

## Positive association in the absence of linkage suggests a minor role for the glucokinase gene in the pathogenesis of Type 2 (non-insulin-dependent) diabetes mellitus amongst South Indians

M. I. McCarthy<sup>1</sup>, M. Hitchins<sup>1</sup>, G. A. Hitman<sup>1</sup>, P. Cassell<sup>1</sup>, K. Hawrami<sup>1</sup>, N. Morton<sup>2</sup>, V. Mohan<sup>3</sup>, A. Ramachandran<sup>3</sup>, C. Snehalatha<sup>3</sup>, M. Viswanathan<sup>3</sup>

<sup>1</sup>The Medical Unit, The Royal London Hospital, London, UK

<sup>2</sup>Genetic Epidemiology Group, Princess Anne Hospital, Southampton, UK

<sup>3</sup>Diabetes Research Centre, MV Hospital for Diabetes, Royapuram, Madras, India

**Summary.** Mutations of the glucokinase gene have been implicated in the development of glucose intolerance in pedigrees with maturity-onset diabetes of the young. However, the contribution of the glucokinase gene to the aetiology of common Type 2 (non-insulin-dependent) diabetes mellitus is uncertain. We have studied the role of the glucokinase gene in the pathogenesis of Type 2 diabetes in South Indians, using both population-association and linkage methodology. A pair of CA-repeat sequences (GCK(3') and GCK(5')) straddling the glucokinase gene were employed as markers, each subject being typed using the polymerase chain reaction and polyacrylamide gel electrophoresis. Comparisons of allele frequencies at these markers were made between 168 Type 2 diabetic subjects and 70 racially-matched control subjects. No differences in allele frequencies were apparent at the GCK(5') marker; however, there were significant differences

in allele frequencies at the GCK(3') marker between the Type 2 diabetic subjects and control subjects ( $\chi^2 = 11.6$ ,  $df = 3$ ,  $p = 0.009$ ) with an increase of the z allele (78.0% vs 66.4%) and a decrease of the z + 2 allele (13.7% vs 25.0%) amongst the diabetic subjects. Linkage between glucose intolerance and the glucokinase gene was studied in 53 nuclear pedigrees under a variety of genetic models. Linkage was excluded (lod score  $< -2$ ) at a recombination fraction of zero under five of the ten models used and highly unlikely ( $-2 < \text{lod score} < -1$ ) under the others. The combination of positive association and negative linkage suggests that glucokinase acts as a minor gene influencing the development of Type 2 diabetes within this population.

**Key words:** Type 2 (non-insulin dependent) diabetes mellitus, glucokinase, genetics, linkage.

The importance of genetic influences in the development of Type 2 (non-insulin-dependent) diabetes mellitus is emphasised by a wide variety of twin, family and population studies [1]. However, the specific genetic determinants of glucose intolerance remain largely unknown and no clear consensus has emerged as to the number of different genes which may contribute to glucose intolerance either within an individual or within a population. Despite evidence for a major gene in certain populations at extreme risk of diabetes [2–6] it is likely that more typically, Type 2 diabetes is a genetically heterogeneous condition and that genetic and environmental influences affecting both insulin action and insulin secretion contribute to the development of the diabetic phenotype [7].

Glucokinase had long been considered a plausible candidate gene for a pathogenetic role in Type 2 diabetes [8, 9]. Expressed in liver and beta cells, this enzyme is responsible for the phosphorylation of glucose to glucose 6-phosphate. In contrast to other hexokinases, it has a high  $K_m$  value and is not inhibited by glucose 6-phosphate; thus, rates of glucose phosphorylation within these tissues vary

in tandem with glucose levels within the physiological range. This has led to the suggestion that, within the beta cell, glucokinase is a critical element of the 'glucose sensor', and that, in the liver, glucokinase is an important arbiter of hepatic glucose flux [8, 9].

Recently, unequivocal evidence has been presented [10, 11] demonstrating linkage between the glucokinase gene and glucose intolerance in several French and one British pedigree segregating for maturity-onset diabetes of the young (MODY). Mutations within the coding sequence of the glucokinase gene have been identified in affected individuals within these pedigrees [12–16] and shown to be associated with impaired glucose-stimulated insulin release [17–19].

The relevance of these findings to Type 2 diabetes as a whole is uncertain. The phenotype of subjects with glucokinase-linked glucose intolerance is predominantly one of insulin secretory disturbance [17–19] leading to hyperglycaemia which is typically mild, of early-onset and only slowly progressive with age [19]. The relative contributions of defects in insulin secretion and insulin action to

**Table 1.** Characteristics of subjects used in the population study

Subject group	<i>n</i>	Sex (% male)	Age <sup>a</sup> (years)	Diabetes onset <sup>a,b</sup> (years)	Family history of diabetes (% positive)	BMI <sup>a</sup> (kg/m <sup>2</sup> )	Treatment with	
							Tablets (%) <sup>c</sup>	Insulin (%) <sup>c</sup>
Control	70	53	36.2 (12.8)	–	0	24.4 (2.7)	0	0
Type 1 diabetes	50	66	19.7 (9.8)	15.4 (8.4)	18	17.6 (3.3)	0	100
Type 2 diabetes								
All	168	60	43.0 (12.1)	36.1 (11.5)	75	25.3 (4.2)	91	21
Probands	52	63	35.9 (6.6)	32.4 (6.6)	96	25.9 (4.4)	96	19
Clinic	116	58	46.3 (12.7)	37.9 (12.9)	65	25.0 (4.1)	89	22

<sup>a</sup> For age, onset and BMI results are given as mean (SD). No details available amongst the blood donor control subjects (*n* = 25). <sup>b</sup> Ranges of onset of Type 2 diabetes were 21–47 years (probands) and 11–72 years (clinic). <sup>c</sup> Many Type 2 diabetic subjects treated with insulin were also receiving oral hypoglycaemic agents

the development of Type 2 diabetes has been a matter of considerable debate [7] but subjects with diabetes necessarily have insulin secretion inappropriately low for the prevailing level of glycaemia. A major role for glucokinase in the development of Type 2 diabetes is thus feasible; given the likely heterogeneity of Type 2 diabetes the importance of glucokinase may vary in different ethnic groups.

