Am J Cancer Res 2021;11(8):3921-3934
www.ajcr.us /ISSN:2156-6976/ajcr0133760

Original Article

Body mass index and type 2 diabetes and breast cancer survival: a Mendelian randomization study

Yi-Shian Liu[1], Pei-Ei Wu[1], Wen-Cheng Chou[1], Rajeev Vikram[1], Wei-Ting Chen[1], Show-Ling Yang[1], Manjeet K Bolla[2], Qin Wang[2], Joe Dennis[2], Tsun L Chan[3,4], Ji-Yeob Choi[5,6,7], Ming-Feng Hou[8], Hidemi Ito[9,10], Daehee Kang[6,11], Sung-Won Kim[12], Ava Kwong[3,13,14], Keitaro Matsuo[10,15], Sue K Park[5,6,11], Xiao-Ou Shu[16], Wei Zheng[16], Alison M Dunning[2,17], Douglas F Easton[2,17], Chen-Yang Shen[1,18]

[1]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; [2]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; [3]Hong Kong Hereditary Breast Cancer Family Registry, Cancer Genetics Centre, Happy Valley, Hong Kong, China; [4]Department of Pathology, Hong Kong Sanatorium and Hospital, Happy Valley, Hong Kong, China; [5]Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea; [6]Cancer Research Institute, Seoul National University, Seoul, Korea; [7]Institute of Health Policy and Management, Seoul National University Medical Research Center, Seoul, Korea; [8]Department of Surgery, Kaohsiung Municipal Hsiao-Kang Hospital, Kaohsiung, Taiwan; [9]Division of Cancer Information and Control, Aichi Cancer Center Research Institute, Nagoya, Japan; [10]Division of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan; [11]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea; [12]Department of Surgery, Daerim Saint Mary's Hospital, Seoul, Korea; [13]Department of Surgery, The University of Hong Kong, Pok Fu Lam, Hong Kong, China; [14]Department of Surgery, Hong Kong Sanatorium and Hospital, Happy Valley, Hong Kong, China; [15]Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan; [16]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA; [17]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom; [18]College of Public Health, China Medical University, Taichung, Taiwan

Received March 31, 2021; Accepted July 10, 2021; Epub August 15, 2021; Published August 30, 2021

Abstract: The causal relationship between body mass index (BMI) and type 2 diabetes (T2D) and breast cancer prognosis is still ambiguous. The aim of this study was to investigate the prognostic effect of BMI and T2D on breast cancer disease-free survival (DFS) among Asian individuals. In this two-sample Mendelian randomization (MR) study, the instrumental variables (IVs) were identified using a genome-wide association study (GWAS) among 24,000 participants in the Taiwan Biobank. Importantly, the validity of these IVs was confirmed with a previous large-scale GWAS (Biobank Japan Project, BBJ). In this study, we found that a genetic predisposition toward higher BMI (as indicated by BMI IVs, F = 86.88) was associated with poor breast cancer DFS (hazard ratio [HR] = 6.11; P < 0.001). Furthermore, higher level of genetically predicted T2D (as indicated by T2D IVs) was associated with an increased risk of recurrence of and mortality from breast cancer (HR = 1.43; P < 0.001). Sensitivity analyses, including the weighted-median approach, MR-Egger regression, Radial regression and Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO) supported the consistency of our findings. Finally, the causal relationship between BMI and poor breast cancer prognosis was confirmed in a prospective cohort study. Our MR analyses demonstrated the causal relationship between the genetic prediction of elevated BMI and a greater risk of T2D with poor breast cancer prognosis. BMI and T2D have important clinical implications and may be used as prognostic indicators of breast cancer.

Keywords: Mendelian randomization, breast cancer progression, body mass index, type 2 diabetes, breast cancer survival analysis

Introduction

Breast cancer is the most frequent malignancy in women and causes ~627,000 deaths annually worldwide [1]. Because of its high incidence, the treatment for breast cancer has been well standardized based on specific subtypes, such as the estrogen receptor (ER) sta-

tus of the tumor [1]. However, breast cancer progression varies substantially, even among patients with the same subtype.

Diabetes and obesity have been suggested to be linked to breast cancer progression [2, 3]. Obese and diabetic breast cancer patients are more likely to have larger-size tumors, and they have a three times higher breast cancer recurrence rate than patients who were neither diabetic nor obese [2]. An increased BMI is associated with higher mortality and recurrence of breast cancer, particularly in postmenopausal women in observational studies [4-6]. Furthermore, diabetes mellitus may reduce breast cancer survival [7-9]. A multiethnic cohort study reported that breast cancer women who have had T2D for > 7 years have higher all-cause mortality than those who have had T2D for < 7 years [10]. Whether putative association between BMI and T2D and breast cancer progression is causal or subject to undetected confounding effects remains unclear. If it can be clarified, proper management of BMI and/or T2D may become an additional therapeutic means for improving breast cancer outcomes.

Unfortunately, conventional epidemiological observation studies are prone to being biased by reverse causal association and unmeasured confounding [11]. In contrast, Mendelian randomization (MR), which is based on the foundation of random assignment of alleles from parent to child, estimates the causal relationship between exposures and diseases outcomes, which can potentially overcome this limitation. There are three major assumptions when performing a MR analysis, which have been described in the Supplementary Material [12]. Our aim was to investigate the prognostic effect of BMI and T2D on breast cancer prognosis among Asian women. To validate our findings, a series of sensitivity analyses was performed. Additionally, the causal relationship between BMI and breast cancer progression was further confirmed in a prospective cohort study.

Methods

Study population

The study participants were from the Breast Cancer Association Consortium (BCAC) [13]. The present study was purposely conducted in a genetically homogenous population; as a result, we included only 8766 women of East Asian ancestry. Detailed descriptions about study participants are provided in the Supplementary Material and Table S1. Genotyping, imputation and genetic quality-control information are presented in the Supplementary Material. We first excluded subjects without complete vital information (dead or alive) and follow-up time. A total of 6642 participants were retained. All studies in BCAC have been approved by institutional review boards and all patients have provided written informed consent.

Selection of genetic instrument variables

We selected genetic IVs using a two-step approach. First, we conducted a genome-wide association study (GWAS) to identify exposure (i.e., BMI and T2D)-associated variants (SNPs) in 24,000 individuals aged 30-70 years from the Taiwan Biobank cohort [14] who had successfully passed the standard genetic quality-control. Second, we replicated these exposure-associated variants to ensure the sufficient strength of IVs by using the Biobank Japan Project (BBJ) [15], which is a publicly available large-scale GWAS. Only SNPs that achieved the genome-wide significance level ($P < 5 \times 10^{-8}$) with a minor allele frequency of > 0.01 were considered as valid IVs. Moreover, the *F* statistic was used to examine the precision and strength of the effect of BMI IVs on BMI measurement [16]. Conventionally, a threshold of *F* statistic > 10 has been considered to be adequate to avoid weak instrument bias [17]. However, the status of T2D (i.e., whether a patient had or did not have T2D) was not available in our studies, we could not perform an *F* statistic analysis for T2D. Summary-level statistics were extracted from the BBJ for female-specific GWASs for BMI (n = 82,438 women) and T2D (n = 102,386 women).

Mendelian randomization analyses

In our MR study, the effect estimates and corresponding SEs of IVs for exposure (BMI and T2D) were extracted from the female-specific GWASs in BBJ; the effect estimates and corresponding SEs of IVs and outcome (breast cancer DFS) were determined from multiple Cox proportional hazard models after adjusting for breast cancer prognostic factors, including age at diagnosis, tumor stage, ER status and genet-

ic admixture using the first two principal components of population structure (to account for genetic heterogeneity among Eastern-Asian women) in BCAC. These breast cancer prognostic factors were adjusted in all of our analyses. We also harmonized our data to ensure consistent directions for the IV-exposure and IV-outcome associations, as described in the Supplementary Material.

We used the inverse-variance weighted (IVW) method by Burgess et al. [18] to obtain MR estimates. To substantiate our analyses, we further performed a series of sensitivity analyses: weighted-median, MR-Egger regression, Radial regression and Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO). The IVW, MR-Egger regression and weighted median methods were conducted using the R package '*MendelianRandomization*'. Radial regression and MR-PRESSO were conducted using the R package '*RadialMR*' and '*MRPRESSO*', respectively. Moreover, we assessed the relationship between BMI and breast cancer DFS by stratifying by ER status as described in the Supplementary Material.

Sensitivity analyses

To account for potential pleiotropy effects that may lead to biased MR results, we carried out additional sensitivity analyses. MR-Egger regression was used to detect the presence of horizontal pleiotropy effects by examining if the intercept deviated significantly from zero [19]. The weighted median method was assumed to provide unbiased estimates of at least 50% of valid IVs [20]. Radial regression was used to detect influential outliers [21]. After correcting for pleiotropy effects by removing all influential outliers until little or no heterogeneity remained, we re-analyzed the MR results. Finally, the MR-PRESSO method [22] was examined whether there were any remaining pleiotropy effects.

Observational analyses for body mass index on the outcome of breast cancer progression

A prospective cohort study in the BCAC was conducted to evaluate the relationship of an increased BMI with reduced breast cancer DFS. We limited the follow-up period to 10 years because it is commonly used in clinical evaluation of breast cancer progression. According to the Asian-specific criteria for BMI as documented by the WHO recommendations [23], we categorized BMI as follows: underweight (BMI < 18.5 kg/m^2), normal weight (18.5 ≤ BMI < 23 kg/m^2), overweight (23 ≤ BMI ≤ 27.5 kg/m^2) and obese (BMI > 27.5 kg/m^2). We combined underweight and normal weight individuals into one group and overweight and obese individuals into a second group for analysis. Survival curves were conducted with the Kaplan-Meier method, and the survival probabilities were compared using the log-rank test between BMI subgroups by stratifying the ER status. A multiple Cox proportional hazard model was used to assess the relationship between BMI subgroups and breast cancer DFS after adjusting for breast cancer prognostic factors in ER-positive and ER-negative patients separately.

Results

Cohort characteristics

In our study, 467 events occurred among 6642 breast cancer women during the follow-up time (6-360 months). Characteristics of this group are shown in **Table 1**. Overall, age at diagnosis, tumor stage and ER status were associated with breast cancer DFS (P < 0.05) (**Table 2**). This is consistent with the current understanding of breast cancer progression, supporting the validity of our initial analysis.

Two-sample Mendelian randomization analysis

We identified 188 BMI-associated SNPs and 251 T2D-associated SNPs in the Taiwan Biobank cohort. Next, we evaluated whether these candidate IVs achieved genome-wide significance in the BBJ cohort and verified 33 and 86 SNPs as IVs for BMI and T2D, respectively. We subsequently excluded 21 pleiotropic outliers for T2D IVs by Radial regression, which showed substantial heterogeneity (**Figure 1**). Finally, we used 33 SNPs as BMI IVs and 45 SNPs as T2D IVs (Tables S2, S3). After data harmonization, the direction of the IVs was positively associated with increased BMI and a higher risk of having T2D.

