

# Searching for Type 2 Diabetes Genes on Chromosome 20

Marshall Alan Permutt, Jonathan Wasson, Latisha Love-Gregory, Jiyan Ma, Gary Skolnick, Brian Suarez, Jennifer Lin, and Benjamin Glaser

Genome scans in families with type 2 diabetes identified a putative locus on chromosome 20q. For this study, linkage disequilibrium mapping was used in an effort to narrow a 7.3-Mb region in an Ashkenazi type 2 diabetic population. The region encompassed a 1-logarithm of odds (LOD) interval around the microsatellite marker D20S107, which gave a LOD score of  $>3$  in linkage analysis of a combined Caucasian population. This 7.3-Mb region contained 25 known and 99 predicted genes. Predicted single nucleotide polymorphisms (SNPs) were chosen from public databases and validated. Two SNPs were unique to the Ashkenazi. Here, 91 SNPs with a minor allele frequency of  $\geq 10\%$  were genotyped in pooled DNA from 150 case subjects and 150 control subjects of Ashkenazi Jewish descent. The SNP association study showed that SNP rs2664537 in the *TIX1* gene had a significant *P* value of 0.035, but the finding did not replicate in an additional case pool. In addition, *HNF4a* and *Mybl2* were screened for mutations and new polymorphisms. No mutations were identified, and a new nonsynonymous SNP (R687C in exon 14 of *Mybl2*) was found. The limits to this type of association study are discussed. *Diabetes* 51 (Suppl. 3): S308–S315, 2002

**T**ype 2 diabetes is a complex metabolic disorder characterized by abnormal hepatic glucose output, insulin resistance, and impaired insulin production (1). Multiple environmental and genetic factors contribute, and genome scans in families with multiple affected individuals from several racial/ethnic groups have been undertaken (Table 1). A number of potential loci have been identified, but in general, the evidence for linkage has not been strong, and the regions identified have been quite broad. A major question remaining is how to proceed with the search for complex disease genes, knowing that a single gene is neither necessary nor sufficient and that recombinant mapping in families will not suffice. The next phase of the search for diabetes susceptibility genes will likely require new strategies.

A genome scan with microsatellite markers at an average distance of 9.5 cM was completed in type 2 diabetic

sibling pairs ( $n = 472$ ) of Ashkenazi Jewish descent (2). The Ashkenazi population is relatively young and homogeneous, having undergone several constrictions and expansions resulting in reduced genetic heterogeneity in comparison to that of most Western Caucasian subjects. Studies of DNA polymorphisms have suggested that present-day Ashkenazi Jews descended from a small founder population, numbering perhaps as few as 10,000 individuals who existed in Eastern Europe at about 1500 AD (3). Today, there are about 10 million, representing a 1,000-fold expansion in roughly 20 generations. In the genome scan, five regions on four chromosomes exhibited nominal evidence for linkage ( $P < 0.05$ ) (2). A maximal signal of  $Z = 2.05$  was observed on chromosome 20 near D20S195. Because four other groups had previously reported evidence for linkage in the same region of chromosome 20q in Caucasians (see summary in Table 1), this region was further considered.

Subsequent to the Ashkenazi type 2 diabetes genome scan, several investigators contributed linkage data for chromosome 20 to the International Type 2 Diabetes Linkage Analysis Consortium (<http://www.sfbr.org/external/diabetes>), a collaborative effort organized principally by the National Institute of Diabetes and Digestive and Kidney Diseases to map genes for the disease. Common markers on chromosome 20 were genotyped, and results were combined for a total of 1,852 families. The initial results of this analysis suggested a peak of linkage at D20S107 with a logarithm of odds (LOD) of  $>3$ . We therefore decided to target this region in our search in Ashkenazi patients with type 2 diabetes.

Gene mapping by linkage disequilibrium analysis in related families identifies broad conserved regions of chromosomal DNA. In contrast, “unrelated” affected individuals share smaller conserved chromosomal regions because there are many more meioses, resulting in greater recombination around the region harboring the disease gene. Theoretically, polymorphic markers in linkage disequilibrium with the disease locus can be used to find associations of regions containing the gene mutation.

Linkage disequilibrium mapping has been shown to be an important tool for fine-mapping of monogenic diseases (10–12), and recently there has been success in identifying a gene involved in a complex disease— inflammatory bowel disease (13). However, there is little doubt that linkage disequilibrium mapping for most other complex diseases will be more difficult (14). The distance over which disequilibrium extends between markers and disease loci is not well understood nor is the degree of genetic risk contributed by any particular locus, suggesting that genotyping closely spaced markers in many case and control subjects would be required. Single nucleotide

From the Washington University School of Medicine, Division of Endocrinology, Diabetes and Metabolism, St. Louis, Missouri.

Address correspondence and reprint requests to M.A. Permutt, 660 S. Euclid Ave., Campus Box 8127, Saint Louis, MO 63110. E-mail: apermutt@im.wustl.edu.

Received for publication 20 March 2002 and accepted in revised form 15 May 2002.

LOD, logarithm of odds; PPAR, peroxisome proliferator-activated receptor; SNP, single nucleotide polymorphism.

