Active cigarette smoking is associated with an exacerbation of genetic susceptibility to diabetes

Wan-Yu Lin 1,2*, Yu-Li Liu 3, Albert C. Yang 4,5, Shih-Jen Tsai 5,6,7, Po-Hsiu Kuo 1,2*

1 Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
2 Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan
3 Center for Neuropsychiatric Research, National Health Research Institutes, Miaoli, Taiwan
4 Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA, USA
5 Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan
6 Division of Psychiatry, National Yang-Ming University, Taipei, Taiwan
7 Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan

Short title: Smoking and genetic risk of diabetes

* Corresponding authors: Po-Hsiu Kuo, Ph.D. and Wan-Yu Lin, Ph.D.

Po-Hsiu Kuo, Ph.D. (http://orcid.org/0000-0003-0365-3587)
Room 521, No. 17, Xu-Zhou Road, Taipei 100, Taiwan
Phone: +886-2-33668015; Fax: +886-2-23511955; E-mail: phkuo@ntu.edu.tw

Wan-Yu Lin, Ph.D. (http://orcid.org/0000-0002-3385-4702)
Room 501, No. 17, Xu-Zhou Road, Taipei 100, Taiwan
Phone/Fax: +886-2-33668106; E-mail: linwy@ntu.edu.tw
Active cigarette smoking is associated with an exacerbation of genetic susceptibility to diabetes. This study included a discovery cohort (TWB1) of 25,460 and a replication cohort (TWB2) of 58,774 unrelated Taiwan Biobank (TWB) subjects. Genetic risk score (GRS) of each TWB2 subject was calculated with weights retrieved from TWB1 analyses. We then assessed the significance of GRS-smoking interactions on FG/HbA1c/diabetes while adjusting for covariates. A total of 5 smoking measurements were investigated respectively, including “active smoking status”, “pack-years”, “years as a smoker”, “packs smoked per day”, and “hours as a passive smoker per week”. Except passive smoking, all smoking measurements were associated with FG/HbA1c/diabetes ($P < 0.0033$) and were associated with an exacerbation of the genetic risk of FG/HbA1c ($P_{Interaction} < 0.0033$). For example, each 1 standard deviation increase in GRS is associated with a $1.68\%$ higher FG in subjects consuming one more pack of cigarettes per day ($P_{Interaction} = 1.9 \times 10^{-7}$). Figure 1 shows that the GRS effects on FG/HbA1c/diabetes were larger in smokers than in non-smokers. One of the mechanisms behind gene-smoking interactions is epigenetics. When one is smoking, carbon monoxide displaces oxygen from his/her hemoglobin. This will decrease the amount of oxygen available for delivery to many tissues. Smoking can therefore affect one’s entire body including cardiovascular and circulatory systems. We analyzed the blood DNA methylation data of 2,091 TWB subjects, and found cigarette consumption is linked to differential DNA methylation of some diabetes susceptibility genes, such as the $KCNQ1$ gene (potassium voltage-gated channel subfamily Q member 1). Through its link with differential DNA methylation, smoking can regulate gene expressions and lead to more damage to people who are more genetically predisposed to diabetes.
Abstract

The heritability levels of two traits for diabetes diagnosis, fasting serum glucose (FG) and glycated hemoglobin (HbA1c), were estimated to be 51% ~ 62%. Studies have shown that cigarette smoking is a modifiable risk factor for diabetes. It is important to uncover whether smoking may modify the genetic risk of diabetes. This study included a discovery cohort (TWB1) of 25,460 and a replication cohort (TWB2) of 58,774 unrelated Taiwan Biobank subjects. Genetic risk score (GRS) of each TWB2 subject was calculated with weights retrieved from TWB1 analyses. We then assessed the significance of GRS-smoking interactions on FG/HbA1c/diabetes while adjusting for covariates. A total of 5 smoking measurements were investigated respectively, including “active smoking status”, “pack-years”, “years as a smoker”, “packs smoked per day”, and “hours as a passive smoker per week”. Except passive smoking, all smoking measurements were associated with FG/HbA1c/diabetes ($P < 0.0033$) and were associated with an exacerbation of the genetic risk of FG/HbA1c ($P_{Interaction} < 0.0033$). For example, each 1 standard deviation increase in GRS is associated with a 1.68% higher FG in subjects consuming one more pack of cigarettes per day ($P_{Interaction} = 1.9 \times 10^{-7}$). Smoking cessation is especially important for people who are more genetically predisposed to diabetes.

Keywords: Gene-smoking interactions, Gene-environment interactions, Genetic risk score, Polygenic risk score.

Introduction

Diabetes have influenced hundreds of millions people around the world, and its prevalence is gradually increasing (1). Genetic factors play an important role in the development of diabetes.
Fasting glucose (FG) and glycated hemoglobin (HbA1c) are commonly used for the diagnosis of diabetes. An FG level > 126 mg/dL (2) and an HbA1c level > 6.5% (48 mmol/mol) (3) have been proposed as diagnostic indicators for diabetes. The heritability levels of FG and HbA1c were estimated to be 51% and 62%, respectively (4). The remaining variability in FG and HbA1c could be explained by lifestyle factors.

Smoking has been found to be associated with a higher risk of insulin resistance and diabetes (5). Nicotine is thought to be responsible for the association between cigarette smoking and the development of diabetes (6). However, smoking has not been uniformly regarded as a risk factor of diabetes (7). Moreover, it is important to uncover whether smoking may modify the genetic risk of diabetes.

To know whether the genetic risk of diabetes may vary with cigarette smoking, we here investigated “gene-smoking interactions” (G×S) on FG, HbA1c, and the dichotomous diabetes status, respectively. Our Taiwan Biobank (TWB) data included a discovery cohort (TWB1) and a replication cohort (TWB2). Genetic risk score (GRS) of each TWB2 subject was calculated with weights retrieved from TWB1 analyses. We tested whether the GRS effects could be modified by “active smoking status”, “the number of pack-years”, “years as a smoker”, “packs smoked per day”, or “hours as a passive smoker per week”. This study aims to uncover whether smoking can modulate the genetic predisposition to diabetes, and which smoking measurement is the most critical effect modifier.

