Genome-wide association study identifies three novel loci for type 2 diabetes

Kazuo Hara1,2,∗, Hayato Fujita1,†, Todd A. Johnson3,‡, Toshimasa Yamauchi1,§, Kazuki Yasuda8, Momoko Horikoshi1, Chen Peng9, Cheng Hu10, Ronald C. W. Ma11,12,13, Minako Imamura4, Minoru Iwata15, Tatsuhiko Tsunoda3, Takashi Morizono3, Nobuhiro Shojima1, Wing Yee So11,12,13, Ting Fan Leung14, Patrick Kwan11, Rong Zhang10, Jie Wang10, Weihui Yu10, Hiroshi Maegawa16, Hiroshi Hirose17, DIAGRAM consortium18, Kohei Kaku19, Chikako Ito20, Hirotaka Watada7, Yasushi Tanaka21, Kazuyuki Tobe15, Atsunori Kashiwagi16, Ryuzo Kawamori7, Weiping Jia10, Juliana C. N. Chan11,12,13, Tik Ying Teo9,22,23,24,25, Tai E. Shyong9,25,26, Naoyuki Kamatani28, Michiaki Kubo5, Shiro Maeda4 and Takashi Kadowaki1

1Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan, 2Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, the University of Tokyo, Tokyo 113-8655, Japan, 3Laboratory for Medical Science Mathematics, 4Laboratory for Endocrinology, Metabolism, and Kidney Diseases and 6Research Group for Genotyping, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan, 5Sportology Center, Graduate School of Medicine and 7Department of Medicine, Metabolism and Endocrinology, School of Medicine, Juntendo University, Tokyo 113-8421, Japan, 8Department of Metabolic Disorder, Diabetes Research Center, National Center for Global Health and Medicine, Research Institute, Tokyo 162-8655, Japan, 9Saw Swee Hock School of Public Health, National University of Singapore, MD3, 16 Medical Drive, Singapore 117597, Singapore, Singapore, 10Shanghai Jiao Tong University Affiliated Sixth People’s Hospital, Shanghai Diabetes Institute, Shanghai Key Laboratory of Diabetes Mellitus, 600 Yishan Road, Shanghai 200233, China, 11Department of Medicine and Therapeutics, Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China, 12Li Ka Shing Institute of Health Sciences, 13Hong Kong Institute of Diabetes and Obesity and 14Department of Paediatrics, Chinese University of Hong Kong, Hong Kong, China, 15First Department of Internal Medicine, University of Toyama, Toyama 930-0194, Japan, 16Department of Medicine, Shiga University of Medical Science, Otsu, Shiga 520-2192, Japan, 17Health Center, Keio University School of Medicine, Tokyo 160-8582, Japan, 18For a Full List of Authors, see Supplementary Material and 19Division of Diabetes, Endocrinology and Metabolism, Department of Internal Medicine, Kawasaki Medical School, Kurashiki, Okayama 701-0192, Japan, 20Medical Court Life Care Clinic, Hiroshima 730-0012, Japan, 21Department of Internal Medicine, Division of Metabolism and Endocrinology, St. Marianna University School of Medicine, Kawasaki, Kanagawa 216-8511, Japan, 22Life Sciences Institute, 23NUS Graduate School for Integrative Science and Engineering, 24Department of Statistics and Applied Probability and 25Department of Medicine, National University of Singapore, Singapore, Singapore, 26Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore, 27Duke-National University of Singapore Graduate Medical School, Singapore, Singapore and 28StaGen, Tokyo 111-0051, Japan

Received March 20, 2013; Revised and Accepted August 9, 2013

Although over 60 loci for type 2 diabetes (T2D) have been identified, there still remains a large genetic component to be clarified. To explore unidentified loci for T2D, we performed a genome-wide association study (GWAS) of 6209637 single-nucleotide polymorphisms (SNPs), which were directly genotyped or imputed using East Asian

∗To whom correspondence should be addressed at: Department of Diabetes and Metabolic Diseases, University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 115-0055, Japan. Tel: 81 358009161; Fax: 81 358009161; Email: thara-tky@umin.ac.jp
†These authors contributed equally to this work.

