Hum Genet (2006) 120:527–542
DOI 10.1007/s00439-006-0231-0

ORIGINAL INVESTIGATION

Association study on chromosome 20q11.21-13.13 locus and its contribution to type 2 diabetes susceptibility in Japanese

Toshihito Tanahashi · Dai Osabe · Kyoko Nomura · Shuichi Shinohara · Hitoshi Kato · Eiichiro Ichiishi · Naoto Nakamura · Toshikazu Yoshikawa · Yoichiro Takata · Tatsuro Miyamoto · Hiroshi Shiota · Parvaneh Keshavarz · Yuka Yamaguchi · Kiyoshi Kunika · Maki Moritani · Hiroshi Inoue · Mitsuo Itakura

Received: 28 April 2006 / Accepted: 3 July 2006 / Published online: 6 September 2006

Abstract Several linkage studies have predicted that human chromosome 20q is closely related to type 2 diabetes, but there is no clear evidence that certain variant(s) or gene(s) have strong effects on the disease within this region. To examine disease susceptibility variant in Japanese, verified SNPs from the databases, with a minor allele frequency larger than 0.15, were selected at 10-kb intervals across a 19.31-Mb region (20q11.21-13.13), which contained 291 genes, including hepatocyte nuclear factor 4α (HNF4α). As a result, a total of 1,147 SNPs were genotyped with TaqMan assay using 1,818 Japanese samples. By searching for HNF4α as a representative disease-susceptible gene, no variants of HNF4α were strongly associated with disease. To identify other genetic variant related with disease, we designed an extensive two-stage association study (725 first and 1,093 second test samples). Although SNP1146 (rs220076) was selected as a landmark within the 19.31 Mb region, the magnitude of the nominal *P* value ($P = 0.0023$) was rather weak. Subsequently, a haplotype-based association study showed that two common haplotypes were weakly associated with disease. All of these tests resulted in non-significance after adjusting for Bonferroni's correction and the false discovery rate to control for the impact of multiple testing. Contrary to the initial expectations, we could not conclude that certain SNPs had a major effect on this promising locus within the framework presented

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00439-006-0231-0 and is accessible for authorized users.

T. Tanahashi · P. Keshavarz · Y. Yamaguchi · K. Kunika · M. Moritani · H. Inoue · M. Itakura (✉)
Division of Genetic Information,
Institute for Genome Research,
The University of Tokushima,
3-18-15, Kuramoto-cho, Tokushima 770-8503, Japan
e-mail: itakura@genome.tokushima-u.ac.jp

D. Osabe · K. Nomura
Department of Bioinformatics,
Division of Life Science Systems,
Fujitsu Limited, 1-5-2, Higashishinbashi,
Minato-ku, Tokyo 105-7123, Japan

S. Shinohara
Division of R&D Solution,
Fujitsu Nagano Systems Engineering Limited,
Nagano 380-3813, Japan

H. Kato
Haplopharma Inc., 2-2-1, Yaesu, Chuo-ku,
Tokyo 104-0028, Japan

E. Ichiishi
New Industry Creation Hatchery Center,
Tohoku University, Aramaki, Aoba-ku,
Sendai, Miyagi 980-8579, Japan

N. Nakamura · T. Yoshikawa
Department of Endocrinology and Metabolism,
Kyoto Prefectural University of Medicine Graduate
School of Medical Sciences, 465, Kajii-cho,
Hirokoji-Kawaramachi, Kamigyo-ku,
Kyoto 602-8566, Japan

Y. Takata
Department of Orthopedics, Institute for Health
Biosciences, The University of Tokushima,
3-18-15, Kuramoto-cho, Tokushima 770-8503, Japan

T. Miyamoto · H. Shiota
Department of Ophthalmology and Visual Neuroscience,
Institute for Health Biosciences,
The University of Tokushima, 3-18-15,
Kuramoto-cho, Tokushima 770-8503, Japan

here. As a way to extend our observations, we emphasize the importance of a subsequent association study including replication and/or meta-analysis in multiple populations.

Abbreviations

FUSION	The Finland-United States Investigation of Non-insulin-dependent diabetes mellitus genetics
UTR	Untranslated region
JPT	Japanese in Tokyo
CHB	Han Chinese in Beijing
CEU	Utah residents with ancestry from northern and western Europe by the Centre d'Etude du Polymorphisme Humain

Introduction

Single-nucleotide polymorphisms (SNPs) are the most abundant form of DNA variation in the human genome. To date, more than one million SNPs have been intensively searched and submitted to public databases as common variants in the genome sequence (Altshuler et al. 2005). Numerous resources of dense SNPs provide a research framework that can be exploited to search for genetic factors relevant to common complex diseases with an association study.

Genetic association studies are mainly implemented using two approaches. On the basis of hypothesis-driven theory, candidate gene analysis is directly connected with systemic knowledge of biological function, whereas a whole-genome association study theoretically investigates the entire genome simultaneously in an unbiased manner (Wang et al. 2005; Hirschhorn et al. 2005). However, no method to analyze the entire genome for complex diseases has been established with respect to the availability of unbiased reliability and cost efficiency (Wang et al. 2005; Hirschhorn and Daly 2005). To improve these methodological problems, we have developed a practical and efficient strategy for an extensive region-wide association study with evenly spaced dense SNPs markers (one marker per 10 kb). This research strategy has been applied to the discovery of disease-causing gene(s) and variant(s) as an effective method (Hamada et al. 2005; Kato et al. 2006).

Type 2 diabetes is a complex metabolic disorder characterized by impaired insulin production, insulin resistance in peripheral tissues, and abnormal glucose output by the liver. It is proposed that multiple genetic and environmental factors contribute to the disease (Permutt et al. 2005; O'Rahilly et al. 2005). Whole-genome linkage studies in families with affected individuals have been undertaken in different ethnic groups, and some mapped loci have been observed. In particular, the suggested linkages overlap on the chromosome 20q locus in Caucasian, African-American, and Asian populations (Ghosh et al. 1999, 2000; Vionnet et al. 2000; Permutt et al. 2001; Duggirala et al. 2001; Luo et al. 2001; Mori et al. 2002; Iwasaki et al. 2003). This region harbors the hepatocyte nuclear factor 4α (HNF4α) gene, mutations of which cause maturity-onset diabetes of the young (MODY) type 1, a dominantly inherited, and early-onset form of type 2 diabetes (OMIM 12850). It is not yet clear whether common variants in the same gene also have similar effects on late-onset forms of type 2 diabetes. Hence, several association studies have further focused on this gene, especially in Caucasians (Winckler et al. 2005; Bagwell et al. 2005; Silander et al. 2004; Love Gregory et al. 2004; Vaxillaire et al. 2005). Until now, however, there has been no clear evidence on chromosome 20q as to whether susceptibility gene(s) or variant(s) definitively contribute to type 2 diabetes in any populations including Japanese.

In this study, we closely examined SNPs within HNF4α as a likely candidate gene for common type 2 diabetes with an association study, using 1,818 Japanese samples. Simultaneously, to prevent gaps when searching for genetic variants on chromosome 20q, verified SNPs were extensively examined with a two-stage association study across 19.31 Mb on a putative candidate locus. We studied the contribution of detected variants statistically and extended the knowledge about type 2 diabetes on this important locus.

Materials and methods

Selection of samples for the association study

We prepared two independent samples from 1,818 Japanese individuals (925 cases and 893 controls). The first test samples (725 samples) consisted of 367 type 2 diabetes cases and 358 healthy controls. Next, we repeated the association test with the second test samples (1,093 samples) of 558 type 2 diabetes subjects and 535 controls. A description of the case and control samples is provided in Table 1.

Type 2 diabetes patients were mainly recruited at the Tokushima University Hospital, Kyoto Prefectural

Table 1 Clinical characteristics of 1,818 Japanese samples in the association study (893 controls and 925 cases)

Phenotype	First test samples (725 samples)		Second test samples (1,093 samples)	
	Cases	Controls	Cases	Controls
Number of samples	367	358	558	535
Gender (male/female)	191/176	145/213	277/281	291/244
Age (years)	64.1 ± 10.1	38.1 ± 14.2	62.9 ± 10.0	37.9 ± 11.4
BMI (kg/m^2)	23.6 ± 3.4	21.9 ± 2.9	23.4 ± 3.3	22.3 ± 2.9
HbAlc (%)	7.4 ± 1.4	4.7 ± 0.3	7.3 ± 1.3	4.9 ± 0.3
Age at onset				
< 50 years (%)	178 (48.5)	–	275 (49.3)	–
> 50 years (%)	187(51.0)	–	281 (50.4)	–
Unknown (%)	2 (0.5)	–	2 (0.3)	–
Positive family history (first degree relatives, %)	132 (36.0)	–	224 (40.1)	–

Age, BMI, and HbAlc are represented as means ± SD. BMI denotes body mass index

University Hospital, and its related hospitals. The diagnosis was based on medical records using the criteria of the World Health Organization (World Health Organization 1985). Type 2 diabetes was clinically defined as gradual adult onset of the disease with medication. Patients with mitochondrial disease or MODY were completely excluded. Control subjects consisted of members of the general population. A large number of control samples (56.7%) were collected by the Pharma SNP consortium (PSC). This consortium works in close collaboration with the National Millennium Project in Japan. Through this consortium, their health conditions and negative family history of diabetes were thoroughly checked. On examining their birthplace information, all subjects had full Japanese ancestry. Hence, the bias of population stratification caused by different ethnic groups was supposed to be low.

Genomic DNA was prepared from peripheral blood leukocytes or from Epstein-Barr virus-immortalized lymphoblast cell lines with a standard protocol. The study protocol was approved by the Institutional Review Board of The University of Tokushima and was in accordance with the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to blood sampling. All personal identifiers associated with medical information and blood samples were carefully eliminated and replaced with anonymous identifiers in each recruiting institution.

Previous linkage reports

Initially, we examined several published genome-wide linkage studies related to chromosome 20q with multiple ethnic groups and extracted nine representative reports in Supplementary Table 1. To date, for Japanese studies, there have been only three linkage reports, including two positive and one negative report on chromosome 20q after searching the PubMed database.

