Young-Onset Type 2 Diabetes Families Are the Major Contributors to Genetic Loci in the Diabetes UK Warren 2 Genome Scan and Identify Putative Novel Loci on Chromosomes 8q21, 21q22, and 22q11

Timothy M. Frayling,1 Steven Wiltshire,2 Graham A. Hitman,3 Mark Walker,4 Jonathan C. Levy,5 Mike Sampson,6 Christopher J. Groves,2 Stephan Menzel,2 Mark I. McCarthy,2,5 Timothy M. Frayling,1 Steven Wiltshire,2 Graham A. Hitman,3 Mark Walker,4 Jonathan C. Levy,5 Mike Sampson,6 Christopher J. Groves,2 Stephan Menzel,2 Mark I. McCarthy,2,5 and Andrew T. Hattersley1

A young onset of type 2 diabetes is likely to result, in part, from greater genetic susceptibility. Young-onset families may therefore represent a group in which genes are more easily detectable by linkage. To test this hypothesis, we conducted age at diagnosis (AAD) stratified linkage analyses in the Diabetes UK Warren 2 sibpairs. In the previously published unstratified analysis, evidence for linkage (logarithm of odds [LOD] >1.18) was found at seven loci. The LOD scores at these seven loci were higher in the 245 families with AAD <55 years (L55) compared with the 328 families with AAD >55 years (G55). Five of these seven loci (1q24–25, 5q13, 8p21–22, 8q24.2, and 10q23.2) had LOD scores >1.18 in the L55 subset but only one (8p21–22) did in the G55 subset. Two additional loci (8q21.13 and 21q22.2) showed evidence for linkage in the L55 subset alone. Another locus (22q11) showed evidence for linkage in a subset of families with AAD <45 years. Using a locus-counting approach, the L55 subset had significantly more loci (\(P \sim 0.01\)) than expected under the null hypothesis of no linkage across the LOD score range 0.59–3.0. In contrast, the G55 subset contained no more susceptibility loci than that expected by chance. In conclusion, young-onset families provide both disproportionate evidence for linkage to known loci and evidence for additional novel loci. Our data confirm our hypothesis that families segregating young-onset type 2 diabetes represent a more powerful resource for defining susceptibility genes by linkage. Diabetes 52:1857–1863, 2003

For many complex diseases, genetic predisposition is stronger in subjects diagnosed at younger ages. There is evidence that this is true in type 2 diabetes. Younger type 2 diabetic subjects show increased familial clustering of diabetes even when monogenic forms are excluded. Several studies have shown high prevalences of parental diabetes in subjects diagnosed at young ages (1–3). Weijnen et al. (4) demonstrated that younger age at diagnosis (AAD) results in an increased relative risk to siblings. This is likely to reflect increased genetic predisposition as well as shared environmental factors.

If type 2 diabetic patients diagnosed at a young age have a stronger genetic predisposition, two questions need to be answered: 1) do these patients have a different genetic predisposition? and 2) are patients diagnosed at young ages more heavily loaded for the predisposing genes that contribute to disease across all ages? The answers to these questions will be crucial in refining strategies for type 2 diabetes gene discovery.

If subjects diagnosed at younger ages have a stronger genetic predisposition, it is then likely that they provide a more powerful resource for mapping susceptibility genes. The hypothesis that families segregating young-onset diabetes are a more powerful resource for mapping genes compared with families with patients diagnosed from across all age ranges has not been tested using genome-wide genotype data. In some studies, such as FUSION (5), subjects diagnosed young contribute disproportionately to the evidence for linkage at some loci. However, these previous studies did not present the evidence for linkage from young-onset families from across all the genome.