© The Author 2013. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
Introduction

T2D is a complex disease characterized by hyperglycemia resulting from impaired pancreatic β-cell function and a decreased action of insulin on target tissues (1,2). Familial aggregation and twin studies have shown that a genetic component plays a major role in the onset of T2D. Although there has been a marked increase in the identification of genetic loci for T2D, with over 60 of the discoveries made through genome-wide association studies (GWASs), it is estimated that at best 10% of the genetic component of T2D can be explained by the loci identified so far (3–5). The 1000 Genomes Project was founded with the aim of characterizing over 95% of variants with a minor allele frequency (MAF) >1% to provide a more extensive catalog of the genetic variations in major ethnic groups (6,7). We reasoned that we could take advantage of this recent advance to explore the remaining unidentified part of the genetic component of T2D by extending our search. This idea is supported by a recent article that reported that imputing the Wellcome Trust Case Control Consortium Phase I genotype data by using references from the 1000 Genomes Project detected loci for diabetes that were not identified in the original case–control study (8). In the meantime, genetic studies have failed to explain the reason why the prevalence of T2D differs across different ethnic groups. In light of the high and rising prevalence of T2D in East Asia (9), there is a huge need for exploring genetic variations vigorously in East Asian populations, although GWASs for T2D have been performed not only in European (10–15), but also in Asian (16–22) populations.

Results

GWAS of typed and imputed SNPs using 1000 Genomes Project data

This is a three-stage study comprised of (i) a discovery GWAS, (ii) follow-up analysis and (iii) validation analysis. During the discovery GWAS, we obtained directly genotyped data with a 610K SNP array in 5976 Japanese patients with T2D and 20829 nondiabetic individuals. Nineteen unreported loci were selected and taken forward to follow-up analyses. Combined discovery and follow-up analyses (30,392 cases and 34,814 controls) identified three new loci with genome-wide significance, which were MIR129-LEP (rs791595; risk allele = A; risk allele frequency (RAF) = 0.080; P = 2.55 × 10^{-13}; odds ratio (OR) = 1.17), GPM1 (rs11787792; risk allele = A; RAF = 0.874; P = 1.74 × 10^{-10}; OR = 1.15) and SLC16A13 (rs312457; risk allele = G; RAF = 0.078; P = 7.69 × 10^{-13}; OR = 1.20). This study demonstrates that GWASs based on the imputation of genotypes using modern reference haplotypes such as that from the 1000 Genomes Project data can assist in identification of new loci for common diseases.

Replication and validation of top SNPs selected in the discovery GWAS

A top SNP at each of the two unreported loci was taken forward to follow-up analysis by in silico (2799 cases and 3793 controls) and de novo (10,319 cases and 6795 controls) genotyping in East Asian populations (Supplementary Material, Table S2). By combining the results obtained in the first discovery GWAS and the second follow-up analyses (in total 19,094 cases and 31,417 controls), four novel SNPs reached genome-wide significance (P < 5 × 10^{-8}), which were near MIR129-LEP (rs791595; P = 5.46 × 10^{-11}; OR = 1.19), GPM1 (rs11787792; P = 7.26 × 10^{-11}; OR = 1.18), MRPS35 (rs7316898; P = 7.36 × 10^{-9}; OR = 1.10) and SLC16A13 (rs312457; P = 2.15 × 10^{-8}; OR = 1.18) (Supplementary Material, Table S2). To validate these associations, the four top SNPs were subsequently genotyped further in another 11,298 cases and 3397 controls. We confirmed that MIR129-LEP (rs791595; risk allele = A; RAF = 0.08; P = 2.55 × 10^{-13}; OR = 1.17), GPM1 (rs11787792; risk allele = A; RAF = 0.874; P = 1.74 × 10^{-10}; OR = 1.15) and SLC16A13 (rs312457; risk allele = G; RAF = 0.078; P = 7.69 × 10^{-13}; OR = 1.20) were associated with T2D, reaching a Bonferroni-adjusted P-value for significance of 8.05 × 10^{-9}, which is more stringent than the conventional cut-off of 5 × 10^{-8} (Table 1) (Fig. 2). In contrast, rs7316898 in MRPS35 did not reach genome-wide significance (P = 5.42 × 10^{-5}; OR = 0.94) after combining three-stage analyses. Rs14768973 in TNKS2 also reached genome-wide significance (P = 2.43 × 10^{-9}) (Supplementary Material, Table S2). However, conditioning for rs12219514 in the previously
reported locus \textit{HHEX} \cite{13,14} abrogated the signal at rs147689733, showing that it was a proxy for the stronger \textit{HHEX} signal. Therefore, this SNP was not taken forward for validation analysis. To further confirm our imputation-based results, we performed additional direct genotyping of the three novel SNPs in a subset of the discovery GWAS samples that consisted of \sim 2600 samples and observed almost perfect concordance with the imputed data (Supplementary Material, Table S3). Combined with the direct genotyping performed in the other two analysis stages, the three novel SNPs were directly genotyped in over 34 000 samples.

