The association between plasma chemokines and breast cancer risk and prognosis: A Mendelian randomization study

Xingxing Yu¹, Yanyu Zhang¹, Yuxiang Lin^{2,3,4}, Shuqing Zou¹, Pingxiu Zhu¹, Mengjie Song¹, Fangmeng Fu^{2,3,4}, Haomin Yang^{1,5}

Author affiliations:

1. Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, 350122, Fuzhou, China
2. Department of Breast Surgery, Fujian Medical University Union Hospital, Fuzhou, 350001, China
3. Department of General Surgery, Fujian Medical University Union Hospital, Fuzhou, 350001, China
4. Breast Cancer Institute, Fujian Medical University, Fuzhou, 350001, China
5. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels Väg 12A, 171 77 Stockholm, Sweden.

***Corresponding author:**

Haomin Yang, PhD. Email: haomin.yang@fjmu.edu.cn; Phone: 0086-13581540402; Fax: 0086-591-22862510. Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, Xuefu North Road 1, University Town, Fuzhou, China.

Fangmeng Fu, MD, PhD. Email: ffm@fjmu.edu.cn. Department of Breast Surgery, Fujian Medical University Union Hospital, Fuzhou, 350001, China

Tables:2; Figures:2; Supplementary Tables: 4; References: 38

Word count: 248 (abstract); 2288 (text excluding abstract, acknowledgement, reference, tables and figures)

Running title: Chemokines and breast cancer

Abstract

Background: Despite the potential role of several chemokines in the migration of cytotoxic immune cells to prohibit breast cancer cell proliferation, a comprehensive view of chemokines and the risk and prognosis of breast cancer is scarce, and little is known about their causal associations.

Methods: With a two-sample Mendelian randomization (MR) approach, genetic instruments associated with 30 plasma chemokines were created. Their genetic associations with breast cancer risk and survival were extracted from the recent genome-wide association study, with available survival information for 96,661 patients. We further tested the associations between the polygenetic risk score (PRS) for chemokines and breast cancer in the UK Biobank cohort using logistic regression models. The association between chemokine expression in tumors and breast cancer survival was analyzed in the TCGA cohort with Cox regression models.

Results: Plasma CCL5 was causally associated with the risk of breast cancer in the MR analysis, which was significant in the luminal and HER-2 enriched subtypes and further confirmed using PRS analysis (OR=0.94, 95% CI=0.89-1.00). A potential causal association with breast cancer survival was only found for plasma CCL19, especially for ER-positive patients. In addition, we also found an inverse association between CCL19 expression in tumors and breast cancer overall and relapse-free survival (HR=0.58, 95% CI=0.35-0.95).

Conclusion: We observed an inverse association between genetic predisposition to CCL5 and the risk of breast cancer, while CCL19 was associated with breast cancer survival. These associations suggested the potential of these chemokines as tools for breast cancer prevention and treatment.

Keywords: chemokines, breast cancer, Mendelian randomization

Introduction

Breast cancer is one of the most common malignant tumors, which seriously affects the physical and mental health of women worldwide and accounts for 24.2% and 15.0% of the total number of new cancers and deaths in women, respectively [1]. Although breast cancer has caused a considerable global burden, its etiology is not yet completely understood. Currently, the known breast cancer risk factors include age, sex, race, family history, genetic mutations, early menarche/late menopause and hormone replacement therapy, while evidence for the causal effect of inflammation in breast cancer carcinogenesis is still limited [2, 3]. To prevent the incidence of breast cancer and find potential therapeutic targets, we need to better understand the role of inflammatory biomarkers in breast cancer.