The symposium and the publication of this article have been made possible by an unrestricted educational grant from Servier, Paris.

TABLE 1  
Results of genome scans for chromosome 20

Reference	Population	Sample	Scores	<i>P</i> <	Marker	Location
4	Caucasian	21 families, 53 ASP	1.48 LOD	0.005	D20S197	46 Mb
5	Caucasian	29 families, 498 individuals, 159 affected	2.36 LOD	0.009	D20S197	46 Mb
6	Caucasian-French	148 families, 301 ASP	2.27 LOD	0.001	RPNII	35.6 Mb
7	Finnish	477 families, 716 ASP	2.06 MLS	0.009	D20S909	34.3 Mb
			2.00 MLS	0.01	D20S107	59 cM
			1.92 MLS	0.013	D20S886	38.7 Mb
					D20S905	69.5 cM
						18.5 cM
2	Ashkenazi Jew	267 families, 896 individuals, 472 ASP	2.05 NPL <i>Z</i>	0.05	D20S195	31.7 Mb
8	Chinese	102 families, 478 individuals, 282 affected	1.641 NPL	0.03	D20S117	2.9 cM
			1.959 NPL	0.013	D20S889	11 cM
			1.517 NPL	0.04	D20S196	73.5 cM
9	Japanese	159 families, 224 ASP	2.32 MLS	0.00102	D20S119	43.5 Mb
						66.15 cM

ASP, affected sib-pair; MLS, maximum linkage score; NPL, nonparametric linkage score.

polymorphisms (SNPs), whereas biallelic, have been preferred over simple sequence repeat polymorphisms for this type of analysis because SNPs are more abundant in the genome (15). As a general rule, the extent of linkage disequilibrium or association between a marker and a disease locus depends on the genetic distance between the two and the number of generations that have occurred since the mutation originated (16). In isolated populations such as the Ashkenazi Jews, linkage disequilibrium has been shown to extend over a broad region (3). We therefore undertook a search for diabetes susceptibility gene(s) on chromosome 20q in Ashkenazi Jewish unrelated patients with diabetes. Here we report the initial results of our association studies with 91 validated SNPs spanning a 1-LOD interval (7.3 Mb) around the microsatellite marker D20S107 candidate region on chromosome 20. DNA from case and control subjects was genotyped in pools by a recently described method involving pyrosequencing technology (17). No significant associations have been found to date.

## RESEARCH DESIGN AND METHODS

**Patient population.** Type 2 diabetic case subjects and control subjects were of Ashkenazi Jewish descent as described (2). As shown in Table 2, the control subjects were significantly older than the case subjects by design because control subjects were selected for old age and absence of diabetes.

**DNA isolation, quantification, and construction of pools.** The DNA samples were isolated from whole blood using the Puregene kit as described (Gentra Systems, Minneapolis, MN). DNA was quantified with the TKO 100 Mini-Fluorometer and Hoechst dye method as described (Hofer Scientific Instruments, San Francisco, CA). For the purposes of creating DNA pools, efforts to accurately determine DNA concentrations for each sample are critical because errors will skew the proportion of each genotype in the pool. Spectrophotometric analysis was avoided because substances such as protein and salts may give spurious results (18). The DNA samples were gently mixed on a rocking platform to ensure homogeneity before pipetting. Equal volumes of each sample were delivered to a sterile 55-ml polypropylene solution basin (Labcor Products, Frederick, MD) using an accurately calibrated multichannel pipette. After mixing gently and thoroughly overnight, the pooled DNA was placed into 1.0-ml aliquots in sterile 1.5-ml polypropylene microtubes and stored at 4°C in the dedicated refrigerator. As a quality control, the uniformity of the mixing procedure was verified by genotyping replicate aliquots of the pools for several SNPs.

**PCR.** The reaction consisted of 2.5 µl GeneAmp 10× Buffer II (Applied Biosystems, Foster City, CA), 2.0–3.0 µl of 25 mmol/l MgCl<sub>2</sub> solution, 0.5 µl of

each 20 mmol/l dNTP (Amersham Pharmacia, Piscataway, NJ), 1 µl of 10 pmol/µl 5' biotin-triethylene glycol-labeled high-performance liquid chromatography (HPLC)-purified primer and 1 µl of 10 pmol/µl unlabeled primer (IDT, Coralville, IA), 0.25 µl (1.25 units) Amplitaq Gold (Applied Biosystems), 2.5 µl of 10 ng/µl DNA (pooled or individual), and sterile water to 25 µl total volume. Thermal cycling was done interchangeably on a GeneAmp 9700 (Applied Biosystems) or PTC-200 (MJ Research, Watertown, MA) using the following profile: heated lid, 95°C for 10 min × 1 cycle/95°C for 45 s/annealing temperature for 45 s/72°C for 1 min × 45 cycles to 4°C. The annealing temperature varied from 56 to 62°C. Forty-five cycles ensured that all PCR components were exhausted. PCR primers were designed with Primer3 Software (code available at [http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html)) (19), and the predicted reaction conditions (annealing temperature and MgCl<sub>2</sub>) were tested on several nonessential DNA samples. Fragment sizes between 100 and 500 bp were successfully analyzed.

