Polygenic Risk Modelling for Prediction of Epithelial Ovarian Cancer Risk

Eileen O. Dareng[1]*, Jonathan P. Tyrer[2]*, Daniel R. Barnes[1], Michelle R. Jones[3], Xin Yang[1], Katja K.H. Aben[4, 5], Muriel A. Adank[6], Simona Agata[7], Irene L. Andrulis[8, 9], Hoda Anton-Culver[10], Natalia N. Antonenkova[11], Gerasimos Aravantinos[12], Banu K. Arun[13], Annelie Augustinsson[14], Judith Balmaña[15, 16], Elisa V. Bandera[17], Rosa B. Barkardottir[18, 19], Daniel Barrowdale[1], Matthias W. Beckmann[20], Alicia Beeghly-Fadiel[21], Javier Benitez[22, 23], Marina Bermisheva[24], Marcus Q. Bernardini[25], Line Bjorge[26, 27], Amanda Black[28], Natalia V. Bogdanova[11, 29, 30], Bernardo Bonanni[31], Ake Borg[32], James D. Brenton[33], Agnieszka Budzilowska[34], Ralf Butzow[35], Saundra S. Buys[36], Hui Cai[21], Maria A. Caligo[37], Ian Campbell[38, 39], Rikki Cannioto[40], Hayley Cassingham[41], Jenny Chang-Claude[42, 43], Stephen J. Chanock[44], Kexin Chen[45], Yoke-Eng Chiew[46, 47], Wendy K. Chung[48], Kathleen B.M. Claes[49], Sarah Colonna[36], **GEMO Study Collaborators**[50-52], **GC-HBOC Study Collaborators**[53], **EMBRACE Collaborators**[1], Linda S. Cook[54, 55], Fergus J. Couch[56], Mary B. Daly[57], Fanny Dao[58], Eleanor Davies[59], Miguel de la Hoya[60], Robin de Putter[49], Joe Dennis[1], Allison DePersia[61, 62], Peter Devilee[63, 64], Orland Diez[65, 66], Yuan Chun Ding[67], Jennifer A. Doherty[68], Susan M. Domchek[69], Thilo Dörk[30], Andreas du Bois[70, 71], Matthias Dürst[72], Diana M. Eccles[73], Heather A. Eliassen[74, 75], Christoph Engel[76, 77], D. Gareth Evans[78, 79], Peter A. Fasching[20, 80], James M. Flanagan[81], Renée T. Fortner[42], Eitan Friedman[83, 84], Patricia A. Ganz[85], Judy Garber[86], Francesca Gensini[87], Graham G. Giles[88-90], Gord Glendon[8], Andrew K. Godwin[91], Marc T. Goodman[92], Mark H. Greene[93], Jacek Gronwald[94], OPAL Study Group[95], AOCS Group[38, 46], Eric Hahnen[53, 96], Christopher A. Haiman[97], Niclas Håkansson[98], Ute Hamann[99], Thomas V.O. Hansen[100], Holly R. Harris[101, 102], Mikael Hartman[103, 104], Florian Heitz[70, 71, 105], Michelle A.T. Hildebrandt[106], Estrid Høgdall[107, 108], Claus K. Høgdall[109], John L. Hopper[89], Ruea-Yea Huang[110], Chad Huff[106], Peter J. Hulick[61, 62], David G. Huntsman[111-114], Evgeny N. Imyanitov[115], **KConFab Investigators**[38], **HEBON Investigators**[116], Claudine Isaacs[117], Anna

Jakubowska[94, 118], Paul A. James[39, 119], Ramunas Janavicius[120, 121], Allan Jensen[107], Oskar Th. Johannsson[122], Esther M. John [123, 124], Michael E. Jones[125], Daehee Kang[126-128], Beth Y. Karlan[129], Anthony Karnezis[130], Linda E. Kelemen[131], Elza Khusnutdinova[24, 132], Lambertus A. Kiemeney[4], Byoung-Gie Kim[133], Susanne K. Kjaer[107, 109], Ian Komenaka[134], Jolanta Kupryjanczyk[34], Allison W. Kurian[123, 124], Ava Kwong[135-137], Diether Lambrechts[138, 139], Melissa C. Larson[140], Conxi Lazaro[141], Nhu D. Le[142], Goska Leslie[1], Jenny Lester[129], Fabienne Lesueur[51, 52, 143], Douglas A. Levine[58, 144], Lian Li[45], Jingmei Li[145], Jennifer T. Loud[93], Karen H. Lu[146], Jan Lubiński[94], Eva Machackova[82], Phuong L. Mai[147], Siranoush Manoukian[148], Jeffrey R. Marks[149], Rayna Kim Matsuno[150], Keitaro Matsuo[151, 152], Taymaa May[25], Lesley McGuffog[1], John R. McLaughlin[153], Iain A. McNeish[154, 155], Noura Mebirouk[51, 52, 143], Usha Menon[156], Austin Miller[157], Roger L. Milne[88-90], Albina Minlikeeva[158], Francesmary Modugno[159, 160], Marco Montagna[7], Kirsten B. Moysich[158], Elizabeth Munro[161, 162], , Katherine L. Nathanson[69], Susan L. Neuhausen[67], Heli Nevanlinna[163], Joanne Ngeow Yuen Yie[164, 165], Henriette Roed Nielsen[166], Finn C. Nielsen[100], Liene Nikitina-Zake[167], Kunle Odunsi[168], Kenneth Offit[169, 170], Edith Olah[171], Siel Olbrecht[172], Olufunmilayo I. Olopade[173], Sara H. Olson[174], Håkan Olsson[14], Ana Osorio[23, 175], Laura Papi[87], Sue K. Park[126-128], Michael T. Parsons[176], Harsha Pathak[91], Inge Sokilde Pedersen[177-179], Ana Peixoto[180], Tanja Pejovic[161, 162], Pedro Perez-Segura[60], Jennifer B. Permuth[181], Beth Peshkin[117], Paolo Peterlongo[182], Anna Piskorz[33], Darya Prokofyeva[183], Paolo Radice[184], Johanna Rantala[185], Marjorie J. Riggan[186], Harvey A. Risch[187], Cristina Rodriguez-Antona[22, 23], Eric Ross[188], Mary Anne Rossing[101, 102], Ingo Runnebaum[72], Dale P. Sandler[189], Marta Santamariña[175, 190, 191], Penny Soucy[192], Rita K. Schmutzler[53, 96, 193], V. Wendy Setiawan[97], Kang Shan[194], Weiva Sieh[195, 196], Jacques Simard[197], Christian F. Singer[198], Anna P Sokolenko[115], Honglin Song[199], Melissa C. Southey[88, 90, 200], Helen Steed[201], Dominique Stoppa-Lyonnet[50, 202, 203], Rebecca Sutphen[204], Anthony J. Swerdlow[125, 205], Yen Yen Tan[206], Manuel R. Teixeira[180, 207], Soo Hwang Teo[208, 209], Kathryn

L. Terry[74, 210], Mary Beth Terry[211], **The OCAC Consortium**, **The CIMBA Consortium**, Mads Thomassen[166], Pamela J. Thompson[92], Liv Cecilie Vestrheim Thomsen[26, 27], Darcy L. Thull[212], Marc Tischkowitz[213, 214], Linda Titus[215], Amanda E. Toland[216], Diana Torres[99, 217], Britton Trabert[28], Ruth Travis[218], Nadine Tung[219], Shelley S. Tworoger[181, 220], Ellen Valen[26, 27], Anne M. van Altena[4], Annemieke H. van der Hout[221], Els Van Nieuwenhuysen[172], Elizabeth J. van Rensburg[222], Ana Vega[223-225], Digna Velez Edwards[226], Robert A. Vierkant[140], Frances Wang[227, 228], Barbara Wappenschmidt[53, 96], Penelope M. Webb[95], Clarice R. Weinberg[229], Jeffrey N. Weitzel[230], Nicolas Wentzensen[28], Emily White[102, 231], Alice S. Whittemore[123, 232], Stacey J. Winham[140], Alicja Wolk[98, 233], Yin-Ling Woo[234], Anna H. Wu[97], Li Yan[235], Drakoulis Yannoukakos[236], Katia M. Zavaglia[37], Wei Zheng[21], Argyrios Ziogas[10], Kristin K. Zorn[147], Zdenek Kleibl[237], Douglas Easton[1,2], Kate Lawrenson[3, 238], Anna DeFazio[46, 47], Thomas A. Sellers [239], Susan J. Ramus[240, 241], Celeste L. Pearce[242, 243], Alvaro N. Monteiro[181], Julie Cunningham[244], Ellen L. Goode[244], Joellen M. Schildkraut[245], Andrew Berchuck[186], Georgia Chenevix-Trench[176], Simon A. Gayther[3], Antonis C. Antoniou[1], Paul D.P. Pharoah[1, 2]

[1] University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Cambridge, UK.
[2] University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Oncology, Cambridge, UK.
[3] Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
[4] Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands.
[5] Netherlands Comprehensive Cancer Organisation, Utrecht, The Netherlands.
[6] The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Family Cancer Clinic, Amsterdam, The Netherlands.
[7] Veneto Institute of Oncology IOV - IRCCS, Immunology and Molecular Oncology Unit, Padua, Italy.
[8] Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Fred A. Litwin Center for Cancer Genetics, Toronto, ON, Canada.
[9] University of Toronto, Department of Molecular Genetics, Toronto, ON, Canada.
[10] University of California Irvine, Department of Epidemiology, Genetic Epidemiology Research Institute, Irvine, CA, USA.
[11] N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus.
[12] 'Agii Anargiri' Cancer Hospital, Athens, Greece.

[13] University of Texas MD Anderson Cancer Center, Department of Breast Medical Oncology, Houston, TX, USA.
[14] Lund University, Department of Cancer Epidemiology, Clinical Sciences, Lund, Sweden.
[15] Vall d'Hebron Institute of Oncology, Hereditary cancer Genetics Group, Barcelona, Spain.
[16] University Hospital of Vall d'Hebron, Department of Medical Oncology, Barcelona, Spain.
[17] Rutgers Cancer Institute of New Jersey, Cancer Prevention and Control Program, New Brunswick, NJ, USA.
[18] Landspitali University Hospital, Department of Pathology, Reykjavik, Iceland.
[19] University of Iceland, BMC (Biomedical Centre), Faculty of Medicine, Reykjavik, Iceland.
[20] University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, Erlangen, Germany.
[21] Vanderbilt University School of Medicine, Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Nashville, TN, USA.
[22] Biomedical Network on Rare Diseases (CIBERER), Madrid, Spain.
[23] Spanish National Cancer Research Centre (CNIO), Human Cancer Genetics Programme, Madrid, Spain.
[24] Ufa Federal Research Centre of the Russian Academy of Sciences, Institute of Biochemistry and Genetics, Ufa, Russia.
[25] Princess Margaret Hospital, Division of Gynecologic Oncology, University Health Network, Toronto, Ontario, Canada.
[26] Haukeland University Hospital, Department of Obstetrics and Gynecology, Bergen, Norway.
[27] University of Bergen, Centre for Cancer Biomarkers CCBIO, Department of Clinical Science, Bergen, Norway.
[28] National Cancer Institute, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA.
[29] Hannover Medical School, Department of Radiation Oncology, Hannover, Germany.
[30] Hannover Medical School, Gynaecology Research Unit, Hannover, Germany.
[31] IEO, European Institute of Oncology IRCCS, Division of Cancer Prevention and Genetics, Milan, Italy.
[32] Lund University and Skåne University Hospital, Department of Oncology, Lund, Sweden.
[33] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK.
[34] Maria Sklodowska-Curie National Research Institute of Oncology, Department of Pathology and Laboratory Diagnostics, Warsaw, Poland.
[35] University of Helsinki, Department of Pathology, Helsinki University Hospital, Helsinki, Finland.
[36] Huntsman Cancer Institute, Department of Medicine, Salt Lake City, UT, USA.
[37] University Hospital, SOD Genetica Molecolare, Pisa, Italy.
[38] Peter MacCallum Cancer Center, Melbourne, Victoria, Australia.
[39] The University of Melbourne, Sir Peter MacCallum Department of Oncology, Melbourne, Victoria, Australia.
[40] Roswell Park Cancer Institute, Cancer Pathology & Prevention, Division of Cancer Prevention and Population Sciences, Buffalo, NY, USA.
[41] Division of Human Genetics, The Ohio State University, Department of Internal Medicine, Columbus, OH, USA.
[42] German Cancer Research Center (DKFZ), Division of Cancer Epidemiology, Heidelberg, Germany.
[43] University Medical Center Hamburg-Eppendorf, Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), Hamburg, Germany.

[44] National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA.
[45] Tianjin Medical University Cancer Institute and Hospital, Department of Epidemiology, Tianjin, China.
[46] The University of Sydney, Centre for Cancer Research, The Westmead Institute for Medical Research, Sydney, New South Wales, Australia.
[47] Westmead Hospital, Department of Gynaecological Oncology, Sydney, New South Wales, Australia.
[48] Columbia University, Departments of Pediatrics and Medicine, New York, NY, USA.
[49] Ghent University, Centre for Medical Genetics, Gent, Belgium.
[50] INSERM U830, Department of Tumour Biology, Paris, France.
[51] Institut Curie, Paris, France.
[52] Mines ParisTech, Fontainebleau, France.
[53] Faculty of Medicine and University Hospital Cologne, University of Cologne, Center for Familial Breast and Ovarian Cancer, Cologne, Germany.
[54] University of New Mexico, University of New Mexico Health Sciences Center, Albuquerque, NM, USA.
[55] Alberta Health Services, Department of Cancer Epidemiology and Prevention Research, Calgary, AB, Canada.
[56] Mayo Clinic, Department of Laboratory Medicine and Pathology, Rochester, MN, USA.
[57] Fox Chase Cancer Center, Department of Clinical Genetics, Philadelphia, PA, USA.
[58] Memorial Sloan Kettering Cancer Center, Gynecology Service, Department of Surgery, New York, NY, USA.
[59] Cambridge, Cambridge, UK.
[60] CIBERONC, Hospital Clinico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Molecular Oncology Laboratory, Madrid, Spain.
[61] NorthShore University Health System, Center for Medical Genetics, Evanston, IL, USA.
[62] The University of Chicago Pritzker School of Medicine, Chicago, IL, USA.
[63] Leiden University Medical Center, Department of Pathology, Leiden, The Netherlands.
[64] Leiden University Medical Center, Department of Human Genetics, Leiden, The Netherlands.
[65] Vall dHebron Institute of Oncology (VHIO), Oncogenetics Group, Barcelona, Spain.
[66] University Hospital Vall dHebron, Clinical and Molecular Genetics Area, Barcelona, Spain.
[67] Beckman Research Institute of City of Hope, Department of Population Sciences, Duarte, CA, USA.
[68] University of Utah, Huntsman Cancer Institute, Department of Population Health Sciences, Salt Lake City, UT, USA.
[69] University of Pennsylvania, Basser Center for BRCA, Abramson Cancer Center, Philadelphia, PA, USA.
[70] Ev. Kliniken Essen-Mitte (KEM), Department of Gynecology and Gynecologic Oncology, Essen, Germany.
[71] Dr. Horst Schmidt Kliniken Wiesbaden, Department of Gynecology and Gynecologic Oncology, Wiesbaden, Germany.
[72] Jena University Hospital - Friedrich Schiller University, Department of Gynaecology, Jena, Germany.
[73] University of Southampton, Faculty of Medicine, Southampton, UK.
[74] Harvard T.H. Chan School of Public Health, Department of Epidemiology, Boston, MA, USA.
[75] Brigham and Women's Hospital and Harvard Medical School, Channing Division of Network Medicine, Boston, MA, USA.