At the first stage of malignant evolution, chronic inflammation may play an important role during the initiation, development and prognosis of breast cancer [4], in which chemokines might regulate the migration of immune cells toward the tumor and act as biomarkers for breast cancer. Studies have shown that CCL-4, -7, -8, -11, -15, -16, -19, -22, -23, -24, -25 and CXCL-5, -8, -9, and -16 were increased in breast cancer compared with normal adjacent breast tissue [5], and genetic polymorphisms near genes encoding CCL-2, -4, -5, -20, -22, CXCL-10, -12 and XCL-1 were found to be associated with the risk and prognosis of breast cancer [6-13]. However, traditional observational studies are susceptible to reverse causation and residual confounding [14]. Contrary to previous studies, Bowen Chen et al [15] used several bioinformatics analysis tools to analyze the expression data of CC chemokines in patients with BC and yielded inconsistent results. Therefore, it is still a challenge to fully understand the causal association between chemokines and breast cancer.

Mendelian randomization (MR) uses genetic variation as an instrumental variable (IV) to

infer the association between exposure factors and disease outcomes. Because the genotypes are presumed to be randomly allocated in the process of gamete formation, MR analyses are largely free of reverse causality and less sensitive to confounding factors, which overcomes the limitations of conventional observational studies [14, 16].

In the present study, we implemented a two-sample MR design to investigate the potential causal effect of plasma chemokines on the risk and prognosis of breast cancer and to analyze their molecular subtypes to provide more credible evidence on the potential role of chemokines in the development of breast cancer. We further used the UK Biobank and TCGA cohorts to validate our findings.

Methods

Study population

Mendelian randomization analysis was used to assess the causal association between plasma chemokines and breast cancer risk. In the first step, the associations were tested using a two-sample approach with summary statistics from genome-wide association studies (GWAS), which does not need individual-level data[17]. Genetic instrumental variables for plasma chemokines were created by identifying single nucleotide polymorphisms (SNPs) associated with 41 plasma chemokines in a recent GWAS on the human plasma proteome at the GWAS significant level ($p < 5 \times 10^{-8}$) [18]. From the 5965 SNPs reported in the original study, 84 SNPs for 30 plasma chemokines were selected after excluding SNPs with linkage disequilibrium (LD threshold: $r^2 < 0.01$) and palindromes. The SNPs used in this study and their corresponding plasma chemokines and proportion of variance explained are listed in sTable 1. We used R^2 and F statistics to estimate the proportion of variance explained by

SNPs, with the following formulas $R^2 = \frac{2\beta^2 * (MAF) * (1-MAF)}{2\beta^2 * (MAF) * (1-MAF) + (se(\beta))^2 * 2N * MAF * (1-MAF)}$:

and $F = \frac{R^2 * (N-1-k)}{(1-R^2) * k}$, in which β is the coefficient for the SNP, MAF is the minor allele frequency, $se(\beta)$ is the standard error of the coefficient, N is the sample size and k is the number of SNPs used.

GWAS summary statistics for breast cancer risk overall and by molecular subtypes (luminal A, luminal B, Her-2 enriched and triple negative) were downloaded from the Breast Cancer Association Consortium [19], with 133,384 breast cancer cases and 113,789 controls. The GWAS summary statistics for breast cancer survival were also downloaded from BCAC, including 96,661 patients with 7,697 breast cancer mortality cases; among them, 64,171 were ER-positive patients, and 16,172 were ER-negative [20]. All studies involved in the GWAS on the human plasma proteome and the Breast Cancer Association Consortium were approved by the relevant research ethics committee, and all participants provided informed consent.

For those identified plasma chemokines in the two sample Mendelian randomization studies for breast cancer risk, a single-sample Mendelian randomization was further applied to validate the findings in UK Biobank. Briefly, the UK Biobank is a prospective cohort comprising approximately 502,505 participants (55% women) of various ethnicities, aged 40-69 at recruitment between 2006 and 2010. At enrolment, information on the participants' health-related factors and socio-demographic characteristics and biological samples were collected in 22 assessment centers around UK.

