

1-1-2009

## **Agreement among type 2 diabetes linkage studies but a poor correlation with results from genome-wide association studies**

A. Wilton  
*University of New South Wales*

S. Lillioja  
*University of Wollongong, lillioja@uow.edu.au*

Follow this and additional works at: <https://ro.uow.edu.au/medpapers>



Part of the [Medicine and Health Sciences Commons](#)

---

### **Citation**

Wilton, A. and Lillioja, S., 2009, Agreement among type 2 diabetes linkage studies but a poor correlation with results from genome-wide association studies, *Diabetologia*, 52(6), 1061-1074.  
<https://ro.uow.edu.au/medpapers/1>

---

## Agreement among type 2 diabetes linkage studies but a poor correlation with results from genome-wide association studies

### Abstract

**AIMS/HYPOTHESIS:** Little of the genetic basis for type 2 diabetes has been explained, despite numerous genetic linkage studies and the discovery of multiple genes in genome-wide association (GWA) studies. To begin to resolve the genetic component of this disease, we searched for sites at which genetic results had been corroborated in different studies, in the expectation that replication among studies should direct us to the genomic locations of causative genes with more confidence than the results of individual studies. **METHODS:** We have mapped the physical location of results from 83 linkage reports (for type 2 diabetes and diabetes precursor quantitative traits [QTs, e.g. plasma insulin levels]) and recent large GWA reports (for type 2 diabetes) onto the same human genome sequence to identify replicated results in diabetes genetic 'hot spots'. **RESULTS:** Genetic linkage has been found at least ten times at 18 different locations, and at least five times in 56 locations. All replication clusters contained study populations from more than one ethnic background and most contained results for both diabetes and QTs. There is no close relationship between the GWA results and linkage clusters, and the nine best replication clusters have no nearby GWA result. **CONCLUSIONS/INTERPRETATION:** Many of the genes for type 2 diabetes remain unidentified. This analysis identifies the broad location of yet to be identified genes on 6q, 1q, 18p, 2q, 20q, 17pq, 8p, 19q and 9q. The discrepancy between the linkage and GWA studies may be explained by the presence of multiple, uncommon, mildly deleterious polymorphisms scattered throughout the regulatory and coding regions of genes for type 2 diabetes.

### Keywords

wide, diabetes, 2, type, among, agreement, genome, results, correlation, poor, but, studies, association, linkage

### Disciplines

Medicine and Health Sciences

# Agreement among type 2 diabetes linkage studies but a poor correlation with results from genome-wide association studies

S. Lillioja · A. Wilton

Received: 18 December 2008 / Accepted: 13 February 2009 / Published online: 19 March 2009  
© Springer-Verlag 2009

## Abstract

**Aims/hypothesis** Little of the genetic basis for type 2 diabetes has been explained, despite numerous genetic linkage studies and the discovery of multiple genes in genome-wide association (GWA) studies. To begin to resolve the genetic component of this disease, we searched for sites at which genetic results had been corroborated in different studies, in the expectation that replication among studies should direct us to the genomic locations of causative genes with more confidence than the results of individual studies. **Methods** We have mapped the physical location of results from 83 linkage reports (for type 2 diabetes and diabetes precursor quantitative traits [QTs, e.g. plasma insulin levels]) and recent large GWA reports (for type 2 diabetes) onto the same human genome sequence to identify replicated results in diabetes genetic ‘hot spots’. **Results** Genetic linkage has been found at least ten times at 18 different locations, and at least five times in 56 locations. All replication clusters contained study populations from more than one ethnic background and most contained results for both diabetes and QTs. There is no close relationship

between the GWA results and linkage clusters, and the nine best replication clusters have no nearby GWA result.

**Conclusions/interpretation** Many of the genes for type 2 diabetes remain unidentified. This analysis identifies the broad location of yet to be identified genes on 6q, 1q, 18p, 2q, 20q, 17pq, 8p, 19q and 9q. The discrepancy between the linkage and GWA studies may be explained by the presence of multiple, uncommon, mildly deleterious polymorphisms scattered throughout the regulatory and coding regions of genes for type 2 diabetes.

**Keywords** Allelic heterogeneity · Cluster analysis · Genetic association · Genetic linkage · Genome-wide association study · MODY · Quantitative trait · Replication · Single nucleotide polymorphisms · Type 2 diabetes

## Abbreviations

GWA Genome-wide association  
LOD Logarithm of odds  
QT Quantitative trait  
SNP Single nucleotide polymorphism

**Electronic supplementary material** The online version of this article (doi:10.1007/s00125-009-1324-9) contains supplementary material, which is available to authorised users.

S. Lillioja (✉)  
Illawarra Health and Medical Research Institute,  
University of Wollongong,  
Room 41.319, Northfields Avenue,  
Wollongong, NSW 2522, Australia  
e-mail: lillioja@uow.edu.au

A. Wilton  
The School of Biotechnology and Biomolecular Science (BABS),  
University of New South Wales,  
Sydney, NSW, Australia

## Introduction

Type 2 diabetes is a chronic metabolic disease affecting the lives of millions and creating major healthcare problems worldwide. The huge personal and financial costs, together with good evidence for a genetic component of the disease, have justified a substantial investment into the search for novel type 2 diabetes genes [1–83]. Linkage studies have shown limited success to date in identifying causative genes. Genome-wide association (GWA) studies have identified multiple novel, apparently causative genes, but the contribution of these to disease risk and predictive value is small

[84, 85]. Pathway-specific candidate gene searches have also had little success. Therefore, most of the genetic contribution to type 2 diabetes remains unknown and the pressure for solutions to the disease remains intense. The environmental contributions to type 2 diabetes may partly be explained by a gene–environmental interaction whereby a particular environment triggers the disease in those with an underlying genetic predisposition. The search for diabetes disease genes is therefore as critical as the current efforts to modify the environmental and lifestyle factors that contribute to the disease.

Family-based studies of the genetic determinants of type 2 diabetes and related precursor quantitative traits (QTs, e.g. plasma insulin and glucose levels) [1–83] and GWA studies have now provided an abundance of evidence for potentially causative genes. These results have been drawn together onto a single map of the human genome sequence [86]. The goal is to look for genomic locations where the presence of a potential underlying type 2 diabetes gene has been attested to repeatedly—diabetes genetic ‘hot spots’. Such replication increases our confidence of the presence of an underlying gene. While GWA studies look for diabetes genes using a different approach to linkage analysis, the ultimate goal is the same—to find the genetic determinants of the disease. Therefore, the results of linkage and association must eventually match each other. The current analysis identifies multiple linkage locations that differ from those found in the recent GWA studies [87–89], and suggests the location of additional major type 2 diabetes susceptibility genes.

## Methods

Linkage studies of type 2 diabetes and related QTs were identified through Medline or other literature searches. In total, 52 genetic linkage projects reported in 83 publications were included in our analysis. Participants originated from multiple ethnic backgrounds. We were aware of some overlap in study populations, mainly the Genetics of NIDDM (GENNID) study [18, 22, 23, 25, 31, 44], but also other studies [4, 17, 30, 59]. The number of individual results removed because of this to avoid redundancy was small. Pima Indian QT studies from clinical research [65] or field studies [33] were treated separately. For studies that updated their analyses we have endeavoured to use the most recent results.

The X chromosome is excluded because of limited results. There were 450 different genetic marker names representing 439 loci (11 markers had two aliases). The physical location of the marker reported for the linkage peak or the 2 point score, was identified on build 36.1 of the Human Genome from UCSC Genome Bioinformatics (<http://genome.ucsc.edu/>, accessed July 2008) [86]. Where the linkage location fell between several markers, the p-terminal marker was used. Inclusion of results is based on logarithm of odds (LOD)

score, not  $p$  value. Linkage peaks for clarity have been plotted as a single point/line. To be included the marker name needed to be explicitly provided in the publication (for study [78] the marker names were provided by the senior investigator).

To avoid mapping multiple results from the same linkage signal, linkage results were sorted by study population and phenotype and then ranked by descending LOD score. Any result with a lower LOD score for the same population/phenotype within 30 Mb in either direction of the signal was deleted. A physical distance of 30 Mb is approximately 33 cM. Using the Kosambi map function, this corresponds to a recombination fraction of 0.29 (maximum is 0.5) [90]. Some studies had both subgroup and combined group analyses. Since the aim of this study was to identify replicated findings, we preferentially used the subgroups (e.g. for the Finland–United States Investigation of NIDDM Genetics [FUSION], select FUSION1 and FUSION2 and delete FUSION1+2 [73, 78] if co-located).

