Predicting Risk of Type 2 Diabetes Mellitus with Genetic Risk Models on the Basis of Established Genome-wide Association Markers: A Systematic Review

Wei Bao, Frank B. Hu, Shuang Rong, Ying Rong, Katherine Bowers, Enrique F. Schisterman, Liegang Liu, and Cuilin Zhang*

* Correspondence to Dr. Cuilin Zhang, Epidemiology Branch, Division of Epidemiology, Statistics, and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, 6100 Executive Boulevard, Rockville, MD 20852 (e-mail: zhangcu@mail.nih.gov).

Initially submitted November 29, 2012; accepted for publication May 21, 2013.

This study aimed to evaluate the predictive performance of genetic risk models based on risk loci identified and/or confirmed in genome-wide association studies for type 2 diabetes mellitus. A systematic literature search was conducted in the PubMed/MEDLINE and EMBASE databases through April 13, 2012, and published data relevant to the prediction of type 2 diabetes based on genome-wide association marker–based risk models (GRMs) were included. Of the 1,234 potentially relevant articles, 21 articles representing 23 studies were eligible for inclusion. The median area under the receiver operating characteristic curve (AUC) among eligible studies was 0.60 (range, 0.55–0.68), which did not differ appreciably by study design, sample size, participants’ race/ethnicity, or the number of genetic markers included in the GRMs. In addition, the AUCs for type 2 diabetes did not improve appreciably with the addition of genetic markers into conventional risk factor–based models (median AUC, 0.79 (range, 0.63–0.91) vs. median AUC, 0.78 (range, 0.63–0.90), respectively). A limited number of included studies used reclassification measures and yielded inconsistent results. In conclusion, GRMs showed a low predictive performance for risk of type 2 diabetes, irrespective of study design, participants’ race/ethnicity, and the number of genetic markers included. Moreover, the addition of genome–wide association markers into conventional risk models produced little improvement in predictive performance.

Abbreviations: AUC, area under the receiver operating characteristic curve; CRM, conventional risk factors–based model; GRM, genome-wide association marker–based risk model; GWA, genome-wide association; IDI, integrated discrimination improvement; NRI, net reclassification improvement; SNP, single nucleotide polymorphism; T2DM, type 2 diabetes mellitus.
developed and validated (7, 8). Recently, the advent of genome-wide association (GWA) studies has presented an exciting opportunity to incorporate novel genetic variants into the risk prediction models for T2DM (9). So far, 59 loci associated with T2DM susceptibility (herein called GWA markers) have been identified and/or confirmed at the genome-wide significance level \(P < 5 \times 10^{-8}\) in GWA studies or meta-analyses of GWA studies (10). To translate emerging genomic knowledge into clinical applications, GWA marker–based genetic risk scores or genotype scores have been developed for the prediction of T2DM risk (7, 8, 11). Meanwhile, the “direct-to-consumer” genetic profiling for the prediction of T2DM risk has been offered by commercial companies (12). Studies based on simulated data have also demonstrated that combined information from multiple common genetic variants could improve the prediction of complex diseases (13, 14). However, empirical studies have not provided clear evidence to support the utility of incorporating genomic information into T2DM risk prediction (15–18).

The primary objective of this systematic review was to summarize the predictive performance of genome-wide association marker–based risk models (GRMs) for T2DM risk. The secondary objective was to evaluate whether adding GWA markers to conventional risk factor–based models (CRMs) improves the prediction of T2DM risk.

**MATERIALS AND METHODS**

**Eligibility criteria**

We adhered to guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (19) when undertaking this study. The C statistic or area under the receiver operating characteristic curve (AUC) (20–22), the most widely used metric in genetic prediction studies, was the main parameter of the population discriminative ability in this systematic review. The AUC allows comparison of the discriminative accuracy of diverse prediction models independent of the choice of cutoff value in different studies. When the sensitivity and specificity of a test are calculated for each possible cutoff value and plotted as receiver operating characteristic curves, the AUCs, which may vary from 0.5 (no discrimination) to 1 (perfect discrimination), measure the discriminative ability of the test. Recently, reclassification measures (20), including the net reclassification improvement (NRI) and integrated discrimination improvement (IDI), have been used as alternatives to the increase of the AUC for evaluating the incremental predictive performance of GWA markers. Thus, studies were eligible if they reported the AUC or these reclassification measures of GRMs for the prediction of T2DM risk. We did not use odds ratios or risk ratios because the magnitude of genetic associations or effect size does not closely correspond to predictive performance (23, 24).

