scientific reports

Check for updates

OPEN

Exploring the genetic associations and causal relationships between antibody responses, immune cells, and various types of breast cancer

Yang Yang[1,3], Jiayi Chen[1,3], Fuhong Gong[1,3], Jingge Miao[1], Mengping Lin[1], Ruimin Liu[1], Chenxi Wang[1], Fei Ge[2]✉ & Wenlin Chen[1]✉

Background: There may be potential associations between various pathogens, antibody immune responses, and breast cancer (BC), but the specific mechanisms and causal relationships remain unclear. Methods: First, multiple Mendelian randomization (MR) methods were used for univariable MR analysis to explore potential causal relationships between 34 antibody immune responses (related to 12 pathogens), 46 antibody immune responses (related to 13 pathogens), antibody responses post-COVID-19 vaccination, 731 immune cell types, and various BC subtypes (including overall BC, ER-positive, ER-negative, Luminal A, Luminal B, Luminal B HER2-negative, HER2-positive, and triple-negative BC). The primary results were then subjected to reverse MR analysis, heterogeneity testing using Cochran's Q, and horizontal pleiotropy testing. Robust findings were further used to design mediation pathways involving antibody immune responses, immune cells, and BC. After adjusting the effect estimates using multivariable MR (MVMR), a two-step mediation analysis was conducted to explore mediation pathways and mediation proportions. Finally, linkage disequilibrium score regression (LDSC) was applied to analyze the genetic correlation between phenotypes along mediation pathways, and cross-phenotype association analysis (CPASSOC) was performed to identify pleiotropic SNPs among three phenotypes along these pathways. Bayesian colocalization tests were conducted on pleiotropic SNPs using the multiple-trait-coloc (moloc). Results: We identified potential causal relationships between 15 antibody immune responses to 8 pathogens (Hepatitis B virus, Herpes Simplex Virus 2, Human Herpesvirus 6, Polyomavirus 2, BK polyomavirus, Cytomegalovirus, Helicobacter pylori, Chlamydia trachomatis), 250 immune cell phenotypes, and various BC subtypes. MVMR-adjusted mediation analysis revealed four potential mediation pathways. LDSC results showed no significant genetic correlation between phenotypes pairwise. CPASSOC analysis identified two potential mediation pathways with common pleiotropic SNPs (rs12121677, rs281378, rs2894250). However, none of these SNPs passed the Bayesian colocalization test by moloc. These results excluded horizontal pleiotropy, stabilizing MR analysis results. Conclusion: This study utilized MR methods to analyze potential causal relationships between various antibody immune responses, immune cell types, and BC subtypes, identifying four potential regulatory mediation pathways. The findings of this study offer potential targets and research directions for virus-related and immunotherapy-related studies, providing a certain level of theoretical support. However, limitations such as GWAS sample size constraints and unclear specific pathophysiological mechanisms need further improvement and validation in future studies.

Keywords Breast cancer, Pathogen, Virus, Antibody immune response, Causal inference, Mediation analysis, Genetic correlation

Breast cancer (BC) is one of the most common malignant tumors among women, and its incidence is gradually increasing[1]. Many studies have observed potential associations between various pathogen infections and

[1]Third Department of Breast Surgery, Peking University Cancer Hospital Yunnan Hospital, The Third Affiliated Hospital of Kunming Medical University & Yunnan Cancer Hospital, Kunming 650118, China. [2]Department of Breast Surgery, First Affiliated Hospital of Kunming Medical University, Kunming 650032, China. [3]Yang Yang, Jiayi Chen and Fuhong Gong contributed equally to this work. ✉email: ajqnadjd@hotmail.com; chenwenlin1@hotmail.com

BC, but the underlying mechanisms and causal relationships remain largely unexplored and require further clarification[2–5].

Various pathogens, including viruses, bacteria, and parasites, have been reported to have carcinogenic effects[6]. Additionally, some microorganisms can exert anti-cancer effects by inducing immune responses[7]. For instance, Epstein Barr virus (EBV) has been reported to be associated with the risk of various cancers, including lymphomas, nasopharyngeal carcinoma, and gastric cancer[3]. The infection rate of EBV in BC cases is significantly higher compared to non-BC controls[3]. Moreover, other viruses, such as human cytomegalovirus (HCMV), mouse mammary tumor virus (MMTV), and human papillomavirus (HPV), have been reported to be associated with an increased risk of BC[2,3,8]. In some cancer patients, spontaneous regression of tumors has been observed following viral infection[9]. This suggests that certain viruses possess anti-cancer activity. Consequently, oncolytic virus (OV) therapy has become a focal point in tumor immunotherapy. Some studies have reported that specific antibodies can recognize tumor-associated antigens and activate immune cells to inhibit tumor growth[10]. Conversely, certain antibodies can promote tumor development and immune escape through immunosuppressive pathways[10]. Although these phenomena have been observed in related research, the exact role of these pathogen immune responses in BC development, and their interaction with the immune system, requires further investigation.

BC can be classified into subtypes based on molecular phenotypes: Luminal A, Luminal B, HER2-enriched-like, and Triple-negative (TNBC)[11]. The biological behaviors exhibited by different subtypes also vary[11]. Therefore, the relationship between immune responses to different pathogens and different BC subtypes may not be uniform. Exploring the relationship between antibody immune responses and BC subtypes may provide valuable insights into the immune mechanisms of BC and further identify potential targets for immunotherapy.

Mendelian randomization (MR) provides a reliable method for inferring causal relationships. MR uses genetic variations as instrumental variables (IVs) to estimate the effects of exposures (e.g., immune responses and immune cells) on outcomes (e.g., BC)[12]. This method can reduce confounding factors such as social and economic environments, effectively minimizing bias in causal estimates and strengthening causal inference[12]. This study used bidirectional MR analysis to explore the potential causal relationships between various antibody immune responses, immune cells, and various BC subtypes. Multivariable MR (MVMR) and mediation analysis were conducted to investigate the mediating role of immune cells in the relationship between antibody immune responses and BC. Additionally, global genetic correlation analysis and SNP-level analysis were performed to explore genetic correlations between pathogen antibody immune responses, immune cells, and BC. This study aims to provide valuable insights into the immune mechanisms of BC and identify new potential targets for immunotherapeutic interventions.

Methods

Study design

We obtained data on 34 antibody immune responses, 46 antibody immune responses, antibody responses post-COVID-19 vaccination, and 731 immune cells from four studies in the GWAS Catalog[13]. Data on BC and its subtypes were obtained from the Breast Cancer Association Consortium (BCAC). First, two-sample bidirectional MR analysis was used to investigate potential causal relationships between antibody immunity, immune cells, and BC. Then, multivariable MR (MVMR) and mediation analyses were conducted to explore the mediating role and proportion of antibody immune responses between immune cells and BC. Next, based on the three phenotypes identified in the potential mediation pathways, linkage disequilibrium (LD) score regression (LDSC) was used to estimate the heritability of each phenotype and the genetic correlations between pairs of phenotypes. Cross-phenotype association analysis (CPASSOC) was then performed to identify key pleiotropic SNPs that may influence all three phenotypes simultaneously. Finally, Bayesian colocalization analysis using the multiple-trait colocalization method (moloc) was conducted to determine whether the identified pleiotropic SNPs could affect all three phenotypes concurrently. The study design followed the STROBE-MR guidelines, with the workflow detailed in Fig. 1.

Data sources

GWAS data on 34 antibody immune responses

The data on 34 antibody immune responses were obtained from the French Milieu Intérieur cohort[14]. This cohort included 1,000 healthy individuals divided into five age groups (20–29, 30–39, 40–49, 50–59, and 60–70 years), with 200 participants in each age group, balanced by sex. Sample quality control was performed using methods such as sample integrity, genotype consistency, and contamination detection. The study included 12 pathogens, resulting in 34 phenotypes. These phenotypes included total IgA, IgE, IgG, and IgM levels, as well as antibody levels and serostatus for common pathogens (e.g., cytomegalovirus, Epstein-Barr virus, etc.; for details, see Supplementary Table 1).

GWAS data on 46 antibody immune responses

The data on 46 antibody immune responses were obtained from the UK Biobank, including 8,735 individuals[15]. The study excluded low-quality genotype data, adjusted for sex and age, and used principal component analysis to adjust for population stratification. A total of 13 pathogens were analyzed, resulting in 46 phenotypes. These phenotypes included 15 serostatus phenotypes and 31 quantitative antibody measurement phenotypes (e.g., herpes simplex virus, human herpesvirus 6 and 7, etc.; for details, see Supplementary Table 1).

