Title: Polygenic type 2 diabetes prediction at the limit of common variant detection

Running title: T2D polygenic prediction

Jason L. Vassy[1,2,3], Marie-France Hivert[1,4,5], Bianca Porneala[6], Marco Dauriz[1,6,7], Jose C. Florez[1,8,9], Josée Dupuis[10,11], David S. Siscovick[12], Myriam Fornage[13], , Laura J. Rasmussen-Torvik[14], Claude Bouchard[15], James B. Meigs[1,6]*

*Corresponding author: James B. Meigs, MD, MPH, General Medicine Division, Massachusetts General Hospital, 50 Staniford Street, 9th floor, Boston, MA 02104; tel. 617-724-3203; fax 617-724-3455; e-mail: jmeigs@partners.org

1. Harvard Medical School, Boston, MA; 2. Section of General Internal Medicine, VA Boston Healthcare System, Boston, MA; 3. Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA; 4. Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA; 5. Division of Endocrinology, Department of Medicine, Université de Sherbrooke, Sherbrooke, QC; 6. General Medicine Division, Massachusetts General Hospital, Boston, MA; 7. Division of Endocrinology and Metabolic Diseases, Department of Medicine, University of Verona Medical School and Hospital Trust of Verona, Verona, Italy; 8. Diabetes Research Center (Diabetes Unit), and Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 9. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 10. Department of Biostatistics, Boston University School of Public Health; 11. National Heart, Lung, and Blood Institute's Framingham Heart Study; 12. Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, WA; 13. Center for Human Genetics, University of Texas Health Science Center at Houston, Houston, TX; 14. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 15. Human Genomics Laboratory, Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, LA

Word count: 4257
Tables: 5
Figures: 2

Diabetes Publish Ahead of Print, published online February 11, 2014

Abstract

Genome-wide association studies (GWAS) may have reached their limit of detecting common type 2 diabetes (T2D)-associated genetic variation. We evaluated the performance of current polygenic T2D prediction. Using data from the Framingham Offspring (FOS) and the Coronary Artery Risk in Young Adults (CARDIA) studies, we tested three hypotheses: 1) a 62-locus genotype risk score (GRS_t) improves T2D prediction compared to previous less inclusive GRS_t; 2) separate β-cell and insulin resistance GRS (GRS_β and GRS_{IR}) independently predict T2D; and 3) the relationships between T2D and GRS_t, GRS_β, or GRS_{IR} do not differ between blacks and whites. Among 1650 young white adults in CARDIA, 820 young black adults in CARDIA, and 3,471 white middle-aged adults in FOS, cumulative T2D incidence was 5.9%, 14.4%, and 12.9%, respectively, over 25 years. The 62-locus GRS_t was significantly associated with incident T2D in all three groups. In FOS but not CARDIA, the 62-locus GRS_t improved the model C statistic (0.698 and 0.726 for models without and with GRS_t, respectively, $p<0.001$); it did not materially improve risk reclassification in either study. Results were similar among blacks compared with whites. The GRS_β, but not GRS_{IR}, predicted incident T2D among FOS and CARDIA whites. At the end of the era of common variant discovery for T2D, polygenic scores can predict T2D in whites and blacks but do not outperform clinical models. Further optimization of polygenic prediction may require novel analytic methods including less common as well as functional variants.

Introduction

Type 2 diabetes (T2D) is a common complex disease with both genetic and environmental determinants. Risk factors including overnutrition, sedentary behavior, and lack of physical exercise, make the disease amenable to prevention through lifestyle modification(1; 2), but the most effective behavior change programs can be cost-intensive(3). As the genome-wide association study (GWAS) era has discovered dozens of genetic loci associated with T2D risk, there has been hope that genotype might help clinicians and public health practitioners target limited prevention resources to those at greatest risk. Although genotype predicts incident T2D(4-9), studies using limited genetic information from the first waves of GWAS have demonstrated that the addition of genotype to T2D prediction models based upon routinely measured clinical risk factors(6; 10; 11) does not substantively improve risk stratification(4; 8; 9).

The DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium recently published the largest T2D GWAS meta-analysis to date (DIAGRAMv3), identifying many additional common variants associated with T2D and bringing the total number of independent T2D loci to 65(12). Together, these loci explained about 5.7% of the variance in genetic susceptibility to T2D. DIAGRAMv3 also modeled the theoretical existence of 488 additional common variants likely associated with T2D on the arrays used in their analyses but with effect sizes too small for detection. These hundreds of single-nucleotide polymorphisms (SNPs) would increase the proportion of explained T2D susceptibility to 10.7%. Subsequent models using genome-wide complex trait analysis suggested that 63% of T2D susceptibility might be

attributable to common genetic variation in the full set of GWAS SNPs(12). Still, current GWAS methodology is likely nearing its limit(13; 14) to identify the additional specific common SNPs associated with T2D. Recent analyses have suggested that even a tripling of the GWAS discovery sample size would not materially increase the C statistic of polygenic T2D models(15). Ongoing next-generation sequencing efforts may identify additional variants with major allele frequency >1%, although SNP genotype and imputation data from GWAS arrays have likely already captured most of this common variation.

Thus, the 65 DIAGRAMv3 loci may represent the majority of common and significant T2D-association genetic variants expected to be identified. If so, it is opportune to evaluate the performance of currently available genetic information for T2D risk prediction and classification. The additional loci discovered in DIAGRAMv3 may improve polygenic T2D prediction over previous attempts using polygenic models with fewer loci(4; 5; 9; 16). Because GWAS use a cross-sectional case-control design, it is important to determine how well these loci prospectively predict incident T2D. Moreover, polygenic models may be improved by taking into consideration the biological pathways underlying these T2D-associated loci. Though most of these remain to be elucidated, some functional studies and analyses of more specific metabolic phenotypes have implicated some loci in pancreatic β-cell dysfunction or, less commonly, insulin resistance(17; 18). Individuals carrying a high genetic burden for both β-cell dysfunction and insulin resistance might be at especially high risk of developing T2D. Finally, although DIAGRAMv3 used data from populations of mostly European ancestry, it is important for clinical practice and public health to know whether these associations hold in non-white populations.

Research design and methods

We used data from the Framingham Offspring (FOS) and the Coronary Artery Risk Development in Young Adults (CARDIA) studies to examine the performance of updated polygenic prediction models for T2D among young and middle-aged adults of European and African ancestry. We tested three primary hypotheses. First, we hypothesized that an updated total genotype risk score (GRS_t) with up to 65 T2D-associated risk loci improves the prediction of incident T2D in young and middle-aged adulthood, compared to previously published scores with fewer loci. We examined both genotype-only and genotype-plus-clinical prediction models. Second, because β-cell dysfunction and insulin resistance represent two distinct pathways in the pathogenesis of T2D, we hypothesized that separate GRS comprised of SNPs postulated to influence β-cell or insulin resistance (GRS_β and GRS_{IR}) independently predict incident T2D. In subsidiary analyses, we investigated whether GRS_β and GRS_{IR} together exhibit a multiplicative effect on T2D risk and whether the association between T2D risk and GRS_β or GRS_{IR} varies between lean and obese individuals. Third, we hypothesized that the relationships between incident T2D and GRS_t, GRS_β, or GRS_{IR} do not differ between black and white individuals.

Study participants

Both FOS and CARDIA are large well-described prospective cohort studies(19-21). The FOS began in 1971 and consists of offspring of the original Framingham Heart Study participants and their spouses. At the first examination, FOS participants were between 5 and 70 years of age. They were examined again after eight years and then every four years thereafter through

examination 8 (2005-2008). The CARDIA Study is a multicenter prospective study of 5,115 white and black participants recruited in 1985-1986 from four United States cities(20; 21). Participants were aged 18 to 30 years at the baseline examination and have been invited to participate in serial follow-up examinations over the subsequent 25 years. Written informed consent was obtained from all FOS and CARDIA participants, and the institutional review board at each participating center approved the original studies. We limited the present analyses to FOS and CARDIA participants with at least two study examinations, genotype information, and baseline data available for all predictors of interest. We excluded any participant with diabetes or pregnancy at the baseline examination. CARDIA participants who reported diabetes treatment exclusively with insulin during the observation period were considered to have type 1 diabetes and were also excluded from analyses. We did not apply this exclusion to the older FOS cohort; greater than 99% of the FOS diabetes cases are type 2(11). The Partners Human Research Committee approved these analyses.

Type 2 diabetes

The primary outcome was incident T2D during the observation period. Each FOS examination included an assessment of medical history, a physical examination, and a fasting blood sample(22). All CARDIA study visits included an updated medical history assessment, including medications, and fasting glucose was measured at Years 0, 7, 10, 15, 20, and 25. We defined T2D in FOS and CARDIA by a fasting plasma glucose ≥7.0 mmol/L (≥126 mg/dL) or report of taking diabetes medications(9; 10).

Clinical risk factors and covariates

Data collection methods in FOS and CARDIA have been described previously(19; 21). We considered a study participant to have a positive parental history of diabetes if he/she reported on a family history questionnaire that one or both parents had diabetes(23). Fasting plasma glucose and lipid levels were measured as described previously(22; 24). All FOS participants were white, and in CARDIA race was determined by self-report (black or white).

Genotyping and genotype risk scores

Details of the genotyping and quality of FOS and CARDIA samples have been published previously(25-27). In previous reports, we calculated GRS_t consisting of all the T2D-associated loci known at the time: 17- and 40-SNP GRS_t in FOS and a 38-SNP GRS_t in CARDIA(4; 9; 16). In the present analyses, we updated these GRS_t to include as many of the 65 index SNPs or their proxies as were available at the confirmed or newly identified loci from DIAGRAMv3(12) (Table 1 and Figure 1), using previously reported methods(4; 9; 16). For each locus for each individual, we prioritized inclusion of the following information into the GRS_t, in order: genotyped data at the index SNP, imputed data at the index SNP, and then genotyped data at a suitable proxy for the index SNP. We used SNAP (http://www.broadinstitute.org/mpg/snap/) to identify proxy SNPs, as needed, defined as being in linkage equilibrium with the index SNP ($r^2 \geq 0.5$) in the HapMap II release 22 CEU reference population. Of the 65 loci, genotyped or imputed data were available for 62 of the index SNPs for the FOS and CARDIA studies. No genotype information was available for rs11063069 at *CCND2*, rs11651052 at *HNF1B (TCF2)*, or rs8108269 at *GIPR*. Whites and blacks in CARDIA had genotyped or imputed data for these same 62 loci. For FOS and CARDIA whites, we calculated GRS_t as the weighted sum of the number of risk alleles (0, 1, or 2) at each of the available loci, weighted by its effect size (beta)

from DIAGRAMv3. Because no sufficiently large T2D GWAS in people of African ancestry exists from which to derive locus effect sizes, we used an unweighted GRS_t for CARDIA blacks, calculated by summing the risk alleles across the loci.

