

Lipidomic markers of processed meat and unprocessed red meat intake and risk of diabetes in American Indians. Xiaoxiao Wen, et al.

Supplemental Methods and Results

Study population

The Strong Heart Family Study (SHFS, 2001-ongoing) is a family-based prospective study designed to identify genetic, metabolic, and behavioral factors for cardiometabolic diseases in American Indians. Briefly, a total of 2,786 tribal members (aged ≥ 14 years) residing in Arizona, North Dakota, South Dakota, and Oklahoma were examined in 2001-2003 (baseline) and re-examined in 2006-2009 (after about 5-year follow-up). Detailed descriptions of the SHFS methods have been published previously (1). All participants received a personal interview and a physical examination at each visit, during which fasting blood samples were collected for laboratory tests. Lipidomics data were collected at both visits for participants. Informed consent was obtained from all participants. The SHFS protocols were approved by the Institutional Review Boards (IRBs) of the participating institutions and tribes.

Inclusion and exclusion criteria

The flowchart of participant inclusion and exclusion is shown in Supplemental Figure S1. In the analysis of lipidomic markers for meat intake, we included 1,816 participants who: (a) were free of overt cardiovascular disease at baseline, and (b) had available dietary and lipidomics data at baseline. In the analyses of meat intake-related lipids with incident T2D by 5-year follow-up, we excluded participants with prevalent T2D or impaired fasting glucose at baseline ($n=740$), resulting in a total of 1,076 participants in the analysis. For the analysis on the longitudinal association between meat intake-related lipids and glucose/insulin metrics, we included all participants

(n=1,492) who had complete information for lipidomics, glucose and insulin measures at both visits.

Assessment of meat intake (red meat and processed meat)

An interviewer-administered Block 119-item Food Frequency Questionnaire (FFQ) was used to assess participants' usual food intake over the past year. Serving sizes were defined using standard units (e.g., one banana, two eggs) or standard volume/weight portions, with photographs provided as visual aids. Participants were asked to report the frequency of consumption for each food item, selecting from categories of seasonally, never, a few times per year, once per month, 2–3 times per month, once per week, twice per week, 2–3 times per week, 5–6 times per week, or daily. Portion sizes were classified as small, medium, or large.

In addition to the standard Block FFQ items, participants answered supplementary questions about the frequency and portion sizes of foods commonly consumed among American Indians, including menudo, pozole, guysava, red or green chili, Indian taco, fry bread, corn tortillas, flour tortillas, and "spam". "Spam" refers to various canned meats available in the United States, typically consisting of a mixture of beef or pork shoulder, salt, sodium nitrate, potato starch, and water.

Daily energy and macronutrient intakes were calculated using the Block database (Block Dietary Systems). To estimate average daily nutrient intake, the reported frequency of consumption for each food item on both the FFQ and the supplementary questionnaire was multiplied by the documented nutrient content of the specified portion size, with values summed across all reported foods. Based on these dietary data, we quantified habitual intake of processed meat (i.e., breakfast sausage, spam, hot dogs, lunch meat and bacon) and unprocessed red meat (i.e., pork chops, pork roast, veal, lamb, deer, ribs, hamburger, cheeseburger, roast beef, steak). We considered 100 g (3.5 oz) and 50 g (1.8 oz) as one serving of unprocessed red meat and processed meat, respectively.

Assessments of covariates

Sociodemographic information, lifestyle factors (cigarette smoking and alcohol use), medical history, and use of prescription medications (e.g., use of anti-hypertensive, anti-diabetic, and/or lipid-lowering drugs) were collected using structured questionnaires at both examinations. For cigarette smoking, participants were classified as current, former, or never smokers. Similarly, alcohol use was categorized as current, former, or never drinkers. To collect physical activity data, participants were asked to wear Accusplit AE120 pedometers for 7 consecutive days, except while bathing or swimming; the mean number of steps per day was calculated by averaging the total number of steps recorded each day during the 7-day period. Anthropometric measures including height and weight, as well as blood pressure measurements, were obtained during the physical examination. BMI was calculated as body weight in kilograms divided by the square of height in meters. Hypertension was defined as measured systolic blood pressure ≥ 140 mmHg, diastolic blood pressure ≥ 90 mmHg, or use of antihypertensive medications.

Blood sample collection

Participants were instructed to fast overnight before their visit, and fasting blood sample was collected into 10ml EDTA tubes at the Strong Heart Study field centers. The tubes were then gently inverted and placed on ice or refrigerated (-4°C) immediately. Plasma sample was obtained by centrifuging the tubes for 10 minutes at 3,000 rpm at -4°C and aliquots (0.5 ml) were immediately stored at -80°C until further analysis. For the current study, 0.5 ml fasting plasma sample, which were never thawed before, was shipped to Dr. Fiehn's lab at the West Coast Metabolomics Center (UC-Davis) on dry ice via FedEx overnight, and stored at -80°C immediately on arrival until further analyses. Samples were randomized before shipping to the Fiehn's laboratory, where randomization was performed again before the lipidomics analysis as described below. As part of

the QC procedures, 109 duplicated samples (55 at baseline and 54 at 5-year follow-up) were included to evaluate analytical or measurement precision. Lipid values of the duplicated samples were highly correlated (Spearman's correlation coefficient, $\rho = 0.95$, $P < 2.2 \times 10^{-16}$). Laboratory technicians were blinded to all clinical data throughout the assays.

Lipidomic data acquisition

Plasma samples were first extracted based on a modified liquid-liquid extraction method (cold methanol/ MTBE/water). The extracted samples were then subjected to lipidomics analysis by LC-MS in both positive and negative ionization modes.

