IJC
International Journal of Cancer

Diabetes and cancer risk: A Mendelian randomization study

Atsushi Goto[1], Taiki Yamaji[1], Norie Sawada[1], Yukihide Momozawa[2], Yoichiro Kamatani[2], Michiaki Kubo[3], Taichi Shimazu[1], Manami Inoue[1], Mitsuhiko Noda[4], Shoichiro Tsugane[1] and Motoki Iwasaki[1]

[1]Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan
[2]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
[3]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
[4]Department of Endocrinology and Diabetes, Saitama Medical University, Saitama, Japan

Earlier cohort studies using conventional regression models have consistently shown an increased cancer risk among individuals with type 2 diabetes. However, reverse causality and residual confounding due to common risk factors could exist, and it remains unclear whether diabetes *per se* contributes to cancer development. Mendelian randomization analyses might clarify the true association between diabetes and cancer risk. We conducted a case–cohort study with 10,536 subcohort subjects and 3,541 newly diagnosed cancer cases derived from 32,949 eligible participants aged 40–69 years within the Japan Public Health Center-based Prospective Study. With 29 known type 2 diabetes susceptibility variants, we used an inverse variance-weighted method to estimate hazard ratios for the associations of diabetes with risks of total and site-specific cancers. The hazard ratios of cancer per doubling of the probability of diabetes were 1.03 (95% confidence interval [CI], 0.92–1.15) overall, 1.08 (95% CI: 0.73–1.59) for the pancreas, 0.80 (95% CI: 0.57–1.14) for the liver and 0.90 (95% CI: 0.74–1.10) for the colorectum. Additional analyses, using publicly available large-scale genome-wide association study data on colorectal cancer in Japan, resulted in a narrower CI (hazard ratio: 1.00; 95% CI: 0.93–1.07). In this prospective Mendelian randomization study with a large number of incident cancer cases, we found no strong evidence to support associations between diabetes and overall and site-specific cancer risks. Our findings suggest that there is little evidence to support the genetic role of type 2 diabetes in cancer development in the Japanese population.

Introduction

Over the decades, a number of cohort studies reported associations between type 2 diabetes and increased cancer risks, especially liver, colorectal and pancreatic cancers.[1,2] More recently, studies have shown that high levels of blood glucose[3,4] or glycated hemoglobin A1c[5] are associated with cancer incidence. These associations are biologically plausible, given that hyperglycemia increases mitochondrial glucose oxidation, thereby promoting DNA damage through oxidative stress. Because diabetes or high glucose levels could be prevented through lifestyle modifications or medications, these could be potential strategies for cancer prevention.

Key words: diabetes, Mendelian randomization, genetics, epidemiology

Abbreviations: BBJ: BioBank Japan; BMI: body mass index; CI: confidence interval; FIOC: Fundamental Innovative Oncology Core; GWAS: genome-wide association study; ICD-O-3: International Classification of Diseases for Oncology, Third Edition; IVW: inverse variance-weighted; JPHC: Japan Public Health Center-based Prospective; SNP: single nucleotide polymorphism

Additional Supporting Information may be found in the online version of this article.

Conflict of interests: YK declares that he has a potential financial conflict of interest as follows: when the colorectal cancer GWAS used in this article was performed, YK received funding as a BioBank Japan project from the Japan Agency for Medical Research and Development (JP17km0305002) until March 2018. Results of the GWAS are publicly available as written in the article. Other authors declare no conflict of interest.
Grant sponsor: Ministry of Health, Labor and Welfare of Japan; **Grant sponsor:** Japan Society for the Promotion of Science (KAKENHI); **Grant number:** 26290046; **Grant sponsor:** National Cancer Center Research and Development Fund; **Grant numbers:** 23-A-31 [toku], 26-A-2, 28-A-19, 29-A-4; **Grant sponsor:** Japan Agency for Medical Research and Development; **Grant numbers:** 15ck0106095h0002, 16ck0106095h0003, 17ck0106266h001, 17ck0106266h0003, 17ck0106370h0001, 18ck0106370h0002
[Correction added on May 10, 2019 after first online publication: Figure 1 caption updated.]
DOI: 10.1002/ijc.32310

History: Received 14 Sep 2018; Accepted 13 Mar 2019; Online 30 Mar 2019.

Correspondence to: Atsushi Goto, Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan. E-mail: atgoto@ncc.go.jp

What's new?

While type 2 diabetes is implicated in cancer development, the two conditions share multiple risk factors, raising questions about the actual contribution of diabetes to cancer risk. Here, a Mendelian randomization (MR) analysis was used to clarify the relationship between diabetes and cancer in a Japanese cohort. Based on data for 32,949 individuals, including 3,541 incident cancer cases, MR analysis revealed no strong genetic evidence supporting a link between diabetes and cancer risk, including site-specific and overall risk. This finding was confirmed by investigation of data from a genome-wide association study of colorectal cancer from the BioBank Japan Project.

However, the causal relationship between diabetes and cancer risk has not been established yet. Diabetes and cancer have many common risk factors such as unhealthy diet, physical inactivity, smoking and adiposity, which raise the likelihood of residual confounding. If diabetes occurs secondary to precancerous state or subclinical cancer, reverse causation could also be present.

A Mendelian randomization (MR) approach may overcome the limitations of traditional observational study designs by using Mendel's law. The MR approach utilizes the random allocation of genotypes at conception, which makes genotypes to be independent of potential confounders and also avoids reverse causation.[6]

Large-scale MR studies have been increasingly reported in recent years.[7,8] However, these studies often consist of mixed populations mainly from traditional case–control studies, with high risk of selection bias and heterogeneity, and lack important information about lifestyle factors prior to the onset of cancer. Thus, MR studies from prospective cohort studies with detailed covariates at baseline would probably give less biased estimates than those from case–control studies. Therefore, we performed an MR analysis to clarify the true association between diabetes and cancer risk using genotyped and imputed genetic data from the Japan Public Health Center-based Prospective (JPHC) Study.

