HHS Public Access
Author manuscript
Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2022 December 01.

Large-scale integrated analysis of genetics and metabolomic data reveals potential links between lipids and colorectal cancer risk

Xiang Shu[1,2], Zhishan Chen[2], Jirong Long[2], Xingyi Guo[2], Yaohua Yang[2], Conghui Qu[3], Yoon-Ok Ahn[4], Qiuyin Cai[2], Graham Casey[5], Stephen B. Gruber[6], Jeroen R. Huyghe[3], Sun Ha Jee[7], Mark A. Jenkins[8], Wei-Hua Jia[9], Keum Ji Jung[7], Yoichiro Kamatani[10,11], Dong-Hyun Kim[12], Jeongseon Kim[13], Sun-Seog Kweon[14], Loic Le Marchand[15], Koichi Matsuda[16], Keitaro Matsuo[17,18], Polly A. Newcomb[3,19], Jae Hwan Oh[20], Jennifer Ose[21], Isao Oze[22], Rish K. Pai[23], Zhi-Zhong Pan[9], Paul D.P. Pharoah[24], Mary C. Playdon[25], Ze-Fang Ren[26], Robert E. Schoen[27], Aesun Shin[28,29], Min-Ho Shin[14], Xiao-ou Shu[2], Xiaohui Sun[1,30], Catherine M. Tangen[31], Chizu Tanikawa[32], Cornelia M. Ulrich[22], Franzel J.B. van Duijnhoven[33], Bethany Van Guelpen[34,35], Alicja Wolk[36], Michael O. Woods[37], Anna H. Wu[38], Ulrike Peters[3,39], Wei Zheng[2]

[1.]Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[2.]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA

[3.]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

[4.]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, South Korea

[5.]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA

[6.]Department of Preventive Medicine & USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[7.]Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea

[8.]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia

[9.]State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University, Guangzhou, China

[10.]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan

Correspondence Dr. Xiang Shu, Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, 485 Lexington Ave, 2nd Floor, New York, NY 10017, USA, ShuX@mskcc.org.

The authors declare no potential conflicts of interest.

[11.] Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

[12.] Department of Social and Preventive Medicine, Hallym University College of Medicine, Okcheon-dong, Korea

[13.] Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Gyeonggi-do, South Korea

[14.] Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, South Korea

[15.] University of Hawaii Cancer Center, Honolulu, Hawaii, USA

[16.] Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan

[17.] Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, Japan

[18.] Department of Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan

[19.] School of Public Health, University of Washington, Seattle, Washington, USA

[20.] Center for Colorectal Cancer, National Cancer Center Hospital, National Cancer Center, Gyeonggi-do, South Korea

[21.] Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, Utah, USA

[22.] Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan

[23.] Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, Arizona, USA

[24.] Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[25.] Cancer Control and Population Sciences, Huntsman Cancer Institute and Department of Nutrition and Integrative Physiology, University of Utah, Salt Lake City, Utah, USA

[26.] School of Public Health, Sun Yat-sen University, Guangzhou, China

[27.] Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

[28.] Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

[29.] Cancer Research Institute, Seoul National University, Seoul, Korea

[30.] Department of Epidemiology, Zhejiang Chinese Medical University, Zhejiang, China

[31.] SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

[32.] Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan

33.Division of Human Nutrition, Wageningen University and Research, Wageningen, The Netherlands

34.Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden

35.Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden

36.Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

37.Memorial University of Newfoundland, Discipline of Genetics, St. John's, Canada

38.University of Southern California, Preventative Medicine, Los Angeles, California, USA

39.Department of Epidemiology, University of Washington, Seattle, Washington, USA

Abstract

Background: The etiology of colorectal cancer is not fully understood.

Methods: Using genetic variants and metabolomics data including 217 metabolites from the Framingham Heart Study (n = 1,357), we built genetic prediction models for circulating metabolites. Models with prediction $R^2 > 0.01$ ($N_{metabolite} = 58$) were applied to predict levels of metabolites in two large consortia with a combined sample size of approximately 46,300 cases and 59,200 controls of European and approximately 21,700 cases and 47,400 controls of East Asian (EA) descent. Genetically predicted levels of metabolites were evaluated for their associations with colorectal cancer risk in logistic regressions within each racial group, after which the results were combined by meta-analysis.

Results: Of the 58 metabolites tested, 24 metabolites were significantly associated with colorectal cancer risk [Benjamini–Hochberg FDR (BH-FDR) < 0.05] in the European population (odds ratios [ORs] ranged from 0.91 to 1.06; P-values ranged from 0.02 to 6.4×10^{-8}). Twenty one of the 24 associations were replicated in the EA population (ORs ranged from 0.26 to 1.69, BH-FDR < 0.05). In addition, the genetically predicted levels of C16:0 cholesteryl ester was significantly associated with colorectal cancer risk in the EA population only (OR_{EA}: 1.94, 95% CI, 1.60–2.36, P = 2.6×10^{-11}; OR_{EUR}: 1.01, 95% CI, 0.99–1.04, P = 0.3). Nineteen of the 25 metabolites were glycerophospholipids and triacylglycerols (TAG). Eighteen associations exhibited significant heterogeneity between the two racial groups ($P_{EUR-EA-Het} < 0.005$), which were more strongly associated in the EA population. This integrative study suggested a potential role of lipids, especially certain glycerophospholipids and TAGs, in the etiology of colorectal cancer.

Conclusions: This study identified potential novel risk biomarkers for colorectal cancer by integrating genetics and circulating metabolomics data.

Impact: The identified metabolites could be developed into new tools for risk assessment of colorectal cancer in both European and East Asian populations.

INTRODUCTION

Colorectal cancer remains a significant health burden in the United States (United States) and many other counties. More than 1.9 million new colorectal cancer cases and 935,000 colorectal cancer deaths occurred worldwide in 2020 (1,2). The incidence varies

significantly across regions (1). For example, the age standardized incidence of colorectal cancer is 36.9 per 100,000 in the United States and 25.3 per 100,000 in China, respectively (3-5). Obesity, cigarette smoking, heavy alcohol consumption, diets high in fat and red meat or processed meat, sedentary lifestyle, and history of adenomatous polyps are established or suspected risk factors for colorectal cancer (6). Genetic factors also play an important role in colorectal carcinogenesis. Genome-wide association studies (GWAS) have uncovered over 100 genetic susceptibility loci of colorectal cancer in European and East Asian populations (7-16). However, the biological mechanisms underlying these associations for most of the identified loci and whether it has a differential impact on colorectal cancer development in different racial groups remain elusive, indicating the need for further investigations.

The advance of omics techniques has enabled a comprehensive and efficient examination of intermediate phenotypic markers such as circulating metabolites within population-based studies (17-19), casting novel insights into cancer etiology and biology. Nevertheless, limitations of traditional observational studies including relatively small sample size, residual confounding, and evident heterogeneity due to differences in research design, study population, 'omics' measurement platform, and statistical analysis, pose challenges for making causal inference.

In recent years, new methods integrating multi-omics data have been developed and applied to uncover novel etiologic factors for cancer. One such method, transcriptome-wide association study (TWAS) (20,21), has been widely implemented to identify novel susceptibility genes for different cancers (22-26). By combining genetic information with transcriptomics data, TWAS assesses the relationship between genetically predicted gene expressions, versus measured expression levels, and cancer risk. Because the approach takes advantage of random assignment of parental genotypes within each locus that occurs at meiosis (27), theoretically, TWAS minimizes the impact of reverse causation and confounding, compared with traditional observational studies. Extending the use of TWAS to data of other omics, such as metabolomics, is promising and important to address the gaps mentioned above. The approach is highly cost efficient at the screening stage for a biomarker study, likely to resulting in promising and high-quality candidates for follow-up investigations.

Here, we extended the application of genetic prediction algorithms to existing metabolomics data, to search for novel risk biomarkers, and facilitate a better understanding of colorectal cancer etiology in two racial groups.

MATERIALS & METHODS

The study flowchart was shown in Supplementary Figure 1.

