Pancreatic beta-cell function and type 2 diabetes risk: quantify the causal effect using a Mendelian randomization approach based on meta-analyses

Yiqing Song1,2,†, Edwina Yeung3, Aiyi Liu4, Tyler J. VanderWeele5,6, Liwei Chen7, Chen Lu8, Chunling Liu9, Enrique F. Schisterman3, Yi Ning10 and Cuilin Zhang3,*†

1Institute of Vascular Medicine, Peking University Third Hospital, Beijing 100191, China, 2Division of Preventive Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA, 3Epidemiology Branch and 4Biostatistics & Bioinformatics Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, MD, USA, 5Department of Biostatistics and 6Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA, 7Program of Biostatistics, School of Public Health, Louisiana State University Health Science Center, New Orleans, LA, USA, 8Department of Biostatistics, Boston University, School of Public Health, Boston, MA, USA, 9Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China and 10Department of Epidemiology and Community Medicine, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

Received March 13, 2012; Revised August 5, 2012; Accepted August 10, 2012

The objective of the study is to quantify the causal effect of β-cell function on type 2 diabetes by minimizing residual confounding and reverse causation. We employed a Mendelian randomization (MR) approach using TCF7L2 variant rs7903146 as an instrument for lifelong levels of β-cell function. We first conducted two sets of meta-analyses to quantify the association of the TCF7L2 variant with the risk of type 2 diabetes among 55,436 cases and 106,020 controls from 66 studies by calculating pooled odds ratio (OR) and to quantify the associations with multiple direct or indirect measures of β-cell function among 35,052 non-diabetic individuals from 31 studies by calculating pooled mean difference. We further applied the method of MR to obtain the causal estimates for the effect of β-cell function on type 2 diabetes risk based on findings from the meta-analyses. The OR [95% confidence interval (CI)] was 0.87 (0.81–0.93) for each five unit increment in homeostasis model assessment of insulin secretion (HOMA-%B) (P = 3.0 × 10−5). In addition, for measures based on intravenous glucose tolerance test, ORs (95% CI) associated with type 2 diabetes risk were 0.24 (0.08–0.74) (P = 0.01) and 0.14 (0.04–0.48) (P = 0.002) for per 1 standard deviation increment in insulin sensitivity index and disposition index, respectively. Findings from the present study lend support to a causal role of pancreatic β-cell function itself in the etiology of type 2 diabetes.

INTRODUCTION

Type 2 diabetes is characterized by insulin resistance in peripheral tissues coupled with a progressive impairment of pancreatic β-cell function. The recent demonstration that most type 2 diabetes-associated variants impact β-cell function provided support of the theory that a background of β-cell insufficiency is more likely to predate and set the stage for type 2 diabetes (1–4). However, the magnitude of the potential causal effect of β-cell function itself on the development of type 2 diabetes remains uncertain (5,6). The majority of previous epidemiological studies on the role of β-cell function in the development of type 2 diabetes were conducted among adults whose β-cell function may have been affected by impaired glucose tolerance from pre-diabetes or undiagnosed diabetes. Therefore, reverse causation is a plausible explanation...

†The authors with equal contributions.

*To whom correspondence should be addressed. Tel: +1 3014356917; Fax: +1 3014022084; Email: zhangcu@mail.nih.gov

Published by Oxford University Press 2012.
Table 1. A meta-analysis involving 66 independent studies for rs7903146 for reliably assessing the overall genetic association between the TCF7L2 common polymorphism and type 2 diabetes risk

<table>
<thead>
<tr>
<th>TCF7L2 Genotypes</th>
<th>n*</th>
<th>Meta-analysis Random-effects OR (95%CI)</th>
<th>P-value</th>
<th>Test of heterogeneity Q-test H</th>
<th>( P^2 (%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7903146</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/T versus C/C</td>
<td>53</td>
<td>1.75 (1.61–1.91)</td>
<td>3.85 × 10(^{-40})</td>
<td>&lt;0.0001</td>
<td>1.4 (1.2–1.6)</td>
</tr>
<tr>
<td>C/T versus C/C</td>
<td>53</td>
<td>1.37 (1.31–1.43)</td>
<td>7.50 × 10(^{-43})</td>
<td>0.0003</td>
<td>1.4 (1.1–1.6)</td>
</tr>
<tr>
<td>Per T allele (versus C allele)</td>
<td>49</td>
<td>1.41 (1.37–1.46)</td>
<td>2.49 × 10(^{-121})</td>
<td>0.11</td>
<td>1.1 (1.0–1.3)</td>
</tr>
</tbody>
</table>

*Total number of independent studies with the data available (a total of studies included: \( n = 66 \) with 55 436 cases and 106 020 controls).

for previous findings. As a result, findings from the conventional population studies cannot tease apart the impact of \( \beta \)-cell function from that of peripheral insulin/glucose action on the pathogenesis of type 2 diabetes. Moreover, residual confounding from poorly measured or unmeasured factors is always a major concern on the associations observed in observational studies (7).

