ARTICLE OPEN

Check for updates

Epidemiology

The impact of circulating protein levels identified by affinity proteomics on short-term, overall breast cancer risk

Felix Grassmann[1,2✉], Anders Mälarstig[1,3], Leo Dahl[4], Annika Bendes[4], Matilda Dale[4], Cecilia Engel Thomas[4], Marike Gabrielsson[1], Åsa K. Hedman[1,3], Mikael Eriksson[1], Sara Margolin[5,6], Tzu-Hsuan Huang[7], Mikael Ulfstedt[8], Simon Forsberg[8], Per Eriksson[8], Mattias Johansson[9], Per Hall[1,5], Jochen M. Schwenk[4] and Kamila Czene[1]

OBJECTIVE: Current breast cancer risk prediction scores and algorithms can potentially be further improved by including molecular markers. To this end, we studied the association of circulating plasma proteins using Proximity Extension Assay (PEA) with incident breast cancer risk.
SUBJECTS: In this study, we included 1577 women participating in the prospective KARMA mammographic screening cohort.
RESULTS: In a targeted panel of 164 proteins, we found 8 candidates nominally significantly associated with short-term breast cancer risk ($P < 0.05$). Similarly, in an exploratory panel consisting of 2204 proteins, 115 were found nominally significantly associated ($P < 0.05$). However, none of the identified protein levels remained significant after adjustment for multiple testing. This lack of statistically significant findings was not due to limited power, but attributable to the small effect sizes observed even for nominally significant proteins. Similarly, adding plasma protein levels to established risk factors did not improve breast cancer risk prediction accuracy.
CONCLUSIONS: Our results indicate that the levels of the studied plasma proteins captured by the PEA method are unlikely to offer additional benefits for risk prediction of short-term overall breast cancer risk but could provide interesting insights into the biological basis of breast cancer in the future.

British Journal of Cancer (2024) 130:620–627; https://doi.org/10.1038/s41416-023-02541-2

INTRODUCTION

Breast cancer is the most common cancer in women worldwide, with incidence rates still increasing in Western countries. While recent advances in therapy have increased the odds of survival after a breast cancer diagnosis, early detection of aggressive breast cancer is paramount to further improve health in our aging population. Current mammographic screening programmes have a number needed to screen around 1000–2000 [1, 2], indicating many women have to be screened every 2 years for 10 years to save a single life. Thus, our current screening programmes need to be improved by better detection of women at risk of developing invasive breast cancer, particularly within the next screening interval.

Traditional risk prediction algorithms are mainly based on reproductive risk factors, genetic risk factors such as aggregate genetic risk scores, family history and lifestyle factors. More recent efforts to identify women with a high short-term or long-term risk for breast cancers used clinical models that additionally included features from mammographic images such as breast density or the presence of microcalcifications [3, 4]. Those models have shown high discriminatory performance compared to traditional risk models and are now suitable for identifying individuals at high risk for breast cancer. Nevertheless, the sensitivity and specificity of the models can potentially be further improved by identifying additional (independent) risk factors.

To this end, there are several approaches to identifying novel (molecular) markers for breast cancer risk. Current large-scale efforts focus on genome-wide scans to identify genetic factors that influence the overall and/or subtype-specific breast cancer risk [5–8]. In addition, other molecular markers such as DNA modifications [9], circulating metabolites [10], and cell-free DNA/RNA [11], as well as proteins [12, 13] are being studied. Apart from inherited genetic markers, only a few other biomarkers have been successfully validated in independent studies [14]. In addition, many past and currently underway studies suffer from several limitations, such as small sample size and lack of available incident cases not confounded by treatment [15].

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [2]Institute for Clinical Research and Systems Medicine, Health and Medical University, Potsdam, Germany. [3]Pfizer Worldwide Research, Development and Medical, Stockholm, Sweden. [4]Science for Life Laboratory, Department of Protein Science, KTH Royal Institute of Technology, Solna, Sweden. [5]Department of Oncology, Södersjukhuset, Stockholm, Sweden. [6]Department of Clinical Science and Education Södersjukhuset, Karolinska Institutet, Stockholm, Sweden. [7]Cancer Immunology Discovery, Pfizer Inc., San Diego, CA, USA. [8]Olink Proteomics, Uppsala Science Park, Uppsala, Sweden. [9]Genomic Epidemiology Branch, International Agency for Research on Cancer (IARC/WHO), Lyon, France. ✉email: felix.grassmann@ki.se

Received: 12 June 2023 Revised: 22 November 2023 Accepted: 1 December 2023
Published online: 22 December 2023

Table 1. Summary statistics of included KARMA participants at baseline exam.

	Stockholm		Skåne	
Variable	**Controls**	**BC cases**	**Controls**	**BC cases**
Number of individuals	410	405	371	391
Mean age (SD) [years], matched*	58.48 (9.56)	58.52 (9.60)	58.66 (9.82)	59.05 (9.66)
Mean body mass index (SD) [kg/m^2]	25.16 (4.19)	25.61 (4.20)	25.33 (4.25)	25.76 (4.13)
Postmenopausal [%]	70.49	69.38	70.35	72.38
Ever smoked [%]	58.00	59.17	50.70	55.59
Lipid medication taken [%]	11.46	11.85	12.40	15.60
Hypertensive medication taken [%]	27.07	22.72	28.30	26.09
Heart medication taken [%]	11.71	10.86	11.86	9.46
Renal failure (prevalent) [%]	0.73	0.25	0.54	0.00
Mean age of plasma [years]	7.92 (0.65)	7.53 (0.71)	8.02 (0.70)	7.61 (0.76)
Average frequency of proteins below LOD (SD)	0.12 (0.02)	0.12 (0.02)	0.14 (0.02)	0.14 (0.02)
Mean 313 SNP Genetic Risk Score (SD)	−0.08 (0.29)	−0.02 (0.30)	−0.07 (0.32)	−0.05 (0.31)

LOD level of detection.
*Variable used for median matching cases and controls.