Research Design and Methods

Taiwan Biobank

The TWB keeps collecting genomic and lifestyle information from Taiwan residents aged
To participate in the TWB, community-based volunteers signed informed consent, took physical examinations, and provided blood and urine samples. Their lifestyle factors were further collected through a face-to-face interview with TWB researchers. TWB received ethical approval from the Ethics and Governance Council of Taiwan Biobank, Taiwan, and also from the Institutional Review Board on Biomedical Science Research/IRB-BM, Academia Sinica, Taiwan. Our study was approved by the Research Ethics Committee of National Taiwan University Hospital (NTUH-REC no. 201805050RINB).

This study included 27,737 and 67,512 subjects who have been whole-genome genotyped by the TWB1 and TWB2 genotyping arrays until February, 2020, respectively. To explore cryptic relatedness, we used PLINK 1.9 (9) to estimate PI-HAT = Probability(IBD = 2) + 0.5 × Probability(IBD = 1), where IBD is the genome-wide identity by descent (IBD) sharing coefficients between any two TWB individuals. Similar to many genetic studies (10; 11), we also excluded relatives by removing one subject from every pair with PI-HAT \( \geq 0.2 \). A total of 25,460 unrelated TWB1 subjects and 58,774 unrelated TWB2 subjects remained after this quality control process.

Most TWB subjects were of Han Chinese ancestry (8). Released in April 2013, the TWB1 genotyping array was designed for Taiwan’s Han Chinese and run on the Axiom Genome-Wide Array Plate System (Affymetrix, Santa Clara, CA, USA). Based on user experience in TWB1 and next-generation sequencing of ~1,000 TWB individuals, the TWB2 genotyping array was released in August 2018 to further cover specific SNPs in Taiwan population. A total of 632,172 and 648,611 autosomal SNPs were genotyped in TWB1 and TWB2 arrays, respectively. In TWB1, we removed 27,628 SNPs with genotyping rates < 95%, and 6,900 SNPs with Hardy-Weinberg test \( P \)-values \( < 5.7 \times 10^{-7} \) (12). In TWB2, we removed 27,859 SNPs with
genotyping rates < 95%, and 14,656 SNPs with Hardy-Weinberg test $P$-values $< 5.7 \times 10^{-7}$ (12). Finally, 597,644 TWB1 SNPs and 606,096 TWB2 SNPs were kept in our analysis and were used to construct ancestry principal components (PCs). A total of 92,052 SNPs were overlapped between these two arrays.

We imputed the genotypes of autosomal SNPs using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html), with the reference panel based on the East Asian (EAS) population from the 1000 Genomes Phase 3 v5. After removing SNPs with Hardy-Weinberg test $P$-values $< 5.7 \times 10^{-7}$ (12) and SNPs with low imputation information score (Rsq $< 0.8$), 7,433,014 and 6,347,468 SNPs remained in TWB1 and TWB2, respectively.

**Three diabetes-related traits**

We here analyzed two continuous traits related to the diagnosis of diabetes, FG and HbA1c. After a minimum 6-hour fast (no calorie intake for at least 6 hours), serum glucose was measured using a Hitachi LST008 analyzer (Hitachi High-Technologies, Tokyo, Japan), whereas HbA1c was measured with the Trinity Biotech Premier Hb9210 analyzer (Bray, Ireland/Kansas City, MO, USA). Same as the criterion used worldwide (2; 3), the Ministry of Health and Welfare in Taiwan defined an FG level higher than 126 mg/dL or an HbA1c level higher than 6.5% (48 mmol/mol) as an indicator of diabetes.

Moreover, diabetes status was also analyzed as a dichotomous trait, where subjects with diabetes included those with physician-diagnosed diabetes, or those having FG $> 126$ mg/dL or HbA1c $> 6.5$ % (48 mmol/mol) according to the TWB test results.

**Definition of smoking, drinking, regular exercise, and educational attainment**

In TWB, “active smoker” was defined as a subject who had actively smoked for at least 6
months and had not quit smoking at the time his/her FG and HbA1c were measured. Moreover, TWB researchers also asked every active smoker “how many packs of cigarettes smoked per day” and “how many years as an active smoker”. “The number of pack-years” was then calculated as multiplying “the number of packs smoked per day” by “years as a smoker”. Furthermore, TWB researchers asked each individual “the number of hours as a passive smoker per week”. In total, 5 smoking measurements will be investigated in the following analyses.

Sex, age, body mass index (BMI), alcohol drinking status, regular exercise, and educational attainment were regarded as covariates and adjusted in all of our regression analyses. Drinking was defined as a subject having a weekly intake of more than 150 cc of alcohol for at least 6 months and having not stopped drinking at the time his/her FG and HbA1c were assessed. Regular exercise was defined as engaging in 30 minutes of “exercise” three times a week. “Exercise” indicates leisure-time activities such as jogging, mountain climbing, yoga, etc (13).

Educational attainment was recorded through a face-to-face interview with TWB researchers. It was represented by a number ranging from 1 to 7, where 1 indicates “illiterate”, 2 means “no formal education but literate”, 3 denotes “primary school graduate”, 4 represents “junior high school graduate”, 5 indicates “senior high school graduate”, 6 means “college graduate”, and 7 denotes “Master’s or higher degree”. Similar to a recent study for diabetes (14), we also considered “educational attainment” as an covariate in all analyses.

**Association between smoking and FG/HbA1c/diabetes**

Initially, to test whether smoking is significantly associated with FG/HbA1c/diabetes, we regressed each trait according to the following model:

$$g[E(Y)] = \beta_0 + \beta_{Smoking}Smoking + \sum_{v=1}^{16} \beta_{Covariate_v}Covariate_v,$$

where $Y$ is natural log transformed FG (or HbA1c) with an identity link $g[\cdot]$, or the diabetes
status with a logit link \( g[\cdot] \). Smoking is coded as 1 for active smokers and 0 otherwise.

Sex, age, BMI, educational attainment, and ancestry PCs are commonly adjusted in models for FG/HbA1c/diabetes (14; 15). Regular physical exercise has been shown to reduce the risk of insulin resistance, metabolic syndrome and diabetes (16). Effects of cigarette smoking and alcohol consumption on diabetes are usually investigated (5-7; 17). Therefore, a total of 16 covariates were adjusted in model (1), including sex, age (in years), BMI, drinking status (yes vs. no), regular exercise (yes vs. no), educational attainment (a value ranging from 1 to 7), and the first 10 ancestry PCs.

