Available online at www.sciencedirect.com
SCIENCE DIRECT
Genomics 87 (2006) 446 – 458
GENOMICS
www.elsevier.com/locate/ygeno

Association of single-nucleotide polymorphisms in the suppressor of cytokine signaling 2 (*SOCS2*) gene with type 2 diabetes in the Japanese

Hitoshi Kato [a,b,1], Kyoko Nomura [b,c], Dai Osabe [b,c], Shuichi Shinohara [b,c,d], Osamu Mizumori [b,e], Rumi Katashima [b,f], Shoji Iwasaki [a], Koichi Nishimura [a], Masayasu Yoshino [a], Masato Kobori [a], Eiichiro Ichiishi [g], Naoto Nakamura [h], Toshikazu Yoshikawa [h], Toshihito Tanahashi [i], Parvaneh Keshavarz [i], Kiyoshi Kunika [i], Maki Moritani [i], Eiji Kudo [i], Kazue Tsugawa [i], Yoichiro Takata [j], Daisuke Hamada [j], Natsuo Yasui [j], Tatsuro Miyamoto [k], Hiroshi Shiota [k], Hiroshi Inoue [i], Mitsuo Itakura [b,i,*]

[a] *Molecular Medicine Research Laboratories, Drug Discovery Research, Astellas Pharma, Inc., Tokyo, Japan*
[b] *Section for Diabetes, Genotyping Division, Genetic Diversification Analysis Project, Japan Biological Information Consortium, Tokyo, Japan*
[c] *Division of Life Science Systems, Fujitsu Nagano Systems Engineering Ltd., Nagano, Japan*
[d] *Division of R&D Solution, Fujitsu Nagano Systems Engineering Ltd., Nagano, Japan*
[e] *Toyobo Co., Ltd., Yokohama, Japan*
[f] *Applied Biosystems Japan, Tokyo, Japan*
[g] *New Industry Creation Hatchery Center, Tohoku University, Sendai, Japan*
[h] *Kyoto Prefectural University of Medicine Graduate School of Medical Sciences, Kyoto, Japan*
[i] *Division of Genetic Information, Institute for Genome Research, University of Tokushima, 3-18-15, Kuramoto-cho, Tokushima City, Tokushima 770-8503, Japan*
[j] *Department of Orthopedics, Institute of Health Biosciences, University of Tokushima, 3-18-15, Kuramoto-cho, Tokushima City, Tokushima 770-8503, Japan*
[k] *Department of Ophthalmology and Visual Neuroscience, Institute for Health Biosciences, University of Tokushima, 3-18-15, Kuramoto-cho, Tokushima City, Tokushima 770-8503, Japan*

Received 11 July 2005; accepted 19 November 2005
Available online 10 January 2006

Abstract

Several previous linkage scans in type 2 diabetes (T2D) families indicated a putative susceptibility locus on chromosome 12q15–q22, while the underlying gene for T2D has not yet been identified. We performed a region-wide association analysis on 12q15–q22, using a dense set of >500 single-nucleotide polymorphisms (SNPs), in 1492 unrelated Japanese individuals enrolled in this study. We identified an association between T2D and a haplotype block spanning 13.6 kb of genomic DNA that includes the entire *SOCS2* gene. Evolutionary-based haplotype analysis of haplotype-tagging SNPs followed by a "sliding window" haplotypic analysis indicated SNPs that mapped to the 5′ region of the *SOCS2* gene to be associated with T2D with high statistical significance. The *SOCS2* gene was expressed ubiquitously in human and murine tissues, including pancreatic β-cell lines. Adenovirus-mediated expression of the *SOCS2* gene in MIN6 cells or isolated rat islets significantly suppressed glucose-stimulated insulin secretion. Our data indicate that *SOCS2* may play a role in susceptibility to T2D in the Japanese.

Keywords: Type 2 diabetes; Association study; Japanese; Single-nucleotide polymorphism; Linkage disequilibrium; Haplotype; Human *SOCS2* gene; Glucose-stimulated insulin secretion

Type 2 diabetes (T2D) is a complex group of disorders characterized by two distinct pathophysiological defects—impaired insulin secretion from pancreatic β cells and insulin resistance in muscle, fat, and the liver [1]. While genetic factors are known to affect both of these components [2], the search for the susceptibility gene(s) has proven to be a challenge. Studies of selected candidate genes for T2D have succeeded in identifying

* Corresponding author. Division of Genetic Information, Institute for Genome Research, University of Tokushima, 3-18-15, Kuramoto-cho, Tokushima City, Tokushima 770-8503, Japan. Fax: +81 88 633 9455.
E-mail address: itakura@genome.tokushima-u.ac.jp (M. Itakura).
[1] Present address: Haplopharma, Inc., 2-2-1, Yaesu, Chuo-ku, Tokyo 104-0028, Japan.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.11.009

at least two common genetic variants accounting for a substantial proportion of common T2D: the Pro12Ala variant of the peroxisome proliferative activated receptor, gamma (*PPARG*) and the Glu23Lys variant of the potassium inwardly-rectifying channel, subfamily J, member 11 (*KCNJ11*) (for a review, see [3]). These variants may contribute significantly to the risk of T2D, conferring insulin resistance on hepatic, muscle, and fat tissues (*PPARG*, Pro12Ala) and a relative insulin secretory deficiency (*KCNJ11*, Glu23Lys). Associations between each variant and T2D were tested in different racial and ethnic groups and widely reproduced with modest estimated odds ratios (ORs) of 1.2–1.5. Such studies of candidate genes are, however, largely limited to genes coding for proteins involved in known glucose homeostasis. In contrast, genome-wide linkage scans for T2D should not presume any prior knowledge of the mechanisms underlying glucose/insulin regulation. To date, multiple studies in different racial groups have been performed and a number of possible susceptibility loci for both T2D and T2D-related traits were identified throughout the genome. However, only the *NIDDM1* locus on chromosome 2q has been mapped to a single gene, calpain 10 [4]. As in many other cases of common multifactorial diseases, phenotypic variability and polygenic inheritance in T2D seem to be adverse factors that complicate and reduce the power of linkage analysis.

The case of the hepatocyte nuclear factor 4, alpha (*HNF4A*) gene is complicated but raises many interesting possibilities. The *HNF4A* is a transcription factor important for the development, differentiation, and function of pancreatic β cells (for a review, see [5]). Initially, mutations in the *HNF4A* gene were identified as a cause of MODY1 (maturity-onset diabetes of the young, type 1), an early onset monogenic subtype of diabetes [6]. A number of T2D linkage studies, on the other hand, found evidence of linkage to 20q12–q13.1 [7], where the *HNF4A* gene is located. Taken together, these observations make *HNF4A* an excellent functional and positional candidate for T2D. Recent studies have found an association between T2D and four common single-nucleotide polymorphisms (SNPs) in the P2 promoter region, which lies 45 kb upstream of the liver-specific promoter and is the primary transcription start site in the β cell [8,9]. Of these, SNP rs2144908 demonstrated the strongest, though marginal, association with T2D, with ORs of 1.33 (95% CI 1.06–1.65) in a Finnish population and 1.46 (1.12–1.91) in an Ashkenazi Jewish population. Of note, in both studies, the risk haplotype was thought to explain a large proportion of the evidence of linkage to 20q12–q13.1.

Genome-wide SNP association studies using genotypes of numerous samples with hundreds of thousands of SNPs and the use of linkage disequilibrium (LD) and haplotype-association mapping have been proposed as a powerful and promising approach to the identification of genes with small effects, often seen in complex traits. In fact, such strategies are beginning to reveal novel and complex disease-related genes, e.g., for myocardial infarction and rheumatoid arthritis [10–12]. However, further development of cost-efficient SNP genotyping and characterization of the haplotype structure of the human genome are also needed. For some time, fine-scale LD and association mapping of susceptibility candidate loci identified by linkage studies could be an effective intermediate technique for identifying T2D genes. Such candidate loci, showing strong evidence of being replicated linkage signals for T2D, include chromosomes 12q [13], 1q [14–16], 8p [17], and 3q [18]. In this study, we performed a region-wide analysis of an association study for T2D with a dense set of >500 SNPs on a 27-Mb region of chromosome 12, using 1492 unrelated Japanese individuals. Our approach, which included identification of the susceptibility LD block followed by evolutionary-based haplotype analysis of haplotype-tagging SNPs (htSNPs), was useful in demonstrating an association between T2D and the suppressor of cytokine signaling 2 (*SOCS2*) gene.

Results

Selection of SNPs

To identify the T2D susceptibility gene on human chromosome 12q, we first examined the results of published linkage studies [16,19–22] and found that linkage peaks on chromosome 12q were quite broad and relatively confusing. At least two separate 12q regions have been identified with supportive evidence, one near the NIDDM2/MODY3 locus (in 12qter) and the other more centromeric (12q15–q22). In this study, we decided to target the 12q15–q22 region. Our target region corresponded to a 27-Mb interval between two STS markers, D12S375 and D12S362, which encompasses the 1-LOD-drop support interval of linkage signals, extending from 80 to 100 cM on chromosome 12 when adjusted for the positions of genetic markers using the Marshfield Genetic Map as a reference. The region thus included neither the linkage signals identified as the *NIDDM2*locus by Mahtani et al. [23] nor signals identified by Shaw et al. [24] in their Australian pedigree. Our target region included a total of 125 genes based on the National Center for Biotechnology Information (NCBI) Build 34 human genome assembly. We subsequently searched for SNP markers within the region with the following criteria: (1) suitability for designing optimal TaqMan high-throughput genotyping [25]; (2) mapping to the gene-centric region, which we assigned to an interval between 10 kb upstream of the transcription start site and 10 kb downstream of the final exon; (3) an average distance of 10 kb between adjacent SNPs whenever possible; and (4) common SNPs with a minor allele frequency (MAF) >0.15 based on the genotype data of 46 Japanese control individuals supplied by Applied Biosystems (unpublished data).

