Human Molecular Genetics, 2003, Vol. 12, No. 15 **1865–1873**
DOI: 10.1093/hmg/ddg195

A meta-analysis of four European genome screens (GIFT Consortium) shows evidence for a novel region on chromosome 17p11.2–q22 linked to type 2 diabetes

Florence Demenais[1], Timo Kanninen[2,3], Cecilia M. Lindgren[3], Steven Wiltshire[4], Stéphane Gaget[5], Candice Dandrieux[1], Peter Almgren[3], Marketa Sjögren[3], Andrew Hattersley[6], Christian Dina[5], Tiinamaija Tuomi[7], Mark I. McCarthy[4,8], Philippe Froguel[5] and Leif C. Groop[3,*]

[1]EMI 0006, INSERM-Université d'Evry, Evry, France, [2]Biocomputing Platforms Ltd Oy, Helsinki, Finland, [3]Department of Endocrinology, Wallenberg Laboratory, Malmö University Hospital, SE-205 02 Malmö, Sweden, [4]Wellcome Trust Centre for Human Genetics, Oxford, UK, [5]Institute of Biology-CNRS 8090, Institut Pasteur, Lille, France, [6]Department of Diabetes and Vascular Medicine, School of Postgraduate Medicine and Health Sciences, University of Exeter, Exeter, UK, [7]Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland and [8]Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK

Received March 11, 2003; Revised May 7, 2003; Accepted June 2, 2003

Positional cloning is expected to identify novel susceptibility genes underlying complex traits, but replication of genome-wide linkage scan findings has proven erratic. To improve our ability to detect and prioritize chromosomal regions containing type 2 diabetes susceptibility genes, the GIFT consortium has implemented a meta-analysis of four scans conducted in European samples. These included the Botnia I and Botnia II scans, with respectively 58 and 353 pedigrees from Finland and Sweden, the Warren 2 scan performed in 573 multiplex sibships from the UK, and a scan of 143 families from France. The meta-analysis was implemented using the genome-search analysis method (GSMA), an exploratory data analysis technique which is robust across study designs. The analysis provided evidence for linkage of type 2 diabetes to six regions, with the strongest evidence on chromosome 17p11.2–q22 ($P = 0.0016$), followed by 2p22.1–p13.2 ($P = 0.027$), 1p13.1–q22 ($P = 0.028$), 12q21.1–q24.12 ($P = 0.029$), 6q21–q24.1 ($P = 0.033$) and 16p12.3–q11.2 ($P = 0.033$). Linkage analysis of the pooled raw genotype data generated maximum LOD scores in the same regions as identified by GSMA. Altogether, our results have indicated that GSMA is a valuable tool to identify chromosomal regions of interest and that accumulating evidence for linkage from small peaks detected across several samples may be more important than getting a high peak in a single sample. This meta-analysis has led to identification of a novel region on chromosome 17 linked to type 2 diabetes; this region has not been highlighted in any published scan to date but on the basis of these data justifies further exploration.

INTRODUCTION

Type 2 diabetes is a complex and heterogeneous disease characterized by insulin resistance and pancreatic β-cell dysfunction. Almost one in 10 of the world population already has this condition, or can be expected to develop it during their lifetime, with prevalence rates forecast to double within the next 15 years. Type 2 diabetes is a potent risk factor for cardiovascular disease, and, uncorrected, the chronic hyperglycaemia leads to various complications, including retinopathy, neuropathy and nephropathy. All the available evidence indicates that type 2 diabetes is a multifactorial disease with individual susceptibility determined by a rich mix of genetic and environmental factors.

*To whom correspondence should be addressed. Tel: +46 40332303; Fax: +46 40337023; Email: leif.groop@endo.mas.lu.se

Considerable effort has been made worldwide during the last decade to identify the genetic determinants of type 2 diabetes. Studies of rare monogenic forms of type 2 diabetes have led to the identification of several genes, including the genes that cause MODY (maturity-onset diabetes of the young), as well as mutations in mitochondrial DNA (1). However, these genes appear to play a minor role in the common form of type 2 diabetes.

In recent years, candidate gene studies have had increasing success in identifying variants that influence the risk of type 2 diabetes, with substantial evidence supporting a role for polymorphisms in the genes for peroxisome proliferator-activated receptor-gamma, insulin receptor substrate 1, glycogen synthase, insulin, the potassium inwardly rectifying channel 11, and insulin promoter factor 1, amongst others (1,2).

The complementary, positional cloning approach is expected to reveal additional novel susceptibility loci and genome-wide scans for linkage have been conducted in a wide range of populations, including African-Americans (3), Ashkenazim (4), British (5), Chinese (6), Finns and Scandinavians (7,8–11), French (12), Mexican Americans (3,13,14), Pima Indians (15) and US Europeans (3,16). These scans have reported a number of chromosomal regions that may harbour genes involved in type 2 diabetes, with the most promising, replicated findings on chromosomes 1q21–q24 (5,12,15,16), 2q37 (13), 12q24 (3,7,10,11) and chromosome 20 (8,9). Linkage to the latter region was also reported by candidate region studies (17–19) and in a meta-analysis of chromosome 20 data from many of the scans described above (20). There has been, to date, one example whereby the positional cloning approach has led to identification of a diabetes-susceptibility gene: evidence for linkage to type 2 diabetes on chromosome 2q37 in Mexican Americans led to the characterization of susceptibility variants within the *CAPN10* gene (13,21). This raises the possibility that other type 2 diabetes genes may be identified in the near future using the same approach.

The variability in linkage findings reported between populations may reflect not only the intrinsic complexity of the disease architecture, but also differences in family ascertainment, phenotype definition, markers genotyped and analytical methods (22). One approach that may assist in the identification of the regions which are most likely to contain susceptibility genes for type 2 diabetes would be to conduct a meta-analysis to assess the evidence for linkage across studies. Given the likely ethnic heterogeneity of type 2 diabetes, it is probably sensible to initiate such meta-analysis in populations of similar ethnic origin. We have formed a European consortium (the Genomics Integrated Force for type 2 Diabetes, GIFT) to combine our efforts in identifying genes involved in type 2 diabetes (more detailed information on GIFT can be found at www.gift.med.ic.ac.uk). Constituent groups within GIFT have previously reported their genome-wide scans conducted in Swedish–Finnish (Botnia I and Botnia II scans) (7,10,11), UK (5) and French family samples (12). These studies have provided evidence for linkage at a number of different chromosomal regions.

