ELECTRONIC SUPPLEMENTARY MATERIAL (ESM)

Abbreviations for ESM Tables

Cohort study name (study acronym) (country): Cardiovascular Health Study (CHS) (USA), Framingham Heart Study (FHS) (USA), Multi-Ethnic Study of Atherosclerosis (MESA) (USA), Rotterdam Study I (RS1) (The Netherlands), Rotterdam Study II (RS2) (The Netherlands), Cardiovascular Risk in Young Finns Study (YFS) (Finland); Atherosclerosis Risk In Communities (ARIC) Study (USA); Malmö Diet and Cancer (MDC) Study (Sweden); Netherlands Epidemiology in Obesity (NEO) Study (The Netherlands); Nurses' Health Study (NHS) (USA); and Western Australian Pregnancy Cohort Study (Raine) (Australia).

Abbreviations: BMI, body mass index; Chr, chromosome; MAF, minor allele frequency; N, total sample size; NA, not available; PCA, principal component analysis.

ESM METHODS

SNP Selection Strategy

Variants in the *CHREBP* locus associate with hypertriacylglycerolaemia and low HDL-cholesterol at genome-wide significance levels, and variants in the FGF21 locus are associated with higher circulating FGF21 concentrations and higher carbohydrate relative to fat intake in humans (see variant-specific references in ESM Table 4). It has long been postulated that insulin resistance represents a common pathogenic mechanism contributing to the development of numerous features of the metabolic syndrome and is a key factor leading to the development of type 2 diabetes. Among the features of the metabolic syndrome, the correlation between elevated fasting serum triacylglycerol and low-HDL cholesterol with insulin resistance is particularly strong. Nevertheless, by genome-wide association studies (GWAS) there has been little overlap between loci associated with hypertriacylglycerolaemia or low HDL-cholesterol and those implicated in the development of hyperglycaemia, hyperinsulinaemia, or diabetes. Thus, either there is no common pathogenic genetic determinant for hypertriacylglycerolaemia and insulin resistance, or unidentified variables may interact with genetic determinants of hypertriacylglycerolaemia to cause insulin resistance. Given a role for ChREBP in contributing to sugar-induced derangements in both lipid and glucose homeostasis, we sought to test the hypothesis that variants associated with hypertriacylglycerolaemia in the ChREBP pathway might interact with SSB consumption to regulate glycaemic traits. Aside from variants in *CHREBP*, we selected SNPs that previously showed significant (i.e., $p < 5 \times 10^{-8}$) or suggestive (i.e., $p < 5 \times 10^{-6}$) association with hypertriacylglycerolaemia or low-HDL cholesterol in humans in genes important for hepatic fructose and glucose metabolism (KHK, ALDOB, GCK, SLC2A2, SLC2A5), which may directly regulate ChREBP activity. We also included other genes implicated in the regulation of both ChREBP and blood serum triacylglycerol levels (FADS1 and TRIB1). Lastly, we included variants in the loci that code for ChREBP-regulated metabolic hormone FGF21 and its obligate receptor KLB.

Genotype Exclusion Criteria

Genotyped SNPs were excluded on the basis of low call rate (<95%) or departure from Hardy-Weinberg equilibrium (p <1x10^{-6}). Imputed SNPs were removed on the basis of low imputation quality (MACH and BIMBAM: R^2 <0.3; IMPUTE: proper info <0.4). Additional genotyping information for each cohort is provided in ESM Table 5.

Genetic Analyses

Main associations between all selected SNPs for discovery cohorts and a subset of selected SNPs for replication cohorts [*i.e.*, those with suggestive interaction results (*i.e.*, p <0.05) from the discovery cohorts] and glycaemic outcomes were investigated by using linear regression models or linear mixed effects models for family data, and an additive genetic model adjusted for age, sex, study site for multi-centered cohorts, and population structure where applicable (in CHS; FHS; MESA; RS1; RS2; YFS for discovery cohorts; and ARIC; NEO; Raine; MDC for replication cohorts). In discovery cohorts, a total of 36 interaction tests (1 environmental variable × 18 SNPs × 2 outcomes) between SSB intake and the selected SNPs on glycaemic outcomes were performed using SSB × SNP cross-product terms and including main-effect terms in linear regression analyses or linear mixed effects models for family data adjusted for age, sex, energy intake, BMI, and study site for multi-centered cohorts, and population structure where applicable.

In a secondary analysis using a similar approach previously described, we examined the sex-stratified SNP associations between 14 out of the 18 selected SNPs (those available in >2 discovery cohorts) and glycaemic outcomes, in addition to sex-stratified interaction tests between SSB intake and the 14 selected SNPs on glycaemic outcomes in the following discovery cohorts (CHS; FHS; MESA; RS1; RS2). Similarly, for SNPs with suggestive interaction results (*i.e.*, p <0.05) from the discovery cohorts, sex-stratified replication was sought among replication cohorts (ARIC; NEO; NHS; RAINE).

Power Calculations

Our study had 80% power to detect a minimal interaction effect between total SSB intake (per 1 serving/day) and SNP (per effect allele) on fasting glucose levels of 0.10 for SNPs with MAF ≥ 0.05, and even smaller interaction effects of 0.05 for SNPs with MAF ≥ 0.25 at the Bonferroni-corrected p-value of p <0.001.

Sensitivity Analyses

Heterogeneity across studies was tested by using Cochran's Q statistic and quantified using the I^2 statistic. Analyses with moderate 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 *metafor* package (R version 3.1.0; https://cran.r-project.org) to assess the effect of the following moderator variables on heterogeneity of association/interaction: geographical location (U.S. vs. northern Europe vs. Australian), BMI (<27 vs. ≥27 kg/m^2), mean age (<60 vs. ≥60 years), blood sample (serum vs. plasma), and sample size (n <1000 vs. ≥1000). In addition, we conducted sensitivity analyses to assess the influence of individual cohorts on the meta-analysed estimates by removing one cohort study at a time for all analyses.

ESM RESULTS

Sex-Stratified Interaction Analyses

In sex-stratified analyses, we observed a suggestive interaction between SSB intake and FGF21-rs838133 for FI among men [β ± SE = -0.0477 ± 0.0168 $\log_e$-pmol/l, p = 0.005] and GCK-rs4607517 for FI among femen [β ± SE = 0.05 ± 0.0226 $\log_e$-pmol/l, p = 0.03] (ESM Table 13). These findings, however, did not replicate among our replication cohorts (ESM Table 14).

Meta-Regression and Sensitivity Analyses

Meta-regression results suggest that cohorts' mean BMI and age, and cohort sample size may influence our meta-analysed estimates for associations between SSB intake and FI (ESM Table 15). Our subsequent subgroup meta-analyses revealed differences in the magnitude, but not statistical significance, of the associations between SSB intake and FI, except when subgrouping by cohort sample size. The association between SSB intake and FI remained significant among the cohorts with a larger sample size (9 cohorts) and not a smaller sample size (2 cohorts). In addition, sensitivity analyses did not identify source of heterogeneity (results not shown) and performing a random-effect meta-analyses did not substantially alter our observed results (ESM Table 16).

ESM Table 1. Descriptions and acknowledgements of participating CHARGE cohorts.

Cohort	Study Description and Acknowledgements	Relevant References
Atherosclerosis Risk In Communities (ARIC) Study[b] **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. [representing authors: MG, KLY, and KEN]	https://www2.cscc.unc.edu/aric/ The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989 Apr;129(4):687-702.
Cardiovascular Health Study (CHS)[a] **USA**	The Cardiovascular Health Study (CHS) is a population-based prospective cohort study of cardiovascular disease, and includes 5,888 participants ≥ 65 years of age identified from four U.S. communities using Medicare eligibility lists (Forsyth County, NC; Sacramento County, CA; Washington County, MD; Pittsburgh, PA). The original cohort included 5201 participants recruited in 1989–1990 and 687 additional subjects were recruited in 1992–1993 to enhance the racial/ethnic diversity of the cohort. The CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, 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 CHS investigators and institutions can be found at http://chs-nhlbi.org/. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, 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. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant HL105756. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. [representing authors: RNL, KJM, KR, BMP, DSS, and DM]	http://www.chs-nhlbi.org/ Ann Epidemiol. 1(3): 263-276, 1991 Fried, L.P. et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol 1, 263-76 (1991).

