

# Linkage and Association Mapping of a Chromosome 1q21-q24 Type 2 Diabetes Susceptibility Locus in Northern European Caucasians

Swapan Kumar Das,<sup>1</sup> Sandra J. Hasstedt,<sup>2</sup> Zhengxian Zhang,<sup>1</sup> and Steven C. Elbein<sup>1,3</sup>

We have identified a region on chromosome 1q21-q24 that was significantly linked to type 2 diabetes in multiplex families of Northern European ancestry and also in Pima Indians, Amish families, and families from France and England. We sought to narrow and map this locus using a combination of linkage and association approaches by typing microsatellite markers at 1.2 and 0.5 cM densities, respectively, over a region of 37 cM (23.5 Mb). We tested linkage by parametric and non-parametric approaches and association using both case-control and family-based methods. In the 40 multiplex families that provided the previous evidence for linkage, the highest parametric, recessive logarithm of odds (LOD) score was 5.29 at marker D1S484 (168.5 cM, 157.5 Mb) without heterogeneity. Nonparametric linkage (NPL) statistics ( $P = 0.00009$ ), SimWalk2 Statistic A ( $P = 0.0002$ ), and sib-pair analyses (maximum likelihood score = 6.07) all mapped to the same location. The one LOD CI was narrowed to 156.8–158.9 Mb. Under recessive, two-point linkage analysis, adjacent markers D1S2675 (171.5 cM, 158.9 Mb) and D1S1679 (172 cM, 159.1 Mb) showed LOD scores  $>3.0$ . Nonparametric analyses revealed a second linkage peak at 180 cM near marker D1S1158 (163.3 Mb, NPL score 3.88,  $P = 0.0001$ ), which was also supported by case-control (marker D1S194, 178 cM, 162.1 Mb;  $P = 0.003$ ) and family-based (marker ATA38A05, 179 cM, 162.5 Mb;  $P = 0.002$ ) association studies. We propose that the replicated linkage findings actually encompass at least two closely spaced regions, with a second susceptibility region located telomeric at 162.5–164.7 Mb. *Diabetes* 53:492–499, 2004

From the <sup>1</sup>Department of Medicine, University of Arkansas for Medical Sciences, Little Rock, Arkansas; the <sup>2</sup>Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, Utah; and the <sup>3</sup>Department of Medicine, Central Arkansas Veterans Healthcare System, Little Rock, Arkansas.

Address correspondence and reprint requests to Steven C. Elbein, MD, Professor of Medicine, University of Arkansas for Medical Sciences, Endocrinology 111J-1/LR, John L. McClellan Memorial Veterans Hospital, 4700 W. 7th St., Little Rock, AR 72205. E-mail: elbeinsteve@uams.edu.

Received for publication 26 August 2003 and accepted in revised form 10 November 2003.

Additional information for this article can be found in an online appendix at <http://diabetes.diabetesjournals.org>.

HLOD, heterogeneity LOD; IDB, identical by descent; LOD, logarithm of odds; MLS, maximum likelihood score; NPL, nonparametric linkage; PKLR, liver- and red cell-type pyruvate kinase; RORC, retinoid-related orphan receptor  $\gamma$ ; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test.

© 2004 by the American Diabetes Association.

**T**ype 2 diabetes (MIM125853) likely encompasses a diverse set of diseases marked by elevated levels of plasma glucose. Among Caucasian populations, individuals with type 2 diabetes, individuals with the intermediate phenotype of impaired glucose tolerance, and likely individuals at risk of diabetes are all characterized by variable degrees of both decreased insulin action, particularly resistance to insulin-mediated muscle glucose uptake, and impaired insulin secretion in response to that decreased insulin action (1). Defects of both insulin action and insulin secretion among individuals with normal glucose tolerance predict later onset of diabetes (2). Despite the diverse phenotypic nature of type 2 diabetes, monozygotic and dizygotic twin studies, family studies, and marked differences in disease prevalence across populations all provide convincing evidence for an important role of genetic susceptibility loci in type 2 diabetes pathogenesis (1). Based on epidemiological data, the total sibling relative risk ( $\lambda_s$ ) has been estimated at 3–4 (3), although the number of loci that contribute to this risk is unclear.

Based on these data supporting type 2 diabetes susceptibility genes, genome scans for both type 2 diabetes and type 2 diabetes-related traits have been undertaken by multiple laboratories in Caucasian, Pima Indian, African-American, and Asian populations (1,4,5), among others. These scans have identified possible susceptibility loci throughout the genome, but to date only the *NIDDM1* locus on chromosome 2q in Mexican-American subjects has been mapped to a single gene, the calpain 10 gene (6). Calpain 10 plays a small role in most other populations, however, and has been inconsistently replicated by linkage and association. Other regions with evidence for replication include chromosome 12q (7–9) and chromosome 20 (10–12). A region on chromosome 1q21-q23 was identified independently among Pima Indian sib-pairs discordant for type 2 diabetes or Pima Indian sib-pairs with onset of diabetes before age 25 years (13) and in studies from our laboratory of 42 multiplex kindreds of Northern European ancestry ascertained in Utah (14). Subsequent studies in French families (15), English sib-pairs (16), and Amish families (17) and in preliminary studies of Chinese sib-pairs (18) have identified linkage of type 2 diabetes to this same region, very near the original Pima and Utah linkage peaks. Furthermore, this region was linked to HbA<sub>1c</sub> in the Framingham Offspring Study (19), to meta-

bolic syndrome traits in nuclear families from Hong Kong (20), and to the possibly related phenotype of familial combined hyperlipidemia (21,22). Given the difficulty in replicating linkage in complex diseases, the finding of diabetes and related traits in at least 10 studies from diverse populations is striking. However, the exact map location of the linkage peaks, the specific trait or disease definition for the study, and the subgroup providing the evidence for linkage differs among studies.

In previous studies from our laboratory (23), the most significant linkage peak (logarithm of odds [LOD] = 4.3) was found using pedigrees trimmed to fit into the Genehunter program (24) under a partially penetrant recessive parametric model. The linkage peak was quite broad, with a 1 LOD CI that extended from between D1S305 and CRP to D1S196, or ~20 cM. A similar location, albeit with lower significance, was identified with both sib-pair analysis (Mapmaker/Sibs) and nonparametric linkage (NPL). Studies in Pima Indians and in French families placed the linkage peak within 5 cM of our data, although initial Amish and English studies placed the peak centromeric or telomeric, respectively. In post hoc analyses from our laboratory, the LOD score was reduced when full families were used for fewer markers, when unaffected individuals were removed from the analysis, and when individuals with intermediate diagnoses were removed. In contrast, removal of two families that segregated hepatocyte nuclear factor 1 $\alpha$  variants increased the LOD score to 4.87 in the remaining 40 families (23,25). Finally, we found no linkage to chromosome 1 in either 21 smaller replication families or when all 63 families were analyzed together without heterogeneity (23). The goal of the present study was to localize the well-replicated type 2 diabetes susceptibility gene in this region using a dense microsatellite map across a 37-cM region for linkage, case-control association studies, and family-based association studies.

