Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose

Laura J. Rasmussen-Torvik[1], **Alvaro Alonso**[2], **Man Li**[3], **Wen Kao**[3], **Anna Köttgen**[3], **Yuer Yan**[4], **David Couper**[4], **Eric Boerwinkle**[5], **Suzette J. Bielinski**[6], and **James S. Pankow**[2]

[1]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL

[2]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN

[3]Division of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD

[4]Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC

[5]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX

[6]Division of Epidemiology, Mayo Clinic, Rochester, MN

Abstract

Although GWAS have been performed in longitudinal studies, most used only a single trait measure. GWAS of fasting glucose have generally included only normoglycemic individuals. We examined the impact of both repeated measures and sample selection on GWAS in ARIC, a study which obtained four longitudinal measures of fasting glucose and included both individuals with and without prevalent diabetes. The sample included Caucasians and the Affymetrix 6.0 chip was used for genotyping. Sample sizes for GWAS analyses ranged from 8372 (first study visit) to 5782 (average fasting glucose). Candidate SNP analyses with SNPs identified through fasting glucose or diabetes GWAS were conducted in 9133 individuals, including 761 with prevalent diabetes. For a constant sample size, smaller p-values were obtained for the average measure of fasting glucose compared to values at any single visit, and two additional significant GWAS signals were detected. For four candidate SNPs (rs780094, rs10830963, rs7903146, and rs4607517), the strength of association between genotype and glucose was significantly (p-interaction < .05) different in those with and without prevalent diabetes and for all five fasting glucose candidate SNPs (rs780094, rs10830963, rs560887, rs4607517, rs13266634) the association with measured fasting glucose was more significant in the smaller sample without prevalent diabetes than in the larger combined sample of those with and without diabetes. This analysis demonstrates the potential utility of averaging trait values in GWAS studies and explores the advantage of using only individuals without prevalent diabetes in GWAS of fasting glucose.

Keywords

GWAS; fasting glucose; type 2 diabetes; sample selection

Several genome-wide association studies (GWAS) of fasting blood glucose have been published. All found SNPs in one or more of three gene regions (*MTNR1B*, *G6PC2*, and

---

Contact information: Dr. Laura Rasmussen-Torvik Department of Preventive Medicine 680 N. Lake Shore Drive Suite 1400 Chicago, IL 60611 ljrtorvik@northwestern.edu p. 312-503-3596 f. 312-908-9588.

GCK) to be significantly ($p < 5 \times 10^{-8}$) associated with fasting glucose [Prokopenko et al., 2009, Chen et al., 2008, Bouatia-Naji et al., 2009, Bouatia-Naji et al., 2008, Chambers et al., 2009]. A recent meta-analysis performed in MAGIC (the Meta-Analysis of Glucose and Insulin traits Consortium) including over 100,000 individuals in discovery and replication samples has identified 13 additional SNPs significantly associated with fasting blood glucose [Dupuis et al., 2010]. However, previous GWAS of fasting glucose have typically restricted the study sample to include only normoglycemic or non-diabetic individuals. Additionally, although many prospective studies with multiple trait measures have published GWAS papers or participated in large GWAS consortia, there has been little investigation of the consistency of GWAS results over multiple study measures or the utility of using average measures in GWAS.

We aimed to address the following questions: (1) Are fasting glucose GWAS results consistent for multiple longitudinal measures of fasting glucose? (2) Does using averaged glucose levels over time provide additional information to single glucose measures in genetic analyses? (3) Are associations between SNPs and fasting glucose different in normoglycemic and diabetic individuals? To address these questions, we used the population-based Atherosclerosis Risk In Communities (ARIC) study. In the ARIC study, four measures of fasting glucose were taken over a period of approximately nine years. Additionally, the ARIC study included individuals with the complete distribution of fasting glucose values (normoglycemic, impaired fasting glucose, and those with diabetes). To evaluate consistency of results across the four time points and explore the potential advantage of having repeated measures, we conducted genome-wide association analyses of fasting glucose values for each of the four time points as well as for the average value. To evaluate the potential impact of sample selection (i.e. including or excluding subjects with diabetes), we analyzed, at a single time point, associations between fasting glucose and 22 candidate SNPs identified through GWAS of fasting glucose or type 2 diabetes in other study populations [Dupuis et al., 2010, Zeggini et al., 2008, Lango et al., 2008]

Materials and Methods

Study Sample

The ARIC study is a multi-center prospective cohort study focused on cardiovascular disease [The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. 1989]. Men and women aged 45–64 years at baseline were recruited from four communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland. A total of 15,792 individuals participated in the baseline examination in 1987–1989. There were three triennial follow-up examinations in 1990–1992 (93% return rate), 1993–1995 (86% return rate), and 1996–1998 (81% return rate). The study was approved by the institutional review board at each field center and all participants gave informed consent.

