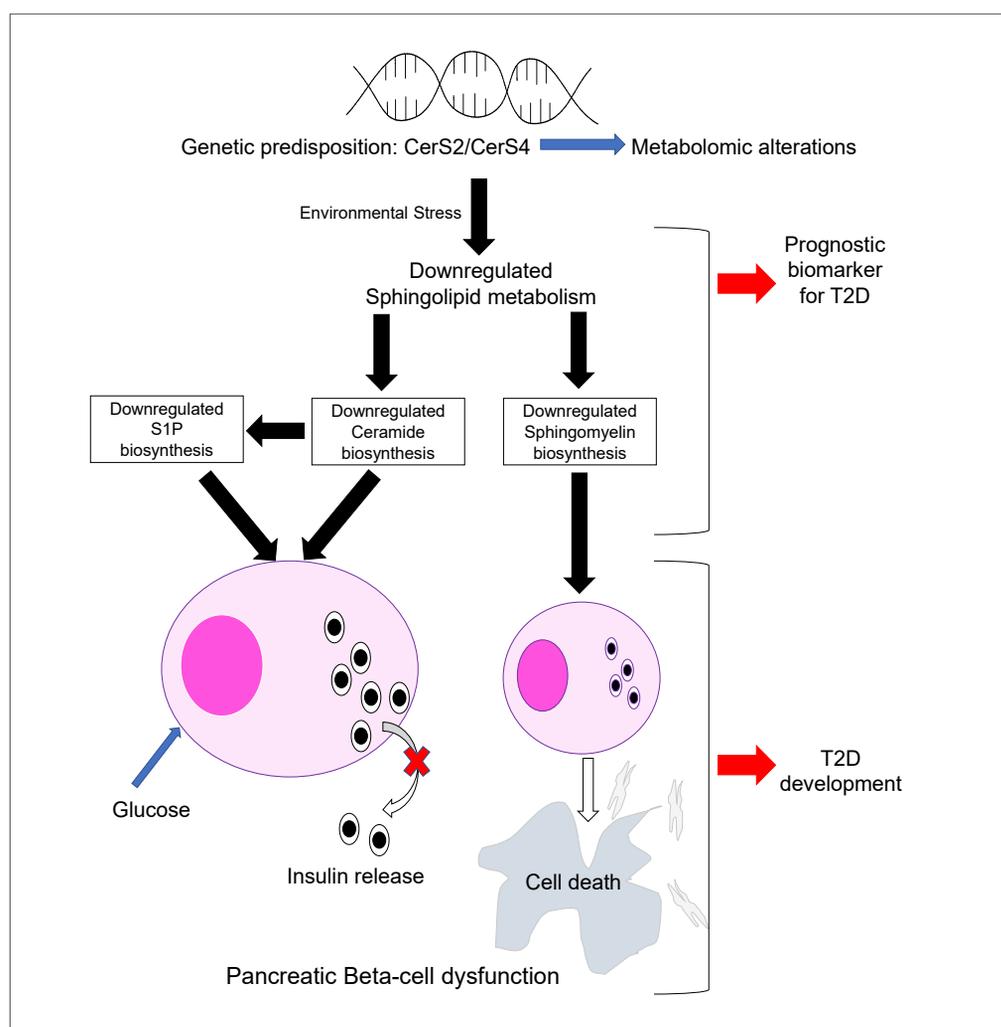


## Article

Diminished Sphingolipid Metabolism, a Hallmark of Future Type 2 Diabetes Pathogenesis, Is Linked to Pancreatic  $\beta$  Cell Dysfunction

Saifur R. Khan,  
Yousef Manialawy,  
Andreea  
Obersterescu,  
Brian J. Cox, Erica  
P. Gunderson,  
Michael B.  
Wheeler

b.cox@utoronto.ca (B.J.C.)  
erica.gunderson@kp.org  
(E.P.G.)  
michael.wheeler@utoronto.ca  
(M.B.W.)

**HIGHLIGHTS**

Diminished sphingolipid metabolism is associated with a transition from GDM to T2D

Merging metabolomics and GWAS data identifies sphingolipid metabolism genes associated with T2D

Inhibiting sphingolipid metabolism impairs pancreatic  $\beta$  cell function

Incorporating sphingolipid SM C16:10 measurements with an OGTT improves T2D prediction

Khan et al., iScience 23,  
101566  
October 23, 2020 © 2020 The  
Author(s).  
[https://doi.org/10.1016/  
j.isci.2020.101566](https://doi.org/10.1016/j.isci.2020.101566)

## Article

Diminished Sphingolipid Metabolism, a Hallmark of Future Type 2 Diabetes Pathogenesis, Is Linked to Pancreatic  $\beta$  Cell Dysfunction

Saifur R. Khan,<sup>1,2</sup> Yousef Manialawy,<sup>1,2</sup> Andreea Obersterescu,<sup>1</sup> Brian J. Cox,<sup>1,3,\*</sup> Erica P. Gunderson,<sup>4,\*</sup> and Michael B. Wheeler<sup>1,2,5,\*</sup>

## SUMMARY

**Gestational diabetes mellitus (GDM) is the top risk factor for future type 2 diabetes (T2D) development. Ethnicity profoundly influences who will transition from GDM to T2D, with high risk observed in Hispanic women. To better understand this risk, a nested 1:1 pair-matched, Hispanic-specific, case-control design was applied to a prospective cohort with GDM history. Women who were non-diabetic 6–9 weeks postpartum (baseline) were monitored for the development of T2D. Metabolomics were performed on baseline plasma to identify metabolic pathways associated with T2D risk. Notably, diminished sphingolipid metabolism was highly associated with future T2D. Defects in sphingolipid metabolism were further implicated by integrating metabolomics and genome-wide association data, which identified two significantly enriched T2D-linked genes, *CERS2* and *CERS4*. Follow-up experiments in mice and cells demonstrated that inhibiting sphingolipid metabolism impaired pancreatic  $\beta$  cell function. These data suggest early postpartum alterations in sphingolipid biosynthesis contribute to  $\beta$  cell dysfunction and T2D risk.**

## INTRODUCTION

The choice of proper intervention for any disease depends on adequate knowledge of disease pathophysiology. With type 2 diabetes (T2D), current drug interventions either directly or indirectly contribute to the management of hyperglycemia, which is the primary endpoint of the underlying metabolic changes leading to T2D. Studies have shown that strict blood glucose management has no substantial beneficial effects on long-term morbidity and mortality (Akirov et al., 2017; Amiel et al., 2019). As such, it is critical to develop a better understanding of the early-stage T2D pathophysiology to devise proper interventions. Since women with gestational diabetes mellitus (GDM) exhibit a very high transition rate (i.e., ~35% women with GDM 10 years postpartum) from postpartum normoglycemia to T2D (Gunderson et al., 2007; Kim et al., 2002), they are ideal models for studying early-stage pathophysiology of T2D and for discovering predictive biomarkers (i.e., prognostic).

The global prevalence of gestational diabetes mellitus (GDM) has risen in recent years to now affect nearly 14% of all pregnancies (Koning et al., 2016; Hunt and Schuller, 2007; Hod et al., 2019; Zhu and Zhang, 2016). Women who experience a GDM pregnancy have a 74% increased age-adjusted risk for subsequent T2D development compared with women with no history of GDM (Ratner, 2007). Although most women with GDM exhibit normoglycemia immediately after delivery, ~35% of them will progress to T2D within 10 years (Gunderson et al., 2007; Kim et al., 2002). As such, GDM is the top risk factor for future T2D development. The transition rate from normoglycemia to T2D (i.e., hyperglycemia) and disease complication patterns (i.e., microvascular complications, macrovascular complications) vary widely among different races and ethnic backgrounds (Arora et al., 2019) (<https://www.cdc.gov/>). This variation is due to the strong influence of race/ethnicity over metabolic variables stemming from genetic, dietary, and cultural differences. This variation in T2D epidemiology and its complications suggests an underlying heterogeneity in the disease among races/ethnicities. Using 510 women from the RADIEL cohort (i.e., a lifestyle intervention study), Huvinen et al. recently demonstrated the existence of GDM heterogeneity as a significant challenge in developing an optimal risk scoring assessment (Huvinen et al., 2018b). A follow-up study of this same group also highlighted the variability in the long-term risk of diabetes and metabolic syndromes due to this

<sup>1</sup>Department of Physiology, University of Toronto, ON, Canada

<sup>2</sup>Advanced Diagnostics, Metabolism, Toronto General Research Institute, ON, Canada

<sup>3</sup>Department of Obstetrics and Gynaecology, University of Toronto, ON, Canada

<sup>4</sup>Kaiser Permanente Northern California, Division of Research, Oakland, CA, USA

<sup>5</sup>Lead Contact

\*Correspondence: [b.cox@utoronto.ca](mailto:b.cox@utoronto.ca) (B.J.C.), [erica.gunderson@kp.org](mailto:erica.gunderson@kp.org) (E.P.G.), [michael.wheeler@utoronto.ca](mailto:michael.wheeler@utoronto.ca) (M.B.W.)

<https://doi.org/10.1016/j.isci.2020.101566>



heterogeneity (Huvinen et al., 2018a). T2D is linked to genetic predisposition (Udler et al., 2018), which is significantly influenced by racial/ethnic origins (Bamshad, 2005). Therefore, the preferred method to address this heterogeneity is adopting a single race/ethnicity-focused precision medicine approach; implementing an additional pair-matching strategy for major clinical covariates further aids in reducing other confounding factors from the study (Khan et al., 2019a). A recent Centers for Disease Control and Prevention (CDC) study has found that Hispanic-race/ethnicity holds the highest risk for T2D development (Cheng et al., 2019). However, the reasons underlying the higher transition rate to T2D in the Hispanic race have not been ascertained. Additionally, the American Diabetes Association (ADA) recommended routine screening protocol (i.e., a 2 h 75 g OGTT at 6–12 weeks postpartum and once per year every 1–3 years), which suffers from low discrimination power (area under the curve [AUC] ~70%) (Khan et al., 2019b; Abdul-Ghani et al., 2009; Allalou et al., 2016) and has not yet been evaluated in Hispanic women with GDM independently.

In the present study, we adopted a single race/ethnicity-focused precision medicine approach by focusing on the Hispanic women in the SWIFT (Study of Women, Infant Feeding and Type 2 Diabetes after GDM) cohort to better understand the early-stage T2D pathophysiology through the identification of major pathways, their genetic predispositions with risk-alleles (i.e., single-nucleotide polymorphisms [SNPs]), and a predictive risk-score panel for T2D up to 2 years post-baseline. This objective was accomplished using both metabolomic and lipidomic analyses of fasting plasma samples obtained from the 2 h 75 g OGTT at 6–9 weeks postpartum (study baseline), along with GDM severity and other clinical and lifestyle behavioral variables. One goal was to integrate genome-wide association study (GWAS) data with metabolic pharmacological inhibition using *in vivo* and *in vitro* models to determine pathways with metabolic dysregulation that are causal in promoting glucose dysmetabolism and impairing pancreatic  $\beta$  cell function. The second goal was to identify a simple metabolite signature with strong predictive power for the development of T2D.