Cohort	Study Description and Acknowledgements	Relevant References
Framingham Heart Study (FHS) [a] **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 every 3–4 y. 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) 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. [representing authors: JD, DR, LAC, and JBM]	Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci. 1963;107:539-556. PMID 14025561 Feinleib m, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med. 1975;4:518-525. PMID 1208363 Splansky GL, Corey D, Yang Q et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol. 2007;165:1328-1335. PMID 17372189
Multi-Ethnic Study of Atherosclerosis (MESA) [a] **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 SNP Health Association Resource (SHARe) project are conducted and supported by contracts N01-HC-95159 through N01-HC- 95169 and RR-024156 from the National Heart, Lung, and Blood Institute. Funding for MESA SHARe genotyping was provided by National Heart, Lung, and Blood Institute [Contract N02-HL-6-4278]. MESA Family is conducted and supported in collaboration with MESA investigators; support is provided by National Heart, Lung, and Blood Institute grants and contracts [grant numbers R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071252, R01HL071258, R01HL071259]. [representing authors: ACW, ATC, LS and SSR]	Bild DE, et al. Am. J. Epidemiol. 156 (9): 871-881.2002

Cohort	Study Description and Acknowledgements	Relevant References
Malmö Diet and Cancer (MDC) Study [b] **Sweden**	The Malmö Diet and Cancer study is a population-based cohort designed to investigate the relationship between dietary factors and certain forms of cancer. At baseline in 1991-1996, a total of 28,098 adults (40% of those invited), aged 45 to 73 years, participated in the study. The cardiovascular sub-cohort consists of 6,103 adults (46-68 y, 58% femen) randomly selected from the parent Malmö cohort. In the present study sample, individuals with reported energy intake above or below the 95% confidence interval for total energy expenditure (i.e., estimated from their calculated basal metabolic rate and self-reports of leisure-time physical activity, work activity, household work, and sleep hours) were excluded. The Malmö Diet and Cancer study was initiated and planned in collaboration with the International Agency for Research on Cancer (IARC), the Swedish Cancer Society, the Swedish Medical Research Council and the Faculty of Medicine, Lund University, Sweden. The study is also funded by Region Skåne, City of Malmö, Påhlsson Foundation and the Swedish Heart and Lung Foundation. [representing authors: ES, CAS, UE, LB and MOM]	
Netherlands Epidemiology in Obesity (NEO) Study [b] **The Netherlands**	The NEO Study 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 panal. 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, Petra Noordijk, Pat van Beelen 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). [representing authors: DOMK, RLG, FRR, and RdM]	

Cohort	Study Description and Acknowledgements	Relevant References
Nurses' Health Study (NHS) [b] **USA**	The NHS is a prospective cohort study of 121,700 female registered nurses who were 30 to 55 years old at study inception in 1976. Information about medical history, lifestyle factors and health conditions has been collected biennially by self-administered questionnaires since inception. For this analysis, we used 1986 as baseline and a total of 676 healthy women of European ancestry with both genotype and dietary data available were included. The Nurses' Health Study was supported by grants of UM1 CA186107, R01 CA49449 from the National Institutes of Health. Current study was supported by the National Heart, Lung, and Blood Institute (HL071981, HL034594, HL126024), the National Institute of Diabetes and Digestive and Kidney Diseases (DK58845, DK091718, DK100383), the Boston Obesity Nutrition Research Center (DK46200), and United States – Israel Binational Science Foundation Grant2011036. Dr. Qi was a recipient of the American Heart Association Scientist Development Award (0730094N). [representing authors: YL, TH, and LQ]	
Western Australian Pregnancy Cohort (Raine) Study [b] **Australia**	The Western Australian Pregnancy Cohort (Raine) Study is a prospective pregnancy cohort where 2,900 mothers were recruited between 1989 and 1991 (http://www.rainestudy.org.au/). Recruitment took place at Western Australia's major perinatal centre, King Edward Memorial Hospital, and nearby private practices. Women were eligible for enrolment if they were between 16 and 20 weeks gestation, had sufficient English language skills, an expectation to deliver at King Edward Memorial Hospital and an intention to reside in Western Australia to allow for future follow-up of their child. The mothers completed questionnaires regarding the children and the children had physical examinations at average ages of 1, 2, 3, 6, 8, 10, 14, 17, 20 and 22 years. This study was supported by the National Health and Medical Research Council of Australia [grant numbers 403981 and 003209] and the Canadian Institutes of Health Research [grant number MOP-82893]. The authors are grateful to the Raine Study participants and their families, and to the Raine Study research staff for cohort coordination and data collection. The authors gratefully acknowledge the NH&MRC for their long term contribution to funding the study over the last 20 years and also the following Institutions for providing funding for Core Management of the Raine Study: The University of Western Australia (UWA), Raine Medical Research Foundation, UWA Faculty of Medicine, Dentistry and Health Sciences, The Telethon Institute for Child Health Research and Women and Infants Research Foundation. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. [representing authors: CAW, WHD, and CEP]	Newnham JP, Evans SF, Michael CA, Stanley FJ, Landau LL. Effects of frequent ultrasound during pregnancy: a randomised controlled trial. *Lancet* 1993;342(8876):887-91.

Cohort	Study Description and Acknowledgements	Relevant References
Rotterdam Study I & II (RS1 & RS2) [a] **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. [representing authors: ETML, JCK-dJ, AH, AGU, TV and OHF]	Hofman et al. Eur J Epidemiol. 2015;30(8):661-708
Cardiovascular Risk in Young Finns Study (YFS) [a] **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. A total of 1,730 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; Kuopio, Tampere and Turku University Hospital Medical Funds (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 expert technical assistance in the statistical analyses by Irina Lisinen, and Mika Helminen are gratefully acknowledged. [representing authors: VM, OR, MK, JV, IS, and TL]	Raitakari OT et al. Cohort profile. Int. J Epidemiol. 2008;37:1220-6.

[a] Discovery cohorts

[b] Replication cohorts

ESM Table 2. Dietary assessment methods of participating CHARGE cohorts.

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Whole Grain Intake	Fish Intake	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)	Cooked cereals/ whole grain cold cereal/ dark or whole grain bread	Canned tuna fish/ dark meat fish/ other light meat fish	Willett WC, *et al. Am J Epidemiol.* 1985; 122(1):51–65. Stevens J, *et al. Nutrition Research* 1996;16:735–745.
CHS	99-item, self-administered, picture-sort version of National Cancer Institute FFQ	Harvard	Tang, Start breakfast drinks; Regular soft drinks	Apple, applesauce, pears; bananas; peaches, apricots (canned, frozen); peaches, apricots, nectarines (in season); cantaloupe; watermelon; strawberries; oranges; grapefruit; any other fruit	String beans, green beans; peas; corn; winter squash, baked squash; tomato, tomato juice; broccoli; cauliflower or Brussels sprouts; spinach (raw); spinach (cooked); mustard greens, turnip greens, collards; coleslaw, cabbage, sauerkraut; carrots, mixed vegies with carrots; green salad; other vegies including onions, summer squash; sweet potatoes yams	Dark bread, like whole wheat, rye, pumpernickel; cooked cereals; high fiber, bran or granola cereals	Tuna; other fish; fried fish	Kumanyika S, et al. J Am Diet Assoc. 1996;96(2):137–144.
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).s; quantified as servings/week	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	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	Rimm et al. Am J Epidemiol 1992;135:1114–26, 1127–36.

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Whole Grain Intake	Fish Intake	Relevant References
MDC	Modified diet history (7-day food diary of prepared meals and cold beverages, combined with a 168-item FFQ of regularly consumed foods not covered by the diary, followed by a 1-hour diet interview)	Swedish National Food Agency Database	Sugar-sweetened soft drinks, sugar-sweetened fruit juice. SSB were reported in 7-day dairy.	From questionnaire: apples, pears, bananas, orange, other citrus fruits, peaches/nectarines, plums, kiwi, grapes, melons, dried fruit, other fruits. Fruits in cooked meals were reported in 7-day diary.	From questionnaire (vegetables on sandwich and between meals): carrot, tomato, cucumber, peppers, green salad, pickled vegetables, other vegetables. Vegetables in cooked meals were reported in 7-day dairy.	Soft bread >6% fiber, crisp bread >10% fiber, biscuits and rusks >10% fiber, cereals (information on bread and cereals were mainly from questionnaire)	All fish and shellfish (information on fish intake was mainly from 7-day dairy)	Elmstahl, et al. Eur J Clin Nutr. 1996; 50:134–142. Elmstahl, et al. Eur J Clin Nutr. 1996; 50:143–151. Riboli, et al. Int J Epidemiol. 1997;26 Suppl 1:S161–173. Callmer, et al. J Intern Med. 1993; 233:53–57.
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	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	Nettleton JA, et al. Am J Clin Nutr 2006;83:1369–79. Norris J et al Am J Public Health 1997;87:740–6. Mayer-Davis EJ, et al Ann Epidemiol 1999;9:314–24.
NEO	Dietary intake was assessed using a semi-quantitative food frequency questionnaire, originally validated in the Dutch general population	Dutch food composition table (NEVO) (version 2011)	Regular soda, lemonades and sportdrinks, excluding diet (light) beverages) as servings/d	All fruits	Cooked, baked and raw vegetables	N/A	All fish, including baked fish (salmon, soul), raw fish (herring), and shell fish	(Verkleij-Hagoort AC, *et al. Eur J Clin Nutr* 2006;61:610-5.)

