

A Common Polymorphism in the Upstream Promoter Region of the Hepatocyte Nuclear Factor-4 α Gene on Chromosome 20q Is Associated With Type 2 Diabetes and Appears to Contribute to the Evidence for Linkage in an Ashkenazi Jewish Population

Latisha D. Love-Gregory,¹ Jonathon Wasson,¹ Jiyang Ma,¹ Carol H. Jin,² Benjamin Glaser,³ Brian K. Suarez,^{2,4} and M. Alan Permutt¹

Variants in hepatocyte nuclear factor-4 α (HNF4 α), a transcription factor that influences the expression of glucose metabolic genes, have been correlated with maturity-onset diabetes of the young, a monogenic form of diabetes. Previously, in a genome scan of Ashkenazi Jewish type 2 diabetic families, we observed linkage to the chromosome 20q region encompassing HNF4 α . Here, haplotype-tag single nucleotide polymorphisms (htSNPs) were identified across a 78-kb region around HNF4 α and evaluated in an association analysis of Ashkenazi Jewish type 2 diabetic ($n = 275$) and control ($n = 342$) subjects. We found that two of nine htSNPs were associated with type 2 diabetes: a 3' intronic SNP, rs3818247 (29.2% case subjects vs. 21.7% control subjects; $P = 0.0028$, odds ratio [OR] 1.49) and a 5' htSNP located ~3.9 kb upstream of P2, rs1884614 (26.9% case subjects vs. 20.3% control subjects; $P = 0.0078$, OR 1.45). Testing of additional SNPs 5' of rs1884614 revealed a >10-kb haplotype block that was associated with type 2 diabetes. Conditioning on the probands' rs1884614 genotype suggested that the chromosomal region identified by the htSNP accounted for the linkage signal on chromosome 20q in families in which the proband carried at least one risk allele. Notably, the associations and the partitioned linkage profiles near P2 were independently observed in a Finnish sample, suggesting the presence of potential regulatory element(s) that may contribute to the risk for type 2 diabetes. *Diabetes* 53:1134–1140, 2004

From the ¹Division of Endocrinology, Diabetes and Metabolism, Washington University School of Medicine, St. Louis, Missouri; the ²Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri; the ³Endocrinology and Metabolism Service, Internal Medicine Department, The Hadassah-Hebrew University Medical Center, Jerusalem, Israel; and the ⁴Department of Genetics, Washington University School of Medicine, St. Louis, Missouri.

Address correspondence and reprint requests to M. Alan Permutt, MD, Division of Endocrinology, Diabetes and Metabolism, Washington University School of Medicine, St. Louis, MO 63110. E-mail: apermitt@im.wustl.edu.

Received for publication 10 September 2003 and accepted in revised form 13 January 2004.

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

dHPLC, denaturing high-performance liquid chromatography; FUSION, Finland-United States Investigation of NIDDM Genetics; htSNP, haplotype-tag single nucleotide polymorphism; HNF hepatocyte nuclear factor; HWE, Hardy-Weinberg equilibrium; LD, linkage disequilibrium; MODY, maturity-onset diabetes of the young; SNP, single nucleotide polymorphism.

© 2004 by the American Diabetes Association.

Regions defined by linkage to complex diseases typically encompass >10 cM, an area that may harbor hundreds of genes, thereby making the identification of disease-causing loci a tedious process. The study of populations that have undergone genetic isolation, as have the Ashkenazim, is thought to be useful in mapping complex disease genes. Taking into consideration that the larger American and European Caucasian populations originated from the Mediterranean basin as did the Ashkenazim, genetic risk factors identified in the Ashkenazi Jewish population may be important in other Caucasian populations (1,2). In a genome-wide scan of 267 multiplex type 2 diabetic Ashkenazi Jewish families, regions on chromosome 20 that exhibited nominal evidence for linkage ($P < 0.05$) were identified (3). The strongest linkage signal on chromosome 20q was observed at D20S195; a weaker signal on chromosome 20p was seen at D20S103. Several other type 2 diabetes studies have also identified linkage to chromosome 20q13.1-13.2 in Caucasian (4–7) and Japanese (8) families. Interestingly, the linkage peaks in these studies overlap at a region near the hepatocyte nuclear factor-4 α (HNF4 α) gene. The gene spans ~29 kb with 12 exons on chromosome 20q13.1-13.2 (9). It encodes for an orphan receptor member of the nuclear receptor superfamily 1.

HNF4 α variants have been shown to cosegregate in an autosomal-dominant manner in families with an atypical form of type 2 diabetes known as maturity-onset diabetes of the young (MODY)-1. MODY is a clinically and genetically heterogeneous form of nonketotic diabetes that presents before age 25 years, usually in nonobese, asymptomatic, hyperglycemic individuals (10–12). HNF4 α 's role in MODY stems from its function as a β -cell transcription factor that influences glucose-induced insulin secretion (13). In contrast to MODY, type 2 diabetes usually occurs between ages 40 and 60 years, with the exception of obesity-related pediatric type 2 diabetes, regardless of family history (14). Both MODY and type 2 diabetic patients have reduced insulin sensitivity as a result of pancreatic islet β -cell dysfunction. In addition, HNF4 α has

been shown to influence lipid transport and metabolism (15,16).

HNF4 α is differentially expressed in mammalian liver, kidney, small intestine, colon, stomach, and pancreas from as many as nine different transcripts (17,18). An alternative promoter, P2, lies 45.6 kb upstream of the proximal P1 promoter (18–20). P2-driven transcripts have been described as the predominant splice variant in pancreatic β -cells (18–21). Although *HNF4 α* intragenic and/or proximal P1 promoter single nucleotide polymorphisms (SNPs) have been described in previous type 2 diabetes studies (4,22–26), a thorough examination of the P2 region has not been reported; thus, association mapping was designed to examine the P2 region in this study.

