

Tests for Genetic Interactions in Type 1 Diabetes Linkage and Stratification Analyses of 4,422 Affected Sib-Pairs

Grant Morahan,^{1,2} Munish Mehta,^{1,2} Ian James,³ Wei-Min Chen,⁴ Beena Akolkar,⁵ Henry A. Erlich,⁶ Joan E. Hilner,⁷ Cécile Julier,^{8,9} Jørn Nerup,¹⁰ Concepcion Nierras,¹¹ Flemming Pociot,¹² John A. Todd,¹³ and Stephen S. Rich,⁴ for the Type 1 Diabetes Genetics Consortium

OBJECTIVE—Interactions between genetic and environmental factors lead to immune dysregulation causing type 1 diabetes and other autoimmune disorders. Recently, many common genetic variants have been associated with type 1 diabetes risk, but each has modest individual effects. Familial clustering of type 1 diabetes has not been explained fully and could arise from many factors, including undetected genetic variation and gene interactions.

RESEARCH DESIGN AND METHODS—To address this issue, the Type 1 Diabetes Genetics Consortium recruited 3,892 families, including 4,422 affected sib-pairs. After genotyping 6,090 markers, linkage analyses of these families were performed, using a novel method and taking into account factors such as genotype at known susceptibility loci.

RESULTS—Evidence for linkage was robust at the HLA and *INS* loci, with logarithm of odds (LOD) scores of 398.6 and 5.5, respectively. There was suggestive support for five other loci. Stratification by other risk factors (including HLA and age at diagnosis) identified one convincing region on chromosome 6q14 showing linkage in male subjects (corrected LOD = 4.49; replication $P = 0.0002$), a locus on chromosome 19q in HLA identical sibs (replication $P = 0.006$), and four other suggestive loci.

CONCLUSIONS—This is the largest linkage study reported for any disease. Our data indicate there are no major type 1 diabetes subtypes definable by linkage analyses; susceptibility is caused by actions of HLA and an apparently random selection from a large number of modest-effect loci; and apart from HLA and *INS*, there is no important susceptibility factor discoverable by linkage methods. *Diabetes* 60:1030–1040, 2011

From the ¹Centre for Diabetes Research, Western Australian Institute for Medical Research, University of Western Australia, Crawley, Australia; the ²Centre for Medical Research, University of Western Australia, Crawley, Australia; the ³Centre for Clinical Immunology and Biomedical Statistics, Murdoch University, Perth, Australia; the ⁴Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia; the ⁵Division of Diabetes, Endocrinology, and Metabolic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland; ⁶Roche Molecular Systems, Pleasanton, California; the ⁷Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama; ⁸INSERM UMR-S 958, Faculté de Médecine Denis-Diderot, Paris, France; the ⁹University Paris 7, Paris, France; the ¹⁰Steno Diabetes Center, Gentofte, Denmark; the ¹¹Juvenile Diabetes Research Foundation, New York, New York; the ¹²Science Park, University Hospital Glostrup, Glostrup, Denmark; and the ¹³Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, U.K.

Corresponding author: Grant Morahan, gem@waimr.uwa.edu.au.

Received 23 August 2010 and accepted 15 December 2010.

DOI: 10.2337/db10-1195

This article contains Supplementary Data online at <http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db10-1195/-/DC1>.

© 2011 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

Type 1 diabetes is an autoimmune disease in which the insulin-producing β -cells are destroyed. The defective immune mechanisms in type 1 diabetes have not been identified, although clearly both genes and environmental factors contribute to risk (1,2). Many studies have been performed on the genetics of type 1 diabetes, and identification of the risk genes is ongoing. The Type 1 Diabetes Genetics Consortium (T1DGC) was established in response to the need to identify type 1 diabetes risk genes (3).

The T1DGC has assembled DNA, sera, cell lines, and data from 17,129 individuals in 3,892 affected-sib-pair (ASP) families, representing the largest family collection in any immune-mediated disease. Using these resources, the T1DGC performed smaller genome-wide linkage studies (4,5) and association studies, finding 21 new loci (6). In addition, the T1DGC conducted the most detailed investigation of the HLA complex in disease, characterizing over 3,000 single nucleotide polymorphisms (SNPs), and independently tested all previously reported type 1 diabetes susceptibility genes (7,8; and associated special reports).

From these and other studies, over 50 loci have been identified that affect the risk of developing type 1 diabetes (6) (www.t1dbase.org). Foremost among these is the HLA complex, which is long recognized as the most important risk factor in type 1 diabetes and other immune diseases. The second locus identified was *INS*, but this and other non-HLA loci have relatively minor risk effects comparable to loci mapped in other common diseases, with risk estimates typically between 1.05 and 2.0 (6).

Despite the T1DGC's success, not all the estimated familial heritability for type 1 diabetes has been found. Twin studies suggest ~80% of clustering is a result of sharing of susceptibility alleles at multiple loci. This issue of "missing heritability" also arises in other complex genetic diseases (9). A number of other issues pertaining to type 1 diabetes remain unresolved. For example, can it be subdivided into disease subtypes with different genetic etiology? Can we identify major genetic interactions in susceptibility? Are there loci with multiple rare risk alleles that may have gone undetected in association studies? Does genomic instability in the form of structural variants play a role?

Motivated by these issues, we report here the genotyping of the final 1,583 T1DGC ASPs and perform linkage analyses on a total of 3,892 families and 17,129 individuals. The T1DGC family collection provides 80% power to search for loci undetectable by genome-wide association SNP scanning with effects of the same magnitude as the

insulin gene with allelic odds ratio of ~ 2 and an allele frequency of 0.3. If there is residual “missing heritability” (i.e., loci that were not found by the SNP studies), this large dataset is a valuable resource to search for it. In this report, we describe a linkage search for genetic interactions in type 1 diabetes.

RESEARCH DESIGN AND METHODS

This study was approved by review boards of all contributing institutions, and appropriate informed consent was obtained from families. Inclusion criteria have been reported (3–5). Briefly, a family must contain at least one affected sib-pair; “affected” indicates a type 1 diabetes diagnosis before the age of 35 years with insulin required within 6 months of diagnosis. Samples from 1,505 newly recruited families, as well as previously untyped members from 84 families recruited earlier (4,5), were available for genotyping. Families came from four T1DGC networks: Asia-Pacific ($n = 116$ families), Europe ($n = 628$), North America ($n = 807$), and the U.K. ($n = 38$). Of samples sent for genotyping, 1,396 (88%) families had two affected full siblings, 63 (4.0%) families had three, and 3 (0.2%) families had four or more. Four pedigrees also provided affected half-siblings. The remaining 117 (7.4%) families included

samples from other family members that were unavailable for genotyping for the previously reported cohorts; thus, they were available for the analyses of the total T1DGC dataset. Both parents were available for 914 families, whereas 397 families consisted of a single parent.

Genotyping. Genotyping of 6,090 SNPs with an average 0.58-cM genome-wide spacing was carried out at the Center for Inherited Disease Research using the Illumina Human Linkage-12 Beadchip. Karyotype locations of SNPs were from the T1DBase.

Pedigree checking. Genotype data were evaluated for Mendelian errors using PedCheck (10) and PREST (11). Merlin’s Pedwipe function (12) identified and resolved inconsistencies within families. Five families (seven ASPs) were removed because of nonresolvable family issues. Another 19 families (19 ASPs) were excluded by Merlin. A total of 1,487 families passed all quality-control filters and were analyzed as T1DGC cohort 3. There were non-ASP samples from 78 previously incomplete families (4,5). Combined with the previously reported cohorts (4,5), there was a total of 4,422 ASPs from 3,892 families.

Pedigree stratification. Sib-pairs were selected for stratification analyses using SibShipper (www.sysgen.org.au), a suite of tools for pedigree manipulation, and PLINK (13).

