

## UC Davis

### UC Davis Previously Published Works

#### Title

In utero metabolomic signatures of refined grain intake and risk of gestational diabetes: A metabolome-wide association study.

#### Permalink

<https://escholarship.org/uc/item/6c21453s>

#### Journal

American Journal of Clinical Nutrition, 117(4)

#### Authors

Chehab, Rana

Ferrara, Assiamira

Zheng, Siwen

et al.

#### Publication Date

2023-04-01

#### DOI

10.1016/j.ajcnut.2023.02.009

Peer reviewed



## Original Research Article

## In utero metabolomic signatures of refined grain intake and risk of gestational diabetes: A metabolome-wide association study

Rana F. Chehab<sup>1,\*</sup>, Assiamira Ferrara<sup>1</sup>, Siwen Zheng<sup>2</sup>, Dinesh K. Barupal<sup>3</sup>, Amanda L. Ngo<sup>1</sup>, Liwei Chen<sup>4</sup>, Oliver Fiehn<sup>5</sup>, Yeyi Zhu<sup>1,6,\*</sup>

<sup>1</sup> Division of Research, Kaiser Permanente Northern California, Oakland, CA; <sup>2</sup> School of Public Health, University of California, Berkeley, CA; <sup>3</sup> Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, NY; <sup>4</sup> Department of Epidemiology, University of California, Los Angeles, CA; <sup>5</sup> West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, CA; <sup>6</sup> Department of Epidemiology and Biostatistics, University of California, San Francisco, CA

### A B S T R A C T

**Background:** Epidemiologic evidence has linked refined grain intake to a higher risk of gestational diabetes (GDM), but the biological underpinnings remain unclear.

**Objectives:** We aimed to identify and validate refined grain-related metabolomic biomarkers for GDM risk.

**Methods:** In a metabolome-wide association study of 91 cases with GDM and 180 matched controls without GDM (discovery set) nested in the prospective Pregnancy Environment and Lifestyle Study (PETALS), refined grain intake during preconception and early pregnancy and serum untargeted metabolomics were assessed at gestational weeks 10–13. We identified refined grain-related metabolites using multivariable linear regression and examined their prospective associations with GDM risk using conditional logistic regression. We further examined the predictivity of refined grain-related metabolites selected by least absolute shrinkage and selection operator regression in the discovery set and validation set (a random PETALS subsample of 38 individuals with and 336 without GDM).

**Results:** Among 821 annotated serum (87.4% fasting) metabolites, 42 were associated with refined grain intake, of which 17 (70.6% in glycerolipids, glycerophospholipids, and sphingolipids clusters) were associated with subsequent GDM risk (all false discovery rate-adjusted  $P$  values  $<0.05$ ). Adding 7 of 17 metabolites to a conventional risk factor-based prediction model increased the C-statistic for GDM risk in the discovery set from 0.71 (95% CI: 0.64, 0.77) to 0.77 (95% CI: 0.71, 0.83) and in the validation set from 0.77 (95% CI: 0.69, 0.86) to 0.81 (95% CI: 0.74, 0.89), both with  $P$ -for-difference  $<0.05$ .

**Conclusions:** Clusters of glycerolipids, glycerophospholipids, and sphingolipids may be implicated in the association between refined grain intake and GDM risk, as demonstrated by the significant associations of these metabolites with both refined grains and GDM risk and the incremental predictive value of these metabolites for GDM risk beyond the conventional risk factors. These findings provide evidence on the potential biological underpinnings linking refined grain intake to the risk of GDM and help identify novel disease-related dietary biomarkers to inform diet-related preventive strategies for GDM.

**Keywords:** refined grains, gestational diabetes, untargeted metabolomics, biomarker, prediction

## Introduction

Gestational diabetes (GDM), one of the most common pregnancy complications [1], affects approximately 8% of pregnancies in the United States [2] and predisposes pregnant individuals and their offspring to a multitude of perinatal complications and long-term

diabetes and cardiovascular disease sequelae [3]. The etiology of GDM is multifactorial and remains poorly characterized, with several risk factors potentially contributing to the onset of GDM, including age, overweight or obesity, and family history of diabetes [4]. Diet and lifestyle factors before and during pregnancy have been also associated with the risk of GDM. One such food group is total grains, which are

**Abbreviations:** aHEI-P, alternate Healthy Eating Index for Pregnancy; ChemRICH, chemical similarity enrichment analysis; EHR, electronic health record; FDR, false discovery rate; GCT, glucose challenge test; GDM, gestational diabetes; KPNC, Kaiser Permanente Northern California; LASSO, least absolute shrinkage and selection operator; MS, mass spectrometry; OGTT, oral glucose tolerance test; PETALS, Pregnancy Environment and Lifestyle Study.

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [rana.chehab@kp.org](mailto:rana.chehab@kp.org) (R.F. Chehab), [yeyi.zhu@kp.org](mailto:yeyi.zhu@kp.org) (Y. Zhu).

<https://doi.org/10.1016/j.ajcnut.2023.02.009>

Received 8 September 2022; Received in revised form 6 January 2023; Accepted 8 February 2023

Available online 11 February 2023

0002-9165/© 2023 American Society for Nutrition. Published by Elsevier Inc. All rights reserved.

among the most commonly consumed foods in the world, contributing to approximately 25% of the total daily energy intake among US adults [5]. In particular, refined grains such as white flour, white rice, and white bread, which are milled for finer texture and longer shelf life but lack dietary fiber, iron, and B vitamins and have a higher glycemic index compared with whole grains, have been implicated in the risk of GDM [6]. However, the biological underpinnings of the association between refined grain intake and the risk of GDM remain understudied.

Metabolomics, the study of small molecules known as metabolites and the omics approach closest to the phenotype, is a powerful tool for precision nutrition [7]. In contrast to the conventional single-factor epidemiological approach, metabolomics can provide an integrated profile of the current biological status, serving as a pathophysiologic read-out reflecting endogenous and exogenous interplay [8]. Few studies have used metabolomics to systematically identify metabolomic markers for grain intake and further examine their associations with type 2 diabetes risk [9,10], whereas data among pregnant individuals are scarce. No previous studies have examined the metabolomic profile related to refined grain intake among pregnant

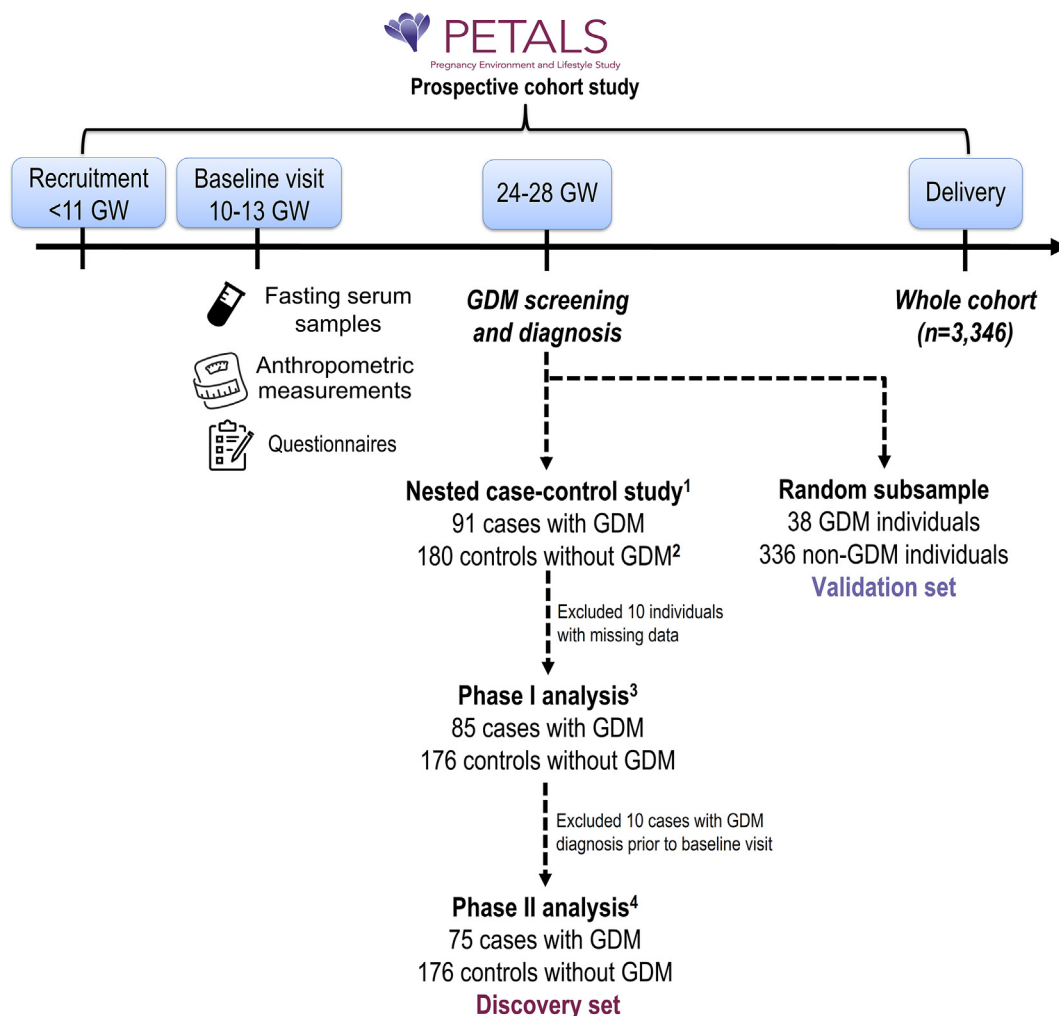
individuals nor has any study investigated the role of refined grain-related metabolites in the risk of GDM.

