博士論文

Identification of *KCNQ1* as a susceptibility gene
for type 2 diabetes mellitus

（２型糖尿病感受性遺伝子としての *KCNQ1* の同定）

安田　和基

Abstract

Although many genes have been reported to be associated with type 2 diabetes

mellitus using a candidate gene approach, most of these results were not

sufficiently convincing. Therefore, with 1,612 diabetic patients and 1,424

non-diabetic control subjects in total, I conducted a multistage genome-wide

association study (GWAS) of type 2 diabetes mellitus in the Japanese population

Among the 100,000 single nucleotide polymorphisms (SNPs) used for screening,

the most significantly associated was rs2237895 located in intron 15 of *KCNQ1* (P

= 3.4×10^{-6}).　I performed dense mapping within *KCNQ1*, and found that

rs2237892, another SNP in intron 15, exhibited the lowest P value of 6.7×10^{-13},

which showed genome-wide significance, and the odds ratio of 1.49 (95%

confidence interval (CI) of 1.34 to 1.66). The association of *KCNQ1* with type 2

diabetes mellitus was confirmed in two independent Japanese panels. In

collaboration it was also reproduced in Asian panels of Korean and Chinese

population, and surprisingly in a Caucasian panel. A meta-analysis was then

performed combing 19,930 individuals (9,569 cases and 10,361 controls) in total,

and rs2237892 was significantly associated with the disease with a P value of 1.7×10^{-42} and the odds ratio of 1.40 (95% CI = 1.34 to 1.47). Among the non-diabetic subjects in the Japanese panels, the risk allele of rs2237892 was related to impaired insulin secretion but not with insulin resistance. These data thus demonstrate *KCNQ1* as a novel diabetes susceptibility gene in multiple ethnic groups.

Introduction

Diabetes mellitus is clinically characterized by chronic hyperglycemia and associated complications, and mainly results from impaired insulin secretion, insulin action or both. In Japan, the prevalence of type 2 diabetes mellitus is on the rapid increase, and it affects more than 10% of subjects over 40 years of age. While diabetes in Caucasians is often associated with obesity and insulin resistance, relatively few diabetic individuals are obese in Japan, and impaired insulin secretion often develops before diabetes onset [1].

Genetic factors are critical for the development of diabetes mellitus. Some forms of

diabetes are caused by mutations in a single gene; in contrast, type 2 diabetes, the

major form of the disease, is a multifactorial disorder. Most previous genetic studies of

polygenic diseases were conducted using the candidate gene approach to investigate

the association between genetic variants in genes known to be involved in the disease

pathogenesis. For type 2 diabetes, more than 100 genes related to insulin secretion,

insulin resistance and energy metabolism have been assessed in case-control studies.

Although several genes such as *PPARG* [2] were found to be associated with diabetes

mellitus in a reproducible manner, most genes investigated exhibited only marginal

effects or even controversial results; thus, the candidate gene approach for type 2

diabetes does not provide convincing results and does not reveal the genetic details of

the disease. Other methods such as sib-pair linkage analysis have been applied in

genetic studies of type 2 diabetes, and *CAPN10* on chromosome 20 was suggested to

be related to diabetes [3]. However, the genetic structure of type 2 diabetes mellitus

remains largely unclear, and is referred to as a "geneticists' nightmare".

In 2007, genome-wide association studies (GWAS) were developed to evaluate

polygenic diseases. This method was used to study diabetes mellitus [4, 5], and several

novel genes such as *TCF7L2* and *CDKAL1* were reported to be associated with

diabetes in different ethnic panels. Although the results were reproducible and

convincing, all the reports were from Caucasians. Thus, it was unclear whether similar

results would be achieved in East Asians.

With this background, a national project aiming to identify disease-associated single

nucleotide polymorphisms (SNPs) for common diseases in Japan was launched and

designated the Millennium Genome Project in Japan. I conducted a multistage GWAS

of type 2 diabetes mellitus in this project. 100,000 SNPs deposited in JSNP database

(http://snp.ims.u-tokyo.ac.jp) [6] were utilized for screening the genome and the scan

was referred to as the JSNP Genome Scan (JGS). I was completely responsible for the

project from its initial study design to the final data analysis.

Methods and materials

Study samples

I assembled three independent sample panels with collaborators in eleven core study

groups in Japan. Panel 1 (for stage1) consisted of 188 diabetic patients, panel 2 (for

stage 2) included 752 diabetics and 752 controls, and panel 3 (for stage3) included 672

diabetics and 672 controls. Since it was decided to target common type 2 diabetes, the

inclusion criteria for diabetic cases were determined as follows : (1) the age of diabetes

onset was between 40 and 55 in order to recruit common type diabetes and to exclude

diabetic patients with a single gene mutation who are often misdiagnosed as type 2

diabetes , (2) maximum BMI of should be below 30 kg/m^2 , to exclude obesity-related

diabetes, (3) insulin treatment had not been started until three years after diagnosis, to

exclude slowly progressive IDDM(insulin-dependent diabetes mellitus), and (4)

antibodies to glutamic acid decarboxylase (GAD) were negative, to exclude type 1

diabetes. For panel 2 and panel3, the criteria for the non-diabetic controls panels 2 and

3 were : (1) the age was above 60, (2) without past history of diabetes diagnosis, and

(3) HbA$_{1c}$ (hemoglobin A$_{1c}$) was below 6%. About 5 to 10 mL of blood was drawn and

genomic DNA was purified using standard methods. Current and maximum BMI,

family history of diabetes and biochemical data of blood were also obtained. Insulin

secretion and insulin senseitivity were assessed by homeostasis model assessment

(HOMA) model. The HOMA-β (β-cell function) and HOMA-IR (insulin resistance)

were calculated as [FIRI (pmol/l) × 20]/{[FPG (mmol/l) − 3.5] × 6} and [FPG (mmol/l)

× FIRI (pmol/l)]/(22.5 × 6), respectively.

Additional replication panels were recruited by my collaborators with basically similar

inclusion criteria. Replication 1 (1,521 cases and 1,544 controls) and Replication 2

(1,433 cases and 1,444 controls) are Japanese panels. Replication 3 (Han Chinese from

Hong Kong) consisted of 1416 diabetic patients (including 626 early-onset diabetes)

and 1,577 controls. Replication 4 (Korean) comprised of 758 diabetic patents and 632

Table 1. Clinical characteristics of subject panels.

| | Panel 1 | | Panel 2 | | Panel 3 | |
	Diabetes		Diabetes	Control	Diabetes	Control
n	187		752	752	672	672
Male participants (%)	59.4		57.8	46.3	57.9	38.7
Age at study (years)	63.2 ± 9.1		62.4 ± 8.9	69.2 ± 6.9	62.6 ± 9.1	70.2 ± 6.4
BMI (kg/m^2)	22.9 ± 2.8		23.3 ± 3.0	22.9 ± 2.8	23.1 ± 3.0	22.5 ± 3.0
HbA$_{1c}$ (%)	7.2 ± 1.3		7.5 ± 1.5	5.0 ± 0.3	7.6 ± 1.5	5.0 ± 0.3

| | Replication 1 (Japanese) | | Replication 2 (Japanese) | |
	Diabetes	Control	Diabetes	Control
n	1521	1544	1433	1444
Male participants (%)	57.9	46.4	51.2	49.1
Age at study (years)	60.2 ± 10.5	65.6 ± 9.9	63.2 ± 10.4	39.9 ± 14.9
BMI (kg/m^2)	24.1 ± 3.9	23.7 ± 3.5	23.6 ± 3.4	22.1 ± 2.9
HbA$_{1c}$ (%)	7.7 ± 2.4	5.0 ± 0.3	7.3 ± 1.4	4.8 ± 0.4

| | Replication 3 (Chinese) | | Replication 4 (Korean) | | Replication 5 (Caucasian) | |
	Diabetes	Control	Diabetes	Control	Diabetes	Control
n	1416	1577	758	632	2830	3740
Male participants (%)	40.4	46.1	46.7	45.4	58.9	37.9
Age at study (years)	50.0 ± 13.7	25.1 ± 14.2	59.2 ± 9.9	64.7 ± 3.6	57.9 ± 11.5	57.4 ± 6.0
BMI (kg/m^2)	25.0 ± 4.0	21.0 ± 3.8	24.4 ± 2.9	23.5 ± 3.1	29.6 ± 5.5	25.1 ± 3.6
HbA$_{1c}$ (%)	8.0 ± 2.1		8.1 ± 1.6	5.3 ± 0.3	7.3 ± 1.9	4.7 ± 0.4

non-diabetic controls. Replication 5 (Caucasian) was recruited from Sweden and

contained 2,830 cases and 3,740 controls. The general features of each panel are shown

in Table 1.

The protocol was approved by the ethics committee of each institution; the study

number in National Center for Global Health and Medicine was D0001. Written

informed consent was obtained from all subjects.

