

# Genetic Variation Near the Hepatocyte Nuclear Factor-4 $\alpha$ Gene Predicts Susceptibility to Type 2 Diabetes

Kaisa Silander,<sup>1</sup> Karen L. Mohlke,<sup>1</sup> Laura J. Scott,<sup>2</sup> Erin C. Peck,<sup>1</sup> Pablo Hollstein,<sup>1</sup> Andrew D. Skol,<sup>2</sup> Anne U. Jackson,<sup>2</sup> Panagiotis Deloukas,<sup>3</sup> Sarah Hunt,<sup>3</sup> George Stavrides,<sup>3</sup> Peter S. Chines,<sup>1</sup> Michael R. Erdos,<sup>1</sup> Narisu Narisu,<sup>1</sup> Karen N. Conneely,<sup>2</sup> Chun Li,<sup>2</sup> Tasha E. Fingerlin,<sup>2</sup> Sharanjeet K. Dhanjal,<sup>4</sup> Timo T. Valle,<sup>5,6</sup> Richard N. Bergman,<sup>7</sup> Jaakko Tuomilehto,<sup>5,6,8</sup> Richard M. Watanabe,<sup>4</sup> Michael Boehnke,<sup>2</sup> and Francis S. Collins<sup>1</sup>

The Finland-United States Investigation Of NIDDM Genetics (FUSION) study aims to identify genetic variants that predispose to type 2 diabetes by studying affected sibling pair families from Finland. Chromosome 20 showed our strongest initial evidence for linkage. It currently has a maximum logarithm of odds (LOD) score of 2.48 at 70 cM in a set of 495 families. In this study, we searched for diabetes susceptibility variant(s) at 20q13 by genotyping single nucleotide polymorphism (SNP) markers in case and control DNA pools. Of 291 SNPs successfully typed in a 7.5-Mb interval, the strongest association confirmed by individual genotyping was with SNP rs2144908, located 1.3 kb downstream of the primary  $\beta$ -cell promoter P2 of hepatocyte nuclear factor-4 $\alpha$  (*HNF4A*). This SNP showed association with diabetes disease status (odds ratio [OR] 1.33, 95% CI 1.06–1.65,  $P = 0.011$ ) and with several diabetes-related traits. Most of the evidence for linkage at 20q13 could

be attributed to the families carrying the risk allele. We subsequently found nine additional associated SNPs spanning a 64-kb region, including the P2 and P1 promoters and exons 1–3. Our results and the independent observation of association of SNPs near the P2 promoter with diabetes in a separate study population of Ashkenazi Jewish origin suggests that variant(s) located near or within *HNF4A* increases susceptibility to type 2 diabetes. *Diabetes* 53:1141–1149, 2004

**E**vidence for a type 2 diabetes locus at chromosome 20q12-q13 (OMIM no. 603694) has been found in several Caucasian (1–5) and Asian (6,7) study populations. This region harbors the transcription factor hepatocyte nuclear factor-4 $\alpha$  (*HNF4A*) gene, mutations of which cause maturity-onset diabetes of the young (MODY) type 1, a dominantly inherited, early-onset type 2 diabetes characterized by defective glucose-dependent insulin secretion (OMIM no. 125850) (8). *HNF4A* has a complex expression pattern, which includes elaborate alternative splicing (9–11), and is expressed in many tissues, including the liver and pancreas (9). Three of the isoforms are transcribed by an alternative P2 promoter, located ~46 kb upstream of the P1 promoter and the rest of the coding exons (10,12). Transcripts from both the P1 and P2 promoters have been detected in pancreatic  $\beta$ -cells (11), but the P2 promoter is suggested to be the primary transcription start site in these cells (10,12). Mutations in *HNF4A* have been identified in MODY type 1 families in both coding and regulatory regions of the gene (13), including the P2 promoter region (10). Among its varied roles, HNF4A is implicated in glucose homeostasis in both the pancreatic  $\beta$ -cells and liver. In the  $\beta$ -cells, HNF4A regulates expression of genes involved in glucose metabolism and insulin secretion (14,15) and directly activates insulin gene expression (16). In liver, HNF4A plays a role in gluconeogenesis (17).

In a genome-wide scan based on 478 affected sibling pair (ASP) families from the Finland-United States Investigation of NIDDM Genetics (FUSION 1) study, the strongest linkage results were on chromosome 20 (4,18), with maximum logarithm of odds (LOD) scores (MLs) of 1.99, 2.04, and 2.15 at 18, 57, and 70 cM, respectively, on the FUSION

From the <sup>1</sup>Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland; the <sup>2</sup>Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan; <sup>3</sup>The Wellcome Trust Sanger Institute, Hinxton, U.K.; the <sup>4</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; the <sup>5</sup>Diabetes and Genetic Epidemiology Unit, Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland; the <sup>6</sup>Department of Biochemistry, National Public Health Institute, Helsinki, Finland; the <sup>7</sup>Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California; and the <sup>8</sup>Department of Public Health, University of Helsinki, Helsinki, Finland.

Address correspondence and reprint requests to Michael Boehnke, PhD, Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029. E-mail: boehnke@umich.edu.

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

K.S. and K.L.M. contributed equally to this work.

The current affiliation for K.S. is with the Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland. The current affiliation for C.L. is with the Department of Molecular Physiology and Biophysics, Program in Human Genetics, Vanderbilt University, Nashville, Tennessee.

Posted on the World Wide Web at <http://diabetes.diabetesjournals.org> on 9 March 2004.

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

ASP, affected sibling pair; EC, elderly control; FUSION, Finland-United States Investigation of NIDDM Genetics; GIST, Genotype–Identity–By–Descent Sharing Test; HNF, hepatocyte nuclear factor; IBD, identity by descent; LD, linkage disequilibrium; LOD, logarithm of odds; MLS, maximum LOD score; MODY, maturity-onset diabetes of the young; OGTT, oral glucose tolerance test; SNP, single nucleotide polymorphism.

© 2004 by the American Diabetes Association.

genetic map. Following fine mapping with microsatellite markers and using 495 updated FUSION 1 families, our highest MLS on chromosome 20 was 2.48 at 70 cM, with a 1-LOD support interval flanked by markers D20S861 and D20S897 (20q13.12–20q13.13), a region of 6.7 Mb (19). We found no evidence for linkage in this region when studying an independent set of 242 Finnish ASP families (FUSION 2), although the lack of linkage could be partially explained by the smaller FUSION 2 sample size (19). In the present study, we aimed to identify evidence for disease susceptibility variant(s) at 20q13.12–20q13.13 by typing a dense set of single nucleotide polymorphism (SNP) markers, using a DNA pool–based, case-control study design.

DNA pools can be used effectively to estimate differences in SNP allele frequencies between case and control populations (20–22). We applied this method in the fine mapping of the 20q13.12–20q13.13 region, successfully typing 291 SNPs, using pools of type 2 diabetes case and control subjects. We identified an SNP, rs2144908, ~1.3 kb downstream of the P2 promoter of *HNF4A*, which shows association with diabetes disease status. We subsequently identified seven additional associated SNPs spanning a 59-kb region that includes both the P2 and P1 promoters and exons 1–3. Interestingly, in a study sample of Ashkenazi Jewish origin, Love-Gregory et al. (23; see this issue of *Diabetes*) independently identified two type 2 diabetes-associated SNPs near the P2 promoter. The associated SNPs were reciprocally tested in both samples, identifying four SNPs flanking the P2 promoter and in near perfect linkage disequilibrium (LD) with each other that are associated with diabetes disease status in both study samples. These results further implicate *HNF4A* as a likely candidate gene for common type 2 diabetes and advocate additional studies of the regulatory region of this gene.