Data set used for model building

Framingham Heart Study (FHS) Offspring Cohort: The FHS Offspring Study was a longitudinal community-based cohort study, which was initiated in 1971 after the establishment of the original FHS cohort (28,29). A total of 2,079 participants of European descent, who underwent metabolic profiling and genome-wide genotyping, were eligible to

be included in the genetic prediction model building process. We further excluded related participants according to their genomic relatedness (> 0.05) using Plink1.9 (30), which resulted in 1,357 unrelated subjects remaining in the current study. As described previously, blood samples were collected from the participants after an overnight fast (31-33). Genome-wide genotyping was conducted using the Affymetrix 500K mapping array and the Affymetrix 50K gene-focused MIP array (31). The called genotypes were then imputed to the 1000 Genome phase 3 reference panel. After quality control (QC) procedures, only genetic variants with minor allele frequency (MAF) > 0.05 and imputation $R^2 > 0.8$, were kept for prediction model building. Ten to 30 μl of plasma from the same set of participants were used to profile circulating metabolites using three different approaches. The details of the procedures to profile circulating metabolites in plasma samples were described in previous literature (31-33). Two hundred and seventeen metabolites (113 polar analytes and 104 lipid analytes) were measured by the LC/MS-based metabolomics platform. Amino acids, amino acid derivatives, urea cycle intermediates, nucleotides, and other positively charged polar metabolites were profiled using a 4000 QTRAP triple quadrupole mass spectrometer that was coupled to a multiplexed LC system with hydrophillic interaction chromatography columns installed. Organic acids, sugars, bile acids, and other negatively charged polar metabolites were profiled using a 5500 QTRAP triple quadrupole mass spectrometer using electrospray ionization (ESI) and multiple reaction monitoring (MRM) in the negative ion mode. For both approaches, isotope standards were added to the samples for generating calibration curves and absolute quantification of metabolites of interest. Plasma lipid profiles were obtained using a 4000 QTRAP Triple Quadrupole Mass Spectrometer, coupled to a reverse-phase chromatography with Prosphere HP C4 columns installed. No isotope-labeled standards were used to determine absolute levels of profiled lipids. All the FHS data are accessible *via* dbGAP (https://www.ncbi.nlm.nih.gov/gap/, study accession: phs000007.v31.p12).

Prediction model building

To address the right-skewed distributions of metabolite levels and differences in scaling, metabolites were log-transformed, then regressed against age and sex to obtain residuals. The residuals were then quantile normalized and standardized (mean of 0 and standard deviation of 1) in the overall study population. We randomly split the unrelated 1,357 participants of FHS Offspring study into training (N = 1,000) and testing set (N = 357), with a rough ratio of 3:1. We specifically aimed to build prediction models for 63 metabolites, selected from the 217 metabolites (Supplementary Table 1), as strong metabolite-quantitative trait loci (QTL); associations were previously reported for these metabolites (31).

Genetic variants passing the QC and located within the proximity of 500 kb both upstream and downstream to each reported metabolite-QTL variant, were subject to a variable selection procedure by the elastic net method (R package glmnet, α = 0.5). Because genetic variants in high linkage disequilibrium (LD) contain redundant information, we performed pairwise pruning (LD $r^2 > 0.9$), prior to implementing the elastic net procedure. We implemented a 5-fold cross-validation in the training set to address the potential issue of overfitting. A tuning parameter of regularization (λ) for a model with the best performance

was determined by minimizing the mean cross-validated error during the cross-validation procedure. The regularized βs of genetic predictors were extracted and applied to the samples in the testing set. Pearson correlation r was calculated between the genetic predicted levels of metabolites ($\beta_i G_i$) and their observed levels in the testing set. In a sensitivity analysis, we observed minimal variation for model performance when changing the fold of cross-validation or the pruning criterion (Supplementary Figure 2). Levels of metabolites predicted well in the training set were also correlated well with the corresponding measured levels in the testing set (Supplementary Figure 2). We combined training and testing sets and repeated the same abovementioned procedures to refine β estimates of genetic predictors for each metabolite. In this procedure, 58 metabolites had models with $R^2 > 0.01$ (correlation coefficient between predicted and measured levels > 0.1 in the cross-validation) (13,22,24) and were considered for downstream analyses with colorectal cancer risk.

Colorectal cancer GWAS consortia

Individual-level genotype data for the selected genetic predictors for metabolites were extracted from several large-scale consortia (Supplementary Table 2).

European data: This study includes GWAS data from the COloREctal Cancer Translational Study (CORECT), the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), the Colorectal Cancer Family Registry (CCFR), and UK Biobank. Detailed descriptions regarding genotype datasets, sample selection, and studies have been published previously (7,9-11). The details of genotype quality control procedures to filter out samples were described in previous publications (7,9-11). Briefly, individuals that were second-degree or more closely related, were excluded based on identity by descent estimates for each pair of samples. Samples with discrepancies between reported and genotypic sex based on X chromosome heterozygosity and the mean values of sex chromosome probe intensities were also excluded. Variants having missing call rate >2%, discordant calls in sample duplicates, and departing from Hardy-Weinberg equilibrium (HWE) (P $< 1 \times 10^{-4}$) based on European-ancestry controls were removed. All GWAS data were imputed to the Haplotype Reference Consortium (HRC) panel (34) using the University of Michigan Imputation Server (35). The current study was restricted to individuals of European descent and invasive cancer cases, leaving 46,323 colorectal cancer and 59,288 controls for downstream analyses. Approximately 10% of the cases were diagnosed at an age younger than 50 years old and 61.1% were diagnosed with colon cancer (Supplementary Table 2). All participants provided written informed consent, and each study was approved by the relevant Institution Review Board (IRB) or research ethics committee.

Asia Colorectal Cancer Consortium: The current study utilized genotyping data from 21,731 colorectal cancer cases and 47,442 controls of East Asian ancestry from studies conducted in the Asia Colorectal Cancer Consortium (ACCC), and some were also included in the CORECT study (Supplementary Table 2). Details of sample selection and matching, genotyping, genotype calling, and QC have been described previously (8,13-16). Briefly, the samples were genotyped using a variety of Illumina assays. Samples or SNPs were excluded if they met any of the following criteria: (i) genotype call rate per sample < 95%, (ii) genetically identical or duplicate samples (i.e., PI_HAT > 0.9), (iii) sex determined

using genotypes inconsistent with epidemiologic or clinical data, (iv) first- or second-degree relatives (i.e., PI_HAT > 0.25), (v) ethnic outliers with a population structure inconsistent with HapMap Asian samples, (vi) genotype call rate per SNP < 95%, (vii) MAF < 1%, (viii) genotyping consistency rates < 95% in quality control samples, (ix) P for HWE $<1\times10^{-5}$ in controls, or (x) SNPs not in autosomes. The genotyping data were imputed using 1000 Genome phase 3 mixed reference haplotypes *via* the Michigan Imputation Server (SHAPIT2 for haplotype phasing and minimac3 for imputation) (35). Nearly 22% of the participants were diagnosed with colorectal cancer at younger than 50 years old. All participants provided written informed consent, and each study was approved by the relevant research ethics committee or IRB.

Statistical analysis

Genetically predicted levels of metabolites ($N_{metabolite} = 58$) were calculated as a genetic score (GS) using the following formula:

$$GS_j = \sum_{k=1}^{n} [\beta_i * x_{ij}]$$

where the β_i is the per-allele log odds ratio (OR) of the variants i from the built model for the corresponding metabolite. The x_{ij} is the allele dosage for variant i of individual j, and n is the total number of variants included in the GS calculation. $\beta_i G_i$ then were modeled as the exposure of interest in the logistic regression models to obtain ORs, 95% confidence interval (CI), and p-values for the association with colorectal cancer risk. Covariates adjusted in the multivariable models included age, sex, top principal components (to adjust for potential population structure), genotyping platform, and substudy, when appropriate. Regression analysis was performed separately for the European-ancestry sample sets, and each substudy in ACCC. The estimates were then combined by meta-analysis within each racial group (European and EA) and across the two groups. Stratified analyses were also conducted by sex (male and female), age at diagnosis (< 50 years and 50 years), and cancer site (colon and rectum, available in European data only). Principal component analysis (PCA) and pairwise partial correlation were performed to show the correlations of measured metabolite levels in the FHS data (Supplementary Figure 3). As FHS data contain a relatively limited number of metabolites used as candidates for our model building strategy, we alternatively performed instrumental analysis using summary statistics of genetic variants reported as metabolite-quantitative trait loci (metaboliteQTL) in a recently published study (36). The study reported 499 associations ($P < 4.9 \times 10^{-10}$) across 142 unique metabolites. We employed inverse-variance approach to evaluate the associations between the 142 metabolites and colorectal cancer risk using data from the two consortia. All statistical analyses were conducted using R 3.4.1 or Stata version 11.

