scientific reports

Check for updates

OPEN

Multilayer analysis of ethnically diverse blood and urine biomarkers for breast cancer risk and prognosis

Jia Feng[1,5]✉, Xing Qi[1,2,5]✉, Chen Chen[1], Baolin Li[1], Min Wang[1], Xuelong Xie[3], Kailan Yang[4], Xuan Liu[1], Rui min Chen[1], Tongtong Guo[1]✉ & Jinbo Liu[1]✉

Breast cancer (BC) is one of the most common malignancies among women globally, characterized by complex pathogenesis involving various biomarkers present in blood and urine. To enhance understanding of the genetic associations between biomarkers and BC via multidimensional, cross ethnic investigations. Based on GWAS data of 35 blood and urine biomarkers from European populations, we adopted multiple analysis strategies including univariable Mendelian randomization (MR) analysis, reverse MR analysis, sensitivity analysis and multivariate MR to identify potential biomarkers associated with BC risk and survival. Our initial analysis included 122,977 BC and 105,974 controls of European ancestry. Building upon these findings, we conducted cross ethnic validation by applying the same analyses to East Asian populations using data from the IEU GWAS database, which included 5,552 BC and 89,731 controls. This step allowed us to investigate the universality and heterogeneity of our identified biomarkers across different ancestries. Subsequently, utilizing clinical laboratory detection data from multiple regions in China, we performed differential analyses and survival assessments on these potential biomarkers to evaluate their clinical relevance and utility. Notably, we leveraged Luzhou's clinical data to integrate HDL-C with conventional tumor markers (CEA, CA125, CA153) into a machine learning model, comparing its diagnostic efficacy against tumor marker combination. Our study validated associations of ALP, HDL-C, TG, SHBG, and IGF-1 with BC risk, reinforcing the reliability of these findings. Moreover, notable interethnic disparities emerged in the association between HDL-C and BC risk, where in HDL-C demonstrates a contrasting role: acting as a genetic protective agent against BC and suggesting promise as an auxiliary diagnostic marker in East Asian populations, yet inversely, it serves as a genetic dangerous predictor in European populations. Analyzing BC subtypes, we identified associations of HDL-C, TG, SHBG, and CRP with ER^{+}BC, while ER^{-}BC showed associations with GLU, urinary creatinine and microalbuminuria, underscoring subtype-specific genetic characteristics critical for personalized prevention and treatment strategies. Overall, this comprehensive study, by traversing the intricate landscape of genetic associations across ethnic boundaries and employing advanced analytical methodologies, not only uncovers the complex interplay between key biomarkers and BC susceptibility but also highlights the significance of ethnic-specific differences in the role of HDL-C. By enhancing the diagnostic power of a tailored biomarker panel through machine learning, this study contributes to the advancement of precision medicine in BC, offering strategies tailored to the unique genetic profiles and biomarker patterns across diverse populations.

Keywords Breast cancer, Blood and urine biomarkers, Mendelian randomization, Machine learning

Breast Cancer (BC), the most prevalent malignancy among women globally, is characterized by a complex etiology involving genetic, environmental, and lifestyle factors[1]. Recent advances in genetics and molecular biology have shed light on numerous genetic biomarkers closely associated with the onset, progression, and

[1]Department of Laboratory Medicine, Sichuan Province Engineering Technology Research Center of Molecular Diagnosis of Clinical Diseases, Molecular Diagnosis of Clinical Diseases Key Laboratory of Luzhou, The Affiliated Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China. [2]Department of Clinical Laboratory Medicine, Ziyang Central Hospital, Ziyang 641300, Sichuan, China. [3]Clinical Laboratory of Yibin Second People's Hospital, Yibin 644000, Sichuan, China. [4]Clinical Laboratory of Zigong First People's Hospital, Zigong 643000, Sichuan, China. [5]Jia Feng and Xing Qi are contributed equally to this work and share first authorship. ✉email: ; ; guotongtong0822@163.com; liujb7203@swmu.edu.cn

prognosis of BC[2,3]. Blood and urinary biomarkers, particularly notable for their roles as indicators of metabolic, inflammatory, and endocrine states, are increasingly pivotal in assessing BC risk and determining prognosis[4]. Nonetheless, current research faces challenges in marker selection, analyzing ethnic disparities, and conducting subtype-specific investigations, highlighting the need for a more systematic and comprehensive exploration framework.

BC exhibits notable heterogeneity in incidence, prognosis, and genetic susceptibility across diverse ethnicities and geographical regions, highlighting the complex interplay of genetic backgrounds, environmental factors, and lifestyle habits in disease progression[5]. Disparities in genetic profiles between European and Asian populations can lead to variations in the association of specific biomarkers with BC, influencing the accuracy of risk assessment and treatment strategies. Although mammography followed by histopathological biopsy remains standard, its utility is constrained by inherent limitations[6,7]. In contrast, liquid biopsy techniques, particularly those utilizing blood and urine samples, are emerging as promising alternatives due to their non-invasive nature and high reproducibility[8,9].

Mendelian randomization (MR) analysis, a robust method for causal inference, employs genetic variants as proxies for exposures, circumventing confounding factors inherent in observational studies, thus elucidating causal relationships between exposures (such as blood and urinary biomarkers) and BC risk[10]. The validity of MR analysis rests on three core assumptions: genetic variants are linked to the exposure, independent of confounders, and do not directly influence the outcome[10]. By meeting these criteria, MR analysis mirrors conditions akin to randomized controlled trials, offering compelling evidence for causality[11]. Through MR analyses, select blood and urinary biomarkers, including high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), alkaline phosphatase (ALP), and sex hormone-binding globulin (SHBG), among others, have been implicated in BC risk[12–17]. These biomarkers, however, often exhibit dual roles and high heterogeneity across BC subtypes. Inconsistencies in study findings can partly be attributed to small sample sizes, confounding interference, and methodological constraints. Research predominantly focuses on individual markers, lacking multi-dimensional, cross-ethnic integration, and in-depth examination of inter-marker interactions, subtype specificity, and their implications for survival outcomes. Notably, prospective studies targeting specific ethnic groups, such as East Asians, are relatively scarce, hindering comprehensive understanding of BC's genetic determinants.

Through an in-depth analysis of trans-ethnic genomic data, this study innovatively integrates genetic data of 35 blood and urine biomarkers, employing advanced Mendelian randomization (MR) methodologies including reverse MR (R-MR), sensitivity analyses, and multivariate MR (MVMR). The research reinforces scientific foundations linking biomarkers such as ALP, HDL-C, TG, and SHBG to BC. Comparative genetic effect analyses between European and East Asian populations reveal ethnic-specific risk patterns. Notably, substantial ethnic disparities in the interplay between HDL-C and BC risk are highlighted, emphasizing the need for diverse ethnic investigations. Machine learning (ML) algorithms assess the clinical utility of HDL-C in East Asian populations. Moreover, the study identifies differential genetic determinants among BC subtypes, such as SHBG's protective role in ER^+BC and revealing a novel association of glucose (GLU) with ER^-BC. These findings underscore the importance of tailored therapeutic and preventive strategies based on BC subtypes. Despite challenges such as limited sample sizes and multiple comparisons, preliminary findings suggest UREA and calcium (Ca) as potential prognostic markers for survival rates in specific BC subtypes, offering new avenues for survival prediction and monitoring strategies. This study has made significant progress in assessing BC risk and survival, providing a comprehensive insight into the complex association between biomarkers and BC.

Research design and methods

Study population and data

The genetic association datasets employed in this study were sourced from the IEU GWAS database, accessible at (https://gwas.mrcieu.ac.uk/), featuring a GWAS investigation by the BC Association Consortium (BCAC) aimed at elucidating genetic susceptibilities underlying BC risk. This GWAS encompassed 122,977 cases and 105,974 European ancestry controls, leveraging over 106 million SNP data points. Integrating findings from 68 distinct studies within BCAC, the iCOGS collaboration, and an additional 11 BC-focused GWASs, the research further conducted stratified analyses by ER status, separately examining ER^+ (comprising 69,501 cases and 105,974 controls) and ER^- (involving 21,468 cases and 105,974 controls) subsets[18]. Survival analyses specific to BC patients entailed 2,900 cases and 35,054 controls, with additional stratification by ER status into ER^+ (1,333 cases, 21,726 controls) and ER^- (920 cases, 5,961 controls) groups[19]. The GWAS summary-level data for BC patients based on East Asian populations was released in the published study, including 5,552 BC cases and 89,731 controls[20]. All research endeavors were conducted under the aegis of the initial ethical approvals obtained for the GWAS, thereby obviating the need for supplementary informed consent procedures or additional ethical reviews.

Exposure data

The phenotypic and genotypic data of 35 blood and urine biomarkers used in this study were obtained from the UK Biobank ($n = 363{,}228$ individuals), which conducted a large prospective cohort study[21,22]. Exposure data for the East Asian population, specifically including HDL-C, TG, ALP, CRP, GLU, Apolipoprotein A (APOA) and SHBG, were primarily downloaded from the IEU GWAS database[23].

Selection criteria for the instrumental variables

In the univariate MR analyses, conditionally uncorrelated variants that were strongly associated with the exposure ($P < 5 \times 10^{-8}$) and independently so according to linkage disequilibrium (LD) ($r^2 < 0.001$, with a window size of 10,000 kb) were selected as instrumental variables (IVs) for the European population. For the East Asian population, adhering to the research criteria put forth by Weng et al.[24], we adjusted the LD filtering criteria to

$r^2 < 0.1$ and a window size of 500 kb. The LD proxies were delineated utilizing 1000 Genomes data encompassing both European and East Asian samples. Palindromic IVs exhibiting intermediate minor allele frequencies were omitted. Following the exclusion of weak IVs based on an F-statistic threshold of less than 10 and the Steiger test[25], exposures with a minimum of three remaining valid IVs proceeded to subsequent analyses. Moreover, we employed the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) method to identify and correct for instances of horizontal pleiotropy and outlier data points[26].

MR analysis

Two-ample MR analysis was performed to verify the potential causal link of 35 blood and urine markers to BC risk and survival[27]. The inverse-variance weighted (IVW) method was used as our main MR method to detect the causal effects of 35 markers on the risk of BC and BC survival[28]. To account for type I error, FDR correction was applied to analysis results for each cancer type to identify significant associations at $P_{FDR} < 0.05$. To assess the robustness of causal associations identified via the IVW method, we subsequently conducted a series of complementary analyses, including the weighted median[29], the maximum likelihood, the weighted mode, the sample median, the sample mode, the penalised weighted median, the MR-Robust Adjusted Profile Scoring (MR-RAPS), MR-Egger regression[30], the constrained maximum likelihood and model averaging (cML-MA)[31], the MR-PRESSO, the Bayesian weighted mendelian randomization (BWMR) approachs[32]. Cochran's Q statistic was applied to estimate the heterogeneity among genetic instruments used in the main analysis[29]. We adopted random-effects IVW model if heterogeneity existed, otherwise fix-effects IVW model was used.

Sensitivity, reverse MR and MVMR analyses

To assess the robustness of significant association identified by IVW, We further conducted sensitivity analyses, including Egger intercept test and MR-PRESSO global test to evaluate horizontal pleiotropy, and LOO analyses to examine the presence of dominant IVs[33,34]. To examine the possible reverse causality of the identified significant associations, we performed R-MR analyses in which cancer was treated as the exposure and metabolites as the exposure. Because many genetic variants are pleiotropic, we performed MVMR analysis to explore potential independent associations of HDL and TG, APOA, and BC risk as proxies for genetic variants.

Clinical laboratory data validation from east Asian populations

We compiled a comprehensive dataset comprising clinical biochemical test results from patients diagnosed with benign breast diseases and BC across multiple medical centers in Sichuan Province, China, spanning from June 2018 to December 2023. Specifically, we documented 9,748 cases of benign breast disease and 9,271 cases of BC at the Affiliated Hospital of Southwest Medical University in Luzhou, with detailed survival outcomes available for 1,290 BC patients. Additionally, data from the First People's Hospital of Zigong (April 2020 to December 2023) included 449 benign cases and 443 BC cases, along with survival status information for an additional 130 patients. Furthermore, biochemical indicators and survival data from 120 BC cases were obtained from Ziyang People's Hospital within the same timeframe. Lastly, the Second People's Hospital of Yibin provided matched biochemical test results for 210 BC patients at initial diagnosis and recurrence, all prior to pharmacological or surgical interventions.

