Multiple Nonglycemic Genomic Loci Are Newly Associated With Blood Level of Glycated Hemoglobin in East Asians

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.2337/db13-1815</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:17820684">http://nrs.harvard.edu/urn-3:HUL.InstRepos:17820684</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Multiple Nonglycemic Genomic Loci Are Newly Associated With Blood Level of Glycated Hemoglobin in East Asians

Glycated hemoglobin A1c (HbA1c) is used as a measure of glycemic control and also as a diagnostic criterion for diabetes. To discover novel loci harboring common variants associated with HbA1c in East Asians, we conducted a meta-analysis of 13 genome-wide association studies (GWAS; N = 21,026). We replicated our findings in three additional studies comprising 11,576 individuals of East Asian ancestry. Ten variants showed associations that reached genome-wide significance in the discovery data set, of which nine (four novel variants at TMEM79 [P value = 1.3 × 10−20], HBS1L/MYB [8.5 × 10−15], MYO9B [9.0 × 10−15], and CYBA [1.1 × 10−8] as well as five variants at loci that had been previously identified [CDKAL1, G6PC2/ABC611, GCK, ANK1, and FN3KI]) showed consistent evidence of association in replication data sets. These variants explained 1.76% of the variance in HbA1c. Several of these variants (TMEM79, HBS1L/MYB, CYBA, MYO9B, ANK1, and FN3KI) showed no association with either blood glucose or type 2 diabetes. Among individuals with nondiabetic levels of fasting glucose (<7.0 mmol/L) but elevated HbA1c (≥6.5%), 36.1% had HbA1c <6.5% after adjustment for these six variants. Our East Asian GWAS meta-analysis has identified novel variants associated with HbA1c as well as demonstrated that the effects of known variants are largely transferable across ethnic groups. Variants affecting erythrocyte parameters rather than glucose metabolism may be relevant to the use of HbA1c for diagnosing diabetes in these populations.
raise the possibility that the impact of these variants on 
HbA1c relates to their effects on erythrocyte half-life. Al-
ternatively, variants near fructosamine 3 kinase (FN3K) 
may act through their effects on protein deglycation (14).
These effects become particularly relevant now in that 
HbA1c has been adopted as a diagnosis criterion of dia-
betes, because it exhibits less intradividual variability than 
either fasting glucose or 2-h post-challenge glucose after an 
oral glucose tolerance test and does not require fasting 
(15,16). It is recognized that the glucose and HbA1c criteria 
are not completely concordant and that genetic variants 
result in a significant reclassification of individuals based 
on HbA1c criteria when compared with fasting glucose 
criteria (17).
To date, with one exception (4), all GWAS for HbA1c 
have been conducted in populations of European ancestry. We have already demonstrated that genetic association 
studies in different ethnic groups offer opportunities
to identify novel loci that harbor variants encoding susceptibility to T2D (18,19). Furthermore, some hereditary anemias are more common in Asians, which may also affect HbA1c (20). In this study, we sought to find common genetic variants at novel loci that are associated with blood HbA1c level in the consortium of the Asian Genetic Epidemiology Network (AGEN; http://www.agenconsortium.org/), which consists of East Asian cohorts.

RESEARCH DESIGN AND METHODS

We conducted a meta-analysis of data from 16 cohorts comprising 32,602 individuals of East Asian ancestry (Table 1). The study was carried out in two stages. Stage 1 involved the meta-analysis of GWAS of 19,017 Chinese, Korean, Japanese, and Malay individuals, as well as an in silico lookup of all index single nucleotide polymorphisms (SNPs) with \( P \) value \(<10^{-4}\) and their proxies (in total 198 SNPs) in 2,009 subjects of Chinese ancestry of the Singapore Chinese Health Study (SCHS) of diabetes. This gave a sample size of 21,026 for the discovery stage. In stage 2, we selected five genome-wide significant (\( P \) value \(<5 \times 10^{-8}\) ) SNPs at potential novel loci from stage 1 and two SNPs that reached suggestive levels of significance at the fatty acid desaturase 2 (FADS2) and proteasome (prosome, macropain) 26S subunit, non-ATPase, 13 (PSMD13) genes in view of their known impact on blood lipid levels (21) and hematologic traits (22). We then carried out de novo genotyping of these SNPs in 9,592 Japanese individuals. In addition, 48 out of these 198 index SNPs were also genotyped in a study of 1,984 Chinese individuals of the TaiChi study. The findings from all these studies were then meta-analyzed.

Most of the studies were population-based cross-sectional studies in adults. For the studies that used a case-control design to study genetic association with diabetes, only nondiabetic control individuals were included in this study. For other case-control studies, the association testing was done in cases and controls separately. The average HbA1c across the participating studies varied between 4.8 and 5.8%, while the SDs were less than 0.4%. All subjects provided written informed consent. Detailed information of each study can be found in Table 1 and the Supplementary Data.