Various methods of meta-analysis have been described (23,24): each is designed to allow a rapid genome-wide assessment of regions of interest as a prelude to embarking on the more demanding task of a pooled analysis of raw genotype data.

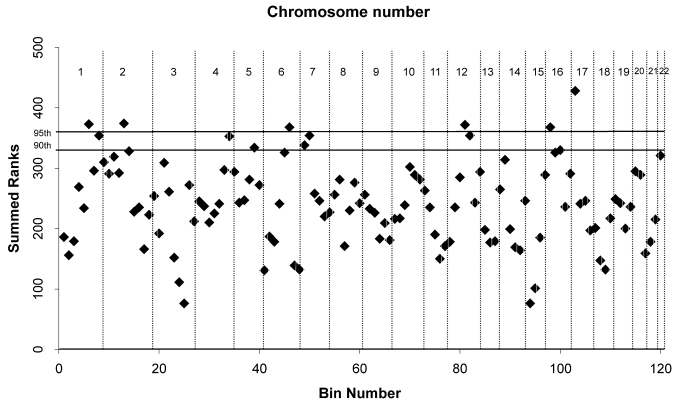

Figure 1. GSMA results from four genome screens for type 2 diabetes (Botnia I, Botnia II, UK and French). The horizontal axis indicates the bin number and the vertical axis shows the summed ranks across the four scans. The two solid lines represent the 90th (summed rank = 332) and 95th (summed rank = 356) percentiles of the summed rank distribution.

Recently, the genome search meta-analysis (GSMA) method has been proposed as an exploratory data analysis technique which is robust across study designs (23). To better characterize the chromosomal regions showing the strongest evidence for linkage across European data sets, we have applied GSMA to the four genome screens of the GIFT consortium. This was then followed by linkage analysis of the pooled raw data for those chromosomes showing significant results by GSMA.

RESULTS

Genome search meta-analysis

Chromosomal regions that show consistently increased linkage statistics (LOD scores or NPL scores) (25) in the four European genome scans were identified by the non-parametric ranking GSMA method. For each scan, genetic regions (bins of equal length defined on each chromosome) were ranked according to the maximum linkage statistic achieved in a given region. The summed ranks across scans were then compared with their probability distribution derived under the null hypothesis that ranks are randomly assigned. The main characteristics of the four genome-wide screens to which GSMA was applied are shown in Table 1. GSMA results are presented in Figure 1. The summed ranks (vertical axis) are plotted against the bin location by chromosome (horizontal axis) with a single point plotted for the summed rank for each bin. The 90th and 95th percentiles for the distribution of summed ranks are also shown in Figure 1. A total of six bins lie above the 95% threshold. The strongest evidence for linkage from the GSMA occurs on chromosome 17p11.2–q22 with a summed rank of 428 ($P = 0.0016$). Other significant regions are on chromosomes 2p22.1–p13.2 ($P = 0.027$), 1p13.1–q22 ($P = 0.028$), 12q21.1–q24.12 ($P = 0.029$), 6q21–q24.1 ($P = 0.033$) and 16p12.3–qll.2 ($P = 0.033$). Results from the GSMA and the original genome screens are presented in Table 2. Note that the *P*-values shown in this table for the original scans corresponded to the

Table 1. Descriptive characteristics of type 2 diabetes genome-wide screens

Characteristics of each scan	Botnia I	Botnia II	United Kingdom	French
Population	Swedish–Finnish	Swedish–Finnish	UK subjects	French
Sample size	58 pedigrees	353 pedigrees	573 multiplex sibships	143 families (148 nuclear)
Total no. of subjects[a]	440 (7.6)[c]	1488 (4.2)	1386 (2.4)	633 (4.4)
Total no. affected[b]	229 (3.9)[d]	959 (2.7)	1223 (2.1)	432 (3.0)
Age at diagnosis of type 2 diabetes (years)	57.5 ± 0.8[e]	52.1 ± 11.5	55.6 ± 8.6	49.5 ± 10.7
Number of markers[f]	387	468	418	401
Programs used in the published papers	Genehunter 2.0	Genehunter Plus	Allegro 1.1	MLBGH 1.0 Mapmaker-Sibs 2.0 Genehunter 2.0
References	(11)	(10)	(5)	(12)

[a]The total number of subjects includes only those with a known affection status and corresponds to siblings in the UK data set.
[b]The total number of affected subjects are the affected sibs in the UK data set.
[c]Average number of subjects per family in parentheses.
[d]Average number of affected subjects per family in parentheses.
[e]Mean age ± standard deviation.
[f]Number of markers used by each initial scan.

maximum NPL scores (25) observed in a given bin, as provided by each group to the GIFT database for GSMA. However, more significant levels may have been published in the same regions by using other linkage statistics (e.g. LOD scores from GENEHUNTER PLUS) (26). Three of the significant results from GSMA ($P \leq 0.05$) were significant at this level (or lower) in the Botnia II (6q21–q24, 12q21–q24, 16p12–q11) and French scans (1p13–q22, 2p22–p13, 17p11–q22), while two were significant in the Botnia I scan (16p12–q11, 17p11–q22) and none in the UK scan (three bins on 1p13–q22, 2p22–p13 and 6q21–q24 showing marginal results with $0.05 < P \leq 0.10$). Contributions to the most significant 17p11–q22 region come mainly from the Botnia I and French scans with marginal contribution from Botnia II. Evidence for other regions comes from different combinations of any two or three scans. Moreover, two bins, one 30 cM telomeric to the significant bin on 1q (1q25–q32) and another one adjacent to the significant bin on 12q (12q24.13–q24.32), have summed ranks that are marginally significant ($0.05 < P < 0.10$) with contributions from the same samples, French and UK for lq, Botnia I and Botnia II for 12q. In addition, three other regions lie between the 90th and 95th percentiles (Fig. 1): 4q mainly due to Botnia I, 5q with contributions from Botnia I and UK screens and 7p with contributions from Botnia II, French and marginally Botnia I scans. However, there were other potential linkages, reported by the published analyses of the individual raw data, that show no evidence for linkage in GSMA: mainly, 3p, 4q and 9p in Botnia I; 4p, 5q and 18p in Botnia II; 3q, 10q, 20p and 20q in the French scan; and 5q, 8p, 8q and 10q in the UK scan. The non-replication of significant results across genome screens may be due to false-positive results, lack of power, stochastic variation in the strength of the linkage signal detected, or may represent population-specific susceptibility genes.