**PCR plate setup, template preparation, and pyrosequencing.** There were 96-well plates (8 × 12) set up with eight replicates of the case and control pools. A variable number of replicates from three to ten were tested, and it was found that eight replicates most consistently resulted in an SD of ≤2%. Template preparation and pyrosequencing (Pyrosequencing AB) was conducted as described (17). Genotypes for a 96-well plate were generated in 10 min.

**Allele quantification software.** Allele frequencies in the samples were assessed by SNP Software Aq (Pyrosequencing AB) as described, and the data were exported to an Excel (Microsoft) spreadsheet for further analysis.

**Denaturing HPLC.** The PCR was as previously described. Heteroduplex DNA was formed and analyzed as described with the Wave (Transgenomic, Omaha, NE).

TABLE 2  
Phenotypic characteristics in case subjects versus control subjects

Averages	Case subjects	Elderly control subjects
M/F	72/78	69/74
Age at diagnosis (years)	48	76
BMI	31 ± 4.49	26 ± 3.99
Triglycerides	2.09 ± 1.24	
Cholesterol	5.41 ± 0.99	
HDL	0.99 ± 0.39	
LDL	4.02 ± 0.87	
HbA <sub>1c</sub>	8.3 ± 2.4	
Fasting blood glucose	164 ± 34.0	
Fastin insulin	19.86 ± 9.18	
Leptin	13.02 ± 6.82	

Data are means ± SD unless otherwise indicated.

**TABLE 3**  
Known and predicted genes in a 1-LOD interval around D20S107 with islet expression

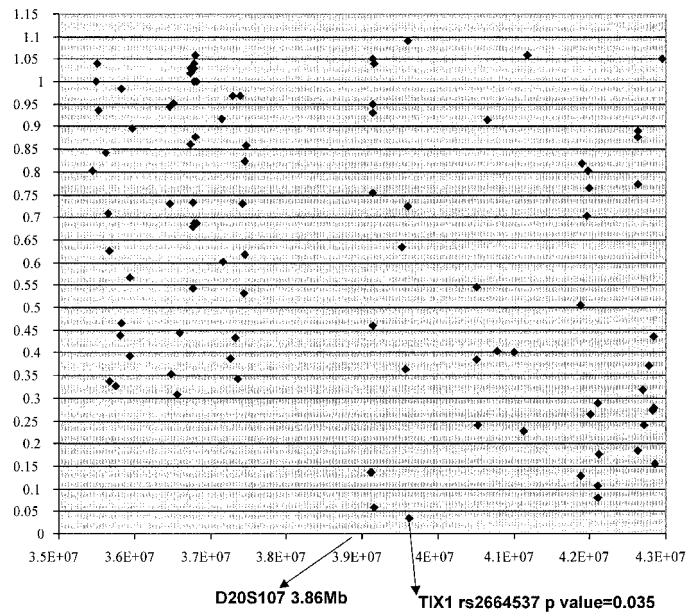
Gene	Description	Number of islet cDNAs in UniGene
<i>RBL-1</i>	Retinoblastoma-like 1	—
<i>RPN2</i>	Ribophorin II	—
<i>GHRH</i>	Growth hormone-releasing hormone	—
<i>SRC</i>	v-src sarcoma	1
<i>BLCAP</i>	Bladder cancer-associated protein	—
<i>NNAT</i>	Neuronatin	—
<i>KIAA1219</i>	Predicted protein	2
<i>TGM2</i>	Transglutaminase 2	1
<i>KIAA0406</i>	Predicted protein	2
<i>BPI</i>	Bactericidal/permeability-increasing protein	—
<i>LBP</i>	Lipopolysaccharide-binding protein	—
<i>VIAAT</i>	Vesicular inhibitory amino acid	—
<i>PPP1R16B</i>	Protein phosphatase 1, regulatory (inhibitor) subunit	2
<i>DDX35</i>	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 35	—
<i>MAFB</i>	v-maf musculoaponeurotic fibrosarcoma oncogene B	8
<i>TOP1</i>	Topoisomerase (DNA) I	2
<i>PRO0628</i>	PRO0628 protein	—
<i>PLCG1</i>	Phospholipase C, gamma 1	—
<i>TIX1</i>	Triple homeobox 1	—
<i>RPL12L2</i>	Ribosomal protein L12-like	—
<i>PTPRT</i>	Protein tyrosine phosphatase, receptor type, T	—
<i>SFRS6</i>	Splicing factor, arginine/serine rich 6	2
<i>H-L(3)MBT</i>	Lethal (3) malignant brain tumor	—
<i>SGK2</i>	Serum glucocorticoid-regulated kinase 2	—
<i>MYBL2</i>	v-myb myeloblastosis viral oncogene (avian)-like 2	—
<i>HNF4α</i>	Hepatocyte nuclear factor 4, alpha	4
<i>TDE1</i>	tumor differentially expressed 1	—

**SNP validation.** For each SNP assay, a reference individual and three DNA pools of 42 individuals from African-American, Asian (10 Chinese/32 Japanese), and Caucasian (the SNP Consortium DNA panels, Coriell Institute, Camden, NJ, <http://arginine.umd.edu>) populations were amplified by PCR. The Caucasian pool was used as a reference for Ashkenazi study. PCR products were cycle-sequenced using one of the original PCR primers and fluorescent dye-terminators and electrophoresed on an ABI 3700 (Applied Biosystems). For analysis, the allele frequencies in the pooled samples were estimated by comparison of the nucleotide peak heights in the electropherogram (20).

