

Segregation analysis of NIDDM in Caucasian families

J. T. E. Cook¹, D. C. Shields², R. C. L. Page¹, J. C. Levy¹, A. T. Hattersley¹, J. A. G. Shaw¹, H. A. W. Neil³, J. S. Wainscoat⁴, R. C. Turner¹

¹ Diabetes Research Laboratories, Radcliffe Infirmary, Oxford, UK

² CRC Genetic Epidemiology Research Group, Department of Child Health, University of Southampton, UK

³ Department of Public Health and Primary Care, University of Oxford, UK

⁴ Department of Haematology, John Radcliffe Hospital, Oxford, UK

Summary Non-insulin-dependent diabetes mellitus (NIDDM) has a substantial genetic component, but the mode of inheritance and the molecular basis are unknown. We have undertaken segregation analysis of NIDDM after studying 247 subjects in 59 Caucasian nuclear pedigrees ascertained without regard to family history of the disorder. The analyses were performed using POINTER and COMDS, which are computer programs which apply statistical models to the data. POINTER analysis was performed defining the phenotype as a presence or absence of hyperglycaemia. Among single locus hypotheses, the analyses rejected a recessive model and favoured a dominant model, but could not statistically show that this fitted better than a mixed model (a single locus against a polygenic background) or a polygenic model. COMDS analysis assumed a continuum of hyperglycaemia from normality to NIDDM, classified family members into a series of diathesis classes with

increasing plasma glucose levels and compared the distribution with that found by screening the normal population. This analysis improved the likelihood of a dominant single locus model and suggested a gene frequency of 7.4%. It raised the possibility of a second locus, but cannot identify or exclude a polygenic model. In conclusion, two types of segregation analyses rejected a recessive model and favoured a dominant model of inheritance, although they could not statistically show that this fitted better than the polygenic model. The results raised the possibility of a common dominant gene with incomplete penetrance, but genetic analysis of NIDDM needs to take into account the likelihood of polygenic inheritance with genetic heterogeneity. [Diabetologia (1994) 37: 1231–1240]

Key words Non-insulin-dependent diabetes mellitus, genetic epidemiology, genetic linkage.

Non-insulin-dependent diabetes mellitus (NIDDM) is a common metabolic disorder with considerable morbidity and mortality. Despite evidence for a substantial genetic component, the mode of inheritance and the molecular basis of this inheritance remain un-

known. The identification of a gene (or genes) contributing to susceptibility to NIDDM would have profound implications upon its prevention and management.

Barnett et al. [1] documented 91% concordance for NIDDM in monozygotic twin pairs. Whilst this study may have included some ascertainment bias, the prospective study of unselected twins by Newman et al. [2] provided supportive data. At the initial examination 58% concordance for NIDDM was found, and only 1 of 15 originally discordant twin pairs remained discordant after 10 years. These studies suggest a strong genetic predisposition to NIDDM, but the lack of complete concordance and variation in age of onset between twins suggest input from environmental influences. No comprehensive

Received: 21 January 1994
and in revised form: 20 June 1994

Corresponding author: Dr. R. C. Turner, Clinical Reader, Radcliffe Infirmary, Diabetes Research Laboratories, Woodstock Road, Oxford OX2 6HE, UK

Abbreviations: MODY, Maturity onset diabetes of the young; IDDM, insulin-dependent diabetes mellitus; NIDDM, non-insulin-dependent diabetes mellitus; FPG, fasting plasma glucose, AIC, Akaike Information criterion.

prospective studies comparing the development of diabetes in monozygotic and dizygotic twins, or the concordance rates in monozygotic twins reared together and apart, have been reported.

Bimodality of glucose tolerance has been demonstrated in three populations where diabetes is common [3–5], and bimodality has been taken to suggest a single major gene influence. An alternative explanation may be that the rate of transition from normality to disease in these populations is rapid, perhaps due to the deleterious effects of hyperglycaemia. Bimodality has not been described in the Caucasian population [7] but has been reported in the first-degree relatives of Caucasian NIDDM subjects [8].

Linkage analysis requires knowledge of genetic parameters such as the mode of inheritance and the gene frequency and the penetrance, and if these are incorrectly specified the sensitivity is reduced. Linkage analysis has been successfully applied to the clinical subtype termed maturity-onset diabetes of the young (MODY), which is characterised by the presentation of diabetes in early adult life and by pedigree structures suggestive of autosomal dominant transmission [9]. The application of linkage analysis to NIDDM has been more problematic because of the lack of data concerning its mode of inheritance.

Segregation analysis is the statistical technique used for the detection of the mode of inheritance of familial disease. Although it would appear to be a necessary prerequisite to more sophisticated genetic studies, segregation analysis data has not been previously reported for NIDDM. Early studies of the familial prevalence of diabetes are difficult to interpret because insulin and non-insulin dependent diabetes were not discriminated [10–14]. Studies documenting family history of the disease are likely to be only partially informative because NIDDM may be subclinical [15]. Studies in which complete nuclear families have been tested were difficult to undertake because of the late age of onset and increased mortality of the disease. The major difficulty is that in most cases one or both parents of a subject with NIDDM are deceased, while their children are not yet old enough to express the disease.

A recent study examined 20 consecutive nuclear families selected without regard to family history of the disorder, in which both parents of probands were alive and available for study [16]. Seven probands were found to have neither parent affected with diabetes or impaired glucose tolerance, ten probands had one affected parent (six with diabetes and four with impaired glucose tolerance) and three had both parents affected (one with concordant diabetes and two with concordant impaired glucose tolerance). These data did not support the assumption of autosomal dominant inheritance with complete penetrance, although the data set was too small for statistical analysis.

This study describes the application of formal segregation analysis to the nuclear families of a series of 59 Caucasian NIDDM probands who were ascertained without regard to family history of the disorder. The implications of our findings for the genetic analysis of NIDDM are discussed.

Methods and materials

The protocol was approved by the Central Oxford Research and Ethics Committee and informed consent obtained from all subjects.

Nuclear pedigrees. We investigated 697 Caucasian NIDDM subjects attending routine diabetic clinics in Oxfordshire concerning the availability of a complete nuclear family for testing. We ascertained 59 nuclear families without regard to family history of the disorder. From 431 subjects 21 probands with both parents alive were ascertained. An additional 38 nuclear pedigrees were identified by seeking elderly probands with a living spouse and offspring aged older than 25 years. All probands were diagnosed after age 35 years, had no history of ketosis and had been treated initially by diet or oral hypoglycaemic agents.

