Original Article

SORCS1: A Novel Human Type 2 Diabetes Susceptibility Gene Suggested by the Mouse

Mark O. Goodarzi,[1,2] Donna M. Lehman,[3] Kent D. Taylor,[2] Xiuqing Guo,[2] Jinrui Cui,[2] Manuel J. Quiñones,[4] Susanne M. Clee,[5] Brian S. Yandell,[5] John Blangero,[6] Willa A. Hsueh,[4] Alan D. Attie,[5] Michael P. Stern,[3] and Jerome I. Rotter[2]

OBJECTIVE—A small number of susceptibility genes for human type 2 diabetes have been identified by candidate gene analysis or positional cloning. Genes found to influence diabetes or related traits in mice are likely to be susceptibility genes in humans. *SorCS1* is the gene identified as responsible for the mouse chromosome 19 *T2dm2* quantitative trait locus for fasting insulin levels, acting via impaired insulin secretion and increased islet disruption in obese females. Genes that impair compensatory insulin secretion in response to obesity-induced insulin resistance may be particularly relevant to human diabetes. Thus, we sought to determine whether variation in the human SORCS1 gene was associated with diabetes-related traits.

RESEARCH DESIGN AND METHODS—We assessed the contribution of variation in SORCS1 to human insulin–related traits in two distinct Mexican-American cohorts. One cohort (the Mexican-American Coronary Artery Disease [MACAD] cohort) consisted of nondiabetic individuals, allowing assessment of genetic association with subclinical intermediate insulin-related traits; the second cohort (the San Antonio Family Diabetes Study [SAFADS]) contained individuals with diabetes, allowing association analyses with overt disease.

RESULTS—We first found association of SORCS1 single nucleotide polymorphisms and haplotypes with fasting insulin levels and insulin secretion in the MACAD cohort. Similar to our results in the mice, the genetic association was strongest in overweight women. We then observed association with diabetes risk and age at diagnosis in women of the SAFADS cohort.

CONCLUSIONS—Identification of SORCS1 as a novel gene affecting insulin secretion and diabetes risk is likely to provide important insight into the biology of obesity-induced type 2 diabetes. ***Diabetes*** **56:1922–1929, 2007**

Obesity and type 2 diabetes have reached epidemic proportions in Western societies. Numerous twin and familial aggregation studies have established an inherited basis for type 2 diabetes susceptibility (1). Variants in a small number of genes have been identified that contribute to diabetes risk. A number of these were examined based on their known function (the candidate gene approach). Examples of such candidate genes include peroxisome proliferator–activated receptor γ (*PPARG*) and the pancreatic β-cell inwardly rectifying potassium channel Kir6.2 (*KCNJ11*) (2–5). Another approach to diabetes gene discovery is the whole-genome approach, using linkage or association to interrogate the entire genome. A benefit of this approach is that it does not require knowledge of the underlying pathophysiology of the disease. This approach led to the discovery of calpain-10 (*CAPN10*) and *TCF7L2* as genes that contribute to diabetes risk (6,7). Each of the validated diabetes genes contributes modestly to overall risk of developing diabetes, consistent with the concept that diabetes is a complex genetic condition wherein multiple genes contribute moderately to overall risk (8).

A relatively unexploited additional approach to diabetes gene discovery is to translate findings from murine genetics to humans. Genes that contribute to diabetes risk in mice may also contribute to this disease in humans. Genetic heterogeneity underlying the phenotype, the segregation of other unknown modifier loci, and a highly variable environment makes the study of diabetes genetics difficult in humans. Genetic analysis in model organisms enables the study of genetically homogeneous strains in a controlled environment. Inbred mouse strains replicate much of the genetic diversity of the human population, and it is likely that variation at many of the same gene loci are responsible for diabetes susceptibility in both organisms (9). The genetic component of complex traits can then be dissected through the creation of congenic strains wherein a chromosomal segment containing a single modifier locus is fixed within a defined genetic background, increasing the sensitivity to detect relatively small independent effects of these loci (10).

We have focused on obesity-induced type 2 diabetes in the mouse as a model of human type 2 diabetes. Obesity leads to insulin resistance; however, most obese individu-

From the [1]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California; the [2]Departments of Pediatrics and Medicine, Medical Genetics Institute, Steven Spielberg Pediatric Research Center, Cedars-Sinai Medical Center, Los Angeles, California; the [3]Department of Medicine/Clinical Epidemiology, University of Texas Health Science Center, San Antonio, Texas; the [4]Division of Endocrinology, Diabetes, and Hypertension, Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California; the [5]Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin; and the [6]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas.

Address correspondence and reprint requests to Mark O. Goodarzi, MD, PhD, Cedars-Sinai Medical Center, Division of Endocrinology, Diabetes, and Metabolism, 8700 Beverly Blvd., Becker B-131, Los Angeles, CA 90048. E-mail: mark.goodarzi@cshs.org.

Received for publication 30 November 2006 and accepted in revised form 3 April 2007.

Published ahead of print at http://diabetes.diabetesjournals.org on 10 April 2007. DOI: 10.2337/db06-1677.

Additional information for this article can be found in an online appendix at http://dx.doi.org/10.2337/db06-1677.

CAD, coronary artery disease; LD, linkage disequilibrium; MACAD, Mexican-American Coronary Artery Disease; QTL, quantitative trait locus; SAFADS, San Antonio Family Diabetes Study; SNP, single nucleotide polymorphism.

als do not develop diabetes. Diabetes is thought to develop in obese insulin-resistant individuals when the β-cells become unable to secrete adequate insulin to overcome the insulin resistance (11,12). Thus, we anticipate that many genes conferring susceptibility to diabetes will turn out to affect the ability of the pancreas to compensate for insulin resistance with an augmented insulin response.

In an F2 intercross between two strains differing in susceptibility to obesity-induced diabetes, we mapped *T2dm2*, a quantitative trait locus (QTL) on mouse chromosome 19 affecting fasting insulin levels (13). Using interval-specific congenic mouse strains, we recently identified *SorCS1* as the gene underlying *T2dm2* (14). Genetic variation in *SorCS1* had its most dramatic effect in obese female mice, where it led to decreased fasting insulin levels via reduced insulin secretion following a glucose challenge and pancreatic islet disruption; insulin sensitivity was not affected.