Case-control studies of unrelated individuals have become the methodology of choice to follow up on linkage findings. The working hypothesis is that variants in linkage disequilibrium (LD) with the susceptibility locus will define the genomic region responsible for the original linkage signal. However, the extent of LD in various regions of genomic DNA has been shown to be highly variable (27–29). Recently, we reported (30) the ongoing evaluation of SNPs across a 7.3-Mb region near microsatellite D20S107 in an association study using pooled DNAs from 150 Ashkenazi Jewish type 2 diabetic patients and 150 control subjects. In the absence of a strong positive association between any of these SNPs and type 2 diabetes, we implemented a more direct candidate gene approach involving the *HNF4 α* gene in this study. We determined the patterns of LD and haplotype block structure to identify the number of haplotype-tag SNPs (htSNPs) required to capture the most common haplotypes across a 78-kb region encompassing *HNF4 α* and P2 preparatory to case-control analysis. An htSNP in the P2 region was associated with type 2 diabetes and appeared to be responsible for the previously defined linkage peak in families in which the probands carried at least one risk allele. Notably, similar findings were independently observed in a Finnish sample from the concurrent FUSION (Finland-United States Investigation of NIDDM Genetics) study (31; see this issue of *Diabetes*).

RESEARCH DESIGN AND METHODS

This study involved DNA from three independent sample sets: 1) 275 previously described multiplex families of Ashkenazi Jewish descent, with 1 affected individual from each family selected for case-control analysis (3); 2) 342 control subjects of Ashkenazi Jewish descent (150 of these were previously described [3] and an additional 192 DNA samples obtained from the National Laboratory for the Genetics of Israeli Populations, Tel Aviv University, Israel); and 3) an independent sample of 68 Ashkenazi Jewish individuals. For the case-control analysis, 275 probands from sample 1 and 342 control subjects from sample 2 were used. Sample 3 was used to assess LD and haplotype block structure. The institutional review boards of Washington University (St. Louis, MO) and Hadassah University Hospital (Jerusalem, Israel) approved the study.

Genotyping and single nucleotide polymorphism discovery. All genotypes in this study were assessed by PCR amplification of genomic DNA and Pyrosequencing technology (Pyrosequencing, Uppsala, Sweden), as previously described (32). To establish an SNP map spanning the 78 kb encompassing *HNF4 α* and P2, SNPs were ascertained by searching the public database, WAVE analysis, and/or sequencing. In all, 35 SNPs were identified and tested for validation through our efforts described below. To achieve an SNP density of 1 SNP every 5 kb, SNPs were dropped if they occurred <2 kb apart. In all, 19 SNPs, resulting in an average density of 1 SNP every 4.1 kb, were tested for Hardy-Weinberg equilibrium (HWE), assessment of minor allele frequency, and characterization of LD structure across the region in the sample of 68 unrelated Ashkenazi Jewish individuals (sample 3). SNPs with a

minor allele frequency of ≤ 0.09 in sample 3 were eliminated from further study.

From the public database, 12 SNPs (rs736821, rs10480819, rs3212210, rs1535337, rs1028583, rs2273618, rs3818247, rs1884614, rs2425639, rs4810424, rs1885088, and rs3761186) were selected and validated in samples of previously described DNA pools (30). Of these, nine SNPs (rs736821, rs3212210, rs1535337, rs1028583, rs3818247, rs1884614, rs2425639, rs1885088, and rs3761186) were further characterized in the 68 independent Ashkenazi Jewish individuals (sample set 3) for this study. An additional five public database SNPs (rs3761184, rs1884613, rs2144908, rs2425637, and rs2425640) were provided by Silander et al. (31).

To screen for SNPs by WAVE analysis (denaturing high-performance liquid chromatography [dHPLC]; Transgenomic, Omaha, NE), 15 primer sets were used to amplify by PCR all 12 exons (1,589 bp), flanking 5' and 3' intronic sequences (3,364 bp), the proximal promoter (800 bp), and 415 bp of the P2 promoter of the *HNF4 α* gene in 96 type 2 diabetic Ashkenazi Jewish probands (from sample 1). These PCR fragments were subsequently screened using a modification of the WAVE analysis. Because SNP-specific heteroduplex and homoduplex controls were not available, dHPLC peaks were visually scored for differences in WAVE patterns. Subsequently, one or two patients specific for each WAVE pattern were sequenced by dye-terminator chemistry (Applied Biosystems, Foster City, CA). A total of 14 variants were identified, of which 7 were novel (Table 1). Because many of the SNPs were <1 kb apart, only five (rs1800963, rs2071197, rs736823, rs3212195, and lg8100208) were further characterized.

To discover variants by direct sequencing, SNPs within the 45-kb gap between P2 and *HNF4 α* were identified as follows: 21 noncontiguous PCR fragments spanning a region 2.8 kb 5' of the P2 promoter to 13.5 kb upstream of *HNF4 α* 's proximal promoter, P1, were sequenced in eight randomly chosen Ashkenazi probands (from sample 1) using the ABI 3100 Avant Genetic Analyzer (Applied Biosystems). PCR fragments (size 0.5–1.0 kb) were spaced at 1-kb intervals and amplified using PCR primers intended to avoid poly A/T and repetitive elements. We identified nine novel SNPs, five of which had allele frequencies >0.09 (Y3, Y2, W1, R1, and R2) (online appendix Table 2, available at <http://diabetes.diabetesjournals.org>).

