

SCIENTIFIC REPORTS

OPEN

Genome-wide association study of 1,5-anhydroglucitol identifies novel genetic loci linked to glucose metabolism

Man Li^{1,2}, Nisa M. Maruthur^{1,3,4}, Stephanie J. Loomis¹, Maik Pietzner^{5,6}, Kari E. North⁷, Hao Mei⁸, Alanna C. Morrison⁹, Nele Friedrich^{5,6}, James S. Pankow¹⁰, Matthias Nauck^{5,6}, Eric Boerwinkle^{9,11}, Alexander Teumer^{6,12}, Elizabeth Selvin^{1,3,4} & Anna Köttgen^{1,13}

1,5-anhydroglucitol (1,5-AG) is a biomarker of hyperglycemic excursions associated with diabetic complications. Because of its structural similarity to glucose, genetic studies of 1,5-AG can deliver complementary insights into glucose metabolism. We conducted genome-wide association studies of serum 1,5-AG concentrations in 7,550 European ancestry (EA) and 2,030 African American participants (AA) free of diagnosed diabetes from the ARIC Study. Seven loci in/near *EFNA1/SLC50A1*, *MCM6/LCT*, *SI*, *MGAM*, *MGAM2*, *SLC5A10*, and *SLC5A1* showed genome-wide significant associations ($P < 5 \times 10^{-8}$) among EA participants, five of which were novel. Six of the seven loci were successfully replicated in 8,790 independent EA individuals, and *MCM6/LCT* and *SLC5A10* were also associated among AA. Most of 1,5-AG-associated index SNPs were not associated with the clinical glycemic markers fasting glucose or the HbA1c, and vice versa. Only the index variant in *SLC5A1* showed a significant association with fasting glucose in the expected opposing direction. Products of genes in all 1,5-AG-associated loci have known roles in carbohydrate digestion and enteral or renal glucose transport, suggesting that genetic variants associated with 1,5-AG influence its concentration via effects on glucose metabolism and handling.

1,5-anhydroglucitol (1,5-AG) is a non-traditional biomarker of hyperglycemia that is of growing clinical interest¹. It is a naturally occurring monosaccharide found in nearly all foods and absorbed in the gut. Under normoglycemic conditions, its concentrations in blood are maintained constant through renal filtration followed by reabsorption in the proximal tubules. Glucose and 1,5-AG share some transport proteins for which they represent competing substrates. When blood glucose concentrations exceed the renal glucose threshold of approximately 180 mg/dL, glucose is excreted in the urine and inhibits tubular re-absorption of 1,5-AG, resulting in lower blood 1,5-AG concentrations². Consequently, glucose peaks can lead to decreased 1,5-AG serum concentrations³, and 1,5-AG has been established as a marker of hyperglycemic excursions and postprandial glucose peaks^{3,4}. Recent

¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

²Division of Nephrology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA.

³Department of Medicine, Division of General Internal Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

⁴The Welch Center for Prevention, Epidemiology, and Clinical Research, The Johns Hopkins University, Baltimore, Maryland, USA.

⁵Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany.

⁶DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany.

⁷Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA.

⁸Department of Data Science, School of Population Health, University of Mississippi Medical Center, Jackson, Mississippi, USA.

⁹Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA.

¹⁰Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA.

¹¹Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA.

¹²Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany.

¹³Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Germany. Correspondence and requests for materials should be addressed to A.K. (email: anna.koettgen@uniklinik-freiburg.de)

	EA	AA
N	7550	2030
Age, years ^a	57.0 (5.7)	55.9 (5.7)
Female	4099	1290
Site		
Forsyth County, NC	2264	215
Jackson, MS	—	1815
Minneapolis Suburbs, MN	2872	—
Washington County, MD	2414	—
1,5-anhydroglucitol, ug/mL ^b	18.9 (15.3, 22.6)	17.4 (13.8, 20.9)
Fasting glucose, mg/dL ^b	101 (95, 109)	104 (97, 114)
Fasting glucose \geq 126 mg/dL, % (N)	5.5% (417)	11.3% (229)
HbA1c, % ^{a*}	5.4 (5.2, 5.6)	5.7 (5.4, 6.0)
HbA1c \geq 6.5%, % (N)	2.5% (187)	10.2 (204)

Table 1. Characteristics of ARIC study participants included in present genetic study. ^aMean (standard deviation). ^bMedian (interquartile range: 25th percentile, 75th percentile). ^{*}82 EA and 39 AA participants were missing hemoglobin A1c (%) values.

studies have demonstrated robust associations of low serum 1,5-AG concentrations with long-term microvascular and macrovascular complications in persons with diabetes^{5,6}, and with major cardiovascular events in persons without diabetes⁷. These observations are supported by complementary evidence linking daily glucose fluctuations to cardiovascular complications⁸. Gaining insights into the genetic underpinnings of a glycemic marker with unique properties, such as 1,5-AG, may improve our understanding not only of the biology of the marker itself, but also of diabetes, hyperglycemia and glucose metabolism.

Genome-wide association studies (GWAS) are a useful tool to identify genetic variants associated with 1,5-AG concentrations free of prior biological hypotheses. Previous GWAS have evaluated 1,5-AG as one of hundreds of metabolites quantified from a non-targeted metabolomics platform and reported two significantly associated loci near *RAB3GAP1* and *MGAM*^{9,10}. The objective of the current study was to carry out the first GWAS of absolute blood 1,5-AG concentrations quantified with a targeted assay in order to better understand mechanisms underlying glucose metabolism as well as to identify determinants of 1,5-AG as a biomarker of hyperglycemic excursions.

