Shu Xiao-Ou (Orcid ID: 0000-0002-0711-8314)
Guo Xingyi (Orcid ID: 0000-0001-5269-1294)
Dörk Thilo (Orcid ID: 0000-0002-9458-0282)
Gago-Dominguez Manuela (Orcid ID: 0000-0001-6713-4351)
Zheng Wei (Orcid ID: 0000-0003-1226-070X)

Evaluation of Associations between Genetically Predicted Circulating Protein Biomarkers and Breast Cancer Risk

Xiang Shu[1], Jiandong Bao[1], Lang Wu[1], Jirong Long[1], Xiao-Ou Shu[1], Xingyi Guo[1], Yaohua Yang[1], Kyriaki Michailidou[2,3], Manjeet K. Bolla[2], Qin Wang[2], Joe Dennis[2], Irene L. Andrulis[4,5], Jose E. Castelao[6], Thilo Dörk[7], Manuela Gago-Dominguez[8,9], Montserrat García-Closas[10], Graham G. Giles[11,12], Artitaya Lophatananon[13,14], Kenneth Muir[13,14], Håkan Olsson[15], Gadi Rennert[16], Emmanouil Saloustros[17], Rodney J. Scott[18,19], Melissa C. Southey[20], Paul D.P. Pharoah[2,21], Roger L. Milne[11,12], Peter Kraft[22,23], Jacques Simard[24], Douglas F. Easton[2,21], Wei Zheng[1]

1 Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA.

2 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

3 Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus.

4 Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada.

5 Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.32542

6 Oncology and Genetics Unit, Instituto de Investigacion Biomedica (IBI) Galicia Sur, Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain.

7 Gynaecology Research Unit, Hannover Medical School, Hannover, Germany.

8 Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain.

9 Moores Cancer Center, University of California San Diego, La Jolla, CA, USA.

10 Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA.

11 Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia.

12 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.

13 Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, UK.

14 Institute of Population Health, University of Manchester, Manchester, UK.

15 Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden.

16 Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel.

17 Hereditary Cancer Clinic, University Hospital of Heraklion, Heraklion, Greece.

18 Division of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, New South Wales, Australia.

19 Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Callaghan, New South Wales, Australia.

20 Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia.

21 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.

22 Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

23 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

24 Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, QC, Canada.

Correspondence

Prof. Wei Zheng
Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center
Vanderbilt University School of Medicine
2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738, USA
Telephone: 615-936-0682
E-mail: wei.zheng@vanderbilt.edu

Keywords: Breast cancer; circulating protein biomarkers; genetics; instruments

Abbreviations:

OR

Odds ratio

CI

Confidence interval

BH-FDR

The Benjamini-Hochberg false discovery rate

BCAC

Breast Cancer Association Consortium

iCOGS

The Collaborative Oncological Gene-environment Study

GWAS

Genome-wide association study

pQTL

Protein quantitative trait loci

Estrogen receptor-positive/-negative

ER+/-

CRP

C-reactive protein

IGF1

Insulin-like growth factor 1

ISLR2

Leucine-rich repeat protein 2

IR

Insulin receptor

MET

Hepatocyte growth factor receptor

NOTCH1

Neurogenic locus notch homolog protein 1

VEGFR2

Vascular endothelial growth factor receptor 2

B3GNT2

Beta-1,3-N-Acetylglucosaminyltransferase 2

RSPO3

R-spondin 3

VCAM1

Vascular cell adhesion protein 1

Category: Research Articles

Novelty and Impact

The study identified 56 circulating proteins, for which their genetically predicted levels were associated with breast cancer risk. These proteins are involved in estrogen receptor signaling, insulin resistance, and other important biological processes, and may serve as candidate biomarkers for further investigations.

Abstract

A small number of circulating proteins have been reported to be associated with breast cancer risk, with inconsistent results. Herein, we attempted to identify novel protein biomarkers for breast cancer via the integration of genomics and proteomics data. In the Breast Cancer

Association Consortium (BCAC), with 122,977 cases and 105,974 controls of European descendants, we evaluated the associations of the genetically predicted concentrations of >1,400 circulating proteins with breast cancer risk. We used data from a large-scale protein quantitative trait loci (pQTL) analysis as our study instrument. Summary statistics for these pQTL variants related to breast cancer risk were obtained from the BCAC and used to estimate odds ratios (OR) for each protein using the inverse-variance weighted method. We identified 56 proteins significantly associated with breast cancer risk by instrumental analysis (false discovery rate < 0.05). Of these, the concentrations of 32 were influenced by variants close to a breast cancer susceptibility locus (*ABO*, 9q34.2). Many of these proteins, such as insulin receptor, insulin-like growth factor receptor 1 and other membrane receptors (OR: 0.82 to 1.18, *P* values: 6.96×10^{-4} to 3.28×10^{-8}), are linked to insulin resistance and estrogen receptor signaling pathways. Proteins identified at other loci include those involved in biological processes such as alcohol and lipid metabolism, proteolysis, apoptosis, immune regulation, and cell motility and proliferation. Consistent associations were observed for 22 proteins in the UK Biobank data ($P < 0.05$). The study identifies potential novel biomarkers for breast cancer, but further investigation is needed to replicate our findings.

Introduction

Breast cancer is the most common malignancy diagnosed among women in many countries [1]. Established risk factors for breast cancer include certain menstrual and reproductive factors, postmenopausal obesity, the use of hormone replacement therapy, family history of the disease, and the carrying of high-penetrance mutations [2, 3]. Circulating protein biomarkers have an important utility in cancer screening and risk assessment [4]. Several circulating protein biomarkers of breast cancer risk have been reported in observational studies. Some examples of these are C-reactive proteins (CRP), insulin-like growth factor 1 (IGF1), and leptin [5-8]. However, conventional observational studies may be influenced by reverse causation, confounding, selection biases, or small sample sizes. Therefore, results from previous studies have been inconsistent.

