ORIGINAL ARTICLE

Clinical Risk Factors, DNA Variants, and the Development of Type 2 Diabetes

Valeriya Lyssenko, M.D., Anna Jonsson, M.Sc., Peter Almgren, M.Sc., Nicoló Pulizzi, M.D., Bo Isomaa, M.D., Tiinamaija Tuomi, M.D., Göran Berglund, M.D., David Altshuler, M.D., Peter Nilsson, M.D., and Leif Groop, M.D.

ABSTRACT

BACKGROUND

Type 2 diabetes mellitus is thought to develop from an interaction between environmental and genetic factors. We examined whether clinical or genetic factors or both could predict progression to diabetes in two prospective cohorts.

METHODS

We genotyped 16 single-nucleotide polymorphisms (SNPs) and examined clinical factors in 16,061 Swedish and 2770 Finnish subjects. Type 2 diabetes developed in 2201 (11.7%) of these subjects during a median follow-up period of 23.5 years. We also studied the effect of genetic variants on changes in insulin secretion and action over time.

RESULTS

Strong predictors of diabetes were a family history of the disease, an increased body-mass index, elevated liver-enzyme levels, current smoking status, and reduced measures of insulin secretion and action. Variants in 11 genes (*TCF7L2*, *PPARG*, *FTO*, *KCNJ11*, *NOTCH2*, *WFS1*, *CDKAL1*, *IGF2BP2*, *SLC30A8*, *JAZF1*, and *HHEX*) were significantly associated with the risk of type 2 diabetes independently of clinical risk factors; variants in 8 of these genes were associated with impaired beta-cell function. The addition of specific genetic information to clinical factors slightly improved the prediction of future diabetes, with a slight increase in the area under the receiver-operating-characteristic curve from 0.74 to 0.75; however, the magnitude of the increase was significant ($P=1.0\times10^{-4}$). The discriminative power of genetic risk factors improved with an increasing duration of follow-up, whereas that of clinical risk factors decreased.

CONCLUSIONS

As compared with clinical risk factors alone, common genetic variants associated with the risk of diabetes had a small effect on the ability to predict the future development of type 2 diabetes. The value of genetic factors increased with an increasing duration of follow-up.

From the Department of Clinical Sciences, Lund University, Malmö, Sweden (V.L., A.J., P.A., G.B., P.N., L.G.); the University of Pisa, Pisa, Italy (N.P.); the Folkhalsan Research Center (B.I., T.T.) and Helsinki University Central Hospital and the University of Helsinki (T.T., L.G.) — all in Helsinki, Finland; and the Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, and Massachusetts General Hospital, Boston (D.A.). Address reprint requests to Dr. Lyssenko at the Department of Clinical Sciences, Diabetes, and Endocrinology, Lund University, CRC, Malmö University Hospital, 20502 Malmö, Sweden, or at valeri.lyssenko@med.lu.se.

N Engl J Med 2008;359:2220-32.

Type 2 diabetes mellitus is a complex polygenic disorder in which common genetic variants interact with environmental factors to unmask the disease. The identification of persons at high risk for the disease may aid in disease prevention. A family history of diabetes, an increase in body-mass index (BMI, the weight in kilograms divided by the square of the height in meters), and impaired insulin secretion and action are risk factors for type 2 diabetes.[1-4] A challenge has been to identify genetic variants that explain the excess risk associated with a family history of diabetes. From a long list of candidate genes, variants in only three have been consistently associated with type 2 diabetes: *TCF7L2*, *KCNJ11*, and *PPARG*.[5-7] However, in 2007, a number of novel genetic variants (*CDKAL1*, *IGF2BP2*, the locus on chromosome 9 close to *CDKN2A/CDKN2B*, *FTO*, *HHEX*, *SLC30A8*, and *WFS1*)[8-14] were shown to increase susceptibility to type 2 diabetes in reproducible studies. Furthermore, a recent meta-analysis identified six novel variants (*JAZF1*, *CDC123/CAMK1D*, *TSPAN8/LGR5*, *THADA*, *ADAMTS9*, and *NOTCH2*) that are associated with type 2 diabetes.[15]

We examined subjects in two large Scandinavian prospective studies with a median follow-up period of 23.5 years to determine whether these genetic variants alone or in combination with clinical risk factors might predict the future development of type 2 diabetes and whether these variants were associated with changes in insulin secretion or action over time.

METHODS

STUDY POPULATIONS

We followed two prospective cohorts from the Malmö Preventive Project (MPP) and the Botnia study in Finland, consisting of 18,831 persons, for a median period of 23.5 years (Fig. 1, and the Methods section and Table 1A in the Supplementary Appendix, available with the full text of this article at www.nejm.org). Among these subjects, diabetes developed in 2201 (11.7%) during this period.

MEASUREMENTS

We measured weight, height, waist and hip circumference, and blood pressure, as reported previously.[16] In the MPP cohort at baseline, blood samples were drawn at 0, 40, and 120 minutes during the 75-g oral glucose-tolerance test for measurements of blood glucose and serum insulin, and fasting samples were drawn at a follow-up visit for measurement of plasma glucose and lipids with the use of standard techniques.[17] In the Botnia study, blood samples were drawn 10 minutes before the glucose-tolerance test and then at 0, 30, 60, and 120 minutes. The insulin sensitivity index (ISI) was calculated from the oral glucose-tolerance test according to the formula[18]:

$$\mathrm{ISI} = 10{,}000 \div \sqrt{([\text{fasting plasma glucose} \times \text{fasting plasma insulin}] \times [\text{mean OGTT}_{\text{glucose}} \times \text{mean OGTT}_{\text{insulin}}])},$$

in which OGTT denotes the oral glucose-tolerance test. We calculated the basal insulin resistance index by the homeostasis model assessment (HOMA) levels of fasting insulin and glucose (www.dtu.ox.ac.uk). Beta-cell function was assessed as corrected incremental insulin response (CIR) during the glucose-tolerance test according to the formula[19]:

$$\mathrm{CIR} = (100 \times \text{insulin at 30 min or 40 min in MPP}) \div ([\text{glucose at 30 min or 40 min in MPP}] \times [\text{glucose at 30 min or 40 min in MPP} - 3.89])$$

or as a disposition index (i.e., insulin secretion adjusted for insulin sensitivity, or CIR × ISI).[20]

Plasma glucose was measured by the hexokinase method in the MPP cohort and by the glucose oxidase method in the Botnia cohort. Plasma insulin was measured with the use of a local radioimmunoassay in the MPP cohort and with the use of enzyme-linked immunosorbent assay (Dako) in the Botnia cohort.[16,21]

GENOTYPING

We genotyped 16 SNPs in 16 genes that in recent cross-sectional genomewide association studies have shown consistent association with type 2 diabetes: *TCF7L2* (rs7903146), *KCNJ11* (rs5219), *PPARG* (rs1801282), *CDKAL1* (rs7754840), *IGF2BP2* (rs4402960), *CDKN2A/CDKN2B* (rs10811661), *FTO* (rs9939609), *HHEX* (rs1111875), *SLC30A8* (rs13266634), *WFS1* (rs10010131), *JAZF1* (rs864745), *CDC123/CAMK1D* (rs12779790), *TSPAN8/LGR5* (rs7961581), *THADA* (rs7578597), *ADAMTS9* (rs4607103), and *NOTCH2* (rs10923931).[5-9,11-15,22]