## RESEARCH DESIGN AND METHODS

The FUSION linkage analysis sample consisted of 737 Finnish ASP families collected in two phases. The FUSION 1 (F1) set included 495 families with 1,129 affected individuals, and the FUSION 2 (F2) set included 242 families with 580 affected individuals (18,19,24). A single affected individual was chosen from each family for case-control and phenotype association analyses. Also, we included an additional 37 FUSION 1 and 21 FUSION 2 affected individuals from families excluded from linkage analysis because they did not have a genotyped affected sibling in the study. We used a diabetes age-of-onset criteria of >35 years in the recruitment of families to exclude likely cases of type 1 diabetes and MODY. For control subjects, we studied 225 elderly Finnish subjects, who had normal glucose tolerance by oral glucose tolerance test (OGTT) at ages 65 and 70 years. In addition, we used a control group of 189 normoglycemic spouses of FUSION 1 index case subjects or affected siblings (24,25). To evaluate association with diabetes-related traits, we also studied 467 unaffected offspring from 185 FUSION 1 families.

Informed consent was obtained from each study participant, and the study protocol was approved by the ethics committee or institutional review board of each of the participating centers.

**Genotyping DNA pools and individuals.** We studied two case DNA pools and two control DNA pools, constructed as previously described (21). Our FUSION 1 case pool included one affected individual from each of 499 FUSION 1 families; our FUSION 2 case pool included one affected individual from each of 249 FUSION 2 families. Our spouse control pool included 182 normoglycemic spouses, and our elderly control (EC) pool included 228 elderly, normal glucose-tolerant individuals. The numbers of individuals in each of the pools differ slightly from the number of individuals used for individual typing because of DNA availability and slightly different inclusion criteria.

We identified available SNPs from public databases, mapped them to unique locations within our region of interest using a combination of electronic PCR (26), RepeatMasker (27), and MegaBlast (28) and annotated the SNPs with validation status, allele frequency, and gene structure informa-

tion. We tested SNPs on a crude pool of ~400 individuals and selected SNPs exhibiting apparent minor allele frequencies >0.05 to genotype on our four carefully quantitated case and control DNA pools. We typed 146 SNPs on the FUSION 1 case pool and both control pools, and 235 additional SNPs on all four pools. The pooled genotyping using mass spectrometry, allele frequency estimation, quality criteria for acceptance of data, and *P* values for comparisons of individual case and control pool allele frequencies have been previously described (21). In this study, we also generalized the test of comparisons between individual pools to compare the allele frequencies in the combined case pools (FUSION 1 and 2) to the combined control pools (EC and spouse control). Of 291 successfully typed SNPs, we selected SNPs for follow-up from those that exhibited allele frequency differences between case and control pools >0.05 or pool-specific *P* values <0.05.

SNPs were typed on individual DNA samples by homogeneous MassExtend reaction using the MassARRAY System (Sequenom, San Diego, CA) as previously described (21). Three of the 62 SNPs genotyped on individual samples were not consistent with Hardy-Weinberg equilibrium: rs878559 (*P* = 0.045), rs6017309 (*P* = 0.020), and rs6093978 (*P* = 0.049); this number of failures is consistent with the number expected by chance, and these SNPs were included in the analysis. On average, 98.6% of attempted genotypes were successful. Among 4,958 blinded duplicate genotype pairs, we observed three discrepancies, for an estimated genotyping reproducibility of 99.97%.

Once the first SNP exhibiting association with diabetes status was identified, additional SNPs were selected for genotyping using five sources. First, 18 SNPs were selected from public databases. Second, 33 SNPs were selected that captured the variability found in haplotype blocks generated from SNPs genotyped by the Sanger Institute in the *HNF4A* region (see below). Third, six SNPs were selected from among eight variants identified by variant screening of the P2 promoter region (see below); two variants have not yet been genotyped on cases and controls due to repetitive sequence surrounding the SNPs. Fourth, one SNP in intron 4 of *HNF4A* was selected for typing after showing association in our previous variant screening effort of the *HNF4A* gene (4). Fifth, four SNPs were genotyped to assess evidence for associations observed independently by Love-Gregory et al. (23).

**Selecting haplotype-tagging SNPs for the *HNF4A* region.** We obtained genotype data for 979 SNPs in a 10-Mb interval, having a minor allele frequency >0.10, and previously typed in the founders of 12 Centre d'Etude du Polymorphisme Humain families and a U.K. sample (P.D., S.H., G.S., manuscript submitted). Sixty-seven SNPs mapped to the region –145 to 102 kb relative to the translation initiation site (the A of the ATG is defined as +1) in exon 1 of the P2 promoter of *HNF4A* and were used to identify blocks of SNPs in strong LD. We used the method of Dawson et al. (29), which adds an SNP to a block when it has an LD value  $D' \geq 0.85$  with the current haplotypes, requires that the five most common predicted haplotypes account for  $\geq 75\%$  of all haplotypes, and allows up to five intervening markers to be excluded per block. The 67 SNPs clustered into 19 overlapping blocks of limited haplotype diversity. We selected haplotype-tagging SNPs that were able to predict the most common haplotypes using the method of Johnson et al. (30) as implemented in Stata. Thirty-three SNPs accounted for 95% of the predicted haplotype diversity within the region.

**Variant screening.** Based on functional studies of the P2 promoter region (10), we attempted to screen for variants in a 2,310-bp region spanning the P2 promoter of *HNF4A*, from –2,196 to 144 relative to the translation initiation site. We performed denaturing high-performance liquid chromatography using a Wave DNA Fragment Analysis system (Transgenomic, Omaha, NE), followed by direct sequencing of PCR products on 12 individuals with type 2 diabetes from families showing evidence for linkage at 20q13.13 and 11 EC individuals. Due to the presence of repeat sequences, we were able to successfully screen 1,835 of the 2,310 bp for variants: 263 bp between positions –2,196 and –1,934, and 1,572 bp between positions –1,458 and 114.