There is compelling evidence that the concentration of many circulating proteins may be determined by genetic variants [9, 10]. A twin study measured 342 proteins in plasma and estimated that the mean heritability was ~14% [11]. Since genetic alleles are randomly distributed during gamete formation, the variations in protein concentration determined by genetic variants should not be affected by environmental exposures or lifestyle factors. Therefore, the use of genetic variants as instruments to investigate circulating proteins in relation to cancer risk can reduce confounding effects, selection biases, and circumvent reverse causation, all of which are frequently encountered in epidemiological studies [12]. Importantly, the genetically determined protein concentrations represent a long-term exposure since birth. Recently, Sun et al. identified 1,927 genome-wide significant protein quantitative trait loci (pQTL) in individuals of European

ancestry [10]. Herein, we have utilized these pQTL variants as instruments to evaluate the genetically predicted concentration of each of the 1,469 proteins in relation to breast cancer risk in the Breast Cancer Association Consortium (BCAC). The identified associations were further assessed using UK Biobank data.

Material and Methods

An inverse-variance weighted method [13] was used to evaluate the associations of predicted circulating protein concentrations with breast cancer risk using summary statistics data from two sources. The first was beta coefficients of the associations between genetic variants and circulating protein concentrations. These were obtained from a recent genome-wide association study (GWAS) to identify protein quantitative trait loci (pQTL) that evaluated 2,994 circulating proteins in 3,301 healthy subjects of European descent [10]. The proteins were quantified using SOMAscan platform. A total of 1,927 associations were identified for 1,478 circulating proteins with a $P < 1.5\times10^{-11}$. The second source of summary statistics for each of these pQTL variants came from the GWAS of breast cancer risk in the BCAC studies that comprised three datasets: 11 individual breast cancer GWAS combined (14,910 cases and 17,588 controls), the Collaborative Oncological Gene-environment Study (iCOGS) (46,785 cases and 42,892 controls), and the OncoArray study (61,282 cases and 45,494 controls) [14-16]. Summary statistics of iCOGS and OncoArray can be accessed through the BCAC website

(http://bcac.ccge.medschl.cam.ac.uk/bcacdata/). All participating studies of the BCAC were approved by their corresponding ethics review boards and all subjects provided informed consent. Our analyses were limited to the women of European ancestry included in the BCAC. Details of the genotyping protocols in the BCAC have been published elsewhere [15-17] (iCOGS: http://ccge.medschl.cam.ac.uk/research/consortia/icogs/). Samples included in the OncoArray, iCOGS, and nine of the individual GWAS datasets were imputed by IMPUTE version 2, using the 1000 Genomes Project (October 2014 version 3 release) dataset as the reference panel [16, 18]. Two of the individual GWAS, BPC3 and EBCG were imputed separately using MACH and Minimac [19, 20].

Approximately 25% of the instruments were constructed using multiple genetic variants. When multiple variants were associated with a single protein, only those with linkage disequilibrium (LD) < 0.1 were retained for downstream analyses. The F-statistic was used to measure the strength of the instruments, with 10 being a commonly used threshold [21]. It was calculated following the formula R^2*(n-1-k)/(1-R^2)/k, where R^2 is percentage of variance explained by used SNPs; n is the sample size of BCAC data (=228,951); and k is the number of SNPs used in the instrument. Thanks to the large sample size of the BCAC, all of the instruments have an F-statistic of > 1,000. The beta coefficient of the association between genetically predicted concentrations of a given protein and breast cancer risk was estimated using $\sum_i \ \beta_{i,GX} * \beta_{i,GY} * \sigma_{i,GY}^{-2} / (\sum_i \ \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})$ and its standard error calculated as $1/(\sum_i \ \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})^{0.5}$, where $\beta_{i,GX}$ is the beta coefficient of the association between *i* th SNP and the protein of interest from

the above-mentioned pQTL study [10]. *GY* represents the association between *i* th SNP and breast cancer risk in the BCAC meta-analysis [overall, estrogen receptor-positive (ER+), or ER-negative (ER-)]; thus, $\beta_{i,GY}$ and $\sigma_{i,GY}$ are the corresponding beta coefficient and standard error for SNP_i, respectively. Odds ratio (OR) was expressed as the exponential of beta coefficients. The Benjamini-Hochberg false discovery rate (FDR) of < 0.05 was used as the significance level for a two-sided test. The ingenuity pathway analysis was employed to visualize the potential interplay of genes and proteins.

We downloaded genetic association summary statistics for each identified risk-associated protein from http://www.phpc.cam.ac.uk/ceu/proteins (EGAS00001002555). We abstracted pQTL variants associated at the level of 5×10^{-8} (LD $R^2 < 0.1$) to construct new instruments, then performed sensitivity analyses.

Summary statistics derived from the associations between genetic variants and breast cancer risk using the UK Biobank samples were obtained to replicate the associations revealed in the BCAC [22]. The imputation was completed by combining the Haplotype Reference Consortium and the UK10K haplotype resource as the reference panel. Genome-wide association analyses for over 2,000 phenotypes were conducted using data from ~337,000 unrelated individuals of British ancestry included in the UK Biobank (http://www.nealelab.is/uk-biobank). The statistics for the associations of SNPs with breast cancer risk were used for our validation study. A highly correlated proxy SNP ($R^2 > 0.9$) was identified and used to construct the genetic instrument if the original SNP was not available (e.g. if insertions/deletions or variants failed in the quality control

assessment). Since the majority of the 2,000 phenotypes were in continuous form, a linear regression model was employed for all phenotypes, including binary outcomes. We obtained summary statistics data derived from the analysis conducted for histologically-confirmed incident breast cancers (ICD-10: C50, N = 5,510) and prevalent breast cancers reported by participants at the baseline interview (N =7,480). We also abstracted pQTL variants from other two genome-wide pQTL studies [9, 23] ($P < 5\times10^{-8}$, LD $R^2 < 0.1$) to construct new instruments for the risk-associated proteins. Both studies were conducted in the populations of European descent and the same SOMAscan platform was used to measure blood proteins. Genome-wide association analysis was performed to identify significant pQTL variants for circulating proteins in the two studies. The aforementioned inverse-variance weighted method was applied to validate our primary findings. The potential pleiotropic effects of our genetic instruments were investigated via Phenoscanner [24]. The lead pQTL variants for the identified proteins were queried.

