Combined associations of a polygenic risk score and classical risk factors with breast cancer risk

Pooja Middha Kapoor, M.P.H.[1, 2], Nasim Mavaddat, Ph.D.[3], Parichoy Pal Choudhury, Ph.D.[4], Amber N. Wilcox, M.P.H.[4, 5], Sara Lindström, Ph.D.[6, 7], Sabine Behrens, Ph.D.[1], Kyriaki Michailidou, Ph.D.[3, 8], Joe Dennis, M.Sc.[3], Manjeet K. Bolla, M.Sc.[3], Qin Wang, M.Sc.[3], Audrey Jung, Ph.D.[1], Zomoroda Abu-Ful, M.D.[9], Thomas Ahearn, Ph.D.[4], Irene L. Andrulis, Ph.D.[10, 11], Hoda Anton-Culver, Ph.D.[12], Volker Arndt, M.D.[13], Kristan J. Aronson, Ph.D.[14], Paul L. Auer, Ph.D.[15, 16], Laura E. Beane Freeman, Ph.D.[4], Heiko Becher, Ph.D.[17, 18], Matthias W. Beckmann, M.D.[19], Alicia Beeghly-Fadiel, Ph.D.[20], Javier Benitez, Ph.D.[21, 22], Leslie Bernstein, Ph.D.[23], Stig E. Bojesen, Dr. Med. Sci.[24-26], Hiltrud Brauch, Ph.D.[27-29], Hermann Brenner, M.D.[13, 29, 30], Thomas Brüning, M.D.[31], Qiuyin Cai, M.D.[20], Daniele Campa, Ph.D.[1, 32], Federico Canzian, Ph.D.[33], Angel Carracedo, M.D.[34, 35], Brian D. Carter, M.P.H.[36], Jose E. Castelao, M.D.[37], Stephen J. Chanock, M.D.[4], Nilanjan Chatterjee, Ph.D.[4, 38, 39], Georgia Chenevix-Trench, Ph.D.[40], Christine L. Clarke, Ph.D.[41], Fergus J. Couch, Ph.D.[42], Angela Cox, Ph.D.[43], Simon S. Cross, M.D.[44], Kamila Czene, Ph.D.[45], James Y. Dai, Ph.D.[15], H. Shelton Earp, M.D.[46], Arif B. Ekici, Ph.D.[47], A. Heather Eliassen, Sc.D. [48, 49], Mikael Eriksson, M.Sc.[45], D. Gareth Evans, M.D.[50, 51], Peter A. Fasching, M.D.[19, 52], Jonine Figueroa, Ph.D.[4, 53, 54], Lin Fritschi, M.B.B.S.[55], Marike Gabrielson, Ph.D.[45], Manuela Gago-Dominguez, M.D.[34, 56], Chi Gao, Sc.D.[49, 57], Susan M. Gapstur, Ph.D.[36], Mia M. Gaudet, Ph.D.[36], Graham G. Giles, Ph.D.[58-60], Anna González-Neira, Ph.D.[22], Pascal Guénel, Ph.D.[61], Lothar Haeberle, Ph.D.[62], Christopher A. Haiman, Sc.D.[63], Niclas Håkansson, Ph.D.[64], Per Hall, Ph.D.[45, 65], Ute Hamann, Ph.D.[66], Sigrid Hatse, Ph.D.[67],

Jane Heyworth, Ph.D.[68], Bernd Holleczek, Ph.D.[69], Robert N. Hoover, M.D.[4], John L. Hopper, Ph.D.[59], Anthony Howell, M.B.B.S.[70], David J. Hunter, Sc.D.[49, 57, 71], ABCTB Investigators[72], kConFab/AOCS Investigators[73, 74], Esther M. John, Ph.D.[75], Michael E. Jones, Ph.D.[76], Rudolf Kaaks, Ph.D.[1], Renske Keeman, M.Sc.[77], Cari M. Kitahara, Ph.D.[78], Yon-Dschun Ko, M.D.[79], Stella Koutros, Ph.D.[4], Allison W. Kurian, M.D.[75], Diether Lambrechts, Ph.D.[80, 81], Loic Le Marchand, M.D.[82], Eunjung Lee, Ph.D.[63], Flavio Lejbkowicz, Ph.D.[9], Martha Linet, M.D.[78], Jolanta Lissowska, Ph.D.[83], Ana Llaneza, Ph.D.[84], Robert J. MacInnis, Ph.D.[58, 59], Maria Elena Martinez, Ph.D.[56, 85], Tabea Maurer, Dipl. Psych.[86], Catriona McLean, M.D.[87], Susan L. Neuhausen, Ph.D.[23], William G. Newman, Ph.D.[50, 51], Aaron Norman, Ph.D.[88], Katie M. O'Brien, M.S.P.H.[89], Andrew F. Olshan, Ph.D.[90], Janet E. Olson, Ph.D.[88], Håkan Olsson, Ph.D.[91], Nick Orr, Ph.D.[92], Charles M. Perou, Ph.D.[93], Guillermo Pita, Ph.D.[94], Eric C. Polley, Ph.D.[88], Ross L. Prentice, Ph.D.[15], Gad Rennert, M.D., Ph.D.[9], Hedy S. Rennert, M.P.H.[9], Kathryn J. Ruddy, M.D.[95], Dale P. Sandler, Ph.D.[89], Christobel Saunders, M.B.B.S.[96], Minouk J. Schoemaker, Ph.D.[76], Ben Schöttker, Ph.D.[13, 97], Fredrick Schumacher, Ph.D.[98], Christopher Scott, M.S.[88], Rodney J. Scott, Ph.D.[99-101], Xiao-Ou Shu, M.D.[20], Ann Smeets, Ph.D.[102], Melissa C. Southey, Ph.D.[60, 103, 104], John J. Spinelli, Ph.D.[105, 106], Jennifer Stone, Ph.D.[59, 107], Anthony J. Swerdlow, D.Sc.[76, 108], Rulla M. Tamimi, Ph.D.[48, 49, 57], Jack A. Taylor, Ph.D.[89, 109], Melissa A. Troester, Ph.D.[90], Celine M. Vachon, Ph.D.[110], Elke M. van Veen, M.Sc.[50, 51], Xiaoliang Wang, Ph.D.[6, 7], Clarice R. Weinberg, Ph.D.[111], Caroline Weltens, Ph.D.[112], Walter Willett, Dr.P.H.[49, 113, 114], Stacey J. Winham, Ph.D.[115], Alicja Wolk, Dr.Med.Sci.[64, 116], Xiaohong R. Yang, Ph.D.[4], Wei Zheng, Ph.D.[20], Argyrios Ziogas, Ph.D.[12], Alison M. Dunning, Ph.D.[117],

Paul D.P. Pharoah, Ph.D.[3, 117], Marjanka K. Schmidt, Ph.D.[77, 118], Peter Kraft[, Ph.D.49, 57], Douglas F. Easton, Ph.D.[3, 117], Roger L. Milne, Ph.D.[58-60], Montserrat García-Closas, Dr.P.H.[4] †, Jenny Chang-Claude, Ph.D.[1, 86] †*

1. Division of Cancer Epidemiology. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
2. Faculty of Medicine. *University of Heidelberg*. Heidelberg: Germany; 69120.
3. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care. *University of Cambridge*. Cambridge: UK; CB1 8RN.
4. Division of Cancer Epidemiology and Genetics. *National Cancer Institute, National Institutes of Health, Department of Health and Human Services*. Bethesda, MD: USA; 20850.
5. Department of Epidemiology, Gillings School of Global Public Health. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
6. Department of Epidemiology. *University of Washington School of Public Health*. Seattle, WA: USA; 98195.
7. Public Health Sciences Division. *Fred Hutchinson Cancer Research Center*. Seattle, WA: USA; 98109.
8. Biostatistics Unit and the Cyprus School of Molecular Medicine. *The Cyprus Institute of Neurology & Genetics*. Nicosia: Cyprus; 1683.
9. Clalit National Cancer Control Center. *Carmel Medical Center and Technion Faculty of Medicine*. Haifa: Israel; 35254.
10. Fred A. Litwin Center for Cancer Genetics. *Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital*. Toronto, ON: Canada; M5G 1X5.
11. Department of Molecular Genetics. *University of Toronto*. Toronto, ON: Canada; M5S 1A8.
12. Department of Epidemiology, Genetic Epidemiology Research Institute. *University of California Irvine*. Irvine, CA: USA; 92617.
13. Division of Clinical Epidemiology and Aging Research, C070. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
14. Department of Public Health Sciences, and Cancer Research Institute. *Queen's University*. Kingston, ON: Canada; K7L 3N6.
15. Cancer Prevention Program. *Fred Hutchinson Cancer Research Center*. Seattle, WA: USA; 98109.
16. Zilber School of Public Health. *University of Wisconsin-Milwaukee*. Milwaukee, WI: USA; 53205.
17. Institute of Medical Biometry and Epidemiology. *University Medical Center Hamburg-Eppendorf*. Hamburg: Germany; 20246.
18. Institute of Biometry and Clinical Epidemiology. *Charité - Universitätsmedizin Berl*Berlin: Germany; 10117.
19. Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN. *University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg*. Erlangen: Germany; 91054.
20. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center. *Vanderbilt University School of Medicine*. Nashville, TN: USA; 37232.
21. *Centro de Investigación en Red de Enfermedades Raras (CIBERER)*. Madrid: Spain; 28029.
22. Human Cancer Genetics Programme. *Spanish National Cancer Research Centre (CNIO)*. Madrid: Spain; 28029.
23. Department of Population Sciences. *Beckman Research Institute of City of Hope*. Duarte, CA: USA; 91010.

24. Copenhagen General Population Study, Herlev and Gentofte Hospital. *Copenhagen University Hospital*. Herlev: Denmark; 2730.
25. Department of Clinical Biochemistry, Herlev and Gentofte Hospital. *Copenhagen University Hospital*. Herlev: Denmark; 2730.
26. Faculty of Health and Medical Sciences. *University of Copenhagen*. Copenhagen: Denmark; 2200.
27. *Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology*. Stuttgart: Germany; 70376.
28. iFIT-Cluster of Excellence. *University of Tuebingen*. Tuebingen: Germany; 72074.
29. German Cancer Consortium (DKTK). *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
30. Division of Preventive Oncology. *German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT)*. Heidelberg: Germany; 69120.
31. Institute for Prevention and Occupational Medicine of the German Social Accident Insurance. *Institute of the Ruhr University Bochum (IPA)*. Bochum: Germany; 44789.
32. Department of Biology. *University of Pisa*. Pisa: Italy; 56126.
33. Genomic Epidemiology Group. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
34. 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; 15706.
35. Centro de Investigación en Red de Enfermedades Raras (CIBERER) y Centro Nacional de Genotipado (CEGEN-PRB2). *Universidad de Santiago de Compostela*. Santiago De Compostela: Spain; 15782.
36. Behavioral and Epidemiology Research Group. *American Cancer Society*. Atlanta, GA: USA; 30303.
37. Oncology and Genetics Unit. *Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS*. Vigo: Spain; 36312.
38. Department of Biostatistics, Bloomberg School of Public Health. *John Hopkins University*. Baltimore, MD: USA; 21205.
39. Department of Oncology, School of Medicine. *John Hopkins University*. Baltimore, MD: USA; 21205.
40. Department of Genetics and Computational Biology. *QIMR Berghofer Medical Research Institute*. Brisbane, Queensland: Australia; 4006.
41. Westmead Institute for Medical Research. *University of Sydney*. Sydney, New South Wales: Australia; 2145.
42. Department of Laboratory Medicine and Pathology. *Mayo Clinic*. Rochester, MN: USA; 55905.
43. Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism. *University of Sheffield*. Sheffield: UK; S10 2TN.
44. Academic Unit of Pathology, Department of Neuroscience. *University of Sheffield*. Sheffield: UK; S10 2TN.
45. Department of Medical Epidemiology and Biostatistics. *Karolinska Institutet*. Stockholm: Sweden; 171 65.
46. Lineberger Comprehensive Cancer Center. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
47. Institute of Human Genetics, University Hospital Erlangen. *Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN*. Erlangen: Germany; 91054.
48. Channing Division of Network Medicine, Department of Medicine. *Brigham and Women's Hospital and Harvard Medical School*. Boston, MA: USA; 02115.
49. Department of Epidemiology. *Harvard T.H. Chan School of Public Health*. Boston, MA: USA; 02115.

50. Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health. *University of Manchester, Manchester Academic Health Science Centre*. Manchester: UK; M13 9WL.
51. North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine. *St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre*. Manchester: UK; M13 9WL.
52. David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology. *University of California at Los Angeles*. Los Angeles, CA: USA; 90095.
53. Usher Institute of Population Health Sciences and Informatics. *The University of Edinburgh Medical School*. Edinburgh: UK; EH16 4TJ.
54. *Cancer Research UK Edinburgh Centre*. Edinburgh: UK; EH4 2XR.
55. School of Public Health. *Curtin University*. Perth, Western Australia: Australia; 6102.
56. Moores Cancer Center. *University of California San Diego*. La Jolla, CA: USA; 92037.
57. Program in Genetic Epidemiology and Statistical Genetics. *Harvard T.H. Chan School of Public Health*. Boston, MA: USA; 02115.
58. Cancer Epidemiology Division. *Cancer Council Victoria*. Melbourne, Victoria: Australia; 3004.
59. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health. *The University of Melbourne*. Melbourne, Victoria: Australia; 3010.
60. Precision Medicine, School of Clinical Sciences at Monash Health. *Monash University*. Clayton, Victoria: Australia; 3168.
61. Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP). *INSERM, University Paris-Sud, University Paris-Saclay*. Villejuif: France; 94805.
62. Department of Gynaecology and Obstetrics, University Hospital Erlangen. *Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN*. Erlangen: Germany; 91054.
63. Department of Preventive Medicine, Keck School of Medicine. *University of Southern California*. Los Angeles, CA: USA; 90033.
64. Institute of Environmental Medicine. *Karolinska Institutet*. Stockholm: Sweden; 171 77.
65. Department of Oncology. *Södersjukhuset*. Stockholm: Sweden; 118 83.
66. Molecular Genetics of Breast Cancer. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
67. Laboratory of Experimental Oncology (LEO), Department of Oncology. *Leuven Cancer Institute, KU Leuven*. Leuven: Belgium; 3000.
68. School of Population and Global Health. *The University of Western Australia*. Perth, Western Australia: Australia; 6009.
69. *Saarland Cancer Registry*. Saarbrücken: Germany; 66119.
70. Division of Cancer Sciences. *University of Manchester*. Manchester: UK; M13 9PL.
71. Nuffield Department of Population Health. *University of Oxford*. Oxford: UK; OX3 7LF.
72. Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research. *University of Sydney*. Sydney, New South Wales: Australia; 2145.
73. Research Department. *Peter MacCallum Cancer Center*. Melbourne, Victoria: Australia; 3000.
74. Sir Peter MacCallum Department of Oncology. *The University of Melbourne*. Melbourne, Victoria: Australia; 3000.
75. Departments of Epidemiology & Population Health and of Medicine, Division of Oncology and Stanford Cancer Institute. *Stanford University School of Medicine*. Stanford, CA: USA; 94304.
76. Division of Genetics and Epidemiology. *The Institute of Cancer Research*. London: UK; SM2 5NG.
77. Division of Molecular Pathology. *The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital*. Amsterdam: The Netherlands; 1066 CX.
78. Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics. *National Cancer Institute*. Bethesda, MD: USA; 20892.

79. Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH. *Johanniter Krankenhaus*. Bonn: Germany; 53177.
80. VIB Center for Cancer Biology. *VIB*. Leuven: Belgium; 3001.
81. Laboratory for Translational Genetics, Department of Human Genetics. *University of Leuven*. Leuven: Belgium; 3000.
82. Epidemiology Program. *University of Hawaii Cancer Center*. Honolulu, HI: USA; 96813.
83. Department of Cancer Epidemiology and Prevention. *M. Sklodowska-Curie National Research Institute of Oncology*. Warsaw: Poland; 02-034.
84. General and Gastroenterology Surgery Service. *Hospital Universitario Central de Asturias*. Oviedo: Spain; 33011.
85. Department of Family Medicine and Public Health. *University of California San Diego*. La Jolla, CA: USA; 92093.
86. Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH). *University Medical Center Hamburg-Eppendorf*. Hamburg: Germany; 20246.
87. Anatomical Pathology. *The Alfred Hospital*. Melbourne, Victoria: Australia; 3004.
88. Department of Health Sciences Research. *Mayo Clinic*. Rochester, MN: USA; 55905.
89. Epidemiology Branch. *National Institute of Environmental Health Sciences, NIH*. Research Triangle Park, NC: USA; 27709.
90. Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
91. Department of Cancer Epidemiology, Clinical Sciences. *Lund University*. Lund: Sweden; 222 42.
92. Centre for Cancer Research and Cell Biology. *Queen's University Belfast*. Belfast, Ireland: UK; BT7 1NN.
93. Department of Genetics, Lineberger Comprehensive Cancer Center. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
94. Human Genotyping-CEGEN Unit, Human Cancer Genetic Program. *Spanish National Cancer Research Centre*. Madrid: Spain; 28029.
95. Department of Oncology. *Mayo Clinic*. Rochester, MN: USA; 55905.
96. School of Medicine. *University of Western Australia*. Perth, Western Australia: Australia.
97. Network Aging Research. *University of Heidelberg*. Heidelberg: Germany; 69115.
98. Department of Population and Quantitative Health Sciences. *Case Western Reserve University*. Cleveland, OH: USA; 44106.
99. Division of Molecular Medicine, Pathology North. *John Hunter Hospital*. Newcastle, New South Wales: Australia; 2305.
100. Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health. *University of Newcastle*. Callaghan, New South Wales: Australia; 2308.
101. Hunter Medical Research Institute. *John Hunter Hospital*. Newcastle, New South Wales: Australia; 2305.
102. Department of Surgical Oncology. *University Hospitals Leuven*. Leuven: Belgium; 3000.
103. Department of Clinical Pathology. *The University of Melbourne*. Melbourne, Victoria: Australia; 3010.
104. Cancer Epidemiology Centre. *Cancer Council Victoria*. Melbourne, Victoria: Australia; 3004.
105. Population Oncology. *BC Cancer*. Vancouver, BC: Canada; V5Z 1G1.
106. School of Population and Public Health. *University of British Columbia*. Vancouver, BC: Canada; V6T 1Z4.
107. The Curtin UWA Centre for Genetic Origins of Health and Disease. *Curtin University and University of Western Australia*. Perth, Western Australia: Australia; 6000.
108. Division of Breast Cancer Research. *The Institute of Cancer Research*. London: UK; SW7 3RP.
109. Epigenetic and Stem Cell Biology Laboratory. *National Institute of Environmental Health Sciences, NIH*. Research Triangle Park, NC: USA; 27709.

110. Department of Health Science Research, Division of Epidemiology. *Mayo Clinic*. Rochester, MN: USA; 55905.
111. Biostatistics and Computational Biology Branch. *National Institute of Environmental Health Sciences, NIH*. Research Triangle Park, NC: USA; 27709.
112. Leuven Multidisciplinary Breast Center, Department of Oncology. *Leuven Cancer Institute, University Hospitals Leuven*. Leuven: Belgium; 3000.
113. Department of Nutrition. *Harvard T.H. Chan School of Public Health*. Boston, MA: USA; 02115.
114. Channing Division of Network Medicine. *Brigham and Women's Hospital and Harvard Medical School*. Boston, MA: USA; 02115.
115. Department of Health Science Research, Division of Biomedical Statistics and Informatics. *Mayo Clinic*. Rochester, MN: USA; 55905.
116. Department of Surgical Sciences. *Uppsala University*. Uppsala: Sweden; 751 05.
117. Centre for Cancer Genetic Epidemiology, Department of Oncology. *University of Cambridge*. Cambridge: UK; CB1 8RN.
118. Division of Psychosocial Research and Epidemiology. *The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital*. Amsterdam: The Netherlands; 1066 CX.

Keywords: Polygenic risk score, Breast cancer, Risk factors, Genetic susceptibility, Gene-environment interactions, Risk prediction, Epidemiology

*Corresponding author.

†These authors contributed equally to this work.

Corresponding author:

Jenny Chang-Claude

Unit of Genetic Epidemiology, Division of Cancer Epidemiology

German Cancer Research Center (DKFZ)

Im Neuenheimer Feld 581, 69120 Heidelberg, Germany

Tel: +49-6221 42 2373, Fax: +49-6221 42 2203

E-mail: j.chang-claude@dkfz.de

Abstract

We evaluated the joint associations between a new 313-variant PRS (PRS_{313}) and questionnaire-based breast cancer risk factors for women of European ancestry, using 72,284 cases and 80,354 controls from the Breast Cancer Association Consortium. Interactions were evaluated using standard logistic regression, and a newly developed case-only method, for breast cancer risk overall and by estrogen receptor status. After accounting for multiple testing, we did not find evidence that per-standard deviation PRS_{313} odds ratio differed across strata defined by individual risk factors. Goodness-of-fit tests did not reject the assumption of a multiplicative model between PRS_{313} and each risk factor. Variation in projected absolute lifetime risk of breast cancer associated with classical risk factors was greater for women with higher genetic risk (PRS_{313} and family history), and on average 17.5% higher in the highest vs lowest deciles of genetic risk. These findings have implications for risk prevention for women at increased risk of breast cancer.

Precision prevention and early-detection of cancer is a key aim of cancer research and utilizes tools such as risk prediction models for risk stratification[1, 2]. Many breast cancer risk prediction models are focused either on classical risk factors or on inherited mutations causing a moderate-to-high risk of cancer, and do not include risk associated with common susceptibility variants[3]. Modeling the joint associations of genetic and classical risk factors could result in substantial improvement in risk stratification and therefore improved prevention and screening modalities for breast cancer[4-7] .

Combined associations of SNPs can be summarized by a polygenic risk score (PRS); women in the top 1% of the newly derived 313-SNP PRS(PRS_{313}) have a four-fold increased risk of breast cancer than women at population-average risk[8]. Previous studies, which evaluated combined associations between classical risk factors and breast cancer PRS based on 77 SNPs[9] and 24 SNPs[10], found weak or no evidence of departure from the multiplicative risk assumption for overall breast cancer. In the current study, we extend these analyses to assess the combined associations of the PRS_{313} and classical risk factors using data from the Breast Cancer Association Consortium (BCAC). This new PRS has been validated by prospective studies and shown to be more predictive than the previously reported 77-SNP PRS[11] for risk of breast cancer overall as well as for estrogen receptor (ER) subtype-specific breast cancer[8]. Additionally, this study found evidence of interaction for ER-positive disease between PRS_{313} and family history, indicating the need to consider the joint effects of these two factors[8].

Detailed information on study samples, genetic data and risk factor data is provided in the **Supplementary Materials**. Briefly, we performed analyses using data from women of European ancestry from 16 prospective cohorts, 14 population-based case-control studies and 16 non-population based studies included in BCAC **(Supplementary Table 1)**. Samples were genotyped

using two arrays, iCOGS[12] and OncoArray[13-15]. Risk factor data were derived with respect to a reference age (date at diagnosis for cases and date at interview for controls). Development of the PRS is briefly explained in **Supplementary Materials**[8]. We standardized the PRS to have unit standard deviation for the controls.

Departure from the assumption of multiplicative combined effects of standardized PRS_{313} and each risk factor was assessed using two methods, unconditional logistic regression model and likelihood ratio test, and a newly developed case-only method, which assumes independence between PRS and risk factors in the underlying population and has greater efficiency compared with logistic regression[16]. Individual models were fitted for each PRS-risk factor combination for overall and ER-specific breast cancer. Models were adjusted for reference age, study, and corresponding ten ancestry-informative principal components for each array. Array-specific results were meta-analyzed using a fixed-effect inverse-variance weighted method. To evaluate global goodness-of-fit of the multiplicative model between PRS_{313} and each risk factor, we performed the Hosmer-Lemeshow test using population-based studies. Moreover, we assessed goodness-of-fit at the extremes of the distribution (tails) using a tail-based test[17]. Using the iCARE-BPC3 model[4], we projected absolute lifetime risk of breast cancer for 50-year old White non-Hispanic US women up to age 80 years. We assessed the distribution of risk due to classical (i.e. menstrual/reproductive, and lifestyle) and modifiable risk factors, respectively, within categories of risk defined by genetic factors (i.e. breast cancer family history and PRS_{313}).