**Data sources and searches**

Genetic association studies regarding T2DM prediction were searched in the PubMed/MEDLINE and EMBASE databases through April 13, 2012, by using a combination of free text and subheadings from MeSH and EMTREE terms. The following terms were used for the PubMed/MEDLINE search: (“diabetes mellitus, type 2[MeSH] or type 2 diabetes[tiab]) and (“polymorphism, single nucleotide” [MeSH] or “genotype”[MeSH] or “alleles”[MeSH] or “genetic variation”[MeSH] or (“genetic risk score” or (“genetic score” or (“genotype score” or (“genetic variant” or genotyper*[tiab] or allele*[tiab]) and (“ROC curve”[MeSH] or “area under curve”[MeSH] or “area under the curve” or “AUC” or “AUCs” or “AROC” or (“C statistic” or predict*[tiab] or discriminat*[tiab] or reclassification or net reclassification improvement or integrated discrimination improvement), not (review[pt] or editorial[pt]). Similar search terms were used for the EMBASE database. In addition, the references listed in relevant articles were screened. No restrictions on language, geographical location, or study design (e.g., cross-sectional, case-control, cohort study) were applied in the literature search process; however, conference abstracts without sufficient data were not included.

**Study selection**

All of the indexed articles were evaluated independently by 2 reviewers (S.R. and Y.R.), and disagreements regarding eligibility were solved in consultation with a third reviewer (W.B.). The process of study selection is depicted in Figure 1. During the screening steps, we excluded review articles, editorials, and protocols, as well as the following study types: nonhuman studies (cell culture or animal studies); studies that did not assess genetic associations; studies with outcomes of obesity, prediabetes, metabolic syndrome, or other diseases but not T2DM; studies on quantitative traits for T2DM (e.g., glucose, insulin, hemoglobin A1c, lipid parameters, insulin sensitivity, β cell function); and studies on T2DM complications or comorbid diseases. In addition, pharmacogenetic or pharmacogenomic studies for antidiabetic drugs and genetic association studies that did not report the predictive performance of GWA markers for T2DM were also excluded. Three additional articles (16, 25, 26) were excluded because either their results (16, 25) were updated by the same group in more recent reports (15, 27), or the results (26) were previously reported in the same population (28). Another article was excluded because it did not report data separately for prediabetes and T2DM (29).

**Data extraction and quality assessment**

Two reviewers (W.B. and S.R.) independently extracted data and evaluated study quality, and disagreements were solved by consensus. The following data were extracted from each published article: author’s name, year of publication, characteristics of study subjects (e.g., age, sex, body mass index), sample size, genetic variants, AUC or reclassification measures, and consistency of genotype frequencies with Hardy-Weinberg equilibrium, if available. In studies that reported combined effects in prediction models incorporating both GWA markers and nongenetic risk factors, information about the nongenetic risk model was also extracted.

To assess study quality, we considered items in the Strengthening the Reporting of Genetic Risk Prediction Studies statement (30). We particularly evaluated the following items: study design (cross-sectional, case-control or prospective
cohort), selection criteria and basic characteristics for participants in the study, genetic variants definition, measurement, coding, and risk model construction.

Data synthesis and analysis

AUCs for T2DM were the main measures in this systematic review because they were reported in almost all of the included studies. In addition, we reviewed the studies in which reclassification measures (i.e., NRI and/or IDI) were reported.

For studies that reported multiple AUCs for different genetic models, the model with the most comprehensive information was used. Statistical meta-analysis of the AUCs and their 95% confidence intervals was conducted to quantitatively summarize the findings in the included studies, and the detailed methods and results are shown in the Web Appendix and in Web Figures 1–4 available at http://aje.oxfordjournals.org/. Because of the various components included in the risk prediction models and the large heterogeneity of data, a descriptive summary (i.e., median) of AUCs and their ranges (minimum to maximum) were used in the main text. Stratification analyses were conducted according to study design (cross-sectional, case-control, and prospective cohort); sample size (i.e., number of T2DM cases); race/ethnicity (Caucasian, Asian, or other); the number of genetic variants included in the GRMs (<10, 10–19, or ≥20); and the average age at diagnosis of T2DM (<50 years or ≥50 years).

For our primary objective, cross-sectional, case-control, and cohort studies were all eligible for inclusion because there was no concern about the temporal causality between exposure (i.e., genotypes that were predetermined during gamete formation and conception) and outcome (i.e., T2DM). For the secondary objective, we restricted the analysis to cohort studies because of the possibility of reverse causality between conventional risk factors and T2DM (e.g., individuals with T2DM may change their lifestyles after diagnosis) that might arise in cross-sectional and case-control studies.

Figure 1. Flow chart for study selection. AUC, area under the receiver operating characteristic curve; T2DM, type 2 diabetes mellitus.
RESULTS