Lipidomics ESI (+) on 6550 Agilent LC-QTOF MS. The injection volume on ESI (+) mode was 3 μ L. The mobile phase compositions were: A) Acetonitrile: water (60:40, v/v) with 10 mM ammonium formate and 0.1% formic acid; and B) Isopropanol: acetonitrile (90:10, v/v) with 10 mM ammonium formate and 0.1% formic acid. The LC gradient consisted of the following elution conditions: 0 min 15% (B); 0–2 min 30% (B); 2–2.5 min 48% (B); 2.5–11 min 82% (B); 11–11.5 min 99% (B); 11.5–12min 99% (B); 12–12.1 min 15% (B); and 12.1–15 min 15% (B).

Lipidomics ESI (-) on 6550 Agilent LC-QTOF MS. The injection volume on ESI (-) mode was 5 μ L. The mobile phase compositions were: A) Acetonitrile: water (60:40, v/v) with 10 mM ammonium acetate; and B) Isopropanol: acetonitrile (90:10, v/v) with 10 mM ammonium acetate. The LC gradient consisted of the following elution conditions: 0 min 15% (B); 0–2 min 30% (B); 2–2.5 min 48% (B); 2.5–11 min 82% (B); 11–11.5 min 99% (B); 11.5–12min 99% (B); 12–12.1 min 15% (B); and 12.1–15 min 15% (B).

The lipids were separated on Agilent 1290 Infinity LC system using an Acquity CSH C18 column (100 mm \times 2.1 mm, 1.7 μ m) with an Acquity CSH C18 guard column (5 mm \times 2.1 mm, 1.7 μ m) (Waters, Milford, MA). The column temperature was 65 $^{\circ}$ C with a flow rate of 0.6

mL/min. Sample temperature was maintained at 4 °C throughout the experiment. The guard column was changed every 300 samples, and a new column was replaced every 1,000 samples. We also measured a Bioreclamation plasma sample per 10 samples and a NIST plasma sample per 40 samples as quality controls for monitoring the instrumental drift.

Lipidomics data processing and normalization

The lipidomics data were pre-processed using a new in-house cloud-based software (LC- BinBase) with peak detection and deconvolution algorithms adapted from MS-DIAL (2). Raw files were automatically converted into correct formats and the LC-BinBase algorithms performed peak picking, retention time alignment using internal standards, and gap filling from raw data for missed peaks. Lipid peak intensity results were manually checked against raw data files. Adducts were combined into single features for statistical assessments. False negative features and peaks with 50% missing values across all samples were removed. The batch effect of reformatted dataset was normalized by SERRF software (Systematic Error Removal using Random Forest) (3), which dramatically reduced the raw data variance coefficient by 23% in positive mode data and 25% in negative mode to less than 10% in result files. The lipid levels of the duplicated samples were highly correlated (Spearman's correlation coefficient $\rho = 0.95$, $p < 2.2 \times 10^{-16}$). Outlier samples were detected by principal component analysis, and those beyond mean ± 5 SD for any of the first three PCs were further removed. After preprocessing and quality control, we obtained 1,542 lipids (518 known, 1,024 unknown) in 1,957 participants at baseline, and 1,948 participants at 5-year follow-up.

The lipid classes include: Acylcarnitine (AC), Cholesterol ester (CE), Ceramide (CER), Cholesterol, Diacylglycerol (DAG), Fatty acid (FA), Glycosylceramide (GlcCer),

Phosphatidylcholine (PC), Phosphatidylethanolamine (PE), Phosphatidylglycerol (PG), Phosphatidylinositol (PI), Phosphatidylserine (PS), Sphingomyelin (SM), Triacylglycerol (TAG).

Statistical analysis

Pathway enrichment analyses

Pathway enrichment analysis was conducted for the identified lipids at $P < 0.05$. Of the 89 lipids (known) associated with red meat consumption at $P < 0.05$, 49 were successfully matched to the Human Metabolome Database (HMDB) or the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. For the 58 lipids associated with processed meat consumption at $P < 0.05$, 40 were matched to HMDB or KEGG.

We performed metabolite set enrichment analysis using the MetaboAnalyst 6.0 platform (4), based on main chemical class metabolite sets or lipid sets. Enrichment ratio was calculated as (number of observed hits)/(number of expected hits). Additionally, potential pathway alterations were identified using Lipid Pathway Enrichment Analysis (LIPEA)(5), a web tool specially designed for lipidomic research based on the KEGG database. The statistical significance of each pathway was evaluated using the Fisher's exact test. A percentage of enrichment was calculated as (number of hits)/(number of lipids in the pathway).

Sensitivity analyses

We performed the following sensitivity analyses for the identification of lipidomic markers: (a) Sex-stratified analyses; (b) Identifying food groups that correlated with unprocessed red or processed meat intake, and additionally adjusting for these food groups in separate models as well as in single models; (c) Assessing the specificity of the identified lipids by examining their associations with food groups. For the associations of lipids with glucose/insulin metrics, we

performed sensitivity analyses by: (a) restricting to baseline data, and (b) additionally adjusting for glucose-lowering medication use.