Focusing on survival data from Luzhou, we assessed the clinical diagnostic value of HDL-C. Initially, missing data points were omitted, and the dataset was partitioned into training (70%) and testing sets (30%) following a 7:3 ratio. Model development and feature selection were performed on the training set. Subsequently, model performance was validated on both the testing set and the entire dataset. Ten machine learning algorithms, including Random Forest (RF), Support Vector Machines (SVM), Neural Networks (NNet), K-Nearest Neighbors (KNN), and others, were employed to construct classification models for BC diagnosis, incorporating HDL-C alongside common female tumor markers (CA125, CA153, and CEA). Model optimization utilized 5-fold cross-validation and benchmark validation. Model evaluations were conducted using Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC), sensitivity, and specificity. Ultimately, for enhanced interpretability, the developed models were explained using the "iml" and "DALEX" R packages.

The collection of retrospective clinical data for this study was approved and consented by the Ethics Review Committees of the following institutions: Ethics Committee of the Affiliated Hospital of Southwest Medical University (NO.KY2025001), Ethics Committee of the First People's Hospital of Zigong (NO.2023-016), Ethics Committee of Ziyang People's Hospital (NO.2021-052), and Ethics Committee of the Second People's Hospital of Yibin (NO.2023-126-01). Informed consent forms were obtained from all patients prior to admission.

Statistical analysis

In IVW analysis, a P value less than 0.05 is considered suggestive of significance, while a $P_{FDR} < 0.05$ is deemed statistically significant. All MR estimates were presented as odds ratios (OR) with 95% confidence intervals (CI) for outcomes. For MR-Egger intercept method, a two-sided $P < 0.05$ indicated potential directional pleiotropy. An unmatched T-test was used to analyze differences between groups, with bilateral $P < 0.05$ considered statistically significant. Survival curves were generated using the Kaplan-Meier (K-M) method, and log-rank tests were used to compare survival probabilities between UREA and Ca subgroups stratified by ER status. All statistical analyses were conducted using R version 4.3.2, with packages including gtx, MendelianRandomization, TwoSampleMR, ggplot2, ggrepel, grid, gridExtra, gtable, qqman, RColorBrewer, mlr3, mlr3verse, mlr3viz, tidyverse, mlr3measures and RGraphics.

Results

Identification of potential blood and urine biomarkers for BC and Its subtypes in the European population.

In this study, we aimed to investigate the causal relationship between blood and urine biomarkers and BC, initially focusing on the European population, with subsequent validation of findings in East Asian populations, supported by clinical laboratory biochemical data collected. The overall study flowchart is illustrated in Fig. 1.

Employing rigorous statistical methodologies, encompassing 12 MR analysis strategies, sensitivity analyses and R-MR analyses, we meticulously selected exposure factors meeting the following criteria: IVW $P < 0.05$, MR-Egger intercept test $P < 0.05$, and R-MR $P < 0.05$, ensuring the robustness and reliability of our results (Fig. 2). For the European population, our findings revealed suggestive statistical associations between overall BC risk and the following biomarkers: ALP, APOA, HDL-C, insulin-like growth factor 1 (IGF-1), and TG. After correction for multiple comparisons (Type I error adjustment), the associations of all markers with BC remained significant except for APOA. Notably, ALP [OR = 0.947; 95% CI: 0.914–0.982; $P_{FDR} = 0.027$] and TG [OR = 0.927; 95% CI: 0.890–0.965; $p_{FDR} = 0.004$] demonstrated significant genetic protective effects, whereas HDL-C [OR = 1.075; 95% CI: 1.033–1.118; $P_{FDR} = 0.004$] and IGF-1 [OR = 1.081; 95% CI: 1.037–1.128; $P_{FDR} = 0.004$] indicated genetic risk predispositions, with these observations further strengthened through graphical representation (Fig. 3A; Supplementary Table S1-2). Moreover, Fig. 3B illustrates the consistent protective or risky trends of the key biomarkers identified in overall BC across all MR analyses.

Delving into subtype analyses, we observed suggestive causal relationships between CRP and SHBG with ER + BC, yet after adjustment, only HDL-C [OR = 1.091; 95% CI: 1.042–1.142; $P_{FDR} = 0.007$] and TG [OR = 0.922; 95% CI: 0.880–0.966; $P_{FDR} = 0.011$] retained significant associations (Fig. 4A; Supplementary Table S3-4). Surprisingly, in ER$^-$BC, besides ALP, creatinine_in_urine, microalbuminuria_in_urine, and glucose emerged as new exposures, with glucose [OR = 1.217; 95% CI: 1.115–1.328; $P_{FDR} < 0.001$] remaining a genetic risk factor post-correction (Supplementary Table S5-6; Fig. 4A). The scatter plots in Fig. 4B confirm the consistent effect directions of the key biomarkers identified across MR analyses within BC subtypes. Collectively, while the ER$^-$BC sample size was relatively small, it displayed starkly different genetic correlation patterns compared to ER$^+$BC, highlighting the complex differences in genetic determinants between BC subtypes.

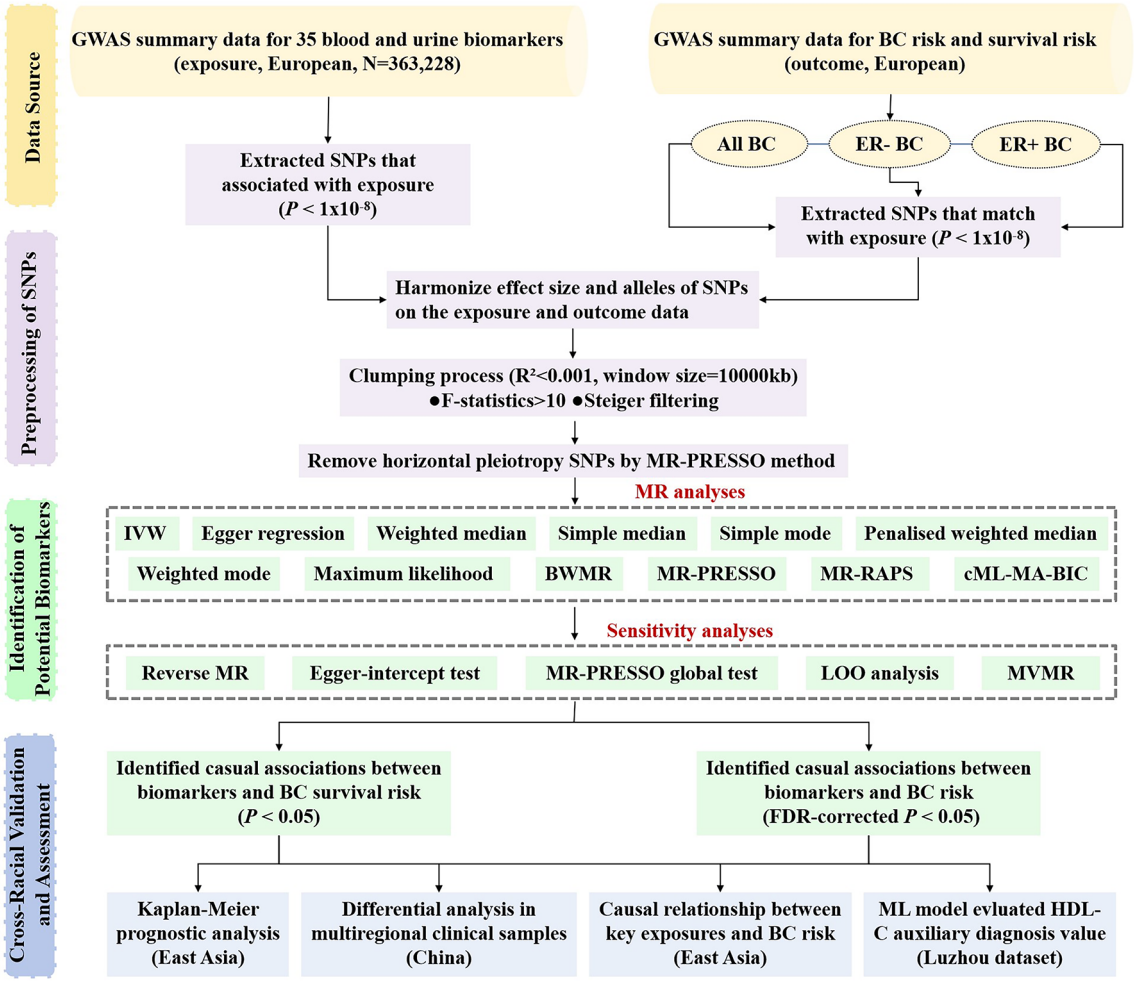


Fig. 1. Flowchart of the analysis process in this study.

outcome	exposure	method	nsnp	pval	pFDR	Q_pval	Q_pFDR	OR(95% CI)
BC	**Alkaline_phosphatase**	Inverse variance weighted	265	0.003	**0.027**	<0.001	<0.001	0.947 (0.914 to 0.982)
	Apolipoprotein_A	Inverse variance weighted	211	0.014	0.081	<0.001	<0.001	1.055 (1.011 to 1.100)
	HDL_cholesterol	Inverse variance weighted	228	<0.001	**0.004**	<0.001	<0.001	1.075 (1.033 to 1.118)
	IGF_1	Inverse variance weighted	285	<0.001	**0.004**	<0.001	<0.001	1.081 (1.037 to 1.128)
	Triglycerides	Inverse variance weighted	205	<0.001	**0.004**	<0.001	<0.001	0.927 (0.890 to 0.965)
ER+_BC	Alkaline_phosphatase	Inverse variance weighted	264	0.027	0.137	<0.001	<0.001	0.955 (0.916 to 0.995)
	Apolipoprotein_A	Inverse variance weighted	212	0.010	0.087	<0.001	<0.001	1.067 (1.016 to 1.121)
	C_reactive_protein	Inverse variance weighted	173	0.015	0.103	<0.001	<0.001	1.076 (1.014 to 1.140)
	HDL_cholesterol	Inverse variance weighted	230	<0.001	**0.007**	<0.001	<0.001	1.091 (1.042 to 1.142)
	IGF_1	Inverse variance weighted	293	0.009	0.087	<0.001	<0.001	1.067 (1.017 to 1.119)
	SHBG	Inverse variance weighted	213	0.026	0.137	<0.001	<0.001	0.946 (0.900 to 0.993)
	Triglycerides	Inverse variance weighted	208	<0.001	**0.011**	<0.001	<0.001	0.922 (0.880 to 0.966)
ER−_BC	Alkaline_phosphatase	Inverse variance weighted	269	0.009	0.089	<0.001	<0.001	0.930 (0.881 to 0.982)
	Creatinine_in_urine	Inverse variance weighted	23	0.044	0.184	0.034	0.045	1.488 (1.010 to 2.193)
	Glucose	Inverse variance weighted_fixed	101	<0.001	**<0.001**	0.105	0.130	1.217 (1.115 to 1.328)
	Microalbumin_in_urine	Inverse variance weighted_fixed	4	0.017	0.072	0.696	0.696	1.738 (1.105 to 2.734)

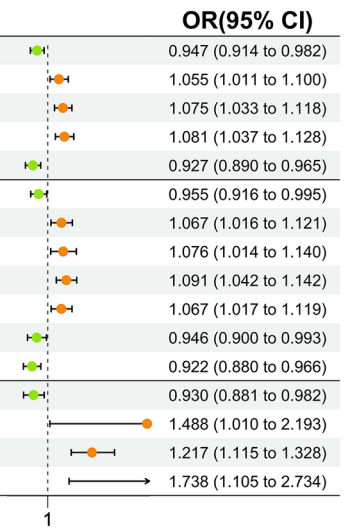


Fig. 2. Mendelian randomization results using the inverse variance weighted method for blood and urine biomarkers in relation to breast cancer.

