International Journal of Epidemiology, 2020, 1117–1131
doi: 10.1093/ije/dyz242
Advance Access Publication Date: 23 December 2019
Original article

Mendelian Randomization

A Mendelian randomization analysis of circulating lipid traits and breast cancer risk

Alicia Beeghly-Fadiel,[1†] Nikhil K Khankari,[1†] Ryan J Delahanty,[2] Xiao-Ou Shu ,[1] Yingchang Lu,[1] Marjanka K Schmidt,[3,4] Manjeet K Bolla,[5] Kyriaki Michailidou,[5,6] Qin Wang,[5] Joe Dennis,[5] Drakoulis Yannoukakos,[7] Alison M Dunning,[8] Paul D P Pharoah,[5,8] Georgia Chenevix-Trench,[9] Roger L Milne,[10,11] David J Hunter,[12,13] Hall Per,[14] Peter Kraft,[12,13] Jacques Simard,[15] Douglas F Easton[5,8] and Wei Zheng [1]*; on behalf of the Breast Cancer Association Consortium

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA, [2]Tenet Healthcare, Nashville, TN, USA, [3]Division of Molecular Pathology, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, [4]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, [5]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK, [6]Department of Electron Microscopy/Molecular Pathology, Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, [7]Molecular Diagnostics Laboratory, INRASTES, National Centre for Scientific Research 'Demokritos', Athens, Greece, [8]Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK, [9]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, [10]Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, VIC, Australia, [11]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, VIC, Australia, [12]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [13]Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [14]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden and [15]Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, QC, Canada

*Corresponding author. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, 2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738, USA. E-mail: wei.zheng@vanderbilt.edu

[†]These authors contributed equally to this work.

Editorial decision 28 October 2019; Accepted 5 November 2019

Abstract

Background: Conventional epidemiologic studies have evaluated associations between circulating lipid levels and breast cancer risk, but results have been inconsistent. As Mendelian randomization analyses may provide evidence for causal inference,

we sought to evaluate potentially unbiased associations between breast cancer risk and four genetically predicted lipid traits.

Methods: Previous genome-wide association studies (GWAS) have identified 164 discrete variants associated with high density lipoprotein-cholesterol (HDL-C), low density lipoprotein-cholesterol (LDL-C), triglycerides and total cholesterol. We used 162 of these unique variants to construct weighted genetic scores (wGSs) for a total of 101 424 breast cancer cases and 80 253 controls of European ancestry from the Breast Cancer Association Consortium (BCAC). Unconditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI) for associations between per standard deviation increase in genetically predicted lipid traits and breast cancer risk. Additional Mendelian randomization analysis approaches and sensitivity analyses were conducted to assess pleiotropy and instrument validity.

Results: Corresponding to approximately 15 mg/dL, one standard deviation increase in genetically predicted HDL-C was associated with a 12% increased breast cancer risk (OR: 1.12, 95% CI: 1.08–1.16). Findings were consistent after adjustment for breast cancer risk factors and were robust in several sensitivity analyses. Associations with genetically predicted triglycerides and total cholesterol were inconsistent, and no association for genetically predicted LDL-C was observed.

Conclusions: This study provides strong evidence that circulating HDL-C may be associated with an increased risk of breast cancer, whereas LDL-C may not be related to breast cancer risk.

Key words: Breast cancer, lipids, cholesterol, genetics, Mendelian randomization, instrumental variable, epidemiology

Key Messages

- We conducted a large Mendelian randomization analysis to provide unbiased estimates of association with breast cancer risk for four lipid traits among 181 677 European-ancestry women from the Breast Cancer Association Consortium.
- One standard deviation increase (representing approximately 15 mg/dL) increase in genetically predicted high density lipoprotein-cholesterol (HDL-C) was associated with a 12% increased risk of breast cancer, whereas no consistent associations were found with low density lipoprotein-cholesterol, triglycerides or total cholesterol.
- This study suggests that circulating HDL-C levels may influence breast cancer susceptibility.

Introduction

Circulating lipids, including high density lipoprotein-cholesterol (HDL-C), low density lipoprotein-cholesterol (LDL-C), triglycerides and total cholesterol, have long been hypothesized to influence the risk of breast, colorectal and other common cancers.[1–4] Early prospective cohort studies reported inverse associations for total cholesterol and cancer risk.[2–4] However, these findings could be due to reverse causation, where disease development or progression leads to lower circulating cholesterol levels years before disease diagnosis.[5–7] It is also possible that confounding factors, such as smoking, alcohol consumption, and socioeconomic status may have biased associations reported in previous epidemiologic studies.[8,9]

The role of HDL-C in disease risk is controversial. Although it is an established risk factor for coronary heart disease,[10] large Mendelian randomization analyses have suggested that the association between low HDL-C and heart disease may not be causal.[11,12] Furthermore, clinical trials designed to increase circulating HDL-C levels pharmacologically have not demonstrated overall benefits in heart disease prevention.[13,14] With regard to breast cancer, multiple studies have found inverse associations between HDL-C and risk.[15,16] Contrary to these findings, HDL-C

was associated with increased breast cancer risk when repeated serum lipid measures were evaluated.[17] Given the current controversy regarding the association between circulating lipid traits and cancer in general, and with breast cancer in particular, this finding has faced scepticism.[18]

Due to methodological limitations such as reverse causation and confounding, it is unlikely that conventional observational studies can resolve the longstanding debate about the role of circulating lipids in breast cancer development. Mendelian randomization analyses can potentially overcome some of the limitations inherent in conventional epidemiologic studies. Taking advantage of the random assortment of alleles which occurs during gametogenesis, thereby resembling randomized clinical trials, Mendelian randomization analysis uses genetic data (i.e. single nucleotide polymorphisms, or SNPs) as genetic instruments to estimate exposures of interest for association analyses with disease outcomes. Results from a Mendelian randomization analysis may provide strong evidence for causality, if the genetic instruments used are associated with the exposure, only affect the outcome via the exposure and are not associated with any of the confounders of the exposure-outcome relationship.[19] To date, genome-wide association studies (GWAS) have linked circulating lipid traits to at least 157 genetic loci.[20,21] A recent Mendelian randomization analysis used only summary statistics approaches and reported that genetically increased LDL-C and HDL-C were associated with increased estrogen receptor (ER)-positive breast cancer risk.[22] We independently conducted a Mendelian randomization analysis that leveraged both individual-level and summary statistics data for lipid-associated variants, and created instrumental variables to evaluate shared genetic components and associations between four circulating lipid traits and breast cancer risk.

Methods

Study population

The Breast Cancer Association Consortium (BCAC) is an international collaboration initiated in 2005 to study genetic susceptibility to breast cancer. First, we included individual-level epidemiologic and genetic data from 62 846 breast cancer cases and 43 207 healthy controls of European ancestry from 67 BCAC studies; genetic data included more than 500 000 variants from a custom OncoArray platform that was designed to provide dense coverage across known cancer susceptibility loci as well as common variants.[23] Second, we included independent data from 38 578 cases and 37 046 BCAC controls that were genotyped on the Illumina iSelect genotyping Collaborative Oncological Gene-Environment Study (iCOGS) array [http://ccge.medschl.cam.ac.uk/research/consortia/icogs/].[24] Demographic and select patient characteristics were harmonized across BCAC studies according to a standardized protocol. All BCAC studies were approved by relevant institutional review boards, and all participants provided written informed consent.