Statistical analyses. Nonparametric linkage analyses were performed using the linear NPL model (14) implemented in Merlin (12), which accounted for linkage disequilibrium between SNPs by selecting $r^2 > 0.10$ to define SNP

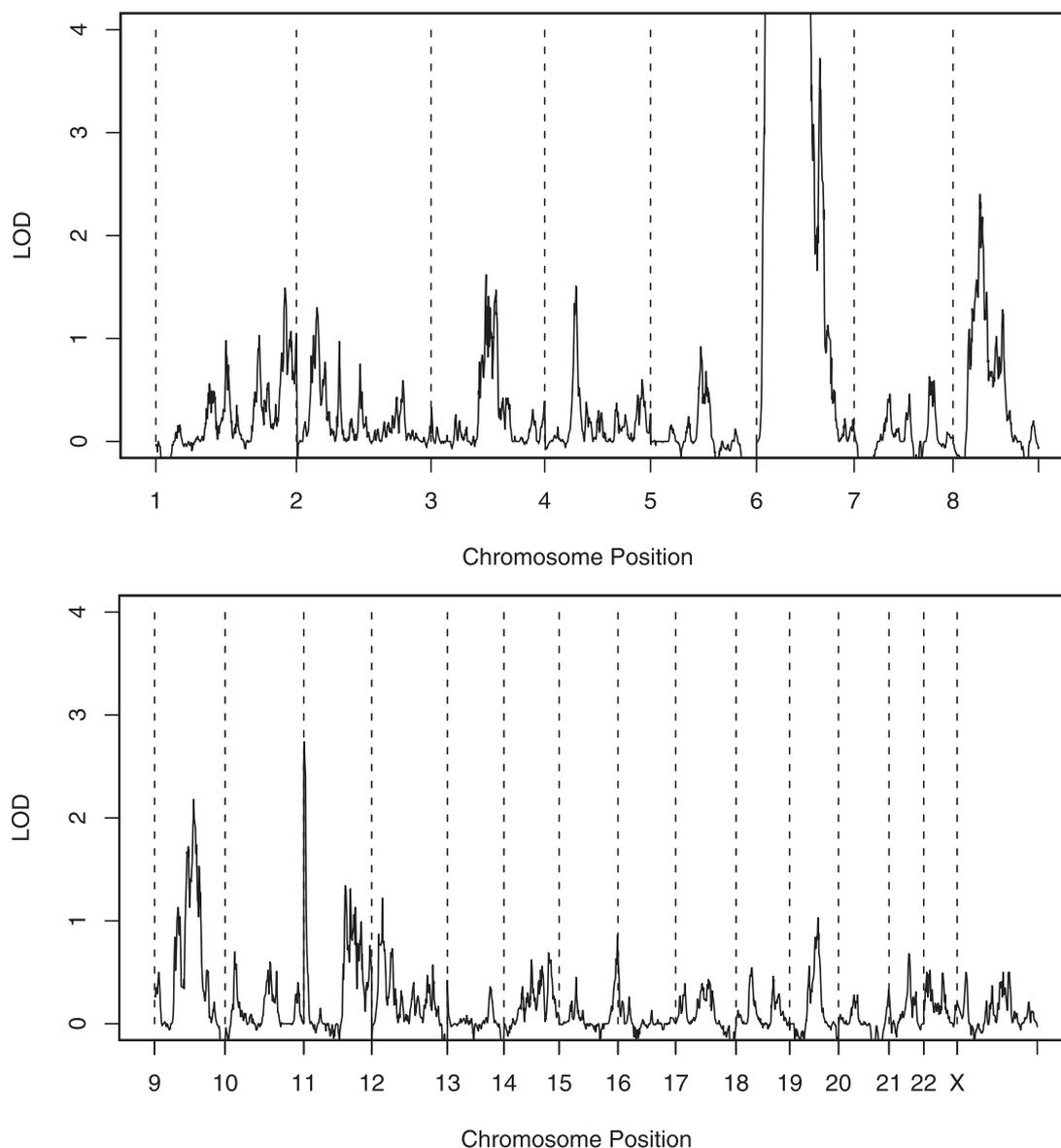


FIG. 1. Genome-wide linkage analyses of the final T1DGC cohort. Nonparametric linkage calculations of 6,090 SNPs segregating in 1,487 new affected sib-pair families were performed using Merlin (12). Kong and Cox LOD scores (14) are shown for each chromosome. The peak LOD on chromosome 6 was over 130, but the scale is truncated to show linkage scores for other loci.

clusters. We developed a novel method for comparing stratified sets of sib-pairs. Profiles of logarithm of odds (LOD) scores were compared statistically by treating scores as correlated time-series, with chromosome position as "time." The automated expert modeler facility in SPSS 17.0 was used to obtain the optimal parsimonious time-series model with approximately independent residuals. For both profiles, this produced a model in which the difference between each successive pair of scores regressed linearly on the preceding difference (i.e., a differenced first-order autoregressive model). Peaks in the original profiles would manifest as increased variation in the residuals at positions spanning the peak. The absolute values of residuals were then compared across sections of the chromosome via ANOVA. More details are provided below.

RESULTS

Linkage results of the third T1DGC cohort. Each member of the 1,583 newly recruited T1DGC families was genotyped at 6,090 SNPs genome-wide. Nonparametric linkage analysis was performed; results are summarized in Fig. 1 and Table 1. Peak linkage was observed on chromosome 6p21.3, the location of the HLA class II genes contributing the main type 1 diabetes risk. Apart from three other peaks on chromosome 6, no other region yielded genome-wide significant linkage (i.e., LOD >3.6). The three peaks on chromosome 6 were at SNPs *rs2296412* (chromosome 6q14.2 and 92.7 cM), *rs6934871* (chromosome 6q15 and 95.7 cM), and *rs873460* (chromosome 6q22.31 and 123.1 cM); these chromosomes had LOD scores (unadjusted for HLA linkage) of 10.71, 10.61, and 3.78, respectively (see Table 1 for corrected LOD scores).

The highest linkage score on other chromosomes was 2.74 at chromosome 11p15, the location of *INS*. The only other locus with an LOD score above the suggestive linkage threshold of 2.2 was on chromosome 8p12, with peak linkage at SNP *rs1836851*. This chromosome region has not been implicated in previous genome-wide linkage or association scans.

Complete T1DGC linkage results. The total T1DGC collection of 4,422 ASPs constitutes the largest linkage study conducted for any disease. Results from the linkage

analysis of the total dataset are presented in Fig. 2, and a summary is shown in Table 1. As expected, the evidence for linkage to the HLA region increased, with a peak LOD score of 398.6 (Fig. 2B), reinforcing the importance of this complex in the etiology of type 1 diabetes.

Chromosome 6 linkage was further examined, taking into account the HLA linkage that may mask other susceptibility loci on this chromosome. One way to account for this is by calculating the expected LOD (ELOD) based on the decay of linkage from HLA, assuming a Kosambi map function (i.e., independent crossovers). Region(s) showing linkage significantly higher than the ELOD may contain gene(s) that affect type 1 diabetes risk. These analyses were carried out as previously described (4,5). This analysis showed additional linkage signals on chromosome 6q (Fig. 2B; Table 1), which overlapped with the previously reported locus *IDDM15* (4,5,15).

The second locus that achieved genome-wide significance was the *IDDM2 (INS)* locus on chromosome 11p15, with an LOD of 5.53. A locus on chromosome 19q almost reached genome-wide significance (LOD = 3.3). Apart from these, across the rest of the genome there was no increase in the evidence for linkage at loci that previously (4,5) had suggestive scores. Support for some loci was reduced (e.g., the chromosome 2q locus including the *CTLA4* gene [4,5] fell from an LOD of 3.35 to an LOD of 3.11). There were two new regions with suggestive support levels defined in this study: the evidence for linkage at chromosomes 2q13 and 8q21 increased to LODs of ~2.7 (SNP *rs1439287* at 111.6 Mb and SNP *rs1902866* at 87.8 Mb). No significantly associated loci had been mapped to these regions (6).