Linkage analysis of the raw data

Linkage analyses of the raw data (Fig. 2) were carried out, using the Genehunter Plus program (26), for the six chromosomes which displayed significant results, as obtained from the previous GSMA. Table 3 shows the multipoint maximum LOD scores (with corresponding pointwise P values) (26) obtained in the pooled raw data and the highest scores observed in each data set in the same region, a region having the same length as a bin. All maximum LOD scores in the pooled data were reached in the same bin as found significant by GSMA, except for 12q where it was reached in the adjacent bin (12q24.13–q24.32) lying between the 90th and 95th percentiles when using GSMA. The highest maximum LOD scores were observed on chromosomes 2p22–p13 (LOD score = 1.82, $P = 0.002$) and 17p11–q22 (LOD score = 1.54; $P = 0.004$), followed by 6q21–q24 (LOD score = 1.32; $P = 0.007$), 12q21–q24 (LOD score = 0.92; $P = 0.020$ in the bin significant with GSMA and LOD score = 1.23; $P = 0.009$ in the adjacent bin), 16p12–q11 (LOD score = 1.09; $P = 0.013$) and 1p13–q22 (LOD score = 0.97; $P = 0.017$). As seen from Table 3, the respective contributions of the data sets to these results are similar to those presented in Table 2, but with lower P-values, due to the use of the more powerful allele sharing LOD score statistic (Table 3) as compared with the NPL score (Table 2).

DISCUSSION

One of the main difficulties in the genetic analysis of multifactorial traits has been the frequent failure to observe signal replication across studies. Replication is regarded as an essential step in the identification of regions meriting further exploration and positional cloning endeavours. Problems with replications arise from false-positive results, from inadequate power and from the genetic, clinical, ethnic and experimental heterogeneity between data sets. Meta-analysis cannot overcome totally the problem of genetic heterogeneity but can identify regions that may contain disease genes in a subset of the pooled studies. It may not only confirm evidence for regions highlighted in at least one scan but also identify novel regions where the genetic effect is too small to be detected in a single study.

Table 2. Results obtained from GSMA (*P* values for the summed ranks) and individual genome screens (*P*-values for NPL scores)

Chromosomal region[a]	Bin (cM)[b]	GSMA	Botnia I	Botnia II	French	UK
17p11.2–q22	29–58	0.0016	0.018	≤0.10	0.04	NS
2p22.1–p13.2	58–87	0.027	≤0.10	NS[c]	0.007	≤0.10
1p13.1–q22	145–174	0.028	NS	NS	0.02	≤0.10
12q21.1–q24.12	87–116	0.029	≤0.10	0.04	NS	NS
6q21–q24.1	116–145	0.033	NS	0.04	NS	≤0.10
16p12.3–q11.2	29–58	0.033	0.03	0.03	NS	NS

[a]The cytogenetic region corresponding to the position of each bin on the Marshfield map was obtained from the National Center for Biotechnology Information (NCBI) web site (www.ncbi.nlm.nih.gov).
[b]The position of each significant bin (29 cM length) is indicated in centimorgans (cM) with respect to the Marshfield map.
[c]NS = not significant.

The present GSMA identified a total of six regions, the most significant result being on chromosomes 17p11.2–q22 followed by 2p22.1–p13.2, 1p13.1–q22, 12q21.1–q24.12, 6q21–q24.1 and 16p12.3–q11.2. Most of these regions did not tend to be the most significant linkages in any single scan, but instead provided modest evidence for linkage in at least two analyses. The contribution of the four scans to the significant results of the GSMA appears highest for Botnia II and French followed by Botnia I and then UK sets. The contribution of any given scan may depend upon many factors including the number of families, the family structure (pedigrees or nuclear families), the number of affected and genotyped subjects, their ages at diagnosis and other clinical characteristics, the mode of family selection and the number of informative markers. Although a different weight could be assigned to each scan in GSMA to take into account the various factors described above, the optimal weighting strategy is unknown and arbitrary choices might be expected to influence the outcome of the analysis. The GSMA results may also depend upon the way the maps used in the different analyses are adjusted for and upon the way the bins are defined on the chromosomal maps. However, use of different ways to adjust the GENEHUNTER maps of the four scans (e.g. to the French/UK map since these two scans used similar marker sets) did not lead to any substantial change in the results obtained. For example, the bins on chromosomes 17 and 2 provided the highest summed ranks in all analyses.

Linkage analysis of the pooled raw data generated maximum LOD scores within the same bins as those found significant with GSMA, except for 12q where the maximum LOD was reached in the adjacent bin. This latter result is in agreement with the wide confidence intervals classically observed around a peak LOD score and the difficulty in defining precisely the limits of a linkage region.

It is important, however, to realize the intrinsic differences between the GSMA methodology and the computation of LOD scores based on the pooled raw data. Most notably, the *P*-values derived from the maximum LOD scores cannot be directly compared with those obtained from GSMA since they are based on different test statistics. The *P*-values from GSMA are derived from the distribution of the summed rank *R*, with a highly significant result ($P = 0.0016$) obtained for the 17p11–q22 region. The *P*-values obtained from the analysis of the pooled raw data need to be interpreted in the light of their genome-wide significance, and to take account of the multiple testing implicit in such a genome scan (27). Note that the LOD scores presented here were computed only for the six chromosomes harbouring significant GSMA bins. Moreover, these LOD scores rely on the available marker information, especially for a late onset disease such as type 2 diabetes where most parents are not genotyped (28). Although a multipoint analysis of the pooled raw data was conducted, it could not use the full multipoint information from all markers since the vast majority of markers were not common to the four data sets. Thus, both non-overlap of marker locations and absence of parental genotype data can reduce the power of the pooled-data linkage analysis. In addition, it is important to recognize that the non-parametric ranking procedure which makes GSMA robust to study design also has its limitations. It tends to amplify the effects of modest changes in linkage statistic scores in the middle or lower part of the LOD score range, and to reduce the impact of changes in the linkage statistic score in the higher part of the range. This can potentially lead to some discordance between peak localization by the two methods, particularly where linkage peaks from different component studies fall into adjacent bins (as on chromosome 1q in the UK data). Nevertheless, the strong correspondence between the initial GSMA results and the subsequent examination of the positive regions based on the pooled raw data indicates that the GSMA-based approach is a valuable tool for the identification of chromosomal regions of interest since it is robust across study designs and does not rely on a common set of markers across studies.