Additionally, we used prior genetic and physiologic evidence to categorize the loci as associated predominantly with β-cell function or insulin resistance (Supplementary Table 1). We identified 20 predominantly β-cell associated SNPs by 1) their significant effect on HOMA-β (beta<-0.008; p<0.05) in the most recent Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC)(12) and/or 2) a significant effect (p<0.05) on one of the β-cell function indices(18): insulinogenic index or acute insulin response. We identified 10 predominantly insulin resistance-related SNPs by 1) their significant association with HOMA-IR (p<0.05) in the MAGIC data(12), 2) significant association with fasting insulin in the MAGIC GWAS conditional on BMI or BMI-SNP interaction(28), and/or 3) evidence of association with insulin resistance-related traits such as lower high-density lipoprotein (HDL) cholesterol, higher triglycerides, higher BMI, and higher waist-to-hip ratio(18). Similar to the GRS_t, we calculated separate β-cell (GRS_β) and insulin resistance (GRS_{IR}) genotype risk scores, with each locus weighted in whites by the same effect size as in the GRS_t. For CARDIA blacks, we calculated unweighted GRS_β and GRS_{IR}.

Statistical analysis

We constructed logistic and proportional-hazards regression models for incident T2D using similar statistical methods as in our previous FOS and CARDIA analyses, respectively (Supplementary methods)(4; 9; 16). In each study, we constructed regression models for

incident T2D as a function of GRS, sex, and age (demographic model) and GRS, sex, age, and risk factors routinely measured in clinical practice (clinical model: parental history of diabetes (yes vs. no), BMI, systolic blood pressure, fasting plasma glucose, and log-transformed HDL cholesterol and triglyceride levels). We used C statistics and continuous net reclassification improvement (NRI) indices to compare prediction models with and without genotype information (29-32). To examine the relationship between β-cell and insulin resistance genotype, we also performed the models above with 1) GRS_{β} alone, 2) GRS_{IR} alone, 3) GRS_{β} and GRS_{IR}, and 4) GRS_{β}, GRS_{IR}, and a GRS_{β} x GRS_{IR} interaction term. Further, we examined the relationship between genotype and BMI in two ways: 1) the inclusion of an interaction term between each GRS and an indicator variable for obesity (BMI $\geq$ 30 kg/m^2 vs. BMI < 30 kg/m^2) and 2) analyses stratified by BMI category (BMI $\geq$ 30 kg/m^2 vs. BMI < 30 kg/m^2). To test the hypothesis that the association between each GRS and T2D risk does not differ between whites and blacks, we meta-analyzed the regression beta coefficients from FOS and CARDIA whites and then used a *t*-test to compare the result to the corresponding beta in CARDIA blacks. We considered odds ratios and hazard ratios as statistically significant at $p<0.05$.

Results

Participant characteristics and incident type 2 diabetes

Among the 3,869 FOS participants, 11,358 person-periods from 3,471 individuals were eligible for the present analyses. In CARDIA, 1650 white and 820 black individuals with 50,309 total person-years of follow-up were eligible. Table 2 shows the baseline participant characteristics. In FOS, there were 446 incident cases of T2D (cumulative incidence 12.9%)

over a mean 25.6 years of follow-up. In the younger CARDIA cohort, among whites there were 97 T2D cases (cumulative incidence 5.9%) over a mean follow-up of 24.2 years, and among blacks there were 118 cases (cumulative incidence 14.4%) over a mean follow-up of 23.4 years.

GRS_t and prediction of incident type 2 diabetes

The mean 62-SNP GRS_t was greater among T2D cases than non-cases in FOS (p<0.001), CARDIA whites (p<0.001), and CARDIA blacks (p=0.01; Table 3). Among all three cohorts, each GRS_t was significantly associated with incident T2D in both the demographic and clinical prediction models (Tables 4 and 5). In the demographic models in FOS, each additional weighted allele in the 17-, 40-, and 62-SNP GRS_t was associated with an increased odds for incident T2D of 11% (7-15%), 8% (6-11%), and 8% (6-10%), respectively. Among CARDIA whites, each additional weighted allele in the 38- and 62-SNP GRS_t was associated with an increase in the adjusted hazard for incident T2D of 12% (6-18%) and 7% (3-12%); the corresponding increases among CARDIA blacks were 5% (0-11%) and 5% (1-9%). The addition of each successive SNP to the GRS_t lowered the per-allele odds ratio for incident T2D in FOS (Figure 1). The addition of the 62-SNP GRS_t to the demographic and clinical prediction models in FOS weakly improved risk reclassification [continuous NRI 0.286 (0.192, 0.380) and 0.256 (0.162, 0.351), respectively] (Table 4). Reclassification was moderate among FOS individuals younger than 50 years and weak among those 50 years or older (Table 4). Reclassification was not markedly higher in the younger CARDIA cohort. Among CARDIA whites, the addition of the 62-SNP GRS_t to the demographic and clinical models resulted in a continuous NRI of 0.311 (0.088, 0.525) and 0.306 (0.073, 0.517), respectively. Similarly, the resulting NRI among CARDIA blacks were 0.243 (0.031, 0.455) and 0.296 (0.098, 0.513),

respectively. Compared to our previous GRS_t consisting of fewer loci, the 62-SNP GRS_t increased model C statistics but did not increase the NRI in FOS (Table 4); NRI in CARDIA whites and blacks were generally higher than with the 38-SNP GRS_t but still indicated weak reclassification improvement (Table 5). The effect size of the 62-SNP GRS_t did not differ between whites (meta-analyzed between FOS and CARDIA) and CARDIA blacks in either the demographic or clinical model (all $p>0.05$) (Supplementary Table 6). The demographic models with the 17-, 40-, and 62-SNP GRS_t explained only 2.0%, 2.1%, and 2.2% of the variance in T2D risk in FOS. In CARDIA whites, the 38- and 62-SNP GRS_t explained 1.7% and 1.5% of risk variance, respectively, and in CARDIA blacks they explained 1.5% and 1.6%, respectively. Figure 2 shows the C statistics the demographic and clinical models with and without the 62-SNP GRS_t.

GRS_β and GRS_{IR} and type 2 diabetes prediction

Among FOS and CARDIA whites, those with incident T2D had a higher mean GRS_β ($p<0.05$ for both cohorts), but not GRS_{IR}, compared with non-cases. In contrast, CARDIA blacks with incident T2D had a higher mean GRS_{IR} ($p=0.03$), but not GRS_β, than non-cases (Supplementary Table 2). Among whites in FOS and CARDIA, GRS_β was associated with incident T2D in the demographic and clinical models (Supplementary Tables 3 and 4). The GRS_β was not associated with T2D among CARDIA blacks, although the between-race difference in effect size was not statistically significant (Supplementary Tables 5 and 6). The GRS_{IR} was associated with T2D among whites after meta-analysis of the FOS and CARDIA results in the demographic model only. It was not associated with T2D among CARDIA blacks, although this effect did not

statistically differ from that in whites (Supplementary Tables 3-6). We found no evidence of a multiplicative interaction between GRS_{β} and GRS_{IR} for T2D risk (all $p>0.05$).

BMI stratification

In BMI-stratified models in both FOS and CARDIA, GRS_{β} was associated with incident T2D among both non-obese and obese subgroups (Supplementary Tables 7 and 8). In contrast, GRS_{IR} was not significantly associated with T2D in either subgroup in either study. In models adjusted for age, sex, and (for CARDIA) race, there were no statistically significant interactions between obesity and GRS_t, GRS_{β}, or GRS_{IR} (Supplementary Tables 9-10). The effect sizes of GRS_{β} were 1.14 (1.09, 1.19) and 1.10 (1.05, 1.15) in lean and obese individuals in FOS, respectively, and 1.08 (1.04, 1.11) and 1.10 (1.06, 1.14) in the lean and obese in CARDIA, respectively.

Discussion

In clinical medicine and public health, there is great interest in identifying individuals and population subgroups at increased T2D risk before disease onset. Genotype has a certain appeal as a risk predictor, as germline genetic code is fixed from birth. The largest T2D GWAS meta-analysis to date(12) may include all of the common T2D-associated loci of at least modest effect size that can be expected to be specifically identified. If so, it marks an appropriate time to evaluate the contribution of known common genetic variation to such risk stratification. Using data from two large well-characterized prospective cohort studies, we have shown that a polygenic score, GRS_t, consisting of 62 of the known T2D-associated loci, is significantly associated with incident T2D over 25 years of observation.

First, we hypothesized that the inclusion of a greater number of T2D-associated loci in the GRS_t would improve T2D prediction, compared to less inclusive GRS_t and to a clinical prediction model. Our prior analyses in FOS and CARDIA demonstrated that GRS_t consisting of up to 40 loci do predict incident T2D from young and middle adulthood but do not improve upon clinical models, as measured by C statistics and NRI indices(4; 9; 16). An updated risk score might improve prediction for at least two reasons. First, a greater number of loci should explain a larger proportion of the heritability of T2D. Second, we updated the weight we used for each locus in our GRS_t based on the effect sizes from the largest T2D GWAS meta-analysis to date(12). For each locus discovered in previous smaller GWAS, the larger sample size of the DIAGRAMv3 discovery set should reduce the error around its effect size on T2D risk(33). The greater precision of these weights might improve the ability of the composite GRS_t to distinguish future T2D cases from non-cases. In the present analyses, we found that the addition of a greater number of loci to the GRS_t steadily improved the C-statistic of the simple demographic prediction model in FOS but not in CARDIA. These polygenic models, using only data available from birth (sex, genotype, and age), achieved C statistics of 0.6-0.7, comparable to other non-genetic T2D prediction models(5-7). However, the inclusion of multiple clinical risk factors to the prediction models overwhelmed any additional improvement in discrimination from genotype information, even though all GRS_t remained significantly associated with incident T2D after adjustment for these factors. Moreover, we did not find evidence that additional SNPs improved risk reclassification over the less inclusive GRS_t. Indeed, among FOS participants, the updated 62-SNP GRS_t lowered the NRI in the demographic and clinical models compared to a 40-SNP GRS_t, although it did perform better than the 17-SNP GRS_t. An exception to this

observation occurred among black young adults in CARDIA. Compared to our previous 38-SNP GRS_t, the 62-SNP GRS_t increased the NRI from 0.083 to 0.243 in the demographic model and from 0.164 to 0.296 in the clinical model. Nonetheless, the magnitudes of these NRI still indicate weak reclassification improvement. Moreover, the relatively small number of cases among CARDIA blacks likely makes these NRI estimates more susceptible to imprecision.

Compared to demographic and clinical prediction models without genotype information, the addition of the 62-SNP GRS_t resulted in relatively small risk reclassification in most of the subgroups examined. Prediction models use risk factors to assign each individual a probability of having the event of interest: here, incident T2D. The continuous NRI measures one model's ability to improve upon the risk classification predicted by another model. Compared to non-genetic models, the addition of a 62-SNP GRS_t generally achieved NRI indices of 0.1 to 0.3, indicative of weak reclassification improvement. The exception was among FOS participants younger than 50 years old at baseline, among whom the 62-SNP GRS_t achieved moderate reclassification improvement (NRI 0.376 compared to the clinical model). Reclassification was much weaker among older FOS participants. This observation suggests that, when added to routine clinical risk factors, genotype information may have greater predictive utility among younger age groups, in whom risk factors such as obesity and impaired fasting glucose might not yet be fully manifest, compared to among older adults. However, we did not observe that the addition of a GRS_t to prediction models among even younger adults in CARDIA resulted in similar reclassification improvement. Because T2D-associated loci included in the GRS_t were discovered in cohorts of largely middle-aged and older adults, they may exert their greatest effect on T2D risk in those decades of life. These loci may only improve T2D prediction among

younger adults when the prediction time horizon is extended beyond the 25 years of follow-up available in the CARDIA Study.