The 21 probands with living parents were aged 44 ± 6 years (mean \pm SD), had a duration of diabetes of 6 ± 5 years and were of body mass index (BMI) 29.4 ± 5.1 kg/m². We studied the 28 siblings and 42 parents of these probands. The 38 probands with a living spouse and offspring were aged 71 ± 8 years, had a duration of diabetes 10 ± 7 years and were of BMI 28.5 ± 4.7 kg/m². We studied the 38 spouses and 82 offspring of these probands. None of the probands had ketonuria greater than 1.5 mmol/l (Bayer Diagnostics, Basingstoke, UK) at presentation, and all were treated by diet or tablets for at least 3 months.

Plasma glucose samples were obtained from the probands, their siblings and both parents after a 12-h overnight fast. Of the 247 subjects 229 were then studied with a continuous infusion of glucose test [17]. This consisted of a continuous intravenous infusion of 5 mg glucose · kg ideal body weight · min⁻¹ for 60 min. Ideal body weight was taken from the Metropolitan Life Insurance tables for a medium frame [18]. The achieved plasma glucose is the mean of the 50-, 55- and 60-min samples. Seventeen of the subjects declined glucose tolerance testing, and only fasting plasma glucose was obtained. Plasma glucose was measured with a hexokinase method using a Cobas MIRA centrifugal analyser (Roche Diagnostica, Welwyn Garden City, UK).

Liability classes. Liability classes were used to specify the age-associated risk of diabetes. The diabetic and the non-diabetic nuclear family members were assigned to liability classes on the basis of age: Class 1, 20–29 years; Class 2, 30–39 years; Class 3, 40–49 years; Class 4, 50–59 years; Class 5, 60–69 years; Class 6, 70–79 years; Class 7, ≥ 80 years. The population frequencies of NIDDM for these seven classes were calculated from two Oxford community survey studies which documented the prevalence of known and newly-diagnosed NIDDM, respectively. The Oxford Community Diabetes Study [19] documented the frequencies of known diabetes in adult subjects at different ages. This study did not discriminate between IDDM and NIDDM. As the prevalence of IDDM is reported to be 0.3% in adults [20], the frequencies in each liability class were lowered by 0.003. The population frequencies of newly-diagnosed NIDDM were taken from the prevalence

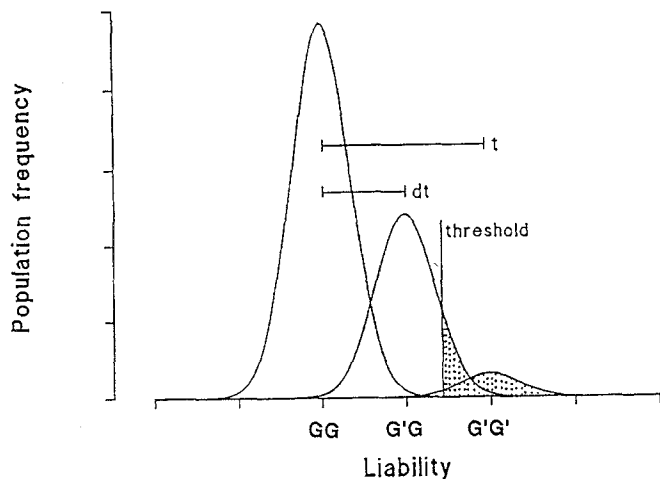


Fig. 1. Calculating penetrances i.e. probability of affection given genotype. From left to right, the three genotypes are homozygote normal (GG), heterozygote (G'G) and susceptible homozygote (G'G'). The size of the distributions is determined by the frequency of the rare allele q , while the displacement between the three genotypes is determined by d (degree of dominance) and t (the displacement between 2 homozygotes of major gene). In this example, d , t and q are 0.5, 3.0 and 0.2, respectively, and the calculations are carried out for the oldest class, where the population frequency of diabetes is 0.1146. A threshold is found, such that the overall probability of affection is 0.1146. The penetrance for each genotype is calculated as the ratio of the shaded to the total area of the genotypic distribution, giving penetrances of 0.0002, 0.2349 and 0.9827 for each genotype. A similar calculation is carried out for diathesis classes in the COMDS analysis, replacing t with Bt

of fasting plasma glucose (FPG) greater than 7.8 mmol/l in a screening study of an Oxford population sample of 4201 subjects without known diabetes [21]. In addition to the 4006 subjects aged 25–60 years described by Neil et al. [21], we included 195 subjects from the screening study who were aged over 60 years. The frequencies of known and newly-diagnosed NIDDM were added together, and the final population frequencies used were: Class 1, 0.0008; Class 2, 0.0042; Class 3, 0.0089; Class 4, 0.0211; Class 5, 0.0283; Class 6, 0.0927 and Class 7, 0.1146.

Affection status. For analysis with POINTER, affection was defined as having hyperglycaemia defined as a fasting plasma glucose, or an achieved plasma glucose level after a continuous infusion of glucose test, more than two standard deviations above the mean normal value for the subject's age and obesity as determined in comparison with a population of 104 normal subjects (age range 21–76 years, ideal body weight range 86%–158%). A comparison study of 30 subjects who had both the glucose infusion test and a standard 75 g oral glucose tolerance test showed that these criteria gave 89% sensitivity and 100% specificity for World Health Organisation (WHO) defined impaired glucose tolerance [22]. NIDDM was diagnosed according to the WHO criterion of fasting plasma glucose greater than 7.8 mmol/l [23]. As IGT and diabetes form a continuum, and the WHO definition of NIDDM related to the increased risk for microvascular disease and not to a specific phenotype, the phenotype chosen for study has been the presence of abnormal hyperglycaemia per se. The use of age-adjusted criteria makes it unlikely that over-diagnosis in old age would occur.

With COMDS, the non-diabetic family members were assigned to classes of increasing FPG, termed diathesis classes. Increasing diathesis class can be used to imply increasing probability of being genetically predisposed to the disease. The population distribution of FPG among non-diabetic individuals was taken from the Oxford population screening study of 4201 subjects [21]. The FPG was shown by linear regression to associate with age. The residual FPG was therefore obtained as $[FPG - 3.91 - 0.013(\text{age})]$. This transformation also fits the age-dependence of FPG in a United States population survey [24]. The age-corrected residual was subdivided into five diathesis classes of increasing FPG, comprising 75%, 10%, 5%, 5% and 5% of the population, respectively, using cut-off points of 0.28, 0.49, 0.63 and 0.90 mmol/l above the age-corrected FPG. Non-diabetic individuals in the 59 families were assigned to one of the five diathesis classes on the basis of their FPG corrected for age.

