J Hum Genet (2008) 53:972–982
DOI 10.1007/s10038-008-0339-2

ORIGINAL ARTICLE

Common coding variant in the TCF7L2 gene and study of the association with type 2 diabetes in Japanese subjects

Kiyoshi Kunika · Toshihito Tanahashi · Shusuke Numata · Shu-ichi Ueno · Tetsuro Ohmori · Naoto Nakamura · Kazue Tsugawa · Katsuyuki Miyawaki · Maki Moritani · Hiroshi Inoue · Mitsuo Itakura

Received: 2 June 2008 / Accepted: 1 October 2008 / Published online: 18 November 2008

Abstract Genetic variants of the transcription factor 7-like 2 (TCF7L2) gene affect the risk of type 2 diabetes in populations with multiple ethnic groups. However, a comprehensive survey of this gene has not been done for a Japanese population. Thus, we conducted this gene-based association study, in which the common genetic variants were analyzed. Using 24 Japanese type 2 diabetic subjects, we first screened a 9.5 kb region, which included the entire coding sequence, to assess potential functional variants of TCF7L2. Sequencing revealed a common coding variant (Pro477Thr) in exon 14 of TCF7L2 that was not enrolled in the public SNP database. Nineteen SNPs and the microsatellite DG10S478 were genotyped across the gene in 2,877 unrelated Japanese subjects. This independent screen identified the previously reported rs7903146 with a strongest association (allele $P = 0.0001$, odds ratio = 1.59 [95% confidence interval 1.25–2.01]), but there was no significant association between Pro477Thr and type 2 diabetes (allele $P = 0.64$). Expression of the Pro477Thr variant did not alter TCF7L2 expression in 30 lymphoblast cells. Although a genotypic effect of Pro477Thr on expression of TCF7L2 was not apparent, Pro477Thr was identified as a common variant of TCF7L2 in 2,877 Japanese subjects. Further functional studies are required to determine the possible effect of this coding variant on type 2 diabetes.

Keywords TCF7L2 · Type 2 diabetes · Coding variant · Association study · Japanese

The authors Kiyoshi Kunika and Toshihito Tanahashi contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10038-008-0339-2) contains supplementary material, which is available to authorized users.

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

S. Numata · S. Ueno · T. Ohmori
Department of Psychiatry, Institute of Health Biosciences, The University of Tokushima, 3-18-15, Kuramoto-cho, Tokushima 770-8503, Japan

S. Ueno
Department of Community and Psychiatric Nursing, The University of Tokushima, 3-18-15, Kuramoto-cho, Tokushima 770-8503, Japan

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

Introduction

Variants of the transcription factor 7-like 2 (TCF7L2) gene, a member of the Wnt signaling pathway (Yi et al. 2005), have been associated with type 2 diabetes in Icelandic, Danish, and European-American populations (Grant et al. 2006). TCF7L2 mediates the regulation of glucagons-like peptide-1 expression (Yi et al. 2005), which may be the mechanism by which this gene influences susceptibility to type 2 diabetes.

Grant and colleagues initially reported a strong association with type 2 diabetes in a tetranucleotide microsatellite marker (DG10S478) located in intron 3 of TCF7L2 (Grant et al. 2006). Additionally, five SNPs correlated with

DG10S478 clustered within the same linkage disequilibrium (LD) block spanning from intron 3 to intron 4 showed a significant association with type 2 diabetes. The two most highly correlated SNPs, rs7903146 and rs12255372, were supported by a replication study, and the associations were subsequently confirmed in populations of Caucasian (Florez 2007; Weedon 2007), African (Helgason et al. 2007), and Asian (Ng et al. 2007; Chang et al. 2007) ancestries.

However, most replication studies have focused on the SNPs within a single LD block that encompasses exon 4 and two large flanking introns, while the other TCF7L2 locus dropped out of the screening. Indeed, in Mexican-Americans, the additional association signals extend the length of the reported LD block (Lehman et al. 2007). In addition, unlike populations of European ancestry, the risk of the T allele frequency of rs7903146 is relatively low in Asian populations, including the Japanese (i.e., 2.3% in the HapMap Japanese Databank), suggesting genetic heterogeneity among racial groups (Chang et al. 2007). This finding raised the question of whether these variants are major contributors to type 2 diabetes in the Japanese population.

This study was designed to find functional variant(s) in Japanese subjects. The sequence of the TCF7L2 coding region was used in an association study to identify the major susceptibility variant in 2,877 Japanese subjects. A coding SNP, possibly a functional one, was detected in exon 14 of TCF7L2, and its potential effect on gene expression was further examined.

Subjects and methods

Subjects

The 2,877 Japanese subjects used for this study (Table 1) were presented in several recently published association studies (Kato et al. 2006; Moritani et al. 2007; Sakamoto et al. 2007; Tanahashi et al. 2006). They consisted of 1,444 healthy controls and 1,433 patients with type 2 diabetes.

Controls were recruited from healthy adult volunteers with less than 5.8% of HbA1c and no other diseases after clinical examination. Of the control samples, 42.8% (618/1,444) were collected by the Pharma SNP Consortium (PSC) (http://www.jpma.or.jp/). Their health conditions and negative family histories of diabetes were checked through the Consortium. All subjects were of full Japanese ancestry, suggesting no systematic confounding factor due to population heterogeneity. Previously, the Japanese population has been shown to be homogenous, except for the Ainus from Hokkaido and the Okinawans from Miyako, using genetic markers of human immunoglobulin (Matsumoto 1988). Using a subset of these samples, the assessment of stratification with the genomic-control method resulted in no harmful inflation factor in a recent study (Moritani et al. 2007).

