ORIGINAL ARTICLE

Shiro Maeda · Shuichi Tsukada · Akio Kanazawa
Akihiro Sekine · Tatsuhiko Tsunoda · Daisuke Koya
Hiroshi Maegawa · Atsunori Kashiwagi
Tetsuya Babazono · Masafumi Matsuda
Yasushi Tanaka · Tomoaki Fujioka · Hiroshi Hirose
Takashi Eguchi · Yoichi Ohno · Christopher J. Groves
Andrew T. Hattersley · Graham A. Hitman
Mark Walker · Kohei Kaku · Yasuhiko Iwamoto
Ryuzo Kawamori · Ryuichi Kikkawa
Naoyuki Kamatani · Mark I. McCarthy
Yusuke Nakamura

Genetic variations in the gene encoding *TFAP2B* are associated with type 2 diabetes mellitus

Abstract To search a gene(s) conferring susceptibility to type 2 diabetes mellitus, we genotyped nearly 60,000 gene-based SNPs for Japanese patients and found evidence that the gene at chromosome 6p12 encoding transcription-factor-activating protein 2β (*TFAP2B*) was a likely candidate in view of significant association of polymorphism in this gene with type 2 diabetes. Extensive analysis of this region identified that several variations within *TFAP2B* were significantly associated with type 2 diabetes [a variable number of tandem repeat

S. Maeda (✉) · S. Tsukada · A. Kanazawa
Laboratory for Diabetic Nephropathy,
SNP Research Center, The Institute of Physical
and Chemical Research, 1-7-22 Suehiro-cho,
Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
E-mail: smaeda@src.riken.jp
Tel.: +81-45-5039595
Fax: +81-45-5039567

A. Sekine
Laboratory for Genotyping,
SNP Research Center,
The Institute of Physical and Chemical Research,
Kanagawa, Japan

T. Tsunoda
Laboratory for Medical Informatics,
SNP Research Center,
The Institute of Physical and Chemical Research,
Kanagawa, Japan

D. Koya · H. Maegawa · A. Kashiwagi · R. Kikkawa
Department of Medicine,
Shiga University of Medical Science, Shiga, Japan

T. Babazono · Y. Iwamoto
Diabetes Center, Tokyo Women's Medical University,
Tokyo, Japan

M. Matsuda · K. Kaku
Division of Endocrinology and Metabolism,
Department of Internal Medicine,
Kawasaki Medical School, Okayama, Japan

Y. Tanaka · R. Kawamori
Department of Medicine, Metabolism and Endocrinology,
School of Medicine, Juntendo University,
Tokyo, Japan

T. Fujioka
Department of Urology,
Iwate Medical University School of Medicine,
Iwate, Japan

H. Hirose · T. Eguchi · Y. Ohno
Department of Internal Medicine,
Keio University School of Medicine, Tokyo, Japan

C. J. Groves · M. I. McCarthy
Oxford Centre for Diabetes,
Endocrinology and Metabolism, Oxford, UK

C. J. Groves · M. I. McCarthy
Wellcome Trust Centre for Human Genetics, Oxford, UK

A. T. Hattersley
Centre for Molecular Genetics,
Peninsula Medical School, Exeter, UK

G. A. Hitman
Centre of Diabetes and Metabolic Medicine,
Bart's and The London,
Queen Mary's School of Medicine and Dentistry, London, UK

M. Walker
School of Clinical Medical Sciences,
University of Newcastle, Newcastle, UK

N. Kamatani
Institute of Rheumatology,
Tokyo Women's Medical University, Tokyo, Japan

N. Kamatani · Y. Nakamura
Research Group for Personalized Medicine,
SNP Research Center, The Institute of Physical and
Chemical Research, Kanagawa, Japan

Y. Nakamura
Laboratory of Molecular Medicine,
Human Genome Center, The Institute of Medical Science,
University of Tokyo, Tokyo, Japan

locus: $\chi^2 = 10.9$, $P = 0.0009$; odds ratio = 1.57, 95% CI 1.20–2.06, intron $1+774$ (G/T); $\chi^2 = 11.6$, $P = 0.0006$; odds ratio = 1.60, 95% CI 1.22–2.09, intron $1+2093$ (A/C); $\chi^2 = 12.2$, $P = 0.0004$; odds ratio = 1.61, 95% CI 1.23–2.11]. The association of *TFAP2B* with type 2 diabetes was also observed in the UK population. These results suggest that *TFAP2B* might be a new candidate for conferring susceptibility to type 2 diabetes and contribute to the pathogenesis of type 2 diabetes.

Keywords Type 2 diabetes · Variable number of tandem repeats (VNTR) · Single nucleotide polymorphism (SNP) · Adipocytes · Association study

Introduction

Type 2 diabetes mellitus (DM) affects more than one hundred million individuals worldwide (Zimmet et al. 2001). Its pathogenesis appears to be the consequence of insulin resistance in peripheral tissues combined with dysfunction of β cells in pancreatic islets although the precise mechanism is still not well known (Kahn 1998; Saltiel 2001).

That genetic factors contribute to the onset and progression of DM is undoubted, and several genes responsible for specific forms of the disease, such as maturity-onset diabetes of the young (MODY) or mitochondrial diabetes, have been identified (Fajans et al. 2001; Kadowaki et al. 1994). However, genetic alterations associated with these specific forms of diabetes account for only a small subset of cases; the gene or genes conferring susceptibility to type 2 diabetes in most patients remain to be identified.

