Family studies of non-insulin-dependent diabetes mellitus in South Indians

M. I. McCarthy[1], G. A. Hitman[1], D. C. Shields[2], N. E. Morton[2], C. Snehalatha[3], V. Mohan[3], A. Ramachandran[3], M. Viswanathan[3]

[1] Cellular Mechanisms Research Unit, Medical Unit, London Hospital Medical College, London, UK
[2] Genetic Epidemiology Group, Princess Anne Hospital, Southampton, UK
[3] Diabetes Research Center, M. V. Hospital for Diabetes, Royapuram, Madras, India

Summary Though a genetic basis for non-insulin-dependent diabetes mellitus (NIDDM) is clear, the likely mode of inheritance is not known. The segregation of NIDDM was studied in 64 nuclear South Indian pedigrees (449 individuals) ascertained through an affected proband having both parents and more than 1 sibling alive and available for oral glucose tolerance testing. A high proportion of parents were found to be of abnormal glucose tolerance [89 of 128 (70 %) diabetic and 11 of 128 (9 %) impaired]. Complex segregation analysis was performed using (1) POINTER which implements the mixed model and distinguishes major gene, multifactorial and non-transmitted environmental contributions to affection and (2) COMDS which implements an oligogenic model with major gene, modifier gene and environmental contributions to a) affection and b) diathesis (an ordered polychotomy amongst non-affected family members, based on 2-h plasma glucose level). Using POINTER, there was no formal support for a major gene and the most parsimonious solutions were achieved with multifactorial models. Using COMDS, we found i) significant improvements in models when information on glucose levels in non-diabetic family members (diathesis) was included, ii) support for segregation of a diallelic gene as well as background familial resemblance, and iii) under the best-supported model, this diallelic locus featured incomplete dominance ($d = 0.8$) and a disease-predisposing allele frequency of 14 %. In South Indians, segregation of NIDDM is inadequately described by simple major gene models: more complex models provide more satisfactory descriptions. This finding, if applicable in other populations, has important implications for the search for diabetes-susceptibility genes. [Diabetologia (1994) 37: 1221–1230]

Key words Non-insulin-dependent diabetes mellitus, segregation analysis, South India, family studies, polygenes, major gene, heritability, diathesis.

Twin, family and population studies provide strong evidence for an important genetic contribution to the development of non-insulin-dependent diabetes mel-

Corresponding author: Dr. M. McCarthy, Medical Unit, The Royal London Hospital, London E1 1BB, UK

Abbreviations: MODY, maturity onset diabetes of the young; DRC-MVH, Diabetes Research Centre-MV Hospital for Diabetes; OGTT, oral glucose tolerance test; AIC, Akaike Information Content; IGT, impaired glucose tolerance; NIDDM, non-insulin-dependent diabetes mellitus; IDDM insulin-dependent diabetes mellitus

litus (NIDDM) but do not imply a model for that genetic predisposition [1]. On theoretical grounds, the development of glucose intolerance in an individual might result from the effect of a single major gene, from the combined action of a number of minor genes, or a combination of the two. Furthermore, different individuals having NIDDM within a single population may be expressing different diabetes-susceptibility genes, and the most influential genes determining glucose tolerance may differ between ethnic groups. Such genetic heterogeneity would explain the considerable phenotypic heterogeneity evident within NIDDM. No gene has yet been identified which explains more than a small proportion of the inherited susceptibility to NIDDM [1].

Evidence that highly penetrant, single major genes may be responsible for glucose intolerance derives principally from studies of populations at extreme risk of diabetes (e.g. Pima Indians and Nauruans) in which a bimodal distribution of glucose levels suggests a major gene effect [2–4]. In addition, maturity-onset diabetes of the young (MODY), a subgroup of NIDDM, is defined on the basis of an early presentation of diabetes and autosomal dominant inheritance [5]: recently, it has been shown that single inherited mutations can explain inheritance of glucose intolerance within such pedigrees [6–8].

In contrast, studies of more typical late-onset NIDDM in populations at lower risk of diabetes have favoured more complex models of inheritance [9–17], invoking genes of incomplete penetrance or a multifactorial (polygenic) basis for the disease or both. If genetic susceptibility to NIDDM in most populations is not described by simple single-gene models, this has crucial implications for the practical application and likely success of the various strategies available in the search for diabetes-susceptibility genes [1, 18]. The power of population-association studies, of linkage analyses and of molecular scanning are all reduced by genetic heterogeneity and further by the interaction of genetic factors with environmental influences.

We have used the techniques of complex segregation analysis to study the segregation of NIDDM in a collection of nuclear pedigrees from South India, in order to establish whether any evidence could be adduced for a major gene determining the diabetic phenotype. Two complementary methods of segregation analysis were used. The program POINTER [19, 20] implements the mixed model [21, 22] and has been widely used in segregation analysis of complex diseases. The program COMDS has recently been developed to implement a two-gene (oligogenic) model of inheritance and to permit combined linkage and segregation analysis [23]: furthermore it allows for a polychotomy of degrees of diathesis within non-affected individuals and of severity within affected individuals [24].

Subjects, materials and methods

Subjects

Sixty-four nuclear pedigrees were recruited from the outpatient department of the Diabetes Research Centre, M.V. Hospital for Diabetes, Madras (DRC-MVH) between September 1990 and April 1993. The families of Dravidian (South Indian) NIDDM subjects were invited to attend for study if both parents and at least one sibling of the proband were alive and available. Recruitment was thus initiated without explicit regard to the family history of diabetes. NIDDM was defined by an insidious onset of disease without ketosis and follow-up of more than 2 years since diagnosis: if patients were on insu-

lin, this had not commenced within 5 years of diagnosis. Insulin-dependent diabetes (IDDM) and fibrocalculous pancreatic diabetes mellitus contribute to a small minority of diabetes in this population [25] and were defined and excluded as previously described [26].

Assessment

All available members of participating families attended the DRC-MVH: if of unknown glucose tolerance, a 75 g OGTT was performed with interpretation according to World Health Organization criteria [27]. Glucose measurements were performed on venous plasma by the glucose oxidase peroxidase method (Hitachi 704 autoanalyser; Boehringer Mannheim, Mannheim, Germany). Subjects found to be diabetic were considered to have NIDDM unless they met diagnostic criteria for IDDM or fibrocalculous diabetes at diagnosis or on follow-up. No family member was pregnant or on medication predisposing to hyperglycaemia when tested. If a subject had a pre-existing diagnosis of NIDDM and was receiving oral hypoglycaemic agents or insulin (provided > 5 years between diagnosis and commencement of insulin) the existing diagnosis was accepted if a diagnostic glucose measurement had been recorded in the past. Information was collected on antecedent generations to identify those nuclear pedigrees in which the mother and father were consanguineous.