Phenotypic Measurements

Blood samples were drawn from an antecubital vein into tubes containing a serum separator gel. Blood samples were shipped to a central lab for analysis. For visits 1 and 2, blood samples were analyzed at a central lab in Minneapolis, MN. For visits 3 and 4, blood samples were analyzed at a central lab in Houston, TX. Glucose was measured by a hexokinase/glucose-6-phosphate dehydrogenase method on a Coulter DACOS device (Beckman Coulter, Fullerton, CA). BMI was calculated from participants' heights and weights measured in scrub suits.

Genotyping

In the ARIC Study, genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0. Subjects who disallowed DNA use, unintentional duplicates with higher

missing genotype rates, suspected mixed/contaminated samples, scans from one problem plate, samples with a mismatch between called and phenotypic sex, samples with genotype mismatch with 39 previously genotyped SNPs, suspected first-degree relative of an included individual, and genetic outliers based on average IBS statistics and principal components analysis using EIGENSTRAT were excluded. SNPs were excluded due to no chromosome location, being monomorphic, having call rate <95%, MAF < 1% or HWE-p < 10-5. After the filtering, 669,450 SNPs were used in the imputation to 2,543,887 autosomal SNPs from HapMap Phase II CEU samples using MACH v1.0.16. A more detailed description can be found in the supplementary methods in the online appendix. We restricted all analyses to Caucasians with cleaned genotype information available from the Affymetrix Genome-Wide Human SNP Array 6.0 (n = 9345).

GWAS Analysis of repeated fasting glucose measures

For GWAS analyses of fasting glucose, we excluded participants not fasting at least 8 hours at the relevant study visit and participants with prevalent diabetes at the relevant study visit to be consistent with previous GWAS analyses. Prevalent diabetes was defined as the presence of any of the following: a fasting serum glucose of $\geq$ 126 mg/dl (7.0mmol/L); a non-fasting serum glucose level of $\geq$200 mg/dl (11.1mmol/L); self-reported physician diagnosis of diabetes; or self-reported pharmacologic treatment of diabetes in the past two weeks. Individuals missing data critical to the definition of prevalent diabetes were also excluded from the analysis. After exclusions, there were 8372, 7871, 7099, and 6421 subjects included at visits 1–4 respectively. We also performed GWAS analyses for fasting glucose measured at visits 1–4 and the average of these four measures excluding all individuals missing a glucose measurement at any visit or having prevalent diabetes at any visit (n=5782).

GWAS analyses were performed in ProbABEL. Untransformed fasting glucose was regressed linearly on genotype adjusted for age, sex, and ARIC study center. Genotypes were modeled additively. For imputed genotypes, a genotype dose was used that accounted for the certainty with which the genotypes were estimated. To test whether the association of some SNPs with fasting glucose differed between measures at different visits, or between visit 1 and the average of fasting glucose, we used a repeated measures regression model implemented in SAS Proc Mixed v 9.1 (Cary, NC) and tested the significance of a measurement * SNP interaction term in the model. To define a significant association for GWAS, we followed the guidance of Frayling et al. [Frayling, 2007] and used a p-value cut-off of $p < 5 \times 10^{-8}$. Post-hoc power analyses were performed using the computer program Quanto [Gauderman and Morrison, 2009, Gauderman, 2002].

Candidate SNP analysis of fasting glucose measured at visit 1 in different study samples

For candidate SNP analyses, individuals were excluded if they were not fasting at least 8 hours prior to the visit 1 examination, missing diabetes status at visit 1, or had a visit 1 fasting glucose value greater than 22 mmol/L (in order to exclude outliers). After these exclusions there were 9133 individuals available for candidate SNP analyses.

Type 2 diabetes candidate gene regions and SNPs were selected from Lango et al [Lango et al., 2008]. Fasting glucose candidate gene regions and SNPs were selected from the MAGIC meta-analysis results [Dupuis et al., 2010]. Using these selection techniques, there was only one overlapping region between the diabetes and fasting glucose candidate SNPs (SNPs near the *SLC30A8* gene). Candidate SNPs and gene regions are listed in Supplementary Table 1 in the online appendix. The estimated r^2 between the imputed genotype and the actual genotype was only .29 for SNP rs757210 so this SNP was excluded. SNP rs5219 was not genotyped or imputed in our sample, so SNP rs5215 was substituted. This SNP was the top signal detected in the *KCNJ11* region in a diabetes GWAS meta-analysis reported by Zeggini [Zeggini et al., 2008]. For typed SNPs, missing genotypes were imputed by the MACH program, such that all

individuals had complete genotyping for all candidate SNPs. For imputed genotypes, most likely estimated genotypes were used.

