Genetic Variants in SGLT1, Glucose Tolerance, and Cardiometabolic Risk

Seidelmann, Sara B.

2018-10-09


http://hdl.handle.net/10138/251200
https://doi.org/10.1016/j.jacc.2018.07.061

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.
Genetic Variants in SGLT1, Glucose Tolerance, and Cardiometabolic Risk

Sara B. Seidelmann, MD, PhD,a,b Elena Feofanova, MS,c Bing Yu, PhD,d Nora Franceschini, MD, e Brian Claggett, PhD, a Mikko Kuokkanen, PhD,e,f Hannu Puolijoki, MD, PhD,g Tapani Ebeling, MD, PhD,h Markus Perola, MD, PhD,e,f Veikko Salomaa, MD, PhD,e Amil Shah, MD,a Josef Coresh, MD, PhD,i Elizabeth Selvin, PhD, MPH,i Calum A. MacRae, MD, PhD,a Susan Cheng, MD,a Eric Boerwinkle, PhD,c,j Scott D. Solomon, MDa

ABSTRACT

BACKGROUND Loss-of-function mutations in the SGLT1 (sodium/glucose co-transporter-1) gene result in a rare glucose/galactose malabsorption disorder and neonatal death if untreated. In the general population, variants related to intestinal glucose absorption remain uncharacterized.

OBJECTIVES The goal of this study was to identify functional SGLT1 gene variants and characterize their clinical consequences.

METHODS Whole exome sequencing was performed in the ARIC (Atherosclerosis Risk in Communities) study participants enrolled from 4 U.S. communities. The association of functional, nonsynonymous substitutions in SGLT1 with 2-h oral glucose tolerance test results was determined. Variants related to impaired glucose tolerance were studied, and Mendelian randomization analysis of cardiometabolic outcomes was performed.

RESULTS Among 5,687 European-American subjects (mean age 54 ± 6 years; 47% male), those who carried a haplotype of 3 missense mutations (frequency of 6.7%)—Asn51Ser, Ala411Thr, and His615Gln—had lower 2-h glucose and odds of impaired glucose tolerance than noncarriers (β-coefficient: −8.0; 95% confidence interval [CI]: −12.7 to −3.3; OR: 0.71; 95% CI: 0.59 to 0.86, respectively). The association of the haplotype with oral glucose tolerance test results was consistent in a replication sample of 2,791 African-American subjects (β = −16.3; 95% CI: −36.6 to 4.1; OR: 0.39; 95% CI: 0.17 to 0.91) and an external European-Finnish population sample of 6,784 subjects (β = −3.2; 95% CI: −6.4 to −0.02; OR: 0.81; 95% CI: 0.68 to 0.98). Using a Mendelian randomization approach in the index cohort, the estimated 25-year effect of a reduction of 20 mg/dl in 2-h glucose via SGLT1 inhibition would be reduced prevalent obesity (OR: 0.43; 95% CI: 0.23 to 0.63), incident diabetes (hazard ratio [HR]: 0.58; 95% CI: 0.35 to 0.81), heart failure (HR: 0.53; 95% CI: 0.24 to 0.83), and death (HR: 0.66; 95% CI: 0.42 to 0.90).

CONCLUSIONS Functionally damaging missense variants in SGLT1 protect from diet-induced hyperglycemia in multiple populations. Reduced intestinal glucose uptake may protect from long-term cardiometabolic outcomes, providing support for therapies that target SGLT1 function to prevent and treat metabolic conditions. (J Am Coll Cardiol 2018;72:1763–73) © 2018 Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
A
fter ingestion, complex carbohydrates are enzymatically broken down to produce monosaccharides (glucose, galactose, and fructose), which are absorbed in the small intestine and used as substrate for the body’s metabolically active tissues. The sodium/glucose co-transporter (SGLT)-1 protein is a rate-limiting factor for absorption of glucose and galactose in the small intestine, and it uses transmembrane sodium gradients to drive the cellular uptake of these molecules. Loss-of-function mutations, including missense, nonsense, and frameshift mutations, of the SGLT1 gene result in impaired cellular glucose transport and cause glucose-galactose malabsorption (GGM). GGM is a rare, autosomal recessive, Mendelian disorder resulting in neonatal onset of severe diarrhea, dehydration, and malabsorption (1-3).

Functional gene variants in SGLT1 associated with altered glucose metabolism in the general population have not been described. However, in the process of identifying causal mutations for GGM, SGLT1 gene variants that are associated with subtle abnormalities of glucose absorption in vivo have been identified; the importance of these variants, which do not result in GGM, is unknown (1). We hypothesized that rare or low-frequency variants in SGLT1 that are predicted to be damaging, but still preserve some of the protein’s function, result in lower postprandial blood glucose levels by decreasing glucose uptake in the small intestine and thereby reduce overall caloric absorption. Rare or low-frequency variant analysis of SGLT1 has not been systematically studied in the general population and could provide insights into the gene’s relationship with diet and disease, as well as potential drug targets for hyperglycemia and obesity and its consequences. We assessed the relationship between protein-altering, nonsynonymous (NS) substitutions in SGLT1 and glycemia and obesity by using whole exome sequencing performed in participants from 4 U.S. communities who were enrolled in the ARIC (Atherosclerosis Risk In Communities) study. External validation was then performed in a European-Finnish population sample (FINRISK and DILGOM [Dietary, Lifestyle and Genetic Predictors of Obesity and Metabolic Syndrome] study). A Mendelian randomization approach was then used to estimate the effect of SGLT1 inhibition on long-term cardiometabolic outcomes.

