SUPPLEMENTAL MATERIAL

Abbreviations for Supplemental Material:

Cohort study name (study acronym) (country): Western Australian Birth Cohort Study (Raine Study) (Australia), Atherosclerosis Risk in Communities Study (ARIC) (USA), Framingham Heart Study (FHS) (USA), Netherlands Epidemiology in Obesity Study (NEO) (The Netherlands), The Fenland Study (Fenland) (United Kingdom), Young Finns Study (YFS) (Finland), Women's Genome Health Study (WGHS) (USA), Women's Health Initiative (WHI) (USA), Multi-Ethnic Study of Atherosclerosis (MESA) (USA), Cardiovascular Health Study (CHS), the Rotterdam Study (RS) (The Netherlands), and the UK Biobank (UKB) (United Kingdom)

Abbreviations: BMI, body mass index; CHARGE, Cohorts for Heart and Aging Research in Genetic Epidemiology; CI, confidence interval; HDL-C, high-density lipoprotein cholesterol concentrations; mo, month; N, total sample size; SE, standard error; SNP, single nucleotide polymorphism; SSB, sugar-sweetened beverages; TG, triglyceride concentrations; wk, week.

I. Methods

Subjects

The primary study population consisted of up to 63,599 study participants of European ancestry from 11 well-characterized, large population-based cohorts participating in the CHARGE consortium's Nutrition Working Group: The Raine Study (Australia), Atherosclerosis Risk in Communities Study (USA), Framingham Heart Study (USA), Netherlands Epidemiology in Obesity Study (The Netherlands), The Fenland Study (United Kingdom), Young Finns Study (Finland), Women's Genome Health Study (USA), Women's Health Initiative (USA), Multi-Ethnic Study of Atherosclerosis (USA), Cardiovascular Health Study (USA), and the Rotterdam Study (The Netherlands). Individuals were excluded if they were missing data for all genotypes, age, sex, total energy intake, SSB intake, HDL-C concentrations, or TG concentrations. To assess reproducibility, we studied participants from the UKB, an open access prospective cohort study of greater than 500,000 participants established in 2006. In the UKB, we included only participants of European ancestry with valid dietary data (n=70,747) and TG and HDL-C concentrations (n=59,220) available at baseline. A full description of each of the 12 cohorts is presented in **Table I**. All study participants provided written informed consent, and approval for all study protocols was granted by local institutional review boards and/or oversight committees.

Dietary Assessment and SSB Consumption

Dietary intakes for all CHARGE cohorts were estimated using cohort-specific food frequency questionnaires that have been previously assessed for validity in each cohort (**Table II**). Food and beverage intakes were estimated from food frequency questionnaire data using cohort-specific groupings and nutrient intakes were calculated from these data by multiplying the

frequency of consumption of a food item by the respective nutrient content per standard serving size. Participants were excluded if they were missing dietary data or had implausible total energy intakes based on cohort-specific cut-points similar to previous studies.[37] Estimates of SSB consumption were derived from cohort-specific beverage items: (1) regular cola/soda with sugar; (2) caffeine free cola/soda with sugar; (3) carbonated non-soda beverages with sugar (e.g., 7-Up, ginger ale); and (4) fruit-flavored drinks (e.g. lemonade, Tang, Hawaiian Punch, squash). One serving of SSB was converted to 360 mL (12 fl oz.) for analysis. In the UKB, participants were either classified as SSB consumers or non-consumers and nutrient intakes were estimated based on their responses to a single 24-hr diet recall collected closest to the time of blood draw during the first in-person baseline assessment (**Table III**).

Genotyping and Imputation in CHARGE and UKB

Genotyping of individuals was carried out in each cohort separately using Affymetrix or Illumina genotyping platforms. Genetic variants from 100-kilobase upstream to 190-kilobase downstream of the transcription start site of the *CHREBP* gene [chr7:72823661-73136661(hg19)] were selected individually in each cohort. Genotyped SNPs were excluded based on cohort-specific low call rates (minimally <90%) or departure from Hardy-Weinberg equilibrium (minimally $p<1x10^{-4}$). Variants in the *CHREBP* region that were not directly genotyped were imputed using algorithms implemented in IMPUTE2 or MACH with the Haplotype Reference Consortium (HRC), 1000G phase 3, and/or UK10K as reference panels (**Table IV)**. After imputation, SNPs were excluded on the basis of low minor allele count (allele frequency*sample size <20) or low imputation quality (MACH: R^2 <0.3; IMPUTE2: INFO score <0.4). After meta-analysis, we excluded SNPs with a low number of contributing cohorts (≤3).

In total, 1,606 SNPs were included in the meta-analysis. For all interaction tests, we excluded SNPs with a minor allele frequency < 1% to ensure enough individuals were included throughout the range of SSB consumption categories.

Outcome and Covariate Definitions

In the CHARGE cohorts, HDL-C and TG concentrations were measured by standard enzymatic methods from venous blood collected after ≥8 hours of fasting. In the UKB, enzyme immunoinhibition and glycerol phosphate oxidase-peroxidase methods were used to measure HDL-C and TG, respectively, from venous blood was collected after a mean fasting time of 3.8 hours.

Relevant covariate data was collected by questionnaire, including information on education, smoking status, physical activity, and alcohol intake. Description of cohort-specific methodologies for all relevant covariates are described in **Table V**. Body mass index was calculated from measured or self-reported weight (kg) divided by height squared (m^2).

Statistical Analyses

All CHARGE cohorts followed a uniform, pre-specified analysis plan as detailed in the data sharing template used for sharing summary data between the CHARGE consortium cohorts (**Appendix I**). Participants were grouped by category of SSB consumption (<1 serving/month, 1-4 serving/month, 1-2 serving/week, 3-7 serving/week, >1 serving/day) and a natural logarithmic transformation was applied to TG concentrations to approximate a normal distribution. Standardized linear regression models (and mixed effect linear regression for the Framingham Heart Study) containing an additive genetic effect were used to examine the associations

between SSB consumption, SNPs, and the interaction between SSB consumption and SNPs on HDL-C and TG concentrations in each cohort (**Figure I**).

Associations between SSB Consumption and Lipid Traits in CHARGE Cohorts

Four models were assessed for the association between SSB consumption and lipid traits. Model 1 adjusted for age (continuous), sex (M/F) (where applicable), total energy intake (continuous), and study site for multi-centered cohorts. Model 2 adjusted for model 1 covariates plus education (cohort-specific definition), smoking status (cohort-specific definition), physical activity (cohort-specific definition), and alcohol intake (grams/day) (**Table V**). Model 3 adjusted for model 2 covariates plus BMI (continuous). Model 4 adjusted for model 3 covariates plus servings per day of vegetables, total fruit, whole grains, nuts/seeds, and seafood, as well as % energy from saturated fat (continuous). Individuals missing covariate data were excluded from those models.

For associations between SSB intake on lipid outcomes, beta coefficients were combined through inverse-variance weighted, random effects meta-analyses. When SSB intake was assessed as a categorical variable, multivariate meta-analyses [42] were conducted using the *mvmeta* R package (https://cran.r-project.org). When SSB intake was assessed using an additive model (per increase in category of SSB intake), univariate meta-analyses were conducted using the *meta* R package (https://cran.r-project.org). Results from analyses between SSB intake and lipid traits were considered statistically significant at $p<0.05$.

The Cochrane Q statistic [43] and the I^2 statistic [44] were used to examine heterogeneity among the cohorts. Analyses with moderate-to-high heterogeneity ($I^2>30\%$) were further assessed for potential sources of heterogeneity by conducting meta-regression and sensitivity

analyses. Meta-regression analyses were conducted using the R *meta* package to assess the effect of potential moderator variables on heterogeneity of associations/interactions. Mean effect sizes were compared in meta-analyses stratified by geographical region (US, Europe, or Australia), age (<40, 40-60, or >60 years), BMI (<27 or ≥27 kg/m^2) SSB intake (<0.4 or ≥0.4 servings/week), study date (≤ 2005 or >2005), proportion of current smokers (<20% or ≥20%), and sample size (≤3,500 or >3,500). A statistically significant moderator was defined at $p_{\text{interaction}} < 0.007$ (0.05/7 potential moderators). The influence of individual cohorts on the meta-analyzed estimates was examined by removing one cohort at a time for statistically significant associations.

Associations of SNPs on Lipid Traits in CHARGE Cohorts

Linear regression models examining the association between SNPs (additive) and HDL-C and TG concentrations were adjusted for age, sex, principal components of ancestry (where applicable), and study site (for multi-centered cohorts). SNP associations were not included in the meta-analysis if they had a low number of contributing cohorts (≤3). In the full sample, regression (β) coefficients for the additive genetic effect of SNPs on lipid outcomes were combined through inverse-variance weighted, fixed-effect meta-analyses using METAL (version released on 2011-03-25). We first conducted a meta-analysis for single SNP analyses, and then implemented conditional and joint association (COJO) analysis with GCTA software [45] to select top distinct signals [to account for SNPs that may be in high linkage disequilibrium (LD)] from the summary statistics (using the 1000 genomes cohort as a reference sample). Similar to the genetic interaction analyses, results from these genetic analyses were considered statistically significant at a Bonferroni-corrected $p<0.0001$ (0.05/499 independent tests).

Interactions between SSB Consumption and SNPs on HDL-C and TG in CHARGE Cohorts

To examine interactions between SNPs and SSB consumption, two separate methods were used: 1) a difference test; and 2) a cross-product interaction test. For the difference test, SNP analyses stratified by category of SSB consumption were conducted in each cohort using two covariate models. Model 1 adjusted for age, sex (where applicable), principal components of ancestry (where applicable), study site (for multi-centered cohorts), and total energy intake. Model 2 adjusted for model 1 covariates plus education, smoking status, physical activity, alcohol intake, and body mass index. Individuals missing covariate data were excluded from those models. Next, stratum-specific regression (β) coefficients for the additive genetic effect of SNPs on lipid outcomes in the lowest (<1 serving SSB per month) and highest (>1 servings SSB per day) categories of SSB intake were separately combined through inverse-variance weighted fixed-effect meta-analyses in METAL (version released on 2011-03-25).[46] Finally, the difference test for the regression coefficients between SSB strata was implemented for each SNP using R (version 3.6.0; http://cran.r-project.org) as previously described.[47] For the cross-product interaction test, regression coefficients for cross-product interaction terms (additive SNP x ordinal SSB category) were obtained from the cohorts using the same covariate models as described above and were combined through inverse-variance weighted, fixed-effect meta-analyses using METAL. Robust standard errors were reported in cross-product interaction models to reduce risk of spurious type-I error inflation in SSBxSNP interaction analyses.[48] In each meta-analysis, the Cochrane Q statistic[43] and the I^2 statistic[44] were calculated as a measure of heterogeneity among the cohorts.

Given the large number of statistical tests conducted, the simpleM[49] method was used to estimate 499 effective independent tests among the 1,606 SNPs within and near the *CHREBP* locus among individuals in the 1000 genomes reference panel.[50] Thus, results from all interaction analyses were considered statistically significant at a Bonferroni-corrected $p<0.0001$ (0.05/499 independent tests), and a suggestive interaction was defined as $p<0.005$. Interaction p-values for the difference test are represented with p_{Diff} and cross-product interaction with $p_{interact}$. To select top distinct signals from the summary statistics in interaction analyses, LD-based result clumping was implemented with PLINK software (version 1.90 released 21 May 2017) (http://pngu.mgh.harvard.edu/purcell/plink/)[51] using 1000 genomes as a reference sample using the following thresholds: $p<0.01$ for index SNPs; $p<0.05$ for clumped SNPs; $r<0.50$ and physical distance >500 kb for clumping. For statistically significant associations, the influence of individual cohorts on the estimates from meta-analyses was examined by removing one cohort at a time.

Interactions between SSB Consumption and SNPs on Lipid Traits in the UKB and Meta-Analysis with CHARGE Cohort Results

Data from participants in the UKB were used to examine significant interactions observed in the CHARGE cohorts. Participants were classified as SSB consumers (report of ≥0.5 servings on a 24-hour recall) or non-consumers (no report of SSB consumption on a 24-hour recall). Natural logarithmic transformation was applied to TG concentrations to approximate a normal distribution. To assess associations between top SNPs and TG and HDL-C using an additive genetic effect stratified by SSB consumption category, SAIGE linear-mixed models were implemented to account for familial relatedness and cryptic population structure.[52] Similar

to the CHARGE analyses, two covariate models were constructed. Model 1 adjusted for age (continuous), sex (M/F), fasting hours prior to blood draw (continuous), total energy intake (continuous), principal components of ancestry, education, smoking status, physical activity, and alcohol intake (see **Table V** for details on categorization and measurement). Model 2 adjusted for model 1 covariates plus body mass index (continuous) to examine the potential influence of adiposity on interaction results. Stratum-specific regression (β) coefficients for the additive genetic effect of top SNPs on lipid outcomes in the lowest (UKB: SSB non-consumers; CHARGE: <1 serving SSB per month) and highest (UKB: SSB consumers; CHARGE: >1 servings SSB per day) categories of SSB intake were combined through inverse-variance weighted fixed-effect meta-analyses in METAL. Because large heterogeneity between the two effect estimates was observed for several SNPs, the final effect estimates for the top SNPs are drawn from a random effects meta-analysis conducted in R. The difference test for stratum-specific regression coefficients among SSB consumers and non-consumers was implemented for each SNP using R. Statistical significance for interactions in the UKB analysis and the meta-analysis of UKB and CHARGE cohort results was set at a Bonferroni-corrected $p_{Diff}<0.01$ (0.05/5 top signals) for HDL-C and $p_{Diff}<0.025$ (0.05/2 top signals) for TG concentrations.

Sensitivity Analyses for Interactions in the UKB

Additional modeling of the top result from the CHARGE meta-analysis and the UKB was pursued to improve interpretation and evaluate the influence of adjustment for additional covariates and stratification by fasting status. Adjusted means were derived from linear regression models examining associations between rs71556729 dosage and HDL-C among unrelated SSB consumers and non-consumers in the UKB adjusting for age, sex, fasting hours

prior to blood draw, total energy intake, current smoking status, education, physical activity, alcohol consumption, intakes of dietary fiber and potassium and percent energy from saturated fat and polyunsaturated fat, and BMI. Because the number of hours that the UKB participants fasted prior to blood draw was variable, we conducted sensitivity analyses examining the different test for top candidate SNPs identified for HDL-C and TG concentrations among participants fasting >6, 4-6, and <4 hours.

II. Supplemental Results

Associations between SSB Consumption and Lipid Traits in CHARGE Cohorts

Figures VIII and IX show multivariate inverse-variance weighted random-effects meta-analyses of SSB intake on lipid traits in fully adjusted models (Model 4), and **Figures XII and XIII** represent forest plots additionally including cohort-specific association estimates. Participants in the highest category of SSB consumption (>1 serving/day) had lower mean HDL-C concentrations [β (95% CI):-2.1 (-2.9, -1.2) mg/dl] and higher mean TG concentrations [β (95% CI): 0.06 (0.03, 0.09) ln-mg/dl] than those in the lowest category of SSB intake (<1 serving/month). Covariate adjustment, including adjustment for BMI, did not change the directionality or significance of results, but effect sizes were attenuated as covariates were added to the models (**Table IX**). Moderate to high heterogeneity was observed in both analyses (HDL-C: I^2=38%; TG: I^2=40%). **Figures X and XI** represent forest plots of regression coefficients in individual cohorts and in univariate random-effects meta-analyses for the association of categorical SSB intake (additive) on lipid traits in fully adjusted models (Model 4). Each

increase in SSB intake category was associated with lower mean HDL-C concentrations [β (95% CI) : -0.55 (-0.72, -0.37) mg/dl; p_{trend} = 0.002] and higher mean TG concentrations [β (95% CI): 0.02 (0.01, 0.02) ln-mg/dl; p_{trend} < 0.0001]. Although the direction of effect sizes was consistent for both HDL-C (negative for 11/11 cohorts) and TG (positive for 9/11 cohorts) concentrations, high heterogeneity was observed in these analyses (HDL-C: I^2=67%; TG: I^2=59%).