Associations between PRS_{313} and overall and ER-specific breast cancer risk are likely to be over-estimated because there was substantial overlap between the SNP discovery samples and our dataset (**Supplementary Figure 1)**. The number of cases and controls varied for each risk factor, ranging from 61,617 cases and 74,698 controls for ever parous to 14,576 cases and 19,640

controls for pack-years smoked for overall breast cancer risk **(Supplementary Table 2)**. Based on the population-based case-control and prospective cohort studies, the associations of the risk factors with overall and ER subtype-specific breast cancer were of the expected magnitude and direction **(Supplementary Table 3)**.

After accounting for multiple testing using Bonferroni adjustment ($p_{int} < 0.05/16 = 0.003$), none of the interactions between PRS_{313} and any classical risk factor was statistically significant except for family history **(Table 1)**. All statistical tests were two-sided. The observed interaction between PRS_{313} and family history for ER-positive breast disease is consistent with what has been previously published based on an overlapping dataset[8]. Such an interaction was also found for overall and ER-negative breast cancer risk. There was no evidence for a clear dose-response in the estimated ORs associated with classical risk factors when stratified by PRS percentiles (**Supplementary Figure 2-4**). Neither global nor tail-based goodness-of-fit tests supported departure from the multiplicative model for any risk factor, for both overall and ER-positive breast cancer (**Supplementary Table 4**). Goodness-of-fit tests were not performed for ER-negative breast cancer due to the relatively small sample size.

Lack of evidence for substantial departure from the multiplicative assumption between the PRS_{313} and risk factors using this large study implies that the absolute risk associated with each classical risk factor is greater for women with higher polygenic risk[5, 18]. This is illustrated by our projections, which show that the lifetime risk due to classical risk factors was higher with a wider variation across women who are at a higher risk due to genetic factors (PRS_{313} and family history) (**Figure 1A**), and consistent with a recent study of BMI combined with a measure of familial risk based on multi-generational family history[18]. The predicted average lifetime risk due to all classical risk factors for women in the lowest and highest deciles

of the genetic risk were 21.9% and 4.4%, respectively, so the difference in risk was17.5%. The difference in risk between these two deciles associated with the subset of modifiable risk factors was 16.5% (**Figure 1B**). However, the absolute risk projections shown in Figure 1 should be viewed with caution since they assume perfect model calibration. In addition, these absolute risk projections require validation.

Our analyses using the current PRS_{313} are based on a sample size three times larger than that used in previously published BCAC analyses[9], although the dataset for ER-negative breast cancer is still limited. Our previous work on the PRS_{313} development[8] and the current analyses are based on European ancestry and may not be generalizable to other populations, highlighting the need for more studies in populations of non-European or mixed ancestry.

Overall, the combined associations of the newly developed PRS_{313} and the classical risk factors on breast cancer risk are well explained by a multiplicative model, except for family history, and will inform the development of overall and ER-specific risk prediction models in future. Most importantly, our findings suggest that preventive strategies aimed at modifying individual risk factors could have stronger impact on absolute risk reduction for women at higher genetic risk.

Funding

This work was supported by following funding agencies.

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-

2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report.

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. 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 Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065.

The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers 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) Senior Principal Research Fellow. 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]. The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute (grant number Z01-CP010119), and the National Institute of Environmental Health Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BCEES was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). The BREast Oncology GAlician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. CBCS is funded by the Canadian Cancer Society (grant

313404) and the Canadian Institutes of Health Research. The CECILE study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), 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 and Gentofte Hospital. The CNIO-BCS was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The CTS was supported by the California Breast Cancer Act of 1993, the California Breast Cancer Research Fund (contract 97-10500) and the National Institutes of Health (R01 CA77398, K05 CA136967, UM1 CA164917, and U01 CA199277). Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. HAC receives support from the Lon V Smith Foundation (LVS39420). The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer

Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). 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 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, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. The GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. 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. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer Fund,

Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, 199600). G.C.T. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The MCBCS was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and 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 augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.The MEC was support by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. The NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family Registry

(NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers 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 Carolina Breast Cancer Study was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065 and the Breast Cancer Research Foundation. The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. PROCAS is funded from NIHR grant PGfAR 0707-10031. Prof. Dr. D Gareth Evans is supported by the NIHR Biomedical Research Centre in Manchester (IS-BRC-1215-20007). 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 (NIH) and the Susan G. Komen Breast Cancer Foundation. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and 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 PDPP from the NHS in the East of England through the Clinical Academic Reserve. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01.

Notes

Role of the funders: The funder had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript.

Acknowledgements: The authors thank all the researchers, clinicians, technicians, and administrative staff involved in the Breast Cancer Association Consortium. The authors would also like to thank all study participants, researchers, clinicians, technicians and administrative

staff who participated in the parent studies (ABCFS, ABCS, ABCTB, AHS, BBCC, BCEES, BCINIS, BREOGAN, CBCS, CECILE, CGPS, CNIO-BCS, CPSII, CTS, EPIC, ESTHER, GENICA, GESBC, KARMA, kConFab /AOCS, LMBC, MARIE, MCBCS, MCCS, MEC, MISS, MMHS, NBHS, NC-BCFR, NCBCS, NHS, NHS2, OFBCR, PBCS, PKARMA, PLCO, PROCAS, SASBAC, SBCS, SEARCH, SISTER, SMC, UCIBCS, UKBGS, USRT, WHI) and have enabled this work to be carried out. The authors thank the staff of the Center for Genetic Epidemiology Laboratory, the CNIO genotyping unit, the McGill University and Génome Québec Innovation Centre, and the Mayo Clinic Genotyping Core Facility. CPSII acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. The GENICA Network thanks Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany, German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - EXC 2180 – 390900677, Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany, Institute of Pathology, University of Bonn, Germany, Molecular Genetics of Breast Cancer, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany, and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany. KARMA and SASBAC studies thank the Swedish Medical Research Counsel. kConFab/AOCS wishes to thank 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. NHS and NHS2 would like to thank the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. PROCAS thanks NIHR for funding. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

Conflict of Interest: none declared

References

1. Rebbeck TR, Burns-White K, Chan AT*, et al.* Precision Prevention and Early Detection of Cancer: Fundamental Principles. Cancer Discov 2018;8(7):803-811.
2. Roberts MC. Implementation Challenges for Risk-Stratified Screening in the Era of Precision Medicine. JAMA Oncol 2018;4(11):1484-1485.
3. Cintolo-Gonzalez JA, Braun D, Blackford AL*, et al.* Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. Breast Cancer Res Treat 2017;164(2):263-284.
4. Choudhury PP, Wilcox AN, Brook MN*, et al.* Comparative validation of breast cancer risk prediction models and projections for future risk stratification. J Natl Cancer Inst 2019; 10.1093/jnci/djz113.
5. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. J Natl Cancer Inst 2014;106(11).
6. Lee A, Mavaddat N, Wilcox AN*, et al.* BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med 2019; 10.1038/s41436-018-0406-9.
7. Pharoah PD, Antoniou AC, Easton DF*, et al.* Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med 2008;358(26):2796-803.
8. Mavaddat N, Michailidou K, Dennis J*, et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet 2019;104(1):21-34.
9. Rudolph A, Song M, Brook MN*, et al.* Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. Int J Epidemiol 2018;47(2):526-536.
10. Maas P, Barrdahl M, Joshi AD*, et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol 2016;2(10):1295-1302.

11. Mavaddat N, Pharoah PD, Michailidou K*, et al.* Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 2015;107(5).
12. 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(4):353-61, 361e1-2.
13. Amos CI, Dennis J, Wang Z*, et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev 2017;26(1):126-135.
14. Michailidou K, Lindstrom S, Dennis J*, et al.* Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551(7678):92-94.
15. Milne RL, Kuchenbaecker KB, Michailidou K*, et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet 2017;49(12):1767-1778.
16. Meisner A, Kundu P, Chatterjee N. Case-Only Analysis of Gene-Environment Interactions Using Polygenic Risk Scores. Am J Epidemiol 2019;188(11):2013-2020.
17. Song M, Kraft P, Joshi AD*, et al.* Testing calibration of risk models at extremes of disease risk. Biostatistics 2015;16(1):143-54.
18. Hopper JL, Dite GS, MacInnis RJ*, et al.* Age-specific breast cancer risk by body mass index and familial risk: prospective family study cohort (ProF-SC). Breast Cancer Res 2018;20(1):132.

Table 1: Odds ratios and 95% confidence intervals for multiplicative interactions between the 313 SNP-polygenic risk score (continuous) and classical risk factors of breast cancer, overall and by estrogen receptor (ER) status

Risk Factors	Controls	Case-control logistic regression method*‡									Case-only linear regression method*†‡					
		Overall breast cancer risk			ER-positive breast cancer risk			ER-negative breast cancer risk			Overall breast cancer risk		ER-positive breast cancer risk		ER-negative breast cancer risk	
		Cases	OR_{int} (95% CI)	p_{int}	Cases	OR_{int} (95% CI)	p_{int}	Cases	OR_{int} (95% CI)	p_{int}	OR_{int} (95% CI)	p_{int}	OR_{int} (95% CI)	p_{int}	OR_{int} (95% CI)	p_{int}
Age at menarche (per 2 years)	64087	52170	1.01 (0.99-1.03)	0.26	36820	1.01 (0.99-1.03)	0.29	8323	1.01 (0.98-1.04)	0.55	1.01 (1.00-1.02)	0.22	1.01 (0.99-1.02)	0.37	1.02 (0.99-1.06)	0.21
Ever parous (yes/no)	72552	59298	0.97 (0.93-1.00)	0.07	41858	0.98 (0.94-1.02)	0.35	9273	0.98 (0.92-1.05)	0.57	0.97 (0.94-1.00)	0.08	0.99 (0.96-1.03)	0.72	1.00 (0.92-1.09)	0.97
Number of children (1,2,3,≥4)§	61654	48786	1.00 (0.99-1.02)	0.96	34666	1.00 (0.99-1.02)	0.73	7552	0.99 (0.96-1.02)	0.53	1.01 (0.99-1.02)	0.38	1.01 (1.00-1.03)	0.11	1.00 (0.97-1.04)	0.90
Age at FFTP (per 5 years)§	53042	41671	1.00 (0.99-1.02)	0.82	29601	1.00 (0.98-1.01)	0.68	6517	0.99 (0.96-1.02)	0.52	1.00 (0.98-1.01)	0.39	0.99 (0.97-1.00)	0.06	1.00 (0.97-1.03)	0.92
Breastfeeding (yes/no)§	37568	34199	1.02 (0.98-1.06)	0.44	24273	1.01 (0.96-1.05)	0.81	5548	1.01 (0.95-1.08)	0.74	1.02 (0.99-1.05)	0.17	1.02 (0.98-1.06)	0.36	1.02 (0.95-1.11)	0.55
Duration of breast feeding (per 12 months)§	26367	27741	1.00 (0.98-1.02)	0.71	19329	1.00 (0.97-1.02)	0.76	4669	0.99 (0.95-1.03)	0.57	1.01 (0.99-1.03)	0.32	1.01 (0.99-1.03)	0.57	0.99 (0.96-1.03)	0.77
Adult height (per 5 cm)	62414	54847	0.99 (0.98-1.00)	0.07	38730	0.99 (0.98-1.00)	0.04	8682	1.00 (0.98-1.02)	0.77	1.00 (0.99-1.01)	0.92	0.99 (0.98-1.01)	0.29	1.01 (0.99-1.03)	0.48
Premenopausal BMI (per 5 kg/m2)‖	15610	12837	0.98 (0.95-1.00)	0.08	8354	0.99 (0.96-1.02)	0.48	2333	0.95 (0.91-1.00)	0.04	0.97 (0.94-1.00)	0.02	1.00 (0.96-1.03)	0.77	0.95 (0.89-1.01)	0.10
Postmenopausal BMI (per 5 kg/m2)¶	46137	37088	1.01 (0.99-1.02)	0.49	27305	1.01 (0.99-1.02)	0.39	5260	1.01 (0.99-1.04)	0.36	1.01 (1.00-1.02)	0.29	1.01 (1.00-1.03)	0.08	0.99 (0.96-1.02)	0.45
Ever use of oral contraceptives (yes/no)	56768	44979	1.01 (0.98-1.04)	0.63	31640	1.02 (0.98-1.05)	0.36	7061	1.02 (0.97-1.08)	0.42	0.99 (0.97-1.02)	0.45	1.00 (0.97-1.02)	0.75	1.01 (0.95-1.08)	0.73
Current use of	20896	19047	1.07	0.02	14465	1.06	0.08	2761	1.05	0.49	1.00	0.93	0.98	0.32	1.04	0.59