RESULTS

Model building

A total of 58 metabolites passed the predefined criterion at cross-validation $R^2 > 0.01$ in the model building process when up to 1,357 unrelated samples were analyzed (Supplementary

Table 3). The number of genetic variants selected as predictors varied from 1 for C54:2 triacylglycerols (TAG) to 67 for β-aminoisobutyric acid with a median of 9. On average, the correlation coefficient between predicted and measured metabolites in the overall study population was 0.155 (or 0.024 if presented as prediction R^2, Supplementary Table 3). Among the 58 metabolites that passed the model accuracy criterion, 41 of them broadly belong to lipids, including glycerophosphocholines (n = 11), glycosphingolipids (n = 4), glycerophosphoethanolamines (n = 3), phosphosphingolipids (n=2), triradylcglycerols (TAG, n = 18), and diradylglycerols (n = 3).

Association findings in Europeans

Genetically predicted levels of 24 metabolites showed a significant association with colorectal cancer risk in individuals of European descent (Benjamini-Hochberg False Discovery Rate [BH-FDR] < 0.05, Table 1). With a few exceptions, i.e., lactate, alanine, α-hydroxybutyrate, and cholesteryl esters, most of the metabolites were glycerophospholipids and their derivatives (n = 13) or TAGs (n = 6). Half of the metabolites (12/24) were positively associated with colorectal cancer risk. The most significant association was observed for C38:4 phosphatidylcholine (PC) (OR = 1.02, 95% CI = 1.01-1.03, P = 6.4×10^{-8}) after adjustment of age, sex, study, and top principal components. Four chromosomal loci, i.e., chr2p23.3 (*GCKR*), chr11q12.2 (*FADS1-3*), chr7p11.2 (*SEC61G*) and chr12p12.1 (*SLCO1B1*), were driving the identified significant associations and may influence colorectal cancer risk through regulating metabolite levels in blood (Table 1). Genetic loci influencing other metabolites lacking a significant association with colorectal cancer risk were also presented (Supplementary Table 4).

Association findings in EA

We replicated 21 of the 24 associations in individuals of EA using ACCC data as all of them remained significant after correction for multiple comparisons (BH-FDR < 0.05, Table 1). One additional association was found in EA that was not observed among Europeans (C16:0 cholesteryl ester [CE], OR $_{EA}$: 1.94, 95% CI, 1.60-2.36, P = 2.60×10^{-11}; OR $_{EUR}$: 1.01, 95% CI, 0.99-1.04, P = 0.30). In addition to the known GWAS loci that regulate metabolites mentioned above, variants in chr18p11.21 (*GNAL*) contributed to the variability of circulating metabolite levels, particularly to the levels of C16:0 CE. Although the size of EA studies included in the current analysis was apparently smaller than that of European studies (Supplementary Table 2), the effect size of many identified associations was markedly greater in EA populations (Table 1 & Figure 1) and the P-values were also lower in the same populations. We further compared the original GWAS estimates for the genetic variants involved in the current study between the two populations (Figure 2 & Supplementary Table 5). The effect sizes were systematically larger for the selected variants in EA populations than that in European populations.

Findings of meta-analysis by combining European and EA data

When meta-analyzing the race-specific association estimates, genetically predicted levels of 24 metabolites were significantly associated with colorectal cancer risk after accounting for multiple comparisons (BH-FDR < 0.05). Strong heterogeneity (P $_{het}$ < 0.005, Table 1) was found for 18 associations including C16:0 CE between the two populations.

Clusters of the identified metabolites

For the identified metabolites, PCA analysis showed distinct clusters based on their measured metabolite levels in the FHS dataset; for example, a group of TAG (i.e., C48:2 TAG, C48:3 TAG, C50:3 TAG and C50:4 TAG) was distinctively separated from cholesteryl esters (Supplementary Figure 3).

Stratified analysis

In the European populations, we observed similar associations across the site of primary tumor (colon and rectum) in the stratified analysis for all the identified risk-associated metabolites (Supplementary Table 6), although the significance of the associations was attenuated due to reduced sample size (Supplementary Table 2). We also evaluated the identified associations by sex (male and female) and age at diagnosis (< 50 and 50 years) (Supplementary Figure 4). All observed associations were consistently associated with colorectal cancer risk in women and men (Table 2). None of the identified associations were significantly associated with risk of young-onset colorectal cancer with small effect sizes (Table 2); however, all these associations were consistent in direction for the two strata. Tests for heterogeneity indicated that identified associations were mainly driven by colorectal cancer cases with an age at diagnosis of 50 years or older, which accounted for approximately 90% of the participants. In contrast, we did not find strong heterogeneity by sex or age at disease diagnosis in the EA population. The effect sizes were comparable or even larger among patients diagnosed at younger than 50 years old in this population (Table 3).

Results from genetic instrumental analysis based on summary statistics

We further conducted additional analysis by employing instrumental analysis using summary statistics from a recently published study (36). Sixty three of the 142 metabolites reported by the original study were significantly associated with colorectal cancer risk in both European and EA populations (Supplementary Table S7; P_{EUR} and $P_{EA} <$ 0.005). All of them belong to lipids particularly fall into subgroups of glycerophospholipids (PCs, lysoPCs, and sphingomyelins) In addition, kynurenine and PC ae C32:2 were found significantly associated with colorectal cancer risk in European data only ($P_{EUR} < 0.005$), while hexadecanoylcarnitine, lysoPC a C28:0, octadecenoylcarnitine, PC aa C36:3, SM C18:0 were significant in EA data only ($P_{EA} < 0.005$).

DISCUSSION

In the current study, we found that genetically predicted levels of 24 metabolites were associated with colorectal cancer risk, after accounting for multiple comparisons (BH-FDR < 0.05) in the populations of European descent, while 21 of them were also replicated in East Asians with the same criteria.

Compelling evidence has shown that many circulating metabolites can be regulated by germline genetic variants (37,38). For example, previous GWAS identified 145 genetic loci associated with approximately 300 metabolites, which covered amino acids, sterols, carnitines and intermediates of metabolisms of inositol, fatty acids, glucose, and nucleosides

in human blood (38). Another recent study reported 588 associations (mainly for lipids) involving a total of 54 independent regions (39). In these studies, heritability of metabolites explained by reported genetic loci varied from an average of 6.9% to over 20%, which serves as a strong foundation for our approach to predict metabolite levels using genetic variants. With an unprecedentedly large sample size, we hence evaluated the associations between genetically predicted metabolites and colorectal cancer risk in both individuals of European and EA descent, particularly focusing on metabolites known to be influenced by genetic variants. By combining GWAS data from several large-scale colorectal cancer consortia, our analysis showed genetically predicted levels of 25 metabolites were significantly associated with colorectal cancer risk in individuals of European and/or EA descent, the majority of which were glycerophospholipids and TAGs.

To our knowledge, this is the first large investigation that evaluates the associations between metabolites and colorectal cancer risk *via* an integrative omics approach. Various methods based on genetic instruments, such as Mendelian Randomization and TWAS (20,40), have been recently developed and widely employed in epidemiologic studies to facilitate causal inference in disease etiology research. The success of these approaches is partly attributable to the rapidly growing publicly available GWAS data. Conceptually, our analysis is an extension of the TWAS approach, by building genetic prediction models for circulating metabolite levels, rather than gene expression levels.

Most metabolites significantly associated with colorectal cancer risk were glycerophospholipids and their downstream derivatives (i.e., lysophospholipids), and TAGs. Previous population-based metabolomics studies including our own, suggested significant associations between glycerophospholipids and colorectal cancer risk (19,41). By utilizing prediagnostic samples, these studies were less prone to reverse causation and other biases. However, since the sample size of these studies remains relatively small, definitive evidence for the observed associations is still lacking (19,41). The current study, on the other hand, has leveraged unprecedentedly large consortium data to evaluate the associations of circulating metabolites with colorectal cancer risk. Importantly, by adopting an integrative design similar to the TWAS approach, we improved the statistical power and minimized the possibility of reverse causation and selection bias, which are limitations often seen in traditional biomarker studies, enhancing the validity of our findings and resulting in promising candidates for follow-up investigations. Furthermore, including data from two populations of different ancestry (i.e., European and EA), has in turn improved generalizability of the study findings. Therefore, the current study could provide strong evidence that glycerophospholipids and TAGs play an important role in colorectal cancer development.

