Philippa J. Talmud,[1] Jackie A. Cooper,[1] Richard W. Morris,[2] Frank Dudbridge,[3] Tina Shah,[4] Jorgen Engmann,[4] Caroline Dale,[3] Jon White,[5] Stela McLachlan,[6] Delilah Zabaneh,[5] Andrew Wong,[7] Ken K. Ong,[7,8] Tom Gaunt,[9,10] Michael V. Holmes,[4,11] Debbie A. Lawlor,[9,10] Marcus Richards,[7] Rebecca Hardy,[7] Diana Kuh,[7] Nicholas Wareham,[8] Claudia Langenberg,[4,8] Yoav Ben-Shlomo,[10] S. Goya Wannamethee,[2] Mark W.J. Strachan,[12] Meena Kumari,[4] John C. Whittaker,[13] Fotios Drenos,[1,10] Mika Kivimaki,[4] Aroon D. Hingorani,[4,14] Jacqueline F. Price,[6] and Steve E. Humphries,[1] on behalf of the UCLEB Consortium

Sixty-Five Common Genetic Variants and Prediction of Type 2 Diabetes

Diabetes 2015;64:1830–1840 | DOI: 10.2337/db14-1504

We developed a 65 type 2 diabetes (T2D) variant–weighted gene score to examine the impact on T2D risk assessment in a U.K.-based consortium of prospective studies, with subjects initially free from T2D (*N* = 13,294; 37.3% women; mean age 58.5 [38–99] years). We compared the performance of the gene score with the phenotypically derived Framingham Offspring Study T2D risk model and then the two in combination. Over the median 10 years of follow-up, 804 participants developed T2D. The odds ratio for T2D (top vs. bottom quintiles of gene score) was 2.70 (95% CI 2.12–3.43). With a 10% false-positive rate, the genetic score alone detected 19.9% incident cases, the Framingham risk model 30.7%, and together 37.3%. The respective area under the receiver operator characteristic curves were 0.60 (95% CI 0.58–0.62), 0.75 (95% CI 0.73 to 0.77), and 0.76 (95% CI 0.75 to 0.78). The combined risk score net reclassification improvement (NRI) was 8.1% (5.0 to 11.2; $P = 3.31 \times 10^{-7}$). While BMI stratification into tertiles influenced the NRI (BMI ≤24.5 kg/m^2, 27.6% [95% CI 17.7–37.5], $P = 4.82 \times 10^{-8}$; 24.5–27.5 kg/m^2, 11.6% [95% CI 5.8–17.4], $P = 9.88 \times 10^{-5}$; >27.5 kg/m^2, 2.6% [95% CI −1.4 to 6.6], $P = 0.20$), age categories did not. The addition of the gene score to a phenotypic risk model leads to a potentially clinically important improvement in discrimination of incident T2D.

Type 2 diabetes (T2D) is an important and increasingly prevalent condition with a high morbidity, resulting in a growing cost to health services. Notably, individuals frequently remain asymptomatic until presenting with complications. Age and obesity are the major environmental risk factors for T2D; the latter is driven by the increased intake of processed food and sedentary behaviors, with commensurate raised calorie intake, influenced by a Western-style diet, and is becoming more prevalent in low- and middle-income countries. However, a subset of T2D patients remain lean and are likely to represent a different subtype of the disease with less macrovascular disease, who, with an extended life span, develop microvascular

[1]Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, London, U.K.
[2]Department of Primary Care and Population Health, University College London, Royal Free Campus, London, U.K.
[3]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, U.K.
[4]Department of Epidemiology and Public Health, University College London Institute of Epidemiology and Health Care, University College London, London, U.K.
[5]University College London Genetics Institute, Department of Genetics, Environment and Evolution, London, U.K.
[6]Centre for Population Health Sciences, University of Edinburgh, Edinburgh, U.K.
[7]Medical Research Council Unit for Lifelong Health and Ageing at University College London, London, U.K.
[8]Medical Research Council Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, U.K.
[9]School of Social and Community Medicine, University of Bristol, Bristol, U.K.
[10]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, U.K.
[11]Division of Transplant Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
[12]Metabolic Unit, Western General Hospital, Edinburgh, U.K.
[13]Genetics Division, Research and Development, GlaxoSmithKline, Harlow, U.K.
[14]Centre for Clinical Pharmacology, University College London, London, U.K.

Corresponding author: Philippa J. Talmud, p.talmud@ucl.ac.uk.

Received 9 October 2014 and accepted 27 November 2014.

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db14-1504/-/DC1.

See accompanying article, p. 1495.

complications (1,2). For these reasons, there is interest in developing tools for the prediction of T2D, with one systematic review identifying 84 different risk algorithms with the area under the receiver operating characteristic curves (A_{ROCs}) ranging from 0.62 to 0.90 (3). It was noted that several of these had not been externally validated and no one algorithm performed best (3).

The expectation, in the early phase of the genome-wide association studies (GWAS), was that this approach would lead to the identification of novel genetic risk loci to aid in risk prediction of complex diseases such as T2D. However, the overall variance in disease risk explained by the identified loci remained low, and there is a pervading negativity about the use of genetic information in risk prediction and clinical utility (4).

In 2010, we compared the performance of a genetic risk score based on 20 known T2D risk alleles in combination with the phenotypic-derived Framingham Offspring T2D risk score (FORS) (5) in the prospective Whitehall II study (WHII) of U.K. civil servants (6). The results were not encouraging; a genetic risk score weighted by the effect size of each of the 20 single nucleotide polymorphisms (SNPs) did not improve discrimination, risk estimation, or reclassification of individuals who went on to develop T2D compared with the FORS alone. A recent review of 19 studies, reported prior to 2013, which used between 2 and 40 risk alleles, providing A_{ROCs} ranging from 0.54 to 0.63, concluded that genetic variants did not improve prediction over established phenotypic predictors.

GWAS since 2012 have identified additional T2D susceptibility loci, and meta-analysis of studies using gene-centric chips (7,8) has brought the total number of known T2D risk variants close to 70. Since these in combination explain more of the variation in T2D risk, using the increased number of risk alleles may also improve risk prediction.