The following simplified phenotypes were created prior to filtering: (1) diabetes (diabetes, diabetes and impaired glucose tolerance combined, diabetes age of onset), (2) glucose (fasting glucose, OGTT glucose levels, HbA<sub>1c</sub>), (3) insulin (any insulin, proinsulin, C-peptide, some bivariate results), (4) acute insulin response (AIR), (5) minimal model insulin sensitivity (Si), (6) minimal model glucose effectiveness (Sg), and (7) euglycaemic–hyperinsulinaemic clamp (glucose infusion rate,  $M$  value). Phenotypes constructed from two categories were filtered against the results in both categories. While obesity also contributes to type 2 diabetes susceptibility, it is a distinct genetic disorder that is not always associated with the disease, and was therefore not included.

For several studies we converted  $p$  values to LOD scores using the assumption that  $\text{LOD} \times 2\log_e 10$  has a  $\chi^2$  distribution [91]. We assumed a two-sided test for [1], according to the authors instructions, and a one-sided test for [43, 46] (i.e.  $p$  value doubled before further calculation). This may introduce some error, but the location estimates are not influenced by the calculation. A number of authors supplied additional data for their studies on request, but we did not systematically request data from all authors.

The SAS statistical package, version 9.1 (SAS Institute, Cary, NC, USA) was used for the analyses. Linkage results for any particular underlying gene are expected to lie close to the genomic location of the gene. Since these results may be scattered locally, rather than falling in exactly the same spot, and to make the identification of any potential clusters of replicated results as objective as possible, the refined results were analysed using MODECLUS (method 1), a SAS statistical clustering procedure. It is not possible to assign a significance value to an individual cluster. The initial seeding radius for clusters provided to MODECLUS was 12 Mb, but lower if there were results not unambiguously assigned to a cluster. Large seeds will put all

observations into one large cluster, which is clearly unhelpful. Three markers fell in the same cluster as a related result (but >30 Mb away). These were deleted from further analyses after the goodness-of-fit tests. The goodness-of-fit tests evaluated the distribution of location results against a uniform distribution (Anderson–Darling statistic).

A quality score was developed and applied to each cluster. Clusters were first ranked by three properties: the sum of all the LOD scores, the density of the cluster (number of studies per 10 Mb of genome), and the number of studies in the cluster with an LOD score of  $\geq 3.0$ . The sum of these ranks was used to calculate a final ranking, which we use as a quality score (the 15 best are in Table 1).

**Calculations** The significance of the relationship between the GWA results and clusters was calculated as:  $P(k) = \{n!/[k!(n-k)!]\}p^k(1-p)^{(n-k)}$ , where  $n$  is the number of trials (i.e. 20 for GWA results),  $k$  is the number of hits (i.e. GWA markers falling into the clusters) and  $p$  is the probability of success in one trial (i.e. amount of genome covered by clusters). The overall probability is the sum over  $k, k+1, k+2 \dots k+20$ .

The  $\lambda_s$  value (sibling recurrence risk ratio, which is the risk to a sibling relative to the general population; Table 3) is calculated for an additive genetic effect as follows [90, 92]:

$$p = \text{risk allele frequency, } q = (1 - p)$$

$$g = \text{odds per allele (for two alleles, } g_2 = (2g - 1); \text{ one allele, } g_1 = g; \text{ no alleles, } g_0 = 1)$$

Note: Since  $f_0$ , the disease penetrance in those without the disease allele, eventually cancels out in these equations, only the genotype relative risks (odds ratios) are included.

$$V_A = 2pq\{[p(g_2 - g_1)] + [q(g_1 - g_0)]\}^2$$

$$V_D = (p^2)(q^2)[(g_2 - 2g_1 + g_0)^2]$$

$$K = (p^2g_2) + (2pqg_1) + (q^2)$$

$$C = (V_A \times 0.5) + (V_D \times 0.25)$$

$$K_s = K + (C/K)$$

$$\lambda_s = K_s/K$$

where  $V_A$  is the additive variance,  $V_D$  is the dominance variance,  $K$  is the population prevalence of the disease,  $C$  is the genetic covariance between two full siblings (epistatic effects ignored) and  $K_s$  is the sibling recurrence risk.

## Results

We identified 560 linkage results, on 22 autosomes, representing 439 different marker locations that met the following criteria: an LOD score of  $\geq 1.18$  (equivalent to  $p \leq$

0.01 in a single-point analysis or the likelihood of about two such results occurring by chance in a genome-wide analysis [93]); a marker with a known genomic location; and each result independent of any other result lying either within 30 Mb or in the same cluster (Tables 1 and 2, Fig. 1, Electronic supplementary material [ESM] Figs 1 and 2). There were 264 linkage results for the type 2 diabetes phenotype, and 296 linkage results for QTs. The number of linkage reports according to racial group were: Europeans 266, Native/Mexican American 124, Chinese/Japanese 78, African 55, Other/mixed/uncertain 37. Among the Europeans there was almost an equal number of results for type 2 diabetes and QTs (124 and 142, respectively), but QT studies dominated in the Native/Mexican American and African groups (>70% of results), and type 2 diabetes analyses predominated among the East Asians (83% of results). LOD scores of  $\geq 2.2$ ,  $\geq 3.0$  and  $\geq 3.6$  were present in 188 (34%), 80 (14%) and 41 (7%) of results, respectively (genome-wide  $p$  values  $\sim 0.2$ ,  $\sim 0.05$ ,  $\sim 0.01$  [93]). By other assessments, an LOD of 2.2 is expected to occur by chance once in every genome-wide scan [91]. Within each major racial group a similar proportion of results (13–17%) had LOD scores  $\geq 3.0$ .

By goodness-of-fit tests against a uniform distribution, the genomic locations of the linkage results were non-random in a number of assessments. Testing all results (chromosomes 1–22) as a continuous block suggested a non-uniform distribution ( $p=0.045$ ). Chromosomes 1, 8, 10, 17 and 18 show a non-uniform distribution ( $p \leq 0.02$ ), as do 2, 5 and 6 ( $p \leq 0.05$ ), while on chromosomes 3 and 21, the evidence is only suggestive ( $p \leq 0.075$ ). These tests may be too crude to truly assess clustering and may obscure finer detail.

Cluster analysis by individual chromosome identified 56 clusters, each containing at least five linkage results. The selection criterion of five members to define a cluster is arbitrary but provides a compromise between showing strong support for a replication locus on the one hand and discarding too many results of modest evidence for replication on the other. The 56 clusters contained 471 results (84% of total), had a mean size of 20.3 Mb ( $\pm 11.7$  Mb), and covered 39.7% of the autosomes (Tables 1 and 2, Fig. 1, ESM Figs 1 and 2). The clusters contained a mixture of type 2 diabetes and QT results, except for three small clusters (QT only). There were 89 (16%) results not assigned to a cluster, with type 2 diabetes and QT equally represented (17% and 15%, respectively). The mean LOD score was not significantly different between these orphaned results and clustered results (ANOVA and Wilcoxon tests). All clusters contained studies from several different racial groups (Tables 1 and 2).

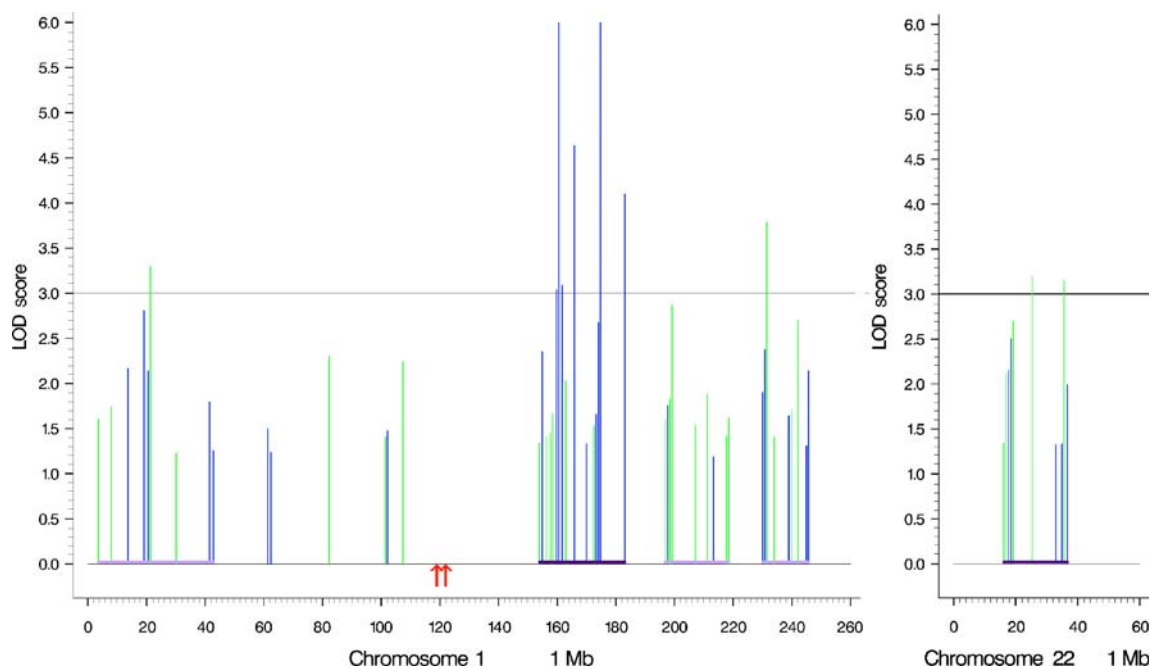
The clusters were given a quality rank (Tables 1 and 2), and according to this, the top three clusters were located on