One positive report identified nominal evidence of linkage with a maximum LOD (logarithm of odds) score of 1.34 for D20S107 (Mori et al. 2002). Alternative linkage analysis also indicated the same region with a maximum LOD score of 1.99 for D20S107 (Iwasaki et al. 2003). In this region, linkage peak has been reported, but this was confusing because the linkage peak was quite broad and rather low. Another recent report failed to confirm the previous two items of linkage evidence on chromosome 20q (Nawata et al. 2004).

SNPs markers selection

Using the previous positive reports, the linkage position was used to adjust the genetic markers to the reference Marshfield Genetic Map. The linkage peaks confirmed that the putative disease-susceptibility locus was mapped to the 50.81–75.01 cM region on chromosome 20q11.21-13.13, corresponding to a 19.31-Mb interval between STS (sequence-tagged site) markers of D20S195 and D20S196. We selected the 19.31-Mb interval as the target region for the extensive association study (Table 2).

To establish a useful marker set for the association study, we searched SNPs located in the interval between D20S195 and D20S196 in public databases [Database of Single-Nucleotide Polymorphisms (dbSNP) and the SNP Consortium] and Celera Discovery System database. (Celera Discovery System database is no longer in use. It was commercially available until June 2005.) Among the numerous candidate SNPs identified, we selected SNPs that mainly

Table 2 Summary of target region on chromosome 20q, and SNPs markers

Target region on chromosome 20q	Results
Cytobands	20qll.21-20ql3.13
STS marker	D20S195-D20S196
Total target size (bp)	19,310,566
NCBI build 35 contig start position (bp)	29,871,045
NCBI build 35 contig end position (bp)	49,181,611
(A) Size of gene-centric region (gene ± 10 kb) (bp)	13,432,173
(B) Size of gene-centric region (gene ± 0 kb) (bp)	10,100,670
Number of RefSeq genes	291
Number of SNPs	1,147
Number of RefSeq genes with > 1 SNP (%)	219 (75.3%)
Average SNPs distance (bp) per (A)	10,820
Average SNPs distance (bp) per (B)	8,797

Gene information and position are based on the NCBI Build 35 human genome

matched the following criteria: (1) dense and evenly spaced SNPs in both coding and non-coding regions, (2) in the coding region, SNPs particularly placed on the gene locus spanning 10 kb upstream of the transcription start site and 10 kb downstream from the end of the last exon, (3) SNPs suitable for developing a TaqMan allele discrimination assay, and (4) to eliminate the effect of rare variants, which were underrepresented in the database, common SNPs that showed minor allele frequencies (MAF) > 0.15 in 46 healthy unrelated Japanese controls (24 males and 22 females) initially tested (Hamada et al. 2005; Kato et al. 2006).

SNPs genotyping

TaqMan MGB (minor groove binder) probe and primer sets were designed by Applied Biosystems (Applied Biosystems, Foster City, CA). We genotyped SNPs with either TaqMan Universal Master Mix (Applied Biosystems) or a QuantiTect Probe polymerase chain reaction (PCR) kit (Qiagen, Stanford, CA) and 5 ng of DNA, 450 nM of each primer, and 100 nM of probe. PCR conditions were as follows: 95°C for 10 min for enzyme activation, followed by 40 cycles of 92°C or 94°C for 15 s and 60°C for 1 min. Thermal cycling was conducted on an ABI GeneAmp PCR System 9700 (Applied Biosystems). Each 384-well plate contained 380 samples of unknown genotypes and four samples of no-template controls. After reaction, the fluorescence of VIC and FAM dyes was measured with an ABI Prism 7900HT using Sequence Detector System (SDS) ver 2.1 software (both from Applied Biosystems).

Data review and quality control criteria to genotyping

In genotyping with the TaqMan allele discrimination assay, genotype calls were identified by clustering the fluorescence intensity measurements for each SNP. Therefore, quality control was assessed by the reliability of the intensity measurements (Hinds et al. 2005).

We first checked the number of observed genotyped clusters and eliminated SNPs that could not be clustered or for which our call rate was lower than 98%. Expected genotyping accuracies were estimated with the quality score algorithm used in SDS ver 2.1 program (Applied Biosystems). The intensity measurements were carefully checked for accuracy by two independent researchers. In addition, the deviation of genotype distribution was checked by the Hardy–Weinberg equilibrium test. The test for deviation from Hardy–Weinberg equilibrium was quite effective for identifying artifacts and improving data quality (Klein et al. 2005).

To summarize, quality control included the call rate, the number of genotyped clusters, and consistency with the Hardy–Weinberg equilibrium test ($P < 0.05$). After the evaluation of quality control, the success rate of allele-identification was over 99%, and there was perfect agreement between the results of genotyping and direct sequencing, as previously described in our laboratory (Hamada et al. 2005; Kato et al. 2006).

Design of the extensive association study and statistical analyses

To bring down the cost and technical burden linked to genotyping, and to further reduce the type 1 error rate (false-positive) without losing any significant quality of data, we designed a two-stage association analysis strategy (Fig. 1). This approach is highly recommended to minimize the chance of false-positive association results (Hirschhorn and Daly 2005).

In the first association test samples, all SNPs were genotyped, and SNPs exhibiting the association with type 2 diabetes ($P < 0.05$) were further examined for replication in independent second test samples (Fig. 1 and Table 3). For each association analysis, the results of SNPs genotyping were calculated with four types of chi-square tests (allele frequency, genotype frequency, dominant model, and recessive model) with 2×2, or 2×3 contingency tables. All analyses were two-tailed. As the cases were older than the controls in this study, a logistic regression model was used to adjust for age. However, this adjustment did not change the

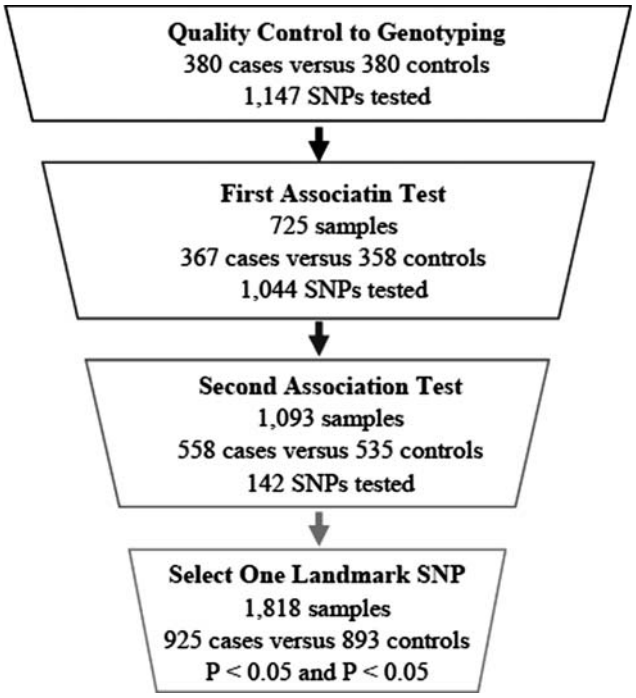

Fig. 1 Scheme of the quality control and two-stage association study. Sample size and the number of SNPs are indicated. Full sets of SNPs are genotyped with a relatively small population (first association test). Only SNPs with P values < 0.05 for association with type 2 diabetes advanced to the next stage with a larger population for the purpose of replication (second association test)

significance level or patterns of association. Analysis was carried out using the SPSS system, release 12.0 J for Windows (SPSS Japan Inc., Tokyo, Japan).

Hardy–Weinberg equilibrium of alleles at individual loci was calculated with genotype frequency obtained by simple gene counting, and was evaluated by the chi-square test to compare observed and expected values. SNPs with a Hardy–Weinberg equilibrium test result of $P < 0.05$ in controls and cases were removed from the analysis (Klein et al. 2005).

To evaluate the type 1 error rate (false-positive) generated by testing a relatively large number of SNPs, standard Bonferroni's correction was used (McIntyre et al. 2000). In addition, we used the false-discovery rate (FDR) approach of Benjamini (Benjamini and Hochberg 1995) and Storey (Storey 2002), as implemented with the FDR control program in the R library. The FDR approach estimates the expected proportion of false positives among significant tests through q values. To obtain the q value for each marker, we used the QVALUE software (Storey and Tibshirani 2003). As a threshold, we used a value under 0.1 for SNPs associated with disease on the basis of a recent report (Shiffman et al. 2005).

As we evaluated numerous SNPs, it was necessary to process an enormous amount of genotyping data correctly. For this purpose, a high-speed resolution program, FGDS (Fujitsu Gene Discovery System; now available as ver 2.0, Fujitsu Ltd, Tokyo, Japan), was newly developed to efficiently process genotyping data in our laboratory. All algorithms using this system could modify the available programs in the public databases, and were strictly verified (Hamada et al. 2005; Kato et al. 2006).

Association study for Hepatocyte Nuclear Factor 4α (HNF4α) gene variants

HNF4α was considered as a putative susceptibility gene with type 2 diabetes, because several studies have observed suggestive linkage peaks around the HNF4α

Table 3 The distribution of P values in 1,044 SNPs markers with the first association test

P value	Number of SNPs									
	Allele model		Genotype model		Dominant model		Recessive model		Overall	
$1 \geqq P \geqq 0.9$	115	11.0%	105	10.1%	104	10.0%	110	10.5%	10	1.0%
$0.9 > P \geqq 0.8$	86	8.2%	89	8.5%	103	9.9%	93	8.9%	33	3.2%
$0.8 > P \geqq 0.7$	99	9.5%	86	8.2%	103	9.9%	90	8.6%	44	4.2%
$0.7 > P \geqq 0.6$	105	10.1%	81	7.8%	97	9.3%	87	8.3%	74	7.1%
$0.6 > P \geqq 0.5$	79	7.6%	91	8.7%	112	10.7%	93	8.9%	79	7.6%
$0.5 > P \geqq 0.4$	93	8.9%	114	10.9%	89	8.5%	99	9.5%	86	8.2%
$0.4 > P \geqq 0.3$	103	9.9%	108	10.3%	95	9.1%	102	9.8%	119	11.4%
$0.3 > P \geqq 0.2$	113	10.8%	117	11.2%	112	10.7%	103	9.9%	144	13.8%
$0.2 > P \geqq 0.1$	126	12.1%	118	11.3%	111	10.6%	131	12.5%	201	19.3%
$0.1 > P \geqq 0.05$	64	6.1%	48	4.6%	62	5.9%	71	6.8%	112	10.7%
$0.05 > P$	61	5.8%	87	8.3%	56	5.4%	65	6.2%	142	13.6%
Total	1,044		1,044		1,044		1,044		1,044	

Data are presented as the number of SNPs. P values are calculated by four chi-square tests, allele, genotype, dominant and recessive models. 142 SNPs revealed a P value of < 0.05 in at least one of the four type chi-square tests

gene on chromosome 20q in multiple populations (Love Gregory et al. 2004; Silander et al. 2004).