DNA was extracted from whole blood with the use of a Plasmid Maxi Kit (Qiagen). Genotyping was performed with the use of matrix-assist-

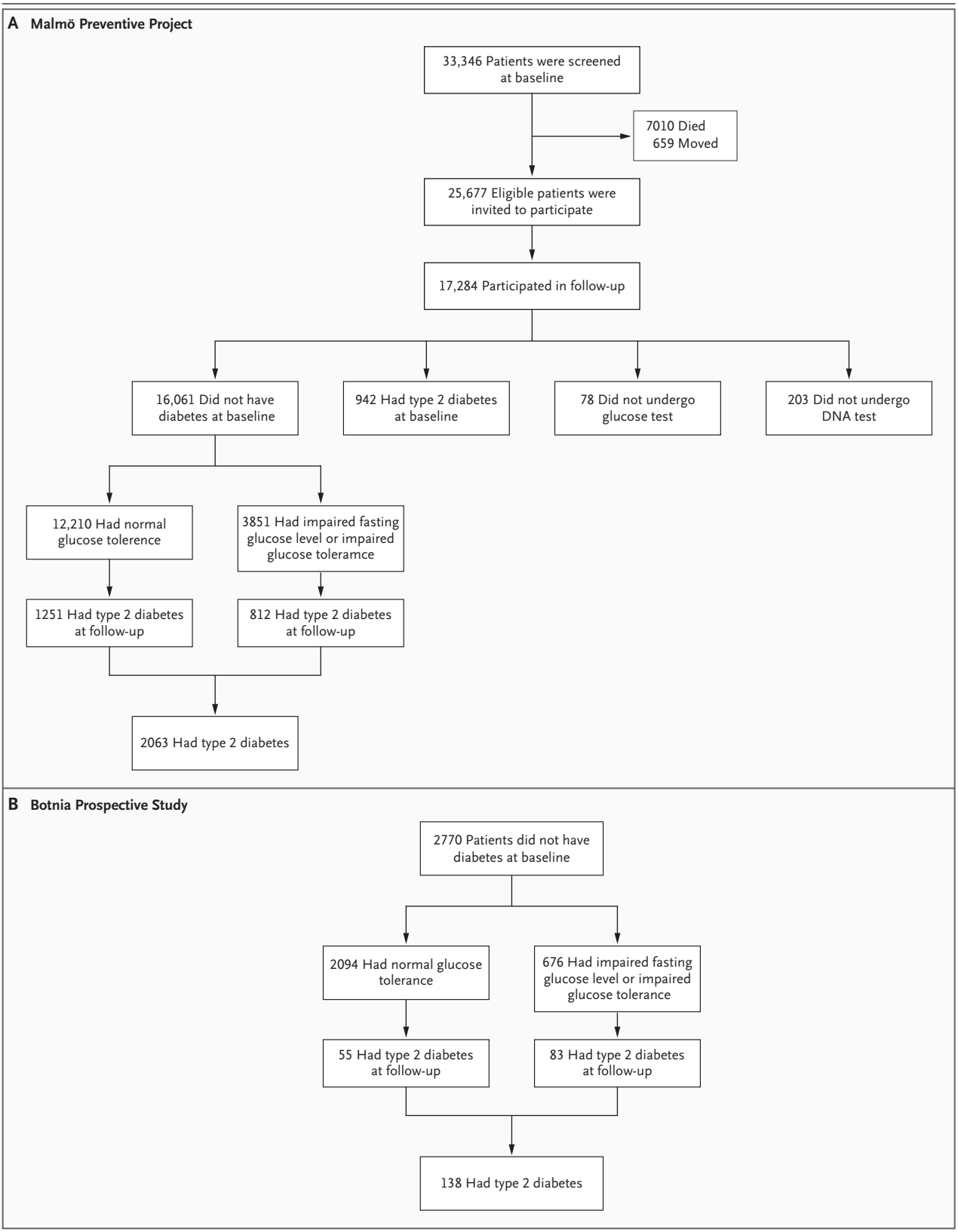

Figure 1 (facing page). Design of the Malmö and Botnia Studies.

Panel A shows an outline of the data collection in the Malmö Preventive Project, in which 16,061 subjects without diabetes were initially eligible for the study of the prediction of future diabetes; type 2 diabetes developed in 2063 (12.8%) of these subjects. Panel B shows the progression to diabetes in the Botnia Prospective Study, which included 2770 family members and spouses without diabetes; type 2 diabetes developed in 138 (5.0%) of these subjects.

ed laser desorption–ionization time-of-flight mass spectrometry on the MassARRAY platform (Sequenom)[23] for rs7903146, rs1801282, rs5219, rs7754840, and rs10811661; with an allelic discrimination assay-by-design method on ABI 7900 (Applied Biosystems) for rs4402960, rs9939609, rs10010131, rs1111875, rs864745, rs12779790, rs7961581, rs7578597, rs4607103, and rs10923931; and with an allele-specific assay (KASPar, KBioscience) for rs13266634. We obtained an average genotyping success rate of more than 95% and an average genotyping accuracy of more than 98% by regenotyping 11% of the samples using the Sequenom platform. All SNPs were in Hardy–Weinberg equilibrium ($P>0.001$), with the exception of rs864745 in the *JAZF1* gene ($P=0.001$). Genotyping errors are an unlikely explanation for this finding, since in the genotyping of 2416 samples (15%) of rs864745 with the use of two different methods (allelic discrimination on ABI7900 and Sequenom), the concordance rate was 98.7%.

STATISTICAL ANALYSIS

We investigated the predictive ability of clinical factors and the specific polymorphisms that we had genotyped as risk factors for future type 2 diabetes using logistic-regression analysis applied to the following models: first, a model using univariate clinical risk factors (with adjustment for age and sex); second, a model using personal factors (age, sex, family history of diabetes, and BMI) and clinical factors (age, sex, family history of diabetes, BMI, and levels of blood pressure, triglycerides, and fasting plasma glucose), as used by Wilson et al. in the Framingham Offspring Study[4]; third, a clinical model in which we replaced the clinical variables suggested by Wilson et al. by measures of insulin secretion; and fourth, a clinical model with the polymorphic gene variants. Since men and women were included at different times, we adjusted for this factor using the participation period (coded 0 or 1), sex, and an interaction term (participation period × sex, which was coded 0 or 1) as covariates in the analyses.