Our second hypothesis was that separate β-cell and insulin resistance polygenic scores independently predict incident T2D. The earliest discoveries among common T2D-associated genetic variants pointed towards genes involved in β-cell function. With the DIAGRAMv3 publication and examination in MAGIC of more refined phenotypes among individuals without diabetes, there are now about ten loci possibly implicated in insulin action as well(18). We also hypothesized that GRS_{β} might have a stronger effect in leaner individuals than in obese individuals. In 2010, the DIAGRAM investigators reported that 23 of 30 T2D-loci investigated showed greater effect sizes among individuals with BMI≤30 kg/m^2 compared to those with BMI>30 kg/m^2, although this difference was statistically significant only for *TCF7L2* and *BCL11A*(34). BMI-stratified GWAS analyses by Perry replicated different sets of previously identified T2D associations among the lean to the obese and identified a novel association with T2D at *LAMA1* only among lean individuals. A polygenic score of 36 known T2D loci had a stronger association with T2D among the lean compared to the obese(35). On the other hand, genetic variants associated with fasting insulin were more easily detected in MAGIC data when BMI was included in the models, and the effect sizes were generally larger in individuals with higher BMI(28). Given this heterogeneous genetic architecture of T2D and related traits, we examined whether the association between T2D risk and GRS_{β} and GRS_{IR} might differ by obesity status. Among whites in FOS and CARDIA, GRS_{β} and GRS_{IR} were associated with incident T2D. Neither score met statistical significance among CARDIA blacks, but the between-race differences were not statistically significant. In contrast to the cross-sectional

analyses by Perry that examined subgroups with BMI < 25 kg/m^2 and BMI ≥ 30 kg/m^2, we found no evidence that GRS_t has a different effect size on incident T2D among individuals with BMI < 30 kg/m^2 compared to those with BMI ≥ 30 kg/m^2. This difference may be due to the lower power from the smaller sample sizes of our analyses, the larger number of loci used in our GRS_t, or our use of prospective data instead of the case-control design used by Perry.

The third aim of our analyses was to examine whether polygenic prediction of T2D differs between individuals of self-reported white and black race. The DIAGRAMv3 meta-analysis consisted predominantly of populations of European ancestry(12). Genome-wide analyses in African populations have been limited by smaller sample sizes(25; 36). First efforts have replicated the association between *TCF7L2* and T2D in populations of African ancestry(36) but have otherwise been largely unrevealing as to the genetic architecture in this group. Examinations of the association between individual European-derived loci and T2D among African populations have inconsistently replicated only a small fraction of these(37; 38), but polygenic scores consisting of these same European-derived loci are nonetheless associated with T2D among African-Americans(8; 9; 38). The biracial composition of the CARDIA Study allowed us to compare the association of the 62-SNP GRS_t with T2D between the two subgroups. The GRS_t was significantly associated with incident T2D among both blacks and whites in the demographic and clinical models, and the effect sizes of the GRS_t, GRS_β, and GRS_{IR} did not differ between the two racial groups. We observed this consistency of effect despite the higher BMI among CARDIA blacks compared to whites (17.3% vs. 6.6% with baseline obesity) and their higher cumulative incidence of T2D (14.4% vs. 5.9%). Most individual European-derived SNPs are only proxies for the true causal variants driving the

associations between given loci and T2D, and differences in linkage disequilibrium between ancestral groups likely magnify this imprecision when examining the relationship between these SNPs and T2D in populations in which they were not originally discovered. While this imprecision may explain why individual European-derived SNPs may not replicate in populations of African ancestry, it remains unclear why a composite polygenic score consisting of these imprecise markers would significantly predict T2D in these same populations. It is likely that the same loci, if not the specific SNPs themselves, are implicated in T2D across ancestral groups(39), and our unweighted GRS_t in CARDIA blacks essentially represents a count of these loci.

Some key lines of inquiry may overcome the limitations of the present analyses and move the field of polygenic risk prediction forward. Polygenic scores such as ours are simple weighted counts of T2D risk alleles across the genome. Such scores significantly predict incident T2D in a number of studies(40). However, other methods of combining genetic risk markers, which do not assume the independence of loci or the additivity of their effects, may improve the performance of prediction models(41; 42). Improved polygenic models may also need to account for epistatic genetic effects and the interactions between loci and environmental factors such as diet and physical activity, although some analyses have suggested that the incremental predictive value of such models may be limited(43). Our examination of the differential effects of β-cell and insulin resistance polygenic scores on T2D risk is a first attempt to account for potential differences at a physiologic level, but more complex molecular pathways may need to be considered. The use of sequencing to identify the causal variants at each T2D-associated locus, for which most of the SNPs included in our GRS are imperfect proxies, should also further

improve the predictive ability of polygenic models(33). In the meantime, except perhaps in younger subgroups, polygenic prediction of T2D using most of the common genetic variation expected to be found in the GWAS era has modest clinical value.

Author contributions

J.L.V, M.-F.H., B.P., M.D., J.C.F, J.D. and J.B.M. conceived the analyses. J.L.V., M.-F.H., B.P. and J.D. performed the analyses. J.L.V, M.-F.H., B.P., M.D., J.C.F, J.D., D.S., M.F., L.J.R.-T., C.B. and J.B.M. analyzed the results. J.L.V, M.-F.H., and J.B.M. wrote the manuscript J.L.V, M.-F.H., B.P., M.D., J.C.F, J.D., D.S., M.F., L.J.R.-T., C.B. and J.B.M. reviewed the manuscript. J.B.M. 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.

Acknowledgements

This work was supported by R01 DK078616 (JBM), K24 DK080140 (JBM), U01 HG006500 (JLV), and L30 DK089597 (JLV). The CARDIA Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging. This manuscript has been reviewed by CARDIA for scientific content. The Framingham Heart Study was supported by the NHLBI (contract number N01-HC-25195) and its contract with Affymetrix, Inc., for genotyping services (contract number

N02-HL-6-4278). A portion of this research was conducted using the Linux Clusters for Genetic Analysis (LinGA) computing resources at the Boston University Medical Campus. The authors declare that they have no conflicts of interest related to this work.

References

1. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM, Diabetes Prevention Program Research G: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. The New England Journal of Medicine 2002;346:393-403
2. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, Keinanen-Kiukaanniemi S, Laakso M, Louheranta A, Rastas M, Salminen V, Uusitupa M, Finnish Diabetes Prevention Study G: Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. The New England Journal of Medicine 2001;344:1343-1350
3. The Diabetes Prevention Program Research Group: The 10-Year Cost-Effectiveness of Lifestyle Intervention or Metformin for Diabetes Prevention: An intent-to-treat analysis of the DPP/DPPOS. Diabetes Care 2012;35:723-730
4. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB, Sr., Cupples LA: Genotype score in addition to common risk factors for prediction of type 2 diabetes. New England Journal of Medicine 2008;359:2208-2219
5. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P, Groop L: Clinical risk factors, DNA variants, and the development of type 2 diabetes. The New England Journal of Medicine 2008;359:2220-2232
6. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, J. BE, M. K, Kivimaki M, Humphries SE: Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ 2010;340:b4838-b4838
7. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC: Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes 2008;57:3122-3128
8. Vassy JL, Mahapatra PD, Meigs JB, Chen W, Schork NJ, Magnussen CG, Raitakari OT, Jamal SM, Berenson GS, Goodman E: Genotype predicts type 2 diabetes in adulthood in a multiracial adolescent population. Pediatrics 2012;130:e1235-1242
9. Vassy JL, Durant NH, Kabagambe EK, Carnethon MR, Rasmussen-Torvik LJ, Fornage M, Lewis CE, Siscovick DS, Meigs JB: A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. Diabetologia 2012;55:2604-2612
10. Wilson PWF, 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. Archives of Internal Medicine 2007;167:1068-1074
11. Meigs JB, Cupples LA, Wilson PW: Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes 2000;49:2201-2207
12. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT,

Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI: Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature Genetics 2012;44:981-990

13. Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, Barrett JC, Blackburn H, Brand O, Burren O, Capon F, Compston A, Gough SC, Jostins L, Kong Y, Lee JC, Lek M, MacArthur DG, Mansfield JC, Mathew CG, Mein CA, Mirza M, Nutland S, Onengut-Gumuscu S, Papouli E, Parkes M, Rich SS, Sawcer S, Satsangi J, Simmonds MJ, Trembath RC, Walker NM, Wozniak E, Todd JA, Simpson MA, Plagnol V, van Heel DA: Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature 2013;498:232-235

14. Morrison AC, Voorman A, Johnson AD, Liu X, Yu J, Li A, Muzny D, Yu F, Rice K, Zhu C, Bis J, Heiss G, O'Donnell CJ, Psaty BM, Cupples LA, Gibbs R, Boerwinkle E: Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. Nature Genetics 2013;45:899-901

15. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH: Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nature Genetics 2013;45:400-405, 405e401-403

16. de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, Dupuis Je, Florez JC, D'Agostino RB, Cupples LA, Meigs JB, the MI, the DI: 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

17. Ingelsson E, Langenberg C, Hivert MF, Prokopenko I, Lyssenko V, Dupuis J, Magi R, Sharp S, Jackson AU, Assimes TL, Shrader P, Knowles JW, Zethelius B, Abbasi FA, Bergman RN, Bergmann A, Berne C, Boehnke M, Bonnycastle LL, Bornstein SR, Buchanan TA, Bumpstead SJ, Bottcher Y, Chines P, Collins FS, Cooper CC, Dennison EM, Erdos MR, Ferrannini E, Fox CS, Graessler J, Hao K, Isomaa B, Jameson KA, Kovacs P, Kuusisto J, Laakso M, Ladenvall C, Mohlke KL, Morken MA, Narisu N, Nathan DM, Pascoe L, Payne F, Petrie JR, Sayer AA, Schwarz P, Scott LJ, Stringham HM, Stumvoll M, Swift AJ, Syvanen AC, Tuomi T, Tuomilehto J, Tonjes A, Valle TT, Williams GH, Lind L, Barroso I, Quertermous T, Walker M, Wareham NJ, Meigs JB, McCarthy MI, Groop L, Watanabe RM, Florez JC, on behalf of the Mi: Detailed

physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. Diabetes 2010;59:1266-1275
18. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, Benazzo A, Rybin D, Jackson AU, Stringham HM, Song C, Fischer-Rosinsky A, Boesgaard TW, Grarup N, Abbasi FA, Assimes TL, Hao K, Yang X, Lecoeur C, Barroso I, Bonnycastle LL, Böttcher Y, Bumpstead S, Chines PS, Erdos MR, Graessler J, Kovacs P, Morken MA, Narisu N, Payne F, Stancakova A, Swift AJ, Tönjes A, Bornstein SR, Cauchi S, Froguel P, Meyre D, Schwarz PE, Häring H-U, Smith U, Boehnke M, Bergman RN, Collins FS, Mohlke KL, Tuomilehto J, Quertemous T, Lind L, Hansen T, Pedersen O, Walker M, Pfeiffer AF, Spranger J, Stumvoll M, Meigs JB, Wareham NJ, Kuusisto J, Laakso M, Langenberg C, Dupuis J, Watanabe RM, Florez JC, Ingelsson E, McCarthy MI, Prokopenko I, investigators obotM: Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 2013;epub ahead of print.
19. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP: An investigation of coronary heart disease in families. The Framingham offspring study. American Journal of Epidemiology 1979;110:281-290
20. Cutter GR, Burke GL, Dyer AR, Friedman GD, Hilner JE, Hughes GH, Hulley SB, Jacobs DR, Jr., Liu K, Manolio TA, et al.: Cardiovascular risk factors in young adults: the CARDIA baseline monograph. Control Clin Trials 1991;12:1S-77S
21. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs Jr DR, Liu K, Savage PJ: CARIDA: study design, recruitment, and some characteristics of the examined subjects. Journal of Clinical Epidemiology 1988;41:1105-1116
22. Meigs JB, Nathan DM, Wilson PW, Cupples LA, Singer DE: Metabolic risk factors worsen continuously across the spectrum of nondiabetic glucose tolerance. The Framingham Offspring Study. Annals of Internal Medicine 1998;128:524-533
23. Murabito JM, Nam BH, D'Agostino RB, Sr., Lloyd-Jones DM, O'Donnell CJ, Wilson PW: Accuracy of offspring reports of parental cardiovascular disease history: the Framingham Offspring Study. Annals of Internal Medicine 2004;140:434-440
24. Bild DE, Jacobs Jr DR, Liu K, Williams OD, Hilner JE, Perkins LL, Marcovina SM, Hulley SB: Seven-year trends in plasma low-density-lipoprotein-cholesterol in young adults: the CARDIA Study. Annals of Epidemiology 1996;6:235-245
25. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, Bennett F, Bowden DW, Chakravarti A, Dreisbach A, Farlow DN, Folsom AR, Fornage M, Forrester T, Fox E, Haiman CA, Hartiala J, Harris TB, Hazen SL, Heckbert SR, Henderson BE, Hirschhorn JN, Keating BJ, Kritchevsky SB, Larkin E, Li M, Rudock ME, McKenzie CA, Meigs JB, Meng YA, Mosley TH, Newman AB, Newton-Cheh CH, Paltoo DN, Papanicolaou GJ, Patterson N, Post WS, Psaty BM, Qasim AN, Qu L, Rader DJ, Redline S, Reilly MP, Reiner AP, Rich SS, Rotter JI, Liu Y, Shrader P, Siscovick DS, Tang WH, Taylor HA, Tracy RP, Vasan RS, Waters KM, Wilks R, Wilson JG, Fabsitz RR, Gabriel SB, Kathiresan S, Boerwinkle E: Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. PLoS genetics 2011;7:e1001300
26. Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, Nettleton JA, King IB, Weng LC, Bhattacharya S, Bandinelli S, Bis JC, Rich SS, Jacobs DR, Jr., Cherubini A, McKnight B, Liang S, Gu X, Rice K, Laurie CC, Lumley T, Browning BL, Psaty BM, Chen YD, Friedlander Y, Djousse L, Wu JH, Siscovick DS, Uitterlinden AG, Arnett DK, Ferrucci L, Fornage M, Tsai MY, Mozaffarian D, Steffen LM: Genetic loci associated with plasma

phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. PLoS genetics 2011;7:e1002193
27. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D: The structure of haplotype blocks in the human genome. Science 2002;296:2225-2229
28. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu CT, Bielak LF, Prokopenko I, Amin N, Barnes D, Cadby G, Hottenga JJ, Ingelsson E, Jackson AU, Johnson T, Kanoni S, Ladenvall C, Lagou V, Lahti J, Lecoeur C, Liu Y, Martinez-Larrad MT, Montasser ME, Navarro P, Perry JR, Rasmussen-Torvik LJ, Salo P, Sattar N, Shungin D, Strawbridge RJ, Tanaka T, van Duijn CM, An P, de Andrade M, Andrews JS, Aspelund T, Atalay M, Aulchenko Y, Balkau B, Bandinelli S, Beckmann JS, Beilby JP, Bellis C, Bergman RN, Blangero J, Boban M, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Borecki IB, Bottcher Y, Bouchard C, Brunner E, Budimir D, Campbell H, Carlson O, Chines PS, Clarke R, Collins FS, Corbaton-Anchuelo A, Couper D, de Faire U, Dedoussis GV, Deloukas P, Dimitriou M, Egan JM, Eiriksdottir G, Erdos MR, Eriksson JG, Eury E, Ferrucci L, Ford I, Forouhi NG, Fox CS, Franzosi MG, Franks PW, Frayling TM, Froguel P, Galan P, de Geus E, Gigante B, Glazer NL, Goel A, Groop L, Gudnason V, Hallmans G, Hamsten A, Hansson O, Harris TB, Hayward C, Heath S, Hercberg S, Hicks AA, Hingorani A, Hofman A, Hui J, Hung J, Jarvelin MR, Jhun MA, Johnson PC, Jukema JW, Jula A, Kao WH, Kaprio J, Kardia SL, Keinanen-Kiukaanniemi S, Kivimaki M, Kolcic I, Kovacs P, Kumari M, Kuusisto J, Kyvik KO, Laakso M, Lakka T, Lannfelt L, Lathrop GM, Launer LJ, Leander K, Li G, Lind L, Lindstrom J, Lobbens S, Loos RJ, Luan J, Lyssenko V, Magi R, Magnusson PK, Marmot M, Meneton P, Mohlke KL, Mooser V, Morken MA, Miljkovic I, Narisu N, O'Connell J, Ong KK, Oostra BA, Palmer LJ, Palotie A, Pankow JS, Peden JF, Pedersen NL, Pehlic M, Peltonen L, Penninx B, Pericic M, Perola M, Perusse L, Peyser PA, Polasek O, Pramstaller PP, Province MA, Raikkonen K, Rauramaa R, Rehnberg E, Rice K, Rotter JI, Rudan I, Ruokonen A, Saaristo T, Sabater-Lleal M, Salomaa V, Savage DB, Saxena R, Schwarz P, Seedorf U, Sennblad B, Serrano-Rios M, Shuldiner AR, Sijbrands EJ, Siscovick DS, Smit JH, Small KS, Smith NL, Smith AV, Stancakova A, Stirrups K, Stumvoll M, Sun YV, Swift AJ, Tonjes A, Tuomilehto J, Trompet S, Uitterlinden AG, Uusitupa M, Vikstrom M, Vitart V, Vohl MC, Voight BF, Vollenweider P, Waeber G, Waterworth DM, Watkins H, Wheeler E, Widen E, Wild SH, Willems SM, Willemsen G, Wilson JF, Witteman JC, Wright AF, Yaghootkar H, Zelenika D, Zemunik T, Zgaga L, Wareham NJ, McCarthy MI, Barroso I, Watanabe RM, Florez JC, Dupuis J, Meigs JB, Langenberg C: A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature Genetics 2012;44:659-669
29. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW: Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30:11-21
30. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS: Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-172
31. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845

32. Pencina MJ, D'Agostino RB: Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med 2004;23:2109-2123
33. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM: Pitfalls of predicting complex traits from SNPs. Nature Reviews Genetics 2013;14:507-515
34. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Bostrom K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI, investigators M, Consortium G: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nature Genetics 2010;42:579-589
35. Perry JR, Voight BF, Yengo L, Amin N, Dupuis J, Ganser M, Grallert H, Navarro P, Li M, Qi L, Steinthorsdottir V, Scott RA, Almgren P, Arking DE, Aulchenko Y, Balkau B, Benediktsson R, Bergman RN, Boerwinkle E, Bonnycastle L, Burtt NP, Campbell H, Charpentier G, Collins FS, Gieger C, Green T, Hadjadj S, Hattersley AT, Herder C, Hofman A, Johnson AD, Kottgen A, Kraft P, Labrune Y, Langenberg C, Manning AK, Mohlke KL, Morris AP, Oostra B, Pankow J, Petersen AK, Pramstaller PP, Prokopenko I, Rathmann W, Rayner W, Roden M, Rudan I, Rybin D, Scott LJ, Sigurdsson G, Sladek R, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Uitterlinden AG, Vivequin S, Weedon MN, Wright AF, Hu FB, Illig T, Kao L, Meigs JB, Wilson JF, Stefansson K, van Duijn C, Altschuler D, Morris AD, Boehnke M, McCarthy MI, Froguel P, Palmer CN, Wareham NJ, Groop L, Frayling TM, Cauchi S: Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. PLoS genetics 2012;8:e1002741
36. Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, An SS, Hester JM, Cooke JN, Bostrom MA, Rudock ME, Talbert ME, Lewis JP, Ferrara A, Lu L, Ziegler JT, Sale MM, Divers J, Shriner D, Adeyemo A, Rotimi CN, Ng MCY, Langefeld CD, Freedman BI, Bowden DW, Consortium D, Investigators M: A genome-wide association search for type 2 diabetes genes in African Americans. PloS one 2012;7:e29202

37. Lewis JP, Palmer ND, Hicks PJ, Sale MM, Langefeld CD, Freedman BI, Divers J, Bowden DW: Association analysis in African Americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. Diabetes 2008;57:2220-2225
38. Cooke JN, Ng MC, Palmer ND, An SS, Hester JM, Freedman BI, Langefeld CD, Bowden DW: Genetic risk assessment of type 2 diabetes-associated polymorphisms in African Americans. Diabetes Care 2012;35:287-292
39. Liu CT, Ng MC, Rybin D, Adeyemo A, Bielinski SJ, Boerwinkle E, Borecki I, Cade B, Chen YD, Djousse L, Fornage M, Goodarzi MO, Grant SF, Guo X, Harris T, Kabagambe E, Kizer JR, Liu Y, Lunetta KL, Mukamal K, Nettleton JA, Pankow JS, Patel SR, Ramos E, Rasmussen-Torvik L, Rich SS, Rotimi CN, Sarpong D, Shriner D, Sims M, Zmuda JM, Redline S, Kao WH, Siscovick D, Florez JC, Rotter JI, Dupuis J, Wilson JG, Bowden DW, Meigs JB: Transferability and fine-mapping of glucose and insulin quantitative trait loci across populations: CARe, the Candidate Gene Association Resource. Diabetologia 2012;55:2970-2984
40. Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisterman EF, Liu L, Zhang C: Predicting risk of type 2 diabetes mellitus with genetic risk models on the basis of established genome-wide association markers: a systematic review. American Journal of Epidemiology 2013;178:1197-1207
41. Abraham G, Kowalczyk A, Zobel J, Inouye M: Performance and robustness of penalized and unpenalized methods for genetic prediction of complex human disease. Genetic Epidemiology 2013;37:184-195
42. Che R, Motsinger-Reif AA: A new explained-variance based genetic risk score for predictive modeling of disease risk. Statistical applications in genetics and molecular biology 2012;11:Article 15
43. Aschard H, Chen J, Cornelis MC, Chibnik LB, Karlson EW, Kraft P: Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. American Journal of Human Genetics 2012;90:962-972
44. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM: Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 2010;42:565-569

