# Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes

**James B. Meigs, M.D., M.P.H.**, **Peter Shrader, M.S.**, **Lisa M. Sullivan, Ph.D.**, **Jarred B. McAteer, B.A.**, **Caroline S. Fox, M.D., M.P.H.**, **Josée Dupuis, Ph.D.**, **Alisa K. Manning, M.A.**, **Jose C. Florez, M.D., Ph.D.**, **Peter W.F. Wilson, M.D.**, **Ralph B. D'Agostino Sr., Ph.D.**, and **L. Adrienne Cupples, Ph.D.**

General Medicine Division (J.B. Meigs, P.S.), the Department of Medicine (J.B. Meigs, P.S., J.C.F.), and the Center for Human Genetic Research and Diabetes Unit, Department of Medicine (J.B. McAteer, J.C.F.), Massachusetts General Hospital, Boston; the Department of Biostatistics, Boston University School of Public Health, Boston (L.M.S., J.D., A.K.M., L.A.C.); the Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA (J.B. McAteer, J.C.F.); the Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Harvard Medical School, Boston (C.S.F.); the National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA (C.S.F.); the Emory Program in Cardiovascular Outcomes Research and Epidemiology, Emory University School of Medicine, Atlanta (P.W.F.W); and the Mathematics and Statistics Department, Boston University, Boston (R.B.D.)

## Abstract

**BACKGROUND—**Multiple genetic loci have been convincingly associated with the risk of type 2 diabetes mellitus. We tested the hypothesis that knowledge of these loci allows better prediction of risk than knowledge of common phenotypic risk factors alone.

**METHODS—**We genotyped single-nucleotide polymorphisms (SNPs) at 18 loci associated with diabetes in 2377 participants of the Framingham Offspring Study. We created a genotype score from the number of risk alleles and used logistic regression to generate C statistics indicating the extent to which the genotype score can discriminate the risk of diabetes when used alone and in addition to clinical risk factors.

**RESULTS—**There were 255 new cases of diabetes during 28 years of follow-up. The mean (±SD) genotype score was 17.7±2.7 among subjects in whom diabetes developed and 17.1±2.6 among those in whom diabetes did not develop (P<0.001). The sex-adjusted odds ratio for diabetes was 1.12 per risk allele (95% confidence interval, 1.07 to 1.17). The C statistic was 0.534 without the genotype score and 0.581 with the score (P=0.01). In a model adjusted for sex and self-reported family history of diabetes, the C statistic was 0.595 without the genotype score and 0.615 with the score (P=0.11). In a model adjusted for age, sex, family history, body-mass index, fasting glucose level, systolic blood pressure, high-density lipoprotein cholesterol level, and triglyceride level, the C statistic was 0.900 without the genotype score and 0.901 with the score (P=0.49). The genotype score resulted in the appropriate risk reclassification of, at most, 4% of the subjects.

**CONCLUSIONS—**A genotype score based on 18 risk alleles predicted new cases of diabetes in the community but provided only a slightly better prediction of risk than knowledge of common risk factors alone.

Address reprint requests to Dr. Meigs at the General Medicine Division, Massachusetts General Hospital, 50 Staniford St., 9th Fl., Boston, MA 02114, or at jmeigs@partners.org..

TYPE 2 DIABETES MELLITUS IS A MAJOR health problem worldwide.[1] Fortunately, its development can be prevented in many instances,[2] and persons at risk can be readily identified with the measurement of a few common risk factors.[3-5] Type 2 diabetes is heritable, with a risk for people with familial diabetes as compared with those without familial diabetes that is increased by a factor of 2 to 6.[6,7] Recent genetic association studies have provided convincing evidence that several novel loci are associated with the risk of diabetes,[8-13] each with a 5 to 37% increase in the relative odds of diabetes per risk allele.[13] These loci may account for the familial basis of diabetes, and their discovery may herald a new era in the prediction of the disease, whereby individual loci might be combined into a genetic risk score for enhanced detection of persons at risk for diabetes.[14-17]

In clinical practice, a few common risk factors that can be easily measured are powerful harbingers of type 2 diabetes.[5] Despite the intuitive appeal of a genetic risk score, it remains an untested hypothesis that genetic information allows better prediction of the risk of diabetes than knowledge of common risk factors alone.[18,19] We tested this hypothesis in the framework of three perspectives. First, we asked whether autosomal genetic information at birth, when only sex (the combination of X and Y chromosomes) is known, improves the ability to identify people who are at increased risk for the development of diabetes by middle age. Second, we asked whether genotype information would add to knowledge of family history, which is commonly considered to represent genetic risk. Finally, we asked whether adding genetic information to information on risk factors that are commonly measured at a clinical examination in adulthood improves the prediction of risk. We used this framework to test the ability of a panel of 18 single-nucleotide polymorphisms (SNPs) that are known to have associations with the risk of diabetes to predict new cases of type 2 diabetes in a large, prospectively examined, community-based cohort.

## METHODS

### STUDY SAMPLE

The Framingham Heart Study commenced in 1948 with the enrollment of 5209 people of European ancestry, 28 to 62 years of age, residing in Framingham, Massachusetts; the participants were subsequently examined every 2 years. The Framingham Offspring Study commenced in 1971 with the enrollment of 5124 offspring of the original cohort and the spouses of the offspring; participants were 5 to 70 years of age at the first examination.[20] Participants were next examined 8 years later and then every 4 years thereafter through examination 7 (which took place in the period between 1999 and 2001). Of the 5124 participants in the Framingham Offspring Study, 2776 were genotyped for 18 SNPs. For 2377 of these participants, complete phenotypic and follow-up data over the course of one or more observational periods were available, as well as complete information on at least 15 of 18 genotyped SNPs. The study was approved by the institutional review board at Boston University, and written informed consent was obtained from all participants.

