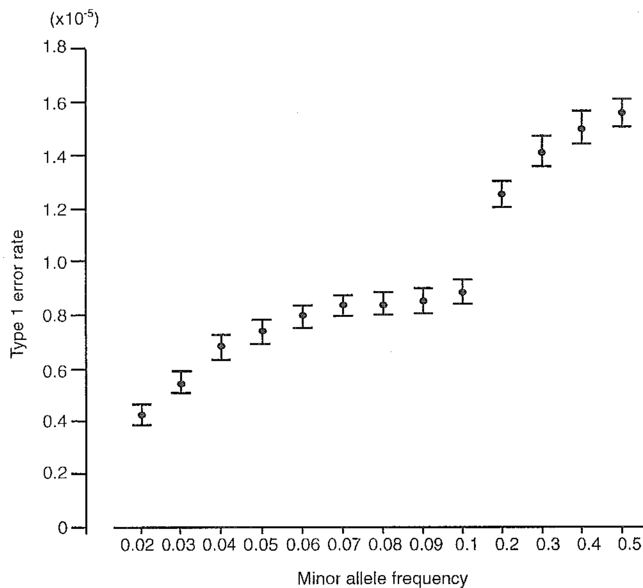

Fig. 1 Empirical type I error rate of the first through third tests. For comparison between cases and controls, these tests were simulated by the Monte-Carlo method. The details were described in the Methods section. The simulation was iterated two hundred million times for each given minor allele frequency, and the mean and 95% CI were indicated for each minor allele frequency

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.

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 ± SD

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

	1 2 3 4 5 6 7 8 9 10 11 12 13	DM	Control	χ2	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

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

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

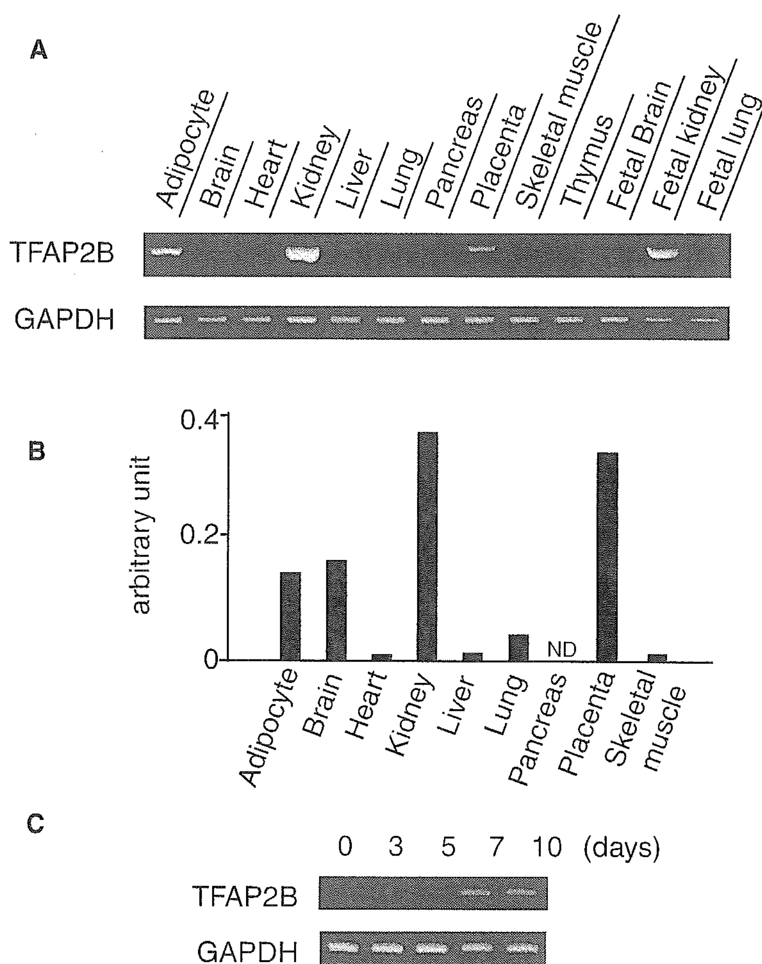

Fig. 3 Expression profiles of the *TFAP2B* gene. **a** The reverse transcription polymerase chain reaction (RT-PCR) using RNAs from multiple human tissues. **b** Results of quantitative real-time PCR. **c** Expression of murine *TFAP2B* in 3T3-L1 cells measured by RT-PCR on the indicated days after induction of differentiation

$<4.67\times10^{-7}$ 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.

Acknowledgements We thank Ms. K. Ohashi for her technical assistance and Drs. S. Saito, T. Tanaka, Y. Ohnishi, R. Yamada, S. Ikegawa, A. Takahashi, and K. Hotta for helpful discussions. This work was supported by a grant from the Japanese Millennium Project.

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 T, 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

Diabetologia (2005) 48: 1315–1322
DOI 10.1007/s00125-005-1797-0

ARTICLE

A. Kanazawa · Y. Kawamura · A. Sekine · A. Iida · T. Tsunoda · A. Kashiwagi · Y. Tanaka · T. Babazono · M. Matsuda · K. Kawai · T. Iizumi · T. Fujioka · M. Imanishi · K. Kaku · Y. Iwamoto · R. Kawamori · R. Kikkawa · Y. Nakamura · S. Maeda

Single nucleotide polymorphisms in the gene encoding Krüppel-like factor 7 are associated with type 2 diabetes

Received: 26 October 2004 / Accepted: 27 January 2005 / Published online: 4 June 2005
© Springer-Verlag 2005

Abstract *Aims/hypothesis:* Although genetic susceptibility plays an important role in the pathogenesis of type 2 diabetes, most of the genes that influence susceptibility to type 2 diabetes have yet to be identified. Krüppel-like transcription factors are known to play important roles in development and cell differentiation, and have recently been implicated in the pathogenesis of type 2 diabetes. The present study aimed to examine the associations of single nucleotide polymorphisms (SNPs) in genes encoding members of the Krüppel-like-factor (KLF) family with type 2 diabetes in a large cohort of Japanese subjects. *Methods:* We genotyped 33 SNP loci found in 12 KLF

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00125-005-1797-0.

A. Kanazawa · Y. Kawamura · S. Maeda (✉)
Laboratory for Diabetic Nephropathy,
SNP Research Centre,
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 Centre,
Institute of Physical and Chemical Research,
Kanagawa, Japan

A. Iida · Y. Nakamura
Research Group for Personalized Medicine,
SNP Research Centre,
Institute of Physical and Chemical Research,
Kanagawa, Japan

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

A. Kashiwagi · R. Kikkawa
Department of Medicine, Shiga University of Medical Science,
Shiga, Japan

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

T. Babazono · Y. Iwamoto
Diabetes Centre,
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

K. Kawai
Kawai Clinic,
Ibaragi, Japan

T. Iizumi
Department of Internal Medicine,
Toride Kyodo Hospital,
Ibaragi, Japan

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

M. Imanishi
Department of Internal Medicine,
Osaka City General Hospital,
Osaka, Japan

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

genes in subjects with type 2 diabetes and in subjects from the general population using the PCR-Invader assay. We also examined the effects of the overexpression of *KLF7* on adipogenesis in 3T3-L1 cells. *Results:* We identified a significant association between an SNP in *KLF7* and type 2 diabetes (A vs C: p=0.004 after Bonferroni's correction, odds ratio=1.59, 95% CI 1.27–2.00). The expression of *Klf7* decreased in response to the differentiation of 3T3-L1 adipocytes, and the overexpression of *KLF7* resulted in significant inhibition of adipogenesis in 3T3-L1 cells. *Conclusions/interpretation:* These results indicate that the gene encoding KLF7 is a novel candidate for conferring genetic susceptibility to type 2 diabetes.

Keywords Adipocyte differentiation · Association study · Diabetes · Gene polymorphisms · Krüppel-like factor 7

Abbreviations KLF: Krüppel-like factor · LD: linkage disequilibrium · SNP: single nucleotide polymorphism

Introduction

Type 2 diabetes affects more than 100 million individuals worldwide [1]. The pathogenesis of this disease appears to involve insulin resistance in peripheral tissues combined with dysfunction of beta cells in pancreatic islets; however, the precise mechanism remains unknown [2, 3].

It is generally accepted that genetic factors contribute to the onset and progression of diabetes, and several genes responsible for specific forms of the disease, such as MODY and mitochondrial diabetes, have been identified [4, 5]. However, genetic alterations associated with these specific forms of diabetes account for only a small percentage of cases, and genes conferring susceptibility to type 2 diabetes have not been identified in the majority of patients.

The Krüppel-like factor (KLF) family of transcription factors was initially reported to play a pivotal role in cellular development and differentiation [6, 7]. Recently, members of the KLF family have been found to be involved in adipogenesis [8] and the expression of GLUT4 in adipocytes [9]. Because of this new role in fat metabolism, the KLF genes are thought to be candidates for conferring susceptibility to type 2 diabetes. However, to date, no genetic association studies have focused on the KLF genes.

In the present study, we present the results of a case–control study of several members of the KLF gene family in Japanese subjects, and provide evidence that *KLF7* is a good candidate for conferring susceptibility to type 2 diabetes.

Subjects and methods

Subjects and DNA preparations DNA samples were obtained from 1,130 patients with type 2 diabetes who regularly attended outpatient clinics at the Shiga University of Medical Science, the Tokyo Women's Medical University, Juntendo University School of Medicine, Kawasaki Medical School, the Kawai Clinic, Toride Kyodo Hospital, Iwate Medical University School of Medicine, and Osaka City General Hospital (672 men, 458 women; age 60.1±11.6 years; duration of diabetes 13.5±10.0 years; HbA_1c 7.2±1.2%; fasting plasma glucose 8.3±2.5 mmol/l; BMI 23.2±3.4 kg/m^2 [all values are means±SD]). Diabetes was diagnosed according to the criteria of the World Health Organization. Type 2 diabetes was clinically defined as a gradual, adult onset of the disease. Subjects who tested positive for anti-glutamic acid decarboxylase antibody, and patients with mitochondrial disease (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes [MELAS]) or MODY were excluded. The control group comprised 564 members of the general population who were recruited through several medical institutes in Japan (272 men, 292 women; age 48.3±23 years, BMI 24.9±3.6 kg/m^2). We repeated the study in a second set of diabetic subjects (n=560; 349 men, 211 women; age 61.2±10.8 years; duration of diabetes 17.2±9.6 years; HbA_1c 7.9±4.2%; fasting plasma glucose 8.4±2.9 mmol/l; BMI 24.6±3.7 kg/m^2) and control subjects (n=359; 193 men 166 women; age 42.7±9.6 years). Written informed consent was obtained from each patient, and DNA extraction was performed using a standard phenol–chloroform procedure.