The IVW method, which provides unbiased estimates when all IVs are valid [24], showed that a higher genetically predicted BMI (as indicated by IVs) was associated with poor breast cancer DFS (HR = 6.11 [95% confidence inter-

Table 1. Characteristics of the study population from the Breast Cancer Association Consortium

Characteristic	Value
Total number of patients	6642
Events, *n* (%)	467 (9.74)
Follow-up, months, median (range)	48 (6-360)
Age at diagnosis, years, mean (SD)	48.3 (11.15)
BMI, kg/m^2, mean (SD)	23.11 (3.52)
underweight (BMI < 18.5 kg/m^2), *n* (%)	317 (4.77)
normal (18.5 ≤ BMI < 23 kg/m^2), *n* (%)	2231 (33.59)
overweight (23 ≤ BMI ≤ 27.5 kg/m^2), *n* (%)	1833 (27.60)
obese (BMI > 27.5 kg/m^2), *n* (%)	515 (7.75)
ER status, *n* (%)	
Positive	4096 (61.66)
Negative	2133 (32.11)
Tumor stage, *n* (%)	
I	2705 (40.72)
II	2561 (38.56)
III	721 (10.86)
IV	77 (1.16)

Abbreviations: BMI: body mass index; ER: estrogen receptor; Tumor stage: tumor size-lymph node involvement-metastasis status of tumor stage.

Table 2. Multivariate Cox regression analyses among BMI and breast cancer prognostic factors among individuals from the Breast Cancer Association Consortium

Factor	HR (95% CI)	*P*
Age at diagnosis	1.013 (1.003-1.023)	0.0067
BMI	1.028 (1.000-1.059)	0.05
ER status	0.594 (0.482-0.778)	< 0.001
Tumor stage	2.646 (2.270-3.083)	< 0.001
PC1	1.456 (1.214-1.746)	< 0.001
PC2	0.662 (0.562-0.778)	< 0.001

Abbreviations: BMI: body mass index; ER: estrogen receptor; Tumor stage: tumor size-lymph node involvement-metastasis status; HR: hazard ratio; 95% CI: 95% confidence interval; PC1: the first principal component of population structure; PC2: the second principal component of population structure.

val (CI): 3.10-12.01]) (**Figure 2**). The *F* statistic for BMI was 86.88 after adjusting for the breast cancer prognostic factors, which strongly indicates the adequate strength of our BMI IVs. Moreover, we performed a stratified analysis according to the ER status of breast cancer. We identified 25 of 33 BMI IVs ($P < 0.05$) among ER-positive patients ($n = 4096$). In contrast, no BMI-associated SNPs were identified in ER-negative patients ($n = 2133$). This suggests that an elevated BMI is causally associated with an increased risk of breast cancer recurrence and mortality, especially among ER-positive patients (**Figure 3**; **Table 3**).

In terms of T2D and breast cancer progression, IVW analysis indicated an increased risk of having T2D with reduced breast cancer DFS (**Table 3**). In addition, the wide confidence intervals for the relationship between BMI and breast cancer DFS in IVW analysis may indicate uncertainty in the precision of the effect size. To verify these results, we used the pleiotropy robust methods (MR-PRESSO and IVW Radial), which provide more precise estimates. It is notable that the findings were consistent across these MR analyses (**Figure 2**). The scatter plots of MR estimates are shown in **Figure 4**.

Sensitivity analyses

Weighted-median analyses showed similar causal effect estimates and were in the same direction as the IVW analyses for BMI and T2D (**Figure 2**). For the T2D analysis, the intercept of the MR-Egger regression deviated from zero ($P = 0.001$), revealing the presence of modest pleiotropy (**Table 3**). Furthermore, after removing pleiotropy outliers among the T2D IVs identified by the Radial regression, we observed a slightly stronger causal estimate for genetically higher T2D risk on breast cancer DFS (before removing the outliers: HR = 1.23 [95% CI: 1.10-1.38]; after removing the outliers: HR = 1.43 [95% CI: 1.24-1.65]) (**Table 3**). In contrast, we found no evidence of horizontal pleiotropy among BMI IVs.

Observational analyses of body mass index and breast cancer disease-free survival

Taking advantage of the availability of information for both BMI and the follow-up period as collected by individual studies in the BCAC, we were able to confirm our MR findings by a prospective cohort study. Survival curves indicated that breast cancer patients who were underweight or of normal weight were more likely to experience longer survival times relative to those who were overweight or obese in ER-positive patients (**Figure 5**). The DFS probability for breast cancer differed significantly between BMI subgroups as determined by the

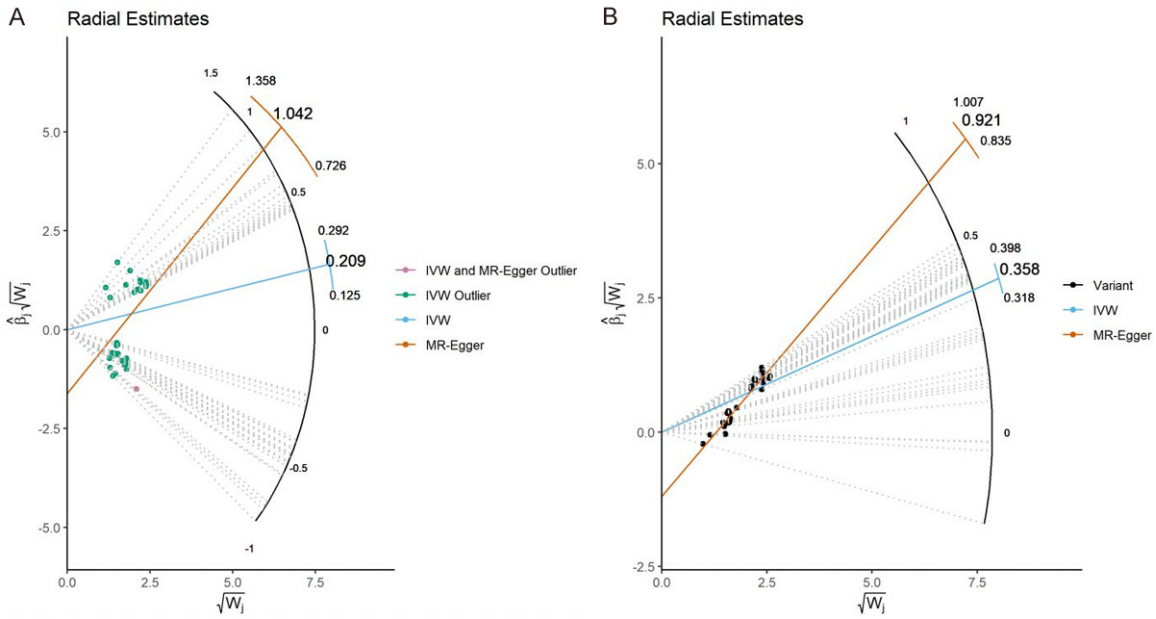

Figure 1. The effect of removing pleiotropy outliers from the T2D IVs. (A, B) Graphs of radial estimates before (A) and after (B) removing pleiotropy outliers from the T2D IVs. In the radial plots, the absolute vertical distance of each variant from the slope is equal to the square root of its contribution to heterogeneity with respect to Cochran's Q statistic. The horizontal axis of the radial plots is the square root of the actual weight from the IVW analysis of each SNP. Its vertical axis scale refers to the ratio estimate for each SNP multiplied by the same square root weight.

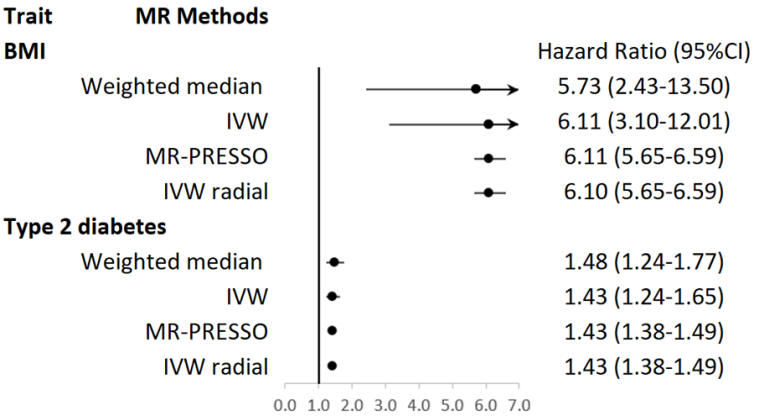

Figure 2. Mendelian randomization (MR) analyses testing the prognostic effect of body mass index (BMI) and type 2 diabetes (T2D) on breast cancer. Results from four MR methods-weighted median, IVW (inverse-variance weighted), MR-PRESSO (Mendelian randomization pleiotropy residual sum and outlier) and IVW Radial-are presented as hazard ratios with 95% confidence intervals (95% CIs).

log-rank test in ER-positive patients (P < 0.001), but this was not the case for ER-negative patients. A multivariate Cox proportional hazard model demonstrated the relationship between increased BMI and poor breast cancer DFS (HR = 1.42 [95% CI: 1.05-1.92]) among ER-positive patients (**Figure 5**). Both MR and observational analyses supported the existence of an adverse effect of an increased BMI on breast cancer progression.

Discussion

To date, this is the first two-sample MR analysis among Asian that attempted to clarify possible metabolic factors (i.e., BMI and T2D) related to breast cancer prognosis. This is also the first MR analysis using both data from an individual GWAS and from a GWAS consortium to select IVs. This careful approach resulted in IVs with high confidence.

Consistent with our findings, a MR study conducted using genetic risk scores based on BMI-associated SNPs in European suggested the causal effect of elevated BMI on reduced breast cancer survival among ER-positive patients, but not found in ER-negative patients [25]. A higher BMI was associated with higher breast cancer mortality and recurrence across different ethnicities [4-6]. Breast cancer patients who were

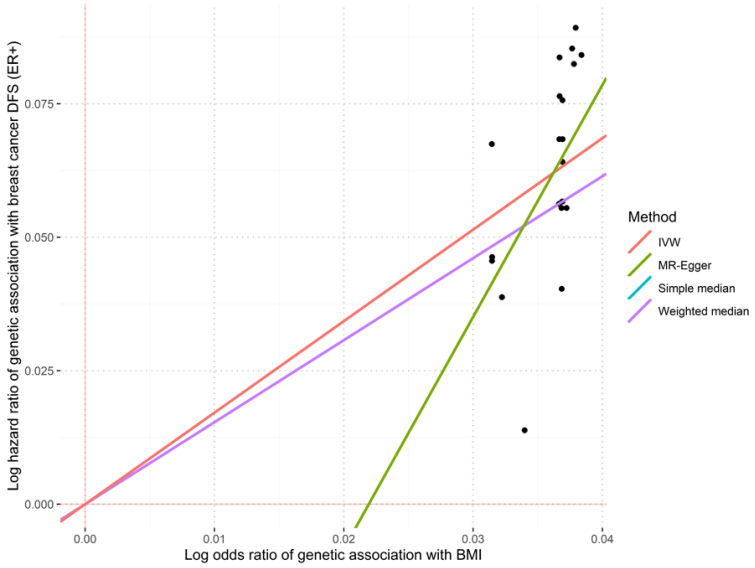

Figure 3. Results of MR analyses testing the prognostic effect of BMI on breast cancer in patients with estrogen receptor-positive (ER+) cancer. Scatter plot of the log odds ratio of genetic association with BMI in ER-positive cancer versus the log hazard ratio of genetic associations with breast cancer disease-free survival (DFS). The line of simple median is almost overlap with the line of weighted median.

overweight during early adulthood as compared with patients who were never overweight had a risk of early death from breast cancer increased by 3% [26]. Similarly, a Japanese study showed that obese patients had reduced breast cancer DFS relative to the individuals with normal range BMIs, especially among postmenopausal women with ER-positive tumors [27]. Biological mechanisms underlying the connection between obesity and breast cancer risk/recurrence have been proposed, including insulin resistance, increased inflammatory cytokines, higher leptin levels and adipokine imbalances [28].