**TABLE 4**  
Estimates of allele frequencies for SNPs in known candidate genes for type 2 diabetes

Gene	Variant	Proband pool*	SD	Normal pool*	SD	P
<i>KIR6.2</i>	Glu23Lys	67.1 G/32.9 A	1.8	66.9 G/33.1 A	2.1	1.026
<i>CALPAIN 63</i>	nt16378	89.6 C/10.4 T	1.8	87.6 C/12.4 T	1.5	0.583
<i>CALPAIN 43</i>	nt4852	67.6 C/32.4 T	0.9	63.8 C/36.2 T	1.2	0.407
<i>CALPAIN 62</i>	nt16112	40 G/60 A	2.5	39.8 G/60.2 A	1.7	1.025
<i>SUR 18</i>	C/T exon 22	97.2 C/2.8 T	0.7	98.0 C/2.0 T	0.4	0.816
<i>PPARγ 21</i>	Pro12Ala	6.6 G/93.4 C	1	7.7 G/92.3 C	2.3	0.770
<i>IRS512</i>	Ala512Pro	4.4 G/95.6 C	0.8	4.6 G/95.3 C	0.6	1.046
<i>IRS972</i>	Gly972Arg	80.9 C/19.1 T	2.9	82.6 C/17.4 T	2.2	0.698

\*Allele frequency (%); n = 300.



**FIG. 1.** P values of differences (93 P values in a 1-LOD interval around D20S107) in minor allele frequencies between case and control pools (n = 300 alleles for each pool), plotted on the y-axis, versus the SNP position in megabases, plotted on the x-axis.

**Statistical analysis.** The P values quantifying the significance of the difference between pools of case and control subjects were calculated using the two-sample test for binomial proportions (normal theory test) (21), including twice the measurement variance in the calculation of the Z statistic.

**RESULTS**

The microsatellite markers bordering the 1-LOD interval around D20S107 include RPN2 and D20S911 at 35.7 and 43 Mb, respectively. This 7.3-Mb region contains a total of 25 known and 99 predicted genes (NCBI Human Genome Map, Build 28) (<http://www.ncbi.nlm.nih.gov>). All of the known genes in this region are shown in Table 3, along with the indication of those genes with expressed sequence tags (ESTs) expressed in pancreatic islets found in UniGene (<http://www.ncbi.nlm.nih.gov/UniGene>). As can be seen, seven of the known genes and two predicted genes were found with islet ESTs, and one gene (*MAFB*) showed relatively high expression in islets.

Because it would be difficult to directly sequence all 124 known and predicted genes within the putative at-risk region for a significant number of patients with diabetes,

linkage disequilibrium mapping was used to evaluate the entire 7.3-Mb region. SNPs were identified in public databases and validated in the Ashkenazi population through a collaborative arrangement with a member of the SNP Consortium (P. Kwok, Washington University Medical School). Direct sequencing of pooled DNA from Caucasian, Asian, and African-American individuals was conducted. SNPs with minor allele frequencies greater than 10% in the Caucasian subjects were then tested in pooled samples of DNA from Ashkenazi subjects. Interestingly, only ~50% of the SNPs submitted for validation were found to have a minor allele frequency of  $\geq 10\%$ . The validation data were entered into the public SNP database 1 week after testing (<http://www.ncbi.nih.gov/SNP>). The results for the validated SNPs for the four racial/ethnic groups are shown in Table A1 (in the APPENDIX). The SNPs rs736823 and rs932440, which were monomorphic in the Caucasian subjects, had minor allele frequencies of 42.2 and 11.1, respectively.

The allele frequencies of SNPs between Ashkenazi case subjects and control subjects ( $n = 300$  each) were tested. As shown in Table A2, a total of 91 SNPs were examined. Of these, 65 SNPs were located in known or predicted genes, and 26 were intragenic. Statistical analyses (Fig. 1) showed that the allele frequency for one SNP at *TIX1* appeared to differ (13.2 vs. 20.8%T for control and case subjects, respectively;  $P = 0.035$ ) but on replication was found to have no difference (17.8%T in case pool 2,  $P = 0.22$  vs. control) (Table A2 and Fig. 1).

Two genes in the region, *Mybl2* and *HNF4a*, were screened by denaturing HPLC for exonic mutations and new polymorphisms. *Mybl2* was screened because of the marginally significant  $P$  value of 0.057 for rs419842. *HNF4a* has been described in maturity-onset diabetes of the young (MODY)-1 (22) and as a possible type 2 diabetes candidate gene and has not yet been examined in Ashkenazi subjects. No coding or obvious splicing mutations were identified in either gene. However, in *Mybl2*, an unpublished nonsynonymous SNP R687C in exon 14 was identified (nt2186 C to T, accession number NM\_002466) (data not shown).