### PHENOTYPIC RISK FACTORS AND DEFINITIONS OF DIABETES

Each examination consisted of a medical history taking, physical examination, and collection of a fasting blood sample.[21] In the sixth examination cycle (1995 through 1998), participants completed a self-administered questionnaire that asked about family history of disease. We defined a positive self-reported family history of diabetes as a report that one or both parents had diabetes; this definition is more than 56% sensitive and 97% specific for confirmed parental diabetes.[22] Parental diabetes was confirmed by means of direct observation of the original cohort, over the course of 46 years of observation after their enrollment in the Framingham Heart Study, at the end of which time the mean age of surviving parents was 83 years. We considered diabetes to be present in a parent when medication was prescribed to control the

diabetes or when the casual plasma glucose level was 11.1 mmol per liter or higher or 200.0 mg per deciliter or higher at any examination. We defined diabetes to be present in an offspring when treatment was prescribed to control the diabetes or when the fasting plasma glucose level was 7.0 mmol per liter or higher or 126.0 mg per deciliter or higher at any examination. More than 99% of the cases of diabetes among the participants in the Framingham Offspring Study are type 2 diabetes.[6]

## SELECTION OF RISK ALLELES, GENOTYPING, AND GENOTYPE-SCORE CONSTRUCTION

We used two recent diabetes genomewide association studies[10,13] to select 17 SNPs confirmed to be associated with type 2 diabetes in populations of European ancestry. We added one SNP, rs689, in the *INS* locus, that we had previously found to be associated with diabetes (P=0.02) in Framingham Offspring Study participants.[23] Using these 18 SNPs, we constructed a genotype score ranging from 0 to 36 on the basis of the number of risk alleles. Genotyping was performed with the use of iPLEX (Sequenom).[24] The minimum call rate was 96.9%, the average consensus rate from 254 duplicates was 99.5%, and all SNPs were in Hardy-Weinberg equilibrium (P>0.02).[25]

## STATISTICAL ANALYSIS

We used mixed-effects models to compare the mean genotype score for persons in whom diabetes developed with the score for those in whom diabetes did not develop. We estimated the cumulative incidence of diabetes by dividing the number of persons in whom diabetes developed as of the end of follow-up by the total number at risk. We used pooled logistic-regression models with generalized estimating equations to examine the association between the genotype score and the risk that diabetes would develop over the course of 28 years. The use of generalized estimating equations accounts for the presence of related persons in the sample,[26] and the method of pooling person-examinations accounts for time-dependent risk factors, providing valid estimates of effect similar to those obtained with the use of time-dependent Cox analyses.[27,28] We pooled three examination periods (examinations 1 and 2, 3 and 4, and 5 through 7) to test the 8-to-10-year risk of diabetes, as we did in our previous diabetes prediction model.[5] For these analyses, subjects with diabetes at the first examination of each period were excluded, and new cases of diabetes were enumerated through the end of the last examination in the period. We constructed a series of genotype-score models that were adjusted for sex, for sex and self-reported family history of diabetes, and for risk factors identified in our previously published and validated "simple clinical model," including sex, family history, age, body-mass index, fasting plasma glucose level, systolic blood pressure, high-density lipoprotein cholesterol level, and triglyceride level.[5,29] We had previously estimated the precision of this model using bootstrap resampling.[5] Since self-report is the method by which information on family history is almost always collected, we used self-reported family history of diabetes in the primary analysis.

We calculated odds ratios and 95% confidence intervals associated with each additional risk allele for each SNP individually and in the genotype score. Using C statistics that were compared with a nonparametric approach, we evaluated the discriminatory capability of the models with the genotype score as compared with the models without the genotype score.[30] We also evaluated risk reclassification with the use of the genotype score, according to the method developed by Pencina et al. for determining net reclassification improvement.[31] We assessed model calibration using the Hosmer-Lemeshow chi-square test.[32] We used categories of genotype score to calculate likelihood ratios and posterior probabilities of diabetes.[33] Statistical analyses were performed with the use of SAS software, version 8 (SAS Institute). A two-tailed P value of less than 0.05 was considered to indicate statistical significance.

# RESULTS

## CHARACTERISTICS OF THE SUBJECTS, RISK ALLELES, AND GENOTYPE SCORE

Characteristics of the 2377 subjects at the baseline of each of the three cross-sectional periods are shown in Table 1. From the inception of the study in 1971 to the baseline of the third period in 1987, subjects gained weight and levels of risk factors became more adverse. Through the end of follow-up, 255 cases of diabetes accumulated over 6130 person-examinations. Characteristics of 18 risk loci are shown in Table 2. The risk alleles were common (with a risk-allele frequency of >10%), genotype frequencies were similar to those in other samples of subjects of European ancestry, and in this population, few individual odds ratios were significant. The mean (±SD) genotype score was 17.7±2.7 among subjects in whom diabetes developed and 17.1±2.6 among those in whom diabetes did not develop (P<0.001), and the cumulative incidence of diabetes increased significantly with increasing score (P<0.001) (Fig. 1, and Fig. 1 and 2 in the Supplementary Appendix, available with the full text of this article at www.nejm.org). Of 2377 subjects, 2.7% were homozygous for any risk allele, 16.2% were heterozygous for the risk allele, and 8.4% had no risk allele at half or more (≥9 of 18 SNPs) of the loci.