Similarities and differences between East Asians and Europeans

To briefly assess the similarity of genetic architecture between East Asians and European populations, we compared the reported effect sizes of loci reported first in Europeans and those obtained in the present study (Supplementary Material, Table S4). A high concordance in the direction of the effects and correlation of ORs between the two populations was observed ($r=0.49$, $P=0.0018$), showing that the two populations shared the same susceptibility genes when the data are restricted to common variants identified by previous GWASs. As for the novel loci identified in the present study, there is no evidence for an association between SNPs in and near \textit{MIR129}-\textit{LEP} (Supplementary Material, Fig. S3A) and \textit{SLC16A13} (Supplementary Material, Fig. S3C) and only modest evidence for association around \textit{GPSM1} (Supplementary Material, Fig. S3B) in European populations \cite{15}. Rs4731420, being in perfect linkage disequilibrium (LD) with rs791595 near \textit{MIR129}-\textit{LEP} ($r^2=1.00$), was not associated with T2D in the DIAGRAMv3 ($P=0.760$). There are no available data on rs312457 in \textit{SLC16A13} and its proxy in European populations. However, rs11652868, which was in modest LD with another SNP that was associated with T2D in the present study (rs312458), was not associated with T2D in the European data. The \textit{GPSM1} locus has been reported to be linked to diabetes-related traits \cite{23,24}; rs3829109, located in the adjacent gene \textit{DNLZ}, was previously associated with the fasting glucose level \cite{23}, and rs60980157, a nonsynonymous SNP (Ser391Leu of \textit{GPSM1}), was previously reported to be associated with an insulin secretion measure, the insulinogenic index \cite{24}. These two SNPs were in modest LD with rs11787792 ($r^2=0.127$ for rs3829109 and 0.344 for rs60980157) and were associated with T2D ($P=6.10 \times 10^{-6}$ and $9.03 \times 10^{-6}$, respectively) in this study, but we did not detect any association between rs11787792 and fasting glucose level or an insulin secretion index, HOMA-\beta (Supplementary Material, Table S5).

Comparison between association statistics of typed and imputed SNPs

Supplementary Material, Figure S4, demonstrates the usefulness of using imputation to explore previously unknown loci; imputed SNPs (grey circles) were more significantly associated with T2D than typed (red circles) SNPs. The signal at \textit{GPSM1} would not have been taken forward to follow-up analysis, and therefore, it would have been missed (Supplementary Material, Fig. S4). Moreover, rs11787792 in \textit{GPSM1} is included only in 1000 Genomes Project data and not in HapMap2 data. We also sought to define the most relevant SNPs for susceptibility to T2D in previously identified loci. We found rs7656416 in \textit{CTBP1} ($P=1.29 \times 10^{-8}$) to be more significantly associated with T2D than the previously reported SNP, rs6815464 in \textit{MAEA} ($P=1.32 \times 10^{-5}$) \cite{20}. Rs7656416 and rs6815464 were in LD ($r^2=0.58$), and conditioning for rs7656416 in the logistic regression abrogated the signal at rs6815464.

BMI-stratifying analysis

It has been reported that the genetic predisposition to T2D is different in lean subjects compared with findings for obese subjects \cite{25,26}. We tested the association between the top SNP in each of the eight autosomal known loci with $P<5 \times 10^{-8}$ (adjusted for sex, age, and the first four principal components from PCA in the present discovery analyses) after dichotomizing the T2D subjects into lean (body mass index, BMI $<25$ kg/m$^2$) and overweight (BMI $\geq 25$ kg/m$^2$) groups. We found that the top SNPs in the known loci were more significantly associated with T2D...
in lean subjects than in overweight subjects (Supplementary Material, Fig. S5). Several loci (KCNQ1, CDC123, IGF2BP2 and CDKAL1) had large enough heterogeneity z-scores to suggest that a substantial difference may exist in the association statistics between lean and overweight groups (Supplementary Material, Table S6).