To fill this gap, we conducted a prospective discovery and validation metabolome-wide association study among pregnant individuals with multiracial and ethnic backgrounds in a large integrated clinical setting where universal screening and standardized diagnosis for GDM were implemented. We aimed to 1) identify metabolites in early pregnancy associated with dietary intake of refined grains, 2) investigate the prospective associations of the refined grain-related metabolites with the risk of GDM, and 3) explore the incremental predictivity of these refined grain-related metabolites for GDM risk beyond conventional risk factors.

## Methods

### Study design and population

This study is a secondary analysis of the Pregnancy Environment and Lifestyle Study (PETALS), designed to examine associations of



**FIGURE 1.** Study flowchart

GDM, gestational diabetes; GW, gestational weeks.

<sup>1</sup>1:2 case-control ratio with cases and controls matched on age at delivery ( $\pm 5$  y), calendar time for enrollment ( $\pm 3$  mo), and gestational weeks at baseline clinic visit ( $\pm 3$  wk).

<sup>2</sup>180 not 182 because 2 selected controls had missing blood samples.

<sup>3</sup>Phase I analysis identified individual metabolites associated with refined grain intake.

<sup>4</sup>Phase II analysis identified refined grain-related metabolites associated with a subsequent risk of GDM.

intrauterine environmental factors and the risk of GDM. The study design has been described in detail elsewhere [11]. Briefly, PETALS is a population-based longitudinal multiracial and -ethnic pregnancy cohort study (Figure 1). The study population was drawn from the membership of Kaiser Permanente Northern California (KPNC), an integrated health care delivery system serving 4.5 million members, about 30% of the population across 14 counties of the Greater Bay Area and the California Central Valley [12]. The KPNC membership is socio-economically diverse and highly representative of the entire population living in the served geographic area [12]. After weekly searches of the electronic health records (EHRs), pregnant individuals aged 18–45 y at delivery, carrying singletons, and without recognized chronic diabetes, cancer, hepatitis C, or liver cirrhosis were invited to participate in the study before gestational week 11, and 3346 were included in the final cohort. Fasting blood draw, anthropometric measurements, and questionnaires on health history, dietary intake, and other lifestyle factors were completed at the baseline clinic visit (gestational weeks 10–13), and GDM screening was conducted at gestational weeks 24–28. The study was approved by the human subjects committee of the Kaiser Foundation Research Institute. Written informed consent was obtained from all the participants.

### Selection of cases with GDM and controls without GDM

Pregnant individuals at KPNC received universal GDM screening (96%) with a 50-g, 24-h glucose challenge test (GCT) around gestational weeks 24–28 [2]. If the screening test was abnormal (GCT  $\geq$ 140 mg/dL), a diagnostic 100-g, 3-h oral glucose tolerance test (OGTT) was performed after an 8–12-h fast. Plasma glucose measurements were performed using the hexokinase method at the KPNC regional laboratory, which participates in the College of American Pathologists' accreditation and monitoring program. GDM was ascertained by meeting any of the following criteria: 1)  $\geq$ 2 OGTT plasma glucose values meeting or exceeding the Carpenter-Coustan thresholds: 1-h 180 mg/dL, 2-h 155 mg/dL, and 3-h 140 mg/dL [13] or 2) 1-h GCT  $\geq$ 180 mg/dL and fasting glucose  $\geq$ 95 mg/dL performed alone or during the OGTT [14].

Among the PETALS participants who delivered between April 2015 and January 2018, we identified 200 individuals with GDM, 91 of whom were diagnosed by Carpenter-Coustan criteria, had serum untargeted metabolomics data, and constituted the cases with GDM in the current nested case-control study. We then identified 180 controls without GDM from the PETALS cohort who delivered during the same time and were matched with the cases at a case:control ratio of 1:2 according to age at delivery ( $\pm$ 5 y), calendar time for enrollment ( $\pm$ 3 mo), and gestational weeks at baseline clinic visit ( $\pm$ 3 wk). Notably, the number of controls was 180 not 182 because 2 selected controls had missing blood samples. We excluded 10 participants with missing data on refined grain intake or other covariates, rendering a sample size of 261 for identification of refined grain-related metabolites. To identify the refined grain-related metabolites associated with GDM risk, we further excluded 10 cases because they were diagnosed with GDM before the timing of FFQ administration and blood sample collection, resulting in a sample of 251 pregnant individuals in the discovery set. To derive the validation set, we randomly selected approximately 10% of participants in the PETALS cohort (38 individuals with GDM and 336 individuals without GDM) who delivered between April 2014 and May 2019, were not included in the discovery set, and had complete data on dietary intake and serum untargeted metabolomics in early pregnancy.

### Dietary exposure assessment

Data on habitual dietary intake during the previous 3 mo were obtained via the Block FFQ administered at the baseline clinic visit (gestational weeks 10–13), reflecting diet during preconception and early pregnancy. The Block FFQ has demonstrated adequate reliability and validity in comparison to multiple dietary records [15], serving as a useful instrument for analysis at the energy, food, and nutrient level among diverse populations including pregnant individuals [16,17]. Study participants reported their usual intake and portion size of foods and beverages, including items modified to accommodate the diverse dietary habits of the multiracial and -ethnic study cohort as used in previous studies [18]. The nutrient and food group analysis database was developed from the US Department of Agriculture Food and Nutrient Database for Dietary Studies version 5.0, the Food Pyramid Equivalents Database, and the Nutrient Database for Standard Reference [19]. Energy intake values  $<$ 400 kcal/d or  $>$ 6000 kcal/d were defined as implausible and excluded.

### Metabolomics data acquisition and pre-processing

Serum samples collected during the baseline visit (gestational weeks 10–13) were stored at  $-80^{\circ}\text{C}$  before analysis. To conduct a metabolome-wide association study, untargeted metabolomics data were acquired at the University of California, Davis West Coast Metabolomics Center, using 3 complementary mass spectrometry (MS) based assays as follows [20]: 1) primary metabolites such as mono- and disaccharides, hydroxyl- and amino acids were measured by gas chromatography/time-of-flight MS [21] including data alignment and compound annotation using the BinBase database algorithm [22], 2) complex lipids ranging from triacylglycerides, phosphoglycerolipids, and sphingolipids to free fatty acids were analyzed by liquid chromatography (LC)/quadrupole time-of-flight (QTOF) MS [23], and 3) biogenic amines including microbial compounds such as trimethylamine N-oxide, methylated and acetylated amino acids, and short di- and tripeptides were measured by hydrophilic interaction LC/QTOF MS. All LC-MS/MS data included diverse sets of internal standards. LC-MS data were processed by MS-DIAL version 4.0 software [24], and compounds were annotated based on accurate mass, retention time, and MS/MS fragment matching using LipidBlast [25] and Massbank of North America libraries [20]. MS-FLO was used to remove erroneous peaks and reduce the false positive peak in LC datasets [26]. A total of 821 known metabolites were annotated and included in our metabolome-wide association analysis, with the raw metabolite concentrations measured as peak intensities. Data in the discovery and validation sets were normalized via systematic error removal using random forest [27] and further transformed using the inverse normal transformation to account for the batch effect and improve normality. Residual technical errors were assessed by coefficients of variation: on average, 5.8% (range 1.2%–17.8%) for primary metabolites, 3.7% (range 0.7%–19.2%) for complex lipids, and 11.8% (range 2.7%–19.5%) for biogenic amines.