Of note, we observed a striking phenomenon where TG and HDL-C, typically regarded as risk and protective factors for BC respectively, exhibited counterintuitive results in the European population analysis. Considering the genetic covariance between TG, HDL-C, and APOA, we devised a MVMR strategy to meticulously disentangle their individual and combined causal relationship with BC. Owing to the significant reverse causality observed between APOA and HDL-C with BC in ER-BC, and the lack of a pronounced direct causal relationship for TG with BC, we refrained from implementing multi-variable analysis in the ER-BC subgroup to avoid potential confounding and interpretational intricacies. Instead, we focused on the total BC and ER + BC subgroups, executing MVMR analysis to precisely delineate the independent and interactive roles of these biomarkers. Through this multi-variable MR analysis, we found that the causal connection between HDL-C and BC became insignificant in the total BC population upon adjusting for the potential impacts of APOA and TG, underlining the importance of biomarker interactions. However, in ER + BC, even after adjusting for APOA and TG effects, HDL-C [OR = 2.363; 95% CI: 1.007–5.543; $P = 0.048$] maintained a significant positive correlation with BC risk, revealing the persistence of its genetic risk role in specific subtypes. Additionally, while the direct causal relationship of TG with BC did not reach statistical significance after correcting for HDL-C and APOA, the observed genetic risk trend suggests a more nuanced role for TG in BC etiology, necessitating larger-scale studies and more refined multi-variable analyses to accurately define its function (Fig. 5; Supplementary Table S7).

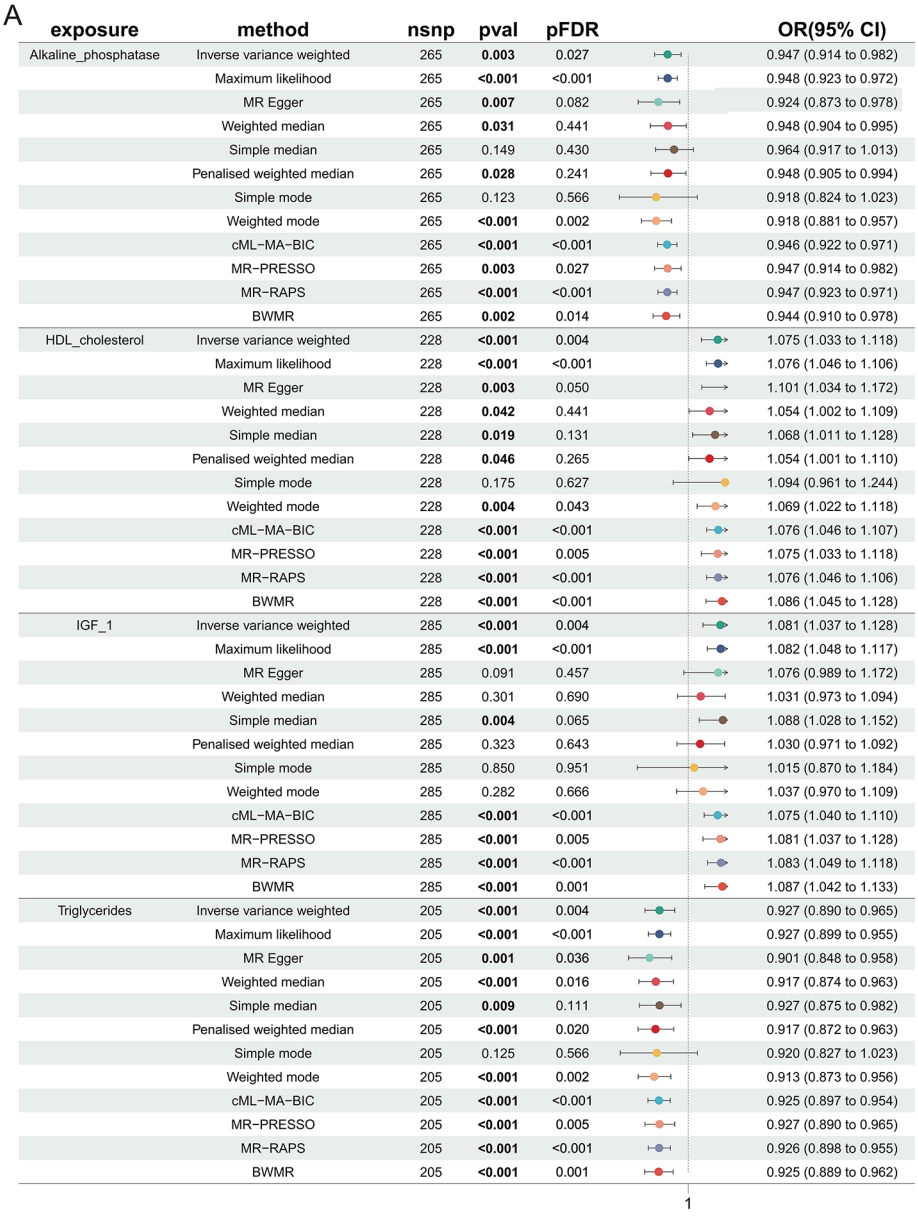
Exploring the performance of potential biomarkers in east Asian populations

To further investigate the genetic causal relationships of the aforementioned suggestive and significant exposure factors in East Asian populations, we obtained GWAS data for ALP, TG, APOA, CRP, SHBG, GLU, and HDL-C from the IEU GWAS platform, designating them as exposure variables with BC in East Asians as the ultimate outcome. Applying a comprehensive suite of 12 MR analysis strategies, we sought to uncover potential causal pathways linking these markers to BC risk. Our analysis revealed that only HDL-C demonstrated a significant genetic protective effect [OR: 0.924, 96% CI: 0.855–0.999, $P = 0.046$], contrasting the observations in European populations and suggesting differing genetic regulatory mechanisms for HDL-C in BC etiology between East Asian and European populations (Fig. 6A; Supplementary Table S8-9). Moreover, a scatter plot analysis (Fig. 6B) showed that all 11 statistical methods, excluding the weighted mode algorithm, consistently supported HDL-C's genetic protective role against BC.

To validate the actual impact of these markers in BC onset, we contrasted serum marker levels between BC patients and those with benign breast diseases utilizing extensive clinical laboratory data. We found elevated concentrations of ALP, GLU, high-sensitivity CRP (hs-CRP), and TG in BC patients compared to controls in the Luzhou dataset, accompanied by a decline in HDL-C levels. These results were replicated in an independent sample set from Zigong, reinforcing the data's reliability and generalizability (Fig. 6C, D). Furthermore, we analyzed differential expression of these markers in BC patients with varying ER statuses within the Luzhou dataset. APOA1 showed a trend towards higher expression in both ER subtypes, while GLU was significantly increased only in ER^+BC. Other exposure variables showed insignificant differences between subtypes, potentially due to small sample sizes insufficient to reveal subtle variations (Supplementary Fig.S1). These findings not only highlight HDL-C's unique role in BC genetic risk but also underscore the importance of trans-ethnic genetic heterogeneity in BC research, offering valuable insights for exploring molecular pathogenesis and formulating population-specific prevention and treatment strategies.

A

exposure	method	nsnp	pval	pFDR	OR(95% CI)
Alkaline_phosphatase	Inverse variance weighted	265	**0.003**	0.027	0.947 (0.914 to 0.982)
	Maximum likelihood	265	**<0.001**	<0.001	0.948 (0.923 to 0.972)
	MR Egger	265	**0.007**	0.082	0.924 (0.873 to 0.978)
	Weighted median	265	**0.031**	0.441	0.948 (0.904 to 0.995)
	Simple median	265	0.149	0.430	0.964 (0.917 to 1.013)
	Penalised weighted median	265	**0.028**	0.241	0.948 (0.905 to 0.994)
	Simple mode	265	0.123	0.566	0.918 (0.824 to 1.023)
	Weighted mode	265	**<0.001**	0.002	0.918 (0.881 to 0.957)
	cML−MA−BIC	265	**<0.001**	<0.001	0.946 (0.922 to 0.971)
	MR−PRESSO	265	**0.003**	0.027	0.947 (0.914 to 0.982)
	MR−RAPS	265	**<0.001**	<0.001	0.947 (0.923 to 0.971)
	BWMR	265	**0.002**	0.014	0.944 (0.910 to 0.978)
HDL_cholesterol	Inverse variance weighted	228	**<0.001**	0.004	1.075 (1.033 to 1.118)
	Maximum likelihood	228	**<0.001**	<0.001	1.076 (1.046 to 1.106)
	MR Egger	228	**0.003**	0.050	1.101 (1.034 to 1.172)
	Weighted median	228	**0.042**	0.441	1.054 (1.002 to 1.109)
	Simple median	228	**0.019**	0.131	1.068 (1.011 to 1.128)
	Penalised weighted median	228	**0.046**	0.265	1.054 (1.001 to 1.110)
	Simple mode	228	0.175	0.627	1.094 (0.961 to 1.244)
	Weighted mode	228	**0.004**	0.043	1.069 (1.022 to 1.118)
	cML−MA−BIC	228	**<0.001**	<0.001	1.076 (1.046 to 1.107)
	MR−PRESSO	228	**<0.001**	0.005	1.075 (1.033 to 1.118)
	MR−RAPS	228	**<0.001**	<0.001	1.076 (1.046 to 1.106)
	BWMR	228	**<0.001**	<0.001	1.086 (1.045 to 1.128)
IGF_1	Inverse variance weighted	285	**<0.001**	0.004	1.081 (1.037 to 1.128)
	Maximum likelihood	285	**<0.001**	<0.001	1.082 (1.048 to 1.117)
	MR Egger	285	0.091	0.457	1.076 (0.989 to 1.172)
	Weighted median	285	0.301	0.690	1.031 (0.973 to 1.094)
	Simple median	285	**0.004**	0.065	1.088 (1.028 to 1.152)
	Penalised weighted median	285	0.323	0.643	1.030 (0.971 to 1.092)
	Simple mode	285	0.850	0.951	1.015 (0.870 to 1.184)
	Weighted mode	285	0.282	0.666	1.037 (0.970 to 1.109)
	cML−MA−BIC	285	**<0.001**	<0.001	1.075 (1.040 to 1.110)
	MR−PRESSO	285	**<0.001**	0.005	1.081 (1.037 to 1.128)
	MR−RAPS	285	**<0.001**	<0.001	1.083 (1.049 to 1.118)
	BWMR	285	**<0.001**	0.001	1.087 (1.042 to 1.133)
Triglycerides	Inverse variance weighted	205	**<0.001**	0.004	0.927 (0.890 to 0.965)
	Maximum likelihood	205	**<0.001**	<0.001	0.927 (0.899 to 0.955)
	MR Egger	205	**0.001**	0.036	0.901 (0.848 to 0.958)
	Weighted median	205	**<0.001**	0.016	0.917 (0.874 to 0.963)
	Simple median	205	**0.009**	0.111	0.927 (0.875 to 0.982)
	Penalised weighted median	205	**<0.001**	0.020	0.917 (0.872 to 0.963)
	Simple mode	205	0.125	0.566	0.920 (0.827 to 1.023)
	Weighted mode	205	**<0.001**	0.002	0.913 (0.873 to 0.956)
	cML−MA−BIC	205	**<0.001**	<0.001	0.925 (0.897 to 0.954)
	MR−PRESSO	205	**<0.001**	0.005	0.927 (0.890 to 0.965)
	MR−RAPS	205	**<0.001**	<0.001	0.926 (0.898 to 0.955)
	BWMR	205	**<0.001**	0.001	0.925 (0.889 to 0.962)



B

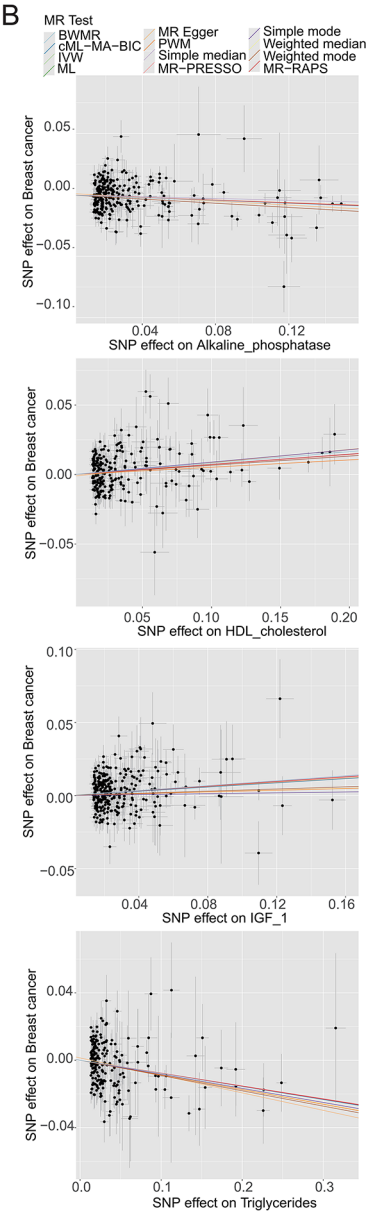


Fig. 3. Results from 12 Mendelian randomization (MR) analyses for blood and urine biomarkers in relation to overall breast cancer. (**A**). The forest plot displays the results of the 12 MR analyses, showing odds ratios (OR), SNPs, and p-values; (**B**). Scatter plots illustrate the genetically predicted causal effects of key biomarkers on overall breast cancer risk.