Seventeen SNPs markers were arranged around the HNF4α locus including the pancreatic beta cell P2 promoter on the basis of two recent publications containing significant SNPs associated with disease (Winckler et al. 2005; Bagwell et al. 2005), and were examined with 1,818 samples (both first and second test samples). In this evaluation, mis-sense variant Thr130Ile (rs1800961) was included (Zhu et al. 2003). The average spacing between SNPs markers was approximately one per 5.9 kb across this locus. After the pattern and extent of linkage disequilibrium (LD) were examined, a haplotype-based association test was also analyzed.

Statistical power and sample size calculation for association study

To ensure adequate statistical results with an association study, power calculation was simulated with the PS power and sample-size program. Power was the probability correctly rejecting the null hypothesis that the odds ratio was equal between cases and controls. Power in the association study was dependent on several factors: allele frequency, the type 1 error (false-positive) rate, the odds ratio, and the sample size.

Our simulation showed that a sample size of 900 cases and 900 controls would reach 60–80% power to detect an odds ratio as small as 1.2–1.5, when predisposing an allele frequency with a type 1 error rate of 0.05 (Supplementary Fig. 1). In this simulation, the power was low for uncommon rare variants. In general, rare alleles require larger samples to attain the same power, as compared to common frequent alleles. When adjusting the type 1 error rate to less than 0.01 to compensate for multiple testing, much larger samples were required than reported here. Recent reports indicated that the odds ratio with type 2 diabetes was likely within 1.2–1.5 (O'Rahilly et al. 2005; Barroso et al. 2003).

Re-sequencing and discovery of SNPs around a landmark SNP

After a landmark SNP was selected as a putative disease-susceptibility variant within the target region, additional SNPs were arranged by searching the public databases. In addition, to discover unidentified SNPs not listed in databases, we screened SNPs by re-sequencing the genes around a landmark SNP in 24 individuals with type 2 diabetes.

On the basis of the genomic sequence in the public databases, primers were designed using Primer 3 program covering all exons and flanking intronic sequences. In addition, we designed primers to amplify 10 kb upstream of the first exon and 10 kb downstream of the last exon. Details of the primer sequences are available upon request. With few exceptions, PCR primers were designed to amplify an 800–900 bp fragment with a more than 100 bp overlap between amplicons. Amplification primers were also used for sequencing. All PCR products were prepared for sequence analysis by treatment with ExoSAP-IT (Amersham Biosciences, Piscataway, NJ) at 37°C for 15 min, followed by incubation at 80°C for 15 min to deactivate the enzyme. PCR products were sequenced using an ABI BigDye terminator cycle sequencing kit, ver 1.1 (Applied Biosystems) in both directions according to standard protocols. Analysis was performed on an ABI 3100 or ABI 3730xl automated sequencer (both from Applied Biosystems).

Pairwise linkage disequilibrium definition

LD was defined as the statistical association between alleles at two or more sites (Wall and Pritchard 2003). Among them, $|D'|$ and r^2 were used as the standard approach in this study. These values were calculated with FGDS and SNPAlyze ver 3.2.2 Pro software (DYNACOM, Yokohama, Japan).

1. The classical statistical $|D'|$ threshold approach (Lewontin 1964): A LD was defined whenever all pairwise $|D'|$ between three or more successive markers exceeded 0.9. Pairwise $|D'|$ coefficients were calculated using simple sliding window assessment.
2. The value of r^2 definition (Hill and Robertson 1968): r^2 was the square of the correlation coefficient between markers and ranged from 0 to 1. The maximum possible value depended on the minor allele frequency of two markers. Thus, the value of r^2 was 1 when two SNPs arose on the same branch of the genealogy and remained undisrupted by recombination. However, the value was less than 1 when SNPs arose on different branches.

Computational haplotype inference and analysis

In this study, we used two different methods for haplotype inference based on the measured genotyped data, each of which provided different and complementary approaches.

1. By the Bayesian statistical method, haplotypes were implemented from genotype data with PHASE ver 2.1 program. This method regarded the unknown haplotypes as unobserved random quantities and aimed to evaluate their conditional distribution in light of the genotype data. We used the default parameter values (100 iterations, a thinning interval of 1, and a burn-in value of 100) specified in the program to estimate haplotype frequency (Stephens et al. 2001).
2. As an alternative method, haplotype frequencies for multiple loci were estimated with the expectation-maximization (EM) algorithm in SNPAlyze ver 3.2.2 Pro software (DYNACOM). This procedure was an interactive method obtaining maximum-likelihood, and haplotype frequencies were estimated from multi-locus genotype data when the gametic phase was unknown (Ohmori et al. 2004).

In addition, the permutation method was used to test the deviation of allelic frequencies of haplotypes. Distribution of a test statistic was estimated by evaluating the statistics for a random sampling of 10,000 iterated permutations with a fixed total number of both cases and controls, which was incorporated in SNPAlyze ver 3.2.2. The measurement of the P value was estimated by the proportion of permutations for which the permutated data test statistic ($P_{\text{permutated}}$) was greater than the initially observed test statistic (P_{observed}), so permutation $P = P(P_{\text{observed}} > P_{\text{permutated}})$ (Jinnai et al. 2004).

Uniform Resource Locators (URLs) of computer programs used in this study

PS power and sample-size program: http://www.biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize.

Primer 3: http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi.

QVALUE (FDR program): http://www.faculty.washington.edu/~jstorey/qvalue/index.html.

PHASE ver 2.1: http://www.stat.washington.edu/stephens/software.html.

SNPAlyze ver 3.2.2: http://www.dynacom.co.jp/e/products/package/snpalyze/index.html.

Pharma SNP consortium (PSC): http://www.jpma.or.jp/psc/frame-e.html.

URLs of public databases used in this study

HapMap project: http://www.hapmap.org/.

dbSNP (now Build 125 is available): http://www.ncbi.nlm.nih.gov/SNP/.

NCBI Build 35 human genomes: http://www.ncbi.nlm.nih.gov/Genomes/.

PubMed database: http://www.ncbi.nlm.nih.gov/entrez/.

Results

Association test for HNF4α gene variants

Single-marker association test

We evaluated 17 SNPs covering a 100-kb region spanning distal to the pancreatic beta cell P2 promoter through to the end of the coding sequence at an average of one per 5.9 kb. All SNPs were located in the non-coding locus except rs1800961 (exon 4), resulting in a Thr130Ile amino acid change (Zhu et al. 2003). In an association study, 893 control and 925 case samples were used. The results of the association study and the Hardy–Weinberg equilibrium test are summarized in Table 4A. In controls, five SNPs were not confirmed in the Hardy–Weinberg equilibrium test. Three consecutive SNPs (rs2273618, rs6073435, and rs6031601) from intron 7 to 9 were found to be associated with disease (P = 0.0458, 0.0426, and 0.0421; allele model). However, the magnitude of P values was extremely weak and spurious when evaluated by multiple testing (corrected for 17 tests). In our study, the variant with rs6031601 was mostly associated with disease, but was not observed in Caucasians (Winckler et al. 2005; Bagwell et al. 2005; Silander et al. 2004; Love Gregory et al. 2004).

In the FUSION (The Finland-United States Investigation of Non-insulin-dependent diabetes mellitus genetics) study and Ashkenazi Jews, associations were observed with variants in the P2 promoter region. However, a recent study failed to replicate the association with these variants in more than 7,000 samples (Winckler et al. 2005). We were also unable to confirm the association of four SNPs in the P2 promoter region (rs4810424, rs1884613, rs1884614, and rs2144908). In this region, there were quite different allele frequencies among our samples (including HapMap JPT samples) and HapMap CEU samples (Supplementary Table 2A).

Recently, a rare loss-of-function SNPs (rs1800961) in exon 4 was associated with type 2 diabetes in Japanese (Zhu et al. 2003). However, our results failed to confirm the association, although we examined more than twice as many samples.