Improvement in area under the receiver-operating-characteristic (ROC) curves (also referred to as C statistics) was assessed after adding the genetic data to the clinical model.[24] To confirm that the addition of genetic data to clinical models improved risk prediction, we tested the ability of the combined clinical and genetic model to reclassify subjects into predefined risk categories on the basis of the percentage likelihood of type 2 diabetes developing (<10%, 10 to 20%, and >20%), using the net-reclassification-improvement approach.[25] Since this method requires predefined risk categories, we also used another approach, without this requirement (i.e., the integrated-discrimination-improvement method).[25]

For the first analysis of the effects of the polymorphic DNA variants, we used additive genetic models. In addition, we tested dominant and recessive alternative models for the best fit (http://pngu.mgh.harvard.edu/~purcell). Multivariate linear regression analyses were used to test correlations between genotype and phenotype.[26] Non-normally distributed variables were log-transformed before analysis. The effect size of a genetic or clinical risk factor on the risk of type 2 diabetes was calculated from multivariate regression analysis, with adjustment for age and sex, with the use of Nagelkerke R square. We estimated the predictive value of a combination of risk alleles (each person could have 0, 1, or 2 of them, for a total of 22) in 11 genes, which significantly predicted the risk of diabetes by defining subjects with more than 12 risk alleles (about 20%) as being at high risk and those with fewer than 8 risk alleles (about 20%) as being at low risk. All statistical analyses were performed with the use of SPSS software, version 14.0; PLINK (http://pngu.mgh.harvard.edu/~purcell/plink); and Stata software.

RESULTS

INCIDENCE OF DIABETES

In the MPP study, diabetes developed in 2063 subjects (12.8%) during a median follow-up period

Table 1. Baseline Clinical Factors Predicting Type 2 Diabetes in the Malmö and Botnia Prospective Studies.*

Variable	Malmö		Botnia	
	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	P Value
Univariate regression analysis				
Age, per SD	1.13 (1.07–1.20)	2.9×10^{-5}	1.87 (1.54–2.27)	2.9×10^{-6}
First-degree family history of diabetes	1.67 (1.46–1.91)	1.0×10^{-13}	2.13 (1.22–3.71)	0.008
Current smoking	1.30 (1.18–1.43)	6.5×10^{-8}	1.20 (0.76–1.89)	0.43
Impaired fasting glucose level	2.58 (2.29–2.91)	5.6×10^{-53}	3.84 (2.22–6.64)	2.0×10^{-6}
Impaired glucose tolerance	1.91 (1.61–2.26)	1.7×10^{-14}	3.67 (2.27–5.93)	1.8×10^{-8}
Both impaired fasting glucose level and impaired glucose tolerance	5.37 (4.26–6.78)	5.9×10^{-47}	7.77 (4.48–13.46)	1.9×10^{-15}
Increased body-mass index, per 1 SD	1.84 (1.76–1.93)	2.1×10^{-153}	1.84 (1.57–2.16)	5.0×10^{-13}
Increased waist circumference, per 1 SD	NA	NA	2.23 (1.85–2.67)	2.9×10^{-17}
Increased blood pressure, per 1 SD				
Systolic	1.34 (1.28–1.41)	2.6×10^{-37}	1.56 (1.31–1.85)	9.8×10^{-7}
Diastolic	1.39 (1.33–1.46)	3.9×10^{-44}	1.52 (1.27–1.83)	2.2×10^{-5}
Increased triglycerides, per 1 SD	1.70 (1.62–1.78)	9.8×10^{-101}	1.53 (1.28–1.82)	8.1×10^{-6}
Increased γ-glutamyltransferase, per 1 SD	1.53 (1.46–1.60)	8.4×10^{-72}	1.44 (1.21–1.71)	3.8×10^{-5}
Increased aspartate aminotransferase, per 1 SD	1.28 (1.22–1.35)	2.2×10^{-24}	0.99 (0.83–1.20)	0.98
Increased alanine aminotransferase, per 1 SD	1.64 (1.56–1.72)	9.1×10^{-86}	NA	NA
Increased apolipoprotein A-I, per 1 SD	0.79 (0.73–0.86)	2.6×10^{-8}	0.75 (0.62–0.91)	0.002
Increased insulin sensitivity index, per 1 SD	0.59 (0.53–0.65)	1.2×10^{-22}	0.43 (0.36–0.52)	1.2×10^{-17}
Increased insulin resistance index, per 1 SD	1.47 (1.33–1.62)	2.4×10^{-14}	2.02 (1.70–2.39)	5.3×10^{-12}
Decreased corrected insulin response, per 1 SD	1.26 (1.10–1.44)	7.1×10^{-4}	1.48 (1.21–1.82)	3.3×10^{-4}
Decreased disposition index, per 1 SD	2.12 (1.82–2.47)	6.7×10^{-22}	3.40 (2.66–4.34)	2.5×10^{-21}
Multivariate regression analysis†				
Age, per SD	0.92 (0.83–1.009)	0.08	1.52 (1.20–1.91)	2.9×10^{-4}
Male sex	0.88 (0.67–1.12)	0.29	NA	NA
Time period for participation in study	0.72 (0.55–0.95)	0.02	NA	NA
Time period for either men or women	1.76 (1.29–2.41)	6.3×10^{-4}	NA	NA
First-degree family history of diabetes	1.62 (1.38–1.89)	2.0×10^{-10}	NA	NA
Current smoking	1.43 (1.25–1.63)	1.4×10^{-9}	NA	NA
Increased body-mass index, per 1 SD	1.45 (1.37–1.55)	4.5×10^{-36}	1.43 (1.20–1.72)	0.002
Increased fasting plasma glucose, per 1 SD	1.54 (1.43–1.65)	1.7×10^{-34}	NA	NA
Increased diastolic blood pressure, per 1 SD	1.15 (1.07–1.22)	8.6×10^{-6}	NA	NA
Increased triglycerides, per 1 SD	1.26 (1.18–1.35)	4.8×10^{-13}	NA	NA
Increased γ-glutamyltransferase, per 1 SD	1.12 (1.04–1.21)	0.002	NA	NA
Increased aspartate aminotransferase, per 1 SD	0.90 (0.83–0.98)	0.01	NA	NA
Increased alanine aminotransferase, per 1 SD	1.37 (1.25–1.50)	2.4×10^{-11}	NA	NA
Increased apolipoprotein A-I, per 1 SD‡	NA	NA	0.76 (0.62–0.92)	0.006
Decreased disposition index, per 1 SD§	NA	NA	3.04 (2.34–3.96)	5.1×10^{-14}

* CI denotes confidence interval, and NA not applicable.
† Missing values from the Botnia study were not part of the multivariate regression analysis because they were not significant.
‡ Values for apolipoprotein A-I from the Malmö study were not calculated in the multivariate regression analysis because they were not significant.
§ The disposition index in the Malmö study was not part of the multivariate regression analysis because of the limited data that were available (see Table 1 in the Supplementary Appendix).