Results

Study population characteristics. In this study population of individuals without diagnosed diabetes, the median (25th percentile, 75th percentile) serum concentrations of 1,5-AG were 18.9 (15.3, 22.6) ug/mL for the 7,550 European ancestry participants (EA) and 17.4 (13.8, 20.9) ug/mL for the 2,030 African American participants (AA) (Table 1). In agreement with higher 1,5-AG concentrations among EA compared to AA participants, median (25th percentile, 75th percentile) of fasting glucose and hemoglobin A1c (HbA1c) concentrations were lower (101 (95, 109) mg/dl and 5.4 (5.2, 5.6) %, respectively, in EA vs. 104 (97, 114) mg/dl and 5.7 (5.4, 6.0)%, respectively, in AA).

Genome-wide association studies. The GWAS of 1,5-AG were based on ~8.5 million autosomal single nucleotide polymorphisms (SNPs) in the EA and ~14.7 million autosomal SNPs in the AA participants. There was no indication for systematic inflation of the resulting association *P*-values (genomic control factor 1.02 for EA and 0.99 for AA). The corresponding quantile-quantile plots for the obtained *P*-values from each GWAS are shown in Supplementary Figure S1. As illustrated by the Manhattan plot in Fig. 1, six loci on chromosomes 1q22, 2q21, 3q26, 7q34, 17p11, and 22q12 contained SNPs associated with 1,5-AG concentrations at genome-wide significance ($P < 5 \times 10^{-8}$) in EA participants. Information about genes mapping into the associated loci can be found in Supplementary Table S1, with roles in carbohydrate digestion and enteral and/or renal monosaccharide uptake emerging as common themes (Fig. 2). No locus reached genome-wide significance in the smaller sample of AA participants (Manhattan plot, Supplementary Figure S2).

For four of the six genome-wide significant loci in EA participants, 2q21, 3q26, 17p11, and 22q12, the minor allele of the index SNPs was associated with lower 1,5-AG concentrations. The association with the lowest *p*-value was observed for the intronic SNP rs182549 located in *MCM6* on chromosome 2, ($\beta = -1.19$, MAF = 0.33, $P = 6.5 \times 10^{-32}$, Table 2). The *MCM6* gene is located upstream of *LCT*, the gene encoding lactase, with associated SNPs spanning the entire region (Supplementary Figure S3). The other five identified genome-wide significant index SNPs were rs11976181 (an intronic SNP located in *MGAM*), rs117355297 (a synonymous variant in *SLC5A10*), rs9330264 (an intergenic variant close to *EFNA1* and *SLC50A1*), rs9825346 (an intergenic variant close to *SI*), and rs117086479 (a variant located in an upstream regulatory region of *SLC5A1*). Regional association plots for each locus are shown in Supplementary Figure S3. Together, the six index SNPs explained 5.13% of the variance in 1,5-AG concentrations among EA participants (Table 2).

The EA index SNP at the *MCM6/LCT* locus was significantly associated with 1,5-AG among the ARIC AA participants (Table 3, $P < 1.19 \times 10^{-5}$, see Methods). In addition, the direction of effect was consistent for five of the six EA index SNPs and the magnitude of effect was larger than in EA and nominally significant ($p = 0.01$) for the index SNP at *SLC5A10* in this smaller AA study sample. Another SNP at the *SLC5A10* locus, rs7214031,

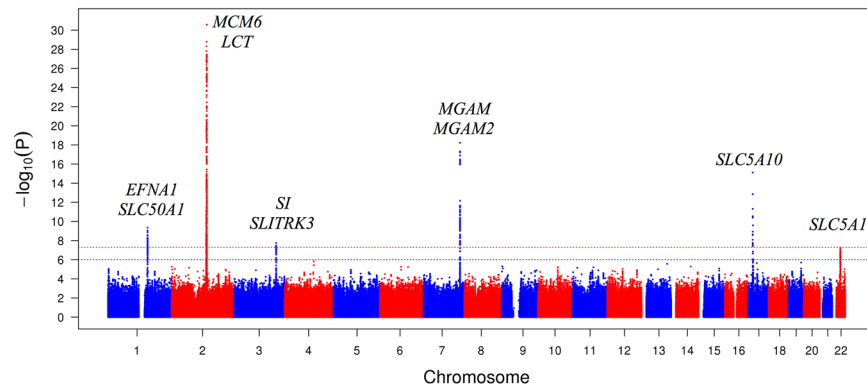


Figure 1. Manhattan plot of the results of the GWAS of 1,5-anhydroglucitol in European American participants from the ARIC study. The y-axis represents the ($-\log_{10}$) association P-values and the x-axis the genomic position (GRCh build37). The red dotted line indicates the threshold for genome-wide significance ($P < 5 \times 10^{-8}$) and the blue dotted line the suggestive significance threshold ($P < 1 \times 10^{-6}$). For intergenic index variants, the flanking genes are provided as annotation.

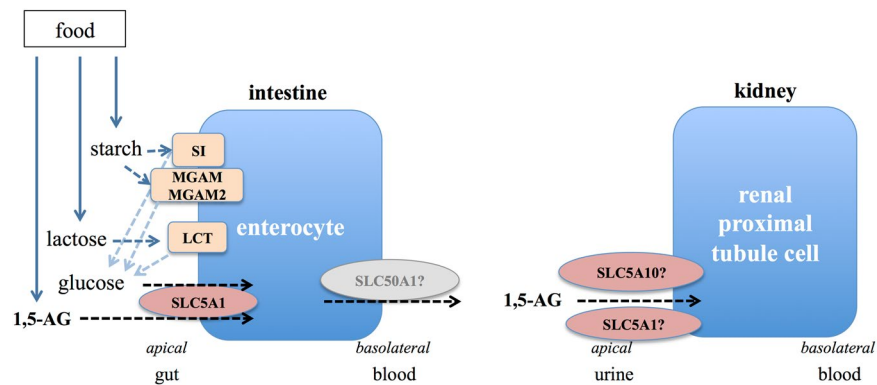


Figure 2. Glucose metabolism as a common and biologically plausible theme among the genes mapping into the identified loci. This figure shows their role in intestinal carbohydrate digestion as well as glucose and 1,5-AG reabsorption in gut and kidney.

reached regional significance among the AA participants ($p = 9.54 \times 10^{-6}$, $D' = 1$ with the EA index SNP in the 1000 Genomes phase 3 AFR data).