Blood samples from women in the UK Biobank were genotyped using a custom UK Biobank Axiom Array, comprising 825,927 SNPs. Details of the array design, sample handling, quality control processes and imputation of non-genotyped variants are described elsewhere [21]. Breast cancer cases were obtained through linkage with the national cancer registries in the UK and updated until 2015. The ICD-10 code C50 and ICD-9 code 174 were used to identify breast cancer cases,

while the controls were women without a diagnosis of breast cancer. To assess whether breast cancer is associated with a genetic predisposition to plasma chemokines, we selected the same genome-wide significant SNPs used in the two-sample MR. For all individuals, a weighted polygenic risk score (PRS) was calculated using the following formula:

$$PRS = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \beta_n x_n$$

where β is the per-allele log odds ratio (OR) of the breast cancer-associated risk allele for SNP k , x_k is the number of alleles for the same SNP (0, 1, 2), and n is the total number of breast cancer SNPs included in the profile. For the analyses, PRS was categorized into tertiles.

To further explore the effect of the identified chemokines associated with breast cancer survival, 1077 women in the TCGA breast cancer cohort were followed for their overall and breast cancer survival. Gene expression data of the chemokines in tumor tissues were also downloaded and categorized into tertiles.

Statistical Analysis

Several methods were used to estimate the potential causal effect of plasma chemokines on breast cancer risk and prognosis. The main analysis was based on a fixed-effects meta-analysis using an inverse-variance weighting model (IVW), in which β and se were calculated using the following formulas $\beta_{IVW} = \frac{\sum_{i=1}^N \beta_{exp.i} \beta_{out.i} \sigma_{out.i}^{-2}}{\sum_{i=1}^N \beta_{exp.i} \sigma_{out.i}^{-2}}$ and $se(\beta_{IVW}) = \frac{1}{\sum_{i=1}^N \beta_{exp.i} \sigma_{out.i}^{-2}}$: N is the total number of SNPs, $\beta_{exp.i}$ is the coefficient of the i -th SNP on the exposure, $\beta_{out.i}$ is the coefficient of the i -th SNP on the outcome, and $\sigma_{out.i}$ is the standard error of $\beta_{out.i}$. If only one SNP was identified for the chemokine, the Wald ratio method was used by calculating the ratio of β_{exp} and β_{out} , while se could be approximately estimated using the Delta method. [22] We further performed sensitivity analysis to test horizontal pleiotropy using weighted median and Egger regression methods. The

weighted median estimator is consistent with the IVW estimation even when up to 50% of the SNPs are invalid instrumental variables, while MR-Egger regression simply adds an intercept term of pleiotropy into the IVW regression model [23].

To test the key assumption of no horizontal pleiotropy in MR, we further analyzed the effect of plasma chemokines on potential risk factors for breast cancer, e.g., age at menarche, age at menopause, age at smoking initiation, and BMI, using their corresponding GWAS summary statistics [24-26].

To validate the discovered potential causal association between plasma chemokines and breast cancer, we used logistic regression models to calculate the ORs of breast cancer among UK biobank women by tertiles of PRS for plasma chemokines. These associations were adjusted for age, the first ten principal components, and the assessment center. We also used the Cox model to estimate the hazard ratios of overall survival and relapse-free survival in the TCGA cohort by tertiles of gene expression of the chemokines in tumor tissues and adjusted for age at diagnosis, estrogen, progesterone, and tumor stage.

All statistical tests were two-sided and performed using R 3.6.3 and Stata 15.1.

Results

In the main analysis, the selected SNPs explained 0.9%–36.3% of the variance in plasma chemokines, and the corresponding F-statistic ranged from 29.7–1879.3 (sTable 1), suggesting no evidence of weak instrument bias. A higher plasma CCL5 was associated with a lower risk of breast cancer (OR=0.89, 95% CI: 0.85 to 0.94, $p<0.001$), which was primarily attributed to the luminal A and luminal B subtypes (OR for luminal A=0.88, 95% CI: 0.82 to 0.94; OR for luminal B=0.80, 95% CI: 0.69 to 0.94). Plasma growth-regulated alpha protein and Ck-beta-8 were also associated with the

risk of breast cancer, although the p value did not survive the Bonferroni correction for multiple testing (Figure 1, sTable 2).