To examine the hypothesis that young-onset families are a more powerful resource for mapping genes, and to define whether the localization of predisposing genes in these families was similar or different to that found in all subjects, we used the Diabetes UK Warren 2 sibling pair collection. A recent genome-wide scan using 573 of these families revealed seven putative loci on chromosomes 1, 5, 7, 8, and 10 (6). None of these loci reached genome-wide levels of significance when using the Lander and Kruglyak guidelines (7), which, though widely used, assume that...
complete genetic information has been extracted from the families. However, as the majority of type 2 diabetes sibling pairs do not have parents available, and because our genome scan had a mean marker spacing of 9 cM, such assumptions are not appropriate. Instead, genome-wide significance levels were determined empirically from the data. In addition, a recent study applied a complementary approach to assess results from the Diabetes UK Warren 2 genome scan (8). Simulations of the genome scan (using the same family structures, maps, and allelic frequencies) were performed under the null hypothesis of no linkage and were then compared with the observed number of loci showing evidence for linkage with the number expected under the null hypothesis of no linkage. The results showed there were more loci with evidence for linkage than expected by chance, even though it was not possible to determine which loci were real and which signals had appeared through stochastic variation.

In this study, we have examined subsets of young AAD sibship families from the Warren 2 genome scan. We have tested whether younger families contribute disproportionately to the evidence for linkage at loci identified in our overall scan and whether they provide evidence for additional loci.

**RESEARCH DESIGN AND METHODS**

**Subjects.** All subjects in this study are from the 573 families analyzed in the previously reported genome-wide scan of type 2 diabetes Diabetes UK Warren 2 sibpairs. Details of the ascertainment and the inclusion and exclusion criteria of these families are given in our previous publication (6). Briefly, index sibships were diagnosed between 35 and 70 years. Extensive efforts were made to exclude other forms of diabetes by standard clinical criteria, including personal and family history and an absence of anti-GAD autoantibodies. All sibships were of British/Irish descent. Sibships with evidence of bilineal inheritance and large sibships with a high proportion of affected subjects were excluded.

**Subsets.** We reanalyzed our genome-wide scan on three subsets of families according to defined AAD cutoffs. These subsets were selected to have an average AAD in affected family members of <45.0 (L45) and ≤55.0 (L55) years. These represented the youngest 39 (7%) and 245 (43%) families, respectively. The subsets were not exclusive, so all members of the L45 subset were also in the L55 subset. In addition to the sibships, DNA was available from two affected and one unaffected parent in the L45 subset and from six affected and six unaffected parents in the L55 subsets. In addition, to assess the contribution of young AAD families to the evidence for linkage, we analyzed the complementary 328 families with an average AAD ≥55 years (G55). Clinical characteristics of affected subjects from these subsets are given in table one along with details of the total dataset for comparison. In addition, because of its potential relevance to increased genetic susceptibility, we calculated the proportion of subjects with a recalled maternal or paternal history of diabetes. No subjects had a bilineal paternal history of diabetes, as this was an exclusion criterion. To minimize bias, we only included data from probands who recalled the diabetes status of both parents. There was no difference in the proportion of subjects recalling the diabetes status of both parents across subsets (∼90% in all groups). Fisher’s exact test and χ² test for trend were used to compare frequencies of parental diabetes between groups.

**Linkage analysis.** Linkage analyses were performed using the genotype data generated from the 573 families as previously reported (6). Briefly, this included 418 autosomal microsatellite markers genotyped on ABI-377 platforms. These markers had an average spacing of 9 Haldane centiMorgans [cM (H)]. To eliminate bias toward loci identified in the overall scan, we did not include the extra markers used to fine map the region identified on chromosome 1q. Analysis was performed exactly as described in the analysis of the 573 sibships: a Haldane map function was used and multipoint model-free linkage statistics (the allele-sharing logarithm of odds [LOD] score of Kong and Cox [9]) was computed using ALLEGRO version 1.1b [10]). All LOD scores are multipoint.

**Significance of individual young AAD loci.** To compare the evidence for linkage obtained in each subset with that obtained from the total analysis, we used permutation methods to derive empirical significance values for any differences in LOD score observed. For each replicate, we selected either 39 or 245 families (equivalent to the numbers of pedigrees in the L45 and L55 subsets, respectively) at random from the 573 family dataset and determined the LOD score. With 10,000 replicates, we were able to establish the frequency with which the selection led to an increase in LOD score threshold of 2 or greater than that detected in the observed data. The number of times the actual LOD was reached or exceeded (within 20 cM of either side of the observed peak) was then used to calculate an empirical P value for the effects of subsetting. Any increase in evidence for linkage in the young families was not due to the extraction of extra information from these families. To ensure family size, as assessed by numbers of affected subjects, and information extraction were identical across the L45, L55, and G55 subsets (data not shown).