The region of mouse chromosome 19 harboring the *SorCS1* gene is syntenic with human chromosome 10q. We previously reported linkage between type 2 diabetes and a region on chromosome 10q (15). Linkage and association of type 2 diabetes and diabetes-related traits to this region were subsequently reported in a number of studies (16–28). Indeed, our prior genome-wide scan in Mexican Americans detected a susceptibility locus for type 2 diabetes and age of diabetes onset to a broad region on chromosome 10q, which includes the human SORCS1 gene (15). In the current study, we examined SORCS1 as a candidate gene for human type 2 diabetes. We utilized two Mexican-American family cohorts: one free of diabetes but extensively phenotyped for insulin-related traits and a second cohort with numerous diabetic individuals. This allowed us to assess whether SORCS1 was associated with not only fasting insulin (the trait used to identify this gene in mice), but also with insulin sensitivity or insulin secretion, two processes that contribute to plasma insulin levels and, when deranged, lead to diabetes. Thus, this study had the advantage of the ability to assess association with intermediate traits as well as with risk and age of onset of overt disease.

Similar to our findings in mice, SORCS1 variants were associated with fasting insulin and insulin secretion in humans. SORCS1 variants were also associated with diabetes risk and age of onset of diabetes. This is the first genetic association study that translated a diabetes susceptibility gene identified in mice to humans. This study identifies a novel potential gene for human diabetes and also suggests that the underlying mechanism is altered insulin secretion into the bloodstream, highlighting the importance of factors affecting β-cell compensatory responses to the development of type 2 diabetes.

RESEARCH DESIGN AND METHODS

Associations with fasting insulin, insulin resistance, and insulin secretion were assessed in participants in the Los Angeles Mexican-American Coronary Artery Disease (MACAD) Study, a study of coronary artery disease (CAD) and insulin resistance in Mexican Americans (29,30). In the present report, 102 two-generation Mexican-American families were included. In total, 548 subjects (adult offspring of probands with CAD and the spouses of those offspring) were extensively phenotyped, and 808 subjects (adult offspring along with their spouses and parents) were genotyped. The phenotyped offspring were healthy and free of clinically overt diabetes and CAD; thus, the observed associations represent preclinical manifestations that may predispose to the onset of diabetes.

The hyperinsulinemic-euglycemic clamp (31) was performed as previously described with an insulin infusion rate of 60 mU/m^2 per min (29). The glucose infusion rate (M, given in mg · kg^{-1} · min^{-1}) over the last 30 min of steady-state insulin and glucose concentrations reflects glucose uptake by all tissues of the body (primarily insulin-mediated glucose uptake in muscle) and is therefore a direct physiologic measurement of tissue insulin sensitivity (31).

The oral glucose tolerance test consisted of baseline glucose and insulin measurements, followed by administration of 75 g oral glucose with blood draws at 30, 60, 90, 120, and 180 min. Insulin secretion indexes were obtained from glucose and insulin measurements at the oral glucose tolerance test 30-min time point to estimate early insulin secretion (30). We used the insulin-to-glucose ratio (IGR30, μIU · dl · mg^{-1} · ml^{-1}), the 30-min insulin value divided by the 30-min glucose value. Such indexes have correlations of 0.31–0.51 with acute insulin secretion directly measured by frequently sampled intravenous glucose tolerance testing (32).

Associations with diabetes risk were assessed in subjects from the San Antonio Family Diabetes Study (SAFADS) (15). The SAFADS cohort comprises 574 individuals from 32 extended Mexican-American families, ascertained from a proband with type 2 diabetes. The prevalence of diabetes in SAFADS is 20.6% (19.6% in female and 22.1% in male subjects) with a mean age of onset of 48.2 years (48.7 years in female and 47.7 years in male subjects). Female subjects comprise 58.9% of the cohort, and their average BMI was 30.6 ± 0.39 kg/m^2. These families ranged in size from 2 to 53 members, with a median size of 16 individuals per family. Diabetes was defined as a fasting plasma glucose level ≥7.0 mmol/l (126 mg/dl) or a self-report of physician-diagnosed diabetes under treatment with either oral antidiabetes agents or insulin.

The institutional review boards of the University of California Los Angeles, Cedars-Sinai Medical Center, and the University of Texas Health Science Center at San Antonio approved all procedures, and all subjects gave informed consent.

In both studies, peripheral blood lymphocytes were transformed and immortalized for all participants to establish a perpetual source of genomic DNA.

Genotyping. In the initial screen, seven single nucleotide polymorphisms (SNPs) within SORCS1 were selected using information from the HapMap Project (33), such that each SNP would *1*) have a high frequency of the minor allele in the Caucasian population, *2*) represent one region of high linkage disequilibrium (LD) (haplotype block) as measured by D′ (http://www.hapmap.org), and thus *3*) contribute to testing the hypothesis that the human genomic region syntenic to the *T2dm2* linkage peak was associated with fasting insulin (Table 1 and Fig. 1). Three SNPs flanking this region (two in the neighboring SORCS3 gene and one ~330 kb telomeric to SORCS1) were also genotyped in the initial screen (Table 1).

Next, 13 additional SNPs were selected to refine the haplotype structure of the syntenic region (Table 1). Genotyping was performed using TaqMan MGB technology as previously described (29,34). PCR primers and TaqMan MGB probes for the SNPs genotyped are listed in online appendix Tables 1 and 2 (available at http://dx.doi.org/10.2337/db06-1677).