	Locus	Chromosome	SNP	Risk allele	Other allele
Used in 17-SNP GRS	*TCF7L2	10	rs7903146	C	T
	*CDKN2A/B	9	rs10811661	T	C
	*CDKAL1	6	rs7756992	G	A
	*THADA	2	rs10203174	C	T
	*IGF2BP2	3	rs4402960	T	G
	*SLC30A8	8	rs3802177	G	A
	**PPARG	3	rs1801282	C	G
	JAZF1	7	rs849135	G	A
	*HHEX/IDE	10	rs1111875	C	T
	ADAMTS9	3	rs6795735	C	T
	*CDC123/CAMK1D	10	rs11257655	T	C
	*KCNJ11	11	rs5215	C	T
	NOTCH2	1	rs10923931	T	G
	BCL11A	2	rs243088	T	A
	TSPAN8/LGR5	12	rs7955901	C	T
Used in 38/40-SNP GRS	**FTO	16	rs9936385	C	T
	*ADCY5	3	rs11717195	T	C
	**HMGA2	12	rs2261181	T	C
	**IRS1	2	rs2943640	C	A
	*MTNR1B	11	rs10830963	G	C
	WFS1	4	rs4458523	G	T
	*ARAP1 (CENTD2)	11	rs1552224	A	C
	*DGKB	7	rs17168486	T	C
	*GCK	7	rs10278336	A	G
	*KCNQ1	11	rs163184	G	T
	ZBED3	5	rs6878122	G	A
	**GCKR	2	rs780094	C	T
	TLE4	9	rs17791513	A	G
	*PROX1	1	rs2075423	G	T
	HNF1A (TCF1)	12	rs12427353	G	C
	PRC1	15	rs12899811	G	A
	TP53INP1	8	rs7845219	T	C
	DUSP8	11	rs2334499	T	C
	RBMS1	2	rs7569522	A	G
	ZFAND6	15	rs11634397	G	A
	**KLF14	7	rs13233731	G	A
Used in 62-SNP GRS	CILP2	19	rs10401969	C	T
	**ANKRD55	5	rs459193	G	A
	BCAR1	16	rs7202877	T	G
	KLHDC5	12	rs10842994	C	T
	**GRB14	2	rs13389219	C	T
	*UBE2E2	3	rs1496653	A	G
	**MC4R	18	rs12970134	A	G
	ANK1	8	rs516946	C	T
	HMG20A	15	rs7177055	A	G
	*MAEA	4	rs6819243	T	C
	GCC1	7	rs17867832	T	G
	TLE1	9	rs2796441	G	A
	ZMIZ1	10	rs12571751	A	G
	GLIS3*	9	rs10758593	A	G
	HNF4A	20	rs4812829	A	G
	SPRY2	13	rs1359790	G	A

***PEPD*	19	rs8182584	T	G
**C2CD4A*	15	rs4502156	T	C
**VPS26A*	10	rs12242953	G	A
KCNK16	6	rs3734621	C	A
PTPRD	9	rs16927668	T	C
SRR	17	rs2447090	A	G
AP3S2	15	rs2007084	G	A
PSMD6	3	rs12497268	G	C
ST64GAL1	3	rs17301514	A	G
ZFAND3	6	rs4299828	A	G

Table 1: Type 2 diabetes-associated loci and corresponding single-nucleotide polymorphisms (SNP) used to calculate genotype risk scores (GRS) in the Framingham Offspring and CARDIA studies, ordered by effect size in DIAGRAMv3 within each of the three waves of discovery (see Figure 1). *Locus also used in a β-cell genotype score (GRS_{β}). **Locus also used in an insulin resistance genotype score (GRS_{IR}).

Table 2: Baseline characteristics of participants in the Framingham Offspring (FOS) and CARDIA Studies

	FOS n=3471	CARDIA whites n=1650	CARDIA blacks n=820
Age (years)	35.9 (9.7)	25.5 (3.3)	24.3 (3.8)
Men	1617 (46.6)	767 (46.5)	318 (38.8)
Parental history of diabetes	383 (11.0)	159 (9.6)	146 (17.8)
BMI (kg/m^2)	25.0 (4.1)	23.7 (4.0)	25.6 (5.7)
Obese	390 (11.2)	109 (6.6)	142 (17.3)
Systolic blood pressure (mmHg)	120.6 (15.7)	109.1 (10.8)	111.4 (10.7)
Fasting plasma glucose (mg/dL)	91.1 (8.1)	82.4 (8.0)	80.9 (8.5)
HDL cholesterol (mg/dL)	51.2 (14.6)	52.1 (12.9)	54.4 (13.0)
Fasting triglycerides (mg/dL)	89.3 (68.6)	78.4 (56.9)	64.9 (32.5)

Data are means (SD) or counts (percentages), as appropriate. BMI: body-mass index; HDL: high-density lipoprotein. Obesity is defined as BMI$\geq$30 kg/m^2.

Table 3: Mean genotype risk scores in FOS and CARDIA

	17-SNP GRS_t	38/40-SNP GRS_t	62-SNP GRS_t
FOS	**17.2 (2.8)**	**39.6 (4.0)**	**66.8 (5.3)**
T2D	17.9 (2.7)	40.7 (4.0)	68.7 (5.2)
No T2D	17.1 (2.8)	39.5 (4.0)	66.7 (5.2)
CARDIA whites	---	**40.8 (3.7)**	**66.4 (5.2)**
T2D	---	42.3 (4.2)	68.4 (4.9)
No T2D	---	40.7 (3.7)	66.3 (5.1)
CARDIA blacks	---	**44.0 (3.4)**	**69.2 (4.5)**
T2D	---	44.6 (3.1)	70.1 (4.1)
No T2D	---	43.9 (3.5)	69.0 (4.6)

Data are mean (SD) genotype risk scores (GRS_t) consisting of increasing numbers of single-nucleotide polymorphisms (SNP) in the overall FOS and CARDIA cohorts and in participants with and without type 2 diabetes (T2D). A 17-SNP GRS_t was published only in FOS(4). 38 SNPs were used in CARDIA(9) and 40 SNPs in FOS(16). Among FOS and CARDIA whites, GRS_t are weighted by the effects sizes from the DIAGRAMv3 meta-analysis(12). GRS_t are unweighted among CARDIA blacks.

Table 4: Prediction models for incident type 2 diabetes without a GRS_t and with a 17-, 40-, and 62-SNP GRS_t in the Framingham Offspring Study

	Without GRS_t	With 17-SNP GRS_t	With 40-SNP GRS_t	With 62-SNP GRS_t
Demographic model				
OR (per GRS_t allele)	---	1.11 (1.07,1.15)	1.08 (1.06, 1.11)	1.08 (1.06, 1.10)
C statistic	0.698 (0.68,0.72)	0.713 (0.692, 0.734)	0.718 (0.697, 0.740)	0.726 (0.705, 0.747)
Continuous NRI	---	0.238 (0.144, 0.332)	0.321 (0.227, 0.414)	0.286 (0.192, 0.380)
Clinical model				
OR (per GRS_t allele)	---	1.10 (1.06, 1.15)	1.07 (1.04, 1.10)	1.06 (1.04, 1.08)
C statistic	0.903 (0.89,0.92)	0.905 (0.891, 0.919)	0.906 (0.892, 0.920)	0.906 (0.892, 0.920)
Continuous NRI	---	0.223 (0.129, 0.312)	0.274 (0.180, 0.368)	0.256 (0.162, 0.351)
Clinical model: age-stratified				
Continuous NRI (<50 years)	---	0.471 (0.310, 0.632)	0.423 (0.261, 0.585)	0.376 (0.213, 0.538)
Continuous NRI (≥50 years)	---	0.091 (-0.026, 0.207)	0.171 (0.055, 0.288)	0.156 (0.039, 0.272)

Data are odds ratios (OR) for type 2 diabetes (T2D) per weighted allele increase in GRS_t, C statistics, and continuous net reclassification improvement (NRI) indices comparing each GRS_t model to the corresponding model without GRS_t. Demographic model is adjusted for age and sex. Clinical models are adjusted for sex, parental T2D (yes. vs. no), body-mass index, systolic blood pressure, fasting glucose, HDL cholesterol, triglyceride levels, and (except for the age-stratified models) age.

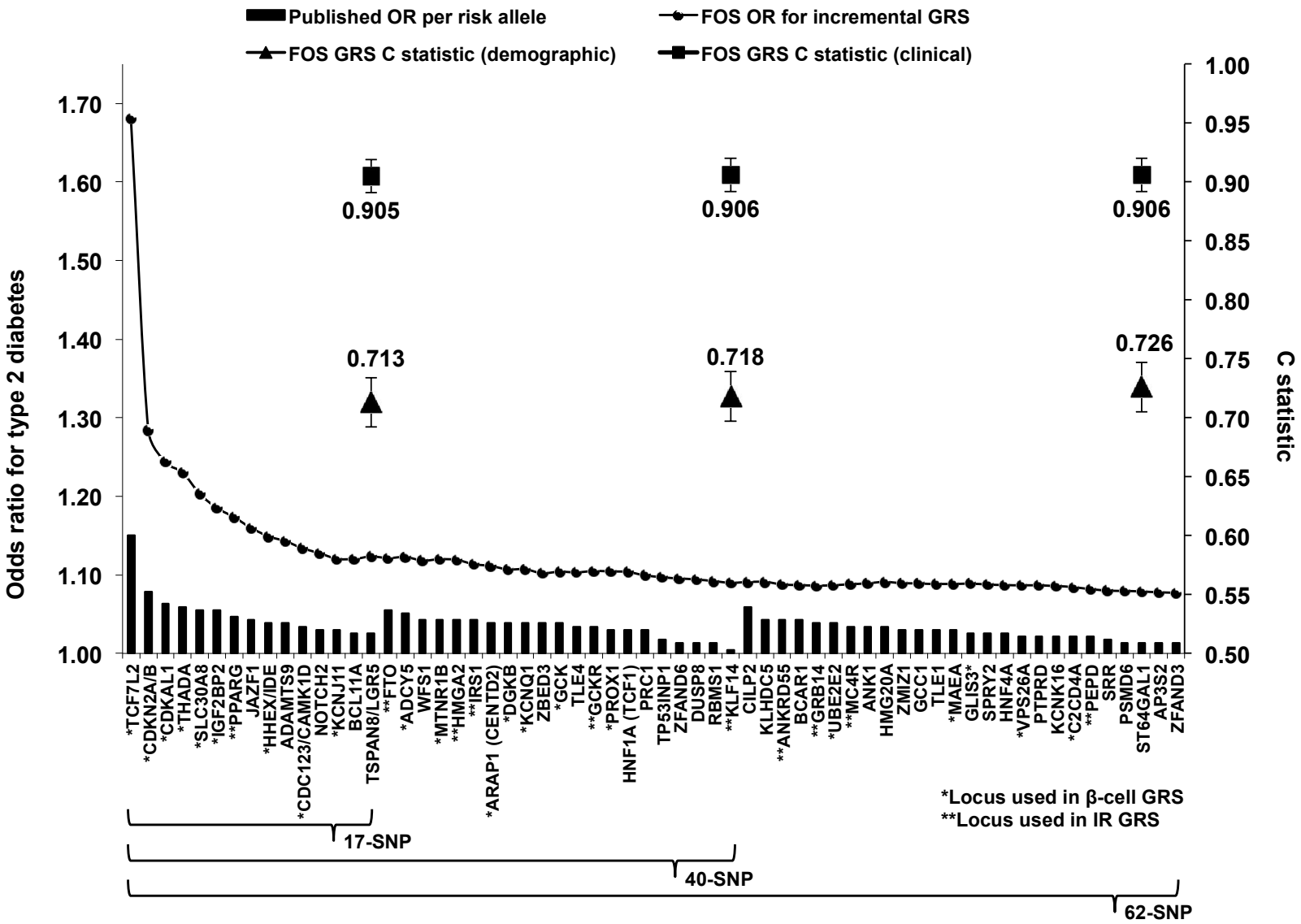
Table 5: Prediction models for incident type 2 diabetes without a GRS_t and with a 38- and 62-SNP GRS_t among whites and blacks in the CARDIA Study

