 Human Molecular Genetics, 1997, Vol. 6, No. 9 **1401–1408**

A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene

Habib Zouali[1,2,+], El Habib Hani[2,+], Anne Philippi[3], Nathalie Vionnet[2], Jacques S. Beckmann[4], Florence Demenais[3,§] and Philippe Froguel[2,§,*]

[1]Fondation Jean Dausset—CEPH, 27 rue Juliette Dodu, 75010 Paris, France, [2]CNRS EP10, Institut Pasteur de Lille, 1 rue du Professeur Calmette, 59019 Lille Cedex, France, [3]INSERM-U358, Hôpital Saint Louis, 1 avenue Claude Vellefaux, 75010 Paris, France and [4]Généthon, 1 rue de l'Internationale, BP 60, 91002 Every Cedex, France

Received November 13, 1996; Revised and Accepted June 20, 1997

Several candidate genes for non-insulin-dependent diabetes mellitus (NIDDM) map on chromosome 20, including the phosphoenolpyruvate carboxykinase gene (*PCK1*) and one of the maturity onset diabetes of the young genes (*MODY1*). Thus, we have investigated the entire long arm of chromosome 20. Linkage analyses were conducted in a total sample of 148 NIDDM families (301 NIDDM sib pairs) and in a subset of 42 early onset NIDDM families, where genetic components are likely to play a more important role (55 NIDDM sib pairs diagnosed at or before 45 years of age), using 10 highly polymorphic markers with an average map density of 7.5 cM. Using affected sib pair methods (two-point linkage and multipoint linkage analyses), significant results were obtained with the 20q13 region, in the vicinity of the *PCK1* locus, only in the subset of 55 early onset NIDDM sib pairs (multipoint MLS = 2.74, *P* = 0.0004; MLS = 2.34, *P* = 0.0009 when using a conservative weighting procedure). Moreover, another region spanning the ribophorin II (*RPNII*), phospholipase C *(PLC1)* and adenosine deaminase (*ADA*) loci suggested linkage with NIDDM (multipoint MLS of 1.81 in all NIDDM sib pairs, *P* = 0.003; MLS = 1.31, *P* = 0.012 when using a conservative weighting procedure). Whereas our study suggests the location of a susceptibility locus for early onset NIDDM in the *PCK1* gene region, further investigation in larger data sets is required to confirm these results and assess the role of other regions on chromosome 20q in human NIDDM.

INTRODUCTION

It is widely accepted that non-insulin-dependent diabetes mellitus (NIDDM) is a heterogeneous metabolic disorder with a complex pattern of inheritance. NIDDM exhibits a combination of two major alterations, insulin resistance and pancreatic β-cell dysfunction, but the primary defect is still largely unknown (1). Candidate gene approaches and genome-wide scans are powerful tools to search for NIDDM susceptibility loci. In this regard, recent genome-wide searches led to mapping of major susceptibility loci for NIDDM to chromosome 2q in Mexican-Americans and to chromosome 12q in Finnish families (2,3). However, these two chromosomal regions do not appear to be major contributors to NIDDM in the French NIDDM families, illustrating the difficulties of mapping common multifactorial traits (4,5).

On the other hand, and so far, the intensive search for NIDDM candidate genes potentially involved in insulin secretion and/or insulin action has only provided clues for limited subtypes of NIDDM. This has led to the identification of rare Mendelian sub-entities and few genetic variants having separately relatively modest effects on insulin secretion and/or insulin action, therefore accounting for no more than 5% of the familial clustering of NIDDM (6–9). Insulin resistance is considered to be the earliest detectable clinical defect in the pre-diabetic state (10) and segregation analyses in several ethnic groups strongly suggest that major loci may be involved in the familial clustering of insulin resistance (11–13). Therefore, investigating genes of the insulin pathways may potentially help to elucidate some forms of human NIDDM. The cytosolic phosphoenolpyruvate carboxykinase gene (*PCK1*) located on chromosome 20q13.3 (14) encodes a key regulatory enzyme (PEPCK) in liver glucose metabolism. This enzyme governs the rate limiting step of gluconeogenesis and thus in part hepatic glucose output. As demonstrated in transgenic mice, over-expression of PEPCK enzyme leads to hyperglycemia, insulin resistance and NIDDM (15), and a similar mechanism may be expected to cause insulin resistance with increased hepatic gluconeogenesis in human NIDDM. Thus, *PCK1* is a strong candidate gene for NIDDM, which is characterized by abnormally increased and non-suppressible hepatic glucose production (1,10).

*To whom correspondence should be addressed. Tel: +33 3 20 87 79 11; Fax: +33 3 20 87 72 29; Email: froguel@xenope.univ-lille2.fr
[+]The two first authors contributed equally to this work
[§]The two last authors were equally responsible for this work

Preliminary genetic studies of several candidate genes for insulin resistance showed some indication of linkage between a *PCK1* intragenic microsatellite marker and NIDDM in a limited set of 42 French NIDDM families with multiple cases of NIDDM (16). It is noteworthy that two other diabetogenes also map to chromosome 20q, both centromeric to the *PCK1* locus, namely the *MODY1* gene, causing a subtype of maturity onset diabetes of the young, and the agouti signaling protein gene (*ASP*), which is the human homolog of a murine gene responsible for an obesity–diabetes syndrome in mice (17,18). This prompted us to investigate the entire chromosome 20q in our collection of French Caucasian NIDDM families. A sample of 148 families, ascertained through at least two NIDDM siblings (including the original set of 42 multiple case NIDDM families), was genotyped for 10 polymorphic markers spanning a 67 cM region on the long arm of chromosome 20q with an average map density of 7.5 cM.

We report here suggestive evidence for linkage of early onset NIDDM to the chromosome 20q13 region, in the vicinity of the *PCK1* locus in a subset of 55 NIDDM sib pairs diagnosed at or before 45 years of age. In addition, our data indicate a possible linkage of NIDDM to the 20q12–13.11 region, proximal to *PCK1*, spanning the ribophorin II (*RPNII*), phospholipase C (*PLC1*) and adenosine deaminase (*ADA*) loci.