Genotyping The single nucleotide polymorphisms (SNPs) that were used for genotyping experiments were selected from the IMS-JST Japanese database of SNPs (http://snp.ims.u-tokyo.ac.jp, last accessed in April 2005) [10, 11]. The genotype at each SNP locus was analysed using the Invader assay (Third Wave Technologies, Madison, WI, USA) or Taqman assay (Applied Biosystems, Foster City, CA, USA), as previously described [12]. The success rates of these assays were >95%, and there was almost 100% agreement between the results of genotyping and the results of direct sequencing [12]. Differences in the genotype and/or allele frequencies between the case group and the control group were analysed using 2×3 or 2×2 contingency tables. The protocol was approved by the Ethics Committee of the Institute of Physical and Chemical Research (Kanagawa, Japan).

Northern blot analysis Mouse 3T3-L1 cells were grown to confluence and induced to differentiate into adipocytes using methods described previously [13]. Total RNA was extracted from the 3T3-L1 cells by Trizol reagent (Invitrogen, Carlsbad, CA, USA) at 0, 3 and 5 days after inducing adipocyte differentiation. Total RNA (20 μg) was separated by 1% denaturing formaldehyde agarose gel electrophoresis, and transferred onto a Hybond-N membrane (Amersham Biosciences, Piscataway, NJ, USA). Hybridisation was carried out using α-^{32}P-labelled complementary DNA (cDNA) probes for mouse *Klf7* and *Gapd*. Mouse cDNAs for *Klf7* and *Gapd* were cloned by RT-PCR using the following primers: *Klf7* sense, 5′-CCA CTG GAT GTC TGC CTT CT-3′, antisense, 5′-CAC TCG CAT CCT TCC CAT GA-3′; *Gapd* sense, 5′-CTT CTG

Table 1 Association between the SNPs within the genes encoding the KLF family of transcription factors and type 2 diabetes

dbSNP ID	Gene	Case[a]	Control[a]	*p* value	dbSNP ID	Gene	Case[a]	Control[a]	*p* value
rs2072596	*KLF1*	0.06	0.06	0.90	rs2296859	*KLF12*	0.10	0.10	0.99
rs448888	*KLF2*	0.02	0.03	0.28	rs2296889	*KLF12*	0.27	0.26	0.66
rs410867	*KLF2*	0.02	0.03	0.28	rs2274086	*KLF12*	0.22	0.23	0.71
rs2242189	*KLF3*	0.26	0.26	0.89	rs3764109	*KLF12*	0.08	0.08	0.94
rs3782933	*KLF5*	0.36	0.39	0.29	rs1426	*KLF12*	0.17	0.22	0.06
rs17731	*KLF6*	0.45	0.45	0.84	rs3764133	*KLF12*	0.48	0.46	0.56
rs3750860	*KLF6*	0.15	0.16	0.60	rs3764134	*KLF12*	0.18	0.18	0.87
rs3750861	*KLF6*	0.03	0.05	0.17	rs2241779	*KLF13*	0.47	0.43	0.19
rs3793919	*KLF6*	0.05	0.06	0.38	rs3800561	*KLF14*	0.02	0.03	0.19
rs3829201	*KLF6*	0.07	0.09	0.35	rs3800562	*KLF14*	0.35	0.32	0.22
rs2269071	*KLF7*	0.10	0.06	0.047	rs3800563	*KLF14*	0.26	0.24	0.53
rs2244696	*KLF7*	0.36	0.36	0.98	rs3800564	*KLF14*	0.25	0.27	0.27
rs2284927	*KLF7*	0.09	0.07	0.14	rs3807137	*KLF14*	0.03	0.03	0.57
rs2302870	*KLF7*	0.07	0.13	0.003	rs3807139	*KLF14*	0.20	0.22	0.32
rs3791997	*KLF7*	0.24	0.23	0.61	rs3746038	*KLF16*	0.45	0.50	0.11
rs3747284	*KLF8*	0.26	0.23	0.24	rs1054972	*KLF16*	0.45	0.48	0.33
rs3780603	*KLF9*	0.15	0.16	0.45					

[a]Minor allele frequencies are presented

Table 2 Association between SNPs within *KLF7* and type 2 diabetes

Position (major/minor) [dbSNP ID]		Major/major	Major/minor	Minor/minor	Total
Intron 1 +7249 (G/C)	Case	795 (74%)	255 (24%)	28 (2%)	1,076
[rs3791996]	Control	391 (70%)	161 (29%)	7 (1%)	559
Intron 2 +7257 (A/C)	Case	602 (56%)	406 (38%)	58 (5%)	1,066
[rs2244089]	Control	340 (62%)	184 (34%)	24 (4%)	548
Intron 2 +35092 (A/C)	Case	888 (83%)	174 (16%)	8 (1%)	1,070
[rs2302870]	Control	419 (75%)	130 (23%)	10 (2%)	559
Intron 3 +5706 (T/G)	Case	785 (74%)	259 (24%)	19 (2%)	1,063
[rs2287504]	Control	381 (69%)	154 (28%)	18 (3%)	553
Intron 3 +5889 (C/G)	Case	641 (60%)	359 (33%)	61 (6%)	1,061
[rs1263596]	Control	355 (63%)	189 (34%)	19 (3%)	563

		Genotype (2×3)	Allele	Major/major vs others	Minor/minor vs others
Intron 1 +7249 (G/C)	χ^2	7.0	1.1	2.9	2.5
	p value	0.03	0.28	0.09	0.11
Intron 2 +7257 (A/C)	χ^2	4.7	4.4	4.8	0.8
	p value	0.09	0.03	0.03	0.36
Intron 2 +35092 (A/C)	χ^2	16.2	16.2	14.9	3.6
	p value	0.0003	0.000057[a]	0.0001	0.06
Intron 3 +5706 (T/G)	χ^2	6.4	5.9	3.5	4.4
	p value	0.04	0.02	0.04	0.06
Intron 3 +5889 (C/G)	χ^2	4.6	2.7	1.1	4.4
	p value	0.08	0.05	0.3	0.04

Odds ratio (95% CI)	Allele	Major/major vs others	Minor/minor vs others
Intron 1 +7249 (G/C)	1.10 (0.90–1.35)	0.83 (0.66–1.04)	2.10 (0.91–4.84)
Intron 2 +7257 (A/C)	1.21 (1.01–1.44)	1.26 (1.02–1.56)	1.26 (0.77–2.05)
Intron 2 +35092 (A/C)	1.59 (1.27–2.00)	1.63 (1.27–2.09)	2.42 (0.95–6.16)
Intron 3 +5706 (T/G)	1.28 (1.05–1.56)	1.27 (1.02–1.60)	1.85 (0.96–3.55)
Intron 3 +5889 (C/G)	1.20 (1.01–1.44)	1.12 (0.91–1.38)	1.75 (1.03–2.95)

Genotype data are presented as number of subjects; percentages of totals are shown in brackets
[a]p=0.004 after Bonferroni's correction

CCG ATG CCC CCA T-3′, antisense, 5′- GCC TGC TTC ACC ACC TTC TTG AT-3′.

Construction of adenovirus vector and viral infection An adenovirus vector encoding human *KLF7* was prepared using the Adenovirus Expression Vector Kit (TaKaRa, Japan). Forty-eight hours before the induction of differentiation, 3T3-L1 cells were transduced with 100 plaque-forming units per cell for 12 h, and samples were obtained at the indicated days after inducing differentiation.

Oil red O staining At the indicated days after inducing differentiation, the cells were washed twice with PBS, and then fixed for 2 h with 3.7% formaldehyde. Fixed cells were incubated with oil red O for 15 min at room temperature. After washing the cells four times with water, the stained lipid droplets in the cells were visualised by light microscopy. For quantification, the dye was extracted with isopropyl alcohol, and the absorbance was measured at 540 nm.

Statistical analysis Statistical methods for determining the genotype–phenotype association and the Hardy–Weinberg equilibrium [14], and to calculate the linkage disequilibrium (LD) coefficients (D' and Δ) have been described previously [15]. Differences between the case and control groups in terms of genotype distribution or allele frequency were analysed by the chi square test using 2×3 or 2×2 contingency tables, and p values below 0.05 were considered significant.