## RESEARCH DESIGN AND METHODS

We performed a number of analyses using both family-based and case-control studies to narrow the regions of susceptibility genes on chromosome 1q. For linkage analyses, we first attempted to replicate the earlier analyses showing linkage under a recessive model (23), but using a dense marker map. Although software is now available that permits multipoint analysis of full families, we included recessive analysis using Genehunter v. 2.1 and Genehunter-sized families to be comparable with our earlier analysis. While our highest linkage peak was under a recessive model, based on the variable location of the linkage peak from other laboratories and unpublished data from our laboratory suggesting associations in multiple locations, we considered the possibility that multiple susceptibility loci might be present and that these loci might have different modes of inheritance. To test this hypothesis, we included two nonparametric (model independent) analyses, one using the statistics implemented in SimWalk 2 (26) and the other using the sib-pair analysis that provided the highest nonparametric score in our previous study (23). By using multiple analytical methods, we were also able to assess whether the localization of the linkage peaks was robust to model assumptions. Finally, based on the success of microsatellite association studies in mapping other complex disease genes (27,28), we included both case-control and family-based studies of a dense microsatellite map as a framework for mapping genes by association.

We included two closely related study populations. Both linkage and family-based association studies were conducted in samples from previously described families (23). Briefly, the primary studies were conducted on 618 members of 42 families (526 nonfounders). The mean number of individuals tested was 13.3 per family, and the mean number of affected individuals per family was 4.0, with a mean age of onset of 50.6 years. An additional 27 smaller families (mean number of individuals tested: 6.6), which included six families that were not previously typed, were used as a replication set and were typed for all markers in the present study. The replication families were ascertained

under the same criteria as the initial families, but the families were smaller and had fewer available members for testing. All families were ascertained for at least two siblings with type 2 diabetes diagnosed before the age of 65 years and with no more than one parent known to have type 2 diabetes. All subjects were ascertained in Utah for Northern European ancestry. All available parents and siblings of the index sib-pair, as well as all available offspring of diabetic siblings, were studied. All nondiabetic individuals underwent a 75-g oral glucose tolerance test. Subjects were classified as affected if they had a previous diagnosis of type 2 diabetes and were on medical therapy. To incorporate young-onset impaired glucose tolerance into the affection status, individuals were considered affected if the fasting glucose exceeded 7.8 mmol/l or if the 2-h postchallenge glucose was >7.8 mmol/l for participants under age 45 years, 11.1 mmol/l for participants aged 45–64 years, or 13.3 mmol/l for those over age 64 years. All other individuals with abnormal glucose tolerance tests were considered to be of unknown affection status. This scheme closely follows the World Health Organization criteria for impaired glucose tolerance (under age 45 years) and type 2 diabetes (age 45–64 years) but raises the postchallenge glucose for elderly subjects based on epidemiological data. All diagnoses were the same as in our previous study (23). Uncertainty was programmed into parametric models for individuals considered affected but who did not meet the criteria for type 2 diabetes.

Case-control association studies were conducted on 150 unrelated individuals with known type 2 diabetes and 150 ethnically matched, unrelated control individuals. Of the type 2 diabetic individuals, 70 were selected from the linkage families and 80 additional individuals were selected from the same population for type 2 diabetes and a family history of type 2 diabetes in a first-degree relative. Control individuals included spouses from linkage families who had normal glucose tolerance tests (108 subjects) and Caucasian individuals ascertained in Utah or Arkansas (42 subjects) who had normal glucose levels or glucose tolerance tests and no family history of diabetes in a sibling, parent, or grandparent.

All individuals provided written informed consent under protocols approved by the University of Utah Institutional Review Board (diabetic kindreds and case-control population) or the University of Arkansas for Medical Sciences Institutional Review Board (additional case-control samples).

**Marker selection and typing.** For linkage studies of chromosome 1, we added 37 microsatellite markers to the 38 markers previously typed (23), with 29 new markers in the region between D1S305 and D1S212, where previous linkage signals were found. Marker order and spacing was derived from published maps (29,30) with reference to the physical map to establish the order and distance for closely spaced markers (National Center for Biotechnology Information [NCBI] build 33). The average marker distance between D1S305 and D1S212 was 1.17 cM. For the population-based case-control association study, we typed 46 microsatellite markers between markers D1S305 and D1S212, with an average inter-marker distance of 0.52 Mb.

Microsatellite markers were amplified in the presence of universal M13 forward primers that were labeled with LI-COR IR700 and IR800 dyes, and the products were separated and detected on LI-COR 4200 sequencers using standard methods (Li-COR, Lincoln, NE). Genotypes were scored automatically using either SAGA<sup>GT</sup> software (31) (Li-COR) or semiautomatically using GeneImage IR 3.56 software (Scanalytics, Fairfax, VA). All readings were reviewed independently, and between 30 and 50 blinded duplicate samples were included for all markers for both linkage and association studies. All gels included at least two additional samples from selected grandparents of CEPH (Centre d'Etude du Polymorphisme Humain) families as an additional quality control. Before linkage analysis, all data were checked for inconsistencies in size, inconsistencies between duplicates, and inconsistencies in Mendelian inheritance using the PEDCHECK program (v. 1.1) (32). All blinded duplicates were in agreement with the exception of four samples that were consistently incorrect and appeared to be incorrectly identified duplicate samples. We identified 0.98% genotyping errors (251 of 28,095 genotypes that were automatically read without reference to pedigree data) that resulted in noninheritance and were changed to unknown before analysis.

**Linkage analysis.** The marker map used for all multipoint studies was derived from primary reference to the Marshfield map (<http://research.marshfieldclinic.org/genetics>), which included all of the typed markers. To properly space markers that were too close to be resolved on the Marshfield linkage map, we set the distance between markers with 0 recombination fractions to 0.5 cM, with marker order based on the physical map. Consequently, our map over the region from D1S305 to D1S212 was expanded by 3 cM from the Marshfield map and by 4 cM from the recently published DeCode map (33). Thus, exact locations used in the current study differ slightly from those cited in the most recent Marshfield map.