Previous studies have reported that being overweight on breast cancer prognosis may relate to menopausal status and ER status. We found that elevated BMI was associated with reduced breast cancer DFS among ER-positive patients, consistent with previous findings among Europeans [25, 27]. Due to the absence of menopausal status data, we could not stratify by it. It is notable that breast cancer in Asians is intrinsically and etiologically different from that in Westerns [29, 30]. Further investigation of how BMI affects breast cancer progression after considering the effect of menopausal status and different ethnicities simultaneously in larger samples is warranted.

Patients who had been diagnosed with pre-existing diabetes before breast cancer compared with those who did not have pre-existing diabetes were linked to poor breast cancer DFS [31]. Diabetic breast cancer patients had significantly larger tumors, a higher rate of lymph node involvement and higher overall mortality as compared with non-diabetic breast cancer patients [32]. There are putative shared metabolic changes between the progression toward diabetes and breast cancer that include dyslipidemia, hyperinsulinemia and hyperglycemia [33]. These shared metabolic mechanisms, which include insulin resistance [34], the activation of transcription factors (such as of NF-κB [35], STAT3 [36] and HIF1α [37]) and related mitogenic and angiogenic signalling mechanisms that may cause the acceleration of breast cancer progression (such as invasion and metastasis of breast cancer and increased breast cancer cell survival, proliferation and migration [7]), supported our findings that having T2D is causally linked to breast cancer progression.

Conventional epidemiological studies are likely to biases by unmeasured confounding factors, different error types (e.g., measurement error) and uncertainty of causality. MR overcomes these limitations and should provide more reliable causal estimates. The causal effect estimates of increased BMI and reduced breast cancer DFS differed from the observational analyses and MR analyses in our study. We suggest that the effect size estimates from our observational analyses are a measure of the total effect of BMI, which may be influenced by environmental and/or lifestyle factors. In contrast, the MR analysis reflected the direct effect of genetic predisposition of BMI on breast cancer progression. The main limitation of this study is the inability to control for the treatment of breast cancer patients, a

Table 3. Mendelian randomization (MR) analyses testing the prognostic effect of body mass index (BMI) and type 2 diabetes (T2D) on breast cancer

Trait	Method	Number of IVs	Estimate	HR (95% CI)	*P*
BMI	Weighted median	33	1.745	5.73 (2.43-13.50)	6.6E-05
BMI	IVW	33	1.809	6.11 (3.10-12.01)	1.6E-07
BMI	Intercept (from MR-Egger)	33	0.055		0.72995
BMI	MR-PRESSO	33	1.890	6.62 (6.13-7.15)	8.2E-31
BMI	IVW Radial	33	1.815	6.14 (5.68-6.64)	0
BMI (ER+)	Weighted median	25	1.536	4.65 (1.28-16.92)	0.01983
BMI (ER+)	IVW	25	1.714	5.55 (1.96-15.70)	0.00123
BMI (ER+)	Intercept (from MR-Egger)	25	-0.095		0.77704
BMI (ER+)	MR-PRESSO	25	1.756	5.79 (5.37-6.24)	6.1E-25
BMI (ER+)	IVW Radial	25	1.755	5.78 (5.37-6.23)	0
T2D	Weighted median	86	0.389	1.48 (1.27-1.72)	6.1E-07
T2D	IVW	86	0.209	1.23 (1.10-1.38)	0.00024
T2D	Intercept (from MR-Egger)	86	-0.127		0.00102
T2D	MR-PRESSO	86	0.209	1.23 (1.14-1.58)	3.42E-06
T2D	IVW Radial	86	0.210	1.23 (1.14-1.34)	2.45E-06
T2D-outlier removed	Weighted median	45	0.394	1.48 (1.24-1.77)	1.7E-05
T2D-outlier removed	IVW	45	0.358	1.43 (1.24-1.65)	6.3E-07
T2D-outlier removed	Intercept (from MR-Egger)	45	-0.103		0.12116
T2D-outlier removed	MR-PRESSO	45	0.358	1.43 (1.38-1.49)	5.4E-22
T2D-outlier removed	IVW Radial	45	0.358	1.43 (1.37-1.49)	9.3E-73

Abbreviations: BMI: body mass index; ER: estrogen receptor; ER+: patients with estrogen receptor positive tumor; HR: hazard ratio; 95% CI: 95% confidence interval; IVW: inverse-variance weighted; MR-PRESSO: Mendelian randomization pleiotropy residual sum and outlier; IVW Radial: inverse-variance weighted Radial regression.

potential confounder, because of incomplete data for our study participants. As breast cancer is the most common female cancer in the world, treatment protocols based on different subtypes and tumor stages have been well-standardized. Therefore, patients included in this study should have received similar treatments. Consequently, the inability to control for treatment in our analyses is non-differential and should tend to be biased toward the null, and, therefore, the estimation of the relationship between exposure (BMI and T2D) and breast cancer progression would be conservative. As a second limitation, the IVs of our study may have missed some information from important SNPs. For instance, we did not include the SNPs in the gene *FTO* (a well-known gene that affects BMI) in our IVs. Because the BCAC population was genotyped using iCOGS and OncoArray, both of which are specifically designed for studying hormone-related cancers, including breast, ovarian and prostate cancer and so on [38, 39], information for other relevant SNPs may have been lost. However, the F statistic for BMI was 86.88, which strongly indicates the adequate strength of our IVs. The third limitation is an absence of T2D status data for our study participants. Without this information, we could not examine the strength of T2D IVs with the *F* statistic [17] and could not perform the survival analysis for breast cancer and T2D. We utilized the external weights from the BBJ, which ensured the representativeness of these IVs in our study population and provided more precise effect sizes of exposure owing to the large population. A possible fourth limitation, the small sample size may have resulted in inadequate statistical power to identify BMI-associated SNPs in ER-negative breast cancer patients. However, our MR results are consistent with the previous finding [25]. As a fifth limitation, we didn't examine whether BMI and T2D was one of the risk factor of breast cancer. Our study population is all breast cancer patients. To recruit the non-breast cancer women as the controls is needed for further examination to address this important ques-

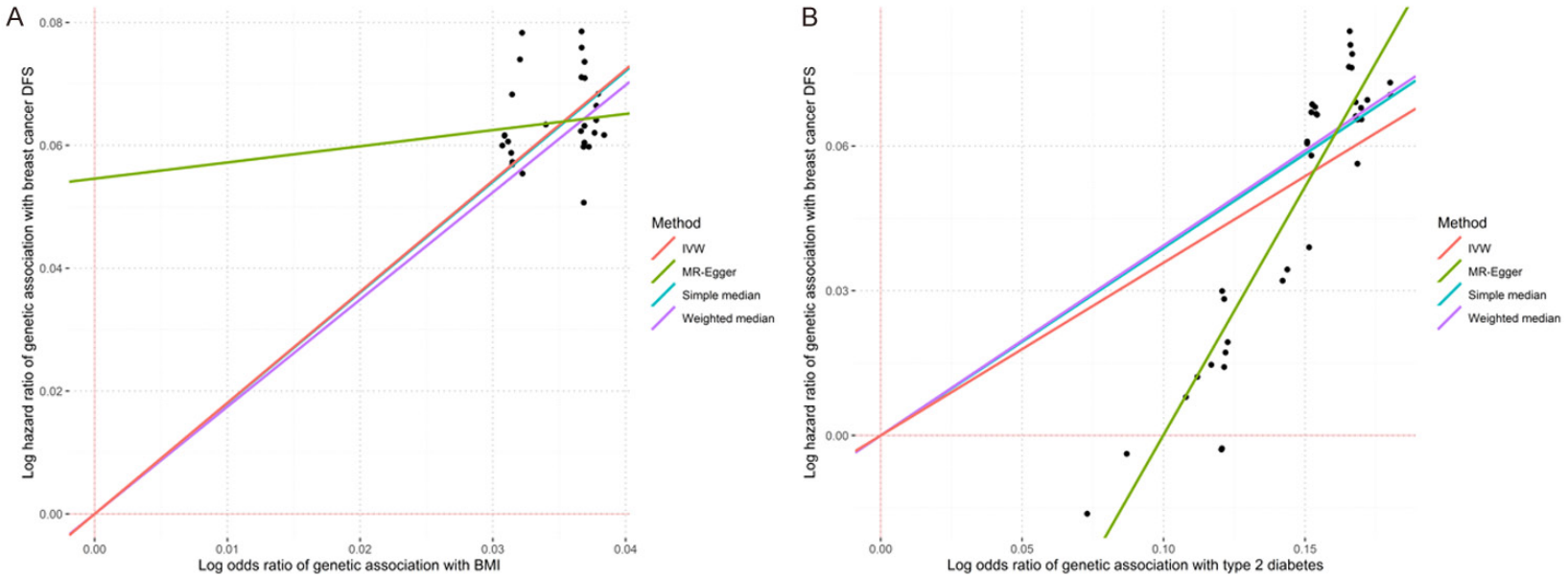

Figure 4. Results of MR analyses testing the breast cancer prognostic effect of BMI and type 2 diabetes. A. Scatter plot of the log odds ratio of genetic association with BMI versus the log hazard ratio of genetic association with breast cancer disease-free survival (DFS). The line of simple median is almost overlap with the line of IVW for BMI. B. Scatter plot of the log odds ratio of genetic association with type 2 diabetes versus the log hazard ratio of genetic association with breast cancer DFS. The line of simple median is almost overlap with the line of weighted median for T2D.

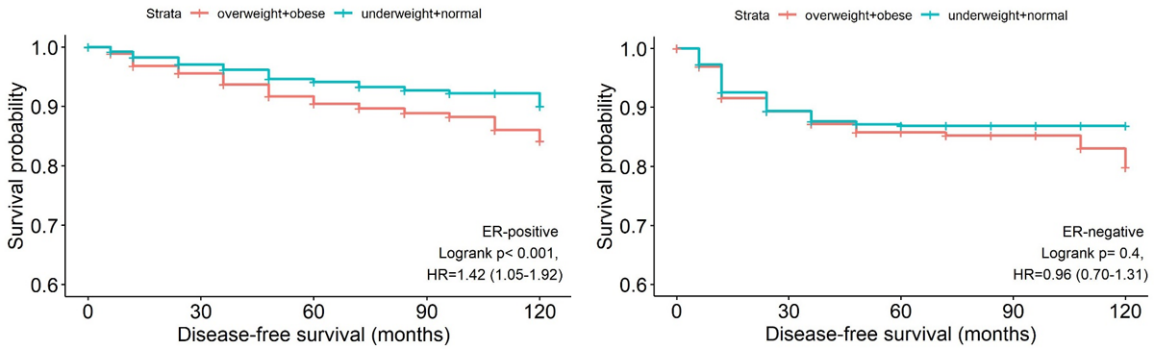

Figure 5. Kaplan-Meier survival curves and the survival probability for different BMI subgroups with a 10-year follow-up. For disease-free survival (DFS) analysis, the follow-up time for each patient was calculated from the date of diagnosis to the date of death or breast cancer-specific recurrence, the date censored or the date of last follow-up. Survival probabilities were compared using the log-rank test.

tion. Finally, it must be mentioned that several genetic variants in our study were not directly genotyped. However, we imputed only variants with high information quality scores (iCOGS: mean r^2 = 0.97, range = 0.61-1.00; OncoArray: mean r^2 = 0.99, range = 0.88-1.00) to reduce the false probability rate.