Table 1. Three new T2D loci reaching genome-wide significance from combined analysis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Chromosome position</th>
<th>Nearby gene</th>
<th>Risk allele</th>
<th>Other allele</th>
<th>RA1 First stage</th>
<th>OR (95%CI)</th>
<th>P-value</th>
<th>RA2 Second stage</th>
<th>OR (95%CI)</th>
<th>P-value</th>
<th>RA3 Third stage</th>
<th>OR (95%CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs791595</td>
<td>5</td>
<td>7976 862 866</td>
<td>MIR129-LEP</td>
<td>A</td>
<td>G</td>
<td>0.08</td>
<td>1.19 (1.11–1.28)</td>
<td>0.00002</td>
<td>1.16 (1.06–1.28)</td>
<td>0.00003</td>
<td>1.17 (1.12–1.22)</td>
<td>0.00002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11787792</td>
<td>9</td>
<td>139 26 52 148</td>
<td>GPSM1</td>
<td>A</td>
<td>G</td>
<td>0.04</td>
<td>1.17 (1.09–1.25)</td>
<td>0.00001</td>
<td>1.15 (1.09–1.20)</td>
<td>0.00006</td>
<td>1.16 (1.11–1.22)</td>
<td>0.00002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs312457</td>
<td>17</td>
<td>69 40 393</td>
<td>SLC16A13</td>
<td>G</td>
<td>A</td>
<td>0.05</td>
<td>1.16 (1.10–1.22)</td>
<td>0.00001</td>
<td>1.15 (1.10–1.20)</td>
<td>0.00003</td>
<td>1.16 (1.11–1.22)</td>
<td>0.00001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Regional plots of the three newly discovered T2D loci: (A) MIR129-LEP, (B) GPSM1 and (C) SLC16A13. Regional associations were plotted for the three novel loci that showed genome-wide significance after combing the results from Stages 1, 2 and 3. Genotyped and imputed SNPs are plotted with the P-values (as $-\log_{10}$ values) from discovery analysis versus their physical position (NCBI Build 37). In each panel, the top SNP is represented by a purple diamond and P-values derived by combining first + second + third stage results are shown. P-values of other SNPs were derived from Stage 1 result alone and are color coded according to their pairwise LD with the top SNP based on 1000 Genomes Project East Asian reference data. Estimated recombination rates are plotted to reflect the local LD structure.
Sex-stratifying analysis

We performed a sex-differentiated analysis to test for sexual dimorphism of associations between SNPs and T2D, allowing for heterogeneity in allelic effects between males (3938 cases and 9553 controls) and females (183 cases and 9220 controls) (Supplementary Material, Table S7). We found modest evidence for sexual dimorphism in 23 genes and regions (heterogeneity $P < 1 \times 10^{-8}$) (Supplementary Material, Table S7). We did not find evidence of sex heterogeneity at previously reported loci such as KCNQ1, DGKB, and GRB14, nor present novel loci.

DISCUSSION

We identified three novel T2D loci in East Asian populations: MIR129-LEP, GPM1 and SLC16A13. GPM1 locus could not have been identified as a T2D locus without imputation using the 1000 Genomes Project reference data, as shown in Supplementary Material, Figure S4. This demonstrates the utility of using up-to-date reference haplotype data such as that from the 1000 Genomes Project to perform imputation to identify novel loci for T2D. Given the fact that whole-genome sequencing for thousands of samples is still highly cost intensive, an imputation-based GWAS is still useful to search for novel common disease loci. At MIR129-LEP and SLC16A13 loci, imputed SNPs were more significantly associated with type 2 diabetes (T2D) than directly genotyped SNPs, but imputation was not mandatory for the identification of these loci, because a directly genotyped SNP reached the $P$-value of $1 \times 10^{-4}$, the cut-off for follow-up analysis. Those two loci could have been found because of improved statistical power by increasing the sample size from ~7000 to 27 000 in the first-stage analysis.