## GENOTYPE SCORE AND RISK OF DIABETES

Three regression models for predicting diabetes are shown in Table 3, and in Figure 3 in the Supplementary Appendix. The C statistic for the sex-adjusted model was low (0.534) but improved significantly with the addition of the genotype score (0.581, P=0.01), and the relative risk for diabetes increased by 12% per risk allele. In a model adjusted for sex and self-reported family history of diabetes, the C statistic was modest without the genotype score (0.595) and did not improve significantly with the score. In a model adjusted for the risk factors included in the simple clinical model, the C statistic was excellent without the genotype score (0.900) and did not improve significantly with the score, and the genetic relative risk remained constant at 11% per risk allele.

Net reclassification by means of the genotype score over the course of 8 to 10 years of follow-up is shown in Table 4, and in Figure 4 in the Supplementary Appendix. In the model that was adjusted for sex alone, the genotype score appropriately reclassified 4.1% of the participants (P=0.004), primarily by reclassifying lower-risk persons into higher-risk categories. In the models adjusted for sex and family history or for simple clinical risk factors, the genotype score appropriately reclassified smaller proportions of participants (≤2.6%, P≥0.17).

Adjustment for age diminished the genotype score's capacity to improve discrimination, with a C statistic in the sex-adjusted model of 0.729 without the score and 0.741 with the score (P=0.05) (Table 1 in the Supplementary Appendix). However, when the sample was stratified by age (<50 years vs. ≥50 years), discrimination with the addition of the genotype score appeared to be substantially better among younger subjects (C statistic, 0.532 to 0.609; P=0.009; net reclassification improvement, 11.9%; P=0.009) than among older subjects (C statistic, 0.530 to 0.558; P=0.20; net reclassification improvement, 0.47%; P=0.92). The interaction of age with genotype score was not significant (P=0.37).

## SUBSIDIARY ANALYSES OF THE GENOTYPE SCORE

C statistics for models classifying the genotype score into three groups were virtually identical to those for models using the genotype score per risk allele. In the model adjusted for sex, the odds ratio for the development of diabetes in the middle-score group (genotype score, 16 to 20) as compared with the low-score group (genotype score, ≤15) was 1.62 (95% confidence interval [CI], 1.14 to 2.29), and the odds ratio for the development of diabetes in the high-score group (genotype score, ≥21) as compared with the low-score group was 2.60 (95% CI, 1.68 to

4.02) (Fig. 1, and Table 2 in the Supplementary Appendix). Positive likelihood ratios for estimating the posterior probability of diabetes were 0.62, 1.04, and 1.73 in the low-score, middle-score, and high-score groups, respectively (Table 3 in the Supplementary Appendix). Models that used a weighted genotype score had discriminatory properties that were similar to those in models that used the unweighted score (Tables 4, 5, and 6 in the Supplementary Appendix). Sex-adjusted odds ratios for diabetes associated with directly observed parental diabetes as compared with no parental diabetes were 1.91 (95% CI, 1.44 to 2.55) without the genotype score and 1.82 (95% CI, 1.37 to 2.43) with the genotype score, and C statistics were 0.576 without the genotype score and 0.604 with the score (P=0.048). Models that used risk factors in clinical categories had results that were similar to those of models that used continuous risk factors (Tables 7 and 8 in the Supplementary Appendix).

## DISCUSSION

In a community-based sample followed for 28 years, we found that a genotype score for type 2 diabetes, based on 18 loci, was associated with a very modest but significant 12% increase in the relative risk of diabetes per risk allele. Adjustment for family history and common risk factors did not diminish the size or significance of this association. Irrespective of clinical variables, people with the highest genotype scores as compared with those with the lowest scores had a risk that was increased by a factor of 2.6. Although the individual risk alleles were common, only a small percentage of people had at least one risk allele at half or more of the loci. It might be expected that a score based on common variants would not be an efficient discriminator of risk, owing to weak effects for individual alleles.[34] However, we found that a combination of risk alleles was a strong risk factor with modest discriminatory ability when sex alone was considered, especially among younger persons. This finding might be useful for genetic screening at birth or in youth, before obvious risk factors have developed.

When familial diabetes or clinical risk factors that are typically documented at a periodic examination in adulthood were considered, the genotype score did not improve risk discrimination. A possible explanation for this finding is that some alleles might increase the risk through these intermediate traits or that phenotypic risk factors are overwhelmingly stronger determinants of the near-term risk of diabetes than are known genetic influences. Findings were similar with use of a score in which loci were weighted according to previous evidence of their association with diabetes. The results suggest that "personalized medicine" that is made possible by the expanded understanding of genetics is not yet as useful for the prediction of the risk of diabetes in adults as it is for other potential applications such as pharmacogenetic analyses of drug toxicity or response.

A few other studies have examined the use of combinations of SNPs to predict the risk of diabetes. In the Botnia study, people with risk alleles in both the gene encoding for the peroxisome proliferator-activated receptor gamma (*PPARG*) and the gene encoding for the cystein protease calpain 10 (*CAPN10*), as compared with people who had no risk alleles, had a risk that was increased by a factor of 2.6,[14] but the use of risk alleles as predictors did not result in a better C statistic for diabetes (0.68) than did the use of fasting glucose level and body-mass index.[18] In a study from the United Kingdom, subjects with all six risk alleles in the gene encoding for potassium inwardly-rectifying channel, subfamily J, member 11 (*KCNJ11*), *PPARG*, and the transcription factor 7-like gene (*TCF7L2*) (1% of the sample), in comparison with subjects who had no risk alleles, had a risk that was increased by a factor of 5 to 7; the C statistic with the three loci as predictors was 0.58.[15] In the Data from an Epidemiological Study on the Insulin Resistance Syndrome (DESIR) study, carriers of at least 4 risk alleles in the genes encoding for glucokinase (*GCK*), interleukin 6 (*IL6*), and *TCF7L2*, as compared with those who had one risk allele or none, had a risk that was increased by a factor of 2.5.[17] The C statistic with these three variants as predictors was 0.56, and the C statistic