SNP Genotyping, Quality Control, and Imputation

Our stage 1 cohorts were genotyped on Affymetrix (Santa Clara, CA) and Illumina SNP arrays. We applied stringent sample and SNP quality control (QC) within each study separately. As a result, the QC criteria varied slightly across these studies, but the following steps have been largely applied (Supplementary Table 1). Firstly, we excluded samples that showed cryptic relatedness, were population outliers, or showed inconsistency between clinical and genetic genders. Secondly, we removed SNPs with minor allele frequency (MAF) <1%, genotype call rate <0.95, or Hardy-Weinberg equilibrium \( P \) value \(<1 \times 10^{-6}\).

We imputed our post-QC array genotypes up to the HapMap phase 2 haplotypes (23) using several widely used

---

**Table 1—Demographics of the participant cohorts**

<table>
<thead>
<tr>
<th>Study name</th>
<th>n</th>
<th>Design</th>
<th>Ethnicity</th>
<th>Age, mean (SD)</th>
<th>Male, %</th>
<th>BMI, mean (SD)</th>
<th>HbA1c, %, mean (SD)</th>
<th>IFCC,1 mean (SD)</th>
<th>( \lambda_{GC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRC</td>
<td>640</td>
<td>Population</td>
<td>Chinese</td>
<td>43.9 (7.7)</td>
<td>24.7</td>
<td>22.4 (2.2)</td>
<td>5.7 (0.6)</td>
<td>39 (6.6)</td>
<td>1.020</td>
</tr>
<tr>
<td>KARE</td>
<td>7,696</td>
<td>Population</td>
<td>Korean</td>
<td>51.6 (8.8)</td>
<td>46.6</td>
<td>24.5 (3.8)</td>
<td>5.6 (0.4)</td>
<td>38 (4.4)</td>
<td>1.052</td>
</tr>
<tr>
<td>CAGE-NCGM</td>
<td>323</td>
<td>Population</td>
<td>Japanese</td>
<td>63.0 (6.1)</td>
<td>53.6</td>
<td>22.9 (2.9)</td>
<td>5.2 (0.3)</td>
<td>33 (3.3)</td>
<td>1.001</td>
</tr>
<tr>
<td>NHAPC</td>
<td>2,507</td>
<td>Population</td>
<td>Chinese</td>
<td>58.5 (6.0)</td>
<td>42.4</td>
<td>24.3 (3.6)</td>
<td>5.7 (0.4)</td>
<td>39 (4.4)</td>
<td>1.005</td>
</tr>
<tr>
<td>SCES</td>
<td>1,580</td>
<td>Population</td>
<td>Chinese</td>
<td>57.7 (9.4)</td>
<td>42.2</td>
<td>23.5 (3.5)</td>
<td>5.8 (0.3)</td>
<td>40 (3.3)</td>
<td>1.002</td>
</tr>
<tr>
<td>SIMES</td>
<td>1,727</td>
<td>Population</td>
<td>Malay</td>
<td>57.6 (11.2)</td>
<td>33.9</td>
<td>25.8 (5.1)</td>
<td>5.7 (0.4)</td>
<td>39 (4.4)</td>
<td>1.007</td>
</tr>
<tr>
<td>SP2–610</td>
<td>797</td>
<td>Population</td>
<td>Chinese</td>
<td>47.4 (10.6)</td>
<td>17.0</td>
<td>22.2 (3.7)</td>
<td>5.6 (0.4)</td>
<td>38 (4.4)</td>
<td>1.027</td>
</tr>
<tr>
<td>SP2–1M</td>
<td>777</td>
<td>Population</td>
<td>Chinese</td>
<td>46.7 (10.1)</td>
<td>61.5</td>
<td>22.8 (3.4)</td>
<td>5.6 (0.4)</td>
<td>38 (4.4)</td>
<td>1.006</td>
</tr>
<tr>
<td>SP2–550</td>
<td>266</td>
<td>Population</td>
<td>Chinese</td>
<td>48.3 (12.2)</td>
<td>66.0</td>
<td>23.3 (3.5)</td>
<td>5.6 (0.4)</td>
<td>38 (4.4)</td>
<td>0.992</td>
</tr>
<tr>
<td>TWSC</td>
<td>920</td>
<td>Population</td>
<td>Chinese</td>
<td>50.0 (17.8)</td>
<td>50.8</td>
<td>23.6 (3.5)</td>
<td>5.2 (0.4)</td>
<td>33 (4.4)</td>
<td>1.010</td>
</tr>
<tr>
<td>SCHS-CHD</td>
<td>1,024</td>
<td>MI controls</td>
<td>Chinese</td>
<td>60.1 (8.0)</td>
<td>67.6</td>
<td>22.9 (3.1)</td>
<td>5.7 (0.4)</td>
<td>39 (4.4)</td>
<td>0.998</td>
</tr>
<tr>
<td>CAGE-Fukuoka</td>
<td>457</td>
<td>MI cases</td>
<td>Chinese</td>
<td>59.8 (7.8)</td>
<td>63.0</td>
<td>22.6 (2.9)</td>
<td>5.7 (0.4)</td>
<td>39 (4.4)</td>
<td>1.008</td>
</tr>
<tr>
<td>SBCS</td>
<td>303</td>
<td>BC controls</td>
<td>Chinese</td>
<td>53.0 (8.4)</td>
<td>0.0</td>
<td>24.7 (5.1)</td>
<td>5.8 (0.4)</td>
<td>40 (4.4)</td>
<td>1.012</td>
</tr>
<tr>
<td>SCHS-DB</td>
<td>2,009</td>
<td>T2D controls</td>
<td>Chinese</td>
<td>55.2 (7.1)</td>
<td>46.7</td>
<td>22.7 (3.1)</td>
<td>5.5 (0.3)</td>
<td>37 (4.4)</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Stage 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TaiChi</td>
<td>1,984</td>
<td>Population</td>
<td>Chinese</td>
<td>68.6 (9.0)</td>
<td>49.7</td>
<td>24.3 (3.4)</td>
<td>5.8 (0.3)</td>
<td>40 (3.3)</td>
<td>1.09</td>
</tr>
<tr>
<td>CAGE-Fukuoka</td>
<td>4,880</td>
<td>Population</td>
<td>Japanese</td>
<td>63.8 (5.8)</td>
<td>46.2</td>
<td>22.7 (2.7)</td>
<td>4.8 (0.2)</td>
<td>29 (2.2)</td>
<td>NA</td>
</tr>
<tr>
<td>JMGP</td>
<td>4,712</td>
<td>Population</td>
<td>Japanese</td>
<td>59.5 (14.3)</td>
<td>35.6</td>
<td>22.8 (3.1)</td>
<td>5.5 (0.4)</td>
<td>37 (4.4)</td>
<td>NA</td>
</tr>
</tbody>
</table>