In this study, we present the results from the KARMA cohort, the largest prospective breast cancer screening cohort in Sweden. We measured plasma protein levels in an exploratory and a targeted panel and analysed their association with incident breast cancer to identify novel markers for breast cancer risk.

METHODS

Study population

Women aged 40–74 years are invited every 18–24 months to the national screening programme in Sweden. Women attending the mammographic screening in two regions (Stockholm and Skåne) in Sweden were invited to participate in the KARMA study between 2011 and 2013. A total of 70,877 women gave informed consent to participate in the KARMA study [16]. Participants answered a comprehensive web-based questionnaire, donated blood, and accepted linkage to national registers. From linkage to the cancer registry, we identified 826 women diagnosed with breast cancer which occurred within 3 years of blood draw between 2012 and 2015. From those, only 804 had plasma specimens and thus were used in our study. We used the *matchit* function from the *MatchIt* library implemented in R to match 804 controls from KARMA study to the incident cases by randomly drawing women without incident breast cancer so that the median age at blood draw in cases and controls was similar (median matching).

We used self-reported questionnaire data to create dichotomous variables for menopausal and smoking status. Family history of first-degree relatives was assessed from the multi-generation registry, as previously described [17]. BMI and age were assessed at the time of study entry and thus at the time of qualifying blood draw. Linkage to the prescription registry was used to determine whether women had taken a lipid medication (ATC code C10) between 2005 and blood draw.

Tumour characteristics such as oestrogen receptor (ER) status, human epidermal growth factor receptor 2 (HER2) status, grade and lymph node involvement were retrieved from medical records or from the Swedish National Cancer Registry. Mode of detection was defined by the timing between the last mammographic screening and time of diagnosis [18, 19]. Briefly, women diagnosed between two scheduled screening intervals without a detectable tumour in the previous screening were deemed to have interval breast cancer (IC). Conversely, women diagnosed with breast cancer at a regularly scheduled mammogram are considered screen-detected (SDC). Patients who did not attend screening or missed their scheduled screening prior to diagnosis were not considered in the analysis of IC vs. SDC (Supplementary Table 1).

Protein measurements—targeted panel

The samples from the Karma cohort were distributed across 96-well plates with samples from the same individual placed on the same plate and the remaining samples randomly distributed. Samples from Skåne and Stockholm were placed on separate plates. Proximity Extension Assay was performed at SciLifeLab's Affinity Proteomics Unit in Stockholm according to instructions from Olink Proteomics AB (Uppsala, Sweden) [20] to measure proteins in EDTA plasma using the Cardiometabolic (v.3603, Lot No A94923) and Immuno-Oncology (v.3111, Lot No B01401) panels. For the cardiometabolic panel, plasma samples were diluted 1:2025, and for the immune-oncology panel 1:1 (undiluted). Normalised protein expression (NPX) values were obtained from the Olink NPX Manager software (version 2.2.1.311) after normalisation using the "Intensity normalisation v2" method to account for the various measurement batches [21]. In addition to standard quality control measures, we removed proteins that had missing values in more than 10% of the samples, either in the Skåne or the Stockholm recruitment centre. After quality control, 163 high-quality protein measurements from the Cardiometabolic and Immuno-Oncology panel were available for 796 incident cases and 781 controls (Table 1) from both cohorts. As an additional quality control maker, we also computed the percentage of proteins that were below the level of detection (LOD) in each participant and recorded the duration the plasma was stored at −80 °C (age of plasma). To account for differences between the protein levels by recruitment centre, we used a rank-based inverse normal transformation on each protein in both cohorts separately with the *qnorm* function in R. From this normalised data, we used the *prcomp* function in R to compute the first ten principal components (PCs) to capture additional underlying data structures represented by those PCs.

Protein measurements—exploratory panel

Similar to the approach for the targeted panel of proteins, we also measured over 3000 proteins from eight different panels with a Proximity Extension Assay (Olink Proteomics AB, Uppsala, Sweden) in a subset of individuals from the Skåne cohort. The raw protein measurements were analysed with the Olink NPX Manager software as described above to yield normalised protein expression values. Proteins with more than 50% missing values were removed from analyses as were those flagged with a warning or error from the NPX Manager software. The less stringent cut-off for the exclusion of proteins was chosen since the exploratory panel contains many proteins only present in minute concentrations and thus can often be below level of detection across the cohort. Furthermore, individuals that were flagged as outliers by principal component analyses were also excluded, yielding a final analytical dataset consisting of 2204 proteins in 303 BC cases and 294 controls (for more details, see ref. [22]). Similar to the small panel, we also computed the first principal components from the protein data and used those as additional exposures in our association analyses.

Breast cancer genetic risk score

All cases and controls were genotyped on the OncoArray genotyping platform and passed standard quality control, as previously described [18].

Briefly, we excluded related individuals, individuals with excessive missingness (> 3% missing sites before variant QC) or heterozygosity as well as individuals not of European descent. In addition, variants with a strong deviation from Hardy–Weinberg equilibrium ($P < 0.00001$) or with a high degree of missingness (missing in more than 10% of individuals) were also excluded. For more details, please see refs. [18, 22].

Breast cancer risk was quantified by a genetic (polygenic) risk score from 313 variants associated with breast cancer risk as previously described with *plink2* [7]. Briefly, the genotype at each variant (coded as the number of risk-increasing alleles 0, 1 or 2) was multiplied with the respective log odds ratio indicating the strength of association with breast cancer risk. Then, for each individual, the weighted alleles were summed up over all variants, yielding a single score for each individual. Higher genetic risk scores indicate a higher genetic risk for breast cancer from common variants, while lower scores indicate lower risk.