[76] University of Leipzig, Institute for Medical Informatics, Statistics and Epidemiology, Leipzig, Germany.
[77] University of Leipzig, LIFE - Leipzig Research Centre for Civilization Diseases, Leipzig, Germany.
[78] University of Manchester, Manchester Academic Health Science Centre, Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester, UK.
[79] St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, North West Genomics Laboratory Hub, Manchester Centre for Genomic Medicine, Manchester, UK.
[80] University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, Los Angeles, CA, USA.
[81] Imperial College London, Division of Cancer and Ovarian Cancer Action Research Centre, Department of Surgery and Cancer, London, UK.
[82] Masaryk Memorial Cancer Institute, Department of Cancer Epidemiology and Genetics, Brno, Czech Republic.
[83] Chaim Sheba Medical Center, The Susanne Levy Gertner Oncogenetics Unit, Ramat Gan, Israel.
[84] Tel Aviv University, Sackler Faculty of Medicine, Ramat Aviv, Israel.
[85] Jonsson Comprehensive Cancer Centre, UCLA, Schools of Medicine and Public Health, Division of Cancer Prevention & Control Research, Los Angeles, CA, USA.
[86] Dana-Farber Cancer Institute, Cancer Risk and Prevention Clinic, Boston, MA, USA.
[87] University of Florence, Department of Experimental and Clinical Biomedical Sciences 'Mario Serio', Medical Genetics Unit, Florence, Italy.
[88] Cancer Council Victoria, Cancer Epidemiology Division, Melbourne, Victoria, Australia.
[89] The University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, Melbourne, Victoria, Australia.
[90] Monash University, Precision Medicine, School of Clinical Sciences at Monash Health, Clayton, Victoria, Australia.
[91] University of Kansas Medical Center, Department of Pathology and Laboratory Medicine, Kansas City, KS, USA.
[92] Cedars-Sinai Medical Center, Samuel Oschin Comprehensive Cancer Institute, Cancer Prevention and Genetics Program, Los Angeles, CA, USA.
[93] National Cancer Institute, Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA.
[94] Pomeranian Medical University, Department of Genetics and Pathology, Szczecin, Poland.
[95] QIMR Berghofer Medical Research Institute, Population Health Department, Brisbane, Queensland, Australia.
[96] Faculty of Medicine and University Hospital Cologne, University of Cologne, Center for Integrated Oncology (CIO), Cologne, Germany.
[97] University of Southern California, Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA.
[98] Karolinska Institutet, Institute of Environmental Medicine, Stockholm, Sweden.
[99] German Cancer Research Center (DKFZ), Molecular Genetics of Breast Cancer, Heidelberg, Germany.
[100] Rigshospitalet, Copenhagen University Hospital, Department of Clinical Genetics, Copenhagen, Denmark.
[101] Fred Hutchinson Cancer Research Center, Program in Epidemiology, Division of Public Health Sciences, Seattle, WA, USA.
[102] University of Washington, Department of Epidemiology, Seattle, WA, USA.

[103] National University of Singapore and National University Health System, Saw Swee Hock School of Public Health, Singapore, Singapore.
[104] National University Health System, Department of Surgery, Singapore, Singapore.
[105] Humboldt-UniversitŠt zu Berlin, and Berlin Institute of Health, Department for Gynecology with the Center for Oncologic Surgery CharitŽ Campus Virchow-Klinikum, CharitŽ Đ UniversitŠtsmedizin Berlin, corporate member of Freie UniversitŠt Berlin, Berlin, Germany.
[106] University of Texas MD Anderson Cancer Center, Department of Epidemiology, Houston, TX, USA.
[107] Danish Cancer Society Research Center, Department of Virus, Lifestyle and Genes, Copenhagen, Denmark.
[108] University of Copenhagen, Molecular Unit, Department of Pathology, Herlev Hospital, Copenhagen, Denmark.
[109] University of Copenhagen, Department of Gynaecology, Rigshospitalet, Copenhagen, Denmark.
[110] Roswell Park Cancer Institute, Center For Immunotherapy, Buffalo, NY, USA.
[111] BC Cancer, Vancouver General Hospital, and University of British Columbia, British Columbia's Ovarian Cancer Research (OVCARE) Program, Vancouver, BC, Canada.
[112] University of British Columbia, Department of Pathology and Laboratory Medicine, Vancouver, BC, Canada.
[113] University of British Columbia, Department of Obstetrics and Gynecology, Vancouver, BC, Canada.
[114] BC Cancer Research Centre, Department of Molecular Oncology, Vancouver, BC, Canada.
[115] N.N. Petrov Institute of Oncology, St. Petersburg, Russia.
[116] Coordinating center: The Netherlands Cancer Institute, The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), Amsterdam, The Netherlands.
[117] Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA.
[118] Pomeranian Medical University, Independent Laboratory of Molecular Biology and Genetic Diagnostics, Szczecin, Poland.
[119] Peter MacCallum Cancer Center, Parkville Familial Cancer Centre, Melbourne, Victoria, Australia.
[120] Vilnius University Hospital Santariskiu Clinics, Hematology, oncology and transfusion medicine center, Dept. of Molecular and Regenerative Medicine, Vilnius, Lithuania.
[121] State Research Institute Centre for Innovative Medicine, Vilnius, Lithuania.
[122] Landspitali University Hospital, Department of Oncology, Reykjavik, Iceland.
[123] Stanford University School of Medicine, Department of Epidemiology & Population Health, Stanford, CA, USA.
[124] Stanford Cancer Institute, Stanford University School of Medicine, Department of Medicine, Division of Oncology, Stanford, CA, USA.
[125] The Institute of Cancer Research, Division of Genetics and Epidemiology, London, UK.
[126] Seoul National University College of Medicine, Department of Preventive Medicine, Seoul, Korea.
[127] Seoul National University Graduate School, Department of Biomedical Sciences, Seoul, Korea.
[128] Seoul National University, Cancer Research Institute, Seoul, Korea.
[129] University of California at Los Angeles, David Geffen School of Medicine, Department of Obstetrics and Gynecology, Los Angeles, CA, USA.
[130] UC Davis Medical Center, Department of Pathology and Laboratory Medicine, Sacramento, CA, USA.
[131] Medical University of South Carolina, Hollings Cancer Center, Charleston, SC, USA.
[132] Saint Petersburg State University, Saint Petersburg, Russia.

[133] Sungkyunkwan University School of Medicine, Department of Obstetrics and Gynecology, Samsung Medical Center, Seoul, Korea.
[134] City of Hope Clinical Cancer Genetics Community Research Network, Duarte, CA, USA.
[135] Cancer Genetics Centre, Hong Kong Hereditary Breast Cancer Family Registry, Happy Valley, Hong Kong.
[136] The University of Hong Kong, Department of Surgery, Pok Fu Lam, Hong Kong.
[137] Hong Kong Sanatorium and Hospital, Department of Surgery, Happy Valley, Hong Kong.
[138] VIB Center for Cancer Biology, Leuven, Belgium.
[139] University of Leuven, Laboratory for Translational Genetics, Department of Human Genetics, Leuven, Belgium.
[140] Mayo Clinic, Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Rochester, MN, USA.
[141] ONCOBELL-IDIBELL-IGTP, Catalan Institute of Oncology, CIBERONC, Hereditary Cancer Program, Barcelona, Spain.
[142] BC Cancer, Cancer Control Research, Vancouver, BC, Canada.
[143] Inserm U900, Genetic Epidemiology of Cancer team, Paris, France.
[144] NYU Langone Medical Center, Gynecologic Oncology, Laura and Isaac Pearlmutter Cancer Center, New York, NY, USA.
[145] Genome Institute of Singapore, Human Genetics Division, Singapore, Singapore.
[146] University of Texas MD Anderson Cancer Center, Department of Gynecologic Oncology and Clinical Cancer Genetics Program, Houston, TX, USA.
[147] Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.
[148] Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Unit of Medical Genetics, Department of Medical Oncology and Hematology, Milan, Italy.
[149] Duke University Hospital, Department of Surgery, Durham, NC, USA.
[150] University of Hawaii Cancer Center, Cancer Epidemiology Program, Honolulu, HI, USA.
[151] Aichi Cancer Center Research Institute, Division of Cancer Epidemiology and Prevention, Nagoya, Japan.
[152] Nagoya University Graduate School of Medicine, Division of Cancer Epidemiology, Nagoya, Japan.
[153] Samuel Lunenfeld Research Institute, Public Health Ontario, Toronto, ON, Canada.
[154] Imperial College London, Division of Cancer and Ovarian Cancer Action Research Centre, Department Surgery & Cancer, London, UK.
[155] University of Glasgow, Institute of Cancer Sciences, Glasgow, UK.
[156] University College London, MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, London, UK.
[157] Roswell Park Cancer Institute, NRG Oncology, Statistics and Data Management Center, Buffalo, NY, USA.
[158] Roswell Park Cancer Institute, Division of Cancer Prevention and Control, Buffalo, NY, USA.
[159] Magee-Womens Research Institute and Hillman Cancer Center, Womens Cancer Research Center, Pittsburgh, PA, USA.
[160] University of Pittsburgh School of Medicine, Division of Gynecologic Oncology, Department of Obstetrics, Gynecology and Reproductive Sciences, Pittsburgh, PA, USA.
[161] Oregon Health & Science University, Department of Obstetrics and Gynecology, Portland, OR, USA.
[162] Oregon Health & Science University, Knight Cancer Institute, Portland, OR, USA.
[163] University of Helsinki, Department of Obstetrics and Gynecology, Helsinki University Hospital, Helsinki, Finland.

[164] National Cancer Centre, Cancer Genetics Service, Singapore, Singapore.
[165] Nanyang Technological University, Lee Kong Chian School of Medicine, Singapore, Singapore.
[166] Odense University Hospital, Department of Clinical Genetics, Odence C, Denmark.
[167] Latvian Biomedical Research and Study Centre, Riga, Latvia.
[168] Roswell Park Cancer Institute, Department of Gynecologic Oncology, Buffalo, NY, USA.
[169] Memorial Sloan Kettering Cancer Center, Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, New York, NY, USA.
[170] Memorial Sloan Kettering Cancer Center, Clinical Genetics Service, Department of Medicine, New York, NY, USA.
[171] National Institute of Oncology, Department of Molecular Genetics, Budapest, Hungary.
[172] University Hospitals Leuven, Division of Gynecologic Oncology, Department of Obstetrics and Gynaecology and Leuven Cancer Institute, Leuven, Belgium.
[173] The University of Chicago, Center for Clinical Cancer Genetics, Chicago, IL, USA.
[174] Memorial Sloan-Kettering Cancer Center, Department of Epidemiology and Biostatistics, New York, NY, USA.
[175] Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain.
[176] QIMR Berghofer Medical Research Institute, Department of Genetics and Computational Biology, Brisbane, Queensland, Australia.
[177] Aalborg University Hospital, Molecular Diagnostics, Aalborg, Denmark.
[178] Aalborg University Hospital, Clinical Cancer Research Center, Aalborg, Denmark.
[179] Aalborg University, Department of Clinical Medicine, Aalborg, Denmark.
[180] Portuguese Oncology Institute, Department of Genetics, Porto, Portugal.
[181] Moffitt Cancer Center, Department of Cancer Epidemiology, Tampa, FL, USA.
[182] IFOM - the FIRC Institute of Molecular Oncology, Genome Diagnostics Program, Milan, Italy.
[183] Bashkir State University, Department of Genetics and Fundamental Medicine, Ufa, Russia.
[184] Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Milan, Italy.
[185] Karolinska Institutet, Clinical Genetics, Stockholm, Sweden.
[186] Duke University Hospital, Department of Gynecologic Oncology, Durham, NC, USA.
[187] Yale School of Public Health, Chronic Disease Epidemiology, New Haven, CT, USA.
[188] Fox Chase Cancer Center, Population Studies Facility, Philadelphia, PA, USA.
[189] National Institute of Environmental Health Sciences, NIH, Epidemiology Branch, Research Triangle Park, NC, USA.
[190] Fundación Pública Galega Medicina Xenómica, Santiago De Compostela, Spain.
[191] Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago De Compostela, Spain.
[192] Centre Hospitalier Universitaire de Québec – Université Laval Research Center, Genomics Center, Québec City, QC, Canada.
[193] Faculty of Medicine and University Hospital Cologne, University of Cologne, Center for Molecular Medicine Cologne (CMMC), Cologne, Germany.
[194] Hebei Medical University, Fourth Hospital, Department of Obstetrics and Gynaecology, Shijiazhuang, China.
[195] Icahn School of Medicine at Mount Sinai, Department of Population Health Science and Policy, New York, NY, USA.
[196] Icahn School of Medicine at Mount Sinai, Department of Genetics and Genomic Sciences, New York, NY, USA.
[197] Centre Hospitalier Universitaire de Québec-Université Laval Research Center, Genomic Center, Québec City, QC, Canada.

[198] Medical University of Vienna, Dept of OB/GYN and Comprehensive Cancer Center, Vienna, Austria.
[199] University of Cambridge, Department of Public Health and Primary Care, Cambridge, UK.
[200] The University of Melbourne, Department of Clinical Pathology, Melbourne, Victoria, Australia.
[201] Royal Alexandra Hospital, Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, Edmonton, Alberta, Canada.
[202] Institut Curie, Service de Génétique, Paris, France.
[203] Université Paris Descartes, Paris, France.
[204] University of South Florida, Epidemiology Center, College of Medicine, Tampa, FL, USA.
[205] The Institute of Cancer Research, Division of Breast Cancer Research, London, UK.
[206] Medical University of Vienna, Dept of OB/GYN and Comprehensive Cancer Center, Vienna, Austria.
[207] University of Porto, Biomedical Sciences Institute (ICBAS), Porto, Portugal.
[208] Cancer Research Malaysia, Breast Cancer Research Programme, Subang Jaya, Selangor, Malaysia.
[209] University of Malaya, Department of Surgery, Faculty of Medicine, Kuala Lumpur, Malaysia.
[210] Brigham and Women's Hospital and Harvard Medical School, Obstetrics and Gynecology Epidemiology Center, Boston, MA, USA.
[211] Columbia University, Department of Epidemiology, Mailman School of Public Health, New York, NY, USA.
[212] Magee-Womens Hospital, University of Pittsburgh School of Medicine, Department of Medicine, Pittsburgh, PA, USA.
[213] McGill University, Program in Cancer Genetics, Departments of Human Genetics and Oncology, Montréal, QC, Canada.
[214] University of Cambridge, Department of Medical Genetics, Cambridge, UK.
[215] Dartmouth College, Geisel School of Medicine, Hanover, NH, USA.
[216] The Ohio State University, Department of Cancer Biology and Genetics, Columbus, OH, USA.
[217] Pontificia Universidad Javeriana, Institute of Human Genetics, Bogota, Colombia.
[218] University of Oxford, Cancer Epidemiology Unit, Oxford, UK.
[219] Beth Israel Deaconess Medical Center, Department of Medical Oncology, Boston, MA, USA.
[220] Harvard T.H. Chan School of Public Health, Department of Epidemiology, Boston, MA, USA.
[221] University Medical Center Groningen, University Groningen, Department of Genetics, Groningen, The Netherlands.
[222] University of Pretoria, Department of Genetics, Arcadia, South Africa.
[223] Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain.
[224] Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain.
[225] Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain.
[226] Vanderbilt University Medical Center, Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Department of Biomedical Sciences, Women's Health Research, Nashville, TN, USA.
[227] Duke Cancer Institute, Cancer Control and Population Sciences, Durham, NC, USA.
[228] Duke University Hospital, Department of Community and Family Medicine, Durham, NC, USA.

[229] National Institute of Environmental Health Sciences, NIH, Biostatistics and Computational Biology Branch, Research Triangle Park, NC, USA.
[230] City of Hope, Clinical Cancer Genomics, Duarte, CA, USA.
[231] Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
[232] Stanford University School of Medicine, Department of Biomedical Data Science, Stanford, CA, USA.
[233] Uppsala University, Department of Surgical Sciences, Uppsala, Sweden.
[234] University of Malaya, Department of Obstetrics and Gynaecology, University of Malaya Medical Centre, Kuala Lumpur, Malaysia.
[235] Hebei Medical University, Fourth Hospital, Department of Molecular Biology, Shijiazhuang, China.
[236] National Centre for Scientific Research 'Demokritos', Molecular Diagnostics Laboratory, INRASTES, Athens, Greece.
[237] Institute of Biochemistry and Experimental Oncology, First Faculty od Medicine, Charles University, Czech Republic
[238] Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Centre, Department of Obstetrics and Gynecology, Los Angeles, CA, USA.
[239] 1032 Royal Pass Road, Tampa, FL, USA.
[240] University of NSW Sydney, School of Women's and Children's Health, Faculty of Medicine, Sydney, New South Wales, Australia.
[241] University of NSW Sydney, Adult Cancer Program, Lowy Cancer Research Centre, Sydney, New South Wales, Australia.
[242] University of Michigan School of Public Health, Department of Epidemiology, Ann Arbor, MI, USA.
[243] University of Southern California Norris Comprehensive Cancer Center, Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA.
[244] Mayo Clinic, Department of Health Science Research, Division of Epidemiology, Rochester, MN, USA.
[245] Emory University, Department of Epidemiology, Rollins School of Public Health, Atlanta, GA, USA.