Tests of association for candidate SNP analyses were performed using linear regression implemented in SAS Proc GLM with SNPs modeled additively. For some analyses in individuals with prevalent diabetes or in the combined (diabetic and non-diabetic) population, glucose was log transformed to account for the skewed distribution of glucose. To test whether the association between candidate SNPs and fasting glucose differed between individuals with and without prevalent diabetes at visit 1, a diabetes * SNP interaction term was included in linear regression models of log-transformed fasting glucose in the combined sample. Because of concerns that use of glucose-lowering medication use among those with prevalent diabetes might bias regression analyses involving fasting glucose, analyses were also conducted excluding all individuals reporting the use of glucose-lowering medications (n = 242). We used a significance cut-off of p < .002 (.05/21) for candidate SNP analyses.

Results

GWAS of repeated fasting glucose measures

Table 1 shows the sample characteristics at each study visit for ARIC participants eligible for GWAS analyses. Because of losses to follow-up and the increase in the prevalence of diabetes as the cohort aged, the number of eligible participants for fasting glucose GWAS analyses decreased across visits. There was a small consistent increase in BMI from the visit 1 to visit 4 samples. Mean fasting glucose increased slightly between visit 1 and visit 2 samples and again between visit 3 and visit 4. Mean fasting glucose decreased between the visit 2 and visit 3 samples; the central laboratory responsible for this measurement changed between visits 2 and 3 and this reduction may thus be due to slight variations in the lab assay. The rank order of glucose values remained fairly consistent across visits; the spearman correlation coefficients with visit 1 fasting glucose were r = .56, .49, and .52 for fasting glucose measured at visits 2–4, respectively. The standard deviation for the average (v1–v4) fasting glucose measure was smaller than the standard deviations for the visit 1, 2, 3, or 4 measures of fasting glucose.

Figure 1 shows Manhattan plots representing all p-values for SNP associations in the GWAS of visits 1, 2, 3, and 4 fasting glucose and average fasting glucose. These plots include only the individuals who both had measures of fasting at all four study visits and did not have prevalent diabetes at all four study visits (n = 5782). Supplementary Figure 1 shows similar Manhattan plots for v1–v4, but includes all individuals with a glucose measurement and without prevalent diabetes at the relevant visit, thus including more individuals than for the plots presented in Figure 1. There were five regions where SNPs exceeded the threshold for significance ($p < 5 \times 10^{-8}$) in at least one GWAS. These regions were in or near genes *GCKR* (chromosome 2), *G6PC2* (chromosome 2), *GCK* (chromosome 7), *SLC30A8* (chromosome 8), and *MTNR1B* (chromosome 11). Supplementary Figure 2 in the online appendix shows the QQ plots for all GWAS analyses presented in Figure 1. For all GWAS analyses, the distribution of p-values was as expected under the null, until the p-values dropped below $p < 10^{-5}$ at which point there was an excess of p-values compared to what would be expected in the absence of any SNP associations.

Table 2 lists the SNP with the smallest p-value in each of the five regions from the GWAS of visits 1–4 fasting glucose and the average measure. In some cases, the SNP with the smallest p-value in each region varied by visit. However, as shown in Supplementary Table 2, there was considerable LD ($r^2 > .7$) between the SNPs with the smallest p-values across visits within these regions. There were significant SNP associations detected in or near *GCK*, *G6PC2*, and *MTNR1B* for all GWAS. We found significant associations for SNPs near *GCKR* and *SLC30A8* only for the GWAS of average fasting glucose. Post-hoc power calculations of SNP

rs13266634 in the gene region *SLC30A8* for a sample size of 5782 individuals revealed 74.3% power to detect a p-value of $5 * 10^{-8}$ for average fasting glucose, but only 28.4% power to detect a p-value of $5 * 10^{-8}$ for visit 1 fasting glucose.