Statistical analysis. Methods of analyses in the two human cohorts were selected based on the characteristics of the different populations. The MACAD cohort is composed of small families and marrying-in spouses; therefore, the generalized estimating equation (35) approach was selected to utilize data from all phenotyped subjects. We evaluated association using this robust variance estimation approach to test hypothesized associations between phenotypes and genotypes while accounting for familial correlations present in the data. The PROC GENMOD procedure in SAS (version 9.0; SAS Institute, Cary, NC) was used for the association analysis in which a sandwich estimator was used, assuming exchangeable correlation. Family was taken as the cluster factor, i.e., members from the same family were assumed to be correlated. Both fasting insulin and IGR30 were log transformed to better approximate conditional normality and homogeneity of variance. A dominant genetic model was assumed in all the association analyses. Analyses were conducted using age and sex as covariates, unless otherwise specified. To confirm the associations detected with the original set of 10 screening SNPs (7 in SORCS1 and 3 flanking), the entire sample was permuted by shuffling the fasting insulin phenotype between individuals, and subsequent association analyses were computed 1,000 times to obtain the distribution of the test statistic under the null hypothesis of no association. The empirical P values were obtained as the proportion of the 1,000 replicates that had P values less than or equal to the nominal one that was obtained from the actual (unshuffled) data. The resulting P values were essentially unchanged. As subsequent analyses were confirmatory of the hypotheses generated from these original SNPs, further permutation testing was not performed.

Haploview 3 was used to determine haplotype frequencies and to delineate haplotype blocks (36). Haploview constructs haplotypes by using an accelerated expectation-maximization algorithm. Haploview was used to calculate LD (the D′ statistic) between each pairwise combination of all 20 SNPs used in haplotype block determination. To determine haplotype blocks, Haploview

TABLE 1
Association of individual SNPs with fasting insulin in MACAD

	SNP position	Location in SORCS1	Major/ minor allele	Minor allele frequency	Fasting insulin (pmol/l)		
					1/1	1/2 and 2/2	*P*
SORCS3							
*rs813756	106,761,882	~1.56 Mb cen	T/G	30.2	109.92 ± 3.80 (265)	112.07 ± 3.66 (282)	0.39
*rs1670008	106,792,219	~1.53 Mb cen	G/C	31.1	109.85 ± 3.80 (262)	112.15 ± 2.80 (285)	0.31
SORCS1							
Haplotype block 1							
rs2788677	108,458,473	Intron 7	T/C	46.0	108.27 ± 5.45 (159)	114.44 ± 3.30 (386)	0.14
rs10748924	108,477,170	Intron 6	C/T	35.5	110.50 ± 4.45 (225)	114.08 ± 3.66 (319)	0.33
Haplotype block 2							
rs7067660	108,492,751	Intron 5	T/C	7.1	113.94 ± 3.09 (469)	103.97 ± 7.10 (76)	0.22
***rs10736189**	**108,502,070**	**Intron 5**	**C/G**	**41.9**	**104.47 ± 4.38 (180)**	**114.73 ± 1.22 (352)**	**0.032**
rs10509818	**108,512,646**	**Intron 4**	**G/A**	**24.4**	**107.19 ± 3.59 (316)**	**120.18 ± 4.45 (229)**	**0.0056**
***rs2249022**	**108,532,319**	**Intron 3**	**A/G**	**40.4**	**103.32 ± 4.09 (193)**	**115.30 ± 3.52 (341)**	**0.017**
rs821994	**108,547,162**	**Intron 3**	**C/T**	**37.4**	**104.61 ± 3.95 (211)**	**117.53 ± 3.87 (334)**	**0.041**
Haplotype block 3							
rs822000	**108,550,027**	**Intron 3**	**A/G**	**41.6**	**101.45 ± 3.87 (192)**	**118.75 ± 3.80 (353)**	**0.0038**
*rs2243454	108,567,901	Intron 3	C/T	47.5	116.45 ± 5.60 (150)	108.99 ± 3.01 (387)	0.49
rs1538417	**108,573,589**	**Intron 3**	**G/A**	**21.3**	**108.56 ± 3.37 (329)**	**118.67 ± 4.88 (213)**	**0.050**
***rs4390282**	**108,579,740**	**Intron 2**	**T/G**	**30.1**	**105.40 ± 3.52 (251)**	**116.16 ± 4.02 (285)**	**0.032**
*rs1537919	108,681,751	Intron 2	G/A	34.0	112.58 ± 4.45 (232)	109.99 ± 3.37 (300)	0.59
Haplotype block 4							
rs10748932	108,805,465	Intron 1	A/G	9.5	114.08 ± 3.16 (447)	105.33 ± 6.24 (98)	0.68
rs11193188	108,805,896	Intron 1	G/A	17.5	113.94 ± 3.529 (374)	109.20 ± 4.52 (170)	0.65
*rs11193190	108,808,277	Intron 1	T/C	39.9	106.69 ± 3.66 (223)	114.87 ± 3.80 (310)	0.43
rs7086426	108,837,197	Intron 1	G/A	20.2	112.15 ± 3.59 (346)	113.37 ± 4.66 (199)	0.87
Haplotype block 5							
rs607437	108,870,445	Intron 1	T/C	40.7	116.52 ± 5.09 (178)	110.78 ± 3.44 (463)	0.51
rs685316	108,883,709	Intron 1	G/A	34.3	117.74 ± 4.88 (226)	108.99 ± 3.37 (317)	0.36
*rs7897974	108,886,554	Intron 1	G/A	25.5	108.13 ± 3.23 (289)	114.73 ± 4.52 (243)	0.18
rs1251753	108,888,293	Intron 1	A/G	47.6	116.16 ± 5.96 (146)	111.50 ± 3.16 (396)	0.43
Flanking SNP							
*rs1530248	109,250,514	~330 kb tel	T/C	44.7	112.86 ± 4.81 (174)	110.28 ± 3.23 (364)	0.88

Data are means ± SE (*n*). *P* values were obtained from generalized estimating equation models. 1/1 are homozygotes for the major allele at each SNP, 1/2 are heterozygotes, and 2/2 are homozygotes of the rare allele. SNP locations are the basepair positions in the University of California Santa Cruz Genome Browser hg17 assembly, which is based on the National Center for Biotechnology Information build 35. Significant associations are indicted in bold. *The 10 SNPs used in the initial screen. cen, centromeric to SORCS1; tel, telomeric to SORCS1.

searches for regions of strong LD ($D' > 0.8$) running from one marker to another, wherein the first and last markers in a block are in strong LD with all intermediate markers (solid spine of LD algorithm). Based on the pedigree structures and genotype data of all individuals in each pedigree, haplotypes were constructed as the most likely set (determined by the maximum likelihood method) of fully determined parental haplotypes of the marker loci for each individual in the pedigree, using the simulated annealing algorithm implemented in the program Simwalk2 (37).