GWAS data on antibody responses post-COVID-19 vaccination

The data on antibody responses post-COVID-19 vaccination were obtained from the UK Biobank, including data from the 200 K self-test antibody study and the 60 K Coronavirus infection study[16], totaling 54,066 individuals.

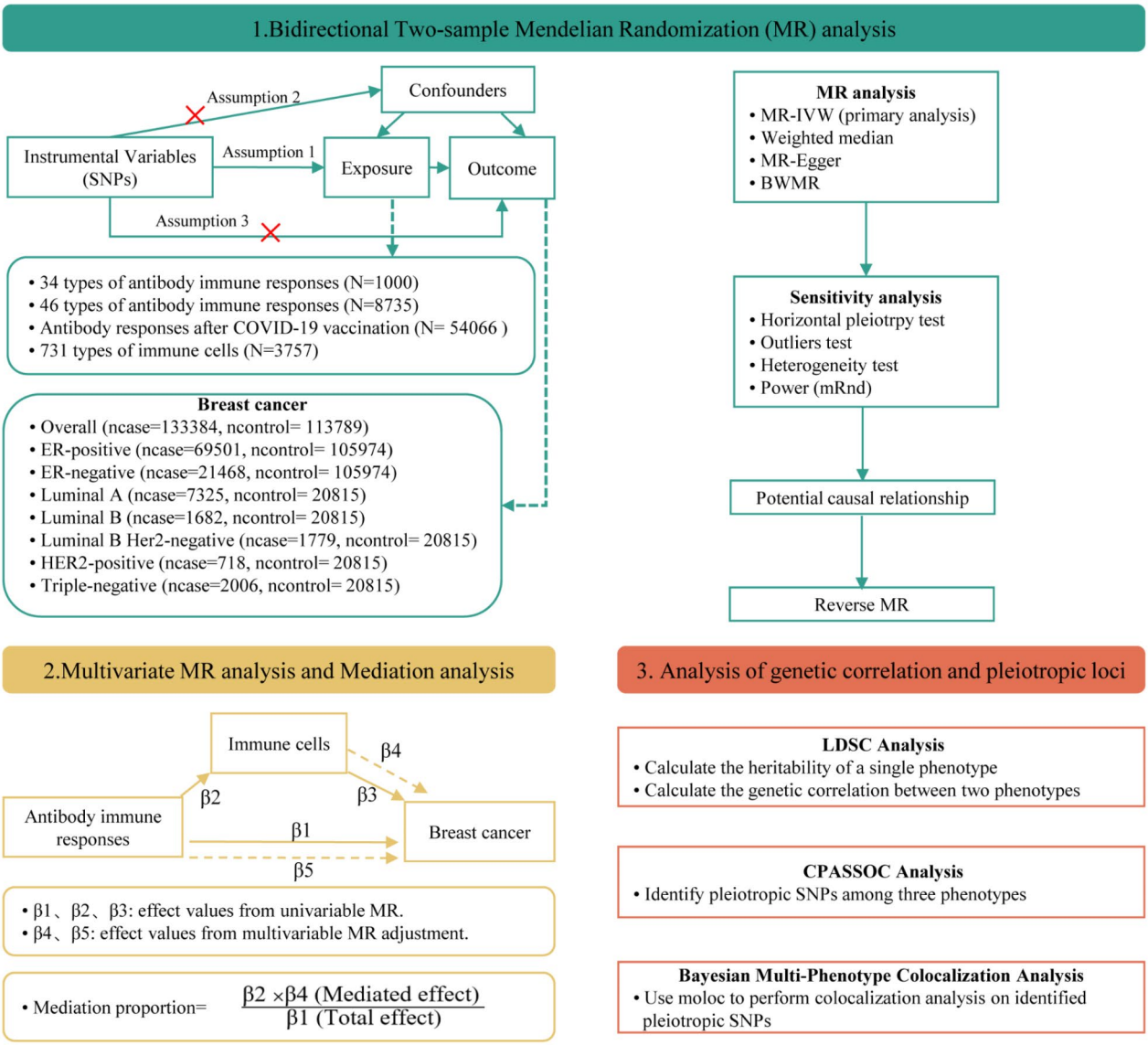

Fig. 1. Flowchart of this study.

The study excluded low-quality genotype data, adjusted for sex and age, and used principal component analysis to adjust for population stratification. The final analysis included two phenotypes: IgG serostatus after the first and second doses of the COVID-19 vaccine (detailed information in Supplementary Table 1).

GWAS data on 731 immune cells

Data on 731 immune cells were obtained from the SardiNIA cohort in Sardinia, Italy[17]. The cohort includes 3,757 individuals aged 18 to 102 years, with a roughly balanced gender ratio. The study analyzed 731 immune phenotypes after adjusting for covariates such as age and sex. These phenotypes cover multiple subpopulations of immune cells, including T cells, B cells, dendritic cells, and monocytes. The data include absolute cell counts ($n=118$), median fluorescence intensity indicating surface antigen levels ($n=389$), morphological parameters ($n=32$), and relative cell counts ($n=192$) (for details, see Supplementary Table 1).

GWAS data on breast cancer risk

The study by Zhang et al.[18] included data on overall BC (Overall), and five subtypes (Luminal A, Luminal B, Luminal B HER2-negative, HER2-positive, and Triple-negative). Additionally, ER-positive and ER-negative BC data were obtained from the study by Michailidou et al.[19]. The data were from European individuals, retaining SNPs with a minor allele frequency (MAF) greater than 0.01, and excluding missing or duplicate reference SNP IDs (rsid). All original GWAS studies obtained ethical approval and informed consent from the relevant institutions, as detailed in the original studies. Therefore, no additional ethical approval was required for this study. Moreover, there was no sample overlap between the exposure and outcome data used in this study.

SNP selection

SNPs as IVs must satisfy three core assumptions: (1) SNPs are associated with exposure factors[20]; (2) SNPs are not associated with confounding factors; (3) SNPs are not associated with outcomes (see Fig. 1).

To ensure an adequate number of IVs for each phenotype, we selected SNPs associated with antibody immune responses and immune cells at a threshold of $P < 1e\text{-}5$, based on previous studies[21,22]. During the selection process, we referred to the whole-genome information from the 1000 Genomes Project[23], setting LD parameters at $r2 < 0.001$ and selecting the SNP with the smallest P-value within a 10,000 kb range as the independent IV. To avoid potential bias from weak IVs, only IVs with an F-statistic > 10 were retained for subsequent analysis[24,25]. Palindromic variants were removed using the harmonized method. Phenotypes with fewer than three IVs were excluded from sensitivity analysis to ensure sufficient IV numbers[26].

Statistical analysis

MR analysis

Four MR methods were used to estimate the association between exposure and outcome: inverse-variance weighted (IVW), MR-Egger, weighted median (WM), and Bayesian weighted MR (BWMR). The IVW method assumes that all IVs are independent and lack horizontal pleiotropy[27]. The MR-Egger method allows for invalid IVs, providing more reasonable causal effect estimates in the presence of pleiotropic effects[28]. When some IVs (< 50%) exhibit directional horizontal pleiotropy, the WM method can provide robust estimates[29]. The BWMR method enhances causal inference robustness by considering the uncertainty of weak effects due to polygenicity[30]. The IVW method, due to its high statistical power, was selected as the primary analysis method[31]. MR-Egger, WM, and BWMR were used as supplementary analyses. Positive results were considered when the estimates from the four methods were consistent and the IVW method was significant ($P_{IVW} < 0.05$)[32].

Finally, the results of IVW were adjusted for the false discovery rate (FDR)[33]. Significant associations were considered when FDR < 0.1; suggestive associations were considered when FDR > 0.1 and $P_{IVW} < 0.05$.

Reverse MR analysis

To further explore whether BC has a reverse causal effect on identified gut microbiota, reverse MR analysis was conducted when $P_{IVW} < 0.05$ in forward MR analysis. In reverse MR analysis, BC was considered an exposure, and its associated SNPs as IVs, with gut microbiota and blood metabolites as outcomes. The analysis process was similar to forward MR analysis.