### Covariates

Potential covariates were selected based on biological plausibility and prior knowledge, including: age at delivery ( $<$ 25, 25–29, 30–34,  $\geq$ 35 y), self-identified race and ethnicity (Asian/Pacific Islander, non-Hispanic Black, Hispanic, non-Hispanic White, and Other/unknown), education (high school or less, some college/associate degree, college degree or higher), prepregnancy BMI ( $<$ 18.5, 18.5–24.9, 25.0–29.9,  $\geq$ 30.0 kg/m<sup>2</sup>), nulliparity (yes/no), family history of diabetes (yes/no), chronic hypertension (yes/no), smoking before and during pregnancy

(yes/no), alcohol use before and during pregnancy (yes/no), gestational weeks at blood collection (continuous), fasting status at blood collection (yes/no), total energy intake (kcal/d) (quartiles), overall dietary quality assessed by alternate Healthy Eating Index for Pregnancy (aHEI-P) score (quartiles), and physical activity during pregnancy assessed as metabolic equivalent of tasks-h/d (continuous).

Information on the pregnant individuals' sociodemographic and lifestyle factors was collected by a structured questionnaire administered at the baseline clinic visit (gestational weeks 10–13) and information on medical history was extracted from the EHR and if missing, supplemented by the study questionnaire. Prepregnancy BMI was calculated as prepregnancy weight (kg) measured by the clinical staff on average 11 wk before conception and abstracted from the EHR (97.5%) or by self-report (2.5%), divided by squared height (m<sup>2</sup>) measured at the baseline visit. The aHEI-P was adapted from the AHEI-2010 by Chiuvè et al [28] and an earlier pregnancy AHEI score by Rifas-Shiman et al. [29], which included a whole grain but not refined grain component. Moderate-to-vigorous physical activity was assessed by the validated Pregnancy Physical Activity Questionnaire in the first trimester [30]. A covariate was included in the final model if the coefficient of exposure of interest changed by 10% or more. Although cases and controls were matched on age at delivery and gestational weeks at blood collection, we adjusted for these factors in the regression models to account for potential residual confounding [31].

## Statistical analysis

Characteristics of study participants included in the nested case-control study (discovery set) and in the validation set were summarized as frequency (%) for categorical variables and mean (SD) by quartiles of refined grain intake.

To systematically identify refined grain-related metabolites associated with GDM risk, the statistical analysis was conducted in 2 phases. In phase I (metabolome-wide association analysis), we identified individual metabolites associated with refined grain intake using multivariable linear regression adjusting for age at delivery, race and ethnicity, education, prepregnancy BMI, chronic hypertension, smoking before and during pregnancy, alcohol use before and during pregnancy, total energy intake, aHEI-P score, gestational weeks at blood collection, and fasting status at blood collection. We considered refined grain intake as the primary exposure and presented the results as percentage difference in metabolite levels per single serving (1 ounce, 28.35 g equivalent/d) increase in refined grain intake, using the following exponential function:  $[\exp(\beta \text{ coefficient}) - 1] \times 100\%$  [32]. In a sensitivity analysis, we tested the linear trend of the association between refined grain intake and metabolite levels (both examined as continuous variables) by applying restricted cubic splines with 3 equally spaced knots at 33rd (reference) and 66th percentiles.

In phase II, we examined the associations of the refined grain-related metabolites identified in phase I with subsequent risk of GDM using conditional logistic regression. In addition to the covariates adjusted for in phase I, we adjusted for nulliparity and family history of diabetes as major risk factors for GDM. We then conducted a chemical similarity enrichment analysis (ChemRICH) of the refined grain-related metabolites associated with GDM risk to facilitate biological interpretation. ChemRICH is a statistical enrichment approach that is based on chemical similarity rather than sparse biochemical knowledge annotations [33]. ChemRICH yields study-specific, nonoverlapping clusters of metabolites, and each cluster has a self-contained size where *P* values of different metabolite clusters were calculated using the Kolmogorov-Smirnov test, not relying on the size of the background database [33].

Finally, we developed 3 sequential prediction models using logistic regression analysis to examine the incremental predictive ability of refined grain-related metabolites above and beyond conventional risk factors in GDM risk prediction. Model 1 included the covariates in phase II as conventional risk factors for GDM. Model 2 included a subset of the refined grain-related metabolites associated with GDM risk in the discovery set identified using the least absolute shrinkage and selection operator (LASSO) regression. A subset instead of all the metabolites was used to develop more interpretable and parsimonious models. Model 3 included a combination of predictors in Models 1 and 2. Model discrimination was evaluated by the C-statistic, which represents the area under the curve of the receiver operating characteristic curve, and which was pair-wisely compared using the nonparametric DeLong test [34]. To avoid overfitting, 10-fold cross-validation was performed to derive conservative estimates within the discovery set. To derive recalibrated results of the prediction models generalizable to the entire PETALS cohort, samples from the nested case-control discovery set were reweighted using sampling weights created via a weighted likelihood approach based on the inverse probability of GDM in the subsample compared with the entire cohort. We further validated the predictive performance of the selected metabolite subset by applying the prediction models derived from the discovery set to the validation set.

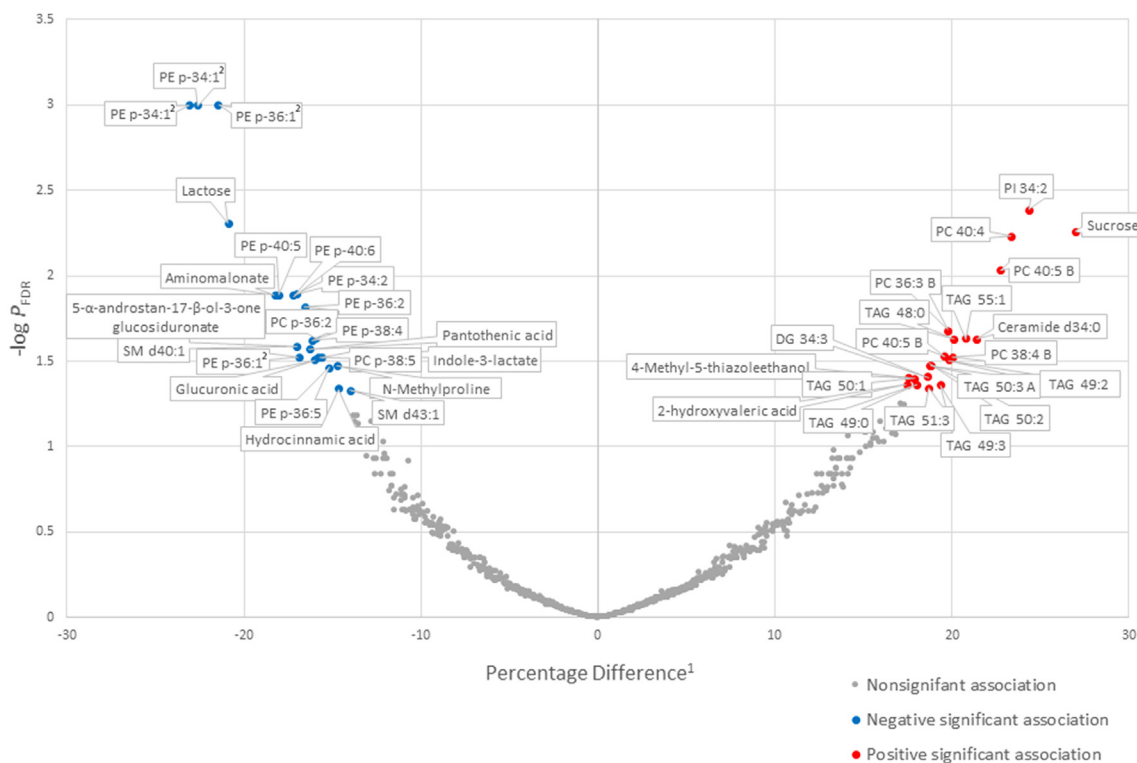
Regarding power calculations, among 91 cases with GDM and 180 matched controls without GDM, with 80% power and a false discovery rate (FDR) of 5%, we can detect a true minimum effect size of 0.15 (15% difference in normalized metabolite concentration between cases and controls) for all annotated metabolites (*n* = 821), assuming that at least 80% (nondifferential proportion) compounds are not associated with the outcomes. Lowering the percentage of associated compounds to only 10% reduces the minimum effect size to 0.14 for all metabolites, suggesting that there is sufficient power to detect even small effects in this study for a large number of compounds and to discover metabolic signatures of interest.

All analyses were performed using SAS version 9.4 (SAS Institute Inc) and R version 3.6. We calculated the FDR-adjusted *P* value to correct for multiple comparisons. The statistically significant level was set at a 2-tailed *P* value <0.05.