Stratification by HLA sharing status. Previously, support for some non-HLA type 1 diabetes susceptibility genes was increased after stratification of ASPs according to biologically relevant criteria, such as HLA genotype or sex (16–20). We tested whether subgroup analyses such as these would provide stronger evidence of linkage. Our strategy was to conduct these tests in the previously

TABLE 1
Suggestive (LOD >2.2) and significant (LOD >3.6) loci from genome linkage scans of the T1DGC family collections

| Families | Position (cM) | LOD | P | LOD-1 support interval | Reported in T1DGC sets 1 and 2 (4,5) |
|------------------|---------------|-------|---------|------------------------|--------------------------------------|
| T1DGC set 3 | | | | | |
| Chromosomes | | | | | |
| 6p21 | 52.5 | 135.7 | | 51.0–52.5 | Yes |
| 6q14 | 92.7 | 5.1* | | 90.7–93.3 | Yes |
| 6q15 | 95.7 | 6.2* | | 93.3–98.1 | Yes |
| 6q22 | 123.1 | 3.23* | 0.00006 | 120.5–126.9 | No |
| 8p12 | 52.7 | 2.4 | 0.0004 | 48.8–60.9 | No |
| 11p15 | 1.05 | 2.74 | 0.0003 | 0–4.7 | Yes |
| Sets 1, 2, and 3 | | | | | |
| Chromosomes | | | | | |
| 2q13 | 123.8 | 2.68 | 0.0002 | 120.3–126.8 | No |
| 2q31 | 194.2 | 3.11 | 0.00008 | 191.8–193.8 | Yes |
| 6p21 | 52.0 | 398.6 | | 51.5–52.5 | Yes |
| 8q21 | 95.5 | 2.65 | 0.0002 | 93.5–101.5 | No |
| 11p15 | 2.5 | 5.53 | | 0–5.0 | Yes |
| 19q13 | 58.0 | 3.30 | 0.00005 | 54.5–63.5 | Yes |

Summary of linkage analyses for the third and final set of T1DGC families. Positions of LOD score peaks are shown for the chromosome band (taken from www.tlbase.org for the nearest SNP) and cM position. P values shown for LODs <5. *LOD scores on chromosome 6q were adjusted for HLA linkage by subtracting the ELOD (see text). Peaks were defined as points at least 1 LOD unit above the surrounding linkage plot; the additional families in the combined dataset obscured the separable peaks observed on chromosomes 6q14 and 6q15 in set 3 alone.

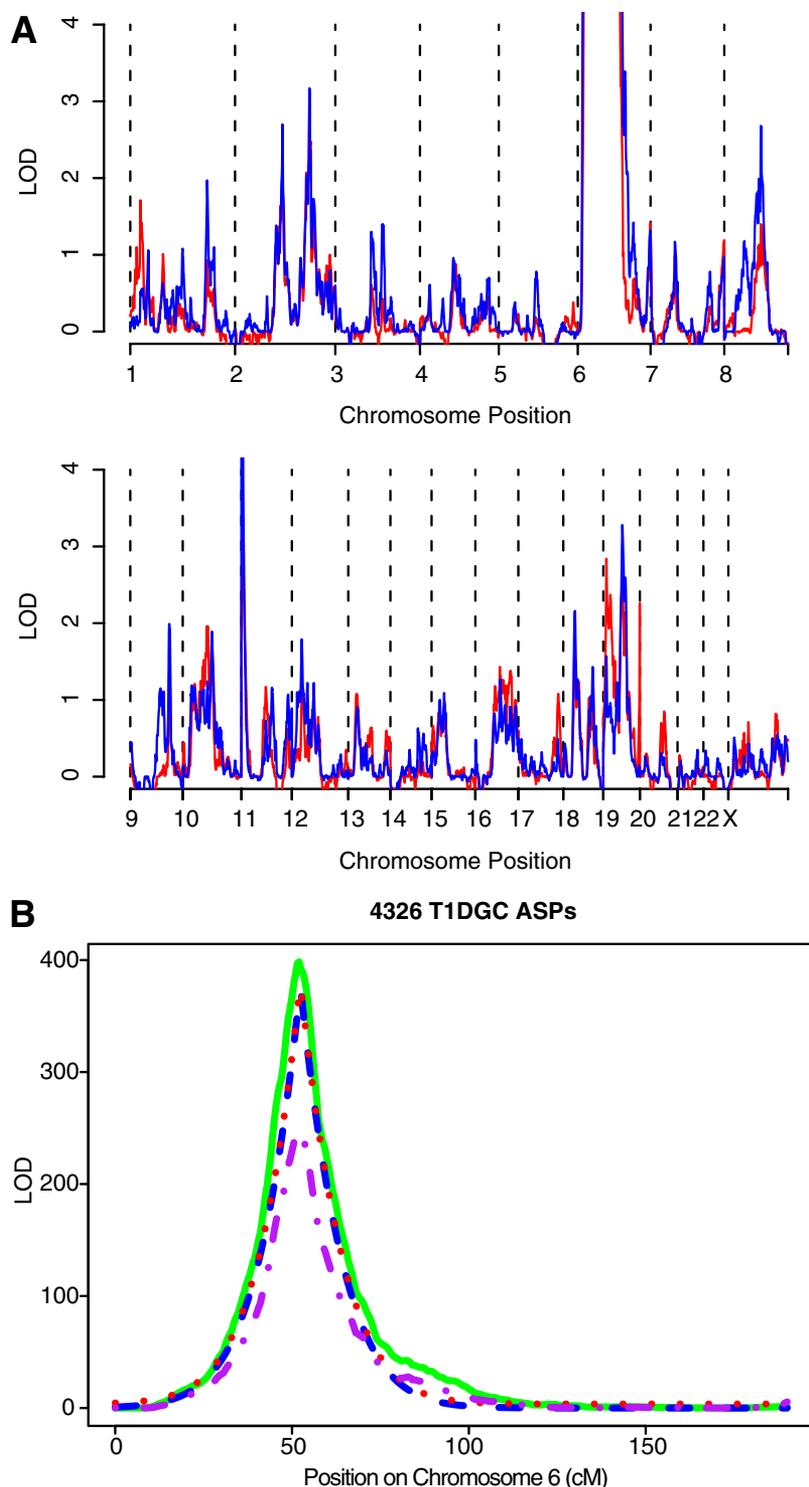


FIG. 2. Evidence of linkage from the entire T1DGC ASP family collection. *A*: Pedigree files were merged from the latest set of T1DGC families and those reported earlier (4,5) (which are shown in red). LOD scores for the entire dataset were calculated as described in Fig. 1 and shown in blue. The scale is truncated at 4, so it does not indicate the height of the peaks on chromosome 6 (~400) (see *B*) and 11 (~5). *B*: Linkage of chromosome 6 markers in the T1DGC family collection. LOD scores were calculated for the complete T1DGC family collection (solid green line) and compared with the cohort 3 families (dashed purple line). In addition, LOD scores were calculated correcting for linkage to HLA: the ELOD score based on decay of linkage from HLA is shown as a dashed blue line and ELOD + 3.6 as a dotted red line. The increase in observed LOD above ELOD in the region ~90–100 cM is consistent with additional type 1 diabetes susceptibility gene(s) mapping to this region.

reported sets of 2,658 families, with replication of any significant findings in the final (set 3) T1DGC cohort. Taking into account multiple testing, a threshold for suggestive linkage was set at an LOD = 3.28.