with these loci plus age, sex, and body-mass index as predictors was 0.82. The Genetics of Diabetes Audit and Research Tayside (GoDARTS) study examined 18 risk loci.[35] Carriers of more than 24 risk alleles (1.2% of the sample), as compared with carriers of 10 to 12 risk alleles, had a prevalence ratio of 4.2. The C statistic with all variants combined as predictors was 0.60; the C statistic with age, body-mass index, and sex as predictors was 0.78; and the C statistic with variants and risk factors as predictors was 0.80. Our data extend these studies to show that individual per-allele effects are small; that people with more risk alleles are at greater risk than those with fewer, no matter how many or which genes are considered; that groups with an apparently greatly increased genetic risk can be identified but are not commonly found; and that the marginal ability of genotype scores to discriminate risk is small, with minimal effect after consideration of even a few common risk factors.

We found that the presence or absence of parental diabetes and the genotype score were independently associated with the risk of diabetes. This suggests that family history as a risk factor for diabetes conveys more than heritable genetic information; it probably includes nongenetic familial behaviors and norms. The lower relative risks for diabetes associated with observed parental diabetes as compared with those associated with self-reported family history (approximately 1.8 vs. approximately 2.2) support the contention that family history contains more risk information than is implied by inheritance of the diabetes phenotype alone.

One of the limitations of our study is that the 18 SNPs we included are probably insufficient to account for the familial risk of diabetes. They account for a minority of diabetes heritability, and the SNP array platforms from which they were chosen capture only approximately 80% of common variants in Europeans. In addition, we have not considered structural variants that might confer a risk of diabetes. It is possible that the addition of rare risk alleles with large effects, or a much larger number of common risk alleles with small individual effects, could improve discrimination.[36] Indeed, as many as 500 loci may underlie the genetic risk of type 2 diabetes.[16] Also, we did not study interactions among genes or between genes and the environment that might alter the genetic risk in exposed persons. As more diabetes risk variants become known, their incorporation into the genotype score may explain more of the genetic risk implied by parental diabetes.

Our study has other limitations. There were few significant associations between individual risk alleles and diabetes in the Framingham Offspring Study cohort, but this finding was expected, given that alleles of small effect were tested in a community-based sample of modest size, and the aggregate set of 18 SNPs was predictive of new cases of diabetes. The participants in the Framingham Offspring Study are essentially all of European ancestry; allelic variation may require that different SNPs be used to generate a genotype score in different ancestry groups.[37] Our genotype score gave all alleles the same weight; this may not be a true reflection of the biologic basis of type 2 diabetes. We considered the marginal value of the genotype score after accounting for only phenotypic risk factors, without consideration of behavioral risk factors for diabetes.[38] We expect that accounting for unhealthful behaviors associated with the risk of diabetes would only further diminish the discriminatory capacity of a genotype score. However, persons with relatively less healthful lifestyle behaviors might be more susceptible to genetic risk than those with more healthful behaviors.[39] Whether the genotype score would have value in predicting the risk of diabetes in specific subgroups that have an elevated risk on the basis of poor health habits remains to be tested.

In summary, a genotype score based on 18 risk alleles predicted new cases of diabetes in the community but did not result in a substantially better prediction of risk than the knowledge of common phenotypic risk factors alone. Although genetic information appeared to be useful when only factors known in youth were considered, genetic information in the context of risk factors measured in adulthood did not help to refine the prediction of diabetes risk. Our findings

underscore the view that identification of adverse phenotypic characteristics remains the cornerstone of approaches to predicting the risk of type 2 diabetes.[19]

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
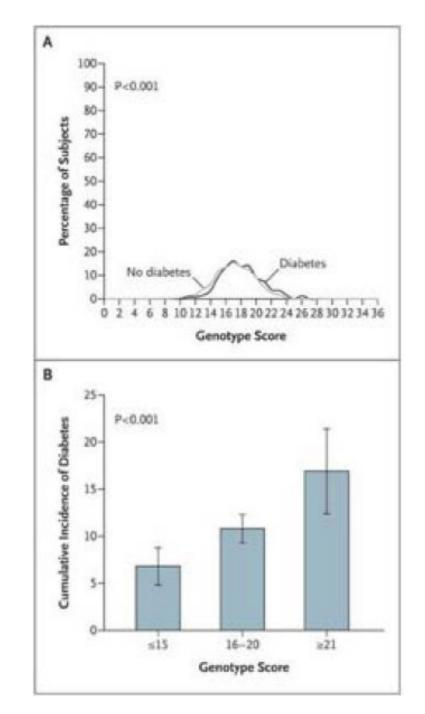
## Acknowledgments

## References

1. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. Diabetes Care 1998;21:1414–31. [PubMed: 9727886]

2. Gillies CL, Abrams KR, Lambert PC, et al. Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. BMJ 2007;334:299. [PubMed: 17237299]

3. Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: the Atherosclerosis Risk in Communities study. Diabetes Care 2005;28:2013–8. [PubMed: 16043747]

4. Stern MP, Williams K, Haffner SM. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? Ann Intern Med 2002;136:575–81. [PubMed: 11955025]

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;167:1068–74. [PubMed: 17533210]

6. Meigs JB, Cupples LA, Wilson PWF. Parental transmission of type 2 diabetes mellitus: the Framingham Offspring Study. Diabetes 2000;49:2201–7. [PubMed: 11118026]

7. Weijnen CF, Rich SS, Meigs JB, Krolewski AS, Warram JH. Risk of diabetes in siblings of index cases with Type 2 diabetes: implications for genetic studies. Diabet Med 2002;19:41–50. [PubMed: 11869302]

8. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 2006;38:320–3. [PubMed: 16415884]

9. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881–5. [PubMed: 17293876]

10. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331–6. [PubMed: 17463246]

11. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341–5. [PubMed: 17463248]

12. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336–41. [PubMed: 17463249] Science 2007;317:1035–6. [PubMed: 17717169]Erratum

13. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638–45. [PubMed: 18372903]

14. Lyssenko V, Almgren P, Anevski D, et al. Genetic prediction of future type 2 diabetes. PLoS Med 2005;2(12):e345. [PubMed: 17570749]
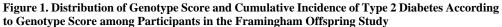
15. Weedon MN, McCarthy MI, Hitman G, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 2006;3(10):e374. [PubMed: 17020404]

16. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 2007;17:1520–8. [PubMed: 17785532]

17. Vaxillaire M, Veslot J, Dina C, et al. Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study. Diabetes 2008;57:244–54. [PubMed: 17977958]

18. Janssens ACJW, Gwinn M, Khoury MJ, Subramonia-Iyer S. Does genetic testing really improve the prediction of future type 2 diabetes? PLoS Med 2006;3(2):e114. [PubMed: 16492078]

19. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle — will we get our wish? N Engl J Med 2008;358:105–7. [PubMed: 18184955]

20. Kannel WB, Feinleib M, McNamara JR, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. Am J Epidemiol 1979;110:281–90. [PubMed: 474565]

21. Meigs JB, Nathan DM, Wilson PWF, Cupples LA, Singer DE. Metabolic risk factors worsen continuously across the spectrum of nondiabetic glucose tolerance: the Framingham Offspring Study. Ann Intern Med 1998;128:524–33. [PubMed: 9518396]

22. 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. Ann Intern Med 2004;140:434–40. [PubMed: 15023709]

23. Meigs JB, Dupuis J, Herbert AG, Liu C, Wilson PW, Cupples LA. The insulin gene variable number tandem repeat and risk of type 2 diabetes in a population-based sample of families and unrelated men and women. J Clin Endocrinol Metab 2005;90:1137–43. [PubMed: 15562019]

24. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9. [PubMed: 12029063]

25. Emigh TH. A comparison of tests for Hardy-Weinberg equilibrium. Biometrics 1980;36:627–42.

26. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30. [PubMed: 3719049]

27. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. Stat Med 1990;9:1501–15. [PubMed: 2281238]

28. Cupples LA, D'Agostino RB, Anderson K, Kannel WB. Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study. Stat Med 1988;7:205–22. [PubMed: 3353604]

29. Li J, Bornstein SR, Landgraf R, Schwarz PE. Validation of a simple clinical diabetes prediction model in a middle-aged, white, German population. Arch Intern Med 2007;167:2528–9. [PubMed: 18071180]

30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45. [PubMed: 3203132]

31. 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;27:157–72. [PubMed: 17569110]

32. Hosmer, DW., Jr; Lemeshow, S. Applied logistic regression. John Wiley; New York: 1989.

33. Sackett, DL.; Haynes, RB.; Guyatt, GH.; Tugwell, P. Clinical epidemiology: a basic science for clinical medicine. Little, Brown; Boston: 1991.

34. Ware JH. The limitations of risk factors as prognostic tools. N Engl J Med 2006;355:2615–7. [PubMed: 17182986]

35. Lango H, Palmer CN, Morris AD, et al. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes. June 30;2008 Epub ahead of print

36. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928–35. [PubMed: 17309939]

37. Lewis JP, Palmer ND, Hicks PJ, et al. Association analysis in African Americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. Diabetes 2008;57:2220–5. [PubMed: 18443202]

38. Hu FB, Manson JE, Stampfer MJ, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. N Engl J Med 2001;345:790–7. [PubMed: 11556298]

39. Florez JC, Jablonski KA, Bayley N, et al. *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. N Engl J Med 2006;355:241–50. [PubMed: 16855264]

**Figure 1. Distribution of Genotype Score and Cumulative Incidence of Type 2 Diabetes According to Genotype Score among Participants in the Framingham Offspring Study**
The panels were drawn on the basis of 2434 participants with complete genotypic data. (Data on 2377 participants with complete genotypic and phenotypic data were included in the remainder of the analysis reported in this article.) Panel A shows the distribution of participants in the Framingham Offspring Study according to genotype score, stratified according to persons in whom diabetes developed over a period of 28 years of follow-up and those in whom diabetes did not develop. The minimum genotype score was 7, and the maximum was 27. The P value is for the difference in the mean genotype score between the two groups, accounting for the presence of related persons in the sample. Panel B shows the 28-year cumulative

incidence of type 2 diabetes grouped according to whether the genotype score was low (≤15, 41 cases of diabetes among 605 participants at risk), medium (16 to 20, 169 cases among 1562 at risk), or high (≥21, 45 cases among 267 at risk). Overall, among the 2434 participants at risk, 24.9% had a low genotype score, 64.2% had a medium score, and 11.0% had a high score. The P value is for the difference in cumulative incidence across genotype-score groups, accounting for the presence of related persons in the sample. I bars indicate standard errors.