Results for sensitivity analyses

In sex-stratified analyses, about 2/3 of significant lipids ($P < 0.05$) from main results remained significant among females. In males, only about 1/4 of the identified lipids remain significant, likely due to the smaller sample size ($n=677$) (**Supplemental Table S7**). We additionally identified food groups that correlated with unprocessed red or processed meat intake (**Supplemental Figure S5**). The top groups highly correlated with the unprocessed red meat were grains (Pearson's $r=0.52$), processed meat (0.44) and vegetables ($r=0.44$). For processed meat, the top correlated groups were grains (0.6), unprocessed red meat (0.44) and vegetables (0.4). Results remain largely consistent after further adjusting for these food groups separately or jointly (**Supplemental Tables S8-S12**). Furthermore, the identified lipids weakly correlated with other food groups rather than with red and/or processed meat, suggesting high specificity of the identified lipidomic markers (**Supplemental Figure S6**). Lastly, for the associations of lipids with glucose/insulin metrics, results remain largely unchanged after either including only baseline data, or additionally adjusting for glucose-lowering medication use (**Supplemental Figures S7 and S8**).

Supplemental Figures

List of figures

Supplemental Figure S1. Participant selection and analyses flowchart

Supplemental Figure S2. Manhattan plot for the association of individual lipid species with meat intake

Supplemental Figure S3. Metabolite set enrichment results for meat intake-related lipids (P<0.05)

Supplemental Figure S4. Pathway enrichment results for meat intake-related lipids (P<0.05)

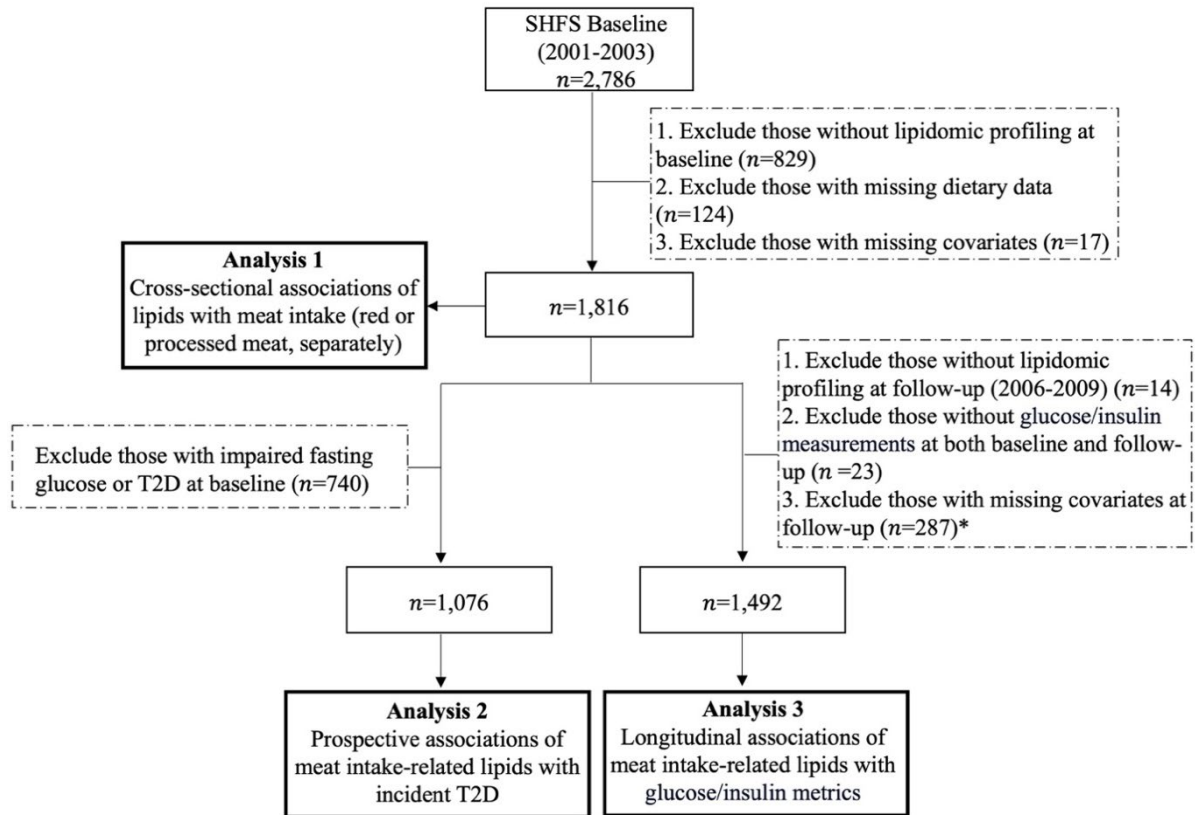
Supplemental Figure S5. Correlation between the consumption of unprocessed red or processed meat and food groups

Supplemental Figure S6. Correlation between the meat intake-related lipids and food group

Supplemental Figure S7. Associations between meat intake-related lipids and glucose/insulin homeostasis metrics using baseline data

