The Role of HNF4A Variants in the Risk of Type 2 Diabetes

Karen L. Mohlke, PhD,* and Michael Boehnke, PhD

Introduction

Susceptibility to type 2 diabetes mellitus (T2DM) is influenced by genetic variation. Approaches to discovery of genes affecting T2DM susceptibility include linkage analysis and candidate gene association analysis. More than 20 genome scans for linkage to T2DM have been reported [1•], and many groups have performed genome-wide quantitative trait locus linkage analysis for T2DM-related traits, such as glucose or insulin metabolism, obesity, energy metabolism, and lipids or lipoproteins [2]. The first gene described based on a genome-wide screen and positional cloning is CAPN1 [3], and recent meta-analyses described a 19% and a 17% increased risk, respectively, to carriers of the C allele of intronic “SNP (single nucleotide polymorphism) 44” [4] and carriers of the GG genotype of intronic “SNP 43” [5].

Evidence for Linkage to T2DM on Chromosome 20q

At least 10 groups have reported evidence for chromosome 20q linkage with T2DM (Table 1) [13,14,15••,16–25]. Initially, three groups tested for linkage on chromosome 20q, motivated in part by the location of the MODY gene HNF4A; these groups described evidence of T2DM linkage in Caucasians [13,18,22]. Other reports of linkage to this region based on genome-wide studies have been described in other populations of Caucasians [14,19,20], Asians [16,17,24], and Africans [25]. Some of these studies have been updated to include additional markers or samples [15••,21,23]. In the FUSION (Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics) study, genotyping additional markers in 495 families increased the maximum logarithm of the odds (LOD) score to 2.48, but a second sample of 242 independent families showed no evidence for linkage on chromosome 20q, and the combined samples showed a maximum LOD score on chromosome 20q of only 0.51 [21]. Chromosome 20q has also been implicated in linkage studies of other diabetes-related traits, including obesity [26,27], insulin [28,29], energy metabolism [30], and lipids [31,32], and there are mouse quantitative trait loci for obesity in the syntenic region of chromosome 2 [28,33].

Identification of Common Variants Near HNF4A Associated with T2DM

At least three groups evaluated evidence that common variants on chromosome 20q may have an impact on risk to T2DM [34,35,36••]. Two groups independently identified equivalent
Table 1. Published genome scans with evidence for linkage to chromosome 20q

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Families/sample</th>
<th>Phenotype</th>
<th>Score</th>
<th>cM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zouali et al. [13]</td>
<td>French</td>
<td>148 families, 301 ASP</td>
<td>T2DM</td>
<td>1.31</td>
<td>MLS</td>
</tr>
<tr>
<td>Permutt et al. [14]</td>
<td>Ashkenazi Jewish</td>
<td>267 families, 472 ASP</td>
<td>T2DM</td>
<td>2.05</td>
<td>NPL</td>
</tr>
<tr>
<td>Permutt et al. [14], updated by Love-Gregory et al. [15]</td>
<td>Ashkenazi Jewish</td>
<td>199 families, 299 ASP</td>
<td>T2DM age of diagnosis &gt; 35 y</td>
<td>2.01</td>
<td>MLS</td>
</tr>
<tr>
<td>Iwasaki et al. [16]</td>
<td>Japanese</td>
<td>164 families, 256 ASP</td>
<td>T2DM</td>
<td>1.99</td>
<td>MLS</td>
</tr>
<tr>
<td>Mori et al. [17]</td>
<td>Japanese</td>
<td>159 families, 224 ASP, 359 affected individuals</td>
<td>Lean T2DM</td>
<td>2.32</td>
<td>MLS</td>
</tr>
<tr>
<td>Mori et al. [17]</td>
<td>Japanese</td>
<td>159 families, 224 ASP, 359 affected individuals</td>
<td>T2DM</td>
<td>1.67</td>
<td>MLS</td>
</tr>
<tr>
<td>Bowden et al. [18]</td>
<td>Caucasian</td>
<td>21 families, 53 ASP</td>
<td>T2DM + diabetic nephropathy</td>
<td>1.48</td>
<td>MLS</td>
</tr>
<tr>
<td>Vionnet et al. [19]</td>
<td>French</td>
<td>143 nuclear families, 677 ASP</td>
<td>Large T2DM families</td>
<td>1.72</td>
<td>MLS</td>
</tr>
<tr>
<td>Ghosh et al. [20], updated by Slender et al. [21]</td>
<td>Finnish</td>
<td>495 families, 1129 affected individuals</td>
<td>T2DM</td>
<td>2.48</td>
<td>MLS</td>
</tr>
<tr>
<td>Ji et al. [22], updated by Klupa et al. [23]</td>
<td>Caucasian</td>
<td>43 families, 241 affected individuals</td>
<td>Middle-age-onset T2DM</td>
<td>5.32</td>
<td>NPL</td>
</tr>
<tr>
<td>Luo et al. [24]</td>
<td>Han Chinese</td>
<td>102 families, 282 affected individuals</td>
<td>T2DM</td>
<td>1.52</td>
<td>NPL</td>
</tr>
<tr>
<td>Rotimi et al. [25]</td>
<td>West African</td>
<td>343 ASPs, 691 affected individuals</td>
<td>T2DM</td>
<td>1.80</td>
<td>MLS</td>
</tr>
<tr>
<td>Zouali et al. [13]</td>
<td>French</td>
<td>42 families, 55 ASP</td>
<td>Early-onset T2DM</td>
<td>2.34</td>
<td>MLS</td>
</tr>
<tr>
<td>Rotimi et al. [25]</td>
<td>West African</td>
<td>343 ASPs, 691 affected individuals</td>
<td>T2DM</td>
<td>2.63</td>
<td>MLS</td>
</tr>
</tbody>
</table>