Study characteristics

We identified 1,234 potentially relevant articles from the PubMed/MEDLINE and EMBASE databases. After screening, we evaluated 259 articles in detail. Finally, 21 articles representing 23 studies (15, 27, 28, 31–49) published through April 13, 2012, were eligible for this systematic review (Figure 1). The studies by Lyssenko et al. (15) and Xu et al. (47) comprised 2 independent populations; therefore, they were treated as 2 studies in each of the articles. The study design, participants’ characteristics, and predictive performance for T2DM by using genetic risk models, conventional risk models, and combined models in the individual studies are shown in Tables 1 and 2. Of these eligible studies, 13 were conducted in Europe, 6 in Asia, 3 in the United States, and 1 in North Africa; 11 were prospective cohort studies, 9 were case-control studies, and 3 were cross-sectional studies. Most studies, but not all (33, 40, 41, 44), reported sufficient detail about selection criteria and basic characteristics for participants in the study population. All studies described the selection and measurement of genetic variants and the construction of GRMs. The number of genetic variants in the GRMs ranged from 3 to 40. The most common GWA markers included in the GRMs were solute carrier family 30 (zinc transporter), member 8 (SLC30A8) rs13266634; cyclin-dependent kinase inhibitor 2A/2B (CDKN2A/2B) rs10811661; transcription factor 7-like 2 (TCF7L2) rs7903146; hematopoietically expressed homeobox (HHEX) insulin-degrading enzyme (IDE) rs1111875; insulin-like growth factor 2 mRNA binding protein 2 (IGF2BP2) rs4402960; JAZF zinc finger 1 (JAZF1) rs864745; thyroid adenoma-associated (THADA) rs7578597; and ADAM metallopeptidase with thrombospondin type 1 motif, 9 (ADAMTS9) rs4607103. The most common components in the CRMs were age, sex, and body mass index. Additional components included blood pressure, waist circumference, family history of diabetes, and biochemical markers (e.g., fasting plasma glucose, triglycerides, high-density lipoprotein cholesterol). In addition, 3 studies included established diabetes risk scores as the CRMs, such as the Finnish Diabetes Risk Score (41), the German Diabetes Risk Score (37), the Cambridge Diabetes Risk Score (40), and the Framingham Offspring Diabetes Risk Score (40) (Web Table 1).

Predictive performance of GRMs

The predictive performance of GRMs for T2DM was reported in 19 of the 23 eligible studies. In general, GRMs showed a relatively poor discrimination for T2DM in all studies (Table 2). The AUCs ranged from 0.55 to 0.68 with a median of 0.60.

In our subgroup analyses stratified by study design, sample size (i.e., number of T2DM cases) and race/ethnicity (Table 3), the median AUCs were as follows: for case-control studies, 0.62 (range, 0.58–0.63); for cohort studies, 0.60 (range, 0.55–0.68); for cross-sectional studies, 0.57 (range, 0.55–0.59); for studies with fewer than 1,000 T2DM cases, 0.57 (range, 0.55–0.68); for studies with 1,000–1,999 T2DM cases, 0.61 (range, 0.59–0.63); for studies with 2,000 or more T2DM cases, 0.60 (range, 0.58–0.63); for studies of Caucasians, 0.59 (range, 0.55–0.68); and for studies of Asians, 0.63 (range, 0.62–0.63).

We also examined whether the AUCs for T2DM varied by the number of genetic variants included in the GRMs, because the Framingham Offspring Study showed that increasing the number of GWA markers in the GRMs from 18 in a previous report (16) to 40 in a recent report (27) improved the AUC for T2DM risk prediction, although the predictive performance of the updated model was still limited, and the magnitude of the change between the median AUCs was small (0.58 (range, 0.55–0.62) vs. 0.63 (range, 0.61–0.66)). In a subgroup study stratified by the number of genetic variants included in the GRMs, the median AUCs for T2DM were 0.57 (range, 0.56–0.58), 0.60 (range, 0.55–0.68), and 0.59 (range, 0.55–0.63) for GRMs including <10, 10–19, and ≥20 genetic variants, respectively.

In addition, a recent report suggested that the predictive performance of GRMs may vary by participant age (27). Therefore, we performed a subgroup analysis of cases by age (mean or median) at diagnosis or testing. We found a slightly higher AUC for studies with an average age of study participants younger than 50 years compared with those that included older study participants (with median AUCs of 0.62 (range, 0.55–0.68) and 0.60 (range, 0.55–0.63), respectively).

Predictive performance of GWA markers versus CRMs

To assess the incremental improvement in predictive performance of GWA markers beyond that of CRMs for T2DM risk, we included prospective cohort studies in which AUCs and 95% confidence intervals for CRMs with and without the addition of GRMs were reported. Among the 11 prospective cohort studies, the median AUCs were 0.78 (range, 0.63–0.90) for the CRMs and 0.79 (range, 0.63–0.91) for combined models that included both conventional risk factors and GWA markers.

In addition to reporting AUCs, several studies (15, 27, 36, 37, 40, 48) also reported reclassification measures as metrics for incremental predictive performance (Table 4). The categorical NRI is a cutoff point–dependent measure; however, the cutoff points were quite different among studies that reported NRI (15, 27, 40). Among studies that reported IDI (36, 37, 48), only 1 provided sufficient information on IDI and its 95% confidence interval (37). The results reported in the included studies were inconsistent; some studies found statistically significant improvement in NRI and/or IDI (15, 27, 36), whereas others did not (40, 48). However, even in studies that reported statistically significant improvement, the magnitude of NRI and/or IDI was modest (~4.5% for NRI (15, 27) and 1.2% for relative IDI (36)) compared with results (NRI of ≤39% and IDI of ≤7.8%) from an empirical evaluation of the use of reclassification for assessment of improved prediction (50). This is consistent with the results of an ad hoc study showing that reclassification observed in the absence of an increase of AUC is unlikely to improve clinical utility (21).

