Development and validation of a multivariable genotype-informed gestational diabetes prediction algorithm for clinical use in the Mexican population: insights into susceptibility mechanisms

Mirella Zulueta,[1] Héctor Gallardo-Rincón ![ORCID] ,[2,3] Luis Alberto Martinez-Juarez,[3] Julieta Lomelin-Gascon,[3] Janinne Ortega-Montiel,[3] Alejandra Montoya,[3] Leire Mendizabal,[1] Maddi Arregi,[1] María de los Angeles Martinez-Martinez,[4] Eneida del Socorro Camarillo Romero,[4] Hugo Mendieta Zerón,[4] José de Jesús Garduño García,[4] Laureano Simón,[1] Roberto Tapia-Conyer[5]

Correspondence to
Dr Héctor Gallardo-Rincón;
hgallardo@
fundacioncarlosslim.org

ABSTRACT

Introduction Gestational diabetes mellitus (GDM) is underdiagnosed in Mexico. Early GDM risk stratification through prediction modeling is expected to improve preventative care. We developed a GDM risk assessment model that integrates both genetic and clinical variables.

Research design and methods Data from pregnant Mexican women enrolled in the 'Cuido mi Embarazo' (CME) cohort were used for development (107 cases, 469 controls) and data from the 'Mónica Pretelini Sáenz' Maternal Perinatal Hospital (HMPMPS) cohort were used for external validation (32 cases, 199 controls). A 2-hour oral glucose tolerance test (OGTT) with 75 g glucose performed at 24–28 gestational weeks was used to diagnose GDM. A total of 114 single-nucleotide polymorphisms (SNPs) with reported predictive power were selected for evaluation. Blood samples collected during the OGTT were used for SNP analysis. The CME cohort was randomly divided into training (70% of the cohort) and testing datasets (30% of the cohort). The training dataset was divided into 10 groups, 9 to build the predictive model and 1 for validation. The model was further validated using the testing dataset and the HMPMPS cohort.

Results Nineteen attributes (14 SNPs and 5 clinical variables) were significantly associated with the outcome; 11 SNPs and 4 clinical variables were included in the GDM prediction regression model and applied to the training dataset. The algorithm was highly predictive, with an area under the curve (AUC) of 0.7507, 79% sensitivity, and 71% specificity and adequately powered to discriminate between cases and controls. On further validation, the training dataset and HMPMPS cohort had AUCs of 0.8256 and 0.8001, respectively.

Conclusions We developed a predictive model using both genetic and clinical factors to identify Mexican women at risk of developing GDM. These findings may contribute to a greater understanding of metabolic functions that underlie elevated GDM risk and support personalized patient recommendations.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Women of Hispanic ancestry have not been adequately represented in datasets investigating genetic associations with gestational diabetes mellitus. Given the high incidence of gestational diabetes mellitus in Mexico, it is important to have risk assessment tools to identify Mexican women at high risk. As such, the integration of both genetic and clinical variables may help improve predictability.

WHAT THIS STUDY ADDS

⇒ This study provides insight into genetic risk factors associated with the development of gestational diabetes mellitus within the Mexican population. We developed a predictive model for gestational diabetes mellitus in Mexican women that integrated genotypic and phenotypic traits. These findings contribute to the understanding of the potential metabolic functions underlying elevated risk and can support further research in this area.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The development of predictive models that incorporate both genetic and clinical factors can potentially support a movement towards personalized recommendations and treatment for individual patients. The use of genomic intelligence-based tools in clinical practice will contribute to advancing diabetes precision medicine. Consequently, patients, clinical practice, and healthcare systems are expected to benefit.

INTRODUCTION

Gestational diabetes mellitus (GDM), defined as hyperglycemia with onset or first recognition during pregnancy, is associated with an increased risk of pregnancy complications

and adverse perinatal outcomes, including pre-eclampsia, stillbirth, large for gestational age, neonatal hypoglycemia, preterm birth, low Apgar scores, and admission to neonatal intensive care.[1–3] Fetal exposure to diabetes in utero has been linked to macrosomia and adiposity in newborns and impaired glucose tolerance and obesity in childhood, thereby increasing risks for adverse cardiometabolic outcomes later in life.[4 5] While hyperglycemia commonly resolves post partum, GDM can reoccur and is often associated with a subsequent diagnosis of type 2 diabetes (T2D) and coronary heart disease.[2 3] Although the global prevalence of GDM is increasing at a concerning rate,[6] it varies according to population characteristics (eg, maternal age, ancestry, and obesity rates) and the criteria used for screening and diagnosis.[7] In Mexico, the estimated prevalence of GDM in 2021 was 11.2%.[8] Unfortunately, GDM is detected in only about 1% of cases in Mexico, and glucometers and glucose strips are generally not available for glucose self-monitoring.

Early risk stratification by prediction modeling might offer opportunities to improve care for those women at high risk of developing GDM. As timely intervention is key to preventing adverse outcomes in GDM,[9] clinicians need simple prediction models that can be used in the first trimester of pregnancy. Clinical multivariate GDM risk prediction models have been proposed.[10–12] However, these novel measures of biochemical and clinical markers have not been thoroughly examined and the equations are complex, making these prediction models difficult to use in routine clinical practice.