The most significant result obtained from GSMA concerns the 17p11–q22 region which was not among the most highlighted results in any of the individual component scans. It was reported at a nominally-significant *P*-value in Botnia I (11) but was not mentioned in the published French scan (12) and the Botnia II linkage peak on 17q was 40 cM telomeric to the upper bound of the significant GSMA bin (10). However, this region has been recently reported to be linked to plasma leptin levels in a genome-wide screen of 507 Caucasian nuclear families (29). This QTL was suggested to act epistatically with a QTL on chromosome 3q27 linked to six traits related to the metabolic syndrome. Linkage of chromosome 17 to total cholesterol and high density lipoprotein-cholesterol (HDL-C) was also found in a genome scan of 232 multigenerational pedigrees randomly selected from the population (30). This region contains several potential candidate genes including the nuclear receptor corepressor 1(NCOR1), MAPKK3, MAPKK4, MAPK7, SREBP1, pancreatic polypeptide 2, very long chain

Table 3. Maximum LOD scores (*P*-values) obtained in the pooled raw data using Genehunter Plus and highest scores observed in each data set within the same bin for the chromosomes displaying significant results with the previous GSMA

Chromosomal region[a]	Bin (cM)[b]	Pooled data	Botnia I	Botnia II	French	UK
17p11.2–q22	29–58	1.54 (0.004) (*D17S921*; 36 cM)[d]	1.05 (0.014)	0.51 (NS)	0.83 (0.025)	0.35 (NS)
2p22.1–p13.2	58–87	1.82 (0.002) (*D2S391*; 70 cM)	0.55 (0.055)	0.05 (NS)	1.63 (0.003)	0.76 (0.03)
1p13.1–q22	145–174	0.97 (0.017) (*D1S1679*; 171 cM)	0.04 (NS)[e]	0.11 (NS)	1.00 (0.015)	0.59 (0.05)
12q21.1–q24.12	87–116	0.92 (0.020) (*D12S84*; 116 cM)	0.77 (0.03)	1.05 (0.014)	0.06 (NS)	0.75 (0.03)
12q24.12–q24.32	116–145[c]	1.23 (0.009) (*D12S342*; 145 cM)	1.21 (0.009)	1.17 (0.010)	0.06 (NS)	0.08 (NS)
6q21–q24.1	116–145	1.32 (0.007) (*D6S262*; 130 cM)	0.43 (NS)	0.80 (0.028)	0.05 (NS)	0.98 (0.017)
16p12.3–q11.2	29–58	1.09 (0.013) (*D16S412*; 43 cM)	0.78 (0.029)	1.41 (0.005)	0.09 (NS)	0.22 (NS)

[a]The cytogenetic region corresponding to the position of each bin on the Marshfield map was obtained from the NCBI web site.
[b]The bins (29 cM length) are those which were significant with GSMA and their position is indicated with respect to the Marshfield map.
[c]For 12q, two bins are shown: the one significant with GSMA and the adjacent one lying between the 90th and 95th percentiles corresponding to the maximum LOD score obtained in the pooled raw data.
[d]Marker and position on Marshfield map (in cM from pter) corresponding to the maximum LOD score reached in the pooled raw data.
[e]NS = not significant.

acyl-CoA dehydrogenase (VLCAD), carboxypeptidase D, syntaxin 8, TCF2 (HNF1β), PPARBP, STAT5B etc. Given the role of variants in the PPARγ gene and its coactivator PGC-1 in type 2 diabetes, NCOR1 is an interesting candidate which has been shown to interact with all three forms of PPAR receptors. It spans 200 kb at 17p12 and contains 48 exons. Work is in progress to explore the role of variants in these genes for the linkage to chromosome 17.

The 2p22–p13 region, which scored mainly in the French screen (12) with marginal contributions from the UK and Botnia I screens, was reported to be linked to obesity related-phenotypes, including leptin levels, fat mass or body mass index in different populations (31–33).

The 1p13–q22 region, with evidence mainly from the French scan and from the UK scan, has been reported by a number of other scans in subjects of Northern European origin (16,34,35) and Pima Indians (15). Note that the peak LOD score in the UK scan (LOD = 1.50; $P = 0.004$) was rather obtained in the adjacent bin (1q23–q25) which was not significant with GSMA since the Botnia scans had much lower ranking scores in this bin (see above). The definition of bin width and boundaries used in GSMA can be crucial to the ranking of scores in situations such as this where different component scans peak in neighbouring regions. This issue was discussed by Wise *et al.* (23), who recommended a bin width of about 30 cM, being wide enough to limit the correlation between adjacent bins and not too wide to avoid including distinct peak LOD scores from different studies within the same bin.

The chromosome 12q21–q24 region, detected by the present analysis, is adjacent to the nearby marginally significant bin (12q24.12–q24.32) that includes the *MODY3* locus (36) and the tightly linked *NIDDM2* locus showing linkage to type 2 diabetes in a subset of Botnia I pedigrees with low insulin levels (7). This region was also implied in a recent screen of four American populations as part of the GENNID study (3), where it was found to interact with chromosome 5 in the phase 1 white samples but was not replicated in phase 2 samples. Linkage to this region was also reported in type 2 diabetic patients with nephropathy (17) and in one large Australian family with late-onset type 2 diabetes (37).

The 6q21–q24 region, scoring positively in the Botnia II and UK scans, has not been highlighted by any published scan of type 2 diabetes to date, but was reported to be linked to fasting insulin concentrations and insulin-resistant phenotypes with a pleiotropic effects on obesity-related phenotypes in a genome scan of non-diabetic Mexican Americans (38). This region contains the gene responsible for transient neonatal diabetes mellitus (TNDM), a rare condition that resolves before one year of age but predisposes to type 2 diabetes later in life (39).