METHODS

STUDY DESIGN AND STUDY POPULATION. Originally comprising 15,792 men and women aged between 45 and 64 years, the ARIC study is an ongoing, prospective observational trial of atherosclerosis risk factors in 4 U.S. communities (Forsyth County, North Carolina; Jackson, Mississippi; suburbs of Minneapolis, Minnesota; and Washington County, Maryland) recruited between 1987 and 1989 (visit 1) (4). ARIC study participants were initially examined every 3 years, with the second examination occurring in the years 1990 to 1992, the third in 1993 to 1995, the fourth in 1996 to 1998, and the fifth in 2011 to 2013. At each participating site, institutional review boards
approved the study protocol. Informed consent was obtained in writing at each study visit. Participants were excluded from this analysis if they restricted use of their deoxyribonucleic acid (DNA) (n = 44).

**MEASUREMENT OF DIET.** At the baseline visit (1987 to 1989), participants were interviewed by using a 66-item semi-quantitative food frequency questionnaire (FFQ), modified from a 61-item validated FFQ (5). Participants reported the occurrence with which they consumed individual foods or beverages in 9 categories (ranging from never or <1 time per month to ≥6 times per day). Nutrient intakes were derived from FFQ responses using the Harvard Nutrient Database.

**PHENOTYPIC MEASUREMENTS.** Age, sex, race (self-reported), total cholesterol (measured by using blood plasma assay), high-density lipoprotein (measured by blood plasma assay), current smoking status (self-reported), hypertension (HTN) (defined as antihypertensive medications or blood pressure >140/90 mm Hg), estimated glomerular filtration rate, chronic kidney disease (defined as an estimated glomerular filtration rate <60 ml/min/1.73 m²), and obesity (defined as a body mass index ≥30 kg/m²) were analyzed at baseline visit. Baseline systolic blood pressure was adjusted for ≥15 mm Hg and diastolic blood pressure ≥10 mm Hg if the subject was taking antihypertensive medication (assessed by self-reported questionnaire). A standard 75-g oral glucose tolerance test (OGTT) was performed after an overnight fast before ARIC visit 4 (1996 to 1998). Participants who were taking diabetes mellitus (DM) medications at visit 4, had a recent history of taking DM medications (at visit 3), reported not fasting for at least 12 h, or had had part of their stomach removed did not qualify to undergo OGTT. Impaired glucose tolerance (IGT) was defined as a 2-h postprandial glucose level ≥140 mg/dl to associate all individuals with potentially abnormal physiological handling of glucose to gene variants of interest. All participants who underwent OGTT were included in the current analysis.

Sensitivity analyses were performed excluding individuals with a diagnosis of DM or with a fasting plasma glucose level >126 mg/dl. All analytes were determined at central laboratories according to standard protocols. Plasma glucose was measured by using a hexokinase assay.

**OUTCOMES.** Outcomes evaluated in this study are incident DM diagnosis, initiation of DM medications, death, and incident heart failure (HF), from the first visit until 2015 (median 25-year follow-up). Because heritable functional variations in SGLT1 could theoretically influence an individual’s success in the dietary management of DM, a diagnosis of DM and initiation of DM medications were explored separately as outcomes. Patients at risk of HF are also at risk of cardiovascular death before development of overt HF; thus, both the individual outcomes as well as a composite of the first occurrence of HF or death was assessed. The methods used for the detection and adjudication of events have been presented previously (6–8). Incident death was determined via death certificates, hospitalized patients, physician questionnaires, and next-of-kin interviews. Incident HF was defined as the first HF hospitalization identified by International Statistical Classification of Diseases and Related Health Problems (ICD) codes 428 (for ICD-9) or 150 (for ICD-10) in any position on hospital discharge or deaths coded as HF (428 or 150) in any position on death certificates (8). Incident cases of DM were identified based on self-reported DM diagnosis or use of DM medications during the scheduled ARIC visits as well as annual telephone interviews conducted with participants for a maximum of 15 years of follow-up (7).

**WHOLE EXOME SEQUENCING.** Whole exome sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center (HGSC), using the HGSC VCRome 2.1 reagent (9) (42 Mb, NimbleGen). All samples underwent paired-end sequencing by using GAII or HiSeq instruments (Illumina, San Diego, California). Atlas2 suite was used to perform the variant calling (10). The DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer’s protocol. The complete protocol and oligonucleotide sequences are accessible from the Baylor College of Medicine-HGSC website. Barcoded pre-capture libraries were pooled and then hybridized to NimbleGen exome capture array (HGSC VCRome 2.1 design [9] (42 Mb, NimbleGen) and sequenced in paired-end mode in a single lane on the Illumina HiSeq 2000 platform. Illumina sequence analysis was performed by using the HGSC Mercury analysis pipeline. Pooled samples were de-multiplexed by using Consensus Assessment of Sequence and Variation software (Illumina). Reads were then mapped to the Genome Reference Consortium Human Build 37 human reference sequence using Burrows-Wheeler Aligner (11) producing binary alignment/map files. Aligned reads were then recalibrated by using the Genome Analysis Toolkit (12) along with binary alignment/map sorting, duplicate read marking, and realignment near insertions or deletions. The Atlas2 (10) suite was used to call variants and produce high-quality variant call files (13).
Each single nucleotide variant call was filtered based on the following criteria to produce a high-quality variant list: low single nucleotide variant posterior probability (<0.95), low variant read count (<3), variant read ratio <0.25 or >0.75, strand-bias of >99% variant reads in a single strand direction, or total coverage <10-fold. All variant calls filtered according to these criteria, and reference calls with <10-fold coverage, were set to missing. The variant call filters were the same for insertions or deletions except a total coverage <30-fold was used. Variant-level quality control steps excluded variants outside the exon capture regions (VChrome 2.1), multi-allelic sites, missing rate >20%, and mean depth of coverage >500-fold. Variants not meeting Hardy-Weinberg equilibrium expectations in ancestry-specific groups (p < 5 × 10^-6) were also excluded. Sample-level quality control metrics were calculated according to cohort and ancestry group. A sample was excluded for missingness >20%, or, if compared with the other samples, it fell <6 SDs for mean depth, >6 SDs for singleton count, or outside of 6 SDs for heterozygote to homozygote ratio or a transition-to-transversion ratio. We selected all NS variants in the SGLT1 gene for this study. Protein function was determined from previous in vivo studies of heterologous expression that determined glucose uptake in cells expressing SGLT1 mutant proteins (1).