Table 4 Results of the association study for 17 SNPs around the HNF4α locus (A) and 20 SNPs around a landmark SNP (B)

Cyto-bands	No.	SNP ID	dbSNP ID	Genes	Location	Position	SNP allele	Minor allele frequency		*P* value (first) (367 cases versus 358 controls)				*P* value (second) (558 cases versus 535 controls)				*P* value (overall) (925 cases versus 893 controls)				Odds ratio	95% CI	*P* value
							Allele 1/2	Control	Case	Allele	Geno-type	Domi-nant	Reces-sive	Allele	Geno-type	Domi-nant	Recessive	Allele	Geno-type	Domi-nant	Reces-sive			Hardy Weinberg equilibrium test
(A) 17SNPs around the HNF4α locus																								
q13.12	1	SNP2258	rs4810424	HNF4α	P2promotor	42,408,437	G/C	0.442	0.462	0.08	0.22	0.09	0.29	0.92	0.77	0.77	0.62	0.24	0.52	0.42	0.29	NS	NS	**0.02**
	2		rs1884613	HNF4α	P2promotor	42,413,829	C/G	0.474	0.477	0.27	0.51	0.28	0.46	0.47	0.57	0.30	0.92	0.89	0.89	0.89	0.69	NS	NS	0.93
	3		rs1884614	HNF4α	P2promotor	42,413,933	C/T	0.476	0.482	0.25	0.46	0.23	0.48	0.68	0.59	0.39	0.82	0.68	0.81	0.96	0.53	NS	NS	1.00
	4		rs2144908	HNF4α	P2promotor	42,419,131	G/A	0.477	0.487	0.25	0.42	0.55	0.19	0.86	0.49	0.57	0.43	0.55	0.72	0.42	0.86	NS	NS	0.97
	5		rs6073418	HNF4α	–	42,434,004	C/T	0.220	0.210	0.29	0.11	0.11	0.38	0.89	0.96	0.97	0.78	0.46	0.50	0.31	0.80	NS	NS	0.64
	6	SNP1383	rs2071197	HNF4α	Intron 1	42,463,849	G/A	0.456	0.446	0.19	0.30	0.62	0.12	0.74	0.26	0.23	0.57	0.56	0.17	0.54	0.16	NS	NS	0.28
	7	SNP1384	rs736824	HNF4α	Intron 1	42,468,074	C/T	0.259	0.255	0.23	0.20	0.11	0.82	0.17	0.42	0.22	0.37	0.76	0.66	0.94	0.40	NS	NS	**0.08**
	8		rs1885088	HNF4α	Intron 3	42,472,454	G/A	0.007	0.003	0.97	0.97	0.97	–	0.07	0.07	0.07	–	0.14	0.14	0.14	–	NS	NS	**0.02**
	9		rs1885089	HNF4α	Intron 3	42,472,677	G/A	0.007	0.003	0.68	0.68	–	0.68	0.07	0.07	–	0.07	0.09	0.09	–	0.09	NS	NS	**0.04**
	10		rs1800961	HNF4α	Exon 4	42,475,778	C/T	0.010	0.016	0.21	0.21	–	0.21	0.19	0.19	–	0.19	0.07	0.07	–	0.07	NS	NS	0.14
	11	SNP1385	rs3212198	HNF4α	Intron 5	42,477,776	C/T	0.331	0.352	0.29	0.29	0.12	0.46	0.61	0.72	0.46	0.72	0.18	0.30	0.12	0.68	NS	NS	0.28
	12	SNP2259	rs1028583	HNF4α	Intron 7	42,484,175	T/G	0.306	0.328	0.26	0.26	0.73	0.10	0.47	0.26	0.59	0.26	0.15	0.08	0.84	**0.04281**	NS	NS	**0.04**
	13		rs1028584	HNF4α	Intron 7	42,484,395	A/C	0.336	0.358	0.54	0.54	0.27	0.56	0.37	0.41	0.21	0.41	0.17	0.24	0.10	0.74	NS	NS	0.44
	14		rs2273618	HNF4α	Intron 7	42,485,984	C/T	0.377	0.410	0.34	0.34	0.50	0.14	0.15	0.31	0.45	0.31	**0.04588**	0.11	0.31	**0.03649**	1.15	1.002–1.309	0.45
	15	SNP1386	rs6073435	HNF4α	Intron 8	42,487,002	A/T	0.343	0.375	0.31	0.31	0.13	0.55	0.15	0.29	0.12	0.29	**0.04266**	0.09	**0.02984**	0.38	1.15	1.005–1.318	0.13
	16	SNP2260	rs6031601	HNF4α	Intron 9	42,491,234	C/A	0.351	0.383	0.31	0.31	0.51	0.13	0.15	0.34	0.20	0.34	**0.04211**	0.13	0.16	0.07	1.15	1.005–1.319	0.57
	17		rs911358	HNF4α	–	42,508,442	T/A	0.117	0.117	0.15	0.15	0.05	0.85	0.64	0.77	0.48	0.77	0.98	0.86	0.62	0.92	NS	NS	0.93
(B) 20 SNPs around a landmark SNP locus																								
q11.23	1	SNP2076	rs2425229	DLGAP4	Intron	34,440,834	A/G	0.355	0.370	0.54	0.38	0.38	0.95	–	–	–	–	–	–	–	–	NS	NS	0.76
	2	SNP113	rs2425236	DLGAP4	Intron	34,445,973	G/C	0.357	0.375	0.48	0.26	0.26	0.13	–	–	–	–	–	–	–	–	NS	NS	0.52
	3	SNP1140	rs2425242	DLGAP4	Intron	34,452,333	T/C	0.359	0.372	0.61	0.29	0.29	0.17	–	–	–	–	–	–	–	–	NS	NS	0.46
	4	SNP1143	rs6124857	DLGAP4	Intron	34,484,621	A/G	0.419	0.450	0.24	0.43	0.43	0.20	–	–	–	–	–	–	–	–	NS	NS	0.96
	5	SNP1144	rs1275396	DLGAP4	Intron	34,501,432	G/A	0.356	0.389	0.11	0.23	0.23	0.09	0.21	0.43	0.28	0.30	**0.04339**	0.12	0.19	0.06	1.15	1.004–1.316	0.91
	6		rs1275404	DLGAP4	Intron	34,530,533	A/G	0.245	0.259	0.14	0.28	0.28	0.20	0.99	0.66	0.49	0.72	0.33	0.54	0.84	0.27	NS	NS	0.72
	7		rs1275398	DLGAP4	Intron	34,540,383	G/A	0.060	0.055	0.38	0.48	0.48	0.32	0.13	0.31	0.14	0.54	0.55	0.83	0.54	0.97	NS	NS	0.66
	8	SNP1145	rs220079	DLGAP4	3′-UTR	34,589,556	G/A	0.425	0.472	0.08	0.18	0.18	0.23	**0.02974**	0.09	0.13	0.04	**0.00461**	**0.01812**	**0.02388**	**0.01797**	1.21	1.060–1.380	1.00
	9	SNP1146	rs220076	MYL9	Intron	34,605,971	C/A	0.432	0.483	**0.04055**	0.07	0.07	0.30	**0.02626**	0.06	017	0.02	**0.00231**	**0.01010**	**0.01157**	**0.01507**	1.23	1.077–1.399	0.93
	10		rs694379	–		34,612,910	C/T	0.428	0.474	0.06	0.15	0.15	0.18	**0.04714**	012		0.20	0.05	**0.00563**	**0.02091**	**0.03319**	**0.01643**	1.20	1.056–1.373	0.94
	11		rs6071089	–		34,617,113	C/T	0.497	0.475	0.10	0.20	0.20	0.35	0.43	0.22	0.13	0.79	0.09	0.07	0.02	0.68	NS	NS	0.13
	12	SNP1147	rs85440	TGIF2	Intron	34,648,765	G/A	0.407	0.432	0.11	0.10	0.10	**0.03463**	0.53	0.14	0.09	0.67	0.12	0.25	0.11	0.30	NS	NS	0.18
	13	SNP1148	rs555394	C20orf24	Intron	34,670,739	G/A	0.233	0.259	0.26	0.48	0.48	0.66	–	–	–	–	–	–	–	–	NS	NS	0.40
	14	SNP2081	rs693361	SLA2	Intron	34,684,087	A/G	0.227	0.255	0.21	0.44	0.44	0.25	–	–	–	–	–	–	–	–	NS	NS	0.42
	15	SNP2082	rs221311	NDRG3	Intron	34,719,874	C/T	0.236	0.271	0.13	0.13	0.13	0.94	–				–	–	–	–	NS	NS	0.52
	16	SNP2083	rs221314	NDRG3	Intron	34,732,082	A/G	0.256	0.247	0.69	0.85	0.85	0.60	–	–	–	–	–	–	–	–	NS	NS	0.70
	17	SNP1149	rs221307	NDRG3	Intron	34,748,455	A/G	0.309	0.309	0.99	0.93	0.93	0.78	–	–	–	–	–	–	–	–	NS	NS	0.96
	18	SNP1150	rs6129459	NDRG3	Intron	34,748,581	T/C	0.312	0.306	0.81	0.95	0.95	1.00	–	–	–	–	–	–	–	–	NS	NS	0.87

Table 4 continued

Cyto-bands	No.	SNP ID	dbSNP ID	Genes	Location	Position	SNP allele	Minor allele frequency		*P* value (first) (367 cases versus 358 controls)				*P* value (second) (558 cases versus 535 controls)				*P* value (overall) (925 cases versus 893 controls)				Odds ratio	95% CI	*P* value
							Allele 1/2	Control	Case	Allele	Geno-type	Domi-nant	Reces-sive	Allele	Geno-type	Domi-nant	Reces-sive	Allele	Geno-type	Domi-nant	Reces-sive			Hardy Weinberg equilibrium test
	19	SNP1151	rs6072018	NDRG3	Intron	34,786,132	G/A	0.338	0.331	0.79	0.53	0.53	0.55	–	–	–	–	–	–	–	–	NS	NS	0.19
	20	SNP2084	rs2425266	NDRG3	Intron	34,794,053	C/T	0.322	0.313	0.72	0.55	0.55	0.61	–	–	–	–	–	–	–	–	NS	NS	0.29

(A) Results of genotyping with 10 dbSNPs were added to the analysis for the HNF4a locus. (B) Results with four dbSNPs were added to the analysis for the locus around a landmark SNP. Genes and SNPs positions were based on the NCBI Build 35 human genome sequence database. *P* values were calculated with four type chi-square tests, allele, genotype, dominant and recessive models. Chi-square tests were calculated with the available genotype results. The nominal *P* values < 0.05 are in boldface with five decimal columns. 95% CI denotes 95% confidence interval. *NS* 'not significant'. *P* values with Hardy–Weinberg equilibrium test are shown only in control samples. The *P* values < 0.05 are in boldface

Haplotype-based association test

Nevertheless, it is possible that haplotypes might be associated with increased or decreased disease risks, although individual SNPs are weakly associated with disease. Similar to other populations (Finnish and Ashkenazi Jews), four SNPs in the P2 promoter region showed high LD values ($|D'| > 0.99$ and $r^2 > 0.89$) (Table 5A). With the interpretation of r^2, the coding region was divided into the short LD blocks. Based on the interpretation of r^2, there were two different haplotype blocks around the HNF4α gene, such as one in the P2 promoter region and another in the coding region.