Results

We constructed genetic instruments for 1,469 out of the 1,478 proteins, using one to six genetic variants that were associated with the circulating concentration of each protein. Of the 375 proteins whose concentrations could be predicted using multiple variants, 27 showed associations with overall breast cancer risk, after accounting for multiple comparisons (FDR <

0.05, Table 1). Of them, the concentrations of 10 proteins were positively associated with breast cancer risk (ORs ranging from 1.03 to 1.08 per unit of increase; *P* values ranging from 1.41×10^{-3} to 1.19×10^{-7}), while the other 17 were inversely associated with the risk (ORs ranging from 0.90 to 0.98; *P* values ranging from 1.56×10^{-3} to 5.20×10^{-8}). The most noticeable association was observed for the immunoglobulin superfamily, particularly the leucine-rich repeat protein 2 (ISLR2), of which the genetically predicted concentration was inversely associated with breast cancer risk (OR: 0.93, $P = 5.20\times 10^{-8}$). Analyses using single-variant instruments identified an additional 29 proteins at 11 loci, with their predicted circulating concentrations associated with overall breast cancer risk after accounting for multiple comparisons (FDR < 0.05, Table 2). The effect sizes of these associations were comparable and consistent in direction across the three independent datasets included in the BCAC for all associated proteins (supplementary Table S1).

A recently reported breast cancer susceptibility locus, 9q34.2 (*ABO*) [16], showed strong pleiotropy with 32 risk-associated proteins. All of the instruments for these 32 protein biomarkers were constructed using genetic variants located at 9q34.2, alone or in combination with variants in other chromosomes. For instance, both genetically predicted concentrations of IGF1 receptors and insulin receptors were associated with a reduced risk of breast cancer (IGF1R: OR = 0.82, IR: OR =0.93, $P = 3.28\times10^{-8}$ for both proteins). Of the 32 proteins, 20 are membrane receptors and 11 are linked to the estrogen receptor via the ingenuity pathway analysis (supplementary Figure S1). The majority of the risk-associated proteins showed consistent associations for ER positive (+) and ER negative (-) breast cancer (supplementary

Table S2). The association for seven of the proteins was stronger in risk of ER- than of ER+ breast cancer ($P_{het} < 0.05$ from heterogeneity tests, Table 3).

In the sensitivity analysis, we constructed new instruments using independent pQTL variants, with $P < 5\times10^{-8}$ for the 56 identified proteins (see Methods). The associations were not materially changed (Supplementary Table S3). Furthermore, we provided another source of validation via new instruments (see Methods). The associations of 27 proteins were replicated at $P < 0.05$ (Table 4). The pleiotropic effects of lead pQTL variants for the 56 proteins are presented in supplementary Table S4 ($P < 5\times10^{-8}$).

We evaluated 55 of the 56 predicted protein biomarkers with breast cancer risk in the UK Biobank data (data for Fas ligand via rs371314787 were not available and no proxy could be identified within 500 Kb). Consistent associations were observed for 22 proteins with either histologically-confirmed or self-reported breast cancer risk (Table 5, $P < 0.05$). We also observed nominally consistent associations for an additional four proteins (Table 5, $P < 0.1$).

Discussion

The use of genetic variants as instrumental variables to assess the exposure-outcome relationship could help reduce confounding and selection bias, and eliminate biases due to reverse causation [12]. Using data from a large-scale consortium (the BCAC), we identified 56 circulating protein biomarkers associated with overall breast cancer risk, after adjusting for multiple comparisons. Of these, 22 associations were nominally replicated using data from the UK Biobank, providing assurance of the validity of our findings. Although the causality cannot be determined for the identified proteins, our study provides substantial new information about protein biomarker candidates for breast cancer risk.

A recently reported breast cancer susceptibility locus, 9q34.2 [16], was related to more than half of the identified protein risk biomarkers. This region is known for its wide spectrum of pleiotropy on its concentrations of metabolites [25], lipids [26], and proteins [9, 10], as well as its risk of coronary artery disease [27] and pancreatic cancer [28]. The biological mechanisms underlying this pleiotropy remain obscure. The ingenuity pathway analysis revealed a network of multiple membrane proteins regulated by genetic variants in 9q34.2, such as insulin-like growth factor 1 receptor (IGF1R), insulin receptor (IR), hepatocyte growth factor receptor (MET), neurogenic locus notch homolog protein 1 (NOTCH1), and vascular endothelial growth factor receptor 2 (VEGFR2). All of these are linked to estrogen receptor signaling (Supplementary Figure S1) and insulin resistance [29-32]. It is possible that the genetic variants in 9q34.2 may affect the concentration of these receptors, subsequently leading to impaired insulin sensitivity and/or

abnormal estrogen signaling. Of note, the biological activities triggered by the contacts between ligands and these receptors are not limited to insulin resistance. Some examples are: 1) insulin/IGF1 has a mitogenic effect on cell proliferation and growth through its binding to IR/IGF1R [33]; 2) hepatocyte growth factor (HGF) and MET are important players in mammary gland development [34]; 3) the notch signaling pathway is widely involved in diverse developmental and homeostatic processes [35]; 4) VEGF/VEGFR2 interaction is responsible for developmental and pathological angiogenesis [36]. Thus, the down-regulation of these receptors may trigger carcinogenic effects through elongated interactions with their ligands. However, the causality between specific proteins and breast cancer risk cannot be established due to the strong pleiotropy observed in this and several other loci. Nevertheless, these proteins can serve as candidate biomarkers for future studies.

