Identification of Type 2 Diabetes Genes in Mexican Americans Through Genome-Wide Association Studies

M. Geoffrey Hayes,[1] Anna Pluzhnikov,[1] Kazuaki Miyake,[1] Ying Sun,[2] Maggie C.Y. Ng,[1] Cheryl A. Roe,[1] Jennifer E. Below,[2] Raluca I. Nicolae,[2] Anuar Konkashbaev,[1] Graeme I. Bell,[1,2] Nancy J. Cox,[1,2] and Craig L. Hanis[3]

OBJECTIVE—The objective of this study was to identify DNA polymorphisms associated with type 2 diabetes in a Mexican-American population.

RESEARCH DESIGN AND METHODS—We genotyped 116,204 single nucleotide polymorphisms (SNPs) in 281 Mexican Americans with type 2 diabetes and 280 random Mexican Americans from Starr County, Texas, using the Affymetrix GeneChip Human Mapping 100K set. Allelic association exact tests were calculated. Our most significant SNPs were compared with results from other type 2 diabetes genome-wide association studies (GWASs). Proportions of African, European, and Asian ancestry were estimated from the HapMap samples using structure for each individual to rule out spurious association due to population substructure.

RESULTS—We observed more significant allelic associations than expected genome wide, as empirically assessed by permutation (14 below a P of 1×10^{-4} [8.7 expected]). No significant differences were observed between the proportion of ancestry estimates in the case and random control sets, suggesting that the association results were not likely confounded by substructure. A query of our top ~1% of SNPs ($P < 0.01$) revealed SNPs in or near four genes that showed evidence for association ($P < 0.05$) in multiple other GWAS interrogated: rs979752 and rs10500641 near *UBQLNL* and *OR52H1* on chromosome 11, rs2773080 and rs3922812 in or near *RALGPS2* on chromosome 1, and rs1509957 near *EGR2* on chromosome 10.

CONCLUSIONS—We identified several SNPs with suggestive evidence for replicated association with type 2 diabetes that merit further investigation. ***Diabetes*** **56:3033–3044, 2007**

From the [1]Department of Medicine, University of Chicago, Chicago, Illinois; the [2]Department of Human Genetics, University of Chicago, Chicago, Illinois; and the [3]Human Genetics Center, University of Texas Health Sciences Center, Houston, Texas.

Address correspondence and reprint requests to Nancy J. Cox, PhD, Department of Medicine, University of Chicago, 5841 S. Maryland Ave., MC6091, Chicago, IL 60637. E-mail: ncox@bsd.uchicago.edu; or Craig L. Hanis, PhD, Human Genetics Center, University of Texas Health Science Center at Houston, P.O. Box 20186, Houston, TX 77225. E-mail: craig.l.hanis@uth.tmc.edu.

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

Published ahead of print at http://diabetes.diabetesjournals.org on 10 September 2007. DOI: 10.2337/db07-0482.

Additional information for this article can be found in an online appendix at http://dx.doi.org/10.2337/db07-0482.

BRLMM, Bayesian robust-fitting linear model with Mahalanobis distance classifier; DGI, Diabetes Genetics Initiative; DM, dynamic modeling; FDR, false discovery rate; FHS, Framingham Heart Study; GEL, genotype calling algorithm using empirical likelihood; GWAS, genome-wide association study; HWE, Hardy-Weinberg equilibrium; LD, linkage disequilibrium; MAF, minor allele frequency; POA, proportion of ancestry; SNP, single nucleotide polymorphism.

© 2007 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Diabetes continues to pose a substantial and increasing burden of morbidity and mortality on society, especially among minority populations. In the U.S., ~18 million people have diabetes, of which one-third remain undiagnosed and most (90–95%) have type 2 diabetes (1). By 2050, rates of diagnosed diabetes are projected to more than double to 39 million, with fully one-third of children born in the year 2000 expected to develop diabetes over their lifetime (1). Minority populations, such as Mexican Americans, have a disproportionate incidence of diabetes (2–5). For example, the Mexican-American population from Starr County, Texas, has the highest diabetes-specific morbidity and mortality of any county in Texas, yet it is only the 53rd largest of Texas' 254 counties. Age-specific prevalences are three- to fivefold higher than the general U.S. population (4,6), and in the last two decades alone there has been a 74% increase in type 2 diabetes prevalence in those aged ≥25 years in this population.

Population studies, pedigree investigations, molecular studies, and animal models consistently implicate a substantial role for genes in determining risk for type 2 diabetes (see 7,8). These studies also establish that no simple genetic model adequately explains risk for diabetes. Rather, there are likely to be multiple genes with small to modest effects that interact with each other and with environmental factors to affect susceptibility (9–11). This view of the genetics of diabetes is able to explain both its population and familial aggregation and implies that we are looking for genes whose effects are neither necessary nor sufficient to cause disease.

A great deal of effort has been expended in identifying genes underlying the risk for type 2 diabetes, including genome linkage scans (see 12,13), candidate gene studies (e.g., 14), and, more recently, genome-wide association studies (GWASs) (15–19). To date, such studies have yielded several replicated type 2 diabetes–associated risk genes including *CAPN10*, *CDKAL1*, *CDKN2A*, *HHEX*, *HNF4A*, *IGF2BP2*, *KCNJ11*, *PPARG*, *SLC30A8*, and *TCF7L2* (20–25), but none account for a large proportion of the risk of developing type 2 diabetes in the particular population under study nor are any seen universally across all populations. Again, this suggests that many more type 2 diabetes susceptibility genes remain undiscovered.

Over the past decade, we have conducted genome-wide linkage scans on Mexican-American families from Starr County, Texas, to localize genes conferring risk to type 2 diabetes and were successful in positionally cloning the *CAPN10* gene as a type 2 diabetes susceptibility locus (6,20). Given the increased power in association studies over linkage studies (26) for complex genetic diseases

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

TABLE 1
Descriptive statistics for the individuals with type 2 diabetes in the primary GWAS set from Starr County, Texas

n	281
Sex (*n* female)	174
Age (years)	57.9 ± 10.7
Age at diagnosis (years)	45.9 ± 10.1
Fasting glucose (mg/dl)	190.3 ± 75.0
A1C (%)	11.6 ± 3.5
BMI (kg/m^2)	31.4 ± 6.2

Data are means ± SD, unless otherwise indicated.

such as type 2 diabetes, we conducted a GWAS of a >600-member case-control set to identify additional genomic regions harboring type 2 diabetes susceptibility loci in the Starr County population. We present the results of this type 2 diabetes GWAS, the first in a non-Caucasian population, along with supporting evidence for replication from available GWASs, primarily the three accompanying this one (27,28,29).

RESEARCH DESIGN AND METHODS

This study was completed in a Mexican-American population from Starr County, Texas. We selected as unrelated cases 291 individuals who represent the youngest age-at-onset individuals from the multiplex families in our previous linkage studies and for whom we have the richest phenotypic data. The comparison individuals are not true control subjects in that their diabetes status is unknown. Rather, they are a representative sample of 323 unrelated individuals drawn from a random survey of Starr County. Of this case and random control set, 281 and 280 individuals were analyzed (see "Quality control" below) and are described in Table 1. An overlapping cohort (online appendix Table 1 [available at http://dx.doi.org/10.2337/db07-0482]) of 760 individuals (including 555 of the 561 individuals analyzed) was used to verify genotypes before single nucleotide polymorphism (SNP) selection for follow-up replication.

Diabetes was classified based on earlier National Diabetes Data Group recommendations (30), namely, previously diagnosed diabetes and current or sustained use of glucose-lowering medications, fasting glucose ≥140 mg/dl on more than one occasion, or a 2-h postload glucose of ≥200 mg/dl. Individuals were considered to have type 2 diabetes unless they were diagnosed before age 30 years, had a BMI <30 kg/m^2, and had used insulin continuously since diagnosis.

