nature genetics

Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus

Kazuki Yasuda[1], Kazuaki Miyake[2], Yukio Horikawa[3], Kazuo Hara[4], Haruhiko Osawa[5], Hiroto Furuta[6], Yushi Hirota[2], Hiroyuki Mori[2], Anna Jonsson[7], Yoshifumi Sato[8], Kazuya Yamagata[8,27], Yoshinori Hinokio[9], He-Yao Wang[1,27], Toshihito Tanahashi[10], Naoto Nakamura[11], Yoshitomo Oka[9], Naoko Iwasaki[12], Yasuhiko Iwamoto[12], Yuichiro Yamada[13,27], Yutaka Seino[13,27], Hiroshi Maegawa[14], Atsunori Kashiwagi[14], Jun Takeda[3], Eiichi Maeda[15], Hyoung Doo Shin[16], Young Min Cho[17], Kyong Soo Park[17], Hong Kyu Lee[17], Maggie C Y Ng[18], Ronald C W Ma[18], Wing-Yee So[18], Juliana C N Chan[18], Valeriya Lyssenko[7], Tiinamaija Tuomi[19,20], Peter Nilsson[21], Leif Groop[7,19], Naoyuki Kamatani[22], Akihiro Sekine[23,27], Yusuke Nakamura[23], Ken Yamamoto[24], Teruhiko Yoshida[25], Katsushi Tokunaga[26], Mitsuo Itakura[10], Hideichi Makino[5], Kishio Nanjo[6], Takashi Kadowaki[4] & Masato Kasuga[2]

We carried out a multistage genome-wide association study of type 2 diabetes mellitus in Japanese individuals, with a total of 1,612 cases and 1,424 controls and 100,000 SNPs. The most significant association was obtained with SNPs in *KCNQ1*, and dense mapping within the gene revealed that rs2237892 in intron 15 showed the lowest *P* value (6.7×10^{-13}, odds ratio (OR) = 1.49). The association of *KCNQ1* with type 2 diabetes was replicated in populations of Korean, Chinese and European ancestry as well as in two independent Japanese populations, and meta-analysis with a total of 19,930 individuals (9,569 cases and 10,361 controls) yielded a *P* value of 1.7×10^{-42} (OR = 1.40; 95% CI = 1.34–1.47) for rs2237892. Among control subjects, the risk allele of this polymorphism was associated with impairment of insulin secretion according to the homeostasis model assessment of β-cell function or the corrected insulin response. Our data thus implicate *KCNQ1* as a diabetes susceptibility gene in groups of different ancestries.

In Japan, the prevalence of type 2 diabetes mellitus is increasing rapidly, and more than 10% of individuals over 40 years of age are affected. Relatively few diabetic individuals in Japan are obese, and impairment of insulin secretion often develops before the onset of diabetes[1]. As part of a national project designated the Millennium Genome Project in Japan, in 2002 we began a multistage genome-wide association study (GWAS) to identify disease-associated SNPs for type 2 diabetes mellitus using 100,000 SNPs from a collection of

[1]Department of Metabolic Disorder, Research Institute, International Medical Center of Japan, Tokyo 162-8655, Japan. [2]Division of Diabetes, Metabolism, and Endocrinology, Department of Internal Medicine, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan. [3]Department of Diabetes and Endocrinology, Division of Molecule and Structure, Gifu University School of Medicine, Gifu 501-1194, Japan. [4]Department of Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan. [5]Department of Molecular and Genetic Medicine, Ehime University Graduate School of Medicine, Ehime 791-0295, Japan. [6]First Department of Medicine, Wakayama Medical University, Wakayama 641-8509, Japan. [7]Department of Clinical Sciences, Diabetes and Endocrinology, University Hospital Malmö, Lund University, S-205 02 Malmö, Sweden. [8]Department of Metabolic Medicine, Graduate School of Medicine, Osaka University, Osaka 565-0871, Japan. [9]Division of Molecular Metabolism and Diabetes, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan. [10]Division of Genetic Information, Institute for Genome Research, University of Tokushima, Tokushima 770-8503, Japan. [11]Department of Endocrinology and Metabolism, Kyoto Prefectural University of Medicine, Graduate School of Medical Sciences, Kyoto 602-8566, Japan. [12]Department of Medicine, Diabetes Center, Tokyo Women's Medical University, Tokyo 162-8666, Japan. [13]Department of Diabetes and Clinical Nutrition, Kyoto University School of Medicine, Kyoto 606-8501, Japan. [14]Division of Endocrinology and Metabolism, Department of Medicine, Shiga University of Medical Science, Shiga 520-2192, Japan. [15]Clinical Genome Informatics Center, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan. [16]Department of Genetic Epidemiology, SNP Genetics Inc., Seoul 110-834, Korea. [17]Department of Internal Medicine, Seoul National University College of Medicine, Seoul 110-744, Korea. [18]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Shatin, Hong Kong. [19]Department of Medicine, Helsinki University Hospital, FIN-00300 Helsinki, Finland. [20]Folkhaelsan Research Center, FIN-00014 Helsinki, Finland. [21]Department of Clinical Sciences, Medicine Research Unit, University Hospital Malmö, Lund University, S-205 02 Malmö, Sweden. [22]Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo 162-8666, Japan. [23]SNP Research Center, Institute of Physical and Chemical Research (RIKEN), Yokohama 230-0045, Japan. [24]Department of Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan. [25]Genetics Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan. [26]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo 113-0033, Japan. [27]Present addresses: Department of Medical Biochemistry, Faculty of Medical and Pharmaceutical Sciences, Kumamoto University, Kumamoto 860-8556, Japan (K. Yamagata), Shanghai Institute of Materia Medica, Chinese Academy of Science, Shanghai 200031, China (H.-Y.W.), Department of Internal Medicine, Akita University School of Medicine, Akita 010-8543, Japan (Y.Y.), Kansai Electric Power Hospital, Osaka 553-0003, Japan (Y. Seino) and Genome Informatics, Center for Genomic Medicine, Graduate School of Medicine and Faculty of Medicine, Kyoto University, Kyoto 606-8501, Japan (A.S.). Correspondence should be addressed to M.K. (kasuga@med.kobe-u.ac.jp).

Received 3 March; accepted 6 June; published online 17 August 2008; doi:10.1038/ng.207

Table 1 Positive SNPs identified in the third screening

					Panel 1 (187 cases)			Panel 2 (752 cases, 752 controls)				Panel 3 (672 cases, 672 controls)			
dbSNP ID	Risk allele	Chr.	Gene	RAF(DM)	RAF(NC)	OR (95% CI)	*P* value	Control	RAF(DM)	RAF(NC)	OR (95% CI)	*P* value	RAF(DM)	RAF(NC)	OR (95% CI)	*P* value
rs151290	C	11	*KCNQ1*	0.63	0.57	1.30 (1.03–1.65)	0.027	ODG	0.62	0.55	1.34 (1.16–1.55)	7.4×10^{-5}	0.61	0.54	1.36 (1.16–1.58)	1.1×10^{-4}
rs163184	G	11	*KCNQ1*	0.51	0.43	1.33 (1.06–1.67)	0.015	JDC	0.49	0.44	1.22 (1.06–1.41)	0.0064	0.48	0.42	1.27 (1.09–1.48)	0.0021
rs2237895	C	11	*KCNQ1*	0.45	0.35	1.53 (1.22–1.93)	2.8×10^{-4}	JDC	0.42	0.33	1.49 (1.28–1.73)	1.4×10^{-7}	0.42	0.33	1.45 (1.24–1.70)	3.4×10^{-6}
rs2250402	C	15	*EIF2AK4*	0.20	0.27	1.45 (1.09–1.93)	0.011	JDC	0.24	0.21	1.20 (1.01–1.43)	0.035	0.26	0.21	1.34 (1.11–1.60)	0.0018
rs2307027	C	12	*KRT4*	0.14	0.22	1.68 (1.20–2.36)	0.0024	ODG	0.20	0.17	1.23 (1.02–1.47)	0.031	0.21	0.16	1.37 (1.12–1.67)	0.0017
rs3741872	C	12	*FAM60A*	0.29	0.23	1.37 (1.06–1.76)	0.015	ODG	0.29	0.24	1.29 (1.09–1.52)	0.0024	0.28	0.23	1.28 (1.07–1.52)	0.0060
rs574628	G	20	*ANGPT4*	0.56	0.64	1.38 (1.09–1.74)	0.0066	ODG	0.65	0.61	1.17 (1.01–1.36)	0.037	0.64	0.59	1.28 (1.10–1.50)	0.0018
rs2233647	G	6	*SPDEF*	0.92	0.86	1.87 (1.07–3.27)	0.026	ODG	0.88	0.86	1.24 (1.00–1.54)	0.047	0.89	0.86	1.29 (1.02–1.62)	0.033
rs3785233[a]	C	16	*A2BP1*	0.20	0.17	1.20 (0.90–1.61)	0.22	ODG	0.19	0.16	1.25 (1.03–1.51)	0.023	0.19	0.16	1.23 (1.01–1.50)	0.039
rs2075931	A	1		0.71	0.64	1.37 (1.07–1.75)	0.013	ODG	0.68	0.65	1.17 (1.01–1.37)	0.038	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. RAF(DM) and RAF(NC), risk allele frequencies in cases and controls, respectively; OR, odds ratio for risk allele.
[a]This SNP was selected for the second stage on the basis of the recessive model (OR = 2.59, CI = 1.20–5.58, *P* = 0.012).

standard Japanese SNPs[2] (which we refer to as the JSNP Genome Scan (JGS)), as part of the multi-disease collaborative genome scan (**Supplementary Fig. 1** online).

Among 100,000 SNPs genotyped by multiplex PCR–based Invader analysis in the first stage of the study, 82,343 autosomal polymorphisms passed our typing quality control in 187 individuals with diabetes (**Supplementary Table 1** online). We then carried out two separate association analyses to compare the 187 individuals with diabetes with two different control groups, which we considered as population controls: one to compare allele frequencies with reference data for 752 individuals representing the general Japanese population deposited in the JSNP database (referred to as the 'JSNP database control' (JDC)), and one to compare allele or genotype frequencies with those of the 752 individuals in the initial panels for the other four disease groups (Alzheimer's disease, gastric cancer, hypertension and asthma) of the national project (referred to as the 'other disease group' (ODG)). The combination of two types of association analysis resulted in the selection of 2,880 SNPs for the second stage of the study. An independent case-control panel (panel 2) was analyzed, and 201 positive SNPs ($P < 0.05$) were selected for the third stage (see **Supplementary Table 2a** online). Ten SNPs yielded a *P* value of <0.05 at the third stage using another case-control panel (panel 3; **Table 1** and **Supplementary Table 2b**). These SNPs showed variable *P* values in the first stage, suggestive of a limited power of the study design. The most significant association ($P = 3.4 \times 10^{-6}$) was obtained with rs2237895, which is located in intron 15 of *KCNQ1*. Another two SNPs (rs151290 and rs163184) were also located in the same intron, yielding *P* values of 1.1×10^{-4} and 0.0021, respectively. Panels 2 and 3 combined (panel 2+3) were analyzed for these 10 SNPs, yielding even lower *P* values for all the SNPs (**Supplementary Table 2b**). The genotype-based Cochran-Armitage trend test gave *P* values similar to those based on the allele data (**Supplementary Table 2b**).