We found no association between the three novel genes identified in this study and T2D in European populations (DIAGRAMv3) (15). We think there could be three sources for the lack of evidence for associations in DIAGRAM. First, in general, the difference in allele frequencies would cause discrepancy between the association results among different ethnic groups. A lower allele frequency reduces statistical power to detect association. However, we found no apparent differences in the allele frequencies of the top SNPs and those in LD with the top SNP between the East Asian populations and European populations. Taking into account that the DIAGRAM conducted the largest GWAS for T2D in European populations covering ~2.5 million SNPs, it is unlikely that the DIAGRAM did not find our loci due to the lack of statistical power. Secondly and most likely is dissimilarities in the patterns of LD among populations, which could lead to a substantial difference in the strength of LD between a causal variant and its proxy SNP between our population and European populations. While in GPM1 locus, which showed modest association with T2D in DIAGRAMv3 (Supplementary Material, Fig. S3B), CEU and JPT + CHB show quite similar haplotype structure using HapMap Project data (Supplementary Material, Fig. S6B), while the haplotype structure around the other two loci appears to be notably different between the two population samples (Supplementary Material, Fig. S6A and S6C). In particular, the region around SLC16A13 shows a large block in which many SNPs are in complete LD ($D' = 1$). Further examination using the HapMap project’s phased haplotype browser shows a single high-frequency haplotype extending roughly 20 kb in each direction from rs321457 in the JPT or CHB haplotypes that is not present in the CEU haplotypes (data not shown).

Lastly and least likely, it is possible that a difference in effect size could lead to a situation where a disease locus would be discovered in a particular group with a strong effect. But no such locus, at least for T2D, has been reported so far, and effect sizes generally appear to be comparable among different ethnic groups. Further study will be needed to clarify whether three loci detected in this study are specific to East Asians.

Of the three newly identified loci, rs791595 is located between MIR129-1 and LEP. The coding product of LEP, leptin, plays a critical role in the regulation of body weight by inhibiting food intake and stimulating energy expenditure, and its deficiency in mice and humans causes morbid obesity and diabetes (27,28). Leptin is a hormone produced and secreted by white adipose tissue, and its circulating levels are closely related to body fat mass (29). Thus, LEP is one of the most plausible candidates for a susceptibility gene of obesity and T2D, although we could not find any association with BMI in the controls, and adjustment for BMI did not influence the strength of the association with T2D ($P = 9.42 \times 10^{-6}$). Instead, we did find rs791595 to be significantly associated with the homeostasis model assessment of insulin resistance (HOMA-IR) index (30) ($P = 0.005$) (Supplementary Material, Table S5). As for MIR129-1, mir129, the mature product of MIR129–1, has been reported to be up-regulated in bladder cancers in accordance with the down regulation of its targeted genes SOX4 and GALNT1, which are involved in cell death processes (31). However, whether mir129 has any role in tissues relevant to T2D is unknown. The Genomic Evolutionary Rate Profiling scores for the position of rs791595 near MIR129-LEP is 2.96, indicating that this site may be under evolutionary constraint.

The coding product of G-protein signaling modulator 1 isoform (GPM1) influences the basal activity of G-protein signaling systems through interaction with G-protein subunits (32). Gpm1 null mice have a lean phenotype with reduced fat mass and increased nocturnal energy expenditure (33), suggesting that GPM1 is a biologically plausible obesity gene. We could not find any association with BMI ($P = 0.11$) and adjustment for BMI did not influence the strength of the association with T2D.

SLC16A13 encodes solute carrier family 16, member 13, which is one of the monocarboxylate transporters (MCTs) (34). The first four MCTs (MCT1-4) catalyze the transport of monocarboxylates, such as lactate and pyruvate, but the functions of MCT13 (encoded by SLC16A13) and MCT11 (encoded by SLC16A11, which is located adjacent to SLC16A13) are unknown, except for one report that showed that intestinal expression of SLC16A13 was up-regulated by peroxisome proliferator-activated receptor-α agonists (35). Rs312457 in SLC16A13 is in perfect LD with rs17203120 ($R^2 = 0.86$, $P = 6.71 \times 10^{-6}$, OR = 1.19), which is located between SLC16A13 and SLC16A11. There is evidence that rs17203120 is associated with the expression level of SLC16A11 in lymphoblastoid cell lines (36).

There was no low-frequency variant with MAF ranging from 1 to 5% that reached genome-wide significance among 967 419 variants tested for association with T2D in this study. The present study yielded ~80% power to detect variants with an
MAF of 1% and OR of 1.6 and retained 80% power to detect association for variants with an MAF of 1% and OR of 2.0 even when imputed with a minimum quality metric $r^2$ of 0.3, which was sufficient for detecting low-frequency variants with a relatively large effect. Further studies that use newer data from the 1000 Genomes Project or large-scale meta-analyses across populations will be required to build on these results and to further elucidate the global genetic architecture of T2D.