It was not possible to assess all living siblings due to distance or infirmity. To avoid a bias towards affection in siblings, all siblings not attending the DRC-MVH were regarded as of unknown glucose tolerance for further analysis whether or not reported as diabetic by relatives. Ethical approval was obtained from the Ethical Committee of the M.V. Hospital for Diabetes, Madras. All individuals gave informed consent for testing.

Recruitment

Of the families meeting the inclusion criteria approximately 30% agreed to participate. The families reported here represent those most fully characterised and comprise 449 individuals (64 probands, 128 parents and 257 siblings: Table 1 and Fig.1). Either due to death or residence distant from Madras 131 siblings currently remain untested: information provided by their siblings indicates that they do not differ significantly in age or sex from those siblings who attended for testing (Table 1). In nine of the pedigrees the parents were related as first-cousins and in two as uncle and niece.

To gauge any bias associated with incomplete participation, reported family histories were obtained from a sample of 136 probands who met all the criteria for inclusion in the study but whose families did not participate.

Confirmation of family relationships

Ten ml of blood was extracted from each family member and stored at –20 °C prior to and during transport to the UK. Paternity of family members was established by typing at several highly-polymorphic loci: HLA-DQα (heterozygosity = 0.81) (Amplitype; Cetus Corporation, Emeryville, Calif., USA), tyrosine hydroxylase (heterozygosity = 0.77)[28] and two microsatellites straddling the glucokinase gene (combined heterozygosity = 0.74)[29,30].

Table 1. Characteristics of family members. Age, age of diagnosis and BMI are given as mean (SD)

	n	Sex (male: female)	Age (years)	Age of diabetes diagnosis (years)	BMI (kg/m²)
Fathers	64	–	65.8 (7.2)	53.6 (11.2)	24.0 (3.5)
Mothers	64	–	58.4 (6.6)	50.3 (8.5)	25.4 (3.9)
Probands	64	41 : 23	36.8 (6.5)	33.3 (5.8)	26.3 (4.1)
Siblings	257	132 : 125	33.8 (7.7)	37.0 (8.6)[a]	25.3 (3.8)[a]
- tested	126	60 : 66[b]	34.5 (7.6)[b]	–	–
- untested	131	72 : 59[b]	33.0 (7.8)[b]	–	–

[a] tested family members only. [b] Tested vs untested, not significant ($p > 0.05$)
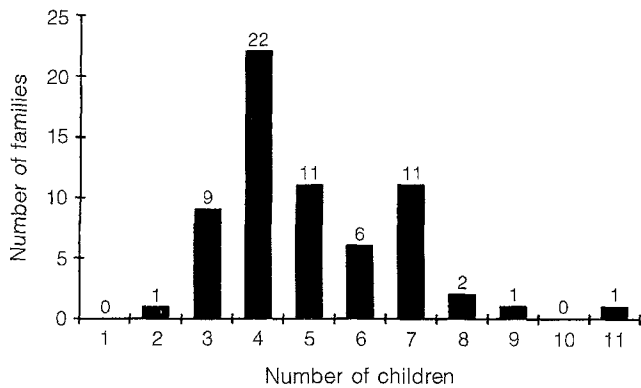

Fig. 1. Distribution of family size in 64 families studied

Segregation analysis

Segregation analysis was performed at the Human Genetics Centre in Southampton, UK using the programs POINTER and COMDS (available from the authors). Initial analysis was performed using the mixed model [21, 22] as implemented by POINTER [19, 20]. The mixed model divides variation in the liability to affection into three additive, independent components: i) a diallelic major gene, ii) a multifactorial, transmissible effect, and iii) a random, non-transmitted (environmental) component. The multifactorial component encompasses genetic effects not attributable to the major gene (i.e. poly-

genes) and culturally-transmitted (shared family environment) effects. Parameters of this model are shown in Table 2. Affection corresponds to a liability to affection exceeding a threshold which varies with age, sex and other covariates of affection: these situational variables are used to define a series of 'liability classes' each associated with a particular threshold.

For complex diseases like NIDDM, dichotomisation into affected and non-affected may not be optimal given the arbitrary nature of the boundary and the loss of information. The program COMDS implements an oligogenic model [23] and introduces an ordered polychotomy of gradations of normality (diathesis classes) and of affection (severity classes) [24]. Liability to affection is determined by the joint, additive contributions of i) a diallelic major gene, ii) a diallelic minor (modifier) gene and iii) a random, non-transmitted (environmental) component [23]. Diathesis and severity classes are considered determined by underlying traits of diathesis and severity, the variance of each of which also has major gene, minor gene and random components. Measures of diathesis and severity are covariance-adjusted, so that they reflect genotype and not other covariates of affection i.e. those used to determine liability classes. Classes of diathesis (or severity) are defined within the distribution of the quantitative trait encoding diathesis (or severity) on a percentile basis: COMDS therefore makes few distributional assumptions [24] in contrast to quantitative trait analyses with POINTER [20]. The parameters of this model are given in Table 2. The parameters, b and b_m are proportionality constants specifying the relationship between liability to affection and diathesis for the major and minor gene respectively: s and s_m are analogous parameters for

Table 2. Parameters used in segregation analysis with POINTER and COMDS. For further details see text

	Parameter	Meaning
POINTER		
Major gene	d	degree of dominance e. g. $d = 1$ (dominant); $d = 0$ (recessive); $d = 0.5$ (additive)
	t	displacement (along liability axis) between homozygotes of the major gene
	q	frequency of disease-causing allele
Multifactorial	h	childhood heritability
	z	ratio of adult:child heritability (hence hz = adult heritability)
COMDS		
Major gene	d	degree of dominance
	t	displacement (along liability axis) between homozygotes
	q	frequency of disease-causing allele
	b	scaling factor for diathesis effects (relative to t)
	s	scaling factor for severity effects (relative to t)
Minor gene	d_m	degree of dominance
	t_m	displacement (along liability axis) between homozygotes
	q_m	frequency of disease-causing allele
	b_m	scaling factor for diathesis effects (relative to t_m)
	s_m	scaling factor for severity effects (relative to t_m)

severity. In each case the parameter scales the genotypic displacement between diathesis (or severity) classes to that between the two homozygous genotypes under the liability model.

