의학박사 학위논문

Effect of Genetic Predisposition and Gene-Environment Interaction on the Incidence of Type 2 Diabetes Mellitus in a Community-based Cohort Study

제 2 형 당뇨병 발생에 대한 유전적 소인 및 유전-환경 상호작용의 영향 조사: 지역사회기반 코호트 연구

2015 년 8 월

서울대학교 대학원
의학과 예방의학 전공
박 혜 인

A thesis of the Degree of Doctor of Philosophy

제 2 형 당뇨병 발생에 대한 유전적 소인 및 유전-환경 상호작용의 영향 조사: 지역사회기반 코호트 연구

Effect of Genetic Predisposition and Gene-Environment Interaction on the Incidence of Type 2 Diabetes Mellitus in a Community-based Cohort Study

August 2015

The Department of Preventive Medicine,
Seoul National University
College of Medicine
Hye Yin Park

Effect of Genetic Predisposition and Gene-Environment Interaction on the Incidence of Type 2 Diabetes Mellitus in a Community-based Cohort Study

by
Hye Yin Park

A thesis submitted to the Department of Preventive Medicine in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Medicine(Preventive Medicine) at Seoul National University College of Medicine

August 2015

Approved by Thesis Committee:

Professor _____**Chairman**

Professor _____**Vice chairman**

Professor _____

Professor _____

Professor _____

제 2 형 당뇨병 발생에 대한 유전적 소인 및 유전-환경 상호작용의 영향 조사: 지역사회기반 코호트 연구

Effect of Genetic Predisposition and Gene-Environment Interaction on the Incidence of Type 2 Diabetes Mellitus in a Community-based Cohort Study

Hye Yin Park

The Department of Preventive Medicine,
Seoul National University College of Medicine

ABSTRACT

Introduction: Contribution of genetic predisposition to risk prediction of type 2 diabetes mellitus (T2DM) was investigated, and genotype variation was tested for gene-environment interaction effect using a prospective study in middle-aged adults in South Korea.

Methods: From a community cohort of 6,910 subjects with 8 years' follow-up, genetic predisposition score with subsets of 4, 16, 36 selected single nucleotide polymorphisms (SNP) (genetic predisposition score, GPS) in association with T2DM were determined, and their effect was evaluated using risk prediction models. Also, genetic and environmental factors were tested for their interaction effect on T2DM risk.

Results: Sixteen SNPs were found to be in significant association with T2DM in the study population, and hazard ratios of GPSs above median risk allele scores were 1.19 (95% confidence intervals: 1.04-1.36), 1.23 (1.08-1.41), 1.42 (1.19-1.70) with GPS-4, GPS-16, GPS-36, respectively. Significant changes in C-statistics for discrimination were observed in both subject groups tested with GPS-16 and GPS-36, and while significant net reclassification index (NRI) in subpopulations weakened upon addition of glycated hemoglobin (HbA1c) in subjects with GPS-4, those with information on GPS-16 and GPS-36 remained robust. Significant positive gene-environment interactions by additional scales were observed for GPS-16 and GPS-36 with environmental variables such as family history of DM, high-density lipoprotein cholesterol and HbA1c, and multiplicative scales were also significant for family history

of DM.

Conclusions: From a cohort of middle-aged Koreans, subjects with genetic predisposition information on multiple numbers of SNPs showed significant model discrimination and reclassification, even after adjusting for HbA1c, a stronger predictor for T2DM risk. For gene-environment interaction, although some effect is suggested from the results of the current study, further investigations with larger number of subjects or in an independent population is suggested.

Keywords: Diabetes mellitus, gene-envrionment interaction; genetic predisposition; risk prediction

Student number: 2012-31144

CONTENTS

LIST OF TABLES AND FIGURES

LIST OF ABBREVIATIONS

AP	attributable proportion
AUC	area under the curve
CI	confidence interval
DBP	diastolic blood pressure
FPG	fasting plasma glucose
GPS	genetic predisposition score
GWAS	genome-wide association studies
HbA1c	glycated hemoglobin A1c
HDL-C	high-density lipoprotein cholesterol
HR	hazard ratio,
IDI	integrated discrimination improvement
KoGES	Korean Genome and Epidemiology Study
NRI	net reclassification improvement
OR	odds ratio
P/E	physical exercise
RAF	risk allele frequency
RERI	relative excess risk due to interaction
RR	relative ratio
S	synergy index
SBP	systolic blood pressure
SNP	single nucleotide polymorphisms
T2DM	type 2 diabetes mellitus

INTRODUCTION

Although type 2 diabetes mellitus (T2DM), a prevalent and complex disease, is known to be caused by combinations of genes and environmental factors, the genetic contribution is not yet fully comprehended. Dozens of single nucleotide polymorphisms (SNPs) in association with T2DM were identified by genome-wide association studies (GWAS), such as *PPAR*, *KCNJ11*, *TCF7L2*, *CDKAL1*, *CDKN2A/B*, and *FTO*[1, 2]. Contribution of SNPs to development of T2DM, however so far, have been shown to be limited, with reported estimates of genetic contribution to heritability for T2DM unveiled by GWAS as 6-15 %.

Accordingly, recommended research expansion are directed to investigating the genetic contribution with increased diversity such as different ethnicities, or to undiscovered genetic field such as rare alleles or copy-number variations, and interaction analyses[3, 4].

1. Effect of genetic predisposition on T2DM prediction

Genetic predisposition, expressed in scores of combined risk alleles of SNPs discovered from GWAS, is a concept used in researches on utilizing genotype information for practical use, including construction of risk prediction models[5-7].

So far, reports on genetic predisposition, compared to common non-genetic risk factors, have been shown to contribute to inconsistent or limited improvement in prediction ability on risk of T2DM incidence[3, 8]. Some well-known prospective studies with relatively long periods of follow-up and information on multiple numbers of SNPs have also examined on this issue (Table 1).

In Framingham Offspring Study, the first investigation on subjects at Period 1 to 3 showed non-significant improvement in discrimination or reclassification (*P* for contrast in change in area under the curve (ΔAUC) = 0.49; *P* for net reclassification index (NRI), 0.17), although the effect of the multiple SNPs were significant (hazard ratio (HR) (95% confidence interval (CI)): 1.11 (1.05-1.17)[5]. In a more recent study with subjects at Period 1 to 4 and with increased number of tested SNPs, authors reported significant NRI at subjects under 50 years of age at baseline (NRI 10.2 %, *P* = 0.001), suggesting for importance of genotype information, especially in the younger population[9]. This hypothesis was also supported by another cohort study on young adults that showed significant improvement in reclassification (continuous NRI (95% CI): 0.285 (0.126-0.433))[10]. In other well-documented studies in Scandinavia and Britain, significant ΔAUCs were reported, although reclassification was omitted or was found to be non-significant[6, 7]. Although not a community cohort, a recent case-registry on gestational DM patients in South Korea with genotype information on 48 SNPs reported significant shift of AUC (0.741 to 0.775, *P* = 0.015) and of NRI improvement (0.430 (0.218-0.642)). Although results may be

exaggerating when interpreting to the general population, these results nonetheless may suggest influence of genetic variation on predicting risk of T2DM incidence[11]. Results form case-cohort or case-control studies are few in number and also inconsistent, while few studies are subjected to Asian population[12-14].

In addition to the ongoing research reports that are mostly based on Caucasian population[15], evaluating the significance of genetic predisposition on the predictive performance on T2DM incidence to wider characteristic range of study populations including non-European subjects has been strongly suggested[16]. Moreover, to my knowledge, investigation of genetic predisposition on T2DM prediction ability in South Koreans at a community level has not been carried out previously. The current study is also one of the few researches that examined combined genetic and non-genetic dataset from a large prospective study constructed to represent national-level health status.

Table 1. Characteristics of studies on influence of GPS on T2DM risk

Author, yr	Study name	Population at baseline	F/U period	Incident T2DM	Variables tested	No. of SNPs	HR/OR (95% CI) by GPS	C-statistics (AUC)	Reclassification (NRI/IDI, *P*/95% CI)
Meigs JB et al., 2008 [5]	FOS (Framingham Offspring Study); Period 1-3	(N=2,377)	28 yrs.	255 (10.7%)	age, sex, family history of DM, BMI, FPG, SBP, HDL-C, TG	18	1.11 (1.05-1.17)	0.900 → 0.901 (*P* = 0.49)	2.13% (*P* = 0.17)
Lyssenko V et al., 2008 [7]	MPP(Malmo Preventive Project), Botnia study	(N=18,831)	23.5 yrs.	2,201 (11.7%)	family history of DM, BMI, liver enzymes, current smoking, insulin deposition index	11	1.12 (1.08-1.15) 0.94 (0.84-1.04)	0.74 → 0.75 (*P* = 0.0001)	N/A
de Miguel-Yanes JM et al., 2011 [9]	FOS (Framingham Offspring Study); Period 1-4	46±12.7 yrs. (N=3,471)	11358 person-time	346 (12.8%)	sex, family history of DM, BMI, FPG, SBP, HDL-C, TG	40	1.15 (1.09-1.22) <50 yrs: 1.24 (1.13-1.36)	0.903 → 0.906 (*P* = 0.04) <50yrs: 0.908 → 0.911 (*P* = 0.3)	1.8% (*P* = 0.2) <50 yrs: 10.2% (*P* = 0.001)
Talmud PJ et al., 2010 [6]	Whitehall II Study	35-55 yrs. (N=5,535)	10 yrs.	302 (5.5%)	Cambridge risk score: age, sex, family history of DM, BMI, smoking, drug treatment FOS risk score: age, sex, parental history of DM, BMI, HDL-C, TG, FPG	20	[OR, unadjusted] 2.3 (1.5-3.8)	0.72 → 0.73 0.78 → 0.78	-1.1% (*P* = 0.66) 0.2% (*P* = 0.94)

Table 1. Characteristics of studies on influence of GPS on T2DM risk (continued)

Author, yr	Study name	Population at baseline	F/U period	Incident T2DM	Variables tested	No. of SNPs	HR/OR (95% CI) by GPS	C-statistics (AUC)	Reclassification (NRI/IDI, *P*/95% CI)
Vassy JL et al., 2012 [10]	CARDIA (Coronary Artery Risk Development in Young Adults)	18-30 yrs. (N=2439)	23.9 yrs.	215 (8.8%)	age, sex, race, parental history of DM, BMI, MAP, FPG, HDL-C, TG	38	1.08 (1.04-1.13)	0.824 → 0.829 (*P*=0.26)	[cNRI] 0.285 (0.126-0.433)
Kwak SH et al., 2013 [17]	Registry of GDM women	(N=395)	med. 45mo.	116 (29.4%)	age, family history of DM, pre-pregnancy BMI, MAP, insulin	48	1.66 (1.30-2.13)	0.741 → 0.775 (*P*=0.015)	[cNRI] 0.430 (0.218-0.642)
Schulze MB et al., 2009 [12]	EPIC-Potsdam Study; case-cohort	579 cases, 7963 controls	N/A	N/A	German DRS(i.e. lifestyle), FPG, HbA1c	20	N/A	0.8926 → 0.8928 (*P*=0.7361)	[IDI] 0.0014 (*P*=0.36)
					German DRS(i.e. lifestyle), FPG, HbA1c, HDL-C, GTP, ALT, hs-CRP			0.9000 → 0.9002 (*P*=0.6868)	[IDI] 0.0015 (*P*=0.034)
Tam CHT et al., 2013 [13]	Hong Kong Diabetes Registry; case cohort	5,882 cases, 2,569 controls	N/A	N/A	N/A (selected by school(adolescent), occupation group(adult), age group(elderly))	8	N/A (individual ORs: 1.07 to 2.09)	N/A	11.0% (*P* <.0001)
Imamura M et al., 2013 [14]	Multi-center registry	2,613 cases, 1,786 controls	N/A	N/A	age, sex, BMI	49	GPS-49: 1.13 (1.11-1.15) GPS-10: 1.26 (1.22-1.31)	0.768 → 0.773 (GPS-10 → GPS-49)	N/A

NA, not applicable; N/A, not available

2. Gene-environment interactions

As a leap from common alleles, suggestions on investigating genetic contribution on development of T2DM have extended to rare/minor allele frequency, copy-number variations (CNVs), epigenetics, or gene-gene interactions and gene-environment interaction[3, 18]. While most of the prominent recommendations usually demand multiple large-scale datasets, number of subjects needed to test for interaction effect may depend on the size of the effect estimate, fraction proportion of cases over the subgroups of gene-gene or gene-environment risk groups[19].

Meanwhile, a few studies have also reported some interaction effect between behavioral risk factors and genetic polymorphisms, as well as significant effect of lifestyle intervention in subjects with high genetic risk[3, 20]. The current research is tended to thoroughly investigate on the genetic variation information in the study population, and there is sufficient information on environmental variables in the analyzing dataset. Thus, expanding the research to confirming the influence of gene-environment (*g*e*) interaction on the present population of interest may be prospective, as any *g*e* interaction effect discovered in the population may be of practical value in the extent of policy making.

3. Hypotheses

With the incidence of T2DM as the outcome, two hypothesis was formulated using a well-designed prospective data from a community-based cohort study of middle-aged South Koreans. First, that contribution of information on genetic variants would significantly influence risk prediction ability of T2DM incidence; for this hypothesis, an 8-years prediction model was constructed. Second, that there would be a significant *g*e* interaction between the investigated genetic variants and known environmental factors in the same study population; for this hypothesis, information on genetic variation validated from the first hypothesis as well as baseline variables of the study subjects were tested.

MATERIALS AND METHODS

1. Study subject

The Ansung-Ansan Cohort Study, one of the three prospective community-based cohort studies from the Korean Genome and Epidemiology Study (KoGES), begun with 10,038 subjects aged 40 to 69 years at baseline (2001-2003). Whole-genome sequencing using Affymetrix 500K Array (Affymetrix, Santa Clara, CA, USA) was performed in 8,842 randomly selected subjects during the baseline investigation period, and unphased genotypes were imputed with Japanese+Chineses HapMap phase 2 haplotype panel using IMPUTE version 2 (http://mathgen.stats.ox.ac.uk/impute). Follow-up studies are carried out in 2-year intervals, at 2003-2005, 2005-2007, and so on. In this study, 8-years follow-up data was used, collected biennially until the 4th follow-up (2009-2011). Details regarding the KoGES, including methods and quality control for the genotyping, have been described in previous reports[21, 22].

At baseline, 2 subjects without any information needed for T2DM definition, 683 subjects with history of DM diagnosis/treatment or in current oral hypoglycemic medication/insulin therapy for DM, and 544 subjects with glycated hemoglobin (HbA1c) ≥6.5% or fasting plasma glucose (FPG) ≥7.0 mM/l or plasma glucose level 2-h after ingestion of 75g oral glucose load (2 hr-OGTT) ≥11.1mM/l were excluded. From 8,809 subjects at baseline, 954

(10.8%) subjects were eliminated due to follow-up loss after fourth follow-up in 2009-2011. Of the remaining 7,855 subjects, another 945 (12.0%) subjects who had not been selected for genotyping procedures at baseline were excluded. Thus 6,910 subjects remained for analysis (Figure 1).