Results

We first genotyped 33 SNP loci within 12 KLF genes in 188 Japanese patients with type 2 diabetes and compared the allelic frequencies at these loci with those in the general population (n=564). Of the 33 SNP loci tested, only two exhibited a p value of less than 0.05 (Table 1). Both loci were found in *KLF7* on chromosome 2q32. We then analysed 36 SNP loci within *KLF7* using the maximum

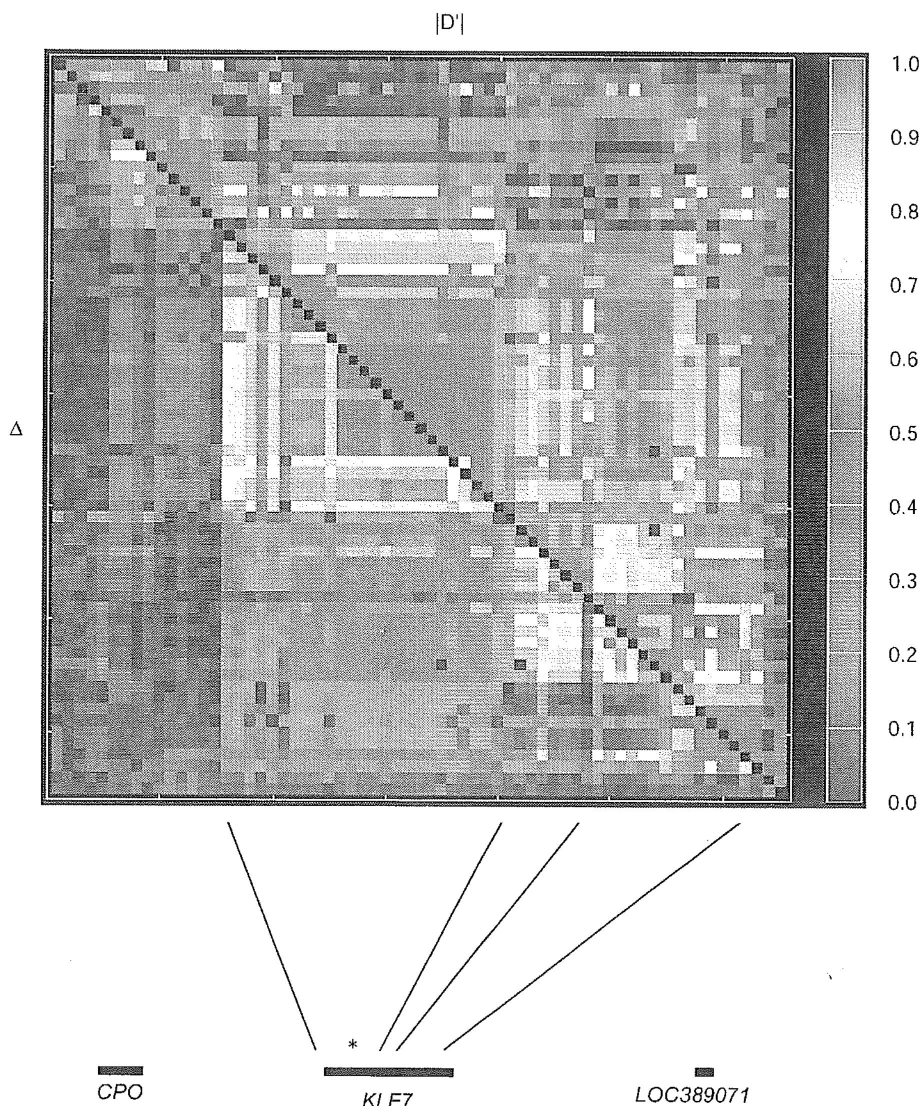

Fig. 1 LD mapping of the area around *KLF7*. LD coefficients (D' and Δ) between every two SNPs were calculated. The *asterisk* indicates the SNP at intron 2 +35092 (A/C). *CPO* Carboxypeptidase O

number of patients (n=1,130; see Table 1 of the Electronic supplementary material for details of these 36 SNPs). Several of these SNPs were found to be significantly associated with type 2 diabetes. In particular, we found that one of the alleles at an SNP locus in the second intron of *KLF7* was strongly associated with type 2 diabetes (A vs C: p=0.004 after Bonferroni's correction, odds ratio=1.59, 95% CI 1.27–2.00; Table 2). Because sex appeared to be distributed differently between the case and control groups, we subdivided the case group into male and female groups, and compared the genotype distribution for these two groups. The results indicated that the *KLF7* genotype was not distributed differently between the male and female group (male: AA=82.8%, AC=16.5%, CC=0.6%; female: AA=83.2%, AC=15.9%, CC=0.9%), and a similar association between *KLF7* and type 2 diabetes was observed in the two groups.

To exclude the possibility that the result obtained reflected associations between type 2 diabetes and other genes located near *KLF7*, we performed LD mapping of this region. This process identified a region of approximately 80 kb (20 kb upstream and 60 kb downstream of the landmark SNP; Fig. 1). We therefore concluded that the critical region for susceptibility to type 2 diabetes must lie within this 80-kb region, which did not contain any genes other than *KLF7*, indicating that *KLF7* is the candidate gene.

We next examined the haplotype structure of the region surrounding *KLF7*. The analysis was performed by estimating haplotype phasing using the Expectation Maximisation (EM) algorithm [16] and by constructing haplotype blocks as previously described [17, 18]. The results indicated that 16 SNPs in *KLF7* constitute one haplotype block with an allelic frequency of greater than 0.05, and that four common haplotypes define more than 90% of the population (Fig. 2). Subsequent examinations of the association of each haplotype with type 2 diabetes identified a significant relationship between haplotype 3 and type 2 diabetes; however, this association was not stronger than that found between the single locus (intron 2 +35092) and type 2 diabetes. Thus, the SNP in intron 2 (+35092, A/C) appeared to be able to explain the majority of the positive association between *KLF7* and type 2 diabetes. We confirmed the association of this SNP with type 2 diabetes in an independent set of case and control subjects (Table 3).

To establish whether this gene is involved in the pathogenesis of type 2 diabetes, we examined the pattern of expression of *Klf7* in mouse 3T3-L1 cells at different stages of differentiation. As shown in Fig. 3, the expression of *Klf7* was most abundant before inducing differentiation, and rapidly decreased after inducing differentiation. Furthermore, the overexpression of *KLF7* in preadipocytes resulted in the remarkable inhibition of adipogenesis in 3T3-L1 cells (Fig. 3b–d).

Discussion

Although the contribution of genetic factors to the pathogenesis of type 2 diabetes is widely acknowledged, to date, only a few genes have been implicated in susceptibility to type 2 diabetes [19–21]. The difficulty involved in identifying alleles responsible for common diseases can be explained by the fact that the effects of individual genes in a complex genetic and environmental background are often too small to be identified. Therefore, a large-scale geno-

	*	Diabetic patients	Control subjects	χ2	*p* value	OR	95% CI
Haplotype 1	A A G T C G C C A A G T A A A C	0.609	0.610	0.003	0.958		
Haplotype 2	C G A A G T G T C G A C G G A T	0.191	0.159	2.537	0.111		
Haplotype 3	A G A A C T G T C G A T G G C C	0.070	0.105	6.893	0.009	1.57	1.12–2.21
Haplotype 4	A G A T G T G T C G A T G G A C	0.032	0.030	0.027	0.870		
Haplotype 5	C G A A G T G T C G A C G G A C	0.021	0.019	0.036	0.850		
Haplotype 6	A G A T C G C C C A G T A A A C	0.015	0.017	0.067	0.796		
Haplotype 7	C G A A G T G T C G A T G G C C	0.014	0.018	0.249	0.618		
Haplotype 8	C A G T C G C C A A G T A A A C	0.008	0.003	2.097	0.148		
Haplotype 9	A G A T C T G T C G A T A A A C	0.008	0.006	0.062	0.804		
Haplotype 10	C G A A G T G T C G A T G G A T	0.006	0.002	1.308	0.253		
Haplotype 11	A G A A C T G T C G A T G G A T	0.003	0.003	0.065	0.799		
Haplotype 12	A A G T C G C C A A G T G G C C	0.003	0.000	1.318	0.251		
Haplotype 13	A G A A G T G T C G A C G G A T	0.002	0.004	0.963	0.326		
Haplotype 14	A G G T C G C C A A G T A A A C	0.002	0.000	0.878	0.349		

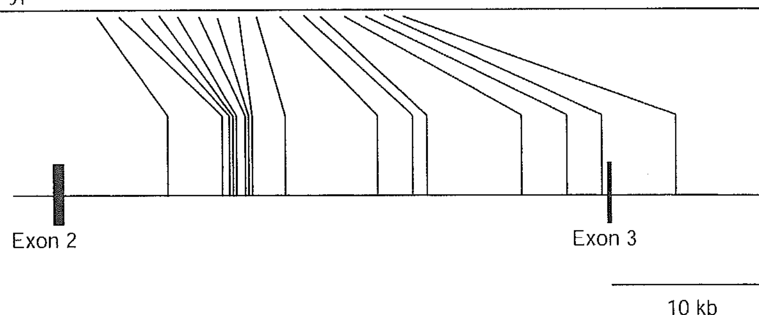

Fig. 2 Analysis of the haplotype structure and the estimated haplotype frequencies in *KLF7*. Sixteen SNPs, including the SNP at intron 2 +35092, constituted one haplotype block. The *asterisk* indicates the SNP at intron 2 +35092 (A/C). The *p* values for the comparison of haplotype frequencies between the case and control groups were generated by the chi square test using a 2×2 contingency table composed of the number of one haplotype and the sum of the other haplotypes

Table 3 Association between the SNP within the second intron of *KLF7* (intron 2 +35092 A/C) and type 2 diabetes in a second case–control group

	AA	AC	CC	A	C
Case	463 (82.7%)	94 (16.8%)	3 (0.5%)	0.91	0.09
Control	279 (77.7%)	74 (20.6%)	6 (1.7%)	0.88	0.12

	Genotype (2×3)	Allele	AA vs others	CC vs others
χ^2	5.3	4.5	3.5	2.9
p value	0.07	0.03	0.06	0.09
Odds ratio	–	1.39	1.37	3.16
95% CI	–	1.02–1.88	0.98–1.91	0.78–12.7