Study design

The multistage GWAS in JGS was designed as collaboration of five diseases, and the

grand design (Figure 1), has been described previously [7]. Briefly, 188 patients for

each disease at the first stage (panel 1 for diabetes) were recruited for genotyping of

100,000 SNPs deposited in the JSNP database [6]. The 187 diabetic subjects were then

compared with two different sets of control groups. One was referred to as the "JSNP

database control" (JDC), which was the reference frequencies of SNP alleles derived

from 752 Japanese subjects. The other was the "other disease group" (ODG), which was

allele or genotype frequencies obtained from four other disease groups (gastric cancer,

hypertension, Alzheimer's dementia and asthma) combined in the first stage of JGS.
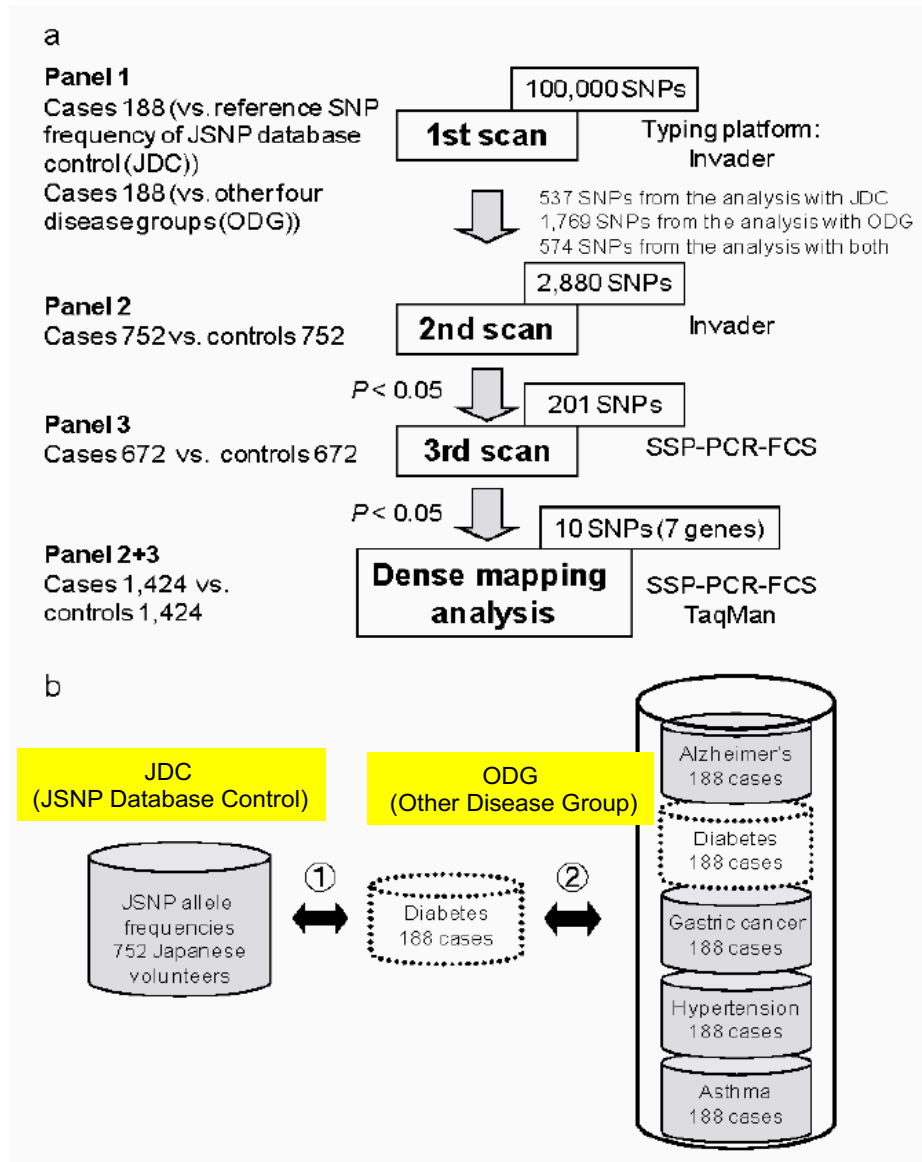
Figure 1. Design of the JSNP Genome Scan (JGS) for type 2 diabetes
a, In the first stage of the JGS, 188 cases for type 2 diabetes were genotyped for 100,000 SNPs by multiplex PCR-based Invader analysis. Two additional rounds of screening followed by dense mapping analysis were then performed as indicated. **b**, The SNP typing in the first stage of the JGS was followed by two case-reference association analyses with either reference allele frequency in the JSNP database(JDC) or allele (or genotype) frequency in the other four disease groups combined(ODG).

By two association studies, 2,880 SNPs were chosen in the order of their *P* value for the

next stage.

Panel 2 consisting of cases and controls was typed in the second stage, and 201 SNPs

positive for association ($P < 0.05$) were selected. In the third stage, another independent

panel (panel 3) was investigated, and for the 10 SNPs positive for association ($P < 0.05$),

I combined the data of panel 2 and panel 3 (panel 2+3) and reassessed the association

with 1,424 cases and 1,424 controls. My collaborators checked for stratification in panel

2 and panel 3 by typing of another 28 'neutral' SNPs followed by several methods

including principal component analysis (PCA), the chi-square test , and STRUCTURE

analysis (http://pritch.bsd.uchicago.edu/software.html), and they reported to me that

there were no stratification suggested between cases and controls(data not shown).

Dense SNP mapping for *KCNQ1*

I picked up 49 SNPs of *KCNQ1* from dbSNP (NCBI), and the interval was around ~10

kbp on average. I then typed them together with the three positive SNPs in the third

stage in panel 2+3. For comprehensive identification of genetic variation in Japanese,

my collaborators re-sequenced all the exons, a putative promotor region (up to 4 kbp

upstream from the transcription start site), and 47-kbp region including intron 15 of

KCNQ1 in twenty-four Japanese people. Ten SNPs were selected on the basis of LD

and MAF (>10%), and these 10 SNPs and the two novel nonsynonymous variants were

genotyped in panel 2+3. In total, 64 SNPs in *KCNQ1* were examined in panel 2+3, and

18 SNPs of them were located between rs151290 and rs2237895.

Typing methods

In the first and second round of scan, genotyping was performed using the multiplex

polymerase chain reaction (PCR)-based Invader assay (Third Wave Technologies,

Madison, WI) as described previously [8]. In the third scan and for dense mapping and

replication panels, genotyping was performed either by sequence-specific primer

(SSP)–PCR analysis followed by fluorescence correlation spectroscopy (FCS) [9] by

my collaborators, or real-time PCR analysis using TaqMan probes (Applied

Biosystems, Foster City, CA).

Statistical analysis and *in silico* analysis

As described, I evaluated two case-control associations in the first screening. For

comparison with ODG, I examined allele or genotype data in 2 × 2 contingency tables

using dominant or recessive models, and for comparison with JDC where genotype

data were unavailable, only allele data were used. In the later stages, allele data were

analyzed using the chi-square test. LD was analyzed with my collaborators using

Haploview 3.31 software. A *P* value of <0.05 was considered statistically significant.

Genotype-based analyses were also performed for 10 positive SNPs in the third scan,

rs2237892 and rs2074196 using the Cochran-Armitage trend test. For meta-analysis,

the Mantel-Haenszel method was applied based on fixed-effects models with the "meta"

package of the R-Project (http://www.r-project.org) with my collaborators. For

analyzing the predicted transcription factor binding sites, SNPInspector in Genomatix

software suite (München, Germany) was used according to the company's instruction.

Results

Identification of *KCNQ1* as a potential susceptibility gene for type 2 diabetes

mellitus by multistage genome-wide association study

In the first screening of the study, data for 82,343 autosomal polymorphisms out of

100,000 SNPs genotyped passed the typing quality control in 187 diabetic subjects; no

genotype call was yielded for one subject in panel1.   The collaborators working on

other diseases reported to me that they no significant population stratification was

detected among the five first-stage panels by standard methods; for genomic control

analysis 1,025 SNPs were selected and inflation factor was 1.06. The genotype-based

analysis was performed with dominant and recessive models. Firstly, I selected SNPs

of following features: 1)MAF was higher than 10% and 2) either an allele OR higher

than 1.3 or a genotype OR of higher than 1.5 in either association analysis. Only one

SNP was chosen among the SNPs in strong LD ($r^2 > 0.9$). By combining the two types

of association analysis I selected 2,880 SNPs for the second screening; 1,111 SNPs

were from comparison with JDC and 2,343 SNPs were recruited by the analyses with

ODG.

In the second stage, an independent panel (panel 2) was assessed by multiplex Invader

analysis. No results were obtained for 38SNPs at all in panel 2, and five and three

SNPs did not work for the case or the control groups, respectively I could not annotate

seven probes on the human genome. As a result, 2,827 SNPs generated valid data, and

their call rate was 0.993. In total, 201 SNPs positive for association ($P < 0.05$) were

selected based on the allelic data (Table 2).