Worldwide efforts to sequence the entire human genome have established a nearly complete blueprint (International Human Genome Sequencing Consortium 2001), providing a large body of information regarding genes whether their functions are already known or not. Single nucleotide polymorphisms (SNPs), the type of genetic variation found most frequently throughout the sequenced genome, have become useful markers for identifying genes involved in common diseases, such as DM. We developed a high-throughput SNP genotyping system that combined the Invader assay with multiplex polymerase chain reactions (PCRs) (Ohnishi et al. 2001) and undertook genome-wide association studies using SNPs to discover loci involved in susceptibility to common diseases.

In the study presented here, we show the results of a large-scale, case-control study using nearly 60,000 gene-based SNPs as genetic markers and provide the evidence that the gene encoding *TFAP2B* at chromosome 6p12 might be a novel candidate conferring susceptibility to type 2 diabetes.

Subjects and methods

Subjects and DNA preparations

DNA samples were obtained from patients with type 2 diabetes who come regularly to the outpatient clinics of Shiga University of Medical Science, Tokyo Women's Medical University, Juntendo University, Kawasaki Medical School, Keio University School of Medicine or Iwate Medical University. Control individuals consisted of 470 members of the general population (control 1) and another set of the general population (control 2, $n = 889$) who were recruited through several medical institutes in Japan. We also used a third set of control subjects with normal plasma glucose levels (HbA1c < 5.5% or fasting plasma glucose < 100 mg/dl and no family history for diabetes, control 3, $n = 598$), for final analysis. Written informed consent was obtained from each patient, and DNA extraction was performed using a standard phenol-chloroform procedure. The UK samples comprised 590 cases with type 2 diabetes enriched for positive family history (probands from the Diabetes UK Warren 2 repository) (Wiltshire et al. 2001) and 549 UK population controls (the ECACC-HRC collection) (Groves et al. 2003).

Genotyping for gene-based SNPs

The SNPs for genotyping experiments were selected randomly from the IMS-JST Japanese SNPs database (http://snp.ims.u-tokyo.ac.jp) (Hirakawa et al. 2002; Haga et al. 2002). The genotype at each SNP locus was analyzed with the Invader assay, as previously described (Ohnishi et al. 2001). We screened 188 diabetic patients at first, and genotype and/or allele frequencies were compared with those of the general population. After evaluating the statistical data using 2×3 or 2×2 contingency tables, SNPs that showed significant differences in genotype or allele frequencies between diabetic patients and the general population were examined further in another larger set of diabetics ($n = 631$). The protocol was approved by the ethics committee of the Institute of Physical and Chemical Research.

Discovery of SNPs in the TFAP2B gene, and genotyping

On the basis of GenBank information about DNA sequences in the genomic region containing the *TFAP2B*

gene (accession number: NT_007592), we designed PCR primers to amplify appropriate fragments of genomic DNA. Repetitive elements were excluded from the search by invoking the REPEAT MASKER computer program, in the manner described previously (Seki et al. 2000). PCR reactions and DNA sequencing were carried out, as previously described (Saito et al. 2001). The SNPs in this region were genotyped by means of Invader (Japanese samples) or Amplifluor assays (UK samples) (Bengra et al. 2002), and VNTR loci were analyzed with respect to allele sizes using the Applied Biosystems ABI PRISM 3700 Automated DNA Sequencer and GeneScan software (GenoTyper program).

Reverse transcription and polymerase chain reactions

First-strand cDNA was prepared by reverse transcription (RT) of total RNA extracted from the murine 3T3-L1 cells by oligo-dT priming using Superscript II reverse transcriptase (Invitrogen). Human cDNAs from multiple tissues were obtained from CLONTECH Inc. (Palo Allto, CA, USA). The first-strand or double-strand cDNAs were amplified by PCR experiments using primers mAP2RT-F (5′-GCG TCC TCA GAA GAG CCA AAT C-3′) and mAP2RT-R (5′-GTG CGT GAT GAG ACT GAA GTG C-3′) for murine *TFAP2B* and hAP2RT-F (5′-CCA AAT CTG TGA CTT CTC TAA TGA-3′) and hAP2RT-R (5′-GTA ACG TGA CAT TTG CTG CTT TG-3′) for human *TFAP2B*. Real time quantitative RT-PCR for human *TFAP2B* was performed by TaqMan assay using primers hAP2BTM-F (5′-TTG AAC CGG CAG CAC ACA-3′), hAP2BTM-R (5′-CTT GGT GGC CAA CAG CAT ATT-3′) and probe hAP2BTM-P (5′(FAM)-CCG AGT GAC CTG CAC TCC CGA AA-(TAMRA)3′).

Statistical analysis

Statistical methods for determining associations, haplotype frequencies and to calculate linkage disequilibrium (LD) coefficients (Δ) were described previously (Yamada et al. 2001). Analysis of haplotype structure was carried out by estimating haplotype phasing using the EM algorithm (Excoffier and Slatkin 1995) and by constructing haplotype blocks, as previously described (Ozaki et al. 2002; Daly et al. 2001).

For the simulation approach, to calculate the actual type 1 error rate in our study, we simulated exactly the process of the third tests (first test: case 1 versus control 1; second test: case 2 versus control 1; third test: case 2 versus control 2) using the Monte-Carlo method. Since the two sets of cases (cases 1 and 2) and the two sets of controls (controls 1 and 2) were all collected independently, two alleles were independently drawn for each subject, assuming that the frequency of the minor allele was the same for all the four groups. This means that the simulations were performed under the null hypothesis.