HDL-C may be serve as a potential auxiliary diagnostic biomarker for BC

Given HDL-C's significant causal relationship with BC in East Asians and notable serum level discrepancies between benign breast diseases and BC patients, we probed HDL-C's potential as an auxiliary diagnostic marker for BC differentiation. Leveraging the Luzhou dataset, we integrated conventional tumor markers (CEA, CA125, and CA153) with HDL-C to formulate a multi-marker diagnostic model, assessing its clinical utility via machine learning techniques. Implementing 10 advanced machine learning algorithms with 5-fold cross-validation strategies involving constant proportion sampling and oversampling, we comprehensively compared each model's performance under different preprocessing conditions (standardized and non-standardized). Based on multi-criterion assessment (Fig. 7A), the NNet model demonstrated the highest AUC and lowest misclassification rate, exhibiting excellent overall performance despite room for improvement in specificity. Confusion matrices detailed the model's classification efficacy across training, validation, and combined datasets (Fig. 7B). ROC curve analysis further revealed that a composite blood biomarker model comprised of CA125, CA153, CEA, and HDL-C (overall AUC = 0.728; training AUC = 0.727; testing AUC = 0.731) significantly outperformed the traditional tumor marker model lacking HDL-C (overall AUC = 0.695; training AUC = 0.696 ; testing AUC = 0.694) in distinguishing BC from non-malignant breast lesions (Fig. 7C). Additionally, the composite model displayed good consistency and clinical practicality across all dataset subsets through calibration and

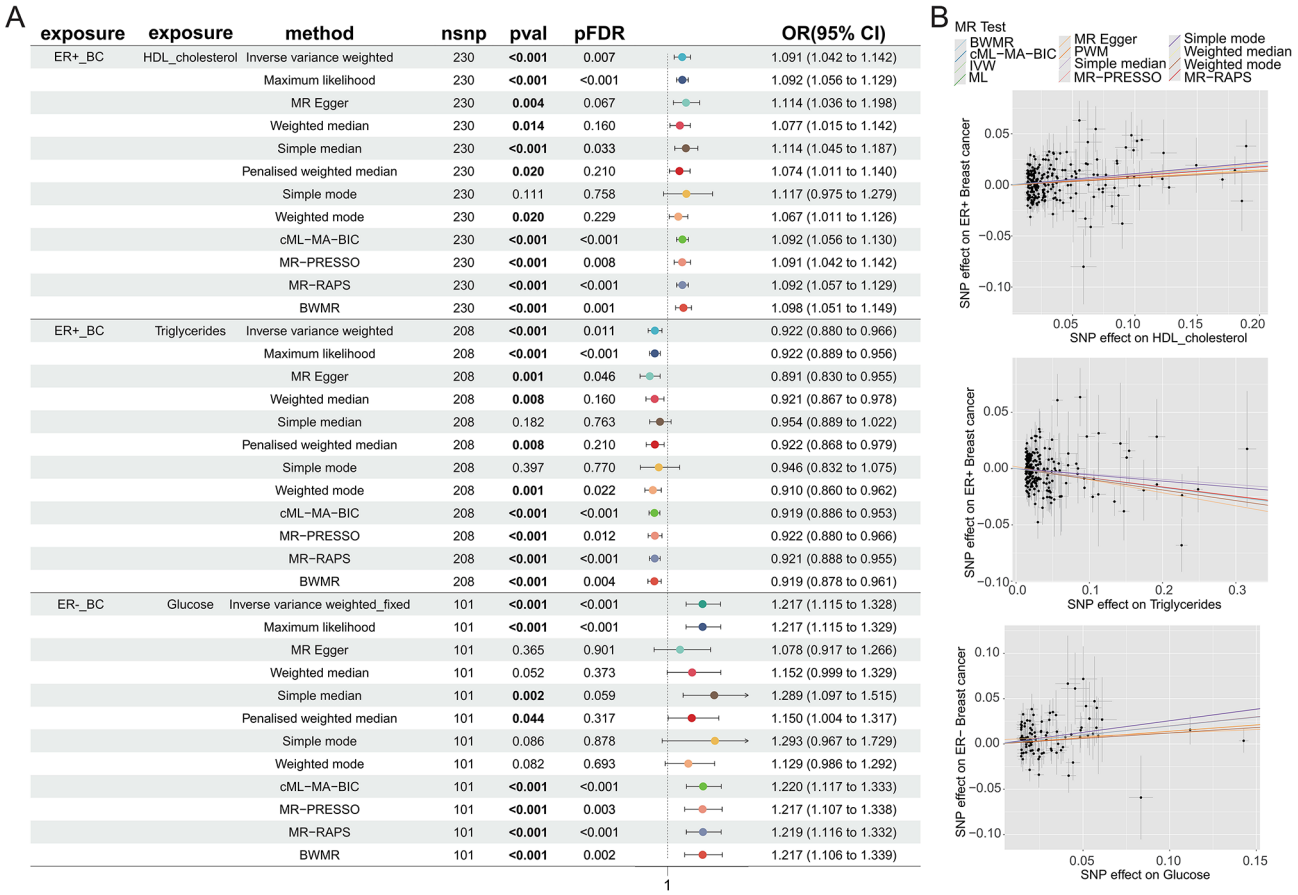


Fig. 4. Results from 12 Mendelian randomization (MR) analyses for blood and urine biomarkers in relation to breast cancer subtypes. (**A**). The forest plot displays the results of the 12 MR analyses, showing odds ratios (OR), SNPs, and p-values; (**B**). Scatter plots illustrate the genetically predicted causal effects of key biomarkers on breast cancer subtypes risk.

decision curves (Fig. 7D, E). Collectively, our study indicates that incorporating HDL-C into existing tumor marker panels significantly enhances BC diagnostic accuracy, providing a promising auxiliary diagnostic tool for early BC recognition, particularly in East Asian health management and cancer screening.

To deepen our understanding of model features' contributions to BC diagnostic prediction, we employed the IML and DALEX R packages for feature importance interpretation. Analyses showed that HDL-C significantly positively influenced model predictive performance across all three datasets, underscoring its value in the diagnostic model (Fig. 7F, G). This finding further supports HDL-C's potential role in BC risk assessment. Accumulated Local Effects (ALE) analysis revealed that CA153, CA125, and CEA's mean predicted values correlated positively with BC prediction probabilities, while HDL-C had a negative correlation, aligning with previous causal analysis results and robustly supporting HDL-C's genetic protective role in East Asians (Supplementary Fig.S2). To gain individual-level insights, we randomly selected one case from the training, validation, and full datasets for single-sample explanation. HDL-C displayed a near-normal distribution pattern across datasets, with generally lower measurements in BC patients. SHapley Additive Explanations (SHAP) and Break Down Plot analyses of these random cases showed that HDL-C exerted varying degrees of negative contribution to each patient's diagnosis, indicating that low HDL-C levels tend to decrease BC prediction probabilities (Supplementary Fig.S3-S5). These results consistently demonstrate HDL-C's practical auxiliary value in BC diagnosis at both population and individual levels, emphasizing its possible protective role in BC development. In conclusion, our study confirms the clinical significance of HDL-C in diagnosing BC in East Asian populations and provides insights into its potential mechanism as a BC genetic protective factor.

Causal effects of BC survival

In the in-depth analysis of BC survival rates, we unveiled a suggestive risk impact of blood UREA on overall BC survival, with a hazard ratio (HR) of 1.230 and a 95% CI ranging from 1.000 to 1.513, reaching marginal significance ($P = 0.049$). More notably, in ER$^-$BC patients, this risk effect was even more pronounced, exhibiting an HR of 1.749, 95% CI of 1.180 to 2.592, with $P = 0.005$ (Fig. 8; Supplementary Table S10). Concurrently, elevated serum Ca levels hinted at a protective effect on the survival of ER$^-$BC patients, with an HR of 0.833, 95% CI of 0.702 to 0.989, and $P = 0.037$. However, similar correlations were not evident in ER$^+$BC. Although these suggestive findings lost statistical significance after multiple comparison adjustments, we proceeded to

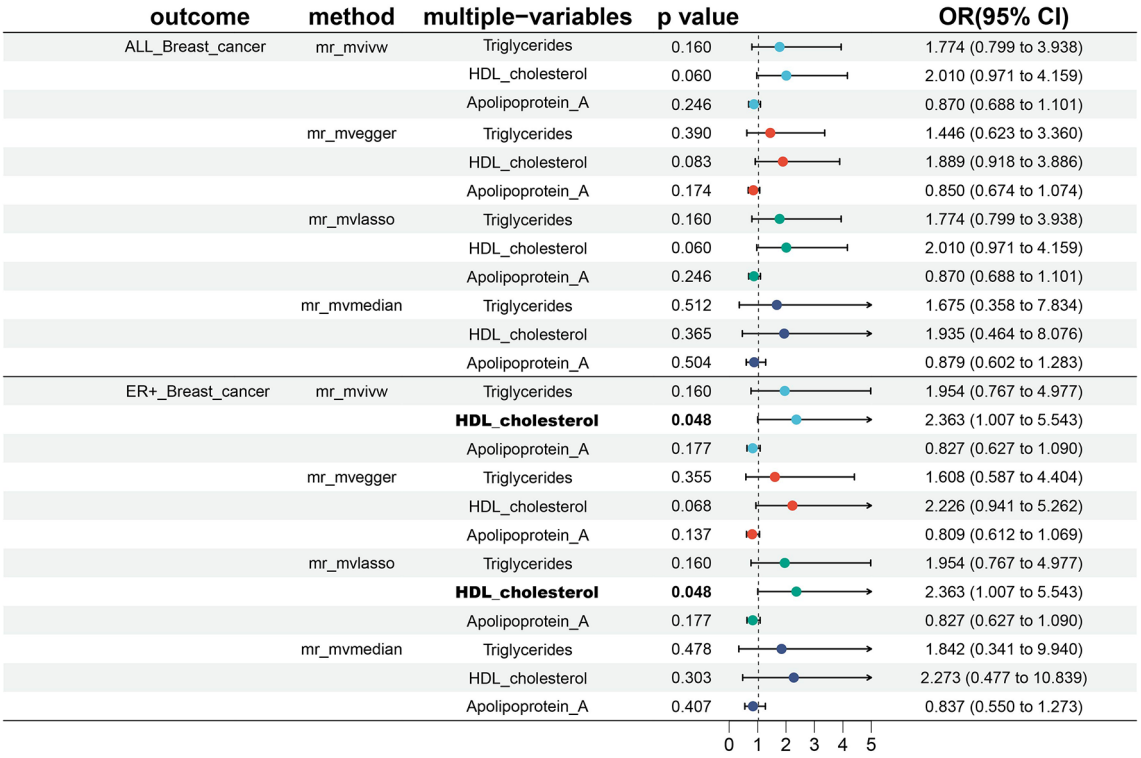


Fig. 5. The forest plot displays the results of the multivariate mendelian randomization analysis between TG, APOA, HDL-C and breast cancer.

examine the clinical relevance of these markers in East Asian populations. Comparisons of UREA and Ca clinical laboratory test results between surviving and deceased BC patients across multiple datasets were conducted. Analysis of the Luzhou dataset revealed significantly lower UREA levels in surviving BC patients compared to those who succumbed to the disease, with no significant difference noted in subtype analyses. Regarding calcium, a significantly higher level was observed only in surviving ER$^-$BC patients. While the Zigong and Ziyang datasets, due to their smaller sample sizes, failed to reveal statistically significant differences, their data trends aligned with the Luzhou dataset, suggesting a potential association between changes in UREA and Ca levels and survival status (Supplementary Fig.S6). Subsequently, K-M survival analysis on the Luzhou dataset further confirmed UREA as a risk factor for overall BC and Ca as a protective factor for ER$^-$BC, aligning with our previous Mendelian randomization findings in European populations (Fig. 9).