Statistical analysis. Statistical significance for type 2 diabetes SNP association was determined by Fisher's exact test, and the 95% CI was calculated using the approximation of Woolf (InStat version 3; GraphPad Software, San Diego, CA). *P* values were corrected for multiple tests using the Benjamini-Hochberg method (33). SNP genotype departures from HWE were tested using a χ^2 test with 1 degree of freedom.

Haplotypes were inferred using the Bayesian method as implemented in phase v1.0.1 (34). Phase-formatted data were run as a single file (case and control subjects combined) to allow for a more conservative estimation of haplotype frequency than would be obtained by separate case and control sample analyses. The program has the potential to optimize to each file separately, possibly skewing the haplotype frequencies. Several runs of phase were performed using the following parameters: iterations = 10 and 20 K, thinning intervals = 100 and 1,000, and burn-ins = 10 and 20 K.

Haplotype block structure was inferred by the greedy algorithm as implemented in HaploBlockFinder (35). In this program, the extent of LD was measured in terms of D' , d^2 , and r^2 (36,37). The significance of LD was assessed by the log likelihood ratio statistic under the assumption of HWE. HaploBlockFinder selects sets of SNPs defining 80% of the haplotypes (i.e., htSNPs) within a block based on r^2 , which represents absolute levels of LD. The parameters for HaploBlockFinder were as follows: block definition = minimum LD range; minimum $D' = 0.80$; genotype quality filter = 0.50 (ambiguous genotypes at a given locus can affect block partitioning, thus loci with ambiguous-to-total ratio genotypes with a threshold >0.50 are ignored); minor allele frequency (lowerbound) = 0.10; and coverage of htSNPs = 0.80–0.90.

Linkage analysis. To test the hypothesis that the "A" allele at SNP rs1884614 (or an allele in strong LD with it) is a risk factor for type 2 diabetes, we partitioned a sample of multiplex type 2 diabetic families that had previously been genotyped for up to 40 chromosome 20 microsatellites into subgroups according to the probands' genotype at rs1884614. The average heterozygosity of these 40 markers was 0.72. To protect against inadvertently including families with MODY, we required that the age at diagnosis of all affected pedigree members be >35 years. In all, 199 multiplex nuclear families met this inclusion criterion. Of these, 4 were maternal half-sibling families, 152 contained a pair of affected full siblings, 37 contained three affected siblings, and 6 contained an affected sibling quartet. A small number of additional non-first-degree genotyped relatives (two affected half-siblings and three affected first cousins) as well as 86 unaffected relatives (mostly siblings) were included in the linkage analysis. All linkage analyses were performed with Genehunter Plus using the "pairs" option under the exponential model (38,39).

TABLE 1

Type 2 diabetes association analysis of nine htSNPs and additional *HNF4 α* -related SNPs in unrelated Ashkenazi case and control subjects

kb position	SNP name	Major/minor allele	Minor allele frequency		Uncorrected <i>P</i>	Corrected <i>P</i>	OR (95% CI)
			Cases (<i>n</i> = 550)	Controls (<i>n</i> = 684)			
-14.647	rs3761184	A/G	0.17 (92)	0.20 (137)	0.141*	—	0.80 (0.60–1.07)
-9.422	rs4810424	C/G	0.26 (140)	0.20 (137)	0.024*	—	1.37 (1.04–1.79)
-4.030	rs1884613	C/G	0.27 (144)	0.21 (141)	0.017†	—	1.38 (1.06–1.80)
-3.926	rs1884614	G/A	0.27 (147)	0.20 (137)	0.0078	0.0352	1.45 (1.10–1.90)
1.272	rs2144908	G/A	0.27 (146)	0.20 (136)	0.006†	—	1.46 (1.12–1.91)
4.396	R1	T/C	0.22 (120)	0.25 (172)	0.2242	0.2883	0.84 (0.64–1.09)
10.685	R2-3	T/C	0.47 (252)	0.51 (344)	0.1659	0.2986	0.85 (0.68–1.07)
39.604	rs2425637	G/T	0.49 (269)	0.46 (313)	0.274†	—	1.14 (0.91–1.43)
43.065	rs2425639	A/G	0.49 (270)	0.46 (309)	0.2743	0.3086	1.14 (0.91–1.42)
43.634	rs2425640	G/A	0.31 (372)	0.32 (460)	0.757†	—	0.96 (0.75–1.23)
44.84	rs1800963	C/A	0.44 (241)	0.50 (330)	0.0424	0.1272	0.79 (0.62–0.99)
54.595	rs1885088	C/T	0.21 (111)	0.24 (159)	0.267†	—	0.85 (0.65–1.12)
58.69	rs3212195	G/A	0.21 (115)	0.24 (160)	0.1883	0.2700	0.86 (0.65–1.12)
66.316	rs1028583	G/T	0.31 (164)	0.26 (164)	0.0582	0.1310	1.29 (1.00–1.66)
73.035	rs3818247	G/T	0.29 (160)	0.22 (145)	0.0028	0.0252	1.49 (1.15–1.90)
74.766	rs3212210	G/T	0.16 (89)	0.16 (110)	1.0000	1.0000	0.99 (0.73–1.34)

Data are % (successfully genotyped chromosomes), unless otherwise noted. kb position indicates position relative to the *HNF4 α* P2 promoter translation initiation site at chromosomal 20q base position 43,622,874 of the human reference sequence (University of California, Santa Cruz [UCSC] Genome Browser, April 2003). *Additional SNPs genotyped to define the 5' boundary of the associated haplotype block "tagged" by rs1884614 †additional SNPs (originally tested for association by FUSION) genotyped in Ashkenazi Jewish case and control subjects to replicate the associations identified by the FUSION study (31).