Table 1 Clinical characteristics of the 2,877 Japanese subjects

	Subjects (2,877)		
	Control	Case	*P* value
Number	1,444	1,433	
Gender (male/female)	709/735	733/700	0.271
Age (years)	39.9 ± 14.9	63.2 ± 10.4	<0.0001
BMI (kg/m^2)	22.1 ± 2.9	23.6 ± 3.4	<0.0001
HbA1c (%)	4.8 ± 0.3	7.3 ± 1.4	<0.0001
Age at onset (%)			
<50 years	–	730 (50.9)	–
>50 years	–	697 (48.6)	–
Unknown	–	6 (0.4)	–
Positive family history	–		–
First-degree relatives (%)		760 (53.0)	
With insulin therapy (%)	–	412 (28.8)	–

Age, BMI, and HbA1c are presented as mean ± SD

Statistical significance was calculated using Student's *t* test

BMI, body mass index

Cases were primarily recruited at Tokushima University Hospital, Kyoto Prefecture University Hospital and its affiliated hospitals. All patients were diagnosed with type 2 diabetes based on the 1985 criteria of the World Health Organization (WHO 1985). Patients with mitochondrial disease or maturity-onset diabetes of the young (MODY) were excluded on a clinical basis.

Genomic DNA was prepared from peripheral blood leukocytes or from Epstein–Barr virus-immortalized lymphoblast cells with a standard protocol. The study protocol was approved by the Institutional Review Board of the University of Tokushima. Written informed consent was obtained from all participants prior to blood sampling.

Microsatellite marker assay

Microsatellite marker DG10S478 was assayed as previously described (Togawa et al. 2006). In brief, 5 ng of genomic DNA was mixed with 10 μl of PCR master mix (Platinum Taq DNA Polymerase; Invitrogen, Carlsbad, CA, USA). The size of the PCR product (384 bp) corresponded to six repeats of a tetranucleotide (CTTT). The primer sequences are shown in Table S1 of the "Electronic supplementary material" (Grant et al. 2006).

PCR conditions were as follows: 30 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min after pre-incubation at 94 °C for 2 min with ABI GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). Diluted PCR product was mixed with Hi-Di Formamide™

(Applied Biosystems) and GeneScan-500 (Rox) Size Standard (Applied Biosystems). The number of microsatellite repeats was determined with 3730xl DNA Analyzer (Applied Biosystems), and the results were analyzed using GeneMapper software (Applied Biosystems). The number of repeats was partly confirmed by sequencing.

DNA sequencing

Using 24 Japanese patients, we searched for SNPs in all of the TCF7L2 exons, exon/intron boundaries extending >100 bp into each intron, 5′ and 3′ untranslated regions (UTR), and in the 1-kb region upstream of the start codon. The NCBI RefSeq for the TCF7L2 gene (NCBI Gene ID 6934 and NM_030756) had 14 exons that spanned 215.9 kb of genomic DNA, with a 2.4-kb coding sequence. Three alternative exons were also sequenced: one in intron 3 and two in intron 13. The total re-sequencing length was approximately 9.5 kb. With few exceptions, PCR primers were designed to amplify 400–1,000 bp fragments that covered all exons and flanking intronic sequences. Details of sequencing primers are available upon request.

All PCR products were prepared for sequence analysis by treatment with ExoSAP-IT (Amersham Biosciences, Piscataway, NJ, USA) at 37 °C for 15 min, followed by incubation at 80 °C for 15 min to deactivate the enzyme. PCR products were sequenced using the ABI BigDye terminator cycle-sequencing kit version 1.1 (Applied Biosystems) according to standard protocols. Cycle-sequencing reactions were performed with 25 cycles at 96 °C for 10 s, 50 °C for 5 s and 60 °C for 4 min. The reaction mixture was then filtered through Sephadex™ G-50 fine (Amersham Biosciences) with multiscreen 96-well plates (Millipore, Molsheim, France). After mixing with Hi-Di Formamide™ (Applied Biosystems), analysis was performed on an ABI 3100 or ABI 3730xl automated sequencer (both from Applied Biosystems).

SNPs genotyping and selection

TaqMan minor groove binder (MGB) probe and primer sets were designed by Applied Biosystems and used in standard TaqMan assays (Kato et al. 2006; Moritani et al. 2007; Tanahashi et al. 2006). The PCR conditions were as follows: 95 °C for 10 min for enzyme activation, followed by 40 cycles of 92 °C or 94 °C for 15 s and 60 °C for 1 min. Thermal cycling was conducted on an ABI GeneAmp PCR System 9700 (Applied Biosystems). Each 384-well plate contained 380 samples of unknown genotypes and four samples of no-template controls. After amplification, fluorescence of the VIC and FAM dyes was measured with an ABI Prism 7900HT using Sequence Detector System (SDS) v.2.1 software (both from Applied Biosystems).

Genotyping results were obtained by auto-call mode with the quality score algorithm used in SDS, and were judged independently by two researchers. The results of genotyping showed 100% concordance with sequencing, as previously described (Kato et al. 2006).

Based on previous reports and the original sequencing, we selected 19 SNPs with an average distance of 11.5 kb across the coding region of TCF7L2 (Table S2 of the "Electronic supplementary material"). In the HapMap JPT database (HapMap Public Release #22), 67 SNPs were genotyped with minor allele frequencies (MAF) exceeding 1%. To identify common haplotype-tagging SNPs, the 19 selected SNPs were entered into the Tagger program (http://www.broad.mit.edu/mpg/tagger/) with a MAF threshold of 20% and an r^2 of 0.8. Nineteen SNPs captured 100% of the variance for haplotypes composed of all SNPs with a MAF of >20% across the TCF7L2 gene. Of these 19 SNPs, four (rs7903146, rs12255372, rs7901695, and rs11196205) were included as they showed the strongest correlation to DG10S478, as previously described (Grant et al. 2006). Marker rs7918599 was selected as a substitute for rs7895340, which was not available in the TaqMan assay.

Statistical analyses

SNP genotyping results were calculated with two types of chi-square tests (allele and genotype model) with 2 × 2, and 2 × 3 contingency tables. Nominal two-sided P values were reported and corrected for multiple testing by conducting 10,000 permutations with Haploview 4.0 (http://www.broad.mit.edu/mpg/haploview/index.php). Hardy–Weinberg equilibrium (HWE) of alleles was calculated with genotype frequency, and evaluated by the chi-square test to compare observed and expected values. In controls, SNPs with a HWE test result of $P < 0.05$ were removed from the analysis (Klein et al. 2005).