The first stratification was HLA linkage status. Siblings sharing two HLA haplotypes that were identical by descent (IBD) formed one group. The second stratification group (HLA non-IBD) consisted of siblings sharing one or zero

haplotypes by descent (but note that these siblings may be identical by state for particular HLA class I or class II alleles). Genome-wide linkage analyses on these two sets showed no locus with an LOD >3.28 (Fig. 3). However, there was some (uncorrected) evidence of two loci contributing to type 1 diabetes on chromosome 19, each of which showed linkage in different sib-pair sets depending on the HLA sharing status (Fig. 4). These peaks were at *rs1548506* (LOD = 2.87; HLA IBD ASPs) and at *rs966591* (LOD = 3.03; HLA non-IBD ASPs). To test the statistical differences in linkage between stratified sets, we applied a novel approach based on time-series analysis. Profile shapes were reflected empirically in changing variability of residuals from fitted-differenced autoregressive models. Chromosome 19 was then divided approximately into thirds, and the relative changes in variation in the residuals between the two strata across these regions were assessed through the interaction term in two-way ANOVA. Square

roots of the absolute values of the residuals were used. The overall analysis ($P < 10^{-15}$) indicated different linkage profiles. In particular, variation was higher in non-IBD than IBD ASPs at 19p loci ($P = 2.0 \times 10^{-5}$), with the converse at 19q loci ($P = 1.3 \times 10^{-12}$) (Wilcoxon tests) (Fig. 4B).

Stratification by HLA-DRB1 status. ASPs were stratified according to HLA-DRB1 haplotype status. Pairs of siblings were selected using the following criteria: both with DRB1*03 but not DRB1*04 alleles ("DR3/x"); both with DRB1*04 but not DRB1*03 alleles (DR4/x); both heterozygous for HLA-DRB1 high-risk alleles (i.e., DRB1*03/04; DR3/4); and neither affected sibling having DRB1*03 nor DRB1*04 (DRx/x). Linkage analyses were performed on each group. Apart from HLA, no locus exceeded our threshold of corrected LOD = 3.28. Only one locus exhibited uncorrected LOD >3 (figure not shown): this had an LOD of 3.18 in the DR3/4 ASP group at the SNP *rs424074*, located at 99 cM on chromosome 16q23.1.

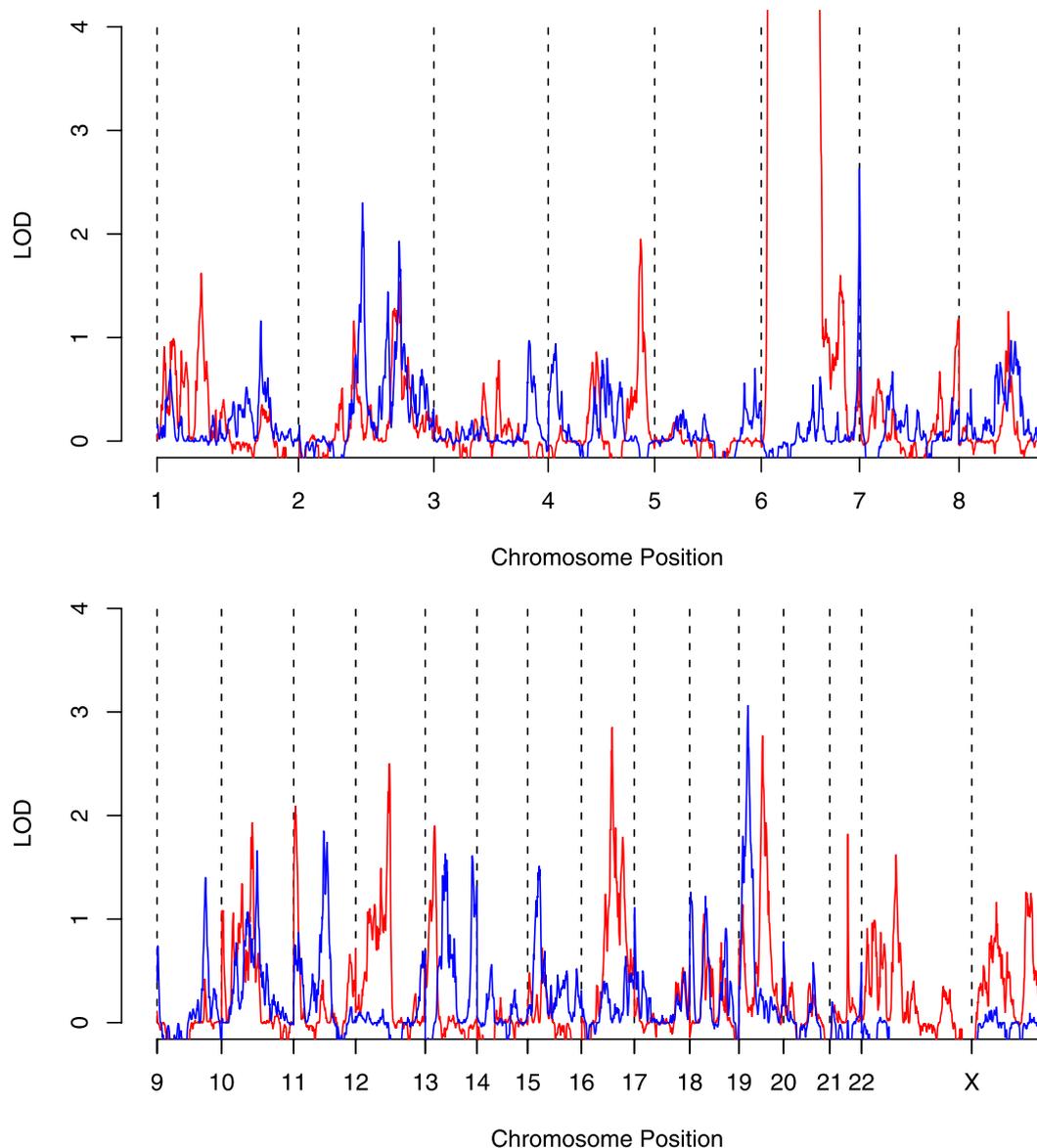


FIG. 3. Genome-wide linkage analysis of siblings sharing 2 (IBD) or 1/0 (non-IBD) HLA haplotypes IBD. Pedigree files of T1DGC sets 1 and 2 were constructed using SibShipper (www.sysgen.org.au) from ASPs that shared 2 HLA haplotypes IBD and from the remaining set of siblings that shared one or zero alleles IBD (for convenience, the latter set is referred to as non-IBD). LOD scores were calculated for each set as described in Fig. 1. The scores for the IBD ASPs are shown in red and non-IBD ASPs are shown in blue.