For the simulations, the frequency of the minor alleles was changed from 0.02 to 0.5. After the genotypes of all subjects were determined, the cases and controls were compared in four different ways (first and second tests). Thus, (1) the differences in allele frequencies were tested using the allele frequency 2×2 contingency tables; (2) the differences in the frequencies of the subjects with minor alleles were then tested using the 2×2 contingency tables; (3) the differences in the frequencies of the subjects with major alleles were tested; and (4) the frequencies of three genotypes were tested using 3×2 contingency tables. All tests were done using the Pearson's chi-square test. The Monte-Carlo simulation was performed using Mersenne Twister uniform pseudo-random number generator (Matsumoto and Nishimura 1998).

For step-wise logistic regression analysis, the probability Pc of an individual of being a case rather than a control was assumed to be affected by a set of SNPs according to the logistic model: $logit(Pc) = a_0 + a_1 x_1 + a_2 x_2$ for single SNP, for example. Here, we used a coding scheme $x_1 = -1, 0, 1$, and $x_2 = -0.5, 0.5, -0.5$ for genotypes 1/1, 1/2 and 2/2, respectively, for representing an additive effect by x_1 and a dominance/recessive effect by x_2 (Cordell and Clayton 2002). The weights were estimated by the maximum-likelihood method and tested by comparison with the null-hypothesis logit $(Pc) = a_0$ (constant). For multiple SNPs, interaction effects were added further in addition to the main effects of additional SNPs and tested step-wise whether their effects were significant or not. The tests were performed using *R*. We applied both strategies of forward selection (starting from one SNP) and backward selection (starting from all SNPs) until the most significant SNP set were obtained (Cordell and Clayton 2002).

Results

Association study

We first genotyped 188 Japanese patients with type 2 diabetes (case 1) at 58,266 SNP loci and compared their allelic or genotype frequencies at these loci with those in the general population (control 1) (first test). At each locus, we tested the differences between the two populations by the four ways described in the Methods section; 1,496 SNP loci revealed a *P* value of <0.01 by at least one of the four tests (Table 1). We then analyzed these 1,496 loci for the second test using another larger group of patients (case 2). When case 2 and control 1 were compared by the four ways described in the Methods section, we found that the distribution of genotypes at a landmark SNP locus in the second intron of the *TFAP2B* gene on chromosome 6p12 was most strongly associated with type 2 diabetes [GG versus GC+CC: $\chi^2 = 15.9$, $P = 0.00007$, odds ratio = 1.65, 95% CI 1.29–2.11, G versus C: $\chi^2 = 11.3$, $P = 0.0007$, odds ratio = 1.38, 95% CI 1.14–1.66, (Tables 2 and 3)]. Furthermore, we compared the frequency of the alleles

Table 1 Distribution of P values in the first test[a]

	Genotype: n (%)	Allele: n (%)	Dominant model: n (%)	Recessive model: n (%)	Overall: n (%)
$P < 0.01$	626 (1.0)	755 (1.3)	605 (1.0)	630 (1.1)	1,496 (2.6)
$0.01 \leqq P < 0.05$	2,435 (4.1)	2,691 (4.5)	2,323 (4.0)	2,380 (4.1)	4,931 (8.5)
$0.05 \leqq P < 0.1$	2,899 (5.0)	3,121 (5.4)	3,146 (5.4)	3,105 (5.3)	5,816 (10.0)
$0.1 \leqq P < 0.2$	5,784 (9.9)	6,138 (10.5)	5,671 (9.7)	5,752 (9.9)	9,421 (16.2)
$0.2 \leqq P < 0.3$	5,762 (9.8)	5,725 (9.8)	5,573 (9.6)	5,836 (10.0)	8,224 (14.1)
$0.3 \leqq P < 0.4$	5,562 (9.5)	5,926 (10.2)	5,405 (9.3)	5,460 (9.4)	6,945 (11.9)
$0.4 \leqq P < 0.5$	5,696 (9.8)	5,935 (10.2)	5,873 (10.0)	5,993 (10.3)	6,504 (11.2)
$0.5 \leqq P < 0.6$	5,810 (10.0)	5,703 (9.8)	5,909 (10.1)	5,842 (10.0)	5,347 (9.2)
$0.6 \leqq P < 0.7$	5,707 (9.8)	5,584 (9.6)	5,175 (8.9)	5,255 (9.0)	3,789 (6.5)
$0.7 \leqq P < 0.8$	5,798 (10.0)	5,623 (9.7)	5,316 (9.1)	5,478 (9.4)	2,943 (5.1)
$0.8 \leqq P < 0.9$	5,760 (9.9)	5,592 (9.6)	5,157 (8.9)	5,356 (9.2)	1,891 (3.2)
$0.9 \leqq P \leqq 1$	6,427 (11.0)	5,523 (9.5)	8,113 (13.9)	7,179 (12.3)	959 (1.6)
Total	58,266	58,266	58,266	58,266	58,266

[a] Data are presented as number of SNPs

Table 2 Distribution of P values in the second test[a]

	Genotype	Allele	Dominant	Recessive	Overall
$P < 0.0001$	0	0	1[b]	0	1[b]
$0.0001 \leqq P < 0.001$	5	15	6	4	17
$0.001 \leqq P < 0.01$	43	44	46	32	85
$0.01 \leqq P$	1,448	1,437	1,443	1,460	1,393

[a] Data are presented as number of SNPs
[b] SNP at *TFAP2B* locus

Table 3 Association of landmark SNP in *TFAP2B* (intron 2 + 58 G/C) during genome-wide screening[a]