RESULTS

This study was conducted in a total sample of 148 multiplex NIDDM pedigrees, including 156 sibships with a maximum of 301 NIDDM sib pairs available for genotyping. Affected status was defined as having overt NIDDM according to the WHO criteria (individuals with fasting plasma glucose >7.8 mM and/or >11 mM after a 2 h glucose load and/or if treated by oral hypoglycemic agents or by insulin at least 1 year after diagnosis) (19). Since NIDDM is a heterogeneous disorder and, as reported in a previous study, the contribution of the genetic component to NIDDM decreases as the age at diagnosis increases (20), we selected a subset of 42 sibships including at least two siblings with NIDDM diagnosed at or before the age of 45 years, where genetic factors may be easier to detect. This subset of families, designated in this paper 'early onset NIDDM', included a total of 55 early onset NIDDM sib pairs. We did not consider those sib pairs diagnosed after the age of 45 years separately, as these are probably a mixture of early and late onset cases (early onset diabetic patients remaining frequently undiagnosed for a long period of time). Indeed, analysis of this late-onset subset would increase the number of statistical tests without any likely gain with respect to the whole sample. Available parents and siblings (affected and unaffected) were genotyped for 10 markers with an average heterozygosity of 0.77 ± 0.06 (mean ± SD), estimated from the data (Fig. 1).

Two-point linkage analyses were conducted using three different sib pair methods which may display different statistical properties (21): the Haseman–Elston method (22,23), based on the estimated mean proportion of alleles shared identical by descent (IBD) among affected sib pairs (π); a method comparing the number of shared and non-shared alleles among affected sib pairs using only informative meioses (Joe Terwilliger, personal communication); the maximum likelihood method (MLS test), based on the estimated IBD probabilities with constraints imposed by the 'possible triangle' (24,26).

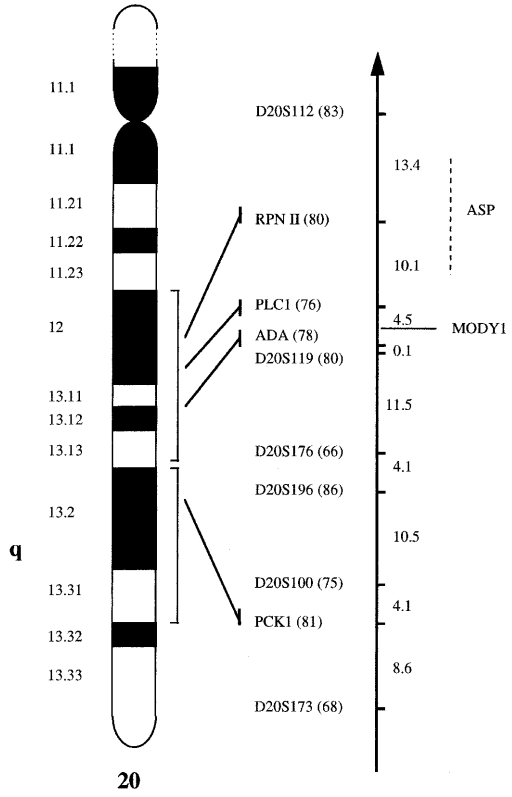

Figure 1. Ideogram of chromosome 20 showing the 10 markers used in the linkage study. The order of markers and genetic distances in cM were obtained from the Généthon map (recombination fractions between adjacent markers estimated from our data agree with those reported on that map). The percent heterozygosity of each marker is indicated in parentheses next to the marker name. The assumed locations of the *MODY1* gene and human homolog of the mouse *agouti* gene, agouti signaling protein gene (*ASP*), are also shown.

In all NIDDM sib pairs (Table 1) the most significant results in favor of linkage were obtained with the *RPNII* locus ($P = 0.001$ without weighting, $P = 0.009$ with weighting), whichever sib pair method was used. Nominal evidence for linkage ($P < 0.05$) was also observed with the *PLC1* marker, the *ADA* marker and the *PCK1* locus (10, 15 and 45 cM distal to *RPNII*, Fig. 1). When considering the early onset NIDDM sib pairs (Table 2), an excess of allele sharing was observed with the following markers, from centromere to telomere: *D20S112* ($0.003 < P < 0.03$, depending on the method used), *RPNII* ($0.005 < P < 0.020$), *ADA* ($0.0002 < P < 0.004$), *D20S196* ($0.004 < P < 0.009$), *D20S100* ($0.004 < P < 0.015$) and *PCK1* ($0.001 < P < 0.01$). To determine empirical *P* values, which may be more accurate in small samples, simulations of 5000 replicates of the data were conducted under the null hypothesis of no linkage with three loci, *RPNII*, *ADA* and *PCK1*. The empirical *P* values calculated using the Haseman–Elston method were very close to those obtained from the observed data using the one-sided *t*-test distribution. The empirical *P* values versus the *P* values provided by the observed data were 0.002 versus 0.001 for *PCK1*, 0.0002 versus 0.0002 for *ADA* and 0.007 versus 0.005 for *RPNII* in early onset NIDDM pairs and 0.002 versus 0.001 for *RPNII* in all NIDDM pairs. Analyses by the shared/not shared allele method led to similar results.

Table 1. Sib pair analyses of 301 NIDDM sib pairs,using three different methods

Locus	n[a]	$\pi \pm$ SD[b]	P[b]	Shared/not shared[c]	P[c]	MLS[d]	P[d]
D20S112	292	0.52 ± 0.30	0.159	254.3/233.4	0.172	0.23	0.211
				181.3/167.1	*0.223*	*0.17*	*0.255*
RPNII	270	0.55 ± 0.29	0.001	233.1/173.2	0.001	2.27	0.001
				173.1/132.3	*0.009*	*1.39*	*0.009*
PLC1	290	0.53 ± 0.31	0.048	246.5/210.9	0.048	1.12	0.019
				177.7/151.3	*0.073*	*0.83*	*0.040*
ADA	288	0.53 ± 0.28	0.023	229.4/191.0	0.030	0.98	0.027
				168.3/143.0	*0.076*	*0.57*	*0.079*
D20S119	290	0.52 ± 0.27	0.061	219.9/190.6	0.073	0.61	0.070
				164.7/144.2	*0.121*	*0.37*	*0.140*
D20S176	276	0.51 ± 0.26	0.253	181.8/168.9	0.246	0.12	0.304
				136.4/127.9	*0.300*	*0.06*	*0.377*
D20S196	284	0.51 ± 0.32	0.329	245.9/235.8	0.322	0.22	0.214
				178.4/170.1	*0.329*	*0.17*	*0.250*
D20S100	281	0.52 ± 0.28	0.106	224.2/199.9	0.119	0.41	0.122
				168.8/157.4	*0.264*	*0.23*	*0.210*
PCK1	294	0.53 ± 0.28	0.035	250.2/215.8	0.055	0.71	0.055
				182.4/158.6	*0.098*	*0.46*	*0.105*
D20S173	273	0.50 ± 0.27	1.000	178.0/181.7	0.923	0.005	0.537
				128.3/138.6	*0.764*	*0.00*	*0.600*

[a]*n* is the total number of affected sib pairs genotyped for a given marker.
[b]Estimated mean proportion ± SD of IBD alleles and the corresponding *P* value, using the Haseman–Elston method.
[c]Number of shared and non-shared IBD alleles among affected sib pairs computed from informative meioses (observed or inferred) and the corresponding *P* value (Terwilliger's approach).
[d]Maximum LOD score using the possible triangle and the corresponding *P* value (Risch and Holmans' approach).
The results in italics shown on the second line for each marker are obtained when weighting for non-independent sib pairs from multiple affected sibships, as made possible by the second and third approaches.