For microsatellite marker, we calculated the nominal P value with the chi-square test for the 2 x 2 contingency tables with each individual allele. The corrected P value was also calculated using the clump v.2.3 program (http://www.smd.qmul.ac.uk/statgen/dcurtis/software.html). This program employs Monte Carlo simulation to assess the significance of case control association with multi-allelic markers.

The Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) was used for a prior simulation. With a MAF of 0.4, the sample size had over 70% power to detect SNPs with a significance level of 0.05. In less frequent SNPs with a MAF of 0.2, the same sample size showed lower power (57.0%) at the same significance level.

Pairwise linkage disequilibrium (LD) between SNPs was estimated with two popular measurements, $|D'|$ and r^2. LD was separately calculated with 1,444 controls and

1,433 cases. Haplotypes were inferred by the expectation-maximization algorithm in SNPAlyze v.3.2.2 Pro software (DYNACOM, Yokohama, Japan), and defined by an r^2 value of >0.8.

Lymphoblast cells and isolation of total RNA

Thirty samples of Epstein–Barr virus-immortalized lymphoblast cells were prepared from Japanese control subjects (14 male and 16 female) for genotyping. Cells were cultured in RPMI1640 (Gibco BRL, Rockville, MD, USA; catalogue number 11875) and supplemented with 10% fetal bovine serum (Gibco) and antibiotics (100 U/ml penicillin G and 100 mg/ml streptomycin) at 37 °C in a humidified incubator under 5% CO_2. Total RNA was isolated using the RNeasy kit (Qiagen, Stanford, CA, USA) following the manufacturer's protocol.

TCF7L2 mRNA quantification and genotype-specific analysis

Real-time quantitative RT-PCR was used to examine genotype-specific expression (Riancho et al. 2007). cDNA was synthesized from 1 μg of total RNA using oligo d(T)12-18 primers and Super Script III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA). Control cDNA was also synthesized using qPCR reference human total RNA (Clontech, Mountain View, CA, USA).

Two ng of cDNA template were combined with 1x SYBER Green Master Mix (Applied Biosystems), and 200 nM of forward and reverse primers. PCR was performed in triplicate in a MicroAmp optical 384-well plate in a total volume 20 μl using an ABI Prism 7900HT. PCR conditions were 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s and 60 °C for 1 min. Amplification of each genotype was determined by monitoring the fluorescence of SYBER Green. Data were collected with SDS 2.1 software (Applied Biosystems), and the cycle threshold (Ct) of each allele was determined. Although Ct values are directly related to the amount of specific cDNA present in a sample, the Ct may be influenced by other factors, such as fluorescence signal efficiency. Therefore, a standard curve was constructed with serial dilutions of control cDNA. The relative expression of the TCF7L2 gene was normalized to the amount of ACTB expression (NM_001101) in the same cDNA with the standard curve. Two primer sets were used for different coding regions. Primer sequences are shown in Table S1 of the "Electronic supplementary material."

Genotype-specific expression levels were stratified into three statuses: Pro/Pro homozygote, Pro/Thr heterozygote, and Thr/Thr homozygote. A one-way ANOVA test was adapted as a parametric method for the assessment of genotype-specific expression levels.

Results

Microsatellite marker DG10S478

We assayed DG10S478 in 897 controls and 872 cases with a success rate of 99.5% (1,760/1,769), and the results are summarized in Table 2A. The initial report showed that DG10S478 contained a tetranucleotide repeat (CTTT) and had seven alleles of different repeats. Its sixth repeat was the major allele defined as reference allele 0. In contrast, only five DG10S478 alleles, −8, 0, 4, 8, and 12, were found in the Japanese subjects. The frequency of reference allele 0 was higher than in the Icelandic population, and the long alleles (allele 16 and 20) were absent in the Japanese subjects. One variant, the short allele (allele −8) was only observed in this study (Table 2A).

As in the initial report (Grant et al. 2006), all of the nonzero alleles (alleles −8, 4, 8, and 12) were integrated into a composite allele referred to as allele X. Allele X had a frequency of 4.0% in controls and 5.6% in cases (Table 2B), which differed from the frequencies of 27.6 and 36.4%, respectively, found in Icelanders. A weak association (nominal $P = 0.021$) was found between allele 0 and allele X that was not significant when assessed with the Monte Carlo simulation (corrected $P = 0.076$). Consistent with another Japanese report (Hayashi et al. 2007), LD values were lower than those of populations of European ancestry (Table 2C). Collectively, it was rather difficult to completely identify DG10S478 as a causal genetic marker of type 2 diabetes in Japanese.

Sequencing of the TCF7L2 coding region

Three reports have used Japanese subjects in TCF7L2 association studies (Hayashi et al. 2007; Horikoshi et al. 2007; Miyake et al. 2008), but all were conducted as replications in which only minimal sets of known variants were examined. Therefore, we screened the entire coding region and 1,000 bp upstream of TCF7L2.

Grant et al. reported three rare SNPs in the 3′-UTR of TCF7L2. In contrast, we detected a common coding SNP in the Japanese, Pro477Thr (c.C1429A) in exon 14 (CC allele: 5 subjects, CA allele: 15 subjects, and AA allele: 3 subjects) (Fig. 1a, b). This variant mapped on chromosome 10q 114,915,359 (NCBI Build 36), but had not been deposited into HapMap Public Release #22 and has no dbSNP ID (Build 128). Additionally, this base position was conserved across other species, such as mouse, rat, and fugu (UCSC Genome Browser Comparative Genomics). However, Pro477Thr was previously identified in early-onset type 2 diabetes (non-MODY) (Cauchi et al. 2007), East Asian type 2 diabetic subjects (Groves et al. 2006), and Japanese renal cell carcinoma samples

Table 2 Association study and linkage disequilibrium pattern of DG10S478 in 1,760 Japanese subjects

A. Genotype frequencies of DG10S478 in the TCF7L2 gene

Genotype	−8/−8	−8/0	−8/4	−8/8	−8/12	0/0	0/4	0/8	0/12	8/8
Control ($n = 891$) (%)	0 (0.0)	19 (2.1)	0 (0.0)	0 (0.0)	1 (0.1)	822 (92.3)	5 (0.6)	41 (4.6)	2 (0.2)	1 (0.1)
Case ($n = 869$) (%)	1 (0.1)	25 (2.9)	0 (0.0)	2 (0.2)	0 (0.0)	775 (89.2)	8 (0.9)	51 (5.9)	6 (0.7)	1 (0.1)