The first study to use the expanded risk SNPs examined whether 40 T2D risk SNPs in a weighted risk score could improve the C-statistic, when added to a phenotypic risk score, on incident T2D in 3,471 individuals, of whom 446 developed T2D over 34-year follow-up. Using age stratification above or below 50 years, there was no improvement to the C-statistic, but there was a significant increase in the net reclassification improvement (NRI) in those below 50 years but not in those 50 years old or above (9). Walford et al. (10) went on to use a 62 SNP–weighted gene score (206 incident T2D cases in a total of 1,622 individuals followed for 13.4 years). This larger genetic risk score did provide improvement to the C-statistic of the combined genetic and phenotypic risk scores over either risk score alone, suggesting complementation for metabolic and genetic information. A second study examined the efficacy of a 62-SNP gene score in T2D prediction in the Framingham Offspring Study (3,869 subjects, of whom 446 developed T2D) and the multiethnic Coronary Artery Risk Development in Young Adults (CARDIA) study (total of 1,650 whites with 97 incident T2D cases, and among the 820 blacks, 118 developed T2D). While the gene score was associated with incident T2D risk and an improved C-statistic when combined with phenotypic risk factors, there was no improvement in net reclassification (11). There was also no evidence of an interaction between the genetic risk score and obesity in the determination of T2D susceptibility (11). The much larger EPIC InterAct case-control cohort with >12,000 T2D cases and >16,000 controls reported that a 49-SNP T2D genetic risk score had a greater effect on the development of T2D in younger and leaner individuals (12), but the investigators did not examine the effect of other parameters of risk assessment.

Our aim was to determine whether using the expanded number of 65 T2D genetic variants improves risk prediction and also to explore whether risk prediction, including a genetic risk score, differs by age and BMI. Since heritability for T2D decreases with increasing age (13), we might anticipate that genetic variants would improve prediction in younger individuals. Moreover, addition of genetic variants to T2D risk might vary by BMI depending on whether the prediction tool is enriched for adiposity-related SNPs. We incorporated seven prospective cohort studies with over 13,000 individuals initially free of T2D, 804 of whom developed incident T2D during follow-up, thus providing information on discrimination, risk estimation, and net reclassification in the largest study to date. These studies were part of the University College London-London School of Hygiene and Tropical Medicine-Edinburgh-Bristol (UCLEB) Consortium.

RESEARCH DESIGN AND METHODS

UCLEB Consortium

A full description of the UCLEB Consortium has been previously published (14). For the current analysis, seven cohorts with genotype and complete incident T2D information were included, comprising a total of 13,294 individuals, of whom 804 developed T2D over the period of study.

Briefly, the 12 UCLEB studies are almost exclusively of European ancestry and cover a wide geographic range within the U.K. Population structure was assessed by principal components analysis, and outliers were excluded. All studies have longitudinal follow-up ranging from 5 to 62 years (for a full description, see the Supplementary Data). MetaboChip genotype information was available on 21,474 individuals. For the current analysis, the following cohorts with genotype and complete incident T2D information were included: British Regional Heart Study (BRHS; N = 2,317), British Women's Heart and Health Study (BWHHS; N = 1,854), Edinburgh Artery Study (EAS; N = 703), Medical Research Council National Survey of Health and Development (MRC NSHD; N = 2,410), WHII (N = 3,045), English Longitudinal Study of Aging (ELSA; N = 1,685), and Caerphilly Prospective Study (CAPS; N = 1,280). A total of 13,294 individuals were included. The 1,542 individuals with prevalent T2D were excluded from the analysis, and over the period of study, 804 developed T2D (for full details of individual studies see the Supplementary Data).

Clinical Characteristics of the Participants

All studies have harmonized information on a wide range of risk factor and disease variables in a shared data set, as previously described (14). For the current analysis, data on blood lipids, fasting glucose, age, sex, blood pressure, and BMI were used, as well as data on incident diabetes; additional data on family history of T2D were requested from individual cohorts. Medication data were also collated, including lipid-lowering drugs (statins or other medication), blood pressure–lowering drugs, and glucose-lowering drugs. Classification as "prevalent T2D" was based on self-report, medical record review, use of glucose-lowering medication, and/or a fasting glucose >7 mmol/L. Within individual cohorts, biochemical measurements were performed in accredited laboratories using international standards. DNA was extracted from blood samples either collected at baseline (BWHHS) or at a subsequent resurvey (BRHS, MRC NSHD, EAS, WHII, ELSA, and CAPS).

Definition of T2D

An individual was coded as developing T2D if the condition was self-reported or recorded by medical record review, if a new prescription of a glucose-lowering medication was recorded, or following a recorded fasting glucose of 7 mmol/L or higher (nonfasting in the case of BRHS). For MRC NSHD, glucose levels were estimated based on fasting HbA_{1c} levels (15).

Genotyping

Genotypes of 13,294 individuals for T2D susceptibility variants were obtained using MetaboChip, an array that includes almost 200,000 SNPs, which cover the loci identified by GWAS in cardiometabolic diseases, including rare variants identified by the 1000 Genomes Project (16). Duplicate samples were genotyped to compute the error rate. Quality control analysis on genotyped samples has been previously reported (14), and all included SNPs had a call rate of >98%. All genotypes were in Hardy-Weinberg Equilibrium in all studies.

We used the list of T2D risk SNPs recently identified in a large meta-analysis (8). The SNP, the nearest gene with chromosome number, the minor allele frequency, and the reported effect size are presented in Supplementary Table 1. For each gene, the lead SNP reported by Morris et al. (8) was chosen. Genotype data were available on all 65 identified T2D SNPs, including the adiposity genes *FTO* and *MC4R*, both previously associated with T2D.

Statistical Analysis

Score Construction

We used the published regression coefficients for age, sex, parental history of T2D, BMI, blood pressure, HDL cholesterol, triglyceride, and fasting glucose level to compute the FORS (5). We computed a genetic risk score using the published coefficients (log odds ratios [ORs]) for 65 SNPs identified by prior GWAS meta-analysis and previously reviewed (8,17). Coefficients were multiplied by 0, 1, or 2 according to the number of risk alleles carried by each person, and the score was centered by subtracting the mean. The two scores were added to produce a combined score. In addition to the weighted genetic risk score, we also calculated an unweighted score by summing the number of risk alleles.

Association Testing

Associations of individual SNPs with risk markers were assessed by regression, and a significance level of $P < 0.001$ was used after Bonferroni correction for the number of SNPs analyzed. Logistic regression models were fitted to obtain the OR per SD increase in the gene score as well as OR associated with each quintile. Association models were fitted using the combined data set with a term for study included in the model.

Model Discrimination

We calculated the A_{ROC} and the detection rate, defined as the proportion of all cases detected for a false-positive rate (FPR) of 5 and 10%. A_{ROCs} were calculated separately for each study and combined using fixed effects meta-analysis. Improvements in the receiver operating characteristic area were assessed by calculating the difference between the two receiver operating characteristic areas in each study along with bootstrap estimates of the CI and then combining these over all the studies.

Model Calibration

Estimates of risk were obtained by converting the logit given by the weighted coefficients back to a probability. Observed and estimated risks were converted to 10-year risk, taking the length of follow-up into account. Observed risks were then compared with predicted risks and the Hosmer-Lemeshow test was used to assess goodness of fit.

