Evidence From a Large U.K. Family Collection That Genes Influencing Age of Onset of Type 2 Diabetes Map to Chromosome 12p and to the MODY3/NIDDM2 Locus on 12q24

Steven Wiltshire,1 Timothy M. Frayling,2 Christopher J. Groves,1,3 Jonathan C. Levy,3 Graham A. Hitman,1 Mike Sampson,3 Mark Walker,6 Stephan Menzel,1 Andrew T. Hattersley,2 Lon R. Cardon,1 and Mark I. McCarthy1,3

Additional information on genetic susceptibility effects relevant to type 2 diabetes pathogenesis can be extracted from existing genome scans by extending examination to related phenotypes such as age at disease onset. In this study, we report the reanalysis of data from 573 U.K. sibships ascertained for multiplex type 2 diabetes, using age at onset (assessed by the proxy measure of age at diagnosis) as the phenotype of interest. Genome-wide evidence for linkage to age at diagnosis was evaluated using both variance components and Haseman-Elston (HECOM) regression approaches, with extensive simulations to derive empirical significance values. There was broad agreement across analyses with six regions of interest (logarithm of odds [LOD] ≥1.18) identified on chromosomes 1qter, 4p15–4q12, 5p15, 12p13–12q13, 12q24, and 14q12–14q21. The strongest empirically “suggestive” evidence for linkage comes from regions on chromosome 12. The first region (12p13–12q13), peaking at D12S310 (variance components LOD [LODVC] = 2.08, empirical pointwise P = 0.0007; HECOM LOD [LODHECOM] = 2.58, P = 0.0010) seems to be novel. The second (12q24) peaking between D12S824 and D12S1659 (LODVC = 1.87, P = 0.0016; LODHECOM = 1.93, P = 0.0027) overlaps a region showing substantial prior evidence for diabetes linkage. These data provide additional evidence that genes mapping to these chromosomal regions are involved in the susceptibility to, and/or development of, type 2 diabetes.

Type 2 diabetes is a multifactorial disease of rising prevalence and increasing medical importance (1). The development of novel preventative and therapeutic strategies is largely contingent on an improved understanding of the molecular events involved in the pathogenesis of diabetes and related phenotypes such as obesity and insulin resistance (2). Identification and characterization of the susceptibility variants underlying predisposition to these conditions remains one of the most powerful routes to such understanding.

Most genome-wide linkage scans to date have concentrated on the analysis of type 2 diabetes as a discrete dichotomous trait (3). However, it seems likely that additional useful information on genetic susceptibility effects can be extracted from existing scans by extending examination to pertinent continuous phenotypes, such as those reflecting the action of modifier genes influencing age at disease onset and presentation (4,5).

We have previously reported (6) our analysis of the genome-wide evidence for linkage to type 2 diabetes in 573 U.K. affected sibships and recently demonstrated that the evidence for linkage to type 2 diabetes in this cohort comes disproportionately from patients with younger ages of onset (7). Here, we present an analysis of the same dataset designed to detect genes influencing age at onset (as assessed by the proxy measure of age at diagnosis) using two powerful complementary statistical approaches.

RESEARCH DESIGN AND METHODS

The subjects for study comprised the 573 full-sib pedigrees previously analyzed for evidence for linkage to type 2 diabetes (6). These pedigrees had been genotyped with 418 autosomal microsatellite markers (mean spacing 9.26 cM [Haldane units]), as previously described (6). Age at diagnosis was available for 1,233 affected individuals (1,223 offspring and 10 parents); mean age at diagnosis was 55.2 (SD 8.6) years in men (n = 661) and 56.2 (SD 8.8) years in women (n = 572). Genotypes were available for these 1,233 individuals and a further 173 individuals who were nondiabetic at the time of ascertainment.

The phenotype for analysis in this study was age at diagnosis in affected individuals, distributed as a continuous quantitative variable. These measures were self-reported by subjects at the time of study recruitment. Though, in principle, it would have been desirable to include information derived from...
the age at study among unaffected relatives (treated as a censored trait, for example using the hazard function based approach of Hanson and Knowler [8]), attempts to do so generated a markedly bimodal distribution of data points, a consequence of the incomplete extraction of inheritance information.

