

Genome-wide Polygenic Risk Score, Cardiometabolic risk Factors and Type 2 Diabetes Mellitus in The Chinese Population

Ninghao Huang¹, Wendi Xiao¹, Jun Lv^{1,2}, Canqing Yu^{1,2}, Yu Guo^{3,4}, Pei Pei², Ling Yang^{5,6}, Iona Y. Millwood^{5,6}, Robin G. Walters^{5,6}, Yiping Chen^{5,6}, Huaidong Du^{5,6}, Simon Gilbert⁶, Daniel Avery⁶, Tingting Ou⁷, Junshi Chen⁸, Zhengming Chen⁶, Tao Huang^{1,9*}, Liming Li^{1,2*}, on behalf of the China Kadoorie Biobank Collaborative Group[†]

1. Department of Epidemiology & Biostatistics, School of Public Health, Peking University, Beijing 100191, China;
2. Peking University Centre for Public Health and Epidemic Preparedness & Response, Beijing 100191, China;
3. Fuwai Hospital Chinese Academy of Medical Sciences, Beijing 100730, China
4. National Center for Cardiovascular Diseases, Beijing 1002300, China
5. Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, United Kingdom
6. Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, United Kingdom
7. NCDs Prevention and Control Department, Hainan CDC, Hainan 570203, China
8. China National Centre for Food Safety Risk Assessment, Beijing, China
9. Centre for Intelligent Public Health, Academy for Artificial Intelligence, Peking University, Beijing 100191, China

[†]The members of steering committee and collaborative group are listed in the Acknowledgements.

***Corresponding authors:**

Tao Huang, PhD

Department of Epidemiology and Biostatistics, Peking University

38 Xueyuan Road, Beijing 100191, China

Email: huang.tao@pku.edu.cn

Liming Li, MD, MPH

Department of Epidemiology and Biostatistics, Peking University

38 Xueyuan Road, Beijing 100191, China

Phone: 86-10-82801528

Email: lmlee@vip.163.com

Abstract

Objective: Type 2 diabetes (T2D) is caused by both genetic and cardiometabolic risk factors. However, the magnitude of T2D's genetic predisposition and the etiological structure of genetic and cardiometabolic risk factors for T2D in the Chinese population remains largely unknown.

Methods: We included 93,488 participants from China Kadoorie Biobank (CKB). Multiple polygenic risk scores (PRS) were calculated and predicted the cumulative T2D risk between the ages of 40 to 80 years based on the genetic risk. The genotypic variance explained (R^2) was used to assess the power of the predictive risk model of T2D. A common cardiometabolic risk score (CRS) using smoking, alcohol consumption, physical activity, diet, obesity, blood pressure, and blood lipids was constructed to investigate the effects of cardiometabolic risk factors on T2D. Furthermore, an equation based on ideal PRS, CRS, and their interaction was established to explore the combined effects on T2D.

Results: We reached an ideally fitting PRS ($R^2 = 7.6\%$) based on multiple PRS calculation methods in Chinese populations. We found that participants aged over 80 got the highest genetic risk with an approximately 31% cumulative prevalence risk of T2D. The lowest PRS group can reduce the risk of T2D by nearly 20% at age 80, compared with participants in the highest PRS group. Per 1-point CRS increment can shrink the T2D risk by about 34% (adjusted hazard ratio: 0.66, 95% confident intervals (CI): 0.61-0.71). We also found an additive interaction between PRS and CRS (coefficient = 29%, 95%CI: 0.22-0.36, $P < 0.001$). The R^2 of the T2D predictive model could increase to 8.3% when the CRS and the interaction terms of PRS×CRS were considered. In the etiological composition of T2D, the ratio of genetic risk effect, cardiometabolic risk effect, and interaction between gene and cardiometabolic factor were 67:16:17.

Conclusions: This study identified an ideally fitting PRS for identifying and predicting the risk of T2D suitable for the Chinese population from multiple PRS calculation methods. We detected the quantified etiological structure of genes, cardiometabolic risk factors, and their interactions (67:16:17), which elucidated the critical effect of genetic factors. An improving lifestyle and physical condition can be helpful in preventing T2D, especially in those with a high genetic risk for T2D.

Keywords: Genetic predisposition; polygenic risk score; Type 2 diabetes; China Kadoorie Biobank

Research in context

What is already known about this subject?

- Type 2 diabetes (T2D) is a combination of genetic and environmental risk factors. The association of individual genetic risk factors or environmental factors with the onset of T2D mellitus has been described previously, while the combined effects from both genetic and environmental influences in the Chinese population remain poorly studied.

What is the key question?

- Which calculation methods of polygenic risk score (PRS) is a suitable estimate of T2D in the Chinese population?
- How is it related to T2D in conjunction with environmental factors?

What are the new findings?

- In this study, based on 93,488 Chinese, it was found that the best proportion of T2D was explained by the way PRS were calculated by single chromosome.
- Adherence to a good healthy lifestyle can significantly reduce the risk of T2D, even if the genetic risk is high.
- There is an interaction between genetic factors and environment. The ratio of genes to environment and their interactions is about 67:16:17.

How might this impact on clinical practice in the foreseeable future?

- This PRS calculation method works best in the Chinese population. Our research could be a milestone to identify disadvantaged people to address the escalating prevalence of T2D

Introduction

Type 2 diabetes (T2D) is one of the most widespread and prevalent metabolic diseases in the world ^[1]. The increasing prevalence of T2D creates a huge burden on individual physiological aspects and national economical aspect^[2]. More than 400 million people suffer from T2D, and the number will climb to 622 million by 2040 ^[3, 4]. Because of the high disease burden of T2D, it is extremely urgent to identify potential high T2D risk groups and achieve primary prevention as soon as possible to reduce quality-year loss associated with diabetes.

The risk of T2D is determined by both genetic risks and cardiometabolic risks rather than by a single factor ^[5]. Polygenic risk scores (PRS) based on genome-wide association studies have been proven to predict T2D ^[6-13]. However, the scale of genetic variants that construct PRS is in the millions of magnitudes, and there existed differences between varied methods ^[14]. There are two main methods to shrink SNPs of a large order of magnitude :(1) screening "priority" SNPs by Linkage Disequilibrium (LD) and p-value; (2) regularization or Bayesian methods are used to shrink the number of SNPs^[15]. While the approaches differ in many ways, such approaches are neither superior nor inferior to each other. In addition to genetic factors, a healthy lifestyle and good physiological indicators can delay or prevent diabetes and be used as predictors of T2D ^[16]. Cardiometabolic risk factors consisting of both lifestyle and metabolic risk factors were highly associated with T2D ^[17]. Finding the most suitable PRS for T2D and the combination of cardiometabolic risk factors is particularly workable for precision medicine.

Existing PRS models for T2D were mainly based on white western populations^[18-20], and the T2D risk prediction model established in China is regional and focuses on economically developed areas ^[21-23]. T2D prediction models that combine genetic and cardiometabolic risk factors to predict diabetes risk in Chinese populations remain unexplored.