All linkage analyses were carried out using marker positions as determined on the Marshfield map (<http://research.marshfieldclinic.org/genetics>).

Linkage analysis was initially performed on three subgroups, depending on the probands' genotype at rs1884614 ("AA" [*n* = 8], "AG" [*n* = 78], and "GG" [*n* = 113]). Visual inspection of the resulting logarithm of odds (LOD) score curves revealed that the "AA" and "AG" partitions were virtually identical (data not shown). Because the sample size of the "AA" subgroup was small, we decided to pool it with the "AG" partition. Accordingly, the comparison was between families with a proband who had at least one "A" allele and families where the proband lacked the "A" allele.

We carried out a randomization test to determine the significance of the partitioning. Subsamples of families (*n* = 86 and its complement [*n* = 199 – 86 = 113]) were drawn at random. For each subsample, we re-estimated the allele frequencies and performed a Genehunter Plus analysis, as above. A total of 10,000 randomizations were used to obtain empirical *P* values.

RESULTS

Genomic DNAs from 68 unrelated Ashkenazi Jewish individuals were genotyped at an average of one SNP every 4.1 kb across an ~78-kb region harboring *HNF4 α* and its

alternative upstream promoter, P2, using the 19 variants identified by extensive SNP discovery (see RESEARCH DESIGN AND METHODS for details). Of the 19 variants, 4 were subsequently eliminated because 2 did not conform to HWE and 2 occurred on <0.09 of the Ashkenazi chromosomes.

As shown in Fig. 1, the remaining 15 informative SNPs were used to determine the pattern of LD in this region. Phase software was used to estimate haplotypes, which were distributed into blocks and "tagged" according to user-defined parameters in HaploBlockFinder. The LD plot, shown in Fig. 2, illustrates seven haplotype blocks spanning the 78-kb region that were identified in the Ashkenazim. These seven blocks included three "singleton" blocks. A singleton refers to a single SNP that is not in LD with neighboring variants. As is seen in Fig. 2, the LD plot of *D'* indicates strong LD among neighboring SNPs across the P2 region and *HNF4 α* ; however, LD tended to

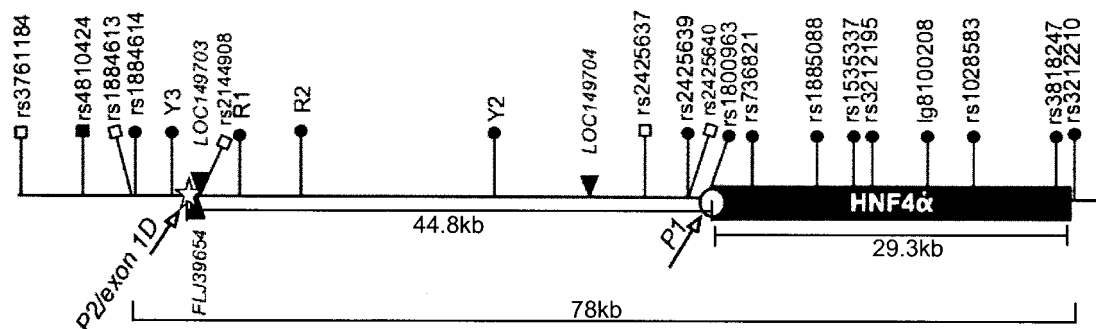


FIG. 1. SNP map of the 78-kb region including *HNF4 α* and P2. Shown is a schematic of the SNPs used to determine the extent of LD (SNPs shown as vertical lines with rounded tail) and identify evidence for association with type 2 diabetes (SNPs shown as vertical lines with square tail; □ indicates those SNPs originally tested for association to type 2 diabetes in the FUSION study) across the 78-kb *HNF4 α* region along chromosome 20q. *HNF4 α* is indicated by the elongated black box (~29 kb). *, Location of P2; ○, location of P1 promoter; ▲, location of predicted genes (FLJ39654, LOC149703, and LOC149704) in the 78-kb region. The location of the genes and predicted genes here are based on Entrez MapViewer (build 33).

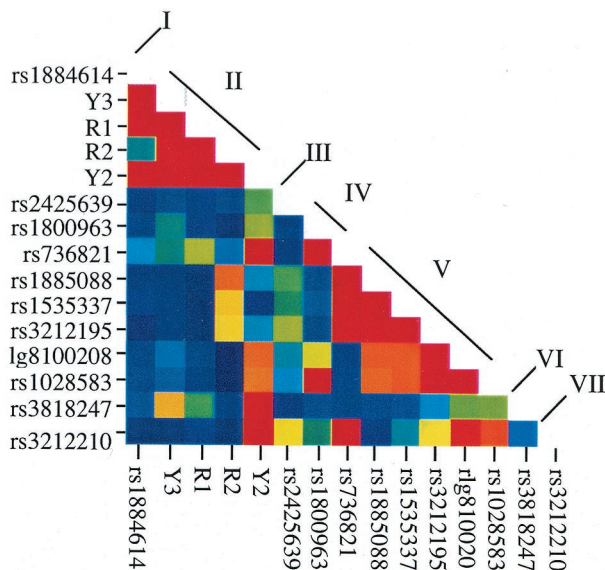


FIG. 2. A) Pairwise D' between all informative SNPs in Ashkenazi Jewish sample. D' values range from 0.0 to 1.0 (blue to red, respectively). Block structure indicated by I-VII. Single tick marks indicate "singleton" blocks as defined in HaploBlockFinder.

decay within the ~ 45 -kb gap separating *HNF4 α* from its alternative upstream promoter. Although the pattern of LD across this region was not striking due to the presence of the singleton blocks, this was not an unusual finding considering that LD and distance are semi-independent over short distances (40). In a separate study, we compared the LD pattern by genotyping these 15 SNPs in a sample of Centre d'Etude du Polymorphisme Humain individuals ($n = 34$) and found that the block structure was not significantly different. P2 and *HNF4 α* were distributed in the same blocks identified in the Ashkenazim, and the singleton blocks located within the 45-kb region were also observed (data not shown).