**Statistical analysis of data.** Disease-marker association and odds ratio (OR) calculations were performed using logistic regression with genotypes coded as none, one, and two copies of the minor allele that corresponds to a multiplicative model on the OR scale. We also tested SNP rs2144908 for association with diabetes-related quantitative traits, which were derived from clinical measures, OGTTs, and frequently sampled intravenous glucose tolerance tests with minimal model analysis (24). Details regarding phenotypes have been previously published (19,24,25). Trait differences among genotypes were determined by ANOVA within case subjects, unaffected spouses, EC subjects, and the unaffected offspring of case subjects. Trait values were transformed to approximate univariate normality and adjusted for age (except for age at diagnosis and duration), sex, and BMI (except for weight-related variables and waist-to-hip ratio). A generalized estimating equations approach was used for the offspring data in order to account for the correlation among related individuals (31). Analyses were performed assuming additive, dominant, recessive, and general (2 degrees of freedom) genetic models. We report

TABLE 1  
Allele frequencies and evidence for association in case and control study samples for SNPs in the *HNF4A* region

kb position*	SNP name	Major/minor allele	Minor allele frequency ( <i>n</i> typed)				<i>P</i> value†	Multiplicative OR (95% CI)‡
			F1	F2	F1 + F2	Control		
-9.422	rs4810424§	G/C	0.20 (530)	0.20 (259)	0.20 (789)	0.16 (409)	0.014	1.32 (1.05–1.64)
-4.030	rs1884613	C/G	0.21 (526)	0.21 (261)	0.21 (787)	0.16 (410)	0.010	1.34 (1.07–1.66)
-3.926	rs1884614§	C/T	0.20 (516)	0.20 (260)	0.20 (776)	0.16 (402)	0.017	1.31 (1.04–1.63)
1.272	rs2144908	G/A	0.21 (530)	0.21 (263)	0.21 (793)	0.16 (413)	0.011	1.33 (1.06–1.65)
5.269	rs6031551	T/C	0.19 (527)	0.18 (262)	0.18 (789)	0.22 (401)	0.030	0.79 (0.64–0.98)
5.349	rs6031552	C/A	0.19 (527)	0.18 (263)	0.18 (790)	0.23 (409)	0.015	0.77 (0.63–0.95)
39.604	rs2425637	G/T	0.53 (528)	0.52 (262)	0.53 (790)	0.47 (406)	0.014	1.23 (1.04–1.47)
43.592	rs2425640	G/A	0.37 (525)	0.34 (261)	0.36 (786)	0.41 (410)	0.021	0.82 (0.68–0.97)
44.840	rs1800963	C/A	0.42 (529)	0.44 (260)	0.43 (789)	0.45 (409)	0.448	0.93 (0.79–1.11)
50.693	rs3212183	T/C	0.42 (515)	0.43 (238)	0.43 (753)	0.38 (400)	0.040	1.20 (1.01–1.44)
54.595	rs1885088	G/A	0.14 (531)	0.16 (262)	0.15 (793)	0.11 (409)	0.017	1.37 (1.05–1.75)
66.316	rs1028583§	G/T	0.43 (512)	0.44 (263)	0.43 (775)	0.45 (392)	0.372	0.92 (0.78–1.10)
73.035	rs3818247§	G/T	0.32 (522)	0.32 (261)	0.32 (783)	0.34 (391)	0.236	0.89 (0.75–1.07)

\*Kilobase position from the *HNF4A* P2 promoter translation initiation site at chromosome 20, base position 43,669,874 of the human reference sequence (UCSC Genome Browser, July 2003); †calculated for F1 + F2 case vs. control subjects; ‡calculated for F1 + F2 case vs. control subjects for the minor allele in control; §SNPs originally tested for association by Love-Gregory et al. (23); ||SNP associated with disease status in Ashkenazi sample (23). F1, FUSION 1; F2, FUSION 2.

the results of the score test. We did not correct the *P* values for the number of traits examined or tests performed.

We performed permutation tests to assess overall significance when accounting for the number of correlated tests. For each of the four groups tested, we created 10,000 permuted samples by randomly shuffling genotypes (leaving intact the vector of correlated traits, including sex and age). For the unaffected offspring, randomization was performed at the sibship level to preserve family structure. For each permuted sample, we performed the tests of association described above assuming additive, dominant, and recessive models for all traits and compared the number of significant associations in the permuted samples with the number of significant associations originally found. We computed the overall *P* value for each group as the proportion of permuted samples having at least as many tests with *P* less than or equal to the largest significant *P* value in the original set of comparisons. To estimate an overall *P* value for the four groups combined, we randomly merged the permutations from all four groups and used the same approach.

Haplotype frequencies were estimated for the combined case and control samples using the expectation-maximization algorithm. Haplotype frequencies and counts for the separate case and control samples were estimated based on the multilocus genotypes of each individual and the estimated haplotype frequencies in the combined sample. All *n* haplotypes were jointly tested for association with disease status using a  $2 \times n \chi^2$  test of independence. Haplotypes with counts  $\leq 10$  were pooled. Each individual haplotype was tested for association with disease status using a  $2 \times 2 \chi^2$  test of independence. To account for the variability introduced by the haplotype frequency estimation, significance was assessed by permuting case and control status and recalculating the test statistic 10,000 times.

We performed linkage analysis using the  $S_{\text{pairs}}$  statistic as programmed in Genehunter Plus (32). Each family-specific statistic was weighted by the square root of the number of affected individuals minus one. To examine the evidence for linkage in families with the risk allele of an associated SNP, families were divided into two subsets based on the presence or absence of the associated allele in the single genotyped affected family member (492 FUSION 1 and 732 FUSION 1 + 2 genotyped individuals). Linkage analysis was run for each subset and for the combined samples. Evidence for association between identity-by-descent (IBD) sharing and possession of the risk allele was also investigated using the Genotype-IBD Sharing Test (GIST) (33). GIST assigns family-specific weights based on the genotype of the affected family members and the model of interest (dominant, recessive, and additive) and tests for a positive correlation between this weight and the family-based IBD sharing as represented by the nonparametric linkage score.

Pairwise marker LD measure  $D'$  was estimated using the ldmx program of Abecasis and Cookson (34). A pictogram describing pairwise marker LD was created using Matlab (The MathWorks, Natick, MA).

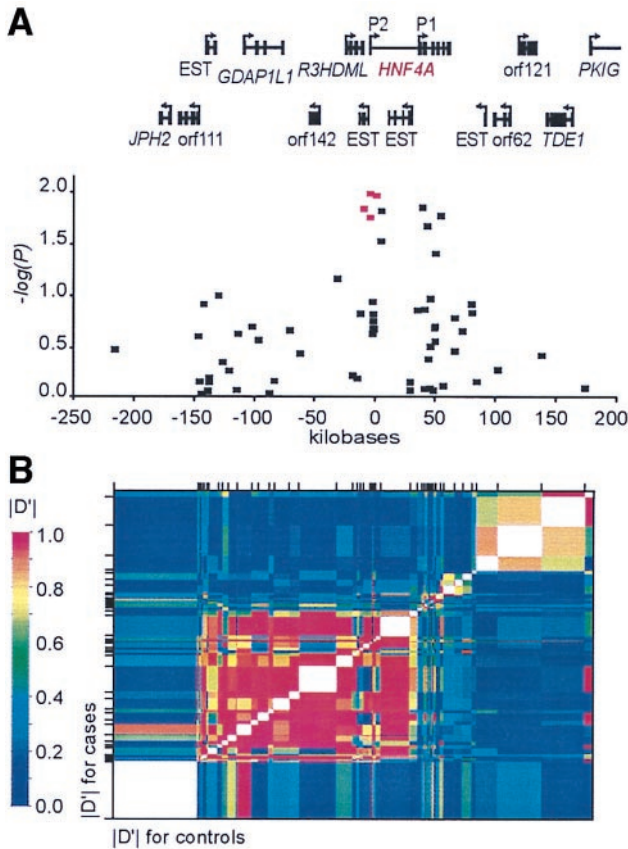
## RESULTS

To fine map our chromosome 20q13 type 2 diabetes linkage peak, we selected SNPs between markers D20S96 and D20S196, a region of 12.1 cM and 7.4 Mb, slightly

larger than our 1-LOD linkage support interval. We successfully assayed on DNA pools 291 SNPs and tested them for association with disease status in the case and control pools (21). Of 21 SNPs selected for follow-up individual genotyping, rs2144908 exhibited the strongest evidence for association based on individual genotypes, with identical allele frequencies for both FUSION 1 and 2 case subjects (21%), a lower allele frequency in control subjects (16%), and OR 1.33 ( $P = 0.011$ , 95% CI 1.06–1.65) for the combined FUSION 1 and 2 sets.