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Whole Grain Intake	Fish Intake	Relevant References
NHS	131-item, self-reported, Willett FFQ	Harvard	Caffeinated colas, caffeine-free colas, carbonated noncola soft drinks, and noncarbonated sugar-sweetened beverages (fruit punches, lemonades, or other fruit drinks).	Apples or pears/ bananas/ peaches, apricots, plums/ oranges/ blue berry/ strawberry/grape/other fruit	Broccoli/ string beans, green beans/ egg plant/cauliflower, cabbage, Brussels sprouts/ spinach, collards, other greens/ corn/ tomatoes	Cooked cereals/ whole grain cold cereal/ dark or whole grain bread	Canned tuna fish/ dark meat fish/ other light meat fish	Giovannucci E, Colditz G, Stampfer MJ, et al. Am J Epidemiol 1991;133:810 –7. RimmEB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Am J Epidemiol 1992;135:1114 – 26.
Raine	FFQ	Cancer Council Australia	Soft drink, energy drink, fruit juice	Tinned fruit, oranges, apples, pears, bananas, melon, pineapple, strawberries, apricots, peaches, mango, avocado, pumpkin	Potatoes, tomatoes, capsicums, lettuce, cucumber, celery, beetroot, carrots, cauliflower, broccoli, spinach, peas, green beans, bean sprouts, baked beans, other beans, onion, garlic, mushrooms, zucchini	Wholemeal, mixed-grain or high fibre slicked bread, oatmeal, muesli, bran, wheat germ, other wholegrain breakfast cereals	Fish, fried fish, tinned fish	NA

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Whole Grain Intake	Fish Intake	Relevant References
RS1	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	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	Klipstein-Grobusch et al. Eur J Clin Nutr. 1998; 52(8):588-96
RS2	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	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	Goldbohm RA, et al. Eur J Clin Nutr 1994; 48(4):253-65 Feunekes GI, et al. Am J Clin Nutr 1993; 58(4):489-96

Cohort	Dietary Assessment Method	Nutrient Database	Sugar-Sweetened Beverages	Fruit Intake	Vegetable Intake	Whole Grain Intake	Fish Intake	Relevant References
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, quantified as servings/d	Grams per day	Grams per day	Servings per day	Grams per day	Paalanen L, et al. J Clin Epid. 2006;59(9):994 – 1001.

ESM Table 3. Assessment of additional characteristics of participating CHARGE cohorts.

Cohort	Fasting Glucose	Fasting Insulin	Education	Smoking Status	Physical Activity
ARIC	≥8-hour fasting blood samples were drawn from an antecubital vein into tubes containing a serum separator gel. Serum glucose concentrations were assessed with a hexokinase/glucose-6-phosphate dehydrogenase method. (units used in current models =mmol/l)	≥8-hour fasting insulin was quantified by radioimmunoassay (125Insulin Kit; Cambridge Medical Diagnosis, Billerica, MA), with a 7 pmol/l lower limit of sensitivity and 33% cross-reactivity with proinsulin (units used in current models = pmol/l).	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).
CHS	≥8-hour fasting glucose was quantified using a Kodak Ektachem 700 analyzer with reagents (Eastman Kodak, Rochester, NY). The overall CV was 1.86%, and the correlation coefficient between 169 pairs of blind replicates was 0.997.	≥8-hour fasting insulin was quantified by radioimmunoassay (Coat-A-Count Insulin assay (Diagnostics Products Corp, Los Angeles, CA)	Categorized into 3 groups: no high school degree, high school or vocational school degree, college degree	Classified as current, former, never smoker	Kilocalories of energy expended per week assessed with a questionnaire of leisure time physical activities
FHS	≥8-hour fasting plasma glucose was quantified with a hexokinase reagent kit (A-gent glucose test, Abbott Laboratories, South Pasadena, CA). Glucose assays were run in duplicate, and the intra-assay coefficient of variation ranged from 2–3%, depending on the assayed glucose concentration.	≥8-hour fasting insulin concentrations were quantified in plasma using human specific RIA at offspring exam 7 and using human-specific insulin ELISA in generation 3 exam 1 (both assays from Linco Inc., St. Louis, MO)	Categorized into 9 groups: no school, grades 1-8, grades 9-11, completed high school, some college but no degree, technical school certificate, associate degree, bachelor's degree, graduate or professional degree.	Regular cigarette smoking in the last year: classified as yes or no/never smoked	Physical activity score taking into account typical number of hours sleeping, sitting, slight activity, moderate activity, and heavy activity
MDC	Fasting glucose was quantified in overnight fasting whole blood samples by a hexokinase-glucose-6-phosphate dehydrogenase method. Blood glucose was converted to plasma glucose using a correction factor of 1.13.	Fasting insulin was quantified by non-specific radioimmunoassay in overnight fasting blood samples.	Categorized into 5 groups: elementary, primary and secondary, upper secondary, further education without a degree, university degree	Classified as current, former, never smoker	Leisure-time physical activity was obtained from a list of 17 different activities. The duration of each activity was multiplied by an intensity factor creating a score. The score was divided into six categories.

Cohort	Fasting Glucose	Fasting Insulin	Education	Smoking Status	Physical Activity
MESA	≥12 hour fasting glucose was measured at each examination by rate reflectance spectrophotometry using thin-film adaptation of the glucose oxidase method on the Vitros analyzer (Johnson & Johnson Clinical Diagnostics, Rochester, NY). Consistency of the serum glucose assay over examinations was established by reanalyzing 200 samples from each of the four examinations over a short time period and then recalibrating the original observations. The laboratory analytical CV was 1.1%.	Insulin was determined by a radioimmunoassay method using the Linco Human Insulin Specific RIA Kit (Linco Research Inc., St. Charles, MO). The laboratory analytical CV was 4.9%.	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
NEO	≥10-hour fasting glucose was measured using enzymatic colorimetry by Roche Modular p800 Analyzer, Roche Diagnostics, Mannheim, Germany.	≥10-hour fasting insulin was measured two-site chemilluminiscent immunometric assay by Siemens Immulite 2500, Siemens Healthcare, Breda, the Netherlands.	Categorized into 9 groups ranging from no education up to graduate school.	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
NHS	NA	≥8-hour fasting insulin was measured by using a radioimmunoassay specific for insulin with ,1% cross-reactivity between insulin and its precursors (Linco Research; CV: 9.5%).	NA	Classified as current, former, never smoker or missing/ unknown	Physical activity [metabolic equivalent tasks (h)/wk, tertiles]
Raine	Overnight fasting blood samples drawn from vein and assessed by an automated Technicon Axon analyser (Bayer Diagnostics, Sydney, Australia) using a hexokinase method	Overnight fasting blood samples drawn from vein and assessed by an automated radioimmunoassay (Tosoh, Tokyo, Japan)	Dichotomized as completed high school and did not complete high school (year 12)	Study participants answered to question: Do you smoke cigarettes? (Yes/No/Not stated)	NA
RS1/RS2	Fasting glucose was enzymatically determined using the Hexokinase method (Boehringer Mannheim) during the third visit to the research center (March 1997 - December 1999) for RS-I and during the third visit (February 2011-February 2012) for RS-II. Venous blood samples were taken at the research center after an overnight fast and stored at −80°C in a number of 5-ml aliquots. Serum glucose levels were determined within 1 week after sampling.	Fasting insulin levels were determined in samples that had been kept frozen from the third visit at the Research center (March 1997– December 1999 for RS-I and February 2011 – February 2012 for RS-II) until usage in 2008 and were measured on a Roche Modular Analytics E170 analyzer (Roche Diagnostics GmbH, Mannheim, Germany) by electrochemiluminescence immunoassay technology.	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

Cohort	Fasting Glucose	Fasting Insulin	Education	Smoking Status	Physical Activity
YFS	≥8-h fasting glucose concentrations were analysed enzymatically (Olympus Diagnostica GmbH, Hamburg, Germany).	≥8-h fasting serum insulin was quantified by microparticle enzyme immunoassay kit (Abbott Laboratories, Diagnostic Division, Dainabot, South Pasadena, CA).	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.

ESM Table 4. Description of selected *CHREBP-FGF21* pathway selected SNPs.