## Results

The Table presents the characteristics of study participants by quartiles of refined grain intake in the nested case-control study (discovery set) and the random subsample of the PETALS cohort (validation set). The weighted participant characteristics in the discovery set after applying sampling weights to account for the oversampling of cases with GDM were similar to those of the PETALS cohort (Supplementary Table 1). The distribution of the 821 annotated metabolites by metabolic superclass was as follows: lipids and lipid-like molecules (71.0%), organic acid and derivatives (13.0%), organoheterocyclic compounds (5.6%), organic oxygen compounds (4.5%), benzenoids (2.2%), organic nitrogen compounds (1.7%), nucleosides, nucleotides, and analogs (0.7%), phenylpropanoids and polyketides (0.6%), alcohols and derivatives and homogeneous nonmetal compounds (0.3% for both), and hydrocarbons (0.1%) (Supplementary Figure 1).

In the phase I metabolome-wide association analysis, among the 821 annotated metabolites, 42 (78.6% lipids and lipid-like molecules, 14.3% organic oxygen compounds, 4.8% organoheterocyclic compounds, and 2.4% phenylpropanoids and polyketides) were associated with refined grain intake after FDR adjustment (*P*<sub>FDR</sub> <0.05) (Figure 2



**FIGURE 2.** Volcano plot depicting the percentage difference in the metabolites per serving increase in refined grain intake among participants in the discovery set of a nested case-control study within the PETALS cohort.

DG, diglyceride; PC, phosphatidylcholine; PE, polyethylene; PI, phosphatidylinositol; SM, sphingomyelin; TAG, triacylglycerol.

<sup>1</sup>Calculated using linear regression analysis adjusted for age at delivery, self-identified race and ethnicity, education, prepregnancy body mass index, chronic hypertension, smoking before and during pregnancy, alcohol use before and during pregnancy, total energy intake, alternate Healthy Eating Index for Pregnancy (aHEI-P) score, gestational weeks at blood collection, and fasting status at blood collection.

<sup>2</sup>Different electrospray ionization (ESI) modes of the indicated metabolite.

and [Supplementary Table 2](#)). Of these 42 metabolites, 47.6% were positively and 52.4% were negatively associated with refined grain intake. In a sensitivity analysis exploring nonlinear associations, regression based on restricted cubic splines showed that most of the metabolites (83%) had linear associations with refined grain intake ( $P$  value for nonlinearity  $>0.05$ ; data not shown).

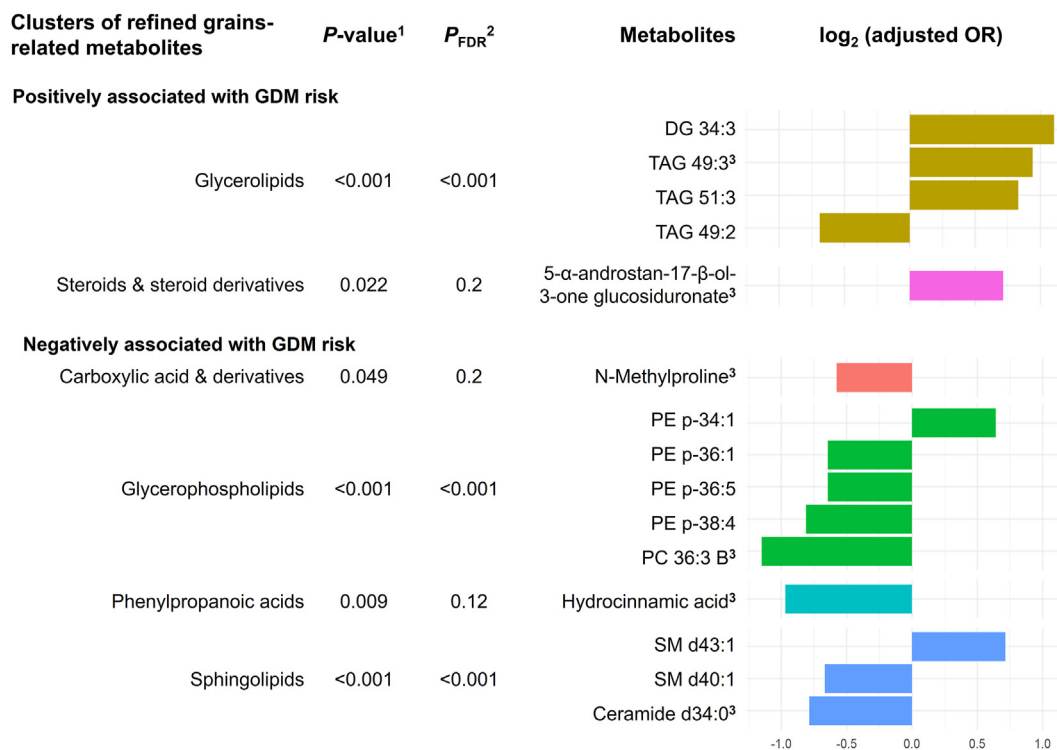
In the phase II analysis, 17 out of the 42 metabolites were associated with the risk of GDM ( $P_{FDR} < 0.05$ ) ([Supplementary Table 3](#)). We further conducted ChemRICH analysis to map biochemical clusters and facilitate biological interpretation of metabolic processes underlying the associations of refined grain metabolites and the risk of GDM. Seven clusters of refined grain-related metabolites significantly associated with GDM risk were enriched via ChemRICH ( $P < 0.05$ ), 3 clusters (glycerolipids, glycerophospholipids, and sphingolipids) of which remained significant after FDR adjustment ( $P_{FDR} < 0.001$ ; [Figure 3](#)). Overall, the glycerolipid cluster was positively associated with the risk of GDM and driven by key metabolite triacylglycerol 49:3 (OR: 2.15; 95% CI: 1.33, 3.48) with the lowest  $P$  value within this cluster. The glycerophospholipids and sphingolipid clusters were overall negatively associated with the risk of GDM with phosphatidylcholine 36:3 B (OR: 0.45; 95% CI: 0.28, 0.73) and ceramide d34:0 (OR: 0.58; 95% CI: 0.38, 0.88) as the key metabolites, respectively. In a sensitivity analysis further adjusting for physical activity, we found similar results to those in the main analysis.

To evaluate the incremental predictability of metabolites beyond conventional risk factors (Model 1 as reference), we developed multimetabolite panels using 10-fold cross-validation (Model 2) and an

additive Model 3 including predictors in Models 1 and 2 in the discovery set ([Figure 4A](#)). The multimetabolite panels selected using LASSO regression included the following 7 of the 17 GDM-associated metabolites: polyethylene p-38:4, sphingolipids d40:1, hydrocinnamic acid, N-methylproline, sphingolipids d43:1, diglyceride 34:3, and triacylglycerol 51:3. Details of feature selection via LASSO regression are shown in [Supplementary Figure 2](#). The C-statistic for Model 1 using the conventional risk factors (0.71; 95% CI: 0.64, 0.77) was similar to that for Model 2 (0.71; 95% CI: 0.64, 0.78;  $P = 0.990$ ). In Model 3, adding the 7 LASSO-selected metabolites to the conventional risk factors model further increased the C-statistic to 0.77 (95% CI: 0.71, 0.83;  $P = 0.008$  comparing Model 3 vs. Model 1). In the validation set ([Figure 4B](#)), 6 out of the 7 metabolites (i.e., polyethylene p-38:4, sphingolipids d40:1, N-methylproline, sphingolipids d43:1, diglyceride 34:3, triacylglycerol 51:3) selected by LASSO in the discovery set were detected and available. The C-statistic for Model 2 (0.65; 95% CI: 0.57, 0.73) was lower compared with Model 1 (0.77; 95% CI: 0.69, 0.86;  $P = 0.038$ ). Combining the conventional risk factors and metabolite biomarkers increased the C-statistic in Model 3 (0.81; 95% CI: 0.74, 0.89) compared to Model 1 ( $P = 0.013$ ) and Model 2 ( $P < 0.001$ ).

## Discussion

In a case-control metabolome-wide association study nested within the prospective PETALS cohort, we identified 42 serum metabolites



**FIGURE 3.** Multivariate ChemRICH enrichment plots depicting the clusters and metabolites associated with intake of refined grains and the risk of gestational diabetes in the discovery set of a nested case-control study within the PETALS cohort.

DG, diglyceride; FDR, false discovery rate; GDM, gestational diabetes; PC, phosphatidylcholine; PE, polyethylene; SM, sphingomyelin; TAG, triacylglycerol.

<sup>1</sup> $P$  value of each metabolite pathway was calculated using the Kolmogorov-Smirnov test.

<sup>2</sup> $P_{FDR}$  was adjusted across the metabolites in all clusters.

<sup>3</sup>Key metabolite: the metabolite with the lowest  $P$  value within each cluster.