We initially obtained a total of 652 TaqMan genotyping probes from Applied Biosystems and genotyped an additional 111 Japanese control individuals to assess the genotyping quality (a total of 157 “genotyping and quality” controls, combining 46 and 111 individuals). After excluding SNPs with ambiguous genotyping qualities, those with significant deviations from the Hardy–Weinberg equilibrium (p value <0.05, χ^2 test) or with MAF <0.10, a total of 585 SNPs were selected as the Gene-Centric, Evenly Spaced Common SNPs and used for further association testing (a complete list of SNPs is available in Appendix A1 and a statistical summary is shown in Appendix

A2; appendices are in the supplementary material). Of these, 536 SNPs (91.6%) were mapped within 91 genes (72.8% of the total target genes), with an average distance of 12.5 kb, and 49 were in intergenic regions. The average MAF of the 585 SNPs was 0.34 ± 0.11 (SD). There were a significant number of allelic associations among the SNPs, mostly between two or three adjacent SNPs (data not shown). When SNP haplotype blocks were generated by the LD-based algorithm of Gabriel et al. [26], we identified 129 independent SNP haplotype blocks.

Association study

To reduce the time and cost of genotyping, we designed a large, modified two-stage case–control association analysis in the 1492 unrelated Japanese individuals with the 585 Gene-Centric, Evenly Spaced Common SNPs on chromosome 12q (Appendix B provides an overview of the study design).

In Stage 1 (Discovery panel), we genotyped 147 Japanese individuals with T2D at all 585 SNP loci and compared allelic frequencies with those of 157 genotyping and quality controls (see above), using a standard χ^2 test. In Stage 1, we selected SNP markers with a relatively high significance level of $\alpha < 0.10$ to increase statistical power for detecting associations. This also meant most of the SNPs selected at this stage were likely to be false positives. Thirty-seven SNPs were selected in Stage 1 (Appendix A1). We noticed that the number of SNPs selected was fewer than expected (see below), and this was probably due to the presence of marker–marker associations among SNPs. For the 37 SNPs selected in Stage 1, we genotyped an additional 259 control and 188 T2D individuals, pooled genotypic data, and regarded this expanded panel as Stage 1E (Expanded panel; a total of 416 control and 335 T2D individuals). We chose a significance level of $\alpha < 0.05$ in this pooled Stage 1E and selected 9 of 37 SNPs (SNP017, SNP152, SNP260, SNP262, SNP378, SNP380, SNP488, SNP489, and SNP514).

Next, we tested the nine SNPs selected in Stage 2, an independent replicative panel consisting of 376 T2D subjects and 365 control subjects, with an increased significance level of $\alpha < 0.01$ (Table 1). Overall, our two-stage, step-wise focusing design was expected to select less than one false-positive SNP on average (since 585 SNPs were initially tested, average numbers of false-positive markers were expected to be 59 in Stage 1 (585 SNPs at $\alpha = 0.1$), 30 in Stage 1E (585 SNPs at $\alpha = 0.05$), and 0.3 in Stage 2 (30 SNPs at $\alpha = 0.01$), without taking account of allelic associations). In Stage 2, only a single SNP, SNP488 (rs10777530), marginally reached the significance level ($p = 0.0095$, Table 1). A retrospective analysis of the power for the significant association detected revealed that the probabilities for detecting the differences between cases and controls for SNP488 were 55.5% at Stage 1 (significance level $\alpha = 0.1$), 79.6% at Stage 1E ($\alpha = 0.05$), and 58.2% at Stage 2 ($\alpha = 0.01$).

When all of the genotypic data were pooled (Stages 1E and 2, a total of 781 controls and 711 T2D individuals), the crude OR for SNP488 was 1.32 (95% CI 1.15–1.53, $p = 0.00013$). Logistic regression analysis revealed that the association between SNP488 and T2D remained statistically significant after adjustment for age and sex (data not shown). It should be mentioned that, when Bonferroni's correction for multiple analyses was applied for 585 independent tests (number of initial SNPs), SNP488 did not reach statistical significance (Bonferroni-corrected $p = 0.076$). However, because significant LD existed in our 585 SNPs, such that they were not completely "independent," Bonferroni's correction was likely to be too conservative. It is noteworthy that, in the pooled data set, two other SNPs, SNP489 (rs10745657) and SNP152 (rs7137054), were significantly associated ($p < 0.05$) with T2D (SNP489, OR 1.28, 95% CI 1.10–1.50, $p = 0.0019$; SNP152, OR 1.20, 95% CI 1.04–1.39, $p = 0.01$). We could not exclude the possibility of the associations for SNP489 and SNP152 being false positives; however, because SNP489 was close to SNP488 and these two SNPs were in strong LD ($|D'| = 1.00$), we assumed both to be associated with the same putative susceptibility variant(s).

To determine the probability that our significant associations were false positives, we applied the false positive report probability (FPRP) method that was recently proposed by

Table 1
Results of association tests for nine SNPs in Stage 2

SNP ID	Gene	RS number	NCBI position (bp)[a]	Alleles (major/minor)	MAF[b]		*p* value			Crude OR (95% CI)[c]
					Control (n = 781)	Case (n = 711)	Stage 1E	Stage 2	Pooled	
SNP017	*CPM*	rs2293637	67,551,545	C/G	0.207	0.182	0.030	0.854	0.084	NA[e]
SNP152	*TRHDE*	rs7137054	71,193,410	C/G	0.448	0.404	0.036	0.203	0.014	1.20 (1.04–1.39)
SNP260[d]	*SYT1*	rs10746105	78,213,527	C/G	0.428	0.451	0.013	0.516	0.199	NA[e]
SNP262[d]	*SYT1*	rs10746114	78,215,902	G/A	0.426	0.449	0.018	0.588	0.202	NA[e]
SNP378[d]	*PPFIA2*	rs11114926	80,491,943	A/T	0.295	0.315	0.042	0.695	0.246	NA[e]
SNP380[d]	*PPFIA2*	rs1922416	80,500,960	T/A	0.296	0.315	0.050	0.714	0.259	NA[e]
SNP488[d]	*SOCS2*	rs10777530	92,463,578	T/C	0.481	0.449	0.0053	0.0095	0.00013	1.32 (1.15–1.53)
SNP489[d]	*SOCS2*	rs10745657	92,475,969	A/G	0.267	0.319	0.025	0.036	0.0019	1.28 (1.10–1.50)
SNP514	*CRADD*	rs10859589	92,704,990	C/T	0.335	0.312	0.014	0.426	0.196	NA[e]

Association results of nine SNPs tested in Stage 2 are shown, along with the results of Stage 1E and the pooled sample (781 controls and 711 cases).

[a] SNP positions were based on NCBI Build 34 human genome sequence information.

[b] MAF, minor allele frequencies.

[c] Crude odds ratio (OR) was calculated against disease-protective allele.

[d] There were close allelic associations between SNP260 and SNP262, SNP378 and SNP380, and SNP488 and SNP489.

[e] NA, not assessed.

Wacholder and colleagues [27]. The FPRP is determined by: (1) the prior probability of a true association, (2) observed *p* value for the association, and (3) statistical power of the study. Since the current study was the hypothesis-driven candidate region scan, we assigned a low to moderate prior probability range (0.1 to 1%). To determine FPRP values, we used our experimentally determined values of OR and the corresponding 95% CI for SNP488, SNP489, and SNP152 in the total study group. For a prior probability of 0.1%, we obtained the FPRP value of 0.193 for SNP488, which satisfied a stringent FPRP value of <0.2 [27] and indicated our findings to be noteworthy (0.701 for SNP488 and 0.938 for SNP152). For a prior probability of 1%, the FPRP value for SNP488 was less than 5% (0.023). In addition, the FPRP value for SNP489 (0.188) was also satisfied at a <0.2 FPRP level (the results of FPRP analysis are summarized in Appendix C).

As a result, our association test identified two associated SNPs, SNP488 and SNP489, as putative candidate SNPs for Japanese T2D. The results are summarized in Table 1 and illustrated graphically in Fig. 1.

Determination of LD block structure around SNP488–SNP489

SNP488 and SNP489 were mapped to the 5′ and 3′ regions of the *SOCS2* gene, respectively, with a distance of 12,391 bp. There was strong LD between the two ($|D'| = 1.00$), indicating that both SNPs are in the same LD block. Since the length of the LD block may have extended beyond 12 kb, and therefore might contain a gene(s) other than *SOCS2*, and to investigate the LD structure in detail and to clarify the susceptibility region, we again searched various databases for ungenotyped SNP markers flanking SNP488 and SNP489. We identified four new SNPs (ZS003, SNP487, A356, and A357) in the public database and genotyped them in all DNA panels (Stage 1E and Stage 2), along with SNP490 used in Stage 1. LD analysis using these seven SNPs indicated that SNP488, A356, and SNP489 belonged to the same LD groups ($|D'| > 0.99$), while four other SNPs clearly lay outside the haplotype block (ZS003–SNP487 in the upstream region and A357–SNP490 in the downstream region; see Appendix D).

We next conducted resequencing analysis in the ~20-kb interval between SNP487 and SNP490 to search for unidentified SNPs and to determine the exact boundaries of a given haplotype block. For this purpose, we chose genomic DNAs from 24 control individuals for direct sequencing analysis and identified 50 new polymorphic sites (49 SNPs and 1 microsatellite, data not shown). We genotyped all DNA panels (Stage 1E and Stage 2), and the subsequent LD analysis using SNPs with MAF > 0.1 using the Haploview software [28] identified three haplotype blocks (Fig. 2). A putative T2D susceptibility LD block as defined by Gabriel's rule [26] spanned a 13,584-bp genomic region between the ZS009 and the SNP489 markers (Fig. 2,