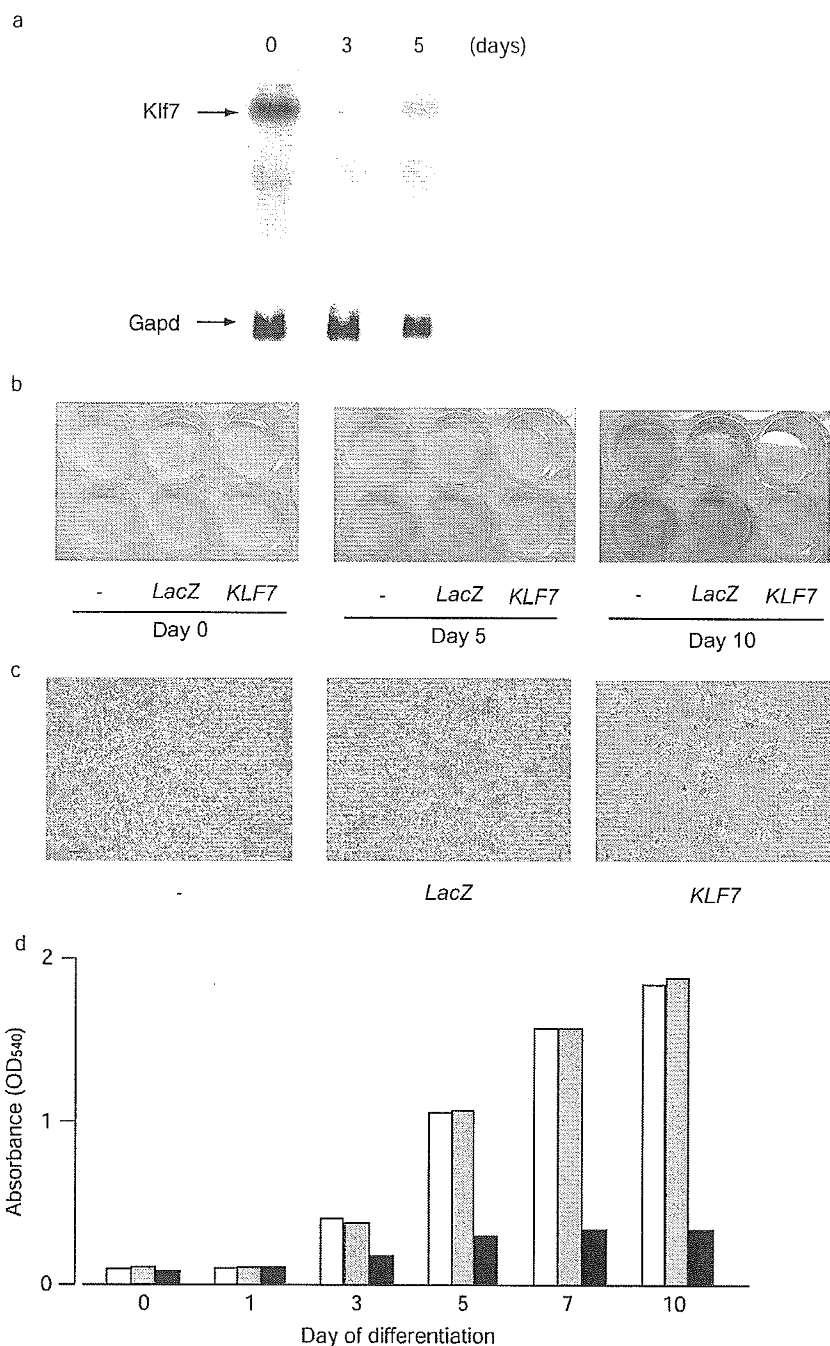

Fig. 3 **a** Northern blot analysis of mouse *Klf7* using total RNA isolated from 3T3-L1 cells at 0, 3 and 5 days after the induction of differentiation. One transcript (8.3 kb, indicated by an arrow) was identified. The same blot was re-hybridised with *Gapd* used as a loading control. The blot shown is representative of the results obtained from three independent experiments. **b** Effect of *KLF7* overexpression on adipogenesis. Cells were induced to differentiate into adipocytes 48 h after adenovirus infection. Oil red O staining was performed at the indicated days after inducing differentiation. – Uninfected cells; *LacZ* cells infected with adenovirus vectors encoding *LacZ* (control); *KLF7* cells infected with adenovirus vectors encoding *KLF7*.**c** Oil red O staining of the cells at 10 days after the induction of differentiation (magnification ×100). The results shown are representative of two independent experiments. **d** Quantification of oil red O staining in uninfected cells (*white bars*), *LacZ*-infected cells (*grey bars*) and *KLF7*-infected cells (*black bars*)

typing strategy, as employed in this study, can be more effective than a single candidate gene approach for the identification of susceptibility genes for common diseases such as diabetes mellitus. Because we examined total of 64 SNP loci, type 1 error due to multiple testing must be taken into account. We therefore corrected the p value for the association between a particular SNP in *KLF7* (intron 2 +35092) and type 2 diabetes by Bonferroni's method, and obtained a p value of approximately 0.004. Based on this, we concluded that the association between this SNP and type 2 diabetes was statistically significant. Because we used individuals from the general population as a control group, it is possible that the potential power of our study to detect an association of a particular locus with the disease was relatively weaker than if we had used subjects with normal glucose tolerance as a control group. Therefore, we examined the relationship between the SNP in *KLF7* with type 2 diabetes in an independent set of case and control subjects to verify the association. Because the *KLF7* locus, chromosome 2q32, was also reported to be a diabetes or relevant metabolic trait locus [22], we concluded that *KLF7* was a good candidate gene for conferring susceptibility to type 2 diabetes.

KLF7 was originally identified by a degenerative PCR approach using a cDNA library from human vascular endothelial cells. Because the expression of *KLF7* was detectable in wide variety of human tissues, it was then named ubiquitous Krüppel-like factor [23]. KLF7 was also identified as a transcriptional activator and was shown to be capable of increasing the expression of $p21^{WAF1/Cip1}$, which encodes a cell cycle regulator [24]. Furthermore, the gene encoding KLF7 is known to be predominantly expressed in the brain and the spinal cord in humans [23], and in the central and peripheral nervous system during embryogenesis in mice [24]. These findings suggest that KLF7 plays a pivotal role in the regulation of neurogenesis; however, the mechanism of its involvement in the pathogenesis of type 2 diabetes is not known.

In the present paper we have demonstrated that *Klf7* expression can be detected in 3T3-L1 preadipocytes, and that expression decreases in response to the differentiation of 3T3-L1 adipocytes (Fig. 3a). Using cells infected with adenoviral vectors encoding *KLF7*, we have also demonstrated that *KLF7* expression has an inhibitory effect on adipogenesis. This pattern of expression and effect on adipogenesis is similar to that previously reported for *KLF2* [8], suggesting that *KLF7* could also play a role in the regulation of adipogenesis. Therefore, it can be hypothesised that *KLF7* may confer susceptibility to type 2 diabetes by affecting adipocyte function.

The exact mechanism by which the polymorphisms in *KLF7* affect susceptibility to type 2 diabetes remains to be elucidated. By searching for additional polymorphisms within exons and the regulatory region of *KLF7*, we identified several additional polymorphisms within the gene (one synonymous SNP within exon 2, three SNPs within the 3′ untranslated region of exon 4, and three SNPs within 2 kb upstream of the transcription initiation site). However, a subsequent association study revealed that none of these SNPs was significantly associated with type 2 diabetes (data not shown).

In conclusion, we have identified *KLF7* as a new candidate gene with specific alleles that can confer susceptibility to type 2 diabetes. Our results also suggest that *KLF7* may contribute to the pathogenesis of type 2 diabetes by regulating the function of adipocytes.

Acknowledgements The authors would like to acknowledge S. Tsukada, K. Ohashi and the technical staff at the SNP Research Centre (Institute of Physical and Chemical Research, Kanagawa, Japan). This work was supported by a grant from the Japanese Millennium Project.

References

1. Zimmet P, Alberti KGMM, Shaw J (2001) Global and societal implications of the diabetes epidemic. Nature 414:782–786
2. Kahn BB (1998) Type 2 diabetes: when insulin secretion fails to compensate for insulin resistance. Cell 92:593–596
3. Saltiel AR (2001) New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. Cell 104:517–529
4. 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
5. Kadowaki T, Kadowaki H, Mori Y et al (1994) A subtype of diabetes mellitus associated with a mutation of mitochondrial DNA. N Engl J Med 330:962–968
6. Black AR, Black JD, Azizkhan-Clifford J (2001) Sp1 and Kruppel-like factor family of transcription factors in cell growth regulation and cancer. J Cell Physiol 188:143–160
7. Bieker JJ (2001) Kruppel-like factors: three fingers in many pies. J Biol Chem 276:34355–34358
8. Banerjee SS, Feinberg MW, Watanabe M et al (2003) The Kruppel-like factor KLF2 inhibits peroxisome proliferator-activated receptor-gamma expression and adipogenesis. J Biol Chem 278:2581–2584
9. Gray S, Feinberg MW, Hull S et al (2002) The Kruppel-like factor KLF15 regulates the insulin-sensitive glucose transporter GLUT4. J Biol Chem 277:34322–34328
10. 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
11. 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. Single-nucleotide polymorphism. J Hum Genet 47:605–610
12. 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
13. Kishida K, Shimomura I, Kondo H et al (2001) Genomic structure and insulin-mediated repression of the aquaporin adipose (AQPap), adipose-specific glycerol channel. J Biol Chem 28:36251–36260
14. Nielsen DM, Ehm MG, Weir BS (1998) Detecting marker-disease association by testing for Hardy–Weinberg disequilibrium at a marker locus. Am J Hum Genet 63:1531–1540
15. Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29:311–322
16. Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

17. Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229
18. Tanaka N, Babazono T, Saito S et al (2003) Association of solute carrier family 12 (sodium/chloride) member 3 with diabetic nephropathy, identified by genome-wide analyses of single nucleotide polymorphisms. Diabetes 52:2848–2853
19. Horikawa Y, Oda N, Cox NJ et al (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26:163–175
20. Ong KK, Phillips DI, Fall C et al (1999) The insulin gene VNTR, type 2 diabetes and birth weight. Nat Genet 21:262–263
21. Deeb SS, Fajas L, Nemoto M et al (1998) A Pro12Ala substitution in PPAR2 associated with decreased receptor activity, lower body mass index, and improved insulin sensitivity. Nat Genet 20:284–287
22. Iwasaki N, Cox NJ, Wang YQ et al (2003) Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects. Diabetes 52:209–213
23. Matsumoto N, Laub F, Aldabe R et al (1998) Cloning the cDNA for a new human zinc finger protein defines a group of closely related Kruppel-like transcription factors. J Biol Chem 273:28229–28237
24. Laub F, Aldabe R, Friedrich V Jr, Ohnishi S, Yoshida T, Ramirez F (2001) Developmental expression of mouse Kruppel-like transcription factor KLF7 suggests a potential role in neurogenesis. Dev Biol 233:305–318