EPT (yes/no)[¶,#]			(1.01-1.14)			(0.99-1.13)			(0.92-1.19)		(0.96-1.04)		(0.93-1.03)		(0.91-1.18)	
Current use of Estrogen-only therapy (yes/no)[¶,#]	20716	18716	0.97 (0.91-1.03)	0.33	14201	0.96 (0.90-1.03)	0.28	2733	1.06 (0.94-1.20)	0.37	0.96 (0.91-1.01)	0.09	0.94 (0.89-0.99)	0.03	1.08 (0.95-1.23)	0.26
Alcohol consumption (per 10g/day)	16851	14484	1.00 (0.97-1.02)	0.75	10253	0.98 (0.96-1.00)	0.07	2259	1.06 (1.01-1.11)	0.03	1.00 (0.99-1.02)	0.71	0.99 (0.97-1.01)	0.19	1.04 (1.00-1.08)	0.06
Current smoking (yes/no)[**]	56308	43303	1.04 (1.00-1.08)	0.07	30486	1.05 (1.00-1.10)	0.03	6813	1.05 (0.97-1.13)	0.25	1.02 (0.98-1.05)	0.42	1.02 (0.98-1.06)	0.40	1.03 (0.95-1.11)	0.52
Pack-years of smoking (per 10 pack-years)[††]	15990	11766	0.99 (0.98-1.01)	0.43	8268	0.99 (0.97-1.01)	0.19	1778	0.99 (0.96-1.02)	0.67	1.00 (0.99-1.01)	0.97	1.00 (0.99-1.01)	0.99	1.00 (0.97-1.03)	0.97
Family history in a first-degree relative (yes/no)[‡‡]	50955	42024	0.93 (0.89-0.96)	0.00003	28909	0.93 (0.90-0.97)	0.0008	6921	0.93 (0.87-0.99)	0.03	—	—	—	—	—	—

[*] Number of cases are same for case-control and case-only method

[†] The case-only analyses do not provide additional evidence to case-control analyses

[‡] Models are adjusted for reference age, study and ten ancestry-informative principal components

[§] Among parous women

[‖] Among premenopausal women

[¶] Among postmenopausal women

[#] Models used to assess association with the use of MHT have been further adjusted for former use of any MHT, and use of other MHT preparations than the MHT preparation of interest

[**] Models used to assess association with current smoking have been further adjusted for former smoking

[††] Among ever smoked

[‡‡] PRS and family history are not independent therefore, case-only analyses were not conducted for family history

OR_{int}: Interaction odds ratio (per SD of PRS_{313}), CI: confidence intervals, SNP: single nucleotide polymorphisms, FFTP: First full-term pregnancy, BMI: Body mass index, MHT: Menopausal hormonal therapy, EPT: Estrogen-progesterone therapy.

Figure 1: Distribution of absolute lifetime risk explained by a) all classical risk factors, b) modifiable classical risk factors within decile categories of genetic risk, due to 313-variant polygenic risk score (PRS) and family history, for 50-year old White non-Hispanic women in the United States before 80 years.

The solid horizontal lines represent the mean risk within each decile, while the dashed horizontal line across the plot represents the population lifetime mean risk (10.9%). Lifetime risk is estimated using the iCARE-BPC3 model and refers to absolute risk from age 50 to 80 years. The genetic component includes the 313-variant polygenic risk score and breast cancer family history. The classical risk factor component includes following risk factors: age at menarche, age at menopause, parity, age at first birth, height, body mass index, alcohol intake, smoking status, ever and current use of hormone replacement therapy (HRT), and HRT type among ever users. The modifiable classical risk factor component includes BMI, ever or current use of HRT, smoking status, and alcohol consumption. Outliers defined as points beyond 1.5 times the interquartile range below the first quartile or above the third quartile were excluded from the plot.

Supplementary Methods

Study participants

Analyses were conducted using data from 46 studies (16 prospective cohorts, 14 population-based case-control studies and 16 non-population based studies) participating in BCAC (Supplementary Table 1). Participants were excluded if they were male, were of non-European descent, had breast cancer of unknown invasiveness or had in-situ breast tumors. Women with unknown reference age (defined as age at diagnosis for cases and age at interview for controls) and women who had prevalent disease at the time of recruitment were also excluded from the analyses. After implementation of the above exclusion criteria, studies with at least 150 cases and 150 controls having genetic data and information on at least one of the lifestyle risk factor were included in the further analyses. All studies were approved by the relevant ethics committee and written informed consent was acquired from the study participants.

Genetic data

Two custom-made genotyping arrays: iSelect genotyping array (iCOGS) and OncoArray 500K (Oncoarray) were used to genotype the samples. Detailed information about genotyping and imputation can be found elsewhere [1-4]. Briefly, 28,176 cases and 32,209 controls of European ancestry were genotyped by the iCOGS array, containing 211,155 SNPs, and 44,109 cases and 48,145 controls were genotyped by Oncoarray, comprising of 533,000 SNPs, of which 230,000 SNPs served as "GWAS backbone" (Illumina HumanCore).

Epidemiological data

Epidemiological data from different studies was centrally quality controlled and harmonized to a common data dictionary and was derived with respect to a reference date (age at diagnosis for cases and age at interview for controls). The mean (standard deviation) of reference age in the iCOGS data set is 57.5 (11.3) years for cases and 56.8 (11.4) years for controls. In the OncoArray dataset, the mean (standard deviation)

reference age is 59.5 (11.7) years for cases and 57.3 (11.9) years for controls. The following lifestyle risk factors variables were used in the analysis: age at menarche (per 2 years), ever parous (yes or no), ever use of oral contraceptives (yes or no), adult body mass index (BMI) separately for pre- and postmenopausal women (per 5 kg/m^2), adult height (per 5 cm), lifetime alcohol consumption (per 10 g/day), current smoking (yes or no), and family history defined as family history of breast cancer in a first-degree relative (yes or no). Further reproductive variables, including number of full-term pregnancies (1, 2, 3 and ≥4), age at first full-term pregnancy (per 5 years), ever breastfed (yes or no), duration of breastfeeding (per 12 months) were assessed in parous women. Current use of combined estrogen-progesterone menopausal hormonal therapy (MHT) (yes or no) and current use of estrogen-only MHT (yes or no) were analyzed for postmenopausal women. Women were categorized as pre- and postmenopausal based on their self-reported menopausal status. In case of missing menopausal status, reference age (<54 years: premenopausal and ≥54 years: postmenopausal) was used as surrogate to assign menopausal status.

Creation of PRS

Detailed information on creation of PRS is explained in Mavaddat *et al.*[5]. Briefly, using the Breast Cancer Association Consortium data from 69 studies comprising of nearly 94,000 cases and 75,000 controls of European descent, a new 313-SNP PRS was developed. SNPs were sorted and ranked based on their p-value of the associations with overall breast cancer risk. SNPs were then filtered in linkage disequilibrium and correlation such that, uncorrelated SNPs with lowest p-values were taken forward. Two approaches were employed to the remaining SNPs after preliminary filtration: (i) hard thresholding and stepwise forward regression model and (ii) penalized lasso regression method. Effect estimates for all the SNPs chosen by these methods were assessed in a logistic regression model in order to develop a best PRS. For ER-subtype specific PRS, effect estimates were obtained from case-only lasso model, otherwise overall estimates were utilized. The best PRS was further validated in an independent dataset of 10 prospective studies (approximately 11,000 cases and 18,000 controls) and also using data from the UK Biobank cohort (nearly 3,000 breast cancer incident cases).

Individual PRS was derived using the formula

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

where β_k is per-allele log risk ratio (in this case, odds ratio) for breast cancer established with the minor allele of SNP k, x_k is the dosage of the allele for SNP k and n is the total number of SNPs (which is 313 in these analyses). The effect estimates used to construct the PRS_{313} are obtained from Supplementary Table 7 of Mavaddat *et al*.[5]. Subtype-specific PRSs were created by incorporating ER-subtype specific weights.

Overall, the 313-SNP PRS showed evidence of increased risk of overall breast cancer with an odds ratio (OR) of 1.65 (95% CI = 1.59-1.72) per 1 SD for the PRS. This PRS was found to be more predictive for ER-positive breast cancer risk with OR of 1.74 (95% CI = 1.66-1.82) per SD of PRS when compared to ER-negative breast cancer risk (OR = 1.65, 95% CI = 1.59-1.72).

Statistical analysis

Interaction odds ratio (OR) and 95% confidence interval were assessed using unconditional logistic regression and likelihood ratio tests. We also conducted a newly developed case-only method [6] to evaluate the departure from multiplicative model between polygenic risk score (PRS) and lifestyle risk factors. This method takes into account the independence between PRS and risk factors, and has been shown to be more efficient over the logistic regression. The interaction between PRS and risk factors is evaluated using a simple linear regression of the PRS on the risk factors in the sample of cases. To check the independence assumption between the PRS and classical risk factors, we calculated pair-wise Spearman correlations for all variables using unaffected controls (Supplementary Figure 5).

Individual models were fitted for each PRS-risk factor combination for overall and ER-specific breast cancer. The ER-specific PRS was used for interaction analyses of the corresponding ER-specific breast cancer risk. Each model was adjusted for reference age (date at diagnosis for cases and date at interview for controls), study and ten array-specific principal components. An indicator variable for study design was created (prospective cohort/population-based case-control vs. non-population-based studies). To

account for potential differential main effects of risk factors by study design (prospective cohort/population-based, non-population based), an interaction term between risk factor and the aforementioned indicator variable was also added to the model, along with main effects. Models assessing current use of menopausal hormonal therapy (MHT) by type (estrogen-progesterone combined (EPT) or estrogen-only therapy (ET)) were further adjusted for former use of any MHT and former use of MHT other than the one being assessed. The association analysis of current smoking was further adjusted for former smoking. Analyses were conducted separately for iCOGS and OncoArray and then results were meta-analyzed using fixed-effect inverse-variance method. Analyses were conducted using SAS 9.4 [7] and R version 3.4.4[8].

Using the population-based studies, we evaluated the goodness-of-fit of a multiplicative model between PRS_{313} and individual risk factors for overall and ER-positive breast cancer risk. Global goodness-of-fit was tested using the Hosmer-Lemeshow (HL) test to compare expected and observed risks by quantiles. Furthermore, goodness-of-fit was tested at the extremes of the distribution (tails) by using the tail-based goodness of fit [9]. Due to relatively small number of cases, goodness of fit was not tested for ER-negative breast cancer risk.

We used the iCARE-BPC3 model [10] to estimate the distribution of lifetime risk of breast cancer for 50-year old White non-Hispanic US women before attaining 80 years. For these calculations, we utilized an individual level reference dataset of risk factors representative of this population [11] as well as breast cancer incidence rates from the US National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER) (2015) and competing mortality rates from the Center for Disease Control (CDC) WONDER database (2015). We assume that the PRS is independent of the other risk factors, conditional on family history. The genetic risk score accounts for the attenuation of the family history association due to its correlation with the PRS.

For computing the genetic risk score, the log relative risks for all the risk factors except family history and PRS was set to zero. We categorized the population into deciles of the genetic risk score based on the 313-

SNP PRS and family history (i.e., presence or absence of breast cancer in first degree relatives) multiplied by the log-relative risk for family history. A new variable was created to record the decile specific average genetic risk score and included as a covariate in the model. The log relative risk for this new variable was set to 1 and the log relative risk for family history was set to 0.