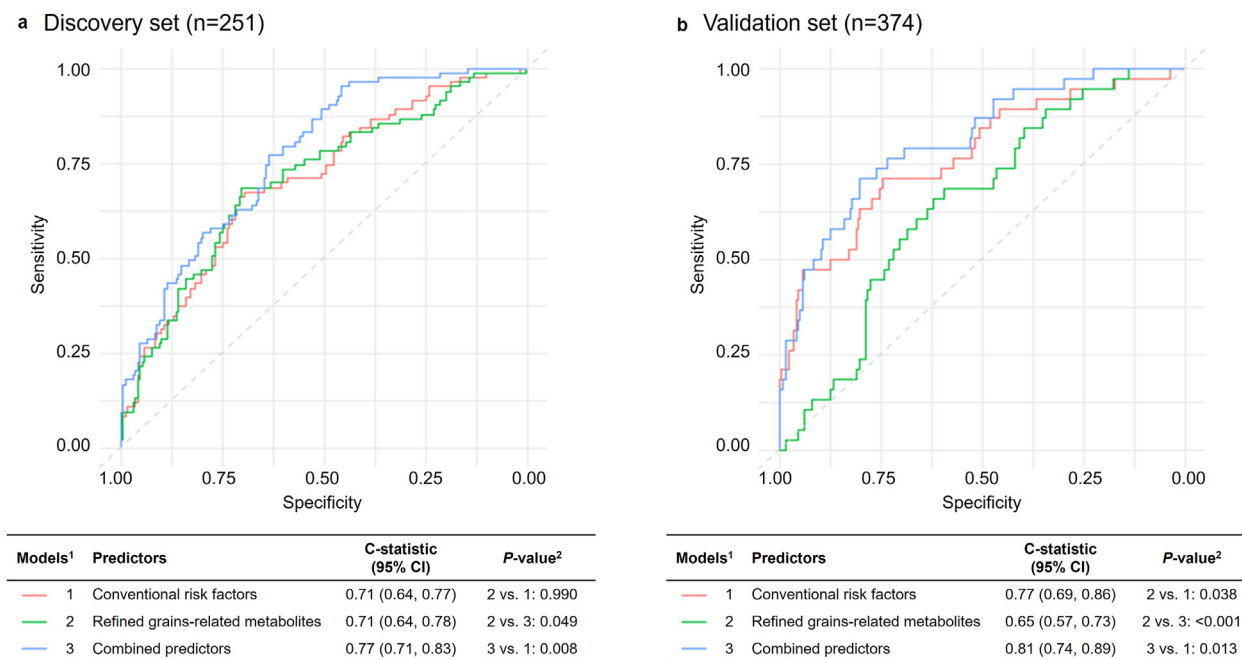
(mainly lipids and derivatives) as biomarkers for the intake of refined grains. Of the 42 metabolites, 17 were also associated with GDM risk. Moreover, 7 refined grain-related metabolites selected by LASSO regression exhibited incremental predictability of GDM risk above and beyond conventional risk factors. These findings provide new insights into the metabolomic profiles of refined grain intake and the underlying mechanisms linking them to the risk of GDM to inform diet-related preventive strategies for GDM.

To the best of our knowledge, no previous study has investigated the metabolomic profile of refined grain intake, nor has any study examined the metabolic pathways linking refined grain intake to the risk of GDM. The majority of the annotated metabolites in our study were lipids and lipid-like molecules, consistent with other studies in the literature among pregnant individuals [35,36], potentially due to the easier identification of lipids compared to other metabolites. We identified 17 metabolites associated with both refined grain intake and the risk of GDM, including those belonging to fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, steroids, carboxylic acids, organooxygen compounds, and phenylpropanoic acid clusters. These findings underscore the key role of lipid metabolism in the refined grain-related GDM etiology and echo those from previous studies on the metabolomic profile of GDM. In a case-control study of 321 pregnant individuals nested in the National Institute of Child Health and Human Development Fetal Growth Studies-Singleton cohort, plasma glycerolipids at the time of GDM diagnosis were positively associated with GDM risk [37]. Two case-control studies in Poland and Germany among 24–40 pregnant individuals showed that plasma glycerophospholipids at the time of GDM diagnosis were

negatively associated with GDM risk [38,39]. One longitudinal study with 61 pregnant individuals in China found that unsaturated glycerophospholipids and sphingolipids were lower in the plasma samples of participants with GDM throughout pregnancy compared with controls [40]. However, none of the previous studies have comprehensively assessed the associations among refined grain intake, serum untargeted metabolites, and the risk of GDM.

Importantly, we identified novel metabolomic markers including N-methylproline and hydrocinnamic acid related to both refined grain intake and GDM risk. Despite the lack of comparable data among pregnant individuals, previous studies among nonpregnant individuals have linked N-methylproline to dietary intake of healthy plant-based foods and lower risk of diabetes and cardiovascular disease [41,42]. Hydrocinnamic acid, as a major component of cinnamon extract, may increase glucose disposal by enhancing glucose transport activity in animal models [43].

Given the positive association between refined grain intake and GDM risk, we expected the associations between refined grain-related metabolites and the risk of GDM to be consistent so that metabolites positively associated with refined grain intake are also positively associated with GDM risk and vice versa. However, 3 metabolites (i.e., triacylglycerol 49:2, phosphatidylcholine 36:3 B, and ceramide d34:0) that were positively associated with refined grain intake were negatively associated with GDM risk, and 4 metabolites (i.e., polyethylene p-34:1, sphingomyelin d43:1, 5- $\alpha$ -androstan-17- $\beta$ -ol-3-one glucosiduronate, and lactose) that were negatively associated with refined grain intake were positively associated with GDM risk. It is plausible that these metabolites do not function individually; therefore, we used



**FIGURE 4.** The C-statistic (95% CI) of prediction models of gestational diabetes risk based on conventional risk factors and/or serum refined grain-related metabolites in the discovery set (a) and validation set (b) within the PETALS cohort.

<sup>1</sup>All models were calculated using logistic regression analysis.

Model 1 adjusted for conventional risk factors for gestational diabetes, including age at delivery, self-identified race and ethnicity, education, prepregnancy body mass index, nulliparity, family history of diabetes, chronic hypertension, smoking before and during pregnancy, alcohol use before and during pregnancy, total energy intake, alternate Healthy Eating Index for Pregnancy (aHEI-P) score, gestational weeks at blood collection, and fasting status at blood collection.

Model 2 adjusted for the 7 refined grain-related metabolites selected via least absolute shrinkage and selection operator (LASSO) regression in the discovery set: polyethylene p-38:4, sphingomyelin d40:1, sphingomyelin d43:1, diglyceride 34:3, triacylglycerol 51:3, N-methylproline, and hydrocinnamic acid. All 7 metabolites except hydrocinnamic acid were available in the validation set.

Model 3 adjusted for the combination of predictors in Models 1 and 2.

<sup>2</sup>P value for pairwise comparisons of C-statistics calculated using DeLong's test.

ChemRICH analysis to overcome the inherent limitation of univariate analysis for individual metabolites. We observed that the glycerolipid cluster, which included metabolites positively associated with refined grain intake, was positively associated with GDM risk. In addition, clusters of glycerophospholipids and sphingolipids, with mixed associations of metabolites with refined grain intake (two-thirds negative and one-third positive), were overall negatively associated with GDM risk. These findings highlight the importance of assessing the comprehensive metabolomic profile using multivariate pathway enrichment methods to account for metabolite interplay.

Our findings were biologically plausible. Triacylglycerols, which constituted the majority of metabolites within the glycerolipid cluster and were positively associated with refined grain intake and GDM risk, might not directly inhibit insulin signaling [44]. Rather, lipid intermediates produced during triacylglycerol synthesis, such as diacylglycerol, which activates insulin-signaling inhibiting protein kinase C $\epsilon$ , and phosphatidic acid, which alters mTOR kinase activity, may interfere with the intracellular signaling pathway leading to insulin resistance [44]. Sphingolipids are involved in the cascade of intracellular signaling and cell recognition [45]. A case-control study and its follow-up experiments in mice and cells showed that a downregulation in sphingolipid metabolism was associated with pancreatic  $\beta$ -cell dysfunction and insulin resistance, potentially disrupting glucose homeostasis [46]. On the other hand, the mechanism behind the negative association of phosphatidylcholine and polyethylene, the most abundant phospholipids in mammalian cells, which constituted most

metabolites in the glycerophospholipids cluster in our study, with insulin sensitivity is less clear [47]. Reduced phosphatidylcholine levels mediated by hepatic knockout of the key enzyme for phosphatidylcholine biosynthesis led to triacylglycerol accumulation but had no effect on insulin sensitivity in a knockout mouse model [48]. Similarly, deletion of the rate-limiting enzyme for polyethylene production in mice increased levels of diacylglycerol but did not lead to insulin resistance [49].