In contrast to T2D, there are relatively few published studies on the genetic susceptibility to GDM, and despite the high incidence in Mexico, studies on the genetic architecture of GDM in the Mexican population are lacking. To our knowledge, only a single study by Huerta-Chagoya et al[13] has provided insight into the genetic factors of GDM in Mexican women, confirming that T2D and GDM share a common genetic background and suggesting that other genetic mechanisms may be in play for GDM. Before that, Watanabe et al identified the association between variants in *TCF7L2* and GDM in a small study of 152 women of Mexican American origin.[14]

A meta-analysis by Lowe et al[15] that included the original genome-wide association study (GWAS) of glycemic traits in pregnancy in the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) Study, multiple ethnic groups from the HAPO Study, and two other pregnancy cohorts[16 17] has significantly contributed to the identification of genetic variants associated with GDM. While these studies were significant and their scope was expansive, women of Hispanic ancestry represented <10% of the participants. More recently, within the GENetics of Diabetes In Pregnancy Consortium, Pervjakova et al conducted the largest and most ancestrally diverse GWAS meta-analysis for GDM that included a total of 5485 women with GDM and 347 856 without; of those, 2.8% were of Hispanic/Latino origin.[18] Powe et al[19–21]

have also provided insight into the genetic heterogeneity among women with GDM. They examined polygenic scores for T2D, fasting glucose, fasting insulin, insulin secretion, and insulin resistance among women with different physiologic subtypes of GDM. Their genotype-based approach to heterogeneity in GDM suggested that genetic data provide both information on GDM risk and distinct genetic information pointing to phenotypic data.

The Nurses' Health Study II and the Danish National Birth Cohort were used by Ding et al[22] to identify eight novel genetic variants and create a genetic risk score based on those variants. Furthermore, in our genetic variant analysis of the Europe-wide vitamin D and lifestyle intervention randomized controlled trial, an interaction between *MTNR1B* variants and lifestyle intervention in regard to maternal and neonatal outcomes was identified.[23]

In the study presented here, our aim was to develop and validate a risk assessment model to identify Mexican women at high risk of GDM, through an algorithm that integrates genetic and clinical variables. In this study, we used data from the 'Cuido mi Embarazo' (CME) cohort, which used an oral glucose tolerance test (OGTT) and was supplied glucose screening through MIDO Embarazo, a module of the Integrated Monitoring for Early Detection system (Medición Integrada para la Detección Oportuna, MIDO).

METHODS
Study populations
Development cohort: CME cohort
This population included 576 Mexican women from the prospective multicenter study CME (research registry no. 7405). Participants were recruited between May 8, 2019 and May 18, 2021, from six participating study sites in Mexico: three primary healthcare facilities in Hidalgo, two in Guanajuato, and one in Mexico City. The CME cohort included pregnant women without T2D who were <28 gestational weeks and had performed a 2-hour 75 g OGTT between gestational weeks 24 and 28. Those diagnosed with pre-gestational diabetes, had multiple pregnancies, or had a previous chronic disease that required their pregnancy to be monitored by secondary care were excluded. All participants received primary antenatal care (standardized prenatal care was offered) and fasting plasma and capillary glucose measurements were performed at any time during the first 28 weeks of gestation.

Information on maternal age, ethnicity, gestational week at the time of the OGTT, body mass index (BMI), family history of T2D, medical history of GDM, obstetric history and parity, gestational weight gain, associated comorbidities, and newborn birth weight was collected. Baseline characteristics of this cohort are described in table 1 and in Martinez-Juarez et al.[8]

Table 1 Baseline characteristics of study participants

Characteristics	Cases	Controls	P value
CME cohort			
Number of participants	107	469	
Age at baseline (years)	28.64±6.56	26.06±5.98	<0.001*
Pre-pregnancy BMI (kg/m^2)	28.05±4.76	25.52±4.88	<0.001*
Previous GDM (yes)	16 (14.95)	2 (0.43)	<0.001†
Family history of T2D (yes)	73 (68.22)	217 (46.27)	<0.001‡
Previous pregnancies (yes)	72 (67.29)	266 (56.72)	0.058‡
HMPMPS cohort			
Number of participants	32	199	
Age at baseline (years)	29.13±6.61	25.60±6.72	0.008*
Pre-pregnancy BMI (kg/m^2)	30.67±6.10	25.00±4.54	<0.001*
Previous GDM (yes)	0	0	–
Family history of T2D (yes)	25 (78.13)	113 (66.83)	0.037‡
Previous pregnancies (yes)	27 (84.38)	126 (63.32)	0.033*

Data are presented as n (%) or mean±SD unless otherwise indicated.
*P values obtained using Student's t-test.
†P values obtained using Fisher's exact test.
‡P values obtained using the X^2 test.
BMI, body mass index; CME, Cuido mi Embarazo; GDM, gestational diabetes mellitus; HMPMPS, 'Mónica Pretelini Sáenz' Maternal Perinatal Hospital; T2D, type 2 diabetes.

External validation cohort: 'Mónica Pretelini Sáenz' Maternal Perinatal Hospital cohort

The 'Mónica Pretelini Sáenz' Maternal Perinatal Hospital (HMPMPS) cohort included 231 women who were recruited by 'Mónica Pretelini Sáenz' Maternal Perinatal Hospital, Toluca, State of Mexico. Participants were recruited between February 1 and September 30, 2018, and the eligibility criteria for the HMPMPS cohort were the same as that of the CME cohort. This study is registered in ClinicalTrials.gov (ID: NCT01649167). As described in online supplemental material 1, these two study cohorts were used to develop and validate an algorithm that could predict the risk of GDM in Mexican women during the early stages of pregnancy or before pregnancy.