This distribution was not amenable to meaningful analysis with any available method; consequently, unaffected relatives were treated as unknown for the age at diagnosis phenotype. There was a marginally significant \(P = 0.049\) effect of sex on age at diagnosis. Heritability estimates for age at diagnosis, both before and after adjustment for sex, were obtained using MERLIN (9). BMI and other anthropometric measures around the time diabetes onset were not available, so adjustment was not possible for these measures. Sex-adjusted age at diagnosis was only slightly skewed (coefficient = \(-0.22\)) and platykurtic (coefficient = \(-0.42\)), and therefore untransformed sex-adjusted values were standardized and used in the linkage analyses.

**Linkage analysis.** We used two complementary approaches to analyze age at diagnosis: variance components analysis, implemented in GENEHUNTER 2 (10), and combined “squared sums-squared differences” Haseman-Elston (HECOM) regression, implemented in MERLIN REGRESS (9). For both, we estimated identity-by-descent (IBD) allele-sharing coefficients from all genotyped individuals. Variance components analysis partitions the variance of the quantitative trait, assumed to be multivariate normal, into components attributable to an additive major gene (\(\sigma^2_a\)), an additive polygenic effect (\(\sigma^2_p\)), and nonshared environmental effects (\(\sigma^2_e\)) at each position in the genome at which IBD allele-sharing coefficients are calculated. HECOM regression regresses the proportion of allele-sharing IBD between relative pairs on the effect of an additive major gene on the trait without explicitly partitioning the variance (9). Both methods produce a test statistic distributed asymptotically under the null hypothesis of no linkage as a 50:50 mixture of 0 and \(\chi^2\) with 1 degree of freedom, which can be expressed as an LOD score (variance components LOD \(\text{LOD}_{VC}\) and Haseman-Elston LOD \(\text{LOD}_{HECOM}\), respectively). In HECOM regression analysis we specify the mean of the \(\sigma^2\) (\(\sigma^2 = 1\), variance (1), and heritability of the sex-adjusted values of age at diagnosis observed in our dataset considered to represent an unselected sample of familial type 2 diabetes.

To detect colocalization of regions of linkage to age at diagnosis and obesity, we also undertook linkage analysis for BMI in diabetic members of the same families. BMI values (measured at the time of study) were logarithmically transformed and adjusted for the effects of age and sex using normative data from the adult U.K. population (11). These data were then standardized against the population data and analyzed using HECOM regression.

**Power calculations.** We simulated 1,000 replicates of a typical Warren 2 chromosome, 20 markers, mean spacing 9.26 cM, mean heterozygosity 78%, and mean missing genotype rate of 15%, for 573 pedigrees with the same pattern of genotyped and phenotyped individuals as that seen in our actual dataset. A single quantitative trait locus (QTL) (with a minor allele frequency of 0.1) influencing a normally distributed trait was simulated, without dominance, linked to these markers. The locus-specific heritability was estimated at 50% of the variance of the trait being equal to the observed values for age at diagnosis in our dataset (see Supplemental Fig. 3). These simulated data were analyzed with both variance components and HECOM regression methods. The proportion of replicates achieving a suggestive or significant linkage was taken as an estimate of the power of the study.

**Empirical significance simulations.** Variance components methods assume multivariate normality, and factors such as intrinsic trait non-normality and selection, and the effects of a major gene (12) that violate this assumption can lead to inaccurate statistical inference when asymptotic distributions are used. We therefore derived by simulation empirical estimates of the pointwise significance of the maximum LOD\(_{VC}\) scores in any regions showing evidence for linkage at LOD \(\geq 2.5\) (asymptotic pointwise \(P \leq 0.01\), vide supra) were observed on chromosomes 1qter, 4p15–4q12, 5p15, 12p13–12q13, 12q24, and 14q12–14q21 (Table 1). Variance components and HECOM regression analyses provide similar evidence for linkage to age at diagnosis at each of these regions. Moreover, the empirical pointwise significance levels of the maximum LOD scores in these regions were comparable with the asymptotic values, indicating no substantial violation of the underlying distributional assumptions regarding the phenotype. None of the regions showing evidence for linkage to type 2 diabetes in our previous analyses (6,7) were represented among the regions showing evidence for linkage to age at diagnosis.