DISCUSSION

In the current systematic review, we found that GRMs showed a relatively low predictive performance for T2DM risk irrespective of study design, participants’ race/ethnicity...
<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Study Location</th>
<th>Ethnic Origin</th>
<th>Study Design</th>
<th>No. of Participants</th>
<th>Age, years&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Male Sex, %</th>
<th>BMI&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weedon, 2006 (31)</td>
<td>United Kingdom</td>
<td>Caucasian</td>
<td>Case-control</td>
<td>2,409 3,668</td>
<td>48.7 31.8</td>
<td>58 50</td>
<td>31.4 27.2</td>
</tr>
<tr>
<td>Cauchi, 2008 (43)</td>
<td>France</td>
<td>Caucasian</td>
<td>Case-control</td>
<td>3,295 3,595</td>
<td>62 56</td>
<td>61.8 42.3</td>
<td>28.3 24.9</td>
</tr>
<tr>
<td>Cauchi, 2008 (43)</td>
<td>France</td>
<td>Caucasian</td>
<td>Case-control</td>
<td>937 1,000</td>
<td>66 50</td>
<td>61.7 42.8</td>
<td>31.1 24.1</td>
</tr>
<tr>
<td>Lango, 2008 (32)</td>
<td>United Kingdom</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>2,309 2,598</td>
<td>55.7 NA</td>
<td>56 51</td>
<td>31.5 26.9</td>
</tr>
<tr>
<td>Lyssenko, 2008 (15)</td>
<td>Sweden</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>2,063 13,998</td>
<td>45.5&lt;sup&gt;c&lt;/sup&gt; 64.9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>24.3&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Lyssenko, 2008 (15)</td>
<td>Finland</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>138 2,632</td>
<td>44.9&lt;sup&gt;c&lt;/sup&gt; 45.5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>25.6&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>van Hoek, 2008 (28)</td>
<td>The Netherlands</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>601 5,221</td>
<td>68.2 69.0</td>
<td>44.3 40.4</td>
<td>28.0 26.0</td>
</tr>
<tr>
<td>Vaxillaire, 2008 (33)</td>
<td>France</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>523 2,919</td>
<td>NA NA</td>
<td>NA NA</td>
<td>NA NA</td>
</tr>
<tr>
<td>Cornelis, 2009 (34)</td>
<td>United States</td>
<td>Caucasian</td>
<td>Cohort (men)</td>
<td>1,197 1,338</td>
<td>55.7 55.4</td>
<td>100 100</td>
<td>27.8 25.1</td>
</tr>
<tr>
<td>Cornelis, 2009 (34)</td>
<td>United States</td>
<td>Caucasian</td>
<td>Cohort (women)</td>
<td>1,612 2,163</td>
<td>44.1 43.6</td>
<td>0 0</td>
<td>27.7 23.9</td>
</tr>
<tr>
<td>Hu, 2009 (35)</td>
<td>China</td>
<td>Asian</td>
<td>Case-control</td>
<td>1,849 1,785</td>
<td>61.2 57.4</td>
<td>52.5 41.2</td>
<td>24.0 23.6</td>
</tr>
<tr>
<td>Lin, 2009 (36)</td>
<td>Switzerland</td>
<td>Caucasian</td>
<td>Cross-sectional</td>
<td>356 5,004</td>
<td>60.7 52.8</td>
<td>67.4 46.0</td>
<td>30.4 25.5</td>
</tr>
<tr>
<td>Miyake, 2009 (45)</td>
<td>Japan</td>
<td>Asian</td>
<td>Case-control</td>
<td>2,316 2,370</td>
<td>61.3 67.5</td>
<td>58.2 45.9</td>
<td>23.6 23.3</td>
</tr>
<tr>
<td>Schulze, 2009 (37)</td>
<td>Germany</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>579 1,962</td>
<td>54.6 49.4</td>
<td>58.7 36.9</td>
<td>30.4 25.9</td>
</tr>
<tr>
<td>Sparso, 2009 (38)</td>
<td>Denmark</td>
<td>Caucasian</td>
<td>Case-control</td>
<td>4,093 5,302</td>
<td>60 47</td>
<td>59.3 46.3</td>
<td>30.6 25.6</td>
</tr>
<tr>
<td>Fontaine-Bisson, 2010 (46)</td>
<td>Sweden</td>
<td>Caucasian</td>
<td>Cross-sectional</td>
<td>1,327 1,424</td>
<td>53.6 53.1</td>
<td>58.4 50.2</td>
<td>29.5 25.8</td>
</tr>
<tr>
<td>Qi, 2010 (39)</td>
<td>China</td>
<td>Asian</td>
<td>Case-control</td>
<td>424 1,908</td>
<td>58.6 58.8</td>
<td>51.2 58.5</td>
<td>25.8 23.8</td>
</tr>
<tr>
<td>Talmud, 2010 (40)</td>
<td>United Kingdom</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>302 5,233</td>
<td>49&lt;sup&gt;c&lt;/sup&gt; 67&lt;sup&gt;c&lt;/sup&gt;</td>
<td>NA NA</td>
<td>NA NA</td>
</tr>
<tr>
<td>Wang, 2010 (41)</td>
<td>Finland</td>
<td>Caucasian</td>
<td>Cross-sectional</td>
<td>518 6,714</td>
<td>45–74&lt;sup&gt;d&lt;/sup&gt;</td>
<td>NA NA</td>
<td>NA NA</td>
</tr>
<tr>
<td>Xu, 2010 (47)</td>
<td>China</td>
<td>Asian</td>
<td>Case-control</td>
<td>1,825 2,200</td>
<td>63.3 59.3</td>
<td>43.9 38.4</td>
<td>26.3 24.3</td>
</tr>
<tr>
<td>Xu, 2010 (47)</td>
<td>China</td>
<td>Asian</td>
<td>Cohort</td>
<td>67 667</td>
<td>61.5 61.0</td>
<td>35.8 36.7</td>
<td>26.1 24.8</td>
</tr>
<tr>
<td>de Miguel-Yanes, 2011 (27)</td>
<td>United States</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>446 3,025</td>
<td>46&lt;sup&gt;c&lt;/sup&gt; 46.5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>26.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Hivert, 2011 (48)</td>
<td>United States</td>
<td>Mixed</td>
<td>Cohort</td>
<td>2,843&lt;sup&gt;d&lt;/sup&gt;</td>
<td>50.6&lt;sup&gt;d&lt;/sup&gt; 33.2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>34.0&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Cauchi, 2012 (49)</td>
<td>Morocco</td>
<td>Arab</td>
<td>Case-control</td>
<td>1,193 1,055</td>
<td>58 54</td>
<td>65.6 69.7</td>
<td>28.4 27.7</td>
</tr>
<tr>
<td>Cauchi, 2012 (49)</td>
<td>Tunisia</td>
<td>Arab</td>
<td>Case-control</td>
<td>1,446 942</td>
<td>61 61</td>
<td>54.2 55.7</td>
<td>27.6 24.4</td>
</tr>
<tr>
<td>Janipalli, 2012 (42)</td>
<td>India</td>
<td>Asian</td>
<td>Case-control</td>
<td>1,808 1,549</td>
<td>47.7 NA</td>
<td>55.8 53.2</td>
<td>25.7 19.4</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; NA, not available.

