Research Paper

Causal role of blood metabolites in HER-positive and HER-negative breast cancer: a Mendelian randomization (MR) study

Jian Yue[1,2,3,*], Huiying Fang[4,*], Sheng Chen[3], Lei Gu[5], Guosheng Ren[1,2]

[1]Chongqing Key Laboratory of Molecular Oncology and Epigenetics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[2]Department of Breast and Thyroid Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[3]Department of Breast Surgery, Gaozhou People's Hospital, Gaozhou, China
[4]Department of Breast Cancer Center, Chongqing Key Laboratory for Intelligent Oncology in Breast Cancer (iCQBC), Chongqing University Cancer Hospital, Chongqing, China
[5]Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Soochow University, Suzhou, China
[*]Share first authorship

Correspondence to: Lei Gu, Guosheng Ren; **email:** gulei116860@163.com, https://orcid.org/0000-0003-4035-3583; Rengs726@126.com, https://orcid.org/0000-0002-6416-8967
Keywords: blood metabolites, breast cancer, HER, Mendelian randomization
Received: January 16, 2024 **Accepted:** July 5, 2024 **Published:** August 2, 2024

ABSTRACT

Background: Previous studies provide evidence that *in vivo* metabolites are associated with breast cancer (BC). However, the causal relationship between blood metabolites and BC remains unclear.

Method: Comprehensive two-sample Mendelian randomization analysis was conducted to determine the causal association between 1400 publicly available genetic data on metabolic factors and human epidermal growth factor receptor positive (HER+) BC or HER- BC in this study.

Result: Epiandrosterone sulfate levels (OR = 1.07, 95% CI = 1.02 ~ 1.10, *p* = 0.0013), 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels (OR = 1.07, 95% CI = 1.03 ~ 1.12, *p* = 0.0012), glycohyocholate levels (OR = 0.85, 95% CI = 0.77 ~ 0.93, *p* = 0.0007) and etiocholanolone glucuronide levels (OR = 1.12, 95% CI = 1.05 ~ 1.20, *p* = 0.0013) were causally correlated with HER+ BC. 5 metabolites were causally correlated with HER- BC: Vanillic acid glycine levels (OR = 1.14, 95% CI = 1.06 ~ 1.22, *p* = 0.0003), Thyroxine levels (OR = 1.26, 95% CI = 1.11 ~ 1.44, *p* = 0.0004), 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels (OR = 0.86, 95% CI = 0.79 ~ 0.94, *p* = 0.0010), N-acetylphenylalanine levels (OR = 1.12, 95% CI = 1.05 ~ 1.19, *p* = 0.0007) and Glucose-to-mannose ratio (OR = 1.15, 95% CI = 1.06 ~ 1.24, *p* = 0.0008). Two common causally related metabolites were identified: Gamma-glutamyl glutamate and X-12849 levels.

Conclusions: Our study has respectively demonstrated the connection between blood metabolites and HER+ or HER- BC by genetic means, thereby offering opportunities for therapeutic targets.

INTRODUCTION

The most recent report from CA (A Cancer Journal for Clinicians) indicates that breast cancer (BC) is currently the most commonly diagnosed cancer among women, accounting for 31% of female cancer cases in 2023. It remains the second leading cause of cancer-related mortality in women, accounting for 15% of such

instances [1]. The ErbB family comprises receptor tyrosine kinases, including human epidermal growth factor receptors (HER) 1/2/3/4, situated on the cellular membrane and responsive to a diverse range of ligands [2]. These receptors, capable of homo- or heterodimerization, play a crucial role in normal cell development but can lead to cancer through dysregulation, driving abnormal cell growth and survival through intricate signaling pathways [3, 4]. Among these receptors, HER2 has been extensively studied and is a primary target for treatment. Overexpression of HER2 is observed in approximately 15–20% of breast cancer cases and is associated with a poorer prognosis [2, 5]. Notably, HER3 lacks intrinsic kinase activity but can form heterodimers with HER2 (and/or HER1), significantly enhancing transphosphorylation and subsequent activation of downstream signaling pathways [3]. In recent years, conflicting data have emerged regarding the role of HER4 in breast cancer. Some studies suggest a negative impact of HER4 expression on disease progression, while others demonstrate beneficial effects [6]. A study involving postmenopausal breast cancer patients with varying levels of HER4 expression revealed significantly improved survival rates in those lacking HER4. These findings may be associated with the intricate interplay among these receptors [7].

There is mounting evidence of a robust association between metabolites and tumorigenesis and progression. In the context of breast cancer, in addition to the well-established pivotal roles of estrogen and progesterone in its development, an increasing body of literature has identified a diverse array of metabolites intricately linked to breast cancer [8–10]. A previous study systematically characterized metabolites in triple-negative breast cancer (TNBC) by profiling the polar metabolome and lipidome in 330 TNBC samples and 149 paired normal breast tissues, highlighting key subtype-specific metabolites as potential therapeutic targets [11]. Therefore, delving into the causal relationship between metabolites and breast cancer in depth is of significant scientific interest, especially considering the wealth of metabolomics data available and the opportunity for causal analysis among different subtypes of breast cancer characterized by distinct HER status.

Mendelian randomization (MR), a method grounded in the principles of Mendelian inheritance, serves as an indispensable analytical tool for inferring causal relationships in epidemiological studies [12]. Given the complex roles that different HER statuses play in the development and progression of breast cancer, there is a pressing need for a more comprehensive exploration. Accordingly, this study employs a comprehensive two-sample MR analysis to establish causal associations between metabolites and HER-positive/negative (HER+/−) breast cancer. The primary objective is to provide a nuanced understanding that can effectively inform clinical practices.

METHODS

We employed a two-sample MR approach utilizing publicly available datasets that provide genome-wide association outcomes for metabolic factors, HER+ breast cancer and HER- breast cancer. Two-sample MR involves the use of distinct datasets or samples to establish the gene–risk factor associations (e.g., blood metabolites and metabolite ratios traits) and the gene–outcome associations (e.g., malignant neoplasm of breast, HER-positive/malignant neoplasm of breast, HER-negative).

Study design

We assessed the causal relationship between 1,400 blood metabolic factors (1,091 blood metabolites and 309 metabolite ratios) and HER+/HER− breast cancer based on a two-sample MR analysis. MR utilizes genetic variation to represent risk factors, and therefore, valid instrumental variables (IVs) in causal inference must satisfy three key assumptions: (1) genetic variation is directly associated with exposure; (2) genetic variation is not associated with possible confounders between exposure and outcome; and (3) genetic variation does not affect outcome through pathways other than exposure [13, 14]. The studies included in our analysis were approved by the relevant institutional review boards (Figure 1).

Genome-wide association study (GWAS) data sources for HER+/HER− breast cancer

The GWAS summary statistics of HER+ breast cancer was obtained from the FinnGen consortium R9 release data. The study performed a GWAS on 176715 European individuals (Ncase = 9698, Ncontrol = 167017). As well as HER- breast cancer which performed a GWAS on 172982 European individuals (Ncase = 5965, Ncontrol = 167017).

Metabolites GWAS data sources

GWAS summary statistics for each blood metabolic factor are publicly available from the GWAS Catalog (accession numbers from GCST90199621 to GCST90201020) [15]. A total of 1400 metabolic factors (1,091 blood metabolites and 309 metabolite ratios) were involved. The original GWAS on blood metabolites was performed using data from 8096 unrelated European subjects in Canadian Longitudinal

Study of Aging (CLSA) who have been genome-wide genotyped and have had circulating plasma metabolites measured [15, 16]. Approximately there are 15.4 million single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) higher than 0.1%, imputation quality score >0.3, and missing rate <0.1 for GWAS testing [15].

Selection of instrumental variables (IVs)

In accordance with recent research [17, 18], the significance level of IVs for each metabolite was set to 1×10^{-5}. To mitigate potential bias arising from strong linkage disequilibrium (LD), we implemented a clumping algorithm with a cutoff of $r^2 < 0.001$ and a distance of 10,000 base pairs (kb) to ensure independence among the included SNPs. For consistency, we harmonized exposures and outcomes in terms of the effect allele and carried out subsequent analyses using the merged exposure-outcome dataset. The F statistic is a measure of instrument strength that is related to the proportion of variance in the phenotype explained by the genetic variants, sample size, and the number of instruments. An F statistic of ≥10 indicates a relatively low risk of weak instrument bias in MR analysis [19].

Statistical analysis

To evaluate the causal association between 1400 metabolic factors and HER+/HER− breast cancer, we utilized five well-established MR methods, comprising inverse-variance weighted (IVW), MR-Egger regression, weighted median, weighted mode, and simple mode, to analyze data involving multiple IVs [20, 21].

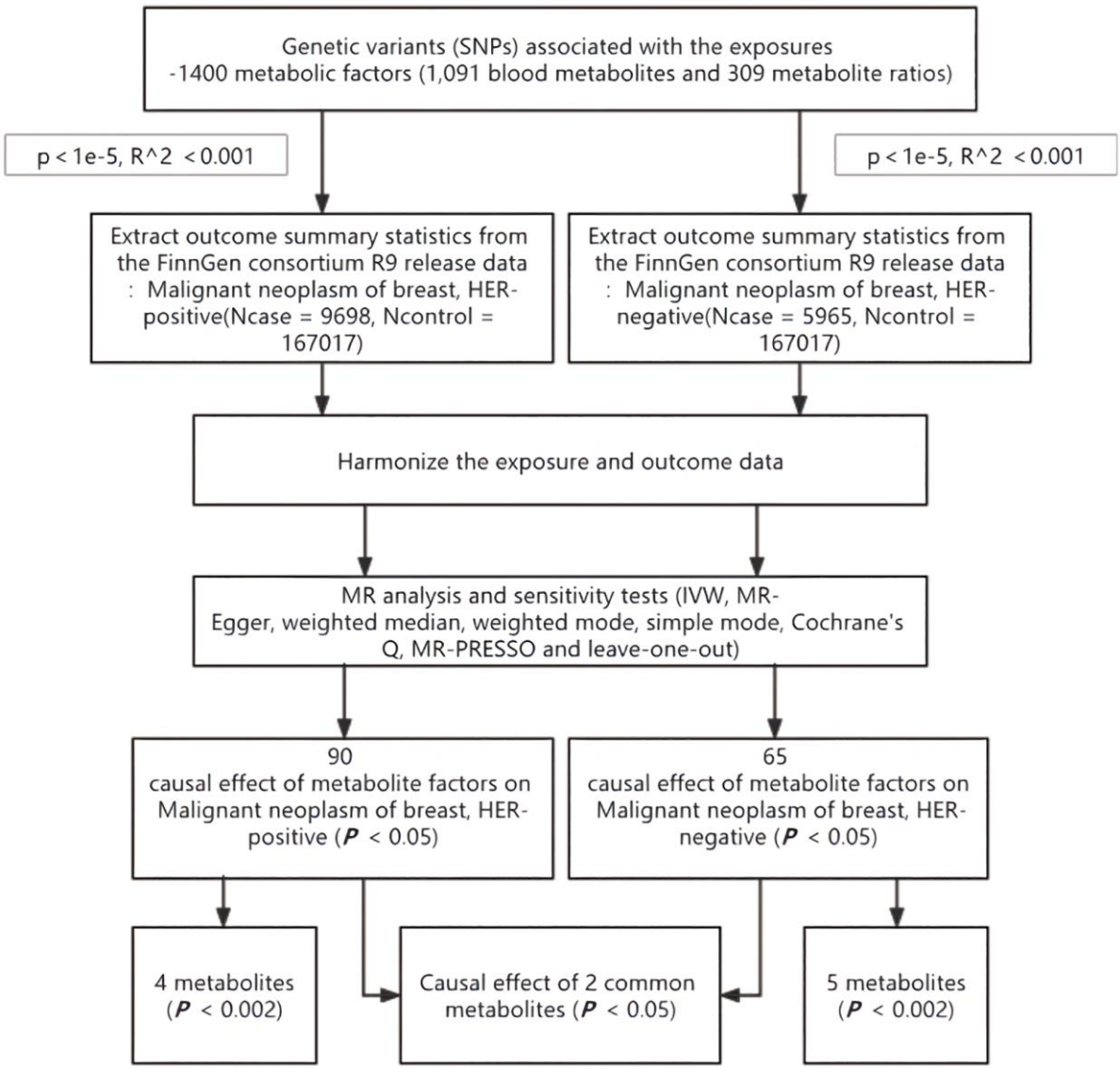


Figure 1. Flow chart for study. Abbreviations: MR: mendelian randomization; IVW: inverse-variance weighted; MR-PRESSO: MR pleiotropy residual sum and outlier.

We did not correct for multiple testing in this exploratory study. The primary emphasis was placed on the IVW method for our main results at significance of 0.05 level [22, 23], with the other methods providing supplementary insights. To gauge the heterogeneity among IVs, we employed Cochrane's Q-statistic, considering $p < 0.05$ as indicative of significant heterogeneity [24]. If the null hypothesis is rejected, random effects IVW was used instead of fixed-effects IVW [24, 25]. In the presence of notable pleiotropy, we conducted the MR-Egger intercept test and MR pleiotropy residual sum and outlier (MR-PRESSO) method to assess directional pleiotropy [26, 27]. Furthermore, MR-PRESSO method was utilized to exclude possible horizontal pleiotropic outliers that could substantially affect the estimation results [28]. To assess result stability, we conducted a leave-one-out sensitivity analysis, systematically excluding individual IVs one at a time [29]. In addition, scatter plots and funnel plots were used. Scatter plots showed that the results were not affected by outliers. Funnel plots demonstrated the robustness of the correlation and no heterogeneity. All statistical analyses were conducted with R software (version 4.3.0) using the "TwoSampleMR" and "MR-PRESSO" packages.

Data availability statement

All data are publicly available.