Multiple glycerophospholipids and TAGs associated with colorectal cancer risk were shared across European and EA populations in the present study. TAGs are main components of very-low-density lipoprotein and chylomicrons, which are a main energy source and depot for the human body. The relationship between TAG and colorectal cancer remains inconclusive as some studies reported that elevated total TAG level was associated with an increased colorectal cancer risk, while others found null associations (42,43). The inconsistent findings could be due to the differences in study design, populations, and

potential residual confounders (44). In addition, few studies have conducted a detailed investigation on individual TAG species, which we reported herein. This highlights the importance of our work, which emphasizes that total TAG level could not serve as a reliable biomarker for colorectal cancer risk and more investigations are warranted for its species.

Glycerophospholipids like PCs are essential for maintaining structural integrity of cell membranes. Lysophosphatidylcholines are derived from the partial hydrolysis of PCs. Previous metabolomics studies have linked PCs to risks of different cancers; overall, an inverse relationship between levels of PCs and cancer risk was reported in the literature (18,19,45,46). One explanation is that the anti-inflammatory property of PCs may play a critical role in lowering cancer risk (47). However, the altered levels of PCs in circulation may be merely a reflection of increased activities of PC-specific phospholipase C and other relevant enzymes in cancer cells (48). Given that most cancers have a long disease latency period, it is conceivable that many colorectal cancer patients remain asymptomatic and undiagnosed for years. This implies that performing a sensitivity analysis to remove patients who are diagnosed shortly after cohort enrollment is critical to minimize the impact of reverse causation. Our study has eliminated such concern since genetically determined phenotypes like the genetically predicated levels of metabolites are not modified by cancer status.

Two chromosomal loci, chr2p23.3 (*GCKR*) and chr11q12.2 (*FADS1-3*), are known GWAS regions exerting strong pleiotropic effects (8,49-53). *GCKR* encodes glucokinase regulator, a protein that inhibits glucokinase by binding non-covalently to form an inactive complex with the enzyme in liver and pancreatic islet cells. Genetic variants in this locus associate with a variety of proteins, metabolites, and other traits. For instance, an early GWAS found that the locus was associated with fasting blood insulin and glucose levels and the findings were successfully validated in other studies (51,54). The locus was also related to C-reactive protein levels (55), amino acids (56), and Crohn's disease (53). A prior study has also shown that genetic variants in chr2p23.3 may exert a similar effect across different racial groups on colorectal cancer risk (57). Chr11q12.2 is a known colorectal cancer susceptibility locus (8), initially identified in EAs, then replicated among European populations. The locus also harbors regulatory variants that altered expression of fatty acid desaturases (*FADS*). As suggested by the name, these genes are key players in unsaturation of fatty acids, converting monounsaturated fatty acids to polyunsaturated fatty acids. It has been reported consistently in prior studies that Chr11q12.2 is associated with a variety of lipids including glycerophospholipids and TAGs in addition to fatty acids (31,38,52,58,59). In the present study, we were not able to evaluate the relationship between unsaturated fatty acids and colorectal cancer risk directly since they were not covered by the metabolomic platform used in the parental FHS Offspring study. However, our study highlighted a potential role of glycerophospholipids and TAGs in colorectal tumorigenesis, providing new evidence that the underlying mechanism linking the susceptibility locus on chr11q12.2 to colorectal cancer development may be mediated through a dysregulated lipid profile.

We observed generally larger effect sizes for the identified associations in EA populations than in European populations. This may be explained by the fact that the effect sizes of individual genetic variants on colorectal cancer risk involved in the current study were

systematically larger in the EA populations compared to European populations. This is not unexpected because the original colorectal cancer susceptibility locus, Chr11q12.2-*FADS1-3*, was initially reported by GWAS conducted in an EA population (8).

Despite many strengths of our study such as large sample size and inclusion of two racial groups, we acknowledge several limitations. First, we lacked an external dataset composed of genetics and metabolite data from independent subjects, which would be ideal for validating performance of our models. Second, the study findings were only generalizable to individuals of European and EA descent but not to other racial/ethnic groups. Furthermore, although the overall sample size was large, the Asian population samples size was smaller than our European cohort. However, the differences in magnitude of associations of genetically predicted metabolites and colorectal cancer found between the two racial groups may not be explained by the disparity of sample size of the two populations. Another limitation is that the variability of the identified metabolites such as TAGs were influenced by dietary intakes, which was not accounted for in the current study. We also lacked data on obesity and type 2 diabetes which are relevant to metabolic alterations in human body and serve as known risk factors for colorectal cancer. Thus, our study was unable to illuminate the interrelationship between metabolites and lifestyle risk factors and their separate and joint impact on colorectal cancer development. Finally, only a small proportion of circulating metabolites were investigated in this study. A more comprehensive analysis will be feasible when GWAS data, coupled with broader coverage of metabolome for global metabolite profiling, become available. For example, several GWAS of circulating metabolome have been published in recent years and summary statistics are accessible via public databases (60,61). Further investigations by including a larger reference panel for model building would be critical next step. On the other hand, metabolites lacking strong genetic determinants cannot be evaluated using our approach.

In conclusion, *via* an integrative approach, our study identified multiple metabolites that may help us better understand etiology of colorectal cancer in individuals of European and EA descent. The current study provided strong evidence to support the important role of certain lipids, particularly glycerophospholipids and TAGs, in colorectal carcinogenesis. Actual measurement of the identified metabolites in the prediagnostic samples and further evaluation for their association with colorectal cancer risk are warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

REFERENCE

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 doi 10.3322/caac.21660.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. Ca-Cancer J Clin 2020;70(1):7–30 doi 10.3322/caac.21590. [PubMed: 31912902]

3. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin 2021;71(1):7–33 doi 10.3322/caac.21654. [PubMed: 33433946]

4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115–32 doi 10.3322/caac.21338. [PubMed: 26808342]

5. Zhang L, Cao F, Zhang G, Shi L, Chen S, Zhang Z, et al. Trends in and Predictions of Colorectal Cancer Incidence and Mortality in China From 1990 to 2025. Front Oncol 2019;9:98 doi 10.3389/fonc.2019.00098. [PubMed: 30847304]

6. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg 2009;22(4):191–7 doi 10.1055/s-0029-1242458. [PubMed: 21037809]

7. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology 2013;144(4):799–807 e24 doi 10.1053/j.gastro.2012.12.020. [PubMed: 23266556]

8. Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet 2014;46(6):533–42 doi 10.1038/ng.2985. [PubMed: 24836286]

9. Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. Nat Commun 2015;6:7138 doi 10.1038/ncomms8138. [PubMed: 26151821]

10. Schmit SL, Edlund CK, Schumacher FR, Gong J, Harrison TA, Huyghe JR, et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst 2019;111(2):146–57 doi 10.1093/jnci/djy099. [PubMed: 29917119]

11. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet 2019;51(1):76–87 doi 10.1038/s41588-018-0286-6. [PubMed: 30510241]

12. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun 2019;10(1):2154 doi 10.1038/s41467-019-09775-w. [PubMed: 31089142]

13. Lu Y, Kweon SS, Tanikawa C, Jia WH, Xiang YB, Cai Q, et al. Large-Scale Genome-Wide Association Study of East Asians Identifies Loci Associated With Risk for Colorectal Cancer. Gastroenterology 2019;156(5):1455–66 doi 10.1053/j.gastro.2018.11.066. [PubMed: 30529582]

14. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet 2013;45(2):191–6 doi 10.1038/ng.2505. [PubMed: 23263487]

15. Lu Y, Kweon SS, Cai Q, Tanikawa C, Shu XO, Jia WH, et al. Identification of Novel Loci and New Risk Variant in Known Loci for Colorectal Cancer Risk in East Asians. Cancer Epidemiol Biomarkers Prev 2020;29(2):477–86 doi 10.1158/1055-9965.EPI-19-0755. [PubMed: 31826910]

16. Zeng C, Matsuda K, Jia WH, Chang J, Kweon SS, Xiang YB, et al. Identification of Susceptibility Loci and Genes for Colorectal Cancer Risk. Gastroenterology 2016;150(7):1633–45 doi 10.1053/j.gastro.2016.02.076. [PubMed: 26965516]

17. Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. Nat Med 2014;20(10):1193–8 doi 10.1038/nm.3686. [PubMed: 25261994]

18. Shu X, Zheng W, Yu D, Li HL, Lan Q, Yang G, et al. Prospective metabolomics study identifies potential novel blood metabolites associated with pancreatic cancer risk. Int J Cancer 2018;143(9):2161–7 doi 10.1002/ijc.31574. [PubMed: 29717485]