**Ordered subset analysis.** As a complementary analysis to those based on predetermined age cutoffs (L45 and L55), we performed ordered subset analyses at each locus. This method can be used for any quantitative trait. It attempts to reduce the likely genetic heterogeneity underlying type 2 diabetes by identifying more homogeneous groups of families on the basis of diabetes-associated intermediate traits. This method was used in analyses by the FUSION study (5). We used mean AAD to rank all 573 families from youngest to oldest and, in a separate analysis, oldest to youngest. The LOD score for linkage to type 2 diabetes was then repeatedly recalculated as families were added in rank order. The overall maximum LOD score for any subset of families is reported. The significance of this maximized LOD score derived from these analyses was assessed by 10,000 permutations adding in families in random order.

**Estimation of expected numbers of independent regions of linkage.** To gain a better understanding of the strength of evidence for linkage obtained from each subset, we compared the results from the subset genome scans with results expected under the null hypothesis of no linkage using previously described methods (8). Briefly, using SIMULATE (11, 10,000 replicates of the genome) and the null hypothesis of no linkage, we calcualted the number of independent loci, assuming no latent genotyping error, using the same map and allele frequencies applied for the 573 families and the same pattern of missing genotype information observed in the subsets of families. Each replicate was then analyzed using ALLEGRO and IRLs, defined as maximum LOD scores >55 cM(H) apart (∼40 cM[K]), counted. To assess the significance of the observed number of loci obtained from across the range of LOD score thresholds compared with the number of loci expected under the null hypothesis of no linkage, we also determined from our simulations the number of independent regions of linkage (IRLs) with a cumulative probability of 0.05 and 0.99. These probabilities are equivalent to a one-tailed P value of 0.05 and 0.01, respectively.

The locus-counting plots in Fig. 1 graphically represent the genome-wide number of independent regions showing evidence for linkage (i.e., independent loci). These plots show the number of IRLs observed in an actual genome-wide scan experiment, in relation to the number expected by chance under the null hypothesis of no linkage at all, across the whole range of feasible LOD score thresholds. Each point on the null IRL distribution represents the mean count of IRLs per genome scan, estimated from 10,000 replicates of the genome under the null, that meet or exceed the given LOD score threshold. The actual IRL distribution trace shows the analogous counts observed in the actual genome scan experiment itself: as it is based on one single experiment, it has a step-like appearance. Any excess of loci observed in the actual genome scan, over that expected by chance under the null, can be clearly visualized from these traces. Each point on the P = 0.05 limit trace represents the number of IRLs at the given LOD threshold, that must be met or exceeded in the actual genome scan for the excess to be significant at the P = 0.05 significance limit. As only a whole number of IRLs can be observed at any LOD threshold in an actual genome-scan experiment, the P = 0.05 limit trace shows individual points instead of a continuous line.

**Power.** To calculate the power of the young subsets, we performed simulations (1,000 replicates of an average Warren 2 chromosome; that is, a chromosome with the mean number of markers, mean marker spacing [9 cM(H)], mean number and distribution of alleles, and mean missing genotype proportions) at various values of λa (relative risk to sibling) as previously described (6). As expected, the young subsets had reduced power to detect significant linkage with the total 573 pedigrees. Moreover, the 245 L55 families still had very good power (>95%) to detect loci at λa values of 2.5 and good power (84%) to detect loci at λa values of 2.0 (at LOD scores of >3.0). The 39 L45 families only had good power to detect loci of larger effect at smaller thresholds (90% at λa values of 7.5 with LOD >1). Despite this low power, we have described the results from these 39 families because we believe it would be important to prioritize to the evidence for linkage at small effects in parental diabetes compared with other subsets, which is consistent with them having a stronger genetic predisposition.
RESULTS

Characteristics of young AAD subsets. Table 1 shows characteristics of affected subjects from the L45, L55, and G55 subsets in comparison to the overall 573 families. Families with younger AADs had increased parental history of diabetes ($P_{/H11005}$ 0.0009 for trend across L45, L55, and G55 groups).