In addition, allele frequencies for reported type 2 diabetes candidate gene SNPs were determined for the islet ATP-sensitive  $K^+$  channel (*KIR6.2* and *SUR*), peroxisome proliferator-activated receptor (PPAR)- $\gamma$ , three SNPs for Calpain 10, and two SNPs in insulin receptor substrate 1, each previously shown in more than one study to be associated with type 2 diabetes (23,24). No differences were found between allele frequencies in case and control subjects for these candidate SNPs (Table 4).

## DISCUSSION

This study involved SNP association in pooled DNA for a 7.3-Mb chromosomal region on 20q containing a total of 124 known and predicted genes. Allele frequencies were assessed in DNA pools rather than individual genotypes to expedite the study and decrease costs. By using pyrosequencing technology, allele frequencies in the pooled DNAs occurred within 2% of those frequencies defined by individual genotypes (17).

One SNP (*TIX1*) out of 91 tested showed marginal

significance in case subjects versus control subjects, which was not replicated in a second pool of type 2 diabetic case subjects. Despite the lack of association to a specific SNP at this preliminary stage in the study, the occurrence of significant LOD scores along chromosome 20q from four racial/ethnic group studies supports the hypothesis that a genetic element contributing to type 2 diabetes is present. However, there are several limiting factors. First, each of the chromosome 20q peaks are fairly broad and encompass ~10–20 Mb of DNA. This in turn makes it difficult to anticipate the risk of not identifying the disease genes in the region around D20S107 and to determine whether a shift in focus more centromeric toward D20S195 or more telomeric toward D20S197 should be indicated. Second, lack of reported SNPs in this region mandates denaturing HPLC and sequencing of the genome in an effort to find more. For example, *MAFB*, a transcription factor, had fourfold more expression in islet cDNAs than any other gene, and only one SNP has been tested. Testing of 91 SNPs does not constitute an adequate evaluation of a 7.3-Mb region, and the SNP association study is still in progress. Further, we used the NCBI Chromosome Mapviewer as a roadmap for our SNP work. There have been at least five different “builds” or updates to the map since we started about 1 year ago. The map has expanded from 25 to possibly 124 genes in the region. It is evolving and so are our studies.

A major question remains as to what SNP density is required for accurate analysis. Does this mean typing one SNP every 10 kb requiring ~730 SNPs or one SNP every 1 kb requiring ~7,300 SNPs? A public project initiative to identify linkage disequilibrium blocks in the human genome may facilitate an answer to this question; however, this project is at least 2 years from completion. Furthermore, the differences in SNP allele frequencies between Ashkenazi and Western Caucasian subjects (Table A1) suggest that a unique haplotype map for isolated Caucasian subjects may be necessary. On the positive side, five groups have identified essentially the same region as harboring a putative diabetes gene(s) on chromosome 20q. In fact, a gene associated with type 2 diabetes, *PPAR- $\gamma$* , on chromosome 3p has not given a positive linkage signal in any of the Caucasian genome scans, suggesting perhaps that the signal on chromosome 20q confers a greater risk than that for *PPAR- $\gamma$* . The results of the current study indicate that the process of linkage disequilibrium mapping to narrow regions of linkage for complex diseases such as type 2 diabetes will be a difficult one.

## ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health Grants DK16746, DK07120, and DK49583 (to M.A.P.).

## APPENDIX

The results for the validated SNPs for the four racial/ethnic groups (Table A1) and all known genes on chromosome 20q (Table A2) are presented on the following pages.

TABLE A1  
Minor SNP allele frequencies in Ashkenazi versus other populations

SNP rs number	Ashkenazi control subjects, pool 1, 300 alleles	Ashkenazi cases, pool 1, 300 alleles	Ashkenazi cases, pool 2, 252 alleles	Caucasian pool, 84 alleles	Asian pool, 84 alleles	African-American pool, 84 alleles
1124692	24.1	25.4	—	20	5	42
1936999	27.2	26.9	—	13	2	36
1892205	44.1	44	—	nr	39	nr
1744765	14.6	15.2	—	12	40	2
765630	21.7	20	—	11	36	12
2012439	18.1	14.5	—	34	48	41
1001351	30.3	28	—	30	25	25
754625	37.8	34.2	37.2	40	0	0
1962746	22.1	21.5	—	11	27	46
990792	38.9	37.1	—	21	24	30
1341024	44.2	43.1	—	36	40	42
2066252	46.8	47	—	41	5	13
1341021	18.4	18.6	18.3	0	40	35
2022498	41.1	42.9	—	50	46	nr
1609800	35.8	35.5	35.1	40	10	15
737091	46.9	49	—	35	0	20
737090	49.6	49.9	48	39	nr	nr
2016946	25	21.4	—	21	31	18
742421	31.6	28.1	—	10	30	6
2064181	38.1	36.3	—	47	23	0
742277	39.4	36.4	—	36	47	48
926393	46.1	43.6	—	29	41	50
1534928	42.3	41.2	—	25	40	50
932440	8.7	13.5	—	0	0	0
926345	43.9	48.2	46.5	40	14	23
753381	34.5	32.7	—	42	11	nr
909775	45.3	39.9	—	30	40	40
1014757	45.2	44.4	—	23	38	0
929071	24.3	27.9	—	40	20	20
1033641	49.5	45.5	—	35	20	15
742420	29	34.3	—	30	40	45
2038668	40.4	40.4	—	35	40	20
916410	36.9	42	—	15	25	35
763227	26.3	31.1	—	17	17	35
442143	33.5	40.7	30.5	0	20	15
956609	16	16.8	—	10	3	46
764383	18.2	22.8	—	15	15	40
717247	32.7	27.9	—	30	30	45
736823	40.6	44.3	41.7	2	nr	0
1028583	26.8	32.9	33.1	40	30	45