Genetic Variations in the Gene Encoding ELMO1 Are Associated With Susceptibility to Diabetic Nephropathy

Atsuyuki Shimazaki,[1] Yoshihiro Kawamura,[1] Akio Kanazawa,[1] Akihiro Sekine,[1] Susumu Saito,[1] Tatsuhiko Tsunoda,[1] Daisuke Koya,[2] Tetsuya Babazono,[3] Yasushi Tanaka,[4] Masafumi Matsuda,[5] Koichi Kawai,[6] Tomohiro Iizumi,[7] Masahito Imanishi,[8] Toshihiro Shinosaki,[9] Toru Yanagimoto,[9] Minoru Ikeda,[9] Shigeki Omachi,[9] Atsunori Kashiwagi,[2] Kohei Kaku,[5] Yasuhiko Iwamoto,[3] Ryuzou Kawamori,[4] Ryuichi Kikkawa,[2] Masatoshi Nakajima,[9] Yusuke Nakamura,[1,10] and Shiro Maeda[1]

To search for a gene(s) conferring susceptibility to diabetic nephropathy (DN), we genotyped over 80,000 gene-based single nucleotide polymorphisms (SNPs) in Japanese patients and identified that the engulfment and cell motility 1 gene (*ELMO1*) was a likely candidate for conferring susceptibility to DN, in view of the significant association of an SNP in this gene with the disease (intron 18+9170, GG vs. GA+AA, χ^2 = 19.9, P = 0.000008; odds ratio 2.67, 95% CI 1.71–4.16). In situ hybridization (ISH) using the kidney of normal and diabetic mice revealed that *ELMO1* expression was weakly detectable mainly in tubular and glomerular epithelial cells in normal mouse kidney and was clearly elevated in the kidney of diabetic mice. Subsequent in vitro analysis revealed that *ELMO1* expression was elevated in cells cultured under high glucose conditions (25 mmol/l) compared with cells cultured under normal glucose conditions (5.5 mmol/l). Furthermore, we identified that the expression of extracellular matrix protein genes, such as type 1 collagen and fibronectin, were increased in cells that overexpress *ELMO1*, whereas the expression of matrix metalloproteinases was decreased. These results indicate that *ELMO1* is a novel candidate gene that both confers susceptibility to DN and plays an important role in the development and progression of this disease. *Diabetes* 54:1171–1178, 2005

From the [1]SNP Research Center, The Institute of Physical and Chemical Research, Yokohama, Kanagawa, Japan; the [2]Department of Medicine, Shiga University of Medical Science, Otsu, Shiga, Japan; [3]Diabetes Center, Tokyo Women's Medical University, Tokyo, Japan; the [4]Department of Medicine, Metabolism & Endocrinology, School of Medicine, Juntendo University, Tokyo, Japan; the [5]Division of Endocrinology and Metabolism, Department of Internal Medicine, Kawasaki Medical School, Kurashiki, Okayama, Japan; the [6]Kawai Clinic, Tsukuba, Ibaragi, Japan; the [7]Department of Internal Medicine, Toride Kyodo Hospital, Toride, Ibaragi, Japan; the [8]Department of Internal Medicine, Osaka City General Hospital, Osaka, Japan; [9]Discovery Research Laboratories, Shionogi Co. Ltd., Toyonaka, Osaka, Japan; and the [10]Laboratory for Molecular Medicine, Human Genome Center, The Institute of Medical Science, University of Tokyo, Tokyo, Japan.

Address correspondence and reprint requests to Shiro Maeda, 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.

Received for publication 11 November 2004 and accepted in revised form 20 December 2004.

Additional information for this article can be found in an online appendix at http://diabetes.diabetesjournals.org.

AER, albumin excretion rate; Alb/Cr, albumin-to-creatinine ratio; DIG, digoxigenine; DN, diabetic nephropathy; ISH, in situ hybridization; LD, linkage disequilibrium; MMP, matrix metalloproteinase; SNP, single nucleotide polymorphism; SSC, sodium chloride–sodium citrate; TGF, transforming growth factor.

© 2005 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Diabetic nephropathy (DN) is the leading cause of end-stage renal disease in Western countries and Japan (1,2). The pathogenesis of DN appears to be multifactorial. Several genetic and environmental factors are likely to contribute to its development and progression, although the precise mechanisms for this contribution are unknown. Because cumulative epidemiological findings have provided evidence that genetic susceptibility plays an important role in the pathogenesis of this renal disease (3,4), there have been extensive efforts to identify the gene(s) involved in the development and progression of nephropathy, in both type 1 and type 2 diabetes (5,6), but no definitive results have yet emerged.

Single nucleotide polymorphisms (SNPs) are the most common genetic variation and are considered to be potentially useful as markers to identify genetic variants that may confer susceptibility to common etiologically complex diseases.

After developing a high-throughput system for genotyping SNPs that combines the Invader assay with multiplex PCRs (7), we carried out genome-wide association studies using SNPs to identify loci involved in susceptibility to common diseases.

In the study reported here, we performed a genome-wide SNP genotyping analysis of a large panel of Japanese patients with type 2 diabetes in an effort to identify the gene(s) conferring susceptibility to DN. Our data suggest that the engulfment and cell motility 1 gene (*ELMO1*) is likely to contribute to genetic susceptibility to DN.

RESEARCH DESIGN AND METHODS

DNA preparation and SNP genotyping. DNA samples were obtained from the peripheral blood of patients with type 2 diabetes who regularly attend outpatient clinics at Shiga University of Medical Science, Tokyo Women's Medical University, Juntendo University, Kawasaki Medical School, Iwate Medical University, Toride Kyodo Hospital, Kawai Clinic, Osaka City General Hospital, or Chiba Tokusyukai Hospital. All subjects provided informed consent before enrolling in this study, and DNA extraction was performed according to standard phenol-chloroform procedures. Diabetic patients were divided into two groups according to the following diagnostic criteria: *1*) cases of DN, i.e., patients with diabetic retinopathy as well as overt nephropathy, indicated by urinary albumin excretion rates (AERs) ≥200 μg/min or urinary albumin-to-creatinine ratios (Alb/Cr) ≥300 mg/gCr, or patients under chronic renal-replacement therapy; and *2*) control patients, or patients with diabetic retinopathy but showing no evidence of renal dysfunction, i.e., AER <20 μg/min or Alb/Cr <30 mg/gCr. The SNPs for genotyping were randomly

TABLE 1
Association of landmark SNP in *ELMO1* (intron 18+9170 A/G) during genome-wide screening

Patient groups	n	AA	AG	GG	A	G
Case 1	87	31 (35.6)	40 (46.0)	16 (18.4)	0.59	0.41
Control 1	92	44 (47.8)	45 (48.9)	3 (3.3)	0.72	0.28
Case 2	459	170 (37.0)	216 (47.0)	73 (15.9)	0.61	0.39
Control 2	242	109 (45.0)	120 (49.6)	13 (5.4)	0.70	0.30
			χ^2	P	OR	95% CI
Case 1 vs. control 1		GG vs. AG + AA	10.8	0.001	6.69	1.87–23.85
Case 2 vs. control 2		GG vs. AG + AA	16.3	0.00005	3.33	1.81–6.15

Genotype data are number of subjects (%).

selected from our gene-based Japanese SNP database (available at http://snp.ims.u-tokyo.ac.jp) (8,9). The genotype of each SNP locus was analyzed with Invader assays, as previously described (7).

Our first screening involved genotyping 94 DN patients (63 men and 31 women, age 57.9 ± 12.5 years, duration of diabetes 18.6 ± 9.7 years, HbA_{1c} 7.7 ± 1.3%, mean ± SD) and 94 control patients (37 men and 57 women, age 62.7 ± 9.9 years, duration of diabetes 16.2 ± 8.4 years, HbA_{1c} 7.4 ± 1.1%) for 81,315 SNP loci. By evaluating the statistical data using 2 × 3 or 2 × 2 contingency tables, we selected SNPs that showed significant differences in genotypic or allelic frequencies between the DN and control groups. Then we analyzed the SNPs in a larger number of subjects—466 DN and 266 control patients (DN patients: 305 men and161 women, age 59.6 ± 13.5 years, duration of diabetes 17.3 ± 10.4 years, HbA_{1c} 7.8 ± 1.1%; control patients: 125 men and 141 women, age 62.9 ± 12.0 years, duration of diabetes 14.6 ± 9.3 years, HbA_{1c} 7.1 ± 1.2%) (10). The ethics committees of the Institute of Physical and Chemical Research and each participating institution approved the study protocol.

Identification of polymorphisms in *ELMO1* and genotyping. PCR primers were designed using GenBank *ELMO1* sequence data (accession nos: AC078843, AC083861, AC007444, AC009196, AC078844, and AC007349) to amplify specific regions of *ELMO1*. Repetitive elements were excluded from the search by invoking the Repeat Masker computer program (available at http://www.repeatmasker.org) in the manner described by Bedell et al. (11). PCR experiments and DNA sequencing were carried out as previously described (12). Genotyping of each SNP in the critical region was performed with the Invader assay or, in some cases, by the TaqMan assay using maximum number of the subjects.

ISH.