Variant genotyping, imputation, and selection

Genetic variants associated with lipid traits were selected from the Global Lipids Genetics Consortium. The first lipid-trait GWAS included approximately 100 000 subjects of European ancestry and identified 102 genetic variants in 95 loci.[20] The second GWAS, conducted among 188 577 subjects predominantly of European ancestry, identified 83 additional variants in 62 loci, resulting in a total of 185 lipid-associated variants in 157 loci.[21] However, among the 83 variants, 21 were associated with more than one lipid trait and were not discrete. Thus, we identified a total of 164 unique lipid trait associated variants, of which 87 were genotyped by OncoArray and 75 were imputed with high information quality scores (mean $r^2 = 0.98$, range = 0.86–0.99). In iCOGS data, 39 selected variants were genotyped, and 123 were successfully imputed (mean $r^2 = 0.82$, range = 0.35–0.99). Two variants (rs2247056 and rs3177928) were not imputed in either dataset, providing a total of 162 in our analysis (Supplementary Table S1, available as Supplementary data at *IJE* online). Except for two variants (rs2814982 and rs2814944) in moderate linkage disequilibrium (LD; $r^2 = 0.51$), all included variants were independent ($r^2 < 0.1$). Because rs2814982 was associated with total cholesterol and rs2814944 was associated with HDL-C, both were retained in our analysis, as no instrumental variable included both SNPs. Thus, based on information available from published GWAS, our instrumental variables for HDL-C, LDL-C, triglycerides and total cholesterol included 74, 57, 43 and 74 variants, respectively.

Mendelian randomization analyses

Our primary analysis used individual-level data from BCAC iCOGS and OncoArray to generate weighted-genetic scores (wGSs) for four lipid traits (HDL-C, LDL-C, triglycerides and total cholesterol). For each lipid trait, we constructed instrumental variables as follows: $wGS=\sum_{i=1}^{n} \beta_{gx} * \alpha_i$, where β_{gx} represents the effect for the genetic variant (g) associated with an increase in lipid levels (x) and α_i is effect allele dosage for each genetic variant (ranging from 0 to 2 for each individual), for n genetic variants from the Global Lipids Genetics Consortium GWAS.[20,21]

Associations between lipid trait wGRs and breast cancer risk factors (conducted separately for iCOGS and OncoArray datasets) were assessed with linear or

logistic regression for continuous or categorical variables, respectively (Supplementary Table S2, available as Supplementary data at *IJE* online). Associations between lipid trait wGSs and breast cancer risk were estimated by odds ratios (ORs) and 95% confidence intervals (95% CIs) from unconditional logistic regression using individual-level data. Analyses were conducted separately for BCAC participants with iCOGS and OncoArray data (Supplementary Table S3, available as Supplementary data at *IJE* online), and then combined by random-effects or fixed-effects meta-analysis (Supplementary Table S4, available as Supplementary data at *IJE* online); Cochran's Q statistic was used to evaluate heterogeneity. Models were adjusted for age, principal components (PCs) for European ancestry (iCOGS: six PCs; OncoArray: 10 PCs), and either study site (iCOGS) or country (OncoArray), as previously described.[23,25] Additional adjustment included breast cancer risk factors that were associated with lipid trait wGSs. We assessed effect measure modification by menopausal status, age (dichotomized at 50 years) and body mass index (dichotomized at 30 kg/m^2) using likelihood ratio tests (LRT) for multiplicative interaction terms in nested models. Polytomous regression was employed to evaluate associations with estrogen receptor (ER) positive (+) and ER negative (-) breast cancer subtypes; tests of equivalence of beta coefficients across subtypes were used to evaluate heterogeneity.

To reduce correlation between instrumental variables, we also constructed amended wGSs that included only genetic variants that were exclusively associated with HDL-C (55 variants), LDL-C (44 variants) or triglycerides (20 variants) at a genome-wide significance level (Supplementary Table S5, available as Supplementary data at *IJE* online) and then re-evaluated associations with breast cancer risk (Supplementary Table S6, available as Supplementary data at *IJE* online). Because total cholesterol includes other lipid traits, no such amended wGS was created. Analyses were completed using SAS (version 9.4) and Stata (version 12.1).

Sensitivity analyses

In addition to individual-level analyses, we also conducted Mendelian randomization analysis using inverse-variance weighted summary statistics (Supplementary Table S7, available as Supplementary data at *IJE* online).[26] Four additional sensitivity analyses were used to assess the influence of genetic pleiotropy and validity of our genetic instruments. First, Mendelian randomization Egger (MR Egger) regression was employed to evaluate the presence of directional pleiotropy by testing whether the intercept was statistically different from zero, and to estimate a bias-reduced Mendelian randomization estimate from the regression slope.[27] Second, a weighted multivariable regression-based approach was used to assess the influence of potential pleiotropic effects of genetic variants included in each instrument on other lipid traits; specifically, we regressed beta-coefficients for associations between genetic variants and breast cancer risk (β_{BC}) on beta-coefficients between genetic variants and lipid traits (HDL-C: $\beta_{HDL\text{-}C}$, LDL-C: $\beta_{LDL\text{-}C}$, triglycerides: β_{TG}, and total cholesterol: β_{TC}), thereby adjusting for the associations between genetic variants and other lipid traits.[28,29] Third, we estimated associations using a weighted-median Mendelian randomization approach where we assumed that 50% of the variants included in each genetic instrument were invalid instruments (i.e. did not meet at least one of the three assumptions necessary for a valid instrumental variable); standard errors were estimated by bootstrapping and were subsequently used to calculate 95% CIs.[30] Fourth, we conducted a leave-one-out analysis where the Mendelian randomization association was re-estimated after removing the strongest SNP (as determined by the largest change in magnitude in comparison with results from instruments with all variants). Sensitivity analyses were conducted using the 'TwoSampleMR' package curated by MR-Base[31] using R version 3.5.1, R Foundation for Statistical Computing [https://www.r-project.org/]. Finally, visual representations of the IVW, MR Egger, and weighted-median approaches were created for comparison (Supplementary Figures S1–S4, available as Supplementary data at *IJE* online), and funnel plots for individual SNP MR estimates in relation to the inverse of the standard errors (Supplementary Figures S5–S8, available as Supplementary data at *IJE* online) were inspected for symmetry to indicate validity of our Mendelian randomization analysis.

Results

One genetically predicted standard deviation increase in HDL-C, LDL-C, triglycerides and total cholesterol was calculated to correspond to approximately 15, 37, 43 and 42 mg/dL increases, respectively. Associations between breast cancer risk factors and lipid trait wGSs were evaluated among all BCAC participants and among only controls (Supplementary Table S2). Several associations were identified; however, the only consistent association across the two populations and genotyping platforms was between increasing age and lower total cholesterol (iCOGS $P = 4.0 \times 10^{-4}$ and OncoArray $P = 0.01$). Similarly, the only consistent association among controls was between increasing age and lower triglycerides (iCOGs $P = 0.04$ and OncoArray $P = 0.03$).

Associations for each standard deviation increase in genetically predicted lipid trait from iCOGS and OncoArray genotyped BCAC participants (Supplementary Table S3) were combined by random-effects meta-analysis (Table 1).