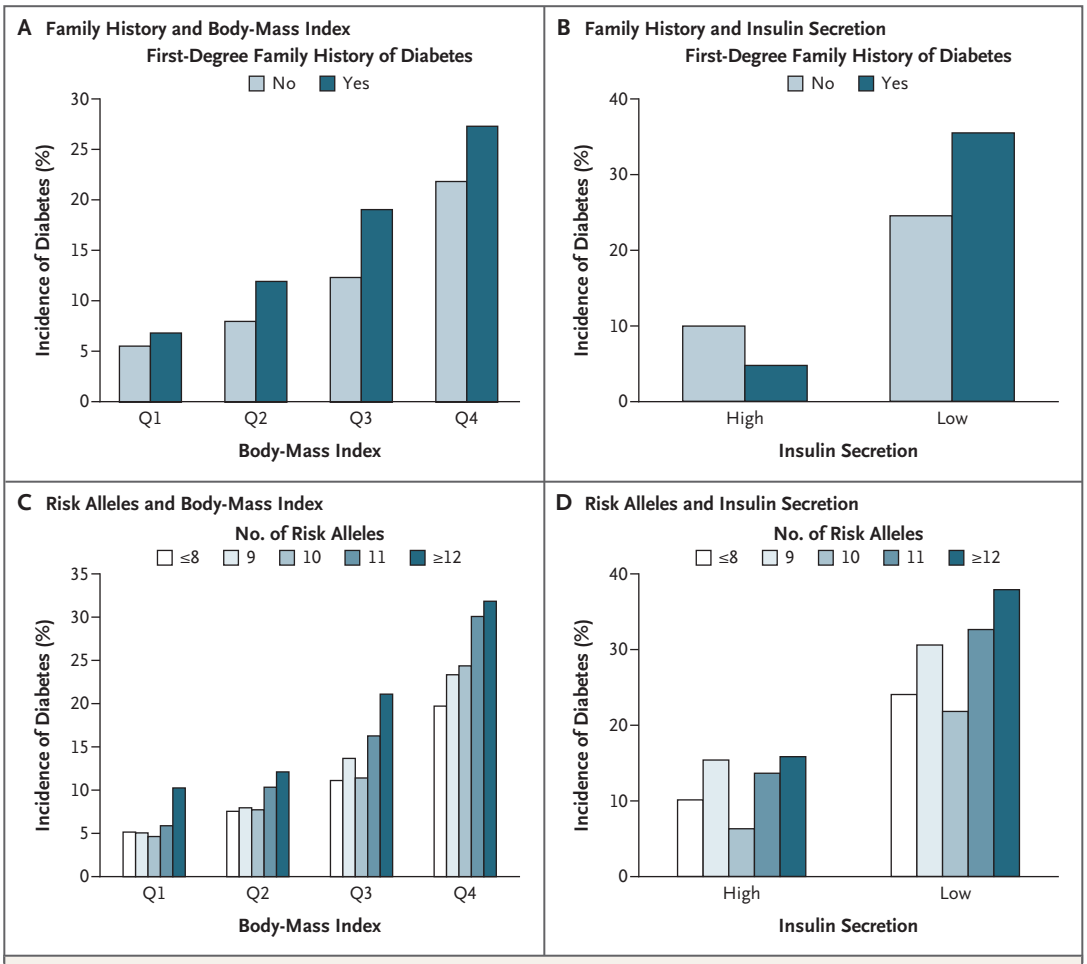

Figure 2. Nongenetic and Genetic Risk Factors for Type 2 Diabetes in the Malmö Study.

Panel A shows the incidence of type 2 diabetes in four quartiles (Q) of body-mass index (BMI) among Malmö subjects who had a family history of diabetes and those without such a history. An increase in the quartile of the BMI gradually increased the risk of diabetes, as compared with the lowest quartile, with an odds ratio of 1.50 for the second quartile (95% confidence interval [CI], 1.26 to 1.78; $P=6.7\times10^{-6}$), of 2.36 for the third quartile (95% CI, 2.00 to 2.78; $P=1.5\times10^{-24}$), and of 4.96 for the fourth quartile (95% CI, 4.25 to 5.79; $P=1.1\times10^{-90}$). Panel B shows the incidence of type 2 diabetes in relation to insulin secretion (disposition index) among subjects with a family history of diabetes and those without such a history. Subjects with a disposition index below the median of 23,393 (26.1% of high-risk subjects and 9.4% of low-risk subjects) had an increase in the risk of type 2 diabetes by a factor of 3.23 (95% CI, 2.41 to 4.34; $P=5.8\times10^{-15}$), as compared with those above the median. A family history of diabetes significantly increased the risk of diabetes in subjects with impaired insulin secretion (35.5% vs. 9.9%), with an odds ratio of 4.86 (3.12 to 7.56, $P=2.3\times10^{-12}$). Panel C shows the incidence of type 2 diabetes in carriers of an increasing number of risk alleles in 11 genes, which individually predicted future risk of type 2 diabetes, in relation to quartiles of BMI. There was a stepwise increase in diabetes risk with an increasing number of risk alleles and increasing quartiles of BMI so that participants carrying more than 12 risk alleles showed a doubling of the risk conferred by BMI alone. In the highest quartile of BMI (31.8% vs. 5.1%), this yielded an odds ratio of 8.0 (95% CI, 5.71 to 11.19; $P=9.1\times10^{-34}$). Panel D shows the incidence of type 2 diabetes in carriers of an increasing number of risk alleles in the 11 genes, which individually predicted future risk of type 2 diabetes, in relation to low insulin secretion. Carriers of more than 12 risk alleles and a low disposition index (37.9% vs. 10.1%) had an odds ratio of 5.81 (95% CI, 3.18 to 10.61; $P=1.1\times10^{-8}$).

of 24.8 years, with the highest conversion rate (21.1%) in those with impaired fasting glucose levels or impaired glucose tolerance at baseline (Fig. 1A). Impaired fasting glucose levels developed in 1400 of 10,933 subjects with normal glucose tolerance at baseline (12.8%). In the Botnia study, diabetes developed during the follow-up period in 138 of all 2770 subjects (5.0%) and in 83 (12.3%) of those with impaired fasting glucose levels or impaired glucose tolerance at baseline

Table 2. Genetic Factors Predicting Type 2 Diabetes in the Malmö and Botnia Studies.*