Reclassification of T2D Risk

We used the NRI that quantifies the extent to which the combined score moved people to risk categories that better reflected their event status (18). As three of the studies were of case-control design, we used a weighted version of the NRI, weighting controls by the inverse of the sampling probability and assigning a weight of 1 to cases (19). We used four 10-year T2D risk categories (≤5, 5–9.9, 10–14.9, and 15% or higher). We also calculated both the continuous NRI, which does not require categories, as changes are defined by any upward or downward change in predicted risks, and the integrated discrimination improvement (IDI) as recommended (18). Analyses were conducted for the entire cohort and then within subgroups stratified by tertiles of age and BMI (<24.5, 24.5–27.5, >27.5 kg/m^2) and in men and women separately.

All analysis was conducted using Stata version 13.1 (StataCorp, TX).

RESULTS

The baseline characteristics and T2D incident rates of the subjects in the individual studies are presented in Table 1.

Of the 13,294 subjects (range of follow-up 4–20 years, median 10 years), 804 (6.1%) developed T2D, but the incidence rate differed among studies, in keeping with variation in mean age and duration of follow-up.

Association and Discrimination Based on the FORS

The OR comparing the top and bottom quintiles of the FORS distribution was 21.07 (95% CI 14.86–29.88), and the OR for a 1 SD increase of the FORS score was 2.70 (95% CI 2.48–2.93; $P = 5.4 \times 10^{-121}$) (Table 2). The ORs for the individual studies are presented in Supplementary Table 2. The A_{ROC} for the entire data set was 0.75 (95% CI 0.74–0.77) (Table 3 and, for the individual studies, Supplementary Fig. 1 and Supplementary Table 3). With a 10% FPR, the Framingham risk model alone identified 30.7% of cases. The corresponding detection rate for a 5% FPR was 18.6%. There was significant heterogeneity between detection rates for the seven studies (Fig. 1*A*). The Forest plot for a 5% FPR for the seven studies is presented in Supplementary Fig. 2*A*. There was no difference when a random-effects model was used (data not shown).

Association and Discrimination Using Genotype-Based Risk Scores

The point estimates for 53 of the 65 SNPs used in this study were consistent with those reported in prior meta-analyses involving many thousands of T2D cases (Supplementary Table 1). After correction for multiple testing, eight of the variants contributing to the T2D genetic risk score were also associated with nongenetic variables included in the FORS algorithm, including BMI and fasting glucose (Supplementary Table 3*A*–*E*).

The distribution of a gene score based on these variants, weighted by the published effect sizes, in participants initially T2D free is shown in Fig. 2. For the unweighted score, see Supplementary Fig. 3. The OR for T2D among individuals in the top versus the bottom quintile of the gene score distribution for the weighted gene score was 2.70 (95% CI 2.12–3.43; $P = 7.03 \times 10^{-16}$). Thirty-one percent of incident T2D individuals were in the top quintile of the weighted gene score compared with 19.3% of T2D-free individuals, and the OR for a 1 SD increase was 1.43 (95% CI 1.33–1.54; $P = 2.25 \times 10^{-22}$). The ORs for the individual studies are presented in Supplementary Table 2 and the A_{ROC}s in Supplementary Table 4. With a 10% FPR, the genetic score alone detected 19.9% of incident cases. The corresponding value for a 5% FPR was 11.8%. The A_{ROC} was 0.60 (95% CI 0.58–0.62).

Effect of Adding Genetic Information to Discrimination Based on the FORS

The OR for the top versus the bottom quintile of the combined FORS and weighted gene score was 22.60 (95% CI 15.80–32.40). The addition of genetic information to the FORS marginally improved discrimination as assessed by the increase in the A_{ROC} from 0.75 to 0.76 (difference 0.012 [95% CI 0.006–0.018]; $P = 0.0003$) (Fig. 3 and Table 4). The disease detection rate for a 5% FPR was increased

Table 1—Total number of individuals, T2D rates, and baseline characteristics of subjects for incident T2D in the seven studies

	BRHS	BWHHS	EAS	MRC NSHD	WHII	ELSA	CAPS	Total
Number included in analysis	2,317	1,854	703	2,410	3,045	1,685	1,280	13,294
Number with incident T2D (%)	150 (6.5)	103 (5.6)	16 (2.3)	118 (4.9)	219 (7.2)	74 (4.4)	124 (9.7)	804 (6.1)
Length of follow-up, years	20	7	12	10	10	4	15.5	11.2
Rate per 1,000 person-years	3.2	7.9	1.9	4.9	7.2	11.0	6.3	5.4
Age, years	49.1 (5.6)	70.8 (5.3)	64.5 (5.6)	53.0 (0)	48.9 (6.0)	73.6 (9.1)	56.7 (4.5)	57.4 (11.2)
Sex, % male (*n*)	100% (2,317)	0% (0)	46.8% (329)	49.9% (1,203)	76.7% (2,336)	51.8% (873)	100% (1,280)	62.7% (8,338)
BMI, kg/m^2	25.4 (2.9)	27.5 (4.8)	25.3 (3.6)	27.3 (4.6)	25.2 (3.5)	27.2 (4.2)	26.6 (3.6)	26.3 (4.1)
HDL, mmol/L	1.15 (0.24)	1.63 (0.44)	1.47 (0.37)	1.68 (0.49)	1.42 (0.40)	1.51 (0.38)	1.03 (0.25)	1.42 (0.44)
Triglyceride, mmol/L*	1.87 (0.81)	1.71 (0.76)	1.35 (0.58)	1.82 (1.01)	1.22 (0.67)	1.56 (0.77)	1.70 (0.85)	1.58 (0.83)
Systolic blood pressure, mmHg	144.6 (20.5)	154.3 (27.2)	142.7 (23.4)	138.0 (21.2)	121.1 (13.9)	146.1 (20.4)	145.9 (22.5)	139.6 (23.6)
Diastolic blood pressure, mmHg	82.2 (13.3)	83.1 (13.0)	77.1 (12.1)	85.8 (13.0)	80.1 (9.7)	77.9 (11.5)	84.5 (11.7)	81.9 (12.3)
Fasting glucose, mmol/L*	5.43 (0.92)	5.79 (0.70)	5.49 (0.56)	5.94† (0.78)	5.19 (0.45)	5.38 (0.77)	5.20 (0.67)	5.48 (0.76)

Data are presented as mean (SD) unless otherwise indicated. Patients with type 1 diabetes or latent autoimmune diabetes of adulthood were excluded. *Geometric mean (approximate SD). †Calculated from nonfasting HbA_{1c}.