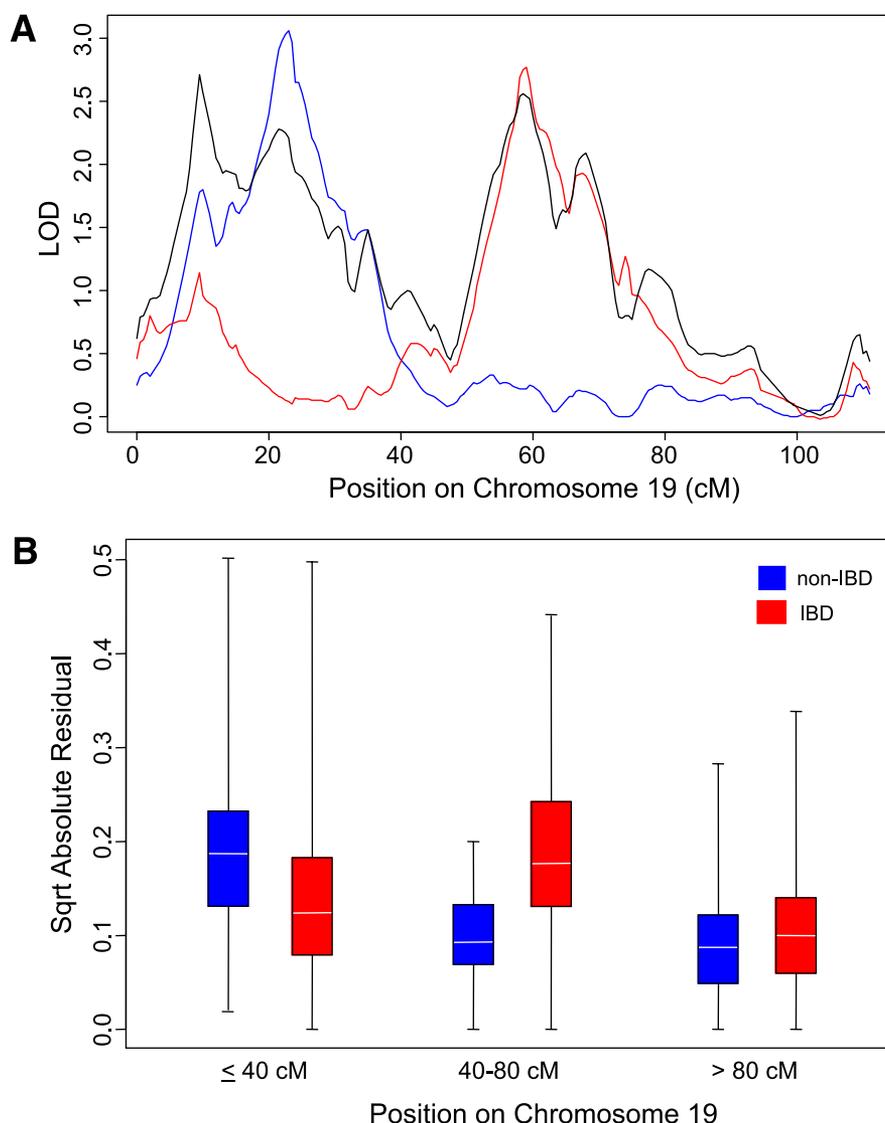


FIG. 4. Linkage analysis of HLA IBD and HLA non-IBD sib-pairs for markers on chromosome 19 of ASPs from T1DGC sets 1 and 2. **A:** LOD scores on chromosome 19 are shown for the total dataset (black), HLA IBD siblings (red), and HLA non-IBD siblings (blue). Stratification according to HLA sharing indicates two different linkage peaks on this chromosome. **B:** ASPs who were HLA IBD (red boxes) or non-IBD (blue boxes) were compared for differences on chromosomes 19p (position <40 cM), 19q (40–80 cM), and 19qtel (>80 cM) via residuals from differenced autoregressive models as described in RESEARCH DESIGN AND METHODS. Relative variation in residuals across the chromosome differed between subgroups ($P < 10^{-15}$), suggesting differing linkage profiles. P values for differences between HLA IBD and non-IBD ASPs for the three regions of chromosome 19 are 19p, $P = 0.00002$; 19q, $P = 1.3 \times 10^{-12}$; and 19tel, $P = 0.06$. Boxes represent the mid-50% of values; horizontal lines indicate means.

Stratification by *INS* genotype status. Many studies have shown that *INS* variants cause differences in insulin gene expression and are significantly associated with type 1 diabetes susceptibility (21,22). We tested whether these variants preferentially interacted with other loci by stratifying ASPs by *INS* genotypes. Two sets of ASPs were studied: those homozygous for the A allele at the –23/HphI site (*rs689*) and those who carried at least one copy of the T allele. The –23/HphI “A” allele shows almost complete linkage disequilibrium with type 1 diabetes-associated *INS* VNTR class I alleles (23). Linkage analysis of the two sets revealed a novel locus in the AA homozygous siblings, with a maximum LOD of 4.19 at 83.5 cM near *rs6988179* on chromosome 8q13.3 (Supplementary Fig. 1; Table 2). ASPs who had at least one T allele showed an apparent linkage peak on chromosome 6q13 (uncorrected LOD = 4.83 near *rs1416546* at 87 cM).

Stratification by sex. Type 1 diabetes, unlike other autoimmune diseases, affects male and female subjects approximately equally. Previous studies (19,20) have suggested susceptibility loci that preferentially affected one sex. We examined this issue on a genome-wide basis. Families were selected in which only male or only female ASPs were affected. There was no difference in LOD scores at HLA for the two sets stratified by sex (female ASPs: LOD = 62.37; male ASPs: LOD = 62.11). Apart from linkage to HLA, there was no other significant linkage in the female-only ASPs set. In the male-only ASPs, five loci yielded LOD scores >3: three loci were mapped to chromosome 6q and the other two loci resided on chromosomes 11p and 19p (Supplementary Fig. 2). Markers at these peaks were *rs1398576* (chromosome 6q14.1; LOD = 5.57), *rs1158747* (chromosome 6q21; LOD = 3.46), *rs1569741* (chromosome 6q22.32; LOD = 3.91; not

TABLE 2
Test for replication of stratified loci in T1DGC set 3 families

| Stratification | cM | LOD | <i>P</i> | Corrected LOD* | LOD-1 interval | <i>P</i> in set 3 | GWAS SNP |
|----------------|-------|-------|----------|----------------|----------------|-------------------|-----------------|
| HLA IBD | | | | | | | |
| Chromosome | | | | | | | |
| 19q13 | 59.0 | 2.83 | 0.0002 | 1.75 | 54.0–64.5 | 0.006 | <i>PRKD2</i> |
| HLA MIS | | | | | | | |
| Chromosome | | | | | | | |
| 19p | 23.0 | 3.04 | 0.00009 | 1.98 | 18.0–27.5 | NS | — |
| HLA DR3/4 | | | | | | | |
| Chromosome | | | | | | | |
| 16q23 | 99.1 | 3.18 | 0.00006 | 2.10 | 96.0–110.0 | 0.06 | <i>CTRB2</i> |
| INS† | | | | | | | |
| Chromosomes | | | | | | | |
| 6q13 | 87.0 | 3.34‡ | 0.00004 | 2.26 | 81.5–90 | 0.004 | |
| 8q13 | 83.5 | 4.19 | 0.00001 | 3.11 | 80.0–88.0 | NS§ | — |
| Male subjects | 90.0 | 2.90‡ | 0.00013 | 1.59 | 86.5–91.5 | 0.0002 | |
| Chromosomes | | | | | | | |
| 6q14 | | | | | | | |
| 6q21 | 116.0 | 3.10‡ | 0.00008 | 2.02 | 109.0–118.5 | 0.03 | — |
| 6q22 | 127.0 | 3.76‡ | 0.00002 | 2.68 | 120.0–130.0 | 0.09 | <i>CENPW</i> |
| 11p15 | 2.5 | 3.18 | 0.00007 | 2.10 | 0–7.0 | NS | <i>INS</i> |
| 19p13 | 9.5 | 3.22 | 0.00006 | 2.14 | 5.0–13.5 | NS | <i>C19orf19</i> |
| AAD | | | | | | | |
| Chromosomes | | | | | | | |
| 7q36 | 172.5 | 3.13 | 0.00009 | 2.05 | 163.5–179.5 | NS§ | — |