We conducted sensitivity analyses to investigate potential sources of heterogeneity among the CHARGE cohorts for associations between SSB consumption and lipid traits. Results were similar when each cohort was removed from the meta-analysis individually, indicating the observed heterogeneity could not be attributed to a single cohort. Stratified analyses of the association between SSB intake (additive) on HDL-C concentrations indicated that geographic region of study [USA (N studies=6), Australia (N studies=1), or Europe (N studies=4)] accounted for 100% of the observed heterogeneity among the cohorts ($p_{interaction}$ < 0.0001; **Table X** and **Figure XIV**). The effect size for additive SSB intake on HDL-C concentrations in the USA cohorts was larger [β (95% CI) : -0.77 (-0.89, -0.65) mg/dl] than that of the European cohorts [β (95% CI) : -0.25 (-0.41, -0.09) mg/dl], and the effect size for the Australian cohort was between the two [β (95% CI) : -0.41 (-1.18, 0.37) mg/dl]. Stratified analyses of the association between additive SSB intake on TG concentrations indicated that study mean age [< 40 years (N studies=2); 40-60 years (N studies=5); >60 years (N studies=4)] accounted for 83% of the observed heterogeneity among the cohorts ($p_{interaction}$ = 0.004; **Figure XV**). The effect size of additive SSB intake on TG concentrations among cohorts with a mean age between 40 – 60 years was larger [β (95% CI): 0.02 (0.02, 0.03) ln-mg/dl] than that of cohorts with mean ages < 40 years [β (95% CI): 0.01 (-0.01, 0.02) ln-mg/dl] or > 60 years [β (95% CI): 0.01 (0.001, 0.01) ln-mg/dl]. However, additional heterogeneity remained among the cohorts with a mean age

between 40 – 60 years (I^2 = 55%), which may be attributed to the larger effect size observed in the FHS cohort (I^2 = 0% when FHS cohort removed).

Associations of SNPs on Lipid Traits in CHARGE Cohorts

Main effect associations of *CHREBP* locus SNPs on lipids traits are presented in **Supplement Table X**. We identified 138 SNPs that were significantly associated with HDL-C concentrations (p<0.0001). Among these top SNPs, the COJO analysis identified one distinct signal for HDL-C (*CHREBP*- rs17145750). We identified 325 SNPs that were significantly associated with TG concentrations (p<0.0001). Among these top SNPs, the COJO analysis identified five distinct signals (*CHREBP*-rs71556736, *CHREBP*-rs13225660, *FZD9*-rs42124, *CHREBP*-rs13240662 and *VPS37D*-rs10245965). Individual cohort estimates are presented in forest plots for the top distinct SNPs on lipid traits in **Figures XVI-XXI**. Consistent direction of effect was observed in the single SNP analyses for most cohorts, but moderate-to-high heterogeneity was observed. Most heterogeneity could be attributed to variability in effect sizes observed among cohorts with larger sample sizes. Each of the five top SNPs was significantly associated with lipid concentrations before implementation of COJO analysis (p<0.0001), except for associations between *VPS37D*-rs10245965 and TG concentrations (p=0.008). This suggests that associations between *VPS37D*-rs10245965 and TG concentrations are only observed when conditioning on other top SNPs in this region.

Sensitivity Analyses for Interactions in the UKB

In fully adjusted models in the UKB examining adjusted means and additionally accounting for other dietary factors, mean HDL-C concentrations were lowest among SSB consumers with the

CC genotype [mean HDL-C (95% CI): 54.6 (52.6, 56.6) mg/dl] and highest among SSB consumers with the TT genotype at rs71556729 [mean HDL-C (95% CI): 57.2 (54.2, 60.2) mg/dl] [β (SE) per additional minor allele: 1.3 (0.6) mg/dl, p_{trend}=0.03] (**Figure VIII**). A positive trend between dosage at rs71556729 and adjusted mean HDL-C was also observed among SSB non-consumers, but the mean difference in HDL-C by genotype was smaller and not statistically significant [β (SE) per additional minor allele: 0.6 (0.3) mg/dl, p_{trend}=0.06]. When limiting to UK Biobank participants fasting >6 hours (9% of all participants) prior to blood draw, difference test interaction results were similar for the top SNPs for TG concentrations and attenuated for HDL-C concentrations (Table S8).

III. Supplemental Discussion

Associations between SSB Consumption and HDL-C and TG in the CHARGE Cohorts

This study assessed the association between SSB consumption and lipid concentrations among adults from multiple geographic areas. Our results are consistent with several smaller studies that have observed higher SSB consumption is associated with lower HDL-C and higher of TG [21,24,53–56] concentrations. Evidence from a meta-analysis of randomized-controlled trials, most of relatively small size, indicate a similar positive association between high sugar consumption and TG concentrations, but also a small positive association between high sugar consumption and HDL-C concentrations [57]. This contrasts with the negative association between SSB consumption and HDL-C concentrations seen in this and other observational studies (5,10–13). Reasons for this discrepancy are unclear but may be due to differences in type of sugar exposure or could be due to residual confounding. Secondary analyses of the heterogeneity of the

association between SSB intake on these lipid traits indicated that the estimated effect size of SSB on HDL-C concentrations was larger among US-based cohorts than the Australian and European cohorts. The estimated effect size of SSB on TG concentrations was larger among cohorts with a mean age of 40-60 years than cohorts with a mean age either less than 40 years or greater than 60 years. While we cannot determine the source of these heterogeneous associations due to our study-design, this observation may be attributed to differences in the information elicited by the questionnaires, societal/generational norms, differences in SSB consumption patterns, and/or difference in the effect of SSB on lipid traits in these groups.

Associations between CHREBP SNPs and HDL-C and TG in the CHARGE Cohorts

Although many studies have reported a significant association between SNPs in the *CHREBP* region on HDL-C and TG concentrations [8–11], we have conducted SNP analyses adjusting for nearby SNPs to identify distinct signals within this region. In this study, we performed conditional analyses utilizing data from all SNPs in this region. Using this approach, we observed one distinct SNP significantly associated with HDL-C concentrations and five distinct SNPs significantly associated with TG concentrations in the *CHREBP* region. Three SNPs significantly associated with TG concentrations (rs42124, rs13240662, and rs10245965) were in low-to-moderate LD with top SNPs from previous GWAS. Another SNP was significantly associated with TG concentrations in this study and in a previous a gene-centric association analysis (rs13225660) [58] is in a high LD region that includes a SNP within the 3'UTR of *CHREBP*. We also replicated the top signal from previous GWAS for both HDL-C and TG concentrations (rs3812316), which is a missense SNP in a high LD region of *CHREBP*. In our analysis, the minor allele at rs71556736 (a SNP in high LD with the previously reported

rs3812316, R^2=0.94) was associated with lower TG concentrations and minor allele at rs1714750 (R^2 with rs3812316 = 0.75) was associated high HDL-C concentrations.

IV. Supplemental Tables

Table I. Descriptions and acknowledgements for participating CHARGE cohorts.

Cohort	Study Description and Acknowledgements	Relevant References
Atherosclerosis Risk In Communities (ARIC) Study **USA**	The ARIC study is a population-based cohort study designed to study new and established risk factors for atherosclerosis and community trends in coronary heart disease. In 1987-89, baseline data was collected on 15,792 adults, aged 45–64 y, living in four U.S. communities (Forsyth County, NC; Jackson, MS; northwest Minneapolis suburbs, MN; Washington County, MD). The baseline exam was conducted in 1987-89 and information was collected on African Americans, Caucasians, and a few adults of other ethnicities, aged 45–64 y. After providing informed consent, 15,792 adults were enrolled (8,710 women and 7,082 men). Up to 8,591, Caucasian adults with available DNA, valid dietary information, and consent to share genetic data were eligible for the current analysis The Atherosclerosis Risk In Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.	https://sites.cscc.unc.edu/aric 59
Cardiovascular Health Study (CHS) **USA**	The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults≥65 conducted across four field centers. The originally, predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons was enrolled in 1992-1993 for a total sample of 5,888. Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples. Genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai among CHS participants who consented to genetic testing and had DNA available using the Illumina 370CNV BeadChip system (for European ancestry participants, in 2007). The current analysis uses diet data collected on the first cohort in 1989-1990 and is restricted to individuals of European ancestry. CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease. This CHS research was supported by NHLBI contracts HHSN268201800001C, HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and R01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR0001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.	http://www.chs-nhlbi.org/ 60

Cohort	Study Description and Acknowledgements	Relevant References
Fenland Study (Fenland) **United Kingdom**	The Fenland Study is a population-based observational study in the East of England (Cambridgeshire), the United Kingdom. Participants born between 1950 and 1975 were recruited from general practice lists near and around Cambridge, Ely, and Wisbech. In total, 12,434 participants were enrolled between 2005 and 2015. Exclusion criteria of the Fenland study included pregnancy, physician-diagnosed diabetes, inability to walk unaided, psychosis, or terminal illness. Participants were excluded for this study if participants reported non-white or uncertain race/ethnic status; provided no data on habitual diet, HDL or trig, or genetic variants; reported energy intake <500 or >3500 kcal/day for women or <800 or >4200 kcal/day for men; or classified as outliers in the genetic variables for population admixture. The authors are grateful to all the Fenland Study volunteers for their time and to the general practitioners and practice staff for help with recruitment. The authors thank the Functional groups of the MRC Epidemiology Unit, including the Fenland Study Co- ordination team, the Field Epidemiology team, the data management team and the laboratory team. The authors also thank Kate Westgate and Stefanie Hollidge for assistance with physical activity data processing. Biochemical assays were performed by the National Institute for Health Research, Cambridge Biomedical Research Centre, Core Biochemistry Assay Laboratory and the Cambridge University Hospitals NHS Foundation Trust, Department of Clinical Biochemistry. The Fenland Study was funded by the Wellcome Trust and the Medical Research Council. Support from Medical Research Council programmes MC_UU_12015/1 and MC_UU_12015/5 is acknowledged. Extension of the GWAS analysis to cover the entire cohort (and metabolomics) were funded by the MRC Omics call (MC_PC_13046). NJW, NGF and AK acknowledge National Institute for Health Research (NIHR) Biomedical Research Centre Cambridge (IS-BRC-1215-20014); NJW is an NIHR Senior Investigator.	https://www.mrc-epid.cam.ac.uk/research/studies/fenland/
Framingham Heart Study (FHS) **USA**	The Framingham Offspring Study is a community-based longitudinal study designed to examine CVD risk in the offspring of the Original Cohort participants of the Framingham Heart Study and their spouses. In 1971, 5,124 individuals were enrolled; since then, the Offspring Cohort has been examined periodically. Between 1998 and 2001, during the 7th examination cycle, 3,539 adults, with a mean age of 61y, underwent a standardized medical history and physical examination. Beginning in 2002, 4,095 Gen III participants, who had at least one parent in the offspring cohort, were enrolled in the Framingham Heart Study. At the first cycle of the Gen III study, 4,095 individuals with a mean age of 40 y, underwent the standard clinic examination. For the present study both cohorts were combined for the analysis. A total of 5,577 adults with available DNA, valid dietary information, and consent to share genetic data were eligible for the current study. This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195, HHSN268201500001I and 75N92019D00031) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). Dr. McKeown is partially supported by the USDA Agricultural Research Service (agreement 58-1950-0-014). Drs Meigs and Dupuis are also supported by National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK078616. Dr. Merino is supported by NIH P30 DK40561 and the European Commission Horizon 2020 program (H2020-MSCA-IF-2015-703787). [representing authors: GP, ANP, JD, JM and JBM]	[61–63]

Cohort	Study Description and Acknowledgements	Relevant References
Multi-Ethnic Study of Atherosclerosis (MESA) **USA**	The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA researchers study a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84 (38 percent of the recruited participants are white, 28 percent African-American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent), as well as 2,128 additional individuals from 594 families recruited through MESA Family by utilizing the existing MESA framework, yielding 3,026 sibpairs divided between African Americans and Hispanic-Americans. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. The authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. MESA and the MESA SHARe projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources, Grant UL1RR033176. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. [representing authors: ACW, ATC, LS and SSR]	https://www.mesa-nhlbi.org/ [64]

Cohort	Study Description and Acknowledgements	Relevant References
Netherlands Epidemiology in Obesity (NEO) Study **The Netherlands**	The Netherlands Epidemiology of Obesity (NEO) study: The NEO was designed for extensive phenotyping to investigate pathways that lead to obesity-related diseases. The NEO study is a population-based, prospective cohort study that includes 6,671 individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. At baseline, information on demography, lifestyle, and medical history have been collected by questionnaires. In addition, samples of 24-h urine, fasting and postprandial blood plasma and serum, and DNA were collected. Genotyping was performed using the Illumina HumanCoreExome chip, which was subsequently imputed to the 1000 genome reference panel. Participants underwent an extensive physical examination, including anthropometry, electrocardiography, spirometry, and measurement of the carotid artery intima-media thickness by ultrasonography. In random subsamples of participants, magnetic resonance imaging of abdominal fat, pulse wave velocity of the aorta, heart, and brain, magnetic resonance spectroscopy of the liver, indirect calorimetry, dual energy X-ray absorptiometry, or accelerometry measurements were performed. The collection of data started in September 2008 and completed at the end of September 2012. Participants are currently being followed for the incidence of obesity-related diseases and mortality. The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).	[65]
UK Biobank **United Kingdom**	A prospective cohort study of approximately 500,000 participants between the ages of 40-69 years. We conducted a cross-sectional analysis among individuals with a single 24-hour recall and fasting blood sample during the initial assessment visit (2006-2010). The UK Biobank was supported by the Wellcome Trust, Medical Research Council, Department of Health, Scottish Government and Northwest Regional Development Agency. It also had funding from the Welsh Assembly Government and British Heart Foundation. The research was designed, conducted, analysed and interpreted by the authors entirely independently of the funding sources. We are grateful to the UK Biobank participants for their participation in the study. The UK Biobank received ethical approval from the North West Multi-centre Research Ethics Committee (REC reference: 12/NW/03820). All participants gave written informed consent before enrolment in the study, which was conducted in accordance with the principles of the Declaration of Helsinki. This research has been conducted using the UK Biobank Resource *under Application Number 3583.*	UK-Biobank. UK Biobank. Protocol for a large-scale prospective epidemiological resources. 2010 (Available at: https://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf) [66]

Cohort	Study Description and Acknowledgements	Relevant References
Western Australian Pregnancy Cohort Study (Raine) Study **Australia**	The Western Australian Pregnancy Cohort (Raine) Study is a prospective pregnancy cohort where 2900 were recruited from King Edward Memorial Hospital between 1989 and 1991. Data were collected throughout pregnancy and the children have been followed-up at ages 1, 2, 3, 5, 8, 10, 14, 17, 18, 20 and 22. Ethnics approval for this study was obtained from King Edward Memorial Hospital and Princess Margaret Hospital. Participants were consented to being involved in this study prior to each follow-up. The Raine Study acknowledges the National Health and Medical Research Council (NHMRC) for their long term contribution to funding the study over the last 29 years. Core Management of the Raine Study has been funded by the University of Western Australia (UWA), Curtin University, the UWA Faculty of Medicine, Dentistry and Health Sciences, the Raine Medical Research Foundation, the Telethon Kids Institute, the Women's and Infants Research Foundation, Edith Cowan University, Murdoch University, and the University of Notre Dame. This study was supported by the National Health and Medical Research Council of Australia [grant numbers 572613, 403981 and 003209] and the Canadian Institutes of Health Research [grant number MOP-82893]. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). All analytic work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.	[67]
Rotterdam Study (RS; RS-I and RS-II were combined prior to CHARGE meta-analysis) **The Netherlands**	The Rotterdam Study is a prospective population-based cohort study in Ommoord, a suburb of Rotterdam, designed to investigate the prevalence and incidence of and risk factors for chronic diseases in the elderly. The baseline exam of the first cohort (RS-I) was conducted between 1990 and 1993. A total of 7,983 adults, aged 55 years and over, participated in the study. In 2000, the study was extended with a second cohort (RS-II) of 3,011 participants who had moved into the area or who had become 55 years since the start of the study. For the current analysis, 3,859 adults were eligible (2,784 from RS-II and 1,075 from RS-II), as they had available data on DNA, dietary intake and outcome information, and consent to share genetic data.. The Rotterdam Study is supported by the Erasmus MC University Medical Center and Erasmus University Rotterdam; The Netherlands Organisation for Scientific Research (NWO); The Netherlands Organisation for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); The Netherlands Genomics Initiative (NGI); the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. The contribution of inhabitants, general practitioners and pharmacists of the Ommoord district to the Rotterdam Study is gratefully acknowledged. ETML, JCK-dJ, TV and OHF work in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.), Metagenics Inc. and AXA. Nestlé Nutrition (Nestec Ltd.), Metagenics Inc. and AXA had no role in design and conduct of the study, collection, management, analysis, and interpretation of the data; and preparation, review or approval of the manuscript.	[68]

Cohort	Study Description and Acknowledgements	Relevant References
Women's Genome Health Study (WGHS) **United States**	The Women's Genome Health Study (WGHS) is a prospective cohort of initially healthy, female North American health care professionals at least 45 years old at baseline representing participants in the Women's Health Study (WHS) who provided a blood sample at baseline and consent for blood-based analyses. The WHS was a 2x2 trial beginning in 1992-1994 of vitamin E and low dose aspirin in prevention of cancer and cardiovascular disease with about 10 years of follow-up. Since the end of the trial, follow-up has continued in observational mode. Additional information related to health and lifestyle were collected by questionnaire throughout the WHS trial and continuing observational follow-up. The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913), the Donald W. Reynolds Foundation, and the Fondation Leducq, with collaborative scientific support and funding for genotyping provided by Amgen. Dr. Mora was supported by the research grants from the National Heart, Lung, and Blood Institute (grants R01HL134811, R01117861, and K24 HL136852); National Institute of Diabetes and Digestive and Kidney Diseases (grant DK112940).	[69]
Women's Health Initiative (WHI) **United States**	WHIMS randomized trials which examined the effects of postmenopausal hormone therapy on cognitive function in women aged 65-80 years. Recruitment began in 1995. The WHI program is supported by contracts from the National Heart, Lung and Blood Institute, NIH. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A listing of WHI investigators can be found at http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator %20Short%20List.pdf. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. NT was supported by NIH HG006915, HL152215, R15006915.	[70]

Cohort	Study Description and Acknowledgements	Relevant References
Cardiovascular Risk in Young Finns Study (YFS) **Finland**	The Cardiovascular Risk in Young Finns (YFS) is a population-based 27 year follow up-study. The first cross-sectional survey was conducted in 1980, when 3,596 Caucasian subjects aged 3-18 years participated. In adulthood, the latest 27-year follow-up study was conducted in 2007 (ages 30-45 years) with 2,204 participants. The study cohort for the present analysis comprised subjects who had participated in the study in 2007 and had validated dietary data from FFQ, available genotype and other risk factor data (Raitakari OT et al. Cohort profile. Int. J Epidemiol. 2008;37:1220-6). The dietary intake of nutrients was assessed using a modified 131-item food frequency questionnaire developed by the Finnish National Institute for Health and Welfare (Paalanen et al. 2006). The study was approved by the local Ethical Committees and was performed according to Helsinki declaration.A total of 1,782 participants with available DNA and who provided complete dietary information were eligible for the current study. The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association**.** The Young Finns Study has been financially supported by the Academy of Finland: grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS and grant 848146 for To Aition); European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation and Finnish Society of Clinical Chemistry.	[71]

Table II. Dietary assessment methods of participating CHARGE cohorts.