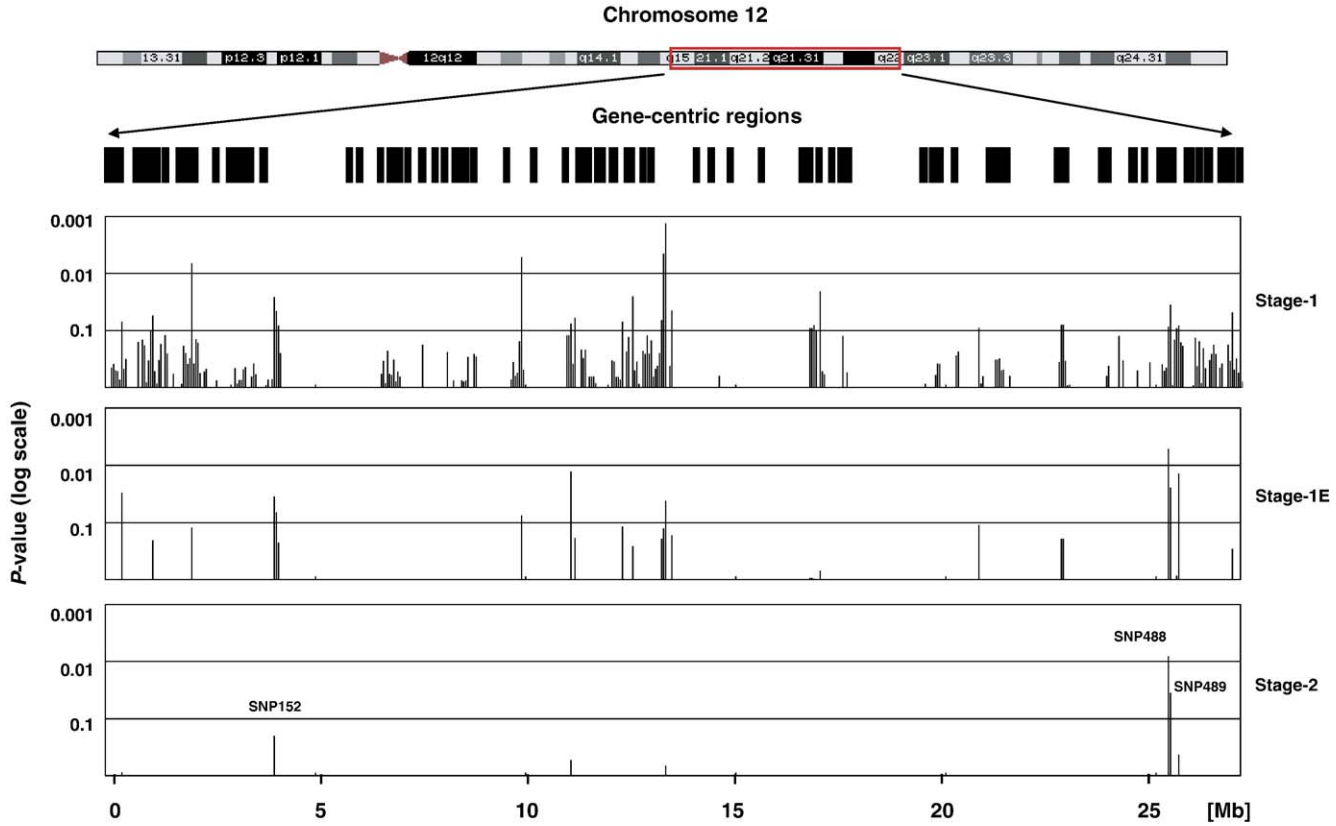


Fig. 1. Results of SNP association study on chromosome 12q. Gene-centric, evenly spaced common SNPs on chromosome 12q15–q22 are shown with *p* values for the association test. For Stages 1, 1E, and 2, 585, 37, and 9 SNPs were analyzed, respectively. The target chromosomal region for the association study is shown above. Gene-centric regions (black boxes) of 128 genes are shown in the middle. Note that SNPs were selected mainly in gene-centric regions (average SNP distance, 12.5 kb) rather than in the intergenic regions.

Block 2). There were 34 polymorphic sites within Block 2 in the Japanese population (Table 2), including one rare synonymous coding SNP in exon 3 (ZS006; Y268Y) and two SNPs in the 5′ untranslated region of the *SOCS2* gene. We also confirmed that *SOCS2* was the only gene that mapped within Block 2.

SNP haplotypes, cladistic analysis, "sliding window" analysis

Results of association tests for the 34 polymorphic sites in Block 2 are shown in Table 2. We observed significant associations with all common SNPs of MAF > 0.25 and a microsatellite MS1, but not with rare SNPs. The most significantly associated SNP was ZS011 (rs12425869, p = 0.000063), which was adjacent to and in complete LD (r^2 = 1) with SNP488.

To identify the putative susceptibility site(s) responsible for the observed association in Block 2, we applied a cladistic approach proposed by Templeton et al. [29]. This approach allows analysis of multilocus data within a genomic segment/gene by considering haplotypes and grouping them in nested clades. The idea is to search for clades with an association between haplotypes and phenotypes and to identify the mutations defining these clades as potential candidate disease susceptibility sites. For the analysis, we first estimated haplotype frequencies of 22 SNPs in Block 2 that were relatively common in terms of their frequencies (MAF > 5%) using SNPAlyze_v.3.2.1PRO software, which is based on the Expectation–Maximization (EM) algorithm. Five major haplotypes that covered more than 97% of all haplotypes were observed (data not shown). The SNPtagger program [30] then revealed that these five haplotypes (Nos. 1–5) were tagged (or identified) by four htSNPs, ZS009, ZS012, ZS004, and ZS016. We found a significant association with T2D by the haplotype association analysis using the four htSNPs and identified a common disease-protective haplotype 1 (Table 3, permutation p = 0.00051) and an at-risk haplotype 2 (permutation p = 0.00097). Finally, a cladistic-based association analysis was performed using the Evolutionary-Based Haplotype Analysis Package (EHAP)_v1.1 program [31]. Results showed that all five haplotypes were connected in a cladogram according to their evolutionary history (Table 3). In addition, the disease-protective haplotype (1) could be discriminated from other haplotype groups (2, 3, 4, and 5, score test p = 0.0001). Because a single mutational step at htSNP ZS009 was responsible for this observation, we concluded that ZS009 was most likely to be etiologically associated with the disease-susceptibility variant.

The htSNP ZS009 represents 17 polymorphic sites within Block 2 (Table 2). However, further application of a conventional fine-mapping strategy appeared to be limited

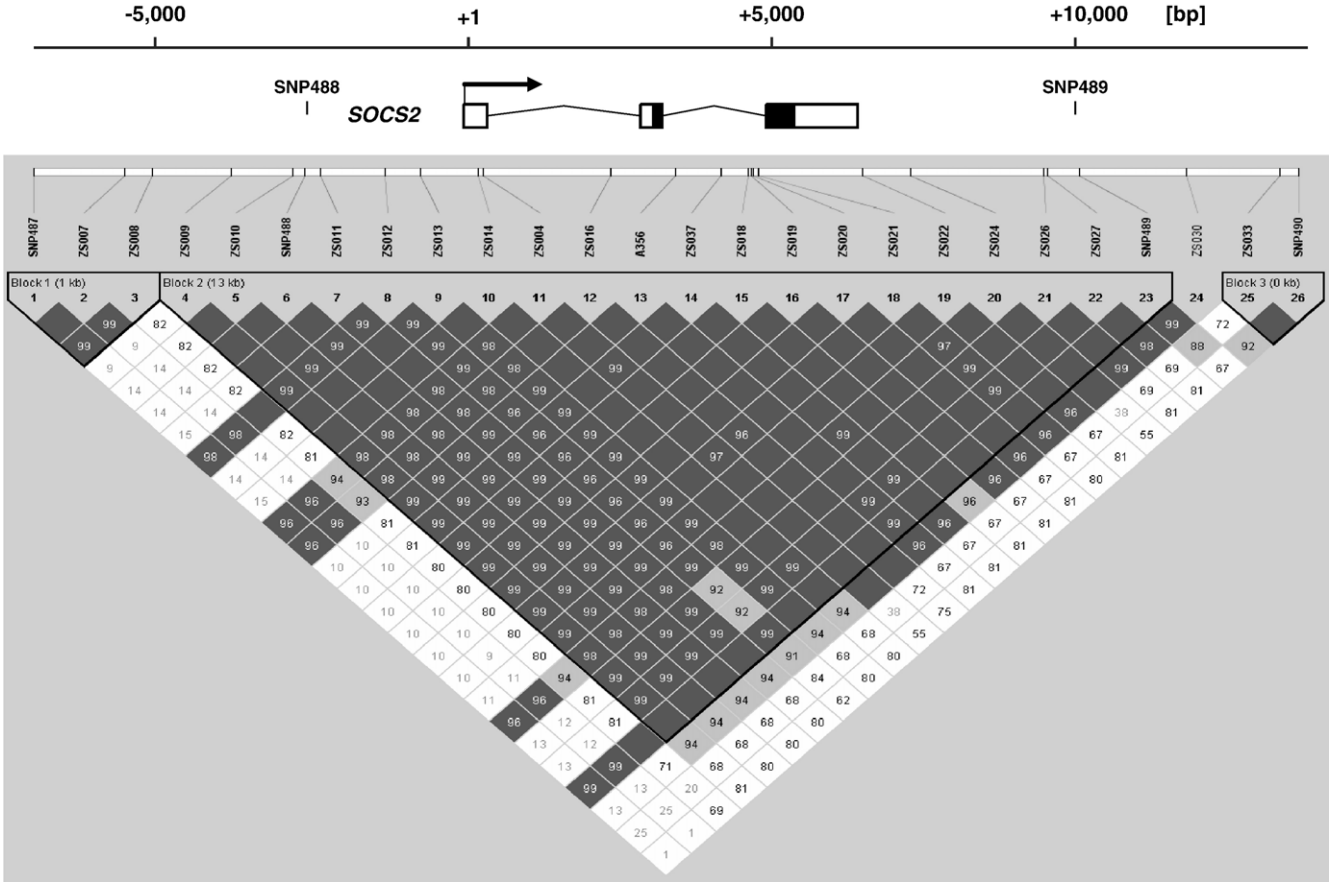


Fig. 2. Haplotype block structure encompassing the human *SOCS2* gene. Haplotype block structure following the Gabriel definition [26] is shown, along with the organization of the *SOCS2* gene (top). Exons are indicated by boxes (black, coding sequences; white, untranslated sequences). A 13,584-bp haplotype block (Block 2) spans the region between markers of ZS009 and SNP489 encompassing the entire *SOCS2* gene.