Despite careful quality control and retyping of markers with excess recombination events, recombination between closely linked markers ex-

ceeded expectations for many intervals. Inspection of genotypes failed to identify errors leading to increased recombination. Consequently, before multipoint analysis we used a mistyping analysis implemented in SimWalk2 (v. 2.82) (26) to remove all genotypes that had a 25% or greater posterior probability of error based on excess recombination. These genotypes were considered missing for all multipoint analyses. We removed a total of 882 of 48,017 genotypes for all 62 families (1.8%). Expected and observed recombination rates for each interval are shown in the online supplemental data (Table 1).

We conducted multipoint linkage analysis under a recessive parametric model that provided the maximum LOD score in our previous studies using Genehunter version 2.1\_r3 beta (24,34) and families trimmed to fit this program. Nonparametric analyses were performed using statistics A through E in SimWalk 2 (v. 2.82) (26). Additionally, based on previous results showing the highest LOD score under a sib-pair analysis, we performed sib-pair linkage analysis using Genehunter (v. 2.1\_r3) under models of dominance variance and no dominance variance (35). The recessive parametric model set the disease allele frequency at 0.25 and included a linear, age-dependent penetrance function that varied from 0.02 below age 30 years to 0.60 over age 65 years (23). The allele frequency of each microsatellite marker used for linkage analysis was estimated from unrelated pedigree members, assuming Hardy-Weinberg equilibrium. Linkage studies were conducted on the full 69-family set (original families and replication families) and on the 40 families that provided the maximum evidence for linkage in our previous study. These 40 families were selected from the 42 families of the previous study but excluded two families that segregated hepatocyte nuclear factor 1 $\alpha$  variants (25). To fit the large families into Genehunter Plus, individuals who were unaffected or of unknown affection status were trimmed before analysis as described previously (23). The location score was also calculated in SimWalk2 using full families. Parametric recessive LOD scores were calculated assuming homogeneity ( $\alpha = 1$ ) and allowing for heterogeneity. The maximum likelihood estimate of alleles shared identically by descent (IBD) among sib-pairs from the 40 kindreds that were primarily responsible for earlier linkage findings was calculated both with and without weighting to correct for multiple sib-pairs and both with and without dominance variance ( $\lambda_S = \lambda_0$ ).

Because of the increased recombination observed in this study despite elimination of clear genotyping errors and to minimize the impact of map errors, particularly between closely spaced markers, we supplemented the multipoint analyses with a two-point linkage analysis of the 40-family set under the recessive model using the FASTLINK program (36). To further minimize the errors in recombination fractions resulting from sex-averaged estimates of recombination, we incorporated sex-specific recombination fractions in these analyses.

**Tests of association.** Population association tests for microsatellite alleles were conducted for 43 markers using CLUMP v. 1.9 software (37). We report the maximized  $\chi^2$  test (T4 statistic), which calculates the maximum  $\chi^2$  value found by collapsing the contingency tables over each allele in turn to form  $2 \times 2$  contingency tables. The significance was assessed using a Monte Carlo approach with 10,000 simulations. Family-based associations with type 2 diabetes were tested in 69 families using a modification of the transmission disequilibrium test (TDT) (38), as implemented in the Pedigree Analysis Package (39) and described previously (40). This analysis tests the probability that a heterozygous parent transmits an allele to an affected offspring more often than expected by chance, similar to the gamete-competition model described by Sinsheimer et al. (41). Increased transmission from parents to affected offspring was tested by maximum likelihood analysis against equal transmission of the alleles. All alleles at a marker were tested simultaneously with  $k-1$  df, where  $k$  represents the number of alleles. The pedigree is analyzed as an intact unit, so that trios and nuclear families were not examined separately. Because linkage in this region was established, this likelihood test was a test of association. Data are presented without correction for the number of markers tested. In a case control study of a two-allele marker, our power for a test of allelic association with 150 individuals in each group exceeds 80% for differences in allele frequency of 12% or greater. Linkage disequilibrium between microsatellite adjacent markers was calculated from the case-control study of unrelated individuals (both case and control subjects included) as a multilocus  $D'$  value using the expectation maximization algorithm as implemented in the 2LD program (<http://linkage.rockefeller.edu>).

**Haplotype estimation and haplotype sharing analysis.** Haplotypes were inferred in 40 large pedigrees using 33 ordered markers from D1S305 to D1S212 and SimWalk2 (v. 2.82) following the methods of Saarela et al. (42). Sharing of maternal and paternal haplotypes between affected siblings within each family was determined by manual inspection (42).

## RESULTS

We examined a total of 75 microsatellite markers on chromosome 1, including 38 markers previously reported (23) and 37 markers newly typed. We typed a total of 33 markers in the region of the previously described linkage peak from marker D1S305 (159 cM) to marker D1S212 (196 cM), with all locations referenced to the Marshfield map (<http://research.marshfieldclinic.org/genetics>) (29). In our earlier analysis, we considered two family sets: the 42 families that constituted our primary genome-wide scan, and 21 smaller replication families. For the present study, we considered all available families (69 families; the original 42 families and 27 replication families, including 6 families not considered in the earlier study) and the 40 families from the original 42 families for which we had not identified another potential diabetes susceptibility gene. Based on our earlier data, we chose the recessive parametric model that provided the best evidence for linkage previously as the primary tool for narrowing the linkage peak. However, to determine whether that localization was robust to model assumptions, we also analyzed the linkage data under nonparametric models.

**Parametric linkage analysis.** As in our previous report of 21 replication families (23), we found no evidence for linkage in the 27 replication families despite the dense map. Using the full available pedigree set (69 families), we only found evidence for linkage under models that incorporated heterogeneity, with a maximum heterogeneity LOD (HLOD) score of 1.42 with 25% of families linked at position 168.5 cM (marker D1S484). When the 40 families from the original linkage study that did not segregate hepatocyte nuclear factor 1 $\alpha$  variants were tested, the maximum LOD score using families trimmed to fit Genehunter requirements was 5.28 at the same location (position 168.5 cM; marker D1S484), which was increased from 4.89 in our previous study. In contrast to the full family set, we found little evidence for heterogeneity (HLOD = 5.29;  $\alpha = 0.96$ , 168.5 cM) using the 40 families that were trimmed of many unaffected individuals. As in our previous analysis (23), inclusion of all unaffected individuals using the Simwalk2 program dropped the nonheterogeneity location score to 2.98 and the heterogeneity LOD score to 4.07 ( $\alpha = 0.65$ ) without moving the location of the peak (marker D1S484; 168.5 cM) (Fig. 1). Based on the Genehunter analysis, the one LOD CI was narrowed to 167.6–170.6 cM, corresponding to locations 156.8–158.9 Mb on the physical map (NCBI Build 33).