The ‘Mendelian randomization’ (MR) approach (8–10), using genetic variants that are causally and specifically related to pancreatic \( \beta \)-cell function and type 2 diabetes as an instrument variable, offers the potential to assess and quantify causation using observational data. Studies exploiting the MR approach build upon the rationale that genotypes formed at conception could be indicators of lifetime exposure that are generally independent of confounders such as lifestyle and environmental factors and also free of reverse causation due to temporal preservation using pre-clinical or undiagnosed disease status (8–10). Genetic association studies have identified and confirmed a number of genetic loci significantly associated with type 2 diabetes, of which the TCF7L2 gene has shown the strongest and most widely replicated association (11–13). Importantly, function data from \textit{in vitro} and \textit{in vivo} studies demonstrated that TCF7L2 specifically affects pancreatic \( \beta \)-cell function through regulation of insulin production and secretion pathways (14–16), incretin action (15,17,18) and \( \beta \)-cell proliferation and apoptosis (15,19). The risk allele T at rs7903146 variant has recently been shown to be in an open chromatin state in human pancreatic islets accompanied by a change in enhancer activity, indicating that it may be a genuine causal variant in regulating TCF7L2 expression and splice pattern in pancreatic islets (20,21).

In addition, it is worth mentioning that direct and reliable measures of \( \beta \)-cell function require dynamic data obtained from some laborious, costly and invasive ‘gold standard’ methods (22). As a result, these measures are not practically feasible in large population studies and their predictive values for type 2 diabetes remain important yet understudied. MR approach based on the totality of evidence provides a unique opportunity to assess an unbiased association between these ‘gold standard’ measures of \( \beta \)-cell function and type 2 diabetes.

In the present study, based on an MR approach using one TCF7L2 gene variant (i.e. rs7903146) as an instrumental variable, we aimed to provide an alternative quantitative assessment of the impact of various surrogate measures of pancreatic \( \beta \)-cell function on the risk of type 2 diabetes by minimizing residual confounding and reverse causation. To maximize statistical power to obtain robust MR estimates, meta-analysis approach was employed to synthesize the evidence from the available genotype-\( \beta \)-cell function phenotype and genotype-type 2 diabetes studies, respectively.

### RESULTS

#### Characteristics of association studies

Of the 66 studies involving a total of 55 436 cases and 106 020 controls, rs7903146 was the most commonly studied TCF7L2 SNP (Supplementary Material, Table S1). Overall, there were significant differences in the minor allele frequency of rs7903146 in TCF7L2 across different populations; the minor allele T in rs7903146 was quite common (0.16–0.48) in all Caucasians, Africans and Hispanics except for Pima Indians but less frequent (0.02–0.04) in all East Asian populations, including Japanese, Chinese and Korean.

#### TCF7L2 genotype and type 2 diabetes risk

The T allele of the rs7903146 variant appeared to confer a greater risk for type 2 diabetes (Table 1). The results showed no evidence of significant between-study heterogeneity for the allelic associations. Overall, the random-effect pooled OR of the allelic association for the T allele at rs7903146 was 1.41 (95% CI, 1.37–1.46; \( P = 2.49 \times 10^{-121} \)), while there was no strong statistical evidence to suggest between-study heterogeneity for all included studies (\( P \)-value for Q statistic \( = 0.11 \), \( H = 1.1 (1.0–1.3) \) and \( P^2 = 20 (0–45) \)).

Despite significant race/ethnicity-specific frequencies of the TCF7L2 rs7903146 SNP, we found no evidence of any divergent results for increased diabetes risk associated with the TCF7L2 rs7903146 SNP across diverse ethnic groups (Table 2 and Supplementary Material, Fig. S1). Similarly, the pooled OR estimate did not differ substantially by the adjustment of body mass index (BMI) (yes versus no), statistical models (crude versus adjusted) or study designs (prospective study versus retrospective study) (Table 2).