Gene	SNPs	Chr	References
ALDOB	rs10819937, rs10819931	9	(1)
FADS1	rs174546	11	(2, 3)
FGF21	rs838133	19	(4, 5)
GCK	rs4607517	7	(6, 7)
GCKR	rs1260326	2	(6, 7)
KHK	rs2119026	2	(8, 9)
KLB	rs1542423	4	(10)
MLXIPL	rs799166, rs799168, rs799160, rs11974409	7	(11, 12)
SLC2A2	rs11920090, rs11924032	3	(13, 14)
SLC2A5	rs5438, rs3820034, rs5840	1	(15)
TRIB1	rs2954029	8	(16)

References:

1. Santer R, Rischewski J, von Weihe M, Niederhaus M, Schneppenheim S, Baerlocher K, Kohlschutter A, Muntau A, Posselt HG, Steinmann B, et al. The spectrum of aldolase B (ALDOB) mutations and the prevalence of hereditary fructose intolerance in Central Europe. *Hum Mutat.* 2005;25(6):594.
2. Dentin R, Benhamed F, Pegorier JP, Foufelle F, Viollet B, Vaulont S, Girard J, and Postic C. Polyunsaturated fatty acids suppress glycolytic and lipogenic genes through the inhibition of ChREBP nuclear protein translocation. *J Clin Invest.* 2005;115(10):2843-54.
3. Tanaka T, Shen J, Abecasis GR, Kisialiou A, Ordovas JM, Guralnik JM, Singleton A, Bandinelli S, Cherubini A, Arnett D, et al. Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. *PLoS Genet.* 2009;5(1):e1000338.
4. Dushay JR, Toschi E, Mitten EK, Fisher FM, Herman MA, and Maratos-Flier E. Fructose ingestion acutely stimulates circulating FGF21 levels in humans. *Mol Metab.* 2015;4(1):51-7.
5. Iizuka K, Takeda J, and Horikawa Y. Glucose induces FGF21 mRNA expression through ChREBP activation in rat hepatocytes. *FEBS Lett.* 2009;583(17):2882-6.
6. Van Schaftingen E. A protein from rat liver confers to glucokinase the property of being antagonistically regulated by fructose 6-phosphate and fructose 1-phosphate. *Eur J Biochem.* 1989;179(1):179-84.
7. Agius L. Glucokinase and molecular aspects of liver glycogen metabolism. *Biochem J.* 2008;414(1):1-18.
8. Bonthron DT, Brady N, Donaldson IA, and Steinmann B. Molecular basis of essential fructosuria: molecular cloning and mutational analysis of human ketohexokinase (fructokinase). *Hum Mol Genet.* 1994;3(9):1627-31.
9. Hayward BE, and Bonthron DT. Structure and alternative splicing of the ketohexokinase gene. *Eur J Biochem.* 1998;257(1):85-91.
10. Adams AC, Cheng CC, Coskun T, and Kharitonenkov A. FGF21 requires betaklotho to act in vivo. *PLoS One.* 2012;7(11):e49977.
11. Koo HY, Wallig MA, Chung BH, Nara TY, Cho BH, and Nakamura MT. Dietary fructose induces a wide range of genes with distinct shift in carbohydrate and lipid metabolism in fed and fasted rat liver. *Biochim Biophys Acta.* 2008;1782(5):341-8.
12. Erion DM, Popov V, Hsiao JJ, Vatner D, Mitchell K, Yonemitsu S, Nagai Y, Kahn M, Gillum MP, Dong J, et al. The role of the carbohydrate response element-binding protein in male fructose-fed rats. *Endocrinology.* 2013;154(1):36-44.
13. Helliwell PA, Richardson M, Affleck J, and Kellett GL. Stimulation of fructose transport across the intestinal brush-border membrane by PMA is mediated by GLUT2 and dynamically regulated by protein kinase C. *Biochem J.* 2000;350 Pt 1(149-54.
14. Corpe CP, Basaleh MM, Affleck J, Gould G, Jess TJ, and Kellett GL. The regulation of GLUT5 and GLUT2 activity in the adaptation of intestinal brush-border fructose transport in diabetes. *Pflugers Arch.* 1996;432(2):192-201.
15. Burant CF, Takeda J, Brot-Laroche E, Bell GI, and Davidson NO. Fructose transporter in human spermatozoa and small intestine is GLUT5. *J Biol Chem.* 1992;267(21):14523-6.
16. Ishizuka Y, Nakayama K, Ogawa A, Makishima S, Boonvisut S, Hirao A, Iwasaki Y, Yada T, Yanagisawa Y, Miyashita H, et al. TRIB1 downregulates hepatic lipogenesis and glycogenesis via multiple molecular interactions. *J Mol Endocrinol.* 2014;52(2):145-58.

ESM Table 5. Genotyping information of participating CHARGE cohorts.

Cohort	Array	Imputed/Genotyped	Imputation Program
ARIC	ARIC samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, CA). Imputation was performed with MACH software.	Imputed	MACH (HapMap II CEU, NCBI 36)
CHS	CHS samples were genotyped using the Illumina HumanCNV370-Duo BeadChip system. Imputation was performed using BIMBAM10 v0.91 with reference to HapMap CEU using release 21A, build 35 using one round of imputations and the default expectation-maximization warm-ups and runs.	Imputed; except rs174546, rs838133, rs4607517 and rs1260326 (Genotyped)	BIMBAM (HapMap II CEU, NCBI 36)
FHS	Framingham study samples were genotyped using Affymetrix 500K (250K Nsp and 250K Sty) and MIPS 50K. Imputation was performed using MACH software. Ratio of variance of dosage to expected variance under binomial model: >0.3.	Imputed	MACH (1000G Phase I Integrated Release Version 3 Haplotypes, NCBI 37)
MDC	The samples were genotyped using the iPLEX Sequenom MassARRAY platform.	Genotyped	NA
MESA	MESA samples were genotyped using the Affymetrix Genome-Wide Human 6.0 array. Genotypes were defined using the Birdseed calling algorithm (Korn et al. 2008). Imputation was performed with IMPUTE software using 1000 Genomes as the reference panel.	Imputed; except rs10819931 and rs2119026 (Genotyped)	IMPUTE2 (1000 Genomes)
NEO	NEO samples were genotyped using the Illumina HumanCoreExome-24 Beadchip Kit at the Centre National de Génotypage (Paris, France). Imputation was performed with IMPUTE2 software. Both SNPs were imputed.	Imputed	IMPUTE2 (Illumina HumanCoreExome) (1000G)
NHS	DNA was extracted from the buffy coat fraction of centrifuged blood using a commercially available kit (QIAmp Blood kit; Qiagen, Chatsworth, California). Samples were genotyped and analysed using the Affymetrix Genome-Wide Human 6.0 array (Santa Clara, California) and the Birdseed calling algorithm.	Genotyped	MACH (HapMap II CEU, NCBI 36)
Raine	RAINE samples were genotyped using Illumina 660W Quad Array. Imputation was performed against the 1000G Phase 1 v3 reference using MACH/Minimac	Imputed	MACH (1000 Genomes)
RS1/2	RS samples were genotyped using the Array type Illumina 550K. Genotype calling was done using Illumina Bead studio. Any samples with a call rate<97.5%, excess autosomal heterozygosity >0.336 (~FDR <0.1%), mismatch between called and phenotypic gender, or if there were outliers identified by the IBS clustering analysis with >3 standard deviations from population mean or IBS probabilities >97% were excluded from the analysis.	Imputed	MACH (HapMap release 22 (build 36)
YFS	Genotyping was performed at the Sanger Institute (UK) using the custom-built Illumina BeadChip Human670K. Genotypes were called using Illumina's clustering algorithm. SNPs that were present on HapMap and that passed quality control measures were used for imputation with MACH version 1.0.	Imputed	MACH 1.0 (HapMap II CEU, NCBI 36)

ESM Table 6. Effect allele frequencies for investigated SNPs in participating CHARGE cohorts.

				Discovery Cohorts						Replication Cohorts				
SNP	**Chr**	**Nearest Gene**	**Allele (effect/ noneffect)**	**CHS**	**FHS**	**MESA**	**RS1**	**RS2**	**YFS**	**NHS**	**ARIC**	**MDC**	**NEO**	**Raine**
rs10819937	9	*ALDOB*	C/G	0.161	0.171	0.185	0.180	0.182	0.200					
rs10819931	9	*ALDOB*	T/C	NA	0.075	0.068	0.068	0.066	0.062					
rs174546	11	*FADS1*	T/C	0.324	0.329	0.328	0.334	0.322	0.408					
rs838133	19	*FGF21*	A/G	0.443	0.463	0.416	0.443	0.419	0.391					
rs4607517	7	*GCK*	A/G	0.189	0.180	0.171	0.176	0.190	0.104					
rs1260326	2	*GCKR*	C/T	0.562	0.548	0.564	0.612	0.627	0.649					
rs2119026	2	*KHK*	C/T	0.390	0.390	0.387	0.378	0.383	0.333					
rs1542423	4	*KLB*	T/C											
All				0.706	0.670	0.667	0.691	0.694	0.674	0.685	0.689	0.718[a]	0.706	0.6835
Male				0.713	0.658	0.628	NA	NA	0.674	NA	0.689	0.716[a]	0.706	0.689
Female				0.701	0.672	0.628	NA	NA	0.674	0.685	0.689	0.719[a]	0.712	0.684
rs799166	7	*CHREBP/MLXIPL*	C/G	NA	NA	0.878	0.873	0.868	0.882					
rs799168	7	*CHREBP/MLXIPL*	G/A	NA	NA	0.785	0.792	0.788	0.804					
rs799160	7	*CHREBP/MLXIPL*	T/C	NA	NA	0.533	0.504	0.498	0.571					
rs11974409	7	*TBL2*	A/G	NA	NA	0.824	0.806	0.802	0.836					
rs11920090	3	*SLC2A2*	A/T	0.136	0.132	0.135	0.136	0.143	0.138					
rs11924032	3	*SLC2A2*	A/G	0.262	0.278	0.271	0.259	0.271	0.217					
rs5438	1	*SLC2A5*	A/G	0.055	0.048	0.060	0.036	0.042	0.040					
rs3820034	1	*SLC2A5*	C/T	0.786	0.809	0.798	0.788	0.789	0.773					
rs5840	1	*SLC2A5*	T/C	0.333	0.329	0.332	0.365	0.366	0.360					
rs2954029	8	*TRIB1*	A/T	0.547	0.561	0.452	0.522	0.524	0.551					

[a] Proxy SNP used rs1979283

Abbreviations: NA, not available/applicable.