**Table 1** Genome locations of type 2 diabetes linkage replication (best 15 locations ranked by quality score)

Chromosome no.	Results per cluster <sup>a</sup>	Cluster size (Mb) <sup>b</sup>	Start–end (Mb) <sup>c</sup>	Chromosome bands <sup>d</sup>	LOD $\geq 2.2^e$	LOD $\geq 3.0^e$	LOD $\geq 3.6^e$	Results per 10 Mb <sup>f</sup>	Quality score rank <sup>g</sup>	GWA result in or near cluster <sup>h</sup>	Ethnic contribution <sup>i</sup>	Phenotypes <sup>j</sup>	References <sup>k</sup>
6	12	9	131–140	q23.1–q24.1	5	3	2	1.4	1		CJ(3)Eu(6)NM(2)Yz(1)	8D4Q	[1, 6, 21, 32, 39, 66, 68, 72, 79, 82]
1	16	29	154–183	q22–q25.3	8	6	4	0.6	2		Af(1)CJ(2)Eu(9)NM(4)	10D6Q	[19, 32, 35, 36, 38, 43, 54, 58, 65, 70, 76, 78, 81, 82]
18	10	11	1–11	p11.32–p11.21	7	2	1	0.9	3		Af(1)CJ(1)Eu(7)Yz(1)	6D4Q	[6, 15, 25, 26, 36, 47, 62, 66, 75, 82]
2	13	24	218–243	q35–q37.3	3	3	1	0.5	4		Af(1)CJ(2)Eu(6)NM(1)Yz(3)	8D5Q	[5, 18, 24, 25, 28, 36, 37, 40, 47, 53, 55, 79, 82]
20	12	21	37–57	q11.23–q13.32	6	2	1	0.6	5		Af(1)CJ(5)Eu(5)NM(1)	9D3Q	[17, 21, 39, 42, 46, 51, 55, 63, 69, 73, 76, 82]
17	13	25	0–25	p13.3–q11.2	4	3	1	0.5	6		Af(1)CJ(1)Eu(7)NM(1)Yz(3)	5D8Q	[8, 18, 27, 53, 68(2), 69, 78, 79, 82]
8	15	33	2–35	p23.3–p12	6	3	2	0.5	7		Af(1)CJ(1)Eu(7)NM(4)Yz(2)	9D6Q	[1, 2, 5, 7, 12, 13, 25, 29, 34, 38, 40, 64, 65, 78, 82]
19	9	11	52–62	q13.32–q13.43	3	2	0	0.8	8		Af(1)Eu(6)NM(1)Yz(1)	1D8Q	[3, 4, 43, 54, 65, 74, 78]
9	7	10	72–82	q21.11–q21.31	5	2	1	0.7	9		CJ(4)Eu(2)NM(1)	4D3Q	[13, 17, 48, 51, 55, 61, 82]
10	12	23	111–134	q25.1–q26.3	5	2	0	0.5	9	TCF7L2 <sup>l</sup>	Af(2)CJ(1)Eu(5)NM(4)	4D8Q	[2, 14, 30, 33, 43, 54, 65, 66, 70, 76, 82]
3	6	8	190–199	q28–q29	4	3	1	0.7	11	IGF2BP2 <sup>m</sup>	Eu(4)NM(1)Yz(1)	2D4Q	[12, 41, 54(2), 65, 76]
19	10	13	3–16	p13.3–p13.12	3	1	1	0.8	12		Af(2)Eu(5)NM(3)	4D6Q	[1, 7, 14, 30, 37, 65–67, 69, 76]
22	10	20	16–36	q11.1–q12.3	4	2	0	0.5	12		Af(1)CJ(1)Eu(5)NM(3)	5D5Q	[7, 26, 29, 34, 39, 61, 65, 70, 78]
14	11	17	20–37	q11.2–q21.1	3	1	0	0.6	14		CJ(2)Eu(7)NM(1)Yz(1)	7D4Q	[6, 7, 23(2), 33, 36, 55, 78, 80]
2	9	18	156–174	q24.1–q31.1	4	2	1	0.5	15		Eu(4)NM(2)Yz(3)	4D5Q	[7, 12, 21, 27, 40, 43, 54, 65, 76]

One row per replication cluster. Each of these clusters is identified by a solid line at the baseline in Fig. 1 and ESM Figs 1 and 2 (coloured dark purple)

<sup>a</sup>The number of linkage results in each cluster. <sup>b</sup>The number of Mb between the results at either end of the cluster. <sup>c</sup>The position of the cluster in Mb from the p-terminus. <sup>d</sup>The chromosome band of the result at either end of the cluster. <sup>e</sup>The number of LOD scores in the cluster that met the threshold criteria. <sup>f</sup>The number of results in the cluster divided by the cluster size (Mb). <sup>g</sup>Rank of quality score based on: the sum of all the LOD scores in the cluster, the number of studies per 10 Mb of genome (i.e. density) and the number of studies in the cluster with an LOD score of 3.0 or more. Rank 1 is the highest quality cluster. <sup>h</sup>Locus identified in recent GWA studies and shown here by the nearest gene. <sup>i</sup>Ethnic/racial sources of the data in the cluster (Af, African; CJ, Chinese/Japanese; Eu, European; NM, Native/Mexican American; Yz, other/combined/uncertain), followed in parentheses by the number of such results. <sup>j</sup>Phenotypes in the cluster (D, type 2 diabetes; Q, quantitative trait) and the number of each type of result. <sup>k</sup>The number following the reference indicates if more than one result came from that publication. <sup>l</sup>SNP located in the cluster but not close to the mean. <sup>m</sup>SNP located outside the cluster but within 5 Mb of the cluster boundary



**Fig. 1** The linkage results for type 2 diabetes and related quantitative traits for chromosomes 1 and 22. The results for all 22 chromosomes plus the reference for the source of the data are given in ESM Fig. 1. The linkage results are plotted as a single line, each line representing a linkage peak or 2 point score, and positioned by the location of the genetic marker. Each result is independent of any other result by population and phenotype, within 30 Mb in either direction, and independent of any results in the same cluster if the cluster exceeds 30 Mb in size. The y-axis shows the LOD score of the results, with values  $\geq 6$  plotted as 6. The x-axis shows the location of the linkage result in megabase pairs from the p-terminus (according to <http://genome.ucsc.edu/>, accessed July 2008). Diabetes/impaired glucose tolerance results are in blue, QT results are in green. A thick solid

horizontal bar at the baseline demarcates each cluster. This is coloured deep purple for the best 15 clusters and pale purple for the remainder. A solid red arrow at the baseline identifies the location of a GWA SNP result (Table 3). For chromosome 1 the locations of the GWA results were adjusted to make the two closely located results visible separately. A number of linkage results were close to, or in an identical position to, another result. Therefore, for plotting purposes only, some results were moved in increments of 0.75 Mb until all values could be seen. The maximum move was 4.5 Mb (chromosome 6). The cluster range (i.e. Start–end) given in the tables is the actual position, whereas the plotted value may be adjusted to make the results visible. ESM Fig. 2 shows the clusters in chromosomes 1, 6, 18 and 22 that contained the most superimposed lines in more detail

chromosome 6q (very dense cluster), 1q (larger number of results and high LOD scores) and 18p (combination of properties). Linkage clusters were compared with results from recently published large GWA studies [87–89] (Table 3, Fig. 1, ESM Figs 1 and 2). There was no close relationship between the GWA and linkage results. Although 15 out of 20 GWA results were in a cluster or within 5 Mb of the cluster edge (56.1% of autosomal genome), this relationship was not significant ( $p=0.067$ ). Only five GWA results fell within 4 Mb of the mean of a cluster (15.6% of autosomal genome), and this is also not significant ( $p=0.19$ ). The best linkage replication clusters had no associated GWA result.

## Discussion

The search for genetic determinants of type 2 diabetes is yet to provide the anticipated insights into the cause of the disease. An abundance of results has not provided a consistent picture or satisfactory set of causative alterations.

The current compilation of linkage results for type 2 diabetes and its precursors suggests locations for type 2 diabetes susceptibility genes based on clusters of replicated results. It suggests that there are major genes for type 2 diabetes on 6q, 1q, 18p, 2q, 20q, 17p, 8p, 19q and 9q, and possibly elsewhere, that are yet to be identified. Because of the limitations of available genetic methods and the likely complexity of the underlying genetic architecture of type 2 diabetes, these genetic locations are currently only broadly defined but do provide promise of further major gene identification.