Population-association studies involving Type 2 diabetic subjects have yielded contradictory results with reports of positive associations in Mauritian Creoles [20] and in US Blacks [21] whereas no association was evident amongst Mauritian Indians [20] and Caucasoids [15, 22]. Linkage analysis of the glucokinase gene in European Type 2 diabetic pedigrees has been reported as negative [23].

In South Indians, the Type 2 diabetic phenotype bears a number of superficial similarities to the phenotype of those individuals with glucokinase-linked glucose intolerance, including a relatively early age of onset [24, 25] and strong familial aggregation [26]. Furthermore, normoglycaemic first-degree relatives of Type 2 diabetic subjects from South India have evidence of a beta-cell defect which may imply a primary abnormality in insulin secretion [27]. In this article, we report studies performed to define the contribution of defects in the glucokinase gene to the development of glucose intolerance in this population.

## Subjects and methods

### Population study

All patient groups used in the population-association studies were unrelated Dravidian (South Indian) subjects recruited via the Diabetes Research Centre in Madras.

Type 2 diabetes was defined by an insidious onset of the disease without ketosis; treatment with insulin was not considered a cause for exclusion provided at least 5 years had elapsed between diagnosis and insulin therapy. The Type 2 diabetic subjects studied were derived from two groups: a) 116 unrelated Type 2 diabetic subjects recruited from the clinic (clinic group), and b) 52 unrelated probands with Type 2 diabetes from the families under study (proband group).

Fifty Type 1 (insulin-dependent) diabetic subjects were so designated on the basis of an acute onset of symptoms before 35 years of age, susceptibility to ketosis and an absolute dependence on exogenously administered insulin; furthermore, to exclude fibrocalculus

pancreatic diabetes (FCPD) all had normal abdominal X-ray and pancreatic ultrasound investigations. Further confirmation of their diagnosis has been afforded by previous studies which indicate that these subjects show the HLA associations typical of Type 1 diabetes [28].

Control subjects were recruited from amongst blood donors (*n* = 25) or from staff and spouses of patients at the MV Hospital for Diabetes (*n* = 45). None had a first-degree family history of diabetes and, in all, random blood glucose estimations were less than 6.7 mmol/l. Characteristics of the different groups are given in Table 1.

### Family study

The nuclear families employed in linkage analysis (*n* = 53) were recruited on the following basis: all individuals were of Dravidian origin and were identified through a proband affected with Type 2 diabetes who had both parents and at least one sibling alive and available for study. Available family members were submitted to clinical assessment including anthropometric measurements and a 75-g oral glucose tolerance test [29]. Where a subject was already receiving oral hypoglycaemic agents or insulin (provided insulin treatment was not instituted within 5 years of diagnosis) a diagnosis of Type 2 diabetes was accepted without repeating a full oral glucose tolerance test, provided a glucose measurement consistent with diabetes had been recorded in the past. Clinical details of the pedigrees are given in Table 2.

Each available family member donated 10 ml of blood which was stored at –20 °C prior to and during transport to the UK.

Paternity of family members was verified by typing all members at a number of polymorphic loci (HLA-DQ $\alpha$ , tyrosine hydroxylase, insulin gene). Two parentage exclusions were identified and these individuals were excluded from analysis.

These studies were approved by the Ethical Committee of the MV Hospital for Diabetes, Madras. All the subjects gave their informed consent to venepuncture.

### Laboratory methods

DNA was extracted from thawed blood either by standard phenol-chloroform methods or by preparation of crude lysates (to a final volume of 250  $\mu$ l from 250  $\mu$ l of thawed blood) [30].

The glucokinase gene on chromosome 7p13 was studied through two polymorphic dinucleotide repeat markers, here termed GCK(3') [31] and GCK(5') [32, 33]. GCK(3') is located 10 kilobase (kb) 3' to the coding region of the glucokinase gene and GCK(5') 6 kb upstream; therefore these markers straddle the coding region of the gene with a physical separation of at least 50 kb [15].

Amplification of the GCK(3') repeat was performed in a 100- $\mu$ l volume containing 45 pmol of each of the primers (9510–5'-CCC ACA CCA AAA CTG CCT GTA TTA G-3' and 9509–5'-TTG

**Table 2.** Characteristics of family members studied

	n	Sex (M:F)	Diagnosis				Age (years)	Age of diabetes onset (years)
			Type 2 diabetes	IGT	Nondiabetic (tested)	Not tested		
Fathers	52 <sup>a</sup>	–	37	7	8	–	64.9 (7.5)	54.0 (12.6)
Mothers	53 <sup>a</sup>	–	42	4	7	–	57.7 (6.9)	48.7 (9.2)
Probands	52 <sup>a</sup>	33:19	52	–	–	–	35.9 (6.6)	32.4 (6.6)
Siblings	206	103:103	25	9	68	104	32.8 (7.4)	35.8 (8.5)

Age and age at onset of diabetes given as mean (SD). IGT, Impaired glucose tolerance. <sup>a</sup> One father and one proband were excluded from analysis as the result of non-paternity