Because FG and HbA1c were skewed to the right, both traits were natural log transformed to improve the R-square of model (1). Such a transformation was commonly used in analyses for FG and HbA1c (18; 19). Based on model (1), we later replaced “active smoking status” with the 4 continuous smoking measurements, respectively.

**Genetic risk score (GRS)**

To build trait-specific GRS, we first identified FG-associated SNPs, HbA1c-associated SNPs, and diabetes-associated SNPs based on the 25,460 unrelated TWB1 subjects. Each trait (denoted by \( Y \)) was regressed on every SNP while adjusting for the 16 covariates, as follows:

\[
g[E(Y)] = \beta_0 + \beta_{SNP,j}SNP_j + \sum_{v=1}^{16} \beta_{C_v}Covariate_v, \quad j = 1, \ldots, 7433014. \tag{2}
\]

\( P_{SNP,j} \), the \( p \)-value of testing \( H_0: \beta_{SNP,j} = 0 \ vs. \ H_1: \beta_{SNP,j} \neq 0 \) from model (2), calibrates the significance of marginal association of SNP \( j \) with \( Y \). Because of ~8 million SNPs in TWB1, SNP \( j \) is significantly associated with a trait if \( P_{SNP,j} < 6.25 \times 10^{-9} = \frac{0.05}{8 \times 10^6} \). We calculated GRS for each of the 58,774 unrelated TWB2 subjects by

\[
GRS' = \sum_{j=1}^{7433014} I(P_{SNP,j} < 6.25 \times 10^{-9})\hat{\beta}_{SNP,j}SNP_j, \tag{3}
\]
where $\hat{\beta}_{SNP,j}$ ($j = 1, \cdots, 7433014$) had been estimated from model (2), and $SNP_j$ is the number of minor alleles at the $j^{th}$ SNP (0, 1, or 2) for the TWB2 individuals. The indicator function $I(P_{SNP,j} < 6.25 \times 10^{-9})$ is 1 if $P_{SNP,j} < 6.25 \times 10^{-9}$ and 0 otherwise.

Because $GRS'$ is composed of SNPs that are more associated with FG/HbA1c/diabetes (satisfying $P_{SNP,j} < 6.25 \times 10^{-9}$), this is the so-called “marginal-association filtering” in gene-environment interaction analyses (20-23). To avoid collinearity in a GRS, we used the default setting of the clumping procedure in PLINK 1.9 (9). To be specific, if there were multiple variants with $P_{SNP,j} < 6.25 \times 10^{-9}$ within a 250-kb region, we kept the most significant one (so-called “index variant”) and removed others in linkage disequilibrium with the index variant ($r^2 > 0.5$). To quantify how many standard deviations (sds) a $GRS'$ is from the mean, we performed the z-score transformation on $GRS'$ and then obtained $GRS$. By fitting model (2) for FG/HbA1c/diabetes separately, we obtained $GRS$ for FG/HbA1c/diabetes, respectively.

$SNP_j$ is the number of minor alleles at the $j^{th}$ SNP (0, 1, or 2). A positive $\hat{\beta}_{SNP,j}$ indicates that the minor allele is trait-increasing, and a subject with more copies of the minor allele (more trait-increasing alleles) will receive an increment in his/her $GRS$. In contrast, a negative $\hat{\beta}_{SNP,j}$ represents that the minor allele is trait-decreasing, and a subject with more copies of the minor allele (more trait-decreasing alleles) will obtain a decrement in his/her $GRS$. Finally, a higher $GRS$ is associated with a larger FG/HbA1c/diabetes.

**European-derived GRS**

In addition to the “TWB1-derived GRS” (for simplicity, abbreviated as “GRS”), we also calculated a “European-derived GRS” (abbreviated as “EuGRS”), based on 38 diabetes-associated SNPs that were previously identified from genome-wide association studies
GWASs) (39-42). Most of these 38 SNPs were identified at the commonly-used genome-wide significance level ($P_{SNP} < 5 \times 10^{-8}$). Individuals of those GWASs (39-42) were overwhelmingly of European descent. The European-derived GRS was calculated as $EuGRS = \sum_{j=1}^{38} w_j SNP_j$, where the weights ($w_j, j = 1, \ldots, 38$) were the effect sizes reported by previous GWASs (39-42) and summarized by Said et al. (14).

To present an analysis parallel to GRS-smoking interaction analysis (where TWB1 was used to find trait-associated SNPs and TWB2 was used to test for interactions), EuGRS-smoking interaction analysis was also performed on the TWB2 cohort with 58,774 subjects. There is one more reason why we did not include the TWB1 cohort into the EuGRS-smoking interaction analysis. While 32 out of the 38 SNPs were genotyped by the TWB2 array, only 27 were genotyped by the TWB1 array. After imputation, still 2 SNPs in TWB1 failed to achieve the criterion of imputation information score (i.e., their Rsq < 0.8). Therefore, we did not compute EuGRS for the 25,460 TWB1 subjects. To sum up, both GRS-smoking interaction and EuGRS-smoking interaction were evaluated in the TWB2 cohort with 58,774 subjects.

**GRS-smoking interactions**

We then used $GRS$ to test for the presence of $G \times S$. The following model was considered for each trait, respectively:

$$g[E(Y)] =$$

$$\phi_0 + \phi_{GRS}GRS + \phi_{Smoking}Smoking + \phi_{INT}GRS \times Smoking + \sum_{v=1}^{16} \phi_{C_v} Covariate_v +$$

$$\sum_{v=1}^{16} \phi_{GC_v} GRS \times Covariate_v + \sum_{v=1}^{16} \phi_{SC_v} Smoking \times Covariate_v,$$

where $Y$ is natural log transformed FG/HbA1c or the diabetes status, and $GRS$ is the standardized GRS z-score of the TWB2 subjects. Covariates adjusted for FG/HbA1c/diabetes have been
described under model (1). To further control confounding (24), \( GRS \times Covariate_\nu \) and \( Smoking \times Covariate_\nu \) \((\nu = 1, \ldots, 16)\) were also included in model (4). To interpret main effects in model (4), we centered all explanatory variables at their means (25).

Let the \( p \)-value of testing \( H_0: \phi_{INT} = 0 \) vs. \( H_1: \phi_{INT} \neq 0 \) be \( P_{INT} \). The presence of \( G \times S \) will be declared if \( P_{INT} < \frac{0.05}{3 \times 5} = 0.0033 \), because 3 diabetes-related traits and 5 smoking measurements were investigated. This significance level for GRS analyses was determined a priori and would be used throughout this study.