Chromosome	Gene	SNP	Risk Allele			Malmö					Botnia					P Value for Heterogeneity†
				RA_{AFF}	No. of Subjects	Primary Model		Alternative Model		RA_{AFF}	No. of Subjects	Primary Model		Alternative Model		
						odds ratio (95% CI)	*P value*	*odds ratio (95% CI)*	*P value*			*odds ratio (95% CI)*	*P value*	*odds ratio (95% CI)*	*P value*	
10	*TCF7L2*	rs7903146	T	0.31	15815	1.30 (1.21–1.40)	9.5×10^{-13}	Dominant 1.42 (1.29–1.55)	5.0×10^{-13}	0.25	2645	1.52 (1.14–2.04)	0.003	Dominant 1.53 (1.07–2.21)	0.02	0.33
3	*PPARG*	rs1801282	C	0.88	15993	1.20 (1.08–1.32)	4.0×10^{-4}	NA	NA	0.90	2544	1.45 (0.96–2.20)	0.08	NA	NA	0.30
16	*FTO*	rs9939609	A	0.44	15931	1.14 (1.07–1.22)	9.2×10^{-5}	Recessive 1.26 (1.12–1.42)	1.3×10^{-4}	0.42	2464	1.04 (0.80–1.36)	0.77	Recessive 1.14 (0.71–1.85)	0.58	0.58
11	*KCNJ11*	rs5219	T	0.41	15600	1.13 (1.06–1.21)	3.6×10^{-4}	NA	NA	0.51	2635	0.98 (0.75–1.26)	0.85	NA	NA	0.31
1	*NOTCH2*	rs10923931	T	0.11	15589	1.13 (1.02–1.26)	0.02	NA	NA	0.11	2642	1.15 (0.77–1.73)	0.45	NA	NA	0.17
4	*WFS1*	rs10010131	G	0.59	15944	1.12 (1.04–1.19)	0.001	Recessive 1.17 (1.07–1.30)	0.001	0.52	2631	0.79 (0.61–1.01)	0.08	Recessive 0.84 (0.56–1.26)	0.40	0.01
6	*CDKAL1*	rs7754840	C	0.33	15487	1.11 (1.03–1.19)	0.004	Recessive 1.24 (1.07–1.44)	0.004	0.34	2495	1.05 (0.80–1.37)	0.74	Recessive 1.15 (0.65–2.01)	0.63	0.87
3	*IGF2BP2*	rs4402960	A	0.32	15157	1.10 (1.03–1.18)	0.008	Recessive 1.23 (1.06–1.44)	0.008	0.27	2500	0.92 (0.69–1.24)	0.59	Recessive 0.78 (0.36–1.71)	0.53	0.30
8	*SLC30A8*	rs13266634	C	0.70	15931	1.10 (1.03–1.18)	0.008	NA	NA	0.57	2497	0.85 (0.66–1.10)	0.21	NA	NA	0.04
7	*JAZF1*	rs864745	A	0.53	15944	1.08 (1.01–1.15)	0.03	NA	NA	0.49	2639	0.99 (0.77–1.28)	0.95	NA	NA	0.63
10	*HHEX*	rs1111875	G	0.60	15942	1.07 (1.00–1.15)	0.03	NA	NA	0.56	2597	0.99 (0.76–1.29)	0.92	NA	NA	0.83
2	*THADA*	rs7578597	T	0.90	15620	1.11 (0.99–1.24)	0.07	NA	NA	0.96	2658	1.29 (0.69–2.39)	0.42	NA	NA	0.48

9	*CDKN2A/ CDKN2B*	rs10811661	T	0.85	15132	1.09 (0.99–1.19)	0.07	NA	NA	0.82	2475	0.84 (0.60–1.17)	0.29	NA	NA	0.18
12	*TSPAN8/ LGR5*	rs7961581	G	0.27	15594	1.04 (0.96–1.12)	0.30	NA	NA	0.26	2656	1.09 (0.83–1.43)	0.55	NA	NA	0.59
3	*ADAMTS9*	rs4607103	C	0.77	15729	1.04 (0.96–1.13)	0.34	NA	NA	0.81	2640	1.23 (0.88–1.73)	0.22	NA	NA	0.27
10	*CDC123/ CAMK1D*	rs12779790	G	0.19	15547	1.02 (0.93–1.11)	0.70	NA	NA	0.22	2642	1.03 (0.76–1.39)	0.87	NA	NA	0.94

* The odds ratios for the risk of type 2 diabetes were calculated with the use of univariate logistic-regression analyses with adjustment for age at participation and sex. The primary genetic models are additive; alternative models are indicated. CI denotes confidence interval, NA not applicable, and RA_{AFF} frequency of the risk allele in affected subjects.
† The test for heterogeneity (interaction) indicates whether the results varied significantly between the Malmö and Botnia studies.

(Fig. 1B), whereas impaired fasting glucose or impaired glucose tolerance developed in 313 of 2039 subjects (15.4%).

CLINICAL FACTORS PREDICTING INCIDENCE OF DIABETES

In both the MPP and Botnia studies, a family history of diabetes, an increased BMI, and increased levels of blood pressure and serum levels of triglycerides, apolipoprotein A-I, and liver enzymes were independent predictors of future type 2 diabetes (Table 1). In the MPP study, current smoking was also associated with a marked increase in the risk of diabetes. Impaired insulin secretion and action, particularly insulin secretion adjusted for insulin resistance (disposition index), were strong predictors of future diabetes. The presence of a first-degree family history of diabetes doubled the risk of the disease that was seen with an increased BMI (Fig. 2A) and a low disposition index (Fig. 2B).

We also constructed models for personal factors (age, sex, a family history of diabetes, and body-mass index) and clinical factors (age, sex, a family history of diabetes, BMI, blood pressure, triglycerides, and fasting plasma glucose) for the risk of type 2 diabetes, as described by Wilson et al.[4] The area under the ROC curve (AUC) for the clinical models was similar in the MPP study (0.74) and the Botnia study (0.79) but was lower than in the Framingham Offspring Study (0.88) (Table 4 in the Supplementary Appendix). The addition of measures of insulin secretion to the clinical model[4] significantly improved values in the AUC in both studies ($P<0.01$).

PREDICTIVE EFFECT OF GENOTYPED DNA VARIANTS

Type 2 Diabetes

Common variants in 11 genes were significantly associated with the risk of future type 2 diabetes in the MPP cohort, including *TCF7L2* (odds ratio, 1.30; $P=9.5\times10^{-13}$), *PPARG* (odds ratio, 1.20; $P=4.0\times10^{-4}$), *FTO* (odds ratio, 1.14; $P=9.2\times10^{-5}$), *KCNJ11* (odds ratio, 1.13; $P=3.6\times10^{-4}$), *NOTCH2* (odds ratio, 1.13; $P=0.02$), *WFS1* (odds ratio, 1.12; $P=0.001$), *CDKAL1* (odds ratio, 1.11; $P=0.004$), *IGF2BP2* (odds ratio, 1.10; $P=0.008$), *SLC30A8* (odds ratio, 1.10; $P=0.008$), *JAZF1* (odds ratio, 1.08; $P=0.03$), and *HHEX* (odds ratio, 1.07; $P=0.03$) (Table 2). Although these findings could not be fully replicated in the smaller Botnia study, there was little heterogeneity between the studies with

respect to the risk conferred by different genotypes.

We also studied whether these variants would predict conversion from normal glucose tolerance and from impaired fasting glucose levels or impaired glucose tolerance to type 2 diabetes in different ways. Variants in most genes predicted progression from normal glucose tolerance to type 2 diabetes, including *TCF7L2* (odds ratio, 1.27; $P=2.7\times10^{-7}$), *PPARG* (odds ratio, 1.15; P=0.03), *FTO* (odds ratio, 1.16; $P=7.2\times10^{-4}$), *KCNJ11* (odds ratio, 1.11; P=0.01), *WFS1* (odds ratio, 1.13; P=0.004), *CDKAL1* (odds ratio, 1.21; P=0.05), *IGF2BP2* (odds ratio, 1.12; P=0.01), and *SLC30A8* (odds ratio, 1.11; P=0.02), whereas four variants predicted transition from impaired fasting glucose levels or impaired glucose tolerance to type 2 diabetes, including *TCF7L2* (odds ratio, 1.30; $P=2.7\times10^{-5}$), *PPARG* (odds ratio, 1.29; P=0.004), *KCNJ11* (odds ratio, 1.15; P=0.02), and *FTO* (odds ratio, 1.13; P=0.03) (Table 3 in the Supplementary Appendix).