Table 2
List of 34 DNA polymorphic sites in the *SOCS2* gene (Block 2)

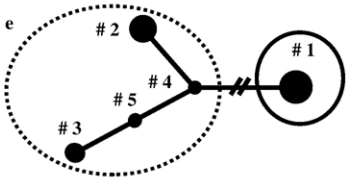
SNP ID	RS number	NCBI position (bp)[a]	Position from *SOCS2* transcription start site	Domain	Allele		Genotyping method	Frequency				*p* value	ZS009-tagged SNPs[b]
					1	2		Control		Case			
								Allele 1	Allele 2	Allele 1	Allele 2		
ZS009	rs3869308	92,462,386	−3680	5′ upstream	T	C	TaqMan	0.476	0.524	0.542	0.458	0.00032	–
ZS010	rs10859532	92,463,385	−2681	5′ upstream	G	C	TaqMan	0.481	0.519	0.552	0.448	0.00011	#
SNP488	rs10777530	92,463,578	−2488	5′ upstream	C	T	TaqMan	0.481	0.519	0.551	0.449	0.00012	#
ZS011	rs12425869	92,463,821	−2245	5′ upstream	A	G	TaqMan	0.479	0.521	0.552	0.448	0.00006	#
ZS035	–	92,464,422	−1644	5′ upstream	C	T	Sequencing	1.000	0.000	0.995	0.005	0.33	
ZS041	–	92,464,752	−1314	5′ upstream	[C]3	[C]2	Sequencing	0.996	0.004	0.996	0.004	0.94	
ZS042	–	92,464,822	−1244	5′ upstream	G	A	Sequencing	0.998	0.002	1.000	0.000	0.33	
ZS012	rs11107111	92,464,851	−1215	5′ upstream	T	G	PCR-ASP	0.723	0.277	0.669	0.331	0.00134	
ZS013	rs2053196	92,465,427	−639	5′ upstream	T	C	TaqMan	0.479	0.521	0.552	0.448	0.00007	#
ZS014	rs2200160	92,466,341	+276	Exon 1 (5′-UTR)	C	T	TaqMan	0.481	0.519	0.552	0.448	0.00010	#
ZS004	rs1498708	92,466,432	+367	Exon 1 (5′-UTR)	C	T	TaqMan	0.833	0.167	0.832	0.168	0.93	
ZS036	–	92,466,775	+710	Intron 1	T	G	Sequencing	0.983	0.017	0.968	0.032	0.35	
ZS015	–	92,467,972	+1907	Intron 1	C	A	TaqMan	0.987	0.013	0.987	0.013	0.97	
ZS016	–	92,468,459	+2394	Intron 1	[T]3	[T]2	TaqMan	0.855	0.145	0.860	0.140	0.66	
ZS005	rs11834113	92,468,686	+2621	Intron 1	T	C	TaqMan	0.034	0.966	0.042	0.958	0.25	
ZS017	–	92,469,295	+3230	Intron 2	G	C	TaqMan	0.985	0.015	0.985	0.015	0.90	
A356	rs930316	92,469,503	+3438	Intron 2	A	G	TaqMan	0.475	0.525	0.538	0.462	0.00065	#
ZS037	rs3782415	92,470,223	+4158	Intron 2	C	T	TaqMan	0.526	0.474	0.461	0.539	0.00034	#
ZS018	rs768775	92,470,665	+4600	Intron 2	T	C	TaqMan	0.474	0.526	0.542	0.458	0.00022	#
ZS019	rs1316739	92,470,724	+4659	Intron 2	G	A	TaqMan	0.474	0.526	0.541	0.459	0.00029	#
ZS020	rs1316740	92,470,756	+4691	Intron 2	C	A	TaqMan	0.474	0.526	0.539	0.461	0.00039	#
ZS021	rs3729654	92,470,822	+4757	Intron 2	insT	–	TaqMan	0.528	0.472	0.462	0.538	0.00035	#
ZS038	–	92,470,858	+4793	Intron 2	[T]12	[T]11	Sequencing	0.528	0.472	0.500	0.500	0.59	#
ZS006	–	92,471,381	+5316	Exon 3 (Y268Y)	A	G	TaqMan	0.998	0.002	1.000	0.000	0.10	
ZS022	rs2072593	92,472,494	+6429	3′ downstream	T	C	TaqMan	0.474	0.526	0.539	0.461	0.00045	#
ZS023	–	92,472,932	+6867	3′ downstream	G	A	TaqMan	0.999	0.001	0.999	0.001	0.62	
ZS024	rs3816997	92,473,254	+7189	3′ downstream	A	C	TaqMan	0.832	0.168	0.832	0.168	0.99	
ZS039	rs10859533	92,473,784	+7719	3′ downstream	[C]5	[C]6	Sequencing	0.523	0.477	0.495	0.505	0.59	#
MS1[c]	rs10594657	92,474,947	+8882	3′ downstream		$[CA]_n = 11-19$	–	–	–	–	0.00761		
ZS025	–	92,475,128	+9063	3′ downstream	G	T	TaqMan	0.987	0.013	0.992	0.008	0.13	
ZS026	rs11107114	92,475,395	+9330	3′ downstream	G	A	TaqMan	0.469	0.531	0.530	0.470	0.00088	#
ZS027	rs11107115	92,475,455	+9390	3′ downstream	A	G	TaqMan	0.470	0.530	0.531	0.469	0.00086	#
ZS028	–	92,475,502	+9437	3′ downstream	C	A	TaqMan	0.965	0.035	0.966	0.034	0.90	
SNP489	rs10745657	92,475,969	+9904	3′ downstream	A	G	TaqMan	0.733	0.267	0.681	0.319	0.00192	

[a] SNP positions were based on NCBI Build 34 human genome sequence information.
[b] Four haplotype-tagging SNPs (htSNPs) were selected with the SNPtagger program (ZS009, ZS012, ZS004, and ZS016). HtSNP ZS009 (–) represents a total of 17 SNPs (#) including ZS009.
[c] MS1 was a $(CA)_n$ microsatellite (n = 11 to 19). The association test was performed using CLUMP_v.2.2 software.

Table 3
Haplotype analysis

Haplotype #	4 ht-SNPs				Frequency[c]			*P*-value	Permutation *P*-value[d]
	ZS009[a]	ZS012	ZS004	ZS016[b]	Overall	Control	Case		
# 1	C	T	C	T	0.491	0.520	0.458	0.00054	0.00051
# 2	T	G	C	T	0.302	0.275	0.331	0.00099	0.00097
# 3	T	T	T	–	0.142	0.143	0.140	0.85	0.87
# 4	T	T	C	T	0.039	0.036	0.043	0.36	0.39
# 5	T	T	T	T	0.025	0.021	0.028	0.21	0.24

e



Solid-line circle, "disease-protective" haplotype; dashed-line circle; "at-risk" haplotype group.
[a] ZS009 discriminates between haplotype 1 and other haplotype groups (2, 3, 4, and 5).
[b] ZS016 is an insertion/deletion polymorphism (TTT and TT, a deletion type, are the major and minor allele types, respectively).
[c] Each haplotype frequency is estimated in 781 controls and 711 cases.
[d] Empirical *p* values are derived by permutation using 100,000 replicates.
[e] Haplotype cladogram and its segregation according to the disease status. Observed haplotypes are depicted as filled circles whose sizes reflect haplotype frequency.

because these 17 SNPs were highly correlated with one another ($r^2 > 0.8$, data not shown) and showed similar results in single-locus association tests at their loci (Table 2). To obtain the best partition of haplotypes associated with T2D, we employed a sliding window analysis using the HTR program [32], which compares, in cases and controls, the frequencies of haplotypes generated with a specified number (window size) of contiguous markers. We assumed that the sliding window approach might better localize the region encompassing disease-susceptibility variants since it would give a quantitative idea of the signal observed in the "background." As shown in Fig. 3, results with window sizes of 2 (two-locus) and 3 (three-locus) clearly indicated SNPs in the 5′ region of the *SOCS2* gene to be associated with an increased risk for T2D with high statistical significance. The strongest association was observed with a pair of SNPs, ZS011–ZS013 ($p = 0.000035$, two-locus) and a trio of SNPs ZS011–ZS013–ZS014 ($p = 0.000019$, three-locus). We tested with window sizes up to 6 and the results were essentially the same (data not shown). SNPs ZS011, ZS013, and ZS014, together with two additional SNPs (ZS010 and SNP488), were in perfect LD with each other ($r^2 = 1$), locating within a 2955-bp genomic segment that included noncoding exon 1 and the 5′ flanking region of the *SOCS2* gene. The locations of the SNPs relative to the putative *SOCS2* transcription start site were as follows: S010, −2681 bp; SNP488, −2488 bp; ZS011, −2245 bp; ZS013, −639 bp; ZS014, +274 bp. Frequencies of the at-risk haplotype (ZS010–SNP488–ZS011–ZS013–ZS014, G-C-A-T-C) and protective haplotype (C-T-G-C-T) were 0.481 and 0.519 in controls and 0.552 and 0.448 in T2D, respectively. The at-risk haplotype was associated with T2D and the association was statistically significant even after adjusting the regression for age and gender (crude OR 1.33, 95% CI 1.08–1.63, $p = 8.8 \times 10^{-5}$, adjusted OR 1.37, 95% CI 1.14–1.66, $p = 0.00098$).

Based on our observations, we concluded that these five SNPs in the regulatory region of the *SOCS2* gene are promising candidates for T2D susceptibility markers in the Japanese population.

Association study of SOCS1 and SOCS3 genes

SOCS2 belongs to the *SOCS* gene family, which has at least eight members (*SOCS1–7* and *CIS*) [33]. Among SOCS proteins, SOCS1 and SOCS3 are known to be natural inhibitors of cytokine signaling and also to influence insulin signaling (for a review, see [34]). Therefore, in addition to the *SOCS2* gene, we speculated that *SOCS1* and *SOCS3* genes might also be good candidates for involvement in the development of both T1D and T2D. In fact, Gylvin et al. [35] previously searched for *SOCS3* mutations in Danish subjects and reported an association between homozygosity for a −920 C-A promoter polymorphism and increased whole-body insulin sensitivity in young healthy unrelated individuals. To test this hypothesis in Japanese subjects, we selected SNPs in the *SOCS1* (SNP SC003, 004, 005, and 006) and *SOCS3* (SNP S3C01, 11, 08, 02, 13, 09, 10, 12, 04) genes, based on our Gene-Centric, Evenly Spaced Common SNPs criteria (see Selection of SNPs) and tested for association in our population (information on SNPs and results of the association study are provided in Appendix E). No significant differences in allelic frequencies of the SNPs between cases and controls were detected, suggesting that neither the *SOCS1* nor the *SOCS3* gene plays a major role in conferring increased susceptibility to T2D in our population.