Statistical analysis and presentation

Unsupervised clustering of the raw NPX values by their Euclidean distance was performed with the *heatmap.2* function from the *gplot* package [23]. All association analyses were carried out with Cox proportional hazard models, as implemented in the *survival* package in R [24]. We adjusted the models for known confounders of protein levels or breast cancer risk. In particular, the models were adjusted for age at blood draw, BMI, lipid and heart medication, renal failure, smoking status, menopause status, family history of breast cancer, breast density, age of plasma (i.e., duration of storage), number of proteins below LOD, the 313 variant genetic risk score and, where appropriate, recruitment centre. The results of the association analyses were plotted as Manhattan plots with the *ggplot* function implemented in the *ggplot2* library [25] or as correlation plots with the *corrplot* function implemented in the *corrplot* library [26].

RESULTS

Quality control and unsupervised clustering

After quality control, a total number of 163 proteins from the Olink Proximity Extension Assay (PEA) Cardiometabolic and Immuno-Oncology panels (targeted panel) were available for analysis in 796 incident cases and 781 controls, which were selected from two cohorts recruited in Stockholm and in the Skåne region in Sweden, respectively (Table 1 and Supplementary Table 1). Baseline characteristics ascertained at blood draw (Table 1) as well as tumour characteristics (Supplementary Table 1) were similar among both cohorts. First, we performed an unsupervised clustering approach and observed that individuals were broadly grouped according to the recruitment centre in which they were recruited (Fig. 1a). To account for this, we normalised protein levels within each region by computing a rank-based inverse normal transformation. This effectively removed the systematic difference in protein levels observed by the recruitment centre (Fig. 1b). The remaining clusters were not representative of other covariates nor of disease status (Fig. 1b and Table 1).

Association of proteins with breast cancer risk in the targeted and exploratory panel

Next, we computed the association of normalised plasma protein levels in the targeted panel with breast cancer risk using Cox regression while adjusting for potential confounders such as age at blood draw, BMI, smoking status, renal failure, family history of breast cancer, the 313 simple nucleotide polymorphism (SNP) genetic risk score for breast cancer and medication. In the targeted panel, we found 8 proteins nominally associated with breast cancer risk ($P < 0.05$), of which two proteins showed a positive and six a negative effect size (Fig. 2). Notably, among the significantly associated proteins was Caspase 8 (CASP8), which has previously been implicated in breast cancer genetic risk as well as other cancers. However, after adjustment for multiple testing (either by controlling the false discovery rate, FDR, at FDR < 0.05 or Bonferroni correction), none of the proteins remained statistically significant (Fig. 2). We did not find a significant association of the principal components computed from the protein data with breast cancer after adjustment for multiple testing (Fig. 2).

Considering the complexity and scale of the circulating proteome, it remained a possibility that none of the 163 proteins analysed on the targeted Olink Cardiometabolic and Immune-oncology panels represented proteins of relevance for breast cancer risk. Therefore, we measured a total of 2950 plasma proteins using the Olink Explore I and II panels in a subset of the Skåne study only to avoid confounding the analysis by the recruitment centre (Supplementary Table 2). After quality control, 2204 proteins remained for statistical analysis in 303 incident breast cancer cases and 294 controls. We performed the association analyses in those sample with the same adjustments as before. Here, 115 proteins were nominally associated with breast cancer risk (Fig. 3). However, none of the 115 proteins nor the first ten principal components survived Bonferroni multiple testing correction ($P < 0.05/2{,}204$, Fig. 3) nor were they significant at FDR < 0.05.

To gain further insights into the nominally significantly associated proteins ($P < 0.05$) from the targeted panel, we investigated the association of those proteins with breast cancer risk after stratifying the patients according to their tumour characteristics (Supplementary Table 1). In general, the proteins are associated similarly in women with different prognostic markers (Supplementary Fig. 1). For particular markers, we see that it is associated more pronounced with more aggressive prognostic markers. Indeed, CXCL13 is statistically significantly more strongly associated with interval compared to screen-detected cancer in a case-only analysis ($P = 0.005$). In addition, CASP8 seems to be more strongly associated with less favourable markers, although it is not significantly different in a case-only design comparing unfavourable against favourable markers (i.e., BC cases with ER-negative, lymph node-positive or high-grade tumours compared to women with ER-positive, lymph node-negative and low-grade tumours).

Lasso regression

While we did not find that proteins were significantly associated with breast cancer risk individually, they could still exhibit combinatorial effects that would help to predict a future breast cancer diagnosis. To this end, we used Lasso regression to detect whether a combination of proteins in addition to established risk factors would improve risk prediction accuracy, as assessed by the area under the receiver operating curve (AUC, Fig. 4a, c). In this analysis, we found that the baseline model containing age at blood draw, BMI, percent mammographic density, menopause status, the 313 BC genetic risk score and family history of breast cancer outperformed models with additional protein level measurements in the targeted (Fig. 4a) and in the exploratory panel (Fig. 4c). This finding strongly indicates that the studied proteins are not useful for risk prediction of overall, short-term breast cancer risk.

Power considerations

Next, we studied whether the lack of association with breast cancer risk could potentially be attributed to a lack of power. Although this study is the largest prospective study investigating the association of proteins with incident breast cancer by a large margin, we could still be underpowered to detect such associations. Therefore, we generated random subsets of our data and artificially inflated the effect sizes of the association of each protein with breast cancer. This allowed us to compute the post hoc power of our study to identify specific proteins with a significant *P* value after accounting for multiple testing (i.e., Bonferroni corrected *P* value < 0.05). We found that we had more than adequate power (> 80%) to detect proteins that were associated with breast cancer risk with a hazard ratio greater than 1.28 per standard deviation (SD, log hazard ratio per SD >0.25, Fig. 4b) in the targeted panel, which is more than adequate to identify risk factors with effect sizes observed for benign breast disease or elevated breast density. Similarly, even though we only measured the exploratory panel in less than half of