<sup>a</sup> Age indicates average age at diagnosis or testing.
<sup>b</sup> Calculated as weight (kg)/height (m)<sup>2</sup>.
<sup>c</sup> Values represent the entire cohort.
<sup>d</sup> Age range for the entire cohort.
and conception, they do not change over time, and the temporal sequence of genotype-phenotype can be clearly established for outcome predictions (56). However, the predictive performance of genetic variants for T2DM may have been overestimated (57). Although an early study indicated an impressive 20-fold increased risk of T2DM by using the combination of 3 genetic variants among individuals who were obese and who had elevated fasting plasma glucose values (25), the discriminative accuracy for T2DM risk prediction did not significantly improve in a reexamination of the same study (23). In the present study, we found that the AUCs of GRMs for T2DM was relatively low, and CRMs, which incorporate age, BMI, and other factors, demonstrated appreciably higher AUCs than did GRMs. One may speculate that genetic profiling, theoretically, could be more useful for predicting

Table 2. AUCs for Genetic Risk Models, Conventional Risk Models, and Combined Models for Predicting Risk of Type 2 Diabetes in the Studies Included in the Systematic Review

<table>
<thead>
<tr>
<th>First Author, Year</th>
<th>Study Location</th>
<th>Genetic Risk Model</th>
<th>Conventional Risk Model</th>
<th>Combined Model</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weedon, 2006 (31)</td>
<td>United Kingdom</td>
<td>0.58 0.57, 0.59</td>
<td>NA  NA</td>
<td>NA  NA</td>
<td>NA  NA</td>
</tr>
<tr>
<td>Cauchi, 2008 (43)</td>
<td>France</td>
<td>NA  NA</td>
<td>NA  NA</td>
<td>NA  NA</td>
<td>NA  NA</td>
</tr>
<tr>
<td>Lango, 2008 (32)</td>
<td>United Kingdom</td>
<td>0.60 0.58, 0.62</td>
<td>0.78 0.77, 0.79</td>
<td>0.80 0.79, 0.81</td>
<td>2.9 × 10^{-12}</td>
</tr>
<tr>
<td>Lyssenko, 2008 (15)</td>
<td>Sweden</td>
<td>0.62 0.61, 0.63</td>
<td>0.74 0.73, 0.75</td>
<td>0.75 0.74, 0.76</td>
<td>1.0 × 10^{-4}</td>
</tr>
<tr>
<td>Lyssenko, 2008 (15)</td>
<td>Finland</td>
<td>0.68 0.63, 0.73</td>
<td>0.79 0.74, 0.84</td>
<td>0.80 0.76, 0.84</td>
<td>NA</td>
</tr>
<tr>
<td>van Hoek, 2008 (28)</td>
<td>The Netherlands</td>
<td>0.60 0.57, 0.63</td>
<td>0.66 0.63, 0.68</td>
<td>0.68 0.66, 0.71</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Vaxillaire, 2008 (33)</td>
<td>France</td>
<td>0.56 0.53, 0.59</td>
<td>0.82 0.82, 0.83</td>
<td>0.83 0.82, 0.83</td>
<td>0.26</td>
</tr>
<tr>
<td>Cornelis, 2009 (34)</td>
<td>United States</td>
<td>0.59 0.57, 0.60</td>
<td>0.78 0.77, 0.79</td>
<td>0.79 0.78, 0.80</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hu, 2009 (35)</td>
<td>China</td>
<td>0.62 0.60, 0.64</td>