Table 2. Sib pair analyses of NIDDM diagnosed at or before 45 years of age using three different methods, in a maximum of 55 sib pairs, available for genotyping

Locus	n[a]	$\pi \pm$ SD[b]	P[b]	Shared/not shared[c]	P[c]	MLS[d]	P[d]
D20S112	54	0.59 ± 0.25	0.003	53.3/32.6	0.013	1.63	0.005
				45.2/29.0	*0.030*	*1.22*	*0.014*
RPNII	54	0.61 ± 0.30	0.005	53.8/30.4	0.005	1.59	0.006
				44.9/26.8	*0.016*	*1.09*	*0.020*
PLC1	55	0.57 ± 0.33	0.061	54.0/38.5	0.053	0.85	0.038
				44.5/36.1	*0.176*	*0.37*	*0.139*
ADA	53	0.64 ± 0.27	0.0002	57.1/27.1	0.0005	2.84	0.0003
				47.4/24.9	*0.004*	*1.91*	*0.002*
D20S119	55	0.58 ± 0.29	0.029	49.5/33.0	0.035	0.85	0.038
				41.5/30.1	*0.088*	*0.44*	*0.112*
D20S176	53	0.54 ± 0.27	0.105	38.9/28.8	0.109	0.54	0.085
				34.1/26.0	*0.146*	*0.42*	*0.118*
D20S196	53	0.61 ± 0.31	0.006	55.4/31.7	0.006	1.72	0.004
				47.6/27.3	*0.009*	*1.55*	*0.006*
D20S100	53	0.60 ± 0.27	0.004	50.7/29.1	0.008	1.55	0.006
				43.4/25.4	*0.015*	*1.31*	*0.012*
PCK1	55	0.60 ± 0.24	0.001	57.8/34.6	0.008	2.04	0.002
				50.3/29.6	*0.010*	*1.92*	*0.003*
D20S173	50	0.50 ± 0.29	1.000	33.3/33.9	0.969	0.05	0.404
				27.4/31.2	*0.808*	*0.00*	*0.600*

[a]*n* is the total number of affected sib pairs genotyped for a given marker.
[b]Estimated mean proportion ± SD of IBD alleles and the corresponding *P* value, using the Haseman–Elston method.
[c]Number of shared and non-shared IBD alleles among affected sib pairs computed from informative meioses (observed or inferred) and the corresponding *P* value (Terwilliger's approach).
[d]Maximum LOD score using the possible triangle and the corresponding *P* value (Risch and Holmans' approach).
The results in italics shown on the second line for each marker are obtained when weighting for non-independent sib pairs from multiple affected sibships, as made possible by the second and third approaches.

Following these two-point linkage analyses, multipoint mapping, based on the MLS method (25), was carried out on the two sets of NIDDM sib pairs, using the whole map of 10 markers. In all NIDDM pairs the MLS reached 1.81 (1.31 when weighted, $0.003 < P < 0.012$) at a location linked to *RPNII*, with a 1 LOD unit confidence interval of 28 cM, contained within the region spanning the loci from *D20S112* to *ADA* (Fig. 2). In the 55 early onset NIDDM pairs the peak MLS was 2.74 (2.34 when weighted, $0.0004 < P < 0.0009$) at a location 7 cM proximal to *PCK1*, with a 1 LOD unit confidence interval of 24 cM, spanning the *D20S176*, *D20S196*, *D20S100* and *PCK1* loci (Fig. 3).

Given this latter result, a direct involvement of *PCK1* in early onset NIDDM was suggested. Since the *PCK1* gene is essentially regulated in a negative manner by insulin, we screened for mutations in the *PCK1* gene promoter sequences. Thus, a 579 bp segment located at the 5′-end of the structural gene and containing the necessary regulatory elements accounting for the normal pattern of gene expression, especially inhibition by insulin, was directly screened by SSCP in early onset NIDDM patients. However, we did not find any DNA polymorphism in the probands studied (data not shown).

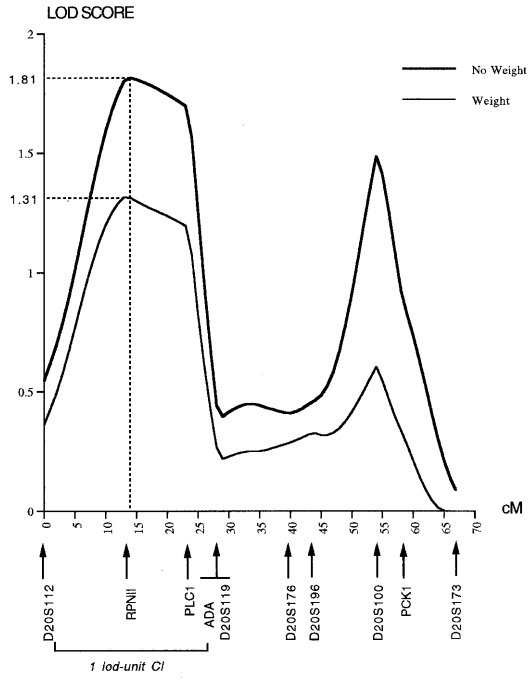

Figure 2. Multipoint LOD scores at different locations of the putative NIDDM susceptibility gene in the total set of NIDDM sib pairs with and without weighting non-independent sib pairs.