We further analyzed *KCNQ1*, which was the only gene that yielded positive results according to the standard criterion (*P* value of $< 5 \times 10^{-7}$) recently proposed for GWAS[3]. The three SNPs of *KCNQ1* that passed the third scan (rs151290, rs163184 and rs2237895) were in moderate linkage disequilibrium (LD) with each other (**Fig. 1**). The SNP with the lowest *P* value, rs2237895, yielded D' and r^2 values of 0.54 and 0.12 with rs151290 and 0.83 and 0.46 with rs163184, respectively. We isolated 49 additional SNPs of *KCNQ1* from dbSNP of NCBI and typed them together with the three originally positive SNPs in panel 2+3 (**Fig. 1**). Among these 52 SNPs, rs2237892, which is also located in intron 15, showed the strongest association with diabetes ($P = 6.7 \times 10^{-13}$), with OR =1.49 and 95% CI = 1.34–1.66; the *P* value for the trend test was 1.7×10^{-12} (**Table 2**). The D' and r^2 values for rs2237895 and rs2237892 were 0.95 and 0.30, respectively.

We also sequenced all the exons and the 47-kb genomic region corresponding to intron 15 of *KCNQ1* in 24 Japanese individuals and identified 212 variations, including three synonymous and two nonsynonymous (P448R and G643S) polymorphisms (**Supplementary Table 3a** online). We then genotyped ten of the newly identified SNPs of intron 15 and the two nonsynonymous polymorphisms in panel 2+3. None of these SNPs showed a stronger association with diabetes than did rs2237892 (**Fig. 1** and **Supplementary Table 3b**).

We next examined the possible association of *KCNQ1* with diabetes in several additional subject panels, including those of other ancestral groups, by genotyping rs2237892, rs2237895 and rs2074196, the three SNPs that showed the strongest association in the original study. Two independent Japanese panels revealed a strong association of these

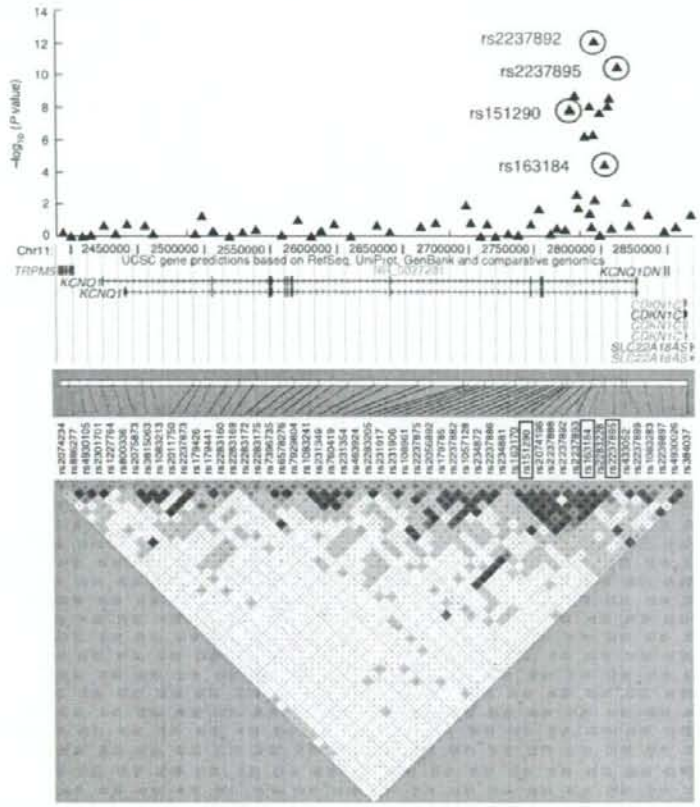

Figure 1 Dense mapping analysis of *KCNQ1*. The top panel shows the association $-\log_{10}$ (*P* value) in panel 2+3 for 64 SNPs of *KCNQ1*. The three blue circles represent the positive SNPs in the third screening. The red circle (rs2237892) indicates the SNP showing the most significant association with type 2 diabetes. The upper middle panel shows the physical position of *KCNQ1* and neighboring genes on chromosome 11 (UCSC Genome Browser). 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 ($n = 1,424$).

polymorphisms with diabetes (**Table 2** and **Supplementary Table 4** online); rs2237892, for example, showed allelic *P* values of 9.6×10^{-10} and 6.9×10^{-10} in the replication 1 and 2 panels, respectively. The three Japanese panels (panel 2+3 and replication 1 and 2), which included a total of 4,378 cases and 4,412 controls, yielded an allelic *P* value of 2.8×10^{-29} and OR of 1.43 (95% CI = 1.34–1.52) for rs2237892. The association was also reproduced in the replication 3 (Chinese) and replication 4 (Korean) panels; the allelic *P* values for rs2237892 in these two panels were 1.3×10^{-8} and 1.7×10^{-5}, respectively (**Table 2** and **Supplementary Table 4**). Meta-analysis of the Asian populations yielded a *P* value of 2.5×10^{-40} and OR of 1.42 (95% CI = 1.34–1.49) for rs2237892. We also examined rs2237892 and rs2074196 in the replication 5 panel (recruited from Sweden), with both SNPs showing a positive association ($P = 7.8 \times 10^{-4}$ and 0.017, respectively). With the inclusion of the replication 5 panel, meta-analysis with a total of 19,930 individuals (9,569 cases and 10,361 controls) yielded a *P* value of 1.7×10^{-42} and OR of 1.40 (95% CI = 1.34–1.47) for rs2237892 (**Table 2** and **Supplementary Fig. 2** online).

We next investigated the relation of rs2237892 to clinical phenotype. Among 1,424 individuals with diabetes in panel 2+3, no association was found between this SNP and clinical parameters such as body mass index (BMI) and the level of insulin resistance. Among the 948 control subjects in panel 2+3 whose fasting plasma glucose and insulin levels were available, homozygotes for the risk allele of rs2237892 (CC) showed a significantly lower homeostasis model assessment of β-cell function (HOMA-β)[4] than did those with the other genotypes (**Supplementary Table 5** online). Among nondiabetic subjects of the Botnia prospective cohort (**Supplementary Methods** online), the corrected insulin response (CIR) at the follow-up visit was significantly lower for individuals with the CC genotype of rs2237892 than for those with the other two genotypes in both an additive and recessive model for this SNP ($P = 0.024$ and 0.010, respectively; **Supplementary Table 5**). These results suggested that the risk allele of *KCNQ1* might contribute to diabetes susceptibility by impairing insulin secretion.

The multistage strategy for GWASs has an advantage in the effective elimination of a large number of false-positive results and has proved to be successful[5]. Indeed, we detected the association of several SNPs of *KCNQ1* with diabetes in the JGS, and this association was reproduced in two independent Japanese panels. *KCNQ1*, which encompasses 404 kb, is located at chromosome 11p15.5, not far from a candidate region at 11p13–p12 with suggestive evidence of linkage to type 2 diabetes in two independent studies of affected Japanese sibpairs[6,7]. We also reproduced the association of *KCNQ1* with diabetes in Chinese and Korean panels, establishing *KCNQ1* as a diabetes susceptibility gene for populations of East Asian descent. We further showed the association to be significant in individuals of European descent. Given that *KCNQ1* was not implicated as a diabetes susceptibility gene in two recent GWASs with individuals of European descent[8,9], we examined SNPs of *KCNQ1* in the available datasets (**Supplementary Fig. 3** and **Supplementary Table 6a,b** online). Within the LD block of *KCNQ1* that includes the SNPs associated with diabetes in Japanese, 11 SNPs in the WTCCC dataset[8] and 9 SNPs in the DGI dataset[9] had been typed, and none of them had been selected for further analysis. This apparent discrepancy may be due mainly to the allele frequencies of the causative SNPs (the minor allele frequency of rs2237892 was 0.28–0.41 and 0.05–0.07 in populations of East Asian and European descent, respectively). Indeed, in a recent meta-analysis of three GWASs (DGI, WTCCC and FUSION; see URLs section in Methods)[10], the risk alleles of both rs2237892 and rs2074196 identified in the present study were associated with an increased risk of type 2 diabetes ($P = 0.01$ and 0.02,

Table 2 Association study results for 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 (4,378 cases, 4,412 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), $P = 4.8 \times 10^{-21}$**
		Replication 3 (Chinese)	0.71	0.63	1.2×10^{-9}	1.40	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 (6,552 cases, 6,621 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), $P = 7.9 \times 10^{-33}$**
		Replication 5 (European)	0.96	0.95	0.017	1.23	1.04	1.46	0.017	
		All	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**1.35 (1.28–1.41), $P = 8.6 \times 10^{-34}$**
rs2237892	C	2+3 (dense mapping)	0.69	0.60	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	0.60	6.9×10^{-10}	1.41	1.26	1.57	1.1×10^{-9}	
		All Japanese (4,378 cases, 4,412 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), $P = 3.0 \times 10^{-29}$**
		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 (6,552 cases, 6,621 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), $P = 2.5 \times 10^{-40}$**
		Replication 5 (European)	0.95	0.93	7.8×10^{-4}	1.29	1.11	1.50	7.2×10^{-4}	
		All	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	**1.40 (1.34–1.47), $P = 1.7 \times 10^{-42}$**
rs2237895	C	2+3 (dense mapping)	0.41	0.33	3.1×10^{-11}	1.44	1.30	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	1.50	5.5×10^{-8}	
		All Japanese (4,378 cases, 4,412 controls)	**0.40**	**0.33**	**1.3×10^{-20}**	**1.34**	**1.26**	**1.43**	**1.7×10^{-20}**	**1.34 (1.26–1.43), $P = 1.4 \times 10^{-20}$**
		Replication 3 (Chinese)	0.40	0.34	3.5×10^{-5}	1.25	1.12	1.39	3.4×10^{-5}	
		Replication 4 (Korean)	0.35	0.30	3.2×10^{-3}	1.27	1.08	1.49	2.7×10^{-3}	
		All Asian (6,552 cases, 6,621 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), $P = 6.1 \times 10^{-26}$**
		Replication 5 (European)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	

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

respectively). These results provide further support for *KCNQ1* as a general susceptibility gene for diabetes, and they also highlight the need to extend GWAS to different populations.

Alternative splicing has been found to generate several variants of *KCNQ1* mRNA (see Accession codes section in Methods), but we do not know whether the identified candidate SNPs in intron 15 affect the splicing pattern of the primary transcript. Although neighboring genes seem to be located outside the LD block containing rs2237892, we are not able to exclude completely the possibility that the SNPs identified in the present study affect the expression of other causative genes. We did not find any microRNA harboring rs2237892 in the miRBase database.

KCNQ1 encodes the pore-forming subunit of a voltage-gated K^+ channel (KvLQT1) that is essential for the repolarization phase of the action potential in cardiac muscle[11]. Mutations in this gene are associated with cardiac diseases such as hereditary long QT syndrome (Romano-Ward syndrome[12] and Jervell and Lange-Nielsen syndrome[13]) and familial atrial fibrillation[14]. This K^+ channel is also expressed in other tissues, including brain, adipose tissues and pancreas[15,16]. The lower HOMA-β or CIR apparent for CC homozygotes of rs2237892 among Japanese and Europeans in the present study may reflect a functional role for this channel in insulin-producing cells. We examined the abundance of *Kcnq1* mRNA by reverse transcription and real-time PCR analysis in the islets of 12-week-old diabetic KK-Ay mice, which manifested both hyperglycemia and hyperinsulinemia. The amount of the mRNA was significantly increased ($P = 0.0004$) by a factor of 1.6 compared with that in the islets of C57BL6 control mice (data not shown). The KCNQ1 protein was previously shown to be expressed in insulin-secreting INS-1 cells, and the KCNQ1 blocker 293B was found to stimulate insulin secretion in the presence of tolbutamide[17]. It is also possible that fine-tuning of the membrane potential by this channel might modulate the survival of pancreatic β cells in the long term. Further studies are necessary to elucidate the precise mechanism by which the risk allele of *KCNQ1* confers susceptibility to diabetes.

