Online Supplementary Materials

Study participants. The Strong Heart Family Study (SHFS, 2001-ongoing), a component of the Strong Heart Study (SHS, 1989-ongoing), is a multicenter, family-based prospective study designed to identify genetic factors for cardiovascular disease (CVD), T2D and their risk factors in American Indians. A total of 2,780 tribal members (≥18 years old) from 12 tribes residing in Arizona, North Dakota, South Dakota, and Oklahoma were initially examined in 2001-2003 and re-examined in 2006-2009 (mean 5.5 years apart) using the same protocols. At each visit, participants received a personal interview to collect data on demographic characteristics, medical history and lifestyle risk factors including smoking, alcohol consumption, diet and physical activity. A physical examination was given to each participant, including anthropometric and blood pressure measurements and an examination of the heart and lungs. Biospecimens including fasting plasma samples were collected at each visit. Laboratory methods were reported previously (1). All living participants are currently being followed through 2026 to collect information for disease morbidity and mortality. All participants have given informed consent. Detailed descriptions of the SHS protocols for the collection of phenotype data have been described previously (2).

Participants in the current study met the following criteria: 1) attended clinical examinations and had available fasting plasma samples at both baseline (2001–2003) and 5-year follow-up (2006–2009); 2) were free of overt CVD at baseline. Participants with missing information for fasting glucose or hypoglycemia medications at either time point were excluded.

Definition of T2D. According to the American Diabetes Association 2003 criteria (3), diabetes was defined as fasting plasma glucose ≥7.0 mmol/L or using hypoglycemic medications. Impaired fasting glucose (IFG) was defined as a fasting glucose of 6.1–6.9 mmol/L and no

hypoglycemic medications, and normal fasting glucose (NFG) was defined as fasting glucose <6.1 mmol/L. Incident diabetes was defined as participants who had NFG at baseline (2001-2003) but developed new diabetes by the end of 5-year follow-up (2006-2009) or 16-year follow-up (December 31, 2017). T2D ascertainment was based on one single blood test at baseline and 5-year follow-up. Incident T2D at 5-year follow-up was detected by physical examination during follow-up visit at SHS field Centers, according to the ADA diagnosis criteria. Because SHS participants did not undergo in-person examination at 16-year visit (thus no blood samples collected for fasting glucose measurement), T2D ascertainment was based on medical records abstracted by trained research staff.

Assessments of clinical factors. Fasting plasma glucose, insulin, lipids, and lipoproteins were measured by standard laboratory methods as previously described (1). Body mass index (BMI) was calculated as body weight in kilograms divided by the square of height in meters. Hypertension was defined as blood pressure levels ≥140/90 mmHg or use of antihypertensive medications. Insulin resistance was assessed using homeostatic model assessment (HOMA) according to the following formula: HOMA-IR = fasting glucose (mg/dL) × insulin (μU/mL)/405.(4) Pancreatic *β*-cell function (HOMA-β) was assessed using the formula: 360 × fasting insulin (μU/mL) / (fasting glucose (mg/dL) - 63) (4). Insulin sensitivity was estimated by calculating the Quantitative Insulin sensitivity check index (QUICKI =1/[log insulin (mU/L) + log baseline glucose (mg/dL)] (5). Renal function was assessed using the estimated glomerular filtration rate (eGFR) calculated by the MDRD equation (6). For cigarette smoking, participants were classified as current smokers, former smokers, and never smokers. Alcohol consumption was determined by self-reported history of alcohol intake, the type of alcoholic beverages consumed, frequency of alcohol consumption, and average quantity consumed per day

and per week. Physical activity was assessed by the mean number of steps per day calculated by averaging the total number of steps recorded each day during a 7-day period (7). Dietary intake, including total protein intake, total calories intake, total fat intake was assessed using the Block Food Frequency questionnaire (8). Diet quality was assessed by the Alternate Healthy Eating Index (AHEI), which showed more advantages than the Healthy Eating Index (HEI) in predicting major chronic disease and CVD risks (9, 10). A higher score of dietary quality based on AHEI are strongly associated with lower risks of cardiovascular mortality (11).

Information on use of medications including anti-hypertensive, hypoglycemic and lipid-lowering drugs was also collected at each visit.

Blood sample collection. Participants were instructed to fast overnight before their visit, and fasting blood sample was collected into 10ml EDTA tubes at the SHS 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 ^{o}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.

Lipidomics data acquisition via liquid chromatograph-mass spectrometry (LC-MS)

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 μL. 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 μL. 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 × 2.1 mm, 1.7 μm) with an Acquity CSH C18 guard column (5 mm × 2.1 mm, 1.7 μm) (Waters, Milford, MA). The column temperature was 65 °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. Detailed methods for sample extraction and lipidomic analysis on ESI (+) and ESI (-) were described in the online supplementary methods.

Lipidomics data pre-processing and quality control

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 (12). 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) (13), 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. After data pre-processing, we obtained 1,809 lipids (579 known and 1,230 unknown lipids) in both positive and negative ionizations (787 positive and 1,022 negative) from 1,983 samples at baseline and 1,994 samples at 5-year follow-up. The lipid levels of the duplicated samples were highly correlated (Spearman's correlation coefficient $\rho = 0.95$, $p < 2.2 \times 10^{-16}$).

References

1. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le N-A, Oopik AJ, et al. 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. North KE, Howard B V, Welty TK, Best LG, Lee ET, Yeh JL, et al. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. American journal of epidemiology. 2003;157(4):303–14.

3. Genuth S, Alberti KG, Bennett P, Buse J, Defronzo R, Kahn R, et al. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus2, the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Follow-up report on the diagnosis of diabetes mellitus. Diabetes Care. 2003;26:3160–7.

4. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412–9.

5. Katz A, Nambi SS, Mather K, Baron AD, Follmann DA, Sullivan G, et al. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. The Journal of Clinical Endocrinology & Metabolism. 2000;85(7):2402–10.

6. Froissart M, Rossert J, Jacquot C, Paillard M, Houillier P. Predictive performance of the modification of diet in renal disease and Cockcroft-Gault equations for estimating renal function. Journal of the American Society of Nephrology. 2005;16(3):763–73.

7. Tudor-Locke C, Ainsworth BE, Thompson RW, Matthews CE. Comparison of pedometer and accelerometer measures of free-living physical activity. Medicine & Science in Sports & Exercise. 2002;34(12):2045–51.

8. Willett W. Nutritional epidemiology. Vol. 40. Oxford university press; 2012.

9. Chiuve SE, Fung TT, Rimm EB, et al. Alternative dietary indices both strongly predict risk of chronic disease. J Nutr. 2012;142(6):1009-1018.

10. McCullough ML, Willett WC. Evaluating adherence to recommended diets in adults: the Alternate Healthy Eating Index. Public Health Nutr. 2006;9(1A):152-157.

11. Morze, J.; Danielewicz, A.; Hoffmann, G.; Schwingshackl, L. Diet Quality as Assessed by the Healthy Eating Index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension Score, and Health Outcomes: A Second Update of a Systematic Review and Meta-Analysis of Cohort Studies. J. Acad. Nutr. Diet. 2020, 120, 1998–2031.

12. 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.

13. 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.