Common breast cancer risk loci predispose to distinct tumor subtypes

Thomas U. Ahearn[1*], Haoyu Zhang[1,2*], Kyriaki Michailidou[3,4,5], Roger L. Milne[6,7,8], Manjeet K. Bolla[3], Joe Dennis[3], Alison M. Dunning[9], Michael Lush[3], Qin Wang[3], Irene L. Andrulis[10], Hoda Anton-Culver[11], Volker Arndt[12], Kristan J. Aronson[13], Paul L. Auer[14,15], Annelie Augustinsson[16], Adinda Baten[17], Heiko Becher[18], Sabine Behrens[19], Javier Benitez[20], Marina Bermisheva[21], Carl Blomqvist[22,23], Stig E. Bojesen[24,25,26,27], Bernardo Bonanni[28], Anne-Lise Børresen-Dale[29,30], Hiltrud Brauch[31,32,33], Hermann Brenner[12,34,33], Angela Brooks-Wilson[35,36], Thomas Brüning[37], Barbara Burwinkel[38,39], Federico Canzian[40], Jose E. Castelao[41], Jenny Chang-Claude[19,42], Stephen J. Chanock[1], Georgia Chenevix-Trench[43], Christine L. Clarke[44], NBCS Collaborators[†], J.Margriet Collée[45], Angela Cox[46], Simon S. Cross[47], Kamila Czene[48], Mary B. Daly[49], Peter Devilee[50,51], Thilo Dörk[52], Miriam Dwek[53], Diana M. Eccles[54], D.Gareth Evans[55,56], Peter A. Fasching[57,58], Jonine Figueroa[59,60], Giuseppe Floris[17], Manuela Gago-Dominguez[61,62], Susan M. Gapstur[63], José A. García-Sáenz[64], Mia M. Gaudet[63], Graham G. Giles[6,7,8], Mark S. Goldberg[65,66,67], David E. Goldgar[68], Anna González-Neira[27], Grethe I. GrenakerAlnæs[29], Mervi Grip[69], Pascal Guénel[70], Christopher A. Haiman[71], Per Hall[48,72], Ute Torres[73], Elaine F. Harkness[74,75], Bernadette A.M. Heemskerk-Gerritsen[76], Bernd Holleczek[77], Antoinette Hollestelle[76], Maartje J. Hooning[76], Robert N. Hoover[1], John L. Hopper[7], Anthony Howell[78], kConFab/AOCS Investigators[†], Milena Jakimovska[79], Anna Jakubowska[80,81], Esther M. John[82], Michael E. Jones[83], Audrey Jung[19], Rudolf Kaaks[19], Saila Kauppila[84], Renske Keeman[85], Elza Khusnutdinova[86], Cari M. Kitahara[87], Yon-Dschun Ko[88], Stella Koutros[1], Vessela N. Kristensen[29,30], Ute Krüger[16], Katerina Kubelka-Sabit[89], Allison W. Kurian[82], Kyriacos Kyriacou[5,90], Diether Lambrechts[91,92], Derrick G. Lee[93,94], Annika Lindblom[95,96], Martha Linet[87], Jolanta Lissowska[97], Ana Llaneza[98], Wing-Yee Lo[31,99], Robert J.

MacInnis[6,7], Arto Mannermaa[100,101,102], Mehdi Manoochehri[73], Sara Margolin[72,103], Maria Elena Martinez[62,104], Catriona McLean[105], Alfons Meindl[106], Usha Menon[107], Heli Nevanlinna[108], William G. Newman[55,56], Jesse Nodora[109,110], Kenneth Offit[111], Håkan Olsson[16,112], Nick Orr[113], Tjoung-Won Park-Simon[52], Julian Peto[3,114], Guillermo Pita[115], Dijana Plaseska-Karanfilska[79], Ross Prentice[14], Kevin Punie[17], Katri Pylkäs[116,117], Paolo Radice[118], Gad Rennert[119], Atocha Romero[120], Thomas Rüdiger[121], Emmanouil Saloustros[122], Sarah Sampson[123], Dale P. Sandler[124], Elinor J. Sawyer[125], Rita K. Schmutzler[126], Minouk J. Schoemaker[83], Ben Schöttker[12], Mark E. Sherman[127], Xiao-Ou Shu[128], Snezhana Smichkoska[129], Melissa C. Southey[8], John J. Spinelli[130,131], Anthony J. Swerdlow[83,132], Rulla M. Tamimi[133,134,135], William J. Tapper[136], Jack A. Taylor[124,137], MaryBeth Terry[138], Diana Torres[73,139], Melissa A. Troester[140], Celine M. Vachon[141], Carolien H.M. van Deurzen[142], Elke M. van Veen[55,56], Philippe Wagner[16], Clarice R. Weinberg[143], Camilla Wendt[103], Jelle Wesseling[85], Robert Winqvist[51,116,117,144], Alicja Wolk[81,145,146,81], Xiaohong R. Yang[1], Wei Zheng[128], Fergus J. Couch[147], Jacques Simard[148], Peter Kraft[134,135], Douglas F. Easton[3,9], Paul D.P. Pharoah[3,9], Marjanka K. Schmidt[85,149], Montserrat García-Closas[1**], Nilanjan Chatterjee[2,150**]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA, [2]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, [3]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, [4]Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, [5]Cyprus School of Molecular Medicine, Nicosia, Cyprus, [6]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia, [7]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia, [8]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia, [9]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK, [10]Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada, [11]Department of Epidemiology, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA, USA, [12]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, [13]Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada, [14]Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [15]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA, [16]Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden, [17]Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium, [18]Institute of Medical Biometry and Epidemiology, University of Hamburg, Hamburg, Germany, [19]Division of Cancer Epidemiology,