Our study is novel in that it investigated potential prognostic factors for breast cancer in Asia, which are distinct from those associated with breast cancer among Western populations. Studies have consistently confirmed a striking difference in the age-specific incidence of Western and Asian breast cancer. Specifically, early onset (< 50 years old) breast cancer has increased markedly in Asia as compared with Western countries [40, 41]. The average age of subjects was 48.30 years in the present study, which reflects the early onset of breast cancer in Asian women. Approximately half of Asian breast cancer patients are premenopausal women, whereas only 15-30% of Western women with breast cancer are premenopausal [42]. Discrepancies in the pathophysiologic factors and biological mechanisms of breast cancer between Asian and Western women have attracted great attention in recent years. Breast cancer is prevalent in Western women who are 60 years or older, and the incidence rate of breast cancer increases among Western women aged 50 to 84 years. Apart from this, there was an upward trend in breast cancer incidence for Asian women before the age of 40 years and a concave downward trend among Asian women aged 50 years and older [29]. Recently, multi-omics studies have highlighted differences in the molecular signature of the tumor immune microenvironment between Asian and Western patients [29, 30]. Asian patients harbor more tumor-infiltrating immune cells than do Western patients [43], which may be linked to a favorable prognosis of breast cancer [30, 44, 45]. However, the composition of infiltrating immune cells did not significantly differ between Asian and Western breast cancer patients [30]. In addition, the relatively high probability of ER-positive breast cancer among Asian women who are < 50 years old as compared with Western women suggests the importance of hormone exposure [46, 47]. In summary, population-specific differences in the biology possibly indicate different environmental variations, molecular signatures and tumorigenic mechanisms underlying breast cancer onset and prognostic patterns.

Given strong causal evidence from MR, we suggest that BMI and T2D may be clinicopathological breast cancer progression predictors. The pathological mechanisms between BMI and T2D and breast cancer progression are certainly worthy of future exploration.

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. The COGS study would not have been possible without the contributions of the following: Per Hall (COGS); Douglas F. Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Douglas F. Easton, Ali Amin

Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. HKBCS thanks Hong Kong Sanatorium and Hospital, Dr Ellen Li Charitable Foundation, The Kerry Group Kuok Foundation, National Institute of Health 1R03CA130065 and the North California Cancer Center for support. We thank all investigators of the KOHBRA (Korean Hereditary Breast Cancer) Study. SBCGS thank study participants and research staff for their contributions and commitment to the studies. BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community's Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C12-87/A10118, C1287/A10710, C12292/A111-74, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112-the GAMEON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065. The HERPACC was supported by MEXT Kakenhi (No. 1701-50181 and 26253041) from the Ministry of Education, Science, Sports, Culture and Technology of Japan, by a Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from Ministry Health, Labour and Welfare of Japan, by Health and Labour Sciences Research Grants for Research on Applying Health Technology from Ministry Health, Labour and Welfare of Japan, by National Cancer Center Research and Development Fund, and "Practical Research for Innovative Cancer Control (15ck0106177h-0001, 20ck0106553h001)" from Japan Agency for Medical Research and development, AMED, and Cancer Bio Bank Aichi. The KOHBRA study was partially supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (HI16-C1127; 1020350; 1420190). The SBCGS was supported primarily by NIH grants R01CA642-77, R01CA148667, UMCA182910, and R37-CA70867. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The scientific development and funding of this project were, in part, supported by the Genetic Associations and Mechanisms in Oncology (GAME-ON) Network U19 CA148065. SEBCS was supported by the BRL (Basic Research Laboratory) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2012-0000347). The TWBCS is supported by the Taiwan Biobank project of the Institute of Biomedical Sciences,

Academia Sinica, Taiwan. This work was supported by grant from Ministry of Science and Technology, Taiwan (MOST 110-2314-B-001-005).

Disclosure of conflict of interest

None.

Address correspondence to: Chen-Yang Shen, Institute of Biomedical Sciences, Academia Sinica, 128 Academia Road, Sec. 2, Taipei, Taiwan. Tel: +886-2-27899036; E-mail: bmcys@ibms.sinica.edu.tw

References

[1] Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, Ruddy K, Tsang J and Cardoso F. Breast cancer. Nat Rev Dis Primers 2019; 5: 66.

[2] Buono G, Crispo A, Giuliano M, De Angelis C, Schettini F, Forestieri V, Lauria R, Pensabene M, De Laurentiis M, Augustin LSA, Amore A, D'Aiuto M, Tortoriello R, Accurso A, Cavalcanti E, Botti G, Montella M, De Placido S and Arpino G. Combined effect of obesity and diabetes on early breast cancer outcome: a prospective observational study. Oncotarget 2017; 8: 115709-115717.

[3] Goodwin PJ, Ennis M, Pritchard KI, Trudeau ME, Koo J, Taylor SK and Hood N. Insulin- and obesity-related variables in early-stage breast cancer: correlations and time course of prognostic associations. J Clin Oncol 2012; 30: 164-171.

[4] Kroenke CH, Chen WY, Rosner B and Holmes MD. Weight, weight gain, and survival after breast cancer diagnosis. J Clin Oncol 2005; 23: 1370-1378.

[5] Chan DS, Vieira AR, Aune D, Bandera EV, Greenwood DC, McTiernan A, Navarro Rosenblatt D, Thune I, Vieira R and Norat T. Body mass index and survival in women with breast cancer-systematic literature review and meta-analysis of 82 follow-up studies. Ann Oncol 2014; 25: 1901-1914.

[6] Loi S, Milne RL, Friedlander ML, McCredie MR, Giles GG, Hopper JL and Phillips KA. Obesity and outcomes in premenopausal and postmenopausal breast cancer. Cancer Epidemiol Biomarkers Prev 2005; 14: 1686-1691.

[7] Samuel SM, Varghese E, Varghese S and Büsselberg D. Challenges and perspectives in the treatment of diabetes associated breast cancer. Cancer Treat Rev 2018; 70: 98-111.

[8] Okon EB, Chung AW, Zhang H, Laher I and van Breemen C. Hyperglycemia and hyperlipidemia are associated with endothelial dysfunction during the development of type 2 diabetes. Can J Physiol Pharmacol 2007; 85: 562-567.

[9] Rodrigues Dos Santos C, Fonseca I, Dias S and Mendes de Almeida JC. Plasma level of LDL-cholesterol at diagnosis is a predictor factor of breast tumor progression. BMC Cancer 2014; 14: 132.

[10] Maskarinec G, Shvetsov YB, Conroy SM, Haiman CA, Setiawan VW and Le Marchand L. Type 2 diabetes as a predictor of survival among breast cancer patients: the multiethnic cohort. Breast Cancer Res Treat 2019; 173: 637-645.

[11] Fewell Z, Davey Smith G and Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am J Epidemiol 2007; 166: 646-655.

[12] Lawlor DA, Harbord RM, Sterne JA, Timpson N and Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008; 27: 1133-1163.

[13] Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb D, Rostamianfar A, Bolla MK, Wang Q, Tyrer J, Dicks E, Lee A, Wang Z, Allen J, Keeman R, Eilber U, French JD, Qing Chen X, Fachal L, McCue K, McCart Reed AE, Ghoussaini M, Carroll JS, Jiang X, Finucane H, Adams M, Adank MA, Ahsan H, Aittomäki K, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Arun B, Auer PL, Bacot F, Barrdahl M, Baynes C, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Børresen-Dale AL, Brand JS, Brauch H, Brennan P, Brenner H, Brinton L, Broberg P, Brock IW, Broeks A, Brooks-Wilson A, Brucker SY, Brüning T, Burwinkel B, Butterbach K, Cai Q, Cai H, Caldés T, Canzian F, Carracedo A, Carter BD, Castelao JE, Chan TL, David Cheng TY, Seng Chia K, Choi JY, Christiansen H, Clarke CL; NBCS Collaborators, Collée M, Conroy DM, Cordina-Duverger E, Cornelissen S, Cox DG, Cox A, Cross SS, Cunningham JM, Czene K, Daly MB, Devilee P, Doheny KF, Dörk T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Elvira M, Engel C, Eriksson M, Fasching PA, Figueroa J, Flesch-Janys D, Fletcher O, Flyger H, Fritschi L, Gaborieau V, Gabrielson M, Gago-Dominguez M, Gao YT, Gapstur SM, García-Sáenz JA, Gaudet MM, Georgoulias V, Giles GG, Glendon G, Goldberg MS, Goldgar DE, González-Neira A, Grenaker Alnæs GI, Grip M, Gronwald J, Grundy A, Guénel P, Haeberle L, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hamel N, Hankinson S, Harrington P, Hart SN, Hartikainen JM, Hartman M, Hein A, Heyworth J, Hicks B, Hillemanns P, Ho DN, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Hou MF, Hsiung CN,