We also identified associations for several other proteins that play important roles in various biological processes. For instance, we found that higher genetically predicted concentrations of copine 1and Fas ligand were associated with a reduced risk of overall breast cancer. Copine 1 belongs to calcium-dependent membrane-binding proteins. It has been shown that the interaction between copine 1 and p65 could repress the transcription of NF-κB, which is essential for cancer initiation and progression [37]. Fas and Fas ligand's role in apoptosis and immune homeostasis have long been acknowledged. By engaging Fas ligand with Fas in the cancer cell membrane, CD8+ cytotoxic lymphocytes can activate caspase 8 and initiate the apoptotic death of cancer cells [38]. Thus, our analyses support the hypothesis that down-regulated

copine 1 and Fas ligands may contribute to breast cancer risk. Another example is beta-1,3-N-Acetylglucosaminyltransferase 2 (B3GNT2), which is the main polylactosamine synthase [39]. Its important role in glycan formation (glycosylation) and its well-known link to aberrant glycosylation and carcinogenesis [40] support the hypothesis that the B3GNT2 protein may have a carcinogenic effect. The positive association we observed for genetically predicted B3GNT2 concentrations and breast cancer risk is consistent with a previous report that *B3GNT2* expression was up-regulated in malignant breast tissues compared to that of normal tissues from healthy women [40].

The interpretation of some of our findings is less straightforward. For example, we identified an inverse correlation for a genetically predicted R-spondin 3 (RSPO3) concentration with breast cancer risk. R-spondins are critical regulators in the canonical Wnt/β-catenin pathway, and they have been shown to be activators of this pathway [41]. It would be expected that excessive R-spondins may be positively related to breast cancer development, as the over-activation of Wnt signaling is generally considered to be mechanistically related to cancer initiation [42]. Similarly, we observed an inverse relationship between genetically predicted vascular cell adhesion protein 1 (VCAM1) concentrations and breast cancer risk, which deserves further investigation. VCAM1 has been widely studied for its role in promoting tumor angiogenesis, progression, and metastasis [43], but less for its effect on tumorigenesis. Thus, some of the associations observed in the present study should be interpreted cautiously, despite their statistical significance.

We evaluated multiple circulating protein biomarkers reported by previous epidemiological studies using genetic instruments. The results were not entirely consistent with previous findings. For example, we observed only a nominal association between genetically predicted CRP concentrations and overall breast cancer risk (OR: 1.04, $P = 0.048$), and the association was trivial after adjusting for multiple comparisons. This contradicts a recent meta-analysis suggesting that an elevated circulating concentration of CRP may be associated with an increased breast cancer risk [5, 44]. Of note, this association varied greatly between retrospective case-control studies and prospective cohort/nested case-control studies. Similarly, no association was observed for the genetically predicted concentrations of IGF1 and HGF, which is in contrast with the findings for measured proteins in previous studies [7, 38, 45]. These inconsistencies could be due to the inaccurate or confounded estimates of associations commonly encountered in traditional observational epidemiological studies. It is also possible that genetic instruments we used in this study are not adequate for the analysis.

Our study is unprecedented in its power to discover novel circulating protein biomarkers for breast cancer risk. Our findings also provide novel evidence revealing mechanistic networks underlying breast carcinogenesis. The identified proteins could serve as candidates in future investigations. Another great strength of this study is that we were able to validate a large number of our identified associations using an independent dataset from the UK Biobank. That we were able to replicate associations with breast cancer for 22 out of 55 proteins from an independent dataset cannot be explained by chance alone (binomial $P = 5\times10^{-10}$ when assuming

5% of the associations could be replicated). We recognize that the number of breast cancer cases was relatively small in the UK Biobank and the SNP-breast cancer associations were derived from a linear regression model instead of a logistic regression model. The association estimates and 95% CIs might be biased; Thus, the design of replication stage was not ideal. We also recognize that the strong genetic pleiotropy, particularly at 9q34.2, prevents us from making causal inferences. Some of the associations may be largely attributable to the correlation between protein concentrations. In addition, our analysis depends on the number of proteins measured and the number of pQTLs identified in the GWAS of circulating protein concentrations. A more comprehensive analysis could be conducted when data becomes available for proteins that have not been evaluated in the current study. The explained variation in protein concentrations is also expected to improve with the identification of additional pQTLs. Lastly, all of the pQTLs involved in the current analysis were identified in the blood. Whether these pQTLs were consistent in breast tissues remains unknown. However, we found that the coding genes of nearly all the proteins analyzed in the current study are expressed in the breast tissues from the Genotype-Tissue Expression project (data not shown).

The Strong pleiotropy of pQTL variants has limited our ability to infer causal roles of the identified proteins in breast cancer development. However, some pleiotropy may reveal potential intermediate phenotypes involved in the causal pathways for breast cancer. For example, rs2489623 (RSPO3) and rs7041 (JAG1) were associated with both central obesity and circulating 25-OH vitamin D concentrations [46-48], respectively. Another example is that pQTLs at 3p21.31

associated with multiple proteins (ADH1B, CRYBB2, DOCK9, STOM, TMPRSS11D, and TNS2) were also linked to age at menarche, which is an established risk factor for breast cancer. Additionally, solid evidence has linked the 9q34.2 (ABO) locus to invasive ovarian cancer risk [49]. Thus, proteins associated with variants in 9q34.2 could be potential biomarkers for both malignancies.

Although sensitivity analysis and extra replication were conducted, our findings should be interpreted with caution. To establish causality and understand the underlying mechanisms for the identified proteins in breast cancer etiology, assays at cell levels such as cell viability and colony formation assays, and whole-transcriptome profiling are potential next steps to depict the impact on breast cancer cell growth and gene-gene interplay networks, after knocking down the encoding genes.

In summary, using genetic instruments, we identified 56 proteins with their genetically predicted circulating levels associated with breast cancer risk in this study. Future investigations are needed to replicate our findings, particularly for the proteins that failed to reach statistical significance in the UK Biobank dataset. Understanding and establishing causality for the identified proteins are important next steps.

Acknowledgement

The current affiliation for JB is College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65: 87-108.

2. Colditz GA, Bohlke K. Priorities for the primary prevention of breast cancer. *CA Cancer J Clin* 2014;64: 186-94.

3. Howell A, Anderson AS, Clarke RB, Duffy SW, Evans DG, Garcia-Closas M, Gescher AJ, Key TJ, Saxton JM, Harvie MN. Risk determination and prevention of breast cancer. *Breast Cancer Res* 2014;16: 446.