German Cancer Research Center (DKFZ), Heidelberg, Germany, [20]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Valencia, Spain, [21]Institute of Biochemistry and Genetics, Ufa Scientific Center of Russian Academy of Sciences, Ufa, Russia, [22]Department of Oncology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, [23]Department of Oncology, Örebro University Hospital, Örebro, Sweden, [24]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [25]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark, [26]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [27]Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain, [28]Division of Cancer Prevention and Genetics, IEO, European Institute of Oncology IRCCS, Milan, Italy, [29]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway, [30]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway, [31]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany, [32]iFIT-Cluster of Excellence, University of Tübingen, Tübingen, Germany, [33]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, [34]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, [35]Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada, [36]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada, [37]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany, [38]Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg,

Germany, [39]Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Germany, [40]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, [41]Oncology and Genetics Unit, Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain, [42]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [43]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, [44]Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia, [45]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands, [46]Sheffield Institute for Nucleic Acids (SInFoNiA), and Weston Park Cancer Centre, Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK, [47]Department of Neuroscience, University of Sheffield, Sheffield, UK, [48]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [49]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA, [50]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, [51]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, [52]Gynaecology Research Unit, Hannover Medical School, Hannover, Germany, [53]Department of Biomedical Sciences, Faculty of Science and Technology, University of Westminster, London, UK, [54]Cancer Sciences Academic Unit, Faculty of Medicine, University of Southampton, Southampton, UK, [55]Division of Evolution and Genomic Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, [56]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester NIHR Biomedical

Research Centre, Manchester University Hospitals NHS, Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK, [57]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA, [58]Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany, [59]Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh Medical School, Edinburgh, UK, [60]Cancer Research UK Edinburgh Centre, Edinburgh, UK, [61]Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain, [62]Moores Cancer Center, University of California San Diego, La Jolla, CA, USA, [63]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA, [64]Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain, [65]Department of Medicine, McGill University, Montréal, QC, Canada, [66]Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montréal, QC, Canada, [67]Breast Cancer Research Unit, Cancer Research Institute, University Malaya Medical Centre, Kuala Lumpur, Malaysia, [68]Department of Dermatology, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT, USA, [69]Department of Surgery, Oulu University Hospital, University of Oulu, Oulu, Finland, [70]Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France, [71]Department of Preventive Medicine, Keck School of Medicine, University of Southern

California, Los Angeles, CA, USA, [72]Department of Oncology, Södersjukhuset, Stockholm, Sweden, [73]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany, [74]Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, [75]Nightingale Breast Screening Centre, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK, [76]Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands, [77]Saarland Cancer Registry, Saarbrücken, Germany, [78]Division of Cancer Sciences, University of Manchester, Manchester, UK, [79]Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov", MASA, Skopje, Republic of North Macedonia, [80]Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland, [81]Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin, Poland, [82]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA, [83]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK, [84]Department of Pathology, Oulu University Hospital, University of Oulu, Oulu, Finland, [85]Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, [86]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia, [87]Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA, [88]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany, [89]Department of Histopathology and Cytology, Clinical Hospital Acibadem Sistina, Skopje, Republic of North Macedonia, [90]Department of Electron

Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus, [91]VIB Center for Cancer Biology, VIB, Leuven, Belgium, [92]Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven, Belgium, [93]Department of Mathematics and Statistics, St. Francis Xavier University, Antigonish, Canada, [94]Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada, [95]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, [96]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden, [97]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Cancer Center, Oncology Institute, Warsaw, Poland, [98]General and Gastroenterology Surgery Service, Hospital Universitario Central de Asturias, Oviedo, Spain, [99]University of Tübingen, Tübingen, Germany, [100]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland, [101]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland, [102]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland, [103]Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden, [104]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA, [105]Department of Anatomical Pathology, The Alfred Hospital, Melbourne, Victoria, Australia, [106]Department of Gynecology and Obstetrics, Ludwig Maximilian University of Munich, Munich, Germany, [107]MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, University College London, London, UK, [108]Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, [109]Moores Cancer Center, University of California, San Diego, La Jolla, CA, USA, [110]Department of Family Medicine and Public Health, School of Medicine, University of California, San Diego, La

Jolla, CA, USA, [111]Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA, [112]Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA, [113]Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Ireland, UK, [114]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK, [115]Human Genotyping-CEGEN Unit, Human Cancer Genetic Program, Spanish National Cancer Research Centre, Madrid, Spain, [116]Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland, [117]Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland, [118]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy, [119]Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel, [120]Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain, [121]Institute of Pathology, Staedtisches Klinikum Karlsruhe, Karlsruhe, Germany, [122]Department of Oncology, University Hospital of Larissa, Larissa, Greece, [123]Prevent Breast Cancer Centre and Nightingale Breast Screening Centre, Manchester University NHS Foundation Trust, Manchester, UK, [124]Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA, [125]Research Oncology, Guy’s Hospital, King's College London, London, UK, [126]Center for Hereditary Breast and Ovarian Cancer, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany, [127]Department of Health Sciences Research, Mayo Clinic College of Medicine, Jacksonville, FL, USA, [128]Division of Epidemiology, Department of

Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA, [129]Ss. Cyril and Methodius University in Skopje, Medical Faculty, University Clinic of Radiotherapy and Oncology, Skopje, Republic of North Macedonia, [130]Population Oncology, BC Cancer, Vancouver, BC, Canada, [131]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada, [132]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK, [133]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, [134]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [135]Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [136]Faculty of Medicine, University of Southampton, Southampton, UK, [137]Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA, [138]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA, [139]Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota, Colombia, [140]Department of Epidemiology, Gilliungs School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [141]Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA, [142]Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands, [143]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA, [144]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada, [145]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, [146]Department of