The participant cohorts are listed in stage 1 (genome-wide association test) and stage 2 (de novo genotyping), respectively. Age, BMI, and HbA1c, are given as the mean value and SE in each cohort. \( \lambda_{GC} \) is the inflation factor calculated as per genomic control. CAGE, Cardiovascular Genomic Epidemiology; CRC, Cardiometabolic Risk in Chinese; JMGP, Japanese Millennium Genome Project; KARE, Korea Association Resource; NCGM, National Center for Global Health and Medicine; NHAPC, Nutrition and Health of Aging Population in China; SBCS, Shanghai Breast Cancer Study; SCES, Singapore Chinese Eye Study; SCHS-CHD, Singapore Chinese Health Study of Coronary Heart Disease; SBS-DB, Singapore Chinese Health Study of Diabetes Mellitus; SIMES, Singapore Malay Eye Study; SP2, Singapore Progressive Study Program; TWSC, Taiwan Super Control Study. 1The International Federation of Clinical Chemistry unit for HbA1c is mmol/mol.
software packages. For the cohorts of Chinese, Korean, and Japanese ancestry, HapMap haplotypes of the Japanese in Tokyo, Japan (JPT) and Chinese in Beijing, China (CHB) were used as the reference panel in the imputation. For Malays, the combined African Yoruba in Ibadan, Nigeria (YRI); Utah residents with ancestry from Northern and Western Europe (CEU); and JPT+CHB panels were used. The combined HapMap panel resulted in about 1.9 million SNPs in the imputed genotypes, while there were 2.4 million in the imputation data when the JPT+CHB panel was used. Imputed SNPs of MAF <1%, Hardy-Weinberg equilibrium $P$ value $\leq 1 \times 10^{-6}$, or poor imputation quality (IMPUTE info $< 0.5$, BEAGLE allelic $R^2 < 0.5$, or MACH $R_{sq} < 0.3$) were removed (24–26).

In stage 2, the follow-up SNPs were genotyped on three platforms. The two Japanese cohorts were genotyped using the TaqMan system (Life Technologies Corporation, Carlsbad, CA). The TaiChi cohort was genotyped using the Illumina CardiO-MetaboChip (San Diego, CA). The same SNP QC criteria as in stage 1 were applied to the two Japanese studies, while similar sample QC and SNP QC as in stage 1 were applied to TaiChi study (Supplementary Table 1).

**Phenotype Definition**

The traditional HbA$_1c$ unit (percentage of HbA$_1c$ in total hemoglobin) was used in the analysis. Subjects with diabetes were excluded. Diabetes was diagnosed if the subject gave a history of physician-diagnosed diabetes, was taking medication for diabetes, or had fasting glucose $\geq 7$ mmol/L or HbA$_1c$ $\geq 6.5\%$ (48 mmol/mol) in those studies where fasting glucose was not available. The HbA$_1c$ measurements were fitted into a linear model that adjusted for the sex and BMI of the individuals. The residuals were then normalized to have a mean of 0 and an SD of 1 using inverse-normal transformation.

**Association With HbA$_1c$ and Meta-analysis**

We tested the association of the SNPs with the normalized HbA$_1c$ residuals using linear regression assuming additive effects of the dosage of the effect allele. The linear regression was evaluated by several widely used software packages (Supplementary Table 1). Study-specific covariates, such as principal components and sample recruitment sites, were also considered as required (Supplementary Data).

The results from separate studies were combined using the fixed-effect scheme weighted by the inverse of the SE as implemented in METAL (27). Genomic control was applied within METAL.