Preliminary investigation of potential biomarker serological alterations in BC recurrence

Following a series of in-depth analyses, we have uncovered a thought-provoking insight: numerous conventionally recognized non-specific indicators, typically deemed to possess singular diagnostic implications in clinical settings, may in fact play more intricate and multidimensional roles throughout the development, progression, and survival outcomes of cancer, transcending our conventional understanding. Although current research falls short of fully elucidating the complex mechanisms and precise functions of these markers, it has sufficiently ignited a reassessment of the functional capacities of these traditional indicators.

In pursuit of a deeper understanding of the potential significance of these markers, we systematically tracked and examined the laboratory records of 210 patients experiencing tumor recurrence. The results revealed notably elevated levels of GLU and TG in the blood tests of recurrent cases, while other conventional indicators did not exhibit significant variations (Fig. 10). This observation intimates a potential correlation between GLU and TG levels and the recurrence dynamics of the tumor. Nonetheless, the limitations of this finding must be acknowledged, chiefly the modest sample size, which constrains the generalizability and robustness of our conclusions.

Discussion

This study, through multidimensional and cross-ethnic explorations, has enriched our understanding of the complex genetic associations between biomarkers in blood and urine and BC. While causal analyses between plasma metabolites and BC have been extensively pursued, the uniqueness of our research lies in the utilization of genetic data of 35 recently blood and urine biomarkers, and the application of multiple advanced MR methodologies, including R-MR, sensitivity analyses, and MVMR approaches, to comprehensively and rigorously assess the causal relations between these markers and BC risk as well as survival outcomes. Not only were these findings validated in European populations, but through comparative analyses, we also evaluated the similarities

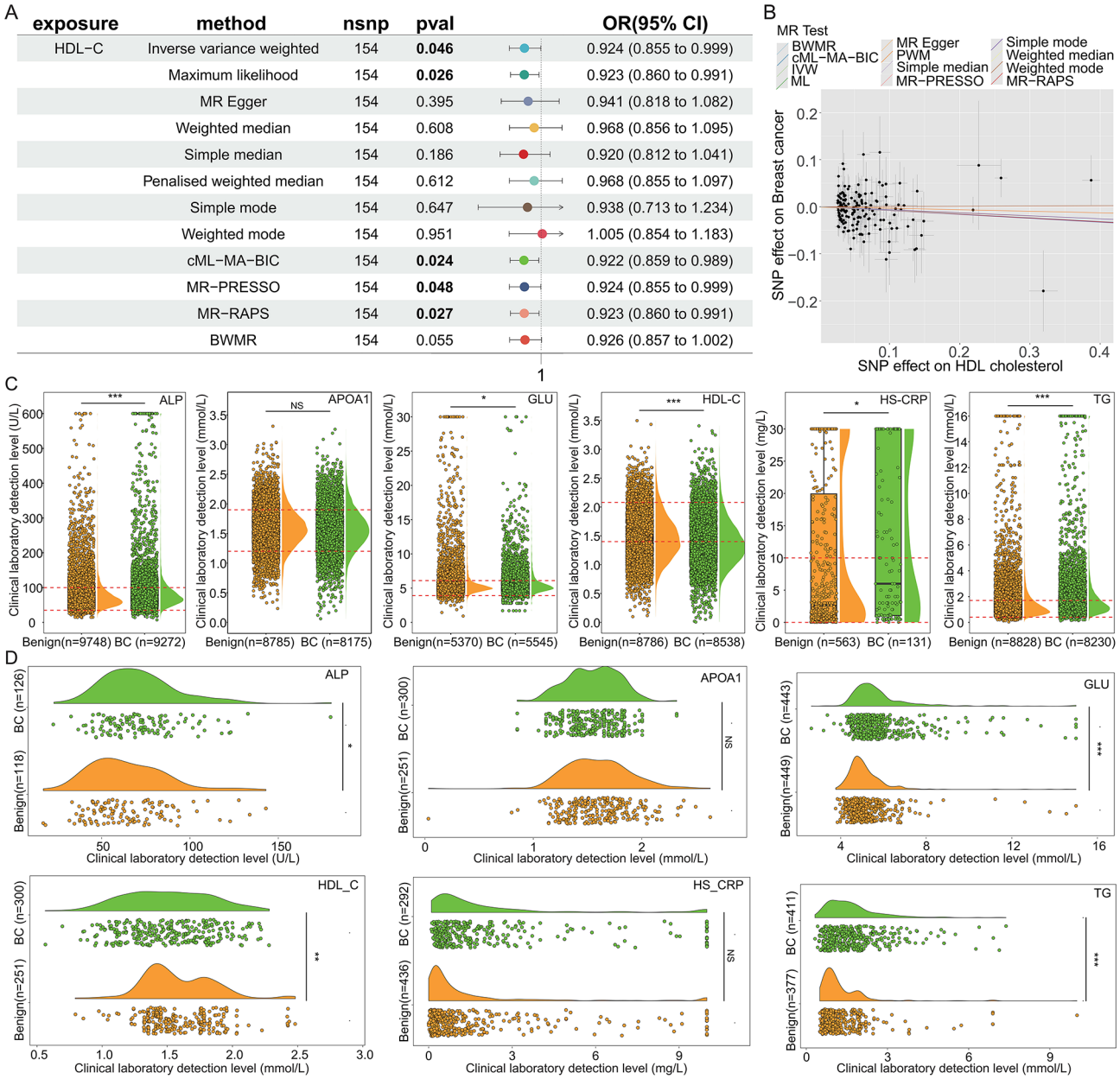


Fig. 6. Exploring the performance of potential biomarkers in East Asian Populations. (**A**). The forest plot displays the results of the 12 MR analyses between HDL-C and overall breast cancer in East Asian Populations; (**B**). Scatter plots illustrate genetically predicted causal effects of HDL-C on breast cancer risk; (**C**). Analysis of differential detection levels of potential biomarkers in the clinical dataset from Luzhou (Breast cancer vs. Benign breast diseases); (**D**). Analysis of differential detection levels of potential biomarkers in the clinical dataset from Zigong (Breast cancer vs. Benign breast diseases). Note: "***": $P<0.001$; "*": $P<0.05$; "NS": No statistical significance.

and discrepancies of these genetic effects in East Asian populations, further harnessing ML algorithms alongside conventional statistical models to gauge the clinical translational value of potential biomarkers, striving to convert fundamental genetic insights into practical clinical applications.

Reinforcing and expanding the scientific basis of known biomarkers

Our study confirms the associations of biomarkers such as ALP, HDL-C, TG, SHBG, IGF2 with BC risk, aligning with a substantial body of previous research and thereby reinforcing the reliability and stability of our findings. ALP, commonly used in clinical monitoring and associated with various diseases, is linked to metastasis and poor prognosis in BC contexts[35]. Our clinical data difference analysis also showed higher levels of ALP detection in BC. Intriguingly, our MR analysis in European populations observed a genetically protective effect of ALP

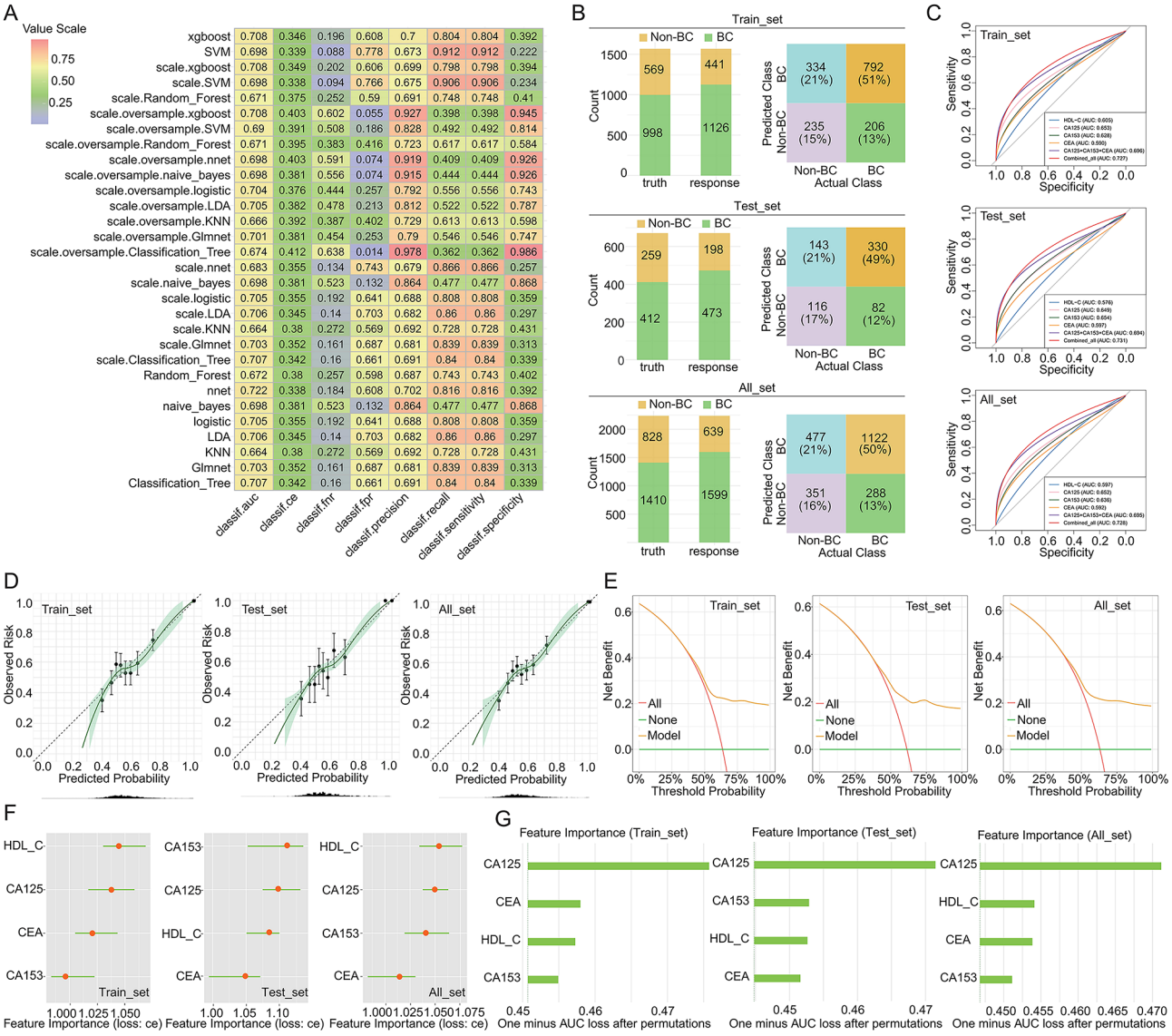


Fig. 7. Machine learning to assess the auxiliary diagnostic value of HDL-C in breast cancer. (**A**). Heatmap displaying performance evaluation of ten machine learning algorithms; (**B**). Confusion matrix evaluation; (**C**). ROC curve evaluation; (**D**). Calibration curve evaluation; (**E**). Decision curve analysis; (**F**). Model feature importance interpretation using IML R package; (**G**). Model feature importance interpretation using DALEX R package.

against BC, contrasting with its traditional clinical interpretation. This discovery underscores a unique role for ALP within the genetic landscape of BC, warranting further mechanistic investigations.