Table 3. Combination of Baseline Clinical and Genetic Factors Predicting Type 2 Diabetes in the Malmö and Botnia Studies.*

	Odds Ratio (95% CI)	P Value
Malmö		
Age, per 1 SD	0.96 (0.87–1.06)	0.44
Male sex	0.94 (0.73–1.22)	0.65
Time period for participation in study	0.69 (0.52–0.92)	0.01
Time period for either men or women	1.74 (1.25–2.41)	0.001
First-degree family history of diabetes	1.65 (1.39–1.95)	1.0×10^{-9}
Current smoking	1.39 (1.29–1.61)	6.3×10^{-8}
Increased body-mass index, per 1 SD	1.49 (1.39–1.59)	4.0×10^{-34}
Increased fasting plasma glucose, per 1 SD	1.51 (1.40–1.59)	1.6×10^{-26}
Increased diastolic blood pressure, per 1 SD	1.16 (1.09–1.25)	0.006
Increased triglycerides, per 1 SD	1.28 (1.19–1.38)	9.3×10^{-13}
Increased γ-glutamyltransferase, per 1 SD	1.10 (1.01–1.19)	0.02
Increased aspartate aminotransferase, per 1 SD	0.91 (0.83–0.99)	0.03
Increased alanine aminotransferase, per 1 SD	1.37 (1.24–1.51)	2.6×10^{-9}
Combination of the risk alleles in 11 SNPs†	1.12 (1.08–1.15)	8.1×10^{-13}
Botnia		
Age, per 1 SD	1.32 (1.20–1.91)	0.05
Increased apolipoprotein A-I, per 1 SD	0.69 (0.54–0.87)	0.002
Increased body-mass index, per 1 SD	1.48 (1.21–1.81)	0.002
Decreased disposition index, per 1 SD‡	3.29 (2.39–4.52)	3.1×10^{-13}
Combination of the risk alleles in 11 SNPs†	0.94 (0.84–1.04)	0.23

* The odds ratios for the risk of type 2 diabetes were calculated with the use of multivariate logistic regression analyses with adjustment for age at participation and sex. CI denotes confidence interval.
† The combination of the risk alleles in 11 SNPs was calculated as sum of the risk alleles (coded as 0, 1, or 2).
‡ The disposition index was not part of the multivariate regression analysis in the Malmö study because limited data were available (see Table 1 in the Supplementary Appendix).

On the basis of the frequency distribution of risk alleles, we defined a low genetic risk group and a high genetic risk group as the quintile with the lowest (≤8) and the highest (≥12) number of risk alleles, respectively (Fig. 1 in the Supplementary Appendix). As expected, more subjects who were diagnosed with diabetes than those without the diagnosis had a high genetic risk (32.0% vs. 22.1%), which translated into an increase in the risk of future diabetes by a factor of 1.95 (95% confidence interval [CI], 1.69 to 2.25; $P=2.5\times10^{-20}$). In addition, in the multivariate regression analysis, the inclusion of the genotyped DNA variants provided information that was independent of clinical risk factors, showing in the Malmö cohort an increase in the risk of type 2 diabetes by a factor of 1.12 per single copy of the risk allele ($P=8.1\times10^{-13}$) (Table 3).

Change in Body-Mass Index and Insulin Secretion and Action

We examined the effect of the genotyped DNA variants on changes in the BMI and insulin secretion (disposition index) and action over time in 2444 subjects from the Botnia study who did not have diabetes. At baseline, carriers of risk genotypes in the *IGF2BP2* and *SLC30A8* genes and at the *CDKN2A/CDKN2B* locus had a lower disposition index, which was maintained unchanged throughout the 8-year observation period (P<0.05) (Fig. 3H, 3I, and 3M in the Supplementary Appendix).

The presence of variants in the *FTO*, *JAZF1*, and *ADAMTS9* genes was associated with changes in the BMI, as compared with the absence of these variants; the BMI was higher in carriers of the *FTO* risk allele by 0.24 (P<0.0001) and lower in carriers of risk genotypes in *JAZF1* and *ADAMTS9* by 0.10 (P=0.003) and 0.13 (P=0.004), respectively (Fig. 3 in the Supplementary Appendix).

In addition, we evaluated changes in insulin secretion and action and BMI over time in carriers of a low and high gene score in the Botnia study

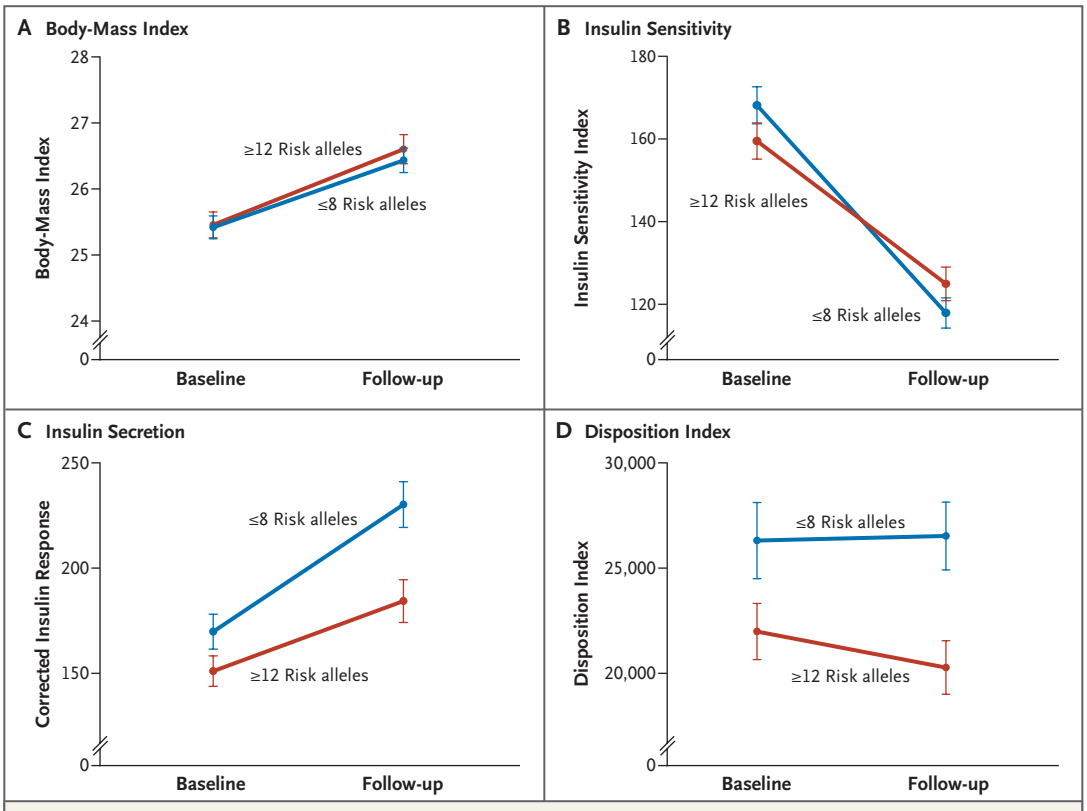

Figure 3. Effect of a Combination of Risk Alleles on Changes in Insulin Secretion and Sensitivity in the Botnia Prospective Study.