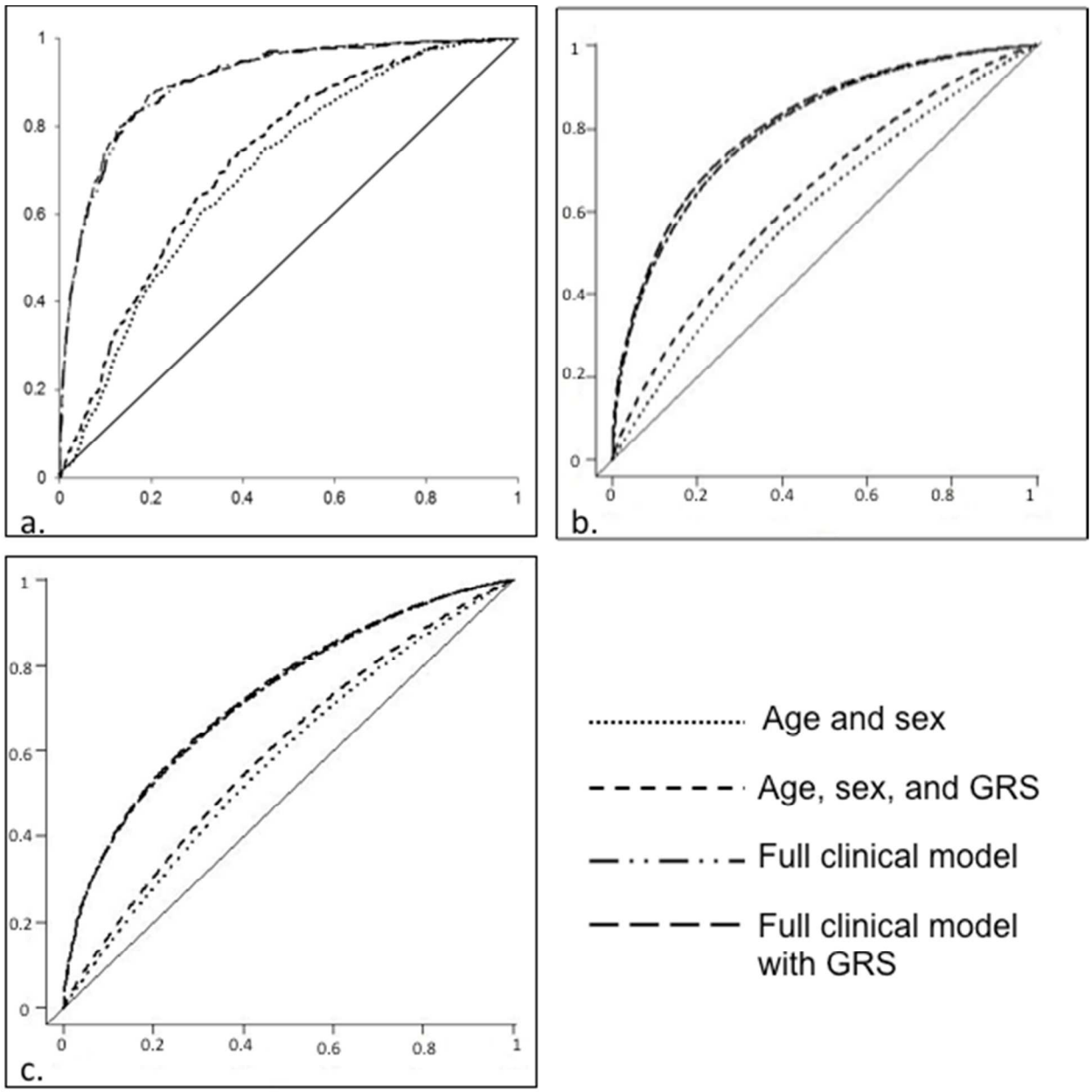
	Without GRS_t	With 38-SNP GRS_t	With 62-SNP GRS_t
Whites			
Demographic model			
HR (per GRS_t allele)	---	1.12 (1.06, 1.18)	1.08 (1.04, 1.12)
C statistic	0.613 (0.548, 0.678)	0.663 (0.604, 0.722)	0.661 (0.604, 0.717)
Continuous NRI	---	0.344 (0.129, 0.556)	0.311 (0.088, 0.525)
Clinical model			
HR (per GRS_t allele)	---	1.10 (1.04, 1.16)	1.06 (1.02, 1.10)
C statistic	0.846 (0.803, 0.889)	0.853 (0.810, 0.896)	0.853 (0.810, 0.896)
Continuous NRI	---	0.219 (-0.011, 0.434)	0.306 (0.073, 0.517)
Blacks			
Demographic model			
HR (per GRS_t allele)	---	1.05 (1.00, 1.11)	1.05 (1.01, 1.09)
C statistic	0.571 (0.515, 0.628)	0.597 (0.546, 0.649)	0.595 (0.544, 0.647)
Continuous NRI	---	0.083 (-0.137, 0.3105)	0.243 (0.031, 0.455)
Clinical model			
HR (per GRS_t allele)	---	1.06 (1.01, 1.12)	1.05 (1.00, 1.09)
C statistic	0.762 (0.717, 0.807)	0.768 (0.724, 0.813)	0.771 (0.727, 0.814)
Continuous NRI	---	0.164 (-0.051, 0.394)	0.296 (0.098, 0.513)

Data are hazard ratios (HR) for type 2 diabetes (T2D), C statistics, and continuous net reclassification improvement (NRI) indices comparing each GRS_t model to the corresponding model without GRS_t. HR are per weighted GRS_t allele in whites and per unweighted allele in blacks. Demographic models are adjusted for age and sex. Clinical models are adjusted for age, sex, parental T2D (yes. vs. no), body-mass index, systolic blood pressure, fasting glucose, log-transformed HDL cholesterol, and log-transformed triglyceride levels.

Figure 1: Type 2 diabetes (T2D)-associated genetic loci. Loci on *x*-axis are ordered by inclusion in published 17-, 40- and 62-SNP genotype risk scores. Black bars (left *y*-axis) indicate published DIAGRAMv3 odds ratio (OR) for T2D per risk allele at each locus. Black line plots the T2D OR in the Framingham Offspring Study (FOS) per allele increase in a genotype risk score (GRS) containing the loci up to that point on the *x*-axis. Points with error bars plot the C statistics (95% confidence intervals) from pooled logistic regression models for T2D in FHS including 17-, 40-, and 62-SNP GRS in demographic (triangles) and clinical (squares) models. Loci used in separate β-cell and insulin resistance (IR) GRS in the present analyses are also indicated.

Figure 2: Receiver operating characteristic (ROC) curves for models predicting incident type 2 diabetes with and without a 62-locus genetic risk score (GRS) among the Framingham Offspring (a) and white (b) and black (c) young adults in the CARDIA Study. Graphs plot the sensitivity vs. (1 − specificity) for diabetes at each possible model cutpoint. The area under a ROC curve corresponds to the C statistic of that model. Full clinical model is adjusted for age, sex, parental diabetes (yes. vs. no), body-mass index, systolic blood pressure, fasting glucose, HDL cholesterol, and triglyceride levels.


Published OR per risk allele
FOS OR for incremental GRS
FOS GRS C statistic (demographic)
FOS GRS C statistic (clinical)
Odds ratio for type 2 diabetes
C statistic
1.70
1.60
1.50
1.40
1.30
1.20
1.10
1.00
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.905
0.906
0.906
0.713
0.718
0.726
*TCF7L2
*CDKN2A/B
*CDKAL1
*THADA
*SLC30A8
*IGF2BP2
**PPARG
JAZF1
*HHEX/IDE
ADAMTS9
*CDC123/CAMK1D
NOTCH2
*KCNJ11
BCL11A
TSPAN8/LGR5
**FTO
*ADCY5
WFS1
*MTNR1B
**HMGA2
**IRS1
*ARAP1 (CENTD2)
*DGKB
*KCNQ1
ZBED3
*GCK
TLE4
**GCKR
*PROX1
HNF1A (TCF1)
PRC1
TP53INP1
ZFAND6
DUSP8
RBMS1
**KLF14
CILP2
KLHDC5
**ANKRD55
BCAR1
**GRB14
*UBE2E2
**MC4R
ANK1
HMG20A
ZMIZ1
GCC1
TLE1
*MAEA
GLIS3*
SPRY2
HNF4A
*VPS26A
PTPRD
KCNK16
*C2CD4A
**PEPD
SRR
PSMD6
ST64GAL1
AP3S2
ZFAND3
*Locus used in β-cell GRS
**Locus used in IR GRS
17-SNP
40-SNP
62-SNP


a.
b.
c.
Age and sex
Age, sex, and GRS
Full clinical model
Full clinical model with GRS

Supplementary Table 1: Rationale for categorizing 30 T2D-associated single-nucleotide polymorphisms (SNP) as affecting β-cell function or insulin resistance, based on known gene function or specific metabolic phenotypes in the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC).