Diagnosis of gestational diabetes

GDM was diagnosed as per the International Association of Diabetes and Pregnancy Study Groups criteria.[1] A 2-hour OGTT with 75 g glucose was performed at 24–28 weeks of gestation. Plasma glucose levels were determined by the glucose oxidase method using fresh plasma samples. A diagnosis of GDM was confirmed if glucose levels were abnormal at one of the three time points (fasting, or 1 hour or 2 hours post-glucose ingestion). The assessment of the GDM outcome was blinded to predictors. Similarly, investigators were blinded to GDM assessment during predictor assessment.

Clinical variables selection

Maternal age, pre-gestational BMI, family history of T2D and previous pregnancies were chosen as the clinical parameters for the model as these parameters have been described as strong risk factors and predictors in the development of GDM.[24–27] Although other clinical variables such as glucose measurement have also been studied as predictors for the development of GDM,[24] we selected variables for which information could be easily obtained from the initial antenatal care questionnaire and that did not require further laboratory testing or specialized trained personnel.

Single-nucleotide polymorphism selection

One hundred and fourteen single-nucleotide polymorphisms (SNPs) were selected based on their predictive power as reported in previously published studies.[13 16 17 19 22 28–35] Specifically, SNPs were prioritized according to the results of a large meta-analysis of GWAS, with the assumption that their effects can be extrapolated and generalized and that large sample sizes allow solid estimations of the true size effect. In addition, significant SNPs that were identified in smaller association studies were also included. The 114 selected SNPs are listed in online supplemental material 2.

SNP genotyping

Genomic DNA was extracted from EDTA-stabilized blood samples taken during the OGTT using the Maxwell RSC instrument (Promega, Dubendorf, Switzerland). Genotyping was performed by iPLEX MassARRAY PCR using the Agena platform (Agena Bioscience, San Diego, California, USA). iPLEX MassARRAY PCR and extension primers were designed from sequences containing each target SNP and 150 upstream and downstream bases with AssayDesign Suite software (http://agenabio.com/assay-design-suite-20-software) using the default settings. Single-base extension reactions were performed on the PCR reactions with the iPLEX Gold Kit (Agena Bioscience) and 0.8 µL of the custom unique variant pool. PCR reactions were dispensed onto SpectroChipArrays with a Nano dispenser (Agena Bioscience). An Agena Bioscience Compact MassArray Spectrometer was used to perform matrix-assisted laser desorption/ionization-time of flight mass spectrometry according to the iPLEX Gold Application Guide[35] The Typer 4 software package, V.4 (Agena Bioscience) was used to analyze the resulting spectra, and the composition of the target bases was determined from the mass of each extended oligo. These

panels were designed in collaboration with Patia, and genotyping was performed using the Agena platform located at the Epigenetics and Genotyping Laboratory, Central Unit for Research in Medicine, Faculty of Medicine, University of Valencia, Valencia, Spain.

Data quality control

Quality control steps included the exclusion of variables with a high absence rate (>30%) to identify attributes and samples that did not provide sufficient information. The remaining missing data were estimated by the most common values of each attribute. The resulting database consisted of 576 samples and 139 attributes.

Statistical analysis

We performed a correlation analysis with the aim of reducing possible redundancies; the similarity between the variables was analyzed by measuring Pearson's correlation coefficient. The decision was made to consider one predictor for each pair/group of highly correlated variables (>0.90). Therefore, we chose to analyze the attribute with the lowest ratio of missing values at the beginning of the study. Comparisons between control and case samples were conducted using the X^2 and Fisher's exact test for qualitative data and Student's t-test for quantitative data (mean±SD). Sample sizes were not calculated or confirmed prior to modeling.

The entire cohort was randomly divided into a training dataset (70% of the cohort) for algorithm development and a testing dataset (30% of the cohort) for validation. We trained the prediction model for GDM using a 10-fold cross-validation logistic regression. In this analysis, the entire cohort was randomly divided into 10 subgroups; 9 of them were used to build the predictive model and 1 was used to validate it. The model was then further validated using the testing dataset (30% of the cohort) and the HMPMPS cohort. Of note, the HMPMPS cohort (validation cohort) differed from the CME cohort in that it was a single-center study versus a multicenter study.

All statistical and model calculations were performed in Python V.3.6, using the scikit-learn package. To validate the performance of the model, a k-fold cross-validation procedure was used to estimate the mean and SD of the values computed in the loop.

RESULTS

The data quality control process retrieved a total of 107 cases and 469 controls from the CME cohort. Baseline characteristics of study participants are shown in table 1. Mean age and BMI were higher in cases than controls (age: 28.64 vs 26.06 years, p=0.0003; BMI: 28.05 vs 25.52 kg/m^2, p=0.00000451).

We examined 114 SNPs that were previously associated with the risk of T2D, GDM, high BMI, and adverse pregnancy traits associated with GDM (online supplemental material 2). A correlation analysis was performed to identify SNPs providing similar information (online supplemental materials 3 and 4). The SNPs rs560887 and rs563694; rs17085593 and rs6235; rs13266634, rs11558471, and rs3802177; rs10814916, rs7041847, and rs7034200; rs4402960 and rs7651090; rs8050136 and rs1421085; and rs1801282 and rs17036328 were shown to be highly correlated.

