A 100K Genome-Wide Association Scan for Diabetes and Related Traits in the Framingham Heart Study

Replication and Integration With Other Genome-Wide Datasets

Jose C. Florez,[1,2,3] Alisa K. Manning,[4] Josée Dupuis,[4] Jarred McAteer,[1,2] Kathryn Irenze,[2] Lauren Gianniny,[2] Daniel B. Mirel,[2] Caroline S. Fox,[5,6] L. Adrienne Cupples,[4] and James B. Meigs[3,7]

OBJECTIVE—To use genome-wide fixed marker arrays and improved analytical tools to detect genetic associations with type 2 diabetes in a carefully phenotyped human sample.

RESEARCH DESIGN AND METHODS—A total of 1,087 Framingham Heart Study (FHS) family members were genotyped on the Affymetrix 100K single nucleotide polymorphism (SNP) array and examined for association with incident diabetes and six diabetes-related quantitative traits. Quality control filters yielded 66,543 SNPs for association testing. We used two complementary SNP selection strategies (a "lowest *P* value" strategy and a "multiple related trait" strategy) to prioritize 763 SNPs for replication. We genotyped a subset of 150 SNPs in a nonoverlapping sample of 1,465 FHS unrelated subjects and examined all 763 SNPs for in silico replication in three other 100K and one 500K genome-wide association (GWA) datasets.

RESULTS— We replicated associations of 13 SNPs with one or more traits in the FHS unrelated sample (16 expected under the null); none of them showed convincing in silico replication in 100K scans. Seventy-eight SNPs were nominally associated with diabetes in one other 100K GWA scan, and two (rs2863389 and rs7935082) in more than one. Twenty-five SNPs showed promising associations with diabetes-related traits in 500K GWA data; one of them (rs952635) replicated in FHS. Five previously reported associations were confirmed in our initial dataset.

CONCLUSIONS— The FHS 100K GWA resource is useful for follow-up of genetic associations with diabetes-related quantitative traits. Discovery of new diabetes genes will require larger samples and a denser array combined with well-powered replication strategies. ***Diabetes*** **56:3063–3074, 2007**

The genetic architecture of type 2 diabetes appears to be composed of several genes, each of which has a modest impact on disease risk (1). Despite significant advances in our understanding of the genetic determinants of the monogenic forms of diabetes (2), the definitive identification of genes that increase risk of common type 2 diabetes in the general population has been far more elusive.

Candidate gene studies have led to the association of several common variants with type 2 diabetes (3). Besides a handful of widely reproduced associations, however, many previously reported associations have not been convincingly replicated despite well-powered attempts to do so. The type 2 diabetes genetics literature is plagued by extensive and often conflicting reports of association. In addition, current gene discovery strategies have frequently focused on coding regions, which overlook regulatory variants that can also influence disease (4,5). Thus, identification of novel type 2 diabetes genes requires complementary approaches that identify high-likelihood variants on the basis of empiric associations derived from well-phenotyped, well-powered cohorts.

It is now possible to perform genome-wide association (GWA) studies, which are agnostic to biological plausibility and to the putative functional status of the assayed variants (6). The development of high-throughput genotyping platforms, the compilation of single nucleotide polymorphisms (SNPs) in public databases (7), the dissemination of new analytical tools and statistical methods (8–13), the assembly of large patient cohorts, and the availability of the HapMap (14,15) have all made it possible to scan the human genome for variants associated with disease, without imposing a priori assumptions that may bias the outcome of the scan. Several GWA studies for type 2 diabetes have been performed in recent months (16–20), making it possible to integrate data, replicate findings, extend them into other populations, and perform more detailed phenotypic characterizations.

Here, we report results from the Framingham Heart Study (FHS) 100K SNP GWA scan for type 2 diabetes and

From the [1]Center for Human Genetic Research and Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts; the [2]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts; the [3]Department of Medicine, Harvard Medical School, Boston, Massachusetts; the [4]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; the [5]Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; the [6]National Heart, Lung, and Blood Institute Framingham Heart Study, Framingham, Massachusetts; and the [7]General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts.

Address correspondence and reprint requests to James B. Meigs, MD, MPH, General Medicine Division, Massachusetts General Hospital, 50 Staniford St., 9th Floor, Boston, MA 02114. E-mail: jmeigs@partners.org.

Received for publication 2 April 2007 and accepted in revised form 5 September 2007.

Published ahead of print at http://diabetes.diabetesjournals.org on 11 September 2007. DOI: 10.2337/db07-0451.

J.B.M. has received research grants from GlaxoSmithKline and Wyeth and serves on safety or advisory boards for GlaxoSmithKline and Lilly.

Additional information for this article can be found in an online appendix at http://dx.doi.org/10.2337/db07-0451.

DGI, Diabetes Genetics Initiative; FPG, fasting plasma glucose; FBAT, family-based association test; FHS, Framingham Heart Study; GEE, generalized estimating equations; GWA, genome-wide association; HOMA-IR, homeostasis model assessment of insulin resistance; ISI, insulin sensitivity index; MAF, minor allele frequency; mFPG, 28-year mean fasting plasma glucose; NIH, National Institutes of Health; SNP, single nucleotide polymorphism.

TABLE 1
Characteristics of the two nonoverlapping independent FHS samples

	Family sample (100K GWA)	Unrelated sample (independent replication)
n	1,087	1,465
Age at exam 5 (years)	1,032 (51.5 ± 9.8)	1,390 (56.1 ± 9.3)
Women	51.5	52.8
Diabetes	8.4	10.8
Mean BMI (kg/m^2)	1,026 (27.5 ± 5.2)	1,384 (27.4 ± 4.8)
Exam 5 FPG (mg/dl)	1,027 (98.9 ± 24.7)	1,383 (101.5 ± 28.7)
Exam 5 A1C (%)	623 (5.3 ± 0.9)	1,135 (5.5 ± 1.0)
28-year mean FPG (mg/dl)	1,087 (97.9 ± 16.2)	1,465 (100.4 ± 17.4)
Exam 5 fasting insulin (μU/ml)	982 (30.1 ± 16.4)	1,337 (30.9 ± 13.4)
Exam 5 HOMA insulin resistance	980 (7.8 ± 7.3)	1,337 (8.1 ± 6.0)
Exam 5 gutt ISI	935 (26.1 ± 7.6)	1,248 (25.4 ± 7.4)

Data are *n* (means ± SD) or %.

related traits; replication of top 100K findings in an independent, unrelated FHS sample; initial integration of FHS 100K data with three other 100K (21–23) and one 500K (http://www.broad.mit.edu/diabetes/) type 2 diabetes GWA scans; and the use of the FHS 100K resource to confirm high-likelihood associations reported by others (16–20). This scan complements other large extant type 2 diabetes GWA studies in three major respects: It is population based (not diabetes proband based), its genetic information comprises two generations, and it is based on compiled data from decades of longitudinal standardized follow-up with detailed phenotyping of the offspring generation. A general and preliminary description of the full FHS 100K GWA resource has been published elsewhere (24).

RESEARCH DESIGN AND METHODS

The FHS. The FHS is a community-based, multigenerational, longitudinal study of cardiovascular disease and its risk factors, including diabetes. The FHS comprises the original cohort, offspring, and generation 3 studies. Subjects described in the present analysis include 1,087 individuals from the FHS offspring "family sample," composed of the 307 largest pedigrees previously selected for linkage analyses (25). These subjects, 560 of whom were women and whose mean age at last follow-up was 59 years, were genotyped on the Affymetrix 100K array (Table 1). The study was approved by Boston University's Institutional Review Board, and informed consent, including consent for genetic analyses, was obtained for all study participants.

Offspring subjects have been examined every 4 years since study onset, except for an 8-year interval between exams 1 and 2, with a standardized medical history and directed physical examination at each exam cycle and collection of an extensive array of diabetes-related quantitative traits and phenotypes (26). In this analysis, our principal diabetes-related quantitative traits come from exam 5 (1991–1994) in which data from a 75-g oral glucose tolerance test are available for all nondiabetic offspring. Diabetes-related quantitative traits include exam 5 fasting plasma glucose (FPG), glycated hemoglobin (A1C), fasting insulin, insulin resistance measured by homeostasis model assessment of insulin resistance (HOMA-IR) (27), Gutt's 0- to 120-min insulin sensitivity index (ISI_0–120) (28), the 28-year time-averaged FPG level obtained from exams 1–7 (mFPG), and incident categorical type 2 diabetes assessed over 28 years of follow-up. Laboratory methods for all quantitative traits have been described previously (26).

We used 2003 American Diabetes Association clinical criteria to define diabetes, in which a case was defined as new use of oral hypoglycemic or insulin therapy or a FPG ≥7.0 mmol/l at the index exam and a FPG ≥7.0 mmol/l on at least one prior exam (29). Age at onset of diabetes was assigned as the exam at which new diabetes therapy or the first FPG ≥7.0 mmol/l was recorded. Those presenting with diabetes at exam 1 underwent chart review to confirm diabetes type and age of onset (30). Among offspring with diabetes, >99% have type 2 diabetes (defined as age >35 years at diagnosis and not requiring continuous insulin therapy after diagnosis) (31). Including all seven exams, 91 offspring of the 1,087 studied here (8.4%) have developed diabetes.

As reported elsewhere (32), this sample size and analytical approach have 97% power to detect a variant that accounts for 2% of the variance in a quantitative trait and 63% power to detect a variant that accounts for 1% of the variance in such a trait.