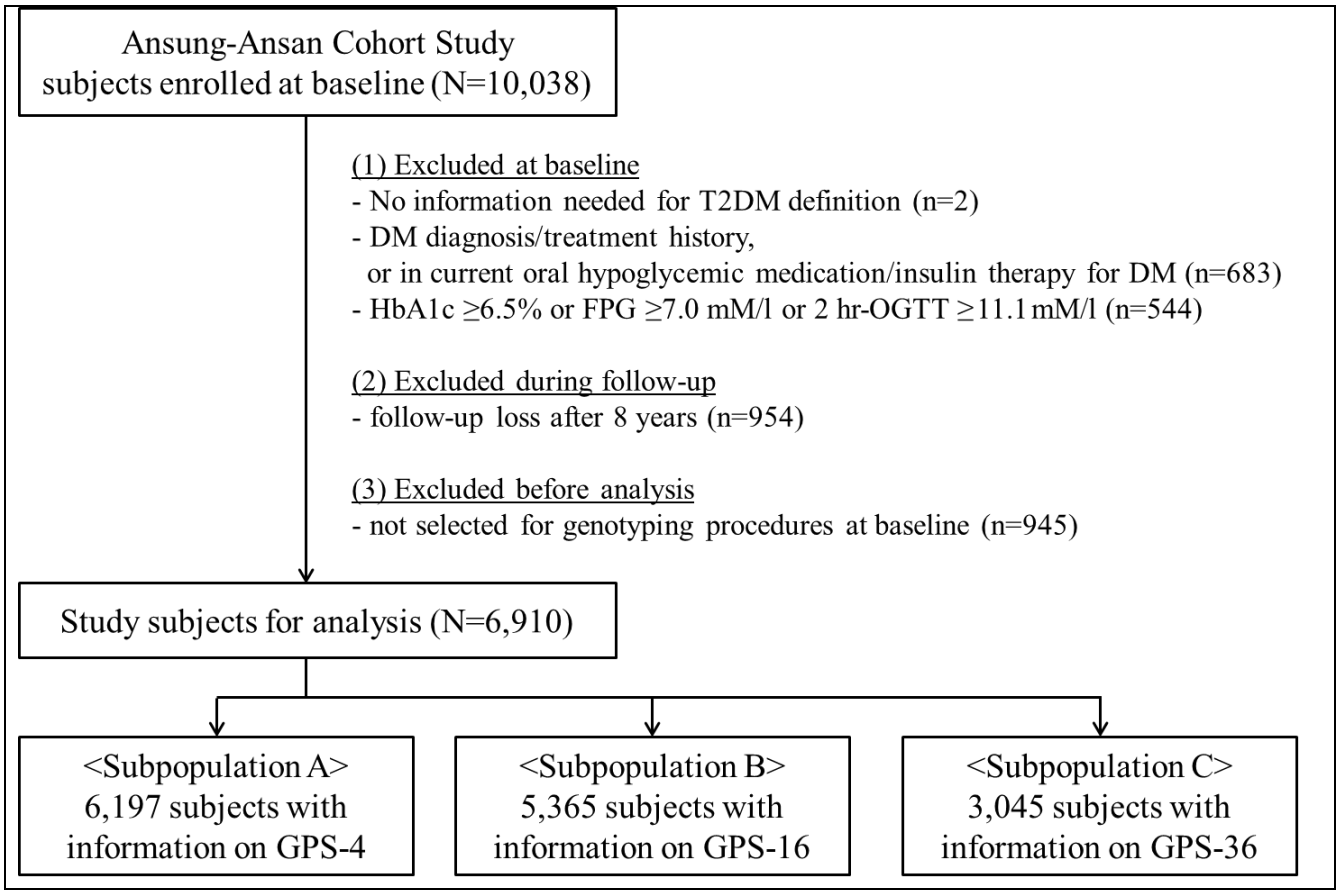

Figure 1. Flow chart showing selection of subjects included in the analysis.

2. Case definition

Incident T2DM cases at each follow-up was identified as corresponding to at least one of the following definitions: HbA1c ≥6.5 %, FPG ≥7.0 mM/l, 2 hr-OGTT ≥11.1 mM/l, or in treatment state for T2DM with insulin or oral hypoglycemic medication since the last follow-up or two years' period.

3. SNP selection

Approximately sixty SNPs have been reported to be in association with T2DM in Korean or East Asian population, from GWAS meta-analysis or candidate gene analysis that partly or entirely used KoGES baseline data[21, 23, 24]. In this study, PLINK 1.7 (http://pngu.mgh.harvard.edu/purcell/plink/) was used to extract relevant SNPs from the raw genotype data of Ansung-Ansan population, from both genotyped and imputed sequencing datasets. Haploview (https://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview) and TagSNP (http://snpinfo.niehs.nih.gov/snpinfo/snptag.htm) was used to calculate rare allele frequency (RAF) and test for linkage disequilibrium of extracted SNPs. Odds ratios (OR) and 95% confidence intervals (CI) of risk alleles of each SNPs on baseline or incident T2DM in the 8,842 cohort population was analyzed using logistic regression analyses.

A score system for scaling the concept of genetic predisposition was constructed using numbers of risk alleles of selected SNPs. The combinations of scores were set as 'genetic predisposition scores (GPSs)', ranging from 0 to number of selected SNPs multiplied by 2. Thus, higher GPS indicate a higher genetic predisposition to T2DM[25]. Different subsets of GPSs were constructed based on p-values or false discovery rate (FDR) of each SNP's effect on T2DM risk.

4. Statistical analysis

Cox's proportional hazard functions were used to estimate HR and their 95% CIs of risk of T2DM incidence in the study population. For selection of variables in the prediction model, selection procedures were carried out using the survival analysis. First, all *a priori* covariates were tested in a univariate Cox regression model at significant level of p-value ≤ 0.2, then all significant and non-significant covariates in multivariate Cox regression models were fitted with p-values ≤ 0.15 required for inclusion in backward and forward selection procedures, respectively. Finally, stepwise selection with the selected covariates with p-value ≤ 0.15 was used to attain the main-effects model. Likelihood ratio test was used for all covariate inclusion/exclusion decisions[26]. Modeling was carried out in each subpopulation of study subjects grouped by availability of information on the various GPSs.

From the full model with all selected variables, each model variables were investigated for their relationship with different GPSs in different

subpopulations using linear or logistic regression analysis. Also, to set the order of subset models for variable adjustments, discrimination statistics upon addition of each variable in comparison to a base model in the logistic function, with incident DM cases as the independent variable, was analyzed.

For all prediction models, discrimination and risk reclassification were evaluated after adjusting for GPS in the models. At the same time, tested models were divided in subsets in accordance with previous literature[5, 10]. C-statistics were used to test for model discrimination, and NRI was analyzed to examine risk reclassification upon addition of selected risk alleles[27].

For gene-environment interaction analyses, all selected variables in the hazard function model were tested as possible environmental factors. Multiplicative interaction was calculated using p-values for trend as well as p-values for interaction in the survival function, and relative excess risk due to interaction (RERI), attributable proportion (AP), and synergy index (S) by logistic regression analysis were also calculated to determine additive interaction[28-31]. Applied from the literature, equations for the additive interaction scales are as follows[32, 33]:

$$RERI = RR_{G+E+} - RR_{G+E-} - RR_{G-E+} + 1$$

$$AP = \frac{RERI}{RR_{G+E+}}$$

$$S = \frac{RR_{A+B+} - 1}{(RR_{G+E-} - 1) + (RR_{G-E+} - 1)}$$

where RR_{G+E+}, RR_{G+E-}, RR_{G-E+} are relative risk of T2DM incidence when genetic (G) and environmental (E) factors are present/absent or at high/low risk (+/-). RERI would indicate part of the total effect due to interaction; RERI equal to 0 means no interaction, while >0 and <0 would mean positive and negative interactions, respectively (range, - ∞ to +∞). AP refers to proportion of the combined effect due to interaction, and result interpretations are same as RERI (range, -1 to +1). S would indicate ratio of combined effect over individual effects; S equal to 1 means no interaction, while >1 and <1 means positive and negative interactions, respectively (range, 0 to +∞).

The analyzed multiplicative interaction effect was used to test the sample size[19]. In brief, with X_1 and X_2 as binary variables tested for interaction, the total number of subjects (n) required to achieve a power of 1 - β at type I error rate α for a two-sided test is as follows:

$$n = \frac{(z1-\alpha/2 + z1-\beta)2G}{[log(\theta)]2\psi(1-p)p(1-\rho2)}$$

ψ : proportion of subjects with incident T2DM

$$\rho = \text{corr}(X_1,X_2) = (p_1 - p_0) \times \sqrt{\frac{q(1-q)}{p(1-p)}}$$

$$p = Pr(X_1 = 1),\ q = Pr(X_2 = 1)$$

$$p_0 = Pr(X_1 = 1 \mid X_2 = 0)$$

$$p_1 = Pr(X_1 = 1 \mid X_2 = 1)$$

$$G = \frac{[(1-q)(1-p0)p0 + q(1-p1)p1]2}{(1-q)q(1-p0)p0(1-p1)p1}$$

A two-tailed p-value <0.05 indicated statistical significance. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC, USA), Stata/SE 13.0 (StataCorp LP, College Station, TX, USA), and R version 3.1.0 (http://cran.rproject.org).

5. Ethical statement

Informed written consent was obtained from all participants, and the study protocol was approved by the institutional review board of the Korea Centers for Disease Control and Prevention (KCDC) as well as Seoul National University Hospital (IRB No. 1306-046-495).

RESULTS

1. Risk loci characteristics

Thirty-eight SNPs reported relevant with T2DM were extracted from the whole genome sequencing data of Ansung-Ansan cohort. By logistic regression analysis, 17 SNPs, rs7593730 (*RBMS1*; HR 1.10 (95% CI 1.003-1.21)), rs1470579 (*IGF2BP2*; 1.13 (1.05-1.21)), rs1801282 (*PPARG*; 1.19 (1.01-1.40)), rs7754840 (*CDKAL1*; 1.22 (1.14-1.30)), rs9465871 (*CDKAL1*; 1.21 (1.13-1.29)), rs864745 (*JAZF1*; 1.11 (1.03-1.20)), rs10811661 (*CDKN2A/B*; 1.15 (1.08-1.24)), rs5015480 (*HHEX*; 1.13 (1.04-1.23)), rs5215 (*KCNJ11*; 1.08 (1.003-1.16)), rs1359790 (*SPRY2*; 1.10 (1.02-1.18)), rs6780569 (*UBE2E2*; 1.12 (1.02-1.23)), rs7756992† (*CDKAL1*; 1.22 (1.13-1.31)), rs4607517 (*GCK*; 1.14 (1.05-1.24)), rs13266634 (*SLC30A8*; 1.10 (1.03-1.18)), rs7903146 (*TCF7L2*; 1.30 (1.05-1.61)), rs1552224 (*CENTD2*; 1.17 (1.01-1.36)), rs2237892 (*KCNQ1*; 1.20 (1.11-1.29)) showed significant association with T2DM incidence in the subjects (Table2). Four of the 17 SNPs, rs7754840, rs9465871, rs10811661, rs7756992, and rs2237892, remained robust even after testing for false discovery rate (FDR) value of 0.00011 (Table 2).

Table 2. Characteristics of selected risk loci for type 2 diabetes mellitus

SNP	Chr	Locus	Risk allele	RAF of case/control	OR (95% CI) in Ansan-Anseong	Reported OR (95% CI) in East Asians*	Reported OR (95% CI) in Caucasians
18 SNPs analyzed by Affymetrix 500K							
rs10923931	1	*NOTCH2*	T	0.04/0.03	1.05 (0.86-1.29)	1.05 (0.92-1.20)	1.13 (1.08-1.17) [1]
rs7593730	2	*RBMS1*	C	0.83/0.83	1.1 (1.003-1.21)	1.03 (0.97-1.09)	1.11 (1.08-1.16) [34]
rs1470579	3	*IGF2BP2*	C	0.33/0.31	1.13 (1.05-1.21)	1.13 (1.08-1.19)	1.17 (1.11-1.23) [35]
rs1801282	3	*PPARG*	C	0.96/0.95	1.19 (1.01-1.4)	1.13 (1.01-1.28)	1.14 (1.08-1.20) [35]
rs4607103	3	*ADAMTS9*	C	0.62/0.61	1.06 (0.99-1.14)	0.99 (0.95-1.04)	1.09 (1.06-1.12) [1]
rs831571	3	*PSMD6*	C	0.63/0.63	1.05 (0.98-1.13)	1.09 (1.06-1.12)	NA
rs7754840	6	*CDKAL1*	C	0.48/0.47	1.22 (1.14-1.3)	1.20 (1.14-1.25)	1.12 (1.08-1.16) [36]
rs9465871	6	*CDKAL1*	C	0.56/0.54	1.21 (1.13-1.29)	1.14 (1.09-1.18)	NA
rs864745	7	*JAZF1*	T	0.74/0.72	1.11 (1.03-1.2)	1.06 (1.00-1.12)	1.10 (1.07-1.13) [1]
rs10811661	9	*CDKN2A/B*	T	0.58/0.56	1.15 (1.08-1.24)	1.21 (1.14-1.28)	1.20 (1.14-1.25) [36]
rs10906115	10	*CDC123/CAMK1D*	A	0.54/0.53	1.05 (0.98-1.13)	1.09 (1.04-1.14)	1.13 (1.08-1.18) [24]
rs5015480	10	*HHEX*	C	0.19/0.19	1.13 (1.04-1.23)	1.16 (1.1-1.23)	1.19 (1.11-1.28) [37]
rs5215	11	*KCNJ11*	C	0.41/0.39	1.08 (1.003-1.16)	1.13 (1.08-1.18)	1.14 (1.10-1.19) [38]
rs1531343	12	*HMGA2*	C	0.11/0.11	1.06 (0.95-1.18)	1.06 (0.99-1.14)	1.10 (1.07-1.14) [2]
rs7961581	12	*TSPAN8/LGR5*	C	0.22/0.23	0.99 (0.91-1.07)	1.01 (0.95-1.06)	1.09 (1.06-1.12) [1]
rs1359790	13	*SPRY2*	G	0.71/0.7	1.1 (1.02-1.18)	1.02 (0.97-1.08)	1.15 (1.10-1.20) [24]
rs1436955	15	*C2CD4A/C2CD4B*	C	0.7/0.69	1.06 (0.98-1.14)	1.13 (1.06-1.21)	NA
rs9939609	16	*FTO*	A	0.87/0.88	1 (0.9-1.11)	1.15 (1.08-1.22)	1.15 (1.09-1.23) [39]

Table 2. Characteristics of selected risk loci for type 2 diabetes mellitus (continued)

SNP	Chr	Locus	Risk allele	RAF of case/control	OR (95% CI) in Ansan-Anseong	Reported OR (95% CI) in East Asians*	Reported OR (95% CI) in Caucasians
20 SNPs from HapMap imputation							
rs340874	1	*PROX1*	C	0.37/0.35	1.06 (0.99-1.15)	1.08 (1.03-1.14)	1.07 (1.05-1.09) [40]
rs243021	2	*BCL11A*	A	0.67/0.66	1.03 (0.96-1.11)	1.05 (1.00-1.10)	1.08 (1.06-1.10) [2]
rs2943641	2	*IRS1*	C	0.94/0.95	0.99 (0.85-1.15)	1.12 (1.03-1.22)	1.19 (1.13-1.25) [41]
rs6780569	3	*UBE2E2*	G	0.83/0.81	1.12 (1.02-1.23)	1.13 (1.07-1.20)	1.21 (1.14-1.30) [42]
rs10010131	4	*WFS1*	G	0.98/0.98	0.92 (0.72-1.17)	1.00 (0.91-1.10)	1.11 (1.05-1.16) [43]
rs7756992†	6	*CDKAL1*	G	0.56/0.54	1.22 (1.13-1.31)	1.14 (1.09-1.18)	1.19 (1.13-1.27) [44]
rs2191349	7	*DGKB*	T	0.68/0.68	1.07 (0.99-1.16)	1.11 (1.05-1.16)	1.06 (1.04-1.08) [40]
rs4607517	7	*GCK*	A	0.23/0.21	1.14 (1.05-1.24)	1.03 (0.97-1.09)	1.07 (1.05-1.10) [40]
rs972283	7	*KLF14*	G	0.7/0.69	1.02 (0.94-1.1)	0.99 (0.93-1.06)	1.07 (1.05-1.10) [2]
rs13266634	8	*SLC30A8*	C	0.59/0.59	1.1 (1.03-1.18)	1.11 (1.06-1.16)	1.15 (1.12-1.19) [45]
rs896854	8	*TP53INP1*	T	0.29/0.29	1.001 (0.93-1.08)	1.07 (1.02-1.12)	1.06 (1.04-1.09) [2]
rs13292136	9	*CHCHD9*	C	0.9/0.89	1.05 (0.94-1.18)	0.99 (0.92-1.07)	1.11 (1.07-1.15) [2]
rs12779790	10	*CDC123/CAMK1D*	G	0.11/0.1	1.09 (0.97-1.22)	1.12 (1.02-1.23)	1.11 (1.07-1.14) [1]
rs7903146	10	*TCF7L2*	T	0.03/0.02	1.3 (1.05-1.61)	1.16 (1.02-1.31)	1.37 (1.31-1.43) [36]
rs10830963	11	*MTNR1B*	G	0.44/0.44	1.06 (0.99-1.14)	0.99 (0.93-1.06)	1.09 (1.06-1.12) [40]
rs1552224	11	*CENTD2*	A	0.94/0.94	1.17 (1.01-1.36)	1.16 (1.06-1.27)	1.14 (1.11-1.17) [2]
rs2237892	11	*KCNQ1*	C	0.64/0.6	1.2 (1.11-1.29)	1.17 (1.11-1.23)	1.40 (1.34-1.47) [46]
rs231362	11	*KCNQ1*	G	0.89/0.88	1.06 (0.95-1.19)	1.10 (1.00-1.20)	1.08 (1.06-1.10) [2]
rs2334499	11	*INS/IGF2B*	T	0.82/0.82	1.01 (0.92-1.1)	NA	1.35 (NA) [47]
rs7172432†	15	*C2CD4A/C2CD4B*	A	0.55/0.55	1.04 (0.97-1.12)	1.09 (1.04--1.15)	1.14 (1.09-1.20) [42]

NA, not available. Adjusted for age, sex, and body mass index.

*Referred from Cho YS et al., 2012[21], Ryoo H et al., 2011[23], and Shu XO et al., 2010[24].

†Rs7756992 and rs 7172432 are in linkage with rs9465871(|D'=0.977) and rs1436955(|D'=1), respectively.