Our study had several notable strengths. The study used untargeted metabolomics profiling, which is a promising tool for a comprehensive measurement of exogenous and endogenous metabolites in a biological fluid [50]. Compared to targeted metabolomics profiling, which is usually less expensive and follows a hypothesis-driven approach of metabolites of known identity, untargeted metabolomics may provide novel information on biological pathways with clinical relevance [51]. Moreover, we performed external validation for the refined grain-related metabolites and their associations with GDM risk, using samples assayed via the same metabolomic platforms following the same data generation and quality control procedures in a single laboratory, which ensures consistent measurements of metabolites and reduces technical variations. There were differences in participant characteristics between the discovery and validation sets; however, the heterogeneity between the 2 sets and higher C-statistics in the validation set demonstrated the robustness of our findings. Finally, our study also collected detailed data on a multitude of covariates to reduce potential residual confounding.



TABLE

Participant characteristics according to quartiles of refined grain intake in the discovery and validation sets of within the PETALS cohort (2014–2019)

Refined grain intake, quartiles	Discovery set (n = 261)				Validation set (n = 374)			
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Quartile 1	Quartile 2	Quartile 3	Quartile 4
<b>Total, n (%)</b>	65 (24.9)	65 (24.9)	66 (25.3)	65 (24.9)	93 (24.9)	93 (24.9)	95 (25.4)	93 (24.9)
<b>Refined grain intake, serving, mean ± SD</b>	1.1 ± 0.4	2.0 ± 0.2	3.0 ± 0.3	5.5 ± 2.1	1.3 ± 0.3	2.2 ± 0.2	3.3 ± 0.4	6.0 ± 2.4
<b>Age at delivery, years, n (%)</b>								
<25	3 (4.6)	6 (9.2)	4 (6.1)	7 (10.8)	12 (12.9)	8 (8.6)	8 (8.4)	19 (20.4)
25–29	12 (18.5)	15 (23.1)	9 (13.6)	14 (21.5)	19 (20.5)	28 (30.1)	24 (25.3)	26 (27.9)
30–34	30 (46.2)	31 (47.7)	31 (47.0)	24 (36.9)	38 (40.8)	37 (39.8)	43 (45.3)	28 (30.1)
≥35	20 (30.8)	13 (20.0)	22 (33.3)	20 (30.8)	24 (25.8)	20 (21.5)	20 (21.1)	20 (21.6)
<b>Race and ethnicity, n (%)</b>								
Asian/Pacific Islander	18 (27.7)	22 (33.8)	28 (42.4)	18 (27.7)	20 (21.4)	22 (23.6)	32 (33.6)	26 (28.1)
Non-Hispanic Black	5 (7.7)	5 (7.7)	5 (7.6)	9 (13.8)	19 (20.3)	12 (12.9)	18 (18.9)	28 (30.1)
Hispanic	25 (38.5)	18 (27.7)	19 (28.8)	18 (27.7)	24 (26.0)	22 (23.6)	27 (28.6)	24 (25.8)
Non-Hispanic White	15 (23.1)	14 (21.5)	11 (16.7)	17 (26.2)	30 (32.3)	37 (39.9)	18 (18.9)	15 (16.1)
Other/unknown	2 (3.1)	6 (9.2)	3 (4.5)	3 (4.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
<b>Education, n (%)</b>								
High school or less	5 (7.7)	6 (9.2)	10 (15.2)	7 (10.8)	7 (7.5)	9 (9.7)	12 (12.6)	19 (20.5)
Some college/associate degree	30 (46.2)	27 (41.5)	24 (36.4)	24 (36.9)	31 (33.4)	31 (33.3)	28 (29.4)	39 (42.0)
College degree or higher	30 (46.2)	32 (49.2)	32 (48.5)	34 (52.3)	55 (59.1)	53 (57.0)	55 (58.0)	35 (37.5)
<b>Nulliparity, n (%)</b>	31 (47.7)	31 (47.7)	28 (42.4)	30 (46.2)	52 (55.8)	48 (51.6)	35 (36.9)	39 (42.0)
<b>Prepregnancy BMI, kg/m<sup>2</sup>, n (%)</b>								
<18.5	2 (3.1)	0 (0.0)	2 (3.0)	1 (1.5)	4 (4.3)	0 (0.0)	2 (2.1)	2 (2.1)
18.5–24.9	24 (36.9)	23 (35.4)	19 (28.8)	28 (43.1)	38 (40.7)	43 (46.2)	33 (34.6)	38 (40.7)
25.0–29.9	26 (40.0)	15 (23.1)	20 (30.3)	16 (24.6)	28 (30.0)	20 (21.4)	32 (33.8)	27 (29.1)
≥30.0	13 (20.0)	27 (41.5)	25 (37.9)	20 (30.8)	23 (25.0)	30 (32.3)	28 (29.5)	26 (28.1)
<b>Chronic hypertension, n (%)</b>	5 (7.7)	6 (9.2)	2 (3.0)	2 (3.1)	7 (7.5)	4 (4.3)	5 (5.2)	8 (8.6)
<b>Family history of diabetes, n (%)</b>	11 (16.9)	18 (27.7)	16 (24.2)	20 (30.8)	20 (21.6)	23 (24.7)	28 (29.5)	19 (20.4)
<b>Smoking before and during pregnancy, n (%)</b>	1 (1.5)	0 (0.0)	4 (6.1)	3 (4.6)	2 (2.1)	4 (4.3)	7 (7.5)	4 (4.4)
<b>Alcohol use before and during pregnancy, n (%)</b>	31 (47.7)	40 (61.5)	39 (59.1)	37 (56.9)	52 (55.9)	51 (54.9)	52 (54.8)	48 (51.7)
<b>Total energy intake, kcal/d, mean ± SD</b>	1065.6 ± 421.0	1283.2 ± 346.1	1485.5 ± 415.6	2275.1 ± 864.1	1077.7 ± 383.3	1401.3 ± 361.4	1716.6 ± 513.0	2495.4 ± 875.0
<b>aHEI-P score, mean ± SD</b>	59.0 ± 11.8	57.5 ± 11.1	58.4 ± 9.9	61.1 ± 12.4	56.2 ± 11.7	60.4 ± 10.8	58.2 ± 9.5	63.3 ± 12.7
<b>Gestational weeks at blood collection, mean ± SD</b>	13.4 ± 2.3	13.5 ± 1.9	13.3 ± 2.1	13.5 ± 2.1	13.6 ± 0.2	13.3 ± 0.3	13.6 ± 0.3	13.5 ± 0.2
<b>Fasting status at blood collection, n (%)</b>	56 (86.2)	56 (86.2)	58 (87.9)	58 (89.2)	92 (98.9)	93 (100.0)	94 (99.0)	92 (98.9)

aHEI-P, Alternate Healthy Eating Index for Pregnancy.

Some study limitations are worth noting. First, refined grain intake was assessed using a conventional dietary assessment tool (FFQ), which may be subject to recall bias and measurement errors. However, the Block FFQ has been validated against 3 4-d diet records and demonstrated applicability to analyses on food group and nutrient levels among diverse populations, including pregnant individuals [16, 17]. Second, many food groups are intercorrelated, resulting in potential residual confounding. Although we controlled for overall dietary quality, changes in the concentrations of certain metabolites may be due to the consumption of food groups other than refined grains. Nonetheless, this metabolome-wide association analysis identified a comprehensive metabolic profile associated with refined grain intake and GDM risk, improving our understanding of the underlying metabolic processes. A short-term controlled feeding study among pregnant individuals can help validate the metabolomic signature of refined grains during pregnancy. Third, although we adjusted our analysis for race and ethnicity, we were not able to identify whether refined grain-related metabolites associated with GDM risk are racially and ethnically specific given our sample size. Future larger-scale studies are needed to help answer this question. Finally, since our study is observational, we cannot infer causality between refined grain intake, the associated metabolites, and the risk of GDM.

In conclusion, we identified 42 metabolites (mainly lipids and derivatives) associated with refined grain intake in early pregnancy, 17 of which were also prospectively associated with the risk of GDM. The significant associations of the metabolites belonging to glycerolipids, glycerophospholipids, and sphingolipid clusters with both refined grains and GDM risk and their incremental predictive value for GDM risk beyond the conventional risk factors potentially implicate these metabolites in the association between refined grain intake and GDM risk.

Our findings provide evidence on the potential biological underpinnings linking refined grain intake to the risk of GDM and help identify novel disease-related dietary biomarkers including N-methylproline and hydrocinnamic acid to inform GDM preventive strategies. Future studies are warranted to validate our findings and further elucidate the biological mechanisms for the associations between dietary intake and the risk of GDM.

We thank all the participants and research team members in the Pregnancy Environment and Lifestyle Study (PETALS). The authors' responsibilities were as follows—SZ: conceptualized and designed the study, conducted statistical analysis, and drafted the initial manuscript; RFC: revised the statistical analysis and the manuscript; ALN: contributed to statistical analysis, conducted a statistical review, and reviewed and revised the manuscript; OF: contributed to statistical

analysis and reviewed and revised the manuscript; EW, ALB, LL: critically reviewed the manuscript for important intellectual content; YZ: conceptualized and designed the study, coordinated and supervised the study, and critically reviewed the manuscript for important intellectual content; and all authors: contributed to data interpretation, editing, and critical review of the manuscript and approved the final manuscript. The authors report no conflicts of interest.

