Genetic Variation Near the Hepatocyte Nuclear Factor-4α Gene Predicts Susceptibility to Type 2 Diabetes

Kaisa Silander,1 Karen L. Mohlke,1 Laura J. Scott,2 Erin C. Peck,1 Pablo Hollstein,1 Andrew D. Skol,2 Anne U. Jackson,2 Panagiotis Deloukas,3 Sarah Hunt,3 George Stavrides,3 Peter S. Chines,1 Michael R. Erdos,1 Narisu Narisu,1 Karen N. Conneely,2 Chun Li,2 Tasha E. Fingerlin,2 Sharanjeet K. Dhanjal,4 Timo T. Valle,5,6 Richard N. Bergman,7 Jaakko Tuomilehto,5,6,8 Richard M. Watanabe,4 Michael Boehnke,2 and Francis S. Collins1

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 pairs 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 455 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 β-cell promoter P2 of hepatocyte nuclear factor-4α (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 increase susceptibility to type 2 diabetes. Diabetes 53:1141–1149, 2004

Evidence 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-4α (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 β-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 β-cells and liver. In the β-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) study, the strongest linkage results were on chromosome 20 (4,18), with maximum logarithm of odds (LOD) scores (MLSs) of 1.99, 2.04, and 2.15 at 18, 57, and 70 cM, respectively, on the FUSION
Also, we included an additional 37 FUSION 1 and 21 FUSION 2 affected collected in two phases. The FUSION 1 (F1) set included 495 families with genetic map. Following of each of the participating centers. studied 467 unaffected offspring from 185 FUSION 1 families. siblings (24,25). To evaluate association with diabetes-related traits, we also test (OGTT) at ages 65 and 70 years. In addition, we used a control group of type 1 diabetes and MODY. For control subjects, we studied 225 elderly, normal glucose-tolerant individuals. The numbers of individuals in normoglycemic spouses, and our elderly control (EC) pool included 228 FUSION 1 families; our FUSION 2 case pool included one affected individual from each of 499 FUSION 1 case pools to the combined control pools (EC individuals). 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: rs875559 (P = 0.045), rs6017500 (P = 0.020), and rs6009978 (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 blindeked 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 previously described families. Second, we selected those that captured the variability found in haplotype blocks generated from SNPs genotyped by the Sanger Institute in the FUSION 1 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. Fifth, four SNPs were genotyped to assess evidence for 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’ >0.85 with the current haplotype, requires that the five most common predicted haplotypes account for ≥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 John et al. (30) as implemented in Stata. Thirty-three SNPs accounted for 96% of the predicted haplotype diversity. 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 and analyzed single markers can be found in the supplemental Material. Prominent SNPs 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 to account for familial dependencies among related individuals (31). Analyses were performed assuming additive, dominant, recessive, and general (2 degrees of freedom) genetic models. We report
Allele frequencies and evidence for association in case and control study samples for SNPs in the HNF4A region