B. Allele frequencies of DG10S478 in the TCF7L2 gene

	Allele						Reference versus the other allele			
	Allele −8	Allele −4	Allele 0	Allele 4	Allele 8	Allele 12	Allele 0	Allele X	Uncorrected *P* value	Empirical *P* value
Control (%)	20 (1.1)	NA	1,711 (96.0)	5 (0.3)	43 (2.4)	3 (0.2)	1,711 (96.0)	71 (4.0)		
Case (%)	29 (1.7)	NA	1,640 (94.4)	8 (0.5)	55 (3.2)	6 (0.3)	1,640 (94.4)	98 (5.6)	0.021	0.076

C. Linkage disequilibrium value (r^2) among DG10S478 and four SNPs

	rs12255372	rs7903146	rs7901695	rs11196205
DG10S478	0.61	0.72	0.72	0.28
rs12255372		0.54	0.52	0.47
rs7903146			0.99	0.25
rs7901695				0.24

Data are presented as *n* (%)

Allele X indicates all alleles except for reference allele 0

Uncorrected and empirical *P* values were calculated with the Clump ver2.3

NA not available

(Shiina et al. 2003). In addition, a recent study showed that this variant is a common SNP in Chinese despite the lack of association with type 2 diabetes (Ren et al. 2008). Pro477Thr was described to be on exon 17 in those previous reports, but the genomic position was the same as this study, since three alternative exons were added. We also found T > C SNPs in the last part of intron 9 with a MAF of 2%, which is mapped on chromosome 10q 114,902,052. No other coding or noncoding variants were found.

Single SNP association study

Nineteen SNPs were genotyped with 1,444 controls and 1,433 cases. The genotype and allele frequency distributions are shown in Table 3. The genotype distributions of all SNPs were in Hardy–Weinberg equilibrium. When non-template controls (NTC) were incorrectly detected, the genotyping was re-assayed. On average, 99.0% of the attempted genotypes were successful (with success rates ranging from 99.6 to 97.9% for each SNP).

Two of the 19 SNPs, rs7903146 and rs12255372, were significantly associated with type 2 diabetes, which was consistent with other reports including Japanese (Hayashi et al. 2007; Horikoshi et al. 2007; Miyake et al. 2008). The T allele of marker rs7903146 was defined as the causal and ancestral variant in the TCF7L2 gene (Grant et al. 2006). We also found that rs7901695 with a MAF of 4%, located 4 kb upstream of rs7903146, was significantly associated with type 2 diabetes (nominal allele $P = 0.0002$ and permuted $P = 0.0035$), with a strong correlation to rs7903146 ($r^2 = 0.996$). In 2,877 subjects, Pro477Thr showed no significant association despite the common SNPs with a MAF of 24% (allele $P = 0.64$ and genotype $P = 0.87$). This screening did not uncover any stronger association than rs7903146, although the allele frequency was relatively low in the Japanese subjects.

LD pattern and haplotype association study

The LD pattern was not significantly different between the patient and control groups (Fig. 2). A large haplotype block containing two representative markers, rs7903146 and rs12255372, encompasses a portion of intron 3, all of exon 4, and part of intron 4 (Grant et al. 2006). However, in