	GG: n (%)	GC: n (%)	CC: n (%)	G	C
Case 1	73 (44)	60 (37)	31 (19)	0.63	0.37
Case 2	267 (42)	285 (45)	79 (13)	0.65	0.35
Case 3	121 (41)	132 (45)	41 (14)	0.64	0.36
Control 1	237 (55)	148 (34)	48 (11)	0.72	0.28
Control 2	423 (48)	373 (42)	93 (10)	0.69	0.31
Control 3	285 (48)	251 (42)	61 (10)	0.69	0.31

P values

		Genotype (2×3)	Allele	GG versus others	CC versus others
First test	Case 1 versus control 1	0.02	0.003	0.03	0.01
Second test	Case 2 versus control 1	0.0002	0.0007	0.00007	0.48
Third test	Case 2 versus control 2	0.1	0.03	0.04	0.21
Fourth test	Case 3 versus control 3	0.09	0.03	0.06	0.1
Total		0.0004	0.00005	0.0002	0.006

[a] Genotype data are presented as number of subjects

at this locus in type 2 diabetic subjects (case 2) with that in a different set of controls (control 2) (third test) and identified significant association with type 2 diabetes (GG versus GC + CC: $P = 0.04$, odds ratio = 1.21, 95% CI: 1.00–1.48, G versus C: $P = 0.03$, odds ratio = 1.18, 95% CI: 1.01–1.37).

We further examined the difference in the allele frequencies at this locus using 349 cases (case 3) and 598 controls (control 3), both of which were unused for the first three tests (fourth test) and confirmed a significant difference with the P value of 0.03.

Since the above three tests (first through third) used the overlapped materials, the P values from the three tests might not be correct. Therefore, the overall empirical type I error rate of these three tests was calculated by the simulation, as described in the Methods section. The simulation was iterated two hundred million times for each given minor allele frequency. As shown in Fig. 1, the type I error rate increased according to the increase of the minor allele frequency from 0.02 to 0.5. Under all conditions tested, the upper limits of the 95% CI of the type I error rates were lower than 1.61×10^{-5} (Fig. 1). This empirical P value obtained by the simulations was multiplied by the P value for the fourth test ($P = 0.03$) because the later test was independent from the former. The resulting P value of $< 4.67 \times 10^{-7}$ is considered to be the probability of the SNP association to pass the four tests. When we test 58,266 SNPs in this way, the probability to judge at least one of the SNPs to be significant was $P < 0.028$, as
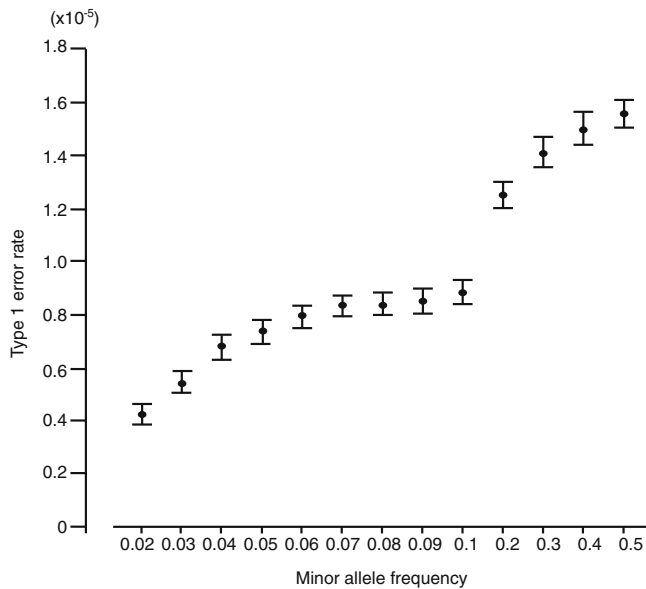
Fig. 1 Empirical type I error rate of the first through third tests. For comparison between cases and controls, these tests were simulated by the Monte-Carlo method. The details were described in the Methods section. The simulation was iterated two hundred million times for each given minor allele frequency, and the mean and 95% CI were indicated for each minor allele frequency

Table 4 Clinical characteristics of the patients

	Type 2 diabetes (case 3)[a]	Control (control 3)[a]	
Number	349	598	
Gender (M:F)	203:146	378:220	NS
Age (years)	59.0 ± 11.1	47.7 ± 8.8	$P < 0.0001$
BMI (kg/m^2)	28.4 ± 3.2	22.9 ± 2.7	$P < 0.0001$
FPG (mg/dl)	154.5 ± 43.7	92.6 ± 8.6	$P < 0.0001$
HbA1c (%)	7.3 ± 1.4	4.7 ± 0.4	$P < 0.0001$
Diabetes duration (years)	9.2 ± 7.9	–	

[a] Values are expressed as mean $\pm$ SD

calculated by Bonferroni's correction for the multiple comparisons.

Thus, the difference in the minor allele frequencies of the selected SNPs between cases and controls is significant at the type I error rate of lower than 0.05.

Subsequent LD mapping of this region using 26 SNPs around the landmark SNP in the *TFAP2B* gene revealed that the LD of this region seemed to extend to an approximately 300-kb region (200 kb upstream and 100 kb downstream to the landmark SNP). Therefore, we thought the critical region for susceptibility to type 2 diabetes lay within this 300-kb block that contained three genes (two confirmed and one predicted), in addition to *TFAP2B*. We further genotyped 188 patients for 33 additional SNP loci present within these three genes but found no significant association between any of the 33 SNPs and type 2 diabetes ($P > 0.05$, data not shown), suggesting the *TFAP2B* gene itself to be the most likely candidate for susceptibility to type 2 diabetes.