We may have missed a substantial number of susceptibility genes in our screening, given that the strategy we adopted seven years ago lacks sufficient analytical power[18] relative to that now achievable as a result of recent progress in genomic studies. The genomic coverage of the SNP set was not robust, in part because the IMS-JST Japanese SNP (JSNP) database was designed to focus on 'gene-centric' SNPs[2]. Several comprehensive studies based on new platforms for GWASs have recently been described, with about ten genes being found to be

reproducibly associated with type 2 diabetes in individuals of European ancestry[8,9,19–23]. None of these genes showed a positive association in our JGS typing data. Given that some of these genes were recently shown to confer susceptibility to diabetes in Japanese[24–26], the lack of association in our study might be due to the limited sample size of the first scan or to weak LD between the SNPs we used and the causative variants; actually, some genes were totally missed in our JGS (**Supplementary Table 6c**).

In summary, with a comprehensive multistage SNP association study in Japanese, we have identified *KCNQ1* as a previously unreported susceptibility gene as well as several other candidate genes for type 2 diabetes mellitus. Replication studies further confirmed the association of *KCNQ1* with diabetes in individuals of East Asian and European descent. Our findings may provide new insight into the pathophysiology of diabetes as well as a basis for the development of new therapeutic agents.

METHODS

Study participants. We assembled three independent subject panels for multistage genome-wide screening. Panel 1 consisted of 188 cases only, panel 2 of 752 cases and 752 controls and panel 3 of 672 cases and 672 controls. The inclusion criteria for diabetic patients were as follows: (i) age of disease onset of 40 to 55 years, (ii) maximum BMI of <30 kg/m^2, (iii) insulin treatment not initiated until at least three years after diagnosis and (iv) absence of antibodies to glutamic acid decarboxylase. Most Japanese diabetic individuals have a BMI of <30 kg/m^2, and we aimed to focus on the most common subtype of type 2 diabetes in Japan. The criteria for controls in panels 2 and 3 were as follows: (i) age of >60 years, (ii) no past history of diagnosis of diabetes and (iii) hemoglobin A_{1c} content of $<5.6\%$. The cases in the three panels and the controls in panels 2 and 3 were recruited at 11 core facilities located in various regions of Japan. Panels 2 and 3 were assembled simultaneously. Genomic DNA was extracted from peripheral blood by standard methods. We also obtained clinical information such as BMI, blood biochemistry (including plasma glucose and insulin levels) and family history of diabetes. The replication panels are described in **Supplementary Methods**. The clinical characteristics of subjects in each panel are summarized in **Supplementary Table 1**. The study protocol was approved by the local ethics committee of each institution, and written informed consent was obtained from all participants.

Study design. The general design and power for the multistage screening in the Millennium Genome Project (**Supplementary Fig. 1**), referred to as the JSNP Genome Scan (JGS), have been described previously[18]. In the first stage, 188 individuals with each disease (panel 1 for diabetes) were genotyped for 100,000 SNPs in the IMS-JST JSNP database (see URLs section below)[2]. The coverage of the nucleotide sequences of the RefSeq NM exonic regions (as defined by 5′ UTR + CDS (coding sequences) + 3′ UTR) achieved by the JSNP 'gene-centric' genome-wide LD mapping is estimated to be ~35%, if we assume an average extent of LD of 10 kb for each SNP with a minor allele frequency (MAF) of $>15\%$. We also previously evaluated the power of the first two stages of the JGS by a simulation experiment[18]. For example, this analysis would yield a sensitivity of ~13% for SNPs with an odds ratio of 1.5 and a disease-associated genotype frequency of 30%.

One subject did not yield a genotype call for any SNP in the first stage. We then carried out two separate association analyses to compare the 187 diabetic individuals with two different control groups, which we referred to as JDC and ODG, respectively. We did not detect significant population stratification among individuals of the initial panels of the five disease groups by standard methods such as genomic control[27] (inflation factor = 1.06 with 1,025 SNPs selected for genomic control analysis). The genotype-based analysis was done with dominant and recessive models. First, SNPs whose MAF was $>10\%$ in the database and which showed either a genotype OR of >1.5 or an allele OR of >1.3 in either association analysis were selected. If multiple SNPs in the same gene with positive association were in strong LD ($r^2 > 0.9$), only one SNP was chosen for the next step to avoid redundancy. A total of 2,880 SNPs for each disease was then selected for the second screening in order of P value; for diabetes, 2,343 and 1,111 SNPs were selected by the association analyses with ODG and JDC, respectively, with 574 SNPs being selected by both analyses.

In the second stage, an independent case-control panel (panel 2) was analyzed, generating valid data for 2,827 SNPs after a quality check. Thirty-eight SNPs gave no results for all the samples in panel 2, whereas five and three SNPs yielded no data for all case or control samples, respectively, by multiplex PCR-based Invader analysis, and seven probes were not annotated on the updated human genome. The call rate for the 2,827 SNPs was 0.993. A total of 201 positive SNPs ($P < 0.05$) was selected for the third stage of the study on the basis of allelic data (**Supplementary Table 2a**). In the third stage, another case-control panel (panel 3) was typed; one SNP could not be typed by SSP-PCR-FCS analysis (see below) for any of the subjects in panel 3, with the call rate for the other 200 SNPs being 0.990. The ten positive SNPs ($P < 0.05$; **Table 1**) were also then analyzed in the combined panels 2 and 3 (panel 2+3, 1,424 cases and 1,424 controls). Panel 2 was genotyped again for these ten SNPs by SSP-PCR-FCS analysis, and the concordance rate with the Invader method used in the second screening was 0.992. The possibility of stratification in panels 2 and 3 was assessed by typing of 28 diabetes-unrelated SNPs followed by (i) comparison of allele and genotype frequencies by the χ^2 test, (ii) principal component analysis or (iii) STRUCTURE analysis (see URLs section below). None of these analyses showed evidence of stratification among cases and controls of panels 2 and 3 (data not shown).

The list of SNPs used for the initial screening and the allele and genotype frequency data for the first and the second stages of the JGS for the five diseases studied in the Millennium Genome Project of Japan, including diabetes, have been deposited in the Genome Medicine Database of Japan (GeMDBJ, see URLs section below).

Dense SNP mapping for *KCNQ1*. We first selected 49 additional SNPs of *KCNQ1* from the dbSNP database of NCBI, with an average interval of ~10 kbp, and typed these polymorphisms in panel 2+3 together with the three positive SNPs originally included in the JGS. We sequenced 24 control Japanese subjects for the gene, including all the exons and the putative promoter region (4 kbp upstream from the transcription start site), in order to comprehensively identify genetic variants in Japanese. We also sequenced the regions surrounding the positive SNPs of *KCNQ1*, spanning 47 kbp (intron 15). Ten of the SNPs identified in the 47-kbp region were selected on the basis of LD and MAF ($>10\%$). These 10 SNPs and the two identified nonsynonymous variants were genotyped in panel 2+3. A total of 64 SNPs was thus genotyped for *KCNQ1*, including 18 SNPs in the 35.6-kbp region between rs151290 and rs2237895, with an average interval of 2 kbp (see **Supplementary Table 3b**).

Typing methods. In the first and second stages of the study, genotyping was done by the multiplex PCR-based Invader assay (Third Wave Technologies) as previously described[28]. In the third stage and for dense mapping, genome-wide amplified DOP degenerate oligonucleotide–primed (DOP)-PCR templates were genotyped by sequence-specific primer (SSP)-PCR analysis followed by fluorescence correlation spectroscopy (FCS)[29]. Some SNPs included in dense mapping were therefore re-genotyped in panel 2 by the SSP-PCR-FCS method. Some SNPs were genotyped by real-time PCR analysis with TaqMan probes (Applied Biosystems). For replication panels, we applied either SSP-PCR-FCS or the TaqMan method.

Statistical analysis. In the first screening, we performed two case-control evaluations as described above. We examined allele or genotype (dominant or recessive models) data in 2×2 contingency tables for comparison with ODG, as well as allele data in 2×2 contingency tables for comparison with JDC (for which genotype data were not available). In the second and third screening and dense mapping, we analyzed allele data in 2×2 contingency tables by the χ^2 test. LD and haplotype analyses were done with Haploview 3.31 software[30]. A P value of <0.05 was considered statistically significant. For ten positive SNPs in the JGS, rs2237892 and rs2074196, genotype-based analyses were also performed by the Cochran-Armitage trend test. Meta-analysis was done by the Mantel-Haenszel method (fixed-effects models) with the "meta" package of the R Project; the P values for heterogeneity among panels joined in the Mantel-Haenszel tests were all >0.05.

URLs. Genome Medicine Database of Japan, https://gemdbj.nibio.go.jp/dgdb/; DGI, WTCCC and FUSION, http://www.well.ox.ac.uk/DIAGRAM/; miRBase database, http://microrna.sanger.ac.uk/sequences/; IMS-JST JSNP database, http://snp.ims.u-tokyo.ac.jp/; STRUCTURE analysis, http://pritch.bsd.uchicago.edu/software.html.

Accession codes. GenBank: *KCNQ1* mRNA, NM_000218.2 and NM_181798.1.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all the participants in the project; S. Sugano and S. Tsuji for support and helpful discussion throughout the project; H. Sakamoto, K. Yoshimura and N. Nishida for genotyping and quality control of the data; M. Yamaoka-Sageshima, K. Nagase, D. Suzuki and A. Berglund for technical assistance; and staff of Mitsui Knowledge Industry Inc. (Tokyo) for help with bioinformatics. This work was supported by a grant from the Program for Promotion of Fundamental Studies in Health Sciences of the Pharmaceuticals and Medical Devices Agency (PMDA) of Japan; a grant from the National Institute of Biomedical Innovation (NIBIO) of Japan; grants from the Ministry of Health, Labour, and Welfare of Japan; a Grant-in-Aid for Scientific Research on Priority Areas (C), "Medical Genome Science (Millennium Genome Project)," "Applied Genomics" and "Comprehensive Genomics" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan; and a grant from New Energy and Industrial Technology Development Organization (NEDO). The replication 2 study was supported by a grant from Cooperative Link of Unique Science and Technology for Economy Revitalization (CLUSTER, Tokushima, Japan). The Hong Kong diabetes case-control study was supported by the Hong Kong Research Grants Committee Central Allocation Scheme CUHK 1/04C. The Korean case-control study was supported by a grant from the Korea Health 21 R&D Project of the Ministry of Health and Welfare of the Republic of Korea (00-PJ3-PG6-GN07-001 to K.S.P.). The replication 5 study and Botnia prospective study were supported by Swedish Research Council (Linne grant), Sigrid Juselius Foundation, Folkhaelsan Research Foundation, European Foundation for the Study of Diabetes and Swedish Diabetes Research Foundation.