Replication samples. Our replication efforts consisted of follow-up genotyping in a nonoverlapping and thus independent sample of unrelated FHS subjects and of in silico integration with other GWA datasets. The replication FHS sample consisted of 1,465 unrelated offspring participants derived from a previously plated set of DNAs in which only one individual from each pedigree was selected (Table 1); an additional sample of 251 offspring who had also been genotyped and analyzed in the 100K array served to check concordance rates between the 100K and follow-up genotyping platforms. The in silico replication effort focused initially on three other datasets from the 100K Type 2 Diabetes Consortium, all of which had been genotyped on the same Affymetrix 100K SNP marker set: a Pima Indian sample of 300 case subjects with type 2 diabetes whose age of onset was <25 years and 334 nondiabetic control subjects older than 45 years of age (including 172 sibships) (21), a Mexican-American sample of 287 case subjects and 316 control subjects from Starr County, Texas (22), and an Old Order Amish sample of 124 genetically enriched type 2 diabetes case subjects and 295 normal glucose-tolerant control subjects (23); reports describing these datasets are published alongside this paper, and the characteristics of each study are summarized in Supplementary Table 3, which is detailed in the online appendix (available at http://dx.doi.org/10.2337/db07-0451). In addition, we used the public resource of the Diabetes Genetics Initiative (DGI) available at http://www.broad.mit.edu/diabetes/ (March 2007 release), comprising 1,464 case subjects and 1,467 matched control subjects from Scandinavia and genotyped on the Affymetrix 500K array (17), for further replication of FHS 100K categorical type 2 diabetes and quantitative trait results.

Genotyping. FHS 100K SNP data are from the Affymetrix 100K SNP GeneChip marker set (116,204 SNPs) genotyped in the Genetics and Genomics Department at Boston University (33). Only genotypes called according to the dynamic modeling algorithm were available to us. We implemented the following quality control filters: SNPs located in autosomes only, genotyping call rate ≥90%, Hardy-Weinberg equilibrium ($P > 0.001$), and minor allele frequency (MAF) ≥10%. We chose this high allele frequency threshold on the basis of power considerations and the observation that SNPs with lower MAFs had a disproportionate amount of *P* values in the tail of the distribution. After quality control, 66,543 SNPs were available for analysis.

Follow-up genotyping was performed by allele-specific multiplex primer extension of PCR-amplified products with detection by matrix-assisted laser desorption ionization–time of flight mass spectroscopy using the Sequenom iPLEX platform (34). Genotyping call rates were 98.3%, and concordance between the Affymetrix and Sequenom platforms on 150 SNPs genotyped on 251 overlapping subjects reached 99.6%.

SNP prioritization in the FHS 100K scan. We pursued two approaches to prioritize SNPs potentially associated with type 2 diabetes or related traits. In the first, we simply ranked *P* values obtained from either generalized estimating equations (GEE) or family-based association test (FBAT) models for association with one or more of the six primary quantitative traits and selected SNPs with $P < 0.001$. In an alternative strategy, we selected SNPs associated with multiple related traits. We selected SNPs with consistent nominal associations ($P < 0.01$ in GEE or FBAT) with all three glucose traits (FPG, mFPG, and A1C) or all three insulin-related traits (fasting insulin, HOMA-IR, and ISI_0–120) or two glucose and two insulin traits. As expected, the correlation among glucose traits in FHS is high (Spearman correlation between mFPG and FPG is 0.83 and between mFPG and A1C is 0.39), as is the correlation among insulin-related traits (Spearman correlation between fast-

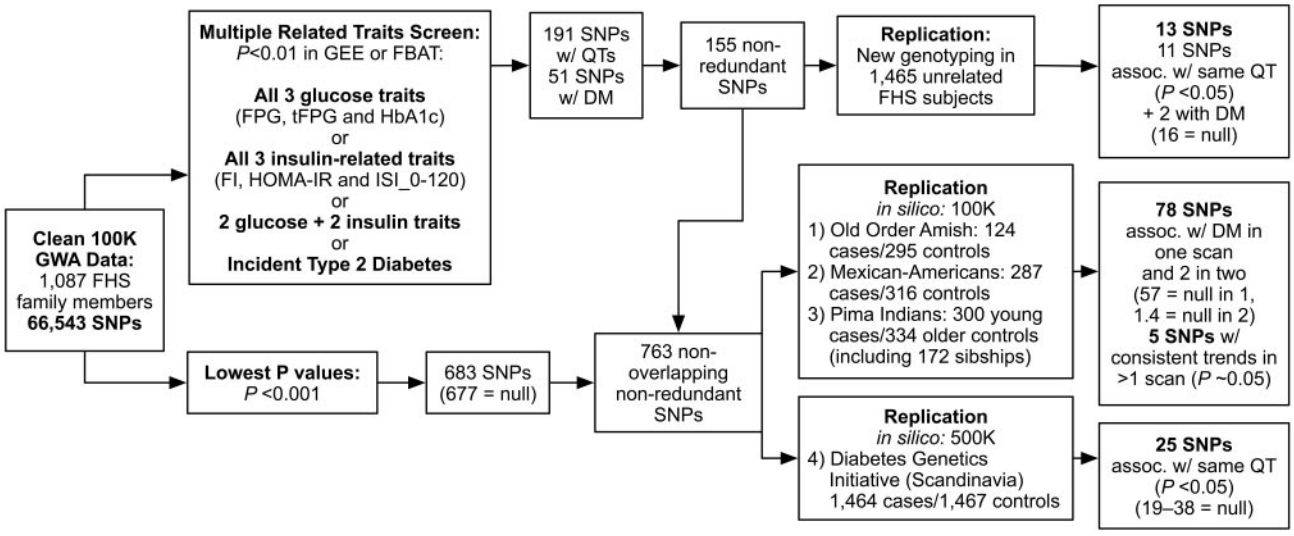

FIG. 1. Overview of SNP selection and replication strategies in a 100K GWA scan for diabetes and related traits in the FHS.

ing insulin and HOMA-IR is 0.94 and between HOMA-IR and ISI_0–120 is −0.54). The correlation between glucose and insulin traits ranges from 0.25 to 0.64 (data not shown). Among selected SNPs, we used extent of linkage disequilibrium to further choose a nonredundant set of SNPs for further replication: when strong linkage disequilibrium was detected ($r^2 \geq 0.8$), only one SNP was promoted to the replication stage, based on the highest genotyping call rate. Our overall strategy is presented in Fig. 1.

Statistical analysis. For quantitative traits, we used additive GEE and FBAT models to test associations between alleles and age-, age^2-, and sex-adjusted residual trait values. In subsidiary models, we also adjusted association results for BMI. The application of these methods to the FHS 100K dataset has been described in detail (32). GEE are a population-based test that takes into account familial correlation of the phenotype: it is prone to increased type 1 error for SNPs with low frequency and in the presence of population admixture, which is not a major concern in the FHS (A.K.M., J.D., L.A.C., unpublished observations). FBAT is a within-family test that controls for population admixture. The test looks for an association between the transmission of an allele and the quantitative trait, that is, it examines whether the transmission of one allele is associated with different levels of the quantitative trait; it is less powerful and more conservative than GEE.

For incident type 2 diabetes, we tested association using two complementary approaches that used longitudinal information on age at onset of diabetes or age through end of follow-up without diabetes. First, we used Cox proportional hazard survival analysis with robust covariance estimates to test SNPs against the hazard of new cases of diabetes over all seven exams, with time failure at the exam when diabetes was diagnosed, or disease-free censoring at last follow-up without diabetes (35). We used Cox models to estimate the hazard ratio (HR) and 95% CIs associated with the risk allele. Second, we created Martingale residuals from a sex-adjusted model in which high negative values indicated young diabetes age of onset and high positive values indicated older age without diabetes at follow-up, and we analyzed residuals using FBAT (36). To replicate 100K associations in the FHS unrelated replication sample, we used the same statistical methods, except that a general regression model was used to explore associations with quantitative traits, FBAT tests were not applied, and no robust covariance estimate was needed for the Cox survival analysis because the sample consists of unrelated participants.

Comparison with other datasets was restricted to a test of whether any SNPs selected from the FHS 100K array were associated with diabetes as a categorical trait in the second dataset at a nominal $P < 0.05$. For the 500K replication analysis, we also tested whether association of any of our selected SNPs with FPG or HOMA-IR were replicated in DGI at a nominal $P < 0.05$. This serial replication strategy yields equivalent power as the joint analysis when <1% of SNPs are promoted to the second stage (13).

To obtain the null expectation of the number of SNPs chosen for replication, we performed a constrained permutation test that both retained the correlation between the traits and attempted to maintain the trait heritability observed in our sample. We permuted the traits together by matching the rank of a phenotype derived from a principal components analysis of the six traits to the rank of a heritable simulated phenotype, thereby maintaining some of the correlation between individuals in the same family (37). The null distribution from 100 replications showed that the overall selection strategy would yield 152 "associated" nonredundant SNPs on average, with a SD of 16 if there were no true positive association to be found anywhere on the genome. The null expectation for each of the various steps in our analysis is shown in Fig. 1.

RESULTS

100K genotyping. Of 116,204 SNPs on the 100K Affymetrix fixed array, 66,543 SNPs passed quality control filters, including genotyping call rate, Hardy-Weinberg equilibrium, and MAF thresholds (Fig. 2*A*). We noted that

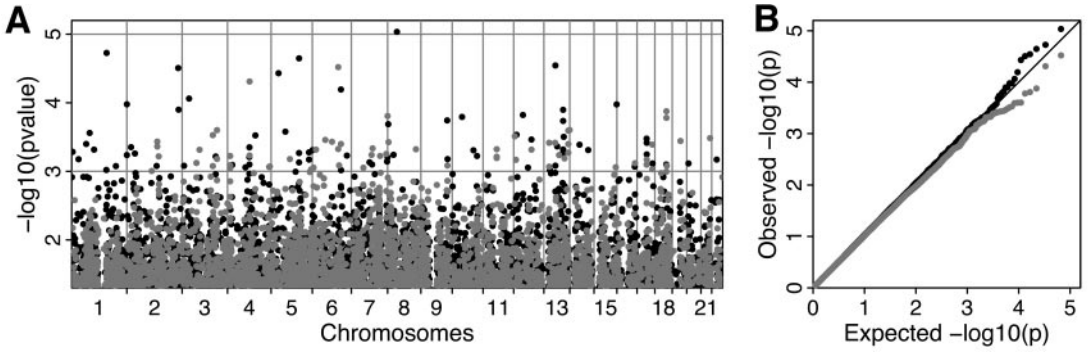

FIG. 2. Association of SNPs with exam 5 FPGs in the FHS. *A*: *P* values obtained with GEE (black) or FBAT (gray) across the autosomes, on a −log(10) scale. *B*: QQ plot of observed versus expected *P* values for both GEE (black) and FBAT (gray). There is an excess of significant *P* values at the end of the distribution for GEE, whereas FBAT yields more conservative results.