In the subtype-specific analysis, plasma CCL19 was associated with luminal A breast cancer, while CXCL16 was associated with HER-2-enriched breast cancer. In addition, plasma CCL16 was positively associated with the risk of luminal A breast cancer but inversely associated with triple-negative breast cancer. Sensitivity analysis using MR Egger and weighted median methods yielded similar results (sTable 2).

To validate the findings in the two-sample MR analysis, 264,187 women who had available genotyped data from UK Biobank were involved in the analysis, with 14760 breast cancer patients. In multivariable logistic regression analysis, a higher polygenic risk scores for plasma CCL5 was inversely associated with breast cancer (OR for the highest tertile =0.94, 95% CI=0.89-1.00, $P=0.045$). However, no statistically significant associations for plasma growth-regulated alpha protein, Ck-beta-8 and breast cancer were found. (Table 1).

In the analysis of breast cancer survival, we found a potential causal effect of plasma CCL19 on breast cancer survival (HR=0.85, 95% CI=0.75-0.96, $P<0.01$), especially for ER-positive breast cancer patients (HR=0.82, 95% CI=0.70-0.97) (Figure 2, sTable 3). Moreover, in the TCGA breast cancer patient cohort, CCL19 expression in tumor tissues was associated with increased overall and relapse-free survival (HR for the last tertile vs. first tertile=0.58, 95% CI=0.35-0.95). The association was also validated for ER-positive patients (Table 2).

MR estimates for the effects of plasma chemokines on breast cancer risk factors indicated little evidence of horizontal pleiotropy (sTable 4).

Discussion

In this Mendelian randomization study of 30 plasma chemokines, we found an inverse association between genetically predicted plasma CCL5 and breast cancer risk in both the two-sample and one-sample analyses. This association was more pronounced in luminal and HER-2-enriched breast cancer patients. An effect on cancer survival was only found for plasma CCL19, especially for ER-positive patients. In addition, we also found an inverse association between CCL19 expression and breast cancer mortality using tumor tissue samples.

As key proteins in immune cell migration, chemokines have been studied for their effect on inflammation-associated cancers. One previous Mendelian randomization analysis on cytokines found an association between genetically predicted CCL2, CCL4, GRO α and the risk of breast cancer. As the SNPs associated with plasma CCL2 and CCL4 did not exceed the GWAS significance threshold, we were not able to replicate these findings in our study. However, both our study and a previous study confirmed the association between GRO α and the risk of breast cancer, confirming the role of GRO α in breast cancer carcinogenesis and probably from the effect of Th17 cells [27].

Our analysis found an inverse association between CCL5 and breast cancer risk, with a similar effect size in a recent phenome-wide Mendelian randomization [28]. We further suggested that the association was driven by luminal and HER-2-enriched subtypes of breast cancer. Although tumor-derived CCL5 in breast cancer patients was found to be associated with poor prognosis [29], plasma CCL5 might still play an antitumor role, as the function of the CCL5/CCR5 axis varies in different conditions[30]. CCL5 can mediate the migration of immune cells and promote lymphocyte migration in breast cancer cells [31]. Many autoimmune diseases are associated with increased levels of CCL5 and inversely associated with breast cancer risk [32, 33]. All this evidence supports the role of plasma CCL5 in breast tumor carcinogenesis.

In addition, both plasma CCL19 and CCL19 gene expression in tumors was associated with increased survival of breast cancer in our study. As downregulation of the CCL19/CCR7 axis has been observed in metastatic breast cancer compared to the primary tumor[34] and CCL19 suppresses angiogenesis and proliferation of several other types of cancers [35, 36], it is biologically plausible that CCL19 inhibits breast cancer metastasis through similar AIM2 and Met/ERK/Elk-1/HIF-1 α /VEGF-A pathways [37] and therefore increases patient survival.