Loci showing LOD scores >3 from the stratified analyses and those showing significant differences between stratified sets (Fig. 4) were tested in the third collection of T1DGC families. Pedigree files of the relevant chromosomes from the third T1DGC cohort were stratified according to criteria presented above, and *P* values were calculated by Merlin. Genome-wide association scan SNPs that showed significant association with type 1 diabetes in or near to these regions are shown as candidates for these loci. NS, not significant; *P* > 0.1. *Final corrected LODs were obtained by subtracting 1.08, the log of the number of stratification tests (*n* = 12) (32). †Note that *INS* stratification shows that the 6q13 score shows a signal from siblings who each have at least one protective T allele, whereas the 8q13 signal comes from those homozygous AA (i.e., the *INS* risk allele). We interpret this as suggesting that the chromosome 6q locus is able to overcome the protective affect of the *INS* T allele. In contrast, the 8q13 locus may interact with the *INS* susceptibility alleles to increase disease risk. ‡LOD scores for chromosome 6q loci were adjusted for residual HLA linkage by subtracting ELOD (see text). §Markers outside the LOD-1 interval gave *P* values of 0.003 and 0.01, respectively, in appropriately stratified set 3 families.

corrected for HLA linkage), *INS* (LOD = 3.18), and *rs1688128* (chromosome 19p13.3; LOD = 3.22), respectively. **Stratification by age at diagnosis.** Type 1 diabetes usually has an onset before puberty, but a significant proportion of cases are diagnosed at later ages. It has been proposed that there may be a different genetic basis for these age effects (24). We were able to perform a genome scan on 304 ASPs who were each diagnosed after 15 years of age. In addition, we analyzed 279 ASPs in which each sibling was diagnosed before the age of 5 years. Only one locus approaching suggestive significance was found. This one locus had an LOD = 3.13 near *rs38993* on chromosome 7q36.2 in the older-onset ASPs (Supplementary Fig. 3). This is a novel region for type 1 diabetes susceptibility.

Testing chromosome 6 loci with correction for HLA linkage. Two stratified sets (male subjects and *INS* AA genotype-affected sib-pairs) showed apparent linkage on chromosome 6q, to which the *IDDM5*, *IDDM8*, and *IDDM15* loci were previously mapped (15,18). ELOD scores were calculated as described above for the stratified sets, showing potential linkage (male sib-pairs; and *INS* T allele-bearing sib-pairs) and, for comparison, the complementary stratified set (female sib-pairs; *INS* AA genotype sib-pairs). The results show that the respective peaks met or exceeded ELOD + 3.6, indicating additional evidence for linkage in the relevant stratified sets (Fig. 5). These peaks map to different locations. Thus, the *IDDM15* locus may reflect a combination of two (or more) loci on

chromosome 6q that interact with different susceptibility loci. Candidates for these loci include the four confirmed genome-wide association study (GWAS) SNP-associated regions on chromosome 6q (6).

Testing stratification loci in T1DGC set 3 families. Families from the final T1DGC cohort were stratified as above, but the analyses were only performed for the loci with LOD >3 or with significant differences between stratified sets, as discussed above. Results of these confirmation tests are summarized in Table 2. Three stratified loci could be confirmed unequivocally: the locus on chromosome 19q for HLA IBD siblings and the loci on chromosome 6q13/14 for male and *INS* “Tx” siblings; the same gene may be responsible for these chromosome 6q loci. Another two loci had significant scores on the same chromosome but outside the region within one LOD unit of the peak (i.e., LOD-1): 8q13 for *INS* AA genotype siblings and 7q36 for older age-of-onset siblings. In complex diseases, the location of causative genes may be displaced compared with both the peak and LOD-1 interval (25); therefore, these regions warrant further investigation.

The correlation with GWAS-identified loci also is presented in Table 2. The following are five of the stratified loci that had (corrected) suggestive LOD scores in regions that gave significant results in the T1DGC GWAS (6): *PRKD2*, *CTRB2*, *CENPW* (formerly *C6orf173*), *INS*, and *C19orf19*. Association of these SNPs listed in Table 2 could be followed-up in future studies of subsets of type 1 diabetes cases.

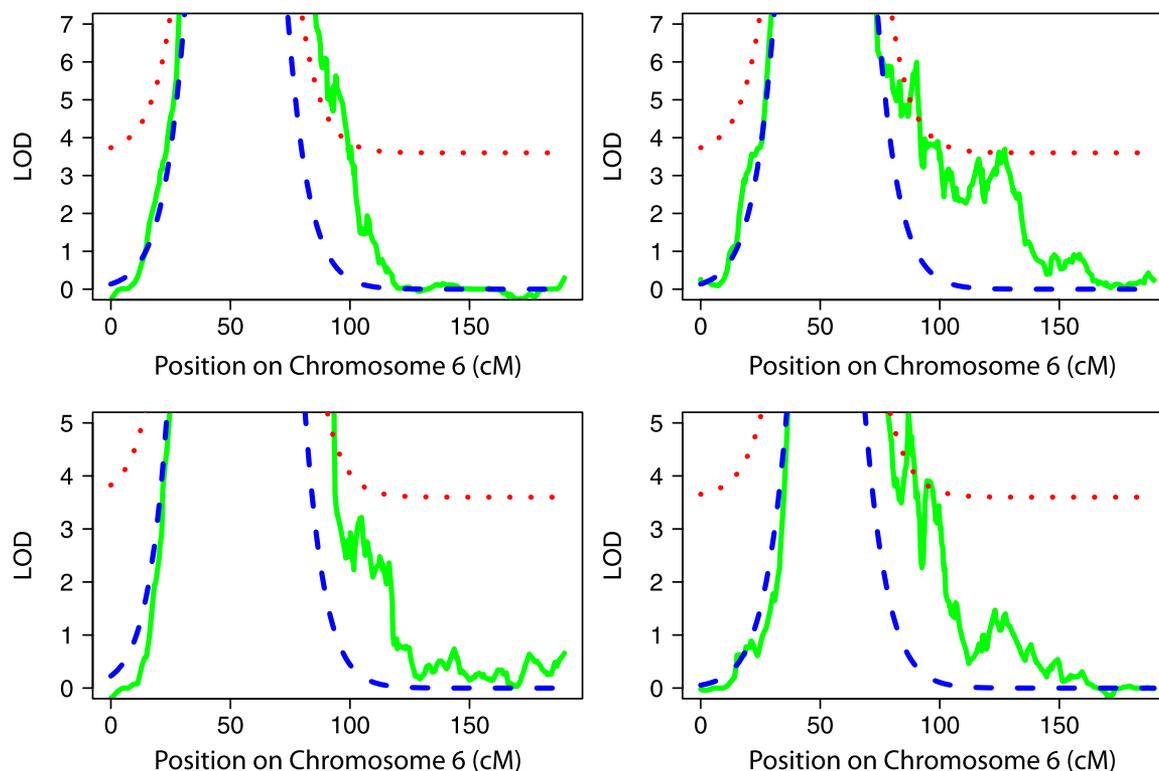


FIG. 5. Chromosome 6 analyses, with correction for HLA linkage. ASP families from T1DGC sets 1 and 2 stratified on the basis of sex (*upper panels: left, female ASPs; right, male ASPs*) or *INS* genotype (*lower panel: left, INS homozygous for the A genotype; right, at least one copy of the protective T allele*). Observed LOD scores are shown in solid green lines. ELOD scores (dashed blue lines) are shown correcting for HLA linkage. The ELOD + 3.6 score (dotted red lines) indicates the threshold for linkage of a type 1 diabetes risk locus independent of HLA.