To assess the association between the SNPs and the traits diabetes and age of diabetes onset in the SAFADS cohort, which is composed of large pedigrees, a measured genotype approach was used in the framework of variance component-based pedigree analysis, as implemented in SOLAR (38). Random effects accounting for residual polygenic effects were included in the model to account for the relatedness among family members. The model was adjusted for the effects of age, age^2, and BMI. The minor allele counts at the SNP (0, or 1, or 2) were included as a covariate in the analysis. The likelihood of the model where the effect of this covariate was estimated by likelihood maximization was compared with the likelihood of the null model where the effect of the genotypes was fixed at zero. Two times the difference between the natural logarithm of the likelihoods of the two models is distributed asymptotically as a χ^2 statistic with 1 d.f. Since these findings represent confirmation of a prior hypothesis, they were not adjusted for multiple comparisons.

RESULTS

Subjects in the MACAD Study were selected to exclude those with diabetes; however, based on their ethnic background, these individuals bear an enrichment of alleles predisposing to type 2 diabetes and have a high frequency of physiologic abnormalities predisposing to type 2 diabetes. This cohort has been extensively phenotyped for both insulin resistance and insulin secretion, two factors contributing to plasma insulin levels. Thus, this is an ideal population to examine the effects of genetic variation in SORCS1 on these intermediate phenotypes that are deranged in type 2 diabetes.

SORCS1 is a very large gene, encompassing >500 kb of human chromosome 10. Our studies in the mice localized the causative variant(s) to a region of 242 kb spanning part of the promoter and first exons of the gene (14). To first test the hypothesis that variation in the human SORCS1 gene contributes to fasting insulin levels in the MACAD population, we chose to focus on this same region of the gene. We performed an initial screen with seven SNPs within the region syntenic to the maximal extent of the linkage peak underlying *T2dm2* in the mouse (14) and three flanking SNPs. We analyzed these SNPs for association with fasting insulin (the trait used to localize the gene in mice) in the MACAD cohort. This screen identified three SNPs within SORCS1 that showed evidence of association with fasting insulin (rs10736189, rs2249022, and rs4390282) (Table 1). To confirm these findings, we used a permutation test to derive empirical *P* values for the

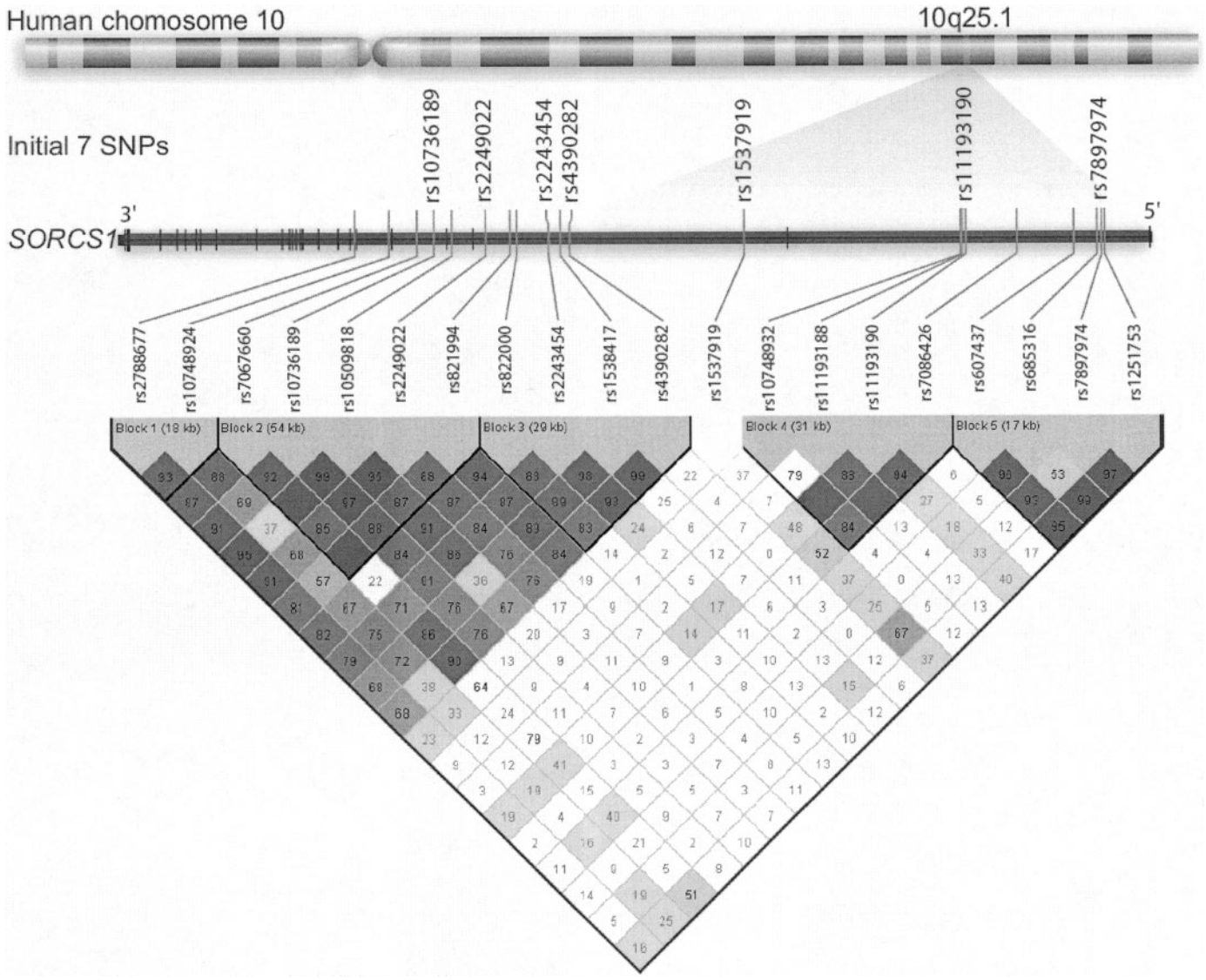

FIG. 1. Location of SORCS1 SNPs and haplotype structure in the MACAD cohort. The upper portion of the figure shows the human SORCS1 gene structure (exons = black lines) on the minus strand of human chromosome 10 and the corresponding location of the SNPs studied. We performed an initial screen of the region syntenic to the *T2dm2* QTL in the MACAD population using seven SNPs representing each major haplotype block. This identified three SNPs (rs10736189, rs2249022, and rs11193190) that were associated with fasting insulin levels. We subsequently screened an additional 13 SNPs in this region to refine the haplotype structure within this population. The LD plot at the bottom of the figure displays D′ values (%) for each pair of SNPs in the box at the intersection of the diagonals from each SNP. Dark solid blocks indicate D′ = 100% for the corresponding pair of variants. Five haplotype blocks were identified.