Table 2: The 201 SNPs selected for the third screening

JSNP ID	dbSNP ID (rs#)	A1	A2	Chr	Gene symbol	*P* value
IMS-JST049162	1884455	T	C	1	*TMEM52*	0.0259
IMS-JST007804	2071983	A	G	1	*PER3*	0.0309
IMS-JST057570	228669	T	C	1	*PER3*	0.0466
IMS-JST021870	2268176	A	C	1	*H6PD*	0.0426
IMS-JST003807	2075931	A	G	1		0.0382
IMS-JST031168	2282231	A	G	1	*MACF1*	0.0342
IMS-JST050273	2296172	A	G	1	*MACF1*	0.02
IMS-JST050275	2296173	A	G	1	*MACF1*	0.033
IMS-JST050277	2296175	A	G	1	*MACF1*	0.017
IMS-JST084449	3738676	T	G	1	*BMP8A*	0.0317
IMS-JST122936	3768320	T	C	1	*PABPC4*	0.0033
IMS-JST014761	2242500	A	G	1	*PABPC4*	0.0471
IMS-JST056578	2300743	A	C	1	*BMP8B*	0.0386
IMS-JST120706	3766321	G	C	1	*SCMH1*	0.0394
IMS-JST148096	3790857	A	G	1	*PGM1*	0.0045
IMS-JST147894	3790684	T	C	1	*NTNG1*	0.0344
IMS-JST107445	3754446	T	G	1	*GSTM5*	0.0032
IMS-JST031268	367199	A	C	1	*LRRC52*	0.0294
IMS-JST031269	2252342	A	G	1	*LRRC52*	0.0188
IMS-JST003629	2067079	A	G	1		0.0431
IMS-JST006299	1042295	T	A	1	*MYBPH*	0.0266
IMS-JST084501	3738722	T	C	1	*ITPKB*	0.0292
IMS-JST084907	3739076	A	C	2	*OTOF*	0.0165

IMS-JST105793	3752916	T	C	2	*SPAST*	0.0219
IMS-JST100718	3748926	A	G	2		0.0315
IMS-JST125211	3770291	A	G	2	*SRBD1*	0.0064
IMS-JST002601	2075171	T	C	2		0.008
IMS-JST126994	3771868	A	C	2	*TACR1*	0.0423
IMS-JST067763	2272535	A	G	2	*FMNL2*	0.0155
IMS-JST079923	3214040	A	G	2	*GALNT5*	0.0078
IMS-JST059133	2228171*	A	G	2	*LRP2**	0.0365
IMS-JST059129	2302693	A	G	2	*LRP2*	0.0068
IMS-JST189945	3828435	C	G	3	*ITPR1*	0.0114
IMS-JST164054	3804992	A	G	3	*ITPR1*	0.0168
IMS-JST043167	2290972	A	G	3	*HHATL*	0.0097
IMS-JST066554	2271615	G	C	3	*SLC6A20*	0.0039
IMS-JST127991	3772759	T	C	3	*KALRN*	0.0085
IMS-JST108765	3755676	G	C	3	*KALRN*	0.0232
IMS-JST108767	3755678	T	C	3	*KALRN*	0.0061
IMS-JST035528	2245285	C	G	3	*PLXND1*	0.0212
IMS-JST129109	3773759	A	G	3		0.0421
IMS-JST046313	2293252	A	G	3	*MRAS*	0.022
IMS-JST034819	2284843	T	C	3	*DGKG*	0.0282
IMS-JST034818	2284842	A	G	3	*DGKG*	0.0422
IMS-JST127578	1016226	A	G	3	*OPA1*	0.0085
IMS-JST081099	3736651	A	T	3	*TFRC*	0.0287
IMS-JST064111	2306475	A	G	4	*CTBP1*	0.026
IMS-JST131594	3775936	C	G	4	*WDR1*	0.0125
IMS-JST013349	2241474	T	C	4	*WDR1*	0.0082
IMS-JST154793	3796822	G	C	4	*WDR1*	0.0115
IMS-JST154792	3796821	A	G	4	*WDR1*	0.0118
IMS-JST109359	3756225	G	C	4	*WDR1*	0.0102
IMS-JST183130	3822239	T	A	4	*WDR1*	0.0097
IMS-JST013357	13441	T	C	4	*WDR1*	0.01
IMS-JST109355	3756222	A	G	4		0.0108

IMS-JST064519	2306802	T	C	4	SMARCAD1	0.027
IMS-JST063967	2069763	T	G	4	IL2	0.0253
IMS-JST073768	2277005	T	C	5	AYTL2	0.0062
IMS-JST087505	476569	A	G	5	C9	0.0278
IMS-JST170668	230755	C	G	5	SKIV2L2	0.0214
IMS-JST059490	2302979	G	C	5	CMYA5	0.0438
IMS-JST171837	2617280	T	G	5	SNCAIP	0.0173
IMS-JST115385	34012	A	G	5	FGF1	0.0174
IMS-JST109533	245821	T	C	5	ARHGAP26	0.0067
IMS-JST024450	2270068	A	C	5	ARHGAP26	0.0355
IMS-JST158738	3800311	A	G	6		0.0291
IMS-JST004143	13408	T	C	6	ZKSCAN4	0.0381
IMS-JST050811	2233647	A	G	6	SPDEF	0.0472
IMS-JST088492	3734687	T	C	6	EGFL9	0.0086
IMS-JST088493	3734688	A	G	6	EGFL9	0.0055
IMS-JST173652	1041523	A	G	6	TMEM63B	0.0399
IMS-JST157238	3798976	A	T	6	BAI3	0.0007
IMS-JST172130	3812141	A	G	6	PRSS35	0.0389
IMS-JST009469	380496	A	G	6		0.013
IMS-JST030819	2281996	A	G	6	SNX9	0.0456
IMS-JST004796	2076573	T	C	6	SNX9	0.0444
IMS-JST015963	2235842	A	G	6	SNX9	0.0437
IMS-JST015941	2235823	T	C	6		0.0128
IMS-JST159553	3801034	T	G	7	ZNF12	0.0066
IMS-JST021272	1541516	A	G	7	ADCYAP1R1	0.0153
IMS-JST088939	3735089	T	G	7	AMPH	0.0187
IMS-JST057858	2301706	T	C	7		0.0356
IMS-JST024358	1990107	T	C	7	SEMA3D	0.0051
IMS-JST135637	3779554	A	G	7	PFTK1	0.0295
IMS-JST027974	2280000	T	C	8	ERICH1	0.0425
IMS-JST066245	2271387	A	C	8	ADAM28	0.0363
IMS-JST047402	2294042	T	C	8	C8orf77	0.0385

IMS-JST050114	734549	A	G	9	*KIAA1432*	0.04
IMS-JST051875	2297349	T	C	9		0.0314
IMS-JST170752	3810920	A	G	9	*WDR31*	0.0357
IMS-JST069940	2274158	T	G	9	*DFNB31*	0.0341
IMS-JST069942	766835	A	G	9	*DFNB31*	0.0448
IMS-JST069944	2274161	T	C	9	*DFNB31*	0.0196
IMS-JST059294	2302827	T	C	9	*ASTN2*	0.0484
IMS-JST052119	2297537	G	C	9	*EGFL7*	0.0015
IMS-JST045854	6605	A	G	9	*NPDC1*	0.0446
IMS-JST063876	2306294	T	C	10	*PFKP*	0.0233
IMS-JST117606	3763788	T	C	10	*PRPF18*	0.0451
IMS-JST014075	2242009	T	C	10	*LOC648701*	0.0172
IMS-JST137402	3781141	A	G	10	*KCNMA1*	0.0153
IMS-JST066780	2271786	T	C	10	*SFTPA2*	0.0033
IMS-JST060356	1713330	T	A	10		0.0271
IMS-JST111814	3758505	T	G	10		0.0046
IMS-JST058710	2302373	A	G	10	*HABP2*	0.0473
IMS-JST137873	3781565	T	C	10	*PNLIPRP1*	0.0467
IMS-JST137869	3781561	T	C	10	*PNLIPRP2*	0.0304
IMS-JST037438	2286779	G	C	10	*PNLIPRP2*	0.0043
IMS-JST104546	3752408	T	G	10	*PNLIPRP2*	0.0043
IMS-JST057145	2301179	T	C	10	*PNLIPRP2*	0.0058
IMS-JST049882	2295873	T	C	10	*TACC2*	0.0155
IMS-JST049887	2295878	A	G	10	*TACC2*	0.0472
IMS-JST029068	9440	T	C	11	*HCCA2*	0.0434
IMS-JST027214	2279478	C	G	11	*DUSP8*	0.0427
IMS-JST003717	2075873	A	G	11	*KCNQ1*	0.001
IMS-JST018590	151290	A	C	11	*KCNQ1*	7.37×10^{-5}
IMS-JST018601	163184	T	G	11	*KCNQ1*	0.0064
IMS-JST018602	2237895	A	C	11	*KCNQ1*	1.37×10^{-7}
IMS-JST007966	7167	T	C	11	*CCDC86*	0.0305
IMS-JST000791	174473	T	C	11	*RAB3IL1*	0.026