**Table 3.** Parameters of models used for linkage analysis

Model	Description	Affection	Frequency of diabetogenic allele (D)	Maximum penetrance (D/D)	Maximum penetrance (D/d)	Maximum penetrance (d/d)	Major gene effect
1	Dominant	Diabetes	0.057	0.995	0.995	0.275	0.503
2	Additive	Diabetes	0.196	1.000	0.999	0.001	0.961
3	Recessive	Diabetes	0.350	1.000	0.262	0.262	0.794
4	Dominant: reduced penetrance	Diabetes	0.057	0.900	0.900	0.275	–
5	Additive: reduced penetrance	Diabetes	0.200	0.900	0.900	0.050	–
6	Recessive: reduced penetrance	Diabetes	0.350	0.900	0.262	0.262	–
7	Dominant: IGT as affected	Diabetes + IGT	0.087	0.998	0.998	0.416	0.574
8	Additive: IGT as affected	Diabetes + IGT	0.193	1.000	1.000	0.250	0.919
9	Recessive: IGT as affected	Diabetes + IGT	0.474	1.000	0.372	0.372	0.913
10	Dominant: glucokinase based	Diabetes + IGT	0.100	0.990	0.990	0.100	–

Major gene effect refers to the proportion of the variance in liability to affection attributed to the major gene in the segregation analysis. Maximum penetrance refers to penetrance of the affected phenotype in the highest liability class and is shown for homozygotes for

the diabetes-susceptibility gene (D/D), heterozygotes (D/d) and for those possessing no diabetogenic allele (d/d)  
IGT, Impaired glucose tolerance

GTC AGT GTA GGCTGA ACT CAT G-3') [31], 200 µmol/l each dNTP, 50 µmol/l tetramethylammonium chloride (TMAC), 50 mmol/l KCl, 10 mmol/l Tris-HCl (pH 8.8), 1.5 mmol/l MgCl<sub>2</sub>, 0.1 % Triton X-100 and typically 0.1 µl (100 ng) of DNA or 0.5 µl of lysate. After denaturation of template DNA at 94 °C (8 min), 3 U of Taq DNA polymerase (Promega, Southampton, Hants., UK) was added and amplification carried out by 32 cycles of 94 °C for 1 min (denaturation), 55 °C for 1 min (annealing) and 72 °C for 1 min (extension) with a final 10-min extension at 72 °C.

The primers used for amplification of GCK(5') were hGK-CA-1 (5'-AAC AGA TAC GCT TCA TCC TG-3') and hGK-CA-2 (5'-TGT CTG CAA CTT ACT CTT AC-3') [32]. Amplification conditions were similar to those for GCK(3') except in the following respects: reaction volume of 50 µl, no TMAC, 1.5 U of Taq DNA polymerase, 10 pmol of each primer, 30 cycles with annealing at 59 °C (2 min) and extension at 72 °C (2 min).

In each case between 3 and 5 µl of polymerase chain reaction product was loaded onto a 10% non-denaturing polyacrylamide gel and visualised by staining with ethidium bromide (0.5 mg/l for 30 min) after 16–20 h electrophoresis at 350 V (11 W). Alleles were sized by reference to 1 kb marker (BRL, Paisley, Strathclyde, UK).

### Analysis of allele and genotype frequencies

Allele frequencies were compared in the various patient groups using chi-square. The three null hypotheses under test were that the distributions of alleles were: i) no different in control and Type 2 diabetic subjects, ii) no different between the two groups of Type 2 diabetic subjects and iii) no different between control and Type 1 diabetic subjects. The significance of deviations from independence for these comparisons was judged at the 5% level. Linkage disequilibrium was tested by chi-squared analysis of the cross-tabulation of alleles in those haplotypes which could be unequivocally assigned (i.e. all except those (18% of the total) heterozygous at both

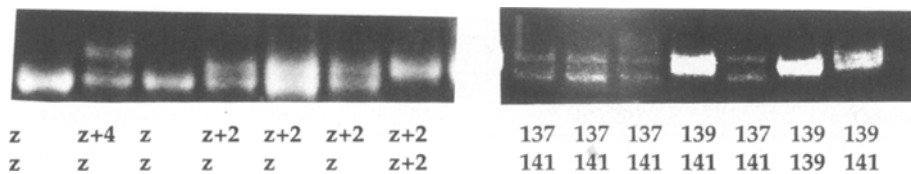
loci). Loglinear analysis was performed on genotype data to explore further the significance of interactions a) between the two marker loci and b) between the marker loci and affection status [34].

Potential allele effects on continuous variables were assessed using parametric statistics (analysis of variance) with logarithmic transformation where appropriate. Non-parametric statistics (Mann-Whitney U) were used when the criteria for parametric analysis could not be established. All the above analyses were performed using SPSS-PC software.

### Linkage analysis

Evidence for linkage was sought using the MLINK and LINKMAP programs of the LINKAGE package [35] accessed remotely at the Human Genome Mapping Project Resource Centre, Harrow, Middx., UK. Liability classes used to code for penetrance were derived from a combination of age, sex and urban/rural habitat, these being the major situational variables determining prevalence in epidemiology-based surveys of Type 2 diabetes in South Indians [24, 25]. The prevalence of diabetes was considered to rise from 0.1% in the lowest liability class (all males and rural/semi-urban females less than 24 years of age) to 35% in liability class 9 (e.g. semi-urban females over 55 years of age).

A set of Dravidian nuclear families in whom the proband was identified without knowledge of the family history ( $n = 48$ ) had previously been used to derive suitable genetic models for Type 2 diabetes in this population, using complex segregation analysis performed using the programs COMDS [36] and POINTER [37] (McCarthy MI et al., unpublished data); estimates of gene frequency and penetrance functions derived for the best-fitting single-gene models were used as parameters to describe the putative diabetes-causing gene (models 1 to 3) (Table 3). However, since many of the models so derived ascribed 100% penetrance to ho-