SNPs under false discovery rate (FDR) = 0.00011 are underlined.

In constructing GPS, rs7756992 and rs7172432 were eliminated as they showed strong linkage with rs9465871 (D'=0.977, r^2=0.933) and rs1436955 (D'=1, r^2=0.627), respectively.

Three sets of GPSs were constructed with differently selected SNPs, i.e. (i) GPS-4 with four SNPs in significant association with T2DM in our study subjects, at p-values less than FDR (i.e. < 0.00011) (range 0-8); (ii) GPS-16 with 16 SNPs in significant association at p-values <0.05 (range 0-32); (iii) GPS-36, with all tested SNPs (range 0-72).

2. Subject characteristics at baseline

Mean age of subjects were 51.8 years at baseline, and males accounted for 47.1 % of the total 6,910 study subjects, and over the 8-year follow-up, 1,240 (17.9 %) were defined as incident T2DM cases. As well as variables tested for prediction modeling (i.e. age, body mass index [BMI], triglyceride [TG], high-density lipoprotein cholesterol [HDL-C], etc.) GPSs were higher in incident diabetic cases compared to those who remained non-diabetic (p-values <.005). T2DM incident cases were older, had higher BMI, TG, FPG, HbA1c and lower HDL-C compared to non-cases. Cases also showed relatively higher risk allele scores, and higher proportion of hypertension (HTN), family history of DM, and so on (Table 3).

Table 3. Baseline characteristics of all study subjects

	All (N=6,910)	Case (n=1,240)	Control (n=5,670)	P-diff
Mean±SD				
Age (years)	51.77±8.79	53.4±8.78	51.41±8.76	<.0001
Body mass index (kg/m^3)	24.47±3.02	25.04±3.21	24.34±2.96	<.0001
HDL cholesterol (mM/L)	1.17±0.26	1.13±0.26	1.17±0.26	<.0001
Triglyceride (mM/L)	1.76±1.1	2.09±1.29	1.69±1.04	<.0001
Fasting glucose (mM/L)	4.61±0.5	4.88±0.61	4.56±0.45	<.0001
HbA1c (%)	5.55±0.35	5.77±0.36	5.51±0.33	<.0001
Risk alleles of 4 SNPs	4.37±1.67	4.51±1.67	4.34±1.66	0.0019
Risk alleles of 8 SNPs	17.72±2.51	17.63±2.52	18.15±2.42	<.0001
Risk alleles of 36 SNPs	40.54±3.53	41.07±3.58	40.43±3.51	0.0002
Average SBP (mmHg)	120.7±18.2	125.0±18.8	119.8±18.0	<.0001
Average DBP (mmHg)	80.0±11.4	82.31±11.5	79.47±11.3	<.0001
Average WC (cm)	82.13±8.68	84.16±8.78	81.68±8.6	<.0001
Average HC (cm)	93.47±5.91	94.32±5.98	93.29±5.87	<.0001
Frequency (%)				
Sex, male	3251 (47.1)	642 (19.7)	2609 (80.3)	0.0002
Current smoking, yes	1702 (24.9)	332 (19.5)	1370 (80.5)	0.0529
Current drinking, yes	3292 (48)	612 (18.6)	2680 (81.4)	0.1829
Regular exercise, no	4001 (58.5)	768 (19.2)	3233 (80.8)	0.0027
Family history of DM, yes	719 (10.4)	174 (24.2)	545 (75.8)	<.0001
Hypertension, yes	1982 (28.7)	490 (24.7)	1492 (75.3)	<.0001
Metabolic syndrome, yes	2256 (32.7)	599 (26.6)	1657 (73.4)	<.0001

Different number of subjects had available information on various GPSs, i.e. subsets of 6,197 subjects, 5,365 subjects, and 3,045 subjects had full information on GPS-4, GPS-16 and GPS-36, respectively. Thus, for carrying out further statistical analyses, we divided the total 6,910 subjects into Subpopulation A, B, and C, with the respective number of subjects with available information on GPS-4, GPS-16 and GPS-36 (Figure 1).

To investigate for any disparity in baseline characteristics among different subpopulations in comparison to total study subjects, we tested for statistical difference between cases with and without information for various GPSs in the three subpopulations. By t-test or chi-square analysis, p-values for difference in baseline characteristics of cases in Subpopulations A, B, and C in comparison to total number of subjects (N = 6,910) minus the subpopulations (numbers of compared subjects: (i) n = 713; (ii) n = 1,545; (iii) n = 3,865) were generally >0.05, except for current smoking, average hip circumferences in testing for cases with GPS-4 (P = 0.0413), GPS-16 (P = 0.0032) and GPS-36 (P = 0.0157) information, respectively. Additional analysis on comparing Subpopulation C with subjects from Subpopulation B minus Subpopulation C ((iv) n = 2,320) also showed non-significant differences of baseline characteristics (Table 4).

Table 4. Baseline characteristics of study subpopulations, case-only

	Testing for cases with GPS-4 information			Testing for cases with GPS-16 information		
	Subpopulation A (N=6,197)	(i) Total subjects – Subpopulation A (n=713)	*P-diff*	Subpopulation B (N=5,365)	(ii) Total subjects – Subpopulation B (n=1,545)	*P-diff*
Mean±SD						
Age (years)	53.42±8.79	53.22±8.74	0.7961	53.51±8.76	53.07±8.86	0.4463
Body mass index (kg/m^3)	25.06±3.19	25.44±3.71	0.2296	25.05±3.19	25.27±3.44	0.2948
HDL cholesterol (mg/dL)	43.67±9.86	44.52±10.57	0.3294	43.48±9.86	44.63±10.16	0.0795
Triglyceride (mg/dL)	184.6±116	185.4±96.83	0.9267	185.1±117.7	183.5±101.8	0.8135
Fasting glucose (mg/dL)	87.86±11.09	87.93±10.71	0.9481	87.88±11.1	87.83±10.86	0.9395
HbA1c (%)	5.77±0.36	5.75±0.38	0.6143	5.77±0.35	5.75±0.38	0.2281
Average SBP (mmHg)	124.9±18.81	125.9±18.5	0.5387	124.7±18.73	126.1±18.89	0.2546
Average DBP (mmHg)	82.32±11.56	82.31±10.94	0.9921	82.03±11.29	83.18±12.01	0.1276
Average WC (cm)	84.08±8.65	84.76±9.69	0.3752	84.1±8.58	84.35±9.34	0.6669
Average HC (cm)	94.24±5.91	94.9±6.48	0.2063	94.03±5.9	95.19±6.15	0.0032
Frequency (%)						
Sex, male	562 (51.5)	80 (53.7)	0.6177	473 (50.9)	169 (54.3)	0.2955
Current smoking, yes	302 (28.1)	30 (20.1)	0.0413	256 (27.9)	76 (24.6)	0.2518
Current drinking, yes	545 (50.4)	67 (45)	0.2125	464 (50.4)	148 (47.9)	0.4501
Regular exercise, no	682 (62.9)	86 (57.7)	0.2199	567 (61.4)	201 (64.8)	0.2842
Family history of DM, yes	149 (13.7)	25 (16.8)	0.3037	125 (13.5)	49 (15.8)	0.3122
Hypertension, yes	428 (39.2)	62 (41.6)	0.5773	354 (38.1)	136 (43.7)	0.0792
Metabolic syndrome, yes	589 (54)	77 (51.7)	0.5961	492 (53)	174 (55.9)	0.3605

Table 4. Baseline characteristics of study subpopulations, case-only (continued)

	Testing for cases with GPS-36 information				
	Subpopulation C (N=3,045)	(iii) Total subjects – Subpopulation C (n=3,865)	*P-diff*	(iv) Subpopulation B – Subpopulation C (n=2,320)	*P-diff*
Mean±SD					
Age (years)	53.75±8.76	53.14±8.8	0.2313	53.2±8.76	0.3439
Body mass index (kg/m^3)	25.03±3.13	25.16±3.35	0.4904	25.07±3.28	0.8401
HDL cholesterol (mg/dL)	43.57±10.08	43.92±9.85	0.5356	43.38±9.59	0.7750
Triglyceride (mg/dL)	191.2±130.7	180±99.62	0.0992	177.3±98	0.0634
Fasting glucose (mg/dL)	88.16±11.35	87.66±10.81	0.4320	87.53±10.79	0.3921
HbA1c (%)	5.78±0.35	5.76±0.37	0.1951	5.76±0.36	0.4116
Average SBP (mmHg)	125.5±18.44	124.7±19.01	0.5036	123.7±19.06	0.1556
Average DBP (mmHg)	82.44±10.94	82.22±11.87	0.7443	81.5±11.73	0.2088
Average WC (cm)	83.92±8.21	84.35±9.17	0.3860	84.34±9.05	0.4578
Average HC (cm)	93.84±5.91	94.67±6.01	0.0157	94.28±5.88	0.2652
Frequency (%)					
Sex, male	264 (50.5)	378 (52.7)	0.4355	209 (51.5)	0.7625
Current smoking, yes	140 (27.4)	192 (26.9)	0.8442	116 (28.6)	0.6769
Current drinking, yes	260 (50.5)	352 (49.2)	0.6643	204 (50.2)	0.9426
Regular exercise, no	317 (61.2)	451 (63.1)	0.5016	250 (61.7)	0.8693
Family history of DM, yes	69 (13.2)	105 (14.6)	0.4676	56 (13.8)	0.7905
Hypertension, yes	208 (39.8)	282 (39.3)	0.8757	146 (36)	0.2359
Metabolic syndrome, yes	282 (53.9)	384 (53.6)	0.8992	210 (51.7)	0.5063

3. Evaluation of effect of genetic predisposition on T2DM risk prediction

Different subpopulations with various GPSs resulted in selecting different model variables for risk prediction by the survival functions. For GPS-4, age, BMI, family history of T2DM, hypertension history, regular physical exercise, and clinical indices such as HDL-cholesterol, triglyceride, FPG, and HbA1c as well as GPS-4 were selected in Subpopulation A (N = 6,197), while HTN history was omitted in Subpopulation B (N= 5,365), and HTN history plus HDL-cholesterol was unselected in Subpopulation C (N = 3,045). For GPS-16, age, BMI, family history of T2DM, regular physical exercise, HDL-cholesterol, triglyceride, FPG, HbA1c were selected in both Subpopulations B and C, while for GPS-36, all variables in the case of GPS-16 except for HDL-cholesterol were selected. The HRs and 95% CIs of each model variables, including the GPSs, in the three subpopulations are shown in Table 5.

Hazard ratios for T2DM incidence for GPS-4, as a binary variable by median number of risk alleles (i.e. below or above 3 risk allele scores), are 1.19 (1.04-1.36), 1.19 (1.02-1.38), and 1.07 (0.88-1.30) in Subpopulations A, B, and C, respectively (1.05 (1.01-1.09), 1.06 (1.02-1.10), and 1.04 (0.98-1.09) with GPS-4 as a continuous variable by per risk allele increase, respectively). Because the effect estimates of GPS-4 were not significant in Subpopulation C, further analysis for investigating influence of GPS-4 was not carried out in this subpopulation. Estimates for T2DM risk above median (≥17 risk allele

scores) of GPS-16 are 1.23 (1.08-1.41) and 1.27 (1.06-1.52) in Subpopulations B and C, respectively (1.07 (1.04-1.09) and 1.08 (1.04-1.11) per risk allele increase, respectively). For GPS-36 in Subpopulation C, HRs were 1.42 (1.19-1.70) and 1.04 (1.02-1.07) for T2DM risk upon treating GPS-36 as categorical (i.e. below or above 40 risk allele scores) and continuous variables, respectively. Apart from GPSs, all selected variables showed significant protective or harmful relationship with risk of T2DM incidence over the follow-up of 8 years in the Ansung-Ansan cohort.

Table 5. Risk (HRs and 95% CI) of T2DM incidence with GPSs across different subpopulations

		with GPS-4		
		Subpopulation A (N=6,197)	Subpopulation B (N=5,365)	Subpopulation C (N=3,045)
Age		1.02 (1.01-1.03)	1.02 (1.01-1.03)	1.03 (1.01-1.04)
Family history of T2DM (Ref: No)	Yes	1.35 (1.13-1.61)	1.33 (1.10-1.62)	1.29 (0.99-1.68)
HTN history (Ref: No)	Yes	1.18 (1.03-1.35)	n/a	n/a
Regular exercise (Ref: Yes)	No	0.78 (0.69-0.89)	0.81 (0.7-0.92)	0.78 (0.65-0.94)
BMI (Ref: <23 kg/m3)	23-25	1.06 (0.90-1.24)	1.10 (0.92-1.30)	1.15 (0.91-1.46)
	25-30	1.13 (0.96-1.33)	1.11 (0.93-1.33)	1.17 (0.92-1.50)
	≥30	1.42 (1.06-1.91)	1.72 (1.25-2.35)	1.79 (1.18-2.72)
HDL-C (Ref: ≥50 mg/dL)	<35	1.29 (1.05-1.60)	1.33 (1.06-1.67)	n/a
	35-49	1.11 (0.95-1.30)	1.11 (0.94-1.32)	n/a
Triglyceride (Ref: <120 mg/dL)	120-150	1.21 (1.004-1.46)	1.21 (0.99-1.48)	1.32 (1.01-1.74)
	≥150	1.72 (1.47-2.00)	1.71 (1.45-2.03)	1.96 (1.57-2.46)
FPG (Ref: 90-100 mg/dL)	<90	0.55 (0.47-0.63)	0.55 (0.47-0.65)	0.57 (0.46-0.71)
	≥100	2.37 (1.94-2.90)	2.55 (2.05-3.17)	3.18 (2.39-4.22)
HbA1c (Ref: <5.5 %)	≥5.5	1.98 (1.69-2.31)	1.96 (1.66-2.33)	2.06 (1.65-2.58)
GPS, binary (Ref: <median)	≥median	1.19 (1.04-1.36)	1.19 (1.02-1.38)	1.07 (0.88-1.30)
GPS, continous		*1.05 (1.01-1.09)*	*1.06 (1.02-1.10)*	*1.04 (0.98-1.09)*

Median value of GPS-4 is 3 risk allele score. n/a, not applicable.

Table 5. Risk (HRs and 95% CI) of T2DM incidence with GPSs across different subpopulations (continued)

		with GPS-16		with GPS-36
		Subpopulation B (N=5,365)	Subpopulation C (N=3,045)	Subpopulation C (N=3,045)
Age		1.02 (1.01-1.03)	1.03 (1.02-1.04)	1.03 (1.02-1.04)
Family history of T2DM (Ref: No)	Yes	1.33 (1.09-1.61)	1.28 (0.98-1.66)	1.30 (1.00-1.69)
HTN history (Ref: No)	Yes	n/a	n/a	n/a
Regular exercise (Ref: Yes)	No	0.80 (0.70-0.92)	0.79 (0.66-0.95)	0.78 (0.65-0.93)
BMI (Ref: <23 kg/m3)	23-25	1.11 (0.93-1.32)	1.15 (0.91-1.45)	1.17 (0.93-1.48)
	25-30	1.13 (0.94-1.35)	1.17 (0.92-1.49)	1.20 (0.94-1.52)
	≥30	1.79 (1.31-2.44)	1.85 (1.23-2.80)	1.94 (1.28-2.93)
HDL-C (Ref: ≥50 mg/dL)	<35	1.32 (1.05-1.66)	1.27 (1.06-1.52)	n/a
	35-49	1.10 (0.93-1.30)	1.08 (1.04-1.11)	n/a
Triglyceride (Ref: <120 mg/dL)	120-150	1.23 (1.00-1.50)	1.32 (1.01-1.72)	1.31 (1.00-1.71)
	≥150	1.74 (1.48-2.05)	1.97 (1.60-2.43)	1.98 (1.61-2.44)
FPG (Ref: 90-100 mg/dL)	<90	0.55 (0.47-0.65)	0.57 (0.46-0.71)	0.58 (0.46-0.72)
	≥100	2.57 (2.07-3.19)	3.20 (2.41-4.26)	3.27 (2.46-4.34)
HbA1c (Ref: <5.5 %)	≥5.5	1.95 (1.65-2.31)	2.03 (1.62-2.54)	2.03 (1.63-2.54)
GPS, binary (Ref: <median)	≥median	1.23 (1.08-1.41)	1.27 (1.06-1.52)	1.42 (1.19-1.70)
GPS, continous		*1.07 (1.04-1.09)*	*1.08 (1.04-1.11)*	*1.04 (1.02-1.07)*

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively. n/a, not applicable.

Results in which the ‘environmental factors (selected variables other than the genetic factor, i.e. GPS)’ in the model were tested for their relationship with GPSs in different subpopulations by regression analysis, with each factor and GPS as independent and dependent variables, are shown in Table 6. Although not all associations showed statistical significance or a certain pattern of consistency, positive relationship was observed for both FPG and HbA1c with GPSs across all subpopulations (all P values for association between HbA1c and GPSs, <.0001; P values for FPG, <.01 in Subpopulations A and B, and <.0001 in Subpopulation C).