Data are %. nr, no result.

TABLE A2

All known genes on chromosome 20q (35.7–43.0 Mb) with validated SNPs listed for allele frequency in probands and normal subjects

Gene	rs number	SNP position	Allele frequency, normal pool 1, <i>n</i> = 300 alleles	SD	Allele frequency, proband pool 1, <i>n</i> = 300 alleles	SD	$\Delta^*$	<i>P</i>	Allele frequency, proband pool 2, <i>n</i> = 252 alleles
<i>RBL1</i>	1124692	35442334	24.1 C	1	25.4 C	1	0.9	0.802	
	1936999	35501910	27.2 G	3.4	26.9 G	3.6	0.3	1	
	1892205	35513009	44.1 G	0.7	44 G	1.3	0.1	1.04	
	1744765	35525386	14.6 G	0.7	15.2 G	1	0.6	0.936	
<i>RPN2</i>	1076686	35613623	47 A	1	48.2 A	0.9	1.2	0.843	
	765630	35653552	21.7 G	1.3 (6)	20 G	1.8	1.7	0.709	
	2012439	35668823	18.1 T	1.9	14.5 T	1.2	3.6	0.337	
G†	1001351	35674649	30.3 T	0.6	28 T	1.1	2.3	0.625	
<i>GHRH</i>		No Gene SNPs							
G	1043415	35747791	31.9 G	2.4	27.6 G	1.5	4.3	0.327	
<i>SRC</i>	754625	35820080	37.8 G	1.3	34.2 G	1.8	3.6	0.439	37.2 G
	1570209	35831387	13.6 C	2.3	11 C	2.1	2.6	0.466	
	1885257	35833495	20.4 G	1.9	20.8 G	3.4	0.4	0.985	
G	1033623	35938915	8.3 A	2.7	11.1 A	2.6	2.8	0.392	8.0 A
	1007373	35944844	24.6 A	0.4	22.1 A	0.8	2.5	0.568	21.9 A
<i>BLCAP/NNAT</i>		No Gene SNPs							
G	2024794	35972553	48.3 C	1.5	49.2 C	0.6	0.9	0.897	48.1 C
<i>C20orf77</i>	1962746	36466943	22.1 C	1.6	21.5 C	1.8	0.6	0.943	
	990792	36474026	38.9 A	1.5	37.1 A	2.8	1.8	0.731	
	910664	36484857	47.3 C	1.2	48.3 C	1	4.4	0.352	
	<i>TGM2</i>	1555074	36563758	15.9 T	1	19.8 T	1.2	3.9	0.308
G	1883484	36601263	19.1 T	1.2	16.1 T	0.5	3.1	0.444	
<i>BPI</i>	1341024	36735232	44.2 C	0.9	43.1 C	1.4	1.1	0.86	
	2066252	36739802	46.8 T	0.6	47 T	0.6	0.2	1.02	
	1341021	36740791	18.4 T	1.7	18.6 T	2.5	0.2	1.03	18.3 T
	154374	36765781	45.8 C	1.8	48.8 C	0.7	5.4	0.542	
G	2022498	36768649	41.1 T	1.2	42.9 T	0.9	1.8	0.734	
	1609800	36781066	35.8 A	1.2	35.5 A	1.5	0.3	1	35.1 A
	737091	36812479	46.9 A	1.2	49 A	1.1	2.1	0.686	
<i>LBP</i>	1780617	36776713	24.1 G	1.3	22.2 G	2	1.9	0.679	
	1739654	36780526	19.4 T	1.3	19.6 T	1.1	0.2	1.03	
	1780623	36791825	49.7 G	1.3	47.6 G	0.7	2.1	0.686	
	1780624	36791858	46.8 G	1.9	46.7 G	1.5	0.1	1.04	
	1780627	36800188	46.7 T	1.3	46.7 T	1.9	0	1.06	
	1739640	36804490	41.8 T	0.9	42.8 T	1.1	1	0.878	
	G	737090	36812180	49.6 G	2.9	49.9 G	1.7	3.1	1
	2016946	37258761	25 A	0.7	21.4 A	1	3.6	0.389	
<i>VIAAT</i>	2667326	37154262	18.3 T	1.2 (7)	17.6 T	1.7	0.7	0.917	
	1321099	37161201	20.6 G	1.4 (7)	18.4 G	1.5	2.2	0.603	
<i>PPP1R16B</i>	208817	37294872	49.5 C/50.5 T	1.1	49 C	0.7	0.5	0.969	
	742421	37321877	31.6 T	1.4	28.1 T	0.7	3.5	0.434	
	10392	37353491	26.8 T	0.8	22.8 T	0.9	4	0.342	
<i>DDX35</i>	974560	37393754	42 A	1.2	42.5 A	2.1	0.5	0.969	40.9 A
	2064181	37415795	38.1 T	1.8	36.3 T	1.9	1.8	0.73	
	742277	37431305	39.4 G	0.5	36.4 G	0.8	3	0.531	
	926393	37448584	46.1 T	1.6	43.6 T	1	2.5	0.619	
	101120	37459975	47.8 A	1.9	46.5 A	1.9	1.3	0.825	
	1534928	37467120	42.3 A	1.3	41.2 A	1.1	1.1	0.859	
<i>KRML</i>	3577	39117226	8.7 T	0.4	13.5 T	0.5	4.8	0.137	
	G	742739	39148496	28.2 C	1.7	28.1 C	1.2	0.1	1.05
	932440	39129071	8.7 T	0.4	13.5 T	0.5	4.8	0.137	
	742740	39148437	23.6 T	1	26.8 T	1.1	3.1	0.46	
<i>TOP 1</i>	2076575	39523386	16.7 G	1.1	18.7 G	1.4	2	0.633	
<i>PRO0628</i>		No Gene SNPs							
<i>PLCG1</i>	926345	39574503	43.9 C	1.9	48.2 C	0.5	4.3	0.363	46.5 C
	753381	39600021	34.5 A	0.3	32.7 A	0.5	1.8	0.725	
<i>TIX1</i>	2235363	39610325							
	2235364	39610606							
	2076147	39615562							
	2664537	39616285	13.2 T	0.9	20.8 T	1.3	7.6	0.035	17.8 T
	2235367	39632678							