Table 1. Overall and stratified analyses for the association between one standard deviation (SD) increase in genetically predicted lipid traits and breast cancer risk using individual-level iCOGS and OncoArray data summarized using random-effects meta-analysis in the Breast Cancer Association Consortium (BCAC)

Subgroup	Cases	Controls	High density lipoprotein (HDL-C) wGS[a]			Low density lipoprotein (LDL-C) wGS[a]			Triglycerides (TG) wGS[a]			Total cholesterol (TC) wGS[a]		
			OR[b]	95% CI[b]	*P*	OR[b]	95% CI[b]	*P*	OR[b]	95% CI[b]	*P*	OR[b]	95% CI[b]	*P*
All women	101 424	80 253	1.12	1.08–1.16	1.7×10^{-9}	1.00	0.96–1.04	0.88	0.93	0.85–1.01	7.5×10^{-2}	1.05	0.99–1.11	0.11
Menopausal status														
Premenopausal	20 782	17 902	1.14	0.96–1.35	0.13	1.00	0.92–1.08	0.89	0.90	0.75–1.08	0.24	1.08	1.00–1.17	0.04
Postmenopausal	43 787	38 847	1.11	1.05–1.17	3.2×10^{-4}	0.99	0.94–1.05	0.80	0.93	0.88–0.99	0.03	1.00	0.95–1.05	0.96
Test for interaction[c]			$P_{iCOGS} = 0.04$, $P_{Onco} = 0.23$			$P_{iCOGS} = 0.31$, $P_{Onco} = 0.41$			$P_{iCOGS} = 0.41$, $P_{Onco} = 0.05$			$P_{iCOGS} = 0.13$, $P_{Onco} = 0.37$		
Age														
<50 years	24 572	22 944	1.17	1.01–1.34	3.4×10^{-2}	0.96	0.90–1.03	0.30	0.90	0.78–1.05	0.17	1.05	0.98–1.12	0.20
≥50 years	71 098	51 700	1.11	1.06–1.16	2.3×10^{-6}	1.01	0.97–1.05	0.65	0.93	0.89–0.98	6.7×10^{-3}	1.04	0.99–1.08	0.12
Test for interaction[c]			$P_{iCOGS} = 0.07$, $P_{Onco} = 0.74$			$P_{iCOGS} = 0.14$, $P_{Onco} = 0.68$			$P_{iCOGS} = 0.78$, $P_{Onco} = 0.13$			$P_{iCOGS} = 0.71$, $P_{Onco} = 0.99$		
Body mass index (BMI)														
BMI <30 kg/m^2	51 114	41 713	1.14	1.08–1.20	1.0×10^{-6}	0.99	0.94–1.04	0.70	0.90	0.83–0.98	1.7×10^{-2}	1.04	0.97–1.12	0.29
BMI ≥30 kg/m^2	12 200	9507	1.06	0.95–1.18	0.31	0.95	0.85–1.05	0.31	1.07	0.95–1.21	0.29	0.96	0.87–1.06	0.41
Test for interaction[c]			$P_{iCOGS} = 0.67$, $P_{Onco} = 0.24$			$P_{iCOGS} = 0.16$, $P_{Onco} = 0.99$			$P_{iCOGS} = 0.22$, $P_{Onco} = 1.9 \times 10^{-2}$			$P_{iCOGS} = 0.10$, $P_{Onco} = 0.68$		
Estrogen receptor (ER) status														
ER- breast cancer	43 039	80 253	1.10	1.03–1.18	7.0×10^{-3}	0.99	0.93–1.05	0.73	0.94	0.87–1.01	8.3×10^{-2}	1.06	0.93–1.21	0.36
ER+ breast cancer	61 140	80 253	1.11	1.07–1.16	8.6×10^{-8}	0.99	0.95–1.03	0.63	0.91	0.85–0.98	1.4×10^{-2}	1.03	0.99–1.07	0.17
Test for heterogeneity[c]			$P_{iCOGS} = 0.95$, $P_{Onco} = 0.52$			$P_{iCOGS} = 0.50$, $P_{Onco} = 0.61$			$P_{iCOGS} = 0.97$, $P_{Onco} = 0.26$			$P_{iCOGS} = 0.16$, $P_{Onco} = 0.69$		

[a]Total number of variants included in each lipid-specific weighted-genetic score (wGS): HDL-C = 74, LDL-C = 57, TG = 43 and TC = 74.

[b]Odds ratios (OR) and 95% confidence intervals (CI) represent results from random-effects meta-analysis summarizing the associations from iCOGS and OncoArray datasets for one SD increase in wGS in relation to breast cancer risk among stratified groups. Polytomous regression was employed to estimate risk for ER+ and ER- breast cancer in the iCOGS and OncoArray datasets. All models were adjusted for age, BCAC study site (iCOGS) or study country (OncoArray), and either top six (iCOGS) or top 10 principal components for European ancestry (OncoArray).

[c]Interactions were assessed on a multiplicative scale using the likelihood ratio test (LRT) for nested models. Differences in stratum-specific ORs and 95% CIs for tumour subtype were assessed using a test for heterogeneity.

Among all women, increased HDL-C levels were associated with increased breast cancer risk (OR: 1.12, 95% CI: 1.08–1.16) in models that included adjustment for age, study site or country and principal components for European ancestry. We found no association for LDL-C (OR: 1.00, 95% CI: 0.96–1.04) modest risk reduction was suggested for increasing triglycerides (OR: 0.93, 95% CI: 0.85–1.01) and a modest increase in risk was suggested for increasing total cholesterol levels (OR: 1.05, 95% CI: 0.99–1.11). Further, we found no significant interactions by menopausal status, age or body mass index (BMI) among either iCOGS or Oncoarray genotyped participants for any lipid trait. For the HDL-C wGS, increased breast cancer risk was observed per one standard deviation increase among postmenopausal women (1.11, 95% CI: 1.05–1.17), women less than 50 years of age (1.17, 95% CI: 1.01–1.34), women age 50 or greater (1.11, 95% CI: 1.06–1.16) and non-obese (BMI $< 30\ kg/m^2$) women (1.14, 95% CI: 1.08–1.20). Associations were also observed for both ER- (1.10, 95% CI: 1.03–1.18) and ER+ (1.11, 95% CI: 1.07–1.16) breast cancers. On the contrary, one standard deviation increase in triglycerides was associated with reduced breast cancer risk among postmenopausal women (OR: 0.93, 95% CI: 0.88–0.99), women age 50 or greater (OR: 0.93, 95% CI: 0.89–0.98) and non-obese women (OR: 0.90, 95% CI: 0.83–0.98); the association was also observed for ER+ breast cancer (OR: 0.91, 95% CI: 0.85–0.91). In these stratified analyses, total cholesterol was associated with breast cancer risk only among premenopausal women (OR: 1.08, 95% CI: 1.00–1.17). Results were materially unaltered when fixed-effect meta-analyses were conducted (Supplementary Table S4).

In addition, we constructed amended instruments with reduced correlation by including only 55, 44 or 20 genetic variants that were exclusively associated with either HDL-C, LDL-C, or triglycerides (Supplementary Table S5); the amended HDL-C wGS was associated with increased breast cancer risk (OR: 1.14, 95% CI: 1.07–1.22) whereas the amended triglyceride wGS was not (OR: 1.00, 95% CI: 0.86–1.16) (Supplementary Table S6). Regardless of adjustment, or whether initial or exclusive variants were included, the LDL-C wGS was not associated with breast cancer risk in our analyses.

We also conducted Mendelian randomization analysis using summary statistics data and included several sensitivity analyses to assess the validity of our instrumental variables (Supplementary Table S7). The inverse-variance weighted Mendelian randomization estimate using summary statistics[26] per standard deviation (SD) increase in circulating lipids confirmed our initial findings: increased HDL-C was associated with increased breast cancer risk (OR_{IVW}: 1.12, 95% CI: 1.08–1.17) whereas increased LDL-C was not associated with breast cancer risk (OR_{IVW}: 0.99, 95% CI: 0.96–1.03). These associations were consistent for HDL-C and LDL-C regardless of dataset. On the contrary, triglycerides were associated with reduced breast cancer risk only in OncoArray data (OR_{IVW}: 0.88, 95% CI: 0.83–0.92) and total cholesterol was associated with increased risk only in iCOGS data (OR_{IVW}: 1.06, 95% CI: 1.01–1.12). The MR Egger regression intercept indicated that the IVW estimate for total cholesterol was potentially biased due to directional pleiotropy (iCOGS: $\beta_{intercept} = 0.0105$, P-value $= 7.3 \times 10^{-2}$; OncoArray: $\beta_{intercept} = 0.0053$, P value $= 7.3 \times 10^{-2}$). The bias-reduced estimate, derived from MR Egger regression, indicated a potential risk reduction for total cholesterol ($OR_{MR\ Egger}$: 0.92, 95% CI: 0.85–1.00). We also conducted weighted multivariable regression with mutual adjustment for other lipid traits.[28,29] Increasing HDL-C was associated with increased breast cancer risk in iCOGS data ($OR_{weighted\text{-}regression}$: 1.16, 95% CI: 1.08–1.25), and higher triglycerides were associated with decreased breast cancer risk in OncoArray data ($OR_{weighted\text{-}regression}$: 0.88, 95% CI: 0.81–0.95). Using a weighted-median approach, which assumes that half of included variants are invalid,[27] only HDL-C was associated with increased breast cancer risk after meta-analysis across our data sources ($OR_{weighted\text{-}median}$: 1.08, 95% CI: 1.02–1.14). Results from our leave-one-out analysis also yielded an association for HDL-C and breast cancer risk ($OR_{Leave\text{-}one\text{-}out}$: 1.13, 95% CI: 1.06–1.20). Finally, associations across approaches were compared visually (Supplementary Figures S1–S4), and symmetry of funnel plots supported the validity of our Mendelian randomization analysis (Supplementary Figures S5–S8).