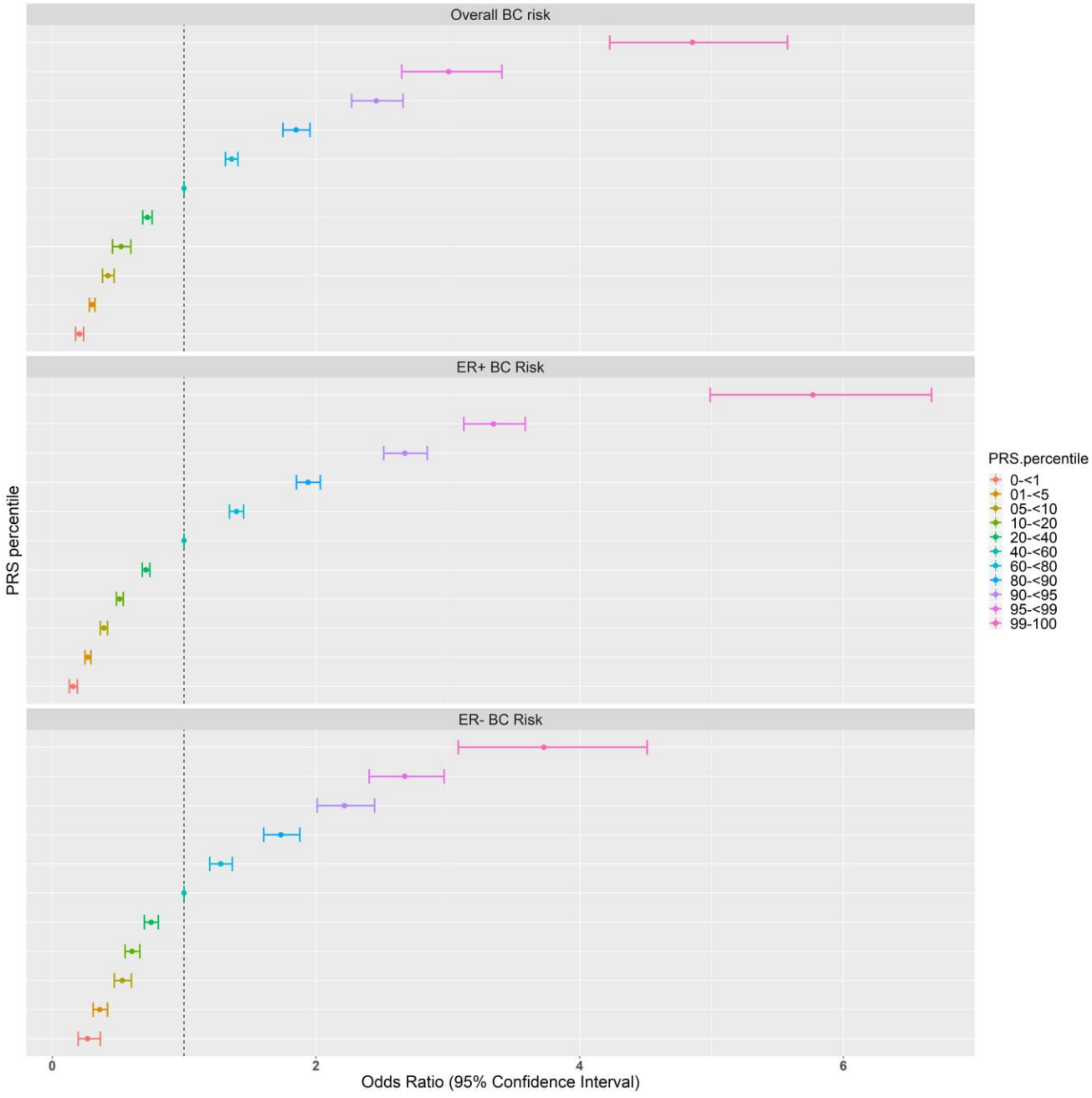
The recently developed Individualized Coherent Absolute Risk Estimator (iCARE) tool was used to perform the above calculations [12]. More specifically the computeAbsoluteRisk() function implemented within the iCARE tool was used. The log relative risk for the risk factors were obtained from Mass, P. et al. [10]. We fitted the multiplicative model presented in this paper and it included an interaction term between BMI and menopausal hormone therapy. Within each category of the genetic risk, we computed the absolute risk in the age range 50-80 years based on a) classical risk factors (i.e., all other risk factors excluding PRS and family history), and b) modifiable classical risk factors (BMI, use of hormonal replacement therapy, smoking status, and alcohol consumption) with the genetic risk score fixed at the category specific average. For computing the absolute risk based on modifiable risk factors, the log relative risk of all the other non-modifiable risk factors are set to zero. More details on the calculation of the absolute risk and the iCARE tool can be found elsewhere [12]. Figure 1a and 1b shows the distribution of this absolute lifetime risk within each category. In calculation of the absolute lifetime risk, we did not include the interaction between family history and PRS, therefore, the absolute lifetime risk may be slightly over-estimated for women with family history and high PRS.

References:

1. Amos CI, Dennis J, Wang Z, *et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev 2017;26(1):126-135.
2. 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(4):353-61, 361e1-2.
3. Michailidou K, Lindstrom S, Dennis J, *et al.* Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551(7678):92-94.
4. Milne RL, Kuchenbaecker KB, Michailidou K, *et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet 2017;49(12):1767-1778.
5. Mavaddat N, Michailidou K, Dennis J, *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet 2019;104(1):21-34.
6. Meisner A, Kundu P, Chatterjee N. Case-Only Analysis of Gene-Environment Interactions Using Polygenic Risk Scores. Am J Epidemiol 2019;188(11):2013-2020.
7. SAS Institute Inc. Cary NC. In.
8. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. In.
9. Song M, Kraft P, Joshi AD, *et al.* Testing calibration of risk models at extremes of disease risk. Biostatistics 2015;16(1):143-54.
10. Maas P, Barrdahl M, Joshi AD, *et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol 2016;2(10):1295-1302.
11. Choudhury PP, Wilcox AN, Brook MN, *et al.* Comparative validation of breast cancer risk prediction models and projections for future risk stratification. J Natl Cancer Inst 2019; 10.1093/jnci/djz113.
12. Choudhury PP, Maas P, Wilcox A, *et al.* iCARE: R package to build, validate and apply absolute risk models. 2018; 10.1101/079954 %J bioRxiv:079954.

Supplementary Figures

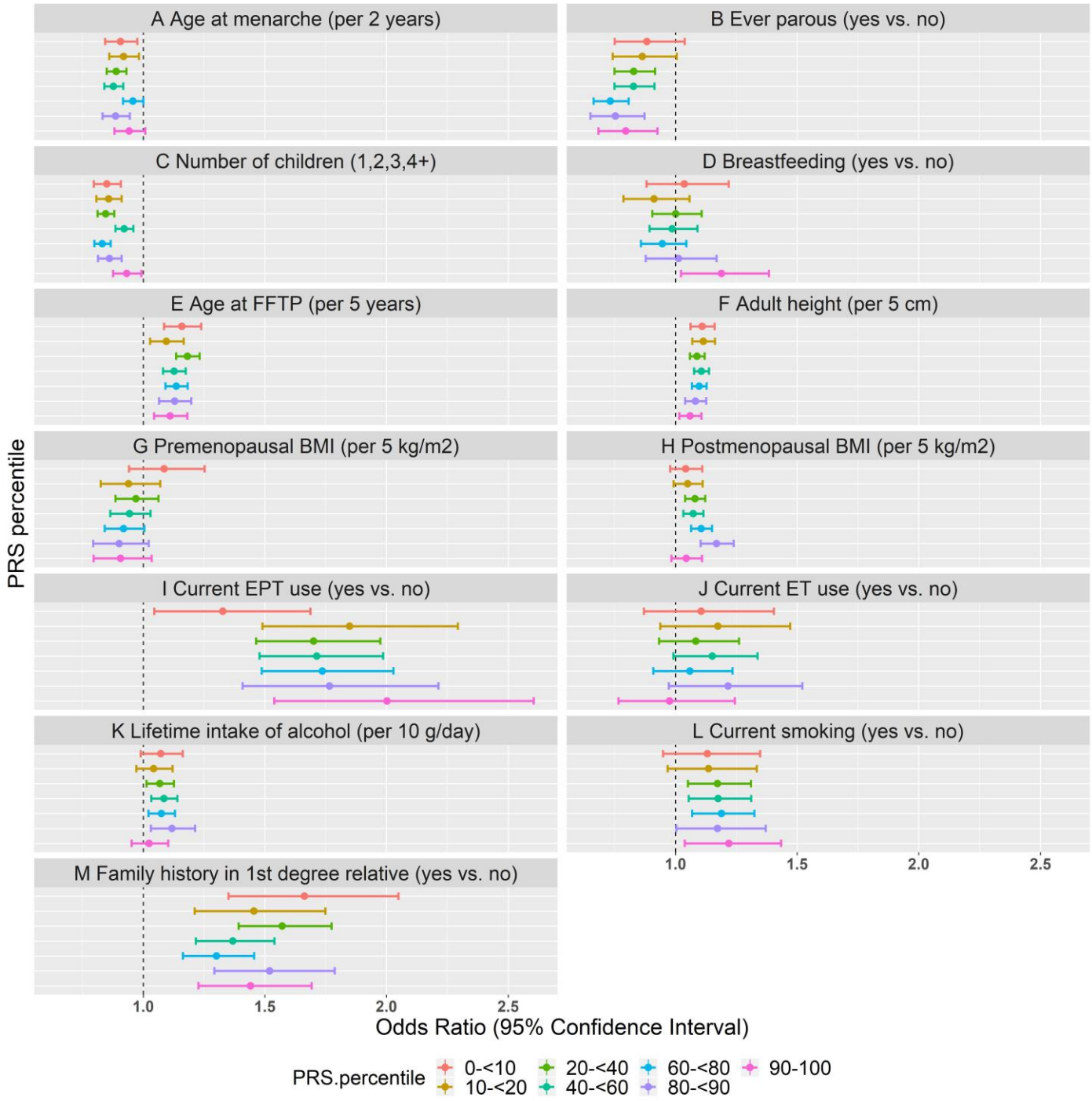
Supplementary figure 1: Associations of main effect of the PRS (by percentiles) on overall and estrogen receptor (ER)-subtype breast cancer risk in this dataset



BC: Breast cancer, PRS: Polygenic risk score, ER: estrogen receptor, ER+: Estrogen receptor positive, ER-: Estrogen receptor negative.

X-axis shows the odds ratio and y-axis shows the PRS percentiles. The legend on the right side shows the corresponding color scheme of PRS percentiles.

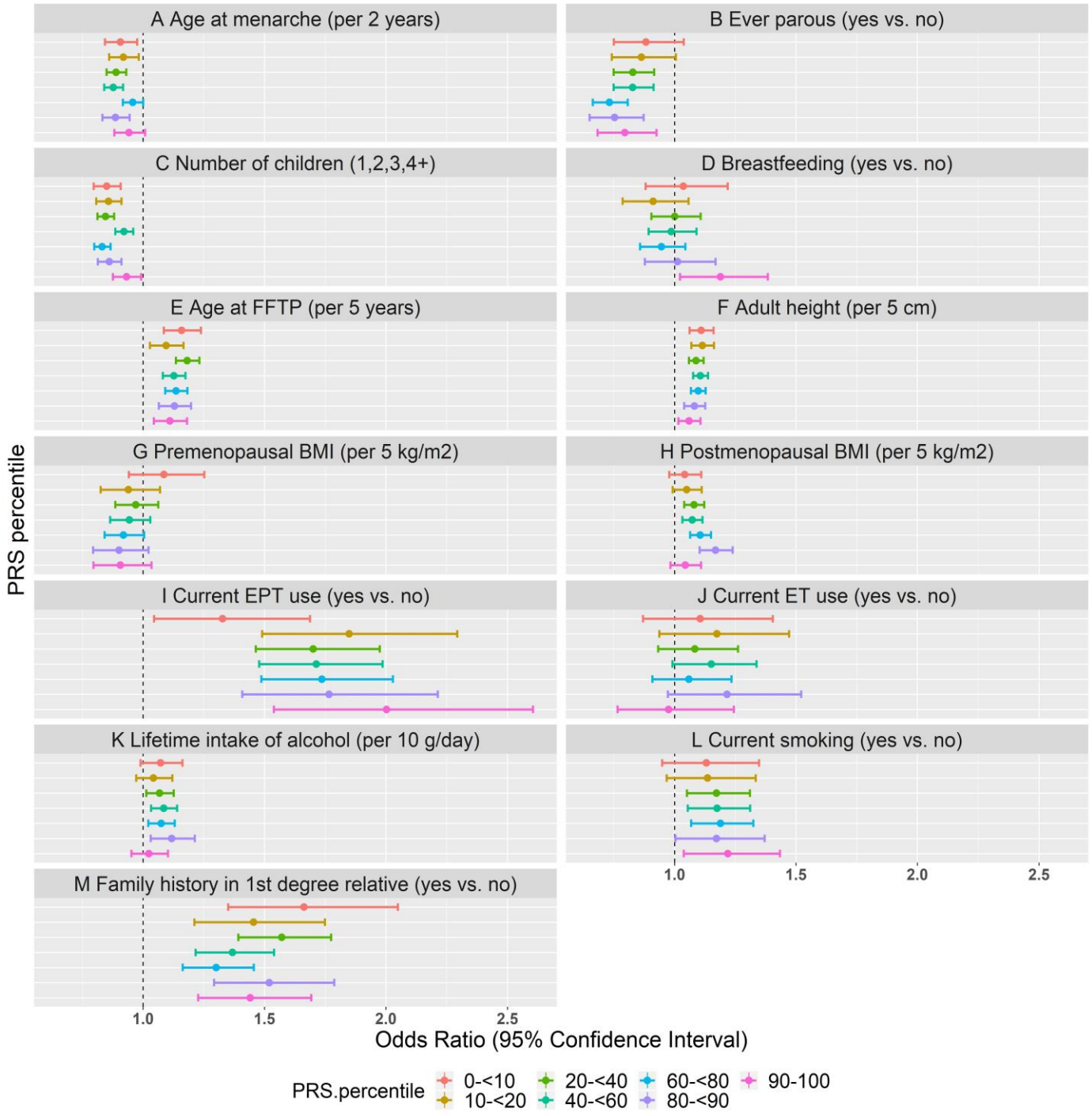
Supplementary Figure 2: Odd ratios and 95% confidence intervals for classical risk factors by percentiles of the 313-SNP polygenic risk score for overall breast cancer risk