Expression and functional testing of the SOCS2 gene product

We analyzed the levels of *SOCS2* mRNA expression in different human and mouse tissues and cell lines (Fig. 4). In adult human tissues, a clear differential expression of *SOCS1–*

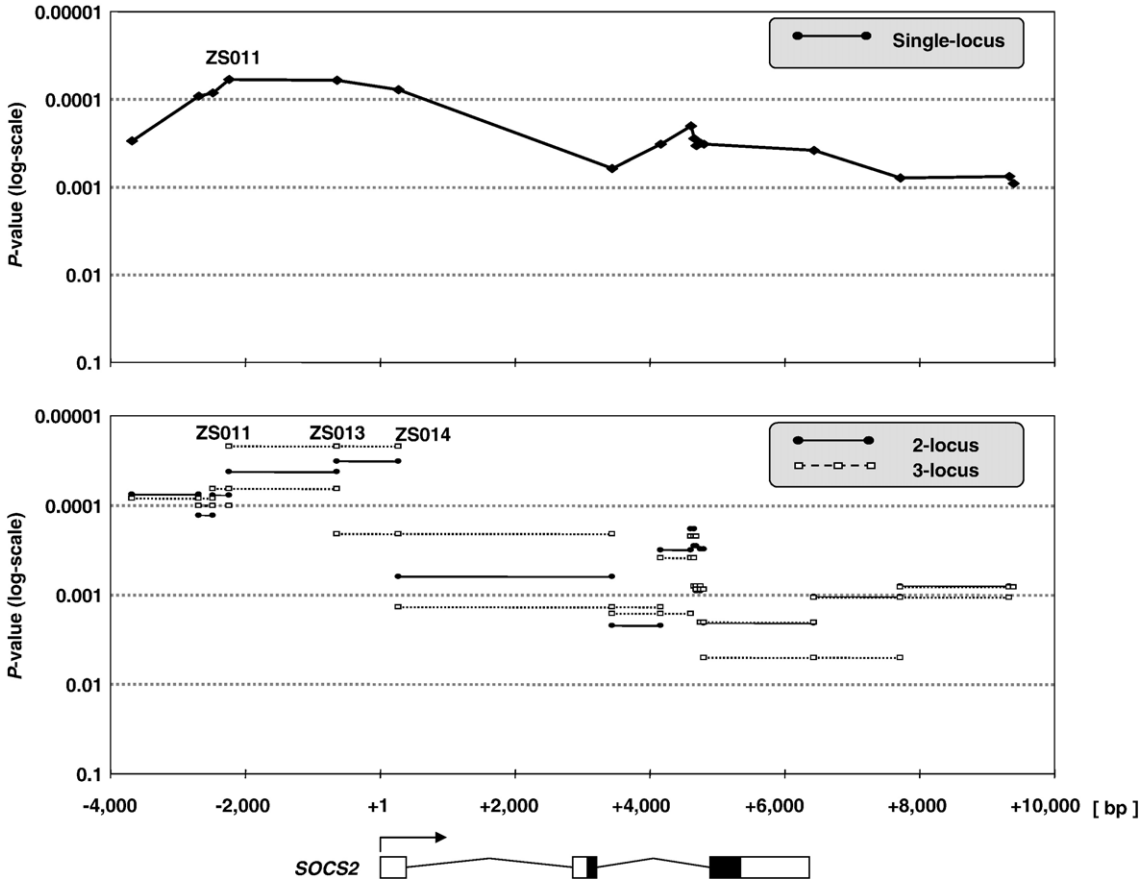


Fig. 3. Results of sliding window analysis in Block 2. Association results for single (top) and sliding windows of two or three SNP haplotypes (bottom) are shown. Each SNP marker is plotted by relative position from the transcription start site of the *SOCS2* gene (shown below). The *p* values for association are plotted on the ordinate by log scale. Note that association results with window sizes of 2 (two-locus) and 3 (three-locus) clearly indicate that SNPs in the 5′ region of the *SOCS2* gene are associated with T2D with high statistical significance.

3 mRNAs was noted (Fig. 4A). *SOCS2* expression was high in the prostate and uterus and relatively low in insulin target tissues such as the liver, skeletal muscle, and adipose. *SOCS1* was more abundantly expressed, with relatively high levels in many other tissues, such as the lung, prostate, uterus, and placenta. *SOCS3* expression was high in adipose, lung, and aortic tissues. In mice, *Socs2* expression was again relatively low in peripheral tissues such as fat, compared to the brain (Fig. 4B). Although we tested only a limited number of cell lines, pancreatic β-cell lines (MIN6, NIT1) apparently expressed relatively high levels of *Socs2* (Fig. 4C).

Because expression data suggest no immediate functional relevance of the *SOCS2* gene in diabetes pathophysiology, we conducted functional screens using an adenovirus-mediated SOCS2 overexpression system. First, glucose-induced insulin secretion was evaluated by overexpressing *SOCS2* in MIN6 or isolated rat islets (Figs. 5A and 5B). Interestingly, *SOCS2* overexpression resulted in significantly decreased glucose-stimulated insulin secretion in both MIN6 (Fig. 5B, Ad-*SOCS2* vs control, 72.9 ± 2.7 vs 94.7 ± 4.0 ng ml^{-1}, $n = 3$, $p < 0.001$) and islets (20.5 ± 2.1 vs 25.9 ± 1.9 ng ml^{-1}, $n = 5$, $p = 0.003$). It is noteworthy that, even in untreated MIN6 cell extracts, endogenous SOCS2 protein was detectable using standard Western blot techniques (Fig. 5A, lane.1), confirming our *SOCS2* expression results. We also tested whether overexpressing *SOCS2* in 3T3-L1 adipocytes affects glucose uptake in response to insulin stimulation; however, no significant difference was observed (data not shown).

Discussion

The current study was designed to determine whether the gene(s) located on chromosome 12q is associated with susceptibility to T2D in the Japanese population. It should be mentioned that, in the Japanese, no linkage study has identified chromosome 12q as a definitive region conferring susceptibility to T2D. However, we hypothesized that the T2D susceptibility gene is likely to be shared among different racial groups, as seen with the *PPARG* and *KCNJ11* genes. When the study was initiated in 2001, there were several uncertainties regarding the accuracy and influencing factors for haplotype inference in random population samples using the EM algorithm, which were later extensively investigated in simulation models and with actual haplotype data [36–38]. Therefore, in this study, we focused on the gene-coding regions on 12q15–q22 and chose SNPs yielding priorities in their locations and allele frequencies, rather than relying on LD information. Recently, Phase I and II results of the International

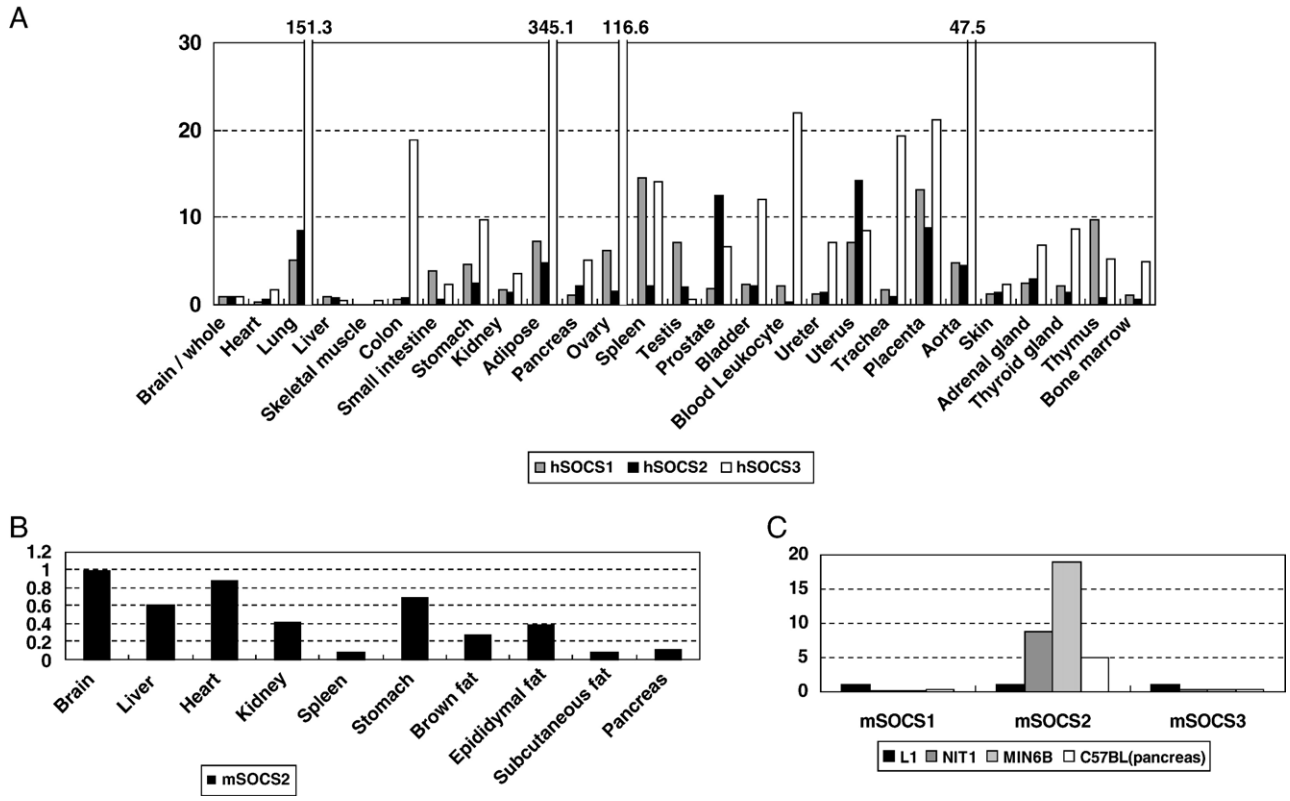


Fig. 4. *SOCS2* gene expression analysis. (A) Quantitative expression analyses of human *SOCS1* (hSOCS1), *2* (hSOCS2), and *3* (hSOCS3) genes in various normal human tissues. Data are presented as the average expression levels after normalization to expression of the human *G3PDH* gene. Expression level of each gene in the brain is set arbitrarily as 1.00. (B) Mouse *Socs2* (mSOCS2) expression in adult mouse tissues. Data are presented as the average expression levels after normalization to expression of the mouse β-actin gene. Relative expression levels are shown with the *Socs2* expression level in the brain set arbitrarily as 1.00. (C) Expression of mouse *Socs1* (mSOCS1), *2* (mSOCS2), and *3* (mSOCS3) in cultured cell lines is shown. L1, 3T3-L1 preadipocytes; NIT1, a pancreatic β-cell line established from a transgenic NOD/Lt mouse; MIN6B, mouse insulinoma-derived cell line; C57BL, C57BL/6 mouse pancreas. The expression levels of mouse *Socs1*, *2*, and *3* were standardized for β-actin. Expression level of each gene in the L1 cells is set arbitrarily as 1.00.