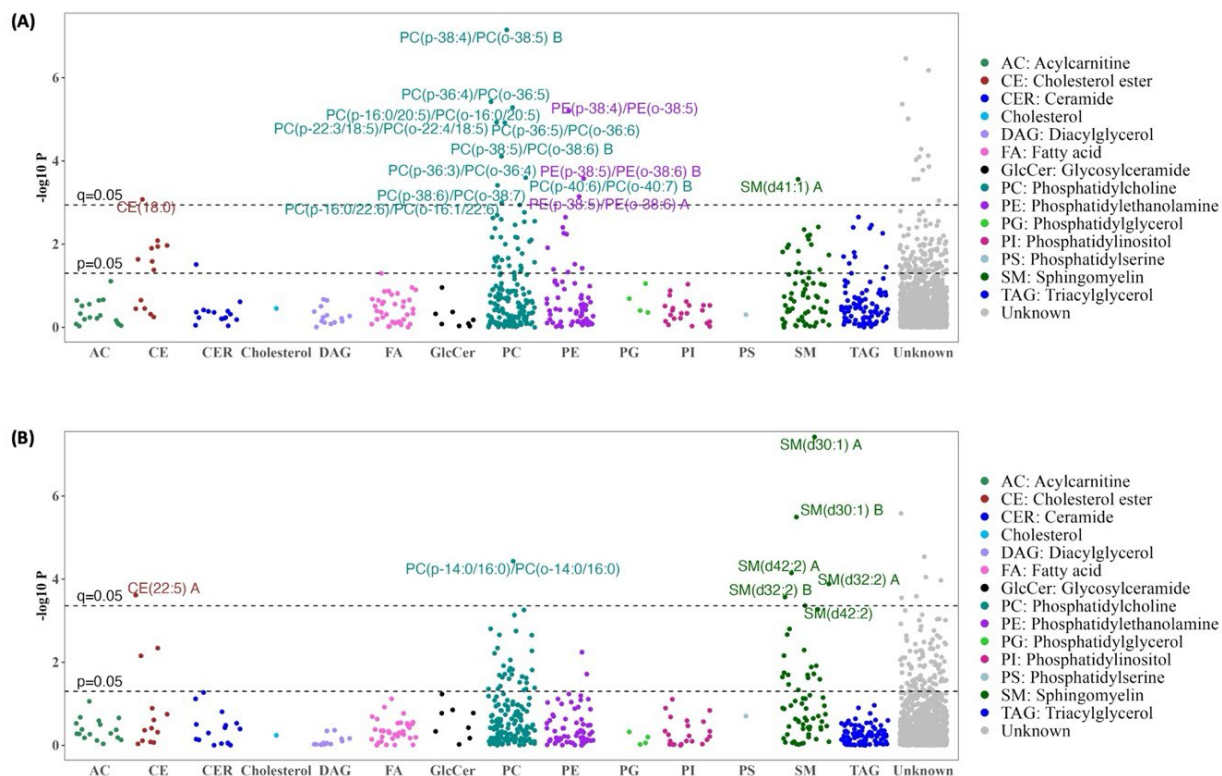
Supplemental Figure S8. Associations between meat intake-related lipids and glucose/insulin homeostasis metrics (sensitivity analyses)

Supplemental Figure S1. Participant selection and analyses flowchart



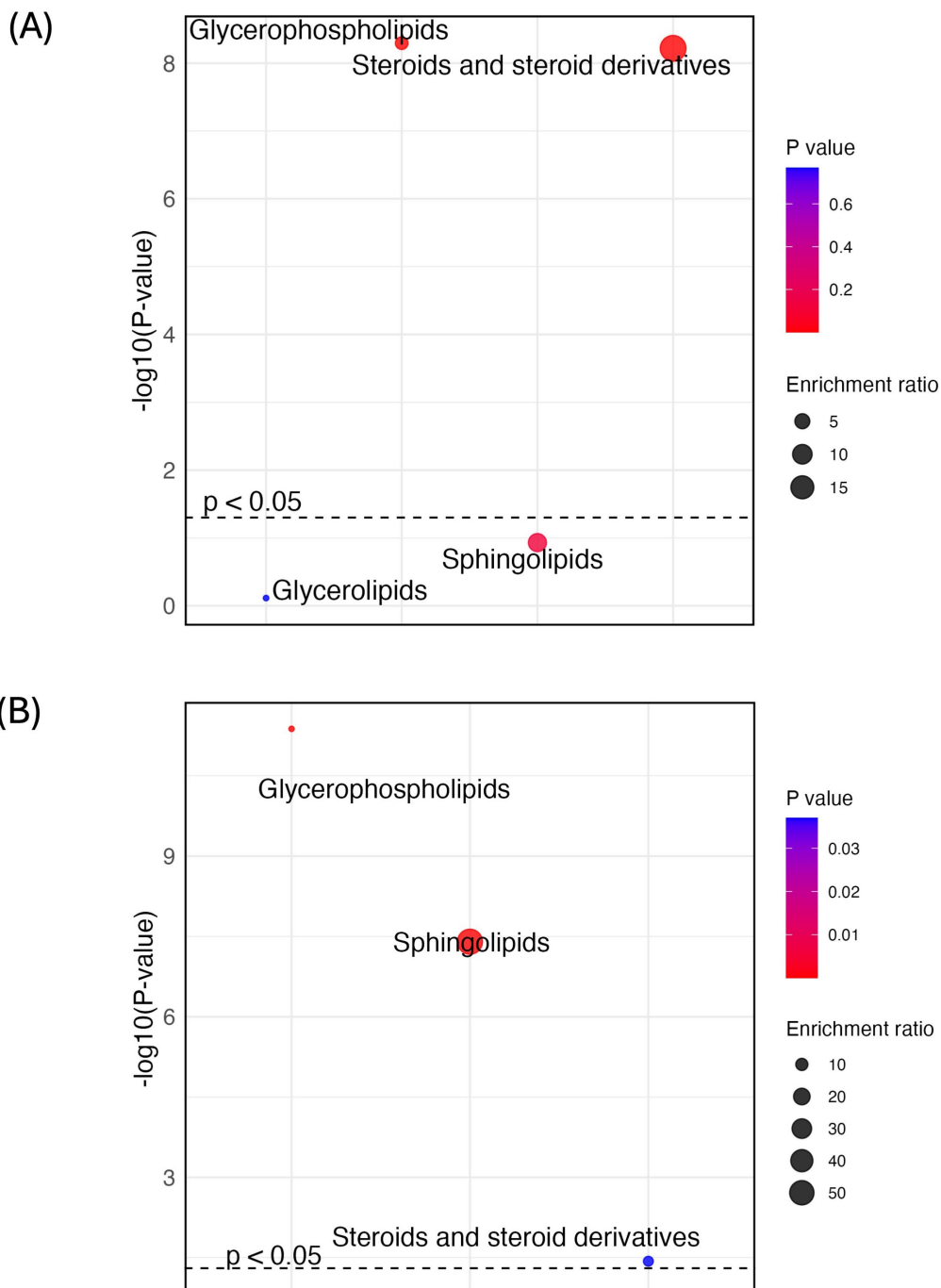
*Including n= 235 with missing total energy intake at follow-up, due to unavailable dietary data.