To further examine heterogeneity in the strength of association between SNPs and fasting glucose measured at multiple visits, one SNP was chosen to represent each region (*GCKR*—rs780094, *G6PC2*—rs560887, *GCK*—rs4607517, *MTNR1B*—rs10830963, *SLC30A8*—rs13266634). Mixed regression models including measurement*SNP interaction terms indicated no significant difference in the strength of association between SNPs and fasting glucose across visits 1–4 or when comparing visit 1 to average fasting glucose for any of these 5 SNPs. To illustrate this consistency across visits, the five above SNPs are highlighted in the Manhattan plots presented in Figures 1A–1E.

The standard errors for SNP associations with average fasting glucose were smaller than for those with individual visit fasting glucose measures due to the smaller standard deviation of the trait. For each region the most significant SNP associations were seen with average fasting glucose. Supplementary Table 3 is similar to Table 2, but includes all individuals with a fasting glucose measurement and without prevalent diabetes at the relevant visit, thus including more individuals than for Table 2. Although the visit 1 fasting glucose GWAS in Supplementary Table 3 includes almost 2600 more individuals that the average fasting glucose GWAS in Table 2, smaller p-values were observed in the GWAS of average fasting glucose for all but one of the five gene regions. When all 8372 individuals without prevalent diabetes at visit 1 were included in a post-hoc power calculation for SNP rs13266634, there was only 66.3% power to detect a p-value of $5 * 10^{-8}$ for visit 1 fasting glucose. This is less power than was found in the similar post-hoc analysis for average fasting glucose with a sample size of 5782 (power = 74.3%).

Candidate SNP analysis of fasting glucose in different study samples

Table 3 shows the characteristics of ARIC participants included in candidate SNP analyses. Mean BMI and fasting glucose were greater in those with prevalent diabetes at visit 1 than in those without. Excluding individuals taking glucose-lowering medications at visit 1 reduced both the mean and standard deviation of fasting glucose in those with prevalent diabetes and in combined study sample. Supplementary Figure 3 in the online appendix shows the distributions of fasting glucose at visit 1 among subjects without diabetes, subjects with diabetes, and the combined sample after excluding those taking glucose lowering medications.

Table 4 shows all significant (p < .002) candidate SNP associations with visit 1 fasting glucose in three subgroups: subjects without prevalent diabetes, subjects with prevalent diabetes, and a combined sample of all subjects regardless of diabetes status. Results are also shown for the diabetes and combined subgroups after excluding subjects taking glucose lowering medications. Associations with only six of 22 SNPs achieved the pre-set level of significance in any subgroup. Log-transforming glucose values for those with prevalent diabetes and the combined sample did not increase the number of SNPs achieving significance (p-values for SNPs not achieving this level of significance are presented in Supplementary Table 4).

Effect sizes (betas) for 4 of the 5 fasting glucose candidate SNPs (rs560887 of *G6PC2*, rs4607517 of *GCK*, rs780094 of *GCKR*, and rs13266634 of *SLC30A8*) were larger in the combined sample than in subjects without prevalent diabetes, but p-values were larger due to the larger standard errors in the combined sample. In contrast, SNP rs7903146 of *TCF7L2* had larger effect sizes and lower p-values in the combined sample compared to the sample without prevalent diabetes. No associations reached the pre-set level of significance in the smaller sub-sample with prevalent diabetes. The strength of association between SNP and fasting glucose was found to be significantly different between those with and those without prevalent diabetes

for rs10830963 (p-interaction, 3.3 E −11), rs780094 (p-interaction, .0067), rs4607517 (p-interaction, .02), and rs7903146 (p-interaction, .0050). In general, associations in the combined sample became more significant after excluding subjects taking glucose lowering medications.

Discussion

This study had two distinct, but related aims: (1) to conduct a GWAS of fasting glucose to evaluate consistency of results across multiple time points and explore the potential advantage of having repeated measures of the trait; and (2) to conduct candidate SNP association analyses with fasting glucose to evaluate the potential impact of sample selection, particularly the inclusion or exclusion of subjects with prevalent diabetes. Our results may have general implications beyond the specific gene variants and phenotypes investigated here.

Impact of Repeated Measures

In this GWAS of fasting glucose in the ARIC study, SNPs in five regions of the genome achieved genome-wide levels of significance for an average measure of fasting glucose. All five regions also reached genome-wide levels of significance in the recent MAGIC meta-analysis, for which the ARIC Study provided results for selected SNPs from visit 1 as an in silico replication sample [Dupuis et al., 2010].