Methods

Study population

The JPHC study was initiated in 1990 (cohort I) and 1993–1994 (cohort II). All subjects were Japanese from 11 public health center (PHC) areas and were aged 40–59 years in 1990 (cohort I) and 40–69 years in 1993 (cohort II) at the time of their first survey. The JPHC Study has been described in detail previously.[9] Our study population was derived from a cohort of 33,736 subjects from nine PHC areas across Japan, who responded to the baseline questionnaire and provided blood samples during the health checkup. We excluded 787 participants who had past history of cancer or cancer incidence before their return of the self-administered questionnaire. Our study was approved by the institutional review board of the National Cancer Center, Japan.

Follow-up and case–cohort selection

We followed up the subjects from the start of the study period until December 31, 2009. Changes in residence status and deaths were ascertained through the residential registry. Among the study population of 32,949 subjects, 1,946 (5.9%) died, 1,085 (3.3%) moved out of the study area and 84 (0.3%) were lost to follow-up. Cancer occurrence was documented through active notifications from the major hospitals in the study areas and data linkage with population-based cancer registries. Death certificates were also used as a supplementary information source. The site and histological features of each cancer case were coded according to the International Classification of Diseases for Oncology, Third Edition (ICD-O-3).[10] For the registry system used, 4.1% of the cases only had information available from the death certificates. For analysis, the earliest date of diagnosis was considered for cases with multiple primary cancers occurring at different times. For the present case–cohort study, all incident cancers were included as case subjects. Cancer was newly diagnosed in 3,750 subjects (2,028 men and 1,722 women) during a median follow-up of 15.9 years. A subcohort sample of 13,024 participants was randomly selected from the base cohort participants irrespective of the cancer and follow-up status, comprising approximately 40% of the base cohort. We further excluded 298 subjects with past history of cancer and four with missing follow-up times, from the subcohort sample, leaving 12,722 subcohort subjects and 3,750 incident cancer cases potentially eligible for the analyses.

Laboratory assays

For the potential gain in power or the reduction in variance, given a fixed maximum cost for our study, a case–cohort design was chosen for the present analyses on genetically predicted diabetes in relation to cancer risk. For each participant, DNA was extracted from a buffy coat of white blood cells using FlexiGene DNA kits (Qiagen, Hilden, Germany). DNA samples of the participants were analyzed using the HumanOmniExpressExome-8 v1.2 BeadChip, the HumanOmniExpress-12 BeadChip, or the HumanOmni2.5-8 BeadChip arrays (Illumina Inc., San Diego, CA). The genotyping for our study was conducted at the Genetics Division or the Department of Clinical Genomics, Fundamental Innovative Oncology Core (FIOC), National Cancer Center Research Institute or at the RIKEN Center for Integrative Medical Sciences. Among the potentially eligible subjects, we excluded those with inappropriate samples, no calls, call rate per sample of <0.99, inconsistent sex, non-Japanese samples based on principal component analysis using the smartpca,[11] or contaminated samples. Applying these sample exclusion criteria, genotype data were available for 12,158 subcohort subjects and 3,680 incident cases. We further excluded closely related samples using concordance

rate, leaving 10,536 subcohort subjects and 3,541 incident cases. For imputation, we excluded single nucleotide polymorphisms (SNPs) using standard SNP quality control (call rate per SNP of <0.99, minor allele frequency of <0.01, *p*-value of Hardy–Weinberg equilibrium in the control group of $<1 \times 10^{-4}$, a large allele frequency difference between the study sample and the reference panel, or SNPs on the Y chromosome). Then, we performed imputation of ungenotyped SNPs using SHAPEIT and IMPUTE2 with the 1,000 Genome Project Phase 3 (all ethnic groups) as a reference panel.

We selected a total of 29 SNPs, identified through published genome-wide association studies (GWASs), up until July 2014, and replicated in Asian populations as follows: *GCKR*, rs780094[12]; *IRS1*, rs2943641[13]; *DNER*, rs1861612[14]; *UBE2E2* rs6780569[15]; *PSMD6*, rs831571[16]; *IGFBP2*, rs4402960[17]; MAEA, rs6815464[16]; *CDKAL1*, rs7756992[18]; *ZFAND3*, rs9470794[16]; *KCNK16*, rs1535500[16]; *JAZF1*, rs864745[19]; *GCC/PAX4*, rs6467136[16]; *ARF5/PAX4/SND1*, rs10229583[20]; *MIR129/LEP*, rs791595[21]; *ANK1*, rs515071[22]; *SLC30A8*, rs13266634[18]; *GLIS3*, rs7041847[16]; *CDKN2A/B*, rs2383208[23]; *GPSM1*, rs11787792[21]; *CDC123/CAMK1D*, rs12779790[19]; *HHEX*, rs1111875[17]; *TCF7L2*, rs7903146[17]; *KCNQ1*, rs2237892[24]; *KCNJ11*, rs5219[17]; *FTO*, rs8050136[25]; *C2CD4A*, rs7172432[15]; *SLC16A13*, rs312457[21]; *PEPD*, rs3786897[16]; *FITM2/R3HDML/HNF4A*, rs6017317.[16] Among the 29 SNPs, 11 were genotyped and 18 were imputed (Supporting Information Table S1). With the exception of rs7041847 (INFO as a measure of imputation accuracy of 0.5168), all imputed SNPs showed high imputation quality (INFO > 0.9). Thus, we performed an analysis excluding rs7041847 on *GLIS3*; however, the results did not materially change, and rs7041847 was included in the main results.