**DNA methylation data**

One of the mechanisms behind \( G \times S \) is epigenetics (33). Cigarette consumption has been found to be linked to differential DNA methylation in some diabetes susceptibility genes, such as \( KCNQ1 \) (potassium voltage-gated channel subfamily Q member 1) (34).

When one is smoking, carbon monoxide displaces oxygen from his/her hemoglobin. This will decrease the amount of oxygen available for delivery to many tissues. Smoking can therefore affect one’s entire body including cardiovascular and circulatory systems (35). TWB also released blood DNA methylation data of 2,091 TWB subjects. These individuals were randomly selected from TWB1 (833 subjects) and TWB2 (1,258 subjects). After GRS-smoking interaction analysis, we also explored the associations between smoking and DNA methylation of diabetes susceptibility genes.

**Data and Resource Availability**

Individual-level TWB data are available upon application to TWB (https://www.twbiobank.org.tw/new_web/). Our application for the TWB data was approved in
February, 2020, with the accession number “TWBR10810-07”.

[Table 1 is approximately here]

**Results**

**Association between smoking and FG/HbA1c/diabetes**

Our discovery cohort contained 25,460 TWB1 subjects, where 3,005 (11.8%) were (active) smokers and 22,455 were non-smokers. The replication cohort included 58,774 TWB2 subjects, 5,173 (8.8%) were (active) smokers and 53,601 were non-smokers. Table 1 presents basic characteristics of these two cohorts. The average FG and HbA1c levels are higher in smokers than in non-smokers. Model (1) was fitted to assess whether smoking is significantly associated with FG/HbA1c/diabetes, while considering the covariates. As shown by Table 2, male, aging, and a larger BMI, are associated with increased FG/HbA1c and risk of diabetes.

[Table 2 is approximately here]

Because FG/HbA1c was natural log transformed, \( \exp(\hat{\beta}_{\text{Smoking}}) - 1 \) \times 100% is shown to represent the percent change in FG/HbA1c that is associated with “active smoking status”, while adjusting for sex, age, BMI, drinking status, regular exercise, educational attainment, and the first 10 PCs. Both our discovery cohort (TWB1) and replication cohort (TWB2) showed, compared with non-smokers, active smokers have an odds ratio (OR) of 1.36 for diabetes (95% confidence interval [C.I.] by TWB1: 1.19-1.57; TWB2: 1.23-1.51).

[Table 3 is approximately here]

Based on model (1), we further analyzed the 4 continuous smoking measurements, respectively. For example, TWB1 (TWB2) showed that, subjects consuming one more pack of cigarettes per day have an OR of 1.51 (1.41) for diabetes (95% C.I. by TWB1: 1.32-1.73; TWB2:
The detailed results of model (1) for the 4 continuous smoking measurements can be found in Supplemental Tables S1-S4.

**GRS-smoking interactions in FG/HbA1c/diabetes**

Analyzing TWB1 according to model (2), a total of 16, 12, and 6 SNPs were identified to have $P_{SNP} < 6.25 \times 10^{-9}$ in FG, HbA1c, and diabetes analyses, respectively. The information of these SNPs was listed in Supplemental Tables S5-S7. A total of 5 out of the 16 FG-associated SNPs, 4 out of the 12 HbA1c-associated SNPs, and 2 out of the 6 diabetes-associated SNPs were also identified by another trait. The GRS of each TWB2 individual was then calculated based on these 16/12/6 SNPs, with weights obtained from TWB1 analysis (i.e., $\hat{\beta}_{SNP}$s from model 2).

[Table 4 is approximately here]

Table 4 presents the results of model (4) when the smoking measurement is “active smoking status”. FG (or HbA1c) was natural log transformed, and therefore $(\exp(\hat{\phi}_{INT}) - 1) \times 100\% = 1.05\%$ represents that each 1 sd increase in GRS was associated with a 1.05% higher FG (or 0.51% higher HbA1c) in active smokers than in non-smokers ($P_{INT-FG} = 2.8 \times 10^{-5}$; $P_{INT-HbA1c} = 0.002$). Each 1 sd increase in GRS was associated with a 1.16 times OR (95% C.I.: 1.05-1.28) for diabetes in active smokers than in non-smokers ($P_{INT-diabetes} = 0.0046$). After analyzing the 4 continuous smoking measurements sequentially according to model (4), we found all smoking measurements, except “hours as a passive smoker per week”, are associated with the exacerbation of the genetic risk of diabetes (Table 5).

[Table 5 is approximately here]

For example, each 1 sd increase in GRS is associated with a 1.68% higher FG (or 0.69% higher HbA1c) in subjects with one more pack of cigarettes per day ($P_{INT-FG} = 1.9 \times 10^{-7}$;
\[ P_{\text{INT-}HbA1c} = 4.5 \times 10^{-4} \]. The detailed results of model (4) for the 4 continuous smoking measurements can be found in Supplemental Tables S8-S11.

Figure 1 shows the average of FG/HbA1c and the prevalence of diabetes stratified by smoking status and the quintiles of the FG/HbA1c/diabetes GRS. The GRS effects on FG/HbA1c/diabetes were larger in smokers than in non-smokers. Supplemental Figures S1-S4 present similar illustrations for the other 4 smoking measurements. Except “passive smoking” (Figure S4), all smoking measurements exhibit interactions with GRS on FG/HbA1c/diabetes. This is in line with the result shown in Table 5.

[Figure 1 is approximately here]

Among the 5 smoking measurements, only “active smoking status” and “years as a smoker” presented interactions with GRS on diabetes under \( \alpha = 0.05 \) (Table 5). The power to identify diabetes-associated SNPs was low, because only 2,207 (8.7%) among the 25,460 TWB1 subjects had diabetes. As a result, we merely detected 6 diabetes-associated SNPs at the significance level of \( 6.25 \times 10^{-9} \) (Supplemental Table S7). The GRS constructed by these 6 SNPs may not be able to well represent the genetic susceptibility to diabetes. Moreover, only 5,308 (9.0%) among the 58,774 TWB2 subjects had diabetes, the power to detect GRS-smoking interactions was limited. Therefore, none of the five smoking measurements were found to be associated with the exacerbation of the genetic risk of diabetes under \( \alpha = \frac{0.05}{15} = 0.0033 \) (the last column of Table 5).