Sensitivity analysis

To ensure the robustness of the results, Cochran's Q statistic was first used to test for heterogeneity[34]. When heterogeneity was significant ($P_{Heterogeneity} < 0.05$), a random-effects IVW method was adopted; when heterogeneity was not significant ($P_{Heterogeneity} > 0.05$), a fixed-effects IVW method was used. MR Egger regression intercept and MR PRESSO global test were then used to assess horizontal pleiotropy[35,36]. When pleiotropy existed ($P_{Pleiotropy} < 0.05$), the result was excluded. Finally, the statistical power was estimated using the mRnd method[37].

MVMR analysis and mediation analysis

Considering the potential association between antibody immune responses and immune cells, multivariable analysis and mediation analysis were conducted to identify potential regulatory pathways. Univariable MR methods were first used to assess the direct causal relationships between antibody immune responses, immune cells, and BC, and obtain corresponding effect values β, where β1 is the effect value of antibody immune responses on BC, β2 is the effect value of antibody immune responses on immune cells, and β3 is the effect value of immune cells on BC. In previous sections of the study, antibody responses and immune cell types were determined based on the condition of $P_{IVW} < 0.05$. Subsequently, additional univariable MR analyses were performed with antibody immune responses as exposures and immune cell types as outcomes. When all three results met the condition of $P_{IVW} < 0.05$, MVMR was conducted using the IVW method, with immune cells and antibody immune responses as exposures, to obtain the effect value β4 of immune cells on BC. After MVMR adjustment, the mediation effect was calculated using the coefficient product method (mediation effect = $\beta2 \times \beta4$). The direct effect was calculated by subtracting the mediation effect from the total effect, and the mediation proportion was quantified [mediation proportion = (mediation effect/total effect) × 100%]. The Delta method and mediation effect were used to estimate the standard error of the mediation proportion. Potential mediation pathways were identified at a 90% CI standard.

Genetic correlation analysis

LDSC was used to estimate heritability and analyze genetic correlation between phenotypes on mediation pathways[38]. Pre-calculated LD scores from the 1000 Genomes Project were used to estimate SNP heritability in HapMap 3 SNPs. Pairwise genetic correlations (rg) between different phenotypes were further estimated, with rg ranging from −1 to 1, representing negative and positive correlations, respectively.

CPASSOC analysis

Genetic correlation analysis cannot detect SNP effects on phenotypes. CPASSOC analysis was performed on exposure, mediation, and outcome phenotypes to explore whether there were pleiotropic SNPs simultaneously affecting these phenotypes in potential mediation pathways. CPASSOC detected common genetic loci through joint analysis of multiple phenotype GWAS data[39]. CPASSOC provides SHom and SHet methods; SHom is similar to a fixed-effect meta-analysis method for consistent effects, while SHet al.lows for effect heterogeneity between phenotypes, with robust statistical power in the presence of heterogeneity. In CPASSOC, $P_{SHet} < 5e\text{-}8$

was considered a SNP significantly associated with two phenotypes. The most relevant independent SNP in a 10,000 kb region for two traits was obtained using PLINK, referencing 1000 Genomes Phase 3 (EUR) data (–clump-p1 5e-8, –clump-p2 1e-5, –clump-r2 0.001, –clump-kb 10000)[40,41]. SNPs with $P_{SHet} < 5e\text{-}8$ and $P_{Single-trait} < 0.05$ in each phenotype were considered pleiotropic SNPs.

Bayesian multiple-trait colocalization analysis

Multiple-trait-coloc (moloc) was used for colocalization analysis of three phenotypes within a 1000 kb region for identified pleiotropic SNPs. Moloc, based on a Bayesian statistical framework, quantifies evidence of shared causal variants in a specific region from GWAS summary data[42]. When colocalizing three traits, possible hypotheses include: H1: phenotypes 1 and 2 share a causal variant, phenotype 3 has an independent causal variant; H2: phenotypes 1 and 3 share a causal variant, phenotype 2 has an independent causal variant; H3: phenotypes 2 and 3 share a causal variant, phenotype 1 has an independent causal variant; H4: phenotypes 1, 2, and 3 each have independent causal variants; H5: phenotypes 1, 2, and 3 share a causal variant. Each hypothesis represents a different shared causal variant scenario. Moloc calculates posterior probability (PPA) for each hypothesis within the Bayesian framework, quantifying evidence of multiple phenotypes sharing a causal variant in a specific region. This study focused on the H5 hypothesis, considered valid when PPA > 0.8.

All statistical analyses and visualizations were performed using R (4.3.1) and R packages such as "TwoSampleMR", "MR-PRESSO", "MendelianRandomization", "mRnd", "moloc", "ggplot2".

Results

Causal analysis of 34 antibody immune responses and BC

After excluding phenotypes without sufficient IVs, univariable MR analysis identified five potential causal relationships between 34 antibody immune responses and BC under $P_{IVW}<0.05$ ($P_{IVW}<0.05$, FDR > 0.1, Supplementary Table 2). The F-statistics range for IVs in MR analysis was 19.7-4659.8. After excluding one non-robust result based on inconsistent effect direction or horizontal pleiotropy ($P<0.05$), potential causal associations between four antibody immune responses to two pathogens [Hepatitis B virus (HBV), Helicobacter pylori (HP)] and total IgA levels with BC and its subtypes were identified (Fig. 2). For example, BC (ER-negative) had potential causal relationships with Anti-hepatitis B virus surface antigen (HBs) IgG seropositivity [OR (95% CI): 0.961 (0.933–0.990), $P_{IVW}=0.009$] and IgA levels [OR (95% CI): 1.196. (1.023–1.398), $P_{IVW}=0.025$].

Reverse MR analysis showed a relationship between BC (Luminal B) and Anti-helicobacter pylori IgG levels [OR (95% CI): 0.891 (0.802–0.990), $P_{IVW}=0.032$] (Supplementary Table 3).

Causal analysis of 46 antibody immune responses and BC

After excluding phenotypes without sufficient IVs, univariable MR analysis identified 17 potential causal relationships between 46 antibody immune responses and BC under $P_{IVW}<0.05$ ($P_{IVW}<0.05$, FDR > 0.1, Supplementary Table 4). The F-statistics range for IVs in MR analysis was 19.5-343.2. After excluding six non-robust results based on inconsistent effect direction or horizontal pleiotropy ($P<0.05$), potential causal

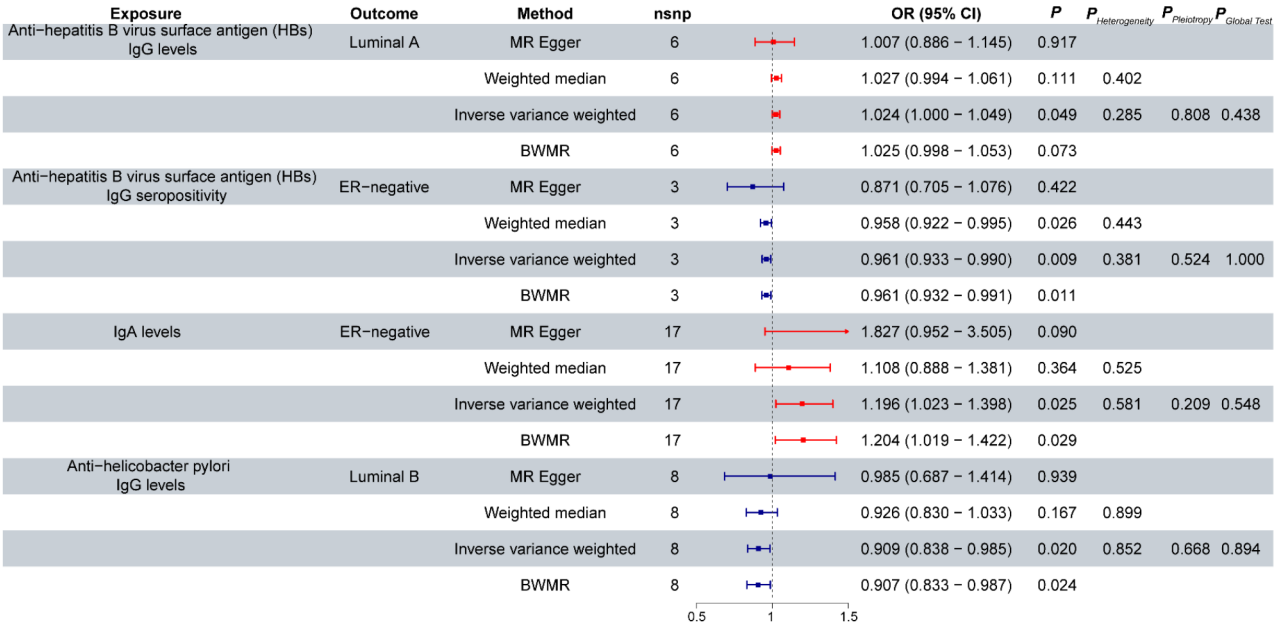

Fig. 2. Results of MR analysis showing potential causal relationships between 34 antibody immune responses and various breast cancer subtypes ($P_{IVW}<0.05$, with red indicating risk factors and blue indicating protective factors in the forest plot).