Definition of diabetes

Diabetes was defined by (*i*) a self-reported diabetes diagnosis and/or use of a glucose-lowering medication via a self-administered questionnaire at baseline or (*ii*) high blood glucose levels during the health checkups at baseline: ≥ 126 mg/dl for those who fasted for ≥8 hr or ≥200 mg/dl for nonfasting subjects[26] (defined by <8 hr or missing data on fasting hours). Blood glucose levels were available for 8,659 subjects among the subcohort and 2,815 incident cancer cases. In a validation study conducted in a subpopulation of the JPHC Study, 94% of subjects with self-report of diabetes were confirmed to have diabetes, after a medical chart review.[27]

Statistical methods

To investigate whether the reported diabetes susceptibility SNPs were associated with diabetes in our study, we fitted the logistic regression model with adjustment for age, sex and study areas under the additive genetic model among the 10,536 subcohort subjects (Supporting Information Table S2). Next, we examined the impact of diabetes susceptibility SNPs on total cancer and cancer risks at major sites: stomach (ICD-O-3: C16), colorectum (C18–C20), liver (C22), pancreas (C25), lung (C34), breast (C50) and prostate (C61) (Supporting Information Tables S3–S10). We further divided the participants into five groups based on quintiles of the total number of risk alleles of the diabetes susceptibility SNPs (Supporting Information Table S1) among the subcohort subjects. Then, we fitted the logistic regression model with adjustment for age, sex and study areas to estimate the odds ratios for diabetes by each quintile among the 10,536 subcohort subjects. We also fitted the weighted Cox model[28] with stratification by study areas to examine hazard ratio for cancer by each quintile, with adjustment for age (continuous) and sex (men *vs.* women) among the 10,536 subcohort subjects and 3,541 incident cases. Furthermore, to compute *p* values for linear trend, we included the total number of risk alleles as a continuous variable to the logistic and Cox models.

For conventional analyses on associations between diabetes and total cancer risk and risks of cancer at major sites, we used the weighted Cox proportional hazards model with stratification by study areas and adjustment for age (continuous), sex (men *vs.* women), body mass index (BMI; continuous), smoking (never, past or current smokers; categorical), alcohol intake (none, occasional or regular: 1–150 g/week; or regular drinkers: >150 g/week; categorical), coffee consumption (almost none, 1–2 days/week, 3–4 days/week, 1–2 cups/day, 3–4 cups/day or ≥5 cups/day; categorical) and physical activity (none, 1–3 days/month or ≥1/week; categorical). To address missing data on BMI, smoking, alcohol intake, coffee consumption and physical activity, we performed multivariate normal imputation (the SAS PROC MI procedure) with 100 rounds of multiple imputations. We included all covariates, follow-up length and outcome status to account for missing data. We then averaged the estimates from each imputed dataset according to Rubin's rules using the SAS PROC MIANALYZE procedure.

For instrumental variable analyses, we used estimates for each SNP-cancer association (the first stage model: Supporting Information Tables S3–S10) and estimates for each SNP-diabetes association (the second stage model: Supporting Information Table S2) as described above. To increase the statistical power of detecting a significant association, we also utilized the publicly available results for SNP-colorectal cancer associations from the BioBank Japan (BBJ) Project (sex-combined: 6,692 colorectal cancer patients and 27,178 controls; Supporting Information Table S11).[29] Using the estimates for 29 SNPs, we employed an inverse variance-weighted (IVW) method, a summary data-based MR analysis,[30,31] to estimate MR estimates for the association between diabetes and cancer risk. The original MR estimate represents the change in the log hazard for cancer per unit increase in the log odds of diabetes. A one-unit increase in the log odds of diabetes corresponds to an exp (1) = 2.72-fold multiplicative increase in the odds of diabetes. To enable clearer interpretation of this relation, we have multiplied the original MR estimate by loge2 (= 0.693), yielding the change in log hazard per twofold increase (doubling) in the odds of diabetes. Because diabetes was a rare disease in Japan during the study period, the odds approximates the probability. We then exponentiated the estimates to obtain the hazard ratios for cancer per doubling in the probability of diabetes.[32]

MR analyses derive valid estimates where the following assumptions are met[33]: (*i*) the SNPs are correlated with diabetes,

(*ii*) the SNPs affect cancer risk only through their effects on diabetes and (*iii*) the SNPs are independent of any confounding factors for the association between diabetes and cancer risk. For assumption (*i*), because we selected SNPs that were identified through the GWAS and replicated in Asian populations, such SNPs are likely to be correlated with diabetes in the source population.

For assumptions (*ii*) and (*iii*), we searched the PhenoScanner database (available at http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner)[34] and the Haploreg database (available at: http://archive.broadinstitute.org/mammals/haploreg/haploreg.php) to examine whether the chosen SNPs had pleiotropic effects on other phenotypes.[35] Some SNPs were associated with traits such as fasting insulin levels at the GWAS significance level ($p < 5.0 \times 10^{-8}$); rs780094 on the *GCKR* was significantly associated with fasting insulin-related traits;[12] rs7756992 on the *CDKAL1* was associated with insulin-related traits;[36] rs864745 on *JAZF1* was associated with Crohn's disease[37] but not likely to directly affect diabetes status; rs7903146 on *TCF7L2* was associated with proinsulin and insulin levels;[38] and rs2237892 on *KCNQ1* was associated with BMI.[39] Furthermore, among the subcohort subjects of our study, we examined whether SNPs were significantly associated with established risk factors for cancer, including BMI, smoking, alcohol intake and physical inactivity after applying the Bonferroni correction by dividing the *p*-value of 0.05 by the number of SNPs examined ($p < 0.05/29 \approx 1.72 \times 10^{-3}$).