SNP discovery in TFAP2B gene and genotyping

We screened genetic polymorphisms in an entire region of the *TFAP2B* gene except repetitive sequences and found 40 additional variations, including 28 SNPs, eight insertion/deletion polymorphisms and four tandem-repeat polymorphisms although no SNP was found in the coding region of the *TFAP2B* gene. We then genotyped these polymorphisms for 349 patients [case 3, lean subjects (BMITable 4. The several variations also revealed a significant association with type 2 diabetes.

Among them, the stronger association was observed at a variable number of tandem repeat (VNTR) loci ($\chi^2 = 10.9$, $P = 0.0009$; odds ratio = 1.57, 95% CI 1.20–2.06) and two SNPs in the first intron ($\chi^2 = 11.6$, $P = 0.0006$; odds ratio = 1.60, 95% CI 1.22–2.09, and $\chi^2 = 12.2$, $P = 0.0004$; odds ratio = 1.61, 95% CI 1.23–2.11) (Table 5). We also analyzed haplotype structure using the EM algorithm and found that 12 SNPs with the allelic frequency of > 0.15 in the *TFAP2B* gene constituted one haplotype block, and five common haplotypes could cover more than 90% of the population (Fig. 2). Subsequent association study for each haplotype with type 2 diabetes identified a significant association of haplotype 4 with type 2 diabetes. However, the association of this haplotype was not stronger than found at the single locus. We also applied a step-wise logistic regression analysis to the SNPs in the block to get the subset of SNPs most significantly associated to the disease. The analysis was based on a full genotype model that included all effects of additive, dominance/recessive, and interaction between SNPs (see Methods). Applying both forward and backward selection strategies, we found that the original SNP itself (SNP at intron 1) was most significantly associated to the disease; other combinations of SNPs revealed less significance, and any additional effects of other SNPs to this original SNP were not significant. Therefore, the variations in intron1 (VNTR and SNPs) seemed to be able to explain most of the positive association of *TFAP2B* with type 2 diabetes.

We next examined the association of this gene with type 2 diabetes in a different ethnic group. As shown in Table 6, the association of this gene with type 2 diabetes could be observed also in UK population. The results indicated that the T allele of SNP at intron $1 + 774$ was shown as a risk allele, which was consistent with the result in the Japanese population although some difference in the allele frequency and in the pattern of LD within this region seemed to be present between these two populations.

Reverse transcription polymerase chain reactions

To investigate the possible biological mechanism of *TFAP2B* involvement in this disease, we examined the

Table 5 Association of polymorphisms in the *TFAP2B* gene with type 2 diabetes in the Japanese population (case 3 versus control 3)

Position (major/minor)	Major/major: *n* (%)	Major/minor: *n* (%)	Minor/minor: *n* (%)	Total
5′ flanking-512 (T/A)				
Case	111 (32)	178 (52)	53 (16)	342
Control	252 (42)	265 (44)	80 (14)	597
Intron 1 + 774 (G/T)				
Case	141 (41)	164 (48)	37 (11)	342
Control	314 (52)	224 (40)	57 (8)	595
Intron 1 VNTR (X/Y)[a]				
Case	138 (41)	159 (48)	36 (11)	333
Control	308 (52)	223 (38)	58 (10)	589
Intron 1 + 2093 (A/C)				
Case	140 (41)	164 (48)	38 (11)	342
Control	314 (52)	224 (40)	57 (8)	595
Intron 2 + 58 (G/C)[b]				
Case	121 (41)	132 (45)	41 (14)	294
Control	285 (48)	251 (42)	61 (10)	597
Intron 3 + 514 (T/C)				
Case	131 (38)	166 (48)	46 (13)	343
Control	285 (48)	251 (41)	61 (10)	597
Intron 3 + 2134 (C/T)				
Case	130 (38)	165 (48%)	47 (14)	342
Control	285 (48)	251 (42)	61 (10)	597

P values

	2×3	Allele	Major/major versus others	Minor/minor versus others
5′ flanking −512 (T/A)	0.01	0.01	0.003	0.37
Intron 1 + 774 (G/T)	0.002	0.004	0.0006	0.54
Intron 1 VNTR (X/Y)	0.005	0.008	0.0009	0.62
Intron 1 + 2093 (A/C)	0.002	0.003	0.0004	0.45
Intron 2 + 58 (G/C)	0.09	0.03	0.06	0.10
Intron 3 + 514 (T/C)	0.01	0.002	0.004	0.09
Intron 3 + 2134 (C/T)	0.01	0.003	0.004	0.1

Odds ratio (95% CI): *n* (range)

	Allele	Major/major versus	Minor/minor versus
5′ flanking −512 (T/A)	1.06 (0.89–1.27)	1.52 (1.15–2.01)	0.85 (0.61∼1.18)
Intron 1 + 774 (G/T)	1.20 (0.99–1.45)	1.60 (1.22–2.09)	1.01 (0.65∼1.56)
Intron 1 VNTR (X/Y)	1.28 (1.06–1.55)	1.57 (1.20–2.06)	1.18 (0.77∼1.81)
Intron 1 + 2093 (A/C)	1.39 (1.15–1.67)	1.61 (1.23–2.11)	1.30 (0.92∼1.84)
Intron 2 + 58 (G/C)	1.35 (1.13–1.63)	1.31 (0.98–1.73)	1.18 (0.82∼1.69)
Intron 3 + 514 (T/C)	1.41 (1.17–1.69)	1.26 (1.02–1.55)	1.34 (0.95∼1.89)
Intron 3 + 2134 (C/T)	1.23 (1.01–1.48)	1.49 (1.14–1.95)	0.98 (0.64∼1.49)