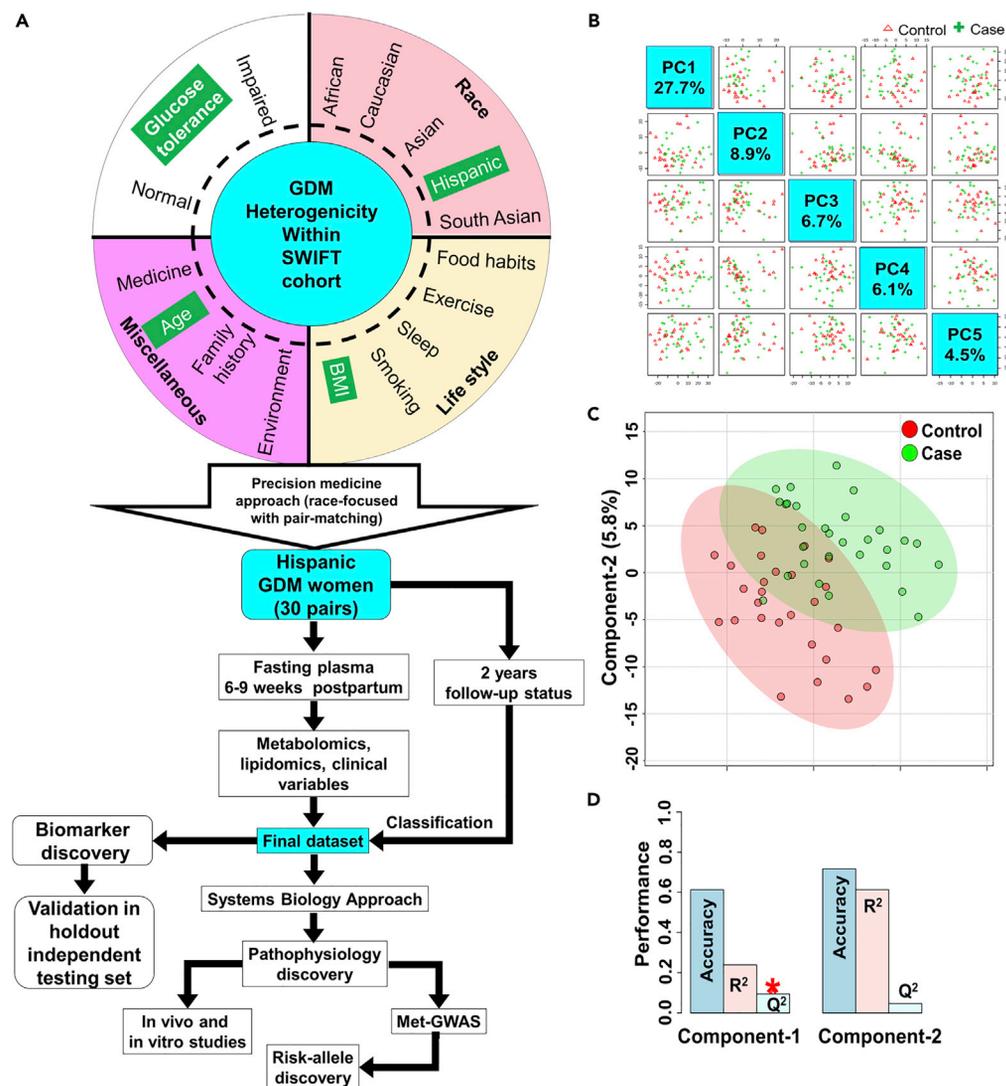
## RESULTS

### Participants of the SWIFT Cohort: Sociodemographic and Clinical Characteristics

The SWIFT study is a prospective, racially and ethnically diverse observational research cohort of 1,035 women with GDM (75% minority; 8% black, 24% white, 34% Hispanic, 30% Asian, and 2% Other) recruited during pregnancy within the Kaiser Permanente Northern California (KPNC) healthcare system from 2008 to 2011. The study participants consented to three in-person research visits and met study eligibility criteria: (1) age 20–45 years at delivery, (2) received prenatal care at Kaiser medical facilities and delivered at Kaiser hospitals, (3) GDM pregnancy diagnosed by 3-h 100 g OGTT using Carpenter and Coustan's criteria, (4) delivered a singleton, live birth  $\geq$  35 weeks gestation, (5) no pre-existing diabetes (DM) or other serious medical conditions prior to index GDM pregnancy, and other criteria as previously described (Gunderson et al., 2011, 2012, 2015). KPNC prenatal standardized treatments for GDM included dietary therapy, oral medication, and insulin treatment (Table S1). Participant clinical characteristics were mean (SD) for pre-pregnancy BMI of 29.6 (7.2) kg/m<sup>2</sup>, total gestational weight gain of 10.3 (6.8) kg, and age at delivery of 33.3 (4.8) years. About 27% of participants had a family income  $\leq$  185% of the federal poverty level, and 33% had impaired glucose tolerance (IGT) and/or impaired fasting glucose (IFG) at study baseline (6–9 weeks post delivery) based on the 2 h 75 g OGTT (Gunderson et al., 2015).

All SWIFT Study research protocols were approved by the KPNC institutional review board requirement, and women provided written informed consent at the 6- to 9-week postpartum research visit (baseline) and attended annual follow-up visits for 2 years. These visits included research 2 h 75 g OGTTs, completion of surveys, and measurement of anthropometry and body composition by trained research staff. Blood glucose and insulin were measured in EDTA-treated plasma by the Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington, Seattle, WA. Other clinical variables, including GDM severity from the prenatal 3-h 100 g OGTT, were obtained from the electronic medical records. Details of the SWIFT Study and participant characteristics have been described elsewhere (Gunderson et al., 2011, 2012, 2015).

This investigation employed a precision medicine approach to develop a nested 1:1 pair-matched (i.e., age, BMI, sex, and glucose tolerance) case-control study using 60 Hispanic women with GDM within the SWIFT cohort. Women who did not progress to T2D within 2 years of postpartum are classified as “Control.” Each control is then pair-matched with case women who developed T2D within 2 years of postpartum.



**Figure 1. Study Design via a Precision Medicine Approach**

(A) A flow diagram of the study design. In brief, a nested 1:1 pair-matched (i.e., age, BMI, and glucose tolerance) case-control study was designed using 60 Hispanic women with GDM within the SWIFT cohort. Women who did not progress to T2D within 2 years postpartum are classified as “Control.” Women who developed T2D within 2 years postpartum are classified as “Case.” This classification was used in the final dataset for both biomarker discovery and pathophysiology discovery. The identified pathways were further investigated in both *in vivo* and *in vitro* models for translation. In addition, a mapping strategy (metGWAS) was devised to identify the genetic predisposition through risk-allele discovery for altered pathophysiology from a stand-alone metabolomic dataset.

(B) PCA analysis revealed the major components of the final dataset.

(C) 2D score plot with a 95% confidence region in PLS-DA analysis.

(D) Ten-fold cross-validation of PLS-DA analysis using two major components showed a significant difference between case and control in terms of  $Q^2$  (\* $p < 0.05$ ).

The study design (including the flow of participants throughout the study) is illustrated in Figure 1A. Baseline (6–9 weeks postpartum) sociodemographic and clinical characteristics of the participants of each group (i.e., case and control) are summarized in Table S1. The case women were more likely to have been treated with oral medications (17 of 30 cases) or insulin (2 cases) during pregnancy than their matched control women. This indicates that case women may possess genetic predispositions for metabolic diseases (i.e., diabetes) (Ding et al., 2018). Neither pre-pregnancy and baseline (6–9 weeks postpartum) BMI nor total energy intake differed significantly between case and control. However, cases engaged in significantly higher levels of regular physical activity ( $p < 0.05$ ) than control in early postpartum and these

(case) women exhibited significantly higher fasting glucose ( $p < 0.001$ ), 2 h OGTT ( $p < 0.001$ ), fasting insulin ( $p < 0.05$ ), and HOMA-IR ( $p < 0.01$ ) at study baseline. For matched pairs, the lactation intensity did not differ significantly, as expected, given the matching on glucose tolerance (normal or impaired GT).

### Evaluation of the Final Dataset—Chemometric Analyses

The final dataset was composed of 824 analytes (i.e., metabolites, lipids, and clinical variables) on which both PCA and PLS-DA were carried out. The unsupervised PCA clustering methods revealed two principal components (i.e., PC1 and PC2) with percent variance values of 28% and 9%, respectively, of the total study population (Figure 1B). No other component was high enough ( $>10\%$ ), indicating the absence or minimal influence of confounding factors in this study. The supervised PLS-DA clustering methods revealed a subtle but distinguishable significant separation between case and control in the two-dimensional score plot (Figure 1C). The performance of PLS-DA was evaluated using 10-fold cross-validation analysis, where these two clusters yielded 61% and 71% accuracies based on  $R^2$  and  $Q^2$  values (Figure 1D). The significant difference between  $Q^2$  ( $p < 0.05$ ) of two components indicates a significant class separation between cases and controls.

### Downregulated Sphingolipid Metabolism—a Major Early-Stage T2D Pathophysiology

In a non-parametric differential analysis of our entire dataset, a total of 130 analytes were found to be significantly altered between the two groups. Among them, 76 analytes were downregulated and 54 were upregulated (Table S2). To robustly determine the statistical significance of the families of altered metabolites, we tested for significant enrichment of KEGG pathways using over-representation analysis (Figure 2A). Only two pathways had a false discovery rate (FDR)  $<0.05$  (or  $-\log_{10}(\text{FDR}) > 1.3$ ); these were downregulated sphingolipid metabolism (FDR  $<0.001$ ) and upregulated fatty acid biosynthesis (FDR  $<0.001$ ) (Table S3). In pathway topology analysis, a relative-betweenness centrality algorithm was applied to determine the impact of each pathway in the study using the likelihood (or importance) parameter in a 0–1 scale. The analysis identified downregulated sphingolipid metabolism as the most impactful among all significant pathways with a value of 0.33, whereas upregulated fatty acid biosynthesis had an importance score of zero. Since sphingolipid metabolism has a higher likelihood or importance value than fatty acid biosynthesis, it is probable that fatty acid biosynthesis upregulation may be due to the downregulation of sphingolipid metabolism. An interaction map between fatty acid biosynthesis and sphingolipid metabolism (Figure 2B) also revealed that the downregulated sphingolipid metabolism could explain the upregulation of fatty acid biosynthesis via the partial downregulation of glycerophospholipid species. Therefore, downregulated sphingolipid metabolism is potentially an early contributor to future T2D onset.

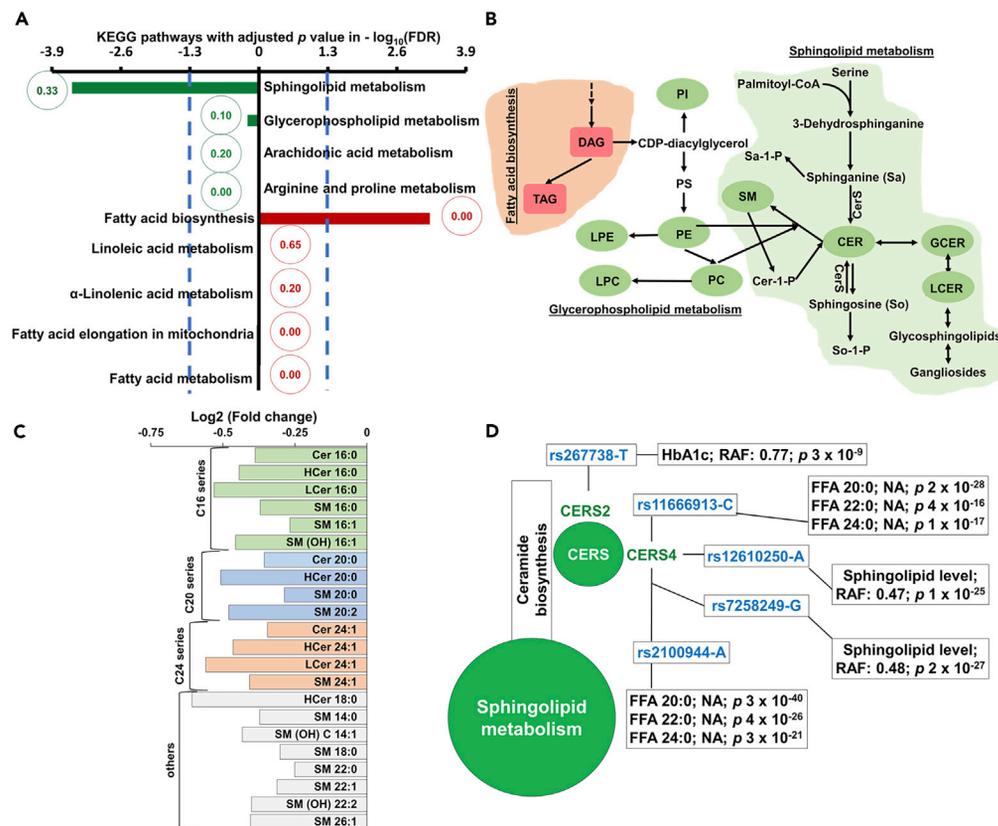
The identification of sphingolipid metabolism as a possible important marker of early-stage T2D pathophysiology warranted further inspection of its significantly altered (i.e., downregulated) metabolites (Figure 2C). It was found that the C16, C20, and C24 series of sphingolipids were significantly downregulated within all parts of the sphingolipid metabolism pathway (i.e., glycosylsphingolipid biosynthesis, ceramide biosynthesis, and sphingomyelin biosynthesis) (Figure 2C).

### Metabolic Disruption Significantly Correlates to T2D Genetic Polymorphisms

As much of sphingolipid metabolite levels are enzymatically regulated, we reasoned that reduced activity levels of enzymes in the sphingolipid pathway could explain an increased risk of T2D. Importantly, both lower expression or loss of function may reduce enzyme activity, and both processes are influenced by genetic polymorphisms. To test polymorphism-metabolite relationships, we mapped our metabolic data to available GWAS data. GWAS data were linked to metabolites by mapping polymorphisms (SNPs) to gene loci and metabolites to interacting enzyme genes. Using MBRole 2.0, we identified ceramide glucosyltransferase (UGCG), several ceramide synthases (i.e., CERS1, CERS2, CERS3, CERS4, CERS5, and CERS6), and sphingomyelin synthases (i.e., SGMS1 and SGMS2) as genes encoding enzymes responsible for sphingolipid production. Mapping these genes to GWAS datasets from studies of T2D identified five SNPs belonging to two of the sphingolipid synthesis genes (i.e., CERS2 and CERS4) (Figure 2D). Applying a hypergeometric test found the intersection of our metabolite sphingolipid pathway with the GWAS data to be significant ( $p < 0.05$ ).

### Translation of Downregulated Sphingolipid Metabolism—an Early-Stage Disease Phenotype

As we found a significant link to the development of T2D based on enriched gene polymorphisms and altered metabolic profiles, we investigated if downregulated sphingolipid metabolism could be causal.