DISCUSSION

This study shows suggestive evidence for linkage of early onset NIDDM with a 24 cM region on 20q13, including the *PCK1* locus. Measurement of the contribution of each sib pair to the multipoint MLS of 2.74 at the *PCK1* locus indicates that the LOD scores corresponding to the initial set of 25 sib pairs and the second set of 30 sib pairs are 1.85 ($P = 0.003$) and 0.88 ($P = 0.035$) respectively, both in favor of linkage. It is common usage to weight non-independent sib pairs from sibships with *s* affecteds if $s > 2$. However, a recent simulation study showed that the weighting procedure, which multiplies the LOD score of each sib pair by $2/s$, as implemented in the computer program Mapmaker/Sibs, leads to conservative tests, especially when parents are not genotyped (27). Since analyses were repeated on two data sets, the early onset sample and the whole sample, it is also common usage to multiply the *P* values by the number of samples analyzed to correct for multiple testing. The *P* value of 0.0004 associated with the MLS of 2.74 would thus become equal to 0.0008, still close to the conservative threshold of suggestive linkage ($P = 0.0007$) recommended by Lander and Kruglyak for a dense map in a complete genome scan (28).

We have chosen *a priori* the cut-off point of 45 years of age to define an early onset subset, where a genetic component is likely to play a more important role, as was done in our previous linkage analyses of NIDDM (4,29). However, analyses were repeated by increasing this cut-off point to 46 and 47 years, since true age at onset of NIDDM is usually not accurately known, the latter corresponding to the mean age of NIDDM diagnosis in the whole sample. The multipoint MLS is 2.32 ($P = 0.001$) at a position 9 cM proximal to *PCK1* when using 46 years as a cut-off point (59 sib pairs) and 1.47 ($P = 0.008$) at a position 14 cM proximal to *PCK1* when using 47 years as a cut-off point (64 sib pairs). In the latter case, there is another maximum of the same order of magnitude (MLS = 1.53, $P = 0.007$) tightly linked to *RPNII*. There is thus a trend to a decrease in the MLS in the *PCK1* region and a rise in another MLS in the *RPNII* region as NIDDM cases with an increasing age at diagnosis are taken into account in the analysis.

Linkage of early onset NIDDM to the chromosome 20q13 region may suggest a possible genetic predisposing defect of the phosphoenolpyruvate carboxykinase gene. It is known that *PCK1* gene expression is regulated by multiple hormones and, especially, is inhibited by insulin. In this regard, the development of NIDDM could be due to variations in the *PCK1* gene *cis*-regulatory elements, rather than to mutations in the coding regions impairing PEPCK activity, as PEPCK deficiency is one of the causes of hypoglycemia in infancy (30). SSCP screening of the regulatory sequences required for normal expression of the *PCK1* gene failed to reveal any DNA abnormality in the early onset NIDDM patients. Similarly, Ludwig *et al.* have recently screened the regulatory sequences of the *PCK1* gene for mutations (SSCP and DNA direct sequencing) and found no DNA variation either in Caucasian patients with early onset or late onset NIDDM or in obese Pima Indian patients (31).

In addition to the *PCK1* gene region, other loci on chromosome 20q showed some evidence for linkage to NIDDM. Multipoint mapping indicated a possible linkage of NIDDM to a 28 cM region on 20q12–13.11, extending from *D20S112* to *ADA* and spanning the *RPNII* and *PLC1* loci (MLS = 1.81 close to *RPNII*, $P < 0.003$). Interestingly, potential candidate genes for NIDDM map near the *RPNII* and *ADA* loci: indeed, the *MODY1* gene is tightly linked to the *ADA* locus (32). Recently this *MODY1* gene has been identified as the hepatocyte nuclear factor-4α encoding gene (*HNF-4α*), with a nonsense mutation co-segregating with the *MODY* phenotype in a large Caucasian pedigree (33). However, investigating the *HNF-4α* gene by direct sequencing in our NIDDM probands showed that *HNF-4α* gene mutations are unlikely to be a common cause of NIDDM in our population (E.H.Hani *et al.*, manuscript in preparation). Furthermore, *RPNII*

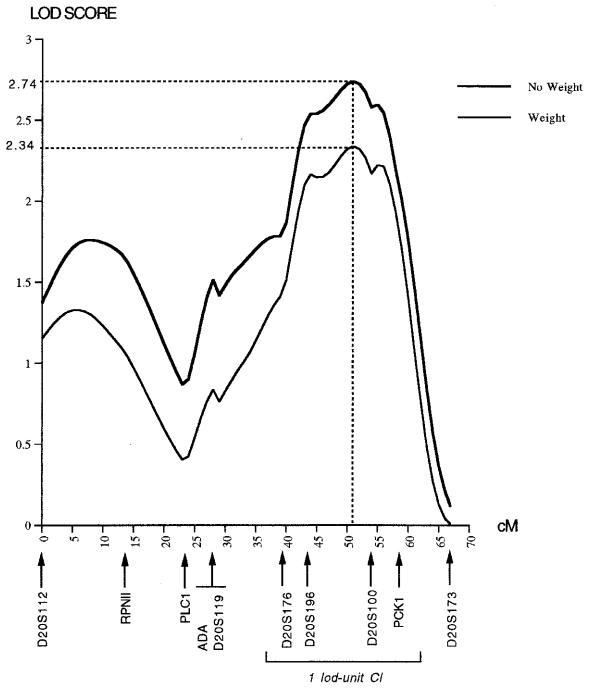

Figure 3. Multipoint LOD scores at different locations of the putative NIDDM susceptibility gene in 55 early onset NIDDM sib pairs with and without weighting non-independent sib pairs.

is located to an interval encompassing <5 cM along with the human homolog of the *agouti* gene, responsible for a form of adult onset diabetes, insulin resistance and obesity in mice (32). It is noteworthy that several additional candidate genes for NIDDM are also located on chromosome 20q: the *PLC1* gene, encoding an intracellular second messenger involved in signal transduction pathways; the CCAAT/enhancer binding protein β gene (*C/EBP*-β), encoding a transcription factor that regulates cAMP-induced expression of the *PCK1* gene in the liver (34,35). Therefore, the chromosome 20q region may contain one or more genes causing monogenic forms of NIDDM and/or contributing to the common (polygenic) forms of NIDDM.

