

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

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,  $P=5.4 \times 10^{-53}$ ). 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* (2009) 54, 236–241; doi:10.1038/jhg.2009.17; published online 27 February 2009

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

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

<sup>20</sup>Current address: Information Center for Medical Sciences, Tokyo Medical and Dental University, Tokyo, Japan.

<sup>21</sup>Current address: Faculty of Medical and Pharmaceutical Sciences, Department of Medical Biochemistry, Kumamoto University, Kumamoto, Japan.

<sup>22</sup>Current address: Department of Internal Medicine, Akita University School of Medicine, Akita, Japan.

<sup>23</sup>Current address: Kansai Electric Power Hospital, Osaka, Japan.

<sup>24</sup>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; published online 27 February 2009

## INTRODUCTION

Genome-wide association studies (GWASs) have identified novel susceptibility genes for type 2 diabetes mellitus in Caucasians.<sup>1–5</sup> *TCF7L2*, *CDKAL1*, *CDKN2A/B*, *IGF2BP2*, *SLC30A8* and *HHEX* have been widely replicated as susceptibility genes for type 2 diabetes in Asian populations<sup>6–12</sup> as well as in populations of European ancestry.<sup>13,14</sup> 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.<sup>15</sup> *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),<sup>15</sup> which was similar to that demonstrated earlier for *TCF7L2* in the Japanese population.<sup>6</sup>

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<sup>16</sup> and an analysis of gene–gene interaction for susceptibility genes<sup>17</sup> 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.<sup>15</sup> 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<sub>1c</sub> 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;<sup>5,18</sup> 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<sup>-1</sup>); 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.<sup>19</sup> 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,<sup>15</sup> 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<sup>20</sup> 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  $\chi^2$  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<sup>15</sup> 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.<sup>15</sup> 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<sup>21</sup> 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.<sup>15</sup> 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.<sup>15</sup> 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 ( $=5 \times 10^{-7}$ ).

### 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*,<sup>1–5</sup> *TCF7L2*,<sup>22</sup> *KCNJ11*,<sup>23</sup> *PPARG*,<sup>24</sup> *WFS1*,<sup>25</sup> *HNF1B*<sup>26</sup> and *KCNQ1*,<sup>15</sup> 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.<sup>6,15,27</sup> 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.<sup>7–9</sup>

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<sup>6–9,15,27–29</sup> 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,<sup>7</sup> was found to be associated with the disease in this enlarged Japanese panel ( $P=1.7 \times 10^{-5}$ ; 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*)  $\times$  rs1470579 (*IGF2BP2*) (nominal  $P=0.0025$ ) and rs1801282  $\times$  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,  $0.05/55=9.1 \times 10^{-4}$ ). 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  $4.7 \times 10^{-56}$  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,  $P=5.4 \times 10^{-53}$ , logistic regression analysis).

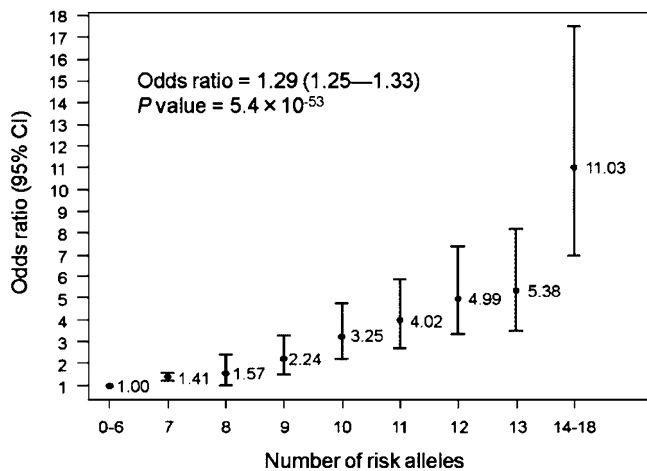
To predict disease status for type 2 diabetes in a given individual, we constructed a prediction model on the basis of the number of risk alleles or the liability value calculated from the number of risk alleles as well as age, sex and BMI. The coefficients to calculate the liability value were estimated 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.

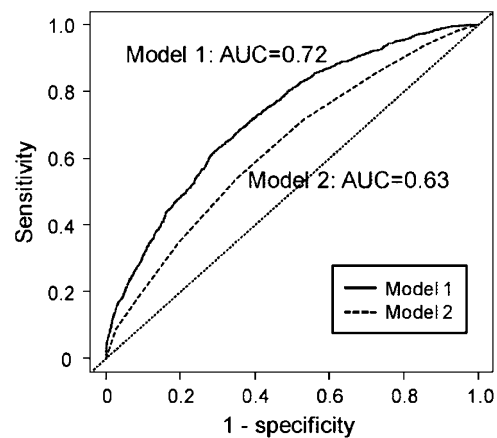
### DISCUSSION

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.<sup>30</sup> 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.<sup>31,32</sup> 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.<sup>12</sup> In this study, we performed replication 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,<sup>26</sup> showed the association with the disease in the Japanese



**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.



**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.

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.<sup>33</sup> 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.<sup>17</sup> 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.<sup>11,34</sup> In our prediction model for type 2 diabetes, the AUC for the ROC curve was lower than that in the earlier study<sup>17</sup> 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.<sup>35–38</sup> A prediction based on 18 loci gave an AUC value of 0.80 for the ROC curve,<sup>35</sup> whereas the corresponding values for a population-based prospective study were 0.68,<sup>36</sup> 0.615<sup>37</sup> and 0.75.<sup>38</sup> 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.

- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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..
- Grup, 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Bannai, M., Higuchi, K., Akasaka, 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).
- Sing, T., Sander, O., Beerwinkel, N. & Lengauer, T. ROCr: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
- 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<sup>-/-</sup>* 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., Wittman, 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>)