Table 2—Quintiles of FORS and externally weighted 65-gene score in the combined studies

Quintile	FORS	Externally weighted gene score	FORS and externally weighted gene score
1	1.00	1.00	1.00
2	2.83 (1.93–4.15)	1.37 (1.05–1.79)	2.62 (1.76–3.92)
3	4.28 (2.97–6.17)	1.36 (1.04–1.78)	4.73 (3.23–6.92)
4	7.76 (5.39–11.16)	2.01 (1.56–2.58)	7.74 (5.32–11.27)
5	21.07 (14.86–29.88)	2.70 (2.12–3.43)	22.59 (15.75–32.41)
Per quintile	2.07 (1.94–2.21); $P = 2.60 \times 10^{-106}$	1.28 (1.21–1.34); $P = 9.03 \times 10^{-20}$	2.12 (1.99–2.27); $P = 7.71 \times 10^{-111}$
Per SD	2.70 (2.48–2.93); $P = 5.40 \times 10^{-121}$	1.43 (1.33–1.54); $P = 2.25 \times 10^{-22}$	2.83 (2.61–3.08); $P = 3.08 \times 10^{-132}$

Data are presented as OR (95% CI).

from 18.6% (95% CI 15.9–21.2) to 23.1% (95% CI 20.2–26.1), and for a 10% FPR, the improvement was even greater from 30.7% (95% CI 27.5–33.8) to 37.3% (95% CI 33.9–40.6) (Table 4 and Supplementary Table 5). Forest plots of the seven studies for these two FPRs are presented in Fig. 1*B* and Supplementary Fig. 2*B*, respectively.

Calibration of the Phenotype-Only and Combined Risk Models

Individuals were assigned into four 10-year T2D risk categories (≤5, 5–9.9, 10–14.9, and 15% or higher) using the FORS only, and then the combined risk scores, and the observed and predicted event rates compared in each category (Supplementary Table 6). The FORS and combined scores both accurately estimated the rates of diabetes in each of the four categories of predicted risk. In addition, we performed the Hosmer-Lemeshow goodness of fit test, which confirmed that there was no significant difference between the observed and predicted risks for either the FORS only ($P = 0.65$) or the combined score ($P = 0.10$) (Supplementary Table 7 and Supplementary Fig. 4).

NRI and IDI

We next tested whether adding the genetic information to the FORS more accurately predicted risk of T2D as assessed by the NRI measure using these absolute risk categories. An individual with incident T2D was considered to be correctly reclassified if they shifted to a higher risk category when the genetic information was added, while a shift to a lower risk score was regarded as incorrect reclassification, with the opposite being the case for participants who remained free of T2D. The addition of the gene score to the FORS resulted in an NRI of 8.1% (95% CI 5.0–11.2; $P = 3.3 \times 10^{-7}$) (Table 4*A*). The addition of the gene score had less effect on the reclassification of those who remained T2D free. This is illustrated in Fig. 4*A* and *B*, plotting the phenotypic score against the combined phenotypic and genetic scores, using as an example the 15% risk category. For most individuals, there is a strong correlation between FORS and the FORS plus gene score (green). In Fig. 4*A*, for those who remained free of T2D, more individuals moved up a risk category (blue) than down a risk category (red) when the gene score was added to the phenotypic risk score resulting in a negative net reclassification. In those who did develop T2D (Fig. 4*B*), addition of the gene score to FORS also resulted in more individuals moving up a risk category (blue), than down (red), leading to a positive net reclassification. The percentage of individuals with T2D moving up was substantially more than the percentage of nondiabetic individuals moving up, resulting in a significant positive NRI. For the continuous NRI (18), which is independent of cut points, the improvement was 29.7% (23.7–35.7; $P = 2.04 \times 10^{-23}$) (Supplementary Fig. 6), composed of an event NRI of 13% and a nonevent NRI of 17%. This 30% improvement corresponds to a small to medium Cohen effect size of 0.37 (18). The discrimination slope improved from 0.076 to 0.089, giving an IDI of 0.013 (0.009–0.017; $P = 6.15 \times 10^{-11}$), and for IDI, the improvement was 0.013 (0.009–0.017; $P = 6.15 \times 10^{-11}$) (Supplementary Fig. 7).

Table 3—A_{ROC} (95% CI) and the false-positive detection rates for the combined data

	OR (95% CI) top vs. bottom quintile	A_{ROC} for combined studies (95% CI)	Detection rate for 5% false positive (95% CI)	Detection rate for 10% false positive (95% CI)
Externally weighted gene score	2.70 (2.12–3.43)	0.60 (0.58–0.62)	11.75% (9.54–13.95)	19.89% (17.14–22.63)
FORS	21.07 (14.86–29.88)	0.75 (0.73–0.77)	18.56% (15.9–21.22)	30.67% (27.51–33.82)
FORS + externally weighted gene score	22.59 (15.75–32.41)	0.76 (0.75–0.78)*	23.14% (20.23–26.05)	37.28% (33.95–40.61)

*$P = 0.003$. P value derived from the comparison with FORS alone.

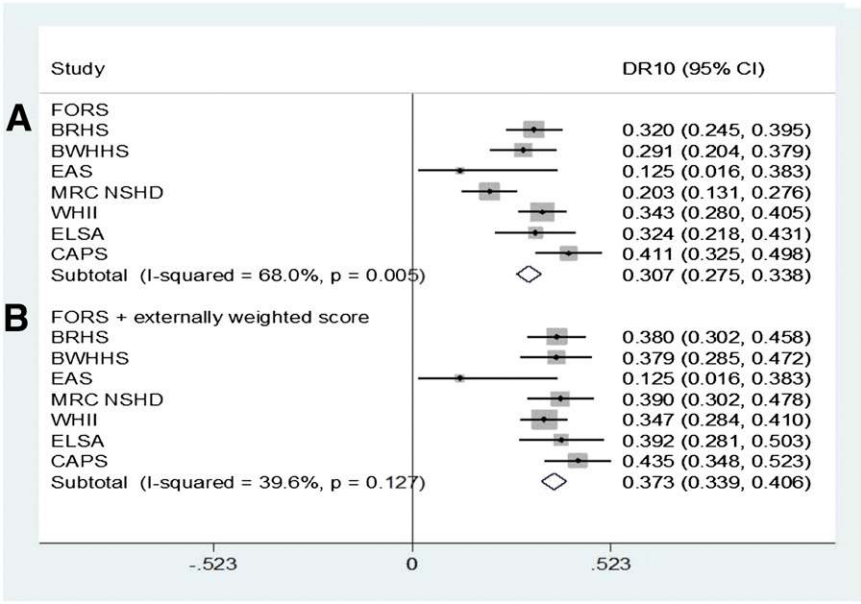

Figure 1—Forest plot showing the 10% detection rate for all seven studies for (*A*) the Framingham phenotypic score alone and (*B*) Framingham T2D score plus the externally weighted gene score. DR10, 10% detection rate.