**RESULTS**

In total, we identified nine loci harboring variants associated with HbA$_1c$ levels in East Asian populations. Among them, four loci were associated with HbA$_1c$ for the first time. The fixed-effect meta-analysis showed no significant evidence of heterogeneity for most of the index SNPs in the cohorts we have recruited. Although the heterogeneity test for variants near the G6PC2/ATP-binding cassette, subfamily B (MDR/TAP), member 11 (ABCB11) locus reached a borderline level of statistical significance, the extent of heterogeneity was moderate (Supplementary Table 2).

**Known Associated Loci Replicated in East Asian Populations**

In stage 1, the maximal inflation factor was 1.052 in the Korea Association Resource (KARE) study, while those of the other studies were all $\sim 1$, indicating that population stratification was unlikely to confound our findings (Table 1). Supplementary Fig. 1 shows the quantile–quantile plots with and without the SNPs known to be associated with HbA$_1c$. This displayed a marked deviation from the null hypothesis of no association that persisted after all known SNPs associated with HbA$_1c$ were removed. In this stage, we had $\sim81\%$ power to discover genomic variants that explain 0.2% phenotype variance at a significance level of $P = 5 \times 10^{-8}$.

The associations between index SNPs at novel and known loci are presented in Table 2. Ten loci showed associations with HbA$_1c$ with $P$ values that met the criteria for genome-wide significance ($P$ value $\leq 5 \times 10^{-8}$) (Fig. 1). Of these, five were at loci that were known to harbor variants associated with HbA$_1c$ levels. The index SNPs at these known loci were rs7772603 in the CDKAL1 gene region, rs3755157 at the G6PC2/ABCB11 locus, rs1799884 near the GCK gene, rs4737009 in the ANK1 gene region, and rs1046875 near the FN3K gene. We investigated the linkage disequilibrium (LD) between our top SNPs, and the reported index SNPs at these loci identified in populations of European ancestry (Supplementary Table 3). At the GCK and FN3K loci, our index SNPs were exactly the same as or were in perfect LD with the index SNPs reported in Europeans. In fact, the SNP identified in populations of European ancestry at both loci (rs1046896 near the FN3K locus and rs730497 near the GCK locus) (3) (3) also showed genome-wide significant associations in our meta-analysis ($P$ value $= 3.4 \times 10^{-13}$ and $1.3 \times 10^{-18}$ after genomic control). For ANK1, we replicated the primary signal rs4737009 reported in populations of European ancestry ($P$ value $= 1.3 \times 10^{-15}$) (3). This same study identified a second variant (rs6474359) in the ANK1 region, which was not in LD with the primary signal at this locus. This second SNP showed only a minor association with HbA$_1c$ in our study ($P$ value $= 9.3 \times 10^{-3}$), with an opposite direction of effect compared with Europeans.

The index SNP identified in European populations at G6PC2/ABCB11 (rs552976) did not show any association with HbA$_1c$ in our meta-analysis ($P$ value $= 0.54$). However, we did identify an association with rs3755157 near this locus that reached genome-wide significance. Although there was no evidence of LD between this and rs552976 in either HapMap panels of European (CEU) or Asian (JPT and CHB) ancestry (Supplementary Table 3), rs3755157 did show a genome-wide significant association with fasting glucose in populations of European
We conducted association tests with and without conditioning on the European SNP rs552976 in Singapore cohorts, including 5,147 Chinese and Malays. 

P values for rs3755157 were $1.5 \times 10^{-24}$ and $1.4 \times 10^{-24}$ with and without conditioning on rs552976, respectively. In addition, the index SNP at the CDKAL1 locus (rs7772603) was different from rs7747752 identified in the Korean cohort that formed part of this analysis. However, these two SNPs showed moderate LD in the JPT+CHB panel, and rs7747752 showed an association with HbA1c that was in the same direction and similar magnitude as that reported previously in the Korean population (4), with a suggestive degree of statistical significance in our study ($P = 1.1 \times 10^{-25}$). This suggests that associations with HbA1c for these two variants may represent the same association signal.

We further compared the direction of the effect for known HbA1c-associated variants between European and East Asian populations (Supplementary Table 6). Of the 17 index SNPs associated with HbA1c in previous publications, rs1800562 (HFE) is monomorphic in East Asians, while rs16926246 (HK1) was excluded because of the poor imputation quality. Most of the remaining 15 index SNPs showed consistent effects between Europeans and East Asians (Fig. 2). The only exception was rs6474359 at the ANK1 locus, for which each copy of the T allele was associated with higher HbA1c level by 0.06% ($P = 1.2 \times 10^{-28}$) but lower HbA1c level by 0.08 SD of the normalized HbA1c, which was equivalent to 0.05% HbA1c ($P = 9.3 \times 10^{-3}$), assuming a normal distribution of HbA1c in our study.