In this study, we combined the results of type 2 diabetes as a discrete trait and the results of QTs. This is justified under the assumption that heritable precursors of type 2 diabetes will share the same genetic determinants as type 2 diabetes itself. Diabetes-related QTs are heritable in multiple ethnic groups (for examples, see [94–96]), and many also predict the subsequent development of diabetes (for examples, see [97–99]). In addition, genetic correlation has been demonstrated between type 2 diabetes and some QTs [100] and between various QTs [14, 37, 96, 100–102].

**Table 2** Genome locations of type 2 diabetes linkage replication (remaining clusters listed by chromosome)

Chromosome	Results per cluster <sup>a</sup>	Cluster size (Mb) <sup>b</sup>	Start–end (Mb) <sup>c</sup>	Chromosome bands <sup>d</sup>	LOD $\geq 2.2^e$	LOD $\geq 3.0^e$	LOD $\geq 3.6^e$	Results per 10Mb <sup>f</sup>	Quality score rank <sup>g</sup>	GWA result in or near cluster <sup>h</sup>	Ethnic contribution <sup>i</sup>	Phenotypes <sup>j</sup>	References <sup>k</sup>
1	9	39	4–43	p36.32–p34.2	2	1	0	0.2	42		AF(2)CJ(3)Eu (2)NM(1)Yz(1)	5D4Q	[2, 3, 7, 35, 43, 55, 77, 82, 83]
1	9	21	197–218	q31.3–q41	1	0	0	0.4	41		CJ(2)Eu(2) NM(5)	2D7Q	[2, 16, 33, 39, 44, 53, 57, 65(2)]
1	9	15	230–245	q42.2–q44	3	1	1	0.6	20		AF(1)Eu(5) NM(1)Yz(2) AF(1)Eu(3)	5D4Q	[3, 7, 28, 53, 62, 65, 76, 78, 80]
2	5	9	2–12	p25.3–p25.1	0	0	0	0.5	47		AF(1)Eu(3) NM(1)	3D2Q	[2, 6, 64, 65, 76]
2	8	22	21–43	p24.1–p21	2	1	0	0.4	32	<i>THADA</i> <sup>m</sup>	AF(1)CJ(2)Eu (4)NM(1)	2D6Q	[2, 3, 20, 36, 39, 43, 57, 76]
2	6	23	62–85	p15–p11.2	1	0	0	0.3	54		AF(2)Eu(4)	1D5Q	[24, 30(3), 54, 78]
3	11	32	0–32	p26.3–p22.3	4	2	1	0.3	22	<i>PPARG</i> <sup>n</sup>	CJ(2)Eu(6) NM(3)	5D6Q	[6, 23, 33, 35, 39, 53, 58, 65, 78 (3)]
3	6	21	65–86	p14.1–p12.1	2	1	1	0.3	47	<i>ADAMTS9</i> <sup>m</sup>	CJ(2)Eu(1) NM(3)	4D2Q	[6, 31, 38, 55, 57, 65]
3	6	30	122–153	q13.33–q25.1	2	0	0	0.2	55		Eu(2)NM(2) Yz(2)	3D3Q	[3, 10, 28(2), 65(2)]
4	6	36	3–40	p16.2–p14	2	1	0	0.2	51	<i>WFS1</i> <sup>l</sup>	AF(1)Eu(4) NM(1)	2D4Q	[15, 35, 36, 65, 76, 78]
4	5	13	131–144	q28.2–q31.21	3	1	1	0.4	37		AF(2)CJ(1)Eu (2)	4D1Q	[10, 36, 58, 67, 69]
4	7	14	176–190	q34.1–q35.2	1	1	0	0.5	32		CJ(1)Eu(4) NM(2)	5D2Q	[2, 38, 48, 55, 63, 65, 76]
5	7	11	0–12	p15.33–p15.2	1	0	0	0.6	34		AF(1)Eu(4) NM(2)	3D4Q	[30(2), 50, 54, 65, 70, 80]
5	5	16	75–91	q13.3–q14.3	3	1	0	0.3	45		CJ(2)Eu(2) NM(1)	3D2Q	[17, 23, 52, 65, 82]
5	7	21	135–156	q31.1–q33.3	2	1	1	0.3	40		CJ(2)Eu(3) NM(1)Yz(1)	5D2Q	[7, 17, 29, 62, 65, 76, 82]
6	7	22	14–36	p23–p21.31	3	0	0	0.3	49	<i>CDKAL1</i> <sup>n</sup>	AF(1)CJ(2)Eu (2)NM(2)	2D5Q	[2, 39, 43, 55, 65, 68, 78]
6	8	27	86–113	q14.3–q21	5	4	3	0.3	20		CJ(2)Eu(4) NM(2)	6D2Q	[1, 10, 34, 50, 56, 57, 72(2)]
6	5	10	154–164	q25.2–q26	3	0	0	0.5	44		AF(3)CJ(1)Eu (1)	3D2Q	[5, 11, 43, 58, 70]
7	11	32	0–32	p22.3–p14.3	3	2	2	0.3	18	<i>JAZF1</i> <sup>l</sup>	AF(2)CJ(3)Eu (3)NM(3)	5D6Q	[2, 17, 24, 29, 39, 55, 65(2), 67, 70, 78]
7	17	71	67–138	q11.22–q34	6	3	2	0.2	16		AF(1)CJ(1)Eu (9)NM(6)	5D12Q	[3, 5, 14, 16, 18, 27, 30, 32, 36, 37, 39, 43, 45, 49, 65, 76, 78]
7	5	5	151–156	q36.1–q36.3	2	0	0	1.0	35		NM(2)Yz(3)	5Q	[4(2), 47, 65(2)]
8	12	45	93–138	q21.3–q24.23	4	2	1	0.3	25	<i>SLC30A8</i> <sup>n</sup>	Eu(8)NM(2) Yz(2)	5D7Q	[5, 7, 27–29, 36, 49, 53, 65(2), 76, 78]
9	11	33	1–34	p24.3–p13.3	1	0	0	0.3	43	<i>CDKN2A/2B</i> <sup>n</sup>	AF(1)CJ(2)Eu (3)NM(5)	3D8Q	[2, 33, 38, 51, 53, 65(3), 78(2), 82]





**Table 3** GWA study results for type 2 diabetes

Chromosome	Risk SNP	Start position	Risk allele frequency	Reference	Risk per allele	$\lambda_s$	Nearest gene	Description
1	rs2641348	120,239,407	0.107	89	1.10	1.001	<i>ADAM30</i>	ADAM metallopeptidase domain 30 preproprotein
1	rs10923931	120,319,482	0.106	89	1.13	1.002	<i>NOTCH2</i>	Notch 2 preproprotein
2	rs7578597	43,586,327	0.902	89	1.15	1.001	<i>THADA</i>	Thyroid adenoma associated isoform 1
3	rs1801282	12,368,125	0.870	87	1.14	1.001	<i>PPARG</i>	Peroxisome proliferator-activated receptor $\gamma$
3	rs4607103	64,686,944	0.761	89	1.09	1.001	<i>ADAMTS9</i>	ADAM metallopeptidase with thrombospondin type 1
3	rs4402960	186,994,381	0.320	87	1.14	1.004	<i>IGF2BP2</i>	Insulin-like growth factor 2 mRNA binding
4	rs10010131	6,343,816	0.600	87	1.11	1.002	<i>WFS1</i>	Wolframin
6	rs10946398	20,769,013	0.320	87	1.14	1.004	<i>CDKAL1</i>	CDK5 regulatory subunit-associated protein
7	rs864745	28,147,081	0.501	89	1.10	1.002	<i>JAZF1</i>	Juxtaposed with another zinc finger protein 1
8	rs13266634	118,253,964	0.690	87	1.15	1.003	<i>SLC30A8</i>	Solute carrier family 30 member 8
9	rs10811661	22,124,094	0.830	87	1.20	1.003	<i>CDKN2A/2B</i>	Cyclin-dependent kinase inhibitor 2A/2B
10	rs12779790	12,368,016	0.183	89	1.11	1.002	<i>CDC123/CAMK1D</i>	Cell division cycle homologue/Ca <sup>2+</sup> calmodulin protein kinase
10	rs1111875	94,452,862	0.650	87	1.15	1.004	<i>HHEX/IDE</i>	Haematopoietically expressed homeobox/insulysin
10	rs7901695	114,744,078	0.310	87	1.37	1.019	<i>TCF7L2</i>	Transcription factor 7-like 2
11	rs2237892	2,796,327	0.690	88	1.42	1.015	<i>KCNQ1</i>	Potassium voltage-gated channel KQT-like
11	rs5215	17,365,206	0.350	87	1.14	1.004	<i>KCNJ11</i>	Potassium inwardly-rectifying channel J11
12	rs1153188	53,385,263	0.733	89	1.08	1.001	<i>DCD</i>	Dermcidin preproprotein
12	rs7961581	69,949,369	0.269	89	1.09	1.001	<i>TSPAN8/LGR5</i>	Tetraspanin-8/leucine-rich G-protein coupled receptor 5
16	rs8050136	52,373,776	0.400	87	1.17	1.005	<i>FTO</i>	Fatso
17	rs4430796	33,172,153	0.470	87	1.10	1.002	<i>TCF2</i>	Transcription factor 2