AUTHOR CONTRIBUTIONS

Principal investigators: K. Yasuda and M.K. Manuscript writing: K. Yasuda., K.M., Y. Horikawa and M.K. Diabetes project planning and design: K. Yasuda, K.M., Y. Hirota, H. Mori, T.Y. and M.K. Ascertainment of study subjects and general data analyses in Japan: K. Yasuda, K.M., Y. Horikawa, K.H., H.O., H.F., Y. Hirota, H. Mori, Y. Sato, K. Yamagata, Y. Hinokio, H.-Y.W., T. Tanahashi, N.N., Y.O., N.I., Y.I., Y.Y., Y. Seino, H. Maegawa, A.K., J.T., E.M., N.K., M.I., H. Makino, K.N., T.K. and M.K. Genotyping and sequencing analyses in Japan: K.M., Y. Horikawa, Y. Hirota, T. Tanahashi, A.S., Y.N., K. Yamamoto, T.Y., K.T. and M.I. Statistical analyses: K.M., Y. Horikawa, Y. Hirota, E.M., T.Y., K.T. and M.I. Genetic analyses in Korea: H.D.S., Y.M.C., K.S.P. and H.K.L. Genetic analyses in Hong Kong: M.C.Y.N., R.C.W.M., W.-Y.S. and J.C.N.C. Genetic analyses in Europe: A.J., V.L., T. Tuomi, P.N. and L.G. Millennium Genome Project Human Genome Variation Team Leader: Y.N. Millennium Genome Project Diabetes Subteam Leader: M.K.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

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. 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. *J. Hum. Genet.* **47**, 605–610 (2002).
3. NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature* **447**, 655–660 (2007).
4. Matthews, D.R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
5. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
6. Mori, Y. *et al.* 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).
7. Nawata, H. *et al.* 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).
8. Zeggini, E. *et al.* Replication of genome-wide association signals in U.K. samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1340 (2007).
9. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
10. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
11. Barhanin, J. *et al.* K(v)LQT1 and IsK (minK) proteins associate to form the I(Ks) cardiac potassium current. *Nature* **384**, 78–80 (1996).
12. Wang, Q. *et al.* Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nat. Genet.* **12**, 17–23 (1996).
13. Neyroud, N. *et al.* A novel mutation in the potassium channel gene KvLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome. *Nat. Genet.* **15**, 186–189 (1997).
14. Chen, Y.H. *et al.* KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science* **299**, 251–254 (2003).
15. Demolombe, S. *et al.* Differential expression of KvLQT1 and its regulator IsK in mouse epithelia. *Am. J. Physiol. Cell Physiol.* **280**, C359–C372 (2001).
16. Chouabe, C. *et al.* Properties of KvLQT1 K+ channel mutations in Romano-Ward and Jervell and Lange-Nielsen inherited cardiac arrhythmias. *EMBO J.* **16**, 5472–5479 (1997).
17. Ullrich, S. *et al.* Effects of I_{Ks} channel inhibitors in insulin-secreting INS-1 cells. *Pflugers Arch.* **451**, 428–436 (2005).
18. Sato, Y. *et al.* Designing a multistage, SNP-based, genome screen for common diseases. *J. Hum. Genet.* **49**, 669–676 (2004).
19. Grant, S.F. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.* **38**, 320–323 (2006).
20. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
21. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
22. Steinthorsdottir, V. *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770–775 (2007).
23. Frayling, T.M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
24. Miyake, K. *et al.* Association of TCF7L2 polymorphisms with susceptibility to type 2 diabetes in 4,087 Japanese subjects. *J. Hum. Genet.* **53**, 174–180 (2008).
25. Horikoshi, M. *et al.* Variations in the HHEX gene are associated with increased risk of type 2 diabetes in the Japanese population. *Diabetologia* **50**, 2461–2466 (2007).
26. Horikawa, Y. *et al.* Replication of genome-wide association studies of type 2 diabetes susceptibility in Japan. *J. Clin. Endocrinol. Metab.* (in the press).
27. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
28. Ohnishi, Y. *et al.* A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46**, 471–477 (2001).
29. Bannai, M. *et al.* Single-nucleotide-polymorphism genotyping for whole-genome-amplified samples using automated fluorescence correlation spectroscopy. *Anal. Biochem.* **327**, 215–221 (2004).
30. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).

Journal of Human Genetics (2009), 1–6

www.nature.com/jhg

ORIGINAL ARTICLE

Construction of a prediction model for type 2 diabetes mellitus in the Japanese population based on 11 genes with strong evidence of the association

Kazuaki Miyake[1], Woosung Yang[2], Kazuo Hara[3], Kazuki Yasuda[4], Yukio Horikawa[5], Haruhiko Osawa[6], Hiroto Furuta[7], Maggie CY Ng[8], Yushi Hirota[1], Hiroyuki Mori[1], Keisuke Ido[2,20], Kazuya Yamagata[9,21], Yoshinori Hinokio[10], Yoshitomo Oka[10], Naoko Iwasaki[11], Yasuhiko Iwamoto[11], Yuichiro Yamada[12,22], Yutaka Seino[12,23], Hiroshi Maegawa[13], Atsunori Kashiwagi[13], He-yao Wang[4,24], Toshihito Tanahashi[14], Naoto Nakamura[15], Jun Takeda[5], Eiichi Maeda[2], Ken Yamamoto[16], Katsushi Tokunaga[17], Ronald CW Ma[8], Wing-Yee So[8], Juliana CN Chan[8], Naoyuki Kamatani[18], Hideichi Makino[6], Kishio Nanjo[7], Takashi Kadowaki[3] and Masato Kasuga[1,19]

Prediction of the disease status is one of the most important objectives of genetic studies. To select the genes with strong evidence of the association with type 2 diabetes mellitus, we validated the associations of the seven candidate loci extracted in our earlier study by genotyping the samples in two independent sample panels. However, except for *KCNQ1*, the association of none of the remaining seven loci was replicated. We then selected 11 genes, *KCNQ1*, *TCF7L2*, *CDKAL1*, *CDKN2A/B*, *IGF2BP2*, *SLC30A8*, *HHEX*, *GCKR*, *HNF1B*, *KCNJ11* and *PPARG*, whose associations with diabetes have already been reported and replicated either in the literature or in this study in the Japanese population. As no evidence of the gene–gene interaction for any pair of the 11 loci was shown, we constructed a prediction model for the disease using the logistic regression analysis by incorporating the number of the risk alleles for the 11 genes, as well as age, sex and body mass index as independent variables. Cumulative risk assessment showed that the addition of one risk allele resulted in an average increase in the odds for the disease of 1.29 (95% CI=1.25–1.33, $P=5.4\times10^{-53}$). The area under the receiver operating characteristic curve, an estimate of the power of the prediction model, was 0.72, thereby indicating that our prediction model for type 2 diabetes may not be so useful but has some value. Incorporation of data from additional risk loci is most likely to increase the predictive power.
Journal of Human Genetics advance online publication, 27 February 2009; doi:10.1038/jhg.2009.17

Keywords: gene–gene interaction; genome-wide association study; prediction model; single nucleotide polymorphism (SNP); type 2 diabetes mellitus

[1]Division of Diabetes, Metabolism and Endocrinology, Department of Internal Medicine, Kobe University Graduate School of Medicine, Kobe, Japan; [2]Clinical Genome Informatics Center, Kobe University Graduate School of Medicine, Kobe, Japan; [3]Department of Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [4]Department of Metabolic Disorder, Research Institute, International Medical Center of Japan, Tokyo, Japan; [5]Division of Molecule and Structure, Department of Diabetes and Endocrinology, Gifu University School of Medicine, Gifu, Japan; [6]Department of Molecular and Genetic Medicine, Ehime University Graduate School of Medicine, Ehime, Japan; [7]First Department of Medicine, Wakayama Medical University, Wakayama, Japan; [8]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Shatin, Hong Kong; [9]Department of Metabolic Medicine, Graduate School of Medicine, Osaka University, Osaka, Japan; [10]Division of Molecular Metabolism and Diabetes, Tohoku University Graduate School of Medicine, Sendai, Japan; [11]Department of Medicine, Diabetes Center, Tokyo Women's Medical University, Tokyo, Japan; [12]Department of Diabetes and Clinical Nutrition, Kyoto University School of Medicine, Kyoto, Japan; [13]Division of Endocrinology and Metabolism, Department of Medicine, Shiga University of Medical Science, Shiga, Japan; [14]Division of Genetic Information, Institute for Genome Research, University of Tokushima, Tokushima, Japan; [15]Department of Endocrinology and Metabolism, Kyoto Prefectural University of Medicine, Graduate School of Medical Sciences, Kyoto, Japan; [16]Department of Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; [17]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [18]Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo, Japan and [19]Research Institute, International Medical Center of Japan, Tokyo, Japan

Correspondence: Dr M Kasuga, Research Institute, International Medical Center of Japan, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan.
E-mail: kasuga@ri.imcj.go.jp

[20]Current address: Information Center for Medical Sciences, Tokyo Medical and Dental University, Tokyo, Japan.
[21]Current address: Faculty of Medical and Pharmaceutical Sciences, Department of Medical Biochemistry, Kumamoto University, Kumamoto, Japan.
[22]Current address: Department of Internal Medicine, Akita University School of Medicine, Akita, Japan.
[23]Current address: Kansai Electric Power Hospital, Osaka, Japan.
[24]Current address: Shanghai Institute of Materia Medica, Chinese Academy of Science, Shanghai, China.
Received 22 December 2008; revised 25 January 2009; accepted 5 February 2009

INTRODUCTION

Genome-wide association studies (GWASs) have identified novel susceptibility genes for type 2 diabetes mellitus in Caucasians.[1–5] *TCF7L2*, *CDKAL1*, *CDKN2A/B*, *IGF2BP2*, *SLC30A8* and *HHEX* have been widely replicated as susceptibility genes for type 2 diabetes in Asian populations[6–12] as well as in populations of European ancestry.[13,14] We recently identified *KCNQ1* as a novel susceptibility gene, as well as seven other candidate susceptibility loci in a multistage GWAS for type 2 diabetes in the Japanese population, in which a total of 1612 cases and 1424 controls and 100 000 single nucleotide polymorphisms (SNPs) were included.[15] *KCNQ1* was found to confer risk of type 2 diabetes with a relatively large effect size in Asian populations (odds ratio (OR) for Japanese, Chinese and Korean individuals of 1.42),[15] which was similar to that demonstrated earlier for *TCF7L2* in the Japanese population.[6]

Follow-up of GWASs includes analysis of second-tier genes, meta-analysis for specific populations, as well as analysis of gene–gene or gene–environment interactions. A large-scale meta-analysis[16] and an analysis of gene–gene interaction for susceptibility genes[17] have been performed for type 2 diabetes in populations of European ancestry.

In this study, we attempted to confirm in independent subject panels of Japanese and Hong Kong Chinese individuals the associations of the seven candidate susceptibility loci that we identified in addition to *KCNQ1* in our GWAS of type 2 diabetes.[15] However, as described in this article, we failed to replicate the associations of the seven loci with diabetes. We then attempted to extract genes with strong evidence of the associations with diabetes, and selected 11 genes, including *KCNQ1*. As we did not detect any gene–gene interaction between the 11 genes, we then attempted to construct a prediction model for this disease by using the data from the 11 genes, as well as age, gender and body mass index (BMI) as independent variables to obtain a comprehensive understanding of the genetic background of diabetes in the Japanese population.