TABLE 2
Replication of 13 100K SNP associations within FHS

						100K FHS family set (n = 1,087)							FHS unrelated replication (n = 1,465)			
SNP	Gene	Location (hg18)	Strand	Alleles	MAF (allele)	DM HR (95% CI)*	Cox P (BMI)	FBAT P (BMI)	Trait	Allele	GEE P (BMI)	FBAT P (BMI)	DM HR (95% CI)*	Cox P	Allele	Regression P
rs625643	AK024684 AK122760	chr1:54409755	−	A/G	0.25 (A)	1.06 (0.73–1.50)	0.77 (0.86)	0.36 (0.34)	Fasting insulin	G	0.009 (0.08)	0.44 (0.32)	1.39 (1.0–1.80)	0.02	G	0.006
									FPG	G	0.004 (0.03)	0.03 (0.03)			G	0.73
									ISI_0–120	A	0.005 (0.02)	0.90 (0.66)			G	0.87
									HOMA-IR	G	0.0008 (0.01)	0.13 (0.06)			G	0.03
									mFPG	G	0.006 (0.03)	0.04 (0.07)			G	0.15
rs952635	*PDE4B*	chr1:66464473	−	A/G	0.31 (G)	0.56 (0.40–0.79)	0.0007 (0.003)	0.16 (0.12)	FPG	A	0.0004 (0.006)	0.02 (0.01)	0.89 (0.69–1.10)	0.34	A	0.03
									ISI_0–120	G	0.003 (0.02)	0.18 (0.08)			G	0.17
									A1C	A	0.003 (0.008)	0.30 (0.21)			A	0.75
									HOMA-IR	A	0.004 (0.051)	0.40 (0.20)			A	0.048
									mFPG	A	0.0003 (0.002)	0.03 (0.02)			A	0.36
rs7531174	*SLC44A3*	chr1:95110679	−	C/T	0.21 (C)	0.58 (0.42–0.80)	0.0009 (0.0002)	0.09 (0.07)	FPG	C	0.006 (0.005)	0.34 (0.41)	0.84 (0.65–1.10)	0.20	C	0.31
									A1C	C	0.0002 (0.0002)	0.005 (0.01)			C	0.02
									mFPG	C	0.0009 (0.001)	0.39 (0.47)			C	0.58
rs1489100	AJ012503	chr3:76215254	−	A/G	0.41 (G)	0.75 (0.54–1.0)	0.09 (0.18)	0.001 (0.002)	FPG	A	0.051 (0.09)	0.003 (0.004)	0.72 (0.57–0.91)	0.007	A	0.11
									A1C	A	0.0007 (0.001)	0.006 (0.01)			A	0.38
									mFPG	A	0.09 (0.12)	0.003 (0.005)			A	0.13
rs2715755	*MORC1*	chr3:110164908	−	C/T	0.39 (C)	1.22 (0.86–1.70)	0.27 (0.10)	0.01 (0.02)	Fasting insulin	T	0.29 (0.33)	0.003 (0.02)	0.98 (0.78–1.20)	0.88	T	0.70
									ISI_0–120	C	0.10 (0.10)	0.006 (0.04)			C	0.01
									HOMA-IR	T	0.20 (0.24)	0.003 (0.03)			T	0.60
rs729511	*SLC9A9*	chr3:144538397	+	A/G	0.44 (A)	1.43 (1.0–2.0)	0.03 (0.03)	0.008 (0.006)	Fasting insulin	G	0.002 (0.046)	0.31 (0.70)	1.12 (0.90–1.40)	0.32	G	0.02
									ISI_0–120	A	0.002 (0.01)	0.054 (0.058)			A	0.75
									HOMA-IR	G	0.004 (0.07)	0.14 (0.31)			G	0.02
rs2545523	Gene desert CNC	chr5:165446786	−	A/T	0.20 (T)	0.95 (0.64–1.40)	0.80 (0.66)	0.52 (0.33)	Fasting insulin	T	0.002 (0.006)	0.75 (0.77)	1.21 (0.94–1.60)	0.13	T	0.07
									ISI_0–120	A	0.01 (0.03)	0.002 (0.003)			A	0.07
									HOMA-IR	T	0.007 (0.02)	0.64 (0.89)			T	0.04
rs1355037	*ZPBP*	chr7:50063624	+	C/G	0.23 (C)	0.84 (0.61–1.20)	0.32 (0.58)	0.006 (0.045)	Fasting insulin	C	0.02 (0.33)	0.0003 (0.04)	0.94 (0.72–1.20)	0.67	C	0.61
									ISI_0–120	G	0.40 (0.95)	0.01 (0.09)			G	0.006
									HOMA-IR	C	0.04 (0.48)	0.001 (0.10)			C	0.86
rs1416406	(*SORCS1*) 120 kb	chr10:109034861	−	C/T	0.27 (T)	1.21 (0.88–1.70)	0.23 (0.31)	0.38 (0.62)	Fasting insulin	T	0.003 (0.01)	0.06 (0.29)	1.22 (0.97–1.50)	0.094	T	0.07
									ISI_0–120	C	0.006 (0.02)	0.005 (0.02)			C	0.14
									HOMA-IR	T	0.001 (0.005)	0.01 (0.07)			T	0.04
rs10500679	(*HNRNPGT*) 44 kb	chr11:7112939	−	C/G	0.28 (C)	0.82 (0.60–1.10)	0.20 (0.28)	0.66 (0.53)	Fasting insulin	C	0.002 (0.009)	0.57 (0.77)	1.25 (0.96–1.60)	0.10	C	0.83
									ISI_0–120	G	0.004 (0.004)	0.59 (0.75)			G	0.31
									A1C	G	0.25 (0.16)	0.001 (0.0004)			G	0.03
									HOMA-IR	C	0.0008 (0.003)	0.73 (0.97)			C	0.91
rs2806739	Gene desert CNC	chr13:53199780	+	C/T	0.22 (T)	1.31 (0.91–1.90)	0.14 (0.12)	0.045 (0.055)	FPG	T	0.006 (0.005)	0.004 (0.01)	1.51 (1.20–1.90)	0.001	T	0.26
									ISI_0–120	C	0.09 (0.07)	0.0006 (0.001)			T	0.77
									HOMA-IR	T	0.08 (0.04)	0.003 (0.009)			T	0.53
									mFPG	T	0.003 (0.0008)	0.03 (0.02)			T	0.27
rs2241119	*TSHR*	chr14:80628718	−	C/T	0.12 (C)	1.99 (1.10–3.50)	0.02 (0.01)	0.44 (0.21)	FPG	T	0.0005 (0.003)	0.09 (0.10)	1.02 (0.73–1.40)	0.92	T	0.27
									A1C	T	0.001 (0.004)	0.13 (0.08)			C	0.77
									HOMA-IR	T	0.0008 (0.003)	0.02 (0.009)			T	0.046
									mFPG	T	0.00004 (0.0003)	0.09 (0.09)			T	0.07
rs2009833	*ATP8B4*	chr15:47980845	+	A/G	0.35 (A)	0.77 (0.56–1.10)	0.10 (0.18)	0.02 (0.03)	FPG	A	0.06 (0.04)	0.005 (0.02)	0.79 (0.63–0.99)	0.037	A	0.001
									A1C	A	0.02 (0.006)	0.004 (0.004)			A	0.001
									HOMA-IR	A	0.07 (0.02)	0.006 (0.003)			A	0.04
									mFPG	A	0.08 (0.03)	0.005 (0.003)			A	0.002

Association of SNPs selected by the multiple related trait strategy with diabetes and/or quantitative traits in the FHS 100K family sample and replication in the nonoverlapping FHS unrelated sample. If the SNP lies within a gene, the gene (italics) or mRNA (no italics) is indicated; if not, the nearest gene is shown in parentheses, with distance in kilobase pairs below the gene name. SNPs located >400 kb from a known gene or mRNA are considered to be in a gene desert; those labeled "CNC" reflect proximity to a conserved noncoding region. The strand on which the SNP was genotyped is stated. Alleles are shown in alphabetical order, with the minor allele in parentheses next to the MAF in FHS, and the HR (*) calculated as second allele versus first allele. The allele associated with the higher level of each trait is shown to the right of the trait. The P values in parentheses indicate adjustment for BMI. Chr, chromosome.

