8-2023

Diet By Genomic Interactions In Determining Metabolic Traits And Profiles On Glucose Metabolism Among Mexican Americans In Starr County, Texas

Shinhye Chung
UTHealth School of Public Health

DIET BY GENOMIC INTERACTIONS IN DETERMINING METABOLIC TRAITS AND PROFILES ON GLUCOSE METABOLISM AMONG MEXICAN AMERICANS IN STARR COUNTY, TEXAS

by

SHINHYE CHUNG, MD, MS, MPH

APPROVED:

Jun, Goo
Digitally signed by Jun, Goo
Date: 2023.07.28 10:52:27 -05'00'

GOO JUN, PHD

Han Chen
Digitally signed by Han Chen
Date: 2023.07.31 15:25:16 -05'00'

HAN CHEN, PHD

yfu
Digitally signed by yfu
DN: dc=edu, dc=uthouston, ou=People, ou=SPH, ou=Faculty, cn=yfu
Date: 2023.07.31 08:08:42 -05'00'

YUN-XIN FU, PHD

chanis
Digitally signed by chanis
Date: 2023.07.28 12:22:08 -05'00'

CRAIG L. HANIS, PHD

DEDICATION

To My family

DIET BY GENOMIC INTERACTIONS IN DETERMINING METABOLIC TRAITS AND PROFILES ON GLUCOSE METABOLISM AMONG MEXICAN AMERICANS IN STARR COUNTY, TEXAS

by

SHINHYE CHUNG
MD, EWHA WOMANS UNIVERSITY, 2006
MS, EWHA WOMANS UNIVERSITY, 2012
MPH, THE UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER, 2017

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH
Houston, Texas
August 2023

ACKNOWLEDGEMENTS

I would like to sincerely express my gratitude to my academic advisor and dissertation supervisor, Dr. Goo Jun. He has dedicated his effort to my academic growth and development, providing me with direction and advice. I would like to express my great appreciation to Dr. Craig L. Hanis for his tremendous guidance and inspiration. I'm also thankful to Dr. Yun-Xin Fu and Dr. Han Chen on my dissertation committee and the external reviewer Dr. Bing Yu. I am always grateful for their assistance whenever I have questions, and it is an honor for me to have them on my dissertation committee. I appreciate all their time and effort once again.

I also thank Dr. Charles F. Burant, Dr. Charles Evans, and all other Michigan Regional Comprehensive Metabolomics Resource Cores staff for the metabolomic data. I am grateful to Dr. Cheryl L. Walker and all the staff in the Training in Precision Environmental Health Sciences (TPEHS) training program for their support. I sincerely appreciate all the professors and colleagues in the Starr Count Lab for their kind support and comments. I am thankful to all the study staff and the participants in Starr County.

This study is funded by The National Institute of Health R01DK11863, and TPEHS Training Program fellowship T32ES027801 supported Shinhye Chung for two years.

I am grateful for all the professors and friends met at UTHealth, including my MPH mentor, Dr. Marlyn A. Allicock. I thank my family, especially Young Un Choi and Samuel W. Choi. They have changed my life totally, and I cannot keep studying without their support. Last, I thank God for leading me to meet all the people above and to study another area, public health.

DIET BY GENOMIC INTERACTIONS IN DETERMINING METABOLIC TRAITS AND PROFILES ON GLUCOSE METABOLISM AMONG MEXICAN AMERICANS IN STARR COUNTY, TEXAS

Shinhye Chung, MD, MS, MPH, PhD
The University of Texas
School of Public Health, 2023

Dissertation Chair: GOO JUN, PhD

Worsening glycemia, prediabetes and diabetes, is one of the essential diseases in public health, considering their high prevalence, the enormous impact on multiple organs, and the economic burden on the community. Various factors can affect the pathogenesis of worsening glycemia, and this study focused on macronutrient intake, genes, and their interactions with metabolome data. 616 self-reported Mexican American participants in Starr County were recruited with informed consent. 308 identified and 2,471 unidentified metabolites were used for the analysis, and all the metabolites were inverse normalized with less than half of the missing rate. Each of the five glycemic and lipid traits was selected, and insulin and HOMA-IR were only log-transformed to correct skewed distribution. Macronutrient intake was calculated from 110-item food frequency questionnaires by the formula of nutrient density. All the analyses were adjusted for age, gender, and BMI as covariates. The analyses to find associations across glycemic and lipid traits, nutrients, and metabolites used linear regression models. We also compared the mean difference of metabolites across the glycemic status group with ANOVA model adjusted covariates. Genetic associations on the metabolites were calculated by GMMAT, and gene-environment

interactions were investigated by MAGEE. 3-hydroxybutyric acid, CAR (5:1), DG (18:1_18:1), DG (32:0), DG (32:1), DG (34:1), DG (34:2), PC (32:1), and 9 unidentified metabolites were associated with macronutrient intake, glycemic traits, and lipid traits. 28 identified and 232 unidentified metabolites were associated with specific SNPs in the cutoff of 5.0E-08. Among the metabolites, DG (32:1) was associated with the SNPs located on the *LRFN2* gene (top signal p-value 8.95E-09), and the other 16 metabolite-gene pairs were newly found. In the SNP-nutrient interactions, 13 SNP-nutrient interaction pairs on identified metabolites and 40 SNP- nutrient interaction pairs on unidentified metabolites were significant, but no significant SNPs overlapped compared to the GMMAT results. Moreover, more than half of the significant signals by MAGEE were located on noncoding DNA regions, so further study should be needed to reveal their functions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

BACKGROUND

Literature Review

Prediabetes and Diabetes: Definition and Classifications

Diabetes mellitus (DM) is defined as a high blood glucose state derived from the malfunction of the endocrine system (1). The essential organ of the endocrine system is the pancreas, so diabetes mellitus has mainly two different types depending on the underlying reason for the beta cell dysfunction in the pancreas (2). If the host immune system attacks an destroys the beta cell in the pancreas, this condition becomes the type I DM. Since this is one of the autoimmune diseases, the onset age of type I DM is usually under the 15-year-old, which is relatively younger than that of type II DM (3). Patients with type II DM do not show structural failure in the pancreas but have a functional loss in insulin secretion in the pancreas combined with the insulin resistance of the tissue (4). This dysfunction in glucose control is a chronic change, so there is a transitional and reversible status between normal and type II DM: prediabetes (3). Compared to diabetes documented in ancient Egyptian manuscripts and recorded since the 1st century, prediabetes is relatively recent, emerging in the late 1970s (5, 6).

Prediabetes and diabetes are diagnosed based on fasting glucose, 2-hour post-load glucose, and HbA1c (7). The American Diabetes Association(ADA) has established the following criteria for the diagnosis of diabetes: A fasting plasma glucose level measured after a minimum of 8 hours of fasting that is equal to or greater than 126 mg/dL (7.0 mmol/L); A 2-hour post-load glucose level checked after taking 75g anhydrous glucose with water that is equal to or greater than 200 mg/dL (11.1 mmol/L); A HbA1c level measured by NGSP

certified and standardized to the DCCT assay that is equal to greater than 6.5 percent (48 mmol/mol); A random plasma glucose that is equal to or greater than 200 mg/dL (11.1 mmol/L) with the symptoms of hyperglycemia or hyperglycemic crisis (7). If one meets any of these criteria, the person is diagnosed with diabetes. Prediabetes is also defined between normoglycemia and the detecting criteria of diabetes. The detailed standards are as follows: A fasting plasma glucose from 100 mg/dL (5.6 mmol/L) to 125 mg/dL (6.9mmol/L) or 2-hour post-load glucose from 140 mg/dL (7.8 mmol/L) to 199 mg/dL (11.0 mmol/L) or HbA1c level from 5.7 percent (39 mmol/mol) to 6.4 percent (47 mmol/mol) (7).

Prevalence and the burden of prediabetes and diabetes

Diabetes mellitus is one of the most common chronic diseases worldwide, and about 424.9 million (8.8 percent) people have diabetes in 2017 (8). Even though various preventive interventions have been developed to decrease worsening glycemia, the International Diabetes Federation (IDF) estimates that about 628.6 million people (9.9 percent) with diabetes based on the increasing trend in 2045 (8). Thus, the healthcare cost worldwide for diabetes was also estimated to increase from 232 billion U.S. dollars in 2007 to 727 billion U.S. dollars in 2017 (8). In both 1980 and 2014, the United States ranked third among 200 countries after China and India regarding the number of individuals affected by diabetes (9).

Diabetes mellitus was the seventh leading cause of death in the United States (U.S.) from 2015 to 2019 (10). Moreover, it is also a major risk factor for heart disease, the leading cause of death in the U.S. (10, 11). Likewise, diabetes mellitus can cause multiple organ damage and other conditions such as kidney failure, blindness, and peripheral neuropathy (12). According to the IDF Report 2017, the U.S. spent the largest healthcare cost for

diabetes compared to other 200 countries, with more than 11000 U.S. dollars as mean per capita spending per year (13). Thus, preventing diabetes is the main public health issue considering the economic and critical burden.

Nevertheless, lay people still fail to recognize the importance of blood sugar regulation. According to the National Diabetes Statistics Report, about one out of five adult patients with diabetes did not know their condition related to glucose control in 2018 (14). Recently, prediabetes has been considered an essential condition in preventing diabetes because it is a reversible condition to normal status by lifestyle modification. However, this specific health condition is also not well known to the public, so more than 80% of adults with prediabetes did not realize their glycemic intolerance status (14).

The risk of diabetes differs depending on the race, and Hispanics showed the second highest prevalence after American Indian/Alaska Native among US adults in 2017-2018 (14). Thus, diabetes is the fifth leading cause of death in the Hispanic U.S. population (11). In particular, Mexican Americans showed the highest prevalence of diabetes among Hispanics (14). Because the Hispanic or Latino population is the second-largest race in the U.S. in 2019, the burden of diabetes from this population needs to be decreased through proper preventive interventions (15).

The conundrum of prediabetes and diabetes

As described above, three different diagnostic criteria of prediabetes exist: HbA1c, fasting blood sugar, and 2-hour post-load oral glucose tolerance. Individuals diagnosed with HbA1c may be considered normal according to other standards, fasting blood sugar and 2-hour post-load glucose, due to variations in glucose metabolism that are not accounted for in

those standards. Table 1 shows the distribution of glucose intolerance groups divided by each diabetes diagnostic criteria in Starr County, Texas. The percentage of the normal group showed a variance from 42.7% to 61.6%, and that of the prediabetes group showed a variance from 27.6% to 51.1%. The number of participants diagnosed with diabetes by all three criteria above was only 15, yet 88 participants met at least one of the criteria.

Moreover, only 58 participants with prediabetes met all three criteria among 433 participants diagnosed with at least one standard. Lay people could be confused by these inconsistent results in understanding their glucose intolerance condition. Incomplete comprehension hinders individuals struggling with high blood glucose from adhering to the treatment, as lifestyle modification is the most important aspect of managing worsening glycemia.

Table 1. The distribution of glycemic status group in Starr County, Texas

	Fasting Glucose n (%)	2-hour OGTT n (%)	HbA1c n (%)
Normal	263 (42.7%)	373 (61.6%)	346 (56.2%)
Prediabetes	315 (51.1%)	167 (27.6%)	235 (38.1%)
Diabetes	38 (6.2%)	66 (10.9%)	35 (5.7%)
Total	616	606[a]	616

*Abbreviations: OGTT: Oral glucose tolerance test; HbA1c: Hemoglobin A1C
a. There were 10 missing people in the 2-hour OGTT criteria since they were already diagnosed with diabetes by fasting glucose, so the 2-hour OGTT did not perform for safety.

Another difficulty in diabetes care is the different diagnostic standards across the expert group. We explained the diagnostic standard from American Diabetes Association above, but WHO does not use HbA1c for the diagnosis of prediabetes, and the cutoff of normoglycemia based on the fasting plasma glucose is up to 110 mg/dL that is 10 mg/dL is higher than the standard from ADA (6). International Expert Committee(IEC) is another expert group that only uses HbA1c to detect prediabetes; the cutoff level of HbA1c is from 6.0 to 6.4 percent (6). Thus, the prevalence of prediabetes in the U.S. adults aged 20 or older can be changed from 4.3 percent to 43.5 percent in the same population analysis based on the National Health and Nutrition Examination Survey 2015-2016 data (6). Thus, we need additional knowledge about prediabetes and diabetes to improve the management of these specific conditions.

Diet and glucose metabolism

A diet should be considered in the analysis of glucose metabolism since glucose comes from the degradation of carbohydrates, fat, and protein, which are the essential nutrients of a diet. Previous studies have established the association between certain diet patterns and diabetes. For example, the “prudent” diet pattern, which consists of higher consumption of fruits and vegetables, shows a reduced risk of diabetes, and the “conservative” pattern, which contains butter, potatoes, and whole milk, is found to be associated with an increased risk of diabetes (16). Low-fat, Mediterranean, low-glycemic index, vegetarian, and lower-carbohydrate eating patterns are also suggested for improving glycemic controls and insulin resistance (17-19). Willett et al. proclaimed that the quality of a vegetable-based diet should be considered to improve glycemic control (20).

On the other hand, Panagiotakos et al. found the Mediterranean diet did not show a statistically significant difference in glucose, insulin, or homeostatic model assessment of insulin resistance (HOMA-IR) in multiple regression analysis (21). However, most of this nutritional research investigated non-Hispanic groups, so these specific diet patterns could be the barrier to dietary preventive intervention for Hispanics since they are unfamiliar with these diet patterns. Titus et al. assessed the barriers to managing type II DM for Hispanics, and the respondents counted their diet and exercise as the most difficult factors in their glucose management (22). The Hispanic Community Health Study/Study of Latinos(HCHS/SOL) searched how much Hispanic and Latino groups can follow the 2010 Dietary Guidelines for Americans (DGA), and sodium and fatty acid guidelines were the most vulnerable parts to follow (23). Since these two nutrients have been known to be important factors in glucose intolerance, these results suggest the need for advanced study and diet guidelines based on the nutrient components.

Gene and glucose metabolism

Another factor, genes, should also be considered since glucose metabolism in individuals can be different depending on the genetic variants of each individual. The heritability of type II DM is estimated from 30 to 70 percent, and about 20 percent can be counted from common variants (24). The variance of the metabolic response in each diet could be related to the genetic factor. So far, more than 700 loci have been recognized as the genetic variants associated with type 2 diabetes in the population-based single nucleotide polymorphisms(SNPs) analysis (24, 25). One of the essential genes related to type II DM is the *TCF7L2* gene located on chromosome 10q25.3, and this gene and Wnt signaling pathway

affect insulin resistance and lipid metabolism in adipocytes (2, 26, 27). Other essential genes associated with type II DM are *PPARG* on chromosome 3, *KCNJ11* on chromosome 11, *CDKAL1* on chromosome 6, *SLC30A8* on chromosome 8, *IGF2BP2* on chromosome 3, and *CDKN2 A/B* on chromosome 9 (2, 28-31). Glycemic traits related to type II DM have also been studied for their genetic association. *MTNR1B* on chromosome 11, *G6PC2* and *SPC25* on chromosome 2, and *GCK* on chromosome 7 are associated with fasting plasma glucose (32-36). GCK and TCF7L2 are also related to 2-hour post-load glucose, HBB on chromosome 11, HK1 on chromosome 10, and TMC6 and TMC8 on chromosome 17 show top signals associated with HbA1c in the Hispanic race (34, 37-40). However, most of the studies considered only genetic associations, and there are few studies investigating gene-environmental interaction on glycemic metabolism due to a lack of method. Recent advancements in statistical techniques have enabled us to identify gene-environmental interactions, so this study will investigate gene-nutrient interactions on metabolites.

Untargeted metabolomics

Metabolomics, a study investigating the metabolites of a biological system from cell to organism, has been used to reveal unknown metabolism in the human body (41). Metabolomics has the advantage of identifying proteins produced from human genes and understanding the byproducts generated in environments such as food or gut microbiomes (41). Because the metabolome has been studied since the late 1990s, researchers have investigated the metabolome to reveal unknown metabolites related to worsening glycemia (42). Leucine, Isoleucine, and aromatic amino acids showed significant associations with insulin resistance and type II DM (43). Wang-Sattler et al. found that glycine,

lysophosphatidylcholine (LPC) (18:2), and acetylcarnitine C2 are associated with prediabetes (44). Zeng et al. reported five biomarkers, 20-Hydroxy-leukotriene E4, Lysopc (20:4), 5-methoxytryptamine, Endomorphine-1, and Lysopc (20:3) are associated with the transition from prediabetes to normal glucose regulation in 108 participants(45). Guasch-Ferre et al. identified that isoleucine, leucine, valine, tyrosine, and phenylalanine were associated with an increased risk of type II DM in 2016 (46). Her research team added additional results in 2022, and alanine, glutamate, lysine, methionine, mannose, trehalose, pyruvate, four different forms of acylcarnitine, diacylglycerol, triacylglycerols, phosphatidylethanolamine, and ceramides were also positive associations with the risk of type II DM (47). Despite these findings, Hispanics were still an understudied population in metabolomics, so additional studies on this population are needed.

Public Health Significance

This study aims to discover how diet and genetic interaction can impact glucose metabolism through untargeted metabolomics. The metabolites that are significantly associated with diet and having genetic interaction could compensate for the gap in our knowledge about glucose metabolism. Diet is one of the controllable essential factors in glucose metabolism, so this additional knowledge could also provide scientific evidence to create more effective prediabetes prevention strategies considering the individual genetic factor. With more effective prediabetes prevention strategies, the prevalence of diabetes in the vulnerable population should decrease because prediabetes is a condition that can be reversed to normal status.

Hypothesis, Research Question, Specific Aims or Objectives

This study investigated the diet and genetic interaction on glycemic metabolism with untargeted metabolomics data. Based on this knowledge, the long-term goal of this study is to establish scientific evidence to develop preventive interventions for prediabetes. In particular, this study includes diet, which can be modified by various methods such as cooking classes and dietary education. Hence, if there are any novel metabolites on glycemic metabolism related to diet and genetic interaction, it could be considerable surrogate outcomes to modify the diet plan for preventing prediabetes. The objective is to identify diet by genetic interaction on glucose metabolism with untargeted metabolomics data. The central hypothesis is that diet and genetic interactions can affect metabolites in glucose metabolism. Based on this hypothesis, the aims are:

Aim 1: To identify plasma metabolites associated with prediabetes-related traits and macronutrient intake among Mexican Americans in Starr County, Texas.

Aim 2: To identify genetic variants affecting plasma metabolites.

Aim 3: To investigate the interaction between genetic variation and nutrient intake on plasma metabolites.

This study is expected to identify diet by genetic interaction on glucose metabolism. Based on this outcome, it could provide additional knowledge on glucose metabolism. As mentioned above, the gap between the different standards and metabolism of glucose intolerance could be explained via diet by genetic interaction. Moreover, this discovery could suggest scientific evidence for developing dietary prevention for prediabetes. For example, if

a high-fat diet with specific genes is attributed to glucose intolerance, dietary prevention could be constructed by a low-fat diet for people with particular risk genes.

METHODS

Study Design

A cross-sectional design is selected for this study, despite its inherent weaknesses such as temporal ambiguity. However, this design offers advantages in terms of cost and time efficiency. Additionally, it enables us to estimate the impact of the disease on the general population. Given that prediabetes and diabetes are common diseases and the participants share common factors such as diet and genes, a cross-sectional design can be suitable for this study. While measurement error resulting from plasma traits may be a significant concern, we can mitigate its impact using a standardized manual for measuring them. The data collected for this study is primary data obtained from participants residing in Starr County.

Study Subjects and Setting

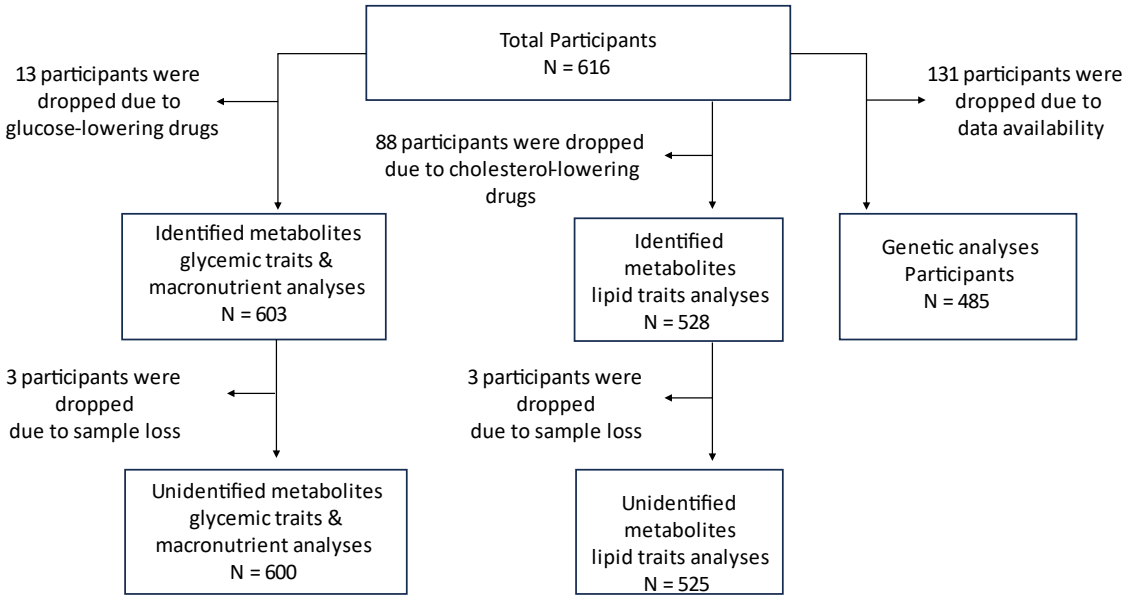
Participants were recruited by the local staff in Starr County from 2018 to 2019. All the local staff were trained to measure the anthropometric measurement and to ask questions about basic descriptive characteristics and food frequency questionnaires in standardized forms. All the standardized questionnaires were provided in English and Spanish.

A total of 616 participants without prior diagnosis of diabetes were recruited. Since we found people newly diagnosed with diabetes through the visits to the local office, the participants were divided into three groups, normal, prediabetes, and diabetes. Normal group means the participants who show a normal range in all three diagnostic criteria of diabetes in the American Diabetes Association. Recruiting criteria excluded those with previously known diagnoses of diabetes, but 83 people showed their glycemic traits in the diagnostic criteria of diabetes on their first visit. In the investigation of medication use, 13 participants

were found using glucose-lowering medicines for losing weight or other purposes. Thus, they were excluded from the analysis, and 603 participants were investigated in the analyses of identified metabolite data. The demographic characteristics of the participants are shown in the supplemental table S1. The normal group shows the largest proportion of the age group under 40 years old (Normal vs. Prediabetes vs. Diabetes: 18.42% vs. 10.87% vs. 4.82%) and the least proportion of the age group above 60 compared to other groups (Normal vs. Prediabetes vs. Diabetes: 11.18% vs. 14.67% vs. 19.28%). 71% of the participants are female, and the normal group shows a larger proportion of females than others (Normal vs. Prediabetes vs. Diabetes: 81.58% vs. 67.12% vs. 68.67%). Since glucose intolerance is associated with obesity, the diabetes group shows the largest proportion of body mass index (BMI) greater than 30, which is the diagnostic standard of obesity (Normal vs. Prediabetes vs. Diabetes: 38.82% vs. 63.32% vs. 80.72%). There was a time gap between receiving identified metabolites and unidentified metabolites data and three missing samples on unidentified metabolites. Therefore, 600 participants were used for the analyses of the unidentified metabolites. For the lipid traits, 88 participants had taken cholesterol-lowering drugs, so only 528 participants were included in the investigation related to lipid traits with identified metabolites, and 525 participants were analyzed with unidentified metabolites. Also, 12 participants showed high triglyceride, so the Fridewald equation could not calculate their low-density lipoprotein (LDL) concentration. Thus, these participants were only removed in the analysis to find the association between metabolites and calculated LDL levels. Lastly, the sample size for the genetic association study (Aim 2) and gene-nutrient

interaction study (Aim 3) was limited to only 485 participants due to the available data. Figure 1 summarizes how the samples were selected in each analysis.

Figure 1. Summary of the sample size in each analysis



The exposure of this study is diet and gene, and the outcome is the plasma metabolites associated with exposure and prediabetes-related traits. Age, sex, and body mass index were used as covariates for all the analyses.

Nutrient intake and food sources were obtained using a semi-quantitative food frequency questionnaire and transformed into nutrient density in macronutrients. The 110-item food frequency questionnaires were previously tested and validated in this same population, and their correlations compared to 3-day food records were 0.77 for energy, 0.76

for total fat, and 0.61 for saturated fat (48). The participants were educated about how to record their food frequency questionnaires from the trained local staff.

Genetic variants were determined previously from imputed genome-wide association study (GWAS) data in this population. The trained nurses collected the blood of the participants to gather metabolomics and glycemic traits data.

