

<https://doi.org/10.1038/s41523-025-00870-4>

# Investigating the relationship between breast cancer risk factors and an AI-generated mammographic texture feature in the Nurses' Health Study II



Xueyao Wu<sup>1</sup>, Shu Jiang<sup>2</sup>, Aaron Ge<sup>3</sup>, Constance Turman<sup>4</sup>, Graham Colditz<sup>2</sup>, Rulla M. Tamimi<sup>5,6</sup> & Peter Kraft<sup>1</sup> ✉

The mammogram risk score (MRS), an AI-driven mammographic texture feature, strongly predicts breast cancer risk independently of breast density, though underlying mechanisms remain unclear. Using data from the Nurses' Health Study II (292 cases, 561 controls), we validated MRS's association with breast cancer and evaluated its relationships with established breast cancer risk factors through observational analyses, polygenic score analyses, and Mendelian randomization. MRS was significantly associated with breast cancer risk before (OR=1.92 per SD increase; 95% CI:1.57 to 2.35; 10-year AUC=0.69) and after adjustment for predicted BI-RADS density (OR=1.85; 95% CI:1.49 to 2.30). Early life body size and adult body mass index (BMI) were inversely associated with MRS, while benign breast disease history and predicted BI-RADS density showed positive associations; after adjusting for density, associations between MRS and the other three risk factors were attenuated. Polygenic score analyses and Mendelian randomization consistently demonstrated significant positive associations between genetic predictors of breast density measures (dense area, percent density, predicted BI-RADS density) and MRS. After adjusting for predicted BI-RADS density and BMI, genetic predictors of higher waist-to-hip ratio were significantly associated with increased MRS. Our findings reveal robust associations between breast density measures and MRS and suggest a potential impact of central obesity on MRS. Future larger-scale validation studies are needed.

Breast cancer remains the most prevalent malignant cancer among women worldwide<sup>1</sup>. While advances in mammographic screening have facilitated early detection and risk stratification by assessing breast density<sup>2</sup>, traditional measures of mammographic density primarily evaluate the relative amounts of fibroglandular tissue (i.e., the functional breast tissue composed of epithelial and stromal cells). This approach limits our ability to fully capture the heterogeneity of individual breast tissue features, such as architecture and spatial relations<sup>3</sup>. Recent research has leveraged accumulating digital mammogram datasets coupled with sophisticated computational techniques to quantify texture features of mammograms, aiming for more precise and individualized risk predictions. These texture features capture detailed

patterns and variations in breast tissue that go beyond simple density measurements<sup>4</sup>. A notable development in this area is the mammogram risk score (MRS), an innovative, artificial intelligence (AI)-driven texture feature derived from whole mammogram images that robustly predicts breast cancer risk independently of breast density (5-year area under the receiver operating characteristic curve [AUC] = 0.75)<sup>5,6</sup>. However, given that the MRS is derived from a deep learning model that lacks inherent interpretability, its biological underpinnings remain unclear.

The risk of breast cancer is influenced by multiple factors beyond age and genetic markers. Lifestyle, behavioral, and developmental factors, such as anthropometric measures and reproductive events, collectively

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA. <sup>2</sup>Washington University School of Medicine in St. Louis, St. Louis, MO, USA. <sup>3</sup>University of Maryland School of Medicine, Baltimore, MD, USA. <sup>4</sup>Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>5</sup>Population Health Sciences Department, Weill Cornell Medical School, New York, NY, USA. <sup>6</sup>Channing Division of Network Medicine, Department of Medicine, Brigham's Hospital and Harvard Medical School, Boston, MA, USA.

✉ e-mail: [phillip.kraft@nih.gov](mailto:phillip.kraft@nih.gov)



contribute to breast cancer susceptibility<sup>7</sup> and may also relate to features in breast tissue. Epidemiological studies have highlighted significant associations between traditional measures of mammographic density and various risk factors, including early life and adult adiposity<sup>8–11</sup>, height<sup>12,13</sup>, age at menarche<sup>12,13</sup>, age at first birth<sup>14,15</sup>, age at natural menopause<sup>16</sup>, and other reproductive/hormonal factors<sup>17</sup>. Utilizing germline genetic variants as instrumental variables (IVs) to strengthen causal inference, Mendelian randomization (MR) studies have reinforced associations with early life and adult adiposity<sup>18,19</sup>, offering protection against confounding and reverse causation typical in observational studies<sup>20</sup>.

Given that texture features capture distinct aspects of breast tissue from summary density measures, investigating how established risk factors relate to these features could improve our understanding of their underlying biology and provide valuable insights into breast cancer pathogenesis. Previous studies have demonstrated phenotypic and genetic relationships between adiposity and V<sup>21,22</sup>, a texture feature reflecting grayscale intensity variations on digitized film mammograms<sup>23</sup>. MRS, by comparison, was developed using supervised machine learning to not only predict variation in whole digital images more accurately but also to capture biological features relevant to breast cancer risk<sup>6</sup>. These characteristics make MRS a promising target for investigation aimed at advancing breast cancer prevention. Yet, to date, no observational or MR study has explored these associations for the MRS.

With an overarching goal of deepening the understanding of the biological underpinnings of MRS and its potential role in breast cancer susceptibility, the present study comprehensively investigates the relationships between established breast cancer risk factors—encompassing anthropometrics, reproductive and hormonal factors, family history, and traditional mammographic density metrics—and MRS, through comprehensive observational and genetic analyses performed within the Nurses' Health Study II (NHS II).

## Results

### Participant characteristics and MRS-breast cancer association in the NHS II

This nested case-control study comprised 853 women (292 cases and 561 controls) with a mean age of 55.3 ( $\pm 5.45$  years) at the time of the mammogram. The majority (64.1%) were postmenopausal. A comparison of risk factor distribution between groups divided by median MRS can be found in Table 1. Notably, compared to those below the median, participants with above-median MRS were younger, more likely to have lower body mass index (BMI) and waist-to-hip ratio (WHR), denser breasts, a history of benign breast disease at the time of mammogram, and more likely to be breast cancer cases (47.8% vs 20.7%) (all  $P < 0.05$ ).