Tissue preparation. Under pentobarbitol anesthesia, the 30-week-old mice were flushed with PBS through the abdominal aorta followed by perfusion

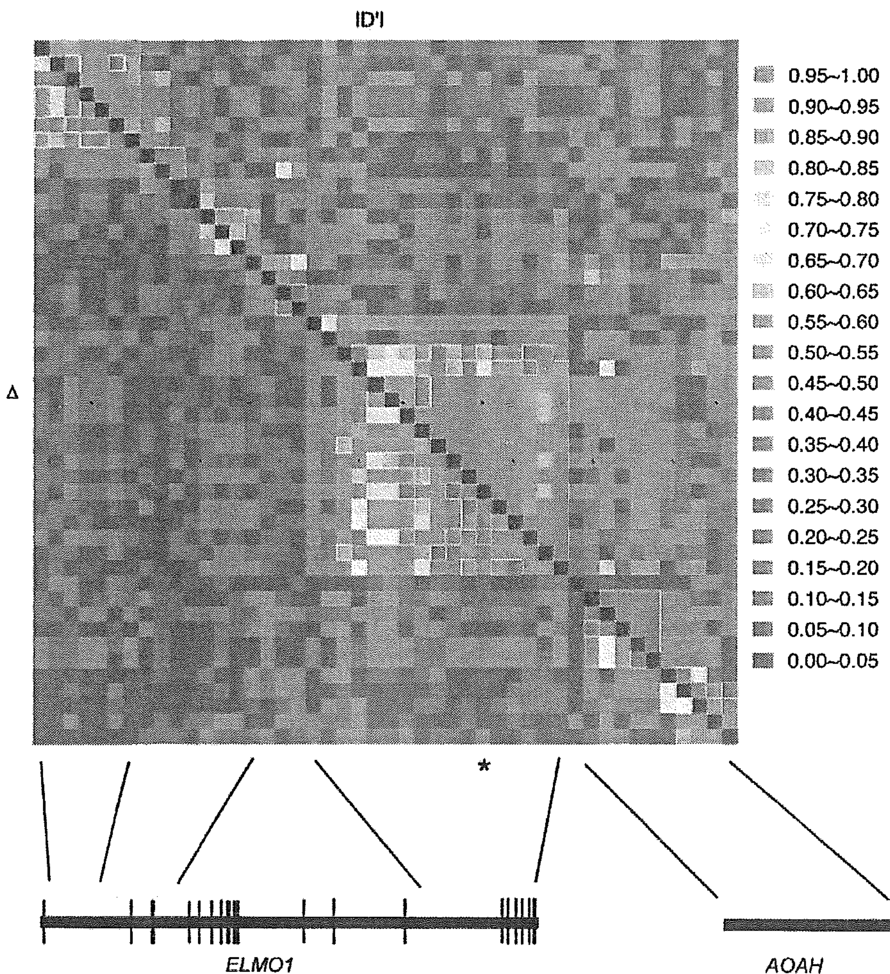

FIG. 1. LD mapping around *ELMO1*. LD coefficients (D′ and Δ) between every two SNPs were calculated. SNPs with minor allele frequencies <0.15 were not included in the calculations. An asterisk denotes the landmark SNP at intron 18+9170 (A/G).

TABLE 2
Genotype distribution of SNPs in the *ELMO1*

Position (major/minor)	Major/ major	Major/ minor	Minor/ minor	Total
Intron 16+101856 (A/G)				
Case group	326 (51.9)	252 (40.1)	50 (8.0)	628
Control group	242 (57.6)	165 (39.3)	13 (3.1)	420
Intron 16+104353 (A/G)				
Case group	296 (46.7)	269 (42.4)	69 (10.9)	634
Control group	211 (49.6)	195 (45.9)	19 (4.5)	425
Intron 16+105608 (C/T)				
Case group	355 (56.2)	237 (37.5)	40 (6.3)	632
Control group	263 (62.8)	148 (35.3)	8 (1.9)	419
Intron 16+114454 (C/T)				
Case group	229 (41.0)	262 (47.0)	67 (12.0)	558
Control group	161 (44.0)	184 (50.3)	21 (5.7)	366
Intron 17+5495 (A/G)				
Case group	286 (45.1)	278 (43.9)	69 (10.9)	633
Control group	212 (49.9)	197 (46.3)	16 (3.8)	425
Intron 18+9170 (A/G)*				
Case group	244 (37.7)	304 (47.0)	99 (15.3)	647
Control group	178 (41.8)	221 (51.9)	27 (6.3)	426
Intron 19+2430 (G/A)				
Case group	239 (37.6)	306 (48.1)	91 (14.3)	636
Control group	182 (42.9)	215 (50.7)	27 (6.4)	424
Intron 20+1239 (A/G)				
Case group	311 (48.7)	265 (41.5)	63 (9.9)	639
Control group	214 (51.0)	189 (45.0)	17 (4.0)	420
3′ flanking + 735 (A/T)				
Case group	316 (50.1)	257 (40.7)	58 (9.2)	631
Control group	221 (52.5)	183 (43.5)	17 (4.0)	421
Intron 1+5853 (T/G)				
Case group	310 (48.7)	278 (43.7)	48 (7.5)	636
Control group	195 (46.2)	174 (41.2)	53 (12.6)	422

Genotype data are presented as number of subjects (%). *Landmark SNP used for the genome-wide screening.

with 4% paraformaldehyde buffered with 0.1 mol/l PBS (pH 7.4). The kidneys were swiftly removed and cut into small pieces. The renal cortex tissue was immediately dissected and immersed into a fresh portion of the same fixative at 4°C overnight. All steps were carried out with care to avoid contamination with RNase. Diethylpyrocarbonate-treated water was used at 0.1% to prepare each buffer. Fixed samples were thoroughly rinsed with 0.1 mol/l PBS (pH 7.4), subsequently dehydrated by passage through an alcohol series, and cleared in xylene. ISH was performed on paraffin-embedded sections.

ISH study. Antisense and sense single-strand cRNAs were synthesized from cDNA fragments encoding *ELMO1* or encoding *ELMO2* using RT-PCR. The *ELMO1* cDNA fragment consisted of 525 bp (mouse sequence nucleotides 51–575, GenBank accession no. AY406934). The *ELMO2* cDNA fragment consisted of 501 bp (mouse sequence nucleotides 51–551, GenBank accession no. AF398884). Each fragment was subcloned into pCRII-TOPO vector (Invitrogen, Carlsbad, CA). The template of *ELMO1* subunit was linearized with the restriction enzyme *Hind*III, and the labeled RNA probe was synthesized with T7 RNA polymerase in the presence of digoxigenin (DIG)-labeled UTP (DIG Labeling Kit; Roche Diagnostics, Basel, Switzerland) for antisense-probe, or it was linearized with the restriction enzyme *Eco*RV, and the labeled RNA probe was synthesized with SP6 RNA polymerase for sense-probe. The template of *ELMO2* subunit was linearized with the restriction enzyme *Eco*RV, and the labeled RNA probe was synthesized with SP6 RNA polymerase for antisense-probe, or the template of *ELMO2* subunit was linearized with the restriction enzyme *Hind*III, and labeled RNA probe was synthesized with T7 RNA polymerase for sense-probe.

The probes were precipitated and DIG incorporation was assessed by dot blotting. After dewaxing, the tissue sections were dipped into 0.01 mol/l citrate buffer (pH6.0) at 95°C for 30 min, rinsed with 0.1 mol/l PBS (pH 7.4), and permeabilized with 2.5 μg/ml proteinase K (Roche Diagnostics) at 37°C for 10 min. Next, they were briefly washed with 0.1 mol/l PBS, rinsed in 0.1 mol/l triethanolamine (pH 8.0), and then acetylated with 0.1 mol/l triethanolamine containing 0.25% acetic anhydride for 10 min. Hybridization was conducted by the method described previously (13). Briefly, sections were prehybridized and then hybridized with 500 ng/ml DIG-labeled antisense or sense cRNA probe for 17 h at 50°C in a hybridization buffer containing 50% deionized formamide, 1 × Denhardt's solution, 10% dextran sulfate, 600 mmol/l NaCl, 0.025% SDS, 5 mmol/l EDTA (pH 8.0), 0.25 mg/ml yeast tRNA, and 10 mmol/l Tris-HCl (pH 7.6). During hybridization, the slides were covered with parafilm and kept in a closed, moist chamber. After hybridization, the samples were rinsed three times with 5 × sodium chloride–sodium citrate (SSC) for 10 min at 50°C. The samples were stringently washed with 2 × SSC containing 50% formamid for 20 min at 50°C and thoroughly washed twice with 0.2 × SSC for 20 min at 50°C. Then the samples were immersed in 1.5% blocking reagent dissolved in DIG buffer 1 (100 mmol/l Tris-HCl, pH 7.5, containing 150 mmol/l NaCl) for 60 min at room temperature, preincubated in normal rabbit serum at a dilution of 1:500 in DIG buffer 1 for 30 min, and subsequently incubated in anti-DIG sheep polyclonal antibodies at a dilution of 1:1,000 in DIG buffer 1 for 30 min at room temperature and rinsed with DIG buffer 1. These samples were then incubated in biotinylated anti-sheep rabbit polyclonal antibodies at a dilution of 1:1,000 in DIG buffer 1 for 30 min at room temperature and rinsed again with DIG buffer 1. The samples were next treated with avidin-biotinylated horseradish peroxidase complex solution (Vector Laboratories, Burlingame, CA) at room temperature for 60 min. After immunological incubation, the samples were extensively washed with DIG buffer 1, then processed using 0.1% 3,3′-diaminobenzidine hydrochloride substrate dissolved in 50 mmol/l Tris-HCl (pH 7.4) containing 0.05% H_2O_2 for 5 min at room temperature in the dark and examined for expression of the specific gene of interest. After a brownish color developed, the reaction was stopped by rinsing in Tris-EDTA buffer (10 mmol/l Tris-HCl, pH 7.6, containing 1 mmol/l EDTA), and the target mRNA signals were visualized. The sections were counterstained with periodic acid-Shiff reagent, dehydrated, and finally mounted in Entellan new (Merck).

Photomicrographic evaluation. The slides were examined under bright-field microscope Optiphoto-equipped Coolpix digital camera system (Nikon), with particular attention paid to evaluation of the glomerular cross views.