Surgical Sciences, Uppsala University, Uppsala, Sweden, [147]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA, [148]Genomics Center, Centre Hospitalier Universitaire de Québec – Université Laval, Research Center, Québec City, QC, Canada, [149]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands, [150]Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

*Contributed equally

**Contributed equally

†Lists of participants and their affiliations appear in the Funding and Acknowledgments

Word count abstract: 250/250

Word count main text: 3,000/3,000

Conflicts of interest: None to report

Corresponding Author

Montserrat Garcia-Closas

National Cancer Institute

Division of Cancer Epidemiology & Genetics

9609 Medical Center Drive

Rockville, MD 20850

montserrat.garcia-closas@nih.gov

Abstract

Background: Genome-wide association studies have identified over 170 common breast cancer susceptibility variants, many of them with differential associations by estrogen receptor (ER). How these variants are related to other tumor features is unclear.

Methods: Analyses included 106,571 invasive breast cancer cases and 95,762 controls of European ancestry with data on 178 genotyped or imputed single nucleotide polymorphisms (SNPs). We used two-stage polytomous logistic regression models to evaluate SNPs in relation to multiple tumor features (ER, progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and grade) adjusting for each other, and to intrinsic-like subtypes.

Results: Nearly half of the SNPs (85 of 178) were associated with at least one tumor feature (false discovery rate <5%). Case-case comparisons identified ER and grade as the most common heterogeneity sources, followed by PR and HER2. Case-control comparisons among these 85 SNPs with intrinsic-like subtypes identified 65 SNPs strongly or exclusively associated at P<0.05 with luminal-like subtypes, 5 SNPs associated with all subtypes at differing strengths, and 15 SNPs primarily associated with non-luminal tumors, especially triple-negative (TN) disease. The I157T *CHEK2* variant (rs17879961) was associated in opposite directions with luminal A-like (odds ratio (OR; 95% confidence interval (CI))=1.44 (1.31 to 1.59); $P=9.26\times10^{-14}$) and TN (OR (95% CI)=0.61 (0.47 to 0.80); $P=2.55\times10^{-4}$).

Conclusion: About half of the breast cancer susceptibility loci discovered in overall and ER-specific risk analyses have differential associations with clinical tumor features. These findings provide insights into the genetic predisposition of breast cancer subtypes and can inform subtype-specific risk prediction.

Introduction

Breast cancer represents a heterogenous group of diseases with different molecular and clinical features [1]. Clinical assessment of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and histological grade are routinely determined to inform treatment strategies and prognostication [2]. Combined, these tumor features define five intrinsic-like subtypes (i.e. luminal A-like, luminal B/HER2-negative-like, luminal B-like, HER2-enriched-like, and basal-like/triple negative) that are correlated with intrinsic subtypes defined by gene expression panels [2]. Most known breast cancer risk or protective factors are related to luminal or hormone receptor (ER or PR; HR) positive tumors, whereas less is known about the etiology of triple-negative (TN) tumors, an aggressive subtype [3, 4].

Breast cancer genome-wide association studies (GWAS) have identified over 170 susceptibility single nucleotide polymorphisms (SNPs), of which many are differentially associated with ER-positive than ER-negative disease [5]. These include 20 SNPs that primarily predispose to ER-negative or TN disease [6, 7]. However, few studies have evaluated SNP associations with other tumor features, or simultaneously studied multiple, correlated tumor markers to identify source(s) of etiologic heterogeneity [6, 8-12]. We recently developed a two-stage polytomous logistic regression method that efficiently characterizes etiologic heterogeneity while accounting for tumor marker correlations and missing tumor data [13]. This method can help describe complex relationships between susceptibility variants and multiple tumor features, helping to clarify breast cancer subtype etiologies and increasing the power to generate more accurate risk estimates between susceptibility variants and less

common subtypes.

In this report, we applied this novel methodology to a large study population from the Breast Cancer Association Consortium (BCAC) to characterize risk associations of 178 known breast cancer susceptibility SNPs with tumor subtypes defined by ER, PR, HER2 and tumor grade.

Methods

Study Population and Genotyping

The study population and genotyping are described in previous publications [5, 6] and in the **Supplemental Methods**. We included invasive cases and controls from 81 BCAC studies with genotyping data from two Illumina genome-wide custom arrays, the iCOGS and OncoArray (106,571 cases (OncoArray: 71,788; iCOGS: 34,783) and 95,762 controls (OncoArray: 58,134; iCOGS: 37,628); **Supplemental Table 1**). We evaluated 178 susceptibility SNPs that were identified in or replicated by prior BCAC analyses [5, 6]. Genotypes for the SNPs marking the 178 susceptibility loci were determined by genotyping with the iCOGS and the OncoArray arrays and imputation to the 1000 Genomes Project (Phase 3) reference panel.