<table>
<thead>
<tr>
<th>kb position</th>
<th>SNP name</th>
<th>Major/minor allele</th>
<th>Minor allele frequency (n typed)</th>
<th>P value†</th>
<th>Multiplicative OR (95% CI)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>F1</td>
<td>F2</td>
<td>F1 + F2</td>
</tr>
<tr>
<td>9.422</td>
<td>rs48104248</td>
<td>G/C</td>
<td>0.20 (530)</td>
<td>0.20 (259)</td>
<td>0.20 (789)</td>
</tr>
<tr>
<td>9.430</td>
<td>rs18846143</td>
<td>C/G</td>
<td>0.21 (526)</td>
<td>0.21 (261)</td>
<td>0.21 (787)</td>
</tr>
<tr>
<td>9.326</td>
<td>rs18846148</td>
<td>C/T</td>
<td>0.20 (516)</td>
<td>0.20 (260)</td>
<td>0.20 (776)</td>
</tr>
<tr>
<td>1.272</td>
<td>rs2144908</td>
<td>G/A</td>
<td>0.21 (530)</td>
<td>0.21 (263)</td>
<td>0.21 (793)</td>
</tr>
<tr>
<td>5.269</td>
<td>rs6031551</td>
<td>T/C</td>
<td>0.19 (527)</td>
<td>0.18 (263)</td>
<td>0.18 (790)</td>
</tr>
<tr>
<td>5.349</td>
<td>rs6031552</td>
<td>C/A</td>
<td>0.19 (527)</td>
<td>0.18 (263)</td>
<td>0.18 (790)</td>
</tr>
<tr>
<td>39.604</td>
<td>rs2425637</td>
<td>G/T</td>
<td>0.53 (528)</td>
<td>0.52 (262)</td>
<td>0.53 (787)</td>
</tr>
<tr>
<td>43.592</td>
<td>rs2425640</td>
<td>G/A</td>
<td>0.37 (525)</td>
<td>0.34 (262)</td>
<td>0.36 (780)</td>
</tr>
<tr>
<td>44.840</td>
<td>rs1800963</td>
<td>C/A</td>
<td>0.42 (529)</td>
<td>0.43 (261)</td>
<td>0.43 (787)</td>
</tr>
<tr>
<td>50.690</td>
<td>rs3212183</td>
<td>T/C</td>
<td>0.42 (515)</td>
<td>0.43 (238)</td>
<td>0.43 (753)</td>
</tr>
<tr>
<td>54.595</td>
<td>rs1885088</td>
<td>G/A</td>
<td>0.14 (531)</td>
<td>0.16 (262)</td>
<td>0.15 (783)</td>
</tr>
<tr>
<td>66.316</td>
<td>rs1028583</td>
<td>G/A</td>
<td>0.21 (527)</td>
<td>0.21 (263)</td>
<td>0.21 (790)</td>
</tr>
<tr>
<td>73.035</td>
<td>rs81182475</td>
<td>G/T</td>
<td>0.32 (522)</td>
<td>0.32 (261)</td>
<td>0.32 (783)</td>
</tr>
</tbody>
</table>

†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 values 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 $2 \times 2 \times n^2$ test of independence. Haplotypes with counts $\leq 10$ were pooled. Each individual haplotype was tested for association with disease status using a $2 \times 2 \times n^2$ 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 $S_{\text{IBD}}$ 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 $D^\prime$ was estimated using the ldmax 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 ~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...
cies between type 2 diabetic case (\(n\)) sample (23). Likewise, we did not con-

trast, SNPs near the P1 promoter and coding exons 1

[available at http://diabetes.diabetesjournals.org]. In con-

our sample, expanding the region harboring associated

and in nearly perfect LD with rs1884613 and rs2144908 in

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 con-

trast, SNPs near the P1 promoter and coding exons 1–3

(rs2425637, rs2425640, and rs1885088) were associated 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 con-

trast, 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 in-

cluding 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 associ-

ated alleles was consistent with that expected based on

random recombination between the second and third

SNPs. We also observed a significantly associated haplo-

type (GCTGCA) (OR 1.51, \(P = 0.010\)) containing the risk

alleles for all SNPs except rs2144908, a nominally protec-

tive 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.