Finally, the chromosome 16p12–q11 region scored mainly in the two Botnia scans. To our knowledge, significant linkage to this region has not been reported in previous scans of type 2 diabetes. The nearby 16q21–q22 region includes the *BBS2* gene responsible for the Bardet–Biedl syndrome, an autosomal-recessive disorder characterized by obesity, retinitis pigmentosa and congenital abnormalities (40) and the *RRAD* (Ras-Related Associated with Diabetes) gene found associated to type 2 diabetes in white Americans (41). This association has not been confirmed in Finns (42).

As is well known and recently discussed by Terwilliger and Goring (43) replication of genome scan results is often difficult to achieve. This paper shows that accumulating evidence for linkage from small peaks from many samples may provide a complement to the more dramatic finding of a major, but unreplicated, peak in a single sample. This approach can also lead to the detection of novel regions, as the 17p11–q22 region found here.

In conclusion, our meta-analysis of four European genome wide scans has provided further insights into the distribution of the evidence for linkage to type 2 diabetes in this population.

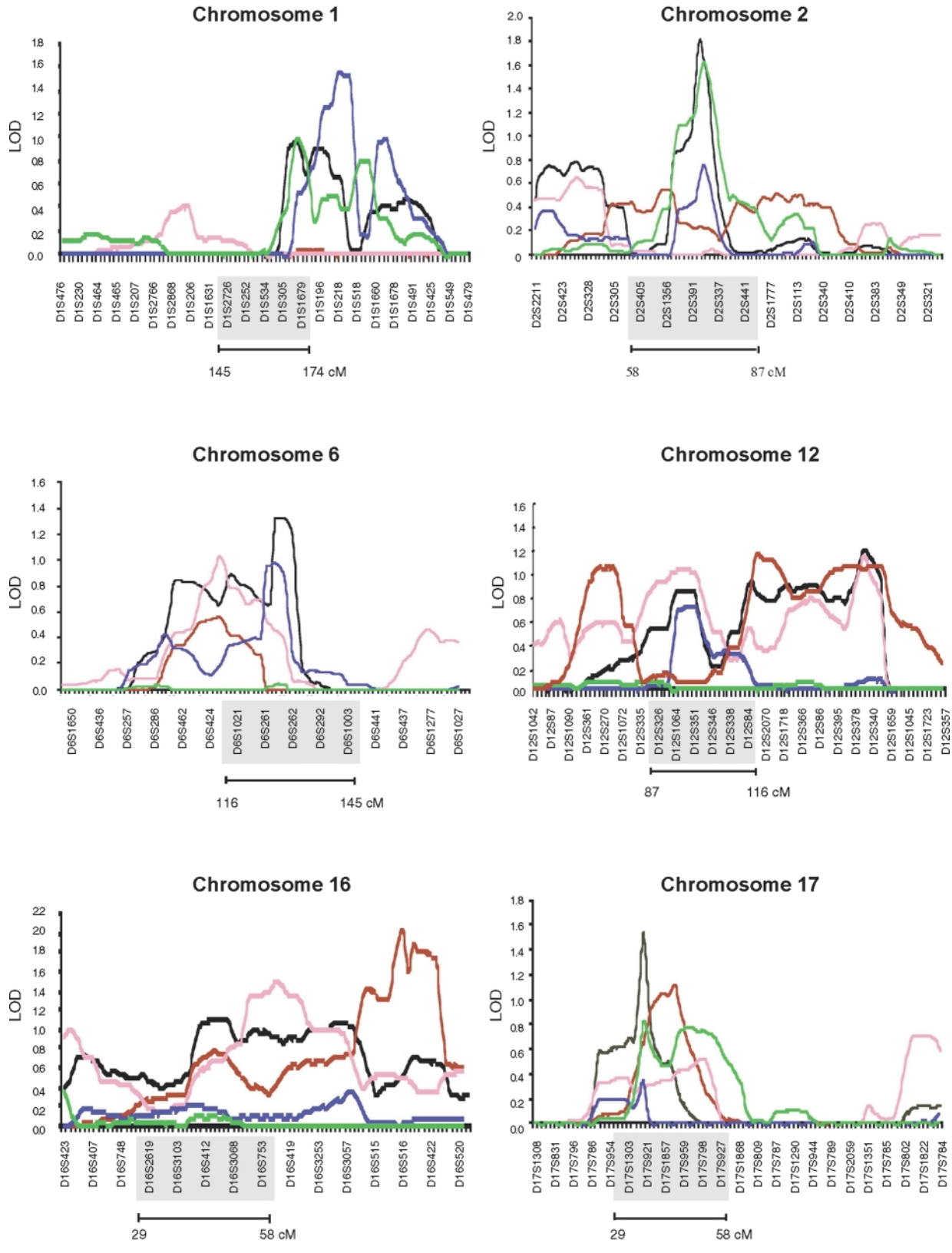

Figure 2. Linkage analysis of the raw data using the Genehunter Plus program. In each graph, the left vertical axis indicates the LOD scores and the horizontal axis shows the microsatellite markers; not all genotyped markers are represented. The LOD scores are represented by a black line for the pooled raw data, a brown line for the BOTNIA I set, a pink line for the BOTNIA II set, a blue line for the UK data and a green line for the French data. On each graph, the shaded area corresponds to the significant GSMA bin with the corresponding location on Marshfield map (lower and upper bounds below the marker names).

This analysis has delineated several new chromosomal regions showing potential linkage to type 2 diabetes and has not confirmed a few other regions reported by the original genome screens. These GSMA results will guide further linkage disequilibrium mapping and positional cloning efforts to identify the susceptibility genes responsible for type 2 diabetes.

MATERIALS AND METHODS

Family samples

The four European data sets included two Swedish–Finnish family samples (Botnia I and Botnia II), one French sample and one UK sample. All these studies had been approved by their local ethical committee and consent forms were obtained from all participating subjects.