Although the response variable in model (4) is natural log transformed FG/HbA1c, these significant findings in GRS-smoking interactions are not scale dependent. Supplemental Table S12 shows that these significant results are still replicable when the response variable is FG/HbA1c without natural log transformation.
Multicollinearity between variables in model (4) has been evaluated via variance inflation factor (VIF), computed by the “car” package (https://cran.r-project.org/web/packages/car/). The VIFs under all models were acceptable (smaller than 10), except the model to assess the interaction between GRS and “hours as a passive smoker per week”. To reduce the multicollinearity between variables in model (4), we further assessed GRS-smoking interactions without controlling for the other interaction terms (i.e., no $GRS \times Covariate_v$ or $Smoking \times Covariate_v$, $v = 1, \ldots, 16$). The VIFs under these simpler models were all smaller than 5. Moreover, Supplemental Table S13 shows that the results from simpler models were similar to Table 5.

We then assessed GRS-smoking interactions within each gender group. Supplemental Table S14 shows that the significant GRS-smoking interactions were mostly driven by the male group. This is because 88.1% and 77.4% smokers were males in TWB1 and TWB2, respectively. The detection of GRS-smoking interactions in females could be hampered by the small sample size of female smokers.

**SNP-smoking interactions in FG/HbA1c/diabetes**

Four SNPs (rs2399794, rs7896600, rs1174605899, and rs11257655) in/near the CDC123 (cell division cycle 123) gene were identified to be associated with FG/HbA1c/diabetes. They presented interactions with most smoking measurements except passive smoking (most $P_{INT} < 0.05$, Tables S5-S7). The CDC123 gene has been found to be associated with the dysfunction of pancreatic β-cells (26). The inability of pancreatic β-cells to secrete adequate levels of insulin is a major cause of diabetes (27).

Two SNPs (rs2233580 and rs61342118) in/near the PAX4 (paired box 4) gene were identified to be associated with FG/HbA1c/diabetes. They presented interactions with some
smoking measurements except passive smoking (some $P_{INT} < 0.05$, Tables S5-S7). The PAX4 gene is necessary for the survival and proliferation of pancreatic $\beta$-cells (28). This gene has been found to be hypermethylated in patients with diabetes (29).

Four SNPs (rs163177, rs60808706, rs11024175, and rs163184) in $KCNQ1$ were identified to be associated with FG/HbA1c/diabetes. SNP rs11024175 was found to interact with all the 5 smoking measurements on HbA1c ($P_{INT} < 0.05$, Table S6). $KCNQ1$ mRNAs were highly expressed in adrenal tissues (30). Disorders of the adrenal cortex can result in glucose intolerance and diabetes (31). Consistent with our finding, $KCNQ1$-by-smoking interaction has been reported by a study in Han Chinese (32).

Smoking is associated with DNA methylation of diabetes susceptibility genes

The 2,091 individuals with methylation measures were randomly selected from TWB1 (833 subjects) and TWB2 (1,258 subjects). Their basic characteristics (Supplemental Table S15) are similar to those of the whole TWB (Table 1). We annotated CpG sites available on the Illumina Infinium MethylationEPIC BeadChip to the abovementioned three genes. A total of 72, 84, and 629 CpG sites are within/near $CDC123$, $PAX4$, and $KCNQ1$, respectively (“near” indicates 50 kb in the 3’ and 5’ regions outside the gene boundary). The methylation percentage of a CpG site was reported as a $\beta$-value ranging from 0 (no methylation) to 1 (full methylation). The $\beta$-value of each CpG site was regressed on “packs smoked per day”, while adjusting for age, sex, and BMI (these three covariates were also adjusted in a related study (34)). Although “packs smoked per day” was served as the smoking factor, “active smoking status” provided similar results.

“Packs smoked per day” is associated with differential DNA methylation of $CDC123$ (cg06335123, $p = 0.00025 < 0.05/72$) and $KCNQ1$ (cg26963277, $p = 3.0 \times 10^{-17}$; cg01744331, $p = 4.5 \times 10^{-10}$; cg16556677, $p = 2.7 \times 10^{-5} < 0.05/629$). More packs smoked per day, aging,
and a larger BMI are associated with decreased levels of DNA methylation at these four sites (Supplemental Table S16). The three CpG sites in KCNQ1 (cg26963277, cg01744331, and cg16556677) have been reported to be associated with smoking in the Rotterdam study (34). Our finding from Han Chinese is in line with that study (34).

Through its link with differential DNA methylation, smoking can modulate gene expressions of diabetes susceptibility genes and lead to more damage to people who are more genetically predisposed to diabetes. However, we have not observed the association between smoking and DNA methylation of PAX4. The mechanism behind PAX4-by-smoking interaction needs further investigation.

A limitation is that our DNA methylation measures were made in peripheral blood, and the more relevant tissue is probably pancreatic β-cell. Although we here found that active smoking is associated with methylation at CpG sites in some diabetes susceptibility genes, additional work is necessary to determine if this mediates the smoking-genotype interaction.

**Discussion**

Smoking is associated with insulin resistance, inflammation, and dyslipidemia (36). Consistent with our finding, KCNQ1-by-smoking interaction has been reported by a study in Han Chinese (32). Moreover, Wu et al. recently identified interactions between smoking status and five SNPs at or near four genes (TCF7L2, CUBN, C2orf63, and FBN1) on the risk of diabetes, in subjects of European and African ancestry (15). Two out of the five SNP-smoking interactions can be replicated in our TWB2 HbA1c/diabetes analysis, at the nominal significance level of 0.05 (Supplemental Table S17). Both of the variants locate at the TCF7L2 (transcription factor 7 like 2) gene. TCF7L2 is a diabetes-associated gene identified in subjects of European ancestry (37), but the variants in this gene are not associated with FG/HbA1c/diabetes in TWB1 and so were not
selected to form our GRSs. Variants in \( TCF7L2 \) are associated with pancreatic \( \beta \)-cell function (38).