Of the 114 SNPs, 105 provided unique information and were used for further analysis. Statistical analysis showed that a total of 19 attributes (5 clinical variables and 14 SNPs) had a significant association (p<0.05) with the outcome (online supplemental material 5). Family history of T2D and personal history of GDM were selected using X^2 test analysis (p=0.0000656 and p=0.0000000000000714, respectively); the other 17 variables were selected using Student's t-test analysis. Data from 70% of the pregnant women in the CME cohort, which included 75 cases and 328 controls with complete SNP genotype and clinical information available, were included in the development set. Fifteen of the 19 attributes that were significantly associated (p<0.05) provided optimal logistic regression performance; of those, 11 were SNPs and 4 were clinical variables (table 2). Of the 11 SNPs selected by the analysis, rs1387153 in LOC100128354/*MTNR1B*, rs4607517 in *GCK*, rs10830963 in *MTNR1B*, rs11715915 in *AMT*, rs340874 in *PROX1*, rs6048205 in *FOXA2*, rs16996148 in *CILP2*, rs2943634 in *IRS1*, rs6742799 in *RBMS1*, and rs2745353 in *RSPO3* correlated with a diagnosis of GDM and rs9379084 in *RREB1* correlated with absence of GDM diagnosis (table 2). The four clinical attributes selected by the analysis were maternal age, pre-gestational BMI, family history of T2D, and previous pregnancies (table 2).

We included the 15 selected attributes in a GDM prediction regression model and applied it to the training dataset. The algorithm showed high predictive ability with an area under the receiver operating curve (AUC) of 0.7507, sensitivity of 79%, and specificity of 71%. The analysis of predictive values is shown in online supplemental material 6. Figure 1 shows violin plots where the number of samples in each risk percentage is represented in terms of density. The algorithm showed adequate power to discriminate between controls and cases, as the area with major density in controls (median, 12.35%) was smaller than that of the cases (median, 31.20%). The prediction model was then internally verified using the validation dataset of the CME cohort and externally validated using the HMPMPS cohort (table 3). The prediction algorithm showed an AUC of 0.8256 in the training dataset and of 0.8001 in the HMPMPS cohort.

Finally, we explored the performance of the risk model including the 11 genetic variables alone, the 4 clinical variables alone, and all 15 variables together. The risk algorithm with only SNPs performed better than the risk algorithm with only clinical factors (table 4), and the robustness of the model increased when all 15 variables were included.

To use this model, clinicians can collect data regarding clinical variables from patient medical histories at the first prenatal examination, and SNP data either through the collection of an epithelial buccal swab sample or

Table 2 Attributes with optimal logistic regression performance using k-fold cross-validation

	Chromosome (GRCh38.p13)	Closest gene	Effect allele	Logistic regression coefficient (β)
Genetic attributes (SNP)				
rs1387153	chr11:92 940 662	*LOC100128354/MTNR1B*	T	0.4307
rs4607517	chr7:44 196 069	*GCK*	A	0.5168
rs10830963	chr11:92 975 544	*MTNR1B*	G	0.5167
rs11715915	chr3:49 417 897	*AMT*	C	0.7845
rs340874	chr1:213 985 913	*PROX1*	C	0.3014
rs6048205	chr20:22 578 963	*FOXA2*	A	0.323
rs9379084	chr6:7 231 610	*RREB1*	A	−0.5717
rs16996148	chr19:19 547 663	*CILP2*	T	0.2868
rs2943634	chr2:226 203 364	*IRS1*	C	0.4524
rs6742799	chr2:160 460 949	*RBMS1*	A	0.7422
rs2745353	chr6:127 131 790	*RSPO3*	T	0.254
Clinical attributes				
Age	–	–	–	0.0333
Pre-gestational BMI	–	–	–	0.1039
Family history of T2D	–	–	–	0.4811
Previous pregnancies	–	–	–	0.0483

BMI, body mass index; SNP, single-nucleotide polymorphism; T2D, type 2 diabetes.

peripheral blood followed by DNA genotyping. Once the data are entered, the model can be used to determine an individual's risk of developing GDM.

DISCUSSION

We have developed a GDM risk model that can be applied during early pregnancy or before pregnancy. AUCs obtained during development were similar to those obtained after development (0.7507 and 0.8256, respectively), supporting the validation of our model. The development of this model is important because early detection of women at high risk of GDM could catalyze timely intervention with the implementation of lifestyle

changes prior to week 20 of pregnancy, or preferably before week 16, when interventions have been shown to be effective.[7 36] The algorithm used in our model includes 11 SNPs and 4 clinical features.

Our study showed that the presence of the G allele at rs10830963 in *MTNR1B*, and the T allele at rs1387153 in *LOC100128354/MTNR1B* are associated with an increased risk of GDM. The association of SNPs in *MTNR1B* with fasting glucose and insulin secretion is well established.[37] Melatonin is the primary hormone secreted by the pineal gland; it regulates sleep, circadian rhythm, and glucose metabolism. MTNR1B is highly expressed in both the placenta and pancreatic islets. Lyssenko *et al* have shown that genetic variants in this melatonin receptor correlate
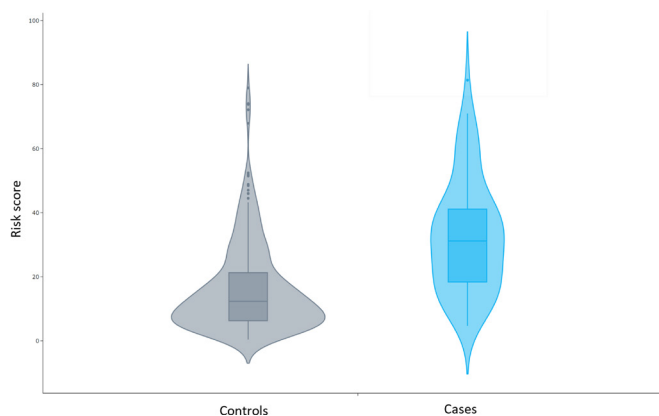

Figure 1 Violin plots of genetic risk scores distribution in cases and controls The distribution of the risk values for the control and case groups is displayed.