Manual pruning of SNPs that potentially have horizontal pleiotropic effects is not generally recommended, because such pruning might lead to an instrument that is not biologically meaningful.[40] Instead, as a sensitivity analysis, we conducted an MR-Egger regression, which provides valid estimate even in the presence of unbalanced pleiotropy.[41] We further performed the leave-one-out analysis by recalculating the MR estimates with the IVW method by sequentially dropping one SNP at a time to examine whether a single SNP that might have a large horizontal pleiotropic effect would influence the MR estimates.[31]

Thresholds for nominal significance were set at a two-sided *p*-value of <0.05. Statistical analyses were conducted using SAS (version 9.4; SAS Institute, Cary, NC) or R (version 3.4.2; R Foundation for Statistical Computing, Vienna, Austria).

Results

Among the subcohort subjects ($n = 10{,}536$), 5.8% ($n = 613$) had diabetes at baseline. Among them, 447 were known to have diabetes, while 166 were identified to have diabetes by blood glucose testing during the health checkups. Among those with incident cancer ($n = 3{,}541$), 8.4% ($n = 298$) had diabetes at baseline. Of them, 230 were known to have diabetes and 68 were identified to have the condition by blood glucose testing. Compared to subjects without diabetes, those with diabetes tended to be older individuals and smokers, had a higher BMI, engaged in more physical activity, drank alcohol more frequently and consumed coffee less frequently (Table 1).

Among the 29 SNPs, four (rs7756992, rs2383208, rs1111875 and rs2237892) were nominally significantly associated with diabetes in our study (Supporting Information Table S2). Quintiles of the total number of risk alleles were positively associated with diabetes in a dose–response manner (*p*-value for liner trend = 1.25×10^{-5}) but were not significantly associated with cancer risk (overall: *p*-value for linear trend = 0.95; stomach: 0.44; colorectum: 0.18; liver: 0.94; pancreas: 0.38; lung: 0.89; breast: 0.35; prostate: 0.52; Fig. 1).

Table 1. Baseline characteristics among subcohort subjects ($n = 10{,}536$)

	All subjects	Subjects without diabetes	Subjects with diabetes
Characteristic[1]	$n = 10{,}536$	$n = 9{,}923$	$n = 613$
Age, years	553.9 ± 7.9	53.7 ± 7.9	56.6 ± 7.7.6
Men, %	35.1	33.7	57.3
BMI, kg/m^2	23.6 ± 3.0	23.6 ± 3.0	24.2 ± 3.3
Missing data, %	0.9	0.8	1.6
Smoking status			
Current smoking, %	17.9	17.2	29.2
Past smoking, %	10.0	9.7	16.2
Missing data, %	0.5	0.5	0.3
Physical activity			
≥1 day/week, %	17.5	17.4	20.4
Missing data, %	1.4	1.4	1.3
Alcohol consumption			
Current drinking,[1] %	28.1	27.2	41.4
Missing data, %	2.7	2.6	3.4
Coffee consumption			
≥1 day/week, %	30.9	31.4	21.9
Missing data, %	1.0	0.9	1.1

Data are presented as mean ± standard deviation.
[1]Alcohol consumption at least once per week.
Abbreviation: BMI, body mass index.

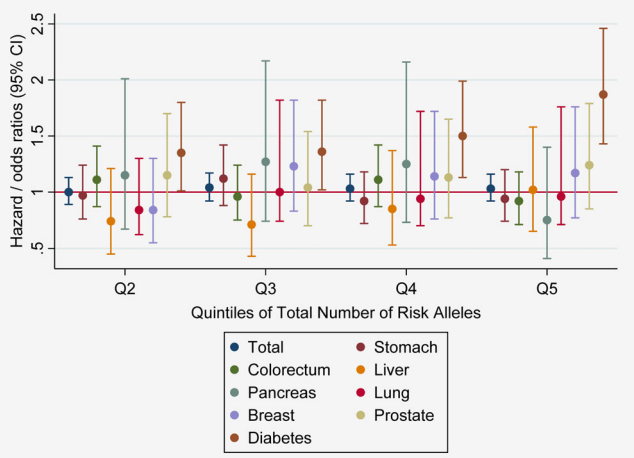

Figure 1. Diabetes/cancer risk by quintiles of total number of risk alleles for diabetes. Abbreviation: CI, confidence interval. Hazard ratios were estimated using the weighted Cox model to examine cancer risk by each quintile comparing to the lowest quintile with adjustment for age (continuous), sex (except for prostate and breast cancer), and study areas among the 10,536 subcohort subjects and 3,541 incident cases. Odds ratios were estimated using the logistic regression model with adjustment for age, sex, and study areas to examine the association of diabetes by each quintile comparing to the lowest quintile among the 10,536 subcohort subjects.