**Figure 2. The Discovery of Early-Stage Pathways via a Systems Biology Approach and a Metabolite-GWAS Analysis for Risk Allele-Containing SNPs Identification**

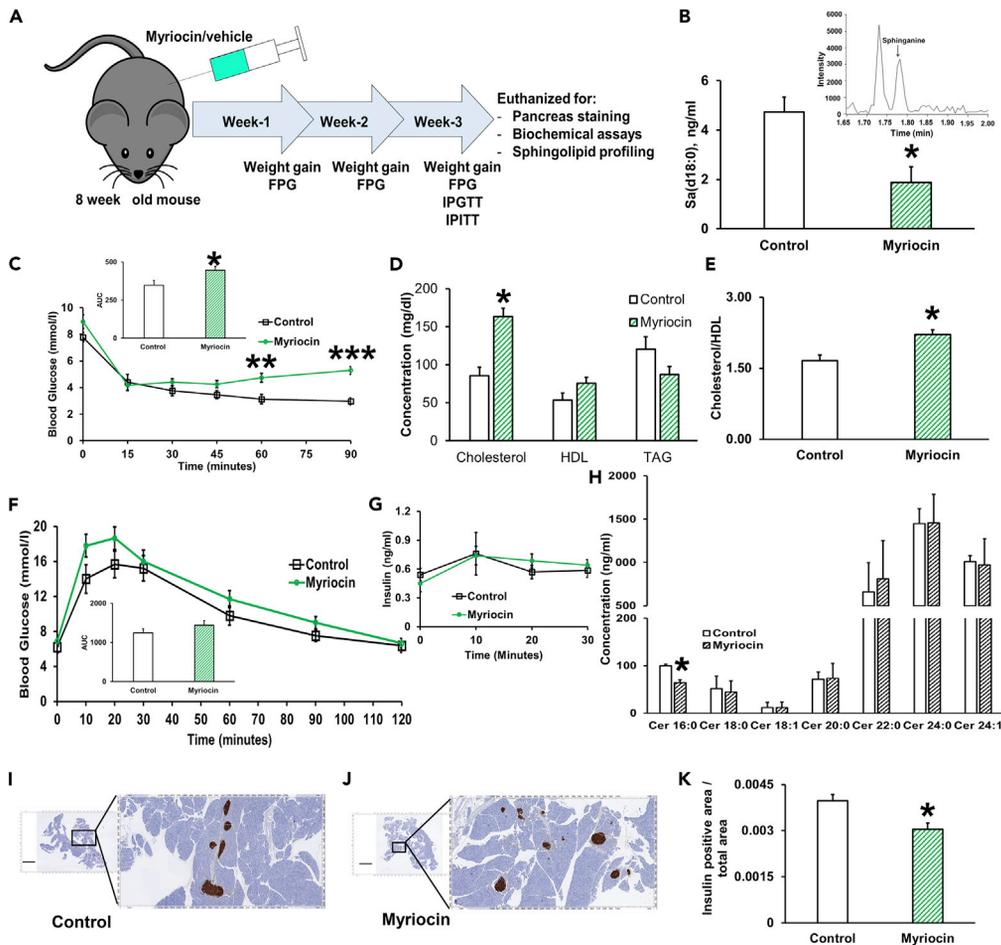
(A) An over-representation pathway analysis accompanied with pathway topology analysis identified the involved KEGG pathways with their significance [based on adjusted  $p$  values in  $-\log_{10}(\text{FDR})$ , where  $-\log_{10}(\text{FDR}) > 1.3$  represents significant; “+” indicates upregulated pathways, whereas “-” indicates downregulated pathways] and importance (in the circle) within this study context. Red indicates upregulation, whereas green indicates downregulation.

(B) The interaction map between fatty acid biosynthesis and sphingolipid metabolism via glycerophospholipid metabolism. The red square indicates the mean upregulation trend, whereas the green circle indicates the mean downregulation trend.

(C) The significantly altered (i.e., downregulated) sphingolipids are presented here based on C-series with their logarithmic fold change.

(D) Discovery of SNPs with T2D risk alleles from identified sphingolipid metabolism pathway using metabolite-GWAS. We identified two genes (CERS2 and CERS4) that contain five T2D SNPs. The reported traits, SNP frequency (RAF) if reported (NA if not reported), and significance ( $p$  value) in the reporting studies are listed here.

*In vivo* and *in vitro* models were employed to examine the role of sphingolipids in glucose homeostasis and  $\beta$  cell function. Initial experiments employed myriocin, a potent inhibitor of serine-palmitoyl-transferase (SPT), which catalyzes the first step of ceramide biosynthesis in sphingolipid metabolism. Myriocin was administered intraperitoneally to mice to examine the effects of downregulated sphingolipid metabolism on glucose homeostasis (Figure 3A). After completion of 3 weeks of treatment, LC-MS analyses on fasting plasma samples of mice showed a significant decrease of sphinganine, Sa(d18:0) (Figure 3B), as anticipated with chronic myriocin treatment. No significant difference with vehicle control was observed for body weight, fasting blood glucose (FBG), and fasting insulin (FI). During the insulin tolerance test (ITT), myriocin-treated mice showed significantly reduced late-phase insulin responsiveness (Figure 3C) with significantly higher AUC (Figure 3C, inset), suggesting latent insulin resistance. Biochemical analyses of fasting plasma showed that both cholesterol and the cholesterol/HDL ratio were significantly increased in myriocin-treated mice (Figures 3D and 3E). However, the GTT did not reveal any differences in glucose excursions (Figures 3F and 3G). Ceramide profiling revealed that ceramide 16:0 (Cer 16:0) was significantly decreased (Figure 3H). Pancreatic staining in myriocin-treated mice revealed a significant reduction in the total insulin-positive area after normalizing by total area (Figures 3I–3K).



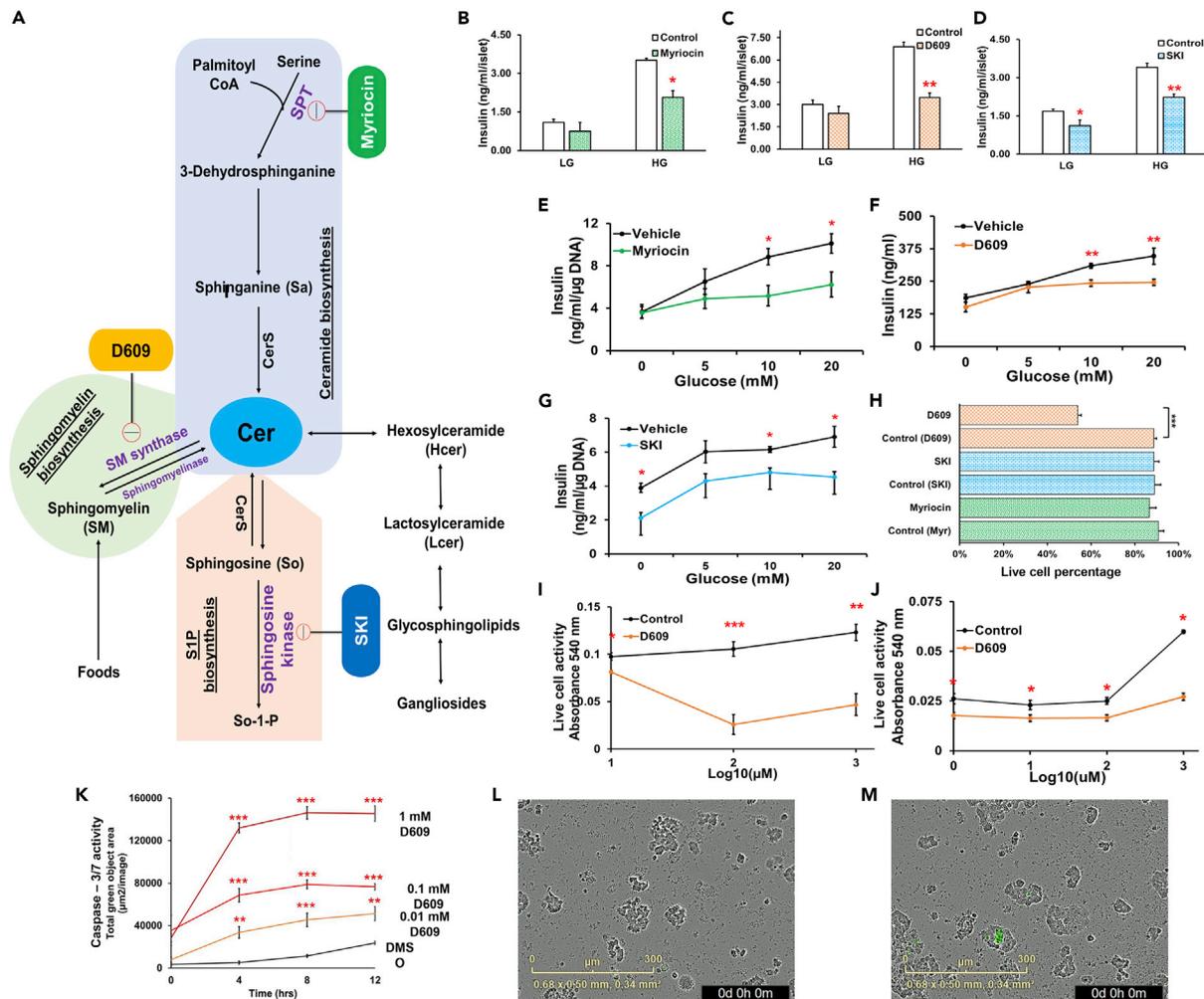
**Figure 3. Translational Approach—In Vivo Studies**

(A) A simplistic diagram of *in vivo* study design and workflow.  
 (B) Myriocin-treated mice exhibited significantly decreased sphinganine [Sa(d18:0)], indicating the successful inhibition of SPT (n = 3). Inset depicts a sphinganine chromatogram.  
 (C) The ITT results showed a significant increase in late-phase insulin intolerance in myriocin-treated mice (n = 11). The bar graph inset shows that the myriocin-treated mice exhibited significantly higher AUC in ITT.  
 (D and E) In biochemical assays, myriocin-treated mice exhibited significantly higher cholesterol and cholesterol/HDL ratios (n = 3). Here, the white bars represent control group, whereas green bars represent myriocin-treated mice.  
 (F) The GTT results showed that myriocin-treated mice were not significantly different in glucose tolerance (n = 11). The AUC of GTT is inset here.  
 (G) Secreted insulin levels during GTT up to 30 min (n = 11).  
 (H) The ceramide profile showed the downregulation of Cer16:0 in the myriocin-treated mice (n = 3).  
 (I and J) Representative insulin-stained pancreas (5 mm thickness, longitudinally sectioned through the pancreatic head-to-tail axis) from control and myriocin-treated mice, respectively; scale bars, 3 mm. Brown-staining identifies insulin-positive pancreatic islets.  
 (K) Myriocin-treated mice exhibited significantly decreased insulin-positive area in pancreatic immunohistochemical studies (n ≥ 7).

A two-tailed, paired t test was carried out for each comparison; unadjusted p values: \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001 versus control. Error bars present mean ± SEM.

### Downregulated Sphingolipid Metabolism in the $\beta$ Cell—Impact on the Function

Given that inhibition of SPT in whole animals produced a less sensitive late-phase insulin response suggesting insulin resistance and was accompanied by an observed overall small but significant reduction in pancreatic  $\beta$  cell area, we employed *in vitro* models to determine the direct effects of downregulated sphingolipid metabolism on pancreatic  $\beta$  cell function and survival using myriocin (SPT inhibitor), D609 (sphingomyelin synthase inhibitor), and SKI (sphingosine kinase inhibitor) (Figure 4A). Both myriocin



**Figure 4. The Role of Different Sphingolipids in Pancreatic  $\beta$  Cell Function and Health—In Vitro Studies**

(A) A simplified diagram of sphingolipid metabolism and inhibitors.

(B) Myriocin treatment, which inhibits SPT to block ceramide biosynthesis, induced a decrease of GSIS without hampering basal insulin secretion in isolated islets ( $n = 3$ ).

(C) D609 treatment, which inhibits sphingomyelin synthase to block sphingomyelin biosynthesis, induced a decrease of GSIS without affecting basal insulin secretion in isolated islets ( $n = 3$ ).

(D) SKI treatment, which inhibits sphingosine kinase to block S1P biosynthesis, induced a decrease of both GSIS and basal insulin secretion in isolated islets ( $n = 3$ ).

(E) Glucose-dependent GSIS in Min6(K8) cells, where myriocin treatment showed a glucose dose-dependent decrease in GSIS ( $n = 4$ ).

(F) Glucose-dependent GSIS in Min6(K8) cells, where D609 treatment showed a glucose dose-dependent decrease in GSIS ( $n = 4$ ).

(G) Glucose-dependent GSIS in Min6(K8) cells, where SKI treatment showed both a decrease in basal insulin secretion and a glucose dose-dependent decrease in GSIS ( $n = 3$ ).

(H) Trypan blue exclusion assays on Min6(K8) cells revealed that both myriocin and SKI were not cytotoxic under the used concentrations; however, D609 treatment induced significant cell death ( $n = 3$ ).

(I and J) MTT assays were carried out for dose-dependent D609 treatment ( $n = 3$ ) on both Min6(K8) cells and isolated murine islets, respectively, revealing that D609 is cytotoxic even at low concentrations. MTT assay measures the metabolic activities of live cells at 540 nm of absorbance. Black line represents control group, whereas red line represents the D609 treated cells.

(K) D609 showed a dose-dependent induction of apoptosis through activation of green-fluorescence of caspase-3/7 in Incucyte live-cell imaging. The apoptosis for 0.01, 0.1, and 1 mM of D609 reached a plateau within 12, 8, and 4 h, respectively ( $n = 3$ ).

(L) A representative video file of Caspase-3/7 activity for DMSO treatment (control).

(M) A representative video file of Caspase-3/7 activity for D609 treatment. Both of these videos were formatted the same way for comparative observation. Green color represents caspase-3/7 activation (i.e., apoptosis).

A two-tailed, paired t test was carried out for each comparison; unadjusted p values: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  versus control. Error bars present mean  $\pm$  SEM. LG, low glucose (0 mM glucose); HG, high glucose (10 mM glucose).

treatment and D609 treatment significantly decreased glucose-stimulated insulin release without altering basal insulin secretion of isolated mouse islets (Figures 4B and 4C). In the case of SKI treatment of isolated mouse cells, both basal insulin secretion and glucose-stimulated insulin secretion were found to be significantly decreased (Figure 4D). When treating Min6(K8) cells, both myriocin and D609 caused a significant decrease in insulin secretion at 10 and 20 mM glucose (Figures 4E and 4F). In the case of SKI treatment, both basal insulin secretion (0 mM glucose) and secretion in the presence of stimulatory glucose (10 and 20 mM) were significantly decreased in Min6(K8) cells (Figure 4G).

To investigate whether this observed pancreatic  $\beta$  cell dysfunction seen with sphingolipid metabolism inhibitors also caused cytotoxicity and potentially cell death, cytotoxicity studies were carried out, and apoptosis was assessed. Cytotoxicity studies revealed that D609 was indeed cytotoxic (Figure 4H) in a dose-dependent manner (Figures 4I and 4J); this was linked to the activation of an apoptosis cascade (Figures 4K–4M; Video S1 and S2).