**Fig. 1.** Examples of results of electrophoresis of polymerase chain reaction products following amplification for GCK(3') (left) and GCK(5') (right). Subjects shown are unrelated. Genotype results are given below each sample. For GCK(3') z represents the most

common allele and others are shown by size relative to z (z + 2, z + 4). For GCK(5') the actual size in base pairs (bp) is given – the most frequent allele is 139 bp

mozygote carriers of the disease-causing allele amongst members of the highest liability class, a figure higher than indicated e.g. by twin studies [1], further models (models 4 to 6) limiting penetrance to 90% and introducing a minimum 5% phenocopy rate throughout the higher liability classes were also used. Other models treating impaired glucose tolerance (IGT) as affected were derived from the preliminary segregation analysis and employed (models 7 to 9). Since GCK mutations in MODY seem to be dominant and associated with early onset of glucose intolerance a further model was introduced derived from the penetrance function employed by Hattersley et al. [11] (model 10): for this model, liability class was recoded solely to reflect age, and affection was considered to incorporate impaired glucose tolerance and diabetes. For all analyses the allele frequencies for GCK(3') and GCK(5') in the control populations were used. All linkage analyses were conducted assuming two different recombination fractions ( $\theta_{3'5'}$ ) between the two markers ( $\theta_{3'5'} = 0$  and  $\theta_{3'5'} = 0.001$ ), the latter chosen to reflect the fact that the two markers are separated by a minimum of 50 kb [15].

The program HOMOG was used to determine any evidence for heterogeneity between the pedigrees [38] although it is recognised that the power to detect heterogeneity is poor when multiple small nuclear families are used [39]. The results of linkage analysis were interpreted in the light of linkage simulation studies with SIMLINK [38, 40] conducted on the same families using a four allele system with polymorphism information content (PIC value) similar to that of the GCK haplotype system. A number of the linkage models were tested to establish the likely range of lod scores obtained for a candidate gene linked to the marker at a recombination fraction ( $\theta_a$ ) of zero under assumptions that the candidate gene was responsible for diabetes in 50%, 75% and 100% of families.

**Results**

*Allele frequencies in control populations*

The alleles identified in South Indian subjects (Fig. 1) were the same as those reported in other Caucasoid groups tested at these loci [22, 31, 32] with no evidence for the larger (z + 10) allele found at the GCK(3') locus in groups with African admixture [20, 21, 31]. The frequencies of alleles at GCK(3') differed between our control group (Table 4) and those of other reported Caucasoid (but Europicid) control groups (Dravidian vs Europicid control subjects [31] ( $n = 142$  chromosomes)  $\chi^2 = 21.8, df = 3, p < 0.001$  and vs Europicid control subjects [22] ( $n = 100$  chromosomes)  $\chi^2 = 14.5, df = 3, p < 0.005$ ). Dravidian control subjects did not differ significantly from Mauritian Indian control subjects ([20] ( $n = 64$  chromosomes)  $\chi^2 = 3.66, df = 3, p = 0.30$ ). At GCK(5') the control population (Table 5) had similar allele frequencies to published frequencies (32) (vs Caucasoid control subjects ( $n = 58$  chromosomes)  $\chi^2 = 6.9, df = 5, p = 0.23$ , vs Asian control subjects ( $n = 28$  chromosomes)  $\chi^2 = 7.3, df = 3, p = 0.06$ ).

Observed polymorphism information content (PIC value) for GCK(3') was 0.43 (heterozygosity = 0.49) and for GCK(5') was 0.41 (heterozygosity = 0.48). This provides a combined PIC of 0.71 (heterozygosity = 0.74) under assumptions of linkage equilibrium.

**Table 4.** Allele frequencies at GCK(3') marker locus in subjects

Allele	Control subjects <i>n</i> (140)	Type 1 diabetes <i>n</i> (100)	Type 2 diabetes <i>n</i> (336)	Clinic group <i>n</i> (232)	Proband group <i>n</i> (104)
z	93 66.4%	72 72%	262 78.0%	181 78.0%	81 77.9%
z + 2	35 25.0%	21 21%	46 13.7%	34 14.7%	12 11.5%
z + 4	11 7.9%	7 7%	28 8.3%	17 7.3%	11 10.6%
z + 6	1 0.7%	–	–	–	–

Numbers (*n*) refer to number of alleles. Significance values: control vs Type 2 diabetic subjects ( $p = 0.009$ ); control vs clinic subjects ( $p = 0.040$ ); control vs proband subjects ( $p = 0.047$ ); controls plus

Type 1 diabetic vs Type 2 diabetic subjects ( $p = 0.015$ ); all other comparisons not significant at 5% level

**Table 5.** Allele frequencies at GCK(5') marker locus in subjects

Allele	Control subjects <i>n</i> (140)	Type 1 diabetes <i>n</i> (100)	Type 2 diabetes <i>n</i> (336)	Clinic group <i>n</i> (232)	Proband group <i>n</i> (104)
127	–	–	1 0.3%	–	1 1.0%
137	6 4.3%	11 11%	29 8.6%	18 7.8%	11 10.6%
139	93 66.4%	57 57%	231 68.8%	165 71.1%	66 63.5%
141	39 27.9%	32 32%	74 22.0%	48 20.7%	26 25.0%
143	2 1.4%	–	1 0.3%	1 0.4%	–

Numbers (*n*) refer to number of alleles. No comparison of subject groups was significant at the 5% level

**Table 6.** Genotype frequencies at the GCK(3') locus

Genotype <i>n</i>	Control subjects (70)	Type 1 diabetes (50)	Type 2 diabetes (168)	Clinic group (116)	Proband group (52)
z,z	34 48.6%	26 52.0%	104 61.9%	71 61.2%	33 63.5%
z,z+2	19 27.1%	14 28.0%	37 22.0%	29 25.0%	8 15.4%
z,z+4	5 7.1%	6 12.0%	17 10.1%	10 8.6%	7 13.5%
z,z+6	1 1.4%	–	–	–	–
z+2,z+2	5 7.1%	3 6.0%	3 1.8%	1 0.8%	2 3.8%
z+2,z+4	6 8.6%	1 2.0%	3 1.8%	3 2.6%	–
z+4,z+4	–	–	4 2.3%	2 1.7%	2 3.8%