Under COMDS, transmissible effects not partitioned to the major gene are assigned to a modifying diallelic gene rather than to a multifactorial transmissible element (as in POINTER). However, a polygenic background may be approximated within COMDS by restricting gene parameters to $q = d = 0.5$ (or $q_m = d_m = 0.5$) ('pseudopolygenic' locus) [24].

Both programs perform iteration to define solutions of maximum likelihood under a series of models each (through restriction of relevant parameters) reflecting a particular hypothesis: inferences are made by standard likelihood ratio tests. For example, evidence for a major gene effect is evaluated by comparison of the likelihoods (L) of the solutions obtained with the frequency of the deleterious allele of the major gene (q) either unrestricted or constrained to zero. For such 'nested' hypotheses, twice the logarithm of the likelihood ratio between the restricted and unrestricted models (i.e. the difference between the $-2\ln(L) + C$ terms) is distributed asymptotically as a chi square function with degrees of freedom equal to the number of restricted parameters. Since (L) varies within the limits $0 < (L) < 1$, the more likely of two models is that with the lower value of the $-2\ln(L) + C$ term. For non-nested models, the Akaike information criterion (AIC: $-2\ln(L) + C$ plus twice the number of estimated parameters) [31] was used to compare solutions for goodness-of-fit and parsimony: the 'best' model is that associated with the smallest AIC.

Assumptions for these analyses

Given the large population served by the DRC-MVH (70,000 patients) probability of ascertainment was set at the arbitrarily low figure of 0.001 since analyses are insensitive to moderate errors in this parameter [24, 32]. Nine liability classes were defined by sex, age and urban/rural residence, based on the prevalence of NIDDM found in more than 2500 South Indian subjects recently investigated in epidemiological surveys [33, 34]. The prevalence of diabetes in urban South Indian populations is moderately high with approximately 30 % of those older than 55 years having diabetes and 50 % diabetes or impaired glucose tolerance (IGT) [33, 34]: in rural populations the peak prevalence of diabetes is 9 % [34].

In non-affected family members, the 2-h plasma glucose level (after OGTT) was used as the basis for diathesis, after covariance-adjustment for situational variables defining liability to affection. Thus, for each non-affected (i.e. non-diabetic) member of the population control (i.e. survey) data [33,34], the 2-h glucose level (logarithmically-transformed) was converted to a Z-score using the mean and standard deviation obtained for all non-affected members of the same liability class: equivalent transformations were then applied in non-affected family members. Whilst in the control data, the distribution of the Z-scores had, by definition, a mean of 0 and SD of 1, in the family members, there was a shift in the 'mean' of the (normal) distribution to 0.57 (SD 0.89). Each non-affected family member was assigned to one of seven diathesis classes defined by the 20th, 40th, 60th, 80th, 90th and 95th centiles of the standardised glucose in the population control data. These percentile divisions were chosen to reflect the fact that the upper part of the distribution would be more informative.

No suitable variable for the definition of severity amongst affected individuals was found. Glucose values were confound-
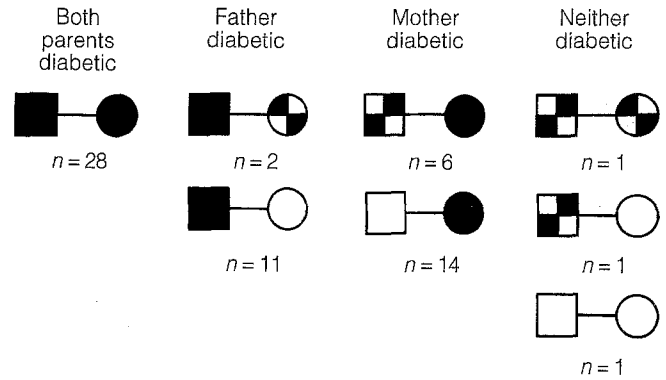

Fig. 2. Parental mating types for the 64 families studied. Fathers are shown as squares, mothers as circles. Filled symbols denote diabetes, chequered symbols impaired glucose tolerance and open symbols normal glucose tolerance

ed by the effects of treatment and age of onset was highly correlated with age of the subjects (and therefore with liability to affection). Thus, in all analyses severity effects were set to zero.

All analyses were conducted under two conditioning paradigms: a) conditioning on ascertainment through probands (joint likelihoods) and, b) conditioning on the parental phenotypes as well as on ascertainment through probands (conditional likelihoods). The first is more informative but is potentially more susceptible to ascertainment bias: the latter is the more conservative approach [24,35].

Statistical analysis

Analyses of the anthropometric and biochemical characteristics of family members were performed with standard parametric tests after logarithmic transformation of variables where appropriate (SPSS/PC+, version 4.0). Dichotomous variables were analysed using StatXact-Turbo (Cytel Corporation, Cambridge, Mass., USA).

Results

Prevalence of diabetes in first-degree relatives

Of 128 parents of diabetic probands, 89 (69.5 %) were diabetic and 11 (8.6 %) had IGT. The parental mating types are shown in Figure 2. In 28 of the families (44 %) both parents were diabetic and in only three (5 %) was neither diabetic. Only one proband had two parents of normal glucose tolerance. Thirteen of the parents with diabetes (14.6 %) and all with IGT were diagnosed during the study OGTT: the remainder had been previously-diagnosed. Amongst siblings of probands who were tested, 23 (18.3 %) were themselves diabetic, 16 (12.7 %) had IGT and 87 (69.0 %) had normal glucose tolerance.