*These authors had equal contributions.

Corresponding Author: Paul D. P. Pharoah (pp10001@medschl.cam.ac.uk)

Running Title: Polygenic Risk Modelling for Prediction of Epithelial Ovarian Cancer Risk

ABSTRACT

Polygenic risk scores (PRS) for epithelial ovarian cancer (EOC) have the potential to improve risk stratification. Joint estimation of Single Nucleotide Polymorphism (SNP) effects in models could improve predictive performance over standard approaches of PRS construction. Here, we implemented computationally-efficient, penalized, logistic regression models (lasso, elastic net, stepwise) to individual level genotype data and a Bayesian framework with continuous shrinkage, "select and shrink for summary statistics" (S4), to summary level data for epithelial non-mucinous ovarian cancer risk prediction. We developed the models in a dataset consisting of 23,564 non-mucinous EOC cases and 40,138 controls participating in the Ovarian Cancer Association Consortium (OCAC) and validated the best models in three populations of different ancestries: prospective data from 198,101 women of European ancestries; 7,669 women of East Asian ancestries; 1,072 women of African ancestries, and in 18,915 *BRCA1* and 12,337 *BRCA2* pathogenic variant carriers of European ancestries. In the external validation data, the model with the strongest association for non-mucinous EOC risk derived from the OCAC model development data was the S4 model (27,240 SNPs) with odds ratios (OR) of 1.38(95%CI:1.28–1.48,AUC:0.588) per unit standard deviation, in women of European ancestries; 1.14(95%CI:1.08–1.19,AUC:0.538) in women of East Asian ancestries; 1.38(95%CI:1.21-1.58,AUC:0.593) in women of African ancestries; hazard ratios of 1.36(95%CI:1.29–1.43,AUC:0.592) in *BRCA1* pathogenic variant carriers and 1.49(95%CI:1.35-1.64,AUC:0.624) in *BRCA2* pathogenic variant carriers. Incorporation of the S4 PRS in risk prediction models for ovarian cancer may have clinical utility in ovarian cancer prevention programs.

Keywords: polygenic risk score, penalized regression model

INTRODUCTION

Rare variants in known high and moderate penetrance susceptibility genes (*BRCA1*, *BRCA2*, *BRIP1*, *PALB2*, *RAD51C*, *RAD51D* and the mis-match repair genes) account for about 40% of the inherited component of EOC disease risk (1,2). Common susceptibility variants, reviewed in Kar et. al. and Jones et. al., explain about 6% of the heritability of EOC (1,3). Polygenic risk scores (PRS) provide an opportunity for refined risk stratification in the general population and in carriers of rare moderate or high risk alleles.

A PRS is calculated as the weighted sum of the number of risk alleles carried for a specified set of variants. The best approach to identify the variant set and their weights to optimize the predictive power of a PRS is unknown. A common approach involves selecting a set of variants that reach a threshold for association based on the p-value for each variant with or without pruning to remove highly correlated variants (4,5). More complex machine learning approaches that do not assume variant independence have also been used (6,7), but these methods have produced only modest gains in predictive power for highly polygenic phenotypes (6,8). Penalized regression approaches such as the lasso, elastic net and the adaptive lasso have also been used with individual level data (9), but a major drawback is the computational burden required to fit the models (9,10).

We present novel, computationally-efficient PRS models using two approaches: 1) penalized regression models including the lasso, elastic net and minimax concave penalty for use with individual genotype data; and 2) a Bayesian regression model with continuous shrinkage priors for use where only summary statistics are available - referred to as the "select and shrink with summary statistics" (S4) method. We compare these models with two commonly used methods, stepwise regression with p-value thresholding and LDPred.

MATERIALS (SUBJECTS) AND METHODS

Model Development Study Population

EOC is a highly heterogeneous phenotype with five major histotypes for invasive disease – high-grade serous, low-grade serous, endometrioid, clear cell and mucinous histotype. The mucinous histotype is the least common and its origin is the most controversial with up to 60% of diagnosed cases of mucinous ovarian cancer often being misdiagnosed metastasis from non-ovarian sites (11). Therefore, in this study, we performed PRS modelling and association testing for all cases of invasive, non-mucinous EOC. We used genotype data from 23,564 invasive non-mucinous EOC cases and 40,138 controls with >80% European ancestries from 63 case-control studies included in the Ovarian Cancer Association Consortium (OCAC) for model development. The distribution of cases by histotype was high-grade serous (13,609), low-grade serous (2,749), endometrioid (2,877), clear cell (1,427), and others (2,902). Sample collection, genotyping and quality control have been previously described (12). Genotype data were imputed to the Haplotype Reference Consortium reference panel using 470,825 SNPs that passed quality control. Of the 32 million SNPs imputed, 10 million had imputation $r^2 > 0.3$ and were included in this analysis.

Model Validation Study Populations

We validated the best-fitting PRS models developed in the OCAC data in 657 prevalent and incident cases of invasive, non-mucinous EOC and 198,101 female controls of European ancestries from the UK Biobank. Samples were genotyped using either the Affymetrix UK BiLEVE Axiom Array or Affymetrix UK Biobank Axiom Array (which share 95% marker content), and then imputed to a combination of the Haplotype Reference Consortium, the 1000Genomes phase 3 and the UK10K reference panels (13). We restricted analysis to genetically confirmed females of European ancestries. We excluded individuals if they were outliers for heterozygosity, had low genotyping call rate <95%, had sex chromosome aneuploidy, or if they were duplicates (cryptic or intended) (12). All SNPs selected in the model development phase were available in the UK Biobank.

We investigated transferability of the best-fitting PRS models to populations of non-European ancestries using genotype data from females of East Asian and African ancestries genotyped as part of the OCAC OncoArray Project (14,15). Women of East Asian ancestries - 2,841 non-mucinous invasive EOC and 4,828 controls - were identified using a criterion of >80% Asian ancestries. This included samples collected from studies in China, Japan, Korea and Malaysia and samples from studies conducted in the US, Europe and Australia (14). Similarly, women of African ancestries - 368 cases of non-mucinous invasive EOC and 704 controls - mainly from studies conducted in the US, were identified using a criterion of >80% African ancestries as described previously (15).

We also assessed the performance of the best-fitting PRS models in women of European ancestries (>80% European ancestries) with the pathogenic *BRCA1* and *BRCA2* variants from the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA). We used genotype data from 18,915 *BRCA1* (2,053 invasive EOC cases) and 12,337 *BRCA2* (717 invasive EOC cases) pathogenic variant carriers from 63 studies contributing to CIMBA (16). Genotyping, data quality control measures, intercontinental ancestries assessment and imputation to the HRC reference panel are as described for the OCAC study population.

Statistical Analysis

Polygenic Risk Models

For all PRS models, we created scores as linear functions of the allele dosage in the general form $PRS_i = \sum_j^p x_{ij}\beta_j$ where genotypes are denoted as x (taking on the minor allele dosages of 0, 1 and 2), with x_{ij} representing the ith individual for the jth SNP (out of p SNPs) on an additive log scale and β_j represents the weight - the log of the odds ratio - of the jth SNP. We used different approaches to select and derive the optimal weights, β_j, in models as described below.

Penalized logistic regression models

A penalized logistic regression model for a set of SNPs aims to identify a set of regression coefficients that minimize the regularized loss function given by

$$plr(x;\lambda,\kappa) = \begin{cases} x - \lambda sign(x)/(1-\kappa) & if|x| < \lambda/\kappa \; and \; |(x)| > \lambda \\ x \; if|x| \geq \dfrac{\lambda}{\kappa} \\ 0 \; if|(x)| < \lambda \end{cases}$$

where x is the effect estimate of a SNP, λ is the tuning parameter and κ is the threshold (penalty) for different regularization paths. λ and κ are parameters that need to be chosen during model development to optimize performance. The lasso, elastic net, minimax concave penalty (MCP), and p-value thresholds are instances of the function with different κ values. We minimized the winner's curse effect on inflated effect estimates for rare SNPs by penalizing rarer SNPs more heavily than common SNPs. Details are provided in the Supplementary Methods.

We used a two-stage approach to reduce computational burden without a corresponding loss in predictive power. The first stage was a SNP selection stage using a sliding windows approach, with 5.5Mb data blocks and a 500kb overlap between blocks. SNP selection was performed for each block and selected SNPs were collated. Single SNP association analyses were then run, and all SNPs with a χ^2 test statistic of less than 2.25 were excluded. The 2.25 cutoff was arbitrary and selected to maximise computational efficiency without loss in predictive power. Penalized regression models were applied to the remaining SNPs using λ values of 3.0 and κ values of 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. SNPs selected in any of these models were included in subsequent analyses. In the second stage, we fit penalized regression models to the training dataset with λ values ranging from 3.0 to 5.5 in increments of 0.1 iterated over κ values from -3.0 to 1 in increments of 0.1. The lasso model ($\kappa = 0$) for each value of λ was fitted first, to obtain a unique maximum. From the fitted maximum the κ value was changed, and the model refitted.

We applied this two-stage approach with five-fold cross-validation (**Figure 1**). In each iteration, the data set was split into five, with one part constituting the the test data and the other four constituting the training data. The variants and their weights from the two-stage penalized logistic regression modelling in the training data were used to calculate the area under the receiver operating characteristic curve (AUC) in the test data in each iteration. AUC estimates for each combination of λ and κ were obtained. We repeated this process for each cross-validation iteration to obtain a mean AUC for each combination of λ and κ. Finally, we selected the tuning and threshold parameters from the lasso, elastic net and minimax concave penalty models with the maximum mean cross-validated AUC and fitted penalized logistic regression models with these parameters to the entire OCAC dataset to obtain SNP weights for PRS scores.

Stepwise logistic regression with variable P-value threshold

This model is a general PLR model with κ=1. As with the other PLR models, we investigated various values for λ values (corresponding to a variable P-value threshold for including a SNP in the model). However, we observed that the implementation of this model on individual level data was more difficult than for other κ values because the model would sometimes converge to a local optimum rather than the global optimum. Therefore, we applied an approximate conditional and joint association analysis using summary level statistics correcting for estimated LD between SNPs, using a reference panel of 5,000 individual level genotype OCAC data as described in Yang et.al. (17). Details are provided in the Supplementary Methods.

LDPred

LDPred is a Bayesian approach that shrinks the posterior mean effect size of each marker based on a point-normal prior and LD information from an external reference panel. We derived seven candidate polygenic risk scores assuming the fractions of associated variants

were 0.001, 0.003, 0.01, 0.03, 0.1, 0.3 and 1.0 respectively using the default parameters as detailed in Vilhjálmsson et.al. (18) using an LD reference panel of 503 samples of European ancestries from the 1000 Genomes phase 3 release and effect estimates from the OCAC model development data.

Select and shrink using summary statistics (S4)

The S4 algorithm is similar to the PRS-CS algorithm (19) - a Bayesian method that uses summary statistics and between-SNP correlation data from a reference panel to generate the PRS scores by placing a continuous shrinkage prior on effect sizes. We adapted this method with penalization of rarer SNPs by correcting for the standard deviation resulting in the selection of fewer SNPs. We varied three parameters, *a, b,* φ, which control the degree of shrinkage of effect estimates. Φ, the overall shrinkage parameter, is influenced by values of *a* which controls shrinkage of effect estimates around 0 and *b* which control shrinkage of larger effect estimates. We generated summary statistics for each cross-validation training set and selected the parameters that gave the best results on average from the cross-validation and applied these to the set of summary statistics for the complete OCAC data set to obtain the final set of weights.

PRS based on meta-analysis of OCAC-CIMBA summary statistics

We conducted a meta-analysis of the EOC associations in *BRCA1* variant carriers, *BRCA2* variant carriers and the participants participating in OCAC (see Supplementary Methods) and constructed two PRS models. An S4 PRS was generated by applying the *a, b* and φ parameters from the S4 model described above. A stepwise PRS was generated by selecting all SNPs that were genome-wide significant ($p < 5 \times 10^{-8}$) in the meta-analysis, along with any independent signals in the same region with $p < 10^{-5}$ from the histotype specific analyses for low-grade serous, high-grade serous, endometrioid, clear cell ovarian cancer and non-mucinous invasive EOC.

Polygenic risk score performance

The best lasso, elastic net, stepwise and S4 models from the model development stage were validated using two independent data sources: the UK Biobank data and *BRCA1/BRCA2* pathogenic variant carriers from the CIMBA. In the UK Biobank data, we evaluated discriminatory performance of the models using the AUC and examined the association between standardized PRS and risk of non-mucinous EOC using logistic regression analysis. For the CIMBA data, we assessed associations for each version of the PRS and invasive non-mucinous EOC risk using weighted Cox regression methods (20). PRSs in the CIMBA data were scaled to the same PRS standard deviations as the OCAC data, meaning that per standard deviation hazard ratios estimated on CIMBA data are comparable to PRS associations in the OCAC and UK Biobank data. The regression models were adjusted for birth cohort (<1920, 1920-1929, 1930-1939, 1940-1949, ≥1950) and the first four ancestries informative principal components (calculated separately by iCOGS/OncoArray genotyping array) and stratified by Ashkenazi Jewish ancestries and country. Absolute risks by PRS percentiles adjusting for competing risks of mortality from other causes were calculated as described in the Supplementary Material.

Transferability of PRS scores to non-European Ancestries

We implemented two straightforward approaches to disentangle the role of ancestries on polygenic risk scoring. We selected homogenous ancestral samples by using a high cut-off criterion of 80% ancestries and we standardized the polygenic risk scores by mean-centering within each population. These approaches led to a more uniform distribution of polygenic risk scores within each ancestral population. Further adjustments using principal components of ancestries did not attenuate risk estimates.

RESULTS

Model development

The results for the models based on individual level genotype data are shown in **Table 1**. The elastic net model had the best predictive accuracy (AUC=0.586). The optimal value of λ obtained from regularization paths for the MCP model was 3.3 meaning the best MCP model was equivalent to the lasso model. The best-fitting model based on summary statistics was the S4 (AUC=0.593) and the LDPred model had the poorest performance of the methods tested (AUC=0.552. It was not considered for further validation in other datasets. All SNPs selected and the associated weights for each model are provided in **Supplementary Tables 1 – 6**.

Model validation in women of European ancestries

Overall the PLR models performed slightly better in the UK Biobank data than the model development data (**Table 2**). Of the models developed using the OCAC model development data, the association was strongest with the S4 PRS. In *BRCA1* and *BRCA2* variant carriers, prediction accuracy was generally higher among *BRCA2* carriers than *BRCA1* carriers. Consistent with results from the general population in the UK Biobank, the S4 PRS model also had the strongest association and predictive accuracy for invasive EOC risk in both *BRCA1* and *BRCA2* carriers. Sensitivity analyses were conducted in which the unadjusted models for *BRCA1* and *BRCA2* carriers were progressively adjusted for birth cohort and 6 principal components. There was little difference in HR estimates and association P-values going from the unadjusted model to the model adjusting for six principal components (Supplementary Table 7). The PRS models developed using the OCAC-CIMBA meta-analysis results had better discriminative ability in the UK Biobank than the PRS models developed using only OCAC data. Compared with the S4 PRS using only OCAC data, the S4 PRS model derived from the meta-analysis had fewer SNPs, a stronger association with invasive EOC risk and better predictive accuracy. Similarly, the stepwise model from the OCAC-CIMBA meta-

analysis performed better than the stepwise model from only OCAC data, but included more SNPs.