**RESULTS**

The heritability of age at diagnosis in these pedigrees was found to be 63.6% before adjustment, confirming the familial clustering of this trait. The effect of sex was marginal (\(P = 0.049\), with age at diagnosis heritability increasing to 64.9% after adjustment for sex. All subsequent analyses were conducted on data adjusted for the effects of sex.

The results of the genome-wide quantitative linkage analysis of age at diagnosis with both variance components and HECOM regression approaches are summarized in Fig. 1 and Table 1. Regions showing evidence for linkage to age at diagnosis (taking a threshold of LOD \(\geq 1.18\), asymptotic pointwise \(P \leq 0.01\), vide supra) were observed on chromosomes 1qter, 4p15–4q12, 5p15, 12p13–12q13, 12q24, and 14q12–14q21 (Table 1). Variance components and HECOM regression analyses provide similar evidence for linkage to age at diagnosis at each of these regions. Moreover, the empirical pointwise significance levels of the maximum LOD scores in these regions were comparable with the asymptotic values, indicating no substantial violation of the underlying distributional assumptions regarding the phenotype. None of the regions showing evidence for linkage to type 2 diabetes in our previous analyses (6,7) were represented among the regions showing evidence for linkage to age at diagnosis.

The empirical genome-wide null distributions of LOD scores and independent regions of linkage, estimated using variance components and HECOM regression methods are shown in Fig. 2. The focus-counting simulations indicate more independent regions showing evidence for linkage than expected by chance for LOD scores from 1.18 to 2.00, with this excess oscillating around the 5% significance level. These simulations also indicate that when our data are analyzed by the variance components method evidence for linkage with LOD\(_{VC}\) = 1.56 occurs once per genome under the null hypothesis (defining the threshold for “suggestive” linkage), whereas a LOD\(_{VC}\) = 2.75 is associated with a genome-wide significance of 0.05 (“significant” linkage). With the HECOM regression analysis, equivalent empirical thresholds are LOD\(_{HECOM}\) = 1.69 and LOD\(_{HECOM}\) = 3.25, respectively.

The most interesting results were obtained for chromosome 12 and are shown in detail in Fig. 3. We observed a broad region showing evidence for linkage to age at diagnosis on 12p13–12q13, peaking at D12S310, (LOD\(_{VC}\) = 2.08, empirical pointwise \(P = 0.0007\); LOD\(_{HECOM}\) = 2.58, empirical pointwise \(P = 0.0010\)). Evidence for linkage was also observed on 12q24 between D12S324 and D12S1659 (LOD\(_{VC}\) = 1.87, empirical pointwise \(P = 0.0016\); LOD\(_{HECOM}\) = 1.93, empirical pointwise \(P = 0.0027\)). Both results are robust to the removal of phenotypic outliers with values \(>3\) SDs from the mean age at diagnosis. Both regions exceed the thresholds for “suggestive” linkage with both analytical approaches.

To detect colocalization between loci influencing age at
diagnosis and obesity, we sought evidence of linkage to BMI measures taken at the time of study recruitment. Modest, nominally significant ($P < 0.05$) peaks for BMI were indeed found on chromosome 5pter (D5S1981, LODHECOM = 0.61) and on chromosome 12 (D12S1725, LODHECOM = 0.80, D12S346-D12S78, LODHECOM = 0.69). Neither of these lie within 20 cM of the maximum age at diagnosis LOD scores on chromosome 12p12 and 12q24, and the LODHECOM scores for linkage to BMI at these age-at-diagnosis peaks were 0.265 and 0.090, respectively. There were no instances of overlap between BMI and age-at-diagnosis loci on other chromosomes (data not shown).

Simulations suggested that our study had good power to detect a suggestive linkage (66–78%) but low power to detect a genome-wide significant linkage (30–41%) to a QTL accounting for 30% of the trait variance with a residual 34.9% polygenic variance component using variance components and HECOM regression methods, respectively.

**DISCUSSION**

We have conducted a genome-wide analysis of a large set of families of U.K. origin seeking evidence for loci influencing age of diagnosis of type 2 diabetes using two complementary statistical approaches. We found evidence for linkage with a LOD ≥1.18 between age at diagnosis and markers on 1qter, 4p15–4q12, 5p15, 12p13–12q13, 12q24,

**FIG. 1.** Genome-wide variance components (A) and HECOM regression (B) multipoint linkage analyses of age at diagnosis in 1,233 affected individuals in 573 Warren 2 sibships.