Segregation analysis

Segregation analysis was performed with POINTER [25] and COMDS [26], which are computer programs which apply statistical models to the nuclear family data.

POINTER segregation analysis. POINTER was used to fit various single gene, polygenic and mixed models to the data. The mixed model assumes that a continuous variable X results from the independent contribution of a major locus, a polygenic component and random environmental effects. The following parameters are provided at maximum likelihood: d – degree of dominance, which ranges between 0 for a recessive gene and 1 for a dominant; t – displacement between the two homozygotes of the major gene; q – gene frequency of allele leading to affection (Fig. 1); H – the heritability of the polygenic component; Z – the ratio of adulthood to childhood heritability. The displacement between homozygote normal and the heterozygote is calculated as (Fig. 1). In POINTER, a quantitative trait may be considered in addition to affection status if it is normally distributed, an example being the consideration of serum iron levels in the segregation analysis of idiopathic haemochromatosis [27]. However, FPG is not normally distributed, and a correction for skewness is unlikely to be sufficient to make the distribution normal. Therefore, the POINTER analysis was restricted to a categorical analysis considering affection status only. It was used because it can model a multifactorial (polygenic) component. POINTER transmission parameters were not reported as recent analysis has shown that their calculation is incorrect [28].

COMDS segregation analysis. In the general population and in families, glucose tolerance and diabetes form a continuum. In COMDS, the phenotype includes increasing degrees of hyperglycaemia as a polychotomy with different diathesis classes. This additional information may increase the power to distinguish between hypotheses. Genotypic displacement is a measure of the difference between two degrees of genetic susceptibility (Fig. 1). Diathesis is introduced into the model through the parameter B which scales the genotypic displacement between diathesis classes between affected and normal (Fig. 2). When $B = 0$ there are no genotypic differences between diathesis classes, and the analysis is equivalent to a categorical dichotomy. When B is less than 1, the genotypic difference between individuals drawn from the different diathesis classes is less than the genotypic difference between an affected and the highest diathesis class (Fig. 2). Holding $B = 1$ is equivalent to assuming that the affected status is an additional equally-

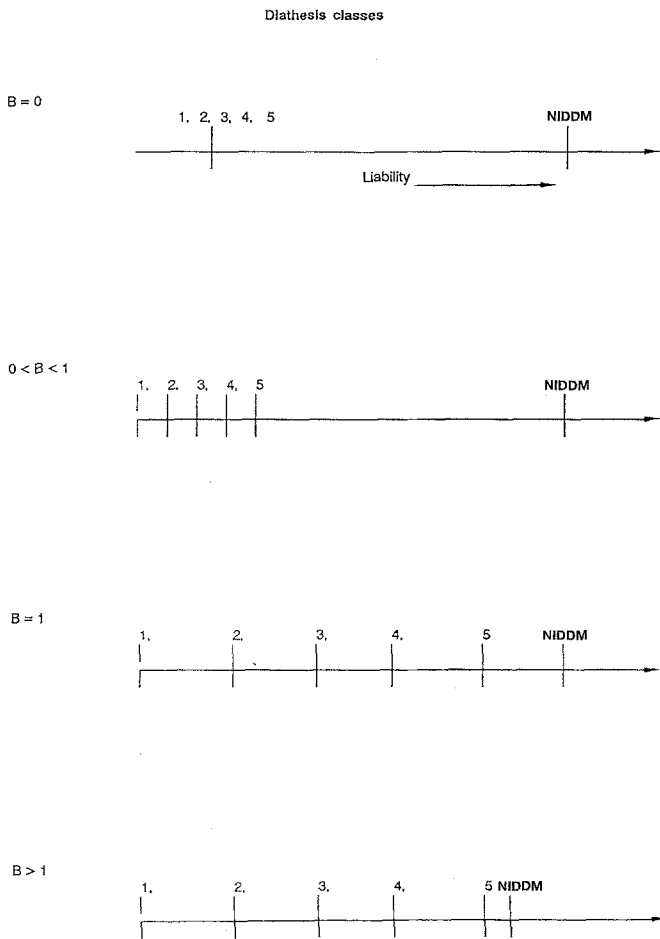


Fig. 2. In COMDS, the phenotype considered includes diathesis class as well as affection status. Diathesis is introduced into the model through the parameter B which scales the genotypic displacement between diathesis classes to the genotypic displacement between affected and normal. When B is held at 0, the analysis is equivalent to one without diathesis information. When B is held at 1, diathesis and NIDDM are assumed to be colinear, and the class of affected is given the same weight as a diathesis class. B can be estimated as less than or greater than 1. When B is estimated as less than 1, the genotypic difference between individuals in diathesis classes is less than the genotypic difference between affected and non-affected. When B is estimated as > 1 , there is a smaller genotypic difference between affected and non-affected than between each of the subsequent diathesis classes

spaced class at the top end of the distribution of diathesis classes. This latter is equivalent to a simpler model, where there is a single underlying distribution split into an ordered polyotomy with affected as the highest class, and with diathesis classes as lower categories. When B is more than 1, there is a smaller genotypic difference between affected and the highest diathesis class than between diathesis classes.

POINTER allows a mixed model, where disease is defined as occurring above a certain threshold of a continuous variable, which comprises the additive effects of a single locus, polygenes and the environment. COMDS differs by having single and two locus models, but no polygenic effect. Two autosomal loci can be considered, each with a high-risk and a low-risk allele, whose effects are additive on a scale of liability. For convenience the loci are termed major and modifier, although the

major locus need not necessarily have the greater influence. The modifier locus has the parameters q_m , d_m , t_m , B_m . A "pseudopolygenic" effect can be approximated in COMDS at one or two loci by holding gene frequency and dominance of 0.5. This does not correspond exactly to the multifactorial effect in POINTER; but it is a more parsimonious representation of a residual familial effect, because it requires fewer parameters than a full two-locus model.