IMS-JST072368	2275998	A	G	11	ACTN3	0.0208
IMS-JST026713	2279124	G	C	11	ALDH3B2	0.0346
IMS-JST053885	620080	A	G	11	CASP4	0.0339
IMS-JST072366	2275997	T	C	11	CADM1	0.0085
IMS-JST034199	2284396	T	C	12	BCL2L14/LRP6	0.0092
IMS-JST021788	2268123	T	C	12	GRIN2B	0.0483
IMS-JST092942	3741872	T	C	12	FAM60A	0.0024
IMS-JST064834	2307027	T	C	12	KRT4	0.0311
IMS-JST043044	2290894	C	G	12	NACA	0.0485
IMS-JST138610	2958149	A	G	12	NACA	0.0326
IMS-JST138611	2926747	T	A	12	NACA	0.0372
IMS-JST079952	3214051	T	C	12	NACA	0.0431
IMS-JST018939	2238153	T	C	12	ATXN2	0.0164
IMS-JST063348	1628639	A	G	12	DTX1	0.0237
IMS-JST063352	1674101	A	G	12	RASAL1	0.0163
IMS-JST027689	515746	A	G	12	TBX3	0.0116
IMS-JST010253	1179445	A	T	12	MSI1	0.0439
IMS-JST070246	2274400	T	C	13	RCBTB2	0.0103
IMS-JST030323	2281678	T	C	14	SLC7A7	0.0384
IMS-JST140012	3783635	T	C	14	CNIH	0.0294
IMS-JST032893	1520332	T	C	14	RGS6	0.0275
IMS-JST032892	1520331	A	G	14	RGS6	0.0316
IMS-JST010357	2239275	C	G	14	RGS6	0.0494
IMS-JST078876	3213729	G	C	14	WDR21A	0.0329
IMS-JST013547	2241622	T	C	14	STON2	0.0076
IMS-JST030870	2149647	T	C	14	C14orf102	0.0107
IMS-JST005764	5224	A	G	14	BDKRB2	0.0341
IMS-JST005982	2070916	T	A	15	FMN1	0.0433
IMS-JST064933	2250402	A	C	15	EIF2AK4	0.0353
IMS-JST061974	9325	A	T	15	ALDH1A2	0.0339
IMS-JST141104	3784610	T	G	15	RORA	0.0329
IMS-JST094435	3743210	T	C	15	C15orf39	0.0246

IMS-JST028440	2280306	G	C	15	*ZNF291*	0.0383
IMS-JST011194	2239858	A	G	15	*PCSK6*	0.0095
IMS-JST140996	756559	T	C	15	*PCSK6*	0.0475
IMS-JST011595	2240143	T	C	16	*SRRM2*	0.0176
IMS-JST188942	3827528	T	C	16	*N-PAC/ROGDI*	0.0121
IMS-JST099282	3747613	G	C	16	*UBN1*	0.0024
IMS-JST044867	2251666	A	G	16	*UBN1*	0.0149
IMS-JST003298	1049212	T	C	16	*PPL*	0.0087
IMS-JST141811	3785264	A	G	16	*A2BP1*	0.0456
IMS-JST141777	3785233	A	C	16	*A2BP1*	0.0231
IMS-JST141732	3785191	A	G	16	*A2BP1*	0.0054
IMS-JST078582	3213646	T	C	16	*EXOD1*	0.0306
IMS-JST013992	2241955	T	C	16	*GNAO1*	0.0319
IMS-JST063083	2305698	T	C	16	*CPNE2*	0.02
IMS-JST162444	3803593	T	C	16	*KATNB1*	0.049
IMS-JST006051	2000999	A	G	16	*HPR*	0.0221
IMS-JST026347	10395	T	C	16		0.0034
IMS-JST044529	2291963	T	C	16	*KIAA0513*	0.0384
IMS-JST042509	2290505	A	G	17	*LLGL1*	0.0132
IMS-JST178518	3817992	T	G	17	*TOP3A*	0.0103
IMS-JST048568	2294913	A	G	17	*TOP3A*	0.0146
IMS-JST068410	2273029	T	C	17	*SHMT1*	0.0168
IMS-JST169500	3809797	A	G	17	*NEK8*	0.0308
IMS-JST113955	550510	A	G	17	*CALCOCO2*	0.034
IMS-JST065141	1994970	A	G	17	*SNF8*	0.0265
IMS-JST044204	2291725	T	C	17	*GIP*	0.0238
IMS-JST044205	2291726	T	C	17	*GIP*	0.0157
IMS-JST095369	3744061	A	G	17	*MFSD11 / SFRS2*	0.0229
IMS-JST068758	1048591	G	C	17		0.0347
IMS-JST025722	8887	A	G	19	*KIAA1881/HDGF2*	0.0465
IMS-JST043425	730291	A	G	19	*JMJD2B*	0.0252
IMS-JST188066	3826740	A	G	19	*ELAVL1*	0.0281

IMS-JST118565	16006	T	C	19	*CACNA1A*	0.0023
IMS-JST040327	2288933	A	G	19	*TIMM50*	0.0096
IMS-JST039026	2002926	A	G	19	*CEACAM5*	0.0161
IMS-JST026395	2278902	T	G	19		0.0067
IMS-JST114219	2547337	T	A	19		0.0487
IMS-JST061536	36615	T	C	19	*RPS9*	0.0485
IMS-JST144420	574628	A	G	20	*ANGPT4*	0.0366
IMS-JST021496	362988	A	G	20	*SNAP25*	0.0049
IMS-JST116298	789434	T	C	20		0.0195
IMS-JST015913	11274	A	G	20	*C20orf111*	0.0127
IMS-JST015915	731498	T	C	20	*C20orf111*	0.0273
IMS-JST152860	2780231	T	C	20	*ZNF334*	0.0084
IMS-JST012770	2241030	A	C	22		0.0373
IMS-JST011340	2239959	A	G	22	*AIFM3*	0.0338
IMS-JST072336	179468	T	G	22	*C22orf16*	0.0432
IMS-JST007170	2071621	A	G	22	*NF2*	0.0139
IMS-JST029513	2246092	A	G	22	*TMPRSS6*	0.0247
IMS-JST075214	3213555	A	G	22	*PSCD4*	0.0009

P values were calculated for allele data in panel 2.

Positions and gene symbols are derived from the dbSNP database (build 127) ,

UCSC Genome Browser, or from the JSNP database.

A1 and A2 represent alleles 1 and 2; Chr, chromosome.

At the third stage, one SNP failed to be typed at all and the call rate for the other 200

SNPs exhibited the call rate of 0.990 by SSP-PCR-FCS analysis. Ten SNPs yielded a *P*

value of <0.05 (Table 3). The most significant association was obtained with

rs2237895 ($P = 3.4 \times 10^{-6}$), which resides in intron 15 of *KCNQ1* gene on chromosome

11. Interestingly, another two SNPs in the same intron (rs151290 and rs163184) were

Table3　10 Positive SNPs identified in the third screening

| dbSNP ID | Risk allele | Chr | Gene | Panel 1 (187 cases) | | | | |
				RAF(DM)	RAF(NC)	OR (95% CI)	*P value*	Control
rs151290	C	11	*KCNQ1*	0.63	0.57	1.30 (1.03–1.65)	0.027	ODG
rs163184	G	11	*KCNQ1*	0.51	0.43	1.33 (1.06–1.67)	0.015	JDC
rs2237895	C	11	*KCNQ1*	0.45	0.35	1.53 (1.22–1.93)	2.8×10^{-4}	JDC
rs2250402	C	15	*EIF2AK4*	0.2	0.27	1.45 (1.09–1.93)	0.011	JDC
rs2307027	C	12	*KRT4*	0.14	0.22	1.68 (1.20–2.36)	0.0024	ODG
rs3741872	C	12	*FAM60A*	0.29	0.23	1.37 (1.06–1.76)	0.015	ODG
rs574628	G	20	*ANGPT4*	0.56	0.64	1.38 (1.09–1.74)	0.0066	ODG
rs2233647	G	6	*SPDEF*	0.92	0.86	1.87 (1.07–3.27)	0.026	ODG
rs3785233[a]	C	16	*A2BP1*	0.2	0.17	1.20 (0.90–1.61)	0.22	ODG
rs2075931	A	1		0.71	0.64	1.37 (1.07–1.75)	0.013	ODG

| dbSNP ID | Risk allele | Chr | Gene | Panel 2 (752 cases, 752 controls) | | | |
				RAF(DM)	RAF(NC)	OR (95% CI)	*P value*
rs151290	C	11	*KCNQ1*	0.62	0.55	1.34 (1.16–1.55)	7.4×10^{-5}
rs163184	G	11	*KCNQ1*	0.49	0.44	1.22 (1.06–1.41)	0.0064
rs2237895	C	11	*KCNQ1*	0.42	0.33	1.49 (1.28–1.73)	1.4×10^{-7}
rs2250402	C	15	*EIF2AK4*	0.24	0.21	1.20 (1.01–1.43)	0.035
rs2307027	C	12	*KRT4*	0.2	0.17	1.23 (1.02–1.47)	0.031
rs3741872	C	12	*FAM60A*	0.29	0.24	1.29 (1.09–1.52)	0.0024
rs574628	G	20	*ANGPT4*	0.65	0.61	1.17 (1.01–1.36)	0.037
rs2233647	G	6	*SPDEF*	0.88	0.86	1.24 (1.00–1.54)	0.047
rs3785233[a]	C	16	*A2BP1*	0.19	0.16	1.25 (1.03–1.51)	0.023
rs2075931	A	1		0.68	0.65	1.17 (1.01–1.37)	0.038

| dbSNP ID | Risk allele | Chr | Gene | Panel 3 (672 cases, 672 controls) | | | |
				RAF(DM)	RAF(NC)	OR (95% CI)	*P value*
rs151290	C	11	*KCNQ1*	0.61	0.54	1.36 (1.16–1.58)	1.1×10^{-4}
rs163184	G	11	*KCNQ1*	0.48	0.42	1.27 (1.09–1.48)	0.0021
rs2237895	C	11	*KCNQ1*	0.42	0.33	1.45 (1.24–1.70)	3.4×10^{-6}
rs2250402	C	15	*EIF2AK4*	0.26	0.21	1.34 (1.11–1.60)	0.0018
rs2307027	C	12	*KRT4*	0.21	0.16	1.37 (1.12–1.67)	0.0017
rs3741872	C	12	*FAM60A*	0.28	0.23	1.28 (1.07–1.52)	0.006
rs574628	G	20	*ANGPT4*	0.64	0.59	1.28 (1.10–1.50)	0.0018
rs2233647	G	6	*SPDEF*	0.89	0.86	1.29 (1.02–1.62)	0.033
rs3785233[a]	C	16	*A2BP1*	0.19	0.16	1.23 (1.01–1.50)	0.039
rs2075931	A	1		0.68	0.64	1.18 (1.00–1.38)	0.048