SNP rs2144908 is located within the *HNF4A* gene, 1.3 kb downstream from the translation initiation site of alternate exon 1 of the P2 promoter (10,12). Given this location, we further evaluated the region for evidence of association with diabetes status. In an  $\sim 390$ -kb region spanning *HNF4A*, we genotyped 18 additional SNPs selected from public databases and 33 SNPs selected to represent the common haplotypes detected in haplotype blocks. We also genotyped six of eight variants (online appendix Table 1 [available at <http://diabetes.diabetesjournals.org>]) identified in a 1,835-bp region surrounding the P2 promoter shown to harbor regulatory elements affecting gene expression (10). SNPs showing the strongest evidence for association with diabetes status are listed in Table 1. Association results for all SNPs studied are shown in Fig. 1. Genotype counts and *P* values for FUSION 1 and 2 case and control subjects for all SNPs studied are detailed in the online appendix (Table 2 [available at <http://diabetes.diabetesjournals.org>]).

We initially identified seven additional noncoding SNPs significantly associated with diabetes status, spanning a 59-kb region that includes the P2 and P1 promoters and exons 1–3 of *HNF4A* (Table 1, Fig. 2). The two SNPs that showed strongest association with disease status, rs1884613 (OR 1.34,  $P = 0.010$ ) and rs2144908 (1.33,  $P = 0.011$ ), flank the P2 promoter and alternate exon 1 and are in near perfect LD with each other (online appendix Fig. 1 [available at <http://diabetes.diabetesjournals.org>]). After learning that a sample of Ashkenazi origin also showed evidence for association to diabetes status in the *HNF4A* region (23), we genotyped four additional SNPs assessed



**FIG. 1.** Association of the *HNF4A* region with type 2 diabetes and intermarker LD. **A:**  $-\log_{10} P$  values of the differences in allele frequencies between type 2 diabetic case ( $n = 795$ ) and control ( $n = 414$ ) subjects were plotted against physical distance. At the top, exons of known and predicted genes and expressed sequence tags are shown, with the direction of transcription indicated by arrows. The two promoters of *HNF4A* are labeled P1 and P2. The location of the four SNPs that are associated with diabetes status in both Finns and Ashkenazi (23) samples are colored red. **B:** The pairwise marker LD between SNPs. The axes are scaled by marker distance (kb), and each SNP location is indicated by a tick mark. The LD between a pair of markers is indicated by the color of the block above and to the right of the intersection of the markers.

in the Ashkenazi sample, to enable a direct comparison. Our results for rs4810424 and rs1884614 confirmed the Ashkenazi association, and these SNPs were within 6 kb of and in nearly perfect LD with rs1884613 and rs2144908 in our sample, expanding the region harboring associated SNPs to 64 kb (Table 1, Fig. 2, and online appendix Fig. 1 [available at <http://diabetes.diabetesjournals.org>]). In contrast, SNPs near the P1 promoter and coding exons 1–3 (rs2425637, rs2425640, and rs1885088) were associated in our Finnish sample but not associated in the Ashkenazi sample (23). Likewise, we did not confirm the Ashkenazi evidence for association for rs1800963 (typed in our *HNF4A*-region fine-mapping step), rs1028583, and rs3818247 (Table 1).

The pattern of LD in the 390-kb region surrounding *HNF4A* is shown in Fig. 1. The P2 promoter and the most strongly associated SNPs are located within a region exhibiting  $D' > 0.6$  for most marker pairs, which extends from  $-147$  to  $30$  kb relative to the *HNF4A* translation initiation site in alternate exon 1. LD extends for more limited distances in the region containing the P1 promoter and most *HNF4A* exons, from  $45$  to  $75$  kb. A region of

moderate LD is located telomeric to *HNF4A*, from  $84$  to  $174$  kb.

We tested whether haplotypes were associated with type 2 diabetes (Table 2). We evaluated SNPs representing those individually associated with diabetes status by including only one SNP from any set of associated SNPs with  $r^2 > 0.95$ . Using a general  $2 \times n$  test for independence, we found significant differences in haplotype frequencies between case and control subjects ( $P = 0.004$ ) for the 21 haplotypes observed at least 10 times in the combined case and control subjects. All haplotypes containing the rs2144908 risk allele (A) had an estimated OR  $> 1$ . The haplotype containing all associated alleles (ACTGCA) had the highest OR (2.11) ( $P = 0.023$ ) but was rare. Strong LD was observed between the first two SNPs (representing the six SNPs closest to the P2 promoter) and the last four SNPs of the haplotype, whereas there was very little LD between the second and third (rs6031552 and rs2425637) SNPs. The frequency of haplotypes containing the associated alleles was consistent with that expected based on random recombination between the second and third SNPs. We also observed a significantly associated haplotype (GCTGCA) (OR 1.51,  $P = 0.010$ ) containing the risk alleles for all SNPs except rs2144908, a nominally protective haplotype containing the nonrisk alleles for all SNPs (GAGATG), and a haplotype with a significantly protective effect containing all nonrisk alleles except rs6031552 (GC GATG).

Based on the presence of the “A” risk allele for rs2144908 in the single genotyped affected individual, we identified 186 of the 492 FUSION 1 families and 270 of the 732 combined FUSION 1 and 2 families as at-risk families. Linkage analysis on at-risk, not-at-risk, and all families (Fig. 3) demonstrated that nearly all of the evidence for

**TABLE 2**  
Haplotype frequencies and evidence for association in F1 + F2 case and control study samples for associated unique SNPs.

rs2144908	rs6031552	rs2425637	rs2425640	rs3212183	rs1885088	Frequency*			
						Case	Control	OR‡	P value§
A	C	T	G	C	A	0.016	0.008	<b>2.11</b>	<b>0.023</b>
G	C	T	G	C	A	0.093	0.063	<b>1.51</b>	<b>0.010</b>
A	C	G	A	T	G	0.036	0.026	<b>1.41</b>	0.087
A	C	T	G	T	G	0.031	0.024	<b>1.28</b>	0.293
A	C	G	G	T	G	0.014	0.011	<b>1.27</b>	0.475
G	C	T	G	C	G	0.142	0.122	<b>1.20</b>	0.109
A	C	T	G	C	G	0.081	0.071	<b>1.14</b>	0.369
G	A	T	G	T	G	0.027	0.025	<b>1.08</b>	0.732
G	C	G	G	T	G	0.047	0.044	<b>1.07</b>	0.715
A	C	G	A	T	A	0.015	0.015	<b>1.01</b>	0.968
G	C	T	G	T	G	0.071	0.078	0.91	0.504
G	C	G	A	C	G	0.010	0.011	0.89	0.732
G	C	G	A	T	G	0.225	0.270	0.78	<b>0.013</b>
G	A	G	G	T	G	0.056	0.071	0.78	0.131
G	A	T	G	C	G	0.047	0.063	0.73	0.058
G	A	G	A	T	G	0.039	0.054	0.71	0.051

\*Haplotypes with estimated frequencies  $> 1\%$  in case and control subjects and the haplotype containing all associated alleles; †risk alleles are shaded; ‡ORs  $> 1$  are in boldface; §P values  $< 0.05$  are in boldface. F1 + F2, FUSION 1 and 2 samples.