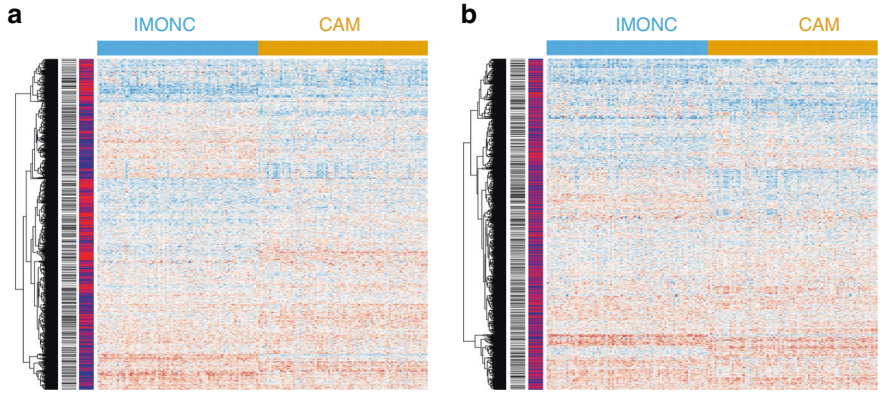

Fig. 1 Unsupervised clustering of patients and controls according to protein levels. a Cases and controls are clustered according to recruitment centre (Skåne in blue and Stockholm in red). **b** After quantile normalisation of protein levels of participants from each centre, no striking differences between centres were obvious. Future disease status is indicated by black (BC) and white (control) bars. IMONC Immuno-Oncology (v.3111) panel, CAM Cardiometabolic (v.3603) panel.

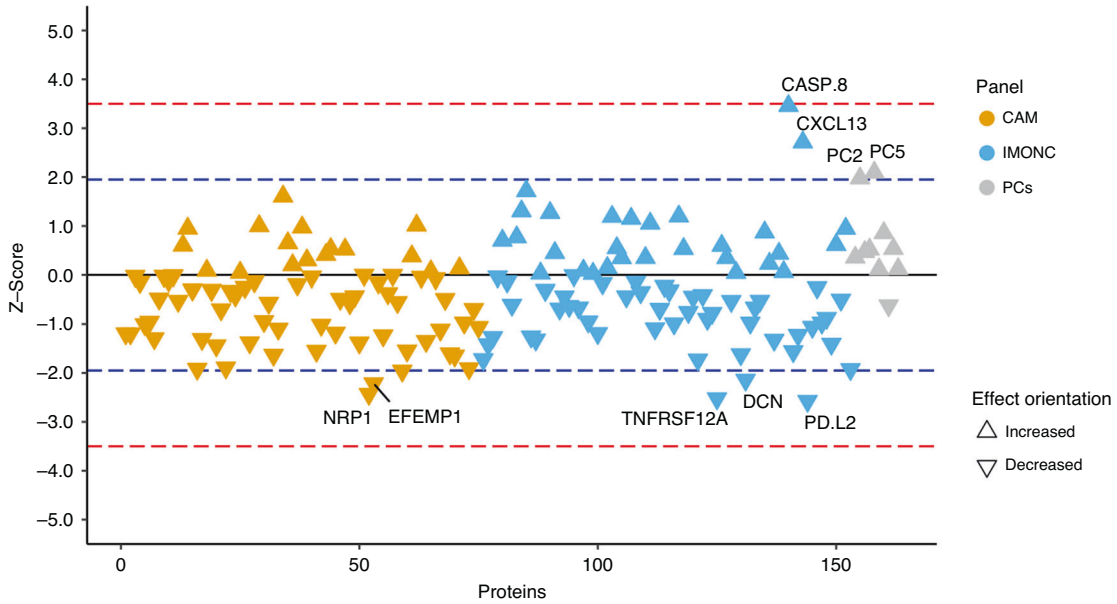

Fig. 2 Association of 163 proteins in the targeted panel with incident breast cancer. We used Cox proportional hazard models to investigate the impact of cardiometabolic (CAM) and immune-oncology (IMONC)-related proteins on the risk to develop breast cancer within 3 years. Triangles pointing up indicate that increased protein levels result in increased risk for breast cancer and vice versa. Significantly associated proteins ($P < 0.05$) are shown above or below the blue dotted line and are labelled. No proteins were found to be significantly associated with incident breast cancer risk after accounting for multiple testing (i.e., $P < 0.05/163$), as indicated by the red dotted line.

our cohort, we still had sufficient power to detect 80% of all associations with a hazard ratio above 1.74 (per SD, Fig. 4d), which corresponds to effect sizes observed for proteins associated with oesophageal carcinoma, or established risk factors such as hormone replacement therapy and family history. These results mean that we would have ample power to detect associations previously reported for other cancers. However, the effect sizes we observed for nominally significantly associated proteins ($P < 0.05$) was 0.10 and 0.18 on average for the targeted and the exploratory panel, respectively. Those findings show that none of the investigated proteins are likely associated with overall breast cancer risk with large and thus relevant effect sizes as observed in previous studies for other cancers.

DISCUSSION

In this study, we present results from the largest study to date about circulating proteins involved in breast cancer risk. We found that neither the pre-selected proteins in the targeted panel nor proteins from the extended Olink panel were significantly associated with short-term general breast cancer risk. The lack of associations was not due to low power but attributable to the small effect sizes observed for even nominally significant proteins. Based on our results, we cannot exclude a role of plasma proteins in breast cancer risk but their impact on risk is likely to be low and thus they are of little value to improve risk prediction efforts.

Interestingly, we found that CASP8 was nominally associated with a short-term risk for breast cancer providing insights into a potential role of plasma CASP8 levels in breast cancer risk. Several in vitro studies have shown that CASP8 is involved in apoptosis and necroptosis [27] in different cell types. Inherited genetic variations in the *CASP8* gene have also previously been found to be associated with breast cancer [28] as well as other types of cancers [29, 30]. Importantly, circulating plasma levels of this protein seem to influence the risk for prostate [31] and oesophageal squamous cell carcinoma (ESCC) [32] as well as type