PRS: Polygenic risk score, FFTP: First full-term pregnancy, BMI: Body mass index, EPT: Estrogen-progesterone therapy, ET: Estrogen-only therapy

X-axis shows the odds ratio and y-axis shows the PRS percentiles. The legend on the right side shows the corresponding color scheme of PRS percentiles.

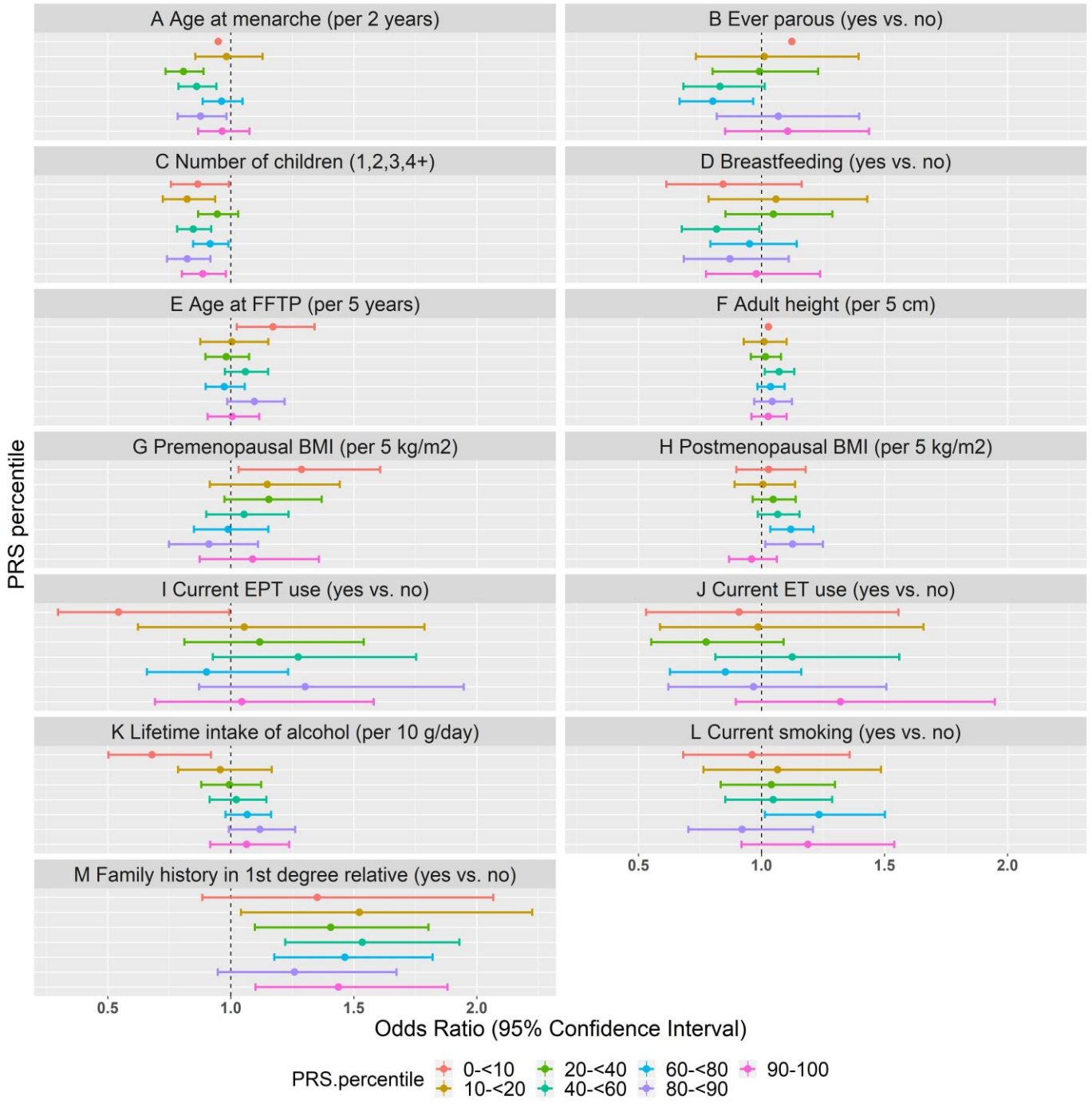
Supplementary Figure 3: Odd ratios and 95% confidence intervals for classical risk factors by percentiles of the 313-SNP polygenic risk score for ER-positive breast cancer risk



PRS: Polygenic risk score, FFTP: First full-term pregnancy, BMI: Body mass index, EPT: Estrogen-progesterone therapy, ET: Estrogen-only therapy

X-axis shows the odds ratio and y-axis shows the PRS percentiles. The legend on the right side shows the corresponding color scheme of PRS percentiles.

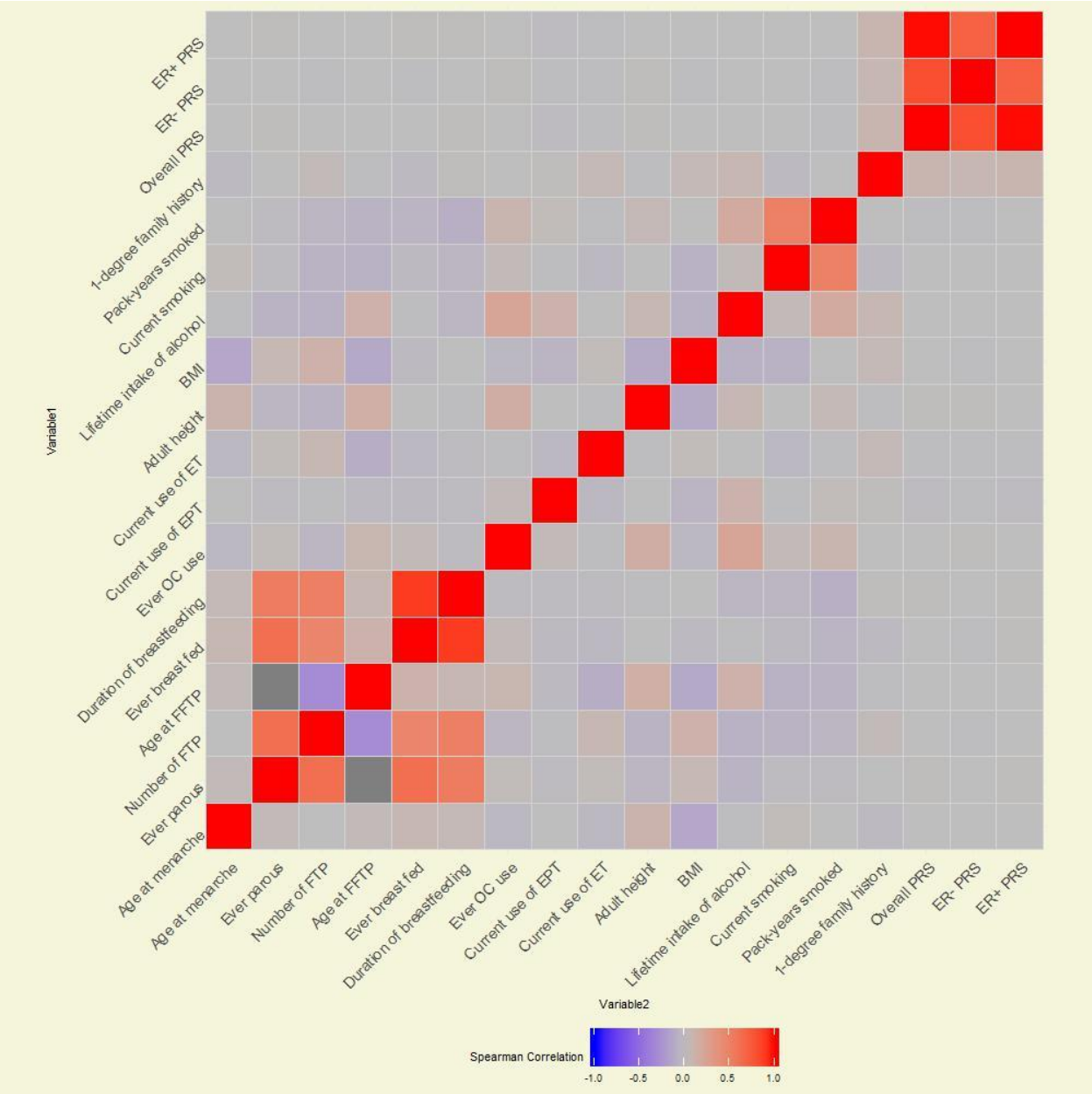
Supplementary Figure 4: Odd ratios and 95% confidence intervals for classical risk factors by percentiles of the 313-SNP polygenic risk score for ER-negative breast cancer risk



PRS: Polygenic risk score, FFTP: First full-term pregnancy, BMI: Body mass index, EPT: Estrogen-progesterone therapy, ET: Estrogen-only therapy

X-axis shows the odds ratio and y-axis shows the PRS percentiles. The legend on the right side shows the corresponding color scheme of PRS percentiles.

Supplementary Figure 5: Heatmap of Spearman pairwise correlation between PRS_{313} (overall and ER-subtype) and all classical risk factors (high positive correlation: red, high negative correlation: blue) using controls.



ER+: Estrogen receptor positive, ER-: Estrogen receptor negative, PRS: Polygenic risk score, BMI: Body mass index, ET: Estrogen-only menopausal hormonal therapy, EPT: Combined estrogen-progesterone therapy, OC: Oral contraceptives, FFTP: First full-term pregnancy, FTP: Full-term pregnancies, 1-degree family history: Family history in first degree relative.