P values were calculated for allele data.
For panel 1, two control groups (ODG, other disease group; JDC, Japanese database control) were used for association studies and the lower P values are listed.
Chr, chromosome. RAF(DM) and RAF(NC), risk allele frequencies in cases and controls, respectively. OR, odds ratio for risk allele.
aThis SNP was selected for the second stage based on the recessive model (OR = 2.59, CI = 1.20–5.58, P = 0.012)

also included in the list. When two panels were analyzed in combination (panel 2+3),

all of the 10 SNPs yielded even lower *P* values. The genotype-based trend test gave

similar results (data not shown).

Dense SNP mapping for *KCNQ1*

The standard criterion of genome-wide significance previously proposed for GWASs

was *P* value of $<5 \times 10^{-7}$ [10]. Since *KCNQ1* was the only gene that closely reached

this level, I decided to further analyze this gene. The three SNPs in *KCNQ1* identified

at the third scan were in moderate LD with each other (Figure 2).
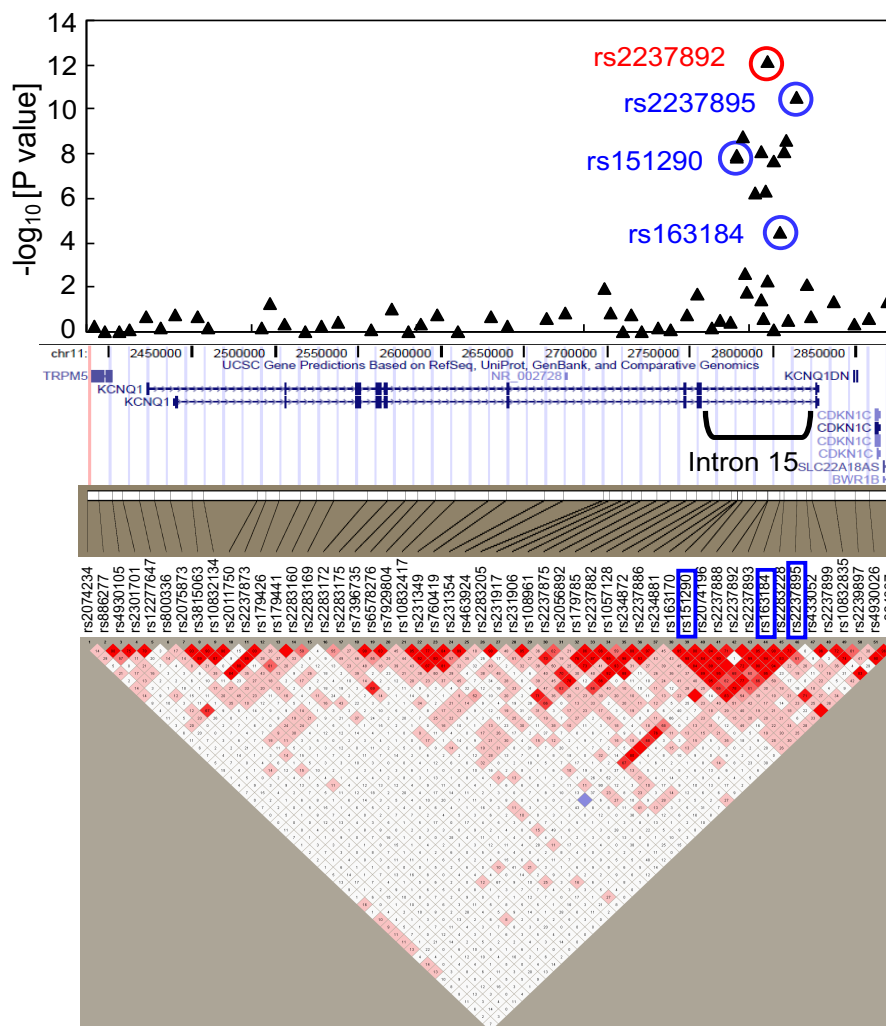
•**Figure 2.** Dense mapping analysis of *KCNQ1*. **The top panel** shows the association $-\log_{10}[P$ value] in panel 2+3 for 64 SNPs of *KCNQ1*. The three blue circles represent the positive SNPs in the third screening. The red circle (rs2237892) indicates the SNP showing the most significant association with type 2 diabetes. **The upper middle panel** shows the physical position of *KCNQ1* and neighboring genes on chromosome 11. **The lower middle panel** shows the positions and rs numbers of the 52 previously identified SNPs. Blue rectangles indicate the positive SNPs in the third screening. **The bottom panel** shows a Haploview representation of LD (*D'*) based on genotyping data from control subjects in panel 2+3 (*n* = 1424).

For rs2237895 and rs163184, D' and r^2 values were 0.83 and 0.46, respectively. When

I genotyped these three SNPs and the additional 49 SNPs picked up from NCBI dbSNP

in panel 2+3 (Figure 2), rs2237892, another SNP located in intron 15, exhibited the

strongest association with the disease ($P = 6.7 \times 10^{-13}$), and an odds ratio (OR) was

1.49 (95% confidence interval (CI) was 1.34 to 1.66) (Table 4). The D' and r^2 values

for rs2237895 and rs2237892 were 0.95 and 0.30, respectively.

Re-sequencing in 24 Japanese subjects of *KCNQ1* gene including the 47-kbp genomic

region corresponding to intron 15 revealed 212 variations. I then typed 10 variants in

intron 15 and the two nonsynonymous variants (P448R and G643S) in panel 2+3, but

none of them exceeded rs2237892 in association (data not shown).

Replication studies and meta-analysis of the association of *KCNQ1* with type 2

diabetes mellitus

  Since the three SNPs, rs2237892, rs2237895, and rs2074196 in *KCNQ1* showed the

strongest association with diabetes. I next asked my collaborators to test the

association in several independent panels. I received the typing data and performed

association analysis. In the two new Japanese panels, replication panel 1 and 2, a

strong association with diabetes was reproduced (Table 4); for example, the allelic P

values of rs2237892 were 9.6×10^{-10} in the replication 1 panel and 6.9×10^{-10} in the

replication 2 panel. When I combined these data with panel 2+3, 4,378 cases and 4,412

controls in total were analyzed and an allelic P value for rs2237892 reached 2.8×10^{-29}

and an OR was 1.43 (95% CI = 1.34 to 1.52).

Table 4 Association study summary for the three SNPs in KCNQ1 and type 2 diabetes