Stratification by BMI, Age, and Sex

To examine the influence of BMI and age on the ability of the genetic score to improve incident case discrimination, we conducted a prespecified analysis by tertiles of BMI and age, and in men and women separately. For BMI tertiles (<24.5, 24.5–27.5, and >27.5 kg/m^2), the calibration of the combined score is presented in Supplementary Fig. 8. Risk tended to be overestimated for those in the bottom tertile (P = 0.0009) and underestimated for those in the middle and top tertiles (P = 0.02 and $P = 9.76 \times 10^{-7}$, respectively). The NRI is presented in Table 4*B–D*. In those who had a BMI <24.5 kg/m^2, the NRI was 27.6% (95% CI 17.7–37.5; $P = 4.82 \times 10^{-8}$). For those in the middle tertile (24.5–27.5 kg/m^2), the NRI was still statistically significant (11.6% [95% CI 5.8–17.4]; $P = 9.88 \times 10^{-5}$). By contrast, those with a BMI above 27.5 kg/m^2 had an NRI of 2.6% (95% CI −1.4 to 6.6; P = 0.20). Adding the gene score to the FORS after stratification improved the A_{ROC} by 0.037

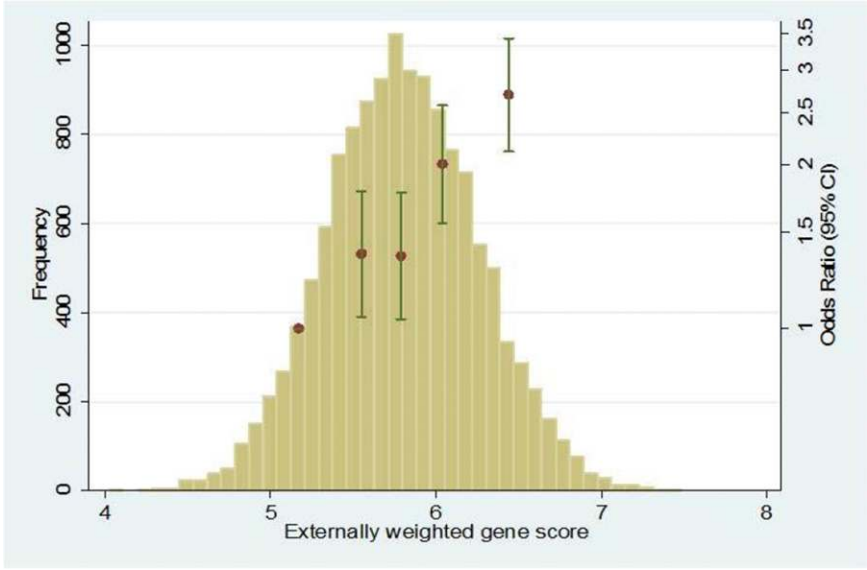

Figure 2—The distribution of the 65-SNP gene score, weighted by the external published β-values in the combined studies. Superimposed are the log ORs for T2D.

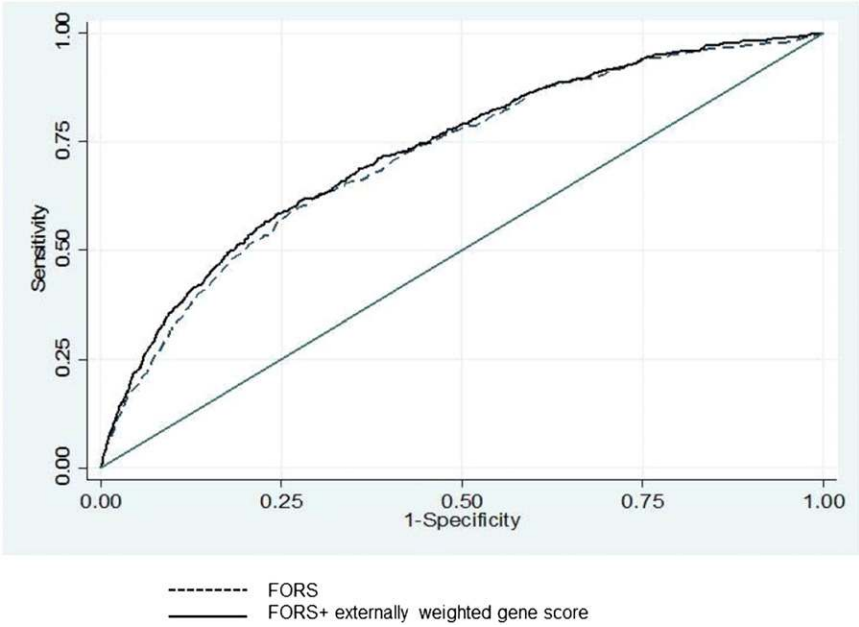

Figure 3—The receiver operating curve for the seven combined studies.

(95% CI 0.018–0.056; P = 0.0001) for those in the lowest tertile of weight but not for those in the middle tertile (0.017 [0.004–0.031]; P = 0.01) or the top tertile (0.013 [95% CI 0.002–0.023]; P = 0.02; P = 0.03 for difference between BMI categories) (Supplementary Table 8). Little difference was seen in model performance for different age categories (see Supplementary Fig. 9 and Supplementary Table 9) or by sex (Supplementary Fig. 10 and Supplementary Table 10). However, while the NRI was significant in men, it reached borderline statistical significance in the women, which might reflect the reduced power in the women, since there were far fewer women in the study (men, 554 with T2D, 7,784 T2D free; women, 250 with T2D, 4,706 T2D free) (Supplementary Table 11).

DISCUSSION

In this study, we have examined the predictive ability of 65 T2D risk variants in predicting T2D risk in the largest number of incident T2D patients reported to date. This included pooled information from multiple prospective cohort studies with over 13,000 individuals with relevant phenotypic data prior to the onset of disease. We used a range of metrics in our analysis: discrimination as assessed by the A_{ROC}, the strength of the association as determined by the OR and quintiles of risk, risk detection at 5% and 10% FPRs, and reclassification based on the NRI index. We also examined whether adding genotypic scores to an established phenotypic risk prediction tool changes prediction differently within prespecified subgroups. This was true for BMI but, contrary to expectation, not for age stratification.

In the prospective analysis of 804 incident cases compared with 12,490 T2D free, the FORS performed reasonably well, with an A_{ROC} of 0.75 for the seven combined studies, compared with an A_{ROC} of 0.85 in the Framingham Offspring Study itself (5), providing external validation for the algorithm. The overall OR for developing T2D was 21.1 in the top quintile versus the bottom quintile for score. The addition of a 65-SNP gene score to the FORS improved the correct classification of individuals with T2D into higher risk categories by 6.2%. When the gene score was added to the FORS score, there was a small, but significant, improvement in the A_{ROC}. For individuals with a BMI of 25 kg/m^2 or below, this improvement was even greater. Examining this BMI effect using repeat BMI measures may be as good, if not better, than a genetic score that is fixed in time. Examination of the BMI changes over time and replication in a larger cohort is required to validate these results.