Changes in diabetes-related phenotypes over time are shown for the 380 subjects in the Botnia study who were at high genetic risk for diabetes (≥12 risk alleles) (red) and for 471 of subjects who were at low risk (≤8 risk alleles) (blue). Panel A shows similar increases in the BMI in subjects at high and low genetic risk. Panel B shows a similar worsening of insulin sensitivity in the two groups. Panel C shows an increase in corrected insulin response over time, an increase that was greater in subjects at low risk than in those at high risk. Panel D shows a decrease in beta-cell function (disposition index) in subjects at high risk, as compared with those at low risk. Among subjects at low risk, the disposition index increased slightly to compensate for an increase in insulin resistance. The I bars denote standard errors.

(Fig. 3). An increase in the BMI and a concomitant decrease in insulin sensitivity during the 8-year period were consistent findings, with no differences between subjects at high and low genetic risk (Fig. 3A and 3B). However, subjects with a high genetic risk did not increase their insulin secretion (disposition index) to compensate for the increase in insulin resistance as efficiently as did those with a low genetic risk (Fig. 3C and 3D).

Combined Effect of Clinical and Genetic Risk Factors

We also evaluated whether genetic risk factors would further increase the risk imposed by an increase in the BMI or a decrease in the disposition index. There was a stepwise increase in diabetes risk with an increasing number of risk alleles and increasing quartiles of BMI (Fig. 2C) or a disposition index above or below the median. Therefore, carriers of more than 12 risk alleles who were in the highest quartile of BMI (263 of 826 subjects vs. 45 of 874 subjects) or who had a low disposition index (58 of 153 subjects vs. 17 of 168 subjects) had an odds ratio for type 2 diabetes of 8.0 (95% CI, 5.71 to 11.19; $P=9.1\times10^{-34}$) and 5.8 (95% CI, 3.18 to 10.61, $P=1.1\times10^{-8}$), respectively (Fig. 2D).

The C statistics of the AUC had minimal yet significant improvement after the addition of data from the genotyped DNA variants to the clinical

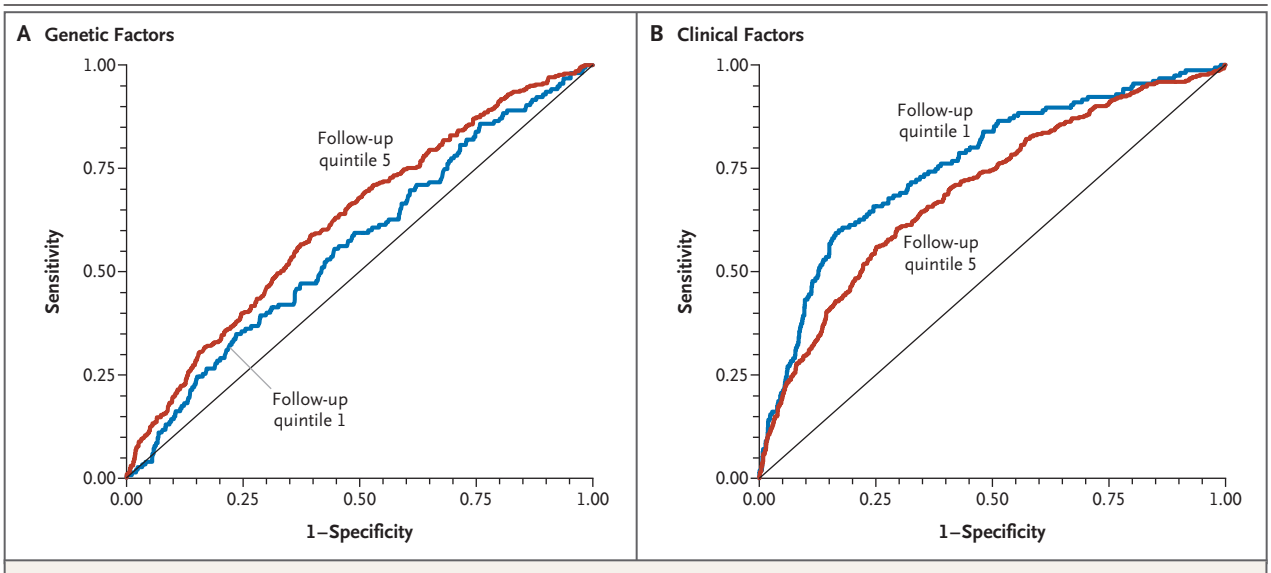

Figure 4. Area under the ROC Curve (C Statistic) for Clinical and Genetic Models Predicting Type 2 Diabetes, According to the Duration of Follow-up.

The effect of genetic risk factors increases with the duration of follow-up, with an area under the ROC curve (AUC) of 0.56 in quintile 1 (blue) and 0.62 in quintile 5 (red) (P=0.01) (Panel A), whereas the effect of clinical risk factors decreased with the duration of follow-up, with an AUC of 0.75 in quintile 1 and 0.67 in quintile 5 (P=0.01) (Panel B). The black line indicates reference values.

model (from 0.74 to 0.75, $P=1.0\times10^{-4}$) (Table 4 in the Supplementary Appendix). Since C statistics could be considered an insensitive method to identify improvement in prediction, we also reclassified subjects into three risk categories (0 to ≤10%, >10 to ≤20%, and >20%) using the net-reclassification-improvement method (Table 5 in the Supplementary Appendix). By adding genetic factors to clinical factors, we could reclassify 9% of the MPP subjects ($P=2.5\times10^{-5}$) and 20% of the Botnia subjects (P=0.05) to a higher risk category. Also, the use of the integrated-discrimination-improvement method, which did not require predefined risk categories, significantly improved the prediction of future diabetes in both the MPP subjects ($P=3.7\times10^{-14}$) and the Botnia subjects (P=0.001).

An important factor defining the discriminative power of clinical risk factors and DNA variants is the duration of follow-up. To address this issue, we defined the AUC for clinical and genetic risk factors in quintiles of time of follow-up. We observed a decrease in the AUC for the clinical model and an increase in the AUC for the genetic risk score (P=0.01 for both comparisons) with increasing duration of follow-up (Fig. 4, and Table 6 in the Supplementary Appendix).

DISCUSSION

Our study provides insight into the relative importance of clinical risk factors and those that are related to a panel of DNA variants associated with type 2 diabetes. Obesity was a strong risk factor for future diabetes, a risk that almost doubled in subjects with a family history of diabetes. However, the addition of data from genotyping of the known DNA variants to clinical risk factors (including a family history of diabetes) had a minimal, albeit statistically significant, effect on the prediction of future type 2 diabetes. Notably, the ability of genetic risk factors to predict future type 2 diabetes improved with an increasing duration of follow-up, suggesting that assessment of genetic risk factors is clinically more meaningful the earlier in life they are measured.

Although subjects in the prediabetic stage showed many features of insulin resistance, beta-cell function that was adjusted for insulin resistance (disposition index) was the strongest predictor of future diabetes. The addition of measures of insulin secretion to the clinical model, which included mostly components of the metabolic syndrome, further improved the discriminatory power of the ROC curve, from 0.70 to 0.74 in the

MPP subjects (P=0.001) and from 0.79 to 0.83 in the Botnia subjects (P=0.006).