HapMap Project, with the increased density of the map to one SNP every kilobase, became available [39,40]. It is now hoped that these results will facilitate the identification of the haplotype blocks and the common haplotypes in the human genome and the definition of an optimum set of tag SNPs for more cost-effective large-scale genetic association studies.

Our results indicate, for the first time, that the *SOCS2* gene is a potential candidate gene for T2D. SNPs significantly associated with Japanese T2D (best SNP: ZS011 (rs12425869), $p = 6.3 \times 10^{-5}$) were located within a 13,584-bp block of LD (Block 2) that contained the entire *SOCS2* gene region. We identified 34 polymorphic sites within Block 2; however, none of these SNPs were predicted to effect amino acid change. Interestingly, a sliding window method of haplotype analysis indicated SNPs mapped to the 5′ region of the *SOCS2* gene to be associated with T2D with a high statistical significance. These findings prompted us to speculate that nucleotide changes in the regulatory region of the *SOCS2* gene might lead to functional consequences in terms of *SOCS2* gene expression and alter susceptibility to T2D. However, a search for potential transcription factor binding motifs in the SNP sites revealed no such motif (TRANSFAC database [41], data not shown). Promoter reporter gene assays using luciferase reporter gene constructs containing the at-risk haplotype (ZS010–SNP488–ZS011–ZS013–ZS014, G-C-A-T-C) are under way in our laboratory, but we have not yet observed any obvious changes in HEK293 or HeLa cells (data not shown). Since an effect on gene function is likely to be small in common disorders like T2D, it is possible that our artificial assay system may not be sensitive enough to detect and assess such subtle effects of SNPs.

The *SOCS2* gene product is a member of the *SOCS* family, which has at least eight members (*SOCS1–7* and *CIS*, the cytokine-inducible SH2 domain-containing protein). It has been suggested that *SOCS* family proteins participate in negative feedback loops in cytokine and growth factor signaling via multiple distinct mechanisms [42,43]. Among the SOCS proteins, SOCS1 and SOCS3 in insulin-sensitive tissues have been shown to increase in insulin-resistant states, such as obesity and endotoxemia. They were also shown to attenuate insulin signaling by binding to the insulin receptor (IR) and inhibiting phosphorylation of insulin receptor substrates 1 and 2 (IRS-1 and IRS-2) [44]. To date, no in vitro experimental data indicating an association between SOCS2 protein and IR signaling have been obtained. SOCS2 protein instead acts as a negative regulator of the growth hormone (GH) and insulin-like growth factor-1 (IGF-1) signaling pathways [45,46]. SOCS2 has been shown to bind to the GH receptor and inhibit GH signaling by multiple mechanisms, for

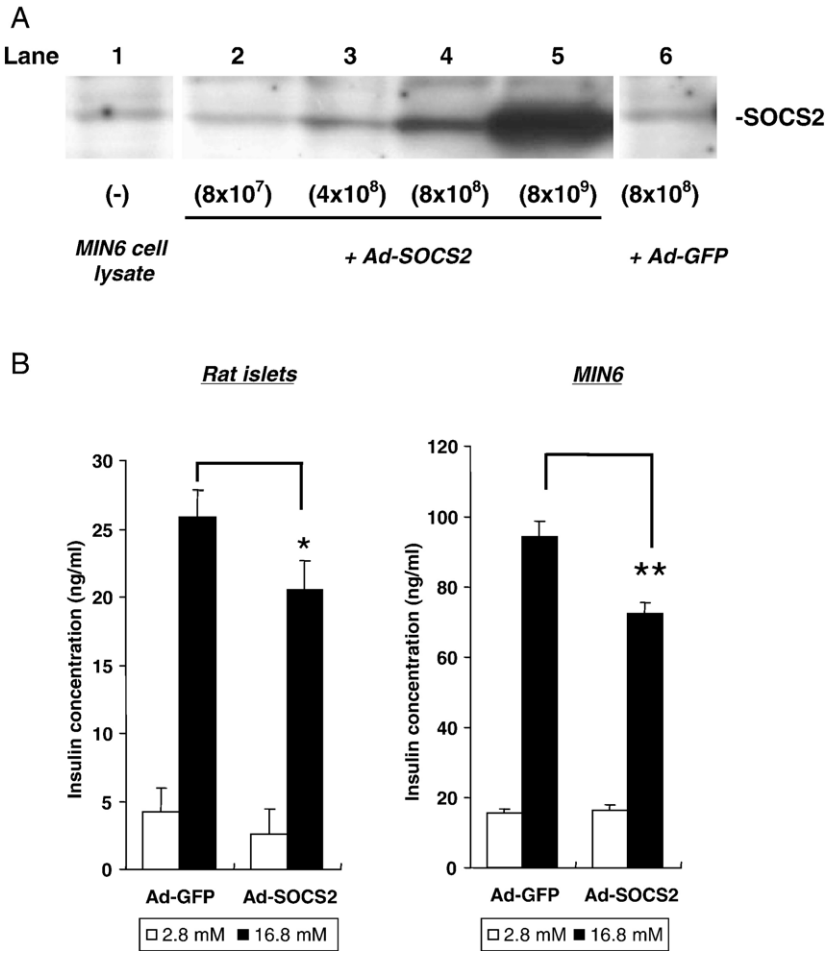


Fig. 5. Functional screen of SOCS2 protein. (A) Construction of the *SOCS2* adenovirus vector (Ad-SOCS2) and endogenously expressed SOCS2 protein in MIN6 cells. Relative levels of SOCS2 protein in MIN6 cells were determined after infecting cells with different amounts of the Ad-SOCS2 adenovirus. Ad-GFP is a control GFP adenovirus. Numbers in parentheses indicate viral titers used for each infection. Whole-cell lysates were prepared at day 3 after infection and were analyzed by immunoblotting using anti-SOCS2 antibody. (B) Effects of *SOCS2* overexpression in MIN6 cells and isolated pancreatic islets on glucose-induced insulin secretion. Insulin release, stimulated with 16.8 mM glucose in MIN6 cells (right) and islets (left) under static incubation conditions, was inhibited by overexpression of *SOCS2* (Ad-SOCS2). Results are expressed as the means ± SE ($n = 3–5$). *$p < 0.05$, **$p < 0.01$.

example by attenuating STAT5 activation. *Socs2* knockout mice exhibit gigantism characterized by increased body weight and length [47], though neither T2D nor diabetes-related phenotypes were reported in these mice. Importantly, the GH/IGF-1 axes are known to be important mitogenic signal transduction pathways in β cells [48]. Pancreatic β cells express both GH and IGF-1 receptors [49]. GH has been shown to stimulate insulin gene transcription, biosynthesis, and secretion and to increase β-cell proliferation. In mice, GH receptor gene deficiency ($GHR^{-/-}$) resulted in elevated insulin sensitivity and diminished pancreatic islet cell mass and serum insulin levels [50]. Mice lacking both IGF-I receptors and IRS-2 show a marked reduction of β-cell mass and die from diabetes due to β-cell insufficiency [51], indicating that IGF-1 receptors couple to IRS-2 in pancreatic islets to mediate β-cell development, proliferation, and survival. From these observations, the dysregulation of SOCS2 might potentially attenuate GH and/or IGF-1 signaling in β cells and thereby lead to loss of β-cell mass and diabetes. Of note, our preliminary experiments with adenoviral-mediated overexpression of *SOCS2* in MIN6 cells or isolated rat islets showed a 25% suppression of glucose-stimulated insulin secretion. This suggests that increased *SOCS2* protein in β cells impairs insulin biosynthesis/storage or glucose sensing, although the exact cellular mechanism is not known at present.

In summary, in the current study, we identified *SOCS2* as a new candidate gene for T2D in the Japanese population. We also obtained detailed information about polymorphic DNA sites and haplotype divergence in the *SOCS2* gene region. Although we carefully designed the study protocol to minimize both false positive and false negative association results, this study was no more than a single population-based case–control study. Therefore, replication studies in other independent populations, especially in those in which linkage has been reported, or family-based tests of association are priorities for determining the consistency of our observations. Obviously, future assessment of a quantitative effect of the associated haplotypes on phenotypes will greatly facilitate understanding the physiology. The function and mechanism of SOCS2 action in pancreatic β cells also need to be elucidated.

Materials and methods

Subjects

The study was conducted in accordance with the tenets of the Declaration of Helsinki. All subjects consented to participate in this study in accordance with the process approved by the Ethical Committee for Human Genome and Gene Research at the University of Tokushima. A total of 711 Japanese individuals with T2D (47.3% women; mean age 62.6 ± 11.4 (SD) years; mean BMI 23.7 ± 3.40 (SD)) and 781 control individuals (54.9% women; mean age 40.6 ± 16.3 (SD) years; mean BMI 22.0 ± 3.05 (SD)) were recruited mainly from the Eye Clinic at Tokushima University Hospital and Kyoto Prefectural University Hospital and its affiliates. The diagnosis of T2D was made either by the 1985 WHO criteria [52] or by being treated with medications for diabetes. Genomic DNA samples were extracted from peripheral blood leukocytes or EB virus-immortalized B lymphoblasts according to a standard method. Some genomic DNA samples were amplified prior to genotyping using the GenomiPhi DNA Amplification Kit (Amersham Biosciences, Uppsala, Sweden).

SNP genotyping

Most SNPs were genotyped by TaqMan assays [25] according to the manufacturer's protocol with modifications, using commercially available reagents (TaqMan Universal PCR Mastermix, No UNG, TaqMan SNP Genotyping Assays, Custom TaqMan SNP Genotyping Assays; Applied Biosystems, Foster City, CA, USA). Our success rate for allele-calling in the Stage 1 study was 99.86% (=177,595/177,840 genotypes, 585 SNPs × 304 subjects). Genotyping accuracy was assessed as follows: 30 SNPs were randomly chosen and genotypes of 32 subjects were determined by direct sequencing analysis and compared with the results of TaqMan assays. Genotype concordance between the two genotyping methods was 100% (959/959 genotypes, a single assay resulted in an "undetermined" genotype and was excluded from the calculation).