ASP—affected sibling pair; cM—centimorgan map position of the maximum logarithm of odds score on the Marshfield genetic map; MLS—maximum logarithm of the odds score; NPL—nonparametric linkage score; T2DM—type 2 diabetes mellitus.

DNA variants near HNF4A that were associated with T2DM in study participants from Finland and Israel [15,36] using somewhat different approaches. In addition, HNF4A was tested as a T2DM candidate gene independent of evidence for linkage, and a haplotype of common variants was found to be associated with T2DM [37]. These three studies will be described in greater detail below.

The HNF4A protein is a widely acting transcription factor in the steroid hormone receptor family. The protein plays an important role in development, metabolism, and differentiation. HNF4A is expressed in the liver, pancreatic islets, kidney, and intestine, has two known promoters designated P1 and P2, and has at least nine splice variants [38]. The P2 promoter is active in pancreatic β cells and hepatocytes and is located 46 kb upstream of the P1 promoter, which is active in hepatocytes [39,40]. The HNF4A protein functions as part of a complex transcriptional network with many downstream targets. Recently, HNF4A was found to be bound to promoters of 42% of genes actively transcribed in liver cells and 43% of genes actively transcribed in pancreatic islet cells [41].

Fine-mapping in the chromosome 20q linkage region
Genetic fine-mapping of a linkage region followed by study of positional candidate genes provides an alternative to the classical candidate gene approach. Fine-mapping acknowledges that our a priori knowledge of the genes involved in diabetes is limited and so more thoroughly investigates the region at the expense of additional genotyping. To thoroughly assess evidence for association across a region, fine-mapping requires densely spaced markers, with more markers in regions of greater recombination. The HapMap project (http://www.hapmap.org) is providing excellent resources and will soon allow a large fraction of common genetic variability to be assayed [42].

The fine-mapping approach used by FUSION investigators (http://fusion.sph.umich.edu) began with a screen for SNP-T2DM association using DNA pools [36]. This approach enabled more SNPs to be screened for a fixed cost, although with the disadvantage of increased variability in allele frequency estimation due to pool construction and measurement error [43]. Using DNA pools of 182 to 499 case or control samples, the investigators screened an initial 291 SNPs for evidence of association with T2DM. This preliminary density of approximately 25 kb per SNP has not captured all common variability between individuals in this region. For the 21 SNPs estimated to have significant allele frequency differences between case pools and control pools, all individuals comprising the pools were genotyped. The most strongly associated SNP based on individual genotypes was rs2144908, located in intron 1D, 1272 bp downstream of the ATG translation initiation site corresponding to the P2 promoter. Once this first associated SNP was identified, 61 other SNPs in the gene region were tested for association. Evidence for association was observed with SNPs spanning a 59-kb region, including both the P2 and P1 promoters and coding exons 1 to 3. Figure 1 shows the location of SNPs with evidence for association. More recently, stronger evidence for T2DM-SNP association was identified with...
rs6031558 ($P = 0.002$, OR = 1.36). This SNP has an allele frequency of 0.749 in FUSION cases and 0.686 in FUSION controls. The SNP is not in strong linkage disequilibrium (LD) with the previously identified SNPs near P2 (rs6031558 with rs2144908, $|D'| < 0.10$).