External validation in NHS II, which included 201 cases and 561 controls after excluding cases diagnosed within 6 months of mammography, demonstrated a strong association between MRS and breast cancer risk (odds ratio [OR] = 1.92 per standard deviation [SD] difference in MRS; 95% confidence intervals [CI]: 1.57 to 2.35;  $P = 1.98 \times 10^{-18}$ ; 10-year AUC = 0.69) (Supplementary Table 1, Supplementary Figs. 1 and 2). Cases were diagnosed 0.5–10.1 years (median 2.6) after the mammogram used for MRS calculation. The association remained robust after adjusting for predicted Breast Imaging Reporting and Data System (BI-RADS) density (OR = 1.85; 95% CI: 1.49 to 2.30;  $P = 2.50 \times 10^{-8}$ ) (Supplementary Table 1).

### Associations between observed breast cancer risk factors and MRS

Both predicted BI-RADS density ( $\beta = 0.31$  SD difference in MRS per SD difference in predicted BI-RADS density; 95% CI: 0.25 to 0.38;  $P = 1.94 \times 10^{-20}$ ) and history of benign breast disease ( $\beta = 0.23$  SD difference in MRS with vs. without history; 95% CI: 0.10 to 0.36;  $P = 4.50 \times 10^{-4}$ ) showed positive associations with MRS. Early life body size ( $\beta = -0.08$  SD difference in MRS per SD difference in body size; 95% CI:  $-0.14$  to  $-0.02$ ;  $P = 9.59 \times 10^{-3}$ ) and adult BMI ( $\beta = -0.08$  SD difference in MRS per SD difference in BMI; 95% CI:  $-0.14$  to  $-0.02$ ;  $P = 1.11 \times 10^{-2}$ ) demonstrated

negative associations. No statistically significant associations were observed for the other examined risk factors (all  $P > 0.05$ , Table 2). These results remained consistent in both direction and statistical significance when restricted to controls only or additionally adjusted for menopausal status (Supplementary Tables 2 and 3). When adjusted for predicted BI-RADS density, associations with history of benign breast disease ( $\beta = 0.11$ ), early life body size ( $\beta = -0.02$ ), and adult BMI ( $\beta = 0.05$ ) were all attenuated towards the null (Table 2, Supplementary Table 2).

### Associations between polygenic scores for breast cancer risk factors and MRS

Linear regressions of MRS on polygenic score (PGS) for risk factors revealed significant positive associations for dense area ( $\beta = 0.16$  SD difference in MRS per SD difference in PGS; 95% CI: 0.06 to 0.25;  $P = 1.37 \times 10^{-3}$ ) and percent density ( $\beta = 0.14$  SD difference in MRS per SD difference in PGS; 95% CI: 0.05 to 0.23;  $P = 3.29 \times 10^{-3}$ ). No significant associations were observed between the PGS for other risk factors and MRS (Table 3). A similar pattern of associations was observed in analyses restricted to controls and in models adjusted for menopausal status (Supplementary Tables 4 and 5). After adjusting for predicted BI-RADS density, the association for percent density remained strong, whereas the association for dense area weakened slightly. A significant association was additionally observed between higher PGS for WHR adjusted for BMI (WHRadjBMI) and increased MRS ( $\beta = 0.12$  SD difference in MRS per SD difference in PGS; 95% CI: 0.03 to 0.21;  $P = 1.18 \times 10^{-2}$ ) (Table 3, Supplementary Table 4).

### Mendelian randomization between breast cancer risk factors and MRS

Linear regressions of risk factors on their corresponding PGS revealed significant genetic associations for 7 risk factors, including height, BMI, age at menarche, early life body size, WHR, predicted BI-RADS density, and age at natural menopause ( $F$ -statistics: 212.56 to 5.36, Supplementary Table 6). While PGS for dense area was significantly associated with predicted BI-RADS density ( $R^2 = 0.03$ ,  $F = 9.09$ ,  $P = 2.76 \times 10^{-3}$ ), PGS for percent density showed no association ( $R^2 = 0.00$ ,  $F = 1.47$ ,  $P = 0.23$ ). The dense area PGS was thus used as an IV for predicted BI-RADS density in subsequent two-stage least squares (2SLS) analyses. WHRadjBMI, age at first birth, and number of children ever born were excluded from 2SLS due to weak instrument strength ( $F$ -statistic  $< 5$ ). Associations adjusted for menopausal status or predicted BI-RADS density are detailed in Supplementary Tables 7 and 8.

2SLS analyses found a significant association between genetically predicted BI-RADS density and MRS ( $\beta = 0.84$  SD difference in MRS per SD difference in predicted BI-RADS density; 95% CI: 0.21 to 1.46;  $P = 8.85 \times 10^{-3}$ ), while identifying no statistically significant associations between the other 6 genetically predicted risk factors and MRS. Among the other risk factors, the strongest effect estimates were observed for age at natural menopause ( $\beta = -0.80$  SD difference in MRS per SD difference in genetically predicted age at natural menopause, 95% CI:  $-2.58$  to  $0.98$ ,  $P = 0.38$ ) and early life body size ( $\beta = -0.13$  SD difference in MRS per SD difference in genetically predicted early life body size, 95% CI:  $-0.59$  to  $0.33$ ,  $P = 0.57$ ), neither of which reached statistical significance (Table 4). All additional adjustments yielded similar results (Table 4, Supplementary Tables 9 and 10).

Two-sample MR analyses identified significant associations between genetically predicted dense area and MRS ( $\beta = 0.83$  SD difference in MRS per SD difference in dense area; 95% CI: 0.39 to 1.27;  $P = 2.09 \times 10^{-4}$ ), and genetically predicted percent density and MRS ( $\beta = 1.14$  SD difference in MRS per SD difference in percent density; 95% CI: 0.55 to 1.74;  $P = 1.61 \times 10^{-4}$ ). No evidence supported significant causal associations with other risk factors (Fig. 1). Sensitivity analyses using MR-Egger regression, weighted median, weighted mode, and inverse-variance weighted (IVW) excluding outlier SNPs yielded consistent results (Supplementary Table 11). MR-Clust analysis found all variants for dense area and percent density clustered into a single group with similar causal effects, suggesting no