**Novel Associated Loci Revealed in East Asian Populations**

The other five genome-wide significant hits in stage 1 were all novel associations (Table 2). These five variants and two additional SNPs (rs540078 at PSMD13 and rs174570 at FADS2) were examined in stage 2. Of these, data for the top SNP at cytochrome b-245, a polypeptide (CYBA; rs9933309) was available only in the TaiChi study, which consists of 1,984 Chinese. Four of these seven SNPs were
replicated successfully in our stage 2 cohorts with consistent effect directions and showed stronger associations after combining the results of stage 2 and stage 1 with $P$ values less than $5 \times 10^{-8}$. They were variants close to the transmembrane protein 79 (TMEM79), Hsp70 subfamily B suppressor 1-like protein/v-myb avian myeloblastosis viral oncogene homolog (HBS1L/MYB), CYBA, and myosin IXB (MYO9B) loci (Table 2). The remaining SNPs,

Figure 1—Manhattan plot of genome-wide meta-analysis of stage 1 cohorts. The $-\log_{10}$ of the association $P$ values (y-axis) are plotted against the genomic coordinates (x-axis). The horizontal line in the plot indicates the genome-wide significance ($5 \times 10^{-8}$). The most relevant gene of each signal was labeled on the top of it, with the novel loci presented in brown and known loci in blue.

Figure 2—The bivariate plot of the effect directions in our meta-analysis and in previous reports. For the index SNPs in the previous GWAS, we plotted the reported effect (x-axis) versus the AGEN effect (y-axis). Whenever available, we used the effects reported by Soranzo et al. (3). The solid dots represent the SNPs that were significant ($P$ value $\leq 0.05$) in our study, while the hollow red dots represent the insignificant ones. The CDKAL1 top SNP reported in Koreans was specially marked by a diamond. EA, East Asian.
rs1467311 at 9q31.2 (stage 1+2 meta-analysis $P$ value = $1.0 \times 10^{-6}$), rs174570 in $FADS2$ ($P$ value = $2.0 \times 10^{-7}$), and rs540078 in $PSMD13$ ($P$ value = $3.2 \times 10^{-5}$) did not show genome-wide significant association with HbA1c in stage 1 and stage 2 meta-analysis. The regional association signals at the novel loci and the known loci were presented in Fig. 3 and Supplementary Fig. 2, respectively.

For the novel loci identified in East Asians, the MAF of the index SNPs in HapMap JPT+CHB panel were comparable to those in CEU panel (Supplementary Table 7). In most cases, the direction of effect was the same in AGEN as it was in the study of European ancestry, with the exception of rs6684514 at $TMEM79$. However, this SNP showed only a nominal degree of statistical significance in Europeans ($P$ value = $4.0 \times 10^{-7}$).

**Association With Fasting Glucose and Erythrocyte Traits**

We also looked up the associations between the variants that reached genome-wide significance with glucose-related traits.
traits in Europeans (28–30) and Asians (19,31) (Supplementary Tables 4 and 5). SNPs near \( GCK \) and \( G6PC2 \) were associated with blood glucose in both Europeans and East Asians. SNPs at the \( CDKL1 \) locus were associated with T2D in East Asians. Among the known loci, \( FN3K \) and \( ANK1 \) were not associated with glucose level and T2D in either East Asians or Europeans. For the novel SNPs associated with \( HbA1c \) that we identified in this study, none showed a statistically significant association with glucose levels except rs9399137 near the \( HBS1L/MYB \) loci, which showed an association with fasting glucose with borderline statistical significance that did not survive Bonferroni correction. We next examined the associations between these variants and red-cell-associated parameters in Europeans (6,8) and East Asians (32) (Supplementary Tables 8 and 9). Variants close to \( TMEM79, HBS1/MYB, CYBA, \) and \( ANK1 \) showed statistically significant associations with red-cell-associated parameters in at least one population.

**Phenotype Variance Explained by the Associated Loci**

We estimated the phenotype variance explained (PVE) by the genome-wide significant index SNPs in our meta-analysis. The PVE in each cohort was calculated by fitting the raw \( HbA1c \) in a linear regression model on the allele dosages of the index SNPs under investigation. The adjusted \( r^2 \)-squared from the linear regression model was used as the estimation of PVE. The estimates were obtained for the novel, known loci and all loci separately. The estimates from separate studies were combined using a sample-size-weighted scheme.

In Singapore Chinese Eye Study (SCES), Singapore Malay Eye Study (SiMES), and the three Singapore Progressive Study Program (SP2) cohorts, the known loci and the novel loci explained 1.01% and 0.75% of the total \( HbA1c \) variance, respectively, while all the index SNPs as a whole explained 1.76% \( HbA1c \) variance.

**Reclassification of Diabetes Diagnosis Using \( HbA1c \)**

Finally, we calculated the proportion of the samples that were reclassified by adjusting for the allele dosage of index SNPs at the six loci that did not show association with glucose or T2D in populations of either European ancestry (28–30) or Asian ancestry (19,31). These included \( TMEM79, HBS1L/MYB, CYBA, MYO9B, ANK1, \) and \( FN3K \).