EXTERNAL VALIDATION COHORTS. FINRISK is a series of cardiovascular risk factor surveys conducted in Finland every 5 years since 1972 (14). DNA has been collected since the 1992 survey. The FINRISK 2002 study is a stratified, random sample of the population 25 to 74 years of age from 6 geographical areas of Finland (15). The original population sample was 13,500, and the overall participation rate was 65.5% (n = 8,798). Participation included a questionnaire and health examination in which blood samples were drawn. Testing for OGTT with fasting and 2-h glucose assessment was conducted in a separate session for participants 45 to 74 years of age (n = 3,738) approximately 2 months after the baseline survey.

The DILGOM study is an extension of the FINRISK 2007 survey and is based on an independent population sample, separate from FINRISK 2002 (16). All participants of FINRISK 2007 were further invited to take part in a more detailed study (i.e., DILGOM). The participation rate was 80% (n = 5,024). The baseline examination of DILGOM in 2007 was conducted about 3 months after the original FINRISK 2007 survey. It included a 2-h OGTT with measurements of fasting and 2-h glucose assessment.

Of the FINRISK and DILGOM participants with OGTT, a total of 6,784 had data on IGT and 2-h glucose assessments available for the present analysis. The analysis on prevalent obesity included FINRISK-92, -97, -02, -07, and -12 cohorts (total N = 27,294). All analyses were performed by using directly genotyped or 1,000 Genomes phase 3 imputed data (quality >0.99).

STATISTICAL ANALYSIS. Baseline characteristics are reported by using descriptive statistics for continuous variables and number counts and percentages for categorical variables. Linear regression and chi-square tests for trend were used to test for a relationship between continuous or categorical baseline variables, respectively, and SGLT1 genotype. All analyses were adjusted for age at the time that the measurement was obtained, sex, and, in a combined analysis, the first 10 principal components from EIGENSTRAT (17) to minimize confounding by ancestry. Genotype-phenotype associations were first examined in a sample of European-American subjects with exome sequencing genotypes and then replicated in African-American subjects. To estimate the contribution of an individual missense mutation to the overall NS variant signal, conditional analyses were performed. These evaluated the association of all NS variants with IGT, further adjusting for each individual missense mutation as a covariate. Survival analysis methods were used for the following outcomes: diagnosis of DM, initiation of DM medications, death, and HF. Follow-up time was calculated as the time from the first examination to the time of the first event or the last known date of follow-up (up to December 31, 2014).

Kaplan-Meier failure curves were created for each outcome, and hazard ratios (HRs) were calculated by using Cox proportional hazards regression. These analyses were first tested in an unadjusted model and then adjusted for age, sex, and the first 10 principal components from EIGENSTRAT (17) to minimize confounding by ancestry. The causal effects of reduced 2-h glucose values on outcomes and quantitative traits were examined according to conventional Mendelian randomization (2-stage least-squares estimator method), which uses predicted values of 2-h glucose per genotype and regresses each outcome or trait against these predicted values. Models were tested for each outcome variable and were adjusted for age, sex, and the first 10 principal components. All data were analyzed by using STATA version 14.0 (StataCorp, College Station, Texas). Values of p < 0.05 were considered statistically significant.
RESULTS

Whole exome sequencing was performed in 8,478 participants in the ARIC cohort; 5,687 were European American and 2,791 were African American. A total of 915 (16%) European-American subjects and 208 (7.5%) African-American subjects carried at least 1 NS substitution in SGLT1 (Table 1, Online Table 1). Overall, baseline characteristics of the study population were similar between those who carried at least 1 NS substitution and those who did not. However, in European-American subjects, diastolic blood pressure, use of antihypertensive medications, prevalent HTN, uric acid levels, and heart rate were lower in those who carried at least 1 NS substitution in SGLT1. The mean number of NS substitutions in European-American and African-American subjects was 0.4 and 0.1, respectively.