For a gene score to be effective, it should improve the reclassification of individuals with T2D into a more accurate risk category over and above the phenotypic risk score. The 65 SNP–weighted gene score did this. In actual terms, at a 10% FPR, the combined phenotypic and genetic risk score led to the correct identification of an additional 53 (6.6%) of the 804 cases. We examined whether genetics would play a bigger role in T2D risk in the absence of the environmental challenge of obesity. With stratification by BMI, individuals in the lowest tertile of BMI (<24.5 kg/m^2) had an NRI of 27.6% compared with those in the top tertile, with an NRI of 2.6%, confirming our hypothesis.

We previously examined the WHII (6) (n = 5,535) and reported that the FORS (5) performed better than the Cambridge T2D score (20), which incorporates only the

Table 4—NRI based on addition of gene score to FORS, calculated using risk cutoffs of 5, 10, and 15% for 10-year risk

	Number of people				Reclassified		Net correctly reclassified
Predicted risk FORS	≤5	5–9.9	10–14.9	≥15	Increased risk	Decreased risk	
A. For the whole cohort							
Plus externally weighted gene score: no diabetes (n = 18,715.81)					1,782.23	2,138.63	1.9% (1.2–32.6)
<5	10,406.62	582.00	36.28	0			
5–9.9	1,064.52	1,967.29	542.56	118.24			
10–14.9	35.16	647.23	682.39	503.15			
≥15	6.55	78.89	306.28	1,738.65			
Plus externally weighted gene score: incident diabetes (n = 1,121.86)					185.96	116.49	6.2% (3.2–9.2)
<5	279.01	34.13	1	0			
5–9.9	36.80	81.20	55.97	15.82			
10–14.9	0	45.93	71.24	79.04			
≥15	0	11.73	22.03	387.96			
B. BMI tertile 1 (BMI <24.5)							
Plus externally weighted gene score: no diabetes (n = 6,267.82)					448.8	616.52	2.7% (1.7–3.7)
<5	4,202.14	95.95	2	0			
5–9.9	540.06	834.1	232.65	63.3			
10–14.9	1	35.27	36.08	54.9			
≥15	0	7.13	33.06	130.18			
Plus externally weighted gene score: incident diabetes (n = 147.53)					45.88	9.13	24.9% (15.0–34.8)
<5	56	3	0	0			
5–9.9	6.13	21.06	26.32	7.82			
10–14.9	0	3	2.21	8.74			
≥15	0	0	0	13.25			
C. BMI tertile 2 (BMI 24.5–27.4)							
Plus externally weighted gene score: no diabetes (n = 6,526.66)					706.02	715.15	0.1% (−1.0 to 1.3)
<5	3,684.61	267.53	22.95	0			
5–9.9	339.91	675.31	185.82	34.24			
10–14.9	20.83	253.93	266.22	195.48			
≥15	6.55	22.56	71.37	479.35			
Plus externally weighted gene score: incident diabetes (n = 308.31)					58.26	22.93	11.5% (5.7–17.2)
<5	112.17	17.13	0	0			
5–9.9	7.13	23.63	12.25	6			
10–14.9	0	11.66	18.64	22.88			
≥15	0	1	3.14	72.68			
D. BMI tertile 3 (BMI ≥27.5)							
Plus externally weighted gene score: no diabetes (n = 5,921.35)					627.42	806.96	3.0% (1.8–4.3)
<5	2,519.87	218.52	11.33	0			
5–9.9	184.55	457.88	124.09	20.71			
10–14.9	13.33	358.03	380.1	252.77			
≥15	0	49.2	201.85	1,129.12			
Plus externally weighted gene score: incident diabetes (n = 666.02)					81.82	84.43	−0.4% (−4.2 to 3.4)
<5	110.83	14	1	0			
5–9.9	23.55	36.52	17.4	2			
10–14.9	0	31.26	50.39	47.42			
≥15	0	10.73	18.89	302.03			

A. Values are weighted to take into account sampling design, thus accounting for the fact that the number of individuals is not an integer. P value for heterogeneity = 0.002. I^2 = 71.1%. NRI (95% CI) = 8.1% (5.0–11.2), no adjustment for study; $P = 3.31 \times 10^{-7}$. NRI (95% CI) = 6.6% (3.6–9.7), results from meta-analysis of individual study results (fixed effects); $P = 2.0 \times 10^{-5}$. NRI (95% CI) = 7.7% (1.7–13.8), results from meta-analysis of individual study results (random effects); P = 0.01. *B*. NRI (95% CI) = 27.6% (17.7–37.5); $P = 4.82 \times 10^{-8}$. *C*. NRI (95% CI) = 11.6% (5.8–17.4); $P = 9.88 \times 10^{-5}$. *D*. Values are weighted to take into account sampling design, thus accounting for the fact that the number of individuals is not an integer. NRI (95% CI) = 2.6% (−1.4 to 6.6); P = 0.20.