The purpose of the present study is to explore the optimal genetic prediction approaches for T2D and combine cardiometabolic risk factors to establish a T2D prediction model suitable for the Chinese population. Our study may help identify the potential high-risk population of T2D and provide scientific evidence of personalized primary prevention for T2D.

Methods

Study Design, Settings and Participants

The China Kadoorie Biobank (CKB) study, which included more than 500 000 participants aged 30-79 years from 10 regions (Harbin, Qingdao, Suzhou, Liuzhou, Haikou, Henan, Gansu, Sichuan, Zhejiang and Hunan) across China, is a prospective cohort study. Detailed information on the CKB has been described in other places previously ^[24]. In short, all participants have asked to fill out the questionnaire, and physical measurements and blood samples were collected at the baseline from 2004 to 2008. Participants were followed up twice in 2008 and from 2013 to 2014, and all participants gave informed written consent. The

Ethical Review Committee of the Chinese Centre for Disease Control and Prevention (Beijing, China) and the Oxford Tropical Research Ethics Committee, the University of Oxford (UK) approved the study.

We included the participants with genetic data. Participants with cancer, missing BMI values, and diabetes diagnosed by physicians were excluded. We also excluded participants who were under insulin therapy and who had more than 4 standard deviations (SD) values based on random glucose and fasting glucose.

Ascertainment of Type 2 Diabetes

Participants' health status information was obtained from the national health insurance database and mortality surveillance data. T2D was recorded as prevalence (T2D diagnosed by a physician at baseline) and incidence (T2D diagnosed during follow-up). The Tenth Revisions of the International Classification of Disease (ICD-10) were used to define T2D. The T2D was composed of the ICD-10 "E11".

Construction of Polygenic Risk Score

The individual-level genotype data in CKB had been conducted by three phases of genotyping. The University of Oxford's Clinical Trial Service Unit and Epidemiological Studies Unit (Oxford, UK), Beijing Genomics Institute (Shenzhen, China) and Thermo Fisher Scientific Inc (CA, USA) cooperated in developing a custom-designed biological chip, which is used to detect genotypes. About 32K CKB participants used the 700K single nucleotide polymorphism (SNP) array to genotype in the first phase. Then the revised and renewed array contained about 803K SNPs and was used to genotype 69K participants in another two phases. The qualified genotypes were defined with call rate >0.98 , plate effect $p > 10^{-6}$, batch effect $p > 10^{-6}$, Hardy–Weinberg equilibrium deviations $p > 10^{-6}$, and minor allele frequency (MAF) difference from 1000 Genomes East Asian frequencies <0.2 . SHAPEIT was used to phase the qualified genotypes for each chromosome^[25], and the 1000 Genomes phase III was used to perform imputation between each 5-Mb interval by IMPUTE 4^[26].

The calculation procession of PRS required quality control in base data and target data, which were described previously^[15]. We used a meta-analysis of genome-wide association studies (GWAS), which identified more than 240 loci that are associated with T2D in the East Asian population without CKB, as the base data of PRS^[27]. We removed SNPs with low MAF (MAF $< 1\%$) and low imputation information score (INFO) (INFO < 0.8) to lessen the probability of false-positive. Mismatching, duplicated, and ambiguous SNPs were also excluded. Similar to the quality control of base data, target data, the individual-level genotype data in CKB, also went through diligent quality control. The detailed quality control of the target data and sample was shown in Figure 1.

We constructed PRS by using ten different PRS calculation methods. PRSice-2 is an efficient and scalable software program for automating and simplifying PRS analyses on large-scale data, and it provides empirical association P-values free from inflation due to overfitting^[28]. LDpred is a method that infers the

posterior mean effect size of each marker by using a prior on effect sizes and LD information from an external reference panel ^[29]. Lassosum can construct PRS using summary statistics and a reference panel in a penalized regression framework ^[30]. PLINK software uses the clump plus threshold (C+T) method similar to PRSice-2, except that PLINK cannot find the fitness P-value and clumping threshold by itself. In this paper, we used PLINK to calculate PRS for SNPs on different chromosomes, and PRS in different chromosomes finally combined.

We use the genetic risk score (GRS) to refer specifically to PRS filtered and manually calculated. We selected the significant signal SNPs with a P-value less than 5×10^{-8} to establish GRS (GRS1) according to the previous study (Table S1) ^[27]. In addition, we searched, screened, and clumped SNPs (GRS2) related to T2D from the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>), the samples of which were all or part of the East Asian population (Table S2-S3). The SNPs from the GWAS catalog were further selected using lasso regression (GRS3) and elastic net regression (GRS4), and the first 100 SNPs were included based on the absolute value of the effect value (Table S4). We constructed unweighted GRS and weighted GRS for each

GRS approach. The weighted GRS is calculated by the formula below: $GRS_{weighted} = \sum_{i=1}^m \hat{b}_i x_{ij}$, where m is the number of SNPs selected, \hat{b}_i is the per allele weight for the SNP, which is the SNPs' effect value (log odds ratio), x_{ij} is a count of the number (0, 1, or 2) of trait-associated alleles of SNP i in individual j. Each personal count of SNPs was imputed by mean.

Cardiometabolic Risk Factors

Cardiometabolic risk factors consist of both lifestyle and metabolism risk factors. All lifestyle risk factors (smoking, alcohol consumption, physical activity, and diet) and metabolic risk factors (obesity, blood pressure, and blood lipid) were collected by baseline questionnaires and physical measurements. The factors were dichotomized into healthy groups and unhealthy groups. Specifically, no current smoking was defined as the healthy group, and current smokers, former smokers with participants who had stopped smoking due to illness, were classified into the unhealthy group. Drinking ≤ 25 g of pure alcohol in men and ≤ 15 g in women per day was defined as the healthy alcohol drinking group, which is according to the Chinese dietary guidelines ^[31]. Participants with a median or higher level of physical activity in by sex-specific were classified into the healthy group. A healthy diet was defined based on food consumption frequency, with vegetables and fruits daily and red meat 1~6 days a week. For obesity factors, both body mass index (BMI; $18.5 \leq \text{BMI} < 24.0$ kg/m²) and waist circumference (WC; WC < 90 cm for men and WC < 85 cm for women) were considered as the healthy group. Healthy blood pressure was defined as untreated blood pressure, by systolic blood pressure <120 mmHg and diastolic blood pressure <80 mmHg. The healthy lipid group is defined as participants without taking lipid-lowering medication currently due to limitations of lipid in the baseline survey of CKB. Each healthy group was scored as 1, and the unhealthy

group was scored as 0. We constructed cardiometabolic risk scores (CRS) by summing each lifestyle and metabolic risk factor above.

Polygenic Risk Score and Cardiometabolic Risk Score Interaction (PRS×CRS) Analysis

To explore the best fitting model consisting of heredity, cardiometabolic risk factors, and their interaction, we constructed a linear regression model by adding PRS, CRS, and the interaction effect between PRS and CRS:

$$\text{Type 2 diabetes} = X \times \text{PRS}_{\text{T2D}} + Y \times \text{CRS}_{\text{T2D}} + Z \times \text{Interaction}_{\text{PRS} \times \text{CRS}} + \text{other covariates}$$

where X is the ratio of PRS_{T2D} ; Y is the ratio of the CRS; Z is the ratio of the interaction between PRS and CRS; other covariates included age, sex, education level, marital status, and family history of diabetes. The PRS_{T2D} is the best fitting model for various PRS approaches. The sum of X, Y, and Z is 100%.