Linkage results. Table 2 gives details of the LOD scores in the L45, L55, and G55 subsets for the seven loci detected at LOD $/H11022$ 1.18 in the overall genome scan of 573 families. This LOD score threshold is used for convenience and consistency with the previous study (6). LOD scores are the maximum observed within 20 cM of either side of the peak identified in the genome scan of 573 families; hence, the LOD scores for the total dataset are not the exact sum of those from the L55 and G55 groups.

For all seven loci highlighted in the genome scan of 573 families, the LOD scores in the L55 group were higher than the LOD scores from the G55 families. Five of the seven loci (1q24–25, 5q13, 8p21–22, 8q24.2, and 10q23.2) occurred at LOD scores >1.18 in the L55 families, but only one (8p21–22) occurred over this nominal threshold in the G55 families. Four of the seven loci exceeded the empirically determined LOD score of 1.56 for declaring suggestive linkage—by chance, the LOD score threshold at which only one locus is expected to occur in a genome-wide scan of the Warren 2 data. This threshold is lower than that indicated by Lander and Kruglyak’s guidelines (LOD 2.2) because of the incomplete inheritance information that can be extracted from the Warren 2 data (8). The increase in LOD score in the L55 subset compared with the G55 subset was attenuated slightly for the 1q locus after the addition of extra markers (LOD 2.0 in L55, 0.78 in G55, and 1.98 in total families [6]). Empirically determined LOD score thresholds for suggestive linkage in the L45, L55, and G55 subsets are 1.60, 1.56, and 1.56, respectively.

Table 3 gives details of three additional loci with some evidence for linkage detected in the young AAD families. These occurred on chromosomes 8q21.13 and 21q22 in the L55 subset and chromosome 22q11 in the L45 subset.

Significance of individual loci: permutation to assess significance of increases in LOD score. Table 4 shows that families segregating young-onset diabetes contribute disproportionately to the evidence for linkage at some loci. Younger AAD families give significantly increased evidence for linkage to the loci on chromosomes 8q24.2 ($P_{/H11005}$ 0.01 for increase in LOD score), 10q23.2 ($P_{/H11005}$ 0.02), 21q22 ($P_{/H11005}$ 0.03), 5q32 ($P_{/H11005}$ 0.006), and 22q11 ($P_{/H11005}$ 0.003) than would be expected if the same number of families had been selected at random from the 573 dataset (Table 4).

Significance of individual loci: ordered subset analyses. Ordered subset analyses were consistent with a subset of the youngest AAD families contributing disproportionately to the evidence for linkage at loci on chromosomes 5q32 (empirical $P = 0.001$) and 22q11 (empirical $P = 0.04$).
Our results show that in the Diabetes UK Warren 2 genome-wide scan for type 2 diabetes, the evidence for linkage comes disproportionately from families diagnosed before 55 years of age. This indicates that the power of linkage analyses is likely to be higher in such families. It is consistent with the hypothesis that genetic susceptibility is enhanced in young-onset families.

For all seven loci with evidence for linkage in the overall genome-wide scan of 573 families, the LOD score was higher in the L55 subset than the G55 subset. Of these seven loci, five were detectable (at LODs > 1.18, nominal P < 0.01) in the L55 subset, namely 1q24–25, 5q13, 8p21–22, 8q24.2, and 10q23.2. In contrast, only one locus highlighted in the overall scan (8p21–22) was detected in the G55 subset. In addition, our results provide evidence for novel loci on chromosomes 8q21, 21q22, and 22q11.

For the loci identified in the L55 subset, on chromosomes 8q24.2, 10q23.2, and 21q22, young families provide significantly more evidence for linkage than would be expected if the contribution to these loci were proportionate across all 573 families.

In the L45 subset, two loci (5q32 and 22q11) were identified above the nominal cutoff of LOD > 1.18. One of these loci (5q32) occurs in the overall genome-wide scan of 573 families. For both these loci, both permutation analyses and ordered subset analyses supported a role for a small number of the youngest AAD families contributing disproportionately to the evidence for linkage.