Regarding TG, typically viewed as a risk factor for conditions like cardiovascular disease, our univariate MR analysis intriguingly suggests a genetic protective trend vis-à-vis BC, particularly in ER$^+$BC. The protective effect could be due to the complex interplay between lipids and other metabolic pathways. Although epidemiological studies present conflicting evidence regarding the relationship between TG and BC risk[14–16,36], our multifactorial analysis discounts significant interactive effects, highlighting the need for cautious interpretation of the TG-BC risk link and advocating for more extensive studies in the future. Clinically, elevated TG levels in BC patients further complicate this area.

The correlation between IGF-1 and BC risk is corroborated in our study, consistent with prior research. For instance, Murphy et al. reported that each 5nmol/L increment in IGF-1 concentration was associated with increased BC risk (OR = 1.11, 95%CI: 1.11–1.16) after adjustment[37], suggesting potential benefits of IGF-1 pathway interventions for breast tumorigenesis prevention. Early and subsequent prospective studies consistently demonstrate a positive correlation between IGF-1 levels and BC risk, particularly for ER$^+$BC[38–40]. These findings highlight IGF-1's role in BC susceptibility and emphasize its importance as both a risk predictive biomarker and a potential therapeutic target, especially in specific subtypes.

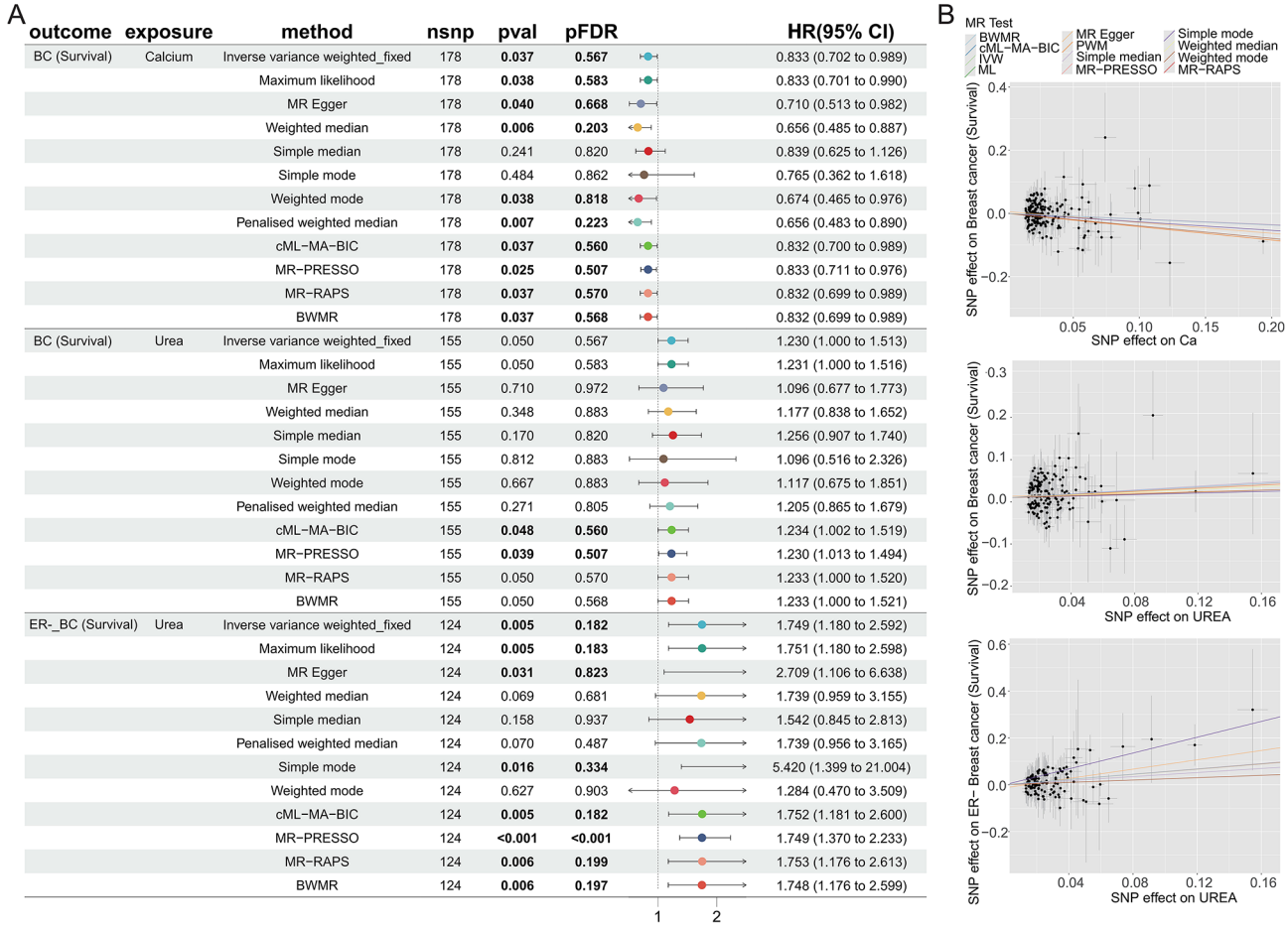


Fig. 8. Results from 12 Mendelian randomization (MR) analyses for blood and urine biomarkers in relation to breast cancer survival. (**A**). The forest plot displays the results of the 12 MR analyses, showing odds ratios (OR), SNPs, and p-values; (**B**). Scatter plots illustrate the genetically predicted causal effects of key biomarkers on breast cancer survival.

Unveiling the dual nature of HDL-C and the complexity of its diagnostic implications

The relationship between HDL-C and BC has been variably depicted in observational epidemiological studies, ranging from positive, negative, to no significant associations[12,13,41–43]. Our study reinforces in European populations the connection between genetically predicted high HDL-C and increased BC risk, particularly in ER^+BC, a finding consistent across multiple analytical methods, demonstrating high robustness and consistency. These results align with the findings of Johnson et al.[44]. In MVMR analyses, when APOA, HDL-C, and TG were concurrently considered, their associations with BC risk generally attenuated, failing to show significant causal effects. However, it is noteworthy that in ER^+BC specifically, the significant correlation between HDL-C and risk persisted, implying HDL-C could be a pivotal driver of the genetic associations between APOA, TG, and ER^+BC risk in a univariate context. After adjusting for APOA and TG, our data strongly support a causal relationship between HDL-C and ER^+BC risk, aligning with a recent study that found a per-standard deviation increase in genetically determined HDL-C associated with increased risk of ER^+BC (OR: 1.13, 95% CI: 1.01–1.26)[41]. Conversely, Zhou et al.'s study observed causal effects of HDL-C only in ER^-BC after multivariate analysis[45]. In contrast, our study identified a significant reverse causal relationship between HDL-C and ER^-BC, highlighting the intricate nature of inferring causality for HDL-C and ER^-BC risk, which challenges straightforward conclusions.

Biologically, the dual nature of HDL-C in BC risk may reflect its functional heterogeneity. As a crucial component in lipid transportation, HDL-C is typically associated with cardiovascular health and lower cardiovascular disease risk[46]. However, recent MR analyses have questioned whether elevating HDL-C levels directly reduces coronary heart disease risk[47,48]. Furthermore, pharmaceutical interventions aimed at increasing HDL-C to improve health outcomes have yielded inconsistent results[49,50], leading the National Lipid Association to explicitly state that HDL-C should not currently be a treatment target[46]. This suggests that the functional evaluation of HDL-C may be more important than its absolute level, given that HDL-C from different sources varies in biological function[46]. For example, oxidized HDL-C and HDL-C in type 2 diabetes patients enhance proliferation, migration, and metastatic abilities of BC cells[51,52]. These findings suggest that the functional quality of HDL-C, rather than its absolute level, may be a more relevant determinant of BC risk. Furthermore,

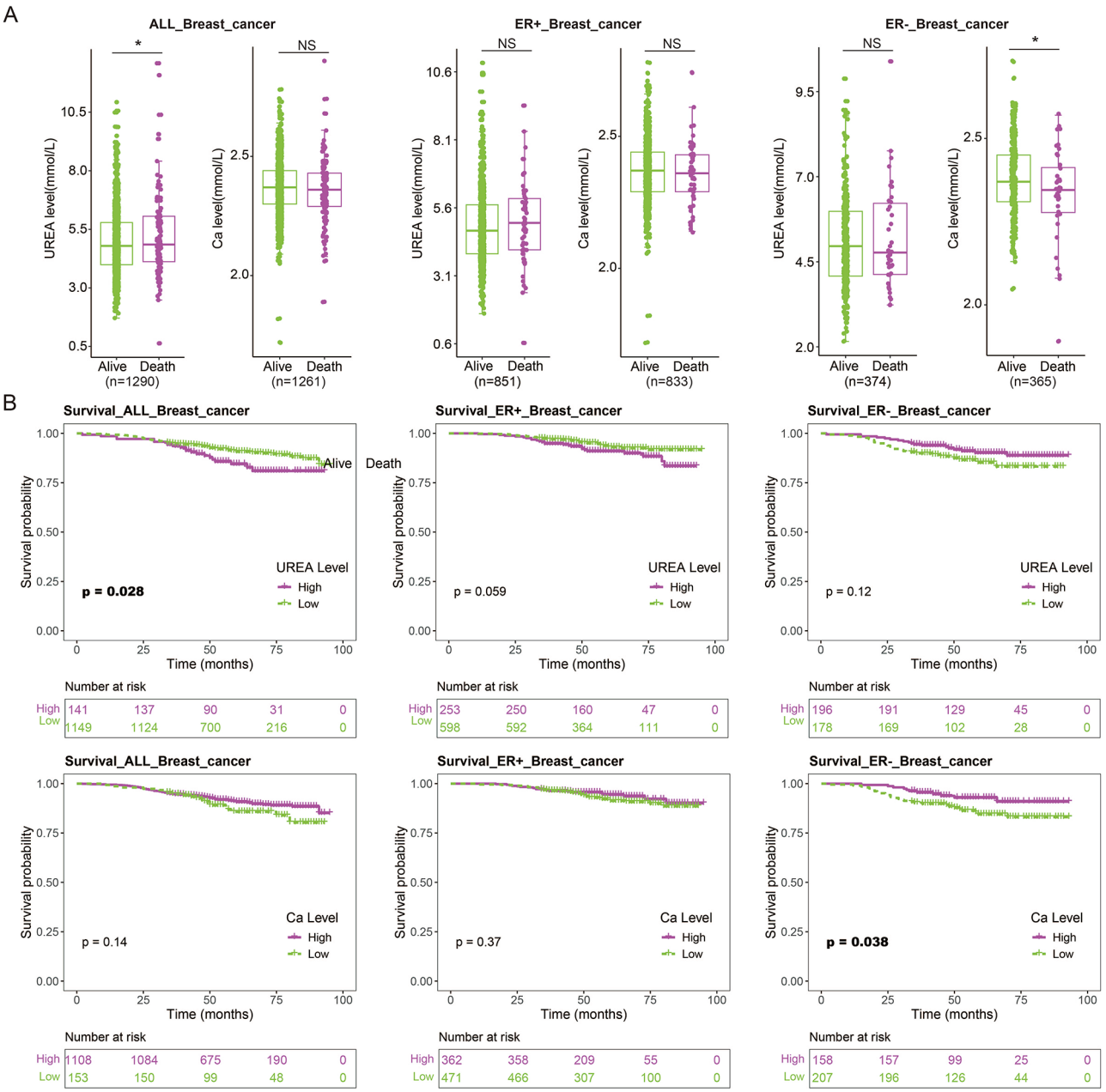


Fig. 9. Analysis of survival disparities and Kaplan-Meier prognostic assessment for UREA and Ca in breast cancer. (**A**). Evaluation of survival differences for UREA and Ca in the total breast cancer population and stratified by ER status within the Luzhou dataset; (**B**). Kaplan-Meier prognostic analysis for UREA and Ca in the overall breast cancer cohort and according to ER status subsets, using data from Luzhou. Note: "*": $P < 0.05$; "NS": No statistical significance.

the interplay between HDL-C and other lipid components, such as TG and APOA, may further complicate its role. In our study, the protective effect of TG against BC risk contrasts with the risk-increasing effect of HDL-C, suggesting that the balance between different lipid fractions may be critical in BC etiology.

