Predicting Type 2 Diabetes Based on Polymorphisms From Genome-Wide Association Studies

A Population-Based Study

Mandy van Hoek,[1,2] Abbas Dehghan,[2] Jacqueline C.M. Witteman,[2] Cornelia M. van Duijn,[2,3] André G. Uitterlinden,[1,2] Ben A. Oostra,[2,3] Albert Hofman,[2] Eric J.G. Sijbrands,[1] and A. Cecile J.W. Janssens[4]

OBJECTIVE—Prediction of type 2 diabetes based on genetic testing might improve identification of high-risk subjects. Genome-wide association (GWA) studies identified multiple new genetic variants that associate with type 2 diabetes. The predictive value of genetic testing for prediction of type 2 diabetes in the general population is unclear.

RESEARCH DESIGN AND METHODS—We investigated 18 polymorphisms from recent GWA studies on type 2 diabetes in the Rotterdam Study, a prospective, population-based study among homogeneous Caucasian individuals of 55 years and older (genotyped subjects, n = 6,544; prevalent cases, n = 686; incident cases during follow-up, n = 601; mean follow-up 10.6 years). The predictive value of these polymorphisms was examined alone and in addition to clinical characteristics using logistic and Cox regression analyses. The discriminative accuracy of the prediction models was assessed by the area under the receiver operating characteristic curves (AUCs).

RESULTS—Of the 18 polymorphisms, the *ADAMTS9*, *CDKAL1*, *CDKN2A/B-rs1412829*, *FTO*, *IGF2BP2*, *JAZF1*, *SLC30A8*, *TCF7L2*, and *WFS1* variants were associated with type 2 diabetes risk in our population. The AUC was 0.60 (95% CI 0.57–0.63) for prediction based on the genetic polymorphisms; 0.66 (0.63–0.68) for age, sex, and BMI; and 0.68 (0.66–0.71) for the genetic polymorphisms and clinical characteristics combined.

CONCLUSIONS—We showed that 9 of 18 well-established genetic risk variants were associated with type 2 diabetes in a population-based study. Combining genetic variants has low predictive value for future type 2 diabetes at a population-based level. The genetic polymorphisms only marginally improved the prediction of type 2 diabetes beyond clinical characteristics.
***Diabetes* 57:3122–3128, 2008**

Type 2 diabetes is a multifactorial disease caused by a complex interplay of multiple genetic variants and many environmental factors. With the recent genome-wide association (GWA) studies, the number of replicated common genetic variants associated with type 2 diabetes has rapidly increased (1–7). A total of 18 polymorphisms have been firmly replicated (1–7). It is unclear whether and how the currently known genetic variants can be used in practice, because the combined effect of these variants has not been investigated in a population-based study. Particularly, because most GWA studies were enriched for patients with a positive family history and early onset of the disease, association of these variants to type 2 diabetes risk in the general population, including elderly individuals, remains to be determined.

Because complex diseases are caused by multiple genetic variants, predictive testing based on a single genetic marker will be of limited value (8,9). Simulation studies suggest that the predictive value could be improved by combining multiple common low-risk variants (10–13). Several empirical studies on the predictive value of genetic polymorphisms have been conducted before the recent GWA data were available (14–16). In a case-control study, Weedon et al. (16) showed that combining the information of three polymorphisms improved disease prediction, albeit to a limited extent. Vaxillaire et al. (15) investigated 19 polymorphisms and found that the predictive value was low compared with clinical characteristics.

Genetic variants associated with risk of type 2 diabetes could potentially be useful for the prediction, prevention, and early treatment of the disease. We investigated whether combining the currently known and well-replicated genetic variants predicts type 2 diabetes in the Rotterdam Study, a prospective population-based follow-up study. We investigated whether these genetic variants improve prediction beyond clinical characteristics.

RESEARCH DESIGN AND METHODS

The design and data collection of the Rotterdam Study have been described previously (17). In short, the Rotterdam Study is a prospective, population-based, cohort study among 7,983 inhabitants of a Rotterdam suburb, designed to investigate determinants of chronic diseases. Participants were aged 55 years and older. Baseline examinations took place from 1990 until 1993. Follow-up examinations were performed in 1993–1994, 1997–1999, and 2002–2004. Between these exams, continuous surveillance on major disease outcomes was conducted. Information on vital status was obtained from municipal health authorities. The medical ethics committee of the Erasmus

From the [1]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands; the [2]Department of Epidemiology and Biostatistics, Erasmus University Medical Center, Rotterdam, the Netherlands; the [3]Department of Clinical Genetics, Genetic Epidemiology Unit, Erasmus University Medical Center, Rotterdam, the Netherlands; and the [4]Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands.

Corresponding author: Dr. Eric J.G. Sijbrands, e.sijbrands@erasmusmc.nl.

Received 27 March 2008 and accepted 1 August 2008.

Published ahead of print at http://diabetes.diabetesjournals.org on 11 August 2008. DOI: 10.2337/db08-0425.

Medical Center approved the study protocol, and all participants gave their written informed consent.

Data collection. At baseline, prevalent cases of diabetes were diagnosed by a nonfasting or postload glucose level (after oral glucose tolerance testing) ≥11.1 mmol/l and/or treatment with antidiabetic medication (oral medication or insulin) and the diagnosis of diabetes as registered by a general practitioner. During follow-up, diabetes was diagnosed at fasting plasma glucose levels ≥7.0 mmol/l and/or a nonfasting plasma glucose levels ≥11.0 mmol/l and/or treatment with antidiabetic medication (oral medication or insulin) (18,19) and the diagnosis of diabetes as registered by a general practitioner. Patients registered in general practitioners' records as type 1 diabetic were excluded from the present analyses (n = 15).