To determine whether there were additional independent signals at each genome-wide significant locus, we performed conditional association analyses. An independent signal was detected at the locus on chromosome 7q34 (regional association plot, Supplementary Figure S3). After conditioning on the lead SNP, rs11976181, the index SNP among the remaining genome-wide significant SNPs was rs13229622, an intronic SNP in neighboring *MGAM2* (β -conditional = -0.73, MAF = 0.22, $P = 8.6 \times 10^{-11}$, P -conditional = 1.1×10^{-8} , 0.54% variance explained, Table 2). *MGAM2* is a paralog of the *MGAM* gene, encodes maltase-glucoamylase 2 and was previously annotated as *LOC93432* in the regional association plot. Our data thus support two independent and novel association signals at chromosome 7q34.

Replication. The six index SNPs identified in our discovery screen were assessed for replication in summary statistics from a publicly available resource from GWAS meta-analyses of >400 metabolites in human blood, including 1,5-AG⁹. The seventh index SNP, the independent variant in *MGAM2*, could not be evaluated in a consistent manner in this resource because we identified it through conditional analyses. Index SNPs or good proxies were available for four of the six evaluated variants, all of which successfully replicated (Table 3). For two loci, *SLC5A10* and *SLC50A1*, a good proxy was not available and the 500 kb flanking region therefore evaluated. For the *EFNA1/SLC50A1* locus, no suggestive significance signal was observed. Conversely, for the *SLC5A10* locus, the 500 kb flanking region contained a suggestive independent association signal in the replication data (index SNP rs2305062 in the neighboring *PRPSAP2*, $P = 1.41 \times 10^{-5}$).

To further confirm the discovery association signals including the conditional finding, the seven index variants were assessed in the SHIP-Trend study ($N = 966$, Supplementary Note). All six loci except the *EFNA1/SLC50A1* locus showed consistent effect directions and association p-values of ≤ 0.05 , despite the much smaller sample (Table 3). Together, the six index variants explained 4.63% of the variance in 1,5-AG concentrations in the independent SHIP-Trend study. Of note, the conditional association signal at *MGAM2* was also replicated.

Index SNP		Discovery GWAS (1,5-AG)					Reported Association with Fasting Glucose*			
Locus [§]	Variant ID	A1/A2	AF	Effect (SE), ug/ mL	P	Expl Var (%)	Proxy SNP	R ² with index SNP [§] (D [†])	Effect (SE), mmol/l	P
<i>EFNA1/SLC50A1</i>	rs9330264	T/C	0.36	0.64 (0.10)	2.96×10^{-10}	0.51	rs11264319	0.53 (0.96)	-0.0065 (0.0036)	0.07
<i>MCM6/LCT</i>	rs182549	C/T	0.33	-1.19 (0.10)	6.54×10^{-32}	1.91	rs4988235	1 (1)	-0.0015 (0.004)	0.7
<i>SI</i>	rs9825346	G/A	0.41	-0.54 (0.10)	1.28×10^{-8}	0.42	rs9825346	1	0.0038 (0.0038)	0.32
<i>MGAM</i>	rs11976181	T/C	0.17	1.11 (0.12)	2.62×10^{-19}	1.03	rs3800993	1 (1)	-0.0078 (0.005)	0.12
<i>MGAM2</i>	rs13229622 [#]	C/G	0.22	-0.73 (0.11)	8.64×10^{-11}	0.54	Not investigated due to lack of conditional estimates			
<i>SLC5A10</i>	rs117355297	T/C	0.04	-2.13 (0.26)	3.83×10^{-16}	0.39	rs1621499	0.28 (0.97)	0.0043 (0.0055)	0.43
<i>SLC5A1</i>	rs117086479	G/A	0.06	-1.09 (0.20)	4.05×10^{-8}	0.87	rs4821013	1 (1)	0.019 (0.0073)	0.0072

Table 2. The seven genome-wide significant index SNPs ($p < 5 \times 10^{-8}$) associated with 1,5-anhydroglucitol in 7,550 European American subjects from the ARIC study. *The association result with fasting glucose was extracted from the published MAGIC Consortium GWAS result¹¹. [§]The gene closest to the variant and other candidate genes are listed (index gene). [†]R² information was calculated from the 1000 Genomes Project Phase 3 in European (EUR) population. [#]rs13229622 is an independent SNP with genome-wide significant association with 1,5-AG found through conditional analysis. A1 is the effect allele, i.e. the allele for which effect estimates are provided. Abbreviations: AF = effect allele frequency; SE = standard error; Expl Var = proportion of 1,5-AG variance explained.

Index SNP Information			Replication, Metabolite GWAS				Replication, SHIP		Meta-analysis		Look-up, ARIC AA		
Locus [§]	Variant ID	A1/A2	Repl SNP	R ² with index SNP [§] (D [†])	Effect [‡] (SE)	P	Effect [‡] (SE)	P	Z-score	P	AF	Effect (SE)	P
<i>EFNA1/SLC50A1</i>	rs9330264	T/C	No good proxy, 500 kb region does not contain suggestive association signals				0.003 (0.01)	0.8	6.02	1.78×10^{-9}	0.09	0.27 (0.35)	0.43
<i>MCM6/LCT</i>	rs182549	C/T	rs4988235	1 (1)	-0.04 (0.003)	4.23×10^{-28}	-0.07 (0.01)	8.93×10^{-8}	-16.91	3.72×10^{-64}	0.88	-1.32 (0.27)	8.51×10^{-7}
<i>SI</i>	rs9825346	G/A	rs9825346	1 (1)	-0.02 (0.003)	8.18×10^{-9}	-0.03 (0.01)	0.02	-8.41	4.07×10^{-17}	0.47	-0.32 (0.18)	0.08
<i>MGAM</i>	rs11976181	T/C	rs11976181	1 (1)	0.03 (0.01)	6.00×10^{-4}	0.05 (0.02)	5.95×10^{-3}	9.96	2.36×10^{-23}	0.26	-0.24 (0.20)	0.22
<i>MGAM2</i>	rs13229622 [#]	C/G	Not investigated due to lack of conditional estimates				-0.04 (0.01)	0.01	-6.95	3.73×10^{-12}	NA	NA	NA
<i>SLC5A10</i>	rs117355297	T/C	No good proxy, 500 kb region contains association signal (rs2305062, $P = 1.41 \times 10^{-5}$)				-0.08 (0.03)	0.01	-8.52	1.62×10^{-17}	0.01	-2.63 (1.05)	0.01
<i>SLC5A1</i>	rs117086479	G/A	rs10483162	1 (1)	-0.02 (0.005)	1.52×10^{-5}	-0.05 (0.03)	0.05	-7.21	5.67×10^{-13}	0.01	-1.49 (0.85)	0.08