Continued on following page

TABLE A2  
Continued

Gene	rs number	SNP position	Allele frequency, normal pool 1, <i>n</i> = 300 alleles	SD	Allele frequency, proband pool 1, <i>n</i> = 300 alleles	SD	$\Delta^*$	<i>P</i>	Allele frequency, proband pool 2, <i>n</i> = 252 alleles
<i>RPL12L2</i>	903374	39983421							
<i>PTPRT</i>	746413	40512545	44.4 A	0.9	40.3 A	0.7	4.1	0.384	41.2 A
	2016647	40517034	23.9 T	2.4	21.3 T	2.2	2.6	0.546	
	909775	40530430	45.3 C	0.7	39.9 C	1.1	5.4	0.242	
	1014757	40664178	45.2 G	0.7	44.4 G	1	0.8	0.915	
	929071	40781985	24.3 T	1.2	27.9 T	1.8	2.6	0.403	
	1033641	41002123	49.5 C	0.9	45.5 T	0.6	5	0.401	
	742420	41137137	29 T	0.7	34.3 T	0.8	5.3	0.227	
	2038668	41180421	40.4 A	1.4	40.4 A	1.3	0	1.06	
G	761039	41885512	10.2 C	1.2	15.3 C	0.6	5.1	0.129	12.5 C
<i>SFRS6</i>	2235611	41892066	15.2 T	1.3 (6)	17.8 T	1.6	2.6	0.506	
G	12264	41893912	76.3 A	1.1	74.5 A	0.9	1.2	0.818	
<i>H-L(3)MBT</i>	2071968	41966323	32.5 C	0.9	30.6 C	1.5	1.9	0.703	
	3205	41982000	18.2 C	1.8	17 C	1	1.2	0.803	
G	2009280	41991469	18.8 T	1.3	19.6 T	1.8	0.8	0.764	
<i>SGK2</i>	916410	42005792	36.9 G	1.7	42 G	0.9	5.1	0.266	
	763227	42007616	26.3 C	1.4	31.1 C	3.3	4.8	0.266	
<i>MYBL2</i>	826950	42104926	7.2 G	2	10.5 G	2	3.3	0.289	
	442143	42113064	33.5 T	0.9	40.7 T	1.6	7.2	0.106	30.5 T
	419842	42113366	20.2 A	0.7	27.8 A	0.6	7.6	0.057	18.2 A
	285186	42117646	7.4 G	0.7	12.9 G	1.5	5.5	0.77	7.5 G
<i>C20orf111</i>	9875	42628042	21.1 G	0.8	26.5 G	2.3	5.4	0.184	
	8268	42628484	22.1 A	1.6 (7)	20.7 A	0.9 (5)	1.4	0.773	
	956609	42639996	16 G	1.5	16.8 G	1.3 (7)	0.8	0.891	
	2143606	42641105	38.3 C	1.2	39.2 C	1.9	0.9	0.876	
<i>MGC3129</i>	761331	42696185	44.6 C	0.8	49.3 C	0.9	4.7	0.317	
G	764383	42719649	18.2 C	1.3	22.8 C	1.2 (7)	4.6	0.242	
	1884614	42783074	23.4 T	1.1	27.2 T	1.3	3.8	0.372	
	717247	42828339	32.7 G	1.9	27.9 G	1.2	4.8	0.272	
<i>HNF4a</i>	736821	42836634	18.5 T	0.8	22.8 T	0.6	4.3	0.278	
	736823	42837338	40.6 A	2.9	44.3 A	1.2	3.7	0.437	41.7 A
	1028583	42853316	26.8 T	1	32.9 T	0.9	6.1	0.155	33.1 T
<i>TDE1</i>	2015437	42944500	24.7 T	2.2	24.8 T	0.9	0.1	1.05	24.3 T

\*Average difference between the normal and proband pool frequencies for the SNP; †intra-genic SNPs. Numbers in parentheses indicate the number of replications if <8.