The *CDKAL1* rs7754840, *CDKN2AB* rs10811661, *FTO* rs8050136, *HHEX* rs1111875, *IGF2BP2* rs4402960, *KCNJ11* rs5219, *PPARG* rs1801282, *SLC30A8* rs13266634, and *TCF7L2* rs7903146 polymorphisms were genotyped by means of TaqMan allelic discrimination assays. DNA material was available for 6,544 of the 7,983 participants for the TaqMan analyses. The assays were designed and optimalized by Applied Biosystems (Foster City, CA; http://store.appliedbiosystems.com). Genotypes were determined in 2-ng genomic DNA. Reactions were performed on the TaqMan Prism 7900HT platform. The analyses were performed as described previously (20). Assays were run on 90 blood bank samples to test for adequate cluster separation. A total of 325 samples were genotyped in duplo. Success rates for TaqMan genotyping ranged from 93.2 to 96.7%, with the exception of 86.1% for IGF2BP2 and 87.4% for HHEX. TaqMan duplicate error rates for the HHEX and IGF2BP2 polymorphisms were 1.2 and 0.6%

The *ADAMTS*9 rs4411878 (proxy for rs4607103, r^2 = 0.95), *CDC123-CAMK1D* rs11257622 (proxy for rs12779790, r^2 = 0.83), *CDKN2A/B* rs1412829 (proxy for rs564398, r^2 = 0.97), *JAZF1* rs1635852 (proxy for rs864745, r^2 = 0.97), *NOTCH2* rs1493694 (proxy for rs10923931, r^2 = 1.0), *TCF2* rs4430796, *THADA* rs7578597, *TSPAN8-LGR5* rs1353362 (proxy for rs7961581, r^2 = 0.96), and *WFS1* rs10012946 (proxy for rs10010131, r^2 = 1.0) genotypes were derived from the genotype data of the version 3 Illumina Infinium II HumanHap550 SNP chip array. From a total of 6,449 subjects, there was sufficient DNA material for the array. Samples with a call rate <97.5% (n = 209), excess autosomal heterozygosity >0.336 (approximate false discovery rate [FDR] <0.1% [n = 21]), or mismatch between called and phenotypic sex (n = 36) or with outliers identified by the identity-by-state (IBS) clustering analysis with >3 SDs from population mean (n = 102) or IBS probabilities >97% (n = 129) were excluded from the analysis; in total, 5,974 samples remained for analyses.

The availability of Illumina 550K array data enabled us to compare genotype calls between TaqMan and Illumina data for the FTO, HHEX, IGF2BP2, SLC30A8, TCF7L2, and CDKAL1 polymorphisms as well. Concordance rates ranged between 98.6 and 99.7%. To increase success rates, we merged the data and deleted pairs that were not concordant. The success rates for the polymorphisms increased to 98.4–99.4%.

Statistical analyses. Associations of individual polymorphisms were investigated using Cox proportional hazards models for the prediction of incident type 2 diabetes and logistic regression analyses for the prediction of prevalent and incident type 2 diabetes together. Analyses were performed crude and adjusted for age, sex, and BMI. We also applied Cox proportional hazards models and logistic regression analyses to investigate the combined predictive value of *1*) the 18 polymorphisms (all polymorphisms included as separate independent categorical variables); *2*) the risk allele score based on the 18 polymorphisms (assuming all effect sizes of equal weight); *3*) age, sex, and BMI; and *4*) age, sex, and BMI and all polymorphisms on type 2 diabetes risk. The risk allele score was calculated by summing up the number of risk alleles for each participant with complete genotype information, with risk alleles being the alleles associated with increased risk of type 2 diabetes (1–5). The risk allele score assumes that all genetic variants have the same effect, i.e., minor differences in effects size are ignored. The association between the risk allele score and the predicted probabilities was quantified by the Spearman correlation coefficient.

The discriminative accuracy was evaluated by the area under the receiver operating characteristic (ROC) curves (AUCs). The AUC can range from 0.5 (total lack of discrimination) to 1.0 (perfect discrimination). AUCs were calculated for the predicted risks of the logistic regression model, the risk allele score, and the linear predictor values of the Cox proportional hazards models. AUCs were compared with Analyze-it version 2.11 (www.analyze-it.com), which uses the method of Hanley and McNeil for ROC curve analyses (21,22).

The analyses were repeated for subgroups for age (cutoff 70 years of age) and BMI (cutoff 26 kg/m^2). All analyses were performed with SPSS software version 12.0.1.

Simulation analyses. A simulation analysis was performed to quantify the expected AUC for prediction of incident type 2 diabetes based on the odds ratios (ORs) of the investigated polymorphisms in literature (OR 1.09 for *ADAMTS9*, 1.11 for *CDC123-CAMK1D*, 1.12 for *CDKAL1*, 1.12 for *CDKN2A/B rs1412829*, 1.20 for *CDKN2A/B rs10811661*, 1.17 for *FTO*, 1.13 for *HHEX*, 1.14 for *IGF2BP2*, 1.10 for *JAZF1*, 1.14 for *KCNJ11*, 1.13 for *NOTCH2*, 1.14 for *PPARG*, 1.12 for *SLC30A8*, 1.10 for *TCF2*, 1.37 for *TCF7L2*, 1.15 for *THADA*, 1.09 for *TSPAN8*, and 1.11 for *WFS1*) (1,2,4,6,7,23). The method of simulation has been described in detail previously (10). In brief, we simulated genetic profiles and type 2 diabetes status for 100,000 individuals, of whom 10.3% were supposed to have incident type 2 diabetes, as observed in our population. Genetic profiles were constructed from the polymorphisms based on observed allele frequencies. Under the assumption that each polymorphism has two alleles and that allele proportions were in Hardy-Weinberg equilibrium, genotype frequencies for the single polymorphisms were calculated. Assuming that the polymorphisms segregate independently, for each individual, a genotype was randomly assigned. Disease risks associated with the genetic profiles were modeled using Bayes' theorem. The likelihood ratio of the genetic profile was calculated by multiplying the likelihood ratios of the single genotypes. The OR of the heterozygous genotypes compared with the homozygous nonrisk genotypes were derived from the three large GWA studies (1,2,4). Finally, disease status was modeled by a procedure that compares disease risk of each subject to a randomly drawn value between 0 and 1 from a uniform distribution. This procedure ensures that for each genomic profile, the percentage of people who will develop the disease equals the disease risk associated with that profile, when the subgroup of individuals with that profile is sufficiently large. The simulation was repeated 10 times to obtain a robust estimate of the AUC. The AUC was obtained as the c-statistic by the function somers2, which is available in the Hmisc library of R software (version 2.5.1; www.R-project.org, accessed December 2007).