Table 6. Association between model variables and GPS, by linear or logistic regression analysis

		Subpopulation A (N=6,197)		Subpopulation B (N=5,365)			
		GPS-4		GPS-4		GPS-16	
		Beta(SE) / OR(95%CI)	*P*	Beta(SE) / OR(95%CI)	*P*	Beta(SE) / OR(95%CI)	*P*
Age	Crude	-0.073 (0.055)	0.1887	-0.053 (0.06)	0.3773	-0.039 (0.065)	0.5472
	Adjusted	-0.074 (0.06)	0.2162	-0.132 (0.068)	0.0522	-0.087 (0.061)	0.154
BMI	Crude	0.145 (0.051)	0.0045	0.173 (0.052)	0.0009	0.038 (0.067)	0.5731
	Adjusted	0.125 (0.055)	0.022	0.156 (0.056)	0.0051	0.037 (0.069)	0.5879
HDL-cholesterol	Crude	0.112 (0.077)	0.1479	0.023 (0.072)	0.751	0.142 (0.074)	0.0558
	Adjusted	0.177 (0.083)	0.0334	0.098 (0.077)	0.2024	0.066 (0.069)	0.3387
TG	Crude	-0.994 (0.746)	0.1827	-0.824 (0.688)	0.2315	0.12 (0.068)	0.0763
	Adjusted	-0.77 (0.793)	0.3312	-0.528 (0.729)	0.469	0.114 (0.071)	0.1072
FPG	Crude	0.291 (0.068)	<.0001	0.212 (0.065)	0.0012	0.292 (0.065)	<.0001
	Adjusted	0.271 (0.073)	0.0002	0.185 (0.07)	0.0086	0.199 (0.063)	0.0015
HbA1c	Crude	0.015 (0.003)	<.0001	0.014 (0.002)	<.0001	0.014 (0.003)	<.0001
	Adjusted	0.015 (0.003)	<.0001	0.014 (0.003)	<.0001	0.012 (0.002)	<.0001
Family history of DM	Crude	1.24 (1.03-1.49)		1.16 (0.95-1.41)		1.05 (0.84-1.33)	
	Adjusted	1.19 (0.99-1.43)		1.12 (0.92-1.37)		1.16 (0.91-1.47)	
HTN history	Crude	0.9 (0.8-1.01)		n/a		n/a	
	Adjusted	0.94 (0.82-1.06)		n/a		n/a	
Physical exercise	Crude	1.04 (0.93-1.16)		1.02 (0.91-1.15)		0.92 (0.79-1.06)	
	Adjusted	1.03 (0.92-1.15)		1.02 (0.9-1.15)		1.01 (0.87-1.17)	

Table 6. Association between model variables and GPS, by linear or logistic regression analysis (continued)

		Subpopulation C (N=3,045)			
		GPS-16		GPS-36	
		Beta(SE) / OR(95%CI)	*P*	Beta(SE) / OR(95%CI)	*P*
Age	Crude	-0.011 (0.048)	0.8165	-0.046 (0.046)	0.3137
	Adjusted	-0.069 (0.045)	0.1262	-0.11 (0.043)	0.0105
BMI	Crude	0.092 (0.05)	0.0682	0.143 (0.067)	0.0322
	Adjusted	0.137 (0.051)	0.0078	0.149 (0.068)	0.0295
HDL-cholesterol	Crude	0.056 (0.055)	0.3116	n/a	
	Adjusted	0.015 (0.051)	0.7645	n/a	
TG	Crude	-0.027 (0.051)	0.6036	0.081 (0.068)	0.2294
	Adjusted	-0.04 (0.053)	0.4543	0.057 (0.07)	0.4155
FPG	Crude	0.271 (0.048)	<.0001	0.267 (0.046)	<.0001
	Adjusted	0.204 (0.047)	<.0001	0.205 (0.044)	<.0001
HbA1c	Crude	0.013 (0.002)	<.0001	0.01 (0.002)	<.0001
	Adjusted	0.01 (0.002)	<.0001	0.008 (0.002)	<.0001
Family history of DM	Crude	1.22 (1.02-1.46)		1.05 (0.84-1.33)	
	Adjusted	1.19 (0.99-1.43)		1.01 (0.8-1.28)	
HTN history	Crude	n/a		n/a	
	Adjusted	n/a		n/a	
Physical exercise	Crude	0.95 (0.85-1.06)		0.92 (0.79-1.06)	
	Adjusted	0.95 (0.85-1.06)		0.9 (0.77-1.04)	

Crude model, adjusted for all selected variables in the relevant subpopulation and GPS, except for the tested independent variable itself.

n/a, not applicable, as the variable is not included in the corresponding model.

Evaluation on discrimination by each model variables in comparison to a crude model with age at baseline in each subpopulation is also investigated, with GPSs also considered in the evaluation. Across all subpopulations, FPG showed greatest increase in AUC, with ΔAUC of 0.0942, 0.0900, and 0.0943, all *P* for contrast <.0001) in Subpopulations A, B, and C, respectively, followed by HbA1c (ΔAUC = 0.0699, 0.0683, 0.0741, all *P-contrast* <.0001) and TG (ΔAUC = 0.0644, 0.0623, 0.0639, all *P-contrast* <.0001). Change of AUC for GPSs were relatively trivial and showed inconsistent significance across the three subpopulations (GPS-4 in Subpopulation A, ΔAUC =0.0015, *P-contrast* = 0.5601; GPS-16 in Subpopulation B, ΔAUC = 0.0147, *P-contrast* = 0.0184; GPS-36 in Subpopulation C, ΔAUC = 0.0151, *P-contrast* = 0.0857) (Table 7).

Table 7. C-statistics in association between each model variable and incident T2DM, with reference model as baseline age.

	Subpopulation A (N=6,197)		Subpopulation B (N=5,365)		Subpopulation C (N=3,045)	
	AUC(95% CI)	*P-contrast*	AUC(95% CI)	*P-contrast*	AUC(95% CI)	*P-contrast*
Age	0.5653 (0.5469-0.5837)		0.5675 (0.5476-0.5873)		0.5779 (0.5517-0.6040)	
F/Hx of DM	0.5778 (0.5593-0.5962)	0.0302	0.5799 (0.5601-0.5997)	0.0476	0.5869 (0.5607-0.2077)	0.2077
HTN history	0.5913 (0.5729-0.6098)	0.0001	n/a		n/a	
Physical exercise	0.5798 (0.5612-0.5985)	0.0229	0.5237 (0.5063-0.541)	<.0001	0.5254 (0.5022-0.5485)	<.0001
BMI	0.5939 (0.5755-0.6123)	0.0003	0.5983 (0.5781-0.6185)	0.0003	0.6108 (0.5841-0.6374)	0.0018
HDL-C	0.5836 (0.5654-0.6018)	0.0058	0.5880 (0.5684-0.6077)	0.0042	n/a	
TG	0.6297 (0.6119-0.6475)	<.0001	0.6298 (0.6105-0.649)	<.0001	0.6418 (0.6163-0.6674)	<.0001
FPG	0.6595 (0.6409-0.6781)	<.0001	0.6575 (0.6373-0.6777)	<.0001	0.6722 (0.6455-0.6988)	<.0001
HbA1c	0.6352 (0.618-0.6523)	<.0001	0.6358 (0.6173-0.6543)	<.0001	0.6520 (0.6271-0.6768)	<.0001
GPS*	0.5668 (0.5484-0.5853)	0.5601	0.5822 (0.5624-0.6019)	0.0184	0.5930 (0.5664-0.6197)	0.0857

* GPSs relevant to Subpopulations A, B, and C are GPS-4, GPS-16, and GPS-36, respectively.

Finally, influence of genotype information on the prediction ability for risk of T2DM incidence was evaluated by investigation on improvements in discrimination (Table 8) and reclassification (Tables 9 & 10) by the models upon addition of GPS. The full models with all selected variables were divided into five subset models, with addition of BMI, TG, FPG and HbA1c in stages, upon judgment by estimated effect scale and statistical significance investigated previously (Tables 6 & 7).

For discrimination evaluation, addition of GPS-4 did not show significant change in C-statistics, in both Subpopulations A and B (*P-contrast* = 0.4905 and 0.6101 in the fully adjusted model (Model 5) in Subpopulations A and B, respectively). Addition of GPS-16 in various subset models showed statistically significant discrimination in both Subpopulations B and C, except for a borderline significance with Model 5 in Subpopulation B (*P-contrast* = 0.0697 and 0.0297 in Subpopulations B and C, respectively). For GPS-36, addition of the genotype variation also showed significant change in C-statistics (*P-contrast* = 0.0190). The significant or borderline-significant discrimination statistics in GPS-16 and GPS-36, calculated by ΔAUC, were 0.0028, 0.0050, and 0.0069 for GPS-16 in Subpopulation B and C, and GPS-36 in Subpopulation C, respectively (Table 8).

Table 8. Discrimination evaluation of T2DM risk prediction with GPSs across different subpopulations

	Subpopulation A (N=6,197)		
	Without GPS (Ref)	With GPS-4	
	AUC (95% CI)	AUC (95% CI)	*P-contrast*
Model 1	0.6219 (0.6036-0.6401)	0.6253 (0.6072-0.6435)	0.1586
Model 2	0.6329 (0.6147-0.6512)	0.6371 (0.6189-0.6553)	0.0859
Model 3	0.6568 (0.6392-0.6745)	0.6610 (0.6435-0.6785)	0.0377
Model 4	0.7066 (0.6889-0.7243)	0.7083 (0.6907-0.7259)	0.2174
Model 5	0.7264 (0.7091-0.7436)	0.7271 (0.7099-0.7443)	0.4905

Table 8. Discrimination evaluation of T2DM risk prediction with GPSs across different subpopulations (continued)

	Subpopulation B (N=5,365)					
	Without GPS (Ref)	With GPS-4		Without GPS (Ref)	With GPS-16	
	AUC (95% CI)	AUC (95% CI)	*P-contrast*	AUC (95% CI)	AUC (95% CI)	*P-contrast*
Model 1	0.6053 (0.5855-0.625)	0.6082 (0.5886-0.6279)	0.3379	0.6067 (0.5868-0.6265)	0.6152 (0.5953-0.635)	0.0006
Model 2	0.6232 (0.6033-0.6432)	0.6276 (0.6077-0.6476)	0.1111	0.6232 (0.6033-0.6432)	0.6334 (0.6135-0.6533)	0.0102
Model 3	0.6513 (0.6322-0.6704)	0.6549 (0.6359-0.6739)	0.1008	0.6513 (0.6322-0.6704)	0.6590 (0.6401-0.678)	0.0113
Model 4	0.7035 (0.6843-0.7226)	0.7052 (0.6861-0.7242)	0.2445	0.7035 (0.6843-0.7226)	0.7076 (0.6887-0.7266)	0.0453
Model 5	0.7236 (0.7049-0.7423)	0.7242 (0.7056-0.7428)	0.6101	0.7236 (0.7049-0.7423)	0.7264 (0.7079-0.7449)	0.0697

Table 8. Discrimination evaluation of T2DM risk prediction with GPSs across different subpopulations (continued)

	Subpopulation C (N=3,045)					
	Without GPS (Ref)	With GPS-16		Without GPS (Ref)	With GPS-36	
	AUC (95% CI)	AUC (95% CI)	*P-contrast*	AUC (95% CI)	AUC (95% CI)	*P-contrast*
Model 1	0.6063 (0.5802-0.6325)	0.6178 (0.5918-0.6438)	0.0039	0.5984 (0.5719-0.625)	0.6154 (0.5891-0.6418)	0.0003
Model 2	0.6275 (0.6012-0.6538)	0.6392 (0.613-0.6653)	0.0423	0.6251 (0.5986-0.6516)	0.6433 (0.6172-0.6693)	0.0093
Model 3	0.6628 (0.6373-0.6883)	0.6727 (0.6477-0.6977)	0.0218	0.6622 (0.6367-0.6878)	0.6747 (0.6496-0.6998)	0.0158
Model 4	0.7173 (0.6918-0.7428)	0.7239 (0.6988-0.7490)	0.0283	0.7173 (0.6918-0.7428)	0.7258 (0.7007-0.7509)	0.0174
Model 5	0.7394 (0.7148-0.7641)	0.7444 (0.7200-0.7688)	0.0297	0.7397 (0.715-0.7644)	0.7466 (0.7222-0.7709)	0.0190

Modeling in Subpopulation A; model 1, adjusted for age, family history of DM, regular physical exercise, HDL-cholesterol, and HTN; model 2, adjusted for model 1 variables plus BMI; model 3, adjusted for model 2 variables plus TG; model 4, adjusted for model 3 variables plus FPG; model 5, adjusted for model 4 variables plus HbA1c.

Modeling in Subpopulation B; similar to modeling in Subpopulation A, except for HTN excluded from the initial model stage.

Modeling in Subpopulation C; similar to modeling in Subpopulation B, except for HDL-C excluded from the initial model stage in subjects with addition of GPS-36 information.

Evaluation for improvement in reclassification upon addition of GPSs with modeling in stages is shown in Table 9.

Reclassification improvements upon addition of GPS-4 showed significant NRI estimates, but the statistical significance disappeared when HbA1c was finally added in the full model, Model 5 (P = 0.1184 and 0.2613 in Subpopulations A and B). In case of GPS-16 and GPS-36, positive NRI estimates remained significant across all subset models, even at adjustment for HbA1c. Estimated NRIs in Model 5 upon addition of GPS-16 in both Subpopulations B and C, as well as GPS-36 in Subpopulation C, were 3.4 % (P = 0.0059), 4.9 % (P = 0.0073), and 6.2 % (P = 0.0019), respectively.

Table 9. Reclassification evaluation of T2DM risk prediction with GPSs across different subpopulations

	Subpopulation A (N=6,197)		Subpopulation B (N=5,365)				Subpopulation C (N=3,045)			
	With GPS-4		With GPS-4		With GPS-16		With GPS-16		With GPS-36	
	NRI (SE)	*P*	NRI (SE)	P	NRI (SE)	*P*	NRI (SE)	P	NRI (SE)	*P*
Model 1	4.6 (1.4)	0.0014	5.6 (1.7)	0.0009	10.3 (2.1)	<.00001	13 (2.8)	<.00001	13 (3.1)	<.0001
Model 2	7.6 (1.4)	<.00001	4.8 (1.6)	0.0022	8.7 (2)	<.0001	10.1 (2.8)	0.0002	14.4 (3.1)	<.00001
Model 3	3.2 (1.4)	0.0194	5 (1.5)	0.0006	7.9 (1.9)	<.0001	8 (2.5)	0.0013	12.1 (2.8)	<.0001
Model 4	4.7 (1.2)	<.0001	1.8 (1.2)	0.1211	4.7 (1.5)	0.0021	7.5 (2.2)	0.0007	7.9 (2.4)	0.0009
Model 5	1.7 (1.1)	0.1184	1.2 (1.1)	0.2613	3.4 (1.2)	0.0059	4.9 (1.8)	0.0073	6.2 (2)	0.0019

Risk classification in NRI analysis: 10%, 15%, 20%.

Modeling in Subpopulation A; model 1, adjusted for age, family history of DM, regular physical exercise, HDL-cholesterol, and HTN; model 2, adjusted for model 1 variables plus BMI; model 3, adjusted for model 2 variables plus TG; model 4, adjusted for model 3 variables plus FPG; model 5, adjusted for model 4 variables plus HbA1c.

Modeling in Subpopulation B; similar to modeling in Subpopulation A, except for HTN excluded from the initial model stage.

Modeling in Subpopulation C; similar to modeling in Subpopulation B, except for HDL-C excluded from the initial model stage in subjects with addition of GPS-36 information.

With the significant GPSs and subpopulations, additional manual calculation for reclassification was carried out, with investigated groups in order of previously estimated NRI (Tab le 9), and results are shown in Table 10 (a-c). Upon addition of GPS-36, net of correctly reclassified risk in incident T2DM cases and non-cases were 0.4 % and 2.2 %, summing up to a total of 2.6 %, which correspond to 95% CI (2.3-10.1 %) of the analyzed NRI. For GPS-16, net of correctly reclassified risk in cases and non-cases were 0.0 % and 1.9 % in Subpopulation C, and 0.2 % and 0.5 % in Subpopulation B. The sum of net correctly reclassified equaled to 1.9 % and 0.8 % for Subpopulations C and B. The calculated estimate corresponds to analyzed 95% CI in case of Subpopulation C (1.4-8.4 %), but results are inconsistent with the analyzed 95% CI in Subpopulation B (1.0-5.8 %).