We first tested whether overall NS variants were associated with phenotypes most likely to be correlated with SGLT1 function in ARIC European-American subjects, given its larger size. Carrying one or more copies of an NS variant was associated with lower fasting and postprandial plasma glucose levels, lower risk of IGT, and obesity in European-American subjects (Online Table 2). Next, to identify specific NS variants that may have contributed to the overall NS substitution results, we tested the relationship of individual variants with IGT, which was the strongest detected association. All participants who underwent OGTT were included to study potentially abnormal physiological handling of glucose in SGLT1 variants. Sensitivity analyses excluding individuals with a DM diagnosis or those with fasting plasma glucose levels >126 mg/dl yielded similar results. Three variants (N51S, A411T, and H615Q) were significantly associated with a 25% to 27% reduction in IGT (odds ratio [OR]: 0.73, 0.72, and 0.75, respectively) (Central Illustration, Table 2). These variants were in high linkage disequilibrium, comprising a risk haplotype with a minor allele frequency of 6.7% (Online Table 3). Conditional analysis showed that the association of NS substitutions with IGT was completely explained by the

| TABLE 1 | Baseline Characteristics of European-American and African-American Participants in the ARIC Cohort Stratified According to the Presence of 1 or >2 NS Substitutions in SGLT1 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | European-American Subjects | | African-American Subjects | | European-American Subjects | | African-American Subjects | |
| | Control NS = 0 | SGLT1 NS = 1 | SGLT1 NS = 2 | p Value for Trend | Control NS = 0 | SGLT1 NS = 1 | SGLT1 NS = 2 | p Value for Trend |
| Patients | 4,772 (84) | 105 (2) | 810 (14) | 2,583 (92.5) | 124 (4.5) | 84 (3) |
| Age, yrs | 54 ± 6 | 54 ± 6 | 54 ± 6 | 0.83 | 53 ± 6 | 54 ± 6 | 54 ± 6 | 0.03 |
| Male | 2,268 (48) | 44 (42) | 385 (48) | 0.88 | 933 (36) | 46 (37) | 30 (36) | 0.97 |
| Smoking | 1,174 (25) | 18 (17) | 187 (23) | 0.25 | 724 (28) | 41 (33) | 28 (33) | 0.13 |
| Current EtOH | 3,262 (68) | 80 (76) | 561 (69) | 0.44 | 794 (31) | 45 (37) | 30 (36) | 0.15 |
| TC, mg/dl | 215 ± 39 | 216 ± 35 | 215 ± 38 | 0.95 | 213 ± 41 | 215 ± 51 | 214 ± 39 | 0.52 |
| HDL, mg/dl | 51 ± 17 | 54 ± 18 | 51 ± 17 | 0.42 | 55 ± 18 | 54 ± 17 | 56 ± 18 | 0.97 |
| LDL, mg/dl | 137 ± 36 | 138 ± 33 | 138 ± 35 | 0.71 | 136 ± 38 | 138 ± 46 | 137 ± 36 | 0.58 |
| Triglycerides, mg/dl | 139 ± 98 | 129 ± 118 | 132 ± 87 | 0.07 | 112 ± 85 | 116 ± 61 | 115 ± 67 | 0.59 |
| SBP, mm Hg | 122.1 ± 19.0 | 120.7 ± 18.3 | 120.8 ± 18.7 | 0.06 | 133.4 ± 21.2 | 133.7 ± 23.1 | 135.7 ± 25.8 | 0.37 |
| DBP, mm Hg | 74.2 ± 11.3 | 73.8 ± 10.8 | 73.3 ± 11.1 | 0.028 | 83.6 ± 12.9 | 84.2 ± 15.6 | 81.3 ± 15.1 | 0.29 |
| Antihypertensive medication | 1,200 (25) | 45 (26) | 156 (21) | 0.029 | 1,080 (42) | 52 (42) | 41 (49) | 0.23 |
| HTN | 1,262 (26) | 46 (27) | 164 (22) | 0.021 | 1,363 (53) | 68 (55) | 47 (56) | 0.45 |
| Urinary acid, mg/dl | 5.17 ± 1.52 | 5.11 ± 1.60 | 4.96 ± 1.39 | <0.001 | 5.43 ± 1.63 | 5.73 ± 1.67 | 5.47 ± 2.07 | 0.26 |
| eGFR | 89.3 ± 17.9 | 88.5 ± 18.6 | 88.9 ± 16.6 | 0.46 | 103.4 ± 26.4 | 101.8 ± 25.2 | 109.9 ± 29.5 | 0.13 |
| CKD | 145 (3.0) | 4 (2.3) | 16 (2.2) | 0.17 | 55 (2.2) | 4 (3.3) | 0 (0.0) | 0.47 |
| Heart rate, beats/min | 66.5 ± 10.0 | 64.9 ± 9.9 | 65.5 ± 9.4 | 0.005 | 66.8 ± 10.8 | 67.6 ± 12.1 | 65.1 ± 10.1 | 0.43 |
| Total energy intake, kcal | 1,644 ± 602 | 1,553 ± 498 | 1,623 ± 597 | 0.28 | 1,580 ± 614 | 1,696 ± 642 | 1,515 ± 568 | 0.85 |
| Sodium intake, mg | 1,513 ± 601 | 1,529 ± 611 | 1,553 ± 639 | 0.28 | 1,354 ± 544 | 1,428 ± 536 | 1,303 ± 571 | 0.99 |
| Sucrose intake, g | 51 ± 36 | 43 ± 25 | 49 ± 33 | 0.05 | 58 ± 37 | 65 ± 45 | 58 ± 37 | 0.30 |
| Glucose intake, g | 22 ± 14 | 20 ± 11 | 22 ± 14 | 0.03 | 24 ± 16 | 27 ± 17 | 24 ± 13 | 0.65 |
| Fructose intake, g | 24 ± 17 | 21 ± 12 | 24 ± 16 | 1.00 | 27 ± 18 | 30 ± 21 | 26 ± 16 | 0.72 |
| Lactose intake, g | 14 ± 13 | 14 ± 11 | 16 ± 14 | 0.04 | 10 ± 10 | 9 ± 8 | 9 ± 8 | 0.41 |

Values are n (%) or mean ± SD.