associations between 11 antibody immune responses to seven pathogens [Herpes Simplex Virus (HSV) 2, Human Herpesvirus 6 (HHV) IE1B, Chlamydia trachomatis, Polyomavirus 2, BK polyomavirus, Helicobacter pylori, Cytomegalovirus (CMV)] and BC and its subtypes were identified (Fig. 3). Among these, three antibody immune responses were associated with more than one BC subtype (Fig. 3). For example, Anti-human herpesvirus 6 IE1B IgG seropositivity was potentially causally related to BC (Overall) [OR (95% CI): 1.021 (1.003–1.039), P_{IVW}=0.023] and BC (Luminal A) [OR (95% CI): 1.044 (1.018–1.070), P_{IVW}<0.001].

Reverse MR analysis showed a relationship between BC (Overall) and Chlamydia trachomatis momp A antibody levels [OR (95% CI): 1.137 (1.004–1.287), P_{IVW}=0.042] (Supplementary Table 5).

Causal analysis of antibody responses post-COVID-19 vaccination and BC

Univariable MR analysis did not identify any potential causal relationships between antibody responses post-COVID-19 vaccination and different BC subtypes under P_{IVW}<0.05 (Supplementary Table 6).

Causal analysis of 731 immune cells and BC

After excluding phenotypes without sufficient IVs, univariable MR analysis identified 321 potential causal relationships between 731 immune cells and BC under P_{IVW}<0.05 (P_{IVW}<0.05, FDR > 0.1, Fig. 4, Supplementary Table 7). The F-statistics range for IVs in MR analysis was 100.0-994.0. After excluding 71 non-robust results based on inconsistent effect direction or horizontal pleiotropy (P < 0.05), potential causal associations between 250 immune cell phenotypes and BC and its subtypes were identified. Among these, 48 immune cell phenotypes were associated with more than one BC subtype (Supplementary Table 8). For example, CD3 on CD28 + CD4-CD8- T cells had potential causal relationships with BC (Overall) [OR (95% CI): 0.953 (0.927–0.98), P < 0.001], BC (ER-positive) [OR (95% CI): 0.935 (0.904–0.967), P_{IVW}<0.001], and BC (Luminal A) [OR (95% CI): 0.942 (0.909–0.976), P_{IVW}=0.001].

Reverse MR analysis under P_{IVW}<0.05 identified bidirectional causal relationships between seven immune cells and BC. For example, BC (Luminal B Her2-negative) was related to CD62L- monocyte [OR (95% CI): 1.112 (1.029–1.203), P_{IVW}=0.007] (Supplementary Table 9).

Power calculations for all univariable MR analyses are provided in Supplementary Table 10.

MVMR analysis and mediation analysis

Mediation analysis was conducted to further explore potential pathways between antibody immune responses, immune cells, and BC. Antibody immune responses were considered exposures, immune cells as mediators, and BC and its subtypes as outcomes. Under P_{IVW}<0.05, 20 potential mediation pathways were identified, with BC [Overall (n = 5), ER-negative (n = 3), Luminal A (n = 2), Luminal B (n = 5), HER2-positive (n = 2), and TNBC (n = 3) (Supplementary Table 11)]. After adjusting effect values using MVMR-IVW, one potential mediation pathway related to BC was identified under P < 0.05. To identify more potential pathways, the P-value threshold was relaxed to 0.1, resulting in four potential mediation pathways related to BC (Fig. 5).

Genetic correlation analysis

LDSC was used to estimate heritability and analyze genetic correlation between phenotypes on the four identified potential mediation pathways. After excluding results with negative heritability due to sample size limitations, Anti-hepatitis B virus surface antigen (HBs) IgG seropositivity had the highest heritability estimate, H2 = 0.41

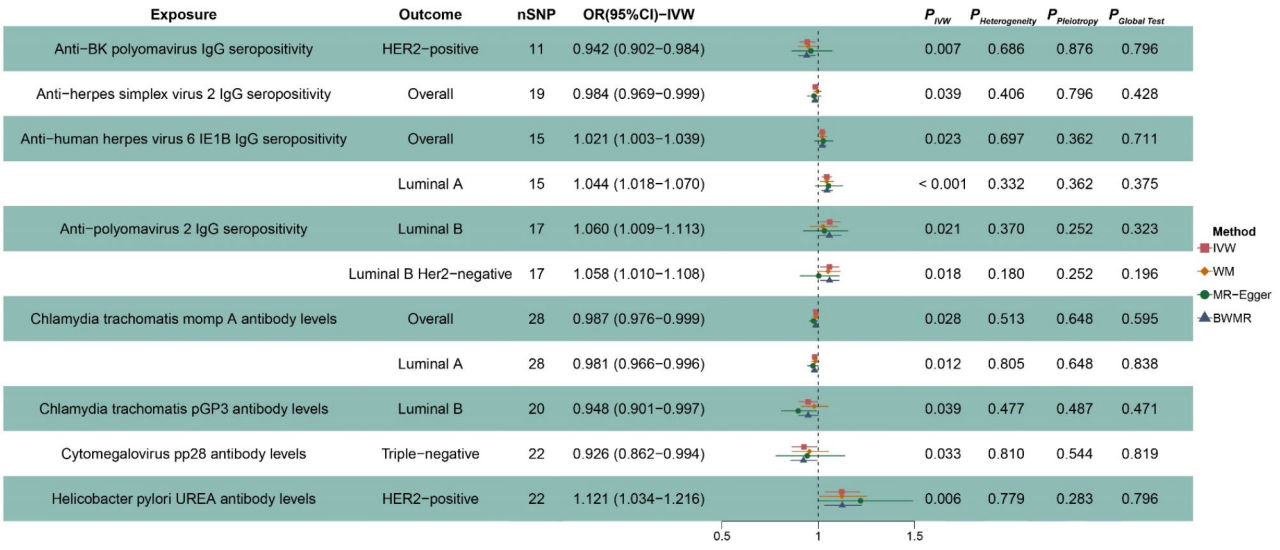

Fig. 3. Results of MR analysis showing potential causal relationships between 46 antibody immune responses and various breast cancer subtypes (P_{IVW} <0.05; the exposure is the same for rows 3 and 4, rows 5 and 6, and rows 7 and 8).