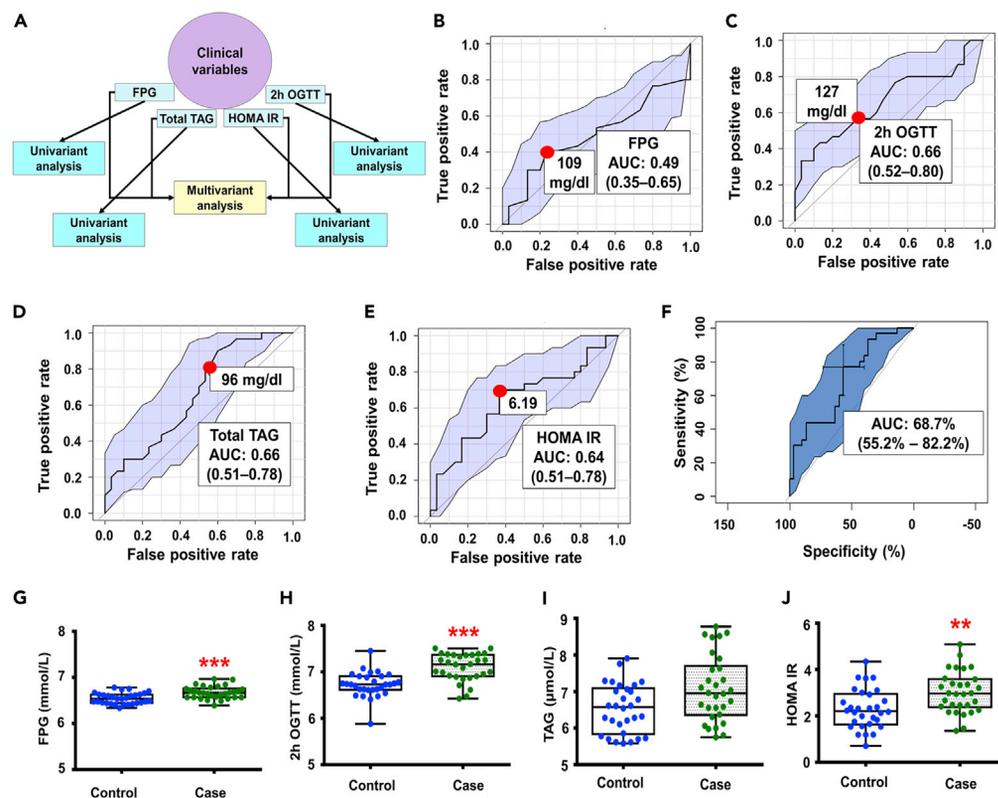
### Evaluation of Clinical Variables as Prognostics

Given the potential to use wide-scale metabolomic analysis to identify metabolites with predictive value prior to T2D onset (i.e., at baseline), this approach was applied to the Hispanic cohort data. However, we first examined the prognostic power of traditional clinical variables such as fasting plasma glucose (FPG), 2 h glucose in OGTT (2 h OGTT), and HOMA-IR using both univariate and multivariate analysis (Figure 5A) on all (60 Hispanic) participants. Total fasting triacylglycerol (Total TAG) was also included, as TAG levels have a strong positive correlation with T2D (Tirosh et al., 2008). Univariate logistic regression of these clinical variables was carried out to reveal their capacity in class separation. The ROC-AUC of FPG, 2 h OGTT, total fasting triglycerides (total TAG), and HOMA-IR were found to be 49%, 66%, 66%, and 64%, respectively (Figures 5B–5E). In sMLR (stepwise multiple logistic regression) analysis, all clinical variables were combined to determine collective class-separation capacity and ROC-AUC was calculated at 68.7% (Figure 5F). The box-whisker plots of these variables are presented in Figures 5G–5J. Three (FPG, 2 h OGTT, and HOMA-IR) of these four clinical variables were significantly higher at baseline in the case group.

### The Discovery of a Novel Prognostic Employing Metabolites

To potentially identify a predictive risk-score panel with high discriminating power (AUC), all analytes (i.e., metabolites, lipids, and clinical variables) were combined for analysis and divided into training and testing sets. Using the training dataset we identified a three-analyte predictive risk-score panel (estimated coefficient of intercept: 0.95, odds ratio: 8.9) using a combination of both the random forest ranking algorithm and univariate ranking system, followed by application of the sMLR algorithm in the training set (i.e., a dataset of 20 pairs of cases and controls) (Figure S1). This three-analyte prognostic comprises one clinical variable (2 h OGTT) and two lipids (SM C16:0, and DAG16:0/16:0). In a looping experiment (described in the Methods section, Figure S1), these three variables showed an incremental increase in AUC (Figure 6A). The details (i.e., cumulative ROC-AUC, mean change direction, unadjusted p value, coefficient, and odds ratio) of this three-analyte predictive risk-score panel are given in Figure 6A. This predictive risk-score panel has a discriminating power (AUC) of 95% (confidence intervals [CIs]: 90%–100%) in the training set, and 90% (CIs: 75%–100%) in the independent testing set (Figures 6B and 6C). A comparison of AUC, accuracy, sensitivity, and specificity of this prognostic between the two sets is presented in Figure 6D.

Correlation analysis of the three analytes from the training set was further carried out, revealing that two lipids (SM C16:0 and DAG 16:0/16:0) have a moderate negative correlation ( $-0.5 > r > -0.8$ ) (Figure 6E). Additionally, the 2 h OGTT was found to have a weak positive relationship with SM C16:0 ( $0.5 > r > 0.25$ ) and a weak negative correlation with DAG 16:0/16:0 ( $0 > r > -0.25$ ). Since 2 h OGTT and SM C16:0 are weakly correlated, these two variables represent two subgroups of the population with marginal overlap. Building on this, the addition (i.e., looping) of SM C16:0 with 2 h OGTT increased ROC-AUC sharply from 69% to 91% (Figure 6A). Furthermore, since DAG 16:0/16:0 has a weak to moderate negative correlation with 2 h OGTT and SM C16:0, respectively, it may loop in an additional subgroup of the population. As before, looping of DAG 16:0/16:0 increased ROC-AUC further from 91% to 95% (Figure 6A). This simple prognostic was shown to have notably high discriminating power. Including an analyte from the sphingolipid family (i.e., SM C16:0) considerably increased the prognostic power of the well-established clinical variable (i.e., 2 h OGTT). The integration of SM C16:0 into the prognostic panel also suggests the importance of sphingolipids in T2D pathogenesis.



**Figure 5. The predictive power of clinical variables**

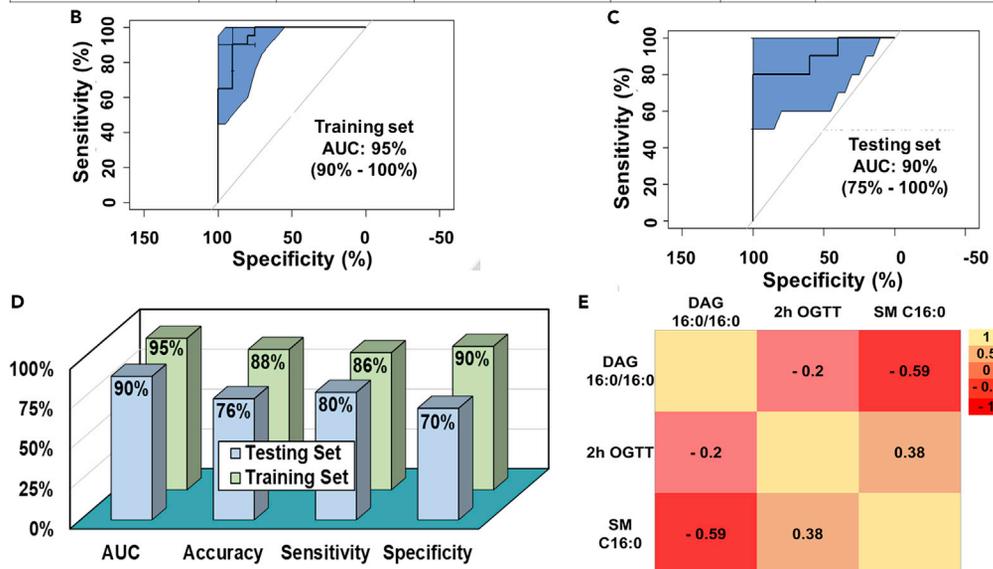
(A) A diagram of different kinds of analysis with clinical variables using 60 Hispanic participants. (B–E) Univariate analysis for clinical variables [i.e., Fasting Plasma Glucose (FPG), 2h-OGTT, total triglyceride (Total TAG), and HOMA-IR, respectively] using logistic regression. (F) Multiple logistic regression for clinical variables using sMLR. (G–J) The box-whisker plots for clinical variables [i.e., Fasting Plasma Glucose (FPG), 2h-OGTT, total triglyceride (Total TAG), and HOMA-IR, respectively]. The bottom and top of the box are the minimum and maximum, respectively, and the central band is the median (Q2 or 50th percentile). The bottom whisker is located within 1.5 IQR of the lower quartile, and the upper whisker is located within 1.5 IQR of the upper quartile. Outliers are presented outside of whiskers. Two sample t-tests for FPG and 2 h OGTT and Wilcoxon rank-sum tests for total TAG and HOMA-IR were carried out. Unadjusted p values: \*\*p < 0.01, and \*\*\*p < 0.001. Error bars present mean ± SEM.

## DISCUSSION

This study intended to gain a better understanding of the early-stage T2D pathophysiology, excluding major confounding factors (i.e., race, age, BMI, sex, and blood glucose). To this end, we designed a 1:1 pair-matched (i.e., race, age, BMI, sex, and glucose tolerance) case-control study that mined metabolomic (Allalou et al., 2016) and lipidomic (Khan et al., 2019b; Lai et al., 2020) datasets of the SWIFT cohort. We identified downregulated sphingolipid metabolism as the major molecular pathway eclipsing upregulated fatty acid biosynthesis in the Hispanic cohort (Figure 2A). Our findings support a previously identified association between sphingolipids with T2D development (Buschard et al., 2005; Khan et al., 2019b; Lai et al., 2020). Experimentally, we show that downregulated sphingolipid metabolism in mice affected insulin sensitivity (Figure 3C) and caused pancreatic β cell dysfunction in both isolated islets and pancreatic β cells (Figure 4). These observations support a role for specific lipid dysmetabolism (i.e., downregulation of sphingolipid metabolism) as a causal factor in early-stage T2D pathophysiology (Alexaki et al., 2017). Holm et al. recently showed that diminished sphingolipids are also observed in T1D through immunostaining of human pancreatic biopsy samples from the DiviD study (Holm et al., 2018; Greenhill, 2018). Interestingly, in the present study, merging metabolomics and diabetes GWAS data identified five SNPs for T2D linked to the sphingolipid metabolism pathway (Figure 2D). These SNPs were associated with two genes of ceramide biosynthesis, CERS2 and CERS4. The CERS2-specific SNP, rs267738, was linked to elevated

A

Predictive panel	Trend	Unadjusted p-value	Co-efficient ± SE	Odds ratio	Looping	cAUC in training
1. 2h OGTT	Up	<0.001*	0.73 ± 0.15	5.4 ± 1.4	1	69% (52%-87%)
2. SM C16:0	Down	< 0.001	0.50 ± 0.17	3.2 ± 1.5	1+2	91% (81%-100%)
3. DAG16:0/16:0	Up	< 0.0001	0.32 ± 0.10	2.1 ± 1.3	1+2+3	95% (90%-100%)



**Figure 6. The Discovery of a Prognostic Risk Score with 95% Discriminating Power**

(A) This table describes both the individual odds ratio of each component of this prognostic risk score panel as well as their collective discriminating strength in cumulative ROC-AUC (through a looping experiment one by one) using a training set (20 pairs). The trend indicates their mean regulation direction (up/down). The unadjusted p values were calculated using non-parametric Wilcoxon-Mann-Whitney test (\* indicates significant parametric unadjusted p value). The odds ratios (with 95% CIs) were calculated from estimated co-efficient (with  $\pm$  SEM) of each component in stepwise (both ways) multiple logistic regression.

(B) A 95% ROC-AUC of this prognostic risk score in the training set.

(C) A 90% ROC-AUC of this prognostic risk score in an independent testing set.

(D) A comparison of AUC, accuracy, sensitivity, and specificity of this prognostic risk score between the training and testing sets.

(E) Pearson's correlation analysis on the components of this prognostic risk score using the training dataset.

HbA1c ( $p = 3 \times 10^{-9}$ ) (Wheeler et al., 2017). The four other SNPs belong to CERS4. Among them, rs11666913 and rs2100944 were associated with increased levels of circulating fatty acids (triglycerides) (Lemaitre et al., 2015), whereas rs12610250 and rs7258249 were linked to decreased circulating sphingolipid levels (Draisma et al., 2015), correlating with our metabolomics data in the SWIFT cohort. Although the pregnancy timeline was beyond the scope of this study, it was able to identify potential genetic predisposition for altered sphingolipids using this innovative bioinformatics method. Since these alternations in sphingolipid metabolism are associated with pancreatic  $\beta$  cell dysfunction and insulin resistance, they may potentially disrupt glucose homeostasis during pregnancy. It is important to note that, since the GWAS results are independent of the Hispanic population studied here, downregulated sphingolipid metabolism may be a more general phenomenon for T2D development.