ARIC = Atherosclerosis Risk in Communities Study; CKD = chronic kidney disease (defined as estimated glomerular filtration rate <60 ml/min/1.73 m²); DBP = diastolic blood pressure (adjusted for the use of antihypertensive medication); eGFR = estimated glomerular filtration rate; EtOH = alcohol; HDL = high-density lipoprotein cholesterol; HTN = hypertension (defined as ≥ antihypertensive medications or blood pressure ≥140/90 mm Hg); LDL = low-density lipoprotein cholesterol; NS = nonsynonymous; SBP = systolic blood pressure (adjusted for the use of antihypertensive medication); SGLT1 = sodium/glucose co-transporter-1; TC = total cholesterol.
Whole exome sequencing of Sgl1 was performed in ARIC (Atherosclerosis Risk in Communities) study participants enrolled from 4 U.S. communities. Among 5,687 European-American subjects (mean age 54 ± 6 years), those who carried a haplotype of 3 missense mutations (Asn51Ser, Ala411Thr, and His615Gln) had lower 2-h glucose and odds of impaired glucose tolerance than noncarriers. The association of the haplotype with oral glucose tolerance test results was consistent in a replication sample of 2,791 African-American subjects and an external European-Finnish population sample of 6,784. Using a Mendelian randomization approach in the index cohort, we estimated the 25-year effect of a 20 mg/dl reduction in 2-h glucose via sodium/glucose cotransporter-1 (SGLT1) inhibition on prevalent obesity, incident diabetes, heart failure, and death. DILGOM = Dietary, Lifestyle and Genetic Predictors of Obesity and Metabolic Syndrome.
N51S/A411T/H615Q haplotype, suggesting that this haplotype was responsible for the original NS signal. Complete characterization of the N51S/A411T/H615Q haplotype in ARIC European-American subjects revealed a significant association with lower levels of postprandial glucose (β = −8.0; 95% confidence interval [CI]: −12.7 to −3.3), risk for IGT (OR: 0.71; 95% CI: 0.59 to 0.86), and prevalent obesity (OR: 0.79; 95% CI: 0.66 to 0.96) (Table 3). To further validate our primary findings, we examined the associations with OGTT phenotypes in an independent sample of individuals of European ancestry living in Finland and participating in the FINRISK and DILGOM studies (mean age 54 ± 12 years; body mass index 27.4 ± 4.8 kg/m²; 47% male). The SGLT1 haplotype (minor allele frequency of 5.6%) was significantly associated with lower 2-h glucose values (β = −3.2; 95% CI: −6.4 to −0.02) and IGT (OR: 0.80; 95% CI: 0.74 to 0.91), and displayed a directionally consistent relationship with obesity in this sample.

We further tested the association of the N51S/A411T/H615Q haplotype with OGTT and obesity in ARIC African-American subjects (minor allele frequency of 1.5%) (Online Table 3). The haplotype was significantly associated with lower risk of IGT (OR: 0.39; 95% CI: 0.17 to 0.91) and obesity (OR: 0.60; 95% CI: 0.37 to 0.98) and a trend for 2-h glucose values (β = −16.3; 95% CI: −36.6 to 4.1) in that population (Central Illustration, Table 3). A combined analysis included ARIC European-American subjects, ARIC African-American subjects, and FINRISK/DILGOM European Finnish subjects with additional adjustment for principal components as covariates to minimize confounding by ancestry. The N51S/A411T/H615Q haplotype was significantly associated with 2-h glucose levels (β = −4.9; 95% CI: −7.5 to −2.3; p = 2.4 × 10⁻⁴), lower risk for IGT (OR: 0.75; 95% CI: 0.66 to 0.85; p = 1.1 × 10⁻⁴) as well as prevalent obesity (OR: 0.80; 95% CI: 0.70 to 0.93; p = 2.5 × 10⁻³) without a significant effect on fasting plasma glucose values (Table 3). The association of IGT and haplotype remained significant after including obesity as a covariate in the ARIC populations (OR: 0.74; 95% CI: 0.61 to 0.90).

To test the association of the N51S/A411T/H615Q haplotype with disease phenotypes that may be