Table 3. Independent replication results for the seven identified index variants in EA populations and evaluation among African Americans. *The gene closest to the variant and other candidate genes are listed (index gene). [§]R² information was extracted from 1000 Genomes Project Phase 3 in European (EUR) population. [‡]Effect sizes in the replication studies are based on semi-quantitative metabolomics measurements of 1,5-AG⁹ and are thus on a different scale compared to the discovery estimates. [#]rs13229622 is an independent SNP with genome-wide significant association with 1,5-AG found through conditional analysis. It was not investigated in the metabolite GWAS because available association results were not conditional estimates. A1 is the effect allele, i.e. the allele for which effect estimates are provided. Abbreviations: Repl SNP = replication SNP; AF = effect allele frequency; SE = standard error.

In the combined analysis of the discovery and the two replication studies, all seven index variants remained genome-wide significant (Table 3), but the signal at *EFNA1/SLC50A1* was purely driven by discovery and should thus not be considered replicated.

Stratified analysis by fasting glucose status. In persons with fasting glucose below the threshold to diagnose diabetes (< 126 mg/dl, $N = 7133$), all seven index SNPs were associated at genome-wide significance with similar p-values compared to the overall GWAS result. In persons with elevated fasting glucose (≥ 126 mg/dl, $N = 417$), i.e., undiagnosed diabetes, the effect of some of the SNPs was smaller compared to persons with non-diabetic fasting glucose concentrations. However, the differences of SNP effects between these two groups were not statistically significant, consistent with wide confidence intervals and non-significant associations in the smaller group with undiagnosed diabetes (Supplementary Table S2).

Associations with traditional glycemic markers. To test whether the 1,5-AG-associated loci were also related to traditional glycemic markers, we investigated the six index SNPs or their good proxies in published large GWAS meta-analysis results for fasting glucose¹¹ ($N = 46,186$) and HbA1c¹² ($N = 46,368$) in the MAGIC Consortium to maximize statistical power. Only the SNP rs117086479 at *SLC5A1* showed a significant association with fasting glucose (Table 2, $P = 0.0072$), showing an opposite effect direction as may be expected in case of competing substrates. None of the SNPs was significantly associated with HbA1c (Supplementary Table S3).

We also investigated known HbA1c and fasting glucose-associated SNPs in the 1,5-AG GWAS result from EA participants (Supplementary Table S4). None of the known SNPs for these traditional glycemic markers was significantly associated with 1,5-AG concentrations after correction for multiple testing.

Discussion

In this GWAS of blood 1,5-AG concentrations, we identified associations at seven independent SNPs mapping into six genomic loci among EA individuals free of diagnosed diabetes. All but one association was successfully replicated. Most of the index SNPs showed strong associations only with 1,5-AG concentrations and not with fasting glucose or HbA1c. Our findings are consistent with known roles of genes mapping into the associated loci in intestinal carbohydrate digestion and enteral and renal glucose handling, supporting shared aspects of 1,5-AG and glucose metabolism.

One previous genome-wide association study of blood metabolite levels has evaluated 1,5-AG as one of over 400 metabolites in EA populations of similar size as our study⁹. The prior study by Shin and colleagues reported significant associations at the defined threshold ($P < 1.03 \times 10^{-10}$) for SNPs in two loci identified in our report, *MGAM* and *MCM6/LCT* (named *RAB3GAP1* in that study). A third SNP, near *SI*, showed $P < 5 \times 10^{-8}$ but was not significant after correction for multiple testing. The high estimated heritability of 1,5-AG concentrations of 61% reported by Shin *et al.* is consistent with important genetic influences on the trait. Although the previous study and ours were of similar sample size, our study identified more than twice as many loci associated at genome-wide significance. Several factors could explain this difference. For one, different sample exclusion criteria were used: while the study by Shin *et al.* did not exclude persons with diagnosed or treated diabetes, these persons were excluded in our study. Secondly, our study used a targeted assay to obtain absolute quantification of blood 1,5-AG concentrations whereas the 1,5-AG concentrations from Shin *et al.* represent semi-quantitative measurements from non-targeted metabolomics experiments.

GWAS can identify genomic loci, but the causal gene(s) and variant(s) often remain unclear and need to be established in follow-up studies. While our scan, therefore, does not allow for drawing definite conclusions regarding causal genes, it is noteworthy that a common and biologically plausible theme among the genes mapping into the identified loci is their role in intestinal carbohydrate digestion as well as glucose and 1,5-AG absorption in gut and kidney, as detailed below.