Numbers (*n*) refer to numbers of individuals

**Table 7.** Genotype frequencies at the GCK(5') locus

Genotype <i>n</i>	Control subjects (70)	Type 1 diabetes (50)	Type 2 diabetes (168)	Clinic group (116)	Proband group (52)
127, 139	–	–	1 0.6%	–	1 1.9%
137, 137	1 1.4%	1 2.0%	4 2.3%	1 0.8%	3 5.8%
137, 139	1 1.4%	6 12.0%	17 10.1%	13 11.2%	4 7.7%
137, 141	3 4.3%	3 6.0%	4 2.3%	3 2.6%	1 1.9%
139, 139	32 45.7%	17 34.0%	79 47.0%	58 50.0%	21 40.4%
139, 141	26 37.1%	17 34.0%	54 32.1%	35 30.2%	19 36.5%
139, 143	2 2.8%	–	1 0.6%	1 0.8%	–
141, 141	5 7.1%	6 12.0%	8 4.8%	5 4.3%	3 5.8%

Numbers (*n*) refer to number of individuals

### Population-association findings

The frequencies of alleles and genotypes for the different subject groups are shown for GCK(3') in Tables 4 and 6 for GCK(5') in Tables 5 and 7. Genotype frequencies revealed no significant departures from Hardy-Weinberg equilibrium.

At the GCK(3') locus there were significant differences between the allele frequencies amongst control subjects and Type 2 diabetic subjects ( $\chi^2 = 11.6$ ,  $df = 3$ ,  $p = 0.009$ ) with a lower frequency of the z+2 allele (13.7% vs 25%) and higher frequency of the z allele (78.0% vs 66.4%) in the Type 2 diabetic subjects. The two subgroups of Type 2 diabetic subjects did not differ significantly ( $\chi^2 = 1.4$ ,  $df = 2$ ,  $p = 0.49$ ) and both were significantly different from the control group (clinic group  $\chi^2 = 8.3$ ,  $df = 3$ ,  $p = 0.040$ ; proband group  $\chi^2 = 7.9$ ,  $df = 3$ ,  $p = 0.047$ ). Allelic frequencies of subjects with Type 1 diabetes were no different from the control subjects ( $\chi^2 = 1.4$ ,  $df = 3$ ,  $p = 0.70$ ). There were no significant associations between disease and allele frequencies at the GCK(5') locus. Given the previously-published relationship between glucokinase and MODY, a subgroup of 31 individuals were identified within the Type 2 diabetic group by virtue of age at diagnosis of diabetes less than 25 years and a positive first-degree family history of diabetes. The allele frequencies in this subgroup (GCK(3') z, 49 (79.0%); z+2, 6 (9.6%); z+4, 7 (11.3%)) revealed a trend towards a lower frequency of the z+2 allele but there was no significant difference between diabetic individuals meeting these criteria for MODY and those not doing so ( $\chi^2 = 1.7$ ,  $df = 2$ ,  $p = 0.43$ ). Comparison of allele frequencies between control and Type 2 diabetic subjects remained significant after excluding the 31 individuals meeting criteria for MODY ( $\chi^2 = 9.1$ ,  $df = 3$ ,  $p = 0.03$ ).

Haplotypes were assigned wherever possible (i.e. excepting those double heterozygotes in whom haplotypes could not be assigned unequivocally); amongst these haplotypes there was no evidence for linkage disequilibrium between alleles at the two loci (control subjects:  $\chi^2 = 4.0$ ,  $df = 6$ ,  $p = 0.68$ ; Type 2 diabetic subjects:  $\chi^2 = 3.6$ ,  $df = 6$ ,  $p = 0.73$ ; all subjects:  $\chi^2 = 7.3$ ,  $df = 8$ ,  $p = 0.51$ ).

Loglinear analysis of genotype data confirmed these results. The interaction term between GCK(5') genotype and disease status (control vs Type 2 diabetic subjects only) was not a significant contributor to the overall model ( $p = 0.27$ ) in contrast to that between GCK(3') genotype and disease status ( $p = 0.023$ ). The interaction of GCK(3') genotype and GCK(5') genotype was also non-significant ( $p = 0.34$ ) in keeping with linkage equilibrium between the two marker loci.

The effects of alleles on continuous variables including age of onset of Type 2 diabetes, body mass index (kg/m<sup>2</sup>) (control subjects alone and all subjects) and random blood glucose (control subjects only) were studied by comparison of those possessing one or more copy of a particular allele with those not possessing the allele. The age of onset of diabetes differed at the GCK(5') locus (139-positive ( $n = 152$ ) 36.7(11.7) years (mean (SD)) vs 139-negative ( $n = 16$ ) 30.7(8.2) years ( $p = 0.049$ ); 141-positive ( $n = 66$ ) 33.8(11.3) years vs 141-negative ( $n = 102$ ) 37.7(11.5) years ( $p = 0.036$ )), but these differences were not significant after correction for multiple comparisons.

### Linkage analysis

The results of linkage analysis are shown in Table 8. Since there were no obligatory recombinants between GCK(3') and GCK(5') (maximal inter-marker lod score = 9.19 at

**Table 8.** Results of linkage analysis (for models shown in Table 3)

Model	Description	Lod scores at $\theta_d =$					
		0.0	0.001	0.005	0.01	0.05	0.1
1	Dominant	-1.08	-1.06	-0.95	-0.83	-0.19	+0.18
2	Additive	-inf	-3.34	-2.54	-2.13	-0.86	-0.24
3	Recessive	-inf	-3.56	-2.79	-2.42	-1.29	-0.70
4	Dominant: reduced penetrance	-1.14	-1.11	-1.00	-0.88	-0.25	+0.13
5	Additive: reduced penetrance	-1.79	-1.73	-1.55	-1.38	-0.58	-0.13
6	Recessive: reduced penetrance	-1.98	-1.95	-1.84	-1.71	-1.02	-0.56
7	Dominant: IGT as affected	-1.53	-1.50	-1.42	-1.33	-0.80	-0.41
8	Additive: IGT as affected	-2.91	-2.88	-2.75	-2.61	-1.72	-1.03
9	Recessive: IGT as affected	-inf	-1.42	-0.72	-0.43	+0.17	+0.33
10	Dominant: glucokinase based	-3.89	-3.85	-3.72	-3.56	-2.53	-1.64