Genotypying. Genomic DNA was isolated from lymphocytes and quantified by picogreen. The genotyping assay was performed according to the manufacturer protocols (Affymetrix, Santa Clara, CA) by the Functional Genomics Core Facility at the University of Chicago. In brief, 250 ng DNA was digested with the restriction enzymes *Xba*I and *Hin*dIII, followed by adaptor ligation. The DNA fragments were then amplified, fragmented, labeled, and hybridized overnight to the Affymetrix GeneChip Human Mapping 100K *Xba*I and *Hin*dIII arrays. The arrays were scanned with the Affymetrix 7G scanner and analyzed with Affymetrix GeneChip DNA Analysis Software to generate hybridization intensity files and subsequent dynamic modeling (DM) algorithm–derived genotypes.

Case and random samples were dispersed randomly throughout the plates to eliminate the possibility of spurious associations due to systematic differences in genotyping conditions between experiments. Genotypes were called using the default Affymetrix DM algorithm and two improved algorithms, (GEL) (31) and Bayesian RLMM (BRLMM) (32,33). After removal of monomorphic markers, we analyzed genotypes for 112,541 autosomal SNPs of the possible 116,204 SNPs interrogated on the array. We anticipate analyzing X chromosome polymorphisms at a later date. Genotyping in verification sets was performed using TaqMan assays on the ABI Prism 7900HT Sequence Detection System.

Statistical methods. We examined the case-random control cohort for evidence of related individuals that went undetected during sample collection using PLINK (34). Pairs with identity-by-descent estimates >0.20 were trimmed, preferentially keeping case rather than control subjects and individuals with higher genotype call rates if the pair was a case-case or control-control.

Fisher's exact tests for allelic associations and departures from Hardy-Weinberg equilibrium (HWE) were calculated for all polymorphic SNPs. We did not remove any of the SNPs for strict quality-control reasons but, rather, cataloged quality-control indicators for each SNP and considered them during the interpretation of the data. We observed that our most significant SNPs, those with *P* values between 5.1×10^{-6} and 6.2×10^{-13}, had highly significant departures from HWE ($P < 0.001$) in random control subjects or call rates <0.85, so we subsequently focused our attention on those that surpassed these thresholds. We also set a minor allele frequency (MAF) ≥0.05 criterion, as the allelic associations at SNPs below this threshold are largely driven by differences in a small number of individuals. We anticipate following-up rare polymorphisms with significant evidence for association separately at a later date. A total of 88,142 SNPs passed these criteria (Fig. 1).

False discovery rates (FDRs) were estimated by conducting the allelic association test in 1,000 permutations (permuting the case and random labels) and tabulating the *P* values at given thresholds. We also conducted logistic regressions between type 2 diabetes status and genotypes under an additive model, with and without a proportion of European ancestry covariate. This was not meant as a substitute for the allelic associations but simply to provide a reasonable approach to investigate how the estimated proportions of ancestry might affect the results when included as a covariate. All statistical analyses were performed using R (available at http://www.rproject.org). Measures of linkage disequilibrium (LD) were calculated using GOLD (35).

Using a population prevalence of 10%, we estimated that a case-random study was sufficiently powered (80%) to detect a genotype relative risk of ~1.6 under dominant, recessive, and additive models in the mid-range of allele frequencies (36).

Assessing admixture proportions. We compared the full set of genotypes for the 116,204 SNPs in the Mexican-American subjects (MA group) and in the unrelated HapMap samples (60 Europeans from Utah from the Centre d'Etude du Polymorphisme Humain [CEU group]; 60 Yoruba from Ibadan, Nigeria [YRI group]; and 89 Asians [ASN group] including Japanese subjects from Tokyo [JPT group] and Han Chinese from Beijing [CHB group]) as proxies for Native Americans (see online appendix). The Asian HapMap samples were chosen as proxies because no 100K data exist for an appropriate Native American population once thought to be ancestral to the Mexican Americans under investigation here. This leaves the Asian samples as the most appropriate proxy. After removing SNPs either not typed or monomorphic in all four populations (CEU, YRI, ASN, and MA groups), we divided the remaining 101,150 SNPs into 10 equal subsets (by taking every 10th SNP) to reduce the degree of LD between SNPs (median intermarker distance ~250 kb in the subsets). To estimate genome-wide proportions of ancestry (POAs) for each individual, we ran structure (37) for each of the 10 subsets using the HapMap populations as learning samples (fixed population identity) and subsequently averaged the estimated POAs across the subsets. The structure runs were conducted under an admixture model with default parameter settings of 10,000 burn-in replications and 10,000 estimating replications after burn in. Altering the prior migration probability from 0.001 to 0.1 had little effect on the results, and we present the POA estimates for the 0.1 runs herein.

In silico replication. We entered into a consortium to share results with three other groups analyzing type 2 diabetes GWAS data in three distinct populations (Amish, Pima Indians, and Framingham Heart Study [FHS]) (Tables 6 and 7), each with different study designs but using the same genotyping platform (27,28,29). Each group requested summary data for their top ~1,000 SNPs following criteria specific to each group in the Type 2 Diabetes 100K GWAS Consortium and shared the same for the other groups' best signals. We requested summary data for our top 1,196 most significant high-quality SNPs (those with $P < 0.01$ and passing the quality-control thresholds described above). We directly compared our Fisher's exact tests for allelic associations to type 2 diabetes in the Mexican Americans with the type 2 diabetes association tests under an additive model in the Amish, the type 2 diabetes association tests by generalized estimating equations and family-based association tests in the FHS, and the case-control and within-family association tests in the Pima Indians. We considered a Mexican-American type 2 diabetes–associated SNP to be in silico replicated if it was associated at $P \leq 0.05$ in the same direction (i.e., the same allele was associated with type 2 diabetes) in at least one other 100K GWAS (Fig. 1).

We also queried our data against the March 2007 prereleased data from a similar study in a Scandinavian cohort (Diabetes Genetics Inititative [DGI]) (available at http://www.broad.mit.edu/diabetes/) but conducted with a denser genotyping platform. We compared our top 1,196 association signals with any SNP reaching nominal significance ($P \leq 0.05$) in the other GWASs that were within 150 kb and had $r^2 \geq 0.8$ in either the HapMap Europeans (CEU group) or Asians (ASN group). Again, we considered a Mexican-American type 2 diabetes–associated SNP to be in silico replicated if another SNP with $r^2 \geq 0.8$ in the CEU or ASN groups to the Mexican-American type 2 diabetes–associated SNP was associated at $P \leq 0.05$ in the same direction (i.e., the same allele was associated with type 2 diabetes) in the DGI GWAS (Fig. 1).