## REFERENCES

- DeFronzo RA: Lilly Lecture 1987: the triumvirate:  $\beta$ -cell, muscle, liver: a collusion responsible for NIDDM. *Diabetes* 37:667-687, 1988
- Permutt MA, Wasson JC, Suarez BK, Lin J, Thomas J, Meyer J, Lewitzky S, Rennich JS, Parker A, Duprat L, Maruti S, Chayen S, Glaser B: A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* 50:681-685, 2001
- Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, Breakefield X, Bressman S: Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet* 9:152-159, 1995
- Bowden DW, Sale M, Howard TD, Qadri A, Spray BJ, Rothschild CB, Akots G, Rich SS, Freedman BI: 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, 1997
- Ji LN, Malecki M, Warram JH, Yang YD, Rich SS, Krolewski AS: New susceptibility locus for NIDDM is localized to human chromosome 20q. *Diabetes* 46:876-881, 1997
- Zouali H, Hani EH, Philippi A, Vionnet N, Beckmann JS, Demenais F, Froguel P: A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenolpyruvate carboxykinase gene. *Hum Mol Genet* 6:1401-1408, 1997
- Ghosh S, Watanabe RM, Hauser ER, Valle T, Magnuson VL, Erdos MR, Langefeld CD, Balow J, Ally DS, Kohtamaki K, Chines P, Birznieks G, Kaleta HS, Musick A, Te C, Tannenbaum J, Eldridge W, Shapiro S, Martin C, Witt A, So A, Chang J, Shurtleff B, Porter R, Kudelko K: Type 2 diabetes: evidence for linkage on chromosome 20 in 176 Finnish affected sib pairs. *Proc Natl Acad Sci U S A* 96:2198-2203, 1999
- Luo TH, Zhao Y, Li G, Yuan WT, Zhao JJ, Chen JL, Huang W, Luo M: A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia* 44:501-506, 2001
- Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecocq C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P: Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* 51:1247-1255, 2002
- Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, et al.: The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* 78:1073-1087, 1994
- Jorde L: Linkage disequilibrium as a genetic-mapping tool. *Am J Hum Genet* 56:11-14, 1995
- Xiong MM, Guo SW: Fine-scale mapping based on linkage disequilibrium: theory and applications. *Am J Hum Genet* 60:1513-1531, 1997
- Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G: Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411:599-603, 2001
- Collins FS, Brooks LD, Chakravarti A: A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* 8:1229-1231, 1998
- Reich DE, Lander ES: On the allelic spectrum of human disease. *Trends Genet* 17:502-510, 2001
- Jorde L: Genetic diversity in the Ashkenazi population: evolutionary considerations. In: *Genetic Diversity Among Jews: Diseases and Markers at the DNA Level*. Bonne-Tamir B, Adam A, Eds. New York, Oxford University Press, 1992, p. 305-318

17. Wasson J, Skolnick G, Love-Gregory L, Permutt MA: Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. *Biotechniques* 32:1144–1146, 2002
18. Wasson J, Brodeur GM: Molecular analysis of gene amplification in tumors. In *Current Protocols in Human Genetics*. Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC, Seidman CE, Seidman JG, Smith DR, Eds. New York, Wiley, 2000, p. 10.5.1–10.5.18
19. Rozen S, Skaletsky HJ: Primer3. Code available at [http://www-genome.wi.mit.edu/gnome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/gnome_software/other/primer3.html), 1998
20. Marth G, Yeh R, Minton M, Donaldson R, Li Q, Duan S, Davenport R, Miller R, Kwok PY: Single-nucleotide polymorphisms in the public domain: how useful are they? *Nat Genet* 27:371–372, 2001
21. Rosner B: *Fundamentals of Biostatistics*. Pacific Grove, CA, Duxbury, 2000, p. 356–371
22. Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ, Fajans SS, Signorini S, Stoffel M, Bell GI: Mutations in the hepatocyte nuclear factor 4 $\alpha$  gene in maturity onset diabetes of the young (MODY1). *Nature* 384:458–460, 1996
23. Altshuler D, Hirschhorn J, Klannemark M, Lindren C, Vohl M, Nemesh J, Lane C, Schaffner S, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson T, Daly M, Groop L, Lander E: The common PPAR $\gamma$  pro12ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80, 2000
24. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner T, Mashima H, Schwarz P, del Bosque-Plata L, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier L, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163–175, 2000