ESM Table 7. General Characteristics of participating CHARGE cohorts stratified by sex.

Male

	Discovery Cohorts						Replication Cohorts				
	CHS	FHS	MESA	RS1	RS2	YFS	ARIC	MDC	NEO	NHS	Raine
Characteristics											
n[a]	1035	2507	721	1150	493	761	4099	1643	2415	N/A	334
Age (yrs)	72.9 (5.7)	48.4 (13.3)	69.90 (9.40)	64.7 (6.3)	61.7 (5.0)	37.66 (5.03)	54.6 (5.6)	57.7 (6.1)	56 (6.0)	N/A	20.1 (0.520)
BMI (kg/m^2)	26.1 (3.4)	27.9 (4.3)	27.91 (4.38)	25.9 (2.8)	26.9 (3.12)	26.43 (3.99)	27.24 (3.87)	25.7 (3.2)	29.4 (3.8)	N/A	24.47 (4.487)
Current Smoker (%)	10.14	15.32%	7.39%	309 (26.9%)	117 (23.7%)	34.1	23.9% (n=981)	27.9	19.2%	N/A	15.27
Completed High School (%)	74.93	92.07%	97.36%	510 (44.7%) [1142]	350 (72.3%) [484]	41.0	82.9% (n=3399)	21.70%	94.2%	N/A	74.85
Fasting Glucose (mmol/l)	5.63 (0.50)	5.4 (0.48)	5.29 (0.53)	5.8 (1.1)	5.8 (1.1)	5.42 (0.45)	5.60 (0.50)	5.67 (0.51)	5.6 (0.5)	N/A	5.074 (0.488)
Fasting Insulin (pmol/l)	95.01 (46.03)	57.16 (45.45)	52.81 (38.20)	78.38 (73.27)	88.06 (58.21)	51.79 (39.63)	79.7 (57.6)	47.11 (31.18)	91.2 (56.4)	N/A	32.38 (41.47)
Dietary Intake											
Energy Intake (kJ/d)	8845 (2791)	8891 (2874)	7832 (3176)	9485 (2138)	9134 (2879)	11196 (3527)	7540 (2720)	11510 (2586)	10711 (3259)	N/A	9473 (4159)
SFA (% total energy)	10.81 (2.15)	11.09 (2.90)	11.35 (3.22)	14.1 (3.0)	11.8 (2.8)	12.12 (2.38)	12.34 (3.05)	16.8 (4.0)	12.3 (2.8)	N/A	16.2 (3.0)
SSB intake (servings/d) [c]	0.18 (0.29)	0.5 (0.8)	0.19 (0.54)	0.14 (0.28)	0.25 (0.35)	0.39 (0.55)	0.59 (0.94)	0.35 (0.63)	0.51 (0.87)	N/A	1.151 (1.2)
Low SSB intake (≤1 serving/d) (%)	98.6	88.8	96.5	98.6	97.8	91.7	84.1	88.3	83.4	N/A	60.5
High SSB intake (>1 serving/d) (%)	1.4	11.2	3.5	1.4	2.2	8.3	15.9	11.7	16.6	N/A	39.5
Fruit intake (servings/d)	2.41 (1.4)	0.18 (0.34)	1.88 (1.6)	1.4 (0.9)	2.1 (2.3)	157.39 (145.56)[b]	1.88 (1.4)	1.7 (1.2)	1.0 (0.9)	N/A	216 (159.5))[b]
Vegetable Intake (servings/d)	2.52 (1.4)	0.56 (0.56)	2.27 (1.4)	3.6 (1.4)	3.9 (3.1)	219.10 (150.76))[b]	1.67 (1.2)	2.4 (1.3)	2.65 (1.4)	N/A	136.9 (76.4))[b]
Whole Grain Intake (servings/d)	0.98 (0.6)	1.18 (1.12)	1.04 (0.8)	5.0 (3.0)	3.8 (2.3)	3.24 (2.07)	1.39 (1.4)	2.0 (2.1)	N/A	N/A	85.4 (83.0))[b]
Fish Intake (servings/d)	0.28 (0.3)	0.18 (0.18)	0.14 (0.2)	0.15 (0.2)	0.3 (0.3)	43.86 (31.19))[b]	0.27 (0.3)	0.58 (0.5)	0.17 (0.16)	N/A	44.4 (57.3))[b]
Alcohol Intake (g/d)	8.57 (16.7)	15.4 (18.5)	12.44 (18.7)	16.2 (17.2)	16.0 (16.8)	13.5 (17.7)	70.82 (120.0)	15.5 (15.1)	21.3 (19.9)	N/A	20.7 (21.3)

Female

	Discovery Cohorts						Replication Cohorts				
	CHS	FHS	MESA	RS1	RS2	YFS	ARIC	MDC	NEO	NHS	Raine
Characteristics											
n[a]	1709	3070	786	1634	582	976	4737	2332	2717	676	371
Age (yrs)	71.9 (5.1)	48.7 (13.6)	70.03 (9.57)	65.3 (6.8)	61.5 (4.9)	37.77 (4.98)	53.8 (5.7)	57.7 (5.9)	56 (5.9)	57.6 (7.0)	20.0 (0.491)
BMI (kg/m^2)	25.9 (4.8)	26.3 (5.7)	27.11 (5.58)	26.4 (3.8)	27.1 (4.07)	25.13 (4.75)	26.29 (5.22)	24.9 (3.8)	29.8 (5.3)	26.1 (4.9)	24.54 (5.658)
Current Smoker (%)	12.35	14.46%	7.02%	298 (18.2%)	111 (19.1%)	22.5	25.14% (n=1190)	25.6	13.0	9.9	14.82
Completed High School (%)	77.01	93.42%	96.95%	1149 (70.7%) [1625]	509 (88.5%) [575]	24.2	86.5% (n=4009)	19.10%	94.5	NA	86.25
Fasting Glucose (mmol/l)	5.47 (0.53)	5.1 (0.49)	5.08 (0.50)	5.7 (1.09)	5.6 (0.9)	5.12 (0.46)	5.37 (0.50)	5.43 (0.50)	5.4 (0.5)	NA	4.843 (0.367)
Fasting Insulin (pmol/l)	92.91 (50.46)	51.65 (39.46)	48.61 (31.17)	88.6 (101.11)	86.58 (72.86)	48.53 (36.22)	67.8 (54.0)	41.00 (33.25)	78.3 (51.2)	51.1 (37.1)	32.99 (35.194)
Dietary Intake											
Energy Intake (kJ/d)	8201 (2636)	7715 (2536)	6519 (2761)	7464 (1628)	7916 (2586)	9000 (2552)	6280 (2180)	8912 (1753)	8573 (2728)	7410 (2029)	6598 (3992)
SFA (% total energy)	9.97 (2.21)	11.15 (3.0)	11.14 (3.48)	14.2 (3.2)	11.7 (3.2)	11.48 (2.27)	12.04 (3.08)	16.5 (3.8)	12.5 (3.0)	10.6 (2.6)	15.385 (3.2)
SSB intake (servings/d) [c]	0.11 (0.23)	0.2 (0.5)	0.11 (0.48)	0.08 (0.21)	0.19 (0.28)	0.22 (0.36)	0.37 (0.82)	0.25 (0.47)	0.27 (0.70)	0.19 (0.4)	0.8212 (1.2)
Low SSB intake (≤1 serving/d) (%)	99.5	96.1	98.2	99.5	97.8	94.7	90.8	5.8	92.2	96.3	24.0
High SSB intake (>1 serving/d) (%)	0.5	3.9	1.8	0.5	2.2	5.3	9.2	94.2	7.8	3.7	76.0
Fruit intake (servings/d)	2.93 (1.51)	0.23 (0.4)	2.29 (1.76)	1.7 (0.9)	2.8 (2.5)	254.84 (215.0))[b]	2.19 (1.50)	2.2 (1.2)	1.3 (0.9)	1.78 (1.2)	176.91 (121.2))[b]
Vegetable Intake (servings/d)	3.02 (1.52)	0.70 (0.7)	2.44 (1.48)	3.6 (1.8)	4.1 (3.2)	280.84 (165.82))[b]	1.78 (1.17)	2.6 (1.3)	2.92 (1.63)	3.57 (1.9)	126.7 (57.8))[b]
Whole Grain Intake (servings/d)	1.04 (0.66)	1.25 (1.2)	0.97 (0.73)	3.9 (2.0)	3.3 (1.9)	3.22 (1.72)	1.33 (1.13)	1.9 (1.6)	N/A	2.0 (1.7)	75.77 (67.6))[b]
Fish Intake (servings/d)	0.35 (0.31)	0.18 (0.2)	0.15 (0.18)	0.14 (0.17)	0.3 (0.3)	32.36 (19.87))[b]	0.32 (0.33)	0.53 (0.36)	0.15 (0.15)	0.30 (0.2)	31.36 (51.5))[b]
Alcohol Intake (g/d)	4.01 (10.1)	7.1 (10.1)	6.23 (11.64)	6.6 (9.9)	8.6 (11.2)	4.8 (6.5)	24.67 (52.64)	7.2 (7.7)	10.2 (12.1)	5.3 (8.9)	13.893 (16.2)