RESULTS

Exploration of the causal effect of metabolites on HER+ breast cancer

The IVW method showed evidence to support that ninety metabolites were identified at a significance of 0.05, so set the *p*-value to 0.002 and detected risk effects of three metabolites on HER+ breast cancer: Epiandrosterone sulfate levels, 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels and Etiocholanolone glucuronide levels. One metabolite was detected as protective effect: Glycohyocholate levels (Figure 2). The odds ratio (OR) of Epiandrosterone sulfate levels on HER+ breast cancer was estimated to be 1.07 (95% CI = 1.02 ~ 1.10, $p = 0.0013$, Supplementary Table 1) by using the IVW method. Similar results were observed by using three more methods: MR Egger (OR =1.06, 95% CI = 1.01 ~ 1.11, $p = 0.0148$), weighted median (OR = 1.06, 95% CI = 1.02 ~ 1.11, $p = 0.0043$) and weighted mode (OR = 1.06, 95% CI = 1.02 ~ 1.11, $p = 0.0061$). However, except the simple mode (OR = 0.98, 95% CI = 0.95 ~ 1.02, $p = 0.3877$). The OR of 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels on HER+ breast cancer was estimated to be 1.07 (95% CI = 1.03 ~ 1.12, $p = 0.0012$, Supplementary Table 2) by using the IVW method. Similar results were observed by using three more methods: MR Egger (OR =1.06, 95% CI = 1.01 ~ 1.12,

Exposures	Methods	nSNP	p.value	or	OR(95%CI)
Epiandrosterone sulfate levels	MR Egger	24	0.014808268	1.0647511	1.06(1.02 to 1.12)
	Weighted median	24	0.004311199	1.0609928	1.06(1.02 to 1.11)
	Inverse variance weighted	24	0.001323512	1.0650839	1.07(1.02 to 1.11)
	Simple mode	24	0.367823592	1.0775864	1.08(0.92 to 1.26)
	Weighted mode	24	0.006171805	1.0628252	1.06(1.02 to 1.11)
5alpha−androstan−3beta,17beta−diol monosulfate (2) levels	MR Egger	31	0.040851566	1.0624726	1.06(1.01 to 1.12)
	Weighted median	31	0.002542732	1.0762979	1.08(1.03 to 1.13)
	Inverse variance weighted	31	0.001192500	1.0737200	1.07(1.03 to 1.12)
	Simple mode	31	0.171744237	1.1218262	1.12(0.96 to 1.32)
	Weighted mode	31	0.003143001	1.0801207	1.08(1.03 to 1.13)
Glycohyocholate levels	MR Egger	16	0.012756403	0.7394958	0.74(0.60 to 0.91)
	Weighted median	16	0.002628614	0.8218046	0.82(0.72 to 0.93)
	Inverse variance weighted	16	0.000670504	0.8493540	0.85(0.77 to 0.93)
	Simple mode	16	0.039569056	0.7572933	0.76(0.59 to 0.96)
	Weighted mode	16	0.032715273	0.7597164	0.76(0.60 to 0.96)
Etiocholanolone glucuronide levels	MR Egger	20	0.049353726	1.1186659	1.12(1.01 to 1.24)
	Weighted median	20	0.049205417	1.0972658	1.10(1.00 to 1.20)
	Inverse variance weighted	20	0.001294616	1.1198071	1.12(1.05 to 1.20)
	Simple mode	20	0.286486016	1.0944444	1.09(0.93 to 1.29)
	Weighted mode	20	0.063513362	1.1013101	1.10(1.00 to 1.21)

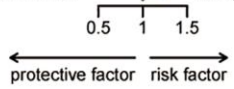


Figure 2. Forest plots showed the causal associations between blood metabolites and HER+ breast cancer by using different methods. Abbreviations: SNP: single nucleotide polymorphism; OR: odds ratio; CI: confidence interval.

p = 0.0409), weighted median (OR = 1.08, 95% CI = 1.03 ~ 1.13, p = 0.0025) and weighted mode (OR = 1.08, 95% CI = 1.03 ~ 1.13, p = 0.0031). But the simple mode (OR = 1.12, 95% CI = 0.96 ~ 1.32, p = 0.1717) did not support this association. The OR of Glycohyocholate levels on HER+ breast cancer was estimated to be 0.85 (95% CI = 0.77 ~ 0.93, p = 0.0007, Supplementary Table 3) by using the IVW method. Similar results were observed by using four more methods: MR Egger (OR = 0.74, 95% CI = 0.60 ~ 0.91, p = 0.0128), weighted median (OR = 0.82, 95% CI = 0.72 ~ 0.94, p = 0.0026), simple mode (OR = 0.76, 95% CI = 0.59 ~ 0.96, p = 0.0396) and weighted mode (OR = 0.76, 95% CI = 0.60 ~ 0.96, p = 0.0327). The OR of Etiocholanolone glucuronide levels on HER+ breast cancer was estimated to be 1.12 (95% CI = 1.05 ~ 1.20, p = 0.0013, Supplementary Table 4) by using the IVW method. Similar results were observed by using two more methods: MR Egger (OR = 1.19, 95% CI = 1.01 ~ 1.24, p = 0.0494) and weighted median (OR = 1.10, 95% CI = 1.00 ~ 1.20, p = 0.0492). However, simple mode (OR = 1.09, 95% CI = 0.93 ~ 1.29, p = 0.2865) and weighted mode (OR = 1.10, 95% CI = 1.00 ~ 1.21, p = 0.0635) did not support this association. These trends are also evident in the forest plots (Supplementary Figure 1A–1D) and scatter plots (Supplementary Figure 2A–2D).

Exploration of the causal effect of metabolites on HER- breast cancer

Sixty-five metabolites were identified to be significant at a p-value of 0.05 using the IVW method; subsequently, setting the p-value to 0.002 revealed the risk effects of four metabolites on HER-negative breast cancer: Vanillic acid glycine levels, Thyroxine levels, N-acetylphenylalanine levels and Glucose-to-mannose ratio. Meanwhile, 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels was detected as a protective effector (Figure 3). The odds ratio (OR) for Vanillic acid glycine levels and the risk of HER- breast cancer was estimated to be 1.14 (95% CI = 1.06 ~ 1.22, p = 0.0003, Supplementary Table 5) using the IVW method. Consistent results were obtained with three other

Exposures	Methods	nSNP	p.value	or	OR(95%CI)
Vanillic acid glycine levels	MR Egger	23	0.005662643	1.1754434	1.18(1.06 to 1.30)
	Weighted median	23	0.001486585	1.1890630	1.19(1.07 to 1.32)
	Inverse variance weighted	23	0.000284828	1.1394710	1.14(1.06 to 1.22)
	Simple mode	23	0.160096138	1.1904636	1.19(0.94 to 1.51)
	Weighted mode	23	0.008081637	1.2066486	1.21(1.06 to 1.37)
Thyroxine levels	MR Egger	21	0.022814092	1.5053019	1.51(1.09 to 2.08)
	Weighted median	21	0.004027232	1.2788773	1.28(1.08 to 1.51)
	Inverse variance weighted	21	0.000406600	1.2631565	1.26(1.11 to 1.44)
	Simple mode	21	0.010607703	1.4878401	1.49(1.13 to 1.96)
	Weighted mode	21	0.010987127	1.4052443	1.41(1.11 to 1.78)
1−palmitoyl−2−linoleoyl−GPI (16:0/18:2) levels	MR Egger	21	0.573468721	0.9441701	0.94(0.78 to 1.15)
	Weighted median	21	0.024695302	0.8723613	0.87(0.77 to 0.98)
	Inverse variance weighted	21	0.000967758	0.8618753	0.86(0.79 to 0.94)
	Simple mode	21	0.013364363	0.7554920	0.76(0.62 to 0.93)
	Weighted mode	21	0.091531781	0.8815685	0.88(0.77 to 1.01)
N−acetylphenylalanine levels	MR Egger	24	0.188790252	1.0732354	1.07(0.97 to 1.19)
	Weighted median	24	0.004282207	1.1323800	1.13(1.04 to 1.23)
	Inverse variance weighted	24	0.000731447	1.1187325	1.12(1.05 to 1.19)
	Simple mode	24	0.418400598	1.0766359	1.08(0.90 to 1.28)
	Weighted mode	24	0.010556455	1.1245105	1.12(1.04 to 1.22)
Glucose−to−mannose ratio	MR Egger	26	0.111773521	1.1415265	1.14(0.98 to 1.34)
	Weighted median	26	0.010763559	1.1749508	1.17(1.04 to 1.33)
	Inverse variance weighted	26	0.000766529	1.1478812	1.15(1.06 to 1.24)
	Simple mode	26	0.074744129	1.2015021	1.20(0.99 to 1.46)
	Weighted mode	26	0.023540992	1.1838512	1.18(1.03 to 1.36)

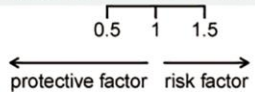


Figure 3. Forest plots showed the causal associations between blood metabolites and HER- breast cancer by using different methods. Abbreviations: SNP: single nucleotide polymorphism; OR: odds ratio; CI: confidence interval.

methods: MR Egger (OR = 1.18, 95% CI = 1.06 ~ 1.30, p = 0.0057), weighted median (OR = 1.19, 95% CI = 1.07 ~ 1.32, p = 0.0015), and weighted mode (OR = 1.21, 95% CI = 1.06 ~ 1.37, p = 0.0080). However, the simple mode did not provide support for this association (OR = 1.19, 95% CI = 0.94 ~ 1.51, p = 0.1601). Thyroxine levels showed an OR of 1.26 (95% CI = 1.11 ~ 1.44, p = 0.0004, Supplementary Table 6) for HER- breast cancer risk using the IVW method. Similar results were observed with four additional methods: MR Egger (OR = 1.51, 95% CI = 1.09 ~ 2.08, p = 0.0228), weighted median (OR = 1.28, 95% CI = 1.08 ~ 1.51, p = 0.0040), simple mode (OR = 1.49, 95% CI = 1.13 ~ 1.96, p = 0.0106), and weighted mode (OR = 1.41, 95% CI = 1.11 ~ 1.78, p = 0.0110). For 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels, the OR for HER- breast cancer risk was estimated as 0.86 (95% CI = 0.79 ~ 0.94, p = 0.0010, Supplementary Table 7) using the IVW method. The weighted median (OR = 0.87, 95% CI = 0.77 ~ 0.98, p = 0.0247) and simple mode (OR = 0.75, 95% CI = 0.62 ~ 0.93, p = 0.0134) also supported this association. However, the MR Egger (OR = 0.94, 95% CI = 0.78 ~ 1.15, p = 0.5735) and weighted mode (OR = 0.88, 95% CI = 0.77 ~ 1.01, p = 0.0915) did not find evidence to support this relationship. N-acetylphenylalanine levels exhibited an OR of 1.12 (95% CI = 1.05 ~ 1.19, p = 0.0007, Supplementary Table 8) for HER- breast cancer risk using the IVW method. Consistent results were obtained with the weighted median (OR = 1.13, 95% CI = 1.04 ~ 1.23, p = 0.0043) and weighted mode (OR = 1.12, 95% CI = 1.04 ~ 1.22, p = 0.0106). However, the MR Egger (OR = 1.07, 95% CI = 0.97 ~ 1.19, p = 0.1888) and simple mode did not support this association (OR = 1.08, 95% CI = 0.90 ~ 1.28, p = 0.4184). The OR for Glucose-to-mannose ratio and HER- breast cancer risk was estimated as 1.15 (95% CI = 1.06 ~ 1.24, p = 0.0008, Supplementary Table 9) using the IVW method. Consistent results were observed with the weighted median (OR = 1.17, 95% CI = 1.04 ~ 1.33, p = 0.0108) and weighted mode (OR = 1.18, 95% CI = 1.03 ~ 1.36, p = 0.0235). However, the MR Egger (OR = 1.14, 95% CI = 0.98 ~ 1.34, p = 0.1118) and simple mode (OR = 1.20, 95% CI = 0.99 ~ 1.46, p = 0.0747) did not support this association. These associations are also noticeable in both the forest plots (Supplementary Figure 1E–1I) and scatter plots (Supplementary Figure 2E–2I).

Exploration of the causal effect of intersecting metabolites on both HER+ and HER- breast cancer

In order to examine the causal effects of common metabolites on both HER+ and HER– breast cancer, we identified a total of 90 metabolites associated with HER+ breast cancer and 65 metabolites associated with HER- breast cancer at a significance level of 0.05 using the IVW method as the primary analysis. By taking the intersection of these sets, we detected two metabolites that were consistently present in both HER+ and HER- breast cancer: Gamma-glutamyl glutamate levels and X-12849 levels (Figure 4A). Interestingly, Gamma-glutamyl glutamate levels were found to act as risk effectors in both HER+ and HER- breast cancer, while X-12849 levels exhibited a protective effect (Figure 4B, 4C). The association between Gamma-glutamyl glutamate levels and HER+ breast cancer risk was estimated to have an odds ratio (OR) of 1.26 (95% CI = 0.67 ~ 1.10, p = 0.0004) using the IVW method. However, other methods such as MR Egger (OR = 0.86, 95% CI = 1.09 ~ 2.08, p = 0.2337), weighted median (OR = 1.09, 95% CI = 0.97 ~ 1.24, p = 0.1567), simple mode (OR = 0.99, 95% CI = 0.79 ~ 1.26, p = 0.9779), and weighted mode (OR = 1.15, 95% CI = 0.97 ~ 1.36, p = 0.1303) did not support this association (Supplementary Table 10). The MR Egger analysis, weighted median, simple mode, and weighted mode did not provide evidence for a causal relationship. Similarly, the OR of X-12849 levels on HER+ breast cancer risk was estimated as 0.92 (95% CI = 0.85 ~ 1.00, p = 0.0479) using the IVW method. However, the MR Egger (OR = 0.85, 95% CI = 0.72 ~ 1.00, p = 0.0677), weighted median (OR = 0.90, 95% CI = 0.80 ~ 1.01, p = 0.0739), simple mode (OR = 1.00, 95% CI = 0.84 ~ 1.19, p = 0.9982), and weighted mode (OR = 0.91, 95% CI = 0.79 ~ 1.04, p = 0.1616) did not support this association (Supplementary Table 11). Regarding HER- breast cancer risk, the OR of Gamma-glutamyl glutamate levels was estimated as 1.15 (95% CI = 1.01 ~ 1.31, p = 0.0309) using the IVW method. However, the MR Egger (OR = 1.01, 95% CI = 0.71 ~ 1.45, p = 0.9380), weighted median (OR = 1.12, 95% CI = 0.95 ~ 1.31, p = 0.1765), simple mode (OR = 1.05, 95% CI = 0.80 ~ 1.37, p = 0.7311), and weighted mode (OR = 1.11, 95% CI = 0.89 ~ 1.37, p = 0.3605) did not support this association (Supplementary Table 12). Similarly, the OR of X-12849 levels on HER- breast cancer risk was estimated as 0.89 (95% CI = 0.81 ~ 0.99, p = 0.0285) using the IVW method. However, the MR Egger (OR = 0.82, 95% CI = 0.67 ~ 1.00, p = 0.0691), weighted median (OR = 0.94, 95% CI = 0.81 ~ 1.08, p = 0.3864), simple mode (OR = 0.96, 95% CI = 0.73 ~ 1.25, p = 0.7448), and weighted mode (OR = 0.95, 95% CI = 0.80 ~ 1.14, p = 0.6093) did not support this association (Supplementary Table 13). These results could also be observed in the forest plots and scatter plots (Supplementary Figure 3).

Sensitivity analysis

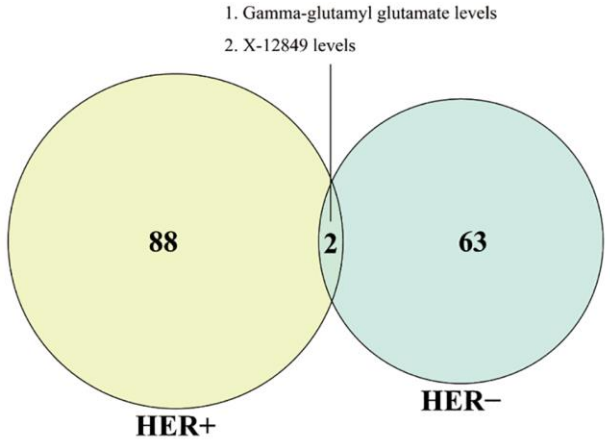
We conducted multiple sensitivity analyses to assess the presence of heterogeneity and pleiotropy in our causal estimates. Cochran's Q-test and MR-PRESSO

test indicated no significant heterogeneity or pleiotropy among the SNPs involved in the causal relationships (Supplementary Tables 14–16). Funnel plots exhibited symmetrical distribution, suggesting no evidence of publication bias across these analyses (Supplementary Figures 4, 5A–5D). Furthermore, the sensitivity analysis, performed through leave-one-out analysis, confirmed the robustness of the causal associations (Supplementary Figures 6, 5E–5H).