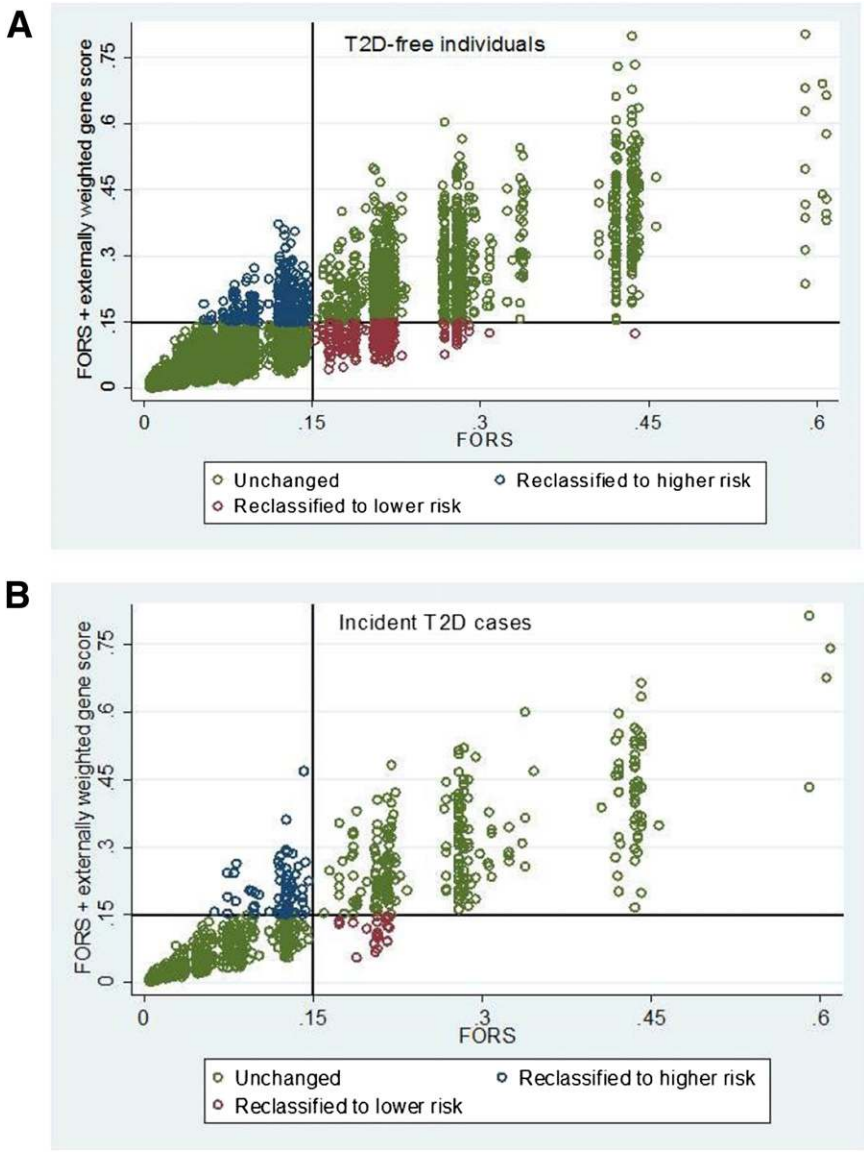

Figure 4—Scatter plot showing the correlation between Framingham T2D score and Framingham T2D score plus the externally weighted gene score. *A*: Those who remained T2D free. *B*: Incident T2D cases.

routinely assessed variables, and that a 20-SNP gene score did not improve prediction when added to the FORS, with the A_{ROC} remaining at 0.78. When repeating analysis on all seven studies reported here, discrimination was not significantly improved (difference in A_{ROC} 0.005 [95% CI 0.003–0.013]; P = 0.23), while the NRI for those same 20 SNPs was 5.9% (95% CI 2.3–9.5) (see Supplementary Table 12), which did not differ significantly from the 8.1% NRI found for the 65 SNPs (P = 0.36).

The 20 SNPs were primarily in genes encoding proteins involved in pancreatic β-cell function (6). The extended 65 variants in the present analysis, however, involve additional pathways, including adipocytokine signaling, cell cycle regulation, and CREBBP-related transcription (8), thus widening the implicated processes leading to T2D. Simulation analysis of GWAS identified additional SNPs associated with T2D susceptibility (8), but with decreasing effect sizes, their impact on risk prediction is likely to be very small. This explains why with ~60, T2D gene score prediction has reached a plateau (21), based on the 62-SNP gene score analysis (11). Of interest, in the search for rare T2D variants of large effect, while exome sequencing has failed to identify these in the case-control setting (22) within a T2D family, exome sequencing has identified a rare cSNP in the gene encoding early endosome antigen 1 (*EEA1*) (23).

Recent assessment of risk scores pooled across studies have highlighted the potential pitfalls, in particular, when assessing the incremental value of adding novel predictors to established predictors (24). These include variations in

distributions of the novel predictors and the variation in impact of these new predictors between studies. Gene distributions are unlikely to vary between the studies presented here, all of which are almost exclusively of white European participants, and there is little reason to suspect their impact would vary between studies. Concerning the overall NRI statistic, the inverse weighting approach we have used to pool across studies may give more weight to studies with few events and bias results toward the null. Accordingly, we have presented event NRI and nonevent NRI both for analysis of all participants and for subgroups.

There are, however, several limitations to our study. Any single genetic risk marker is limited by effect size, overcome to some extent by using them in combination in a gene score. Better diagnosis of the subtypes of T2D, e.g., in lean individuals, is likely to make risk prediction more precise. To confirm the generalizability of our findings, replication in an independent set of cohort studies is needed. One major problem in developing a clinically useful SNP gene score is the underlying genetic architecture of T2D. While individuals carrying many risk alleles are at a much higher risk of T2D than those carrying fewer alleles, they represent only a small proportion of the population. The consequence of this is that individuals with an intermediate number of risk alleles will account for the majority of cases of T2D because of the large number of people at intermediate risk in the population (see Fig. 2). Because of this, there is a substantial overlap of the distribution of risk alleles among individuals who develop diabetes and those who remained disease free; thus it is difficult to set a cut point of a gene score that reliably discriminates T2D cases. Fifty-three out of the 65 SNPs used in this study had effect sizes in concordance with those published in meta-analysis (17). We used these external weights from published meta-analyses of >100,000 subjects (8) to minimize the sampling errors and to avoid overfitting the genetic risk.

The previous reports using risk scores of 40–62 SNPs, from de Miguel-Yanes et al. (9), Walford et al. (10), and Vassy et al. (11), have to be considered in the context that they all performed their analysis in the setting of the same study, the Framingham Offspring Study, and furthermore, the phenotypic risk score was derived from the same Framingham Offspring Study. When applied to the Framingham Offspring Study itself, the phenotypic risk score provides a better C-statistic or equivalent A_{ROC} (0.85) (5) than in the combined studies presented here (0.75). This is not surprising, as the phenotypic risk score always performs better in the study in which it originated. Our study findings, although confirmatory, take this analysis forward in that we examined the predictive impact of 65 SNPs in a cohort of almost double the number of incident cases and in a data set that was independent of the Framingham Offspring Study, thus validating the phenotypic risk score.

In conclusion, an increase in the number of genetic risk variants for T2D to 65 risk SNPs slightly improved discrimination and classification of individuals with the disease into a higher risk category, thus demonstrating incremental value for prediction. Although these results require further independent validation and any suggestion of including genetic variants in risk prediction tools would need to be assessed for clinical and cost-effectiveness in randomized controlled trials, our findings suggest that there is potential for common variants of small effect in combination to aid in risk prediction for T2D. Unlike statins used prophylactically in coronary disease heart prevention, metformin is not used in the same way to prevent T2D. Although it is hoped that those with a high T2D genetic risk might be especially motivated to make lifestyle changes, this has not always proved to be so (25). Genetic variants need to be measured only once for each person, and most primary health care practitioners in high-income countries make use of electronic records. Thus taking into account the genetic component (if this is recorded, once obtained) should be feasible.