Statistical Analysis

We used the genotypic variance explained, R-Square (R^2), and the area under the curve (AUC) to assess the PRS models. A higher value of R^2 and AUC means the fitter the models. The optimal PRS was divided into 5 groups by quintile. Cox proportional hazard ratio model was used to assess the association between PRS groups and T2D. Several adjusted models were built: model 1 was adjusted for sex, age, region, ten genetic principal components, and regional principal components; model 2 was adjusted for model 1 plus education levels, marital status, and family history of diabetes; model 3 additionally adjusted for blood pressure, lipid-lowering drug-taking, body mass index, waist circumference, smoking status, drinking status, diet, and physical activities. Hazard ratios (HR) and their 95% confidence intervals (CI) for each PRS group were evaluated compared with those in the first groups. The population-attributable risk (PAR %) was also calculated to quantify the proportion of T2D attributed to the first quintile PRS group (PRS Q1). We drew a restricted cubic spline to show the nonlinear associations between PRS and T2D. The cumulative incidence of T2D during a 12-year follow-up was estimated by the Cox model in different PRS groups. The predicted incidence of T2D from 40 to 80 (5-year a group) was estimated in different PRS groups, respectively. The difference in predicted incidence between different lowest and highest PRS groups was also tested. We also estimated the multiplicative interactions between PRS and CRS, and calculated the relative excess risk due to interaction (RERI) and attributable proportions (AP) for PRS alone, CRS alone, and their interaction effects on an additive scale.

Several sensitivity analyses were conducted. Firstly, the sex-specific models were constructed respectively in the process of PRS construction to test the robustness of PRS. Besides, we randomly split the overall samples into several subsets as the training set (10%, 20%, 30%, and 50%). Each subset has a similar incident of T2D ($P = 0.63$). We also used a larger sample of GWAS (including the CKB population) to perform the PRS calculation. We further conducted a sensitivity analysis for PRS construction by considering the prevalence at baseline and incidence after follow-up. Subgroups analyses among covariates were performed to explore the potential association between PRS and T2D and to examine its interaction

effect (P for interaction). A two-sided p-value of less than 0.05 was considered to indicate a significant difference.

Patient Involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

A total of 93,488 participants were finally enrolled, including 2,485 T2D cases (Figure 1). Participants with T2D tended to be older (56.4 vs. 53.2), live in urban (61.3% vs. 57.5%), have higher incomes (≥ 20 k yuan, 48.6% vs. 38.5%), BMI (25.6 kg/m² vs. 23.5 kg/m²), and systolic blood pressure (141.8mmHg vs. 132.6mmHg), and have more family history of diabetes (8.9% vs. 6.0%) (Table 1).

Performance of PRS

We executed the strict quality control process and finally obtained 8,802,880 genetic variants (Table S1-S4). The performance of PRS approaches is shown in Figure 2A. With the increase in the sample size, the time and memory required for PRS calculation sharply increased. LDpred taken up the most time and memory, by more than 3k minutes and 20Gb memory (Table S5). PRSice-2 performed stably in different sample sizes and occupied few resources. For the R^2 of PRS models, PRS using PLINK for chromosome division calculation had the most fitting results, explaining by $R^2 = 7.6\%$ in different sample subsets. Lassosum got the second fitness PRS by $R^2 = 2.8\%$ (Table S6-S7). The PRS results obtained by LDpred and PRSice2 were similar ($R^2 \sim 1.5$). There was no particularly significant difference between weighted and unweighted GRS. The loci selected by elastic net and lasso were not effective. From the perspective of performance assessed by AUC, the fitting results of various methods were consistent with R^2 (Table S8).

In the sensitivity analysis, we carried out sex-specific PRS. The R^2 of PRSice-2 in PRS sex-specific calculation was higher than that of the whole population. In contrast, the R^2 of PLINK calculation by gender was not different from that of the entire population (Table S6). When using GWAS, including the CKB population, the calculated R^2 of PRS was significantly improved, increasing by at least two times (Table S9). To sum up, we elected the PRS model with the highest discriminatory performance, which was calculated by meta-chromosome, as the optimal PRS model to join the subsequent analysis (the following PRS refers to the optimal PRS above without explanation).

Genetic risk of T2D

The PRS was divided into five groups (Q1 to Q5) (Table S10). Participants in the highest PRS group had a large proportion living in the urban (59.2%), tended to be older (53.7, SD: 0.08), and had a family history

of diabetes (8.2%) compared with participants who were in the lowest PRS group (Table S11). Per 1-PRS group increment was associated with a 209% higher risk of T2D (adjusted HR (aHR), 3.09; 95% CI: 2.93-3.25), and the HR was 35.30 (95% CI: 27.26-45.70) for the highest PRS group when compared with the PRS lowest group. Table 2 presents the PAR% for each PRS group. Compared with the lowest PRS group, the highest PRS group was estimated to explain 88% of the population's risk of developing T2D. We found a positive linear association detected by restricted cubic spline regression, which is shown in Figure 3, with $P < 0.001$. The 10-year cumulative incidence in the highest PRS was steeper than in other PRS groups, with a 12% cumulative rate for T2D in 12 years of follow-up. We predicted the probability of T2D in people aged 40-80 (Table 3) based on the PRS. The prevalence of T2D at age 40 is less than 0.5%, regardless of the PRS group. With the increment of age, the risk of T2D increased gradually, The risk of T2D rocketed to 0.43% in the lowest PRS group, and 23.91% in the highest PRS group with participants at 80.

The baseline characteristics of CRS were tested (Table S12). Per 1-point CRS increment can shrink the T2D risk by about 34% (aHR: 0.66, 95% CI: 0.61-0.71) (Table S14). The healthier the CRS, the lower the risk of T2D, with 86% lower risks of T2D (aHR: 0.14, 95% CI: 0.09-0.23) in the healthier CRS group (7-8) when compared with the unhealthier CRS group (0-2). We computed the difference in predicted T2D prevalence between the healthiest and unhealthiest CRS groups (supplementary figure 2). People in a healthy CRS group had a maximum reduction of about 20% in the prevalence of T2D compared with those in an unhealthy CRS group (Table S15). Even in the healthy CRS group, participants in the highest PRS group still increased the risk of T2D by about 18%. Participants in the highest PRS group who stayed in an unhealthy CRS group had a 23.22% risk of T2D at age 60, and these figures will rise to 34.36% at age 70 and 43.59% at age 80.