DISCUSSION

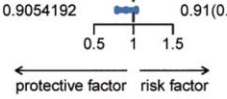
In this study, we conducted separate analyses to investigate the causal associations between 1400

A



B

Exposures	Methods	nSNP	p.value	or	OR(95%CI)
Gamma−glutamyl glutamate levels	MR Egger	25	0.23365369	0.8581873	0.86(0.67 to 1.10)
	Weighted median	25	0.15664982	1.0940811	1.09(0.97 to 1.24)
	Inverse variance weighted	25	0.04408212	1.0967296	1.10(1.00 to 1.20)
	Simple mode	25	0.97787632	0.9966854	1.00(0.79 to 1.26)
	Weighted mode	25	0.13027878	1.1477329	1.15(0.97 to 1.36)
X−12849 levels	MR Egger	22	0.06768462	0.8495871	0.85(0.72 to 1.00)
	Weighted median	22	0.07392522	0.9011699	0.90(0.80 to 1.01)
	Inverse variance weighted	22	0.04785516	0.9204429	0.92(0.85 to 1.00)
	Simple mode	22	0.99818711	1.0002017	1.00(0.84 to 1.19)
	Weighted mode	22	0.16157859	0.9054192	0.91(0.79 to 1.04)



C

Exposures	Methods	nSNP	p.value	or	OR(95%CI)
Gamma−glutamyl glutamate levels	MR Egger	25	0.93797143	1.0143505	1.01(0.71 to 1.45)
	Weighted median	25	0.17645095	1.1150387	1.12(0.95 to 1.31)
	Inverse variance weighted	25	0.03088075	1.1532512	1.15(1.01 to 1.31)
	Simple mode	25	0.73106640	1.0487527	1.05(0.80 to 1.37)
	Weighted mode	25	0.36046882	1.1063328	1.11(0.89 to 1.37)
X−12849 levels	MR Egger	22	0.06907025	0.8202025	0.82(0.67 to 1.00)
	Weighted median	22	0.38640770	0.9389077	0.94(0.81 to 1.08)
	Inverse variance weighted	22	0.02852164	0.8943401	0.89(0.81 to 0.99)
	Simple mode	22	0.74483891	0.9565177	0.96(0.73 to 1.25)
	Weighted mode	22	0.60926806	0.9535110	0.95(0.80 to 1.14)

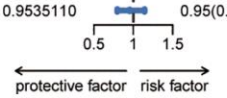


Figure 4. (**A**) Venn diagram showed the intersection of blood metabolites that are causally involved in HER+ and HER-breast cancers. (**B**) Forest plots showed the causal associations between blood metabolites (Gamma-glutamyl glutamate levels and X-12849 levels) and HER+ breast cancer by using different methods. (**C**) Forest plots showed the causal associations between blood metabolites (Gamma-glutamyl glutamate levels and X-12849 levels) and HER- breast cancer by using different methods. Abbreviations: SNP: single nucleotide polymorphism; OR: odds ratio; CI: confidence interval.

metabolic factors, comprising 1,091 blood metabolites and 309 metabolite ratios, and HER+ and HER– breast cancer. We utilized large publicly available genetic data to explore these associations comprehensively. By examining a wide range of metabolic factors, including individual metabolites and their ratios, we aimed to gain insights into the potential causal relationships between these factors and breast cancer subtypes. Due to the large number of metabolites identified at a significance level of $p < 0.05$ in the result, we made the decision to narrow down our focus by selecting only the four to five metabolites with the smallest *p*-values ($p < 0.002$) within each isoform for detailed presentation and discussion. This approach allows us to concentrate on the most statistically significant metabolites and facilitate a more targeted analysis. The findings of our study reveal several significant associations with respect to the causal risk factors for different breast cancer subtypes. For HER+ breast cancer, elevated levels of epiandrosterone sulfate, 5alpha-androstan-3beta,17beta-diol monosulfate (2), and etiocholanolone glucuronide were found to be causally associated with an increased risk. Conversely, increased glycohyocholate levels were found to be causally associated with a decreased risk for HER- breast cancer. In the case of HER- breast cancer, we identified several causal risk factors. These include elevated levels of vanillic acid glycine, thyroxine, N-acetylphenylalanine, and a higher glucose-to-mannose ratio. On the other hand, we observed that increased levels of 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) were causally associated with a decreased risk for HER- breast cancer. Furthermore, when considering the intersection of both breast cancer subtypes, we found a causal association between elevated levels of Gamma-glutamyl glutamate and an increased risk for both HER+ and HER– breast cancer. Additionally, increased levels of X-12849 were causally associated with a decreased risk for both subtypes.

Metabolomics analysis currently relies on a range of detection techniques, notably Liquid Chromatography with tandem mass spectrometry (LC-MS/MS) and others. These methodologies have identified numerous metabolites linked to tumor development, which are being extensively studied as potential therapeutic targets [30, 31]. A prospective study revealed a strong correlation between plasma concentrations of metabolites and breast cancer risk. Specifically, concentrations of arginine, asparagine, and phosphatidylcholine were found to be negatively associated with breast cancer risk, while acylcarnitines exhibited a positive association [31]. In the quest for diagnostic biomarkers in early breast cancer, Wei et al. leveraged untargeted liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QTOF-MS) data to uncover a spectrum of promising metabolites including ethyl (R)-3-hydroxyhexanoate, caprylic acid, hypoxanthine, and others [32]. These findings have provided valuable insights into the potential pathogenesis of early-stage breast cancer. Aligned with this objective, our study also aimed to elucidate metabolites that play a causal role in breast carcinogenesis.

Among the metabolites causally linked to HER+ breast cancer, epiandrosterone sulfate and 5alpha-androstan-3beta-17beta-diol play crucial roles in androgen metabolism [33, 34]. Androgens have been extensively researched and validated for their involvement in various tumorigenesis processes. However, the specific relationships between epiandrosterone sulfate, 5alpha-androstan-3beta-17beta-diol monosulfate, and the risk of HER+ breast cancer remain inconclusive and warrant further investigation. Understanding the significant impact of androgen excess on diverse breast cancer subtypes holds substantial clinical implications for treatment and prevention [35]. Therefore, epiandrosterone sulfate and 5alpha-androstan-3beta-17beta-diol monosulfate show promise as potential therapeutic targets for HER+ breast cancer. Etiocholanolone glucuronide (Etio-G) is a primary testosterone metabolite, alongside androsterone glucuronide (ADT-G), Testosterone glucuronide (TG), and dihydrotestosterone glucuronide (DHTG). The liver and intestines are key sites for Etio-G formation, which is subsequently released into the bloodstream [36]. While previous studies indicate that elevated androsterone-glucuronide levels are linked to an increased risk of non-serous ovarian cancer, the association between Etio-G and cancer remains largely unexplored. Our findings shed new light on the relationship between Etio-G and breast cancer. Belonging to the primary bile acid (BA) family, glycohyocholate has been shown to significantly reduce the risk of nonalcoholic fatty liver disease (NAFLD) [37]. Although bile acids were traditionally viewed as pro-carcinogenic agents (e.g., esophageal cancer), recent evidence suggests that physiological concentrations of bile acids possess anti-cancer properties in certain cancers such as prostate, ovarian, and breast cancer [38]. Notably, breast cancer patients exhibit reduced hepatic bile acid production, reflected in lower serum and fecal bile acid levels. Furthermore, the transformation of bile acids into secondary forms by gut bacteria is also diminished [38–40]. Our discovery that Glycohyocholate acts protectively in HER+ breast cancer aligns with existing research and implies its potential utility as both a diagnostic tool and therapeutic target for breast cancer.

For HER- breast cancer, our study identified five metabolites with causal links. Our findings indicate that the level of vanillic acid glycine may lean towards acting as a risk factor for tumor development. The focus of current study primarily centers around vanillic acid.

The impact of vanillic acid on tumors appears to be multifaceted. Zhu et al. demonstrated its potential as an antitumor agent by activating the stimulator of interferon genes (STING) signaling pathway in macrophages [41]. In colon cancer cells, vanillic acid exerts inhibitory effects on HIF-1alpha expression through the mTOR/p70S6K/4E-BP1 and Raf/MEK/ERK pathways [42]. However, Ujlaki et al. observed hyperproliferative effects when vanillic acid was administered to the mouse breast cancer cell line 4T1 [43]. Epidemiological studies have established associations between thyroid function and breast cancer, suggesting that hormones can play a supportive role in breast cancer development. L-thyroxine (T4) has been demonstrated to induce the proliferation of various types of cancer. This T4-induced activity is facilitated by a cell surface receptor located on the extracellular domain of integrin αvβ3. Subsequently, the T4 signal is transduced by mitogen-activated protein kinase (MAPK/ERK1/2) or phosphatidylinositol 3-kinase (PI3-K) pathways, leading to gene transcription associated with cancer [44]. Notably, T4 has been also identified as a proliferative factor for breast cancer cells in laboratory experiments [45, 46]. Additionally, T4 upregulates the accumulation of checkpoint programmed death-ligand 1 (PD-L1) in cancer cells [47]. However, uncertainties remain regarding whether circulating endogenous T4 levels act as a risk factor for breast cancer among individuals with normal thyroid function but a positive family history. Further research is crucial to unravel the intricate relationship between thyroid hormone levels and breast cancer risk in this specific subgroup of patients.

Currently, there is a scarcity of research on 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2). However, a study by Poupore et al. delved into metabolite distinctions between patients with ischemic stroke and control subjects. In the female group, a total of 1322 biochemicals were identified, comprising 1062 named compounds with known identities and 260 unnamed compounds with unidentified structural features. Notably, among these compounds, 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) displayed significant differences and might hold promise as a diagnostic indicator for ischemic stroke [48]. Our own research findings suggest that 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) functions as a protective element against HER- breast malignancies. Therefore, further investigation into this metabolite could offer a fruitful path for future exploration. While limited information exists on the association between N-acetylphenylalanine and cancer, insights from a study conducted by Tsamouri suggest potential implications. The study suggested that urinary N-acetylphenylalanine levels could serve as a diagnostic marker for uroepithelial carcinoma of the bladder in dogs [49]. Nevertheless, further investigations are necessary to fully grasp the role of N-acetylphenylalanine in cancer development and progression in humans. In our study, the Glucose-to-mannose ratio was identified as a risk factor for HER- breast cancer. This suggests that an increased glucose level or decreased mannose level would elevate the ratio. Glucose metabolism significantly supports tumor cell growth and proliferation [50]. Conversely, mannose ($C_6H_{12}O_6$) has demonstrated tumor growth inhibition in both *in vitro* and *in vivo* studies [50]. Regarding diagnosis, the serum free glucose to mannose ratio holds promise as a potential biomarker for ovarian cancer [51]. Noteworthy results showed a 49% reduction in recurrence risk and a 56% decrease in death risk for esophageal adenocarcinoma (EAC) cases among patients with elevated mannose levels compared to those with lower levels [52, 53]. These findings underscore the potential utility of serum mannose as a diagnostic or prognostic tool for various tumor types.

According to our study, Gamma-glutamylglutamate levels and X-12849 levels are identified as common factors with a causal relationship in both HER+ and HER– breast cancer subtypes. Gamma-glutamylglutamate is a dipeptide formed by the condensation of the gamma-carboxy group of glutamic acid with the amino group of another glutamic acid. Notably, a metabolomic analysis involving 1812 Finnish men and Huang's Cox proportional hazards regression model revealed an association between gamma-glutamylglutamate and an increased risk of prostate cancer-specific mortality [54]. However, whether Gamma-glutamylglutamate acts as a risk factor for breast cancer remains uncertain. In our investigation, we found supporting evidence for a causal link between Gamma-glutamylglutamate and both HER+ and HER– breast cancers. This suggests that Gamma-glutamylglutamate may play a significant role in breast cancer, warranting further exploration and clarification through additional studies. In the realm of untargeted metabolomics, identifying metabolites continues to pose a significant challenge. Typically, metabolites lacking a known chemical structure are denoted with the prefix "X-" followed by a number [55]. Based on our research outcomes, we observed that X-12849 levels act as protective elements against both HER+ and HER– breast carcinogenesis. These findings not only enhance our comprehension of the impact of X-12849 but also offer valuable insights into the connection between unidentified metabolites and human diseases.

Our study employed a two-sample MR analysis, utilizing data from large-scale GWAS cohorts to ensure statistical robustness. This approach allowed us to minimize the impact of confounding factors, such as horizontal pleiotropy and related variables, on our

results. However, it is important to acknowledge several limitations in our study. Firstly, despite conducting multiple sensitivity analyses, fully assessing the presence of horizontal pleiotropy remains challenging. This potential source of bias should be considered when interpreting the results. One major problem is that we found evidence for pleiotropy in our MR of Gamma-glutamyl glutamate levels on HER+ breast cancer, but pleiotropy was eliminated by MR PRESSO. Secondly, due to the lack of individual-level data (e.g., stage, grade, and hormone receptor status), we were unable to perform further stratified analyses within the population. This limits our ability to draw conclusions specific to certain subgroups. Thirdly, it is important to note that in our study, we opted to use a more lenient threshold and did not correct for multiple testing when evaluating the results. While this approach aimed to maximize the detection of potential associations, it also introduces the possibility of increased false positives. Therefore, caution should be exercised when interpreting these findings, and further validation studies are necessary to confirm the observed associations. Lastly, it is worth noting that the external validity of our findings may be limited since the data source for this study primarily consisted of a European population. Generalizing the results to other populations should be done cautiously. Despite these limitations, our study provides valuable insights into the causal effects of hub metabolites on both HER+ and HER− breast cancer, highlighting the need for further research in this area.

In summary, our MR analysis has revealed significant causal relationships between various metabolites and breast cancer characterized by HER+ or HER– expression. This finding not only sheds light on the intricate interactions between metabolites and breast carcinogenesis but also advances our comprehension of the realm of breast cancer and metabolomics. Importantly, certain metabolites that have been overlooked in terms of their association with tumors indicate promising avenues for further investigation. Therefore, further research is essential to elucidate the intricate mechanisms involving metabolites in breast carcinogenesis and to evaluate the feasibility of clinical interventions. These endeavors will not only yield new insights into the origins of breast cancer but also enhance treatment strategies.

AUTHOR CONTRIBUTIONS

JY, LG, SC and GSR conceived and designed the studies. JY, HYF and LG performed the data acquisition and analysed the experiments. JY and HYF wrote the manuscript. JY and HYF prepared figures and tables. All authors approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank the participants of all GWAS cohorts included in the present work, the investigators of the GWAS Catalog produced by EMBL-EBI/NHGRI-EBI and the FinnGen consortium for sharing the GWAS summary statistics.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

ETHICAL STATEMENT

In this Mendelian randomization study, ethical approval is not required due to the use of publicly available genetic data.

FUNDING

This work was supported by the Natural Science Foundation of Chongqing (No. cstc2020jcyj-msxmX0367).

REFERENCES

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023; 73:17–48. https://doi.org/10.3322/caac.21763 PMID:36633525

2. Sirhan Z, Thyagarajan A, Sahu RP. The efficacy of tucatinib-based therapeutic approaches for HER2-positive breast cancer. Mil Med Res. 2022; 9:39. https://doi.org/10.1186/s40779-022-00401-3 PMID:35820970

3. Drago JZ, Ferraro E, Abuhadra N, Modi S. Beyond HER2: Targeting the ErbB receptor family in breast cancer. Cancer Treat Rev. 2022; 109:102436. https://doi.org/10.1016/j.ctrv.2022.102436 PMID:35870237

4. Olsen DA, Ostergaard B, Bokmand S, Wamberg PA, Jakobsen EH, Jakobsen A, Brandslund I. HER1-4 protein concentrations in normal breast tissue from breast cancer patients are expressed by the same profile as in the malignant tissue. Clin Chem Lab Med. 2009; 47:977–84. https://doi.org/10.1515/CCLM.2009.214 PMID:19548848

5. Harbeck N, Gnant M. Breast cancer. Lancet. 2017; 389:1134–50. https://doi.org/10.1016/S0140-6736(16)31891-8 PMID:27865536

6. Brockhoff G. "Shedding" light on HER4 signaling in

normal and malignant breast tissues. Cell Signal. 2022; 97:110401.
https://doi.org/10.1016/j.cellsig.2022.110401
PMID:35820544

7. Wege AK, Chittka D, Buchholz S, Klinkhammer-Schalke M, Diermeier-Daucher S, Zeman F, Ortmann O, Brockhoff G. HER4 expression in estrogen receptor-positive breast cancer is associated with decreased sensitivity to tamoxifen treatment and reduced overall survival of postmenopausal women. Breast Cancer Res. 2018; 20:139.
https://doi.org/10.1186/s13058-018-1072-1
PMID:30458882

8. Vaezi MA, Safizadeh B, Eghtedari AR, Ghorbanhosseini SS, Rastegar M, Salimi V, Tavakoli-Yaraki M. 15-Lipoxygenase and its metabolites in the pathogenesis of breast cancer: A double-edged sword. Lipids Health Dis. 2021; 20:169.
https://doi.org/10.1186/s12944-021-01599-2
PMID:34838055

9. Moore SC, Matthews CE, Ou Shu X, Yu K, Gail MH, Xu X, Ji BT, Chow WH, Cai Q, Li H, Yang G, Ruggieri D, Boyd-Morin J, et al. Endogenous Estrogens, Estrogen Metabolites, and Breast Cancer Risk in Postmenopausal Chinese Women. J Natl Cancer Inst. 2016; 108:djw103.
https://doi.org/10.1093/jnci/djw103
PMID:27193440