associations of each of the SNPs with fasting insulin. These *P* values were essentially unchanged from those obtained by our association tests. None of the three flanking SNPs showed evidence of association with fasting insulin (Table 1), suggesting that the causal variation lies within SORCS1 and is not due to long-range LD to neighboring genes.

To confirm the hypothesis that variation in SORCS1 is associated with fasting insulin, we then selected and genotyped 13 additional SNPs in this region to determine the haplotype structure within this population. We identified five haplotype blocks in the MACAD cohort (Fig. 1). Only SNPs located in blocks 2 and 3 displayed association with fasting insulin (Table 1). Thus, we analyzed haplotypes from only these two blocks for association with fasting insulin and other insulin-related traits. Fasting insulin can be a reflection of insulin sensitivity or insulin secretion. To examine the mechanism behind the alterations in fasting insulin of the associated block 2 and 3 SNPs, we also tested the haplotypes for association with indexes of insulin secretion (IGR30) and insulin sensitivity (*M* value). Three haplotypes within blocks 2 and 3 were significantly associated with fasting insulin levels or IGR30 (Table 2). For each associated haplotype, the direction of effect was consistent between fasting insulin and IGR30. No haplotypes were associated with the *M* value, suggesting that, as observed in the mice, variation in SORCS1 affects fasting insulin primarily by influencing insulin secretion.

Because the effects of *T2dm2* were most pronounced in obese female mice (14), we stratified the cohort by weight and sex and then examined the associations of fasting insulin and IGR30 with the three associated haplotypes in blocks 2 and 3. Despite reduced sample size, the effects of haplotype 1-2-2-2-2 within block 2 were still observed in overweight females (Table 2), whereas no significant as-

TABLE 2
Association of haplotypes in SORCS1 with fasting insulin and insulin secretion in MACAD

Block	Associated haplotype	Haplotype frequency (%)	Fasting insulin (pmol/l)			IGR30		
			Noncarrier	Carrier	*P*	Noncarrier	Carrier	*P*
Whole cohort								
Block 2	1-2-2-2-2	21.0	106.19 ± 3.52 (290)	121.90 ± 5.24 (163)	0.0032	0.66 ± 0.03 (282)	0.71 ± 0.04 (159)	0.20
Block 2	1-2-1-2-1	3.5	111.64 ± 3.01 (421)	114.80 ± 13.78 (32)	0.86	0.67 ± 0.02 (411)	0.86 ± 0.09 (30)	0.049
Block 3	2-1-2-2	17.7	106.05 ± 3.09 (327)	120.11 ± 5.38 (156)	0.010	0.66 ± 0.02 (316)	0.71 ± 0.04 (152)	0.086
Female (BMI ≥25 kg/m^2)								
Block 2	1-2-2-2-2	21.9	115.88 ± 5.67 (124)	134.24 ± 7.32 (75)	0.0098	0.73 ± 0.04 (121)	0.83 ± 0.06 (74)	0.085
Block 2	1-2-1-2-1	2.3	120.90 ± 4.52 (190)	162.66 ± 30.71 (9)	0.12	0.75 ± 0.03 (188)	1.13 ± 0.15 (7)	0.059
Block 3	2-1-2-2	17.9	116.74 ± 4.66 (147)	132.74 ± 8.32 (68)	0.11	0.72 ± 0.04 (143)	0.80 ± 0.06 (68)	0.30

Data are means ± SE (*n*). SNPs contained within these blocks are shown in Fig. 1. Allele 1 represents the common allele; allele 2 is the rare allele. IGR30 is given in $\mu IU \cdot dl \cdot mg^{-1} \cdot ml^{-1}$.

sociations of these haplotypes with insulin were observed in either males or lean individuals (data not shown).

The MACAD Study design focused on intermediate insulin-related traits and thus specifically excluded individuals with type 2 diabetes. To ask whether variation in the SORCS1 haplotype blocks are also related to diabetes susceptibility, we then genotyped SNPs from block 2 in the SAFADS cohort, which includes individuals with type 2 diabetes. This block was chosen because it contained the haplotype associated with fasting insulin and IGR30 in overweight women.

We detected associations with diabetes risk and age at diagnosis for most block 2 SNPs in this cohort (Table 3) as well as haplotypes based on these SNPs (Table 4). Consistent with our findings in mice and in the MACAD cohort, these associations were only observed in females. Within block 2, haplotype 1-2-2-2-2 (comprised of SNPs rs7067660, rs10736189, rs10509818, rs2249022, and rs821994) was associated with decreased diabetes risk and a later age of diabetes diagnosis in SAFADS. This same haplotype was associated with increased fasting insulin and IGR30 in MACAD. These data suggest that SORCS1 variation present on this haplotype acts to increase insulin secretion and plasma insulin levels, resulting in protection from developing type 2 diabetes and a later onset of type 2 diabetes in those who do develop it. While some but not all of the five SNPs in block 2 displayed significant associations (Tables 1 and 3), the mean trait levels for each of the five SNP alleles that define haplotype 1-2-2-2-2 were in the same direction as the haplotype itself.

DISCUSSION

We previously positionally cloned *SorCS1* as the gene underlying a QTL affecting plasma insulin levels in obese mice (14). The primary mechanism for this reduction was reduced secretion of insulin into the bloodstream, accompanying an increased disruption of islet morphology. These studies localized the causative variant to the first few exons of *SorCS1*. Here, we have sought to address the question of whether genetic variation within the same region of human SORCS1 is associated with plasma insulin levels and insulin secretion and to determine whether these variants are associated with altered risk of developing type 2 diabetes.