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

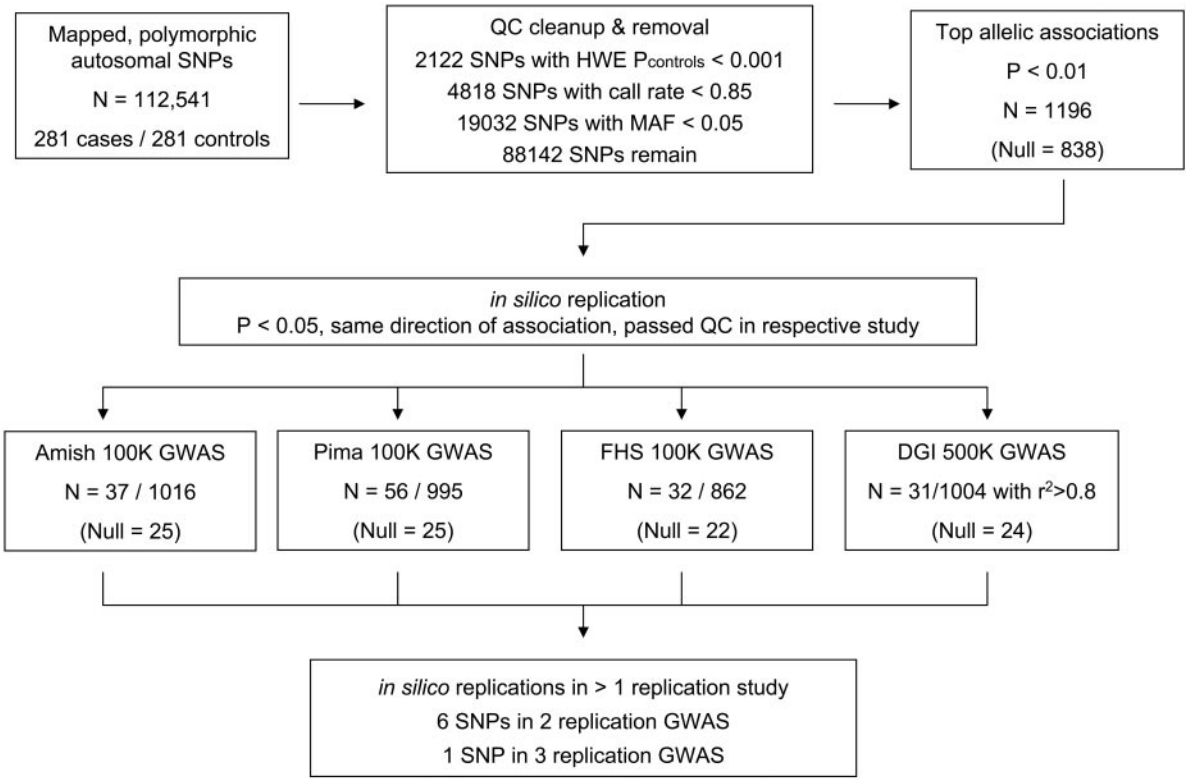

FIG. 1. Schematic of analysis and in silico replication plan.

RESULTS

Quality control. We selected for subsequent analysis the *Xba*I and *Hin*dIII chip experiment with the highest call rate for each individual and two or less discordant genotype calls for the 31 SNPs duplicated on the two chips. Of 323 random control and 291 case subjects for which genotyping was attempted, 316 and 287, respectively, met these criteria. The mean per-chip call rate using the DM algorithm was >95%, although the *Xba*I chip performed slightly better than the *Hin*dIII chip (95.8 and 95.2%, respectively). For both chips, >92% of experiments had call rates >90% (92.4% for *Xba*I and 93.3% for *Hin*dIII). Using the DM algorithm calls, we observed significant ($P < 0.001$) departures from HWE in a substantial number of SNPs (9.8% all samples, 5.1% random subjects only, and 5.2% case subjects only). This is largely attributable to SNPs with excess homozygosity, consistent with nonrandom missing data (heterozygotes have more "no-calls" since their intermediacy between the two homozygote classes renders them more difficult to call than the two homozygote classes).

Using GEL, both increased the call rate (97.2% mean *Xba*I call rate with 95.4% of experiments having >90% call rates and 96.7% mean *Hin*dIII call rate with 95.2% of experiments having >90% call rates) and reduced nonrandom missing data by increasing the proportion of heterozygote genotype calls (online appendix Table 2), which subsequently reduced the number of SNPs showing significant ($P < 0.001$) departures from HWE (4.0% all samples, 2.2% random subjects only, and 2.0% case subjects only). With either genotype calling algorithm (GEL or DM), there was no substantial case-random control difference throughout the majority of the distribution of per-chip genotype call rates, although there were some outliers in the tails of the distribution (online appendix Fig. 1).

For comparative purposes, we also called the genotypes with the BRLMM algorithm, which again yielded an increased proportion of heterozygotes (online appendix Table 2). In contrast to GEL and DM, which call the genotypes for each chip experiment individually, BRLMM normalizes the intensity patterns across all chip experiments, and therefore it is recommended that only chips with DM call rates >90% be used. To do this would require 83 chip-genotyping experiments (7.1%) to be removed from consideration, substantially reducing our power. We experimented with lowering the DM algorithm call rate threshold and found that BRLMM overcompensates for missing genotypes in the heterozygote class and increases the proportion of heterozygotes to unrealistic levels in chips with DM call rates <90% (online appendix Fig. 2). Given the limited number of samples under investigation, the marginal increase in genotype calls using BRLMM over GEL and the high concordance rates between GEL and BRLMM (online appendix Tables 2 and 3), we decided to report results using the GEL algorithm to retain maximal power.

Allelic associations. The chromosomal distribution of Fisher's exact test *P* values for the 88,142 SNPs passing our quality-control thresholds are presented in Fig. 2. A total of 1,196 had allelic association $P < 0.01$ and are presented in online appendix Table 4. The 14 best ($P < 10^{-4}$) SNPs (Table 2) survey 13 different regions of the genome and are in or near *ANKRD50*, *DYRK2*, *EPB41L3*, *GRIK1*, *HPSE2*, *ICA1*, *IFNG*, *NXPH1*, *OR13D1*, *SDF2L1*, *SORBS1*, *SPRY1*, *SLC24A3*, and *TMEFF2*. Two adjacent SNPs on the Affymetrix GeneChip Human Mapping 100K set (rs10518442 and rs1498024) on chromosome 4, in and near *ANKRD50*, respectively, are in perfect ($r^2 = 1$) LD with each other and are associated at $P < 10^{-5}$. Our most significantly associated SNP, rs1932465, has a *P* value of

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

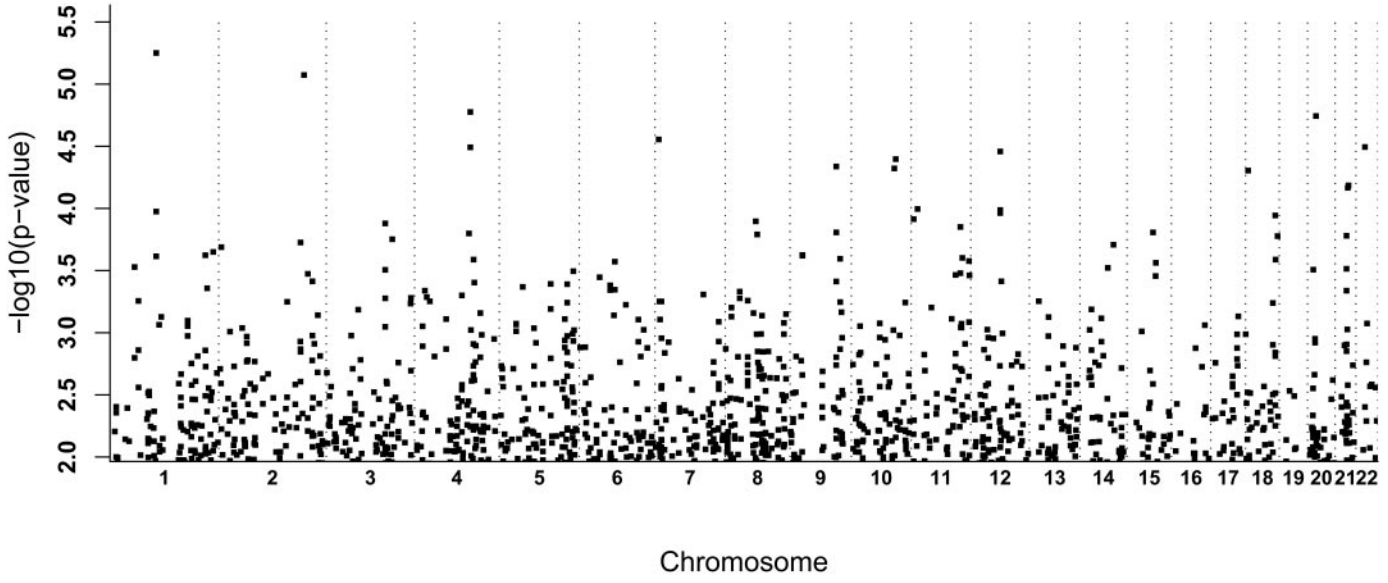

FIG. 2. Fisher's exact test $-\log_{10}(P$ values) for tests of association between the 88702 high-quality autosomal SNPs (SNPs with HWE departure $P > 0.001$ in random subjects, call rates ≥0.85, and MAF ≥0.05) and type 2 diabetes affection status.