**Nonparametric analyses.** To determine whether localization of the chromosome 1q type 2 diabetes susceptibility locus was robust to model assumptions, we tested linkage also using nonparametric approaches (Fig. 2). Our primary nonparametric analyses used SimWalk2, which could handle full families (Fig. 2), and affected sib-pair analysis using the Genehunter program (Fig. 3). Although these analysis corroborated the location of the first peak at 168.5 cM (Marker D1S484, 157.5 Mb; Genehunter NPL score 4.30;  $P = 0.00009$ ), they showed a prominent second peak not seen in the parametric analysis ~12 cM telomeric to the first peak at 180 cM, between markers D1S1158 (163.3 Mb) and D1S2762 (163.6 Mb; NPL score 3.88;  $P = 0.0001$ ) (Fig. 2). Using no weighting for sibships and assuming dominance variance, the highest maximum like-

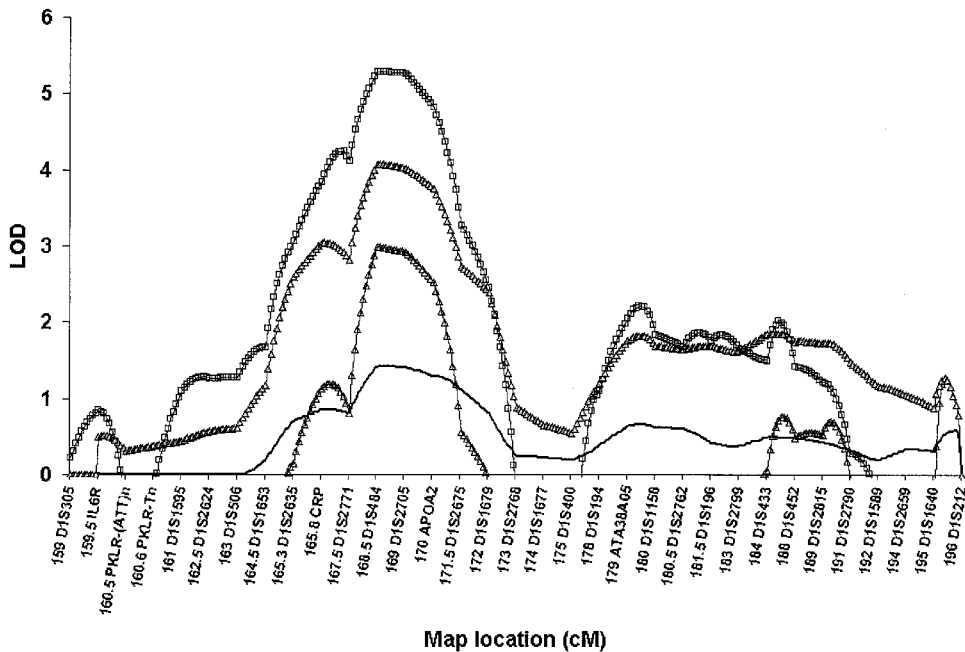


FIG. 1. Multipoint parametric linkage tests. Curves are listed from highest to lowest.  $\square$ , 40-family recessive LOD score in Genehunter-sized families;  $\blacktriangle$ , 40-family HLOD scores in full families;  $\triangle$ , 40-family recessive LOD score, full families; —, HLOD in 69 families (full families).

likelihood score (MLS) was 6.07 at 168.5 cM and 5.25 at 180 cM (Fig. 3). Additionally, a third peak was evident on sib-pair analysis (unweighted; MLS = 2.98) centromeric to the larger peaks at 152.8 cM and just proximal to marker D1S305 (151.0 Mb) and near candidate genes liver- and red cell-type pyruvate kinase (PKLR) (152.0 Mb), retinoid-related orphan receptor  $\gamma$  (RORC), and interleukin 6 receptor (IL6R; 151.1 Mb). Nonparametric statistics examined in SimWalk2, which did not permit simultaneous consideration of the full map region, nonetheless showed similar trends for location (Fig. 2). The most significant SimWalk2 results were seen with statistic A, which is strongest under recessive models, at  $P = 0.0002$  and location 168.5 cM at marker D1S484. The second peak was less obvious with the SimWalk2 statistics but was most significant near marker D1S433 (184 cM, 165.0 Mb;  $P =$

0.001 for statistic C,  $P = 0.002$  for statistic D) (Fig. 2), which was  $\sim 4$  cM or 1.4 Mb telomeric to the Genehunter NPL and sib-pair analyses. When the full family set (69 families) was examined together, the highest MLS scores on sib-pair analysis were 1.73 at 170 cM (APOA2; 158.0 Mb) and 2.46 at 180 cM (D1S1158; 163.3 Mb). Thus, when all families were considered, the proximal peak moved slightly telomeric and the distal peak slightly centromeric but retained approximately the same locations.

**Two-point LOD score.** We observed an unexpectedly high recombination fractions between closely spaced markers despite retyping several markers and careful scrutiny of recombination events (online Supplemental Data, Table 1). To reduce the effect of these potential errors and to incorporate sex-specific recombination fractions, we calculated two-point parametric LOD scores