**Table 1 | Baseline characteristics of participants according to median mammogram risk score**

Characteristic	All N = 853	MRS < the median level N = 426	MRS ≥ the median level N = 427	P
Age, years	55.3 (5.45)	56.4 (5.21)	54.3 (5.48)	<0.001
Early life body size	2.56 (1.16)	2.64 (1.21)	2.49 (1.10)	0.059
Body mass index, kg/m <sup>2</sup>	26.8 (6.06)	27.4 (6.45)	26.2 (5.57)	0.003
Waist-to-hip ratio	0.83 (0.07)	0.83 (0.07)	0.82 (0.07)	0.035
Waist-to-hip ratio adjusted for BMI	0.00 (0.07)	0.00 (0.07)	0.00 (0.07)	0.648
Height, inch	64.9 (2.52)	64.9 (2.48)	65.0 (2.55)	0.523
Age at menarche, years	13.2 (2.95)	13.2 (2.98)	13.2 (2.92)	0.913
Age at first birth, years	26.3 (4.93)	26.0 (4.71)	26.7 (5.12)	0.075
Menopausal status				<0.001
Premenopausal	233 (27.8%)	86 (20.6%)	147 (34.8%)	
Postmenopausal	538 (64.1%)	302 (72.4%)	236 (55.9%)	
Unsure	68 (8.10%)	29 (6.95%)	39 (9.24%)	
Age at natural menopause, years	49.7 (4.14)	49.5 (4.37)	50.0 (3.80)	0.094
Postmenopausal hormone use				0.108
Premenopausal or postmenopausal not on therapy	626 (77.4%)	322 (79.3%)	304 (75.4%)	
Postmenopausal on therapy	183 (22.6%)	84 (20.7%)	99 (24.6%)	
No. pregnancies ≥6 months	1.94 (1.24)	1.96 (1.25)	1.92 (1.22)	0.664
History of benign breast disease				0.013
Yes	525 (61.5%)	244 (57.3%)	281 (65.8%)	
No	328 (38.5%)	182 (42.7%)	146 (34.2%)	
Predicted BI-RADS density <sup>a</sup>				<0.001
a, almost entirely fatty	40 (5.10%)	33 (8.44%)	7 (1.78%)	
b, scattered areas of fibroglandular tissue	320 (40.8%)	209 (53.5%)	111 (28.2%)	
c, heterogeneously dense	396 (50.4%)	140 (35.8%)	256 (65.0%)	
d, extremely dense	29 (3.69%)	9 (2.30%)	20 (5.08%)	
Family history of breast cancer				0.14
Yes	160 (18.8%)	71 (16.7%)	89 (20.8%)	
No	693 (81.2%)	355 (83.3%)	338 (79.2%)	
Breast cancer cases				<0.001
Yes	292 (34.2%)	88 (20.7%)	204 (47.8%)	
No	561 (65.8%)	338 (79.3%)	223 (52.2%)	
Mammogram risk score	0.00 (1.00)	−0.84 (0.58)	0.84 (0.49)	<0.001

<sup>a</sup>Predicted BI-RADS density was assessed using a deep learning algorithm previously developed to predict mammographic breast density from digital mammograms. The algorithm categorizes breasts from a (almost entirely fatty) to d (extremely dense), matching an experienced mammographer's evaluation.

evidence of heterogeneous causal mechanisms; for other risk factors, no variants showed significant effects (Supplementary Fig. 3). Chi-square tests on Wald ratios across all IVs for each risk factor revealed no statistically significant associations between risk factor-associated genetic variants and MRS (all  $P > 0.05$ , Supplementary Table 12).

The patterns of associations were robust to both restriction to control subjects (Supplementary Table 11) and adjustment for menopausal status (Supplementary Tables 13 and 14, Supplementary Fig. 4). Utilizing IV-outcome associations adjusting for predicted BI-RADS density, two-sample MR showed a significant association between genetically predicted WHRadjBMI and MRS (IVW:  $\beta = 0.51$  SD difference in MRS per SD difference in WHRadjBMI; 95% CI: 0.15 to 0.87;  $P = 6.10 \times 10^{-3}$ ) (Fig. 1), consistent across all sensitivity analyses. Association with percent density remained substantially unchanged, while association for genetically predicted dense area was attenuated (Fig. 1, Supplementary Tables 15 and 16, Supplementary Fig. 5).

## Discussion

To the best of our knowledge, this study presents one of the first and most comprehensive examinations to date of the relationships between known breast cancer risk factors and MRS—an AI-generated mammographic

texture feature derived from full-field digital mammograms. Our analyses revealed robust phenotypic and genetic associations between various mammographic density measures—including predicted BI-RADS density, absolute dense area, and percent density—and the MRS, as well as a suggestive association between higher WHRadjBMI and increasing MRS.

Our external validation of the MRS in the predominantly White NHS II cohort demonstrated its robust predictive capability for breast cancer risk, supporting the generalizability of the algorithm to an independent population. The MRS maintained good discriminatory power for long-term risk, achieving a 10-year AUC of 0.69, which compares favorably with the 5-year AUCs reported in the original validation cohorts (0.75 in the Joanne Knight Breast Health Cohort at Washington University [WashU cohort]; 0.74 in the Emory Breast Imaging Dataset [EMBED]; 27–46% Non-Hispanic Black women)<sup>6</sup>. This result underscores the stability of the MRS as a risk marker over an extended follow-up period. Notably, our mutual adjustment analyses provide insight into the relationship between MRS, breast density, and breast cancer risk. The association between MRS and breast cancer incidence remained strong after adjusting for predicted BI-RADS density; conversely, the association for predicted BI-RADS density was substantially attenuated after adjusting for MRS. This pattern suggests that while both are important risk factors, MRS may capture mammographic information that

**Table 2 | Linear regression of mammogram risk score on each breast cancer risk factor**

Risk factor <sup>a</sup>	Crude model <sup>b</sup>				Model adjusted for predicted BI-RADS density <sup>c</sup>			
	N	BETA	SE	P	N	BETA	SE	P
Early life body size	837	−0.082	0.032	$9.59 \times 10^{-3}$	770	−0.022	0.032	0.479
Body mass index	849	−0.082	0.032	$1.11 \times 10^{-2}$	781	0.049	0.035	0.165
Waist-to-hip ratio	728	−0.059	0.034	0.085	668	−0.006	0.035	0.868
Waist-to-hip ratio adjusted for BMI	725	−0.006	0.034	0.869	665	0.008	0.034	0.807
Height	853	−0.024	0.032	0.440	785	−0.011	0.031	0.720
Age at menarche	850	0.003	0.032	0.922	782	−0.016	0.032	0.619
Age at first birth	699	0.031	0.035	0.384	649	0.037	0.034	0.277
Age at natural menopause <sup>d</sup>	535	0.075	0.042	0.074	489	0.059	0.041	0.152
Postmenopausal hormone use <sup>d</sup>	518	0.139	0.094	0.139	475	0.022	0.093	0.815
No. pregnancies $\geq 6$ months	761	−0.005	0.033	0.876	722	0.029	0.033	0.387
History of benign breast disease	853	0.229	0.065	$4.50 \times 10^{-4}$	785	0.111	0.065	0.091
Predicted BI-RADS density	785	0.314	0.033	$1.94 \times 10^{-20}$	-	-	-	-
Family history of breast cancer	853	0.017	0.081	0.838	785	0.003	0.080	0.974

BMI body mass index.