TABLE 3
Mean diabetes-related quantitative trait levels by genotype for eleven 100K SNPs with significant associations in the FHS family sample that replicated the same quantitative trait in the nonoverlapping FHS unrelated sample

	FHS 100K family sample								FHS unrelated replication sample					
	Subjects with diabetes (%)					Mean trait values			Subjects with diabetes (%)			Mean trait values		
SNP	1/2	11	12	22	Trait	11	12	22	11	12	22	11	12	22
rs625643	A/G	7	7.9	8.9	Fasting insulin (μU/ml)	30.6	28.1	31.4	8.6	8.6	12.7	27.1	30.5	31.5
					HOMA-IR	7.4	6.87	8.38				7.1	7.93	8.36
rs952635	A/G	10.9	7.9	1	FPG (mg/dl)	100	99	93	11.6	10.1	10.2	103	101	98
					HOMA-IR	8.4	7.4	6.7				8.3	8	7.7
rs7531174	C/T	13.3	12.7	6	A1C (%)	5.75	5.4	5.18	17.7	10.9	10.4	5.82	5.56	5.48
rs2715755	C/T	7.6	7.1	10.3	ISI_0–120	27.2	26.1	25.7	10.5	11.2	10.5	26.2	25.5	24.9
rs729511	A/G	6.9	7.5	11.4	Fasting insulin (μU/ml)	27.5	30.8	31	9.6	10.9	11.5	28.8	31	31.7
					HOMA-IR	7.02	7.87	8.15				7.21	8.24	8.39
rs2545523	A/T	9.1	7.2	12.1	HOMA-IR	7.7	7.9	7.5	10	12.7	11.8	8	8.1	8.9
rs1355037	C/G	9.4	9.9	7.6	ISI_0–120	26.3	25.9	26.3	12.3	10.7	10.8	22.6	25.1	25.8
rs1416406	C/T	7.5	8.6	9.4	HOMA-IR	7.1	8.1	9.7	9.1	13.1	10.6	7.8	8.5	7.9
rs10500679	C/G	6.3	10.5	6.8	A1C (%)	5.08	5.32	5.26	6.9	10.6	11.8	5.28	5.5	5.56
rs2241119	C/T	0	5.1	9.4	HOMA-IR	6.6	6.7	8	5.9	11.7	10.7	7.1	7.9	8.2
rs2009833	A/G	9.7	9.8	6.4	FPG (mg/dl)	101	100	97	14.6	11.4	9.5	108	102	99
					A1C (%)	5.63	5.28	5.19				5.72	5.55	5.42
					HOMA-IR	8.6	7.9	7.3				9.1	8	8
					mFPG (mg/dl)	100	98	97				104	101	99

SNPs significantly associated with quantitative traits in the FHS family 100K sample were tested for replication in a nonoverlapping unrelated FHS sample. Mean measurements for each trait are presented by genotype, with alleles shown in alphabetical order as "1/2".

the GEE P value distribution deviated from the null expectation for any single quantitative trait: up to 28% more P values were <0.001 than expected if no SNPs were associated. The deviation was more extreme for smaller significance levels. Nevertheless, this deviation did not change significantly when analyses were restricted to increasingly stringent call rate cutoffs, suggesting that it was independent from call rate and not due to nonrandom missing data. Such deviation was not present for the FBAT analyses (Fig. 2*B*).

SNP selection. The "pure P value" strategy yielded 683 SNPs associated with any of six primary quantitative traits or diabetes in either GEE or FBAT at $P < 0.001$. No result achieved conventional genome-wide significance ($P \sim 5 \times 10^{-8}$) (14,15). The "multiple related trait" strategy yielded 191 SNPs, 51 of which also showed $P < 0.01$ for incident diabetes, and 111 of which had $P < 0.001$ for at least one trait (thus overlapping with the first set). We used linkage disequilibrium between SNPs (pairwise $r^2 > 0.8$) to select a nonredundant subset of 155 SNPs for further replication (of which 41 also showed $P < 0.01$ for incident diabetes and 85 had $P < 0.001$ for at least one trait). The probability of selecting 155 or more nonredundant SNPs if there were no true association to be detected anywhere on the genome was estimated to be 50% by permutation. Hence, the number of SNPs chosen by our selection strategy does not differ substantially from the expectation of 152 SNPs expected under the null hypothesis (Fig. 1). The combination of these two approaches yielded 763 unique SNPs with evidence for association with diabetes or related traits (Supplementary Table 1).

Follow-up genotyping. We successfully genotyped 150 (148 nonredundant) of the 155 SNPs obtained from the multiple related trait strategy in a nonoverlapping replication sample of 1,465 FHS unrelated subjects. Eleven SNPs were associated with at least one of the same traits (at nominal $P < 0.05$) in the replication dataset (10 expected under the null); 4 of them (rs2009833, rs625643, rs729511, and rs952635) were associated with more than one trait, and 2 of these (rs2009833 and rs625643) were also associated with incident diabetes. Four SNPs (the aforementioned SNPs, rs2009833 and rs625643, plus rs1489100 and rs2806739) showed association with diabetes incidence in replication (six expected under the null); rs1489100 had also been associated with diabetes incidence initially. These 13 SNPs are presented in Table 2, and for the 11 SNPs with nominally significant replication of quantitative trait associations, the distribution of quantitative trait levels and proportion of subjects with diabetes by SNP genotype is presented in Table 3.

In silico replication. Our 100K Type 2 Diabetes Consortium collaborators tested all 763 FHS-associated SNPs for association with type 2 diabetes in their respective datasets. Of the 13 SNPs obtained from the multiple related trait strategy and replicated in the follow-up FHS unrelated sample (Table 2), none showed a nominal P value <0.05 consistent with the expected direction of effect (Table 4). Six of these 13 SNPs were also present in the Affymetrix 500K array used by the DGI, 2 of them had perfect proxies (pairwise $r^2 = 1.0$), and an adequate proxy ($r^2 \geq 0.6$) could be obtained for an additional 4 SNPs based on Phase II HapMap CEU genotypes; none of them showed association with type 2 diabetes in the DGI, although 2 of them (rs6664618 and rs17281232, see below) did show a suggestive association with insulin resistance, and rs6664618 also had a nominal association with FPG (Table 4).

These data do not offer consistent evidence for association with any one SNP across all datasets. For instance, nominal P values for the association of the minor C allele at rs10500679 with higher insulin resistance measures and lower A1C in FHS are mutually inconsistent, as is the association of its major G allele with diabetes in the Mexican-American dataset, whereas the minor T allele of SNP rs17281232 (which is in strong linkage disequilibrium with rs10500679 in Europeans, $r^2 = 0.92$) is associated

TABLE 4
Attempt at replication of 13 FHS genetic associations in external GWA datasets

SNP	Mexican Americans OR	Mexican Americans *P*	Pima Indian case-control OR	Pima Indian case-control *P*	Pima Indian sibs OR	Pima Indian sibs *P*	Amish OR	Amish *P*	DGI SNP	r^2	DGI OR (95% CI)	DGI OR *P* value	DGI QT	DGI QT *P* value
rs625643	0.70	0.32	1.51	0.21	0.33	0.18	1.11	0.36	Same	1.0	1.12 (0.97–1.29)	0.12	FPG	0.17
											T vs. C		HOMA-IR	0.21
rs952635	0.84	0.54	1.12	0.37	1.16	0.46	0.89	0.33	rs6664618	0.60	1.03 (0.91–1.18)	0.60	FPG	**0.04 (G)**
														0.057
											T vs. G		HOMA-IR	**(G)**
rs7531174	0.85	0.63	0.94	0.62	1.00	0.98	1.00	0.99	rs12565150	0.79	0.98 (0.84–1.14)	0.75	FPG	0.57
											A vs. T		HOMA-IR	0.86
rs1489100	0.89	0.71	0.80	0.09	1.04	0.87	1.13	0.28	rs7620001	1.0	0.92 (0.81–1.04)	0.53	FPG	0.55
											T vs. G		HOMA-IR	0.67
rs2715755	1.41	0.34	Low MAF		Low MAF		1.23	0.08	Same	1.0	1.0 (0.88–1.14)	0.91	HOMA-IR	0.49
											G vs. A			
rs729511	0.89	0.67	1.24	0.09	0.76	0.16	1.01	0.96	Same	1.0	0.96 (0.85–1.09)	0.79	HOMA-IR	0.47
											A vs. G			
rs2545523	0.99	1.00	1.07	0.59	1.10	0.63	0.96	0.76	Same	1.0	1.02 (0.88–1.18)	0.91	HOMA-IR	0.86
											A vs. T			
rs1355037	0.99	1.00	0.99	0.96	0.82	0.33	0.83	0.21	Same	1.0	1.05 (0.91–1.21)	0.82	HOMA-IR	0.55
											C vs. G			
rs1416406	0.82	0.52	0.99	0.92	0.92	0.65	1.07	0.59	rs10787019	1.0	0.95 (0.83–1.09)	0.40	HOMA-IR	0.37
											T vs. G			
rs10500679	2.19	0.02	Low MAF		Low MAF		1.02	0.87	rs17281232	0.92	1.01 (0.86–1.19)	0.98	FPG	0.68
														0.004
											T vs. C		HOMA-IR	**(T)**
rs2806739	1.06	0.89	0.75	0.052	0.86	0.49	1.22	0.11	rs4242932	0.87	0.91 (0.78–1.06)	0.08	FPG	0.93
											T vs. G		HOMA-IR	0.63
rs2241119	1.10	0.85	0.80	0.18	0.91	0.71	1.02	0.90	Same	1.0	1.18 (1.0–1.41)	0.19	FPG	0.52
											G vs. A		HOMA-IR	0.75
rs2009833	0.67	0.16	0.88	0.52	1.40	0.32	0.90	0.31	rs8023809	0.39	1.07 (0.92–1.24)	0.69	FPG	0.74
											A vs. G		HOMA-IR	0.37

The 100K FHS SNPs that showed replication in a nonoverlapping unrelated FHS cohort (Table 2) were examined for association with diabetes in three other 100K datasets, and for replication of their association with diabetes, FPG or HOMA-IR in the 500K DGI dataset; please see their respective publications for a description the statistical methods of each study. DGI data were obtained from http://www.broad.mit.edu/diabetes/ (March 2007). SNPs that did not meet MAF thresholds in the Pima Indian case-control or sib datasets were not analyzed. If the FHS SNP was not present in the 500K DGI array, the SNP in strongest linkage disequilibrium (as measured by r^2 using data from phase 2 of the HapMap in the CEU population) was examined. In the DGI, the associated alleles are indicated below the OR or, for relevant quantitative traits (QT), next to the corresponding *P* value. Nominally significant associations are shown in boldface.

with insulin resistance by HOMA-IR in the 500K DGI scan ($P = 0.004$). In an analogous manner, the nominal *P* value of 0.052 obtained for rs2806739 in the Pima Indian case-control dataset indicates that the T allele would be protective for diabetes (odds ratio [OR] 0.75), whereas this same allele is associated with higher FPG in FHS in the initial 100K dataset and with higher incidence of diabetes on replication (HR 1.51 [95% CI 1.2–1.9], Cox $P = 0.001$). Taken together, these conflicting nominal results caution that the suggestive associations found here could be statistical fluctuations rather than indicating true genetic risk for diabetes.