Statistical Analysis

The statistical methods, including a detailed discussion of the two-stage polytomous logistic regression, are provided in the **Supplemental Methods** and elsewhere [13]. Briefly, we identified SNPs showing evidence for heterogeneity by using a mixed-effects two-stage polytomous model to evaluate a global heterogeneity test that assesses whether a SNP's case-

control risk-estimates vary by at least one of the underlying tumor characteristics. We accounted for multiple testing of the global heterogeneity test using a false discovery rate (FDR) <0.05 under the Benjamini-Hochberg procedure [14]. Among SNPs with evidence for heterogeneity, we used a fixed-effects two-stage model to evaluate a case-case marker-specific tumor heterogeneity test, which identifies the specific tumor marker(s) contributing to the observed heterogeneity, adjusting for the other tumor markers in the model. Marker-specific P<0.05 was considered statistically significant. Our primary analyses evaluated heterogeneity by ER, PR, HER2 and grade. As a secondary analysis, we fit an extended model with an additional term for TN status to test for differences between TN vs non-TN subtypes. The two-stage model implements an efficient expectation-maximization algorithm [15] to essentially perform iterative "imputation" of missing tumor characteristics [13]. We fit an additional two-stage model to estimate case-control ORs and 95% confidence intervals (CI) between the SNPs and five intrinsic-like subtypes defined by combinations of ER, PR, HER2 and grade (see **Supplemental Methods**): (1) luminal A-like, (2) luminal B/HER2-negative-like, (3) luminal B-like, (4) HER2-enriched-like and (5) TN. For all analyses we analyzed OncoArray and iCOGS array data separately, adjusting for the first 10 principal components for ancestry-informative SNPs, and then meta-analyzed the results. We used Euclidean distance in cluster analyses to help describe results and identify common heterogeneity patterns.

Results

The mean (SD) ages at diagnosis (cases) and enrollment (controls) were 56.6 (12.2) and 56.4 (12.2) years, respectively. Eighty-one percent of tumors were ER-positive, 68% PR-positive, 83% HER2-negative and 69% grade 1 or 2 (**Table 1; Supplemental Table 1**). The most

common intrinsic-like subtype was luminal A-like (59%), followed by TN (13%), luminal B/HER2-negative-like (12%), Luminal B-like (12%) and HER2-enriched-like (5%; **Table 1**).

The two-stage models including terms for ER, PR, HER2 and grade, simultaneously adjusting for each other, identified 85 of 178 SNPs (47.7%) with evidence for heterogeneity by at least one tumor feature (FDR<5%). ER and grade most often contributed to the observed heterogeneity (45 and 34 SNPs respectively had case-case marker-specific P<0.05), and 30 SNPs were significantly associated with more than one tumor characteristic (**Figure 1; Supplementary Figure 1)**. Seventeen of these 85 SNPs showed no associations with any individual tumor marker at P<0.05 in their corresponding fixed-effect two-stage models (**Supplementary Figure 1**). Twelve SNPs (**Supplemental Figure** 1) were significantly associated exclusively with grade (1p22.3-rs17426269 (P_{grade}=1.52x10^{-02}), 1q21.2-rs12048493 (P_{grade}=3.09x10^{-03}), 1q22-rs4971059 (P_{grade}=1.84x10^{-02}), 3p.24.1-rs12493607 (P_{grade}=7.78x10^{-10}), 3q26.31-rs58058861 (P_{grade}=1.54x10^{-03}), 5p13.3-rs2012709 (P_{grade}=6.25x10^{-03}), 10q22.3-rs704010 (P_{grade}=2.87x10^{-04}), 11q24.3-rs11820646 (P_{grade}=3.18x10^{-02}), 13q13.1-rs11571833 (P_{grade}=1.78x10^{-03}), 17q22-rs2787486 (P_{grade}=1.70x10^{-04}), 19p13.11-rs4808801 (P_{grade}=5.10x10^{-04}), 22q13.1-rs738321 (P_{grade}=2.80x10^{-02})), four SNPs were associated exclusively with PR (5q11.1-rs72749841 (P_{PR}=5.46x10^{-03}), 9q31.2-rs10816625 (P_{PR}=3.70x10^{-02}), 9q31.2-rs10759243 (P_{PR}=2.18x10^{-05}), 10q26.12-rs11199914 (P_{PR}=4.39x10^{-02})) and one SNP was associated exclusively with HER2 (10q21.2-rs10995201 (P_{HER2}= 2.60x10^{-02})).

Case-control comparisons for the 85 SNPs with evidence for global heterogeneity identified four main clusters of SNPs according to the p-values for risk associations with each subtype (**Figure 2 and Supplemental Figure 2**). Sixty-five SNPs in cluster 1 (n=3 SNPs) and

cluster 4 (n=62 SNPs) showed the strongest evidence for associations with risk for luminal-like subtypes, cluster 2 (n=5 SNPs) was associated with risk for all subtypes at varying strengths and cluster 3 (n=15 SNPs) with stronger evidence for TN or non-luminal subtype associations. **Supplemental Table 2** shows the associations between all 178 SNPs and the intrinsic-like subtypes.

Cluster 1 included two correlated (r^2=0.73) SNPs at 10q26.13, rs2981578 and rs35054928, that were strongly associated with risk of all the luminal-like subtypes (e.g. OR (95%CI)=1.29 (1.27 to 1.31), P=2.01x10^{-231} and OR (95%CI)=1.35 (1.33 to 1.37), P=5.00x10^{-300} for luminal A-like, respectively), weakly associated with risk of HER2-enriched-like subtype (OR (95%CI)=1.11 (1.05 to1.16), P=6.32x10^{-05} and OR (95% CI)=1.10 (1.04 to 1.15), P=3.07x10^{-4}, respectively), but not significantly associated (P>0.05) with TN tumors (**Figures 2-3 and Supplemental Figure 2**). Case-case comparisons showed the strongest evidence for associations with ER (P_{ER}=1.27x10^{-30} for rs2981578 and P_{ER}=9.98x10^{-38} for rs35054928 (**Supplemental Figure 1**). In the extended two-stage model that additionally included a term for TN status (**Supplemental Figure 3**), both SNPs were associated with TN status (P_{TN}=3.79x10^{-06} for rs2981578 and P_{TN}=1.27x10^{-06} for rs35054928). A third SNP in 10q26.13, rs45631563, showed similar association patterns, but fell into cluster 4 since, unlike rs2981578 and rs2981578, it was not significantly associated with risk of the HER2-enriched-like subtype (**Figure 2 and Supplemental Figures 2, 4**). Case-case comparisons for rs45631563 showed associations with ER (P_{ER}=9.09x10^{-7}) and grade (P_{grade}=2.56x10^{-3}). A third SNP in cluster 1, 16q12.1-rs4784227, was most strongly associated with luminal-like subtypes (e.g. OR (95%CI)=1.28 (1.26 to 1.30), P=6.68x10^{-176} for luminal A-like), and the HER2-enriched-like tumors (OR (95%CI)=1.26 (1.19 to