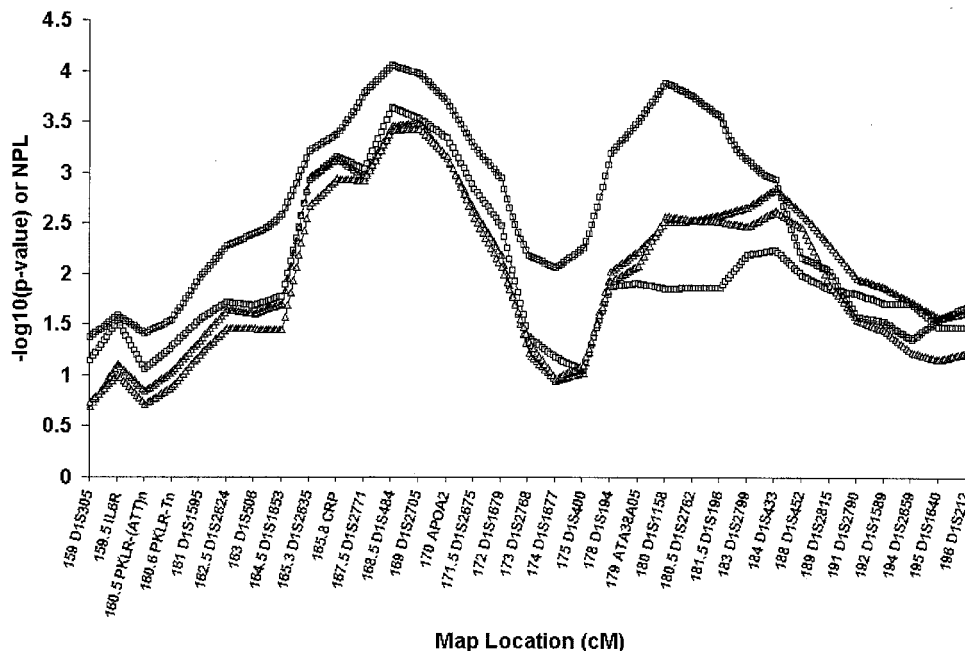


FIG. 2. Multipoint nonparametric linkage tests. Scores are shown for 40 families using SimWalk2 (statistics A, C, and D) in full families or Genehunter NPL statistic in families trimmed to fit the Genehunter program. All scores are  $-\log_{10}$  of the  $P$  value except for the Genehunter NPL score, which is shown as the NPL statistic. Curves are listed from highest to lowest.  $\square$ , Genehunter NPL;  $\blacksquare$ , Simwalk2 statistic A;  $\triangle$ , Simwalk2 statistic C;  $\blacktriangle$ , Simwalk2 statistic D.

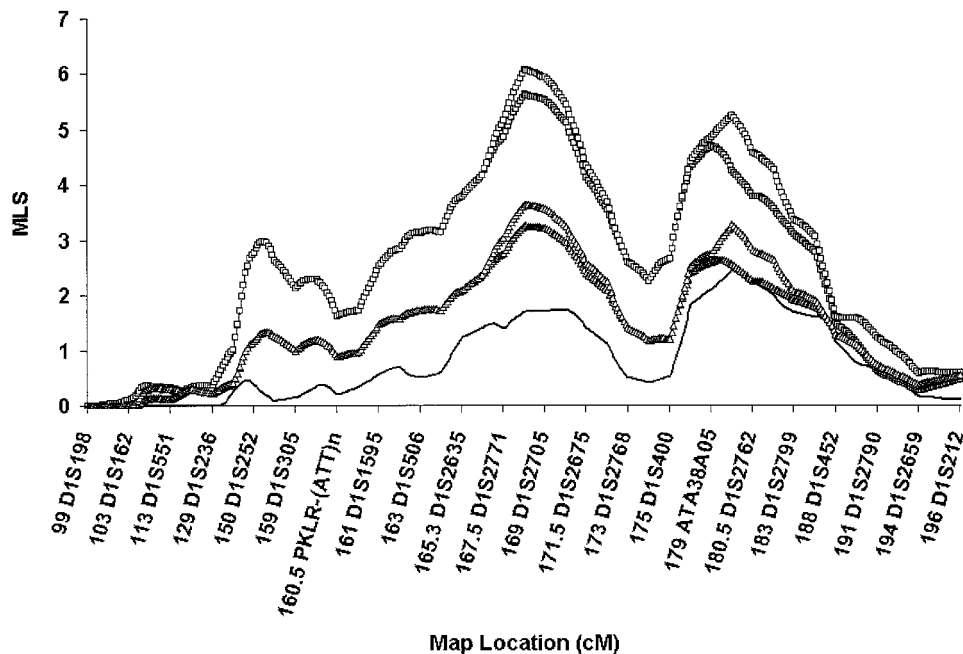


FIG. 3. MLS of IBD sharing by sib-pair analysis.  $\square$ , 40-family analysis with dominance variance, no weighting;  $\blacksquare$ , 40-family analysis with no dominance variance, no weighting;  $\triangle$ , 40-family analysis, dominance variance with weighted sib-pairs;  $\blacktriangle$ , 40-family analysis, no dominance variance, weighted sib-pairs; —, 69-family analysis, dominance variance, and no weighting.

using recessive parametric model described above. As shown in Table 2 of the online Supplemental Data, LOD scores exceeded 3.0 for markers D1S2675 (LOD 3.06) and D1S1679 (3.45), which are located at 171.5 cM (158.9 Mb) and 172 cM (159 Mb), respectively, just telomeric to the recessive multipoint linkage peak near markers D1S484 and D1S2705 (168.5 cM or 157.5 Mb and 169 cM or 157.6 Mb, respectively).

**Association studies.** To further localize the type 2 diabetes susceptibility locus, we tested association in a case-control population comprising diabetic case subjects and nondiabetic control subjects ascertained in Utah or Arkansas for 46 microsatellite markers. We also tested the 33 markers used in the linkage studies for excess transmission of any allele from parents to affected offspring using maximum likelihood methods. In case control studies, markers D1S194 (178 cM, 162.1 Mb) and D1S1677 (176 cM, 160.2 Mb) were nominally significant at  $P = 0.003$  and  $P = 0.012$ , respectively, based on Monte Carlo assessment of significance tested using the CLUMP statistic T4 to examine all alleles simultaneously (37). Marker ATA38A05 at 179 cM (162.5 Mb) was most strongly associated by TDT ( $P = 0.002$ ). These markers fall under the second linkage peak, with both D1S194 and ATA38A05 falling within the 1 LOD CI for the sib-pair analysis. The data for all 46 microsatellites is shown in Table 2 of the online Supplemental Data. Multipoint linkage disequilibrium between adjacent pairs of markers ranged from not significantly different from 0 to the highest  $D'$  value of 0.483 (Table 3 of online Supplemental Data).

**Haplotype sharing.** We followed the methods of Saarela et al. (42) to establish shared haplotypes for the 33 markers that spanned the 37-cM region between markers D1S305 and D1S212. Haplotypes were inferred in SimWalk2 and were examined manually for sharing among the 58 sibships that had two or more affected individuals from the 40 families. Although no single haplotype was shared by all sibships, a 1.16-cM region centered on the first linkage peak and flanked by markers D1S2771 and

D1S2705 was shared by 32 of 58 sibships (Table 4 of online Supplemental Data).