The results presented in this paper represent one of the first attempts to conduct GWAS on multiple or average trait measures. The use of average and multiple measures in this analysis has demonstrated the potential utility of using average trait measures in GWAS analyses; the average fasting glucose trait had a smaller standard deviation than any single visit measure of fasting glucose which resulted in smaller standard errors in regression with average glucose and, accordingly, smaller p-values, despite the smaller sample size for average analyses (compared to individual visit analyses). In our data, for a set sample size (n = 5782), two regions (*GCKR* and *SLC30A8*) were detected with a threshold of $p < 5 * 10^{-8}$ in a GWAS of average glucose, that were not detected in GWAS analyses of any single visit measures of glucose. Additionally, post-hoc power calculations demonstrated that models using the average measure of fasting glucose had considerably greater power than models using a single measure of fasting glucose. Clearly, some caution must be taken in averaging traits which could change meaningfully over long-term follow-up, but when a trait can be expected to maintain relatively stable ranks across visits, the use of an average measure may represent an important tool for detecting GWAS signals not otherwise possible in smaller samples.

Impact of Sample Selection

The absence of significant associations between measured fasting glucose and type 2 diabetes candidate SNPs in the combined sample was interesting, given our relatively low threshold for significance (p < .002). Although Dupuis et al. previously reported only very modest associations between diabetes candidate SNPs and fasting glucose in a large GWAS meta-analysis [Dupuis et al., 2010] of non-diabetic individuals, we hypothesized we might observe more robust associations in the combined sample with the full range of fasting glucose values. The lack of association between most diabetes candidate SNPs and fasting glucose in the combined sample may reflect the absence of an association between the SNPs and fasting glucose in the normal range. Rather than promote a small increase in fasting glucose across the entire range of glucose values, many of these SNPs may increase the likelihood of conversion to diabetes without altering glucose levels in the normal range for most individuals. This hypothesis is supported by the pattern of association seen with *TCF7L2* SNP rs7903146, the only diabetes SNP significantly associated with fasting glucose in this analysis; the association of this SNP and fasting glucose was much stronger in the entire study sample, when those with prevalent diabetes were included, than in the subset of the sample without prevalent

diabetes. Additionally, the association was stronger when glucose-lowering medication users (who, on average, had higher levels of glucose) were included. For some other diabetes candidate SNPs with smaller effects on diabetes risk, the association with diabetes may not be sufficient to produce a significant association with fasting glucose in the combined sample absent a measureable effect of glucose in the normal range, where most individuals in the population are found.

The results of the fasting glucose candidate SNP associations with fasting glucose in different study sub-samples were surprising. Initially, we assumed that associations between fasting glucose candidate SNPs and fasting glucose would be most significant in the entire study sample where the full range of glucose values was present, but this was not the case for any fasting glucose candidate SNPs. Effect (beta) estimates were generally larger in the entire study sample with the complete trait distribution, but the increase in standard errors driven by the increase in the standard deviation of glucose in the entire sample overwhelmed the increased beta estimates such that p-values for association with fasting glucose candidate SNPs were larger in the entire sample, even after log transformation of the trait. When medication users were removed, this generally attenuated the beta estimate the in the combined study sample, possibly because medication users had higher fasting glucose levels, on average, then those with prevalent diabetes not on medication.

Another intriguing explanation for the lack of association in the combined sample is the possibility of different associations between candidate SNPs and fasting glucose in the sub-sample with prevalent diabetes. Interestingly, for four SNPs (rs10830963, rs780094, rs4607517, and rs7903146) there was a statistically significant difference in the strength of the association between SNP and fasting glucose in those with and without prevalent diabetes. For three of the four SNPs (all but rs780094) associations between SNP and genotype were either attenuated or in the opposite direction (i.e. minor allele decreasing rather than increasing glucose) among those with prevalent diabetes compared to those without. This suggests perhaps other metabolic derangements associated with diabetes overwhelm or alter the effects of the variants in some individuals with prevalent diabetes. The number of individuals with prevalent diabetes at visit 1 in ARIC was relatively low (n = 761) and still smaller after individuals reporting glucose-lowering medication use were excluded (n = 519), and no candidate SNP associations with fasting glucose were even nominally significant (p < .05) in those with prevalent diabetes. Nonetheless, the significance of the interaction terms testing the difference of strength association between those with and without prevalent diabetes is compelling and warrants further study.