The index SNP on chromosome 2q21 maps into *MCM6*, upstream of the lactase gene *LCT*. The index variant is in perfect linkage disequilibrium (LD) with rs4988235, a variant associated with lactose intolerance in ClinVar and OMIM (#223100), which functions as an enhancer of the *LCT* gene promoter in intestinal cell culture¹³. Lactase is a glucosidase enzyme located in the brush border of human small intestine, where it is involved in the hydrolysis of lactose into glucose and galactose. The index SNP on chromosome 3q25.2-q26.2 maps upstream of *SI*, which encodes the glucosidase enzyme sucrase-isomaltase, preferentially expressed in the apical brush border membrane of enterocytes. Rare mutations in *SI* are a cause of autosomal-recessive congenital sucrase-isomaltase deficiency (OMIM #222900). The locus on chromosome 7q34 contains two independent index variants, mapping into the genes *MGAM* and *MGAM2*. *MGAM* encodes maltase-glucoamylase, another intestinal brush border membrane enzyme involved in carbohydrate digestion that is the target of alpha-glucosidase inhibitors such as acarbose^{14,15}. The protein is 60% homologous to sucrase-isomaltase, with the two enzymes having complementary roles in starch digestion. The index SNP in *MGAM* is an eQTL for the neighboring *MGAM2*. *MGAM2* is predicted to encode maltase-glucoamylase 2, but has not been studied functionally so far.

Neither maltase-glucoamylase, nor lactase and sucrase-isomaltase are known to directly interact with the monosaccharide 1,5-AG. Several mechanisms by which genetic variation in these genes may relate to 1,5-AG concentrations are therefore conceivable: first, 1,5-AG in food may also occur within macromolecules metabolized by these enzymes. Second, genetic variants in these enzymes may lead to differential availability of glucose from ingested starch that may then compete with 1,5-AG for enteral uptake by shared transport proteins. This mechanism receives some support by the shared association of the *SLC5A1* index variant with higher 1,5-AG and lower fasting glucose concentrations (see below). Third, genetic variants in these genes could lead to higher post-prandial glucose peaks that exceed the renal glucose threshold, leading to competition for reuptake with 1,5-AG by tubular *SLC5A10*. Previous 24-hour blood glucose studies in healthy individuals, however, suggest that post-prandial excursions of blood glucose >180 mg/dl, the renal threshold, do not occur commonly^{16,17}.

The index SNP on chromosome 17p11.2 is a low frequency exonic synonymous variant in *SLC5A10*. According to the Human Protein Atlas¹⁸, *SLC5A10* transcript is exclusively found in human kidney cortex. The protein is a Na⁺-dependent transporter of mannose, fructose, galactose and glucose, responsible for their reabsorption from urine in the brush border of renal proximal tubule cells^{19,20}. Because of its exclusive expression in kidney, genetic variation in this gene is likely related to 1,5-AG concentrations either because it also transports 1,5-AG or because it influences the amount of urinary glucose that competes with 1,5-AG for renal reuptake through *SLC5A9*. The latter protein is thought to be the main renal re-uptake mechanism for 1,5-AG. Of note, we did not observe any association between variants in *SLC5A9* and 1,5-AG concentrations in our study, suggesting that *SLC5A9* may not be the main transporter for renal 1,5-AG reuptake or that variants impacting *SLC5A9* function were not present or detectable in our population. *SLC5A9* shows high similarity to *SLC5A10*²¹, suggesting that *SLC5A10* may represent a novel 1,5-AG transport protein. This hypothesis is supported by a whole-genome sequencing study published while our manuscript was in revision, which reported two rare loss of function mutations in *SLC5A10* that were associated with blood 1,5-AG concentrations²².

The index SNP in *SLC5A1* on chromosome 22q12.3 is in high LD with a missense variant, p.Asn51Ser. *SLC5A1* encodes the sodium/glucose cotransporter 1, which is primarily expressed in the brush border membrane of enterocytes where it mediates glucose absorption. Rare mutations in this gene can cause glucose/galactose malabsorption (OMIM #606824). The protein also mediates glucose re-uptake in the S3 segment of the proximal tubule.

Finally, the index SNP on chromosome 1q22 maps upstream of both flanking genes *EFNA1* and *SLC50A1*. While there is no clear biological connection of the *EFNA1* gene product to 1,5-AG concentrations, *SLC50A1* encodes a glucose transporter. The protein is strongly expressed in absorptive enterocytes where it may mediate glucose efflux across the basolateral membrane^{21,23}. However, as this locus did not show evidence for external replication, this finding needs to be treated as preliminary and confirmed in further studies.

Together, our results suggest that 1,5-AG blood concentrations in individuals without diagnosed diabetes are closely linked to glucose metabolism and can deliver insights complementary to those from the study of fasting glucose concentrations. Genetic variability in carbohydrate digestion and glucose uptake may lead to measurable changes in 1,5-AG concentrations previously thought to only derive from glucose excursions >180 mg/dl that define the renal glucose threshold, supporting the use of 1,5-AG as a marker of postprandial hyperglycemia and glucose variability.

There are several strengths of this study. This is the first dedicated GWAS of 1,5-AG blood concentrations in a large sample of EA participants with detailed characterization of glycemic status based on both traditional and non-traditional glycemic measures. Findings were externally replicated and also studied among African American participants, although statistical power was lower because of smaller sample size. In contrast to one prior study that evaluated 1,5-AG based on semi-quantitative data from a metabolomics experiment, we used a targeted assay to obtain absolute quantitation of 1,5-AG and detected twice as many genomic loci. Some limitations of our study also warrant mention: because we used a GWAS array for genotyping, we were not able to comprehensively examine rare variants. Future studies in larger samples will therefore be needed to determine the effect of rare genetic variants on 1,5-AG concentrations.

In conclusion, we found four novel index variants at *SI*, *EFNA1/SLC50A1*, *MGAM2*, and *SLC5A1* associated with 1,5-AG concentrations in EA study participants and confirmed two known associations at *MGAM* and *MCM6/LCT* from GWAS. Associations for two rare presumably functional variants at *SLC5A10* were reported from a whole-genome sequencing association study while this manuscript was in revision. These loci highlight a putative role of carbohydrate digestion and intestinal and renal glucose and 1,5-AG handling for determining blood 1,5-AG concentrations in individuals without diabetes. Future experimental studies may investigate whether *SLC5A10* is responsible for renal 1,5-AG reuptake, and future genetic association studies in larger sample sizes may investigate determinants of 1,5-AG concentrations among individuals with diabetes as well as the potential effect of genetic interactions of 1,5-AG associated variants on blood glucose levels and diabetes risk.