10. Zhao H, Shen J, Moore SC, Ye Y, Wu X, Esteva FJ, Tripathy D, Chow WH. Breast cancer risk in relation to plasma metabolites among Hispanic and African American women. Breast Cancer Res Treat. 2019; 176:687–96.
https://doi.org/10.1007/s10549-019-05165-4
PMID:30771047

11. Xiao Y, Ma D, Yang YS, Yang F, Ding JH, Gong Y, Jiang L, Ge LP, Wu SY, Yu Q, Zhang Q, Bertucci F, Sun Q, et al. Comprehensive metabolomics expands precision medicine for triple-negative breast cancer. Cell Res. 2022; 32:477–90.
https://doi.org/10.1038/s41422-022-00614-0
PMID:35105939

12. Sekula P, Del Greco M F, Pattaro C, Köttgen A. Mendelian Randomization as an Approach to Assess Causality Using Observational Data. J Am Soc Nephrol. 2016; 27:3253–65.
https://doi.org/10.1681/ASN.2016010098
PMID:27486138

13. Emdin CA, Khera AV, Kathiresan S. Mendelian Randomization. JAMA. 2017; 318:1925–6.
https://doi.org/10.1001/jama.2017.17219
PMID:29164242

14. Burgess S, Thompson SG, and CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol. 2011; 40:755–64.
https://doi.org/10.1093/ije/dyr036
PMID:21414999

15. Chen Y, Lu T, Pettersson-Kymmer U, Stewart ID, Butler-Laporte G, Nakanishi T, Cerani A, Liang KYH, Yoshiji S, Willett JDS, Su CY, Raina P, Greenwood CMT, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. Nat Genet. 2023; 55:44–53.
https://doi.org/10.1038/s41588-022-01270-1
PMID:36635386

16. Raina P, Wolfson C, Kirkland S, Griffith LE, Balion C, Cossette B, Dionne I, Hofer S, Hogan D, van den Heuvel ER, Liu-Ambrose T, Menec V, Mugford G, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). Int J Epidemiol. 2019; 48:1752–53j.
https://doi.org/10.1093/ije/dyz173
PMID:31633757

17. Liu X, Tong X, Zou Y, Lin X, Zhao H, Tian L, Jie Z, Wang Q, Zhang Z, Lu H, Xiao L, Qiu X, Zi J, et al. Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. Nat Genet. 2022; 54:52–61.
https://doi.org/10.1038/s41588-021-00968-y
PMID:34980918

18. Yu XH, Yang YQ, Cao RR, Bo L, Lei SF. The causal role of gut microbiota in development of osteoarthritis. Osteoarthritis Cartilage. 2021; 29:1741–50.
https://doi.org/10.1016/j.joca.2021.08.003
PMID:34425228

19. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, Davey Smith G, Sterne JA. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012; 21:223–42.
https://doi.org/10.1177/0962280210394459
PMID:21216802

20. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016; 40:304–14.
https://doi.org/10.1002/gepi.21965
PMID:27061298

21. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017; 46:1985–98.
https://doi.org/10.1093/ije/dyx102
PMID:29040600

22. Xiang Y, Zhang C, Wang J, Cheng Y, Wang L, Tong Y,

Yan D. Identification of host gene-microbiome associations in colorectal cancer patients using mendelian randomization. J Transl Med. 2023; 21:535.
https://doi.org/10.1186/s12967-023-04335-9
PMID:37563724

23. Yuan S, Titova OE, Zhang K, Gou W, Schillemans T, Natarajan P, Chen J, Li X, Åkesson A, Bruzelius M, Klarin D, Damrauer SM, Larsson SC. Plasma protein and venous thromboembolism: prospective cohort and mendelian randomisation analyses. Br J Haematol. 2023; 201:783–92.
https://doi.org/10.1111/bjh.18679
PMID:36734038

24. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013; 37:658–65.
https://doi.org/10.1002/gepi.21758
PMID:24114802

25. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, and EPIC- InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015; 30:543–52.
https://doi.org/10.1007/s10654-015-0011-z
PMID:25773750

26. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017; 32:377–89.
https://doi.org/10.1007/s10654-017-0255-x
PMID:28527048

27. 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.
https://doi.org/10.1093/ije/dyv080
PMID:26050253

28. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018; 50:693–8.
https://doi.org/10.1038/s41588-018-0099-7
PMID:29686387

29. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, Evans DM, Smith GD. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep. 2017; 4:330–45.
https://doi.org/10.1007/s40471-017-0128-6
PMID:29226067

30. Yang T, Hui R, Nouws J, Sauler M, Zeng T, Wu Q. Untargeted metabolomics analysis of esophageal squamous cell cancer progression. J Transl Med. 2022; 20:127.
https://doi.org/10.1186/s12967-022-03311-z
PMID:35287685

31. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, Ferrari P, Romieu I, Onland-Moret NC, Weiderpass E, Dahm CC, Overvad K, Olsen A, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med. 2019; 17:178.
https://doi.org/10.1186/s12916-019-1408-4
PMID:31547832

32. Wei Y, Jasbi P, Shi X, Turner C, Hrovat J, Liu L, Rabena Y, Porter P, Gu H. Early Breast Cancer Detection Using Untargeted and Targeted Metabolomics. J Proteome Res. 2021; 20:3124–33.
https://doi.org/10.1021/acs.jproteome.1c00019
PMID:34033488

33. Vehof J, Hysi PG, Hammond CJ. A Metabolome-Wide Study of Dry Eye Disease Reveals Serum Androgens as Biomarkers. Ophthalmology. 2017; 124:505–11.
https://doi.org/10.1016/j.ophtha.2016.12.011
PMID:28139245

34. Sun L, Guo D, Jia Y, Shi M, Yang P, Wang Y, Liu F, Chen GC, Zhang Y, Zhu Z. Association between Human Blood Metabolome and the Risk of Alzheimer's Disease. Ann Neurol. 2022; 92:756–67.
https://doi.org/10.1002/ana.26464
PMID:35899678

35. Secreto G, Girombelli A, Krogh V. Androgen excess in breast cancer development: implications for prevention and treatment. Endocr Relat Cancer. 2019; 26:R81–94.
https://doi.org/10.1530/ERC-18-0429
PMID:30403656

36. Li CY, Basit A, Gupta A, Gáborik Z, Kis E, Prasad B. Major glucuronide metabolites of testosterone are primarily transported by MRP2 and MRP3 in human liver, intestine and kidney. J Steroid Biochem Mol Biol. 2019; 191:105350.
https://doi.org/10.1016/j.jsbmb.2019.03.027
PMID:30959153

37. Caussy C, Hsu C, Singh S, Bassirian S, Kolar J, Faulkner C, Sinha N, Bettencourt R, Gara N, Valasek MA, Schnabl B, Richards L, Brenner DA, et al. Serum bile acid patterns are associated with the presence of NAFLD in twins, and dose-dependent changes with increase in fibrosis stage in patients with biopsy-proven NAFLD. Aliment Pharmacol Ther. 2019; 49:183–93.
https://doi.org/10.1111/apt.15035
PMID:30506692

38. Režen T, Rozman D, Kovács T, Kovács P, Sipos A, Bai P,

Mikó E. The role of bile acids in carcinogenesis. Cell Mol Life Sci. 2022; 79:243. https://doi.org/10.1007/s00018-022-04278-2 PMID:35429253

39. Mikó E, Vida A, Kovács T, Ujlaki G, Trencsényi G, Márton J, Sári Z, Kovács P, Boratkó A, Hujber Z, Csonka T, Antal-Szalmás P, Watanabe M, et al. Lithocholic acid, a bacterial metabolite reduces breast cancer cell proliferation and aggressiveness. Biochim Biophys Acta Bioenerg. 2018; 1859:958–74. https://doi.org/10.1016/j.bbabio.2018.04.002 PMID:29655782

40. Murray WR, Blackwood A, Calman KC, MacKay C. Faecal bile acids and clostridia in patients with breast cancer. Br J Cancer. 1980; 42:856–60. https://doi.org/10.1038/bjc.1980.333 PMID:7459219

41. Zhu M, Tang X, Zhu Z, Gong Z, Tang W, Hu Y, Cheng C, Wang H, Sarwar A, Chen Y, Liu F, Huo J, Wang X, Zhang Y. STING activation in macrophages by vanillic acid exhibits antineoplastic potential. Biochem Pharmacol. 2023; 213:115618. https://doi.org/10.1016/j.bcp.2023.115618 PMID:37211172

42. Gong J, Zhou S, Yang S. Vanillic Acid Suppresses HIF-1α Expression via Inhibition of mTOR/p70S6K/4E-BP1 and Raf/MEK/ERK Pathways in Human Colon Cancer HCT116 Cells. Int J Mol Sci. 2019; 20:465. https://doi.org/10.3390/ijms20030465 PMID:30678221

43. Ujlaki G, Kovács T, Vida A, Kókai E, Rauch B, Schwarcz S, Mikó E, Janka E, Sipos A, Hegedűs C, Uray K, Nagy P, Bai P. Identification of Bacterial Metabolites Modulating Breast Cancer Cell Proliferation and Epithelial-Mesenchymal Transition. Molecules. 2023; 28:5898. https://doi.org/10.3390/molecules28155898 PMID:37570868

44. Davis PJ, Lin HY, Hercbergs A, Mousa SA. Actions of L-thyroxine (T4) and Tetraiodothyroacetic Acid (Tetrac) on Gene Expression in Thyroid Cancer Cells. Genes (Basel). 2020; 11:755. https://doi.org/10.3390/genes11070755 PMID:32645835

45. Hercbergs A, Mousa SA, Leinung M, Lin HY, Davis PJ. Thyroid Hormone in the Clinic and Breast Cancer. Horm Cancer. 2018; 9:139–43. https://doi.org/10.1007/s12672-018-0326-9 PMID:29441459

46. Tang HY, Lin HY, Zhang S, Davis FB, Davis PJ. Thyroid hormone causes mitogen-activated protein kinase-dependent phosphorylation of the nuclear estrogen receptor. Endocrinology. 2004; 145:3265–72. https://doi.org/10.1210/en.2004-0308 PMID:15059947

47. Wang K, Chen YF, Yang YSH, Huang HM, Lee SY, Shih YJ, Li ZL, Whang-Peng J, Lin HY, Davis PJ. The power of heteronemin in cancers. J Biomed Sci. 2022; 29:41. https://doi.org/10.1186/s12929-022-00816-z PMID:35705962

48. Poupore N, Chosed R, Arce S, Rainer R, Goodwin RL, Nathaniel TI. Metabolomic Profiles of Men and Women Ischemic Stroke Patients. Diagnostics (Basel). 2021; 11:1786. https://doi.org/10.3390/diagnostics11101786 PMID:34679483

49. Tsamouri MM, Durbin-Johnson BP, Culp WTN, Palm CA, Parikh M, Kent MS, Ghosh PM. Untargeted Metabolomics Identify a Panel of Urinary Biomarkers for the Diagnosis of Urothelial Carcinoma of the Bladder, as Compared to Urolithiasis with or without Urinary Tract Infection in Dogs. Metabolites. 2022; 12:200. https://doi.org/10.3390/metabo12030200 PMID:35323643

50. Tang M, Ren X, Fu C, Ding M, Meng X. Regulating glucose metabolism using nanomedicines for cancer therapy. J Mater Chem B. 2021; 9:5749–64. https://doi.org/10.1039/d1tb00218j PMID:34196332

51. Chen Y, Yao Q, Zhang L, Zeng P. HPLC for simultaneous quantification of free mannose and glucose concentrations in serum: use in detection of ovarian cancer. Front Chem. 2023; 11:1289211. https://doi.org/10.3389/fchem.2023.1289211 PMID:38025059

52. Gu J, Liang D, Pierzynski JA, Zheng L, Ye Y, Zhang J, Ajani JA, Wu X. D-mannose: a novel prognostic biomarker for patients with esophageal adenocarcinoma. Carcinogenesis. 2017; 38:162–7. https://doi.org/10.1093/carcin/bgw207 PMID:28062409

53. Jin H, Liu X, Liu HX. Biological function, regulatory mechanism, and clinical application of mannose in cancer. Biochim Biophys Acta Rev Cancer. 2023; 1878:188970. https://doi.org/10.1016/j.bbcan.2023.188970 PMID:37657682

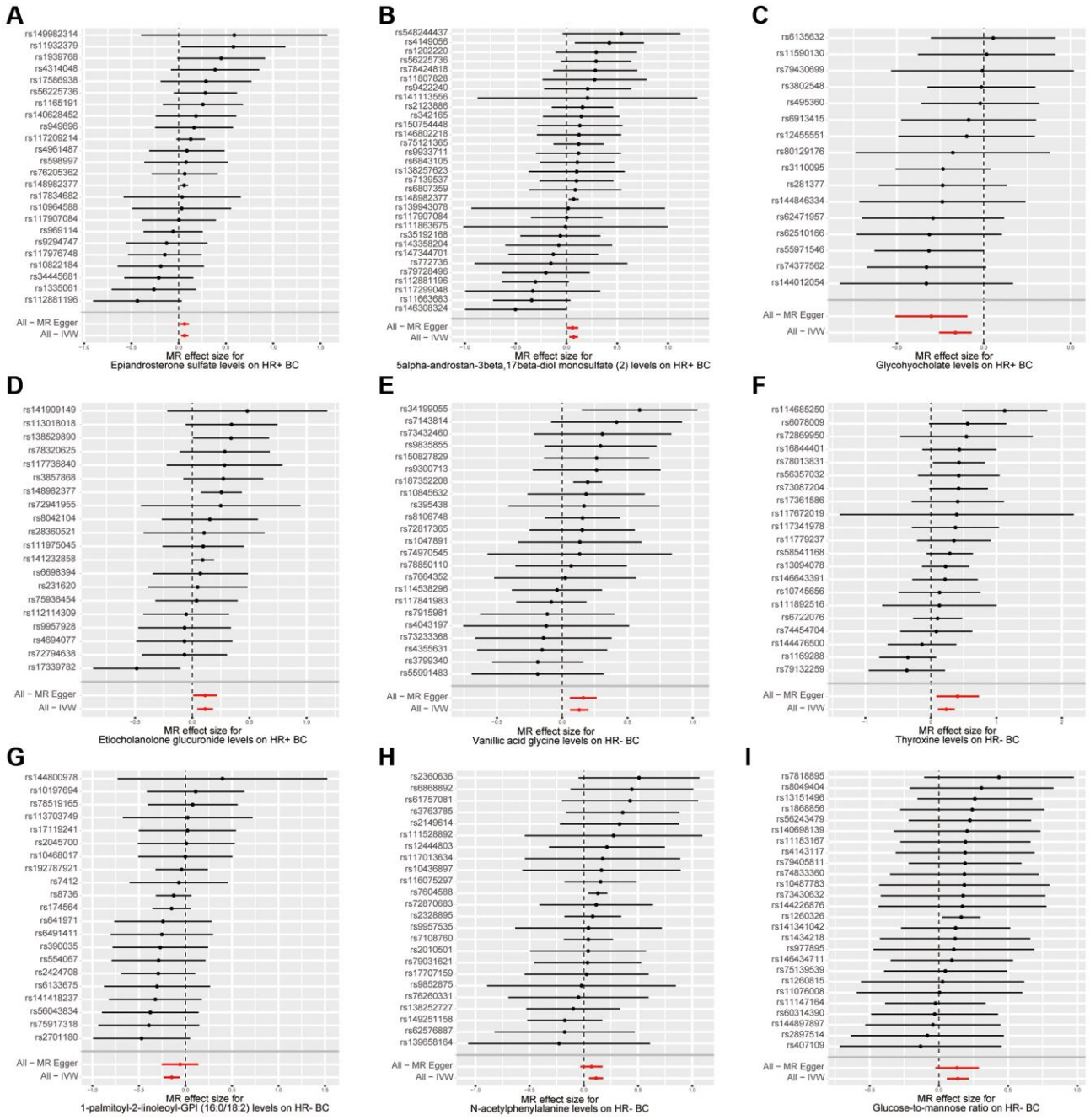
54. Huang J, Zhao B, Weinstein SJ, Albanes D, Mondul AM. Metabolomic profile of prostate cancer-specific survival among 1812 Finnish men. BMC Med. 2022; 20:362. https://doi.org/10.1186/s12916-022-02561-4 PMID:36280842