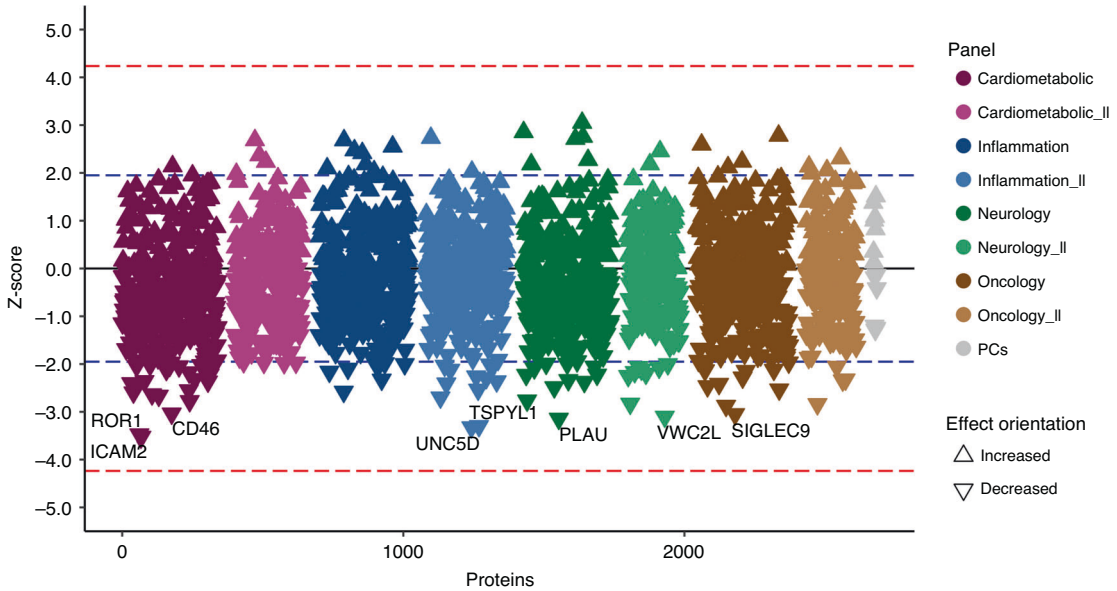

Fig. 3 Association of 2,204 proteins in the exploratory panel with incident breast cancer. We used Cox proportional hazard models to investigate the impact of over 2204 proteins on the risk to develop breast cancer. Triangles pointing up indicate that increased protein levels result in increased risk for breast cancer, while triangles pointing down signify lower risk. Significantly associated proteins ($P < 0.05$) are shown above or below the blue dotted line. No proteins were found to be significantly associated with incident breast cancer risk after accounting for multiple testing (i.e., $P < 0.05/2204$), as indicated by the red dotted line.

2 diabetes [33] and coronary events [34]. Given its role in diverse phenotypes and involvement in multiple pathways leading to apoptosis, the observed association with breast cancer is still not easily understood. Thus, additional studies designed to provide mechanistic insights into the precise role of CASP8 or related genes and downstream targets in breast cancer risk are needed. Similar to genome-wide association studies, which generally identify variants weakly associated with disease risk, our results are encouraging that high-throughput protein panels can potentially identify important proteins and thus pathways in breast cancer even if individual proteins only show weak associations.

Previous studies using Olink protein panels have often found stronger effect sizes than those we observed in our cohort for nominally significant proteins [31, 32, 35–37]. This could be attributed to multiple factors, such as smaller sample size in preceding studies or the winner's curse effect, both of which would strongly bias and inflate the estimates, as much as twice their true value [38]. However, even though those previous studies reported exaggerated effect sizes, we would still have enough power to pick up even signals with smaller impact. Conversely, the effect sizes we observed in our cohort are much smaller than most established risk factors, indicating that protein levels are not strongly associated with general breast cancer risk. This is in agreement with the results from the lasso regression, which showed that the addition of proteins does not improve risk prediction accuracy beyond a baseline model containing established hormonal, reproductive, lifestyle, and family history/genetic risk markers. Thus, the proteins we studied are not only weakly associated with breast cancer risk, but they appear unlikely to be informative for future efforts to predict overall breast cancer risk.

Our results do not preclude that proteins might be useful in predicting certain types of breast cancer, either defined by its aggressiveness (i.e., grade and lymph node involvement) or by cell surface markers (such as oestrogen receptor or human epidermal growth factor receptor 2 (HER2) status) [39]. Indeed, we found that the association strength can differ by breast cancer subtype (see Supplementary Fig. 1), although we only observed a statistically significant difference in the association signals for CXCL13 with interval cancer vs. screen-detected cancer. Therefore, larger studies of incident breast cancer cases would be necessary to include a sufficient number of cases with rarer but highly relevant characteristics, such as triple-negative tumours and those that spread beyond the breast. In addition, a combination of proteins and established or yet unknown risk factors could help to predict breast cancer risk in an individualised fashion for certain subtypes of breast cancer. Such efforts, however, require even more extensive studies and, importantly, independent replication in incident breast cancer cases. Such efforts could be possible by including data from individuals recruited in the UK Biobank, which is currently measuring the 3000 proteins in >50,000 randomly selected individuals [40]. Given the incidence of breast cancer in the UK Biobank of around 700 women per year since recruitment, this would translate to around 140 incident breast cancer patients with such protein measurements within 2 years since recruitment. Hence, our study is still at least twice as large and therefore more powered to detect associations with short-term risk, which would be most useful for risk prediction in current screening programmes.