4. Borrebaeck CA. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. *Nat Rev Cancer* 2017;17: 199-204.

5. Chan DS, Bandera EV, Greenwood DC, Norat T. Circulating C-Reactive Protein and Breast Cancer Risk-Systematic Literature Review and Meta-analysis of Prospective Cohort Studies. *Cancer Epidemiol Biomarkers Prev* 2015;24: 1439-49.

6. Schernhammer ES, Holly JM, Pollak MN, Hankinson SE. Circulating levels of insulin-like growth factors, their binding proteins, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005;14: 699-704.

7. Endogenous H, Breast Cancer Collaborative G, Key TJ, Appleby PN, Reeves GK, Roddam AW. 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* 2010;11: 530-42.

8. Ollberding NJ, Kim Y, Shvetsov YB, Wilkens LR, Franke AA, Cooney RV, Maskarinec G, Hernandez BY, Henderson BE, Le Marchand L, Kolonel LN, Goodman MT. Prediagnostic leptin, adiponectin, C-reactive protein, and the risk of postmenopausal breast cancer. *Cancer Prev Res (Phila)* 2013;6: 188-95.

9. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, Sarwath H, Thareja G, Wahl A, DeLisle RK, Gold L, Pezer M, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017;8: 14357.

10. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, Oliver-Williams C, Kamat MA, et al. Genomic atlas of the human plasma proteome. *Nature* 2018;558: 73-9.

11. Liu Y, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, Vitek O, Mouritsen J, Lachance G, Spector TD, Dermitzakis ET, Aebersold R. Quantitative variability of 342 plasma proteins in a human twin population. *Molecular systems biology* 2015;11: 786.

12. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* 2017;26: 2333-55.

13. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37: 658-65.

14. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, Casey G, Hunter DJ, Sellers TA, Gruber SB, Dunning AM, Michailidou K, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* 2017;26: 126-35.

15. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;45: 353-61, 61e1-2.

16. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, Lemacon A, Soucy P, Glubb D, Rostamianfar A, Bolla MK, Wang Q, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017.

17. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015;47: 373-80.

18. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5: e1000529.

19. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010;34: 816-34.

20. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012;44: 955-9.

21. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. *International journal of epidemiology* 2011;40: 755-64.

22. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* 2017.

23. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, Hoover H, Gudmundsdottir V, Horman SR, Aspelund T, Shu L, Trifonov V, et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* 2018;361: 769-73.

24. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, Young R, Butterworth AS. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016;32: 3207-9.

25. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, Arnold M, Erte I, Forgetta V, Yang TP, Walter K, Menni C, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;46: 543-50.

26. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45: 1274-83.

27. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, Zeng L, Ntalla I, Lai FY, Hopewell JC, Giannakopoulou O, Jiang T, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet* 2017;49: 1385-91.

28. Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, Arslan AA, Beane-Freeman L, Bracci PM, Buring J, Canzian F, Duell EJ, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet* 2014;46: 994-1000.

29. Clemmons DR. The relative roles of growth hormone and IGF-1 in controlling insulin sensitivity. *J Clin Invest* 2004;113: 25-7.

30. Fafalios A, Ma J, Tan X, Stoops J, Luo J, Defrances MC, Zarnegar R. A hepatocyte growth factor receptor (Met)-insulin receptor hybrid governs hepatic glucose metabolism. *Nat Med* 2011;17: 1577-84.

31. Pajvani UB, Shawber CJ, Samuel VT, Birkenfeld AL, Shulman GI, Kitajewski J, Accili D. Inhibition of Notch signaling ameliorates insulin resistance in a FoxO1-dependent manner. *Nat Med* 2011;17: 961-7.

32. Robciuc MR, Kivela R, Williams IM, de Boer JF, van Dijk TH, Elamaa H, Tigistu-Sahle F, Molotkov D, Leppanen VM, Kakela R, Eklund L, Wasserman DH, et al. VEGFB/VEGFR1-Induced Expansion of Adipose Vasculature Counteracts Obesity and Related Metabolic Complications. *Cell Metab* 2016;23: 712-24.

33. Wilcox G. Insulin and insulin resistance. *Clin Biochem Rev* 2005;26: 19-39.

34. Yant J, Buluwela L, Niranjan B, Gusterson B, Kamalati T. In vivo effects of hepatocyte growth factor/scatter factor on mouse mammary gland development. *Exp Cell Res* 1998;241: 476-81.

35. Bray SJ. Notch signalling in context. *Nat Rev Mol Cell Biol* 2016;17: 722-35.

36. Shibuya M. Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies. *Genes Cancer* 2011;2: 1097-105.

37. Ramsey CS, Yeung F, Stoddard PB, Li D, Creutz CE, Mayo MW. Copine-I represses NF-kappaB transcription by endoproteolysis of p65. *Oncogene* 2008;27: 3516-26.

38. Peter ME, Hadji A, Murmann AE, Brockway S, Putzbach W, Pattanayak A, Ceppi P. The role of CD95 and CD95 ligand in cancer. *Cell Death Differ* 2015;22: 549-59.

39. Togayachi A, Kozono Y, Kuno A, Ohkura T, Sato T, Hirabayashi J, Ikehara Y, Narimatsu H. Beta3GnT2 (B3GNT2), a major polylactosamine synthase: analysis of B3GNT2-deficient mice. *Methods Enzymol* 2010;479: 185-204.

40. Potapenko IO, Haakensen VD, Luders T, Helland A, Bukholm I, Sorlie T, Kristensen VN, Lingjaerde OC, Borresen-Dale AL. Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Mol Oncol* 2010;4: 98-118.

41. Jin YR, Yoon JK. The R-spondin family of proteins: emerging regulators of WNT signaling. *Int J Biochem Cell Biol* 2012;44: 2278-87.

42. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene* 2017;36: 1461-73.

43. Schlesinger M, Bendas G. Vascular cell adhesion molecule-1 (VCAM-1)--an increasing insight into its role in tumorigenicity and metastasis. *Int J Cancer* 2015;136: 2504-14.