55. Vike NL, Bari S, Stetsiv K, Talavage TM, Nauman EA,

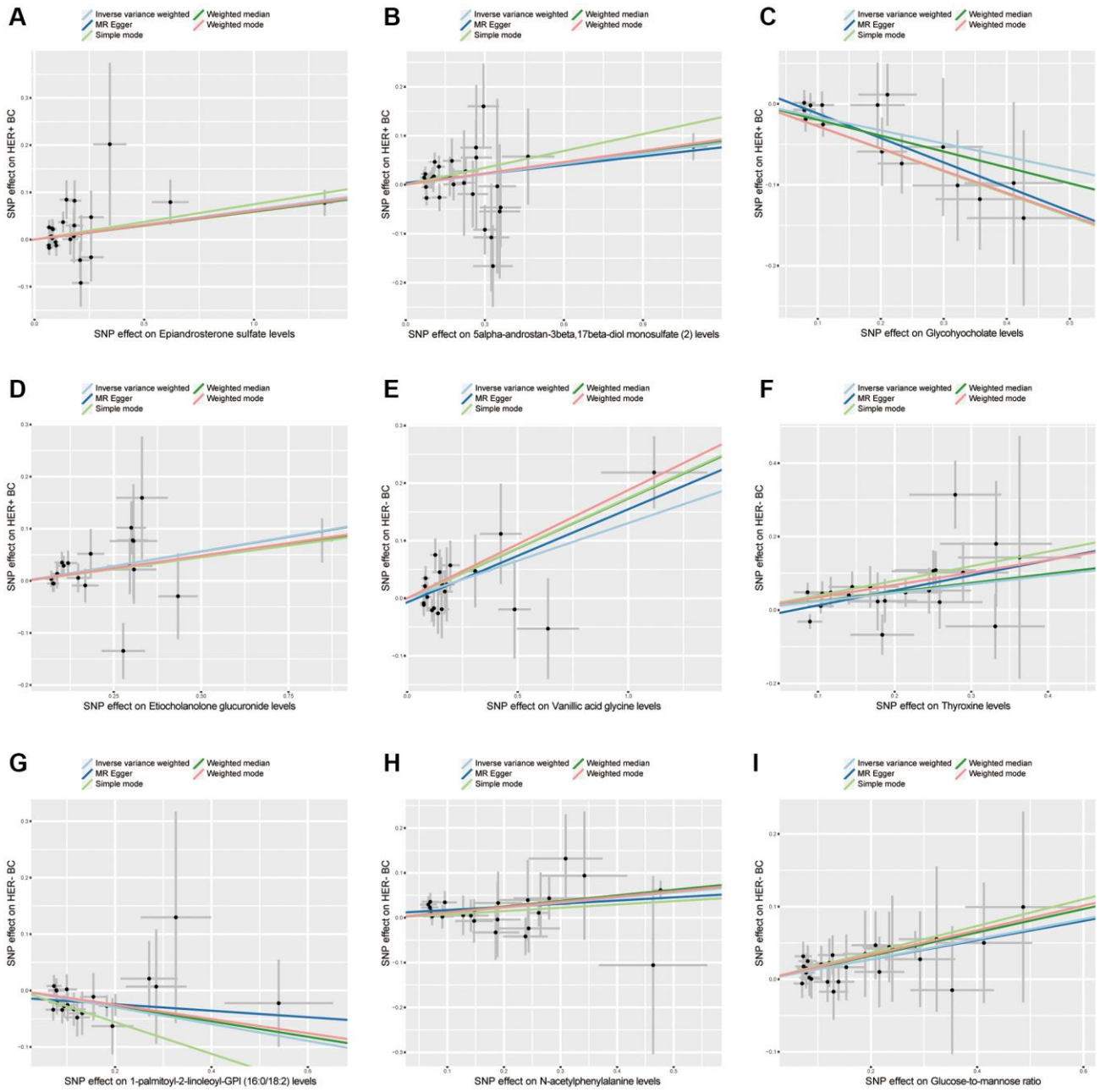
Papa L, Slobounov S, Breiter HC, Cornelis MC. Metabolomic response to collegiate football participation: Pre- and Post-season analysis. Sci Rep. 2022; 12:3091.
https://doi.org/10.1038/s41598-022-07079-6
PMID:35197541

SUPPLEMENTARY MATERIALS

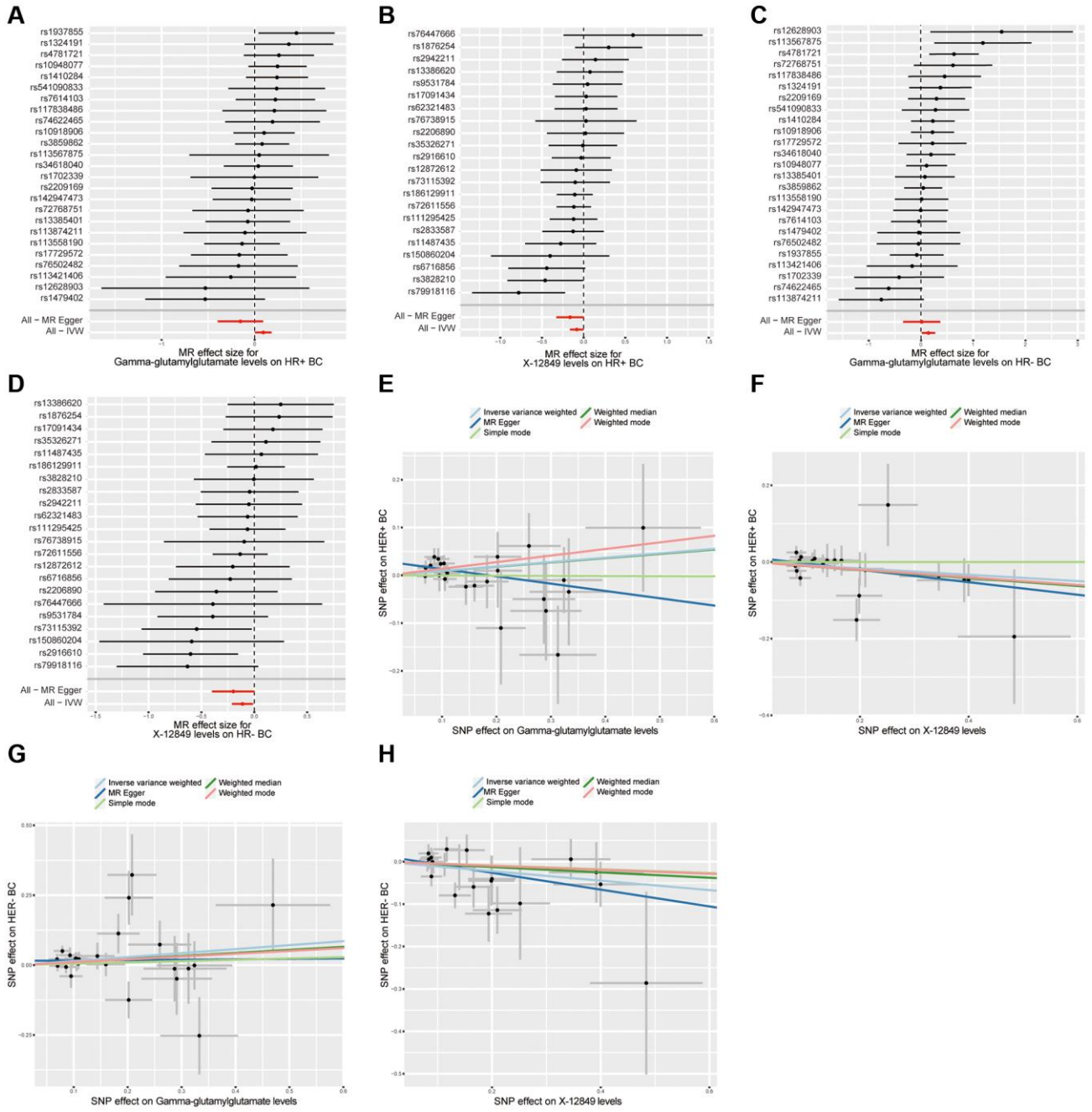
Supplementary Figures



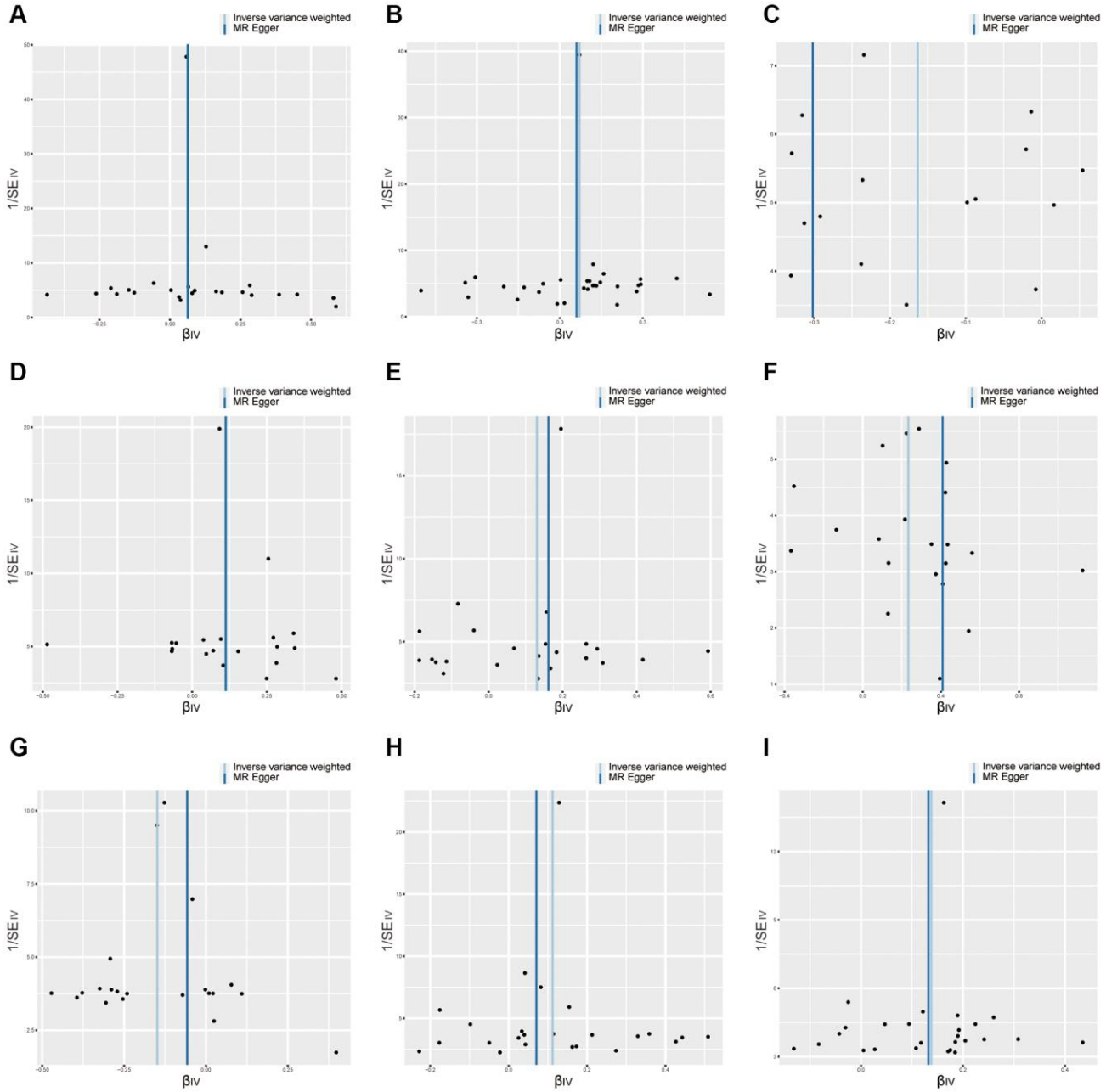
Supplementary Figure 1. Forest plots of SNPs associated with metabolic factors and the risk of HER+/HER− BC. (**A**) Epiandrosterone sulfate levels on HER+ BC. (**B**) 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels on HER+ BC. (**C**) Glycohyocholate levels on HER+ BC. (**D**) Etiocholanolone glucuronide levels on HER+ BC. (**E**) Vanillic acid glycine levels on HER- BC. (**F**) Thyroxine levels on HER- BC. (**G**) 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels on HER-BC. (**H**) N-acetylphenylalanine levels on HER- BC. (**I**) Glucose-to-mannose ratio on HER- BC. Abbreviations: SNP: single nucleotide polymorphisms; HER: human epidermal growth factor receptor; BC: breast cancer; IVW: inverse-variance weighted.



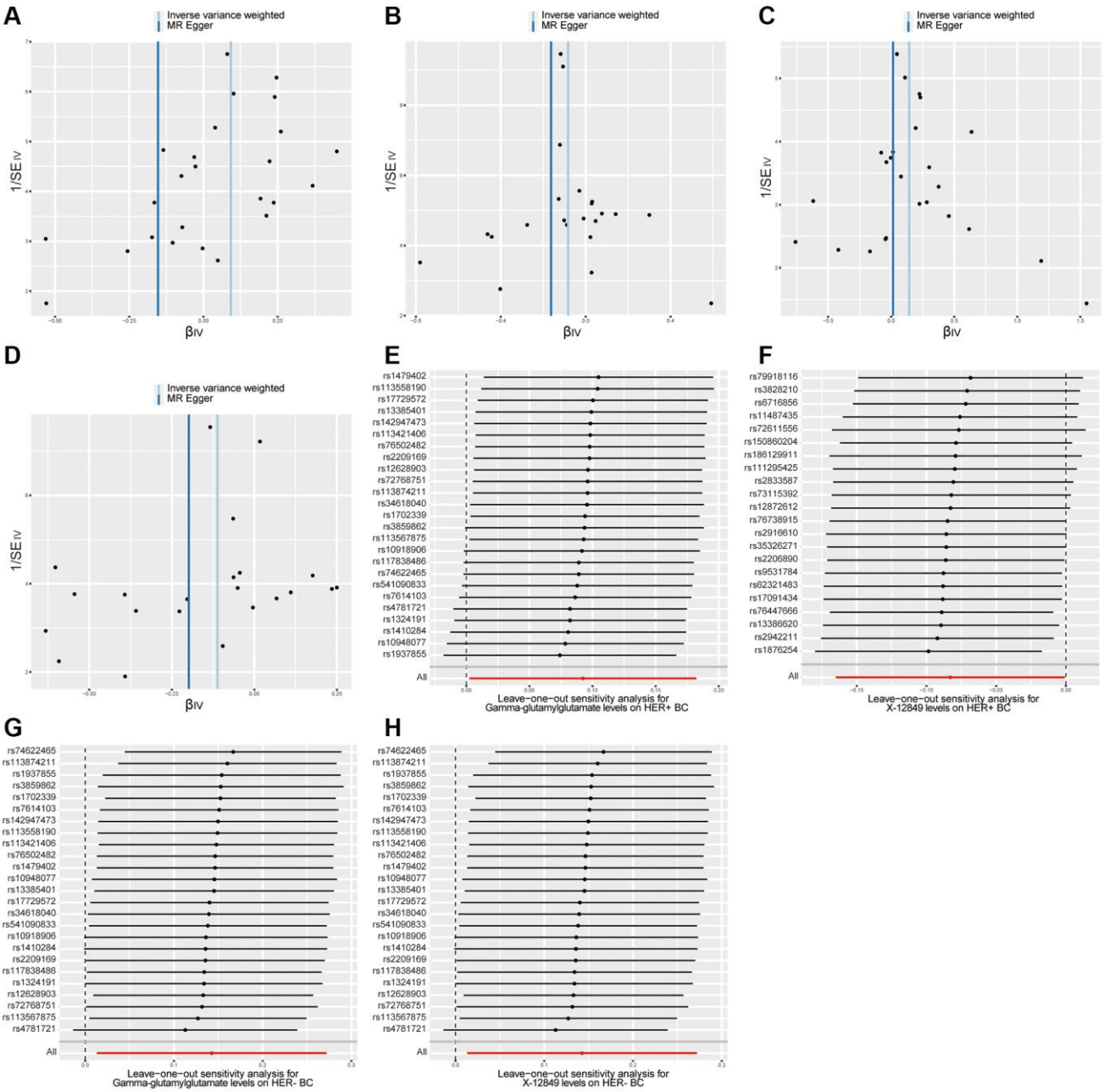
Supplementary Figure 2. Scatter plots showed the genetic associations of metabolic factors and HER+/HER− BC. (**A**) Epiandrosterone sulfate levels on HER+ BC. (**B**) 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels on HER+ BC. (**C**) Glycohyocholate levels on HER+ BC. (**D**) Etiocholanolone glucuronide levels on HER+ BC. (**E**) Vanillic acid glycine levels on HER- BC. (**F**) Thyroxine levels on HER- BC. (**G**) 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels on HER-BC. (**H**) N-acetylphenylalanine levels on HER- BC. (**I**) Glucose-to-mannose ratio on HER- BC. Abbreviations: SNP: single nucleotide polymorphisms; HER: human epidermal growth factor receptor; BC: breast cancer.