Prediabetes-associated glycemic traits are HbA1c, fasting glucose, fasting insulin, HOMA-IR, and 2-hour post-load glucose, and these data were gathered from the EDTA plasma sample. Before collecting the glycemic trait samples, the participants were directed to fast overnight and come to the assigned sampling location. The blood of participants was obtained for their fasting glucose status at first, and the participants ingested the 75g glucose load immediately after the fasting blood test. After 2 hours from the ingestion of the glucose load, the blood of participants had collected once again for assessing 2-hour post glucose loads. The collected blood was centrifuged, and the plasma sample was extracted from the EDTA tube and used for further analysis. HOMA-IR is calculated by the following formula (49):

HOMA-IR = fasting glucose in mmol/L* fasting insulin in μU/mL/22.5

Since worsening glycemia is also associated with obesity, five different lipid traits – cholesterol, triglycerides (TG), calculated LDL, high-density lipoprotein(HDL), and non-HDL- were also investigated to find the associations with metabolites(50). LDL is computed using the Fridewald equation (51):

Calculated LDL = Total cholesterol − HDL − (Triglycerides/5)

Untargeted metabolomic profiles are obtained from EDTA plasma analyzed at the Michigan Regional Comprehensive Metabolomics Resource Core(Ann Arbore, MI, USA). The obtained EDTA samples were stored in the -80 ℃ freezer and sent in a frozen state to the metabolomics core. The details on the metabolomic data process using Liquid Chromatography with Tandem Mass Spectrometry (LC-MS) are available in (52). The metabolomic data for this study were checked to determine whether there was any batch effect across the tests, and there was no specific batch effect on the whole metabolomic data in both identified and unidentified metabolites. During the metabolomic data generation, we found the isomers of some metabolites were measured separately. Thus, those metabolites were marked with "_a" or "_b," and the tested reversed-phase method was shown as "rp".

Data Analysis

Aim 1. Identifying plasma metabolites associated with prediabetes-related traits and macronutrients

Formatted self-administered questionnaires collected descriptive data such as age and sex, and body mass index was calculated based on the measured height and weight by the standardized measurement equipment. This descriptive data is shown with mean and standard deviation in Table 2. ANOVA tests were used to test whether there is any significant statistical difference in the quantitative data, such as age and BMI, between the (pre)diabetes groups. Sex was described with the proportion of each sex, and the statistical difference was tested with a chi-square test. The significance level of all the analyses in descriptive data is 0.05. This table is summarized in the appendix table, and age, sex, and BMI show statistically significant differences across the glucose intolerance groups.

Metabolites with more than half of the missing values are omitted from the analysis, and all other metabolites are inverse-normalized. Thus, 308 identified metabolites and 2,471 unidentified metabolites were used for all the analyses in this study. The Bonferroni method was used to correct all the significance levels to adjust for multiple testing. Therefore, 1.62E-04 (0.05/308) was used as the significance threshold for the identified metabolite data, and 2.02E-05 (0.05/2,471) was used as the significance threshold for unidentified metabolite data. Missing values in each variable were coded as NA and modified by available case analysis. The strength of this method is that it can provide more power for each analysis than a complete case analysis and use real data.

Five prediabetes-related glycemic traits were tested in this study. Fasting plasma glucose, 2-hour post-load glucose, and HbA1c showed normal distributions, but insulin and HOMA-IR showed right-skewed distributions (Insulin: Shapiro-Wilk test p-value: <2.2E-16, Skewness 2.07, Kurtosis 9.46; HOMA-IR Shapiro-Wilk test p-value: <2.2E-16, Skewness 2.0, Kurtosis 10.10). Thus, these two traits were log-transformed and applied to the analyses. Linear regression models were used to identify the association between the plasma metabolites and prediabetes-associated glycemic and lipid traits adjusted to age, sex, and BMI. Each group with different glycemic status was also coded using discrete values for different linear regression analyses by coding the group with normal glycemia as 1, prediabetes as 2, and diabetes as 3. Each glycemic group was also compared by the ANOVA model adjusted by age, sex, and BMI.

Diet data was transformed into the nutrient density in macronutrient categories such as carbohydrates, fat, and protein. Nutrient density was calculated based on the energy from each macronutrient (53). The formula for nutrient density for macronutrients is as follows:

(Grams of total component intake *X) / Kcal of total energy intake

X is the kcal for 1 gram of each macronutrient. Thus, X becomes four kcal/g for carbohydrates and protein and nine kcal/g for all kinds of fat; total fat, saturated fat, monounsaturated fat, and polyunsaturated fat. In this study, the energy from alcohol was not included in the total energy intake. The associations between metabolites and nutrient density of macronutrients were analyzed with linear regression models adjusted age, sex, and BMI. All the analyses were calculated with R 4.1.2. version.

Aim 2. Identifying genetic variants affecting plasma metabolites

Generalized linear Mixed Model Association Tests (GMMAT) were used to investigate genetic variants affecting plasma metabolites to address the cryptic relatedness and heterogeneity in the population structure (54). We used GWAS array data imputed into the 1000 Genomes Project Phase 3 reference panel. The imputation was done by the Michigan imputation server. Metabolites and covariates – age, sex, and BMI- were constructed as a data frame column, and each participant became a row of the data frame. The kinship matrix was built by the GEMMA R package (55). Genetic data were converted to a PLINK bed file for analysis, and major allele frequency was limited from 0.01 to 0.5. Since the metabolite is the quantitative trait, the GMMAT model set up a Gaussian family. All 308 identified metabolites and 2,471 unidentified metabolites were analyzed, and the cutoff value to determine significant genetic-associated metabolites was set to the nominal

genome-wide significance threshold of 5.0E-08. The Manhattan plots were generated by the R package qqman. After finding significantly associated single nucleotide polymorphisms(SNPs), each SNP position was checked to investigate the gene in the UCSC genome browser on Human (GRCh37/hg19) (56). R 4.2.2 version was used for the analyses.

Aim 3. Investigating the interaction between genetic variation and nutrient intake on plasma metabolites

Diet by gene interaction effects on metabolites was modeled using generalized linear models and analyzed by Mixed Model Association Test for GEne-Environment Interaction (MAGEE) (57). Since MAGEE also examines the gene-environment interactions based on the GLMMs from the GMMAT R package, the analyses could be consistent with the previous aim. The running time of MAGEE took more than that of GMMAT, so we selected the metabolites that showed significant associations with any macronutrients, glycemic traits, or genes. Therefore, 145 identified metabolites and 687 unidentified metabolites were tested to find the gene-nutrient intake interaction on metabolites. The same imputed GWAS data on aim 2 was used for this gene-nutrient interaction, and major allele frequencies were limited from 0.01 to 0.5. MAGEE can generate p-value for both interaction and joint tests, so the first cutoff p-value was based on the p-value of the interaction test. However, the p-value of the joint test was also considered to select the significant gene-environment signals. The cutoff value of the p-value is 5.0E-08. After the analysis, we found that the average genomic control(lambda) values were slightly more inflated than 1.0. Therefore, the metabolites were selected based on lambda values from 0.95 to 1.15 in the Q-Q plot of p-value interaction. R 4.2.2 version was used for the analyses.

Ethical Considerations

The data of this study is a part of NIH grants DK118631 Project with IRB approval. All participants in this study willingly enrolled and were informed of their right to withdraw from the study at any time and to decline invasive procedures, such as blood sampling. They were provided with informed consent, which outlined the benefits and risks associated with the study. In case of any uncertainties regarding the study process, participants were encouraged to ask questions to the local staff for clarification. Individuals displaying significantly abnormal laboratory results, such as excessively high plasma glucose levels, were promptly referred to a local physician to prevent potential life-threatening events. The local staff made efforts to minimize invasiveness during sample collection procedures. Furthermore, all data obtained from the study was treated as confidential.

RESULTS

Aim 1. Plasma metabolites associated with prediabetes-related traits and macronutrient intake

Table 2 is the summary of the demographic characteristics of the participants in this study. The group with diabetes showed older age and higher BMI than the group with normal glycemic status, and the percentage of males in the group with diabetes was higher than the group with normal glycemic levels. Although the mean energy intake of the group with diabetes is higher than that of the group with normal glycemic status, it did not show statistically significant differences across the glycemic level. Each macronutrient nutrient density was tested to determine whether there was any significant difference, but any macronutrient nutrient density did not show significant differences across the glycemic status.

Table 2. Demographic characteristics of the participants based on glycemic status

Characteristics	Normal (N= 152)	Prediabetes (N=368)	Diabetes (N=83)	Univariate P-value	Adjusted P-value
Age(years)[a]	47.97 ± 8.13	50.47± 7.93	51.29 ± 7.95	**0.0005[b]**	
Sex (no.(%))					
Male	28 (18.42)	121 (32.88)	26 (31.3)	**0.0038[c]**	
Female	124 (81.58)	247 (67.12)	57 (68.7)		
BMI (kg/m^2)[a]	29.25 ± 5.58	32.64 ± 6.34	35.54 ± 7.21	**<0.0001[b]**	
Energy intake and Nutrient density					
Energy (Kcal)[a]	2001.73 ± 1173.54	2190.02 ± 1323.95	2232.94 ± 1346.12	0.134[b]	0.116[d]
Protein (%)[a]	16.75 ± 5.00	17.22 ± 3.97	16.72 ± 3.64	0.880[d]	0.879[e]
Carbohydrates (%)[a]	46.24 ± 10.46	45.02 ± 8.86	45.58 ± 7.81	0.507[d]	0.503[e]
Total Fat (%)[a]	37.01 ± 7.20	37.76 ± 6.37	37.69 ± 5.76	0.406[d]	0.403[e]
Saturated Fat (%)[a]	11.03 ± 2.45	11.17 ± 2.27	11.04 ± 1.85	0.977[d]	0.977[e]
Monounsaturated Fat (%)[a]	15.40 ± 3.33	15.65 ± 2.91	15.76 ± 2.61	0.390[d]	0.390[e]
Polyunsaturated Fat (%)[a]	7.30 ± 1.82	7.64 ± 1.79	7.63 ± 2.35	0.102[d]	0.101[e]

a Mean ± S.D.
b p-value from univariate ANOVA model
c p-value from the chi-square test
d p-value from ANOVA model adjusted age, sex, and BMI
e p-value from ANOVA model adjusted energy intake, age, sex, and BMI

a. Identified metabolites

a.1.Glycemic traits and glycemic status group

A total of 23 metabolites, including branched amino acids such as leucine and isoleucine, showed associations with fasting glucose. Additionally, 49 metabolites were found to be associated with 2-hour post-load glucose, while 14 metabolites were related to HbA1c. 49 metabolites were associated with log-transformed insulin, and 51 metabolites were related to log-transformed HOMA-IR (Supplemental Table S2-S6). Notably, all metabolites associated with HbA1c were also related to at least one other measure of glycemic traits. Table 3 summarizes the metabolites list associated with three diagnostic glycemic features; fasting glucose, 2-hour post-load glucose, and HbA1c. Behenic acid was positively associated with fasting glucose, but choline was negatively associated with fasting glucose (Behenic acid: β 0.00676, p-value 7.67E-05; Choline: β -0.007, p-value 2.28E-05). Most metabolites were positively associated with each glycemic trait, but phosphatidylcholine (PC) (35:3) showed a negative association with 2-hour post-load glucose (β -0.00385, p-value 2.79E-05). 2-deoxyglucose was also negatively associated with fasting glucose and HbA1c, and glutamine exhibited negative associations with all three diagnostic glycemic traits (2-deoxyglucose: fasting glucose β -0.0089 p-value 6.57E-07, HbA1c β -0.2445 p-value 3.23E-07; glutamine: fasting glucose β -0.0069 p-value 5.53E-05, 2-hour post-load glucose β -0.0039 p-value 2.21E-05, HbA1c β -0.1962 p-value 2.69E-05). 3-methyl-2-oxovaleric acid, DG (18:1_18:1), DG (32:0), DG (34:1), DG (34:2), glutamine, ketoleucine, and leucine were found to be associated with all five glycemic traits. These eight metabolites also showed significant associations with the hyperglycemic groups in the linear

models with glycemic status, ANOVA models comparing the normal and diabetes groups, and ANOVA models comparing the non-diabetes and diabetes groups (Supplemental tables 7-10). Additionally, 34 metabolites showed significant differences when comparing the normal group to all hyperglycemic statuses, including prediabetes and diabetes.

Table 3. Summary of three diagnostic glycemic traits associated with identified metabolites list

Fasting glucose only associated metabolites	2-hour post-load glucose only associated metabolites	Fasting glucose and 2-hour post-load glucose associated metabolites	Fasting glucose and HbA1c associated metabolites	All three diagnostic glycemic traits associated metabolites
Behenic acid Choline	Y-Glutamylisoleucine Arachidic acid Citramalic acid DG(32:1) Docosapentaenoic acid Docosatrienoic acid Docosenoic acid Dodecenoic acid Eicosadienoic acid Eicosatetraenoic acid Eicosatrienoic acid Eicosenoic acid EPA Hydroxydodecanoic acid Hydroxyhexadecanoic acid Hydroxyphenyllactic acid Margaric acid MG (14:0) Myristoleic acid N-Acetylneuraminic acid Octadecadienoic acid Octadecatrienoic acid Oleic acid Palmitic acid Palmitoleic acid PC (32:1) PC (34:4) PC (35:3) Stearic acid	2-Hydroxybutyric acid Y-Glutamylleucine DHA Docosatetraenoic acid Heptadecanedioic acid Hydroxytetradecanoic acid Nonadecenoic acid	2-Deoxyglucose	3-Hydroxybutyric acid 3-Methyl-2-oxovaleric acid DG (18:1_18:1) DG (32:0) DG (34:1) DG (34:2) DG (36:3) Glutamine Isoleucine Ketoleucine Leucine Leucine-Isoleucine MG (16:0)_rp_a

a. underlined metabolites: negative association
b. Abbreviation: DG: Diacylglycerol; EPA: Eicosapentaenoic acid; MG: Monoacylglycerol; PC Phosphatidylcholine; DHA: Docosahexaenoic acid

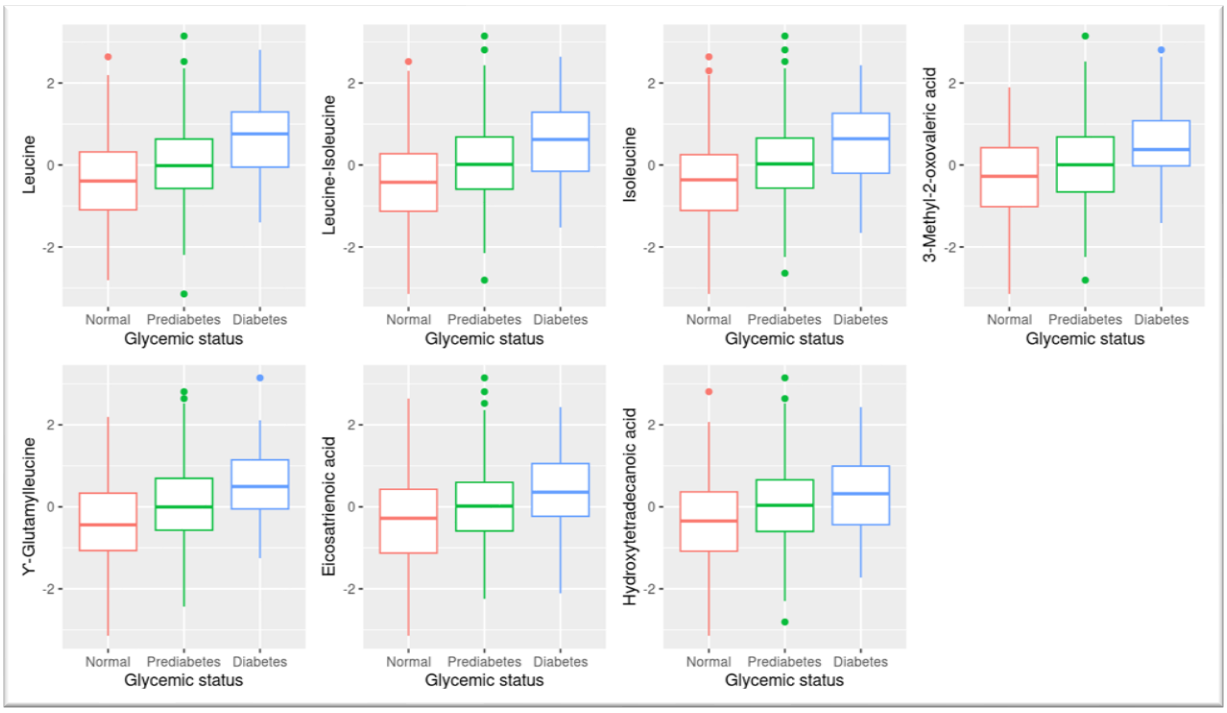
Table 4 summarizes 19 metabolites that exhibit a significant mean difference between the normal and prediabetes groups, as determined by an ANOVA model adjusted for age, sex, and BMI. Among these metabolites, leucine, leucine-isoleucine, isoleucine, 3-methyl-2-oxovaleric acid, γ-glutamylleucine, eicosatrienoic acid, and hydroxytetradecanoic acid demonstrated to increase as worsening glycemia. Figure 2 provides a visual representation of the distribution of these seven metabolites across the various glycemic groups. On the other hand, N-Acetylserine, S-Allylcysteine, N-Acetylleucine, Proline-Phenylalanine, and Phenylalanine-Tryptophan exhibited a significant difference solely in the comparison between the normal and prediabetes groups.

Table 4. Comparison of mean difference of identified metabolites between normal and prediabetes groups

Metabolites	p-value
Leucine-Isoleucine	7.08E-08
5'-Methylthioadenosine	1.84E-07
Isoleucine	2.39E-07
Ɣ-Glutamyltyrosine	2.45E-07
Ɣ-Glutamylleucine	3.43E-07
N-Acetylserine	5.28E-07
Uric acid	9.24E-07
Leucine	1.23E-06
S-Allylcysteine	1.27E-05
N-Acetylleucine	1.39E-05
Ɣ-Glutamylisoleucine	2.85E-05
Proline-Phenylalanine	4.57E-05
Phenylalanine -Tryptophan	5.01E-05
3-Methyl-2-oxovaleric acid	5.83E-05
Glutamic acid- Phenylalanine	5.91E-05
Hydroxytetradecanoic acid	9.57E-05
Eicosatrienoic acid	0.000106
3-Methylbutyrylcarnitine	0.000118
CAR(5:0) isomers	0.000152

a Abbreviation: CAR: Carnitine

Figure 2. Identified metabolites significantly differed in comparing normal and prediabetes with a linear trend



a.2. Lipid traits

We found 37 identified metabolites associated with cholesterol, 14 metabolites related to HDL, 39 metabolites associated with triglyceride, 14 metabolites associated with calculated LDL, and 42 metabolites related to non-HDL (Supplemental tables 11-15). No metabolites were associated with all five lipid traits together, but DG (18:1_18:1), DG (32:0), DG (32:1), DG (34:1), DG (34:2), DG (36:3), and MG (18:1) were found to be statistically positively associated with plasma cholesterol, triglycerides, non-HDL and negatively associated with HDL. Table 5 summarizes these associations in detail. Sphingomyelin (SM)(d32:2) was the only metabolite that overlapped between HDL and calculated LDL, but it did not show any significant association with triglycerides. A comprehensive summary of all the results about

the association between the metabolites and lipid traits can be found in supplemental tables 11-15.

Table 5. Summary of identified metabolites associated with cholesterol, triglycerides, non-HDL, and HDL

	Cholesterol	Triglycerides	Non-HDL	HDL
DG (18:1_18:1)				
β	0.0082	0.0073	0.0125	-0.0396
p-value	5.50E-10	9.53E-67	2.07E-21	6.18E-21
DG (32:0)				
β	0.0081	0.0073	0.0113	-0.0276
p-value	4.55E-10	3.06E-68	8.57E-18	1.06E-10
DG (32:1)				
β	0.0060	0.0075	0.0100	-0.0368
p-value	4.41E-06	2.21E-72	5.10E-14	2.27E-18
DG (34:1)				
β	0.0077	0.0077	0.0120	-0.0394
p-value	2.95E-09	1.26E-78	2.61E-20	2.29E-21
DG (34:2)				
β	0.0063	0.0079	0.0115	-0.0493
p-value	1.57E-06	5.39E-84	2.15E-18	1.09E-33
DG (36:3)				
β	0.0066	0.0075	0.0119	-0.0497
p-value	8.63E-07	3.56E-68	1.14E-18	1.37E-32
MG (18:1)				
β	0.0083	0.0067	0.0113	-0.0219
p-value	2.20E-08	1.66E-42	1.76E-13	1.08E-05

a Abbreviation: DG: Diacylglycerol; MG: Monoacylglycerol

a.3. Macronutrients

Table 6 summarizes the metabolites associated with each nutritional intake, with 9 metabolites showing associations with specific macronutrient nutrient densities. Diacylglycerol (DG) (32:1) showed a positive association with carbohydrate intake but a

negative association with total fat, monounsaturated fat, and cholesterol. Another form of diacylglycerol, DG (18:1_18:1), was also positively associated with carbohydrate nutrient density. Similarly, phosphatidylcholine (PC) (32:1) showed a positive association with carbohydrate nutrient density but negative associations with total fat, monounsaturated fat, and polyunsaturated fat. Conversely, aminobutyric acid exhibited a negative association with carbohydrate nutrient density but a positive association with total fat nutrient density. Carnitine (5:1) and 3-hydroxybutyric acid were positively associated with protein nutrient density. Although DG (32:0), DG (32:1), DG (34:1), DG (34:2), DG (18:1_18:1), and 3-hydroxybutyric acid did not show a statistically significant difference in the comparison between normal and prediabetes, these metabolites exhibited a continuous increase as the glycemic status worsened (Figure 3).

Figure 3. Identified metabolites associated with nutrient intake and worsening glycemia status

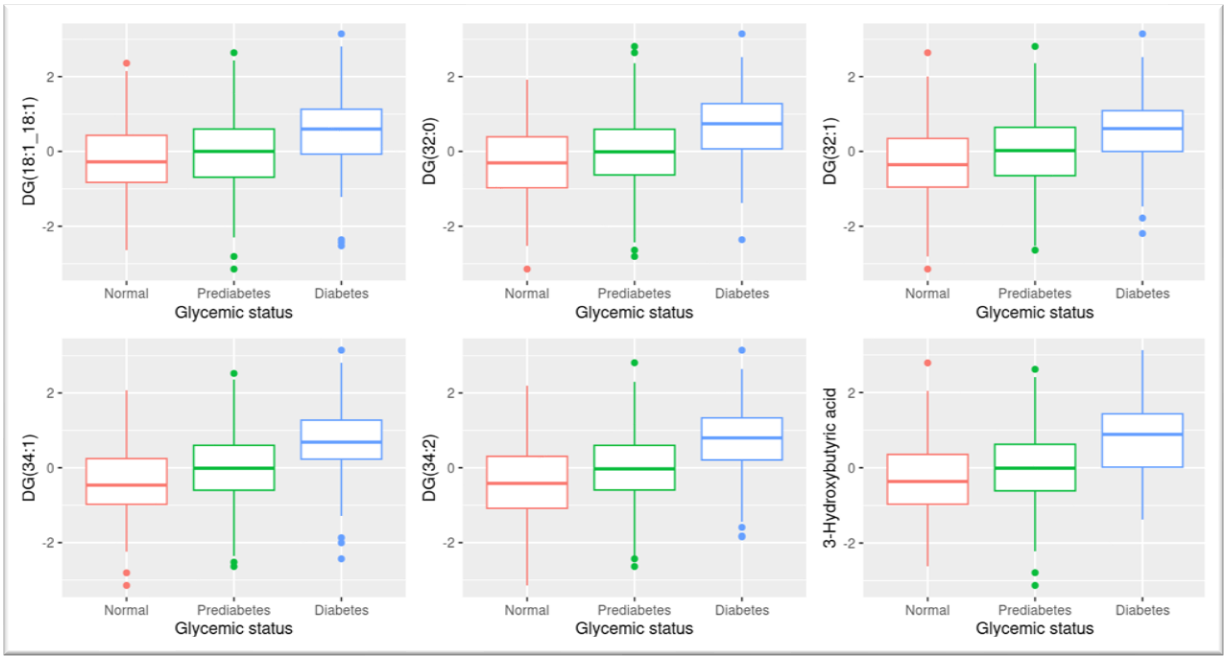


Table 6. Association between nutrient density and identified metabolites

Nutrient density	Metabolites	Standardized β	P-value
Carbohydrates	DG(32:1)	0.0207	1.67E-06
	Aminobutyric acid	-0.0186	2.70E-05
	PC(32:1)	0.0168	1.27E-04
	DG(18:1_18:1)	0.0168	1.37E-04
Protein	CAR(5:1)	0.0427	4.15E-06
	3-Hydroxybutyric acid	0.0380	1.06E-04
Total fat	DG(32:1)	-0.0278	4.78E-06
	PC(32:1)	-0.0252	4.05E-05
	Aminobutyric acid	0.0246	7.68E-05
Monounsaturated fat	DG(32:1)	-0.0715	5.12E-08
	PC(32:1)	-0.0559	2.74E-05
	DG(32:0)	-0.0524	7.28E-05
	DG(34:2)	-0.0521	8.30E-05
	DG(34:1)	-0.0507	1.21E-04
Polyunsaturated fat	PC(32:1)	-0.0808	1.23E-04

a Abbreviation: CAR Carnitine; DG Diacylglycerol; PC Phosphatidylcholine

b. Unidentified metabolites

b.1. Glycemic traits and glycemic status group

Among 2,471 unidentified metabolites, 47 were associated with fasting glucose, and 208 were related to 2-hour post-load glucose. 29 metabolites were significantly associated with HbA1c, and 253 were associated with log-transformed insulin. 241 metabolites were related to log-transformed HOMA-IR. Table 7 and Figure 4 summarize unidentified metabolites associated with three diagnostic glycemic traits. 19 metabolites were associated with all three diagnostic glycemic traits and summarized in table 6, but no metabolites were related to 2-hour post-load glucose and HbA1c.