5.6×10^{-6}, approximately one order of magnitude below a conservative Bonferroni correction for multiple tests ($0.05/88,142 = 5.7 \times 10^{-7}$). We note that none of the most significant signals are SNPs with low MAFs (0.05–0.10; we excluded SNPs with MAFs <0.05). While this observation is not unexpected given the reduced power for detecting susceptibility loci with allele frequencies at the tail of the MAF distribution, it remains noteworthy since nonrandom patterns of missing data and other genotyping errors not detected in quality-control analysis often lead to SNPs with low MAFs being disproportionately found among those with the most significant *P* values, which are subsequently poorly replicated.

Using permutations, we empirically estimated the FDR at various thresholds (online appendix Table 5) and found that we observe many more significant allelic associations than expected genome wide. For our best signals, those meeting a $P \leq 10^{-4}$ significance threshold, the FDR is estimated to be 62%. This suggests that 8–9 of the 14 SNPs will likely turn out to be false-positives. We also compared the distribution of allelic association *P* values against a uniform distribution (online appendix Fig. 3). Our observed distribution begins to depart from the expected uniform one at approximately $P = 10^{-2}$, suggesting an appropriate threshold for investigating in silico replication in order to prioritize SNPs for follow-up.

Ancestry estimates in case and random control samples. The Starr County Mexican-American population is a relatively homogeneous (97.5% Hispanic by self-report [available at factfinder.census.gov]) yet highly admixed population with contributions to the contemporary gene pool from individuals of Spanish, Native American, and African ancestry. Previous estimates using classical markers suggest ancestry proportions of 61, 31, and 8%, respectively (38). Since population substructure can yield spurious case-control associations, we investigated the patterns of ancestry in the case and random control subjects used in the GWAS. We observed no significant difference between the 10 subsets (online appendix Fig. 4), which permitted us to average the admixture proportions over them. The ancestry estimates observed using the 100K SNP sets (68% European, 27% Asian, and 6% African)

TABLE 2
SNPs most significantly associated with type 2 diabetes

dbSNP rs	Chr.	Position*	Gene*	Allele 1/2	Control frequency (a2)	Case frequency (a2)	*P*†	OR†
rs1932465	1	104418527		C/G	0.881	0.958	5.61×10^{-6}	3.102
rs10497723	2	192817829	*TMEFF2*	A/G	0.195	0.095	8.45×10^{-6}	0.431
rs1498024	4	125912629	*SPRY1/ANKRD50*	C/T	0.844	0.735	1.68×10^{-5}	0.513
rs6136651	20	19144096	*SLC24A3*	A/G	0.834	0.725	1.81×10^{-5}	0.525
rs757705	7	8313535	*ICA1/NXPH1*	A/G	0.498	0.63	2.79×10^{-5}	1.713
rs861844	22	20330773	*SDF2L1*	G/T	0.086	0.172	3.21×10^{-5}	2.214
rs10518442	4	125951873	*ANKRD50*	A/C	0.861	0.761	3.22×10^{-5}	0.516
rs10492202	12	66628005	*DYRK2/IFNG*	C/T	0.11	0.203	3.48×10^{-5}	2.054
rs1159006	10	100396273	*HPSE2*	C/T	0.266	0.39	4.01×10^{-5}	1.765
rs10512332	9	104554018	*OR13D1*	C/T	0.789	0.678	4.60×10^{-5}	0.564
rs1536558	10	97222258	*SORBS1*	G/T	0.261	0.155	4.77×10^{-5}	0.519
rs1941011	18	5649096	*EPB41L3*	A/T	0.675	0.551	4.96×10^{-5}	0.590
rs458685	21	30099382	*GRIK1*	A/G	0.105	0.194	6.54×10^{-5}	2.048
rs2831605	21	28467064		A/G	0.945	0.874	6.82×10^{-5}	0.405

*Affymetrix NetAffx annotation; †allele 2 vs. 1. Chr., chomosome.

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

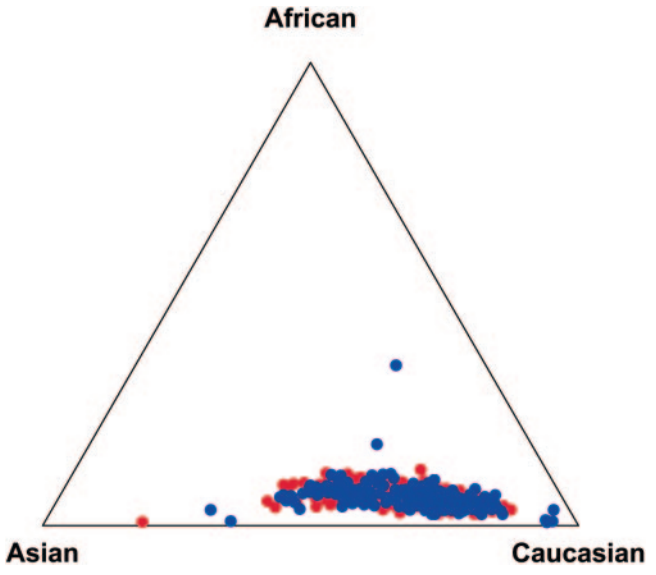

FIG. 3. Estimates of ancestry proportions in case and random control subjects. The genome-wide average ancestral proportions of ancestry for each individual is plotted in a triangular matrix in which each point of the triangle represents 100% ancestry for the indicated ancestral population. The average proportion of each ancestral population in the Starr County Mexican Americans is listed after each ancestral population. Red = case subjects; blue = random control subjects.

were consistent with the previous estimates from classical markers. More importantly, for the purposes here, estimates of the proportion of African, Asian, and European ancestry for the case and random control subjects were indistinguishable from each other (Fig. 3). Formal comparisons by Q-Q plots show no significant differences in the case and random control distributions of ancestry proportions (online appendix Fig. 5), suggesting that spurious associations due to different ancestries of the case and random control subjects are unlikely.

We used these POA estimates as covariates in logistic regressions between type 2 diabetes status and genotype. The POA estimates indicate that *1*) there is very little difference from one individual to the next in the African POA and *2*) the difference in POA estimates per individual lie along an Asian versus European axis of variation. This suggests that the POA variation could be efficiently captured by using the European or Asian POA as a covariate, and we chose to use the former. Including the CEU group covariate had little impact on the association results. The *P* value for nearly all SNP × genotype regressions increased or decreased by less than one-half an order of magnitude (online appendix Fig. 6). We did not observe any highly significant regressions disappearing after including the CEU group POA covariate, again suggesting that spurious associations due to different ancestries of the case and random samples are highly unlikely. Instead, the difference in the regression *P* values distributions was skewed toward increased significance when using the European POA as a covariate.