19. Shu X, Xiang YB, Rothman N, Yu D, Li HL, Yang G, et al. Prospective study of blood metabolites associated with colorectal cancer risk. Int J Cancer 2018;143(3):527–34 doi 10.1002/ijc.31341. [PubMed: 29479691]

20. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47(9):1091–8 doi 10.1038/ng.3367. [PubMed: 26258848]

21. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48(3):245–52 doi 10.1038/ng.3506. [PubMed: 26854917]

22. Wu L, Shi W, Long J, Guo X, Michailidou K, Beesley J, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet 2018;50(7):968–78 doi 10.1038/s41588-018-0132-x. [PubMed: 29915430]

23. Gusev A, Lawrenson K, Lin X, Lyra PC, Jr., Kar S, Vavra KC, et al. A transcriptome-wide association study of high-grade serous epithelial ovarian cancer identifies new susceptibility genes and splice variants. Nat Genet 2019;51(5):815–23 doi 10.1038/s41588-019-0395-x. [PubMed: 31043753]

24. Lu Y, Beeghly-Fadiel A, Wu L, Guo X, Li B, Schildkraut JM, et al. A Transcriptome-Wide Association Study Among 97,898 Women to Identify Candidate Susceptibility Genes for Epithelial Ovarian Cancer Risk. Cancer Res 2018;78(18):5419–30 doi 10.1158/0008-5472.CAN-18-0951. [PubMed: 30054336]

25. Mancuso N, Gayther S, Gusev A, Zheng W, Penney KL, Kote-Jarai Z, et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. Nat Commun 2018;9(1):4079 doi 10.1038/s41467-018-06302-1. [PubMed: 30287866]

26. Wu L, Wang J, Cai Q, Cavazos TB, Emami NC, Long J, et al. Identification of Novel Susceptibility Loci and Genes for Prostate Cancer Risk: A Transcriptome-Wide Association Study in Over 140,000 European Descendants. Cancer Res 2019;79(13):3192–204 doi 10.1158/0008-5472.CAN-18-3536. [PubMed: 31101764]

27. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27(8):1133–63 doi 10.1002/sim.3034. [PubMed: 17886233]

28. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol 2015;44(6):1800–13 doi 10.1093/ije/dyv337. [PubMed: 26705418]

29. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 1979;110(3):281–90 doi 10.1093/oxfordjournals.aje.a112813. [PubMed: 474565]

30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7 doi 10.1186/s13742-015-0047-8. [PubMed: 25722852]

31. Rhee EP, Ho JE, Chen MH, Shen D, Cheng S, Larson MG, et al. A genome-wide association study of the human metabolome in a community-based cohort. Cell Metab 2013;18(1):130–43 doi 10.1016/j.cmet.2013.06.013. [PubMed: 23823483]

32. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17(4):448–53 doi 10.1038/nm.2307. [PubMed: 21423183]

33. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121(4):1402–11 doi 10.1172/JCI44442. [PubMed: 21403394]

34. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48(10):1279–83 doi 10.1038/ng.3643. [PubMed: 27548312]

35. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48(10):1284–7 doi 10.1038/ng.3656. [PubMed: 27571263]

36. Lotta LA, Pietzner M, Stewart ID, Wittemans LBL, Li C, Bonelli R, et al. A cross-platform approach identifies genetic regulators of human metabolism and health. Nat Genet 2021;53:54–64. [PubMed: 33414548]

37. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature 2011;477(7362):54–60 doi 10.1038/nature10354. [PubMed: 21886157]

38. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nat Genet 2014;46(6):543–50 doi 10.1038/ng.2982. [PubMed: 24816252]

39. Gallois A, Mefford J, Ko A, Vaysse A, Julienne H, Ala-Korpela M, et al. A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. Nat Commun 2019;10(1):4788 doi 10.1038/s41467-019-12703-7. [PubMed: 31636271]

40. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 2017;26(5):2333–55 doi 10.1177/0962280215597579. [PubMed: 26282889]

41. Geijsen A, Brezina S, Keski-Rahkonen P, Baierl A, Bachleitner-Hofmann T, Bergmann MM, et al. Plasma metabolites associated with colorectal cancer: A discovery-replication strategy. Int J Cancer 2019;145(5):1221–31 doi 10.1002/ijc.32146. [PubMed: 30665271]

42. van Duijnhoven FJ, Bueno-De-Mesquita HB, Calligaro M, Jenab M, Pischon T, Jansen EH, et al. Blood lipid and lipoprotein concentrations and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition. Gut 2011;60(8):1094–102 doi 10.1136/gut.2010.225011. [PubMed: 21383385]

43. Borena W, Stocks T, Jonsson H, Strohmaier S, Nagel G, Bjorge T, et al. Serum triglycerides and cancer risk in the metabolic syndrome and cancer (Me-Can) collaborative study. Cancer Causes Control 2011;22(2):291–9 doi 10.1007/s10552-010-9697-0. [PubMed: 21140204]

44. Pakiet A, Kobiela J, Stepnowski P, Sledzinski T, Mika A. Changes in lipids composition and metabolism in colorectal cancer: a review. Lipids Health Dis 2019;18(1):29 doi 10.1186/s12944-019-0977-8. [PubMed: 30684960]

45. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med 2019;17(1):178 doi 10.1186/s12916-019-1408-4. [PubMed: 31547832]

46. Kuhn T, Floegel A, Sookthai D, Johnson T, Rolle-Kampczyk U, Otto W, et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. BMC Med 2016;14:13 doi 10.1186/s12916-016-0552-3. [PubMed: 26817443]

47. Treede I, Braun A, Sparla R, Kuhnel M, Giese T, Turner JR, et al. Anti-inflammatory effects of phosphatidylcholine. J Biol Chem 2007;282(37):27155–64 doi 10.1074/jbc.M704408200. [PubMed: 17636253]

48. Iorio E, Ricci A, Bagnoli M, Pisanu ME, Castellano G, Di Vito M, et al. Activation of phosphatidylcholine cycle enzymes in human epithelial ovarian cancer cells. Cancer Res 2010;70(5):2126–35 doi 10.1158/0008-5472.CAN-09-3833. [PubMed: 20179205]

49. Tanaka T, Shen J, Abecasis GR, Kisialiou A, Ordovas JM, Guralnik JM, et al. Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. PLoS Genet 2009;5(1):e1000338 doi 10.1371/journal.pgen.1000338. [PubMed: 19148276]

50. Chen HY, Cairns BJ, Small AM, Burr HA, Ambikkumar A, Martinsson A, et al. Association of FADS1/2 Locus Variants and Polyunsaturated Fatty Acids With Aortic Stenosis. JAMA Cardiol 2020;5(6):694–702 doi 10.1001/jamacardio.2020.0246. [PubMed: 32186652]

51. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42(2):105–16 doi 10.1038/ng.520. [PubMed: 20081858]

52. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 2009;41(1):56–65 doi 10.1038/ng.291. [PubMed: 19060906]

53. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet 2010;42(12):1118–25 doi 10.1038/ng.717. [PubMed: 21102463]

54. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012;44(6):659–69 doi 10.1038/ng.2274. [PubMed: 22581228]

55. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation 2011;123(7):731–8 doi 10.1161/CIRCULATIONAHA.110.948570. [PubMed: 21300955]

56. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet 2012;44(3):269–76 doi 10.1038/ng.1073. [PubMed: 22286219]

57. Ollberding NJ, Cheng I, Wilkens LR, Henderson BE, Pollak MN, Kolonel LN, et al. Genetic variants, prediagnostic circulating levels of insulin-like growth factors, insulin, and glucose and the risk of colorectal cancer: the Multiethnic Cohort study. Cancer Epidemiol Biomarkers Prev 2012;21(5):810–20 doi 10.1158/1055-9965.EPI-11-1105. [PubMed: 22354904]

58. Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;42(2):137–41 doi 10.1038/ng.507. [PubMed: 20037589]

59. Tabassum R, Ramo JT, Ripatti P, Koskela JT, Kurki M, Karjalainen J, et al. Genetic architecture of human plasma lipidome and its link to cardiovascular disease. Nat Commun 2019;10(1):4329 doi 10.1038/s41467-019-11954-8. [PubMed: 31551469]

60. Lotta LA, Pietzner M, Stewart ID, Wittemans LBL, Li C, Bonelli R, et al. A cross-platform approach identifies genetic regulators of human metabolism and health. Nat Genet 2021;53(1):54–64 doi 10.1038/s41588-020-00751-5. [PubMed: 33414548]