SNP ID	Risk allele	Panel	RAF(DM)	RAF(NC)	P_{allele}	OR	95% CI		P_{trend}	Meta-analysis OR (95% CI), P value
rs2074196	G	2+3 (dense mapping)	0.63	0.55	1.7×10^{-9}	1.39	1.25	1.54	1.8×10^{-9}	
		Replication 1 (Japanese)	0.61	0.54	1.4×10^{-7}	1.32	1.19	1.46	2.1×10^{-7}	
		Replication 2 (Japanese)	0.62	0.55	4.7×10^{-7}	1.31	1.18	1.46	6.2×10^{-7}	
		All Japanese (4378 cases, 4412 controls)	**0.62**	**0.55**	**4.6×10^{-21}**	**1.34**	**1.26**	**1.42**	**9.8×10^{-21}**	**1.34 (1.26–1.42),** *P = 4.8 ×10⁻²¹*
		Replication 3 (Chinese)	0.71	0.63	1.2×10^{-9}	1.4	1.26	1.56	9.8×10^{-10}	
		Replication 4 (Korean)	0.66	0.58	3.0×10^{-5}	1.39	1.19	1.62	2.1×10^{-5}	
		All Asian (6552 cases, 6621 controls)	**0.64**	**0.57**	**9.9×10^{-32}**	**1.35**	**1.28**	**1.42**	**2.1×10^{-31}**	**1.36 (1.29–1.42),** *P = 7.9 ×10⁻³³*
		Replication 5 (Caucasian)	0.96	0.95	0.017	1.23	1.04	1.46	0.017	
		All	NA	NA	NA	NA	NA	NA	NA	**1.35 (1.28–1.41),** *P = 8.6 ×10⁻³⁴*
rs2237892	C	2+3 (dense mapping)	0.69	0.6	6.7×10^{-13}	1.49	1.34	1.66	1.7×10^{-12}	
		Replication 1 (Japanese)	0.66	0.59	9.6×10^{-10}	1.39	1.25	1.54	1.6×10^{-9}	
		Replication 2 (Japanese)	0.68	0.6	6.9×10^{-10}	1.41	1.26	1.57	1.1×10^{-9}	
		All Japanese (4378 cases, 4412 controls)	**0.68**	**0.59**	**2.8×10^{-29}**	**1.43**	**1.34**	**1.52**	**1.7×10^{-28}**	**1.43 (1.34–1.52),** *P = 3.0 ×10⁻²⁹*
		Replication 3 (Chinese)	0.72	0.65	1.3×10^{-8}	1.38	1.24	1.55	4.2×10^{-9}	
		Replication 4 (Korean)	0.69	0.61	1.7×10^{-5}	1.41	1.21	1.65	1.0×10^{-5}	
		All Asian (6552 cases, 6621 controls)	**0.69**	**0.61**	**2.0×10^{-39}**	**1.41**	**1.34**	**1.48**	**2.5×10^{-39}**	**1.42 (1.34–1.49),** *P = 2.5 ×10⁻⁴⁰*
		Replication 5 (Caucasian)	0.95	0.93	7.8×10^{-4}	1.29	1.11	1.5	7.2×10^{-4}	
		All	NA	NA	NA	NA	NA	NA	NA	**1.40 (1.34–1.47),** *P = 1.7 ×10⁻⁴²*
rs2237895	C	2+3 (dense mapping)	0.41	0.33	3.1×10^{-11}	1.44	1.3	1.61	4.0×10^{-11}	
		Replication 1 (Japanese)	0.38	0.33	4.5×10^{-5}	1.25	1.12	1.38	4.7×10^{-5}	
		Replication 2 (Japanese)	0.41	0.34	5.8×10^{-8}	1.35	1.21	1.5	5.5×10^{-8}	
		All Japanese (4378 cases, 4412 controls)	**0.4**	**0.33**	**1.3×10^{-20}**	**1.34**	**1.26**	**1.43**	**1.7×10^{-20}**	**1.34 (1.26–1.43),** *P = 1.4 ×10⁻²⁰*
		Replication 3 (Chinese)	0.4	0.34	3.5×10^{-5}	1.25	1.12	1.39	3.4×10^{-5}	
		Replication 4 (Korean)	0.35	0.3	3.2×10^{-3}	1.27	1.08	1.49	2.7×10^{-3}	
		All Asian (6552 cases, 6621 controls)	**0.39**	**0.33**	**2.7×10^{-25}**	**1.31**	**1.24**	**1.38**	**2.7×10^{-25}**	**1.31 (1.25–1.38),** *P = 6.1 ×10⁻²⁶*
		Replication 5 (Caucasian)	NA	NA	NA	NA	NA	NA	NA	

RAF(DM) and RAF(NC), risk allele frequencies in cases and controls, respectively.
Pallele values were calculated for allele data. OR, odds ratio for risk allele.
Ptrend values were calculated by the Cochran-Armitage trend test. Meta-analysis was performed by the Mantel-Haenszel method (fixed-effects models). NA, not applicable.

The association was also confirmed by my collaborators in East Asia. The allelic P

value for rs2237892 in the replication 3 panel (Han Chinese) was 1.3×10^{-8} and in

replication 4 panel (Korean) was 1.7×10^{-5} (Table 4). Then I performed a

meta-analysis of the Asian panels for the association of rs2237892 with diabetes, and a

P value was 2.5×10^{-40} and an OR was 1.42 (95% CI = 1.34 to 1.49). My collaborators

in North Europe also genotyped rs2237892 and rs2074196 in their panel (replication 5

panel), and found both SNPs showed a significant association in their populations ($P =$

7.8×10^{-4} and 0.017, respectively).

When I re-conducted a meta-analysis adding panel 5, a total number of the subjects

were 19,930 (9,569 cases and 10,361 controls) and the association for rs2237892 was

further confirmed; the P value now reached 1.7×10^{-42} and an OR was 1.40 (95% CI =

1.34 to 1.47) (Table 4, Figure 3).
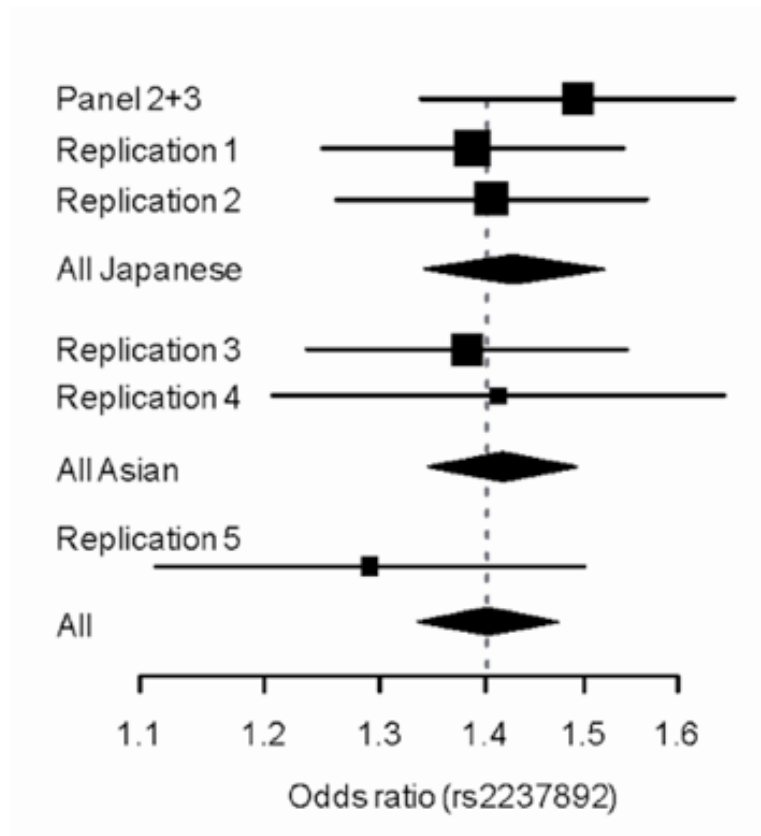
Figure 3: Association of rs2237892 of *KCNQ1* with type 2 diabetes. Allelic odds ratios and 95% CI (confidence intervals) for the indicated sample panels are indicated by squares and horizontal lines. Odds ratios and 95% CI obtained by meta-analysis, as performed by the Mantel-Haenszel method (fixed-effects models), are indicated by diamonds.

The impact of *KCNQ1* SNP on the clinical phenotypes and potential transcription

binding sites

I next investigated the relation of rs2237892 with clinical phenotypes in panel 2+3.

Among 1,424 cases, this SNP was not apparently associated with body mass index

(BMI). Fasting plasma glucose and insulin levels were available with 948 non-diabetic

control subjects, and CC homozygotes (C for the risk allele) exhibited a significantly

lower HOMA-β compared with those with the other genotypes (Table 5). This result

suggested that the risk allele of *KCNQ1* might increase the risk of diabetes through

impaired secretion of insulin.

Table5 The relation of rs2237892 of KCNQ1 to clinical phenotype
Clinical characteristics according to genotype for rs2237892 of KCNQ1
(Japanese controls in panel 2+3)

	CC ($n = 353$)	CT + TT ($n = 595$)	P	Adjusted P
BMI (kg/m^2)	22.3 ± 2.8	22.7 ± 2.9	0.035	0.16[a]
HOMA-IR	1.50 ± 1.03	1.57 ± 1.00	0.13	0.41[b]
HOMA-β	81.7 ± 57.9	94.3 ± 84.3	0.0024	0.021[b]

Data are means ± SD. [a]*P* value after adjustment for sex and age.
  [b]*P* value after adjustment for sex, age, and BMI.

Finally, I investigated whether the risk alleles of rs2237892 and rs2237895 might

change the predicted transcription factor (TF) binding motifs. By SNPInspector, each

risk allele either generated or extinguished several TF binding sites (Table 6), and

among them there were some binding motifs for TFs which may have some functional

implications in pancreatic β cells, such as PTF1 (pancreas transcription factor 1) or

IA-1 (insulinoma-associated 1).

Table 6 Changes in predicted transcription factor (TF) binding motifs by the risk alleles of rs2237892 and rs2237895.
SNPInspector in Genomatix software suite was used.