Huang G, Humphreys K, Ishiguro J, Ito H, Iwasaki M, Iwata H, Jakubowska A, Janni W, John EM, Johnson N, Jones K, Jones M, Jukkola-Vuorinen A, Kaaks R, Kabisch M, Kaczmarek K, Kang D, Kasuga Y, Kerin MJ, Khan S, Khusnutdinova E, Kiiski JI, Kim SW, Knight JA, Kosma VM, Kristensen VN, Krüger U, Kwong A, Lambrechts D, Le Marchand L, Lee E, Lee MH, Lee JW, Neng Lee C, Lejbkowicz F, Li J, Lilyquist J, Lindblom A, Lissowska J, Lo WY, Loibl S, Long J, Lophatananon A, Lubinski J, Luccarini C, Lux MP, Ma ESK, MacInnis RJ, Maishman T, Makalic E, Malone KE, Kostovska IM, Mannermaa A, Manoukian S, Manson JE, Margolin S, Mariapun S, Martinez ME, Matsuo K, Mavroudis D, McKay J, McLean C, Meijers-Heijboer H, Meindl A, Menéndez P, Menon U, Meyer J, Miao H, Miller N, Taib NAM, Muir K, Mulligan AM, Mulot C, Neuhausen SL, Nevanlinna H, Neven P, Nielsen SF, Noh DY, Nordestgaard BG, Norman A, Olopade OI, Olson JE, Olsson H, Olswold C, Orr N, Pankratz VS, Park SK, Park-Simon TW, Lloyd R, Perez JIA, Peterlongo P, Peto J, Phillips KA, Pinchev M, Plaseska-Karanfilska D, Prentice R, Presneau N, Prokofyeva D, Pugh E, Pylkäs K, Rack B, Radice P, Rahman N, Rennert G, Rennert HS, Rhenius V, Romero A, Romm J, Ruddy KJ, Rüdiger T, Rudolph A, Ruebner M, Rutgers EJT, Saloustros E, Sandler DP, Sangrajrang S, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schürmann P, Scott RJ, Scott C, Seal S, Seynaeve C, Shah M, Sharma P, Shen CY, Sheng G, Sherman ME, Shrubsole MJ, Shu XO, Smeets A, Sohn C, Southey MC, Spinelli JJ, Stegmaier C, Stewart-Brown S, Stone J, Stram DO, Surowy H, Swerdlow A, Tamimi R, Taylor JA, Tengström M, Teo SH, Beth Terry M, Tessier DC, Thanasitthichai S, Thöne K, Tollenaar RAEM, Tomlinson I, Tong L, Torres D, Truong T, Tseng CC, Tsugane S, Ulmer HU, Ursin G, Untch M, Vachon C, van Asperen CJ, Van Den Berg D, van den Ouweland AMW, van der Kolk L, van der Luijt RB, Vincent D, Vollenweider J, Waisfisz Q, Wang-Gohrke S, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Wu AH, Xia L, Yamaji T, Yang XR, Har Yip C, Yoo KY, Yu JC, Zheng W, Zheng Y, Zhu B, Ziogas A, Ziv E; ABCTB Investigators; ConFab/AOCS Investigators, Lakhani SR, Antoniou AC, Droit A, Andrulis IL, Amos CI, Couch FJ, Pharoah PDP, Chang-Claude J, Hall P, Hunter DJ, Milne RL, García-Closas M, Schmidt MK, Chanock SJ, Dunning AM, Edwards SL, Bader GD, Chenevix-Trench G, Simard J, Kraft P and Easton DF. Association analysis identifies 65 new breast cancer risk loci. Nature 2017; 551: 92-94.

[14] Chen CH, Yang JH, Chiang CWK, Hsiung CN, Wu PE, Chang LC, Chu HW, Chang J, Song IW, Yang SL, Chen YT, Liu FT and Shen CY. Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan Biobank project. Hum Mol Genet 2016; 25: 5321-5331.

[15] Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakaue S, Matoba N, Low SK, Okada Y, Terao C, Amariuta T, Gazal S, Kochi Y, Horikoshi M, Suzuki K, Ito K, Koyama S, Ozaki K, Niida S, Sakata Y, Sakata Y, Kohno T, Shiraishi K, Momozawa Y, Hirata M, Matsuda K, Ikeda M, Iwata N, Ikegawa S, Kou I, Tanaka T, Nakagawa H, Suzuki A, Hirota T, Tamari M, Chayama K, Miki D, Mori M, Nagayama S, Daigo Y, Miki Y, Katagiri T, Ogawa O, Obara W, Ito H, Yoshida T, Imoto I, Takahashi T, Tanikawa C, Suzuki T, Sinozaki N, Minami S, Yamaguchi H, Asai S, Takahashi Y, Yamaji K, Takahashi K, Fujioka T, Takata R, Yanai H, Masumoto A, Koretsune Y, Kutsumi H, Higashiyama M, Murayama S, Minegishi N, Suzuki K, Tanno K, Shimizu A, Yamaji T, Iwasaki M, Sawada N, Uemura H, Tanaka K, Naito M, Sasaki M, Wakai K, Tsugane S, Yamamoto M, Yamamoto K, Murakami Y, Nakamura Y, Raychaudhuri S, Inazawa J, Yamauchi T, Kadowaki T, Kubo M and Kamatani Y. Large scale genome-wide association study in a Japanese population identified novel susceptibility loci for across diseases. Nat Genet 2020; 52: 669-679.

[16] Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, Davey Smith G and Sterne JA. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res 2012; 21: 223-242.

[17] Burgess S and Thompson SG; CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol 2011; 40: 755-764.

[18] Bowden J, Davey Smith G, Haycock PC and Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol 2016; 40: 304-314.

[19] Burgess S and Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol 2017; 32: 377-389.

[20] Bowden J, Davey Smith G and Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015; 44: 512-525.

[21] Bowden J, Spiller W, Del Greco MF, Sheehan N, Thompson J, Minelli C and Davey Smith G. Im-

proving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. Int J Epidemiol 2018; 47: 1264-1278.

[22] Verbanck M, Chen CY, Neale B and Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018; 50: 693-698.

[23] WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004; 363: 157-163.

[24] Burgess S, Butterworth A and Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013; 37: 658-665.

[25] Guo Q, Burgess S, Turman C, Bolla MK, Wang Q, Lush M, Abraham J, Aittomäki K, Andrulis IL, Apicella C, Arndt V, Barrdahl M, Benitez J, Berg CD, Blomqvist C, Bojesen SE, Bonanni B, Brand JS, Brenner H, Broeks A, Burwinkel B, Caldas C, Campa D, Canzian F, Chang-Claude J, Chanock SJ, Chin SF, Couch FJ, Cox A, Cross SS, Cybulski C, Czene K, Darabi H, Devilee P, Diver WR, Dunning AM, Earl HM, Eccles DM, Ekici AB, Eriksson M, Evans DG, Fasching PA, Figueroa J, Flesch-Janys D, Flyger H, Gapstur SM, Gaudet MM, Giles GG, Glendon G, Grip M, Gronwald J, Haeberle L, Haiman CA, Hall P, Hamann U, Hankinson S, Hartikainen JM, Hein A, Hiller L, Hogervorst FB, Holleczek B, Hooning MJ, Hoover RN, Humphreys K, Hunter DJ, Hüsing A, Jakubowska A, Jukkola-Vuorinen A, Kaaks R, Kabisch M, Kataja V; kConFab/AOCS Investigators, Knight JA, Koppert LB, Kosma VM, Kristensen VN, Lambrechts D, Le Marchand L, Li J, Lindblom A, Lindström S, Lissowska J, Lubinski J, Machiela MJ, Mannermaa A, Manoukian S, Margolin S, Marme F, Martens JWM, McLean C, Menéndez P, Milne RL, Marie Mulligan A, Muranen TA, Nevanlinna H, Neven P, Nielsen SF, Nordestgaard BG, Olson JE, Perez JIA, Peterlongo P, Phillips KA, Poole CJ, Pylkäs K, Radice P, Rahman N, Rüdiger T, Rudolph A, Sawyer EJ, Schumacher F, Seibold P, Seynaeve C, Shah M, Smeets A, Southey MC, Tollenaar RAEM, Tomlinson I, Tsimiklis H, Ulmer HU, Vachon C, van den Ouweland AMW, Van't Veer LJ, Wildiers H, Willett W, Winqvist R, Zamora MP, Chenevix-Trench G, Dörk T, Easton DF, García-Closas M, Kraft P, Hopper JL, Zheng W, Schmidt MK and Pharoah PDP. Body mass index and breast cancer survival: a Mendelian randomization analysis. Int J Epidemiol 2017; 46: 1814-1822.

[26] Arnold M, Charvat H, Freisling H, Noh H, Adami HO, Soerjomataram I and Weiderpass E. Adult overweight and survival from breast and colorectal cancer in Swedish women. Cancer Epidemiol Biomarkers Prev 2019; 28: 1518-1524.

[27] Gondo N, Sawaki M, Hattori M, Yoshimura A, Kotani H, Adachi Y, Kataoka A, Sugino K, Mori M, Horisawa N, Terada M, Ozaki Y and Iwata H. Impact of BMI for clinical outcomes in Japanese breast cancer patients. Jpn J Clin Oncol 2020; 50: 230-240.

[28] Li YR, Ro V and Tchou JC. Obesity, metabolic syndrome, and breast cancer: from prevention to intervention. Curr Surg Rep 2018; 6: 7.

[29] Lin CH, Yap YS, Lee KH, Im SA, Naito Y, Yeo W, Ueno T, Kwong A, Li H, Huang SM, Leung R, Han W, Tan B, Hu FC, Huang CS, Cheng AL and Lu YS; Asian Breast Cancer Cooperative Group. Contrasting epidemiology and clinicopathology of female breast cancer in Asians vs the US population. J Natl Cancer Inst 2019; 111: 1298-1306.

[30] Chen CH, Lu YS, Cheng AL, Huang CS, Kuo WH, Wang MY, Chao M, Chen IC, Kuo CW, Lu TP and Lin CH. Disparity in tumor immune microenvironment of breast cancer and prognostic impact: Asian versus Western populations. Oncologist 2020; 25: e16-e23.

[31] Zhao XB and Ren GS. Diabetes mellitus and prognosis in women with breast cancer: a systematic review and meta-analysis. Medicine (Baltimore) 2016; 95: e5602.

[32] Schrauder MG, Fasching PA, Haberle L, Lux MP, Rauh C, Hein A, Bayer CM, Heusinger K, Hartmann A, Strehl JD, Wachter DL, Schulz-Wendtland R, Adamietz B, Beckmann MW and Loehberg CR. Diabetes and prognosis in a breast cancer cohort. J Cancer Res Clin Oncol 2011; 137: 975-983.

[33] Kang C, LeRoith D and Gallagher EJ. Diabetes, obesity, and breast cancer. Endocrinology 2018; 159: 3801-3812.

[34] Ferroni P, Riondino S, Buonomo O, Palmirotta R, Guadagni F and Roselli M. Type 2 diabetes and breast cancer: the interplay between impaired glucose metabolism and oxidant stress. Oxid Med Cell Longev 2015; 2015: 183928.

[35] Shostak K and Chariot A. NF-κB, stem cells and breast cancer: the links get stronger. Breast Cancer Res 2011; 13: 214.

[36] Sharaf H, Matou-Nasri S, Wang Q, Rabhan Z, Al-Eidi H, Al Abdulrahman A and Ahmed N. Advanced glycation endproducts increase proliferation, migration and invasion of the breast cancer cell line MDA-MB-231. Biochim Biophys Acta 2015; 1852: 429-441.

[37] Wisastra R and Dekker FJ. Inflammation, cancer and oxidative lipoxygenase activity are intimately linked. Cancers (Basel) 2014; 6: 1500-1521.

[38] Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, Casey G, Hunter DJ, Sellers TA, Gruber SB, Dunning AM, Michailidou K, Fachal L, Doheny K, Spurdle AB, Li Y, Xiao X, Romm J, Pugh E, Coetzee GA, Hazelett DJ, Bojesen SE, Caga-Anan C, Haiman CA, Kamal A, Luccarini C, Tessier D, Vincent D, Bacot F, Van Den Berg DJ, Nelson S, Demetriades S, Goldgar DE, Couch FJ, Forman JL, Giles GG, Conti DV, Bickeböller H, Risch A, Waldenberger M, Brüske-Hohlfeld I, Hicks BD, Ling H, McGuffog L, Lee A, Kuchenbaecker K, Soucy P, Manz J, Cunningham JM, Butterbach K, Kote-Jarai Z, Kraft P, FitzGerald L, Lindström S, Adams M, McKay JD, Phelan CM, Benlloch S, Kelemen LE, Brennan P, Riggan M, O'Mara TA, Shen H, Shi Y, Thompson DJ, Goodman MT, Nielsen SF, Berchuck A, Laboissiere S, Schmit SL, Shelford T, Edlund CK, Taylor JA, Field JK, Park SK, Offit K, Thomassen M, Schmutzler R, Ottini L, Hung RJ, Marchini J, Amin Al Olama A, Peters U, Eeles RA, Seldin MF, Gillanders E, Seminara D, Antoniou AC, Pharoah PD, Chenevix-Trench G, Chanock SJ, Simard J and Easton DF. The oncoarray consortium: a network for understanding the genetic architecture of common cancers. Cancer Epidemiol Biomarkers Prev 2017; 26: 126-135.