Figure 4. The number of unidentified metabolites associated with three diagnostic glycemic traits

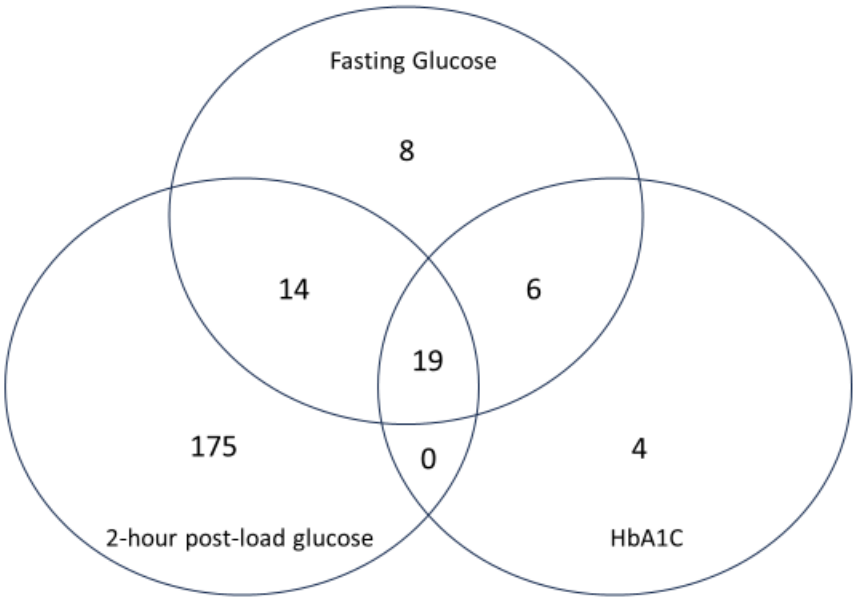


Table 7. 19 unidentified metabolites associated with all three diagnostic glycemic traits

Unidentified metabolites	Fasting plasma glucose		2-hour post-load glucose		HbA1c	
	β	p-value	β	p-value	β	p-value
UNK_441.0745_0.632	0.0167	3.47E-24	0.0045	4.43E-07	0.3466	3.30E-14
UNK_340.1072_2.37	0.0170	2.73E-25	0.0048	5.47E-08	0.3381	1.27E-13
UNK_277.0742_0.946	0.0167	1.22E-26	0.0051	9.88E-10	0.3483	1.11E-15
UNK_115.0401_2.585	0.0103	3.48E-10	0.0063	4.65E-13	0.2345	1.7E-07
UNK_154.0839_1.838	0.0084	1.56E-07	0.0049	1.37E-08	0.2058	2.37E-06
UNK_86.0967_1.836	0.0078	4.72E-07	0.0044	1.41E-07	0.1879	8.62E-06
UNK_132.1023_1.96	0.0086	1.81E-08	0.0045	5.67E-08	0.2138	3.44E-07
UNK_634.4449_12.16	-0.0090	1.04E-07	-0.0045	6.19E-07	-0.2572	2.22E-08
UNK_86.0968_1.96	0.0077	3.61E-07	0.0045	3.13E-08	0.1894	5.07E-06
UNK_154.0839_1.959	0.0088	1.84E-08	0.0043	2.14E-07	0.2056	1.40E-06
UNK_432.2383_1.956	0.0078	2.59E-07	0.0040	1.06E-06	0.1773	1.91E-05
UNK_455.2266_1.957	0.0079	2.99E-07	0.0042	4.56E-07	0.1874	8.64E-06
UNK_183.0274_2.587	0.0093	2.38E-08	0.0064	3.49E-13	0.2119	3.05E-06
UNK_253.0695_2.583	0.0101	7.87E-10	0.0069	1.52E-15	0.2256	4.97E-07
UNK_315.0407_2.584	0.0101	8.88E-10	0.0066	4.72E-14	0.2180	1.30E-06
UNK_349.0819_2.12	0.0102	3.29E-09	0.0068	4.62E-14	0.2207	2.70E-06
UNK_568.3814_1.928	-0.0086	7.20E-08	-0.0043	5.14E-07	-0.1956	7.57E-06
UNK_229.0695_2.119	0.0091	1.21E-07	0.0060	4.55E-11	0.2030	1.72E-05
UNK_411.044_1.065	0.0108	1.30E-10	0.0039	1.48E-05	0.2465	9.13E-08

In the analyses of the comparison across the glycemic statuses, the linear model with each glycemic group identified the significant associated 73 metabolites, and 87 metabolites showed significant mean difference with the comparison between normal and prediabetes groups by the ANOVA model. Additionally, the analyses found 196 metabolites associated with the comparison between normal versus diabetes, 140 metabolites related to the comparison between normal versus the group with hyperglycemia, including prediabetes and diabetes together, and 87 metabolites associated with the comparison between non-diabetic group (Normal and the group with prediabetes) and the group with diabetes. Table 8 summarizes the 17 unidentified metabolites list associated with all these comparisons across the groups with different glycemic statuses.

Table 8. Unidentified metabolites significantly associated with the comparisons across the groups with different glycemic statuses

Unidentified Metabolites	Associated other traits
UNK_594.3774_11.499	2HR, log-transformed insulin, log-transformed HOMA-IR, BMI
UNK_788.5804_14.159	2HR, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, BMI
UNK_146.0462_0.63	2HR, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, BMI
UNK_477.1044_1.958	2HR, log-transformed insulin, log-transformed HOMA-IR, non-HDL
UNK_409.2345_9.292	2HR, log-transformed insulin, log-transformed HOMA-IR, TG, BMI
UNK_439.255_1.956	2HR, log-transformed insulin, log-transformed HOMA-IR, TG, non-HDL, BMI
UNK_273.1109_8.681	2HR, BMI
UNK_613.3594_9.663	2HR, BMI
UNK_1036.0408_2.585	FPG, 2HR, log-transformed insulin, log-transformed HOMA-IR, Cholesterol, HDL, TG, non-HDL, BMI
UNK_132.1023_1.836	FPG, 2HR, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, BMI
UNK_463.0889_0.959	FPG, 2HR, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, non-HDL, BMI
UNK_202.1086_5.755	FPG, 2HR, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, non-HDL, BMI
UNK_215.0333_0.64	FPG, HbA1c, BMI
UNK_432.2382_1.836	FPG, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, BMI
UNK_439.255_1.837	FPG, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, BMI
UNK_301.1431_1.836	FPG, log-transformed insulin, log-transformed HOMA-IR, HDL, TG, BMI
UNK_547.3289_1.927	FPG, log-transformed insulin, log-transformed HOMA-IR, TG, non-HDL, BMI

*Abbreviation: FPG: Fasting plasma glucose; 2HR: 2-hour post-load glucose

b.2. Lipid traits

Five different lipid traits were tested to find the association between unidentified metabolites and lipid traits by linear regression models adjusting covariates. 148 metabolites were associated with cholesterol, and 45 metabolites were related to HDL. The analyses also found 231 metabolites associated with triglyceride, 70 metabolites related to calculated LDL, and 149 metabolites associated with non-HDL. Among the unidentified metabolites associated with any lipid traits, the metabolites related to any glycemic traits were summarized in Table 8 above.

b.3. Macronutrients

A total of 30 different unidentified metabolites were associated with at least one macronutrient intake by the linear regression models adjusted covariates. 15 unidentified metabolites were significantly associated with more than one macronutrient intake. Thus, 20 unidentified metabolites were associated with carbohydrate intake, and 5 were related to protein intake. Total fat intake was associated with 12 unidentified metabolites, saturated fat with 4, monounsaturated fat with 15, and polyunsaturated fat with 4. Table 9 summarizes the results of the significant association between macronutrients and unidentified metabolites.

Table 9. Unidentified metabolites associated with macronutrients intake

Nutrient	Unidentified Metabolites	β	p-value
Carbohydrates	UNK_812.5807_13.719	-0.029	2.55E-11
Carbohydrates	UNK_838.596_14.365	-0.025	1.68E-08
Carbohydrates	UNK_840.6118_14.553	-0.023	1.58E-07
Carbohydrates	UNK_810.5651_13.586	-0.023	2.36E-07
Carbohydrates	UNK_842.5312_14.302	-0.022	4.33E-07
Carbohydrates	UNK_818.5314_14.432	-0.022	7.93E-07
Carbohydrates	UNK_522.1793_7.702	-0.021	1.31E-06
Carbohydrates	UNK_854.592_13.697	0.021	1.86E-06
Carbohydrates	UNK_512.1888_8.096	-0.021	1.97E-06
Carbohydrates	UNK_836.5795_13.891	-0.021	2.41E-06
Carbohydrates	UNK_908.5997_14.55	-0.021	2.73E-06
Carbohydrates	UNK_172.133_1.067	-0.020	4.55E-06
Carbohydrates	UNK_764.5585_14.913	-0.020	8.23E-06
Carbohydrates	UNK_688.492_13.189	0.020	8.54E-06
Carbohydrates	UNK_816.5879_14.523	-0.020	9.54E-06
Carbohydrates	UNK_211.1702_9.742	-0.020	1.01E-05
Carbohydrates	UNK_768.5546_14.354	0.019	1.07E-05
Carbohydrates	UNK_488.1922_8.038	-0.019	1.28E-05
Carbohydrates	UNK_856.6088_14.313	0.019	1.95E-05
Carbohydrates	UNK_712.4921_13.037	0.019	1.96E-05
Protein	UNK_812.5807_13.719	0.050	1.72E-07
Protein	UNK_842.5312_14.302	0.046	1.09E-06
Protein	UNK_838.596_14.365	0.043	7.84E-06
Protein	UNK_818.5314_14.432	0.042	1.36E-05
Protein	UNK_255.0878_6.244	0.041	1.77E-05
Total Fat	UNK_812.5807_13.719	0.037	2.89E-09
Total Fat	UNK_838.596_14.365	0.031	5.55E-07
Total Fat	UNK_840.6118_14.553	0.031	7.04E-07
Total Fat	UNK_211.1702_9.742	0.030	1.80E-06
Total Fat	UNK_908.5997_14.55	0.028	5.43E-06
Total Fat	UNK_836.5795_13.891	0.028	5.43E-06
Total Fat	UNK_810.5651_13.586	0.028	5.75E-06
Total Fat	UNK_688.492_13.189	-0.027	7.99E-06
Total Fat	UNK_768.5546_14.354	-0.027	1.14E-05
Total Fat	UNK_692.487_10.741	-0.026	1.71E-05
Total Fat	UNK_854.592_13.697	-0.027	1.80E-05
Total Fat	UNK_522.1793_7.702	0.026	2.00E-05
Saturated Fat	UNK_522.1793_7.702	0.088	8.19E-07

Saturated Fat	UNK_571.2809_9.24	0.083	5.05E-06
Saturated Fat	UNK_838.596_14.365	0.079	1.31E-05
Saturated Fat	UNK_812.5807_13.719	0.078	1.87E-05
Monounsaturated Fat	UNK_812.5807_13.719	0.081	1.42E-09
Monounsaturated Fat	UNK_838.596_14.365	0.070	1.85E-07
Monounsaturated Fat	UNK_840.6118_14.553	0.070	2.42E-07
Monounsaturated Fat	UNK_836.5795_13.891	0.068	3.89E-07
Monounsaturated Fat	UNK_810.5651_13.586	0.067	8.31E-07
Monounsaturated Fat	UNK_761.5806_13.137	0.064	2.85E-06
Monounsaturated Fat	UNK_211.1702_9.742	0.064	2.93E-06
Monounsaturated Fat	UNK_908.5997_14.55	0.062	4.81E-06
Monounsaturated Fat	UNK_688.492_13.189	-0.060	6.81E-06
Monounsaturated Fat	UNK_748.5149_12.666	-0.060	1.06E-05
Monounsaturated Fat	UNK_764.5585_14.913	0.060	1.10E-05
Monounsaturated Fat	UNK_692.487_10.741	-0.059	1.15E-05
Monounsaturated Fat	UNK_768.5546_14.354	-0.059	1.16E-05
Monounsaturated Fat	UNK_722.4985_12.556	-0.058	1.78E-05
Polyunsaturated Fat	UNK_508.3414_10.993	-0.105	8.67E-07
Polyunsaturated Fat	UNK_790.5593_13.439	-0.095	5.18E-06
Polyunsaturated Fat	UNK_723.4015_6.717	-0.096	6.80E-06
Polyunsaturated Fat	UNK_1040.6362_12.167	-0.089	7.33E-06

Aim 2. Genetic variants affecting plasma metabolites

a. Identified metabolites

A total of 28 metabolites showed statistically significant genetic associations on 1337 SNPs and 34 genes by GMMAT adjusted age, gender, and BMI. The lambda values of the QQ plot of significant identified metabolites were from 0.970 to 1.015. Among these 28 metabolites, biliverdin, bilirubin, CAR(4:0), DG (32:1), eicosatetraenoic acid, EPA, N-acetylleucine, and 2-hydroxybutyric acid were associated with glycemic traits, CAR(6:0), DG(32:1), eicosatetraenoic acid, LPC(10:4)_rp_b, LPE(18:0)_rp_a, PC(34:2), and pipecolic acid were associated with lipid traits, and LPC(20:0) and 2-aminooctanoic acid were related to BMI only. 23 metabolites-gene association pairs were repetitive from previous other studies. Moreover, there were novel findings between the identified metabolites and specific SNPs. DG (32:1) showed significant signals in *Leucine Rich Repeat And Fibronectin Type III Domain Containing 2(LRFN2) gene* on chromosome 6, and CAR (4:0) was associated with multiple SNPs on the *CABP1* and *SPPL3* genes. CAR (6:0) was known to be related to *SLAC44A5, ACADM, RABGGTB,* and *MSH4* genes, and a new association was found on 1:76392437 on the *ASB17* gene on chromosome 1. CAR (9:0), LPC(20:0), LPC(20:4)_rp_b, LPE(18:0)_rp_a, LPE(18:2)_rp_b_1, methyl-3-hydroxybenzoate, and sphingosine also found new genetic associations on single variants. Table 10 summarizes the top signals on GMMAT, and Figure 5 illustrates the example QQ plots and Manhattan plots for each metabolite.

Table 10. Summary of top signals of genetic-associated identified metabolites and related genes

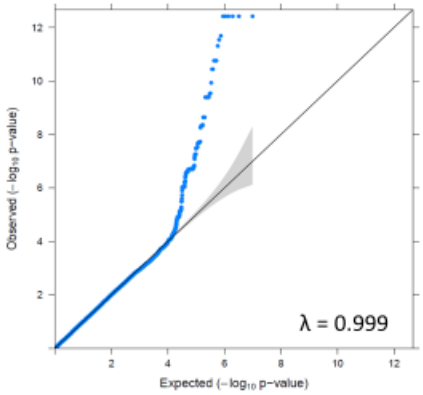
Metabolite	SNP	A1	A2	AF	SCORE	P-value	Gene
Bilirubin	2:234664586	A	ATC	0.70	-104.42	3.78E-13	*UGT1A*
Bilirubin	2:234665983	G	A	0.70	-104.42	3.78E-13	*UGT1A*
Biliverdin	2:234664586	A	ATC	0.70	-117.164	2.96E-15	*UGT1A*
Biliverdin	2:234665983	G	A	0.70	-117.164	2.96E-15	*UGT1A*
CAR(10:1)	1:76125211	A	C	0.71	82.7998	2.88E-09	*SLC44A5*
CAR(10:1)	1:76159225	A	G	0.71	81.4464	5.26E-09	*CR936677*
CAR(10:1)	1:76203479	AT	A	0.72	78.6027	5.97E-09	*ACADM*
CAR(4:0)	12:121176083	A	G	0.67	-191.866	4.67E-41	*ACADS*
CAR(4:0)	12:121155622	T	C	0.71	-169.191	1.03E-34	*UNC119B*
CAR(4:0)	12:121130046	G	A	0.70	-170.211	5.23E-34	*MLEC*
CAR(4:0)	12:121084587	G	A	0.67	99.3045	1.40E-11	***CABP1***
CAR(4:0)	12:121200609	C	T	0.60	99.1632	2.44E-11	*SPPL3*
CAR(4:0)	12:121144144	C	G	0.58	88.9084	1.59E-08	-
CAR(6:0)	1:76168340	G	A	0.71	98.7168	4.80E-12	*CR936677*
CAR(6:0)	1:76106961	T	A	0.70	98.8327	4.91E-12	*SLC44A5*
CAR(6:0)	1:76192582	T	C	0.71	97.0787	7.16E-12	*ACADM*
CAR(6:0)	1:76259677	A	ACCTAA GAGTGA GACTTAA CCCACTT TTAAATT GTTCT	0.73	84.0802	1.29E-09	*RABGGTB*
CAR(6:0)	1:76392437	C	T	0.74	81.3674	4.79E-09	***ASB17***
CAR(6:0)	1:76353294	C	T	0.72	78.3059	3.78E-08	*MSH4*
CAR(8:0)	1:76125211	A	C	0.71	84.6046	1.47E-09	*SLC44A5*
CAR(8:0)	1:76203479	AT	A	0.72	80.8179	2.43E-09	*ACADM*
CAR(8:0)	1:76159225	A	G	0.71	82.1185	4.44E-09	*CR936677*
CAR(9:0)	2:211074909	C	T	0.72	-134.98	1.82E-23	*ACADL*
CAR(9:0)	2:211007287	C	T	0.66	-141.331	3.48E-22	***KANSL1L***
CAR(9:0)	2:210878117	T	C	0.65	-132.894	4.09E-20	***RPE***
CAR(9:0)	2:210846713	C	T	0.59	-117.31	9.58E-15	***UNC80***
CAR(9:0)	2:211156513	T	TA	0.60	-118.271	2.79E-14	***MYL1***
Cholic acid	9:98344706	T	A	0.83	63.229	2.97E-08	-
DG(32:1)	6:40532893	C	G	0.57	87.0426	8.95E-09	***LRFN2***
Dodecadienoic acid	4:149061102	T	C	0.97	30.2017	3.93E-08	***NR3C2***
Eicosatetraenoic acid	11:61603510	C	A	0.57	-85.2784	2.61E-09	*FADS2*
EPA	11:61609750	C	T	0.58	-78.5615	4.37E-08	*FADS2*
LPC(20:0)	2:70693652	C	A	0.84	53.4816	2.78E-08	***TGFA***
LPC(20:4)_rp_b	11:61603510	C	A	0.57	-151.442	1.43E-25	*FADS2*

LPC(20:4)_rp_b	11:61537648	T	TG	0.76	-81.1916	2.12E-10	*MYRF*
LPC(20:4)_rp_b	11:46130135	TATAA	T	0.65	84.5867	7.21E-09	***PHF21A***
LPC(20:4)_rp_b	11:45939934	G	C	0.75	72.7936	2.36E-08	***PEX17***
LPC(20:4)_rp_b	11:45944186	TCGGGGG	T	0.75	72.7936	2.36E-08	***LARGE2***
LPC(20:4)_rp_b	11:45934549	AT	A	0.72	76.9388	3.02E-08	***PEX16***
LPE(18:0)_rp_a	1:115252909	TATAA	T	0.75	58.6886	2.02E-08	***NRAS***
LPE(18:2)_rp_b_1	11:61603510	C	A	0.57	95.1667	4.37E-11	*FADS2*
Methyl-3-hydroxybenzoate	22:34113903	C	A	0.85	64.3352	2.87E-08	***LARGE1***
N-Acetylleucine	2:73634622	TA	T	0.78	-74.2579	3.23E-12	*ALMS1*
N-Acetylleucine	2:73867862	C	G	0.77	-74.5299	1.15E-11	*NAT8*
PC(34:2)	11:61603510	C	A	0.57	87.8791	1.42E-09	*FADS2*
PC(36:2)	11:46250265	GTAGA	G	0.64	79.2683	3.94E-08	-
PC(40:8)	11:61603510	C	A	0.57	-116.079	5.30E-16	*FADS2*
Pipecolic acid	4:181225524	T	C	0.62	-80.4734	4.10E-08	-
Sphingosine	1:200702431	C	T	0.59	95.2974	7.89E-10	***CAMSAP2***
Tetracosenoic acid	22:23786757	A	T	0.54	-87.3145	2.27E-08	-
1-Methylxanthine	8:18272377	C	T	0.66	87.4537	4.79E-09	-
1,3-Dimethyluric acid	13:112843248	T	C	0.63	-74.7115	4.44E-08	-
2-Aminooctanoic acid	2:73805172	G	A	0.78	93.8725	2.74E-14	*ALMS1*
2-Aminooctanoic acid	2:73867862	C	G	0.78	92.4061	8.21E-14	*NAT8*
2-Aminooctanoic acid	2:73845709	T	C	0.57	99.6788	7.40E-12	-
2-Aminooctanoic acid	2:73928366	G	C	0.56	94.1025	1.97E-10	*NAT8B*
2-Aminooctanoic acid	2:73959960	G	T	0.57	91.2276	7.66E-10	*TPRKB*
2-Hydroxybutyric acid	8:76334681	G	C	0.95	37.6949	2.40E-08	***HNF4G***
2-Hydroxybutyric acid	8:76319588	A	G	0.95	35.7059	4.89E-08	-

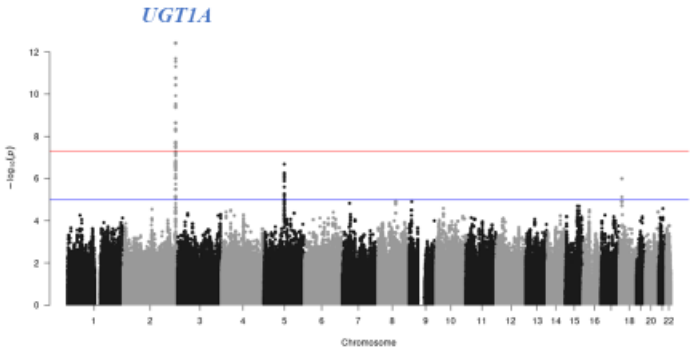
a. Boldly marked gene: possible novel association in this study.

Figure 5. Examples of QQ and Manhattan plots of genetically associated identified metabolites

(a) Bilirubin

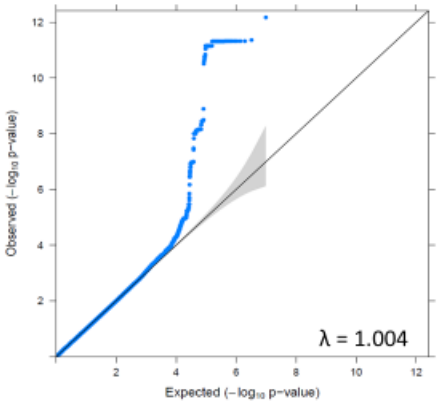


QQ plot of bilirubin (GMMAT)

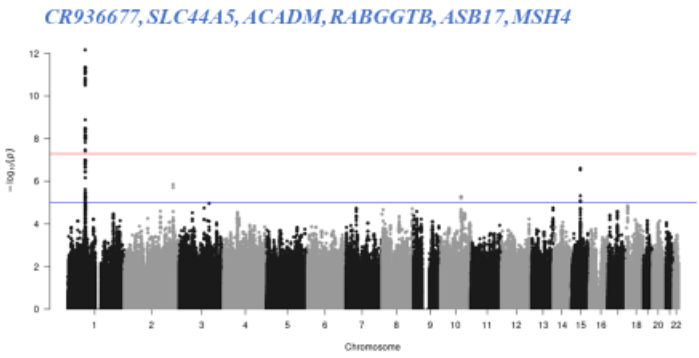


Manhattan plot on bilirubin (GMMAT)

(b) CAR (6:0)

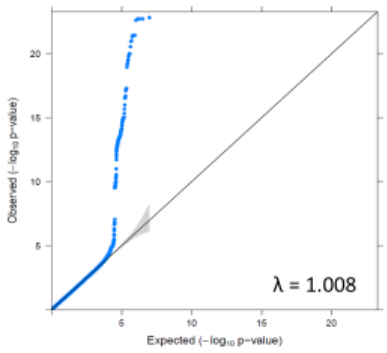


QQ plot of CAR(6:0) (GMMAT)

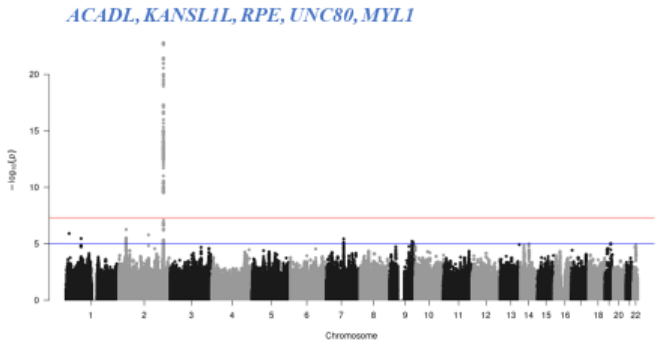


Manhattan plot on CAR(6:0) (GMMAT)

(c) CAR (9:0)

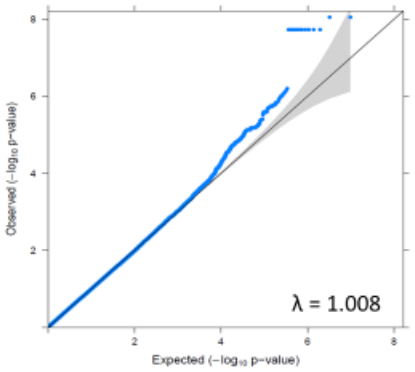


QQ plot of CAR(9:0) (GMMAT)

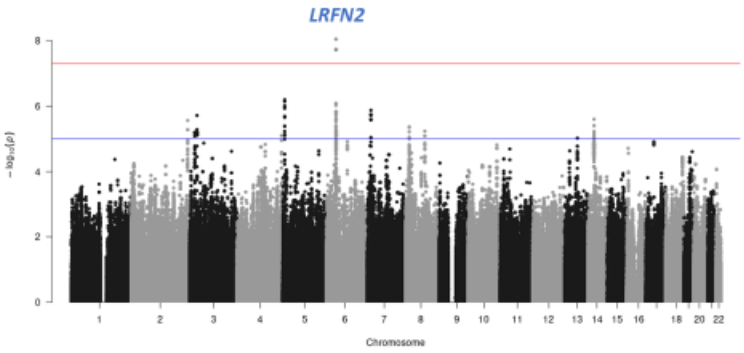


Manhattan plot on CAR(9:0) (GMMAT)

(d) DG (32:1)



QQ plot of DG(32:1) (GMMAT)



Manhattan plot on DG(32:1) (GMMAT)

b. Unidentified metabolites

A total of 232 metabolites showed significant genetic associations with 2,117 distinct SNPs by GMMAT adjusted age, gender, and BMI. Since these unidentified metabolites did not have detailed information yet, we selected 32 metabolites associated with any glycemic traits or the comparison across the groups with glycemic status. A total of 11 different genes were located on the significant SNPs associated with these 32 metabolites. These metabolites list and their genetically related SNPs are summarized in Table 11. Bilirubin and biliverdin were significantly associated with the SNPs located in *UGT1A* family genes on chromosome 2 among identified metabolites, and UNK_255.1138_6.985, UNK_299.1393_7.568, UNK_341.1108_6.44, UNK_341.1109_6.621, UNK_341.1109_6.735, and UNK_357.0772_6.62 were also genetically associated with the SNPs on these genes. Figure 6 illustrates the examples of QQ and Manhattan plots of unidentified metabolites.