Interestingly, we also found that many of the associations between chemokines and breast cancer risk or survival were more pronounced in luminal A- or ER-positive cancers, suggesting that the effect of chemokines might be related to estrogen-related carcinogenesis [38]. However, we still found a potential causal association between CCL16 and triple-negative breast cancer. Further study is therefore needed to confirm this association and test its potential role as a biomarker or treatment target for triple-negative breast cancer.

The strength of our study included the use of the most recent and largest summary statistics for breast cancer with subtype-specific information and a comprehensive dataset for plasma chemokines. Our study had several limitations. As no sex-specific GWAS for chemokines has been reported, we assumed that the sex difference in the chemokine GWAS was small and that the original GWAS had already adjusted for age, sex and principal components. The generalizability of the findings is also affected by the use of European ancestry, and the results should be replicated in other populations.

Conclusion

In conclusion, we found that lower plasma CCL5 was causally associated with higher breast cancer risk, while plasma and tumor-derived CCL19 was associated with increased survival of breast cancer patients. Both findings were more pronounced in ER-positive patients, indicating the potential

use of CCL5 and CCL19 as biomarkers or drug targets in future studies for breast cancer prevention and treatment.

CONFLICT OF INTEREST STATEMENT

None declared.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All studies used for the calculation of breast cancer GWAS were approved by the relevant research ethics committee, and all participants provided informed consent.

AVAILABILITY OF DATA AND MATERIALS

Summary statistics for pQTLs of plasma chemokines can be downloaded from the IEU open GWAS project (<https://gwas.mrcieu.ac.uk/>). Summary statistics for breast cancer risk and survival are available on the BCAC website. (<http://bcac.ccge.medschl.cam.ac.uk/>). Data from the UK Biobank (<http://www.ukbiobank.ac.uk/>) are available to all researchers upon making an application. Part of this research was conducted using the UK Biobank Resource under Application 61083.

AUTHOR CONTRIBUTIONS:

XY and HY had full access to all data and take responsibility for the integrity of the data and the accuracy of the analysis.

Study concept and design: HY

Acquisition, analysis, and interpretation of data: XY

Drafting of the manuscript: XY, YZ

Critical revision of the manuscript for important intellectual content: HY, XY, YL, FF

Statistical analysis: XY, YZ

Funding: HY.

Role of the Funding Source

This work is supported by the Technology Development Fund from the Department of Education of Fujian Province [grant no: 2019L3010008]. Haomin Yang is supported by the Natural Science Foundation of Fujian Province [grant no: 2021J01721], the Startup Fund for High-level Talents of Fujian Medical University [grant no: XRCZX2020007], Startup Fund for Scientific Research, Fujian Medical University [grant no: 2019QH1002] and Laboratory Construction Program of Fujian Medical University [grant no: 1100160208]. The sponsor of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Reference