Methods

Study Population. The ARIC Study is an ongoing prospective cohort originally designed to study risk factors for clinical and subclinical cardiovascular disease²⁴. Participants were middle-aged adults recruited from four U.S. communities (Jackson, Mississippi; Forsyth, North Carolina; Washington County, Maryland; and suburbs of Minneapolis, Minnesota). A total of 15,792 participants attended the first visit (1987–1989) with subsequent in-person visits in 1990–1992 (visit 2), 1993–1995 (visit 3), 1996–1998 (visit 4), and 2011–2013 (visit 5). A sixth visit is ongoing. All study participants provided written informed consent, and the study protocols were approved by the relevant Institutional Review Boards. All methods were performed in accordance with the relevant guidelines and regulations for human subject research, in accordance with the Declaration of Helsinki.

There were 14,348 participants who attended visit 2, the visit at which 1,5-AG was measured. In the present study, all persons were excluded who did not consent to genetic or non-cardiovascular-disease research (N = 119), whose race/ethnicity was recorded as other than white or black (N = 41), who were African Americans recruited at the Minnesota and Washington County sites (N = 45), who were fasting less than 8 hours (N = 506), who were missing 1,5-AG measurements (N = 803), or who were missing genotype data and/or did not pass the genotype data quality control (N = 2,495). In addition, individuals were excluded who had a history of diagnosed diabetes (N = 759), defined as self-reported physician diagnosis of diabetes or self-reported use of glucose-lowering medication at visit 2, to avoid the secondary influence of glucose lowering medications or diagnosis-related lifestyle changes on 1,5-AG concentrations. Individuals with high blood glucose concentrations in whom diabetes was not known or treated were included in the analysis in order to avoid a selective exclusion of untreated individuals with low 1,5-AG levels. The final analytic sample consisted of 7,550 EA and 2,030 AA participants. Supplementary Figure S4 shows the distribution of 1,5-AG concentrations in EA and AA participants.

Genotyping. Genotyping in the ARIC Study was performed using the Affymetrix genome-wide Human SNP Array 6.0. Extensive quality control was performed and has been reported in detail previously²⁵. Genotype imputation in EA and AA participants was conducted separately using genotyped SNPs with minor allele frequency (MAF) > 0.01, call rate > 0.95, and Hardy-Weinberg equilibrium $P > 0.00001$ using the 1000 Genomes March 2012 ALL reference haplotype panel. Pre-phasing was performed using SHAPEIT²⁶, and subsequent imputation was carried out using IMPUTE2²⁷. The imputed genotype dosages were filtered to retain only SNPs with MAF ≥ 0.01 and of imputation quality ≥ 0.3 . To account for potential population stratification, ten principal components were estimated separately for EA and AA participants using a subset of the GWAS SNPs with the software EIGENSTRAT²⁸. The first 10 principal components were included in the respective linear regression models to control for population stratification.

Measurement of glycemic markers. In 2012–2013, we used the Roche Modular P800 system to measure 1,5-AG concentrations (GlycoMark, Winston-Salem, NC) from stored serum specimens obtained from participants at ARIC visit 2. The inter-assay coefficient of variation was 5%, and the reliability coefficient for 610 masked duplicate specimen pairs was 0.99²⁹.

Fasting glucose and HbA1c were also measured using blood samples collected at visit 2. Glucose was measured using a hexokinase method. HbA1c was measured using high-performance liquid chromatography using

the Tosoh A1c 2.2 Plus Glycohemoglobin and Tosoh G7 Analyzers (Tosoh Bioscience, South San Francisco, CA), and was standardized to the Diabetes Control and Complications Trial assay³⁰.

Statistical Methods for GWAS and candidate region interrogation. GWAS of serum 1,5-AG concentrations were conducted in EA and AA participants separately using the software SNPTEST version 2.4.1³¹. Linear regression models were calculated with 1,5-AG as the dependent variable adjusting for age, ARIC study center, sex, and first 10 principal components and assuming an additive genetic model³². We set $P < 5 \times 10^{-8}$ as the genome-wide significance threshold³³, and suggestive significance was defined by a $P < 1 \times 10^{-6}$. LocusZoom³⁴ was used to generate regional association plots for the SNP with the lowest association P at a given locus, the index SNP. To test whether there were any additional associated genetic variants independent of the index SNP in a given gene region, the regression analyses were repeated conditioning on allele dosage at the index SNP in each candidate region. To assess whether the association signals discovered among the EA participants were generalizable to other groups, we examined the association of the EA index SNPs and SNPs in the 500 kb flanking region with 1,5-AG concentrations in AA ARIC participants. For these targeted evaluations, the statistical significance threshold in each region was obtained using a region-specific Bonferroni correction based on the total number of independent SNPs in the region, calculated recursively from a sliding window with size 50 SNPs and pairwise r^2 value of 0.2 using PLINK³⁵ based on genotype information from 1000 genomes phase 1 African (AFR) participants. Supplementary Figure S4 shows a flow chart detailing the study design and significance thresholds.

Replication. To replicate our findings in independent study populations, we first investigated the association between 1,5-AG concentrations and the index SNP or proxy SNPs in high linkage disequilibrium (LD, $r^2 > 0.8$) within a 500 kb flanking region among publicly available results from a large GWAS meta-analysis project investigating more than 400 metabolites in human blood, including 1,5-AG, from 7,824 adult participants of two previously described European population studies, the TwinsUK³⁶ and the KORA³⁷ cohorts, (“Metabolomics GWAS server”, (<http://metabolomics.helmholtz-muenchen.de/>)⁹). The TwinsUK cohort is a British adult-twin registry in the age range 17–85 years, 93% of which are female⁹. The KORA study consists of population-based epidemiological surveys of participants living in southern Germany in the age range 32–77 years, 50% of which are female⁹. Metabolic profiling was done on fasting serum from participants in the TwinsUK and the KORA cohorts, using ultrahigh-performance liquid-phase and gas-chromatography separation coupled with tandem mass spectrometry (Metabolon). Metabolite concentrations were analyzed using log transformation as described previously¹⁰.