Many of the circulating proteins targeted by the PEA approach originate from organs involved in metabolic or inflammatory processes predominantly due to active secretion [41]. In addition, the proteins can also be potentially contained in diverse extracellular vesicles (EVs) to facilitate intercellular communication, immune response, blood coagulation, and tissue repair [42]. Alternatively, the proteins can originate from an (undetected) tumour which either actively or passively secretes those proteins. Our approach, however, is not suited to easily distinguish between those processes. When we stratified our breast cancer patients by time between blood draw and diagnosis according to the median (in this case, 19 months), we found similar associations in both groups, indicating that leakage of proteins from the tumour is unlikely to be the main driver of the blood proteome in breast cancer captured by our PEA approach. Thus, detecting cancer risk-related proteins originating from the tumour, remains a known challenge if these are not enhanced by systemic involvement of metabolic or inflammatory processes. Compared to classical proteomics platforms using mass spectrometry, the chosen

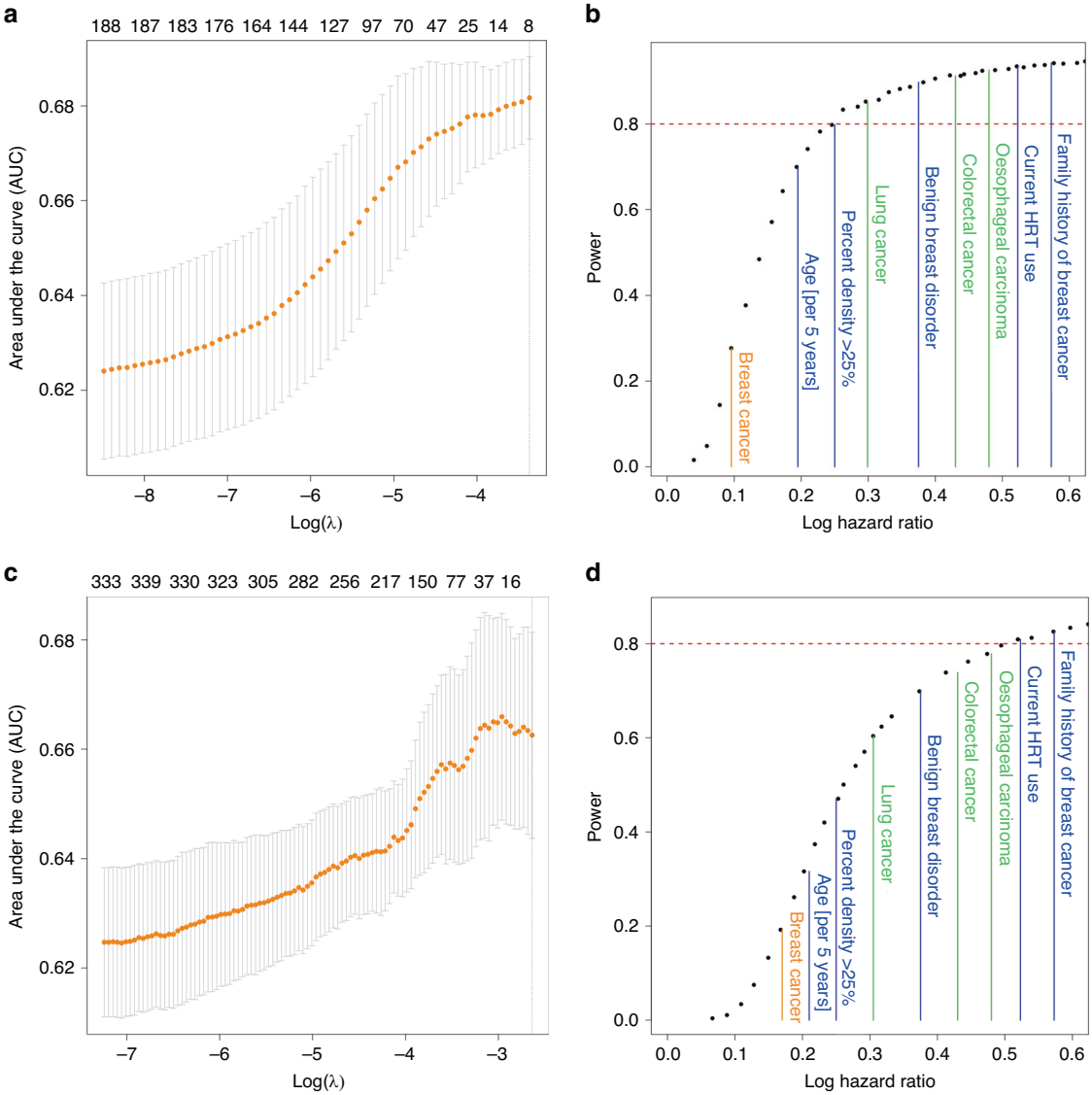

Fig. 4 Lasso regression and power analysis of the targeted and exploratory panel. a (targeted), **c** (exploratory) We used lasso regression to quantify the potential impact of proteins on the accuracy to predict incident breast cancer. While accounting for (and thus not regularising) known breast cancer risk factors, we computed the area under the receiver operating curve (AUC) for each model using a fivefold cross-validation. The mean AUC as well as the standard error of the AUC estimate from the five cross validations is plotted against the penalty parameter λ and the number of proteins/parameters in the model (numbers at the top). Generally, adding proteins to the model did not improve prediction accuracy. **b** (targeted), **d** (exploratory) The power to detect proteins significantly associated with BC was estimated from generating random data with distributions similar to the observed data. By artificially increasing the effect size, we estimated at which effect size we would have had a power of 80% (red line) to detect significant effects (i.e., observed a Bonferroni corrected *P* value below 0.05). The effect sizes of known risk factors for breast cancer are indicated in blue. In green, we highlighted the average absolute effect sizes observed for nominally significant proteins ($P < 0.05$) for other cancers such as Oesophageal Squamous Cell Carcinoma [36], Colorectal Cancer [36] and Lung Cancer [37] from previous publications. The average absolute effect size observed for proteins with $P < 0.05$ in our dataset for Breast Cancer are shown in orange.

affinity-based PEA assay enables highly sensitive analysis of low-abundant blood proteins [43]. However, there remain relevant functional protein characteristics involved in governing human health [44] that were not resolved with this approach such as the presence of different proteoforms [45] and interacting proteins [46]. These protein traits may still harbour information relevant to breast cancer risk but remain out of reach when analysing systemic blood fluid.

In conclusion, our results indicate that the levels of the investigated proteins captured by a Proximity Extension Assay are unlikely to be informative to improve risk prediction of short-term breast cancer risk. Still, our study was designed to study overall breast cancer risk. Thus, a study focused on breast cancer subtypes (i.e., tumour characteristics or survival) could identify more specific associations which could prove useful for predicting certain types of breast cancer. Finally, despite low effect sizes observed in our dataset, we note that protein data can potentially be leveraged in future studies to gain important insights the biology underlying breast cancer, thus enabling the identification of novel preventive targets.