**Table 1**

Characteristics of Participants in the Framingham Offspring Study at the Baseline of Each of Three Examination Periods in the Pooled Analysis[*]

| Characteristic | Examination Period | | |
|---|---|---|---|
| | Period 1, Examinations 1 and 2 (1971-1983) | Period 2, Examinations 3 and 4 (1979-1991) | Period 3, Examinations 5-7 (1987-2001) |
| Duration of period (yr)[†] | 7.9±0.5 | 7.8±0.6 | 10.6±0.9 |
| No. of subjects | 2188 | 1916 | 2026 |
| Age (yr) | 35±9.1 | 42±9.6 | 50±9.7 |
| Female sex (%) | 53.2 | 52.7 | 52.2 |
| Self-reported family history of diabetes (%) | 17.9 | 18.5 | 18.0 |
| Body-mass index[‡] | 24.9±4.1 | 25.2±4.2 | 26.6±4.6 |
| Systolic blood pressure (mm Hg) | 120±14.7 | 121±17.3 | 128±21.5 |
| High-density lipoprotein cholesterol (mg/dl)[§] | 51.2±14.4 | 49.0±13.0 | 50.2±14.6 |
| Triglycerides (mg/dl)[¶] | 88±66.8 | 99±86.5 | 118±88.1 |
| Fasting plasma glucose (mg/dl)[‖] | 91±8.0 | 91±9.0 | 91±8.9 |

[*] The sample comprised 2377 participants for whom no data were missing, with 6130 person-examinations and 255 cases of diabetes observed over three periods (35 cases in the first period, 40 in the second, and 180 in the third). This sample represents 1335 unrelated participants, 918 sibling pairs, 63 avuncular pairs, and 612 cousins from 279 pedigrees.

[†] Duration of period is the mean number of years between the first and last examinations in each period.

[‡] The body-mass index is the weight in kilograms divided by the square of the height in meters.

[§] To convert the values for high-density lipoprotein cholesterol to millimoles per liter, multiply by 0.02586.

[¶] To convert the values for triglycerides to millimoles per liter, multiply by 0.01129.

[‖] To convert the values for glucose to millimoles per liter, multiply by 0.05551.

**Table 2**

Genetic Loci, Genotype Frequencies, and Individual Associations with Diabetes among Participants in the Framingham Offspring Study for 18 Single-Nucleotide Polymorphisms (SNPs) Associated with Type 2 Diabetes in Previous Studies[*]

| SNP | Locus | Chromosome | Locus Relative to Gene | Risk Allele | Published Odds Ratio | Published P Value | Source | Genotype Frequencies[†] | | | | | | Risk-Allele Frequency in Those without Diabetes % | Odds Ratio per Risk Allele for Diabetes in Framingham Offspring Study Participants (95% CI) | P Value[†] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | No Risk Allele | | 1 Risk Allele | | 2 Risk Allele | | | | |
| | | | | | | | | No Diabetes | Diabetes | No Diabetes | Diabetes | No Diabetes | Diabetes | | | |
| | | | | | | | | *no. of patients (%)* | | | | | | | | |
| rs1092931 | NOTCH2 | 1 | Intron 5 | T | 1.13 | $6.7\times10^{-6}$ | Zeggini et al.[13] | 1716 (80.9) | 206 (80.8) | 379 (17.9) | 44 (17.3) | 26 (1.2) | 5 (2.0) | 10.2 | 1.11 (0.83-1.48) | 0.48 |
| rs10490072 | BCL11A | 2 | 3' of gene | T | 1.05 | $4.9\times10^{-5}$ | Zeggini et al.[13] | 129 (6.1) | 15 (5.9) | 788 (37.3) | 102 (40.0) | 1197 (56.6) | 137 (53.7) | 75.3 | 0.94 (0.77-1.16) | 0.57 |
| rs7578597 | THADA | 2 | Missense, exon 24 | T | 1.15 | $1.1\times10^{-4}$ | Zeggini et al.[13] | 17 (80.0) | 2 (0.8) | 372 (17.6) | 36 (14.1) | 1730 (81.6) | 217 (85.1) | 90.4 | 1.30 (0.93-1.83) | 0.12 |
| rs1470579 | IGF2BP2 | 3 | Intron 2 | C | 1.14 | $8.9\times10^{-16}$ | Saxena et al.[10] | 971 (46.1) | 103 (40.4) | 896 (42.5) | 120 (47.1) | 240 (11.4) | 31 (12.2) | 32.7 | 1.20 (1.00-1.44) | 0.054 |
| rs1801282 | PPARg | 3 | Intron 1 | C | 1.14 | $1.7\times10^{-6}$ | Saxena et al.[10] | 21 (1.0) | 5 (2.0) | 411 (19.5) | 38 (14.9) | 1681 (79.6) | 210 (82.4) | 89.3 | 1.13 (0.83-1.55) | 0.43 |
| rs4607103 | ADAMTS9 | 3 | Intron 2 | C | 1.09 | $3.1\times10^{-4}$ | Zeggini et al.[13] | 161 (7.6) | 11 (4.3) | 802 (37.8) | 103 (40.4) | 1157 (54.6) | 141 (55.3) | 73.5 | 1.11 (0.90-1.36) | 0.32 |
| rs7754840 | CDKAL1 | 6 | Intron 5 | C | 1.12 | $4.1\times10^{-11}$ | Saxena et al.[10] | 990 (47.0) | 111 (43.5) | 890 (42.2) | 104 (40.8) | 227 (10.8) | 37 (14.5) | 31.9 | 1.15 (0.96-1.39) | 0.14 |
| rs9472138 | VEGFA | 6 | 3' of gene | T | 1.06 | $4.1\times10^{-5}$ | Zeggini et al.[13] | 1098 (51.7) | 122 (47.8) | 849 (40.0) | 106 (41.6) | 175 (8.3) | 27 (10.6) | 28.3 | 1.15 (0.95-1.39) | 0.16 |
| rs864745 | JAZF1 | 7 | Intron 1 | T | 1.10 | $4.6\times10^{-5}$ | Zeggini et al.[13] | 575 (27.3) | 69 (27.1) | 1011 (47.9) | 110 (43.1) | 524 (24.8) | 76 (29.8) | 48.8 | 1.04 (0.87-1.23) | 0.68 |
| rs13266634 | SLC30A8 | 8 | Missense, exon 8 | C | 1.12 | $5.3\times10^{-8}$ | Saxena et al.[10] | 175 (8.3) | 19 (7.5) | 851 (40.1) | 97 (38.0) | 1094 (51.6) | 139 (54.5) | 71.7 | 1.08 (0.89-1.33) | 0.43 |
| rs10811661 | CDKN2A/2B | 9 | 5' of gene | T | 1.20 | $7.8\times10^{-15}$ | Saxena et al.[10] | 79 (3.7) | 4 (1.6) | 630 (29.7) | 48 (18.8) | 1410 (66.5) | 202 (79.2) | 81.4 | 1.76 (1.33-2.33) | <0.001 |
| rs1111875 | HHEX | 10 | 3' of gene | C | 1.13 | $5.7\times10^{-10}$ | Saxena et al.[10] | 339 (16.0) | 44 (17.3) | 1031 (48.7) | 126 (49.4) | 748 (35.3) | 85 (33.3) | 59.7 | 0.98 (0.81-1.17) | 0.78 |
| rs12779790 | CDC123,CAMK1D | 10 | 3' of gene | G | 1.11 | $4.7\times10^{-5}$ | Zeggini et al.[13] | 1449 (68.4) | 162 (63.5) | 589 (27.8) | 85 (33.3) | 80 (3.8) | 8 (3.1) | 17.7 | 1.08 (0.86-1.35) | 0.52 |
| rs7903146 | TCF7L2 | 10 | Intron 6 | T | 1.37 | $1.0\times10^{-48}$ | Saxena et al.[10] | 1028 (48.9) | 94 (36.9) | 868 (41.3) | 123 (48.2) | 207 (9.8) | 37 (14.5) | 30.5 | 1.34 (1.12-1.61) | 0.002 |
| rs5219 | KCNJ11 | 11 | Missense, exon 1 | T | 1.14 | $6.7\times10^{-11}$ | Saxena et al.[10] | 897 (42.5) | 108 (42.4) | 950 (45.0) | 108 (42.4) | 266 (12.6) | 39 (15.3) | 35.1 | 1.09 (0.91-1.31) | 0.35 |
| rs689 | INS | 11 | Intron 1 | T | 1.93 | $2.0\times10^{-2}$ | Meigs et al.[23] | 1101 (51.9) | 124 (48.6) | 857 (40.4) | 106 (41.6) | 164 (7.7) | 25 (9.8) | 27.9 | 1.16 (0.95-1.41) | 0.14 |
| rs1153188 | DCD | 12 | 5' of gene | A | 1.08 | $3.2\times10^{-5}$ | Zeggini et al.[13] | 1176 (55.5) | 138 (54.1) | 784 (37.0) | 99 (38.8) | 161 (7.6) | 18 (7.1) | 26.1 | 1.04 (0.85-1.27) | 0.73 |
| rs7961581 | TSPAN8, LGR5 | 12 | 5' of gene | C | 1.09 | $3.7\times10^{-5}$ | Zeggini et al.[13] | 1038 (49.0) | 116 (45.5) | 870 (41.1) | 110 (43.1) | 210 (9.9) | 28 (11.0) | 30.5 | 1.08 (0.90-1.31) | 0.4 |