								Physiology based on MAGIC analyses				
	SNP	Locus	Chr	Risk allele	T2D effect	HOMA-β effect	*p*	HOMA-β effect <-0.008	IGI *p*<0.05	AIR *p*<0.05	Proinsulin *p*<0.05	Physiology clustering
β-cell function	rs10830963	*MTNR1B*	11	G	0.0414	-0.0394	8.6E-23	xx	xx			
	rs10203174	*THADA*	2	C	0.0569	-0.0262	9.8E-06	x				x
	rs6819243	*MAEA*	4	T	0.0294	-0.0249	9.5E-03	x				
	rs7903146	*TCF7L2*	10	T	0.1399	-0.0200	1.4E-07	x	x		x	x
	rs11717195	*ADCY5*	3	T	0.0492	-0.0181	2.7E-05	x			x	x
	rs1552224	*ARAP1*	11	A	0.0374	-0.0166	9.4E-05	x	x		xx	
	rs3802177	*SLC30A8*	8	G	0.0531	-0.0160	2.0E-05	x	x	x	x	x
	rs10758593	*GLIS3*	9	A	0.0253	-0.0145	1.3E-05	x				
	rs10278336	*GCK*	7	A	0.0374	-0.0128	2.1E-04	x	x			
	rs17168486	*DGKB*	7	T	0.0374	-0.0126	3.0E-03	x	x			x
	rs2075423	*PROX1*	1	G	0.0294	-0.0125	3.9E-04	x	x			x
	rs4402960	*IGF2BP2*	3	T	0.0531	-0.0115	1.2E-03	x	x	x		
	rs4502156	*VPS13C*	15	T	0.0212	-0.0099	3.6E-03	x				
	rs7756992	*CDKAL1*	6	G	0.0607	-0.0095	7.5E-03	x	xx	x		x
	rs11257655	*CDC123*	10	T	0.0334	-0.0091	2.5E-02	x	x			
	rs1496653	*UBE2E2*	3	A	0.0374	-0.0088	1.9E-02	x				
	rs163184	*KCNQ1*	11	G	0.0374	-0.0086	1.6E-02	x		x		
	rs10811661	*CDKN2A/B*	9	T	0.0755	-0.0085	5.1E-02	x	x	x		x
	rs1111875	*HHEX/IDE*	10	C	0.0374	-0.0042	2.0E-01		xx	x		x
	rs5215	*KCNJ11*	11	C	0.0294	0.0009	7.8E-01			x		
						HOMA-IR effect	*p*	HOMA-IR *p*<0.05	FI $p<10^{-8}$	Obesity $p<10^{-8}$	IR lipid profile	Physiology clustering
Insulin resistance	rs12970134	*MC4R*	18	A	0.0334	0.0084	7.6E-02			x		x
	rs13233731	*KLF14*	7	G	0.0043	0.0077	5.1E-02	x			x	
	rs13389219	*GRB14*	2	C	0.0374	0.0124	2.2E-03	x	x			
	rs1801282	*PPARG*	3	C	0.0453	0.0161	5.6E-03	x	x		x	
	rs2261181	*HMGA2*	12	T	0.0414	0.0135	4.9E-02	x				
	rs2943640	*IRS1*	2	C	0.0414	0.0086	3.6E-02	x	x		x	
	rs459193	*ANKRD55*	5	G	0.0414	0.0115	1.1E-02	x				
	rs780094	*GCKR*	2	C	0.0334	0.0201	7.6E-07	x	x		x	
	rs8182584	*PEPD*	19	T	0.0212	0.0122	3.9E-03	x	x			
	rs9936385	*FTO*	16	C	0.0531	0.0148	3.3E-04	x		x		x

Physiology clustering as β-cell function or insulin resistance based on MAGIC analyses(1). Fasting insulin (FI) *p*-values based on body-mass index*gene analyses in (2). Obesity defined as association with risk of increased body-mass index in Genetic Investigation of ANthropometric Traits (GIANT) data(3). Insulin resistance (IR) lipid profile defined as high triglyceride and low HDL levels as reported in (2). AIR—acute insulin response; Chr—chromosome; FI: fasting insulin; HOMA—homeostasis model of assessment; IGI—insulinogenic index; MAGIC—Meta-Analysis of Glucose and Insulin-related traits Consortium; SNP—single-nucleotide polymorphism; T2D—type 2 diabetes.

Supplementary Table 2: Mean β-cell (GRS_{β}) and insulin resistance (GRS_{IR}) genotype risk scores in the Framingham Offspring and CARDIA Studies

	Total	BMI<30 kg/m^2	BMI≥30 kg/m^2
GRS_{β}			
FOS	21.6 (3.0)	21.6 (3.0)	21.6 (2.9)
T2D	22.6 (3.0)	23.2 (3.1)	22.4 (2.6)
No T2D	21.6 (3.0)	21.6 (3.0)	21.6 (2.9)
CARDIA Whites	21.2 (3.1)	21.2 (3.1)	21.0 (3.2)
T2D	22.1 (3.3)	22.3 (3.4)	21.6 (2.9)
No T2D	21.2 (3.1)	21.2 (3.1)	20.7 (3.3)
CARDIA Blacks	21.3 (2.4)	21.4 (2.4)	21.1 (2.4)
T2D	21.6 (2.5)	21.6 (2.5)	21.7 (2.5)
No T2D	21.3 (2.4)	21.3 (2.4)	20.8 (2.3)
GRS_{IR}			
FOS	10.4 (2.0)	10.4 (2.0)	10.5 (2.0)
T2D	10.3 (2.4)	10.3 (2.3)	10.3 (2.7)
No T2D	10.4 (2.0)	10.4 (2.0)	10.5 (2.0)
CARDIA Whites	10.4 (2.0)	10.4 (2.0)	10.3 (2.1)
T2D	10.6 (1.9)	10.5 (1.9)	10.8 (2.0)
No T2D	10.4 (2.0)	10.4 (2.0)	10.1 (2.1)
CARDIA Blacks	11.1 (1.9)	11.1 (1.9)	11.0 (1.8)
T2D	11.4 (1.8)	11.5 (1.8)	11.3 (1.7)
No T2D	11.0 (1.9)	11.0 (1.9)	10.9 (1.9)

Data are mean (SD) weighted genotype risk scores (GRS) consisting of 20 single-nucleotide polymorphisms (SNP) associated with β-cell dysfunction (GRS_{β}) and 10 SNP associated with insulin resistance (GRS_{IR}) in the overall FOS and CARDIA cohorts and in participants with and without type 2 diabetes (T2D). Among FOS and CARDIA whites, GRS are weighted by the effects sizes from the DIAGRAM v3 meta-analysis(4). GRS are unweighted among CARDIA blacks.

Supplementary Table 3: Odds ratios for GRS_{β} and GRS_{IR} in prediction models for incident type 2 diabetes in the Framingham Offspring Study

	GRS_{β} model	GRS_{IR} model	GRS_{β} + GRS_{IR} model
Demographic model			
GRS_{β}	1.11 (1.08, 1.15)*	---	1.11 (1.08, 1.15)*
GRS_{IR}	---	1.04 (1.00, 1.10)	1.05 (1.00, 1.10)
Clinical model			
GRS_{β}	1.10 (1.06, 1.14)*	---	1.10 (1.06, 1.14)*
GRS_{IR}	---	0.98 (0.93, 1.04)	0.99 (0.93, 1.04)

Data are odds ratios from pooled logistic regression models for incident type 2 diabetes and correspond to a 1-allele increase in the GRS. Demographic models are adjusted for age and sex. Clinical models are adjusted for age, sex, parental history of diabetes (yes vs. no), body-mass index, systolic blood pressure, fasting plasma glucose, high-density lipoprotein (HDL), and fasting triglycerides. GRS_{β} and GRS_{IR} models include only the GRS_{β} and GRS_{IR}, respectively. The GRS_{β} + GRS_{IR} model contains both terms. *$p<0.001$

Supplementary Table 4: Hazard ratios for GRS_{β} and GRS_{IR} in prediction models for incident type 2 diabetes among whites in the CARDIA Study

	GRS_{β} model	GRS_{IR} model	GRS_{β} + GRS_{IR} model
Demographic model			
GRS_{β}	1.09 (1.02, 1.16)*	---	1.09 (1.02, 1.16)**
GRS_{IR}	---	1.06 (0.96, 1.17)	1.06 (0.96, 1.17)
Clinical model			
GRS_{β}	1.09 (1.02, 1.17)**	---	1.09 (1.02, 1.17)**
GRS_{IR}	---	1.01 (0.91, 1.12)	1.01 (0.91, 1.11)

Data are hazard ratios from Cox regression models for incident type 2 diabetes and correspond to a 1-allele increase in the GRS. Demographic models are adjusted for age and sex. Clinical models are adjusted for age, sex, parental history of diabetes (yes vs. no), body-mass index, systolic blood pressure, fasting plasma glucose, log-transformed high-density lipoprotein (HDL), and log-transformed fasting triglycerides. GRS_{β} and GRS_{IR} models include only the GRS_{β} and GRS_{IR}, respectively. The GRS_{β} + GRS_{IR} model contains both terms. *p<0.05, **p<0.01

Supplementary Table 5: Hazard ratios for GRS_{β} and GRS_{IR} in prediction models for incident type 2 diabetes among blacks in the CARDIA Study

	GRS_{β} model	GRS_{IR} model	GRS_{β} + GRS_{IR} model
Demographic model			
GRS_{β}	1.06 (0.98, 1.14)	---	1.06 (0.98, 1.14)
GRS_{IR}	---	1.09 (1.00, 1.20)	1.10 (1.00, 1.20)
Clinical model			
GRS_{β}	1.06 (0.99, 1.15)	---	1.07 (0.99, 1.15)
GRS_{IR}	---	1.05 (0.96, 1.15)	1.05 (0.96, 1.16)

Data are hazard ratios from Cox regression models for incident type 2 diabetes and correspond to a 1-allele increase in the GRS. Demographic models are adjusted for age and sex. Clinical models are adjusted for age, sex, parental history of diabetes (yes vs. no), body-mass index, systolic blood pressure, fasting plasma glucose, log-transformed high-density lipoprotein (HDL), and log-transformed fasting triglycerides. GRS_{β} and GRS_{IR} models include only the GRS_{β} and GRS_{IR}, respectively. The GRS_{β} + GRS_{IR} model contains both terms.

Supplementary Table 6: Racial differences in the associations between GRS and incident type 2 diabetes

	GRS_t	GRS_β	GRS_{IR}
Demographic model			
Whites	1.077 (1.059, 1.095)	1.109 (1.079, 1.139)	1.047 (1.003, 1.093)
Blacks	1.046 (1.005, 1.088)	1.058 (0.982, 1.140)	1.095 (0.997, 1.202)
p	0.19	0.25	0.39
Clinical model			
Whites	1.060 (1.040, 1.080)	1.098 (1.063, 1.133)	0.990 (0.945, 1.038)
Blacks	1.046 (1.003, 1.090)	1.063 (0.986, 1.147)	1.049 (0.957, 1.151)
p	0.57	0.45	0.28

Data are effect sizes of the association between each GRS and incident T2D among FOS and CARDIA whites (meta-analyzed) and CARDIA blacks. Demographic models are adjusted for age and sex. Clinical models are adjusted for age, sex, parental history of diabetes (yes vs. no), body-mass index, systolic blood pressure, fasting plasma glucose, log-transformed high-density lipoprotein (HDL), and log-transformed fasting triglycerides. *P* values correspond to *t*-tests comparing the effect sizes between whites and blacks.

Supplementary Table 7: *P*-values for GRS_{β} and GRS_{IR} regression terms in prediction models for incident type 2 diabetes in the Framingham Offspring Study, stratified by body-mass index (BMI)

	GRS_{β} model	GRS_{IR} model	GRS_{β} + GRS_{IR} model
BMI≥30 kg/m^2			
GRS_{β}	<0.001	---	<0.001
GRS_{IR}	---	0.427	0.426
BMI<30 kg/m^2			
GRS_{β}	<0.001	---	<0.001
GRS_{IR}	---	0.223	0.199

Data are *p*-values from pooled logistic regression models for incident type 2 diabetes, stratified by BMI category. Models are adjusted for age and sex. GRS_{β} and GRS_{IR} models include only the GRS_{β} and GRS_{IR}, respectively. The GRS_{β} + GRS_{IR} model contains both terms.