[39] Bahcall O. COGS project and design of the iCOGS array. Nat Genet 2013.

[40] Huang CS, Lin CH, Lu YS and Shen CY. Unique features of breast cancer in Asian women--breast cancer in Taiwan as an example. J Steroid Biochem Mol Biol 2010; 118: 300-303.

[41] Matsuno RK, Anderson WF, Yamamoto S, Tsukuma H, Pfeiffer RM, Kobayashi K, Devesa SS and Levine PH. Early- and late-onset breast cancer types among women in the United States and Japan. Cancer Epidemiol Biomarkers Prev 2007; 16: 1437-1442.

[42] Min SY, Kim Z, Hur MH, Yoon CS, Park EH and Jung KW. The basic facts of Korean breast cancer in 2013: results of a nationwide survey and breast cancer registry database. J Breast Cancer 2016; 19: 1-7.

[43] Kan Z, Ding Y, Kim J, Jung HH, Chung W, Lal S, Cho S, Fernandez-Banet J, Lee SK, Kim SW, Lee JE, Choi YL, Deng S, Kim JY, Ahn JS, Sha Y, Mu XJ, Nam JY, Im YH, Lee S, Park WY, Nam SJ and Park YH. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. Nat Commun 2018; 9: 1725.

[44] Yu X, Zhang Z, Wang Z, Wu P, Qiu F and Huang J. Prognostic and predictive value of tumor-infiltrating lymphocytes in breast cancer: a systematic review and meta-analysis. Clin Transl Oncol 2016; 18: 497-506.

[45] Denkert C, Loibl S, Noske A, Roller M, Müller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, von Törne C, Weichert W, Engels K, Solbach C, Schrader I, Dietel M and von Minckwitz G. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J Clin Oncol 2010; 28: 105-113.

[46] Lin CH, Chuang PY, Chiang CJ, Lu YS, Cheng AL, Kuo WH, Huang CS, Lai MS, You SL and Tang CH. Distinct clinicopathological features and prognosis of emerging young-female breast cancer in an East Asian country: a nationwide cancer registry-based study. Oncologist 2014; 19: 583-591.

[47] Kurebayashi J, Miyoshi Y, Ishikawa T, Saji S, Sugie T, Suzuki T, Takahashi S, Nozaki M, Yamashita H, Tokuda Y and Nakamura S. Clinicopathological characteristics of breast cancer and trends in the management of breast cancer patients in Japan: based on the breast cancer registry of the Japanese Breast Cancer Society between 2004 and 2011. Breast Cancer 2015; 22: 235-244.

Supplementary material

The main assumptions of MR

When a MR is conducted, there are three assumptions that must be met [48]. (i) The relevance assumption states that genetic variants with well-known effects of exposure must be used as the instrumental variables (IVs) (i.e., BMI and T2D in this study). (ii) The exclusion restriction assumption states that the IVs are related to the outcome (i.e., breast cancer progression) solely via exposure. (iii) The independence assumption states that the IVs should be independent from any other risk factors that affect outcome.

Study population

The study participants were from the Breast Cancer Association Consortium (BCAC) [49]. We included 8,766 women of East Asian ancestry to carry out the study in a genetically homogenous population. Each of the participants had had breast cancer and had been enrolled in one of the six studies in the BCAC-Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC), Hong Kong Breast Cancer Study (HKBCS), Korean Hereditary Breast Cancer Study (KOHBRA), Shanghai Breast Cancer Genetic Study (SBCGS), Seoul Breast Cancer Study (SEBCS) and Taiwanese Breast Cancer Study (TWBCS).

Genotyping and imputation information for the BCAC

Genotyping was conducted using iCOGS and OncoArray; the 1,000 Genomes Phase 3 reference panel was used for imputation after quality control. The iCOGS is an Illumina array specifically designed for studying hormone-related cancers, including breast, ovarian and prostate cancer [50], and the OncoArray is also an Illumina array specifically designed for multiple cancer types, including breast, ovarian, prostate, colorectal and lung cancer [51]. Genotyping and quality control procedures have been described elsewhere [51, 52].

Standard quality-control procedures

Standard and stringent quality control procedures were performed to eliminate errors that could lead to spurious associations. These included the removal of individuals with discordance of genetically inferred sex versus self-reported sex, individuals with an extreme rate of heterozygosity, individual relatedness and individuals with an excess (> 5%) of missing genotype calls. For quality control of single-nucleotide polymorphisms (SNPs), we removed SNPs with low genotype call rates (missing rate of > 5%), SNPs violating the Hardy-Weinberg equilibrium and SNPs with a minor allele frequency of < 5%.

Assessing the relationship between BMI and breast cancer disease-free survival by stratifying based on estrogen receptor (ER) status

Among all of the IVs related to BMI that both identified in Taiwan Biobank and Biobank Japan Project, we performed a linear regression of these IVs to identify BMI-associated SNPs specific in individuals with ER-positive or ER-negative breast cancer tumours from the BCAC cohort. These IVs were then used in the MR analysis to investigate the causal effect between BMI and breast cancer disease-free survival among individuals with ER-positive or ER-negative breast cancer.

Data harmonization for MR

Data harmonization is a necessary process when two or more independently generated datasets are combined. In this two-sample MR analysis, two non-overlapping sets of individuals are used. Inappropriate data harmonization could distort the results of such an analysis. In this study, we harmonized our data to comply with the guidelines developed by Fortier [53, 54]. First, we standardized the direction of all IVs to ensure that all were positively associated with the exposure (which means the exposure-increasing allele is the effect allele) in individual SNP-exposure datasets. Second, we con-

firmed that the exposure-increasing alleles and the reference alleles in the SNP-exposure dataset and SNP-outcome dataset were coded identically. If the variant was not coded in the same direction, we then transformed the effect allele into the reference allele and *vice versa*. To do this, the regression coefficient was multiplied by -1, and the effect allele frequency was subtracted from 1. After data harmonization, the direction of IVs was positively associated with increased BMI and higher risk of having T2D.

References

[48] Lawlor DA, Harbord RM, Sterne JA, Timpson N and Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008; 27: 1133-1163.

[49] Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb D, Rostamianfar A, Bolla MK, Wang Q, Tyrer J, Dicks E, Lee A, Wang Z, Allen J, Keeman R, Eilber U, French JD, Qing Chen X, Fachal L, McCue K, McCart Reed AE, Ghoussaini M, Carroll JS, Jiang X, Finucane H, Adams M, Adank MA, Ahsan H, Aittomäki K, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Arun B, Auer PL, Bacot F, Barrdahl M, Baynes C, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Børresen-Dale AL, Brand JS, Brauch H, Brennan P, Brenner H, Brinton L, Broberg P, Brock IW, Broeks A, Brooks-Wilson A, Brucker SY, Brüning T, Burwinkel B, Butterbach K, Cai Q, Cai H, Caldés T, Canzian F, Carracedo A, Carter BD, Castelao JE, Chan TL, David Cheng TY, Seng Chia K, Choi JY, Christiansen H, Clarke CL; NBCS Collaborators, Collée M, Conroy DM, Cordina-Duverger E, Cornelissen S, Cox DG, Cox A, Cross SS, Cunningham JM, Czene K, Daly MB, Devilee P, Doheny KF, Dörk T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Elvira M, Engel C, Eriksson M, Fasching PA, Figueroa J, Flesch-Janys D, Fletcher O, Flyger H, Fritschi L, Gaborieau V, Gabrielson M, Gago-Dominguez M, Gao YT, Gapstur SM, García-Sáenz JA, Gaudet MM, Georgoulias V, Giles GG, Glendon G, Goldberg MS, Goldgar DE, González-Neira A, Grenaker Alnæs GI, Grip M, Gronwald J, Grundy A, Guénel P, Haeberle L, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hamel N, Hankinson S, Harrington P, Hart SN, Hartikainen JM, Hartman M, Hein A, Heyworth J, Hicks B, Hillemanns P, Ho DN, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Hou MF, Hsiung CN, Huang G, Humphreys K, Ishiguro J, Ito H, Iwasaki M, Iwata H, Jakubowska A, Janni W, John EM, Johnson N, Jones K, Jones M, Jukkola-Vuorinen A, Kaaks R, Kabisch M, Kaczmarek K, Kang D, Kasuga Y, Kerin MJ, Khan S, Khusnutdinova E, Kiiski JI, Kim SW, Knight JA, Kosma VM, Kristensen VN, Krüger U, Kwong A, Lambrechts D, Le Marchand L, Lee E, Lee MH, Lee JW, Neng Lee C, Lejbkowicz F, Li J, Lilyquist J, Lindblom A, Lissowska J, Lo WY, Loibl S, Long J, Lophatananon A, Lubinski J, Luccarini C, Lux MP, Ma ESK, MacInnis RJ, Maishman T, Makalic E, Malone KE, Kostovska IM, Mannermaa A, Manoukian S, Manson JE, Margolin S, Mariapun S, Martinez ME, Matsuo K, Mavroudis D, McKay J, McLean C, Meijers-Heijboer H, Meindl A, Menéndez P, Menon U, Meyer J, Miao H, Miller N, Taib NAM, Muir K, Mulligan AM, Mulot C, Neuhausen SL, Nevanlinna H, Neven P, Nielsen SF, Noh DY, Nordestgaard BG, Norman A, Olopade OI, Olson JE, Olsson H, Olswold C, Orr N, Pankratz VS, Park SK, Park-Simon TW, Lloyd R, Perez JIA, Peterlongo P, Peto J, Phillips KA, Pinchev M, Plaseska-Karanfilska D, Prentice R, Presneau N, Prokofyeva D, Pugh E, Pylkäs K, Rack B, Radice P, Rahman N, Rennert G, Rennert HS, Rhenius V, Romero A, Romm J, Ruddy KJ, Rüdiger T, Rudolph A, Ruebner M, Rutgers EJT, Saloustros E, Sandler DP, Sangrajrang S, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schürmann P, Scott RJ, Scott C, Seal S, Seynaeve C, Shah M, Sharma P, Shen CY, Sheng G, Sherman ME, Shrubsole MJ, Shu XO, Smeets A, Sohn C, Southey MC, Spinelli JJ, Stegmaier C, Stewart-Brown S, Stone J, Stram DO, Surowy H, Swerdlow A, Tamimi R, Taylor JA, Tengström M, Teo SH, Beth Terry M, Tessier DC, Thanasitthichai S, Thöne K, Tollenaar RAEM, Tomlinson I, Tong L, Torres D, Truong T, Tseng CC, Tsugane S, Ulmer HU, Ursin G, Untch M, Vachon C, van Asperen CJ, Van Den Berg D, van den Ouweland AMW, van der Kolk L, van der Luijt RB, Vincent D, Vollenweider J, Waisfisz Q, Wang-Gohrke S, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Wu AH, Xia L, Yamaji T, Yang XR, Har Yip C, Yoo KY, Yu JC, Zheng W, Zheng Y, Zhu B, Ziogas A, Ziv E; ABCTB Investigators; ConFab/AOCS Investigators, Lakhani SR, Antoniou AC, Droit A, Andrulis IL, Amos CI, Couch FJ, Pharoah PDP, Chang-Claude J, Hall P, Hunter DJ, Milne RL, García-Closas M, Schmidt MK, Chanock SJ, Dunning AM, Edwards SL, Bader GD, Chenevix-Trench G, Simard J, Kraft P and Easton DF. Association analysis identifies 65 new breast cancer risk loci. Nature 2017; 551: 92-94.