Several of the 13 SNPs showed some consistent trends in the replication samples, albeit not nominally significant. The G allele of rs1489100 was associated with protection from diabetes both in the initial FHS 100K scan (HR 0.75 [95% CI 0.54–1.0], Cox $P = 0.089$, FBAT $P = 0.001$) and in the FHS replication sample (0.72 [0.57–0.91], Cox $P = 0.007$); consistent with this effect, the G allele was associated with lower glucose levels (as measured by all three glucose-related traits) in the initial scan. The association of rs1489100 with diabetes trended in the same direction in the Pima Indian case-control dataset (OR 0.80, $P = 0.09$). However, a 500K array SNP in perfect linkage disequilibrium with rs1489100 in Europeans (rs7620001, $r^2 = 1.0$) was not associated with diabetes (OR 0.92 [95% CI 0.81–1.04], $P = 0.53$), FPG ($P = 0.55$), or HOMA-IR ($P = 0.67$) in the DGI dataset (Table 4).

The G allele at rs729511 was associated with diabetes incidence in the initial 100K FHS scan (HR 1.43 [95% CI 1.0–2.0], Cox $P = 0.03$, FBAT $P = 0.008$) and with insulin resistance as measured by all three insulin traits ($P = 0.002$–0.004); fasting insulin and HOMA-IR also showed nominal association in the FHS replication sample ($P = 0.02$ for both). The direction of effect for this SNP was consistent in the Pima Indian case-control sample (OR 1.24, $P = 0.09$), but there was no association with diabetes (OR 0.96 for the A allele [95% CI 0.85–1.09], $P = 0.79$) or HOMA-IR ($P = 0.47$) for the same SNP in the DGI (Table 4).

The minor G allele at rs952635 was associated with lower diabetes incidence in the initial 100K FHS dataset (HR 0.56 [95% CI 0.40–0.79], Cox $P = 0.0007$); it was also associated with lower glucose levels and greater insulin

sensitivity in both the initial and follow-up FHS genotyping. Interestingly, the DGI 500K SNP which had the strongest linkage disequilibrium with rs952635 in Europeans (rs6664618, $r^2 = 0.60$) showed nominally significant lower FPG ($P < 0.04$) and a trend toward greater insulin sensitivity for the tagging allele ($P = 0.057$).

Of all 763 FHS-associated SNPs, 78 showed nominal association with type 2 diabetes in one other dataset (57 expected under the null), and 2 (rs2863389 and rs7935082) showed nominal association in more than one (1.4 expected under the null); all results are presented in Supplementary Table 1. The T allele at SNP rs2863389 was protective against diabetes (HR 0.41 [95% CI 0.25–0.69], Cox $P = 0.0006$), whereas the alternate C allele was associated with higher FPG and mFPG in the FHS sample ($P = 0.005$ and 0.0005, respectively); the T allele also showed consistent protection from type 2 diabetes in Mexican Americans (OR 0.43, nominal $P = 0.03$) and in the Amish (OR 0.71, nominal $P = 0.04$) with similar trends in the Pima Indians. At SNP rs7935082, the C allele was associated with higher FPG in FHS (FBAT $P = 0.0006$), whereas the alternate T allele was nominally protective from diabetes in the Mexican Americans (OR 0.53, $P = 0.049$) and in the Pima Indians (OR 0.58, $P = 0.009$).

Of the others, three SNPs revealed suggestive trends: Two SNPs in perfect linkage disequilibrium with each other (rs2378199 and rs6059961, $r^2 = 1.0$) showed nominally significant association with type 2 diabetes in both tests of association used by the Pima Indian investigators for their overlapping (i.e., nonindependent) case-control and sibship samples; this association followed the same direction as that seen in FHS and was consistent with the expected changes in quantitative traits resulting from altered glycemic pathophysiology. Similarly, one other SNP (rs6058115) that was associated with all three insulin traits in FHS showed nominal association with diabetes in the Pima Indian sibs ($P = 0.042$) and neared nominal association in the overlapping Pima Indian case-control sample ($P = 0.054$).

We further examined our 763 SNPs for replication in the public 500K DGI resource. Of the 763 SNPs, 206 (27.0%) were present in both Affymetrix genotyping arrays, an adequate proxy ($r^2 \geq 0.6$) could be found for 443 SNPs (58.1%), and only 26 SNPs (3.4%) could not be captured at all (Supplementary Table 2). Five SNPs (or their proxies) were also nominally associated with type 2 diabetes in the DGI, 8 SNPs with FPG, and 12 SNPs with HOMA-IR, all with consistent direction of effects (Table 5). In all of these analyses, adjustment of the associations with diabetes-related traits for BMI attenuated the associations in some instances and strengthened them in others (Tables 2 and 5).

Positive controls. The 100K FHS resource also serves as a resource in which to pursue phenotypic characterization and further validation of putative diabetes risk SNPs reported in other datasets. We therefore sought to replicate the widely reproduced *TCF7L2* association, and the top findings reported in five recent high-density GWA scans (16–20). The 100K SNP rs7100927 was in moderate linkage disequilibrium ($r^2 = 0.50$) with the diabetes-associated *TCF7L2* SNP rs7903146 and was associated with risk of diabetes (HR 1.56 [95% CI 1.1–2.1], Cox $P = 0.007$) and with mFPG (GEE $P = 0.03$) in the FHS 100K dataset. We confirmed this association by directly genotyping rs7903146 in both the family and unrelated samples, obtaining association with diabetes incidence (1.28 [1.08–1.52], Cox $P = 0.005$). Interestingly, the risk T allele at rs7903146 was directly associated with mFPG and inversely associated with insulin sensitivity adjusted for β-cell function as measured by the ISI_0–120 (nominal GEE $P = 0.03$ for both), an effect that persisted after adjustment for BMI. The *TCF7L2* 100K SNP rs7100927 is also in strong linkage disequilibrium ($r^2 = 0.93$ in HapMap CEU) with SNPs rs7924080 and rs10885406, which tag a putative obesity-associated haplotype (HapA) in Caucasians; we found no statistically significant association between rs7100927 and BMI.

Among other SNPs reported to be highly associated with diabetes in the recently published GWA scan, we noted moderate linkage disequilibrium with SNPs present in our 100K array (Table 6). The two *HHEX* SNPs were in moderate linkage disequilibrium with 100K FHS SNP rs10509645 ($r^2 = 0.57$ and 0.70, respectively), but we found no nominal associations with diabetes incidence or related traits in the FHS. Similarly, the *CDKAL1* SNP rs7754840 was in weak linkage disequilibrium with 100K FHS SNP rs2328545 ($r^2 = 0.35$), and no nominal associations with diabetes incidence or related traits were found in the FHS. On the other hand, the FHS SNP rs1995222 was in weak linkage disequilibrium with the original SNP in *SLC30A8* ($r^2 = 0.20$), and yet it showed nominal associations with diabetes incidence (FBAT $P = 0.01$), FPG (FBAT $P = 0.006$), and mFPG (FBAT $P = 0.008$); the FHS SNP rs10501278 weakly tagged LOC387761 SNP rs7480010 ($r^2 = 0.28$) and showed nominal associations with fasting insulin (FBAT $P = 0.008$), HOMA-IR (FBAT $P = 0.01$), and ISI 01–20 (FBAT $P = 0.047$); the risk alleles at the three *EXT2* SNPs (rs1113132, rs11037909, and rs3740878) were captured by the T allele of rs962848 in the 100K array ($r^2 = 0.47$), which was associated with higher FPG and mFPG (FBAT $P = 0.002$ and 0.007, respectively) and lower insulin sensitivity (nominal $P = 0.049$) in FHS; and the FHS SNP rs10513800 showed modest linkage disequilibrium with two *IGF2BP2* SNPs ($r^2 = 0.33$), and was nominally associated with mFPG in FHS (GEE $P = 0.03$).

DISCUSSION

We present initial associations with type 2 diabetes and related quantitative traits using the FHS 100K GWA resource, with replication and integration of initial associations within FHS and in silico with external GWA datasets. We did not find any single variant to be associated with diabetes or related traits in the FHS 100K sample and all replication samples, but we found a number of consistent associations worthy of follow-up. We were also able to replicate association with the confirmed diabetes risk SNP in *TCF7L2* and with SNPs recently identified in high-density GWA scans (16–20). These results demonstrate the contribution that a community-based sample rich with diabetes-related quantitative trait data can make to type 2 diabetes gene discovery.

GWA scans provide a powerful tool with which to query the genome for common variants that confer modest effects on polygenic traits (6). Because of the many statistical tests involved and the high likelihood of obtaining a large number of false-positive results, it is crucial to perform rigorous genotyping quality control and set stringent statistical thresholds. Thus, unless risk variants are very common and/or have a relatively large effect on the trait under study, true results can only be detected with large sample sizes. In instances in which sample size is limiting, a replication strategy with other similarly con-

TABLE 5
SNPs associated with type 2 diabetes and/or related traits in FHS that show nominally significant associations in the DGI dataset