Discussion

In this large-scale Mendelian randomization study using 162 lipid-associated GWAS variants, we found that higher levels of genetically predicted HDL-C were associated with an increased risk of breast cancer. This finding was robust and consistent across a variety of analytical approaches. Genetically predicted triglyceride and total cholesterol levels were also associated with breast cancer risk in some analyses, but these findings were not consistent and varied by data source and statistical adjustment. Genetically predicted LDL-C was not associated with breast cancer risk in any analyses. Traditional epidemiologic studies that have measured circulating lipids and evaluated breast cancer risk have had conflicting results, likely due to reverse causation, confounding and selection bias. By using a Mendelian randomization approach, we aimed to overcome limitations inherent in traditional studies and to provide strong evidence supporting a possibly causal association between high HDL-C levels and increased breast cancer risk.

Another Mendelian randomization analysis on lipids and breast cancer risk with BCAC data was recently published; their primary findings include an increased risk of ER-positive breast cancer risk per standard deviation of genetically raised HDL-C (OR: 1.13, 95% CI: 1.01–1.26) or LDL-C (OR: 1.14, 95% CI: 1.05–1.24).[22] Several methodological differences may explain why these results differ from ours, most notably for LDL-C. First, in addition to using summary statistics approaches, our analysis included individual-level BCAC data, which enabled us to control for potential confounding by breast cancer risk factors and to conduct stratified analyses. Second, we selected GWAS-significant variants from primary tables in published GWAS, which had slight differences in information available from the Global Lipids Genetics Consortium.[21] Third, although we both started with 185 variants in 157 loci, the number of SNPs included in the final genetic instruments differed considerably. Rather than a subset, we included all available independent variants in our primary instruments: for example, 57 versus 44 SNPs for LDL-C.

Other relevant publications include a study of serum lipids found that higher HDL-C was associated with increased breast cancer risk when serial measurements were assessed, but not when only one baseline measure was evaluated.[17] This contrasts with a meta-analysis of prospective studies that found modest inverse associations with breast cancer risk for both total cholesterol and HDL-C.[15] Given that circulating cholesterol levels are often decreased several years before cancer diagnosis, inverse associations for this trait could be attributable to bias from reverse causation. In addition, residual confounding from factors such as mammographic breast density or alcohol intake, and effect modification by menopausal status, may also likely influence associations between circulating lipids and breast cancer risk.[18]

Plasma lipoproteins transport triglycerides and cholesterol between the liver and tissues. HDL-C is the smallest and most dense lipoprotein, and accounts for approximately 30% of total cholesterol, with levels ranging between 40–60 mg/dL. Higher HDL-C concentrations are associated with better cardiovascular health and lower coronary heart disease risk.[10] However, recent Mendelian randomization analyses have suggested that high HDL-C may not be causally related to reduced coronary heart disease risk.[11,12] Furthermore, pharmacological interventions to increase HDL-C levels have not consistently translated to improved health outcomes,[13,14] and a consensus statement from the National Lipid Association concluded that HDL-C is not currently a therapeutic target.[10] Instead, measures of HDL functionality may be more important than absolute levels, as not all HDL-C functions the same way.[10] For example, oxidized HDL-C and HDL-C from patients with type 2 diabetes had greater capacity to promote proliferation, migration and metastasis of breast cancer cells.[32] Thus, in addition to a major role in reverse cholesterol transport and anti-atherogenic effects, HDL-C also seems to have other functions, including the potential to enhance proliferation of breast cancer cells.[32,33] These data provide possible biological mechanisms supporting the increased breast cancer risk seen with increasing levels of genetically predicted HDL-C in our study.

In the current study, associations for triglycerides and total cholesterol in relation to breast cancer risk were inconsistent. The total cholesterol wGS was associated with breast cancer risk only among iCOGS genotyped participants, and multivariate adjustment attenuated this association. Similarly, genetically predicted triglycerides were associated with reduced breast cancer risk only among OncoArray genotyped participants, and the exclusive variant instrument did not influence breast cancer risk. This suggests that some previously reported associations may be due to residual confounding, and that additional evaluation to understand these discrepant findings may be warranted.

Strengths of this study include a very large sample size, strong instrumental variables for all four lipid traits (F-statistics all >10: HDL-C = 190, LDL-C = 266, TG = 274 and TC = 364),[34] and multiple analytical approaches to assess instrument validity. We included 162 variants reported by the Global Lipids Genetics Consortium, which account for approximately 13.7%, 14.6%, 11.7% and 15.0% of the variance in HDL-C, LDL-C, triglycerides and total cholesterol, respectively.[20,21] Given the large number of variants used to construct our instruments, our wGSs are likely to be most strongly associated with lipids, and not as strongly associated with other traits, satisfying one of the assumptions for a valid Mendelian randomization analysis. Given that pleiotropy remains a concern for Mendelian randomization analyses, we carefully evaluated this possibility using several analytical approaches.[27,28] Associations from exclusive variant wGS, inverse-variance weighted Mendelian randomization and weighted-median regression analysis were consistent, showing associations between HDL-C and breast cancer risk. Furthermore, our estimates were consistent after considering the potential influence of pleiotropy via MR Egger regression and multivariable weighted-regression. In addition, our results were also unaltered whether fixed-effect or random-effect meta-analysis was conducted.

Limitations of our study include that we did not have direct measurements of circulating lipid levels from our study population to further confirm the validity of our instrumental variables. However, in Mendelian randomization analyses, it is preferable to use externally-derived

weights for constructing genetic scores rather than internally-derived weights from the same study population.[35] We included external weights from a single large GWAS that was conducted predominantly among Europeans,[20,21] and included only women of European descent in the current analysis. Additional lipid trait genetic variants have also been reported;[36] however, weights from this multi-ethnic GWAS would not be applicable to our Caucasian study population. Additional limitations include incomplete information on all confounding factors, and that we could not evaluate or adjust for all such possible covariates, so whether our findings were influenced by residual confounding or another potential source of systematic bias cannot be determined. However, when we adjusted for known breast cancer risk factors that were associated with our lipid trait wGSs, our results for HDL-C were unaltered. In addition, many of the variants included in our analysis were not directly genotyped. However, we used only imputed variants of high quality (iCOGS: mean $r^2 = 0.82$, range = 0.35–0.99; OncoArray; mean $r^2 = 0.98$, range = 0.86–0.99). Any misclassification of genetic variants would be expected to be non-differential regarding our outcome, which would be more likely to attenuate rather than amplify an association. We also note that our genetic instruments may still include correlated SNPs (with r^2 threshold <0.1), so future analyses may consider more stringent LD thresholds or pruning methods. Finally, multiple comparisons were made in our analysis. When we amended our significance threshold using a Bonferroni correction, our primary finding regarding HDL-C and increased breast cancer risk remained unaltered.