Interaction Analysis of Gene and Cardiometabolic Risk Factors

Table 4 shows the subgroup association between cardiometabolic risk factors and T2D. Generally, compared with the reference group, the highest CRS has the lowest risk of T2D (aHR: 0.13, 95% CI: 0.09-0.18 for $\text{CRS} \geq 7$). Participants in the healthy group of obesity and blood pressure had also observed a significant association with T2D (aHR: 0.40, 95% CI: 0.36-0.43 for obesity, aHR: 0.61, 95% CI: 0.54-0.70 for blood pressure). Similar results were found when adjusting for PRS. The relative excess risk, which means the difference between the combined effect of PRS and CRS, was 1.07 (95% CI: 0.35-1.78, $P < 0.001$) (Table S16). We also found that the attributable proportion due to additive interaction was significant (coefficient = 29%, 95% CI: 0.22-0.36, $P < 0.001$). In the linear mixed model of gene-environment, we obtained the R^2 of 7.9% when the etiological composition of the PRS to the CRS was 77:33, without considering the interaction term of $\text{PRS} \times \text{CRS}$. We found a significant additive interaction between PRS and CRS ($P \leq 0.001$). When the interaction term of $\text{PRS} \times \text{CRS}$ was considered, the estimated magnitude of genotypic explanation was slightly climbed ($R^2 = 8.3\%$, compared with 7.6%), and the etiological composition of T2D involved PRS, CRS, and their interaction, is 67:16:17.

Discussion

We identified the ideally fitting PRS for T2D in the Chinese population using ten different PRS calculation approaches with an increasing R^2 of 7.6%. Per 1-group of PRS increment was associated with about three times higher risk for T2D. Cumulative prevalence predictions were made for every 5-year-old group from age 40 onwards, and the highest genetic risk group was associated with an approximately 31% risk of T2D at age 80. The highest CRS can reduce the risk of T2D by nearly 15% at age 80, compared with participants in the lowest CRS group. A significant additive interaction between PRS and CRS was found. The ratio of the etiological composition of T2D involved PRS, CRS, and their interaction was 67:16:17, with genetic risk contributing more than half of the T2D prevalence. Furthermore, the R^2 of T2D prevalence model could increase to 8.3% when the CRS and the gene-environment interaction were considered.

PRS is frequently employed to assess the genetic risk of diseases because it can provide guidelines on how high-risk groups should be identified by the government^[32]. Prior studies have assessed the predictive capability of PRS in T2D, but those studies have primarily relied on populations of European ancestry and do not have evidence of Chinese ancestry^[33-35]. Our study explored the ideally fitting PRS of T2D suitable for the Chinese population among several approaches and provided evidence of the population with high T2D risk, and thus can provide individuals with individualized primary preventive strategies. The PRS can be viewed as a sum of many small SNP effects, which are derived from several GWAS studies, so results may appear to be different for diverse people, diverse programs, or diverse diseases of calculation. As expected, the performance of the four PRS software was in line with previous studies, with LDpred displaying the highest time and memory consumption, and PRSice-2 exhibiting a low time and memory consumption^[28]. PRS calculations were made using different screening methods, and the results obtained were generally similar. Previous studies that used the UK Biobank genotype data and PLINK and LDpred to compute PRS have yielded a high degree of differentiation (AUC=0.709) for participants with European ancestry^[33]. It also achieved a 9.2% R^2 as a part of a similar PRS model comparison study, which tested for differences in psychiatric disorders using dozens of PRS-built methods^[36]. The comparison of PRS risk prediction methods shows that, except for PRS calculated by chromosome, there is not much difference between PRS calculation methods. Based on the fact that we divided the chromosomes, we calculated the best PRS, and each chromosome had the best threshold value, so the result was the most accurate^[37].

We observe that PRS can guide primary prevention, namely, assessing genetic risk can be used to determine if a population is at risk. In light of recent advances in genomics and the decreased cost of gene sequencing in high-risk populations, PRS may be helpful as a way to screen for the risk of developing T2D. Nevertheless, it is far from sufficient to define the disease risk of T2D solely by genetic factors. Indeed, lifestyle, metabolic, and other factors need to be taken to predict T2D. There is well-established evidence that lifestyle and cardiometabolic modifications can increase the risk of β cell damage in genetically at-risk individuals, leading to T2D^[38]. A variety of prevention and treatment guidelines for

T2D recommend lifestyle and metabolic adjustments to complement genetic risk reduction in cases where genetic factors cannot be corrected ^[39]. Researchers from He Y et al. have developed the polyexposure score for non-genetic exposure and lifestyle in the UK biobank and have added PRS to predict the clinical risk of T2D better and better differentiate the risk of disease ^[33]. We developed a CRS that can be generalized to general screening and does not require any other invasive or expensive medical procedures in addition to blood sampling (to test lipid metabolites and to conduct genetic sequencing), based on questionnaires and data collected from standard physical examinations. The findings of this study provide a T2D cumulative incidence prediction model for several genetic and cardiometabolic risk groups between 40 and 80 years of age, which can guide individualized strategies for interventions in clinical practice. We recommend, in general, that those in the top 20% of genetic risk groups, regardless of their lifestyle, be screened for T2D and given guidance regarding lifestyle interventions that may be necessary. Nearly 15% of T2D risk reduction may be achieved if these populations had a healthy lifestyle and metabolic profile.

Additionally, the additive interaction effect between genes and the cardiometabolic risk factors was examined. Approximately 28% of T2D events are caused by additive interactions between PRS and CRS, indicating that improving lifestyles and physical conditions are critical in preventing T2D, especially for those with a high genetic risk for T2D. Importantly, the interaction between cardiometabolic risk factors and T2D cases played a significant role in explaining the proportion of T2D patients. Our research revealed the composition of T2D etiology, pointing out that genetic risk contributes to more than half of T2D prevalence. There existed evidence from multiple twin studies that the contributions of T2D heritability, a measure of the contribution of genetic factors to disease, is more than 50% ^[40, 41]. Similarly, studies involving the Chinese twin cohort have demonstrated about 56% heritability of T2D, which is in line with the present study ^[42]. However, there remained unmeasurable and unexplainable confounders. We still yield only about 8.3% of R^2 even by combining the effects of heredity and cardiometabolic factors. Hence, it is imperative to include other omics data available to improve the rate of etiological contribution to T2D and better understand the composition of etiological causes for T2D in future studies.

Currently, this is the first study to construct and evaluate multiple PRS in a Chinese population using GWAS summary statistics and individual data, and to explore the etiology of T2D in conjunction with cardiometabolic risk factors. Our study still has many limitations. Firstly, we used a unified quality control standard for base data and target data in PRS construction, and did not adjust the parameters during the quality control phase. Though there may be better quality control parameters to increase the fitness of PRS, unified quality control can increase the comparability of PRS. Second, we did not include all possible risk factors for T2D because we used limited lifestyle and metabolic factors because of the restricted baseline questionnaires. In addition, we only considered the linear model when fitting the model and did not include the higher-dimensional model. It is expected that future studies will consider the possible risk factors for T2D, which will increase the proportion of explanations for its etiology.

Conclusion

In summary, we obtained ideally fitting PRS of T2D suitable for the Chinese population from multiple PRS calculation methods. The findings of the quantified etiological structure of genes, cardiometabolic risk factors, and their interactions (67:16:17) were detected and elucidated the role and underlying effect of genetic factors. Strategies for improving lifestyles and physical conditions can be helpful in preventing T2D, especially in those with a high genetic risk for T2D.