**MATERIALS AND METHODS**

**Study design**

The present study was comprised of a three-stage analysis: a discovery stage (first stage), de novo genotyping and *in silico* follow-up (second stage), and de novo genotyping (third stage) analysis. The discovery stage was a GWAS that tested directly genotyped and imputed variants for association with T2D. We phased SNPs from the Illumina 610K SNP array and performed imputation using the 1000 Genomes East Asian reference panel. We selected 39 loci that had a top SNP with a $P < 1 \times 10^{-4}$ (adjusted for age, sex, BMI and the first four principal components from PCA), and filtered out 19 loci that were located in or near genes that were found to be associated with T2D in previous GWAS. The remaining 20 top SNPs were subsequently investigated for association with T2D by de novo genotyping analysis or *in silico* imputation analysis. We combined the results of the discovery and follow-up analyses to identify new T2D loci. An outline of the study design is shown in Supplementary Material, Figure S1.

**Subjects**

Subjects for the discovery analysis [BioBank Japan (BBJ) 1; 5976 cases with T2D and 20,829 controls] were recruited at several medical institutions in Japan (37,38). The follow-up analysis consisted of five studies, which were BBJ2, Shanghai Jiao Tong University (SJTU), Singapore Diabetes Cohort Study (SDCS)/Singapore Prospective Study Project (SP2), Singapore Chinese Eye Study (SCES) and Chinese University of Hong Kong (CUHK). We obtained a total of 2799 cases and 3793 controls *for in silico* follow-up analysis and a total of 10,319 cases and 6795 controls for the follow-up analysis by de novo genotyping. Samples in the validation analysis included an independent case-control sample (11,298 cases and 3,397 controls) from subjects enrolled in the BioBank Japan (BBJ3). T2D cases were from individuals registered as having T2D. Diabetes was diagnosed according to the WHO criteria. The exclusion criteria for cases were individuals positive for antibody to glutamic acid decarboxylase (GAD) or those with diabetes due to liver dysfunction, steroids and other drugs that might raise glucose levels, malignancy or monogenic disorder known to cause diabetes. Control subjects were healthy volunteers confirmed at annual health check-up or individuals not having T2D but with diseases other than T2D: bronchial asthma, myocardial infarction, breast cancer, Basedow’s disease, cerebral infarction, cerebral aneurism, osteoporosis, heart failure, unstable angina, poliomyelitis, arteriosclerosis obliterance, emphysema or atopic dermatitis. Note that BBJ1, BBJ2 and BBJ3 did not overlap for samples. Ethnicity was self-reported by the enrolled individuals. For each study, approval was obtained from the appropriate institutional review boards of the participating institutions, and a written informed consent was obtained from all participants.

Genotyping and imputation: in the discovery (BBJ1) and *in silico* follow-up analyses (SDCS/SP2, SCES, and CUHK), genotyping was done with genome-wide SNP arrays. In the de novo follow-up analysis, genotyping was carried out by using a multiplex polymerase chain reaction invader assay (BBJ2) and Mass ARRAY (SJTU). The typing platforms and quality control methods for each study are described in Supplementary Material, Table S1. We included SNPs from the SNP array for imputation and the association analysis with a call rate of ≥0.99 and a Hardy–Weinberg equilibrium (HWE) $P \geq 1 \times 10^{-6}$, selecting 6209 637 imputed SNPs with an MAF of $\geq 0.01$ and $r^2$ higher than a set of MAF specific thresholds as described in the previous study: MAF 0–0.1 = 0.75, MAF 0.1–0.2 = 0.70, MAF 0.2–0.3 = 0.66, MAF 0.3–0.4 = 0.60, MAF 0.4–0.5 = 0.55 for the discovery analysis. In the discovery and *in silico* analyses, SNP imputation was done using 572 East Asian haplotypes (194 CHB, 200 CHS and 178 JPT) from the 1000 Genomes Project data (June 2011 release).

**eQTL**

Potential candidates for association with T2D were pursued with eQTL studies in available datasets in lymphoblastoid cell lines (36) and a liver tissue gene expression database (39).