MATERIALS AND METHODS

Validation of the results from a multistage GWAS in the Japanese population

Study subjects. We assembled two independent subject panels for our replication study: replication-Japanese and replication-Chinese. The 1000 cases and 1000 controls for the replication-Japanese panel were recruited by the Study Group of the Millennium Genome Project for Diabetes Mellitus. The inclusion criteria for diabetic patients were (i) an age at disease onset of 30–60 years and (ii) the absence of antibodies to GAD. Types of diabetes other than type 2 were excluded on the basis of clinical data. The criteria for controls included (i) an age of >50 years, (ii) no past history of a diagnosis of diabetes and (iii) an HbA_{1c} content of <5.8%.

For the replication-Chinese panel, subjects of southern Han Chinese ancestry, who resided in Hong Kong, were recruited. The cases consisted of 1416 individuals with type 2 diabetes selected from the Prince of Wales Hospital Diabetes Registry;[5,18] 626 of these subjects had early-onset diabetes (age at diagnosis of <40 years) and a positive family history, whereas the remaining 790 patients were randomly selected from the registry. Patients with classic type 1 diabetes with acute ketotic presentation or a continuous requirement for insulin within 1 year of diagnosis were excluded. The controls consisted of 1577 subjects with normal glucose tolerance (fasting plasma glucose concentration of $<6.1\,mmol\,l^{-1}$); 596 of these individuals were recruited either from the general population participating in a community-based screening program for cardiovascular risk or from hospital staff, whereas the remaining 981 subjects were recruited from a population-based screening program for cardiovascular risk in adolescents.[19] The clinical characteristics of the subjects in each panel are summarized in Supplementary Table 1A. The study protocol was approved by the local ethics committee of each institution. Written informed consent was obtained from each subject.

Study design and statistical analysis. For the validation of the results from our earlier multistage GWAS,[15] seven SNPs (rs2250402, rs2307027, rs3741872, rs574628, rs2233647, rs3785233 and rs2075931) were genotyped in the two panels either by sequence-specific primer–PCR analysis followed by fluorescence correlation spectroscopy[20] or by real-time PCR analysis with TaqMan probes (Applied Biosystems, Foster City, CA, USA). Differences in allele frequency between cases and controls for each SNP were evaluated by χ^2 with one degree of freedom. Meta-analysis was performed by the Mantel–Haenszel method (fixed-effects models) with the 'meta' package of the R-Project (http://www.r-project.org). A *P*-value of <0.05 was considered statistically significant.

Examination of gene–gene interaction and construction of a prediction model

Study subjects. In total, 2424 cases and 2424 controls of the Japanese population obtained by combining the second and third screening panels in our original study[15] and the replication-Japanese panel of this study were included in this analysis (analysis-panel). The criteria for the second and third screening panels were described in the earlier report.[15] The clinical characteristics of the subjects are summarized in Supplementary Table 1B.

Selection of the loci included in this study. Prediction of the phenotypes on the basis of genetic polymorphisms should include the genetic data from the loci with strong evidence of the association. Starting from 15 genes described in earlier reports, we selected 11 genes with strong evidence of the association on the basis of the data in the literature and on the results of the replication experiments in this study. Process of the selection of the 11 genes will be described in detail in Results.

Statistical methods. Multiplicative gene–gene interaction was evaluated for each pair of the 11 genes using an interaction term in addition to the terms for the pair of the genes in the logistic regression model. The genotypes for each locus were coded by 0, 1 and 2. Correction for multiple testing was performed by Bonferroni's method.

As there was no evidence for the presence of gene–gene interactions, we attempted to construct a phenotype prediction model by incorporating the number of risk alleles for the 11 loci as an independent variable in addition to age, gender and BMI. The Cochran–Armitage test was used to examine the trend of the increase in the odds by increasing the number of the risk alleles. To construct a prediction model, the log of odds was expressed by the linear combination of the independent variables. Coefficients for the variables were estimated by the logistic regression analysis after making disease (cases) or nondisease (controls) as the dependent variable. Using the coefficients estimated by the logistic regression analysis, we constructed a phenotype prediction model. To evaluate the prediction model, receiver operating characteristic (ROC) curves[21] for the sensitivity and specificity of the prediction model with or without adjustment for age, sex and BMI were generated, and the area under the curve (AUC) was calculated from the ROC curve.

RESULTS

Validation of the results from a multistage GWAS in the Japanese population

We identified earlier 10 loci associated with type 2 diabetes by three-staged GWAS starting from 100 000 SNPs. Among the 10 loci, 3 SNPs were located in an intron of *KCNQ1*, and the association of this gene with diabetes was confirmatory.[15] To validate the other seven loci for the association with type 2 diabetes, we analyzed them in two independent replication panels of Japanese and Han-Chinese individuals (Table 1, Supplementary Table 2). Only one SNP, rs2250402, which is located in *EIF2AK4*, was found to be significantly associated in the replication-Japanese panel ($P=0.039$, OR=1.17, 95% CI=1.01–1.36). However, neither this SNP ($P=0.41$, OR=1.05) nor any of the other six SNPs showed such an association in the replication-Chinese panel. Meta-analyses for these SNPs showed that rs2307027 in *KRT4* and rs3785233 in *A2BP1* yielded *P*-values of <0.05 and ORs between 1.12 and 1.13 (Table 1). When the original second and third screening

Table 1 Association study for the candidate susceptibility genes for type 2 diabetes selected by multistage screening in the Japanese population

SNP ID	Chr	Gene	Risk allele	Panel	RAF (DM)	RAF (NC)	P	OR	95% CI
rs2250402	15	*EIF2AK4*	C	Replication-Japanese	0.23	0.20	0.04	1.17	1.01–1.36
				Replication-Chinese	0.24	0.23	0.41	1.05	0.93–1.19
				Meta-analysis			0.05	1.10	1.00–1.20
rs2307027	12	*KRT4*	C	Replication-Japanese	0.18	0.17	0.17	1.12	0.95–1.32
				Replication-Chinese	0.14	0.13	0.16	1.11	0.96–1.29
				Meta-analysis			0.05	1.12	1.00–1.25
rs3741872	12	*FAM60A*	C	Replication-Japanese	0.25	0.24	0.18	1.11	0.96–1.28
				Replication-Chinese	0.23	0.22	0.21	1.08	0.96–1.22
				Meta-analysis			0.07	1.09	0.99–1.20
rs574628	20	*ANGPT4*	G	Replication-Japanese	0.60	0.61	0.46	0.95	0.84–1.08
				Replication-Chinese	0.65	0.65	0.59	1.03	0.93–1.15
				Meta-analysis			0.96	1.00	0.92–1.08
rs2233647	6	*SPDEF*	G	Replication-Japanese	0.86	0.87	0.70	0.97	0.81–1.16
				Replication-Chinese	0.94	0.93	0.54	1.07	0.87–1.31
				Meta-analysis			0.90	1.01	0.88–1.16
rs3785233	16	*A2BP1*	C	Replication-Japanese	0.18	0.16	0.19	1.12	0.95–1.32
				Replication-Chinese	0.13	0.12	0.10	1.14	0.97–1.34
				Meta-analysis			0.04	1.13	1.01–1.27
rs2075931	1	Intergenic	A	Replication-Japanese	0.67	0.66	0.85	1.01	0.89–1.16
				Replication-Chinese	0.73	0.74	0.27	0.94	0.84–1.05
				Meta-analysis			0.48	0.97	0.89–1.06

Abbreviations: Chr, chromosome; OR, odds ratio for risk allele frequency.
Assignment of risk alleles was based on the original study.[15] Numbers of cases versus control subjects in the replication-Japanese and replication-Chinese panels were 1000 versus 1000 and 1416 versus 1577, respectively. RAF (DM) and RAF (NC) denote risk allele frequencies in cases and controls, respectively. *P* values were calculated for allele frequency. Meta-analysis was performed by the Mantel–Haenszel method (fixed-effects models). *P*-values for the test of heterogeneity among panels joined in the Mantel–Haenszel tests were all >0.05.

panels were included in the meta-analyses, these two loci, as well as the SNPs in *EIF2AK4* (rs2250402) and *FAM60A* (rs3741872), gave *P*-values of <0.001 and ORs between 1.15 and 1.18 (Supplementary Table 3). However, the *P*-values did not reach the proposed significance of GWAS ($=5\times10^{-7}$).

Selection of polymorphisms for the prediction model

To construct a reliable prediction model for diabetes, polymorphisms with strong evidence of association should be used. From the previous literature, we selected 15 genes (including one intergenic marker), that is, *SLC30A8*, *HHEX*, *LOC387761*, *EXT2*, *CDKN2A/B*, *GCKR*, *IGF2BP2*, *CDKAL1*, *FTO*,[1–5] *TCF7L2*,[22] *KCNJ11*,[23] *PPARG*,[24] *WFS1*,[25] *HNF1B*[26] and *KCNQ1*,[15] as candidate genes to be included in both gene–gene interaction analysis and construction of a prediction model. Starting from 23 SNPs in these 15 genes, we selected 11 SNPs in 11 genes according to the following process. There is sufficient evidence of the associations of *KCNQ1* and *TCF7L2* genes with diabetes as supported by replication studies in the Japanese population.[6,15,27] In addition, *SLC30A8*, *HHEX*, *CDKN2A/B*, *IGF2BP2* and *CDKAL1* associated with the disease in the European population were found in our earlier study to be associated with the disease in the Japanese population as well.[7–9]

To further extract genes with strong evidence of the association with diabetes, we attempted to replicate the associations reported earlier using our own data (analysis panel with 2424 cases and 2424 controls). For the 19 SNPs in *SLC30A8*, *HHEX*, *LOC387761*, *EXT2*, *CDKN2A/B*, *GCKR*, *IGF2BP2*, *CDKAL1*, *FTO*, *TCF7L2*, *KCNJ11*, *PPARG* and *KCNQ1*, we extracted genotyping data from our earlier studies[6–9,15,27–29] and, if necessary, genotyped additional subjects to obtain a data set for 2424 cases and 2424 controls of the Japanese population (analysis panel). The SNPs in *WFS1* (rs6446482, rs734312) and *HNF1B* (rs7501939, rs4430796) were genotyped for this study in the same individuals. SNPs with *P*-values for the test of deviation from the Hardy–Weinberg equilibrium of <0.01 were excluded for further analysis. When two SNPs were located in the same genomic region, the one with the lower *P*-value for the association test was selected for further analysis. *GCKR*, for which we earlier reported the marginal association with type 2 diabetes,[7] was found to be associated with the disease in this enlarged Japanese panel ($P=1.7\times10^{-5}$; Supplementary Table 4). *KCNJ11* and *PPARG*, which have been included in the genes associated with diabetes in Caucasians, showed marginal associations ($P=0.066$ and $P=0.075$, respectively; Supplementary Table 4) in our panel. Two SNPs in *WFS1* and two SNPs in *HNF1B* were newly genotyped in the analysis panel. Although no association was apparent between *WFS1* and type 2 diabetes, both SNPs in *HNF1B* exhibited *P*-values of <0.05 (Supplementary Table 4). From these data, we included 11 SNPs in 11 genes as described above for the source of genotype data to be analyzed in both the examination of gene–gene interaction and the prediction of phenotypes.