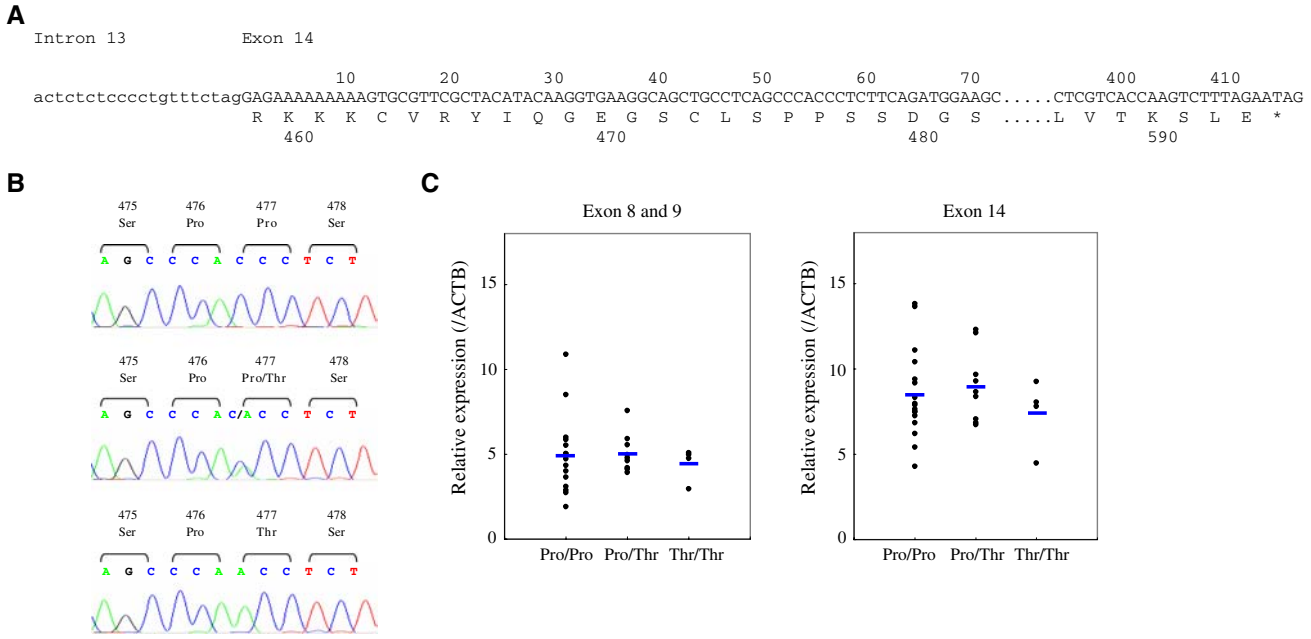

Fig. 1 **a** Partial sequence of the last part of TCF7L2 intron 13 and exon 14. Sequence information was based on the NCBI RefSeq (NM_030756). *Large and small characters* represent the coding and the noncoding sequences, respectively. The coding variant c.C1429A is located at codon 477 of the predicted proline. The stop codon is indicated as an *asterisk*. **b** Mutation analysis of Pro477Thr in exon 14 of TCF7L2. DNA sequence electropherograms from 24 analyzed Japanese type 2 diabetic subjects. Human NCBI RefSeq of TCF7L2 (NM_030756) has a homozygous CC at codon 477 of the predicted proline. Of the 24 analyzed subjects, five were CC homozygous (*upper part*), 15 subjects were CA heterozygous (*middle part*), and three were AA homozygous (*lower part*). In one subject, the sequence was not detected by direct sequencing. **c** TCF7L2 mRNA expression stratified by the Pro477Thr genotypes. TCF7L2 expression level was measured in lymphoblast cells. Expression for the three genotypes of Pro477Thr was analyzed with two primer sets for the middle region (exon 8 and 9) and the 3′ end of the gene (exon 14). A one-way ANOVA test was used to assess significance. Relative expression was calculated by the standard curve method. The expression of TCF7L2 showed no significant difference among the three genotypes with two primer sets for the middle region and 3′ end of the gene ($P = 0.84$ and 0.54, respectively). The means are indicated by *blue horizontal bars*

the Japanese, this region was divided by rs7918599, and rs7903146 and rs12255372 exhibited modest LD values of r^2 (control: 0.54 and case: 0.48). The Pro477Thr variant showed rather low LD values for the other SNPs within this locus. Defined by r^2 values of >0.8, four haplotype blocks were found within the TCF7L2 locus.

Block 2, which constituted of rs7901695 and rs7903146, showed a significant association with type 2 diabetes (Table 4): the T-C haplotype was associated with reduced risk (nominal $P = 0.0004$ and permuted $P = 0.0001$), while the C–T haplotype was associated with an increased disease risk (nominal $P = 0.0005$ and permuted $P = 0.0001$). Thus, these associations remained significant after the multiple corrections with permutation method. In block 3, only the T-G haplotype showed significant association, but its frequency was rather low. The remaining two blocks showed no significant association.

TCF7L2 mRNA expression in three genotypes

As multiple splice forms of TCF7L2 have been reported (Shiina et al. 2003), we designed two primer sets for both the middle region (exons 8 and 9) and the 3′ end of the gene (exon 14). Gene expression was analyzed in 30 lymphoblast cells using real-time quantitative RT-PCR, and the relative expressions of the three Pro477Thr genotypes were evaluated (Fig. 1c).

The primer sets for the middle region of the gene (exons 8 and 9) yielded a relative expression value of 4.93 ± 2.17 (mean ± SD) in 17 Pro/Pro subjects. The 5.07 ± 1.14 in nine Pro/Thr and 4.44 ± 0.99 in four Thr/Thr subjects showed no significant statistical differences (one-way ANOVA test $P = 0.84$). Similarly, no difference was observed with the use of the primer sets for the 3′ end of the gene ($P = 0.54$). In lymphoblast cells, the three Pro477Thr genotypes did not differ in their expression levels.

Discussion

DG10S478, a common microsatellite in TCF7L2 (previously known as TCF-4), has been strongly associated with type 2 diabetes (Grant et al. 2006), and variants of TCF7L2 have been replicated in multiple populations (Florez 2007;

Table 3 Nineteen SNPs of the TCF7L2 gene and their association with type 2 diabetes

	dbSNP ID	Genome position	Location	Allele	Control			Case			MAF (%)		Allele *P*	Genotype *P*	Permuted *P*	Odds ratio (95% CI)
				1/2	1 1	1 2	2 2	1 1	1 2	2 2	Control	Case				
1	rs6585194	114,707,461	Intron 3	C/G	636	631	154	659	599	159	0.33	0.32	0.58	0.52	1	–
2	rs2296784	114,713,549	Intron 3	G/A	719	605	107	726	560	130	0.29	0.29	0.78	0.14	1	–
3	rs7081062	114,730,735	Intron 3	A/G	749	581	97	785	535	96	0.27	0.26	0.20	0.26	0.94	–
4	rs7895307	114,733,951	Intron 3	G/A	752	579	98	777	526	96	0.27	0.26	0.21	0.27	0.95	–
5	rs7901695	114,744,078	Intron 3	T/C	1308	112	3	1240	169	5	0.04	0.06	**0.0002**	**0.0010**	**0.0038**	1.56 (1.23 − 1.98)
6	rs7903146	114,748,339	Intron 3	C/T	1309	111	3	1246	171	5	0.04	0.06	**0.0001**	**0.0006**	**0.0021**	1.59 (1.25 − 2.01)
7	rs7918599	114,767,386	Intron 3	C/T	1260	169	5	1233	181	12	0.06	0.07	0.15	0.17	0.88	–
8	rs11196205	114,797,037	Intron 4	G/C	1253	166	5	1244	167	10	0.06	0.07	0.54	0.43	1	–
9	rs12255372	114,798,892	Intron 4	G/T	1341	78	2	1302	104	2	0.03	0.04	**0.047**	0.12	0.47	1.34 (1.002 − 1.80)
10	rs11196213	114,811,544	Intron 4	C/T	1247	183	10	1218	200	8	0.07	0.08	0.45	0.54	1	–
11	rs7919409	114,814,966	Intron 4	T/C	829	511	95	769	554	102	0.24	0.27	0.06	0.12	0.56	–
12	rs3750804	114,823,840	Intron 4	C/T	953	440	45	898	462	64	0.18	0.21	**0.029**	0.07	0.35	1.16 (1.02 − 1.32)
13	rs3814572	114,837,713	Intron 4	A/G	1103	310	17	1120	298	11	0.12	0.11	0.33	0.44	0.99	–
14	rs11196224	114,845,387	Intron 4	C/T	784	536	113	757	566	95	0.27	0.27	0.95	0.25	1	–
15	rs290477	114,856,622	Intron 4	C/A	772	546	116	761	556	95	0.27	0.26	0.54	0.35	1	–
16	rs911768	114,864,761	Intron 4	G/C	956	426	47	915	447	48	0.18	0.19	0.31	0.53	0.99	–
17	rs7906315	114,877,208	Intron 4	A/G	732	567	109	712	591	105	0.28	0.28	0.64	0.65	1	–
18	rs7903424	114,894,026	Intron 6	G/A	1096	327	19	1097	303	23	0.13	0.12	0.65	0.56	1	–
19	Pro477Thr	114,915,359	Exon 14	C/A	822	536	81	828	518	79	0.24	0.24	0.64	0.87	1	–