The observed distribution of the OR estimates within centiles of the PRS distribution were consistent with ORs from predicted values under the assumption that all SNPs interact multiplicatively (**Figure 3**), with all 95% confidence intervals intersecting with the theoretical estimates for women of European ancestries. Compared with women in the middle quintile, women in the top 95th percentile of the lasso derived PRS model had a 2.23-fold increased odds of non-mucinous EOC (**Table 3**).

Absolute Risk of Developing Ovarian Cancer by PRS percentiles

We estimated cumulative risk of EOC within PRS percentiles for women in the general population (**Figure 2**), by applying the odds ratio from the PRS models to age-specific population incidence and mortality data for England in 2016. For *BRCA1* and *BRCA2* pathogenic variant carriers, we applied the estimated hazard ratios from PRS models to age-specific incidence rates obtained from Kuchenbaecker et al. (21). For women in the general population, the estimated cumulative risks of EOC by age 80 for women at the 99th centile of the PRS distribution were 2.24%, 2.18%, 2.54% and 2.81% for the lasso, elastic net, stepwise and S4 models, respectively. In comparison, the absolute risks of EOC by age 80 for women at the 1st centile were 0.76%, 0.78%, 0.64% and 0.56% for the lasso, elastic net, stepwise and S4 models, respectively.

The absolute risks of developing EOC in *BRCA1* and *BRCA2* pathogenic variant carriers were considerably higher than for women in the general population (Figures S1 and S2). The estimated absolute risk of developing ovarian cancer by age 80 for *BRCA1* carriers at the 99th PRS centiles were 63.2%, 66.3%, 59.0% and 68.4% for the lasso, elastic net, stepwise and S4 models, respectively. The corresponding absolute risks for women at the 1st PRS centile were 27.7%, 25.6%, 30.8% and 24.2%. For *BRCA2* carriers the absolute risks for women at the

714 99th centile were 36.3%, 36.3%, 33.0% and 36.9%; and 7.10%, 7.12%, 8.24% and 6.92% at the
715 1st centile for the lasso, elastic net, stepwise and S4 models, respectively.

716 ***PRS distribution and ancestries***

717 To investigate the transferability of the PRS to other populations, we applied the scores
718 to women of African (N=1,072) and Asian (N=7,669) ancestries genotyped as part of the
719 OncoArray project. In general, the distributions of the raw PRS were dependent on both the
720 statistical methods used in SNP selection and ancestral group. PRS models that included more
721 variants had less dispersion, such that the elastic net models had the least between individual
722 variation in all ancestral groups (standard deviation=0.15, 0.19 and 0.22 for individuals of
723 Asian, African and European ancestries respectively), while the distributions from the stepwise
724 models were the most dispersed (standard deviation = 0.23, 0.27 and 0.30 for individuals of
725 Asian, African and European ancestries respectively). As expected, given the variation in
726 variant frequencies by population, the distribution of polygenic scores was significantly
727 different across the three ancestral groups, with the least dispersion among women of Asian
728 ancestries and the most variation in women of European ancestries. The difference in polygenic
729 risk score distribution was minimized after correction for ancestry by standardizing the PRS to
730 have unit standard deviation using the control subjects for each ancestral group.

731 High PRSs were significantly associated with risk of non-mucinous EOC in both Asian
732 and African ancestries (**Table 4**), although the effects were weaker than in women of European
733 ancestries. For example, with the lasso model, the odds ratio per unit standard deviation
734 increment in polygenic score was 1.16 (1.11–1.22) in women of East Asian ancestries, 1.28
735 (1.13–1.45) in women of African ancestries and 1.37 (1.27–1.48) in women of European
736 ancestries (p for heterogeneity < 0.0001). Variability in effect sizes among ancestral groups
737 was highest for the stepwise model (I^2 = 92%) versus 84% and 83% for elastic net and lasso
738 derived polygenic scores respectively. The best discriminative model among women of East

Asian and African ancestries were the elastic net PRS (AUC=0.543) and the S4 PRS derived from OCAC-CIMBA meta-analysis (AUC=0.596). Women of African ancestries in the top 5% of the PRS had about two-fold increased risk compared to women in the middle quintile (lasso OR:1.64,95%CI: 0.90–3.00; elastic net OR:1.64,95%CI:0.90–3.00; stepwise OR:2.15, 95%CI:1.17–3.95; S4 OR:1.80, 95%CI:0.99–3.31). Effect estimates were smaller in women of East Asian ancestries with women in the top 5% of the PRS, having about a 1.5 fold increased risk compared to women in the middle quintile (lasso OR:1.40, 95%CI:1.12–1.76; elastic net OR:1.60, 95% CI:1.28–2.01; stepwise OR:1.32, 95%CI:1.04–1.65; S4 OR:1.32, 95%CI:1.05–1.66).

DISCUSSION

Genetic risk profiling with polygenic risk scores has led to actionable outcomes for cancers such as breast and prostate (22,23). Previous PRS scores for invasive EOC risk in the general population and *BRCA1/BRCA2* pathogenic variant carriers have been based on genetic variants for which an association with EOC risk had been established at nominal genome-wide significance (20,24,25). Here, we explored the predictive performance of computationally-efficient, penalized, regression methods in modelling joint SNP effects for EOC risk prediction in diverse populations and compared them with common approaches. By leveraging the correlation between SNPs which do not reach nominal genome-wide thresholds and including them in PRS models, the polygenic risk scores derived from penalized regression models provide stronger evidence of association with risk of non-mucinous EOC than previously published PRSs in both the general population and in *BRCA1/BRCA2* pathogenic variant carriers.

Recently, Barnes et. al derived a PRS score using 22 SNPs that were significantly associated with high-grade serous EOC risk (PRS_{HGS}) to predict EOC risk in *BRCA1/BRCA2* pathogenic variant carriers (20). To make effect estimates obtained in this analysis comparable

764 to the effect estimates obtained from the PRS_{HGS}, we standardized all PRSs using the standard
765 deviation from unaffected *BRCA1/BRCA2* carriers and provide estimates which are directly
766 comparable to the PRS_{HGS} in Supplementary Table 9. All PRS models in this analysis except
767 the Stepwise (OCAC only) had higher effect estimates (20). The AUC estimates from the
768 adjusted PLR methods implemented in this analysis, are higher than the corresponding PRS_{HGS}
769 estimates for *BRCA1* carriers (0.604). In *BRCA2* carriers, the AUC estimates for the lasso and
770 S4 models did slightly better than the PRS_{HGS} AUC estimate (0.667), while the stepwise did
771 slightly worse and the elastic net estimate was comparable. The AUC estimates for women in
772 the general population, as estimated from the UK Biobank, are slightly higher than estimates
773 from previously published PRS models for overall EOC risk by Jia et al (AUC=0.57) and Yang
774 et al (AUC=0.58) (25,26).

775 The level of risk for women above the 95th percentile of the PRS is similar to that
776 conferred by pathogenic variants in moderate penetrance genes such as *FANCM* (RR=2.1,
777 95%CI=1.1–3.9) and *PALB2* (RR=2.91 95%CI=1.40–6.04) (27,28). The inclusion of other risk
778 factors such as family history of ovarian cancer, presence of rare pathogenic variants, age at
779 menarche, oral contraceptive use, hormone replacement therapy, parity, and endometriosis in
780 combination with the PRS could potentially improve risk stratification as implemented in the
781 CanRisk tool (www.canrisk.org), which currently uses a 36-SNP PRS with the potential to use
782 other PRS models (29,30).

783 We found that the discrimination of the PRS varied by ancestry with greater
784 discrimination in women of European ancestries than in women of African and East Asian
785 ancestries. The better performance in African than East Asian populations is in contrast to what
786 one would expect given human demographic history, and the performance of PRS for other
787 phenotypes in African populations. This may simply be the play of chance given the small
788 number of samples of African ancestries. Alternatively it reflects the fact that the allele

789 frequencies of the PRS SNPs were more similar between the African and European populations
790 than they were with the East Asian population (Supplementary Tables 10-14).

791 Further optimization of the models could be achieved by varying the penalization
792 function based on prior knowledge. For example, varying the penalty function to select more
793 SNPs from genomic regions with know susceptibility variants given that susceptibility variants
794 tend to cluster together. Alternatively, the penalty functions could be modified to incorporate
795 information about functionally active regions of the genome such a promoters, enhancers and
796 transcription factor binding sites. However, incorporating functional annotation has resulted in
797 limited gains in prediction accuracy for complex traits such as breast cancer, celiac disease,
798 type 2 diabetes and rheumatoid arthritis (31).

799 Machine/deep learning approaches are alternative ways to constructing PRS, but
800 methods such as the neural net, support vector machine and random forest have been shown to
801 be computationally prohibitive or produce inferior results to other approaches (32,33).Other
802 machine learning methods, such as those based on gradient boosting do not perform well in
803 genomic regions where strong genetic interactions are present, for which alternative
804 approaches such as the LDPred may perform better (18). Our approach has several benefits
805 over alternative machine learning methods, including its simplicity, and intrinsic robustness to
806 minor misspecification of LD or association strength.

807 In conclusion, our results indicate that using the lasso model for individual level
808 genotype data and the S4 model for summary level data in polygenic risk score construction
809 provide an improvement in risk prediction for non-mucinous EOC over more common
810 approaches. Our approach overcomes the computational limitations in the use of penalized
811 methods for large scale genetic data, particularly in the presence of highly-correlated SNPs and
812 the use of cross-validation for parameter estimation is preferred. In practical terms, the
813 polygenic risk score provides sufficient discrimination, particularly for women of European

ancestries, to be considered for inclusion in risk prediction and prevention approaches for EOC in the future. Further studies are required to optimize these polygenic risk scores in ancestrally diverse populations and to validate their performance with the inclusion of other genetic and lifestyle risk factors.

Data Availability

OncoArray germline genotype data for the OCAC studies have been deposited at the European Genome-phenome Archive (EGA; https://ega-archive.org/), which is hosted by the EBI and the CRG, under accession EGAS00001002305. Summary statisitics for the Ovarian Cancer Association Consortium are available in the NHGRI-EBI GWAS catalogue (https://www/ebi.ac.uk/gwas/home) under the accession number GCST90016665. A subset of the OncoArray germline genotype data for the CIMBA studies are publically available through the database of Genotypes and Phenotypes (dbGaP) under accession phs001321.v1.p1. The complete data set will not be made publically available because of restraints imposed by the ethics committees of individual studies; requests for further data can be made to the Data Access Coordination Committee (http://cimba.ccge.medschl.cam.ac.uk/)

Acknowledgements and Funding

See Supplementary Material

Conflicts of Interest

Anna DeFazio has received a research grant from AstraZeneca, not directly related to the content of this manuscript. Matthias W. Beckmann conducts research funded by Amgen, Novartis and Pfizer. Peter A. Fashing conducts research funded by Amgen, Novartis and Pfizer. He received Honoraria from Roche, Novartis and Pfizer. Allison W. Kurian reports research funding to her institution from Myriad Genetics for an unrelated project. Usha Menon owns stocks in Abcodia Ltd. Rachel A. Murphy is a consultant for Pharmavite. The other authors declare no conflicts of interest.

Ethics Statement

All study participants provided written informed consent and participated in research or clinical studies at the host institute under ethically approved protocols. The studies and their approving institutes are listed in the Supplementary Material (Ethics Statement)

References

1. Jones MR, Kamara D, Karlan BY, Pharoah PDP, Gayther SA. Genetic epidemiology of ovarian cancer and prospects for polygenic risk prediction. Gynecol Oncol. 2017 Dec;147(3):705–13.

2. Lyra PCM, Rangel LB, Monteiro ANA. Functional Landscape of Common Variants Associated with Susceptibility to Epithelial Ovarian Cancer. Curr Epidemiol Rep. 2020 Mar 1;7(1):49–57.

3. Kar SP, Berchuck A, Gayther SA, Goode EL, Moysich KB, Pearce CL, et al. Common Genetic Variation and Susceptibility to Ovarian Cancer: Current Insights and Future Directions. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2018 Apr;27(4):395–404.

4. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007 Oct;17(10):1520–8.

5. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug 6;460(7256):748–52.

6. Abraham G, Kowalczyk A, Zobel J, Inouye M. Performance and robustness of penalized and unpenalized methods for genetic prediction of complex human disease. Genet Epidemiol. 2013 Feb;37(2):184–95.

7. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics. 2011 May 23;12:186.

8. Szymczak S, Biernacka JM, Cordell HJ, González-Recio O, König IR, Zhang H, et al. Machine learning in genome-wide association studies. Genet Epidemiol. 2009;33 Suppl 1:S51-57.

9. Privé F, Aschard H, Blum MGB. Efficient Implementation of Penalized Regression for Genetic Risk Prediction. Genetics. 2019 May;212(1):65–74.

10. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genet Epidemiol. 2017 Sep;41(6):469–80.

11. Perren TJ. Mucinous epithelial ovarian carcinoma. Ann Oncol Off J Eur Soc Med Oncol. 2016 Apr;27 Suppl 1:i53–7.

12. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet. 2017 May;49(5):680–91.

13. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.

14. Lawrenson K, Song F, Hazelett DJ, Kar SP, Tyrer J, Phelan CM, et al. Genome-wide association studies identify susceptibility loci for epithelial ovarian cancer in east Asian women. Gynecol Oncol. 2019 May;153(2):343–55.

15. Manichaikul A, Peres LC, Wang X-Q, Barnard ME, Chyn D, Sheng X, et al. Identification of novel epithelial ovarian cancer loci in women of African ancestry. Int J Cancer. 2020 Jun 1;146(11):2987–98.

16. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet. 2017 May;49(5):680–91.

17. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012 Mar 18;44(4):369–75, S1-3.

18. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet. 2015 Oct 1;97(4):576–92.

19. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019 Apr 16;10(1):1776.

20. Barnes DR, Rookus MA, McGuffog L, Leslie G, Mooij TM, Dennis J, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. Genet Med Off J Am Coll Med Genet. 2020 Jul 15;

21. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017 20;317(23):2402–16.

22. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2019 Jan 3;104(1):21–34.

23. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018 Jul;50(7):928–36.

24. Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. J Natl Cancer Inst. 2017 Jul 1;109(7).

25. Yang X, Leslie G, Gentry-Maharaj A, Ryan A, Intermaggio M, Lee A, et al. Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. J Med Genet. 2018 Aug;55(8):546–54.

26. Jia G, Lu Y, Wen W, Long J, Liu Y, Tao R, et al. Evaluating the Utility of Polygenic Risk Scores in Identifying High-Risk Individuals for Eight Common Cancers. JNCI Cancer Spectr. 2020 Jun;4(3):pkaa021.

27. Song H, Dicks EM, Tyrer J, Intermaggio M, Chenevix-Trench G, Bowtell DD, et al. Population-based targeted sequencing of 54 candidate genes identifies PALB2 as a susceptibility gene for high-grade serous ovarian cancer. J Med Genet. 2021 May;58(5):305–13.

28. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, et al. Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. J Clin Oncol Off J Am Soc Clin Oncol. 2020 Mar 1;38(7):674–85.

926 29. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al.
927 BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic
928 and nongenetic risk factors. Genet Med Off J Am Coll Med Genet. 2019 Aug;21(8):1708–
929 18.

930 30. Welcome to CanRisk [Internet]. [cited 2021 Aug 15]. Available from: https://canrisk.org/

931 31. Hu Y, Lu Q, Powles R, Yao X, Yang C, Fang F, et al. Leveraging functional annotations
932 in genetic risk prediction for human complex diseases. PLoS Comput Biol. 2017
933 Jun;13(6):e1005589.

934 32. Gola D, Erdmann J, Müller-Myhsok B, Schunkert H, König IR. Polygenic risk scores
935 outperform machine learning methods in predicting coronary artery disease status. Genet
936 Epidemiol. 2020 Mar;44(2):125–38.

937 33. Paré G, Mao S, Deng WQ. A machine-learning heuristic to improve gene score prediction
938 of polygenic traits. Sci Rep. 2017 Oct 4;7(1):12665.