Of the 16 loci that have been associated with type 2 diabetes previously,[8-15] we showed that 11 — *TCF7L2*, *PPARG*, *FTO*, *KCNJ11*, *NOTCH2*, *WFS1*, *CDKAL1*, *IGF2BP2*, *SLC30A8*, *JAZF1*, and *HHEX* — were associated with an enhanced risk of future diabetes. Many of the variants that we genotyped appear to influence beta-cell function, possibly through effects on proliferation, regeneration, and apoptosis. There was a time-dependent increase in the BMI and a decrease in insulin sensitivity in the subjects from the Botnia study, an increase in insulin resistance that was reflected by an increase in insulin secretion. However, this increase was inadequate to compensate for the increase in insulin resistance in carriers with a high genetic risk, which resulted in a markedly impaired disposition index. Only variants in *FTO* were associated with an increased BMI. Both *FTO* and *PPARG* together with *TCF7L2* and *KCNJ11* predicted transition from impaired fasting glucose levels or impaired glucose tolerance to manifest diabetes, which suggests that a combination of increased obesity and insulin resistance with a deterioration in beta-cell function contribute to the manifestation of diabetes in these subjects. Collectively, our findings emphasize the critical role of inherited defects in beta-cell function for the development of type 2 diabetes.

Given the large number of subjects with a long follow-up, we were in a position to determine whether the addition of genotyping data at known loci associated with diabetes to clinical risk factors could improve the ability of models to predict future diabetes. To this end, we confirmed that clinical risk factors were good predictors of future diabetes. However, our AUC values for the MPP subjects (0.74) and for the Botnia subjects (0.79) were lower than the value reported in the Framingham Offspring Study (0.88).[4] The application of the coefficients that were derived from the Framingham study to our populations decreased the values of the AUC (0.60 for the MPP subjects and 0.75 for the Botnia subjects), which suggests that different clinical variables have different discriminatory value in different studies.[27]

The addition of DNA data to the clinical model improved not only the discriminatory power, as assessed by ROC curves, but also the reclassification of the subjects into different risk strategies, with the use of net-reclassification-improvement and integrated-discrimination-improvement approaches. However, the discriminatory power of genes alone was relatively low (0.62) but in keeping with findings from two previous studies.[28,29] In contrast, one recent case–control study showed a very high value for the AUC of 0.86 for 15 novel gene loci.[30] The most likely explanation is that in this study, subjects with diabetes were compared with those who maintained completely normal glucose tolerance. If we were to restrict our analysis of subjects without diabetes to those with normal glucose tolerance, the value for the AUC would increase to 0.82. These data thus emphasize the need for population-based studies for assessment of diabetes risk.

One of the strengths of our study was its prospective nature. Cross-sectional studies often include case subjects and control subjects who were ascertained in different ways, thereby limiting their predictive value. Prospective studies have the advantage that all subjects have been ascertained and followed up in the same way. One caveat was that in the MPP study, men and women were included at different times. However, we adjusted for this variable using the participation period and sex as covariates in the analyses.

In conclusion, the inclusion of common genetic variants that are associated with type 2 diabetes very slightly improved the prediction of future type 2 diabetes, as compared with the inclusion of clinical risk factors alone. Although this effect might be too small to allow for individual risk prediction, it could be useful in reducing the number of subjects who would need to be included in intervention studies aimed at the prevention of type 2 diabetes.

Supported by grants from the Swedish Research Council (including Linné grant 31475113580), the Heart and Lung Foundation, the Swedish Diabetes Research Society, a Nordic Center of Excellence Grant in Disease Genetics, the Diabetes Program at the Lund University, the Finnish Diabetes Research Society, the Sigrid Juselius Foundation, the Påhlsson Foundation, the Crafoord Foundation, the Folkhälsan Research Foundation, the Novo Nordisk Foundation, the European Network of Genomic and Genetic Epidemiology, the Wallenberg Foundation, and the European Foundation for the Study of Diabetes.

Dr. Lyssenko reports receiving consulting fees from Tethys; Dr. Altshuler, receiving consulting fees from Merck, Eisai, and MPM Capital and grant support from Novartis, receiving an unrestricted grant from the Bristol-Myers Squibb Foundation, and having an equity interest in MPM Capital; and Dr. Groop, receiving consulting fees from Tethys, Eli Lilly, GlaxoSmithKline, Aventis, and Sanofi, lecture fees from Aventis and Sanofi, and grant support from Novartis. No other potential conflict of interest relevant to this article was reported.

We thank the patients for their participation and Anna Berglund, Margareta Svensson, and Malin Neptin for their excellent technical assistance.

REFERENCES

1. Lyssenko V, Almgren P, Anevski D, et al. Predictors of and longitudinal changes in insulin sensitivity and secretion preceding onset of type 2 diabetes. Diabetes 2005;54:166-74.
2. Tirosh A, Shai I, Tekes-Manova D, et al. Normal fasting plasma glucose levels and type 2 diabetes in young men. N Engl J Med 2005;353:1454-62. [Erratum, N Engl J Med 2006;354:2401.]
3. 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.
4. 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.
5. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 2000;26:76-80.
6. Gloyn AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 2003;52:568-72.
7. 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.
8. 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.
9. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316:889-94.
10. Sandhu MS, Weedon MN, Fawcett KA, et al. Common variants in WFS1 confer risk of type 2 diabetes. Nat Genet 2007;39:951-3.
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.
12. 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.
13. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007;39:770-5.
14. 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.
15. 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.
16. Groop L, Forsblom C, Lehtovirta M, et al. Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. Diabetes 1996;45:1585-93.
17. Eriksson KF, Lindgarde F. Prevention of type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise: the 6-year Malmö feasibility study. Diabetologia 1991;34:891-8.
18. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 1999;22:1462-70.
19. Hanson RL, Pratley RE, Bogardus C, et al. Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. Am J Epidemiol 2000;151:190-8.
20. Bergman RN, Ader M, Huecking K, Van Citters G. Accurate assessment of beta-cell function: the hyperbolic correction. Diabetes 2002;51:Suppl 1:S212-S220.
21. Heding LG. A simplified insulin radioimmunoassay method. In: Donato L, Hulhaud G, Sirchis J, eds. Labelled proteins in tracer studies. Brussels: European Atomic Energy Community, 1966:345-50.
22. Lyssenko V, Almgren P, Anevski D, et al. Genetic prediction of future type 2 diabetes. PLoS Med 2005;2(12):e345.
23. Tang K, Fu D, Kötter S, Cotter RJ, Cantor CR, Köster H. Matrix-assisted laser desorption/ionization mass spectrometry of immobilized duplex DNA probes. Nucleic Acids Res 1995;23:3126-31.
24. 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;23:2109-23.
25. 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.
26. Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd ed. Hillsdale, NJ: Lawrence Erlbaum, 2003.
27. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180-7.
28. 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.
29. 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.
30. Cauchi S, Meyre D, Durand E, et al. Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. PLoS ONE 2008;3(5):e2031.

POWERPOINT SLIDES OF *JOURNAL* FIGURES AND TABLES

At the *Journal*'s Web site, subscribers can automatically create PowerPoint slides. In a figure or table in the full-text version of any article at **www.nejm.org**, click on Get PowerPoint Slide. A PowerPoint slide containing the image, with its title and reference citation, can then be downloaded and saved.