Recently published results of large GWA studies [87–89]. These results are plotted in Fig. 1 and ESM Figs 1 and 2 by the location of the risk SNP. Each of the genes adjacent to the risk SNP, and that are also in or near a cluster, are listed in Tables 1 and 2. We have used risk allele frequencies and risk per allele data from one review [87] and two original publications [88, 89]), to calculate the effect on the sibling recurrence risk ratio ( $\lambda_s$ ) (see Methods, Calculations)

On the other hand, by providing twice the reservoir of data to analyse, the combination of evidence from both type 2 diabetes and precursor QTs should improve the likelihood of identifying type 2 diabetes gene locations. Almost all replication clusters contained both types of results, supporting this approach (Tables 1 and 2).

A recent meta-analysis [103] evaluated 23 type 2 diabetes linkage (but not QT) studies with the aim of determining gene locations. The genome was divided into bins based on genetic distance, and LOD scores in each bin were evaluated using three different scoring schemes. The study identified 24 linkage loci, with three of these loci roughly consistent across scoring schemes. These three loci corresponded to one of our clusters, chromosome 4 ~176 cM/176 Mb, chromosome 6 ~125–150 cM/125–150 Mb, and chromosome 10 ~138 cM/119 Mb. The three other results given special comment by the authors do not correspond to a cluster in our results. Our study is more qualitative than that of Guan et al. [103], but does allow for a visual evaluation

of the raw data, is not constrained by selection of bin size or location (we used cluster analysis instead), and the inclusion of QT doubles the number of results that can inform the gene location evaluation.

The individual linkage results show a wide scatter of locations (Fig. 1, ESM Figs 1 and 2). There are several potential explanations for this. Firstly, the data may merely be randomly distributed without any underlying genetic loci (see goodness-of-fit testing results). Alternatively, the scatter could reflect the limitations inherent in location estimates from linkage studies and the density of markers. Finally, the scatter could reflect the presence of multiple co-located independent genes.

We expect that both the linkage location estimate and the LOD score will be subject to error. This location error will determine the likely size of a cluster of results around a locus that is creating linkage signals. Genetic loci are unlinked when the recombination rate at meiosis between two loci reaches the maximum 50%, a distance of

approximately 100 cM, about 100 Mb [90, 104]. Accurately defined linkage signals this far apart are not linked, suggesting an upper limit for replication cluster sizes. The multipoint linkage peaks observed in genetic analyses are typically broad, with the width of a peak easily reaching 40 cM or more (see also [105]). This indicates the uncertainty in the location estimate for the underlying gene. Confidence intervals for the location of the susceptibility gene have been examined by modelling for affected sib-pairs [106, 107] and for family studies [108–110]. These intervals are often surprisingly large, i.e. tens of centimorgans. For a  $\lambda_s$  of 1.24 (i.e. tenfold larger than that for *TCF7L2* [Table 3]) and 400 sib-pairs, modelling suggested the standard deviation of the location estimate to be 13.11 cM, i.e. a 95% confidence interval of 51 cM ( $1.96 \times 2 \times 13.11$ ) [106]. Indeed, the chance of finding the susceptibility gene under the region of maximum allele sharing is quite small [111]. While the use of a 1 LOD distance on either side of the linkage peak might provide tight confidence limits, such limits can be deceptively narrow and will often completely miss the true location [107]. Hence, we should reasonably expect a scatter of results around a susceptibility locus, with a larger scatter observed when the gene has a small effect or the study population is small [108–110, 112]. Cluster analysis tries to take this scatter into account. Studies with larger samples might serve to better define the susceptibility gene locations by genuinely narrowing the confidence intervals.

Results with lower LOD scores are included here to incorporate as much corroborative evidence as possible. Including lower LOD scores may, however, increase the scatter of the data and increase the number of false-positive results. The orphaned (non-clustered) results did not have lower LOD scores, suggesting that they are not false-positives and instead reflect random error in the location estimates. Because of the large number of tests typically performed in a genome-wide scan, the cutoff values for genome-wide significance for linkage are stricter than the point-wise significance limits. An LOD of  $\geq 3.6$  (theoretical) [91] or  $\geq 3.0$  (modelling) [93] suggests genome-wide significance ( $p < 0.05$ ). Linkage results were centred around LOD scores of  $\geq 3.0$  in 40 (71%) of the clusters, and around LOD scores of  $\geq 3.6$  in 27 (48%) of the clusters. Inspection of the results can help the reader evaluate the linkage signals in these locations (Tables 1 and 2, Fig. 1, ESM Figs 1 and 2). When defining the correct position of the locus underlying a cluster, for positional cloning or bioinformatics purposes, it may be beneficial to weight each result.

In the current analysis, we excluded results within 30 Mb in either direction of the selected result if they were derived from the same study population and phenotype, because it was necessary to set limits to avoid spurious replication. This may have excluded a genuine second

linkage signal, lying nearby, although it is doubtful that signals in such close proximity can be adequately resolved in most studies. However, it is possible that the wide clusters represent multiple co-located genes, as seen in prostate cancer (8q24) [113] and several other diseases. For the moment, we invoke Occam's razor in assuming that, generally, a single cluster represents a single underlying gene, with the linkage signal picked up over a substantial length of the genome. The scatter also indicates that positional cloning efforts based on the location estimates from a single study alone may not correctly target the underlying gene.

The mean number of linkage results in a cluster was 8.4, and it is clear that the linkage signal for each putative underlying gene is not identified by every study (most, but not all, studies were genome-wide scans). The absence of a linkage signal could reflect the different population structures in the studies, the genetic heterogeneity of type 2 diabetes or the lack of power in individual studies.

If only a subset of type 2 diabetes susceptibility genes was required for the disease in any individual and the frequencies of these susceptibility genes were different in each population, linkage results would be variable. This might easily arise if hyperglycaemia was a collection of subtly different phenotypes, each resulting from different subsets of underlying genes. Heterogeneity for diabetes as a broad phenotype is already apparent in the distinct features of type 1 diabetes, type 2 diabetes and MODY/monogenic diabetes [114]. The non-monogenic form of type 2 diabetes is likely to feature further levels of heterogeneity. Phenotypic heterogeneity may be largely independent of the ethnic background however, since there was a mixture of racial groups in all replication clusters (Tables 1 and 2). Even though association studies [88, 115] suggest that there will be some differences in the frequency of individual type 2 diabetes genes between ethnic backgrounds, many type 2 diabetes genes may be shared between individuals of different continents of origin.

Studies with an insufficient number of participants may also fail to detect a linkage signal for locations where the genetic effect is weak. Given that  $>2,500$  families may be needed to detect loci conferring a genotype risk ratio of less than 2 ([116] compare with Table 3), it is expected that many studies will not be able to identify the location of some genes. The lack of support for a particular location in some studies does not discount the potential importance of the existing data, but does emphasise the importance of replication and the need for larger studies.

Fifteen GWA results were in or near a cluster. Given the size of clusters and the amount of the genome they cover, we were unable to demonstrate that this was a significant correlation. The GWA type 2 diabetes loci may be the source of the adjacent linkage signals in some cases, but

demonstrating this would require a specific re-analysis of the linkage data. Our nine best replication clusters are not associated with a GWA result, and these clusters quite likely reflect genes with even bigger effects than do the current GWA candidates. This is because the 20 loci identified by GWA studies have small  $\lambda_s$  values (Table 3), with a  $\lambda_s$  value in combination of only about 1.08–1.10. The expected  $\lambda_s$  value for type 2 diabetes appears to be somewhere between 1.2 and 6 [96, 117–120]. Therefore, we must conclude that much of the genetic cause of type 2 diabetes remains unidentified and the GWA results should be seen as complementing, not replacing, the linkage analyses.

While MODY genes are not considered responsible for typical type 2 diabetes, five of the nine known MODY (<http://www.ncbi.nlm.nih.gov/omim/>, accessed December 2008) genes fall into the following replication clusters: *MODY1/HNF4A* (chromosome 20/42.4 Mb); *MODY3/HNF1A* (chromosome 12/119.9 Mb); *MODY7/KLF11* (chromosome 2/10.1 Mb); *MODY8/CEL* (chromosome 9/134.9 Mb); and *MODY9/PAX4* (chromosome 7/127.0 Mb). Since some loci may contain both weak and strong alleles [121], it remains possible that variants in these genes could be influencing typical type 2 diabetes [122].