44. Guo L, Liu S, Zhang S, Chen Q, Zhang M, Quan P, Lu J, Sun X. C-reactive protein and risk of breast cancer: A systematic review and meta-analysis. *Sci Rep* 2015;5: 10508.

45. Sheen-Chen SM, Liu YW, Eng HL, Chou FF. Serum levels of hepatocyte growth factor in patients with breast cancer. *Cancer Epidemiol Biomarkers Prev* 2005;14: 715-7.

46. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, Kilpelainen TO, Esko T, Magi R, Li S, Workalemahu T, Feitosa MF, et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* 2013;9: e1003500.

47. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, Kiel DP, Streeten EA, Ohlsson C, Koller DL, Peltonen L, Cooper JD, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 2010;376: 180-8.

48. Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, Jacobs EJ, Ascherio A, Helzlsouer K, Jacobs KB, Li Q, Weinstein SJ, et al. Genome-wide association study of circulating vitamin D levels. *Human molecular genetics* 2010;19: 2739-45.

49. Kuchenbaecker KB, Ramus SJ, Tyrer J, Lee A, Shen HC, Beesley J, Lawrenson K, McGuffog L, Healey S, Lee JM, Spindler TJ, Lin YG, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet* 2015;47: 164-71.

Figure legend

Figure S1. Multiple identified proteins are relevant to estrogen receptor signaling

The interplay of identified proteins were generated using the Ingenuity Pathway Analysis software. Several proteins associated with genetic variants in 9q34.2 (ABO) are linked to estrogen receptor signaling.

Tables

Table 1. Associations between genetically predicted concentrations of circulating proteins and breast cancer risk using multi-variant instruments

Protein	Protein-associated SNPs	OR (95% CI)	*P* value
ISLR2	rs115478735, rs2959011, rs4055121	0.93 (0.90-0.95)	5.20×10^{-8}
C1GALT1C1	rs2519093, rs7787942	1.06 (1.04-1.09)	1.19×10^{-7}
ALPI	rs550057, rs679574	0.90 (0.87-0.94)	8.61×10^{-7}

FAM177A1	rs550057, rs679574	0.91 (0.87-0.94)	1.28×10^{-6}
SELP	rs2519093, rs6136, rs74227709	0.96 (0.94-0.98)	1.69×10^{-5}

Table 2. Associations between genetically predicted concentrations of circulating proteins and breast cancer risk using single-variant instruments

CPNE1	rs12481228, rs62143206	0.96 (0.95-0.98)	4.69×10^{-5}
RELT	rs3741148, rs7952686	1.07 (1.03-1.10)	6.06×10^{-5}
CTSF	rs11347749, rs1791679	1.08 (1.04-1.11)	8.68×10^{-5}
TPST2	rs115478735, rs2283824, rs34436714	1.06 (1.03-1.10)	1.22×10^{-4}
VEGFR3/FLT4	rs10935473, rs2519093, rs34221241	0.97 (0.96-0.99)	1.56×10^{-4}
KIN	rs149092047, rs62143198, rs7412	0.94 (0.91-0.97)	1.64×10^{-4}
QSOX2	rs10858248, rs149092047	1.03 (1.01-1.04)	1.67×10^{-4}
B3GNT2	rs2519093, rs67047091	1.05 (1.02-1.07)	2.29×10^{-4}
VEGFR2/KDR	rs34231037, rs34336071, rs635634	0.96 (0.94-0.98)	2.54×10^{-4}
KLRF1	rs11708955, rs62143194	1.10 (1.05-1.16)	2.72×10^{-4}
MAN1A2	rs35505705, rs8176643	1.08 (1.04-1.13)	3.07×10^{-4}
TIE1 (soluble)	rs10935473, rs2275180, rs8176743	0.97 (0.95-0.99)	3.32×10^{-4}
CAMK1	rs4525, rs61751507	0.98 (0.96-0.99)	3.53×10^{-4}
GOLM1	rs149092047, rs601338	1.04 (1.02-1.06)	4.11×10^{-4}
BCAM	rs144579705, rs8176747	0.92 (0.88-0.96)	5.48×10^{-4}
AKR1A1	rs62143198, rs72688441	0.97 (0.96-0.99)	9.68×10^{-4}
THSD1	rs2519093, rs41292808	0.95 (0.92-0.98)	1.14×10^{-3}
PSD	rs1303, rs429358	0.96 (0.94-0.99)	1.14×10^{-3}
SULF2	rs10424405, rs7614709, rs7971133	0.94 (0.91-0.98)	1.15×10^{-3}
PRDM1	rs13093385, rs2232613	1.03 (1.01-1.05)	1.41×10^{-3}
SEMA6A	rs3733724, rs56278466, rs8176743	0.94 (0.91-0.98)	1.52×10^{-3}
JAG1	rs550057, rs7041	0.93 (0.90-0.97)	1.56×10^{-3}

Protein-associated SNPs	Locus	Protein	ORs	*P* values
rs2519093[a]	9q34.2, *ABO*	IGF1R (soluble), IR, MET, IL3RA, SELE (soluble), ENG, LIFR (soluble), FAM20B, ICAM2 (soluble), CHST15	0.82 to 1.18	1.39×10^{-6} to 3.28×10^{-8}
rs3184504	12q24.12, *SH2B3*	VCAM1	0.85	4.09×10^{-7}
rs3197999	3p21.31, *MST1*	TMPRSS11D, DOCK9, TNS2, CRYBB2	1.06 to 1.16	2.51×10^{-5}
rs151288400	7q22.1, *PILRB*	HTN1	1.09	4.62×10^{-5}
rs6770670[a]	3p21.31, *BSN*	STOM, ADH1B	1.06	7.17×10^{-5}
rs371314787	3p21.31, *APEH*	FASLG (soluble)	0.89	1.58×10^{-4}
rs2205895	1q24.2, *SELP*	GAL	0.88	3.90×10^{-4}
rs1800594[a]	1q24.2, *F5*	SEC13, TFPI	0.88 to 1.08	1.46×10^{-3} to 5.21×10^{-4}
rs8176693[a]	9q34.2, *ABO*	CD36, NOTCH1, TLL1, CDH5	0.90 to 0.93	6.96×10^{-4} to 6.41×10^{-4}
rs2489623	6q22.33, *RSPO3*	RSPO3	0.92	6.82×10^{-4}
rs1378892	15q21.2	SCG3	1.05	1.24×10^{-3}
rs148410779	9q31.3, *LOC107987116*	PTGR1	0.93	1.89×10^{-3}

[a] SNPs are in strong LD with other pQTL SNPs in CEU population. Only a representative SNP was presented.