In this regard, two North American studies have recently reported linkage between NIDDM and markers on chromosome 20q (36,37). The first, conducted in large Caucasian pedigrees with multiple NIDDM cases, showed suggestive evidence for linkage in a subset of 14 families with age at diagnosis of NIDDM >47 years (36). The multipoint non-parametric linkage score reached a maximum of 3.3 ($P < 0.009$) for two tightly linked markers, *D20S178* and *D20S197*, located ~8 cM centromeric of *D20S196* (37). Moreover, one allele of the *D20S197* marker was found to be transmitted from heterozygous parents to NIDDM offspring more frequently than expected ($P < 0.01$), suggesting a possible linkage disequilibrium between this allele and a disease locus and making the *MODY1* gene (8 cM centromeric of *D20S197*) an unlikely candidate for NIDDM in these families. The second study, conducted in African-American and Caucasian pedigrees with multiple NIDDM affected members and a history of diabetic nephropathy, suggested linkage with the same markers, *D20S178* and *D20S197*, in only 53 Caucasian NIDDM sib pairs (multipoint MLS = 1.48, $P = 0.008$) (38). As compared with our own results, the peak multipoint scores in the two North American studies occur at a position which is ~15 cM centromeric of the peak obtained in our early onset subset and 22 cM telomeric of the peak observed in our whole sample. The confidence intervals surrounding these peaks may be wide and we found that the centromeric bound of the l LOD unit confidence interval of the maximum LOD score obtained in our early onset subgroup corresponded to the position of the *D20S197* and *D20S178* markers. Moreover, whereas evidence for linkage was stronger in the late onset subset of families (mean age at NIDDM diagnosis ranging from 48 to 63 years) in the former study, the average age of diagnosis of all Caucasian NIDDM patients with a history of nephropathy in the latter study was 45.7 years, which is in closer agreement with our findings in the early onset subset. A meta-analysis of these different samples may be useful to clarify these issues and assess whether one or more NIDDM susceptibility genes are located on chromosome 20q.

Stern *et al.* have recently performed a segregation analysis on Mexican-Americans that showed evidence for a major locus influencing early age of onset of NIDDM (39). Therefore, the early onset NIDDM subgroup may be more homogeneous than the late onset forms of NIDDM, which could increase the power of linkage analysis. In this case, investigating these pedigrees may help to map NIDDM susceptibility genes, at least in European Caucasians, where confounding factors like severe obesity (which may reveal earlier and prematurely worsen diabetes in individuals with a family history of diabetes) are not very prevalent.

In summary, these results suggest the location of a susceptibility locus for early onset NIDDM in the *PCK1* region but confirmation in larger data sets is required. Whether this locus and other loci on chromosome 20q are involved in NIDDM requires further investigation in larger data sets.

MATERIALS AND METHODS

Subjects

From an initial pool of 278 multiplex NIDDM families, we have selected 46 multiple case pedigrees. Family information was collected in France through a multimedia campaign and clinical data for each family member were obtained during a standardized clinical examination performed at the Endocrinology Department of the Hôpital Saint Louis, Paris, or by the subject's personal physician. Four of these families were excluded from the study because both parents had NIDDM. Subjects were considered affected if they presented NIDDM as defined by WHO criteria (19). The 42 selected families presented NIDDM cases in at least two consecutive generations. To increase the sample size, a second set of 116 NIDDM families (smaller in size than those in the first set), with at least one NIDDM sib pair, was selected. Ten families with both parents affected were excluded and only 106 families were included in the study, making a total of 148 families when added to the initial set. Criteria used to select our family samples comprised exclusion of linkage to the *MODY3* locus on chromosome 12q (4) and absence of mutations in the glucokinase gene and a $tRNA^{Leu}$ mutation in mitochondrial DNA when testing at least one affected individual from each family. In the total set of 148 NIDDM families studied there were 393 family members presenting NIDDM; the first set of 42 NIDDM families and the total set have been previously described (4,16). All 393 NIDDM subjects were tested for fasting glucose levels, and 16%

of the affected members from these families who were not being treated for NIDDM consented to oral glucose tolerance testing. Among all NIDDM patients of the total set, 69% were being treated with oral hypoglycemic agents and 15% with insulin. The total sample of 148 families included 156 sibships with at least two NIDDM cases available for genotyping, making an overall sample size of 301 NIDDM sib pairs. Since the genetic component is suggested to play a more important role in early onset NIDDM, we selected a subset of 42 sibships with at least two sibs with NIDDM diagnosed at or before age 45 years, including a total of 55 early onset NIDDM sib pairs (39% of affected family members in the whole sample were diagnosed for NIDDM at or before the age of 45 years). The cut-off point of 45 years of age was chosen as in our previous linkage studies (4,29), since true age at onset of NIDDM is difficult to know accurately and 45 years seemed a reasonable choice to define an actual early onset subgroup. Characteristics of NIDDM patients in this subset and the whole sample of sib pairs are as follows: mean age at diagnosis of NIDDM 36.9 ± 6.9 in the early onset group and 47.7 ± 11.8 in the whole set; proportions of males among all patients 57 and 51% respectively (not significantly different, $P = 0.47$); mean body mass indices (BMI) similar, being respectively 27.1 ± 4.0 and 27.0 ± 4.4. Moreover, comparison of the early onset (55 NIDDM sib pairs diagnosed at or before age 45) and late onset (246 NIDDM sib pairs diagnosed after age 45) subgroups did not show any significant difference in terms of sex ratio, mean BMI, waist/hip ratio, blood glucose and insulin levels or cholesterol and triglycerides levels. The proportions of families with no, one or two parents genotyped are respectively 77, 19 and 4% in the whole sample of 156 families and 69, 26 and 5% in the subset of 42 early onset families. Moreover, the distributions of these sibships according to the number of affected sibs s available for genotyping is as follows: the total of 156 sibships includes 109 sibships with $s = 2$, 37 with $s = 3$, six with $s = 4$, three with $s = 5$ and one with $s = 6$, while the 42 early onset sibships include 37 sibships with $s = 2$, four with $s = 3$ and one with $s = 4$. The number of genotyped sib pairs differed slightly according to the marker studied and are presented in Tables 1 and 2.

Genotyping

DNA from individuals of NIDDM families was extracted from peripheral blood leukocytes by standard procedures (40). All family members were genotyped by PCR using highly polymorphic simple tandem repeat DNA polymorphisms (STRPs) located within or in the vicinity of the *PCK1*, *ADA*, *PLC1* and *RPNII* genes as previously described (41–44). Six Généthon anonymous markers (*D20S173*, *D20S112*, *D20S119*, *D20S176*, *D20S100* and *D20S196*) were also investigated (45). The PCR reactions were carried out with unlabeled primers and dNTPs and the PCR products were separated by electrophoresis on a 6% denaturing polyacrylamide gel containing 8 M urea. Electrophoresed PCR products were then transferred by blotting onto Hybond N^+ nylon membranes which were then hybridized in a non-radioactive labeling procedure using one of the primers used in DNA amplification (46).