Supplementary Table 8: *P*-values for GRS_{β} and GRS_{IR} regression terms in prediction models for incident type 2 diabetes in the overall CARDIA Study, stratified by body-mass index (BMI)

	GRS_{β} model	GRS_{IR} model	GRS_{β} + GRS_{IR} model
BMI≥30 kg/m^2			
GRS_{β}	0.018	---	0.021
GRS_{IR}	---	0.221	0.263
BMI<30 kg/m^2			
GRS_{β}	0.015	---	0.013
GRS_{IR}	---	0.084	0.070

Data are *p*-values from Cox regression models for incident type 2 diabetes, stratified by BMI category. Models are adjusted for age, sex, and race. GRS_{β} and GRS_{IR} models include only the GRS_{β} and GRS_{IR}, respectively. The GRS_{β} + GRS_{IR} model contains both terms.

Supplementary Table 9: Prediction models for incident type 2 diabetes in the Framingham Offspring Study, examining the interaction between genotype risk score and obesity

	GRS model	GRS + obesity model	GRS*obesity interaction model
GRS_t model			
GRS_t	1.08 (1.06, 1.10)	1.08 (1.06, 1.10)	1.09 (1.07, 1.12)
Obesity	---	4.46 (3.66, 5.43)	22.39 (1.66, 301.34)
GRS_t*obesity interaction	---	---	0.98 (0.94, 1.01)
GRS_β model			
GRS_β	1.11 (1.08,1.15)	1.13 (1.09, 1.16)	1.14 (1.09, 1.19)
Obesity	---	4.46 (3.66, 5.43)	8.44 (1.97, 36.16)
GRS_β*obesity interaction	---	---	0.97 (0.91, 1.04)
GRS_{IR} model			
GRS_{IR}	1.04 (1.00, 1.10)	1.03 (0.98, 1.09)	1.04 (0.98, 1.11)
Obesity	---	4.31 (3.54, 5.25)	5.11 (1.82, 14.36)
GRS_{IR}*obesity interaction	---	---	0.98 (0.89, 1.08)

Data are odds ratios (OR) from pooled logistic regression models for type 2 diabetes per weighted allele increase in 62-SNP GRS (GRS_t), β-cell GRS (GRS_β), and insulin resistance GRS (GRS_{IR}), or for obesity (BMI≥30 kg/m^2). All models are adjusted for age and sex. The GRS models include the corresponding GRS. GRS + obesity models include both the corresponding GRS and a term for obesity. GRS*obesity interaction models include the corresponding GRS, an obesity term, and an interaction term between GRS and obesity.

Supplementary Table 10: Prediction models for incident type 2 diabetes in the overall CARDIA Study, examining the interaction between genotype risk score and obesity

	GRS model	GRS + obesity model	GRS*obesity interaction model
GRS_t model			
GRS_t	1.06 (1.03, 1.09)	1.07 (1.04, 1.10)	1.07 (1.03, 1.10)
Obesity	---	6.11 (4.52, 8.26)	9.99 (0.19, 517.69)
GRS_t*obesity interaction	---	---	0.99 (0.94, 1.05)
GRS_β model			
GRS_β	1.09 (1.02, 1.16)	1.09 (1.04, 1.14)	1.07 (1.01, 1.14)
Obesity	---	6.17 (4.56, 8.34)	2.60 (0.28, 23.81)
GRS_β*obesity interaction	---	---	1.04 (0.94, 1.15)
GRS_{IR} model			
GRS_{IR}	1.08 (1.01, 1.15)	1.08 (1.01, 1.15)	1.10 (1.01, 1.19)
Obesity	---	5.94 (4.40, 8.02)	9.61 (2.01, 45.85)
GRS_{IR}*obesity interaction	---	---	0.96 (0.83, 1.10)

Data are odds ratios (OR) from Cox regression models for type 2 diabetes per weighted allele increase in 62-SNP GRS (GRS_t), β-cell GRS (GRS_β), and insulin resistance GRS (GRS_{IR}), or for obesity (BMI≥30 kg/m^2). All models are adjusted for age, sex, and race. The GRS models include the corresponding GRS. GRS + obesity models include both the corresponding GRS and a term for obesity. GRS*obesity interaction models include the corresponding GRS, an obesity term, and an interaction term between GRS and obesity.

References

1. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, Benazzo A, Rybin D, Jackson AU, Stringham HM, Song C, Fischer-Rosinsky A, Boesgaard TW, Grarup N, Abbasi FA, Assimes TL, Hao K, Yang X, Lecoeur C, Barroso I, Bonnycastle LL, Böttcher Y, Bumpstead S, Chines PS, Erdos MR, Graessler J, Kovacs P, Morken MA, Narisu N, Payne F, Stancakova A, Swift AJ, Tönjes A, Bornstein SR, Cauchi S, Froguel P, Meyre D, Schwarz PE, Häring H-U, Smith U, Boehnke M, Bergman RN, Collins FS, Mohlke KL, Tuomilehto J, Quertemous T, Lind L, Hansen T, Pedersen O, Walker M, Pfeiffer AF, Spranger J, Stumvoll M, Meigs JB, Wareham NJ, Kuusisto J, Laakso M, Langenberg C, Dupuis J, Watanabe RM, Florez JC, Ingelsson E, McCarthy MI, Prokopenko I, investigators obotM: Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes (in resubmission)
2. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu CT, Bielak LF, Prokopenko I, Amin N, Barnes D, Cadby G, Hottenga JJ, Ingelsson E, Jackson AU, Johnson T, Kanoni S, Ladenvall C, Lagou V, Lahti J, Lecoeur C, Liu Y, Martinez-Larrad MT, Montasser ME, Navarro P, Perry JR, Rasmussen-Torvik LJ, Salo P, Sattar N, Shungin D, Strawbridge RJ, Tanaka T, van Duijn CM, An P, de Andrade M, Andrews JS, Aspelund T, Atalay M, Aulchenko Y, Balkau B, Bandinelli S, Beckmann JS, Beilby JP, Bellis C, Bergman RN, Blangero J, Boban M, Boehnke M, Boerwinkle E, Bonnycastle LL, Boomsma DI, Borecki IB, Bottcher Y, Bouchard C, Brunner E, Budimir D, Campbell H, Carlson O, Chines PS, Clarke R, Collins FS, Corbaton-Anchuelo A, Couper D, de Faire U, Dedoussis GV, Deloukas P, Dimitriou M, Egan JM, Eiriksdottir G, Erdos MR, Eriksson JG, Eury E, Ferrucci L, Ford I, Forouhi NG, Fox CS, Franzosi MG, Franks PW, Frayling TM, Froguel P, Galan P, de Geus E, Gigante B, Glazer NL, Goel A, Groop L, Gudnason V, Hallmans G, Hamsten A, Hansson O, Harris TB, Hayward C, Heath S, Hercberg S, Hicks AA, Hingorani A, Hofman A, Hui J, Hung J, Jarvelin MR, Jhun MA, Johnson PC, Jukema JW, Jula A, Kao WH, Kaprio J, Kardia SL, Keinanen-Kiukaanniemi S, Kivimaki M, Kolcic I, Kovacs P, Kumari M, Kuusisto J, Kyvik KO, Laakso M, Lakka T, Lannfelt L, Lathrop GM, Launer LJ, Leander K, Li G, Lind L, Lindstrom J, Lobbens S, Loos RJ, Luan J, Lyssenko V, Magi R, Magnusson PK, Marmot M, Meneton P, Mohlke KL, Mooser V, Morken MA, Miljkovic I, Narisu N, O'Connell J, Ong KK, Oostra BA, Palmer LJ, Palotie A, Pankow JS, Peden JF, Pedersen NL, Pehlic M, Peltonen L, Penninx B, Pericic M, Perola M, Perusse L, Peyser PA, Polasek O, Pramstaller PP, Province MA, Raikkonen K, Rauramaa R, Rehnberg E, Rice K, Rotter JI, Rudan I, Ruokonen A, Saaristo T, Sabater-Lleal M, Salomaa V, Savage DB, Saxena R, Schwarz P, Seedorf U, Sennblad B, Serrano-Rios M, Shuldiner AR, Sijbrands EJ, Siscovick DS, Smit JH, Small KS, Smith NL, Smith AV, Stancakova A, Stirrups K, Stumvoll M, Sun YV, Swift AJ, Tonjes A, Tuomilehto J, Trompet S, Uitterlinden AG, Uusitupa M, Vikstrom M, Vitart V, Vohl MC, Voight BF, Vollenweider P, Waeber G, Waterworth DM, Watkins H, Wheeler E, Widen E, Wild SH, Willems SM, Willemsen G, Wilson JF, Witteman JC, Wright AF, Yaghootkar H, Zelenika D, Zemunik T, Zgaga L, Wareham NJ, McCarthy MI, Barroso I, Watanabe RM, Florez JC, Dupuis J, Meigs JB, Langenberg C: A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature Genetics 2012;44:659-669
3. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, Speliotes EK, Thorleifsson G, Willer CJ, Herrera BM, Jackson AU, Lim N, Scheet P, Soranzo N, Amin N,

Aulchenko YS, Chambers JC, Drong A, Luan J, Lyon HN, Rivadeneira F, Sanna S, Timpson NJ, Zillikens MC, Zhao JH, Almgren P, Bandinelli S, Bennett AJ, Bergman RN, Bonnycastle LL, Bumpstead SJ, Chanock SJ, Cherkas L, Chines P, Coin L, Cooper C, Crawford G, Doering A, Dominiczak A, Doney AS, Ebrahim S, Elliott P, Erdos MR, Estrada K, Ferrucci L, Fischer G, Forouhi NG, Gieger C, Grallert H, Groves CJ, Grundy S, Guiducci C, Hadley D, Hamsten A, Havulinna AS, Hofman A, Holle R, Holloway JW, Illig T, Isomaa B, Jacobs LC, Jameson K, Jousilahti P, Karpe F, Kuusisto J, Laitinen J, Lathrop GM, Lawlor DA, Mangino M, McArdle WL, Meitinger T, Morken MA, Morris AP, Munroe P, Narisu N, Nordstrom A, Nordstrom P, Oostra BA, Palmer CN, Payne F, Peden JF, Prokopenko I, Renstrom F, Ruokonen A, Salomaa V, Sandhu MS, Scott LJ, Scuteri A, Silander K, Song K, Yuan X, Stringham HM, Swift AJ, Tuomi T, Uda M, Vollenweider P, Waeber G, Wallace C, Walters GB, Weedon MN, Witteman JC, Zhang C, Zhang W, Caulfield MJ, Collins FS, Davey Smith G, Day IN, Franks PW, Hattersley AT, Hu FB, Jarvelin MR, Kong A, Kooner JS, Laakso M, Lakatta E, Mooser V, Morris AD, Peltonen L, Samani NJ, Spector TD, Strachan DP, Tanaka T, Tuomilehto J, Uitterlinden AG, van Duijn CM, Wareham NJ, Hugh W, Waterworth DM, Boehnke M, Deloukas P, Groop L, Hunter DJ, Thorsteinsdottir U, Schlessinger D, Wichmann HE, Frayling TM, Abecasis GR, Hirschhorn JN, Loos RJ, Stefansson K, Mohlke KL, Barroso I, McCarthy MI: Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. PLoS genetics 2009;5:e1000508

4. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Muller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stancakova A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutskov K, Langford C, Leander K, Lindholm E, Lobbens S, Mannisto S, Mirza G, Muhleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurethsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvanen AC, Eriksson JG, Peltonen L, Nothen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njolstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyovalti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jockel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI: Large-scale

association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature Genetics 2012;44:981-990