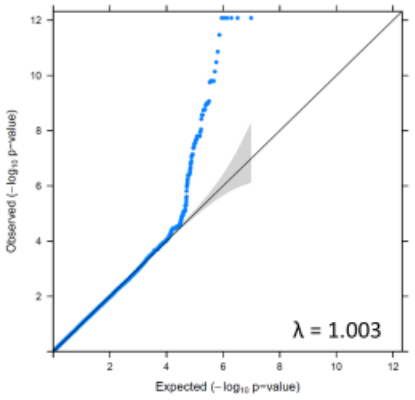
Table 11. Summary of top signals of genetic-associated unidentified metabolites and related genes

Metabolites	SNP	A1	A2	AF	SCORE	VAR	p-value	Gene
UNK_1195.8522_11.509	1:40922240	C	T	0.65	87.90	255.493	3.81E-08	*ZFP69B*
UNK_139.113_8.618	18:58392436	A	G	0.51	-87.58	247.574	2.60E-08	-
UNK_152.009_1.214	4:10005435	G	A	0.74	85.57	233.877	2.20E-08	*SLC2A9*
UNK_255.1138_6.985	2:234664586	A	ATC	0.70	-82.36	213.214	1.70E-08	*UGT1A*
UNK_255.1138_6.985	2:234665983	G	A	0.70	-82.36	213.214	1.70E-08	*UGT1A*
UNK_257.2473_11.687	7:26381618	C	T	0.82	-69.50	159.092	3.58E-08	*SNX10*
UNK_259.2431_11.486	11:61603510	C	A	0.57	-90.73	205.622	2.49E-10	*FADS2*
UNK_285.2071_10.349	12:21331549	C	T	0.86	-58.96	112.344	2.66E-08	*SLCO1B1*
UNK_297.2784_11.374	6:132711678	G	T	0.88	54.48	97.5177	3.45E-08	*MOXD1*
UNK_299.1393_7.568	2:234664586	A	ATC	0.70	-98.22	219.423	3.33E-11	*UGT1A*
UNK_299.1393_7.568	2:234665983	G	A	0.70	-98.22	219.423	3.33E-11	*UGT1A*
UNK_301.2174_11.263	11:61609750	C	T	0.58	-88.59	217.889	1.95E-09	*FADS2*
UNK_313.2732_11.408	6:116491302	A	G	0.74	-74.37	172.554	1.50E-08	*NT5DC1*
UNK_341.1108_6.44	2:234664586	A	ATC	0.70	-105.90	218.973	8.27E-13	*UGT1A*
UNK_341.1108_6.44	2:234665983	G	A	0.70	-105.90	218.973	8.27E-13	*UGT1A*
UNK_341.1109_6.621	2:234664586	A	ATC	0.7	-114.28	220.268	1.36E-14	*UGT1A*
UNK_341.1109_6.621	2:234665983	G	A	0.7	-114.28	220.268	1.36E-14	*UGT1A*
UNK_341.1109_6.735	2:234664586	A	ATC	0.70	-110.89	219.972	7.64E-14	*UGT1A*
UNK_341.1109_6.735	2:234665983	G	A	0.70	-110.89	219.972	7.64E-14	*UGT1A*
UNK_343.1946_11.516	11:61603510	C	A	0.57	-86.74	208.554	1.89E-09	*FADS2*
UNK_343.1946_11.516	11:44712698	C	G	0.75	74.91	179.21	2.20E-08	-
UNK_356.0376_1.226	4:10004805	C	T	0.74	86.76	243.848	2.76E-08	*SLC2A9*
UNK_357.0772_6.62	2:234664586	A	ATC	0.70	-106.33	217.123	5.35E-13	*UGT1A*
UNK_357.0772_6.62	2:234665983	G	A	0.70	-106.33	217.123	5.35E-13	*UGT1A*
UNK_359.1674_11.513	11:61603510	C	A	0.57	-95.70	207.84	3.18E-11	*FADS2*

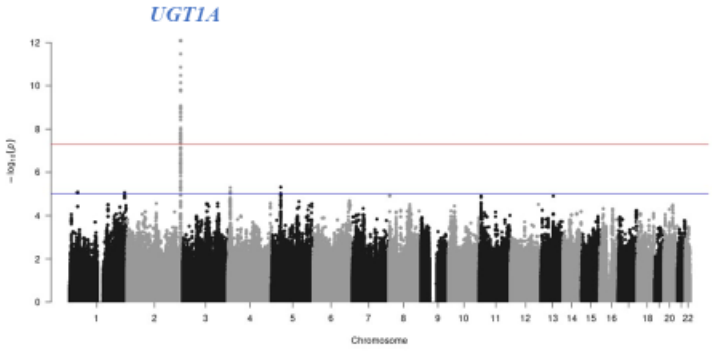
UNK_371.2206_11.488	11:61603510	C	A	0.57	-89.84	202.362	2.70E-10	*FADS2*
UNK_371.2206_11.488	11:44712698	C	G	0.75	72.34	175.805	4.87E-08	-
UNK_400.1706_11.512	11:61603510	C	A	0.57	-103.92	198.627	1.67E-13	*FADS2*
UNK_403.3205_9.707	12:21331549	C	T	0.86	-74.82	116.914	4.51E-12	*SLCO1B1*
UNK_403.3205_9.707	12:21399955	A	G	0.86	-67.33	116.91	4.75E-10	-
UNK_554.3474_1.927	3:189534621	T	C	0.72	90.59	256.163	1.51E-08	*TP63*
UNK_568.3814_1.928	12:4940628	A	G	0.88	63.44	130.829	2.91E-08	*GALNT8*
UNK_613.3594_9.663	12:21331549	C	T	0.86	-93.82	116.281	3.31E-18	*SLCO1B1*
UNK_613.3594_9.663	12:21399955	A	G	0.86	-88.88	116.61	1.86E-16	-
UNK_637.3558_9.707	12:21331549	C	T	0.86	-89.67	117.483	1.30E-16	*SLCO1B1*
UNK_651.371_9.996	12:21331549	C	T	0.86	-64.23	119.252	4.07E-09	*SLCO1B1*
UNK_681.3474_9.665	12:21331549	C	T	0.86	-87.83	117.545	5.44E-16	*SLCO1B1*
UNK_681.3474_9.665	12:21399955	A	G	0.86	-83.10	117.686	1.86E-14	-
UNK_762.5079_13.262	15:58726744	G	C	0.56	-98.07	241.358	2.74E-10	-
UNK_810.5651_13.586	11:61603510	C	A	0.57	-92.93	203.218	7.09E-11	*FADS2*
UNK_838.596_14.365	11:61609750	C	T	0.58	-87.54	199.652	5.80E-10	*FADS2*
UNK_850.5618_13.042	11:61593816	C	T	0.57	-112.25	209.746	9.15E-15	*FADS2*
UNK_850.5618_13.042	11:61640379	A	G	0.58	-96.12	242.243	6.58E-10	-
UNK_890.6257_14.785	14:67975822	C	A	0.54	168.49	244.628	4.65E-27	*TMEM229B*
UNK_890.6257_14.785	14:67987543	C	T	0.59	137.69	230.648	1.23E-19	-

Figure 6. Examples of QQ and Manhattan plots of genetically associated unidentified metabolites

(a) UNK_341.1108_6.44: associated with log-transformed insulin and log-transformed HOMA-IR

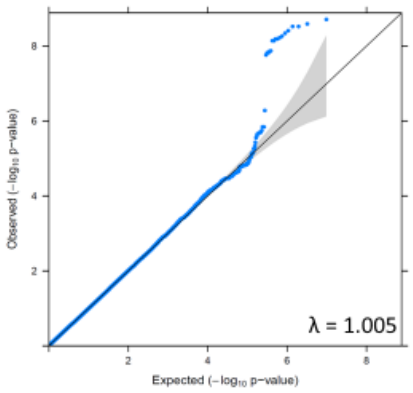


QQ plot of UNK_341.1108_6.44 (GMMAT)

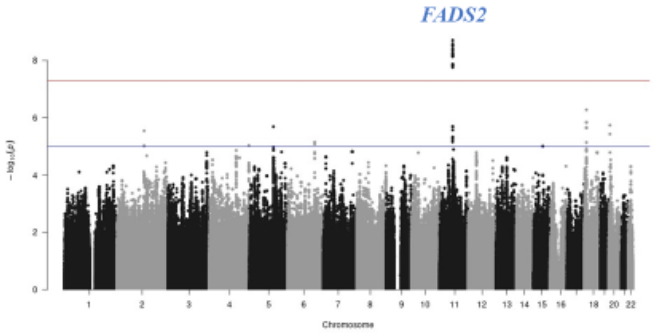


Manhattan plot on UNK_341.1108_6.44 (GMMAT)

(b) UNK_301.2174_11.263: associated with 2-hour post-load glucose

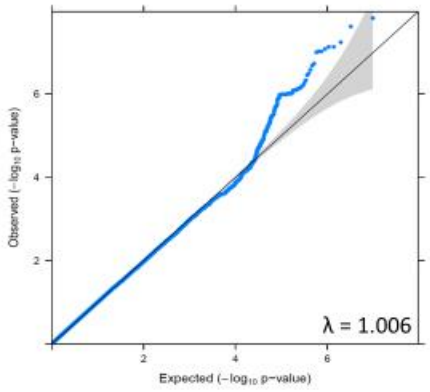


QQ plot of UNK_301.2174_11.263 (GMMAT)

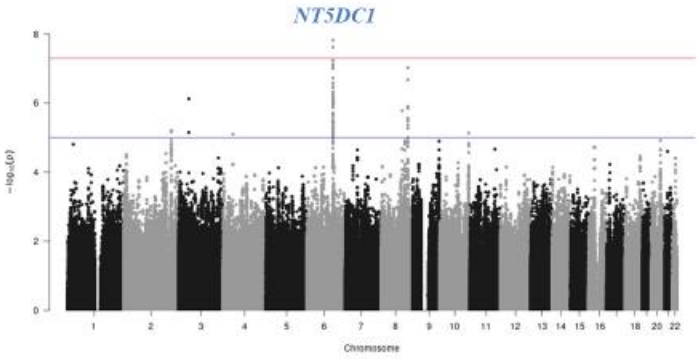


Manhattan plot on UNK_ 301.2174_11.263 (GMMAT)

(c) UNK_313.2732_11.408: associated with fasting plasma glucose, HbA1c, log-transformed insulin,log-transformed HOMA-IR, groups with different glycemic groups by the linear model, the group with non-diabetes Vs. the group with diabetes, and triglyceride

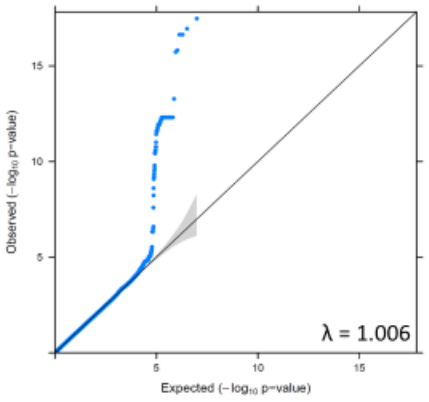


QQ plot of UNK_313.2732_11.408 (GMMAT)

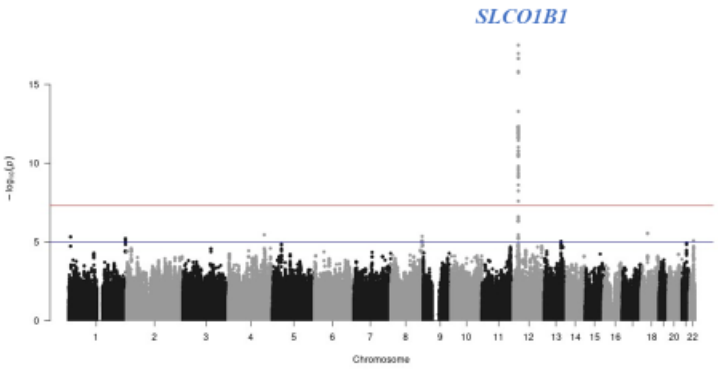


Manhattan plot on UNK_313.2732_11.408 (GMMAT)

(d) UNK_613.3594_9.663: associated with 2-hour post-load glucose, groups with different glycemic groups by the linear model, normal Vs. the group with diabetes, normal Vs. the group with prediabetes, normal Vs. the group with hyperglycemia, the group with non-diabetes Vs. the group with diabetes, and BMI



QQ plot of UNK_613.3594_9.663 (GMMAT)



Manhattan plot on UNK_613.3594_9.663 (GMMAT)

Aim 3. Interaction between genetic variation and nutrient intake on plasma metabolites

a. Identified metabolites

Carbohydrates, protein, total fat, saturated fat, monounsaturated fat, and polyunsaturated fat were tested separately to find the gene-nutrient density interaction on 145 identified metabolites by MAGEE-adjusted age, gender, and BMI. A total of 30 metabolites showed statistically significant interactions on each macronutrient nutrient density on 195 different SNPs. However, we also found severe inflations on the QQ plot of the interaction p-values on 23 metabolites. Thus, the metabolites showing the lambda value of QQ plot above 1.15 were dropped to decrease the false positive. 13 distinct SNPs- carbohydrate on Glutamic acid-Phenylalanine dipeptide, 3 distinct SNPs-protein on 3-methylbutyrylcarnitine, 3 different SNPs – protein on PC (38:6), 9 SNPs- saturated fat on docosatetraenoic acid, 3 SNPs-saturated fat on phenylacetic acid, 5:176735612 SNP- monounsaturated fat on eicosadienoic acid, and 3:55577710 SNP - polyunsaturated fat on 3-Methyl-2-oxovaleric acid were identified to show gene-nutrient interactions, and total fat nutrient density did not have any significant interaction on single variants. Especially, 3:6170316-protein on 3-methylbutyrylcarnitine, 2 different SNPs – protein on PC (38:6), and 9 SNPs- saturated fat on docosatetraenoic acid, 5:176735612 SNP- monounsaturated fat on eicosadienoic acid also showed statistical significance on gene-nutrient density interaction in the join test. The single variants that showed significant interactions with each nutrient intake did not show significant genetic associations by GMMAT in the aim2. Table 12 summarizes the identified metabolites with single variant-nutrient density interaction results by MAGEE. Figure 7

shows the QQ plots and the Manhattan plots of the interaction p-values and Manhattan plots of the same SNPs for comparison.

Table 12. Significant identified metabolites with single variant-nutrient density interaction

Metabolites	Nutrient	SNP	AF	β G-E	P-Value Interaction	P-Value Joint	GMMAT p-value	Gene
Glu-Phe	Carb.	19:20046102	0.03	0.122	1.36E-08	9.84E-08	0.82	*ZNF93*
Glu-Phe	Carb.	19:20043175	0.03	0.121	3.70E-08	2.57E-07	0.78	*ZNF93*
Glu-Phe	Carb.	19:20044886	0.03	0.121	3.70E-08	2.57E-07	0.78	*ZNF93*
Glu-Phe	Carb.	19:20019931	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20022056	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20023643	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20023925	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20033650	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20036176	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20039068	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20041972	0.02	0.127	4.35E-08	2.58E-07	0.50	*ZNF93*
Glu-Phe	Carb.	19:20054746	0.02	0.127	4.35E-08	2.58E-07	0.50	-
Glu-Phe	Carb.	19:20055370	0.02	0.127	4.35E-08	2.58E-07	0.50	-
3-Methylbutyrylcarnitine	Protein	3:6170316	0.37	-0.091	5.65E-09	4.22E-08	0.87	-
3-Methylbutyrylcarnitine	Protein	3:6170374	0.37	-0.092	6.96E-09	5.19E-08	0.90	-
3-Methylbutyrylcarnitine	Protein	3:6169451	0.40	-0.092	9.59E-09	7.05E-08	0.93	-
PC(38:6)	Protein	6:129907587	0.26	-0.103	5.24E-09	3.94E-08	0.92	*ARHGAP18*
PC(38:6)	Protein	6:129907597	0.26	-0.103	5.24E-09	3.94E-08	0.92	*ARHGAP18*
PC(38:6)	Protein	6:129902753	0.26	-0.102	7.90E-09	5.83E-08	0.99	*ARHGAP18*
Docosatetraenoic acid	Sat. Fat	10:1514015	0.64	-0.168	4.52E-09	1.99E-08	0.30	*ADARB2*
Docosatetraenoic acid	Sat. Fat	10:1514105	0.64	-0.168	4.52E-09	1.99E-08	0.30	*ADARB2*
Docosatetraenoic acid	Sat. Fat	10:1512458	0.63	-0.167	6.33E-09	2.61E-08	0.27	*ADARB2*
Docosatetraenoic acid	Sat. Fat	10:1511786	0.65	-0.167	6.38E-09	3.20E-08	0.37	*ADARB2*
Docosatetraenoic acid	Sat. Fat	10:1511150	0.65	-0.167	6.41E-09	2.95E-08	0.32	*ADARB2*
Docosatetraenoic acid	Sat. Fat	10:1513202	0.65	-0.167	6.41E-09	2.95E-08	0.32	*ADARB2*
Docosatetraenoic acid	Sat. Fat	10:1513616	0.65	-0.167	6.41E-09	2.95E-08	0.32	*ADARB2*
Docosatetraenoic acid	Sat. Fat	10:1514074	0.65	-0.167	6.41E-09	2.95E-08	0.32	*ADARB2*

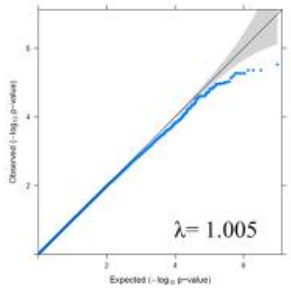
Docosatetraenoic acid	Sat. Fat	10:1510047	0.63	-0.165	9.76E-09	4.34E-08	0.31	*ADARB2*
Phenylacetic acid	Sat. Fat	3:53345623	0.24	-0.180	2.74E-08	1.89E-07	0.78	*DCP1A*
Phenylacetic acid	Sat. Fat	3:53332363	0.24	-0.180	3.11E-08	2.09E-07	0.73	*DCP1A*
Phenylacetic acid	Sat. Fat	3:53360412	0.24	-0.180	3.11E-08	2.09E-07	0.73	*DCP1A*
Eicosadienoic acid	Mon. Fat	5:176735612	0.33	0.131	9.88E-09	2.84E-08	0.16	*MXD3*
3-Methyl-2-oxovaleric acid	Poly. Fat	3:55577710	0.71	-0.198	3.03E-08	1.63E-07	0.44	*ERC2*

*Abbreviation: Carb: Carbohydrate; Sat.Fat: Saturated Fat; Mon.Fat: Monounsaturated Fat; Poly. Fat: Polyunsaturated Fat; PC: Phosphatidylcholine

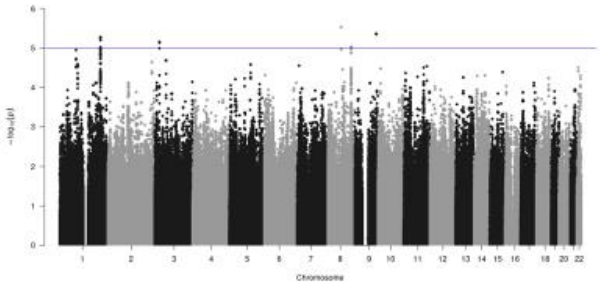
Figure 7. Example QQ plots and the Manhattan plots of the interaction p-values for the significant gene-nutrient interactions on identified metabolites by MAGEE compared to the results of GMMAT

(a) Protein

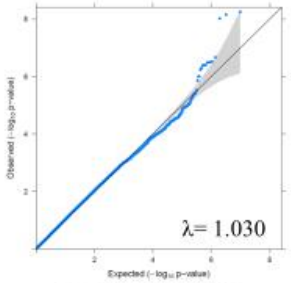
a.(1) Protein on 3-Methylbutyrylcarnitine



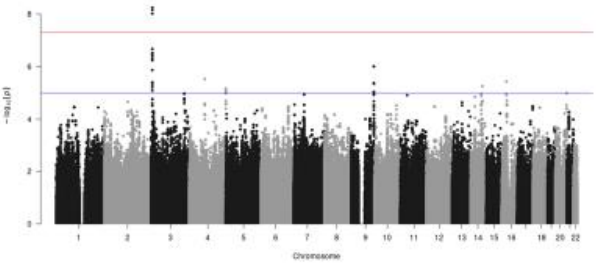
QQ plot of GMMAT



GMMAT : 3 - Methylbutyrylcarnitine

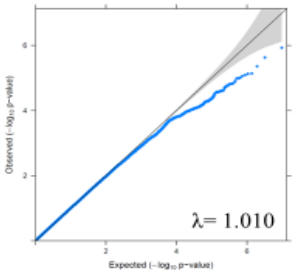


QQ plot of MAGEE

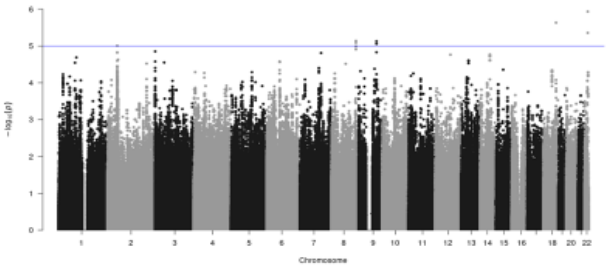


MAGEE : SNP – protein interaction on 3- Methylbutyrylcarnitine

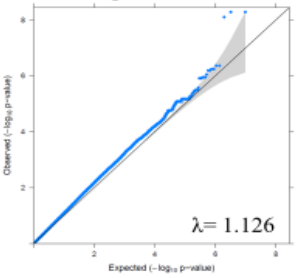
a.(2) Protein on PC (38:6)



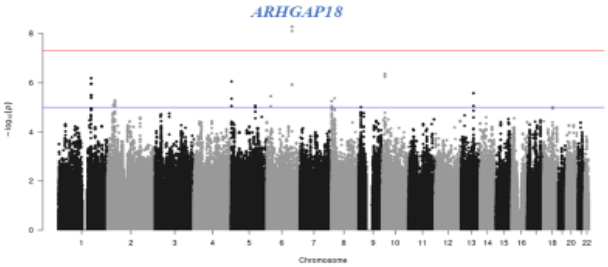
QQ plot of GMMAT



GMMAT on PC (38:6)

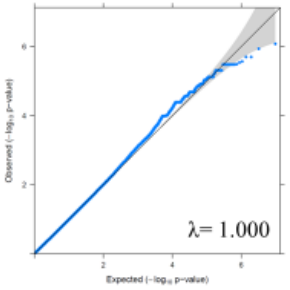


QQ plot of MAGEE

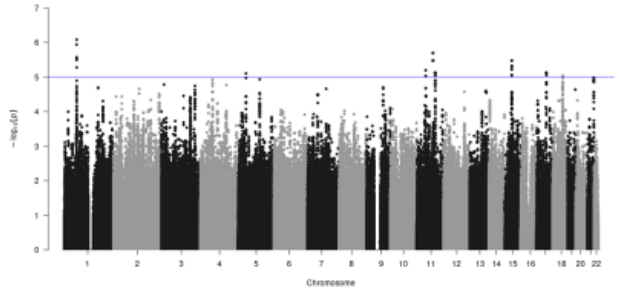


MAGEE : SNP – protein interaction on PC (38:6)