RESULTS

Baseline characteristics. A total of 6,544 participants were successfully genotyped for at least one polymorphism. Complete genotype information on all polymorphisms was present in 5,297 subjects (of whom 490 were incident cases and 545 were prevalent cases). Age (P = 0.11), sex (P = 0.22), BMI (P = 0.30), and presence of type 2 diabetes (P = 0.20) were not significantly different between successfully genotyped individuals or individuals with one or more missing genotypes. General characteristics of the population are shown in Table 1. Individuals with type 2 diabetes had higher BMI, higher waist circumference, more often hypertension, and lower HDL cholesterol compared with individuals without type 2 diabetes. All polymorphisms were in Hardy Weinberg equilibrium in the total population and in individuals without type 2 diabetes (highest χ^2 3.58, 2 d.f., P = 0.06 for *PPARG rs1801282*).

Individual effects of clinical characteristics and polymorphisms on type 2 diabetes risk. Table 2 shows the effect of each polymorphism on type 2 diabetes risk (prevalent and incident type 2 diabetes). The minor alleles of the *CDKAL1*, *FTO*, *IGF2BP2*, and *TCF7L2* variants were associated with increased risk of type 2 diabetes. The *ADAMTS*9, *CDKN2A/B* rs1412829, *JAZF*, *SLC30A8*, and *WFS1* minor alleles decreased type 2 diabetes risk. Cox regression analyses restricted to incident type 2 diabetes results gave similar results. Adjustment for age, sex, and BMI did not materially change the results, except for the FTO polymorphism, for which the effect on type 2 diabetes risk disappeared after adjustment for BMI.

In a Cox regression analysis, age (hazard ratio [HR] 1.02 [95% CI 1.01–1.03]), sex (0.67 [0.57–0.79]), and BMI (1.14 [1.12–1.16]) affected prospective type 2 diabetes risk.

Risk allele score and risk of type 2 diabetes. Figure 1 shows the ORs associated with increasing risk allele scores compared with the reference group (0–12 risk alleles) in a logistic regression model. Individuals carrying 21 risk alleles or more (14.4% of the population) had

TABLE 1
General characteristics of genotyped participants at study baseline by type 2 diabetes status

	All participants	Subjects without type 2 diabetes	Incident case subjects with type 2 diabetes	Prevalent cases with type 2 diabetes
n	6,544	5,221	601	686
Age (years)	69.5 ± 0.11	69.0 ± 0.13	68.2 ± 0.32*	73.6 ± 0.35†
Men (%)	40.7	40.4	44.3	39.8
BMI (kg/m^2)	26.3 ± 0.05	26.0 ± 0.05	28.0 ± 0.15†	26.8 ± 0.15†
Waist circumference (cm)	90.5 ± 0.14	89.6 ± 0.16	94.7 ± 0.45†	93.8 ± 0.46†
Systolic blood pressure (mmHg)	139.3 ± 0.3	137.9 ± 0.31	143.5 ± 0.85†	146.8 ± 0.93†
Diastolic blood pressure (mmHg)	73.7 ± 0.1	73.6 ± 0.2	75.5 ± 0.5†	73.1 ± 0.5
Hypertension (%)	33.4	30.5	46.9†	52.9†
Total cholesterol (mmol/l)	6.6 ± 0.02	6.6 ± 0.02	6.6 ± 0.05	6.5 ± 0.05
HDL cholesterol (mmol/l)	1.34 ± 0.005	1.37 ± 0.005	1.25 ± 0.01†	1.25 ± 0.01†
Current smoking (%)	22.1	22.5	25.5	22.1
Former smoking (%)	40.7	42.0	42.8	39.0

Continuous variables are expressed as means ± SE. Type 2 diabetes status was missing for 36 individuals. *$P < 0.05$, †$P < 0.001$ for comparison with subjects without type 2 diabetes. Comparisons between groups were performed using ANOVA for continuous variables and χ^2 test for categorical variables.

significantly higher type 2 diabetes risk (7.2% of the population carried 21 alleles, OR 1.90 [1.07–3.40]; 4.4% had 22 alleles, 2.11 [1.15–3.86]; 2.0% had 23 alleles, 2.11 [1.07–4.18]; and 1.8% had 24–32 alleles, 2.10 [1.04–4.22]) compared with the reference group of 0–12 alleles ($n = 109$; 2.0% of the population). In a Cox regression analysis on incident cases of diabetes, this figure was similar (data not shown). The per-allele HR was 1.04 (95% CI 1.02–1.07) ($P = 0.001$).

Risk allele score. Figure 2 shows the predicted type 2 diabetes risks from the logistic regression model that included all 18 genetic polymorphisms in relation to the risk allele score. The Spearman correlation coefficient was 0.60, indicating a wide range of predicted risks for each value of the risk allele scores. When analyzing only incident type 2 diabetes, this figure was similar (Spearmen rho 0.59; figure not shown).