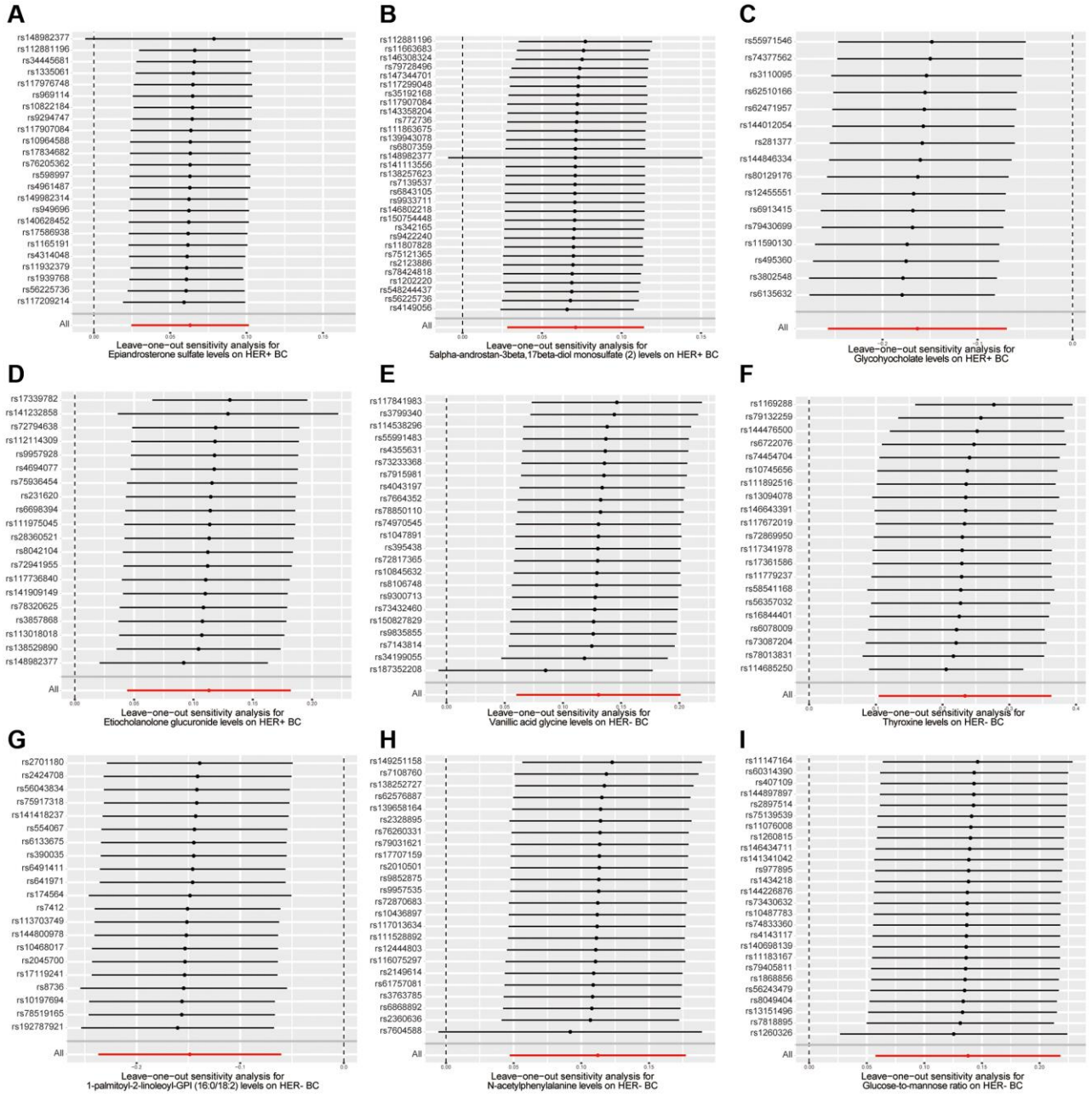
Supplementary Figure 3. Forest plots and scatter plots of metabolic factors on HER+/HER− BC. (**A**) Forest plot of Gamma-glutamylglutamate levels on HER+ BC. (**B**) Forest plot of X-12849 levels on HER+ BC. (**C**) Forest plot of Gamma-glutamylglutamate levels on HER- BC. (**D**) Forest plot of X-12849 levels on HER- BC. (**E**) Scatter plot of Gamma-glutamylglutamate levels on HER+ BC. (**F**) Scatter plot of X-12849 levels on HER+ BC. (**G**) Scatter plot of Gamma-glutamylglutamate levels on HER- BC. (**H**) Scatter plot of X-12849 levels on HER- BC. Abbreviations: SNP: single nucleotide polymorphisms; HER: human epidermal growth factor receptor; BC: breast cancer.



Supplementary Figure 4. Funnel plots showed the heterogeneity between metabolic factors and HER+/HER− BC. (**A**) Epiandrosterone sulfate levels on HER+ BC. (**B**) 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels on HER+ BC. (**C**) Glycohyocholate levels on HER+ BC. (**D**) Etiocholanolone glucuronide levels on HER+ BC. (**E**) Vanillic acid glycine levels on HER- BC. (**F**) Thyroxine levels on HER- BC. (**G**) 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels on HER-BC. (**H**) N-acetylphenylalanine levels on HER- BC. (**I**) Glucose-to-mannose ratio on HER- BC. Abbreviations: Beta: risk index; Se: standard error.



Supplementary Figure 5. Funnel plots and leave-one-out plots of metabolic factors on HER+/HER− BC. (**A**) Funnel plot of Gamma-glutamylglutamate levels on HER+ BC. (**B**) Funnel plot of X-12849 levels on HER+ BC. (**C**) Funnel plot of Gamma-glutamylglutamate levels on HER-BC. (**D**) Funnel plot of X-12849 levels on HER- BC. (**E**) Leave-one-out plot of Gamma-glutamylglutamate levels on HER+ BC. (**F**) Leave-one-out plot of X-12849 levels on HER+ BC. (**G**) Leave-one-out plot of Gamma-glutamylglutamate levels on HER- BC. (**H**) Leave-one-out plot of X-12849 levels on HER- BC. Abbreviations: SNP: single nucleotide polymorphisms; HER: human epidermal growth factor receptor; BC: breast cancer.



Supplementary Figure 6. Leave-one-out plots for the causal effects of metabolic factors on HER+/HER− BC. (**A**) Epiandrosterone sulfate levels on HER+ BC. (**B**) 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels on HER+ BC. (**C**) Glycohyocholate levels on HER+ BC. (**D**) Etiocholanolone glucuronide levels on HER+ BC. (**E**) Vanillic acid glycine levels on HER- BC. (**F**) Thyroxine levels on HER- BC. (**G**) 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels on HER-BC. (**H**) N-acetylphenylalanine levels on HER- BC. (**I**) Glucose-to-mannose ratio on HER- BC. Abbreviations: HER: human epidermal growth factor receptor; BC: breast cancer.

Supplementary Tables

Supplementary Table 1. Genetic data for the MR on the effect of epiandrosterone sulfate levels on malignant neoplasm of breast, HER-positive.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs10822184	−0.188421569	0.23294118	0.418583909	0.828265463	0.524669739	1.307534292	20.73820934
rs10964588	0.032182983	0.266742083	0.903966629	1.032706456	0.612239642	1.741936571	21.43573389
rs112881196	−0.435423443	0.239048617	0.068533182	0.646990646	0.404963341	1.033666145	28.35063804
rs1165191	0.257408695	0.215939321	0.23324486	1.293573696	0.847188233	1.975160704	24.91289764
rs117209214	0.127939656	0.076937053	0.096329772	1.136484421	0.977402274	1.321458803	55.18570351
rs117907084	0.003556025	0.19970231	0.985793116	1.003562355	0.678506987	1.484343449	20.06779613
rs117976748	−0.145734469	0.198434995	0.462693292	0.864387194	0.585864315	1.275321267	20.2000859
rs11932379	0.57928347	0.280559558	0.038947294	1.784759139	1.029822788	3.09311973	21.63336488
rs1335061	−0.261670892	0.227828544	0.25074445	0.769764318	0.492522646	1.203065706	20.53411176
rs140628452	0.183972838	0.217284855	0.397168127	1.201983175	0.785130387	1.840157479	20.19661709
rs148982377	0.058740377	0.020914412	0.004975624	1.060499875	1.017906564	1.104875462	1232.653533
rs149982314	0.588319122	0.501964201	0.241183397	1.800958677	0.673322603	4.817084918	20.48065325
rs17586938	0.289261488	0.244790928	0.237337653	1.335440884	0.82652201	2.157719132	21.92739899
rs17834682	0.037573221	0.316268616	0.905432532	1.038288019	0.55860467	1.929883635	21.56000559
rs1939768	0.45061658	0.236461542	0.05669363	1.569279474	0.987234142	2.494482274	34.83365566
rs34445681	−0.209590616	0.185887977	0.259526773	0.810916154	0.563306693	1.167365873	22.74255116
rs4314048	0.386768204	0.238304444	0.104589323	1.472215198	0.922831603	2.348659909	20.06542188
rs4961487	0.087461739	0.203206663	0.666899002	1.091400505	0.732843343	1.625388392	26.52156846
rs56225736	0.283035001	0.170650237	0.097202899	1.327151613	0.94986115	1.854304076	38.52168963
rs598997	0.07868796	0.224981019	0.726523497	1.081866683	0.696091029	1.681440317	21.2381214
rs76205362	0.064688883	0.17818401	0.716570982	1.066827064	0.752351399	1.512750539	20.13317837
rs9294747	−0.126756	0.220332878	0.565093039	0.880948598	0.572004488	1.356755843	26.37805868
rs949696	0.163540851	0.209913237	0.43592797	1.177673464	0.780446495	1.777078628	20.53960905
rs969114	−0.057751433	0.159153317	0.716704492	0.943884537	0.690947155	1.289415569	41.18745558
All - Inverse variance weighted	0.063053547	0.019637758	0.001323512	1.065083869	1.024867709	1.106878125	NA
All - MR Egger	0.062741091	0.023725518	0.014808268	1.06475113	1.016371646	1.115433486	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Epiandrosterone sulfate levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 2. Genetic data for the MR on the effect of 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels on malignant neoplasm of breast, HER-positive.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs111863675	−0.009326836	0.513639177	0.985512535	0.990716524	0.362018623	2.71123961	23.07545101
rs112881196	−0.30581922	0.167895557	0.068533182	0.736519756	0.529991393	1.023528603	48.04621488
rs11663683	−0.3421822	0.194710939	0.078852106	0.710218791	0.484898523	1.040239776	24.65996095
rs117299048	−0.331387347	0.33877409	0.327977559	0.717927028	0.36958129	1.394603115	22.85136511
rs117907084	0.003197174	0.179549622	0.985793116	1.00320229	0.705590607	1.426343867	20.91978384
rs11807828	0.277783356	0.261994272	0.28902406	1.320200152	0.789997563	2.20624534	24.84529912
rs1202220	0.293139268	0.204683138	0.152097665	1.340629484	0.897591978	2.002343446	19.88100982
rs138257623	0.101011682	0.240881413	0.674966239	1.106289566	0.68996398	1.773826808	23.80669888
rs139943078	0.016669697	0.486205337	0.972649645	1.016809412	0.392078673	2.636974289	20.93272933

rs141113556	0.207232641	0.552168523	0.707432256	1.230268749	0.416854657	3.630908692	20.33773588
rs143358204	−0.075303679	0.268048007	0.778761366	0.927461793	0.548439793	1.568422622	19.66687909
rs146308324	−0.502091443	0.253103617	0.047284976	0.605263461	0.368551606	0.99400966	19.67591162
rs146802218	0.12393409	0.211983744	0.55879008	1.131941262	0.747101632	1.71501569	22.04457795
rs147344701	−0.129526678	0.225266144	0.565295364	0.878511151	0.564932897	1.366147814	21.38021978
rs148982377	0.071205907	0.025352743	0.004975624	1.073802307	1.021747638	1.128508989	424.9030325
rs150754448	0.132156646	0.214030393	0.536926923	1.141287083	0.750254404	1.736126038	21.45243931
rs2123886	0.15836279	0.154357665	0.304916548	1.171591159	0.865733431	1.58550634	36.55078129
rs342165	0.146255663	0.192725862	0.447924777	1.157492078	0.793352987	1.688766453	20.43767199
rs35192168	−0.060529064	0.201004385	0.763313036	0.941266411	0.634766758	1.395760642	21.44885522
rs4149056	0.422960835	0.173259937	0.014638855	1.52647451	1.086945381	2.143736449	27.01568196
rs548244437	0.542581896	0.29577842	0.066591649	1.720443138	0.963537563	3.071934823	24.45539684
rs56225736	0.291948112	0.176024217	0.097202899	1.339033536	0.948323725	1.890715968	32.08251119
rs6807359	0.086811931	0.231770218	0.707987476	1.090691535	0.692492621	1.717863829	20.08732606
rs6843105	0.107429998	0.185915153	0.563369293	1.113412917	0.773396286	1.602914762	23.44067997
rs7139537	0.099068951	0.185554757	0.593406002	1.104142428	0.767498792	1.588446152	25.17973585
rs75121365	0.120666996	0.126305891	0.339398079	1.128249138	0.880828311	1.445169394	20.73057531
rs772736	−0.152966163	0.383980642	0.690357862	0.858158758	0.404312081	1.82145548	21.26363088
rs78424818	0.28440662	0.211148087	0.177995262	1.328973207	0.878584113	2.010245531	20.18660501
rs79728496	−0.202806567	0.220141111	0.356916198	0.816436151	0.530315512	1.256927196	19.94569702
rs9422240	0.20860023	0.218748032	0.340281279	1.231952403	0.802401708	1.891454998	19.96396173
rs9933711	0.119312043	0.213005107	0.57538599	1.126721449	0.742169247	1.710527928	22.56486913
All - Inverse variance weighted	0.07112927	0.021948953	0.0011925	1.073720017	1.028508118	1.120919373	NA
All - MR Egger	0.060598849	0.028310409	0.040851566	1.062472617	1.005123527	1.123093859	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on 5alpha-androstan-3beta,17beta-diol monosulfate (2) levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 3. Genetic data for the MR on the effect of glycohyocholate levels on malignant neoplasm of breast, HER-positive.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs11590130	0.016348	0.201423	0.935312	1.016483	0.684928	1.508534	20.20511
rs12455551	−0.09816	0.199881	0.623366	0.906505	0.612671	1.341259	19.98324
rs144012054	−0.33053	0.254222	0.193543	0.718542	0.43657	1.182634	22.42442
rs144846334	−0.23794	0.243737	0.328958	0.788251	0.488868	1.270976	27.61669
rs281377	−0.23615	0.187632	0.208176	0.78966	0.546669	1.140658	24.39526
rs3110095	−0.2343	0.139719	0.093561	0.791128	0.60161	1.040346	43.51957
rs3802548	−0.01367	0.158	0.931068	0.986426	0.723723	1.344487	31.89631
rs495360	−0.02041	0.173046	0.906093	0.979793	0.697966	1.375417	27.15487
rs55971546	−0.31575	0.159354	0.04754	0.72924	0.533612	0.996586	37.17743
rs6135632	0.05405	0.182769	0.767437	1.055537	0.737731	1.510252	20.60889
rs62471957	−0.29196	0.208418	0.161258	0.746796	0.496356	1.123598	26.20432
rs62510166	−0.31292	0.212843	0.141511	0.731309	0.481864	1.109882	20.35706
rs6913415	−0.08699	0.197944	0.660333	0.916689	0.621912	1.351186	20.66338
rs74377562	−0.32945	0.174815	0.059487	0.719318	0.51064	1.013273	25.87804
rs79430699	−0.00756	0.267961	0.977496	0.99247	0.586981	1.678072	20.31629
rs80129176	−0.17817	0.285062	0.531954	0.836799	0.478599	1.46309	22.44105