---

**TABLE 2** List of NS Variants From Exome Sequencing of SGLT1 in ARIC European-American and African-American Subjects

<table>
<thead>
<tr>
<th>Allele</th>
<th>Count</th>
<th>Chrom</th>
<th>Position</th>
<th>RSID</th>
<th>Ref</th>
<th>Alt</th>
<th>Amino Acid Change</th>
<th>Transcript Consequence</th>
<th>CADD Raw</th>
<th>CADD Phred</th>
<th>Association of Single Variant With IGT* OR (95% CI)</th>
<th>Conditional Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>European-American subjects (n = 5,687)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>834</td>
<td>22</td>
<td>32506050</td>
<td>rs33954001</td>
<td>C G</td>
<td>p.His615Gln</td>
<td>c.1845C&gt;G</td>
<td>0.74</td>
<td>11.22</td>
<td>0.75</td>
<td>0.63-0.90</td>
<td>0.69</td>
<td>0.52-0.93</td>
</tr>
<tr>
<td>829</td>
<td>22</td>
<td>32445946</td>
<td>rs17683011</td>
<td>A G</td>
<td>p.Asn61Ser</td>
<td>c.152A&gt;G</td>
<td>1.23</td>
<td>14.55</td>
<td>0.73</td>
<td>0.61-0.87</td>
<td>0.81</td>
<td>0.59-1.11</td>
</tr>
<tr>
<td>773</td>
<td>22</td>
<td>32487000</td>
<td>rs17683430</td>
<td>G A</td>
<td>p.Ala411Thr</td>
<td>c.1231G&gt;A</td>
<td>2.94</td>
<td>22.80</td>
<td>0.72</td>
<td>0.60-0.87</td>
<td>0.77</td>
<td>0.54-1.09</td>
</tr>
<tr>
<td>23</td>
<td>22</td>
<td>32479102</td>
<td>rs15017594</td>
<td>G T</td>
<td>p.Val209Leu</td>
<td>c.625G&gt;T</td>
<td>0.76</td>
<td>11.39</td>
<td>1.00</td>
<td>0.38-2.62</td>
<td>0.83</td>
<td>0.75-0.92</td>
</tr>
<tr>
<td>19</td>
<td>22</td>
<td>32506051</td>
<td>rs16733910</td>
<td>G A</td>
<td>p.Gly615Ser</td>
<td>c.1845G&gt;A</td>
<td>1.14</td>
<td>13.97</td>
<td>1.07</td>
<td>0.38-3.08</td>
<td>0.83</td>
<td>0.75-0.92</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>32439318</td>
<td>rs33951240</td>
<td>T C</td>
<td>p.VAL7Ala</td>
<td>c.50T&gt;C</td>
<td>0.17</td>
<td>5.82</td>
<td>1.00</td>
<td>0.17-6.04</td>
<td>0.83</td>
<td>0.75-0.92</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>32439303</td>
<td>rs15028967</td>
<td>C T</td>
<td>p.Ala12Val</td>
<td>c.35C&gt;T</td>
<td>0.41</td>
<td>8.63</td>
<td>1.57</td>
<td>0.31-7.96</td>
<td>0.83</td>
<td>0.75-0.92</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>32505988</td>
<td>rs147453689</td>
<td>C A</td>
<td>p.Pro595Thr</td>
<td>c.1783C&gt;G</td>
<td>0.80</td>
<td>11.66</td>
<td>4.78</td>
<td>0.49-46.3</td>
<td>0.83</td>
<td>0.75-0.92</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>32439369</td>
<td>rs139760182</td>
<td>T C</td>
<td>p.Ile34Thr</td>
<td>c.101T&gt;C</td>
<td>0.101</td>
<td>2.50</td>
<td>0.62</td>
<td>0.41-0.95</td>
<td>0.62</td>
<td>0.41-0.95</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>32482251</td>
<td>rs141412905</td>
<td>G A</td>
<td>p.Gly356Ser</td>
<td>c.1066G&gt;A</td>
<td>0.80</td>
<td>11.66</td>
<td>4.78</td>
<td>0.49-46.3</td>
<td>0.83</td>
<td>0.75-0.92</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>32439357</td>
<td>rs201800716</td>
<td>C T</td>
<td>p.Ser30Phe</td>
<td>c.89C&gt;T</td>
<td>0.89</td>
<td>1.24</td>
<td>0.62</td>
<td>0.41-0.95</td>
<td>0.62</td>
<td>0.41-0.95</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>32479100</td>
<td>rs33951240</td>
<td>C T</td>
<td>p.Thr208Met</td>
<td>c.625G&gt;C</td>
<td>0.50</td>
<td>1.04</td>
<td>0.92</td>
<td>0.69-1.11</td>
<td>0.92</td>
<td>0.69-1.11</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>32479101</td>
<td>rs33954001</td>
<td>C G</td>
<td>p.His615Gln</td>
<td>c.1845C&gt;G</td>
<td>0.74</td>
<td>11.22</td>
<td>0.75</td>
<td>0.63-0.90</td>
<td>0.69</td>
<td>0.52-0.93</td>
</tr>
</tbody>
</table>

24 additional variants in European-American subjects and 15 additional variants in African-American subjects only represented in a single individual are not listed. *Adjusted for age and sex. †Conditional analysis tests the association of all NS variants with impaired glucose tolerance (IGT) on oral glucose tolerance test adjusting for each individual mutation as a covariate. This estimates the contribution of the individual mutation in the overall NS signal. IGT was defined as 2-h glucose level ≥140 mg/dL.

Alt = alternate nucleic acid; CADD = Conditional Annotation Dependent Depletion; Chrom = chromosome; CI = confidence interval; OR = odds ratio; Ref = reference nucleic acid; other abbreviations as in Table 1.
affected by alteration of the gene’s function, we focused on long-term clinical outcomes using the combined ARIC European- and African-American samples. After a 25-year median follow-up, there were 1,943 incident cases of DM, 1,556 incident HF events, and 3,319 deaths. The N51S/A411T/H615Q haplotype was significantly associated with a reduced incidence of diagnosis of DM and initiation of DM medications in unadjusted models (HR: 0.77 [95% CI: 0.66 to 0.89], p = 3.8 × 10⁻⁴; HR: 0.72 [95% CI: 0.61 to 0.85], respectively) and in age-, sex-, and principal component-adjusted models (HR: 0.81 [95% CI: 0.69 to 0.94], p = 0.006; HR: 0.78 [95% CI: 0.66 to 0.92], p = 0.004) (Online Figure 1). This haplotype was also associated with reduced risk of death or HF (HR: 0.80; 95% CI: 0.71 to 0.90; p = 2.5 × 10⁻⁴) in the combined ARIC sample, after adjustment for age, sex, and principal components (p = 8.9 × 10⁻⁵). This finding was also observed when assessing the outcomes of HF and death separately in both unadjusted models (HR: 0.71 [95% CI: 0.58 to 0.87], p = 6.4 × 10⁻⁴; HR: 0.81 [95% CI: 0.71 to 0.93], p = 0.0018) and adjusted models (HR: 0.74 [95% CI: 0.60 to 0.91], p = 0.0046; HR: 0.81 [95% CI: 0.71 to 0.94], p = 0.0049).