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Nuts/ Seeds Intake	Whole Grain Intake	Fish Intake	% Energy Saturated Fat	Relevant References
ARIC	66-item, interviewer-administered, modified Willett FFQ	Harvard	Regular soda/ fruit-flavored drink, quantified as servings/d	Apples or pears/ bananas/ peaches, apricots, plums/ oranges/ other fruit	Broccoli/ string beans, green beans/ cauliflower, cabbage, Brussels sprouts/ spinach, collards, other greens/ corn/ tomatoes (potatoes not included)	peanuts, almonds, peanut butter, cashews, sesame seeds, walnuts	Cooked cereals/ whole grain cold cereal/ dark or whole grain bread	Canned tuna fish/ dark meat fish/ other light meat fish	As a percent of total energy, where 1 g saturated fat has 9 kcal.	[72,73]
CHS	99-item, self-administered, picture-sort version of National Cancer Institute FFQ	Harvard	Tang, Regular soft drinks	apples, applesauce, pears/bananas/peaches, apricots, nectarines (canned, frozen, whole year/in season), cantaloupe/watermelon/strawberries/oranges/grapefruit/any other fruit.	string beans, green beans/peas/corn/winter squash, baked squash/tomatoes, tomato juice/broccoli/cauliflower, brussel sprouts/spinach (raw/cooked)/mustard greens, turnip greens, collards/cole slaw, cabbage, sauerkraut/carrots, mixed vegetables containing carrot/green salad/sweet potatoes, yams/other veg including onions, summer squash	peanuts	dark bread (whole wheat, rye, pumpernickel)/ cooked cereals/high fiber, bran or granola cereals.	tuna fish, tuna salad, tuna casserole/other fish, broiled, baked	As a percent of total energy	[74]

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Nuts/ Seeds Intake	Whole Grain Intake	Fish Intake	% Energy Saturated Fat	Relevant References
Fenland	160 -item, self-administered, modified EPIC-Norfolk FFQ	FETA	Fizzy soft drinks, eg. Coca cola, lemonade (glass) + Fruit squash or cordial (glass)	apples/pears/oranges satsumas mandarins/grapefruits/bananas/grapes/melon/peaches plums apricots/strawberries raspberries kiwi fruit/tinned fruit/dried fruit eg raisins prunes	carrots/spinach/broccoli spring greens kale/brussels sprouts/cabbage/peas/green beans broad beans runner beans/marrow courgettes/cauliflower/parsnips turnips swedes/leeks/onions/garlic/mushrooms/sweet peppers/beansprouts/green salad lettuce cucumber celery/watercress/tomatoes/sweetcorn/beetroot/coleslaw/avocado/baked beans/driedLnetils/tofu	peanuts/ peanut butter	Brown bread and rolls/wholemeal bread and rolls/brown rice/wholemeal pasta/porridge/breakfast creals	fried fish in batter as in fish and chips/fish fingers fish cakes/other white fish fresh or rozen eg. Cod haddock plaice sole halibut/oily fish fresh or canned eg mackerel kippers tuna salmon sardines herring/shellfish eg crab prawns mussels/fish roe taramasalata	As a percent of total energy, where 1 g saturated fat has 9 kcal.	[75,76]
FHS	126-item, semi-quantitative FFQ	USDA	Caffeinated colas, caffeine-free colas, carbonated noncola soft drinks, and noncarbonated sugar-sweetened beverages (fruit punches, lemonades, or other fruit drinks).	Raisins, prunes, bananas, cantaloupe, watermelon, apples, pears, oranges, grapefruit, fruit juices, strawberries, blueberries, peaches	Tomatoes, tomato juice, red chili sauce, tofu, soybeans, string beans, broccoli, cabbage, coleslaw, cauliflower, Brussels sprouts, carrots, corn, peas, lima beans, mixed vegetables, beans/lentils, squashes, potatoes, spinach, kale/mustard/chard, lettuce, celery, beets, alfalfa, garlic	nuts, peanut butter	Quantified as g/day (ready to eat breakfast cereal, cooked oatmeal, dark bread, brown rice, other grains (i.e., bulgur, kasha, couscous), popcorn, bran, wheat germ, other (fill-in) whole grain foods)	Canned tuna fish; dark meat fish; other fish; shrimp, lobster, scallops as main dish	As a percent of total energy, where 1 g saturated fat has 9 kcal.	[77]

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Nuts/ Seeds Intake	Whole Grain Intake	Fish Intake	% Energy Saturated Fat	Relevant References
MESA	120-item, self-administered, modified-Block FFQ	NDSI	Coke, Pepsi, 7-up or other carbonated (not diet) + other juice (not fruit) quantified as servings / d	Peaches, apricots, nectarines, plums / cantaloupe mango, papaya / strawberries, blueberries, other berries / apples, applesauce, pears / bananas, plantains / oranges, grapefruit, tangerines, kiwi / any other fruit	Tomato / cruciferous veg / yellow veg / green leafy veg / other veg	serv/day	Whole grain breakfast cereal; oatmeal; dark bread; bran muffins; brown or wild rice	Weighted amounts defined from Tuna, salmon, sardines; including sashimi or sushi / Other broiled, steamed, baked or raw fish– trout, sole, halibut, poke, grouper /tuna / boiled fish	As a percent of total energy, where 1 g saturated fat has 9 kcal	[78]
NEO	semi-quantitative food frequency questionnaire, originally validated in the Dutch general population	Dutch food composition table (NEVO) (version 2011)	regular soda/ fruit-flavored drink, quantified as servings/d	apples or pears/ bananas/ peaches, apricots, plums/ oranges/ other fruit	Cooked, baked and raw vegetables	peanuts, almonds, peanut butter, cashews, sesame seeds, walnuts	N/A	canned tuna fish/ dark meat fish/ other light meat fish		[79]
Raine Study	self-administered, 74-item FFQ designed to measure dietary intake over the past 12 months	Australian NUTTAB 1995 food composition database (Cancer Council Victoria Australia)	Soft drink, regular energy drinks	oranges, apples, pears, bananas, melon, pineapples, strawberries, apricots, peaches, mangoes, avocado	tomatoes, pumpkin, zucchini, cucumber, capsicum, lettuce, celery, beetroot, carrots, cabbage, cauliflower, broccoli, spinach, peas, green beans, bean sprouts, onion, garlic	mixed nut, peanut butter	Ryebread, multigrain bread, All bran, bran flakes, porridge, muesli.	grilled fish, fried fish, tinned fish	As a percent of total energy, where 1 g saturated fat has 9 kcal	

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Nuts/ Seeds Intake	Whole Grain Intake	Fish Intake	% Energy Saturated Fat	Relevant References
RS-I	170-item semi-quantitative food frequency questionnaire	Dutch Food Composition Table 1993 (NEVO 1993).	Soft drink	Apple without skin / Strawberries / Banana / Grapes white/black with skin / Grapefruit / Tangerines / Pear without skin / Plums with skin / Orange / Apple sauce canned / Fruit cocktail in syrup canned / Kiwi fruit	Endive boiled / Cauliflower boiled / Kale curly boiled / Mushrooms boiled / Cabbage green boiled / Cucumber raw / Cucumber boiled / Leek boiled / Rhubarb raw / Cabbage red boiled / Cabbage Savoy boiled / Lettuce raw / Spinach boiled / Cabbage oxhair boiled / Brussel sprouts boiled / Tomatoes raw / Onions boiled / Chicory raw / Chicory boiled / Cabbage white cooked / Carrots raw / Carrots boiled / Cabbage sauerkraut cooked / Vegetables mixture raw / Gherkins sweet pickled / Sweet pepper red boiled without salt / Broccoli boiled / Beetroot boiled	serv/day	Wheat germ / Bread brown wheat / Rye bread dark / Rye bread light / Bread brown wholemeal / Porridge oatmeal / Wheat bran / Muesli without sugar / Rice brown boiled	Herring salted / Mackerel raw / Sardines/pilchards in oil canned / Salmon canned / Eel smoked / Plaice raw / Fish fingers uncooked / Cod raw / Herring marinated	As a percent of total energy, where 1 g saturated fat has 9 kcal	[80,81]
RS-II	389-item semi-quantitative food frequency questionnaire	Dutch Food Composition Table 2011 (NEVO 2011).	Soft drinks	Apple / Strawberries / Banana / Grapes / Grapefruit / Tangerines / Pear / Plums / Orange / Kiwi fruit / Other fruit	Cauliflower / Kale / Cabbage / Cucumber / Leek / Lettuce / Spinach / Brussel sprouts / Tomatoes / Onions / Broccoli / Peppers / Carrots / Other vegetables	serv/day	Whole-wheat crackers / Wheat germ / Wheat bread / Rye bread / Whole-wheat bread / Oatmeal / Whole-wheat cereal / Muesli without sugar / Whole-grain rice	Herring / Mackerel / Sardines/pilchards in oil canned / Salmon / Eel / Trout / Flounder / Fish fingers / Other fish	As a percent of total energy, where 1 g saturated fat has 9 kcal.	[82,83]

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Nuts/ Seeds Intake	Whole Grain Intake	Fish Intake	% Energy Saturated Fat	Relevant References
WGHS	126-item, semi-quantitative FFQ	Harvard	Caffeinated colas, caffeine-free colas, carbonated noncola soft drinks, and noncarbonated sugar-sweetened beverages (fruit punches, lemonades, or other fruit drinks).	apples, apple juice, avocados, bananas, blueberries, cantaloupes, grapefruit, grapefruit juice, oranges, orange juice, peaches, prunes, raisins, strawberries, other juice	beans, beets, broccoli, brussels sprouts, cabbage, cooked & raw carrots, cauliflower, celery, red chili sauce, corn, eggplant, green pepper, iceberg lettuce, kale, mixed vegetables, cooked & raw onions, dark orange squash, peas, romaine lettuce, cooked & raw spinach, string beans, tofu, tomatoes, tomato juice, tomato sauce, yams	peanuts, peanut butter, other nuts	whole grain cold cereal, dark bread, popcorn, oatmeal, wheat germ, brown rice, oat bran, other grain	dark meat fish, canned tuna fish, other fish, shrimp, lobster, scallops as main dish	As a percent of total energy, where 1 g saturated fat has 9 kcal.	[77]
WHI	126-item, semi-quantitative FFQ		Tang/Kool-Aid/Hi-C/other fruit drinks,Regular soft drinks (not diet)	serv/day	serv/day	serv/day	serv/day	serv/day	As a percent of total energy, where 1 g saturated fat has 9 kcal.	
YFS	131-item FFQ	Finnish food composition database Fineli, maintained by the Nutrition Unit, National Institute of Health and Welfare, Finland.	Sugar-sweetened soda/ sugar-sweetened cola drinks/ sugar-sweetened fruit or berry drinks	Grams per day (approximate conversions used for presentation in Table 1)	Grams per day (approximate conversions used for presentation in Table 1)	Grams per day (approximate conversions used for presentation in Table 1)	Servings per day	Grams per day (approximate conversions from CSFII Pyramid Servings Database used for presentation in Table 1)		[84]

Table III. Dietary assessment methods of UK Biobank cohort.

Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Energy	Saturated Fat	Polyunsaturated fat	Dietary fibre	Magnesium	Potassium	Relevant References
314-item, Oxford WebQ, web-based 24-hour dietary recall, administered at Assessment visit	UK Food composition database McCance and Widdowson's The Composition of Foods	Squash and fizzy drinks, quantified as glasses/cans consumed yesterday	Estimated intake of total energy, based on food and beverage consumption yesterday, excluding supplementation.	Estimated intake based on food and beverage consumption yesterday, excluding supplementation.	Estimated intake based on food and beverage consumption yesterday, excluding supplementation.	Non-starch polysaccharide, estimated intake based on food and beverage consumption yesterday, excluding supplementation.	Estimated intake based on food and beverage consumption yesterday, excluding supplementation.	Estimated intake based on food and beverage consumption yesterday, excluding supplementation.	.[85,86]

Table IV. Genotyping information of participating cohorts.

Cohort	Genotyping Array & Calling Algorithm	Sample QC	SNP Inclusion Criteria	Phasing & Imputation	Correction for Ancestry/ Relatedness
ARIC	Affymetrix 6.0	>95% call rate	≥ 90% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	IMPUTE2; HRC r1.1 2016 on the Michigan imupation server	Principal Components
CHS	Illumina CNV370; Illumina BeadStudio	>95% call rate; Individuals additionally excluded for: Presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack; lack of available DNA; genotype discordant with known sex or prior genotyping.	≥ 97% call rate; Hardy-Weinberg equilibrium $\geq 10^{-5}$, ≤2 duplicate errors or Mendelian inconsistencies, heterozygote frequency ≠0	SHAPE-IT; HRC r1.1 2016 on the Michigan imputation server	None
Fenland	Affymetrix Axiom UKBiobank	≥95% call rate; >0.19 and <0.21 calculated on SNPs with MAF > 1%, >0.004 and <0.0125 calculated on SNPs with MAF < 1%; Channel contrast (DishQC <0.82), sex discrepancy, unusually high number of singleton genotypes, impossible IBD values	>0% MAF; ≥ 95% call rate; Hardy-Weinberg equilibrium $<10^{-6}$		
	Illumina Metabochip	≥95% call rate (for including individuals) and sex mismatch	>0% MAF; ≥ 90% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	HRC + UK10K + 1000G phase 3 (IMPUTE, http://mathgen.stats.ox.ac.uk/impute/impute.html)	Kinship matrix
	Infinium Core Exome 24 v1	≥98% call rate; Sex call based on X heterozygosity consistent with self-reported sex; ethnically aware rare variant heterozogosity (SNPs with MAF <=1%) Heterozygosity > 0.005; Ethnically aware common variant heterozogosity (SNPs with MAF >1%) Heterozygosit >0.363 or <0.336; Ethnically aware rare allele count outliers: > 60 singleton calls, >60 doubleton calls or > 60 tripleton calls; Ethnic outliers with inconsistent heterozygosity patterns; Duplicates: removed sample with lower call rate where both passed	≥ 95% call rate; Hardy-Weinberg equilibrium $<10^{-6}$;		
FHS	Affymetrix 500K (250K Nsp and 250K Sty) and MIPS 50K	Sample call rate >95%; Number of Mendelian errors greater or equal to 1000	≥0.01% MAF; ≥ 90% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	ShapeIT v2.r837; HRC r1.1 2016 on the Michigan imputation server	Kinship Matrix
MESA	Affymetrix Genome-Wide Human 6.0 array	Sample call rate >95%;	>97% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	HRC r1.1 2016: Michigan Imputation server	Principal Components

NEO	Illumina HumanCoreExome-24v1; Illumina GenomeStudio	1) call rate <98%; 2) principal component analysis outliers; 3) sex mismatch; 4) heterozygosity +/-3SD from mean of ancestry distribution; 5) first degree relatedness;	>0% MAF; ≥% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	Impute version 2.2 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/)	Principal Components
Raine Study	Illumina Human660W Quad BeadChip; GenomeStudio	1) call rate >97%; 2) duplicates and plate controls; 3) gender discrepency when compared to core data; 4) cryptic relatedness (IBD), pi > 0.1875, 4) heterozygosity, h>0.32	≥ 95% if MAF ≥ 5%, ≥ 99% if MAF < 5%; Hardy-Weinberg equilibrium $<10^{-6}$	ShapeIT v2.r790; 1000G Phase 3 v5, HRC r1.1 2016: Michigan Imputation server	Principal Components
RS	Array type Illumina 550K	≥ 97.5% call rate	≥ 95% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	SHAPEIT2; HRC r1.1 2016 on the Michigan imputation server	None
UKB	Affymetrix UK BiLEVE Axiom array on an initial 50,000 participants; the remaining 450,000 participants were genotyped using the Affymetrix UK Biobank Axiom® array	SNPs exclude for non-white participants; significant differences between batches ; heterozygosity=0; aneuploidy=1; inferred gender different from submitted gender; >95% call rate	>95% call rate; min MAF=0.01; min MAC=20	http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015-1.pdf https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/genotyping_qc.pdf [87]	None
WGHS	Illumina HumanHap300 Duo & HumanHap300 Duo and iSelect chips with the Infinium II protocol; Illumina BeadStudio	Sample call rate >95%	≥ 1% MAF; ≥ 90% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	HRC r1.1 2016 on the Michigan imputation server	Principal Components
WHI	1000 genomes; Illumina, Affymetrix, HumanOmni	Sample call rate >95%	>0.5% MAF; ≥ 90% call rate; Hardy-Weinberg equilibrium $<10^{-4}$	Beagle/MACH	None
YFS	Illumina 670K; Illuminus clustering algorithm	1) call rate ≥95%; Sanger genotyping pipeline QC: heterozygosity, call rate, duplicate/related and Sequenom fingerprint discrepancy. Sex status match genotyping sex. Cryptic relatedness: pi-hat<0.2	≥ 1% MAF; ≥ 95% call rate; Hardy-Weinberg equilibrium $<10^{-6}$	SHAPEIT v1; IMPUTE v2.2.2	Principal Components

Table V. Assessment of additional characteristics of participating cohorts.