[a] X: 10 or 8 repeats, Y: 9 repeats
[b] Landmark SNP used for the genome-wide screening

expression pattern of this gene by RT-PCR using RNAs from various human tissues and found the pattern similar to that reported previously (Moser et al. 1995). However, we identified a high level of *TFAP2B* expression in the adipose tissue that had not been examined in the previous studies (Fig. 3a, b). Furthermore, we interestingly found that expression of m *TFAP2B* increased in mouse 3T3-L1 cells according to the degree of differentiation (Fig. 3c).

Discussion

In the report presented here, we performed a genome-wide, case-control association study using gene-based SNPs and identified the *TFAP2B* gene as a candidate gene conferring susceptibility to type 2 diabetes.

The contribution of genetic factors to pathogenesis of type 2 diabetes is well accepted, but only a few genes have been implicated in playing significant roles in susceptibility to type 2 diabetes so far (Horikawa et al. 2000; Ong et al. 1999; Altshuler et al. 2000). The difficulty of identifying alleles responsible for common diseases is explained by the fact that effects of individual genes in a complex genetic and environmental background are often too small to be identified with classical approaches. Our successful results presented here, as well as the recent publication for the susceptibility genes for myocardial infarction (Ozaki et al. 2002) and rheumatoid arthritis (Suzuki et al. 2003), provide solid

Fig. 2 Analysis of haplotype structure and estimated haplotype frequencies in the *TFAP2B* gene. Thirteen common variations constituted one haplotype block. *1* 5′-flanking –512, *2* intron 1+774, *3* VNTR, **X*: 10 or 8 repeat, *Y*: 9 repeat, *4* intron 1+1697, *5* intron 1+2093, *6* intron 1+2491, *7* intron 2+58, *8* intron 2+2093, *9* intron 3+242, *10* intron 3+514, *11* intron 3+2134, *12* intron 4+528, *13* intron 6+1710. *P* values for comparing haplotype frequencies between case and control groups were generated by chi-square test using 2×2 contingency table comprised of the number of one haplotype and the sum of other haplotypes

	1 2 3 4 5 6 7 8 9 10 11 12 13	DM	Control	χ^2	*p* value
haplotype 1	T G X C A G G A A T C G C	0.356	0.392	1.3	0.25
haplotype 2	A T Y G C A C G G C T A T	0.20	0.219	0.48	0.49
haplotype 3	T G X G A A G G G T C A T	0.188	0.158	1.67	0.20
haplotype 4	A T Y G C A C G G C T G C	0.146	0.082	11.3	0.0008
haplotype 5	A G X G A A G G G T C A T	0.068	0.076	0.16	0.69
haplotype 6	T G X C A G C G G C T A T	0.029	0.033	0.09	0.76

Exon 1 2 3 4 5 6 7

Table 6 Association of SNPs in the *TFAP2B* gene with type 2 diabetes in the UK population

Position (major/minor)	Major/major: *n* (%)	Major/minor: *n* (%)	Minor/minor: *n* (%)	Total
Intron 1+774 (G/T)				
Case	442 (79.1)	104 (18.6)	13 (2.3)	559
Control	421 (82.4)	89 (17.4)	1 (0.2)	511
Intron 1+2093 (A/C)				
Case	387 (68.5)	159 (28.1)	19 (3.4)	565
Control	357 (68.0)	157 (29.9)	11 (2.1)	525
Intron 2+58 (G/C)				
Case	375 (67.8)	162 (29.3)	16 (2.9)	553
Control	336 (66.9)	154 (30.7)	12 (2.4)	502

P values	2×3	Allele	Major/major versus others	Minor/minor versus others
Intron 1+774 (G/T)	0.0073	0.039	0.17	0.002
Intron 1+2093 (A/C)	0.39	0.81	0.86	0.20
Intron 2+58 (G/C)	0.81	0.95	0.76	0.61
	Odds ratio (95% CI)			
	Allele	Major/major versus	Minor/minor versus	
Intron 1+774 (G/T)	1.35 (1.01–1.79)	1.24 (0.91–1.68)	12.1 (1.58–93.16)	

| Haplotype frequency[a]	GAG	TCC	GCC	GAC	GCG
Case	0.823	0.119	0.058	0.001	0
Control	0.820	0.090	0.080	0.006	0.002

[a] Permutation analysis (1,000 replicates) demonstrated a significant difference in haplotype frequency between case and control groups ($P = 0.007$)

evidences that a genome-wide approach using SNPs as genetic markers is a useful and powerful tool to identify genes conferring susceptibility to common diseases, such as DM.