We analyzed these two haplotypes to certify the haplotype-based association test. No haplotypes in the P2 promoter region were associated with disease (data not shown), similar to the results in the single-marker association test. When one haplotype with three SNPs encompassing intron 7–9 was examined, it was weakly associated with controls and cases (Table 6A). However, this association could be spurious because the magnitude of the *P* value failed to satisfy Bonferroni's correction (corrected for 17 plus 8 tests).

Target region on chromosome 20q and SNPs markers

To widen the observations, we enlarged the analysis region on the basis of previous linkage reports. The putative disease susceptibility locus on chromosome 20q was certainly mapped to the 50.81–75.01 cM region, corresponding to a 19.31-Mb interval. Within the genomic region, 291 genes were found in the National Center for Biotechnology Information (NCBI) Build 35 human genome assembly (Table 2).

To establish a useful and dense SNPs marker set, we first searched the public and Celera databases and developed TaqMan assays for 631 SNPs markers. In addition, to collect enough SNPs markers for fine analysis, custom SNPs markers (Custom TaqMan SNP Genotyping Assays) were especially designed for the other 516 positions. As a result, 1,147 SNPs marker probes were assayed, covering the target region with an average of one per 10.8-kb interval. Of the 1,147 SNPs genotyped in the target region, 103 SNPs were excluded by quality control criteria (see Materials and methods), mainly because of deviation in the Hardy–Weinberg equilibrium test (Figs. 1, 2). Of 291 genes within the target region, 219 genes (75.3%) were arranged with at least one SNPs marker. All information regarding the 1,147 SNPs markers is listed in Supplementary Table 3.

Table 5 Pairwise LD values between SNPs, as defined by $|D'|$ and r^2 around the HNF4α locus (A) and a landmark SNP locus (B)

(A) Pairwise LD of 17 SNPs around the HNF4α locus

No.	SNP ID	dbSNP ID	Gene	Location	Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
						rs4810424	rs1884613	rs1884614	rs2144908	rs6073418	rs2071197	rs736824	rs1885088	rs1885089	rs1800961	rs3212198	rs1028583	rs1028584	**rs2273618**	**rs6073435**	**rs6031601**	rs911358
1	SNP2258	rs4810424	HNF4α	P2 promotor	42,408,437		0.991	1	1	0.590	0.077	0.461	0.427	0.445	0.576	0.010	0.013	0.035	0.021	0.018	0.002	0.148
2		rs1884613	HNF4α	P2 promotor	42,413,829	0.893		1	1	0.594	0.112	0.505	0.366	0.372	0.568	0.000	0.021	0.026	0.031	0.036	0.002	0.209
3		rs1884614	HNF4α	P2 promotor	42,413,933	0.902	0.992		1	0.599	0.120	0.511	0.356	0.360	0.581	0.010	0.012	0.020	0.021	0.026	0.008	0.196
4		rs2144908	HNF4α	P2 promotor	42,419,131	0.890	0.978	0.986		0.594	0.125	0.508	0.345	0.345	0.596	0.018	0.004	0.011	0.018	0.022	0.013	0.204
5		rs6073418	HNF4α	-	42,434,004	0.073	0.081	0.083	0.083		0.225	0.188	0.358	0.677	0.017	0.118	0.189	0.150	0.118	0.161	0.003	0.553
6	SNP1383	rs2071197	HNF4α	Intron 1	42,463,849	0.006	0.011	0.013	0.014	0.011		1	0.99986	0.99995	0.99998	0.573	0.492	0.547	0.465	0.414	0.446	0.120
7	SNP1384	rs736824	HNF4α	Intron 1	42,468,074	0.062	0.082	0.085	0.086	0.003	0.295		0.99999	1	0.998	0.048	0.064	0.051	0.053	0.246	0.105	0.065
8		rs1885088	HNF4α	Intron 3	42,472,454	0.001	0.001	0.001	0.001	0.00013	0.004	0.013		1	1	0.99999	1	0.99999	0.99998	1	0.99999	1
9		rs1885089	HNF4α	Intron 3	42,472,677	0.001	0.001	0.001	0.001	0.001	0.004	0.014	1		1	0.99999	1	0.99999	0.99999	0.99999	0.99999	1
10		rs1800961	HNF4α	Exon 4	42,475,778	0.003	0.003	0.003	0.004	0.00001	0.011	0.004	0.00004	0.00005		0.99999	1	0.99999	0.99999	1	0.99999	0.014
11	SNP1385	rs3212198	HNF4α	Intron 5	42,477,776	0.00006	0	0.00005	0.00016	0.007	0.144	0.002	0.008	0.010	0.020		0.892	0.941	0.934	0.779	0.882	0.082
12	SNP2259	rs1028583	HNF4α	Intron 7	42,484,175	0.00009	0.00021	0.00006	0.00001	0.018	0.098	0.001	0.009	0.011	0.023	0.726		0.997	1	0.978	0.841	0.244
13		rs1028584	HNF4α	Intron 7	42,484,395	0.001	0.00038	0.00020	0.00006	0.010	0.136	0.002	0.008	0.009	0.020	0.847	0.871		1	0.851	0.860	0.073
14		**rs2273618**	**HNF4α**	**Intron 7**	**42,485,984**	0.00034	0.001	0.00030	0.00021	0.005	0.123	0.001	0.006	0.007	0.016	0.675	0.709	0.803		1	0.990	0.242
15	**SNP1386**	**rs6073435**	**HNF4α**	**Intron 8**	**42,487,002**	0.00021	0.001	0.00040	0.00027	0.011	0.084	0.011	0.008	0.009	0.019	0.543	0.783	0.668	0.867		0.863	0.038
16	**SNP2260**	**rs6031601**	**HNF4α**	**Intron 9**	**42,491,234**	0	0	0.00004	0.00010	0	0.101	0.002	0.007	0.008	0.018	0.679	0.565	0.667	0.869	0.729		0.247
17		rs911358	HNF4α	-	42,508,442	0.004	0.007	0.006	0.006	0.011	0.002	0.00021	0.001	0.001	0.00001	0.002	0.004	0.002	0.013	0.00039	0.016	

$|D'|$: 0.95 < | ; 0.90 < | ; 0.85 < |

r^2: 0.95 < r^2; 0.90 < r^2; 0.85 < r^2

Haplotype blocks $|D'|$

Haplotype blocks r^2

(B) Pairwise LD of 20 SNPs around a landmark SNP locus

No.	SNP ID	dbSNP ID	Genes	Location	Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
						rs2425229	rs2425236	rs2425242	rs6124857	rs1275396	rs1275404	rs1275398	**rs220079**	**rs220076**	**rs694379**	rs6071089	rs85440	rs555394	rs693361	rs221311	rs221314	rs221307	rs6129459	rs6072018	rs2425266
1	SNP2076	rs2425229	DLGAP4	Intron	34,440,834		1	0.997	0.298	0.096	0.006	0.541	0.100	0.092	0.086	0.051	0.092	0.115	0.134	0.034	0.075	0.025	0.060	0.031	0.109
2	SNP1139	rs2425236	DLGAP4	Intron	34,445,973	0.986		1	0.298	0.104	0.013	0.554	0.109	0.101	0.098	0.069	0.103	0.113	0.139	0.041	0.079	0.010	0.042	0.015	0.095
3	SNP1140	rs2425242	DLGAP4	Intron	34,452,333	0.983	0.997		0.301	0.103	0.013	0.556	0.110	0.103	0.097	0.066	0.100	0.118	0.145	0.042	0.073	0.018	0.053	0.028	0.102
4	SNP1143	rs6124857	DLGAP4	Intron	34,484,621	0.067	0.068	0.069		0.524	0.541	0.555	0.404	0.436	0.407	0.427	0.345	0.596	0.589	0.386	0.158	0.316	0.311	0.290	0.259
5	SNP1144	rs1275396	DLGAP4	Intron	34,501,432	0.009	0.010	0.010	0.216		0.985	0.975	0.889	0.868	0.881	0.946	0.640	0.850	0.821	0.663	0.175	0.462	0.452	0.377	0.376
6		rs1275404	DLGAP4	Intron	34,530,533	0.00002	0.0001	0.0001	0.128	0.535		0.982	0.852	0.823	0.832	0.924	0.838	0.840	0.822	0.598	0.703	0.593	0.591	0.579	0.592
7		rs1275398	DLGAP4	Intron	34,540,383	0.033	0.033	0.033	0.025	0.100	0.183		0.653	0.619	0.653	1	0.635	0.725	0.712	0.667	0.99999	0.894	0.889	0.885	0.892
8	**SNP1145**	**rs220079**	**DLGAP4**	**3′-UTR**	**34,589,556**	0.007	0.008	0.009	0.153	0.580	0.295	0.032		0.978	0.973	0.978	0.892	0.955	0.926	0.767	0.137	0.452	0.440	0.376	0.376
9	**SNP1146**	**rs220076**	**MYL9**	**Intron**	**34,605,971**	0.006	0.007	0.007	0.173	0.541	0.268	0.029	0.933		0.969	0.946	0.890	0.954	0.925	0.751	0.163	0.438	0.421	0.349	0.349
10		**rs694379**	**-**		**34,612,910**	0.005	0.007	0.007	0.155	0.572	0.281	0.032	0.943	0.919		1	0.920	0.981	0.953	0.779	0.119	0.474	0.458	0.374	0.353
11		rs6071089	-		34,617,113	0.001	0.003	0.002	0.138	0.537	0.280	0.061	0.776	0.742	0.814		0.917	0.977	0.944	0.778	0.173	0.730	0.736	0.618	0.593
12	SNP1147	rs85440	TGIF2	Intron	34,648,765	0.007	0.009	0.008	0.111	0.344	0.325	0.035	0.696	0.673	0.738	0.597		0.989	0.976	0.789	0.169	0.512	0.502	0.437	0.416
13	SNP1148	rs555394	C20orf24	Intron	34,670,739	0.007	0.007	0.008	0.151	0.390	0.694	0.102	0.363	0.354	0.382	0.305	0.442		0.970	0.735	0.656	0.629	0.621	0.608	0.620
14	SNP2081	rs693361	SLA2	Intron	34,684,087	0.010	0.010	0.011	0.144	0.352	0.648	0.102	0.332	0.324	0.351	0.277	0.419	0.914		0.770	0.675	0.657	0.649	0.656	0.671
15	SNP2082	rs221311	NDRG3	Intron	34,719,874	0.001	0.001	0.001	0.066	0.248	0.349	0.083	0.245	0.231	0.252	0.203	0.296	0.512	0.542		0.604	0.810	0.806	0.796	0.808
16	SNP2083	rs221314	NDRG3	Intron	34,732,082	0.003	0.004	0.003	0.006	0.006	0.055	0.021	0.005	0.007	0.004	0.010	0.006	0.046	0.047	0.041		0.639	0.611	0.640	0.836
17	SNP1149	rs221307	NDRG3	Intron	34,748,455	0.0002	0.00003	0.000	0.058	0.159	0.261	0.112	0.112	0.103	0.121	0.233	0.163	0.288	0.303	0.505	0.061		0.997	0.987	0.916
18	SNP1150	rs6129459	NDRG3	Intron	34,748,581	0.001	0.0005	0.001	0.057	0.155	0.259	0.109	0.107	0.096	0.115	0.241	0.158	0.279	0.292	0.495	0.056	0.987		0.990	0.908
19	SNP1151	rs6072018	NDRG3	Intron	34,786,132	0.0003	0.0001	0.0002	0.055	0.118	0.222	0.098	0.086	0.073	0.085	0.188	0.134	0.240	0.270	0.434	0.068	0.870	0.883		0.994
20	SNP2084	rs2425266	NDRG3	Intron	34,794,053	0.003	0.002	0.003	0.040	0.108	0.252	0.109	0.080	0.068	0.070	0.159	0.112	0.269	0.304	0.481	0.108	0.805	0.800	0.918	