Cohort	Education	Smoking Status	Physical Activity	Alcohol Intake	Relevant References
ARIC	Categorized into 6 groups: grade school or none, some high school, high school graduate, vocational school, college, graduate/ professional school	Classified as current, former, never smoker or missing/ unknown; current smokers given in descriptive	Assessed as both sport and leisure time using the Baecke questionnaire. A sports activity score and a leisure activity score ranged from low to high (five quintile categories).	grams/day	
CHS	Classified as having completed high school or GED based on self-report	Classified as current, former, never smoker by self report	Assessed as kcal expended in physical activities excluding household chores based on self-report of specific activities, used as a continuous variable.	drinks/day converted to grams/day by multiplying by 14.	
Fenland	Participants were asked to confirm 12 types of British educational certificates. If they had a school leaving certificate or a GCE AS level, AS level or higher, they were classified as having an educatino level of high school or above.	Classified as current, former, and never smoker, according to questions about current smoking and smoking history	Physical activity was objectively assessed over 6 days using a combined heart rate and movement sensor (Actiheart, CamNTech, Cambridge, UK), with individual calibration of heart rate performed using a treadmill test. Data from free-living was pre-processed and modelled using a branched equation framework to estimate intensity time-series, which were summarised over time as daily Physical Activity Energy Expendi- ture (PAEE) (kJ/kg/d). Then, kj/d was computed by PAEE times a body weight for each individual.	grams/day	[88]
FHS	Categorized into 4 groups: less than high school, graduated high school, some college, or graduated college	Regular cigarette smoking in the last year: classified as yes (current) or no (never)	Physical activity score taking into account typical number of hours sleeping, sitting, slight activity, moderate activity, and heavy activity	grams/day	[89]
MESA	Measured as 'highest level achieved' for 9 categories: no school, grades 1-8, grades 9-11, completed high school / GED, some college but no degree, technical school certificate, associate's degree, Bachelor's degree, graduate school	Classified as current, former, never smoker or missing/ unknown	Sport and leisure time as METs / week	grams/day	
NEO	Classified as having completed high school based on self-report	Classified as current, former, never smoker or missing/ unknown	Assessed using the SQUASH questionnaire. For the current analyses, MET-hours were calculated for sports/leisure activities		

Cohort	Education	Smoking Status	Physical Activity	Alcohol Intake	Relevant References
Raine Study	Classified as having completed high school based on self-report	Classified as current, never smoker or missing/ unknown	Self-reported using 'The International Physical Activity Questionnaire'. METMINperWeek calculated for activities of different intensities - Vigorous, moderate, walking.	Expressed as precentage of total daily energy intake	
RS	2 Categories: 1) primary education, lower vocational training or lower general education, 2) intermediate vocational training, higher general education, higher vocational training, college, or university.	Classified as current, former, never smoker	NA	grams/day	
UKB	Participants asked to choose from 6 types of British qualifications. If they had a school leaving certificate or a GCE AS level, AS level or higher, they were classified as having an education level of high school or above.	Classified by self-report as prefer not to answer, never, previous, current, or NA	Classified as low, moderate, or high. Assessed by derived MET (Metabolic Equivalent Task) scores based on IPAQ (International Physical Activity Questionnaire) guidelines	Grams/d	
WGHS	Categorized into three groups: less than Bachelor's degree, Bachelor's degree, or graduate degree	three category smoking status variable (never/past/current)	Total energy expended from recreational physical activity (MET-hours/week)	grams/day	
WHI	Classified as having completed high school based on self-report	three category smoking status variable (never/past/current)	Total energy expended from recreational physical activity (MET-hours/week)	alcohol from beer, wine, and liquor, as well as from foods (some foods contain alcohol due to minute amounts of alcohol in vanilla extract, almond extract etc used in baking)	
YFS	Continuous, years of total education	Classified as current, former, never smoker	Continuous: MET-exercise index, which refers to energy consumption of 1 kcal per hour, per 1kg of person weight.	grams/day	

Table VI. General characteristics of participants included from the UK Biobank[a]

	Overall	SSB non-consumers	SSB consumers
n	59,220	46,784	12,436
SSB Consumption (drinks/day)	0.29	0	1.43
Age (years)	56.3 (8.2)	56.6 (8.0)	54.9 (8.5)
Women	55	56	49
Body Mass Index (kg/m^2)	27.1 (4.7)	27.0 (4.7)	27.6 (4.8)
Triglycerides (mg/dl)	146 (77)	144 (76)	155 (82)
Mean TG (ln-mg/dl)	4.86 (0.49)	4.85 (0.49)	4.91 (0.50)
Mean HDL-C (mg/dl)	57 (15)	58 (15)	55 (14)
Completed high school	89	89	89
Smoking Status			
Never	56	56	57
Former	35	35	33
Current	8	8	9
Prefer not to answer	1	1	1
Physical Activity Level (IPAQ)			
High	35	35	36
Moderate	35	35	33
Low	15	14	15
Not Available	15	16	16
Energy Intake (kcal)	2083 (673)	2040 (657)	2246 (709)
Alcohol Intake (g)	15.8 (23.9)	16.4 (24.2)	13.6 (22.4)
Dietary Fiber Intake (g)	16.6 (7.6)	16.7 (7.6)	16.1 (7.6)
Polyunsaturated Fat Intake (g)	14.1 (8.5)	13.9 (8.3)	15.0 (8.8)
Saturated Fat Intake (g)	29.0 (14.3)	28.5 (14.1)	31.1 (14.9)
Potassium Intake	3723 (1,323)	3721 (1308)	3727 (1381)

Table VII. Top signals of difference test interaction between SSB consumptions and *CHREBP* SNPs on lipid traits for SSB consumers vs non-consumers in the UKB.[a]

SNP	Location (Hg19)	Nearest gene	Alleles (E/A)[b]	Minor Allele Frequency	P_{Diff}/ Model[c]	SSB Intake Category	*n*	Effect Size (SE)	*P-value*
HDL-C (mg/dl)									
rs35709627	72999171	*TBL2*	A/G	0.05	0.01/	non-consumers	46,261	0.23 (0.22)	0.30
					Model 1	consumers	12,051	1.13 (0.40)	0.004
					0.01/	non-consumers	45,866	0.19 (0.20)	0.34
					Model 2	consumers	11,928	1.03 (0.36)	0.005
rs71556729	72989516	*TBL2*	T/C	0.03	0.001/	non-consumers	46,261	0.07 (0.29)	0.82
					Model 1	consumers	12,051	1.60 (0.53)	0.002
					0.0004/	non-consumers	45,866	0.06 (0.26)	0.82
					Model 2	consumers	11,928	1.65 (0.49)	0.0008
rs71556736	73034929	*CHREBP*	T/C	0.13	0.2/	non-consumers	46,261	0.44 (0.14)	0.001
					Model 1	consumers	12,051	0.73 (0.25)	0.003
					0.2/	non-consumers	45,866	0.46 (0.12)	0.0002
					Model 2	consumers	11,928	0.70 (0.23)	0.002
rs73137017	72974413	*BCL7BI*	G/A	0.04	0.2/	non-consumers	46,261	-0.005 (0.24)	0.98
					Model 1	consumers	12,051	-0.50 (0.44)	0.27
					0.4/	non-consumers	45,866	-0.07 (0.22)	0.74
					Model 2	consumers	11,928	-0.36 (0.40)	0.36
rs79578725	73002455	*CHREBP*	A/G	0.05	0.4/	non-consumers	46,261	-0.35 (0.24)	0.14
					Model 1	consumers	12,051	-0.04 (0.43)	0.93
					0.7/	non-consumers	45,866	-0.25 (0.21)	0.25
					Model 2	consumers	11,928	-0.11 (0.40)	0.79
TG (mg/dl)									
rs799157	73020301	*CHREBP*	T/C	0.04	0.2/	non-consumers	46,261	0.03 (0.008)	0.003
					Model 1	consumers	12,051	0.05 (0.02)	0.008
					0.4/	non-consumers	46,261	0.04 (0.008)	2.6E-06
					Model 2	consumers	12,051	0.05 (0.02)	0.002
rs55673514	73021456	*CHREBP*	G/A	0.03	0.003/	non-consumers	46,261	0.004 (0.01)	0.69
					Model 1	consumers	12,051	0.05 (0.02)	0.004
					0.003/	non-consumers	46,261	0.006 (0.009)	0.49
					Model 2	consumers	12,051	0.06 (0.02)	0.002

[a] Regression coefficients are shown as β (SE). β represents the direction and magnitude of change in the outcome trait with each additional effect allele. [b] Alleles presented as effect (E)/alternative (A) alleles. [c] Regression coefficient in linear regression models used an additive genetic effect and adjusted for age (years), sex (male/female), IPAQ (low, medium, high), and smoking status (never, previous, current), energy, alcohol, BMI, and PCs. [d] Linkage Disequilibrium (R^2) with rs71556729 (Table 3) = 0.94 in European ancestry groups of Phase 3 (Version 5) of the 1000 Genomes Project

Table VIII. Difference test interaction between SSB consumption and top SNPs for TG (rs799157) and HDL-C (rs71556729) for SSB consumers vs non-consumers and by fasting status in the UKB.[a]

Associations between rs71556729 and HDL-C (mg/dl)						Associations between rs799157 and TG (ln-mg/dl)					
Fasting Status	***P*$_{Diff}$**	**SSB Intake Category**	***n***	**Effect Size (SE)**	***P-value***	**Fasting Status**	***P*$_{Diff}$**	**SSB Intake Category**	***n***	**Effect Size (SE)**	***P-value***
>6 hours	0.95	non-consumers	3,860	1.34 (0.93)	0.15	>6 hours	0.008	non-consumers	3,860	-0.01 (0.03)	0.75
		consumers	1,103	1.23 (1.87)	0.51			consumers	1,103	0.14 (0.06)	0.02
4-6 hours	0.12	non-consumers	17,901	0.40 (0.43)	0.34	4-6 hours	0.0006	non-consumers	17,901	0.001 (0.01)	0.97
		consumers	4,609	1.51 (0.79)	0.05			consumers	4,609	0.09 (0.03)	0.002
<4 hours	0.002	non-consumers	24,103	-0.10 (0.35)	0.77	<4 hours	0.88	non-consumers	24,103	0.01 (0.01)	0.29
		consumers	6,213	1.76 (0.67)	0.008			consumers	6,213	0.02 (0.02)	0.52

[a] Regression coefficients are shown as β (SE). β represents the direction and magnitude of change in the outcome trait with each additional effect allele. Regression coefficient in linear regression models used an additive genetic effect and adjusted for age (years), sex (male/female), IPAQ (low, medium, high), and smoking status (never, previous, current), energy, alcohol, BMI, and PCs. Fasting status represents hours fasting prior to blood draw.

Table IX. Meta-analysis of the associations between SSB intake and lipid traits.[a]

		<1 serv/mo	1-4 serv/mo	1-2 serv/wk	3-7 serv/wk	>1 serv/day		
	n		*β2* (SE)	*β3* (SE)	*β4* (SE)	*β5* (SE)	p_{trend}	I^2
HDL-C (mg/dl)								
Model 1[b]	63,527	Ref.	-0.9 (0.2)	-2.1 (0.3)	-2.7 (0.2)	-4.3 (0.5)	<0.0001	49%
Model 2[c]	61,168	Ref.	-0.5 (0.2)	-1.7 (0.3)	-2.0 (0.2)	-3.1 (0.4)	<0.0001	31%
Model 3[d]	60,991	Ref.	-0.9 (0.1)	-1.8 (0.3)	-2.1 (0.3)	-3.0 (0.5)	<0.0001	47%
Model 4[e]	60,908	Ref.	-0.8 (0.1)	-1.6 (0.2)	-1.8 (0.2)	-2.1 (0.4)	<0.0001	40%
TG (ln-mg/dl)								
Model 1[b]	61,879	Ref.	0.01 (0.01)	0.04 (0.01)	0.06 (0.01)	0.11 (0.02)	<0.0001	41%
Model 2[c]	60,716	Ref.	0.01 (0.01)	0.03 (0.01)	0.05 (0.01)	0.10 (0.02)	<0.0001	45%
Model 3[d]	60,539	Ref.	0.02 (0.01)	0.04 (0.01)	0.06 (0.01)	0.09 (0.02)	<0.0001	48%
Model 4[e]	60,455	Ref.	0.02 (0.01)	0.03 (0.01)	0.05 (0.01)	0.06 (0.02)	<0.0001	38%

[a] Regression coefficients are shown as *βs* (SEs). *β2, β3, β4,* and *β5* represent beta-coefficients for 1-4 servings SSB/month (n=12,581), 1-2 servings SSB/week (n=7,185), 3-7 servings SSB/week (n=10,725), and >1 servings SSB/day (n=4,685) compared to <1 serving SSB/month (reference; n=28,356), respectively. I^2 represents heterogeneity in multivariate meta-analysis of categorical SSB intake on lipid traits, and p_{trend} represents statistical significance in univariate meta-analysis when SSB intake was assessed using an additive model (per increase in category).

[b] Model 1: adjusted for age, sex, total energy intake and study site for multi-centered cohorts (CHS, FHS, YFS, Fenland, RS, MESA).

[c] Model 2: adjusted for Model 1 covariates and smoking status, education status, physical activity, and alcohol intake.

[d] Model 3: adjusted for Model 2 covariates and BMI.

[e] Model 4: adjusted for Model 3 covariates and fruit intake, vegetable intake, whole grains intake (except in YFS), fish intake, and saturated fatty acids (% of total energy).