In addition to phenotypic heterogeneity, we should expect to see allelic heterogeneity in type 2 diabetes. There are ten million SNPs in the human genome, and in the two individuals whose whole genomes have been sequenced, over 3 million SNP differences were detected between each individual and the reference human genome sequence [123]. The majority of these SNPs will be rare in the whole population [124, 125]. Individual genes can harbour hundreds of genetic variants. As an illustration, the *MODY3* gene (*HNF1A*, the commonest monogenic cause of diabetes) has 200–300 known mutations (missense, nonsense, splicing defects, insertions and deletions) [126]. Therefore, we should expect the genetic architecture of typical type 2 diabetes to be much more complex still, and this will make it harder to identify genetic causes of the disease.

Both association studies and linkage analysis are looking for new causative genes using different principles [92, 121, 127–129]. Association analysis is considered more powerful than linkage analysis [92, 116], though the difference may have been inflated [128]. In linkage analysis, gene transmission is assessed one family at a time. If every family studied is affected by a different rare disease allele elsewhere in the same gene (allelic heterogeneity), a linkage signal should still be apparent [92]. Disease mutations need not be identical, only close enough to a marker that recombination over one or two generations rarely separates them in any particular family. A GWA study, however, would need to either type the same rare SNP directly in each individual, which is unlikely with current markers, or to indirectly type one or more rare SNPs

by typing a representative or ‘tag’ SNP [125] on the same haplotype block. The disease allele and the typed tag SNP also need to occur at about the same frequency for a study to identify the relationship between them [127]. The gene typing arrays currently available are largely based on the HapMap results [124, 125], which specifically targeted common SNPs. Investigators made the assumption that common diseases would be due to common variants [128]. It is not surprising therefore that the current set of GWA results are associated with SNPs with high-risk allele frequencies. As a corollary, the future discovery of new rare disease variants may be difficult using this methodology. Current GWA studies may have also missed copy number variants [130]. Therefore, GWA studies may be identifying genes that linkage studies have been insufficiently powered to detect, while linkage studies may be detecting genes with multiple rare variants within the same gene, variants rare enough individually to go undetected by current GWA tools.

Linkage studies have helped in the discovery of monogenic forms of type 2 diabetes (MODY), and most of these have probably now been identified. GWA studies have identified a number of common genes with low penetrance, but current methods may soon reach a limit here, too. This leaves the moderately rare genetic variants with modest penetrance left to be identified (see box 7 in [129]). Since one gene could harbour multiple, different, rare, modestly penetrant variants, we suggest that the most likely interpretation of strong linkage clusters with no associated GWA result is the presence of variants of this sort. The discovery of genes for type 2 diabetes now appears to require an increased catalogue of rare variants or large-scale re-sequencing of well-defined diabetes linkage ‘hotspots’ [131].

**Acknowledgement** The authors wish to thank S. Wiltshire for his helpful advice.

**Duality of interest** The authors declare that there is no duality of interest associated with this manuscript.

## References

1. Abney M, Ober C, McPeck MS (2002) Quantitative-trait homozygosity and association mapping and empirical genome-wide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites. *Am J Hum Genet* 70:920–934
2. An P, Hong Y, Weisnagel SJ et al (2003) Genomic scan of glucose and insulin metabolism phenotypes: the HERITAGE Family Study. *Metab Clin Exper* 52:246–253
3. An P, Teran-Garcia M, Rice T et al (2005) Genome-wide linkage scans for prediabetes phenotypes in response to 20 weeks of endurance exercise training in non-diabetic whites and blacks: the HERITAGE Family Study. *Diabetologia* 48:1142–1149
4. An P, Freedman BI, Hanis CL et al (2005) Genome-wide linkage scans for fasting glucose, insulin, and insulin resistance in the

- National Heart, Lung, and Blood Institute Family Blood Pressure Program: evidence of linkages to chromosome 7q36 and 19q13 from meta-analysis. *Diabetes* 54:909–914
5. An P, Freedman BI, Rich SS et al (2006) Quantitative trait loci on chromosome 8q24 for pancreatic beta-cell function and 7q11 for insulin sensitivity in obese nondiabetic white and black families: evidence from genome-wide linkage scans in the NHLBI Hypertension Genetic Epidemiology Network (HyperGEN) study. *Diabetes* 55:551–558
  6. Aulchenko YS, Vaessen N, Heutink P et al (2003) A genome-wide search for genes involved in type 2 diabetes in a recently genetically isolated population from The Netherlands. *Diabetes* 52:3001–3004
  7. Avery CL, Freedman BI, Heiss G et al (2004) Linkage analysis of diabetes status among hypertensive families: the Hypertension Genetic Epidemiology Network Study. *Diabetes* 53:3307–3312
  8. Avery CL, Freedman BI, Kraja AT et al (2006) Genotype-by-sex interaction in the aetiology of type 2 diabetes mellitus: support for sex-specific quantitative trait loci in Hypertension Genetic Epidemiology Network participants. *Diabetologia* 49:2329–2336
  9. Bektas A, Suprenant ME, Wogan LT et al (1999) Evidence of a novel type 2 diabetes locus 50 cM centromeric to NIDDM2 on chromosome 12q. *Diabetes* 48:2246–2251
  10. Bowden DW, Rudock M, Ziegler J et al (2006) Coincident linkage of type 2 diabetes, metabolic syndrome, and measures of cardiovascular disease in a genome scan of the diabetes heart study. *Diabetes* 55:1985–1994
  11. Broeckel U, Hengstenberg C, Mayer B et al (2002) A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet* 30:210–214
  12. Busfield F, Duffy DL, Kesting JB et al (2002) A genomewide search for type 2 diabetes-susceptibility genes in indigenous Australians. *Am J Hum Genet* 70:349–357
  13. Cai G, Cole SA, Freeland-Graves JH, MacCluer JW, Blangero J, Comuzzie AG (2004) Genome-wide scans reveal quantitative trait Loci on 8p and 13q related to insulin action and glucose metabolism: the San Antonio Family Heart Study. *Diabetes* 53:1369–1374
  14. Cai G, Cole SA, Butte NF, Voruganti VS, Comuzzie AG (2007) A quantitative trait locus on chromosome 13q affects fasting glucose levels in Hispanic children. *J Clin Endocrinol Metab* 92:4893–4896
  15. Chen G, Adeyemo A, Zhou J et al (2007) Genome-wide search for susceptibility genes to type 2 diabetes in West Africans: potential role of C-peptide. *Diabetes Res Clin Pract* 78:e1–6
  16. Cheng LS, Davis RC, Raffel LJ et al (2001) Coincident linkage of fasting plasma insulin and blood pressure to chromosome 7q in hypertensive Hispanic families. *Circulation* 104:1255–1260
  17. Chiu Y-F, Chuang L-M, Hsiao C-F et al (2005) An autosomal genome-wide scan for loci linked to pre-diabetic phenotypes in nondiabetic Chinese subjects from the Stanford Asia-Pacific Program of Hypertension and Insulin Resistance Family Study. *Diabetes* 54:1200–1206
  18. Cox NJ, Frigge M, Nicolae DL et al (1999) Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nat Genet* 21:213–215
  19. Das SK, Hasstedt SJ, Zhang Z, Elbein SC (2004) Linkage and association mapping of a chromosome 1q21-q24 type 2 diabetes susceptibility locus in northern European Caucasians. *Diabetes* 53:492–499
  20. Diego VP, Goring HHH, Cole SA et al (2006) Fasting insulin and obesity-related phenotypes are linked to chromosome 2p: the Strong Heart Family Study. *Diabetes* 55:1874–1878
  21. Duggirala R, Blangero J, Almasy L et al (2001) A major locus for fasting insulin concentrations and insulin resistance on chromosome 6q with strong pleiotropic effects on obesity-related phenotypes in nondiabetic Mexican Americans. *Am J Hum Genet* 68:1149–1164
  22. Duggirala R, Almasy L, Blangero J et al (2003) Further evidence for a type 2 diabetes susceptibility locus on chromosome 11q. *Genet Epidemiol* 24:240–242
  23. Ehm MG, Karnoub MC, Sakul H et al (2000) Genomewide search for type 2 diabetes susceptibility genes in four American populations. *J Hum Genet* 66:1871–1881
  24. Einarsdottir E, Mayans S, Ruikka K et al (2006) Linkage but not association of calpain-10 to type 2 diabetes replicated in northern Sweden. *Diabetes* 55:1879–1883
  25. Elbein SC, Hoffman MD, Teng K, Leppert MF, Hasstedt SJ (1999) A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes* 48:1175–1182
  26. Falchi M, Wilson SG, Paximadas D, Swaminathan R, Spector TD (2008) Quantitative linkage analysis for pancreatic B cell function and insulin resistance in a large twin cohort. *Diabetes* 57:1120–1124
  27. Fradin D, Heath S, Lathrop M, Bougneres P (2007) Quantitative trait loci for fasting glucose in young Europeans replicate previous findings for type 2 diabetes in 2q23-24 and other locations. *Diabetes* 56:1742–1745
  28. Francke S, Manraj M, Lacquemant C et al (2001) A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum Mol Genet* 10:2751–2765
  29. Frayling TM, Wiltshire S, Hitman GA et al (2003) Young-onset type 2 diabetes families are the major contributors to genetic loci in the Diabetes UK Warren 2 genome scan and identify putative novel loci on chromosomes 8q21, 21q22, and 22q11. *Diabetes* 52:1857–1863
  30. Freedman BI, Rich SS, Sale MM et al (2005) Genome-wide scans for heritability of fasting serum insulin and glucose concentrations in hypertensive families. *Diabetologia* 48:661–668
  31. Hanis CL, Boerwinkle E, Chakraborty R 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
  32. Hanson RL, Ehm MG, Pettitt DJ 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
  33. Hanson RL, Imperatore G, Narayan KM et al (2001) Family and genetic studies of indices of insulin sensitivity and insulin secretion in Pima Indians. *Diabetes/Metab Res Rev* 17:296–303
  34. Hegele RA, Sun F, Harris SB, Anderson C, Hanley AJ, Zinman B (1999) Genome-wide scanning for type 2 diabetes susceptibility in Canadian Oji-Cree, using 190 microsatellite markers. *J Hum Genet* 44:10–14
  35. Hoffmann K, Mattheisen M, Dahm S et al (2007) A German genome-wide linkage scan for type 2 diabetes supports the existence of a metabolic syndrome locus on chromosome 1p36.13 and a type 2 diabetes locus on chromosome 16p12.2. *Diabetologia* 50:1418–1422
  36. Hsueh W-C, St Jean PL, Mitchell BD 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
  37. Hsueh W-C, Silver KD, Pollin TI et al (2007) A genome-wide linkage scan of insulin level derived traits: the Amish Family Diabetes Study. *Diabetes* 56:2643–2648
  38. Hunt KJ, Lehman DM, Arya R et al (2005) Genome-wide linkage analyses of type 2 diabetes in Mexican Americans: the