All analyses shown were performed using MLINK with the recombination fraction ( $\theta_{35}$ ) between the two markers set to zero. IGT, impaired glucose tolerance; inf, infinity

$\theta_{35} = 0$ ) results are shown for an inter-marker distance of zero. However, analysis was also performed with  $\theta_{35} = 0.001$  (i.e. 0.1% recombination) given the likely physical distance between the markers: lod scores (as derived from location scores) were not appreciably altered except in the case of models 2,3 and 9 where the lod scores changed to (values given for lod scores at the two markers)  $-\infty/-3.34$  (model 2),  $-\infty/-3.27$  (model 3) and  $-\infty/-1.42$  (model 9).

Given that glucokinase is a candidate gene, the lod scores at the lowest recombination fractions between disease and marker loci can be used to assess evidence for and against linkage. Significant evidence against linkage between glucokinase and diabetes in these families (cumulative lod score  $< -2.0$  at  $\theta_d = 0$ ) is thus available for 5 of the 10 models employed: linkage is unlikely under the other models (lod score in the range  $-1$  to  $-2$ ).

No evidence for genetic heterogeneity could be adduced using the program HOMOGEN on a selection of those models with the most significantly negative lod scores. For example, using the output from model 2 there was no significant evidence for heterogeneity over homogeneity ( $\chi^2 = 0.8$ ,  $df = 1$ ,  $p = 0.39$ ).

Analysis of the power of this study to exclude linkage (using SIMLINK) indicated that, even when the simulated marker locus was assumed to be linked to a diabetes-susceptibility gene in only 50% of families, the interval of lod scores lying within 3 SD of the mean lod score obtained in the SIMLINK analysis did not include the lod score obtained for the corresponding model using the observed data from the two glucokinase markers.

## Discussion

The results obtained in this study of Type 2 diabetes in South Indian subjects indicate that the glucokinase gene is not a major gene responsible for the majority of the genetic susceptibility to this form of diabetes in South India. However, the evidence of a positive population-association together with negative linkage is consistent with a minor role for glucokinase in this population.

Dravidians are considered on anthropological and genetic grounds to share common ancestry with Europid

and other Caucasoid groups [41] and the fact that the alleles seen at these markers were also those reported in other Caucasoid populations [22, 31, 32] (and distinct from those reported in other ethnic groups [20, 21, 31]) is in keeping with this. These results, therefore, extend the positive findings for the GCK(3') marker observed in Mauritian Creoles [20] and US Blacks [21] to a Caucasoid population for the first time. In contrast to our findings, a study of Indian subjects living in Mauritius has failed to find any association between diabetes and the GCK(3') marker [20]. The Indian community on Mauritius originates from several areas of the subcontinent, including Bombay and Calcutta as well as Madras (and therefore includes Dravidians and non-Dravidians) [42] and these differences in ethnic background offer an explanation for the contrasting findings.

The two markers under study are separated by at least 50 kb and straddle the coding region of the gene [15]. In South Indians these two loci are in linkage equilibrium, as has been suggested in other populations [33]. The positive association is consistent with the interpretation that a mutation in the glucokinase gene, which is in linkage disequilibrium with the GCK(3') but not the GCK(5') marker, contributes to the genetic susceptibility to Type 2 diabetes in South India.

Although population-association studies have the capacity to demonstrate quite subtle genetic effects [43] they are susceptible to a number of well-documented methodological difficulties which may result in false positive findings [1, 43]. Inappropriate matching of the ethnic origin of control and diabetic subjects is one potential cause of spurious positive results. In distinction to the other groups in which positive results have been described, the South Indian population has not been subject to major migration, invasion or interethnic mating in recent generations. Consequently, Dravidians are more likely to represent an ethnically homogeneous group with fewer opportunities for unobserved ethnic differences between the diabetic and control groups. Selection of an appropriate control group is a critical matter in any population-association study, a problem worsened in studies of Type 2 diabetes by the relatively late age of onset of disease and the absence of a suitable biochemical marker of diabetes risk. Choosing a control group which is relatively old (with respect to the typical age of onset of disease)

minimises the possibility that individuals destined to develop diabetes in later life will be included within the control group and thereby reduces the risk of a false negative result. However, if the control group is too old they may represent a group selected for longevity, in which case any genetic effects uncovered may be related to survival rather than to the disease under study. The control subjects studied here had a mean age of 36 years: the mean age of diagnosis of Type 2 diabetes in the outpatient clinic in Madras is 44 years. It is inevitable therefore that the control group includes some individuals who will develop diabetes in the future but their numbers are likely to be modest given the slow progression of glucose intolerance and the normal glucose results obtained at entry to the study, together with the absence of a positive family history. Furthermore, the effect of any such individuals would be to reduce, rather than exaggerate, the magnitude of differences between the groups at the GCK(3') locus.

In the Type 1 diabetic individuals, the allele frequencies observed at the GCK(3') locus were intermediate between those of the control and the Type 2 diabetic subjects. Given the high prevalence of Type 2 diabetes (and hence of any diabetes-susceptibility genes) in South India [24, 25] and the differing genetic bases of the two forms of glucose intolerance, this is entirely consistent with our main findings. The control subjects have been selected to represent a population with a particularly low prevalence of Type 2 diabetes-susceptibility genes, as evidenced by the fact that none has a family history of Type 2 diabetes. The Type 1 diabetic subjects (18% of whom reported a family history of Type 2 diabetes) would be expected to have a prevalence of Type 2 diabetes-susceptibility genes similar to that of the general population, and intermediate between those of the non-diabetic control and the Type 2 diabetic subjects.