In *in vivo* studies in mice, although short-term myriocin-treated animals did not exhibit a so-called diabetes phenotype, they possessed some pre-T2D stage-like pathogenesis, including late-phase insulin intolerance (suggesting problems in hepatic glucose uptake/gluconeogenesis) (Figure 3C), significantly higher blood cholesterol and cholesterol/HDL ratios (Figures 3D and 3E), and a significant decrease of pancreatic  $\beta$  cell area (Figures 3I–3K). With glucose tolerance maintained under the conditions tested,

this early-stage T2D phenotype did not lead to a prediabetes/T2D-like condition. However, the reduction of sphingolipid metabolism under myriocin treatment may accelerate the emergence of factors like  $\beta$  cell dysfunction and insulin resistance that lead to overt T2D over time. We have previously shown that short-term inhibition of ceramide biosynthesis decreases pancreatic  $\beta$  cell area *in vivo* using fumonisin B1 (i.e., an inhibitor of ceramide synthase, another strategy for ceramide biosynthesis inhibition) (Khan et al., 2019b). In keeping with our results, several knockout mouse studies have shown that the inhibition of ceramide biosynthesis causes both insulin resistance and the disruption of glucose homeostasis associated with fatty liver (Alexaki et al., 2017; Lee et al., 2017; Park et al., 2013). A recent study has shown that downregulation of very-long-chain sphingolipids causes insulin-resistance and liver steatosis during a high-fat-diet intervention in CERS2 haploinsufficient mice (Raichur et al., 2014). Interestingly, a follow-up study showed that CERS2 inhibition could activate the fatty acid biosynthesis regulatory mechanisms by upregulating CERS6 (Kim et al., 2019).

Both disruption of pancreatic  $\beta$  cell function (i.e., glucose responsiveness) and attenuation of  $\beta$  cell mass are crucial factors for the onset and progression of T2D (Park and Woo, 2019; Maedler and Donath, 2004). We have previously shown that 24 h of ceramide biosynthesis inhibition causes pancreatic  $\beta$  cell dysfunction (Khan et al., 2019b). In the case of sphingosine-1-phosphate (S1P) biosynthesis, Stanford et al. have previously induced murine pancreatic  $\beta$  cell dysfunction with SKI, an S1P inhibitor (Cantrell Stanford et al., 2012). Additionally, Hong et al. observed both decreased insulin secretion and reduced cell survival when Min6 cells were treated with 15  $\mu$ M SKI for 48 h (Hong et al., 2018). In the case of sphingomyelin biosynthesis, Subathra et al. found it to be essential for insulin secretion in rat  $\beta$  cells (i.e., INS-1) using a high concentration of D609 under 4 h of incubation (Subathra et al., 2011); in addition, Kavishwar et al. showed a significant loss of sphingolipid patch regions (sphingomyelin-rich areas associated with cholesterol) of  $\beta$  cells in both type 1 and type 2 diabetic subjects (Kavishwar and Moore, 2013). Altogether these studies highlight the potential importance of different sphingolipids in pancreatic  $\beta$  cell function; however, they employed different cell models and protocols. Such discrepancies make it a challenge to understand the roles of different sphingolipid biosynthetic processes (i.e., ceramide biosynthesis, S1P biosynthesis, sphingomyelin biosynthesis) in health and disease. In this present study, we kept the inhibition period constant (i.e., 48 h) for all treatments and used glucose-responsive Min6K8 cells and isolated mouse pancreatic islets. We have found that under the conditions studied, both ceramide biosynthesis and S1P biosynthesis are important for  $\beta$  cell function, whereas sphingomyelin biosynthesis appears to be important for  $\beta$  cell viability (Figure 4).

To uncover a metabolic prognostic for Hispanic women with GDM with strong discrimination power (AUC), we used datasets employing metabolomics/lipidomics alone or in combination with common clinical variables. This study discovered a unique three-analyte predictive risk-score panel with 95% discriminating power (AUC) using the sMLR algorithm under internal validation through bootstrapping (Figures 6A–6C). This prognostic was then validated using an independent testing set where its discriminating power (AUC) was also highly robust (90%) (Figure 6D). Previously, Urpi-Sarda et al. identified an eight-metabolite prognostic with 96.4% ROC-AUC using urine samples from 154 multi-race participants of the PREDIMED trial, an observational, longitudinal study. In addition, they showed that this prognostic was better than glucose, which had 89.9% ROC-AUC (Urpi-Sarda et al., 2019). In a previous study, we identified a more complex seven-lipid prognostic panel with 92% ROC-AUC from fasting plasma samples of 140 mixed-race participants of the SWIFT study. This seven-analyte prognostic exhibited better predictive power than OGTT, which had 71% ROC-AUC (Khan et al., 2019b), but with seven metabolites needed, it is not considered practical in application. The main differences between the identified prognostic of the current study and all other metabolomic prognostics including our previous studies (Allalou et al., 2016; Leong et al., 2018; Papandreou et al., 2019; Yang et al., 2018; Merino et al., 2018; Khan et al., 2019b) include its race specificity (i.e., precision medicine approach by design), minimal size with outstanding discrimination power (i.e., three analytes exhibited 95% ROC-AUC), and integration with a known clinical variable (i.e., OGTT). Although this prognostic meets the criteria of the phase 1 stage of biomarker discovery (Khan et al., 2019a), its stability beyond 2 years has not yet been evaluated. Notably, one of the three analytes was a sphingolipid (i.e., SM C16:0) whose addition enhanced the discriminating power of the predictive panel substantially, from 69% to 91% (Figure 6A). Hilvo et al. have recently shown that the Cer(d18:1/18:0)/Cer(d18:1/16:0) ratio is an independent predictive biomarker for incident diabetes (Hilvo et al., 2018). This observation further suggests the importance of sphingolipids in T2D pathogenesis.

In conclusion, downregulated sphingolipid metabolism was a critical indicator of early-stage T2D pathophysiology within this Hispanic GDM cohort. The centrality of sphingolipids in the transition of GDM to T2D is indicated by observations that downregulated sphingolipid metabolism in mouse and cell models contributes to insulin resistance and pancreatic  $\beta$  cell dysfunction. The early disruption of sphingolipid metabolism prior to the onset of T2D may result from a genetic predisposition, demonstrating that mapping metabolomics and GWAS data in tandem is a powerful tool to identify potential genetic markers of T2D risk. We have also shown that the inclusion of a specific sphingolipid metabolite in the early post-partum period with standard glucose testing enhances its ability to predict future T2D diabetes risk.

### Limitations of the Study

Our work aimed to present a proof-of-concept study focusing on the role of sphingolipid metabolism in the pathophysiological context of T2D development. Although we used a case-control study design, we did not use a propensity score matching strategy in pair matching of participants owing to the unavailability of some samples. Instead of propensity score matching, we had used a strategy described in the [Transparent Methods](#) section. This strategy has less immunity against biased participant selection, which may result in a higher degree of confounding factor influence. To estimate the effect of confounding factors, we ran PCA on the metabolomics datasets of the participants. PCA found a minimum effect from confounding factors. Although the identified pathophysiology may be interpreted for generalized T2D, the discovered biomarkers are based on a more uniform ethnicity. Additionally, larger sample size studies may find other associated factors with smaller effect sizes. Lastly, our identified biomarkers need to be further validated in other observational cohorts.

### Resource Availability

#### Lead Contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the Lead Contact, Michael B. Wheeler ([Michael.wheeler@utoronto.ca](mailto:Michael.wheeler@utoronto.ca)).

#### Materials Availability

All unique materials generated in this study are available from the Lead Contact.

#### Data and Code Availability

Both metabolomics and lipidomics data have been deposited in Harvard Dataverse: <https://doi.org/10.7910/DVN/Q4PEQR>. Codes are available in [Data S1](#) and [S2](#) of [Supplemental Information](#). All of these will be accessible to readers upon publication of this manuscript.

## METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101566>.

## ACKNOWLEDGMENTS

These studies are supported by the Canadian Institutes of Health Research (CIHR), FRN 143219 (M.B.W.), the National Institute of Child Health and Human Development (NICHD)R01 HD050625 (E.P.G.), the National Institute of Diabetes and Digestive and Kidney Diseases R01DK118409 (E.P.G.), and Janssen Pharmaceuticals 430086739 (M.B.W., E.P.G.). S.R.K. is supported by a Diabetes Canada post-doctoral fellowship followed by a Banting and Best post-doctoral fellowship. B.J.C. is supported by a Tier II Canada Research Chair. A.O. is supported by a CGS-M NSERC scholarship. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## AUTHOR CONTRIBUTIONS

S.R.K. and M.B.W. conceptualized and designed the research work. E.P.G. designed, recruited, and conducted all in-person clinical research examinations, administered the multiple 2 h 75 g research OGTTs to measure glucose and insulin annually post delivery, and performed behavioral, sociodemographic, and other surveys under standardized research protocols in the SWIFT Study participants ([Table S1](#)), and

provided human fasting plasma stored samples for the current analysis. All predictive analytics and bioinformatics were performed by S.R.K. and supervised by B.J.C. The metGWAS was conceptualized by S.R.K., performed by S.R.K. and A.O., supervised by B.J.C. S.R.K. conducted all *in vivo* and *in vitro* studies. The manuscript was written by S.R.K. and edited by Y.M., M.B.W., E.P.G., B.J.C., and A.O. All authors assisted in reviewing the manuscript and gave final approval of the version to be published.

## DECLARATION OF INTERESTS

M.B.W. and E.P.G. have declared a research grant from Janssen Pharmaceuticals Company. They have no other financial interests to declare. The other authors declare that no competing interests exist.

Received: May 19, 2020

Revised: July 20, 2020

Accepted: September 11, 2020

Published: October 23, 2020

## REFERENCES

- Abdul-Ghani, M.A., Lyssenko, V., Tuomi, T., DeFronzo, R.A., and Groop, L. (2009). Fasting versus postload plasma glucose concentration and the risk for future type 2 diabetes, results from the Botnia study. *Diabetes Care* 32, 281–286.
- Akirov, A., Grossman, A., Shochat, T., and Shimon, I. (2017). Mortality among hospitalized patients with hypoglycemia: insulin related and noninsulin related. *J. Clin. Endocrinol. Metab.* 102, 416–424.
- Alexaki, A., Clarke, B.A., Gavriloova, O., Ma, Y., Zhu, H., Ma, X., Xu, L., Tuymetova, G., Larman, B.C., Allende, M.L., et al. (2017). De novo sphingolipid biosynthesis is required for adipocyte survival and metabolic homeostasis. *J. Biol. Chem.* 292, 3929–3939.
- Allalou, A., Nalla, A., Prentice, K.J., Liu, Y., Zhang, M., Dai, F.F., Ning, J.X., Osborne, L.R., Cox, B.J., Gunderson, E.P., and Wheeler, M.B. (2016). A predictive metabolic signature for the transition from gestational diabetes to type 2 diabetes. *Diabetes* 65, db151720.
- Amiel, S.A., Aschner, P., Childs, B., Cryer, P.E., De Galan, B.E., Frier, B.M., Gonder-Frederick, L., Heller, S.R., Jones, T., Khuntji, K., et al. (2019). Hypoglycaemia, cardiovascular disease, and mortality in diabetes: epidemiology, pathogenesis, and management. *Lancet Diabetes Endocrinol.* 7, 385–396.
- Arora, G.P., Åkerlund, M., and Brøns, C. (2019). Phenotypic and genotypic differences between Indian and Scandinavian women with gestational diabetes mellitus. *J. Intern. Med.* 286, 192–206.
- Bamshad, M. (2005). Genetic influences on health: does race matter? *JAMA* 294, 937–946.
- Buschard, K., Fredman, P., Bøg-Hansen, E., Blomqvist, M., Hedner, J., Råstam, L., and Lindblad, U. (2005). Low serum concentration of sulfatide and presence of sulfated lactosylceramide are associated with Type 2 diabetes. The Skaraborg Project. *Diabet. Med.* 22, 1190–1198.
- Cantrell Stanford, J., Morris, A.J., Sunkara, M., Popa, G.J., Larson, K.L., and Özcan, S. (2012). Sphingosine 1-phosphate (S1P) regulates glucose-stimulated insulin secretion in pancreatic beta cells. *J. Biol. Chem.* 287, 13457–13464.
- Cheng, Y.J., Kanaya, A.M., Araneta, M.R.G., Saydah, S.H., Kahn, H.S., Gregg, E.W., Fujimoto, W.Y., and Imperatore, G. (2019). Prevalence of diabetes by race and ethnicity in the United States, 2011–2016. *JAMA* 322, 2389.
- Ding, M., Chavarro, J., Olsen, S., Lin, Y., Ley, S.H., Bao, W., Rawal, S., Grunnet, L.G., Thuesen, A.C.B., Mills, J.L., et al. (2018). Genetic variants of gestational diabetes mellitus: a study of 112 SNPs among 8722 women in two independent populations. *Diabetologia* 61, 1758–1768.
- Draisma, H.H.M., Pool, R., Kobl, M., Jansen, R., Petersen, A.-K., Vaarhorst, A.A.M., Yet, I., Haller, T., Demirkan, A., Esko, T., et al. (2015). Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat. Commun.* 6, 7208.
- Greenhill, C. (2018). Sphingolipids involved in T1DM. *Nat. Rev. Endocrinol.* 14, 381.
- Gunderson, E.P., Hedderson, M.M., Chiang, V., Crites, Y., Walton, D., Azevedo, R.A., Fox, G., Elmasian, C., Young, S., Salvador, N., et al. (2012). Lactation intensity and postpartum maternal glucose tolerance and insulin resistance in women with recent GDM: the SWIFT cohort. *Diabetes Care* 35, 50–56.
- Gunderson, E.P., Hurston, S.R., Ning, X., Lo, J.C., Crites, Y., Walton, D., Dewey, K.G., Azevedo, R.A., Young, S., Fox, G., et al. (2015). Lactation and progression to type 2 diabetes mellitus after gestational diabetes mellitus. *Ann. Intern. Med.* 163, 889.
- Gunderson, E.P., Lewis, C.E., Tsai, A.-L., Chiang, V., Carnethon, M., Quesenberry, C.P., and Sidney, S. (2007). A 20-year prospective study of childbearing and incidence of diabetes in young women, controlling for glycemia before conception: the coronary artery risk development in young adults (CARDIA) study. *Diabetes* 56, 2990–2996.
- Gunderson, E.P., Matias, S.L., Hurston, S.R., Dewey, K.G., Ferrara, A., Quesenberry, C.P., Lo, J.C., Sternfeld, B., and Selby, J.V. (2011). Study of women, infant feeding, and type 2 diabetes mellitus after GDM pregnancy (SWIFT), a prospective cohort study: methodology and design. *BMC Public Health* 11, 952.
- Hilvo, M., Salonurmi, T., Havulinna, A.S., Kauhanen, D., Pedersen, E.R., Tell, G.S., Meyer, K., Teerinemi, A.M., Laatikainen, T., Jousilahti, P., et al. (2018). Ceramide stearic to palmitic acid ratio predicts incident diabetes. *Diabetologia* 61, 1424–1434.
- Hod, M., Kapur, A., and McIntyre, H.D. (2019). Evidence in support of the International Association of Diabetes in Pregnancy study groups' criteria for diagnosing gestational diabetes mellitus worldwide in 2019. *Am. J. Obstet. Gynecol.* 227, 109–116.
- Holm, L.J., Krogvold, L., Hasselby, J.P., Kaur, S., Claessens, L.A., Russell, M.A., Mathews, C.E., Hanssen, K.F., Morgan, N.G., Koeleman, B.P.C., et al. (2018). Abnormal islet sphingolipid metabolism in type 1 diabetes. *Diabetologia* 61, 1650–1661.
- Hong, S.W., Lee, J., Kwon, H., Park, S.E., Rhee, E.J., Park, C.Y., Oh, K.W., Park, S.W., and Lee, W.Y. (2018). Deficiency of sphingosine-1-phosphate reduces the expression of prohibitin and causes  $\beta$ -cell impairment via mitochondrial dysregulation. *Endocrinol. Metab.* (Seoul, Korea) 33, 403–412.
- Hunt, K.J., and Schuller, K.L. (2007). The Increasing prevalence of diabetes in pregnancy. *Obstet. Gynecol. Clin. North Am.* 34, 173–199.
- Huvinen, E., Eriksson, J.G., Koivusalo, S.B., Grotenfelt, N., Tiitinen, A., Stach-Lempinen, B., and Rono, K. (2018a). Heterogeneity of gestational diabetes (GDM) and long-term risk of diabetes and metabolic syndrome: findings from the RADIEL study follow-up. *Acta Diabetol.* 55, 493–501.
- Huvinen, E., Eriksson, J.G., Stach-Lempinen, B., Tiitinen, A., and Koivusalo, S.B. (2018b). Heterogeneity of gestational diabetes (GDM) and challenges in developing a GDM risk score. *Acta Diabetol.* 55, 1251–1259.
- Kavishwar, A., and Moore, A. (2013). Sphingomyelin patches on pancreatic beta-cells are indicative of insulin secretory capacity. *J. Histochem. Cytochem.* 61, 910–919.