All - Inverse variance weighted	−0.16328	0.048004	0.000671	0.849354	0.773084	0.933148	NA
All - MR Egger	−0.30179	0.105753	0.012756	0.739496	0.601059	0.909817	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Glycohyocholate levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X

Supplementary Table 4. Genetic data for the MR on the effect of etiocholanolone glucuronide levels on malignant neoplasm of breast, HER-positive.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs111975045	0.095969	0.181569	0.597115	1.100725	0.771123	1.571209	20.9241
rs112114309	−0.05323	0.191212	0.780736	0.948166	0.65181	1.379264	20.02932
rs113018018	0.343348	0.204587	0.093299	1.409659	0.943986	2.105051	22.52342
rs117736840	0.282134	0.258356	0.274818	1.325956	0.79912	2.200119	22.65326
rs138529890	0.339581	0.169448	0.045065	1.404359	1.007491	1.957561	50.75351
rs141232858	0.091693	0.050264	0.068119	1.096028	0.993198	1.209504	750.6996
rs141909149	0.48145	0.356618	0.177002	1.618419	0.80451	3.255746	20.04527
rs148982377	0.255139	0.090842	0.004976	1.290641	1.080138	1.542168	49.29313
rs17339782	-0.48544	0.194454	0.012546	0.615428	0.420392	0.900949	19.7898
rs231620	0.047124	0.222408	0.832199	1.048252	0.677873	1.621001	19.54917
rs28360521	0.10378	0.269853	0.700549	1.109356	0.653684	1.882672	21.11227
rs3857868	0.271677	0.178246	0.127466	1.312163	0.925256	1.86086	32.03785
rs4694077	-0.06806	0.21386	0.750293	0.934203	0.614327	1.420635	20.24474
rs6698394	0.070593	0.211988	0.739129	1.073145	0.708289	1.625945	22.94598
rs72794638	−0.06825	0.19026	0.719797	0.934025	0.643289	1.35616	59.20234
rs72941955	0.249276	0.356158	0.483989	1.283096	0.638397	2.578857	20.43926
rs75936454	0.037777	0.183449	0.836848	1.0385	0.724856	1.487856	20.21996
rs78320625	0.28462	0.200691	0.156133	1.329257	0.896968	1.969884	19.72805
rs8042104	0.154115	0.214331	0.47211	1.166625	0.766459	1.775718	21.61875
rs9957928	−0.06675	0.206792	0.746847	0.935427	0.623713	1.402926	24.8461
All - Inverse variance weighted	0.113156	0.035173	0.001295	1.119807	1.04521	1.199728	NA
All - MR Egger	0.112137	0.053207	0.049354	1.118666	1.007882	1.241627	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Etiocholanolone glucuronide levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 5. Genetic data for the MR on the effect of vanillic acid glycine levels on malignant neoplasm of breast, HER-negative.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs1047891	0.135427	0.241475	0.574912	1.145026	0.713292	1.838074	23.2309
rs10845632	0.18359	0.228901	0.422525	1.201523	0.767163	1.881815	24.04263
rs114538296	−0.0399	0.176104	0.82074	0.960882	0.680404	1.356978	43.54908
rs117841983	−0.08292	0.13728	0.545846	0.920428	0.703291	1.204605	20.53855
rs150827829	0.263888	0.205362	0.198797	1.301982	0.870557	1.94721	20.18708
rs187352208	0.195583	0.056102	0.00049	1.21602	1.089396	1.357363	21.58818
rs34199055	0.593037	0.225676	0.008593	1.809476	1.162662	2.816127	29.9908
rs3799340	−0.18685	0.17779	0.29328	0.829569	0.585484	1.175413	47.17725
rs395438	0.167138	0.294723	0.570646	1.181917	0.663305	2.106011	21.08149

rs4043197	−0.12212	0.324078	0.706305	0.885042	0.468924	1.670417	20.64513
rs4355631	−0.15312	0.254099	0.546778	0.858028	0.521444	1.41187	20.46171
rs55991483	−0.18764	0.257874	0.466845	0.828917	0.500039	1.3741	20.2511
rs7143814	0.416955	0.25529	0.102414	1.517334	0.919971	2.502581	21.81484
rs72817365	0.153579	0.205552	0.454969	1.166	0.779345	1.744487	20.86364
rs73233368	−0.1427	0.266131	0.591816	0.867013	0.514624	1.460701	20.93719
rs73432460	0.308528	0.269	0.251405	1.361419	0.803553	2.306585	20.47054
rs74970545	0.133812	0.359918	0.710053	1.143178	0.564606	2.314634	25.15526
rs7664352	0.023403	0.277305	0.932743	1.023679	0.594452	1.762831	20.71519
rs78850110	0.068632	0.217174	0.751983	1.071042	0.699753	1.639338	22.50003
rs7915981	−0.11412	0.262514	0.663762	0.89215	0.533312	1.492431	20.14542
rs8106748	0.156164	0.147007	0.288102	1.169018	0.876368	1.559395	59.55688
rs9300713	0.263685	0.249369	0.290325	1.301718	0.798454	2.122187	23.38335
rs9835855	0.293402	0.218799	0.179932	1.340981	0.873327	2.059057	21.18513
All - Inverse variance weighted	0.130564	0.035981	0.000285	1.139471	1.061881	1.22273	NA
All - MR Egger	0.161645	0.052462	0.005663	1.175443	1.060585	1.30274	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Vanillic acid glycine levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 6. Genetic data for the MR on the effect of thyroxine levels on malignant neoplasm of breast, HER-negative.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs10745656	0.132547	0.317067	0.675917	1.141733	0.613298	2.125481	19.99626
rs111892516	0.129683	0.444183	0.770319	1.138467	0.476677	2.719045	20.35275
rs114685250	1.125	0.331035	0.000678	3.080218	1.6099	5.893373	21.67381
rs1169288	−0.35188	0.221182	0.111631	0.703365	0.455939	1.085062	29.97077
rs117341978	0.37451	0.33825	0.268207	1.454279	0.749417	2.822095	19.96895
rs117672019	0.394313	0.9104	0.664927	1.483365	0.249058	8.834787	20.47427
rs11779237	0.352126	0.286673	0.219326	1.422088	0.810785	2.494291	23.69207
rs13094078	0.223959	0.183097	0.221264	1.25102	0.873794	1.791098	20.30568
rs144476500	−0.13466	0.267005	0.614019	0.874011	0.51789	1.475014	26.22235
rs146643391	0.216047	0.25445	0.395839	1.241161	0.753765	2.043713	20.16352
rs16844401	0.434456	0.287088	0.130198	1.544122	0.879646	2.710538	21.95951
rs17361586	0.4097	0.359475	0.254404	1.506366	0.744628	3.047344	22.47474
rs56357032	0.425139	0.317509	0.180576	1.529803	0.821045	2.850391	21.10301
rs58541168	0.28832	0.180453	0.110096	1.334184	0.936723	1.900292	20.47729
rs6078009	0.560027	0.300191	0.062101	1.750721	0.972052	3.153147	23.04986
rs6722076	0.102134	0.190838	0.592519	1.107532	0.761925	1.609907	38.40639
rs72869950	0.542499	0.514439	0.291634	1.7203	0.627632	4.715236	20.67701
rs73087204	0.42326	0.226859	0.062078	1.526931	0.978842	2.381913	21.17641
rs74454704	0.082849	0.279324	0.766767	1.086378	0.62837	1.878222	21.44002
rs78013831	0.428069	0.202577	0.034591	1.534291	1.031504	2.282152	23.51178
rs79132259	−0.36707	0.296642	0.215937	0.692764	0.387328	1.239058	19.56237
All - Inverse variance weighted	0.233614	0.066072	0.000407	1.263156	1.109726	1.4378	NA
All - MR Egger	0.408993	0.16511	0.022814	1.505302	1.089129	2.080501	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Thyroxine levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 7. Genetic data for the MR on the effect of 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels on malignant neoplasm of breast, HER-negative.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs10197694	0.109425	0.267058	0.681994	1.115637	0.660995	1.882988	21.69864
rs10468017	−0.00197	0.257232	0.993877	0.998028	0.602813	1.652353	21.12909
rs113703749	0.024403	0.355626	0.945293	1.024703	0.510367	2.057374	20.6024
rs141418237	−0.32441	0.255004	0.203308	0.722952	0.438577	1.191718	20.0654
rs144800978	0.397625	0.5747	0.489011	1.488286	0.482494	4.59072	20.02776
rs17119241	0.021194	0.266099	0.936518	1.02142	0.606312	1.72073	22.05837
rs174564	−0.14958	0.105179	0.154987	0.861071	0.700663	1.058203	128.2794
rs192787921	−0.04159	0.143314	0.771667	0.959264	0.724348	1.270367	22.96185
rs2045700	0.009159	0.265864	0.972519	1.009201	0.599335	1.699362	21.36032
rs2424708	−0.29237	0.202258	0.14831	0.746493	0.502181	1.109663	36.53821
rs2701180	−0.4721	0.265588	0.075475	0.62369	0.370592	1.049645	21.15642
rs390035	−0.27102	0.26163	0.300255	0.762602	0.456662	1.273509	21.8915
rs554067	−0.28897	0.257274	0.261355	0.749036	0.452383	1.240219	22.33709
rs56043834	−0.37799	0.264855	0.153532	0.685236	0.407748	1.151567	19.55945
rs6133675	−0.30561	0.290871	0.29341	0.736674	0.416563	1.302776	21.32047
rs641971	−0.24098	0.266641	0.366123	0.785858	0.465988	1.325297	19.77715
rs6491411	−0.25384	0.280509	0.365514	0.77582	0.447699	1.34442	21.33727
rs7412	−0.07137	0.270175	0.791666	0.931121	0.548313	1.581188	31.0331
rs75917318	−0.39402	0.276574	0.154265	0.674344	0.392154	1.159595	21.64591
rs78519165	0.07761	0.246901	0.753266	1.080701	0.6661	1.753362	20.92825
rs8736	−0.12695	0.097302	0.191987	0.880775	0.727846	1.065837	161.469
All - Inverse variance weighted	−0.14864	0.045047	0.000968	0.861875	0.78904	0.941433	NA
All - MR Egger	−0.05745	0.100287	0.573469	0.94417	0.775682	1.149256	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on 1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 8. Genetic data for the MR on the effect of N-acetylphenylalanine levels on malignant neoplasm of breast, HER-negative.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs10436897	0.162257	0.371858	0.662588	1.176163	0.567461	2.437806	19.53623
rs111528892	0.273874	0.417734	0.512069	1.31505	0.579909	2.982116	20.47289
rs116075297	0.154515	0.169156	0.361007	1.167092	0.837754	1.625901	19.79232
rs117013634	0.172937	0.365399	0.636012	1.188791	0.58086	2.432984	19.79306
rs12444803	0.212904	0.272664	0.434903	1.237266	0.725048	2.111346	21.77973
rs138252727	−0.09767	0.221159	0.658771	0.906951	0.587935	1.399068	19.50911
rs139658164	−0.22814	0.427795	0.593832	0.796013	0.34417	1.841056	23.5126
rs149251158	−0.17558	0.17629	0.319264	0.83897	0.593862	1.185244	36.48982
rs17707159	0.025432	0.291646	0.93051	1.025758	0.57915	1.816768	20.4308
rs2010501	0.039769	0.272372	0.883914	1.04057	0.610132	1.774676	19.66764
rs2149614	0.329545	0.280591	0.240207	1.390336	0.802188	2.409702	20.63435
rs2328895	0.082115	0.133183	0.537526	1.085581	0.83617	1.409386	98.4315
rs2360636	0.508759	0.284186	0.073416	1.663226	0.9529	2.903052	21.86603
rs3763785	0.358775	0.266672	0.178502	1.431575	0.848826	2.414402	20.44586

rs61757081	0.426654	0.320268	0.182803	1.532122	0.817853	2.870193	22.4386
rs62576887	−0.17678	0.329819	0.59197	0.837966	0.439015	1.599463	20.03196
rs6868892	0.44287	0.288824	0.125188	1.557171	0.884064	2.742765	20.16622
rs7108760	0.041507	0.115648	0.719663	1.042381	0.830968	1.30758	179.082
rs72870683	0.114036	0.266764	0.669031	1.120792	0.664433	1.890599	25.41432
rs7604588	0.128698	0.044702	0.00399	1.137346	1.041937	1.241493	926.7864
rs76260331	−0.04938	0.329861	0.881006	0.951821	0.498623	1.816931	19.96712
rs79031621	0.033579	0.252513	0.894209	1.03415	0.630434	1.696395	20.27854
rs9852875	−0.02209	0.443771	0.960295	0.97815	0.409883	2.334268	21.30936
rs9957535	0.042629	0.344942	0.901645	1.043551	0.530754	2.051795	21.09939
All - Inverse variance weighted	0.112196	0.033219	0.000731	1.118733	1.048214	1.193995	NA
All - MR Egger	0.070678	0.052114	0.18879	1.073235	0.969024	1.188655	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on N-acetylphenylalanine levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 9. Genetic data for the MR on the effect of glucose-to-mannose ratio on malignant neoplasm of breast, HER-negative.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs10487783	0.184665	0.313782	0.556187	1.202816	0.650283	2.224824	19.95859
rs11076008	0.004923	0.305119	0.987127	1.004935	0.552605	1.827515	19.71775
rs11147164	−0.02491	0.185323	0.89307	0.975396	0.678314	1.402593	21.8003
rs11183167	0.192068	0.239798	0.423157	1.211753	0.757345	1.938806	21.86084
rs1260326	0.162462	0.070699	0.021566	1.176403	1.024179	1.351254	335.8678
rs1260815	0.027013	0.300905	0.928469	1.027381	0.569634	1.852964	19.59771
rs13151496	0.260273	0.211747	0.219007	1.297285	0.856629	1.964616	21.39685
rs140698139	0.204595	0.270158	0.44886	1.227028	0.722589	2.083616	19.95859
rs141341042	0.121498	0.20112	0.545774	1.129187	0.761323	1.674799	20.49849
rs1434218	0.117687	0.277163	0.671119	1.124892	0.653409	1.936585	20.55981
rs144226876	0.170973	0.309431	0.580578	1.186459	0.646933	2.175934	20.58319
rs144897897	−0.04263	0.249589	0.864376	0.958265	0.587532	1.56293	20.20088
rs146434711	0.094234	0.22554	0.676081	1.098817	0.706223	1.709656	19.63824
rs1868856	0.241477	0.265734	0.363499	1.273128	0.756267	2.143232	20.31357
rs2897514	−0.08285	0.281854	0.768795	0.920488	0.529785	1.599325	20.02076
rs407109	−0.13179	0.298444	0.658786	0.876525	0.488342	1.573275	22.50079
rs4143117	0.190045	0.255668	0.457284	1.209304	0.732666	1.996019	19.71288
rs56243479	0.224412	0.225723	0.32013	1.251586	0.804121	1.94805	21.52789
rs60314390	−0.03039	0.234005	0.896663	0.970065	0.613213	1.534583	20.90927
rs73430632	0.175201	0.303778	0.564114	1.191486	0.656914	2.161073	20.27928
rs74833360	0.185068	0.274262	0.499813	1.2033	0.702938	2.059827	23.65132
rs75139539	0.046835	0.226002	0.83583	1.047949	0.672919	1.631989	21.07131
rs7818895	0.434803	0.275897	0.115035	1.544659	0.899465	2.652654	20.06782
rs79405811	0.189572	0.207939	0.361941	1.208733	0.804135	1.816902	21.77785
rs8049404	0.308263	0.265143	0.244981	1.361058	0.809435	2.288608	22.3594
rs977895	0.108016	0.296469	0.715602	1.114066	0.623091	1.991914	20.22298