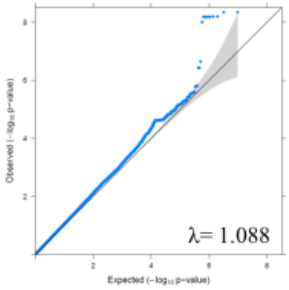
(b) Saturated fat on docosatetraenoic acid



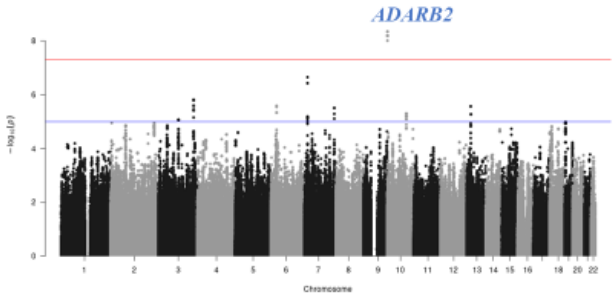
QQ plot of GMMAT



GMMAT on Docosatetraenoic acid

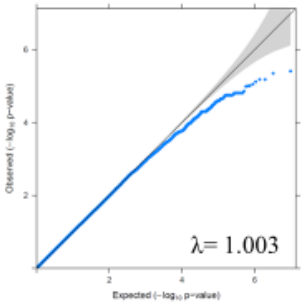


QQ plot of MAGEE

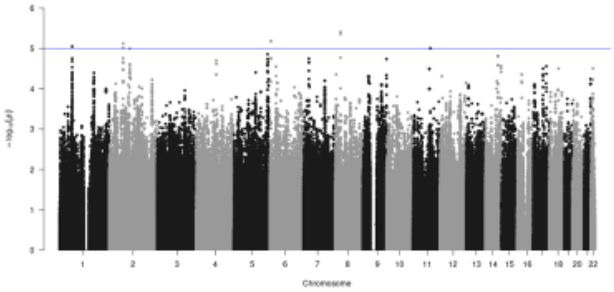


MAGEE: SNP-Sat. Fat interaction on Docosatetraenoic acid

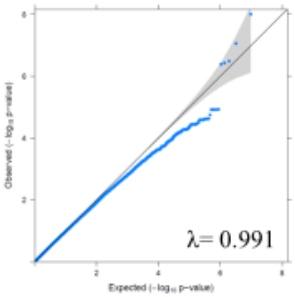
(c) Monounsaturated fat on eicosadienoic acid



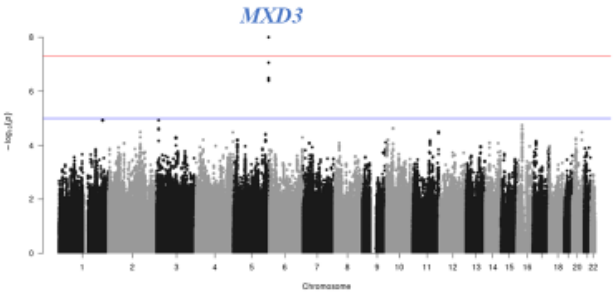
QQ plot of GMMAT



GMMAT on Eicosadienoic acid



QQ plot of MAGEE



MAGEE: SNP- Mono. Fat interaction on Eicosadienoic acid

b. Unidentified metabolites

687 unidentified metabolites that showed any association with glycemic traits, lipid traits, genetic association by GMMAT, and macronutrients were selected to find gene-nutrient interaction. 112 unidentified metabolites were related to one of 807 single variants-nutrient interaction pairs. After removing the severely inflated unidentified metabolites by lambda value of the QQ plots of p-value interactions, 32 unidentified metabolites were left. Thus, there were 39 single variants-carbohydrate interaction pairs associated with seven distinct unidentified metabolites, eleven single variants-protein interaction pairs on six metabolites, 58 SNPs - total fat interaction pairs on five metabolites, 17 SNPs – saturated fat interaction pairs on five metabolites, 38 single variants – monounsaturated fat interaction pairs on five metabolites, and 20 SNPs – polyunsaturated fat interaction pairs on four metabolites. 1:14463067 interacted with total fat and monounsaturated fat on UNK_422.2188_3.625. This unidentified metabolite was related to log-transformed insulin and log-transformed HOMA-IR. 14 same SNPs also showed significant interactions with total fat and monounsaturated fat on UNK_881.6741_15.005, and this unidentified metabolite was related to 2-hour post-load glucose, log-transformed insulin, log-transformed HOMA-IR, the comparison of glycemic status by the linear model, the comparison between normal Vs. diabetes, normal Vs. hyperglycemic group, HDL, and triglyceride. 14 metabolites having genetic associations were significant interactions with nutrient density by MAGEE, but no SNPs overlapped across two different methods. 25 SNPs – carbohydrate interaction pairs on UNK_319.129_6.441, 3:6249007 SNP – protein interaction fairs on UNK_625.204_12.115, 12 distinct SNPs on chromosome 11– saturated fat interaction pairs

on UNK_1008.7684_0.55, 1:187747584 SNP – total fat interaction pair, and 20:40382240 SNP – saturated fat interaction pairs on UNK_1195.8522_11.509 also showed significant interaction results with joint tests, none of them were located on the coding region. Table 13 summarizes important top signals on one macronutrient by MAGEE.

13 SNPs showed significant interactions with total fat and monounsaturated fat intake on UNK_881.6741_15.005, and 1:14463067 located on *the KAZN* gene also showed significant interactions with total fat and monounsaturated fat intake on UNK_422.2188_3.625. Table 14 summarizes these significant same SNP-nutrient interactions on total fat and monounsaturated fat by MAGEE. Figure 8 shows the example QQ and Manhattan plots of the gene-nutrient interaction on unidentified metabolites.

Table 13. Significant unidentified metabolites with single variants- one nutrient interaction

Metabolites	Nutrient	SNP	Effect Allele	AF	β G-nut.	P-Value Interaction	P-Value Joint	GMMAT p-value
UNK_319.129_6.441	Carb.	5:30599586	G	0.21	0.048	1.84E-08	2.08E-08	0.055
UNK_319.129_6.441	Carb.	5:30600589	A	0.22	0.048	1.65E-08	1.79E-08	0.052
UNK_319.129_6.441	Carb.	5:30602495	C	0.22	0.048	1.65E-08	1.79E-08	0.052
UNK_319.129_6.441	Carb.	5:30602910	T	0.22	0.048	1.65E-08	1.79E-08	0.052
UNK_319.129_6.441	Carb.	5:30604771	C	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30607792	C	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30608507	T	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30611219	TA	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30611639	A	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30612899	T	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30613836	A	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30615569	T	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30617557	C	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30617938	T	0.22	0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30623762	C	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30628417	T	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30629256	T	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30629646	A	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30630563	T	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30634884	G	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30635164	G	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30641038	T	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30641512	C	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_319.129_6.441	Carb.	5:30645124	T	0.78	-0.048	2.03E-08	1.98E-08	0.046
UNK_319.129_6.441	Carb.	5:30649097	A	0.78	-0.048	1.90E-08	1.97E-08	0.050
UNK_625.204_12.115	Protein	3:6249007	T	0.10	0.158	2.17E-08	1.38E-08	0.030
UNK_1008.7684_0.55	Sat. Fat	11:103681759	T	0.04	0.370	3.32E-08	3.09E-08	0.048
UNK_1008.7684_0.55	Sat. Fat	11:103683709	G	0.04	0.370	3.32E-08	3.09E-08	0.048
UNK_1008.7684_0.55	Sat. Fat	11:103689442	A	0.04	0.364	4.53E-08	4.11E-08	0.046
UNK_1008.7684_0.55	Sat. Fat	11:103693019	G	0.04	0.364	4.53E-08	4.11E-08	0.046
UNK_1008.7684_0.55	Sat. Fat	11:103694149	A	0.04	0.370	3.32E-08	3.09E-08	0.048
UNK_1008.7684_0.55	Sat. Fat	11:103700803	A	0.04	0.370	3.32E-08	3.09E-08	0.048
UNK_1008.7684_0.55	Sat. Fat	11:103702022	C	0.04	0.364	4.53E-08	4.11E-08	0.046
UNK_1008.7684_0.55	Sat. Fat	11:103702225	C	0.04	0.370	3.32E-08	3.09E-08	0.048
UNK_1008.7684_0.55	Sat. Fat	11:103710896	T	0.04	0.364	4.53E-08	4.11E-08	0.046
UNK_1008.7684_0.55	Sat. Fat	11:103714138	C	0.04	0.370	3.32E-08	3.09E-08	0.048
UNK_1008.7684_0.55	Sat. Fat	11:103715726	C	0.04	0.364	4.53E-08	4.11E-08	0.046
UNK_1008.7684_0.55	Sat. Fat	11:103717473	T	0.04	0.370	3.32E-08	3.09E-08	0.048

UNK_1195.8522_11.509	Sat. Fat	20:40382240	C	0.51	0.147	3.84E-08	4.71E-08	0.063

*Abbreviation: Carb: Carbohydrate; Sat.Fat: Saturated Fat

Table 14. Significant unidentified metabolites with single variants- two nutrient interactions

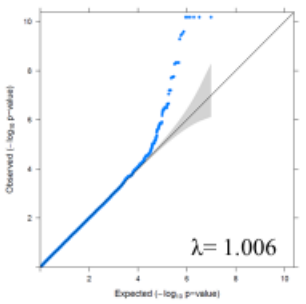
Metabolites	SNP	Effect Allele	AF	Nutrient	β G-nut.	P-Value Interaction	P-Value Joint	GMMAT p-value
UNK_881.6741_15.005	1:187747584	T	0.27	Total Fat	-0.063	3.69E-09	2.63E-08	0.638
				Mon. Fat	-0.144	3.05E-09	2.11E-08	
UNK_881.6741_15.005	1:187737758	T	0.28	Total Fat	-0.062	9.48E-09	6.93E-08	0.806
				Mon. Fat	-0.133	3.83E-08	2.66E-07	
UNK_881.6741_15.005	1:187737754	G	0.28	Total Fat	-0.060	1.66E-08	1.19E-07	0.782
				Mon. Fat	-0.130	3.91E-08	2.69E-07	
UNK_881.6741_15.005	1:187689364	C	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187689404	A	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187692534	A	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187693050	A	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187695514	G	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187698135	T	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187699089	A	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187699094	A	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_881.6741_15.005	1:187699305	C	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	

UNK_881.6741_15.005	1:187701595	C	0.24	Total Fat	-0.060	3.51E-08	2.19E-07	0.549
				Mon. Fat	-0.132	4.62E-08	2.76E-07	
UNK_422.2188_3.625	1:14463067	A	0.77	Total Fat	0.060	4.21E-08	2.85E-07	0.808
				Mon. Fat	0.136	1.15E-08	8.20E-08	

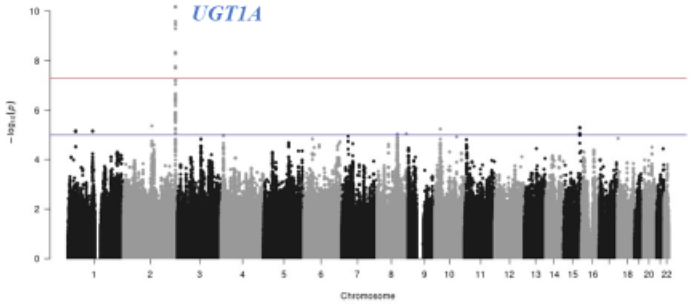
*Abbreviation: Mon.Fat: Monounsaturated Fat

Figure 8. Example QQ plots and the Manhattan plots of the interaction p-values for the significant gene-nutrient interactions on unidentified metabolites by MAGEE compared to the results of GMMAT

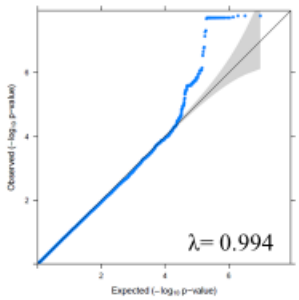
(a) Carbohydrate on UNK_319.129_6.441



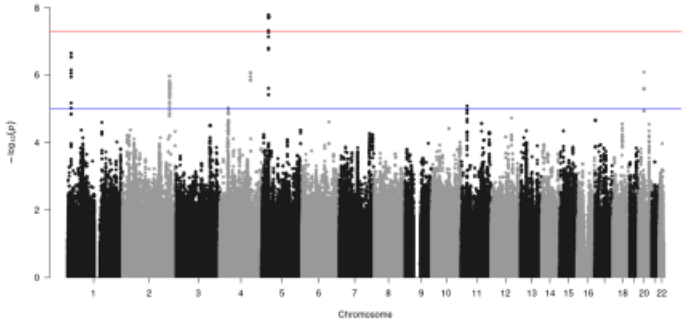
QQ plot of GMMAT



GMMAT on UNK_319.129_6.441

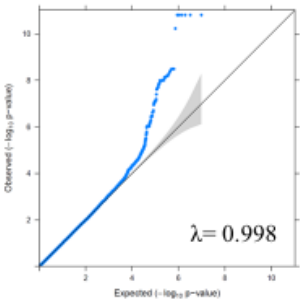


QQ plot of MAGEE

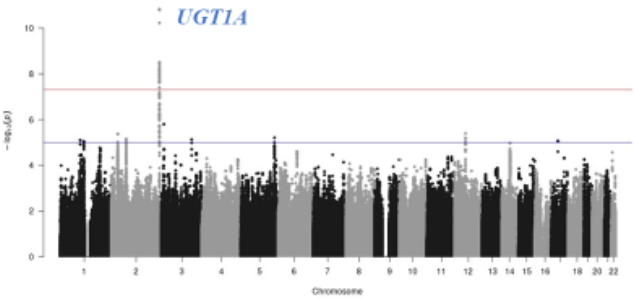


MAGEE: SNP- Carbohydrate interaction on UNK_319.129_6.441

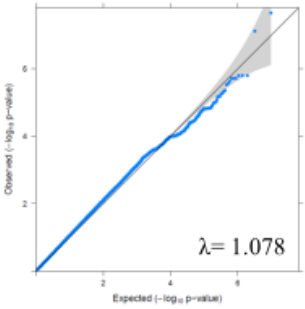
(b) Protein on UNK_625.204_12.115



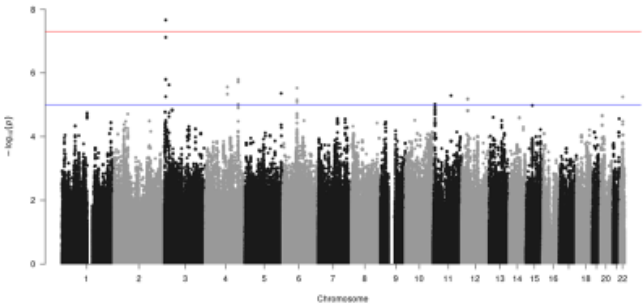
QQ plot of GMMAT



GMMAT on UNK_625.204_12.115

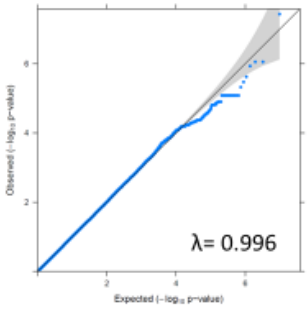


QQ plot of MAGEE

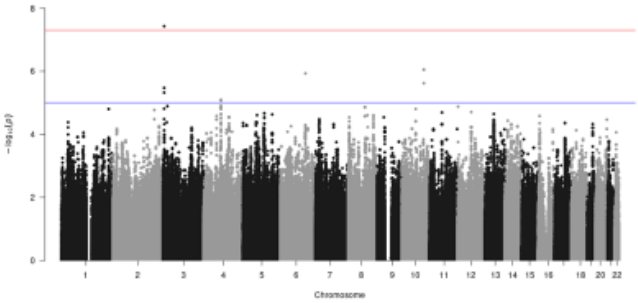


MAGEE: SNP- Protein interaction on UNK_625.204_12.115

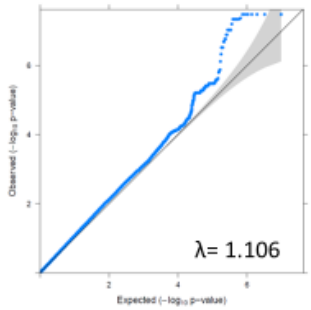
(c) Saturated fat on UNK_1008.7684_0.55



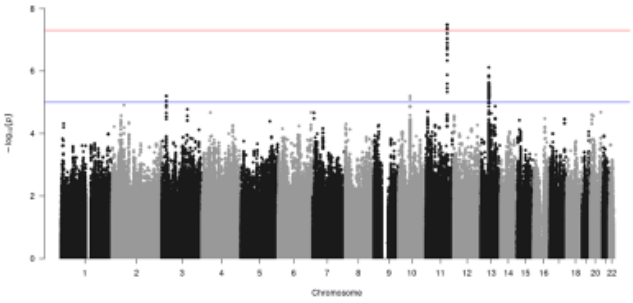
QQ plot of GMMAT



GMMAT on UNK_1008.7684_0.55

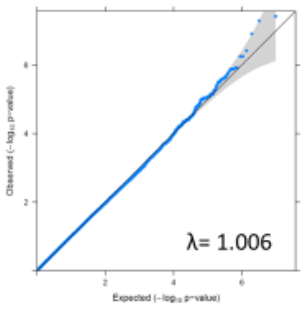


QQ plot of MAGEE

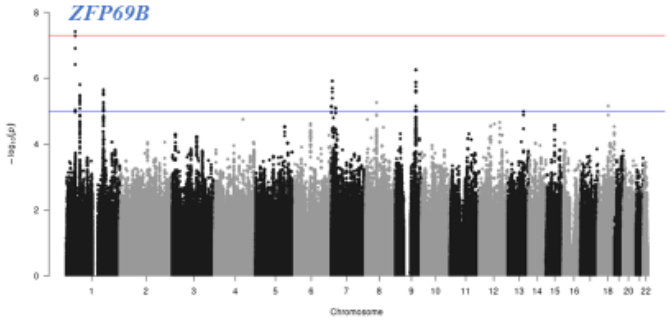


MAGEE: SNP- Sat. Fat interaction on UNK_1008.7684_0.55

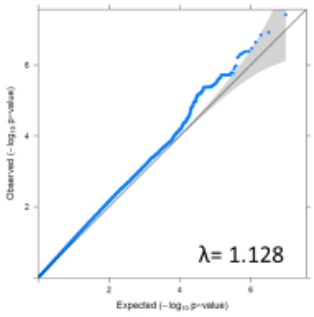
(d) Saturated fat on UNK_1195.8522_11.509



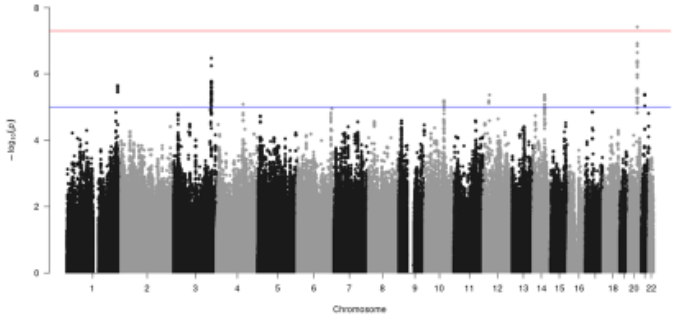
QQ plot of GMMAT



GMMAT on UNK_1195.8522_11.509

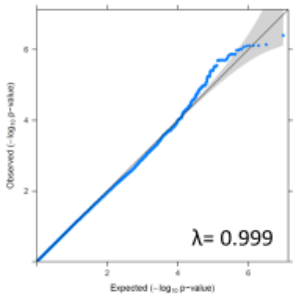


QQ plot of MAGEE

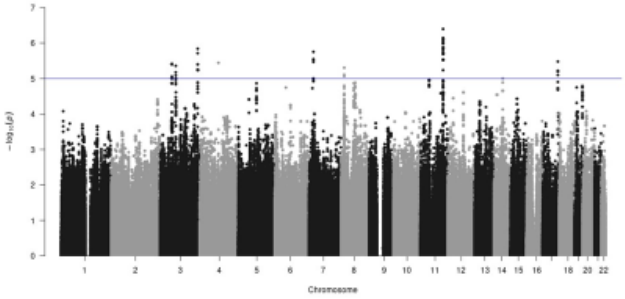


MAGEE: SNP- Sat. Fat interaction on UNK_1195.8522_11.509

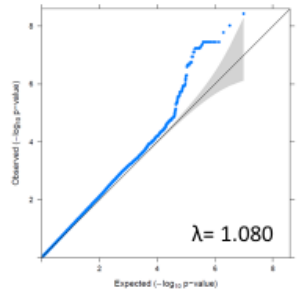
(e) Total fat and monounsaturated fat on UNK_881.6741_15.005



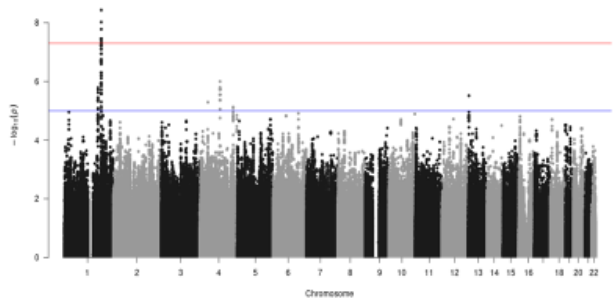
QQ plot of GMMAT



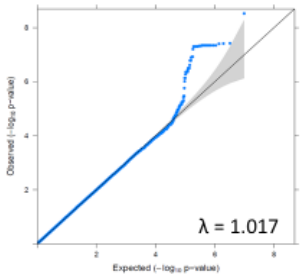
GMMAT on UNK_881.6741_15.005



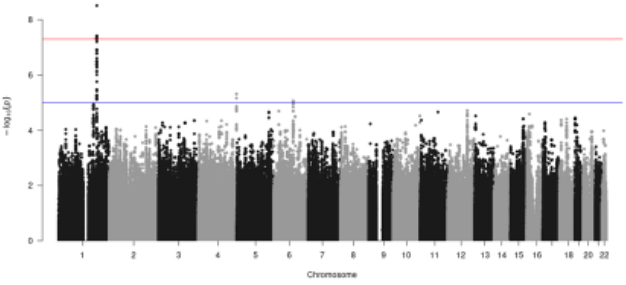
QQ plot of MAGEE (Total fat)



MAGEE: SNP- Tot. Fat interaction on UNK_881.6741_15.005

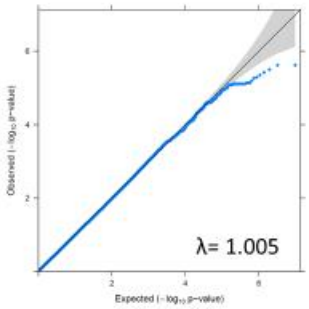


QQ plot of MAGEE (Mono. Fat)

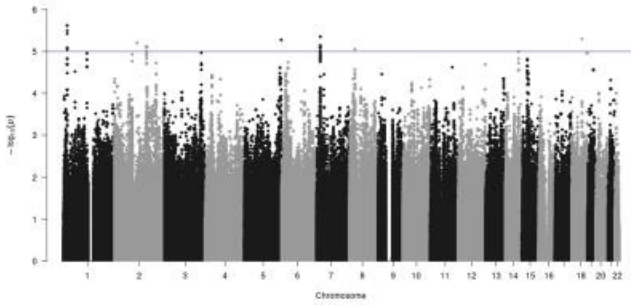


MAGEE: SNP- Mono. Fat interaction on UNK_881.6741_15.005

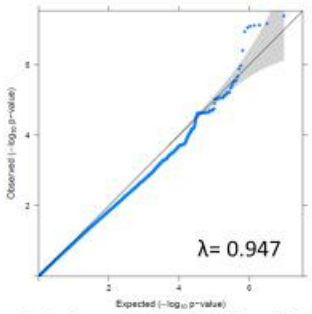
(f) Total fat on monounsaturated fat on UNK_422.2188_3.625



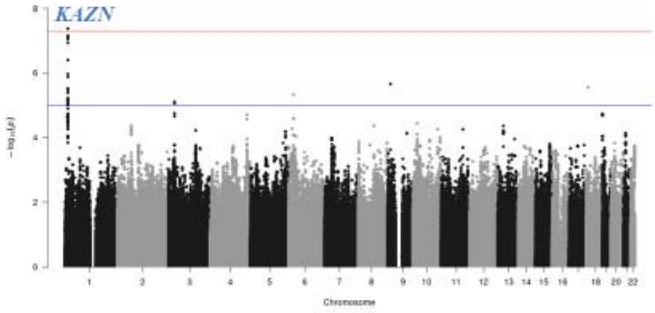
QQ plot of GMMAT



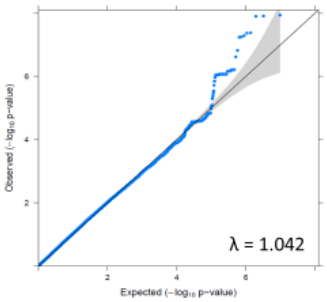
GMMAT on UNK_422.2188_3.625



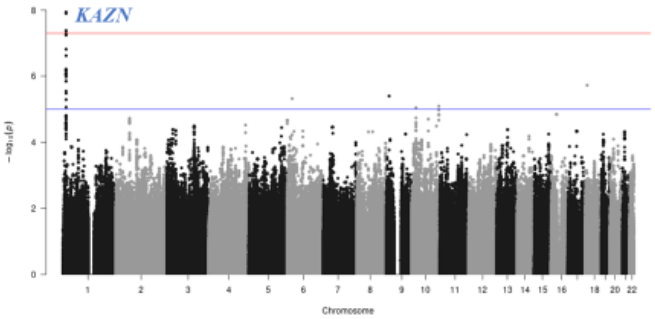
QQ plot of MAGEE (Total fat)