## DISCUSSION

Multiple genome-wide scans for type 2 diabetes have implicated a large number of regions for possible susceptibility genes. To date, only a single gene has been cloned, *NIDDM1* or calpain 10 on chromosome 2q, but the at-risk haplotype at this locus is rare outside of Hispanic populations. Other regions with evidence for replication include chromosomes 12q and 20, but the replication has generally been at some distance from the original description. Chromosome 1 has now been identified in Pima Indians (13), our studies described here, Amish Caucasians (17), British Caucasians (16), French Caucasians (15), and in preliminary reports of both Chinese and African Americans (5). Furthermore, a syntenic region was identified in the GK rat (43). The location of these linkage peaks is remarkably consistent but nonetheless spans the three peaks observed in the present study (5). Thus among Amish with both type 2 diabetes and impaired glucose homeostasis, the peak was near 159 cM (marker D1S2858), with a second peak that was centromeric on the *P* arm. This first peak falls just centromeric to our primary peak at 169 cM. Among Pima Indians, the highest scores were at 175 cM (sib-pairs discordant for diabetes) and 200 cM (sib-pairs with onset before age 25 years), and thus fall more into our second linkage peak. Initial reports from Wiltshire et al. (16) placed their linkage in the region of our second peak at 181 cM (D1S196), although additional markers are reported to have moved the highest score more centromeric to the location of our first and largest linkage peak. The results of Vionnet et al. (15) place their chromosome 1 peak in nearly the same location as our first peak, albeit only in lean ( $BMI < 27 \text{ kg/m}^2$ ) individuals. The loci reported in other studies are not precisely localized (5). These studies thus support the possibility that several

susceptibility loci account for the apparent replication across studies, as suggested by our distinct linkage peaks.

We focused the current study on the 40 multiplex families that provided the majority of the evidence for linkage in our initial report and that did not segregate other known mutations. Unlike the original study, the dense map of approximately one marker every centiMorgan has resolved the broad linkage peak observed initially into at least two narrow peaks. The first of these peaks has moved slightly centromeric from the original peak at APOA2 (170 cM) to the present location of D1S484 (168.5 cM). With additional markers, the LOD score has increased to 5.28 under the recessive model and using Genehunter-sized pedigrees. Similarly, using multipoint sib-pair analysis, the MLS has increased from 2.98 in the original study to 6.07 in the present study. Despite the large variation in significance levels for the first peak with different analytical methods, the location of this peak was remarkably consistent. Both the SimWalk2 statistic *A* and the parametric analysis continue to support a recessive-like mode of inheritance for the susceptibility gene or genes that accounts for the first peak. Based on the present analyses, we have narrowed the 1 LOD CI for this peak to a region from 167.6 cM (156.8 Mb) to 170.6 cM (158.9 Mb). This peak includes at least 60 RefSeq genes, including a number of strong candidate genes, many of which have been evaluated by our laboratory and others. Among the candidate genes previously evaluated in this region are apolipoprotein A2 (APOA2) at 170 cM (157.9 Mb) (44); phosphoprotein enriched in astrocytes (PEA15), which may be involved in insulin action (45); C-reactive protein, which may be involved in inflammation (46); and two inwardly rectifying potassium channel genes, KCNJ9 and KCNJ10 (47,48). None of the reported associations of single nucleotide polymorphisms (SNPs) in these genes can convincingly account for the strong linkage signal in our families, however. In contrast, we have identified two regions within the 1 LOD support interval in which a cluster of SNPs shows strong associations with type 2 diabetes in case-control studies. These associations thus appear to support the linkage findings. Neither region falls close to a strong candidate gene, but work is in progress to identify additional polymorphisms within these regions and to evaluate nearby coding genes. Additional support for an association under this peak has come from other groups with linkage in this region (17).

Unlike our original report, the present study suggests a second peak at 180 cM, ~10 cM from the first peak at 169 cM. Based on the 40-family sib-pair analysis, the 1 LOD support interval is 177.7–181.6 cM, or ~162.5 to 164.7 Mb. Unlike the first peak, this second region is much less prominent using the recessive parametric models and is most prominent using multipoint sib-pair analysis, under which this peak nearly equals the first peak with a MLS of 5.247. These data suggest that the susceptibility locus accounting for the second peak acts less like a recessive locus. Furthermore, this peak has a higher MLS score than the first peak when all 69 families are considered, thus suggesting that the susceptibility allele accounting for this peak may be more prevalent than that accounting for the first peak. The most prominent candidate genes for type 2 diabetes in the 1 LOD support interval are the RXR $\gamma$  (49),

for which we found an association with lipid abnormalities but a less prominent association with type 2 diabetes, and the overlapping homeobox transcription factor LMX1A (50). The microsatellite associations found in the present study also support one or more susceptibility genes that account for this peak. Marker D1S194, which was associated with type 2 diabetes in the case-control study, lies just telomeric to RXR $\gamma$  (162.06 Mb), whereas marker D1S1677 lies nearly 2 Mb telomeric to the 1 LOD CI (160.2 Mb). However, the only marker identified as overtransmitted in family members in a TDT-like test, marker ATA38A05, also lies within this second peak (162.5 Mb). We cannot exclude the possibility that one or more of the associations are spurious, particularly given the modest *P* values and the span of nearly 2.5 Mb between associated microsatellite markers. Additional SNP typing in these regions will be needed to confirm these associations and to narrow the genes responsible for these associations.

Although this study narrowed the most prominent linkage and association signals to the region between 156 and 168 Mb, we have previously demonstrated an association of multiple noncoding SNPs within the PKLR gene with type 2 diabetes (51), which is centromeric to the first linkage peak. This association would fall under the most centromeric linkage peak that was observed only on the unweighted sib-pair analysis (Fig. 3). The physical distance encompassed by this peak might extend from 117 Mb to at least 152.5 Mb. Among possible candidates in this region besides PKLR are RORC (52),  $\alpha$ -endosulfine (ENSA) (53), and interleukin-6 receptor (S.C.E., unpublished data). Of these candidates, a prominent association in this population was observed only with PKLR. Because of unusually strong linkage disequilibrium extending for large distances in this centromeric region, the actual genes accounting for the linkage peak and the association may lie at some physical distance from the observed association.

Were a single variant responsible for our linkage signal on 1q21-q24, we would expect to identify one haplotype of the microsatellite markers across the linkage peak that was shared among affected individuals. In contrast, in the region between D1S305 to D1S212, no single haplotype was shared. This finding is consistent with the existence of at least two and possibly three linkage peaks, suggesting more than one susceptibility gene in this region. The finding of several association peaks in this region offers further support for multiple susceptibility loci. We did identify a 1.16-cM region flanked by markers D1S2771 and D1S2705 in which affected siblings of 55% of the 58 sibships from the 40 families shared the same haplotype, but no single haplotype was shared, even in this narrow region. This finding is consistent with other common disease susceptibility genes and suggests that even within this first linkage peak, multiple at-risk haplotypes contribute to the linkage signal.