61. Feofanova EV, Yu B, Metcalf GA, Liu X, Muzny D, Below JE, et al. Sequence-Based Analysis of Lipid-Related Metabolites in a Multiethnic Study. Genetics 2018;209(2):607–16 doi 10.1534/genetics.118.300751. [PubMed: 29610217]
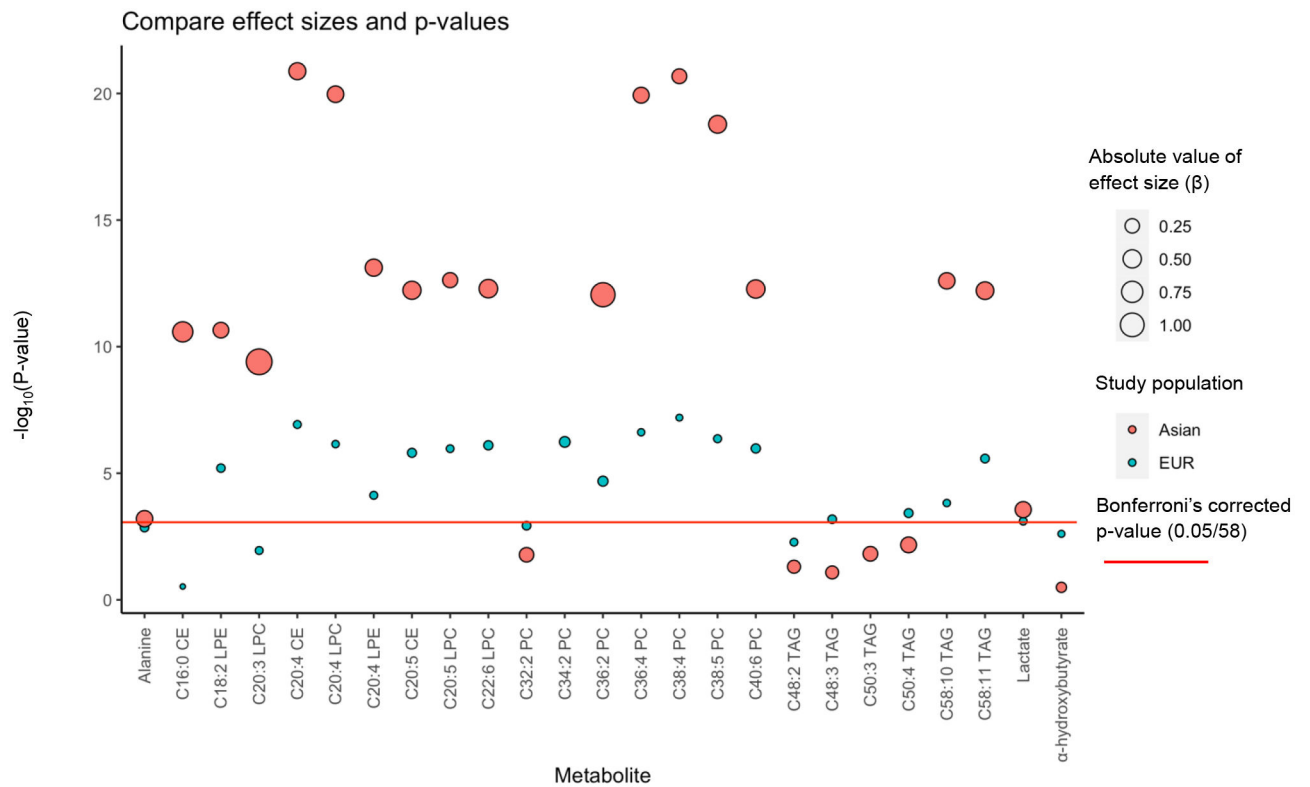
Figure 1. Comparisons of effect size for identified associations between European and East Asian populations

Comparisons of effect size for genetically predicted metabolites on colorectal cancer risk between European and Asian populations.

Difference of EUR and EA populations: Mean of $|\beta_{EUR}|$- $|\beta_{EA}|$: −0.390, $P_{\text{paired t-test}} = 9.91\text{x}10^{-8}$

Figure 2. Comparisons of effect size for individual genetic variants on colorectal cancer risk between European and East Asian populations

Comparisons of effect size for individual genetic variants used as genetic instruments on colorectal cancer risk between European and Asian populations.

Difference of EUR and EA populations: Mean of $|\beta_{EUR}| - |\beta_{EA}|$: -0.0135, $P_{\text{paired t-test}} = 2.2 \times 10^{-16}$
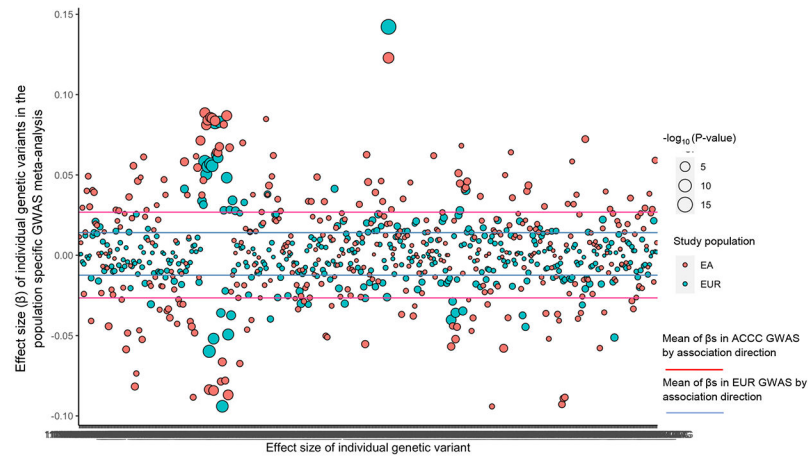
Table 1.

Associations between genetically predicted metabolite levels and colorectal cancer risk in individuals of European and EA descent, meta-analysis