1.33), P=3.04x10^{-16}), and weaker so with TN tumors (OR (95%CI)=1.11 (1.08 to 1.15), P=6.05x10^{-10}; **Figures 2-3 and, Supplemental Figure 2**). In case-case analyses, rs4784227 was significantly associated with all tumor markers, particularly PR (P_{PR}=2.16x10^{-4}; **Supplemental Figure 1**).

Five SNPs in cluster 2 were associated, to different extents, with all five intrinsic-like subtypes. 6q25-rs2747652 and 1p36.22-rs616488 were associated particularly with risk of HER2-positive subtypes (**Figures 2-3**; **Supplemental Figure 2**). In case-case comparisons these SNPs were associated with HER2 status (P_{HER2}=1.84x10^{-7} for rs2747652 and P_{HER2}=2.51x10^{-6} for rs616488), and grade (P_{Grade}=3.82x10^{-5} for rs2747652 and P_{Grade}=0.02 for rs616488) (**Supplemental Figure 1**). Two additional SNPs in 6q25 showed the strongest evidence for being associated with TN disease (OR (95%CI)=1.30 (1.24 to 1.38), P=8.26x10^{-23} for rs9397437 and OR (95%CI)=1.15 (1.12 to 1.19), P=1.24x10^{-19} for rs3757322; **Figures 2-3**; **Supplemental Figure 2**), and in case-case comparisons were associated with ER (P_{ER}=4.72x10^{-3} for rs9397437 and P_{ER}=3.64x10^{-2} for rs3757322) and grade (P_{grade}=2.87x10^{-5} for rs9397437 and P_{grade}=2.34x10^{-3} for rs3757322; **Supplemental Figure 1**). 13q13.1-rs11571833 was associated with risk of all subtypes, but case-case comparisons showed an association only with grade (P_{Grade}=1.78x10^{-3}; **Supplemental Figure 1**). In case-control comparisons the ORs (95% Cis) for rs11571833 with grade 3, grade 2 and grade 1 subtypes were: 1.48 (1.36 to 1.62), P=2.3x10^{-19}; 1.27 (1.18 to 1.35), P=5.0x10^{-12}; and 1.08 (0.97-1.20), P=0.15, respectively (**Supplemental Figure 5**).

Cluster 3 included 15 SNPs most strongly associated with risk of HR-negative subtypes, including three SNPs with the strongest evidence for associations with TN disease: 19p13.11-rs67397200 (OR (95% CI)=1.27 (1.23 to 1.31), P=1.07x10^{-50}), 5p15.33-rs10069690 (OR (95% CI)=1.27 (1.23 to 1.31), P=3.79x10^{-48}) and 1q32.11-rs4245739 (OR (95% CI)=1.18 (1.14 to 1.22),

P=2.72x10^{-23}). Two SNPs at 11q22.3, rs11374964 (OR (95% CI)=0.90 (0.88-0.93), P=2.71x10^{-11}) and rs74911261 (OR (95% CI)=0.93 (0.90-0.96), P=2.71x10^{-11}), were significantly associated (P<0.05) only with TN disease (**Figures 2-3; Supplemental Figure 2**). In the extended model these five SNPs were associated with TN status (P_{TN}<0.05; **Supplemental Figure 3**). The remaining 10 SNPs in case-control comparisons were all associated with TN disease (P<0.05; **Figures 2-3; Supplemental Figure 2**), and in case-case comparisons seven of these 10 SNPs were exclusively associated with ER (P_{ER}<0.05): 1q32.1-rs6678914, 2p23.2-rs4577244, 8p23.3-rs66823261, 13q22.1-rs6562760, 16q12.2-rs11075995, 16p13.3-rs11076805 and 18q12.1-rs36194942. 5p15.33-rs3215401 was associated with both ER (P_{ER}=2.22x10^{-03}) and PR (P_{PR}=4.65x10^{-02}). 2p24.1-rs12710696 and 19q12-rs113701136 showed no significant associations (P>0.05) with any of the individual tumor markers (**Supplemental Figure 1**). Besides rs11374964 and rs74911261, SNPs in this cluster were not HR-negative or TN-specific SNPs as they were also associated with luminal-like subtypes. Three SNPs showed weak associations with luminal A-like disease (OR (95% CI)=0.98 (0.97 to 1.00); P=0.039 for rs67397200; OR (95% CI)=1.02 (1.00 to 1.03); P=0.024 for rs6678914; and OR (95% CI)=1.02 (1.00 to 1.04); P=0.02 for rs4577244) in an opposite direction to their associations with TN disease (OR (95% CI)=0.93 (0.91 to 0.96); P=1.07x10^{-4} for rs6678914 and OR (95% CI)=0.90 (0.86 to 0.93); P=2.99x10^{-9} for rs4577244; **Figures 2-3**; **Supplemental Figure 2**).