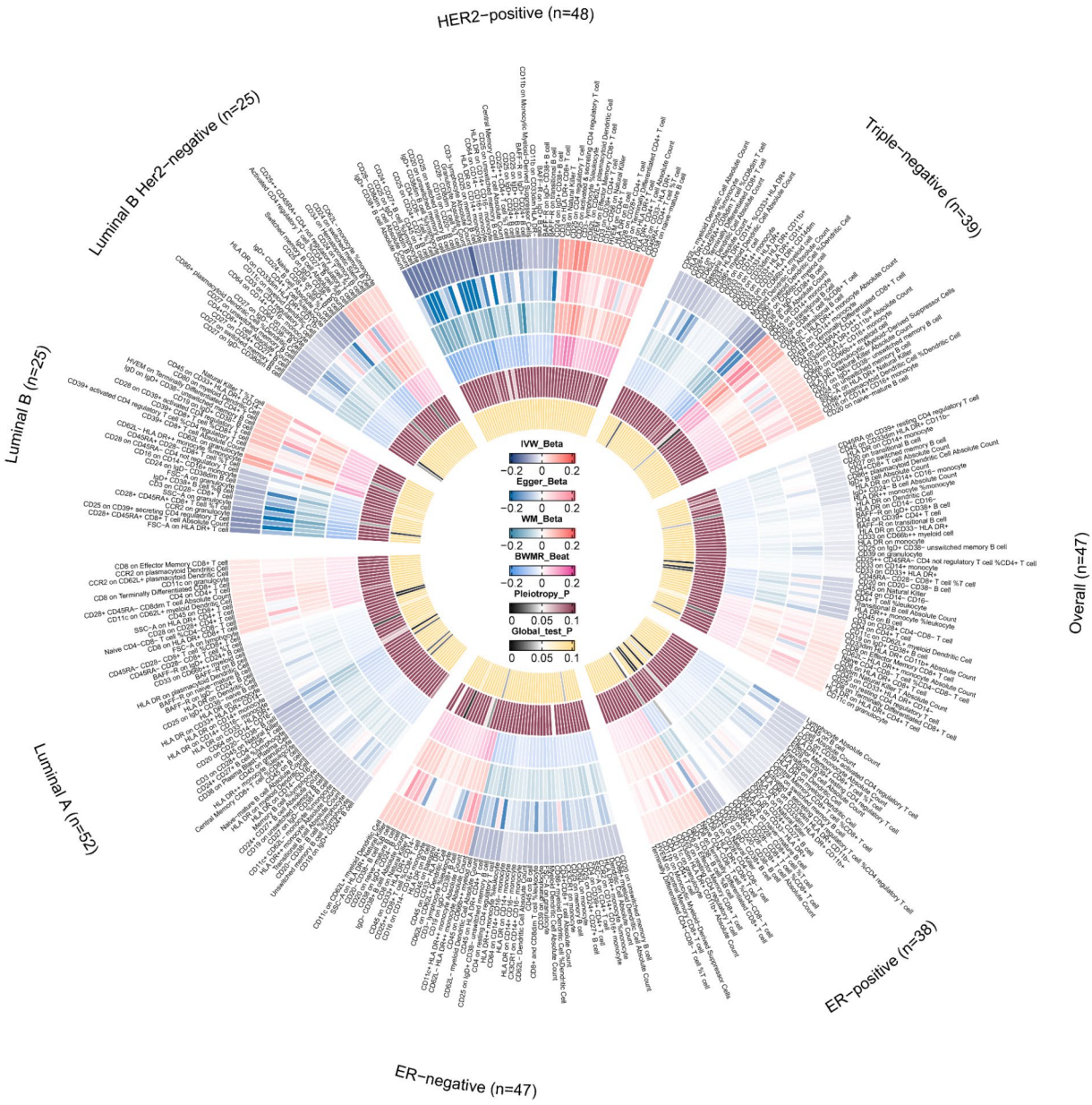

Fig. 4. Results of MR analysis showing potential causal relationships between 731 immune cells and various breast cancer subtypes (P_{IVW} <0.05). The circular heatmap was generated using the "ComplexHeatmap" R package.

(Supplementary Table 12). The heritability estimates for Luminal B, ER-negative, and HER2-positive were 0.23, 0.07, and 0.09, respectively. Genetic correlation analysis showed no significant genetic correlation between phenotypes under $P < 0.05$, considering sample size limitations (Supplementary Table 13).

CPASSOC and colocalization analysis

CPASSOC analysis identified three potential pleiotropic SNPs (Fig. 6, Supplementary Table 14). Two potential pleiotropic SNPs (rs281378, rs2894250) were identified on the mediation pathway from Anti-polyomavirus 2 IgG seropositivity to CD80 on myeloid dendritic cells to BC (Luminal B, Fiugure6A, Fiugure6B). One potential pleiotropic SNP (rs12121677) were identified on the mediation pathway from Anti-hepatitis B virus surface antigen (HBs) IgG seropositivity to CD25 on CD24 + CD27 + B cells to BC (ER-negative) (Fig. 6C).

However, none of these pleiotropic SNPs passed the Bayesian colocalization test (PPA < 0.75), indicating the absence of pleiotropic SNPs within 1000 kb affecting the three phenotypes.

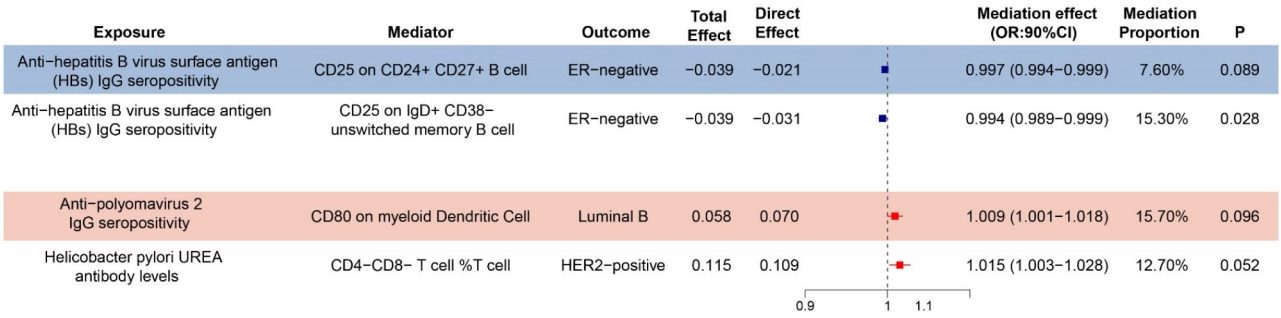

Fig. 5. Analysis of four potential mediatory pathways from antibody immune responses to immune cell types to breast cancer. The exposure data in rows 1 and 2 are from 34 antibody immune responses, while the data in rows 3 and 4 are from 46 antibody immune responses.

Discussion

This study used the largest and most comprehensive BC GWAS summary data currently available to explore potential causal relationships between various pathogen antibody immune responses, immune cell types, and BC subtypes through MR analysis, mediation analysis, and genetic correlation analysis. A total of 15 antibody immune responses to 8 pathogens (6 viruses, 1 bacterium, 1 Chlamydia) and 250 immune cell phenotypes were identified as potentially causally related to various BC subtypes. Reverse MR analysis revealed bidirectional causal relationships between BC and three antibody immune responses and seven immune cells. After adjusting effect values through MVMR, four mediation pathways related to BC were identified. These findings provide new insights and directions for clinical and experimental research on the immune system and BC, offering a theoretical basis for the regulatory pathways of pathogen antibody response-immune cell-BC axis.

In this study, when pathogen antibody immune responses were considered exposures, potential causal associations between antibody responses to six viruses and BC were identified. The results showed that HBV was related to increased risk of BC (Luminal A) and decreased risk of BC (ER-negative), though caution is needed as the result for BC (Luminal A) was at the margin of significance. Luminal A is an ER-positive, progesterone receptor-positive BC. Studies have reported that long-term HBV infection may indirectly affect BC occurrence by elevating estrogen levels in the body [43]. This might be how HBV influences BC. Furthermore, the study found that HSV2 antibody immune response might reduce the risk of BC (Overall). Related studies have shown that oHSV2 constructed with HSV2 exhibits strong anti-tumor activity and stable biological properties[44,45]. The oncolytic virus FusOn-H2, derived from HSV-2, induces a robust T-cell response against both primary and metastatic breast tumors in a murine BC model[46]. This suggests that oHSV2 may be an effective treatment for BC. Whether HHV-6 has direct carcinogenic ability is unclear[47]. It may indirectly promote tumor growth by collaborating with other viruses or as an opportunistic virus in an immunodeficient environment[47]. Additionally, it may contribute to tumor progression by suppressing the host's immune system[48]. The study results showed that Anti HHV-6 IE1B IgG seropositivity was positively related to increased BC (Overall) risk. Studies have reported an association between Polyomaviruses infection and BC risk[49]. The JC polyomavirus has been detected in some breast cancers[50]. It can integrate into the genomic DNA of eukaryotic cells and target signaling pathways such as p53, β-catenin, IRS, Rb, TGF-β1, PI3K/Akt, and AMPK, ultimately leading to the development of breast cancer[51].The study results showed that Anti-polyomavirus 2 IgG seropositivity was positively related to increased risk of BC (Luminal B, Luminal B HER2-negative), and Anti-BK polyomavirus IgG seropositivity was negatively related to decreased risk of BC (HER2-positive). Multiple studies have reported associations between CMV infection and BC development[52,53], possibly through regulating inflammation markers and activating immunity[53]. pp28 is a protein encoded by the human cytomegalovirus (HCMV) UL99 gene, important in the virus's assembly, transport, and maturation process[54]. Previous studies have reported differences in immune levels among individuals against the same pathogen[14]. Higher pp28 antibody levels represent stronger immune responses and immunity against the pathogen. The study results showed a negative association between CMV pp28 antibody levels and TNBC.