In the case of SNPs being problematic in designing TaqMan probes or optimizing assays, PCR allele-specific primer (PCR-ASP) assays [53] or direct sequencing assays were used for genotyping. In PCR-ASP assays, the amount of PCR product amplified with allele-specific primers was monitored using the ABI Prism 7900HT sequence detection system (Applied Biosystems) with fluorescent double-strand DNA binding dye (SYBR Green PCR Master Mix; Applied Biosystems). Genotypes were identified by comparing the amount of each allele-specific amplicon. For the PCR direct sequencing analysis, PCR amplicons (averaging 500–600 bases in length) were sequenced directly from both sense and antisense ends with an ABI Prism BigDye Terminator Cycle Sequencing-Ready Reaction Kit (version 1.1 or 3.1; Applied Biosystems) using automated sequencers (ABI Prism 3100 genetic analyzer and Applied Biosystems 3730xl DNA analyzer; Applied Biosystems). One microsatellite (MS1) was genotyped by electrophoresing the fluorescence-labeled PCR amplicons with an automated sequencer (Applied Biosystems 3730xl DNA analyzer; Applied Biosystems) followed by analysis using GeneMapper_v3.0 (Applied Biosystems).

Genetic computer software

Computer software programs used in this study were as follows:

(1) CLUMP_v.2.2 (http://www.mds.qmw.ac.uk/statgen/dcurtis/software.html): we used this software to test for association with a microsatellite marker, MS1. The CLUMP program is based on the Monte Carlo method of assessing the significance of association with each allele of a microsatellite [54].
(2) Haploview_v.3.0 (http://www.broad.mit.edu/mpg/haploview/index.php): we used this free software to analyze and visualize the LD and haplotypic patterns [28].
(3) SNPAlyze_v.3.2.1PRO (http://www.dynacom.co.jp/products/package/snpalyze/): we purchased this software from Dynacom Co., Ltd. (Yokohama, Japan). The SNPAlyze program estimates haplotype frequencies by employing the EM algorithm and compares haplotype frequencies in cases and controls using a likelihood ratio test [55]. In the present study, we calculated empirical p values by using permutation tests for 1,000,000 replicates.
(4) HTR (http://statgen.ncsu.edu/zaykin/htr.html): the HTR program estimates haplotype frequencies using the EM algorithm and then relates the inferred haplotype frequencies to the observed phenotype using a regression model [32]. In the present study, we calculated empirical p values by using permutation tests for 1,000,000 replicates. The HTR also provides a sliding window mode of haplotype association analysis, which compares, in cases and controls, the frequencies of haplotypes generated with a specified number (window size) of contiguous markers.
(5) Haplo.stats_v.1.1.1 (http://www.mayo.edu/hsr/people/schaid.html): this free software requires the open source statistical programming package *R* (available at http://cran.r-project.org/) to run. The Haplo.stats programs assigns a probability for each haplotype pair in each individual and then models an individual's phenotype directly as a function of each inferred haplotype pair to account for haplotype ambiguity [56,57]. With this program, analysis with adjustment for covariates and computation of simulation p values for each haplotype can be performed. In the present study, we controlled for age and sex as covariates and compared unadjusted and adjusted results.
(6) EHAP_v.1.1 (http://wpicr.wpic.pitt.edu/WPICCompGen/): we used this program for the cladistic-based association analysis, pioneered by Templeton et al. [29]. The EHAP utilizes information contained in the evolutionary relationships among haplotypes and in the sample using generalized linear models. A sequential test series is performed first considering haplotypes that fall into the external nodes of the cladogram (zero-step clades) and then considering one-step clades (produced by moving backward one mutational step from the zero-step clades toward internal nodes), two-step clades, and so forth. From an ancestral haplotype, polymorphisms lead to the observed diversity of haplotypes, which is summarized in a cladogram. By assessing the distribution of phenotypes within a cladogram, it is possible to search for the putative causal polymorphism (mutation) with an increased statistical power.
(7) SNPtagger (http://www.well.ox.ac.uk/~xiayi/haplotype/index.html): we used this World Wide Web-based program to search for minimal htSNPs from inferred haplotypes.
(8) FPRP calculation sheet (http://jncicancerspectrum.oxfordjournals.org/cgi/content/full/jnci;96/6/434/DC1): Wacholder et al. recently proposed a measure called false positive report probability to assess the likelihood that a molecular epidemiological finding is a false positive [27]. The FPRP requires prior estimation of the probability that a gene is associated with the phenotype and is compatible with the Bayesian view that the prior credibility of the hypothesis should influence evaluation of association studies. To obtain FPRP values, the estimated prior probability range, the statistical power to detect an OR of 1.5, and observed ORs and 95% confident intervals were entered into the calculation sheet provided by Wacholder et al. [27].

Expression analyses

Quantitative RT-PCR analyses to estimate the expression of human and mouse *SOCS1–3* were performed using an ABI Prism 7900 sequence detection system (Applied Biosystems) with SYBR Green PCR Master Mix (Applied Biosystems), following the manufacturer's instructions. The primers for detecting mRNA of human *SOCS1–3* were *SOCS1* (NCBI Reference Sequence NM_003745), hSOCS1F, 5′-TAGCACACAACCAGGTGGCA-3′, and hSOCS1R, 5′-GCTCTGCTGCTGTGGAGACTG-3′; *SOCS2* (NM_003877), hSOCS2F, 5′-CCTTTATCTGACCAAACCGCTCTA-3′, and hSOCS2R, 5′-TGTTAATGGTGAGCCTACAGAGATG-3′; *SOCS3* (NM_003955), hSOCS3F, 5′-TCTCTGTCGGAAGACCGTCAA-3′, and hSOCS3R, 5′-CGGCAGCTGGGTGACTTT-3′. The primers for detecting mouse *Socs1–3* mRNAs were *Socs1* (NM_009896), mSOCS1F, 5′-GGCACCTTC-TTGGTGCGC-3′, and mSOCS1R, 5′-AAGCCATCTTCACGCTGAGC-3′; *Socs2* (NM_007706), mSOCS2F, 5′-ATATCCGTTAAGACGTCAGCTGG-3′, and mSOCS2R, 5′-TATGATAGAATCCAATCTGAATTTCCC-3′; *Socs3* (NM_007707), mSOCS3F, 5′-CAAGACCTTCAGCTCCAAAAGC-3′, and mSOCS3R, 5′-CTCCAGTAGAATCCGCTCTCCT-3′. *SOCS1–3* gene expres-

sion was normalized with the expression of *G3PDH* for human and β-actin for mouse mRNA. The primer pairs used for human *G3PDH* and mouse β-actin were hG3PDH-F, 5′-GGGAAGGTGAAGGTCGGA-3′, and hG3PDH-R, 5′-GCAGCCCTGGTGACCAG-3′; mβ-actinF, 5′-TCTGTGTGGATCGGTGGC-3′, and mβ-actinR, 5′-GCTGATCCACATCTGCTGGA-3′. All oligonucleotides were purchased from ProOligo (Lismore, NSW, Australia).

Total RNA was extracted from the pancreas of C57BL/6 mice and various cell lines (L1, NIT1, and MIN6), using the RNeasy kit (Qiagen N.V., Venlo, The Netherlands) or Isogen reagents (Nippongene, Toyama, Japan). Total RNA was reverse-transcribed using Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and random hexamer primers to generate first-strand cDNA. Human cDNA was purchased from BD Biosciences Clontech (Palo Alto, CA, USA) or the Biochain Institute (Hayward, CA, USA).

Adenovirus construction

Adenoviruses containing mouse *Socs2* were constructed as previously described [58]. Full-length mouse *Socs2* cDNAs were generated by RT-PCR with forward primer 5′-CGGGGTACCGCCATGACCCTGCGGTGCCTG-GAGCCCTCC-3′ and reverse primer 5′-CCGCTCGAGTTATACCTGGAATT-TATATTCTTCCAA-3′; the *Kpn*I and *Xho*I sites included in these primers are underlined. The cDNA fragment was then inserted into the *Kpn*I–*Xho*I site of the pAdTrack-CMV vector [58], and the adenoviruses were replicated and amplified in HEK293 cells.

Adenovirus-mediated overexpression of *SOCS2*in MIN6 cells or rat islets was achieved as previously described [59], with concentrations of 10^8 pfu/ml for MIN6 cells and 10^6 pfu/ml for islets. For Western blotting, the MIN6 cell lysate was boiled in the SDS sample buffer with 10 mM dithiothreitol and then applied to SDS–PAGE and transferred onto to a polyvinylidene difluoride membrane (Immun-Blot PVDF Membrane; Bio-Rad Laboratories, Hercules, CA, USA). The membrane was incubated with primary anti-SOCS2 antibody (AnaSpec, Inc., San Jose, CA, USA) and thereafter with horseradish peroxidase-conjugated secondary antibody.

Glucose-induced insulin secretion test in MIN6 and isolated rat islets

The mouse insulinoma cell line MIN6 [60] was grown in DMEM containing 15% (v/v) heat-inactivated fetal calf serum (Invitrogen), 100 IU ml^{-1} penicillin, and 100 mg ml^{-1} streptomycin. Cells were cultured in a humidified atmosphere with 5% CO_2 at 37°C. Pancreatic islets were isolated from 6- to 8-week-old fed Wistar rats (CLEA Japan, Inc., Tokyo, Japan) by collagenase digestion in conjunction with Ficoll gradient separation. Islets were hand-picked under the microscope and cultured overnight in RPMI 1640 medium containing 10% (v/v) heat-inactivated fetal calf serum (Invitrogen).

For glucose-induced insulin secretion testing, MIN6 cells or isolated rat islets were washed in Krebs–Ringer bicarbonate (KRB) buffer supplemented with 2.8 mM glucose, 0.1% BSA, and 10 mM Hepes (pH 7.4) in 24-well dishes. Cells and islets were then preincubated at 37°C for 1 h and further incubated for 20 min (MIN6) or 1.5 h (islets) in the KRB buffer containing 2.8 or 16.8 mM glucose. Insulin concentrations in the medium were determined with an insulin radioimmunoassay kit (Amersham Biosciences, Buckinghamshire, England) using rat insulin as a standard. All data are presented as means ± SE. Student's *t* test was used to determine the significance of differences between groups.