DATA AVAILABILITY

Access to phenotypes, biospecimens and genotypes from the KARMA study can be requested from https://karmastudy.org/contact/data-access/.

REFERENCES

1. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2013;CD001877. https://doi.org/10.1002/14651858.CD001877.pub5.
2. Keen JD, Keen JE. What is the point: will screening mammography save my life? BMC Med Inf Decis Mak. 2009;9:18. https://doi.org/10.1186/1472-6947-9-18.
3. Eriksson M, Czene K, Pawitan Y, Leifland K, Darabi H, Hall P. A clinical model for identifying the short-term risk of breast cancer. Breast Cancer Res BCR. 2017;19:29. https://doi.org/10.1186/s13058-017-0820-y.
4. Eriksson M, Czene K, Strand F, Zackrisson S, Lindholm P, Lång K, et al. Identification of women at high risk of breast cancer who need supplemental screening. Radiology. 2020;297:327–33. https://doi.org/10.1148/radiol.2020201620.
5. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45:353–61. https://doi.org/10.1038/ng.2563.
6. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, 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. https://doi.org/10.1038/ng.3242.
7. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet. 2019;104:21–34. https://doi.org/10.1016/j.ajhg.2018.11.002.
8. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet. 2020;52:572–81. https://doi.org/10.1038/s41588-020-0609-2.
9. Joo JE, Dowty JG, Milne RL, Wong EM, Dugué PA, English D, et al. Heritable DNA methylation marks associated with susceptibility to breast cancer. Nat Commun. 2018;9:867. https://doi.org/10.1038/s41467-018-03058-6.
10. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med. 2019;17:178. https://doi.org/10.1186/s12916-019-1408-4.
11. Page K, Martinson LJ, Fernandez-Garcia D, Hills A, Gleason KLT, Gray MC, et al. Circulating tumor DNA profiling from breast cancer screening through to metastatic disease. JCO Precis Oncol. 2021;1768–76.
12. Thomas CE, Dahl L, Byström S, Chen Y, Uhlén M, Mälarstig A, et al. Circulating proteins reveal prior use of menopausal hormonal therapy and increased risk of breast cancer. Transl Oncol. 2022;17:101339. https://doi.org/10.1016/j.tranon.2022.101339.
13. Veyssière H, Bidet Y, Penault-Llorca F, Radosevic-Robin N, Durando X. Circulating proteins as predictive and prognostic biomarkers in breast cancer. Clin Proteom. 2022;19:25. https://doi.org/10.1186/s12014-022-09362-0.
14. Nassar FJ, Chamandi G, Tfaily MA, Zgheib NK, Nasr R. Peripheral blood-based biopsy for breast cancer risk prediction and early detection. Front Med. 2020;7. https://doi.org/10.3389/fmed.2020.00028.
15. Fichtali K, Bititi A, Elghanmi A, Ghazi B. Serum lipidomic profiling in breast cancer to identify screening, diagnostic, and prognostic biomarkers. BioResearch Open Access. 2020;9:1–6. https://doi.org/10.1089/biores.2018.0022.
16. Gabrielson M, Eriksson M, Hammarström M, Borgquist S, Leifland K, Czene K, et al. Cohort profile: the Karolinska mammography project for risk prediction of breast cancer (KARMA). Int J Epidemiol. 2017;46:1740–1g. https://doi.org/10.1093/ije/dyw357.
17. Grassmann F, Yang H, Eriksson M, Azam S, Hall P, Czene K. Mammographic features are associated with cardiometabolic disease risk and mortality. Eur Heart J. 2021;42:3361–70. https://doi.org/10.1093/eurheartj/ehab502.
18. Grassmann F, He W, Eriksson M, Gabrielson M, Hall P, Czene K. Interval breast cancer is associated with other types of tumors. Nat Commun. 2019;10:4648. https://doi.org/10.1038/s41467-019-12652-1.
19. Ugalde-Morales E, Grassmann F, Humphreys K, Li J, Eriksson M, Tobin NP, et al. Interval breast cancer is associated with interferon immune response. Eur J Cancer. 2022;162:194–205. https://doi.org/10.1016/j.ejca.2021.12.003.
20. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Thorsen SB, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS ONE. 2014;9:e95192. https://doi.org/10.1371/journal.pone.0095192.
21. Olink Proteomics. Data normalization and standardization [white paper]. 2021. https://www.olink.com/application/data-normalization-and-standardization/.
22. Mälarstig A, Grassmann F, Dahl L, Dimitriou M, McLeod D, Gabrielson M, et al. Evaluation of circulating plasma proteins in breast cancer using Mendelian randomisation. Nat Commun. 2023;14:7680. https://doi.org/10.1038/s41467-023-43485-8.
23. Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: various R programming tools for plotting data.
24. Therneau TM. A package for survival analysis in S. 2015.
25. Wickham H. Ggplot2: elegant graphics for data analysis. New York: Springer New York; 2009.
26. Wei T, Simko V. R package "corrplot": visualization of a correlation matrix. 2021.
27. Fritsch M, Günther SD, Schwarzer R, Albert MC, Schorn F, Werthenbach JP, et al. Caspase-8 is the molecular switch for apoptosis, necroptosis and pyroptosis. Nature. 2019;575:683–7. https://doi.org/10.1038/s41586-019-1770-6.
28. Shephard ND, Abo R, Rigas SH, Frank B, Lin WY, Brock IW, et al. A breast cancer risk haplotype in the caspase-8 gene. Cancer Res. 2009;69:2724–8. https://doi.org/10.1158/0008-5472.CAN-08-4266.
29. Lubahn J, Berndt SI, Jin CH, Klim A, Luly J, Wu WS, et al. Association of CASP8 D302H polymorphism with reduced risk of aggressive prostate carcinoma. Prostate. 2010;70:646–53. https://doi.org/10.1002/pros.21098.
30. Bethke L, Sullivan K, Webb E, Murray A, Schoemaker M, Auvinen A, et al. The common D302H variant of CASP8 is associated with risk of glioma. Cancer Epidemiol Biomark Prev. 2008;17:987–9. https://doi.org/10.1158/1055-9965.EPI-07-2807.
31. Liu S, Garcia-Marques F, Zhang CA, Lee JJ, Nolley R, Shen M, et al. Discovery of CASP8 as a potential biomarker for high-risk prostate cancer through a high-multiplex immunoassay. Sci Rep. 2021;11:7612. https://doi.org/10.1038/s41598-021-87155-5.
32. Aversa J, Song M, Shimazu T, Inoue M, Charvat H, Yamaji T, et al. Prediagnostic circulating inflammation biomarkers and esophageal squamous cell carcinoma: a case-cohort study in Japan. Int J Cancer. 2020;147:686–91. https://doi.org/10.1002/ijc.32763.
33. Svensson T, Svensson AK, Kitlinski M, Almgren P, Engström G, Nilsson J, et al. Plasma concentration of caspase-8 is associated with short sleep duration and the risk of incident diabetes mellitus. J Clin Endocrinol Metab. 2018;103:1592–600. https://doi.org/10.1210/jc.2017-02374.
34. Xue L, Borné Y, Mattisson IY, Wigren M, Melander O, Ohro-Melander M, et al. FADD, caspase-3, and caspase-8 and incidence of coronary events. Arterioscler Thromb Vasc Biol. 2017;37:983–9. https://doi.org/10.1161/ATVBAHA.117.308995.
35. Camargo MC, Song M, Ito H, Oze I, Koyanagi YN, Kasugai Y, et al. Associations of circulating mediators of inflammation, cell regulation and immune response with esophageal squamous cell carcinoma. J Cancer Res Clin Oncol. 2021;147:2885–92. https://doi.org/10.1007/s00432-021-03687-3.
36. Sun X, Shu XO, Lan Q, Laszkowska M, Cai Q, Rothman N, et al. Prospective proteomic study identifies potential circulating protein biomarkers for colorectal cancer risk. Cancers. 2022;14 https://doi.org/10.3390/cancers14133261.
37. Dagnino S, Bodinier B, Guida F, Smith-Byrne K, Petrovic D, Whitaker MD, et al. Prospective identification of elevated circulating CDCP1 in patients years before onset of lung cancer. Cancer Res. 2021;81:3738–48. https://doi.org/10.1158/0008-5472.CAN-20-3454.
38. Ioannidis JPA. Why most discovered true associations are inflated. Epidemiol Camb Mass. 2008;19:640–8. https://doi.org/10.1097/EDE.0b013e31818131e7.
39. Shu X, Zhou Q, Sun X, Flesaker M, Guo X, Long J, et al. Associations between circulating proteins and risk of breast cancer by intrinsic subtypes: a Mendelian randomisation analysis. Br J Cancer. 2022;127:1507–14. https://doi.org/10.1038/s41416-022-01923-2.
40. Sun B, Chiou J, Traylor M, Benner C, Hsu YH, Richardson T, et al. Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. 2022. https://doi.org/10.1101/2022.06.17.496443.
41. Uhlén M, Karlsson MJ, Hober A, Svensson AS, Scheffel J, Kotol D, et al. The human secretome. Sci Signal. 2019;12. https://doi.org/10.1126/scisignal.aaz0274.
42. Mallia A, Gianazza E, Zoanni B, Brioschi M, Barbieri SS, Banfi C. Proteomics of extracellular vesicles: update on their composition, biological roles and potential use as diagnostic tools in atherosclerotic cardiovascular diseases. Diagnostics. 2020;10:843. https://doi.org/10.3390/diagnostics10100843.
43. Deutsch EW, Omenn GS, Sun Z, Maes M, Pernemalm M, Palaniappan KK, et al. Advances and utility of the human plasma proteome. J Proteome Res. 2021;20:5241–63. https://doi.org/10.1021/acs.jproteome.1c00657.
44. Conibear AC. Deciphering protein post-translational modifications using chemical biology tools. Nat Rev Chem. 2020;4:674–95. https://doi.org/10.1038/s41570-020-00223-8.
45. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, et al. How many human proteoforms are there? Nat Chem Biol. 2018;14:206–14. https://doi.org/10.1038/nchembio.2576.
46. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. Nature. 2020;580:402–8. https://doi.org/10.1038/s41586-020-2188-x.