Gene–gene interaction

We evaluated multiplicative gene–gene interaction for each pair of the 11 loci as described in Materials and methods. Two combinations, rs1801282 (*PPARG*) ×rs1470579 (*IGF2BP2*) (nominal $P=0.0025$) and rs1801282×rs3802177 (*SLC30A8*) (nominal $P=0.018$), showed *P*-values of less than 0.05 (Supplementary Figure 1). However, these *P*-values were not significant when Bonferroni's correction for multiple testing was applied (significance level, $0.05/55=9.1\times10^{-4}$). Although *PPARG* and *IGF2BP2* are located on the same chromosome (3p25 and 3q28, respectively), it is unlikely that loci on different arms of the same chromosome show significant linkage disequilibrium. *SLC30A8* is located on a different chromosome (8q24.11) from

PPARG. The reason why nominal *P*-values of these combinations showed less than 0.05 may be because of the low minor allele frequency of rs1801282.

Cumulative risk assessment for type 2 diabetes on the basis of susceptibility genes

As there was no evidence of gene–gene interaction between 11 SNPs of 11 genes, *SLC30A8*, *HHEX*, *CDKN2A/B*, *GCKR*, *IGF2BP2*, *CDKAL1*, *TCF7L2*, *KCNJ11*, *PPARG*, *KCNQ1* and *HNF1B*, they were included in the prediction model as independent variables with the additive effect (additive effect in the liability and multiplicative effect in the odds) without interaction terms. Effective numbers of cases and controls whose genotypes for the 11 loci were successfully obtained were 2316 and 2370, respectively. The Cochran–Armitage trend test gave a *P*-value of 4.7×10^{-56} for the trend in the increase in the odds for cases relative to controls with an increasing number of risk alleles for the 11 susceptibility loci (Supplementary Table 5). We then estimated ORs for type 2 diabetes in subjects with different numbers of risk alleles on the basis of the multiplicative model by logistic regression analysis with adjustment for age, sex and BMI. The ORs for type 2 diabetes in subjects with 7–18 risk alleles in comparison with those harboring 0–6 risk alleles are shown in Figure 1. An increase of one risk allele resulted in an average increase in the odds of 1.29 (95% CI=1.25–1.33, $P=5.4\times10^{-53}$, logistic regression analysis).

To predict disease status for type 2 diabetes in a given individual, we constructed a prediction model on the basis of the number of risk alleles or the liability value calculated from the number of risk alleles as well as age, sex and BMI. The coefficients to calculate the liability value were estimated with the logistic regression model. To estimate the predictive power of the model, we generated ROC curves as described in Materials and methods. The AUC was 0.63 when only the number of risk alleles was used for the prediction. When age, sex and BMI were also included, the AUC increased to 0.72 (Figure 2). Meanwhile, an AUC value for the ROC curve based on only age, sex and BMI was 0.68, which was better than that based on only the number of risk alleles (data now shown). The model incorporating age, sex and BMI as well as the number of risk alleles thus showed moderate power for the prediction of type 2 diabetes. The best accuracy was 0.66 at the threshold between non-diabetic and diabetic status of 0.52 (non-diabetic status=0, diabetic status=1), for which the specificity and the sensitivity were 0.71 and 0.61, respectively.

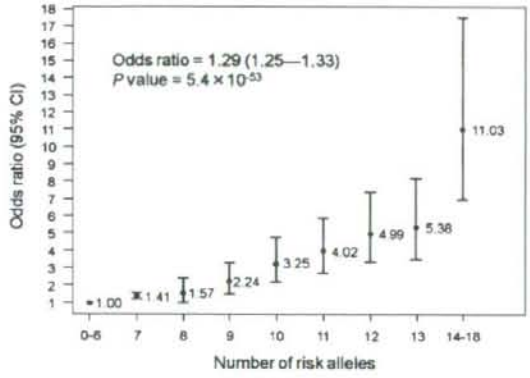

Figure 1 Odds ratios for subjects with different numbers of risk alleles for 11 susceptibility loci for type 2 diabetes. The cumulative effect of the 11 loci on type 2 diabetes was tested by counting the number of risk alleles associated with type 2 diabetes with a logistic regression model with adjustment for age, sex and BMI. The ORs for subjects with each number of risk alleles are expressed relative to individuals with 0–6 risk alleles.

DISCUSSION

By the validation of the results from our multistage GWAS, we detected only marginal associations of *EIF2AK4*, *KRT4* and *A2BP1* with type 2 diabetes in meta-analyses with two subject panels of Japanese or Chinese individuals. Relations of *KRT4* (keratin 4 gene) and *A2BP1* (ataxin-2-binding protein 1 gene, also known as *FOX1*) to glucose or lipid metabolism are unknown. Deletion of *EIF2AK4* (eukaryotic translation initiation factor 2 alpha kinase 4 gene, also known as *GCN2*) in mice resulted in liver steatosis during leucine deprivation as a result of unrepressed expression of lipogenic genes.[30] The functionally related gene, *EIF2AK3* (also known as *PERK* or *PEK*), has been shown to cause diabetes mellitus both in humans (Wolcott–Rallison syndrome, OMIM604032) and in rodent models.[31,32] Taken together, *EIF2AK4* may be a good candidate for the diabetes susceptibility gene. The sample size required for a statistical power of 0.80 with equal numbers of cases and controls is 10 505 when the frequency of the risk allele, OR and type I error probability are assumed to be 0.20, 1.10 (the value for *EIF2AK4* in the meta-analysis in Table 1) and 0.05, respectively. Further studies of these genes in other Asian populations as well as in other ethnic groups are needed for confirmation of their association with type 2 diabetes. Given this uncertainty, we did not include these genes in the assessments of cumulative risk and gene–gene interaction.

Among tens of type 2 diabetes susceptibility genes identified by recent GWASs in Caucasians, the associations of six genes, that is, *TCF7L2*, *CDKAL1*, *CDKN2A/B*, *IGF2BP2*, *SLC30A8* and *HHEX*, have been replicated in Asian populations as well as in populations of European ancestry. A recent meta-analysis in Japanese subjects also supported the associations.[12] In this study, we performed replication

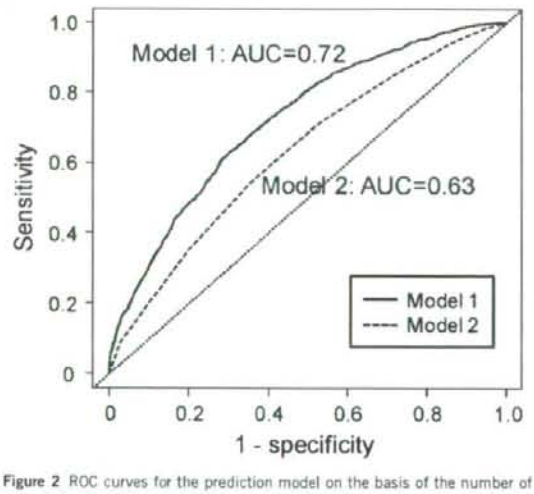

Figure 2 ROC curves for the prediction model on the basis of the number of risk alleles for 11 susceptibility loci for type 2 diabetes. The prediction model for type 2 diabetes was constructed using the logistic regression model, and ROC curves for the model were generated. In model 1, the number of risk alleles was used as an independent variable together with age, sex and BMI as covariates, whereas only the number of risk alleles was used as an independent variable in model 2.

study, and, on the basis of the results, we added five more genes, that is, *KCNJ11*, *PPARG*, *GCKR*, *KCNQ1* and *HNF1B*, for the cumulative risk assessment for type 2 diabetes. Thus, the SNPs of *HNF1B*, which were earlier associated with type 2 diabetes in Chinese as well as in Caucasians,[26] showed the association with the disease in the Japanese population in this study. In addition, the *C* allele of rs780094 in *GCKR* was associated with increased risk of type 2 diabetes in this study, which is consistent with a recent study in Caucasians.[33] The associations of *KCNJ11* and *PPARG* with diabetes were marginal in this study; however, they were included for the prediction model, as the associations were replicated in some studies of Caucasians.

Our gene–gene interaction analysis showed no significant interaction for any of the 55 possible pairs of genes when corrected for multiple testing. When the significance level was set at 0.05, two pairs were judged to be significant. However, such gene–gene interactions were not supported from the functional point of view. A large-scale study may provide more convincing evidence for such interactions.

As no confirmatory evidence for gene–gene interaction was observed, we treated the 11 genes as independent variables in the prediction model. The addition of one risk allele was estimated to increase the odds by an average of 1.29 according to the multiplicative model. This value is similar to that (1.24) estimated for type 2 diabetes in Caucasians.[17] Two earlier cumulative risk assessments for type 2 diabetes in Asian populations with relatively small numbers of associated loci yielded values of 1.17 and 1.24 for the fold increase in risk for each additional risk allele.[11,34] In our prediction model for type 2 diabetes, the AUC for the ROC curve was lower than that in the earlier study[17] based on 15 loci in Caucasians (0.72 and 0.86, respectively). However, the number of loci in our study (11 loci) was lower than that in the study for Caucasians. The inclusion of additional loci in our model should improve its ability to predict type 2 diabetes in Asian populations. Several reports of the prediction of type 2 diabetes using ~18 loci were recently described for populations of European ancestry.[35–38] A prediction based on 18 loci gave an AUC value of 0.80 for the ROC curve,[35] whereas the corresponding values for a population-based prospective study were 0.68,[36] 0.615[37] and 0.75.[38] They concluded that genetic variations associated with diabetes had a small effect on the ability to predict the development of type 2 diabetes as compared with clinical characteristics alone. In fact, the AUC value (0.72) based on both the genetic variations and the clinical characteristics was slightly better than that based on only the clinical characteristics (0.68). We admit that the evidence of the association with diabetes is a little weaker for *KCNJ11* and *PPARG* in the Japanese population than for the other nine genes. If *KCNJ11* and *PPARG* were excluded from the analysis, the AUC for the ROC curve in the prediction model incorporating age, sex and BMI remained unchanged at 0.72, probably because of the relatively large effects of *KCNQ1* and *TCF7L2*.

Finally, our prediction model for type 2 diabetes achieved limited success even though it has some value. Given that GWASs for diabetes in Asians have not been as extensive as those in Caucasians, many risk loci for diabetes in Asians remain most likely to be undiscovered. Considering that the average increase in OR conferred by each additional risk allele was similar between Caucasians and Japanese, incorporation of data from additional risk loci is most likely to increase the predictive power.

ACKNOWLEDGEMENTS

We thank all the subjects who joined this project; Sumio Sugano and Shoji Tsuji for support and helpful discussion throughout the project; and Megumi Yamaoka-Sageshima for technical assistance. This work was supported by a grant from the Program for Promotion of Fundamental Studies in Health Sciences of the Pharmaceuticals and Medical Devices Agency (PMDA) of Japan; a grant from the National Institute of Biomedical Innovation (NIBIO) of Japan; grants from the Ministry of Health, Labour and Welfare of Japan; a Grant-in-Aid for Scientific Research on Priority Areas (C), 'Medical Genome Science (Millennium Genome Project)', 'Applied Genomics', and 'Comprehensive Genomics', from the Ministry of Education, Culture, Sports, Science and Technology of Japan; and a grant from the Cooperative Link of Unique Science and Technology for Economy Revitalization (CLUSTER, Tokushima, Japan). The Hong Kong diabetes case–control study was supported by the Hong Kong Research Grants Committee Central Allocation Scheme CUHK 1/04C.