As increasing levels of HDL-C are generally thought to be healthier, an association with increased breast cancer risk was somewhat surprising. Although we could identify potential mechanisms in support of this finding from the literature, additional research to elucidate underlying factors is needed. For example, if lipid trait associations vary by sex, then using sex-specific weights would be preferable. Similarly, associations with breast cancer risk may vary by clinical characteristics, such as stage or grade of disease; future studies should be undertaken to address these possibilities. Additional future directions include a bidirectional Mendelian randomization analysis to test whether breast cancer risk GWAS variants are associated with lipid traits, and mediation analysis to evaluate whether covariates such as BMI or smoking are confounding or mediating factors in the relationship between lipids and breast cancer risk.

Our results suggest that increased HDL-C levels are associated with a 12% increased risk of breast cancer. Given contradictory evidence in terms of the beneficial effects of modifying HDL-C to affect cardiovascular disease risk, our findings provide some additional support against the broad use of therapeutic approaches to increase HDL-C in the general population. Instead, our results may be most useful for precision medicine, such as identifying women at increased risk of breast cancer as predicted by HDL-C-associated SNPs. In addition, although Mendelian randomization analysis may provide evidence for causality, interpreting our results as causal is not recommended; it is impossible to confirm that all three assumptions for Mendelian randomization assumptions were met, or that our findings were not influenced by pleiotropy.

In conclusion, our Mendelian randomization analysis of circulating lipids demonstrated that genetically predicted levels of increasing HDL-C were associated with increased breast cancer risk. Given the strong methodology used in this study, our results may help to clarify the inconsistencies observed across previous conventional observational studies and to support the hypothesis that circulating lipids may influence breast cancer risk.

Supplementary data

Supplementary data are available at IJE online.

Funding

This work was supported by a National Cancer Institute at the National Institutes of Health research grant (R37CA070867) and endowment funds for the Ingram Professorship and Anne Potter Wilson Chair. N.K.K. was supported by a National Institutes of Health training grant (R25CA160056). BCAC is funded by Cancer Research UK (C1287/A16563, C1287/A12014) and by the European Communityś Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009–223175) (COGS). Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009–223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defense (W81XWH-10–1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. The DRIVE Consortium was funded by U19 CA148065. The Australian Breast Cancer Family Study (ABCFS) was supported by the National Institutes of Health (USA) (UM1 CA164920). The content of this manuscript does not necessarily reflect the views or policies of the National Institutes of Health or any

of the collaborating centres in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Australia Fellow and M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society (grants NKI 2007–3839; 2009 4363) and BBMRI-NL, which is a Research Infrastructure financed by the Dutch government (NWO 184.021.007). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BBCS is funded by Cancer Research UK and Breakthrough Breast Cancer and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). E.J.S. (BIGGS) is supported by NIHR Comprehensive Biomedical Research Centre, Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London, UK; I.T. is supported by the Oxford Biomedical Research Centre. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). The CECILE study was funded by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire (ANSES) and Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital. The CNIO-BCS was supported by the Spanish Network on Rare Diseases (CIBERER) and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). The Human Genotyping-CEGEN Unit (CNIO) is supported by the Instituto de Salud Carlos III. The CTS was initially supported by the California Breast Cancer Act of 1993 and the California Breast Cancer Research Fund (contract 97–10500) and is currently funded through the National Institutes of Health (R01 CA77398). Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting programme mandated by California Health and Safety Code Section 103885. H.A.C. receives support from the Lon V Smith Foundation (LVS39420). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). The GC-HBOC was supported by Deutsche Krebshilfe (107 352). The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance and Institute of the RuhrUniversity Bochum (IPA), as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus Bonn, Germany. HEBCS was supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (266528), the Finnish Cancer Society, and the Sigrid Juselius Foundation. The HMBCS was supported by the Rudolf Bartling Foundation. Financial support for KARBAC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, the Gustav V Jubilee foundation and Bert von Kantzows Foundation. The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the Academy of Finland and by the strategic funding of the University of Eastern Finland. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia and the Cancer Foundation of Western Australia. LMBC is supported by the 'Stichting tegen Kanker' (232–2008 and 196–2010); D.L.s is supported by the FWO and the KULPFV/10/016-SymBioSysII and by an ERC consolidator grant. The MARIE study was supported by the Deutsche Krebshilfe e.V. (70–2892-BR I, 106332, 108253, 108419), the Hamburg Cancer Society, the German Cancer Research Center (DKFZ), and the Federal Ministry of Education and Research (BMBF) Germany (01KH0402). MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects '5x1000'). The MCBCS was supported by National Institutes of Health grants (CA116167, CA176785, CA192393) a National Institutes of Health Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. The MEC was support by National Institutes of Health grants (CA63464, CA54281, CA098758 and CA132839). The work of MTLGEBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the 'CIHR Team in Familial Risks of Breast Cancer' program (grant # CRN-87521) and the Ministry of Economic Development, Innovation and Export Trade (grant # PSR-SIIRI-701). The NBCS was supported by grants from the Norwegian Research council, (155218/V40, 175240/S10 to A.L.B.D., FUGE-NFR 181600/V11 to V.N.K. and a Swizz Bridge Award to ALBD). The NBHS was supported by National Institutes of Health (R01CA100374); biological sample preparation was conducted by the Survey and Biospecimen Shared Resource, which is supported by National Institutes of Health (P30 CA68485). The OBCS was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant number 250083, 122715 and Center of Excellence grant number 251314), the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation and the special Governmental EVO funds for Oulu University Hospital-based research activities. The OFBCR was supported by National Institutes of Health (UM1 CA164920). The content of this manuscript does not necessarily reflect the views or policies of the National Institutes of Health or any of the collaborating centres in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products or organizations imply endorsement by the US Government or the BCFR. The ORIGO study was supported by the Dutch Cancer Society (RUL 1997–1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS was funded by

Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The pKARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. The RBCS was funded by the Dutch Cancer Society (DDHK 2004–3124, DDHK 2009–4318). The SASBAC study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health and the Susan G Komen Breast Cancer Foundation. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and the Breast Cancer Now Tissue Bank. SEARCH is funded by Cancer Research UK (C490/A10124, C490/A16561) and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support for P.D.P.P. from the NHS in the East of England through the Clinical Academic Reserve. SKKDKFZS is supported by the DKFZ. The SZBCS was supported by Grant PBZ_KBN_122/P05/2004. The TNBCC was supported by: a Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), a grant from the Breast Cancer Research Foundation, a generous gift from the David F and Margaret T Grohne Family Foundation, the Hellenic Cooperative Oncology Group research grant (HR R_BG/04) and the Greek General Secretary for Research and Technology (GSRT) Program, Research Excellence II, the European Union (European Social Fund – ESF), and Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF) - ARISTEIA. The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. I.C.R. acknowledges NHS funding to the NIHR Biomedical Research Centre.