Linkage analyses

Evidence for linkage was sought using three non-parametric affected sib pair methods which were shown to display different statistical properties (21). The Haseman–Elston sib pair method (22,23) consists of estimating the mean proportion of alleles shared identical by descent (IBD) among affected sib pairs, using all marker information available in parents and siblings (affected and unaffected). This mean proportion is tested against the expected value of 0.5, under the null hypothesis of no linkage, by a one-sided t-test with degrees of freedom (df) equal to $(n - 2)$ (n being the total number of affected pairs). This method is implemented in the computer program Sibpal of the Sage package (47). The second method, based on the number of shared and non-shared alleles among affected sib pairs computed from informative meioses (heterozygous parents and IBD status unambiguously determined in sib pairs) either observed or inferred from marker information in the data (J.Terwilliger, personal communication). The number of shared and non-shared alleles are equal under the null hypothesis of no linkage, whereas there are more shared than non-shared alleles when there is linkage. This is tested by a one-sided χ^2 test with 1 df. This method has been recently implemented in the program Sibpair of the Analyze package (distributed by J.Terwilliger). The third method is a maximum likelihood method which estimates the IBD probabilities among affected sib pairs from observed marker information and compares it with the expected distribution (1/4, 1/2, 1/4) when there is no linkage using a LOD score test (MLS method; 24). The power of this method can be increased by imposing constraints on IBD probabilities based on the 'possible triangle' (25) and 2ln(10) times the MLS is distributed as a mixture of χ^2 values with 1 df and 2 df. The MLS was computed using the computer program Mapmaker/Sibs (25). The latter two methods, as implemented in the computer programs, make it possible to correct for the non-independence of multiple affected sib pairs from the same sibship by weighting the contribution of each sib pair by $2/s$ (s being the number of affected sibs). To compute empirical P values, marker data were simulated in affected sib pairs under the null hypothesis of no linkage using observed marker information in parents and unaffected siblings. Multipoint mapping was conducted using the MLS approach and the computer program Mapmaker/Sibs with and without weighting sib pairs from multiple affected sibships (26).

Marker allelic frequencies were estimated using all family information from the data set with the ILink program of the Linkage package (48).

Table 3. PCR amplification primers for SSCP screening of the *PCK1* gene promoter

Fragment	Forward primer	Reverse primer	Position	Fragment size (bp)
1	5′-ACAGAGCAGACAATCAATAC-3′	5′-TAGCAGCACATTTTGTGTAC-3′	–547/–348	199
2	5′-GACTGTGACCTTTGACTATG-3′	5′-CCAACTGACTAAACCTTGAC-3′	–434/–246	188
3	5′-TGACCCACCTGCCTGTTAAG-3′	5′-CAACTCACTGCAACACGCCC-3′	–322/–137	185
4	5′-CAATGATTATCTCCCTGGGG-3′	5′-CAGCAAGTTTGTGTTCCCAG-3′	–174/+32	206

LOD score calculations

In this study we have genotyped the *PCK1* and *RPNII* markers in the CEPH pedigrees and positioned them with respect to a subset of polymorphic markers using multipoint mapping analyses with the Linkage package of programs (48).

SSCP screening of the *PCK1* gene promoter

A 579 bp segment of the *PCK1* promoter region (from –547 to +32), encompassing all the *PCK1* regulatory elements, including the insulin response element (49), was subdivided into four small overlapping segments (<250 bp, Table 3). Each fragment of this regulatory sequence was screened for mutations by PCR amplification followed by SSCP analysis in 41 affected individuals with NIDDM diagnosed at or before 45 years of age and 40 unrelated normoglycemic control individuals. Genomic DNA extracted using the standard procedures was amplified in a total volume of 25 µl under the following PCR conditions: an initial denaturation at 94°C for 6 min, 30 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 30 s for all fragments except fragment 2 (amplified at 58°C in a 35 cycle PCR reaction) and a final extension step at 72°C for 6 min. For SSCP analysis, 1 µl PCR product was denatured in 1 µl formamide loading buffer and then directly loaded onto a preprepared polyacrylamide gel (Pharmacia, PhastSystem™) and run under each fragment's optimized conditions. Gels were silver stained and checked for the presence of an abnormal pattern of migration of the PCR fragments.

ACKNOWLEDGEMENTS

We thank our patients for their cooperation. We are indebted to Dr G.I.Bell for his helpful reading and suggestions on the manuscript and to L.Zekiri for secretarial assistance. This work was supported by the French Ministry of Research, the Centre National de la Recherche Scientifique, the Institut de Recherches Internationales—Servier, the Institut National de la Santé et de la Recherche Médicale, the Direction of Clinical Research of Assistance Publique-Hôpitaux de Paris and Glaxo-Welcome. This work was supported partially by the BIOMED II program no. BMH4-CT950662 and the Groupe de Recherches et d'Etudes sur les Génomes. The results in this work were obtained using the SAGE program package, which is supported by a US Public Health Service Resource Grant (1-P41-RR03655). We are also grateful to GIS Infobiogen for providing us with additional computer facilities.