In our study, families segregating young-onset type 2

---

**DISCUSSION**

Our results show that in the Diabetes UK Warren 2 genome-wide scan for type 2 diabetes, the evidence for linkage comes disproportionately from families diagnosed before 55 years of age. This indicates that the power of linkage analyses is likely to be higher in such families. It is consistent with the hypothesis that genetic susceptibility is enhanced in young-onset families.

For all seven loci with evidence for linkage in the overall genome-wide scan of 573 families, the LOD score was higher in the L55 subset than the G55 subset. Of these seven loci, five were detectable (at LODs > 1.18, nominal P < 0.01) in the L55 subset, namely 1q24–25, 5q13, 8p21–22, 8q24.2, and 10q23.2. In contrast, only one locus highlighted in the overall scan (8p21–22) was detected in the G55 subset. In addition, our results provide evidence for novel loci on chromosomes 8q21, 21q22, and 22q11.

For the loci identified in the L55 subset, on chromosomes 8q24.2, 10q23.2, and 21q22, young families provide significantly more evidence for linkage than would be expected if the contribution to these loci were proportionate across all 573 families.

In the L45 subset, two loci (5q32 and 22q11) were identified above the nominal cutoff of LOD > 1.18. One of these loci (5q32) occurs in the overall genome-wide scan of 573 families. For both these loci, both permutation analyses and ordered subset analyses supported a role for a small number of the youngest AAD families contributing disproportionately to the evidence for linkage.

In our study, families segregating young-onset type 2

---

**TABLE 1**
Details of affected siblings from all 573 families and subsets

<table>
<thead>
<tr>
<th>L45</th>
<th>L55</th>
<th>G55</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>573</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Loci (marker)</th>
<th>Position*</th>
<th>Subset (number of families)†</th>
<th>L45</th>
<th>L55</th>
<th>G55</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q24–25 (D1S218)</td>
<td>209.8</td>
<td>0.06</td>
<td>2.38</td>
<td>0.50</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>5q13 (D5S647)</td>
<td>82.1</td>
<td>0.18</td>
<td>1.28</td>
<td>0.57</td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td>5q22 (D5S346)</td>
<td>160.8</td>
<td>2.13</td>
<td>0.70</td>
<td>0.54</td>
<td>1.22</td>
<td></td>
</tr>
<tr>
<td>7p15.3 (D7S493)</td>
<td>38.0</td>
<td>1.17</td>
<td>0.77</td>
<td>0.56</td>
<td>1.31</td>
<td></td>
</tr>
<tr>
<td>8p21–22 (D8S228)</td>
<td>42.2</td>
<td>0.13</td>
<td>1.68</td>
<td>1.43</td>
<td>2.55</td>
<td></td>
</tr>
<tr>
<td>8q24.2 (D8S272)</td>
<td>162.6</td>
<td>0.03</td>
<td>2.48</td>
<td>0.08</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td>10q23.2 (D10S1765)</td>
<td>125.2</td>
<td>0.56</td>
<td>3.07</td>
<td>0.12</td>
<td>1.99</td>
<td></td>
</tr>
</tbody>
</table>

*Position of peak LOD in centiMorgans in 573 families. †Empirically determined LOD score thresholds for suggestive linkage in L45, L55, and G55 subsets are 1.60, 1.56, and 1.56, respectively.

---

**TABLE 2**
Multipoint LODs for young AAD subsets for loci where LOD > 1.18 in complete set of 573 sibships

<table>
<thead>
<tr>
<th>Locus (marker)</th>
<th>Position*</th>
<th>Subset (number of families)†</th>
<th>L45</th>
<th>L55</th>
<th>G55</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>8q21.13 (D8S275)</td>
<td>106.6</td>
<td>0.36</td>
<td>1.44</td>
<td>0.00</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>21q22 (D21S266)</td>
<td>54.2</td>
<td>0.13</td>
<td>1.67</td>
<td>0.03</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>22q11 (D22S420)</td>
<td>0</td>
<td>2.51</td>
<td>0.47</td>
<td>0.08</td>
<td>0.34</td>
<td></td>
</tr>
</tbody>
</table>

*Position of peak LOD in centiMorgans.
TABLE 4
Multipoint LODs for all peaks with significance of increases compared to total families and ordered subset analyses