939

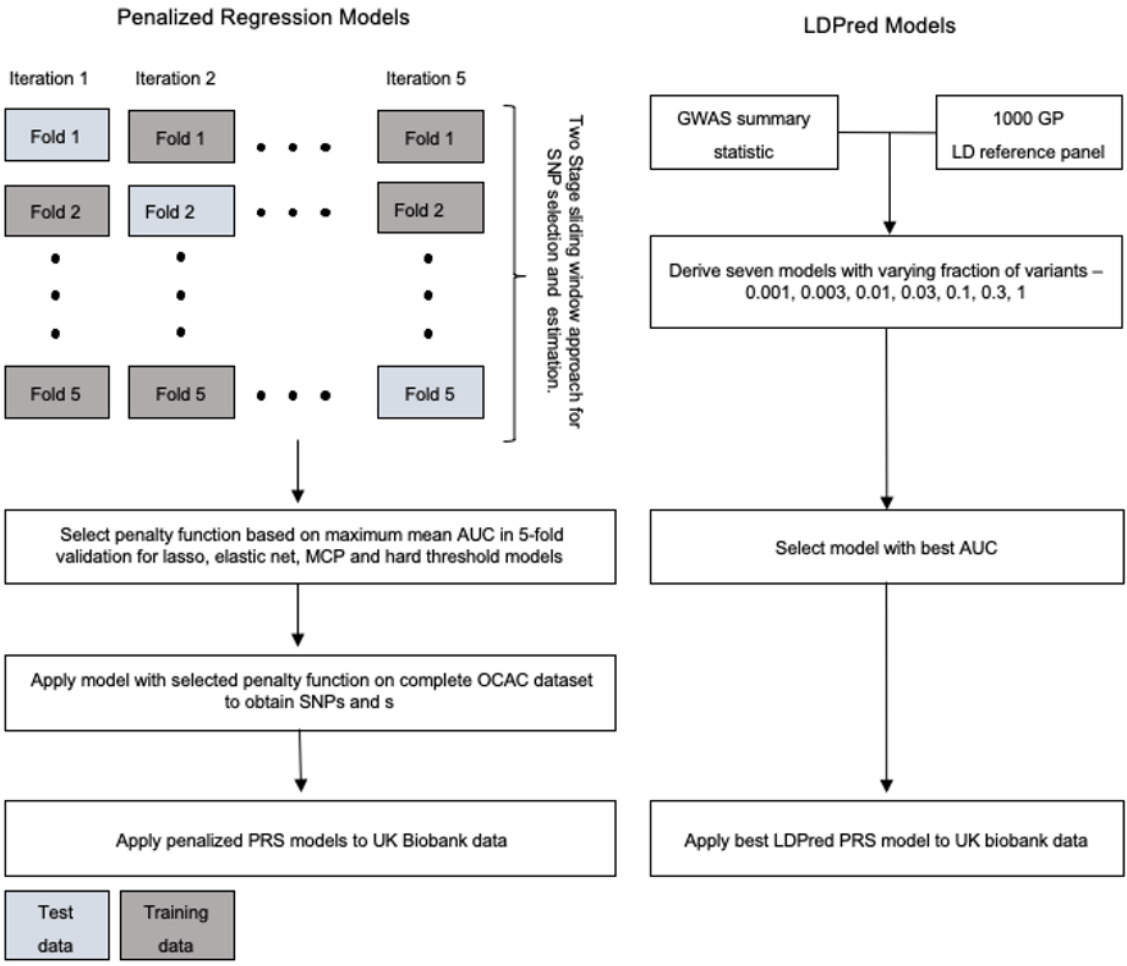
940

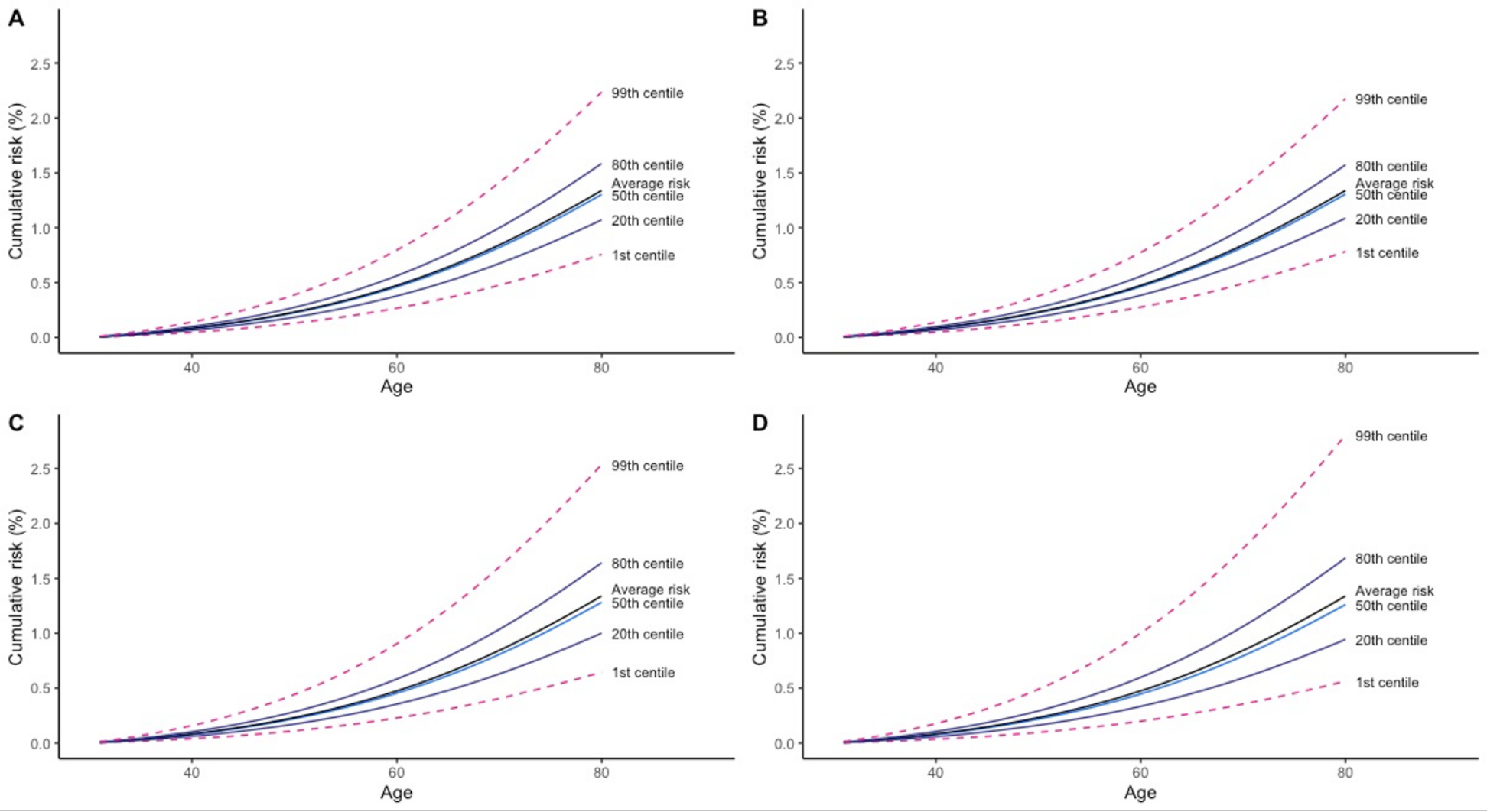
Figure captions

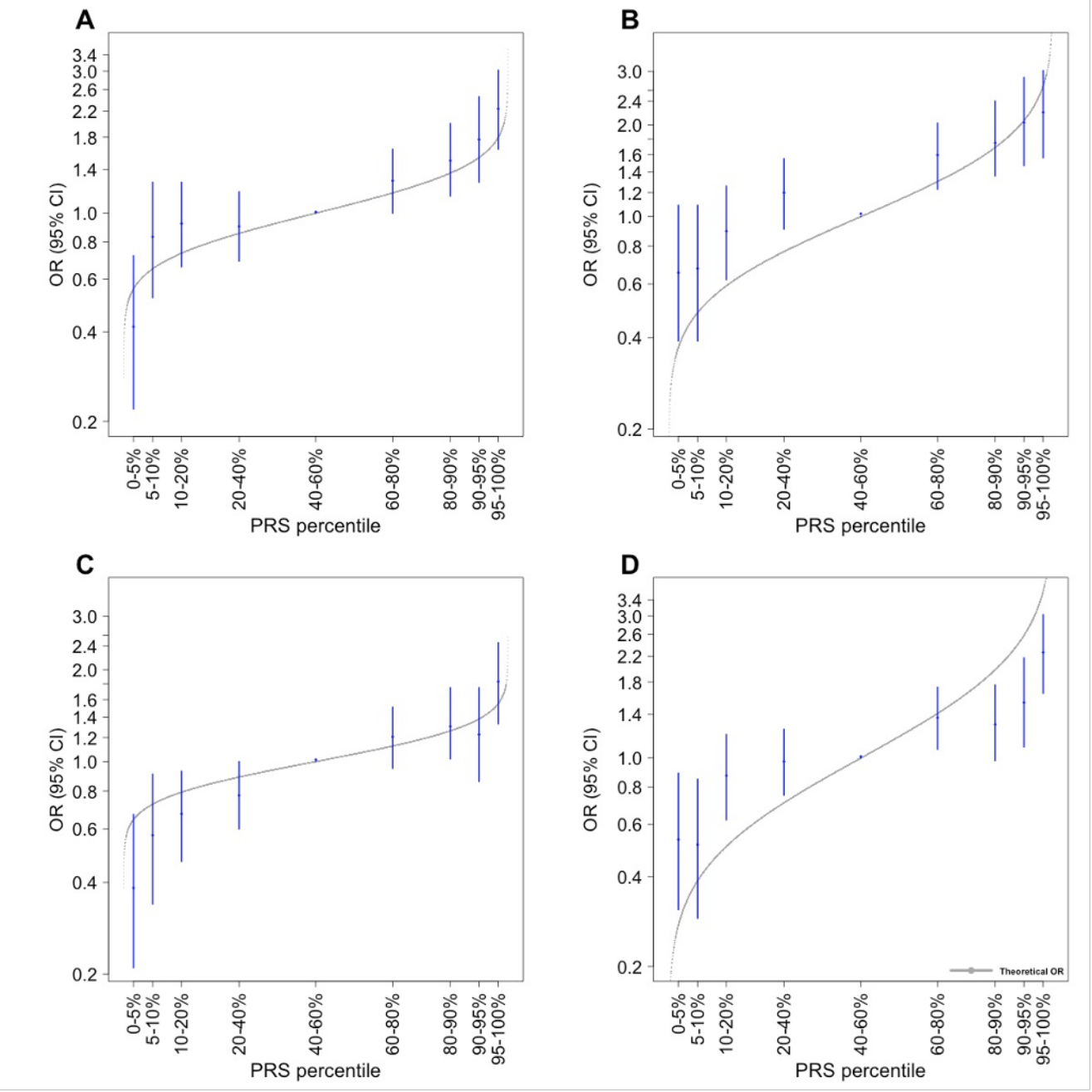
Figure 1: PRS model development using penalized regression and LDPred Bayesian approach

Figure 2: Cumulative risk of ovarian cancer between birth and age 80 by PRS percentiles and PRS models. Shown are the cumulative risk of ovarian cancer risk in UK women by polygenic risk score percentiles. The lasso (A) and elastic net (B) penalized regression models were applied to individual level genotype data, while the stepwise (C) and S4 (D) models were applied to summary level statistics**.** Note that the median and the mean risk differ because the distribution of the relative risk in the population is left-skewed (the log relative risk is a Normal distribution)

Figure 3: Association between the PLR PRS models and non-mucinous ovarian cancer by PRS percentiles. Shown are estimated odds ratios (OR) and confidence intervals for women of European ancestries by percentiles of polygenic risk scores derived from lasso (A), elastic net (B), stepwise (C) and S4 (D) models relative to the middle quintile.


Penalized Regression Models
Iteration 1
Iteration 2
Iteration 5
Fold 1
Fold 2
Fold 5
Two Stage sliding window approach for SNP selection and estimation.
Select penalty function based on maximum mean AUC in 5-fold validation for lasso, elastic net, MCP and hard threshold models
Apply model with selected penalty function on complete OCAC dataset to obtain SNPs and s
Apply penalized PRS models to UK Biobank data
Test data
Training data
LDPred Models
GWAS summary statistic
1000 GP LD reference panel
Derive seven models with varying fraction of variants – 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1
Select model with best AUC
Apply best LDPred PRS model to UK biobank data


A
B
C
D
Cumulative risk (%)
0.0
0.5
1.0
1.5
2.0
2.5
40
60
80
Age
99th centile
80th centile
Average risk
50th centile
20th centile
1st centile


A
B
C
D
OR (95% CI)
PRS percentile
0-5%
5-10%
10-20%
20-40%
40-60%
60-80%
80-90%
90-95%
95-100%
Theoretical OR

SUPPLEMENTARY MATERIAL

Supplementary Method

Penalized logistic regression models: the lasso, elastic net and minimax concave penalty.

A penalized logistic regression model for a set of SNPs aims to identify a set of regression coefficients that minimize the regularized loss function. The aim is to correct the effect estimates towards zero in order to account for prior likelihood. The regularized loss function is given by

$$plr(x;\lambda,\kappa) = \begin{cases} x - \lambda sign(x)/(1-\kappa) & if|x| < \lambda/\kappa \; and \; |(x)| > \lambda \\ x \; if|x| \geq \dfrac{\lambda}{\kappa} \\ 0 \; if|(x)| < \lambda \end{cases}$$

where x is the estimate of SNP effect, λ is the tuning parameter and κ is the threshold (penalty) for different regularization paths. λ and κ are parameters that need to be chosen during model development to optimize performance. The lasso, elastic net, minimax concave penalty (MCP), and p-value thresholds are instances of the function with different κ values. At $\kappa = 0$, the function is a lasso regularization path and several regression coefficients are shrunk to zero. When $\kappa < 0$, the function is an elastic-net solution. The MCP model is obtained when $0 \leq \kappa \leq 1$. When $\kappa = 1$, the function is equivalent to a P-value threshold model.

As the winner's curse effect can result in inflated estimates for rarer SNPs, it is desirable to penalize rarer SNPs more heavily than common SNPs. We therefore rescaled and normalized the imputed SNP genotypes (coded as number of alternate alleles carried) as follows:

The alternate allele dose, d_i, is first rescaled so that the mean value is zero

$$d_i^{rescaled} = d_i - mean(d)$$

The rescaled alternate allele dose is then normalized:

$$d_i^{normalized} = \frac{d_i^{rescaled}}{\sqrt{(\sum_i p_i(1-p_i))(var(d)+0.008)/var(d)}}$$

where p_i is the estimated probability of individual i being a case based on a simple logistic regression model. This conversion penalizes rarer SNPs and also means that the tuning parameter, λ, used in a penalized logistic regression model corresponds closely to a threshold for including a SNP if the SNP z-score for association is greater than λ.

There are two general problems with fitting penalized models to high dimensional genetic data: the large number of SNPs which need to be fitted into memory and the correlation between SNPs. We used a two-stage approach to reduce computational burden without a corresponding loss in predictive power. The first stage was a selection stage to reduce the set of SNPs to a manageable number by excluding those unlikely to be included in a final model. We used a sliding windows approach and split the data into 5.5Mb blocks with a 500kb overlap between blocks. The SNP selection analysis was carried out for each block and the selected SNPs collated. Single SNP association analyses were run, and all SNPs with a X^2 test statistic of less than 2.25 were excluded. Lasso regression models were then run on the remaining SNPs using lambda values of 3.0 and kappa values (penalty score) of 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. SNPs selected in any of these models were included in subsequent analyses. In the second stage, we fit penalized regression models to the training dataset with λ values from 3.0 to 5.5 in increments of 0.1 iterated over κ values from -3.0 to 1 in increments of 0.1. The lasso model ($\kappa = 0$) for each value of lambda was fitted first, which has a unique maximum. From the fitted maximum the kappa value is changed, and the model fitted again. This helps to ensure that the final fitted model is close to the global maximum for non-zero

values of kappa. For small lambda values the larger kappa values were not analysed as these would not fit the data well and require significant computational resources and time.