Acknowledgments

The most important acknowledgment is to the participants in the study and the members of the survey teams in each of the 10 regional centres, as well as to the project development and management teams based in Beijing, Oxford and the 10 regional centres.

Members of the China Kadoorie Biobank collaborative group

International Steering Committee: Junshi Chen, Zhengming Chen (PI), Robert Clarke, Rory Collins, Yu Guo, Liming Li (PI), Jun Lv, Richard Peto, Robin Walters. **International Co-ordinating Centre, Oxford:** Daniel Avery, Derrick Bennett, Ruth Boxall, Sue Burgess, Ka Hung Chan, Yumei Chang, Yiping Chen, Zhengming Chen, Johnathan Clarke; Robert Clarke, Huaidong Du, Ahmed Edris Mohamed, Zammy Fairhurst-Hunter, Hannah Fry, Simon Gilbert, Alex Hacker, Mike Hill, Michael Holmes, Pek Kei Im, Andri Iona, Maria Kakkoura, Christiana Kartsonaki, Rene Kerosi, Kuang Lin, Mohsen Mazidi, Iona Millwood, Sam Morris, Qunhua Nie, Alfred Pozarickij, Paul Ryder, Saredo Said, Sam Sansome, Dan Schmidt, Paul Sherliker, Rajani Sohoni, Becky Stevens, Iain Turnbull, Robin Walters, Lin Wang, Neil Wright, Ling Yang, Xiaoming Yang, Pang Yao.

National Co-ordinating Centre, Beijing: Yu Guo, Xiao Han, Can Hou, Jun Lv, Pei Pei, Chao Liu, Canqing Yu, Qingmei Xia. **10 Regional Co-ordinating Centres:** **Qingdao CDC:** Zengchang Pang, Ruqin Gao, Shanpeng Li, Haiping Duan, Shaojie Wang, Yongmei Liu, Ranran Du, Yajing Zang, Liang Cheng, Xiaocao Tian, Hua Zhang, Yaoming Zhai, Feng Ning, Xiaohui Sun, Feifei Li. **Licang CDC:** Silu Lv, Junzheng Wang, Wei Hou. **Heilongjiang Provincial CDC:** Wei Sun, Shichun Yan, Xiaoming Cui. **Nangang CDC:** Chi Wang, Zhenyuan Wu, Yanjie Li, Quan Kang. **Hainan Provincial CDC:** Huiming Luo, Tingting Ou. **Meilan CDC:** Xiangyang Zheng, Zhendong Guo, Shukuan Wu, Yilei Li, Huimei Li. **Jiangsu Provincial CDC:** Ming Wu, Yonglin Zhou, Jinyi Zhou, Ran Tao, Jie Yang, Jian Su. **Suzhou CDC:** Fang Liu, Jun Zhang, Yihe Hu, Yan Lu, Liangcai Ma, Aiyu Tang, Shuo Zhang, Jianrong Jin, Jingchao Liu. **Guangxi Provincial CDC:** Mei Lin, Zhenzhen Lu. **Liuzhou CDC:** Lifang Zhou, Changping Xie, Jian Lan, Tingping Zhu, Yun Liu, Liuping Wei, Liyuan Zhou, Ningyu Chen, Yulu Qin, Sisi Wang. **Sichuan Provincial CDC:** Xianping Wu, Ningmei Zhang, Xiaofang Chen, Xiaoyu Chang. **Pengzhou CDC:** Mingqiang Yuan, Xia Wu, Xiaofang Chen, Wei Jiang, Jiaqiu Liu, Qiang Sun. **Gansu Provincial CDC:** Faqing Chen, Xiaolan Ren, Caixia Dong. **Maiji CDC:** Hui Zhang, Enke Mao, Xiaoping Wang, Tao Wang, Xi zhang. **Henan Provincial CDC:** Kai Kang, Shixian Feng, Huizi Tian, Lei Fan. **Huixian CDC:** XiaoLin Li, Huarong Sun, Pan He, Xukui Zhang. **Zhejiang Provincial CDC:** Min Yu,

Ruying Hu, Hao Wang. **Tongxiang CDC:** Xiaoyi Zhang, Yuan Cao, Kaixu Xie, Lingli Chen, Dun Shen. **Hunan Provincial CDC:** Xiaojun Li, Donghui Jin, Li Yin, Huilin Liu, Zhongxi Fu. **Liuyang CDC:** Xin Xu, Hao Zhang, Jianwei Chen, Yuan Peng, Libo Zhang, Chan Qu.

Funding

This work was supported by the National Natural Science Foundation of China (82192901, 82192904, 82192900). The CKB baseline survey and the first re-survey were supported by a grant from the Kadoorie Charitable Foundation in Hong Kong. The long-term follow-up is supported by grants from the UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), grants (2016YFC0900500) from the National Key R&D Program of China, National Natural Science Foundation of China (91846303, 81390540), and Chinese Ministry of Science and Technology (2011BAI09B01). The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, or the decision to submit the article for publication.

Disclaimer

The sponsors had no role in the design, analysis, interpretation or drafting of this manuscript.

Conflict of interest

None declared.

Data sharing

The access policy and procedures are available at www.ckbiobank.org.

or,

Details of how to access China Kadoorie Biobank data and details of the data release schedule are available from www.ckbiobank.org/site/Data+Access.