Verification. Before genotyping any SNP in a larger collection of individuals for replication, we wanted to first verify the association in the same set of individuals using a different genotyping platform (TaqMan). To identify the most robust SNPs for verification genotyping, we selected a subset of 10 SNPs from the 50 highly associated SNPs (Table 2) that met our quality-control criteria (HWE departure $P > 0.001$ in random subjects, call rates ≥0.85, and MAF ≥0.05) using both the DM and GEL algorithms. All SNPs remained significant at a $P < 0.01$ (8/11 $P \leq 10^{-3}$, 4/11 $P \leq 10^{-4}$, and 1/11 $P \leq 10^{-5}$), with the exception of rs861844 (near *SDF2L*), which dropped to $P = 0.02$ (online appendix Table 6). Since the overall genotyping concordance between the genotyping platforms was 99.2% (per-marker range 98.6–99.8%), the decrease in allelic association significance is not a function of differential genotyping but rather the increase in sample size.

In silico replications in other 100K type 2 diabetes GWASs. A total of 120 SNPs (online appendix Table 4) associated in the Mexican-American subjects ($P < 0.01$) had the same allele associated ($P < 0.05$) in one of the other 100K GWASs (27,28,29). At the more stringent $P < 0.001$ level (Table 3), six were replicated in the Amish, three were replicated in the Pima Indians (all by case-control tests), and four were replicated in the FHS (one by generalized estimating equations alone and three by family-based association test alone). These included SNPs in or near the following genes: *RALGPS2* and *ANGPTL1* (chromosome 1); *LCORL*, *NCAPG*, and *CSN3* (chromosome 4); *HTR4* and *ADRB2* (chromosome 5); *UTRN* (chromosome 6); *LINGO2* (chromosome 9); *EGR2* (chromosome 10); *UBQLNL* and *OR52H1* (chromosome 11); and *RORA* (chromosome 15). Of these, one was replicated in multiple studies: rs979752*T ($P = 0.0012$; odds ratio [OR] 0.562) near *UBQLNL* and *OR52H1* in the Amish ($P = 0.03$; 0.764) and FHS ($P = 0.04$; hazard rate ratio 0.709). Additionally, two nonredundant SNPs ($r^2 < 0.8$) in or near *RALGPS2* were independently replicated in the Amish (rs2773080*G; $P = 0.00080$ and OR 0.628 in Mexican Americans; $P = 0.033$ and OR 0.793 in Amish) and Pima Indians (rs3922812*G; $P = 0.00088$ and OR 1.523 in Mexican Americans; $P = 0.028$ and OR 1.311 in Pima Indian case-control subjects).

Replication in non-100K type 2 diabetes GWASs. We also observed 31 SNPs associated ($P < 0.01$) with type 2 diabetes in the Mexican Americans in high LD in either the HapMap Europeans or Asians, also showing evidence for association with type 2 diabetes in a GWAS ($P < 0.05$) in a Scandinavian cohort (DGI; online appendix Table 7). Four of these are significant in the Mexican Americans at a more stringent $P < 0.001$ level and are located in or near *ACTN2* on chromosome 1, *GDNF* and *EGFLAM* on chromosome 5, *EGR2* on chromosome 10, and a nongenic region on chromosome 11 (Table 4).

Replication in more than one other GWAS. We investigated the intersection of the in silico replications in the other GWAS examined and found that six SNPs associated in Mexican Americans ($P < 0.01$) replicated in multiple studies ($P < 0.05$). SNPs in or near *GYPC* (chromosome 2), *EGR2* (chromosome 10), and a nongenic region (chromosome 18) replicated in the Pima Indians and DGI, *DBC1* (chromosome 9) in the Pima Indians and FHS, and *PHLDB1* (chromosome 11) in the Amish and Pima Indians (Table 5). rs10504319*T in or near *MGC34646* and *CHD7* was found to decrease risk in the three (Amish, Pima Indians, and DGI) of four comparative cohorts as well as the Mexican Americans. An additional region on chromosome 11 contains two redundant SNPs ($r^2 > 0.8$) that show evidence for replication: rs979752 in or near *UBQLNL* and *OR52H1* is replicated in the Amish and FHS and nearby rs10500641 is replicated in the DGI study. This is in addition to the multiple *RALGPS2* replications discussed above.

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

TABLE 3
SNPs associated with type 2 diabetes in Mexican Americans ($P < 0.001$) with replication in at least one other 100K GWAS

dbSNP rs	Chr.	Position*	Gene*	Allele 1/2	Mexican Americans (P value)†	Mexican Americans (OR)†	Amish (P value)‡	Amish (OR)‡	Pima Indians (CC P)§	Pima Indians (CC OR)§	Pima Indian sibs (P)‖	Pima Indian sibs (OR)‖	FHS GEE (P)¶	FHS FBAT (P)#	FHS HRR**
rs979752	11	5495715	*UBQLNL/OR52H1*	C/T	0.00012	0.562	0.030	0.764					0.040	0.658	0.709
rs7164773	15	58855240	*RORA*	C/T	0.00016	1.639	0.021	1.300							
rs981864	9	28293495	*LINGO2*	G/T	0.00024	2.138								0.020	1.035
rs10498761	6	45850484		A/C	0.00036	1.675	0.041	1.290							
rs1517645	2	223998255		C/G	0.00072	0.475								0.005	0.766
rs3775745	4	71293834	*CSN3*	A/C	0.00078	0.636	0.002	0.698							
rs1833714	5	148035093	*HTR4/ADRB2*	C/T	0.00078	0.643								0.026	0.713
rs2773080	1	175428407	*RALGPS2*	A/G	0.00080	0.628	0.033	0.793							
rs1509957	10	64280724	*EGR2*	A/G	0.00084	0.652			0.032	0.721					
rs3922812	1	175391590	*RALGPS2/ANGPTL1*	A/G	0.00088	1.523			0.028	1.311					
rs10516322	4	18099836	*LCORL/NCAPG*	C/G	0.00089	0.656	0.050	0.806							
rs6929370	6	145395271	*UTRN*	C/T	0.00095	1.533			0.033	3.988					

*Affymetrix NetAffx annotation; †allele 2 vs. 1; ‡genotype 22 vs. 12; §case-control tests, allele 2 vs. 1; ‖within-family tests, allele 2 vs. 1; ¶generalized estimating equations (GEEs), allele 2 vs. 1; #family-based association test (FBAT), allele 2 vs. 1; **hazard rate ratio (HRR), allele 2 vs. 1. Chr., chromosome.

TABLE 4
Replication of Mexican-American SNPs ($P < 0.001$) with non-100K Type 2 Diabetes GWAS results

Mexican American (dbSNP rs)	Mexican American (chromosomes)	Mexican American (position)*	Mexican American (gene)†	Mexican American (1/2)	Mexican American (P)‡	Mexican American (OR)‡	DGI (dbSNP rs)	DGI (allele)	DGI (chromosomes)	DGI (position)*	DGI (P)	DGI (OR)	HapMap (r^2)	
													CEU	ASN
rs819639	1	233219640	*ACTN2*	C/T	0.00022	1.611	rs819639	T	1	233219640	0.01731	1.194	N/A	N/A
rs1509957	10	64280724	*EGR2*	A/G	0.00084	0.652	rs1509957	G	10	64280724	0.00820	0.831	N/A	N/A
rs270568	5	38058636	*GDNF/EGFLAM*	C/T	0.00085	0.448	rs270565	C	5	38061529	0.01440	0.766	1.000	1.000
rs951432	11	111854649		C/T	0.00090	0.639	rs7116022	T	11	111856182	0.02057	0.866	0.964	0.910

*Position in March 2006 University of California Santa Cruz build; †Affymetrix annotations; ‡allele 2 vs. 1. N/A, not applicable (i.e., identical SNP).