## Data Availability

Extracted data are available within the publication and its online supplement. A deidentified analytic dataset with the code used in this study can be shared with qualified researchers subject to approval by the Kaiser Foundation Research Institute Human Subjects Committee and by the Human Subjects Committee at the institutions requesting the data and a signed data sharing agreement. Please send all requests to the corresponding author of this article. Data will be available to requesters from 1 y after the date of publication of this article.

## Funding

This research was supported by the US National Institutes of Health Building Interdisciplinary Research Careers in Women's Health (BIRCWH) Program (grant K12HD052163) to YZ, National Institute of Diabetes and Digestive and Kidney Diseases (grant K01DK120807) to YZ, National Institute of Environmental Health Sciences (grants R01ES019196, U2CES026561, U2CES026555, P30ES023515, and U2CES030859) to AF, National Institute of Child Health and Human Development (grant R01HD073572) to AF, and the National Institutes of Health Office of Directors (grants UG3OD023289 and UH3OD023289) to AF. DKB was supported by National Center for Advancing Translational Sciences (grant UL1TR001433).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ajcnut.2023.02.009>.

## References

- [1] National Institute of Diabetes and Digestive and Kidney Diseases. [Internet]. Available from: <https://www.niddk.nih.gov/health-information/health-statistics/diabetes-statistics> (accessed September 29, 2021).
- [2] A. Ferrara, H.S. Kahn, C.P. Quesenberry, C. Riley, M.M. Hedderson, An increase in the incidence of gestational diabetes mellitus: Northern California, 1991-2000, *Obstet. Gynecol.* 103 (3) (2004) 526–533, <https://doi.org/10.1097/01.Aog.0000113623.18286.20>.
- [3] Y. Zhu, C. Zhang, Prevalence of gestational diabetes and risk of progression to type 2 diabetes: a global perspective, *Curr. Diab. Rep.* 16 (1) (2016) 7, <https://doi.org/10.1007/s11892-015-0699-xv>.
- [4] K. Giannakou, E. Evangelou, P. Yiallourou, C.A. Christophi, N. Middleton, E. Papatheodorou, et al., Risk factors for gestational diabetes: an umbrella review of meta-analyses of observational studies, *PLOS ONE* 14 (4) (2019), e0215372, <https://doi.org/10.1371/journal.pone.0215372>.
- [5] C.E. O'Neil, T.A. Nicklas, D.R. Keast, V.L. Fulgoni, Ethnic disparities among food sources of energy and nutrients of public health concern and nutrients to limit in adults in the United States: NHANES 2003-2006, *Food Nutr. Res.* 58 (2014), 15784, <https://doi.org/10.3402/fnr.v58.15784>.
- [6] D.A. Schoenaker, G.D. Mishra, L.K. Callaway, S.S. Soedamah-Muthu, The role of energy, nutrients, foods, and dietary patterns in the development of gestational diabetes mellitus: a systematic review of observational studies, *Diabetes Care* 39 (1) (2016) 16–23, <https://doi.org/10.2337/dc15-0540>.
- [7] M.J. Gibney, M. Walsh, L. Brennan, H.M. Roche, B. German, B. van Ommen, Metabolomics in human nutrition: opportunities and challenges, *Am. J. Clin. Nutr.* 82 (3) (2005) 497–503, <https://doi.org/10.1093/ajcn.82.3.497>.
- [8] O. Fiehn, Metabolomics—the link between genotypes and phenotypes, *Plant Mol. Biol.* 48 (1–2) (2002) 155–171, <https://doi.org/10.1007/978-94-010-0448-0>.
- [9] T. Sun, Y. Rong, X. Hu, Y. Zhu, H. Huang, L. Chen, et al., Plasma alkylresorcinol metabolite, a biomarker of whole-grain wheat and rye intake, and risk of type 2 diabetes and impaired glucose regulation in a Chinese population, *Diabetes Care* 41 (3) (2018) 440–445, <https://doi.org/10.2337/dc17-1570>.
- [10] X. Li, X. Cai, X. Ma, L. Jing, J. Gu, L. Bao, et al., Short- and long-term effects of wholegrain oat intake on weight management and glucolipid metabolism in overweight type-2 diabetics: a randomized control trial, *Nutrients* 8 (9) (2016) 549, <https://doi.org/10.3390/nu8090549>.
- [11] Y. Zhu, M.M. Hedderson, J. Feng, A.A. Mevi, A. Ferrara, The Pregnancy Environment and Lifestyle Study (PETALS): a population-based longitudinal multi-racial birth cohort, *BMC Pregnancy Childbirth* 17 (1) (2017) 122, <https://doi.org/10.1186/s12884-017-1301-0>.
- [12] N. Gordon, T. Lin, The Kaiser Permanente Northern California Adult Member Health Survey, *Perm. J.* 20 (4) (2016) 15–225, <https://doi.org/10.7812/tp/15-225>.
- [13] M.W. Carpenter, D.R. Coustan, Criteria for screening tests for gestational diabetes, *Am. J. Obstet. Gynecol.* 144 (7) (1982) 768–773, [https://doi.org/10.1016/0002-9378\(82\)90349-0](https://doi.org/10.1016/0002-9378(82)90349-0).
- [14] International Association of Diabetes and Pregnancy Study Groups Consensus Panel, B.E. Metzger, S.G. Gabbe, B. Persson, T.A. Buchanan, P.A. Catalano, et al., International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia during pregnancy, *Diabetes Care* 33 (3) (2010) 676–682, <https://doi.org/10.2337/dc09-1848>.
- [15] G. Block, M. Woods, A. Potosky, C. Clifford, Validation of a self-administered diet history questionnaire using multiple diet records, *J. Clin. Epidemiol.* 43 (12) (1990) 1327–1335, [https://doi.org/10.1016/0895-4356\(90\)90099-b](https://doi.org/10.1016/0895-4356(90)90099-b).
- [16] G. Block, C. DiSogra, WIC dietary assessment validation study, Final report. US Department of Agriculture, Food and Nutrition Service, Alexandria, VA, 1994.
- [17] B.A. Johnson, A.H. Herring, J.G. Ibrahim, A.M. Siega-Riz, Structured measurement error in nutritional epidemiology: applications in the Pregnancy, Infection, and Nutrition (PIN) study, *J. Am. Stat. Assoc.* 102 (479) (2007) 856–866, <https://doi.org/10.1198/016214506000000771>.
- [18] A. Ferrara, M.M. Hedderson, S.D. Brown, C.L. Albright, S.F. Ehrlich, A.L. Tsai, et al., The comparative effectiveness of diabetes prevention strategies to reduce postpartum weight retention in women with gestational diabetes mellitus: the Gestational Diabetes' Effects on Moms (GEM) cluster randomized controlled trial, *Diabetes Care* 39 (1) (2016) 65–74, <https://doi.org/10.2337/dc15-1254>.
- [19] NutritionQuest Assessment & Analysis Services. [Internet]. Available from: <https://www.nutritionquest.com/assessment/list-of-questionnaires-and-screens/> (accessed January 18 2021).
- [20] West Coast Metabolomics Center. Assays and services. [Internet]. Available from: <https://metabolomics.ucdavis.edu/core-services/assays-and-services> (accessed November 30 2020).
- [21] O. Fiehn, Metabolomics by gas chromatography-mass spectrometry: combined targeted and untargeted profiling, *Curr. Protoc. Mol. Biol.* 114 (2016) 30, <https://doi.org/10.1002/0471142727.mb3004s114>, 4.1–30.4.2.
- [22] K. Skogerson, G. Wohlgemuth, D.K. Barupal, O. Fiehn, The volatile compound BinBase mass spectral database, *BMC Bioinformatics* 12 (1) (2011) 321, <https://doi.org/10.1186/1471-2105-12-321>.
- [23] T. Cajka, J.T. Smilowitz, O. Fiehn, Validating quantitative untargeted lipidomics across nine liquid chromatography-high-resolution mass spectrometry platforms, *Anal. Chem.* 89 (22) (2017) 12360–12368, <https://doi.org/10.1021/acs.analchem.7b03404>.
- [24] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, et al., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, *Nat. Methods* 12 (6) (2015) 523–526, <https://doi.org/10.1038/nmeth.3393>.
- [25] T. Kind, K.H. Liu, D.Y. Lee, B. DeFelice, J.K. Meissen, O. Fiehn, LipidBlast in silico tandem mass spectrometry database for lipid identification, *Nat. Methods* 10 (8) (2013) 755–758, <https://doi.org/10.1038/nmeth.2551>.
- [26] B.C. DeFelice, S.S. Mehta, S. Samra, T. Cajka, B. Wanciewicz, J.F. Fuhrmann, O. Fiehn, Mass spectral feature list optimizer (MS-FLO): a tool to minimize false positive peak reports in untargeted liquid chromatography–mass spectrometry (LC-MS) data processing, *Anal. Chem.* 89 (6) (2017) 3250–3255, <https://doi.org/10.1021/acs.analchem.6b04372>.