We applied this two-stage approach with five-fold cross-validation by splitting the dataset into five groups, and randomly assigning each individual to one of the five groups. We restricted to OCAC studies with at least five controls and five cases, so that each group in the cross-validation analyses would have cases and controls from all studies in the OCAC data (**Figure 1**). In each iteration, four-groups were used as the training set and the fifth group used as the test set. The variants and their weights from the two-stage penalized logistic regression modelling in the training data were used to calculate the area under the receiver operating characteristic curve (AUC) in the test data. We repeated this process for each cross-validation iteration to obtain a mean AUC for each combination of λ and κ.

Finally, we selected the tuning and threshold parameters from the lasso, elastic net and minimax concave penalty models with the maximum mean cross-validated AUC and fitted penalized logistic regression models with these parameters to the entire OCAC dataset to obtain SNP weights for PRS scores. Details of the workflow are provided in Figure 1.
We developed a custom program written in C++ to implement the penalized logistic regression PLR models using individual level genotype data. This used the coordinate descent algorithm where the loss of function for each variable is optimized using a univariable Newton step while keeping other variables constant to obtain the optimum penalized likelihood function (36).

Stepwise logistic regression with variable P-value threshold

This model is a general PLR model with $\kappa = 1$. As with the other PLR models, we investigated various values for λ values (corresponding to a variable P-value threshold for including a SNP in the model). However, we observed that the implementation of this model on individual level data was more difficult than for other κ values because the model would sometimes converge to a local optimum rather than the global optimum. Therefore, we applied an approximate conditional and joint association analysis using summary level statistics correcting for estimated LD between SNPs using a reference panel of 5,000 individual level genotype OCAC data as described in Yang et.al (37). In brief, to initiate the model, we identified the most significant SNP at a specified λ value threshold, and subsequently investigated all other SNPs across the genome conditional on the SNP already selected. If a new SNP being jointly tested had correlation $r^2 > 0.9$, it was left out of the analysis. The SNP with the minimum conditional SNP lower than the cut-off threshold was selected. This process was repeated until no new SNPs could be selected. All selected SNPs were jointly fitted in the model and SNPs with the largest p-value greater than the cut-off threshold were dropped. This process was repeated until no new SNPs were selected or dropped using the training data. As with the PLR models, we investigated several λ values. We used the same cross-validated sets as were used by the PLR methods. The lambda coefficient that gave the best average likelihood increase in the cross-validation test set was applied to the full OCAC data set to obtain the final set of weights.

Meta-analysis of OCAC-CIMBA summary statistics

We conducted a meta-analysis of the EOC associations in *BRCA1* pathogenic variant carriers, *BRCA2* pathogenic variant carriers and the participants participating in OCAC using previously described methodological approaches (3). In brief, the primary analyses of SNPs

and EOC risk in *BRCA1/BRCA2* pathogenic variant carriers included samples genotyped on the iCOGS and OncoArray genotyping chips (with a preference for data from the OncoArray chip when samples were genotyped on both platforms), excluded all overlapping OCAC samples (148 *BRCA1* and 51 *BRCA2*) and SNPs not included in the OncoArray chip. Association analyses were carried out separately for *BRCA1* pathogenic variant carriers and *BRCA2* pathogenic variant carriers, within a survival analysis framework, with time to ovarian cancer diagnosis as the end point. Hazards ratios (HRs) from BRCA1 and BRCA2 pathogenic variants carriers were pooled to give an overall *BRCA1+BRCA2* effect estimate. SNPs exhibiting evidence of heterogeneity ($P_{het}<1x10^{-4}$) were not considered for downstream analyses. Per-allele HR estimates from *BRCA1+BRCA2* pathogenic variant carriers were combined with per allele odds ratios from OCAC to provide an overall relative risk invasive EOC. SNPs with allele frequencies <0.5% were excluded from consideration. All meta-analyses were fixed-effects inverse-variance weighted and implemented using the METAL software (39).

Absolute risk calculations

We calculated the absolute risk of EOC by PRS percentiles adjusting for competing risks of mortality from other causes. We obtained incidence rates of EOC associated with an average risk profile and mortality estimates from 2016 population-based registries in the UK. The absolute risk of EOC in PRS percentiles at a given age (AR_{prs} (t)) is calculated as

$$AR_{PRS}(t) = \sum_{u=0}^{t} I_{PRS}(u).S_{PRS}(u).S_m(u)$$

Where $I_{PRS}(u)$ is the EOC incidence associated with the PRS at age t, $S_{PRS}(t)$ is the PRS-percentile specific probability of being free of ovarian cancer up to age t and $S_m(t)$ is the probability of surviving to age t, *i.e.,* not dying from a cause other than EOC

SUPPLEMENTARY TABLES

Table S1: Effect estimates (weights) of SNPs selected for lasso PRS model.

Table S2: Effect estimates (weights) of SNPs selected for elastic net PRS model.

Table S3: Effect estimates (weights) of SNPs selected for Stepwise (OCAC) PRS model.

Table S4: Effect estimates (weights) of SNPs selected for Select and Shrink (OCAC) PRS model.

Table S5: Effect estimates (weights) of SNPs selected for Stepwise (OCAC-CIMBA) PRS model.

Table S6: Effect estimates (weights) of SNPs selected for Slect and Shrink (OCAC-CIMBA) PRS model.

SUPPLEMENTARY FIGURES

Figure S1: Cumulative risk of ovarian cancer risk in *BRCA1* carriers by polygenic risk score percentiles. The lasso (A) and elastic net (B) penalized regression models were applied to individual level genotype data, while the stepwise (C) and S4 (D) models were applied to summary level statistics.

Figure S2: Cumulative risk of ovarian cancer risk in *BRCA2* carriers by polygenic risk score percentiles. The lasso (A) and elastic net (B) penalized regression models were applied to individual level genotype data, while the stepwise (C) and S4 (D) models were applied to summary level statistics.

ETHICS STATEMENT

All study participants provided written informed consent and participated in research studies at the host institute under ethically approved protocols.

OCAC FUNDING AND ACKNOWLEDGEMENT

Funding

The Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PPD/RPCI.07). The scientific development and funding for this project were in part supported by the US National Cancer Institute GAME-ON Post-GWAS Initiative (U19-CA148112). This study made use of data generated by the Wellcome Trust Case Control consortium that was funded by the Wellcome Trust under award 076113. The results published here are in part based upon data generated by The Cancer Genome Atlas Pilot Project established by the National Cancer Institute and National Human Genome Research Institute (dbGap accession number phs000178.v8.p7).

The OCAC OncoArray genotyping project was funded through grants from the U.S. National Institutes of Health (CA1X01HG007491-01 (C.I.A.), U19-CA148112 (T.A.S.), R01-CA149429 (C.M.P.) and R01-CA058598 (M.T.G.); Canadian Institutes of Health Research (MOP-86727 (L.E.K.) and the Ovarian Cancer Research Fund (A.B.). The COGS project was funded through a European Commission's Seventh Framework Programme grant (agreement number 223175 - HEALTH-F2-2009-223175).

Funding for individual studies: **AAS:** National Institutes of Health (RO1-CA142081); **AOV:** The Canadian Institutes for Health Research (MOP-86727); **AUS:** The Australian Ovarian Cancer Study Group was supported by the U.S. Army Medical Research and Materiel Command under DAMD17-01-1-0729, The Cancer Council Tasmania and The Cancer Foundation of Western Australia and the National Health and Medical Research Council of Australia (NHMRC; ID400413 and ID400281). The Australian Ovarian Cancer Study

gratefully acknowledges additional support from S. Boldeman, the Agar family, Ovarian Cancer Action (UK), Ovarian Cancer Australia and the Peter MacCallum Foundation; **BAV:** ELAN Funds of the University of Erlangen-Nuremberg; **BEL:** National Kankerplan; **BGS:** Breast Cancer Now, Institute of Cancer Research; **BVU:** Vanderbilt University Medical Center's BioVU is supported by the 1S10RR025141-01 instrumentation award and Vanderbilt CTSA grant from the National Institutes of Health (NIH)/National Center for Advancing Translational Sciences (NCATS) (ULTR000445); **CAM:** National Institutes of Health Research Cambridge Biomedical Research Centre and Cancer Research UK Cambridge Cancer Centre; **CHA:** Innovative Research Team in University (PCSIRT) in China (IRT1076); **CNI:** Instituto de Salud Carlos III (PI 12/01319); Ministerio de Economía y Competitividad (SAF2012); **DKE:** Ovarian Cancer Research Fund; **DOV:** National Institutes of Health R01-CA112523 and R01-CA87538; **EMC:** Dutch Cancer Society (EMC 2014-6699); **EPC:** 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 Danish Cancer Society (Denmark) (EMC 2014-6699); 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 (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); ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition and Health (Norway); 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); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); 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); **GER:** German Federal Ministry of Education and Research, Programme of Clinical Biomedical Research (01 GB 9401) and the German Cancer Research Center (DKFZ); **GRC:** This research has been co-financed by 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) - Research Funding Program of the General Secretariat for Research & Technology: SYN11_10_19 NBCA. Investing in knowledge society through the European Social Fund; **GRR:** Roswell Park Cancer Institute Alliance Foundation, P30 CA016056; **HAW:** U.S. National Institutes of Health (R01-CA58598, N01-CN-55424 and N01-PC-67001); **HJO:** Intramural funding; Rudolf-Bartling Foundation; **HMO:** Intramural funding; Rudolf-Bartling Foundation; **HOC:** Helsinki University Hospital Research Fund; **HOP:** University of Pittsburgh School of Medicine Dean's Faculty Advancement Award (F. Modugno), Department of Defense (DAMD17-02-1-0669) and NCI (K07-CA080668, R01-CA95023, P50-CA159981 MO1-RR000056 R01-CA126841); **HUO:** Intramural funding; Rudolf-Bartling Foundation; **JPN:** Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare; **KRA:** This study (Ko-EVE) was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (HI16C1127; 0920010); **LAX:** American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN) and the National Center for Advancing Translational Sciences (NCATS), Grant UL1TR000124; **LUN:** ERC-2011-AdG

294576-risk factors cancer, Swedish Cancer Society, Swedish Research Council, Beta Kamprad Foundation; **MAC:** National Institutes of Health (R01-CA122443, P30-CA15083, P50-CA136393); Mayo Foundation; Minnesota Ovarian Cancer Alliance; Fred C. and Katherine B. Andersen Foundation; Fraternal Order of Eagles; **MAL:** Funding for this study was provided by research grant R01- CA61107 from the National Cancer Institute, Bethesda, MD, research grant 94 222 52 from the Danish Cancer Society, Copenhagen, Denmark; and the Mermaid I project; **MAS:** Malaysian Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Initiatives Foundation; **MAY:** National Institutes of Health (R01-CA122443, P30-CA15083, P50-CA136393); Mayo Foundation; Minnesota Ovarian Cancer Alliance; Fred C. and Katherine B. Andersen Foundation; **MCC:** MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. Cancer Council Victoria, National Health and Medical Research Council of Australia (NHMRC) grants number 209057, 251533, 396414, and 504715; **MDA:** DOD Ovarian Cancer Research Program (W81XWH-07-0449); **MEC:** NIH (CA54281, CA164973, CA63464); **MOF:** Moffitt Cancer Center, Merck Pharmaceuticals, the state of Florida, Hillsborough County, and the city of Tampa; **NCO:** National Institutes of Health (R01-CA76016) and the Department of Defense (DAMD17-02-1-0666); **NEC:** National Institutes of Health R01-CA54419 and P50-CA105009 and Department of Defense W81XWH-10-1-02802; **NHS:** UM1 CA186107, P01 CA87969, R01 CA49449, R01-CA67262, UM1 CA176726; **NJO:** National Cancer Institute (NIH-K07 CA095666, R01-CA83918, NIH-K22-CA138563, and P30-CA072720) and the Cancer Institute of New Jersey; If Sara Olson and/or Irene Orlow is a co-author, please add NCI CCSG award (P30-CA008748) to the funding sources; **NOR:** Helse Vest, The Norwegian Cancer Society, The Research Council of Norway; **NTH:** Radboud University Medical Centre; **OPL:** National Health and Medical Research Council (NHMRC) of Australia (APP1025142) and Brisbane Women's Club; **ORE:** Sherie Hildreth Ovarian Cancer (SHOC) Foundation;

OVA: This work was supported by Canadian Institutes of Health Research grant (MOP-86727) and by NIH/NCI 1 R01CA160669-01A1; **PLC:** Intramural Research Program of the National Cancer Institute; **POC:** Pomeranian Medical University; **POL:** Intramural Research Program of the National Cancer Institute; **PVD:** Canadian Cancer Society and Cancer Research Society GRePEC Program; **RBH:** National Health and Medical Research Council of Australia; **RMH:** Cancer Research UK, Royal Marsden Hospital; **RPC:** National Institute of Health (P50 CA159981, R01CA126841); **SEA:** Cancer Research UK (C490/A10119 C490/A10124); UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge; **SIS:** 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); **SMC:** 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); **SRO:** Cancer Research UK (C536/A13086, C536/A6689) and Imperial Experimental Cancer Research Centre (C1312/A15589); **STA:** NIH grants U01 CA71966 and U01 CA69417; **SWH:** NIH (NCI) grant R37-CA070867; **TBO:** National Institutes of Health (R01-CA106414-A2), American Cancer Society (CRTG-00-196-01-CCE), Department of Defense (DAMD17-98-1-8659), Celma Mastry Ovarian Cancer Foundation; **TOR:** NIH grants R01 CA063678 and R01 CA063682; **UCI:** NIH R01-CA058860 and the Lon V Smith Foundation grant LVS-39420; **UHN:** Princess Margaret Cancer Centre Foundation-Bridge for the Cure; **UKO:** The UKOPS study was funded by The Eve Appeal (The Oak Foundation) with investigators supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre and MRC core funding (MR_UU_12023); **UKR:** Cancer Research UK (C490/A6187), UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge; **USC:** P01CA17054, P30CA14089, R01CA61132, N01PC67010, R03CA113148,

R03CA115195, N01CN025403, and California Cancer Research Program (00-01389V-20170, 2II0200); **VAN:** BC Cancer Foundation, VGH & UBC Hospital Foundation; **VTL:** NIH K05-CA154337; **WMH:** National Health and Medical Research Council of Australia, Enabling Grants ID 310670 & ID 628903. Cancer Institute NSW Grants 12/RIG/1-17 & 15/RIG/1-16; **WOC:** National Science Centre (N N301 5645 40), The Maria Sklodowska-Curie Memorial Cancer Centre and Institute of Oncology, Warsaw, Poland.

This work was supported by the Cedars-Sinai Precision Health Initiative (awarded to Simon Gayther and Michelle Jones.

Additional research funding and support included: UM1 CA186107, P01 CA87969, (Nurses' Health Study); U01 CA176726, R01 CA67262, (Nurses' Health Study II).

Acknowledgements

We are grateful to the family and friends of Kathryn Sladek Smith for their generous support of the Ovarian Cancer Association Consortium through their donations to the Ovarian Cancer Research Fund. The OncoArray and COGS genotyping projects 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, Marjorie J. Riggan (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, Jonathan P. Tyrer, Siddhartha Kar, 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. We pay special tribute to the contribution of Professor Brian Henderson to the GAME-ON consortium; to Olga M. Sinilnikova for her contribution to CIMBA and for her part in the initiation and coordination of GEMO until she sadly passed away on the 30th June 2014 and to Catherine M. Phelan for her contribution to OCAC and coordination of the OncoArray until she passed away on 22 September 2017. We thank the study participants, doctors, nurses, clinical and scientific collaborators, health care providers and health information sources who have contributed to the many studies contributing to this manuscript.

