5 estaban inversamente relacionadas con el riesgo de diabetes (25).

Las diferencias en los metabolitos encontrados en distintas poblaciones podrían explicarse por las distintas metodologías utilizadas para la medición de los metabolitos. La espectroscopía acoplada a RMN es útil para la medición de muestras de cohortes grandes, la preparación de la muestra es mínima y la muestra una vez usada puede preservarse para análisis subsecuentes, mientras que la espectroscopía de masas acoplada a cromatografía líquida es una técnica más sensible (9). Por otro lado, algunos estudios exploran a los metabolitos de manera no dirigida (untargeted) con el objetivo de descubrir nuevos biomarcadores, a diferencia de otros estudios quienes usan a la metabolómica con un panel con un blanco dirigido (targeted) lo que también explica los diferentes hallazgos entre estudios. Por último, el estudiar distintas poblaciones también es un factor que contribuye a los distintos hallazgos encontrados.

En nuestro modelo multivariante con regresión logística el cual incluyó solo metabolitos, los metabolitos con mayor OR fueron la glucosa (OR = 10.37, $p < 0.001$) y la lipoproteína XL-HDL-TG (OR = 1.24×10^{36} , $p = 0.007$). Otros de los metabolitos asociados a menor probabilidad de diabetes incidente fueron los triglicéridos en las HDL pequeñas (S HDL TG OR = 5.53×10^{-19} , $p = 0.009$) y el colesterol HDL (OR = 0.434, $p = 0.001$). Esto podría estar en relación con la fisiopatogenia de la dislipidemia en la diabetes tipo 2 (20). El ácido linoleico, también lo encontramos como metabolito

relevante en la predicción de la diabetes tipo 2 en mexicanos (OR = 0.436, $p = <0.001$), el cual parece asociarse con menor probabilidad de tener diabetes, similar a lo encontrado en el estudio de población finlandesa (24). Por último los ésteres de colesterol, fue otro de los metabolitos relacionado a diabetes incidente (OR = 1.94, $p = 0.006$). No sabemos si los triacilgliceroles descritos en la cohorte de FOS también son metabolitos predictores de diabetes en mexicanos, ya que no fueron medidos en el panel de metabolómica (12, 17). La glucemia de ayuno fue la variable con mayor capacidad de predicción en los modelos clínicos en población mexicana.

13. FORTALEZAS Y DEBILIDADES DEL ESTUDIO

Dentro de las fortalezas, podemos señalar que se trata del primer estudio en mexicanos en el cual se crea un modelo clínico y otro modelo de metabolómica de predicción de diabetes incidente. Otra fortaleza es el tamaño de muestra y el seguimiento de los pacientes de la cohorte. Por último, el haber obtenido la medición basal de todas las variables en una misma cohorte y el haber obtenido información longitudinal.

Las debilidades del estudio fue no contar con la curva de tolerancia oral a la glucosa (CTOG) o la hemoglobina glucosilada (HbA1c) de todos los pacientes para definir a los pacientes diabéticos. No contar con la medición de todos los metabolitos en todos los pacientes de la cohorte para obtener una muestra más grande. Por último, el no contar con la medición de algunos de los metabolitos que predicen diabetes incidente, ya descritos en otras poblaciones, debido a que se usó un panel de metabolómica específico (targeted).

14. CONCLUSION

Las mediciones específicas de la plataforma de metabolómica mejoran de manera modesta la predicción de diabetes tipo 2 al compararlo con los modelos de predicción previos los cuales incluyen solo a los factores clínicos. Se necesitan estudios a futuro para probar los modelos de predicción y cómo esta información puede ayudar a la prevención de diabetes tipo 2 en aquellos sujetos con más alto riesgo metabólico.

15. BIBLIOGRAFIA

1. Whiting DR, et al. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract* 2011; 94: 311 – 321.

2. Jiménez-Corona A. et al. Diabetes mellitus tipo 2 y frecuencia de acciones para su prevención y control. *Salud Pública de México* 2013; 55(2): s137 – s143.

3. Knowler WC, et al. Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002; 346: 393 – 403.

4. Wilson PW, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007; 167: 1068 – 1074.

5. Tabák AG, et al. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *Lancet* 2009; 373: 2215 – 2221.

6. Soininen P et al. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circ Cardiovasc Genet* 2015; 8: 192 – 206.

7. Vassy JL, et al. Polygenic Type 2 Diabetes Prediction at the Limit of Common Variant Detection. *Diabetes* 2014; 63: 2172 – 2182.
8. Larive CK, et al. NMR Spectroscopy for Metabolomics and Metabolic Profiling. *Anal Chem* [dx.doi.org/10.1021/ac504075g](https://doi.org/10.1021/ac504075g)/*Anal. Chem.* XXXX, XXX, XXX-XXX
9. Brennan L. NMR-based metabolomics: From sample preparation to applications in nutrition research. *Progress in Nuclear Magnetic Resonance Spectroscopy* 2014; 83: 42 – 49.
10. Kettunen et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature Genetics* 2011; 44(3): 269-276.
11. Wang TJ, et al. Metabolite profiles and the risk of developing diabetes. *Nature Medicine* 2011; 17(4): 448 – 453.
12. Rhee EP, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* 2011; 121(4): 1402 – 1411.
13. Cheng S, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation* 2012; 125: 2222 – 2231.
14. Wang TJ, et al. 2-Aminoadipic acid is a biomarker for diabetes risk. *J Clin Invest* 2013; 123(10): 4309 – 4317.
15. Zhao, X. et al. Changes of the plasma metabolome during an oral glucose tolerance test: is there more than glucose to look at? *Am J Physiol Endocrinol Metab* 2009; 296: E384 – E393.
16. Menni C, et al. Biomarkers for Type 2 Diabetes and Impaired Fasting Glucose Using a Nontargeted Metabolomics Approach. *Diabetes* 2013; 62: 4270 – 4276.

17. Walford GA, et al. Metabolite Traits and Genetic Risk Provide Complementary Information for the Prediction of Future Type 2 Diabetes. *Diabetes Care* 2014; 37: 2508 – 2514.
18. Chen L, et al. AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. *Med J Aust* 2010; 192(4): 197 – 202.
19. Rahman M, et al. A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study. *Family Practice* 2008; 25: 191 – 196.
20. Aguilar-Salinas CA, et al. Management of the metabolic syndrome as a strategy for preventing the macrovascular complications of type 2 diabetes: controversial issues. *Current Diabetes Reviews*. 2005; 1: 145 – 158.
21. Jinying Zhao, et al. Novel Metabolic Markers for the Risk of Diabetes Development in American Indians. *Diabetes Care* 2015; 38: 220 – 227.
22. Navdeep Gogna, et al. Investigating correlations in the altered metabolic profiles of obese and diabetic subjects in a South Indian Asian population using an NMR-based metabolomics approach. *Mol. BioSyst.* 2015; 11: 595 – 606.
23. Mahendran Y, et al. Association of ketone body levels with hyperglycemia and type 2 diabetes in 9398 Finnish men. *Diabetes* 2013; 62: 3618 – 3626.
24. Mahendran Y, et al. Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. *Diabetes Care* 2013; 36: 3732 – 3738.
25. Floegel A, et al. Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolic Approach. *Diabetes* 2013; 62: 639 – 648.

26.Samuel VT, et al. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. *J Clin Invest* 2016; 126(1):12 – 22.

27.Floyd JC, et al. Stimulation of insulin secretion by amino acids. *J Clin Invest* 1966; 45: 1487 – 1502.

28.Newgard CB et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2009; 9: 311 – 326.

29.Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research* 2011; 46: 399 – 424.