Analyses of discriminative accuracy. Figure 3 shows the ROC curves for the prediction of incident type 2 diabetes based on the genetic polymorphisms, clinical characteristics, and both. The AUC was 0.60 (95% CI 0.57–0.63) for prediction based on the genetic polymorphisms; 0.66 (0.63–0.68) for age, sex, and BMI; and 0.68 (0.66–0.71) for the genetic polymorphisms and clinical characteristics combined. The difference between the AUCs for clinical characteristics with and without the genetic polymorphisms was significant ($P < 0.0001$). The AUC of the risk allele score was 0.56 (0.53–0.59). When including only the significantly associated genetic variants of the current study (*ADAMTS9*, *CDKAL1*, *CDKN1412829*, *FTO*, *IGFBP2*, *JAZF1*, *TCF7L2*, *SLC30A8*, and *WFS1*), the AUC was 0.58 (0.56–0.61). Combining the *KCNJ11*, *PPARG*, and *TCF7L2* variants resulted in an AUC of 0.53 (0.50–0.55). Based on the simulation study, the expected AUC for all genetic polymorphisms using the effect sizes described in literature was 0.57. When combining incident and prevalent type 2 diabetes cases, the AUC of all polymorphisms was 0.60 (0.58–0.62).

In subgroup analyses, the AUC of the all polymorphisms was 0.62 (95% CI 0.58–0.67) in the low BMI group and 0.59 (0.56–0.62) in the high BMI subgroup. The AUC was 0.61 (0.59–0.65) in the low age-group and 0.63 (0.58–0.67) in the high age-groups.

DISCUSSION

We investigated the predictive value of 18 type 2 diabetes risk polymorphisms from the recent GWA studies for the prediction of type 2 diabetes in a large, prospective, population-based study of elderly individuals. Our study shows that combining information of these 18 well-replicated variants has relatively low discriminative accuracy for the prediction of type 2 diabetes in a general population (AUC 0.60). The 18 genetic variants identified to date did not substantially improve the discriminative accuracy of disease prediction based on clinical characteristics.

In line with the results of the GWA studies (1–5,24), *ADAMTS9*, *CDKAL1*, *CDKN2A/B*, *FTO*, *IGFBP2*, *JAZF1*, *SLC30A8*, *TCF7L2*, and *WFS1* were associated type 2 diabetes risk in our population. Some of these effects were slightly stronger than the effects described previously. In contrast to most previous studies, the *KCNJ11* polymorphism was not associated. The ORs of the other polymorphisms were similar to previously published results but not statistically significant, which may be explained by the smaller number of type 2 diabetes cases and therefore smaller power to reach significance in our prospective study.

A risk allele score, obtained by counting the number of risk alleles, can be used as a simple proxy of the combined effect of multiple polymorphisms. Risk allele scores ignore the effect sizes of the individual genetic variants, but a previous simulation study has shown that this has limited impact on the discriminative accuracy (11). In contrast, we found a wide range of predicted risks for each value of the risk allele score, suggesting that differences in the variants carried result in substantial differences in actual disease risks. The risk allele score associated with modest increases in disease risk and the AUC for prediction was 0.56. When taking effect size differences between polymorphisms into account, the AUC was 0.60, showing that the risk allele score predicted less accurately. Other empirical studies have not investigated the differences in diagnostic accuracy between simple risk allele scores and predicted risks obtained from regression models (14–16,25).

The discriminative accuracy of predictive genetic testing in complex diseases depends on the number of genes involved, the risk allele frequencies, and the size of the associated risks (10). The maximum discriminative accu-

TABLE 2
Individual effects of nine polymorphisms on type 2 diabetes risk with and without adjustment for covariates