Seq. name	Non-risk to risk allele	lost/new	Family/matrix	Further Information	Matrix similarity
rs2237892	T -> C	lost	V$CTCF/CTCFL.01	CCCTC-binding factor (zinc finger protein)-like (BORIS)	0.821
rs2237892	T -> C	lost	V$ZF02/ZNF300.01	KRAB-containing zinc finger protein 300	0.993
rs2237892	T -> C	lost	V$EBOX/USF.04	Upstream stimulating factor 1/2	0.916
rs2237892	T -> C	lost	V$KLFS/EKLF.01	Erythroid krueppel like factor (EKLF)	0.901
rs2237892	T -> C	lost	V$SREB/SREBP.04	Sterol regulatory element binding transcription factor	0.961
rs2237892	T -> C	new	V$HICF/HIC1.01	Hypermethylated in cancer 1	0.918
rs2237892	T -> C	new	V$SNAP/PSE.01	Proximal sequence element (PSE) of RNA polymerase II-transcribed snRNA genes	0.776
rs2237892	T -> C	new	V$ZTRE/ZTRE.01	ZTRE motifs (1 bp spacer), ZNF658 binding site	0.86
rs2237895	A -> C	lost	V$RXRF/RAR_RXR.01	Retinoic acid receptor / retinoid X receptor heterodimer, DR1 sites	0.78
rs2237895	A -> C	lost	V$NOLF/EBF1.01	Early B-cell factor 1	0.89
rs2237895	A -> C	lost	V$RORA/RORA.02	RAR-related orphan receptor alpha, homodimer DR5 binding site	0.751
rs2237895	A -> C	lost	V$INSM/INSM1.01	Zinc finger protein insulinoma-associated 1 (IA-1)	0.907
rs2237895	A -> C	new	V$GCF2/LRRFIP1.01	Leucine rich repeat (in FLII) interacting protein 1	0.845
rs2237895	A -> C	new	V$PTF1/PTF1.01	PTF1 (Pancreas transcription factor 1) binding sites	0.775

Discussion

Although this study started with a small number of subjects for GWAS, the multistage

strategy for GWASs is expected to be effective to reduce false positive associations

and has proven much successful [11]. Indeed, I was able to identify *KCNQ1* as a

susceptibility gene for type 2 diabetes in the JGS. *KCNQ1* is located at chromosome

11p15.5, and it is relatively close to 11p13-p12 region which was linked to type 2

diabetes in two reports of affected Japanese sib-pair analyses [12,13]. *KCNQ1* was also

associated with diabetes in two Asian panels, and it was thus established that *KCNQ1*

confers susceptibility to type 2 diabetes in East Asians. I further confirmed the

significant association in Caucasians. Considering two early GWAS reports with

Caucasians [4, 5] failed to identify *KCNQ1* as a diabetes gene, I decided to examine

the available data sets with my collaborators. Within the LD block of *KCNQ1* showing

association with diabetes in Japanese, 11 SNPs in the WTCCC data set [4] and nine

SNPs in the DGI data set [5], respectively, had been genotyped in the projects, but the

association was only marginal if any. This apparent discrepancy may be mainly

attributable to the allele frequencies of the causative SNPs (for example, in East Asians

the MAF of rs2237892 was 0.28 to 0.41 but in Caucasians as low as 0.05 to 0.07).

Indeed, when I looked into a large dataset of recent meta-analysis of three big GWASs

(WTCCC, DGI and FUSION: http://www.well.ox.ac.uk/DIAGRAM),[10] the risk alleles

in *KCNQ1* were now associated with diabetes in Caucasians ($P = 0.01$ for rs2237892

and $P = 0.012$ for rs2074196). These results not only provide further support that

KCNQ1 is associated with diabetes in multiple ethnic groups, and but also demonstrate the need to extend GWASs to every population.

KCNQ1 encodes a subunit of a voltage-gated potassium channel (KvLQT1); this channel plays a pivotal role in regulating the membrane potential in the repolarization phase in heart. It was already reported that mutations in this gene can cause long QT syndrome (for example, Romano-Ward syndrome [14]) and familial atrial fibrillation. The expression of this potassium is also reported in a variety of tissues including the pancreas. The KCNQ1 protein is expressed in insulin-secreting INS-1 cells, and seems functional since in the presence of tolbutamide, a KCNQ1 blocker augmented insulin secretion [15]. The lower HOMA-β in the homozygotes of rs2237892 risk allele among Japanese may be interesting in this sense.

However, as the susceptibility variants are located in introns, there are several other possibilities. Alternative splicing can generate several variants for many kinds of ion channels, and multiple splice variants *KCNQ1* mRNA were also reported; it is not known whether the identified SNPs in intron 15 can modulate the alternative splicing pattern in human pancreatic islets. It is also interesting to speculate that the SNPs

conferring susceptibility of diabetes may affect the expression of other causative genes

located in the genome close to *KCNQ1*. By *in silico* assay, I found several transcription

factor (TF) binding sites to be generated or disrupted by SNPs (Table6), and some of

these TFs are expected to play functional roles in pancreatic β cells. My lab also

screened the proteins in a pancreatic β cell line that differentially bind to the alleles in

the SNP region using magnetic nanobeads, and reported several candidate proteins

with several SNPs [16]. The genomic region including *KCNQ1* is genetically imprinted,

and my colleagues published a paper describing that genetic manipulation of

imprinting in this locus in mice may lead to impaired insulin secretion by increased

expression of a nearby gene Cdkn1c [17]. I did not find any microRNA harboring

rs2237892 in the miRBase database (http://microrna.sanger.ac.uk/sequemces). Further

studies are necessary to elucidate the molecular mechanism how the risk allele of

KCNQ1 confers susceptibility to diabetes.

I may have missed a substantial number of susceptibility genes in our screening, since

the strategy adopted in 2002 lacks sufficient analytical power [7] compared with

currently available technologies. In this sense, the detection power of the first two

stages for JGS had been previously simulated by my collaborators [7]. If risk SNPs had

an odds ratio of 1.5 and a risk genotype frequency was 30%, detection sensitivity will

be expected as approximately 13%. Thus, *KCNQ1* turned out to be a good (and maybe

the only) target of this kind of strategy for type 2 diabetes in Japanese.

Several comprehensive studies based on new platforms for GWASs were published

just before this report, with approximately 10 genes reproducibly found to be

associated with type 2 diabetes in Caucasians [18-21] . None of the these genes

showed a positive association in the JGS typing data. Given that some of these genes

conferred susceptibility to diabetes in Japanese subjects [22, 23], the lack of

association in this study may be related to the limited sample size of the first scan or to

the weak LD between the SNPs we used and the causative variants.

When this study was published [24] in 2008, an independent GWAS from Japanese

was also reported [25] at the same time. Surprisingly, the other report also identified

SNPs in *KCNQ1* intron 15 with very similar effect sizes, strongly suggesting that

KCNQ1 is a susceptibility gene for type 2 diabetes in Japanese. Since then, numerous

studies supporting the association of *KCNQ1* with type 2 diabetes mellitus were

reported in both Asians and Caucasians. *KCNQ1* remains the most relevant

susceptibility gene for type 2 diabetes in most studies in Asians. Several genetic

prediction models of diabetes with multiple SNPs including *KCNQ1* variants were

constructed, and subtle but distinct prediction powers were demonstrated [26]. My

colleagues and I also found that *KCNQ1* is associated in an increased risk of diabetes

incidence in a prospective cohort [27]. The variants are also reported to be associated

with other related disorders such as diabetic complications and responses to

anti-diabetic drugs including sulfonylureas and DPP4 inhibitors. Thus, the role of

KCNQ1 as the major diabetes susceptibility gene has been established.

In genetic studies of type 2 diabetes, this study can be regarded as a kind of landmark

paper for several reasons.

1) This is the first report of a GWAS to examine type 2 diabetes in a non-European

   population, which identified the novel susceptibility gene *KCNQ1* ; this gene was

   not previously identified in GWAS of Caucasians.

2) The association of *KCNQ1* was reproduced not only in East Asians but also in

   Caucasians. Thus, *KCNQ1* is a diabetes susceptibility gene in multiple ethnic

groups.

3) The main reason that *KCNQ1* was not detected in previous GWAS is its lower minor allele frequency (MAF) in Caucasians. Interestingly, in Caucasians *TCF7L2* is the most important diabetes gene and the MAF of risk SNPs is approximately 30%, while that Japanese people is only 5%. This may be considered as a mirror image of the case of *KCNQ1*. Thus, even trans-ethnic susceptibility genes with substantial effect sizes can be identified only by GWAS in specific populations, and this study clearly revealed the value of conducting GWAS for each ethnic group. This has important implication for genetic studies of various common diseases. This may also partly explain the ethnic difference in the genetic structure of the same disease.

4) Finally, although *KCNQ1* is the strongest diabetes gene detected in Asians, the effect size of the variants is still modest; the odds ratio was approximately 1.4. This implies a universal rule for which the most common variants are associated with common diseases.

Since this study was published, genetic studies of type 2 diabetes mellitus by GWAS

have shown numerous important rules. GWAS is based on the 'common disease common variant hypothesis' and has been extended mainly in two directions; one is to increase the number of SNP markers and density of coverage, while the other is to include a much larger number of subjects. The former was made possible by improved platform for microarrays and an imputation method which takes the linkage disequilibrium information of the human genome into consideration. The other was achieved by incorporating and combining many GWAS data by meta-analysis. Genes with smaller effect sizes or more ethnicity-specific factors have been identified, and approximately 100 diabetes susceptibility genes have been reported. I had the opportunity to join several big GWAS projects after this study [28-30]. However, even when all the GWAS-identified SNPs are combined, only a limited fraction of genetic backgrounds for type 2 diabetes can be explained (some estimates were approximately 15%), and "missing heritability" has attracted increasing attention in genetic studies of common diseases. This has led to the search for rare variants with large effect, and whole exome sequencing (WES) and whole genome sequencing (WGS) are now being performed extensively to this end. The interaction between genetic and environmental

factors and the involvement of epigenetic modification are of great interest, and will become major research targets in the future. In any case, the advantages and disadvantages of GWAS are now clearly recognized, and this study may play an important role in these advances.