Candidate gene in the chromosome 20q linkage region

The investigators studying Ashkenazi Jewish individuals from Israel also observed evidence for linkage on chromosome 20q (Table 1) [14]. In their linkage region, an initial screen for T2DM association using a similar approach with DNA pools did not identify reproducible associations [35], so they pursued candidate genes using more closely spaced SNPs [15••]. These investigators tested SNPs spanning 78 kb around $HNF4A$ in 275 Ashkenazi Jewish individuals with T2DM and 342 control individuals. They identified evidence for association initially with SNP rs1884614 located near the P2 promoter and rs3818247 located in intron 9 near the 3’ end of the gene. Once these initial associated SNPs were identified, additional SNPs were tested for association. As in the FUSION study, evidence for SNP-T2DM association was observed with SNPs spanning both the P2 and P1 promoter regions and coding exons, a region of 82 kb (Fig. 1) [15••].

These two groups shared preliminary results and discovered that the T2DM-associated SNPs each group had identified near the P2 promoter were in perfect LD ($r^2 = 1$) with one another. That is, the same alleles at SNPs rs2144908 and rs1884614, as well as SNPs rs4810424 and rs1884613, were always found together on the chromosomes in both the Finnish and Ashkenazi Jewish samples. In contrast, the T2DM-associated SNPs that each group had identified in intron 1D, the P1 promoter region, and among the distal coding exons were not associated in the other study [15••, 36••]. These results may be explained, in part, by the significant evidence for historic recombination between the P2 and P1 promoters. It remains to be determined whether there is one common underlying variant responsible for evidence of association in both populations, or whether there is more than one underlying variant in one or both populations.

Candidate gene independent of evidence for linkage

Common SNPs in $HNF4A$ were also detected in a candidate gene survey that did not use prior evidence for linkage as a criterion for gene selection [37••]. These investigators tested 152 SNPs in 71 genes for evidence of association with T2DM in 517 Caucasian individuals from the United Kingdom with T2DM and 517 control subjects, and they tested the same SNPs in a cohort of 1100 Caucasians for evidence of association with related quantitative traits. In this study, three SNPs in $HNF4A$ were tested individually and as haplotypes. The SNPs are located in exon 1C and intron 1B, downstream of the P1 promoter. Although neither SNP showed individual evidence for association, a common haplotype with population frequency 0.33 of G at SNP rs2071197 (Val49Met) and C at SNP rs736824 showed decreased risk of T2DM (OR 0.83; 95% confidence interval, 0.68 to 1.00). In addition, this same haplotype was significantly associated with increased insulin secretion compared with the other two haplotypes observed. The SNPs on this haplotype are located near T2DM-associated SNPs identified separately by the studies discussed earlier (Fig. 1). In the Finnish sample, this haplotype shows the same trend but is not significantly associated.
with T2DM (case frequency = 0.306, control frequency = 0.331, OR = 0.89, \( P = 0.32 \)). These variants may be in LD with causative variants shared across studies.

### Confirmation of association between T2DM and HNF4A SNPs near the P2 promoter

Recently, other groups have tested the described SNPs near the P2 promoter for evidence of association (Table 2). As expected for a complex trait, whereas some of these replication studies confirm significance others may not. Consistent with the modest significance level in the original two studies and the winner’s curse [46,47], replication studies have reported lower ORs and larger \( P \) values. The original studies may have overestimated the ORs associated with the risk allele due to bias or population diversity [46,48], the use of familial cases of T2DM, or because the original studies have some evidence for linkage at HNF4A, which strengthened the power to detect the association [49]. A meta-analysis is underway to determine whether these variants are associated with T2DM in combined population samples. Although these DNA variants show evidence for association with T2DM, functional studies will be needed to identify the putative causal variant(s) and it remains possible that a nearby gene is affected rather than (or in addition to) HNF4A.