Supplementary Tables

Supplementary Table 1: List of participating studies with number of total cases and controls

Study name	*Study acronym*	*Country*	*Study Design*[1]	*Cases*	*Controls*
Australian Breast Cancer Family Study	ABCFS	Australia	Population-based case-control study	1317	738
Amsterdam Breast Cancer Study	ABCS	Netherlands	Non population-based study	442	1376
Australian Breast Cancer Tissue Bank	ABCTB	Australia	Non population-based study	947	375
Agricultural Health Study	AHS	USA	Prospective cohort study	513	1137
Bavarian Breast Cancer Cases and Controls	BBCC	Germany	Non population-based study	809	706
Breast Cancer Employment and Environment Study	BCEES	Australia	Population-based case-control study	783	834
Breast Cancer in Northern Israel Study	BCINIS	Israel	Population-based case-control study	1315	724
Breast Oncology Galicia Network	BREOGAN	Spain	Non population-based study	1265	725
Canadian Breast Cancer Study	CBCS	Canada	Population-based case-control study	568	817
CECILE Breast Cancer Study	CECILE	France	Population-based case-control study	910	1002
Copenhagen General Population Study	CGPS	Denmark	Non population-based study	4064	5241
Spanish National Cancer Centre Breast Cancer Study	CNIO-BCS	Spain	Non population-based study	746	829
Cancer Prevention Study-II Nutrition Cohort	CPSII	USA	Prospective cohort study	2546	3323
California Teachers Study	CTS	USA	Prospective cohort study	1156	610
European Prospective Investigation Into Cancer and Nutrition	EPIC	France, Germany, Greece, Italy, Spain, The Netherlands, and UK	Prospective cohort study	3436	3597
ESTHER Breast Cancer Study	ESTHER	Germany	Population-based case-control study	476	505
Gene Environment Interaction and Breast Cancer in Germany	GENICA	Germany	Population-based case-control study	912	710
Genetic Epidemiology Study of Breast Cancer by Age 50	GESBC	Germany	Population-based case-control study	316	181
Karolinska Mammography Project for Risk Prediction of Breast Cancer - Cohort Study	KARMA	Sweden	Prospective cohort study	1415	6026
Kathleen Cuningham Foundation Consortium for research into Familial Breast Cancer/Australian Ovarian Cancer Study	KCONFAB/AOCS	Australia and New Zealand	Non population-based study	251	896
Leuven Multidisciplinary Breast Centre	LMBC	Belgium	Non population-based study	3003	1821

Mammary Carcinoma Risk Factor Investigation	MARIE	Germany	Population-based case-control study	1643	2065
Mayo Clinic Breast Cancer Study	MCBCS	USA	Non population-based study	2062	2041
Melbourne Collaborative Cohort Study	MCCS	Australia	Prospective cohort study	1002	1206
Multiethnic Cohort	MEC	USA	Prospective cohort study	668	724
Melanoma Inquiry of Southern Sweden	MISS	Sweden	Prospective cohort study	599	1529
Mayo Mammography Health Study	MMHS	USA	Prospective cohort study	276	1635
Nashville Breast Health Study	NBHS	USA	Population-based case-control study	482	652
Northern California Breast Cancer Family Registry	NC-BCFR	USA	Non population-based study	696	150
North Carolina Breast Cancer Study	NCBCS	USA	Population-based case-control study	2074	1006
Nurses' Health Study	NHS	USA	Prospective cohort study	1103	1804
Nurses' Health Study 2	NHS2	USA	Prospective cohort study	1112	1905
Ontario Familial Breast Cancer Registry	OFBCR	Canada	Non population-based study	1934	728
NCI Polish Breast Cancer Study	PBCS	Poland	Population-based case-control study	1768	2082
Karolinska Mammography Project for Risk Prediction of Breast Cancer - Case-Control Study	PKARMA	Sweden	Non population-based study	3115	5464
The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	PLCO	USA	Prospective cohort study	1822	2595
Predicting the Risk Of Cancer At Screening Study	PROCAS	UK	Population-based case-control study	342	1656
Singapore and Sweden Breast Cancer Study	SASBAC	Sweden	Population-based case-control study	1129	1373
Sheffield Breast Cancer Study	SBCS	UK	Non population-based study	594	848
Study of Epidemiology and Risk factors in Cancer Heredity	SEARCH	UK	Non population-based study	12571	8889
The Sister Study	SISTER	USA	Prospective cohort study	1501	1562
Swedish Mammography Cohort	SMC	Sweden	Prospective cohort study	1349	661
UCI Breast Cancer Study	UCIBCS	USA	Non population-based study	427	258
UK Breakthrough Generations Study	UKBGS	UK	Prospective cohort study	1047	1032
US Radiologic Technologists Study	USRT	USA	Non population-based study	848	1699
Women's Health Initiative Observational Study	WHI	USA	Prospective cohort study	4930	4617
Total				72284	80354

[1]Population-based design was defined as recruiting a random sample of all cases occurring in a geographically defined population during a specified period of time, and recruiting controls that were a random sample of the same source population as cases during the same period of time. Non-population-based design was defined as not strictly population-based (e.g. due to oversampling of selected participant groups for genotyping) or hospital-based.

1

Supplementary Table 2: Characteristics of the study population by study design.

Characteristics	Population-based studies				Non population-based studies			
	Cases N (%)	Controls N (%)	Cases Mean (S.D.)	Controls Mean (S.D.)	Cases N (%)	Controls N (%)	Cases Mean (S.D.)	Controls Mean (S.D.)
Reference age	38510	48308	61.03 (11.27)	58.66 (10.71)	33774	32046	56.10 (11.38)	54.80 (12.58)
ER status								
Positive	27830 (72.27)				22385 (66.28)			
Negative	5783 (15.02)				5113 (15.14)			
Missing	4897 (12.72)				6276 (18.58)			
Menopausal status								
Premenopausal	9045 (23.49)	12047 (24.94)			12556 (37.18)	13424 (41.89)		
Postmenopausal	29465 (76.51)	36261 (75.06)			21218 (62.82)	18622 (58.11)		
Family history in a first-degree relative								
Yes	7226 (18.76)	6784 (14.04)			5396 (15.98)	2060 (6.43)		
No	19564 (50.80)	28860 (59.74)			19764 (58.52)	15895 (49.60)		
Missing	11720 (30.43)	12664 (26.22)			8614 (25.50)	14091 (43.97)		
Reproductive risk factors								
Age at menarche (years)	36893	46855	12.86 (1.53)	12.96 (1.56)	22415	19439	12.99 (1.57)	12.99 (1.55)
Ever parous								
Yes	32025 (83.16)	41555 (86.02)			20442 (60.53)	23398 (73.01)		
No	5217 (13.55)	5618 (11.63)			3933 (11.65)	4127 (12.88)		
Missing	1268 (3.29)	1135 (2.35)			9399 (27.83)	4521 (14.11)		
Number of full-term pregnancies[a]								
1	5572 (17.40)	6182 (14.88)			3912 (19.14)	4151 (17.74)		
2	13004 (40.61)	17091 (41.13)			9811 (47.99)	11263 (48.14)		
3	7735 (24.15)	10427 (25.09)			4511 (22.07)	5140 (21.97)		
≥4	5323 (16.62)	7652 (18.41)			2048 (10.02)	2146 (9.17)		
Missing	391 (1.22)	203 (0.49)			160 (0.78)	698 (2.98)		
Ever breastfed[a]								

Yes	17358 (54.20)	19953 (48.02)			11298 (55.27)	9543 (40.79)		
No	6155 (19.22)	6557 (15.78)			3167 (15.49)	2409 (10.30)		
Missing	8512 (26.58)	15045 (36.21)			5977 (29.24)	11446 (48.92)		
Duration of breast feeding[a] (months)	20737	22183	7.83 (10.39)	8.30 (10.77)	9201	5555	6.99 (10.22)	7.32 (10.67)
Age at FFTP[a] (years)	30412	39987	24.92 (4.65)	24.67 (4.54)	17883	16192	25.09 (5.05)	25.48 (4.82)
Anthropometric risk factors								
Adult height (cm)	35767	46506	163.58 (6.50)	163.62 (6.49)	23642	18359	164.13 (6.78)	164.54 (6.88)
Premenopausal BMI[b] (kg/m^2)	8509	11510	24.85 (4.78)	25.26 (5.12)	8467	9464	25.28 (4.95)	24.98 (4.78)
Postmenopausal BMI[c] (kg/m^2)	28069	35112	26.52 (5.29)	26.05 (4.98)	14877	15508	26.45 (4.99)	26.30 (4.85)
Hormonal risk factors								
Ever use of oral contraceptives								
Yes	19632 (50.98)	26311 (54.47)			11018 (32.62)	12356 (38.56)		
No	15750 (40.90)	18441 (38.17)			5080 (15.04)	3419 (10.67)		
Missing	3128 (8.12)	3556 (7.36)			17676 (52.34)	16271 (50.77)		
Current use of EPT[c]								
Yes	3490 (11.84)	2758 (7.61)			258 (1.22)	174 (0.93)		
No	13525 (45.90)	16757 (46.21)			3962 (18.67)	3406 (18.29)		
Missing	12450 (42.25)	16746 (46.18)			16998 (80.11)	15042 (80.78)		
Current use of ET[c]								
Yes	2736 (9.29)	3236 (8.92)			185 (0.87)	240 (1.29)		
No	14072 (47.76)	16180 (44.62)			3929 (18.52)	3282 (17.62)		
Missing	12657 (42.96)	16845 (46.45)			17104 (80.61)	15100 (81.09)		
Lifestyle risk factors								
Lifetime intake of alcohol (g/day)	15829	18723	6.55 (12.57)	5.79 (10.33)	1461	1376	10.84 (14.81)	33.60 (63.31)
Current smoking								
Yes	4762 (12.37)	5630 (11.65)			2505 (7.42)	14681 (45.81)		
No	28975 (75.24)	37592 (77.82)			11965 (35.43)	15214 (47.48)		
Missing	4773 (12.39)	5086 (10.53)			19304 (57.16)	14681 (45.81)		

Pack-years smoked[d]	11607	15660	17.82 (18.27)	15.64 (16.44)	2969	3980	18.31 (17.63)	16.03 (15.58)

This table shows the number of cases and controls for each risk factor after all exclusions except for the exclusion of 150 cases and 150 controls for the variable of interest. This exclusion was conducted individually for each risk factor at the time of fitting logistic regression models. For continuous variables mean and standard deviation are reported, whereas, for categorical variables numbers and percentage are reported.

N: Number; %: Percentage; S.D.: Standard deviation; ER: Estrogen receptor; FFTP: First full-term pregnancy; BMI: Body mass index; EPT: Combined estrogen-progesterone menopausal hormonal therapy; ET: Estrogen-only menopausal hormonal therapy

[a] Among parous women, [b] Among premenopausal women, [c] Among postmenopausal women, [d] Among women who were ever smokers

Supplementary Table 3: Associations of epidemiological risk factors for overall and ER-specific subtype breast cancer risk in population-based and cohort studies

Environmental risk factor	*Overall breast cancer risk OR (95% CI)*	*ER-positive breast cancer risk OR (95% CI)*	*ER-negative breast cancer risk OR (95% CI)*
Univariate models[a]			
Age at menarche (per 2 years)	0.91 (0.89-0.92)	0.91 (0.89-0.93)	0.89 (0.85-0.93)
Ever parous (yes/no)	0.81 (0.77-0.84)	0.78 (0.74-0.81)	0.94 (0.85-1.04)
Number of full-term pregnancies (1,2,3,≥4)[1]	0.87 (0.85-0.88)	0.86 (0.84-0.87)	0.90 (0.86-0.94)
Age at first full-term pregnancy (per 5 years)[1]	1.14 (1.12-1.16)	1.17 (1.14-1.19)	1.02 (0.97-1.06)
Ever breastfed (yes/no)[1]	0.91 (0.88-0.95)	0.92 (0.88-0.96)	0.96 (0.88-1.03)
Duration of breastfeeding (per 12 months)[1]	0.96 (0.93-0.98)	0.95 (0.93-0.98)	0.98 (0.94-1.03)
Adult height (per 5 cm)	1.09 (1.08-1.10)	1.10 (1.09-1.12)	1.03 (1.00-1.05)
Premenopausal BMI (per 5 kg/m^2)	0.95 (0.92-0.98)	0.92 (0.89-0.95)	1.07 (0.98-1.16)
Postmenopausal BMI (per 5 kg/m^2)	1.07 (1.05-1.09)	1.07 (1.05-1.09)	1.05 (1.00-1.11)
Ever use of oral contraceptives (yes/no)	1.22 (1.18-1.26)	1.24 (1.20-1.29)	1.14 (1.05-1.23)
Current use of EPT (yes/no)[2,3]	1.75 (1.65-1.87)	1.93 (1.81-2.06)	1.11 (0.92-1.34)
Current use of ET (yes/no)[2,3]	1.10 (1.03-1.17)	1.11 (1.03-1.19)	1.35 (1.11-1.64)
Lifetime intake of alcohol (per 10 g/day)	1.07 (1.05-1.10)	1.09 (1.07-1.11)	1.03 (0.98-1.08)
Current smoking (yes/no)[4]	1.18 (1.13-1.24)	1.18 (1.12-1.25)	1.06 (0.96-1.18)
Pack years smoked (per 10 pack-years)[5]	1.02 (1.00-1.04)	1.02 (1.00-1.04)	1.00 (0.95-1.04)
Family history (yes/no)	1.56 (1.49-1.64)	1.54 (1.46-1.62)	1.53 (1.39-1.68)
Multivariate model 1[b]			
Age at menarche (per 2 years)	0.89 (0.86-0.94)	0.90 (0.86-0.95)	0.85 (0.79-0.93)
Number of full-term pregnancies (per 1 unit)	0.89 (0.86-0.93)	0.88 (0.85-0.92)	0.89 (0.82-0.95)
Age at first full-term pregnancy (per 5 years)	1.07 (1.02-1.12)	1.08 (1.03-1.14)	0.96 (0.88-1.04)
Ever breastfed (yes/no)	0.97 (0.89-1.05)	0.98 (0.89-1.07)	0.96 (0.82-1.12)
Adult height (per 5 cm)	1.05 (1.01-1.08)	1.07 (1.03-1.10)	0.99 (0.94-1.05)
BMI (<18.5 kg/m^2)	1.28 (0.92-1.76)	1.35 (0.94-1.92)	1.13 (0.66-1.96)
BMI (25-30 kg/m^2)	1.04 (0.93-1.16)	1.00 (0.88-1.13)	1.12 (0.92-1.36)
BMI (≥30 kg/m^2)	1.14 (0.98-1.33)	1.02 (0.86-1.22)	1.03 (0.76-1.38)
Current use of MHT (yes/no)	1.30 (1.15-1.46)	1.39 (1.22-1.58)	1.00 (0.80-1.24)
Interaction between current use of MHT and BMI (<18.5 kg/m^2)	0.61 (0.33-1.13)	0.55 (0.28-1.07)	1.12 (0.39-3.23)
Interaction between current use of MHT and BMI (25-30 kg/m^2)	0.91 (0.74-1.12)	0.93 (0.74-1.16)	0.82 (0.54-1.26)
Interaction between current use of MHT and BMI (≥30 kg/m^2)	0.98 (0.72-1.34)	1.02 (0.72-1.43)	0.69 (0.33-1.42)