Real-time quantitative RT-PCR. Total RNA was prepared from COS cells cultured under normal (5.5 mmol/l) or high (25 mmol/l) glucose conditions. Quantitative real-time PCR was performed using the primers described below. For *ELMO1*: sense 5′-CCG GAT TGT GCT TGA GAA CA-3′, antisense 5′-CTC ACT AGG CAA CTC GCC CA-3′; fibronectin: sense 5′-GCT CAG AAT CCA AGC GGA GA-3′, antisense 5′-CTT TCC CAA GCA ATT TTG ATG G-3′; for collagen I (α1): sense 5′-CAC CAA TCA CCT GCG TAC AGA-3′, antisense 5′-TCA CAG ATC ACG TCA TCG CAC-3′; for transforming growth factor (TGF)-β1: sense 5′-AGG TCA CCC GCG TGC TAA T-3′, antisense 5′-GGT TCA GGT ACC GCT TC TCG -3′; for matrix metalloproteinase (MMP)-2: sense 5′-GAT GCC GCC TTT AAC TGG AG-3′, antisense 5′-CAT CTG CGA TGA GCT TGG G-3′; for MMP3: sense 5′-TTT CTC GTT GCT GCT CAT GAA-3′, antisense 5′-GAG ACA GGC GGA ACC GAG T-3′; and for GAPDH: sense 5′-GCT CAG AAT CCA AGC GGA GA-3′, antisense 5′-CTT TCC CAA GCA ATT TTG ATG G-3′. The amplifications were carried out in a 25-μl reaction volume containing 1 × EX *Taq* buffer, 200 nmol/l dNTPs, 1/20,000 SYBR Green, 800 nmol/l of each primer, 0.05 units EX *Taq*DNA polymerase, 2.75 ng TaqStart antibody (Clontech), and 5 ng of template. The thermal profile used was 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 30 s, 63°C for 30 s, and 72°C for 30 s. The amplification and quantification was performed in an ABI PRISM 7000 Sequence Detection System (PE Applied Biosystems) and normalized to GAPDH.

Statistical analysis. Statistical analyses for the association study, haplotype frequencies, and calculation of linkage disequilibrium (LD) coefficients (D′ or Δ) were described previously (14). Analysis of haplotype structure was carried out by estimating haplotype phasing using an expectation maximization algorithm (15) and by constructing haplotype blocks as previously described (10,16). For quantitative RT-PCR experiments, comparisons among three or more groups were analyzed by one-way ANOVA, followed by Scheffe's tests to evaluate statistical differences between two groups.

RESULTS

Genome-wide association study. Among the 81,315 SNP loci we tested in our original screening, 1,615 SNP loci had significant P values <0.01 between DN and control patients (see supplementary Table 1 in the online appendix available at http://diabetes.diabetesjournals.org) and were analyzed further in another larger group of patients (10). Of these SNPs, one SNP locus in the 18th intron of *ELMO1* on chromosome 7p14 was strongly associated with DN (GG vs. GA+AA, $\chi^2 = 16.3$, $P = 0.00005$; odds ratio [OR] 3.33, 95% CI 1.81~6.15) (Table 1). Correction of multiple

TABLE 3
Association of SNPs in the *ELMO1* with DN

Position	Genotype (2×3)	Allele	Major/major vs. others	Minor/minor vs. others
Intron 16+101856 (A/G)				
χ^2	11.5	7.3	3.3	10.5
P	0.003	0.007	0.07	0.001
Intron 16+104353 (A/G)				
χ^2	13.7	5.3	0.9	13.7
P	0.001	0.02	0.3	0.0002
Intron 16+105608 (C/T)				
χ^2	13.0	8.7	4.5	11.3
P	0.002	0.003	0.03	0.0008
Intron 16+114454 (C/T)				
χ^2	10.1	4.2	0.8	10.1
P	0.006	0.04	0.4	0.002
Intron 17+5495 (A/G)				
χ^2	17.7	8.4	2.3	17.5
P	0.0001	0.004	0.13	0.00002
Intron 18+9170 (A/G)*				
χ^2	19.9	9.4	1.8	19.9
P	0.00005	0.002	0.2	0.000008
Intron 19+2430 (G/A)				
χ^2	16.6	9.8	3.0	16.2
P	0.0003	0.002	0.08	0.00006
Intron 20+1239 (A/G)				
χ^2	12.3	4.0	0.5	12.3
P	0.002	0.04	0.5	0.0005
3′flanking + 735 (A/T)				
χ^2	10.1	3.6	0.6	10.1
P	0.006	0.06	0.4	0.001
Intron 1+5853 (T/G)				
χ^2	7.4	3.4	0.7	7.4
P	0.02	0.07	0.4	0.007

	OR (95% CI)		
	Allele	Major/major vs.	Minor/minor vs.
Intron 16+101856 (A/G)	1.32 (1.08–1.62)	1.26 (0.98–1.62)	2.71 (1.45–5.05)
Intron 16+104353 (A/G)	1.25 (1.03–1.51)	1.13 (0.88–1.44)	2.61 (1.55–4.41)
Intron 16+105608 (C/T)	1.38 (1.11–1.70)	1.32 (1.02–1.69)	3.47 (1.61–7.49)
Intron 16+114454 (C/T)	1.23 (1.01–1.50)	1.13 (0.86–1.47)	2.24 (1.35–3.73)
Intron 17+5495 (A/G)	1.33 (1.10–1.61)	1.21 (0.94–1.54)	3.13 (1.79–5.47)
Intron 18+9170 (A/G)*	1.32 (1.11–1.60)	1.19 (0.92–1.52)	2.67 (1.71–4.16)
Intron 19+2430 (G/A)	1.34 (1.12–1.61)	1.25 (0.97–1.60)	2.46 (1.57–3.84)
Intron 20+1239 (A/G)	1.22 (1.00–1.48)	1.10 (0.80–1.41)	2.59 (1.50–4.50)
3′flanking + 735 (A/T)	1.21 (0.99–1.47)	1.10 (0.86–1.41)	2.41 (1.38–4.19)
Intron 1+5853 (T/G)	0.83 (0.70–1.01)	0.90 (0.71–1.16)	0.57 (0.38–0.86)

*Landmark SNP used for the genome-wide screening.

testing error was made using the following calculation: Overall *P* values (*P*1st × *P*2nd) × number of tests (1st + 2nd) = 0.001 × 0.00005 × (81,315 × 4 + 1,615 × 4) = 0.017. After performing this correction, we concluded that the association of this landmark SNP with DN was statistically significant.

Subsequent mapping of LD around the landmark SNP in *ELMO1* revealed that LD in this region extended ~100 kb upstream and 100 kb downstream of the landmark site (Fig. 1). Therefore, the critical region for susceptibility to DN seemed likely to lie within this 200-kb high-LD block. Because *ELMO1* is the only gene within this LD block, it was concluded that *ELMO1* was the most likely candidate for conferring susceptibility to DN.

We next screened for additional polymorphisms in *ELMO1* and identified another 516 polymorphisms (448 SNPs, 49 insertion/deletion polymorphisms, 18 tandem repeat polymorphisms, and 1 other polymorphism) (see supplementary Fig. 1 in the online appendix). Genotyping of these variations using the maximum number of DN (n = 640) and control (n = 426) subjects showed that several of them were associated with DN (Tables 2 and 3; information on those SNPs can be found in supplementary Table 2 in the online appendix); in particular, SNP in intron 18 (+9170 A/G) had the strongest association with DN (GG vs. AG+AA, χ^2 = 19.9, P = 0.000008; OR 2.67, 95% CI 1.71–4.16). We also analyzed haplotype structure using the expectation maximization algorithm and found that nine SNPs in *ELMO1*, with an allelic frequency of >0.15, constituted one haplotype block, and that the five common haplotypes accounted for >90% of the population (Fig. 2). Subsequent association studies with DN for each haplo-

	1 2 3 4 5 6 7 8 9	case	control	χ2	*P*	odds ratio (95% CI)
Haplotype 1	T C C T A G A T G	0.297	0.322	1.4	0.24	
Haplotype 2	C C C C G A G T A	0.296	0.256	3.8	0.05	1.22 (1.00~1.49)
Haplotype 3	C T A C A G A C A	0.217	0.240	1.4	0.23	
Haplotype 4	C C C C G A A T A	0.060	0.051	0.7	0.40	
Haplotype 5	C C C C A G A T A	0.056	0.072	2.2	0.14	
Haplotype 6	C C C C G A A C A	0.019	0.006	5.9	0.015	3.14 (1.19~8.29)
Haplotype 7	C C A C A G A C A	0.011	0.009	0.2	0.63	
Haplotype 8	C T A C A G A C G	0.005	0.013	2.46	0.12	

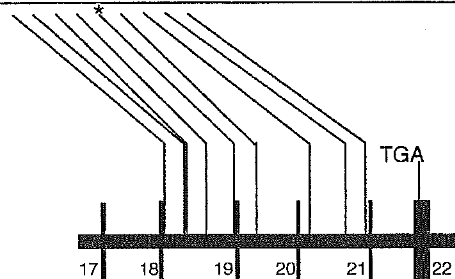

FIG. 2. Analysis of haplotype structure and estimated haplotype frequencies within *ELMO1*. Nine variations including landmark SNP constituted one haplotype block. An asterisk highlights the landmark SNP at intron 18+9170 (A/G). 1: intron18+324 (C/T); 2: intron18+2583 (C/T); 3: intron 18+2634 (C/A); 4: intron 18+5417 (C/T); 5: intron 18+9170 (A/G); 6: intron 19+2430 (G/A); 7: intron 20+1239 (A/G); 8: intron 20+5475 (T/C); 9: intron 20+8419 (A/G). *P* values for comparing haplotype frequencies between case and control groups were generated by χ^2 test using a 2 × 2 contingency table composed of the number of one haplotype and the sum of other haplotypes (e.g., haplotype 2/other haplotypes).

type identified a significant association of haplotypes 2 and 6 with DN. However, the association of these haplotypes was not stronger than that found at the +9,170 A/G locus.