<sup>a</sup>All risk factors reflect measurements taken at adulthood, either at cohort baseline or at the time of the mammogram, except for early life body size, which reflects recalled values from childhood.<sup>b</sup>Models include adjustments for age and breast cancer case-control status.<sup>c</sup>Models include adjustments for age, predicted BI-RADS density, and breast cancer case-control status.<sup>d</sup>Analyses of age at natural menopause and postmenopausal hormone use were limited to postmenopausal women.**Table 3 | Linear regression of mammogram risk score on polygenic score for each breast cancer risk factor**

Risk factor <sup>a</sup>	Crude model <sup>b</sup>				Model adjusted for predicted BI-RADS density <sup>c</sup>			
	N	BETA	SE	P	N	BETA	SE	P
Early life body size	383	−0.027	0.050	0.588	353	0.035	0.050	0.482
Adult body size	383	−0.003	0.049	0.950	353	0.037	0.049	0.444
Waist-to-hip ratio	383	0.008	0.048	0.870	353	0.073	0.048	0.124
Waist-to-hip ratio adjusted for BMI	383	0.071	0.048	0.139	353	0.118	0.047	$1.18 \times 10^{-02}$
Height	383	−0.004	0.049	0.932	353	−0.027	0.048	0.580
Age at menarche	383	−0.022	0.049	0.658	353	−0.059	0.049	0.223
Age at first birth	383	−0.006	0.049	0.908	353	−0.033	0.047	0.489
Age at natural menopause	383	−0.034	0.049	0.492	353	−0.045	0.047	0.344
Number of children ever born	383	−0.055	0.049	0.261	353	−0.044	0.048	0.357
Dense area	383	0.155	0.048	$1.37 \times 10^{-03}$	353	0.093	0.049	0.062
Non-dense area	383	0.001	0.048	0.979	353	0.009	0.047	0.850
Percent density	383	0.141	0.048	$3.29 \times 10^{-03}$	353	0.126	0.048	$9.36 \times 10^{-03}$

BMI body mass index.

<sup>a</sup>All risk factors reflect measurements taken at adulthood, except for early life body size, which reflects recalled values from childhood.<sup>b</sup>Models include adjustments for age, genotyping platform, the top 10 genetic principal components, and breast cancer case-control status.<sup>c</sup>Models include adjustments for age, predicted BI-RADS density, genotyping platform, the top 10 genetic principal components, and breast cancer case-control status.

is more proximally located on the causal pathway to breast cancer than summary density measures alone. These findings collectively underscore MRS's potential to enhance breast cancer risk stratification across diverse clinical settings and populations.

Our findings align with previously reported significant phenotypic relationships between mammographic density and breast texture features<sup>17,21,24–28</sup>. The moderate correlation between predicted BI-RADS density and MRS ( $r \sim 0.31$ ) further corroborates that while these measures are related, they are likely to reflect distinct aspects of mammographic information. Beyond the phenotypic association, our genetic analyses provide converging evidence for a shared genetic architecture and potential causal link between mammographic density and MRS. These findings corroborate and extend previous evidence from different texture measures and study designs that demonstrated similar causal relationships<sup>29</sup>,

enhancing the credibility of MRS as a biologically plausible risk factor for breast cancer. Future studies should aim to elucidate the specific biological processes reflected by MRS and their implications for breast cancer etiology.

Moving beyond density, investigating the effects of lifestyle, behavioral, and developmental/biological risk factors on breast tissue characteristics, as summarized in mammograms, is crucial for extracting biological insights into modifiable factors for prevention studies and understanding pathways for potential preventive drug targets. While MRS itself represents a novel feature with limited existing literature, it is important to contextualize our findings within the existing body of research on other mammographic features. For instance, previous studies have demonstrated associations between various breast cancer risk factors and mammographic density, and between risk factors and other texture features such as  $V^{21,22}$ . Our study of MRS builds upon previous findings by focusing on MRS—a supervised,

**Table 4 | Two-stage least squares regression between each breast cancer risk factor (exposure) and mammogram risk score (outcome)**