Gene variant	Risk allele	Genotype (*n*)	Percent case subjects		Percent control subjects	All case subjects	Incident type 2 diabetes	
			Prevalent	Incident		OR (95% CI)	Crude HR (95% CI)	Adjusted HR (95% CI)*
ADAMTS9	C	CC (3,450)	61.9	61.1	57.4	1.0	1.0	1.0
rs4411878†		CT (2,159)	32.5	34.7	37.1	0.84 (0.74–0.97)	0.88 (0.73–1.05)	0.87 (0.73–1.04)
		TT (321)	5.6	4.2	5.5	0.84 (0.62–1.13)	0.68 (0.45–1.04)	0.68 (0.45–1.04)
CDC123/	C	TT (3,952)	64.4	67.4	67.9	1.0	1.0	1.0
CAMK1D		TC (1,758)	31.5	28.8	29.6	1.04 (0.90–1.20)	0.98 (0.82–1.19)	0.98 (0.81–1.18)
rs11257622‡		CC (213)	4.2	3.8	3.5	1.18 (0.84–1.64)	1.07 (0.69–1.67)	1.05 (0.68–1.63)
CDKAL1	C	GG (3,097)	47.1	45.4	48.7	1.0	1.0	1.0
rs7754840		GC (2,692)	41.1	43.0	41.9	1.05 (0.93–1.20)	1.09 (0.92–1.29)	1.10 (0.93–1.31)
		CC (633)	11.8	11.6	9.4	1.31 (1.07–1.61)	1.31 (1.001–1.72)	1.38 (1.06–1.80)
CDKN2A/B	A	AA (1,915)	31.3	35.2	32.1	1.0	1.0	1.0
rs1412829§		AG (2,910)	47.7	50.4	49.2	0.97 (0.84–1.12)	0.94 (0.78–1.13)	0.89 (0.74–1.08)
		GG (1,098)	21.0	14.5	18.7	0.93 (0.77–1.12)	0.72 (0.56–0.94)	0.70 (0.54–0.92)
CDKN2A/B	T	TT (4,131)	72.0	63.1	66.0	1.0	1.0	1.0
rs10811661		TC (1,865)	24.7	34.5	30.1	0.95 (0.83–1.09)	1.17 (0.99–1.39)	1.13 (0.95–1.34)
		CC (232)	3.2	2.4	3.9	0.70 (0.49–1.02)	0.64 (0.37–1.08)	0.68 (0.40–1.17)
FTO	A	CC (2,526)	38.3	38.4	39.8	1.0	1.0	1.0
rs8050136		CA (2,944)	44.0	46.3	46.3	1.01 (0.88–1.16)	1.03 (0.86–1.23)	1.00 (0.84–1.20)
		AA (927)	17.7	15.3	14.0	1.23 (1.02–1.47)	1.12 (0.88–1.43)	1.06 (0.83–1.36)
HHEX	C	CC (2,235)	36.1	36.3	34.8	1.0	1.0	1.0
rs1111875		CT (3,097)	50.1	46.1	48.4	0.96 (0.84–1.10)	0.91 (0.76–1.09)	0.93 (0.78–1.12)
		TT (1,052)	13.7	17.6	16.8	0.89 (0.74–1.07)	1.01 (0.80–1.27)	1.01 (0.79–1.28)
IGF2BP2	T	GG (3,101)	46.9	47.7	49.4	1.0	1.0	1.0
rs4402960		GT (2,650)	41.8	41.5	42.0	1.04 (0.91–1.18)	1.01 (0.85–1.19)	1.01 (0.85–1.20)
		TT (575)	11.3	10.8	8.6	1.35 (1.09–1.66)	1.24 (0.95–1.63)	1.23 (0.94–1.62)
JAZF1	T	TT (1,646)	31.0	29.8	27.1	1.0	1.0	1.0
rs1635852¶		TC (2,927)	46.5	48.6	49.8	0.85 (0.73–0.98)	0.88 (0.73–1.07)	0.85 (0.70–1.04)
		CC (1,357)	22.5	21.6	23.1	0.85 (0.71–1.02)	0.86 (0.68–1.09)	0.84 (0.66–1.06)
KCNJ11	G	AA (2,394)	37.7	39.9	39.2	1.0	1.0	1.0
rs5219		AG (2,925)	48.3	46.6	47.8	1.00 (0.88–1.15)	0.97 (0.81–1.16)	0.97 (0.81–1.16)
		GG (807)	14.0	13.5	13.0	1.07 (0.88–1.30)	1.02 (0.79–1.32)	1.02 (0.79–1.32)
NOTCH2	T	CC (4,670)	77.8	79.0	78.8	1.0	1.0	1.0
rs1493692‖		CT (1,168)	20.7	19.7	19.6	1.04 (0.89–1.22)	1.03 (0.84–1.27)	1.07 (0.86–1.32)
		TT (92)	1.4	1.3	1.6	0.86 (0.50–1.49)	0.75 (0.35–1.57)	0.74 (0.35–1.56)
PPARG	C	CC (4,888)	79.7	76.9	77.1	1.0	1.0	1.0
rs1801282		CG (1,322)	19.1	21.7	21.1	0.95 (0.81–1.10)	1.03 (0.84–1.25)	0.99 (0.81–1.21)
		GG (108)	1.2	1.4	1.8	0.70 (0.41–1.20)	0.76 (0.38–1.54)	0.66 (0.33–1.34)
SLC30A8	C	CC (3,176)	49.8	54.1	48.8	1.0	1.0	1.0
rs13266634		CT (2,709)	43.7	38.2	42.4	0.91 (0.80–1.04)	0.82 (0.69–0.97)	0.84 (0.70–0.99)
		TT (546)	6.6	7.7	8.8	0.75 (0.59–0.96)	0.80 (0.59–1.09)	0.84 (0.62–1.14)
TCF2	G	AA (1,560)	25.9	24.1	26.8	1.0	1.0	1.0
rs4430796		AG (2,924)	48.1	50.7	49.6	1.06 (0.91–1.24)	1.11 (0.90–1.36)	1.14 (0.92–1.40)
		GG (1,417)	26.1	25.2	23.6	1.16 (0.97–1.39)	1.16 (0.91–1.47)	1.16 (0.91–1.48)
TCF7L2	T	CC (3,292)	43.7	47.5	52.6	1.0	1.0	1.0
rs7903146		CT (2,587)	42.4	42.3	39.7	1.23 (1.08–1.41)	1.19 (1.01–1.41)	1.20 (1.01–1.42)
		TT (554)	13.9	10.2	7.7	1.82 (1.48–2.24)	1.48 (1.12–1.95)	1.62 (1.22–2.14)
THADA	T	TT (4,608)	79.9	78.8	77.3	1.0	1.0	1.0
Rs7578597		TC (1,227)	19.0	19.0	21.1	0.88 (0.75–1.03)	0.91 (0.73–1.12)	0.90 (0.72–1.12)
		CC (94)	1.1	2.2	1.6	1.01 (0.60–1.67)	1.32 (0.74–2.34)	1.51 (0.85–2.67)
TSPAN8/ LGR5	C	TT (3,018)	50.4	48.3	51.4	1.0	1.0	1.0
*Rs1353362***		TC (2,409)	40.3	41.8	40.6	1.05 (0.92–1.20)	1.10 (0.92–1.31)	1.08 (0.90–1.29)
		CC (493)	9.3	9.9	8.0	1.25 (0.99–1.57)	1.29 (0.96–1.73)	1.21 (0.90–1.62)
WFS1	C	CC (2,179)	40.7	40.4	35.8	1.0	1.0	1.0
Rs1412829††		CT (2,801)	44.6	45.7	47.8	0.83 (0.73–0.96)	0.86 (0.72–1.04)	0.87 (0.73–1.05)
		TT (949)	14.7	13.9	16.4	0.77 (0.63–0.93)	0.77 (0.59–1.00)	0.76 (0.58–0.99)