MAGEE: SNP- Tot. Fat interaction on UNK_422.2188_3.625



QQ plot of MAGEE (Mono. Fat)

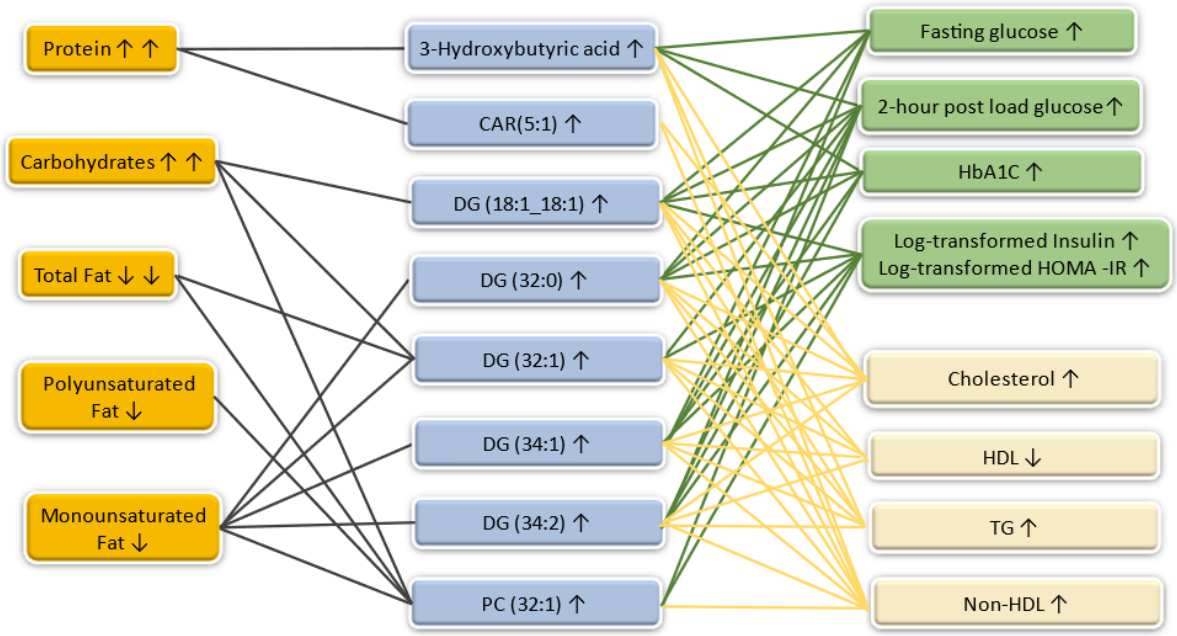


MAGEE: SNP- Mono. Fat interaction on UNK_422.2188_3.625

DISCUSSION

We found 23 identified and 47 unidentified metabolites associated with fasting glucose, 49 identified and 208 unidentified metabolites associated with 2-hour post-load glucose, 14 identified and 29 unidentified metabolites related to HbA1c, 49 identified and 253 unidentified metabolites related to log-transformed insulin, and 51 identified and 241 unidentified metabolites associated with log-transformed HOMA-IR. In the comparison related to five different lipid traits, 37 identified and 148 unidentified metabolites were related to cholesterol, 14 identified and 45 unidentified metabolites were associated with HDL, 39 identified and 231 unidentified metabolites were associated with triglyceride, 14 identified and 70 unidentified metabolites were related to calculated LDL, and 42 identified and 149 unidentified metabolites were associated with non-HDL. When we compared the participants divided into groups based on their glycemic statuses with linear and ANOVA models, 19 identified and 87 unidentified metabolites had significant associations in comparing the group with normal glycemia and prediabetes. 9 identified and 2 unidentified metabolites were associated with macronutrient nutrient density.

Figure 9. Summary of association across macronutrient intake, identified metabolites, plasma glycemic traits, and plasma lipid traits



a Abbreviation: CAR Carnitine; DG Diacylglycerol; PC Phosphatidylcholine; TG Triglyceride

Figure 9 summarizes these associations among the identified metabolites across nutrient intake, metabolites, glycemic, and lipid traits. Orange cells mean macronutrient intake, blue cells are the identified metabolites, green cells point to five glycemic traits, and yellow cells represent plasma lipid traits in Figure 9. Increased protein intake showed positive associations with 3-hydroxybutyric acid and CAR (5:1), and an increase of 3-hydroxybutyric acid was also positively related to the rise in fasting glucose, 2-hour post-load glucose, HbA1c, cholesterol, triglyceride, and non-HDL. CAR (5:1) was also positively associated with cholesterol and non-HDL but had no significant association with glycemic traits. We did not find any metabolites related to saturated fat, and the total fat was negatively

associated with DG (32:1) and PC (32:1) in the same direction as the association with monounsaturated fat. Although it may appear to contradict previous research results, the higher average intake of monounsaturated fat compared to Hispanics in the United States could be the reason for this result. According to the recent NHANES statistics from 2017 to 2020, the average monounsaturated fat intake per 1,000 kcal in Hispanic adults aged 20 and above in their study is 11.7 % (58). Still, the participants in our study reported 15.62 % of the energy coming from monounsaturated fat, so the higher portion of monounsaturated fat could impact the direction of association between total fat and identified metabolites since total fat included all the kinds of fat, including monounsaturated fat. Moreover, monounsaturated fat showed significant associations with five different metabolites, DG (32:0), DG (32:1), DG (34:1), DG (34:2), and PC (32:1), and the absolute value of each beta on the monounsaturated fat is greater than that of carbohydrates and total fat. Thus, decreasing monounsaturated fat intake could impact the glycemic and lipid traits more extensively than changing other macronutrient intakes.

We found five different diacylglycerols associated with nutrient intake and glycemic and lipid traits; DG (18:1_18:1), DG (32:0), DG (32:1), DG (34:1), and DG (34:2). Diacylglycerol is marked as DG or DAG in various articles and a neutral lipid formed by glycerol connected to two fatty acids chains by ester bonds (59). It involves multiple metabolic pathways, such as lipogenesis in the endoplasmic reticulum and activation of protein kinase C in the plasma membrane in the cell (60). Increased intracellular diacylglycerol impacts protein kinase C(PKC) and D(PKD) and generates different responses depending on the types of tissues. For example, PKCε inhibits ketogenesis and increases

gluconeogenesis in the liver, and it also inhibits insulin sensitivity in the skeletal muscle (59). In previous experimental studies, diacylglycerol provoked by a high-fat diet can affect insulin resistance (61, 62). Most of the studies have studied intracellular diacylglycerols since extracellular diacylglycerols can be digested into monoacylglycerol, but our study showed carbohydrate and fat intake were associated with these extracellular diacylglycerols.

3-hydroxybutyric acid, known as β – hydroxybutyrate, is a metabolite produced during ketone body metabolism (63). This compound increases in concentration during prolonged periods of fasting and serves as a crucial energy source for the human body, particularly the brain. According to Møller, during starvation, 3-hydroxybutyric acid supplies approximately 60 percent of the brain's energy compared to normal feeding conditions (64). Furthermore, this energy resource plays a significant role in various cellular processes, including cell signaling and metabolic pathways associated with inflammation, cancer cell cycle, oxidative stress, and angiogenesis (63-65). The activation of hydroxy-carboxylic acid receptors (HCAR) and inhibition of free fatty acid receptors (FFAR) have been linked to increased inflammation and the development of type 2 diabetes (64). Thus, these cellular pathways might explain the positive correlation between 3-hydroxybutyric acid and three diagnostic markers for prediabetes and diabetes. The result of this study suggests that a high protein intake could potentially stimulate these signaling pathways. Additionally, branched amino acids, such as leucine, exhibit a statistically significant association with each glycemic trait, although they are not directly linked to protein intake. This finding implies that the plasma concentration of branched amino acids could be influenced by factors other than food intake, such as the gut microbiome.

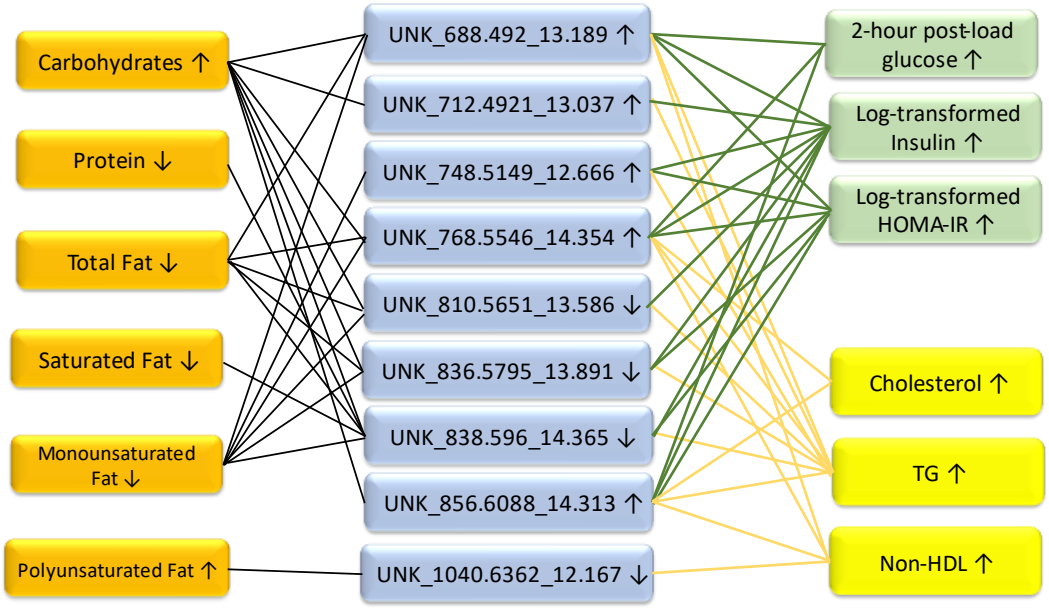
In our study, CAR (5:1), known as tiglyl carnitine related to leucine, isoleucine, and valine metabolism, was only associated with cholesterol and non-HDL. CAR (5:1) was also found as a biomarker of protein intake and low-fat milk intake, so this replication could validate the study result (66, 67). This metabolite also showed associations with short-term diet resistance and metabolic change after a short-term high-fat diet (68, 69).

PC (32:1) was associated with 2-hour post-load glucose levels and the logarithmically transformed insulin levels. Furthermore, this metabolite serves as a predictive lipid biomarker for both type 2 diabetes and gestational diabetes (70, 71). It is also among the metabolites that significantly change in response to prolonged sitting in individuals with type 2 diabetes, suggesting a potential influence on skeletal muscle activity (72). Kumar et al. demonstrated that phosphatidylcholine derived from phosphatidylethanolamine in intestinal epithelial exosomes could impact insulin response, triggering these pathways by a high-fat diet (73). Gao et al. studied PC derived from various food sources, such as squid, soy, and eggs, revealing differing effects on inducing insulin resistance depending on the specific food source (74). We observed a negative association between monounsaturated fat intake and PC 32:1, indicating that further experimental investigations focusing on specific dietary fat compositions could unveil the associated biological pathways.

Among 30 unidentified metabolites associated with macronutrient intake, 9 were related to 2-hour post-load glucose, log-transformed insulin, log-transformed HOMA-IR, cholesterol, triglyceride, and non-HDL. Figure 10 summarizes the associations across nutrient intake, unidentified metabolites, and glycemic and lipid traits. We found that UNK_838.596_14.365 and UNK_810.5651_13.586 were negatively associated with

carbohydrates but positively associated with total and monounsaturated fat. These unidentified metabolites were also significantly associated with SNP 11:61609750 and 11:61603510, located on the *FADS2* gene. However, UNK_838.596_14.365 were also positively associated with protein and saturated fat intake, and these results need additional studies since low saturated fat intake is usually considered beneficial for glucose intolerance. We also found that increased polyunsaturated fat intake could induce the decrease of the UNK_1040.6362_12.167 level, but this was associated with increased non-HDL levels. Since polyunsaturated fat is also considered healthy, this result should be investigated with additional studies.

Figure 10. Summary of association across macronutrient intake, unidentified metabolites, plasma glycemic traits, and plasma lipid traits



We found 28 genetically associated identified metabolites and 232 genetically associated unidentified metabolites. Considering the previous results in Aim 1, biliverdin, CAR(4:0), CAR (6:0), DG (32:1), eicosatetraenoic acid, EPA, LPC (20:0), LPC (20:4), LPE (18:0), N-acetylleucine, PC (34:2), pipecolic acid, 2-aminooctanoic acid, 2-hydroxybutyric acid, and 32 unidentified metabolites were overlapped metabolites associated with any glycemic or lipid traits. Especially the SNP located on 6:40532893 associated with DG (32:1) is on the *LRFN2* gene and has not been reported yet. The *LRFN2* gene is known to encode synaptic cell adhesion protein in the brain and be related to memory deficit and Alzheimer's disease (75). Still, multiple SNPs on this gene have been repetitively reported related to type 2 diabetes, BMI, waist-hip ratio, visceral adipose tissue adjustment, triglyceride, HDL, and non-HDL (76-88). Thus, the novel finding about the genetic association of the SNP on the *LRFN2* gene with DG (32:1) could be the clue why the multiple SNPs on this gene have been found the significant associations with the traits related to metabolic disease.

33 SNP-nutrient significant interactions on identified metabolites and 183 SNP-nutrient interactions on unidentified metabolites were found in this study. Among these SNP-nutrition interaction pairs, 13 SNP-nutrient pairs on identified metabolites and 40 SNP – nutrient pairs on unidentified metabolites also showed significantly low p-value on the joint tests. However, none of the same SNPs exhibited significant results in both GMMAT and MAGEE. Thus, we can conclude that the SNPs genetically associated with specific metabolites by GMMAT in this study had no gene-nutrient interactions, and the SNPs interacted with nutrition by MAGEE could suggest having SNP-nutrient interactions on the particular metabolites.

Strengths and limitations

One of the strengths of this study is the participants consisted of the Mexican American population in the U.S. borderline area. The Mexican American population is relatively understudied compared to the European in genetic studies, but this population has increased in the U.S. (15, 89). Specifically, 45 percent of the participants in Starr County reported uninsured health insurance, and more than half of the participants answered their household incomes were below $30,000 per year (90). Thus, the economic burden related to health on this population could be greater than others in the U.S. These combined analyses of nutrient intake, metabolomics, and genetics in this study could provide the essential evidence to develop additional preventive interventions on worsening glycemia targeting this population. Since we focused on macronutrients rather than specific diet patterns, the change in the nutrition component based on the Mexican diet could be more applicable in developing preventive interventions for worsening glycemia (91). Another strength of this study is that we could find novel associations between metabolite and distinct SNPs. Among the results of identified metabolites in Aim 2, CAR (9:0), DG (32:1), dodecadienoic acid, LPC (20:0), LPC (20:4), LPE (18:2), sphingosine, and 2-hydroxybutyric acid have not been reported any genetic associations before to my knowledge. These novel associations could be reliable since we also found other repetitive genetic associations on bilirubin, biliverdin, CAR (10:1), CAR (4:0), CAR (6:0), CAR (8:0), Eicosatetraenoic acid, EPA, N-Acetylleucine, PC (34:2), PC(40:8), 2-aminooctanoic acid. To my knowledge, this is the first study to test single variant-nutrient interaction on the metabolites of Mexican Americans. Previous gene-environmental interaction studies usually tested specific target genetic regions, but this study

tested single variant–nutrient interactions across whole imputed GWAS data. Thus, the results from MAGEE in Aim 3 could be found on the gene-nutrient interactions in different SNPs.

Still, this study has some limitations. First, recall bias can still disrupt the macronutrient data from the food frequency questionnaire even though it was previously validated in this population (92). Even though we educated the participants about checking their food frequency questionnaire to decrease the impact of recall bias, it was inevitable in the nutrient study. Second, the cross-sectional design cannot determine the causality due to the temporal ambiguity, so additional studies need to reveal the causal associations. Third, the statistical power could be weakened since the sample size of the study decreased from the original proposal due to unpredictable medicine use and sample loss. During the follow-up period on the participants, we found that 13 people used glucose-lowering drugs such as metformin, and 88 people used cholesterol-lowering medications. We did not find the genetic data from 131 participants in this study, so the decreased statistical power from the sample size might be one of the reasons for exhibiting unstable MAGEE results. MAGEE was originally developed for the large sample-size investigation of more than 2000, so we need additional studies with increasing sample sizes in the Mexican-American population to validate our study results(93). The decreased sample size was another barrier to finding rare variants, so we limited the allele frequency from 0.01 to 0.5 and did not proceed with the investigation with exome sequencing data. Last, we have not determined the name of unidentified metabolites yet, so identifying unknown metabolites should be added to this study.

CONCLUSION

Diabetes mellitus is one of the main issues in public health because of the perspective of high prevalence, complications in multiple organs, economic burden, and presence of reversible intermediate stage by lifestyle modification. Since diet and genes can cause worsening glycemia, this study investigated how these factors affect metabolome data and are associated. 308 identified metabolites and 2,471 unidentified metabolites were used for the analyses, and 3-hydroxybutyric acid, CAR (5:1), DG (18:1_18:1), DG (32:0), DG(32:1), DG(34:1), DG(34:2), PC (32:1), and 9 unidentified metabolites had significant associations between macronutrient intake and glycemic and lipid traits. 19 identified and 87 unidentified metabolites were significant mean differences in the comparison between the group with normoglycemia and the group with prediabetes. Among 28 identified and 232 unidentified metabolites related to SNPs, the SNPs on the *LRFN2* gene were associated with DG (32:1), which had associations with carbohydrate, total fat, and monounsaturated fat intake. Although we can test SNP-macronutrient interactions by MAGEE, all these significant SNPs were located in the noncoding DNA regions. Thus, additional studies are needed to reveal the exact information on these loci.

We have not determined the unidentified metabolites yet, so identifying unknown metabolites should be the next step for this study. Moreover, this study was obtained from the baseline data of a 3-year longitudinal study, so we could validate the novel findings with additional longitudinal data. The food intake could be changed over time, so additional food frequency data could improve the reliability and validity of the nutrient data. If we could collaborate with other Mexican American study groups to reach a sample size enough to

confirm the genetic association, we can check whether the novel results from this study can be validated and test it for the rare variants showing allele frequencies of less than one percent. If significant associations between macronutrients and metabolites can be confirmed with these further studies, we could investigate the specific metabolites, such as DG (32:1), and glycemic traits with experimental studies with animals to establish causality.

APPENDICES

Appendix A: Supplemental Tables

Appendix A. Supplemental tables

Table S1. The demographic characteristics of the participants

Characteristics	Normal N(%)	Prediabetes N(%)	Diabetes N(%)	Total N(%)
Age(years)				
< 40	28 (18.42)	40 (10.87)	4 (4.82)	72 (11.94)
40-49	59 (38.82)	132 (35.87)	31 (37.35)	222 (36.82)
50-59	48 (31.58)	142 (38.59)	32 (38.55)	222 (36.82)
≥ 60	17 (11.18)	54 (14.67)	16 (19.28)	87 (14.43)
Gender				
Male	28 (18.42)	121 (32.88)	26 (31.3)	175 (29.02)
Female	124 (81.58)	247 (67.12)	57 (68.7)	428 (70.98)
BMI(kg/m^2)				
≤ 24.9	34 (22.37)	26 (7.07)	3 (3.61)	63 (10.45)
25.0-29.9	59 (38.82)	109 (29.62)	13 (15.66)	181 (30.02)
≥ 30.0	59 (38.82)	233 (63.32)	67 (80.72)	359 (59.54)
Total	152 (25.21)	368 (61.03)	83 (13.76)	603 (100.00)

Table S2. Association between fasting plasma glucose levels and metabolites

Metabolites	β	p-value
3-Hydroxybutyric acid	0.0122	5.54E-11
Leucine	0.0093	1.40E-09
3-Methyl-2-oxovaleric acid	0.0091	3.71E-09
Leucine Isoleucine	0.0090	4.08E-09
Ketoleucine	0.0090	5.61E-09
DG(34:1)	0.0097	7.40E-09
DG(34:2)	0.0091	7.17E-08
Isoleucine	0.0081	1.66E-07
DG(18:1_18:1)	0.0088	2.50E-07
2-Deoxyglucose	-0.0089	6.57E-07
DG(32:0)	0.0079	2.66E-06
MG(16:0)_rp_a	0.0084	3.03E-06
DG(36:3)	0.0079	4.31E-06
Heptadecanedioic acid	0.0075	1.26E-05
Choline	-0.0070	2.28E-05
Docosatetraenoic acid	0.0070	4.02E-05
Glutamine	-0.0069	5.53E-05
Behenic acid	0.0068	7.67E-05
Nonadecenoic acid	0.0065	8.63E-05
DHA	0.0068	8.73E-05
Ɣ-Glutamylleucine	0.0059	9.45E-05
Hydroxytetradecanoic acid	0.0067	0.000119
2-Hydroxybutyric acid	0.0066	0.000162

Table S3. Association between 2-hour post-load glucose levels and metabolites

Metabolites	β	p-value
Docosatetraenoic acid	0.0071	2.80E-15
Docosatrienoic acid	0.0070	6.21E-15
3-Hydroxybutyric acid	0.0072	3.57E-14
DG(34:1)	0.0066	7.14E-14
DG(34:2)	0.0065	2.45E-13
Eicosatrienoic acid	0.0066	3.03E-13
DHA	0.0066	1.15E-12
Eicosadienoic acid	0.0064	1.19E-12
Nonadecenoic acid	0.0062	2.16E-12
Docosapentaenoic acid	0.0063	4.71E-12
DG(32:0)	0.0060	1.34E-11
DG(32:1)	0.0059	2.72E-11
3-Methyl-2-oxovaleric acid	0.0054	3.61E-11
Ketoleucine	0.0053	8.63E-11
Palmitic acid	0.0057	3.84E-10
Hydroxytetradecanoic acid	0.0056	1.04E-09
Stearic acid	0.0056	1.60E-09
Leucine	0.0048	4.00E-09
Margaric acid	0.0053	1.55E-08
Leucine Isoleucine	0.0046	1.90E-08
Isoleucine	0.0046	4.46E-08
DG(18:1_18:1)	0.0048	1.66E-07
MG(16:0)_rp_a	0.0049	3.47E-07
Citramalic acid	0.0068	3.86E-07
Palmitoleic acid	0.0043	5.53E-07
Eicosatetraenoic acid	0.0046	6.19E-07
DG(36:3)	0.0045	1.07E-06
Y-Glutamylleucine	0.0040	1.17E-06
Oleic acid	0.0043	1.61E-06
Hydroxydodecanoic acid	0.0044	2.77E-06
Octadecatrienoic acid	0.0042	3.36E-06
PC(34:4)	0.0042	7.64E-06
PC(32:1)	0.0041	7.79E-06
Hydroxyphenyllactic acid	0.0037	1.09E-05
Octadecadienoic acid	0.0039	1.21E-05
2-Hydroxybutyric acid	0.0041	1.27E-05
Dodecenoic acid	0.0040	1.40E-05
EPA	0.0040	1.41E-05

Ɣ-Glutamylisoleucine	0.0037	1.52E-05
Arachidic acid	0.0040	1.94E-05
Hydroxyhexadecanoic acid	0.0040	2.02E-05
Glutamine	-0.0039	2.21E-05
PC(35:3)	-0.0039	2.79E-05
Eicosenoic acid	0.0037	4.88E-05
MG(14:0)	0.0053	0.000104
Heptadecanedioic acid	0.0035	0.000109
Myristoleic acid	0.0034	0.000109
Docosenoic acid	0.0052	0.000145
N-Acetylneuraminic acid	0.0050	0.000158

Table S4. Association between HbA1c levels and metabolites

Metabolites	β	p-value
DG(34:1)	0.2631	8.02E-09
Leucine	0.2378	1.56E-08
3-Hydroxybutyric acid	0.2841	4.74E-08
Leucine Isoleucine	0.2240	1.03E-07
DG(34:2)	0.2405	1.66E-07
2-Deoxyglucose	-0.2444	3.23E-07
DG(18:1_18:1)	0.2363	4.18E-07
DG(32:0)	0.2243	1.04E-06
Ketoleucine	0.2014	1.88E-06
3-Methyl-2-oxovaleric acid	0.1991	2.36E-06
Isoleucine	0.2006	2.55E-06
MG(16:0)_rp_a	0.2129	1.13E-05
Glutamine	-0.1962	2.69E-05
DG(36:3)	0.1943	3.86E-05