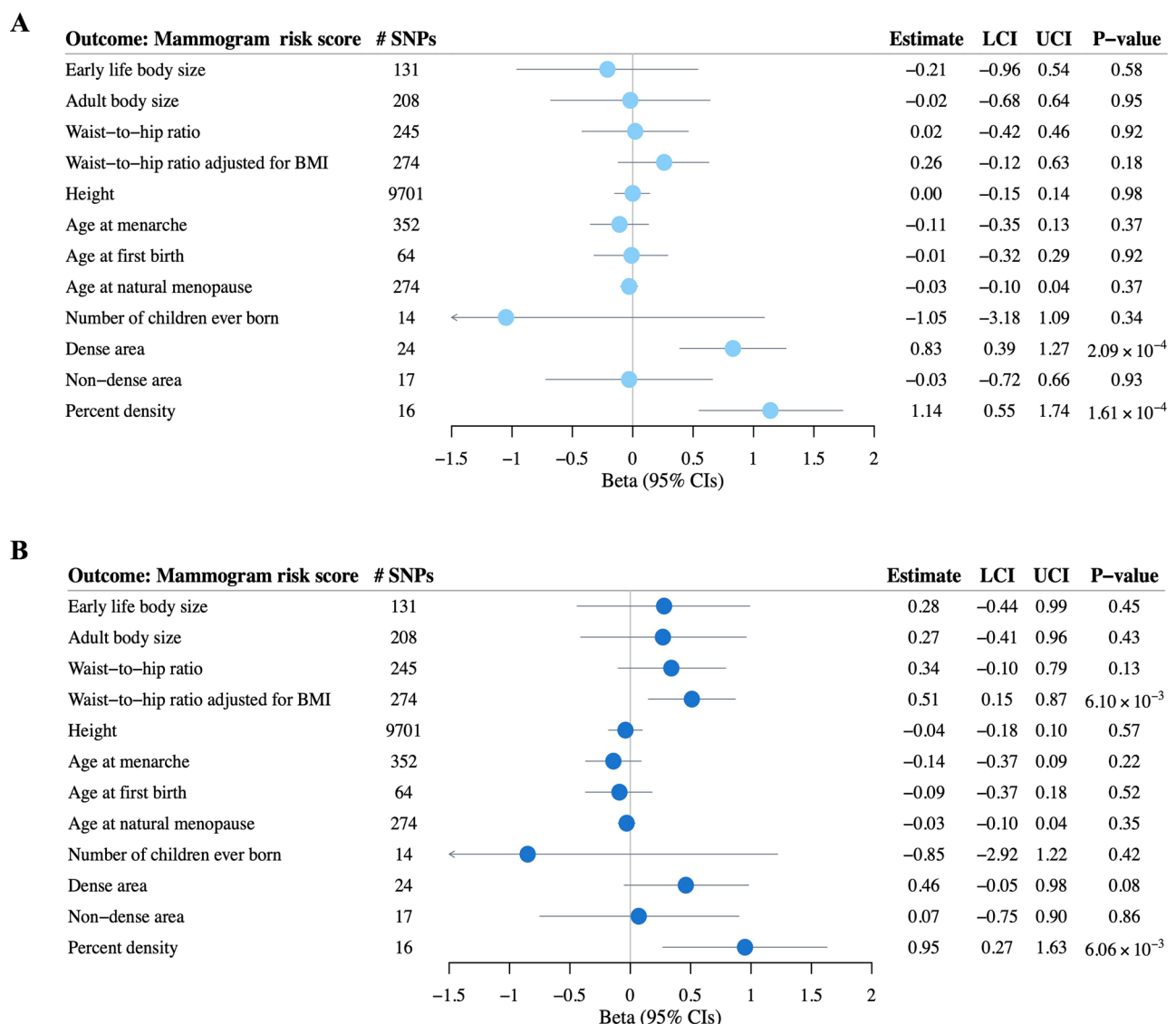
Risk factor <sup>a</sup>	Crude model <sup>b</sup>				Model adjusted for predicted BI-RADS density <sup>c</sup>			
	N	BETA	SE	P	N	BETA	SE	P
Early life body size	370	−0.132	0.234	0.574	341	0.215	0.269	0.424
Body mass index	381	−0.012	0.175	0.946	351	0.153	0.201	0.446
Waist-to-hip ratio	336	0.081	0.304	0.790	307	0.702	0.542	0.196
Height	383	−0.007	0.084	0.932	353	−0.045	0.080	0.579
Age at menarche	383	−0.081	0.183	0.659	353	−0.222	0.186	0.232
Age at natural menopause <sup>d</sup>	259	−0.800	0.907	0.378	240	−1.068	1.388	0.442
Predicted BI-RADS density	353	0.835	0.317	$8.85 \times 10^{-03}$	-	-	-	-

<sup>a</sup>All risk factors reflect measurements taken at adulthood, except for early life body size, which reflects recalled values from childhood.

<sup>b</sup>Models include adjustments for age, genotyping platform, the top 10 genetic principal components, and breast cancer case-control status.

<sup>c</sup>Models include adjustments for age, predicted BI-RADS density, genotyping platform, the top 10 genetic principal components, and breast cancer case-control status.

<sup>d</sup>Analyses of age at natural menopause were restricted to postmenopausal women.



**Fig. 1 | Two-sample Mendelian randomization analysis examining associations between genetically predicted risk factors (exposures) and mammogram risk score (outcome).** Results are shown **A** before and **B** after adjusting for predicted BI-RADS density in Nurses' Health Study II (NHS II). Point estimates (circles) and 95% confidence intervals (error bars) were calculated using the inverse-variance weighted

approach. Effect estimates are interpreted as the standard deviation change in MRS per one-unit increase in the exposure as defined in the original GWAS. All risk factors represent adult measurements, except for early life body size, which was retrospectively reported for childhood. BMI body mass index, LCI lower confidence interval, UCI upper confidence interval.



risk-optimized score trained via a ResNet-18 convolutional neural network and validated in large, diverse cohorts to identify tissue patterns most predictive of future breast cancer incidence. This approach offers a complementary perspective for investigating the biology of risk-relevant mammographic changes.

The emergence of a statistically significant association between genetic predictors of WHRadjBMI and MRS only after adjusting for predicted BI-RADS density suggests that fat distribution, independent of overall body mass, might influence breast tissue characteristics in ways not fully captured by mammographic density alone. The ability of MRS to reveal this relationship indicates its value as an advanced imaging feature in reflecting nuanced aspects of breast tissue composition that may be relevant to cancer risk assessment. Future studies are needed to validate our results and investigate the biological mechanisms underlying the complex interplay between fat distribution, breast tissue texture features, and breast cancer susceptibility.

Several limitations of our study should be acknowledged. First, our sample size was relatively limited, which may have led to insufficient statistical power to detect associations with some risk factors, particularly those with smaller effect sizes. We emphasize that null findings observed should not be interpreted as definitive evidence of no association, and that larger studies with greater statistical power are needed. Second, our analysis was limited by the availability of only craniocaudal (CC)-view mammograms. As the MRS algorithm is optimized using four views, our reported predictive accuracy likely represents a conservative estimate of its full potential. Nevertheless, the strong performance achieved with this two-view approach underscores the algorithm's utility in common, real-world scenarios where imaging sets may be incomplete, thus broadening its applicability in diverse data settings. Third, our analysis was based on a nested case-control design, which could potentially introduce ascertainment bias. However, we expect that this design would not substantially affect our results, given our careful adjustment for case-control status and the consistency of our findings in control-only analyses<sup>30</sup>. Fourth, while we evaluated predicted BI-RADS density (which mimics qualitative visual assessments by radiologists)<sup>31</sup>, we were unable to adjust for quantitative density measures due to data unavailability. This limitation may have reduced our statistical power to detect associations, potentially underestimating relationships between other breast cancer risk factors and MRS. Additionally, our 2SLS analyses may have been affected by phenotype mismatches between the traits used to derive PGS from genome-wide association studies (GWAS) and the corresponding phenotypes measured in our study. For example, the use of quantitative dense area PGS to instrument qualitative BI-RADS density categories could potentially violate the restriction exclusion assumption and introduce bias in our causal estimates.