				FHS						DGI							
Chr	SNP	Gene	Position	Alleles (MAF)	HR *P* (BMI)	HR (95% CI)	Trait	QT GEE *P* (BMI)	QT FBAT *P* (BMI)	SNP on 500K	r^2	m	Type 2 diabetes *P* value	OR (95% CI)	Trait	*P*	β
1	rs952635	*PDE4B*	64982242	A/**G**	0.0007	0.56	FPG	0.0004 (0.006)	0.02 (0.01)	rs6664618	0.60	T	0.60	1.03	FPG	0.04	−0.08
				0.31	0.003	(0.40–0.79)	HOMA-IR	0.004 (0.051)	0.40 (0.20)					(0.91–1.18)	HOMA-IR	0.057	−0.07
1	rs963328	*FLVCR*	186263657	A/**C**	0.37	0.87	FPG	0.38 (0.60)	0.25 (0.78)	Same	1.0	G	0.79	0.99	FPG	0.64	−0.02
				0.48	0.55	(0.64–1.20)	HOMA-IR	0.30 (0.66)	0.0003 (0.02)					(0.87–1.12)	HOMA-IR	0.03	−0.09
2	rs385909	LOC400937	256807	C/**T**	0.11	0.71	FPG	0.0001 (0.002)	0.07 (0.32)	rs389621	1.0	T	0.97	0.98	FPG	0.35	−0.05
				0.13	0.27	(0.47–1.10)	HOMA-IR	0.00006 (0.0002)	0.03 (0.20)					(0.82–1.16)	HOMA-IR	0.04	−0.11
2	rs300703	*SH3YL1*	302047	**C**/T	0.08	1.54	FPG	0.0006 (0.008)	0.12 (0.24)	rs4241316	1.0	C	0.96	0.96	FPG	0.35	−0.05
				0.11	0.16	(0.94–2.50)	HOMA-IR	0.00003 (0.0006)	0.11 (0.20)					(0.80–1.15)	HOMA-IR	0.046	−0.11
2	rs10497719	(*MYO1B*)	185972926	C/**T**	0.00003	2.03	FPG	0.04 (0.10)	0.07 (0.07)	rs17351803	1.0	C	0.16	1.09	FPG	0.01	0.15
				0.11	0.004	(1.50–2.80)	HOMA-IR	0.13 (0.18)	0.22 (0.27)					(0.91–1.32)	HOMA-IR	0.91	0.01
5	rs10491394	*EFNA5*	102784623	A/**G**	0.02	0.58	FPG	0.001 (0.008)	0.005 (0.007)	Same	1.0	C	0.47	0.94	FPG	0.01	−0.12
				0.18	0.06	(0.37–0.90)	HOMA-IR	0.003 (0.02)	0.02 (0.01)					(0.80–1.11)	HOMA-IR	0.86	−0.01
5	rs1918159	(BC045192)	112977574	**C**/G	0.002	0.56	FPG	0.002 (0.003)	0.01 (0.02)	rs17142889	1.0	T	0.63	1.07	FPG	0.04	0.12
				0.14	0.007	(0.39–0.81)	HOMA-IR	0.009 (0.03)	0.04 (0.04)					(0.89–1.30)	HOMA-IR	0.32	0.06
5	rs1990930	*SIL1*	134614609	A/**G**	0.03	0.71	FPG	0.009 (0.02)	0.03 (0.03)	rs11744695	1.0	G	0.03	0.85	FPG	0.16	−0.05
				0.46	0.055	(0.52–0.98)	HOMA-IR	0.02 (0.12)	0.07 (0.0505)					(0.75–0.97)	HOMA-IR	0.25	−0.04
6	rs10498861	BC031312	69343049	**A**/G	0.09	1.52	FPG	0.005 (0.07)	0.03 (0.13)	rs6939862	0.93	T	0.56	0.96	FPG	0.04	−0.12
				0.12	0.14	(0.93–2.50)	HOMA-IR	0.0002 (0.04)	0.049 (0.24)					(0.79–1.16)	HOMA-IR	0.56	−0.03
6	rs6910169	(BC036223)	113625254	**C**/T	0.90	0.98	FPG	0.006 (0.04)	0.00003 (0.0002)	rs1299158	1.0	C	0.48	0.96	FPG	0.30	0.04
				0.43	0.85	(0.74–1.30)	HOMA-IR	0.06 (0.35)	0.008 (0.02)					(0.84–1.09)	HOMA-IR	0.04	−0.08
7	rs2024265	*NOM1*	155987004	**A**/G	0.047	1.43	FPG	0.0007 (0.03)	0.13 (0.38)	rs10256184	1.0	C	0.007	0.81	FPG	0.11	−0.07
				0.22	0.14	(1.0–2.0)	HOMA-IR	0.16 (0.82)	0.38 (0.87)					(0.70–0.93)	HOMA-IR	0.56	−0.03
10	rs2025463	*SFMBT2*	7199259	**C**/T	0.02	0.66	FPG	0.02 (0.03)	0.15 (0.12)	rs11255049	0.84	T	0.04	1.15	FPG	0.30	−0.05
				0.29	0.01	(0.46–0.94)	HOMA-IR	0.03 (0.06)	0.053 (0.03)					(1.0–1.33)	HOMA-IR	0.28	−0.05
10	rs7089102	(BC017976)	51109048	**G**/T	0.008	0.68	FPG	0.008 (0.01)	0.86 (0.93)	rs7082607	1.0	G	0.84	1.01	FPG	0.45	−0.03
				0.47	0.006	(0.51–0.90)	HOMA-IR	0.001 (0.02)	0.56 (0.93)					(0.89–1.14)	HOMA-IR	0.02	0.09
10	rs1591565	(AK056904)	79197636	C/**G**	0.86	0.97	FPG	0.33 (0.35)	0.01 (0.003)	Same	1.0	G	0.33	1.06	FPG	0.62	0.02
				0.33	0.86	(0.72–1.30)	HOMA-IR	0.02 (0.02)	0.001 (0.0004)					(0.93–1.21)	HOMA-IR	0.001	0.13
10	rs10509927	(*SMC3*)	106203638	A/**G**	0.054	1.5	FPG	0.03 (0.04)	0.37 (0.61)	**rs4918573**	0.70	T	0.61	1.05	FPG	0.22	0.06
				0.13	0.059	(1.0–2.30)	HOMA-IR	0.002 (0.003)	0.13 (0.64)					(0.89–1.23)	HOMA-IR	0.05	0.10
10	rs10509928	(*SMC3*)	106203679	G/**T**	0.03	1.56	FPG	0.02 (0.03)	0.35 (0.62)	**rs4918573**	0.70	T	0.61	1.05	FPG	0.22	0.06
				0.13	0.04	(1.0–2.30)	HOMA-IR	0.001 (0.002)	0.07 (0.50)					(0.89–1.23)	HOMA-IR	0.05	0.10
10	rs845080	*CR607950*	118930209	**A**/G	0.50	0.9	FPG	0.08 (0.09)	0.009 (0.02)	rs705146	0.68	T	0.84	1.05	FPG	0.94	0
				0.39	0.79	(0.65–1.20)	HOMA-IR	0.11 (0.20)	0.009 (0.02)					(0.93–1.19)	HOMA-IR	0.03	0.08
11	rs10500679	*HNRNPG-T*	7275478	**C**/G	0.20	0.82	FPG	0.05 (0.08)	0.71 (0.65)	rs17281232	0.92	T	0.98	1.01	FPG	0.68	0.02
				0.28	0.28	(0.60–1.10)	HOMA-IR	0.0008 (0.003)	0.73 (0.97)					(0.86–1.19)	HOMA-IR	0.004	0.14
11	rs1151488	*MAP3K11*	62698321	**C**/T	0.00005	0.54	FPG	0.02 (0.004)	0.94 (0.61)	Same	1.0	C	0.12	1.09	FPG	0.94	0
				0.32	0.00002	(0.40–0.73)	HOMA-IR	0.17 (0.06)	0.42 (0.11)					(0.95–1.25)	HOMA-IR	0.03	0.09
11	rs1387153	LOC647483	89978446	C/**T**	0.16	1.28	FPG	0.004 (0.01)	0.13 (0.17)	rs3847554	0.59	T	0.54	1.04	FPG	0.04	0.08
				0.28	0.21	(0.90–1.80)	HOMA-IR	0.07 (0.23)	0.59 (0.91)					(0.92–1.18)	HOMA-IR	0.97	0
12	rs1730454	LOC643063	46532538	A/**G**	0.047	0.54	FPG	0.01 (0.02)	0.38 (0.56)	rs2078033	0.95	A	0.95	1.02	FPG	0.05	−0.11
				0.12	0.046	(0.29–1.0)	HOMA-IR	0.002 (0.0005)	0.09 (0.10)					(0.85–1.22)	HOMA-IR	0.18	−0.07
12	rs10506806	*SYT1*	79121551	C/**T**	0.02	0.66	FPG	0.0003 (0.0009)	0.0005 (0.0003)	Same	1.0	T	0.38	1.03	FPG	0.02	−0.10
				0.30	0.06	(0.46–0.94)	HOMA-IR	0.03 (0.06)	0.003 (0.002)					(0.90–1.18)	HOMA-IR	0.35	−0.04
13	rs478243	*LATS2*	2640558	**C**/T	0.08	0.77	FPG	0.01 (0.01)	0.11 (0.11)	Same	1.0	C	0.57	1.05	FPG	0.86	−0.01
				0.34	0.049	(0.58–1.0)	HOMA-IR	0.01 (0.02)	0.003 (0.007)					(0.91–1.21)	HOMA-IR	0.01	0.10
20	rs6074234	LOC339593	11098703	C/**T**	0.0009	2.12	FPG	0.08 (0.18)	0.91 (0.43)	**rs6131145**	1.0	T	0.01	1.26	FPG	0.31	0.06
				0.11	0.0008	(1.40–3.30)	HOMA-IR	0.01 (0.06)	0.59 (0.12)					(1.05–1.51)	HOMA-IR	0.48	0.04
20	rs4142363	LOC339593	11099044	C/**T**	0.0009	2.05	FPG	0.08 (0.15)	0.87 (0.45)	**rs6131145**	1.0	T	0.01	1.26	FPG	0.31	0.06
				0.11	0.0009	(1.30–3.10)	HOMA-IR	0.02 (0.06)	0.49 (0.12)					(1.05–1.51)	HOMA-IR	0.48	0.04

Of 763 SNPs selected for presumptive association with diabetes and/or related traits in the FHS 100K dataset, 25 SNPs examined by the DGI investigators (either the same SNPs or SNPs from the 500K Affymetrix array in linkage disequilibrium with the FHS 100K SNPs) showed consistent nominal associations with the same quantitative traits in the DGI. The degree of linkage disequilibrium is indicated by r^2; DGI SNPs in bold served as proxies for more than one 100K SNP. In the FHS dataset, alleles are shown in alphabetical order with the minor allele in bold and the MAF below; HRs with 95% CI are estimated as second allele versus first allele. In FHS, quantitative trait (QT) *P* values are shown without and with adjustment for BMI (in parentheses). In the DGI, ORs are estimated as the minor allele versus the major allele: the minor allele (m) is shown. In the DGI, quantitative traits are estimated for the minor allele, with the direction of the effect indicated by the positive or negative sign of the β estimate. When alleles for the same SNP are not identical between both datasets, they were genotyped on opposite strands. The genes (italicized) or mRNAs (not italicized) in which SNPs lie are shown; otherwise, the nearest gene or mRNA is indicated in parentheses. DGI data were obtained from http://www.broad.mit.edu/diabetes/ (March 2007). Chr, chromosome.