*HR adjusted for age, sex, and BMI; †proxy for rs4607103, $r^2 = 0.95$; ‡proxy for rs12779790, $r^2 = 0.83$; §proxy for rs564398, $r^2 = 0.97$; ¶proxy for rs864745, $r^2 = 0.97$; ‖proxy for rs10923931, $r^2 = 1.0$; **proxy for rs7961581, $r^2 = 0.96$; ††proxy for rs10010131, $r^2 = 1.0$.

racy is determined by the heritability of the disease (10). Based on previously published effect sizes for the 18 polymorphisms, we predicted that the AUC would be 0.57; and based on our empirical data, we found that it was 0.60. This difference is explained by the fact that some polymorphisms had a slightly larger effect in our population than

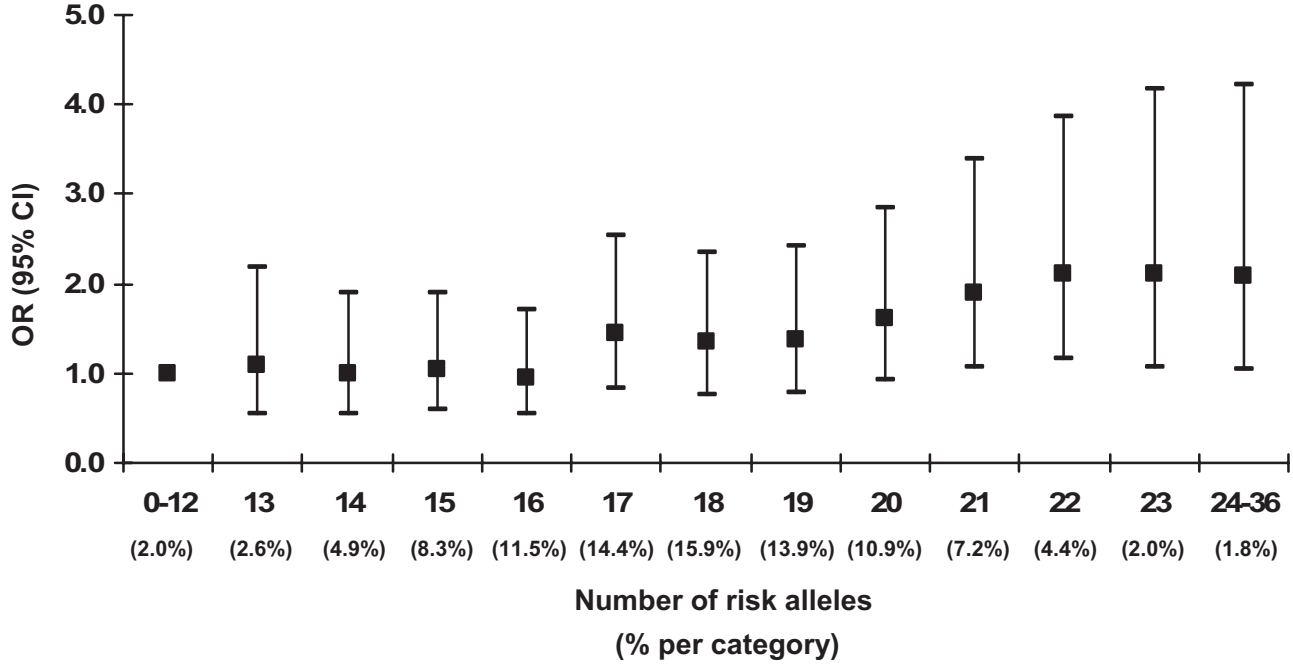

FIG. 1. Odds ratios for type 2 diabetes according to the number of risk alleles carried.

described in the literature (1–4,6,7,16,23). Nonetheless, the discriminative accuracy of all known replicated type 2 diabetes susceptibility variants to date was rather low. However, our analysis was based on 18 common variants with relatively small effects. The heritability of type 2 diabetes is estimated to range from 30 to 70% depending on the population investigated (26), and many common variants with small effects or fewer rare variants with stronger effects are still to be discovered. These may further improve the discriminative accuracy of predictive genetic testing for type 2 diabetes.

Several previous studies have investigated the predictive value of multiple genetic variants in type 2 diabetes, either alone or in addition to clinical characteristics. An overview of the studies performed so far in Caucasian populations and the number of genes investigated is provided in Table 3. Weedon et al. (16) investigated three variants that were also in our study and reported an AUC of 0.58, whereas we found an AUC of 0.53 for the same polymorphisms. The population of Weedon et al. consisted in large part of patients who had early onset of type 2 diabetes or a positive family history of type 2 diabetes, whereas our population included elderly subjects from the general population. The percentage of variance of the disease explained by genetic factors is expected to be higher in populations with a positive family history for the