This reclassification analysis was done in 15,150 individuals with fasting glucose measurements available, which came from SP2, KARE, Nutrition and Health of Aging Population in China (NHAPC), National Center for Global Health and Medicine (NCGM), Cardiometabolic Risk in Chinese (CRC), and TaiChi. Individuals with known diabetes (defined by having diabetic history or using diabetic medication) were removed. In the remaining subjects, undiagnosed diabetes was defined as those having fasting glucose \( \geq 7 \) mmol/L, whereas the nondiabetic individuals had fasting glucose \( < 7 \) mmol/L. We adjusted the raw \( HbA1c \) levels using a linear regression, including the allele dosages of the six SNPs as covariates. We then classified individuals into those with and without diabetes based on \( HbA1c \geq 6.5\% \) (15). We compared the concordance between these three methods for diagnosing diabetes (fasting glucose \( \geq 7 \) mmol/L, whereas the nondiabetic individuals had fasting glucose \( < 7 \) mmol/L). We adjusted the raw \( HbA1c \) levels using a linear regression, including the allele dosages of the six SNPs as covariates. We then classified individuals into those with and without diabetes based on \( HbA1c \geq 6.5\% \) (15).

The reclassification analysis was done in cohorts with fasting glucose measurement. Individuals with known diabetic history or diabetic medication were removed. In the remaining subjects, undiagnosed diabetes was defined as those having fasting glucose \( \geq 7 \) mmol/L, whereas the nondiabetic individuals had fasting glucose \( < 7 \) mmol/L. We adjusted the raw \( HbA1c \) levels using a linear regression, including the allele dosages of the six SNPs as covariates. We then classified individuals into those with and without diabetes based on \( HbA1c \geq 6.5\% \) (15).

We compared the concordance between these three methods for diagnosing diabetes (fasting glucose \( \geq 7 \) mmol/L, whereas the nondiabetic individuals had fasting glucose \( < 7 \) mmol/L). We adjusted the raw \( HbA1c \) levels using a linear regression, including the allele dosages of the six SNPs as covariates. We then classified individuals into those with and without diabetes based on \( HbA1c \geq 6.5\% \) (15).

One hundred eighty-one subjects had undiagnosed diabetes based on fasting glucose, of which 123 had \( HbA1c \geq 6.5\% \), of whom three had adjusted \( HbA1c < 6.5\% \). Fifty-eight individuals had \( HbA1c < 6.5\% \), of which one had adjusted \( HbA1c \geq 6.5\% \).

In contrast, 14,96 individuals were nondiabetic, with fasting glucose \( < 7 \) mmol/L. Of these, 476 were classified as having diabetes based on \( HbA1c \geq 6.5\% \). However, 172 of the 476 individuals had adjusted \( HbA1c < 6.5\% ; 14,493 \) had \( HbA1c < 6.5\% \). Among these, 40 had adjusted \( HbA1c \geq 6.5\% \). As such, some nonconcordance was observed when diabetes was diagnosed based on fasting plasma glucose \( \geq 7 \) mmol/L and \( HbA1c \geq 6.5\% \). Among those with fasting plasma glucose \( \geq 7 \) mmol/L, very little of the nonconcordance was explained by these six variants. However, among those with fasting plasma glucose \( < 7 \) mmol/L, 36.13% of those with elevated \( HbA1c \) had \( HbA1c < 6.5\% \) after these six variants were taken into consideration.

In populations of European ancestry, the equivalent proportion was 15.18% (3).

**DISCUSSION**

In this East Asian GWAS meta-analysis, we showed that most of the variants identified in populations of European ancestry had similar effects in East Asians. Variants close

<table>
<thead>
<tr>
<th>Table 3—Individual reclassification using raw ( HbA1c ) or adjusted ( HbA1c )</th>
<th>Adjusted ( HbA1c )</th>
<th>Subtotal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting plasma glucose ( \geq 7 )</td>
<td>( HbA1c &lt; 6.5 )</td>
<td>57</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( HbA1c \geq 6.5 )</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>Fasting plasma glucose ( &lt; 7 )</td>
<td>( HbA1c &lt; 6.5 )</td>
<td>13,813</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>( HbA1c \geq 6.5 )</td>
<td>172</td>
<td>304</td>
</tr>
</tbody>
</table>
to the G6PC2/ABCA11, GCK, ANKI, and FN3K that were first identified in populations of European ancestry also showed associations that reached genome-wide significance in our study. However, at the G6PC2/ABCB11, our index SNP was not in LD with the index SNP (rs552976) identified in populations of European ancestry. The latter was not associated with HbA1c in our study (P value = 0.54). Therefore, rs3755157 could represent an additional signal at this locus, particularly given that we did not observe obvious attenuation of the associations in the analysis conditioning on rs552976. Another SNP at this locus, particularly given that we did not observe obvious attenuation of the associations in the analysis conditioning on rs552976. Another SNP at this locus (rs560887) was associated with fasting glucose in Europeans (28) but not with HbA1c in our study (P value = 0.10). The majority of the other known SNPs showed similar direction of effect and effect size in our population as in populations of European Ancestry. The exception was a variant at the ANKI locus with showed an association in the opposite direction to that observed in populations of European ancestry. We are not able to explain this finding at this time. However, we would point out that the association in our study, although in a different direction, was far from reaching genome-wide significance. Larger sample sizes will be required to formally test for heterogeneity of effect of this variant in populations of European and East Asian ancestry.