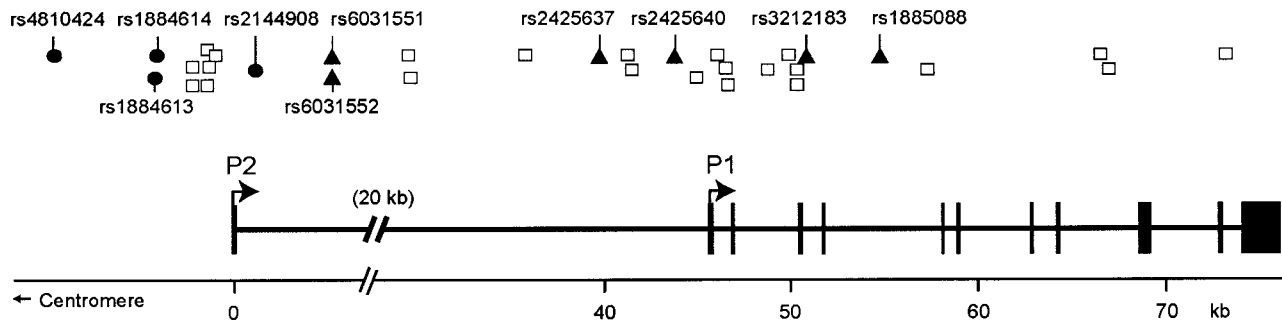


FIG. 2. *HNF4A* gene structure and location of SNPs studied. ●, SNPs associated with diabetes disease status in both Finn and Ashkenazi samples (23); ▲, SNPs associated only in Finns; □, SNPs not associated with disease status.

linkage in FUSION 1 at 64 cM was observed in the at-risk families (MLS 1.86), with little evidence in the not-at-risk families (MLS 0.28). A similar trend was seen in the combined FUSION 1 and 2 samples. Interestingly, the risk allele had an even stronger effect on linkage evidence in the centromeric peak region at 52.5 cM (FUSION 1 sample, at-risk families versus not-at-risk families, MLS 3.07 vs. 0.02, respectively). In contrast, almost all evidence for 20p linkage was observed in the not-at-risk families. Using GIST to test the SNPs showing significant disease-marker association, only those surrounding the P2 promoter and in (near) perfect LD with rs2144908 accounted for a significant fraction of the FUSION 1 linkage evidence.

Dominant and additive models showed greater ability to explain the linkage signal (rs2144908,  $P = 0.04$  and  $0.05$ , respectively) than the recessive model ( $P = 0.41$ ), although none were significant after accounting for the three-way comparison. In the combined FUSION 1 and 2 sample, evidence that these SNPs could account for a portion of the linkage signal did not reach significance (rs2144908,  $P = 0.09$ ,  $0.24$ , and  $0.89$  for the dominant, additive, and recessive models, respectively), although the result for the dominant model is still suggestive.

We also examined whether rs2144908 was associated with diabetes-related quantitative traits in unaffected offspring, unaffected spouses, EC subjects, and FUSION 1 and 2 case subjects (Table 3 and online appendix Table 3 [available at <http://diabetes.diabetesjournals.org>]). Unaffected offspring (34.9 ± 7.4 years of age) showed the strongest phenotype associations (Table 3). The 14 unaffected offspring homozygous for the risk allele exhibited significantly lower BMI and higher fasting insulin, 2-h glucose and insulin, and  $\Delta$  glucose and insulin (2 h – fasting) from the OGTT (Table 3, recessive model). Under the dominant and additive models, offspring with the risk allele had lower acute insulin response to glucose (AIRg) and disposition index from the frequently sampled intravenous glucose tolerance test. There was no association with insulin resistance, fasting lipids, or blood pressure. Even after accounting for the number of correlated tests performed, we observed more significant associations for unaffected offspring than would be expected by chance ( $P = 0.007$ ). Fewer associations were detected for other groups (online appendix Table 3 [available at <http://diabetes.diabetesjournals.org>]), but the number of associations was marginally significant for unaffected spouses ( $P = 0.062$ ) and EC subjects ( $P = 0.053$ ), although not for affected case subjects ( $P = 0.743$ ). For the four groups combined, more significant associations were observed than would be expected by chance ( $P = 0.003$ ).

#### DISCUSSION

Using a DNA pool-based, case-control study approach on chromosome 20, we identified a variant near the P2 promoter of *HNF4A* that was significantly associated with type 2 diabetes in our Finnish study sample. We further scrutinized this region and initially identified a total of eight SNPs significantly associated ( $P < 0.05$ ) with diabetes disease status, spanning a 59-kb region that includes the P2 and P1 promoters of *HNF4A*. Love-Gregory et al. (23) independently identified two SNPs associated with

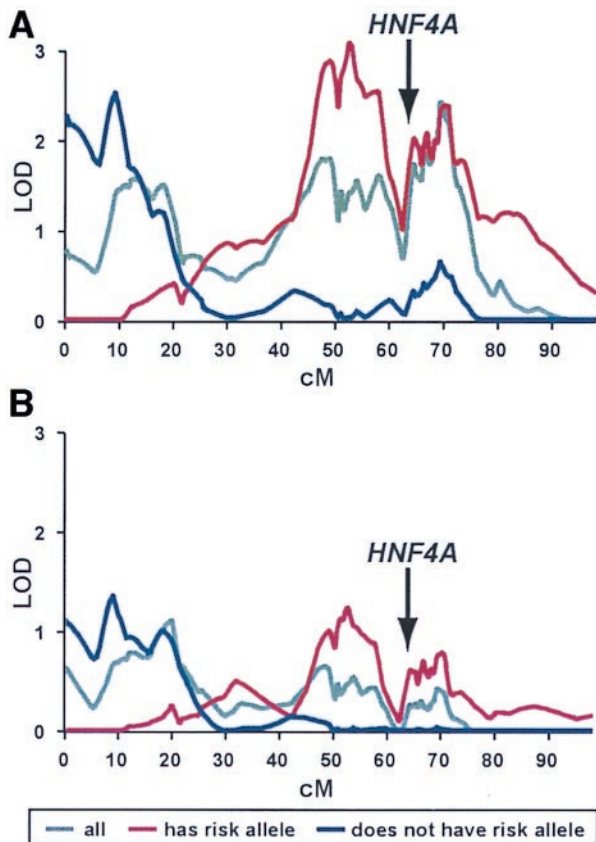


FIG. 3. Linkage evidence for chromosome 20 by presence/absence of the risk allele for marker rs2144908 in a single affected family member. A: Four-hundred ninety-two FUSION 1 families (gray line), 186 at-risk families (red line), and 306 not-at-risk families (blue line). B: Seven-hundred thirty-two combined FUSION 1 and 2 families (gray line), 270 at-risk families (red line), and 462 not-at-risk families (blue line).