$|D'|$

r^2

Haplotype blocks $|D'|$

Haplotype block r^2

Results of genotyping data using 725 Japanese samples (367 cases and 358 controls). Each column shows the values of linkage disequilibrium (LD) between pairwise SNPs: $|D'|$ (upper right triangle) and r (lower left triangle). Blue coloring indicates values over 0.95, light blue indicates values ranging from 0.95 and 0.90, and light green indicates values ranging from 0.90 to 0.85. Based on the values of $|D'|$ and r^2 , inferred haplotype block(s) are shown as a gray scale bar at the bottom of each table. SNPs with the nominal $P < 0.05$ by a single-marker association test are in boldface

Selection of a landmark SNP with an extensive association study

In the first association test, we used a total of 1,044 SNPs markers, consisting of 581 TaqMan assays and 463 custom-designed SNPs assays for analysis (Supplementary Table 4). After the first association test, we found a significant association ($P < 0.05$) with 142 SNPs markers (13.6%) in at least one of the four chi-square tests (Table 3).

Subsequently, the initial 142 candidate SNPs were genotyped in the second test samples (independent samples) with the number larger than the first test samples. The genotyping data from each association test were combined and calculated. Of these 142 SNPs, SNP1146 (rs220076) was mostly associated with type 2 diabetes in four chi-square tests ($P = 0.00231$, 0.01010, 0.01157, and 0.01507; allele model) (Table 4B). The unadjusted odds ratio for SNP1146 (rs220076) was 1.23 (95% confidence interval 1.077–1.399). Logistic regression analysis indicated that the association with SNP1146 (rs220076) remained statistically significant after adjusting for age, sex, and BMI ($P = 0.002$). The same result was obtained when statistical significance was assessed with two independent samples, indicating that SNP1146 (rs220076) was a putative susceptibility variant.

With the large number of SNPs examined, it was necessary to exclude false-positive results by multiple testing. We applied standard Bonferroni's correction for multiple testing (corrected for 1,044 tests), but the association with SNP1146 (rs220076) was no longer significant. In addition, we calculated the FDR with 142 SNPs in the second association test as a threshold value of under 0.1 (Shiffman et al. 2005). Unfortunately, among 142 SNPs, there was no SNP under this threshold (Supplementary Fig. 2. and Supplementary Table 5). Taken together, the magnitude of the nominal P value was statistically rather weak after the evaluation of multiple testing, in

Table 6 Inferred haplotype frequency and haplotype-based association test on the HNF4α locus (A) and a landmark SNP locus (B)

(A) HNF4α locus

Haplotype	rs2273618	rs6073435	rs6031601	Frequency (Bayesian method)			Frequency (EM algorithm)			*P* value	Permutation *P* value
	C/T	A/T	C/A	Overall (n = 1,818)	Control (n = 893)	Case (n = 925)	Overall (n = 1,818)	Control (n = 893)	Case (n = 925)		
#1	T	A	A	0.602	0.618	0.586	0.602	0.618	0.586	**0.0403**	**0.0438**
#2	C	T	C	0.328	0.312	0.343	0.328	0.311	0.343	**0.0403**	**0.0436**
#3	C	A	C	0.033	0.034	0.033	0.033	0.034	0.033	0.92	0.93
#4	C	T	A	0.032	0.032	0.033	0.033	0.032	0.033	0.86	0.93
#5	T	A	C	0.005	0.005	0.005	0.005	0.005	0.005	0.69	0.82
#6	C	A	A	< 0.001	< 0.001	< 0.001					
#7	T	T	C								
#8	T	T	A								

(B) Landmark SNP locus

Haplotype	rs220079	rs220076	rs694379	Frequency (Bayesiian method)			Frequency (EM algorithm)			*P* value	Permutation *P* value
	G/A	C/A	C/T	Overall (n = 1,818)	Control (n = 893)	Case (n = 925)	Overall (n = 1,818)	Control (n = 893)	Case (n = 925)		
#1	G	C	C	0.531	0.556	0.507	0.531	0.555	0.508	**0.0044**	**0.0047**
#2	A	A	T	0.443	0.418	0.467	0.444	0.418	0.469	**0.0023**	**0.0028**
#3	G	A	C	0.012	0.010	0.013	0.011	0.010	0.012	0.59	0.58
#4	G	C	T	0.006	0.006	0.005	0.005	0.006	0.005	0.76	0.72
#5	A	C	C	0.004	0.004	0.004	0.004	0.004	0.004	0.95	0.87
#6	G	A	T	0.002	0.003	0.001	0.002	0.003	0.001	0.14	0.12
#7	A	A	C	0.002	0.002	0.001	0.002	0.002	0.001	0.37	0.29
#8	A	C	T	0.001	0.001	0.001	0.001	0.001	0.001	0.55	0.60

Haplotype frequencies were estimated in 1,818 subjects with PHASE ver 2.1 (Bayesian method) and SNPAlyze ver 3.2.2 Pro (EM algorithm). *P* value and permutation *P* value are calculated with SNPAlyze ver 3.2.2 Pro (permutation replicates 10,000). The nominal *P* values < 0.05 are shown in boldface

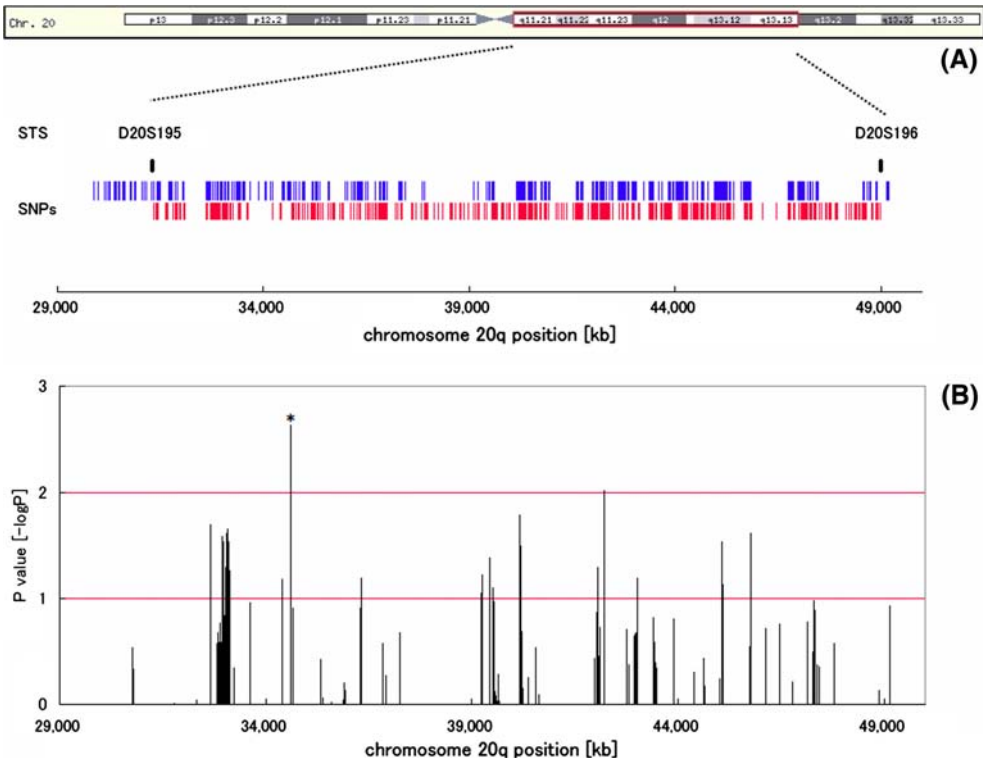

Fig. 2 **a** Map of chromosome 20, STS and 1,044 SNPs markers (after quality control). *Blue bars* indicate 581 SNPs marker (TaqMan SNP Genotyping Assays) positions. *Red bars* indicate additional custom 463 SNPs marker positions, especially designed for this study. *Black bars* show two STS marker positions. **b** Results of *P* value plots by allele frequency chi-square test with 142 SNPs using 925 cases and 893 controls. *Asterisk* indicates SNP1146 (rs220076) with the most significant *P* value ($P = 0.0023$) as a landmark SNP. *Red lines* indicate reference *P* values ($P = 0.1$ and 0.01, respectively). **a** and **b** are illustrated to the same physical scale. The *horizontal axis* shows the physical position on chromosome 20q based on NCBI Build 35 human genome

which the selected variant could not reach the statistical confidence level.