In summary, using combined linkage mapping, haplotype sharing and association studies with a dense marker map, we were able to confirm and narrow our original peak of linkage to a 3.3-cM region or ~2.1 Mb. We have resolved a second linkage peak that is ~10 cM telomeric to our largest peak but in a region of both association and linkage in other studies. Our analysis strongly suggests

that the replication in this region comes in part from the coalescence of several susceptibility loci in a region that could not be resolved on a 10-cM genome scan. The region harbors many strong candidate genes for type 2 diabetes, as well as a large number of poorly characterized transcripts that may also be good candidates. International collaborative efforts are underway to map these loci using positional candidate and linkage disequilibrium approaches in the populations with linkage to this region.

#### ACKNOWLEDGMENTS

This work was supported by grant DK39311 from the National Institutes of Health/NIDDK. Subject ascertainment was supported in part by the Research Service of the Department of Veterans Affairs, by the American Diabetes Association, and by National Institutes of Health/NCRR support of the General Clinical Research Centers of University of Arkansas for Medical Sciences (M01RR14288) and the University of Utah (M01RR03655). We thank Demond Williams and Winston Chu for technical assistance, Terri Hale and Judith Cooper for assistance with subject ascertainment, and the GCRC nursing and laboratory staff for assistance with subject assessment.

#### REFERENCES

- Elbein SC, Chiu KC, Permut MA: Type 2 diabetes mellitus. In *The Genetic Basis of Common Diseases*. King RA, Rotter JI, Motulsky AG, Eds. New York, Oxford University Press, 2002, p. 457–480
- Weyer C, Bogardus C, Mott DM, Pratley RE: The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest* 104:787–794, 1999
- Rich SS: Mapping genes in diabetes: genetic epidemiological perspective. *Diabetes* 39:1315–1319, 1990
- Elbein SC: Perspective: the search for genes for type 2 diabetes in the post-genome era. *Endocrinology* 143:2012–2018, 2003
- McCarthy MI: Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* 3:159–167, 2003
- Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hani CL, Bell GI: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163–175, 2000
- Bektas A, Hughes JN, Warram JH, Krolewski AS, Doria A: Type 2 diabetes locus on 12q15: further mapping and mutation screening of two candidate genes. *Diabetes* 50:204–208, 2001
- Ehm MG, Karnoub MC, Sakul H, Gottschalk K, Holt DC, Weber JL, Vaske D, Briley D, Briley L, Kopf J, McMillen P, Nguyen Q, Reisman M, Lai EH, Joslyn G, Shepherd NS, Bell C, Wagner MJ, Burns DK: Genomewide search for type 2 diabetes susceptibility genes in four American populations. *Am J Hum Genet* 66:1871–1881, 2000
- Mahtani MM, Widen E, Lehto M, Thomas J, McCarthy M, Brayer J, Bryant B, Chan G, Daly M, Forsblom C, Kanninen T, Kirby A, Kruglyak L, Munnely K, Parkkonen M, Reeve-Daly MP, Weaver A, Brettin T, Duyk G, Lander ES, Groop LC: Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat Genet* 14:90–94, 1996
- Fossey SC, Mychaleckyj JC, Pendleton JK, Snyder JR, Bensen JT, Hirakawa S, Rich SS, Freedman BI, Bowden DW: A high-resolution 6.0-megabase transcript map of the type 2 diabetes susceptibility region on human chromosome 20. *Genomics* 76:45–57, 2001
- Ghosh S, Watanabe RM, Hauser ER, Valle T, Magnuson VL, Erdos MR, Langefeld CD, Balow J Jr, Ally DS, Kohtamaki K, Chines P, Birznieks G, Kaleta HS, Musick A, Te C, Tannenbaum J, Eldridge W, Shapiro S, Martin C, Witt A, So A, Chang J, Shurtliff B, Porter R, Boehnke M: Type 2 diabetes: evidence for linkage on chromosome 20 in 716 Finnish affected sib-pairs. *Proc Natl Acad Sci U S A* 96:2198–2203, 1999
- Klupa T, Malecki MT, Pezzolesi M, Ji L, Curtis S, Langefeld CD, Rich SS, Warram JH, Krolewski AS: Further evidence for a susceptibility locus for type 2 diabetes on chromosome 20q13.1-q13.2. *Diabetes* 49:2212–2216, 2000
- Hanson RL, Ehm MG, Pettitt DJ, Prochazka M, Thompson DB, Timberlake D, Foroud T, Kobes S, Baier L, Burns DK, Almasy L, Blangero J, Garvey WT, Bennett PH, Knowler WC: An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* 63:1130–1138, 1998
- Elbein SC, Hasstedt SJ, Wegner K, Kahn SE: Heritability of pancreatic beta-cell function among nondiabetic members of Caucasian familial type 2 diabetic kindreds. *J Endocrinol Metab* 84:1398–1403, 1999
- 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
- 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, 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 U.K. 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
- Hsueh WC, St Jean PL, Mitchell BD, Pollin TI, Knowler WC, Ehm MG, Bell CJ, Sakul H, Wagner MJ, Burns DK, Shuldiner AR: Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the old order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. *Diabetes* 52:550–557, 2003
- Xiang K, Wang Y, Zheng T, Shen K, Jia W, Li J, Lin X, Wu S, Zhang G, Wang S, Lu H: Genome wide scan for type 2 diabetes susceptibility loci in Chinese (Abstract). *Diabetes* 51 (Suppl. 2):A262, 2002
- Meigs JB, Panhuysen CI, Myers RH, Wilson PW, Cupples LA: A genome-wide scan for loci linked to plasma levels of glucose and HbA<sub>1c</sub> in a community-based sample of Caucasian pedigrees: the Framingham Offspring Study. *Diabetes* 51:833–840, 2002
- Ng M, So W-Y, Lam V, Cockram C, Chan J: Identification of a susceptibility locus on chromosome 1q for metabolic syndrome traits in Hong Kong Chinese diabetic families (Abstract). *Diabetes* 52 (Suppl. 1):A35, 2003
- Coon H, Myers RH, Borecki IB, Arnett DK, Hunt SC, Province MA, Djousse L, Leppert MF: Replication of linkage of familial combined hyperlipidemia to chromosome 1q with additional heterogeneous effect of apolipoprotein A-I/C-III/A-IV locus: the NHLBI family heart study. *Arterioscler Thromb Vasc Biol* 20:2275–2280, 2000
- Pajukanta P, Nuotio I, Terwilliger JD, Porkka KV, Ylitalo K, Pihlajamaki J, Suomalainen AJ, Syvanen AC, Lehtimaki T, Viikari JS, Laakso M, Taskinen MR, Ehnholm C, Peltonen L: Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. *Nat Genet* 18:369–373, 1998
- Elbein SC, Hoffman MD, Teng K, Leppert MF, Hasstedt SJ: A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes* 48:1175–1182, 1999
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347–1363, 1996
- Elbein SC, Teng K, Yount P, Scroggin E: Linkage and molecular scanning analyses of MODY3/hepatocyte nuclear factor-1 alpha gene in typical familial type 2 diabetes: evidence for novel mutations in exons 8 and 10. *J Endocrinol Metab* 83:2059–2065, 1998
- 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
- Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, Kocher K, Miller K, Guschwan S, Kulbokas EJ, O'Leary S, Winchester E, Dewar K, Green T, Stone V, Chow C, Cohen A, Langelier D, Lapointe G, Gaudet D, Faith J, Branco N, Bull SB, McLeod RS, Griffiths AM, Bitton A, Greenberg GR, Lander ES, Siminovitch KA, Hudson TJ: Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 29:223–228, 2001
- Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML,