We performed Mendelian randomization analysis, instrumented by the SGLT1 haplotype, of 2-h glucose values with cardiometabolic outcomes to estimate the potential effect of pharmacological SGLT1 inhibition on disease phenotypes. A decrease of 20 mg/dl in 2-h glucose levels, instrumented by the SGLT1 haplotype, was associated with lower odds of prevalent obesity (OR: 0.43; 95% CI: 0.23 to 0.62) after adjustment for age, sex, and principal components (Table 4). In long-term outcomes, each decrease of 20 mg/dl in 2-h glucose levels corresponded to a reduced incidence of DM (HR: 0.58; 95% CI: 0.35 to 0.81), HF (HR: 0.53; 95% CI: 0.24 to 0.83), and death (HR: 0.66; 95% CI: 0.42 to 0.90) after adjustment for age, sex, and principal components.

**DISCUSSION**

In European-American subjects and in a separate replication sample of African-American subjects, individuals with NS mutations in SGLT1 had a lower risk of IGT and obesity despite equivalent estimated intake of total calories, sodium, and sugars at baseline, and that this outcome was primarily driven by the N51S/A411T/H615Q haplotype (Central Illustration). We further found that individuals with this haplotype had a lower incidence of DM, initiation of DM therapies, death, and HF. These data suggest that genetic variations in this gene influence the risk of metabolic disease in relation to carbohydrate intake and that selective inhibitors of SGLT1 may be useful in prevention by reducing the burden of metabolic disease and its cardiovascular disease consequences.
There are several lines of evidence which suggest that the N51S/A441T/H615Q haplotype is likely to be functional. The N51S, A441T, and H615Q variants have been expressed in oocytes in a previous study and were found to reduce uptake of glucose by 70% to 80% compared with controls, suggesting that these variants, in isolation, have slightly reduced function (1). Adding further support to the in vitro studies, Combined Annotation Dependent Depletion scores (18), a measure of pathogenicity, of all 3 variants are high; these findings suggest that the altered amino acid sequences are likely to have functional consequence. Also similar to the in vivo findings, in our work, we observed the strongest association of this haplotype with reduced plasma glucose levels after oral glucose exposure in 3 independent populations, suggesting lower intestinal glucose uptake in haplotype carriers. Of interest, this haplotype includes 3 mutations in SGLT1, which may have originated from a single founder. Data from 1,000 Genomes shows that the highest frequency of the haplotype is carried in participants from England and Scotland (12%), followed by those of Indo-European ancestry (between 2% and 7%), South American ancestry (approximately 1% to 2%), and lowest in those with African heritage (0.008%). In a study of the evolutionary history of genes encoding brush-border proteins involved in carbohydrate digestion/absorption, the SGLT1 A441T and H615Q variants were both detected as positively selected sites in interspecies analysis, with A441T being a modern-human-specific site (19). These data support either a functional role for these haplotype variants or suggest that they are in strong linkage disequilibrium with a functional variant. In analyses of a separate population-based sample comprising individuals of European and specifically Finnish descent, we observed similarly significant findings for postprandial glucose levels and IGT.

To date, there have been no studies, to our knowledge, comprehensively evaluating genetic variation in SGLT1 with disease in a general population-based cohort of any ethnic/racial composition. Previous research has focused on identifying causal mutations in SGLT1 for GGM. This research has established that mutations which result in complete loss of function of SGLT1 lead to the extreme phenotype of GGM: severe neonatal diarrhea causing death from dehydration unless glucose and galactose are removed from the diet. Similar to the presentation in humans, Sglt1−/− mice develop GGM but are healthy when given a diet absent of glucose and galactose (20,21). When fed a Western-style diet, Sglt1−/− mice display improved OGTT (20,21). Support of the critical role of SGLT1 in maintaining body weight comes from RS1 knockout (RS1−/−) mice. In the absence of RS1, SGLT1 protein is overexpressed 7-fold, resulting in obesity, 30% increased body weight, and 80% increased total body fat compared with that of adult, wild-type mice (22). Although complete loss of function of the protein causes GGM, we would expect that mutations that cause even a subtle change in function in the SGLT1 protein that decreases absorption of intestinal glucose could affect an individual’s propensity to develop obesity and thereby reduce risk for DM. True caloric consumption (calories absorbed) would not necessarily reflect the number of calories that have been eaten, and very small changes to the percentage of total glucose eaten/total glucose absorbed could explain a protection or susceptibility to obesity. Adding further support that SGLT1 is important in determining postprandial glucose levels in human populations, a genome-wide association study of 1,5-anhydroglucitol levels (a biomarker of hyperglycemic excursions) identified significant linkage to SGLT1 in 2 independent European-American populations (23).