Searching around a landmark SNP

As a landmark, SNP1146 (rs220076) was mapped to the first intron of MYL9 (myosin light polypeptide 9, regulatory: NCBI Gene ID; 10398). Four additional SNPs markers (rs1275404, rs1275398, rs694379, and rs6071089) were selected from the public database and placed around SNP 1146 (rs220076) for fine analysis (Table 4B). SNP1145 (rs220079) and rs694379 adjacent to the landmark SNP also revealed association. SNP1145 (rs220079) was mapped to the 3′-UTR (untranslated region) of DLGAP4 (discs, large [Drosophila] homolog-associated protein 4: Gene ID; 22839).

With the re-sequencing method, we potentially searched for the additional variants in two genes, MYL9 (full length, 10 kb) and DLGAP4 (exons 10, 11, 12, and 3′-UTR, 8 kb). We discovered 13 novel SNPs that were not registered in the public database, but there were no common SNPs (MAF > 0.15) or functional variants (data not shown). Consequently, all novel SNPs were considered to be rare and synonymous variants, and were excluded from further analysis.

By two pairwise measurements of LD ($|D'|$ and r^2), the property of LD was examined around the landmark SNP (Table 5B). In the interpretation of r^2, there was one LD block with extremely high correlation (range from 0.919 to 0.943) among three common SNPs including a landmark SNP. Because LD did not extend outside the region by the interpretation of r^2, one haplotype block was predicted to consist of three common SNPs (rs220079, rs220076, and rs694379).

Haplotype-based association test including landmark SNP

We examined the haplotype-based test by a multi-locus approach, which was a more sensitive method for detecting associations than the assessment of individual SNPs. In the interpretation of r^2, one haplotype block was recognized to show a clear transition zone, which was 23 kb in full length including the landmark SNP (Table 6B). When haplotype frequencies were estimated with both Bayesian method and EM algorithm, two common haplotypes were observed to cover more than 95% data. Because there was no major difference in the results with two methods, the haplotype phase was apparent.

In each haplotype analysis, #1 haplotype was inferred as a protective haplotype, and #2 haplotype was inferred as a risk haplotype. However, the magnitudes of the nominal *P* values (0.0044 and 0.0023) were weak after Bonferroni's correction (corrected for 1,044 plus

eight tests), suggesting that there was no significant difference in each haplotype.

Discussion

Although a large quantity of information about the human genome is now available in genetic studies, tremendous efforts have been made to detect genetic polymorphisms contributing to the susceptibility to common complex human diseases. Because complex diseases are caused by several interacting genetic determinants, most of which have small effects, it is difficult to clearly identify the susceptibility locus (Wang et al. 2005; Hirschhorn and Daly 2005; Togawa et al. 2006). However, an association study using common variants has been proposed as a powerful strategy to search for genuine variants that underlie complex traits. With intensive association studies, common variants have important roles in common complex diseases, such as rheumatoid arthritis (Suzuki et al. 2003), ischemic heart disease (Ozaki et al. 2002), inflammatory bowel disease (Hugot et al. 2001), and diabetes mellitus (Grant et al. 2006). Under the current tendency, we designed an extensive association study and succeeded in identifying the susceptibility gene(s) and variant(s) related to common complex diseases as previously reported (Hamada et al. 2005; Kato et al. 2006).

Type 2 diabetes is a complex disease influenced by multiple environmental and genetic factors, and a number of potential loci have been proposed by whole-genome linkage analysis. In particular, several groups have focused on chromosome 20q, in which evidence of linkage was shown in the same region in different populations (Ghosh et al. 1999, 2000; Vionnet et al. 2000; Permutt et al. 2001; Duggirala et al. 2001; Luo et al. 2001; Mori et al. 2002; Iwasaki et al. 2003). It was suggested that chromosome 20q was a promising locus including certain susceptibility variant(s) or gene(s) with type 2 diabetes. In this study, we regarded HNF4α as a candidate gene with type 2 diabetes to undertake a detailed association study. Simultaneously, to reveal whether other variant(s) were associated with disease on chromosome 20q11.21-13.13, a total of 1,044 SNPs were extensively examined in 1,818 Japanese samples. In part of all the tests, the nominal *P* values were weakly associated with disease. Although we could not completely rule out the possibility of multiple and minor genetic contributions on this locus, we propose that these results are statistically non-significant findings after adjusting strictly for multiple testing.

There are several explanations for our inability to detect the disease-susceptibility variant with a statistical confidence level. First, it is possible that the initial linkage studies might overestimate chromosome 20q as a candidate locus. This hypothesis seems most likely since a recent Japanese study failed to confirm the previous linkage evidence on chromosome 20q. Similarly, in Chinese, two recent reports indicated a lack of evidence for previous linkage on chromosome 20q (Xiang et al. 2004; Ng et al. 2004). In addition, several groups have regarded HNF4α as a disease susceptibility gene due to its chromosomal location under a linkage peak and extensively searched for susceptibility variants (Winckler et al. 2005; Bagwell et al. 2005; Silander et al. 2004; Love Gregory et al. 2004; Vaxillaire et al. 2005). However, in an association study involving more than 7,000 subjects, HNF4α variants showed a complete lack of confirmation of disease susceptibility (Winckler et al. 2005). We also propose that variants in the HNF4α gene have no major effect on the risk of disease in Japanese. Despite successful cases (Grant et al. 2006), it will be difficult to narrow down the regions of linkage evidence for type 2 diabetes (Permutt et al. 2002, 2005).

A second possibility is that our results are false negative (type 2 error). In the setting of an association study, several factors could cause us to miss the true variant. It is likely that we lacked the power to detect a subtle association due to relatively small sample numbers in the first association test (380 vs.380 samples). With prior simulation, our samples (1,818 samples) were sufficient to detect the susceptibility locus with a power less than 80% as the maximum. Nevertheless, a more powered study involving thousands of subjects may uncover the genuine susceptibility variant(s), because it is predicted that true variant(s) have only a modest effect on type 2 diabetes. However, it will be a laborious procedure to collect enough late onset type 2 diabetes samples, making it difficult to replicate the findings.

As the major weak point, controls were unmatched with cases for age. We suppose the interpretation to be reliable since logistic regression analysis showed a significant association after adjusting for age. However, this analysis could not be entirely correct because of the sampling bias. If the power is affected by bias and is less than the value with prior simulation, it is possible to overlook the genuine association on the 20q locus. Within the framework of this study, it was difficult to evaluate the actual effect of sampling bias, even if it was present. Thus, as one favored approach, it is proposed to confirm the results with other independent

populations, such as controls in a large cohort study (Hattersley and McCarthy 2005).

We supposed that marker density must be relatively dense to completely avoid false negatives. It was not possible to place SNPs markers at less than 10-kb intervals due to the experimental cost and time, but our approach is expected to ensure a practical solution (Hamada et al. 2005; Kato et al. 2006). Additionally, an unknown causal variant was clearly detected with this approach using evenly spaced SNPs markers at 8–5 kb density (Edwards et al. 2005).

It is also proposed that an evenly spaced marker might not be a good guide for the design of an association study, because there is massive variability in the extent of LD that overstates or understates the useful range (Reich et al. 2002). To elucidate this question, the HapMap project established the generality of human haplotypes in multiple populations. Unfortunately, the HapMap resource was not completely available at the onset of this study, and it was necessary to design a different strategy, as presented here.

Furthermore, it was supposed that Bonferroni's correction would be too conservative and would likely to lead to a false negative result, because SNPs were not independent due to strong LD. To examine this point, the number of LD blocks was calculated by the definition of $|D'|$, which indicated that the target region contained 207 LD blocks (unpublished data). However, to evaluate Bonferroni's correction with the number of LD blocks (i.e., 0.05 divided by the number of LD blocks), the nominal P value with a landmark SNP did not reach the statistical confidence level.

The other possibility is that there is heterogeneity among different populations. For instance, the allele frequencies of some variants in HNF4α are quite different between Caucasians and Asians. The G allele for rs1884613 had a frequency of 0.19 in HapMap CEU, which showed a causative variant in the initial two reports (Love Gregory et al. 2004; Silander et al. 2004). In contrast, the same allele had a different frequency of 0.47, indicating a lack of association in our study. A difference in allele frequencies between populations may suggest specific population-attributed risks of type 2 diabetes. Hence, our observations could not completely rule out evidence of association on chromosome 20q in other ethnic groups. To explain this hypothesis, an additional proof of heterogeneity should be documented.

In this study, we presented a systematic gene-based and two-stage association study for the assessment of multiple markers across a relatively long range on chromosome 20q, and showed novel data within the framework. In spite of the initial notable prediction with linkage evidence, our results are lacking in evidence that certain SNPs had a strong effect on the risk of type 2 diabetes within the target region. Under these association study conditions, it is likely that many reasonably powered studies would produce negative results, which do not attract attention due to publication bias (Blomqvist et al. 2006; Freimer and Sabatti 2005). However, we are convinced that these data could help to resolve the important role of variant(s) on human chromosome 20q.

Acknowledgments We are very grateful to Dr. Naoyuki Kamatani (Division of Statistical Genetics, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University) for the critical reading of the manuscript and helpful comments. We thank the members of the Institute for Genome Research for helpful discussions and assistance. This study was supported by a grant from the Cooperative Link of Unique Science and Technology for Economy Revitalization (CLUSTER) and Takeda Science Foundation (Osaka, Japan).