REFERENCES

1. De Fronzo,R.A. (1988) The triumvirate, β-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes*, **37**, 667–687.
2. Hanis,C.L., Boerwinkle,E., Chakraborty,R., Ellsworth,D.L., Concannon,P., Stirling,B., Morrison,V.A., Wapelhorst,B., Spielman,R.S., Godolin-Ewens,K.J., Shepard,J.M., Williams,S.R., Risch,N., Hinds,D., Iwasaki,N., Ogata,M., Omori,Y., Petzold,C., Rietzch,H., Schroder,H.E., Schulze,J., Cox,N.J., Menzel,S., Boriraj,V.V., Chen,X., Lim,L.R., Lindner,T., Mereu,L.E., Wang,Y.-Q., Xiang,Y., Yang,Y. and Bell,G.I. (1996) A genome-wide search for human non-insulin-dependent (type2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nature Genet*., **13**, 161–166.
3. Mahtani,M.M., Widen,E., Lehto,M., Thomas,J., McCarthy,M., Brayer,J., Bryant,B., Chan,G., Daly,M., Forsblom,C., Kanninen,T., Kirby,A., Kruglyak,L., Munnelly,K., Parkkonen,M., Reeve-Daly,M., Weaver,A., Brettin,T., Duyk,G., Lander,E.S. and Groop,L.C. (1996) Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nature Genet*., **14**, 90–94.
4. Lesage,S., Hani,E.H., Philippi,A., Vaxillaire,M., Hager,J., Passa,P., Demenais,F., Froguel,P. and Vionnet,N. (1995) Linkage analyses of the MODY3 locus on chromosome 12q with late-onset NIDDM. *Diabetes*, **44**, 1243–1247.
5. Hani,E.H., Hager,J., Philippi,A., Demenais,F., Froguel,P. and Vionnet,N. (1997) Mapping NIDDM susceptibility loci in French families: studies with markers in the region of *NIDDM1* on chromosome 2q. *Diabetes*, **46**, 1225–1226.
6. Froguel,Ph., Zouali,H., Vionnet,N., Velho,G., Vaxillaire,M., Sun,F., Lesage,S., Stoffel,M., Takeda,J., Passa,P., Permutt,M.A., Beckmann,JS., Bell,G.I. and Cohen,D. (1993) Familial hyperglycemia due to mutations in glucokinase: definition of a subtype of diabetes mellitus. *New Engl. J. Med*., **328**, 697–702.
7. Hager,J., Hansen,L., Vaisse,C., Vionnet,N., Philippi,A., Poller,W., Velho,G., Carcassi,C., Contu,L., Julier,C., Cambien,F., Passa,P., Lathrop,M., Kindsvogel,W., Demenais,F., Nishimura,E. and Froguel,P. (1995) A missense mutation in the glucagon receptor gene is associated with non-insulin-dependent diabetes mellitus. *Nature Genet*., **9**, 299–304.
8. Baier,L.J., Sacchettini,J.C., Knowler,W.C., Eads,J., Paolisso,G., Tataranni, P.A., Mochizuki,H., Bennett,P.H., Bogardus,C. and Prochazka,M. (1995) An amino acid substitution in the human intestinal fatty acid binding is associated with increased fatty acid binding, increased fat oxidation and insulin resistance. *J. Clin. Invest*., **3**, 1281–1287.
9. Clément,K., Vaisse,C., Manning, B.S., Basdevant,A., Guy-Grand,B., Ruiz,J., Silver,K.D., Shuldiner,A.R., Froguel,P. and Strosberg,A.D. (1995) Genetic variation in the beta-3-adrenergic receptor and an increased capacity to gain weight in patients with morbid obesity. *New Engl. J. Med*., **333**, 352–354.
10. Beck-Nielsen,H. and Groop,L.C. (1994) Metabolic and genetic characterization of prediabetic states. *J. Clin. Invest*., **94**, 1714–1721.
11. Martin,B.C., Warram,J.H., Krolewski A.S., Bergman,R.N., Soeldner,J.S and Kahn,C.R. (1992) Role of glucose and insulin resistance in the development of type 2 diabetes mellitus: results of a 25-year follow-up study. *Lancet*, **340**, 925–929.
12. Schumacher,M.C., Hasstedt,S.J., Hunt,S.C., Williams,R.R. and Elbein,S.C. (1992) Major gene effect for insulin levels in familial NIDDM pedigrees. *Diabetes*, **41**, 416–423.
13. Mitchell,B.D., Kammerer,C.M., Hixson,J.E., Atwood,L.D., Hackleman,S., Blangero,J., Haffner,S.M., Stern,M.P. and MacCluer,J.W. (1995) Evidence for a major gene affecting postchallenge insulin levels in Mexican-Americans. *Diabetes*, **44**, 284–289.
14. Stoffel,M., Xiang,K.S., Espinosa,R., Cox,N.J., Le Beau,M.M. and Bell,G.I. (1993) cDNA sequence and localization of polymorphic human cytosolic phosphoenolpyruvate carboxykinase gene (*PCK1*) to chromosome 20, band q13.31: *PCK1* is not tightly linked to maturity-onset diabetes of the young. *Hum. Mol. Genet*., **2**, 1–4.
15. Valera,A., Pujol,A., Pelegrin,M. and Bosch,F. (1994) Transgenic mice overexpressing phosphoenolpyruvate carboxykinase develop non-insulin-dependent diabetes mellitus. *Proc. Natl. Acad. Sci. USA*, **91**, 9151–9154.
16. Hani,E.H., Zouali,H., Vionnet,N., Philippi,A., Demenais,F. and Froguel,P. (1996) Linkage of human non-insulin-dependent diabetes mellitus to the phosphoenolpyruvate carboxykinase gene locus chromosome 20. *Eur. J. Hum. Genet*., **4** (suppl. 1), 147 (abstract).
17. Bell,G.I., Xiang,K.S., Newman,M.W., Wu,S.H., Wright,L.G., Fajans,S.S., Spielman,R.S. and Cox,N.J. (1991) Gene for non-insulin dependent diabetes mellitus (maturity-onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. *Proc. Natl. Acad. Sci. USA*, **88**, 1484–1488.
18. Wilson,B.D., Ollmann,M.M., Kang,L., Stoffel,M., Bell, G.I. and Barsh,G.S. (1995) Structure and function of ASP, the human homologue of the mouse *agouti* gene. *Hum. Mol. Genet*., **2**, 223–230.
19. WHO Expert Committee on Diabetes Mellitus (1980) *Second Report, Technical Report Series 646*. World Health Organization, Geneva, Switzerland.
20. Thomas,F., Balkau,B., Vauzelle-Kervroedan,F., Papoz,L. and the CODIAB–INSERM–ZENEKA Study Group. (1994) Maternal effect and familial aggregation in NIDDM—the CODIAB study. *Diabetes*, **43**, 63–67.