TABLE 6
Association of confirmed SNPs from previous high-density GWA scans in FHS

Chr	Gene	Original SNP	100K SNP	r^2	FHS MAF	Alleles	HR (95% CI)	Cox *P* value	FBAT *P* value	Trait	GEE Allele	GEE *P*	FBAT Allele	FBAT *P*
10	*TCF7L2*	rs7903146	rs7100927	0.50	0.49	A/G	1.56 (1.1–2.1)	**0.007**	**0.04**	FPG	G	0.08	G	0.60
										mFPG	G	**0.03**	G	0.13
										A1C	G	0.67	A	0.15
										Fasting insulin	A	0.94	A	0.30
										HOMA-IR	G	0.43	A	0.65
										ISI_0–120	A	0.12	A	0.59
8	*SLC30A8*	rs13266634	rs1995222	0.197	0.44	A/G	0.96 (0.69–1.3)	0.80	**0.01**	FPG	A	0.43	A	**0.006**
										mFPG	A	0.46	A	**0.008**
										A1C	A	0.82	A	0.07
										Fasting insulin	G	0.93	G	0.99
										HOMA-IR	A	0.75	A	0.23
										ISI_0–120	A	0.40	G	0.69
10	*HHEX*	rs7923837 rs1111875	rs10509645	0.702	0.33	G/T	1.04 (0.75–1.5)	0.81	0.49	FPG	T	0.83	T	0.35
										mFPG	T	0.67	T	0.43
										A1C	T	0.07	T	0.27
										Fasting insulin	T	0.13	T	0.87
										HOMA-IR	T	0.19	T	0.65
										ISI_0–120	G	0.48	T	0.99
11	LOC387761	rs7480010	rs10501278	0.284	0.11	A/T	0.75 (0.43–1.3)	0.30	0.89	FPG	T	0.83	T	0.28
										mFPG	A	0.79	T	0.36
										A1C	A	0.71	A	0.75
										Fasting insulin	T	0.10	T	**0.008**
										HOMA-IR	T	0.13	T	**0.01**
										ISI_0–120	A	0.30	A	**0.047**
11	*EXT2*	rs1113132 rs11037909 rs3740878	rs962848	0.474	0.33	C/T	1.16 (0.84–1.6)	0.37	0.13	FPG	T	**0.002**	T	**0.002**
										mFPG	T	**0.04**	T	**0.007**
										A1C	T	0.70	T	0.35
										Fasting insulin	T	0.98	C	0.83
										HOMA-IR	T	0.22	T	0.27
										ISI_0–120	C	0.21	C	**0.049**
6	*CDKAL1*	rs7754840	rs2328545	0.347	0.13	C/G	0.98 (0.67–1.5)	0.93	0.73	FPG	C	0.90	C	0.29
										mFPG	G	0.97	C	0.31
										A1C	G	0.19	G	0.80
										Fasting insulin	C	0.97	C	0.76
										HOMA-IR	C	0.83	C	0.43
										ISI_0–120	C	0.09	C	0.19
3	*IGF2BP2*	rs1470579 rs4402960	rs10513800	0.328	0.22	A/C	0.77 (0.55–1.1)	0.12	0.71	FPG	A	0.28	A	0.21
										mFPG	A	**0.03**	A	0.37
										A1C	A	0.26	A	0.49
										Fasting insulin	C	0.12	C	0.55
										HOMA-IR	C	0.45	C	0.94
										ISI_0–120	A	0.78	C	0.60

SNPs shown to be significantly associated with type 2 diabetes in recent high-density GWA scans (16–19) were examined for association with diabetes incidence or related quantitative traits within the FHS 100K dataset. A 100K SNP was found to be in moderate to low linkage disequilibrium with the previously associated SNP in each case. Nominally significant associations in a consistent direction were found in FHS for five of these SNPs (shown in boldface). Chr, chromosome.

ducted datasets is essential. This can take the form of a staged approach or a joint analysis of several stages, which requires statistical integration of disparate datasets (13). It is estimated that ~30% of common variants are captured by the 100K array (12); thus, our genotyping density and moderate sample size do not represent a comprehensive assessment of common variants in the genome. In addition, because of the relatively low number of incident diabetes events in the FHS on one hand and the rich trove of longitudinal phenotypic data on the other, we are best poised to detect associations with quantitative glycemic traits. The availability of other datasets genotyped on the same platform (21–23), as well as larger and denser GWA scans that also contain quantitative trait data (16) (http://www.broad.mit.edu/diabetes/), allow us to compare results, validate our SNP selection strategies, and replicate and/or extend findings from other groups.

To prioritize SNPs from the 100K array results and to maximize the likelihood of selecting true positive associations, we developed a method that harnesses the wealth of phenotypic data in FHS while recognizing the limited statistical power of this modestly sized sample. In addition to choosing SNPs based solely on small *P* values, we selected SNPs that showed consistent nominal associations with multiple related traits. We reasoned that such a SNP is less likely to be a spurious finding and more likely to represent a real association with hyperglycemia/insulin resistance, at least in the FHS. We tested this latter strategy by seeking replication in a nonoverlapping cohort of unrelated FHS par-

ticipants and both approaches by in silico comparisons with three 100K and one 500K datasets.

None of the primary FHS results achieved convincing replication across multiple datasets, although the two SNPs rs2863389 (not near a known gene) and rs7935082 (in intron 4 of the ubiquitous membrane-spanning 4-domain subfamily A member 7, *MS4A7*) showed consistent associations in two other populations (Supplementary Table 4). This low yield could be due to either initial false-positive associations or false-negative follow-up testing. In regard to the former, we note that our set of positive results did not depart significantly from the null expectation. A fraction of false-positive results may have been introduced by systematic enrichment of low *P* values in FHS; although this might have affected the multiple related trait selection strategy, theoretically, it should not have distorted the *P* value ranks used in our pure *P* value approach. Alternatively, true positives may have been missed because of low power. Given the emerging notion that a ceiling for the combination of effect size/allele frequency in type 2 diabetes seems to hover around that of *TCF7L2* rs7903146 (16) and that diabetes-related polymorphisms may only explain a small fraction of the variance in quantitative glycemic traits, it is not surprising that our initial sample of ~1,000 individuals was insufficient to detect a large number of novel findings and that none of our *P* values achieved genome-wide significance.

In regard to the absence of replication, differences in ancestry among cohorts and the relatively small sample sizes of the other 100K datasets may have also precluded us from obtaining significant *P* values in replication, even among true positive findings. A planned joint meta-analysis of all four datasets where all test statistics are combined may help prioritize the few true positive results that remain consistent across populations. Nevertheless, the strength of the FHS resource lies in its quantitative trait database rather than in diabetes incidence; thus, such integration may be more fruitful when limited to such phenotypes.

The larger DGI 500K dataset, which contains publicly available diabetes and glycemic trait statistics for a European population similar to FHS, provides another convenient replication venue. Here, we have tested our top results and obtained a ~3% yield of SNPs that show suggestive evidence of replication. Of these 25 SNPs, one of them (rs952635, an intronic SNP in the *PDE4B* gene encoding a cAMP-specific phosphodiesterase expressed in brain, heart, lung, and skeletal muscle) holds particular promise in that it showed remarkably consistent associations with multiple glycemic traits in the 100K scan and replication in the FHS unrelated sample, and its proxy rs6664618 was also associated with FPG and HOMA-IR in the DGI dataset (Supplementary Table 4).

The worst-case scenario would dictate that fundamental flaws in the 100K genotyping process, in the genotype-calling algorithm, in our phenotypic characterization, or in our statistical procedures prevented us from making striking discoveries; if that were the case, we would not expect to be able to detect any real associations. The convincing results we have obtained for SNPs in *TCF7L2* and other genes reported by others (16–19) indicate that FHS is a viable sample in which to replicate real results of adequate magnitude and characterize the phenotypic effects of such variants on glycemic traits and their change over time.

The particular utility of the population characteristics of the FHS cohort is illustrated by our attempt to clarify the effects of *TCF7L2* variants on diabetes while accounting for obesity. The association of *TCF7L2* rs7903146 with type 2 diabetes is incontrovertible, having reached a *P* value $<10^{-80}$ after meta-analysis of nearly 50,000 samples (38). This variant appears to confer risk of diabetes by causing an impairment in insulin secretion (39–41). Recently, DECODE investigators have suggested that a haplotype largely defined on the basis of the alternate C allele at rs7903146 (HapA) is associated with obesity, when case and control subjects are analyzed separately (42). However, that strategy also imposes constraints in ascertainment: Control subjects who carry the diabetes risk allele must be protected from diabetes by other factors, including a lower BMI (thus resulting in an apparent association of the C allele with BMI), whereas case subjects who carry the protective C allele must have diabetes on the basis of other components of risk, including BMI (thus resulting in the same apparent association). Therefore, population samples free of diabetes ascertainment criteria such as FHS are needed to verify whether these associations are real. We did not observe a significant association of the 100K SNP rs7100927 (which is in strong linkage disequilibrium with the variants that tag HapA) with BMI, and we found that rs7903146 was nominally associated with higher rather than lower insulin resistance. These results are consistent with those reported in other large population samples (43,44) and indicate that ascertainment on diabetes case-control status may introduce spurious associations when neither phenotypic traits nor haplotype variants are independent.