[a] Maximum available observations, *n*, for interactions between SSB intake and single nucleotide polymorphisms in glucose outcomes analyses. Sample sizes vary in some cohorts depending on availability of genotype information.
Values are means (standard deviations) or percentages (%).
[b] in grams/day
[c] One serving is equivalent to 360 ml (12 fl oz.)
Abbreviations: BMI, body mass index; *n*, total sample size; SFA, saturated fatty acids; SSB, sugar-sweetened beverages.

ESM Table 8. Meta-analysed main associations between SNPs and glycaemic trait outcomes[a].

SNP	Chr	Gene	Alleles[b]	Fasting Glucose (mmol/l)		Fasting Insulin (log_e-pmol/l)	
				β (SE)	*p*	β (SE)	*p*
rs10819937	9	*ALDOB*	C/G	0.0028 (0.0088)	0.75	-0.0060 (0.0087)	0.49
rs10819931	9	*ALDOB*	T/C	-0.0105 (0.0149)	0.48	-0.0122 (0.0159)	0.44
rs174546	11	*FADS1*	T/C	-0.0228 (0.0066)	$6.00x10^{-4}$***	0.0099 (0.0065)	0.13
rs838133	19	*FGF21*	A/G	0.0003 (0.0072)	0.97	0.0043 (0.0069)	0.53
rs4607517	7	*GCK*	A/G	0.0723 (0.0084)	$5.75x10^{-18}$***	0.0036 (0.0080)	0.66
rs1260326	2	*GCKR*	C/T	0.0255 (0.0064)	$6.00x10^{-5}$***	0.0244 (0.0063)	$1.00x10^{-4}$***
rs2119026	2	*KHK*	C/T	0.0077 (0.0066)	0.24	0.0010 (0.0064)	0.88
rs1542423	4	*KLB*	T/C	0.0178 (0.0072)	0.013*	0.0132 (0.0072)	0.07
rs799166	7	*CHREBP/MLXIPL*	C/G	0.0005 (0.0123)	0.97	-0.0026 (0.0125)	0.84
rs799168	7	*CHREBP/MLXIPL*	G/A	0.0074 (0.0106)	0.49	-0.0049 (0.0109)	0.65
rs799160	7	*CHREBP/MLXIPL*	T/C	0.0011 (0.0091)	0.90	-0.0082 (0.0092)	0.37
rs11974409	7	*TBL2*	A/G	-0.0070 (0.092)	0.45	0.0030 (0.0094)	0.75
rs11920090	3	*SLC2A2*	A/T	-0.0313 (0.0091)	$6.00x10^{-4}$***	0.0111 (0.0091)	0.22
rs11924032	3	*SLC2A2*	A/G	-0.0144 (0.0072)	0.047*	0.0018 (0.007)	0.80
rs5438	1	*SLC2A5*	A/G	0.0016 (0.0168)	0.92	0.0070 (0.0170)	0.68
rs3820034	1	*SLC2A5*	C/T	-0.0034 (0.0088)	0.70	0.0009 (0.0089)	0.92
rs5840	1	*SLC2A5*	T/C	0.0044 (0.0066)	0.66	-0.0082 (0.0065)	0.21
rs2954029	8	*TRIB1*	A/T	0.0039 (0.0064)	0.54	0.0059 (0.0063)	0.35

[a] Additive allele mode, adjusted for age, sex, study site for multi-centered cohorts (in CHS; MESA; YFS), and family or population structure (in FHS; MESA; RS1; RS2; YFS). Association coefficients are shown as β (SE). β represents the direction and magnitude of the change in outcome trait per each additional copy of the effect allele.

[b] Alleles presented as effect/noneffect alleles.

*p<0.05, **p<0.01, ***p<0.001

Abbreviations: SNP, single nucleotide polymorphism.

ESM Table 9. Meta-analysed associations between rs1542423 and glycaemic traits[a].

	Discovery Cohorts				Replication Cohorts				All Cohorts			
	n	*β* (SE)	*p*	I^2	*n*	*β* (SE)	*p*	I^2	*n*	*β* (SE)	*p*	I^2
All	15590	0.0132 (0.0072)	0.07	0%	18338	0.0031 (0.0071)	0.66	68%	33928	0.0081 (0.0051)	0.11	38%
Male	6392	0.0194 (0.0110)	0.08	8.7%	8238	-0.0131 (0.0098)	0.18	56.0%	14630	0.0014 (0.0073)	0.8529	47.6%
Female	8398	0.0101 (0.0095)	0.29	0%	10477	0.0081 (0.0099)	0.42	60.8%	18875	0.0091 (0.0069)	0.1839	33.3%

[a] Additive allele mode, adjusted for age, sex, family or population structure (in MESA; RS1; RS2; YFS), and study site (in FHS; CHS; MESA; YFS). Association coefficients are shown as *β* (SE). *β* represents the direction and magnitude of the change in outcome trait per each additional copy of the effect allele.
**p*<0.05, ***p*<0.01, ****p*<0.001
Abbreviations: SNP, single nucleotide polymorphism.

ESM Table 10. Meta-analysed main associations between SNPs and fasting glucose (mmol/l) stratified by sex[a].

				Male		Female	
SNP	Chr	Gene	Alleles[b]	β (SE)	*p*	β (SE)	*p*
rs10819937	9	*ALDOB*	C/G	0.0234 (0.0146)	0.11	-0.0157 (0.0129)	0.22
rs10819931	9	*ALDOB*	T/C	-0.0292 (0.0235)	0.22	-0.0067 (0.0201)	0.74
rs174546	11	*FADS1*	T/C	-0.0178 (0.011)	0.11	-0.0238 (0.0096)	**0.01**
rs838133	19	*FGF21*	A/G	-0.0034 (0.0123)	0.78	-0.0016 (0.0106)	0.88
rs4607517	7	*GCK*	A/G	0.0732 (0.0132)	3.01×10^{-8}***	0.0723 (0.0117)	6.80×10^{-10}***
rs1260326	2	*GCKR*	C/T	0.0132 (0.0105)	0.21	0.0327 (0.0091)	3.11×10^{-4}***
rs2119026	2	*KHK*	C/T	0.0119 (0.011)	0.28	0.009 (0.0093)	0.33
rs1542423	4	*KLB*	T/C	0.0096 (0.0119)	0.42	0.0232 (0.0104)	0.03*
rs11920090	3	*SLC2A2*	A/T	-0.0349 (0.0151)	0.02*	-0.0193 (0.0133)	0.15
rs11924032	3	*SLC2A2*	A/G	-0.0165 (0.0118)	0.16	-0.0014 (0.0103)	0.89
rs5438	1	*SLC2A5*	A/G	0.0332 (0.0281)	0.24	-0.0038 (0.0244)	0.88
rs3820034	1	*SLC2A5*	C/T	0.0079 (0.0149)	0.59	-0.0115 (0.014)	0.41
rs5840	1	*SLC2A5*	T/C	0.0081 (0.011)	0.46	-0.0015 (0.0096)	0.87
rs2954029	8	*TRIB1*	A/T	0.0012 (0.0107)	0.91	0.0013 (0.0092)	0.89

[a]Additive allele mode, adjusted for age, sex, study site for multi-centered cohorts (in CHS and MESA), and family or population structure (in FHS; MESA; RS1; RS2). Association coefficients are shown as β (SE). β represents the direction and magnitude of the change in outcome trait per each additional copy of the effect allele.

[b] Alleles presented as effect/noneffect alleles.

*p<0.05, **p<0.01, ***p<0.001

Abbreviations: SNP, single nucleotide polymorphism.

ESM Table 11. Meta-analysed main associations between SNPs and fasting insulin (log_e-pmol/l) stratified by sex[a].