Table X. Meta-analysis of associations between SSB intake (per category) and lipid traits stratified by potential moderators*

Stratification criteria		HDL-C (mg/dl)				TG (ln-mg/dl)			
		β (95% CI)	$p_{interaction}$	Direction	I^2	β (95% CI)	$p_{interaction}$	Direction	I^2
Region	Australia	-0.41 (-1.18, 0.37)	**<0.0001**	-	0%	0.001 (-0.03, 0.03)	0.30	+	0%
	USA	-0.77 (-0.89, -0.65)		------	0%	0.02 (0.01, 0.03)		++++++	65%
	Europe	-0.25 (-0.41, -0.09)		----	0%	0.01 (0.002, 0.02)		++++	0%
Age	< 40 yrs	-0.28 (-0.83, -0.26)	0.50	--	0 %	0.01 (-0.01, 0.02)	**0.004**	++	0 %
	40-60 yrs	-0.64 (-0.91, -0.37)		-----	88 %	0.02 (0.02, 0.03)		+++++	55 %
	> 60 yrs	-0.52 (-0.88, -0.15)		----	0 %	0.01 (0.001, 0.01)		++++	0%
BMI	< 27 kg/m^2	-0.46 (-0.72, -0.21)	0.25	------	80%	0.01 (0.01,0.02)	0.10	++++++	61 %
	≥ 27 kg/m^2	-0.71 (-1.04, -0.38)		-----	0 %	0.02 (0.01, 0.03)		+++++	44 %
SSB Intake	< 0.4 serv/d	-0.49 (-0.74, -0.24)	0.36	-------	76 %	0.01 (0.01, 0.02)	0.12	+++++++	65 %
	≥ 0.4 serv/d	-0.70 (-1.06, -0.33)		----	0 %	0.02 (0.01, 0.03)		++++	52 %
Study Date	≤ 2005	-0.65 (-0.88, -0.42)	0.24	------	29 %	0.01 (0.01, 0.02)	0.51	++++++	67 %
	> 2005	-0.43 (-0.71, -0.15)		-----	77 %	0.02 (0.01, 0.03)		+++++	62 %
Smoking	< 20 %	-0.60 (-0.86, -0.34)	0.54	--------	76 %	0.02 (0.01, 0.02)	0.51	++++++++	64 %
	≥ 20 %	-0.46 (-0.83, -0.09)		---	51 %	0.01 (0.002, 0.02)		+++	48 %
Sample Size	≤ 3,500	-0.46 (-0.81, -0.12)	0.52	-----	0 %	0.01 (-0.003, 0.02)	0.05	+++++	0 %
	> 3,500	-0.60 (-0.85, -0.35)		------	85 %	0.02 (0.01, 0.02)		++++++	55 %

* Regression coefficients represent beta-coefficients for each increase in category of SSB intake [<1 serving SSB/month (reference; n=28,356), 1-4 servings SSB/month (n=12,581), 1-2 servings SSB/week (n=7,185), 3-7 servings SSB/week (n=10,725), and >1 servings SSB/day (n=4,685) in meta-analysis. Regression models were adjusted for age, sex, total energy intake, study site for multi-centered cohorts, smoking status, education status, physical activity, alcohol intake, BMI, fruit intake, vegetable intake, whole grains intake (except in YFS), fish intake, and saturated fatty acids (% of total energy). Stratification criteria are based on mean values in cohorts, and $P_{interaction}$ represents statistical significance of the stratification criteria as a moderator in meta-regression

Table XI. Results from meta-analysis and joint association analysis of *CHREBP* SNPs on lipid traits in CHARGE consortium cohorts.[a]

SNP	Basepair (hg19; chr7)	Nearest gene	Nearest Reported SNP	R^2 with Reported SNP[b]	Alleles (E/A)[c]	Minor Allele Frequency	Effect Size (SE)[d]	*p*	Direction[e]	I^2	Joint Effect Size (SE)[f]	Joint *p*
HDL-C (mg/dl)												
rs17145750[g]	73026378	*CHREBP*	rs3812316[i]	0.75	T/C	0.16	0.85 (0.12)	4.16E-12	+++++?+++++	0%	0.85 (0.12)	4.16E-12
TG (ln-mg/dl)												
rs71556736[g]	73034929	*CHREBP*	rs3812316[i]	0.94	T/C	0.13	-0.07 (0.004)	6.20E-61	----------	62%	-0.04 (0.007)	1.40E-08
rs13225660	73006388	*CHREBP*	rs1051921[h]	1.00	T/C	0.19	-0.06 (0.004)	4.18E-57	----------	69%	-0.03 (0.005)	2.16E-07
rs42124	72832340	*FZD9*	rs799160[i]	0.02	A/G	0.04	0.05 (0.01)	4.19E-06	+-+++++++++	14%	0.05 (0.01)	7.71E-06
rs13240662	73032823	*CHREBP*	rs17145738[j]	0.40	A/G	0.05	-0.08 (0.007)	2.52E-34	---------+-	24%	-0.04 (0.008)	1.23E-05
rs10245965	73063515	*VPS37D*	rs799160[g]	0.45	T/C	0.36	0.01 (0.003)	0.008	++++++++++-	72%	0.01 (0.004)	3.00E-05

[a] Regression coefficients are shown as β (SE). β represents the direction and magnitude of change in the outcome trait with each additional effect allele.
[b] Population genotype data from European ancestry groups in Phase 3 (Version 5) of the 1000 Genomes Project.
[c] Alleles presented as effect (E)/alternative (A) alleles
[d] Regression coefficient in linear regression models used an additive genetic effect and adjusted for age (years), sex (male/female) and cohort/field center (CHS, FHS, YFS, Fenland, RS, MESA); where applicable) while accounting for family, or population structure (where applicable: CHS, FHS, YFS, Fenland, RS, MESA, WGHS).
[e] Order of cohorts for regression coefficient directions: FHS, YFS, Fenland, CHS, NEO, RS, MESA, WHI, ARIC, Raine Study, WGHS (+, positive effect size; -, negative effect size; ?, SNP not available in cohort).
[f] Approximate regression coefficients in joint model of all top signals on lipid outcome
[g] Linkage disequilibrium (R^2) between rs17145750 and rs71556736 = 0.72 in European ancestry groups of Phase 3 (Version 5) of the 1000 Genomes Project.
[h] Talmud et al. *Am J Hum Genet.* 2009 Nov 13;85(5):628–42.
[i] Kooner et al. *Nat Genet.* 2008 Feb;40(2):149–51.
[j] Kathiresan et al. *Nat Genet.* 2008 Feb;40(2):189–97 and Willer et al. *Nat Genet.* 2008 Feb;40(2):161–9.

Abbreviations: ARIC, Atherosclerosis Risk in Communities Study; CHARGE, Cohorts for Heart and Aging Research in Genetic Epidemiology; CHS, Cardiovascular Health Study; Fenland, The Fenland Study; FHS, Framingham Heart Study; HDL-C, high-density lipoprotein cholesterol concentrations; MESA, Multi-Ethnic Study of Atherosclerosis; NEO, Netherlands Epidemiology in Obesity Study; Raine Study, The Raine Study; RS, The Rotterdam Study; SE, standard error; SNP, single nucleotide polymorphism; TG, triglyceride concentrations; WGHS, Women's Genome Health Study; WHI, Women's Health Initiative; YFS, Young Finns Study.

V. Supplemental Figures and Figure Legends

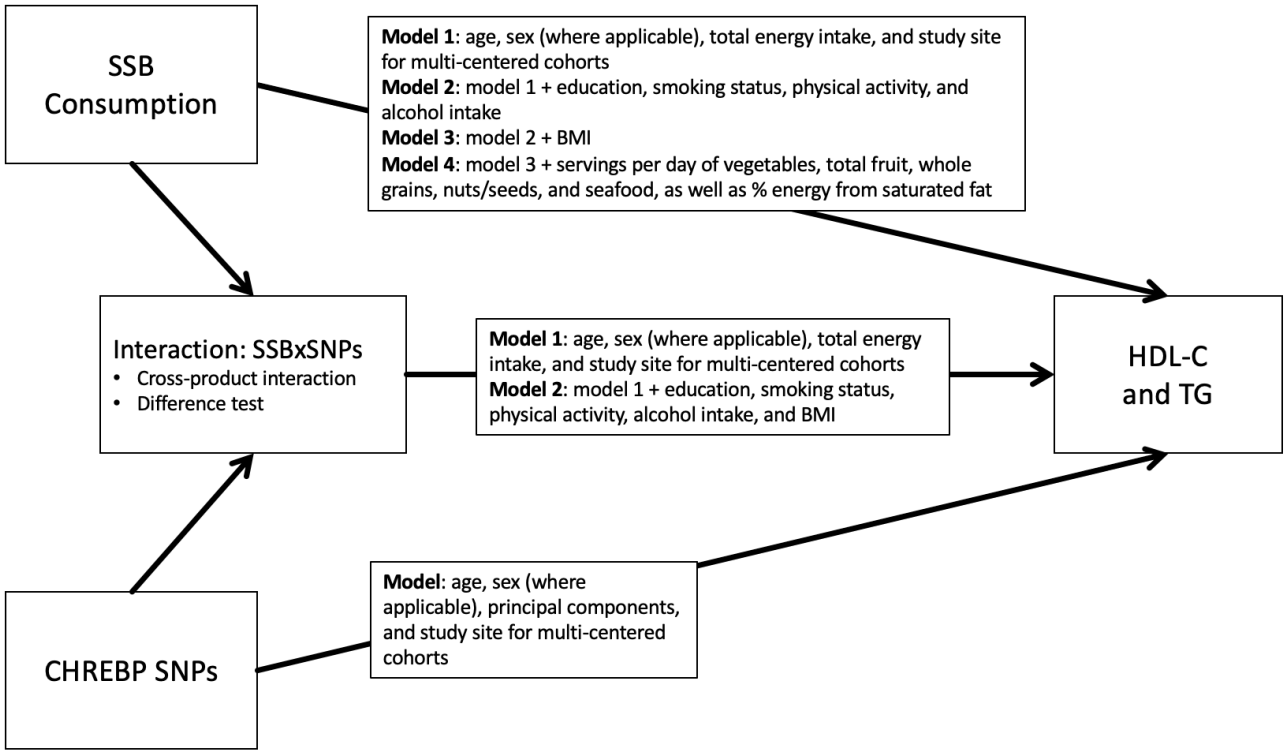


Figure I. Summary of applied linear regression models

Abbreviations: BMI, body mass index; CHREBP, carbohydrate responsive element binding protein; HDL-C, high-density lipoprotein cholesterol concentrations; SNP, single nucleotide polymorphism; SSB, sugar-sweetened beverages; TG, triglyceride concentrations.

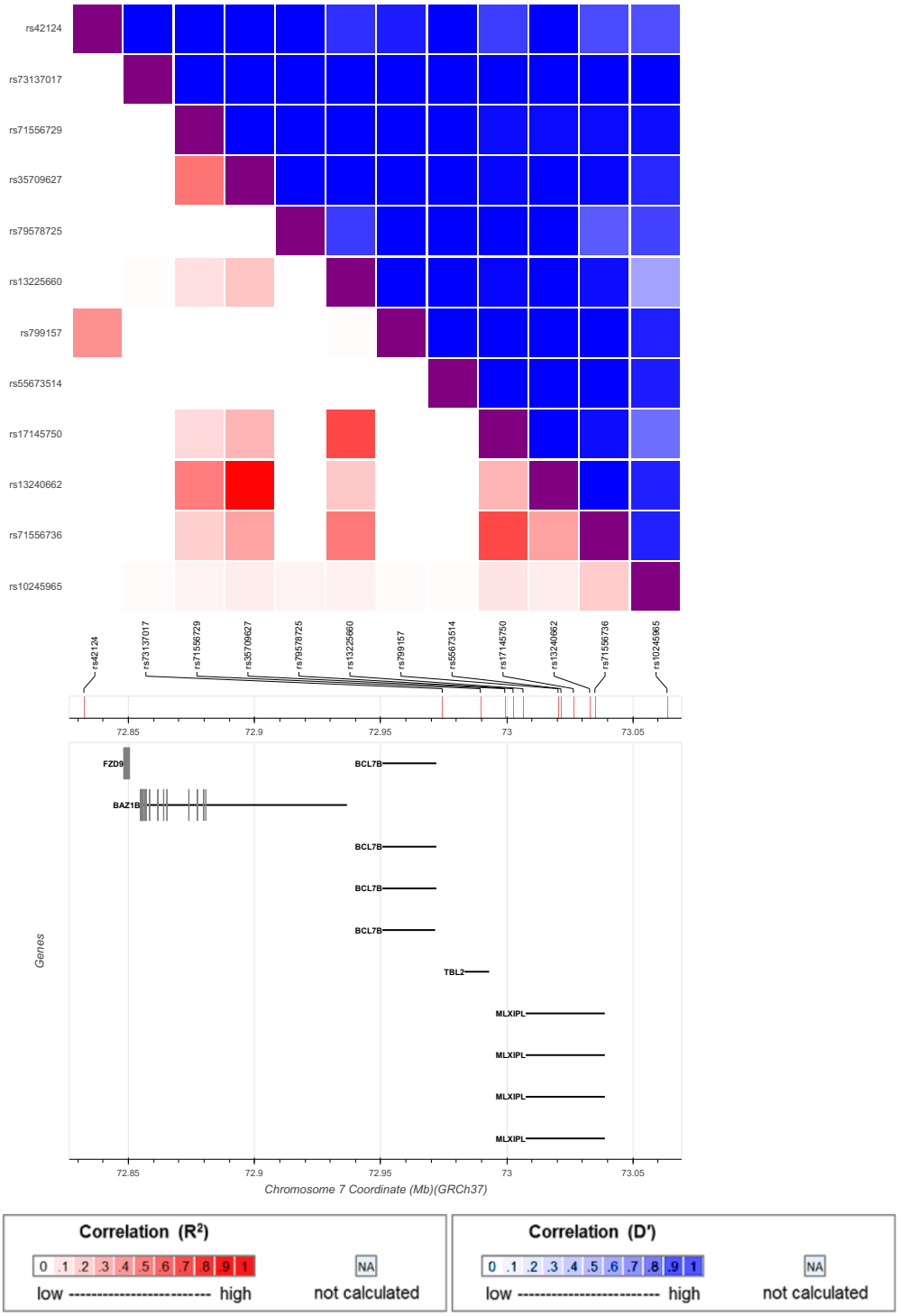


Figure II. Heatmap of pairwise linkage disequilibrium among top SNP and interaction signals generated using LDpair[90]

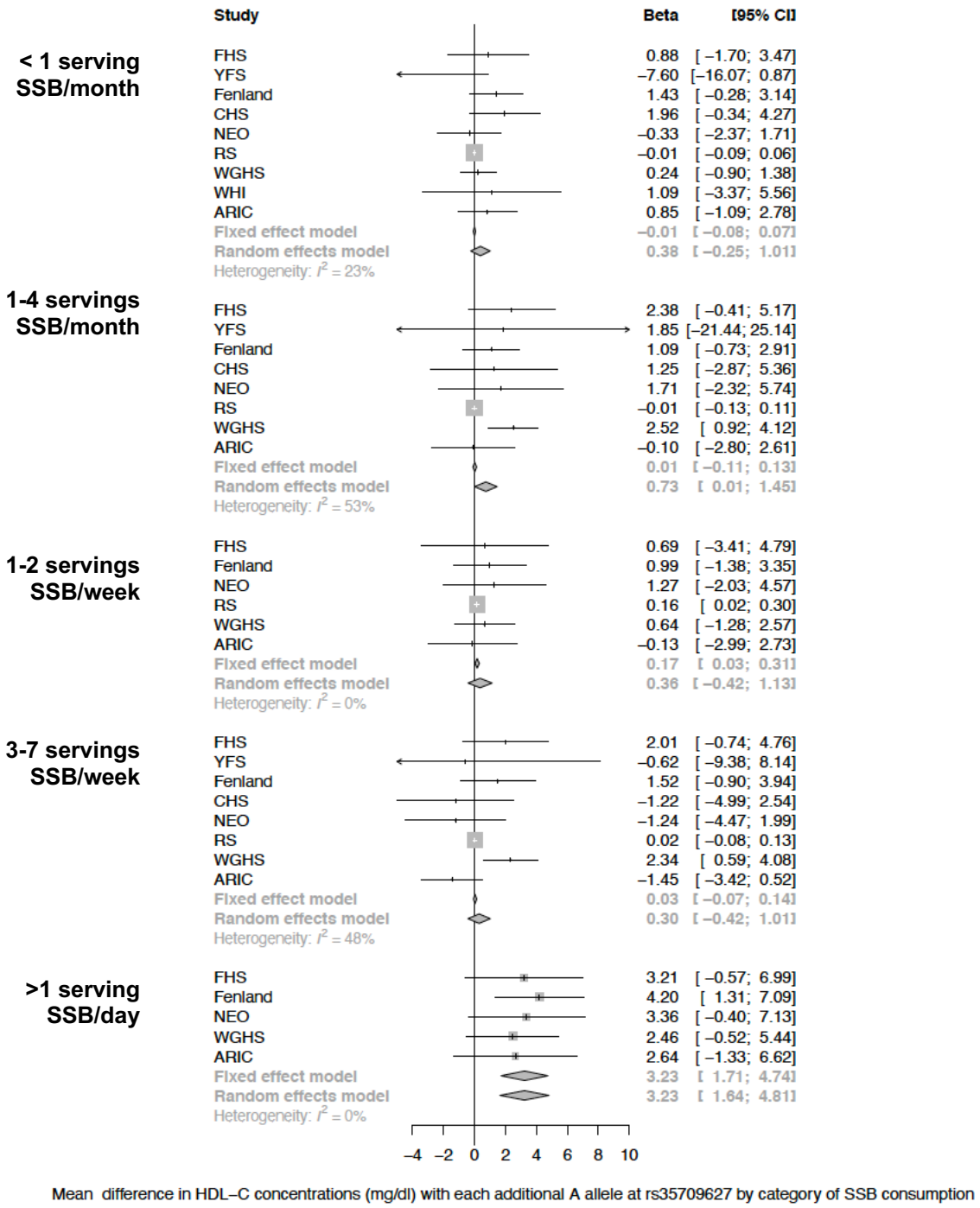


Figure III. Forest plot of association between rs35709627 and HDL-C concentrations stratified by category of SSB intake (Model 1)