Our study has several notable strengths. A key advantage is the availability of genetic data, digital mammogram data, and comprehensive covariate data on the same set of samples, allowing for integrated analyses across multiple domains. Triangulating evidence from both observational and genetic studies mitigated biases inherent in each study design, providing a multi-perspective evaluation of associations. The MRS algorithm was developed independently of the NHS II cohort, reducing the possibility for overfitting or circular reasoning in our analyses. Developed using standard digital mammograms, sophisticated statistical methods, and large-scale populations, the MRS itself represents an advanced texture feature with significant potential in clinical settings.

To conclude, this study provides initial insights into the etiologic underpinnings of MRS. We validated that MRS serves as a robust predictor of breast cancer risk, providing information independent of and beyond that captured by traditional density measurements. Our investigation further revealed robust associations between breast density measures and MRS and suggests a potential impact of central obesity on MRS. Future research should encompass larger-scale studies to definitively characterize these associations and elucidate the underlying biological mechanisms. As our understanding of mammographic texture features advances, tools like MRS that offer a more nuanced view of risk beyond conventional measures may

become integral to personalized breast cancer assessment and prevention strategies.

## Methods

### Study participants

The current study leverages resources from the NHS II, a large prospective cohort established in 1989 with 116,429 female and predominantly White (>90%) registered nurses aged 25–42 from 14 states<sup>32</sup>. Between 1996 and 1999, blood samples were collected from 29,611 women, forming a blood subcohort<sup>33</sup>. Genotype data from four platforms (Affymetrix 6.0, Illumina HumanHap, Illumina OmniExpress, and Illumina OncoArray) imputed to the 1000 Genomes Phase 3 version 5 reference panel were used in this study. Pre-diagnostic screening mammograms, conducted as close as possible to the blood draw date, were collected as part of a breast cancer case-control study nested within the blood subcohort. Participants have been followed up biennially through self-administered questionnaires to update exposure information and disease diagnoses. For this study, we initially included 853 women (292 cases and 561 controls) with eligible full-field digital mammograms. Among these, 383 women (143 cases and 240 controls) had available imputed genotype data and were included in the genetic analyses. Detailed descriptions of the full genotyping and quality control pipeline<sup>34</sup>, as well as the mammogram collection and processing procedure<sup>21,24</sup>, are available in previous publications. Cohort participants provided written informed consent. The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

### Risk factors measurement

Information on various established risk factors for breast cancer was collected for NHS II women. These factors included early life and adult body size, fat distribution, height, reproductive characteristics, and family history of breast cancer. Early life body size, WHR, height (inches), and age at menarche were reported via the baseline questionnaire in 1989. Body sizes at ages 5 and 10 years were recalled using Stunkard's nine-level pictogram (levels 1–9: most lean to most overweight)<sup>35</sup>. The average of these two measurements was used to represent early life body size. For other covariates, we used the most recent information from the biennial questionnaires preceding the date of the mammogram. These covariates included: BMI, age at first birth, menopausal status, age at natural menopause, current postmenopausal hormone use, parity (number of pregnancies  $\geq 6$  months), history of benign breast disease, and family history of breast cancer. BMI ( $\text{kg}/\text{m}^2$ ) was calculated by dividing weight (kg) by the square of baseline height (m). WHRadjBMI was further calculated by regressing WHR on BMI and using the residuals from this regression.

We also assessed predicted BI-RADS density using a deep learning algorithm, which was previously developed to predict mammographic breast density from digital mammograms<sup>31</sup>. The algorithm categorizes breasts from a (almost entirely fatty) to d (extremely dense), matching an experienced mammographer's evaluation (weighted  $\kappa$  for agreement with radiologists = 0.85). We coded these categories as 1, 2, 3, 4, with higher numbers indicating denser breasts. The digital mammograms used for MRS calculation were used to assess predicted BI-RADS density. Other quantitative density measures, such as absolute dense area and percent density, were not directly measured for the NHS II participants included in this study.

### Mammogram risk score measurement

The MRS is an AI-derived score capturing the texture information embedded in the whole digital mammograms, represented by millions of pixels<sup>5,6</sup>. It was developed utilizing 220,868 mammograms from 10,126 racially diverse, initially cancer-free women in the WashU cohort<sup>36</sup>, of whom 505 developed breast cancer during follow-up. Validation was performed using 150,352 mammograms from 15,885 women in EMED, demonstrating consistently robust predictive performance (5-year

AUC = 0.74)<sup>6</sup>. The algorithm, previously described in detail<sup>6</sup>, takes all standard mammogram views (CC and/or mediolateral oblique) from both breasts as input with the option of additional clinical risk factors. The outputs of the algorithm include MRS, which is a transparent weighted sum of feature coefficients, probability of 5-year breast cancer onset, and relative risk for each woman that can be used for risk calibration. For the current study, we generated an MRS for each of the 853 women by applying the algorithm to their pair of digital CC-view mammograms (one from each breast), totaling 1706 images. We used the earliest digital mammograms available for each woman.

### Genetic variants and polygenic scores for risk factors

We selected the largest available GWAS conducted among women of European ancestry for early life and adult body size, WHR, WHRadjBMI, height, age at menarche, age at first birth, age at natural menopause, number of children ever born, dense area, non-dense area, and percent density. For each risk factor, we collected lists of genetic variants reported as genome-wide significant ( $P < 5.0 \times 10^{-8}$ ) in the original female-specific GWAS, along with their beta coefficients. When such variants were not explicitly reported, we applied PLINK's clumping function<sup>37</sup> (parameters:  $P < 5.0 \times 10^{-8}$ , linkage disequilibrium  $r^2 < 0.001$  within a 10 Mb window) to obtain this information. For height, for which no female-specific GWAS is known to be publicly accessible, we used genetic variant information from the largest available sex-combined GWAS. This approach was justified as no statistically significant evidence for sex differences in height genetics has been reported<sup>38</sup>.

To ensure these established variants were reliably imputed in the NHS II data, we included only those with matching alleles, non-ambiguous SNPs, a minimum imputation score  $>0.3$  across all genotyping platforms, and a minor allele frequency  $>0.005$ . These selected variants were used for the PGS calculation and as IVs in causal inference analyses. Detailed information on GWAS sources and quality control of genetic variants is provided in Supplementary Table 17.

### Statistical analysis

We generated descriptive statistics for all variables. Continuous variables were described using mean and SD, while categorical variables were described using frequency and percentage. We assessed differences between higher and lower MRS groups (using the median value as the cutoff) using Student's *t*-test or Wilcoxon rank-sum test for continuous variables and Chi-square test for categorical variables.