We evaluated seven haplotype blocks anchored by nine common polymorphisms (htSNPs) in 275 Ashkenazi Jewish type 2 diabetic probands and 342 control subjects (Table 1). Association with type 2 diabetes was observed with two htSNPs (Table 1). The minor allele frequency of htSNP rs3818247 (located in a 3' intronic region of *HNF4 α*) occurred on 29.2% of the proband chromosomes versus 21.7% of the control subject chromosomes ($P = 0.0028$, uncorrected). Subsequent analysis of the distribution of the probands' rs3818247 genotypes showed that the number of heterozygotes was greater than that expected by chance, consequently resulting in a failure to achieve HWE ($\chi^2 = 6.96$, $P < 0.01$). Accordingly, this SNP was dropped from further analysis. The minor allele frequency of haplotype tag rs1884614 (located ~ 3.9 kb 5' to P2) occurred in 26.9% of the case subjects vs. 20.3% of the control subjects ($P = 0.0078$, uncorrected). When corrected for multiple tests (i.e., nine htSNPs), the associations remained significant ($P < 0.05$). To determine the physical length of the associated haplotype block identified by htSNP rs1884614, we tested an additional SNP, rs4810424 (located upstream of rs1884614), and found it to be associated with type 2 diabetes. LD measures between these two SNPs indicated strong LD ($D' = 0.98$, $r^2 = 0.91$,

$d^2 = 0.85$ in the case subjects and $D' = 0.99$, $r^2 = 0.98$, $d^2 = 0.98$ in the control subjects).

In an independent association analysis of type 2 diabetes in a Finnish sample from the FUSION study (31), several additional SNPs (Fig. 1) were observed to be associated with type 2 diabetes. We tested five of these SNPs and an additional SNP, rs3761184, to further define the length of the associated haplotype block in the Ashkenazim (Table 1). The two P2 proximal SNPs (rs1884613 and rs2144908) were found to be associated with type 2 diabetes in the Ashkenazim. Furthermore, these SNPs were found to be in strong LD with rs1884614 ($D' = 0.98$, $r^2 = 0.95$, and $d^2 = 0.95$ between rs1884613 and rs1884614 in the probands; D' , r^2 , and $d^2 = 1.0$ between rs2144908 and rs1884614 in the probands). Consequently, results from both studies identified a haplotype block spanning >10 kb of DNA that was associated with type 2 diabetes. In contrast, the FUSION-associated SNPs located near P1 (promoter proximal to the gene) were not associated with type 2 diabetes in the Ashkenazi sample.

Figure 3 reports the LOD score curves for the total sample of 199 multiplex families and the two subpartitions. As can be seen, the profiles for the two partitions were dramatically different. Indeed, it appears that the "A+" partition accounted for virtually all of the linkage signal on 20q12-13 present in our earlier analysis (3). For all 199 families, the maximum LOD score of 2.01 was located at D20S195. In the "A+" partition (families with the risk allele), the maximum LOD (2.72) also occurred at D20S195. By contrast, the "A-" partition (families without the risk allele) attained a LOD score of 0.17 at D20S195.

Figure 4 reports the results of the randomization tests. The A+ partition resulted in a significant enhancement ($P < 0.05$) of the LOD scores over two broad intervals. The most significant interval covered ~ 16.5 cM (D20S470 to D20S107). The greatest difference occurred at D20S477, where only 0.29% of the randomizations resulted in a larger LOD score than that observed in the true partition. The second interval extended from D20S100 (84.78 cM) to the most distal microsatellite we genotyped (D20S171 at 95.7 cM). This interval lies well outside the region of linkage we originally reported (3) and could easily have been a false-positive. Figure 4 also reveals a significant enhancement in LOD scores in the A- partition over a 10-cM interval on 20p (D20S103 to D20S482), where we had previously reported a weak linkage signal (3). The findings for this interval, however, were less clear than for the interval near D20S477 on 20q because the interval on 20p is punctuated by two adjacent groups of two microsatellites, each with $P > 0.05$.

DISCUSSION

The genetic dissection of any complex heterogeneous disease, for which type 2 diabetes certainly qualifies, is a slow and arduous process. Having previously defined a linkage peak encompassing *HNF4 α* (1), the goal of this study was to examine 78 kb of the gene region by htSNP analysis. The approach taken here began with the identification of nine SNPs that define the common haplotypes encompassing the candidate region of *HNF4 α* and its alternative promoter. This was followed by case-control studies in which we identified an htSNP (rs1884614)