<td>0.61 0.60, 0.63</td>
<td>0.67 0.65, 0.69</td>
<td>0.0002</td>
</tr>
<tr>
<td>Lin, 2009 (36)</td>
<td>Switzerland</td>
<td>0.57 0.54, 0.60</td>
<td>0.86 0.84, 0.88</td>
<td>0.87 0.85, 0.89</td>
<td>0.002</td>
</tr>
<tr>
<td>Miyake, 2009 (45)</td>
<td>Japan</td>
<td>0.63 NA</td>
<td>0.68 NA</td>
<td>0.72 NA</td>
<td>NA</td>
</tr>
<tr>
<td>Schulze, 2009 (37)</td>
<td>Germany</td>
<td>0.55 0.53, 0.58</td>
<td>0.90 0.89, 0.91</td>
<td>0.90 0.89, 0.91</td>
<td>0.6868</td>
</tr>
<tr>
<td>Sparso, 2009 (38)</td>
<td>Denmark</td>
<td>0.60 0.59, 0.61</td>
<td>0.92 0.91, 0.93</td>
<td>0.93 0.92, 0.94</td>
<td>NA</td>
</tr>
<tr>
<td>Fontaine-Bisson, 2010 (46)</td>
<td>Sweden</td>
<td>0.59 NA</td>
<td>0.65 NA</td>
<td>0.72 NA</td>
<td>NA</td>
</tr>
<tr>
<td>Qi, 2010 (39)</td>
<td>China</td>
<td>0.62 0.59, 0.65</td>
<td>0.77 0.74, 0.80</td>
<td>0.79 0.76, 0.81</td>
<td>0.007</td>
</tr>
<tr>
<td>Talmud, 2010 (40)</td>
<td>United Kingdom</td>
<td>0.55 0.51, 0.59</td>
<td>0.78 0.75, 0.82</td>
<td>0.78 0.74, 0.81</td>
<td>0.10</td>
</tr>
<tr>
<td>Wang, 2010 (41)</td>
<td>Finland</td>
<td>0.55 0.53, 0.58</td>
<td>0.77 0.75, 0.79</td>
<td>0.77 0.75, 0.79</td>
<td>NA</td>
</tr>
<tr>
<td>Xu, 2010 (47)</td>
<td>China</td>
<td>NA  NA</td>
<td>0.71 NA</td>
<td>0.73 NA</td>
<td>NA</td>
</tr>
<tr>
<td>Xu, 2010 (47)</td>
<td>China</td>
<td>NA  NA</td>
<td>0.63 NA</td>
<td>0.66 NA</td>
<td>NA</td>
</tr>
<tr>
<td>de Miguel-Yanes, 2011 (27)</td>
<td>United States</td>
<td>0.63 0.61, 0.66</td>
<td>0.90 0.89, 0.92</td>
<td>0.91 0.89, 0.92</td>
<td>0.01</td>
</tr>
<tr>
<td>Hivert, 2011 (48)</td>
<td>United States</td>
<td>NA  NA</td>
<td>0.63 NA</td>
<td>0.63 NA</td>
<td>0.34</td>
</tr>
<tr>
<td>Cauchi, 2012 (49)</td>
<td>Morocco and Tunisia</td>
<td>0.60 NA</td>
<td>0.64 NA</td>
<td>0.67 NA</td>
<td>0.004</td>
</tr>
<tr>
<td>Janipalli, 2012 (42)</td>
<td>India</td>
<td>0.63 0.62, 0.65</td>
<td>0.96 0.95, 0.97</td>
<td>0.96 0.96, 0.97</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the curve; CI, confidence interval; GWA, genome-wide association; NA, not available; T2DM, type 2 diabetes mellitus.

a Risk prediction model based on genetic variants identified and/or confirmed in GWA studies of T2DM.
b Risk prediction model based on conventional risk factors of T2DM (e.g., age, sex, body mass index (weight (kg)/height (m)^2), family history of diabetes).
c Risk prediction model based on both genetic variants identified or confirmed in GWA studies of T2DM and conventional risk factors of T2DM.
d P value for difference between the AUC for type 2 diabetes with a conventional risk factor–based model and with a combined model, indicating the incremental value when adding GWA markers into the conventional risk factor–based model.