**TABLE 1**

Multipoint linkage analysis of age at diagnosis in 1,233 individuals with type 2 diabetes indicating regions for which LOD ≥1.18 ($P ≤ 0.01$)

<table>
<thead>
<tr>
<th>Marker(s) at or flanking maximum LOD score [position(s) in cM from p-terminal of chromosome]</th>
<th>Maximum LOD$_{VC}$ score (variance components analysis)</th>
<th>Empirical pointwise $P$ value (asymptotic $P$ value)</th>
<th>Maximum LOD$_{HECOM}$ score (HECOM analysis)</th>
<th>Empirical pointwise $P$ value (asymptotic $P$ value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1S2836 (316.2)</td>
<td>1.21</td>
<td>0.0103 (0.0091)</td>
<td>1.31</td>
<td>0.0089 (0.0070)</td>
</tr>
<tr>
<td>D4S2971 (66.2)</td>
<td>1.49</td>
<td>0.0042 (0.0044)</td>
<td>1.61</td>
<td>0.0034 (0.0032)</td>
</tr>
<tr>
<td>D5S1981–D5S406 (0.6–11.7)</td>
<td>1.17*</td>
<td>0.0111 (0.0100)</td>
<td>1.56</td>
<td>0.0055 (0.0036)</td>
</tr>
<tr>
<td>D12S310 (39.1)</td>
<td>2.08</td>
<td>0.0007 (0.0010)</td>
<td>2.58</td>
<td>0.0010 (0.0003)</td>
</tr>
<tr>
<td>D12S324–D12S1659 (161.8–171.5)</td>
<td>1.87</td>
<td>0.0016 (0.0017)</td>
<td>1.93</td>
<td>0.0027 (0.0014)</td>
</tr>
<tr>
<td>D14S70–D14S288 (36.7–43.3)</td>
<td>1.43</td>
<td>0.0058 (0.0051)</td>
<td>1.71</td>
<td>0.0030 (0.0025)</td>
</tr>
</tbody>
</table>

Empirical pointwise $P$ values are estimated from 10,000 replicates of the data for each analytical method. *The LOD$_{VC}$ score for chromosome 5 is <1.18 but is included in the table as the LOD$_{HECOM}$ score is >1.18.
and 14q12–21. These results are consistent across both analyses. Although none of these regions achieves genome-wide significance individually, locus counting simulations indicate that we observed significantly more evidence for linkage to age at diagnosis, genome wide, than expected by chance. These data imply that several of the regions detected in this scan may harbor susceptibility loci influencing age at diagnosis. Only replication by, or concordance with, other genome scans can indicate which regions are likely to represent the true positives.

We have found no regions of concordance between the present age at diagnosis analysis and the previous genome-wide scans for type 2 diabetes in this dataset, either in the complete (6) or age-stratified (7) analyses, even in regions such as chromosome 1q that have been widely replicated (3) and therefore have the strongest likelihood of containing type 2 diabetes susceptibility genes. This lack of concordance may reflect intrinsic methodological differences between analyses that use age at diagnosis as a quantitative variable (where the aim, as here, is to detect genes influencing the age at onset of type 2 diabetes in susceptible individuals) and those that consider type 2 diabetes itself as the trait of interest, as in an affected subjects–only analysis.

An alternative conclusion is that these findings indicate that the genetic determinants of type 2 diabetes susceptibility and of disease progression are, at least in the U.K. population, distinct. There are few populations for which linkage data are available for both type 2 diabetes and age at diagnosis (5,15). Consequently, it is not yet possible to obtain a comprehensive view of the relationship between the genetic basis of susceptibility and that of diabetes onset or progression. Indeed, the influence of a type 2 diabetes disease gene on the spectrum of clinical phenotypes, from a strict susceptibility effect to a strict progression effect, may depend on genetic and environmental

---

**FIG. 2.** Actual and expected distributions of independent regions showing evidence for linkage from the genome-wide multipoint linkage analyses using variance components (A) and HECOM regression (B) in the 573 Warren 2 sibships.
background; a gene principally influencing susceptibility in one population may influence progression in another. The strongest evidence for linkage to age at diagnosis in the present study was seen in the two regions on chromosome 12. The first of these covered the pericentromeric region (12p13–12q13), with evidence peaking on 12p12.