Previous GWAS studies of fasting glucose have been conducted only in normogylcemic individuals to allow the use of some case-control samples [Prokopenko et al., 2009, Bouatia-Naji et al., 2009], to investigate variation in the normal range, and to eliminate potential confounding by treatment effects [Dupuis et al., 2010]. The results of this study seem to validate this approach, as the associations of fasting glucose candidate genes are less significant in the combined sample than when restricted non-diabetic subjects. However, it is of interest to further examine the association of both type 2 diabetes and fasting glucose candidate genes with measured fasting glucose in those with prevalent diabetes, given the significant interaction p-values for diabetes status found in our candidate gene analysis. Cases from type 2 diabetes case-control studies excluded from initial GWAS examinations of fasting glucose would be logical populations for study. Such analyses will require careful consideration of medication effects in these populations.

Conclusions

In conclusion, the analyses reported herein suggest that the use of average traits in GWAS analyses may be advantageous in appropriate situations. Additionally, SNPs identified in

previous fasting glucose GWAS were more significantly associated with fasting glucose in those without prevalent diabetes than in those with prevalent diabetes or those in a combined sample, demonstrating the critical importance of sample selection in genetic association studies. Strengths of the analysis include relatively comprehensive coverage of the genome, with 2.5 million imputed and typed SNPs, large sample size, and a population-based study with long-term follow-up and excellent retention. Weaknesses include the restriction to Caucasians, and small sample size in the sub-sample with prevalent diabetes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129:687–702. [PubMed: 2646917]

Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C, et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet 2009;41:89–94. [PubMed: 19060909]

Bouatia-Naji N, Rocheleau G, Van Lommel L, et al. A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. Science 2008;320:1085–8. [PubMed: 18451265]

Chambers JC, Zhang W, Zabaneh D, et al. Common genetic variation near melatonin receptor MTNR1B contributes to raised plasma glucose and increased risk of type-2 diabetes amongst Indian Asians and European whites. Diabetes 2009;58:2703–8. [PubMed: 19651812]

Chen WM, Erdos MR, Jackson AU, et al. Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. J Clin Invest 2008;118:2620–8. [PubMed: 18521185]

Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42:105–16. [PubMed: 20081858]

Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet 2007;8:657–62. [PubMed: 17703236]

Gauderman WJ, Morrison JM. Quanto 1.1: A computer program for power and sample size calculations for genetic epidemiology studies. 2009 2010.

Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol 2002;155:478–84. [PubMed: 11867360]

Lango H, UK Type 2 Diabetes Genetics Consortium. Palmer CN, et al. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes 2008;57:3129–35. [PubMed: 18591388]

Prokopenko I, Langenberg C, Florez JC, et al. Variants in MTNR1B influence fasting glucose levels. Nat Genet 2009;41:77–81. [PubMed: 19060907]

Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638–45. [PubMed: 18372903]
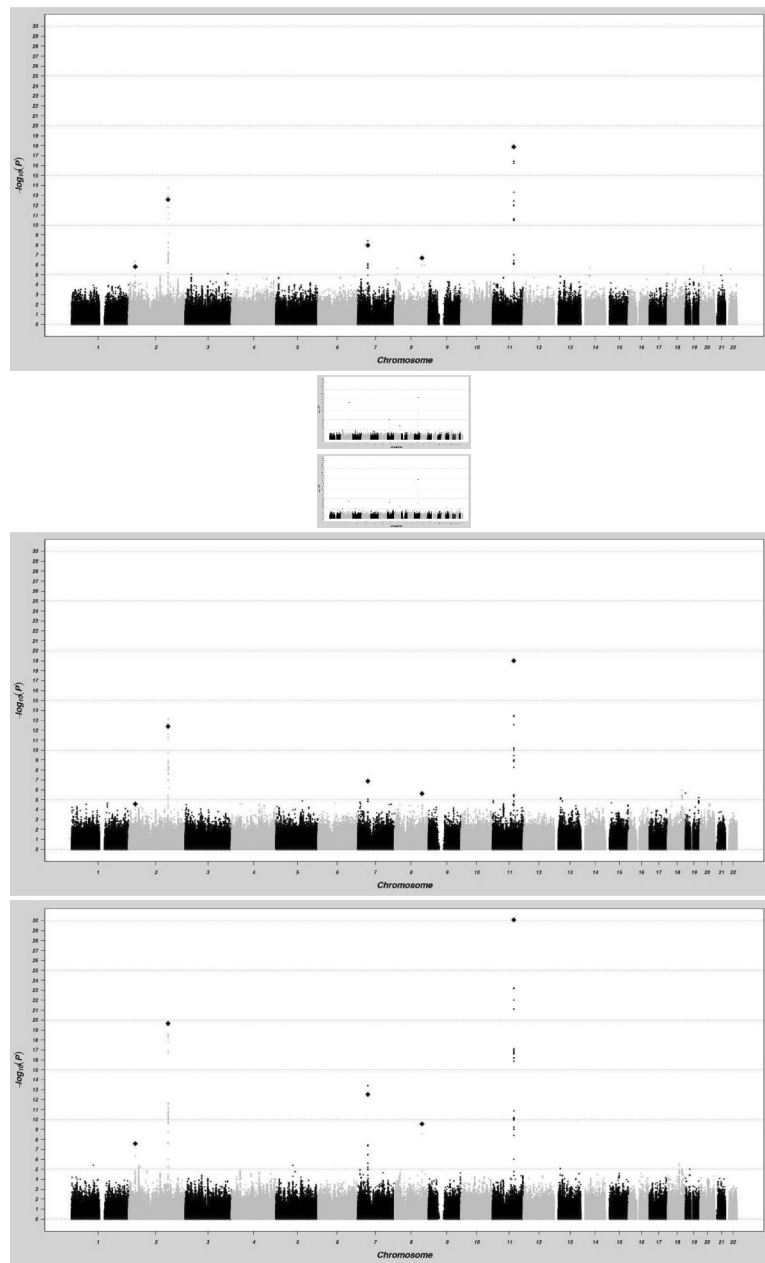
Figure 1.
Manhattan plots for GWAS of fasting glucose, ARIC study. Plots display the p-values for the association of approximately 2.5 million SNPs with fasting glucose; each dot represents a single SNP association plotted by chromosome position (x-axis) and log(10)-transformed p-value (y-axis). P-values are from regression equations with SNPs modeled additively, adjusted for age, sex, and study center. For all plots only individuals without prevalent diabetes at all four study visits included (n = 5782). A) GWAS of fasting glucose measured at visit 1, B) GWAS of fasting glucose measured at visit 2, C) GWAS of fasting glucose measured at visit 3, D) GWAS of fasting glucose measured at visit 4, E) GWAS of average fasting glucose (visit – visit 4). To demonstrate the consistency of certain SNPs associations across the trait measures, p-values for five SNPs are highlighted with black diamonds across all five plots. From left to right, the highlighted SNPs are: rs780094 (*GCKR* gene region), rs560887 (*G6PC2* gene region),

rs4607517 (*GCK* gene region), rs13266634 (*SLC30A8* gene region), rs10830963 (*MTNR1B* gene region).

Table 1

Characteristics of ARIC participants eligible for GWAS of fasting glucose

	visit 1	visit 2	visit 3	visit 4	average v1–v4[*]
n (individuals without prevalent diabetes)	8372	7871	7099	6421	5782
age (years)	54.1 (5.7)	57.0 (5.7)	60.1 (5.6)	62.9 (5.6)	54.0 (5.6)
% male	46.4	46.0	45.7	45.1	44.6
BMI (mg/kg/m^2)	26.7 (4.6)	26.9 (4.7)	27.5 (4.9)	27.8 (5.0)	26.3 (4.3)
fasting glucose	5.47 (.50)	5.62 (.52)	5.45 (.55)	5.48 (.52)	5.46 (.40)

Mean (SD)

[*] Average value of fasting glucose (v1–v4) presented—only individuals without prevalent diabetes at all visits were included. For all other covariates in this column, visit 1 characteristics are listed.

Table 2

GWAS analysis of fasting glucose across 4 study visits[*]: SNPs with smallest p-values for association in regions exceeding genome-wide threshold for significance.

region		visit 1	visit 2	visit 3	visit 4	average visit 1–4
MTN1RB	SNP	rs10830963	rs10830963	rs10830963	rs10830963	rs10830963
	$\beta^{\dagger}$	.082	.096	.100	.0094	.093
	SE	.0093	.0100	.0109	.0103	.0080
	p-value	1.34 E-18	8.64 E-22	3.79 E-20	1.01 E-19	8.44 E-31
G6PC2	SNP	rs853787	rs560887	rs560887	rs853787	rs853787
	$\beta^{\dagger}$	−.066	−.086	−.063	−.071	−.069
	SE	.0086	.0095	.0104	.0095	.0074
	p-value	1.58 E-14	2.84 E-19	1.52 E-9	7.65 E-14	1.6 E-20
GCK	SNP	rs2971669	rs2971669	rs2971669	rs2971669	rs2971669
	$\beta^{\dagger}$	.064	.080	.076	.064	.071
	SE	.0108	.0116	.0127	.0120	.0094
	p-value	3.77 E-9	7.60E-12	1.54 E-9	4.92 E-8	3.91 E-14
GCKR	SNP	rs1260326	rs1260326	rs780094	rs780094	rs780094
	$\beta^{\dagger}$	−.042	−.040	−.043	−.039	−.040
	SE	.0084	.0090	.0097	.0092	.0072
	p-value	4.46 E-7	9.29 E-6	8.22 E-6	2.66 E-5	2.69 E-8
SLC30A8	SNP	rs13266634	rs3802177	rs3802177	rs11774700	rs1326634
	$\beta^{\dagger}$	−.047	−.051	−.052	−.050	−.049
	SE	.0090	.0096	.0105	.0105	.0078
	p-value	2.05 E-7	1.35 E-7	7.48 E-7	1.73 E-6	2.81 E-10