[50] Bahcall O. COGS project and design of the iCOGS array. Nat Genet 2013.

[51] Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, Casey G, Hunter DJ, Sellers TA, Gruber SB, Dunning AM, Michailidou K, Fachal L, Doheny K, Spurdle AB, Li Y, Xiao X, Romm J, Pugh E, Coetzee GA, Hazelett DJ, Bojesen SE, Caga-Anan C, Haiman CA, Kamal A, Luccarini C, Tessier D, Vincent D, Bacot F, Van Den Berg DJ, Nelson S, Demetriades S, Goldgar DE, Couch FJ, Forman JL, Giles GG, Conti DV, Bickeböller H, Risch A, Waldenberger M, Brüske-Hohlfeld I, Hicks BD, Ling H, McGuffog L, Lee A, Kuchenbaecker K, Soucy P, Manz J, Cunningham JM, Butterbach K, Kote-Jarai Z, Kraft P, FitzGerald L, Lindström S, Adams M, McKay JD, Phelan

CM, Benlloch S, Kelemen LE, Brennan P, Riggan M, O'Mara TA, Shen H, Shi Y, Thompson DJ, Goodman MT, Nielsen SF, Berchuck A, Laboissiere S, Schmit SL, Shelford T, Edlund CK, Taylor JA, Field JK, Park SK, Offit K, Thomassen M, Schmutzler R, Ottini L, Hung RJ, Marchini J, Amin Al Olama A, Peters U, Eeles RA, Seldin MF, Gillanders E, Seminara D, Antoniou AC, Pharoah PD, Chenevix-Trench G, Chanock SJ, Simard J and Easton DF. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. Cancer Epidemiol Biomarkers Prev 2017; 26: 126-135.

[52] Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N; Breast and Ovarian Cancer Susceptibility Collaboration, Fletcher O, Peto J, Gibson L, Dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Czene K, Irwanto A, Liu J, Waisfisz Q, Meijers-Heijboer H, Adank M; Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Müller-Myhsok B, Lichtner P, Hopper JL, Southey MC, Makalic E, Schmidt DF, Uitterlinden AG, Hofman A, Hunter DJ, Chanock SJ, Vincent D, Bacot F, Tessier DC, Canisius S, Wessels LF, Haiman CA, Shah M, Luben R, Brown J, Luccarini C, Schoof N, Humphreys K, Li J, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang X, Vachon C, Stevens KN, Lambrechts D, Moisse M, Paridaens R, Christiaens MR, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken Z, Aaltonen K, Heikkinen T, Broeks A, Veer LJ, van der Schoot CE, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Zamora MP, Perez JI, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MW, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Henderson BE, Schumacher F, Le Marchand L, Andrulis IL, Knight JA, Glendon G, Mulligan AM; kConFab Investigators; Australian Ovarian Cancer Study Group, Lindblom A, Margolin S, Hooning MJ, Hollestelle A, van den Ouweland AM, Jager A, Bui QM, Stone J, Dite GS, Apicella C, Tsimiklis H, Giles GG, Severi G, Baglietto L, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Brenner H, Müller H, Arndt V, Stegmaier C, Swerdlow A, Ashworth A, Orr N, Jones M, Figueroa J, Lissowska J, Brinton L, Goldberg MS, Labrèche F, Dumont M, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Brüning T; GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network, Radice P, Peterlongo P, Manoukian S, Bonanni B, Devilee P, Tollenaar RA, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dörk T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo KY, Noh DY, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng CC, Van Den Berg D, Stram DO, Shu XO, Lu W, Gao YT, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng JH, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Shen CY, Hsiung CN, Wu PE, Ding SL, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Blot WJ, Signorello LB, Cai Q, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Simard J, Garcia-Closas M, Pharoah PD, Chenevix-Trench G, Dunning AM, Benitez J and Easton DF. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013; 45: 353-361, 361e1-2.

[53] Hartwig FP, Davies NM, Hemani G and Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int J Epidemiol 2016; 45: 1717-1726.

[54] Fortier I, Raina P, Van den Heuvel ER, Griffith LE, Craig C, Saliba M, Doiron D, Stolk RP, Knoppers BM, Ferretti V, Granda P and Burton P. Maelstrom research guidelines for rigorous retrospective data harmonization. Int J Epidemiol 2017; 46: 103-105.

Table S1. Description of Breast Cancer Association Consortium studies from which participants were enrolled

Study acronym	Full study name	Country	Study design	Case definition	Sample size in our study	Age (years)	Reference
HERPACC	Hospital-based Epidemiologic Research Program at Aichi Cancer Center	Japan	Hospital-based case-control study	Incident breast cancer cases that first visited Aichi Cancer Center between 2001 and 2013 and were diagnosed within 1 year from the first visit. No previous history of any type of cancer.	793	23-79	Kawase T, Matsuo K, Suzuki T, Hiraki A, Watanabe M, Iwata H, Tanaka H and Tajima K. FGFR2 intronic polymorphisms interact with reproductive risk factors of breast cancer: results of a case control study in Japan. Int J Cancer 2009; 125: 1946-1952.
HKBCS	Hong Kong Breast Cancer Study	Hong Kong	Hospital-based case-control study	Genetic screening of breast cancer patients at high risk from all Hong Kong hospitals. Incidence cases classified as high-risk group: 1) first-degree relative with breast and/or ovarian cancer, 2) cases where age is ≤ 45 years, 3) bilateral breast cancer, 4) triple-negative breast cancer, 5) family history of breast and/or ovarian cancer. Cases were recruited 2006-2014.	547	18-82	1) Kwong A, Ng EK, Law FB, Wong HN, Wa A, Wong CL, Kurian AW, West DW, Ford JM and Ma ES. Novel BRCA1 and BRCA2 genomic rearrangements in Southern Chinese breast/ovarian cancer patients. Breast Cancer Res Treat 2012; 136: 931-3. 2) Kwong A, Ng EK, Wong CL, Law FB, Au T, Wong HN, Kurian AW, West DW, Ford JM and Ma ES. Identification of BRCA1/2 founder mutations in Southern Chinese breast cancer patients using gene sequencing and high resolution DNA melting analysis. PLoS One 2012; 7: e43994.
KOHBRA	Korean Hereditary Breast Cancer Study	Korea	Population-based case-control study	Breast cancer patients at high risk were recruited from nationwide University Hospitals from May 2007 to May 2012. High-risk status means 1) familial breast cancer, 2) early-onset breast cancer (age < 40 years), 3) breast and past/current ovarian cancer, 4) past/current double primary cancers, 5) bilateral breast cancer, 6) male breast cancer cases.	1432	19-81	Han SA, Park SK, Ahn SH, Lee MH, Noh DY, Kim LS, Noh WC, Jung Y, Kim KS and Kim SW; Korean Breast Cancer Study Group. The Korean Hereditary Breast Cancer (KOHBRA) study: protocols and interim report. Clin Oncol (R Coll Radiol) 2011; 23: 434-41.
SBCGS	Shanghai Breast Cancer Genetic Study	China	Population-based case-control study, cohort study	Newly diagnosed breast cancer cases recruited from 1996 to 2009. Cases were identified mostly from the Shanghai Cancer Registry. Some cases were identified from the Shanghai Women's Health Study.	1243	25-80	Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W and Shu XO. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009; 41: 324-8.
SEBCS	Seoul Breast Cancer Study	Korea	Hospital-based case-control study	Consecutive incident cases from two hospitals in Seoul recruited 2001-2005.	2123	19-89	1) Lee KM, Choi JY, Park SK, Chung HW, Ahn B, Yoo KY, Han W, Noh DY, Ahn SH, Kim H, Wei Q and Kang D. Genetic polymorphisms of ataxia telangiectasia mutated and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2005; 14: 821-5. 2) Han S, Lee KM, Choi JY, Park SK, Lee JY, Lee JE, Noh DY, Ahn SH, Han W, Kim DH, Hong YC, Ha E, Yoo KY and Kang D. CASP8 polymorphisms, estrogen and progesterone receptor status, and breast cancer risk. Breast Cancer Res Treat 2008; 110: 387-93.
TWBCS	Taiwanese Breast Cancer Study	Taiwan	Hospital-based case-control study	Incident cases diagnosed & treated at two major teaching hospitals in Taiwan. Cases recruited between March 2002 and August 2005.	504	18-90	1) Hsu HM, Wang HC, Chen ST, Hsu GC, Shen CY and Yu JC. Breast cancer risk is associated with genes encoding the DNA double-strand break repair Mre11/Rad50/Nbs1 complex. Cancer Epidemiol Biomarkers Prev 2007; 16: 2024-32. 2) Ding SL, Yu JC, Chen ST, Hsu GC, Kuo SJ, Lin YH, Wu PE and Shen CY. Genetic variants of BLM interact with RAD51 to increase breast cancer susceptibility. Carcinogenesis 2009; 30: 43-9.