21. Tores,F., Barrillot,A., Philippi,A., Vaysseix,G., Demenais,F. and Martinez,M. (1996) Affected sib-pair strategies for late onset diseases: type I error and marker allele frequencies. *Genet. Epidemiol*., **13**, 302 (abstract).
22. Haseman,J.K. and Elston,R.C. (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet*., **2**, 3–19.
23. Elston,RC. (1984) Genetic analysis workshop II: sib pair screening tests for linkage. *Genet. Epidemiol*., **1**, 175–78.
24. Risch,N. (1990) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am. J. Hum. Genet*., **46**, 242–253.
25. Holmans,P. (1993) Asymptotic properties of affected-sib-pair linkage analysis. *Am. J. Hum. Genet*., **52**, 362–374.
26. Kruglyak,L. and Lander,S.E. (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet*., **57**, 439–454.
27. Meunier,F., Philippi,A., Martinez,M. and Demenais.,F. (1997) Affected sib pairs tests: type I errors with dependent sib pairs. *Genet. Epidemiol*., in press.
28. Lander,E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet*., **11**, 241–247.
29. Vionnet,N., Hani,E.H., Lesage, S., Philippi,A., Hager,J., Varret,M., Stoffel,M., Tanizawa,Y., Chiu,K.C., Glaser,B., Permutt,M.A., Passa,P., Demenais,F. and Froguel,P. (1997) Genetics of non insulin dependent diabetes mellitus (NIDDM) in France: studies with nineteen candidate genes in affected sib-pairs. *Diabetes*, **46**, 1062–1068.
30. Fiser,R.H.,Jr, Melsher,H.L. and Fisher,D.A. (1974) Hepatic phosphoenolpyruvate carboxykinase (PEPCK) deficiency: a new cause of hypoglycaemia in childhood. *Pediat. Res*., **8**, 432 (abstract).
31. Ludwig,S.D., Vidal-Puig,A., O'Brien,R.M, Printz,R.L., Granner,D.K., Moller,D.E. and Flier,J.S. (1996) Examination of the phosphoenolpyruvate carboxykinase gene promoter in patients with non-insulin-dependent diabetes mellitus. *J. Clin. Endocrinol. Metab*., **81**, 503–506.
32. Stoffel,M., LeBeau,M.M., Espinosa,R., Bohlander,S.F., Le paslier,D., Cohen,D., Xiang,K.S., Cox,N.J., Fajans,S.S. and Bell,G.I. (1996) A yeast artificial chromosome-based map of chromosome 20 containing the diabetes-susceptibility gene, *MODY1*, and a myeloid leukemia related gene. *Proc. Natl. Acad. Sci. USA*, **93**, 3937–3941.
33. Yamagata,K., Furuta,H., Oda,N., Kaisaki,P.J., Menzel,S., Cox,N.J., Fajans,S.S., Signorini,S., Stoffel,M. and Bell,G.I. (1996) Mutations in the hepatocyte nuclear factor-4α gene in maturity-onset diabetes of the young (MODY1). *Nature*, **384**, 458–460.
34. Hendricks-Taylor,R.L., Bachinski,L.L., Siciliano,M.J., Fertitta,A., Trask,B., de Jong,P.J., Ledbetter,D.H. and Darlington,G.J. (1992) The CCAAT/enhancer binding protein a (C/EBPa) gene (CEBPA) maps to human chromosome 19q13.1 and the related nuclear factor NF-IL6 (C/EBPb) gene (CEBPB) maps to human chromosome 20q13.1. *Genomics*, **14**, 12–17.
35. Park,E.A., Gurney,A.L., Nizielski,S.E., Hakimi,P., Cao,Z., Moorman,A. and Hanson,R.W. (1993) Relative roles of CCAAT/enhancer-binding protein beta and cAMP regulatory element-binding protein in controlling transcription of the gene for phosphoenolpyruvate carboxykinase (GTP). *J. Biol. Chem*., **268**, 613–619.
36. Ji,L., Malecki,M., Warram,H.J., Yang,Y., Rich,S.S. and Krolewski,A.S. (1997) New susceptibility locus for NIDDM is localized to human chromosome 20q. *Diabetes*, **46**, 876–881.
37. Kruglyak,L., Daly,M.J., Reeve-Daly,M.P. and Lander,E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet*., **58**, 1347–1363.
38. Bowden,W.D. *et al*. (1997) Linkage of genetic markers on human chromosomes 20 and 12 to NIDDM in Caucasian sib pairs with a history of diabetic nephropathy. *Diabetes*, **46**, 882–886.
39. Stern,M.P., Mitchell,B.D., Blangero,J., Reinhart,L., Krammerer,C.M., Harrison,C.R, Shipman,P.A., O'Connell, P., Frazier,M.L. and MacCluer,J.W. (1996) Evidence for a major gene for type II diabetes and linkage analyses with selected candidate genes in Mexican-Americans. *Diabetes*, **45**, 563–568.
40. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
41. Ting,C.N., Burgess,D.L., Chamberlain,J.S., Keith,T.P., Falls,K. and Meisler,M.H. (1993) Phosphoenolpyruvate carboxykinase (GTP): characterisation of the human *PCK1* gene and localization distal to MODY on chromosome 20. *Genomics*, **16**, 698–706.
42. Rothschild,C.B., Akots,G., Hayworth,R., Pettenati,M.J., Rao,P.N., Wood,P., Stolz,F.-M., Hansmann,I., Serino,K., Keith,T.P., Fajans,S.S. and Bowden,D.W. (1993) A genetic map of chromosome 20q12-q13.1: multiple highly polymorphic microsatellite and RFLP markers linked to the maturity-onset diabetes of the young (MODY) locus. *Am. J. Hum. Genet*., **52**, 110–123.
43. Rothschild,C.B., Akots,G., Fajans,S.S. and Bowden,D.W. (1992) A microsatellite polymorphism associated with the *PLC1* (phospholipase C) locus: identification, mapping, and linkage to MODY locus on chromosome 20. *Genomics*, **13**, 560–564.
44. Stoffel,M., Xiang,K. and Bell,G.I. (1992) Dinucleotide repeat polymorphism at the human ribophorin II locus on chromosome 20q. *Hum. Mol. Genet*., **1**, 656.
45. Gyapay,G., Morissette,J., Dib,C., Fizames,P., Millasseau,P., Marc,S., Bernardi,G., Lathrop,G.M. and Weissenbach,J. (1994) The 1993–1994 Généthon human genetic linkage map. *Nature Genet*., **7**, 246–339.
46. Vignal,A., Gyapay,G., Hazan,J., Nguyen,S., Dupraz,C., Cheron,N., Becuwe,N., Tranchant,M. and Weissenbach,J. (1993) A non-radioactive multiplex procedure for genotyping of microsatellite markers. In Adolph,K.W. (ed.), *Methods in Molecular Genetics: Gene and Chromosome Analysis*. Academic Press, San Diego, CA, pp. 211–221.
47. Elston,R.C., Bailey-Wilson,J.E., Bonney,G.E., Keats,B.J. and Wilson,A.F. (1986) A package of computer programs to perform statistical analysis in genetic epidemiology. *The Seventh International Congress of Human Genetics*, Berlin, Germany, September 22–26.
48. Lathrop,G.M. and Lalouel,J.M. (1984) Easy calculations of Lod scores and genetic risks on small computers. *Am. J. Hum. Genet*., **36**, 460–465.
49. O'Brien,R.M., Printz,R.L., Halimi,N., Tiesinga,J.J. and Granner,D.K. (1995) Structural and functional analysis of the human phosphoenolpyruvate carboxykinase gene promoter. *Biochem. Biophys. Acta*, **1264**, 284–288.