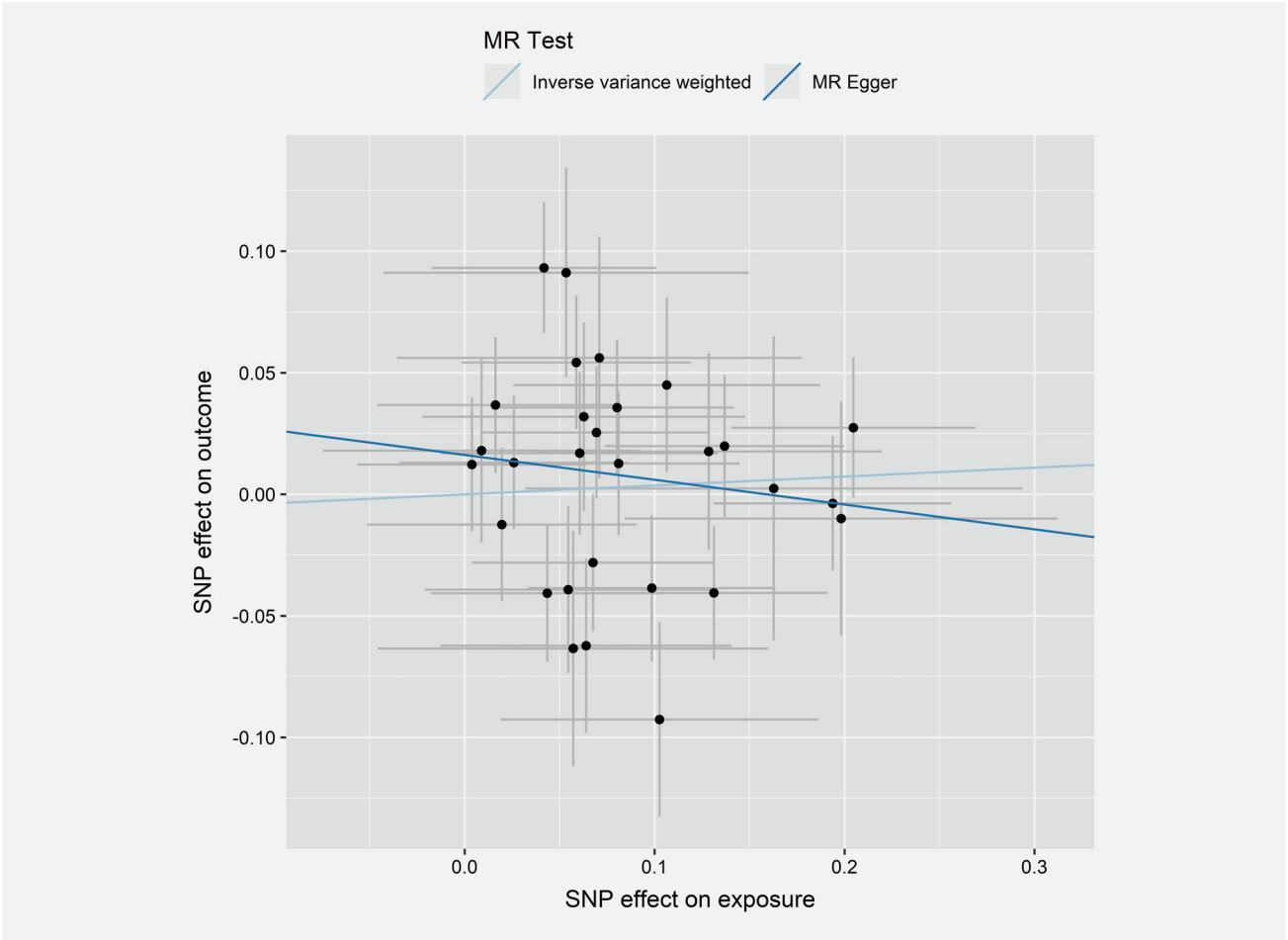

Figure 2. Scatter plot showing the associations of the SNP effects on the diabetes (log odds ratio) against the SNP effects on the total cancer (log hazard ratio). Abbreviations: MR, Mendelian randomization; SNP, single nucleotide polymorphisms. The inverse variance-weighted estimates are represented by a dotted light blue line, and the Egger regression estimates by a dark blue line.

In contrast to conventional analyses showing a positive association between diabetes and cancer risks, the MR analysis indicated no strong evidence to support associations between diabetes and risks of total cancer and cancer subtypes (Figs. 2 and 3). The IVW hazard ratios per doubling of the probability of diabetes for cancer were 1.03 (95% confidence interval [CI], 0.92–1.15) overall, 1.08 (0.73–1.59) for the pancreas, 0.80 (0.57–1.14) for the liver and 0.90 (0.74–1.10) for the colorectum (Fig. 3). The Egger estimates resulted in slightly wider confidence intervals. We also examined whether the selected variants were known associated risk factors of cancer. After the Bonferroni correction, statistically significant findings were observed between three SNPs and associated risk factors: rs780094 at the *GCKR* and BMI, p-value = 5.8×10^{-4}; rs8050136 at the *FTO* and BMI, p-value = 6.87×10^{-12}; and rs5219 at the *KCNJ11* and regular alcohol drinking, p-value = 1.64×10^{-3}; data not shown). However, p values for intercepts from the Egger regression did not suggest the possibility of pleiotropic effects (Fig. 3). Furthermore, the leave-one-out analyses, performed by sequentially dropping one SNP at a time, resulted in similar findings (Fig. 4). Additional analyses, using publicly available genetic data on SNP-colorectal cancer associations from the BBJ Project (29 SNPs), resulted in a narrower confidence interval (hazard ratio for colorectal cancer = 1.00 [0.93–1.07]; Fig. 3 and Supporting Information Table S11).

Discussion

In this population-based prospective study, we utilized the MR methodology and examined the relation between diabetes and cancer risk in a large number (3,541) of incident cancer cases from 32,949 people as a source population, during a median follow-up of 15.9 years. In our MR analysis, no strong evidence was found to support associations between diabetes and the risks of total cancer, colon cancer, pancreatic cancer or liver cancer. These findings were robust to a number of sensitivity analyses. To further complement our results, we utilized publicly available large-scale GWAS results of colorectal cancer (the BBJ Project) and confirmed that genetically predicted diabetes was not associated with colorectal cancer. The results of our study support the notion that diabetes *per se* might not be responsible for the reported positive association between diabetes and cancer.