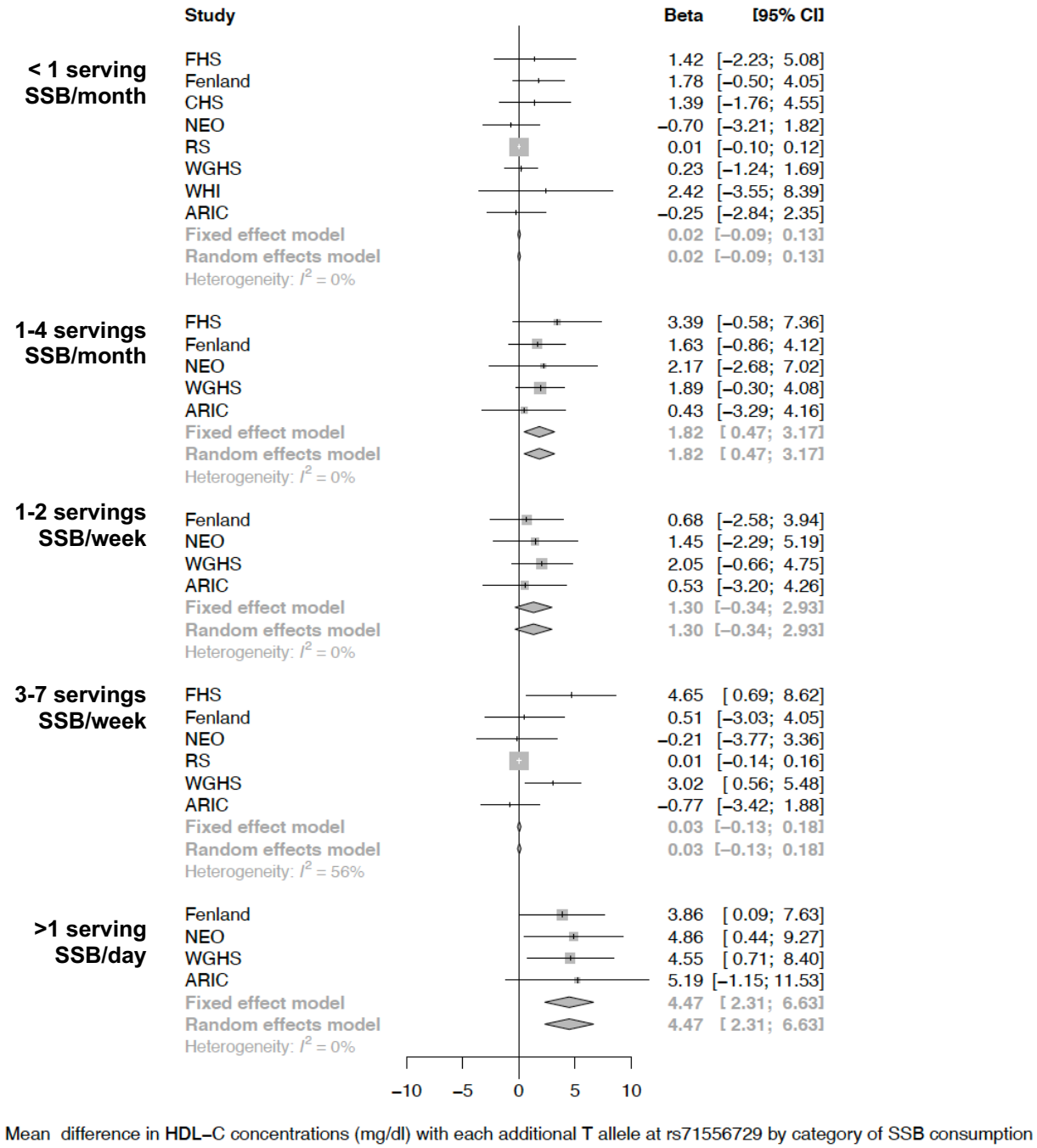


Figure IV. Forest plot of association between rs71556729 and HDL-C concentrations stratified by category of SSB intake (Model 1)

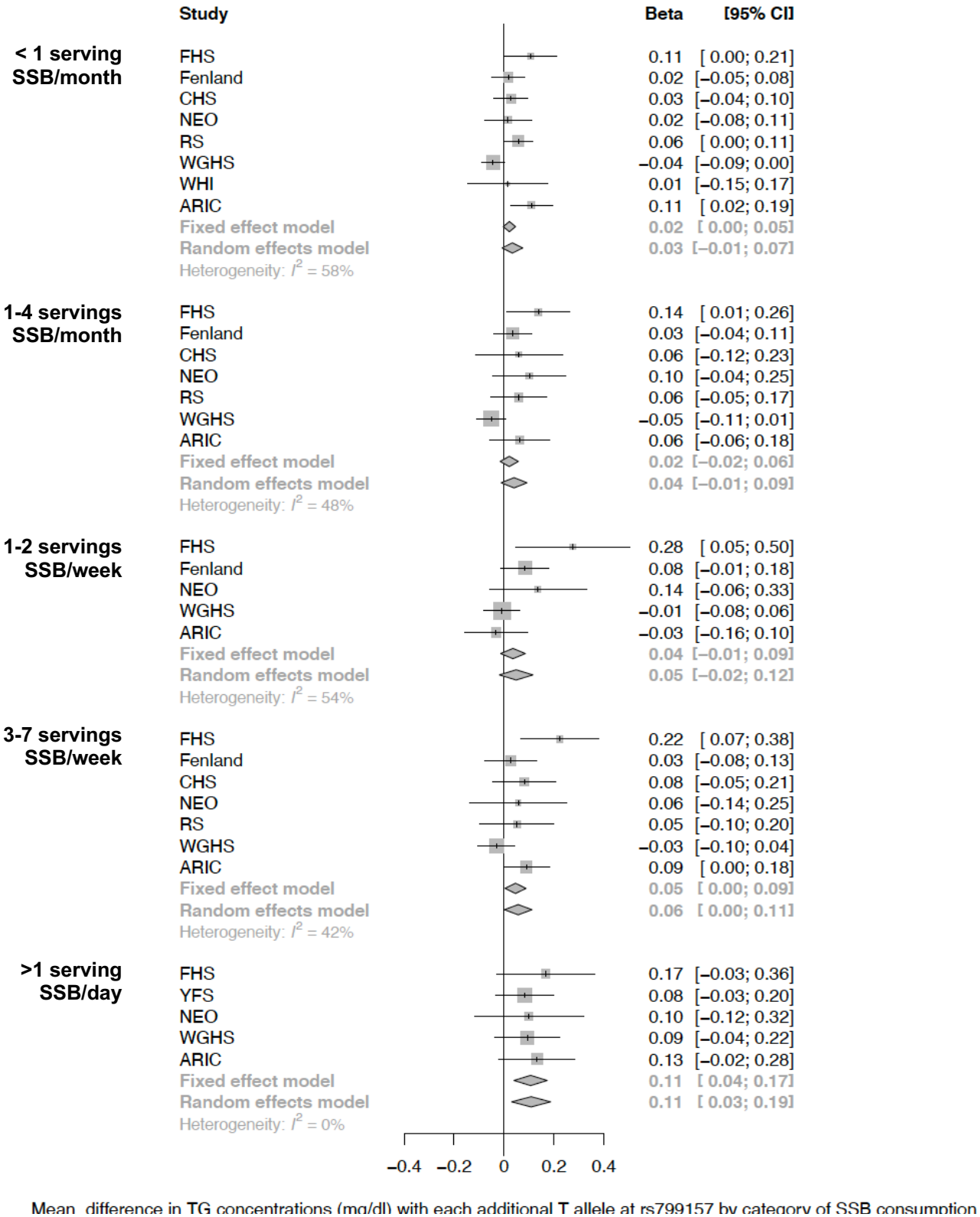


Figure V. Forest plot of association between rs799157 and TG concentrations stratified by category of SSB intake (Model 1)

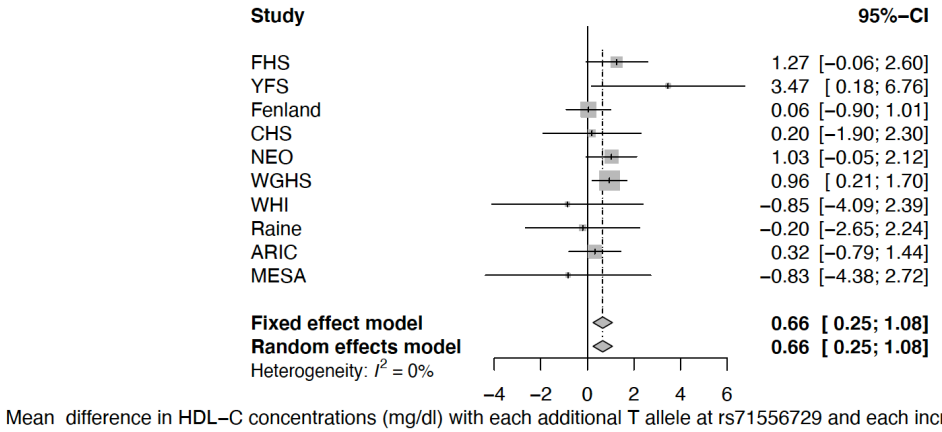


Figure VI. Forest plot of multiplicative interaction between SSB intake and rs71556729 on HDL-C concentrations (Model 1)

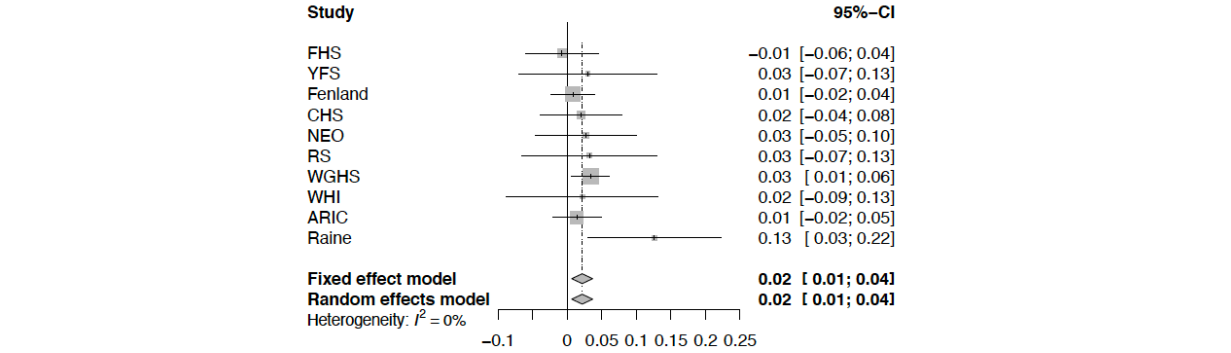


Figure VII. Forest plot of multiplicative interaction between SSB intake and rs55673514 on TG concentrations (Model 2)

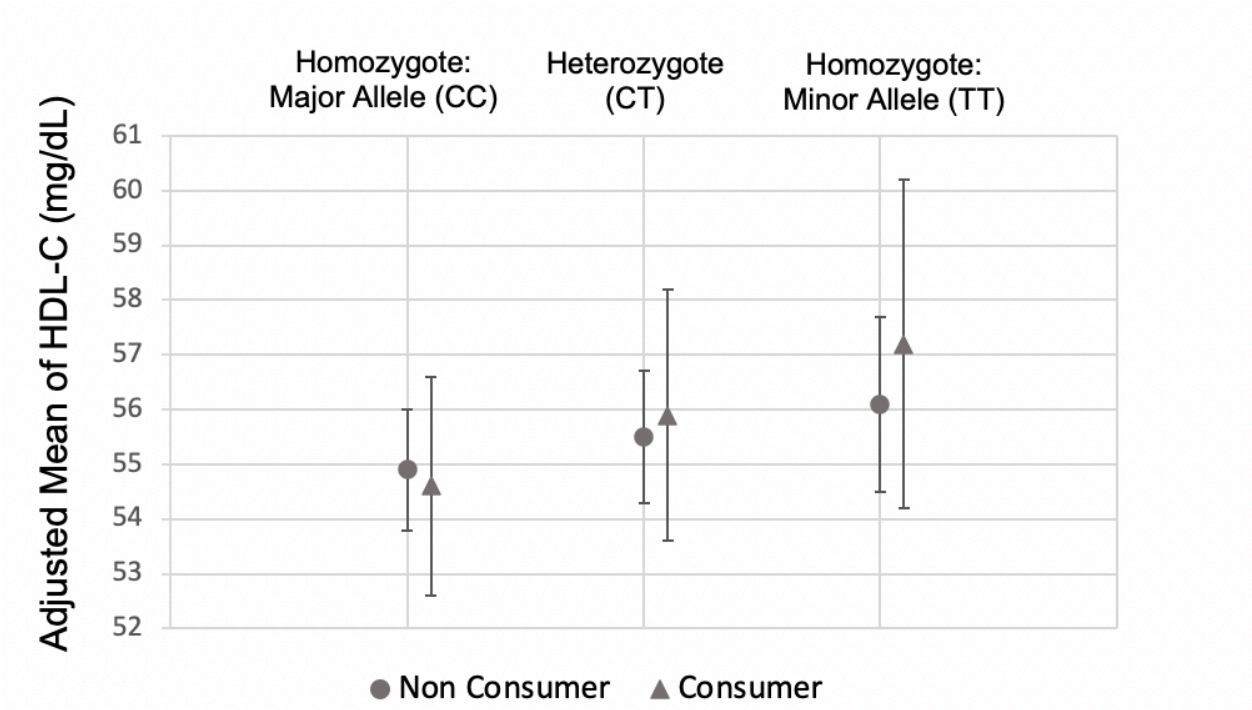


Figure VIII. Adjusted means of HDL-C concentrations among SSB consumers and SSB non-consumers by genotype at rs71556729 among UK Biobank participants

Adjusted mean HDL-C concentrations were lowest among SSB consumers with the CC genotype [mean HDL-C (95% CI): 54.6 (52.6, 56.6) mg/dl] and highest among SSB consumers with the TT genotype at rs71556729 [mean HDL-C (95% CI): 57.2 (54.2, 60.2) mg/dl] [β (SE) per additional minor allele: 1.3 (0.6) mg/dl, p_{trend}=0.03]. A positive trend between dosage at rs71556729 and adjusted mean HDL-C was also observed among SSB non-consumers, but the mean difference in HDL-C by genotype was smaller and not statistically significant [β (SE) per additional minor allele: 0.6 (0.3) mg/dl, p_{trend}=0.06]. Adjusted means were derived from linear regression models examining associations between rs71556729 dosage and HDL-C among unrelated SSB consumers and non-consumers in the UKB adjusting for age, sex, fasting hours prior to blood draw, total energy intake, current smoking status, education, physical activity, alcohol consumption, intakes of dietary fiber and potassium and percent energy from saturated fat and polyunsaturated fat, and body mass index. **Abbreviations**: HDL-C, high-density lipoprotein cholesterol concentrations; SE, standard error; SSB, sugar-sweetened beverage consumption.

Figure IX. Main association between category of SSB intake and HDL-C concentrations among CHARGE consortium cohorts

Compared to the lowest SSB consumers (<1 serving/month), participants consuming 1-4 servings SSB/month [β (95% CI):-0.8 (0.00, -0.5)], 1-2 servings SSB/week [β (95% CI):-1.6 (-2.1, -1.1)], 3-7 servings SSB/week [β (95% CI): -1.8 (-2.2, -1.3)], and >1 serving SSB/day [β (95% CI): -2.1 (-2.9, -1.2)] displayed lower mean HDL-C concentrations (mg/dl) in a dose-response manner. All regression models were adjusted for age, cohort (where applicable), sex (where applicable), total energy, education, current smoking status, physical activity, alcohol intake, body mass index, percent energy from saturated fat, and servings per day of vegetables, total fruits, whole grains (except the Young Finns Study), nuts/seeds, and seafood. I^2 represents heterogeneity for individual regression coefficients and I^2 overall represents heterogeneity in multivariate meta-analysis. Abbreviations: CHARGE, Cohorts for Heart and Aging Research in Genetic Epidemiology; CI, confidence interval; HDL-C, high-density lipoprotein cholesterol concentrations; SSB, sugar-sweetened beverages.