Cluster 4 (n=62 SNPs) showed evidence for associations with risk of luminal-subtypes, especially luminal A-like disease. The five SNPs in this cluster showing the strongest evidence for associations with risk of luminal A-like disease were: 5q11.2-rs62355902 (OR (95% CI)=1.22 (1.20 to 1.25), P=3.94x10^{-85}), 11q13.3-rs75915166 (OR (95% CI)=1.44 (1.40 to 1.48), P=2.26x10^{-}

[131]), 11q13.3-rs554219 (OR (95% CI)=1.33 (1.30 to 1.37), P=1.31x10^{-98}), 1p11.2-rs11249433 (OR (95% CI)=1.17 (1.15 to 1.19), P=5.25x10^{-90}) and 5p12-rs10941679 (OR (95% CI)=1.20 (1.18 to 1.22), P=1.39x10^{-93}; **Figure 2 and Supplemental Figures 2,4**). In case-case comparisons these five SNPs were associated with grade (P_{grade}=2.32x10^{-02} for rs62355902, P_{grade}=1.74x10^{-13} for rs75915166, P_{grade}=2.14x10^{-12} for rs554219, P_{grade}=2.20x10^{-12} for rs11249433 and P_{grade}=2.47x10^{-06} for rs10941679) and with at least one other tumor marker (**Supplemental Figure 1**). Eighteen of the 62 SNPs also showed weaker evidence for associations with non-luminal subtypes, 11 of which were associated with risk of TN disease (P<0.05): 5q11.2-rs62355902, 5p12-rs10941679, 12q22-rs17356907, 10q21.2-rs10995201, 10p12.31-rs7072776, 19p13.11-rs4808801, 10p12.31-rs11814448, 8q21.11-rs6472903, 11q24.3-rs11820646, 19p13.13-rs78269692 and 22q12.1-rs17879961 (I157T; **Figure 2 and Supplemental Figures 2,4**). Notably, rs7072776 and rs17879961 were associated in opposite directions with luminal A-like disease (OR (95% CI)=1.10 (1.08 to 1.11); P=4.96x10^{-27} for rs7072776; and OR (95% CI)=1.44 (1.31 to 1.59); P=9.26x10^{-14} for rs17879961) and TN disease (OR (95% CI)= 0.96 (0.93 to 0.99); P=0.02 for rs7072776 and OR (95% CI)=0.61 (0.47 to 0.80); P=2.55x10^{-4} for rs17879961). In case-case comparisons rs7072776 was associated with ER (P_{ER}=8.27x10^{-5}) and grade (P_{grade}=0.01; **Supplemental Figure 1**), and rs17879961 was associated with the TN phenotype in the extended model (P_{TN}=1.45x10^{-5}; **Supplemental Figure 3**).

Supplemental Figure 5 shows case-control associations by tumor grade for 12 SNPs associated exclusively with grade in case-case comparisons (P_{grade}<0.05). rs11571833, rs17426269 and rs11820646 showed stronger evidence for predisposing to risk of high-grade subtypes, and the remaining SNPs showed stronger evidence for predisposing to risk of low-

grade subtypes.

Discussion

We found compelling evidence that about half of the investigated breast cancer susceptibility loci (85 of 178 SNPs) predispose to tumors of different characteristics. We identified tumor grade, along with confirming ER and TN status, as important determinants of etiologic heterogeneity. Associations with individual tumor features translated into differential associations with the risk of intrinsic-like subtypes defined by their combinations.

Many of the SNPs showing subtype heterogeneity predisposed to risk of all subtypes, but with different magnitudes. For example, 21 of 65 SNPs found predominately associated with risk of luminal-like subtypes were also associated with risk of at least one of the non-luminal subtypes. These include SNPs identified in early GWAS for overall breast cancer, such as SNPs in the loci for *FGFR2* (rs35054928 and rs2981578)[16, 17] and 8q24.21 (rs13281615) [16] that were associated with luminal-like and HER2-enriched-like subtypes. rs4784227 located near *TOX3* [16, 18] and rs62355902 located in a *MAP3K1* [16] regulatory element, were associated with risk of all five subtypes. Of the five SNPs found associated in opposite directions with luminal A-like and TN disease, we previously reported rs6678914 and rs4577244 to have opposite effects between ER-negative and ER-positive tumors [6]. rs17879961 (I157T), a likely causal [19] mis-sense variant located in a *CHEK2* functional domain that reduces or abolishes substrate binding [20], was previously reported to have opposite directions of effects on lung adenocarcinoma and lung squamous cell carcinoma and for lung cancer between smokers and

non-smokers [21, 22]. However, further studies are required to follow-up and clarify the mechanisms for these apparent cross-over effects.

In prior ER-negative GWAS we identified 20 SNPs that predispose to ER-negative disease, of which five SNPs were only or most strongly associated with risk of TN disease (rs4245739, rs10069690, rs74911261, rs11374964, and rs67397200) [6, 7]. We further confirmed these five SNPs to be most strongly associated with TN disease and that they were associated with TN status in the extended model. The remaining previously identified 15 SNPs all showed associations with risk of HR-negative disease, and for all but four SNPs (rs17350191, rs200648189, rs6569648, and rs322144) evidence of global heterogeneity was observed. Among the SNPs in cluster 3, rs3215401 was the only SNP that was not identified in a prior ER-negative GWAS [6, 7]. rs3215401 was identified in a fine-mapping analysis of *TERT* and, consistent with our findings, reported to be most strongly associated with ER-negative disease but was also associated with ER-positive disease [23].