Table 3. Characteristics of probands according to the number and sex of diabetic parents

Parents		n	Age (years)	Age at diabetes diagnosis (years) (n)	BMI (kg/m²)	Weight (kg)	Height (m)
All offspring							
Both diabetic		145	33.9 (7.7)	35.1 (8.3) (39)	25.9 (3.9)	63.5 (11.3)	1.56 (0.08)
One diabetic		166	34.7 (7.5)	33.4 (5.4) (45)	25.6 (4.0)	63.4 (11.2)	1.57 (0.08)
	– mother	97	34.4 (7.2)	32.7 (5.2) (25)	25.4 (4.2)	62.9 (11.9)	1.57 (0.08)
	– father	69	35.2 (8.0)	34.4 (5.6) (20)	26.0 (3.7)	64.1 (10.2)	1.56 (0.08)
None diabetic		10	35.6 (6.5)	35.3 (4.2) (3)	22.2 (2.8)	59.5 (9.1)	1.63 (0.08)
Probands only							
Both diabetic		28	36.1 (6.2)	33.4 (6.5)	27.0 (4.2)	67.8 (11.3)	1.59 (0.09)
One diabetic		33	37.4 (6.9)	32.9 (5.3)	25.9 (4.0)	64.0 (11.5)	1.57 (0.08)
	– mother	20	36.8 (7.0)	32.6 (5.6)	25.4 (4.6)	63.1 (13.1)	1.57 (0.08)
	– father	13	38.4 (6.9)	33.4 (5.1)	26.5 (3.0)	65.4 (8.8)	1.57 (0.06)
None diabetic		3	36.7 (4.6)	35.3 (4.2)	23.5 (3.6)	63.3 (6.2)	1.65 (0.07)

All data as mean (SD). BMI, weight and height only available for those offspring tested ($n = 190$). All comparisons not significant at $p = 0.05$

Table 4. Segregation analysis of families with POINTER (conditioning on ascertainment through the proband)

model	no	d	t	q	h	z	-2ln(L) + C	AIC
No familial resemblance								
Null	1	–	–	–	–	–	168.4	168.4
Single-gene models								
Dominant	2	(1.0)	2.26	0.06	–	–	14.9	18.9
Recessive	3	(0.0)	2.56	0.38	–	–	37.8	41.8
Additive	4	(0.5)	3.56	0.18	–	–	7.6	11.6
General	5	0.63	3.24	0.08	–	–	3.4	9.4
Multifactorial models								
z = 1.0	6	–	–	–	0.99	(1.0)	0.4	2.4
z unrestricted	7	–	–	–	0.95	1.06	0.2	4.2
Mixed models								
Additive major	8	(0.5)	2.74	0.09	0.59	(1.0)	2.8	8.8
General solution	9	0.71	2.46	0.12	0.32	(1.0)	0	8.0

Bracketing of parameter values implies that the parameter was fixed at that value and not iterated upon.

Characteristics of family members

Affected (diabetic, $n = 89$) and non-affected (normal and IGT, $n = 39$) parents did not differ for measures including age (diabetic 61.6 (8.0) vs non-diabetic 63.2 (7.4) years, $p = 0.31$), weight (59.1 (10.3) vs 58.3 (8.5)kg, $p = 0.70$), height (1.54 (0.08) vs 1.55 (0.09)m, $p = 0.48$) and BMI (24.9 (3.7) vs 24.3 (3.9)kg/m², $p = 0.46$) [all figures as mean (SD)]. The excess of maternal diabetes over paternal was non-significant (48 vs 41, likelihood ratio $\chi^2 = 1.81$, exact $p = 0.19$).

To assess whether the high parental prevalence of diabetes was related to an early age of onset of diabetes in the children, proband and sibling characteristics were studied in relation to the number and sex of their diabetic parents. There was no association between the 'density' of parental diabetes (none, one or both parents diabetic) and the following variables amongst offspring: age of onset of diabetes, age, height, weight, BMI (Table 3) and prevalence of diabetes in the offspring [39 of 85 (45.9 %) tested off-

spring of 'both parents diabetic' themselves diabetic: 45 of 98 (45.9 %) of 'one parent diabetic': 3 of 7 (42.9 %) of 'neither parent diabetic'].

Complex segregation analysis using POINTER

Results obtained using POINTER (conditioning on ascertainment through the probands) are shown in Table 4. A model admitting no familial resemblance was firmly rejected (no. 9 vs no. 1 χ^2 (4) = 168.4, $p < 0.0001$). In comparison with the general solution under the mixed model (no. 9), dominant and recessive single-gene models were rejected (vs no. 2, χ^2 (2) = 14.9, $p = 0.0006$; vs no. 3, χ^2 (2) = 37.8, $p < 0.0001$), but it was not possible to reject the general single-gene model (no. 9 vs no. 5, χ^2(1) = 3.4, $p = 0.07$). This is equivalent to a failure to reject the null hypothesis of no multifactorial component ($h = 0$). Similarly, the hypothesis of no major gene ($q = 0$) could not be rejected (no. 9 vs no. 6, χ^2 (3) = 0.4, $p = 0.94$). Using

Table 5. Segregation analysis of oligogenic models under COMDS, conditioning on ascertainment through the proband

Model	no.	d	t	q	b	d_m	t_m	q_m	b_m	-2ln(L) + C	AIC
No familial resemblance											
Null	1	–	–	–	–	–	–	–	–	230.0	230.0
'Best' model	2	(0.5)	2.20	(0.5)	(1)	0.83	1.65	0.14	(0)	0	8.0
Changes in major gene											
Dominant major	3	(1)	1.49	0.15	(1)	0.69	2.60	0.10	(0)	18.1	28.1
Recessive major	4	(0)	1.32	0.68	(1)	0.65	2.24	0.12	(0)	17.5	27.5
Unrestricted major gene	5	0.55	2.20	0.50	(1)	0.79	1.67	0.14	(0)	-0.2	11.8
Changes in minor gene											
Recessive minor	6	(0.5)	2.19	(0.5)	(1)	(0)	1.40	0.51	(0)	17.7	23.7
Dominant minor	7	(0.5)	2.19	(0.5)	(1)	(1)	1.43	0.12	(0)	5.7	11.7
Additive minor	8	(0.5)	2.25	(0.5)	(1)	(0.5)	2.60	0.11	(0)	4.9	10.9
No minor gene	9	(0.5)	2.13	(0.5)	(1)	–	–	–	–	65.7	67.7
General single-gene ($b = 1$)	10	0.57	2.49	0.08	(1)	–	–	–	–	38.4	44.4
General single-gene	11	0.67	2.90	0.19	0.5	–	–	–	–	16.4	24.4
Changes in diathesis parameters											
No diathesis effect ($b = 0$)	12	(0.5)	1.35	(0.5)	(0)	0.67	2.74	0.10	(0)	56.3	64.3
Diathesis effect ($b = 0.5$)	13	(0.5)	2.25	(0.5)	(0.5)	0.64	1.61	0.20	(0)	25.3	33.3
Unrestricted diathesis *(b)*	14	(0.5)	2.20	(0.5)	1.0	0.83	1.65	0.14	(0)	0	10.0
Diathesis effect ($b_m = 1$)	15	(0.5)	1.34	(0.5)	(1)	0.62	2.19	0.17	(1)	31.3	39.3
Unrestricted diathesis (b_m)	16	(0.5)	2.18	(0.5)	(1)	0.82	1.69	0.14	0.11	-0.6	9.4

Bracketing of parameter values implies that the parameter was fixed at that value and not iterated upon
Likelihood values have been rescaled to facilitate comparison with the best supported model (no. 2)

POINTER, therefore, no formal support for the presence of a major gene could be found: indeed solutions admitting no major gene effect (nos. 6, 7) were judged by AIC to provide the most parsimonious descriptions of the data. These multifactorial models did however generate solutions of high heritability (0.99) and there were problems with convergence of the solutions as h approached 1. There was no evidence for an intergenerational difference in heritability of the multifactorial component [i.e. z did not differ significantly from 1.0 (no. 7 vs no. 6, χ^2 (1) = 0.3, $p = 0.58$)]. In model 9, variance in the liability to affection was partitioned 55 % to the major gene and 32 % to the multifactorial component.