In a large-scale, genome-wide association study like our present one, the type 1 error should be minimized. The design of this study to select a particular SNP associated with diabetes was complicated because the same controls or cases were used in the first (case 1 versus control 1), the second (case 2 versus control 1), and the third (case 2 versus control 2) tests. Hence, the ordinary methods for the statistical tests could not be applied. Since these three tests were not independent of each other, one could easily overestimate the significance. To estimate the probability of an SNP passing the three tests under the conditions used in this study, we simulated exactly the process of the three tests using the Monte-Carlo method, and the final *P* value of

Fig. 3 Expression profiles of the *TFAP2B* gene. **a** The reverse transcription polymerase chain reaction (RT-PCR) using RNAs from multiple human tissues. **b** Results of quantitative real-time PCR. **c** Expression of murine *TFAP2B* in 3T3-L1 cells measured by RT-PCR on the indicated days after induction of differentiation
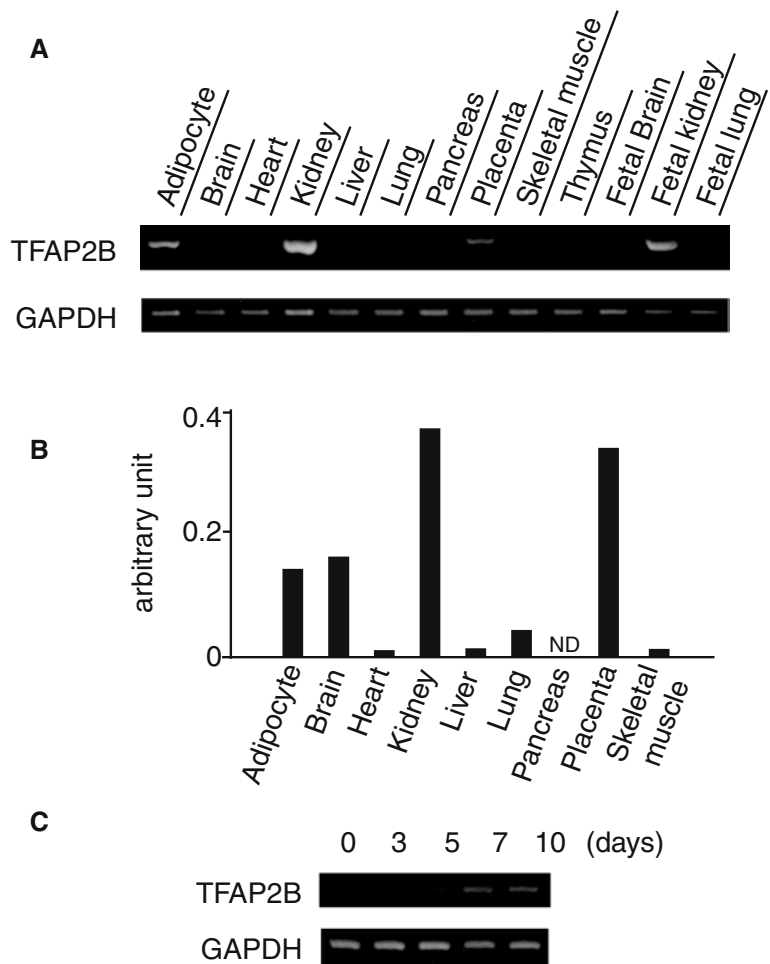
$< 4.67 \times 10^{-7}$ was obtained as the probability of the particular SNP to pass the four tests ($P < 0.028$ after Bonferroni's correction).

To evaluate the association of *TFAP2B* with type 2 diabetes further, we examined the association of this gene with type 2 diabetes in a different ethnic group. The results indicated that the SNP at the first intron of *TFAP2B* (intron $1 + 774$) was significantly associated with type 2 diabetes in the UK population also (GG 0.791, GT 0.186, TT 0.023 in type 2 DM, GG 0.824, TG 0.174, TT 0.002 in control, $P = 0.002$, Table 6). Subsequent haplotype analysis revealed a significant difference in haplotype frequency between type 2 DM and controls ($P = 0.007$, Table 6), with an increase in the TCC haplotype in the case. This result in the UK population was almost consistent with that in the Japanese population, further supporting a positive association of the *TFAP2B* gene with type 2 diabetes although there seemed to be some differences in the allele frequency and in the pattern of LD in this region between these two populations.

TFAP2B is a well-known transcription factor and has been reported to play an important role in embryonic development. In mice, expression of m *TFAP2B* decreases significantly after birth (Moser et al. 1995).

Mice lacking *TFAP2B* die within 1 or 2 days after birth from renal failure due to polycystic kidney disease (Moser et al. 1997). In humans, mutation of *TFAP2B* causes Char syndrome, a condition characterized by patent ductus arteriosus and variable degrees of facial dysmorphism and hand abnormalities (Satoda et al. 2000); those features suggest that *TFAP2B* plays an important role in the embryonic development of various tissues. However, until now, no evidence has emerged to suggest a role of *TFAP2B* in the pathogenesis of type 2 diabetes. To investigate its possible roles in this disease, we examined the expression pattern of this gene by RT-PCR using RNAs from various human tissues.

Our report is the first to show that *TFAP2B* is expressed in differentiated adipocytes that are well known as a target of insulin and as cells associated with insulin resistance. Differentiated adipocytes can function in an endocrine manner to secrete several cytokines, called "adipokines," which include TNF-α, IL-6, leptin, adiponectin, and others (Spiegelman and Flier 1996; Matsuzawa et al. 1999). These genes are found to contain binding sites for TFAP2 in their promoter (Kroeger and Abraham 1996; Isse et al. 1995; Takahashi et al. 2000). Given such observations, we suggest that *TFAP2B* plays a key role in the pathogenesis of type 2

diabetes by affecting insulin responsiveness through the transcriptional regulation of genes involved in insulin response of differentiated adipocytes.