- Khan, S.R., Manialawy, Y., Wheeler, M.B., and Cox, B.J. (2019a). Unbiased data analytic strategies to improve biomarker discovery in precision medicine. *Drug Discov. Today* 24, 1735–1748.
- Khan, S.R., Mohan, H., Liu, Y., Batchuluun, B., Gohil, H., Al Rijjal, D., Manialawy, Y., Cox, B.J., Gunderson, E.P., and Wheeler, M.B. (2019b). The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes. *Diabetologia* 62, 687–703.
- Kim, C., Newton, K.M., and Knopp, R.H. (2002). Gestational diabetes and the incidence of type 2 diabetes. *Syst. Rev.* 25, 1862–1868.
- Kim, Y.-R., Lee, E.-J., Shin, K.-O., Kim, M.H., Pewzner-Jung, Y., Lee, Y.-M., Park, J.-W., Futerman, A.H., and Park, W.-J. (2019). Hepatic triglyceride accumulation via endoplasmic reticulum stress-induced SREBP-1 activation is regulated by ceramide synthases. *Exp. Mol. Med.* 51, 1–16.
- Koning, S.H., Hoogenberg, K., Lutgers, H.L., Van Den Berg, P.P., and Wolffenbuttel, B.H.R. (2016). Gestational Diabetes Mellitus: current knowledge and unmet needs. *J. Diabetes* 8, 770–781.
- Lai, M., Al Rijjal, D., Röst, H.L., Dai, F.F., Gunderson, E.P., and Wheeler, M.B. (2020). Underlying dyslipidemia postpartum in women with a recent GDM pregnancy who develop type 2 diabetes. *eLife* 9, e59153.
- Lee, S.-Y., Lee, H.-Y., Song, J.-H., Kim, G.-T., Jeon, S., Song, Y.-J., Lee, J.S., Hur, J.-H., Oh, H.H., Park, S.-Y., et al. (2017). adipocyte-specific deficiency of de novo sphingolipid biosynthesis leads to lipodystrophy and insulin resistance. *Diabetes* 66, 2596–2609.
- Lemaitre, R.N., King, I.B., Kabagambe, E.K., Wu, J.H.Y., Mcknight, B., Manichaikul, A., Guan, W., Sun, Q., Chasman, D.I., Foy, M., et al. (2015). Genetic loci associated with circulating levels of very long-chain saturated fatty acids. *J. Lipid Res.* 56, 176–184.
- Leong, A., Daya, N., Porneala, B., Devlin, J.J., Shiffman, D., Mcphaul, M.J., Selvin, E., and Meigs, J.B. (2018). Prediction of type 2 diabetes by hemoglobin a<sub>1c</sub> in two community-based cohorts. *Diabetes Care* 41, 60–68.
- Maedler, K., and Donath, M.Y. (2004). Beta-cells in type 2 diabetes: a loss of function and mass. *Horm. Res.* 62 (Suppl 3), 67–73.
- Merino, J., Leong, A., Liu, C.-T., Porneala, B., Walford, G.A., Von Grotthuss, M., Wang, T.J., Flannick, J., Dupuis, J., Levy, D., et al. (2018). Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. *Diabetologia* 61, 1315–1324.
- Papandreou, C., Bulló, M., Ruiz-Canela, M., Dennis, C., Deik, A., Wang, D., Guasch-Ferré, M., Yu, E., Razquin, C., Corella, D., et al. (2019). Plasma metabolites predict both insulin resistance and incident type 2 diabetes: a metabolomics approach within the Prevención con Dieta Mediterránea (PREDIMED) study. *Am. J. Clin. Nutr.* 109, 626–634.
- Park, J.-W., Park, W.-J., Kuperman, Y., Boura-Halfon, S., Pewzner-Jung, Y., and Futerman, A.H. (2013). Ablation of very long acyl chain sphingolipids causes hepatic insulin resistance in mice due to altered detergent-resistant membranes. *Hepatology* 57, 525–532.
- Park, Y.J., and Woo, M. (2019). Pancreatic  $\beta$  cells: gatekeepers of type 2 diabetes. *J. Cell Biol.* 218, 1094–1095.
- Raichur, S., Wang, Siew T., Chan, Puck W., Li, Y., Ching, J., Chaurasia, B., Dogra, S., Öhman, Miina K., Takeda, K., Sugii, S., et al. (2014). CerS2 haploinsufficiency inhibits  $\beta$ -oxidation and confers susceptibility to diet-induced steatohepatitis and insulin resistance. *Cell Metab.* 20, 687–695.
- Ratner, R.E. (2007). Prevention of type 2 diabetes in women with previous gestational diabetes. *Diabetes Care* 30, S242–S245.
- Subathra, M., Qureshi, A., and Luberto, C. (2011). Sphingomyelin synthases regulate protein trafficking and secretion. *PLoS one* 6, e23644.
- Tirosh, A., Shai, I., Bitzur, R., Kochba, I., Tekes-Manova, D., Israeli, E., Shochat, T., and Rudich, A. (2008). Changes in triglyceride levels over time and risk of type 2 diabetes in young men. *Diabetes Care* 31, 2032–2037.
- Udler, M.S., Kim, J., Von Grotthuss, M., Bonàs-Guarch, S., Cole, J.B., Chiou, J., Christopher D. Anderson on behalf of METASTROKE and the ISGC, The, I., Boehnke, M., Laakso, M., et al. (2018). Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med.* 15, e1002654.
- Urpi-Sarda, M., Almanza-Aguilera, E., Llorach, R., Vázquez-Fresno, R., Estruch, R., Corella, D., Sorli, J.V., Carmona, F., Sanchez-Pla, A., Salas-Salvadó, J., and Andres-Lacueva, C. (2019). Non-targeted metabolomic biomarkers and metabolotypes of type 2 diabetes: a cross-sectional study of PREDIMED trial participants. *Diabetes Metab.* 45, 167–174.
- Wheeler, E., Leong, A., Liu, C.-T., Hivert, M.-F., Strawbridge, R.J., Podmore, C., Li, M., Yao, J., Sim, X., Hong, J., et al. (2017). Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS Med.* 14, e1002383.
- Yang, S.J., Kwak, S.-Y., Jo, G., Song, T.-J., and Shin, M.-J. (2018). Serum metabolite profile associated with incident type 2 diabetes in Koreans: findings from the Korean Genome and Epidemiology Study. *Sci. Rep.* 8, 8207.
- Zhu, Y., and Zhang, C. (2016). Prevalence of gestational diabetes and risk of progression to type 2 diabetes: a global perspective. *Curr. Diab. Rep.* 16, 7.

iScience, Volume 23

## **Supplemental Information**

### **Diminished Sphingolipid Metabolism, a Hallmark of Future Type 2 Diabetes Pathogenesis, Is Linked to Pancreatic $\beta$ Cell Dysfunction**

**Saifur R. Khan, Yousef Manialawy, Andreea Obersterescu, Brian J. Cox, Erica P. Gunderson, and Michael B. Wheeler**

## TRANSPARENT METHODS

### Study population and study design

The SWIFT study is a prospective, racially and ethnically diverse observational research cohort of 1,035 women with GDM (75% minority; 8% black, 24% white, 34% Hispanic, 30% Asian and 2% Other) recruited during pregnancy within the Kaiser Permanente Northern California (KPNC) healthcare system from 2008 to 2011. The study participants consented to three in-person research visits and met study eligibility criteria: (1) Age 20 to 45 years at delivery, (2) Received prenatal care at Kaiser medical facilities and delivered at Kaiser hospitals, (3) GDM pregnancy diagnosed by 3-hr 100 g OGTT using Carpenter and Coustan's criteria, (4) Delivered a singleton, live birth  $\geq$  35 weeks gestation, (5) No pre-existing diabetes (DM) or other serious medical conditions prior to index GDM pregnancy, and other criteria as previously described (Gunderson et al., 2012; Gunderson et al., 2015; Gunderson et al., 2011). KPNC prenatal standardized treatments for GDM included dietary therapy, oral diabetes medication, and insulin treatment (**Supplemental Table-1**). All SWIFT Study research protocols were approved by the KPNC institutional review board requirement, and women provided written informed consent at the 6–9 weeks postpartum research visit (baseline) and attended annual follow-up visits for 2 years. These visits included research 2h 75 g OGTTs, completion of surveys, and measurement of anthropometry and body composition by trained research staff. Blood-glucose and insulin were measured in EDTA-treated plasma by the Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington, Seattle, WA. Other clinical variables, including GDM severity from the 3-h 100 g OGTT, were obtained from the electronic medical records. Details of the SWIFT Study and participant characteristics have been described elsewhere (Gunderson et al., 2012; Gunderson et al., 2015; Gunderson et al., 2011).

A nested 1:1 pair-matched case-control study was designed using a precision medicine approach, where a Hispanic population within the SWIFT cohort was targeted to avoid the influence of race/ethnicity-related bias due to genetic predisposition (**Figure-1A**). At baseline (6-9 weeks postpartum), all participants (60 women) were confirmed not to have type 2 diabetes using the 2h 75g OGTT. Trained research staff collected and processed the fasting and 2h OGTT plasma samples at each visit, which were stored  $-80^{\circ}\text{C}$  for future biochemical analyses (i.e., metabolomics and lipidomics). In the follow-up examinations (1<sup>st</sup> year and 2<sup>nd</sup>-year post-baseline), all

participants were reassessed for type 2 diabetes using a 2h 75g OGTT. Women who met the ADA criteria for diabetes diagnosis at follow-up examinations were designated as “case” (also referred to as T2D interchangeably). Women who did not meet the ADA diagnostic criteria at follow-up examinations were labeled as “control.” We selected 30 Hispanic cases (i.e., progressed to incident type 2 diabetes after baseline), and then selected matched pairs (1:1) of 30 Hispanic controls (i.e., without progression to type 2 diabetes after baseline) during the 2-year follow-up (up to 32 months post-baseline). Matching variables included maternal age ( $\pm 2$  years), pre-pregnancy BMI ( $\pm 0.96$  kg/m<sup>2</sup>), and glucose intolerance at 6-9 weeks postpartum (normal or impaired glucose tolerance based on ADA criteria; completely matched).

## **Broad-spectrum omics**

### ***Targeted metabolomics***

Single-blind targeted metabolomic analyses were performed on baseline fasting plasma samples at the Analytical Facility for Bioactive Molecules (The Hospital for Sick Children, Toronto, Canada) where a total of 188 metabolites were assayed using p180 AbsoluteIDQ™ plate technology according to the manufacturer’s instructions (Biocrates Life Sciences AG, Austria). The details of the metabolomic analyses are described elsewhere (Allalou et al., 2016).

### ***Targeted lipidomics***

Single-blind targeted-lipidomic analyses of 873 lipid species were carried out on baseline fasting plasma samples at Metabolon (Morrisville, NC, USA). This lipidomics data has been taken from our recently published multiracial lipidomics papers where the details of lipidomics analysis are described (Khan et al., 2019b; Lai et al., 2020).