<table>
<thead>
<tr>
<th>Locus</th>
<th>Position (cM from pter)‡</th>
<th>LOD in young families (P)*</th>
<th>No. families</th>
<th>LOD (P)†</th>
<th>No. families (AAD cut-off)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q24–25 (D1S218)</td>
<td>214.2</td>
<td>2.38 (0.09)</td>
<td>245</td>
<td>2.68 (0.14)</td>
<td>376 (59)</td>
</tr>
<tr>
<td>5q13 (D5S418)</td>
<td>64.1</td>
<td>1.28 (0.43)</td>
<td>245</td>
<td>1.63 (0.52)</td>
<td>257 (56)</td>
</tr>
<tr>
<td>5q32 (D5S410)</td>
<td>170.3</td>
<td>2.13 (0.006)</td>
<td>39</td>
<td>3.67 (0.001)</td>
<td>23 (43)</td>
</tr>
<tr>
<td>7p15.3 (D7S493)</td>
<td>38.0</td>
<td>0.77 (NA)</td>
<td>245</td>
<td>2.55 (0.07)</td>
<td>5 (39)</td>
</tr>
<tr>
<td>8p12–22 (D8S258)</td>
<td>44.3</td>
<td>1.68 (NA)</td>
<td>245</td>
<td>2.46 (0.78)</td>
<td>572 (72)</td>
</tr>
<tr>
<td>8q21.13 (D8S275)</td>
<td>106.6</td>
<td>1.44 (0.38)</td>
<td>245</td>
<td>1.55 (0.41)</td>
<td>257 (56)</td>
</tr>
<tr>
<td>8q24.2 (D8S272)</td>
<td>166.9</td>
<td>2.45 (0.01)</td>
<td>245</td>
<td>3.00 (0.08)</td>
<td>284 (56)</td>
</tr>
<tr>
<td>10q23.2 (D10S1765)</td>
<td>125.2</td>
<td>3.07 (0.02)</td>
<td>245</td>
<td>3.19 (0.10)</td>
<td>315 (57.5)</td>
</tr>
<tr>
<td>21q22 (D21S666)</td>
<td>54.2</td>
<td>1.67 (0.03)</td>
<td>245</td>
<td>1.76 (0.10)</td>
<td>245 (55)</td>
</tr>
<tr>
<td>22q11 (D22S420)</td>
<td>0</td>
<td>2.51 (0.003)</td>
<td>245</td>
<td>1.89 (0.04)</td>
<td>39 (45)</td>
</tr>
</tbody>
</table>

*P for increase in LOD in young subset compared with 573 families; †P for maximized LOD score in analysis of young to old families; ‡position of peak LOD in young subset; §OSA uses linear function of KAC Maximum LOD score statistic and all other analyses are performed using exponential function so differences in LODs occur at chrom 21 and 22 despite same number of families in analyses. NA, not applicable as young subset LOD not higher than peak in 573 families.

diabetes provide more evidence for linkage than expected under the null hypothesis of no linkage, even though no loci individually reach genome-wide levels of significance. For example, in the L55 subset, we saw significantly (P = 0.01) more loci across the whole LOD score threshold range of 0.59–3.0 than expected by chance. For the L45 subset, only loci at the higher end of the LOD score range (LOD > 2.0) were observed in significant excess. In contrast, the G55 subset yielded no more evidence for linkage than that expected by chance under the null hypothesis of no genetic predisposition. This suggests families with subjects diagnosed at older ages have reduced power to detect linkage, which raises the question as to whether these families have any genetic component to their diabetess or whether their etiology is entirely environmental. We would argue that there is at least some genetic component in older families but that the genes involved may have too small an effect to be detected by linkage. It is well known that the detection of linkage is very sensitive to the relative risk associated with the locus and that most genome scans operate close to the limits of detection for loci of modest effect. For example, our genome scan of 573 families had ~76 and 55% power to detect loci of β values of 1.37 and 1.28, respectively (6). This is not sufficiently powerful to detect the predisposing variants described to date, such as the E23K variant in the KCNJ11 gene (Kir6.2) that predisposes to diabetes with an odds ratio of 1.23 (1.12–1.36) (12). In our large studies, the frequency of the predisposing K allele was very similar (40%) across all type 2 diabetes cases, regardless of AAD (10).