Genome positions were based on NCBI Human Build 36

Pro477Thr was detected by sequencing

P values <0.05 are shown in boldface

MAF, minor allele frequency

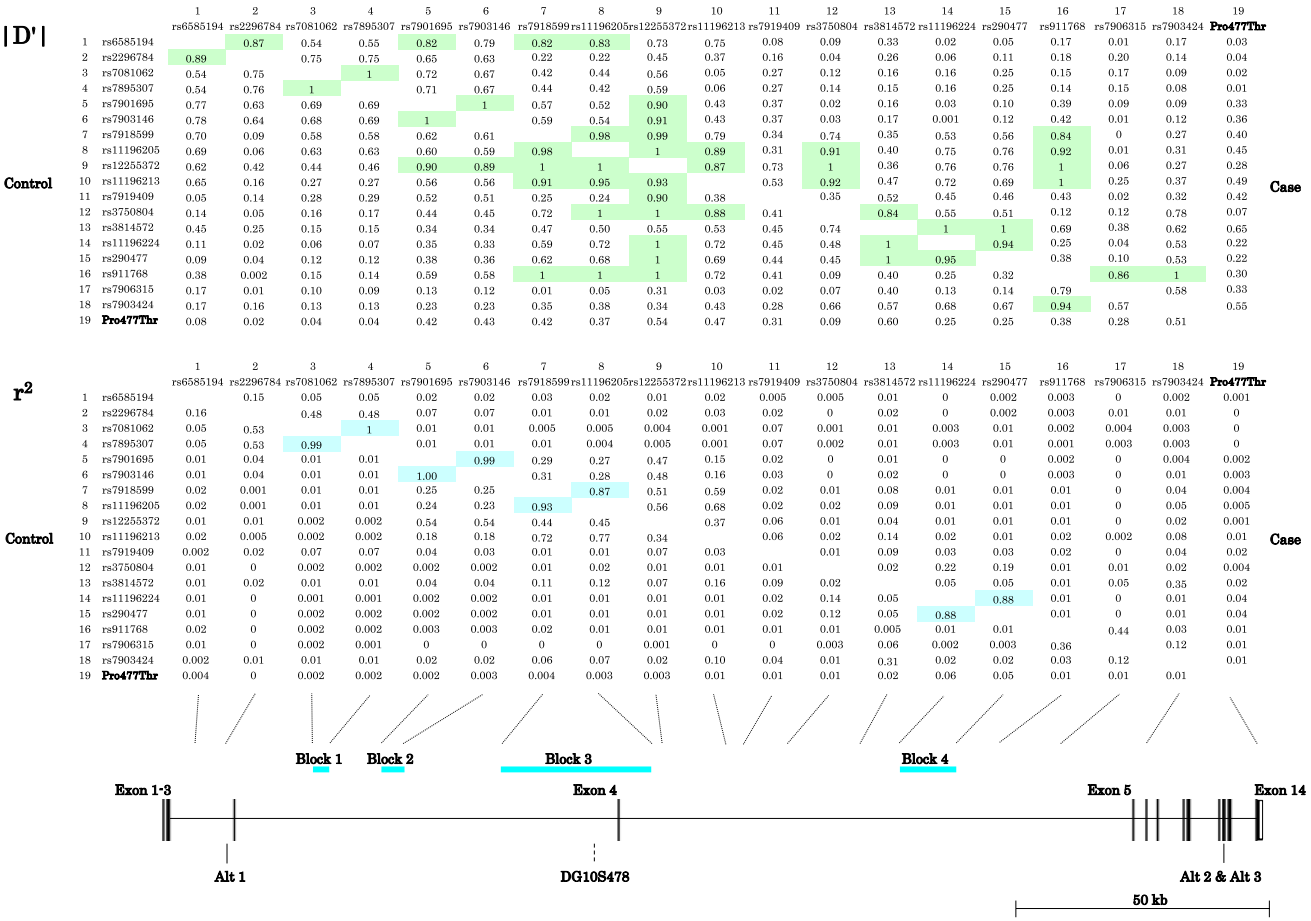

Fig. 2 Pairwise LD values of |D′| (*upper part*) and r^2 (*middle part*), and structure of the TCF7L2 gene (*lower part*). LD values with control are shown *below the diagonal*, and cases are shown *above the diagonal*. *Light green* coloring indicates a |D′| of >0.8 and *light blue* indicates r^2 values over 0.8. Four haplotype blocks are shown as *blue horizontal bars* defined by r^2 over 0.8. The exon regions of TCF7L2 are shown with *rectangles* and numbered in order from 1 to 14 based on the NCBI RefSeq (NM_030756). Three alternative exons are also shown as *black vertical bars*. Microsatellite marker DG10S478 is indicated as a *black broken bar*

Weedon 2007). Subsequently, TCF7L2 was most strongly associated with type 2 diabetes in recent genome-wide association studies in Caucasians (Sladek et al. 2007; The Wellcome Trust Case Control Consortium 2007; Saxena et al. 2007; Zeggini et al. 2007; Scott et al. 2007). However, the specific etiological variant of TCF7L2 has not been clearly identified. Marker rs7903146 displays a strong association and remains the most likely candidate (Florez 2007; Weedon 2007). The risk T allele was recently associated with impaired insulin secretion and enhanced rate of hepatic glucose production (Lyssenko et al. 2007), but we could not examine the correlation between candidate variants of TCF7L2 and clinical phenotype due to the restricted clinical information. There was no apparent mechanism by which it affects the function of TCF7L2, since this variant is located in intron 3 (a non-coding section of the gene). Moreover, published reports from other ethnic groups suppose that different variants of the TCF7L2 gene conferred the risk of type 2 diabetes in different populations (Chang et al. 2007; Lehman et al. 2007).