We have detected associations of SNPs within two haplotype blocks with diabetes or diabetes-related traits (fasting insulin and insulin secretion) in two different Mexican-American populations. These SNPs are distinct from those in mice yet were associated with similar phenotypes as those observed in mice (14), suggesting that the effect of SORCS1 on insulin secretion is directly relevant to the development of type 2 diabetes in humans. The risk ratios for type 2 diabetes observed for these SORCS1 variants are similar to those reported for SNPs in other type 2 diabetes susceptibility genes (8,39).

A strength of our study was the use of two distinct but complementary Mexican-American cohorts: one that specifically excluded subjects with type 2 diabetes but underwent detailed phenotyping of intermediate traits for diabetes (MACAD) and another containing diabetic subjects (SAFADS), enabling us to test for association with diabetes and age of diabetes onset. Since diabetic subjects may have either higher or lower fasting insulin levels, depending on the stage of their disease and medication history, assessment of relationships with fasting insulin must be confined to nondiabetic subjects. Thus, we did not test for association of SORCS1 variants with fasting insulin in SAFADS. That the effect is sex specific imposes a

TABLE 3
Association of block 2 SNPs with diabetes risk in SAFADS female subjects

SNP	Major/minor allele	Minor allele frequency (%)	Relative risk for diabetes		Age at diagnosis	
			2/2*	*P*	Effect on age at diagnosis	*P*
rs7067660	T/C	5.2	1.09	0.55	Decrease	0.43
rs10736189	C/G	40.7	0.86	0.041	Increase	0.17
rs10509818	G/A	23.7	0.80	0.0018	Increase	0.0055
rs2249022	A/G	42.3	0.86	0.040	Increase	0.16

*Relative risk of 2/2 compared with 1/1. Allele 1 represents the common allele; allele 2 is the rare allele. Associations were observed in female subjects only and have been adjusted for BMI.

TABLE 4
Association of block 2 haplotypes in SORCS1 with diabetes risk in SAFDS female subjects

Block	Associated haplotype*	Minor allele frequency (%)	Relative risk for diabetes		Age at diagnosis	
			Risk†	*P*	Effect on age at diagnosis	*P*
Block 2	1-1-1-1-1	51.7	1.08	0.067	NA	0.23
Block 2	1-2-2-2-2	22.3	0.78	0.0020	Increased	0.010

*Allele at fifth position inferred from MACAD haplotypes. †Relative risk of bearing 2 copies of haplotype vs. no copies of haplotype. Allele 1 represents the common allele; allele 2 is the rare allele. Associations were observed in female subjects only and have been adjusted for BMI.

further reduction in the number of subjects available for analysis; hence, we lacked power to assess the relationship between SORCS1 variants and fasting insulin in the SAFADS cohort.

The human SORCS1 gene is located on chromosome 10q25. In the SAFADS cohort, we had previously observed linkage in this population between type 2 diabetes and age of type 2 diabetes onset with a broad region of human chromosome 10 encompassing SORCS1 (15). Other rodent (9) and human (16–28) linkages to type 2 diabetes-related traits have been mapped to regions containing SORCS1, raising the possibility that this gene may have broad relevance to the development of type 2 diabetes.

Recently, variants within *TCF7L2* have been identified as predictors of type 2 diabetes (7). This gene resides 5.8 Mb telomeric to SORCS1. Thus, it was possible that the effects of SORCS1 variants on insulin levels and diabetes could have been a consequence of variants of the *TCF7L2* gene. We tested this possibility by determining whether known SNPs within the two genes are in LD; they were not (data not shown). In addition, results from the SNPs flanking SORCS1 suggest that our results are due solely to variation within the SORCS1 gene itself.

SORCS1 belongs to a family of five proteins (sortilin, SORLA [LR11], and SORCS-1, -2, and -3) that are composed of a vacuolar protein sorting-10 domain, a single transmembrane domain and a short cytoplasmic tail (40). A leucine rich-repeat domain between the vacuolar protein sorting-10 and transmembrane segments characterizes the SORCS family. SORCS1 is most highly expressed in the brain, heart, kidney, pancreatic islets, and β-cell lines (14,41). Our congenic mice have reduced insulin levels due to impaired insulin secretion in vivo, but not in isolated islets in vitro, suggesting impaired delivery of insulin to the bloodstream. SORCS1 is known to bind platelet-derived growth factor-BB (42), which plays a role in blood vessel development. Our working hypothesis is that SORCS1 plays a role in the maintaining or expanding islet vasculature during islet growth and compensation for insulin resistance. Impaired vascular structure may directly affect insulin secretion to the blood and may also lead to disrupted islet architecture in areas of insufficient vascular flow (43,44).

The insulin deficiency and the pancreatic islet phenotypes associated with *T2dm2* were only evident in obese mice, not in their lean littermates, suggesting the mouse *SorCS1* variants are insufficient to affect islet morphology in the absence of a challenge to increase β-cell mass and/or function (14). Also, we only observed the genotype effect on plasma insulin levels in female mice. Thus, it is striking that in humans we also observed a genotype effect primarily in overweight female subjects. This may be coincidental or may be reflective of a sexual dimorphism resulting from a complex interaction among SORCS1 genotype, other genes, and/or the hormonal milieu present in women but not men. Future studies will be required to resolve this issue. It is possible that other variants of SORCS1 may influence islet biology in a wider array of individuals.

Block 2 haplotype 1-2-2-2-2 was associated with increased fasting insulin and IGR30 in MACAD and with decreased diabetes risk in SAFADS; no association with insulin sensitivity was observed. This may be interpreted as inherited preservation in insulin secretion leading to protection against developing diabetes for individuals bearing this particular haplotype. Until further genotyping and/or sequencing is conducted, we cannot determine which SNP(s) on this haplotype are responsible for the observed associations. As this study represents our initial exploration of this very large (>500 kb) gene, we cannot rule out the possibility that other regions of the gene will also harbor functional variants or that additional SNPs showing associations (either protective or predisposing to diabetes) will be detected within this region. Any such additional associations will become evident as more complete genotyping and analysis are carried out. The goal of this study is to present the successful translation of a murine diabetes gene to humans, to stimulate further investigation of SORCS1 in other human cohorts.