- [27] S. Fan, T. Kind, T. Cajka, S.L. Hazen, W.H.W. Tang, R. Kaddurah-Daouk, et al., Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data, *Anal. Chem.* 91 (5) (2019) 3590–3596, <https://doi.org/10.1021/acs.analchem.8b05592>.
- [28] S.E. Chiuev, T.T. Fung, E.B. Rimm, F.B. Hu, M.L. McCullough, M. Wang, et al., Alternative dietary indices both strongly predict risk of chronic disease, *J. Nutr.* 142 (6) (2012) 1009–1018, <https://doi.org/10.3945/jn.111.157222>.
- [29] S.L. Rifas-Shiman, J.W. Rich-Edwards, K.P. Kleinman, E. Oken, M.W. Gillman, Dietary quality during pregnancy varies by maternal characteristics in Project Viva: a US cohort, *J. Am. Diet. Assoc.* 109 (6) (2009) 1004–1011, <https://doi.org/10.1016/j.jada.2009.03.001>.
- [30] L. Chasan-Taber, M.D. Schmidt, D.E. Roberts, D. Hosmer, G. Markenson, P.S. Freedson, Development and validation of a pregnancy physical activity questionnaire, *Med. Sci. Sports Exerc.* 36 (10) (2004) 1750–1760, <https://doi.org/10.1249/01.mss.0000142303.49306.0d>.
- [31] N. Pearce, Analysis of matched case-control studies, *BMJ* 352 (2016) i969, <https://doi.org/10.1136/bmj.i969>.
- [32] D. Hang, O.A. Zeleznik, X. He, M. Guasch-Ferre, X. Jiang, J. Li, et al., Metabolomic signatures of long-term coffee consumption and risk of type 2 diabetes in women, *Diabetes Care* 43 (10) (2020) 2588–2596, <https://doi.org/10.2337/dc20-0800>.
- [33] D.K. Barupal, O. Fiehn, Chemical similarity enrichment analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets, *Sci. Rep.* 7 (1) (2017), 14567, <https://doi.org/10.1038/s41598-017-15231-w>.
- [34] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics* 44 (3) (1988) 837–845.
- [35] L. Liang, M.H. Rasmussen, B. Piening, X. Shen, S. Chen, H. Röst, et al., Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women, *Cell* 181 (7) (2020) 1680–1692.e15, <https://doi.org/10.1016/j.cell.2020.05.002>.
- [36] I. Diboun, M. Ramanjaneya, Y. Majeed, L. Ahmed, M. Bashir, A.E. Butler, et al., Metabolic profiling of pre-gestational and gestational diabetes mellitus identifies novel predictors of pre-term delivery, *J. Transl. Med.* 18 (1) (2020) 366, <https://doi.org/10.1186/s12967-020-02531-5>.
- [37] M.L. Rahman, Y.A. Feng, O. Fiehn, P.S. Albert, M.Y. Tsai, Y. Zhu, et al., Plasma lipidomics profile in pregnancy and gestational diabetes risk: a prospective study in a multiracial/ethnic cohort, *BMJ Open Diabetes Res. Care* 9 (1) (2021), e001551, <https://doi.org/10.1136/bmjdr-2020-001551>.
- [38] D. Dudzik, M. Zorawski, M. Skotnicki, W. Zarzycki, G. Kozłowska, K. Bibik-Malinowska, et al., Metabolic fingerprint of gestational diabetes mellitus, *J. Proteomics* 103 (2014) 57–71, <https://doi.org/10.1016/j.jprot.2014.03.025>.
- [39] R. Lehmann, T. Friedrich, G. Krebichl, D. Sonntag, H.U. Häring, A. Fritsche, et al., Metabolic profiles during an oral glucose tolerance test in pregnant women with and without gestational diabetes, *Exp. Clin. Endocrinol. Diabetes* 123 (7) (2015) 483, <https://doi.org/10.1055/s-0035-1549887>, 438.
- [40] K.P. Law, X. Mao, T.L. Han, H. Zhang, Unsaturated plasma phospholipids are consistently lower in the patients diagnosed with gestational diabetes mellitus throughout pregnancy: a longitudinal metabolomics study of Chinese pregnant women part 1, *Clin. Chim. Acta* 465 (2017) 53–71, <https://doi.org/10.1016/j.cca.2016.12.010>.
- [41] J.C. Chai, G.C. Chen, B. Yu, J. Xing, J. Li, T. Khambaty, et al., Serum metabolomics of incident diabetes and glycemic changes in a population with high diabetes burden: the Hispanic Community Health Study/Study of Latinos, *Diabetes* 71 (6) (2022) 1338–1349, <https://doi.org/10.2337/db21-1056>.
- [42] E. Smith, F. Ottosson, S. Hellstrand, U. Ericson, M. Orho-Melander, C. Fernandez, et al., Ergothioneine is associated with reduced mortality and decreased risk of cardiovascular disease, *Heart* 106 (9) (2020) 691–697, <https://doi.org/10.1136/heartjnl-2019-315485>.
- [43] W. Kim, L.Y. Khil, R. Clark, S.H. Bok, E.E. Kim, S. Lee, et al., Naphthalenemethyl ester derivative of dihydroxyhydrocinnamic acid, a component of cinnamon, increases glucose disposal by enhancing translocation of glucose transporter 4, *Diabetologia* 49 (10) (2006) 2437–2448, <https://doi.org/10.1007/s00125-006-0373-6>.
- [44] C. Zhang, E.L. Klett, R.A. Coleman, Lipid signals and insulin resistance, *Clin. Lipidol.* 8 (6) (2013) 659–667, <https://doi.org/10.2217/clp.13.67>.
- [45] B.Y. Gueuvoghlian-Silva, F.B. Cordeiro, T.F. Lobo, T.R. Cataldi, E.G. Lo Turco, R.P. Bertolla, et al., Lipid fingerprinting in mild versus severe forms of gestational diabetes mellitus, *PLOS ONE* 10 (12) (2015), e0144027, <https://doi.org/10.1371/journal.pone.0144027>.
- [46] S.R. Khan, Y. Manialawy, A. Obersterescu, B.J. Cox, E.P. Gunderson, M.B. Wheeler, Diminished sphingolipid metabolism, a hallmark of future type 2 diabetes pathogenesis, is linked to pancreatic  $\beta$  cell dysfunction, *iScience* 23 (10) (2020), 101566, <https://doi.org/10.1016/j.isci.2020.101566>.
- [47] W. Chang, G.M. Hatch, Y. Wang, F. Yu, M. Wang, The relationship between phospholipids and insulin resistance: from clinical to experimental studies, *J. Cell Mol. Med.* 23 (2) (2019) 702–710, <https://doi.org/10.1111/jcmm.13984>.
- [48] R.L. Jacobs, C. Devlin, I. Tabas, D.E. Vance, Targeted deletion of hepatic CTP: phosphocholine cytidylyltransferase alpha in mice decreases plasma high density and very low density lipoproteins, *J. Biol. Chem.* 279 (45) (2004) 47402–47410, <https://doi.org/10.1074/jbc.M404027200>.
- [49] A. Selathurai, G.M. Kowalski, M.L. Burch, P. Sepulveda, S. Risis, R.S. Lee-Young, et al., The CDP-ethanolamine pathway regulates skeletal muscle diacylglycerol content and mitochondrial biogenesis without altering insulin sensitivity, *Cell Metab* 21 (5) (2015) 718–730, <https://doi.org/10.1016/j.cmet.2015.04.001>.
- [50] I. Bondia-Pons, T. Barri, K. Hanhineva, K. Juntunen, L.O. Dragsted, H. Mykkänen, et al., UPLC-QTOF/MS metabolic profiling unveils urinary changes in humans after a whole grain rye versus refined wheat bread intervention, *Mol. Nutr. Food. Res.* 57 (3) (2013) 412–422, <https://doi.org/10.1002/mnfr.201200571>.
- [51] M. Guasch-Ferré, S.N. Bhupathiraju, F.B. Hu, Use of metabolomics in improving assessment of dietary intake, *Clin. Chem.* 64 (1) (2018) 82–98, <https://doi.org/10.1373/clinchem.2017.272344>.