In summary, I conducted a multistage GWAS for type 2 diabetes in Japanese and have identified *KCNQ1* as a susceptibility gene, and confirmed association not only in East Asians but also in Caucasians. These findings may provide new insight into the pathophysiology of diabetes, and this study also elucidated the usefulness of GWAS performed in different ethnic groups.

References

1.Fukushima M, Suzuki H, Seino Y. Insulin secretion capacity in the development from normal glucose tolerance to type 2 diabetes. *Diabetes Res. Clin. Pract.* **66**, S37–S43, 2004.

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

3. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Horikawa Y, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet.* **26**:163-75, 2000.

4. 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; Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT. Replication of genome-wide association signals in U.K. samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341, 2007.

5. 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 Boström K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Råstam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjögren 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. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336, 2007.

6. Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T. Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. Single-nucleotide polymorphism. *J. Hum. Genet.* **47**, 605–610, 2002.

7. Sato Y, Suganami H, Hamada C, Yoshimura I, Yoshida T, Yoshimura K. Designing a multistage, SNP-based, genome screen for common diseases. *J. Hum. Genet.* **49**, 669–676, 2004.

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

9. Bannai M, Higuchi K, Akesaka T, Furukawa M, Yamaoka M, Sato K, Tokunaga K. Single-nucleotide-polymorphism genotyping for whole-genome-amplified samples using automated fluorescence correlation spectroscopy. *Anal. Biochem.* **327**, 215–221, 2004.

10. NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature* **447**, 655–660, 2007.

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

12. Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecoeur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P. Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* **51**, 1247–1255, 2002.

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

14. Wang Q, Curran ME, Splawski I, Burn TC, Millholland JM, VanRaay TJ, Shen J, Timothy KW, Vincent GM, de Jager T, Schwartz PJ, Toubin JA, Moss AJ, Atkinson DL, Landes GM, Connors TD, Keating MT. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nat. Genet.* **12**, 17–23, 1996.

15. Ullrich S, Su J, Ranta F, Wittekindt OH, Ris F, Rösler M, Gerlach U, Heitzmann D, Warth R, Lang F. Effects of I_{Ks} channel inhibitors in insulin-secreting INS-1 cells. *Pflugers Arch.* **451**, 428–436, 2005.

16. Hiramoto M, Udagawa H, Watanabe A, Miyazawa K, Ishibashi N, Kawaguchi M, Uebanso T, Nishimura W, Nammo T, Yasuda K. Comparative analysis of type 2 diabetes-associated SNP alleles identifies allele-specific DNA-binding proteins for the *KCNQ1* locus. *Int J Mol Med.* **36**:222-30, 2015.

17. Asahara S, Etoh H, Inoue H, Teruyama K, Shibutani Y, Ihara Y, Kawada Y, Bartolome A, Hashimoto N, Matsuda T, Koyanagi-Kimura M, Kanno A, Hirota Y, Hosooka T, Nagashima K, Nishimura W, Inoue H, Matsumoto M, Higgins MJ, Yasuda K, Inagaki N, Seino S, Kasuga M, Kido Y. Paternal allelic mutation at the Kcnq1 locus

reduces pancreatic β-cell mass by epigenetic modification of Cdkn1c. *Proc Natl Acad Sci U S A*. **112**:8332-7, 2015.

18. 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. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345, 2007.

19. 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. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885, 2007.

20. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **9**, 770–775, 2007.

21. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Boström KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jørgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM,

Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA,

Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen

E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A,

Shields B, Sjögren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G,

Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM,

Weedon MN, Willer CJ; Wellcome Trust Case Control Consortium, Illig T, Hveem K,

Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT,

Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of

genome-wide association data and large-scale replication identifies additional

susceptibility loci for type 2 diabetes. *Nat*. *Genet*. **40**:638-45, 2008.

22. 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. Association of TCF7L2 polymorphisms with susceptibility to

type 2 diabetes in 4,087 Japanese subjects. *J*. *Hum*. *Genet*. **53**, 174–180, 2008.

23. Horikawa Y, Miyake K, Yasuda K, Enya M, 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, Kasuga M. Replication of genome-wide

association studies of type 2 diabetes susceptibility in Japan. *J. Clin. Endocrinol.*

Metab. **93**:3136-41, 2008.

24. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H,

Jonsson A, Sato Y, Yamagata K, Hinokio Y, Wang HY, Tanahashi T, Nakamura N,

Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A,

Takeda J, Maeda E, Shin HD, Cho YM, Park KS, Lee HK, Ng MC, Ma RC, So WY,

Chan JC, Lyssenko V, Tuomi T, Nilsson P, Groop L, Kamatani N, Sekine A,

Nakamura Y, Yamamoto K, Yoshida T, Tokunaga K, Itakura M, Makino H, Nanjo K,

Kadowaki T, Kasuga M. Variants in KCNQ1 are associated with susceptibility to type

2 diabetes mellitus. *Nat Genet.* **40**:1092-7, 2008.

25. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, Ng DP,

Holmkvist J, Borch-Johnsen K, Jørgensen T, Sandbaek A, Lauritzen T, Hansen T,

Nurbaya S, Tsunoda T, Kubo M, Babazono T, Hirose H, Hayashi M, Iwamoto Y,

Kashiwagi A, Kaku K, Kawamori R, Tai ES, Pedersen O, Kamatani N, Kadowaki T,

Kikkawa R, Nakamura Y, Maeda S. SNPs in KCNQ1 are associated with susceptibility

to type 2 diabetes in East Asian and European populations. *Nat Genet*. **40**:1098-1102,

2008.

26. Imamura M, Shigemizu D, Tsunoda T, Iwata M, Maegawa H, Watada H, Hirose H,

Tanaka Y, Tobe K, Kaku K, Kashiwagi A, Kawamori R, Maeda S. Assessing the

clinical utility of a genetic risk score constructed using 49 susceptibility alleles for type

2 diabetes in a Japanese population.J Clin Endocrinol Metab. 98(10):E1667-73,2013

27. Goto A, Noda M, Goto M, Yasuda K, Mizoue T, Yamaji T, Sawada N, Iwasaki M,

Inoue M, Tsugane S; JPHC Study Group. Predictive performance of a genetic risk

score using 11 susceptibility alleles for the incidence of Type 2 diabetes in a general

Japanese population: a nested case-control study. *Diabet Med*. **35**:602-611, 2018

28. Imamura M, Maeda S, Yamauchi T, Hara K, Yasuda K, Morizono T, Takahashi A,

Horikoshi M, Nakamura M, Fujita H, Tsunoda T, Kubo M, Watada H, Maegawa H,

Okada-Iwabu M, Iwabu M, Shojima N, Ohshige T, Omori S, Iwata M, Hirose H, Kaku

K, Ito C, Tanaka Y, Tobe K, Kashiwagi A, Kawamori R, Kasuga M, Kamatani

N; Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium,

Nakamura Y, Kadowaki T. A single-nucleotide polymorphism in ANK1 is associated

with susceptibility to type 2 diabetes in Japanese populations. *Hum Mol Genet*.

21:3042-9, 2012.

29. Hara K, Fujita H, Johnson TA, Yamauchi T, Yasuda K, Horikoshi M, Peng C, Hu

C, Ma RC, ImamuraM, Iwata M, Tsunoda T, Morizono T, Shojima N, So WY, Leung

TF, Kwan P, Zhang R, Wang J, Yu W, Maegawa H, Hirose H; DIAGRAM consortium,

Kaku K, Ito C, Watada H, Tanaka Y, Tobe K, Kashiwagi A, Kawamori R, Jia W, Chan

JC, Teo YY, Shyong TE, Kamatani N, Kubo M, Maeda S, Kadowaki T. Genome-wide

association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet*.

23:239-46, 2014.

30. Imamura M, Takahashi A, Yamauchi T, Hara K, Yasuda K, Grarup N, Zhao W,

Wang X, Huerta-Chagoya A, Hu C, Moon S, Long J, Kwak SH, Rasheed A, Saxena R,

Ma RC, Okada Y, Iwata M, Hosoe J, Shojima N, Iwasaki M, Fujita H, Suzuki K,

Danesh J, Jørgensen T, Jørgensen ME, Witte DR, Brandslund I, Christensen C, Hansen

T, Mercader JM, Flannick J, Moreno-Macías H, Burtt NP, Zhang R, Kim YJ, Zheng W,

Singh JR, Tam CH, Hirose H, Maegawa H, Ito C, Kaku K, Watada H, Tanaka Y, Tobe

K, Kawamori R, Kubo M, Cho YS, Chan JC, Sanghera D, Frossard P, Park KS, Shu

XO, Kim BJ, Florez JC, Tusié-Luna T, Jia W, Tai ES, Pedersen O, Saleheen

D, Maeda S, Kadowaki T. Genome-wide association studies in the Japanese population

identify seven novel loci for type 2 diabetes. *Nat Commun*.**7**:10531, 2016.