The identification of a positive association between the glucokinase marker and Type 2 diabetes has two important corollaries. In each population which has revealed a positive association [20, 21], diabetes has been associated with a different allele at the GCK(3') locus. This indicates that the polymorphic repeat itself is unlikely to have any direct effect on the functioning of the glucokinase gene but instead acts as a marker in linkage disequilibrium with a pathogenic mutation within the coding or regulatory region of the glucokinase gene.

Secondly, the presence of a positive association implies, not only that glucokinase plays a role in common Type 2 diabetes, but also that a single mutation is responsible for a substantial component of that effect within the population under study. If a number of mutations of the glucokinase gene were common within South Indian diabetic subjects, it is likely that each would be in linkage disequilibrium with a different allele at the GCK(3') locus, and that an association with one particular allele would be lost. This is in contrast to the evidence from pedigree studies in British and French Caucasoids [12–16] which have shown that a wide variety of different mutations underlie glucokinase-linked glucose intolerance in MODY, and that many of these mutations are observed in only one pedigree.

In order to evaluate further the positive association detected in Dravidians, linkage analysis was performed on a large number of nuclear Dravidian pedigrees. We found that linkage between the glucokinase gene and glucose intolerance (whether or not defined as including IGT) is highly unlikely under some models and excluded under others.

There are several potential explanations for the combination of positive association with negative linkage observed in South Indian subjects with Type 2 diabetes. The association demonstrated may be spurious, reflecting a type 2 error resulting from sampling bias or methodological deficiencies (e.g. hidden population stratification, survivor effects). Although we have no reason to suspect that any of these applies in this study, we would consider a positive population-association study such as this as a preliminary indicator of likely genetic involvement which requires confirmation, both by population-association studies in other populations and ultimately by other methodologies (by linkage analysis or molecular scanning, or both).

Alternatively, our findings would also be obtained if glucokinase is acting as a minor gene for Type 2 diabetes. Though linkage analysis has undisputed power to identify major genes segregating within genetically-homogeneous families, linkage analysis is highly sensitive to non-allelic genetic heterogeneity; this rapidly increases the number of pedigrees required to demonstrate linkage as the proportion of linked families declines [44, 45]. Linkage analysis is also limited in its capacity to detect the effects of minor or 'susceptibility' loci (alleles which increase risk of a disease but which are neither necessary nor sufficient for disease expression) [46]. The failure to detect positive evidence for linkage in this study may therefore indicate one or both of the following possibilities: a) the glucokinase gene acts as a 'major' gene but only in a minority of families (multigenic heterogeneity) and/or b) the glucokinase gene acts as a minor 'susceptibility' locus in a substantial number of pedigrees (polygenic heterogeneity). Either of these could result in positive association but negative linkage. Tests of homogeneity, such as that implemented by HOMOG [38], do, in principle, allow an assessment of multigenic heterogeneity by testing the hypothesis that the pedigrees under study can be divided into those linked and those not linked to the locus concerned. However, the power of this method is poor when small nuclear pedigrees are employed [39] since each pedigree makes only a small (and individually statistically insignificant) contribution to the cumulative lod score. The negative results of homogeneity tests here may therefore reflect insufficient power to demonstrate multigenic heterogeneity rather than its absence. No firm conclusions can therefore be drawn as to the manner in which the glucokinase locus may exert a 'minor' effect, although the linkage simulation studies using SIMLINK effectively exclude the possibility that the glucokinase gene contributes as a major gene to diabetes-susceptibility in more than 50% of the families under study.

In summary, we have demonstrated that in South Indian subjects, an association exists between a marker adjacent to the glucokinase gene and Type 2 diabetes. Link-

age analysis indicates that glucokinase is not a major gene contributing to the development of diabetes in this population, although the minor role in the development of Type 2 diabetes suggested by the positive association is not excluded by the results of the linkage analysis. Molecular screening techniques, such as single-stranded conformational polymorphism (SSCP)[47], will permit us to screen the glucokinase gene for pathogenic mutations and more precisely define the relationship of the glucokinase gene to Type 2 diabetes in this population.

*Acknowledgements.* This study was supported by grants from the Wellcome Trust, and the Central Research Fund of the University of London. M.McC. is a Medical Research Council Training Fellow. We acknowledge helpful discussions with Drs. A. Hattersley and R. Turner (Oxford, UK).