TABLE 3

Phenotype associations with rs2144908 in unaffected offspring under recessive, dominant, additive, and general (2 degrees of freedom) genetic models.

	AA	AG	GG	P value*			2 df
				Recessive	Dominant	Additive	
<i>n</i>	14	141	316				
BMI (kg/m <sup>2</sup> )	22.9 (4.0)	25.1 (5.1)	25.6 (5.7)	<b>0.007</b>	0.358	0.091	<b>0.021</b>
Fasting glucose (mmol/l)	4.96 (0.30)	5.02 (0.52)	4.97 (0.67)	0.812	0.142	0.133	0.322
2-h glucose (mmol/l)	5.72 (1.53)	5.25 (1.74)	5.11 (1.86)	<b>0.025</b>	<b>0.050</b>	<b>0.017</b>	<b>0.029</b>
Δ Glucose (mmol/l)†	0.81 (1.26)	0.28 (1.80)	0.05 (1.82)	<b>0.029</b>	0.055	<b>0.019</b>	<b>0.036</b>
Fasting insulin (pmol/l)	69.0 (30.0)	60.0 (42.0)	60.0 (42.0)	<b>0.013</b>	0.426	0.983	<b>0.020</b>
2-h insulin (pmol/l)	345 (432)	228 (330)	231 (288)	<b>0.010</b>	0.441	0.139	<b>0.037</b>
Δ Insulin (pmol/l)†	276 (426)	174 (276)	168 (270)	<b>0.019</b>	0.371	0.122	0.062
Insulin sensitivity (10 <sup>-5</sup> · min <sup>-1</sup> · pmol <sup>-1</sup> · l)‡	8.6 (8.18)	6.13 (5.07)	6.87 (6.02)	0.684	0.319	0.395	0.555
AIRg (10 <sup>-2</sup> · pmol · l <sup>-1</sup> · 8 min)§	1.74 (1.88)	1.74 (1.90)	1.89 (1.86)	0.403	<b>0.009</b>	<b>0.023</b>	<b>0.022</b>
Disposition index (×10 <sup>-3</sup> )	11.6 (13.6)	10.7 (11.7)	12.5 (10.9)	0.583	<b>0.003</b>	<b>0.008</b>	<b>0.007</b>

Data are unadjusted median values (interquartile range). \*P values calculated for age- and sex-adjusted data and, for traits not involving weight, also for BMI-adjusted data; P values <0.05 are in boldface; †difference between 2-h and fasting values; ‡derived from minimal model analysis of frequently sampled intravenous glucose tolerance tests; §computed from the frequently sampled intravenous glucose tolerance test as the incremental area under the insulin curve from 0 to 10 min; ||product of insulin sensitivity and AIRg, a measure of β-cell function. AIRg, acute insulin response to glucose.

type 2 diabetes near the P2 promoter region in a sample of Ashkenazi Jewish origin. These SNPs are in near perfect LD with the two FUSION SNPs showing strongest evidence for association, rs2144908 and rs1884613. Including these two SNPs, we have observed in our sample a total of 10 associated SNPs in a 64-kb region that spans the P2 and P1 promoters and exons 1–3 of *HNF4A*. The associations observed for SNPs closer to the P1 promoter and exons 1–3 of the gene were not replicated in the Ashkenazi sample. Together these results suggest that variant(s) near the P2 promoter of *HNF4A* increase susceptibility to type 2 diabetes.

The four SNPs that showed evidence for association in both Finns and Ashkenazim span >10 kb around the *HNF4A* P2 promoter. Individually, these SNPs appear to explain a significant portion of the observed evidence for FUSION 1 linkage. One or more of the four associated SNPs described here may directly influence diabetes susceptibility, or these SNPs may be in LD with the true susceptibility variant(s). None of these SNPs are located in confirmed transcription factor binding sites or sequences conserved with mouse and rat (35), suggesting that other, as yet unknown, SNPs may be the causal variant(s). The location of the associated SNPs in a 177-kb region of LD including >140 kb upstream of the P2 promoter and only ~30 kb within *HNF4A* (but not including most coding exons) (Fig. 1) suggests that susceptibility SNP(s) could be located anywhere in this large interval; the lack of other associated SNPs outside of the P2 region could be due to low SNP density. The larger region of LD contains several other known and predicted genes and expressed sequence tags (Fig. 1). Although it is possible that the true susceptibility variant(s) may affect these genes or expressed sequence tags, our current (lack of) knowledge of their function makes them much less attractive candidates than *HNF4A*.

Both the Finns and Ashkenazim show evidence of association with SNPs located near the *HNF4A* P1 pro-

motor and coding exons. However, unlike the SNPs flanking the P2 promoter region, SNPs near the P1 promoter and coding exons do not show the same patterns of association in the Finnish and Ashkenazi samples, even though their locations overlap. It is thus possible that there is a single shared susceptibility variant, but recombination and/or other factors contributing to disequilibrium, such as genetic drift, population history and structure, and gene conversion (36), have resulted in different patterns of association in the two study populations. Our haplotype analysis and an examination of the LD plots indicate that in the Finnish population there is substantial recombination between the associated SNPs located in the P1 and coding exon regions and the SNPs in the P2 region. Thus, it is also possible that there are actually two (or more) disease-predisposing alleles, and only the one near the P2 promoter is shared in common between the Ashkenazi and Finnish samples. Further SNP identification, genotyping, and study of additional samples and populations will be necessary to delimit the region that could harbor the susceptibility SNP(s). Also, our finding of association between variants near or within the *HNF4A* gene and type 2 diabetes does not rule out the possibility that other susceptibility genes are present at 20q13.12-13.13.

When testing for association between diabetes-related traits and the risk allele for SNP rs2144908, we found more significant phenotype results than we would have expected by chance in the unaffected offspring and in the four groups combined. We found traits related to insulin secretion to be associated with the risk allele, consistent with the known function of *HNF4A* and the P2 promoter. Offspring homozygous for the risk allele had higher fasting, 2-h, and Δ insulin concentrations from the OGTT; however, these were in the presence of higher glucose levels. Careful examination of the relationship between Δ glucose and Δ insulin from the OGTT reveals that offspring with the risk allele had a lower plasma insulin response to

glucose. Furthermore, offspring with at least one copy of the risk allele had lower acute insulin response to glucose and disposition index, both measures of  $\beta$ -cell function. Because we did not measure fractional hepatic insulin extraction or directly assess hepatic glucose production, we cannot fully dismiss an effect of this putative variant at the level of the liver. However, the lack of difference in insulin sensitivity between genotypes suggests an effect at the level of the  $\beta$ -cell, and not of the liver. More direct clinical investigation of individuals with the at-risk allele is justified. The observed associations with insulin secretion-related traits suggest abnormal insulin secretory response to oral or intravenous glucose and are consistent with our previous observation of strong heritability of such traits in FUSION 1 families (37).