Application of the models. COMDS and POINTER were used to calculate the likelihood of the offsprings' phenotypes, conditional on parental phenotypes. The nuclear families were sampled through an affected proband. The likelihood was corrected for ascertainment by conditioning on whether the family was identified through a parent or offspring. The probability of ascertaining an individual in the population was assumed to be small, and taken to be 0.001. The likelihood is presented as twice the natural log likelihood, plus a constant. The significance of adding parameter(s) to a model may be evaluated by taking the difference in $-2 \ln(L)$ between the two models as a chi-square with the number of degrees of freedom equal to the number of extra parameters.

The Akaike Information Criteria (AIC) [29, 30] were calculated from the POINTER analysis likelihood by adding twice the degrees of freedom to the likelihood and comparing the overall values. The lowest number represents the best model fit. AIC was also calculated for COMDS except for $B = 1$ and $B = B_m = 1$.

Results

The distribution of affection in relatives. Of the probands with living parents, 7 had neither parent affected, 11 had one parent affected and 3 had both parents affected. The probands with affected parents and those with unaffected parents had similar age at diagnosis (39 ± 5 , 42 ± 5 years) and obesity (30 ± 3 , 29 ± 8 kg/m²). In the 7 families with neither parent affected, 30% of the siblings of the probands were affected. In the 11 families with one affected parent, 46% of the siblings were affected. In the 3 families with both parents affected, each of the three siblings available for testing had impaired glucose tolerance.

Of the probands with a living spouse and offspring, 32 had a normoglycaemic spouse. In these families, 26% of the offspring were affected. Six of the probands had an affected spouse, and 65% of the offspring in these families were affected (Table 1). A summary of the pedigree structures is provided (Fig. 3).

Formal segregation analysis

POINTER segregation analysis. Likelihood ratio tests were constructed for the general model where all relevant parameters were estimated except that Z was fixed to 1 (Table 2). The hypothesis of no major gene was rejected ($\chi^2_3 = 6.4$, $p < 0.05$) and the hypothesis of no multifactorial component was rejected ($\chi^2_1 = 0$), however, both could not be dropped from

Table 1. Description of families; the frequency with which neither, one or both parents were affected, by impaired glucose tolerance or diabetes, and the proportion of affected first degree relatives

Probands with living parents		Affected siblings
Diabetic status of parents	<i>n</i>	
Neither affected	7	3/10 (30%) 1 DM, 2 IGT
One parent affected	11	6/13 (46%) 2 DM, 4 IGT
Both parents affected	3	3/3 (100%) 3 IGT
2 IGT/IGT, 1 DM/DM		
Probands with living spouse and offspring		Affected offspring
Normoglycaemic spouse	32	17/65 (26%) 3 DM, 14 IGT
– one parent affected		
Affected spouse	6	11/17 (65%)
– both parents affected		3 DM, 3 IGT
		1 DM, 10 IGT

DM, Diabetes mellitus; IGT, impaired glucose tolerance

converged to a single dominant locus with no polygenic effect (i.e., *d* converged to 1 and *H* to 0). For comparison with the polygenic model, the dominant model may be taken to be the mixed model with a dominant major gene effect. While the polygenic and multifactorial models had a somewhat lower likelihood, they were not significantly different from the mixed model ($\chi^2_3 = 6.4, p < 0.10$). Thus, the analysis suggested a single dominant locus but could not statistically show that this fitted better than a mixed model or a polygenic model. There was no significant evidence that childhood and adult heritability differ; when their ratio, *Z*, is allowed to depart from 1, $\chi^2_1 = 6.4 - 4.8 = 1.6$. Using AIC, the autosomal dominant model provided the best fit.

COMDS analysis. When the parameter *B* is held at 0, COMDS is equivalent to the categorical affection status of POINTER, with all non-diabetic patients being included together. When *B* and *B_m* are held at 0, the genetic displacement represented by *B_t* and *B_{m,t_m}* in the model are also 0 (Fig. 1). With a single locus model when *B* = 0 (analysis of affection status only) the likelihood of differences were identical to the POINTER results (Table 3). The addition of the FPG diathesis (estimating *B*) to allow for the continuum of FPG as a polychotomy improved the likelihood of the general (dominant) single locus model ($\chi^2_1 = 239.4 - 105.2 = 134.2 = p < 0.001$) (Table 3). The dominant model was favoured and both the recessive model ($\chi^2_1 = 123.0 - 105.2 = 17.8, p < 0.001$) and the co-dominant model were rejected ($\chi^2_1 = 119.5 - 105.2 = 14.3$). When *d* was estimated, it converged to 1. The addition of the diathesis classes thus increased the power to distinguish between models. The families ascertained through a parent and those ascertained through an offspring both contributed to the evidence favouring a single dominant gene. Under a single, dominant gene model, there was no evidence for heterogeneity (holding *d* = 1, *q* = 0.075, *B* = 1, estimating *t*, $\chi^2_1 = 2.16$).

When a second locus was added to the model an improved fit was obtained, but this was not significant ($\chi^2_4 = 105.2 - 100.0 = 5.2$) (Table 4) and it was not possible to ascertain whether a dominant or recessive second locus was present. Using AIC the same conclusion was reached, with AIC equal to 111.2 for a dominant single locus and 112.0 for the best 2 locus model. This eight-parameter model is quite complex for a relatively small data set, so models with a second locus effect and fewer parameters were also considered to determine whether a significant residual genetic effect could be modelled. A dominant/pseudopolygenic model, with only one estimated second locus parameter failed to provide an improvement in fit ($\chi^2_1 = 105.2 - 104.6 = 0.6$). Other simple models tested also failed to provide a significant improvement: those which did not simply reduce to the dominant single locus model are presented in Table 4.