Testing GWAS SNPs in the stratified datasets. To test whether any of the linkage peaks observed in the stratified sets could be explained by SNPs associated with type 1 diabetes susceptibility in the T1DGC GWAS study (6), families were stratified according to criteria above and tested by the transmission disequilibrium test (TDT). Twenty SNPs were significantly associated in the overall dataset (uncorrected P values $< 10^{-3}$) (Table 3). However, for the linked regions listed in Tables 1 and 2, only regions 6p21, 11p15, 19q13, and 2q33 contained SNPs showing associations with type 1 diabetes by analysis of allele transmissions (Table 3). There were significant differences in transmission for the HLA SNPs between the HLA identical and mismatched ASP sets, confirming a similar earlier analysis of 3,000 HLA SNPs in the T1DGC families (26). There also was an apparent bias in transmission of the *HLA-DQA1* SNP in affected brothers. However, there were no significant differences in the non-HLA SNPs between any of the sets stratified by the other criteria (sex, *INS* genotype, or age of diagnosis).

DISCUSSION

Comparison with previous type 1 diabetes linkage scans. The results from the third and final T1DGC family collection include strong confirmation of linkage to the HLA complex (LOD = 135.7) and suggestive linkage to *INS*. Additional peaks on chromosome 6q also were evident. Three *IDDM* loci have been mapped to chromosome 6q (6,15,16,18). Part of the linkage signal detected on chromosome 6q may reflect these or the 6q GWAS loci, including *TAGAP* SNPs, which showed evidence of association in these families (Table 3). In addition, there was

a suggestive linkage peak on chromosome 8p12, a region not previously reported.

Considering the entire T1DGC family collection of 4,422 ASPs, there was a total of six loci with LOD scores over the suggestive threshold of 2.2. Aside from *HLA* and *INS*, the combined dataset provided the 19q13 locus with an increased score of 3.3. However, a previously reported locus including 2q31 and *CTLA4* had a reduced, though still suggestive (3.11), LOD score in the final dataset. This chromosomal region contains four associated loci, including *CTLA4*; these and possibly other undetected loci could have accumulated effects to generate this linkage signal.

In the combined dataset, an additional locus on 6q22 at ~123 Mb reached nominal significance (unadjusted LOD = 3.47); this may be attributed to the effect of the GWAS SNP near *CENPW* (6) and/or other associated SNPs in the region, as well as *IDDM15*. The *IDDM15* locus was previously mapped with a peak at ~102 Mb on 6q, but its location is difficult to define because of the very large HLA signal, residual linkage to this region, male/female recombination differences, and because this locus may reflect multiple susceptibility genes indicated by the GWAS chromosome 6q SNPs (6).

Comparison with GWAS data. The findings from these linkage studies, summarized in Table 3, contrast with those from the T1DGC GWAS (6) in which over 50 type 1 diabetes risk genes were defined. One reason for these contrasting results could be that familial clustering of type 1 diabetes is predominantly attributed to HLA-linked genes with little influence of non-HLA genes. This is supported by evidence that, although non-HLA type 1 diabetes risk loci identified in GWAS could be replicated in ASP families, the

TABLE 3
Transmission of type 1 diabetes-associated SNP alleles in total and stratified sets of T1DGC families

| Chromosome | SNP | Gene | Method for affected sib-pair family stratification | | | | | | | | | | | | | |
|------------|------------|---------|--|------------------------|------------------------|-----|------------------|--------------------|---------------------|-------------------|-----------------------|---------------------|---------------------|---------------------|--------------------|------------------|
| | | | None (n = 4,663) | HLA IBD (n = 1,558) | Non-IBD (n = 2,026) | HLA | DR3 (n = 635) | DR4 (n = 1,088) | DR34 (n = 1,246) | DRxx (n = 283) | INS_AA (n = 1,647) | INS_Tx (n = 641) | Female (n = 968) | Male (n = 1,108) | Young (n = 429) | Old (n = 269) |
| 1p13 | rs2476601 | PTPN22 | 50.3 | 4.4 | 12.1 | | 4.2 | 3.5 | — | — | — | 2.3 | 2.7 | 3.5 | 3.2 | — |
| 2q24 | rs3747517 | IFIH1 | 5.3 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 2q33 | rs231727 | CTLA4 | 11.7 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 6p21 | rs2187668 | DQAI | 202.1 | 84.7 | 8.6 | | 44.2 | 88.2 | — | — | 11.8 | 19.2 | 39.5 | 14 | 5.1 | — |
| 6p21 | rs7454108 | DRB1 | 230.9 | 124.1 | 58.3 | | 104.4 | 97 | 2.2 | — | 31.3 | 51.5 | 50.3 | 19.7 | 7.8 | — |
| 6q25 | rs1738074 | TAGAP | 3.3 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 10p15 | rs12722495 | IL2RA | 10.6 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 10p15 | rs947474 | PRKCC | 3.2 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 10q23 | rs10509540 | RNL5 | 14.1 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 11p15 | rs689 | INS | 77.2 | 6.5 | 15.3 | | 4 | 7.3 | 4 | — | NA | 5.1 | — | 5.2 | — | — |
| 12p13 | rs4763879 | CD69 | 3.8 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 12q13 | rs2292239 | IKZF4 | 9.9 | — | 2.3 | | — | — | — | — | — | — | — | — | — | — |
| 12q24 | rs3184504 | SH2B3 | 14.5 | — | — | | 2.5 | — | — | — | — | — | — | — | — | — |
| 15q25 | rs3825932 | CTSH | 3.9 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 16p13 | rs12708716 | CLECI6A | 8.9 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 17q12 | rs2290400 | GSDML | 4.9 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 18p11 | rs478582 | PTPN2 | 3.9 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 18q22 | rs763361 | CD226 | 5.4 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 19q13 | rs425105 | PRKD2 | 3.7 | — | — | | — | — | — | — | — | — | — | — | — | — |
| 20p13 | rs2281808 | SIRPG | 5.6 | — | — | | — | — | — | — | — | — | — | — | — | — |

ASPs were previously genotyped at 53 SNPs as part of the T1DGC GWAS (6). Transmission of alleles was analyzed in these families by TDT. Results are shown as the negative log(P). Only SNPs that had $-\log P$ values >3 in the total family collection were studied further. Families were stratified according to criteria described in the text and TDT scores calculated for the 20 SNPs shown. In this case, scores were corrected for multiple testing by subtracting 2.38, the log of the number of tests ($n = 240$); only values >2.2 are shown. "Gene" does not necessarily indicate the causal gene but the gene either containing or very near the most disease-associated SNP or a gene of strong functional candidacy in the associated region.

odds ratios estimated from these families had smaller effect sizes (6). Lack of significant evidence for linkage in the regions corresponding to GWAS-detected associated SNPs was expected as a consequence of the small individual contribution of these loci and their often high allele frequency in these reported associations, so that siblings could have a genotype identical by state rather than IBD, thereby not showing linkage and not increasing the observed LOD scores. In addition, the sporadic cases for the GWAS (6) were drawn from only two countries (U.K. and U.S.), whereas the T1DGC families were recruited from many different countries and diverse recruitment networks (Asia-Pacific, Europe, and North America) that differ markedly in environmental influences that may affect penetrance of susceptibility alleles. Thus, the increased genetic and environmental heterogeneity could limit the ability to detect genes that may affect risk in some populations but not others. The non-HLA loci that survived the total analyses may have done so because they had consistent, albeit weak, effects across cohorts, rather than showing significant linkage in individual cohorts.

Genetic heterogeneity and epistasis. Another reason for the scant returns from this study could be disease heterogeneity, which could be increased by the requirement of worldwide recruitment of families to achieve sufficient statistical power. If there were multiple type 1 diabetes subtypes involving different genetic pathways, then the numbers of sib-pairs sharing haplotypes at relevant genes would be swamped by the remaining siblings with Mendelian-sharing ratios. Support for a notion of different genetically determined type 1 diabetes subtypes comes from the variation seen in disease, with some families having an earlier onset, and from studies that suggest that the increase in type 1 diabetes in recent decades is associated with lower risk HLA alleles (24). To address this issue, we compared sib-pairs selected for relevant criteria that have been shown previously to increase linkage evidence for particular loci (16–20). Although seven loci were implicated using this approach, the LOD scores for these could not support a hypothesis that they played a major role even in the type 1 diabetes subtype studied.