- San Antonio Family Diabetes/Gallbladder Study. *Diabetes* 54:2655–2662
39. Iwasaki N, Cox NJ, Wang Y-Q et al (2003) Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects. *Diabetes* 52:209–213
  40. Kim S-H, Ma X, Weremowicz S et al (2004) Identification of a locus for maturity-onset diabetes of the young on chromosome 8p23. *Diabetes* 53:1375–1384
  41. Kissebah AH, Sonnenberg GE, Myklebust J 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
  42. Klupa T, Malecki MT, Pezzolesi M et al (2000) Further evidence for a susceptibility locus for type 2 diabetes on chromosome 20q13.1-q13.2. *Diabetes* 49:2212–2216
  43. Lakka TA, Rankinen T, Weinsagel SJ et al (2003) A quantitative trait locus on 7q31 for the changes in plasma insulin in response to exercise training: the HERITAGE Family Study. *Diabetes* 52:1583–1587
  44. Langefeld CD, Wagenknecht LE, Rotter JI et al (2004) Linkage of the metabolic syndrome to 1q23-q31 in Hispanic families: the Insulin Resistance Atherosclerosis Study Family Study. *Diabetes* 53:1170–1174
  45. Lehman DM, Arya R, Blangero J et al (2005) Bivariate linkage analysis of the insulin resistance syndrome phenotypes on chromosome 7q. *Hum Biol* 77:231–246
  46. Lembergas AV, Perusse L, Chagnon YC et al (1997) Identification of an obesity quantitative trait locus on mouse chromosome 2 and evidence of linkage to body fat and insulin on the human homologous region 20q. *J Clin Invest* 100:1240–1247
  47. Li W-D, Dong C, Li D, Garrigan C, Price RA (2004) A quantitative trait locus influencing fasting plasma glucose in chromosome region 18q22-23. *Diabetes* 53:2487–2491
  48. Lindgren CM, Mahtani MM, Widen E 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
  49. Lindgren CM, Widen E, Tuomi T et al (2002) Contribution of known and unknown susceptibility genes to early-onset diabetes in Scandinavia: evidence for heterogeneity. *Diabetes* 51:1609–1617
  50. Lindsay RS, Kobes S, Knowler WC, Bennett PH, Hanson RL (2001) Genome-wide linkage analysis assessing parent-of-origin effects in the inheritance of type 2 diabetes and BMI in Pima Indians. *Diabetes* 50:2850–2857
  51. Luo TH, Zhao Y, Li G et al (2001) A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia* 44:501–506
  52. Martin LJ, Comuzzie AG, Dupont S et al (2002) A quantitative trait locus influencing type 2 diabetes susceptibility maps to a region on 5q in an extended French family. *Diabetes* 51:3568–3572
  53. Meigs JB, Panhuysen CIM, Myers RH, Wilson PWF, Cupples LA (2002) A genome-wide scan for loci linked to plasma levels of glucose and HbA<sub>1c</sub> in a community-based sample of Caucasian pedigrees: the Framingham Offspring Study. *Diabetes* 51:833–840
  54. Meigs JB, Manning AK, Fox CS et al (2007) Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Medical Genetics* 8(Suppl 1):S16
  55. Mori Y, Otabe S, Dina C et al (2002) Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate Loci on 7p and 11p. *Diabetes* 51:1247–1255
  56. Nawata H, Shirasawa S, Nakashima N et al (2004) Genome-wide linkage analysis of type 2 diabetes mellitus reconfirms the susceptibility locus on 11p13-p12 in Japanese. *J Hum Genet* 49:629–634
  57. Ng MCY, So W-Y, Lam VKL et al (2004) Genome-wide scan for metabolic syndrome and related quantitative traits in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. *Diabetes* 53:2676–2683
  58. Ng MCY, So W-Y, Cox NJ et al (2004) Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. *Diabetes* 53:1609–1613
  59. North KE, Franceschini N, Borecki IB et al (2007) Genotype-by-sex interaction on fasting insulin concentration: the HyperGEN Study. *Diabetes* 56:137–142
  60. Palmer ND, Langefeld CD, Campbell JK et al (2006) Genetic mapping of disposition index and acute insulin response loci on chromosome 11q. The Insulin Resistance Atherosclerosis Study (IRAS) Family Study. *Diabetes* 55:911–918
  61. Panhuysen CIM, Cupples LA, Wilson PWF, Herbert AG, Myers RH, Meigs JB (2003) A genome scan for loci linked to quantitative insulin traits in persons without diabetes: the Framingham Offspring Study. *Diabetologia* 46:579–587
  62. Parker A, Meyer J, Lewitzky S et al (2001) A gene conferring susceptibility to type 2 diabetes in conjunction with obesity is located on chromosome 18p11. *Diabetes* 50:675–680
  63. Permutt MA, Wasson JC, Suarez BK et al (2001) A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* 50:681–685
  64. Pezzolesi MG, Nam M, Nagase T et al (2004) Examination of candidate chromosomal regions for type 2 diabetes reveals a susceptibility locus on human chromosome 8p23.1. *Diabetes* 53:486–491
  65. Pratley RE, Thompson DB, Prochazka M et al (1998) An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians. *J Clin Invest* 101:1757–1764
  66. Reynisdottir I, Thorleifsson G, Benediktsson R et al (2003) Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *Am J Hum Genet* 73:323–335
  67. Rich SS, Bowden DW, Haffner SM et al (2004) Identification of quantitative trait loci for glucose homeostasis: the Insulin Resistance Atherosclerosis Study (IRAS) Family Study. *Diabetes* 53:1866–1875
  68. Rich SS, Bowden DW, Haffner SM et al (2005) A genome scan for fasting insulin and fasting glucose identifies a quantitative trait locus on chromosome 17p: the insulin resistance atherosclerosis study (IRAS) family study. *Diabetes* 54:290–295
  69. Rotimi CN, Chen G, Adeyemo AA et al (2004) A genome-wide search for type 2 diabetes susceptibility genes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. *Diabetes* 53:838–841
  70. Sale MM, Freedman BI, Langefeld CD et al (2004) A genome-wide scan for type 2 diabetes in African-American families reveals evidence for a locus on chromosome 6q. *Diabetes* 53:830–837
  71. Shaw JT, Lovelock PK, Kesting JB et al (1998) Novel susceptibility gene for late-onset NIDDM is localized to human chromosome 12q. *Diabetes* 47:1793–1796
  72. Shtir C, Nagakawa IS, Duren WL et al (2007) Subsets of Finns with high HDL to total cholesterol ratio show evidence for linkage to type 2 diabetes on chromosome 6q. *Hum Hered* 63:17–25
  73. Silander K, Scott LJ, Valle TT et al (2004) A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* 53:821–829
  74. van Tilburg JH, Sandkuijl LA, Strengman E et al (2003) A genome-wide scan in type 2 diabetes mellitus provides independent replication of a susceptibility locus on 18p11 and suggests the existence of novel loci on 2q12 and 19q13. *J Clin Endocrinol Metab* 88:2223–2230