In addition, neither a funnel plot analysis nor formal tests (Begg’s and Egger’s tests) showed evidence for the presence of substantial publication bias for the TCF7L2-type 2 diabetes association in different inheritance models (data not shown).

#### TCF7L2 and quantitative measures of \( \beta \)-cell function among non-diabetic individuals

Overall, carriers of the T allele at rs7903146 (TT and CT) consistently had lower levels of \( \beta \)-cell function measures than non-carriers (Table 3). Compared with those homozygous
indicated that the gold standard measures of (95% CI, 0.81–0.93). More importantly, MR analysis results associated with per 5% increase in HOMA-%B was 0.87 associated with risk of type 2 diabetes; the estimated causal risk. Moreover, HOMA-%B was significantly and inversely associated with lower diabetes risk. The concomitant markers are not very specific and may to some extent reflect responses, such associations were expected because these bio-

Findings from the present study provide some evidence for a causal role of pancreatic β-cell function in the development of type 2 diabetes. First, our systematic analysis of up to 55 436 diabetes cases and 106 020 controls demonstrated robust associations of the genetic variants of TCF7L2 gene with both measures of β-cell function and the risk of type 2 diabetes. Moreover, our MR causal estimates by leveraging available genetic association evidence showed that lifelong averaged levels of various β-cell function measures were significantly and inversely associated with risk of type 2 diabetes. Such significant MR estimates indicate a causal role of β-cell function alone in the etiology of type 2 diabetes free of reverse causation and confounding from many factors after birth, especially during adulthood.

**Causal association of β-cell function with risk of type 2 diabetes by MR**

Table 4 shows the predicted OR of type 2 diabetes associated with per unit increase in direct or indirect measures of β-cell function using rs7903146 variant as an instrumental variable for β-cell function. Increased fasting glucose and post-load glucose levels were strongly associated with increased risk of type 2 diabetes, whereas fasting insulin and 2 h post-load insulin levels were modestly associated with lower diabetes risk. Moreover, HOMA-%B was significantly and inversely associated with risk of type 2 diabetes; the estimated causal OR associated with per 5% increase in HOMA-%B was 0.87 (95% CI, 0.81–0.93). More importantly, MR analysis results indicated that the gold standard measures of β-cell function (i.e. AIR, SI and DI based on IVGTT), were also significantly and inversely associated with diabetes risk, although the estimates may be relatively less precise due to small number and low statistical power. The causal ORs were significant per 1-SD increment in SI (0.24; 95%, 0.08–0.74) and DI (0.14; 95% CI, 0.04–0.48).
Table 3. Pooled estimates for effects of TCF7L2 polymorphisms on phenotype traits of insulin homeostasis in healthy non-diabetic controls

<table>
<thead>
<tr>
<th>Traits</th>
<th>rs7903146: T/T versus C/C</th>
<th>Pooled mean difference</th>
<th>P-valuea</th>
<th>% heterogeneityb</th>
<th>% heterogeneityb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting insulin, pmol/l</td>
<td>25</td>
<td>3.42 (6.12 - 0.19)</td>
<td>&lt;0.001</td>
<td>0.0003</td>
<td>0.02</td>
</tr>
<tr>
<td>Glucose, mmol/l</td>
<td>26</td>
<td>0.08 (0.05 - 0.10)</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>28</td>
<td>0.00 (0.00 - 0.00)</td>
<td>&lt;0.001</td>
<td>0.0003</td>
<td>0.02</td>
</tr>
<tr>
<td>OGTT measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIR</td>
<td>4</td>
<td>0.00 (0.00 - 0.00)</td>
<td>&lt;0.001</td>
<td>0.0003</td>
<td>0.02</td>
</tr>
<tr>
<td>SI</td>
<td>4</td>
<td>0.00 (0.00 - 0.00)</td>
<td>&lt;0.001</td>
<td>0.0003</td>
<td>0.02</td>
</tr>
<tr>
<td>2 h post-load glucose, mmol/l</td>
<td>8</td>
<td>0.00 (0.00 - 0.00)</td>
<td>&lt;0.001</td>
<td>0.0003</td>
<td>0.02</td>
</tr>
<tr>
<td>HOMA-%B</td>
<td>9</td>
<td>0.00 (0.00 - 0.00)</td>
<td>&lt;0.001</td>
<td>0.0003</td>
<td>0.02</td>
</tr>
<tr>
<td>HOMA-IR and HOMA-%B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled mean difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-valuea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: P-values for pooled mean differences.

a P-values for heterogeneity between studies.

b Data are the standardized effect sizes used.

cData are the weighted mean differences for those parameters labeled with units.