Mutations in *HNF4A* are well documented to be a cause of the MODY phenotype (8,13). A few previous studies have also identified variants in *HNF4A* that appeared significantly associated with common type 2 diabetes in small study samples (38) or in single families (39,40). To our knowledge, however, none of these have been confirmed to be associated in additional study samples. Other studies have sought but not identified significant association between *HNF4A* variants and common type 2 diabetes (41–43), including variants within the P2 promoter and alternate exon 1 (44). We previously screened the P1 promoter region and exons of *HNF4A* and identified one variant showing borderline significant association with disease status (4). This variant was not significantly associated with disease status when analyzing all of our case and control subjects (SNP rs736824, online appendix Table 2 [available at <http://diabetes.diabetesjournals.org>]). In our current study, we studied a new set of SNPs, mostly located at the P2 promoter region and within introns.

*HNF4A* interacts with multiple transcription factors, including *HNF-1 $\alpha$*  (*HNF1A*) and *-1 $\beta$* , and insulin promoter factor 1 (10,12); variants in each cause specific subtypes of MODY (8). It is possible that coding or noncoding variants in other MODY genes could influence type 2 diabetes susceptibility. In fact, a G319S mutation in *HNF1A* was significantly associated with early-onset type 2 diabetes, which is distinct from MODY (45). Also, evidence for linkage at chromosome 12q24 near the *HNF1A* gene has been found in a Finnish sample and in several other studies (46). *HNF4A* also regulates and interacts with several other proteins that have a major role in the maintenance of glucose homeostasis and that have been implicated in type 2 diabetes pathogenesis, such as peroxisome proliferator-activated receptor- $\gamma$  (47) and peroxisome proliferator-activated receptor coactivator-1 $\alpha$  (48–50).

In summary, we found four variants in a >10-kb region that includes the P2 promoter of *HNF4A*, which are associated with type 2 diabetes and explain, in part, the evidence for linkage in our sample and in an Ashkenazi Jewish sample (23). Additional variant identification and genotyping, study of other populations, and, finally, functional studies will be needed to identify the true type 2 diabetes susceptibility variant(s) at 20q13.

#### ACKNOWLEDGMENTS

The FUSION study was funded by intramural funds from the National Human Genome Research Institute project

OH95-C-N030 and by National Institutes of Health (NIH) Grants HG00376 and DK62370 to M.B. J.T. has been partially supported by the Academy of Finland (38387 and 46558). K.L.M. is the recipient of a Burroughs Wellcome Career Award in the Biomedical Sciences. T.E.F., K.N.C., and A.D.S. have been partially supported by NIH training Grant HG00040. R.N.B. was supported by NIH Grants DK27619 and DK29867. R.M.W. is supported by the American Diabetes Association.

We are indebted to the Finnish individuals who volunteered to participate in our study. We thank Konstantinos Lazaridis, Amy Carver, Lori Bonnycastle, Abby Woodroffe, Peggy White, and Darryl Leja for their technical assistance. We also thank all past and present FUSION investigators for their hard work and enthusiasm. We thank two anonymous reviewers for their helpful suggestions.

#### REFERENCES

- 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
- 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
- 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
- Ghosh S, Watanabe RM, Hauser ER, Valle T, Magnuson VL, Erdos MR, Langefeld CD, Balow J, Ally DS, Kohtamaki K, Chines P, Birznieks G, Kaleta H-S, Musick A, Te C, Tannenbaum J, Eldridge W, Shapiro S, Martin C, Witt A, So A, Chang J, Shurtleff B, Porter R, Kudelko K, Unni A, Segal L, Sharaf R, Blaschak-Harvan J, Eriksson J, Tenkula T, Vidgren G, Ehnholm C, Tuomilehto-Wolf E, Hagopian W, Buchanan TA, Tuomilehto J, Bergman RN, Collins FS, 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
- 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
- Luo TH, Zhao Y, Li G, Yuan WT, Zhao JJ, Chen JL, Huang W, Luo M: A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia* 44:501–506, 2001
- Mori Y, Otobe S, Dina C, Yasuda K, Populaire C, Lecoeur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki P, 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
- 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
- Nakhei H, Lingott A, Lemm I, Ryffel GU: An alternative splice variant of the tissue specific transcription factor *HNF4 $\alpha$*  predominates in undifferentiated murine cell types. *Nucleic Acid Res* 26:497–504, 1998
- 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-4 $\alpha$*  gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum Mol Genet* 10:2089–2097, 2001
- Eckhoute J, Moerman E, Bouckenoghe 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
- 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
- Ryffel GU: Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (*HNF*)1 and *HNF4* families:

