Introduction The American Indian or Alaska Native population experienced substantially higher death rates than other populations in the United States.[1-4] Cardiovascular disease (CVD) is the leading causes of deaths in American Indians.[5, 6] Epidemiological studies have shown that individuals with metabolic diseases suffer from higher mortality rate than those without.[7-11] In addition, high levels of low-density lipoprotein cholesterol (LDL-c) and total cholesterol, and a low level of high-density lipoprotein cholesterol (HDL-c) have previously been associated with increased cardiovascular disease (CVD) mortality or all-cause mortality in different populations,[12-16] although results are mixed.[17-20] These findings suggest that dyslipidemia may contribute to all-cause or CVD mortality. Deciphering the metabolic pathways underlying the association between lipid metabolism and mortality is likely to lead to novel biomarkers for risk prediction and stratification. However, routine lipid panel cannot capture all molecular lipids in blood (i.e., blood lipidome), and thus has limited ability to detect perturbed lipid metabolism associated with mortality. Lipidomics is an emerging high-throughput biochemical technique that can identify and quantify hundreds to thousands of lipid species in the biospecimens, and thus is well suited to characterize disturbed lipid metabolism that precedes mortality. There has been a growing interest in using lipidomics to better understand the associations between altered lipid metabolism and mortality. Several studies have reported associations of altered blood lipid species, such as phosphatidylcholines, ceramides, ratio of polyunsaturated fatty acids to total fatty acids, phosphatidylethanolamines, with all-cause or cause-specific mortality in various populations.[21-26] However, existing studies were either limited to smaller sample size or only examined a smaller number of lipids at a single time point. To date, no study has investigated the relationship between fasting plasma lipid species and mortality in American Indians. Moreover, no large-scale longitudinal epidemiological studies have examined the association between the longitudinal change in lipid species and mortality in any racial/ethnic group. Here we report findings from the first longitudinal lipidomic profile in 3,821 fasting blood samples from 1,930 American Indians attending two clinical examinations (mean follow-up: 5.5 years) in the Strong Heart Family Study (SHFS). Our primary goals here are to (1) identify individual lipid species associated with risk of all-cause and CVD mortality beyond traditional risk factors, and (2) examine the association between longitudinal changes in plasma lipidome and risk of all-cause and CVD mortality. Research design and methods Study populations Discovery cohort. The Strong Heart Family Study (SHFS, 2001-ongoing), a family-based prospective study designed to identify genetic and lifestyle factors for CVD and risk factors in American Indians, as previously described.[5, 27, 28] Briefly, 2,780 tribal members (aged 14 and older) in three geographic regions (Arizona, North/South Dakota, Oklahoma) were initially examined in 2001-2003 and re-examined in 2006-2009 (mean 5.5 years apart) using the same protocols. Information for demography, family history, medical records, and lifestyle

was collected at each visit. A total of 1,930 individuals (1,203 women, mean age at baseline: 40.4) with complete information for clinical and lipidomic data were included in the current analysis. More information for covariate assessments was described elsewhere.[29] All the participants provided informed consents. The SHFS protocols were approved by the Institutional Review Boards of participating institutions and the American Indian tribes. Replication cohort. The Malmö Diet and Cancer-Cardiovascular Cohort (MDC-CC), a subcohort of the Malmö Diet and Cancer (MDC) Study,[30] is a prospective population-based cohort designed to study the epidemiology of carotid artery disease (CAD) and its risk factors in European Caucasians in Sweden, as previously described.[31, 32] A subset of 3,943 participants (2,319 women, mean age at baseline: 57.7) who had complete information for clinical phenotypes and lipidomic data were included in the analysis. All MDC-CC participants provided written informed consent, and the study was approved by the Ethics Committee at Lund University. Follow-up and ascertainment of mortality In the SHFS, baseline information was collected in 2001-2003 and all living participants were followed through 2020. Detailed methods for ascertainment of deaths and causes of deaths in the SHFS have been described previously.[5, 6, 27] Briefly, death of a participant was identified from the Indian Health Service hospital records and direct contact with the family members. The cause of death was determined by physicians on the SHS Mortality Review Committee using medical records, autopsy reports, and informant interview. Information on mortality was retrieved from the mortality decision form used for cardiovascular disease surveillance in the SHS. The cause of death was determined independently by two members of the SHS Mortality Committee after reviewing information including death certificates and medical records including pathology reports as well as informant interviews, when needed. Death certificate codes were recorded according to the International Classification of Diseases Injuries, and Causes of Death, 9th Revision (ICD-9).[5] CVD mortality was defined as death caused by myocardial infarction, stroke, sudden death from coronary heart disease, or congestive heart failure (ICD-9 codes 390-448). Cancer mortality was defined as death with a primary cause of death listed as malignant neoplasm (ICD-9 codes 140-239). In the MDC-CC, all subjects were followed from the baseline examination until death, emigration from Sweden, or December 31 2020, whichever came first. Causes of death and the vital status of participants were retrieved through record linkage of the personal identification number and the Swedish Cause of Death Register and the National Tax Board.[33, 34] CVD mortality was defined as death caused by CVD on the basis of ICD-9 codes 390-459 or ICD-10 codes I00-I99. Cancer mortality was defined based on ICD-9 codes 140-239 or ICD-10 codes D00-D48. Assessments of clinical covariates In the SHFS, demographic information (age and sex), lifestyle habits (smoking/drinking status, physical activity), medical history, family history of illnesses, and use of prescription medications were collected using structured