**Statistical analysis**

Associations between SNPs and T2D were tested by logistic regression analysis using an additive model with or without adjustment for age, sex, BMI and the first four principal components from PCA. A quantile–quantile plot was constructed by plotting the distribution of the observed $P$-values for the SNPs against the theoretical distribution of the expected $P$-values for T2D. In the discovery analysis, the genomic control inflation factor ($\lambda$) was calculated as the median $\chi^2$ statistic divided by 0.456. Meta-analysis was performed by an inverse variance method assuming fixed effects using R software. Quantitative trait analyses were done for BMI, FPG, HbA1c, log-transformed HOMA-β, log-transformed HOMA-IR, total cholesterol, HDL cholesterol and triglycerides by multiple linear regression analysis, employing an additive association model with or without adjustment for the relevant covariates. The power of detecting previously reported loci in the present study was estimated by using QUANTO, employing the RAF and the sample sizes in the discovery stage, the reported ORs, an assumed T2D prevalence of 10%, and $\alpha = 0.05$. For the analysis of lean versus obese subjects, samples were dichotomized into lean (BMI $< 25$) and obese (BMI $\geq 25$) groups and SNPs analyzed for each group using a logistic regression base model adjusted for gender, age and four principal components. Heterogeneity between lean and obese samples was calculated as a z-score using the beta and standard error (s.e.) from the logistic regression estimates as: $z = (\beta_{\text{lean}} - \beta_{\text{obese}})/\text{s.e.(\beta_{\text{lean}})} = (\beta_{\text{lean}} - \beta_{\text{obese}})/\text{s.e.}/\text{s.e.(\beta_{\text{obese}})}$. To test whether the reduced sample size of the obese sample group could affect its observed decrease in significance, we performed a re-sampling-based analysis, whereby samples with
sizes matching that of the obese case and control sizes were drawn without replacement 10 000 times from the complete set of samples. We estimated the probability of observing the obese group’s P-value due to random sample size differences using R’s empirical distribution function.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS
B.B.J.: We thank all the participants and the staff of the BioBank Japan project. We thank all participating doctors and staff from collaborating institutes for providing DNA samples. We also thank the technical staff of the Laboratory for Endocrinology and Metabolism at the RIKEN Center for Genomic Medicine for providing the technical assistance. Likewise, we thank the technical staff of the Laboratory for Genotyping Development at RIKEN Center for Genomic Medicine for performing SNP genotyping.

Conflict of Interest statement. None declared.

FUNDING
This work was supported by a grant from the Leading Project of Ministry of Education, Culture, Sports, Science and Technology, Japan. S.J.T.U.: We thank all medical staff of the Shanghai Clinical Center for Diabetes. This work was supported by grants from the National 973 Program (2011CB504001), 863 Program (2006AA02A409), National Science Foundation of China (30800617, 81170735), Excellent Young Medical Expert of Shanghai (XYQ2011041) and the major program of the Shanghai Municipality for Basic Research (08dj1400601), China. SP2: The Singapore BioBank and the Genome Institute of Singapore, Agency for Science, Technology, and Research provided services for tissue archival and genotyping, respectively. This work was supported by grants from the Biomedical Research Council of Singapore (BMRC 05/1/36/19/413 and 03/1/27/18/1102) and the National Medical Research Council of Singapore (NMRC/1174/2008). C.U.H.K.: We thank all medical and nursing staff of the Prince of Wales Hospital Diabetes Mellitus Education Centre for their commitment and professionalism. We would like to thank the Genome Institute at Quebec for help with replication genotyping, and the Chinese University of Hong Kong Information Technology Services Centre for their support with computing resources. This work was supported by the Hong Kong Foundation for Research and Development in Diabetes established under the auspices of the Chinese University of Hong Kong, the Hong Kong Government Research Grant Committee Central Allocation Scheme (CUHK 1/04C), Research Grants Council Earmarked Research Grant (CUHK4727/0M), the Innovation and Technology Fund (ITS/088/08 and ITS/487/09FP), a Chinese University Direct Grant, the Research Fund of the Department of Medicine and Therapeutics and the Diabetes and Endocrine Research Fund of the Chinese University of Hong Kong. The summary results of DIAGRAMv3 study were downloaded via the website at http://diagram-consortium.org/index.html.

REFERENCES