Chromosome 20

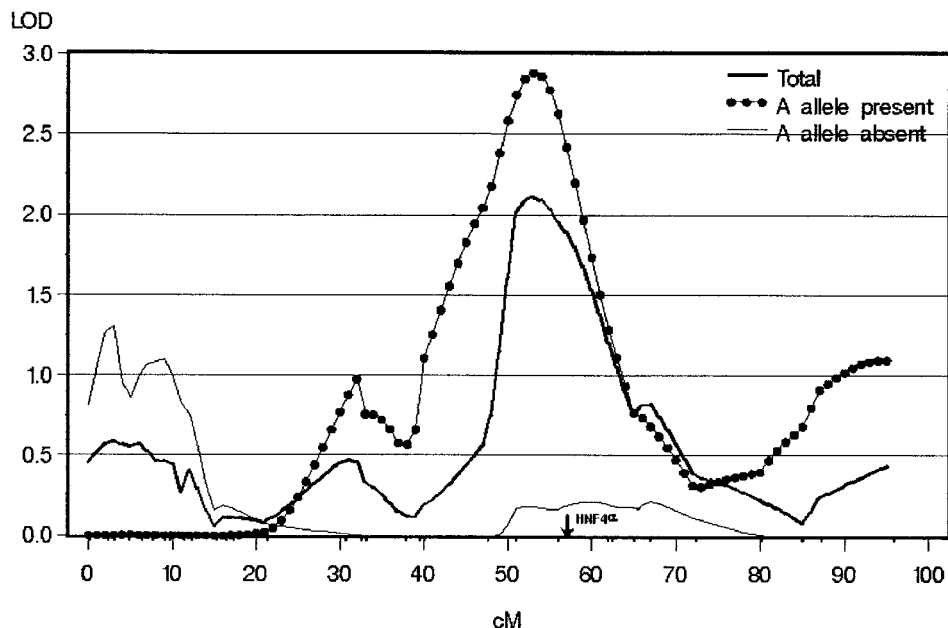


FIG. 3. LOD score curves for 199 multiplex type 2 diabetic families and two partitions defined by the proband's genotype at rs1884614 (A+, $n = 86$ families; A-, $n = 113$ families).

located in the P2 region that was associated with type 2 diabetes. Collaborative efforts between the Ashkenazi Jewish and FUSION studies led to the discovery of four SNPs (rs4810424, rs1884613, rs1884614, and rs2144908) located within a >10-kb haplotype block encompassing the P2 promoter that were associated with type 2 diabetes in both study populations. Furthermore, the risk attributed by each SNP, estimated by the ORs for the associated SNPs, was remarkably similar between the two study populations, lending further support to the evidence that this region contributes to the risk for type 2 diabetes. Although we found associations in common with the FUSION study in the P2 region, we did not replicate FUSION findings in the P1 region. Several of the allelic differences of the SNPs in case versus control subjects tended in the same direction in both groups, favoring

differences in sample size as a possible explanation. However, the absence of an association in the P1 region may have been due to a population-specific event in which the SNPs arose on different haplotypes in the two groups.

Significant SNPs were then tested to determine if they could resolve the etiologic heterogeneity by partitioning the families that provided the original linkage signal into homogeneous subgroups. The demonstration that partitioning our sample of multiplex families according to probands' genotype at rs1884614 gave rise to significantly different LOD score profiles is *prima facie* evidence that the "A" allele (or an allele in strong LD with it) is a potent genetic risk factor predisposing to type 2 diabetes. We note that the P2 promoter of *HNF4 α* was not located directly under our peak LOD score in the A+ partition. It would, indeed, be remarkable if the partitioning event

Chromosome 20

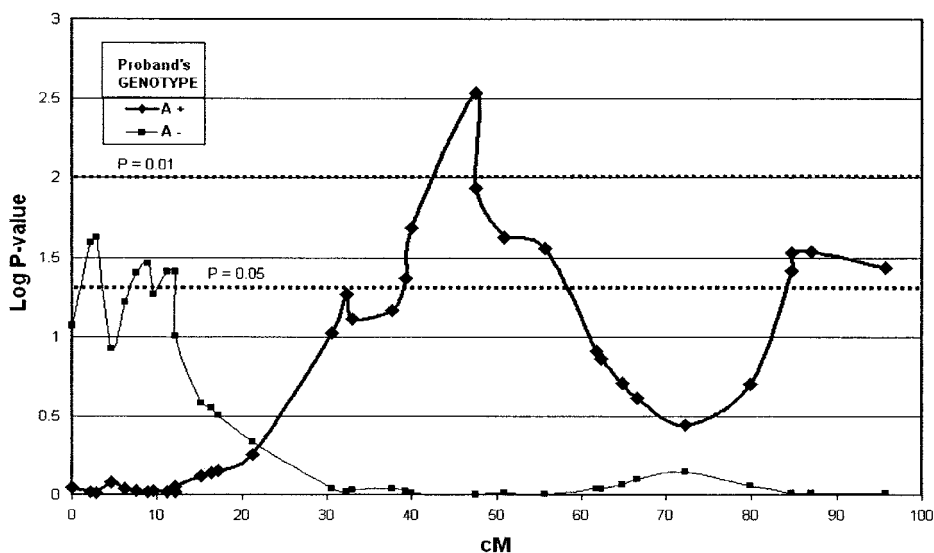


FIG. 4. Empirical P values obtained by randomly sampling A+ ($n = 86$) families and A- ($n = 113$) families. The P values are based on 10,000 randomizations and report for each microsatellite marker the number of randomizations that attained a LOD score higher than that obtained when the families were partitioned according to the proband's genotype at rs1884614. Some of the markers lie close to one another and are not clearly separated in the figure.

enhanced the LOD score in the *HNF4 α* interval (bounded in our data by D20S107 and D20S119) to a greater extent than in the more centromeric region where our original signal was maximized in these same families.