The Botnia I set included 440 individuals (229 affected; average family size of 7.6) from 58 multiplex pedigrees (at least two individuals with type 2 diabetes) from western Finland (11). Families with type 1 diabetes mellitus or MODY were excluded. Diabetes was diagnosed on the basis of World Health Organization guidelines (44): either (a) a previous diagnosis of type 2 diabetes mellitus with treatment with oral agents and/or insulin, or (b) either fasting blood (fB)-glucose >6.7 mmol/l or 2 h blood (2hrB) glucose ≥ 8.5 mmol/l. This lower threshold for 2 h glucose was used as it has been shown that these individuals have a very high risk of developing diabetes over the next 5 years (11).

The Botnia II set included 1488 individuals (959 affected) from 353 pedigrees (from different parts of Finland and southern Sweden) with at least two affected members (being mainly two affected siblings) diagnosed with type 2 diabetes before the age of 70 years (10). Families with early-onset diabetes (before 25 years of age) and with a dominant type of inheritance consistent with MODY were excluded as well as families including patients with type 1 diabetes. Type 2 diabetes in this sample was defined using the WHO criteria, as defined above, or based on treatment (oral hypoglycaemic drugs and/or insulin).

The French sample comprised 143 pedigrees (148 nuclear families), ascertained for an affected sibpair and including 432 diabetic patients among a total of 633 subjects of known affection status (12). Diabetes status was based upon clinical records and contemporary glucose results interpreted according to the American Diabetes Association criteria (45). Subjects were considered to have diabetes if receiving oral hypoglycaemic agents or insulin ≥ 1 year after the diagnosis, or when fasting glycemia was ≥ 7 mmol/l or 2 h glycaemia after an oral glucose load was ≥ 11.1 mmol/l. As in the other samples, potential MODY families and families with a mixture of type 1 and type 2 diabetes were discarded.

The UK screen was conducted in 573 nuclear families with at least two affected siblings and including a total of 1223 affected sibs (5). Diagnosis of diabetes was based either on current treatment (oral hypoglycaemics and/or insulin) or, for subjects with diet alone, historical or contemporary laboratory evidence of hyperglycaemia (as defined by WHO). Other forms of diabetes (MODY, mitochondrial diabetes, type 1 diabetes, families containing subjects positive for anti-GAD antibodies) were excluded.

Genome search meta-analysis method

Meta-analysis of these four genome screens was carried out using the GSMA method. This method is a data exploratory tool that combines genome-wide statistics (NPL scores, LOD scores or other statistics) for each chromosomal region across studies, using a non-parametric ranking method, after dividing the genome into a series of bins of equal length. It does not use raw genotypic data. In our analysis, we used 120 bins of about 29 cM each, as recommended by Wise *et al.* (23). For each scan to be included, the most significant result (e.g. maximum NPL score) within each bin was recorded. The bins were then ranked according to this maximum value, the bin containing the most significant result in the study being awarded the highest rank. Once this ranking procedure was completed for all screens, the ranks for each bin were summed across the screens. Under the null hypothesis of no linkage in any chromosomal bin, the ranks are randomly distributed within each study. For m studies and n chromosomal bins, the probability that the ranks X_i $(i = 1, \ldots m)$ from a specific bin sum to R is:

$$P\left(\sum_{i=1}^{m} X_i = R\right) = 0 \text{ for } R < m$$

$$= \frac{1}{n^m} \sum_{k=0}^{d} (-1)^k \binom{R - kn - 1}{m - 1} \times \binom{m}{k} \text{ for } m \leq R \leq nm$$

$$= 0 \text{ for } R > mn$$

where d is the integer part of $(R - m)/n$ and index k varies from 0 to d. This equation allows the distribution of R to be derived and the type 1 error rate to be computed analytically, i.e. the probability that a summed rank of R or greater is obtained within a bin under the null hypothesis. For example, for four studies and 120 bins, as in the present analysis, a summed rank $R \geq 356$ corresponds to a type 1 error of 5%.

Linkage results from the four European scans were entered by each group into the common GIFT database. These results consisted of multipoint NPL scores computed at different positions on each chromosome using the Genehunter program (25). Since markers and maps differed among studies, the different maps used by the Genehunter analyses were adjusted to the reference Marshfield map (this genetic map can be found at http://research.marshfieldclinic.org/genetics). Each group indicated the first and last markers genotyped on each chromosome, which were placed on the Marshfield map. Let x be a given Genehunter map position and $m(x)$ the position adjusted for Marshfield map: $m(x) = m(\text{first}) + fx$, where $m(\text{first})$ is the position of the first marker on Marshfield map and $f = m(\text{last}) - m(\text{first})/y$, $m(\text{last})$ being the position of the last marker on Marshfield map and y being the total map length used by a given Genehunter run. The adjusted maps were divided into bins of equal length (about 29 cM), the number of bins for a given chromosome being determined from the

Marshfield map. Software to adjust map positions and to conduct GSMA were integrated into the GIFT database.

Linkage analysis of the raw data

Linkage analysis of the pooled raw data was carried out for the chromosomes which displayed significant results, as obtained from the previous GSMA. All data, including pedigree information, clinical data and marker genotypes together with allele frequencies and the maps used by each scan were uploaded to the GIFT database. The pooled data files were built by renumbering sequentially the alleles of markers that were common to at least two scans and using allele frequencies specific to each data-set, as allowed by Genehunter (25). For example, if a given marker has i alleles in one scan and k alleles in another scan, alleles are numbered 1 to i for the first data set and $i + 1$ to $i + k$ in the second data set. Allele frequencies were re-estimated from each data set using a maximum likelihood-based method, as implemented in the Vitesse software (46). This takes into account all the available genotypic information within families. We checked that these frequencies were in agreement with those used in the original scans.

To build the maps for the selected chromosomes, inter-marker distances were estimated in each data set using Vitesse and compared with those provided by each group to the GIFT database and those in the published Marshfield map. A common map for the pooled analysis was constructed by using all markers genotyped across the four data sets, the marker order being taken from the published Marshfield map. The inter-marker distances were basically set at the values taken from the Marshfield map with slight modifications to fit the distances estimated from the data sets. We did not compute a map from the pooled raw data themselves since most markers were not shared by all four data sets; only the French and UK scans had used similar marker sets. Model-free multipoint linkage analyses were performed with the Genehunter Plus program (26). We used the S_{all} scoring statistic for calculation of the NPL Z-scores (25) and the exponential model to compute the allele sharing LOD scores (26).