Table 3. Associations between genetically predicted concentrations of circulating proteins and breast cancer risk differed by ER status: analysis using genetic instruments

	Estrogen receptor +		Estrogen receptor -		
Protein	OR (95% CI)	*P*	OR (95% CI)	*P*	P_{het}
TMPRSS11D	1.06 (1.03-1.09)	1.41×10^{-4}	1.13 (1.08-1.19)	4.00×10^{-7}	0.026
DOCK9	1.11 (1.05-1.17)	1.41×10^{-4}	1.25 (1.14-1.36)	4.00×10^{-7}	0.024
TNS2	1.09 (1.04-1.13)	1.41×10^{-4}	1.19 (1.11-1.27)	4.00×10^{-7}	0.03
CRYBB2	1.17 (1.08-1.27)	1.41×10^{-4}	1.39 (1.22-1.58)	4.00×10^{-7}	0.027
STOM	1.06 (1.03-1.10)	6.11×10^{-4}	1.14 (1.08-1.20)	1.50×10^{-6}	0.022
ADH1B	1.07 (1.03-1.11)	6.10×10^{-4}	1.16 (1.09-1.23)	1.60×10^{-6}	0.026
PRDM1	1.03 (1.01-1.05)	0.013	1.08 (1.05-1.12)	6.20×10^{-6}	0.014

Table 4. Validation of primary results with instruments constructed using pQTL variants from independent studies

Protein	Protein-associated SNPs	OR (95% CI)	*P* value	Consistent in direction
SELP	rs10800462, rs651007	0.94 (0.92-0.97)	1.89×10^{-4}	Yes
CPNE1	rs2425143	0.97 (0.95-0.98)	4.10×10^{-5}	Yes
RELT	rs7119167	1.06 (1.02-1.10)	3.91×10^{-3}	Yes
VEGFR3/FLT4	rs10935480, rs651007	0.97 (0.96-0.99)	5.50×10^{-4}	Yes
VEGFR2/KDR	rs34231037, rs651007	0.96 (0.94-0.99)	7.49×10^{-3}	Yes
TIE1 (soluble)	rs10935480, rs8176749	0.97 (0.95-0.99)	6.57×10^{-4}	Yes
BCAM	rs8176749	0.93 (0.89-0.97)	7.49×10^{-4}	Yes
AKR1A1	rs6662572	0.97 (0.93-1.01)	1.72×10^{-1}	Yes
JAG1	rs4588, rs651007	0.95 (0.92-0.98)	1.07×10^{-3}	Yes
IR	rs651007	0.95 (0.93-0.97)	8.30×10^{-7}	Yes
MET	rs35349146, rs651007	0.93 (0.90-0.96)	7.20×10^{-6}	Yes
SELE (soluble)	rs651007	0.96 (0.95-0.98)	8.30×10^{-7}	Yes
ENG	rs651007	0.93 (0.91-0.96)	8.30×10^{-7}	Yes
ICAM2 (soluble)	rs651007	0.91 (0.88-0.95)	8.30×10^{-7}	Yes
CHST15	rs651007	0.90 (0.87-0.94)	8.30×10^{-7}	Yes
TFPI	rs6027	0.97 (0.92-1.01)	1.76×10^{-1}	No
CD36	rs651007	0.93 (0.91-0.96)	8.30×10^{-7}	Yes
NOTCH1	rs8176749	0.92 (0.87-0.96)	7.49×10^{-4}	Yes
CDH5	rs8176749	0.97 (0.95-0.99)	7.49×10^{-4}	Yes
ALPI	rs4942471	1.07 (0.99-1.15)	9.51×10^{-2}	No
B3GNT2	rs1800470, rs492488	1.04 (0.99-1.08)	1.05×10^{-1}	Yes
C1GALT1C1	rs579459	1.12 (1.07-1.17)	7.40×10^{-7}	Yes
CTSF	rs607736	1.18 (1.10-1.27)	1.30×10^{-5}	Yes
FAM177A1	rs492488, rs799498	0.99 (0.96-1.01)	3.18×10^{-1}	-
GOLM1	rs492488, rs7854118	1.03 (1.00-1.06)	4.58×10^{-2}	Yes
ISLR2	rs579459, rs923118	0.94 (0.91-0.97)	1.42×10^{-4}	Yes
KIN	rs579459, rs7412	0.95 (0.92-0.97)	1.90×10^{-4}	Yes
MAN1A2	rs1289863	0.95 (0.89-1.02)	1.41×10^{-1}	No
PRDM1	rs9852529	1.08 (1.04-1.13)	7.60×10^{-5}	Yes
QSOX2	rs12378344, rs492488	1.03 (1.01-1.05)	2.22×10^{-3}	Yes
RSPO3	rs3734626	0.94 (0.91-0.98)	1.51×10^{-3}	Yes
SCG3	rs1456297	1.05 (1.02-1.09)	1.08×10^{-3}	Yes
SEC13	rs62295996	0.99 (0.92-1.07)	8.45×10^{-1}	-

SULF2	rs7485577	0.97 (0.91-1.03)	3.43×10^{-1}	-
THSD1	rs704	0.97 (0.92-1.02)	1.90×10^{-1}	Yes
TMPRSS11D	rs9852529	1.09 (1.04-1.13)	7.60×10^{-5}	Yes
TPST2	rs9608491	1.00 (0.92-1.08)	9.10×10^{-1}	-

Table 5. Protein biomarkers replicated using data from the UK Biobank: associations of genetically predicted concentrations of circulating proteins with breast cancer risk