Supplemental Figure S2. Manhattan plot for the association of individual lipid species with meat intake. (A) unprocessed red meat; (B) processed meat



X-axis: lipid classes; Y-axis: $-\log_{10} P$. Different colors represent different lipid categories. The dashed line represents significance level at $q=0.05$ or $P=0.05$. The letter A or B in the name of lipids represents isomers.

Supplemental Figure S3. Metabolite set enrichment results for meat intake-related lipids (P<0.05). (A) unprocessed red meat; (B) processed meat

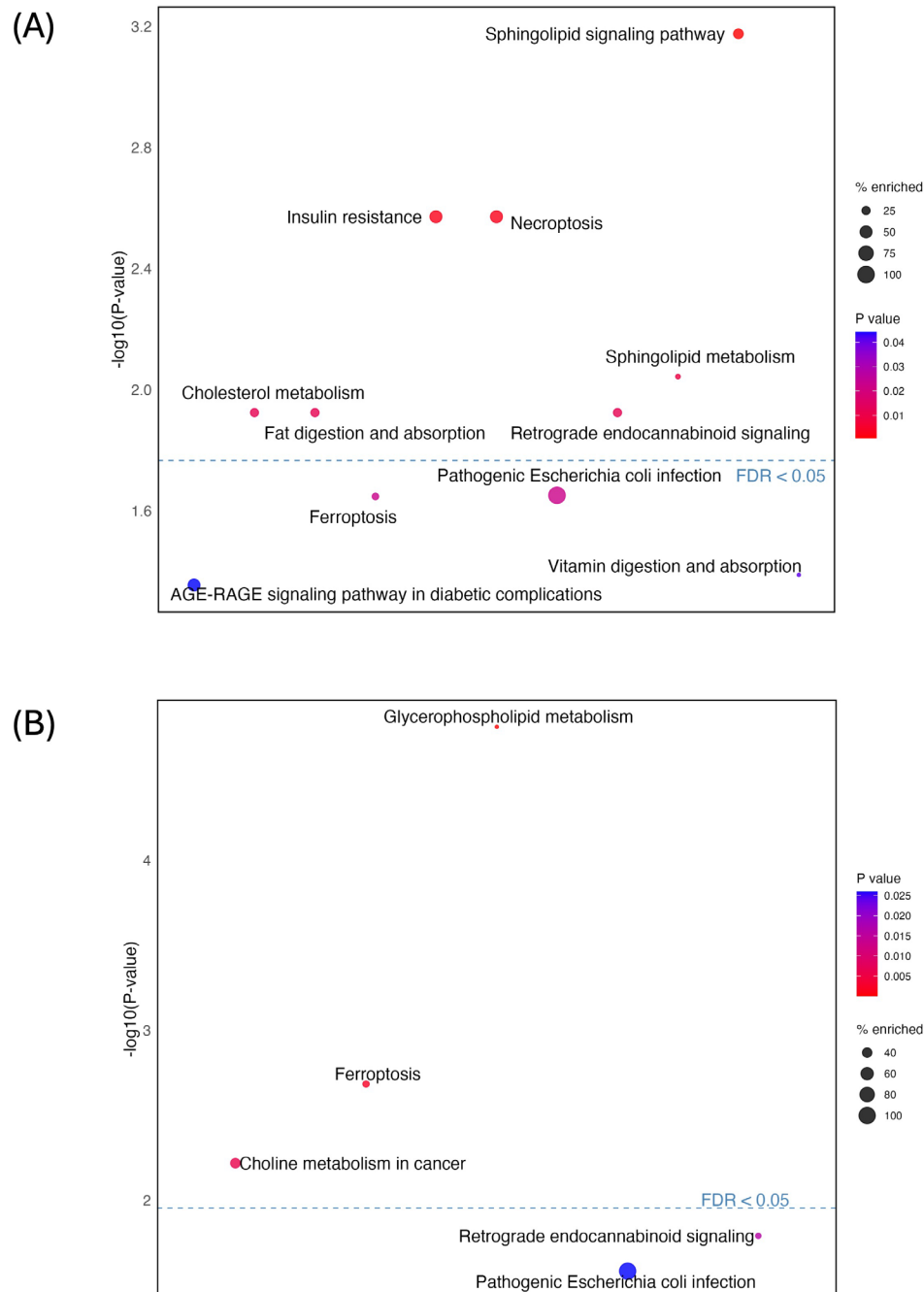


X-axis: Metabolite sets; Y-axis: $-\log_{10}P$. Analyses were conducted using MetaboAnalyst 6.0.

Enrichment ratio is calculated as (number of observed hits)/(number of expected hits)

Supplemental Figure S4. Pathway enrichment results for meat intake-related lipids (P<0.05).