ACKNOWLEDGEMENTS

This work was partly funded by the European Community (Framework Program 5—GIFT Consortium grant QLG2-CT-1999-00546). We are very grateful to Cathryn Lewis who introduced us to the GSMA method and gave us very helpful comments. We wish also to thank the Finnish Supercomputer Centre for biocomputing support and all the co-workers on the different genome-wide scans.

REFERENCES

1. McCarthy, M.I. and Froguel, P. (2002) Genetic approaches to the molecular understanding of type 2 diabetes. *Am. J. Physiol. Endocrinol. Metab.*, **283**, E217–225.
2. Altshuler, D., Hirschhorn, J.N., Klannemark, M., Lindgren, C.M., VohI, M.C., Nemesh J., Lane, C.R., Schaffner, S.F., Bolk, S., Brewer, C. *et al.* (2000) The common PPARgamma Pro l2Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.*, **26**, 76–80.
3. Ehm, M.G., Karnoub, M.C., Sakul, H., Gottschalk, K., Holt, D.C., Weber, J.L., Vaske, D., Briley, D., Briley, L., Kopf, J. *et al.* (2000) Genomewide search for type 2 diabetes susceptibility genes in four American populations. *Am. J. Hum. Genet.*, **66**, 1871–1881.
4. Permutt, M.A., Wasson, J.C., Suarez, B.K., Lin, J., Thomas, J., Meyer, J., Lewitzky S., Rennich, J.S., Parker, A., DuPrat, L. *et al.* (2001) A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes*, **50**, 681–685.
5. Wiltshire, S., Hattersley, A.T., Hitman, G.A., Walker, M., Levy, J.C., Sampson, M., O'Rahilly, S., Frayling, T.M., Bell, J.I., Lathrop, G.M. *et al.* (2001) A genomewide scan for loci predisposing to type 2 diabetes in a UK population (the Diabetes UK Warren 2 repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am. J. Hum. Genet.*, **69**, 553–569.
6. Luo, T.H., Zhao, Y., Li, G., Yuan, W.T., Zhao, J.J., Chen, J.L., Huang, W. and Luo, M. (2001) A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia*, **44**, 501–506.
7. Mahtani, M.M., Widen, E., Lehto, M., Thomas, J., McCarthy, M., Brayer, J., Bryant, B., Chan, G., Daly, M., Forsblom, C. *et al.* (1996) Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat. Genet.*, **14**, 90–94.
8. Ghosh, S., Watanabe, R.M., Valle, T.T., Hauser, E.R., Magnuson, V.L., Langefeld C.D., Ally, D.S., Mohike, K.L., Silander, K., Kohtamaki, K. *et al.* (2000) The Finland–United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am. J. Hum. Genet.*, **67**, 1174–1185.
9. Watanabe, R.M., Ghosh, S., Langefeld, C.D., Valle, T.T., Hauser, E.R., Magnuson V.L., Mohlke, K.L., Silander, K., Ally, D.S., Chines, P. *et al.* (2000) The Finland–United States investigation of non-insulin-dependent diabetes quantitative-trait loci. *Am. J. Hum. Genet.*, **67**, 1186–1200.
10. Parker, A., Meyer, J., Lewitzky, S., Rennich, J.S., Chan, G., Thomas, J.D., Orho-Melander, M., Lehtovirta, M., Forsblom, C., Hyrkko, A. *et al.* (2001) A gene conferring susceptibility to type 2 diabetes in conjunction with obesity is located on chromosome 18p11. *Diabetes*, **50**, 675–680.
11. Lindgren, C.M., Mahtani, M.M., Widén, E., McCarthy, M.I., Daly, M.J., Kirby, A., Reeve, M.P., Kruglyak, L., Parker, A., Meyer, J. *et al.* (2002) Genomewide search for type 2 diabetes mellitus susceptibility loci in Finnish families: the Botnia study. *Am. J. Hum. Genet.*, **70**, 509–516.
12. Vionnet, N., Hani, E.H., Dupont, S., Gallina, S., Francke, S., Dotte, S., De Matos, F., Durand, E., Lepretre, F., Lecoeur, C. *et al.* (2000) Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. *Am. J. Hum. Genet.*, **67**, 1470–1480.
13. Hanis, C.L., Boerwinkle, E., Chakraborty, R., Ellsworth, D.L., Concannon, P., Stirling B., Morrison, V.A., Wapelhorst, B., Spielman, R.S., Gogolin-Ewens, K.J. *et al.* (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat. Genet.*, **13**, 161–166.
14. Duggirala, R., Blangero, J., Almasy, L., Dyer, T., Williams, K., Leach, R., O'Connell, P. and Stern, M. (1999) Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am. J. Hum. Genet.*, **64**, 1127–1140.
15. Hanson, R.L., Ehm, M.G., Pettitt, D.J., Prochazka, M., Thompson, D.B., Timberlake, D., Foroud, T., Kobes, S., Baier, L., Burns, D.K. *et al.* (1998) An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am. J. Hum. Genet.*, **63**, 1130–1138.
16. Elbein, S., Hoffman, M., Teng, K., Leppert, M. and Hasstedt, S. (1999) A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes*, **48**, 1175–1182.
17. Bowden, D.W., Sale, M., Howard, T.D., Qadri, A., Spray, B.J., Rothschild, C.B., Akots, G., Rich, S.S. and Freedman, B.I. (1997) Linkage of genetic markers on human chromosomes 20 and 12 to NIDDM in Caucasian sib pairs with a history of diabetic nephropathy. *Diabetes*, **46**, 882–886.
18. Ji, L., Malecki, M., Warram, J.H., Yang, Y., Rich, S.S. and Krowlewski, A.S. (1997) New susceptibility locus for NIDDM is localized to human chromosome 20q. *Diabetes*, **46**, 876–881.
19. Zouali, H., Hani, E.H., Philippi, A., Vionnet, N., Beckmann, J., Demenais, F. and Froguel, P. (1997) A susceptibility locus for early-onset non-insulin dependent (type 2) mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene. *Hum. Mol. Genet.*, **6**, 1401–1408.