Due to lack of good proxies for two index variants and because of the conditional analyses that were part of our discovery analysis, findings were further tested for replication among 966 adult EA participants of the Study of Health in Pomerania (SHIP)-Trend study³⁸. The SHIP-Trend study is a population-based cohort study with participants from the North-East region of Germany aged 20–79 years, 56% of which are female. The 1,5-AG concentrations were measured similarly to the studies included in Shin *et al.* (Metabolon)³⁹, and analyzed as runday-scaled intensities from raw ion counts. The criteria to define successful replication were set as effect direction consistency and $P < 0.017$, corresponding to one-sided hypothesis testing after applying a Bonferroni correction for the evaluation of six SNPs. When neither the index SNP nor a good proxy were available, the regional association within a 500 kb flanking region in the replication dataset was examined to assess whether there were any significant signals in the region. Due to different scaling of 1,5-AG in these two replication studies, we used the weighted Z-score method for the joint meta-analysis of these two replication studies and the ARIC EA cohort⁴⁰.

Additional analyses. For each index SNP, we calculated the proportion of variance explained by the SNP genotype by using the difference of deviance between the full model and the reduced regression model, which excludes the SNP genotype, divided by the total deviance of 1,5-AG.

For all 7 index SNPs, we performed stratified analyses by separately studying participants with fasting glucose < 126 mg/dl and fasting glucose ≥ 126 mg/dl. The difference between the effects of the same SNP on 1,5-AG concentrations across the two groups was assessed by a two-sample t-test. Statistical significance was set at $P < 0.05$.

To evaluate whether the 1,5-AG-associated index SNPs had similar effects on traditional glycemic markers, we searched for the index SNPs' effects on HbA1c¹² and fasting glucose¹¹ in publicly available summary statistics from large GWAS meta-analyses to maximize power to detect such associations. If the index SNP did not exist in the publicly available dataset, we used the same proxy SNP as used for replication. The criteria to define significance were set as $P < 8.33 \times 10^{-3}$ (0.05/6).

SNPs known to be associated with HbA1c and fasting glucose were identified by searching “A1c” or “fasting glucose” in the GWAS catalogue, followed by exclusion of SNPs from non-European ancestry studies and SNPs with $P > 5 \times 10^{-8}$. We further excluded two SNPs for high LD with another SNP in the region, leaving 26 fasting glucose and 13 HbA1c-associated index SNPs. Statistical significance to detect association with 1,5-AG was set as $P < 0.002$ (0.05/26) for fasting glucose and $P < 0.004$ (0.05/13) for HbA1c.

References

- Parrinello, C. M. & Selvin, E. Beyond HbA1c and glucose: the role of nontraditional glycemic markers in diabetes diagnosis, prognosis, and management. *Current diabetes reports* **14**, 548, doi:10.1007/s11892-014-0548-3 (2014).
- Dworacka, M. *et al.* 1,5-anhydro-D-glucitol: a novel marker of glucose excursions. *International journal of clinical practice. Supplement*, 40–44 (2002).
- Buse, J. B., Freeman, J. L., Edelman, S. V., Jovanovic, L. & McGill, J. B. Serum 1,5-anhydroglucitol (GlycoMark): a short-term glycemic marker. *Diabetes technology & therapeutics* **5**, 355–363, doi:10.1089/152091503765691839 (2003).
- Dungan, K. M. 1,5-anhydroglucitol (GlycoMark) as a marker of short-term glycemic control and glycemic excursions. *Expert review of molecular diagnostics* **8**, 9–19, doi:10.1586/14737159.8.1.9 (2008).