Table 10a. Reclassification of predicted risk of incident T2DM between cases and non-cases, with GPS-36 in Subpopulation C (N=3,045)

Incident DM at 8yrs f/u, or predicted risk (without GPS-36)	Predicted risk (with GPS-36)				Reclassified		Net correctly reclassified (%)
	<10%	10-15%	15-20%	≥20%	Increased risk	Decreased risk	
DM cases (n=505)							
<10%	64	13			42	40	0.4
10-15%	10	40	12				
15-20%		10	40	17			
≥20%			20	279			
Non-DM (n=2,453)							
<10%	951	87			261	316	2.2
10-15%	121	270	86				
15-20%		98	134	88			
≥20%			97	521			
Net reclassification index							2.6

Table 10b. Reclassification of predicted risk of incident T2DM between cases and non-cases, with GPS-16 in Subpopulation C (N=3,045)

Incident DM at 8yrs f/u, or predicted risk (without GPS-16)	Predicted risk (with GPS-16)				Reclassified		Net correctly reclassified (%)
	<10%	10-15%	15-20%	≥20%	Increased risk	Decreased risk	
DM cases (n=505)							
<10%	66	13			34	34	0
10-15%	10	42	10				
15-20%		11	45	11			
≥20%			13	284			
Non-DM (n=2,453)							
<10%	964	74			210	257	1.9
10-15%	92	321	63				
15-20%		75	174	73			
≥20%			90	527			
Net reclassification index							1.9

Table 10c. Reclassification of predicted risk of incident T2DM between cases and non-cases, with GPS-16 in Subpopulation B (N=5,365)

Incident DM at 8yrs f/u, or predicted risk (without GPS-16)	Predicted risk (with GPS-16)				Reclassified		Net correctly reclassified (%)
	<10%	10-15%	15-20%	≥20%	Increased risk	Decreased risk	
DM cases (n=905)							
<10%	110	16			52	50	0.2
10-15%	9	102	22				
15-20%		15	89	14			
≥20%			26	502			
Non-DM (n=4,328)							
<10%	1519	100			335	358	0.5
10-15%	133	682	119				
15-20%		109	423	116			
≥20%			116	1011			
Net reclassification index							0.8

4. Evaluation of gene-environment interaction on risk of T2DM incidence

For GPS-16 and GPS-36 in Subpopulations B and C that showed significant discrimination and reclassification statistics, all 'environmental factors (i.e. age, family history of DM, regular physical exercise BMI, HDL-C, FPG, TG, HbA1c)' in the relevant prediction model were tested for interaction analysis.

Before statistical analysis with application of interaction terms, HRs and 95% CIs of environmental factors, stratified into a binary variable of low or high risk, on risk of T2DM incidence across GPS quartiles were investigated to test for trend. In general, *P* for trend was significant across all GPSs and subpopulations in case of age, family history of DM, regular physical exercise, BMI, and HDL-C while increase in HRs were observed across GPS quartiles, in both low and high risk groups. In case of TG and FPG, HRs over GPS quartiles were statistically significant in low risk groups (i.e. TG <150 mg/dL, FPG <100 mg/dL), and *P-trends* were also only significant in the low risk groups (TG; *P-trend* = 0.0006, <.0001, and <.0001 in low risk groups with GPS-36, and GPS in Subpopulations C and B; FPG, *P-trend* = 0.0003, <.0001, and <.0001 in high risk groups with GPS-36, and GPS in Subpopulations C and B).). For HbA1c, significant HRs and *P-trends* were observed in the high risk group (i.e. HbA1c ≥5.5 %) (*P-trend* = 0.0025,

0.0005, and <.0001 in high risk groups with GPS-36, and GPS in Subpopulations C and B) (Table 11).

Table 11. Effect of selected variables on T2DM incidence in low or high risk groups, across quartiles of GPS-36 or GPS-16

		With GPS-36 in Subpopuation C (N=3,045)		With GPS-16 in Subpopulation C (N=3,045)		With GPS-16 in Subpopulation B (N=5,365)	
		Low risk	High risk	Low risk	High risk	Low risk	High risk
Age	Q1	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	0.68 (0.40-1.15)	1.18 (0.80-1.73)	1.56 (0.94-2.58)	1.33 (0.91-1.95)	1.22 (0.86-1.73)	1.49 (1.13-1.98)
	Q3	1.07 (0.70-1.63)	1.46 (1.03-2.06)	1.58 (0.96-2.58)	1.36 (0.94-1.97)	1.36 (0.97-1.90)	1.41 (1.07-1.87)
	Q4	1.32 (0.88-1.99)	1.43 (1.00-2.03)	2.05 (1.25-3.36)	1.62 (1.11-2.36)	1.58 (1.12-2.22)	1.73 (1.30-2.30)
	P-trend	*0.0522*	*0.0075*	*0.0014*	*0.0148*	*0.0033*	*0.0002*
F/Hx of DM	Q1	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	0.90 (0.64-1.24)	1.41 (0.59-3.37)	1.58 (1.15-2.18)	0.54 (0.21-1.41)	1.49 (1.18-1.88)	0.79 (0.41-1.52)
	Q3	1.26 (0.95-1.67)	1.46 (0.65-3.29)	1.46 (1.06-2.00)	1.31 (0.58-2.96)	1.42 (1.12-1.79)	1.17 (0.64-2.14)
	Q4	1.24 (0.93-1.65)	2.50 (1.16-5.40)	1.76 (1.27-2.43)	1.63 (0.73-3.64)	1.68 (1.33-2.12)	1.52 (0.84-2.75)
	P-trend	*0.0089*	*0.0494*	*0.0009*	*0.0268*	*<.0001*	*0.0288*
Regular P/E	Q1	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	1.73 (1.01-2.97)	0.71 (0.48-1.04)	1.75 (1.06-2.87)	1.23 (0.84-1.81)	1.63 (1.14-2.33)	1.24 (0.94-1.64)
	Q3	2.22 (1.35-3.63)	1.02 (0.74-1.41)	2.03 (1.25-3.31)	1.21 (0.83-1.76)	1.49 (1.04-2.13)	1.35 (1.03-1.77)
	Q4	2.03 (1.25-3.32)	1.18 (0.86-1.63)	1.75 (1.06-2.90)	1.81 (1.25-2.63)	1.66 (1.16-2.38)	1.68 (1.28-2.21)
	P-trend	*0.005*	*0.0422*	*0.022*	*0.0007*	*0.0081*	*<.0001*

Table 11. Effect of selected variables on T2DM incidence in low or high risk groups, across quartiles of GPS-36 or GPS-16 (continued)

		With GPS-36 in Subpopuation C (N=3,045)		With GPS-16 in Subpopulation C (N=3,045)		With GPS-16 in Subpopulation B (N=5,365)	
		Low risk	High risk	Low risk	High risk	Low risk	High risk
BMI	Q1	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	0.83 (0.53-1.30)	1.11 (0.73-1.67)	1.15 (0.76-1.75)	1.73 (1.13-2.64)	1.20 (0.88-1.65)	1.54 (1.14-2.09)
	Q3	1.25 (0.86-1.81)	1.27 (0.87-1.85)	1.43 (0.96-2.14)	1.45 (0.95-2.20)	1.33 (0.98-1.80)	1.43 (1.06-1.93)
	Q4	1.29 (0.89-1.86)	1.44 (0.99-2.09)	1.60 (1.07-2.40)	1.87 (1.22-2.89)	1.60 (1.19-2.16)	1.63 (1.20-2.23)
	P-trend	*0.0313*	*0.025*	*0.003*	*0.0144*	*0.0002*	*0.0048*
HDL-C	Q1	n/a	n/a	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	n/a	n/a	1.15 (0.73-1.83)	1.65 (1.11-2.47)	1.39 (1.00-1.93)	1.38 (1.03-1.86)
	Q3	n/a	n/a	1.35 (0.87-2.10)	1.55 (1.04-2.30)	1.32 (0.95-1.84)	1.44 (1.08-1.92)
	Q4	n/a	n/a	1.59 (1.02-2.49)	1.94 (1.30-2.91)	1.50 (1.08-2.10)	1.80 (1.35-2.40)
	P-trend			*0.0061*	*0.0036*	*0.0119*	*<.0001*
TG	Q1	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	1.05 (0.65-1.68)	0.88 (0.59-1.32)	1.72 (1.05-2.84)	1.29 (0.88-1.89)	1.38 (0.99-1.92)	1.41 (1.05-1.89)
	Q3	1.24 (0.81-1.89)	1.31 (0.93-1.85)	2.14 (1.33-3.45)	1.10 (0.75-1.61)	1.54 (1.11-2.14)	1.30 (0.97-1.73)
	Q4	1.71 (1.14-2.56)	1.13 (0.79-1.62)	2.40 (1.49-3.86)	1.48 (1.00-2.19)	1.86 (1.34-2.57)	1.55 (1.15-2.08)
	P-trend	*0.0006*	*0.1754*	*<.0001*	*0.1755*	*<.0001*	*0.0159*

Table 11. Effect of selected variables on T2DM incidence in low or high risk groups, across quartiles of GPS-36 or GPS-16 (continued)

		With GPS-36 in Subpopuation C (N=3,045)		With GPS-16 in Subpopulation C (N=3,045)		With GPS-16 in Subpopulation B (N=5,365)	
		Low risk	High risk	Low risk	High risk	Low risk	High risk
FPG	Q1	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	1.01 (0.73-1.40)	0.81 (0.32-2.02)	1.59 (1.15-2.20)	0.63 (0.28-1.44)	1.35 (1.07-1.70)	1.52 (0.76-3.02)
	Q3	1.29 (0.97-1.72)	1.45 (0.70-3.03)	1.65 (1.19-2.27)	0.86 (0.40-1.85)	1.45 (1.16-1.82)	1.25 (0.63-2.49)
	Q4	1.43 (1.07-1.90)	1.40 (0.68-2.86)	2.03 (1.47-2.81)	0.85 (0.38-1.86)	1.72 (1.37-2.17)	1.45 (0.73-2.90)
	P-trend	*0.0003*	*0.4433*	*<.0001*	*0.7073*	*<.0001*	*0.5396*
HbA1c	Q1	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).	1 (Ref).
	Q2	0.64 (0.33-1.25)	1.08 (0.76-1.53)	1.02 (0.57-1.82)	1.57 (1.10-2.25)	0.84 (0.55-1.28)	1.65 (1.27-2.14)
	Q3	0.92 (0.53-1.58)	1.45 (1.07-1.97)	0.87 (0.48-1.59)	1.65 (1.17-2.33)	0.83 (0.54-1.27)	1.64 (1.27-2.12)
	Q4	1.29 (0.75-2.19)	1.43 (1.05-1.95)	1.58 (0.90-2.77)	1.86 (1.31-2.65)	1.28 (0.84-1.93)	1.86 (1.44-2.41)
	P-trend	*0.2321*	*0.0025*	*0.0935*	*0.0005*	*0.1375*	*<.0001*

Estimates in HRs and 95% CIs, and adjusted for other tested variables included in the full model.

Low vs. high risk groups by variables: age, <50yrs vs. ≥50 yrs; regular physical exercise, yes vs. no; F/Hx of DM, no vs. yes; BMI, <25 kg/m3 vs. ≥25 kg/m3; FPG, <100 mg/dL vs. ≥100 mg/dL; TG, <150 mg/dL vs. ≥150 mg/dL; HbA1c, <5.5% vs. ≥5.5%.

Quartile groups of risk allele scores: 0~37/38~39/40~42/ ≥43 and 0~15/16-17/18~19/≥20 for GPS-36 and GPS-16, respectively.

Investigation on gene*environment (*g***e*) interaction effect began with treating genetic and environmental factors to a 2×2 stratification for each examined GPS in the corresponding subpopulation. Hazard ratios and their 95% CIs were calculated in two ways; firstly by the 2×2 stratification with group of low environmental risk and low GPS risk group as the reference, secondly by setting low genetic/environmental risk as within-strata reference in deriving estimates for both risk groups of environmental/genetic factors, vice versa (Tables 12a to 19a). Measures of interaction, in both multiplicative and additive scales, are arranged in Tables 12b to 19b. Multiplicative interaction was analyzed using the hazard function. For additive interaction, RERI, AP, and S with their 95% CIs were derived. Several statistical packages were tried in testing for additive interaction, and after reviewing robustness of the outcomes, results analyzed by package "*add_int*" derived from logistic regression in Stata/SE 13.0 was selected. By this method, parameter estimates were attained by bootstrapping for 500 loops.

Table 12a shows age as a dichotomous variable divided by below or above 50 years old, while GPSs are also dichotomized as below or above median (17 and 40 risk allele scores for GPS-16 and GPS-36, respectively). In 2×2 stratification with group age <50 years and GPSs <median as a reference, all three high/low, low/high, or high/high genetic/environmental (*g/e*) risk groups (i.e. GPS ≥median/age <50 years, GPS<median/age ≥50 years, or GPS ≥median/age ≥50 years) showed significant HRs >1, with highest HRs at high/high *g/e* risk groups (2.26 (1.72-2.97), 2.09 (1.25-2.01), and 1.84 (1.50-2.26) for GPS-36 and GPS-16 in two Subpopulations C and B). When

stratified by GPSs, HRs for high risk of age ≥50 years were also significant in both groups of high/low GPSs, but greater HR in high *g* risk in comparison to low *g* risk was observed only in Subpopulation C tested for GPS-36 (1.60 (1.20-2.12) in GPS-36 <40 vs. 1.69 (1.32-2.15) in GPS-36 ≥40). When stratified by age groups, HRs for high risk of GPSs ≥median were not as consistently significant as in the vice versa. By statistical significance at *P* <.05, HRs in subjects tested with GPS-36 were only significant, and greater HR in high *e* risk was also observed only in the same subjects (1.34 (1.01-1.79) in age <50 years vs. 1.42 (1.13-1.77) in age ≥50 years).

Measures of interaction in multiplicative interaction were not significant in both GPS-36 and GPS-16 in the two subpopulations. In additive scale, subjects tested with GPS-36 information showed significantly positive RERI (0.63 (95% LCI, UCI: 0.01, 1.26) and AP (24 % (2 %, 46 %)). (Table 12b).

Table 12a. Effect of interaction between GPSs and age on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of age
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-36 (Subpopulation C, N=3,045)					
Age <50 yrs	85/734	1 (Ref)	113/774	1.34 (1.01-1.79), *P*=0.045	1.34 (1.01-1.79), *P*=0.045
Age ≥50 yrs	138/795	1.60 (1.20-2.12), *P*=0.001	187/742	2.26 (1.72-2.97), *P*<.0001	1.42 (1.13-1.77), *P*=0.002
HR(95% CI) for age ≥50 yrs within strata of GPS-36		1.60 (1.20-2.12), *P*=0.001		1.69 (1.32-2.15), *P*<.0001	
GPS-16 (Subpopulation C, N=3,045)					
Age <50 yrs	76/689	1 (Ref)	122/819	1.32 (0.98-1.77), *P*=0.063	1.32 (0.98-1.77), *P*=0.063
Age ≥50 yrs	132/730	1.72 (1.28-2.31), *P*<.0001	193/807	2.09 (1.58-2.77), *P*<.0001	1.22 (0.97-1.52), *P*=0.091
HR(95% CI) for age ≥50 yrs within strata of GPS-16		1.72 (1.28-2.31), *P*<.0001		1.58 (1.25-2.01), *P*<.0001	
GPS-16 (Subpopulation B, N=5,365)					
Age <50 yrs	143/1223	1 (Ref)	223/1409	1.30 (1.05-1.61), *P*=0.016	1.30 (1.05-1.61), *P*=0.016
Age ≥50 yrs	235/1295	1.57 (1.26-1.94), *P*<.0001	328/1438	1.84 (1.5-2.26), *P*<.0001	1.18 (0.99-1.39), *P*=0.062
HR(95% CI) for age ≥50 yrs within strata of GPS-16		1.57 (1.26-1.94), *P*<.0001		1.42 (1.19-1.7), *P*<.0001	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 12b. Measure of interaction between GPSs and age.

	GPS-36 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	1.06 (0.73-1.52)	0.92 (0.64-1.33)	0.91 (0.69-1.19)
Additive scale: RERI (95% LCI, UCI)	0.63 (0.01, 1.26)	0.27 (-0.37, 0.9)	0.08 (-0.39, 0.56)
Additive scale: AP (95% LCI, UCI)	0.24 (0.02, 0.46)	0.1 (-0.14, 0.34)	0.04 (-0.17, 0.24)
Additive scale: S (95% LCI, UCI)	1.62 (-0.31, 3.54)	1.19 (0.43, 1.96)	1.07 (0.61, 1.53)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-36 and GPS-16 are 40 and 17 risk allele scores, respectively.