Table S5. Association between log-transformed insulin levels and metabolites

Metabolites	β	p-value
DG(32:1)	0.6498	4.54E-25
DG(34:1)	0.6496	7.24E-25
DG(32:0)	0.6263	3.78E-23
DG(34:2)	0.6248	5.84E-23
Isoleucine	0.5724	1.65E-21
Leucine-Isoleucine	0.4826	5.63E-17
3-Methyl-2-oxovaleric acid	0.4859	6.46E-17
N.Acetylglycine	-0.5868	6.46E-16
Leucine	0.4221	3.18E-13
Ɣ-Glutamyltyrosine	0.4306	1.33E-12
Hydroxyphenyllactic acid	0.4179	4.35E-12
Ɣ-Glutamylisoleucine	0.3818	1.66E-10
Ketoleucine	0.3736	3.11E-10
3-Methylbutyrylcarnitine	0.3846	3.34E-10
Phenylalanine	0.3898	3.91E-10
DG(18:1_18:1)	0.4039	1.00E-09
Glu Phe	0.3521	2.07E-09
CAR(4:0)	0.3901	2.14E-09
Ɣ-Glutamylleucine	0.3432	3.34E-09
CAR(5:0) isomers	0.3466	7.42E-09
CAR(3:0)	0.3758	8.81E-09
DG(36:3)	0.3797	1.10E-08
CAR.5:0)	0.3165	1.52E-07
Adipic acid	0.3382	1.61E-07
Tryptophan	0.3392	2.05E-07
Kynurenine	0.3272	5.41E-07
PC(32:1)	0.3316	5.51E-07
N-Acetylleucine	0.3569	8.49E-07
Taurochenodeoxycholic acid	0.3071	4.63E-06
Taurocholic acid	0.3053	4.97E-06
MG(14:0)	0.3444	6.59E-06
MG(18:1)	0.3380	6.99E-06
Glycocholic acid-Glycohyocholic acid	0.3354	1.02E-05
Pro Phe	0.2584	1.13E-05
Tyrosine	0.2869	1.19E-05
Caffeine	0.2846	2.14E-05
Behenic acid	0.2759	3.07E-05
MG(16:0)_rp_a	0.2970	3.58E-05

Biliverdin	-0.2624	3.66E-05
Glutamine	-0.2707	3.88E-05
PC(40:6)	0.2687	5.46E-05
N-Acetylserine	0.2641	6.06E-05
PC(34:3)	0.2659	7.30E-05
Uric acid	0.2346	7.32E-05
PC(36:3)	0.2635	7.89E-05
Glycochenodeoxycholic acid	0.2604	9.55E-05
dUMP	0.2830	9.57E-05
PC(32:2)	0.2593	0.000103
N-Methyl-D-aspartic acid	0.2541	0.000119

Table S6. Association between log-transformed HOMA-IR levels and metabolites

Metabolites	β	p-value
DG(34:1)	0.6881	6.53E-27
DG(32:1)	0.6722	7.87E-26
DG(34:2)	0.6608	9.55E-25
DG(32:0)	0.6537	2.81E-24
Isoleucine	0.5812	2.47E-21
3-Methyl-2-oxovaleric acid	0.5269	4.50E-19
Leucine-Isoleucine	0.5038	9.06E-18
N-Acetylglycine	-0.5948	1.09E-15
Leucine	0.4456	4.15E-14
Ɣ-Glutamyltyrosine	0.4406	1.09E-12
Hydroxyphenyllactic acid	0.4236	5.89E-12
Ketoleucine	0.4146	6.35E-12
DG(18:1_18:1)	0.4408	5.55E-11
Ɣ-Glutamylisoleucine	0.3966	7.27E-11
3-Methylbutyrylcarnitine	0.3980	1.79E-10
Phenylalanine	0.3983	3.59E-10
Ɣ-Glutamylleucine	0.3672	5.07E-10
DG(36:3)	0.4147	8.55E-10
CAR(5:0) isomers	0.3674	1.76E-09
CAR(4:0)	0.3917	3.78E-09
Glu Phe	0.3490	5.83E-09
CAR(3:0)	0.3731	2.18E-08
CAR(5:0)	0.3351	4.84E-08
Tryptophan	0.3516	1.27E-07
Adipic acid	0.3455	1.51E-07
N-Acetylleucine	0.3763	3.58E-07
PC(32:1)	0.3368	6.07E-07
Tryptophan	0.3350	7.34E-07
Kynurenine	0.3159	2.14E-06
Taurocholic acid	0.3127	4.51E-06
Taurochenodeoxycholic acid	0.3114	5.25E-06
MG(16:0)_rp_a	0.3305	6.62E-06
Glutamine	-0.3010	7.01E-06
MG(14:0)	0.3476	8.42E-06
MG(18:1)	0.3414	8.79E-06
Behenic acid	0.2989	9.26E-06
Glycocholic acid-Glycohyocholic acid	0.3418	1.06E-05

Tyrosine	0.2919	1.24E-05
3-Hydroxybutyric acid	0.2947	1.64E-05
Pro-Phe	0.2586	1.65E-05
Caffeine	0.2899	2.19E-05
PC(34:3)	0.2888	2.34E-05
PC(32:2).	0.2829	3.18E-05
Biliverdin	-0.2626	5.12E-05
dUMP	0.2954	6.50E-05
Glycochenodeoxycholic acid	0.2673	8.53E-05
PC(40:6)	0.2659	9.00E-05
N-Acetylserine	0.2570	0.000134
PC(36:3)	0.2597	0.000137
Chenodeoxycholic acid-Deoxycholic acid	0.2955	0.000153
Glycodeoxycholic acid	0.2571	0.000159

Table S7. Associated metabolites across glucose intolerance status with a linear model

Metabolites	β	p-value
DG(34:2)	0.4364	1.63E-10
Leucine	0.3875	6.28E-10
Leucine Isoleucine	0.3811	1.20E-09
Isoleucine	0.3698	5.40E-09
DG(34:1)	0.3970	5.67E-09
3-Hydroxybutyric acid	0.4190	9.72E-09
3-Methyl-2-oxovaleric acid	0.3493	2.67E-08
DG(32:0)	0.3469	4.19E-07
Ketoleucine	0.3114	8.00E-07
DG(32:1)	0.3303	1.45E-06
Docosatetraenoic acid	0.3311	1.88E-06
MG(16:0)_rp_a	0.3521	1.91E-06
DHA	0.3319	2.65E-06
Ɣ-Glutamylleucine	0.2785	6.30E-06
Glutamine	-0.3026	1.45E-05
Docosatrienoic acid	0.2973	1.91E-05
DG(36:3)	0.2971	2.52E-05
Eicosatrienoic acid	0.2893	2.83E-05
Hydroxytetradecanoic acid	0.2920	3.52E-05
Nonadecenoic acid	0.2798	3.57E-05
Stearic acid	0.2871	4.59E-05
DG(18:1_18:1)	0.2814	5.89E-05

Table S8. Comparison of mean difference of metabolites between normal and all hyperglycemic status

Metabolites	p-value
Leucine Isoleucine	6.42E-11
Isoleucine	2.76E-10
Leucine	7.68E-10
Ɣ-Glutamyltyrosine	9.67E-10
5'-Methylthioadenosine	3.25E-08
Uric acid	2.38E-07
DG(34:1)	4.00E-07
N-Acetylserine	4.37E-07
CAR(5:0) isomers	5.45E-07
DG(34:2)	7.52E-07
3-Methylbutyrylcarnitine	1.44E-06
Glu-Phe	2.83E-06
CAR(5:0)	4.05E-06
S-allylcysteine	4.11E-06
DG(32:1)	4.26E-06
N-Acetylleucine	5.86E-06
Pro-Phe	6.96E-06
Phe-Trp	3.13E-05
DG(32:0)	4.64E-05
CAR(4:0)	6.79E-05
Ɣ-Glutamylleucine	5.05E-10
Ɣ-Glutamylisoleucine	2.17E-07
3-Methyl-2-oxovaleric acid	4.06E-07
Eicosatrienoic acid	1.87E-06
3-Hydroxybutyric acid	4.18E-06
Hydroxytetradecanoic acid	6.29E-06
Phenylalanine	1.10E-05
Hydroxyphenyllactic acid	1.56E-05
Docosatetraenoic acid	3.03E-05
Docosatrienoic acid	4.13E-05
Octadecatrienoic acid	4.73E-05
Nonadecenoic acid	8.53E-05
Ketoleucine	0.000104
Palmitic acid	0.000121

Table S9. Comparison of mean difference of metabolites between normal and diabetes groups

Metabolites	p-value
Leucine	1.89E-14
Leucine Isoleucine	1.77E-13
DG(34:1)	2.29E-13
DG(34:2)	7.18E-13
Isoleucine	1.23E-12
3-Hydroxybutyric acid	1.41E-12
DG(32:0)	3.54E-12
CAR(5:0) isomers	4.66E-12
Ɣ-Glutamylleucine	5.54E-12
Ɣ-Glutamyltyrosine	1.19E-11
DG(32:1)	3.08E-10
CAR(5:0)	4.80E-10
3-Methyl-2-oxovaleric acid	9.42E-10
Ɣ-Glutamylisoleucine	1.21E-09
Docosatetraenoic acid	2.97E-09
3-Methylbutyrylcarnitine	5.39E-09
Ketoleucine	7.10E-09
Nonadecenoic acid	2.75E-08
Docosatrienoic acid	3.10E-08
Tyrosine	5.24E-08
Stearic acid	8.08E-08
MG(16:0)_rp_a	1.38E-07
DG(18:1_18:1)	1.57E-07
Eicosatrienoic acid	3.17E-07
Palmitic acid	4.90E-07
Hydroxyphenyllactic acid	6.39E-07
DHA	1.19E-06
DG(36:3)	3.51E-06
Glu Phe	3.85E-06
Phenylalanine	3.93E-06
Palmitoleic acid	5.89E-06
PC(35:3)	6.67E-06
Tyrosine	6.86E-06
Eicosadienoic acid	1.20E-05
Heptadecanedioic acid	1.27E-05
Glutamine	1.29E-05
Hydroxytetradecanoic acid	1.42E-05
L-Urobilin	1.82E-05

Naproxen	2.04E-05
5’-Methylthioadenosine	2.24E-05
Oleic acid	2.52E-05
2-Hydroxybutyric acid	3.83E-05
N-Acetylneuraminic acid	5.52E-05
Octadecadienoic acid	5.94E-05
Uric acid	6.06E-05
Mesobilirubinogen	9.75E-05
Docosapentaenoic acid	0.000102
Octadecatrienoic acid	0.000138

Table S10. Comparison of mean difference of metabolites between non-diabetes and diabetes groups

Metabolites	p-value
DG(34:2)	1.03E-12
DG(34:1)	2.03E-12
3-Hydroxybutyric acid	4.80E-12
Leucine	5.25E-12
DG(32:0)	7.58E-12
Leucine Isoleucine	6.95E-10
MG(16:0)_rp_a	7.10E-10
CAR(5:0) isomers	1.63E-09
Isoleucine	2.25E-09
DG(32:1)	6.27E-09
Docosatetraenoic acid	1.23E-08
Ɣ-Glutamylleucine	1.29E-08
Nonadecenoic acid	6.24E-08
Ketoleucine	9.66E-08
3-Methyl-2-oxovaleric acid	1.22E-07
Docosatrienoic acid	1.70E-07
Ɣ-Glutamylisoleucine	2.75E-07
DG(18:1_18:1)	3.57E-07
Stearic acid	4.93E-07
CAR(5:0)	6.29E-07
PC(35:3)	7.34E-07
Naproxen	7.40E-07
Glutamine	1.09E-06
3-Methylbutyrylcarnitine	3.26E-06
Mesobilirubinogen	3.97E-06
Ɣ-Glutamyltyrosine	4.79E-06
Ile.Ile	5.25E-06
L-Urobilin	7.75E-06
Heptadecanedioic acid	9.81E-06
DHA	1.65E-05
SM(d35:1)	1.78E-05
Eicosatrienoic acid	2.18E-05
Tyrosine	2.23E-05
Hydroxyphenyllactic acid	2.42E-05
Tyrosine	3.15E-05
DG(36:3)	3.16E-05

Palmitic acid	4.08E-05
2-Hydroxybutyric acid	4.99E-05
Eicosadienoic acid	5.10E-05
Palmitoleic acid	0.000133

Table S11. Association between measured plasma cholesterol levels and metabolites

Metabolites	β	p-value
LPC(20:4)_rp_b	0.0076	4.23E-10
DG(32:0)	0.0081	4.55E-10
DG(18:1_18:1)	0.0081	5.50E-10
LPC(20:3)_rp_b	0.0078	1.21E-09
DG(34:1)	0.0077	2.95E-09
Ketoleucine	0.0069	6.58E-09
SM(d32:1)	0.0074	1.40E-08
PC(34:2)	-0.0071	2.10E-08
MG(18:1)	0.0083	2.20E-08
LPC(20:3)_rp_a	0.0078	6.40E-08
PC(32:0)	-0.0071	1.20E-07
PC(34:4)	0.0068	2.62E-07
SM(d32:2)	0.0060	2.93E-07
SM(d35:1)	-0.0068	3.02E-07
CAR(18:0)	0.0064	5.35E-07
PC(28:0)	0.0068	7.06E-07
DG(36:3)	0.0066	8.63E-07
3-Methyl-2-oxovaleric acid	0.0058	1.38E-06
DG(34:2)	0.0063	1.57E-06
SM(d38:1)	-0.0062	1.98E-06
3-Hydroxybutyric acid	0.0063	2.16E-06
SM(d36:1)	-0.0065	2.35E-06
DG(32:1)	0.0060	4.41E-06
LPC(14:0)	0.0071	4.94E-06
CAR(20:0)	0.0055	8.60E-06
LPC(15:0)_rp_a	0.0058	8.96E-06
LPC(16:0)_rp_a	0.0064	1.18E-05
PC(32:2)α	0.0058	2.00E-05
LPE(18:0)_rp_a	0.0063	2.31E-05
LPC(20:4)_rp_a	0.0063	2.53E-05
LPC(16:1)	0.0063	2.90E-05
Leucine	0.0049	5.32E-05
CAR(5:1)	0.0051	8.02E-05
Deoxyguanosine	0.0062	8.84E-05
Cortisol	0.0052	9.19E-05
Uridine	0.0051	0.000145
SM(d42:2)	-0.0050	0.000158

Table S12. Association between HDL levels and metabolites

Metabolites	β	p-value
DG(34:2)	-0.0493	1.09E-33
DG(36:3)	-0.0497	1.37E-32
DG(34:1)	-0.0394	2.29E-21
DG(18:1_18:1)	-0.0395	6.18E-21
DG(32:1)	-0.0368	2.27E-18
DG(32:0)	-0.0276	1.06E-10
Ɣ-Glutamylisoleucine	-0.0198	4.82E-07
Isoleucine	-0.0199	6.53E-07
α-Tocopherol	-0.0213	1.65E-06
MG(18:1)	-0.0219	1.08E-05
SM(d32:2)	0.0166	1.51E-05
Ɣ-Glutamylleucine	-0.0158	3.21E-05
Cytidine	-0.0182	5.04E-05
Leucine Isoleucine	-0.0155	0.000101

Table S13. Association between Triglyceride levels and metabolites

Metabolites	β	p-value
DG(34:2)	0.0079	5.39E-84
DG(34:1)	0.0077	1.26E-78
DG(32:1)	0.0075	2.21E-72
DG(32:0)	0.0073	3.06E-68
DG(36:3)	0.0075	3.56E-68
DG(18:1_18:1)	0.0073	9.53E-67
MG(18:1)	0.0067	1.66E-42
MG(14:0)	0.0049	2.35E-20
MG(16:0)_rp_a	0.0043	3.94E-17
SM(d35:1)	-0.0039	2.45E-16
3-Methyl-2-oxovaleric acid	0.0032	1.87E-13
PC(32:1)	0.0032	1.82E-11
Isoleucine	0.0029	4.36E-11
Ketoleucine	0.0028	4.38E-11
Leucine Isoleucine	0.0029	4.97E-11
Leucine	0.0027	3.41E-10
SM(d36:1)	-0.0029	3.06E-09
PC(32:2)	0.0029	4.76E-09
LPE(18:0)_rp_a	0.0030	1.16E-08
3-Hydroxybutyric acid	0.0027	1.42E-08
PC(34:3)	0.0027	4.63E-08
PC(34:4)	0.0026	4.65E-08
SM(d40:2)	-0.0025	6.63E-08
Cholesterol (-H2O)	-0.0025	1.60E-07
Ƴ-Glutamylleucine	0.0021	6.72E-07
Citramalic acid	0.0026	1.82E-06
Ƴ-Glutamylisoleucine	0.0020	3.30E-06
SM(d38:1)	-0.0022	3.43E-06
PC(28:0)	0.0023	5.74E-06
Cerotic acid	-0.0020	2.43E-05
PC(30:0)	0.0021	2.56E-05
PC(36:3)	0.0020	3.01E-05
Pipecolic acid	-0.0020	3.90E-05
3-Methylbutyrylcarnitine	0.0018	4.61E-05
SM(d36:2)	-0.0019	4.86E-05
Hydroxyphenyllactic acid	0.0017	5.13E-05
CAR(5:0) isomers	0.0017	7.87E-05
LPC(14:0)	0.0022	8.43E-05

SM(d42:2)	-0.0019	9.61E-05

Table S14. Association between calculated LDL levels and metabolites

Metabolites	β	p-value
LPC(20:4)_rp_b	0.0094	3.02E-10
CAR(18:0)	0.0080	3.32E-07
SM(d32:1)	0.0080	5.44E-07
LPC(20:3)_rp_b	0.0077	9.06E-07
CAR(20:0)	0.0072	2.18E-06
SM(d36:3)	0.0072	3.20E-06
PC(34:2)	-0.0072	3.94E-06
SM(d32:2)	0.0061	1.67E-05
LPC(20:3)_rp_a	0.0079	1.72E-05
CAR(12:0)	0.0068	2.20E-05
Eicosatetraenoic acid	0.0067	2.75E-05
CAR(14:0)	0.0062	0.00012
PC(32:0)	-0.0063	0.000121
CAR(10:0)	0.0061	0.000155

Table S15. Association between non-HDL levels and metabolites

Metabolites	beta	p-value
DG(18:1_18:1)	0.0125	2.07E-21
DG(34:1)	0.0120	2.61E-20
DG(36:3)	0.0119	1.14E-18
DG(34:2)	0.0115	2.15E-18
DG(32:0)	0.0113	8.57E-18
DG(32:1)	0.0100	5.10E-14
MG(18:1)	0.0113	1.76E-13
LPC(20:3)_rp_b	0.0081	7.02E-10
Ketoleucine	0.0075	1.07E-09
LPC(20:4)_rp_b	0.0076	1.97E-09
CAR(18:0)	0.0076	6.35E-09
3-Methyl-2-oxovaleric acid	0.0071	7.70E-09
3-Hydroxybutyric acid	0.0076	2.45E-08
LPC(20:3)_rp_a	0.0083	4.10E-08
SM(d35:1)	-0.0074	7.03E-08
Leucine	0.0065	1.69E-07
PC(34:2)	-0.0068	2.17E-07
Leucine Isoleucine	0.0064	2.90E-07
CAR(20:0)	0.0063	8.14E-07
SM(d32:1)	0.0066	1.17E-06
PC(34:4)	0.0065	1.83E-06
SM(d38:1)	-0.0062	4.36E-06
PC(28:0)	0.0065	4.73E-06
LPC(14:0)	0.0073	6.23E-06
Isoleucine	0.0057	6.58E-06
PC(32:0)	-0.0063	6.89E-06
Ɣ-Glutamylleucine	0.0053	7.70E-06
LPC(16:0)_rp_a	0.0067	1.31E-05
LPC(15:0)_rp_a	0.0058	2.04E-05
PC(32:2)	0.0059	2.09E-05
SM(d36:1)	-0.0059	2.84E-05
LPE(18:0)_rp_a	0.0065	2.87E-05
Eicosatrienoic acid	0.0054	5.78E-05
SM(d32:2)	0.0047	9.18E-05
CAR(5:1)	0.0052	9.59E-05
CAR(14:0)	0.0052	9.98E-05
Cortisone	0.0051	0.000121
Indolelactic acid	0.0048	0.000124

Docosatetraenoic acid	0.0051	0.000127
Deoxyguanosine	0.0063	0.000134
Eicosatetraenoic acid	0.0052	0.000154
CAR(6:0)	0.0051	0.000162

S16. Unidentified metabolites list associated with three diagnostic glycemic traits

Fasting glucose only associated metabolites	2-hour post-load glucose only associated metabolites	HbA1C only associated metabolites	Fasting glucose and 2-hour post-load glucose associated metabolites	Fasting glucose and HbA1C associated metabolites	All three diagnostic glycemic traits associated metabolites
UNK_432.2382_1.836	UNK_323.2263_12.102	UNK_207.0139_0.613	UNK_129.0607_2.142	UNK_323.0798_0.668	UNK_441.0745_0.632
UNK_439.255_1.837	UNK_434.2102_12.104	UNK_191.0402_0.615	UNK_117.0558_3.253	UNK_215.0333_0.64	UNK_340.1072_2.37
UNK_301.1431_1.836	UNK_343.1946_11.516	UNK_801.612_13.735	UNK_792.5547_14.394	UNK_145.0621_0.623	UNK_277.0742_0.946
UNK_547.3289_1.927	UNK_345.1934_12.103	UNK_1197.7095_12.081	UNK_1036.0408_2.585	UNK_527.0944_0.624	UNK_115.0401_2.585
UNK_597.3346_1.927	UNK_349.2403_12.19		UNK_132.1023_1.836	UNK_313.2732_11.408	UNK_154.0839_1.838
UNK_763.5605_12.723	UNK_400.1706_11.512		UNK_463.0889_0.959	UNK_231.0001_2.181	UNK_86.0967_1.836
UNK_262.1287_1.22	UNK_565.5184_11.782		UNK_202.1086_5.755		UNK_132.1023_1.96
UNK_217.0651_2.985	UNK_305.2476_11.514		UNK_329.2696_12.045		UNK_634.4449_12.16
	UNK_359.1674_11.513		UNK_482.024_2.585		UNK_86.0968_1.96
	UNK_172.9971_2.586		UNK_131.0715_5.141		UNK_154.0839_1.959
	UNK_395.2201_11.472		UNK_387.1641_11.499		UNK_432.2383_1.956
	UNK_657.5048_12.045		UNK_351.2301_11.5		UNK_455.2266_1.957
	UNK_659.5057_11.472		UNK_405.1906_1.845		UNK_183.0274_2.587
	UNK_725.4913_12.046		UNK_285.2071_10.349		UNK_253.0695_2.583
	UNK_514.1668_11.472				UNK_315.0407_2.584
	UNK_963.7574_12.045				UNK_349.0819_2.12
	UNK_301.2174_11.263				UNK_568.3814_1.928
	UNK_811.6771_11.605				UNK_229.0695_2.119
	UNK_385.3109_12.278				UNK_411.044_1.065
	UNK_187.1341_8.764				
	UNK_197.1546_10.498				
	UNK_371.2206_11.488				
	UNK_271.2282_10.48				
	UNK_494.1489_7.143				
	UNK_259.2431_11.486				
	UNK_831.6585_14.897				
	UNK_363.1764_11.257				
	UNK_203.0143_1.069				
	UNK_882.6231_14.642				
	UNK_768.5546_14.354				
	UNK_856.6088_14.313				
	UNK_201.0385_1.063				
	UNK_790.5963_14.271				
	UNK_594.3774_11.499				
	UNK_788.5804_14.159				
	UNK_146.0462_0.63				

UNK_477.1044_1.958

UNK_409.2345_9.292

UNK_439.255_1.956

UNK_245.129_5.953

UNK_816.6104_14.5

UNK_444.254_12.103

UNK_805.6396_14.771

UNK_881.6741_15.005

UNK_204.0487_0.972

UNK_187.0427_1.066

UNK_187.0429_0.967

UNK_799.596_13.234

UNK_1010.0267_0.973

UNK_357.2994_11.514

UNK_379.2817_11.507

UNK_790.5392_13.932

UNK_204.0485_1.062

UNK_688.492_13.189

UNK_197.0804_7.936

UNK_527.2142_7.893

UNK_236.0931_5.91

UNK_259.1302_4.131

UNK_682.5178_12.103

UNK_786.5652_13.647

UNK_629.3542_9.116

UNK_217.0375_5.086

UNK_339.1982_12.102

UNK_131.0715_5.033

UNK_716.5235_13.801

UNK_776.5462_13.155

UNK_878.5931_13.884

UNK_356.19_11.688

UNK_301.2111_11.687

UNK_239.0582_11.541

UNK_809.4335_11.784

UNK_317.1619_11.688

UNK_935.5938_11.541

UNK_311.1672_11.687

UNK_527.1757_11.786

UNK_267.1634_11.281

UNK_941.6412_11.782

UNK_313.1678_11.688

UNK_321.2108_11.782

UNK_641.429_11.415

UNK_573.1808_11.787

UNK_603.4653_11.782

UNK_327.2268_11.782

UNK_287.1319_11.281

UNK_299.1954_11.397

UNK_219.2106_11.398

UNK_341.1794_11.784

UNK_285.1363_11.281

UNK_943.6406_11.782

UNK_313.148_11.397

UNK_283.1363_11.281

UNK_225.0421_11.328

UNK_339.1836_11.784

UNK_363.2517_11.929

UNK_399.2517_11.785

UNK_852.6497_11.65

UNK_805.6592_11.651

UNK_770.3734_11.651

UNK_511.4728_11.651

UNK_883.7055_11.743

UNK_655.3664_11.303

UNK_573.4917_11.256

UNK_644.4324_11.742

UNK_599.5076_11.373

UNK_899.6794_11.743

UNK_1179.881_11.51

UNK_213.1499_9.45

UNK_623.5061_11.305

UNK_349.2364_11.744

UNK_649.4426_11.51

UNK_347.2208_11.51

UNK_321.2051_11.37

UNK_1195.8522_11.509

UNK_400.1484_11.745

UNK_930.6967_11.742

UNK_415.2074_11.51

UNK_265.217_11.33

UNK_1176.2343_11.655

UNK_253.217_11.551

UNK_877.6587_11.51

UNK_398.1327_11.509

UNK_762.5079_13.262

UNK_792.5546_14.167

UNK_749.5809_13.18

UNK_385.1678_11.499

UNK_279.035_1.091

UNK_305.1568_4.502

UNK_383.1682_11.499

UNK_424.1703_11.499

UNK_899.5064_7.228

UNK_889.7527_12.045

UNK_567.5352_12.045

UNK_347.2259_11.908

UNK_273.1109_8.681

UNK_613.3594_9.663

UNK_327.1833_11.888

UNK_850.5955_11.468

UNK_257.2473_11.687

UNK_366.2227_12.101

UNK_325.1836_11.891

UNK_648.4623_12.045

UNK_973.7129_12.045

UNK_957.7393_12.045

UNK_402.1639_12.045

UNK_936.7426_12.045

UNK_905.726_12.045

UNK_203.1807_10.716

UNK_376.1705_11.539

UNK_387.1991_11.824

UNK_389.199_11.825

UNK_338.1914_11.688

UNK_416.2227_11.688

UNK_295.1892_11.256

UNK_826.5618_13.082

UNK_293.2482_11.69

UNK_355.1587_11.651

UNK_592.4016_11.65

UNK_381.1744_11.744

UNK_365.199_11.906

UNK_304.1154_2.887
UNK_372.1171_11.371
UNK_814.5604_13.177
UNK_696.4421_11.78
UNK_385.1252_2.631
UNK_293.2124_10.775
UNK_247.1701_10.693
UNK_371.2257_11.823
UNK_337.1482_11.327
UNK_454.2224_12.046
UNK_345.2046_11.304
UNK_681.3474_9.665
UNK_505.1462_6.638
UNK_374.1325_11.65
UNK_335.1522_11.327
UNK_533.4552_11.65