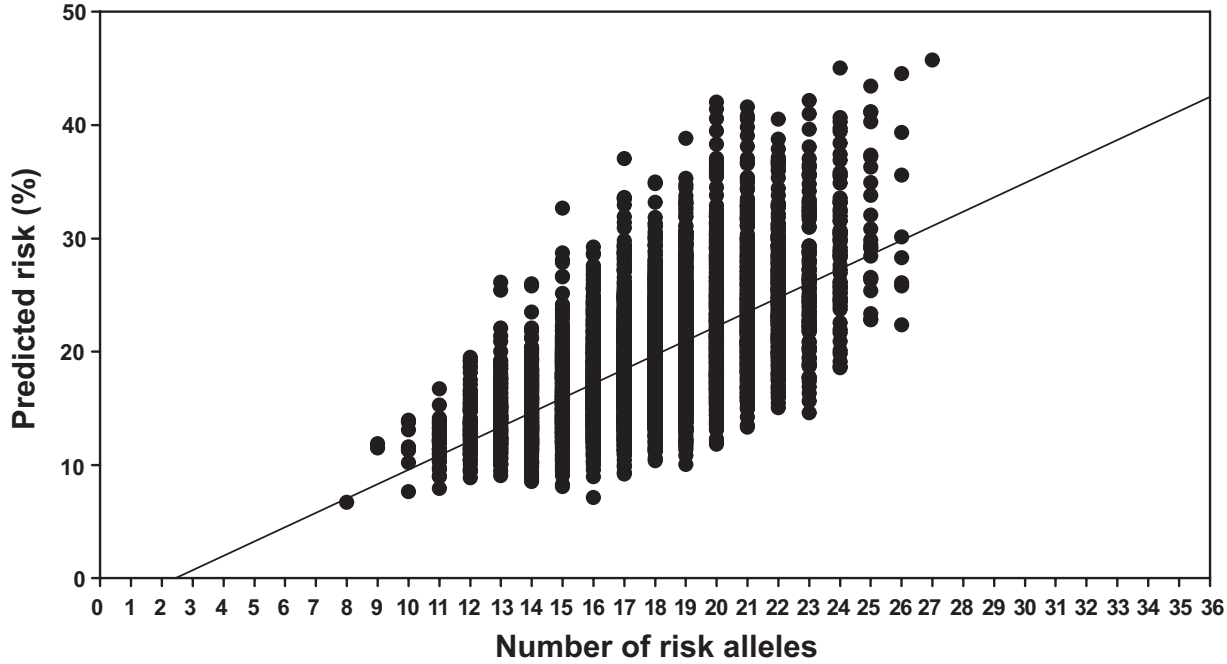

FIG. 2. Correlation of predicted type 2 diabetes risks with the risk allele score. Predicted risks of type 2 diabetes were obtained from the logistic regression model with 18 genetic polymorphisms as independent categorical variables.

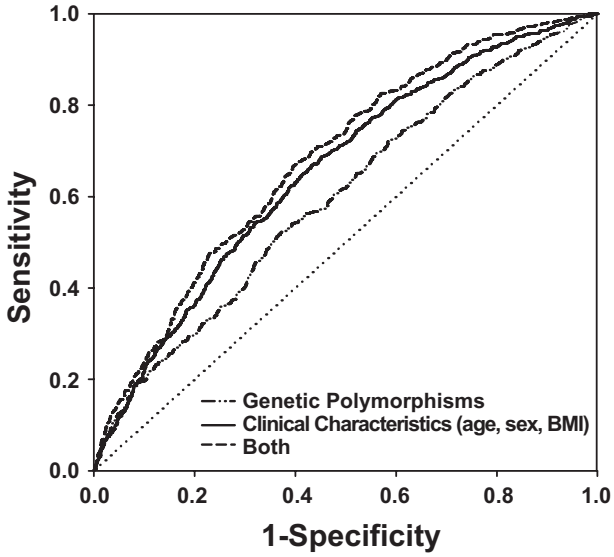

FIG. 3. ROC curves for prediction of incident type 2 diabetes based on 18 genetic polymorphisms, clinical characteristics (age, sex, and BMI), and both.

disease, and this may lead to a higher diagnostic accuracy for genetic variants. Recently, Cauchi et al. (25) investigated 15 genetic variants that were associated in GWA analyses in a French case-control study. The AUC of the genetic variants together with age, sex, and BMI was 0.86. Unfortunately, the AUC was not calculated for genes and clinical characteristics separately to assess the additive value of genetic information. Some of the included variants were specifically identified in this case-control study and had relatively large effects on type 2 diabetes risk compared with effects found in meta-analyzed GWA studies (1,2,4,6). It is therefore difficult to generalize these findings to other populations, and we expect that our population-based prospective study yielded more realistic estimates. Two other studies investigated the improvement of the discriminative accuracy by adding genetic test results to clinical characteristics (14,15). Even though we included the 18 firmly replicated polymorphisms to date and they predominantly tested other genetic variants, our findings were similar, demonstrating no substantial added value of genetic information beyond clinical characteristics (14,15,27).

We can only speculate on the reasons why these genetic variants have little added value beyond clinical characteristics. First, the effects of the genetic variants on type 2 diabetes risk could be exerted through clinical characteristics such as BMI, which implies that including both genes and intermediate factors in the regression model will reduce the effect of the gene. However, adjustment for clinical characteristics did not substantially change the effect of the genetic variants on type 2 diabetes risk (Table 2). Second, the effects of age, sex, and BMI on type 2 diabetes risk in our population may outweigh the contribution of the genetic variants. Such an effect was illustrated in an earlier study, which showed that a genetic predisposition became apparent in subjects with less other risk factors (28). In our elderly population, one may expect that nongenetic risk factors are more prevalent compared with younger populations. However, the AUC for the genetic variants was higher than expected from the simulation analyses. This makes an underestimation of the contribution of the genetic variants in our population unlikely.

Obvious strengths of our study are the large size of the study population, the population-based design, and the long period of follow-up. Despite these advantages, the number of incident type 2 diabetes cases was still relatively low to demonstrate statistically significant effects of low-risk susceptibility genes. Furthermore, we had insufficient statistical power to formally investigate gene-gene and gene-environment interactions. In age and BMI subgroup analyses, the estimates for prediction based on the genetic variants were similar and showed overlapping CIs. However, because of smaller numbers of cases in the subgroups, these results should be interpreted with caution. Cauchi et al. (25) reported gene-gene interactions of recently discovered loci but did not report on the effects of these interactions on the AUC. Earlier studies found no evidence for strong gene-gene interactions (15,16). Taking into account these interactions may further improve the discriminative accuracy.