Table 13a shows family history of DM (F/Hx of DM) as the dichotomous variable divided by yes or no, and GPSs are also dichotomized by the median. In 2×2 stratification with group with no F/Hx of DM and GPSs <median as a reference, all three subject groups showed non-significant HRs in low/high *g/e* groups, i.e. HRs for stratified group with GPS below median and with F/Hx of DM were not statistically significant, while other groups with high/low and high/high *g/e* show increased HRs with P <.0001 at high/high *g/e* groups. Highest HRs were observed at high/high *g/e* risk groups (1.91 (1.35-2.72), 1.83 (1.33-2.52), and 1.73 (1.36-2.21) for GPS-36 and GPS-16 in two Subpopulations C and B). Hazard ratios for high risk of F/Hx of DM stratified by GPS were significant at GPS ≥medium risk allele scores, and estimates for high risk of GPS stratified by F/Hx of DM were mostly significant or borderline-significant. In both, high *g* or *e* risk group in stratification of *e* or *g* showed higher HRs than in low *g* or *e* risk groups (e.g. 0.87 (0.54-1.42) in GPS <median vs. 1.57 (1.15-2.14) in GPS ≥median for subjects tested for GPS-16 in Subpopulation C, at high risk in F/Hx of DM; 1.17 (0.96-1.41) in F/Hx of DM (-) vs. 2.10 (1.22-3.59) in F/Hx of DM at GPS-16 ≥17).

Significant measures of interaction were observed in subjects tested for GPS-16 in Subpopulations B and C. In Subpopulation B, AP was positively significant (30 % (4 %, 57%)). In Subpopulation C, both multiplicative and additive interaction scales showed positive interactions; HR by *g*e* interaction was 1.80 (1.01-3.19), whereas RERI and AP were found to be 1.01 (0.12, 1.89) and 45 % (14 %, 75 %) (Table 13b).

Table 13a. Effect of interaction between GPSs and family history of DM on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of F/Hx of DM
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-36 (Subpopulation C, N=3,045)					
F/Hx of DM, no	194/1373	1 (Ref)	260/1354	1.36 (1.13-1.65), *P*=0.002	1.36 (1.13-1.65), *P*=0.002
F/Hx of DM, yes	29/156	1.20 (0.81-1.79), *P*=0.360	40/162	1.91 (1.35-2.72), *P*<.0001	1.59 (0.98-2.57), *P*=0.060
HR(95% CI) for F/Hx DM(+) within strata of GPS-36		1.20 (0.81-1.79), *P*=0.360		1.40 (1.00-1.97), *P*=0.051	
GPS-16 (Subpopulation C, N=3,045)					
F/Hx of DM, no	189/1284	1 (Ref)	265/1443	1.17 (0.96-1.41), *P*=0.114	1.17 (0.96-1.41), *P*=0.114
F/Hx of DM, yes	19/135	0.87 (0.54-1.42), *P*=0.587	50/183	1.83 (1.33-2.52), *P*<.0001	2.1 (1.22-3.59), *P*=0.007
HR(95% CI) for F/Hx DM(+) within strata of GPS-16		0.87 (0.54-1.42), *P*=0.587		1.57 (1.15-2.14), *P*=0.004	
GPS-16 (Subpopulation B, N=5,365)					
F/Hx of DM, no	338/2283	1 (Ref)	466/2529	1.18 (1.02-1.36), *P*=0.022	1.18 (1.02-1.36), *P*=0.022
F/Hx of DM, yes	40/235	1.11 (0.8-1.56), *P*=0.527	85/318	1.73 (1.36-2.21), *P*<.0001	1.55 (1.06-2.27), *P*=0.023
HR(95% CI) for F/Hx DM(+) within strata of GPS-16		1.11 (0.8-1.56), *P*=0.527		1.46 (1.16-1.86), *P*=0.002	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 13b. Measure of interaction between GPSs and family history of DM.

	GPS-36 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	1.17 (0.69-1.96)	1.80 (1.01-3.19)	1.31 (0.88-1.97)
Additive scale: RERI (95% LCI, UCI)	0.11 (-0.89, 1.11)	1.01 (0.12, 1.89)	0.64 (-0.03, 1.31)
Additive scale: AP (95% LCI, UCI)	0.05 (-0.46, 0.57)	0.45 (0.14, 0.75)	0.30 (0.04, 0.57)
Additive scale: S (95% LCI, UCI)	1.12 (-14.59, 16.82)	5.11 (-157.76, 167.98)	2.33 (-45.0, 49.67)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-36 and GPS-16 are 40 and 17 risk allele scores, respectively.

Table 14a shows regular physical exercise (regular P/E) as the dichotomous variable divided by yes or no, and GPSs are also dichotomized by the median. In 2×2 stratification with group with no regular P/E and GPSs <median as a reference, all HRs in high/low, low/high and high/high *g/e* risk groups are significant in subjects tested for GPS-36, while only high/high *g/e* risk groups showed statistical significance in subjects tested for GPS-16. Highest HRs were observed at high/high *g/e* risk groups (1.84 (1.40-2.42), 1.60 (1.22-2.08), and 1.49 (1.22-1.81) for GPS-36 and GPS-16 in two Subpopulations C and B). When stratified by GPSs, HRs for high risk of P/E showed increased estimate from GPS <median to GPS ≥median in subjects tested for GPS-16 in Subpopulation B (1.16 (0.94-1.43) in GPS <17 vs. 1.31 (1.10-1.57) in GPS ≥17). The same subjects, when stratified by P/E, showed increased HR from non-regular to regular P/E at GPSs ≥median (1.13 (0.91-1.40) in no regular P/E vs. 1.29 (1.08-1.52) in regular P/E).

For measure of interaction, neither multiplicative nor additive scales showed statistically significant estimates (Table 14b).

Table 14a. Effect of interaction between GPSs and regular physical exercise on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of regular P/E
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-36 (Subpopulation C, N=3,045)					
Regular P/E, yes	77/633	1 (Ref)	124/661	1.51 (1.13-2.01), *P*=0.005	1.51 (1.13-2.01), *P*=0.005
Regular P/E, no	142/877	1.39 (1.04-1.85), *P*=0.024	175/838	1.84 (1.40-2.42), *P*<.0001	1.32 (1.06-1.66), *P*=0.014
HR(95% CI) for P/E(+) within strata of GPS-36		1.39 (1.04-1.85), *P*=0.024		1.22 (0.96-1.54), *P*=0.099	
GPS-16 (Subpopulation C, N=3,045)					
Regular P/E, yes	81/604	1 (Ref)	120/690	1.27 (0.96-1.69), *P*=0.097	1.27 (0.96-1.69), *P*=0.097
Regular P/E, no	124/796	1.28 (0.96-1.71), *P*=0.087	193/919	1.60 (1.22-2.08), *P*=0.001	1.24 (0.99-1.57), *P*=0.064
HR(95% CI) for P/E(+) within strata of GPS-16		1.28 (0.96-1.71), *P*=0.087		1.25 (0.99-1.59), *P*=0.059	
GPS-16 (Subpopulation B, N=5,365)					
Regular P/E, yes	148/1043	1 (Ref)	208/1214	1.13 (0.91-1.40), *P*=0.255	1.13 (0.91-1.40), *P*=0.255
Regular P/E, no	226/1450	1.16 (0.94-1.43), *P*=0.180	341/1606	1.49 (1.22-1.81), *P*<.0001	1.29 (1.08-1.52), *P*=0.004
HR(95% CI) for P/E(+) within strata of GPS-16		1.16 (0.94-1.43), *P*=0.180		1.31 (1.1-1.57), *P*=0.003	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 14b. Measure of interaction between GPSs and regular physical exercise.

	GPS-36 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	0.88 (0.61-1.27)	0.98 (0.68-1.41)	1.14 (0.86-1.49)
Additive scale: RERI (95% LCI, UCI)	-0.17 (-0.8, 0.45)	0.18 (-0.36, 0.71)	0.27 (-0.08, 0.62)
Additive scale: AP (95% LCI, UCI)	-0.09 (-0.41, 0.23)	0.10 (-0.20, 0.40)	0.17 (-0.04, 0.38)
Additive scale: S (95% LCI, UCI)	0.84 (-0.10, 1.78)	1.31 (-30.27, 32.9)	1.74 (-12.16, 15.63)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-36 and GPS-16 are 40 and 17 risk allele scores, respectively.

Table 15a shows aBMI as the dichotomous variable divided by below or above 25 kg/m^3, and GPSs also dichotomized by the median. In 2×2 stratification with group with low risk of BMI and GPSs <median as a reference, all three subject groups showed non-significant HRs in low/high *g/e* groups, i.e. HRs for stratified group with GPS below median and with BMI ≥25 kg/m^3 were not statistically significant, while other groups with high/low and high/high *g/e* show increased HRs with P <.05 at high/high *g/e* groups. Highest HRs were observed at high/high *g/e* risk groups (1.50 (1.16-1.95), 1.36 (1.05-1.77), and 1.35 (1.11-1.65) for GPS-36 and GPS-16 in two Subpopulations C and B). In groups stratified by GPS, HRs for BMI at ≥25 kg/m^3 were all non-significant, and size of estimate did not increase from GPS <median to ≥median risk allele scores. When stratified by BMI, HRs at GPS ≥medium only remained significant in groups with BMI < 25 kg/m^3, and increase in HRs from BMI <25 kg/m^3 to ≥25 kg/m^3 was also not observed.

Measures of interaction either by multiplicative or additive indices were non-significant, in all three tested subjects (Table 15b).

Table 15a. Effect of interaction between GPSs and BMI on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of BMI
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-36 (Subpopulation C, N=3,045)					
BMI <25 kg/m3	106/877	1 (Ref)	164/929	1.47 (1.15-1.89), *P*=0.002	1.47 (1.15-1.89), *P*=0.002
BMI ≥25 kg/m3	117/652	1.19 (0.91-1.56), *P*=0.204	136/587	1.50 (1.16-1.95), *P*=0.002	1.26 (0.98-1.62), *P*=0.068
HR for BMI within strata of GPS-36		1.19 (0.91-1.56), *P*=0.204		1.02 (0.81-1.29), *P*=0.864	
GPS-16 (Subpopulation C, N=3,045)					
BMI <25 kg/m3	101/830	1 (Ref)	169/976	1.37 (1.06-1.75), *P*=0.014	
BMI ≥25 kg/m3	107/589	1.22 (0.92-1.61), *P*=0.161	146/650	1.36 (1.05-1.77), *P*=0.022	1.11 (0.86-1.44), *P*=0.404
HR for BMI within strata of GPS-16		1.22 (0.92-1.61), *P*=0.161		0.99 (0.79-1.25), *P*=0.964	
GPS-16 (Subpopulation B, N=5,365)					
BMI <25 kg/m3	179/1453	1 (Ref)	292/1708	1.28 (1.06-1.54), *P*=0.01	1.28 (1.06-1.54), *P*=0.01
BMI ≥25 kg/m3	199/1065	1.18 (0.96-1.45), *P*=0.112	259/1139	1.35 (1.11-1.65), *P*=0.003	1.14 (0.95-1.38), *P*=0.156
HR for BMI within strata of GPS-16		1.18 (0.96-1.45), *P*=0.112		1.06 (0.89-1.26), *P*=0.522	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 15b. Measure of interaction between GPSs and BMI.

	GPS-36 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	0.86 (0.6-1.22)	0.82 (0.57-1.16)	0.89 (0.69-1.17)
Additive scale: RERI (95% LCI, UCI)	-0.03 (-0.68, 0.61)	-0.06 (-0.75, 0.63)	-0.02 (-0.47, 0.42)
Additive scale: AP (95% LCI, UCI)	-0.02 (-0.3, 0.27)	-0.03 (-0.35, 0.29)	-0.01 (-0.23, 0.2)
Additive scale: S (95% LCI, UCI)	0.97 (0.37, 1.58)	0.95 (0.28, 1.62)	0.98 (0.53, 1.43)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-36 and GPS-16 are 40 and 17 risk allele scores, respectively.

Test for *g*e* interaction with HDL-C as an environment factor was carried out for GPS-16 only, as HDL-C was not a selected variable included in the prediction model with GPS-36.

Table 16a shows HDL-C as the dichotomous variable divided by below or above 40 mg/dL (in males) or 50 mg/dL (in females), and GPSs also dichotomized by the median. In 2×2 stratification with group with low risk of HDL-C and GPSs <median as a reference, only significance was observed in high/high *g/e* risk groups in subjects tested for GPS-16 in Subpopulation B (1.28 (1.06-1.55)). The same subjects, upon stratification by high HDL-C risk, showed statistical significance at GPS-16 ≥17 with HR of 1.30 (1.09-1.56), and this was comparable to HR of GPS-16≥17 at low HDL-C risk (1.12 (0.92-1.37)).

Again, some significant measures of interaction were observed in the same subjects tested for GPS-16 in Subpopulation B. Significant positive values for additive interaction was found, with RERI (0.38 (0.05, 0.71)) and AP (24 % (3 %, 44 %)) (Table 16b).

Table 16a. Effect of interaction between GPSs and HDL-cholesterol on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of HDL-C
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-16 (Subpopulation C, N=3,045)					
HDL-C <40/50 mg/dL (M/F)	84/610	1 (Ref)	135/767	1.29 (0.98-1.69), *P*=0.074	1.29 (0.98-1.69), *P*=0.074
HDL-C ≥40/50 mg/dL(M/F)	124/809	0.98 (0.74-1.31), *P*=0.915	180/859	1.22 (0.93-1.59), *P*=0.151	1.24 (0.98-1.57), *P*=0.08
HR(95% CI) for HDL-C ≥40/50 mg/dL within strata of GPS-16		0.98 (0.74-1.31), *P*=0.915		0.95 (0.75-1.19), *P*=0.641	
GPS-16 (Subpopulation B, N=5,365)					
HDL-C <40/50 mg/dL (M/F)	169/1160	1 (Ref)	224/1347	1.12 (0.92-1.37), *P*=0.264	1.12 (0.92-1.37), *P*=0.264
HDL-C ≥40/50 mg/dL(M/F)	209/1358	0.98 (0.79-1.21), *P*=0.857	327/1500	1.28 (1.06-1.55), *P*=0.012	1.3 (1.09-1.56), *P*=0.003
HR(95% CI) for HDL-C ≥40/50 mg/dL within strata of GPS-16		0.98 (0.79-1.21), *P*=0.857		1.14 (0.96-1.36), *P*=0.142	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 16b. Measure of interaction between GPSs and HDL-cholesterol.

	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	0.96 (0.67-1.38)	1.16 (0.89-1.52)
Additive scale: RERI (95% LCI, UCI)	0.19 (-0.29, 0.66)	0.38 (0.05, 0.71)
Additive scale: AP (95% LCI, UCI)	0.11 (-0.18, 0.41)	0.24 (0.03, 0.44)
Additive scale: S (95% LCI, UCI)	1.41 (-557.75, 560.58)	2.63 (-30.78, 36.04)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-16 is 17 risk allele scores.

Table 17a shows TG as the dichotomous variable divided by below or above 150 mg/dL, and GPSs dichotomized by the median. In 2×2 stratification with group with TG <150 mg/dL and GPSs <median as a reference, all three high/low, low/high, or high/high genetic/environmental (g/e) risk groups (i.e. GPS ≥median/TG <150 mg/dL, GPS<median/TG ≥150 mg/dL, or GPS ≥median/TG ≥150 mg/dL) showed significant HRs >1, with highest HRs at high/high g/e risk groups (2.54 (1.95-3327), 2.39 (1.23-1.96), and 2.00 (1.63-2.44) for GPS-36 and GPS-16 in two Subpopulations C and B). When stratified by GPSs, HRs for high risk of TG ≥150 mg/dL were also significant in both groups of high/low GPSs, but greater HR in high *g* risk compared to low *g* risk was not observed in all three subjects. When stratified by TG, HRs for high risk of GPSs ≥median were not as consistently significant as in the vice versa, and increase in HRs from low *e* risk to high *e* risk was also not observed in all three subjects.

Measures of interaction either by multiplicative or additive indices were non-significant, in all three tested subjects (Table 17b).

Table 17a. Effect of interaction between GPSs and TG on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of TG
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-36 (Subpopulation C, N=3,045)					
TG <150 mg/dL	91/934	1 (Ref)	144/958	1.52 (1.16-1.99), *P*=0.002	1.52 (1.16-1.99), *P*=0.002
TG ≥150 mg/dL	132/595	1.96 (1.48-2.58), *P*<.0001	156/558	2.54 (1.95-3.32), *P*<.0001	1.30 (1.02-1.65), *P*=0.031
HR for TG ≥150 mg/dL within strata of GPS-36		1.96 (1.48-2.58), *P*<.0001		1.67 (1.33-2.11), *P*<.0001	
GPS-16 (Subpopulation C, N=3,045)					
TG <150 mg/dL	79/861	1 (Ref)	156/1031	1.54 (1.17-2.03), *P*=0.002	1.54 (1.17-2.03), *P*=0.002
TG ≥150 mg/Dl	129/558	2.25 (1.67-3.03), *P*<.0001	159/595	2.39 (1.80-3.18), *P*=<.0001	1.06 (0.84-1.35), *P*=0.619
HR for TG ≥150 mg/dL within strata of GPS-16		2.25 (1.67-3.03), *P*<.0001		1.55 (1.23-1.96), *P*<.0001	
GPS-16 (Subpopulation B, N=5,365)					
TG <150 mg/dL	164/1557	1 (Ref)	262/1745	1.35 (1.11-1.65), *P*=0.003	1.35 (1.11-1.65), *P*=0.003
TG ≥150 mg/dL	214/961	1.78 (1.44-2.20), *P*<.0001	289/1102	2.00 (1.63-2.44), *P*<.0001	1.12 (0.94-1.34), *P*=0.205
HR for TG ≥150 mg/dL within strata of GPS-16		1.78 (1.44-2.20), *P*<.0001		1.48 (1.24-1.76), *P*<.0001	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 17b. Measure of interaction between GPSs and TG.