(i.e., Caucasian or Asian), and the number of genetic markers included, despite the fact that the associations of the included GWA markers with T2DM risk have been well established and replicated in previous studies (51–54). The risk prediction models for T2DM have been previously reviewed (7, 8, 11, 55); however, all of these reviews except Mihaescu et al. (11) focused mainly on the conventional risk factor–based models. Our results support the notion that known T2DM GWA markers add minimally to the predictive performance for T2DM beyond that of conventional risk factors (15, 16).

Genetic testing has been suggested for identifying individuals at risk of developing T2DM (25). Indeed, compared with nongenetic risk factors, genetic variants, such as single nucleotide polymorphisms (SNPs), have some unique features in that they are predetermined during gamete formation.
T2DM risk among younger individuals who have not yet developed conventional risk factors. This hypothesis was tested recently among white and black adolescents (58) and young adults (59); both studies found that GRMs did not improve the predictive performance of T2DM compared with assessment of clinical risk factors.

The predictive performance of genetic markers for T2DM risk could be improved in future studies through several approaches. First, simulation studies suggest that the cumulative effect of a large number of common genetic variants could lead to an increased AUC for complex disease prediction (13, 14). According to mathematical modeling by Janssens et al. (60), to increase the AUC of genetic profiling to 0.80 or greater, 400 genetic variants with minor allele frequencies of 10% and odds ratios of the heterozygous genotypes for each variant greater than 1.25 are needed. Currently identified and/or confirmed GWA markers for T2DM are still limited, and the majority of them have a modest association with T2DM (odds ratios of heterozygous genotypes are less than 1.15 for most GWA markers); thus, it is not surprising that the AUC values for T2DM did not vary substantially according to the number of GWA markers (n ≤ 40) in this study.

Although it has been found that known GWA markers explain only a limited proportion of the estimated genetic variation for T2DM, which suggests the existence of "missing" heritability (61), whether hundreds of common genetic variants for T2DM will be identified through GWA studies is unknown. Second, with the application of large-scale exon resequencing and next-generation sequencing technologies (62), rare variants for T2DM are likely to be uncovered (63). An empirical analysis suggested that the inclusion of rare variants might have appreciable effects on disease risk prediction (64). However, whether adding rare variants will improve the predictive performance for T2DM remains to be evaluated. Third, whether the incorporation of additional susceptibility loci discovered through novel liability methods (65) and obesity-predisposing SNPs (66) will improve the predictive performance for T2DM warrants further investigation.

It has been speculated that genetic profiling of GWA markers might motivate people who carry the risk variants to change their dietary and lifestyle habits that lead to T2DM (57); however, emerging evidence, although still limited, does not support this notion (67). A recent observational study showed that genome-wide profiling did not result in any measurable short-term changes in diet or exercise behavior (68), although long-term effects remain unknown. Moreover, a randomized trial also showed that diabetes genetic risk counseling with currently available variants does not significantly alter self-reported motivation or prevention program adherence for overweight individuals at risk for diabetes (67). Although it has been demonstrated that broad, population-based lifestyle interventions are effective at reducing the risk of T2DM in high-risk individuals (4), whether personalized interventions based on individual genetic backgrounds may increase the effectiveness warrants further evaluation (69, 70).