questionnaires.[5, 28] Smoking status will be categorized as current versus non-current smokers (former and never smokers combined). Drinking status will be categorized as current versus non-current drinkers (former and never drinkers combined). Anthropometric measures including height, weight and waist circumferences were obtained through physical examinations at each visit. Body mass index (BMI) was calculated as body weight in kilograms divided by the square of height in meters. Fasting glucose and blood lipids, including total cholesterol, triglycerides, LDL-c and HDL-c were measured by standard laboratory methods.[5] Hypertension will be defined as blood pressure levels ≥140/90 mmHg or use of antihypertension medications. Type 2 diabetes will be defined as fasting plasma glucose ≥126 mg/dL or use of hypoglycemic medications. Estimated glomerular filtration rate (eGFR) was calculated using the chronic kidney disease (CKD) Epidemiology Collaboration (CKD-EPI).[35] Physical activity was assessed by the mean number of steps per day calculated by averaging the total number of steps recorded each day during the 7-day period. Information on use of lipid-lowering drugs was also collected at each visit. In the MDC-CC, information on lifestyle and clinical factors, such as BMI, blood pressure, use of antihypertensive treatment, current smoking, clinical lipids, and blood glucose, were obtained as previously described.[36] Lipidomic data acquisition, pre-processing and quality control In the SHFS, relative abundance of fasting plasma lipid species at two time points (mean followup: 5.5 years) was quantified by untargeted liquid chromatography-mass spectrometry (LC-MS) as described previously.[29] After pre-processing and quality control, we obtained 1,542 lipids (518 known, 1,024 unknown) in 3,977 plasma samples (1,983 at baseline, 1,994 at follow-up). After further excluding outlier samples or those with missing covariates, 1,930 participants (1,930 at baseline, 1,891 at follow-up) with complete clinical and lipidomic data were included in the analysis (Figure S1). Lipidomic profiling in the MDC-CC was performed using fasting plasma samples collected at enrollment (i.e., baseline) by mass spectrometry as previously described.[37, 38] Spectra were analyzed with in-house developed lipid identification software based on LipidXplorer.[39] Data processing, normalization and batch correction was performed as described previously.[31, 38, 40] Lipids that were present in at least 70% of the participants were included in the subsequent statistical analysis, resulting in a total of 184 lipids.[31, 38] Missing values of the remaining lipids were imputed using the NIPALS algorithm.[41] A total of 3,943 participants (58.8% females) with available lipidomic data were included in the current study to replicate findings from the SHFS. The mean age of MDC-CC participants was 57.7 years at baseline and 81.3 years at follow-up (mean 23.7 years of follow-up). Statistical analyses Figure S1 illustrates the procedures for participant selection and statistical analyses. All the continuous variables including lipids were standardized to zero mean and unit variance. Multiple testing was controlled by false discovery rate (FDR) using the Storey's q-value method.[42, 43] Prospective association analysis. To identify baseline

plasma lipids predictive of risk for allcause or CVD mortality, we constructed Cox proportional hazards model (SHFS, n=1,930; MDC-CC, n=3,943), in which baseline lipid was the predictor and time to all-cause or CVD mortality was the outcome, adjusting for age, sex, BMI, smoking, hypertension, diabetes, cardiovascular disease and LDL-c at baseline. The analysis in the SHFS additionally adjusted for eGFR. Family relatedness in the SHFS was accounted for by including a frailty term in the model. The analysis was first conducted in the SHFS, followed by replication of top hits (P0 and loss of connectivity (LOC) if MDC