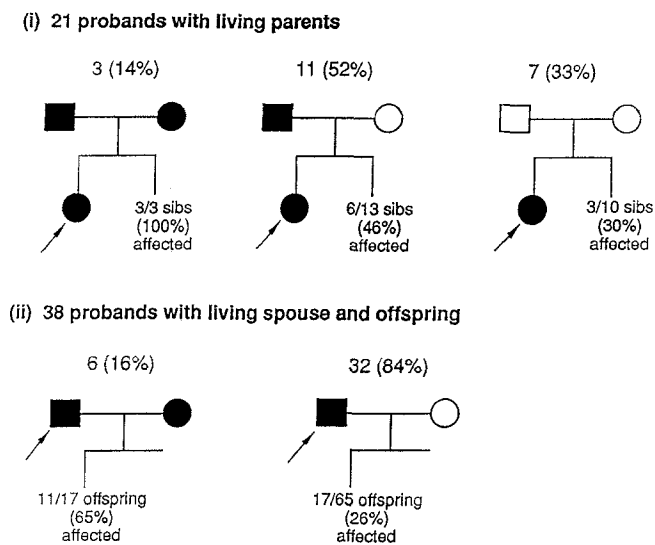


Fig. 3. Summary of structure of 59 NIDDM nuclear families. ■, NIDDM or glucose intolerance; □, normoglycaemia; arrow indicates proband

the model simultaneously ($\chi^2_4 = 28.1$). Among single locus hypotheses, the analysis favoured a dominant model. Assuming there is a major gene, the recessive model was rejected ($\chi^2_1 = 5.5, p < 0.025$) but a co-dominant model was not rejected ($\chi^2_1 = 0.4 - 0 = 0.4$). The families ascertained through an offspring and those ascertained through a parent both contributed to the evidence for a single dominant gene. Under the dominant gene model, there was no evidence of heterogeneity (holding *d* = 1, *q* = 0.013 estimating *t*, $\chi^2_1 = 0.5$). A mixed model of a single locus plus a multifactorial effect could not be fitted that was better than the dominant gene alone. The general model

Table 2. POINTER segregation analysis of affection status alone: single locus, polygenic and mixed models

Hypothesis	d	t	q	H	Z	-2ln(L) + C	AIC
Sporadic	(0)	(0)	(0)	(0)	(0)	28.1	
<i>Single gene models</i>							
Autosomal dominant	(1)	2.33	0.013	(0)	(0)	0	6.0
Co-dominant	(0.5)	4.78	0.010	(0)	(0)	0.4	6.4
Autosomal recessive	(0)	2.78	0.124	(0)	(0)	5.5	11.5
General	1.00	2.33	0.013	(0)	(0)	0	6.0
<i>Polygenic models</i>							
Polygenic	(0)	(0)	(0)	0.68	(1)	6.4	8.4
Multifactorial	(0)	(0)	(0)	0.98	0.41	4.8	8.8
<i>Mixed models</i>							
General	1.00	2.33	0.013	0.00	(1)	0	8.0
(Recessive mixed)	(0)	2.39	0.123	0.57	(1)	4.5	12.5

POINTER is used to fit various single gene, polygenic and mixed models to the data. Estimates of parameters such as degree of dominance (d), gene frequency (q) are provided at maximum likelihood; d, Degree of dominance of major locus; t, displacement between 2 homozygotes of major gene; q, gene frequency at major locus; H, childhood heritability; Z, ratio of adult: childhood heritability; B, diathesis parameter at

major locus; d_m , dominance parameter of modifier locus; t_m , displacement or scale parameter of modifier locus; q_m , gene frequency at modifier locus; B_m , diathesis parameter at modifier locus; $-2\ln(L) + C$, minus twice log-likelihood plus constant; Parenthesis indicates fixed parameter; AIC, Akaike's information criterion

Table 3. COMDS segregation analysis of affection status and fasting glucose diathesis: single locus models

Hypothesis	d	t	q	B	-2ln(L) + C	AIC
Sporadic	(0)	(0)	(0)	(0)	259.6	
<i>Affection status alone (B = 0)</i>						
Dominant	(1)	2.33	0.013	(0)	239.4	243.4
Co-dominant	(0.5)	4.78	0.010	(0)	239.9	243.9
Recessive	(0)	2.78	0.123	(0)	245.0	249.0
General	1.00	2.33	0.013	(0)	239.4	245.4
<i>Affection status co-linear with fasting glucose (B = 1)</i>						
Dominant	(1)	2.09	0.075	(1)	105.8	n/a
Co-dominant	(0.5)	3.64	0.078	(1)	119.9	n/a
Recessive	(0)	1.91	0.385	(1)	124.9	n/a
General	1.00	2.09	0.075	(1)	105.8	n/a
<i>Affection and diathesis scaled (B estimated)</i>						
Dominant	(1)	2.76	0.074	0.75	105.2	111.2
Co-dominant	(0.5)	3.38	0.077	1.11	119.5	125.2
Recessive	(0)	2.80	0.385	0.67	123.0	129.0
General	1.00	2.76	0.074	0.75	105.2	113.2

The susceptibility gene frequency for the dominant single locus model in POINTER, considering a categorical affection status alone, was 1.3% (Table 2). The penetrance factors are a function of the population frequency of diabetes in each liability class and the parameters of the model and are shown in Table 5. When the continuum of glucose from normal to diabetic was modelled with COMDS, the susceptibility gene frequency was 7.4% (Table 3) and the penetrance factors are in Table 5. A second locus, if present, would reduce the first dominant susceptibility gene frequency to 5.9%, and the second locus may have a gene frequency in the order of 4.3% for a dominant and 22.1% for a recessive gene effect. It must be emphasised that the models are based on a moderately-sized data set, and the gene frequencies and penetrances are approximate.

Discussion

The segregation analysis of NIDDM in Caucasian nuclear families rejected both no genetic component and a recessive model. A dominant single gene model gave the best fit to the data, but the analysis could not statistically show that this fitted better than the mixed model or the polygenic model.

The families were ascertained without regard to family history, and those ascertained through a parent and those ascertained through an offspring both contributed to the evidence favouring a single dominant gene. A heterogeneity analysis found no difference between families ascertained through a diabetic parent and ascertained through an adult offspring. However certain biases may have arisen from the need to determine accurately the affection status of

Table 4. COMDS segregation analysis of affection status and fasting plasma glucose diathesis: two locus models