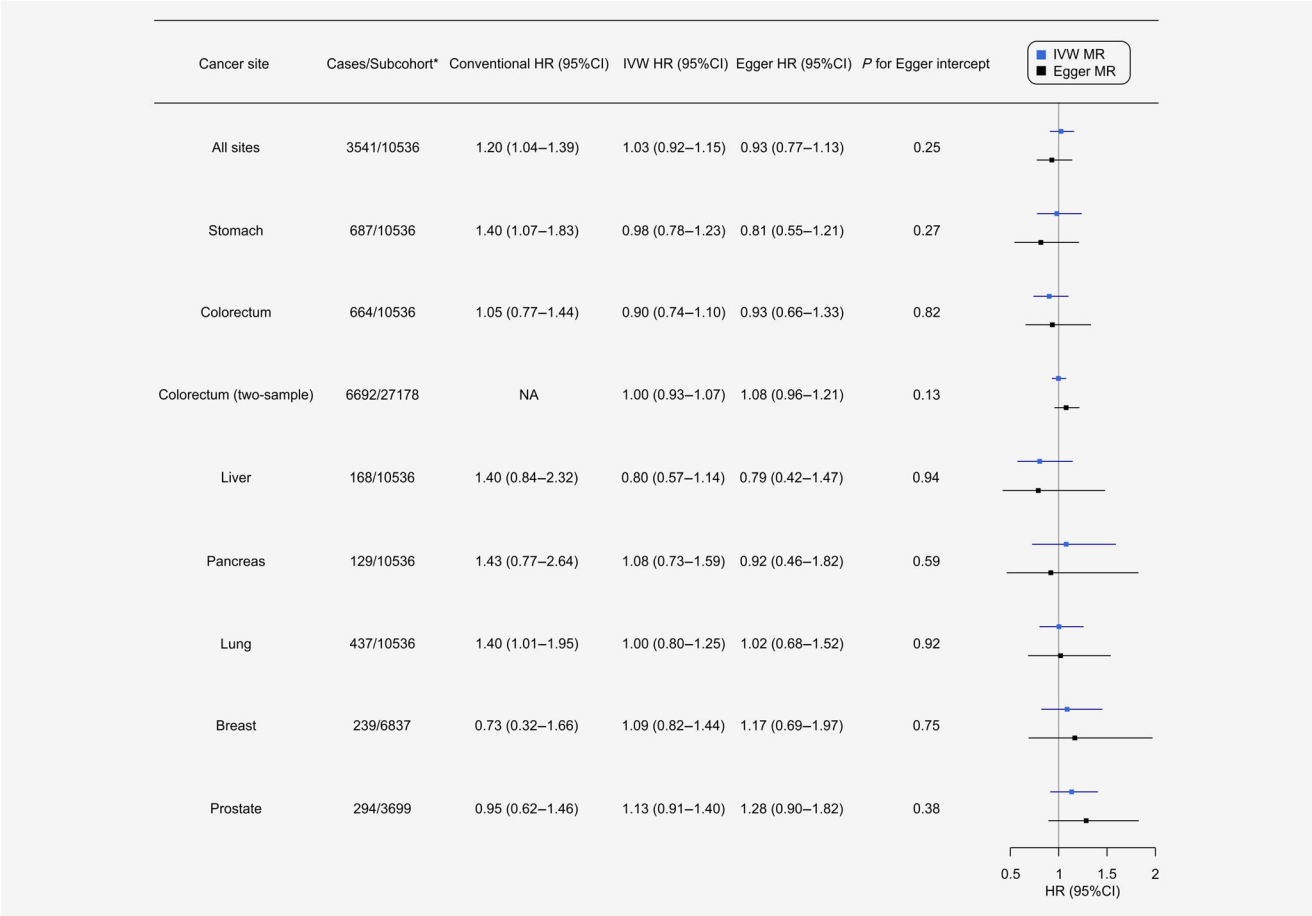

Figure 3. Association between diabetes and cancer risk. Abbreviations: BBJ, BioBank Japan; CI, confidence interval; HR, hazard ratio; IVW, inverse variance-weighted; JPHC, Japan Public Health Center-based Prospective; MR, Mendelian randomization; NA, not applicable; SNP, single nucleotide polymorphism. *Number of cases/controls are shown for the colorectum (two-sample). The conventional HRs for cancer comparing people with diabetes to those without diabetes were estimated using the weighted Cox model with adjustment for age, sex (except for breast and prostate cancer), study areas, body mass index, smoking, alcohol intake, coffee intake and physical activity. The IVW HRs per doubling of the probability of diabetes were estimated using the IVW method with the estimates from the first and second stage models for 29 SNPs. The Egger HRs per doubling of the probability of diabetes and *p*-values for the Egger intercept were estimated using the MR-Egger method with the estimates from the first and second stage models for 29 SNPs. All results, except for the colorectum (two-sample), are summary data-based MR analyses in one sample. For the colorectum (two-sample), we employed a summary data-based MR in two samples (i.e., the JPHC data for SNP-diabetes associations and the BBJ data for SNP-cancer associations).

To the best of our knowledge, this is the first prospective MR study that examined the association between type 2 diabetes and risks of total cancer and site-specific cancer.

Our results are in general agreement with those reported in earlier MR studies with case–control data mainly conducted among European populations using a case–control strategy.[7,8] A recent MR study led by researchers from the International Agency for Research on Cancer investigated the associations of type 2 diabetes and metabolic factors with pancreatic cancer among 7,110 pancreatic cancer patients and 7,264 controls.[7] Using 43 of the susceptibility SNPs for type 2 diabetes, their MR analysis indicated that diabetes was not significantly associated with pancreatic cancer (odds ratio = 1.03; 95% CI = 0.85–1.11; $p = 0.47$). In contrast, their MR analysis found BMI and hyperinsulinemia to be significantly associated with pancreatic cancer. These findings and our prospective findings seem to suggest that there is little evidence to support the genetic role of type 2 diabetes in cancer development.

The null findings observed in our prospective MR analyses could be explained by several interpretations. One possibility is that our study did not have sufficient power to detect a significant association between type 2 diabetes and cancer risk. As shown in *post hoc* power calculations, we had 80% power to detect relatively small effect sizes for total cancer but not for cancer sites (Supporting Information Table S12). To complement the analysis, we used the two-sample MR analysis by using the JPHC data for SNP-diabetes associations and the BBJ data for SNP-colorectal cancer associations, which had 80% power to detect a small effect size (Supporting Information Table S11); however, the findings were consistent overall.