Little is known regarding PR and HER2 as sources of etiologic heterogeneity independent of ER or TN status. Of the four SNPs significantly associated only with PR, rs10759243 [5, 24], rs11199914 [5] and rs72749841 [5] were previously found primarily associated with risk of ER-positive disease, and rs10816625 was found to be associated with risk of ER-positive/PR-positive tumors, but not other ER/PR combinations [11]. rs10995201 was the only variant found to be solely associated with HER2 status, although the evidence was not strong, requiring further confirmation. Previously rs10995201 showed no evidence of being associated with ER status [25]. Among all SNPs found with PR or HER2 associations, few have been investigated for PR or HER2 heterogeneity while adjusting for ER [8-12]. We previously reported rs10941679 to

be associated with PR-status, independent of ER, and also with grade [9]. We also found suggestive evidence of PR-specific heterogeneity for 16q12-rs3803662 [12], which is in high LD (r^2= 0.78) with rs4784227 (*TOX3*), which was strongly associated with PR status. Our findings for rs2747652 are also consistent with a prior BCAC fine-mapping analysis across the *ESR1* locus, which found rs2747652 to be associated with risk of the HER2-enriched subtype and high grade independent of ER [8]. rs2747652 overlaps an enhancer region and is associated with reduced *ESR1* and *CCDC170* expression [8].

Histologic grade is a composite of multiple tumor characteristics including mitotic count, nuclear pleomorphism, and degree of tubule or gland formation [26]. Among the 12 SNPs identified with evidence of heterogeneity by grade only, rs17426269, rs11820646, and rs11571833 were found most strongly associated with grade 3 disease. rs11571833 lies in the *BRCA2* coding region and produces a truncated form of the protein [27] and has been shown to be associated with both risk of TN disease and risk of serous ovarian tumors, both of which tend to be high-grade [28]. To our knowledge, rs17426269 and rs11820646 have not been investigated in relation to grade heterogeneity. The remaining 9 SNPs were all more strongly associated with grade 1 or grade 2 disease. Five of these SNPs were previously reported to be associated primarily with ER-positive disease [5, 29, 30], highlighting the importance of accounting for multiple tumor characteristics to better illuminate heterogeneity sources.

A major strength of our study is our large sample size of over 100,000 breast cancer cases with tumor marker information, and a similar number of controls, making this the largest, most comprehensive breast cancer heterogeneity investigation. Our application of the novel two-stage polytomous logistic regression enabled adjusting for multiple, correlated tumor

markers and accounting for missing tumor marker data. This is a more powerful and efficient modeling strategy for identifying heterogeneity sources among highly correlated tumor markers, compared with standard polytomous logistic regression. However, we identified 17 SNPs with evidence of heterogeneity for which we did not identify specific tumor characteristic(s) contributing to observed heterogeneity. This is likely explained by the fact that the fixed-effects models evaluating specific tumor markers as heterogeneity sources were less statistically powerful compared with the mixed-effects models that evaluated for evidence of global heterogeneity [13]. Our approach to cluster SNPs helped describe common heterogeneity patterns; however, these clusters should not be interpreted as strictly defined categories. Our study was limited by investigating only ER, PR, HER2, and grade as heterogeneity sources and future studies with more detailed tumor characterization could reveal additional etiologic heterogeneity sources.

In summary, our findings provide insights into the complex etiologic heterogeneity patterns of common breast cancer susceptibility loci. These findings may inform fine-mapping and functional analyses to identify the underlying causal variants, clarifying biological mechanisms that drive genetic predisposition to breast cancer subtypes. Moreover, these analyses provide precise estimates of relative risk for different intrinsic-like subtypes that could improve the discriminatory accuracy of subtype-specific polygenic risk scores [31].

References

1. The Cancer Genome Atlas N, Koboldt DC, Fulton RS*, et al.* Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61.

2. Curigliano G, Burstein HJ, E PW*, et al.* De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. Ann Oncol 2017;28(8):1700-1712.

3. Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk of intrinsic tumor subtypes. Biochim Biophys Acta 2015;1856(1):73-85.

4. Yang XR, Chang-Claude J, Goode EL*, et al.* Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 2011;103(3):250-63.

5. Michailidou K, Lindstrom S, Dennis J*, et al.* Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551(7678):92-94.

6. Milne RL, Kuchenbaecker KB, Michailidou K*, et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet 2017;49(12):1767-1778.

7. Garcia-Closas M, Couch FJ, Lindstrom S*, et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet 2013;45(4):392-8, 398e1-2.

8. Dunning AM, Michailidou K, Kuchenbaecker KB*, et al.* Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nat Genet 2016;48(4):374-86.

9. Milne RL, Goode EL, Garcia-Closas M*, et al.* Confirmation of 5p12 as a susceptibility locus for progesterone-receptor-positive, lower grade breast cancer. Cancer Epidemiol Biomarkers Prev 2011;20(10):2222-31.

10. Figueroa JD, Garcia-Closas M, Humphreys M*, et al.* Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. Hum Mol Genet 2011;20(23):4693-706.

11. Orr N, Dudbridge F, Dryden N*, et al.* Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. Hum Mol Genet 2015;24(10):2966-84.

12. Broeks A, Schmidt MK, Sherman ME*, et al.* Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Hum Mol Genet 2011;20(16):3289-303.

13. Zhang H, Zhao N, Ahearn TU*, et al.* A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics. bioRxiv 2018; 10.1101/446039:446039.

14. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 1995;57(1):289-300.

15. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via Em Algorithm. Journal of the Royal Statistical Society Series B-Methodological 1977;39(1):1-38.

16. Easton DF, Pooley KA, Dunning AM*, et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447(7148):1087-93.

17. Hunter DJ, Kraft P, Jacobs KB*, et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39(7):870-4.

18. Stacey SN, Manolescu A, Sulem P*, et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39(7):865-9.

19. Fachal L, Aschard H, Beesley J*, et al.* Fine-mapping of 150 breast cancer risk regions identifies 178 high confidence target genes. bioRxiv 2019; 10.1101/521054:521054.