- Kong A, Stefansson K, Gulcher JR: The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* 35:131–138, 2003
29. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 63:861–869, 1998
  30. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissensbach J: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152–154, 1996
  31. McIndoe RA, Bumgarner RE, Welti R, Hood L: High throughput genotyping: practical considerations concerning day to day application. *SPIE* 2680:341–348, 1996
  32. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266, 1998
  33. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K: A high-resolution recombination map of the human genome. *Nat Genet* 31:241–247, 2002
  34. Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179–1188, 1997
  35. Kruglyak L, Lander ES: Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439–454, 1995
  36. Cottingham RW Jr, Idury RM, Schaffer AA: Faster sequential in genetic linkage computations. *Am J Hum Genet* 53:252–263, 1993
  37. Sham PC, Curtis D: Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Intern Med* 59:97–105, 1995
  38. Ewens WJ, Spielman RS: Disease associations and the transmission disequilibrium tests (TDT and S-TDT). *Current Protocols in Hum Genet* (Suppl. 20):1.12.1–1.12.15, 1999
  39. Hasstedt SJ: PAP: Pedigree Analysis Package, v. 5, 2001. Salt Lake City, UT, University of Utah, Department of Hum Genet, 2001
  40. Elbein SC, Chu W, Ren Q, Hemphill C, Schay J, Cox NJ, Hanis CL, Hasstedt SJ: Role of calpain-10 gene variants in familial type 2 diabetes in Caucasians. *J Endocrinol Metab* 87:650–654, 2002
  41. Sinsheimer JS, Blangero J, Lange K: Gamete-competition models. *Am J Hum Genet* 66:1168–1172, 2000
  42. Saarela J, Schoenberg FM, Chen D, Finnila S, Parkkonen M, Kuokkanen S, Sobel E, Tienari PJ, Sumelahti ML, Wikstrom J, Elovaara I, Koivisto K, Pirttila T, Reunanen M, Palotie A, Peltonen L: Fine mapping of a multiple sclerosis locus to 2.5 Mb on chromosome 17q22–q24. *Hum Mol Genet* 11:2257–2267, 2002
  43. Gauguier D, Froguel P, Parent V, Bernard C, Bihoreau MT, Portha B, James MR, Penicaud L, Lathrop M, Ktorza A: Chromosomal mapping of genetic loci associated with non-insulin dependent diabetes in the GK rat. *Nat Genet* 12:38–43, 1996
  44. Elbein SC, Chu W, Ren Q, Wang H, Hemphill C, Hasstedt SJ: Evaluation of apolipoprotein A-II as a positional candidate gene for familial type II diabetes, altered lipid concentrations, and insulin resistance. *Diabetologia* 45:1026–1033, 2002
  45. Wolford JK, Bogardus C, Ossowski V, Prochazka M: Molecular characterization of the human PEA15 gene on 1q21–q22 and association with type 2 diabetes mellitus in Pima Indians. *Gene* 241:143–148, 2000
  46. Wolford JK, Gruber JD, Ossowski VM, Vozarova B, Antonio TP, Bogardus C, Hanson RL: A C-reactive protein promoter polymorphism is associated with type 2 diabetes mellitus in Pima Indians. *Mol Genet Metab* 78:136–144, 2003
  47. Wolford JK, Hanson RL, Kobes S, Bogardus C, Prochazka M: Analysis of linkage disequilibrium between polymorphisms in the KCNJ9 gene with type 2 diabetes mellitus in Pima Indians. *Mol Genet Metab* 73:97–103, 2001
  48. Farook VS, Hanson RL, Wolford JK, Bogardus C, Prochazka M: Molecular analysis of KCNJ10 on 1q as a candidate gene for type 2 diabetes in Pima Indians. *Diabetes* 51:3342–3346, 2002
  49. Wang H, Chu W, Hemphill C, Hasstedt SJ, Elbein SC: Mutation screening and association of human retinoid X receptor  $\gamma$  variation with lipid levels in familial type 2 diabetes. *Mol Genet Metab* 76:3–12, 2002
  50. Thameem F, Wolford JK, Wang J, German MS, Bogardus C, Prochazka M: Cloning, expression and genomic structure of human LMX1A, and variant screening in Pima Indians. *Gene* 290:217–225, 2002
  51. Wang H, Chu W, Das SK, Ren Q, Hasstedt SJ, Elbein SC: Liver pyruvate kinase polymorphisms are associated with type 2 diabetes in northern European Caucasians. *Diabetes* 51:2861–2865, 2002
  52. Wang H, Chu W, Das SK, Zheng Z, Hasstedt SJ, Elbein SC: Molecular screening and association studies of retinoid-related orphan receptor gamma (RORC): a positional and functional candidate for type 2 diabetes. *Mol Genet Metab* 79:176–182, 2003
  53. Wang H, Craig RL, Schay J, Chu W, Das SK, Zhang Z, Elbein SC: Alpha endosulfine, a positional and functional candidate gene for type 2 diabetes: molecular screening, association studies, and role in reduced insulin secretion. *Mol Genet Metab*. In press