Acknowledgements

Other authors from the Breast Cancer Association Consortium: ABCTB Investigators[1]: Irene L Andrulis,[2] Natalia N Antonenkova,[3] Hoda Anton-Culver,[4] Volker Arndt,[5] Kristan J Aronson,[6] Paul L Auer,[7,8] Caroline Baynes,[9] Laura E Beane Freeman,[10] Matthias W Beckmann,[11] Sabine Behrens,[12] Javier Benitez,[13,14] Marina Bermisheva,[15] Carl Blomqvist,[16] Natalia V Bogdanova,[3,17,18] Stig E Bojesen,[19–21] Bernardo Bonanni,[22] Hiltrud Brauch,[23–25] Hermann Brenner,[5,25,26] Annegien Broeks,[27] Sara Y Brucker,[28] Thomas Brüning,[29] Barbara Burwinkel,[30,31] Qiuyin Cai,[32] Trinidad Caldés,[33] Federico Canzian,[34] Brian D Carter,[35] Jose E Castelao,[36] Jenny Chang-Claude,[12,37] Ting-Yuan David Cheng,[38] Christine L Clarke,[39] Don M Conroy[9], Fergus J Couch,[40] Angela Cox,[41] Simon S Cross,[42] Julie M Cunningham,[40] Kamila Czene,[43] Mary B Daly,[44] Peter Devilee,[45,46] Kimberly F Doheny,[47] Thilo Dörk,[18] Isabel dos-Santos-Silva,[48] Martine Dumont,[49] Miriam Dwek,[50] H Shelton Earp,[51] Diana M Eccles,[52] Heather Eliassen,[53,54] Christoph Engel,[55,56] Mikael Eriksson,[43] D Gareth Evans,[57,58] Peter A Fasching,[11,59] Jonine Figueroa,[60–62] Olivia Fletcher,[63] Henrik Flyger,[64] Lin Fritschi,[65] Manuela Gago-Dominguez,[66,67] Susan M Gapstur,[35] Montserrat García-Closas,[60] Mia M Gaudet,[35] Graham G Giles,[68,69] Mark S Goldberg,[70,71] David E Goldgar,[72] Anna González-Neira,[13] Grethe I Grenaker Alnæs,[73] Pascal Guénel,[74] Lothar Haeberle,[11] Eric Hahnen,[75–77] Christopher A Haiman,[78] Niclas Håkansson,[79] Ute Hamann,[80] Patricia Harrington,[9] Andreas Hartkopf,[24] Alexander Hein,[11] Belynda Hicks,[81] Peter Hillemanns,[18] Antoinette Hollestelle,[82] Maartje J Hooning,[82] Robert N Hoover,[60] Anthony Howell,[83] Guanmengqian Huang,[80] kConFab/Aocs Investigators,[84,85] Anna Jakubowska,[86] Wolfgang Janni,[87] Esther M John,[88] Nichola Johnson,[63] Kristine Jones,[81] Audrey Jung,[12] Rudolf Kaaks,[12] Pooja Middha Kapoor,[12] Siddhartha Kar,[9] Vesa Kataja,[89,90] Elza Khusnutdinova,[15,91] Cari M Kitahara,[60] Julia A Knight,[92,93] Yon-Dschun Ko,[94] Veli-Matti Kosma,[90,95,96] Stella Koutros,[10] Vessela N Kristensen,[73,97,98] France Labrèche,[99] Diether Lambrechts,[100,101] Flavio Lejbkowicz,[102] Annika Lindblom,[103] Sara Lindström,[104,105] Martha S Linet,[60] Jolanta Lissowska,[106] Wing-Yee Lo,[23,24] Sibylle Loibl,[107] Jirong Long,[32] Jan Lubinski,[86] Craig Luccarini,[9] Michael P Lux,[11] Robert J MacInnis,[68,69] Tom Maishman,[52,108] Ivana Maleva Kostovska,[109] Arto Mannermaa,[90,95,96] JoAnn E Manson,[54,110] Sara Margolin,[111] Tabea Maurer,[45] Dimitrios Mavroudis,[112] Alfons Meindl,[113] Usha Menon,[114] Jeffery Meyer,[40] Susan L Neuhausen,[115] Heli Nevanlinna,[116] Patrick Neven,[117] William G Newman,[57,58] Sune F Nielsen,[19,20] Børge G Nordestgaard,[19–21] Olufunmilayo I Olopade,[118] Andrew F Olshan,[119] Janet E Olson,[120] Håkan Olsson,[121] Curtis Olswold,[120] Nick Orr,[63] Charles M Perou,[122] Julian Peto,[48] Mila Pinchev,[102] Dijana Plaseska-Karanfilska,[109] Ross Prentice,[7] Nadege Presneau,[50] Katri Pylkäs,[123,124] Brigitte Rack,[125] Paolo Radice,[126] Gadi Rennert,[102] Hedy S Rennert,[102] Valerie Rhenius,[9] Atocha Romero,[33,127] Emmanouil Saloustros,[128] Dale P Sandler,[129] Elinor J Sawyer,[130] Rita K Schmutzler,[75–77] Andreas Schneeweiss,[30,131] Fredrick Schumacher,[132] Rodney J Scott,[133,134] Christopher Scott,[120] Mitul Shah,[9] Martha J Shrubsole,[32] Ann Smeets,[117] Melissa C Southey,[135] John J Spinelli,[136,137] Jennifer Stone,[69,138] Harald Surowy,[30,31] Anthony Swerdlow,[139,140] Rulla Tamimi,[53,54,104] Jack A Taylor,[129,141] Mary Beth Terry,[142] Daniel C Tessier,[143] Kathrin Thöne,[45] Madeleine M A Tilanus-Linthorst,[82] Rob A E M Tollenaar,[144] Ian Tomlinson,[145] Diana Torres,[80,146] Melissa A Troester,[119] Thérèse Truong,[74] Hans-Ulrich Ulmer,[147] Michael Untch,[148] Celine Vachon,[120] David Van Den Berg,[78] Elke M van Veen,[57,58] Daniel Vincent,[143] Clarice R Weinberg,[149] Camilla Wendt,[111] Hans Wildiers,[117] Robert Winqvist,[123,124] Alicja Wolk,[79] Xiaohong R Yang,[60] Argyrios Ziogas[4] and Elad Ziv[150]

[1]Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, NSW, Australia, [2]Department of Molecular Genetics, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, University of Toronto, ON, Canada, [3]N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus, [4]Department of Epidemiology, University of California Irvine, Irvine, CA, USA, [5]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, [6]Department of Public Health Sciences Cancer Research Institute, Queen's University, Kingston, ON, Canada, [7]Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [8]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA, [9]Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK, [10]Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, Bethesda, MD, USA, [11]Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany, [12]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, [13]Human Cancer Genetics Program, Spanish National Cancer Research Centre, Madrid, Spain, [14]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia, Spain, [15]Institute of Biochemistry and Genetics, Ufa Scientific Center of Russian Academy of Sciences, Ufa, Russia, [16]Department

of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, [17]Department of Radiation Oncology, Hannover Medical School, Hannover, Germany, [18]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany, [19]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [20]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [21]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [22]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia, Milan, Italy, [23]Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany, [24]University of Tübingen, Tübingen, Germany, [25]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, [26]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, [27]Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, [28]Department of Women's Health, University of Tübingen, Tübingen, Germany, [29]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Bochum, Germany, [30]Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany, [31]Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg, Germany, [32]Division of Epidemiology, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA, [33]Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain, [34]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, [35]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA, [36]Oncology and Genetics Unit, Instituto de Investigacion Biomedica (IBI) Galicia Sur, Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain, [37]Research Group Genetic Cancer Epidemiology, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [38]Division of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA, [39]Westmead Institute for Medical Research, University of Sydney, Sydney, NSW, Australia, [40]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, [41]Department of Oncology and Metabolism, Sheffield Institute for Nucleic Acids (SInFoNiA), University of Sheffield, Sheffield, UK, [42]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK, [43]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [44]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA, [45]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, [46]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, [47]Center for Inherited Disease Research (CIDR), Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, [48]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK, [49]Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, QC, Canada, [50]Department of Biomedical Sciences, Faculty of Science and Technology, University of Westminster, London, UK, [51]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [52]Cancer Sciences Academic Unit, University of Southampton, Southampton, UK, [53]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [54]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [55]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany, [56]LIFE - Leipzig Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany, [57]Division of Evolution and Genomic Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, [58]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester Academic Health Science Centre, Manchester, UK, [59]Division of Hematology and Oncology, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA, [60]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA, [61]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh Medical School, Edinburgh, UK, [62]Cancer Research UK Edinburgh Centre, Edinburgh, UK, [63]The Breast Cancer Now Toby Robins Research Centre, Institute of Cancer Research, London, UK, [64]Department of Breast Surgery, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [65]School of Public Health, Curtin University, Perth, WA, Australia, [66]Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain, [67]Moores Cancer Center, University of California San Diego, La Jolla, CA, USA, [68]Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, VIC, Australia, [69]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, VIC, Australia, [70]Department of Medicine, McGill University, Montréal, QC, Canada, [71]Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, Canada, [72]Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA, [73]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway, [74]Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France, [75]Center for Hereditary Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany, [76]Center for Integrated Oncology (CIO), University Hospital of Cologne, Cologne, Germany, [77]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany, [78]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, [79]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, [80]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany, [81]Cancer Genomics Research Laboratory (CGR), Division of Cancer Epidemiology and Genetics, Frederick National Laboratory for Cancer Research, Rockville, MD, USA, [82]Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands, [83]Institute of Cancer studies, University of Manchester, Manchester, UK, [84]Peter MacCallum Cancer Center, Melbourne, VIC, Australia, [85]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia,