Table 3 Performance of GDM prediction algorithm in development and validation cohorts

	Development	Validation	Validation
Cohort	CME 70%	CME 30%	HMPMPS
No of cases	75	32	32
No of controls	328	141	199
Location	Mexico	Mexico	Mexico
Diagnostic criteria	IADPSG	IADPSG	IADPSG
AUC	0.7507	0.8256	0.8001

AUC, area under the curve; CME, Cuido mi Embarazo; GDM, gestational diabetes mellitus; HMPMPS, 'Mónica Pretelini Sáenz' Maternal Perinatal Hospital; IADPSG, International Association of Diabetes and Pregnancy Study Groups.

5

Table 4 Performance of GDM risk algorithm in development and validation sets considering only SNPs, only clinical variables, or both

Variables	AUC development	AUC validation
SNPs only	0.7136	0.7694
Clinical variables only	0.6526	0.6824
SNPs+clinical variables	0.7507	0.8256

AUC, area under the curve; GDM, gestational diabetes mellitus; SNPs, single-nucleotide polymorphisms.

with impaired glucose-stimulated insulin secretion.[38] Furthermore, interactions between variants in *MTNR1B*, GDM risk, and physical activity and healthy eating interventions in pregnant women have been proposed.[22] MTNR1B regulates circadian rhythmicity and influences energy metabolism.[37] Furthermore, associations have been found between relative macronutrient intake, higher fasting plasma glucose, short sleep duration (<7 hours), and *MTNR1B* genetic variants.[39] It has been proposed that lower carbohydrate intake and normal sleep duration may ameliorate cardiometabolic abnormalities conferred by common circadian rhythm-related genetic variants.[39] In addition, carriers of the CC genotype tend to respond more favorably to a hypocaloric diet enriched with monounsaturated fats.[40] Thus, recommendations regarding diet, particularly for carbohydrate and fat consumption, and sleep duration should be emphasized to women who are carriers of *MTNR1B* gene variants and at high risk of GDM.

Another variant included in our model was rs11715915 in *AMT*, a gene that encodes aminomethyltransferase, which is a critical component of the glycine cleavage system in mitochondria, where energy production occurs.[41] The breakdown of glycine produces a methyl group, which is added to and used by folate. rs11715915 is located either in the 3' untranslated region or within coding regions of *AMT*, depending on the transcript, and upstream of *RHOA* (ras homolog family member A).[41] RHOA is a signaling molecule that activates Rho kinase, a regulator of insulin transcription that is differentially regulated in T2D and thought to play a role in glucose homeostasis.[42]

Our study also identified variants in genes encoding transcription factors (*FOXA2*, *PROX1*, *RBMS1*, and *RREB1*) that regulate basic processes in the embryonic development of pancreatic beta cells, cell cycle progression in the pancreas, and insulin response in peripheral tissues. *FOXA2* encodes the forkhead box protein A2, a member of the forkhead class of DNA-binding proteins. FOXA2 has been previously identified as a master regulator in pancreatic development and is involved in regulating both the glucose-sensing apparatus and insulin release.[43] In a study by Yu and Zhong, it was shown that the microRNA *miR-141*, a post-transcriptional regulator in the pathophysiology of T2D, may lead to impaired glucose-stimulated insulin secretion and beta cell proliferation by targeting *FOXA2* at the 3' untranslated region; a potential role for the antidiabetic drug pioglitazone in regulating the *miR-141/FOXA2* axis was also identified.[44]

Another variant of interest identified in our study is the C allele at rs340874 in *PROX1* (Prospero homeobox 1), a transcription factor involved in the embryonic development of the pancreas, liver, and nervous system. Carriers of the CC genotype have been previously shown to have higher non-esterified fatty acid levels after a high-fat meal and lower glucose oxidation after a high-carbohydrate meal in comparison with subjects who have other *PROX1* genotypes.[45] Subjects with the CC variant also had higher accumulation of visceral fat and, surprisingly, lower daily food consumption.