Reference

1. Berg, D.D., et al., *A Biomarker-Based Score for Risk of Hospitalization for Heart Failure in Patients With Diabetes*. *Diabetes Care*, 2021. **44**(11): p. 2573-2581.
2. Wilkinson, L., et al., *Development and validation of a model for predicting incident type 2 diabetes using quantitative clinical data and a Bayesian logistic model: A nationwide cohort and modeling study*. *PLoS Med*, 2020. **17**(8): p. e1003232.
3. Mullee, A., et al., *Association Between Soft Drink Consumption and Mortality in 10 European Countries*. *JAMA Intern Med*, 2019. **179**(11): p. 1479-1490.
4. Umpierrez, G.E. and F.J. Pasquel, *Management of Inpatient Hyperglycemia and Diabetes in Older Adults*. *Diabetes Care*, 2017. **40**(4): p. 509-517.
5. Singer, R.A., et al., *The Long Noncoding RNA Paupar Modulates PAX6 Regulatory Activities to Promote Alpha Cell Development and Function*. *Cell Metab*, 2019. **30**(6): p. 1091-1106.e8.
6. Bao, W., 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): p. 1197-207.
7. de Miguel-Yanes, J.M., et al., *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): p. 121-5.
8. Imamura, M., 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): p. E1667-73.
9. Liao, L.N., et al., *Genetic risk score for risk prediction of diabetic nephropathy in Han Chinese type 2 diabetes patients*. *Sci Rep*, 2019. **9**(1): p. 19897.
10. Lyssenko, V., et al., *Clinical risk factors, DNA variants, and the development of type 2 diabetes*. *N Engl J Med*, 2008. **359**(21): p. 2220-32.
11. Meigs, J.B., et al., *Genotype score in addition to common risk factors for prediction of type 2 diabetes*. *N Engl J Med*, 2008. **359**(21): p. 2208-19.
12. Vassy, J.L., et al., *Genotype prediction of adult type 2 diabetes from adolescence in a multiracial population*. *Pediatrics*, 2012. **130**(5): p. e1235-42.
13. Vassy, J.L., et al., *A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study*. *Diabetologia*, 2012. **55**(10): p. 2604-2612.
14. Yang, J., et al., *Common SNPs explain a large proportion of the heritability for human height*. *Nat Genet*, 2010. **42**(7): p. 565-9.
15. Choi, S.W., T.S. Mak, and P.F. O'Reilly, *Tutorial: a guide to performing polygenic risk score analyses*. *Nat Protoc*, 2020. **15**(9): p. 2759-2772.
16. Zhang, Y., et al., *Combined lifestyle factors and risk of incident type 2 diabetes and prognosis among individuals with type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies*. *Diabetologia*, 2020. **63**(1): p. 21-33.
17. Han, Y., et al., *Lifestyle, cardiometabolic disease, and multimorbidity in a prospective Chinese study*. *Eur Heart J*, 2021. **42**(34): p. 3374-3384.
18. Lagani, V., et al., *Development and validation of risk assessment models for diabetes-related complications based on the DCCT/EDIC data*. *J Diabetes Complications*, 2015. **29**(4): p. 479-87.
19. Pollock, B.D., et al., *Utility of existing diabetes risk prediction tools for young black and white adults: Evidence from the Bogalusa Heart Study*. *J Diabetes Complications*, 2017. **31**(1): p. 86-93.

20. Vettoretti, M., et al., *Importance of Recalibrating Models for Type 2 Diabetes Onset Prediction: Application of the Diabetes Population Risk Tool on the Health and Retirement Study*. Annu Int Conf IEEE Eng Med Biol Soc, 2018. **2018**: p. 5358-5361.
21. Schaefer, K.K., et al., *Prediction of gestational diabetes mellitus in the Born in Guangzhou Cohort Study, China*. Int J Gynaecol Obstet, 2018. **143**(2): p. 164-171.
22. Wan, E.Y.F., et al., *Ten-year risk prediction models of complications and mortality of Chinese patients with diabetes mellitus in primary care in Hong Kong: a study protocol*. BMJ Open, 2018. **8**(10): p. e023070.
23. Xue, M., et al., *A nomogram model for screening the risk of diabetes in a large-scale Chinese population: an observational study from 345,718 participants*. Sci Rep, 2020. **10**(1): p. 11600.
24. Chen, Z., et al., *China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up*. Int J Epidemiol, 2011. **40**(6): p. 1652-66.
25. Delaneau, O., J. Marchini, and J.F. Zagury, *A linear complexity phasing method for thousands of genomes*. Nat Methods, 2011. **9**(2): p. 179-81.
26. Bycroft, C., et al., *The UK Biobank resource with deep phenotyping and genomic data*. Nature, 2018. **562**(7726): p. 203-209.
27. Spracklen, C.N., et al., *Identification of type 2 diabetes loci in 433,540 East Asian individuals*. Nature, 2020. **582**(7811): p. 240-245.
28. Choi, S.W. and P.F. O'Reilly, *PRSice-2: Polygenic Risk Score software for biobank-scale data*. Gigascience, 2019. **8**(7).
29. Vilhjálmsson, B.J., et al., *Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores*. Am J Hum Genet, 2015. **97**(4): p. 576-92.
30. Mak, T.S.H., et al., *Polygenic scores via penalized regression on summary statistics*. Genet Epidemiol, 2017. **41**(6): p. 469-480.
31. Yang, Y.X., et al., *New Chinese dietary guidelines: healthy eating patterns and food-based dietary recommendations*. Asia Pac J Clin Nutr, 2018. **27**(4): p. 908-913.
32. Torkamani, A., N.E. Wineinger, and E.J. Topol, *The personal and clinical utility of polygenic risk scores*. Nat Rev Genet, 2018. **19**(9): p. 581-590.
33. He, Y., et al., *Comparisons of Polyexposure, Polygenic, and Clinical Risk Scores in Risk Prediction of Type 2 Diabetes*. Diabetes Care, 2021. **44**(4): p. 935-943.
34. Li, J.H., et al., *A Polygenic Score for Type 2 Diabetes Risk Is Associated With Both the Acute and Sustained Response to Sulfonylureas*. Diabetes, 2021. **70**(1): p. 293-300.
35. Liu, W., et al., *An Improved Genome-Wide Polygenic Score Model for Predicting the Risk of Type 2 Diabetes*. Front Genet, 2021. **12**: p. 632385.
36. Ni, G., et al., *A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts*. Biol Psychiatry, 2021. **90**(9): p. 611-620.
37. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet, 2007. **81**(3): p. 559-75.
38. Kolb, H. and S. Martin, *Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes*. BMC Med, 2017. **15**(1): p. 131.
39. Forouhi, N.G., et al., *Dietary and nutritional approaches for prevention and management of type 2 diabetes*. Bmj, 2018. **361**: p. k2234.
40. Lønnberg, A.S., et al., *Association of Psoriasis With the Risk for Type 2 Diabetes Mellitus and Obesity*. JAMA Dermatol, 2016. **152**(7): p. 761-7.

41. Wang, Z., et al., *Blood DNA methylation markers associated with type 2 diabetes, fasting glucose, and HbA1c levels: An epigenome-wide association study in 316 adult twin pairs*. Genomics, 2021. **113**(6): p. 4206-4213.
42. Chen, S.J., et al., *Modification of physical activity on genetic effects of type 2 diabetes mellitus: a twin pairs cohort study in China*. Zhonghua Liu Xing Bing Xue Za Zhi, 2021. **42**(7): p. 1167-1173.

Figure legends

Figure 1. Flow chart of quality control

CKB: China Kadoorie Biobank; BMI: body mass index; SD: standard deviation; MAF: minor allele frequency.

Figure 1. Flow chart of quality control

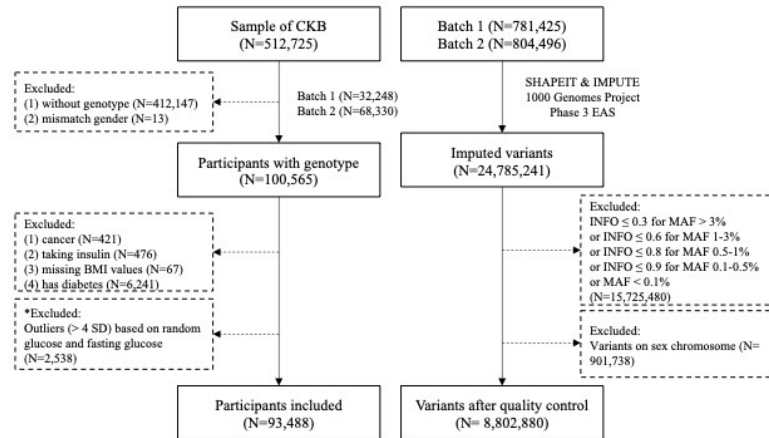


Figure 2. The performance of polygenic risk score.