[86]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland, [87]Department of Gynecology and Obstetrics, University Hospital Ulm, Ulm, Germany, [88]Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA, [89]Cancer Center, Kuopio University Hospital, Kuopio, Finland, [90]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland, [91]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia, [92]Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada, [93]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada, [94]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany, [95]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland, [96]Department of Clinical Pathology, Imaging Center, Kuopio University Hospital, Kuopio, Finland, [97]Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [98]Department of Clinical Molecular Biology, Oslo University Hospital, University of Oslo, Oslo, Norway, [99]Institut de recherche Robert-Sauvé en santé et en sécurité du travail, Montréal, QC, Canada, [100]VIB Center for Cancer Biology, VIB, Leuven, Belgium, [101]Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium, [102]Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel, [103]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, [104]Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [105]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA, [106]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center & Institute of Oncology, Warsaw, Poland, [107]German Breast Group, GmbH, Neu Isenburg, Germany, [108]Southampton Clinical Trials Unit, Faculty of Medicine, University of Southampton, Southampton, UK, [109]Research Centre for Genetic Engineering and Biotechnology 'Georgi D. Efremov', Macedonian Academy of Sciences and Arts, Skopje, Republic of Macedonia, [110]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [111]Department of Oncology - Pathology, Karolinska Institutet, Stockholm, Sweden, [112]Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece, [113]Department of Gynaecology and Obstetrics, Technical University Munich, Germany, [114]Gynaecological Cancer Research Centre, Women's Cancer, Institute for Women's Health, University College London, London, UK, [115]Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA, [116]Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, [117]Department of Oncology, Leuven Multidisciplinary Breast Center, University Hospitals Leuven, Leuven, Belgium, [118]Center for Clinical Cancer Genetics and Global Health, University of Chicago, Chicago, IL, USA, [119]Department of Epidemiology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [120]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA, [121]Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden, ,[122]Department of Genetics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [123]Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland, [124]Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland, [125]Department of Gynecology and Obstetrics, Ludwig-Maximilians University of Munich, Munich, Germany, [126]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS (Istituto Di Ricovero e Cura a Carattere Scientifico) Istituto Nazionale dei Tumori (INT), Milan, Italy, [127]Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain, [128]Hereditary Cancer Clinic, University Hospital of Heraklion, Heraklion, Greece, [129]Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA, [130]Research Oncology, Guy's Hospital, King's College London, London, UK, [131]National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany, [132]Department of Population and Quantitative Health Sciences, Case Western Research University, Cleveland, OH, USA, [133]Division of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, NSW, Australia, [134]Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, NSW, Australia, [135]Department of Pathology, University of Melbourne, Melbourne, VIC, Australia, [136]Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada, [137]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada, [138]Curtin UWA Centre for Genetic Origins of Health and Disease, Curtin University and University of Western Australia, Perth, WA, Australia, [139]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK, [140]Division of Breast Cancer Research, Institute of Cancer Research, London, UK, [141]Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA, [142]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA, [143]McGill University and Génome Québec Innovation Centre, Montréal, QC, Canada, [144]Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands, [145]Wellcome Trust Centre for Human Genetics and Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK, [146]Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota, Colombia, [147]Frauenklinik der Stadtklinik Baden-Baden, Baden-Baden, Germany, [148]Department of Gynecology and Obstetrics, Helios Clinics Berlin-Buch, Berlin, Germany, [149]Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA and [150]Institute for Human Genetics, UCSF Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA

BCAC thanks all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. This study would not have been possible without the contributions of the following: Per Hall (COGS), Douglas F Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A Eeles, Douglas F Easton, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis, Alison M Dunning, Andrew Lee and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C Tessier, Francois Bacot, Daniel Vincent, Sylvie

LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E Bojesen, Sune F Nielsen, Borge G Nordestgaard and the staff of the Copenhagen DNA laboratory, and Julie M Cunningham, Sharon A Windebank, Christopher A Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. The ABCFS wishes to thank Maggie Angelakos, Judi Maskiell and Gillian Dite. The ABCS would like to thank C Ellen van der Schoot, Femke Atsma and the Sanquin Bloodbank. The BBCC wishes to thank Matthias Rübner, Silke Landrith, Alexander Hein, Michael Schneider and Sonja Oeser. The BBCS thanks Eileen Williams, Elaine Ryder-Mills and Kara Sargus. The BIGGS wishes to thank Michael Kerin, Nicola Miller, Niall McInerney and Gabrielle Colleran. The BSUCH would like to thank Peter Bugert and the Medical Faculty Mannheim. The CECILE study thanks Pierre Kerbrat, Patrick Arveux, Romuald Le Scodan, Yves Raoul, Pierre Laurent-Puig and Claire Mulot. We thank the staff and participants of the Copenhagen General Population Study (CGPS) for the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Breast Cancer Group (DBCG) is acknowledged for the tumour information. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. CNIO-BCS wishes to thank Guillermo Pita, Charo Alonso, Daniel Herrero, Nuria Álvarez, Pilar Zamora, Primitiva Menendez and the Human Genotyping-CEGEN Unit (CNIO). The CTS Steering Committee includes Leslie Bernstein, Susan Neuhausen, James Lacey, Sophia Wang, Huiyan Ma, Yani Lu and Jessica Clague DeHart at the Beckman Research Institute of City of Hope, Dennis Deapen, Rich Pinder, Eunjung Lee and Fred Schumacher at the University of Southern California, Pam Horn-Ross, Peggy Reynolds, Christina Clarke Dur and David Nelson at the Cancer Prevention Institute of California, and Hoda Anton-Culver, Argyrios Ziogas and Hannah Park at the University of California Irvine. The ESTHER study thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier and Katja Butterbach. GC-HBOC thanks Heide Hellebrand, Stefanie Engert and GC-HBOC (Supported by Deutsche Krebshilfe). The GENICA network is acknowledged: Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany (HB, Wing-Yee Lo, Christina Justenhoven), Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany (YDK, Christian Baisch), Institute of Pathology, University of Bonn, Germany (Hans-Peter Fischer), Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ) Heidelberg, Germany (Ute Hamann), Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Germany (Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz), Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany (Volker Harth)]. HEBCS thanks Kristiina Aittomäki and Taru A Muranen. HMBCS thanks Natalia Antonenkova, Peter Hillemanns, Hans Christiansen and Johann H Karstens. KBCP wishes to thank Eija Myöhänen and Helena Kemiläinen. We wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study [which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)] for their contributions to this resource, and the many families who contribute to kConFab. LMBC thanks Gilian Peuteman, Dominiek Smeets, Thomas Van Brussel and Kathleen Corthouts. The MARIE study wishes to thank Dieter Flesch-Janys, Petra Seibold, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. MBCSG wishes to thank Siranoush Manoukian, Bernard Peissel and Daniela Zaffaroni of the Fondazione IRCCS Istituto Nazionale dei Tumori (INT); Paolo Peterlongo of the FIRC Institute of Molecular Oncology (IFOM); Monica Barile and Irene Feroce of the Istituto Europeo di Oncologia (IEO); and Loris Bernard and the personnel of the Cogentech Cancer Genetic Test Laboratory. MCBCS would like to thank Emily Hallberg, Curtis Olswold and Susan Slager. MCCS cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index. MTLGEBCS would like to thank Martine Tranchant (Cancer Genomics Laboratory, CHU de Québec Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management and skillful technical assistance. J.S. is Chairholder of the Canada Research Chair in Oncogenetics. Silje Nord (NBCS) has a carrier grant from the Health Region South East (HSØ, grant nr 2014061). The NBHS wishes to thank study participants and research staff for their contributions and commitment to the study. The OBCS wishes to thank Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri Otsukka and Kari Mononen. The OFBCR thanks Teresa Selander and Nayana Weerasooriya. ORIGO thanks E Krol-Warmerdam and J Blom for patient accrual, administering questionnaires and managing clinical information. The LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCDOC) with the help of Dr J Molenaar. PBCS wishes to thank Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao and Michael Stagner. The pKARMA study wishes to thank the Swedish Medical Research Council. The RBCS thanks Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huijskens, Annette Heemskerk and the Erasmus MC Family Cancer Clinic. The SASBAC study wishes to thank the Swedish Medical Research Council. The SBCS thanks Sue Higham, Helen Cramp, Ian Brock, Sabapathy Balasubramanian and Dan Connley. We wish to thank the SEARCH and EPIC teams. We thank all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to the SKKDKFZS study. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study, and the study participants, study staff and doctors, nurses and other health care providers and health information sources who have contributed to the study. WHI programme is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the programme possible. A full listing of WHI investigators can be found at: [http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf]. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. Finally, the authors wish to acknowledge Samantha P Stansel, Kim Kreth and Lisa Long for assistance with preparing and submitting this manuscript, and numerous colleagues in the Vanderbilt Division of