ACKNOWLEDGEMENTS

We thank the Affinity Proteomics Unit at SciLifeLab in Stockholm for data generation and quality control.

AUTHOR CONTRIBUTIONS

Conception and design: FG, KC, JS and AM. Financial support: KC, JS and AM. Collection and assembly of data: ME and MG. Data analysis and interpretation: FG, KC and LD. Manuscript writing: FG and KC. Critical review of draft manuscripts and approval of final version: all authors.

FUNDING

This work was financed by the Swedish Research Council (Grant 2022-00584), the Swedish Cancer Society (Grants 22 2207 and 19 0267), the Stockholm County Council (Grant 20200102) and the Karolinska Institutet's Research Foundation (Grant 2018-02146). This work was also supported by a grant from the Stockholm County Council (FoU-954555 and FoUI-978540). Some of the computations were performed on resources provided by SNIC through the Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under Project SNIC 2022/23-504. Open access funding provided by Karolinska Institute.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All women gave written informed consent to participate in the study, to the retrieval of information from medical records, national registries and mammographic images; donated blood at enrolment for genetic analysis and answered a detailed questionnaire about background and lifestyle risk factors. The study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Regional Ethical Review Board in Stockholm, Sweden (Dnr 2010/958-31/1, 2012/217/-32/2 and 2014/1401-32).

CONSENT FOR PUBLICATION

Not applicable.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-023-02541-2.

Correspondence and requests for materials should be addressed to Felix Grassmann.

Reprints and permission information is available at http://www.nature.com/reprints

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