Additionally, rs6742799, mapping to *RBMS1* (RNA binding motif, single-stranded interacting protein 1) was found to have a significant association with GDM. *RBMS1* is expressed in the placenta and has a possible anti-inflammatory role. Alvine *et al* proposed that increased expression of placental RBMS1 in obese women may serve as an adaptive response to reduce oxidative stress in a maternal obesogenic environment.[46] Oxidative stress is now recognized as playing an essential role in certain pregnancy-related disorders such as GDM, pre-eclampsia, and intrauterine growth retardation.[47] The maternal obesity associated with metabolic alterations seems to lead to the appearance of an elevated placental oxidative stress, compromising both placental metabolism and antioxidant status.[48]

The A allele at rs9379084 in *RREB1* (Ras-responsive element binding protein 1) was found to have a protective effect in our study. RREB1 is a member of zinc finger transcription factors and functions both as a transcriptional activator and repressor, and its role in target gene regulation may depend on its binding partner and the status of epigenetic modifications.[49] The cell cycle regulator *CDKN2A* increases susceptibility to T2D and is regulated by RREB1. Furthermore, RREB1 also directly promotes the expression of insulin genes.[49]

Our GDM risk algorithm also included genetic variants in genes with a signaling function and association with insulin resistance (*IRS1*, *RSPO3*, *CILP2*). IRS1 is a signaling intermediate downstream of activated cell-surface insulin receptors.[48] *RSPO3* encodes R-Spondin-3, which regulates Wnt and beta-catenin signaling pathways; *RSPO3* gene knockdown results in abnormal adipogenesis, lipid metabolism, and insulin signaling.[49] In addition, *CILP2* encodes cartilage intermediate layer protein 2, a glycoprotein initially identified in collagen. *CILP2* is located in the NCAN-CILP2-PBX4 region, an intergenic region spanning 300 kb associated with serum cholesterol, low-density lipoprotein and triglyceride concentrations, cardiovascular disease, and non-alcoholic fatty liver disease.[50]

The 11 SNPs identified in this analysis are located in genetic loci that have been reported to participate in molecular processes related to fasting glucose (*MTNR1B, GCK, AMT, PROX1,* and *FOXA2*), insulin resistance (*CILP2, IRS1,* and *RBMS1*), insulin secretion (*MTNR1B*), and fasting insulin (*IRS1*). Four of these SNPs have previously been associated with T2D (*LOC100128354/ MTNR1B, PROX1, CILP2,* and *RBMS1*), while two other SNPs have previously been reported in GDM (*LOC100128354/MTNR1B* and *RREB1*). Overall, this initial annotation of potential genetic loci characteristics, as reported in the literature, is just an initial investigation into how genetic variants may contribute to GDM susceptibility.

The GDM risk algorithm described in this study also included four phenotypic variables: maternal age, pregestational BMI, family history of T2D, and previous pregnancies. Each of these is a well-known risk factor for GDM. The four phenotypic variables alone yielded an AUC of 0.65 and 0.68 in the development and validation sets, respectively. The 11 SNPs alone yielded respective AUCs of 0.71 and 0.77. The additive contributions of phenotype and genotype increased the overall AUCs to a respective 0.75 and 0.83. To our knowledge, this is the highest performance for a genotype-informed GDM prediction algorithm reported in the literature to date. Although the current rise in GDM prevalence is driven mainly by changes in lifestyle, complex genetic determinants contribute to the inherent susceptibility of this disease. Inclusion of genotype-based susceptibility information will support the use of precision medicine, the identification of women at high risk of GDM during the early stages of pregnancy, and the application of personalized preventive interventions. Translation of new findings from genetic studies to the clinic is the most attractive aspect of genome research. One potential clinical application is the development of genetically informed personalized susceptibility profiles and lifestyle recommendations. However, at present, precision medicine has not yet fulfilled such expectations,[51] as it requires a much-needed process of internal and external validation and calibrations to target specific populations. It is therefore necessary to apply sufficient funding and infrastructure to promote the transfer of knowledge, such as the findings presented herein, to society as a whole.

The strengths of this study include a robust modeling strategy for significant attributes, as well as the analysis of a carefully selected list of 114 SNPs according to their reported predictive value. It is worth noting that we did not simply focus on the correlation of each SNP with GDM, but rather on the combined effect of the significant SNPs. Our analysis yielded both a combination and predictive weight of variables that were predictive of the population studied. Our study had some limitations. The analysis was based on data from two cohorts of women and, as such, the results may not be applicable to the entire Mexican population. Ancestry markers were also not genotyped because our aim was to identify markers with a predictive power in the global Mexican population; we were not evaluating variants specific to any particular subethnicity. This, however, could be considered a limitation of our study and should be evaluated in future analyses. Data regarding patient lifestyle, such as diet and sleep duration, which are associated with *MTNR1B* genetic variants, were not collected in this study. Additionally, the small sample size may have affected the accuracy and reliability of the model to an extent. Large-scale multicenter studies need to be performed to further verify this prediction model for GDM.

CONCLUSIONS

This study demonstrated progress towards adapting global findings on genetic variants that predict the risk of developing GDM to the Mexican population. In addition to having developed a good predictive model with the capacity for timely identification of women who require intervention and treatment for GDM, this study may contribute to the understanding of the potential metabolic functions underlying elevated risk. Translation of novel findings from genetic studies to the clinic is the most attractive aspect of genome research. One potential clinical application for the findings of the present study is the development of genetically informed personalized susceptibility profiles alongside lifestyle recommendations. Our findings will potentially support a movement towards personalized recommendations and treatments for each patient. However, the study of the metabolic pathways that underlie GDM susceptibilities is still limited and requires additional research to improve the accuracy, efficiency, and impact on women's care.