We identified variants at or close to four novel loci associated with HbA1c in this meta-analysis. TMEM79 is a transmembrane protein that is highly expressed in liver, erythrocytes, and adipose tissue (The European Bioinformatics Institute, Expression Atlas database, http://www.ebi.ac.uk/gxa). The index SNP rs6684514 was also associated with mean corpuscular hemoglobin concentration in a GWAS conducted in Japanese (32). Although this suggests that this SNP may exert its effects in HbA1c through its effects on erythrocyte biology, we cannot exclude the possibility that this variant could alter glucose regulation. TMEM79 is downregulated by a high-fat diet in adipose tissue in mice (33). TMEM79 is also differentially methylated in adipose tissue in twins discordant for T2D. However, this association did not survive multiple testing (34). The index SNP at the TMEM79 locus rs6684514 was a common missense variant (Val147Met). However, this was predicted as benign by sorting intolerant from tolerant (SIFT) (35) and polymorphism phenotyping (PolyPhen) (36). We also looked for other functional variants in the nearby genes. Three missense SNPs were found to be in high LD (r² = 0.97; MAF = 0.22) with rs6684514. They were rs10908495 and rs10908496 at the CIrfr85 locus and rs2230194 at the CCT3 locus. However, their P values were all less than 3.7 × 10⁻⁹ in stage 1. Haplotype analysis showed that there were only two common haplotypes (frequency >0.01) in Chinese cohorts of SCES and SP2. They were A-C-A-A and G-T-G-G for rs6684514-rs10908495-rs10908496-rs2230194, while G-T-G-G carried all the alleles associated with elevated HbA1c. Hence, we are unable to discriminate the genuine causal variant among these four missense variants. rs2230194 did not exist in the combined reference panel of HapMap CEU, JPT+CHB, and YRI and hence was absent in Malay cohort of SIMES.

The index SNP rs9399137 is located in the intergenic region between HBS1L and MYB genes and resides in a LD block (HMIP 2), which contains SNPs associated with various hematol-o-gy traits (5–8,22,37,38), including HbA2 (39) and HbF (40). Another SNP in the same LD block was associated with mean corpuscular hemoglobin, erythrocyte count, and other hematol-o-gy traits significantly in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (6). The index SNPs reported by CHARGE were in moderate/high LD with rs9399137 (r² ranges from 0.45 to 1.00 in HapMap release 22 CEU panel). This observation was also supported by recent GWAS in African Americans (9,41). A study has shown a distal regulatory element may exist here to control the expression of MYB gene (42), which plays an important role in erythrocyte formation and differentiation.

MYO9B codes for a single-head myosin isoform that was found to be expressed primarily in peripheral blood white cells (43). The product of the gene is important for actin remodeling of epithelial enterocytes, hence it has been associated with several inflammatory bowel diseases, such as Crohn disease and celiac disease, in which the impairment of the enterocyte function explained part of the disease syndrome (44). MYO9B function could have an impact on both glycemic and nonglycemic determinants of HbA1c. Celiac disease can result in anemia (45) due to various nutrient deficiencies. This can alter red cell turnover and thus HbA1c level. In addition, celiac disease is also associated with lower BMI and a lower prevalence of T2D (46). Actually, the T allele of our index SNP rs11667918 was associated with lower BMI in another meta-analysis conducted by AGEn (P value = 0.03) (47), as well as lower HbA1c, level in our study. Furthermore, intestinal permeability may have a role in the pathogenesis of type 1 diabetes (T1D) (48). MYO9B’s expression is altered by a hydrolyzed casein diet, which is also able to prevent diabetes in the DP-BB mouse, a mouse model of T1D (49). Variants at this locus showed an association with T1D in a Spanish population (50). However, this finding was not replicated in Dutch and British populations (51). Another SNP rs2279008 at MYO9B was associated with human height (52), which was weakly associated with T2D (53), but this SNP is not in LD with our index SNP rs11667918 in both CEU and JPT+CHB panels.

Although, the association of the index SNP of CYBA rs9933309 in our replication cohort was not statistically significant, this SNP was only genotyped in a small number of individuals as part of the replication cohorts due to limitations of resources. However, the association after meta-analysis did reach GWAS significance, and the same SNP showed an association with HbA1c in the Meta-Analyses of Glucose and Insulin-Related Traits Consortium
Genetic Associations With HbA1c in East Asians

fasting plasma glucose.

Approximately 3.3% of the populations studied showed no evidence of association with glucose traits or this variant is very likely to represent a true positive (MAGIC) meta-analysis (3), with an effect direction consistent with that found in our study and P value of $1.2 \times 10^{-4}$ (Supplementary Table 7). For this reason, we believe that this variant is very likely to represent a true positive finding in our study. CYBA encodes p22phox, a subunit of NADPH oxidase. NADPH oxidase produces superoxide after catalyzing the reaction from NADPH to NADPH$^\ast$. In phagocytes, these superoxides are released to kill bacteria and fungi. Since p22phox is the primary component of this microbicidal system, mutations in CYBA gene were reported to cause immunodeficiency diseases, such as the recessive chronic granulomatous disease (54). Of relevance to glucose metabolism, higher NADPH oxidase activity, together with higher levels of its subunits, including p22phox, was found to be induced by high environmental glucose in a T2D rat model (55). This increased oxidative stress has been observed in both animal (56) and human (57) $\beta$-cell models. In fact, the $\beta$-cell is particularly susceptible to the effects of oxidative stress because of the relatively lower expression level of superoxide dismutase, which is protective against oxidation damage, as compared with other tissues (58). It has been argued that high-glucose-induced superoxide was the major cause of $\beta$-cell dysfunction and death (59). Thus variants at the CYBA locus may result in oxidative stress in the $\beta$-cell, leading to glucose intolerance. However, this variant did not show any association with glycemic traits or T2D in any of the populations examined (Supplementary Table 3). Instead, it showed an association with mean cell hemoglobin concentration and mean corpuscular volume in Japanese population, as well as hemoglobin level in a population of European ancestry (Supplementary Table 8 and 9).