All - Inverse variance weighted	0.137918	0.040991	0.000767	1.147881	1.059266	1.24391	NA
All - MR Egger	0.132366	0.080175	0.111774	1.141526	0.975527	1.335773	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Glucose-to-mannose ratio; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 10. Genetic data for the MR on the effect of gamma-glutamyl glutamate levels on malignant neoplasm of breast, HER-positive.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs10918906	0.224046	0.210365	0.286859	1.251129	0.828392	1.889591	23.65602
rs10948077	0.109634	0.199475	0.582584	1.11587	0.754774	1.64972	41.30163
rs113421406	−0.16811	0.442487	0.703996	0.845257	0.355088	2.012061	19.82452
rs113558190	0.01412	0.260835	0.95683	1.01422	0.608282	1.691061	20.01361
rs113567875	1.18976	0.474075	0.012085	3.286294	1.297674	8.322374	20.41966
rs113874211	−0.75929	0.414896	0.06724	0.467999	0.207529	1.055388	21.30396
rs117838486	0.457242	0.354576	0.197208	1.579712	0.788418	3.165184	19.57548
rs12628903	1.549331	0.696083	0.026029	4.708318	1.203229	18.42398	21.02661
rs1324191	0.376135	0.304447	0.216656	1.456644	0.802053	2.645476	19.93714
rs13385401	0.077626	0.290218	0.789103	1.080719	0.611891	1.908759	23.87765
rs1410284	0.230933	0.212766	0.277752	1.259775	0.8302	1.911628	27.36363
rs142947473	−0.00357	0.267129	0.989333	0.996435	0.590289	1.682029	21.23784
rs1479402	−0.03973	0.404449	0.92174	0.961045	0.434981	2.12333	19.67343
rs1702339	−0.41855	0.437569	0.338807	0.658003	0.279102	1.551294	20.34053
rs17729572	0.22393	0.331842	0.499797	1.250984	0.652803	2.397294	20.49173
rs1937855	−0.0792	0.261496	0.761977	0.923852	0.553367	1.542382	22.24685
rs2209169	0.301682	0.278392	0.278516	1.352132	0.783514	2.333411	20.23599
rs34618040	0.194237	0.237317	0.413087	1.214384	0.76269	1.93359	20.76366
rs3859862	0.045667	0.185696	0.805742	1.046726	0.727387	1.506261	46.12258
rs4781721	0.637161	0.240791	0.008142	1.891104	1.179641	3.031662	25.23359
rs541090833	0.282286	0.32908	0.391	1.326158	0.695788	2.52763	19.67458
rs72768751	0.617449	0.382692	0.10665	1.854193	0.875791	3.925628	21.30536
rs74622465	−0.61806	0.327021	0.05876	0.538987	0.283931	1.023161	21.42821
rs7614103	−0.0383	0.272212	0.888105	0.962423	0.564488	1.640882	21.59485
rs76502482	−0.04564	0.407468	0.910824	0.95539	0.42987	2.123363	24.79271
All - Inverse variance weighted	0.142585	0.066054	0.030881	1.153251	1.013205	1.312655	NA
All - MR Egger	0.014249	0.181106	0.937971	1.014351	0.711259	1.4466	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Gamma-glutamyl glutamate levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 11. Genetic data for the MR on the effect of X-12849 levels on malignant neoplasm of breast, HER-positive.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs111295425	−0.06402	0.182714	0.72607	0.93799	0.655645	1.341925	20.29968
rs11487435	0.067219	0.272789	0.805362	1.06953	0.6266	1.825556	22.24892
rs12872612	−0.2033	0.274097	0.458272	0.816036	0.476862	1.396451	20.78645
rs13386620	0.249797	0.255629	0.328479	1.283764	0.777838	2.118758	22.27942

rs150860204	−0.59147	0.445726	0.184516	0.553514	0.231057	1.325982	21.76679
rs17091434	0.176696	0.238975	0.459668	1.193268	0.746997	1.906152	21.36393
rs186129911	0.017042	0.138513	0.902078	1.017188	0.775347	1.334463	22.59548
rs1876254	0.234897	0.257684	0.361995	1.264778	0.763255	2.095845	20.97361
rs2206890	−0.3582	0.295554	0.225522	0.698931	0.39161	1.247425	21.18312
rs2833587	−0.04387	0.23544	0.852179	0.957077	0.603304	1.5183	24.97789
rs2916610	−0.60233	0.228838	0.008485	0.547533	0.349639	0.857435	20.17495
rs2942211	−0.05053	0.256245	0.843681	0.950727	0.575355	1.570999	19.55932
rs35326271	0.11037	0.262913	0.674635	1.116691	0.667018	1.869513	21.37654
rs3828210	−0.00436	0.289035	0.987953	0.995645	0.565032	1.754432	20.36693
rs62321483	−0.06282	0.24128	0.79459	0.939114	0.585244	1.506953	19.66226
rs6716856	−0.22686	0.296503	0.444199	0.797031	0.445745	1.425161	21.91351
rs72611556	−0.13348	0.132526	0.313833	0.875044	0.674872	1.134587	21.28985
rs73115392	−0.54438	0.265697	0.040473	0.5802	0.344677	0.97666	21.38965
rs76447666	−0.3912	0.526486	0.457459	0.676246	0.240963	1.897834	20.85577
rs76738915	−0.09551	0.386448	0.804789	0.908907	0.426156	1.938521	20.90233
rs79918116	−0.6315	0.341097	0.064113	0.531792	0.272518	1.037741	20.14291
rs9531784	−0.39232	0.26643	0.140889	0.675491	0.40071	1.1387	20.4065
All - Inverse variance weighted	−0.11167	0.05099	0.028522	0.89434	0.809281	0.988339	NA
All - MR Egger	−0.1982	0.103165	0.06907	0.820203	0.670047	1.004007	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on X-12849 levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 12. Genetic data for the MR on the effect of gamma-glutamyl glutamate levels on malignant neoplasm of breast, HER-negative.

SNP	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs10918906	0.101604	0.167793	0.544828	1.106944	0.796705	1.537992	23.65602
rs10948077	0.246031	0.159203	0.122253	1.278939	0.936124	1.747296	41.30163
rs113421406	−0.25568	0.356973	0.473842	0.77439	0.384679	1.558913	19.82452
rs113558190	−0.13508	0.207103	0.514245	0.873645	0.582164	1.311066	20.01361
rs113567875	0.048458	0.382438	0.899172	1.049651	0.496029	2.221174	20.41966
rs113874211	−0.1038	0.336623	0.757813	0.901406	0.465995	1.743652	21.30396
rs117838486	0.21209	0.284623	0.456174	1.236259	0.707675	2.159658	19.57548
rs12628903	−0.52952	0.569359	0.352353	0.588886	0.192922	1.797542	21.02661
rs1324191	0.367854	0.243098	0.130231	1.444632	0.897073	2.32641	19.93714
rs13385401	−0.07421	0.232106	0.749172	0.928475	0.589112	1.463331	23.87765
rs1410284	0.240185	0.16971	0.15699	1.271484	0.911698	1.773254	27.36363
rs142947473	−0.03101	0.213287	0.884417	0.96947	0.638235	1.47261	21.23784
rs1479402	−0.53154	0.327986	0.105101	0.5877	0.309007	1.117746	19.67343
rs1702339	−0.00303	0.349937	0.993084	0.996971	0.502123	1.979499	20.34053
rs17729572	−0.16542	0.264892	0.532323	0.847541	0.504289	1.424432	20.49173
rs1937855	0.449422	0.208256	0.030925	1.567406	1.042103	2.357504	22.24685
rs2209169	−0.02698	0.222324	0.903418	0.973383	0.62956	1.504977	20.23599
rs34618040	0.039344	0.189479	0.83551	1.040128	0.717462	1.507907	20.76366
rs3859862	0.080187	0.148064	0.588116	1.08349	0.810569	1.448304	46.12258
rs4781721	0.260756	0.19233	0.175171	1.29791	0.890287	1.892166	25.23359
rs541090833	0.236721	0.264961	0.371635	1.267087	0.753819	2.129835	19.67458

rs72768751	−0.07131	0.304854	0.815062	0.931177	0.512313	1.692501	21.30536
rs74622465	0.192225	0.259384	0.458642	1.211944	0.728938	2.014997	21.42821
rs7614103	0.222893	0.217264	0.304935	1.249687	0.816324	1.913111	21.59485
rs76502482	−0.17329	0.324639	0.593476	0.84089	0.445042	1.588832	24.79271
All - Inverse variance weighted	0.092333	0.045861	0.044082	1.09673	1.002448	1.199879	NA
All - MR Egger	−0.15293	0.125031	0.233654	0.858187	0.671666	1.096505	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on Gamma-glutamyl glutamate levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 13. Genetic data for the MR on the effect of X-12849 levels on malignant neoplasm of breast, HER-negative.

	beta_X	se_X	p_X	or	or_lci95	or_uci95	F
rs111295425	−0.12127	0.145564	0.404772	0.885792	0.665925	1.178251	20.29968
rs11487435	−0.27573	0.218024	0.205987	0.759018	0.49507	1.163691	22.24892
rs12872612	−0.08852	0.217982	0.684689	0.915288	0.597045	1.403163	20.78645
rs13386620	0.076545	0.203647	0.707012	1.079551	0.724261	1.609131	22.27942
rs150860204	−0.40261	0.362336	0.266509	0.668576	0.328643	1.36012	21.76679
rs17091434	0.03062	0.190521	0.872317	1.031094	0.709779	1.497866	21.36393
rs186129911	−0.10676	0.109875	0.331241	0.898745	0.724618	1.114714	22.59548
rs1876254	0.300179	0.205333	0.143765	1.350101	0.902783	2.01906	20.97361
rs2206890	0.022434	0.23593	0.924247	1.022687	0.644044	1.623941	21.18312
rs2833587	−0.12703	0.187742	0.498658	0.88071	0.60957	1.272455	24.97789
rs2916610	−0.02975	0.179851	0.868629	0.970691	0.682321	1.380934	20.17495
rs2942211	0.141589	0.204329	0.488343	1.152103	0.771904	1.719569	19.55932
rs35326271	−0.00962	0.209668	0.963397	0.990424	0.656672	1.493806	21.37654
rs3828210	−0.46252	0.231383	0.045617	0.629697	0.400105	0.991034	20.36693
rs62321483	0.028356	0.192333	0.882793	1.028761	0.705663	1.499796	19.66226
rs6716856	−0.44285	0.235426	0.059964	0.642204	0.404832	1.018759	21.91351
rs72611556	−0.11802	0.105711	0.264224	0.888676	0.722372	1.093268	21.28985
rs73115392	−0.10132	0.21201	0.632719	0.903644	0.596391	1.369189	21.38965
rs76447666	0.591765	0.425366	0.164168	1.807175	0.785093	4.159867	20.85577
rs76738915	0.027982	0.30992	0.928059	1.028377	0.5602	1.887823	20.90233
rs79918116	−0.78018	0.284595	0.006118	0.458323	0.262373	0.800616	20.14291
rs9531784	0.047001	0.212845	0.82523	1.048123	0.690614	1.590703	20.4065
All - Inverse variance weighted	−0.0829	0.041897	0.047855	0.920443	0.847877	0.999219	NA
All - MR Egger	−0.163	0.084382	0.067685	0.849587	0.720079	1.002387	NA

Abbreviations: F: F-statistic; Beta_X: genetic effect on X-12849 levels; SE_X: standard error of Beta_X; p_X: *p*-value for Beta_X.

Supplementary Table 14. Sensitivity analyses of metabolites on malignant neoplasm of breast, HER-positive.

Exposure	Outcome	Number of IVs	Method	Heterogeneity test		MR-Egger pleiotropy test	MR-PRESSO global outlier test			
				Q	*p*-value	Intercept	*p*-value	RSSobs	*p*-value	Outlier
Epiandrosterone sulfate levels	Malignant neoplasm of breast, HER-negative	24	MR Egger	25.87971	0.256882	0.000152	0.9805	27.17973	0.487	None
			IVW	25.88043	0.306543					
5alpha-androstan-	Malignant neoplasm of	31	MR Egger	31.6985	0.333262	0.00356	0.553914	32.80531	0.52	None

3beta,17beta-diol monosulfate (2) levels	breast, HER-negative		IVW	32.09051	0.36333					
Glycohyocholate levels	Malignant neoplasm of breast, HER-negative	16	MR Egger	5.852484	0.970074	0.018193	0.163704	9.274069	0.919	None
			IVW	8.013065	0.923258					
Etiocholanolone glucuronide levels	Malignant neoplasm of breast, HER-negative	20	MR Egger	21.79846	0.241056	0.000248	0.979458	24.06592	0.347	None
			IVW	21.79928	0.294359					

Supplementary Table 15. Sensitivity analyses of metabolites on malignant neoplasm of breast, HER-negative.

Exposure	Outcome	Number of IVs	Method	Heterogeneity test		MR-Egger pleiotropy test		MR-PRESSO global outlier test		
				Q	*p*-value	Intercept	*p*-value	RSSobs	*p*-value	Outlier
Vanillic acid glycine levels	Malignant neoplasm of breast, HER-negative	23	MR Egger	20.02221	0.519855	−0.00709	0.42472	24.52811	0.521	None
			IVW	20.68498	0.540273					
Thyroxine levels	Malignant neoplasm of breast, HER-negative	21	MR Egger	24.20825	0.188285	−0.02834	0.261528	28.62793	0.215	None
			IVW	25.91447	0.168655					
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2) levels	Malignant neoplasm of breast, HER-negative	21	MR Egger	8.720561	0.977838	−0.0128	0.321558	10.44594	0.982	None
			IVW	9.756483	0.972376					
N-acetylphenylalanine levels	Malignant neoplasm of breast, HER-negative	24	MR Egger	11.15759	0.972428	0.010031	0.31239	13.29427	0.973	None
			IVW	12.22665	0.966824					
Glucose-to-mannose ratio	Malignant neoplasm of breast, HER-negative	26	MR Egger	6.401489	0.99987	0.000908	0.936455	6.924869	1	None
			IVW	6.40798	0.999935					

Supplementary Table 16. Sensitivity analyses of metabolites on malignant neoplasm of breast, HER-positive/negative.

Exposure	Outcome	Number of IVs	Method	Heterogeneity test		MR-Egger pleiotropy test		MR-PRESSO global outlier test		
				Q	*p*-value	Intercept	*p*-value	RSSobs	*p*-value	Outlier
Gamma-glutamyl glutamate levels	Malignant neoplasm of breast, HER-positive	25	MR Egger	13.74344	0.933983	0.028479	0.046076	19.61343	0.83	None
			IVW	18.18968	0.793711					
X-12849 levels	Malignant neoplasm of breast, HER-positive	22	MR Egger	21.04402	0.394546	0.012455	0.287808	23.64717	0.447	None
			IVW	22.29876	0.382472					
Gamma-glutamyl glutamate levels	Malignant neoplasm of breast, HER-negative	25	MR Egger	30.99398	0.122941	0.014923	0.453749	33.63561	0.166	None
			IVW	31.77663	0.132629					
X-12849 levels	Malignant neoplasm of breast, HER-negative	22	MR Egger	19.83662	0.468191	0.013453	0.34612	0.514	22.64835	None
			IVW	20.76764	0.473211					