	Major locus					Modifier locus				
	d	t	q	B	d _m	t _m	q _m	B _m	-2ln(L) + C	A1C
<i>B = B_m = 0</i>										
Pseudopolygenic ¹	(0.5)	1.75	(0.5)	(0)	(0.5)	1.75	(0.5)	(0)	247.7	251.7
Dominant/Dominant	(1)	3.10	0.048	(0)	(1)	0.99	0.040	(0)	238.6	246.6
Dominant/Recessive	(1)	2.33	0.013	(0)	(0)	0.09	0.075	(0)	239.4	247.4
<i>B, B_m estimated</i>										
Pseudopolygenic	(0.5)	1.81	(0.5)	0.92	(0.5)	1.81	(0.5)	0.92	124.6	132.6
Dominant/Pseudopolygenic ²	(1)	2.44	0.081	0.84	(0.5)	0.66	(0.5)	0.84	104.1	114.1
Dominant/Dominant	(1)	3.18	0.047	0.64	(1)	0.74	0.043	2.36	100.6	112.6
Dominant/Recessive	(1)	2.94	0.059	0.68	(0)	0.82	0.221	2.38	100.0	112.0
<i>B = B_m = 1</i>										
Pseudopolygenic ¹	(0.5)	1.72	(0.5)	(1)	(0.5)	1.72	(0.5)	(1)	125.3	n/a
Dominant/Pseudopolygenic ²	(1)	2.06	0.083	(1)	(0.5)	0.77	(0.5)	(1)	104.6	n/a
Dominant/Dominant	(1)	0.78	0.084	(1)	(1)	2.10	0.082	(1)	103.0	n/a
Dominant/Recessive	(1)	2.10	0.077	(1)	(0)	0.98	0.267	(1)	104.0	n/a
<i>B = 1, B_m = 0</i>										
Dominant/Dominant	(1)	2.10	0.079	(1)	(1)	1.04	0.054	(0)	103.2	n/a
Recessive/Recessive	(0)	1.87	0.385	(1)	(0)	34.80	0.019	(0)	122.9	n/a

¹ d, q, d_m and q_m = 0.05

² d_m and q_m = 0.5

Table 5. Penetrance factors and gene frequency derived from models: a) POINTER single dominant locus susceptibility allele G' frequency = 0.013; b) COMDS single dominant locus model estimating t, q, and B, allele G' frequency = 0.074

Liability class	Penetrances			
	(a) POINTER		(b) COMDS	
	GG' or G'G'	GG	GG' or G'G'	GG
Age (years)				
1 (20-29)	0.03	0.00	0.01	0.00
2 (30-39)	0.15	0.00	0.03	0.00
3 (40-49)	0.30	0.00	0.06	0.00
4 (50-59)	0.53	0.01	0.15	0.00
5 (60-69)	0.61	0.01	0.20	0.00
6 (70-79)	0.85	0.07	0.65	0.00
7 (≥ 80)	0.88	0.09	0.80	0.00

study participants. Firstly, families motivated to attend for testing may include an excess of those with a positive family history. Secondly, the increased mortality of NIDDM could bias towards the ascertainment of families with unaffected parents. Thirdly, the probands with living parents had an early age of onset of NIDDM and this has been associated with a higher incidence of diabetic parents in Caucasian subjects [31]. This effect was not as apparent as in the data of O'Rahilly et al. [31] who reported 92 % affection in the 23 available parents of 13 NIDDM subjects who presented in this age group. An increased chance of finding diabetes in the relatives was possible in that study, as six probands were ascertained through an affected family member.

The effects of obesity and disease-specific mortality are potential confounding factors in the application of segregation analysis to NIDDM. The analysis

presented makes no correction for obesity, and in the absence of a precise understanding of the interaction between genetic factors, obesity and diabetes this is appropriate [32]. If the obesity that predisposes an individual to NIDDM has a genetic component, this could be regarded as part of the genetic predisposition to diabetes studied in the pedigrees. However, if obesity were purely an environmental risk factor, a correction for this precipitating factor would be desirable; failure to correct would not bias the genetic models obtained, although it would reduce their statistical power for detecting genetic determinants. The effect of specific mortality due to diabetes would be to reduce the gene frequency in the older age groups, whereas the POINTER and COMDS programs assume a constant gene frequency through all age groups. The data of Panzram et al. [33] indicate that the risk of death is only doubled in NIDDM therefore this factor should not seriously bias the analysis.

The COMDS analysis considered models where genetic factors influence the range of FPG levels from the non-diabetic first degree relatives through to diabetes. This is theoretically advantageous as there is a continuous distribution of plasma glucose values in the Caucasian population [7], and subjects with impaired glucose tolerance have an increased risk of progression to NIDDM [34-37]. The likelihood of single locus models was significantly improved by considering the FPG diathesis by COMDS in addition to affection status using POINTER or COMDS. With inclusion of the FPG diathesis, a second major locus was suggested but not supported statistically, although a larger sample might reveal such an effect. COMDS was particularly applicable to ana-

lysis of the trait of FPG, as it is difficult to transform it to allow analysis by methods which assume normality e. g., YPOINT developed by Lalouel et al. [27]. Either method could be used to investigate traits associated with diabetes which can be treated as being normal, such as obesity. The mixed model approach of YPOINT provides a more rigorous methodology for confirming or rejecting the hypothesis of single gene inheritance. The two-locus modelling of COMDS is most useful when segregation analysis is to be extended to a linkage analysis, where linkage could be detected to either of the two loci, with the other locus acting as a surrogate for residual familial components. Examples of other disorders to which COMDS analysis has been applied include Graves' disease, where normal subjects were classified into diathesis classes of increasing thyroid autoantibody titre [38], and schizophrenia, where auditory P300 latency has been measured as a possible correlate of the genetic predisposition to schizophrenia [39].

The segregation analysis data reject the autosomal recessive hypothesis for the inheritance of NIDDM in Caucasian families and one of the models, COMDS, also rejected co-dominant inheritance. This finding is consistent with the relatively low reported prevalence of NIDDM in the offspring of Caucasian conjugal diabetic parents, compared with the 100% expected with a recessive model [40–43]. However, the present data do not exclude the possibility that a recessive gene or co-dominant gene may play a crucial role in some pedigrees, in combination with other genetic or environmental factors. An example of the potential interactions that might occur is the report that some patients with extreme insulin resistance are compound heterozygotes for different mutant alleles that impair insulin receptor function by different mechanisms [44]. The parents who were heterozygous carriers demonstrated less severe insulin resistance. In a similar manner, NIDDM in some pedigrees may be due to combinations of mutations in one or more genes. It is possible that NIDDM is inherited in a dominant manner in some families, is recessively inherited in others, and in other families is polygenic. Segregation analysis is not a sensitive tool for the detection of such heterogeneity.