Some findings presented here appear promising and merit further exploration, including the 13 SNPs that replicated across the two FHS samples, the 5 SNPs that stood out within the Type 2 Diabetes 100K Consortium in silico replication effort, and the 25 SNPs that show suggestive replication in the DGI. None of these 41 unique SNPs lies in genes that would be considered high-likelihood biological candidates, and none represents a coding change. The precedent afforded by *TCF7L2* reassures us that nonbiased genetic screens can uncover novel biology. The upcoming high-density GWA scan in ~9,000 Framingham participants (the FHS SHARe Study) and its integration with other new, public GWA resources should provide well-powered tools with which to continue to draft, and perhaps complete, the genetic architecture of common genetic variants predisposing to type 2 diabetes.

ACKNOWLEDGMENTS

J.C.F. has received National Institutes of Health (NIH) Research Career Award K23 DK-65978-04. J.B.M. has received an American Diabetes Association Career Development Award. This study has been supported by the National Heart, Lung, and Blood Institute FHS (contract no. N01-HC-25195) and the Boston University Linux Cluster for Genetic Analysis funded by the NIH National Center for Research Resources Shared Instrumentation Grant (1S10RR163736-01A1). The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278-01 from the National Center for Research Resources.

We thank our co-investigators of the 100K Type 2 Diabetes Consortium at the University of Maryland, the University of Chicago, the University of Texas-Houston, and National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, for their generous, trusting, and industrious collaboration and our colleagues at the FHS,

the Broad Institute, and Massachusetts General Hospital for many helpful and constructive discussions framing the conduct and implications of type 2 diabetes GWA scans.

REFERENCES

1. Florez JC, Hirschhorn JN, Altshuler D: The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genomics Hum Genet* 4:257–291, 2003
2. Fajans SS, Bell GI, Polonsky KS: Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* 345:971–980, 2001
3. Barroso I: Genetics of type 2 diabetes. *Diabet Med* 22:517–535, 2005
4. Beysen D, Raes J, Leroy BP, Lucassen A, Yates JRW, Clayton-Smith J, Ilyina H, Brooks SS, Christin-Maitre S, Fellous M, Fryns JP, Kim JR, Lapunzina P, Lemyre E, Meire F, Messiaen LM, Oley C, Splitt M, Thomson J, Van de Peer Y, Veitia RA, De Paepe A, De Baere E: Deletions involving long-range conserved nongenic sequences upstream and downstream of *FOXL2* as a novel disease-causing mechanism in blepharophimosis syndrome. *Am J Hum Genet* 77:205–218, 2005
5. Drake JA, Bird C, Nemesh J, Thomas DJ, Newton-Cheh C, Reymond A, Excoffier L, Attar H, Antonarakis SE, Dermitzakis ET, Hirschhorn JN: Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet* 38:223–227, 2006
6. Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6:95–108, 2005
7. Reich DE, Gabriel SB, Altshuler D: Quality and completeness of SNP databases. *Nat Genet* 33:457–458, 2003
8. Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989, 2001
9. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265, 2005
10. de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: Efficiency and power in genetic association studies. *Nat Genet* 37:1217–1223, 2005
11. de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, Bersaglieri T, Penney KL, Butler J, Young S, Onofrio RC, Lyon HN, Stram DO, Haiman CA, Freedman ML, Zhu X, Cooper R, Groop L, Kolonel LN, Henderson BE, Daly MJ, Hirschhorn JN, Altshuler D: Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 38:1298–1303, 2006
12. Pe'er I, de Bakker PIW, Maller J, Yelensky R, Altshuler D, Daly MJ: Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet* 38:663–667, 2006
13. Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. 38:209–213, 2006
14. The International HapMap Consortium: The International HapMap Project. *Nature* 426:789–796, 2003
15. The International HapMap Consortium: A haplotype map of the human genome. *Nature* 437:1299–1320, 2005
16. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:828–830, 2007
17. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University and Novartis Institutes for BioMedical Research: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336, 2007
18. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JRB, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney ASF, The Wellcome Trust Case Control Consortium, McCarthy MI, Hattersley AT: Replication of genome-wide association signals in U.K. samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341, 2007
19. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding C-J, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li X-Y, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345, 2007
20. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K: A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775, 2007
21. Hanson RL, Bogardus C, Duggan D, Kobes S, Infante AM, Baier LJ, Knowler WC: A search for variants associated with young-onset type 2 diabetes in American Indians among 80,044 single nucleotide polymorphisms. *Diabetes* 56:3045–3052, 2007
22. Hayes MG, Pluzhnikov A, Miyake K, Sun Y, Ng MCY, Roe CA, Below JE, Nicolae RI, Konkashbaev A, Bell GI, Cox NJ, Hanis CL: Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* 56:3033–3044, 2007
23. Rampersaud E, Damcott CM, O'Connell J, McArdle P, Shen H, Fu M, Shelton J, Ying J, Shi X, Ott SH, Zhang L, Zhao Y, Mitchell BD, Shuldiner AR: Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes–related quantitative traits and from independent populations. *Diabetes* 56:3053–3062, 2007
24. Meigs JB, Manning AK, Fox CS, Florez JC, Cupples LA, Dupuis J: Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Med Genet* 8 (Suppl. 1):S16, 2007
25. Meigs JB, Panhuysen CIM, Myers RH, Wilson PWF, Cupples LA: A genome-wide scan for loci linked to plasma levels of glucose and HbA1c in a community-based sample of Caucasian pedigrees: the Framingham Offspring Study. *Diabetes* 51:833–840, 2002
26. Meigs JB, Nathan DM, Wilson PWF, Cupples LA, Singer DE: Metabolic risk factors worsen continuously across the spectrum of nondiabetic glucose tolerance: the Framingham Offspring Study. *Ann Intern Med* 128:524–533, 1998
27. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC: Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28:412–419, 1985
28. Gutt M, Davis CL, Spitzer SB, Llabre MM, Kumar M, Czarnecki EM, Schneiderman N, Skyler JS, Marks JB: Validation of the insulin sensitivity index (ISI0,120): comparison with other measures. *Diabetes Res Clin Pract* 47:177–184, 2000
29. American Diabetes Association: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 20:1183–1197, 1997
30. Fox CS, Sullivan L, D'Agostino RB Sr, Wilson PW: The significant effect of diabetes duration on coronary heart disease mortality: the Framingham Heart Study. *Diabetes Care* 27:704–708, 2004
31. Meigs JB, Cupples LA, Wilson PWF: Parental transmission of type 2 diabetes mellitus: the Framingham Offspring Study. *Diabetes* 49:2201–2207, 2000
32. Cupples LA, Arruda H, Benjamin EJ, D-Agostino RB Sr, Demissie S, DeStefano AL, Dupuis J, Falls K, Fox CS, Gottlieb DJ, Govindaraju DR, Guo C-Y, Heard-Costa N, Hwang S-J, Kathiresan S, Kiel DP, Laramie JM, Larson MG, Levy D, Liu C-Y, Lunetta KL, Mailman MD, Manning AK, Meigs JB, Murabito JM, Newton-Cheh C, O'Connor GT, O'Donnell CJ, Pandey M, Seshadri S, Vasan RS, Wang ZY, Wilk JB, Wolf PA, Yang Q, Atwood LD: The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. *BMC Med Genet* 8 (Suppl. 1):S1, 2007
33. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB, Colditz G, Hinney A, Hebebrand J, Koberwitz K, Zhu X, Cooper R, Ardlie K, Lyon H, Hirschhorn JN, Laird NM, Lenburg ME, Lange C, Christman MF: A common genetic variant 10 kb upstream of *INSIG2* is associated with adult and childhood obesity. *Science* 312:279–283, 2006
34. Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Koster H: Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci U S A* 96:10016–10020, 1999
35. Therneau TM, Grambsch PM, Pankratz VS: Penalized survival models and frailty. *J Comput Graph Stat* 12:156–175, 2003
36. Horvath S, Wei E, Xu X, Palmer LJ, Baur M: Family-based association test method: age of onset traits and covariates. *Genet Epidemiol* 21:S403–S408, 2001
37. Iturria SJ, Williams JT, Almasy L, Dyer TD, Blangero J: An empirical test of the significance of an observed quantitative trait locus effect that preserves additive genetic variation. *Genet Epidemiol* 17:S169–S173, 1999

38. Florez JC: The new type 2 diabetes gene *TCF7L2*. *Curr Opin Clin Nutr Metab Care* 10:391–396, 2007
39. Florez JC, Jablonski KA, Bayley N, Pollin TI, de Bakker PIW, Shuldiner AR, Knowler WC, Nathan DM, Altshuler D, the Diabetes Prevention Program Research Group: TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 355:241–250, 2006
40. Damcott CM, Pollin TI, Reinhart LJ, Ott SH, Shen H, Silver KD, Mitchell BD, Shuldiner AR: Polymorphisms in the transcription factor 7-like 2 (*TCF7L2*) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* 55:2654–2659, 2006
41. Saxena R, Gianniny L, Burtt NP, Lyssenko V, Giuducci C, Sjogren M, Florez JC, Almgren P, Isomaa B, Orho-Melander M, Lindblad U, Daly MJ, Tuomi T, Hirschhorn JN, Ardlie KG, Groop LC, Altshuler D: Common single nucleotide polymorphisms in *TCF7L2* are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* 55:2890–2895, 2006
42. Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, Gunnarsdottir S, Adeyemo A, Chen Y, Chen G, Reynisdottir I, Benediktsson R, Hinney A, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Schafer H, Faruque M, Doumatey A, Zhou J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Sigurdsson G, Hebebrand J, Pedersen O, Thorsteinsdottir U, Gulcher JR, Kong A, Rotimi C, Stefansson K: Refining the impact of *TCF7L2* gene variants on type 2 diabetes and adaptive evolution. *Nat Genet* 39:218–225, 2007
43. Cauchi S, Meyre D, Choquet H, Dina C, Born C, Marre M, Balkau B, Froguel P, the DESIR Study Group: *TCF7L2* variation predicts hyperglycemia incidence in a French general population: the Data from an Epidemiological Study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes* 55:3189–3192, 2006
44. Loos RJF, Franks PW, Francis RW, Barroso I, Gribble FM, Savage DB, Ong KK, O'Rahilly S, Wareham NJ: TCF7L2 polymorphisms modulate proinsulin levels and β-cell function in a British Europid population. *Diabetes* 56:1943–1947, 2007