In the present study, we applied a gene-based approach, in which the common genetic variants within the TCF7L2 locus were considered in association analyses, and explored the effect of these TCF7L2 variants in Japanese subjects. We found that the risk-conferring alleles in TCF7L2, such as rs7903146 T and rs12255372 T, are present at relatively low levels in the Japanese. The low allele frequencies prompted a search for other SNPs, especially common variants. Despite the sequencing effort, we did not detect any variants with a stronger association than rs7903146. However, the re-sequencing focused on only the coding region of TCF7L2. It is possible that the important functional variant occurs in unsequenced regions, such as the putative regulatory locus (Hudson 2003). Indeed, one SNP with an as-yet unclear role in gene function was detected at an exon/intron boundary. Further

Table 4 Association analysis of TCF7L2 haplotypes with type 2 diabetes

Block	dbSNP ID		Haplotype	Frequency		Nominal *P*	Permutated *P*
	Major > minor			Control	Case		
Block 1	rs7081062	rs7895307	A–G	0.728	0.742	0.22	0.21
	A > G	G > A	G–A	0.270	0.257	0.27	0.26
			G–G	0.002	0.001	0.17	0.18
			A–A	0	0	–	–
Block 2	rs7901695	rs7903146	T–C	0.959	0.938	**0.0004**	**0.0001**
	T > C	C > T	C–T	0.041	0.062	**0.0005**	**0.0001**
			C–C	0	0.0004	–	–
			T–T	0	0	–	–
Block 3	rs7918599	rs11196205	C–G	0.936	0.927	0.21	0.20
	C > T	G > C	T–C	0.060	0.064	0.55	0.57
			T–G	0.002	0.007	**0.008**	**0.009**
			C–C	0.001	0.001	0.71	1
Block 4	rs11196224	rs290477	C–C	0.720	0.724	0.74	0.74
	C > T	C > A	T–A	0.256	0.252	0.75	0.75
			C–A	0.014	0.011	0.26	0.27
			T–C	0.010	0.013	0.28	0.30

Haplotype frequencies under 0.0001 are shown as 0

P values <0.05 are shown in boldface

The number of permutation = 10,000

extensive searches are recommended if the functional variant in the population of interest is to be identified.

Sequencing revealed a common Pro477Thr variant in the exon 14 of TCF7L2. Notably, this variant was not deposited in the HapMap database, but had been previously identified in type 2 diabetic subjects (Cauchi et al. 2007; Groves et al. 2006). If we had designed this study to be based on the available information in the established SNP database, we would have failed to notice a common coding variant with a potential functional effect. Since the identification of the functional variant is central to the elucidation of the genetic components of complex diseases, actual sequencing remains an important element of the search for disease-conferring variants (Rebbeck et al. 2004).

Markers rs7903146 and rs12255372 were found to be in modest LD (r^2 for control: 0.54; case: 0.48), consistent with Chinese (Ren et al. 2008), but in high LD with the HapMap Caucasians ($r^2 = 0.75$). Although there were different genetic backgrounds among different ethnic groups, LD in this locus was supposed to be preserved in at least Asians including Japanese and Chinese. Except for ours, there was no information about the LD pattern related to Pro477Thr.

The genomic structure of TCF7L2 contains three major domains: the β-catenin binding domain (exon 1), DNA-binding HMG (high mobility group) boxes (exons 10 and 11), and CtBP (COOH-terminal-binding protein) binding sites (exon 14) (Shiina et al. 2003; Duval et al. 2000). CtBP reportedly represses the transcriptional activity of the target genes (Shiina et al. 2003), and Pro477Thr was located in exon 14 along with binding sites for this repressor. In addition, the human TCF7L2 gene displays a high degree of alternative splicing near the 3′ end of the gene (Shiina et al. 2003). Unfortunately, we could not verify any relationship with the expression level in lymphoblast cells, and the present data do not support a major role for Pro477Thr in the development of type 2 diabetes in Japanese. Functional evidence for the TCF7L2 variant has been explored since the initial report of association with type 2 diabetes, but few effects have been published (Cauchi et al. 2008; Shu et al. 2008). The Wnt signaling pathway, in which TCF7L2 plays a central role, is known to regulate adipogenesis and adipose cell differentiation (Cauchi et al. 2008). This study did not cover every cell model related to type 2 diabetes, such as pancreatic beta cell, liver, and skeletal muscle. Further functional studies will be necessary to identify the effect, if any, of Pro477Thr on TCF7L2 expression.

In conclusion, this study did not reveal SNPs with a stronger association with type 2 diabetes than that of rs7903146. However, the common Pro477Thr variant was located in the last exon of TCF7L2, which is a locus with a high degree of alternative splicing. The study also showed the importance of applying sequencing to original samples in the population of interest when conducting a search for variants that affect disease risk. Further functional studies

are required to define the role of the Pro477Thr variant in TCF7L2 gene expression.

Acknowledgments This study was supported by a grant from Cooperative Link of Unique Science and Technology for Economy Revitalization (CLUSTER, Tokushima, Japan). We thank the members of the Institute of Genome Research for helpful discussions and assistance.