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

TABLE 5
SNPs associated with type 2 diabetes in Mexican Americans ($P < 0.01$) with replication in more than one other GWAS

dbSNP rs	Chr.	Position*	Gene*	Allele (1/2)	Mexican American (P)†	Mexican American (OR)†	Amish (P)‡	Amish (OR)‡	Pima Indian case-control (P)§	Pima Indian case-control (OR)§	Pima Indian sibs (P)‖	Pima Indian sibs (OR)‖	FHS GEE (P)¶	FHS FBAT (P)#	FHS HRR**	DGI (rs)	DGI (P)†	DGI (OR)†
rs979752	11	5495715	*UBQLNL/OR52H1*	C/T	0.00012	0.562	0.030	0.764					0.040	0.658	0.709			
rs2773080	1	175428407	*RALGPS2*	A/G	0.00080	0.628	0.033	0.793										
rs1509957	10	64280724	*EGR2*	A/G	0.00084	0.652			0.032	0.721						64280724	0.00820	0.831
rs3922812	1	175391590	*RALGPS2/ANGPTL1*	A/G	0.00088	1.523			0.028	1.311								
rs2191678	9	119433899	*DBC1*	A/G	0.00316	0.678			0.035	0.543				0.009	0.797			
rs2077173	11	117997801	*PHLDB1*	A/G	0.00503	0.703	0.030	0.796			0.035	0.651						
rs360248	2	126731609	*GYPC*	C/G	0.00602	0.676			0.014	0.618						rs360256	0.017	0.881
rs10504319	8	62089326	*MGC34646/CHD7*	C/T	0.00656	0.571	0.010	0.549	0.043	0.726						rs1509957	0.008	0.831
rs10500641	11	5426152		C/T	0.00656	1.522										rs1391619	0.01489	0.746
rs1845386	18	29572254		C/G	0.00901	1.398			0.045	1.487						rs7237379	0.005	1.203

*Affymetrix NetAffx annotation; †allele 2 vs. 1; ‡genotype 22 vs. 12; §case-control tests, allele 2 vs. 1; ‖within-family tests, allele 2 vs. 1; ¶generalized estimating equations (GEEs), allele 2 vs. 1; #family-based association test (FBAT), allele 2 vs. 1; **hazard rate ratio (HRR), allele 2 vs. 1. Chr., chromosome.

DISCUSSION

We have carried out a GWAS of type 2 diabetes in Mexican Americans from Starr County, Texas. We observed a number of allelic associations showing replication in one of the other GWAS, and a limited number of which show multiple lines of evidence for replication. The association signals that appear to be the most robust would be the three that are significant at $P < 10^{-3}$ in the Mexican-American subjects and are replicated ($P < 0.05$) in at least two of four other GWASs interrogated (rs979752 and rs10500641 near *UBQLNL* and *OR52H1* on chromosome 11, rs2773080 and rs3922812 in or near *RALGPS2* on chromosome 1, and rs1509957 near *EGR2* on chromosome 10). These SNPs and many other significantly associated SNPs will be prioritized for further follow-up genotyping in a larger Mexican-American case-random control cohort. Our FDR estimate suggests that if we followed up the 141 associations significant at the $P \leq 10^{-3}$ threshold, a little less than half would not be false-positives. The broad replication of these three signals meeting this significance threshold suggests that they may be true rather than false-positive associations, but confirmation of such will await the results of the follow-up genotyping in the more numerous Mexican-American case-random sample cohort. Even though our most promising SNPs may turn out to be false-positives, it is tempting nonetheless to query whether any of these putative type 2 diabetes susceptibility genes identified in this GWAS have supporting biological evidence for their candidacy as type 2 diabetes genes. Of the genes implicated and discussed above, no direct links to a diabetes-related phenotype were found.

Given the large amount of data generated in a GWAS, one might naively think that this study represents a comprehensive interrogation of the human genome for type 2 diabetes susceptibility genes, but this is simply not true (39). Although the mean intermarker distance for the 116,204 SNPs genotyped on the Affymetrix GeneChip Human Mapping 100K set is only 8.5 kb, the 100K platform does not completely cover the genome given the patterns of LD and uneven SNP density (40). Nowhere is this more evident than searching for associations at previously identified and replicated type 2 diabetes genes. For example, the Mexican-American population under investigation here is the same in which *CAPN10* was identified through positional cloning studies subsequent to genome linkage scans. However, the nearest SNPs to *CAPN10* on the 100K platform are 187 and 250 kb in either direction, well beyond the LD block in which *CAPN10* resides. The results for other "known" type 2 diabetes genes in our study are presented in online appendix Table 8. Like *CAPN10*, there are no SNPs on the Affymetrix GeneChip Human Mapping 100K set near *HNF4A* or *KCNJ11* and *HHEX*. The previously identified type 2 diabetes–associated variant (rs1801282) in *PPARG* is included on the 100K set but is not associated with type 2 diabetes in Mexican Americans ($P = 1.0$). For *TCF7L2*, the SNP (rs7100927) in highest LD ($r^2 = 0.5$) with the previously identified type 2 diabetes–associated variant (rs7903146) also shows no significant associations to type 2 diabetes in the Mexican-American subjects ($P = 0.952$). The SNPs in or near two genes (*IGF2BP2* and *SLC30A8*), previously identified in other GWASs as containing type 2 diabetes risk alleles, show no evidence of association and have modest LD between the previously associated variant and the SNPs on the 100K platform. However, we did observe significant

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

TABLE 6
Comparison of four GWA studies

	FHS	Mexican American	Pima C/C	Pima family	Amish
n	1,087	281 case/280 control	300 case/334 control	172 sibships	124 case/295 control
Type	Population, family-based	Case/control	Case/control	Family	Case/control
Ethnicity	Non-Hispanic white	Mexican American	American Indian	American Indian	Non-Hispanic white
Male/female	527/560				
Case		108/173	114/186	48/92	41/83
Control		69/211	160/174	57/64	153/142
Mean age (years)	51.5 ± 9.8				
Case		57.7 ± 10.8	19.2 ± 4.5	41.0 ± 8.3	51.3 ± 10.5
Control		—	55.5 ± 9.8	27.8 ± 7.9	64.4 ± 12.9
Mean BMI (kg/m^2)	27.5 ± 5.2				
Case		31.5 ± 6.2	38.9 ± 8.4	38.9 ± 9.3	29.3 ± 5.8
Control		—	35.4 ± 8.0	36.9 ± 8.9	27.4 ± 4.7
Quantitative glycemic traits analyzed	FPG, tFPG, A1C, FI, HOMA-IR, Gutt ISI 0_120				FASTG, glucose AUC, insulin AUC, HOMA-IR, insulin secretion index
Notes	91 cases of incident diabetes	Controlled for admixture	Age of onset <25 years	Sibships overlap with C/C	Cases have (+) FH; active lifestyle
SNPs analyzed	66,543	88,702	80,044	80,044	82,485
SNPs failed	39,205 MAF <10%; 4,064 HWE $P < 0.001$; 10,438 call rate <90%	19,032 MAF <5%; 1,562 HWE $P < 0.001$; 4,818 call rate <85%	28,215 MAF <1%; 2,429 HWE $P < 0.001$; 5,122 call rate <85% and/or error rate >3% in duplicate samples	28,215 MAF <1%; 2,429 HWE $P < 0.001$; 5,122 call rate <85% and/or error rate >3% in duplicate samples	26,816 MAF <5%; 2,573 HWE $P < 0.001$; 1,866 call rate <90%
P value thresholds	Strategy 1: $P < 0.01$ in all three glucose traits (FPG, tFPG, and A1C) or in all 3 insulin traits (FI, HOMA-IR, and Gutt) or two glucose and two insulin traits or incident type 2 diabetes; Strategy 2: $P < 0.001$	Primary analysis: $P < 0.01$ for type 2 diabetes; In silico replication: $P < 0.05$	In silico replication: $P < 0.007$ in combined within-family and case-control analysis (weighted to give priority to within-family test); Additional genotyping: $P < 0.001$ in combined within-family and case-control analysis as above	In silico replication: $P < 0.007$ in combined within-family and case-control analysis (weighted to give priority to within-family test); Additional genotyping: $P < 0.001$ in combined within-family and case-control analysis as above	Primary analysis: $P < 0.01$ for type 2 diabetes; Internal replication: type 2 diabetes $P < 0.01$ and one glucose trait (FASTG or GAUC) or one insulin trait (ISI, IAUC, or HOMA-IR) $P < 0.01$
Replication cohort	1,465 unrelated FHS participants (non-overlapping)	760 overlapping individuals for SNP verification	Non-overlapping Pima Indians: 1,207 case/1,627 control	Non-overlapping Pima Indians: 1,207 case/1,627 control	427 nondiabetic Amish participants (295 control from primary type 2 analysis)
In silico replication	T2D 100K Consortium + DGI 500K	T2D 100K Consortium + DGI 500K	T2D 100K Consortium + DGI 500K	T2D 100K Consortium + DGI 500K	T2D 100K Consortium + DGI 500K

T2D, type 2 Diabetes.