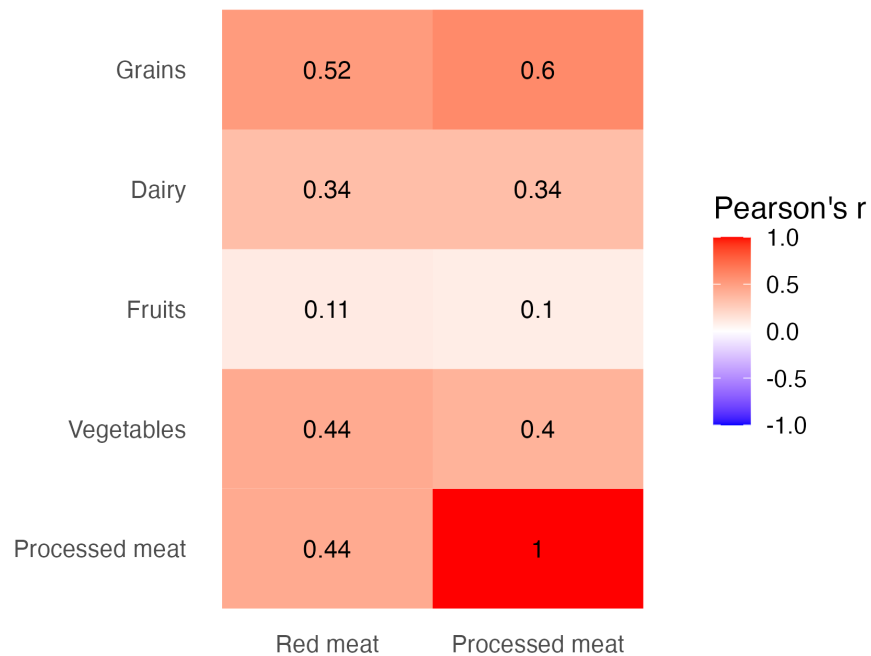
(A) unprocessed red meat; (B) processed meat



X-axis: Metabolic pathways ; Y-axis: $-\log_{10}P$. Analyses were conducted using LIPEA platform.

Percentage of enrichment is calculated as (number of hits)/(number of lipids in the pathway).

Supplemental Figure S5. Correlation between the consumption of unprocessed red or processed meat and food groups

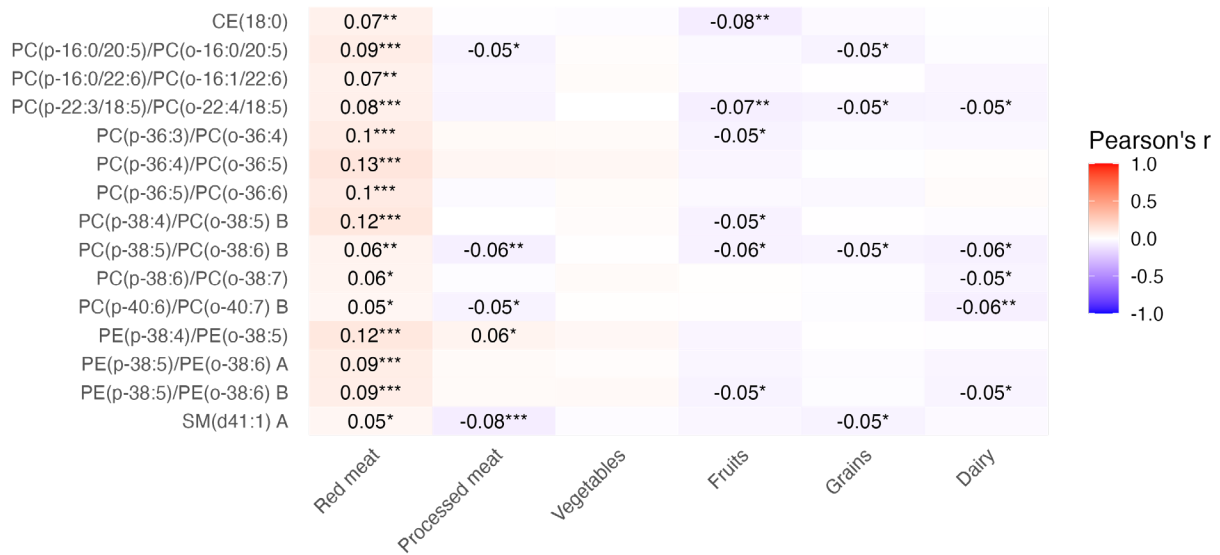


Correlation is calculated with Pearson's correlation coefficients between food consumptions in servings/day. Red color represents positive associations and blue represents negative associations.

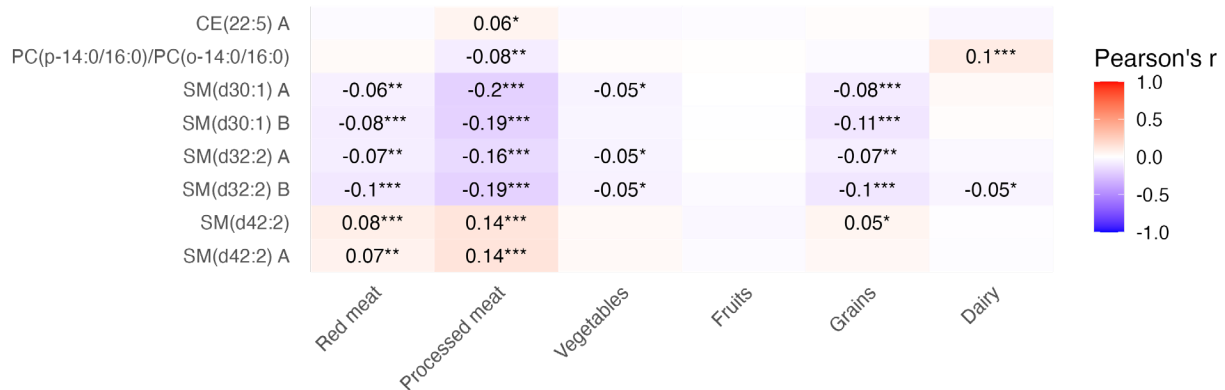
Grains: breads, cereals, rice, pasta; dairy: milk, yogurt, cheese.