We first validated the association between MRS and breast cancer in NHS II using logistic regression after excluding 91 cases diagnosed within 6 months after the mammogram used for calculation. To evaluate the association between breast cancer risk factors and MRS, four main analyses were performed: (i) linear regressions of MRS on each observed risk factor to quantify their observational association without accounting for genetic predisposition; (ii) linear regressions of MRS on the PGS associated with each risk factor to evaluate the relationship between genetic predisposition to each risk factor and MRS; (iii) MR analysis via 2SLS regressions of MRS on each genetically predicted risk factor, and (iv) two-sample MR of MRS using GWAS summary statistics of each risk factor, to evaluate potential causal associations. For all analyses, we standardized MRS and all non-binary variables for easier comparison across risk factors. Binary variables included postmenopausal hormone use, history of benign breast disease, and family history of breast cancer, each categorized as "Yes" or "No." In two-sample MR, we retained the original scale of genetic associations from the source GWAS.

For each risk factor, we calculated its weighted PGS using PLINK's "--score" function<sup>37</sup>, summing the products of effect allele dosage and corresponding beta coefficient across all selected genetic variants for each woman. Prior to 2SLS regression, we assessed instrument strength by regressing each risk factor on its corresponding PGS, obtaining *F*-statistics and correlation coefficient estimates. To minimize weak instrument bias, we excluded PGS with an *F*-statistic  $< 5$  or a correlation  $P > 0.05$  from the 2SLS analysis. The 2SLS procedure involved two stages: first, regressing each risk

factor on its PGS; second, using the predicted values as independent variables in a regression model with MRS as the dependent variable.

For two-sample MR, we obtained the "IVs-exposure" associations directly from the corresponding GWAS. The "IV-outcome" associations were estimated from the NHS II dataset using PLINK's "--glm" function<sup>37</sup>; for our primary analysis, this was performed in the full genetic dataset ( $N = 383$ ) with adjustment for case-control status. Our primary method was the random-effect IVW approach<sup>39</sup>, which assumes a zero intercept and estimates causality using random-effects meta-analysis. To validate MR model assumptions<sup>20</sup> and assess the robustness of our findings, we applied complementary methods including MR-Egger regression (which detects and accounts for directional pleiotropy)<sup>40</sup>, weighted median (robust to up to 50% invalid instruments)<sup>41</sup>, weighted mode (identifies the causal effect estimate that is most consistent across all variants)<sup>42</sup>, and IVW excluding outlier SNPs detected using Radial MR's iterative Cochran's Q method<sup>43</sup>. We considered a causal association significant if it reached statistical significance in the IVW analysis and maintained a consistent direction across all sensitivity analyses. Following two-sample MR, we performed two additional analyses: an MR-Clust analysis to cluster genetic variants with similar causal estimates, which may reflect heterogeneous causal mechanisms<sup>44</sup>, and a Chi-square test on Wald ratios estimated in two-sample MR across all IVs for each risk factor to test if any of the risk factor-associated genetic variants associate with MRS.

To mitigate confounding, we employed three adjustment sets across all analyses. The crude model included age at mammogram and, where appropriate, genotyping platform and the top 10 genetic principal components. The second and third sets were additionally adjusted for menopausal status and predicted BI-RADS density, respectively. To address potential ascertainment bias arising from investigating MRS in a case-control study design that implicitly conditions on breast cancer status, we conducted all analyses using two approaches as previously recommended<sup>30</sup>: (1) including case-control status as an additional covariate (our primary approach to maintain sample size), and (2) restricting analyses to controls only. 2SLS and TSMR analyses were conducted using packages "ivreg", "TwoSampleMR", and "RadialMR" in R (v4.1.0). We used the conventional *P*-value threshold of 0.05 to define statistical significance, given the relatively limited sample size and the exploratory nature of our study.

### Data availability

The data that support the findings of this study are available from the Nurses' Health Studies; however, they are not publicly available. Investigators interested in using the data can request access, and feasibility will be discussed at an investigator's meeting. Limits are not placed on scientific questions or methods, and there is no requirement for co-authorship. Additional data sharing information and policy details can be accessed at <http://www.nurseshealthstudy.org/researchers>. All GWAS summary statistics used in this study are publicly available.

### Code availability

Analysis scripts used to generate the results of this study are available from the corresponding author upon reasonable request.

Received: 25 February 2025; Accepted: 18 November 2025;

Published online: 23 December 2025

### References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 (2024).
2. Vilmun, B. M. et al. Impact of adding breast density to breast cancer risk models: a systematic review. *Eur. J. Radiol.* **127**, <https://doi.org/10.1016/j.ejrad.2020.109019> (2020).
3. Gastouni, A., Conant, E. F. & Kontos, D. Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast

- cancer risk assessment. *Breast Cancer Res.* **18**, <https://doi.org/10.1186/S13058-016-0755-8> (2016).
4. Anandarajah, A. et al. Studies of parenchymal texture added to mammographic breast density and risk of breast cancer: a systematic review of the methods used in the literature. *Breast Cancer Res.* **24**, <https://doi.org/10.1186/S13058-022-01600-5> (2022).
5. Jiang, S., Bennett, D. L. & Colditz, G. A. Deriving a Mammogram-Based Risk Score from Screening Digital Breast Tomosynthesis for 5-Year Breast Cancer Risk Prediction. *Cancer Prev Res (Phila.)* **18**, 347–354 (2025).
6. Jiang, S., Bennett, D., Rosner, B., Tamimi, R. & Colditz, G. Development and validation of a 5-year dynamic risk model using repeated mammograms. *JCO Clin. Cancer Inform.* **8**, e2400200 (2024).
7. Loibl, S., Poortmans, P., Morrow, M., Denkert, C. & Curigliano, G. Breast cancer. *Lancet* **397**(10286), 1750–1769 (2021).
8. Hopper, J. L. et al. Childhood body mass index and adult mammographic density measures that predict breast cancer risk. *Breast Cancer Res. Treat.* **156**, 163–170 (2016).
9. Andersen, Z. J. et al. Birth weight, childhood body mass index, and height in relation to mammographic density and breast cancer: a register-based cohort study. *Breast Cancer Res.* **16**, <https://doi.org/10.1186/BCR3596> (2014).
10. Barnard, M. E. et al. Body mass index and mammographic density in a multiracial and multiethnic population-based study. *Cancer Epidemiol. Biomark. Prev.* **31**, 1313–1323 (2022).
11. Boyd, N. F. et al. Body size, mammographic density, and breast cancer risk. *Cancer Epidemiol. Biomark. Prev.* **15**, 2086–2092 (2006).
12. Dite, G. S. et al. Predictors of mammographic density: insights gained from a novel regression analysis of a twin study. *Cancer Epidemiol. Biomark. Prev.* **17**, 3474–3481 (2008).
13. Ward, S. V. et al. The association of age at menarche and adult height with mammographic density in the International Consortium of Mammographic Density. *Breast Cancer Res.* **24**, <https://doi.org/10.1186/S13058-022-01545-9> (2022).
14. El-Bastawissi, A. Y., White, E., Mandelson, M. T. & Taplin, S. H. Reproductive and hormonal factors associated with mammographic breast density by age (United States). *Cancer Causes Control* **11**, 955–963 (2000).
15. Yaghjian, L., Colditz, G. A., Rosner, B., Bertrand, K. A. & Tamimi, R. M. Reproductive factors related to childbearing and mammographic breast density. *Breast Cancer Res. Treat.* **158**, 351–359 (2016).
16. Rice, M. S. et al. Reproductive and lifestyle risk factors and mammographic density in Mexican women. *Ann. Epidemiol.* **25**, 868–873 (2015).
17. Yaghjian, L. et al. Relationship between breast cancer risk factors and mammographic breast density in the Fernald Community Cohort. *Br. J. Cancer* **106**, 996–1003 (2012).
18. Haas, C. B. et al. Disentangling the relationships of body mass index and circulating sex hormone concentrations in mammographic density using Mendelian randomization. *Breast Cancer Res. Treat.* **206**, 295–305 (2024).
19. Vabistsevits, M. et al. Mammographic density mediates the protective effect of early-life body size on breast cancer risk. *Nat. Commun.* **15**, 1–15 (2024).
20. Smith, G. D. & Ebrahim, S. Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease?. *Int. J. Epidemiol.* **32**, 1–22 (2003).
21. Oh, H. et al. Early-life and adult anthropometrics in relation to mammographic image intensity variation in the nurses’ health studies. *Cancer Epidemiol. Biomark. Prev.* **29**, 343–351 (2020).
22. Liu, Y. et al. A genome-wide association study of mammographic texture variation. *Breast Cancer Res.* **24**, 1–15 (2022).
23. Heine, J. J. et al. A novel automated mammographic density measure and breast cancer risk. *J. Natl. Cancer Inst.* **104**, 1028–1037 (2012).
24. Warner, E. T. et al. Automated percent mammographic density, mammographic texture variation, and risk of breast cancer: a nested case-control study. *npj Breast Cancer* **7**, <https://doi.org/10.1038/S41523-021-00272-2> (2021).
25. Manduca, A. et al. Texture features from mammographic images and risk of breast cancer. *Cancer Epidemiol. Biomark. Prev.* **18**, 837 (2009).
26. Wanders, J. O. P. et al. The combined effect of mammographic texture and density on breast cancer risk: a cohort study. *Breast Cancer Res.* **20**, <https://doi.org/10.1186/S13058-018-0961-7> (2018).
27. Watt, G. P. et al. Mammographic texture features associated with contralateral breast cancer in the WECARE Study. *npj Breast Cancer* **7**, 1–4 (2021).
28. Malkov, S. et al. Mammographic texture and risk of breast cancer by tumor type and estrogen receptor status. *Breast Cancer Res.* **18**, 1–11 (2016).
29. Ye, Z. et al. Causal relationships between breast cancer risk factors based on mammographic features. *Breast Cancer Res.* **25**, 127 (2023).
30. Monsees, G. M., Tamimi, R. M. & Kraft, P. Genome-wide association scans for secondary traits using case-control samples. *Genet Epidemiol.* **33**, 717 (2009).
31. Lehman, C. D. et al. Mammographic breast density assessment using deep learning: clinical implementation. *Radiology* **290**, 52–58 (2019).
32. Colditz, G. A. & Hankinson, S. E. The Nurses’ Health Study: lifestyle and health among women. *Nat. Rev. Cancer* **5**(5), 388–396 (2005).
33. Tworoger, S. S. et al. The association of plasma DHEA and DHEA sulfate with breast cancer risk in predominantly premenopausal women. *Cancer Epidemiol. Biomark. Prev.* **15**, 967–971 (2006).
34. Lindström, S. et al. A comprehensive survey of genetic variation in 20,691 subjects from four large cohorts. *PLoS ONE* **12**, e0173997 (2017).
35. Stunkard, A. J., Sørensen, T. & Schulsinger, F. Use of the Danish Adoption Register for the study of obesity and thinness. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* **60**, 115–120 (1983).
36. Colditz, G. A. et al. Joanne Knight Breast Health Cohort at Siteman Cancer Center. *Cancer Causes Control* **33**, 623 (2022).
37. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559 (2007).
38. Randall, J. C. et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9**, <https://doi.org/10.1371/JOURNAL.PGEN.1003500> (2013).
39. Burgess, S., Scott, R. A., Timpson, N. J., Smith, G. D. & Thompson, S. G. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
40. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377 (2017).
41. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
42. Hartwig, F. P., Smith, G. D. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
43. Bowden, J. et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J. Epidemiol.* **47**, 1264–1278 (2018).
44. Foley, C. N., Mason, A. M., Kirk, P. D. W. & Burgess, S. MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates. *Bioinformatics* **37**, 531–541 (2021).



## Acknowledgements

The Nurses' Health Study II is supported by the National Cancer Institute (U01CA176726 and R01CA67262). This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming.

## Author contributions

P.K. and X.W. conceived and designed the study. R.M.T. prepared the phenotype data for NHS II. C.T. prepared the genotype data for NHS II. S.J. and G.C. prepared the mammogram risk score for NHS II. X.W. prepared the risk factors GWAS data. X.W. analyzed the data with the assistance of A.G. and C.T. X.W. and P.K. interpreted the results, with significant inputs and comments from R.M.T., S.J., and G.C. X.W. was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

## Funding

Open access funding provided by the National Institutes of Health.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41523-025-00870-4>.

**Correspondence** and requests for materials should be addressed to Peter Kraft.

**Reprints and permissions 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.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2025