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

TABLE 7
Replication evidence across four GWA studies

Study	SNP	Chr.	Position (hg17)	Gene	Gene name	OMIM	Initial finding
FHS	rs952635	1p31.2	66403906	*PDE4B*	cAMP-specific phosphodiesterase	600127	G-allele protective for type 2 diabetes (HR 0.56 [0.40–0.79], Cox $P = 0.0007$)
FHS	rs2863389	3q26.1	167631594	200 kb from nearest gene		N/A	Minor T-allele at SNP rs2863389 protective against diabetes (HR 0.41 [0.25–0.69], Cox $P = 0.0006$)
FHS	rs7935082	11q12.2	59911576	*MS4A7*	Membrane-spanning 4-domains subfamily A member 7	606502	T-allele associated with lower FPG in FHS (FBAT $P = 0.0006$)
Amish	rs2237457	7p12.2	50693638	*GRB10*	Growth factor receptor bound protein 10	601523	G-allele protective for type 2 diabetes (OR 0.61, $P = 0.00001$)
Amish	rs3845971	3p14.2	59975712	*FHIT*	Fragile histidine triad gene	601153	T-allele increased type 2 diabetes risk (OR 1.42, $P = 0.004$)
Pima	rs10500938	11p14.3	22601179	*FANCF*	Fanconia anemia, complementation group F	603467	A-allele increased type 2 diabetes risk (OR 2.14, $P = 0.0004$)
Pima	rs686989	11q23.1	113544435	*ZTBT16*	Zinc finger and BTB domain containing-16	176797	A-allele increased type 2 diabetes risk (OR 3.26, $P = 0.0004$)
MA	rs979752	11p15	4326380	*UBQLNL*	Ubiquilin3	605473	T-allele protective for type 2 diabetes (OR 0.562, $P = 0.00012$)
MA	rs2773080	1q25	176963373	*RALGPS2*	Ral GEF with PH domain and SH3 binding motif 2	611154	G-allele protective for type 2 diabetes (OR 0.628, $P = 0.0008$)
MA	rs1509957	10q21.1	64280724	*EGR2*	Early growth response 2	129010	G-allele protective for type 2 diabetes (OR 0.652, $P = 0.00084$)

Chr., chromosome.

Continued on following page

associations with SNPs in *CDKAL1* and *CDKN2A* ($P < 0.01$) but only at SNPs not in LD with the originally associated SNP, so these could not be considered direct replication of the original signal but may point to other variation contributing to risk of type 2 diabetes in Mexican Americans.

The lack of difference in the POA estimates between the case and random samples speaks not just to a reduced likelihood of spurious associations due to substructure but also to a larger issue. Given the high prevalence of type 2 diabetes among Native Americans, it has been previously hypothesized that the high prevalence of type 2 diabetes in Mexican Americans may be due to their Native American ancestry (41,42). In support of this hypothesis is our estimate that ~30% of the contemporary Mexican-American gene pool is Native American derived. Given the prevalence of diabetes among Native Americans, the predicted prevalence in Mexican Americans parallels that expected based on this degree of admixture (43). However, if type 2 diabetes in Mexican Americans was largely Native American derived, a higher proportion of Asian (proxy for Native American) ancestry would have been observed in the case subjects than in the random control subjects; we did not observe this.

The POAs are genome-wide estimates. We assume these may be highly variable from one genomic region to the next, so it remains possible that for any given gene associated with type 2 diabetes in Mexican Americans, it is the Native American–derived variant that is the risk allele. We also noted that the difference in the distributions of the regression *P* values was skewed toward increased significance when using the European POA as a covariate. This suggests that we may be able to exploit this when admixture mapping methods are used in the future.

In conclusion, we observed many SNPs associated with type 2 diabetes, some of which were replicated in at least one of four other GWASs we queried. This study represents our initial examination of the Mexican-American 100K GWAS data; more sophisticated approaches will follow, including a meta-analysis of four 100K GWASs. It may also be that subsequent investigations of this GWAS

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

TABLE 7
Continued

QT	Internal replication	External replication 100K T2D Consortium	External replication 500K DGI
Same protective allele associated with higher Gutt ISI and lower FPG, mFPG, A1C, and HOMA-IR (GEE $P \sim 0.003$)	FPG and HOMA-IR ($P < 0.05$)	No	rs6664618 in moderate LD ($r^2 = 0.6$); consistent trends in FPG ($P = 0.04$) and HOMA-IR ($P = 0.057$)
Same protective allele associated with lower FPG and mFPG in the FHS sample (GEE $P =$ 0.005 and 0.0005, respectively)	No	T-allele protects 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 Pimas	rs9829442 in perfect LD ($r^2 = 1.0$); minor T-allele shows opposite nominal trend in increasing risk of type 2 diabetes (OR 1.15, $P = 0.059$)
Same	No	T-allele was nominally protective from diabetes in the Mexican Americans (OR 0.53, $P = 0.049$) and in the Pimas (OR 0.58, $P = 0.009$)	rs950803 in perfect LD ($r^2 = 1.0$); minor T-allele shows consistent nominal trend in protection from type 2 diabetes (OR 0.89, $P = 0.097$)
Same protective allele associated with lower GAUC ($P = 0.001$)	No	No	Not with rs2237457; six additional *GRB10* SNPs in LD (r^2 0.16–0.78) are associated with type 2 diabetes ($P < 0.05$)
Same risk allele associated with increased GAUC ($P = 0.0004$)	No	Mexican-American type 2 diabetes OR 1.46 ($P = 0.004$)	No
	A-allele OR=1.65 (p=0.0010)	No	(A-allele absent in CEU)
	A-allele OR 1.27 ($P = 0.0333$)	No	No
N/A	N/A	Amish type 2 diabetes OR 0.764 ($P = 0.030$) and FHS HRR 0.709 ($P = 0.040$)	No
N/A	N/A	Amish type 2 diabetes OR 0.793 ($P = 0.033$)	No
N/A	N/A	Pima type 2 diabetes OR 0.721 (P=0.032)	Same SNP OR 0.831 ($P = 0.0082$)

with haplotypes or genes as the unit of investigation, rather than SNPs, will prove to be more informative. Nonetheless, we have highlighted several interesting putative type 2 diabetes genes for follow-up in the hopes that it may further elucidate the etiology of type 2 diabetes and identify new avenues for both the treatment and prevention of this complex disease.

ACKNOWLEDGMENTS

This study was supported in part by U.S. Public Health Service Grants DK-20595, DK-47486, DK-47487, DK-55889, and HL-84715 and a gift from the Kovler Family Foundation. M.G.H. was supported by a mentor-based fellowship from the American Diabetes Association.

We thank Laura Martinolich, Xinmin Li, Edwin Cook, and Carole Ober for providing technical assistance with the Affymetrix genotyping assays. We also thank the 100K Type 2 Diabetes Consortium members and authors of the three other 100K GWAS studies (27,28,29) for valuable discussion and comments regarding the data, analysis, and manuscript preparation.