Metabolite	OR (95%CI) $_{Eur}$	P $_{Eur}$	OR (95% CI) $_{EA}$	P $_{EA}$	OR (95% CI) $_{meta}$	P $_{meta}$	P $_{Eur-EA-Het}$	Chr	Locus
C38:4 PC [†‡¶]	1.02 (1.01-1.03)	6.38×10^{-8}	1.32 (1.24-1.39)	2.09×10^{-21}	1.02 (1.02-1.03)	5.55×10^{-11}	2.17×10^{-18}	11	*FADS1-3*
C20:4 CE [†‡¶]	1.03 (1.02-1.04)	1.19×10^{-7}	1.66 (1.45-1.91)	5.29×10^{-13}	1.04 (1.03-1.05)	5.79×10^{-11}	2.42×10^{-18}	7, 11	*SEC61G, FADS1-3*
C36:4 PC [†‡¶]	1.02 (1.01-1.03)	2.41×10^{-7}	1.41 (1.31-1.51)	1.17×10^{-20}	1.03 (1.02-1.04)	3.18×10^{-10}	7.95×10^{-18}	11	*FADS1-3*
C34:2 PC [‡]	0.91 (0.87-0.94)	5.79×10^{-7}	NA	NA	NA	NA	NA	11	*FADS1-3*
C38:5 PC [†‡¶]	1.03 (1.02-1.05)	4.33×10^{-7}	1.60 (1.44-1.77)	1.65×10^{-19}	1.04 (1.03-1.06)	7.06×10^{-10}	9.08×10^{-17}	11	*FADS1-3*
C20:4 LPC [†‡¶]	1.03 (1.02-1.04)	7.07×10^{-7}	1.46 (1.35-1.59)	1.07×10^{-20}	1.03 (1.02-1.04)	1.21×10^{-9}	5.58×10^{-18}	11	*FADS1-3*
C22:6 LPC [†‡¶]	1.06 (1.03-1.08)	7.89×10^{-7}	1.69 (1.47-1.95)	5.11×10^{-13}	1.07 (1.05-1.09)	1.93×10^{-9}	1.97×10^{-10}	11	*FADS1-3*
C40:6 PC [†‡¶]	1.06 (1.03-1.08)	1.05×10^{-6}	1.51 (1.39-1.64)	1.32×10^{-21}	1.07 (1.05-1.10)	2.09×10^{-9}	2.49×10^{-10}	11	*FADS1-3*
C20:5 LPC [†‡¶]	1.03 (1.02-1.04)	1.07×10^{-6}	1.33 (1.24-1.44)	2.36×10^{-13}	1.04 (1.02-1.05)	3.05×10^{-9}	7.72×10^{-11}	11	*FADS1-3*
C20:5 CE [†‡¶]	1.05 (1.03-1.08)	1.56×10^{-6}	1.64 (1.43-1.87)	5.93×10^{-13}	1.07 (1.04-1.09)	4.22×10^{-9}	2.05×10^{-10}	11	*FADS1-3*
C58:11 TAG [†‡¶]	1.05 (1.03-1.07)	2.63×10^{-6}	1.58 (1.39-1.78)	6.14×10^{-13}	1.06 (1.04-1.08)	9.77×10^{-9}	1.53×10^{-10}	11	*FADS1-3*
C18:2 LPE [†‡¶]	0.96 (0.94-0.98)	6.27×10^{-6}	0.72 (0.66-0.80)	2.23×10^{-11}	0.95 (0.94-0.97)	1.56×10^{-8}	8.38×10^{-9}	11	*FADS1-3*
C36:2 PC [†‡¶]	0.93 (0.89-0.96)	2.06×10^{-5}	0.35 (0.26-0.47)	8.89×10^{-13}	0.91 (0.88-0.94)	3.21×10^{-7}	5.23×10^{-11}	11	*FADS1-3*
C20:4 LPE [†‡¶]	1.03 (1.02-1.05)	7.43×10^{-5}	1.53 (1.37-1.71)	7.58×10^{-14}	1.04 (1.02-1.06)	6.28×10^{-7}	7.92×10^{-12}	11, 12	*FADS1-3, SLCO1B1*
C58:10 TAG [†‡¶]	1.03 (1.01-1.04)	1.50×10^{-4}	1.44 (1.31-1.59)	2.52×10^{-13}	1.03 (1.02-1.04)	2.49×10^{-6}	1.35×10^{-11}	11	*FADS1-3*
C50:4 TAG [†‡¶]	0.95 (0.93-0.98)	3.75×10^{-4}	0.72 (0.56-0.91)	0.007	0.95 (0.92-0.97)	1.23×10^{-4}	0.022	2	*GCKR*
lactate [†‡¶]	0.97 (0.95-0.99)	7.83×10^{-4}	0.70 (0.58-0.85)	2.74×10^{-4}	0.97 (0.95-0.99)	2.38×10^{-4}	9.02×10^{-4}	2	*GCKR*
alanine [†‡¶]	0.96 (0.93-0.98)	1.18×10^{-3}	0.69 (0.56-0.85)	6.25×10^{-4}	0.95 (0.93-0.98)	3.19×10^{-4}	0.003	2	*GCKR*
C48:3 TAG [†‡]	0.96 (0.93-0.98)	6.56×10^{-4}	0.83 (0.68-1.02)	0.083	0.96 (0.93-0.98)	3.28×10^{-4}	0.190	2	*GCKR*
C32:2 PC [†‡¶]	0.96 (0.93-0.98)	1.40×10^{-3}	0.77 (0.63-0.95)	0.017	0.95 (0.93-0.98)	4.21×10^{-4}	0.050	2	*GCKR*
α-hydroxybutyrate [†‡]	0.98 (0.97-0.99)	2.48×10^{-3}	0.92 (0.78-1.08)	0.321	0.98 (0.96-0.99)	0.002	0.463	2	*GCKR*
C48:2 TAG [†‡]	0.97 (0.95-0.99)	5.29×10^{-3}	0.83 (0.69-1.00)	0.049	0.97 (0.95-0.99)	0.003	0.103	2	*GCKR*

Metabolite	OR (95%CI) _Eur_	P _Eur_	OR (95% CI) _EA_	P _EA_	OR (95% CI) _meta_	P _meta_	P _Eur-EA-Het_	Chr	Locus
C20:3 LPC _†‡¶_	0.97 (0.95-0.99)	0.0113	0.29 (0.20-0.43)	3.96×10^{-10}	0.96 (0.94-0.99)	0.004	1.16×10^{-9}	11	_FADS1-3_
C50:3 TAG _†‡¶_	0.97 (0.95-1.00)	0.0201	0.76 (0.62-0.95)	0.015	0.97 (0.95-0.99)	0.010	0.030	2	_GCKR_
C16:0 CE _¶_	1.01 (0.99-1.04)	0.298	1.94 (1.60-2.36)	2.60×10^{-11}	1.03 (1.00-1.05)	0.057	9.57×10^{-11}	11, 18	_FADS1-3, GNAL_

EA: East Asian; Eur: European.

_†_ Association passed the BH-FDR threshold based on meta-analysis of European and EA populations.

_‡_ Association passed the BH-FDR threshold based on European data only.

_¶_ Association passed the BH-FDR threshold based on EA data only.

NA: not applicable due to extremely unstable estimates.

Table 2.

Stratified analysis by sex and age in individuals of European descent

Metabolite	Female		Male			Age at diagnosis < 50 yrs		Age at diagnosis 50 yrs		
	OR (95CI)	P	OR (95%CI)	P	P_{het}	OR (95CI)	P	OR (95CI)	P	P_{het}
C38:4 PC	1.02 (1.01-1.03)	4.51×10^{-4}	1.02 (1.01-1.03)	2.62×10^{-5}	0.757	1.002 (0.998-1.005)	0.353	1.02 (1.01-1.03)	3.45×10^{-8}	3.45×10^{-6}
C20:4 CE	1.03 (1.01-1.05)	2.63×10^{-4}	1.03 (1.02-1.05)	8.15×10^{-5}	0.987	1.001 (0.996-1.006)	0.711	1.03 (1.02-1.05)	5.55×10^{-8}	1.61×10^{-6}
C36:4 PC	1.02 (1.01-1.04)	4.62×10^{-4}	1.02 (1.01-1.04)	1.10×10^{-4}	0.929	1.002 (0.997-1.006)	0.470	1.03 (1.02-1.03)	1.29×10^{-7}	6.32×10^{-6}
C34:2 PC	0.91 (0.86-0.97)	0.002	0.90 (0.86-0.95)	7.93×10^{-5}	0.714	0.992 (0.975-1.009)	0.354	0.91 (0.87-0.94)	3.19×10^{-7}	1.95×10^{-5}
C38:5 PC	1.03 (1.01-1.05)	7.72×10^{-4}	1.04 (1.02-1.05)	1.20×10^{-4}	0.861	1.002 (0.996-1.008)	0.560	1.04 (1.02-1.05)	2.09×10^{-7}	7.06×10^{-6}
C20:4 LPC	1.03 (1.01-1.04)	9.33×10^{-4}	1.03 (1.01-1.04)	1.55×10^{-4}	0.865	1.002 (0.998-1.007)	0.344	1.03 (1.02-1.04)	5.21×10^{-7}	2.89×10^{-5}
C22:6 LPC	1.05 (1.02-1.09)	0.002	1.07 (1.03-1.10)	5.53×10^{-5}	0.588	1.003 (0.993-1.014)	0.511	1.06 (1.04-1.08)	4.92×10^{-7}	1.62×10^{-5}
C40:6 PC	1.05 (1.02-1.09)	0.002	1.06 (1.03-1.10)	1.19×10^{-4}	0.734	1.002 (0.992-1.013)	0.670	1.06 (1.04-1.09)	4.20×10^{-7}	9.30×10^{-6}
C20:5 LPC	1.02 (1.01-1.04)	0.011	1.04 (1.02-1.05)	1.15×10^{-5}	0.257	1.004 (0.998-1.009)	0.187	1.03 (1.02-1.04)	1.95×10^{-6}	1.44×10^{-4}
C20:5 CE	1.05 (1.02-1.09)	9.55×10^{-4}	1.05 (1.02-1.09)	3.55×10^{-4}	0.976	1.003 (0.993-1.013)	0.571	1.06 (1.03-1.08)	7.88×10^{-6}	1.97×10^{-5}
C58:11 TAG	1.05 (1.02-1.08)	0.001	1.05 (1.02-1.08)	4.34×10^{-4}	0.947	1.004 (0.995-1.013)	0.387	1.05 (1.03-1.07)	1.70×10^{-6}	6.18×10^{-5}
C18:2 LPE	0.96 (0.94-0.99)	0.002	0.96 (0.94-0.98)	6.60×10^{-4}	0.942	1.002 (0.994-1.010)	0.634	0.96 (0.94-0.97)	1.11×10^{-6}	3.56×10^{-6}
C36:2 PC	0.93 (0.88-0.98)	0.004	0.92 (0.88-0.97)	0.001	0.927	1.001 (0.985-1.017)	0.935	0.92 (0.89-0.95)	5.92×10^{-6}	3.16×10^{-5}
C20:4 LPE	1.04 (1.01-1.06)	0.002	1.03 (1.01-1.05)	0.011	0.581	1.003 (0.995-1.010)	0.463	1.03 (1.02-1.05)	7.13×10^{-5}	9.24×10^{-4}
C58:10 TAG	1.02 (1.01-1.04)	0.011	1.03 (1.01-1.04)	0.004	0.920	1.003 (0.998-1.009)	0.246	1.02 (1.01-1.04)	3.29×10^{-4}	0.005
C50:4 TAG	0.96 (0.92-1.00)	0.054	0.94 (0.91-0.98)	0.002	0.491	0.986 (0.974-0.998)	0.027	0.96 (0.93-0.98)	0.002	0.058
lactate	0.97 (0.95-1.00)	0.024	0.97 (0.95-1.00)	0.024	0.733	0.993 (0.986-1.001)	0.100	0.97 (0.96-0.99)	0.003	0.037
alanine	0.96 (0.93-1.00)	0.055	0.95 (0.92-0.99)	0.008	0.705	0.988 (0.976-1.000)	0.052	0.96 (0.94-0.99)	0.007	0.095
C48:3 TAG	0.97 (0.93-1.00)	0.082	0.95 (0.92-0.98)	0.002	0.403	0.988 (0.977-0.999)	0.036	0.96 (0.94-0.99)	0.003	0.073
C32:2 PC	0.96 (0.92-1.00)	0.065	0.95 (0.91-0.98)	0.005	0.591	0.988 (0.975-1.000)	0.052	0.96 (0.94-0.99)	0.004	0.074
α-hydroxybutyrate	0.98 (0.96-1.00)	0.095	0.97 (0.96-0.99)	0.008	0.561	0.996 (0.990-1.002)	0.198	0.98 (0.97-0.99)	0.006	0.047
C48:2 TAG	0.98 (0.95-1.01)	0.194	0.96 (0.93-0.99)	0.007	0.387	0.991 (0.981-1.001)	0.064	0.98 (0.95-1.00)	0.019	0.171
C20:3 LPC	0.95 (0.92-0.99)	0.011	0.98 (0.95-1.01)	0.273	0.266	0.999 (0.988-1.010)	0.824	0.97 (0.95-0.99)	0.010	0.023
C50:3 TAG	0.98 (0.95-1.02)	0.298	0.96 (0.94-1.00)	0.023	0.434	0.992 (0.982-1.003)	0.148	0.98 (0.96-1.00)	0.052	0.243
C16:0 CE	1.02 (0.98-1.05)	0.424	1.01 (0.98-1.05)	0.471	0.927	0.997 (0.985-1.009)	0.607	1.02 (0.99-1.05)	0.128	0.110