Acknowledgements for individual studies: **AOV:** We thank Jennifer Koziak, Mie Konno, Michelle Darago, Faye Chambers and the Tom Baker Cancer Centre Translational Laboratories; **AUS:** The AOCS also acknowledges the cooperation of the participating institutions in Australia, and the contribution of the study nurses, research assistants and all clinical and scientific collaborators. The complete AOCS Study Group can be found at www.aocstudy.org. We would like to thank all of the women who participated in this research program; **BEL:** We would like to thank Gilian Peuteman, Thomas Van Brussel, Annick Van den Broeck and Joke De Roover for technical assistance; **BGS:** The Institute of Cancer Research (ICR) acknowledges NHS funding to the NIHR Biomedical Research Centre. We thank the Study staff, study participants , doctors, nurses, health care providers and health information sources who have contributed to the study and Breast Cancer Now and The Institute of Cancer Research for their support; **BVU:** The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU; **CHA:** Innovative Research Team in University (PCSIRT) in China (IRT1076); **CHN:** To thank all members of Department of Obstetrics and Gynaecology, Hebei Medical University, Fourth

Hospital and Department of Molecular Biology, Hebei Medical University, Fourth Hospital; **COE:** Gynecologic Cancer Center of Excellence; **EPC:** To thank all members and investigators of the Rotterdam Ovarian Cancer Study; **GER:** The German Ovarian Cancer Study (GER) thank Ursula Eilber for competent technical assistance; **MAS:** We would like to thank Famida Zulkifli and Ms Moey for assistance in patient recruitment, data collection and sample preparation; **MCC:** 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 and the Australian Cancer Database; **MOF:** the Total Cancer Care™ Protocol and the Collaborative Data Services and Tissue Core Facilities at the H. Lee Moffitt Cancer Center & Research Institute, an NCI designated Comprehensive Cancer Center (P30-CA076292), Merck Pharmaceuticals and the state of Florida; **NHS:** The NHS/NHSII studies 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, and WY; **OPL:** Members of the OPAL Study Group (http://opalstudy.qimrberghofer.edu.au/); **SEA:** SEARCH team, Craig Luccarini, Caroline Baynes, Don Conroy; **SRO:** To thank all members of Scottish Gynaecological Clinical Trails group and SCOTROC1 investigators; **SWH:** We thank the participants and the research staff of the Shanghai Women's Health Study for making this study possible; **UHN:** Princess Margaret Cancer Centre Foundation-Bridge for the Cure; **UKO:** We particularly thank I. Jacobs, M.Widschwendter, E. Wozniak, A. Ryan, J. Ford and N. Balogun for their contribution to the study**; UKR:** Carole Pye; **VAN:** BC Cancer Foundation, VGH & UBC Hospital Foundation; **WMH:** We thank the Gynaecological Oncology Biobank at Westmead, a member of the Australasian Biospecimen Network-Oncology group.

The authors would like to thank the participants and staff of the NHS/NHSII for their valuable contributions as well as 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. The authors assume full responsibility for analyses and interpretation of these data.

CIMBA FUNDING AND ACKNOWLEDGEMENT

Funding

The CIMBA data management and data analysis were supported by Cancer Research – UK grants C12292/A20861, C12292/A11174. GCT and ABS are NHMRC Research Fellows. iCOGS: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, 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 (CRN-87521), and the Ministry of Economic Development, Innovation and Export Trade (PSR-SIIRI-701), Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The PERSPECTIVE and PERSPECTIVE I&I projects were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministry of Economy and Innovation through Genome Québec, and The Quebec Breast Cancer Foundation and the Ontario Research Fund. BCFR: UM1 CA164920 from the National Cancer Institute. 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 US Government or the BCFR. BFBOCC: Lithuania (BFBOCC-LT): Research Council of Lithuania grant SEN-18/2015. BIDMC: Breast Cancer Research Foundation. BMBSA: Cancer Association of South Africa (PI Elizabeth J. van Rensburg). CNIO: Spanish Ministry of Health PI16/00440 supported by FEDER funds, the Spanish Instituto de Salud Carlos III (grant PI19/00640) supported by FEDAR funds and the Spanish Research Network on Rare diseases (CIBERER). COH-CCGCRN: Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under grant number R25CA112486, and RC4CA153828 (PI: J. Weitzel) from the National Cancer Institute and the Office of the Director, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. CONSIT TEAM: Associazione Italiana Ricerca sul Cancro (AIRC; IG2014 no.15547) to P. Radice. Funds from Italian citizens who allocated the 5x1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects ‘5x1000’) to S. Manoukian. Associazione Italiana Ricerca sul Cancro (AIRC; IG2015 no.16732) to P. Peterlongo. DEMOKRITOS: 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) - Research Funding Program of the General Secretariat for Research & Technology: SYN11_10_19 NBCA. Investing in knowledge society through the European Social Fund. DFKZ: German Cancer Research Center. EMBRACE: Cancer Research UK Grants C1287/A17523, C1287/A26886 and C1287/A23382. D. Gareth Evans is supported by an NIHR grant to the Biomedical Research Centre, Manchester (NIHR grant IS-BRC- 1215-20007). The Investigators at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust are supported by an NIHR

grant to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. Ros Eeles and Elizabeth Bancroft are supported by Cancer Research UK Grant C5047/A8385. Ros Eeles is also supported by NIHR support to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. FCCC: The University of Kansas Cancer Center (P30 CA168524), the Kansas Institute for Precision Medicine COBRE (P20 GM130423), the Kansas Bioscience Authority Eminent Scholar Program, R01 CA140323, R01 CA214545, and by the Chancellors Distinguished Chair in Biomedical Sciences Professorship to A.K.G. Ana Vega is supported by the Spanish Health Research Foundation, Instituto de Salud Carlos III (ISCIII) through Research Activity Intensification Program (contract grant numbers: INT15/00070, INT16/00154, INT17/00133), and through Centro de Investigación Biomédica en Red de Enferemdades Raras CIBERER (ACCI 2016: ER17P1AC7112/2018); Autonomous Government of Galicia (Consolidation and structuring program: IN607B), and by the Fundación Mutua Madrileña (call 2018). GC-HBOC: German Cancer Aid (grant no 110837, Rita K. Schmutzler) and the European Regional Development Fund and Free State of Saxony, Germany (LIFE - Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). GEMO: Ligue Nationale Contre le Cancer; the Association “Le cancer du sein, parlons-en!” Award, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program and the French National Institute of Cancer (INCa grants 2013-1-BCB-01-ICH-1 and SHS-E-SP 18-015). GEORGETOWN: the Non-Therapeutic Subject Registry Shared Resource at Georgetown University (NIH/NCI grant P30-CA051008), the Fisher Center for Hereditary Cancer and Clinical Genomics Research, and Swing Fore the Cure. G-FAST: Bruce Poppe is a senior clinical investigator of FWO. Mattias Van Heetvelde obtained funding from IWT. HCSC: Spanish Ministry of Health PI15/00059, PI16/01292, and CB-

161200301 CIBERONC from ISCIII (Spain), partially supported by European Regional Development FEDER funds. HEBCS: Helsinki University Hospital Research Fund, the Finnish Cancer Society and the Sigrid Juselius Foundation. The HEBON study is supported by the Dutch Cancer Society grants NKI1998-1854, NKI2004-3088, NKI2007-3756, the Netherlands Organisation of Scientific Research grant NWO 91109024, the Pink Ribbon grants 110005 and 2014-187.WO76, the BBMRI grant NWO 184.021.007/CP46 and the Transcan grant JTC 2012 Cancer 12-054. HRBCP: Hong Kong Sanatorium and Hospital, Dr Ellen Li Charitable Foundation, The Kerry Group Kuok Foundation, National Institute of Health1R 03CA130065, and North California Cancer Center. HUNBOCS: Hungarian Research Grants KTIA-OTKA CK-80745 and NKFI_OTKA K-112228. ICO: The authors would like to particularly acknowledge the support of the Asociación Española Contra el Cáncer (AECC), the Instituto de Salud Carlos III (organismo adscrito al Ministerio de Economía y Competitividad) and "Fondo Europeo de Desarrollo Regional (FEDER), unamanera de hacer Europa" (PI10/01422, PI13/00285, PIE13/00022, PI15/00854, PI16/00563 and CIBERONC) and the Institut Català de la Salut and Autonomous Government of Catalonia (2009SGR290, 2014SGR338 and PERIS Project MedPerCan). IHCC: PBZ_KBN_122/P05/2004. ILUH: Icelandic Association "Walking for Breast Cancer Research" and by the Landspitali University Hospital Research Fund. INHERIT: 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. IOVHBOCS: Ministero della Salute and "5x1000" Istituto Oncologico Veneto grant. IPOBCS: Liga Portuguesa Contra o Cancro. kConFab: 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. KOHBRA: the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (HI16C1127; 1020350; 1420190). MAYO: NIH grants CA116167, CA192393 and CA176785, an NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), and a grant from the Breast Cancer Research Foundation. MCGILL: Jewish General Hospital Weekend to End Breast Cancer, Quebec Ministry of Economic Development, Innovation and Export Trade. Marc Tischkowitz is supported by the funded by the European Union Seventh Framework Program (2007Y2013)/European Research Council (Grant No. 310018). MODSQUAD: MH CZ - DRO (MMCI, 00209805), MEYS - NPS I - LO1413 to LF, and by Charles University in Prague project UNCE204024 (MZ). MSKCC: the Breast Cancer Research Foundation, the Robert and Kate Niehaus Clinical Cancer Genetics Initiative, the Andrew Sabin Research Fund and a Cancer Center Support Grant/Core Grant (P30 CA008748). NAROD: 1R01 CA149429-01. NCI: the Intramural Research Program of the US National Cancer Institute, NIH, and by support services contracts NO2-CP-11019-50, N02-CP-21013-63 and N02-CP-65504 with Westat, Inc, Rockville, MD. NICCC: Clalit Health Services in Israel, the Israel Cancer Association and the Breast Cancer Research Foundation (BCRF), NY. NNPIO: the Russian Foundation for Basic Research (grants 17-00-00171, 18-515-45012 and 19-515-25001). NRG Oncology: U10 CA180868, NRG SDMC grant U10 CA180822, NRG Administrative Office and the NRG Tissue Bank (CA 27469), the NRG Statistical and Data Center (CA 37517) and the Intramural Research Program, NCI. OSU CCG: Ohio State University Comprehensive Cancer Center. PBCS: Italian Association of Cancer Research (AIRC) [IG 2013 N.14477] and Tuscany Institute for Tumors (ITT) grant 2014-2015-2016. SEABASS: Ministry of Science, Technology and Innovation, Ministry of Higher Education (UM.C/HlR/MOHE/06) and Cancer Research Initiatives Foundation. SMC: the Israeli Cancer

Association. SWE-BRCA: the Swedish Cancer Society. UCHICAGO: NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA125183), R01 CA142996, 1U01CA161032 and by the Ralph and Marion Falk Medical Research Trust, the Entertainment Industry Fund National Women's Cancer Research Alliance and the Breast Cancer research Foundation. OIO is an ACS Clinical Research Professor. UCLA: Jonsson Comprehensive Cancer Center Foundation; Breast Cancer Research Foundation. UCSF: UCSF Cancer Risk Program and Helen Diller Family Comprehensive Cancer Center. UKFOCR: Cancer Research UK. UPENN: National Institutes of Health (NIH) (R01-CA102776 and R01-CA083855; Breast Cancer Research Foundation; Susan G. Komen Foundation for the cure, Basser Research Center for BRCA. UPITT/MWH: Hackers for Hope Pittsburgh. VFCTG: Victorian Cancer Agency, Cancer Australia, National Breast Cancer Foundation. WCP: Dr Karlan is funded by the American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN) and the National Center for Advancing Translational Sciences (NCATS), Grant UL1TR000124. This work was supported by the U.S. National Institute of Health (NIH), National Cancer Institute [1RO1CA159868]; as well as by the Breast Cancer Research Foundation; Australia National Health and Medical Research Council [454508, 288704, 145684]; Victorian Health Promotion Foundation; Victorian Breast Cancer Research Consortium; Cancer Australia [1100868, 809195]; National Breast Cancer Foundation [IF 17]; Queensland Cancer Fund; Cancer Councils of New South Wales, Victoria, Tasmania, and South Australia, and Cancer Foundation of Western Australia. We also thank Heather Thorne, Eveline Niedermayr, Sharon Guo, Stephanie Nesci, Lucy Stanhope, Sarah O'Connor, Sandra Picken, the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the many families who contribute to kConFab. SGBCC was supported by the National Research Foundation Singapore (NRF-NRFF2017-02, awarded to J Li), NUS start-up Grant, National University

Cancer Institute Singapore (NCIS) Centre Grant [NMRC/CG/NCIS/2010, NMRC/CG/012/2013, CGAug16M005], Breast Cancer Prevention Programme (BCPP), Asian Breast Cancer Research Fund, and the NMRC Clinician Scientist Award (SI Category) [NMRC/CSA-SI/0015/2017].

KAP is a National Breast Cancer Foundation (Australia) Practitioner Fellow [grant number PRAC-17-004]. M.A. Caligo was supported by Grant 2016 (prog.127/16) from the Fondazione Pisa and by research funding 2017 from the Susan G. Komen Italia onlus.

Acknowledgements

All the families and clinicians who contribute to the studies; Catherine M. Phelan for her contribution to CIMBA until she passed away on 22 September 2017; Sue Healey, in particular taking on the task of pathogenic variant classification with the late Olga Sinilnikova; Maggie Angelakos, Judi Maskiell, Gillian Dite, Helen Tsimiklis; members and participants in the New York site of the Breast Cancer Family Registry; members and participants in the Ontario Familial Breast Cancer Registry; Vilius Rudaitis and Laimonas Griškevičius; Drs Janis Eglitis, Anna Krilova and Aivars Stengrevics; Yuan Chun Ding and Linda Steele for their work in participant enrollment and biospecimen and data management; Alicia Barroso, Rosario Alonso and Guillermo Pita; all the individuals and the researchers who took part in CONSIT TEAM (Consorzio Italiano Tumori Ereditari Alla Mammella), in particular: Daniela Zaffaroni, Irene Feroce, Mariarosaria Calvello, Davide Bondavalli, Aliana Guerrieri Gonzaga, Monica Marabelli, A. Viel, Laura Ottini, Giuseppe Giannini, Gabriele Lorenzo Capone, Liliana Varesco, Viviana Gismondi, Maria Grazia Tibiletti, Ileana Carnevali, Antonella Savarese, Aline Martayan, Stefania Tommasi, Brunella Pilato and the

personnel of the Cogentech Cancer Genetic Test Laboratory, Milan, Italy. Ms. JoEllen Weaver and Dr. Betsy Bove; The FPGMX group acknowledges members of the Cancer Genetics group (IDIS): Ana Blanco, Marta Santamariña, Miguel Aguado and Belinda Rodríguez-Lage; IFE - Leipzig Research Centre for Civilization Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber); We thank all participants, clinicians, family doctors, researchers, and technicians for their contributions and commitment to the DKFZ. Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers (GEMO) study is a study from the National Cancer Genetics Network UNICANCER Genetic Group, France. We wish to pay a tribute to Olga M. Sinilnikova, who with Dominique Stoppa-Lyonnet initiated and coordinated GEMO until she sadly passed away on the 30th June 2014. The team in Lyon (Olga Sinilnikova, Mélanie Léoné, Laure Barjhoux, Carole Verny-Pierre, Sylvie Mazoyer, Francesca Damiola, Valérie Sornin) managed the GEMO samples until the biological resource centre was transferred to Paris in December 2015 (Noura Mebirouk, Fabienne Lesueur, Dominique Stoppa-Lyonnet). We want to thank all the GEMO collaborating groups for their contribution to this study: Coordinating Centre, Service de Génétique, Institut Curie, Paris, France: Muriel Belotti, Ophélie Bertrand, Anne-Marie Birot, Bruno Buecher, Sandrine Caputo, Anaïs Dupré, Emmanuelle Fourme, Marion Gauthier-Villars, Lisa Golmard, Claude Houdayer, Marine Le Mentec, Virginie Moncoutier, Antoine de Pauw, Claire Saule, Dominique Stoppa-Lyonnet, and Inserm U900, Institut Curie, Paris, France: Fabienne Lesueur, Noura Mebirouk.Contributing Centres : Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon - Centre Léon Bérard, Lyon, France: Nadia Boutry-Kryza, Alain Calender, Sophie Giraud, Mélanie Léone. Institut Gustave Roussy, Villejuif, France: Brigitte Bressac-de-Paillerets, Olivier Caron, Marine Guillaud-Bataille. Centre Jean Perrin, Clermont–Ferrand, France: Yves-Jean Bignon, Nancy Uhrhammer. Centre Léon Bérard, Lyon, France: Valérie Bonadona, Christine Lasset. Centre