20. Boehnke, M., Cox, N.J. and International Type 2 Diabetes Linkage Analysis Consortium (1998) Lessons learned in a combined linkage analysis (24 datasets, >2000 families) of type 2 diabetes on chromosome 20. Paper presented at the *48th annual meeting of the American Society of Human Genetics*, Denver, October 27–31.
21. Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E. *et al.* (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.*, **26**, 163–175.
22. Glazier, A.M., Nadeau, J.H. and Aitman, T.J. (2002) Finding genes that underlie complex traits. *Science*, **298**, 2345–2349.
23. Wise, L.H., Lanchbury, J.S. and Lewis, C.M. (1999) Meta-analysis of genome searches. *Ann. Hum. Genet.*, **63**, 263–272.
24. Loesgen, S., Dempfle, A., Golle, A. and Bickeboller, H. (2001) Weighting schemes in pooled linkage analysis. *Genet. Epidemiol.*, **21** (Suppl. 1), S142–147.
25. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. and Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347–1363.
26. Kong, A. and Cox, N.J. (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am. J. Hum. Genet.*, **61**, 1179–1188.
27. Lander, E.S. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.*, **11**, 241–247.
28. Wilsthire S., Cardon, L.R. and McCarthy, M.I. (2002) Evaluating the results of genomewide linkage scans of complex traits by locus counting. *Am. J. Hum. Genet.*, **71**, 1175–1182.
29. Kissebah, A.H., Sonnenberg, G.E., Myklebust, J., Goldstein, M., Broman, K., James R.G., Marks J.A., Krakower, G.R., Jacob, H.J., Weber, J., Martin, L. *et al.* (2000) Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc. Natl Acad. Sci. USA*, **97**, 14478–14483.
30. Klos, K.L., Kardia, S.L., Ferrell, R.E., Turner, S.T., Boerwinkle, E. and Sing, C.F. (2001) Genome-wide linkage analysis reveals evidence of multiple regions that influence variation in plasma lipid and apolipoprotein levels associated with risk of coronary heart disease. *Arterioscler. Thromb. Vasc. Biol.*, **21**, 971–978.
31. Commuzzie, A.G., Hixson, J.E., Almasy, L., Mitchell, B.D., Mahaney, M.C., Dyer, T.D., Stern, M.P., MacCluer, J.W. and Blangero, J. (1997) A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat. Genet.*, **15**, 273–276.
32. Hager, J., Dina, C., Francke, S., Dubois, S., Houari, M., Vatin, V., Vaillant, E., Lorentz, N., Basdevant, A., Clement, K. *et al.* (1998) A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. *Nat. Genet.*, **20**, 304–308.
33. Rotimi, C.N., Comuzzie, A.G., Lowe, W.L., Luke, A., Blangero, J. and Cooper, R.S. (1999) The quantitative trait locus on chromosome 2 for serum leptin levels is confirmed in African-Americans. *Diabetes*, **48**, 643–644.
34. Meigs, J.B., Panhuysen, C.I., Myers, R.H., Wilson, P.W. and Cupples, L.A. (2002) A genome-wide scan for loci linked to plasma levels of glucose and HbA(lc) in a community-based sample of Caucasian pedigrees: the Framingham Offspring Study. *Diabetes*, **51**, 833–840.
35. Hsueh, W.C., St Jean, P.L., Mitchell, B.D., Pollin, T.I., Knowler, W.C., Ehm, M.G., Bell, C.J., Sakul, H., Wagner, M.J., Burns, D.K. *et al.* (2003) Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. *Diabetes*, **52**, 550–557.
36. Vaxillaire, M., Boccio, V., Philippi, A., Vigouroux, C., Terwilliger, J., Passa, P., Beckmann, J.S., Velho, G., Lathrop, G.M. and Froguel P. (1995) A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q. *Nat. Genet.*, **9**, 418–423.
37. Shaw, J.T., Lovelock, P.K., Kesting, J.B., Cardinal, J., Duffy, D., Wainwright, B. and Cameron, D.P. (1998) Novel susceptibility gene for late-onset NIDDM is localized to human chromosome 12q. *Diabetes*, **47**, 1793–1796.
38. Duggirala, R., Blangero, J., Almasy, L., Arya, R., Dyer, T.D., Williams, K.L., Leach, R.J., O'Connell, P. and Stern, M.P. (2001) A major locus for fasting insulin concentrations and insulin resistance on chromosome 6q with strong pleiotropic effects on obesity-related phenotypes in non-diabetic Mexican Americans. *Am. J. Hum. Genet.*, **68**, 1149–1164.
39. Gardner, R.J., Mungall, A.J., Dunham, I., Barber, J.C., Shield, J.P., Temple, I.K. and Robinson, D.O. (1999) Localisation of a gene for transient neonatal diabetes mellitus to an 18.72 cR3000 (approximately 5.4 Mb) interval on chromosome 6q. *J. Med. Genet.*, **36**, 192–196.
40. Nishimura, D.Y., Searby, C.C., Carmi, R., Elbedur, K., Van Maldergem, L., Fulton, A.B., Lam, B.I., Powell B.R., Swiderski R.E., Bugge, K.E. *et al.* (2001) Positional cloning of a novel gene on chromosome 16q causing Bardet–Biedl syndrome (*BBS2*). *Hum. Mol. Genet.*, **10**, 865–874.
41. Doria, A., CaIdwell, J.S., Ji, L., Reynet, C., Rich, S.S., Weremowicz, S., Morton, C.C., Warram, J.H., Kahn, C.R. and Krolewski, A.S. (1995) Trinucleotide repeats at the rad locus. Allele distributions in NIDDM and mapping to a 3-cM region on chromosome 16q. *Diabetes*, **44**, 243–247.
42. Orho, M., Carlsson, M., Kanninen, T. and Groop, L.C. (1996) Polymorphism at the rad gene is not associated with NIDDM in Finns. *Diabetes*, **45**, 429–433.
43. Terwilliger, J. and Goring, H.H. (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum. Biol.*, **72**, 63–132.
44. World Health Organization Study Group (1985) *Diabetes Mellitus*. WHO Technical Report Series no. 727. World Health Organization, Geneva.
45. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (1997) Report on the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*, **20**, 1183–1197.
46. O'Connell, J. and Weeks, D. (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nat. Genet.*, **11**, 402–408.