[*] For all results presented in table, only individuals without diabetes at all visits included (n = 5782)

[†] mmol/l change per 1 increase copy in minor allele

Table 3

Characteristics of ARIC participants included in candidate SNP analyses

	visit 1 without prevalent diabetes	visit 1 with prevalent diabetes	visit 1 combined sample[*]
n	8372	761	9133
age (years)	54.1 (5.7)	56.2 (5.6)	54.3 (5.7)
% male	46.4	52.8	46.9
BMI (mg/kg/m^2)	26.7 (4.6)	30.4 (5.5)	27.0 (4.8)
fasting glucose	5.47 (.50)	9.24 (3.42)	5.79 (1.51)
fasting glucose with those on medication excluded[†]		8.37 (2.78)	5.64 (1.07)

[*] this column contains all individuals in the "visit 1 without prevalent diabetes" column, and the "visit 1 with prevalent diabetes" column, combined

[†] users of glucose-lowering medication excluded as described in the methods section; n excluded= 242

Table 4

Effect of sample selection on associations between SNPs and fasting glucose at visit 1: β-coefficients[*] (β), standard errors (SE), and p-values, for significant (p< .002) candidate SNP associations, controlled for age, sex, and field center, by sub-sample.

Gene	SNP	exclusions	no prevalent diabetes			prevalent diabetes				combined sample				
			β[*]	SE	p-value	β[*]	SE	p-value[††]	p-value[†††]	β[*]	SE	p-value[††]	p-value[†††]	interaction p-value[††††]
MTNR1B	rs10830963	none	0.081	0.009	2.9E-21	−0.228	0.198	0.25	0.16	0.069	0.025	0.007	2.0E-5	
MTNR1B	rs10830963	medication users[†]				−0.331	0.196	0.09	0.07	0.067	0.018	0.0002	5.5E-8	3.1 E -11
G6PC2	rs560887	none	−0.068	0.008	1.0E-16	−0.118	0.195	0.55	0.37	−0.091	0.024	2.0E-4	1.8E-7	
G6PC2	rs560887	medication users[†]				−0.048	0.195	0.81	0.77	−0.077	0.017	9.1E-6	3.3E-9	.23
GCK	rs4607517	none	0.068	0.010	1.0E-11	−0.067	0.228	0.77	0.65	0.095	0.030	0.001	1.7E-5	
GCK	rs4607517	medication users[†]				0.005	0.228	0.98	0.72	0.082	0.022	0.001	1.3E-6	.02
GCKR	rs780094	none	−0.046	0.008	1.7E-9	−0.150	0.192	0.44	0.41	−0.073	0.023	0.001	4.0E-5	
GCKR	rs780094	medication users[†]				−0.320	0.188	0.08	0.11	−0.082	0.016	4.7E-7	1.3E-8	.0067
SLC30A8	rs13266634	none	−0.044	0.008	6.0E-8	−0.051	0.190	0.79	0.74	−0.098	0.024	4.5E-5	3.6E-7	
SLC30A8	rs13266634	medication users[†]				−0.079	0.187	0.67	0.73	−0.074	0.017	1.5E-5	1.6E-7	.80
TCF7L2	rs7903146	none	0.0240	0.008	0.003	−0.001	0.186	0.99	0.99	0.127	0.024	1.4E-7	4.9E-9	
TCF7L2	rs7903146	medication users[†]				−0.187	0.190	0.32	0.39	0.059	0.017	.0007	5.3E-5	.0050

[*] mmol/l change per 1 increase copy in minor allele

[†] users of glucose-lowering medication excluded as described in the methods section; n excluded=242

[††] p-value for the regression of untransformed fasting glucose on genotype

[†††] p-value for the regression of log-transformed fasting glucose on genotype

[††††] A statistical test to determine if the strength of association between log-transformed fasting glucose and SNP differed between those with and without prevalent diabetes.