75. van Tilburg JH, Sandkuijl LA, Franke L et al (2003) Genome-wide screen in obese pedigrees with type 2 diabetes mellitus from a defined Dutch population. *Eur J Clin Invest* 33:1070–1074
76. Vionnet N, Hani El H, Dupont S 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
77. Voruganti VS, Goring HHH, Diego VP et al (2007) Genome-wide scan for serum ghrelin detects linkage on chromosome 1p36 in Hispanic children: results from the Viva La Familia Study. *Pediatr Res* 62:445–450
78. Watanabe RM, Ghosh S, Langefeld CD et al (2000) The Finland–United States Investigation of Non-insulin-dependent Diabetes Mellitus Genetics (FUSION) Study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. *Am J Hum Genet* 67:1186–1200
79. Wiltshire S, Hattersley AT, Hitman GA et al (2001) A genome-wide scan for loci predisposing to type 2 diabetes in a U.K. 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
80. Wiltshire S, Frayling TM, Groves CJ et al (2004) Evidence from a large U.K. family collection that genes influencing age of onset of type 2 diabetes map to chromosome 12p and to the MODY3/NIDDM2 locus on 12q24. *Diabetes* 53:855–860
81. Wiltshire S, Bell JT, Groves CJ et al (2006) Epistasis between type 2 diabetes susceptibility Loci on chromosomes 1q21-25 and 10q23-26 in northern Europeans. *Ann Hum Genet* 70:726–737
82. Xiang K, Wang Y, Zheng T et al (2004) Genome-wide search for type 2 diabetes/impaired glucose homeostasis susceptibility genes in the Chinese: significant linkage to chromosome 6q21-q23 and chromosome 1q21-q24. *Diabetes* 53:228–234
83. Zhao JY, Xiong MM, Huang W et al (2005) An autosomal genomic scan for loci linked to type 2 diabetes in northern Han Chinese. *J Mol Med* 83:209–215
84. Meigs JB, Shrader P, Sullivan LM et al (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 359:2208–2219
85. Lyssenko V, Jonsson A, Almgren P et al (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359:2220–2232
86. Kent WJ, Sugnet CW, Furey TS et al (2002) The human genome browser at UCSC. *Genome Res* 12:996–1006
87. Frayling TM (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 8:657–662
88. Yasuda K, Miyake K, Horikawa Y et al (2008) Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 40:1092–1097
89. Zeggini E, Scott LJ, Saxena R et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645
90. Sham P (1997) *Statistics in human genetics*. Arnold, London
91. Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
92. Risch NJ (2000) Searching for genetic determinants in the new millennium. *Nature* 405:847–856
93. Wiltshire S, Cardon LR, McCarthy MI (2002) Evaluating the results of genomewide linkage scans of complex traits by locus counting. *Am J Hum Genet* 71:1175–1182
94. Mills GW, Avery PJ, McCarthy MI et al (2004) Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to type 2 diabetes. *Diabetologia* 47:732–738
95. Hong Y, Weisnagel SJ, Rice T et al (2001) Familial resemblance for glucose and insulin metabolism indices derived from an intravenous glucose tolerance test in Blacks and Whites of the HERITAGE Family Study. *Clin Genet* 60:22–30
96. Li JKY, Ng MCY, So WY et al (2006) Phenotypic and genetic clustering of diabetes and metabolic syndrome in Chinese families with type 2 diabetes mellitus. *Diabetes/Metab Rev* 22:46–52
97. Lyssenko V, Almgren P, Anevski D et al (2005) Predictors of and longitudinal changes in insulin sensitivity and secretion preceding onset of type 2 diabetes. *Diabetes* 54:166–174
98. Hara H, Egusa G, Yamakido M (1996) Incidence of non-insulin-dependent diabetes mellitus and its risk factors in Japanese-Americans living in Hawaii and Los Angeles. *Diabet Med* 13: S133–S142
99. Hanson RL, Pratley RE, Bogardus C et al (2000) Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *Am J Epidemiol* 151:190–198
100. Elbein SC, Hasstedt SJ, Wegner K, Kahn SE (1999) Heritability of pancreatic beta-cell function among nondiabetic members of Caucasian familial type 2 diabetic kindreds. *J Clin Endocrinol Metab* 84:1398–1403
101. Bergman RN, Zaccaro DJ, Watanabe RM et al (2003) Minimal model-based insulin sensitivity has greater heritability and a different genetic basis than homeostasis model assessment or fasting insulin. *Diabetes* 52:2168–2174
102. Rasmussen-Torvik LJ, Pankow JS, Jacobs DR et al (2007) Heritability and genetic correlations of insulin sensitivity measured by the euglycaemic clamp. *Diabet Med* 24:1286–1289
103. Guan W, Pluzhnikov A, Cox NJ, Boehnke M, International Type 2 Diabetes Linkage Analysis C (2008) Meta-analysis of 23 type 2 diabetes linkage studies from the International Type 2 Diabetes Linkage Analysis Consortium. *Hum Hered* 66:35–49
104. Kong A, Gudbjartsson DF, Sainz J et al (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241–247
105. Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439–454
106. Cordell HJ (2001) Sample size requirements to control for stochastic variation in magnitude and location of allele-sharing linkage statistics in affected sibling pairs. *Ann Hum Genet* 65:491–502
107. Papachristou C, Lin S (2006) A comparison of methods for intermediate fine mapping. *Genet Epidemiol* 30:677–689
108. Atwood LD, Heard-Costa NL (2003) Limits of fine-mapping a quantitative trait. *Genet Epidemiol* 24:99–106
109. Hsueh W-C, Goring HH, Blangero J, Mitchell BD (2001) Replication of linkage to quantitative trait loci: variation in location and magnitude of the LOD score. *Genet Epidemiol* 21 (Suppl 1):S473–S478
110. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999) Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* 65:876–884
111. Kruglyak L, Lander ES (1996) Limits on fine mapping of complex traits. *Am J Hum Genet* 58:1092–1093
112. Hauser ER, Boehnke M, Guo SW, Risch N (1996) Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol* 13:117–137
113. Haiman CA, Patterson N, Freedman ML et al (2007) Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39:638–644
114. American Diabetes Association (2008) Standards of medical care in diabetes 2008. *Diabetes Care* 31:S12–S54

115. Lewis JP, Palmer ND, Hicks PJ et al (2008) Association analysis in African Americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. *Diabetes* 57:2220–2225
116. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273:1516–1517
117. Florez JC, Hirschhorn J, Altshuler D (2003) The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annu Rev Genom Hum Genet* 4:257–291
118. Hsueh W-C, Mitchell BD, Aburomia R et al (2000) Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care* 23:595–601
119. Shaw JT, Purdie DM, Neil HA, Levy JC, Turner RC (1999) The relative risks of hyperglycaemia, obesity and dyslipidaemia in the relatives of patients with type II diabetes mellitus. *Diabetologia* 42:24–27
120. Weijnen CF, Rich SS, Meigs JB, Krolewski AS, Warram JH (2002) Risk of diabetes in siblings of index cases with type 2 diabetes: implications for genetic studies. *Diabet Med* 19:41–50
121. Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. *Science* 322:881–888
122. Holmkvist J, Almgren P, Lyssenko V et al (2008) Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes. *Diabetes* 57:1738–1744
123. Wheeler DA, Srinivasan M, Egholm M et al (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452:872–876
124. International HapMap C, Frazer KA, Ballinger DG et al (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861
125. International HapMap C (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
126. Bellanne-Chantelot C, Carette C, Riveline J-P et al (2008) The type and the position of HNF1A mutation modulate age at diagnosis of diabetes in patients with maturity-onset diabetes of the young (MODY)-3. *Diabetes* 57:503–508
127. Zondervan KT, Cardon LR (2004) The complex interplay among factors that influence allelic association. *Nat Rev Genet* 5:89–100
128. Blangero J (2004) Localization and identification of human quantitative trait loci: king harvest has surely come. *Curr Opin Genet Dev* 14:233–240
129. McCarthy MI, Abecasis GR, Cardon LR et al (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9:356–369
130. Kidd JM, Cooper GM, Donahue WF et al (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64
131. Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. *Nature* 456:728–731