[*] P values and odds ratios for type 2 diabetes associated with individual risk SNPs are estimated from additive pooled logistic-regression models with generalized estimating equations adjusted for age, age squared, and sex.

[†] Genotype frequencies may not add to 100% owing to missing genotypes for some SNPs.

**Table 3**

Risk of Incident Type 2 Diabetes Associated with Genotype Score, Family History of Diabetes, and Phenotypic Risk Factors for Diabetes[*]

| Model | Prediction Model | |
|---|---|---|
| | Without Genotype Score | With Genotype Score |
| **Model 1: adjusted for sex** | | |
| Male sex vs. female sex | | |
| Odds ratio | 1.31 | 1.31 |
| 95% CI | 1.02-1.68 | 1.02-1.69 |
| Genotype score | | |
| Odds ratio | — | 1.12 |
| 95% CI | — | 1.07-1.17 |
| Discriminatory capability | | |
| C statistic | 0.534 | 0.581 |
| 95% CI | 0.502-0.565 | 0.546-0.617 |
| P value for difference | 0.01 | |
| Accuracy of calibration | | |
| Chi-square | | 3.15 |
| P value | | 0.92 |
| **Model 2: adjusted for sex and self-reported family history** | | |
| Male sex vs. female sex | | |
| Odds ratio | 1.35 | 1.35 |
| 95% CI | 1.05-1.74 | 1.05-1.74 |
| Family history of diabetes vs. no family history of diabetes | | |
| Odds ratio | 2.26 | 2.16 |
| 95% CI | 1.72-2.97 | 1.64-2.85 |
| Genotype score | | |
| Odds ratio | — | 1.11 |
| 95% CI | — | 1.06-1.16 |
| Discriminatory capability | | |
| C statistic | 0.595 | 0.615 |
| 95% CI | 0.560-0.630 | 0.579-0.652 |
| P value for difference | 0.11 | |
| Accuracy of calibration | | |
| Chi-square | | 9.57 |
| P value | | 0.30 |
| **Model 3: adjusted for simple clinical risk factors** | | |
| Age, per year | | |
| Odds ratio | 1.05 | 1.05 |
| 95% CI | 1.03-1.06 | 1.03-1.06 |
| Male sex vs. female sex | | |
| Odds ratio | 0.66 | 0.66 |
| 95% CI | 0.48-0.91 | 0.48-0.92 |
| Family history of diabetes vs. no family history of diabetes | | |
| Odds ratio | 1.72 | 1.64 |