There were no substantive changes in the results obtained when analysis was repeated excluding the 11 consanguineous pedigrees (data not shown). When POINTER analysis was performed under the alternative, less informative conditioning paradigm (using conditional likelihoods), it was not possible to discriminate between the models except to reject the model admitting no familial resemblance (data not shown).

Complex segregation analysis using COMDS

With COMDS, solutions obtained for single-gene models were identical to those with POINTER when diathesis effects were set to zero. Under all single-gene models, inclusion of diathesis effects ($b = 1$, or b unrestricted) produced significant improvements with an additive single-gene ($d = 0.67$, $q = 0.19$: Table 5, no. 11) the favoured solution.

To deal effectively with the many parameters of the two-locus models and to ensure that global maxima were found, the likelihood space was explored using all combinations of major and modifier gene models (dominant, additive, recessive, pseudopolygenic and unrestricted) to determine that subset providing solutions of greatest likelihood: diathesis effects (b, b_m) were varied to values 0, 0.5 and 1.0 and unrestricted (Table 5). The best-fitting model judged by AIC (no. 2) featured a pseudopolygenic 'major' gene influencing diathesis as well as affection ($b = 1$) combined with a modifier gene, of intermediate dominance ($d_m = 0.83$) with a deleterious allele frequency of 14 % and no effect on diathesis ($b_m = 0$). The terms 'major' and 'modifier' are arbitrary in that equivalent solutions were obtained with the loci reversed (i.e. $q_m = d_m = 0.5$): the term 'major' has been reserved for the pseudopolygenic locus here as it was both megaphenic ($t > t_m$) and responsible for the majority of the variance in the liability to affection (60.3 % vs 37.3 % for the modifier gene).

The other models shown in Table 5 emphasise the effects of deviations from model 2. Changes in the major gene parameters were associated with less satisfactory solutions (nos. 3 and 4 vs no. 2, AIC), and iteration on d and q failed to lead to further improvement [no. 5 vs no. 2, χ^2 (2) = 0.2, $p = 0.90$]. Similar

considerations applied to changes in the modifier gene [no. 6 vs no. 2 χ^2 (1) = 17.7, $p < 0.0001$; no. 7 vs no. 2, χ^2 (1) = 5.7, $p = 0.02$; no. 8 vs no. 2, χ^2 (1) = 4.9, $p = 0.03$]. Most notably, and in contrast to the POINTER analysis, these analyses provided support for the segregation of a single diallelic gene (the modifier gene) whilst allowing for background familial resemblance (the pseudopolygenic locus) since the hypothesis of no single-gene ($q_m = 0$, on a pseudopolygenic background) was firmly rejected (no. 9 vs no. 2, χ^2 (3) = 65.7, $p < 0.0001$).

Although single-gene solutions favoured $b = 0.5$ (no. 11), iteration on b in the two-gene models tended towards $b = 1$ [identity of no. 14 and 2; no. 14 vs no. 12, χ^2(1) = 56.3, $p < 0.0001$; no. 14 vs no. 13, χ^2(1) = 25.3, $p < 0.0001$]. Similarly, models with $b_m \neq 0$ were associated with a reduction in likelihood [no. 15 vs no. 2, AIC; no. 16 vs no. 2, χ^2 (1) = 0.6, $p = 0.44$].

When COMDS analysis was repeated under conditional likelihoods, single-gene models again favoured an additive major gene with a strong effect on diathesis ($b = 1$). Addition of a second locus significantly improved the likelihoods of all major gene models. The characteristics of the best-fitting oligogenic model were less easy to specify under this less informative conditioning paradigm, but a pseudopolygenic 'major' gene and a dominant 'modifier' gene explained the data most economically, although in this case the fit was slightly improved with $b_m = 1$ rather than $b_m = 0$. Analyses performed after excluding the 11 consanguineous families had no effect on the overall relationship of the models nor on the interpretations therefrom (data not shown).

Reported family histories in non-participating families

Family histories obtained from 136 probands meeting the inclusion criteria whose families did not assent revealed that 156 parents (57.4 %) were reported to be diabetic (13 pedigrees with neither parent diabetic, 52 with father alone, 38 with mother alone and 33 with both parents diabetic). There was no difference in the parental prevalence of previously-diagnosed NIDDM between the participating and non-participating families ($\chi^2 = 0.15$, exact $p = 0.71$) indicating that motivational factors related to a high family awareness of diabetes were not influencing recruitment.

Discussion

In this sample of South Indian pedigrees, a high frequency of NIDDM was found amongst parents of diabetic probands. Since probands were ascertained blind to family history these findings confirm the strong familial aggregation of diabetes previously reported in South India [12, 36]. The high parental prevalence should be interpreted in the context of the high rates of diabetes found amongst middle-aged and elderly urban South Indians [33, 34]. Such family structures as these are difficult to reconcile with simple modes of inheritance (dominant or recessive major genes) unless high gene frequency and incomplete penetrance are accepted.

Complex segregation analysis confirmed the inadequacy of simple models of inheritance to describe the segregation of glucose intolerance in this population. Using POINTER, it was not possible to reject either the hypothesis of no major gene or that of no multifactorial effect when such restricted models were compared with the general (mixed) model. In fact, multifactorial models provided the most parsimonious solutions in this analysis. Using COMDS, oligogenic models were favoured over simple models, and the best-supported model indicated support for a diallelic locus acting on a pseudopolygenic background: this solution represents an approximation to the mixed model implemented under POINTER.