For diabetes, Langenberg et al. have shown no interactions between genetic risk and physical activity or dietary habits (1). Said et al.’s recent study based on the UK Biobank further found no significant interactions between lifestyle and GRS on diabetes (14). However, their lifestyle is a composite measure of active smoking status, BMI, and physical activity. It remains unclear whether smoking alone may modify the genetic predisposition to diabetes. Moreover, their GRS for diabetes was calculated by 38 diabetes-associated SNPs with effect sizes retrieved from previously published GWASs (39-42). Most of these 38 SNPs were identified at the commonly-used genome-wide significance level \( (P_{SNP} < 5 \times 10^{-8}) \).

Individuals of those GWASs (39-42) were overwhelmingly of European descent. A SNP in the above-mentioned \( TCF7L2 \) gene, rs7903146, is also among the list of the 38 SNPs (Supplemental Tables S18-S20). One of the 38 SNPs, rs13266634 (in the \( SLC30A8 \) gene), is also among our 6 diabetes-associated SNPs (Table S7). Moreover, the 38 SNPs included variants in/near \( CDC123 \) and \( KCNQ1 \), which were also identified to be associated with FG/HbA1c/diabetes according to our discovery cohort (TWB1), as listed in Tables S5-S7.

We then found three smoking factors (“active smoking status”, “the number of pack-years”, and “years as a smoker”) were associated with an exacerbation of \( EuGRS \) on both FG and HbA1c (Supplemental Tables S21-S22), at the nominal significance of 0.05 \( (P_{INT} < 0.05) \). “\( EuGRS \times \) Packs smoked per day” interaction presented a borderline significance on FG \( (P_{INT} = 0.0582) \). Similar with the results based on TWB1-GRS, “\( EuGRS \times \) Hours as a passive smoker per week” interaction was not significant on FG or HbA1c (Supplemental Table S22).

We also analyzed diabetes as a dichotomous trait. Through logistic regression models, we
did not detect any significant EuGRS-smoking interactions. In addition to low statistical power (< 10% subjects with diabetes), this negative finding may also result from the lack of transferability of a European-derived GRS to TWB. Comparing Table 4 with Table S21, we can see that TWB1-GRSs are more significant (smaller p-values) and more predictive (larger R-squares) than EuGRS for all the 3 traits, despite fewer SNPs used for TWB1-GRSs. The detection of EuGRS-smoking interactions could be hampered by the inferior transferability of EuGRS to TWB.

Supplemental Table S23 shows that the results of EuGRS-smoking interactions are still replicable when the response variable is FG/HbA1c without natural log transformation. When assessing EuGRS-smoking interactions without controlling for the other interaction terms (i.e., no EuGRS × Covariate$_v$ or Smoking × Covariate$_v$, $v = 1, \ldots, 16$), Supplemental Table S24 shows that the results from simpler models were similar to Table S22. Moreover, in line with the results from TWB1-GRS, Supplemental Table S25 also shows that the significant EuGRS-smoking interactions were mostly driven by the male group.

Diabetes is a growing health crisis around the world. According to the Centers for Disease Control and Prevention and many previous studies, smoking increases inflammation in the body (43) and also causes oxidative stress (44). Both inflammation and “oxidative stress” have been shown to be linked to an increased risk of diabetes (45). “Oxidative stress” is an imbalance between oxidants and antioxidants in favor of the oxidants (46), potentially modulating gene expressions (47) and leading to more damage to people who are more genetically predisposed to diabetes. Indeed, we here found smoking is associated with DNA methylation at KCNQ1 and CDC123, implying that smoking can play a role in regulating gene expressions of these diabetes susceptibility genes.
According to previous studies (48), the amount of nicotine (the main chemical in cigarettes) absorbed by a passive smoker was between 1/10 and 1/3 of the amount in a cigarette. Therefore, the concentration of nicotine in passive smokers is smaller than that in active smokers (48). This may explain why passive smoking does not significantly modulate the genetic predisposition to diabetes.

The harm of smoking is more impactful in subjects who are more genetically predisposed to diabetes. This study shows that active cigarette smoking is associated with an exacerbation of genetic risk of diabetes. Through constructing a GRS based on diabetes associated SNPs, it is worthwhile to investigate whether smoking may exacerbate the genetic susceptibility to diabetes, even for populations where smoking has not been found to be associated with diabetes.

**Acknowledgements**

The authors would like to thank the editors and three anonymous reviewers for their insightful and constructive comments; Mr. Tsung-Hao Lee and Mr. Yu-Sheng Lee for imputation of the TWB data; Mr. Ya-Chin Lee for assisting with the acquisition of the TWB data.

**Funding**

This study was supported by the Ministry of Science and Technology of Taiwan (grant number MOST 107-2314-B-002-195-MY3 to Wan-Yu Lin). The acquisition of TWB data was supported by two MOST grants (MOST 107-2314-B-002-195-MY3 to Wan-Yu Lin; MOST 102-2314-B-002-117-MY3 to Po-Hsiu Kuo) and a collaboration grant (National Taiwan University Hospital: grant number UN106-050 to Shyr-Chyr Chen and Po-Hsiu Kuo).
Duality of Interest

No potential conflicts of interest relevant to this article were reported.

Author Contributions

W.-Y.L. conceived the study design, developed the analysis tool, analyzed the TWB data, and wrote the manuscript. W.-Y.L. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. P.-H.K., S.-J.T., Y.-L.L., and A.C.Y. contributed to the writing of the manuscript. All authors provided the TWB data and reviewed the manuscript.

References


9. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 2007;81:559-575


11. Calabro M, Drago A, Sidoti A, Serretti A, Crisafulli C: Genes involved in pruning and inflammation are enriched in a large mega-sample of patients affected by Schizophrenia and Bipolar Disorder and controls. Psychiatry Res 2015;228:945-949

12. WTCCC: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-678


24. Keller MC: Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biol Psychiatry 2014;75:18-24
28. Brun T, Franklin I, St-Onge L, Biasone-Louber A, Schoenle EJ, Wollheim CB, Gauthier BR: The diabetes-linked transcription factor PAX4 promotes beta-cell proliferation and survival in rat and...