Performance of the various polygenic risk score (PRS) approaches. (A) Mean time (in minutes) and mean memory (in GB) of PRS calculation across 10 repeats when applied to different sizes of the target sample. (B) Predictive accuracy of the 12 PRS methods for T2D in different sample sizes. The PRS programs (the first 4 programs) were run using their default parameter settings. The number in the panels represents the genotypic variance explained (R^2) by the PRS generated from each program, and the color represents the size of R^2 . The Y-axis represents the different sizes of the target sample, while the X-axis represents the different PRS calculation methods.

Figure 2. The performance of polygenic risk score.

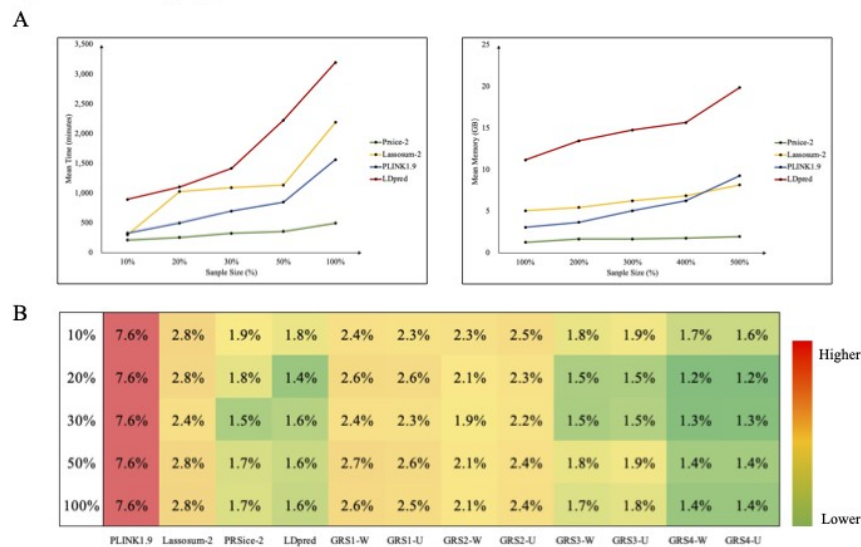


Figure 3. Relationship of the standard polygenic risk score to incident T2D.

The bar chart shows the number and proportion of participants for polygenic risk score (PRS), and the cubic spline modes for the association of PRS with the incidence risks of T2D. Models were adjusted by sex, age, region, ten genetic principal components, and regional principal components. The solid line denotes the adjusted hazard ratio, and the green area denotes 95% confidence intervals. The standardized cumulative incidence of T2D in five PRS groups in the CKB cohort during about 12 years of follow-up; shaded regions represent the 95% confidence intervals.

Figure 3. Relationship of standard polygenic risk score to incident T2D.

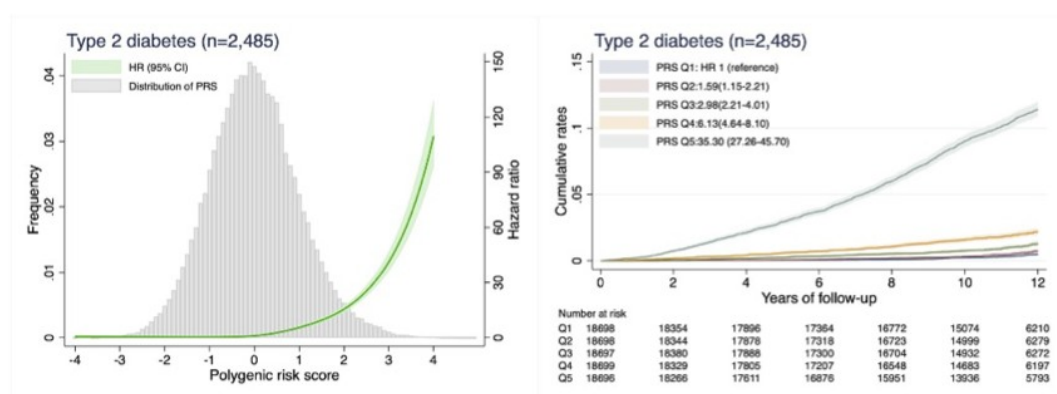


Figure 4. Polygenic risk score \times cardiometabolic risk score (PRS \times CRS) analysis

The Y-axis represents the genotypic variance explained (R^2), while the X-axis represents the ratio of the linear regression model of PRS and cardiometabolic score. The red line represents R^2 when only PRS and

CRS are considered, and the orange line represents R^2 when PRS, CRS and their interaction are considered.

Figure 5. Polygenic risk score \times cardiometabolic risk score (PRS \times CRS) analysis

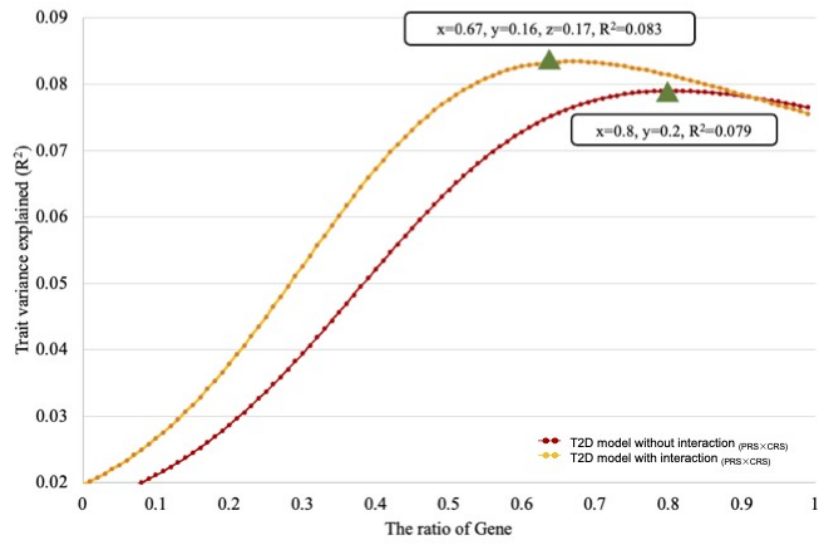


Table 1. Baseline characteristics of participants with and without type 2 diabetes during follow-up