Table 3.

Stratified analysis by sex and age in individuals of Asian descent

Metabolite	Female		Male		Age at diagnosis < 50 yrs		Age at diagnosis 50 yrs	
	OR (95CI)	P	OR (95%CI)	P	OR (95CI)	P	OR (95CI)	P
C38:4 PC	1.16 (1.07-1.26)	6.14×10^{-4}	1.15 (1.05-1.26)	0.002	1.17 (1.02-1.35)	0.022	1.15 (1.08-1.23)	4.92×10^{-5}
C20:4 CE	0.97 (0.71-1.33)	0.853	1.27 (0.92-1.75)	0.154	0.93 (0.56-1.54)	0.773	1.11 (0.86-1.42)	0.437
C36:4 PC	1.21 (1.09-1.35)	5.03×10^{-4}	1.19 (1.06-1.34)	0.003	1.27 (1.06-1.51)	0.009	1.19 (1.09-1.30)	1.00×10^{-4}
C38:5 PC	1.34 (1.14-1.57)	4.32×10^{-4}	1.31 (1.11-1.56)	0.002	1.40 (1.08-1.82)	0.011	1.31 (1.15-1.49)	4.19×10^{-5}
C20:4 LPC	1.24 (1.10-1.40)	5.95×10^{-4}	1.23 (1.08-1.40)	0.002	1.27 (1.04-1.55)	0.018	1.23 (1.12-1.36)	3.78×10^{-5}
C22:6 LPC	1.64 (1.23-2.19)	7.78×10^{-4}	1.63 (1.20-2.22)	0.002	1.78 (1.12-2.84)	0.015	1.62 (1.28-2.04)	5.21×10^{-5}
C40:6 PC	1.63 (1.23-2.15)	6.17×10^{-4}	1.61 (1.20-2.17)	0.002	1.74 (1.11-2.73)	0.016	1.61 (1.29-2.02)	2.87×10^{-5}
C20:5 LPC	1.26 (1.09-1.46)	0.002	1.31 (1.12-1.53)	7.76×10^{-4}	1.40 (1.10-1.77)	0.007	1.26 (1.12-1.42)	1.30×10^{-4}
C20:5 CE	1.63 (1.25-2.14)	3.74×10^{-4}	1.57 (1.18-2.09)	0.002	1.76 (1.13-2.72)	0.012	1.58 (1.27-1.96)	4.17×10^{-5}
C58:11 TAG	1.59 (1.23-2.04)	3.14×10^{-4}	1.53 (1.17-1.99)	0.002	1.71 (1.14-2.57)	0.009	1.53 (1.25-1.87)	3.88×10^{-5}
C18:2 LPE	0.75 (0.62-0.91)	0.004	0.73 (0.59-0.90)	0.003	0.71 (0.52-0.98)	0.035	0.74 (0.63-0.87)	1.72×10^{-4}
C36:2 PC	0.35 (0.20-0.62)	3.16×10^{-4}	0.38 (0.21-0.70)	0.002	0.31 (0.12-0.79)	0.014	0.37 (0.24-0.59)	3.07×10^{-5}
C20:4 LPE	1.39 (1.12-1.72)	0.003	1.44 (1.15-1.80)	0.002	1.45 (1.03-2.05)	0.032	1.40 (1.18-1.66)	1.09×10^{-4}
C58:10 TAG	1.32 (1.10-1.59)	0.003	1.35 (1.11-1.64)	0.002	1.42 (1.06-1.91)	0.020	1.33 (1.14-1.54)	1.76×10^{-4}
C50:4 TAG	0.83 (0.55-1.24)	0.358	0.72 (0.47-1.10)	0.130	0.74 (0.39-1.41)	0.357	0.76 (0.55-1.05)	0.101
Lactate	0.74 (0.53-1.03)	0.073	0.78 (0.55-1.11)	0.164	0.76 (0.45-1.28)	0.305	0.78 (0.60-1.02)	0.069
Alanine	0.72 (0.48-1.08)	0.108	0.73 (0.48-1.12)	0.149	0.59 (0.31-1.13)	0.113	0.77 (0.56-1.07)	0.119
C48:3 TAG	0.92 (0.63-1.34)	0.663	0.74 (0.50-1.10)	0.136	0.85 (0.46-1.54)	0.588	0.80 (0.59-1.08)	0.140
C32:2 PC	0.80 (0.50-1.27)	0.347	0.67 (0.41-1.08)	0.102	0.86 (0.41-1.80)	0.694	0.70 (0.48-1.02)	0.061
α-hydroxybutyrate	0.91 (0.65-1.26)	0.564	0.95 (0.68-1.34)	0.790	1.05 (0.64-1.75)	0.838	0.95 (0.73-1.24)	0.714
C48:2 TAG	0.92 (0.65-1.30)	0.623	0.75 (0.52-1.08)	0.122	0.86 (0.49-1.50)	0.597	0.82 (0.62-1.08)	0.151
C20:3 LPC	0.18 (0.09-0.39)	1.19×10^{-5}	0.42 (0.19-0.93)	0.031	0.26 (0.08-0.86)	0.028	0.30 (0.16-0.55)	9.57×10^{-5}
C50:3 TAG	0.79 (0.52-1.22)	0.291	0.73 (0.46-1.15)	0.172	0.85 (0.43-1.69)	0.652	0.75 (0.53-1.06)	0.102
C16:0 CE	1.67 (1.17-2.40)	0.005	1.81 (1.23-2.65)	0.003	1.75 (0.98-3.13)	0.057	1.70 (1.27-2.27)	3.50×10^{-4}