Supplemental Figure S6. Correlation between the meat intake-related lipids and food groups. (A) unprocessed red meat; (B) processed meat

(A)

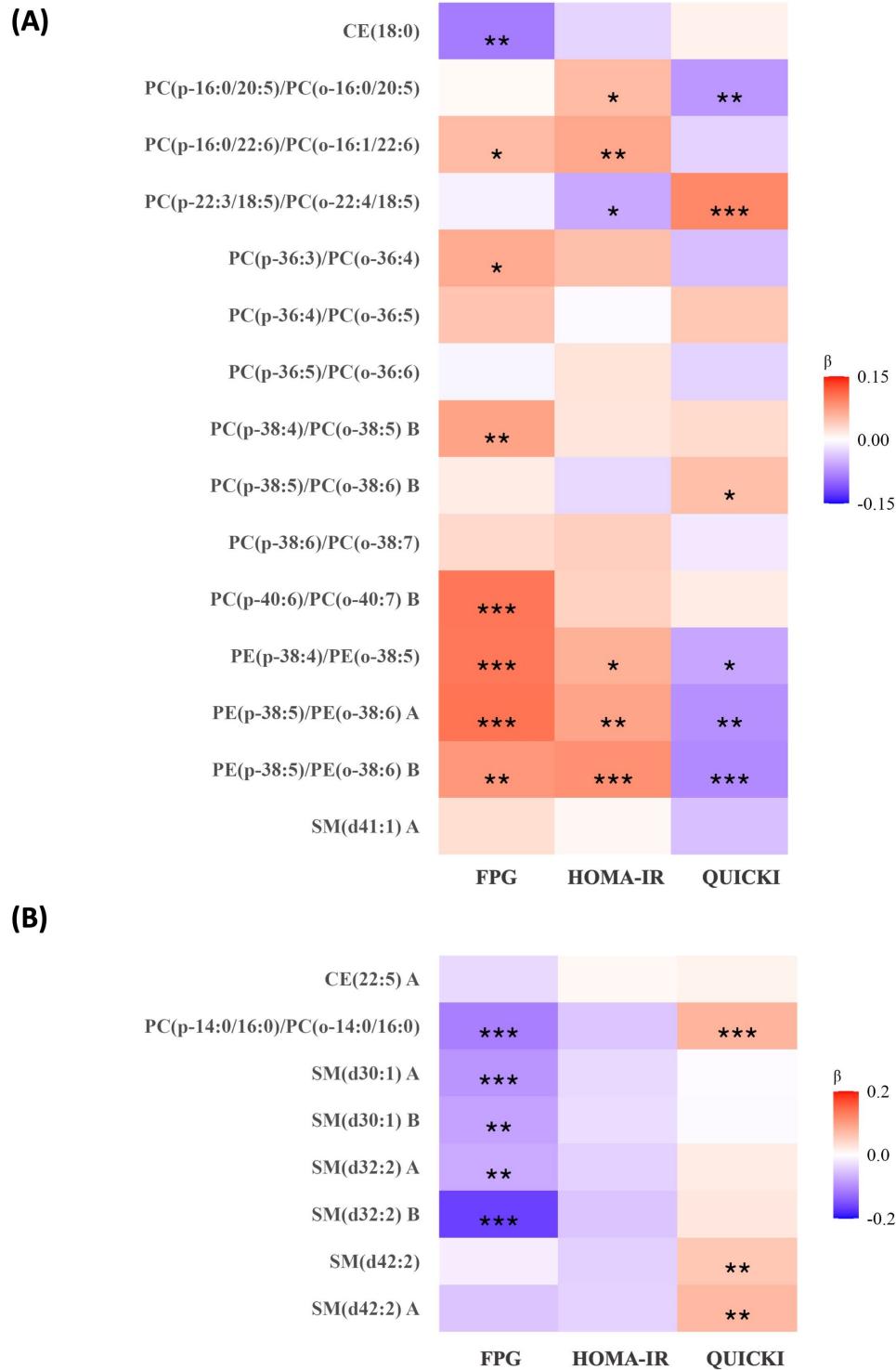


(B)