François Baclesse, Caen, France: Pascaline Berthet, Laurent Castera, Dominique Vaur. Institut Paoli Calmettes, Marseille, France: Violaine Bourdon, Catherine Noguès, Tetsuro Noguchi, Cornel

Popovici, Audrey Remenieras, Hagay Sobol. CHU Arnaud-de-Villeneuve, Montpellier, France: Isabelle Coupier, Pascal Pujol. Centre Oscar Lambret, Lille, France: Claude Adenis, Aurélie Dumont, Françoise Révillion. Centre Paul Strauss, Strasbourg, France: Danièle Muller. Institut Bergonié, Bordeaux, France: Emmanuelle Barouk-Simonet, Françoise Bonnet, Virginie Bubien, Michel Longy, Nicolas Sevenet, Institut Claudius Regaud, Toulouse, France: Laurence Gladieff, Rosine Guimbaud, Viviane Feillel, Christine Toulas. CHU Grenoble, France: Hélène Dreyfus, Christine Dominique Leroux, Magalie Peysselon, Rebischung. CHU Dijon, France: Amandine Baurand, Geoffrey Bertolone, Fanny Coron, Laurence Faivre, Caroline Jacquot, Sarab Lizard. CHU St-Etienne, France: Caroline Kientz, Marine Lebrun, Fabienne Prieur. Hôtel Dieu Centre Hospitalier, Chambéry, France: Sandra Fert Ferrer. Centre Antoine Lacassagne, Nice, France: Véronique Mari. CHU Limoges, France: Laurence Vénat-Bouvet. CHU Nantes, France: Stéphane Bézieau, Capucine Delnatte. CHU Bretonneau, Tours and Centre Hospitalier de Bourges France: Isabelle Mortemousque. Groupe Hospitalier Pitié-Salpétrière, Paris, France: Chrystelle Colas, Florence Coulet, Florent Soubrier, Mathilde Warcoin. CHU Vandoeuvre-les-Nancy, France: Myriam Bronner, Johanna Sokolowska. CHU Besançon, France: Marie-Agnès Collonge- Rame, Alexandre Damette. CHU Poitiers, Centre Hospitalier d'Angoulême and Centre Hospitalier de Niort, France: Paul Gesta. Centre Hospitalier de La Rochelle: Hakima Lallaoui. CHU Nîmes Carémeau, France: Jean Chiesa. CHI Poissy, France: Denise Molina- Gomes. CHU Angers, France : Olivier Ingster; Ilse Coene en Brecht Crombez; Ilse Coene and Brecht Crombez; Alicia Tosar and Paula Diaque; Drs. Sofia Khan, Taru A. Muranen, Carl Blomqvist, Irja Erkkilä and Virpi Palola; The Hereditary Breast and Ovarian Cancer Research Group

Netherlands (HEBON) consists of the following Collaborating Centers: Netherlands Cancer Institute (coordinating center), Amsterdam, NL: M.A. Rookus, F.B.L. Hogervorst, F.E. van Leeuwen, M.A. Adank, M.K. Schmidt, D.J. Jenner; Erasmus Medical Center, Rotterdam, NL: J.M. Collée, A.M.W. van den Ouweland, M.J. Hooning, I.A. Boere; Leiden University Medical Center, NL: C.J. van Asperen, P. Devilee, R.B. van der Luijt, T.C.T.E.F. van Cronenburg; Radboud University Nijmegen Medical Center, NL: M.R. Wevers, A.R. Mensenkamp; University Medical Center Utrecht, NL: M.G.E.M. Ausems, M.J. Koudijs; Amsterdam Medical Center, NL: T.A.M. van Os; VU University Medical Center, Amsterdam, NL: K. van Engelen, J.J.P. Gille; Maastricht University Medical Center, NL: E.B. Gómez-Garcia, M.J. Blok, M. de Boer; University of Groningen, NL: L.P.V. Berger, A.H. van der Hout, M.J.E. Mourits, G.H. de Bock; The Netherlands Comprehensive Cancer Organisation (IKNL): S. Siesling, J. Verloop; The nationwide network and registry of histo- and cytopathology in The Netherlands (PALGA): E.C. van den Broek. HEBON thanks the study participants and the registration teams of IKNL and PALGA for part of the data collection; Hong Kong Sanatorium and Hospital; the Hungarian Breast and Ovarian Cancer Study Group members (Janos Papp, Aniko Bozsik, Timea Pocza, Zoltan Matrai, Miklos Kasler, Judit Franko, Maria Balogh, Gabriella Domokos, Judit Ferenczi, Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary) and the clinicians and patients for their contributions to this study; the Oncogenetics Group (VHIO) and the High Risk and Cancer Prevention Unit of the University Hospital Vall d'Hebron, Miguel Servet Progam (CP10/00617), and the Cellex Foundation for providing research facilities and equipment; the ICO Hereditary Cancer Program team led by Dr. Gabriel Capella; the ICO Hereditary Cancer Program team led by Dr. Gabriel Capella; Catarina Santos and Pedro Pinto; members of the Center of Molecular Diagnosis, Oncogenetics Department and Molecular Oncology Research Center of Barretos Cancer Hospital; 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; the KOBRA Study Group; Csilla Szabo (National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA); Eva Machackova (Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute and MF MU, Brno, Czech Republic); and Michal Zikan, Petr Pohlreich and Zdenek Kleibl (Oncogynecologic Center and Department of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic); Anne Lincoln, Lauren Jacobs; the participants in Hereditary Breast/Ovarian Cancer Study and Breast Imaging Study for their selfless contributions to our research; the NICCC National Familial Cancer Consultation Service team led by Sara Dishon, the lab team led by Dr. Flavio Lejbkowicz, and the research field operations team led by Dr. Mila Pinchev; the investigators of the Australia New Zealand NRG Oncology group; members and participants in the Ontario Cancer Genetics Network; Kevin Sweet, Caroline Craven, Julia Cooper, Amber Aielts, and Michelle O'Conor; Yip Cheng Har, Nur Aishah Mohd Taib, Phuah Sze Yee, Norhashimah Hassan and all the research nurses, research assistants and doctors involved in the MyBrCa Study for assistance in patient recruitment, data collection and sample preparation, Philip Iau, Sng Jen-Hwei and Sharifah Nor Akmal for contributing samples from the Singapore Breast Cancer Study and the HUKM-HKL Study respectively; the Meirav Comprehensive breast cancer center team at the Sheba Medical Center; Christina Selkirk; Håkan Olsson, Helena Jernström, Karin Henriksson, Katja Harbst, Maria Soller, Ulf Kristoffersson; from Gothenburg Sahlgrenska University Hospital: Anna Öfverholm, Margareta Nordling, Per Karlsson, Zakaria Einbeigi; from Stockholm and Karolinska University Hospital: Anna von

Wachenfeldt, Annika Lindblom, Brita Arver, Gisela Barbany Bustinza; from Umeå University Hospital: Beatrice Melin, Christina Edwinsdotter Ardnor, Monica Emanuelsson; from Uppsala University: Hans Ehrencrona, Maritta Hellström Pigg, Richard Rosenquist; from Linköping University Hospital: Marie Stenmark-Askmalm, Sigrun Liedgren; Cecilia Zvocec, Qun Niu; Joyce Seldon and Lorna Kwan; Dr. Robert Nussbaum, Beth Crawford, Kate Loranger, Julie Mak, Nicola Stewart, Robin Lee, Amie Blanco and Peggy Conrad and Salina Chan; Simon Gayther, Paul Pharoah, Carole Pye, Patricia Harrington and Eva Wozniak; Geoffrey Lindeman, Marion Harris, Martin Delatycki, Sarah Sawyer, Rebecca Driessen, and Ella Thompson for performing all DNA amplification.SGBCC thanks the participants and all research coordinators for their excellent help with recruitment, data and sample collection.

Ethics Statement

CIMBA Ethics Statement

All study participants provided written informed consent and participated in research or clinical studies at the host institute under ethically approved protocols. The studies and their approving institutes are: Australian site of the Breast Cancer Family Registry (BCFR-AU) - The University of Melbourne Health Sciences Human Ethics Sub-Committee; Northern California site of the Breast Cancer Family Registry (BCFR-NC) - Northern California Cancer Center Institutional Review Board; New York site of the Breast Cancer Family Registry (BCFR-NY) - Columbia University Medical Center Institutional Review Board; Ontario site of the Breast Cancer Family Registry (BCFR-ON) - Mount Sinai Hospital Research Ethics Board; Philadelphia site of the Breast Cancer Family Registry (BCFR-PA) - Institutional Review Board Fox Chase Cancer Center; Utah site of the Breast Cancer Family Registry (BCFR-UT) - Institutional Review Board University of Utah; Baltic Familial Breast

and Ovarian Cancer Consortium (BFBOCC) - Centrālā medicīnas ētikas Komiteja; Lietuvos Bioetikos Komitetas; BRCA-gene mutations and breast cancer in South African women (BMBSA) - University of Pretoria and Pretoria Academic Hospitals Ethics Committee; Beckman Research Institute of the City of Hope (BRICOH) - City of Hope Medical Center Institutional Review Board; Copenhagen Breast Cancer Study (CBCS) - De Videnskabsetiske Komiteer I Region Hovedsladen; Spanish National Cancer Centre (CNIO) - Instituto de Salud Carlos III Comité de Bioética y Bienestar Animal; City of Hope Cancer Center (COH) - City of Hope Institutional Review Board; CONsorzio Studi ITaliani sui Tumori Ereditari Alla Mammella (CONSIT TEAM) - Comitato Etico Indipendente della Fondazione IRCCS "Istituto Nazionale dei Tumori"; National Centre for Scientific Research Demokritos (DEMOKRITOS) - Bioethics committee of NCSR ‘‘Demokritos’’, 240/EHΔ/11.3; National Centre for Scientific Research Demokritos (DEMOKRITOS) - Papageorgiou Hospital Ethics Committee; Dana Farber Cancer Institute (DFCI) - Dana Farber Cancer Institute Institutional Review Board; Deutsches Krebsforschungszentrum (DKFZ) - Ethik-Kommission des Klinikums der Universität; Deutsches Krebsforschungszentrum (DKFZ) - Hospital Universitario de San Ignacio Comité de Investigaciones y Etica; Deutsches Krebsforschungszentrum (DKFZ) - Shaukat Khanum Memorial Cancer Hospital and Research Centre Institutional Review Board; Epidemiological study of BRCA1 and BRCA2 mutation carriers (EMBRACE) - Anglia & Oxford MREC; Fox Chase Cancer Center (FCCC) - Institutional Review Board Fox Chase Cancer Center; Fundación Pública Galega de Medicina Xenómica - Comite Autonomico de Etica da Investigacion de Galicia; German Consortium of Hereditary Breast and Ovarian Cancer (GC-HBOC) - Ethik-Kommission der Medizinischen Fakultät der Universät zu Köln; Genetic Modifiers of cancer risk in BRCA1/2 mutation carriers (GEMO) - Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé; Georgetown University (GEORGETOWN) -

MedStar Research Institute - Georgetown University Oncology Institutional Review Board; Ghent University Hospital (G-FAST) - Universitair Ziekenhuis Gent - Ethics Committee; Hospital Clinico San Carlos (HCSC) - Comité Ético de Investigación Clínia Hospital Clínico San Carlos; Helsinki Breast Cancer Study (HEBCS) - Helsingin ja uudenmaan sairaanhoitopiiri (Helsinki University Central Hospital ethics committee); HEreditary Breast and Ovarian study Netherlands (HEBON) - Protocol Toetsingscommissie van het Nederlands Kanker Instituut/Antoni van Leeuwenhoek Ziekenhuis; Molecular Genetic Studies of Breast- and Ovarian Cancer in Hungary (HUNBOCS) - Institutional Review Board of the Hungarian National Institute of Oncology; University Hospital Vall d'Hebron (HVH) - The Hospital Universitario Vall d'Hebron Clinical Research Ethics Committee; Institut Català d'Oncologia (ICO) - Catalan Institute of Oncology Institutional Review Board; International Hereditary Cancer Centre (IHCC) - Komisji Bioetycznej Pomorskiej Akademii Medycznej (Pomeranian Medical University Bioethics Committee); Iceland Landspitali - University Hospital (ILUH) - Vísindasiđanefnd National Boethics Committee; Interdisciplinary Health Research International Team Breast Cancer Susceptibility (INHERIT) - Comité d'éthique de la recherche du Centre Hospitalier Universitaire de Québec; Istituto Oncologico Veneto Hereditary Breast and Ovarian Cancer Study (IOVHBOCS) - Centro Oncologico Regionale Azienda Ospedale Di Padova Comitato Etico; Portuguese Oncology Institute-Porto Breast Cancer Study - COMISSÃO DE ÉTICA PARA A SAÚDE (CES) ; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (KCONFAB) - Queensland Institute of Medical Research - Human Research Ethics Committee; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (KCONFAB) - Peter MacCallum Cancer Centre Ethics Committee; University of Kansas Medical Center(KUMC) - The University of Kansas Medical Center Human Research Protection Program; Mayo Clinic (MAYO) - Mayo Clinic Institutional Review Boards; McGill University (MCGILL) -

McGill Faculty of Medicine Institutional Review Board; Modifier Study of Quantitative Effects on Disease (MOD-SQUAD) - Mayo Clinic Institutional Review Boards; Memorial Sloane Kettering Cancer Center (MSKCC) - Human Biospecimen Utilization Committee; Memorial Sloane Kettering Cancer Center (MSKCC) - Memorial Sloan-Kettering Cancer Center IRB; General Hospital Vienna (MUV) - Ethikkommission der Medizinischen Universität Wien; Women's College Research Institute Hereditary Breast and Ovarian Cancer Study - University of Toronto Health Sciences Review Ethics Board; National Cancer Institute (NCI) - NIH Ethics Office; National Israeli Cancer Control Center (NICCC) - Carmel Medical Center Institutional Review Board (Helsinki Committee); N.N. Petrov Institute of Oncology (NNPIO) - N.N. Petrov Institional Ethical Committee; NorthShore University HealthSystem (NORTHSHORE) - Institutional Review Board of NorthShore University HealthSystem; NRG Oncology (NRG_ONCOLOGY) - Cancer Prevention and Control Protocol Review Committee; Ontario Cancer Genetics Network (OCGN) - University Health Network Research Ethics Board; The Ohio State University Comprehensive Cancer Center (MACBRCA) - The Ohio State University Cancer Institutional Review Board; Odense University Hospital (OUH) - Den Videnskabsetiske Komité for Region Syddanmark; Pisa Breast Cancer Study (PBCS) - Azienda Ospedaliera Pisana Comitato Etico per lo studio del farmaco sull'uomo; Sheba Medical Centre - Chaim Sheba Medical Center IRB; Swedish Breast Cancer Study (SWE-BRCA) - Regionala Etikprövningsnämnden Stockholm; University of Chicago (UCHICAGO) - The University of Chicago Biological Sciences Division.

Institutional Review Board (BSD IRB); University of California Los Angeles (UCLA) - UCLA Institutional Review Board (UCLA IRB); University of California San Francisco (UCSF) - Human Research Protection Program Institutional Review Board (IRB); UK and

Gilda Radner Familial Ovarian Cancer Registries (UKGRFOCR) - Roswell Park Cancer Institute IRB; UK and Gilda Radner Familial Ovarian Cancer Registries (UKGRFOCR) - Cambridge Local Research Ethics Committee; University of Pennsylvania (UPENN) - University of Pennsylvania Institutional Review Board; Cancer Family Registry University of Pittsburg (UPITT) - University of Pittsburgh Institutional Review Board; University of Texas MD Anderson Cancer Center (UTMDACC) - University of Texas MD Anderson Cancer Center Office of Protocol Research Institutional Review Board; Victorian Familial Cancer Trials Group (VFCTG) - Peter MacCallum Cancer Centre Ethics Committee; Women's Cancer Program at Cedars-Sinai Medical Center (WCP) - (Cedars-Sinai Medical Center) CSMC Institutional Review Board.