### Tables

<table>
<thead>
<tr>
<th>(Discovery cohort / Replication cohort)</th>
<th>Overall</th>
<th>Smokers</th>
<th>Non-smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total, n</td>
<td>25,460 / 58,774</td>
<td>3,005 / 5,173</td>
<td>22,455 / 53,601</td>
</tr>
<tr>
<td>Males, n (%)</td>
<td>12,800 (50.3) / 19,017 (32.4)</td>
<td>2,647 (88.1) / 4,002 (77.4)</td>
<td>10,153 (45.2) / 15,015 (28.0)</td>
</tr>
<tr>
<td>Age (years), mean (s.d.)</td>
<td>48.9 (11.1) / 50.4 (10.6)</td>
<td>46.4 (9.9) / 47.6 (10.3)</td>
<td>49.2 (11.2) / 50.7 (10.6)</td>
</tr>
<tr>
<td>BMI (kg/m(^2)), mean (s.d.)</td>
<td>24.3 (3.7) / 24.2 (3.8)</td>
<td>25.4 (3.9) / 25.3 (4.1)</td>
<td>24.2 (3.7) / 24.1 (3.7)</td>
</tr>
<tr>
<td>Drinking, n (%)</td>
<td>1,799 (7.1) / 3,229 (5.5)</td>
<td>728 (24.2) / 1,218 (23.5)</td>
<td>1,071 (4.8) / 2,011 (3.8)</td>
</tr>
<tr>
<td>Regular exercise, n (%)</td>
<td>10,423 (40.9) / 24,325 (41.4)</td>
<td>858 (28.6) / 1,473 (28.5)</td>
<td>9,565 (42.6) / 22,852 (42.6)</td>
</tr>
<tr>
<td>Educational attainment, mean (s.d.)</td>
<td>5.5 (1.0) / 5.5 (1.0)</td>
<td>5.4 (0.9) / 5.4 (0.9)</td>
<td>5.5 (1.0) / 5.5 (1.0)</td>
</tr>
<tr>
<td>Fasting glucose (mg/dL), mean (s.d.)</td>
<td>96.3 (21.0) / 95.8 (20.7)</td>
<td>99.4 (26.2) / 99.3 (27.6)</td>
<td>95.9 (20.1) / 95.5 (19.9)</td>
</tr>
<tr>
<td>HbA1c (%), mean (s.d.)</td>
<td>5.73 (0.80) / 5.80 (0.81)</td>
<td>5.80 (0.93) / 5.91 (1.07)</td>
<td>5.72 (0.78) / 5.79 (0.78)</td>
</tr>
<tr>
<td>HbA1c (mmol/mol), mean (s.d.)</td>
<td>39 (8.7) / 40 (8.9)</td>
<td>40 (10.2) / 41 (11.7)</td>
<td>39 (8.5) / 40 (8.5)</td>
</tr>
<tr>
<td>Fasting glucose &gt; 126 mg/dL, n (%)</td>
<td>1,114 (4.4) / 2,478 (4.2)</td>
<td>201 (6.7) / 344 (6.6)</td>
<td>913 (4.1) / 2,134 (4.0)</td>
</tr>
<tr>
<td>HbA1c &gt; 6.5 % (48 mmol/mol), n (%)</td>
<td>1,752 (6.9) / 4,334 (7.4)</td>
<td>266 (8.9) / 530 (10.2)</td>
<td>1,486 (6.6) / 3,804 (7.1)</td>
</tr>
<tr>
<td>Subjects with physician-diagnosed diabetes, n (%)</td>
<td>1,260 (4.9) / 3,099 (5.3)</td>
<td>183 (6.1) / 330 (6.4)</td>
<td>1,077 (4.8) / 2,769 (5.2)</td>
</tr>
<tr>
<td>Subjects with diabetes, n (%)</td>
<td>2,207 (8.7) / 5,308 (9.0)</td>
<td>335 (11.1) / 635 (12.3)</td>
<td>1,872 (8.3) / 4,673 (8.7)</td>
</tr>
</tbody>
</table>
Table 1. Basic characteristics stratified by “active smoking status”

1. Subjects with diabetes included those with physician-diagnosed diabetes, or those having FG > 126 mg/dL or HbA1c > 6.5 % (48 mmol/mol) according to the TWB test results.
Table 2. Results of regression model (1) (prior to GRS analysis)

1. Natural log transformed fasting glucose (or HbA1c) was regressed on sex, age, body mass index, active smoking status, drinking status, regular exercise, educational attainment, and the first 10 PCs. To save space, we here omit the results of the 10 PCs.
2. Because fasting glucose (or HbA1c) was natural log transformed, $\left(\exp\left(\hat{\beta}_{Sex}\right) - 1\right) \times 100\%$ is shown to represent the change in
fasting glucose (or HbA1c) between females and males, while adjusting for age, body mass index, active smoking status, drinking status, regular exercise, educational attainment, and the first 10 PCs. For example, women on average have lower fasting glucose than men by 3.79% ($P_{Sex-FG} = 8.4 \times 10^{-79}$).

3. A $p$-value of 0 means that the test is extremely significant.

4. For continuous traits, R-square is the proportion of variance in natural log transformed fasting glucose (or HbA1c) that can be explained by sex, age, body mass index, active smoking status, drinking status, regular exercise, educational attainment, and the first 10 PCs. For the dichotomous trait (diabetes status), we present pseudo R-square, defined as one minus the ratio of the log likelihood with intercepts only, and the log likelihood with all predictors.
<table>
<thead>
<tr>
<th>(Discovery cohort / Replication cohort) (D/R)</th>
<th>Fasting glucose (mg/dL)</th>
<th>HbA1c (%)</th>
<th>Diabetes (dichotomous trait)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking measurements in regression model (1)</td>
<td>Percent change (%)</td>
<td>p-value</td>
<td>Percent change (%)</td>
</tr>
<tr>
<td>Active smoking status (1: yes vs. 0: no)</td>
<td>0.81 (0.88)</td>
<td>0.009 (1.3E-4)</td>
<td>0.83 (1.32)</td>
</tr>
<tr>
<td>The number of pack-years (mean±sd, D: 19.6±17.0; R: 19.5±17.6)</td>
<td>0.06 (0.05)</td>
<td>2.3E-7 (8.8E-9)</td>
<td>0.06 (0.06)</td>
</tr>
<tr>
<td>Years as a smoker (mean±sd, D: 25.6±10.1; R: 26.3±10.8)</td>
<td>0.04 (0.04)</td>
<td>8.6E-4 (3.2E-6)</td>
<td>0.04 (0.05)</td>
</tr>
<tr>
<td>Packs smoked per day (mean±sd, D: 0.72±0.51; R: 0.69±0.52)</td>
<td>1.70 (1.33)</td>
<td>4.9E-7 (3.6E-7)</td>
<td>1.63^2 (1.70)</td>
</tr>
<tr>
<td>Hours as a passive smoker per week (mean±sd, D: 5.6±11.0; R: 5.4±10.6)</td>
<td>0.03 (0.04)</td>
<td>0.18 (0.02)</td>
<td>0.01 (0.03)</td>
</tr>
</tbody>
</table>