1 Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445,** 881–885 (2007).
2 Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316,** 1331–1336 (2007).
3 Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316,** 1336–1341 (2007).
4 Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316,** 1341–1345 (2007).
5 Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G. B. *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39,** 770–775 (2007).
6 Miyake, K., Horikawa, Y., Hara, K., Yasuda, K., Osawa, H., Furuta, H. *et al.* Association of TCF7L2 polymorphisms with susceptibility to type 2 diabetes in 4087 Japanese subjects. *J. Hum. Genet.* **53,** 174–180 (2008).
7 Horikawa, Y., Miyake, K., Yasuda, K., Enya, M., Hirota, Y., Yamagata, K. *et al.* Replication of genome-wide association studies of type 2 diabetes susceptibility in Japan. *J. Clin. Endocrinol. Metab.* **93,** 3136–3141 (2008).
8 Horikoshi, M., Hara, K., Ito, C., Shojima, N., Nagai, R., Ueki, K. *et al.* Variations in the HHEX gene are associated with increased risk of type 2 diabetes in the Japanese population. *Diabe* **50,** 2461–2466 (2007).
9 Furukawa, Y., Shimada, T., Furuta, H., Matsuno, S., Kusuyama, A., Doi, A. *et al.* Polymorphisms in the IDE-KIF11-HHEX gene locus are reproducibly associated with type 2 diabetes in a Japanese population. *J. Clin. Endocrinol. Metab.* **93,** 310–314 (2008).
10 Omori, S., Tanaka, Y., Takahashi, A., Hirose, H., Kashiwagi, A., Kaku, K. *et al.* Association of CDKAL1, IGF2BP2, CDKN2A/B, HHEX, SLC30A8, and KCNJ11 with susceptibility to type 2 diabetes in a Japanese population. *Diabetes* **57,** 791–795 (2008).
11 Ng, M. C., Park, K. S., Oh, B., Tam, C. H., Cho, Y. M., Shin, H. D. *et al.* Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2 and FTO in type 2 diabetes and obesity in 6719 Asians. *Diabetes* **57,** 2226–2233 (2008).
12 Tabara, Y., Osawa, H., Kawamoto, R., Onuma, H., Shimizu, I., Miki, T. *et al.* Replication study of candidate genes associated with type 2 diabetes based on genome-wide screening. *Diabetes* **58,** 493–498 (2009)10.2337/db07-1785..
13 Grarup, N., Rose, C. S., Andersson, E. A., Andersen, G., Nielsen, A. L., Albrechtsen, A. *et al.* Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10705 Danish subjects: validation and extension of genome-wide association studies. *Diabetes* **56,** 3105–3111 (2007).
14 Cauchi, S., Proença, C., Choquet, H., Gaget, S., De Graeve, F., Marre, M. *et al.* 2008 Analysis of novel risk loci for type 2 diabetes in a general French population: the D.E.S.I.R. study. *J. Mol. Med.* **86,** 341–348 (2008).
15 Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H. *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **40,** 1092–1097 (2008).
16 Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40,** 638–645 (2008).
17 Cauchi, S., Meyre, D., Durand, E., Proença, C., Marre, M., Hadjadj, S. *et al.* Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. *PLoS ONE* **3,** e2031 (2008).
18 Yang, X., So, W. Y., Kong, A. P., Ho, C. S., Lam, C. W., Stevens, R. J. *et al.* Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes Care* **30,** 65–70 (2007).
19 Ozaki, R., Qiao, Q., Wong, G. W., Chan, M. H., So, W. Y., Tong, P. C. *et al.* Overweight, family history of diabetes and attending schools of lower academic grading are independent predictors for metabolic syndrome in Hong Kong Chinese adolescents. *Arch. Dis. Child.* **92,** 224–228 (2007).
20 Bannai, M., Higuchi, K., Akesaka, T., Furukawa, M., Yamaoka, M., Sato, K. *et al.* Single-nucleotide-polymorphism genotyping for whole-genome-amplified samples

using automated fluorescence correlation spectroscopy. *Anal. Biochem.* **327**, 215–221 (2004).
21 Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
22 Grant, S. F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.* **38**, 320–323 (2006).
23 Gloyn, A. L., Weedon, M. N., Owen, K. R., Turner, M. J., Knight, B. A., Hitman, G. *et al.* Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* **52**, 568–572 (2003).
24 Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J. *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
25 Sandhu, M. S., Weedon, M. N., Fawcett, K. A., Wasson, J., Debenham, S. L., Daly, A. *et al.* Common variants in WFS1 confer risk of type 2 diabetes. *Nat. Genet.* **39**, 951–953 (2007).
26 Gudmundsson, J., Sulem, P., Steinthorsdottir, V., Bergthorsson, J. T., Thorleifsson, G., Manolescu, A. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
27 Horikoshi, M., Hara, K., Ito, C., Nagai, R., Froguel, P. & Kadowaki, T. 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 (2007).
28 Yokoi, N., Kanamori, M., Horikawa, Y., Takeda, J., Sanke, T., Furuta, H. *et al.* Association studies of variants in the genes involved in pancreatic beta-cell function in type 2 diabetes in Japanese subjects. *Diabetes* **55**, 2379–2386 (2006).
29 Mori, H., Ikegami, H., Kawaguchi, Y., Seino, S., Yokoi, N., Takeda, J. *et al.* The Pro12→Ala substitution in PPAR-gamma is associated with resistance to development of diabetes in the general population: possible involvement in impairment of insulin secretion in individuals with type 2 diabetes. *Diabetes* **50**, 891–894 (2001).
30 Guo, F. & Cavener, D. R. The GCN2 eIF2alpha kinase regulates fatty-acid homeostasis in the liver during deprivation of an essential amino acid. *Cell Metab.* **5**, 103–114 (2007).
31 Harding, H. P., Zeng, H., Zhang, Y., Jungries, R., Chung, P., Plesken, H. *et al.* Diabetes mellitus and exocrine pancreatic dysfunction in $Perk^{-/-}$ mice reveals a role for translational control in secretory cell survival. *Mol. Cell.* **7**, 1153–1163 (2001).
32 Delépine, M., Nicolino, M., Barrett, T., Golamaully, M., Lathrop, G. M. & Julier, C. EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott–Rallison syndrome. *Nat. Genet.* **25**, 406–409 (2000).
33 Sparsø, T., Andersen, G., Nielsen, T., Burgdorf, K. S., Gjesing, A. P., Nielsen, A. L. *et al.* The GCKR rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes. *Diabetologia* **51**, 70–75 (2008).
34 Wu, Y., Li, H., Loos, R. J., Yu, Z., Ye, X., Chen, L. *et al.* Common variants in CDKAL1, CDKN2A/B, IGF2BP2, SLC30A8 and HHEX/IDE genes are associated with type 2 diabetes and impaired fasting glucose in a Chinese Han population. *Diabetes* **57**, 2834–2842 (2008).
35 Lango, H., Palmer, C. N., Morris, A. D., Zeggini, E., Hattersley, A. T., McCarthy, M. I., *et al.* The UK Type 2 Diabetes Genetics Consortium Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. *Diabetes* **57**, 3129–3135 (2008).
36 van Hoek, M., Dehgan, A., Witteman, J. C., van Duijn, C. M., Uitterlinden, A. G., Oostra, B. A. *et al.* Predicting type 2 diabetes based on polymorphisms from genome wide association studies: a population-based study. *Diabetes* **57**, 3122–3128 (2008).
37 Meigs, J. B., Shrader, P., Sullivan, L. M., McAteer, J. B., Fox, C. S., Dupuis, J. *et al.* Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med.* **359**, 2208–2219 (2008).
38 Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T. *et al.* Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N. Engl. J. Med.* **359**, 2220–2232 (2008).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

Ⅱ．分担研究報告書

厚生労働科学研究費補助金（厚生労働科学特別研究事業）
平成２０年度分担研究報告書

遺伝子解析及びバイオマーカー測定

分担研究者　安田　和基　国立国際医療センター研究所・部長

概要：ヒト糖尿病の多彩な病態を解析し、画期的な診断・治療法を開発するために、疾患の「様々な階層」を解析できるような臨床試料リソースを収集した「バイオリソース」の構築を目指した。そのために、既に病態への関与が示唆されたSNPや生理活性物質、マーカーなどの「生体情報」について、収集した試料を測定（検査、タイピング）する「ボトムアップ型」の情報収集を行った。糖尿病患者のゲノムを用いて、我々が報告した*KCNQ1*をはじめとする11遺伝子の11SNPについてのタイピングを行った。また、教育入院の入・退院時のペア血清を用いて、血清タンパクのマルチプレックスな測定系を利用して、糖尿病・代謝に関連する14の項目（C-peptide、Insulin、Adiponectin、Leptin、Resistin、PAI-1、TNFα、Visfatin、Adipsin、Ghrelin、Glucagon、GLP-1、GIP、IL-6）の測定を行った。いくつかの未報告と思われる知見も得られつつあり、今後、よりサンプル数を増やして検証する必要はあるものの、こうしたバイオリソースの有用性および課題が、明らかになってきたと思われる。

（研究協力者：野田光彦：国立国際医療センター戸山病院　糖尿病・代謝症候群診療部長）

A. 研究目的

ヒト糖尿病の多彩な病態を解析し、画期的な診断・治療法を開発するためには、臨床の現場と密に連携した「バイオリソース」の構築が不可欠である。しかも多元的な解析、具体的には、DNA（ゲノム、エピゲノム）、タンパク・代謝物やその修飾（プロテオーム、グライコーム、メタボロームほか）、さらに相互作用から環境因子も含めて、最終的に個体レベルの表現型（フェノーム、及び病態）に至るまで、疾患の「様々な階層」を解析できるような臨床試料リソースが必要である。

網羅的な「トップダウン型」のオミックス解析は、事前に予想できない全く新

しい知見を与えてくれるが、得られた分子の機能的意義の検証には別の困難がともなう。一方、総括研究報告書にもあるように、既に病態への関与が示唆されたSNPや生理活性物質、マーカーなどの「生体情報」について、収集した試料を測定（検査、タイピング）する「ボトムアップ型」の情報収集は不可欠である。なぜなら、これまで研究により確立されてきた遺伝因子、あるいはバイオマーカーは、個々の意義は論じられているものの糖尿病の病態や治療経過とどのように関連するのか、複数の因子を同時に捉えた場合の意義はどうなのか、については解答が出ていない。特にトップダウン型の解析から得られた因子（GWASから得られた遺伝因子など）では、機能的意義が不明の点が多く、臨床情報と突き合わせた検証が必要である。収集したリソースの価値を確定し、その特徴を把握し分類しておくためには、できるだけ早急にこうしたSNPのタイピングや遺伝子スクリーニング、バイオマーカーの測定を行わなければならない。またGWASでも明らかになってきたように、糖尿病の病態や成因には、人種差が存在し、日本人についての多次元的な解析が必要である。
そこで、既に収集しているサンプルを用いて、既報マーカーの有用性の検証を試みるとともに、重層化した「糖尿病バイオリソース」の構築と、その有用性・問題点の検討を行った。すなわち、

1）糖尿病と関連することが報告されている既報の「生体情報」（特にSNPと血中生理活性物質濃度）について我々の試料で解析を行い、その意義を検討する。
2）上記「生体情報」付加により、我々の試料（バイオリソース）の科学的価値を高める。
3）上記「生体情報」の解析における、試料及び臨床情報の問題点を検討する。

最近確立された遺伝因子を含め、本研究により得られたデータは今後、国がテーラーメイド医療の研究・開発を進めるにあたり、貴重な基盤データとなると期待され、同時に日本人糖尿病を多元的に解析するための基盤リソースの構築にも役立つといえる。