Epidemiology, for their many discussions and helpful insight into breast cancer epidemiology and Mendelian randomization analysis.

Author contributions

W.Z. conceived and directed the study. N.K.K., R.J.D. and A.B.F. conducted the statistical analysis. A.B.F., R.J.D., N.K.K. and W.Z. wrote the manuscript. D.F.E. led the BCAC and COGS. K.M., M.K.B., Q.W., J.D., R.L.M., M.K.S., J.S., P.D.P.P. and G.C.T. contributed significantly to the BCAC and COGS. All authors contributed to collection of the data and biological samples in the original studies, and data preparation for collaboration in BCAC. All authors reviewed the manuscript and approved its submission for publication.

Conflict of interest: None declared.

References

1. Hiatt RA, Friedman GD, Bawol RD, Ury HK. Breast cancer and serum cholesterol. *J Natl Cancer Inst* 1982;**68**:885–89.
2. Törnberg SA, Holm LE, Carstensen JM, Eklund GA. Risks of cancer of the colon and rectum in relation to serum cholesterol and beta-lipoprotein. *N Engl J Med* 1986;**315**:1629–33.
3. Williams RR, Sorlie PD, Feinleib M, McNamara PM, Kannel WB, Dawber TR. Cancer incidence by levels of cholesterol. *JAMA* 1981;**245**:247–52.
4. Schatzkin A, Hoover RN, Taylor PR *et al*. Serum cholesterol and cancer in the NHANES I epidemiologic follow-up study. National Health and Nutrition Examination Survey. *Lancet* 1987;**2**:298–301.
5. Rose G, Shipley MJ. Plasma lipids and mortality: a source of error. *Lancet* 1980;**1**:523–26.
6. International Collaborative Group . Circulating cholesterol level and risk of death from cancer in men aged 40 to 69 years: experience of an international collaborative group. *JAMA* 1982;**248**: 2853–59.
7. Winawer SJ, Flehinger BJ, Buchalter J, Herbert E, Shike M. Declining serum cholesterol levels prior to diagnosis of colon cancer. A time-trend, case-control study. *JAMA* 1990;**263**: 2083–85.
8. Davey Smith G, Shipley MJ, Marmot MG, Rose G. Plasma cholesterol concentration and mortality. The Whitehall Study. *JAMA* 1992;**267**:70–76.
9. Iribarren C, Reed DM, Burchfiel CM, Dwyer JH. Serum total cholesterol and mortality. Confounding factors and risk modification in Japanese-American men. *JAMA* 1995;**273**:1926–32.
10. Toth PP, Barter PJ, Rosenson RS *et al*. High-density lipoproteins: a consensus statement from the National Lipid Association. *J Clin Lipidol* 2013;**7**:484–525.
11. Voight BF, Peloso GM, Orho-Melander M *et al*. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* 2012;**380**:572–80.
12. Holmes MV, Asselbergs FW, Palmer TM *et al*. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* 2015;**36**:539–50.
13. Boekholdt SM, Arsenault BJ, Hovingh GK *et al*. Levels and changes of HDL cholesterol and apolipoprotein A-I in relation to risk of cardiovascular events among statin-treated patients: a meta-analysis. *Circulation* 2013;**128**:1504–12.
14. Kühnast S, Fiocco M, van der Hoorn JWA, Princen HMG, Wouter Jukema J. Innovative pharmaceutical interventions in cardiovascular disease: focusing on the contribution of non-HDL-C/LDL-C-lowering versus HDL-C-raising. A systematic review and meta-analysis of relevant preclinical studies and clinical trials. *Eur J Pharmacol* 2015;**763**(**Pt A**):48–63.
15. Touvier M, Fassier P, His M *et al*. Cholesterol and breast cancer risk: a systematic review and meta-analysis of prospective studies. *Br J Nutr* 2015;**114**:347–57.
16. Borgquist S, Butt T, Almgren P *et al*. Apolipoproteins, lipids and risk of cancer. *Int J Cancer* 2016;**138**:2648–56.
17. Martin LJ, Melnichouk O, Huszti E *et al*. Serum lipids, lipoproteins, and risk of breast cancer: a nested case-control study using multiple time points. *J Natl Cancer Inst* 2015;**107. doi**: 10.1093/jnci/djv032.
18. Byers T, Goff D. Breast cancer, heart disease, and whispering 'fire' in a public theater. *J Natl Cancer Inst* 2015;**107.** doi: 10.1093/jnci/djv076..
19. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89–98.
20. Teslovich TM, Musunuru K, Smith AV *et al*. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;**466**:707–13.
21. Willer CJ, Schmidt EM, Sengupta S *et al*.; Global Lipids Genetics Consortium, Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;**45**:1274–83.
22. Nowak C, Ärnlöv J. A Mendelian randomization study of the effects of blood lipids on breast cancer risk. *Nat Commun* 2018; **9**:3957.
23. Michailidou K, Lindström S, Dennis J *et al*. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;**551**:92–94.
24. Michailidou K, Beesley J, Lindstrom S *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.
25. Michailidou K, Hall P, Gonzalez-Neira A *et al*. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;**45**:353–61.
26. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;**37**:658–65.
27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–25.
28. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;**181**:251–60.
29. Burgess S, Dudbridge F, Thompson SG. Re: 'Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects'. *Am J Epidemiol* 2015;**181**:290–91.
30. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;**40**:304–14.
31. Hemani G, Zheng J, Elsworth B *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018;**7**:e34408.

32. Pan B, Ren H, He Y *et al*. HDL of patients with type 2 diabetes mellitus elevates the capability of promoting breast cancer metastasis. *Clin Cancer Res* 2012;**18**:1246–56.
33. Rotheneder M, Kostner GM. Effects of low- and high-density lipoproteins on the proliferation of human breast cancer cells in vitro: differences between hormone-dependent and hormone-independent cell lines. *Int J Cancer* 1989;**43**:875–79.
34. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011;**40**: 740–52.
35. Zhang B, Shu XO, Delahanty RJ *et al*. Height and breast cancer risk: evidence from prospective studies and Mendelian randomization. *J Natl Cancer Inst* 2015;**107**. doi: 10.1093/jnci/djv219.
36. Hoffmann TJ, Theusch E, Haldar T *et al*. A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet* 2018;**50**:401–13.