Studies have found significant associations between past Chlamydia infections and BC mortality[55]. Furthermore, studies have reported a potential association between Chlamydia trachomatis infection and an increased risk of breast cancer, particularly in women with elevated levels of IL-12[56]. In this study, Chlamydia trachomatis pGP3 antibody levels were negatively related to Luminal B risk, and Chlamydia trachomatis momp A antibody levels were negatively related to BC (Overall, Luminal A) risk. This result suggests that a strong immune response to Chlamydia trachomatis may reduce the risk of breast cancer in the context of infection. However, reverse MR analysis indicated a positive correlation between BC (Overall) and Chlamydia trachomatis MOMP A antibody, implying an increased risk of multiple Chlamydia trachomatis infections in BC patients. Therefore, forward and reverse MR results partially explain the association between BC and Chlamydia trachomatis. For HP antibody immune responses, the results showed that HP was related to decreased risk of BC (Luminal B) and increased risk of BC (HER2-positive). UREA antibody levels indicate chronic Helicobacter pylori infection[57]. Studies have reported that CagA-positive HP might reduce estradiol levels leading to osteoporosis[58]. Estradiol is a risk factor for ER-positive BC[59]. Multiple studies have also reported that HP infection can enhance HER-2 protein expression[60,61], suggesting HP may influence the risk of HER-2 positive BC through unknown pathways.

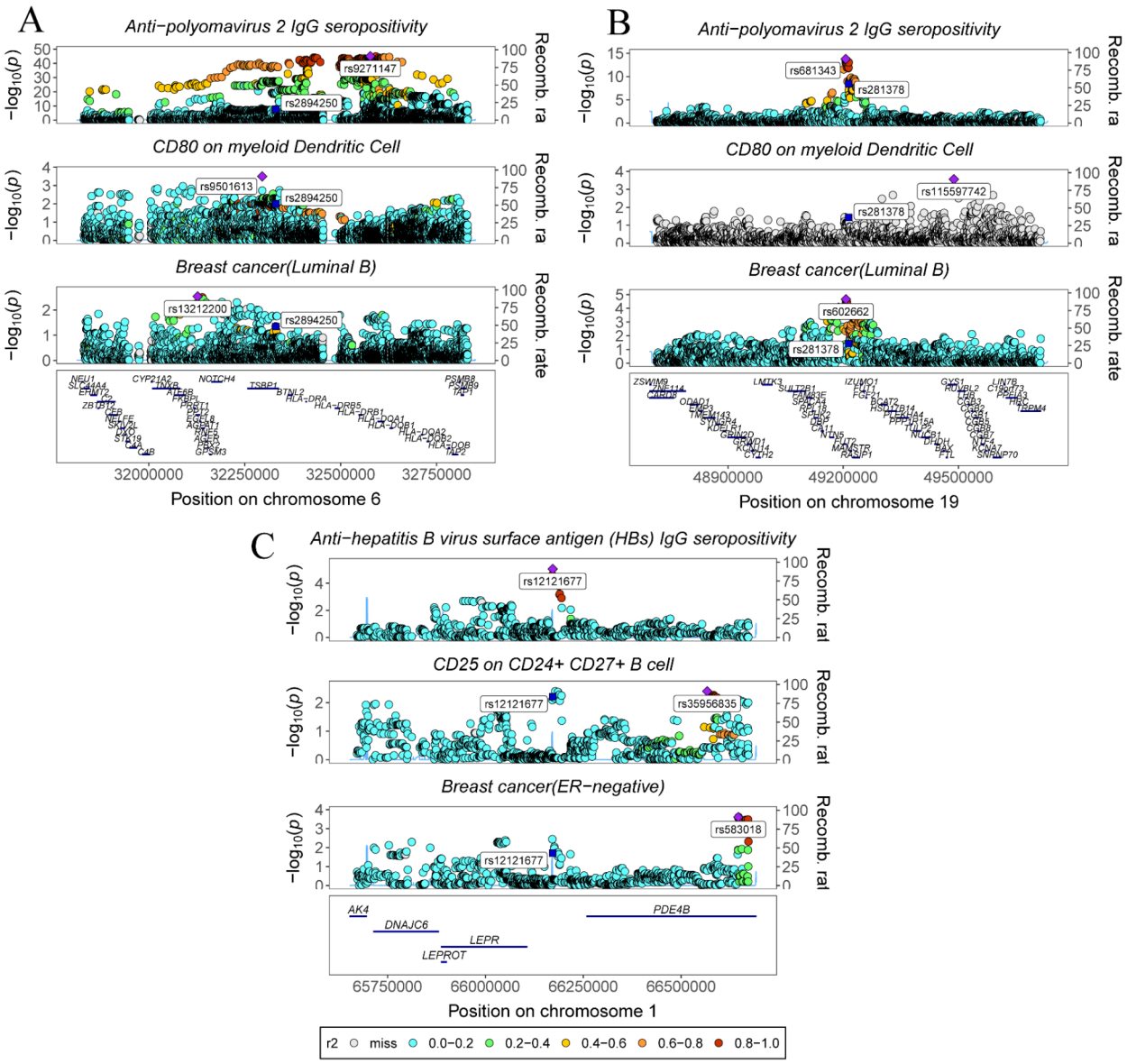

Fig. 6. Moloc colocalization diagram of three pleiotropic SNPs (rs2894250, rs281378, rs12121677) identified by CPASSOC analysis. (**A**) rs2894250 and (**B**) rs281378 showing the relationship between anti-polyomavirus 2 IgG seropositivity, CD80 expression on myeloid dendritic cells, and Luminal B; (**C**) rs12121677 illustrating the association between anti-hepatitis B virus surface antigen (HBs) IgG seropositivity, CD25 expression on CD24 + CD27 + B cells, and ER-negative.

Furthermore, studies have reported that Helicobacter hepaticus interacts with myeloid-derived suppressor cells to promote BC development[62,63]. Additionally, HP HopH protein can reduce VEGF expression and tumor size in BC mouse models, indicating HP also has potential in BC treatment[64]. These study results suggest potential associations between HP and BC, partially explaining the biological rationale of the MR results.

The above results suggest differences in immune responses to various pathogens among individuals. These differences have varying associations with different BC types, necessitating more detailed research to explore the relationship between pathogens and various BC types. For COVID-19 vaccination phenotypes, only antibody responses post-vaccination can be represented, not exposure to the COVID-19 virus. Therefore, no evidence was found for an association between COVID-19 vaccination and BC risk. This study used antibody immune status (positive) to represent lifetime exposure to pathogens. Related antibody levels represent the strength of immune responses to pathogens, but may be influenced by confounding factors such as vaccination. Therefore, the results of this part of the study should be interpreted with caution.

To explore potential mediation pathways and mechanisms between pathogen immune responses and BC, univariable MR analysis was first conducted on 731 immune cells, identifying 250 immune cell phenotypes potentially causally related to different BC types. After adjusting effect values using MVMR, CD25 on

CD24 + CD27 + B cells, CD25 on IgD + CD38- unswitched memory B cells, CD80 on myeloid dendritic cells (mDC), and CD4-CD8- T cells were identified as potential mediators. CD25 + B cells play important roles in normal immune responses, responding to interleukin (IL)-2[65]. CD25 + B cells have stronger antigen-presenting and cytokine-producing abilities compared to CD25- B cells[65,66]. mDCs are important antigen-presenting cells that participate in initiating and regulating immune responses by activating T cells[67]. Knockdown of CD80 and CD86 in mDCs alters Th1/Th2 cytokine production, indicating CD80 and CD86 in mDCs have cytokine-regulating functions[68]. CD4-CD8- T cells, known as double-negative T cells, do not express CD4 or CD8 molecules, can secrete cytokines, and directly contact cells to regulate the functions of other immune cells, playing important roles in controlling inflammation and anti-tumor responses[69]. The study results suggest that antibody responses to pathogens may influence the expression of these immune cell phenotypes, ultimately leading to BC. This implies that regulating antibody levels may influence the expression of different immune cells, potentially preventing BC. However, it should be noted that the P-value threshold was relaxed to identify more potential mediation pathways, which may lead to Type I errors. Therefore, the results should be interpreted with caution.