Author affiliations
[1]Research and Development Department, Patia Europe, San Sebastian, Spain
[2]Health Sciences University Center, University of Guadalajara, Guadalajara, Mexico
[3]Operative Solutions, Carlos Slim Foundation, Mexico City, Mexico
[4]Faculty of Medicine, Autonomous University of the State of Mexico, Toluca, Mexico
[5]Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico

ORCID iD
Héctor Gallardo-Rincón http://orcid.org/0000-0002-0811-4606

REFERENCES

1 American Diabetes Association. 2. classification and diagnosis of diabetes. *Diabetes Care* 2016;39 Suppl 1:S13–22.
2 Kramer CK, Campbell S, Retnakaran R. Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis. *Diabetologia* 2019;62:905–14.
3 Lowe WL, Scholtens DM, Kuang A, *et al*. Hyperglycemia and adverse pregnancy outcome follow-up study (HAPO FUS): maternal gestational diabetes mellitus and childhood glucose metabolism. *Diabetes Care* 2019;42:372–80.
4 Nehring I, Chmitorz A, Reulen H, *et al*. Gestational diabetes predicts the risk of childhood overweight and abdominal circumference independent of maternal obesity. *Diabet Med* 2013;30:1449–56.
5 Yu Y, Arah OA, Liew Z, *et al*. Maternal diabetes during pregnancy and early onset of cardiovascular disease in offspring: population based cohort study with 40 years of follow-up. *BMJ* 2019;367:l6398.
6 Shah NS, Wang MC, Freaney PM, *et al*. Trends in gestational diabetes at first live birth by race and ethnicity in the US, 2011-2019. *JAMA* 2021;326:660–9.
7 Farrar D. Hyperglycemia in pregnancy: prevalence, impact, and management challenges. *Int J Womens Health* 2016;8:519–27.
8 Martinez-juarez LA, Lomelin-gascon J, Ortega-montiel J, *et al*. 1280-PUB: the impact of proactive and personalized strategies in the self-monitoring of blood glucose in women with diabetes in pregnancy: the experience of the cohort cuido MI embarazo in mexico. *Diabetes* 2021;70:Supplement_1.
9 Egan AM, Simmons D. Lessons learned from lifestyle prevention trials in gestational diabetes mellitus. *Diabet Med* 2019;36:142–50. 10.1111/dme.13772 Available: http://doi.wiley.com/10.1111/dme.2019.36.issue-2
10 Savvidou M, Nelson SM, Makgoba M, *et al*. First-Trimester prediction of gestational diabetes mellitus: examining the potential of combining maternal characteristics and laboratory measures. *Diabetes* 2010;59:3017–22.
11 Sweeting AN, Appelblom H, Ross GP, *et al*. First trimester prediction of gestational diabetes mellitus: a clinical model based on maternal demographic parameters. *Diabetes Res Clin Pract* 2017;127:44–50.
12 Sweeting AN, Wong J, Appelblom H, *et al*. A first trimester prediction model for gestational diabetes utilizing aneuploidy and pre-eclampsia screening markers. *J Matern Fetal Neonatal Med* 2018;31:2122–30.
13 Huerta-Chagoya A, Vázquez-Cárdenas P, Moreno-Macías H, *et al*. Genetic determinants for gestational diabetes mellitus and related metabolic traits in Mexican women. *PLoS One* 2015;10:e0126408.
14 Watanabe RM, Allayee H, Xiang AH, *et al*. Transcription factor 7-like 2 (TCF7L2) is associated with gestational diabetes mellitus and interacts with adiposity to alter insulin secretion in Mexican Americans . *Diabetes* 2007;56:1481–5.
15 Lowe WL Scholtens DM, Kuang A, *et al*. Hyperglycemia and adverse pregnancy outcome follow-up study (HAPO FUS): maternal gestational diabetes mellitus and childhood glucose metabolism. *Diabetes Care* 2019;42:372–80.
16 Hayes MG, Urbanek M, Hivert M-F, *et al*. Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. *Diabetes* 2013;62:3282–91.
17 Lowe WL, Scholtens DM, Sandler V, *et al*. Genetics of gestational diabetes mellitus and maternal metabolism. *Curr Diab Rep* 2016;16:15.
18 Pervjakova N, Moen G-H, Borges M-C, *et al*. Multi-ancestry genome-wide association study of gestational diabetes mellitus highlights genetic links with type 2 diabetes. *Human Molecular Genetics* 2022;31:3377–91.
19 Powe CE, Nodzenski M, Talbot O, *et al*. Genetic determinants of glycemic traits and the risk of gestational diabetes mellitus. *Diabetes* 2018;67:2703–9.
20 Powe C, Allard C, Perron P, *et al*. SAT-123 burden of type 2 diabetes genetic risk alleles differs among physiologic subtypes of gestational diabetes mellitus [abstract]. *Journal of the Endocrine Society* 2019;3(Supplement_1):67.
21 Powe CE, Udler M, Allard C, *et al*. 354-or: physiologic pathways in pregnancy glycemic regulation implicated through genetic clustering analysis [abstract]. *Diabetes* 2019;68.
22 Ding M, Chavarro J, Olsen S, *et al*. Genetic variants of gestational diabetes mellitus: a study of 112 snps among 8722 women in two independent populations. *Diabetologia* 2018;61:1758–68.
23 van Poppel MN, Corcoy R, Hill D, *et al*. Interaction between rs10830962 polymorphism in MTNR1B and lifestyle intervention on maternal and neonatal outcomes: secondary analyses of the DALI lifestyle randomized controlled trial. *The American Journal of Clinical Nutrition* 2022;115:388–96.
24 Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, *et al*. Prediction models for the risk of gestational diabetes: a systematic review. *Diagn Progn Res* 2017;1:3.
25 Gao S, Leng J, Liu H, *et al*. Development and validation of an early pregnancy risk score for the prediction of gestational diabetes mellitus in chinese pregnant women. *BMJ Open Diabetes Res Care* 2020;8:e000909.
26 Aydın H, Çelik Ö, Yazıcı D, *et al*. Prevalence and predictors of gestational diabetes mellitus: a nationwide multicentre prospective study. *Diabet Med* 2019;36:221–7.
27 Wang J, Lv B, Chen X, *et al*. An early model to predict the risk of gestational diabetes mellitus in the absence of blood examination indexes: application in primary health care centres. *BMC Pregnancy Childbirth* 2021;21:814.
28 de la Torre NG, Assaf-Balut C, Jiménez Varas I, *et al*. Effectiveness of following mediterranean diet recommendations in the real world in the incidence of gestational diabetes mellitus (gdm) and adverse maternal-foetal outcomes: a prospective, universal, interventional study with a single group. the st carlos study. *Nutrients* 2019;11:1210.
29 Cho NH, Jang HC. Better understanding and new insight of genetic risk loci for gestational diabetes mellitus. *J Diabetes Investig* 2012;3:424–6.
30 Manning AK, Hivert M-F, Scott RA, *et al*. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012;44:659–69.
31 Saxena R, Elbers CC, Guo Y, *et al*. Large-Scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *The American Journal of Human Genetics* 2012;90:410–25.
32 Williams AL, Jacobs SBR, Moreno-Macías H, *et al*. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in mexico. *Nature* 2014;506:97–101.
33 Zhang C, Bao W, Rong Y, *et al*. Genetic variants and the risk of gestational diabetes mellitus: a systematic review. *Hum Reprod Update* 2013;19:376–90.
34 Wu L, Cui L, Tam WH, *et al*. Genetic variants associated with gestational diabetes mellitus: a meta-analysis and subgroup analysis. *Sci Rep* 2016;6:30539.