OGTT, oral glucose tolerance test; HOMA-IR and HOMA-%B, homeostasis model assessment of insulin resistance and insulin secretion; IVGTT, intravenous glucose tolerance test; AIR, a measure of the first-phase insulin secretion; SI, insulin resistance index; DI, disposition index.

TCF7L2 alleles associated with diabetes risk have pleiotropic effects on metabolic or physiological systems other than pancreatic β-cell function (15,21).

The robustness of associations between TCF7L2 variants and type 2 diabetes in genetically heterogeneous populations based on our systematic analyses of up to 55,436 diabetes cases and 106,020 controls further strengthens the precision and certainty of our inferences. In previous studies, the heterogeneous nature of observational studies and limited sample sizes makes the results from individual studies less reliable. Furthermore, ethnic/racial differences in the frequencies of these known TCF7L2 variants are pronounced, with the allele frequency of the minor T allele at rs7903146 being low in East Asians but common in other populations. Such genetic heterogeneity of TCF7L2 in populations of various ethnic origins not only affects the power of individual studies but also indicates different patterns and magnitudes of linkage disequilibrium (LD) between variants in this gene among different ethnic groups (i.e. population-specific LD). Despite substantial ethnic difference in the frequency of TCF7L2 risk allele, we observed a robust association between TCF7L2 genotypes and type 2 diabetes risk in a strong genotype–dose manner with similar magnitude across subgroups. In addition, the assumption of no LD with causal variants or population stratification required for genetic instruments seemed not to be violated in our MR approach. Our meta-analysis synthesizing the final result from each ethnically homogenous population minimized the potential of population stratification.

Meeting all three basic assumptions (8,9), in the present study, the MR analysis based on the evidence from the genotype–intermediate phenotype and the genotype–disease relations provided strong evidence supporting and quantifying the causal role of β-cell function in the etiology of type 2 diabetes. It has been hypothesized that impairments in early β-cell development can lead to fetal malnutrition and predisposes individuals to the development of type 2 diabetes later in life (24–26). Beyond obesity and insulin resistance, β-cell function has been increasingly recognized as a central cause of type 2 diabetes. Because of measurement error, however, a single measure of β-cell function is not sufficient in capturing lifelong β-cell function levels. Dynamic and longitudinal measures of β-cell function are usually not practically feasible in large population studies. In addition, clinical assessment of insulin function is usually conducted during adulthood and has a hyperbolic relation with insulin action in peripheral tissues. As genetic variants in the TCF7L2 gene exert their effects across a lifetime, our estimates of genotype–phenotype relation may reflect the vast majority of variation of β-cell function measures attributable to TCF7L2 polymorphisms. Our MR estimates of the impact of β-cell function on type 2 diabetes risk may therefore largely represent a lifelong effect—a measure of average lifelong exposure, which incorporates various mitigating, compensating and attenuating patterns and lifestyle-related changes over a lifetime. In other words, our MR analysis provided precise estimates of the causal link from genetically determined lifetime levels of β-cell function to adult-onset type 2 diabetes.
which were free of reverse causation or confounding due to insulin resistance or impaired glucose metabolisms or other diabetes risk factors in adulthood.

Like all studies adopting MR approach, some limitations inherent in MR analysis merit consideration (8–10). First, hypothetically, MR analysis is analogous to a randomized controlled trial of a drug treatment or intervention—where there is always a degree of non-adherence to the treatment drug, non-compliance to protocol, drop-out, drop-in and necessary compensatory/rescue drugs used in placebo/control groups that mimic canalization and incomplete penetrance. While MR analysis does require several important assumptions that need to be carefully evaluated prior to making causal inference, some of these assumptions cannot be empirically verified and requires subject-matter knowledge (e.g. absence of pleiotropy/heterogeneity of effects and canalization/the buffering effects of germline variants by developmental compensation via other biologic pathways) (8–10). For example, it remains difficult to evaluate the assumptions of no canalization and no gene–environment interactions, which would require further experimental and intervention work. In the present study, however, the robustness and the strength of associations of multiple measures of β-cell function with type 2 diabetes risk using both conventional multivariable method and the MR analysis provide further some assurance of precision and certainty of our inferences. Secondly, β-cell function measures derived from IVGTT are usually regarded as the gold standard measures. Due to the limited number of available studies with IVGTT measures and the heterogeneity in IVGTT test protocols, confidence intervals (CI) of odds ratio (OR) or relative risk (RR) estimates of the genotype–phenotype association and the MR estimates of causal effect based on the IVGTT measures were somewhat wider and may be relatively less precise.