- functional and pathological consequences. *J Mol Endocrinol* 27:11–29, 2001
14. Stoffel M, Duncan SA: The maturity-onset diabetes of the young (MODY1) transcription factor HNF4alpha regulates expression of genes required for glucose transport and metabolism. *Proc Natl Acad Sci U S A* 94:13209–13214, 1997
  15. 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
  16. 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
  17. Rhee J, Inoue Y, Yoon JC, Puigserver P, Fan M, Gonzalez FJ, Spiegelman BM: Regulation of hepatic fasting response by PPARgamma coactivator-1alpha (PGC-1): requirement for hepatocyte nuclear factor 4alpha in gluconeogenesis. *Proc Natl Acad Sci U S A* 100:4012–4017, 2003
  18. Ghosh S, Watanabe RM, Valle TT, Hauser ER, Magnuson VL, Langefeld CD, Ally DS, Mohlke KL, Silander K, Kohtamäki K, Chines P, Balow J, Birznieks G, Chang J, Eldridge W, Erdos MR, Karanjawala ZE, Knapp JJ, Kudelko K, Martin C, Morales-Mena A, Musick A, Musick T, Pfahl C, Porter R, Rayman JB, Rha D, Segal L, Shapiro S, Sharaf R, Shurtleff B, So A, Tannenbaum J, Te C, Tovar J, Unni A, Welch C, Whiten R, Witt A, Blaschak-Harvan J, Douglas JA, Duren WL, Epstein MP, Fingerlin TE, Kaleta HS, Lange EM, Li C, McEachin RC, Stringham HM, Trager E, White PP, Eriksson J, Toivanen L, Vidgren G, Nylund SJ, Tuomilehto-Wolf E, Ross EH, Demirchyan E, Hagopian WA, Buchanan TA, Tuomilehto J, Bergman RN, Collins FS, Boehnke M: The Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics (FUSION) study. I: an autosomal genome scan for genes that predispose to type 2 diabetes. *Am J Hum Genet* 67:1174–1185, 2000
  19. Silander K, Scott LJ, Valle TT, Mohlke KL, Stringham HM, Wiles KR, Duren WL, Doheny KF, Pugh EW, Chines P, Narisu N, White PP, Fingerlin TE, Jackson AU, Li C, Ghosh S, Magnuson VL, Colby K, Erdos MR, Hill JE, Hollstein P, Humphreys KM, Kasad RA, Lambert J, Lazaridis KN, Lin G, Morales-Mena A, Patzkowski K, Pfahl C, Porter R, Rha D, Segal L, Suh YD, Tovar J, Unni A, Welch C, Douglas JA, Epstein MP, Hauser ER, Hagopian W, Buchanan TA, Watanabe RM, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* 53:821–829, 2004
  20. Kwok PY: Approaches to allele frequency determination. *Pharmacogenomics* 1:231–235, 2000
  21. Mohlke KL, Erdos MR, Scott L, Fingerlin TE, Jackson AU, Silander K, Hollstein P, Boehnke M, Collins FS: High-throughput screening for evidence of association using mass spectrometry genotyping of single nucleotide polymorphisms. *Proc Natl Acad Sci U S A* 99:16928–16933, 2002
  22. Sham P, Bader JS, Craig I, O'Donovan M, Owen M: DNA pooling: a tool for large-scale association studies. *Nat Rev Genet* 3:862–871, 2002
  23. Love-Gregory LD, Wasson J, Ma J, Jin CH, Glaser B, Suarez BK, Permutt MA: A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4alpha gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an Ashkenazi Jewish population. *Diabetes* 53:1134–1140, 2004
  24. Valle T, Tuomilehto J, Bergman RN, Ghosh S, Hauser ER, Eriksson J, Nylund SJ, Kohtamäki K, Toivanen L, Vidgren G, Tuomilehto-Wolf E, Einholm C, Blaschak J, Langefeld CD, Watanabe RM, Magnuson V, Ally DS, Hagopian WA, Ross E, Buchanan TA, Collins F, Boehnke M: Mapping genes for NIDDM: design of the Finland-United States Investigation of NIDDM Genetics (FUSION) study. *Diabetes Care* 21:949–958, 1998
  25. Watanabe RM, Ghosh S, Langefeld CD, Valle TT, Hauser ER, Magnuson VL, Mohlke KL, Silander K, Ally DS, Chines P, Blaschak-Harvan J, Douglas JA, Duren WL, Epstein MP, Fingerlin TE, Kaleta HS, Lange EM, Li C, McEachin RC, Stringham HM, Trager E, White PP, Balow J, Birznieks G, Chang J, Eldridge W, Erdos MR, Karanjawala ZE, Knapp JJ, Kudelko K, Martin C, Morales-Mena A, Musick A, Musick T, Pfahl C, Porter R, Rayman JB, Rha D, Segal L, Shapiro S, Sharaf R, Shurtleff B, So A, Tannenbaum J, Te C, Tovar J, Unni A, Welch C, Whiten R, Witt A, Kohtamäki K, Einholm C, Eriksson J, Toivanen L, Vidgren G, Nylund SJ, Tuomilehto-Wolf E, Ross EH, Demirchyan E, Hagopian WA, Buchanan TA, Tuomilehto J, Bergman RN, Collins FS, Boehnke M: The Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics (FUSION) study. II: an autosomal genome scan for diabetes-related quantitative-trait loci. *Am J Hum Genet* 67:1186–1200, 2000
  26. Schuler GD: Sequence mapping by electronic PCR. *Genome Res* 7:541–550, 1997
  27. Smit AFA, Green P: RepeatMasker [article online], 1997. Available from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>. Accessed January 2002
  28. Zhang Z, Schwartz S, Wagner L, Miller W: A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203–221, 2000
  29. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Lohmussaer E, Zernant J, Tonisson N, Remm M, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Dunham I: A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 418:544–548, 2002
  30. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA: Haplotype tagging for the identification of common disease genes. *Nat Genet* 29:233–237, 2001
  31. Zeger SL, Liang KY: Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42:121–130, 1986
  32. Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179–1188, 1997
  33. Li C, Scott LJ, Boehnke M: Assessing whether an allele can account in part for a linkage signal: the genotype-IBD sharing test (GIST). *Am J Hum Genet* 74:418–431, 2004
  34. Abecasis GR, Cookson WOC: GOLD: graphical overview of linkage disequilibrium. *Bioinformatics* 16:182–183, 2000
  35. Margulies EH, Blanchette M, Haussler D, Green ED, NISC Comparative Sequencing Program: Identification and characterization of multi-species conserved sequences. *Genome Research* 13:2507–2518, 2003
  36. Ardlie KG, Kruglyak L, Seielstad M: Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299–309, 2002
  37. Watanabe RM, Valle T, Hauser ER, Ghosh S, Eriksson J, Kohtamäki K, Einholm C, Tuomilehto J, Collins FS, Bergman RN, Boehnke M, Finland-United States Investigation of NIDDM Genetics (FUSION) Study Investigators: Familiality of quantitative metabolic traits in Finnish families with non-insulin-dependent diabetes mellitus. *Hum Hered* 49:159–168, 1999
  38. Hani EH, Suaud L, Boutin P, Chevre JC, Durand E, Philippi A, Demeis 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
  39. Price JA, Fossey SC, Sale MM, Brewer CS, Freedman BI, Wuertth JP, Bowden DW: Analysis of the HNF4 alpha gene in Caucasian type II diabetic nephropathic patients. *Diabetologia* 43:364–372, 2000
  40. 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-4alpha 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
  41. Malecki MT, Antonellis A, Casey P, Ji L, Wantman M, Warram JH, Krolewski AS: Exclusion of the hepatocyte nuclear factor 4alpha as a candidate gene for late-onset NIDDM linked with chromosome 20q. *Diabetes* 47:970–972, 1998
  42. Rissanen J, Wang H, Miettinen R, Karkkainen P, Kekalainen P, Mykkanen L, Kuusisto J, Karhapää P, Niskanen L, Uusitupa M, Laakso M: Variants in the hepatocyte nuclear factor-1alpha and -4alpha genes in Finnish and Chinese subjects with late-onset type 2 diabetes. *Diabetes Care* 23:1533–1538, 2000
  43. Mitchell SM, Gloyd AL, Owen KR, Hattersley AT, Frayling TM: The role of the HNF4alpha enhancer in type 2 diabetes. *Mol Genet Metab* 76:148–151, 2002
  44. Mitchell SM, Vaxillaire M, Thomas H, Parrizas M, Benmezroua Y, Costa A, Hansen T, Owen KR, Tuomi T, Pirie F, Ryffel GU, Ferrer J, Froguel P, Hattersley AT, Frayling TM: Rare variants identified in the HNF-4 alpha beta-cell-specific promoter and alternative exon 1 lack biological significance in maturity onset diabetes of the young and young onset type II diabetes. *Diabetologia* 45:1344–1348, 2002
  45. Hegele RA, Cao H, Harris SB, Hanley AJ, Zimman B: Hepatocyte nuclear factor-1alpha G319S: a private mutation in Oji-Cree associated with type 2 diabetes (Letter). *Diabetes Care* 22:524, 1999
  46. McCarthy MI: Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* 3:159–167, 2003
  47. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES: The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80, 2000



48. Ek J, Andersen G, Urhammer SA, Gaede PH, Drivsholm T, Borch-Johnsen K, Hansen T, Pedersen O: Mutation analysis of peroxisome proliferator-activated receptor-gamma coactivator-1 (PGC-1) and relationships of identified amino acid polymorphisms to type II diabetes mellitus. *Diabetologia* 44:2220–2226, 2001
49. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC: PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34:267–273, 2003
50. Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, Miyazaki Y, Kohane I, Costello M, Saccone R, Landaker EJ, Goldfine AB, Mun E, DeFronzo R, Finlayson J, Kahn CR, Mandarino LJ: Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A* 100:8466–8471, 2003