Lifetime intake of alcohol (per 10 g/day)	1.02 (0.99-1.06)	1.02 (0.99-1.06)	1.04 (0.99-1.10)
Current smoking (yes/no)	1.28 (1.16-1.42)	1.36 (1.22-1.52)	1.01 (0.85-1.20)
Family history (yes/no)	1.75 (1.57-1.94)	1.73 (1.54-1.94)	1.72 (1.42-2.08)
Multivariate model 2[c]			
BMI (<18.5 kg/m^2)	1.10 (0.87-1.39)	1.22 (0.94-1.58)	1.04 (0.67-1.61)
BMI (25-30 kg/m^2)	0.99 (0.92-1.07)	0.96 (0.88-1.04)	1.04 (0.90-1.21)
BMI (≥30 kg/m^2)	1.10 (1.00-1.22)	1.05 (0.94-1.18)	1.13 (0.91-1.39)
Current use of MHT (yes/no)	1.45 (1.34-1.57)	1.58 (1.45-1.72)	1.09 (0.92-1.28)
Interaction between current use of MHT and BMI (<18.5 kg/m^2)	0.75 (0.49-1.16)	0.63 (0.39-1.01)	0.99 (0.43-2.31)
Interaction between current use of MHT and BMI (25-30 kg/m^2)	0.88 (0.77-1.00)	0.90 (0.78-1.04)	0.94 (0.70-1.27)
Interaction between current use of MHT and BMI (≥30 kg/m^2)	0.86 (0.72-1.03)	0.88 (0.72-1.07)	0.60 (0.37-2.61)
Lifetime intake of alcohol (per 10 g/day)	1.05 (1.03-1.08)	1.07 (1.04-1.09)	1.02 (0.98-1.07)
Current smoking (yes/no)	1.28 (1.19-1.38)	1.31 (1.21-1.42)	1.12 (0.97-1.28)

BMI: Body mass index, EPT: Estrogen-Progesterone menopausal hormonal therapy, ET: Estrogen-only menopausal hormonal therapy; MHT: Menopausal hormonal therapy

[a]All OR estimates are based on a model with single risk factor analyses adjusted for reference age and study

[b] Model includes all classical risk factors: age at menarche, age at first full-term pregnancy, number of children, ever breastfed, height, BMI (18.5-<25 kg/m^2: reference category), current MHT use, current smoking, and lifetime alcohol consumption. This model is adjusted for reference age, study, menopausal status, former smoking, former use of menopausal hormonal therapy, interaction between BMI and current MHT use, and interaction between BMI and former MHT use

[c] Model includes modifiable risk factors: BMI (18.5-<25 kg/m^2: reference category), current MHT use, current smoking and lifetime alcohol consumption. This model is adjusted for reference age, study, menopausal status, former smoking, former use of menopausal hormonal therapy, interaction between BMI and current MHT use, and interaction between BMI and former MHT use.

[1] among parous women

[2] among postmenopausal women

[3] Additionally, models were adjusted for former use of menopausal hormonal therapy and use of any other menopausal hormonal therapy preparations

[4] Additionally, model was adjusted for former smoking

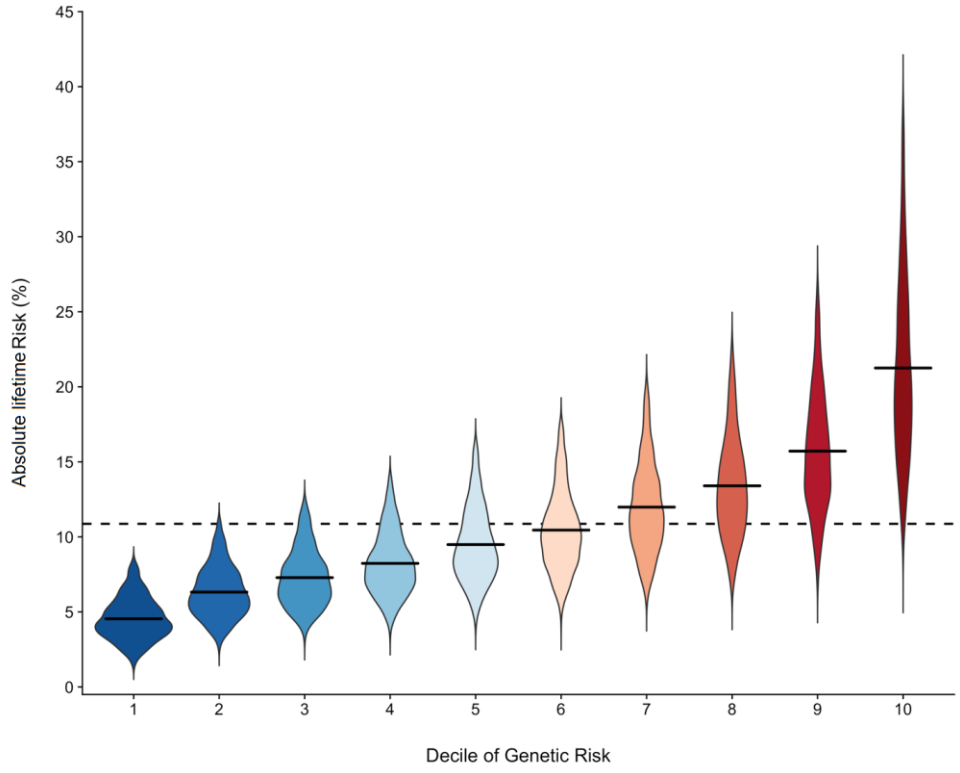
[5] among ever smokers

Supplementary table 4: Goodness of fit test p-values for overall breast cancer and estrogen receptor (ER) positive breast cancer, based on population-based studies

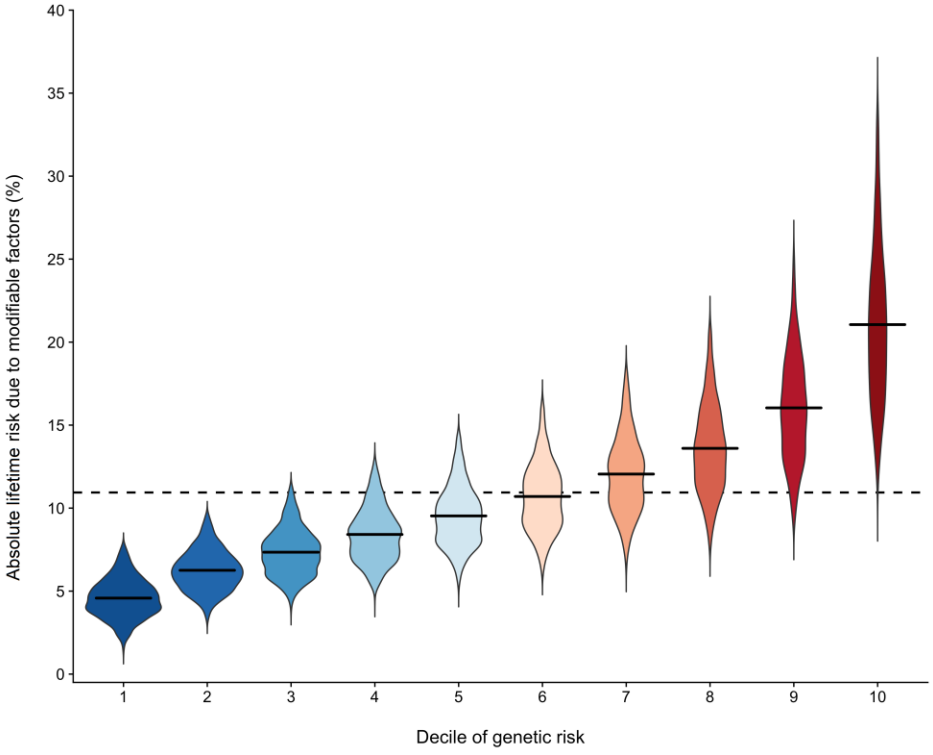
Single risk factor models with 313-SNP PRS[1]	**Overall breast cancer risk**			**ER-positive breast cancer risk**		
	Cases/Controls	*Tail-based goodness-of-fit test*	*Global goodness-of-fit test*	*Cases/Controls*	*Tail-based goodness-of-fit test*	*Global goodness-of-fit test*
Age at menarche (per 2 years)	36893/46855	0.75	0.47	26331/46855	0.41	0.67
Ever parous (yes/no)	37242/47173	0.68	0.28	26938/47173	0.34	0.62
Number of children (1, 2, 3, ≥4)[2]	31634/41352	0.40	0.19	22860/41352	0.29	0.08
Age at first full-term pregnancy (per 5 years)[2]	30412/39987	0.87	0.42	22009/39987	0.78	0.40
Breastfeeding (yes/no)[2]	23513/26510	0.23	0.32	16616/26510	0.45	0.76
Duration of breastfeeding (per 12 months)	20737/22183	0.09	0.08	14615/22183	0.16	0.73
Adult height (per 5 cm)	35767/46506	0.51	0.60	25764/46506	0.18	0.44
Premenopausal BMI (per 5kg/m^2)[3]	8509/11510	0.51	0.56	5615/11510	0.74	0.31
Postmenopausal BMI (per 5kg/m^2)[4]	28069/35112	0.31	0.65	20884/35112	0.09	0.24
Ever use of oral contraceptives (yes/no)	35382/44752	0.30	0.38	25417/44752	0.13	0.53
Current use of EP therapy (yes/no)[4,5]	17015/19515	0.43	0.31	12777/19515	0.55	0.78
Current use of E-only therapy (yes/no)[4,5]	16808/19416	0.48	0.26	12614/19416	0.63	0.79
Alcohol consumption (per 10g/day)	15829/18723	0.40	0.11	11304/18723	0.07	0.44
Current smoking (yes/no)[6]	33737/43222	0.51	0.82	24124/43222	0.20	0.68
Pack-years (per 10 pack-years)	11607/15660	0.71	0.95	8373/15660	0.46	0.80

[1]always adjusted for study; [2]in parous women only; [3] in premenopausal women; [4] in postmenopausal women; [5] adjusted for former use of MHT and use of any other MHT than the preparation of interest; [6] adjusted for former smoking.

a)


Absolute lifetime Risk (%)
0
5
10
15
20
25
30
35
40
45
1
2
3
4
5
6
7
8
9
10
Decile of Genetic Risk

b)


Absolute lifetime risk due to modifiable factors (%)
0
5
10
15
20
25
30
35
40
1
2
3
4
5
6
7
8
9
10
Decile of genetic risk