Acknowledgments

The authors thank the patients and volunteer blood donors for participating in this study. We also thank Dr. Masayuki Saito, Dr. Kenji Suzuki, and Dr. Takao Suzuki (Genomics Research Molecular Medicine Laboratories, Institute for Drug Discovery Research, Astellas Pharmaceutical Co., Ltd.) for helpful discussions. We are grateful to Dr. Hidetoshi Inoko (Tokai University School of Medicine) for his cooperation. This study was supported by grants from the Japan Biological Information Consortium affiliated with the New Energy and Industrial Technology Development Organization, Japan Society for the Promotion of Science (Grant for Genome Research of the Research for the Future Program), and the Ministry of Education, Science, and Technology (Knowledge Cluster Initiative) of Japan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2005.11.009.

References

[1] R.A. DeFronzo, Lilly lecture 1987. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM, Diabetes 37 (1988) 667–687.
[2] K. Almind, A. Doria, C.R. Kahn, Putting the genes for type II diabetes on the map, Nat. Med. 7 (2001) 277–279.
[3] L. Hansen, O. Pedersen, Genetics of type 2 diabetes mellitus: status and perspectives, Diabetes Obes. Metab. 7 (2005) 122–135.
[4] Y. Horikawa, et al., Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus, Nat. Genet. 26 (2000) 163–175.
[5] J.M. Servitja, J. Ferrer, Transcriptional networks controlling pancreatic development and beta cell function, Diabetologia 47 (2004) 597–613.
[6] K. Yamagata, et al., Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1), Nature 384 (1996) 458–460.
[7] S. Ghosh, et al., Type 2 diabetes: evidence for linkage on chromosome 20 in 716 Finnish affected sib pairs, Proc. Natl. Acad. Sci. USA 96 (1999) 2198–2203.
[8] L.D. Love-Gregory, et al., A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 alpha gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an Ashkenazi Jewish population, Diabetes 53 (2004) 1134–1140.
[9] K. Silander, et al., Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes, Diabetes 53 (2004) 1141–1149.
[10] K. Ozaki, et al., Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction, Nat. Genet. 32 (2002) 650–654.
[11] S. Tokuhiro, et al., An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis, Nat. Genet. 35 (2003) 341–348.
[12] A. Suzuki, et al., Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis, Nat. Genet. 34 (2003) 395–402.
[13] M.P. Stern, The search for type 2 diabetes susceptibility genes using whole-genome scans: an epidemiologist's perspective, Diabetes Metab. Res. Rev. 18 (2002) 106–113.
[14] R.L. Hanson, et al., An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians, Am. J. Hum. Genet. 63 (1998) 1130–1138.
[15] S.C. Elbein, M.D. Hoffman, K. Teng, M.F. Leppert, S.J. Hasstedt, A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians, Diabetes 48 (1999) 1175–1182.
[16] S. Wiltshire, et al., 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 (2001) 553–569.
[17] M.G. Pezzolesi, et al., Examination of candidate chromosomal regions for type 2 diabetes reveals a susceptibility locus on human chromosome 8p23.1, Diabetes 53 (2004) 486–491.

[18] N. Vionnet, et al., Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27–qter and independent replication of a type 2-diabetes locus on chromosome 1q21–q24, Am. J. Hum. Genet. 67 (2000) 1470–1480.
[19] R.E. Pratley, et al., An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians, J. Clin. Invest. 101 (1998) 1757–1764.
[20] A. Bektas, et al., Evidence of a novel type 2 diabetes locus 50 cM centromeric to NIDDM2 on chromosome 12q, Diabetes 48 (1999) 2246–2251.
[21] M.G. Ehm, et al., Genomewide search for type 2 diabetes susceptibility genes in four American populations, Am. J. Hum. Genet. 66 (2000) 1871–1881.
[22] A. Bektas, J.N. Hughes, J.H. Warram, A.S. Krolewski, A. Doria, Type 2 diabetes locus on 12q15: further mapping and mutation screening of two candidate genes, Diabetes 50 (2001) 204–208.
[23] M.M. Mahtani, et al., Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families, Nat. Genet. 14 (1996) 90–94.
[24] J.T. Shaw, et al., Novel susceptibility gene for late-onset NIDDM is localized to human chromosome 12q, Diabetes 47 (1998) 1793–1796.
[25] K.J. Livak, Allelic discrimination using fluorogenic probes and the 5′ nuclease assay, Genet. Anal. 14 (1999) 143–149.
[26] S.B. Gabriel, et al., The structure of haplotype blocks in the human genome, Science 296 (2002) 2225–2229.
[27] S. Wacholder, S. Chanock, M. Garcia-Closas, L. El Ghormli, N. Rothman, Assessing the probability that a positive report is false: an approach for molecular epidemiology studies, J. Natl. Cancer Inst. 96 (2004) 434–442.
[28] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics 21 (2005) 263–265.
[29] A.R. Templeton, E. Boerwinkle, C.F. Sing, A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila, Genetics 117 (1987) 343–351.
[30] X. Ke, L.R. Cardon, Efficient selective screening of haplotype tag SNPs, Bioinformatics 19 (2003) 287–288.
[31] H. Seltman, K. Roeder, B. Devlin, Evolutionary-based association analysis using haplotype data, Genet. Epidemiol. 25 (2003) 48–58.
[32] D.V. Zaykin, et al., Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals, Hum. Hered. 53 (2002) 79–91.
[33] L. Larsen, C. Ropke, Suppressors of cytokine signalling: SOCS, APMIS 110 (2002) 833–844.
[34] D.L. Krebs, D.J. Hilton, A new role for SOCS in insulin action: suppressor of cytokine signaling, Sci. STKE 169 (2003) PE6.
[35] T. Gylvin, et al., Mutation analysis of suppressor of cytokine signalling 3, a candidate gene in Type 1 diabetes and insulin sensitivity, Diabetologia 47 (2004) 1273–1277.
[36] S.A. Tishkoff, A.J. Pakstis, G. Ruano, K.K. Kidd, The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus, Am. J. Hum. Genet. 67 (2000) 518–522.
[37] K.M. Kirk, L.R. Cardon, The impact of genotyping error on haplotype reconstruction and frequency estimation, Eur. J. Hum. Genet. 10 (2002) 616–622.
[38] D. Fallin, N.J. Schork, Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data, Am. J. Hum. Genet. 67 (2000) 947–959.
[39] The International HapMap Consortium, A haplotype map of the human genome, Nature 437 (2005) 1299–1320.
[40] The International HapMap Project, HapMap Public Release No. 19 (2005).
[41] E. Wingender, et al., TRANSFAC: an integrated system for gene expression regulation, Nucleic Acids Res. 28 (2000) 316–319.
[42] T.A. Endo, et al., A new protein containing an SH2 domain that inhibits JAK kinases, Nature 387 (1997) 921–924.
[43] J. Elliott, J.A. Johnston, SOCS: role in inflammation, allergy and homeostasis, Trends Immunol. 25 (2004) 434–440.
[44] K. Ueki, T. Kondo, C.R. Kahn, Suppressor of cytokine signaling 1 (SOCS-1) and SOCS-3 cause insulin resistance through inhibition of tyrosine phosphorylation of insulin receptor substrate proteins by discrete mechanisms, Mol. Cell. Biol. 24 (2004) 5434–5446.
[45] C.J. Greenhalgh, et al., SOCS2 negatively regulates growth hormone action in vitro and in vivo, J. Clin. Invest. 115 (2005) 397–406.
[46] B.R. Dey, S.L. Spence, P. Nissley, R.W. Furlanetto, Interaction of human suppressor of cytokine signaling (SOCS)-2 with the insulin-like growth factor-I receptor, J. Biol. Chem. 273 (1998) 24095–24101.
[47] D. Metcalf, et al., Gigantism in mice lacking suppressor of cytokine signalling-2, Nature 405 (2000) 1069–1073.
[48] C.J. Rhodes, IGF-I and GH post-receptor signaling mechanisms for pancreatic beta-cell replication, J. Mol. Endocrinol. 24 (2000) 303–311.
[49] P.J. Miettinen, T. Otonkoski, R. Voutilainen, Insulin-like growth factor-II and transforming growth factor-alpha in developing human fetal pancreatic islets, J. Endocrinol. 138 (1993) 127–136.
[50] F.P. Dominici, G. Arostegui Diaz, A. Bartke, J.J. Kopchick, D. Turyn, Compensatory alterations of insulin signal transduction in liver of growth hormone receptor knockout mice, J. Endocrinol. 166 (2000) 579–590.
[51] R.N. Kulkarni, et al., Beta-cell-specific deletion of the Igf1 receptor leads to hyperinsulinemia and glucose intolerance but does not alter beta-cell mass, Nat. Genet. 31 (2002) 111–115.
[52] WHO, Diabetes mellitus: report of a WHO study group, Tech. Rep. Ser. 727 (1985).
[53] B.F. Main, P.J. Jones, R.T. MacGillivray, D.K. Banfield, Apolipoprotein E genotyping using the polymerase chain reaction and allele-specific oligonucleotide primers, J. Lipid Res. 32 (1991) 183–187.
[54] P.C. Sham, D. Curtis, Monte Carlo tests for associations between disease and alleles at highly polymorphic loci, Ann. Hum. Genet. 59 (1995) 97–105.
[55] D. Fallin, et al., Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease, Genome Res. 11 (2001) 143–151.
[56] D.J. Schaid, C.M. Rowland, D.E. Tines, R.M. Jacobson, G.A. Poland, Score tests for association between traits and haplotypes when linkage phase is ambiguous, Am. J. Hum. Genet. 70 (2002) 425–434.
[57] H. Zhao, R. Pfeiffer, M.H. Gail, Haplotype analysis in population genetics and association studies, Pharmacogenomics 4 (2003) 171–178.
[58] T.C. He, et al., A simplified system for generating recombinant adenoviruses, Proc. Natl. Acad. Sci. USA 95 (1998) 2509–2514.
[59] H. Ishihara, et al., Effect of mitochondrial and/or cytosolic glycerol 3-phosphate dehydrogenase overexpression on glucose-stimulated insulin secretion from MIN6 and HIT cells, Diabetes 45 (1996) 1238–1244.
[60] J. Miyazaki, et al., Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms, Endocrinology 127 (1990) 126–132.