B. 研究方法

我々は既に、糖尿病患者９００名以上、非糖尿病対照者（原則６０才以上）６００名以上について、研究への同意を得てゲノム及び血清を採取しており、

また教育入院前後（約２−３週間のインターバルで代謝状態は劇的に改善）の同一人ペア血清を約２００例収集している。この試料を対象に次のことを行う。（なお、これらと平行して、さらに臨床試料の収集も継続する。）

（１）ゲノム解析

多因子遺伝病としての２型糖尿病の遺伝因子の研究は、近年いわゆるGWAS（Genome-wide association study：ゲノムワイド相関解析）により、急速に進展した。我々が日本人２型糖尿病で最も重要な遺伝因子として同定した*KCNQ1*をはじめ、世界で１８ー２０の遺伝因子がほぼ確立している。これらは、個々の遺伝的効果はあまり大きくなく（オッズ比は最大でも*TCF7L2*と*KCNQ1*の1.4程度）、また特定のパスウェイに属するわけでもないため、単独では病態予測の意義は限られている。これらを合わせての発症リスク予測の検出力は、従来の危険因子を凌駕できないとされ、日本人についての我々の研究も同様の結果を得ている（Miyake et al 2009）。また遺伝的背景の不均一な患者群を一括して扱うことは、病態解析の上で大きな問題となるが、この点で整理・区分けする現時点で唯一の手段は、確立された遺伝因子をタイピングしておくことである。従って、遺伝因子と病態との関係、あるいは遺伝因子一環境因子相互作用を検討するためには、複数の遺伝因子を同時にタイピングしておく必要がある。

糖尿病の遺伝因子は、GWASを中心に「common disease-common variant」仮説にもとづいて解析されてきたが、得られた遺伝因子は、オッズ比が最大で１.５程度と低いため、より効果の強いrare variantの存在が想定される。上記のような多因子遺伝病タイプの遺伝因子の臨床的意義の正確な解析のためには、common variantの意義の検証が必須であるとともに、こうしたrare variantの同定が世界的にも急務になっており、諸外国では既に次世代シークエンサーをここに投入して研究が進められ始めているほどである。

１．我々がゲノムワイド相関解析（GWAS）により、最も重要な２型糖尿病遺伝子の一つとして最近報告した*KCNQ1*遺伝子のSNP（Yasuda K et al. Nature Genetics 2008）をはじめ、日本で糖尿病との関連が確認された１１の糖尿病感受性SNPについて、TaqMan probeを作成してタイピングし、genotypeを決定する（表１）。384サンプルを同時測定可能な機器（ABI 7900HT）を用いて、かなり簡便におこなうことが可能である。

２．収集試料のうち、３５才以下発症の糖尿病を対象として、報告されているMODY遺伝子について、全エクソンのリシークエンス解析を行う。こ

れにより、新規のMODY遺伝子異常が同定される可能性が期待され、rare variants による糖尿病の特徴を明らかにすることが可能になる。

このような解析は、複数回に分けて行うよりも、今回のようにできる限り多くの検体について、重要と思われる遺伝子・SNP を対象に短時間に集中的に行うことで、貴重な臨床試料や、反応試薬の使用量を最小限に抑えることができ、コスト、時間、労力の面できわめて効率的であるとともに、かつ最も信頼性の高いデータを得ることができる。

（2）血清を用いた解析

血清は、病態を反映するバイオマーカーの宝庫であるが、現在のプロテオミクス技術では、微量タンパクを網羅性よく解析することは難しい。一方で、代謝や炎症関連などの生理活性物質は多数報告されており、個別の測定よりもその組み合わせが、よりよく病態を反映する可能性が指摘されている。

使用した方法は、バイオラッド社 BioPlex システムを用いた multiplex 測定である。これは、複数の特異抗体をビーズに固定し、多項目を同時かつ簡便に測定するものであり、従来の方法と比較して、はるかに微量の試料を用いて、多項目を短時間に測定できる利点があるため、重要なヒト由来臨床試料の測定には特に有用である。最近、特定の病態や機能に関連した複数の項目に対する抗体があらかじめ固定されたビーズが市販されて入手可能となり、今回の目的に有用と考えた。複数の項目を同一試料で測定することで、隠された病態や新たなマーカーの抽出が可能性となることもありうる。

今回当初計画したアッセイは以下の通りである（表２）。

1）Bio-Plex Pro Diabetes Assay Panel：糖尿病及び代謝疾患に関連する１２項目を同時測定（12-plex）し、また Adiponectin・Adipsin を測定する。同時に測定できる１２項目は、以下の通りである。血中インスリン動態をあらわす C-peptide 及び Insulin・脂肪組織由来のアディポサイトカイン（アディポカイン）に分類される、Leptin、Resistin、PAI-1、TNFα、Visfatin。代謝関連ホルモン及びインクチンとして、Ghrelin、Glucagon、GLP-1、GIP、及び代謝だけでなく炎症とも関連する IL-6 である。同一患者の教育入院前後の血清で比較し、①短期間の治療による効果、②使用した薬剤の影響、③退院後の血糖コントロール維持、などとの関係を調べる。

2）Bio-Plex Pro Angiogenesis Panel：血管新生に関係する液性因子９項目を

同時測定する。糖尿病の細小血管障害、特に網膜症は脆弱な新生血管が形成され、局所において血管新生促進因子（VEGF、Angiopoientin-2、Erythropoietin など）と抑制因子（PEDF など）のバランスが崩れると考えられている。これが血中マーカーに反映するかどうかは十分調べられておらず、今回「網膜症あり」「血糖コントロール不良だが網膜症なし」「血糖コントロール良好で網膜症なし」の３群で比較する計画を立てた。

3）Bio-Plex Pro サイトカインアッセイキット及び Bio-Plex Pro Acute Phase アッセイキット：これらは、炎症・免疫関連の因子を多項目同時測定できる。最近糖尿病・代謝疾患において、脂肪組織へのマイクロファージの浸潤および活性化・脂肪組織からの種々のサイトカイン産生が知られている。また、その合併症としての動脈硬化も「慢性炎症性疾患」と考えられつつあり、炎症性タンパクである CRP などが疫学的に動脈硬化のバイオマーカーとして期待されているが、系統的に検索された研究は乏しい。そこで、「糖尿病 vs 非糖尿病」、「心疾患あり vs なし」、「腎症あり vs なし」、などとで比較する計画を立てた。

最近本システムはポリスリレンビーズから磁気ビーズに変更となり、磁気洗浄装置を用いた簡便な自動洗浄装置も導入したため、スループット高いスクリーニングも可能となった。

なお実際には、納品の遅れ、測定系の立ち上げなどの問題もあり、研究期間内には Diabetes Assay kit を先行して使用した。

（3）臨床情報

これについては既に糖尿病の必須項目として収集してある。上述した SNP タイピングデータ、生理活性物質やバイオマーカー濃度のデータなどとともに、改めてデータベース化しておく。

（倫理面への配慮）

ヒトゲノム及び血清の収集、解析については、「ヒトゲノム・遺伝子解析研究に関する倫理指針」「臨床研究に関する倫理指針」に準拠し、国立国際医療センター倫理審査委員会の承認を得ており、インフォームドコンセントの取得、試料の匿名化等による個人情報の保護に十分留意するなど、既取得サンプルを

含め倫理面には十分な配慮を行っている。

C. 研究結果と D. 考察

限られた期間であったため、ここでは、主に血中タンパク測定の部分を中心に報告する。

（1）ゲノム解析

1．2型糖尿病感受性11SNPのタイピング

既報に基づき特に我々が日本人糖尿病との相関をこれまでに確認した11SNPについて、糖尿病者のサンプルのDNAでTaqManプローブを用いてタイピングを行った。いくつかのSNPについては、genotypeのclusteringに苦労したが、概ねタイピングは良好であり、そのアリル頻度も我々の既報（Yasuda et al 2008; Miyake et al 2008; Horikawa et al 2008）とほぼ一致した。現在までのところ、各SNPのgenotype別にみて、発症年齢、既往最大BMI、治療法（インスリン治療の有無）、心血管障害の合併の有無、などに、明らかな差は見られていない。

まだ全症例のタイピングが進行中であり、今後複数のSNPの組合せなどにより、遺伝因子の意義を評価したい。また、今後臨床的なデータを含めた解析を行う場合も、こうした遺伝因子情報を基盤にしてサンプルを分類することも考えられる。

2．若年発症糖尿病を対象とした候補遺伝子リシークエンシング

35才発症の糖尿病約35名を対象に、まず日本人MODYでもっとも報告の多いMODY3（*HNF1A*）のエクソンリシークエンシングを行った。まだ進行中であるが、いくつか新規SNPと思われる多型はみられるものの、病態との関連は否定的であり、現在アミノ酸変異を伴う新規の遺伝子変異は同定されていない。今後、MODY1（*HNF4A*）、MODY5（*HNF1B*）の順に、解析を進めてゆく

今後の方向性として、エクソン部分に限定しないリシークエンシングが期待される。これは、イントロンや非翻訳領域などの多型、変異が疾患発症に効果を持つ例が非常に多いからである。また日本人の場合、臨床的にMODYと診断される症例でも、その半分以上で、既報のMODY遺伝子に

明らかな変異が見られないとされている。こうした症例の集積から、より網羅的なゲノム解析（全ゲノムリシークエンシングを含む）などにより、新規のMODY遺伝子、糖尿病の病態に重要な遺伝子が同定される可能性があり、この点でも、将来非常に有用なリソースとして注目される。

（2）血中バイオマーカー解析

先行して使用したDiabetes Assay Panelについて、preliminaryな測定結果を紹介する。

このパネルに含まれる14項目のうち、Insulin、C peptideは、臨床的にもよく測定され、内因性インスリン分泌能（特に空腹時や食後）やインスリン抵抗性（HOMA-IR）などの指標とされてきた。インスリン投与がなされている場合、Insulin値と、内因性分泌を反映するC peptideとを比較することで有用な情報が得られることもある。しかし、これらの血中の濃度は分泌と抵抗性との動的なバランスにより決まること、治療法（インスリン分泌促進薬や、インスリン治療の有無）により変動し、必ずしも内因性の代謝状態ばかりを反映しないこと、などが問題である。

Glucagonは、血糖調節についてはインスリンと拮抗する膵島ホルモンであるが、1型糖尿病では高値を示すのに対し、2型糖尿病では高目という報告が多いものの、必ずしも傾向が一致せず病態との関係は十分調べられていない。これは、膵β細胞特異的な破壊が生じる1型糖尿病と異なり、2型糖尿病では膵島全体が萎縮・変性する場合もあるからかもしれない。今回、Glucagonの値は非常に広い分布を示したが、同一人では入退院時にほぼ同程度の値を示し、個体差の重要性を示した。これは代謝による影響よりも、膵島の変性状態を反映する可能性が高い。

脂肪組織由来の液性因子の中で、Adiponectinは、インスリン感受性作用、抗動脈硬化作用などをもつ「善玉」、TNFαやResistinは、インスリン抵抗性や催炎症作用をもつ「悪玉」とされる。一方Leptinのように作用が多彩で、こうした「善悪」の軸でとらえられないものもある。これらは、基盤的研究によりその生理学的意義は、よく研究されているが、血中濃度と病態との関係、複数のマーカー同士の関係などについては、必ずしも十分わかっていなかった。