SNPs' effects on multiple phenotypes can be divided into horizontal pleiotropy and vertical pleiotropy[70]. Pleiotropic SNPs can simultaneously affect multiple phenotypes, potentially serving as therapeutic targets. After identifying four potential mediation pathways, LDSC analysis was conducted at the global level and CPASSOC and moloc colocalization analyses were conducted at the SNP level to explore whether pleiotropic associations existed. CPASSOC analysis identified three potential pleiotropic SNPs. Although these SNPs did not pass the Bayesian colocalization test, indicating that these SNPs are unlikely to simultaneously affect the three phenotypes within 1000 kb, the analysis results excluded horizontal pleiotropy, strengthening the credibility of causal inference.

The main strengths of this study include the innovative use of bidirectional MR methods, mediation analysis, and genetic correlation analysis. These methods provided genetic evidence for potential causal relationships between antibody immune responses and BC and revealed possible mediation pathways. MR methods have multiple advantages: first, they follow Mendelian principles, ensuring random allele distribution in offspring, effectively reducing bias from confounding factors such as reverse causality and environmental factors. Second, the BC GWAS data used in this study include the largest sample size of European populations, providing generalizability. However, limitations of MR analysis should be noted. MR results only represent genetic causal relationships, while exposure-outcome effects often depend on multiple factors, including genetic, environmental, social, and economic factors. Therefore, the specific effect sizes obtained in this study do not equate to actual effect sizes. Second, MR results represent lifelong exposure effects, unable to explain intervention effects at specific ages or times. Thus, results should be interpreted cautiously. However, MR analysis introduces triangulation evidence through a genetic perspective, contributing significantly to advancing related fields.

This study has several limitations that warrant consideration. First, at the whole-genome level ($P < 5e\text{-}8$), some phenotypes did not have sufficient IVs for analysis. Based on previous experience, a P-value threshold of $< 1e\text{-}5$ was used to select IVs. To avoid bias from weak IVs due to the relaxed threshold, only IVs with $F > 10$ were retained. Second, the study primarily explored associations between antibody immune responses, immune cell types, and BC without delving into underlying pathophysiological mechanisms. Moreover, many genetic variants' specific biological functions remain unclear, potentially influencing final results through unknown pleiotropy. Furthermore, while multiple pathogens were included in this study, many potential causal relationships between pathogen types and BC remain unexplored. Additionally, due to data availability and original study sample size, the statistical power of MR analysis might be insufficient, and sex- and age-stratified analyses were not conducted. The two-sample MR methods assume linear relationships between exposure and outcome, unable to analyze nonlinear associations directly. The GWAS summary data used in this study included only European participants, potentially limiting the generalizability of results to other populations. While correcting for multiple testing reduces Type I error probability, it increases Type II error possibility[71]. In this study, multiple BC subtypes were tested, and a strict significance threshold might be too conservative, potentially missing important causal relationships[72]. As a result, we prioritized results with $P_{IVW} < 0.05$, despite applying FDR adjustments. Therefore, when interpreting these findings, the possibility of false positives must be carefully considered. Some results showed confidence intervals close to 1, which may be due to smaller sample sizes, necessitating cautious interpretation of these specific findings. Given these limitations, translating current research findings into clinical practice requires further in-depth research and validation, including studies with larger and more diverse populations, sex- and age-stratified analyses, exploration of nonlinear relationships, investigation of biological mechanisms, examination of a broader range of pathogen-BC relationships, and validation through alternative methodologies and replication studies.

Conclusion

This study explored the relationships between various antibody immune responses, immune cell types, and BC subtypes, identifying potential causal relationships between pathogen antibody immune responses, immune cell types, and BC subtypes. The study further investigated potential regulatory mechanisms and mediation pathways of antibody immune responses-immune phenotypes-BC, providing possible directions for identifying immune targets. However, the specific mechanisms by which antibody immune responses and immune cells influence BC and their interactions require further in-depth research.

Data availability

The GWAS data on antibody immune responses and 731 types of immune cells can be accessed using PMID in the GWAS Catalog (https://www.ebi.ac.uk/gwas/, accessed on June 1, 2024). The GWAS data for breast cancer can be obtained from the BCAC website (https://bcac.ccge.medschl.cam.ac.uk/, accessed on July 10, 2023).

Received: 29 August 2024; Accepted: 11 November 2024
Published online: 19 November 2024

References

1. Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J. Clin.* **74** (1), 12–49 (2024).
2. Lawson, J. S. & Glenn, W. K. Catching viral breast cancer. *Infect. Agent Cancer*. **16** (1), 37 (2021).
3. Afzal, S. et al. Interrelated oncogenic viruses and breast cancer. *Front. Mol. Biosci.* **9**, 781111 (2022).
4. Yang, Z. et al. Human cytomegalovirus seropositivity and viral DNA in breast tumors are associated with poor patient prognosis. *Cancers (Basel)* **14**(5) (2022).
5. Xuan, C. et al. Microbial dysbiosis is associated with human breast cancer. *PLoS One*. **9** (1), e83744 (2014).
6. Hatta, M. N. A., Mohamad Hanif, E. A., Chin, S. F. & Neoh, H. M. Pathogens and carcinogenesis: A review. *Biology (Basel)* ;**10**(6) (2021).
7. De Mandal, S., Xu, X., Jin, F., Panda, A. K. & Ibrahim, K. S. *Recent Advances and Future Perspectives of Microbial Metabolites: Applications in Biomedicine* (2022).
8. Khasawneh, A. I. et al. Association of human papilloma virus, cytomegalovirus, and Epstein-Barr virus with breast cancer in Jordanian women. *Med. (Kaunas)* **60**(5) (2024).
9. Su, Y., Su, C. & Qin, L. Current landscape and perspective of oncolytic viruses and their combination therapies. *Transl. Oncol.* **25**, 101530 (2022).
10. Karagiannis, S. N. & Arnold, J. N. Immune cell-antibody interactions in health and disease. *Clin. Exp. Immunol.* **209** (1), 1–3 (2022).
11. Barnard, M. E., Boeke, C. E. & Tamimi, R. M. Established breast cancer risk factors and risk of intrinsic tumor subtypes. *Biochim. Biophys. Acta*. **1856** (1), 73–85 (2015).
12. Davey Smith, G. & Ebrahim, S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ*. **330** (7499), 1076–1079 (2005).
13. Sollis, E. et al. The NHGRI-EBI GWAS catalog: Knowledgebase and deposition resource. *Nucleic Acids Res.* **51** (D1), D977–D85 (2023).
14. Scepanovic, P. et al. Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med.* **10** (1), 59 (2018).
15. Butler-Laporte, G. et al. Genetic determinants of antibody-mediated Immune responses to infectious diseases agents: A genome-wide and HLA association study. *Open. Forum Infect. Dis.* **7** (11), ofaa450 (2020).
16. Bian, S. et al. Genetic determinants of IgG antibody response to COVID-19 vaccination. *Am. J. Hum. Genet.* **111** (1), 181–199 (2024).
17. Orru, V. et al. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. *Nat. Genet.* **52** (10), 1036–1045 (2020).
18. Zhang, H. et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* **52** (6), 572–581 (2020).
19. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. **551** (7678), 92–94 (2017).
20. Zheng, J. et al. Recent developments in mendelian randomization studies. *Curr. Epidemiol. Rep.* **4** (4), 330–345 (2017).
21. Ni, J-J. et al. Gut microbiota and psychiatric disorders: A two-sample mendelian randomization study. *Front. Microbiol.* **12**, 737197 (2021).
22. Lv, W-Q. et al. Human gut microbiome impacts skeletal muscle mass via gut microbial synthesis of the short-chain fatty acid butyrate among healthy menopausal women. *J. Cachexia Sarcopenia Muscle*. **12** (6), 1860–1870 (2021).
23. Byrska-Bishop, M. et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**(18) (2022).
24. Bowden, J. et al. Improving the accuracy of two-sample summary-data mendelian randomization: Moving beyond the NOME assumption. *Int. J. Epidemiol.* **48** (3), 728–742 (2019).
25. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in mendelian randomization studies. *Int. J. Epidemiol.* **40** (3), 755–764 (2011).
26. Hong, W. et al. Gut microbiome causal impacts on the prognosis of breast cancer: A mendelian randomization study. *BMC Genom.* **24** (1), 497 (2023).
27. Slob, E. A. W. & Burgess, S. A comparison of robust mendelian randomization methods using summary data. *Genet. Epidemiol.* **44** (4), 313–329 (2020).
28. Burgess, S. & Thompson, S. G. Interpreting findings from mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32** (5), 377–389 (2017).
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. Zhao, J. et al. Bayesian weighted mendelian randomization for causal inference based on summary statistics. *Bioinformatics*. **36** (5), 1501–1508 (2020).
31. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37** (7), 658–665 (2013).
32. Wang, S. et al. Systemic inflammatory regulators and risk of acute-on-chronic liver failure: A bidirectional mendelian-randomization study. *Front. Cell. Dev. Biol.* **11**, 1125233 (2023).
33. Korthauer, K. et al. A practical guide to methods controlling false discoveries in computational biology. *Genome Biol.* **20** (1), 118 (2019).
34. Bowden, J. et al. Improving the visualization, interpretation and analysis of two-sample summary data mendelian randomization via the radial plot and radial regression. *Int. J. Epidemiol.* **47** (6), 2100 (2018).
35. 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).
36. 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).
37. Burgess, S. Sample size and power calculations in mendelian randomization with a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43** (3), 922–929 (2014).
38. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47** (11), 1236–1241 (2015).
39. Li, X. & Zhu, X. Cross-phenotype association analysis using summary statistics from GWAS. *Methods Mol. Biol.* **1666**, 455–467 (2017).

40. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81** (3), 559–575 (2007).
41. Byrska-Bishop, M. et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell.* **185** (18), 3426–40e19 (2022).
42. Giambartolomei, C. et al. A bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics*. **34** (15), 2538–2545 (2018).
43. Adhikari, V. P., Lu, L. J. & Kong, L. Q. Does hepatitis B virus infection cause breast cancer? *Chin. Clin. Oncol.* **5** (6), 81 (2016).
44. Wang, Y. et al. Stability and anti-tumor effect of oncolytic herpes simplex virus type 2. *Oncotarget*. **9** (37), 24672–24683 (2018).
45. Wang, Y. et al. Preclinical safety evaluation of oncolytic herpes simplex virus type 2. *Hum. Gene Ther.* **30** (5), 651–660 (2019).
46. Li, H., Dutuor, A., Fu, X. & Zhang, X. Induction of strong antitumor immunity by an HSV-2-based oncolytic virus in a murine mammary tumor model. *J. Gene Med.* **9** (3), 161–169 (2007).
47. Eliassen, E. et al. Human herpesvirus 6 and malignancy: A review. *Front. Oncol.* **8**, 512 (2018).
48. Alibek, K. et al. Implication of human herpesviruses in oncogenesis through immune evasion and supression. *Infect. Agent Cancer*. **9** (1), 3 (2014).
49. Sahibsharrif, H., Almohsen, M. A. K., Al-Husseini, Y. & AL-Erjan, A. M. (eds) Correlation of polyomaviruses (PyV) infection with the incidence of breast cancer in Iraqi women. In *International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (IEEE, 2022).
50. Hachana, M. et al. Investigation of human JC and BK polyomaviruses in breast carcinomas. *Breast Cancer Res. Treat.* **133** (3), 969–977 (2012).
51. Zheng, H. C., Xue, H. & Zhang, C. Y. The oncogenic roles of JC polyomavirus in cancer. *Front. Oncol.* **12**, 976577 (2022).
52. Touma, J. et al. Detection of human cytomegalovirus proteins in paraffin-embedded breast cancer tissue specimens-A novel, automated immunohistochemical staining protocol. *Microorganisms* **9**(5) (2021).
53. Youssry, S., Hussein, A., Ramadan, R., Alkarmouty, A. & Elsheredy, A. The association of human cytomegalovirus with biomarkers of inflammation and immune activation in breast cancer. *Breast Dis.* **41** (1), 229–239 (2022).
54. Seo, J. Y., Heo, J. A. & Britt, W. J. Phosphorylation of tegument protein pp28 contributes to trafficking to the assembly compartment in human cytomegalovirus infection. *J. Microbiol.* **58** (7), 624–631 (2020).
55. Umberto, C., Giovanni, B. & Martino, R. Breast cancer analysis in 49 countries in relation to iron deficiency anemia (IDA), chlamydia and HIV infections. *J. Med. - Clin. Res. Rev.* **4** (5), 1–7 (2020).
56. Li, N. et al. Association of Chlamydia trachomatis infection with breast cancer risk and the modification effect of IL-12. *Clin. Breast Cancer* (2024).
57. Hua, Z. et al. Helicobacter pylori infection altered gastric microbiota in patients with chronic gastritis. *Front. Cell. Infect. Microbiol.* **13**, 1221433 (2023).
58. Figura, N. et al. Prevalence of Helicobacter pylori infection in male patients with osteoporosis and controls. *Dig. Dis. Sci.* **50** (5), 847–852 (2005).
59. Johansson, A. et al. Investigating the effect of estradiol levels on the risk of breast, endometrial, and ovarian cancer. *J. Endocr. Soc.* **6** (8), bvac100 (2022).
60. Lv, Z., Zhao, L. & Jin, W. Protein changes in gastric epithelial cells RGM-1 in response to Helicobacter pylori infection. *J. Cell. Biochem.* **120** (3), 3197–3202 (2019).
61. Shim, J. H. et al. The effect of Helicobacter pylori CagA on the HER-2 copy number and expression in gastric cancer. *Gene*. **546** (2), 288–296 (2014).
62. Deng, H. et al. Translocation of Helicobacter hepaticus synergizes with myeloid-derived suppressor cells and contributes to breast carcinogenesis. *Oncoimmunology*. **11** (1), 2057399 (2022).
63. Falsafi, T. & Mahboubi, M. Helicobacter hepaticus, a new pathogenic species of the Helicobacter Genus: similarities and differences with H. Pylori. *Iran. J. Microbiol.* **5** (3), 185–194 (2013).
64. Soleimani, N., Farhangi, B. & Tavakoli Yaraki, M. The effect of recombinant HopH protein of Helicobacter pylori on the VEGF expression in metastatic breast cancer model. *Acta Med. Iran.* **55** (12), 744–750 (2017).
65. Amu, S., Gjertsson, I. & Brisslert, M. Functional characterization of murine CD25 expressing B cells. *Scand. J. Immunol.* **71** (4), 275–282 (2010).
66. Amu, S. & Brisslert, M. Phenotype and function of CD25-expressing B lymphocytes isolated from human umbilical cord blood. *Clin. Dev. Immunol.* **2011**, 481948 (2011).
67. Moret, F. M. et al. Intra-articular CD1c-expressing myeloid dendritic cells from rheumatoid arthritis patients express a unique set of T cell-attracting chemokines and spontaneously induce Th1, Th17 and Th2 cell activity. *Arthritis Res. Ther.* **15** (5), R155 (2013).
68. Li, J. G. et al. CD80 and CD86 knockdown in dendritic cells regulates Th1/Th2 cytokine production in asthmatic mice. *Exp. Ther. Med.* **11** (3), 878–884 (2016).
69. Wu, Z. et al. CD3(+)CD4(-)CD8(-) (double-negative) T cells in inflammation, immune disorders and cancer. *Front. Immunol.* **13**, 816005 (2022).
70. Zhu, X. et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96** (1), 21–36 (2015).
71. Feise, R. J. Do multiple outcome measures require p-value adjustment? *BMC Med. Res. Methodol.* **2**, 8 (2002).
72. Burgess, S. et al. Guidelines for performing mendelian randomization investigations: Update for summer 2023. *Wellcome Open. Res.* **4**, 186 (2019).

Acknowledgements

Not Applicable.

Author contributions

YY, JC and FG wrote the main manuscript text and prepared the figures; JM, ML, RL and CW searched the literature and critically revised the manuscript. FG and WC conceived the idea for the review and provided the final approval. All authors reviewed the manuscript and approved the submitted version.

Funding

This study was funded by the National Natural Science Foundation of China (No. 82060538; No. 82060543), the Middle-aged and Young Academic and Technical Leader Reserve Talent Program of Yunnan Province Science and Technology Department (No. 202205AC160008), the 535 Talent Project of the First Affiliated Hospital of Kunming Medical University (No. 2023535D10). The key Project of Yunnan Applied Basic Research (202401AS070023).

Declarations

Ethics approval and consent to participate

Since this study utilized de-identified public data, there was no need for additional approval. All original ethical approvals are available in the original literature and on the website.

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/s41598-024-79521-w.

Correspondence and requests for materials should be addressed to F.G. or W.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.