Table S2. Instrumental variables for BMI

SNP	CHR	Position	Related gene	EA in BBJ	OA in BBJ	Effect estimate (IV-exposure)	SE (IV-exposure)	*P*-value	EA in BCAC	Effect estimate (IV-outcome)	SE (IV-outcome)
rs10946398	6	20661034	CDKAL1	C	A	-0.03685	0.00531	3.96E-12	A	0.05068	0.07049
rs2206734	6	20694884	CDKAL1/RPL36AP25	T	C	-0.03766	0.00546	5.30E-12	C	0.06204	0.06979
rs2328529	6	20631953	CDKAL1	A	C	-0.03399	0.00558	1.08E-09	C	0.06339	0.07278
rs2328545	6	20653550	CDKAL1	C	G	-0.03667	0.00532	5.60E-12	G	0.07109	0.06912
rs2328548	6	20716958	CDKAL1/RPL36AP25	A	G	-0.03114	0.00525	2.99E-09	A	-0.06062	0.06873
rs4710940	6	20658012	CDKAL1	C	A	-0.03691	0.00531	3.69E-12	A	0.07360	0.06945
rs4712522	6	20656800	CDKAL1	G	C	-0.03692	0.00531	3.64E-12	C	0.07095	0.06944
rs4712525	6	20662966	CDKAL1	T	C	-0.03683	0.00532	4.20E-12	C	0.05977	0.07015
rs4712526	6	20663035	CDKAL1	A	T	-0.03684	0.00532	4.19E-12	T	0.05977	0.07015
rs6456367	6	20659587	CDKAL1	A	T	-0.03690	0.00531	3.80E-12	T	0.06045	0.07016
rs6456368	6	20659806	CDKAL1	C	T	-0.03689	0.00531	3.82E-12	T	0.06044	0.07016
rs6906327	6	20659459	CDKAL1	A	G	-0.03690	0.00531	3.77E-12	G	0.06317	0.07017
rs742642	6	20665081	CDKAL1	A	G	-0.03664	0.00533	6.44E-12	G	0.06232	0.07002
rs7751957	6	20638009	CDKAL1	G	C	-0.03145	0.00542	6.60E-09	C	0.06829	0.07028
rs7752780	6	20666022	CDKAL1	A	G	-0.03723	0.00536	3.67E-12	G	0.05977	0.07015
rs7768642	6	20633907	CDKAL1	A	G	-0.03147	0.00542	6.48E-09	G	0.05730	0.07038
rs7774594	6	20661143	CDKAL1	A	T	-0.03684	0.00531	4.00E-12	T	0.05977	0.07015
rs9295474	6	20652717	CDKAL1	G	C	-0.03838	0.00529	4.07E-13	C	0.06168	0.06902
rs9350271	6	20683164	CDKAL1/RPL36AP25	A	G	-0.03793	0.00544	3.20E-12	G	0.06838	0.06966
rs9356747	6	20725007	CDKAL1/RPL36AP25	A	T	-0.03220	0.00527	9.83E-10	T	0.07832	0.06923
rs9356748	6	20725097	CDKAL1/RPL36AP25	A	T	-0.03204	0.00527	1.20E-09	T	0.07400	0.06912
rs9358355	6	20654897	CDKAL1	C	T	-0.03668	0.00532	5.58E-12	T	0.07855	0.06924
rs9358356	6	20667382	CDKAL1	C	T	-0.03778	0.00541	2.89E-12	T	0.06411	0.07046
rs9358357	6	20719145	CDKAL1/RPL36AP25	G	A	-0.03089	0.00525	4.03E-09	G	-0.06164	0.06880
rs9358358	6	20719393	CDKAL1/RPL36AP25	C	A	-0.03071	0.00525	4.93E-09	C	-0.05997	0.06873
rs9368216	6	20655110	CDKAL1	G	A	-0.03669	0.00533	5.58E-12	A	0.07592	0.06923
rs9368219	6	20674691	CDKAL1/RPL36AP25	T	C	-0.03780	0.00551	6.70E-12	C	0.06645	0.07014
rs9460544	6	20661529	CDKAL1	T	G	-0.03683	0.00531	4.05E-12	G	0.05977	0.07015
rs9460545	6	20661550	CDKAL1	C	T	-0.03683	0.00531	4.06E-12	T	0.05977	0.07015
rs9460550	6	20719561	CDKAL1/RPL36AP25	A	G	-0.03088	0.00525	4.08E-09	A	-0.06157	0.06880
rs9465837	6	20624179	CDKAL1	G	C	-0.03223	0.00546	3.63E-09	C	0.05541	0.07120
rs9465847	6	20634428	CDKAL1	T	G	-0.03146	0.00542	6.54E-09	G	0.05689	0.07039

rs9465871	6	20717255	CDKAL1/RPL36AP25	C	T	-0.03138	0.00525	2.32E-09	C	-0.05877	0.06906

BMI: body mass index; SNP: single-nucleotide polymorphism; CHR: chromosome; BBJ: Biobank Japan Project; BCAC: Breast Cancer Association Consortium; EA in BBJ: effect allele of exposure-associated variant from the BBJ; OA in BBJ: other allele of exposure-associated variant from the BBJ; Effect estimate (IV-exposure): the effect estimate of the association of the exposure-associated variant and exposure from the BBJ; SE (IV-exposure): the standard error of the association of the exposure-associated variant and exposure from the BBJ; *P*-value: the genome-wide *P*-value of the exposure-associated variant from the BBJ; EA in BCAC: the effect allele of the exposure-associated variant from the BCAC; Effect estimate (IV-outcome): the effect estimate of the association of the exposure-associated variant and breast cancer disease-free survival from the BCAC; SE (IV-outcome): the standard error of the association of the exposure-associated variant and breast cancer disease-free survival from the BCAC.

Table S3. Instrumental variables for T2D

SNP	CHR	Position	Related gene	EA in BBJ	OA in BBJ	Effect estimates (IV-exposure)	SE (IV-exposure)	*P*-value	EA in BCAC	Effect estimates (IV-outcome)	SE (IV-outcome)
rs4481184	3	185505787	IGF2BP2/MIR548AQ	C	T	0.120618	0.014223	2.25E-17	C	-0.00269	0.07968
rs7646518	3	185514931	IGF2BP2/MIR548AQ	T	C	0.120414	0.014226	2.57E-17	T	-0.00291	0.07970
rs4686696	3	185516520	IGF2BP2/MIR548AQ	G	A	0.120500	0.014227	2.46E-17	G	-0.00291	0.07970
rs9465837	6	20624179	CDKAL1	C	G	0.152245	0.013805	2.79E-28	C	0.05803	0.07142
rs9356741	6	20625100	CDKAL1	T	C	0.073012	0.013328	4.30E-08	C	0.01621	0.07431
rs7768642	6	20633907	CDKAL1	G	A	0.150770	0.013649	2.28E-28	G	0.06091	0.07058
rs9465847	6	20634428	CDKAL1	G	T	0.150743	0.013648	2.31E-28	G	0.06053	0.07059
rs9295474	6	20652717	CDKAL1	C	G	0.154258	0.013403	1.18E-30	C	0.06647	0.06916
rs2328545	6	20653550	CDKAL1	G	C	0.165541	0.013501	1.47E-34	G	0.07639	0.06930
rs9358355	6	20654897	CDKAL1	T	C	0.165816	0.013502	1.15E-34	T	0.08378	0.06940
rs9368216	6	20655110	CDKAL1	A	G	0.166044	0.013515	1.08E-34	A	0.08094	0.06940
rs4712522	6	20656800	CDKAL1	C	G	0.166469	0.013494	5.76E-35	C	0.07622	0.06961
rs4710940	6	20658012	CDKAL1	A	C	0.166697	0.013495	4.74E-35	A	0.07907	0.06961
rs6906327	6	20659459	CDKAL1	G	A	0.167856	0.013515	2.04E-35	G	0.06908	0.07033
rs6456367	6	20659587	CDKAL1	T	A	0.167934	0.013522	2.06E-35	T	0.06614	0.07033
rs6456368	6	20659806	CDKAL1	T	C	0.167999	0.013524	1.98E-35	T	0.06614	0.07032
rs10946398	6	20661034	CDKAL1	A	C	0.168528	0.013553	1.69E-35	A	0.05632	0.07067
rs7774594	6	20661143	CDKAL1	T	A	0.168360	0.013532	1.55E-35	T	0.06550	0.07031
rs9460544	6	20661529	CDKAL1	G	T	0.168427	0.013533	1.48E-35	G	0.06550	0.07031
rs9460545	6	20661550	CDKAL1	T	C	0.168435	0.013534	1.48E-35	T	0.06550	0.07031
rs4712525	6	20662966	CDKAL1	C	T	0.168796	0.013542	1.17E-35	C	0.06550	0.07031
rs4712526	6	20663035	CDKAL1	T	A	0.168822	0.013543	1.15E-35	T	0.06550	0.07031
rs742642	6	20665081	CDKAL1	G	A	0.169795	0.013603	9.34E-36	G	0.06790	0.07019
rs7752780	6	20666022	CDKAL1	G	A	0.169947	0.013572	5.67E-36	G	0.06550	0.07031
rs9358356	6	20667382	CDKAL1	T	C	0.172022	0.013627	1.57E-36	T	0.06954	0.07062

rs9368219	6	20674691	CDKAL1/RPL36AP25	C	T	0.180110	0.013568	3.25E-40	C	0.07065	0.07030
rs9350271	6	20683164	CDKAL1/RPL36AP25	G	A	0.180108	0.013491	1.18E-40	G	0.07313	0.06981
rs2328548	6	20716958	CDKAL1/RPL36AP25	G	A	0.153592	0.013203	2.80E-31	A	-0.06807	0.06897
rs9465871	6	20717255	CDKAL1/RPL36AP25	T	C	0.153930	0.013205	2.11E-31	C	-0.06668	0.06930
rs9358357	6	20719145	CDKAL1/RPL36AP25	A	G	0.152544	0.013202	7.03E-31	G	-0.06864	0.06903
rs9358358	6	20719393	CDKAL1/RPL36AP25	A	C	0.152246	0.013203	9.16E-31	C	-0.06700	0.06896
rs9460550	6	20719561	CDKAL1/RPL36AP25	G	A	0.152597	0.013203	6.72E-31	A	-0.06856	0.06903
rs62481355	7	127201664	CDKAL1/RPL36AP25	C	T	0.086969	0.014236	1.00E-09	T	0.00381	0.07579
rs11558471	8	118185733	SLC30A8	A	G	0.111946	0.013346	4.94E-17	A	0.01214	0.07056
rs11774700	8	118220270	SLC30A8	T	C	0.107908	0.014330	5.07E-14	T	0.00796	0.07274
rs11187007	10	94214580	MARK2P9/IDE	A	G	0.116861	0.014190	1.79E-16	G	-0.01464	0.08040
rs11187033	10	94262359	IDE	T	A	0.120688	0.014077	1.01E-17	A	-0.02993	0.07617
rs10509645	10	94277866	IDE	C	A	0.121433	0.014073	6.19E-18	A	-0.02828	0.07595
rs10882074	10	94281685	IDE	T	G	0.122650	0.014222	6.48E-18	G	-0.01933	0.07541
rs2421943	10	94311815	IDE	G	A	0.121901	0.013991	2.97E-18	A	-0.01719	0.07542
rs7076966	10	94325511	IDE	C	T	0.121408	0.013979	3.79E-18	T	-0.01418	0.07548
rs12778642	10	94464307	KIF11	G	T	0.151447	0.015259	3.24E-23	T	-0.03899	0.08492
rs10748582	10	94477219	EIF2S2P3	T	A	0.143743	0.016688	7.07E-18	A	-0.03441	0.09154
rs7923837	10	94481917	HHEX	G	A	0.142074	0.016619	1.24E-17	A	-0.03208	0.09096
rs7923866	10	94482076	HHEX	C	T	0.142117	0.016622	1.23E-17	T	-0.03208	0.09096

T2D: type 2 diabetes; SNP: single-nucleotide polymorphism; BBJ: Biobank Japan Project; BCAC: Breast Cancer Association Consortium; CHR: chromosome; EA in BBJ: effect allele of the exposure-associated variant from the BBJ; OA in BBJ: other allele of the exposure-associated variant from the BBJ; Effect estimate (IV-exposure): the effect estimate of the association of the exposure-associated variant and exposure from the BBJ; SE (IV-exposure): the standard error of the association of the exposure-associated variant and exposure from the BBJ; *P*-value: the genome-wide *P*-value of the exposure-associated variant from the BBJ; EA in BCAC: the effect allele of the exposure-associated variant from the BCAC; Effect estimate (IV-outcome): the effect estimate of the association of the exposure-associated variant and breast cancer disease-free survival from the BCAC; SE (IV-outcome): the standard error of the association of the exposure-associated variant and breast cancer disease-free survival from the BCAC.