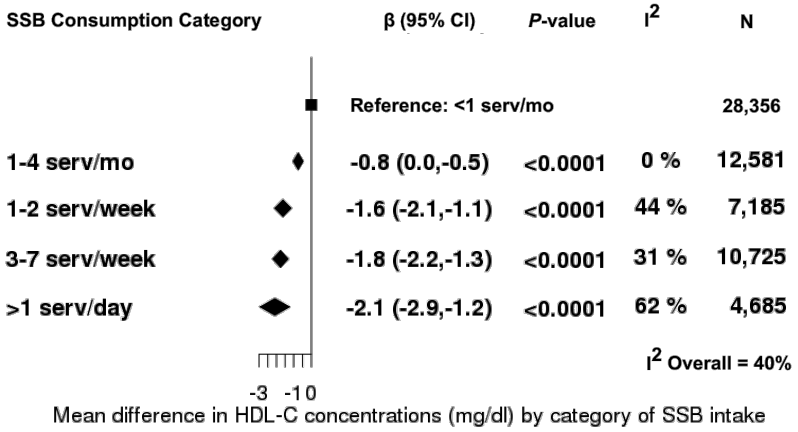


Figure X. Main association between category of SSB intake and TG concentrations among CHARGE consortium cohorts

Compared to the lowest SSB consumers (<1 serving/month), participants consuming 1-4 servings SSB/month [β (95% CI): 0.02 (0.00, 0.03)], 1-2 servings SSB/week [β (95% CI): 0.03 (0.02, 0.05)], 3-7 servings SSB/week [β (95% CI): 0.05 (0.03, 0.06)], and >1 serving SSB/day [β (95% CI): 0.06 (0.03, 0.09)] displayed higher mean TG concentrations (ln-mg/dl) in a dose-response manner. All regression models were adjusted for age, cohort (where applicable), sex (where applicable), total energy, education, current smoking status, physical activity, alcohol intake, body mass index, percent energy from saturated fat, and servings per day of vegetables, total fruits, whole grains (except Young Finns Study), nuts/seeds, and seafood. I^2 represents heterogeneity for individual regression coefficients and I^2 overall represents heterogeneity in multivariate meta-analysis. Abbreviations: CHARGE, Cohorts for Heart and Aging Research in Genetic Epidemiology; CI, confidence interval; SSB, sugar-sweetened beverages; TG, triglyceride.

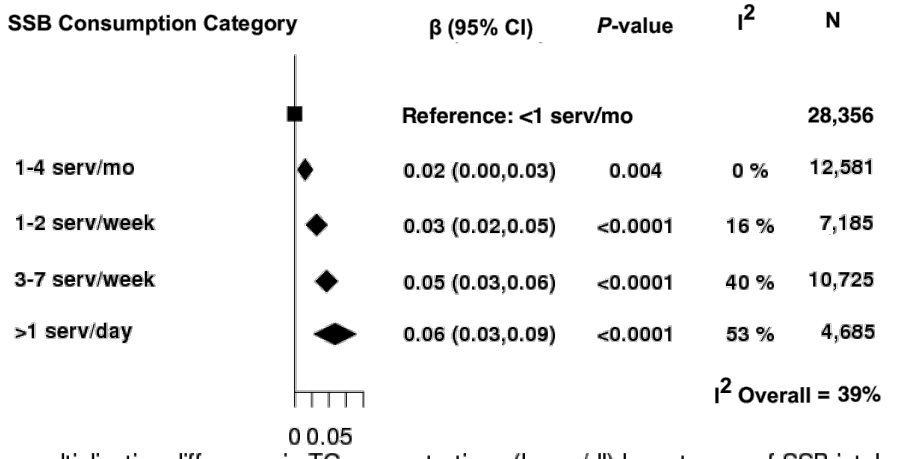


Figure XI. Mean difference in HDL-C concentrations (mg/dl) per increase in category of SSB intake among CHARGE consortium cohorts

In random-effects meta-analysis, each increase in SSB intake category was associated with lower mean HDL-C concentrations [β (95% CI): -0.55 (-0.72, -0.37) mg/dl; p_{trend} = 0.002]. All regression models were adjusted for age, cohort (where applicable), sex (where applicable), total energy, education, current smoking status, physical activity, alcohol intake, body mass index, percent energy from saturated fat, and servings per day of vegetables, total fruits, whole grains (except Young Finns Study), nuts/seeds, and seafood. Abbreviations: CHARGE, Cohorts for Heart and Aging Research in Genetic Epidemiology; CI, confidence interval; HDL-C, high-density lipoprotein cholesterol concentrations; SSB, sugar-sweetened beverages.

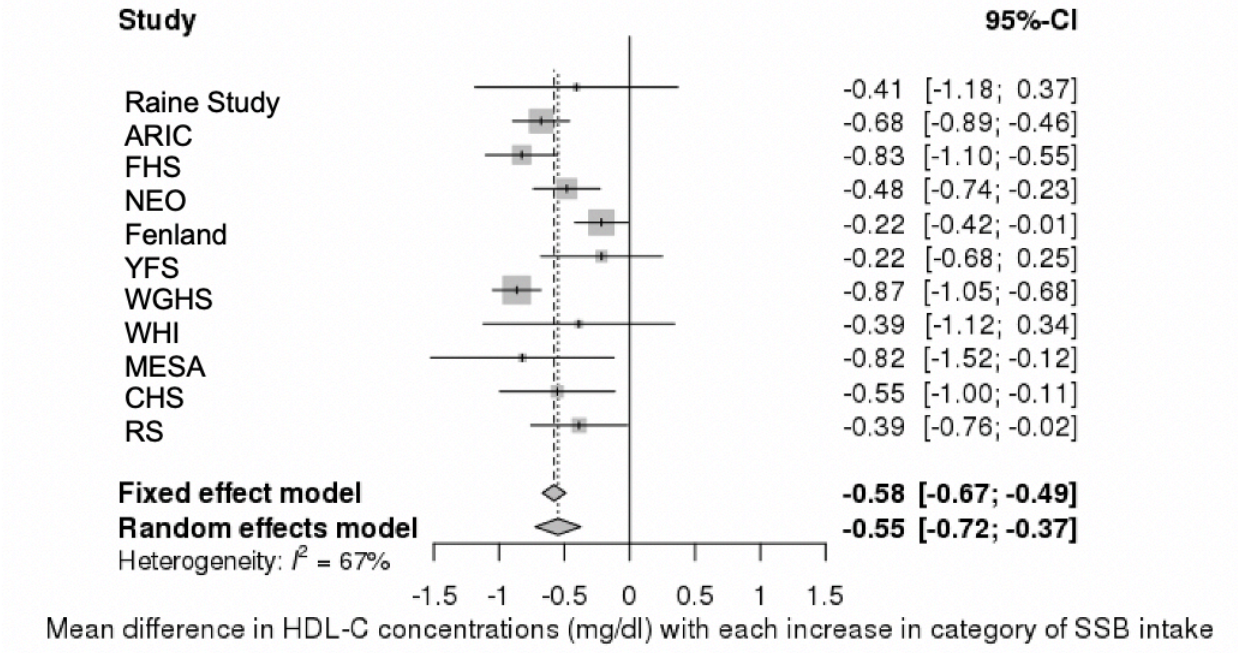
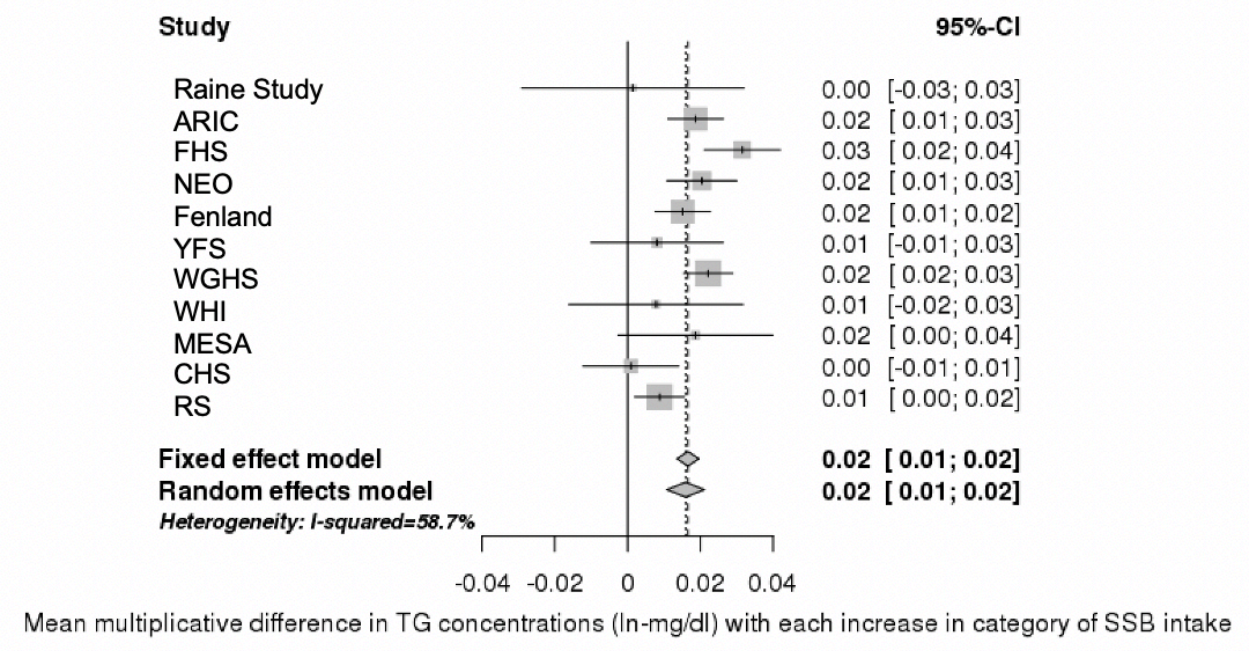


Figure XII. Mean multiplicative difference in TG concentrations (ln-mg/dl) per increase in category of SSB intake among CHARGE consortium cohorts

In random-effects meta-analysis, each increase in SSB intake category was associated with higher mean TG concentrations [β (95% CI): 0.02 (0.01, 0.02) ln-mg/dl; $p_{trend} < 0.0001$]. All regression models were adjusted for age, cohort (where applicable), sex (where applicable), total energy, education, current smoking status, physical activity, alcohol intake, body mass index, percent energy from saturated fat, and servings per day of vegetables, total fruits, whole grains (except Young Finns Study), nuts/seeds, and seafood. Abbreviations: CHARGE, Cohorts for Heart and Aging Research in Genetic Epidemiology; CI, confidence interval; SSB, sugar-sweetened beverages; TG, triglyceride.



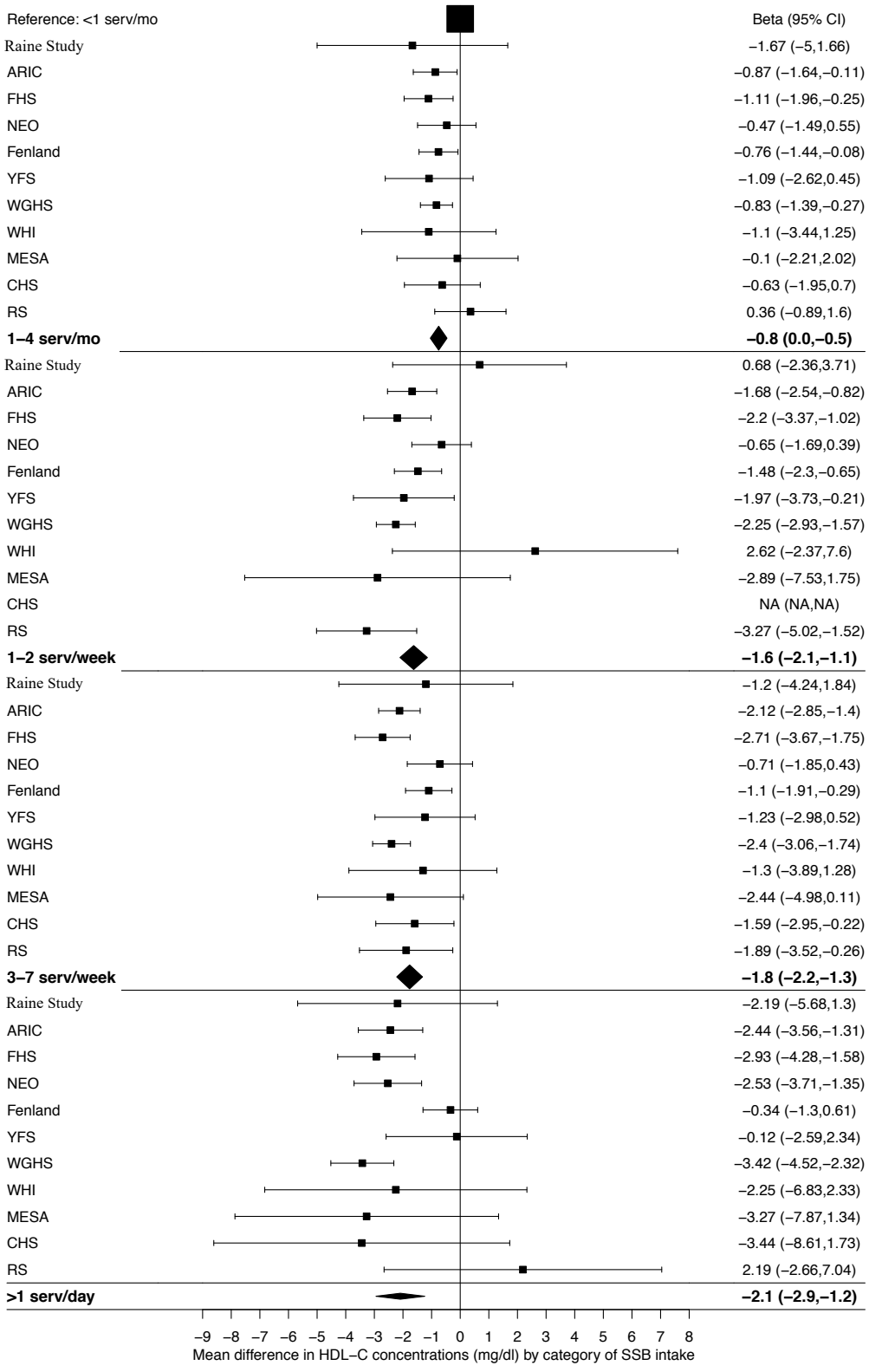


Figure XIII. Forest plot of main association between SSB intake and HDL-C concentrations

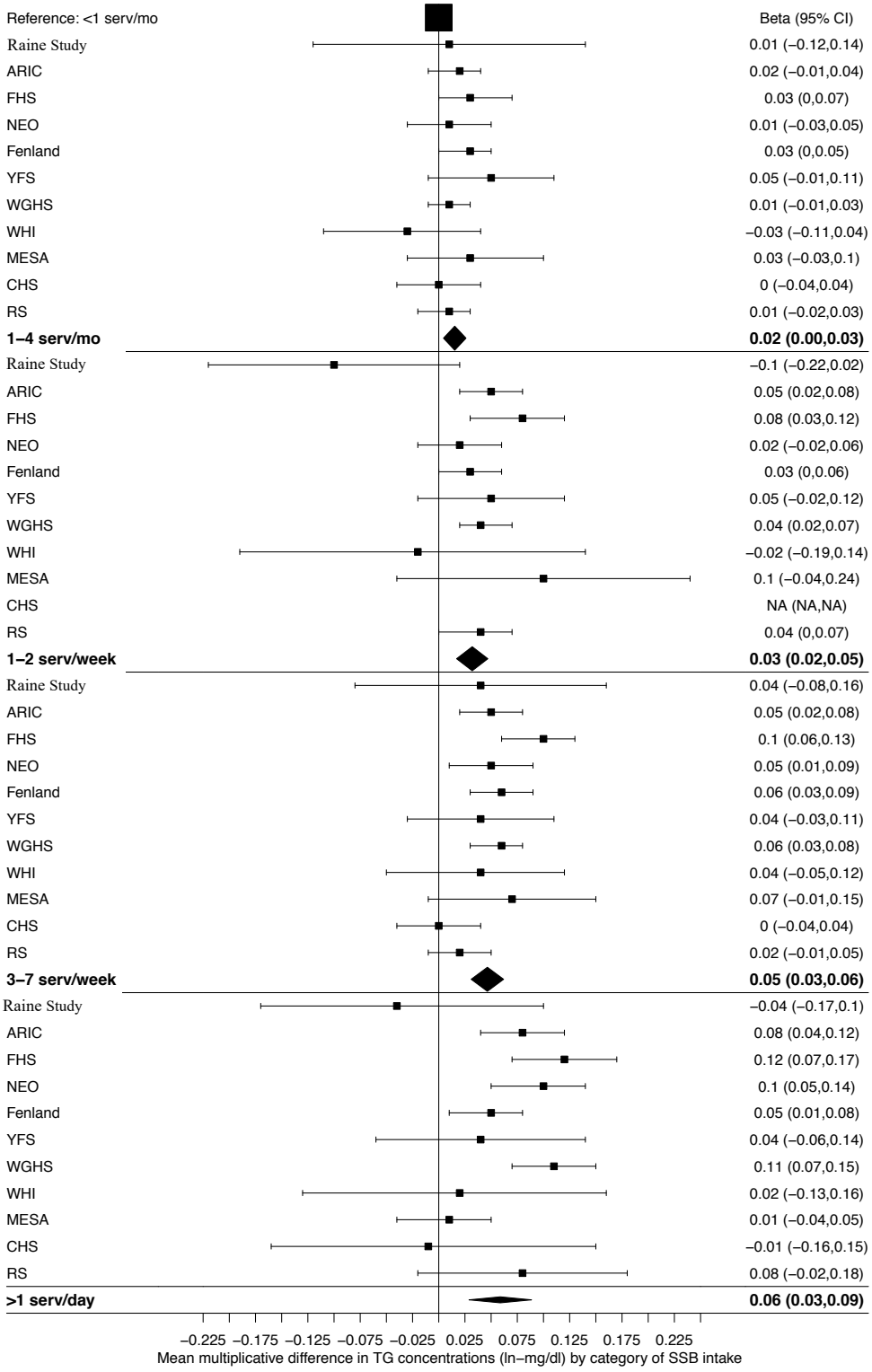


Figure XIV. Forest plot of main association between SSB intake and TG concentrations