REFERENCES

1. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. Nature Reviews Nephrology 2020;16(7):377-390

2. Khan RMM, Chua ZJY, Tan JC, Yang Y, Liao Z, Zhao Y. From pre-diabetes to diabetes: Diagnosis, treatments and translational research. Medicina 2019;55(9):546

3. Forouhi NG, Wareham NJ. Epidemiology of diabetes. Medicine 2019;47(1):22-27

4. Galicia-Garcia U, Benito-Vicente A, Jebari S, Larrea-Sebal A, Siddiqi H, Uribe KB, Ostolaza H, Martín C. Pathophysiology of type 2 diabetes mellitus. International Journal of Molecular Sciences 2020;21(17):6275

5. Lakhtakia R. The history of diabetes mellitus. Sultan Qaboos University Medical Journal 2013;13(3):368

6. Echouffo-Tcheugui JB, Selvin E. Prediabetes and what it means: The epidemiological evidence. Annu Rev Public Health 2021;42:59-77

7. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, Collins BS, Hilliard ME, Isaacs D, Johnson EL. 2. classification and diagnosis of diabetes: Standards of care in diabetes—2023. Diabetes Care 2023;46(Supplement_1):S19-S40

8. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, Stein C, Basit A, Chan JC, Mbanya JC. IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119

9. Zhou B, Lu Y, Hajifathalian K, Bentham J, Di Cesare M, Danaei G, Bixby H, Cowan MJ, Ali MK, Taddei C. Worldwide trends in diabetes since 1980: A pooled analysis of 751 population-based studies with 4· 4 million participants. The Lancet 2016;387(10027):1513-1530

10. Ahmad FB, Anderson RN. The leading causes of death in the US for 2020. Jama 2021;325(18):1829-1830

11. Heron MP. Deaths: Leading causes for 2017. National Vital Statistics Reports 2019;68(6):1-76

12. Tomic D, Shaw JE, Magliano DJ. The burden and risks of emerging complications of diabetes mellitus. Nature Reviews Endocrinology 2022;18(9):525-539

13. Standl E, Khunti K, Hansen TB, Schnell O. The global epidemics of diabetes in the 21st century: Current situation and perspectives. European Journal of Preventive Cardiology 2019;26(2_suppl):7-14

14. Centers for Disease Control and Prevention. National diabetes statistics report, 2020. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services 2020:1-30

15. QuickFacts United States - Population estimates [article online], 2022. Available from https://www.census.gov/quickfacts/fact/table/US/PST045222. Accessed Mar 20 2022

16. Montonen J, Knekt P, Härkänen T, Järvinen R, Heliövaara M, Aromaa A, Reunanen A. Dietary patterns and the incidence of type 2 diabetes. Am J Epidemiol 2005;161(3):219-227

17. Wheeler ML, Dunbar SA, Jaacks LM, Karmally W, Mayer-Davis EJ, Wylie-Rosett J, Yancy WS. Macronutrients, food groups, and eating patterns in the management of diabetes. Diabetes Care 2012;35(2):434-445

18. Yuan X, Wang J, Yang S, Gao M, Cao L, Li X, Hong D, Tian S, Sun C. Effect of the ketogenic diet on glycemic control, insulin resistance, and lipid metabolism in patients with T2DM: A systematic review and meta-analysis. Nutrition & Diabetes 2020;10(1):38

19. Goldenberg JZ, Day A, Brinkworth GD, Sato J, Yamada S, Jönsson T, Beardsley J, Johnson JA, Thabane L, Johnston BC. Efficacy and safety of low and very low carbohydrate diets for type 2 diabetes remission: Systematic review and meta-analysis of published and unpublished randomized trial data. Bmj 2021;372

20. Willett WC, Bhupathiraju SN, Chiuve SE, Borgi L, Manson JE, Spiegelman D, Satija A, Sun Q, Hu FB, Rimm EB. Plant-based dietary patterns and incidence of type 2 diabetes in US men and women: Results from three prospective cohort studies. PLoS Medicine 2016;13(6)

21. Panagiotakos DB, Tzima N, Pitsavos C, Chrysohoou C, Zampelas A, Toussoulis D, Stefanadis C. The association between adherence to the mediterranean diet and fasting indices of glucose homoeostasis: The ATTICA study. J Am Coll Nutr 2007;26(1):32-38

22. Titus SK, Kataoka-Yahiro M. A systematic review of barriers to access-to-care in hispanics with type 2 diabetes. Journal of Transcultural Nursing 2019;30(3):280-290

23. Siega-Riz AM, Pace ND, Butera NM, Van Horn L, Daviglus ML, Harnack L, Mossavar-Rahmani Y, Rock CL, Pereira RI, Sotres-Alvarez D. How well do US hispanics adhere to the dietary guidelines for americans? results from the hispanic community health study/study of latinos. Health Equity 2019;3(1):319-327

24. DeForest N, Majithia AR. Genetics of type 2 diabetes: Implications from large-scale studies. Current Diabetes Reports 2022;22(5):227-235

25. Norris JM, Rich SS. Genetics of glucose homeostasis: Implications for insulin resistance and metabolic syndrome. Arterioscler Thromb Vasc Biol 2012;32(9):2091-2096

26. Mambiya M, Shang M, Wang Y, Li Q, Liu S, Yang L, Zhang Q, Zhang K, Liu M, Nie F. The play of genes and non-genetic factors on type 2 diabetes. Frontiers in Public Health 2019;7:349

27. del Bosque-Plata L, Martínez-Martínez E, Espinoza-Camacho MÁ, Gragnoli C. The role of TCF7L2 in type 2 diabetes. Diabetes 2021;70(6):1220-1228

28. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S. Large-scale association studies of variants in genes encoding the pancreatic β-cell KATP channel subunits Kir6. 2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 2003;52(2):568-572

29. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316(5829):1331-1336

30. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445(7130):881-885

31. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42(2):105-116

32. Rasmussen-Torvik LJ, Guo X, Bowden DW, Bertoni AG, Sale MM, Yao J, Bluemke DA, Goodarzi MO, Chen YI, Vaidya D. Fasting glucose GWAS candidate region

analysis across ethnic groups in the multiethnic study of atherosclerosis (MESA). Genet Epidemiol 2012;36(4):384-391

33. Comuzzie AG, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, Butte NF. Novel genetic loci identified for the pathophysiology of childhood obesity in the hispanic population. PloS One 2012;7(12):e51954

34. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL. Genetic analyses of diverse populations improves discovery for complex traits. Nature 2019;570(7762):514-518

35. Wan JY, Goodman DL, Willems EL, Freedland AR, Norden-Krichmar TM, Santorico SA, Edwards KL. Genome-wide association analysis of metabolic syndrome quantitative traits in the GENNID multiethnic family study. Diabetology & Metabolic Syndrome 2021;13(1):1-15

36. DiCorpo D, Gaynor SM, Russell EM, Westerman KE, Raffield LM, Majarian TD, Wu P, Sarnowski C, Highland HM, Jackson A. Whole genome sequence association analysis of fasting glucose and fasting insulin levels in diverse cohorts from the NHLBI TOPMed program. Communications Biology 2022;5(1):756

37. Moon J, Louie TL, Jain D, Sofer T, Schurmann C, Below JE, Lai C, Aviles-Santa ML, Talavera GA, Smith CE. A genome-wide association study identifies blood Disorder–Related variants influencing hemoglobin A1c with implications for glycemic status in US hispanics/latinos. Diabetes Care 2019;42(9):1784-1791

38. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, Willems SM, Wu Y, Zhang X, Horikoshi M. The trans-ancestral genomic architecture of glycemic traits. Nat Genet 2021;53(6):840-860

39. Downie CG, Dimos SF, Bien SA, Hu Y, Darst BF, Polfus LM, Wang Y, Wojcik GL, Tao R, Raffield LM. Multi-ethnic GWAS and fine-mapping of glycaemic traits identify novel loci in the PAGE study. Diabetologia 2022;65(3):477-489

40. Li JH, Brenner LN, Kaur V, Figueroa K, Schroeder P, Huerta-Chagoya A, Udler MS, Leong A. Genome-wide association analysis identifies ancestry-specific genetic variation associated with acute response to metformin and glipizide in SUGAR-MGH. Diabetologia 2023:1-13

41. Li S. *Computational methods and data analysis for metabolomics. Computational methods and data analysis for metabolomics.*, Springer, 2020

42. Kell DB, Oliver SG. The metabolome 18 years on: A concept comes of age. Metabolomics 2016;12:1-8

43. Pallares-Méndez R, Aguilar-Salinas CA, Cruz-Bautista I, del Bosque-Plata L. Metabolomics in diabetes, a review. Ann Med 2016;48(1-2):89-102

44. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B. Novel biomarkers for pre-diabetes identified by metabolomics. Molecular Systems Biology 2012;8(1):615

45. Zeng H, Tong R, Tong W, Yang Q, Qiu M, Xiong A, Sun S, Ding L, Zhang H, Yang L, Tian J. Metabolic biomarkers for prognostic prediction of pre-diabetes: Results from a longitudinal cohort study. Sci Rep 2017;7(1):6575

46. Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J, Hu FB. Metabolomics in prediabetes and diabetes: A systematic review and meta-analysis. Diabetes Care 2016;39(5):833-846

47. Morze J, Wittenbecher C, Schwingshackl L, Danielewicz A, Rynkiewicz A, Hu FB, Guasch-Ferré M. Metabolomics and type 2 diabetes risk: An updated systematic review and meta-analysis of prospective cohort studies. Diabetes Care 2022;45(4):1013-1024

48. Mcpherson RS, Kohl HW, Garcia G, Nichaman MZ, Hanis CL. Food-frequency questionnaire validation among mexican-americans: Starr county, texas. Ann Epidemiol 1995;5(5):378-385

49. Sarafidis PA, Lasaridis AN, Nilsson PM, Pikilidou MI, Stafilas PC, Kanaki A, Kazakos K, Yovos J, Bakris GL. Validity and reproducibility of HOMA-IR, 1/HOMA-IR, QUICKI and McAuley's indices in patients with hypertension and type II diabetes. J Hum Hypertens 2007;21(9):709-716

50. Boles A, Kandimalla R, Reddy PH. Dynamics of diabetes and obesity: Epidemiological perspective. Biochimica Et Biophysica Acta (BBA)-Molecular Basis of Disease 2017;1863(5):1026-1036

51. Jung K. Tietz textbook of clinical chemistry and molecular diagnostics, fourth edition. Clinical Chemistry (Baltimore, Md.) 2006;52(6):938-952

52. Jun G, Aguilar D, Evans C, Burant CF, Hanis CL. Metabolomic profiles associated with subtypes of prediabetes among mexican americans in starr county, texas, USA. Diabetologia 2020 Feb;63(2):287-295

53. Willett W. *Nutritional epidemiology.* 3rd ed. *Nutritional epidemiology.* Oxford, Oxford University Press, 2012 STAT!Ref electronic medical library.; ID: cdi_askewsholts_vlebooks_9780190240844

54. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, Szpiro AA, Chen W, Brehm JM, Celedón JC. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. The American Journal of Human Genetics 2016;98(4):653-666

55. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet 2012;44(7):821-824

56. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res 2002;12(6):996-1006

57. Wang X, Lim E, Liu C, Sung YJ, Rao DC, Morrison AC, Boerwinkle E, Manning AK, Chen H. Efficient gene–environment interaction tests for large biobank-scale sequencing studies. Genet Epidemiol 2020;44(8):908-923

58. Nutrient Intakes per 1000 kcal from Food and Beverages: Mean Energy and Mean Nutrient Amounts per 1000 kcal
Consumed per Individual, by Race/Ethnicity and Age, What We Eat in America, NHANES 2017-March 2020 Prepandemic. [article online], 2022. Available from https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/1720/Table_42_DEN_RAC_1720.pdf. Accessed June, 22 2023

59. Kolczynska K, Loza-Valdes A, Hawro I, Sumara G. Diacylglycerol-evoked activation of PKC and PKD isoforms in regulation of glucose and lipid metabolism: A review. Lipids in Health and Disease 2020;19:1-15

60. Eichmann TO, Kumari M, Haas JT, Farese RV, Zimmermann R, Lass A, Zechner R. Studies on the substrate and stereo/regioselectivity of adipose triglyceride lipase, hormone-sensitive lipase, and diacylglycerol-O-acyltransferases. J Biol Chem 2012;287(49):41446-41457

61. Schweiger M, Romauch M, Schreiber R, Grabner GF, Hütter S, Kotzbeck P, Benedikt P, Eichmann TO, Yamada S, Knittelfelder O. Pharmacological inhibition of adipose triglyceride lipase corrects high-fat diet-induced insulin resistance and hepatosteatosis in mice. Nature Communications 2017;8(1):14859

62. Erion DM, Shulman GI. Diacylglycerol-mediated insulin resistance. Nat Med 2010;16(4):400-402

63. Newman JC, Verdin E. B-hydroxybutyrate: Much more than a metabolite. Diabetes Res Clin Pract 2014 Nov;106(2):173-181

64. Møller N. Ketone body, 3-hydroxybutyrate: Minor metabolite - major medical manifestations. J Clin Endocrinol Metab 2020;105(9):2884-2892

65. Newman JC, Verdin E. Ketone bodies as signaling metabolites. Trends in Endocrinology & Metabolism 2014;25(1):42-52

66. Kim H, Lichtenstein AH, White K, Wong KE, Miller III ER, Coresh J, Appel LJ, Rebholz CM. Plasma metabolites associated with a Protein-Rich dietary pattern: Results from the OmniHeart trial. Molecular Nutrition & Food Research 2022;66(6):2100890

67. Bernard L, Chen J, Kim H, Huang Z, Bazzano LA, Qi L, He J, Rao V, Kelly TN, Potts K. Abstract P204: Associations between dairy intake and serum metabolites in middle-aged adults in the US. Circulation 2023;147(Suppl_1):AP204

68. Whytock KL, Corbin KD, Parsons SA, Pachori A, Bock CP, Jones KP, Smith JS, Yi F, Xie H, Petucci CJ. Metabolic adaptation characterizes short-term resistance to weight loss induced by a low-calorie diet in overweight/obese individuals. Am J Clin Nutr 2021;114(1):267-280

69. Kim J, Hong S, Park J, Kim S. Metabolic and transcriptomic changes in the mouse brain in response to short-term high-fat metabolic stress. Metabolites 2023;13(3):407

70. Suvitaival T, Bondia-Pons I, Yetukuri L, Pöhö P, Nolan JJ, Hyötyläinen T, Kuusisto J, Orešič M. Lipidome as a predictive tool in progression to type 2 diabetes in finnish men. Metab Clin Exp 2018;78:1-12

71. Lu L, Koulman A, Petry CJ, Jenkins B, Matthews L, Hughes IA, Acerini CL, Ong KK, Dunger DB. An unbiased lipidomics approach identifies early second trimester lipids predictive of maternal glycemic traits and gestational diabetes mellitus. Diabetes Care 2016;39(12):2232-2239

72. Grace MS, Dempsey PC, Sethi P, Mundra PA, Mellett NA, Weir JM, Owen N, Dunstan DW, Meikle PJ, Kingwell BA. Breaking up prolonged sitting alters the postprandial plasma lipidomic profile of adults with type 2 diabetes. J Clin Endocrinol Metab 2017 Jun 1;102(6):1991-1999

73. Kumar A, Sundaram K, Mu J, Dryden GW, Sriwastva MK, Lei C, Zhang L, Qiu X, Xu F, Yan J, Zhang X, Park JW, Merchant ML, Bohler HCL, Wang B, Zhang S, Qin C, Xu Z, Han X, McClain CJ, Teng Y, Zhang H. High-fat diet-induced upregulation of exosomal

phosphatidylcholine contributes to insulin resistance. Nature Communications; Nat Commun 2021;12(1):213

74. Gao X, Du L, Randell E, Zhang H, Li K, Li D. Effect of different phosphatidylcholines on high fat diet-induced insulin resistance in mice. Food & Function 2021;12(4):1516-1528

75. Thevenon J, Souchay C, Seabold GK, Dygai-Cochet I, Callier P, Gay S, Corbin L, Duplomb L, Thauvin-Robinet C, Masurel-Paulet A. Heterozygous deletion of the LRFN2 gene is associated with working memory deficits. European Journal of Human Genetics 2016;24(6):911-918

76. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Wuttke M, Sarnowski C, Magill R, Nano J, Gieger C, Trompet S, Prins BP, Guo XQ, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng M, Petty LE, Sim XL, Zhang WH, Brummett CM, Kardia S, Kronenberg F, Lall K, Liu CT, Locke AE, Luan JA, Ntalla L, Schonherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Frayling TM, Gieaitis V, Hackinger S, Hattersley AT, Herder C, Ikram A, Jorgensen ME, Jorgensen T, Kriebel J, Kuusisto J, Ligthart S, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Sattar N, Stancakova A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte, Dupuis J, Peyser PA, Zeggini E, Loos R, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer C, Grallert H, Metspalu A, Dehghan A, Kottgen A, Abecasis GR, Meigs JB, Rotter JI, Marchini J, Pedersen O, Hansen T, Langenberg C, Wareham NJ, Stefansson K, Gloyn AL, Morris AP, Boehnke M, McCarthy MI. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 2018;50(11):1505-+

77. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, Huffman JE, Assimes TL, Lorenz K, Zhu X, Hilliard AT, Judy RL, Huang J, Lee KM, Klarin D, Pyarajan S, Danesh J, Melander O, Rasheed A, Mallick NH, Hameed S, Qureshi IH, Afzal MN, Malik U, Jalal A, Abbas S, Sheng X, Gao L, Kaestner KH, Susztak K, Sun YV, DuVall SL, Cho K, Lee JS, Gaziano JM, Phillips LS, Meigs JB, Reaven PD, Wilson PW, Edwards TL, Rader DJ, Damrauer SM, O'Donnell CJ, Tsao PS, Chang K, Voight BF, Saleheen D. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet 2020;52(7):680-691

78. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M, Ishigaki K, Suzuki A, Suzuki K, Obara W, Yamaji K, Takahashi K, Asai S, Takahashi Y, Suzuki T, Shinozaki N, Yamaguchi H, Minami S, Murayama S, Yoshimori K, Nagayama S, Obata D, Higashiyama M, Masumoto A,

Koretsune Y, Ito K, Terao C, Yamauchi T, Komuro I, Kadowaki T, Tamiya G, Yamamoto M, Nakamura Y, Kubo M, Murakami Y, Yamamoto K, Kamatani Y, Palotie A, Rivas MA, Daly MJ, Matsuda K, Okada Y. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet 2021;53(10):1415-1424

79. Mahajan A, Petty LE, Chiou J, Miguel-Escalada I, Nakatochi M, Huerta-Chagoya A, Bielak LF, Hai Y, Kals M, Grarup N, Wuttke M, Nousome D, Long J, Ahmad M, Jensen RA, An P, Cade BE, Abaitua F, Akiyama M, Bertoni A, Bian Z, Brummett CM, Chee M, Das SK, Du S, Duan Q, Eckardt K, Fischer K, Floyd JS, Genter P, González-Villalpando ME, Gordon-Larsen P, Gorkin D, Herder C, Hung Y, Hwu C, Ikram MA, Ingelsson M, Islam MT, Isono M, Kamatani Y, Katsuya T, Khor C, Kohara K, Kriebel J, Kuusisto J, Läll K, Leong A, Luan J, Lv J, Lyssenko V, Meitinger T, Morris AD, Nadkarni GN, Nayak U, Ntalla I, Peters A, Preissl S, Rasmussen-Torvik L, Roden M, Sabanayagam C, Schönherr S, Shahriar M, Shi J, Stančáková A, Strauch K, Thorleifsson G, Thorsteinsdottir U, Tsai F, Tusie-Luna T, Udler MS, Valladares-Salgado A, Yajnik CS, Yoon K, Igase M, Hanis CL, Ingelsson E, Zeggini E, Rich SS, Kooperberg C, Engert JC, Wilson JG, Kardia SLR, Groop L, Bharadwaj D, Sale MM, Cruz M, Tai E, Kato N, Laakso M, Köttgen A, Loos RJF, Saleheen D, Hansen T, Kadowaki T, Walters RG, Stefansson K, Meigs JB, Gloyn AL, Chambers JC. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 2022;54(5):560-572

80. Mansour Aly D, Dwivedi OP, Prasad RB, Käräjämäki A, Hjort R, Thangam M, Åkerlund M, Mahajan A, Udler MS, Florez JC, McCarthy MI, Brosnan J, Melander O, Carlsson S, Hansson O, Tuomi T, Groop L, Ahlqvist E. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. Nat Genet 2021;53(11):1534-1542

81. Mansour A, Mousa M, Abdelmannan D, Tay G, Hassoun A, Alsafar H. Microvascular and macrovascular complications of type 2 diabetes mellitus: Exome wide association analyses. Frontiers in Endocrinology 2023;14:1143067

82. Dehghan A, Bis JC, White CC, Smith AV, Morrison AC, Cupples LA, Trompet S, Chasman DI, Lumley T, Völker U. Genome-wide association study for incident myocardial infarction and coronary heart disease in prospective cohort studies: The CHARGE consortium. PloS One 2016;11(3):e0144997

83. Yu X, Cao R, Yang Y, Deng F, Bo L, Lei S. Body surface area is a potential obesity index: Its genetic determination and its causality for later-life diseases. Obesity 2023;31(1):256-266

84. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, Yengo L, Ferreira T, Marouli E, Ji Y. Meta-analysis of genome-wide association studies for body

fat distribution in 694 649 individuals of european ancestry. Hum Mol Genet 2019;28(1):166-174

85. Karlsson T, Rask-Andersen M, Pan G, Höglund J, Wadelius C, Ek WE, Johansson Å. Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. Nat Med 2019;25(9):1390-1395

86. Deng X, Wang B, Fisher V, Peloso G, Cupples A, Liu C. Genome-wide association study for multiple phenotype analysis. In *BMC proceedings*, Springer, p. 139-144

87. Graham SE, Clarke SL, Wu KH, Kanoni S, Zajac GJ, Ramdas S, Surakka I, Ntalla I, Vedantam S, Winkler TW. The power of genetic diversity in genome-wide association studies of lipids. Nature 2021;600(7890):675-679

88. Tachmazidou I, Süveges D, Min JL, Ritchie GR, Steinberg J, Walter K, Iotchkova V, Schwartzentruber J, Huang J, Memari Y. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. The American Journal of Human Genetics 2017;100(6):865-884

89. Spear ML, Diaz-Papkovich A, Ziv E, Yracheta JM, Gravel S, Torgerson DG, Hernandez RD. Recent shifts in the genomic ancestry of mexican americans may alter the genetic architecture of biomedical traits. Elife 2020;9:e56029

90. Essigmann HT, Aguilar DA, Perkison WB, Bay KG, Deaton MR, Brown SA, Hanis CL, Brown EL. Epidemiology of antibiotic use and drivers of cross-border procurement in a mexican american border community. Frontiers in Public Health 2022;10

91. McCurley JL, Gutierrez AP, Gallo LC. Diabetes prevention in US hispanic adults: A systematic review of culturally tailored interventions. Am J Prev Med 2017;52(4):519-529

92. McPherson RS, Kohl III HW, Garcia G, Nichaman MZ, Hanis CL. Food-frequency questionnaire validation among mexican-americans: Starr county, texas. Ann Epidemiol 1995;5(5):378-385

93. Wang X, Lim E, Liu C, Sung YJ, Rao DC, Morrison AC, Boerwinkle E, Manning AK, Chen H. Efficient gene–environment interaction tests for large biobank-scale sequencing studies. Genet Epidemiol 2020;44(8):908-923