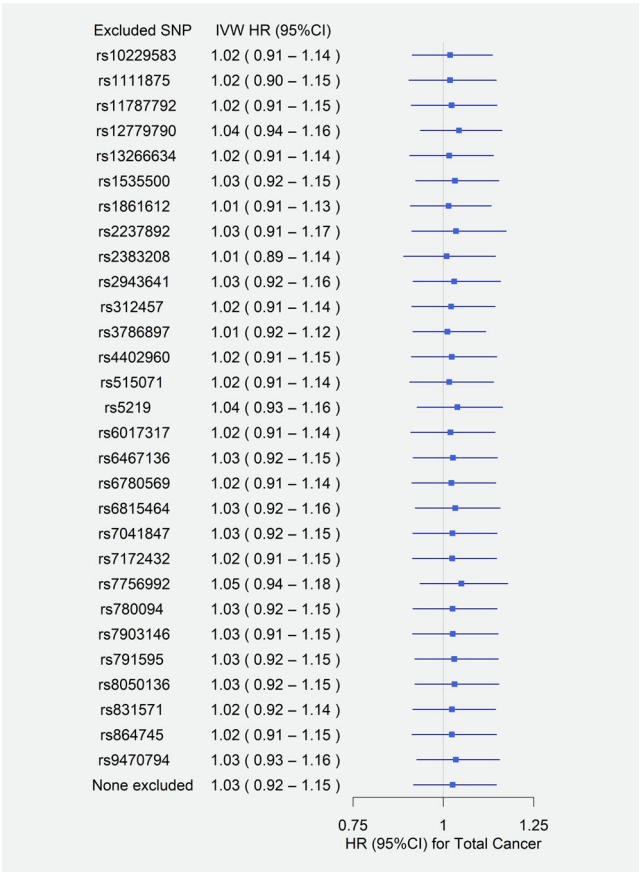

Figure 4. Leave-one-out analysis. Abbreviations: CI, confidence interval; HR, hazard ratio; IVW, inverse variance-weighted; MR, Mendelian randomization; SNP, single nucleotide polymorphism. The leave-one-out analysis was performed by recalculating the MR estimates using the IVW method, by sequentially dropping one SNP at a time to examine whether a single SNP that might have a large horizontal pleiotropic effect would influence the MR estimates. The IVW HRs per doubling of the probability of diabetes were estimated using the IVW method with the estimates from the first and second stage models for 29 SNPs. [Color figure can be viewed at wileyonlinelibrary.com]

Given the relatively wide confidence intervals for other cancer sites, a small effect of diabetes on these cancer sites cannot be ruled out. Another possibility is that type 2 diabetes susceptibility variants tended to be associated with both lower insulin levels and higher glucose levels. This may lead to vertical pleiotropy in which such genetic variants are associated with insulin and glucose levels on the same biological pathway from diabetes to cancer, thus not violating the MR assumptions. However, such genetic variants could also be associated with cancer through discrete pathways, which may lead to horizontal pleiotropy that violates the MR assumptions. For instance, variants in the *KCNQ1* gene might affect insulin secretion, leading to hyperglycemia and hypoinsulinemia, both of which could affect cancer risks. In this case, a possible carcinogenicity of hyperglycemia might have been wiped out by lower insulin levels, leading to null associations. However, the leave-one-out analysis also indicated null associations (Fig. 4). Furthermore, the Egger regression did not suggest the possibility of pleiotropic effects (Fig. 3). Finally, the impact of diabetes *per se* on cancer development, if any, may be smaller than we thought. The conventional regression analyses might have overestimated the true association, possibly due to uncontrolled confounding by common risk factors or reverse causation.

Our MR analysis provides evidence that the genetic mechanisms responsible for type 2 diabetes may not play major roles in cancer development. However, from a public health perspective, it is certainly important to control such shared risk factors for the prevention of both diseases, because diabetes and cancer share a number of established modifiable risk factors such as obesity, physical inactivity and smoking.

The strengths of our study include its prospective, nationwide, population-based design and high rate of follow-up. Our findings might be generalizable to non-Japanese populations, such as Asian populations with similar genetic and environmental background. Nevertheless, several limitations need to be addressed. First, we were unable to verify assumptions (*ii*) and (*iii*) of MR analyses listed in the "Statistical Methods" section. Although we exerted our best efforts to address potential pleiotropic or confounding effects, the possibility remains. Most importantly, as stated above, some variants included in our study are known to be associated with traits related to insulin rather than glucose, which may lead to horizontal pleiotropy. This potentially violates the assumption (*ii*) that SNPs affect cancer risk only through their effects on diabetes. Instead of using diabetes susceptibility variants, the use of variants related to glucose levels with such known function as glucose transporters might be better to meet this assumption. Second, although we included a large number of incident cases, the number of cases for each cancer site was limited. Therefore, we were probably underpowered to detect small effects for several cancer sites. However, the use of a publicly available large-scale GWAS on colorectal cancer from the BBJ Project resulted in a similar finding. Thus, similar analyses with the use of large-scale GWAS results among Japanese or Asian populations for other cancer sites should be performed, once the SNP-outcome associations become available.

In conclusion, in this prospective MR study with a large number of incident cancer cases, we found no strong evidence to support the associations between diabetes and overall and site-specific cancer risks. Our findings suggest that there is little evidence to support the genetic role of type 2 diabetes in cancer development in a Japanese population.