Conflict of interest statement None

References

Cauchi S, Choquet H, Gutierrez-Aguilar R, Capel F, Grau K, Proenca C, Dina C, Duval A, Balkau B, Marre M, Potoczna N, Langin D, Horber F, Sorensen TI, Charpentier G, Meyre D, Froguel P (2008) Effects of TCF7L2 polymorphisms on obesity in European populations. Obesity 16:476–482

Cauchi S, Vaxillaire M, Choquet H, Durand E, Duval A, Polak M, Froguel P (2007) No major contribution of TCF7L2 sequence variants to maturity onset of diabetes of the young (MODY) or neonatal diabetes mellitus in French white subjects. Diabetologia 50:214–216

Chang YC, Chang TJ, Jiang YD, Kuo SS, Lee KC, Chiu KC, Chuang LM (2007) Association study of the genetic polymorphisms of the transcription factor 7-like 2 (TCF7L2) gene and type 2 diabetes in the Chinese population. Diabetes 56:2631–2637

Duval A, Rolland S, Tubacher E, Bui H, Thomas G, Hameli R (2000) The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines. Cancer Res 60:3872–3879

Florez JC (2007) The new type 2 diabetes gene TCF7L2. Curr Opin Clin Nutr Metab Care 10:391–396

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

Groves CJ, Zeggini E, Minton J, Frayling TM, Weedon MN, Rayner NW, Hitman GA, Walker M, Wiltshire S, Hattersley AT, McCarthy MI (2006) Association analysis of 6,736 UK subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. Diabetes 55(9):2640–2644

Hayashi T, Iwamoto Y, Kaku K, Hirose H, Maeda S (2007) Replication study for the association of TCF7L2 with susceptibility to type 2 diabetes in a Japanese population. Diabetologia 50:980–984

Helgason A, Pálsson S, Thorleifsson G, Grant SF, Emilsson V, Gunnarsdottir S, Adeyemo A, Chen Y, Chen G, Reynisdottir I, Benediktsson R, Hinney A, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Schäfer H, Faruque M, Doumatey A, Zhou J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Sigurdsson G, Hebebrand J, Pedersen O, Thorsteinsdottir U, Gulcher JR, Kong A, Rotimi C, Stefánsson K (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nat Genet 39:218–225

Horikoshi M, Hara K, Ito C, Nagai R, Froguel P, Kadowaki T (2007) A genetic variation of the transcription factor 7-like 2 gene is associated with risk of type 2 diabetes in the Japanese population. Diabetologia 50:747–751

Hudson TJ (2003) Wanted: regulatory SNPs. Nat Genet 33:439–440

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

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

Lehman DM, Hunt KJ, Leach RJ, Hamlington J, Arya R, Abboud HE, Duggirala R, Blangero J, Goring HH, Stern MP (2007) Haplotypes of transcription factor 7-like 2 (TCF7L2) gene and its upstream region are associated with type 2 diabetes and age of onset in Mexican Americans. Diabetes 56:389–393

Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, Sjögren M, Ling C, Eriksson KF, Lethagen AL, Mancarella R, Berglund G, Tuomi T, Nilsson P, Del Prato S, Groop L (2007) Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest 117:2155–2163

Matsumoto H (1988) Characteristics of Mongoloid and neighboring populations based on the genetic markers of human immunoglobulins. Hum Genet 80:207–218

Miyake K, Horikawa Y, Hara K, Yasuda K, Osawa H, Furuta H, Hirota Y, Yamagata K, Hinokio Y, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Yamamoto K, Tokunaga K, Takeda J, Makino H, Nanjo K, Kadowaki T, Kasuga M (2008) Association of TCF7L2 polymorphisms with susceptibility to type 2 diabetes in 4,087 Japanese subjects. J Hum Genet 53:174–180

Moritani M, Nomura K, Tanahashi T, Osabe D, Fujita Y, Shinohara S, Yamaguchi Y, Keshavarz P, Kudo E, Nakamura N, Yoshikawa T, Ichiishi E, Takata Y, Yasui N, Shiota H, Kunika K, Inoue H, Itakura M (2007) Genetic association of single nucleotide polymorphisms in endonuclease G-like 1 gene with type 2 diabetes in a Japanese population. Diabetologia 50:1218–1227

Ng MC, Tam CH, Lam VK, So WY, Ma RC, Chan JC (2007) Replication and identification of novel variants at TCF7L2 associated with type 2 diabetes in Hong Kong Chinese. J Clin Endocrinol Metab 92:3733–3737

Rebbeck TR, Spitz M, Wu X (2004) Assessing the function of genetic variants in candidate gene association studies. Nat Rev Genet 5:589–597

Ren Q, Han XY, Wang F, Zhang XY, Han LC, Luo YY, Zhou XH, Ji LN (2008) Exon sequencing and association analysis of polymorphisms in TCF7L2 with type 2 diabetes in a Chinese population. Diabetologia 51(7):1146–1152

Riancho JA, Valero C, Naranjo A, Morales DJ, Sanudo C, Zarrabeitia MT (2007) Identification of an aromatase haplotype that is associated with gene expression and postmenopausal osteoporosis. J Clin Endocrinol Metab 92:660–665

Sakamoto Y, Inoue H, Keshavarz P, Miyawaki K, Yamaguchi Y, Moritani M, Kunika K, Nakamura N, Yoshikawa T, Yasui N, Shiota H, Tanahashi T, Itakura M (2007) SNPs in the KCNJ11-ABCC8 gene locus are associated with type 2 diabetes and blood pressure levels in the Japanese population. J Hum Genet 52:781–793

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345

Shiina H, Igawa M, Breault J, Ribeiro-Filho L, Pookot D, Urakami S, Terashima M, Deguchi M, Yamanaka M, Shirai M, Kaneuchi M, Kane CJ, Dahiya R (2003) The human T-cell factor–4 gene splicing isoforms, Wnt signal pathway, and apoptosis in renal cell carcinoma. Clin Cancer Res 9:2121–2132

Shu L, Sauter NS, Schulthess FT, Matveyenko AV, Oberholzer J, Maedler K (2008) Transcription factor 7-like 2 regulates beta-cell survival and function in human pancreatic islets. Diabetes 57:645–653

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885

Tanahashi T, Osabe D, Nomura K, Shinohara S, Kato H, Ichiishi E, Nakamura N, Yoshikawa T, Takata Y, Miyamoto T, Shiota H, Keshavarz P, Yamaguchi Y, Kunika K, Moritani M, Inoue H, Itakura M (2006) Association study on chromosome 20q11.21-13.13 locus and its contribution to type 2 diabetes susceptibility in Japanese. Hum Genet 120:527–542

The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678

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

Weedon MN (2007) The importance of TCF7L2. Diabet Med 24:1062–1066

World Health Organization (1985) Diabetes mellitus: report of a WHO study group. WHO, Geneva

Yi F, Brubaker PL, Jin T (2005) TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3 beta. J Biol Chem 280:1457–1464

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341