				Male		Female	
SNP	Chr	Gene	Alleles[b]	β (SE)	*p*	β (SE)	*p*
rs10819937	9	*ALDOB*	C/G	0.0153 (0.0135)	0.26	-0.021 (0.0119)	0.08
rs10819931	9	*ALDOB*	T/C	0.0069 (0.0247)	0.78	-0.0324 (0.0207)	0.12
rs174546	11	*FADS1*	T/C	0.0169 (0.0101)	0.09	0.0046 (0.0088)	0.6
rs838133	19	*FGF21*	A/G	0.0045 (0.011)	0.68	0.0066 (0.0093)	0.48
rs4607517	7	*GCK*	A/G	0.0021 (0.0122)	0.86	0.0052 (0.0107)	0.63
rs1260326	2	*GCKR*	C/T	0.0307 (0.0097)	$1.60x10^{-3}$**	0.0226 (0.0084)	$7.11x10^{-3}$**
rs2119026	2	*KHK*	C/T	0.006 (0.01)	0.55	0.0001 (0.0085)	0.99
rs1542423	4	*KLB*	T/C	0.0166 (0.0112)	0.14	0.0141 (0.0098)	0.15
rs11920090	3	*SLC2A2*	A/T	-0.0098 (0.014)	0.48	0.0268 (0.0122)	0.03*
rs11924032	3	*SLC2A2*	A/G	-0.0118 (0.0109)	0.28	0.0089 (0.0093)	0.34
rs5438	1	*SLC2A5*	A/G	-0.0118 (0.0265)	0.65	0.0368 (0.023)	0.11
rs3820034	1	*SLC2A5*	C/T	-0.0009 (0.0138)	0.95	0.0001 (0.0137)	1.00
rs5840	1	*SLC2A5*	T/C	-0.0052 (0.0101)	0.61	-0.0098 (0.0088)	0.26
rs2954029	8	*TRIB1*	A/T	-0.0082 (0.0098)	0.4	0.0163 (0.0084)	0.05

[a] Additive allele mode, adjusted for age, sex, study site for multi-centered cohorts (in CHS and MESA), and family or population structure (in FHS; MESA; RS1; RS2). Association coefficients are shown as β (SE). β represents the direction and magnitude of the change in outcome trait per each additional copy of the effect allele.

[b] Alleles presented as effect/noneffect alleles.

*p<0.05, **p<0.01, ***p<0.001

Abbreviations: SNP, single nucleotide polymorphism.

ESM Table 12. Meta-analysed interactions between SSB intake and SNPs on fasting glucose (mmol/l) stratified by sex[a].

				Male		Female	
SNP	Chr	Gene	Alleles[b]	β (SE)	*p*	β (SE)	*p*
rs10819937	9	*ALDOB*	C/G	-0.0258 (0.0203)	0.20	0.031 (0.0277)	0.26
rs10819931	9	*ALDOB*	T/C	-0.0212 (0.0279)	0.45	-0.032 (0.0377)	0.40
rs174546	11	*FADS1*	T/C	0.0132 (0.0159)	0.41	0.0082 (0.0212)	0.70
rs838133	19	*FGF21*	A/G	-0.0264 (0.0176)	0.13	0.0007 (0.0247)	0.98
rs4607517	7	*GCK*	A/G	-0.0003 (0.0195)	0.99	0.0075 (0.0241)	0.76
rs1260326	2	*GCKR*	C/T	0.0047 (0.0145)	0.74	-0.0212 (0.0182)	0.24
rs2119026	2	*KHK*	C/T	-0.0151 (0.0166)	0.36	-0.0045 (0.0212)	0.83
rs1542423	4	*KLB*	T/C	-0.0161 (0.015)	0.28	-0.0096 (0.0195)	0.62
rs11920090	3	*SLC2A2*	A/T	0.0057 (0.0229)	0.80	0.0582 (0.0292)	0.05
rs11924032	3	*SLC2A2*	A/G	0.0161 (0.0153)	0.29	0.0352 (0.0231)	0.13
rs5438	1	*SLC2A5*	A/G	0.0146 (0.0395)	0.71	0.0342 (0.0493)	0.49
rs3820034	1	*SLC2A5*	C/T	0.0187 (0.0204)	0.36	0.0159 (0.0255)	0.53
rs5840	1	*SLC2A5*	T/C	0.0175 (0.0155)	0.26	0.0054 (0.02)	0.79
rs2954029	8	*TRIB1*	A/T	-0.0009 (0.0159)	0.95	0.0346 (0.0186)	0.06

[a] Additive allele mode, adjusted for age, sex, BMI, study site for multi-centered cohorts (in CHS and MESA), and family or population structure where applicable (in FHS; MESA; RS1; RS2). Interaction coefficients are shown as β (SE). β represents the direction and magnitude of the change in outcome trait with each additional effect allele, per each additional serving/d of SSB intake.

[b] Alleles presented as effect/noneffect alleles.

*p<0.05, **p<0.01, ***p<0.001

Abbreviations: BMI, body mass index; Chr, chromosome; SNP, single nucleotide polymorphism; SSB, sugar-sweetened beverages.

ESM Table 13. Meta-analysed interactions between SSB intake and SNPs on fasting insulin (log_e-pmol/l) stratified by sex[a].

				Male		Female	
SNP	Chr	Gene	Alleles[b]	β (SE)	*p*	β (SE)	*p*
rs10819937	9	*ALDOB*	C/G	-0.0222 (0.0192)	0.25	-0.0044 (0.0253)	0.86
rs10819931	9	*ALDOB*	T/C	-0.0199 (0.0269)	0.46	0.0564 (0.0351)	0.11
rs174546	11	*FADS1*	T/C	-0.0061 (0.0149)	0.68	0.0174 (0.0195)	0.37
rs838133	19	*FGF21*	A/G	-0.0477 (0.0168)	0.005**	0.013 (0.0228)	0.57
rs4607517	7	*GCK*	A/G	-0.0006 (0.0187)	0.97	0.05 (0.0226)	0.03*
rs1260326	2	*GCKR*	C/T	0.0008 (0.0137)	0.95	-0.0322 (0.0169)	0.06
rs2119026	2	*KHK*	C/T	0.0222 (0.0153)	0.15	-0.0264 (0.0198)	0.18
rs1542423	4	*KLB*	T/C	0.0258 (0.0143)	0.07	0.0291 (0.0182)	0.11
rs11920090	3	*SLC2A2*	A/T	0.0185 (0.0215)	0.39	-0.0095 (0.0277)	0.73
rs11924032	3	*SLC2A2*	A/G	-0.0101 (0.0145)	0.48	0.0143 (0.0223)	0.52
rs5438	1	*SLC2A5*	A/G	0.0037 (0.0366)	0.92	0.0089 (0.0459)	0.85
rs3820034	1	*SLC2A5*	C/T	0.0089 (0.0193)	0.65	0.0029 (0.0233)	0.9
rs5840	1	*SLC2A5*	T/C	-0.0071 (0.0147)	0.63	0.0205 (0.0183)	0.26
rs2954029	8	*TRIB1*	A/T	0.0178 (0.015)	0.24	-0.0209 (0.0175)	0.23

[a] Additive allele mode, adjusted for age, sex, BMI, study site for multi-centered cohorts (in CHS and MESA), and family or population structure where applicable (in FHS; MESA; RS1; RS2). Interaction coefficients are shown as β (SE). β represents the direction and magnitude of the change in outcome trait with each additional effect allele, per each additional serving/d of SSB intake.

[b] Alleles presented as effect/noneffect alleles.

*p<0.05, **p<0.01, ***p<0.001

Abbreviations: BMI, body mass index; Chr, chromosome; SNP, single nucleotide polymorphism; SSB, sugar-sweetened beverages.

ESM Table 14. Meta-analysed interactions between SSB intake and rs838133 on fasting insulin (log_e-pmol/l) in male, and SSB intake and rs4607517 on fasting insulin (log_e-pmol/l) in female[a].

			Discovery Cohorts			Replication Cohorts		
	SNP	Alleles[b]	*n*	*β* (SE)	*p*	*n*	*β* (SE)	*p*
Male	rs838133	A/G	5,631	-0.0477 (0.0168)	0.005***	6,684	0.0091 (0.0136)	0.50
Female	rs4607517	A/G	7,544	0.05 (0.0226)	0.03*	8,312	-0.0256 (0.0195)	0.19

[a] Additive allele mode, adjusted for age, sex, BMI, study site for multi-centered cohorts (in CHS and MESA), and family or population structure where applicable (in FHS; MESA; RS1; RS2; ARIC; RAINE). Interaction coefficients are shown as *β* (SE). *β* represents the direction and magnitude of the change in outcome trait with each additional effect allele, per each additional serving/d of SSB intake.

[b] Alleles presented as effect/non-effect alleles.

*p<0.05, **p<0.01, ***p<0.001

Abbreviations: SNP, single nucleotide polymorphism.

ESM Table 15. Meta-regression results and subgroup fixed-effect meta-analysed interactions between SSB intake and fasting insulin (log_e-pmol/l)[a].