Acknowledgements

We thank the participants of the JPHC Study and all the staff members in each study area as well as the central offices for their cooperation and technical assistance. We thank Drs. Hiromi Sakamoto and Teruhiko Yoshida and the staff at the Genetics Division, Department of Clinical Genomics, FIOC, National Cancer Center Research Institute and the RIKEN Center for Integrative Medical Sciences for some of the genotyping analyses in our study. We appreciate Dr. Akira Saito and

Mr. Mitsuhiro Amemiya at the StaGen Co. Ltd for their support on genetic analysis. We are grateful for the valuable scientific suggestions of Dr. Stephen Burgess at the University of Cambridge, Dr. Simin Liu at the Brown University and Dr. Daniel Shriner at the National Human Genome Research Institute. We also thank Ms. Rieko Kanehara, Ms. Shiori Tanaka, Dr. Sanjeev Budhathoki and Dr. Tsuyuka Otsuki at the National Cancer Center and Dr. Akihisa Hidaka at the Jikei University for their contribution to our study. This work was supported by the National Cancer Center Research and Development Fund (23-A-31 [toku], 26-A-2, 28-A-19 and 29-A-4; since 2011); a Grant-in-Aid for Cancer Research from the Ministry of Health, Labor and Welfare of Japan (from 1989 to 2010); a Grant-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (grant number 26290046); and the Practical Research for Innovative Cancer Control (15ck0106095h0002, 16ck0106095h0003, 17ck0106266h001, 17ck0106266h0003, 17ck0106370h0001 and 18ck0106370h0002; since 2015) from the Japan Agency for Medical Research and Development. The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review and approval of the article; or the decision to submit the article for publication.

References

1. Inoue M, Iwasaki M, Otani T, et al. Diabetes mellitus and the risk of cancer: results from a large-scale population-based cohort study in Japan. *Arch Intern Med* 2006;166:1871–7.
2. Sasazuki S, Charvat H, Hara A, et al. Research Group for the Development Evaluation of cancer prevention strategies in Japan. Diabetes mellitus and cancer risk: pooled analysis of eight cohort studies in Japan. *Cancer Sci* 2013;104:1499–507.
3. Jee SH, Ohrr H, Sull JW, et al. Fasting serum glucose level and cancer risk in Korean men and women. *JAMA* 2005;293:194–202.
4. Stocks T, Rapp K, Bjorge T, et al. Blood glucose and risk of incident and fatal cancer in the metabolic syndrome and cancer project (me-can): analysis of six prospective cohorts. *PLoS Med* 2009;6:e1000201.
5. Goto A, Noda M, Sawada N, et al. High hemoglobin A1c levels within the non-diabetic range are associated with the risk of all cancers. *Int J Cancer* 2016;138:1741–53.
6. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
7. Carreras-Torres R, Johansson M, Gaborieau V, et al. The role of obesity, type 2 diabetes, and metabolic factors in pancreatic cancer: a Mendelian randomization study. *J Natl Cancer Inst* 2017;109:djx012.
8. Nead KT, Sharp SJ, Thompson DJ, et al. Evidence of a causal association between insulinemia and endometrial cancer: a Mendelian randomization analysis. *J Natl Cancer Inst* 2015;107:djv178.
9. Tsugane S, Sawada N. The JPHC study: design and some findings on the typical Japanese diet. *Jpn J Clin Oncol* 2014;44:777–82.
10. World Health Organization. *International classification of diseases for oncology*, 3rd edn. Geneva: World Health Organization, 2000. 240.
11. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
12. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010;42:105–16.
13. Rung J, Cauchi S, Albrechtsen A, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* 2009;41:1110–5.
14. Hanson RL, Muller YL, Kobes S, et al. A genome-wide association study in American Indians implicates DNER as a susceptibility locus for type 2 diabetes. *Diabetes* 2014;63:369–76.
15. Yamauchi T, Hara K, Maeda S, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 2010;42:864–8.
16. Cho YS, Chen CH, Hu C, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 2011;44:67–72.
17. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–6.
18. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007;39:770–5.
19. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40:638–45.
20. Ma RC, Hu C, Tam CH, et al. Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4. *Diabetologia* 2013;56:1291–305.
21. Hara K, Fujita H, Johnson TA, et al. Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet* 2014;23:239–46.
22. Imamura M, Maeda S, Yamauchi T, et al. A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Hum Mol Genet* 2012;21:3042–9.
23. Takeuchi F, Serizawa M, Yamamoto K, et al. Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* 2009;58:1690–9.
24. Yasuda K, Miyake K, Horikawa Y, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 2008;40:1092–7.
25. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;316:1341–5.
26. The Committee of the Japan Diabetes Society on the Diagnostic Criteria of Diabetes Mellitus, Seino Y, Nanjo K, et al. Report of the committee on the classification and diagnostic criteria of diabetes mellitus. *J Diabetes Investig* 2010;1:212–28.
27. Kato M, Noda M, Inoue M, et al. Psychological factors, coffee and risk of diabetes mellitus among middle-aged Japanese: a population-based prospective study in the JPHC study cohort. *Endocr J* 2009;56:459–68.
28. Kulathinal S, Karvanen J, Saarela O, et al. Case-cohort design in practice—experiences from the MORGAM project. *Epidemiol Perspect Innov* 2007;4:15.
29. Tanikawa C, Kamatani Y, Takahashi A, et al. GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. *Carcinogenesis* 2018;39:652–60.
30. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37:658–65.
31. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife* 2018;7:e34408.
32. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 2018;33:947–52.
33. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;16:309–30.
34. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016;32:3207–9.
35. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–4.
36. Prokopenko I, Poon W, Magi R, et al. A central role for GRB10 in regulation of islet function in man. *PLoS Genet* 2014;10:e1004235.
37. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–24.
38. Strawbridge RJ, Dupuis J, Prokopenko I, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 2011;60:2624–34.
39. Wen W, Zheng W, Okada Y, et al. Meta-analysis of genome-wide association studies in east Asian-ancestry populations identifies four new loci for body mass index. *Hum Mol Genet* 2014;23:5492–504.
40. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017;14:577–90.
41. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *Int J Epidemiol* 2015;44:512–25.