Results from previous genome scans of type 2 diabetes indicate that there is no one major predisposing locus in type 2 diabetes (6,13–20). Our results confirm that this is also the case in subjects affected at young ages: it is not possible to conclude which linkage signals are real susceptibility loci and which occur by chance. To gain a better understanding of which loci are likely to be real, we compared our results to those of other genome scans. The presence of replication across populations adds weight to the body of evidence for that locus. Of the loci in our study, those on chromosomes 1q24–25 and 8p21–22 have been observed in several other studies and noted before (6,17,19). In addition, a locus on chromosome 5q35 has been identified with a heterogeneity-LOD score of 1.37 in a study of 14 maturity-onset diabetes of the young (MODY) families (21) and with a nonparametric LOD score of 2.1 in 26 early-onset (≥2 members diagnosed ≤45 years) Scandinavian families. There were, however, three families in both the MODY and Scandinavian studies. These regions are close enough to our peak that they could represent the same gene. The locus recently identified on 5q in a large French family is likely to be too proximal to represent the same gene (22).

Other genome-wide scans of type 2 diabetes families have included analyses by young AAD. Unlike our study, however, none has shown that most of the evidence for linkage across the genome comes from the younger-onset subjects. Ghosh et al. (5) performed ordered subset analyses on regions identified in a total genome scan of 719 Finnish sibpair-based type 2 diabetic families (index case diagnosed 35–60 years). This showed that the 74 youngest AAD families provided increased evidence for linkage to a region on chromosome 6 (nominal P = 0.041 at 104.5 cM). A recent study of 472 Ashkenazi Jewish type 2 diabetic sibpairs also used ordered subset analysis to dissect identified loci but did not find any evidence for heterogeneity (20).

Other studies have focused exclusively on younger subjects. These include genome-wide scans in French (19), Utah Caucasian (17), and Scandinavian (23) type 2 diabetic families. Results from the well-replicated locus on 1q21–24 are particularly noteworthy. Subjects from the French and Utah Caucasian families linked to 1q were diagnosed on average at the relatively young ages of 49.5 and 50.6 years, respectively. In addition, in the Pima Indian genome scan (15), the evidence for linkage at this locus is increased in a subset of 55 sibpairs diagnosed <25 years.

It is not known whether subsetting by AAD and reassessing the evidence for linkage across the whole genome in other populations will produce similar results to ours. Such analyses will depend on the initial ascertainment criteria.
for the collection of families, ethnicity, and data on when collections and diagnoses were made. It is likely that the more affected family members ascertainment, the lower the AAD compared with a general type 2 diabetic population. Therefore, if the initial collection focused on younger subjects or larger families, there may not be much value in further stratification by young AAD. Different age cutoffs may be required in non-Caucasian populations as the AAD is generally lower compared with Caucasians. In addition, the prevalence of type 2 diabetes is increasing in young people due to changes in obesity and physical activity, which may mean even younger ages at diagnosis are needed to pull out the families who are most powerful for mapping genetic risk factors by linkage. Despite these caveats, we would argue that analyses by AAD are needed. Our results and the consistent finding that young-onset type 2 diabetic subjects have an increased family history provide a good a priori hypothesis that younger subjects will provide the strongest evidence for linkage. This could both increase the evidence for loci already identified and provide evidence for novel loci specific to younger-onset subjects.

In conclusion, families segregating young-onset type 2 diabetes provide disproportionate evidence for linkage in our U.K. Caucasian genome-wide scan. Younger subjects provide both a disproportionate amount of evidence for linkage to previously identified loci and evidence for additional loci. This suggests that, at least in Caucasian populations, efforts to define type 2 diabetes genes by linkage may be more powerful if focused on young AAD subjects.

ACKNOWLEDGMENTS
We thank William Duren (Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI) for the ordered subset software. T.M.F. is a career scientist of the South and West National Health Service Research Directorate. A.T.H. is a Wellcome Trust Research Leave Fellow. Diabetes U.K. and the Wellcome Trust kindly provided financial support for the genome scan. We thank all of the patients, nurses, and doctors who contributed to the Diabetes UK Warren 2 collection.

REFERENCES