Interestingly, our study reveals new evidence in East Asian populations, indicating that higher genetically predicted HDL-C is associated with reduced BC risk, which is entirely opposite to the findings in European populations. Our research is the first to suggest that the impact of HDL-C on BC risk may be profoundly influenced by population-specific factors, such as genetic background, dietary patterns, and metabolic profiles, may modulate the relationship between lipids and BC risk. This finding emphasizes the necessity of evaluating HDL-C and BC subtype associations in different population contexts, particularly in ER^+ and ER^- BC subtypes, with a call for more East Asian data to elucidate the precise role of HDL-C in BC etiology in this region. Our study further validates the potential of HDL-C as an auxiliary diagnostic tool through machine learning techniques, paving the way for advancements in precision medicine.

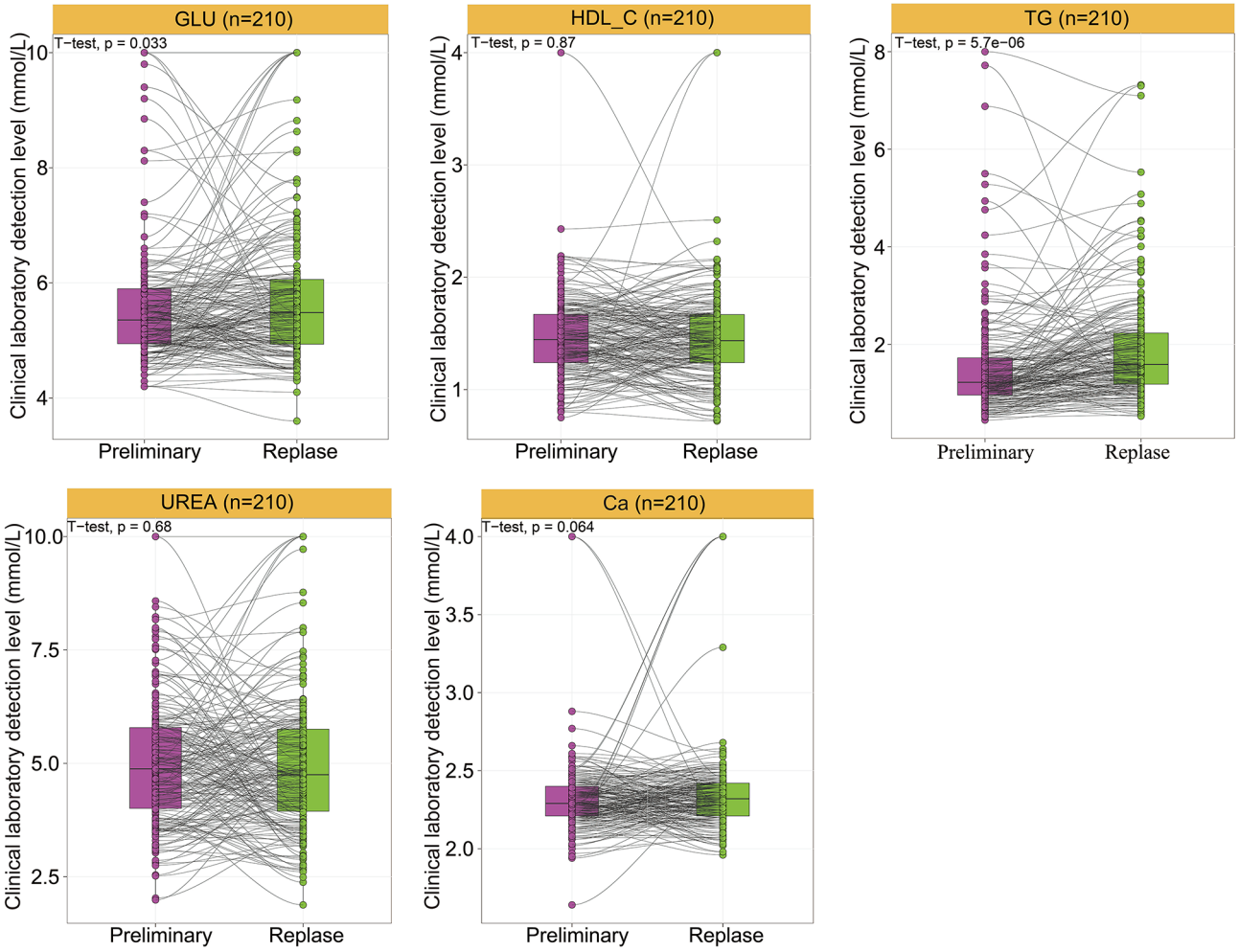


Fig. 10. Comparative analysis of potential biomarker levels in paired recurrent samples from the Yibin dataset.

Elucidating subtype-specific genetic markers in BC

In the analysis of BC subtypes, we uncovered distinct genetic determinants varying across subtypes. ER$^+$BC exhibited significant or suggestive associations with HDL-C, TG, SHBG, and CRP, whereas ER$^-$BC revealed novel links with GLU urinary creatinine and microalbuminuria. These findings highlight the importance of subtype-specific genetic characteristics and personalized prevention and treatment strategies for BC.

Of particular note in ER$^+$BC, SHBG exerts its protective role by regulating the effective concentration of estrogens. By binding to sex hormones, SHBG reduces free hormone levels, which is associated with a decreased risk of estrogen-mediated BC[53–55]. While some argue that the association between SHBG levels and BC risk may not be causal[17], a wealth of literature supports the significance of SHBG in the course of BC[54,56–63]. Studies indicate that the association between SHBG levels and BC risk is influenced by menopausal status and tumor molecular subtypes, commonly manifesting a protective effect in postmenopausal women[56–58,62,63]. Moreover, SHBG inhibits BC cell proliferation and promotes apoptosis through multiple pathways, including direct estrogenic action inhibition, modulation of intracellular signaling (such as cAMP increase and anti-apoptotic gene suppression), and impacting estrogen-dependent gene expression[61]. Consequently, SHBG exerts a complex and profound impact on the initiation and progression of BC through these mechanisms.

Previous studies on circulating GLU have shown inconsistent associations with BC development and progression[64,65]. Recently, an MR study by Jung et al. reported no significant association between genetically determined blood glucose levels and BC risk in European women[66]. In contrast, our analysis uncovers an intriguing discrepancy: while no causal relationship between GLU and BC was detected in the overall BC population, a clear causal association emerged in the ER$^-$BC subtype. This finding not only supplements existing knowledge gaps but also underscores the complexity and specificity of metabolic markers' actions across different BC subtypes, suggesting a potentially more specific role for GLU in the etiology of ER$^-$BC. Therefore, our results imply that considering subtype specificity and variations in metabolic markers is crucial when assessing risk factors for BC.

Exploring new directions in prognostic biomarkers

While our study's findings on UREA and Ca in BC survival analyses are constrained by limited sample sizes and issues of multiple comparisons, their potential roles in different BC subtypes open up new avenues for research in survival prediction and monitoring treatment responses. BUN levels are closely tied to the development and progression of various diseases, long-term mortality, and survival rates[67,68], with high BUN levels having been implicated in increased cancer risk, particularly in BC[69]. Our study is the first to reveal an association between UREA levels and BC survival, notably in ER⁻BC, filling a gap in the field's understanding.

Ca, as the most abundant divalent cation in the human body, plays a central regulatory role in cellular proliferation, differentiation, and apoptosis, among other vital processes[69]. Prior research has associated BC onset and more aggressive tumor characteristics with elevated pre-diagnostic serum calcium levels[70–72]. Although studies examining the relationship between calcium intake and BC survival outcomes are scarce and yield mixed results[73–76], randomized controlled trials have not provided conclusive evidence for a protective effect of Ca and vitamin D supplementation on BC risk or overall cancer mortality[77,78]. An exploratory analysis by Huss et al. suggested a correlation between higher prediagnostic serum calcium levels and lower BC-specific mortality[79]. In our investigation, although no significant association between calcium and survival was observed in the overall BC cohort, a suggestive correlation was found in the ER⁻BC subtype, implying heterogeneous effects of calcium across different BC subtypes.

Despite the use of advanced analytical methods and multiple calibration strategies, potential confounding factors and other biases that were not fully controlled for in the study could impact the interpretation and generalizability of the results. Specifically, due to limited sample sizes, particularly in East Asian populations, the representativeness and generalizability of the study findings may be constrained. Therefore, future research should aim to increase sample sizes, further validate and replicate these findings to confirm the exact associations between biomarkers and breast cancer, and promote the application and implementation of precision medicine across diverse populations.

Conclusion

Our study provides crucial scientific underpinnings and directional guidance for the development of precision diagnostics and treatment strategies in BC, optimization of prognostic assessment systems, and exploration of novel therapeutic targets. By integrating cutting-edge genetic research with clinical practice, our study not only enhances understanding of the complex genetic factors underlying BC but also lays a robust foundation for advancing personalized management and therapeutic strategies for BC.

Data availability

The genetic association datasets employed in this study were sourced from the IEU GWAS database, accessible at (https://gwas.mrcieu.ac.uk/); The clinical data involved in this study is available upon request from the corre sponding author.

Received: 1 August 2024; Accepted: 13 February 2025
Published online: 25 February 2025

References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74** (3), 229–263 (2024).
2. Skvortsova, L. et al. Association of genetic markers with the risk of early-onset breast cancer in Kazakh women. *Genes (Basel)* **15**(1). (2024).
3. Wu, B., Peng, Y., Eggert, J. & Alexov, E. Novel genetic markers for early detection of elevated breast cancer risk in women. *Int. J. Mol. Sci.* **20**(19). (2019).
4. Shu, X. et al. Associations of obesity and circulating insulin and glucose with breast cancer risk: a mendelian randomization analysis. *Int. J. Epidemiol.* **48** (3), 795–806 (2019).
5. Armstrong, K., Micco, E., Carney, A., Stopfer, J. & Putt, M. Racial differences in the use of BRCA1/2 testing among women with a family history of breast or ovarian cancer. *JAMA* **293** (14), 1729–1736 (2005).
6. Nelson, H. D., O'Meara, E. S., Kerlikowske, K., Balch, S. & Miglioretti, D. Factors associated with rates of false-positive and false-negative results from digital mammography screening: an analysis of registry data. *Ann. Intern. Med.* **164** (4), 226–235 (2016).
7. Miglioretti, D. L. et al. Radiation-induced breast cancer incidence and mortality from digital mammography screening: a modeling study. *Ann. Intern. Med.* **164** (4), 205–214 (2016).
8. Subramani, R., Poudel, S., Smith, K. D., Estrada, A. & Lakshmanaswamy, R. Metabolomics of breast cancer: a review. *Metabolites* **12**(7). (2022).
9. Umemori, Y. et al. Evaluating the utility of N1,N12-diacetylspermine and N1,N8-diacetylspermidine in urine as tumor markers for breast and colorectal cancers. *Clin. Chim. Acta*. **411** (23–24), 1894–1899 (2010).
10. Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27** (8), 1133–1163 (2008).
11. Palmer, T. M. et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* **21** (3), 223–242 (2012).
12. Martin, L. J. et al. RE: serum lipids, lipoproteins, and risk of breast cancer: a nested case-control study using multiple time points. *J. Natl. Cancer Inst.* **108**(8). (2016).
13. Beeghly-Fadiel, A. et al. A mendelian randomization analysis of circulating lipid traits and breast cancer risk. *Int. J. Epidemiol.* **49** (4), 1117–1131 (2020).
14. Borena, W. et al. Serum triglycerides and cancer risk in the metabolic syndrome and cancer (Me-Can) collaborative study. *Cancer Causes Control.* **22** (2), 291–299 (2011).
15. Yang, C. et al. Causal relevance of circulating high-density lipoprotein cholesterol with cancer: a mendelian randomization meta-analysis. *Sci. Rep.* **5**, 9495 (2015).
16. Orho-Melander, M. et al. Blood lipid genetic scores, the HMGCR gene and cancer risk: a mendelian randomization study. *Int. J. Epidemiol.* **47** (2), 495–505 (2018).