## **Data preparation and statistical analysis**

Analytes (variables) with more than 5% of values missing were removed. Otherwise, the missing values were estimated as half of the lowest positive values. Both datasets were merged together with some clinical variables (i.e., baseline fasting plasma glucose [FPG], 2h glucose in OGTT [2h OGTT], HOMA-IR, and total fasting triglycerides). A quantile normalization followed by log transformation were carried out to get a normal/semi-normal data distribution. Principal component analysis (PCA) was carried out to determine the influence of confounding factors within the dataset (Khan et al., 2019a). A partial least squares-discriminant analysis (PLS-DA)

was carried out to understand the class-separation using accuracy,  $R^2$ , and  $Q^2$  calculation under 10-fold cross-validation at the baseline.  $Q^2$  is used to determine significant differences between case and control (Khan et al., 2019a). All operations were carried out using MetaboAnalyst 4.0 (<https://www.metaboanalyst.ca/>) using default settings unless otherwise specified.

### **Systems biology approach: differential expression and pathway analysis**

Significant differentially expressed analytes between case and control were found using the Wilcoxon-Mann-Whitney test (a non-parametric test,  $\alpha$  value set at  $p < 0.05$ ) followed by false discovery rate (FDR) analysis ( $\alpha$  value set at  $p < 0.1$ , due to the small dataset). Significant over-representation KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were identified with their pathway importance using a cut-off of  $FDR < 0.05$  under a hypergeometric test and relative-betweenness centrality algorithm using the Homo sapiens KEGG pathway library. All operations were carried out using MetaboAnalyst 4.0 (<https://www.metaboanalyst.ca/>) with default settings (unless stated otherwise).

### **Metabolite - Genome-Wide Association Study (met-GWAS)**

Significantly altered metabolites of the most crucial pathways found in systems biology approaches were used to identify the interacting enzymes using MBRole2.0 (López-Ibáñez et al., 2016). If any of these enzymes contradicted the directions of the metabolite profile, it was excluded. For example, sphingomyelinase is an interacting enzyme of sphingomyelins. It breaks down sphingomyelins to ceramides. Since ceramides were one of our downregulated metabolites in this study, the action of sphingomyelinase was contradicting, so we excluded from our list. We also created a curated T2D-related gene database using the GWAS catalogue (Buniello et al., 2019). To identify the number of unique genes with known T2D pathophysiology-related risk-factors, we used the following search criteria in GWAS catalogue: Type II Diabetes, fasting blood glucose measurement, glucose tolerance test, A1C measurement, HOMA-B, HOMA-IR, fasting blood insulin measurement, insulin secretion rate measurement, insulin response measurement, insulinogenic index measurement, fatty acid measurement, sphingolipid measurement, sphingomyelin measurement, obesity, leucine measurement, hydroxy-leucine measurement, and valine measurement. These key words created a list of 3463 associations, of which 108 associations were not mapped to any genes and were removed, leaving 3355 associations. The 3355 associations were mapped to 2107 unique genes (**Data S1**). From EBI, who also maintains

the GWAS catalog ([www.ebi.ac.uk/gwas/](http://www.ebi.ac.uk/gwas/)), we used 19,901 protein-coding genes and 15,779 non-coding genes as the estimate for the total number of human genes. There were nine genes mapped to the enriched metabolites for the sphingolipid metabolism pathway, of which two were in the T2D pathophysiology database we created with a size of 2107 genes out of an estimated 35680 genes in the genome. A significance test (enrichment of T2D associated genes among the interacting genes of the most critical differential pathway) was carried out using the R function `phyper()`, a hypergeometric distribution analysis, and a  $p < 0.05$  was considered significant (**Data S2**).

### **Translational studies: *in vivo* and *in vitro* studies**

#### **Animal care**

Eight-week-old C57BL/6 J male mice were obtained from Charles River (Sherbrook, QB, Canada), and were housed in the Division of Comparative Medicine facility, University of Toronto, as per the approved protocols of the Animal Care Committee and the Canadian Council of Animal Care guidelines.

#### **Intraperitoneal injections and monitoring**

Every other day, mice were given an intraperitoneal injection containing either  $0.3 \text{ mg kg}^{-1}$  Myriocin (Cayman, Michigan, USA) or vehicle (DMSO–saline [154 mmol/l NaCl]) for 3 weeks ( $n = 11$ ). At the end of each week, weight gain and blood glucose were measured.

#### **Insulin tolerance test, glucose tolerance test, and insulin ELISA**

A standardized protocol was followed for ITTs, GTTs, and insulin ELISA as described elsewhere (Khan et al., 2019b).

#### **Sphingolipid profiling and insulin staining of pancreas**

Mice were euthanized after 3 weeks of treatment to collect plasma and pancreatic tissue. Plasma samples ( $n = 3$ ) were sent to the Analytical Facility for Bioactive Molecules, SickKids, Toronto for sphingolipid profiling through LC-MS/MS. The pancreases ( $n = 7$ ) were fixed using a standard protocol (Khan et al., 2019b) and sent to the Centre for Phenogenomics (TCP), Sinai Health

System Institute, Toronto, for insulin staining. The 40× images of pancreatic slices were produced at TCP and analyzed using the Aperio ImageScope software package (Wetzlar, Germany).

### **Biochemical assays**

The serum concentrations of total cholesterol, HDL-c, and LDL-c were measured using Piccolo Xpress Chemistry Analyzer (Abbott Point of Care Inc., Princeton, NJ).

### **In vitro glucose-stimulated insulin secretion**

Glucose-stimulated insulin secretion (GSIS) was assessed as previously described (Khan et al., 2019b) in isolated male murine C57BL/6 islets *in vitro* after treatment with either 50 nmol/L myriocin (Cayman, Ann Arbor, MI, USA), 0.5 μmol/L sphingosine kinase inhibitor (Millipore Sigma, ON, Canada) or 100 μmol/L D609 (Millipore Sigma, ON, Canada) for 48 h. Glucose-dependent GSIS was also carried out using 0, 5, 10, and 20 mM glucose under treatment conditions mentioned above in Min6(K8) cells. This cell line was a gift from S. Seino [Kobe University, Kobe, Japan] and J. Miyazaki [Osaka University, Suita, Japan] (Khan et al., 2019b).

### **In vitro cell viability studies**

All inhibitors were evaluated for cytotoxicity using the standard trypan blue exclusion method in Min6(K8) cells. MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) assays were conducted for dose-dependent D609-treatments using both Min6(K8) cells and isolated mouse islets. In brief, cells were seeded in a 96-well plate in triplicate for each reaction. Following 48 hrs of D609 treatment at varying concentrations, MTT was added at 0.5mg/ml, and cells were incubated for 1 hr at 37°C. DMSO was then added, followed by 10 minutes of incubation. Thereafter, the plate was read for absorbance at 540 nm with reference absorbance at 630nm using a PHERAstar microplate reader (BMG Labtech, Ortenberg, Germany).

### **Apoptosis studies**

Apoptosis studies were carried out using IncuCyte® Caspase-3/7 green in the Incucyte® live-cell analysis system (Essen BioScience, Inc., Michigan, USA). In brief, Min6(K8) cells were seeded in 96-well format at 5,000-10,000 cells per well and placed inside the Incucyte. The next day,

Incucyte® Caspase-3/7 green reagent was added after dissolution in regular growth media at 1:1000 final dilution. Later, different inhibitors (i.e., myriocin, D609, and SKI) were added in the wells, and the caspase green fluorescence (excitation: 440 nm – 480 nm, emission: 504 nm – 544 nm) was continuously monitored for 48 hrs.

### **Biomarker discovery: predictive analytics**

An independent testing set was created by randomly (using RAND function of excel) holding out one-third of participants (10 cases and 10 controls) from the final dataset. The remaining two-thirds of participants (20 cases and 20 controls) were deployed for training of stepwise (both ways) multiple logistic regression (sMLR) using bootstrapping (internal validation) protocol and 95% confidence interval calculation (Khan et al., 2019b). Initially, variables of the training set were ranked using a random forest algorithm (i.e., 1000 trees were randomly generated), from which the top 50 variables were selected for algorithm training. These 50 variables were re-ranked using their individual ROC-AUC in logistic regression (univariate analysis), and then organized into 10 groups where each group consisted of 5 consecutively ranked variables (e.g., group-1: ranked 1 to 5, group-2: ranked 6 to 10, etc.). A stepwise (both ways) multiple logistic regression (sMLR) algorithm was applied to group-1 (composed of top 5 variables) under bootstrapping protocol and confidence interval calculation. This operation selected 2 out of 5 variables, while the rest were deemed redundant. These selected variables were then merged with group-2 (i.e., another 5 variables), and sMLR was applied to remove redundant variables with increasing predictability. After repeating this process (i.e., groupwise looping) for all 10 groups, the algorithm ultimately selected 8 out of 50 variables. A similar protocol was then adopted for these 8 variables to remove redundancy using a one-by-one looping strategy. A schematic flow-diagram of this predictive analysis is given in **Supplementary Figure-1**.

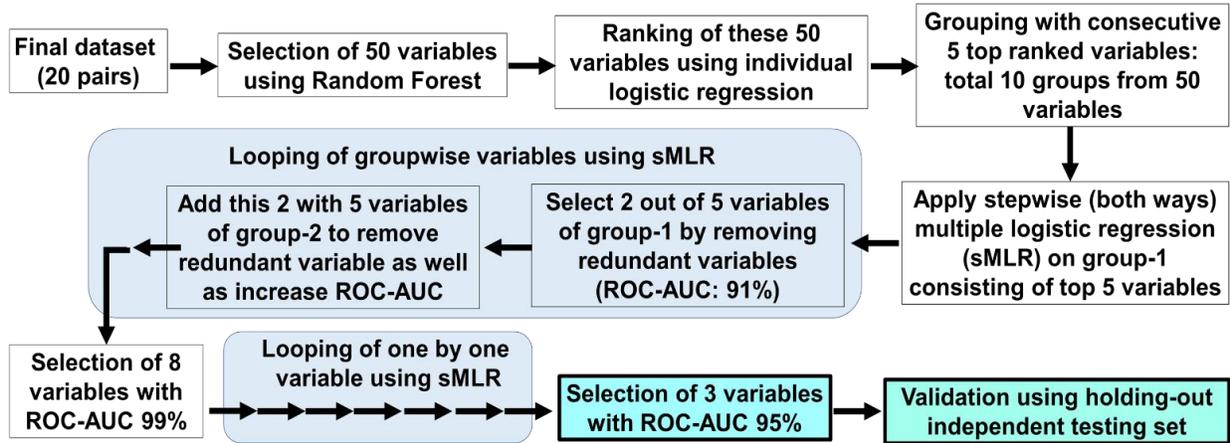
Although sMLR is a form of logistic regression that is inherently less prone to data-overfitting and bias selection, an internal validation protocol was additionally integrated here using bootstrapping and confidence intervals calculation to minimize those problems. Moreover, the final signature was further validated using the hold-out independent testing set. Furthermore, a Pearson's correlation analysis on the components of the prognostic risk score panel was carried on using the training dataset. All operations were carried out using either R-studio (Boston, MA, USA) or MetaboAnalyst 4.0 (<https://www.metaboanalyst.ca/>). The R-code for sMLR is provided elsewhere

(Khan et al., 2019b). ROC-AUC analysis was used for assessing model performance. There was no difference between model development (i.e., training) data, and model validation (i.e., testing) data in terms of participant eligibility criteria, settings, outcomes, and predictors.

## TRANSPARENT METHODS REFERENCES

- Allalou, A., Nalla, A., Prentice, K.J., Liu, Y., Zhang, M., Dai, F.F., Ning, J.X., Osborne, L.R., Cox, B.J., Gunderson, E.P., *et al.* (2016). A Predictive Metabolic Signature for the Transition from Gestational Diabetes to Type 2 Diabetes. *Diabetes*, db151720.
- Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., *et al.* (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research* 47, D1005-D1012.
- Gunderson, E.P., Hedderson, M.M., Chiang, V., Crites, Y., Walton, D., Azevedo, R.A., Fox, G., Elmasian, C., Young, S., Salvador, N., *et al.* (2012). Lactation Intensity and Postpartum Maternal Glucose Tolerance and Insulin Resistance in Women With Recent GDM: The SWIFT cohort. *Diabetes Care* 35, 50-56.
- Gunderson, E.P., Hurston, S.R., Ning, X., Lo, J.C., Crites, Y., Walton, D., Dewey, K.G., Azevedo, R.A., Young, S., Fox, G., *et al.* (2015). Lactation and Progression to Type 2 Diabetes Mellitus After Gestational Diabetes Mellitus. *Annals of Internal Medicine* 163, 889.
- Gunderson, E.P., Matias, S.L., Hurston, S.R., Dewey, K.G., Ferrara, A., Quesenberry, C.P., Lo, J.C., Sternfeld, B., and Selby, J.V. (2011). Study of Women, Infant feeding, and Type 2 diabetes mellitus after GDM pregnancy (SWIFT), a prospective cohort study: methodology and design. *BMC Public Health* 11, 952.
- Khan, S.R., Manialawy, Y., Wheeler, M.B., and Cox, B.J. (2019a). Unbiased data analytic strategies to improve biomarker discovery in precision medicine. *Drug Discov Today* 24, 1735-1748.
- Khan, S.R., Mohan, H., Liu, Y., Batchuluun, B., Gohil, H., Al Rijjal, D., Manialawy, Y., Cox, B.J., Gunderson, E.P., and Wheeler, M.B. (2019b). The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes. *Diabetologia* 62, 687-703.
- Lai, M., Al Rijjal, D., Röst, H.L., Dai, F.F., Gunderson, E.P., and Wheeler, M.B. (2020). Underlying dyslipidemia postpartum in women with a recent GDM pregnancy who develop type 2 diabetes. *eLife* 9.
- López-Ibáñez, J., Pazos, F., and Chagoyen, M. (2016). MBROLE 2.0—functional enrichment of chemical compounds. *Nucleic acids research* 44, W201-W204.