References

Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. Nature 437:1299–1320

Bagwell AM, Bento JL, Mychaleckyj JC, Freedman BI, Langefeld CD, Bowden DW (2005) Genetic analysis of HNF4α polymorphisms in Caucasian-American type 2 diabetes. Diabetes 54:1185–1190

Barroso I, Luan J, Middelberg RP, Harding AH, Franks PW, Jakes RW, Clayton D, Schafer AJ, O'Rahilly S, Wareham NJ (2003) Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. PLoS Biol 1:E20

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B 57:289–300

Blomqvist ME, Reynolds C, Katzov H, Feuk L, Andreasen N, Bogdanovic N, Blennow K, Brookes AJ, Prince JA (2006) Towards compendia of negative genetic association studies: an example for Alzheimer disease. Hum Genet 119:29–37

Duggirala R, Blangero J, Almasy L, Arya R, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP (2001) A major locus for fasting insulin concentrations and insulin resistance on chromosome 6q with strong pleiotropic effects on obesity-related phenotypes in nondiabetic Mexican Americans. Am J Hum Genet 68:1149–1164

Edwards AO, Ritter R 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA (2005) Complement factor H polymorphism and age-related macular degeneration. Science 308:421–424

Freimer NB, Sabatti C (2005) Guidelines for association studies in human molecular genetics. Hum Mol Genet 14:2481–2483

Ghosh S, Watanabe RM, Hauser ER, Valle T, Magnuson VL, Erdos MR, Langefeld CD, Balow Jr 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, Unni A, Segal L, Sharaf R, Blaschak Harvan J, Eriksson J, Tenkula T, Vidgren G, Ehnholm C, Tuomilehto Wolf E, Hagopian W, Buchanan TA, Tuomilehto J, Bergman RN, Collins FS, Boehnke M (1999) Type 2 diabetes: evidence for linkage on

chromosome 20 in 716 Finnish affected sib pairs. Proc Natl Acad Sci USA 96:2198–2203

Ghosh S, Watanabe RM, Valle TT, Hauser ER, Magnuson VL, Langefeld CD, Ally DS, Mohlke KL, Silander K, Kohtamaki K, Chines P, Balow Jr J, Birznieks G, Chang J, Eldridge W, Erdos MR, Karanjawala ZE, Knapp JI, Kudelko K, Martin C, Morales Mena A, Musick A, Musick T, Pfahl C, Porter R, Rayman JB (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. Am J Hum Genet 67:1174–1185

Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38:320–323

Hamada D, Takata Y, Osabe D, Nomura K, Shinohara S, Egawa H, Nakano S, Shinomiya F, Scafe CR, Reeve VM, Miyamoto T, Moritani M, Kunika K, Inoue H, Yasui N, Itakura M (2005) Association between single-nucleotide polymorphisms in the SEC8L1 gene, which encodes a subunit of the exocyst complex, and rheumatoid arthritis in a Japanese population. Arthritis Rheum 52:1371–1380

Hattersley AT, McCarthy MI (2005) What makes a good genetic association study? Lancet 366:1315–1323

Hill WG, Robertson A (1968) The effects of inbreeding at loci with heterozygote advantage. Genetics 60:615–628

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307:1072–1079

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108

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 (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411:599–603

Iwasaki N, Cox NJ, Wang YQ, Schwarz PE, Bell GI, Honda M, Imura M, Ogata M, Saito M, Kamatani N, Iwamoto Y (2003) Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects. Diabetes 52:209–213

Jinnai N, Sakagami T, Sekigawa T, Kakihara M, Nakajima T, Yoshida K, Goto S, Hasegawa T, Koshino T, Hasegawa Y, Inoue H, Suzuki N, Sano Y, Inoue I (2004) Polymorphisms in the prostaglandin E2 receptor subtype 2 gene confer susceptibility to aspirin-intolerant asthma: a candidate gene approach. Hum Mol Genet 13:3203–3217

Kato H, Nomura K, Osabe D, Shinohara S, Mizumori O, Katashima R, Iwasaki S, Nishimura K, Yoshino M, Kobori M, Ichiishi E, Nakamura N, Yoshikawa T, Tanahashi T, Keshavarz P, Kunika K, Moritani M, Kudo E, Tsugawa K, Takata Y, Hamada D, Yasui N, Miyamoto T, Shiota H, Inoue H, Itakura M (2006) Association of single-nucleotide polymorphisms in the suppressor of cytokine signaling 2 (SOCS2) gene with type 2 diabetes in the Japanese. Genomics 87:446–458

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389

Lewontin RC (1964) The interaction of selection and linkage I. General considerations. Genetics 49:49–67

Love Gregory LD, Wasson J, Ma J, Jin CH, Glaser B, Suarez BK, Permutt MA (2004) A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 alpha gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an Ashkenazi Jewish population. Diabetes 53:1134–1140

Luo TH, Zhao Y, Li G, Yuan WT, Zhao JJ, Chen JL, Huang W, Luo M (2001) A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. Diabetologia 44:501–506

McIntyre LM, Martin ER, Simonsen KL, Kaplan NL (2000) Circumventing multiple testing: a multilocus Monte Carlo approach to testing for association. Genet Epidemiol 19:18–29

Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecoeur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P (2002) 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

Nawata H, Shirasawa S, Nakashima N, Araki E, Hashiguchi J, Miyake S, Yamauchi T, Hamaguchi K, Yoshimatsu H, Takeda H, Fukushima H, Sasahara T, Yamaguchi K, Sonoda N, Sonoda T, Matsumoto M, Tanaka Y, Sugimoto H, Tsubouchi H, Inoguchi T, Yanase T, Wake N, Narazaki K, Eto T, Umeda F, Nakazaki M, Ono J, Asano T, Ito Y, Akazawa S, Hazegawa I, Takasu N, Shinohara M, Nishikawa T, Nagafuchi S, Okeda T, Eguchi K, Iwase M, Ishikawa M, Aoki M, Keicho N, Kato N, Yasuda K, Yamamoto K, Sasazuki T (2004) Genome-wide linkage analysis of type 2 diabetes mellitus reconfirms the susceptibility locus on 11p13-p12 in Japanese. J Hum Genet 49:629–634

Ng MC, So WY, Cox NJ, Lam VK, Cockram CS, Critchley JA, Bell GI, Chan JC (2004) Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. Diabetes 53:1609–1613

Ohmori H, Ando Y, Makita Y, Onouchi Y, Nakajima T, Saraiva MJ, Terazaki H, Suhr O, Sobue G, Nakamura M, Yamaizumi M, Munar Ques M, Inoue I, Uchino M, Hata A (2004) Common origin of the Val30Met mutation responsible for the amyloidogenic transthyretin type of familial amyloidotic polyneuropathy. J Med Genet 41:e51

O'Rahilly S, Barroso I, Wareham NJ (2005) Genetic factors in type 2 diabetes: the end of the beginning? Science 307:370–373

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654

Permutt MA, Wasson J, Cox N (2005) Genetic epidemiology of diabetes. J Clin Invest 115:1431–1439

Permutt MA, Wasson J, Suarez BK, Lin J, Thomas J, Meyer J, Lewitzky S, Rennich JS, Parker A, DuPrat L, Maruti S, Chayen S, Glaser B (2001) A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. Diabetes 50:681–685

Permutt MA, Wasson J, Love Gregory LD, Ma J, Skolnick G, Suarez B, Lin J, Glaser B (2002) Searching for type 2 diabetes genes on chromosome 20. Diabetes 51(Suppl 3):S308–S315

Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 32:135–142

Shiffman D, Ellis SG, Rowland CM, Malloy MJ, Luke MM, Iakoubova OA, Pullinger CR, Cassano J, Aouizerat BE, Fenwick RG, Reitz RE, Catanese JJ, Leong DU, Zellner C, Sninsky JJ, Topol EJ, Devlin JJ, Kane JP (2005) Identification of four gene variants associated with myocardial infarction. Am J Hum Genet 77:596–605

Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS (2004) Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. Diabetes 53:1141–1149

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Storey J (2002) A direct approach to false discovery rates. J Roy Stat Soc Ser B 64:479–498

Storey J, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100:9440–9445

Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, Nagasaki M, Nakayama Hamada M, Kawaida R, Ono M, Ohtsuki M, Furukawa H, Yoshino S, Yukioka M, Tohma S, Matsubara T, Wakitani S, Teshima R, Nishioka Y, Sekine A, Iida A, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet 34:395–402

Togawa K, Moritani M, Yaguchi H, Itakura M (2006) Multidimensional genome scans identify the combinations of genetic loci linked to diabetes-related phenotypes in mice. Hum Mol Genet 15:113–128

Vaxillaire M, Dina C, Lobbens S, Dechaume A, Vasseur Delannoy V, Helbecque N, Charpentier G, Froguel P (2005) Effect of common polymorphisms in the HNF4α promoter on susceptibility to type 2 diabetes in the French Caucasian population. Diabetologia 48:440–444

Vionnet N, Hani El H, Dupont S, Gallina S, Francke S, Dotte S, De Matos F, Durand E, Lepretre F, Lecoeur C, Gallina P, Zekiri L, Dina C, Froguel P (2000) Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. Am J Hum Genet 67:1470–1480

Wall JD, Pritchard JK (2003) Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet 4:587–597

Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genomewide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118

Winckler W, Graham RR, de Bakker PI, Sun M, Almgren P, Tuomi T, Gaudet D, Hudson TJ, Ardlie KG, Daly MJ, Hirschhorn JN, Groop L, Altshuler D (2005) Association testing of variants in the hepatocyte nuclear factor 4alpha gene with risk of type 2 diabetes in 7,883 people. Diabetes 54:886–892

Xiang K, Wang Y, Zheng T, Jia W, Li J, Chen L, Shen K, Wu S, Lin X, Zhang G, Wang C, Wang S, Lu H, Fang Q, Shi Y, Zhang R, Xu J, Weng Q (2004) Genome-wide search for type 2 diabetes/impaired glucose homeostasis susceptibility genes in the Chinese: significant linkage to chromosome 6q21-q23 and chromosome 1q21-q24. Diabetes 53:228–234

World Health Organization (1985) Diabetes mellitus: report of a WHO study group. Geneva World Health Org, Tech. Rep. Ser., no. 727

Zhu Q, Yamagata K, Miura A, Shihara N, Horikawa Y, Takeda J, Miyagawa J, Matsuzawa Y (2003) T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. Diabetologia 46:567–573