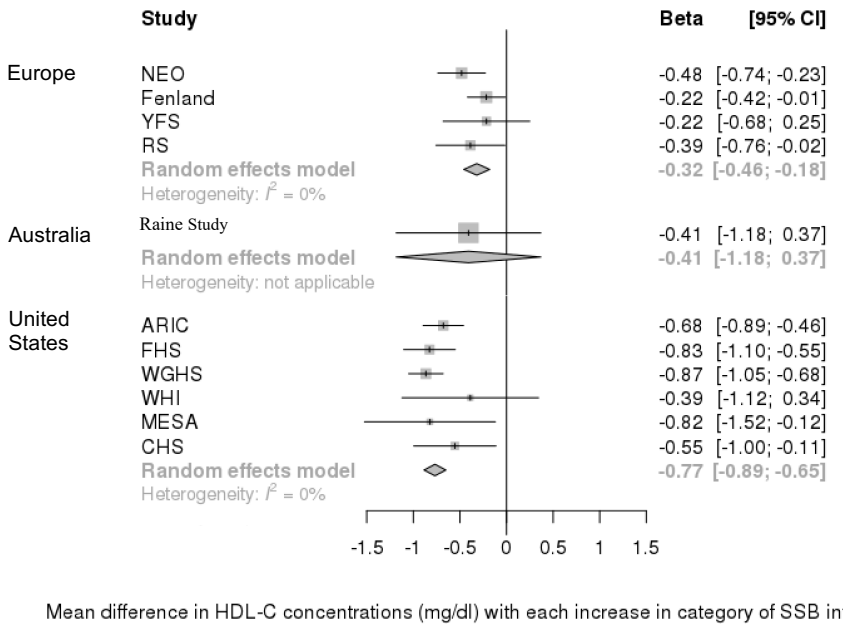


Figure XV. Forest plot of main association between SSB intake and HDL-C concentrations stratified by study region

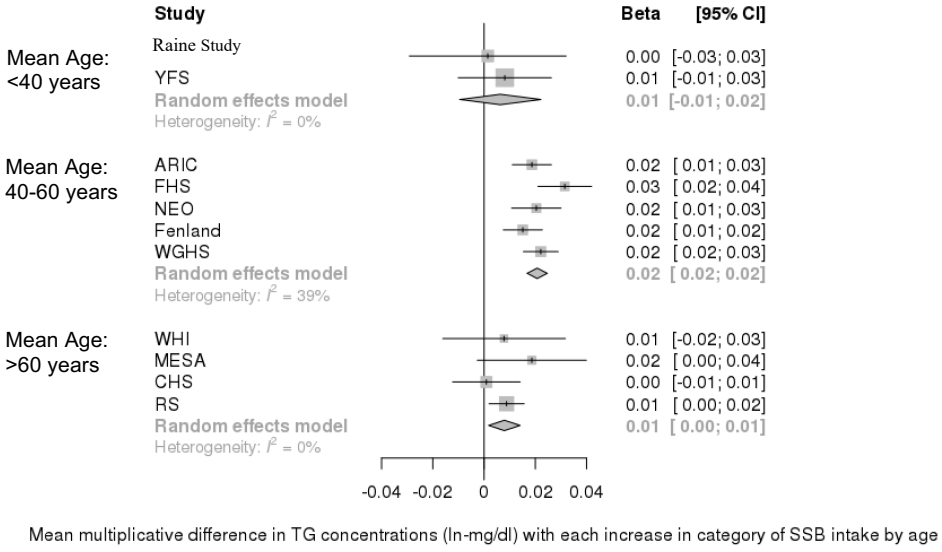


Figure XVI. Forest plot of main association between SSB intake and TG concentrations stratified by mean study age

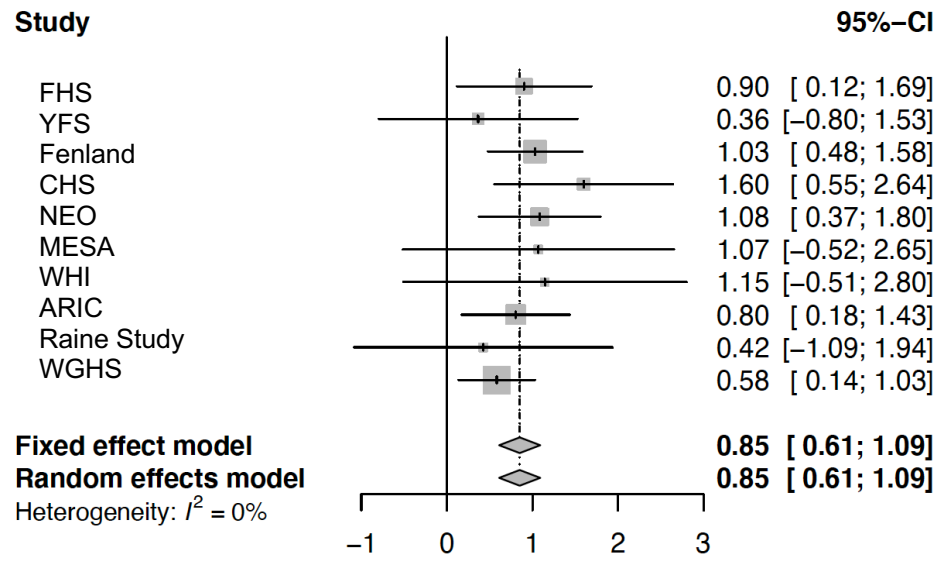


Figure XVII. Forest plot of association between rs17145750 and HDL-C concentrations

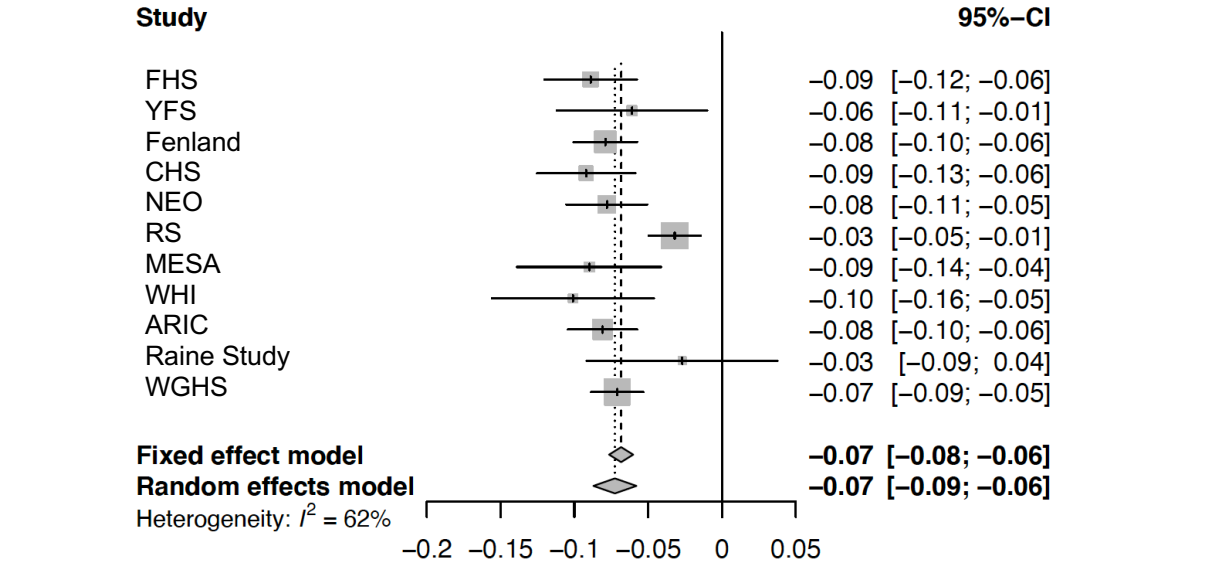


Figure XVIII. Forest plot of association between rs71556736 and TG concentrations

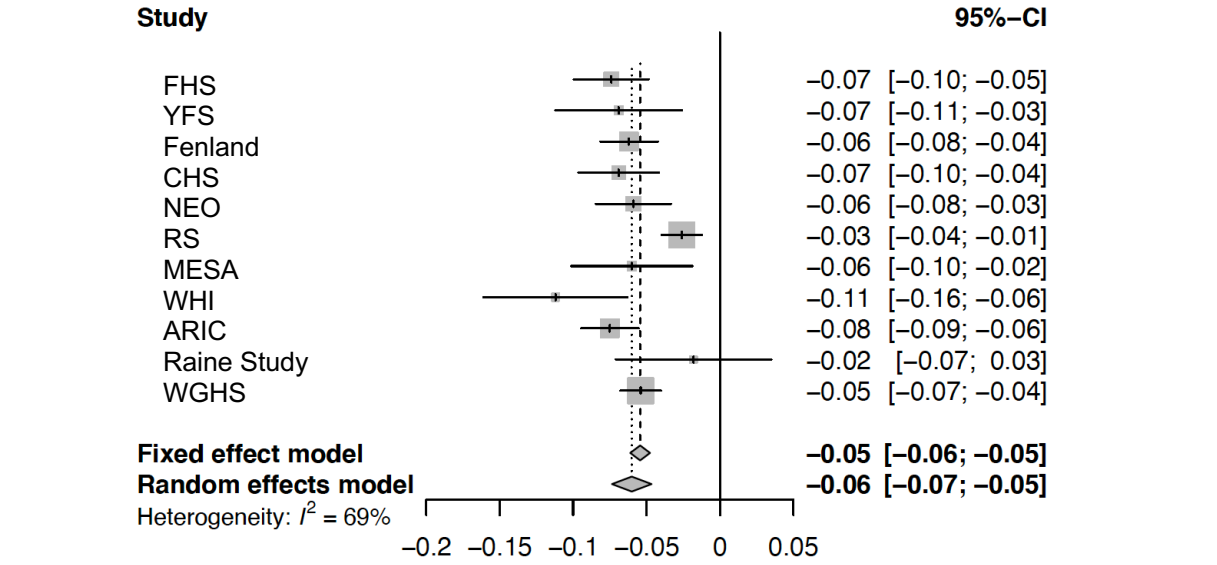


Figure XIX. Forest plot of association between rs13225660 and TG concentrations

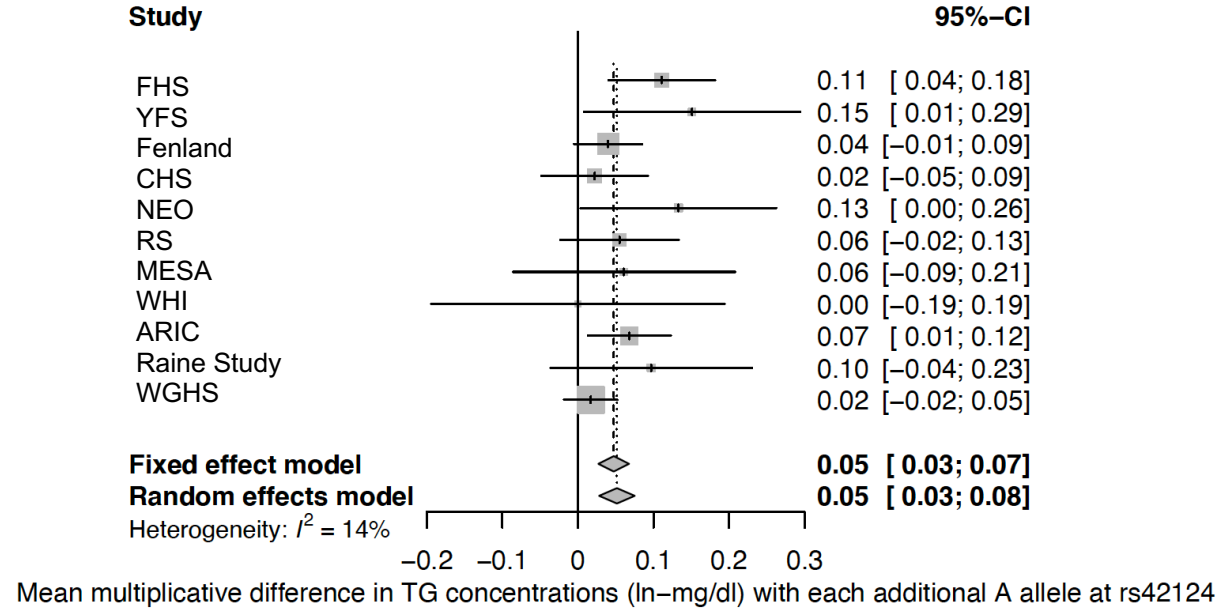


Figure XX. Forest plot of association between rs42124 and TG concentrations

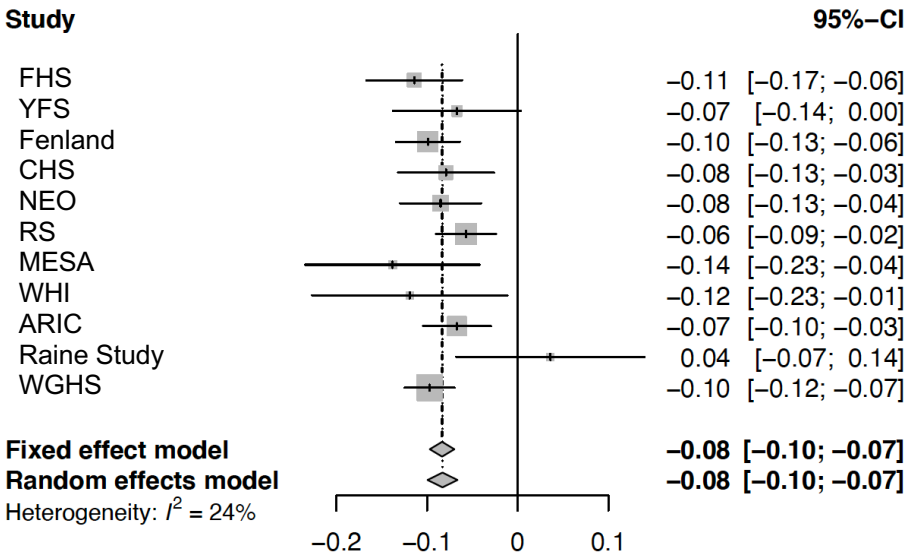


Figure XXI. Forest plot of association between rs13240662 and TG concentrations

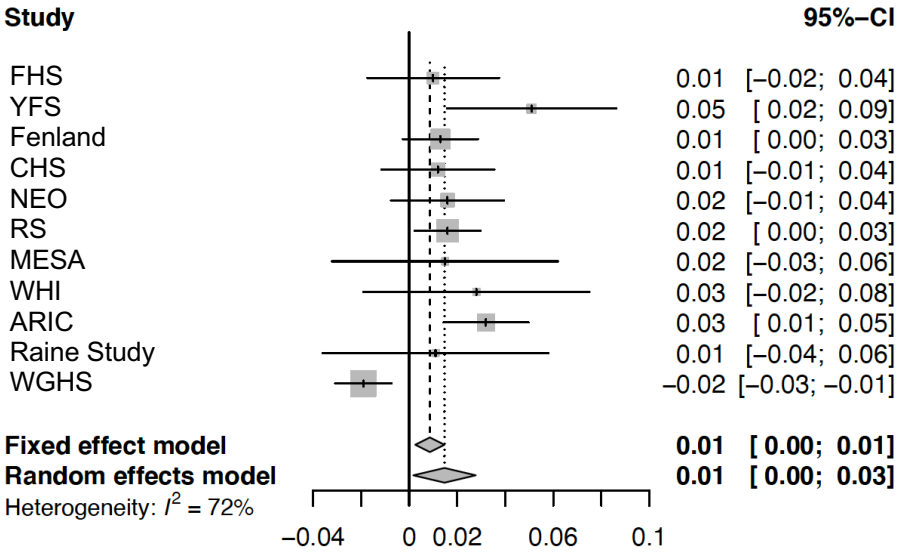


Figure XXII. Forest plot of association between rs10245965 and TG concentrations