There have been no previous reports of evidence for linkage to type 2 diabetes (or related traits) to 12p. However, a locus in the pericentromeric region of 12q has been implicated in two previous studies. In European-American pedigrees segregating early-onset autosomal-dominant type 2 diabetes, a heterogeneity LOD score of 2.5 (\(\alpha = 0.15\)) was observed between D12S1052 and D12S375 on 12q15 (16), and the GENNID study (17) observed a LOD score of 2.81 at D12S853 on 12q21, ~14 cM away, in European-American nuclear families with type 2 diabetes/impaired glucose homeostasis. Nevertheless, the region highlighted in these studies (12q15–21) lies some 40–55 cM (Kosambi units) from our peak on 12p12, and our maximal LOD score was only 0.759 (LODVC) in this region. Given the imprecision of QTL localization by linkage mapping (18), we cannot definitely discount the possibility that the peaks on 12p12 and 12q15–21 reflect the same genetic effect. Even so, given the sizeable difference between observed locations, we consider this unlikely and therefore believe that our findings on 12p12 provide evidence for a novel locus influencing type 2 diabetes progression.

The second region of interest on chromosome 12 lies on 12q24. This region represents one of the best replicated intervals for type 2 diabetes susceptibility with evidence for linkage generated in an appreciable number of previous studies (17,19–22). The present study provides further support for functional variants on chromosome 12q influencing type 2 diabetes susceptibility and/or progression. This region includes the gene TCF1 (encoding hepatocyte nuclear factor-1\(\alpha\)) responsible for a subtype of maturity onset diabetes of the young (23), although there is no evidence to suggest that TCF1 variation explains the chromosome 12 linkage.

The other regions highlighted in our scan have limited support from previous scans of type 2 diabetes susceptibility. The San Antonio Family Heart Study (15) observed evidence for linkage to type 2 diabetes and age at diagnosis on 1q43–44, some 15 cM(K) from our present evidence on 1q. The Amish Family Heart Study (24) observed evidence for linkage of HbA1c levels to markers on 4p, 20 cM telomeric from the peak LOD score on 4p in the present study, and evidence for linkage of glucose levels to markers overlapping our region of interest on 14q. The latter region was also implicated in linkage to type 2 diabetes in the GENNID study (17). A modifier locus influencing age at diagnosis in MODY3 families has been observed on 5p15 (25), some 15 cM from our evidence on 5p15.

Obesity is a known risk factor for type 2 diabetes and likely to influence the age of disease onset. This raises the possibility of overlap between genetic variation influencing obesity susceptibility and that involved in age at diabetes onset. As we did not have access to measures of obesity taken at the pertinent time point (i.e., around the time of disease onset), we were unable to test this possibility directly (for example, by adjusting for BMI at disease onset in the analyses). However, comparison of our linkage findings for age at diagnosis with those for BMI measures taken at the time of study recruitment fail to provide convincing evidence for colocalization of the
susceptibility regions for these two phenotypes. Though there were some modest signals for BMI on chromosomes 5q and 12, even allowing for the imprecision in linkage location estimates, the overlap with those for age at diagnosis is limited. Nevertheless, without access to BMI measurements earlier in life, we cannot entirely discount the possibility that some of the linkage to age at diagnosis is mediated through the effects of variation in BMI.

In summary, this study provides evidence for two loci on chromosome 12 that influence age at diagnosis of type 2 diabetes in a northern European population. For both of these, the current data add to the growing weight of evidence indicating that these regions harbor genes influencing susceptibility to, and/or development of, type 2 diabetes.

ACKNOWLEDGMENTS

The Warren 2 collection and genome scan were funded by Diabetes U.K. S.W. is a Wellcome Trust Career Development Fellow and L.R.C. is a Wellcome Trust Principal Fellow. This work was funded in part through the NIDDK award to the International Type 2 Diabetes Linkage Analysis Consortium (U01 DK068026).

We thank the physicians, nurses, and subjects who participated in the ascertainment and the researchers who contributed to the genotyping.

REFERENCES