35 Gabriel S, Ziaugra L, Tabbaa D. Snp genotyping using the sequenom massarray iPLEX platform. *Curr Protoc Hum Genet* 2009;Chapter 2:Unit

36 Koivusalo SB, Rönö K, Klemetti MM, *et al*. Gestational diabetes mellitus can be prevented by lifestyle intervention: the Finnish gestational diabetes prevention study (RADIEL). *Diabetes Care* 2016;39:24–30.

37 Renström F, Koivula RW, Varga TV, *et al*. Season-dependent associations of circadian rhythm-regulating loci (CRY1, CRY2 and MTNR1B) and glucose homeostasis: the glacier study. *Diabetologia* 2015;58:997–1005.

38 Lyssenko V, Nagorny CLF, Erdos MR, *et al*. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* 2009;41:82–8.

39 Dashti HS, Follis JL, Smith CE, *et al*. Gene-environment interactions of circadian-related genes for cardiometabolic traits. *Diabetes Care* 2015;38:1456–66.

40 Wu H, Wylie-Rosett J, Qi Q. Dietary interventions for weight loss and maintenance: preference or genetic personalization? *Curr Nutr Rep* 2013;2:189–98.

41 Scott RA, Lagou V, Welch RP, *et al*. Large-Scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 2012;44:991–1005.

42 Furukawa N, Ongusaha P, Jahng WJ, *et al*. Role of rho-kinase in regulation of insulin action and glucose homeostasis. *Cell Metab* 2005;2:119–29.

43 Baroukh N, Ravier MA, Loder MK, *et al*. Microrna-124A regulates Foxa2 expression and intracellular signaling in pancreatic beta-cell lines. *J Biol Chem* 2007;282:19575–88.

44 Yu X, Zhong L. Pioglitazone/microrna-141/FOXA2: a novel axis in pancreatic β-cells proliferation and insulin secretion. *Mol Med Report* 2018;17:7931–8.

45 Kretowski A, Adamska E, Maliszewska K, *et al*. The rs340874 PROX1 type 2 diabetes mellitus risk variant is associated with visceral fat accumulation and alterations in postprandial glucose and lipid metabolism. *Genes Nutr* 2015;10:4.

46 Alvine T, Dhasarathy A, Bundy A, *et al*. RBMS1 methylation and mRNA expression are differentially regulated in placenta tissue from obese women (P11-131-19). *Current Developments in Nutrition* 2019;3 nzz048.

47 Coughlan MT, Vervaart PP, Permezel M, *et al*. Altered placental oxidative stress status in gestational diabetes mellitus. *Placenta* 2004;25:78–84.

48 Santos-Rosendo C, Bugatto F, González-Domínguez A, *et al*. Placental adaptive changes to protect function and decrease oxidative damage in metabolically healthy maternal obesity. *Antioxidants (Basel)* 2020;9:794.

49 Deng Y-N, Xia Z, Zhang P, *et al*. Transcription factor RREB1: from target genes towards biological functions. *Int J Biol Sci* 2020;16:1463–73.

50 Chen Z, Yu H, Shi X, *et al*. Functional screening of candidate causal genes for insulin resistance in human preadipocytes and adipocytes. *Circ Res* 2020;126:330–46.

51 Tusié-Luna MT, Aguilar-Salinas CA. PRECISION medicine for metabolic disorders in low- and middle-income countries: areas of opportunity and challenges for the future. *Rev Invest Clin* 2021;73:316–20.