### SNP association and evidence for linkage

As expected for genetic variants with modest effect, tests for association can detect variants in regions without significant evidence for linkage. Within the FUSION study, the two groups of families, designated FUSION 1 and FUSION 2, showed very different evidence for linkage, with maximum LOD scores of 2.48 and 0.00, respectively, near HNF4A [21]. Yet, the same allele frequency of 0.21 was observed in the genotyped cases from the 532 FUSION 1 and 263 FUSION 2 families, with one case genotyped from each family. The modest ORs are also consistent with the observation of significant evidence for association in at least two studies [44,45] where no evidence for linkage was observed on chromosome 20q [50–52].

### HNF4A in MODY Versus T2DM

At least 20 possible mutations in HNF4A have been described to cause MODY [53]. These variants include nonsense, frameshift mutations, as well as an in-frame insertion and a putative splicing variant. In addition, MODY mutations that indicate the importance of the P2 promoter include a translocation disrupting the spacing between the P2 and P1 promoters, and mutations in the IPF1 and HNF1A transcription factor binding sites in the P2 promoter. None of the mutations have been determined to have a dominant negative effect, suggesting that the likely molecular mechanism for MODY is haploinsufficiency.

In addition to severe hyperglycemia, patients with MODY carrying HNF4A mutations exhibit impaired insulin secretion, suggesting that the primary defect is in pancreatic β cells [12]. Serum triglyceride concentration and apolipoproteins AI and CII have been reported as reduced in some individuals with MODY due to HNF4A mutations [54]. In comparison, HNF4A SNPs associated with T2DM also show evidence for association with similar quantitative traits. In FUSION, SNP rs2144908 shows evidence for association with acute insulin response to glucose, a measure of β-cell function, in unaffected at-risk offspring [36]. The haplotype identified by Barroso et al. [37] is also associated with insulin secretion. In a study of the Amish, inheritance of the A "risk" allele at rs1884614 is associated with increased glucose area under the curve during an oral glucose tolerance test, consistent with decreased insulin secretion [45].

### Table 2. Published reports testing evidence of association between T2DM and HNF4A SNPs near the P2 promoter

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Samples†</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silander et al. [36]</td>
<td>Finnish</td>
<td>795 cases, 414 controls</td>
<td>( P = 0.011, \text{OR} = 1.33 )</td>
</tr>
<tr>
<td>Love-Gregory et al. [15]</td>
<td>Ashkenazi Jewish</td>
<td>275 cases, 342 controls</td>
<td>( P = 0.008, \text{OR} = 1.45 )</td>
</tr>
<tr>
<td>Weedon et al. [44]</td>
<td>U.K. Caucasians</td>
<td>2004 cases, 1635 controls, 509 trio families</td>
<td>( P = 0.02, \text{OR} = 1.15 )</td>
</tr>
<tr>
<td>Damcott et al. [45]</td>
<td>Amish</td>
<td>137 T2DM, 139 IGT, 342 NGT individuals</td>
<td>T2DM vs NGT: Trend toward association, ( P = 0.09, \text{OR} = 1.40 ) T2DM + IGT vs NGT: Trend toward association, ( P = 0.07, \text{OR} = 1.35 )</td>
</tr>
</tbody>
</table>

†When unspecified, cases are individuals with T2DM.
*rs1884613, rs1884614, rs2144908, or rs4810424.

IGT—impaired glucose tolerance; NGT—normal glucose tolerance; OR—odds ratio; SNPs—single nucleotide polymorphisms; T2DM—type 2 diabetes mellitus.
mutations and the common variants described earlier is not yet known.

Taken together, these data suggest possible mechanisms for how HNF4A variants may increase susceptibility to T2DM. The T2DM-associated variants identified to date are located mostly in noncoding nonpromoter regions and the many previous screens of promoters, exons, and intron/exon boundaries suggest that T2DM susceptibility variants would regulate gene expression through enhancers or chromosomal biology. Loss-of-function mutations leading to MODY should decrease gene function by 50%, suggesting that T2DM susceptibility variants decrease gene function by less than 50%. Given the very large number of promoters to which HNF4A protein binds [41••], the T2DM susceptibility variants may have a very small effect on HNF4A gene expression, an effect that is amplified by downstream genes. Alternate mechanisms include altered HNF4A expression timing during pancreas or liver development, an altered ratio of splice variants, or feedback from an isoform of HNF4A protein that had been transcribed by the P1 promoter [66].