The remarkable similarity between our linkage partitioning findings and those reported by Silander et al. (31) for the FUSION study allows us to speculate that chromosome 20 may actually contain two distinct type 2 diabetes–predisposing regions. In our families, the strongest signal and the most significant partitioning occurred on 20q. The signal on 20p was less persuasive in terms of the absolute LOD scores. In addition, the interval on 20p is sufficiently distant from the location of the partitioning event at *HNF4 α* that it is unlikely that the effects of the partitioning could propagate over such a large distance. Nonetheless, the partitioning based on rs1884614 in our study or, equivalently, given the degree of LD, on rs2144908 in the FUSION study suggests that the partitioning appears to account for the original signal on 20q and that the region immediately upstream of the P2 promoter of *HNF4 α* is an important contributor to risk.

It is reasonable to suggest that any of the four associated SNPs flanking the P2 promoter could have functional implications. For example, the expression of *HNF4 α* P2-driven transcripts may be affected. Similarly, expression of adjacent hypothetical genes in the region may be affected. According to the current Entrez MapViewer (build 33), there are at least three predicted genes within the 78-kb region examined in this study (Fig. 1). These SNPs could be coding variants in yet-to-be-defined genes in this region. For example, SNP rs2144908 is positioned within the untranslated region of a predicted gene (FLJ39654) in which expressed sequence tags have been isolated from liver, kidney, and spleen. However, these predicted genes have not been described in pancreas.

In conclusion, it appears more likely that the four associated SNPs are regulatory variants or in LD with a coding or regulatory variant that predisposes to type 2 diabetes. These SNPs do not appear to be in linkage equilibrium with coding variants within *HNF4 α* , as extensive dHPLC and sequence analysis failed to identify common nonsynonymous SNPs. A likely explanation for our results is that these SNPs are in fact markers for a chromosomal region that regulates expression of either *HNF4 α* or one of the neighboring genes. This hypothesis can now be tested by measuring allele-specific transcription.

ACKNOWLEDGMENTS

This study was supported by National Institutes of Health Grants R01-DK-49583 and U01-DK-58026. The authors acknowledge the support of the Washington University Diabetes Research & Training Center for oligonucleotide supply.

We are greatly indebted to the investigators of the FUSION study for sharing their data before publication and helpful discussions of the manuscript. We thank Dr. Anthony Hinrichs for access to his 120 processor Beowulf-class computer cluster used to obtain the empiric *P* values from 10,000 randomizations. We also thank Mark Daly and Jeffrey Barrett of the Whitehead Institute for Biomedical Research for assistance with computing the measures of

LD and Kun Zhang of the Center for Genome Information at University of Cincinnati School of Medicine for assistance with the HaploBlockFinder software. Finally, the authors would like to thank Gary Skolnick for assistance with preparation of the manuscript.

REFERENCES

1. Santachiara Benerecetti AS, Semino O, Passarino G, Torroni A, Brdicka R, Fellous M, Modiano G: The common, Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. *Ann Intern Med* 57:55–64, 1993
2. Tikochinski Y, Ritte U, Gross SR, Prager EM, Wilson AC: mtDNA polymorphism in two communities of Jews. *Am J Hum Genet* 48:129–136, 1991
3. Permutt MA, Wasson JC, Suarez BK, Lin J, Thomas J, Meyer J, Lewitzky S, Rennich JS, Parker A, DuPrat L, Maruti S, Chayen S, Glaser B: A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* 50:681–685, 2001
4. 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, et al.: 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
5. Bowden DW, Sale M, Howard TD, Qadri A, Spray BJ, Rothschild CB, Akots G, Rich SS, Freedman BI: Linkage of genetic markers on human chromosomes 20 and 12 to NIDDM in Caucasian sib pairs with a history of diabetic nephropathy. *Diabetes* 46:882–886, 1997
6. Ji L, Malecki M, Warram JH, Yang Y, Rich SS, Krolewski AS: New susceptibility locus for NIDDM is localized to human chromosome 20q. *Diabetes* 46:876–881, 1997
7. Zouali H, Hani EH, Philippi A, Vionnet N, Beckmann JS, Demenais F, Froguel P: A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene. *Hum Mol Genet* 6:1401–1408, 1997
8. Mori Y, Otake S, Dina C, Yasuda K, Populaire C, Lecoecur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P: Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* 51:1247–1255, 2002
9. Argyrokastritis A, Kamakari S, Kapsetaki M, Kritis A, Talianidis I, Moschos NK: Human hepatocyte nuclear factor-4 (hHNF-4) gene maps to 20q12–q13.1 between PLCG1 and D20S17. *Hum Genet* 99:233–236, 1997
10. Winter WE: Newly defined genetic diabetes syndromes: maturity onset diabetes of the young. *Rev Endocr Metab Disord* 4:43–51, 2003
11. Stride A, Hattersley AT: Different genes, different diabetes: lessons from maturity-onset diabetes of the young. *Ann Med* 34:207–216, 2002
12. Fajans SS, Bell GI, Polonsky KS: Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* 345:971–980, 2001
13. Byrne MM, Sturis J, Fajans SS, Ortiz FJ, Stoltz A, Stoffel M, Smith MJ, Bell GI, Halter JB, Polonsky KS: Altered insulin secretory responses to glucose in subjects with a mutation in the MODY1 gene on chromosome 20. *Diabetes* 44:699–704, 1995
14. Aye T, Levitsky LL: Type 2 diabetes: an epidemic disease in childhood. *Curr Opin Pediatr* 15:411–415, 2003
15. Bartoov-Shifman R, Hertz R, Wang H, Wollheim CB, Bar-Tana J, Walker MD: Activation of the insulin gene promoter through a direct effect of hepatocyte nuclear factor 4 alpha. *J Biol Chem* 277:25914–25919, 2002
16. Wang H, Maechler P, Antinozzi PA, Hagenfeldt KA, Wollheim CB: Hepatocyte nuclear factor 4alpha regulates the expression of pancreatic beta-cell genes implicated in glucose metabolism and nutrient-induced insulin secretion. *J Biol Chem* 275:35953–35959, 2000
17. Nakhei H, Lingott A, Lemm I, Ryffel GU: An alternative splice variant of the tissue specific transcription factor HNF4alpha predominates in undifferentiated murine cell types. *Nucleic Acid Res* 26:497–504, 1998
18. Boj SF, Parrizas M, Maestro MA, Ferrer J: A transcription factor regulatory circuit in differentiated pancreatic cells. *Proc Natl Acad Sci U S A* 98:14481–14486, 2001
19. Thomas H, Jaschowitz K, Bulman M, Frayling TM, Mitchell SM, Roosen S, Lingott-Frieg A, Tack CJ, Ellard S, Ryffel GU, Hattersley AT: A distant upstream promoter of the HNF-4alpha gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum Mol Genet* 10:2089–2097, 2001
20. Hansen SK, Parrizas M, Jensen ML, Pruhova S, Ek J, Boj SF, Johansen A,