Inhibition of SGLT1 may also have secondary effects that could account for some of the observations of the present study. In studies of humans and model organisms, inhibiting SGLT1 has shown that decreased glucose uptake in the small intestine results in increased luminal glucose in the more distal gut and colon, stimulating a sustained increase in circulating glucagon-like peptide (GLP)-1 and peptide YY from L cells (24-26). This finding may be mediated by bacterial fermentation of luminal glucose to short-chain-fatty-acids that trigger GLP-1 secretion (25,27), leading to beneficial effects such as decreased appetite and improved cardiac function (28,29). Notably, GLP-1 and peptide YY are increased with SGLT1 but not SGLT2-specific inhibition (25). In addition, in our study, lowering of postprandial glucose levels via SGLT1 was associated with protection from HF. In murine models, cardiac knockdown of SGLT1 in PRKAG2 cardiomyopathy attenuated disease expression, suggesting a role of SGLT1 in disease pathogenesis (30). Consistent with this finding, transgenic overexpression of cardiac SGLT1 was associated with myocardial hypertrophy and left ventricular dysfunction that was reversible by SGLT1 suppression. In another study, human cardiac fibroblasts exposed to high glucose levels increased expression of proteins involved in cardiac fibrosis (i.e., matrix metalloproteinase-2) that was reversed by phlorizin (a dual SGLT1 and SGLT2 inhibitor) but not by dapagliflozin (an SGLT2 inhibitor) (31). In addition to our data, these studies
provide support for the role of SGLT1 as a cardiac sodium/glucose transporter that may be involved in cardiomyopathy.

Marketed SGLT2 inhibitors for DM are a relatively new class of drugs approved by the U.S. Food and Drug Administration that vary in their selectivity for the primarily renally expressed SGLT2, relative to the primarily intestinally expressed SGLT1 (32). In addition to pharmacologically inhibiting renal glucose reabsorption, SGLT2 inhibitors have shown the benefits of decreased body weight and blood pressure, and reduced mortality, cardiovascular events, and hospitalization for incident HF, in 2 recently published trials (EMP-A-REG [Empagliflozin Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients] and CANVAS [Canagliflozin Cardiovascular Assessment Study]) (33,34). Unwanted side effects of SGLT2 inhibition include a higher likelihood of developing genital and urinary tract infections (due to urinary excretion of glucose), osmotic diuresis, and euglycemic ketoacidosis, all of which would be expected to be diminished with increased SGLT1 > SGLT2 inhibition. On a population level, the data presented here indicate that more selective SGLT1 inhibition may be of benefit in attenuating the development of metabolic diseases, suggesting overall cardioprotection. Some of these effects may be mediated through SGLT1-specific mechanisms such as alteration of the microbiome and increased incretin secretion, as described earlier. Although data on SGLT1-specific inhibitors are lacking, initial data from Phase II and III clinical trials of a dual SGLT1/SGLT2 inhibitor is consistent with our findings, with an effect size that was similar to the estimated effect size in this Mendelian randomization analysis (35,36). Our findings further suggest that there may be long-term, protective effects of lowering intestinal glucose absorption via SGLT1 inhibition on cardiometabolic outcomes.

**STUDY LIMITATIONS.** We did not extensively assess the relationship of SGLT1 mutants in other population studies; however, the ARIC is composed of participants from 4 distinct communities, and we have replicated our findings in 2 diverse racial groups. The association of the N51S/A411T/H615Q haplotype with reduced DM, mortality, and HF should be considered hypothesis generating and must be tested in additional prospective large-sized cohorts with similarly longer term outcomes. Given the putative role of SGLT1 in mediating diet-induced metabolic risk, mutation-related effects on outcomes may vary based on chronic dietary variation within or between populations; thus, possible nutrient intake interactions warrant further investigation in cohorts with detailed dietary data in addition to prospective outcomes data.

**CONCLUSIONS**

The SGLT1 N51S/A411T/H615Q haplotype was associated with protection from postprandial hyperglycemia and obesity in European- and African-American subjects from 4 U.S. communities and 2 European cohorts from Finland. These data provide support that reduction of postprandial glucose levels via SGLT1 inhibition may be associated with reduced incidence of DM, HF, and/or mortality in certain populations at risk. These data argue for further research into the role of natural variation in SGLT1 in metabolic disease and suggest that more selective ways to inhibit SGLT1 may lead to therapies that reduce the deleterious effects of Western-style diets.

**ACKNOWLEDGMENTS** The authors thank the staff and participants of the ARIC study for their important contributions, and CSC–IT Center for Science, Finland, for providing computational resources for the FINRISK study.

**ADDRESS FOR CORRESPONDENCE:** Dr. Scott D. Solomon, Cardiovascular Division, Brigham and Women’s Hospital, 75 Francis Street, Boston, Massachusetts 02115. E-mail: ssolomon@rics.bwh.harvard.edu. Twitter: @BrighamWomens.

---

**PERSPECTIVES**

**COMPETENCY IN MEDICAL KNOWLEDGE:**
SGLT1 absorbs glucose from the intestine, while both SGLT2 and SGLT1 are involved in renal reabsorption. Genetic mutations associated with loss of SGLT1 function are associated with lower post-prandial glucose levels and avoidance of obesity.

**COMPETENCY IN PATIENT CARE AND PROCEDURAL SKILLS:** SGLT2 inhibitors increase urinary glucose excretion, reduce body weight, and lower mortality and hospitalization for heart failure. Potential adverse effects include urogenital infections and euglycemic ketoacidosis.

**TRANSLATIONAL OUTLOOK:** Additional studies are needed to determine whether drugs that target SGLT1 could reduce the incidence of diabetes and heart failure and improve longevity in the general population.
REFERENCES


KEY WORDS: glucose tolerance, Mendelian randomization, SGLT1

APPENDIX For supplemental tables and a figure, please see the online version of this paper.