Characteristic	T2D incidence	Non-T2D
No	2,485 (2.7)	91,003 (97.3)
Age, mean (SD)	56.4 (9.9)	53.2 (11.0)
Sex		
Male	1,021 (41.1)	39,127 (43.0)
Female	1,464 (58.9)	51,876 (57.0)
City		
Urban	1,523 (61.3)	52,311 (57.5)
Rural	962 (38.7)	38,692 (42.5)
Family incomes (yuan)		
<10000	630 (25.4)	38,612 (31.4)
10000-19999	647 (26.0)	27,319 (30.0)
>=20000	1,208 (48.6)	35,072 (38.5)
Education level		
Primary and below	1,684 (67.8)	47,862 (52.6)
High school	717 (28.9)	37,735 (41.5)
College or above	84 (3.4)	5,406 (5.9)
Marital status		
Unmarried	278 (11.2)	9,589 (10.5)
Married	2,207 (88.8)	81,414 (89.5)
Family history		
Acute myocardial infarction	74 (3.0)	3,048 (3.4)
Stroke	434 (17.5)	16,450 (18.1)
Malignant tumor	411 (16.5)	15,012 (16.5)
Diabetes	222 (8.9)	5,424 (6.0)
Lifestyle, %		
No-smoker current	1,705 (68.6)	62,436 (68.6)
No-drinker	2,278 (91.7)	84,083 (92.4)
Healthy exercise	1,278 (51.4)	46,893 (51.5)
Healthy diet	158 (6.4)	6,222 (6.8)
Mean metabolism, mean (SD)		
BMI (Kg/m ²)	25.6 (3.6)	23.5(3.4)
SBP (mmHg)	141.8 (22.9)	132.6 (22.4)
DBP (mmHg)	81.6 (11.2)	78.4 (11.7)
FG (mmol/L)	5.78 (0.77)	5.48 (0.71)
RG (mmol/L)	6.33 (1.38)	5.63 (1.15)

T2D, type 2 diabetes; SD, standard deviation; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; FG, fasting glucose; RG, random glucose

Table 2. Risk of type 2 diabetes according to polygenic risk score groups

Outcome	PRS groups					P trend	Per group increment	PAR% [#]
	Q1	Q2	Q3	Q4	Q5			
cases	72 (201152)	108 (200744)	192 (200743)	342 (199373)	1771 (195093)			
model 1 ^a	Reference	1.56 (1.12-2.16)	2.90 (2.15-3.90)	5.87 (4.44-7.74)	35.66 (27.55-46.17)	<0.001	3.15 (2.99-3.31)	88% (85,90)
model 2 ^b	Reference	1.55 (1.12-2.15)	2.89 (2.14-3.89)	5.82 (4.41-7.69)	35.17 (27.17-45.54)	<0.001	3.13 (2.98-3.30)	88% (85,90)
model 3 ^c	Reference	1.59 (1.15-2.21)	2.98 (2.21-4.01)	6.12 (4.64-8.09)	35.30 (27.26-45.70)	<0.001	3.09 (2.93-3.25)	88% (86,90)

PRS, polygenic risk score; HR, Hazard ratio; PAR%, percentage of population attributable risk. All HRs (95% CI) were derived from Cox proportional hazards regression. ^aModel 1 were adjusted for sex, age, region, genetic principal components, and regional principal components; ^bModel 2 was adjusted for model 1 plus education levels, marital status, family history of diabetes; ^cModel 3 additionally adjusted for blood pressure, lipid-lowering drug taking, body mass index, waist circumference, smoking status, drinking status, diet, and physical activities. [#]Percentage (95% CI) of incident cases theoretically attributable to disadvantageous PRS group (Q1).

Table 3. Associations between lifestyle factors, metabolism factors and risk of type 2 diabetes.

Items	T2D incidence		T2D incidence (adjusted for PRS)	
	HR (95% CI)*	PAR%#	HR (95% CI)*	PAR%#
Total scores				
≤2	1.00(ref)		1.00(ref)	
3	0.85(0.65,1.11)		0.95(0.73,1.24)	
4	0.54(0.42,0.70)	66 (58,74)	0.69(0.54,0.89)	61(51,70)
5	0.35(0.27,0.45)		0.49(0.38,0.63)	
6	0.19(0.15,0.26)		0.30(0.22,0.39)	
≥7	0.13(0.09,0.18)		0.20(0.14,0.29)	
Smoking ^a				
Unhealthy	1.00(ref)	3 (-1,6)	1.00(ref)	0 (-4,4)
Healthy	0.92 (1.81,1.05)		1.00 (0.96,1.04)	
Alcohol ^b				
Unhealthy	1.00(ref)	-1 (-3,0)	1.00(ref)	-1 (-3,0)
Healthy	1.14 (0.97-1.34)		1.15 (0.97-1.35)	
Physical activities ^c				
Unhealthy	1.00(ref)	-2 (-11,5)	1.00(ref)	-2 (-10,5)
Healthy	1.05 (0.89-1.23)		1.05 (0.89-1.23)	
Diet ^d				
Unhealthy	1.00(ref)	1 (-18,16)	1.00(ref)	-1 (-20,15)
Healthy	0.99 (0.83-1.19)		1.01 (0.84-1.22)	
Obesity ^e				
Unhealthy	1.00(ref)	42 (56,61)	1.00(ref)	38 (35,41)
Healthy	0.40 (0.36-0.43)		0.45 (0.41-0.49)	
Blood pressure ^f				
Unhealthy	1.00(ref)	34 (26,41)	1.00(ref)	26 (17,33)
Healthy	0.61 (0.54-0.70)		0.71 (0.62-0.81)	
Blood lipid ^g				
Unhealthy	1.00(ref)	0 (0,0)	1.00(ref)	0 (0,0)
Healthy	0.60 (0.33-1.09)		0.76 (0.42-1.38)	

T2D, type 2 diabetes; HR, Hazard ratio; PAR%, percentage of population attributable risk; CI, confidence interval. ^aQuit smoking because of illness; ^bNon-daily or daily alcohol consumption (<25 g/ day for men and <15 g/ day for women); ^cRanked in the top 50% of the same sex population in terms of metabolic equivalent -hour/day; ^dEat vegetables and fruits daily and limit red meat intake (1-6 days/week); ^eBody mass index (18.5-23.9 kg/m²), waist circumference (male <90 cm, female <85 cm); ^fNo history of hypertension, untreated blood pressure <120/80 mmHg; ^gDid not receive lipid-lowering drugs. All HRs (95% CI) were derived from Cox proportional hazards regression. *Model adjusted for age, sex and

recruitment assessment centre, body mass index, smoking status, alcohol daily intake, family disease history (diabetes, CVD and cancer). #Percentage (95% CI) of incident cases theoretically attributable to disadvantageous CRS group (score \geq 3).

Table 4. The difference of incidence of type 2 diabetes between PRS groups

PRS groups	Predicted incidence of T2D (%)								
	40	45	50	55	60	65	70	75	80
PRS Q1	0.05%	0.05%	0.12%	0.26%	0.37%	0.55%	0.85%	1.24%	1.83%
PRS Q2	0.05%	0.11%	0.25%	0.38%	0.59%	0.90%	1.29%	1.78%	2.25%
PRS Q3	0.05%	0.11%	0.41%	0.73%	1.25%	1.84%	2.82%	3.28%	4.13%
PRS Q4	0.12%	0.40%	0.88%	1.59%	2.40%	3.47%	4.78%	5.91%	6.84%
PRS Q5	0.49%	1.95%	4.26%	7.61%	11.91%	16.53%	21.73%	26.77%	31.31%

PRS, polygenic risk scores; T2D, type 2 diabetes;