## References

- Hitman GA, McCarthy MI (1991) Genetics of non-insulin dependent diabetes mellitus. *Bailliere Clin Endocrinol Metab* 5: 455–476
- Serjeantson SW, Zimmet P (1991) Genetics of non-insulin dependent diabetes mellitus in 1990. *Bailliere Clin Endocrinol Metab* 5: 477–493
- 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
- Zimmet P, Whitehouse S (1978) Bimodality of fasting and two-hour glucose tolerance distributions in a Micronesian population. *Diabetes* 27: 793–800
- 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 Res* 1: 19–26
- 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
- Turner R, O'Rahilly S, Levy J, Rudenski A, Clark A (1989) Does type II diabetes arise from a major gene defect producing insulin resistance or  $\beta$ -cell dysfunction? In: Nerup J, Mandrup-Poulsen T, Hökfelt B (eds) *Genes and gene products in the development of diabetes mellitus*. Elsevier, Amsterdam, pp 171–183
- Magnuson MA (1990) Glucokinase gene structure: functional implications of molecular genetic studies. *Diabetes* 39: 523–527
- Matschinsky FM (1990) Glucokinase as glucose sensor and metabolic signal generator in pancreatic beta-cells and hepatocytes. *Diabetes* 39: 647–652
- 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
- Hattersley AT, Turner RC, Permutt MA et al. (1992) Linkage of type 2 diabetes to the glucokinase gene. *Lancet* 339: 1307–1310
- Vionnet M, Stoffel M, Takeda J et al. (1992) Nonsense mutation in the glucokinase gene causes early-onset non insulin-dependent diabetes mellitus. *Nature* 356: 721–723
- Stoffel M, Patel P, Lo YMD et al. (1992) Missense glucokinase mutation in maturity-onset diabetes of the young and mutation screening in late-onset diabetes. *Nature Genetics* 2: 153–156
- Vionnet N, Stoffel M, Takeda J et al. (1992) Mutations in glucokinase gene cause early onset type 2 diabetes. *Diabetologia* 35 [Suppl 1]: A62 (Abstract)
- Permutt MA, Chiu KC, Tanizawa Y (1992) Glucokinase and NIDDM: a candidate gene that paid off. *Diabetes* 41: 1367–1372
- Froguel Ph, Vionnet N, Stoffel M et al. (1992) Different phenotypic expression by three mutant alleles of glucokinase gene in MODY. *Diabetologia* 35 [Suppl 1]: A63 (Abstract)
- Velho G, Froguel Ph, Clement K et al. (1992) Primary pancreatic beta-cell secretory defect caused by mutations in glucokinase gene in kindreds of maturity onset diabetes of the young. *Lancet* 340: 444–448
- Page RCL, Hattersley A, Turner R (1992) Beta-cell secretory defect caused by mutations in the glucokinase gene. *Lancet* 340: 1162 (Letter)
- Page RCL, Hattersley AT, Barrow B et al. (1992) Clinical characteristics of type 2 diabetes linked to the glucokinase gene. *Diabetologia* 35 [Suppl 1]: A62 (Abstract)
- Chiu KC, Province MA, Dowse GK et al. (1992) A genetic marker at the glucokinase gene locus for Type 2 (non-insulin-dependent) diabetes mellitus in Mauritian Creoles. *Diabetologia* 35: 632–638
- Chiu KC, Province MA, Permutt MA (1992) Glucokinase gene is genetic marker for NIDDM in American Blacks. *Diabetes* 41: 843–849
- Saker PJ, Hattersley AT, Patel P et al. (1992) The contribution of glucokinase to type 2 (non-insulin-dependent) diabetes – a population association study. *Diabetologia* 35 [Suppl 1]: A139 (Abstract)
- Cook JTE, Hattersley AT, Christopher P et al. (1992) Linkage analysis of glucokinase gene with NIDDM in Caucasian pedigrees. *Diabetes* 41: 1496–1500
- 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
- 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
- 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
- Snehalatha C, Mohan V, Ramachandran A, Jayashree R, Viswanathan M (1984) Pancreatic beta cell function in offspring of conjugal diabetic parents. Assessment by IRI and C-peptide ratio. *Horm Metab Res* 16 [Suppl]: 142–144
- Hitman GA, Karir PK, Sachs JA et al. (1988) HLA-D region RFLPs indicate that susceptibility to insulin-dependent diabetes mellitus in south India is located in the HLA-DQ region. *Diabetic Med* 5: 57–60
- World Health Organisation Study Group (1985) *Diabetes Mellitus*. WHO Tech Rep Ser, no 727, Geneva
- Balnaves ME, Nasioulas S, Dahl HHM, Forrest S (1991) Direct PCR from CVS and blood lysates for detection of cystic fibrosis and Duchenne muscular dystrophy deletions. *Nucleic Acids Res* 19: 1155
- Matsutani A, Janssen R, Donis-Keller H, Permutt MA (1992) A polymorphic (CA)<sub>n</sub> repeat element maps the human glucokinase gene (GCK) to chromosome 7p. *Genomics* 12: 319–325
- 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
- Tanizawa Y, Matsutani A, Chiu KC, Permutt MA (1992) Human glucokinase gene: isolation, structural characterization, and identification of a microsatellite repeat polymorphism. *Mol Endocrinol* 6: 1070–1081
- Norusis MJ (1990) *SPSS/PC+ advanced statistics 4.0*. SPSS Inc. USA, pp B137–B158
- Lathrop GM, Lalouel JM (1984) Easy calculation of lod score and genetic risk on small computers. *Am J Hum Genet* 36: 460–465
- Morton NE, Shields DC, Collins A (1991) Genetic epidemiology of complex phenotypes. *Ann Hum Genet* 55: 301–314
- Lalouel JM, Morton NE (1981) Complex segregation analysis with pointers. *Hum Hered* 31: 312–321
- Ott J (1991) *Analysis of human genetic linkage (revised edition)*. Johns Hopkins University Press, Baltimore



39. Martinez MM, Goldin LR (1989) The detection of linkage and heterogeneity in nuclear families for complex disorders: one versus two marker loci. *Am J Hum Genet* 44: 552–559
40. Ploughman LM, Boehnke M (1989) Estimating the power of a proposed linkage study for a complex genetic trait. *Am J Hum Genet* 44: 543–551
41. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain A (1988) Reconstruction of human evolution: bringing together genetic, archeological and linguistic data. *Proc Natl Acad Sci USA* 85: 6002–6006
42. Dowse GK, Gareeboo H, Zimmet PZ et al. (1990) High prevalence of NIDDM and impaired glucose tolerance in Indian, Creole and Chinese Mauritians. *Diabetes* 39: 390–396
43. Cox NJ, Bell GI (1989) Disease associations: chance, artifact or susceptibility genes? *Diabetes* 38: 947–950
44. 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
45. Gershon ES, Martinez M, Goldin LR, Gejman PV (1990) Genetic mapping of common diseases: the challenges of manic-depressive illness and schizophrenia. *Trends Genet* 6: 282–287
46. Greenberg DA (1993) Linkage analysis of “necessary” disease loci versus “susceptibility” loci. *Am J Hum Genet* 52: 135–143
47. 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

Received: 21 December 1992  
and in revised form: 4 March 1993

Dr. M. McCarthy  
Medical Unit  
The Royal London Hospital  
London E1 1BB  
UK