5. Selvin, E. *et al.* Association of 1,5-Anhydroglucitol With Cardiovascular Disease and Mortality. *Diabetes* **65**, 201–208, doi:10.2337/db15-0607 (2016).
6. Selvin, E. *et al.* Association of 1,5-anhydroglucitol with diabetes and microvascular conditions. *Clin Chem* **60**, 1409–1418, doi:10.1373/clinchem.2014.229427 (2014).
7. Ikeda, N., Hara, H., Hiroi, Y. & Nakamura, M. Impact of serum 1,5-anhydro-d-glucitol level on prediction of major adverse cardiac and cerebrovascular events in non-diabetic patients without coronary artery disease. *Atherosclerosis* **253**, 1–6, doi:10.1016/j.atherosclerosis.2016.08.016 (2016).
8. Kuroda, M. *et al.* Effect of daily glucose fluctuation on coronary plaque vulnerability in patients pre-treated with lipid-lowering therapy: a prospective observational study. *JACC. Cardiovascular interventions* **8**, 800–811, doi:10.1016/j.jcin.2014.11.025 (2015).
9. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543–550, doi:10.1038/ng.2982 <http://www.nature.com/ng/journal/v46/n6/abs/ng.2982.html> - supplementary-information (2014).
10. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54–60, doi:10.1038/nature10354 (2011).
11. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature genetics* **42**, 105–116, doi:10.1038/ng.520 (2010).
12. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229–3239, doi:10.2337/db10-0502 (2010).
13. Olds, L. C. & Sibley, E. Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a cis regulatory element. *Hum Mol Genet* **12**, 2333–2340, doi:10.1093/hmg/ddg244 (2003).
14. Joshi, S. R. *et al.* Therapeutic potential of alpha-glucosidase inhibitors in type 2 diabetes mellitus: an evidence-based review. *Expert opinion on pharmacotherapy* **16**, 1959–1981, doi:10.1517/14656566.2015.1070827 (2015).
15. Ghani, U. Re-exploring promising alpha-glucosidase inhibitors for potential development into oral anti-diabetic drugs: Finding needle in the haystack. *European journal of medicinal chemistry* **103**, 133–162, doi:10.1016/j.ejmech.2015.08.043 (2015).
16. Polonsky, K. S., Given, B. D. & Van Cauter, E. Twenty-four-hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. *J Clin Invest* **81**, 442–448, doi:10.1172/jci113339 (1988).
17. Tsukuda, K. *et al.* Evaluation of the 24-hour profiles of physiological insulin, glucose, and C-peptide in healthy Japanese volunteers. *Diabetes technology & therapeutics* **11**, 499–508, doi:10.1089/dia.2009.0029 (2009).
18. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419–1260419, doi:10.1126/science.1260419 (2015).
19. Grempler, R. *et al.* Functional characterisation of human SGLT-5 as a novel kidney-specific sodium-dependent sugar transporter. *FEBS Lett* **586**, 248–253, doi:10.1016/j.febslet.2011.12.027 (2012).
20. Fukuzawa, T. *et al.* SGLT5 reabsorbs fructose in the kidney but its deficiency paradoxically exacerbates hepatic steatosis induced by fructose. *PLoS One* **8**, e56681, doi:10.1371/journal.pone.0056681 (2013).
21. Wright, E. M. Glucose transport families SLC5 and SLC50. *Molecular aspects of medicine* **34**, 183–196, doi:10.1016/j.mam.2012.11.002 (2013).
22. Long T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet* **49**(4), 568–578 (2017).
23. Chen, L. Q. *et al.* Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532, doi:10.1038/nature09606 (2010).
24. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *American Journal of Epidemiology* **129**, 687–702 (1989).
25. Maruthur, N. M. *et al.* Genetics of Plasma Soluble Receptor for Advanced Glycation End-Products and Cardiovascular Outcomes in a Community-based Population: Results from the Atherosclerosis Risk in Communities Study. *PLoS One* **10**, e0128452, doi:10.1371/journal.pone.0128452 (2015).
26. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods* **10**, 5–6, doi:10.1038/nmeth.2307 (2013).
27. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**, 955–959, doi:10.1038/ng.2354 (2012).
28. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904–909, doi:10.1038/ng1847 (2006).
29. Selvin, E. *et al.* Racial differences in glycemic markers: a cross-sectional analysis of community-based data. *Ann Intern Med* **154**, 303–309, doi:10.7326/0003-4819-154-5-201103010-00004 (2011).
30. Selvin, E., Coresh, J., Zhu, H., Folsom, A. & Steffes, M. W. Measurement of HbA1c from stored whole blood samples in the Atherosclerosis Risk in Communities study. *Journal of diabetes* **2**, 118–124, doi:10.1111/j.1753-0407.2010.00070.x (2010).
31. Marchini, J. *et al.* A new multipoint method for genome-wide association studies via imputation of genotypes. *Nature Genetics* **39**, 906–913 (2007).
32. Bush, W. S. & Moore, J. H. Chapter 11: Genome-Wide Association Studies. *PLoS Comput Biol* **8**, e1002822, doi:10.1371/journal.pcbi.1002822 (2012).
33. Peér, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* **32**, 381–385, doi:10.1002/gepi.20303 (2008).
34. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)* **26**, 2336–2337, doi:10.1093/bioinformatics/btq419 (2010).
35. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575, doi:10.1086/519795 (2007).
36. Moayyeri, A., Hammond, C. J., Hart, D. J. & Spector, T. D. The UK Adult Twin Registry (TwinsUK Resource). *Twin research and human genetics: the official journal of the International Society for Twin Studies* **16**, 144–149, doi:10.1017/thg.2012.89 (2013).
37. Wichmann, H. E., Gieger, C. & Illig, T. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))* **67**(Suppl 1), S26–30, doi:10.1055/s-2005-858226 (2005).
38. Volzke, H. *et al.* Cohort profile: the study of health in Pomerania. *Int J Epidemiol* **40**, 294–307, doi:10.1093/ije/dyp394 (2011).
39. Knacke, H. *et al.* Metabolic Fingerprints of Circulating IGF-1 and the IGF-1/IGFBP-3 Ratio: A Multifluid Metabolomics Study. *The Journal of clinical endocrinology and metabolism* **101**, 4730–4742, doi:10.1210/jc.2016-2588 (2016).
40. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190–2191, doi:10.1093/bioinformatics/btq340 (2010).

Acknowledgements

The Atherosclerosis Risk in Communities 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. Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). This research was also supported by NIH/NIDDK grant R01DK089174 to Dr. Selvin. Dr. Selvin was also supported by NIH/NIDDK grant K24DK106414. The work of Dr. Köttgen was supported by the German Research Foundation (Heisenberg professorship and CRC 1140). Dr. Li was supported by a National Heart, Lung, and Blood Institute T32-HL0072024 Cardiovascular Epidemiology Training Grant. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants nos 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network ‘Greifswald Approach to Individualized Medicine (GANI_MED)’ funded by the Federal Ministry of Education and Research (grant 03IS2061A). The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH. The authors thank the ARIC Study staff and participants for their important contributions. Reagents for the 1,5-anhydroglucitol assays were donated by the GlycoMark Corporation. Data on glycemetic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org. This work is dedicated to the memory of our friend and colleague, Dr. Wen Hong Linda Kao (1972–2014).

Author Contributions

N.M.M., E.S. and A.K. designed this study. N.M.M., M.P., K.E.N., H.M., A.C.M., N.F., J.P., M.N., E.B., A.T., E.S. and A.K. were involved in the study management. M.L., N.M.M., J.S.P., A.T., E.S. and A.K. interpreted the results. M.L., N.M.M., S.J.L., E.S. and A.K. drafted the manuscript. M.L., N.M.M., A.T., E.S. and A.K. developed statistical methods and performed the analyses. E.B., A.M. performed genotyping. M.L. and A.K. conducted the bioinformatics analyses. All authors critically reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-02287-x](https://doi.org/10.1038/s41598-017-02287-x)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017