	GPS-36 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	0.85 (0.6-1.22)	0.69 (0.48-1.0002)	0.83 (0.64-1.08)
Additive scale: RERI (95% LCI, UCI)	0.27 (-0.67, 1.2)	-0.12 (-1.14, 0.89)	0.07 (-0.55, 0.69)
Additive scale: AP (95% LCI, UCI)	0.07 (-0.18, 0.33)	-0.03 (-0.3, 0.24)	0.02 (-0.18, 0.23)
Additive scale: S (95% LCI, UCI)	1.11 (0.66, 1.57)	0.95 (0.59, 1.32)	1.03 (0.69, 1.38)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-36 and GPS-16 are 40 and 17 risk allele scores, respectively.

Table 18a shows the dichotomous variable, FPG divided by below or above 100 mg/dL, and GPSs by the median. In 2×2 stratification with group with FPG <100 mg/dL and GPSs <median as a reference, all three high/low, low/high, or high/high genetic/environmental (g/e) risk groups (i.e. GPS ≥median/FPG <100 mg/dL, GPS<median/FPG ≥100 mg/dL, or GPS ≥median/FPG ≥100 mg/dL) showed significant HRs >1, with highest HRs at high/high *g/e* risk groups (7.06 (5.08-9.82), 5.80 (4.18-8.05), and 4.62 (3.58-5.96) for GPS-36 and GPS-16 in two Subpopulations C and B). Hazard ratios for high risk of FPG stratified by GPS were significant at GPS ≥medium risk allele scores, and estimates for high risk of GPS stratified by FPG showed significant estimates at FPG <100 mg/dL and borderline-significant or non-significant estimates at FPG ≥100 mg/dL. In both, high *g* or *e* risk group in stratification of *e* or *g* showed higher HRs than in low *g* or *e* risk groups (e.g. 4.45 (2.99-6.61) in GPS-36 <40 vs. 4.90 (3.56-6.75) in GPS-36 ≥40 at FPG ≥100 mg/dL; 1.44 (1.19-1.74) in FPG <100 mg/dL vs. 1.59 (0.99-2.54) in FPG ≥100 mg/dL at GPS-36 ≥40).

Measures of interaction either by multiplicative or additive indices were non-significant, in all three tested subjects (Table 18b).

Table 18a. Effect of interaction between GPSs and FPG on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of FPG
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-36 (Subpopulation C, N=3,045)					
FPG <100 mg/dL	194/1476	1 (Ref)	254/1445	1.44 (1.19-1.74), *P*<.0001	1.44 (1.19-1.74), *P*<.0001
FPG ≥100 mg/dL	29/53	4.45 (2.99-6.61), *P*<.0001	46/71	7.06 (5.08-9.82), *P*<.0001	1.59 (0.99-2.54), *P*=0.054
HR for FPG ≥100 mg/dL within strata of GPS-36		4.45 (2.99-6.61), *P*<.0001		4.90 (3.56-6.75), *P*<.0001	
GPS-16 (Subpopulation C, N=3,045)					
FPG <100 mg/Dl	181/1371	1 (Ref)	267/1550	1.33 (1.1-1.61), *P*=0.004	1.33 (1.1-1.61), *P*=0.004
FPG ≥100 mg/dL	27/48	5.21 (3.45-7.86), *P*<.0001	48/76	5.80 (4.18-8.05), *P*<.0001	1.11 (0.69-1.8), *P*=0.655
HR for FPG ≥100 mg/dL within strata of GPS-16		5.21 (3.45-7.86), *P*<.0001		4.36 (3.18-5.98), *P*<.0001	
GPS-16 (Subpopulation B, N=5,365)					
FPG <100 mg/dL	331/2430	1 (Ref)	475/2716	1.29 (1.12-1.49), *P*<.0001	1.29 (1.12-1.49), *P*<.0001
FPG ≥100 mg/dL	47/88	4.32 (3.17-5.89), *P*<.0001	76/131	4.62 (3.58-5.96), *P*<.0001	1.07 (0.74-1.54), *P*=0.722
HR for FPG ≥100 mg/dL within strata of GPS-16		4.32 (3.17-5.89), *P*<.0001		3.57 (2.79-4.57), *P*<.0001	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 18b. Measure of interaction between GPSs and FPG.

	GPS-36 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	1.1 (0.66-1.83)	0.84 (0.5-1.4)	0.83 (0.56-1.22)
Additive scale: RERI (95% LCI, UCI)	4.07 (-4.43, 12.57)	2.66 (-5.98, 11.31)	0.98 (-3.91, 5.86)
Additive scale: AP (95% LCI, UCI)	0.32 (-0.23, 0.87)	0.23 (-0.4, 0.86)	0.11 (-0.4, 0.62)
Additive scale: S (95% LCI, UCI)	1.53 (0.08, 2.98)	1.33 (0.01, 2.65)	1.14 (0.35, 1.93)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-36 and GPS-16 are 40 and 17 risk allele scores, respectively.

Table 19a shows HbA1c as the dichotomous variable divided by below or above 5.5 %, and GPSs also dichotomized by the median. In 2×2 stratification with group with HbA1c <5.5 % and GPSs <median as a reference, all three subject groups showed non-significant HRs in high/low *g/e* groups, i.e. HRs for stratified group with GPS ≥median and with HbA1c <5.5% were not statistically significant, while other groups with low/high and high/high *g/e* show increased HRs with *P* <.0001 at high/high *g/e* groups. Highest HRs were observed at high/high *g/e* risk groups (2.83 (2.05-3.91), 2.45 (1.79-3.36), and 2.31 (1.83-2.52) for GPS-36 and GPS-16 in two Subpopulations C and B). Hazard ratios for high risk of HbA1c stratified by GPS were significant at GPS ≥medium risk allele scores, while estimates for high risk of GPS stratified by HbA1c were also significant at HbA1c ≥5.5 %. Also, in subjects tested for GPS-16 in Subpopulations C and B, high *g* or *e* risk group in stratification of *e* or *g* showed higher HRs than in low *g* or *e* risk groups (GPS <median vs. ≥median at HbA1c≥5.5 %, 1.94 (1.39-2.69) to 2.06 (1.53-2.77) in Subpopulation C and 1.86 (1.46-2.38) to 2.02 (1.61-2.52) in Subpopulation B; HbAQ1c <5.5 % vs. ≥5.5 % at GPS-16 ≥17, 1.19 (0.81-1.75) to 1.27 (1.04-1.55) in Subpopulation C and 1.15 (0.86-1.53) to 1.24 (1.07-1.44) in Subpopulation B.

For measure of interaction, some significant measures of additive interaction were observed in all three subjects, with more significant scales in subjects tested for GPS-16 in Subpopulations B and C. In subjects with GPS-16 in Subpopulation C, AP (22% (1%, 43%)) showed statistical significance, while some RERI (0.71 (0.11, 1.31)) and AP (20 % (4 %, 36 %)) showed

positive significance in subjects tested for GPS-16 in Subpopulation B (Table 19b).

Table 19a. Effect of interaction between GPSs and HbA1c on the risk of T2DM incidence

	GPS <median		GPS ≥median		HR(95% CI) for GPS ≥median within strata of HbA1c
	case/n	HR (95% CI), *P*	case/n	HR (95% CI), *P*	
GPS-36 (Subpopulation C, N=3,045)					
HbA1c <5.5 %	46/658	1 (Ref)	59/561	1.43 (0.97-2.11), *P*=0.073	1.43 (0.97-2.11), *P*=0.073
HbA1c ≥5.5 %	177/871	2.06 (1.47-2.87), *P*<.0001	241/955	2.83 (2.05-3.91), *P*<.0001	1.38 (1.13-1.68), *P*=0.002
HR for HbA1c ≥5.5 % within strata of GPS-36		2.06 (1.47-2.87), *P*<.0001		1.98 (1.48-2.65), *P*<.0001	
GPS-16 (Subpopulation C, N=3,045)					
HbA1c <5.5 %	50/628	1 (Ref)	55/591	1.19 (0.81-1.75), *P*=0.380	1.19 (0.81-1.75), *P*=0.380
HbA1c ≥5.5 %	158/791	1.94 (1.39-2.69), *P*<.0001	260/1035	2.45 (1.79-3.36), *P*<.0001	1.27 (1.04-1.55), *P*=0.022
HR for HbA1c ≥5.5 % within strata of GPS-16		1.94 (1.39-2.69), *P*<.0001		2.06 (1.53-2.77), *P*<.0001	
GPS-16 (Subpopulation B, N=5,365)					
HbA1c <5.5 %	88/1063	1 (Ref)	97/1005	1.15 (0.86-1.53), *P*=0.358	1.15 (0.86-1.53), *P*=0.358
HbA1c ≥5.5 %	290/1455	1.86 (1.46-2.38), *P*<.0001	454/1842	2.31 (1.83-2.93), *P*<.0001	1.24 (1.07-1.44), *P*=0.005
HR for HbA1c ≥5.5 % within strata of GPS-16		1.86 (1.46-2.38), *P*<.0001		2.02 (1.61-2.52), *P*<.0001	

Estimates are adjusted for other tested variables included in the full model.

Median values of GPS-16 and GPS-36 are 17 and 40 risk allele scores, respectively.

Table 19b. Measure of interaction between GPSs and HbA1c.

	GPS-36 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation C, N=3,045)	GPS-16 (Subpopulation B, N=5,365)
Multiplicative scale: HR (95% CI)	0.96 (0.62-1.49)	1.07 (0.69-1.65)	1.08 (0.78-1.5)
Additive scale: RERI (95% LCI, UCI)	0.65 (-0.35, 1.64)	0.85 (-0.04, 1.75)	0.71 (0.11, 1.31)
Additive scale: AP (95% LCI, UCI)	0.14 (-0.07, 0.36)	0.22 (0.01, 0.43)	0.2 (0.04, 0.36)
Additive scale: S (95% LCI, UCI)	1.23 (0.78, 1.67)	1.42 (0.79, 2.04)	1.37 (0.92, 1.83)

Estimates are adjusted for other tested variables included in the full model.

RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index.

Median value for GPS-36 and GPS-16 are 40 and 17 risk allele scores, respectively.

As none of the tested multiplicative interaction showed significance in the current study except for *g*e* interaction between GPS-16 and F/Hx of DM in Subpopulation C (N=3,045), we applied 1.30 as the minimum *a priori* RR and 1.8 as the maximum RR in confirming sample size calculations. Estimate of 1.30 was derived from the maximum OR with significant 95% CIs analyzed for single individual SNPs (Table 2), and 1.80 from the current significant result with the F/Hx of DM.

On assumption that fraction proportions of high/low *g/e* risk groups are similar to that found for F/Hx of DM, number of the subjects needed in investigating multiplicative interaction effect are shown in Table 20. In the case of testing for statistically robust *g*e* interaction in Subpopulation B (N=5,365) and C, the appropriate HRs by the interaction may be above 1.60 and 1.80 for the two subpopulations, respectively. In the same sense, approximately 15,000 study subjects may be needed in verifying *g*e* interaction effect of at least 1.30.

Table 20. Sample size calculation testing interaction effect of binary variables for Cox proportional hazards regression

Gene-environment interaction effect (RR)	Calculated sample size (N)
1.30	15,272-20,445
1.40	9,286-12,431
1.50	6,395-8,561
1.60	4,759-6,371
1.70	3,734-4,999
1.80	3,043-4,074

Range of sample size depend on power (80 % to 90 %).

DISCUSSION

From a community cohort of 8-years follow-up in Korea, influence of genetic predisposition drawn from genotype information on 16 and 36 validated SNPs on risk of T2DM incidence was observed, even at full models adjusted with strong predictors of T2DM risk. Significant improvement in discrimination and reclassification of the prediction models were found in subpopulation of 5,365 and 3,045 subjects with GPS-16 and GPS-36, and greatest improvement was found in subjects investigated for GPS-36. Gene-environment interaction was tested using the same GPSs and model variables in the prediction model, and some patterns of significant interaction effect was observed, most of which were in additive scales.

1. Effect of genetic predisposition on T2DM risk prediction

Risk prediction modeling for T2DM on the same Ansung-Ansan cohort population had been carried out previously, at 4-years follow-up and without considering for genetic predisposition. The authors had focused on the HbA1c variable, which substantially increased NRI (12.8%) upon addition to the prediction model[22]. Another 5-year follow-up cohort study on Japanese population also reported FPG and HbA1c together were effective predictors for T2DM incidence[48]. A case-cohort research from European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study

that utilized metabolic markers including HbA1c as well as genetic markers in predicting T2DM risk, found that addition of genetic information to metabolic markers, age, anthropometry, and lifestyle characteristics, did not significantly improve disease prediction, while FPG and HbA1c considerably contributed to the prediction[12]. The current results, where reclassification indices after HbA1c adjustment showed non-significance in subpopulations tested for GPS-4 but remained significant in subpopulations with GPS-16 and GPS-36 information, may partially support the previous findings. As an indicator of chronic glycemia, it is convincible that HbA1c is a strong indicator of T2DM prediction, well over information on genetic predisposition[49]. The decrease in prediction in subpopulations with GPS-4 may imply that HbA1c is a phenotype already inherent and reflected by the genetic predisposition. While model variables such as FPG and HbA1c are stronger predictors than genotype information as shown in Table 7, it is also notable that with sufficient number of validated SNPs, the effect of genetic predisposition could still be evaluated with statistically robust estimates for reclassification improvement over the fully adjusted models.

In the similar context, observed HRs as well as increase in AUC and NRI improvement in full models were greater when testing for effect of GPS-36, than in GPS-16 in both tested subpopulations, although the evaluated indices are not absolute values that allow comparability of different GPSs in different subpopulations. This phenomenon may also be related to the increased volume of pertained information with the increased number of genotyped SNPs. Reported number of SNPs in relation to T2DM in East

Asian population, including South Koreans, are approximately 60 in number, and the observed improvements in prediction ability would be more conspicuous had the information on genetic variation not yet provided been provided[21, 23, 24]. On the other hand, similar results on estimate effect, discrimination and reclassification were found in testing for GPS-16 in the two subpopulations with large difference in number (N = 5,365 vs. 3,045), which could imply the robustness of the analyzed results.

As younger population are subject to less developed clinical risk factors, confirming the findings in a younger population would be meaningful, i.e. environmental or acquired factors such as BMI, HbA1c may be a less important factor in predicting T2DM in younger adults, and the influence by genetic variation may persist even after multiple-variable adjustment[10]. However, results were inconsistent and non-significant in subjects ≤50 years old in the current study, and this may be explained by poor validity due to much decreased number and also due to the middle-aged baseline characteristic of the participants, who may have already begun developing subclinical metabolic disorders.

In the prediction models that included GPSs, we found independent effects of F/Hx of DM and GPS on T2DM risk, with greater HRs by family history than GPS in most cases except for Subpopulation C with GPS-36. While further investigations may be required on the impact of GPS-36 over family history in different, larger populations, the current results may support speculations that family history may provide more information from shared environmental influence, i.e. non-genetic familial behaviors such as lifestyle

and dietary habits, than inherited genetic influence alone[5, 6]. On the other hand, while considering F/Hx of DM is necessary in investigating genetic influence by the risk alleles, continuous encouragement on gene-environment interaction and epigenetics research is suggested to further reveal the missing heritability[3, 20].

2. SNP selection/validation

The selected SNPs tested in the risk prediction have already been validated from previous studies that included genetic information from the same Ansung-Ansan cohort subjects for GWAS or meta-GWAS analyses, and the association tests between the SNPs and T2DM incidence (or prevalence) were restricted to East Asian populations. This method has advantage over a single GWAS in the study population, which face insufficient validity of results due to small number of subjects and limited resource for independent population with identical ethnicity for replication.

The pros and cons of using validated SNPs for which information from same subjects were utilized as subset data warrant further investigation, and attempts for replication in an independent population of identical ethnicity are also suggested.