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Winters S, Martin C, Murphy D, Shokar NK: **Breast Cancer Epidemiology, Prevention, and Screening.** *Prog Mol Biol Transl Sci* 2017, **151**:1-32.
3. Deshmukh SK, Srivastava SK, Poosarla T, Dyess DL, Holliday NP, Singh AP, Singh S: **Inflammation, immunosuppressive microenvironment and breast cancer: opportunities for cancer prevention and therapy.** *Ann Transl Med* 2019, **7**(20):593.
4. Danforth DN: **The Role of Chronic Inflammation in the Development of Breast Cancer.** *Cancers* 2021, **13**(15):21.
5. Abrahamsson A, Rzepecka A, Dabrosin C: **Equal Pro-inflammatory Profiles of CCLs, CXCLs, and Matrix Metalloproteinases in the Extracellular Microenvironment In Vivo in Human Dense Breast Tissue and Breast Cancer.** *Front Immunol* 2017, **8**:1994.
6. He S, Zhang X: **The rs1024611 in the CCL2 gene and risk of gynecological cancer in Asians: a meta-analysis.** *World J Surg Oncol* 2018, **16**(1):34.
7. Hu GN, Tzeng HE, Chen PC, Wang CQ, Zhao YM, Wang Y, Su CM, Tang CH: **Correlation between CCL4 gene polymorphisms and clinical aspects of breast cancer.** *Int J Med Sci* 2018, **15**(11):1179-1186.
8. Eskandari-Nasab E, Hashemi M, Ebrahimi M, Amininia S, Bahari G, Mashhadi MA, Taheri M: **Evaluation of CCL5 -403 G>A and CCR5 Delta32 gene polymorphisms in patients with breast cancer.** *Cancer Biomark* 2014, **14**(5):343-351.
9. Jafarzadeh A, Fooladseresht H, Nemati M, Assadollahi Z, Sheikhi A, Ghaderi A: **Higher circulating levels of chemokine CXCL10 in patients with breast cancer: Evaluation of the influences of tumor stage and chemokine gene polymorphism.** *Cancer Biomark* 2016, **16**(4):545-554.
10. Li J, Lindstrom LS, Foo JN, Rafiq S, Schmidt MK, Pharoah PD, Michailidou K, Dennis J, Bolla MK, Wang Q *et al*: **2q36.3 is associated with prognosis for oestrogen receptor-negative breast cancer patients treated with chemotherapy.** *Nature communications* 2014, **5**:4051.
11. Jafarzadeh A, Fooladseresht H, Minaee K, Bazrafshani MR, Khosravimashizi A, Nemati M, Mohammadzadeh M, Mohammadi MM, Ghaderi A: **Higher circulating levels of chemokine CCL22 in patients with breast cancer: evaluation of the influences of tumor stage and chemokine gene polymorphism.** *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 2015, **36**(2):1163-1171.
12. Stephens KE, Levine JD, Aouizerat BE, Paul SM, Abrams G, Conley YP, Miaskowski C: **Associations between genetic and epigenetic variations in cytokine genes and mild persistent breast pain in women following breast cancer surgery.** *Cytokine* 2017, **99**:203-213.
13. Chou WC, Hsiung CN, Chen WT, Tseng LM, Wang HC, Chu HW, Hou MF, Yu JC, Shen CY: **A functional variant near XCL1 gene improves breast cancer survival via promoting cancer immunity.** *Int J Cancer* 2020, **146**(8):2182-2193.
14. Smith GD, Ebrahim S: **'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?** *Int J Epidemiol* 2003, **32**(1):1-22.
15. Chen B, Zhang S, Li Q, Wu S, He H, Huang J: **Bioinformatics identification of CCL8/21 as potential prognostic biomarkers in breast cancer microenvironment.** *Biosci Rep* 2020, **40**(11).
16. Davies NM, Holmes MV, Davey Smith G: **Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians.** *BMJ* 2018, **362**:k601.
17. Pierce BL, Burgess S: **Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators.** *Am J Epidemiol* 2013, **178**(7):1177-1184.
18. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P *et al*: **Genomic atlas of the human plasma proteome.** *Nature* 2018, **558**(7708):73-79.
19. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, Jiang X, O'Mara TA, Zhao N, Bolla MK *et al*: **Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses.** *Nature genetics* 2020, **52**(6):572-581.
20. Escala-Garcia M, Guo Q, Dörk T, Canisius S, Keeman R, Dennis J, Beesley J, Lecarpentier J, Bolla MK, Wang Q *et al*: **Genome-wide association study of germline variants and breast cancer-specific mortality.** *Br J Cancer* 2019, **120**(6):647-657.

21. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J *et al*: **The UK Biobank resource with deep phenotyping and genomic data.** *Nature* 2018, **562**(7726):203-209.
22. Pagoni P, Dimou NL, Murphy N, Stergiakouli E: **Using Mendelian randomisation to assess causality in observational studies.** *Evidence Based Mental Health* 2019, **22**(2):67.
23. Bowden J, Davey Smith G, Haycock PC, Burgess S: **Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator.** *Genetic epidemiology* 2016, **40**(4):304-314.
24. Wood AR, Tyrrell J, Beaumont R, Jones SE, Tuke MA, Ruth KS, Yaghootkar H, Freathy RM, Murray A, Frayling TM *et al*: **Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively.** *Diabetologia* 2016, **59**(6):1214-1221.
25. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, Datta G, Davila-Velderrain J, McGuire D, Tian C *et al*: **Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use.** *Nature genetics* 2019, **51**(2):237-244.
26. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, Bates P, Palmer T, Haberland V, Smith GD *et al*: **The MRC IEU OpenGWAS data infrastructure.** *bioRxiv* 2020:2020.2008.2010.244293.
27. Ma K, Yang L, Shen R, Kong B, Chen W, Liang J, Tang G, Zhang B: **Th17 cells regulate the production of CXCL1 in breast cancer.** *International immunopharmacology* 2018, **56**:320-329.
28. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, Gutteridge A, Erola P, Liu Y, Luo S *et al*: **Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases.** *Nature genetics* 2020, **52**(10):1122-1131.
29. Soria G, Ben-Baruch A: **The inflammatory chemokines CCL2 and CCL5 in breast cancer.** *Cancer letters* 2008, **267**(2):271-285.
30. Fujimoto Y, Inoue N, Morimoto K, Watanabe T, Hirota S, Imamura M, Matsushita Y, Katagiri T, Okamura H, Miyoshi Y: **Significant association between high serum CCL5 levels and better disease-free survival of patients with early breast cancer.** *Cancer Science* 2020, **111**(1):209-218.
31. Wang Z, Han J, Cui Y, Zhou X, Fan K: **miRNA-21 inhibition enhances RANTES and IP-10 release in MCF-7 via PIAS3 and STAT3 signalling and causes increased lymphocyte migration.** *Biochemical and biophysical research communications* 2013, **439**(3):384-389.
32. Zeng Z, Lan T, Wei Y, Wei X: **CCL5/CCR5 axis in human diseases and related treatments.** *Genes & Diseases* 2021.
33. Wadström H, Pettersson A, Smedby KE, Askling J: **Risk of breast cancer before and after rheumatoid arthritis, and the impact of hormonal factors.** *Annals of the rheumatic diseases* 2020, **79**(5):581-586.
34. Szekely B, Bossuyt V, Li X, Wali VB, Patwardhan GA, Frederick C, Silber A, Park T, Harigopal M, Pelekanou V *et al*: **Immunological differences between primary and metastatic breast cancer.** *Annals of Oncology* 2018, **29**(11):2232-2239.
35. Xu Z, Zhu C, Chen C, Zong Y, Feng H, Liu D, Feng W, Zhao J, Lu A: **CCL19 suppresses angiogenesis through promoting miR-206 and inhibiting Met/ERK/Elk-1/HIF-1 α /VEGF-A pathway in colorectal cancer.** *Cell Death & Disease* 2018, **9**(10):974.
36. Zhou R, Sun J, He C, Huang C, Yu H: **CCL19 suppresses gastric cancer cell proliferation, migration, and invasion through the CCL19/CCR7/AIM2 pathway.** *Human Cell* 2020, **33**(4):1120-1132.
37. Chen I-F, Ou-Yang F, Hung J-Y, Liu J-C, Wang H, Wang S-C, Hou M-F, Hortobagyi GN, Hung M-C: **AIM2 suppresses human breast cancer cell proliferation *in vitro* and mammary tumor growth in a mouse model.** *Molecular Cancer Therapeutics* 2006, **5**(1):1-7.
38. Lim CL, Lin VCL: **Estrogen markedly reduces circulating low-density neutrophils and enhances pro-tumoral gene expression in neutrophil of tumour-bearing mice.** *BMC cancer* 2021, **21**(1):1017.