Any comparison of analyses by the two programs must consider the inherent differences between them: COMDS does not explicitly provide for a multifactorial element but does include information on diathesis. However, it is reassuring to note that the parameters of the diallelic locus in the best-supported COMDS model are similar to those obtained in the mixed model solutions. The major difference evident is in the displacement (t/t_m) which is greater in the POINTER solution: this probably reflects the relative platykurtosis of the pseudopolygenic locus (compared to the multifactorial element in POINTER). Under COMDS it was possible to advance support for the existence of a segregating diallelic locus in addition to background family resemblance: the added power of COMDS to discriminate between hypotheses results from the extra information in diathesis. In general, it appears from both analyses that more complex solutions are favoured over simple single-gene solutions and that recessive modes of inheritance are rejected for the most significant contributing loci. It is not possible to state from these analyses the total number of genes that contribute to the development of NIDDM in this population, but it is at least two.

The diallelic locus identified by COMDS has plausible parameters for use within this population in tests of linkage to candidate loci for NIDDM or as a basis for combined segregation and linkage analysis using COMDS [23]. In this regard, the precise parameters of other genetic determinants (subsumed here within the pseudopolygenic locus) do not require precise delineation: they provide a mechanism whereby transmitted effects due to genes other than that of current interest may be accommodated [23, 37].

There are considerable methodological difficulties with segregation analysis performed in complex diseases such as NIDDM [38,39]: late onset of disease, premature mortality, absence of a biochemical marker of future diabetes, high prevalence of undiagnosed glucose intolerance (which mandates testing of relatives) and uncertainties over the appropriate diagnostic criteria all contribute to the difficulty with which suitable multigenerational pedigrees can be ascertained and characterised. Segregation analyses also require ascertainment of families within a defined protocol and appropriate ascertainment corrections: bias which leads to enrichment of particular segregation patterns may yield misleading conclusions.

Ascertainment in this study was made via an affected proband, irrespective of the family history of diabetes, in an effort to avoid serious ascertainment bias. Nevertheless, it is important to recognise that certain, conflicting biases are possible in any study of this condition. Potential sources of bias include (a) selection for living parents (b) motivational effects and (c) selection of early onset probands. The first could lead to preferential ascertainment of non-diabetic parents (due to diabetes-associated mortality), the others to the opposite effect. Evidence that these biases are not significant is provided, respectively, by (a) the high parental prevalence of diabetes observed, (b) the survey of reported family histories in non-participating families and (c) the absence of any relationship between the 'density' of parental diabetes and age of diagnosis in probands and their siblings. Finally, analyses conducted using the more conservative conditioning paradigm were broadly similar to those obtained conditioning on ascertainment through the proband: this indicates that any minor biases are unlikely to invalidate our general conclusions.

One further possible bias lies in the consanguinity evident in some pedigrees. The inbreeding coefficient for the study pedigrees was 0.0127 which is low compared to other South Asian populations [40] and reflects the predominantly urban study population. Inclusion of consanguineous pedigrees did not prejudice our results, since similar findings were obtained when analyses were repeated after excluding them. Furthermore, the effects of consanguinity would be expected to be most influential with rare, recessively-inherited diseases and it is clear from these analyses that there is no evidence for recessive loci contributing to NIDDM in South India.

The earliest attempts to identify genetic models for the inheritance of diabetes predated identification of the major subtypes: no consensus emerged [41] with advocates for all possible modes of inheritance including multifactorial [9]. Inadequate allowance for age-related penetrance and use of different measures of glucose tolerance make later studies of the prevalence of diabetes in the relatives of diabetic subjects difficult to evaluate [10–13] although support for a major gene has been adduced from the observation that approximately 40 % of the first-degree relatives of NIDDM probands may be expected to develop diabetes by the age of 80 [42]. However, the non-linear decrease in estimated prevalence rates of NIDDM seen amongst increasingly remote relatives of diabetic probands is not explicable under simple modes of inheritance and polygenic and oligogenic solutions have been proposed [14].

Evidence for major gene effects has been clearest in populations at high risk of NIDDM. Bimodality of blood glucose levels has been reported in several populations, including Asian Indian migrants to South Africa [2–4, 43, 44] and is consistent with, though not proof of, an underlying major gene effect. Analysis of Seminole Indian families provided evidence for an autosomal recessive gene determining hyperglycaemia [41] and support for a co-dominant gene has been reported in Pima Indians [45]. Serjeantson and Zimmet [46] used POINTER to adduce evidence for an autosomal dominant gene operating on a multifactorial background in Nauruan families. In contrast, segregation studies in populations at lower risk of diabetes (Japanese [47, 48] and Caucasians [49, 50]) have not found clear evidence for major gene effects.

The results obtained with COMDS suggest that genetic factors play a role in determining glucose levels throughout the range of glucose tolerance. Familial aggregation of glucose levels has been reported before [49–51] although in some studies aggregation has been weak [48, 52]. The path analysis methodology used in these studies is better-suited than segregation analysis to separation of genetic and cultural heritability (i.e. shared family environment) and several studies have suggested that the familial resemblance in glucose levels is predominantly cultural [48, 52, 53]. However, in two studies of Caucasian pedigrees, polygenic effects were more pronounced [50, 51]. POINTER does not permit discrimination of polygenic and transmitted environmental influences: evidence for intergenerational differences in the heritability of the multifactorial component (i.e. $z \neq 1$) may be taken to indicate support for additional sharing of environment between siblings but in practise the interpretation of z presents difficulties. In these data, z did not differ from 1 and there is therefore no support for unique environment common to siblings nor for alternative explanations (e.g. dominance deviations in polygenes). The difficulty in disentangling cultural from genetic influences is highlighted by considering obesity, which is strongly associated with diabetes, and results from the interplay of genetic [54], environmental and behavioural factors.

The failure of simple models of inheritance to describe segregation within these pedigrees fits with current views of the pathophysiology of NIDDM

and with experience in the testing of candidate genes in this disease [1]. The development of marked glucose intolerance requires defects in both insulin action and secretion [55]: genetic and environmental factors may contribute to both. Genetic analysis of glucose intolerance might be simplified if these intermediate traits are analysed separately since there is evidence that a major gene determines insulin action in certain populations [56, 57]. The high prevalence of diabetes in our families makes them unsuitable for analysis of intermediate traits because of the effects of hyperglycaemia and treatment on measures of insulin secretion and action in glucose intolerant subjects [58]: we are therefore unable to establish whether similar major gene effects are discernible at the level of intermediate traits in South Indians.