We conclude that there are no major genetically discernible type 1 diabetes subtypes defined by interaction of HLA and non-HLA genes. Furthermore, it appears that there are no major epistatic interactions in type 1 diabetes, at least based on interactions of individual non-HLA genes with HLA, *INS*, and sex-specific factors. This is consistent with a recent analysis of HLA and *PTPN22* alleles (27) and suggests that epistasis is unlikely to be a major contributor to the familial clustering in type 1 diabetes.

Implications. This largest and most robustly powered linkage study of type 1 diabetes found no loci with genome-wide significant linkage beyond those in the major histocompatibility complex (*HLA*) and *INS*. The absence of additional major loci also suggests that the remaining genetic susceptibility (missing heritability) may well reside in many other non-HLA genes, each contributing such a low risk that it escapes detection by linkage even in as large a sample collection as in this study. Importantly, it seems unlikely that the remaining familial clustering is a result of major loci with complex variants not tagged by current SNP maps or by such loci with multiple rare moderate-effect variants.

As previously identified by the T1DGC GWAS, numerous genes contribute to type 1 diabetes risk yet these have relatively small individual effects. It should be emphasized, however, that the size of the effect does not correlate with

the potential importance of the gene in a critical pathway or therapeutic target. By comparison with the NOD mouse model, an individual locus may have a small effect but could nevertheless have a major outcome on disease, as shown in many studies of congenic strains that differ only at a single susceptibility locus (e.g., [28,29]).

The remaining genetic variation contributing to type 1 diabetes risk may be determined by numerous other mechanisms, such as shared intrafamilial environmental factors, which require different research approaches. Another mechanism could involve structural (e.g., copy number) variants not tagged by the current SNP arrays. In type 1 diabetes, one of the most important risk factors is the *INS* variable number of tandem repeats, a copy-number variant. Other copy-number or structural variants may contribute to risk; thus, a genome-wide assessment for these effects is necessary, though our linkage results, and the recent study of copy-number variants in 2,000 type 1 diabetes cases (30), indicate that a structural variant with a common susceptibility allele with a large effect is unlikely to exist in the populations studied. Another mechanism is the contribution of low-frequency sequence susceptibility variants that may cluster in genes but not occur at the same position in the gene, as observed for *IFIH1* (31). However, our linkage results suggest that genes with multiple rare variants with high disease penetrance are not a major contributor to the inheritance of type 1 diabetes. The T1DGC families were recruited for detection of major gene linkage effects and are among the most appropriate resources for future studies addressing the effects of structural variants and rare variants on type 1 diabetes risk.

ACKNOWLEDGMENTS

G.M. and M.M. are supported by program Grant 516700 from the National Health and Medical Research Council of Australia and by the Diabetes Research Foundation (Western Australia). This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Allergy and Infectious Diseases, the National Human Genome Research Institute, the National Institute of Child Health and Human Development, and the Juvenile Diabetes Research Foundation International and is supported by U01-DK-062418. J.A.T. is supported by the Juvenile Diabetes Research Foundation, the National Institute for Health Research, and the Wellcome Trust. No potential conflicts of interest relevant to this article were reported.

G.M. researched data, contributed to discussion, and wrote the manuscript. M.M. wrote the specialist software, researched data, and reviewed the manuscript. I.J. analyzed data, reviewed the manuscript, and contributed to discussion. W.-M.C. analyzed data and reviewed the manuscript. B.A., H.A.E., J.E.H., C.J., J.N., C.N., F.P., and J.A.T. reviewed and edited the manuscript and contributed to discussion. S.S.R. analyzed data, reviewed the manuscript, and contributed to discussion.

REFERENCES

1. Rich SS. Mapping genes in diabetes. Genetic epidemiological perspective. *Diabetes* 1990;39:1315–1319
2. Rich SS, Akolkar B, Concannon P, et al. Current status and the future for the genetics of type I diabetes. *Genes Immun* 2009;10(Suppl. 1):S128–S131

3. Rich SS, Concannon P, Erlich H, et al. The Type 1 Diabetes Genetics Consortium. *Ann N Y Acad Sci* 2006;1079:1–8
4. Concannon P, Erlich HA, Julier C, et al. Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes* 2005;54:2995–3001
5. Concannon P, Chen WM, Julier C, et al. Genome-wide scan for linkage to type 1 diabetes in 2,496 multiplex families from the Type 1 Diabetes Genetics Consortium. *Diabetes* 2009;58:1018–1022
6. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009;41:703–707
7. Rich SS, Akolkar B, Concannon P, et al. Results of the MHC fine mapping workshop. *Diabetes Obes Metab* 2009;11(Suppl. 1):108–109
8. Julier C, Akolkar B, Concannon P, Morahan G, Nierras C, Pugliese A. The Type 1 Diabetes Genetics Consortium 'rapid response' family-based candidate gene study: strategy, genes selection, and main outcome. *Genes Immun* 2009;10(Suppl. 1):S121–S127
9. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–753
10. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;63:259–266
11. Sun L, Wilder K, McPeck MS. Enhanced pedigree error detection. *Hum Hered* 2002;54:99–110
12. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin: rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97–101
13. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575
14. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997;61:1179–1188
15. Delépine M, Pociot F, Habita C, et al. Evidence of a non-MHC susceptibility locus in type 1 diabetes linked to HLA on chromosome 6. *Am J Hum Genet* 1997;60:174–187
16. Davies JL, Kawaguchi Y, Bennett ST, et al. A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994;371:130–136
17. Field LL, Tobias R, Thomson G, Plon S. Susceptibility to insulin-dependent diabetes mellitus maps to a locus (IDDM11) on human chromosome 14q24.3-q31. *Genomics* 1996;33:1–8
18. Luo DF, Bui MM, Muir A, Maclaren NK, Thomson G, She JX. Affected-sib-pair mapping of a novel susceptibility gene to insulin-dependent diabetes mellitus (IDDM8) on chromosome 6q25-q27. *Am J Hum Genet* 1995;57:911–919
19. Morahan G, Huang D, Tait BD, Colman PG, Harrison LC. Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. *Science* 1996;272:1811–1813
20. Morahan G, Huang D, Ymer SI, et al. Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. *Nat Genet* 2001;27:218–221
21. Bennett ST, Todd JA. Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu Rev Genet* 1996;30:343–370
22. Pugliese A, Miceli D. The insulin gene in diabetes. *Diabetes Metab Res Rev* 2002;18:13–25
23. Lucassen AM, Julier C, Beressi JP, et al. Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nat Genet* 1993;4:305–310
24. Fourlanos S, Varney MD, Tait BD, et al. The rising incidence of type 1 diabetes is accounted for by cases with lower-risk human leukocyte antigen genotypes. *Diabetes Care* 2008;31:1546–1549
25. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* 1999;65:876–884
26. Morahan G, Mehta M, McKinnon E, James I. Effect of linkage status of affected sib-pairs on the search for novel type 1 diabetes susceptibility genes in the HLA complex. *Diabetes Obes Metab* 2009;11(Suppl. 1):67–73
27. Clayton DG. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet* 2009;5:e1000540
28. Leiter EH. Nonobese diabetic mice and the genetics of diabetes susceptibility. *Curr Diab Rep* 2005;5:141–148
29. Ridgway WM, Peterson LB, Todd JA, et al. Gene-gene interactions in the NOD mouse model of type 1 diabetes. *Adv Immunol* 2008;100:151–175
30. Craddock N, Hurles ME, Cardin N, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010;464:713–720
31. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009;324:387–389
32. Risch N. A note on multiple testing procedures in linkage analysis. *Am J Hum Genet* 1991;48:1058–1064