ISH. To identify a possible role of *ELMO1* in the pathogenesis of DN, we examined the expression pattern of this gene in normal (*db*/+m) and diabetic (*db/db*) mouse kidney by ISH and identified that the expression of *ELMO1* could be observed mainly in glomerular epithelial cells and tubular epithelial cells (Fig. 3). The signal of *ELMO1* mRNA was remarkably increased in the kidney of diabetic mice (*db/db* 30 weeks old) (Fig. 3*E* and *F*) compared with that of controls (*db*/+m 30 weeks, Fig. 3*B* and *C*), whereas the expression of *ELMO2* was not different between these mice (Fig. 3*H*, *I*, *K*, and *L*). We also examined *ELMO1* expression by RT-PCR in various kinds of cultured cells, including glomerular mesangial cells, proximal tubular epithelial cells, COS cells, and retinal pigment epithelial cells, and could only detect definite expression in COS cells.

Real-time quantitative RT-PCR for *ELMO1*. We next examined the effects of glucose on the expression of *ELMO1* in COS cells cultured under either normal (5.5 mmol/l) or high (25 mmol/l) concentrations of glucose. As shown in Fig. 4, the expressions of *ELMO1* were significantly increased in the cells cultured under high glucose conditions compared with those in the cells under normal conditions or conditions in which osmolarity had been adjusted with mannitol (+19.5 mmol/l). These results suggested that the increase in *ELMO1* expression under high glucose conditions might contribute to the development and progression of DN.

Expression of extracellular matrix genes in *ELMO1*-overexpressing cells. From the results described above, it was suggested that the excess expression of *ELMO1* under high glucose conditions might contribute to the development and progression of DN. Although the precise mechanism of the development and progression of DN is still unknown, it has been suggested that TGF-β1–induced overaccumulation of extracellular matrix proteins is, at least in part, involved in the pathogenesis of DN (17). Therefore, we investigated the effects of excess expression of *ELMO1* on the expression of extracellular matrix protein genes, as well as on that of the TGF-β1 gene, using cells that overexpressed *ELMO1*. As shown in Fig. 5, the expression of these genes in cells stably transfected with pcDNA3.1-*ELMO1* were remarkably increased compared with those in the cells transfected with pcDNA3.1-LacZ (TGF-β1 1.36-fold vs. 1.5-fold; collagen type 1 [α1] 5.5-fold vs. 5.8-fold; fibronectin 11.1-fold vs. 12.1-fold; ELMO1-line 1 and ELMO1-line 2 vs. LacZ, respectively) (Fig. 5*B–D*), whereas the expression of MMPs were reduced (MMP2 0.66 vs. 0.27 and MMP3 0.54 vs. 0.57; ELMO1-line 1 and ELMO1-line 2 vs. LacZ, respectively) (Fig. 5*E* and *F*).

DISCUSSION

In a genome-wide case-control association study using SNPs as genetic markers, we identified *ELMO1* as a candidate gene conferring susceptibility to DN. Our data also suggest that one intronic SNP (intron 18+9170) might affect susceptibility to DN.

The results presented here, as well as in recent publications (10,18,19), provide evidence that a genome-wide case-control association study using gene-based SNPs as genetic markers is a powerful strategy for identifying genes associated with susceptibility to common diseases. Epidemiological findings had strongly suggested a contribution of genetic factors to the development and progression of DN (3,4), but worldwide efforts have so far failed to identify any solid evidence to indicate a genetic susceptibly to the disease. In many cases, the results were conflicting, probably because sample sizes were often inappropriate. However, the major cause of failures to obtain solid conclusions is possibly that multiple genetic factors are involved in DN in a complex manner and the influence of each individual factor is too weak to be identified. Therefore, approaches other than standard candidate-gene analysis or family-based linkage analysis are required to identify genes that are involved in susceptibility to common diseases such as DN. As we began this

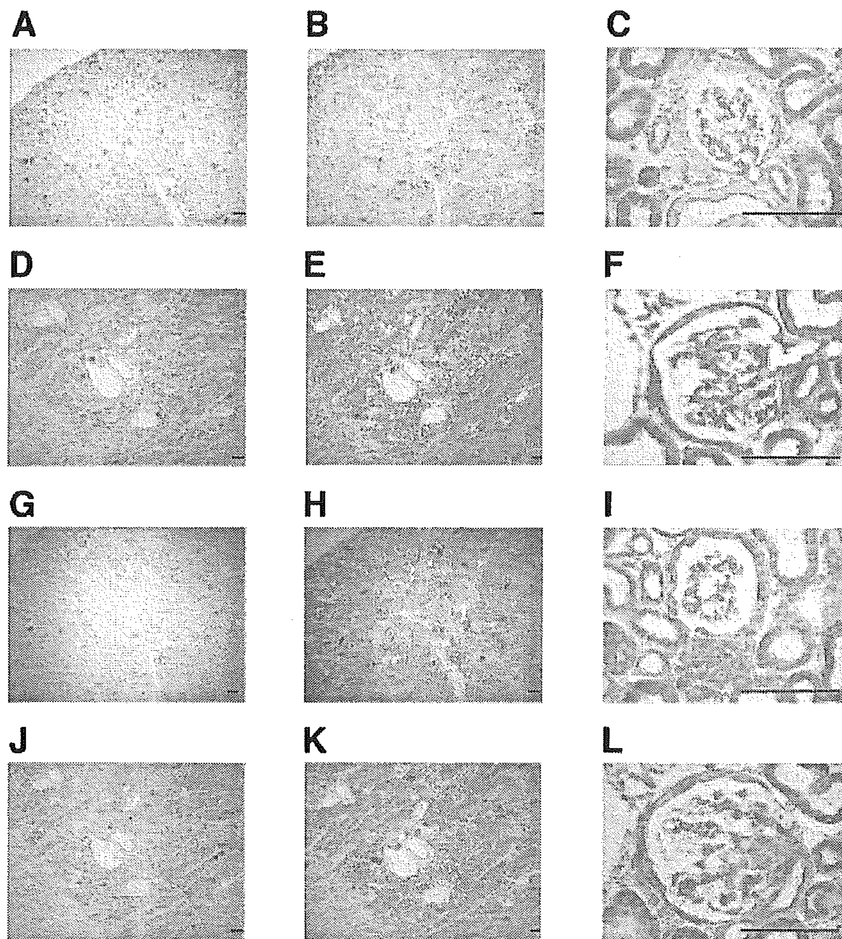

FIG. 3. Results of ISH using the kidney of 30-week-old nondiabetic (*db*/+m) and diabetic (*db/db*) mice. *A*: Sense probe for *ELMO1* (*db*/+m). *B*: Antisense probe for *ELMO1* (*db*/+m). *C*: Antisense probe for *ELMO1* (*db*/+m) *D*: Sense probe for *ELMO1* (*db/db*). *E*: Antisense probe for *ELMO1* (*db/db*). *F*: Antisense probe for *ELMO1* (*db/db*). *G*: Sense probe for *ELMO2* (*db*/+m). *H*: Antisense probe for *ELMO2* (*db*/+m). *I*: Antisense probe for *ELMO2* (*db*/+m). *J*: Sense probe for *ELMO2* (*db/db*). *K*: Antisense probe for *ELMO2* (*db/db*). *L*: Antisense probe for *ELMO2* (*db/db*). Scale bar = 100 μm.

study by analyzing a huge number of loci (81,315 SNPs) on a genome-wide basis, using a high-throughput genotyping system developed in our institute, and because all of the SNPs contributing to this report are the gene-based SNPs from the Japanese population (8,9), we were able to screen for candidate genes for disease susceptibility more efficiently than would be possible using other SNP databases.

The *ELMO1* gene, on chromosome 7p14, is a known mammalian homologue of the *C. elegans* gene, *ced-12*, which is required for engulfment of dying cells and for cell migration (20). *ELMO1* has also been reported to cooperate with CrkII and Dock180, which are homologues of *C. elegans ced-2* and *ced-5*, respectively, to promote phagocytosis and cell shape changes (20,21). However, until now no evidence has been reported to suggest a role for this gene in the pathogenesis of DN.

In this study, we showed that the expression of *ELMO1* was increased in the kidney of diabetic mice compared with that of control mice, whereas the expression of *ELMO2* was not different between diabetic and nondiabetic mice. We also identified that the increased expression of *ELMO1* could be observed in the glomeruli isolated from diabetic mice by real-time quantitative PCR (data not shown). These results suggested some role of *ELMO1* in the pathogenesis of DN.

We next examined the effects of glucose concentrations on the expression of this gene in COS cells, and we found that the expression of *ELMO1* was significantly elevated in cells under high glucose conditions compared with cells under normal glucose conditions. The effect of glucose on the expression of *ELMO1* was not due to increased osmolality, because the expression was not elevated in cells under normal glucose conditions with 19.5 mmol/l mannitol. From these observations, we speculated that increases in the expression of *ELMO1* under high glucose

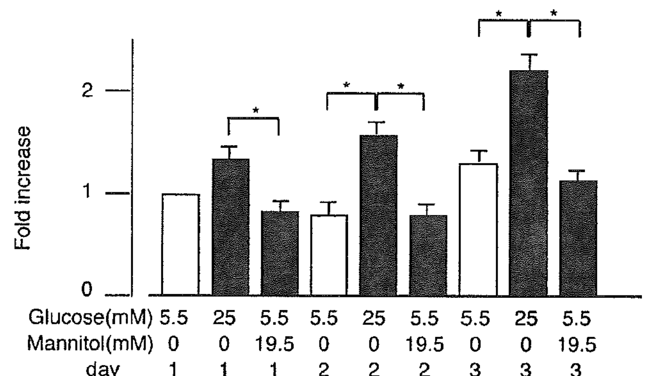

FIG. 4. Effects of glucose on the expression of *ELMO1* in COS cells. Data are presented as fold increase compared with the cells cultured under normal glucose conditions for 1 day. Results were obtained from three independent experiments. Data are means ± SD. *$P < 0.0001$.