**Table 3.** Results of regression model (1) when replacing "active smoking status" with other smoking measurements (prior to GRS analysis)

1. A total of 15 trait-smoking associations were tested here. A trait-smoking association is significant if p < \( \frac{0.05}{15} = 0.0033 \).
2. Because HbA1c (or FG) was natural log transformed, \(\exp(\beta_{\text{packs}}) - 1\) × 100% = 1.63% \((P_{\text{packs}-\text{HbA1c}} = 3.4 \times 10^{-11})\) is the change in HbA1c that is associated with a pack increase of active cigarette smoking per day, while adjusting for sex, age, body mass index, drinking status, regular exercise, educational attainment, and the first 10 PCs. The detailed result of regression model (1) can be found in Supplemental Table S3.
<table>
<thead>
<tr>
<th>Explanatory variables in regression model (4)</th>
<th>Fasting glucose (mg/dL)</th>
<th>HbA1c (%)</th>
<th>Diabetes (dichotomous trait)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent change (%)</td>
<td>p-value</td>
<td>Percent change (%)</td>
</tr>
<tr>
<td>Genetic risk score (GRS) (z-score standardized)</td>
<td>1.72</td>
<td>2.2E-170</td>
<td>1.30</td>
</tr>
<tr>
<td>Sex (1: female vs. 0: male)</td>
<td>-3.32</td>
<td>1.3E-121</td>
<td>-0.81</td>
</tr>
<tr>
<td>Age (in years, continuous variable)</td>
<td>0.36</td>
<td>0²</td>
<td>0.29</td>
</tr>
<tr>
<td>Body mass index (in kg/m², continuous variable)</td>
<td>0.85</td>
<td>0²</td>
<td>0.74</td>
</tr>
<tr>
<td>Active smoking status (1: yes vs. 0: no)</td>
<td>0.60</td>
<td>0.08</td>
<td>0.88</td>
</tr>
<tr>
<td>GRS × Active smoking status (continuous variable)</td>
<td>1.05³</td>
<td>2.8E-5</td>
<td>0.51</td>
</tr>
<tr>
<td>Drinking status (1: yes vs. 0: no)</td>
<td>2.09</td>
<td>2.4E-11</td>
<td>-0.93</td>
</tr>
<tr>
<td>Regular exercise (1: yes vs. 0: no)</td>
<td>-0.92</td>
<td>1.1E-12</td>
<td>-1.02</td>
</tr>
<tr>
<td>Educational attainment (a value ranging from 1 to 7)</td>
<td>-0.50</td>
<td>8.7E-14</td>
<td>-0.21</td>
</tr>
<tr>
<td>R-square</td>
<td>14.3 %</td>
<td>15.4 %</td>
<td>14.0 %</td>
</tr>
</tbody>
</table>

Table 4. Results of regression model (4) when the smoking measurement is “active smoking status” (including GRS and GRS-smoking interaction)
1. TWB1 was used to find trait-associated SNPs and TWB2 was used to test for interactions. Natural log transformed fasting glucose (or HbA1c), or diabetes status, was regressed by model (4). To save space, we here omit the results of the 10 PCs, $GRS \times Covariates$, and $Smoking \times Covariates$.

2. A $p$-value of 0 means that the test is extremely significant.

3. Because fasting glucose (or HbA1c) was natural log transformed, $\left(\exp(\hat{\phi}_{INT}) - 1\right) \times 100\% = 1.05\%$ represents that each 1 sd increase in GRS is associated with a 1.05% higher fasting glucose in (active) smokers than in non-smokers ($P_{INT} = 2.8 \times 10^{-5}$).

4. For continuous traits, R-square is the proportion of variance in natural log transformed fasting glucose (or HbA1c) that can be explained by the explanatory variables shown in model (4). For the dichotomous trait (diabetes status), we present pseudo R-square, defined as one minus the ratio of the log likelihood with intercepts only, and the log likelihood with all predictors.
<table>
<thead>
<tr>
<th>Smoking measurements in regression model (4)</th>
<th>Fasting glucose (mg/dL)</th>
<th>HbA1c (%)</th>
<th>Diabetes (dichotomous trait)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent change (%)</td>
<td>Percent change (%)</td>
<td>*</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>GRS × Active smoking status</td>
<td>1.05</td>
<td>2.8E-5 *</td>
<td>1.16</td>
</tr>
<tr>
<td>GRS × The number of pack-years</td>
<td>0.06</td>
<td>1.6E-7 *</td>
<td>1.002</td>
</tr>
<tr>
<td>GRS × Years as a smoker</td>
<td>0.04</td>
<td>8.5E-6 *</td>
<td>1.004</td>
</tr>
<tr>
<td>GRS × Packs smoked per day</td>
<td>1.68</td>
<td>1.9E-7 *</td>
<td>1.09</td>
</tr>
<tr>
<td>GRS × Hours as a passive smoker per week</td>
<td>0.02</td>
<td>0.25</td>
<td>0.996</td>
</tr>
</tbody>
</table>

Table 5. Results of regression model (4) when replacing “active smoking status” with the 4 continuous smoking measurements (including GRS and GRS-smoking interaction)

* TWB1 was used to find trait-associated SNPs and TWB2 was used to test for interactions. A total of 15 GRS-smoking interactions were tested here. A GRS-smoking interaction is significant if $P_{INT} < \frac{0.05}{15} = 0.0033$. 
Figure legends

Figure 1  Average of FG/HbA1c and the prevalence of diabetes stratified by smoking status and the quintiles of the FG/HbA1c/diabetes genetic risk score

The solid lines are for smokers, whereas the dotted lines are for non-smokers. The black lines depict predicted mean FG/HbA1c or predicted prevalence of diabetes based on model (4). Only subjects without any missing in covariates can be predicted. The blue lines mark crude mean FG/HbA1c or crude prevalence of diabetes, without adjusting for any covariates. The number shown around each point represents the sample size of that category.