3. Significant genotypes in the study population

Odds ratios of the SNPs with adjustment for age, sex and BMI ranged from 1.08 to 1.30. Gene names of the sixteen SNPs with significant HRs found in this study were *RBMS1* (rs7593730), *IGF2BP2* (rs1470579), *PPARG* (rs1801282), *CDKAL1* (rs7754840, rs9465871), *JAZF1* (rs864745), *CDKN2A/B* (rs10811661), *HHEX* (rs5015480), *KCNJ11* (rs5215), *SPRY2* (rs1359790), *UBE2E2* (rs6780569), *CDKAL1* (rs7756992), *GCK* (rs4607517), *SLC30A8* (rs13266634), *TCF7L2* (rs7903146), *CENTD2* (rs1552224), *KCNQ1* (rs2237892). The effect of most of the selected SNPs on T2DM risk have been tested in East Asian population[21, 50, 51]. As previously arranged by a well-documented literature, biological mechanism related to most of the significant SNPs discovered in the current study is explained by relations to insulin secretion, either by β-cell dysfunction (e.g. *IGF2BP2*, *JAZF1*, *GCK*, *SLC30A8*, *TCF7L2*, *KCNQ1*, *KCNJ11*, *CENTD2*, *C2CD4A*) or impaired β-cell development(e.g. *CDKAL1*, *CDKN2A/B*, *HHEX*). Two genes, *FTO* and *PPARG* are related with insulin resistance, with obesity or insulin actions[11].

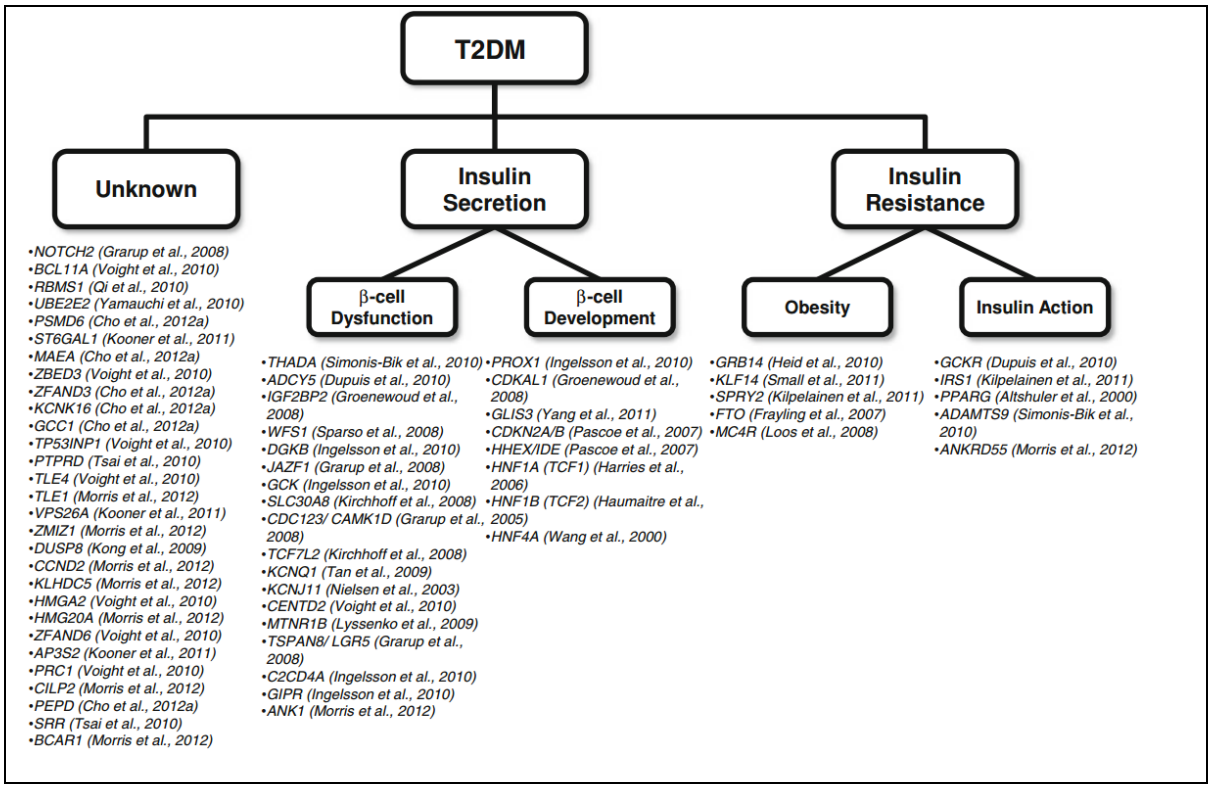

Figure 2. Suggested function of genes associated with T2DM in GWA studies. References indicate the first association results for variants with unknown function or the reports for the physiologic functional analysis (Excerpt from Kwak SH et al., 2013[11])

4. Gene-environment interaction effect

In the gene-environment interaction analysis, both GPS-16 and GPS-36 showed some significant associations with environmental factors such as age (in subjects tested withGPS-36), F/Hx of DM (with GPS-16 in Subpopulations C and B), HDL-C (with GPS-16 in Subpopulation B), and HbA1c (with GPS-16 in Subpopulations C and B). Measures of interaction, either by multiplicative or additive scales, were non-significant with regular P/E, BMI, TG and FPG in the current study. Most of the significant interaction effects were in additive scales, and RERI >1 was found in the interaction analysis for F/Hx of DM (1.01 (1.02-1.89)), implying that there is stronger ground for evidence for the observed interaction[31]. While largest significant AP (45%) was also observed with F/Hx of DM, multiplicative interaction was found significant with F/Hx of DM.

While it may be simply interpreted that the risk allele carriers of GPS-16 may benefit by protection from T2DM risk when they do not have F/Hx of DM but also are positioned at increased risk with the known F/Hx, application of these results should be very carefully considered, as F/Hx of DM shares inherited genetic influence with the GPSs.

As for other environmental factors with significant positive additive interaction (i.e. HDL-C and HbA1c), extended analyses in independent population of same ethnicity is suggested as any confirmed *g*e* interaction effect may be able to contribute to public health measures[31].

Further expanded investigations are required on the analyses, in accordance with availability of validation population. In particular, it should be noted that the significant HRs for F/Hx of DM in multiplicative interaction was as large as 1.80. According to our calculations, smaller size of estimate would require much large number of subjects than in the current study. In the same context, more robust results could have been expected with strong predictors such as FPG and HbA1c at sufficient number of study subjects, i.e. results by interaction analyses in the current study could be improved in their statistical reliability if the larger number of study subjects had been available.

5. Limitations

Although risk prediction models were constructed from a prospective cohort study, duration of follow-up was relatively short. Longer follow-up duration could improve prediction ability of genetic variants relative to time-varying factors (e.g. clinical examination findings), as discrimination power of GPS increase with extended follow-up period[7, 16]. Also, lifestyle risk factors such as smoking and diet could not be considered in the prediction models due to statistical insignificance of their influence on T2DM and subsequent elimination by statistical procedures, despite the alleged influence to the disease[52].

CONCLUSION

Influence of genetic predisposition on modeling risk prediction of T2DM incidence in an 8-years cohort of middle-aged Koreans was robust in a subpopulation with multiple number of 16 and 36 validated genetic variants, while other subpopulations with little genetic variation information showed weakened reclassification ability when strong environmental predictor variables such as FPG and HbA1c were added.

While some gene-environment interaction effect was verified with environmental factors such as family history of DM, HDL-C and HbA1c, further interaction analyses in an independent population of same ethnicity is suggested.

REFERENCES

1. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, 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(5):638-45.
2. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579-89.
3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.
4. Herder C, Roden M. Genetics of type 2 diabetes: pathophysiologic and clinical relevance. Eur J Clin Invest. 2011;41(6):679-92.
5. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med. 2008;359(21):2208-19.
6. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ. 2010;340:b4838.
7. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med. 2008;359(21):2220-32.

8. Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisterman EF, et al. Predicting risk of type 2 diabetes mellitus with genetic risk models on the basis of established genome-wide association markers: a systematic review. Am J Epidemiol. 2013;178(8):1197-207.

9. de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, et al.; MAGIC Investigators; DIAGRAM+ Investigators. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. Diabetes Care. 2011;34(1):121-5.

10. Vassy JL, Durant NH, Kabagambe EK, Carnethon MR, Rasmussen-Torvik LJ, Fornage M, et al. A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. Diabetologia. 2012;55(10):2604-12.

11. Kwak SH, Park KS. Genetics of type 2 diabetes and potential clinical implications. Arch Pharm Res. 2013;36(2):167-77.

12. Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E, et al. Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study. Diabetes Care. 2009;32(11):2116-9.

13. Tam CH, Ho JS, Wang Y, Lam VK, Lee HM, Jiang G, et al. Use of net reclassification improvement (NRI) method confirms the utility of combined genetic risk score to predict type 2 diabetes. PLoS One. 2013;8(12):e83093.

14. Imamura M, Shigemizu D, Tsunoda T, Iwata M, Maegawa H, Watada

H, et al. Assessing the clinical utility of a genetic risk score constructed using 49 susceptibility alleles for type 2 diabetes in a Japanese population. J Clin Endocrinol Metab. 2013;98(10):E1667-73.

15. Willems SM, Mihaescu R, Sijbrands EJ, van Duijn CM, Janssens AC. A methodological perspective on genetic risk prediction studies in type 2 diabetes: recommendations for future research. Curr Diab Rep. 2011;11(6):511-8. Review.

16. Vassy JL, Meigs JB. Is genetic testing useful to predict type 2 diabetes? Best Pract Res Clin Endocrinol Metab. 2012;26(2):189-201.

17. Kwak SH, Choi SH, Kim K, Jung HS, Cho YM, Lim S, et al. Prediction of type 2 diabetes in women with a history of gestational diabetes using a genetic risk score. Diabetologia. 2013;56(12):2556-63.

18. Shea J, Agarwala V, Philippakis AA, Maguire J, Banks E, Depristo M, et al. Myocardial Infarction Genetics Consortium. Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. Nat Genet. 2011;43(8):801-5.

19. Schmoor C, Sauerbrei W, Schumacher M. Sample size considerations for the evaluation of prognostic factors in survival analysis. Stat Med. 2000;19(4):441-52.

20. Patel CJ, Chen R, Kodama K, Ioannidis JP, Butte AJ. Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. Hum

Genet. 2013;132(5):495-508.

21. Cho YS, Chen CH, Hu C, Long J, Ong RT, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2011;44(1):67-72.

22. Lim NK, Park SH, Choi SJ, Lee KS, Park HY. A risk score for predicting the incidence of type 2 diabetes in a middle-aged Korean cohort: the Korean genome and epidemiology study. Circ J. 2012;76(8):1904-10.

23. Ryoo H, Woo J, Kim Y, Lee C. Heterogeneity of genetic associations of CDKAL1 and HHEX with susceptibility of type 2 diabetes mellitus by gender. Eur J Hum Genet. 2011;19(6):672-5.

24. Shu XO, Long J, Cai Q, et al. Identification of new genetic risk variants for type 2 diabetes. PLoS Genet. 2010;6(9):e1001127.

25. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet. 2010;42(11):937-48.

26. Collett D. Modelling Survival Data in Medical Research, Second Edition. Boca Raton: Chapman and Hall/CRC. 2003.

27. 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(2):157-72.

28. Hosmer DW, Lemeshow S. Confidence interval estimation of

interaction. Epidemiology. 1992;3(5):452-6.

29. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. Epidemiology. 1996;7(3):286-90.

30. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol. 2012;41(2):514-20.

31. VanderWeele TJ, Knol MJ. A tutorial on interaction. Epidemiol Method. 2014;3:33-72.

32. Rothman KJ. The estimation of synergy or antagonism. Am J Epidemiol. 1976;103(5):506-11.

33. Knol MJ, VanderWeele TJ, Groenwold RH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. Eur J Epidemiol. 2011;26(6):433-8.

34. Qi L, Cornelis MC, Kraft P, Stanya KJ, Linda Kao WH, Pankow JS, et al. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum Mol Genet 2010;19(13):2706-15.

35. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316(5829):1331-6.

36. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316(5829):1341-5.

37. Li X, Li Y, Song B, Guo S, Chu S, Jia N, et al. Hematopoietically-

Expressed Homeobox Gene Three Widely-Evaluated Polymorphisms and Risk for Diabetes: A Meta-Analysis. PloS One. 2012;7(11):e49917.

38. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316(5829):1336-41.

39. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, 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(5826):889-94.

40. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42(2):105-16.

41. Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet 2009;41(10):1110-5.

42. Yamauchi T, Hara K, Maeda S, Yasuda K, Takahashi A, Horikoshi M, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet 2010;42(10):864-8.

43. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL,

Daly A, et al. Common variants in WFS1 confer risk of type 2 diabetes. Nat Genet 2007;39(8):951-3.

44. Peng F, Hu D, Gu C, Li X, Li Y, Jia N, et al. The relationship between five widely-evaluated variants in CDKN2A/B and CDKAL1 genes and the risk of type 2 diabetes: a meta-analysis. Gene. 2013;531(2):435-43.

45. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445, 881-5

46. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 2008;40(9):1092-7. Epub 2008/08/20. doi: 10.1038/ng.207. PubMed PMID: 18711367.

47. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. Nature 2009;462(7275):868-74.

48. Heianza Y, Arase Y, Hsieh SD, Saito K, Tsuji H, Kodama S, et al. Development of a new scoring system for predicting the 5 year incidence of type 2 diabetes in Japan: the Toranomon Hospital Health Management Center Study 6 (TOPICS 6). Diabetologia. 2012;55(12):3213-23.

49. Selvin E, Steffes MW, Zhu H, Matsushita K, Wagenknecht L, Pankow J, et al. Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. N Engl J Med. 2010;362(9):800-11.

50. Tabara Y, Osawa H, Kawamoto R, Onuma H, Shimizu I, Makino H, et al. Genotype risk score of common susceptible variants for prediction of type 2 diabetes mellitus in Japanese: the Shimanami Health Promoting Program (J-SHIPP study). Development of type 2 diabetes mellitus and genotype risk score. Metabolism. 2011;60(11):1634-40.

51. Yang L, Zhou X, Luo Y, Sun X, Tang Y, Guo W, et al. Association between KCNJ11 gene polymorphisms and risk of type 2 diabetes mellitus in East Asian populations: a meta-analysis in 42,573 individuals. Mol Biol Rep. 2012;39(1):645-59.

52. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369(10):954-64.

국문 초록

서론: 우리나라 중년 인구집단을 대상으로 한 전향적 연구에서 제 2 형 당뇨병(T2DM)의 위험 예측에 대해 유전적 소인(genetic predisposition)의 기여 정도 및 유의한 유전자형 변이에 대해 유전-환경 상호작용 영향을 조사하였다.

방법: 8 년 간 추적 조사한 지역사회기반 코호트 연구 대상자 6,910 명에 대해 4, 16, 36 개의 단일염기변이(SNP)를 선정하여 유전적 소인 점수(GPS)를 구성하고, 위험 예측 모형을 이용하여 이들의 영향을 평가하였다. 또한, 분석을 통해 유의한 변수로 확인된 유전 및 환경 요인에 대해 T2DM 발생에 대한 상호작용 영향을 조사하였다.

결과: 연구 인구집단에서 16 개의 SNP 이 T2DM 과 유의한 관련성을 나타내는 것을 관찰하였으며, GPS-4, GPS-8, GPS-36 의 위험대립유전자가 중위수 이상 증가 시 위험비(hazard ratio)는 각각 1.19 (95% 신뢰구간: 1.04-1.36), 1.23 (1.08-1.41), 1.42 (1.19-1.70)으로 확인되었다. GPS-16 및 GPS-36 의 정보를 가진 집단에서 GPS 의 추가에 따른 곡선하면적(AUC)의 변화가 유의했으며, GPS-4, 8, 36 에서 모두 유의하게 나타난 재분류 지표(NRI)도 당화혈색소(HbA1c)를 추가한 모형에서도 GPS-16 및 GPS-36 정보를 가진 집단에서만 유의성이 유지되었다. GPS-16, GPS-

36 과 환경요인 중 가족력, 고밀도 지질단백질 콜레스테롤, HbA1c 는 양의 방향의 유의한 덧셈상호작용을 나타냈으며, 가족력은 곱셈 상호작용도 유의하게 나타났다.

결론: 우리나라 중년 코호트 연구에서, 다수의 SNP 에 대한 정보를 가진 대상집단에서 T2DM 의 위험 예측 모형에 유전적 소인을 추가하는 것에 대해 차별(discrimination) 및 재분류(reclassification) 기량이 유의한 것을 확인하였으며, T2DM 의 강력한 예측인자인 HbA1c 를 추가한 모형에서도 유의성이 지속되었다. 한편, 본 연구에서 유전-환경 상호작용에 대해 추측할 만한 유의한 결과를 도출하였지만, 이에 대해 더 많은 수의 대상자 또는 본 연구대상자와는 독립적인 다른 집단에서의 확대연구를 제안하고자 한다.

주요어 : 당뇨병, 유전-환경 상호작용, 유전적 소인, 위험 예측
학 번 : 2012-31144