Using a mouse model that mimics common obesity-associated diabetes, we identified *SorCS1* as the gene underlying the *T2dm2* locus (14). This locus affects in vivo insulin secretion, an important process in compensating for the insulin resistance of obesity. Factors affecting the ability of the pancreatic islets to maintain sufficient compensation for insulin resistance are likely important determinants of diabetes risk. Here, we demonstrate that distinct allelic variants of the SORCS1 gene are associated with alterations in fasting insulin levels, insulin secretion (intermediate traits), and type 2 diabetes risk (overt disease) in Mexican Americans. These findings represent the first example of variants in a gene identified through positional cloning in a model organism affecting type 2 diabetes in humans. Combined, these data strongly suggest that alterations in SORCS1 directly affect these phenotypes and provide validation that SORCS1 is a novel type 2 diabetes susceptibility gene.

ACKNOWLEDGMENTS

This work was supported by National Heart, Blood, and Lung Institute Grant HL60030 to the MACAD project; the National Center for Research Resources Grant MO1-RR00425 to the General Clinical Research Center genotyping/phenotyping core at Cedars-Sinai Medical Center; the Diabetes Endocrinology Research Center Grant DK063491

and the Cedars-Sinai Board of Governor's Chair in Medical Genetics (to J.I.R.); the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grants DK42273, DK47482, and DK53889 to the San Antonio Family Diabetes Study project; NIDDK Grants DK58037 and DK66369 (to A.D.A.); and a Junior Faculty Award from the American Diabetes Association (to D.M.L.). S.M.C. is supported by an American Heart Association Scientist Development grant (0635234N).

We thank Adam Steinberg for his assistance in the preparation of the figure.

This work was presented in part at the American Diabetes Association 66th Scientific Sessions, Washington, DC, 9–13 June 2006.

REFERENCES

1. Elbein SC, Chiu KC, Permutt MA: Type 2 diabetes mellitus. In *The Genetic Basis of Common Diseases.* 2nd ed. King RA, Rotter JI, Motulsky AG, Eds. New York, Oxford University Press, 2002, p. 457–480
2. Hani EH, Boutin P, Durand E, Inoue H, Permutt MA, Velho G, Froguel P: Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of type II diabetes mellitus in Caucasians. *Diabetologia* 41:1511–1515, 1998
3. 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 K_{ATP} 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
4. Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J: A Pro12Ala substitution in PPAR-gamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet* 20:284–287, 1998
5. 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
6. 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
7. 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
8. Das SK, Elbein SC: The genetic basis of type 2 diabetes. *Cellscience* 2:100–131, 2006
9. Clee SM, Attie AD: The genetic landscape of type 2 diabetes in mice. *Endocr Rev* 28:48–83, 2006
10. Darvasi A: Experimental strategies for the genetic dissection of complex traits in animal models. *Nat Genet* 18:19–24, 1998
11. Bell GI, Polonsky KS: Diabetes mellitus and genetically programmed defects in beta-cell function. *Nature* 414:788–791, 2001
12. Rhodes CJ: Type 2 diabetes-a matter of beta-cell life and death? *Science* 307:380–384, 2005
13. Stoehr JP, Nadler ST, Schueler KL, Rabaglia ME, Yandell BS, Metz SA, Attie AD: Genetic obesity unmasks nonlinear interactions between murine type 2 diabetes susceptibility loci. *Diabetes* 49:1946–1954, 2000
14. Clee SM, Yandell BS, Schueler KM, Rabaglia ME, Richards OC, Raines SM, Kabara EA, Klass DM, Mui ET, Stapleton DS, Gray-Keller MP, Young MB, Stoehr JP, Lan H, Boronenkov I, Raess PW, Flowers MT, Attie AD: Positional cloning of Sorcs1, a type 2 diabetes quantitative trait locus. *Nat Genet* 38:688–693, 2006
15. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP: Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am J Hum Genet* 64:1127–1140, 1999
16. An P, Hong Y, Weisnagel SJ, Rice T, Rankinen T, Leon AS, Skinner JS, Wilmore JH, Chagnon YC, Bergman RN, Bouchard C, Rao DC: Genomic scan of glucose and insulin metabolism phenotypes: the HERITAGE Family Study. *Metabolism* 52:246–253, 2003
17. Aulchenko YS, Vaessen N, Heutink P, Pullen J, Snijders PJ, Hofman A, Sandkuijl LA, Houwing-Duistermaat JJ, Edwards M, Bennett S, Oostra BA, van Duijn CM: A genome-wide search for genes involved in type 2 diabetes in a recently genetically isolated population from the Netherlands. *Diabetes* 52:3001–3004, 2003
18. An P, Teran-Garcia M, Rice T, Rankinen T, Weisnagel SJ, Bergman RN, Boston RC, Mandel S, Stefanovski D, Leon AS, Skinner JS, Rao DC, Bouchard C: Genome-wide linkage scans for prediabetes phenotypes in response to 20 weeks of endurance exercise training in non-diabetic whites and blacks: the HERITAGE Family Study. *Diabetologia* 48:1142–1149, 2005
19. Frayling TM, Wiltshire S, Hitman GA, Walker M, Levy JC, Sampson M, Groves CJ, Menzel S, McCarthy MI, Hattersley AT: Young-onset type 2 diabetes families are the major contributors to genetic loci in the Diabetes UK Warren 2 genome scan and identify putative novel loci on chromosomes 8q21, 21q22, and 22q11. *Diabetes* 52:1857–1863, 2003
20. Freedman BI, Rich SS, Sale MM, Heiss G, Djousse L, Pankow JS, Province MA, Rao DC, Lewis CE, Chen YD, Beck SR: Genome-wide scans for heritability of fasting serum insulin and glucose concentrations in hypertensive families. *Diabetologia* 48:661–668, 2005
21. Ghosh S, Watanabe RM, Valle TT, Hauser ER, Magnuson VL, Langefeld CD, Ally DS, Mohlke KL, Silander K, Kohtamaki K, Chines P, Balow J Jr, Birznieks G, Chang J, Eldridge W, Erdos MR, Karanjawala ZE, Knapp JI, Kudelko K, Martin C, Morales-Mena A, Musick A, Musick T, Pfahl C, Porter R, Rayman JB: The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am J Hum Genet* 67:1174–1185, 2000
22. Hanson RL, Imperatore G, Narayan KM, Roumain J, Fagot-Campagna A, Pettitt DJ, Bennett PH, Knowler WC: Family and genetic studies of indices of insulin sensitivity and insulin secretion in Pima Indians. *Diabete Metab Res Rev* 17:296–303, 2001
23. Lakka TA, Rankinen T, Weisnagel SJ, Chagnon YC, Rice T, Leon AS, Skinner JS, Wilmore JH, Rao DC, Bouchard C: A quantitative trait locus on 7q31 for the changes in plasma insulin in response to exercise training: the HERITAGE Family Study. *Diabetes* 52:1583–1587, 2003
24. Reynisdottir I, Thorleifsson G, Benediktsson R, Sigurdsson G, Emilsson V, Einarsdottir AS, Hjorleifsdottir EE, Orlygsdottir GT, Bjornsdottir GT, Saemundsdottir J, Halldorsson S, Hrafnkelsdottir S, Sigurjonsdottir SB, Steinsdottir S, Martin M, Kochan JP, Rhees BK, Grant SF, Frigge ML, Kong A, Gudnason V, Stefansson K, Gulcher JR: Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34–q35.2. *Am J Hum Genet* 73:323–335, 2003
25. Sale MM, Freedman BI, Langefeld CD, Williams AH, Hicks PJ, Colicigno CJ, Beck SR, Brown WM, Rich SS, Bowden DW: A genome-wide scan for type 2 diabetes in African-American families reveals evidence for a locus on chromosome 6q. *Diabetes* 53:830–837, 2004
26. Silander K, Scott LJ, Valle TT, Mohlke KL, Stringham HM, Wiles KR, Duren WL, Doheny KF, Pugh EW, Chines P, Narisu N, White PP, Fingerlin TE, Jackson AU, Li C, Ghosh S, Magnuson VL, Colby K, Erdos MR, Hill JE, Hollstein P, Humphreys KM, Kasad RA, Lambert J, Lazaridis KN, Lin G, Morales-Mena A, Patzkowski K, Pfahl C, Porter R, Rha D, Segal L, Suh YD, Tovar J, Unni A, Welch C, Douglas JA, Epstein MP, Hauser ER, Hagopian W, Buchanan TA, Watanabe RM, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* 53:821–829, 2004
27. Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M, O'Rahilly S, Frayling TM, Bell JI, Lathrop GM, Bennett A, Dhillon R, Fletcher C, Groves CJ, Jones E, Prestwich P, Simecek N, Rao PV, Wishart M, Bottazzo GF, Foxon R, Howell S, Smedley D, Cardon LR, Menzel S, McCarthy MI: A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* 69:553–569, 2001
28. Vionnet N, Hani EH, Dupont S, Gallina S, Francke S, Dotte S, De Matos F, Durand E, Lepretre F, Lecoeur C, Gallina P, Zekiri L, Dina C, Froguel P: Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21–q24. *Am J Hum Genet* 67:1470–1480, 2000