17. Thompson, D. J. et al. Identification of common variants in the SHBG gene affecting sex hormone-binding globulin levels and breast cancer risk in postmenopausal women. *Cancer Epidemiol. Biomarkers Prev.* **17** (12), 3490–3498 (2008).
18. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **551** (7678), 92–94 (2017).
19. Guo, Q. et al. Identification of novel genetic markers of breast cancer survival. *J. Natl. Cancer Inst.* ; **107**(5). (2015).
20. Ishigaki, K. et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat. Genet.* **52** (7), 669–679 (2020).
21. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562** (7726), 203–209 (2018).
22. Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53** (2), 185–194 (2021).
23. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53** (10), 1415–1424 (2021).
24. Weng, H. et al. Association between uric acid and risk of venous thromboembolism in east Asian populations: a cohort and mendelian randomization study. *Lancet Reg. Health West. Pac.* **39**, 100848 (2023).
25. Long, Y., Tang, L., Zhou, Y., Zhao, S. & Zhu, H. Causal relationship between gut microbiota and cancers: a two-sample mendelian randomisation study. *BMC Med.* **21** (1), 66 (2023).
26. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat. Genet.* **50** (5), 693–698 (2018).
27. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* ; 7. (2018).
28. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37** (7), 658–665 (2013).
29. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some Invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** (4), 304–314 (2016).
30. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* **44** (2), 512–525 (2015).
31. Xue, H., Shen, X. & Pan, W. Constrained maximum likelihood-based mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *Am. J. Hum. Genet.* **108** (7), 1251–1269 (2021).
32. Zhao, J. et al. Bayesian weighted mendelian randomization for causal inference based on summary statistics. *Bioinformatics* **36** (5), 1501–1508 (2020).
33. Li, Y. et al. The effects of coagulation factors on the risk of endometriosis: a mendelian randomization study. *BMC Med.* **21** (1), 195 (2023).
34. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in mendelian randomization studies. *Hum. Mol. Genet.* **27** (R2), R195–R208 (2018).
35. Jiang, C., Hu, F., Li, J., Gao, G. & Guo, X. Diagnostic value of alkaline phosphatase and bone-specific alkaline phosphatase for metastases in breast cancer: a systematic review and meta-analysis. *Breast Cancer Res. Treat.* **202** (2), 233–244 (2023).
36. Kapil, U., Bhadoria, A. S., Sareen, N., Singh, P. & Dwivedi, S. N. Total cholesterol and triglyceride levels in patients with breast cancer. *J. Breast Cancer*. **16** (1), 129–130 (2013).
37. Murphy, N. et al. Insulin-like growth factor-1, insulin-like growth factor-binding protein-3, and breast cancer risk: observational and mendelian randomization analyses with approximately 430 000 women. *Ann. Oncol.* **31** (5), 641–649 (2020).
38. Peyrat, J. P. et al. Plasma insulin-like growth factor-1 (IGF-1) concentrations in human breast cancer. *Eur. J. Cancer*. **29A** (4), 492–497 (1993).
39. Hankinson, S. E. et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. *Lancet* **351** (9113), 1393–1396 (1998).
40. Endogenous, H. et al. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. *Lancet Oncol.* **11** (6), 530–542 (2010).
41. Nowak, C. & Arnlov, J. A mendelian randomization study of the effects of blood lipids on breast cancer risk. *Nat. Commun.* **9** (1), 3957 (2018).
42. Touvier, M. et al. Cholesterol and breast cancer risk: a systematic review and meta-analysis of prospective studies. *Br. J. Nutr.* **114** (3), 347–357 (2015).
43. His, M. et al. Associations between serum lipids and breast cancer incidence and survival in the E3N prospective cohort study. *Cancer Causes Control.* **28** (1), 77–88 (2017).
44. Johnson, K. E. et al. The relationship between circulating lipids and breast cancer risk: a mendelian randomization study. *PLoS Med.* **17** (9), e1003302 (2020).
45. Zhou, M. et al. Mendelian randomization analysis revealed potential metabolic causal factors for breast cancer. *Sci. Rep.* **13** (1), 14290 (2023).
46. Toth, P. P. et al. High-density lipoproteins: a consensus statement from the national lipid association. *J. Clin. Lipidol.* **7** (5), 484–525 (2013).
47. Voight, B. F. et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* **380** (9841), 572–580 (2012).
48. Holmes, M. V. et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur. Heart J.* **36** (9), 539–550 (2015).
49. Boekholdt, S. M. et al. Levels and changes of HDL cholesterol and apolipoprotein A-I in relation to risk of cardiovascular events among statin-treated patients: a meta-analysis. *Circulation* **128** (14), 1504–1512 (2013).
50. Kuhnast, S., Fiocco, M., van der Hoorn, J. W., Princen, H. M. & Jukema, J. W. Innovative pharmaceutical interventions in cardiovascular disease: focusing on the contribution of non-HDL-C/LDL-C-lowering versus HDL-C-raising: a systematic review and meta-analysis of relevant preclinical studies and clinical trials. *Eur. J. Pharmacol.* **763** (Pt A), 48–63 (2015).
51. Pan, B. et al. HDL of patients with type 2 diabetes mellitus elevates the capability of promoting breast cancer metastasis. *Clin. Cancer Res.* **18** (5), 1246–1256 (2012).
52. Rotheneder, M. & Kostner, G. M. Effects of low- and high-density lipoproteins on the proliferation of human breast cancer cells in vitro: differences between hormone-dependent and hormone-independent cell lines. *Int. J. Cancer*. **43** (5), 875–879 (1989).
53. Kaaks, R. et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr. Relat. Cancer*. **12** (4), 1071–1082 (2005).
54. Key, T. et al. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J. Natl. Cancer Inst.* **94** (8), 606–616 (2002).
55. Zeleniuch-Jacquotte, A. et al. Postmenopausal levels of oestrogen, androgen, and SHBG and breast cancer: long-term results of a prospective study. *Br. J. Cancer*. **90** (1), 153–159 (2004).
56. Tin Tin, S., Reeves, G. K. & Key, T. J. Endogenous hormones and risk of invasive breast cancer in pre- and post-menopausal women: findings from the UK Biobank. *Br. J. Cancer*. **125** (1), 126–134 (2021).
57. Chen, F. et al. Mendelian randomization analyses of 23 known and suspected risk factors and biomarkers for breast cancer overall and by molecular subtypes. *Int. J. Cancer*. **151** (3), 372–380 (2022).
58. He, X. Y., Liao, Y. D., Yu, S., Zhang, Y. & Wang, R. Sex hormone binding globulin and risk of breast cancer in postmenopausal women: a meta-analysis of prospective studies. *Horm. Metab. Res.* **47** (7), 485–490 (2015).
59. Dimou, N. L. et al. Sex hormone binding globulin and risk of breast cancer: a mendelian randomization study. *Int. J. Epidemiol.* **48** (3), 807–816 (2019).
60. Fortunati, N. & Catalano, M. G. Sex hormone-binding globulin (SHBG) and estradiol cross-talk in breast cancer cells. *Horm. Metab. Res.* **38** (4), 236–240 (2006).

61. Fortunati, N., Catalano, M. G., Boccuzzi, G. & Frairia, R. Sex hormone-binding globulin (SHBG), estradiol and breast cancer. *Mol. Cell. Endocrinol.* **316** (1), 86–92 (2010).
62. Arthur, R. S., Xue, X. & Rohan, T. E. Prediagnostic circulating levels of sex steroid hormones and SHBG in relation to risk of ductal carcinoma in situ of the breast among UK women. *Cancer Epidemiol. Biomarkers Prev.* **29** (5), 1058–1066 (2020).
63. Drummond, A. E. et al. Linking physical activity to breast cancer via sex steroid hormones, part 2: the effect of sex steroid hormones on breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **31** (1), 28–37 (2022).
64. Sieri, S. et al. Prospective study on the role of glucose metabolism in breast cancer occurrence. *Int. J. Cancer*. **130** (4), 921–929 (2012).
65. Boyle, P. et al. Blood glucose concentrations and breast cancer risk in women without diabetes: a meta-analysis. *Eur. J. Nutr.* **52** (5), 1533–1540 (2013).
66. Jung, S. Y., Mancuso, N., Han, S. & Zhang, Z. F. The role of genetically determined glycemic traits in breast Cancer: a mendelian randomization study. *Front. Genet.* **11**, 540724 (2020).
67. Lindenfeld, J. & Schrier, R. W. Blood urea nitrogen a marker for adverse effects of loop diuretics? *J. Am. Coll. Cardiol.* **58** (4), 383–385 (2011).
68. Du, J., Niu, J., Ma, L., Sui, Y. & Wang, S. Association between blood urea nitrogen levels and length of stay in patients with pneumonic chronic obstructive pulmonary disease exacerbation: a secondary analysis based on a multicentre, retrospective cohort study. *Int. J. Chron. Obstruct Pulmon Dis.* **17**, 2847–2856 (2022).
69. Wang, C., Sun, H. & Liu, J. BUN level is associated with cancer prevalence. *Eur. J. Med. Res.* **28** (1), 213 (2023).
70. Almquist, M., Bondeson, A. G., Bondeson, L., Malm, J. & Manjer, J. Serum levels of vitamin D, PTH and calcium and breast cancer risk-a prospective nested case-control study. *Int. J. Cancer*. **127** (9), 2159–2168 (2010).
71. Almquist, M., Manjer, J., Bondeson, L. & Bondeson, A. G. Serum calcium and breast cancer risk: results from a prospective cohort study of 7,847 women. *Cancer Causes Control*. **18** (6), 595–602 (2007).
72. Almquist, M. et al. Serum calcium and tumour aggressiveness in breast cancer: a prospective study of 7847 women. *Eur. J. Cancer Prev.* **18** (5), 354–360 (2009).
73. McEligot, A. J., Largent, J., Ziogas, A., Peel, D. & Anton-Culver, H. Dietary fat, fiber, vegetable, and micronutrients are associated with overall survival in postmenopausal women diagnosed with breast cancer. *Nutr. Cancer*. **55** (2), 132–140 (2006).
74. Holmes, M. D. et al. Dietary factors and the survival of women with breast carcinoma. *Cancer* **86** (5), 826–835 (1999).
75. Saquib, J. et al. Dietary intake, supplement use, and survival among women diagnosed with early-stage breast cancer. *Nutr. Cancer*. **63** (3), 327–333 (2011).
76. Beasley, J. M. et al. Post-diagnosis dietary factors and survival after invasive breast cancer. *Breast Cancer Res. Treat.* **128** (1), 229–236 (2011).
77. Cauley, J. A. et al. Calcium plus vitamin D supplementation and health outcomes five years after active intervention ended: the women's Health Initiative. *J. Womens Health (Larchmt)*. **22** (11), 915–929 (2013).
78. Bristow, S. M. et al. Calcium supplements and cancer risk: a meta-analysis of randomised controlled trials. *Br. J. Nutr.* **110** (8), 1384–1393 (2013).
79. Huss, L. et al. Serum levels of vitamin D, parathyroid hormone and calcium in relation to survival following breast cancer. *Cancer Causes Control*. **25** (9), 1131–1140 (2014).

Author contributions

JF, XQ and CC was contributed to reviewing the literature and writing the original manuscript. JF, KY, XX, XL, RC, and MW were contributed to revising and editing the manuscript. JF, BL, TG and JL were contributed to revising and reviewing the manuscript. All authors have read and approved the final submitted manuscript.

Funding

This work was supported by the Luzhou City Science and Technology Plan Project Program, China (2024JYJ030, 2024JYJ014); the 2024 Open Research Project of the Sichuan Provincial Engineering Technology Research Center of Molecular Diagnosis of Clinical Diseases (24GCZXZD03) and the 2023 Open Fund Project of the Luzhou Key Laboratory of Molecular Diagnosis of Clinical Diseases (FZZD2023-01); the Sichuan Provincial Department of Science and Technology (2022YFS0152, 22ZYZF0007); the Jiucheng Talent·Science and Technology Innovation Team (Luzutong [2021] No. 162); the Student innovation and entrepreneurship project of Southwest Medical University (2023 and 2024).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-90447-9.

Correspondence and requests for materials should be addressed to J.F., X.Q., T.G. or J.L.