| Model | Prediction Model | |
| --- | --- | --- |
| | Without Genotype Score | With Genotype Score |
| 95% CI | 1.24-2.38 | 1.18-2.27 |
| Body-mass index, per unit increase[†] | | |
| Odds ratio | 1.14 | 1.14 |
| 95% CI | 1.10-1.17 | 1.11-1.18 |
| Fasting plasma glucose level, per unit increase | | |
| Odds ratio | 1.14 | 1.13 |
| 95% CI | 1.12-1.16 | 1.11-1.15 |
| Systolic blood pressure, per unit increase | | |
| Odds ratio | 1.01 | 1.01 |
| 95% CI | 1.00-1.02 | 1.00-1.02 |
| High-density lipoprotein cholesterol, per unit increase | | |
| Odds ratio | 0.99 | 0.99 |
| 95% CI | 0.97-1.00 | 0.97-1.00 |
| Fasting triglycerides, per unit increase | | |
| Odds ratio | 1.00 | 1.00 |
| 95% CI | 1.00-1.00 | 1.00-1.00 |
| Genotype score | | |
| Odds ratio | — | 1.11 |
| 95% CI | — | 1.05-1.17 |
| Discriminatory capability | | |
| C statistic | 0.900 | 0.901 |
| 95% CI | 0.880-0.919 | 0.881-0.920 |
| P value for difference | 0.49 | |
| Accuracy of calibration | | |
| Chi-square | | 1.87 |
| P value | | 0.99 |

[*] Odds ratios, 95% confidence intervals (CIs), and C statistics for 255 cases of diabetes identified among 6130 person-examinations were calculated with the use of pooled logistic regression with generalized estimating equations.

[†] The body-mass index is the weight in kilograms divided by the square of the height in meters.

**Table 4**

Reclassification of Risk According to the Genotype Score for Subjects in Whom Diabetes Developed and for Those in Whom Diabetes Did Not Develop[*]

| Model without Genotype Score | Model with Genotype Score | | | |
|---|---|---|---|---|
| | 0 to <2% Risk | 2 to <8% Risk | ≥8% Risk | Total at Risk |
| | *number (percent)* | | | |
| **Model 1: adjusted for sex** | | | | |
| Subjects in whom diabetes developed | | | | |
| 0 to <2% risk | **0** | 0 | 0 | 0 |
| 2 to <8% risk | 1 (0.4) | **242 (94.9)** | 12 (4.7) | 255 |
| ≥8% risk | 0 | 0 | **0** | 0 |
| Total no. | 1 | 242 | 12 | 255 |
| Subjects in whom diabetes did not develop | | | | |
| 0 to <2% risk | **0** | 0 | 0 | 0 |
| 2 to <8% risk | 65 (1.1) | **5731 (97.6)** | 79 (1.3) | 5875 |
| ≥8% risk | 0 | 0 | **0** | 0 |
| Total no. | 65 | 5731 | 79 | 5875 |
| Net reclassification improvement — % | 4.1 | | | |
| P value | 0.004 | | | |
| **Model 2: adjusted for sex and self-reported family history** | | | | |
| Subjects in whom diabetes developed | | | | |
| 0 to <2% risk | **0** | 0 | 0 | 0 |
| 2 to <8% risk | 2 (0.9) | **202 (94.8)** | 9 (4.2) | 213 |
| ≥8% risk | 0 | 17 (40.5) | **25 (59.5)** | 42 |
| Total no. | 2 | 219 | 34 | 255 |
| Subjects in whom diabetes did not develop | | | | |
| 0 to <2% risk | **0** | 0 | 0 | 0 |
| 2 to <8% risk | 252 (4.6) | **5102 (93.8)** | 87 (1.6) | 5441 |
| ≥8% risk | 0 | 218 (50.2) | **216 (49.8)** | 434 |
| Total no. | 252 | 5320 | 303 | 5875 |
| Net reclassification improvement — % | 2.60 | | | |
| P value | 0.22 | | | |
| **Model 3: adjusted for simple clinical risk factors** | | | | |
| Subjects in whom diabetes developed | | | | |
| 0 to <2% risk | **20 (83.3)** | 4 (16.7) | 0 | 24 |
| 2 to <8% risk | 4 (7.4) | **45 (83.3)** | 5 (9.2) | 54 |
| ≥8% risk | 0 | 2 (1.1) | **175 (98.9)** | 177 |
| Total no. | 24 | 51 | 180 | 255 |
| Subjects in whom diabetes did not develop | | | | |
| 0 to <2% risk | **3924 (97.5)** | 101 (2.5) | 0 | 4025 |
| 2 to <8% risk | 156 (12.2) | **1063 (83.4)** | 56 (4.4) | 1275 |
| ≥8% risk | 0 | 57 (9.9) | **518 (90.1)** | 575 |
| Total no. | 4080 | 1221 | 574 | 5875 |
| Net reclassification improvement — % | 2.13 | | | |
| P value | 0.17 | | | |

\* Net reclassification improvement for 255 cases of diabetes among 6130 person-examinations was based on pooled logistic-regression models with generalized estimating equations including or not including the genotype score. Net reclassification improvement is better if more people in whom diabetes develops are reclassified as being at higher risk when the genotype score is added to the model and more people who remain free of diabetes are classified as being at lower risk when the genetic score is added. The net reclassification improvement is worse when there is erroneous reclassification — for example, if many people in whom diabetes develops are classified as being at lower risk when the genetic score is added to the model. The number and proportion of people who were not reclassified by the genotype score are indicated by the bold font.