<table>
<thead>
<tr>
<th>Alleles†</th>
<th>Frequency*</th>
<th>OR‡</th>
<th>P value§</th>
</tr>
</thead>
<tbody>
<tr>
<td>A CT GC A</td>
<td>0.016</td>
<td>2.11</td>
<td>0.023</td>
</tr>
<tr>
<td>G CT GC A</td>
<td>0.003</td>
<td>1.51</td>
<td>0.010</td>
</tr>
<tr>
<td>A CT TG G</td>
<td>0.063</td>
<td>1.41</td>
<td>0.087</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.031</td>
<td>1.28</td>
<td>0.293</td>
</tr>
<tr>
<td>A CT TG G</td>
<td>0.014</td>
<td>1.27</td>
<td>0.475</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.142</td>
<td>1.20</td>
<td>0.109</td>
</tr>
<tr>
<td>A CT TG G</td>
<td>0.081</td>
<td>1.14</td>
<td>0.369</td>
</tr>
<tr>
<td>A CT TG G</td>
<td>0.027</td>
<td>1.08</td>
<td>0.732</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.047</td>
<td>1.07</td>
<td>0.715</td>
</tr>
<tr>
<td>A CT TG G</td>
<td>0.015</td>
<td>1.01</td>
<td>0.668</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.071</td>
<td>0.91</td>
<td>0.504</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.101</td>
<td>0.89</td>
<td>0.732</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.225</td>
<td>0.78</td>
<td>0.013</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.056</td>
<td>0.78</td>
<td>0.131</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.047</td>
<td>0.73</td>
<td>0.058</td>
</tr>
<tr>
<td>G CT TG G</td>
<td>0.039</td>
<td>0.71</td>
<td>0.051</td>
</tr>
</tbody>
</table>

*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.
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 Δ 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
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 LD with the two FUSION SNPs showing strongest evidence for association, rs2144908 and rs1884613. Including LD with the two FUSION SNPs showing strongest evidence for association, rs2144908 and rs1884613.

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 promoter 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.

### Table 3

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

<table>
<thead>
<tr>
<th>Trait</th>
<th>AA</th>
<th>AG</th>
<th>GG</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>14</td>
<td>141</td>
<td>316</td>
<td></td>
</tr>
<tr>
<td>Fasting glucose (mmol/l)</td>
<td>22.9 (4.0)</td>
<td>25.1 (5.1)</td>
<td>25.6 (5.7)</td>
<td>0.007</td>
</tr>
<tr>
<td>2-h glucose (mmol/l)</td>
<td>4.96 (0.30)</td>
<td>5.02 (0.52)</td>
<td>4.97 (0.67)</td>
<td>0.012</td>
</tr>
<tr>
<td>Δ Glucose (mmol/l)†</td>
<td>5.72 (1.53)</td>
<td>5.25 (1.74)</td>
<td>5.11 (1.86)</td>
<td>0.025</td>
</tr>
<tr>
<td>Fasting insulin (pmol/l)</td>
<td>0.81 (1.26)</td>
<td>0.28 (1.80)</td>
<td>0.05 (1.82)</td>
<td>0.029</td>
</tr>
<tr>
<td>2-h insulin (pmol/l)</td>
<td>60.0 (30.0)</td>
<td>60.0 (42.0)</td>
<td>60.0 (42.0)</td>
<td>0.029</td>
</tr>
<tr>
<td>Δ Insulin (pmol/l)‡</td>
<td>345 (432)</td>
<td>228 (330)</td>
<td>291 (288)</td>
<td>0.019</td>
</tr>
<tr>
<td>Insulin sensitivity (10⁻⁸)</td>
<td>276 (426)</td>
<td>174 (276)</td>
<td>168 (270)</td>
<td>0.019</td>
</tr>
<tr>
<td>min⁻¹ · pmol⁻¹ · l⁻¹ · 8 min§</td>
<td>8.6 (8.18)</td>
<td>6.13 (5.07)</td>
<td>6.87 (6.02)</td>
<td>0.084</td>
</tr>
<tr>
<td>Disposition index (×10⁻³)</td>
<td></td>
<td>11.6 (13.6)</td>
<td>10.7 (11.7)</td>
<td>12.5 (10.9)</td>
</tr>
</tbody>
</table>

*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.
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 β-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 β-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 rs768824, 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-1α (HNF1A) and -1β, 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-γ (47) and peroxisome proliferator–activated receptor coactivator-1α (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 GM00040. R.N.B. was supported by NIH Grants DK27619 and DK20967. 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

12. Ryffel GU: Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families.

Note: The above text includes multiple references to studies and research papers, which are not fully cited here due to space constraints. For the full references, please refer to the original sources.