It is noteworthy that six of the nine variants that showed genome-wide significant associations with HbA1c showed no evidence of association with glucose traits or T2D. This included three out of four of the variants at novel loci identified in our study (TMEM79, HBS1L/MYB, and MYO9B). Several of these showed strong associations with red-cell-associated parameters (TMEM79, HBS1L/MYB, ANKI, and CYBA). Another (FNK3) is thought to influence HbA1c by impacting the deglycation of proteins (14) rather than through an impact on glucose metabolism.

Approximately 3.3% of the populations studied with fasting plasma glucose <7.0 mmol/L would have been diagnosed as having diabetes based on HbA1c $\geq$6.5%. Over one-third of this nonconcordance could be explained by these six variants. These findings may have implications for the diagnosis of diabetes using HbA1c in East Asians. It may be that the threshold for diagnosing diabetes based on HbA1c should be slightly higher in these populations. However, an evaluation of appropriate thresholds for the diagnosis of diabetes really requires a careful examination of the relationship between HbA1c and microvascular complications associated with diabetes in these populations. Most of the studies included in this analysis did not have data on microvascular complications, and this is outside the scope of this study. These variants may also contribute to the observation that in individuals with diabetes, Southeast and East Asians exhibit HbA1c that is $\sim$0.2–0.5% higher compared with Caucasians with similar mean blood glucose level (60).

In conclusion, common genetic variants associated with HbA1c levels in populations of European ancestry have similar effects on HbA1c levels in East Asians. We identified four novel associated genomic loci in our East Asian populations. Existing data point to an effect that is mediated by the effect of these variants on nonglycemic factors that affect the erythrocyte for at least three of these variants. However, we cannot conclusively exclude the possibility that they may have an impact on glucose regulation. These findings may have implication on the use of HbA1c to diagnose diabetes in East Asian populations.

Acknowledgments. The authors thank the subjects for their participation and collaborators for their efforts in this study. The complete acknowledgments and the list of collaborators can be found in the Supplementary Data.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. P.C., J.-Y.L., and M.O.G. analyzed and interpreted the data and drafted or revised the manuscript. F.T., H.L., Y.T., Y.J.K., M.J.G., C.-H.C., L.-C.C., Q.Q., M.D.G., T.M., and Y.-Y.T. analyzed and interpreted the data. J.-Y.-W., J.Lian., J.Lo., E.V., C.-C.K., J.Liu, X.Y., Y.W., L.L., T.L.Y., A.C.T., C.-Y.C., W.-H.P., M.I., K.O., Y.K., H.U., Y.O., M.K., A.T., T.T., F.J.A.v.R., S.K.G., T.L.A., Y.-T.C., T.A., and T.-Y.W. performed data acquisition. M.A.P., D.O.S., J.Liao, K.Y., J.C., C.H., and J.-Y.H. analyzed and interpreted the data and performed data acquisition. J.Le. drafted or revised the manuscript. R.M.v.D., Y.F., C.-K.H., K.Y., J.C., C.H., and J.-Y.H. analyzed and interpreted the data and performed data acquisition. J.Le. performed the conception and design of the work and data acquisition and drafted or revised the manuscript. W.-P.K., Q.C., Y.G., C.A.H., W.H.-H.S., J.I.R., Y.-D.I.C., J.Y., T.L.A., and L.Q. performed data acquisition and drafted or revised the manuscript. W.Z., X.-O.S., B.-J.K., and E.S.T. performed the conception and design of the work and data acquisition and drafted or revised the manuscript. R.T., T.H., and X.L. performed the conception and design of the work. J.H. analyzed and interpreted the data and performed data acquisition, and drafted or revised the manuscript. E.S.T. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References


22. Teslovich TM, Musunuru K, Smith AV, et al.; GIANT; MAGIC; ANTHROPOMETRIC TRAITS (GANT) Consortium; Asian Genetic Epidemiology Network—Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Diabetics Genetics Replication And Meta-analysis (DiAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981–990


29. Morris AP, Voight BF, Teslovich TM, et al.; Wellcome Trust Case Control Consortium; Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANTHROPOMETRIC TRAITS (GANT) Consortium; Asian Genetic Epidemiology Network—Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Diabetics Genetics Replication And Meta-analysis (DiAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981–990


53. Lawlor DA, Ebrahim S, Davey Smith G. The association between components of adult height and Type II diabetes and insulin resistance: British Women’s Heart and Health Study. Diabetologia 2002;45:1097–1106