In conclusion, we showed that 9 of 18 currently well-established genetic risk variants were associated with type 2 diabetes in a population-based study. The currently known and replicated genetic variants found in GWA studies contributed modestly to the prediction of type 2 diabetes in a population-based setting and marginally improved the risk prediction when added to clinical characteristics. Future research should aim at identifying and replicating new genetic susceptibility variants and gene-gene and gene-environment interactions to approach levels of discriminative accuracy that enable the identification of at-risk groups.

TABLE 3
Overview of diagnostic accuracies obtained from earlier empirical studies on genetic risk variants and type 2 diabetes

Study	Genetic variants (*n*)	AUC genetic variants	AUC clinical characteristics	AUC combined
Weedon et al. (16)	3	0.58	NR	NR
Lyssenko et al. (14,27)	2	NR	0.68*	0.68
Vaxillaire et al. (15)	3	0.56	0.82†	0.84
Cauchi et al. (25)	15	NR	NR†	0.86

NR, not reported. *Clinical characteristics: fasting plasma glucose and BMI. †Clinical characteristics: age, sex, and BMI.

ACKNOWLEDGMENTS

The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University Rotterdam; the Netherlands Organization for Scientific Research; the Netherlands Organization for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly; the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission; and the Municipality of Rotterdam. This study was further supported by the Centre for Medical Systems Biology in the framework of the Netherlands Genomics Initiative. The genome-wide association database of the Rotterdam Study was made possible through funding from the Dutch Research Organisation (nr.175.010.2005.011).

We thank Dr. Fernando Rivadeneira for making the Illumina 550K data from the Rotterdam Study available.

REFERENCES

1. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson BK, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336, 2007
2. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345, 2007
3. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885, 2007
4. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT: Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1336–1341, 2007
5. Welcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678, 2007
6. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645, 2008
7. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So WY, Tong PC, Ng MC, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van BE, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen TO, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CN, Rotimi C, Chan JC, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K: Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39:977–983, 2007
8. Holtzman NA, Marteau TM: Will genetics revolutionize medicine? *N Engl J Med* 343:141–144, 2000
9. Janssens AC, Gwinn M, Valdez R, Narayan KM, Khoury MJ: Predictive genetic testing for type 2 diabetes. *Br Med J* 333:509–510, 2006
10. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM: Predictive testing for complex diseases using multiple genes: fact or fiction? *Genet Med* 8:395–400, 2006
11. Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ: The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genet Med* 9:528–535, 2007
12. Yang Q, Khoury MJ, Botto L, Friedman JM, Flanders WD: Improving the prediction of complex diseases by testing for multiple disease-susceptibility genes. *Am J Hum Genet* 72:636–649, 2003
13. Wray NR, Goddard ME, Visscher PM: Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 17:1520–1528, 2007
14. Lyssenko V, Almgren P, Anevski D, Orho-Melander M, Sjogren M, Saloranta C, Tuomi T, Groop L: Genetic prediction of future type 2 diabetes. *PLoS Med* 2:e345, 2005
15. Vaxillaire M, Veslot J, Dina C, Proenca C, Cauchi S, Charpentier G, Tichet J, Fumeron F, Marre M, Meyre D, Balkau B, Froguel P: Impact of common type 2 diabetes risk polymorphisms in the D.E.S.I.R. Prospective Study. *Diabetes* 57:244–254, 2008
16. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, Rayner NW, Shields B, Owen KR, Hattersley AT, Frayling TM: Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med* 3:e374, 2006
17. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC: The Rotterdam Study: objectives and design update. *Eur J Epidemiol* 22:819–829, 2007
18. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 20:1183–1197, 1997
19. Alberti KG, Zimmet PZ: Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 15:539–553, 1998
20. Fang Y, van Meurs JB, d'Alesio A, Jhamai M, Zhao H, Rivadeneira F, Hofman A, van Leeuwen JP, Jehan F, Pols HA, Uitterlinden AG: Promoter and 3′-untranslated-region haplotypes in the vitamin D receptor gene predispose to osteoporotic fracture: the Rotterdam Study. *Am J Hum Genet* 77:807–823, 2005
21. Hanley JA, McNeil BJ: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843, 1983
22. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36, 1982
23. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, Lango H, Frayling TM, Neumann RJ, Sherva R, Blech I, Pharoah PD, Palmer CN, Kimber C, Tavendale R, Morris AD, McCarthy MI, Walker M, Hitman G, Glaser B, Permutt MA, Hattersley AT, Wareham NJ, Barroso I: Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 39:951–953, 2007
24. Frayling TM: A new era in finding type 2 diabetes genes: the unusual suspects. *Diabet Med* 24:696–701, 2007
25. Cauchi S, Meyre D, Durand E, Proenca C, Marre M, Hadjadj S, Choquet H, De GF, Gaget S, Allegaert F, Delplanque J, Permutt MA, Wasson J, Blech I, Charpentier G, Balkau B, Vergnaud AC, Czernichow S, Patsch W, Chikri M, Glaser B, Sladek R, Froguel P: Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. *PLoS Med* 3:e2031, 2008
26. Stumvoll M, Goldstein BJ, van Haeften TW: Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 365:1333–1346, 2005
27. Janssens AC, Gwinn M, Khoury MJ, Subramonia-Iyer S: Does genetic testing really improve the prediction of future type 2 diabetes? *PLoS Med* 3:e114, 2006
28. Pardo Silva MC, Janssens AC, Hofman A, Witteman JC, van Duijn CM: Apolipoprotein E gene is related to mortality only in normal weight individuals: the Rotterdam Study. *Eur J Epidemiol* 23:135–142, 2008