29. Goodarzi MO, Guo X, Taylor KD, Quinones MJ, Saad MF, Yang H, Hsueh WA, Rotter JI: Lipoprotein lipase is a gene for insulin resistance in Mexican Americans. *Diabetes* 53:214–220, 2004
30. Goodarzi MO, Taylor KD, Guo X, Quinones MJ, Cui J, Li Y, Saad MF, Yang H, Hsueh WA, Hodis HN, Rotter JI: Association of the diabetes gene calpain-10 with subclinical atherosclerosis: the Mexican-American Coronary Artery Disease Study. *Diabetes* 54:1228–1232, 2005
31. DeFronzo RA, Tobin JD, Andres R: Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am J Physiol* 237:E214–E223, 1979
32. Hanson RL, Pratley RE, Bogardus C, Narayan KM, Roumain JM, Imperatore G, Fagot-Campagna A, Pettitt DJ, Bennett PH, Knowler WC: Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *Am J Epidemiol* 151:190–198, 2000
33. The International HapMap Consortium: The International HapMap Project. *Nature* 426:789–796, 2003
34. Livak KJ: Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet Anal* 14:143–149, 1999
35. Zeger SL, Liang KY: Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42:121–130, 1986
36. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265, 2005
37. Sobel E, Lange K: Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 58:1323–1337, 1996
38. Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198–1211, 1998
39. Parikh H, Groop L: Candidate genes for type 2 diabetes. *Rev Endocr Metab Disord* 5:151–176, 2004
40. Hampe W, Rezgaoui M, Hermans-Borgmeyer I, Schaller HC: The genes for the human VPS10 domain-containing receptors are large and contain many small exons. *Hum Genet* 108:529–536, 2001
41. Hermey G, Keat SJ, Madsen P, Jacobsen C, Petersen CM, Gliemann J: Characterization of sorCS1, an alternatively spliced receptor with completely different cytoplasmic domains that mediate different trafficking in cells. *J Biol Chem* 278:7390–7396, 2003
42. Hermey G, Sjogaard SS, Petersen CM, Nykjaer A, Gliemann J: Tumour necrosis factor alpha-converting enzyme mediates ectodomain shedding of Vps10p-domain receptor family members. *Biochem J* 395:285–293, 2006
43. Brissova M, Shostak A, Shiota M, Wiebe PO, Poffenberger G, Kantz J, Chen Z, Carr C, Jerome WG, Chen J, Baldwin HS, Nicholson W, Bader DM, Jetton T, Gannon M, Powers AC: Pancreatic islet production of vascular endothelial growth factor-a is essential for islet vascularization, revascularization, and function. *Diabetes* 55:2974–2985, 2006
44. Lammert E, Gu G, McLaughlin M, Brown D, Brekken R, Murtaugh LC, Gerber HP, Ferrara N, Melton DA: Role of VEGF-A in vascularization of pancreatic islets. *Curr Biol* 13:1070–1074, 2003