A limiting factor in the analysis of genetic linkage with NIDDM has been the requirement to specify a genetic model. The present study indicates that a dominant model is most applicable, but the possibility of a polygenic component implies that linkage analysis could model more than one locus; when a single locus model is used, lower penetrance parameters should be included [45, 46]. The best fit model with COMDS gave a susceptibility gene frequency 7.4% and the penetrance factors derived in this study may be applied to future linkage analysis. Nevertheless, the likelihood of polygenic probably limits the applicability of formal linkage analysis to classic NIDDM

pedigrees. In any case, the premature disease mortality and late-onset of this disease means that suitable pedigrees with one affected and one unaffected parent are unusual [47]. Thus, linkage analysis is most appropriate in pedigrees with specific monogenic disorders such as MODY [48, 49] although it can be useful in excluding a major dominant gene effects in a series of NIDDM nuclear families [46]. If sufficient numbers of pedigrees are being studied, the two-locus model could be applied, although combined segregation and linkage analysis of plasma glucose levels and diabetes with candidate genes in a given data set is an alternative approach [38].

Other robust methods of analysis are available, but they are less statistically powerful than classic linkage analysis. The affected sibling-pair approach [50, 51] does not require assumptions about the mode of inheritance, but the collection of a suitably large number of NIDDM sibling-pairs with living parents for identity by descent analysis is difficult [47]. In the absence of parental information, the statistical power of sibling-pair analyses is reduced, and affected-pedigree-member analysis based on identity-by-state comparisons requires a large number of sibling-pairs to overcome the possibility of heterogeneity between sibling-pairs.

Genetic heterogeneity has recently been demonstrated within a large pedigree with NIDDM, with some diabetic individuals not having the glucokinase mutation found in other family members [50]. In such cases, linkage in a pedigree may be missed despite a significant role for the gene under consideration. By analogy, in hypertensive patients, mutations of the angiotensinogen gene account for genetic linkage in only a subset of the population study [51]. In the context of anticipated polygenicity and heterogeneity, the direct search for mutations in candidate genes in subjects with NIDDM is likely to prove a valuable alternative approach. Mutations can be detected using the polymerase chain reaction [52] with electrophoresis for single-strand conformation polymorphisms [53] or with heteroduplex scanning [54], followed by direct DNA sequencing. These techniques can be applied to individual patients or specific cohorts chosen for their pathophysiological characteristics (e. g., beta-cell dysfunction or insulin insensitivity), and do not depend on the availability of large pedigrees or on the mode of inheritance.

In conclusion, the segregation analysis of NIDDM in Caucasian pedigrees favoured a dominant model of inheritance, rejected a recessive model and suggested a co-dominant model was unlikely. The analysis was also in accord with a mixed model or a polygenic model. The genetic analysis of NIDDM in Caucasian subjects needs to deal with the potential complexities of polygenicity and genetic heterogeneity.

Acknowledgements. We are grateful to the families for their collaboration, to Ms. B. Barrow and Ms. N. Walravens for their help with glucose tolerance tests, to Ms. M. Burnett, Ms. P. Sutton and Mr. D. Jelfs for their technical assistance. We thank Professor N.E. Morton for advice. We are grateful to the Alan and Babette Sainsbury Charitable Trust for financial support. Dr. J. Cook was a Rhodes scholar.

References

- Barnett AH, Eff C, Leslie RDG, Pyke DA (1981) Diabetes in identical twins: a study of 200 pairs. *Diabetologia* 20: 87–93
- Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD (1987) Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia* 30: 763–768
- Zimmet P, Whitehouse S (1978) Bimodality of fasting and two hour glucose tolerance distributions in a Micronesian population. *Diabetes* 27: 793–800
- 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
- Raper LR, Taylor R, Zimmet P, Milne B, Balkan B (1984) Bimodality in glucose tolerance distributions in the urban Polynesian population of Western Samoa. *Diabetes Res* 1: 19–26
- Unger RH, Grundy S (1985) Hyperglycaemia as an inducer as well as a consequence of impaired islet cell function and insulin resistance: implications for the management of diabetes. *Diabetologia* 28: 119–121
- Thompson GS (1965) Genetic factors in diabetes mellitus studied by the oral glucose tolerance test. *J Med Genet* 2: 221–226
- Tillil H, Richter K, Kobberling J (1985) Bimodal distribution of the two-hour blood glucose value during OGTT among first-degree relatives of type 2 diabetics in a Caucasian population. *Diabetes Res Clin Pract* 1 [Suppl 1]: S560 (Abstract)
- Tattersall RB (1974) Mild familial diabetes with dominant inheritance. *Q J Med* 43: 339–357
- Pincus G, White P (1933) On the inheritance of diabetes mellitus. 1. An analysis of 675 family histories. *Am J Med Sci* 186: 1–14
- Nilsson SE (1964) On the heredity of diabetes mellitus and its interrelationship with some other diseases. *Acta Genetica* 14: 97
- Thompson MW, Watson EM (1952) The inheritance of diabetes mellitus. An analysis of the family histories of 1631 diabetics. *Diabetes* 1: 268–275
- Steinberg AG (1961) Heredity in diabetes mellitus. *Diabetes* 10: 269–274
- Falconer DS, Duncan LJP, Smith C (1971) A statistical and genetic study of diabetes. *Ann Hum Genet* 34: 347–367
- Kobberling J, Tillil H (1982) Empirical risk figures for the first-degree relatives of non-insulin-dependent diabetics. In: Kobberling J, Tattersall R (eds) *The genetics of diabetes mellitus*. Academic Press, London, pp 201–209
- Cook JTE, Hattersley AT, Levy JC, Patel P, Wainscoat JS, Hockaday TDR, Turner RC (1993) The distribution of type 2 diabetes in nuclear families. *Diabetes* 42: 106–112
- Hosker JP, Matthews DR, Rudenski AS, Burnett MA, Darling P, Bown EG, Turner RC (1985) Continuous infusion of glucose with model assessment: Measurement of insulin resistance and Beta-cell function in man. *Diabetologia* 28: 401–411
- Metropolitan Life Insurance Company (1959) National weights standards for men and women. *Stat Bull Metropol Life Insurance Company* 40: 1–4
- Neil HAW, Gatling W, Mather HM, Thompson AV, Thoroughgood M, Fowler GH, Hill RD, Mann JI (1987) The Oxford Community Diabetes Study: Evidence for an increase in the prevalence of known diabetes in Great Britain. *Diabetic Medicine* 4: 539–543
- Welborn TA, Glatthaar C, Whittall D, Bennett S (1989) An estimate of diabetes prevalence from a national population sample: a male excess. *Med J Aust* 150: 78–81
- Neil HAWN, Mant D, Jones L, Morgan B, Mann JI (1990) Lipid screening: is it enough to measure total cholesterol concentration? *BMJ* 301: 584–587
- Hammersley MS, Levy JC, Volpicelli G, Barrow B, Turner RC (1992) Assessment of impaired glucose tolerance and beta-cell function with a continuous infusion of glucose test and an oral glucose tolerance test. *Diabet Med* 9 [Suppl 1]: P17 (Abstract)
- WHO Technical Report Series (1985) No 727 Geneva
- Harris MI, Hadden WC, Knowler WC, Bennett PH (1987) Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in U.S. population aged 20–74 yr. *Diabetes* 36: 523–534
- Lalouel JM, Morton NE (1981) Complex segregation analysis with pointers. *Hum Hered* 31: 312–321
- Morton NE, Shields DC, Collins A (1991) Genetic epidemiology of complex genotypes. *Ann Hum Genet* 55: 301–314
- Lalouel JM, Le Mignon L, Simon M, Fauchet R, Bourel M, Rao DC, Morton NE (1985) Genetic analysis of idiopathic hemochromatosis using both qualitative (disease status) and quantitative (serum iron) information. *Am J Hum Genet* 37: 700–718
- Iselius L, Morton NE (1991) Transmission probabilities are not correctly implemented in the computer program POINTER. *Am J Hum Genet* 49: 459
- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Craki F (eds) *Proceedings of the 2nd International Symposium of Information theory*. Budapest, Academai Kiado, pp 267–283
- Akaike H (1987) Factor analysis and AIC. *Psychometrika* 52: 317–332
- O’Rahilly S, Spivey RS, Holman RR, Nugent Z, Clark A, Turner RC (1987) Type 2 diabetes of early onset: a distinct clinical and genetic syndrome? *Br J Med* 294: 923–928
- Baird JD (1982) Is obesity a factor in the aetiology of non-insulin-dependent diabetes? In: Kobberling J, Tattersall R (eds) *The genetics of diabetes mellitus*. Academic Press, London, pp 233–241
- Panzram G (1987) Mortality and survival in type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia* 30: 123–131
- Jarrett RJ, Keen H, Fuller JH, McCartney M (1979) Worsening to diabetes in men with impaired glucose tolerance (“borderline diabetes”). *Diabetologia* 16: 25–30
- Keen H, Jarrett RJ, McCartney P (1982) The ten-year follow-up of the Bedford survey (1962–1972): glucose tolerance and diabetes. *Diabetologia* 22: 73–78
- Sartor G, Schersten B, Carlstrom S, Melander A, Norden A, Persson G (1980) Ten-year follow-up of subjects with impaired glucose tolerance: prevention of diabetes by tolbutamide and diet regulation. *Diabetes* 29: 41–49
- Charles MA, Fontbonne A, Thibault N, Warnet J-M, Rosselin GE, Eschwege E (1991) Risk factors for type 2 diabetes