Model	Moderator p-value	Group 1			Group 2		
		β (SE)	p	I^2	β (SE)	p	I^2
		Cohort Mean BMI (<27 kg/m^2)			Cohort Mean BMI ≥27 kg/m^2)		
Model 4[e]	0.044	0.040 (0.014)	0.0034	59%	0.020 (0.007)	0.0027	0%
		Cohort Mean Age (<60 years)			Cohort Mean Age (≥60 years)		
Model 2[c]	0.046	0.036 (0.001)	$2.4x10^{-4}$	61%	0.075 (0.026)	0.0037	44%
Model 3[d]	0.044	0.031 (0.008)	$3.5x10^{-5}$	49%	0.068 (0.017)	$4.5x10^{-5}$	0%
		Cohort Sample Size (n <1000)			Cohort Sample Size (n ≥1000)		
Model 4[e]	0.040	0.020 (0.007)	0.70	64%	0.035 (0.007)	$2.0x10^{-7}$	29%

[a] Association coefficients are shown as βs (SEs). β represents the change in fasting insulin (log_e-pmol/l) per additional serving/d of sugar-sweetened beverages. Cohort distribution as follows: BMI (<27 kg/m^2: ARIC, CHS, MDC, NHS, Raine, RS1, YFS; ≥27 kg/m^2: FHS, MESA, NEO, RS2); Age (<60 years: ARIC, FHS, MDC, NEO, NHS, Raine, YFS; ≥60 years: CHS, MESA, RS1, RS2); and sample size (n <1000: NHS, Raine; n ≥1000: ARIC, CHS, FHS, MDC, MESA, NEO, RS1, RS2, YFS)

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

[c] Model 2: adjusted for Model 1 covariates and smoking status, education status, physical activity (except in RS1; RS2), 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, fish intake, and saturated fatty acids (% of total energy).

ESM Table 16. Random-effect meta-analysed main associations between SSB intake and glycaemic traits and interactions between SSB and SNPs on glycaemic traits in discovery cohorts[a].

	Fasting Glucose (mmol/l)				Fasting Insulin (log_e-pmol/l)			
Model	n[f]	β (SE)	p	I^2	n[f]	β (SE)	p	I^2
Discovery Cohorts								
Model 1[b]	16739	0.013 (0.015)	0.38	34%	16304	0.044 (0.018)	$1.6x10^{-2}$	39%
Model 2[c]	16097	0.009 (0.015)	0.52	32%	15668	0.047 (0.022)	$3.3x10^{-2}$	65%
Model 3[d]	16024	0.011 (0.008)	0.20	0%	15594	0.045 (0.014)	$1.1x10^{-3}$	35%
Model 4[e]	15885	0.009 (0.009)	0.29	0%	15467	0.031 (0.011)	$4.7x10^{-3}$	13%
Replication Cohorts								
Model 1[b]	18719	0.020 (0.005)	$2.9x10^{-6}$	0%	19265	0.057 (0.009)	$9.0x10^{-10}$	45%
Model 2[c]	18527	0.019 (0.009)	$8.6x10^{-5}$	0%	19075	0.045 (0.010)	$1.0x10^{-5}$	53%
Model 3[d]	18508	0.015 (0.005)	$1.4x10^{-3}$	0%	19059	0.033 (0.010)	$1.5x10^{-3}$	63%
Model 4[e]	18505	0.015 (0.005)	$2.3x10^{-3}$	0%	19056	0.030 (0.012)	0.015	69%
All Cohorts								
Model 1[b]	35458	0.017 (0.005)	$3.6x10^{-4}$	15%	35569	0.051 (0.009)	$1.3x10^{-8}$	51%
Model 2[c]	34624	0.016 (0.005)	$1.4x10^{-3}$	14%	34743	0.044 (0.010)	$9.5x10^{-6}$	60%
Model 3[d]	34532	0.014 (0.004)	$6.6x10^{-4}$	0%	34653	0.037 (0.007)	$6.5x10^{-7}$	46%
Model 4[e]	34390	0.014 (0.004)	$1.5x10^{-3}$	0%	34523	0.031 (0.008)	$8.0x10^{-4}$	48%

[a] Association coefficients are shown as βs (SEs). β represents the change in outcome per additional serving/d of sugar-sweetened beverages.

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

[c] Model 2: adjusted for Model 1 covariates and smoking status, education status, physical activity (except in RS1; RS2), 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, fish intake, and saturated fatty acids (% of total energy).

[f] The number of independent observations in each association analysis.

ESM Table 17. GWA meta-analysis association results between SNPs and carbohydrate intake from CHARGE macronutrient investigation[a].

				Carbohydrate Intake (% energy)	
SNP	Chr	Gene	Alleles[b]	β (SE)	*p*
rs10819937	9	*ALDOB*	C/G	0.0547 (0.079)	0.4892
rs10819931	9	*ALDOB*	T/C	0.0997 (0.1379)	0.4697
rs174546	11	*FADS1*	T/C	0.0093 (0.0596)	0.8762
rs838133	19	*FGF21*	A/G	0.3823 (0.0673)	1.34×10^{-8}
rs4607517	7	*GCK*	A/G	-0.0262 (0.075)	0.7269
rs1260326	2	*GCKR*	C/T	-	-
rs2119026	2	*KHK*	A/G	-0.0009 (0.059)	0.9875
rs1542423	4	*KLB*	T/C	-0.0552 (0.0646)	0.3928
rs799166	7	*CHREBP/MLXIPL*	C/G	-0.0927 (0.0962)	0.3352
rs799168	7	*CHREBP/MLXIPL*	G/A	-	-
rs799160	7	*CHREBP/MLXIPL*	T/C	-	-
rs11974409	7	*TBL2*	A/G	-	-
rs11920090	3	*SLC2A2*	A/T	-0.1122 (0.0834)	0.1786
rs11924032	3	*SLC2A2*	A/G	-0.0901 (0.0649)	0.1651
rs5438	1	*SLC2A5*	T/C	0.0229 (0.1422)	0.8722
rs3820034	1	*SLC2A5*	C/T	-0.0479 (0.0756)	0.5263
rs5840	1	*SLC2A5*	T/C	-0.0177 (0.0643)	0.7832
rs2954029	8	*TRIB1*	A/T	0.026 (0.0573)	0.6499

[a] Analysis adjusted for age, sex, BMI, and study-specific covariates (eg study site, populations stratification principal components when applicable). Association coefficients are shown as β (SE). β represents the direction and magnitude of the change in % energy from carbohydrate intake per each additional copy of the effect allele.
[b] Alleles presented as effect/noneffect alleles.
Abbreviations: BMI, body mass index; CHARGE, Cohorts for Heart and Aging Research in Genomic Epidemiology, GWA, genome-wide association.

Reference:

Tanaka T., Ngwa J.S., van Rooij F.J.A., Zillikens M.C., Wojczynski M.K., Frazier-Wood A.C., Houston D.K., Kanoni S., Lemaitre R.N., Luan J., et al. Genome-wide meta-analysis of observational studies reveals common genetic variants associated with macronutrient intake. *Am. J. Clin. Nutr.* 2013;97(6):1395-402.

ESM Table 18. Comparison of General Characteristics between participating Discovery and Replication CHARGE cohorts[a].

	Discovery Cohorts	Replication Cohorts	*p*[b]
Characteristics			
n	15424	19324	
Age (yrs)	57.6 (8.9)	54.2 (5.7)	<0.0001
Sex (% women)	56.8	56.1	0.1912
BMI (kg/m^2)	26.6 (4.5)	27.1 (4.4)	<0.0001
Current Smoker (%)	16.6	21.9	<0.0001
Completed High School (%)	77.2	80.9	<0.0001
Fasting Glucose (mmol/l)	5.4 (0.7)	5.4 (0.5)	1
Fasting Insulin (pmol/l) †	68.1 (53.0)	70.3 (49.9)	<0.0001
Dietary Intake			
SSB intake (servings/d) †	0.2 (0.5)	0.4 (0.8)	<0.0001
Low SSB intake (≤1 serving/d) (%)	95.9	87.2	<0.0001
High SSB intake (>1 serving/d) (%)	4.1	12.8	<00001
Energy Intake (kJ/d)	8393 (2636)	8301 (2745)	0.0017
SFA (% total energy)	11.7 (2.8)	12.7 (3.1)	<0.0001
Fruit intake (servings/d)	1.4 (1.0)	1.7 (1.3)	<0.0001
Vegetable Intake (servings/d)	2.1 (1.3)	2.3 (1.6)	<0.0001
Whole Grain Intake (servings/d)	2.1 (1.4)	1.6 (1.5)	<0.0001
Fish Intake (servings/d)	0.2 (0.2)	0.3 (0.3)	<0.0001
Alcohol Intake (g/d)	9.5 (14.4)	10.1 (14.2)	<0.0001

[a] Values are averages of discovery/replication cohorts' general characteristics (Table 1) and presented as mean (SD) for continuous traits and % for dichotomous traits. We've excluded cohorts with different methods for ascertainment of trait from the computed mean.

[b] Computed by using t-tests for continuous traits and chi-square tests for dichotomous traits.