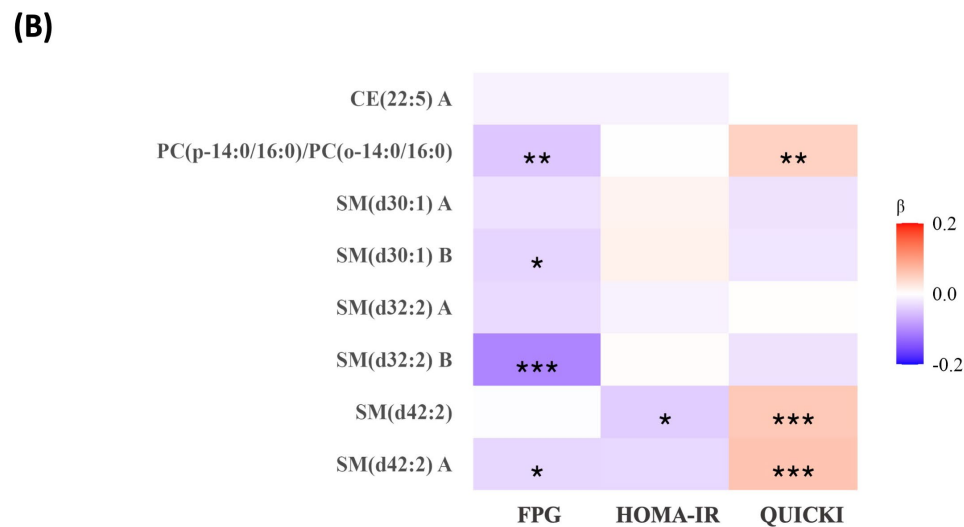
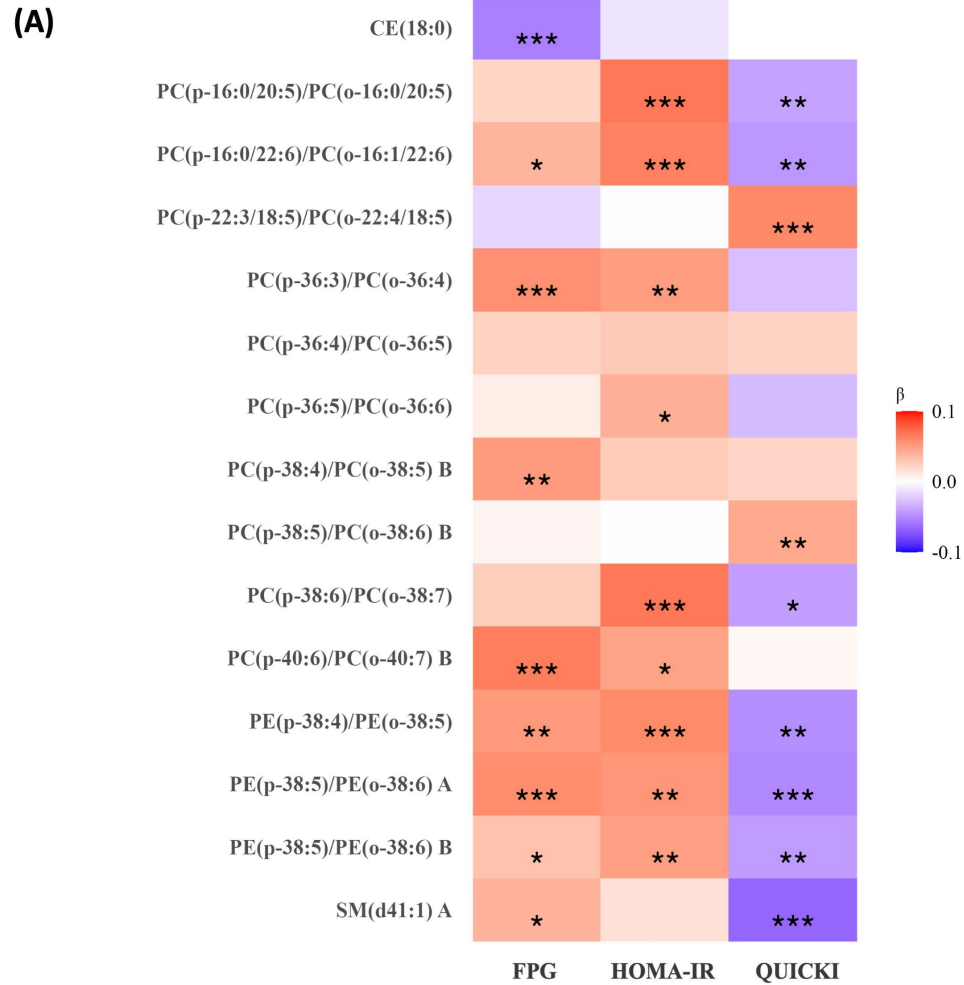
Correlation is calculated with Pearson's correlation coefficients between lipids and food consumptions in servings/day. Red color represents positive associations and blue represents negative associations. Grains: breads, cereals, rice, pasta; dairy: milk, yogurt, cheese. *P < 0.05, **P < 0.01, ***P < 0.001.

Supplemental Figure S7. Associations between meat intake-related lipids and glucose/insulin homeostasis metrics using baseline data. (A) unprocessed red meat (B) processed meat



Regression coefficients were obtained from mixed-effects linear regression models, adjusting for age, sex, study center, education, BMI, total energy intake, smoking, alcohol drinking, physical activity levels, lipid-lowering medication use, and hypertension at baseline. Red color represents positive associations and blue represents negative associations. As sensitivity analyses, this analysis was restricted to baseline data. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Supplemental Figure S8. Associations between meat intake-related lipids and glucose/insulin homeostasis metrics (sensitivity analyses). (A) unprocessed red meat (B) processed meat



Regression coefficients were obtained from mixed-effects linear regression models, adjusting for age, sex, study center, education, BMI, total energy intake, smoking, alcohol drinking, physical activity levels, lipid-lowering medication use, and hypertension at baseline. Red color represents positive associations and blue represents negative associations. As sensitivity analyses, glucose lowering medication use was additionally adjusted. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

References

- [1] Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, Cucchiara AJ, Savage PJ, Howard BV. The Strong Heart Study A study of cardiovascular disease in American Indians: design and methods. *American journal of epidemiology*. 1990;132(6):1141-55.
- [2] Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods*. 2015;12(6):523–526.
- [3] Fan S, Kind T, Cajka T, Hazen SL, Tang WW, Kaddurah-Daouk R, et al. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. *Analytical chemistry*. 2019;91(5):3590–3596.
- [4] Pang Z, Lu Y, Zhou G, Hui F, Xu L, Viau C, Spigelman AF, MacDonald PE, Wishart DS, Li S, Xia J. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic acids research*. 52(W1):W398-406.
- [5] Acevedo A, Durán C, Ciucci S, Gerl M, Cannistraci CV. LIPEA: lipid pathway enrichment analysis. *BioRxiv*. 2018 Mar 3:274969.