- Maestro MA, Rivera F, Eiberg H, Andel M, Lebl J, Pedersen O, Ferrer J, Hansen T: Genetic evidence that HNF-1 α -dependent transcriptional control of HNF-4 α is essential for human pancreatic beta cell function. *J Clin Invest* 110:827–833, 2002
21. Eeckhoutte J, Moerman E, Bouckennooghe T, Lukoviak B, Pattou F, Formstecher P, Kerr-Conte J, Vandewalle B, Laine B: Hepatocyte nuclear factor 4 alpha isoforms originated from the P1 promoter are expressed in human pancreatic beta-cells and exhibit stronger transcriptional potentials than P2 promoter-driven isoforms. *Endocrinology* 144:1686–1694, 2003
 22. Moller AM, Urhammer SA, Dalggaard LT, Reneland R, Berglund L, Hansen T, Clausen JO, Lithell H, Pedersen O: Studies of the genetic variability of the coding region of the hepatocyte nuclear factor-4 α in Caucasians with maturity onset NIDDM. *Diabetologia* 40:980–983, 1997
 23. Hani EH, Suaud L, Boutin P, Chevre JC, Durand E, Philippi A, Demenais F, Vionnet N, Furuta H, Velho G, Bell GI, Laine B, Froguel P: A missense mutation in hepatocyte nuclear factor-4 alpha, resulting in a reduced transactivation activity, in human late-onset non-insulin-dependent diabetes mellitus. *J Clin Invest* 101:521–526, 1998
 24. Price JA, Fossey SC, Sale MM, Brewer CS, Freedman BI, Wuerth JP, Bowden DW: Analysis of the HNF4 alpha gene in Caucasian type II diabetic nephropathic patients. *Diabetologia* 43:364–372, 2000
 25. Malecki MT, Antonellis A, Casey P, Ji L, Wantman M, Warram JH, Krolewski AS: Exclusion of the hepatocyte nuclear factor 4 α as a candidate gene for late-onset NIDDM linked with chromosome 20q. *Diabetes* 47:970–972, 1998
 26. Sakurai K, Seki N, Fujii R, Yagui K, Tokuyama Y, Shimada F, Makino H, Suzuki Y, Hashimoto N, Saito Y, Egashira T, Matsui K, Kanatsuka A: Mutations in the hepatocyte nuclear factor-4 α gene in Japanese with non-insulin-dependent diabetes: a nucleotide substitution in the polypyrimidine tract of intron 1b. *Horm Metab Res* 32:316–320, 2000
 27. Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO: Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet* 68:191–197, 2001
 28. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES: Linkage disequilibrium in the human genome. *Nature* 411:199–204, 2001
 29. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF: Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293:489–493, 2001
 30. Permutt MA, Wasson J, Love-Gregory L, Ma J, Skolnick G, Suarez B, Lin J, Glaser B: Searching for type 2 diabetes genes on chromosome 20. *Diabetes* 51 (Suppl 3):S308–S315, 2002
 31. Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS: Genetic variation near the hepatocyte nuclear factor-4 α gene predicts susceptibility to type 2 diabetes. *Diabetes* 53:1141–1149, 2004
 32. Wasson J, Skolnick G, Love-Gregory L, Permutt MA: Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. *Biotechniques* 32:1144–1146, 1148, 1150 passim, 2002
 33. Sabatti C, Service S, Freimer N: False discovery rate in linkage and association genome screens for complex disorders. *Genetics* 164:829–833, 2003
 34. Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989, 2001
 35. Zhang K, Jin L: HaploBlockFinder: haplotype block analyses. *Bioinformatics* 19:1300–1301, 2003
 36. Awadalla P, Eyre-Walker A, Smith JM: Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286:2524–2525, 1999
 37. Devlin B, Risch N: A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29:311–322, 1995
 38. Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179–1188, 1997
 39. 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
 40. Dunning AM, Durocher F, Healey CS, Teare MD, McBride SE, Carlomagno F, Xu CF, Dawson E, Rhodes S, Ueda S, Lai E, Luben RN, Van Rensburg EJ, Mannernmaa A, Kataja V, Rennart G, Dunham I, Purvis I, Easton D, Ponder BA: The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am J Hum Genet* 67:1544–1554, 2000