REFERENCES

1. Centers for Disease Control and Prevention: *Chronic Disease Prevention: Preventing Diabetes and Its Complications.* Atlanta, GA, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2006
2. Flegal KM, Ezzati TM, Harris MI, Haynes SG, Juarez RZ, Knowler WC, Perez-Stable EJ, Stern MP: Prevalence of diabetes in Mexican Americans, Cubans, and Puerto Ricans from the Hispanic Health and Nutrition Examination Survey, 1982–1984. *Diabetes Care* 14:628–638, 1991
3. Hamman RF, Marshall JA, Baxter J, Kahn LB, Mayer EJ, Orleans M, Murphy JR, Lezotte DC: Methods and prevalence of non-insulin-dependent diabetes mellitus in a biethnic Colorado population: the San Luis Valley Diabetes Study. *Am J Epidemiol* 129:295–311, 1989
4. Hanis CL, Ferrell RE, Barton SA, Aguilar L, Garza-Ibarra A, Tulloch BR, Garcia CA, Schull WJ: Diabetes among Mexican Americans in Starr County, Texas. *Am J Epidemiol* 118:659–672, 1983
5. Samet JM, Coultas DB, Howard CA, Skipper BJ, Hanis CL: Diabetes, gallbladder disease, obesity, and hypertension among Hispanics in New Mexico. *Am J Epidemiol* 128:1302–1311, 1988
6. Hanis CL, Boerwinkle E, Chakraborty R, Ellsworth DL, Concannon P, Stirling B, Morrison VA, Wapelhorst B, Spielman RS, Gogolin-Ewens KJ, Shepard JM, Williams SR, Risch N, Hinds D, Iwasaki N, Ogata M, Omori Y, Petzold C, Rietzch H, Schroder HE, Schulze J, Cox NJ, Menzel S, Boriraj VV, Chen X, Lim LR, Lindner T, Mereu LE, Wang YQ, Xiang K, Yamagata K,

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

Yang Y, Bell GI: A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat Genet* 13:161–166, 1996
7. Genetics of Diabetes—Part I. *Diabetes Reviews* 5:105–174, 1997
8. Genetics of Diabetes—Part II. *Diabetes Reviews* 5:175–291, 1997
9. Hanis C: Genetics of non-insulin-dependent diabetes mellitus among Mexican Americans: approaches and perspectives. In *Genetic Approaches to Noncommunicable Diseases*. Berg K, Boulyjenkov V, Christen Y, Eds. Berlin, Springer-Verlag, 1996
10. Das SK, Elbein SC: The genetic basis of type 2 diabetes. *Cell Sci* 2:100–131, 2006
11. McIntyre EA, Walker M: Genetics of type 2 diabetes and insulin resistance: knowledge from human studies. *Clin Endocrinol (Oxf)* 57:303–311, 2002
12. Elbers CC, Onland-Moret NC, Franke L, Niehoff AG, van der Schouw YT, Wijmenga C: A strategy to search for common obesity and type 2 diabetes genes. *Trends Endocrinol Metab* 18:19–26, 2007
13. McCarthy MI: Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* 3:159–167, 2003
14. Willer CJ, Bonnycastle LL, Conneely KN, Duren WL, Jackson AU, Scott LJ, Narisu N, Chines PS, Skol A, Stringham HM, Petrie J, Erdos MR, Swift AJ, Enloe ST, Sprau AG, Smith E, Tong M, Doheny KF, Pugh EW, Watanabe RM, Buchanan TA, Valle TT, Bergman RN, Tuomilehto J, Mohlke KL, Collins FS, Boehnke M: Screening of 134 single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes replicates association with 12 SNPs in nine genes. *Diabetes* 56:256–264, 2007
15. 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* 445:881–885, 2007
16. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, 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 AS, the Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341, 2007
17. 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
18. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, 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
19. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336, 2007
20. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163–175, 2000
21. Love-Gregory LD, Wasson J, Ma J, Jin CH, Glaser B, Suarez BK, Permutt MA: A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 α gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an Ashkenazi Jewish population. *Diabetes* 53:1134–1140, 2004
22. Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS: Genetic variation near the hepatocyte nuclear factor-4 α gene predicts susceptibility to type 2 diabetes. *Diabetes* 53:1141–1149, 2004
23. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES: The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80, 2000
24. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S, McCarthy MI, Hattersley AT, Frayling TM: Large-scale association studies of variants in genes encoding the pancreatic β-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 52:568–572, 2003
25. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K: Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38:320–323, 2006
26. Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 273:1516–1517, 1996
27. Florez JC, Manning AK, Dupuis J, McAteer J, Irenze K, Gianniny L, Mirel DB, Fox CS, Cupples LA, Meigs JB: A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. *Diabetes* 56:3063–3074, 2007
28. Hanson RL, Bogardus C, Duggan D, Kobes S, Knowlton M, Infante AM, Marovich L, Benitez D, 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
29. Rampersaud E, Damcott CM, Fu M, Shen H, McArdle P, Shi X, Shelton J, Yin J, Chang CY, Ott SH, Zhang L, Zhao Y, Mitchell BD, O'Connell J, 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
30. National Diabetes Data Group: Classification and diagnosis of diabetes and other categories of glucose intolerance. *Diabetes* 28:1039–1057, 1979
31. Nicolae DL, Wu X, Miyake K, Cox NJ: GEL: a novel genotype calling algorithm using empirical likelihood. *Bioinformatics* 22:1942–1947, 2006
32. Rabbee N, Speed TP: A genotype calling algorithm for Affymetrix SNP arrays. *Bioinformatics* 22:7–12, 2006
33. Affymetrix: *BRLMM: An Improved Genotype Calling Method for the GeneChip Human Mapping 500K*. Santa Clara, CA, Affymetrix, Inc., 2006
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, de Bakker P, Daly M, Sham P: PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–575, 2007
35. Abecasis GR, Cookson WO: GOLD: graphical overview of linkage disequilibrium. *Bioinformatics* 16:182–183, 2000
36. Purcell S, Cherny SS, Sham PC: Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–150, 2003
37. Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 155:945–959, 2000
38. Cerda-Flores RM, Kshatriya GK, Bertin TK, Hewett-Emmett D, Hanis CL, Chakraborty R: Gene diversity and estimation of genetic admixture among Mexican-Americans of Starr County, Texas. *Ann Hum Biol* 19:347–360, 1992
39. Pe'er I, de Bakker PI, 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
40. Nicolae DL, Wen X, Voight BF, Cox NJ: Coverage and characteristics of the Affymetrix GeneChip Human Mapping 100K SNP set. *PLoS Genet* 2:e67, 2006
41. Gardner LI Jr, Stern MP, Haffner SM, Gaskill SP, Hazuda HP, Relethford

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024

JH, Eifler CW: Prevalence of diabetes in Mexican Americans: relationship to percent of gene pool derived from native American sources. *Diabetes* 33:86–92, 1984
42. Lorenzo C, Serrano-Rios M, Martinez-Larrad MT, Gabriel R, Williams K, Gonzalez-Villalpando C, Stern MP, Hazuda HP, Haffner SM: Was the historic contribution of Spain to the Mexican gene pool partially responsible for the higher prevalence of type 2 diabetes in Mexican-origin populations? The Spanish Insulin Resistance Study Group, the San Antonio Heart Study, and the Mexico City Diabetes Study. *Diabetes Care* 24:2059–2064, 2001
43. Hanis CL, Hewett-Emmett D, Bertin TK, Schull WJ: Origins of U.S. Hispanics: implications for diabetes. *Diabetes Care* 14:618–627, 1991

Downloaded from http://diabetes.diabetesjournals.org/diabetes/article-pdf/56/12/3033/386976/zdb01207003033.pdf by guest on 25 April 2024