Next Steps Toward Identification of Causative Alleles

The common T2DM-associated SNPs identified to date were more or less randomly chosen and are not likely the putative functional variants. To determine the location of all possible susceptibility variants requires additional testing of previously discovered variants, and may require intense resequencing if several rare risk alleles exist (allelic heterogeneity). Functional studies of potential variants will be necessary to determine the mechanism by which the variants act. These studies are challenging because there could be more than one predisposing SNP, the predisposing SNP(s) could be rare, and there is extensive LD, suggesting that the predisposing alleles could be located tens of kilobases away from the original SNPs. In addition, the functional effect may be too small to measure and may be limited by tissue or timing of expression.

Are There Other Chromosome 20q Genes for T2DM?

The evidence for SNP-T2DM association at HNF4A may represent one of several chromosome diabetes susceptibility genes on chromosome 20q. For example, there is evidence for association between obesity [67], T2DM [68], insulin resistance [69], and SNPs in the PTPN1 gene (protein tyrosine phosphatase, nonreceptor type 1), located approximately 6.2 Mb from HNF4A. The presence of multiple underlying genes could contribute to evidence for T2DM linkage observed in multiple studies (Table 1).

The SNPs near the P2 promoter appeared to partition the evidence for linkage in the FUSION 1 families and the Ashkenazi Jewish sample [15••,36••]. That is, families in which the genotyped individual carried the SNP risk allele exhibit substantially greater evidence for linkage on chromosome 20q than sibling pairs, where a genotyped individual did not carry the SNP risk allele. However, these results do not mean that the SNP fully explained the evidence for linkage. Full explanation of a linkage signal may not be reached if sampling variability by chance-generated excess allele sharing, so that even predisposing SNPs wouldn’t necessarily account fully for the observed linkage [3].

Conclusions

Several years after the identification of HNF4A as a gene causing MODY, there is now accumulating evidence that HNF4A may play a role in cases of T2DM as well. Initially, HNF4A was weakened as a T2DM candidate because no clearly predisposing variants were found in coding regions or exon/intron boundaries or proximal promoter regions [56,58,59,61,62]. As genotyping technology has improved, the cost of screening genes and chromosomal regions more thoroughly has decreased, which improves the chances of identifying noncoding susceptibility alleles. These factors, combined with the larger sample sizes necessary for good power, may lead to further replications of genetic association with T2DM. Perhaps other candidate genes previously excluded as T2DM genes on the basis of screening exons, exon/intron boundaries, and promoters will be found to harbor noncoding susceptibility variants.

Future work will determine whether the HNF4A variants play a significant role in additional populations, which allele(s) is functional, and how the variants impact susceptibility. Additional studies may also identify relevant upstream and downstream genes, as well as possible interacting genes and environmental exposures.

Acknowledgments

We thank our colleagues in the FUSION study group for their many contributions to this work. We thank Alan Permutt, Latisha Love-Gregory, and colleagues for generously sharing results from their study prior to publication. We gratefully acknowledge support from grant R01 DK62370 (MB). KLM is supported, in part, by a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund.

Dr. Michael Boehnke can be reached at the Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, USA; E-mail: boehnke@umich.edu.
References and Recommended Reading

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

A recent review of 20 T2DM genome scans, highlighting regions of interest on chromosomes 1q, 2q, 3q, 8p, 12q, and 20q, with a particular emphasis on 1q.


A recent paper describing genes in which variants are known to cause MODY, and describing how these genes form an integrated transcriptional network important in many metabolic pathways. Also described are other early-onset maternally inherited and recessive forms of T2DM.


Description of the identification of variants in and around HNF4A that are associated with T2DM in an Ashkenazi Jewish sample from Israel using a positional-candidate and haplotype-map–based strategy. These variants explain a substantial portion of the 20q linkage signal in these families.


Follow-up study of Silander et al. [36••] and Love-Gregory et al. [15••] that confirms the association between HNF4A promoter variants and T2DM in a study of 5256 U.K. Caucasian subjects, but suggests a weaker HNF4A effect on T2DM risk.