In summary, by means of a large-scale, gene-based SNP approach, we have identified *TFAP2B* as a novel susceptibility gene for type 2 diabetes. These results suggest that *TFAP2B* itself, as well as molecules upstream or downstream of its function, might represent novel targets for treatment or prevention of this common disorder.

References

Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES (2000) The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 26:76–80

Bengra C, Mifflin TE, Khripin Y, Manunta P, Williams SM, Jose PA, Felder RA (2002) Genotyping of essential hypertension single-nucleotide polymorphisms by a homogeneous PCR method with universal energy transfer primers. Clin Chem 48:2131–2140

Cordell HJ, Clayton DG (2002) A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet 70:124–141

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229–232

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

Fajans SS, Bell GI, Polonsky KS (2001) Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. N Engl J Med 345:971–980

Groves CJ, Wiltshire S, Smedley D, Owen KR, Frayling TM, Walker M, Hitman GA, Levy JC, O'Rahilly S, Menzel S, Hattersley AT, McCarthy MI (2003) Association and haplotype analysis of the insulin-degrading enzyme (IDE) gene, a strong positional and biological candidate for type 2 diabetes susceptibility. Diabetes 52:1300–1305

Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T (2002) Gene-based SNP discovery as part of the japanese millennium genome project: identification of 190,562 genetic variations in the human genome. J Hum Genet 47:605–610

Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y (2002) JSNP: a database of common gene variations in the japanese population. Nucleic Acids Res 30:158–162

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y et al (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26:163–175

International Human Genome Sequencing Consortium (2001) Initial sequencing and the analysis of human genome. Nature 409:860–921

Isse N, Ogawa Y, Tamura N, Masuzaki H, Mori K, Okazaki T, Satoh N, Shigemoto M, Yoshimasa Y, Nishi S, Hosoda K, Inazawa J, Nakao K (1995) Structural oganization and chromosomal assignment of the human obese gene. J Biol Chem 270:27728–27733

Kadowaki T, Kadowaki H, Mori Y, Tobe K, Sakuta R, Suzuki Y, Tanabe Y, Sakura H, Awata T, Goto Y, Hayakawa T, Matsuoka K, Kawamori R, Kamada T, Horai S, Nonaka I, Hagura R, Akanuma Y, Yazaki Y (1994) A subtype of diabetes mellitus associated with a mutation of mitochondrial DNA. N Engl J Med 330:962–968

Kahn BB (1998) Type 2 diabetes: when insulin secretion fails to compensate for insulin resistance. Cell 92:593–596

Kroeger KM, Abraham LJ (1996) Identification of an AP-2 element in the −323 to −285 region of the TNF-alpha gene. Biochem Mol Biol Int 40:43–51

Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul 8:3–30

Matsuzawa Y, Funahashi T, Nakamura T (1999) Molecular mechanism of metabolic syndrome X: contribution of adipocytokines, adipocyte-derived bioactive substances. Ann NY Acad Sci 892:146–154

Moser M, Imhof A, Pscherer A, Bauer R, Amselgruber W, Sinowatz F, Hofstadter F, Schule R, Buettner R (1995) Cloning and characterization of a second AP-2 transcription factor: AP-2β. Development 121:2779–2788

Moser M, Pscherer A, Roth C, Becker J, Mucher G, Zerres K, Dixkens C, Weis J, Guay-Woodford L, Buettner R, Reinhard Fassler (1997) Enhanced apoptotic cell death of renal epithelial cells in mice lacking transcription factor AP-2β. Genes Dev 11:1938–1948

Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y (2001) A high-throughput SNP typing system for genome-wide association studies. J Hum Genet 46:471–477

Ong KKL, Phillips DI, Fall C, Poulton J, Bennett ST, Golding J, Todd JA, Dunger DB (1999) The insulin gene VNTR, type 2 diabetes and birth weight. Nat Genet 21:262–263

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-α gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654

Saito S, Iida A, Sekine A, Miura Y, Sakamoto C, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y (2001) Identification of 197 genetic variations in six human methyltransferase genes in the Japanese population. J Hum Genet 46:529–537

Saltiel AR (2001) New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. Cell 104:517–529

Satoda M, Zhao F, Diaz GA, Burn J, Goodship J, Davidson HR, Pierpont ME, Gelb BD (2000) Mutation in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. Nat Genet 25:42–46

Seki T, Tanaka T, Nakamura Y (2000) Genomic structure and multiple single-nucleotide polymorphisms (SNPs) of the thiopurine S-methyltransferase (TPMT) gene. J Hum Genet 45:299–302

Spiegelman BM, Flier JS (1996) Adipogenesis and Obesity: Rounding out the big picture. Cell 87:377–389

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

Takahashi M, Arita Y, Yamagata K, Matsukawa Y, Okutomi K, Horie M, Shimomura I, Hotta K, Kuriyama H, Kihara S, Nakamura T, Yamashita S, Funahashi T, Matsuzawa Y (2000) Genomic structure and mutations in adipose-specific gene, adiponectin. Int J Obes 24:861–868

Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M, O'Rahilly S et al (2001) A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. Am J Hum Genet 69:553–569

Yamada R, Tanaka T, Unoki M, Nagai T, Sawada T, Ohnishi Y, Tsunoda T, Yukioka M, Maeda A, Suzuki, K, Tateishi H, Ochi T, Nakamura Y, Yamamoto K (2001) Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinational effect that two genetic loci have on susceptibility to the disease. Am J Hum Genet 68:674–685

Zimmet P, Alberti KGMM, Shaw J (2001) Global and societal implications of the diabetes epidemic. Nature 414:782–786