In conclusion, findings from the present study based on MR approach and systematic analyses of available population data lend support to the causal role of pancreatic β-cell function in the pathogenesis of type 2 diabetes. These findings not only further highlight the compelling need for the development and application of validated β-cell function measures for population studies, but also further support using preservation of pancreatic β-cell function as a target for the prevention or treatment of type 2 diabetes in general population.

**MATERIALS AND METHODS**

**Study selection**

Relevant studies were identified by searching MEDLINE, PubMed, EMBASE and NCBI OMIM databases for all published genetic association studies up to February 1, 2010, using the search terms ‘TCF7L2’, ‘pancreatic beta-cell function’, ‘insulin secretion’, ‘insulin resistance’, ‘glucose metabolism’, ‘diabetes mellitus’, ‘type 2 diabetes’ and ‘diabetes’. Additional studies were retrieved through a hand-search of references from original reports. Two common variants (i.e. rs7903146 and rs12255372) have been widely studied for TCF7L2 gene. Overall, more studies are available for the rs7903146 variant such that statistical power from available data was much higher for rs7903146 than rs12255372 in terms of both genotype–phenotype and genotype–diabetes associations. Moreover, the rs7903146 variant is generally in strong LD with the rs12255372 variant except in several African groups. In addition, it has been demonstrated that the variant affects β-cell function by altering different genomic regulatory functions in pancreatic islets (27). In the present study, our main analysis focused on the rs7903146 variant only. All studies on this topic were considered eligible if they had data on the associations of the rs7903146 variant with the risk of type 2 diabetes and/or diabetes-related quantitative traits. Only English language articles were identified and included.

Two authors (Y.S. and C.L.) independently reviewed each published paper and extracted relevant information examining the associations of TCF7L2 gene variants with the risk of type 2 diabetes. Two separate authors (E.Y. and L.C.) independently reviewed each published paper and extracted relevant information examining the associations of TCF7L2 variants with intermediate metabolic phenotypes. When a study reported results on different subpopulations, we considered each subpopulation as a separate study in the meta-analysis. When necessary, missing information (such as genotype distributions,
standard deviations or 95% CI) was obtained by direct contact with the original authors.

Of the 647 retrieved articles that were evaluated and abstracted, 575 studies were excluded (Fig. 1). These included non-human studies, review/editorials, case only studies, studies with abstract only, meta-analysis, redundant reports of the same population and reports that did not provide relevant data. The final data set of our meta-analyses included 72 published articles that provided 66 independent studies for rs7903146. Data on TCF7L2 genotypes and diabetes-related metabolic traits were also available depending on the phenotype (e.g. for rs7903146 genotypes, the sample sizes were 160 and 1045 for the IVGTT measures comparing TT with CC and 18209 and 27446 for the comparison of CT and CC on fasting glucose).

Data extraction

For each of the articles reporting the TCF7L2-type 2 diabetes association, we included studies that provided estimates of RRs or ORs of type 2 diabetes or data that permitted estimation of these parameters. We developed a database that included the first author’s name, year of publication, study population, sample size, mean age for cases and controls, mean age at diagnosis, mean BMI comparisons of allele frequencies between cases and controls, genotype distribution in cases and controls and the estimates of RRs or ORs of type 2 diabetes associated with the genotypes of TCF7L2. For studies reporting the relation of TCF7L2 with diabetes-related phenotypes, we extracted the number of subjects, the means and standard deviations of quantitative metabolic phenotypes that may indirectly or directly reflect β-cell function, including fasting plasma glucose and insulin levels, HOMA-B, plasma levels of glucose and insulin at 2 h after a 75 g oral glucose tolerance test (OGTT) and acute insulin response (AIR), insulin sensitivity index (SI) and disposition index (DI) derived from intravenous glucose tolerance test (IVGTT). Information on homeostasis model assessment index of insulin resistance (HOMA-IR) was also extracted to evaluate the specificity of the association with insulin sensitivity when compared with the β-cell function parameters discussed earlier. When results were not presented in the published papers, attempts were made to obtain additional data by directly contacting 24 corresponding authors.