- in a white population: Paris prospective study. *Diabetes* 40: 796-799
38. Shields DC, Ratanachaiyavong S, McGregor A, Collins A, Morton NE (1994) Combined segregation and linkage analysis of Graves' disease with a thyroid antibody diathesis. *Am J Hum Genet* 55: 540-554
 39. Sham PC, Morton NE, Muir WJ, et al. (1994) Segregation analysis of complex phenotypes: an application to schizophrenia and auditory P300 latency. *Psychiatric Genetics* 4: 29-38
 40. Cooke AM, Fitzgerald MG, Malins JM, Pyke DA (1966) Diabetes in children of diabetic couples. *BMJ* II: 674-676
 41. Kahn CB, Soeldner JS, Gleason RE, Rojas L, Camerini-Davalos RA, Marble A (1969) Clinical and chemical diabetes in the offspring of diabetic couples. *N Engl J Med* 281: 343-346
 42. Tattersall R, Fajans SS (1975) Diabetes and carbohydrate intolerance in 199 offspring of 37 conjugal diabetic parents. *Diabetes* 24: 452-462
 43. Ganda OP, Soeldner SS (1977) Genetic, acquired and related factors in the aetiology of diabetes mellitus. *Arch Intern Med* 137: 461-469
 44. Kadowaki T, Kadowaki H, Rechler MM, Serrano-Ris M, Roth J, Gorden P, Tatlor SI (1990) Five mutant alleles of the insulin receptor gene in patients with genetic forms of insulin resistance. *J Clin Invest* 86: 254-264
 45. Cook JTE, Patel P, Clark A et al. (1991) Non-linkage of the islet amyloid polypeptide gene with type 2 diabetes. *Diabetologia* 34: 103-108
 46. Cook JTE, Hattersley AT, Christopher P et al. (1992) Linkage analysis of glucokinase gene with NIDDM in Caucasian pedigrees. *Diabetes* 41: 1496-1500
 47. Cook JTE, Page RCL, O'Rahilly S et al. (1992) The availability of NIDDM families for the detection of diabetes susceptibility genes. *Diabetes* 1993; 42: 1536-1543
 48. Froguel PH, Vaxilaire M, Sun F (1992) Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin dependent diabetes mellitus. *Nature* 356: 162-164
 49. Hattersley AT, Turner RC, Permutt MA (1992) Linkage of type 2 diabetes to the glucokinase gene. *Lancet* 339: 1307-1310
 50. Suarez BK (1983) A sib pair strategy for the use of restriction fragment length polymorphisms to study the mode of transmission of type 2 diabetes. *Am J Hum Genet* 35: 34-48
 51. Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. *Am J Hum Genet* 42(2): 315-326
 52. Stoffel M, Patel P, Lo Y-MD et al. (1992) Characterization of a missense glucokinase mutation in maturity-onset diabetes of the young (MODY) and mutation screening in late-onset diabetes. *Nature Genetics* 2: 153-156
 53. Jeunemaitre X, Soubrier F, Kotelevstev et al. (1992) Molecular basis of human hypertension: role of angiotensin. *Cell* 71: 169-180
 54. Saiki RK, Gelfand DH, Stoffel et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487-491
 55. Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5: 874-879
 56. White MB, Carvalho M, Derse D, O'Brien SJ, Dean M (1992) Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 12: 301-306