Acknowledgments. The data were collected by the National Centre for Social Research. Research clinics were held at the Wellcome Trust Clinical Research Facility and Princess Alexandra Eye Pavilion in Edinburgh. DNA standardization was conducted at the Genetics Core of the Wellcome Trust Clinical Research Facility in Edinburgh. The data archive is maintained by the University of Bristol. ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London, and the Institute of Fiscal Studies.

Funding. The UCLEB Consortium is supported by a British Heart Foundation program grant (RG/10/12/28456). The Northwick Park Heart Study II (NPHS-II) is supported by the MRC and by the National Institutes of Health (grant NHLBI 33014). BRHS is a British Heart Foundation Research Group and is supported by the British Heart Foundation (RG/08/013/25942). The WHII study is supported by grants from the MRC (K013351, ID85374); the British Heart Foundation (RG/07/008/23674); the Stroke Association; the National Heart, Lung, and Blood Institute (HL-036310); the National Institute on Aging (5R01-AG-13196); Agency for Health Care Policy Research (HS-06516), and the John D. and Catherine T. MacArthur Foundation Research Networks on Successful Midlife Development and Socio-economic Status and Health. Samples from the ELSA DNA Repository received support under a grant (AG1764406S1) awarded by the National Institute on Aging. MRC NSHD is funded by the U.K. MRC (MC_UU_12019/1). BWHHS is supported by funding from the British Heart Foundation and the Department of Health Policy Research Program (England). EAS is funded by the British Heart Foundation (program grant RG/98002), with MetaboChip genotyping funded by a project grant from the Chief Scientist Office of Scotland (project grant CZB/4/672). CAPS was funded by the MRC and undertaken by the former MRC Epidemiology Unit (South Wales). The DNA bank was established with funding from an MRC project grant. S.E.H. holds a chair funded by the British Heart Foundation. P.J.T., M.Ku., A.D.H., and S.E.H. were supported by the British Heart Foundation (grant numbers PG07/133/24260, BHFPG08/008).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Duality of Interest. NPHS-II is supported by Du Pont Pharma. J.C.W. is an employee of GlaxoSmithKline. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. P.J.T., F.Du., and S.E.H. interpreted the data and wrote the manuscript, read the manuscript, and contributed to the final version of the manuscript. J.A.C. and F.Dr. interpreted the data, wrote the manuscript, contributed to the data analysis, read the manuscript, and contributed to the final

version of the manuscript. R.W.M., M.Ki., A.D.H., and J.F.P. interpreted the data, wrote the manuscript, provided study samples for the analysis, read the manuscript, and contributed to the final version of the manuscript. T.S., M.V.H., R.H., S.G.W., M.W.J.S., J.C.W., and all coauthors read the manuscript and contributed to the final version of the manuscript. J.E., C.D., J.W., S.M., D.Z., and K.K.O. contributed to the data analysis, read the manuscript, and contributed to the final version of the manuscript. A.W., T.G., D.A.L., M.R., D.K., N.W., C.L., Y.B.-S., and M.Ku. provided study samples for the analysis, read the manuscript, and contributed to the final version of the manuscript. P.J.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

1. Færch K, Witte DR, Tabák AG, et al. Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: a post-hoc analysis of the longitudinal Whitehall II cohort study. Lancet Diabetes Endocrinol 2013;1:43–51
2. Sinharoy K, Mandal L, Chakrabarti S, Paul UK, Bandyopadhyay R, Basu AK. A study on clinical and biochemical profile of low body weight type 2 diabetes mellitus. J Indian Med Assoc 2008;106:747–750
3. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ 2011;343:d7163
4. Lyssenko V, Laakso M. Genetic screening for the risk of type 2 diabetes: worthless or valuable? Diabetes Care 2013;36(Suppl. 2):S120–S126
5. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007;167:1068–1074
6. Talmud PJ, Hingorani AD, Cooper JA, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ 2010;340:b4838
7. Saxena R, Elbers CC, Guo Y, et al.; Look AHEAD Research Group; DIAGRAM consortium. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci [published correction appears in Am J Hum Genet 2012;90:753]. Am J Hum Genet 2012;90:410–425
8. Morris AP, Voight BF, Teslovich TM, et al.; Wellcome Trust Case Control Consortium; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; DIAbetes 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
9. de Miguel-Yanes JM, Shrader P, Pencina MJ, et al.; MAGIC Investigators; DIAGRAM+ Investigators. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. Diabetes Care 2011;34:121–125
10. Walford GA, Porneala BC, Dauriz M, et al. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. Diabetes Care 2014;37:2508–2514
11. Vassy JL, Hivert MF, Porneala B, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. Diabetes 2014;63:2172–2182
12. Langenberg C, Sharp SJ, Franks PW, et al. Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. PLoS Med 2014;11:e1001647
13. Almgren P, Lehtovirta M, Isomaa B, et al.; Botnia Study Group. Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia 2011;54:2811–2819
14. Shah T, Engmann J, Dale C, et al.; UCLEB Consortium. Population genomics of cardiometabolic traits: design of the University College London-London School of Hygiene and Tropical Medicine-Edinburgh-Bristol (UCLEB) Consortium [published correction appears in PLoS One 2014;8]. PLoS ONE 2013;8:e71345
15. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ; A1c-Derived Average Glucose Study Group. Translating the A1C assay into estimated average glucose values. Diabetes Care 2008;31:1473–1478
16. Voight BF, Kang HM, Ding J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet 2012;8:e1002793
17. Sanghera DK, Blackett PR. Type 2 Diabetes Genetics: Beyond GWAS. J Diabetes Metab 2012;3:6948–6963
18. Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med 2012;31:101–113
19. Ganna A, Reilly M, de Faire U, Pedersen N, Magnusson P, Ingelsson E. Risk prediction measures for case-cohort and nested case-control designs: an application to cardiovascular disease. Am J Epidemiol 2012;175:715–724
20. Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study. Fam Pract 2008;25:191–196
21. Hivert MF, Vassy JL, Meigs JB. Susceptibility to type 2 diabetes mellitus–from genes to prevention. Nat Rev Endocrinol 2014;10:198–205
22. Lohmueller KE, Sparsø T, Li Q, et al. Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet 2013;93:1072–1086
23. Tanaka D, Nagashima K, Sasaki M, et al. Exome sequencing identifies a new candidate mutation for susceptibility to diabetes in a family with highly aggregated type 2 diabetes. Mol Genet Metab 2013;109:112–117
24. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM; Emerging Risk Factors Collaboration. Assessing risk prediction models using individual participant data from multiple studies. Am J Epidemiol 2014;179:621–632
25. Grant RW, O'Brien KE, Waxler JL, et al. Personalized genetic risk counseling to motivate diabetes prevention: a randomized trial. Diabetes Care 2013;36:13–19