Statistical analyses

**Association of TCF7L2 variants with intermediate phenotypes and type 2 diabetes**

To maximize statistical power, a meta-analysis was performed to provide robust effect estimates of the associations of TCF7L2 variant with measures of β-cell function and risk of type 2 diabetes, which are required for the MR algorithm. We calculated the ORs and 95% CI for the associations between TCF7L2 genotypes and risk of type 2 diabetes in three different inheritance models. When the adjusted ORs were not available, we used the crude ORs reported or calculated from the raw data. We performed pre-defined stratified meta-analyses to explore potential heterogeneity (effect modification) of the TCF7L2 genotype-diabetes association by
TCF7L2
Genotype (G) → β Cell Function Pheno-
type (X) → Type 2 Diabetes (Y)

Three major criteria when evaluating the suitability of
genotype as randomization instrument

1. G is independent of U
2. G is robustly associated with X
3. G is independent of Y given X and U (i.e. G affects Y only through X)

Additional assumptions:
The absence of strong LD, gene-environment interaction, population stratification, pleiotropy, and canalization.

Figure 2. Directed acyclic graph (DAG) encoding the causal relationship between TCF7L2 genetic instrumental variables G, intermediate β-cell function phenotype X and type 2 diabetes outcome Y that satisfies three key assumptions.

Causal inference using MR analysis

The MR approach incorporating information on both the genotype–intermediate phenotype association and genotype–disease association into one analytical framework may allow for an unbiased estimate of the intermediate phenotype–disease association (Figs 2 and 3). In the MR paradigm, an instrumental variable has to satisfy the following three criteria: (i) the genotype should be robustly associated with intermediate phenotype (e.g. TCF7L2 gene – β-cell function measures); (ii) the genotype should not be associated with confounding factors that bias the association between intermediate phenotype and disease outcome and (iii) absence of pleiotropy, as the genotype should exert its effect on the clinical outcome only through the specific intermediate phenotype (e.g. effect of TCF7L2 gene on type 2 diabetes risk via β-cell function alone). Based on available evidence (1–3,10,14–16,23,36, TCF7L2 genotypes seemed to meet these assumptions very well. Thus, the MR coefficient estimates using TCF7L2 genotypes as instruments would provide the causal association between β-cell function and type 2 diabetes risk free of bias due to reverse causation and residual confounding.

β-cell function is evaluated using multiple indexes on a continuous scale, and the TCF7L2 genotype is considered as binary (either TT versus CC or TC versus CC). Based on the estimates of the log odds ratio (log OR) of type 2 diabetes given TCF7L2 genotype and the mean difference in the
phenotype levels between the two TCF7L2 genotypes, we estimated the MR causal effect of the phenotype levels on type 2 diabetes using two models (A and B) described by Thompson et al. (also summarized in one Supplementary Material) (37). The two models take into account the correlations between genotype–phenotype and genotype–disease associations from the same studies under different assumptions. Model B assumes independence between the heterogeneities on the genotype–phenotype association and phenotype–disease association (i.e. the between-studies correlation of heterogeneities equals to zero) while Model A assumes only no within-study correlation, i.e. the study specific OR estimates are independent of the phenotype level differences. We conducted further analyses to justify the use of Model B by first testing the independence assumption and then by comparing Model B and Model A using the observed $-2\log$ likelihood principal. Our analyses found no evidence to reject the independence assumption. Furthermore, Model B is either substantially better than or about equivalent to Model A by the $-2\log$ likelihood criterion. In addition, under the assumption that the diabetes prevalence is relatively low in the general population, the OR obtained from our MR approach of case–control data can be interpreted as a causal OR (38). All MR estimations were conducted using the R software.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

**FUNDING**

C.Z., E.Y., C.L. and E.F.S. are supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health & Human Development, National Institutes of Health. T.J.V. is supported by grants from the National Institutes of Health (ES017876 and HD060696).

**REFERENCES**