· Our observations have important implications for the search for genes involved in NIDDM pathogenesis, particularly in populations at intermediate prevalence of disease such as South Indians. The fact that simple modes of inheritance fail satisfactorily to explain the segregation of glucose intolerance implies that the simple genetic models often used to define the NIDDM-locus in linkage studies are inaccurate. Although linkage analysis is fairly robust to misspecifications of the genetic model [59], nevertheless, if NIDDM is appreciably polygenic with no single-gene contributing to a major proportion of the genetic susceptibility, the number of pedigrees required to demonstrate linkage may prove prohibitive [60]. Furthermore, application of inappropriate models may lead to apparent exclusion of loci or chromosomal regions which in fact contribute significantly to affection [18, 61, 62].

One approach to these problems is to employ non-parametric methods of linkage analysis [63] which are model-independent: these may detect but cannot characterise a linked locus. A more informative technique may be to perform combined segregation and linkage analysis using programs such as COMDS, a potentially powerful approach which has proved useful in studies of other complex diseases [23, 64].

In summary, a high prevalence of abnormal glucose tolerance was found amongst the parents of 64 NIDDM-affected South Indian probands recruited without regard to their family history of diabetes. Using complex segregation analysis, the segregation of NIDDM in these pedigrees could not be satisfactorily explained by simple models of inheritance: under the program COMDS, the best-supported model for the segregation of NIDDM featured a diallelic gene accounting for approximately 35 % of the variance in the liability to diabetes, acting on a background of transmitted family resemblance. These results emphasise the complexity of the genetic architecture of NIDDM: the suitability of the various strategies employed in the search for the specific genetic determi-

nants of NIDDM requires re-evaluation in this context.

References

1. McCarthy MI, Hitman GA (1993) The genetic aspects of non-insulin-dependent diabetes mellitus. In: Leslie RDG (ed) The causes of diabetes. Wiley, London, pp157–183
2. Raper LR, Taylor R, Zimmet P, Milne B, Balkau B (1984) Bimodality in glucose tolerance distributions in the urban Polynesian population of Western Samoa. Diabetes Research 1:19–26
3. Rushforth NB, Bennett PH, Steinberg AG, Burch TA, Miller M (1971) Diabetes in the Pima Indians. Evidence of bimodality in glucose tolerance distributions. Diabetes 20:756–765
4. Zimmet P, Whitehouse S (1978) Bimodality of fasting and two-hour glucose tolerance distributions in a Micronesian population. Diabetes 27:793–800
5. Fajans S (1990) Scope and heterogeneous nature of MODY. Diabetes Care 13:49–64
6. Bell GI, Xiang K-S, Newman MV et al. (1991) Gene for non-insulin-dependent diabetes mellitus (maturity onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. Proc Natl Acad Sci USA 88:1484–1488
7. Hattersley AT, Turner RC, Permutt MA et al. (1992) Linkage of type 2 diabetes to the glucokinase gene. Lancet 339:1307–1310
8. Froguel Ph, Vaxillaire M, Sun F et al. (1992) Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. Nature 356:162–165
9. Simpson NE (1964) Multifactorial inheritance. A possible hypothesis for diabetes. Diabetes 13:462–471
10. Cooke AM, Fitzgerald MG, Malins JM, Pyke DA (1966) Diabetes in children of diabetic couples. BMJ 2:674–676
11. Kahn CB, Soeldner JS, Gleason RE, Rojas L, Camerini-Davalos RA, Marble A (1969) Clinical and chemical diabetes in offspring of diabetic couples. New Engl J Med 281:343–347
12. Viswanathan M, Mohan V, Snehalatha C, Ramachandran A (1985) High prevalence of type 2 (non-insulin-dependent) diabetes among the offspring of conjugal type 2 diabetic parents in India. Diabetologia 28:907–910
13. Tattersall RB, Fajans SS (1975) Prevalence of diabetes and glucose intolerance in 199 offspring of thirty-seven conjugal diabetic parents. Diabetes 24:452–462
14. Rich SS (1990) Mapping genes in diabetes: genetic epidemiological perspective. Diabetes 39:1315–1319
15. O'Rahilly S, Turner R (1988) Early onset type 2 diabetes vs maturity onset diabetes of youth: evidence for the existence of two distinct diabetic syndromes. Diabet Med 5:224–229
16. Thompson GS (1965) Genetic factors in diabetes mellitus studied by the oral glucose tolerance test. J Med Genet 2:221–226
17. Cook JTE, Hattersley AT, Levy JC et al. (1993) Distribution of type II diabetes in nuclear families. Diabetes 42:106–113
18. Clerget-Darpoux F, Bonaïti-Pellié C (1992) Strategies based on marker information for the study of human diseases. Ann Hum Genet 56:145–153
19. Lalouel JM, Morton NE (1981) Complex segregation analysis with pointers. Hum Hered 31:312–321
20. Morton NE, Rao DC, Lalouel J-M (1983) Methods in genetic epidemiology. Karger, Basel
21. Morton NE, MacLean CJ (1974) Analysis of family resemblance III. Complex segregation of quantitative traits. Am J Hum Genet 26:489–503

22. Lalouel JM, Rao DC, Morton NE, Elston RC (1983) A unified model for complex segregation analysis. Am J Hum Genet 35:816–826

23. MacLean CJ, Morton NE, Yee S (1984) Combined analysis of genetic segregation and linkage under an oligogenic model. Comput Biomed Res 17:471–480

24. Morton NE, Shields DC, Collins A (1991) Genetic epidemiology of complex phenotypes. Ann Hum Genet 55: 301–314

25. Mohan V, Ramachandran A, Viswanathan M (1988) Diabetes in the tropics. In: Alberti KGMM, Krall LP (eds) Diabetes annual 4. Elsevier Science Publishers, Amsterdam pp 46–55

26. Kambo P, Hitman G, Mohan V et al. (1989) The genetic predisposition to fibrocalculous pancreatic diabetes. Diabetologia 32:45–51

27. World Health Organisation Study Group (1985) Diabetes mellitus. WHO Tech Rep Ser, no 727

28. Polymeropoulos MH, Xiao H, Rath DS, Merril CR (1991) Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). Nucleic Acids Res 19: 3753

29. Matsutani A, Janssen R, Donis-Keller H, Permutt MA (1992) A polymorphic (CA)n repeat element maps the human glucokinase gene (GCK) to chromosome 7p. Genomics 12:319–325