## SUPPLEMENTAL FIGURES



**Supplementary Figure-1, related to Figure-6A:** A flow diagram of step-by-step predictive biomarker analysis.

**Supplemental Table 1, related to Figure-1A. Characteristics of Matched-Pairs of Hispanic Women with GDM (n=60) without diabetes at 6 to 9 weeks postpartum (study baseline), and tested annually up to 2 years post-baseline for incident diabetes.**

<i>Characteristics</i>	<i>Incident T2D (N=30)</i>	<i>Not T2D (N=30)</i>	<i>p-value</i>
<b><u>Socio-demographic and Clinical</u></b>			
Age (years)	34.2 (5.9)	34.7 (5.0)	0.74
Parity n (%)			0.25
Primiparous (1 birth)	8 (26.7)	7 (23.3)	
Biparous (2 birth)	12 (40.0)	7 (23.3)	
Multiparous (>2 birth)	10 (33.3)	16 (53.3)	
GDM Prenatal treatment, n (%) <sup>a</sup>			0.04
Diet only	11 (36.7)	20 (66.7)	
Oral medications	17 (56.7)	10 (33.3)	
Insulin	2 (6.7)	0 (0.0)	
Gestational age at GDM diagnosis (weeks)	21.7 (9.5)	23.8 (8.3)	0.35
Pre-pregnancy BMI (kg/m <sup>2</sup> )	34.3 (4.7)	34.7 (5.1)	0.76
Hypertension history, n (%) <sup>a</sup>	0 (0.0)	1 (3.3)	0.33
Family history of diabetes, n (%)	21 (70.0)	17 (56.7)	0.28
<b><u>Baseline 6-9 Weeks Postpartum, Lifestyle</u></b>			
Physical activity, met-h/week	58.9 (27.0)	43.7 (19.0)	0.01
Total Energy intake, KJ/day	883.7 (414.4)	761.3 (335.6)	0.21
Lactation intensity categories, n (%) <sup>a</sup>			0.74
Exclusive lactation	4 (13.3)	5 (16.7)	
Mostly lactation	12 (40.0)	15 (50.0)	
Mostly formula/Mixed	8 (26.7)	4 (13.3)	
Exclusively formula	6 (20.0)	6 (20.0)	

**Supplemental Table 1, related to Figure-1A. Characteristics of Matched-Pairs of Hispanic Women with GDM (n=60) without diabetes at 6 to 9 weeks postpartum (study baseline), and tested annually up to 2 years post-baseline for incident diabetes.**

<i>Characteristics</i>	<i>Incident T2D (N=30)</i>	<i>Not T2D (N=30)</i>	<i>p- value</i>
<b><u>Baseline 6-9 Weeks Postpartum</u></b>			
BMI (kg/m <sup>2</sup> )	34.6 (4.7)	34.1 (5.2)	0.69
<b><u>Plasma:</u></b>			
Fasting glucose, mg/dl	102.5 (9.8)	94.1 (7.8)	<.001
2-hr Post-load glucose 75 g OGTT, mg/dl	140.0 (27.0)	106.9 (22.6)	<.001
Fasting insulin, μU/ml, Median (IQR) <sup>b</sup>	29.5 (30.4)	22.1 (15.7)	0.01
Fasting Triglycerides, mg/dl, Median (IQR) <sup>b</sup>	124.0 (116.0)	98.0 (70.0)	0.07
Fasting HDL-C, mg/dl	46.7 (11.3)	48.2 (12.2)	0.62
HOMA-IR, Median (IQR) <sup>b</sup>	7.8 (6.7)	4.9 (4.0)	0.007
HOMA-B, Median (IQR) <sup>b</sup>	274.3 (167.2)	274.6 (199.4)	0.44
<b><u>Post-baseline Up to 2-Year Follow Up</u></b>			
Any Subsequent birth, n (%) <sup>a</sup>	4 (13.3)	5 (16.7)	1.00
Follow up Time to Outcome (months), median (IQR) <sup>b</sup>	11.2 (9.1)	21.8 (1.4)	<.001

Note: All participants are “non-smokers”. Data are presented as the mean (SD) unless otherwise noted. Two sample t-test for continuous variables. Chi-square test for categorical variables.<sup>a</sup>

Fisher’s exact test.<sup>b</sup> Wilcoxon sum rank test for medians.

**Supplemental Table 2, related to Figure-2A: Significantly altered analytes (FDR < 0.1).**

Analytes	p.value	FDR	log2(FC)	AUC
CE15:0	0.0062228	0.076	-0.37	0.71
CE17:0	0.00094974	0.074	-0.41	0.75
CE18:0	0.0033006	0.074	-0.28	0.72
CE18:2	0.005466	0.076	-0.12	0.62
CE20:0	0.01499	0.099	-0.37	0.68
CE20:1	0.0098731	0.087	-0.33	0.69
CE24:0	0.012464	0.093	-0.48	0.69
CER16:0	0.0036671	0.074	-0.39	0.72
CER20:0	0.014675	0.098	-0.36	0.68
CER24:1	0.0065074	0.076	-0.35	0.71
Creatinine	0.011409	0.090	-0.25	0.69
DAG20:0/20:0	0.0044244	0.074	-0.37	0.71
FFA20:4	0.0040111	0.074	-0.37	0.72
HCER16:0	0.0054396	0.076	-0.44	0.71
HCER18:0	0.0038421	0.074	-0.61	0.72
HCER20:0	0.0029273	0.074	-0.51	0.72
HCER24:1	0.012465	0.093	-0.46	0.69
LCER16:0	0.0055666	0.076	-0.53	0.71
LCER24:1	0.0031068	0.074	-0.56	0.72
LPC15:0	0.0068145	0.077	-0.38	0.70
LPC17:0	0.0020494	0.074	-0.47	0.73
LPC18:0	0.00723	0.078	-0.21	0.70
LPC18:1	0.0081218	0.083	-0.31	0.70
LPC20:1	0.0043169	0.074	-0.49	0.72
LPC20:2	0.0083029	0.083	-0.35	0.70
LPE20:4	0.015605	0.099	-0.39	0.68
lysoPC a C17:0	0.0045283	0.075	-0.35	0.71
lysoPC a C18:1	0.011698	0.090	-0.24	0.69
lysoPC a C28:1	0.0048507	0.075	-0.48	0.71
PC aa C28:1	0.0059571	0.076	-0.42	0.71
PC aa C38:0	0.0084794	0.083	-0.38	0.70
PC aa C42:0	0.0063714	0.076	-0.44	0.71
PC aa C42:4	0.0058166	0.076	-0.50	0.71
PC ae C30:0	0.0083101	0.083	-0.40	0.70
PC ae C34:0	0.013828	0.096	-0.45	0.69
PC ae C34:1	0.011212	0.090	-0.30	0.69
PC ae C34:2	0.010977	0.090	-0.28	0.69
PC ae C36:1	0.0071121	0.078	-0.34	0.70
PC ae C36:2	0.0047356	0.075	-0.36	0.71

PC ae C38:2	0.0026833	0.074	-0.46	0.73
PC ae C38:4	0.0018984	0.074	-0.42	0.73
PC ae C38:5	0.0077735	0.081	-0.35	0.70
PC ae C40:2	0.0038466	0.074	-0.54	0.72
PC ae C40:3	0.0065164	0.076	-0.45	0.71
PC ae C40:4	0.0032586	0.074	-0.50	0.72
PC ae C40:5	0.004032	0.074	-0.46	0.72
PC ae C40:6	0.0028197	0.074	-0.45	0.73
PC ae C42:5	0.003106	0.074	-0.52	0.72
PC ae C44:5	0.0059585	0.076	-0.50	0.71
PC ae C44:6	0.0092568	0.084	-0.46	0.70
PC14:0/18:1	0.014097	0.096	-0.28	0.69
PC17:0/18:1	0.0065171	0.076	-0.38	0.71
PC17:0/18:2	0.0035809	0.074	-0.45	0.72
PC17:0/20:4	0.0027496	0.074	-0.51	0.73
PC18:1/18:1	0.013814	0.096	-0.33	0.69
PC18:1/18:2	0.0043082	0.074	-0.38	0.72
PC18:1/20:4	0.0092451	0.084	-0.40	0.70
PI18:0/18:1	0.010981	0.090	-0.33	0.69
PI18:1/18:1	0.012461	0.093	-0.40	0.69
SM C16:0	0.00030465	0.074	-0.37	0.77
SM C16:1	0.015275	0.099	-0.27	0.68
SM C18:0	0.015604	0.099	-0.30	0.68
SM C20:2	0.0032569	0.074	-0.48	0.72
SM C24:1	0.0015093	0.074	-0.41	0.74
SM OH C14:1	0.00087552	0.074	-0.43	0.75
SM OH C16:1	0.0049605	0.076	-0.46	0.71
SM OH C22:2	0.0036662	0.074	-0.40	0.72
SM14:0	0.0021471	0.074	-0.37	0.73
SM16:0	0.001461	0.074	-0.30	0.74
SM20:0	0.014647	0.098	-0.29	0.68
SM22:0	0.006294	0.076	-0.25	0.71
SM22:1	0.0034907	0.074	-0.31	0.72
SM24:1	0.0050573	0.076	-0.35	0.71
SM26:1	0.013536	0.096	-0.40	0.69
Total LCER	0.003415	0.074	-0.52	0.72
TAG50:5-FA18:1	0.0090563	0.084	-0.19	0.70
DAG16:0/16:0	0.000071855	0.059	0.67	0.80
DAG16:0/16:1	0.00282	0.074	0.61	0.73
DAG16:0/18:0	0.0016334	0.074	0.26	0.74
DAG16:0/18:1	0.0026163	0.074	0.27	0.73

DAG16:0/20:4	0.01468	0.098	0.35	0.68
DAG18:0/18:1	0.0019009	0.074	0.25	0.73
DAG18:0/18:2	0.010523	0.088	0.20	0.69
Total TAG	0.012958	0.095	0.46	0.67
TAG46:0-FA14:0	0.0092646	0.084	1.03	0.70
TAG46:0-FA16:0	0.010984	0.090	0.98	0.69
TAG48:0-FA14:0	0.014097	0.096	0.63	0.69
TAG48:0-FA16:0	0.0041201	0.074	1.15	0.72
TAG48:1-FA16:0	0.014091	0.096	0.80	0.69
TAG48:1-FA16:1	0.013537	0.096	0.94	0.69
TAG49:0-FA16:0	0.0071264	0.078	0.66	0.70
TAG49:0-FA17:0	0.0044266	0.074	0.76	0.71
TAG50:0-FA14:0	0.012201	0.093	0.35	0.69
TAG50:0-FA16:0	0.0026189	0.074	0.83	0.73
TAG50:0-FA18:0	0.0042233	0.074	0.81	0.72
TAG50:1-FA16:0	0.0062156	0.076	0.74	0.71
TAG50:1-FA18:1	0.0098496	0.087	0.68	0.69
TAG50:2-FA18:2	0.0066537	0.076	0.62	0.70
TAG50:2-FA20:2	0.0098752	0.087	0.45	0.69
TAG50:3-FA18:3	0.0083074	0.083	0.62	0.70
TAG50:3-FA20:3	0.0038471	0.074	0.63	0.72
TAG51:0-FA16:0	0.011462	0.090	0.47	0.69
TAG51:0-FA17:0	0.0090659	0.084	0.46	0.70
TAG51:0-FA18:0	0.015632	0.099	0.44	0.68
TAG52:0-FA16:0	0.0042227	0.074	0.54	0.72
TAG52:0-FA18:0	0.0031039	0.074	0.59	0.72
TAG52:1-FA16:1	0.0072859	0.078	0.36	0.70
TAG52:1-FA18:0	0.010293	0.088	0.48	0.69
TAG52:1-FA20:1	0.0013689	0.074	0.51	0.74
TAG52:2-FA18:0	0.0086686	0.084	0.44	0.70
TAG52:2-FA20:2	0.00092605	0.074	0.60	0.75
TAG52:3-FA20:3	0.00081132	0.074	0.71	0.75
TAG52:4-FA20:3	0.0092559	0.084	0.37	0.70
TAG52:4-FA20:4	0.0043217	0.074	0.67	0.72
TAG52:6-FA22:6	0.013262	0.096	0.48	0.69
TAG52:7-FA16:0	0.010092	0.088	0.49	0.69
TAG54:1-FA16:0	0.0074395	0.079	0.27	0.70
TAG54:1-FA20:1	0.0063729	0.076	0.26	0.71
TAG54:2-FA20:2	0.0047431	0.075	0.27	0.71
TAG54:3-FA20:3	0.00043296	0.074	0.41	0.77
TAG54:4-FA16:0	0.0033281	0.074	0.30	0.72

TAG54:4-FA20:3	0.01143	0.090	0.27	0.69
TAG54:4-FA20:4	0.0054383	0.076	0.36	0.71
TAG54:4-FA22:4	0.0015964	0.074	0.64	0.74
TAG54:5-FA16:0	0.011683	0.090	0.29	0.69
TAG54:5-FA22:5	0.0021018	0.074	0.62	0.73
TAG54:6-FA22:6	0.010311	0.088	0.59	0.69
TAG56:4-FA18:0	0.013256	0.096	0.22	0.69
TAG56:4-FA22:4	0.0060922	0.076	0.34	0.71
TAG56:5-FA16:0	0.015615	0.099	0.27	0.68

**Supplemental Table 3, related to Figure-2A: The identified KEGG pathways**

	Raw <i>p</i>	FDR	Impact
<b>Spingolipid metabolism</b>	0.00000424	0.0003	0.3367
Glycerophospholipid metabolism	0.01524	0.6097	0.1037
Arachidonic acid metabolism	0.0365	0.9733	0.2167
Arginine and proline metabolism	0.32366	1	0.0065
<b>Over-representation KEGG pathways (upregulated)</b>			
<b>Fatty acid biosynthesis</b>	0.00000732	0.0006	0
Linoleic acid metabolism	0.08976	1	0.6563
Fatty acid elongation in mitochondria	0.15609	1	0
alpha-Linolenic acid metabolism	0.16669	1	0.2034
Fatty acid metabolism	0.2708	1	0.0296