Protein	Protein-associated SNPs	Histologically confirmed breast cancer		Self-reported breast cancer	
		Association direction	*P* values	Association direction	*P* values
B3GNT2	rs2519093, rs2231940 [a]	+	0.003	+	0.007
HTN1	rs1063945 [a]	+	0.003	+	5.67×10^{-4}
CPNE1	rs62143206, rs12481228	-	0.007	-	0.044
VCAM1	rs3184504	-	0.010	-	0.063
RSPO3	rs2489623	-	0.014	-	0.240
CHST15	rs550057 [a]	-	0.015	-	0.006
KIN	rs7412, rs62143198, rs550057 [a]	-	0.016	-	0.011
JAG1	rs7041, rs550057 [a]	-	0.017	-	0.068
ALPI	rs679574, rs550057 [a]	-	0.025	-	0.068
FAM177A1	rs679574, rs550057 [a]	-	0.029	-	0.086
IGFIR (soluble), MET, ICAM2 (soluble), LIFR (soluble), ENG, FAM20B	rs635634 [a]	-/+ [b]	0.040	-/+ [b]	0.012
IR	rs507666	-	0.042	-	0.012
MAN1A2	rs532436 [a]	+	0.044	+	0.015
SELE (soluble), IL3RA	rs2519093	-	0.046	-	0.012
C1GALT1C1	rs7787942, rs2519093	+	0.048	+	0.029
GOLM1	rs601338, rs550057 [a]	+	0.066	+	0.009
AKR1A1	rs72688441, rs62143198	-	0.077	-	0.061
VEGFR2/KDR	rs34231037, rs635634	-	0.084	-	0.053
QSOX2	rs550057, rs10858248	+	0.197	+	0.055
ISLR2	rs4055121, rs2959011, rs532436	-	0.207	-	0.088

[a] Proxy SNPs were used ($R^2 > 0.9$ in the 1000 Genome Project CEU population)

[b] Inverse associations (-) for IGF1R, MET, ICAM2, LIFR, ENG; positive association (+) for FAM20B.

Funding

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This work was supported by the U.S. National Institutes of Health [R01CA158473, R01CA148677], as well as funds from the Anne Potter Wilson endowment. Genotyping of the OncoArray was principally funded by three sources: the PERSPECTIVE project, funded from the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie, de la Science et de l'Innovation du Québec through Genome Québec, and the Quebec Breast Cancer Foundation; the NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) initiative and Discovery, Biology and Risk of Inherited Variants in Breast Cancer (DRIVE) project [NIH Grants U19 CA148065, X01HG007492]; and Cancer Research UK [C1287/A10118, C1287/A16563]. The BCAC is funded by Cancer Research UK [C1287/A16563], the European Community's Seventh Framework Programme under grant agreement 223175 [HEALTH-F2-2009-223175] (COGS).

Competing Interests: None declared

Author Contributions

Study concept and design: Wei Zheng

Acquisition of data: Xiao-ou Shu, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang, Joe Dennis, Irene L. Andrulis, Jose E. Castelao, Thilo Dörk, Manuela Gago-Dominguez, Montserrat García-Closas, Graham G. Giles, Artitaya Lophatananon, Kenneth Muir, Håkan Olsson, Gadi Rennert, Emmanouil Saloustros, Rodney J. Scott, Melissa C. Southey, Paul D.P. Pharoah, Roger L. Milne, Peter Kraft, Jacques Simard, Douglas F. Easton, Wei Zheng

Statistical analysis: Xiang Shu, Jiandong Bao

Interpretation of data: Xiang Shu, Xiao-Ou Shu, Wei Zheng

Drafting of the manuscript: Xiang Shu, Wei Zheng

Critical revision of the manuscript for important intellectual content: Lang Wu, Jirong Long, Xingyi Guo, Yaohua Yang, Xiao-ou Shu, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang, Joe Dennis, Irene L. Andrulis, Jose E. Castelao, Thilo Dörk, Manuela Gago-Dominguez, Montserrat García-Closas, Graham G. Giles, Artitaya Lophatananon, Kenneth Muir, Håkan Olsson, Gadi

Rennert, Emmanouil Saloustros, Rodney J. Scott, Melissa C. Southey, Paul D.P. Pharoah, Roger L. Milne, Peter Kraft, Jacques Simard, Douglas F. Easton

Study supervision: Wei Zheng

IJC-18-2429.R2

Reliable biomarkers for breast cancer are critically needed, but results from existing studies have been inconsistent. Here the authors combined genomics and proteomics expertise and identified 56 circulating proteins, for which genetically predicted levels were associated with breast cancer risk. These proteins are involved in relevant biological processes such as estrogen receptor signaling and insulin resistance and will serve as candidates for further evaluative investigations.

Author/s:
Shu, X;Bao, J;Wu, L;Long, J;Shu, X-O;Guo, X;Yang, Y;Michailidou, K;Bolla, MK;Wang, Q;Dennis, J;Andrulis, IL;Castelao, JE;Doerk, T;Gago-Dominguez, M;Garcia-Closas, M;Giles, GG;Lophatananon, A;Muir, K;Olsson, H;Rennert, G;Saloustros, E;Scott, RJ;Southey, MC;Pharoah, PDP;Milne, RL;Kraft, P;Simard, J;Easton, DF;Zheng, W

Title:
Evaluation of associations between genetically predicted circulating protein biomarkers and breast cancer risk

Date:
2020-04-15

Citation:
Shu, X., Bao, J., Wu, L., Long, J., Shu, X. -O., Guo, X., Yang, Y., Michailidou, K., Bolla, M. K., Wang, Q., Dennis, J., Andrulis, I. L., Castelao, J. E., Doerk, T., Gago-Dominguez, M., Garcia-Closas, M., Giles, G. G., Lophatananon, A., Muir, K. ,... Zheng, W. (2020). Evaluation of associations between genetically predicted circulating protein biomarkers and breast cancer risk. INTERNATIONAL JOURNAL OF CANCER, 146 (8), pp.2130-2138. https://doi.org/10.1002/ijc.32542.