30. Nishi S, Stoffel M, Xiang K, Shows TB, Bell GI, Takeda J (1992) Human pancreatic beta-cell glucokinase: cDNA sequence and localization of the polymorphic gene to chromosome 7,band p 13. Diabetologia 35:743–747

31. Akaike H (1987) Factor analysis and AIC. Psychometrika 52:317–332

32. Morton NE, Chung CS (1959) Formal genetics of muscular dystrophy. Am J Hum Genet 11:360–379

33. Ramachandran A, Jali MV, Mohan V, Snehalatha C, Viswanathan M (1988) High prevalence of diabetes in an urban population in South India. BMJ 297:587–589

34. Ramachandran A, Snehalatha C, Dharmaraj D, Viswanathan M (1992) Prevalence of glucose intolerance in Asian Indians – urban-rural difference and significance of upper body adiposity. Diabetes Care, 15:1348–1355

35. Morton NE (1982) Outline of genetic epidemiology. Karger, Basel

36. Ramachandran A, Mohan V, Snehalatha C, Viswanathan M (1988) Prevalence of non-insulin-dependent diabetes mellitus in Asian Indian families with a single diabetic parent. Diabetes Res Clin Pract 4:241–245

37. Vieland VJ, Hodge SE, Greenberg DA (1992) Adequacy of single-locus approximations for linkage analyses of oligogenic traits. Genet Epidemiol 9:45–59

38. Ott J (1990) Cutting a Gordian knot in the linkage analysis of complex human traits. Am J Hum Genet 46:219–221

39. O'Rahilly S, Wainscoat JS, Turner RC (1988) Type 2 (non-insulin-dependent) diabetes mellitus: new genetics for old nightmares. Diabetologia 31:407–414

40. Devi AR, Rao NA, Bittles AH (1988) Consanguinity, fecundity and post-natal mortality in Karnataka, South India. Ann Hum Biol 8:469–472

41. Elston RC, Namboodiri KK, Nino HV, Pollitzer WS (1974) Studies on blood and urine glucose in Seminole Indians: indications for segregation of a major gene. Am J Hum Genet 26:13–34

42. Köbberling J, Tillil H (1982) Empirical risk figures for first degree relatives of non-insulin-dependent diabetics. In: Köbberling J, Tattersall R (eds) The genetics of diabetes mellitus. Academic Press London, pp 201–209

43. Steinberg AG, Rushforth NB, Bennett PH, Burch TA, Miller M (1970) On the genetics of diabetes. Nobel Symposium 13:237–264

44. Omar MAK, Seedat MA, Dyer RB, Motala AA, Knight LT, Becker PJ (1994) South African Indians show a high prevalence of NIDDM and bimodality in plasms glucose distribution pattern. Diabetes Care 17:70–73

45. Yamashita T, Mackay W, Rushforth N, Bennett P, Houser H (1984) Pedigree analyses of non-insulin dependent diabetes mellitus (NIDDM) in the Pima Indians suggest dominant mode of inheritance. Am J Hum Genet 36 [Suppl]:183S (Abstract)

46. Serjeantson SW, Zimmet P (1991) Genetics of non-insulin dependent diabetes mellitus in 1990. Baillieres Clin Endocrinol Metabol 5:477–493

47. Mimura G, Oshiro S, Koganemaru K, Haraguchi T, Hashihuchi J (1964) Studies on the heredity of diabetes mellitus in Japan II. Inheritance of the fasting blood sugar value and the blood sugar value two hours after meal in Uto and Tomiai inhabitants. Kumamoto Med J 17:50–57

48. Williams WR, Morton NE, Rao DC, Gulbrandsen CL, Rhoads GG, Kagan G (1983) Family resemblance for fasting blood glucose in a population of Japanese Americans. Clinical Genetics 23:287–293

49. Rice T, Laskarzewski PM, Rao DC (1992) Commingling and complex segregation analysis of fasting plasma glucose in the Lipid Research Clinics Family Study. Am J Med Genet 44:399–404

50. Laskarzewski PM, Rao DC, Glueck CJ (1984) The Cincinnati Lipid Research Clinic Family Study: analysis of commingling and family resemblance for fasting blood glucose. Genet Epidemiol 1:341–355

51. Boehnke M, Moll PP, Kottke BA, Weidman WH (1987) Partitioning the variance of fasting plasma glucose levels in pedigrees. Am J Epidemiol 125:679–689

52. Friedlander Y, Kark JD, Bar-On H (1987) Family resemblance for fasting blood glucose: the Jerusalem Lipid Research Clinic. Clinical Genetics 32:222–234

53. King H, Rao DC, Bhatia K, Koki G, Collins A, Zimmet P (1989) Family resemblance for glucose tolerance in a Melanesian population, the Tolai. Hum Hered 39:212–217

54. Stunkard AJ, Harris JR, Pedersen NL, McClearn GE (1990) The body-mass index of twins who have been reared apart. New Engl J Med 322:1483–1487

55. Cahill GF (1988) Beta-cell deficiency, insulin resistance or both? New Engl J Med 318:1268–1270

56. Schumacher MC, Hasstedt SJ, Hunt SC, Williams RR, Elbein SC (1992) Major gene effect for insulin levels in familial NIDDM pedigrees. Diabetes 41:416–423

57. Lillioja S, Mott D, Zawadzki JK et al. (1987) In vivo insulin action is familial characteristic in nondiabetic Pima Indians. Diabetes 36:1329–1335

58. Rossetti L, Giaccori A, De Fronzo RA (1990) Glucose toxicity. Diabetes Care 13:610–630

59. Ott J (1991) Analysis of human genetic linkage. Revised edn. Johns Hopkins University Press, Baltimore

60. Lathrop GM (1989) The power of linkage studies of Mendelian and multifactorial genetic disease in man. In: Nerup J, Mandrup-Poulsen T, Hökfelt B (eds) Genes and gene products in the development of diabetes mellitus. Elsevier, Amsterdam, pp 373–379

61. Clerget-Darpoux F, Bonaïti-Pellié C (1993) An exclusion map covering the whole genome: a new challenge for genetic epidemiologists? Am J Hum Genet 52:442–443 (Letter)

62. Greenberg DA (1993) Linkage analysis of "necessary" disease loci versus "susceptibility" loci. Am J Hum Genet 52:135–143

63. Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326

64. Risch N (1984) Segregation analysis incorporating linkage markers. I. single-locus models with an application to type 1 diabetes. Am J Hum Genet 36:363–386