FIGURE LEGENDS

Figure 1. The causal association between plasma chemokines and breast cancer risk overall and by molecular subtype.

OR: odds ratio. Statistical significance is indicated by *** ($p < 0.001$), ** ($p < 0.01$) and * ($p < 0.05$). The estimates are based on the fixed effect of the inverse variance weighted method or Wald ratio method for meta-analyzing individual SNP results. ORs show the risk of breast cancer overall and by subtypes per SD increase of ranked-inverse normalized concentration in exposure to plasma chemokines.

Figure 2. The causal association between plasma chemokines and breast cancer survival

HR: hazard ratio. Statistical significance is indicated by ** ($p < 0.01$) and * ($p < 0.05$). The estimates are based on the fixed effect of the inverse variance weighted method or Wald ratio method for meta-analyzing individual SNP results. HRs show the risk of breast cancer overall and by subtypes per SD increase of ranked-inverse normalized concentration in exposure to plasma chemokines.

Supplementary Tables

sTable 1. List of the SNPs used as the genetic instrument for two-sample MR analysis

sTable 2. Causal association between plasma chemokines and overall and subtypes of breast cancer using different methods of two-sample MR analysis

sTable 3. Causal association between plasma chemokines and overall and subtypes of breast cancer survival using different methods of two-sample MR analysis

sTable 4. The potential causal association between plasma chemokine and major risk factors of breast cancer

Figure 1.

Table 1. The association between polygenic risk scores for chemokines and breast cancer in the UK Biobank

Chemokine	No.	No. of cases	OR (95% CI)	<i>P</i> value
Growth-regulated alpha protein				
Tertile 1	131908	7461	1.00(REF)	
Tertile 2	91150	5184	1.02 (0.98-1.06)	0.251
Tertile 3	41129	2115	0.97 (0.92-1.02)	0.194
C-C motif chemokine 5				
Tertile 1	136038	7636	1.00(REF)	
Tertile 2	98239	5503	0.98 (0.95-1.02)	0.273
Tertile 3	29910	1621	0.94 (0.89-1.00)	0.045
Ck-beta-8-1				
Tertile 1	91937	5073	1.00(REF)	
Tertile 2	121826	6799	0.97 (0.93-1.00)	0.073
Tertile 3	50424	2888	0.99 (0.94-1.04)	0.664

The model was adjusted for age at recruitment, assessment center and top 10 principal genetic components.

Figure 2.

Table 2. The association between CCL19 and survival of breast cancer patients in TCGA cohort.

	Total No.	Overall survival		Relapse free survival	
		No. of cases	HR (95% CI)	No. of cases	HR (95% CI)
Breast cancer overall					
Tertile 1	358	64	1.00(REF)	39	1.00(REF)
Tertile 2	358	46	0.64 (0.43-0.96)	23	0.49 (0.29-0.85)
Tertile 3	357	43	0.66 (0.43-1.00)	34	0.58 (0.35-0.95)
ER+ breast cancer					
Tertile 1	264	46	1.00(REF)	26	1.00(REF)
Tertile 2	275	33	0.52 (0.32-0.84)	19	0.51 (0.27-0.97)
Tertile 3	249	22	0.44 (0.26-0.77)	17	0.43 (0.22-0.82)
ER- breast cancer					
Tertile 1	72	12	1.00(REF)	10	1.00(REF)
Tertile 2	72	10	0.92 (0.38-2.24)	4	0.40 (0.12-1.38)
Tertile 3	94	20	1.09 (0.49-2.44)	16	0.81 (0.33-1.98)

The model was adjusted for age at diagnosis, ER, PR and stage.