It should be noted that the ideal statistical measure of the incremental predictive performance of novel risk markers has been controversial. Most previous studies about the effect of GWA markers on T2DM risk prediction have focused on the AUC, which has been regarded as a standard measure of the effect of a new marker in risk prediction (71). However, the AUC is relatively insensitive to change if a few risk factors with strong associations with T2DM are already included in the model (20, 71). In addition, statistical issues regarding hypothesis testing of changes in AUCs have been documented (72–74). For example, under the null hypothesis, the DeLong test (75), the widely used nonparametric test for evaluating incremental AUCs in prediction models, has an exceptiona...
<table>
<thead>
<tr>
<th>First Author, Year (Reference No.)</th>
<th>Study Location</th>
<th>Ethnic Origin</th>
<th>Study Type</th>
<th>No. of Variants</th>
<th>Conventional T2DM Risk Factors</th>
<th>Cutoff Points, %a</th>
<th>NRI IDI</th>
<th>%</th>
<th>P Value</th>
<th>%</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyssenko, 2008 (15)</td>
<td>Sweden Finland</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>11</td>
<td>Age, sex, BMI, FH, FPG, BP, TG</td>
<td>10 and 20</td>
<td>4.5</td>
<td>2.5 x 10^{-15}</td>
<td>NA</td>
<td>3.7 x 10^{-14}</td>
<td>NA</td>
</tr>
<tr>
<td>Lin, 2009 (36)</td>
<td>Switzerland</td>
<td>Caucasian</td>
<td>Cross-sectional</td>
<td>15</td>
<td>Age, BMI, FH, WHR, TG/HDL-C</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1.2</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>Schulze, 2009 (37)</td>
<td>Germany</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>20</td>
<td>German Diabetes Risk Score, FPG, HbA1c, TG, HDL-C, GGT, ALT</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.34</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Talmud, 2010 (40)</td>
<td>United Kingdom</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>20</td>
<td>Cambridge Diabetes Risk Score</td>
<td>5, 10, and 15</td>
<td>4.6</td>
<td>0.17</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Talmud, 2010 (40)</td>
<td>United Kingdom</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>20</td>
<td>Framingham Offspring Study T2DM Risk Score</td>
<td>5, 10, and 15</td>
<td>-3.2</td>
<td>0.35</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>de Miguel-Yanes, 2011 (27)</td>
<td>United States</td>
<td>Caucasian</td>
<td>Cohort</td>
<td>40</td>
<td>Age, sex, BMI, FH, FPG, BP, HDL-C, TG</td>
<td>2 and 8</td>
<td>4.3</td>
<td>0.004</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Hivert, 2011 (48)</td>
<td>United States</td>
<td>Mixed</td>
<td>Cohort</td>
<td>34</td>
<td>Age, sex, ethnic background, treatment arm, and waist circumference</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-0.007</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ALT, alanine transaminase; BMI, body mass index; BP, blood pressure; FH, family history of diabetes; FPG, fasting plasma glucose; GGT, gamma-glutamyltransferase; GWA, genome-wide association; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; IDI, integrated discrimination improvement; NA, not applicable or not available; NRI, net reclassification improvement; T2DM, type 2 diabetes mellitus; TG, triglycerides; WHR, waist-hip ratio.

a Predefined cutoff point for reclassification based on the likelihood of developing T2DM.

b Calculated as weight (kg)/height (m)^2.

c The German Diabetes Risk Score includes age, waist circumference, height, history of hypertension, physical activity, smoking, and consumption of red meat, whole-grain bread, coffee, and alcohol.

d The Cambridge Diabetes Risk Score includes age, sex, drug treatment, FH, BMI, and smoking status.

e The Framingham Offspring Study Type 2 Diabetes Risk Score includes age, sex, parental history of T2DM, BMI, HDL-C, TG, and FPG.
to be a promising complement to the AUC; however, a recent study showed that current methods for hypothesis testing with the IDI are invalid (78). Other novel measures, such as the continuous NRI (79) and the net benefit plotted by the “decision curves” (80), have not yet widely been used in practice. Thus, an empirical evaluation of the utility of these novel measures for assessing the incremental predictive performance of GWA markers in T2DM prediction is warranted.

Some limitations should be acknowledged. First, current GRMs are based on common SNPs, which represent only part of the genetic variation in the human genome (81). Whether the addition of rare SNPs and other genetic variants, such as copy number variations, which account for more than 12% of the assembled human genome sequence (82), will improve the predictive performance for risk of T2DM remains unanswered. It should be noted that a recent GWA study showed that common copy number variants that can be genotyped on existing platforms are unlikely to contribute substantially to the genetic basis of T2DM, and most of them are well tagged by SNPs (83). Second, most of the available studies were performed among Caucasians and Asians. Whether these results can be generalized to other ethnic groups warrants further investigation. Third, the role of gene-environment interactions in the prediction of T2DM was not addressed in this review because of limited available data. Both genetic and environmental factors (e.g., diet, lifestyle) and their complex interactions are implicated in the development of T2DM (5, 6), and available evidence suggests that individuals with higher genetic susceptibility of T2DM may benefit more from dietary and lifestyle changes. For instance, in the US Health Professionals Follow-up Study, the positive association between the Western dietary pattern and the risk of T2DM was more pronounced among men with a higher genetic risk score (≥12) than in those with a lower score (84). Moreover, the Diabetes Prevention Program suggested that lifestyle intervention might mitigate the elevated risk of T2DM conferred by variants of the TCF7L2 gene (85). However, a recent study showed that the inclusion of gene-environment interactions was unlikely to dramatically improve risk prediction for several types of complex diseases, including T2DM (86).

In summary, GRMs showed a relatively low predictive performance for T2DM risk regardless of study design, sample size, participants’ race/ethnicity (i.e., Caucasian, Asian), and the number of genetic markers included. Moreover, the addition of GWA markers to CRMs produced a minor improvement in predictive performance. Therefore, although the identification of GWA markers could help improve our understanding of the pathophysiology of T2DM, its clinical utility in improving the prediction of T2DM beyond that of conventional risk factors may be limited. Further investigation of the predictive performance of the genetic factors and their interactions with environmental factors is warranted.

REFERENCES


GENETIC RISK PREDICTION MODELS FOR TYPE 2 DIABETES