20. Li J, Williams BL, Haire LF*, et al.* Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. Mol Cell 2002;9(5):1045-54.

21. Wang Y, McKay JD, Rafnar T*, et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 2014;46(7):736-41.

22. McKay JD, Hung RJ, Han Y*, et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet 2017;49(7):1126-1132.

23. Bojesen SE, Pooley KA, Johnatty SE*, et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet 2013;45(4):371-84, 384e1-2.

24. Li X, Zou W, Liu M*, et al.* Association of multiple genetic variants with breast cancer susceptibility in the Han Chinese population. Oncotarget 2016;7(51):85483-85491.

25. Darabi H, McCue K, Beesley J*, et al.* Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate NRBF2 Expression. Am J Hum Genet 2015;97(1):22-34.

26. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19(5):403-10.

27. Mazoyer S, Dunning AM, Serova O*, et al.* A polymorphic stop codon in BRCA2. Nat Genet 1996;14(3):253-4.

28. Meeks HD, Song H, Michailidou K*, et al.* BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. J Natl Cancer Inst 2016;108(2).

29. Darabi H, Beesley J, Droit A*, et al.* Fine scale mapping of the 17q22 breast cancer locus using dense SNPs, genotyped within the Collaborative Oncological Gene-Environment Study (COGs). Sci Rep 2016;6:32512.

30. Michailidou K, Hall P, Gonzalez-Neira A*, et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45(4):353-61, 361e1-2.

31. Mavaddat N, Michailidou K, Dennis J*, et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet 2019;104(1):21-34.

Table 1. Distribution of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and grade and the intrinsic-like subtypes[*,†] among invasive cases of breast cancer in studies from the Breast Cancer Consortium Association.

	N (%)
ER	
Negative	16,900 (19)
Positive	70,030 (81)
Unknown	19,641
PR	
Negative	24,283 (32)
Positive	51,603 (68)
Unknown	30,685
HER2	
Negative	47,693 (83)
Positive	9,529 (17)
Unknown	49,349
Grade	
1	15,583 (20)
2	37,568 (49)
3	24,382 (31)
Unknown	29,038
Intrinsic-like subtypes	
Luminal A-like	33,083 (59)
Luminal B/Her2-negative-like	6,804 (12)
Luminal B-like	6,511 (12)
HER2-enriched-like	2,797 (5)
Triple-negative	7,178 (13)

*= luminal A-like (ER+ and/or PR+, HER2-, grade 1 & 2); luminal B/HER2-negative-like (ER+ and/or PR+, HER2-, grade 3); luminal B-like (ER+ and/or PR+, HER2+); HER2-enriched-like (ER- and PR-, HER2+); TN (ER-, PR-, HER2-)
†= Intrinsic subtypes defined among 56,373 cases with available tumor marker data

Figure 1. Heatmap and clustering of the P-values from the two-stage model's case-case marker-specific heterogeneity test for associations between each of the 178 breast cancer susceptibilty SNPs and estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) or grade, adjusting for principal components and each tumor marker. Columns represent individual SNPs. For more detailed information on the context of figure see **Supplemental Figure 1.**

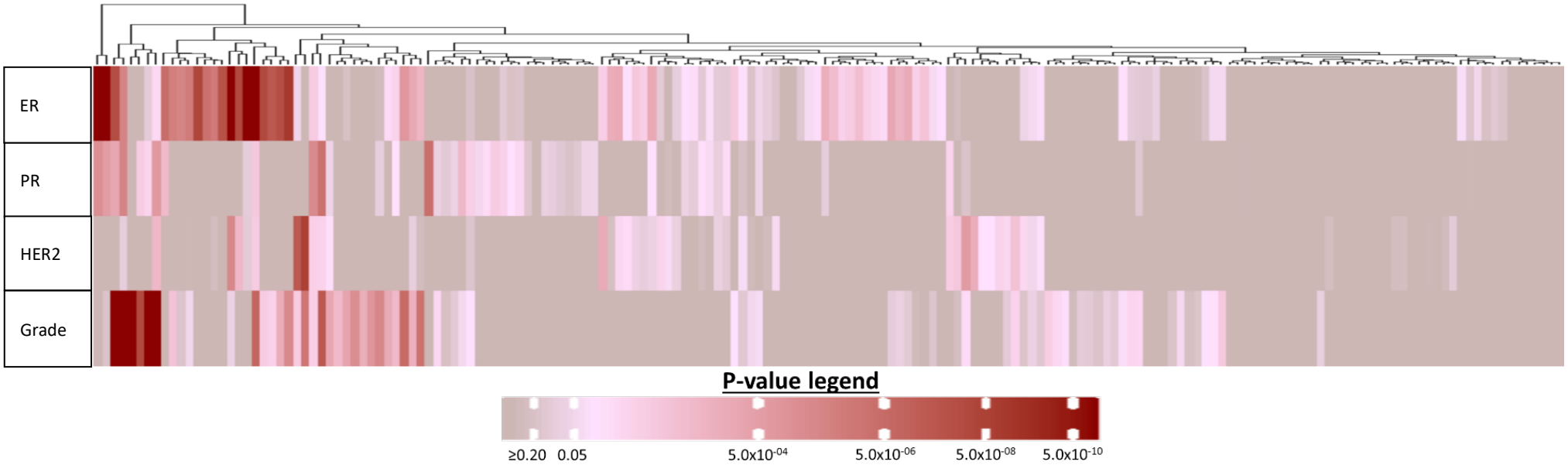
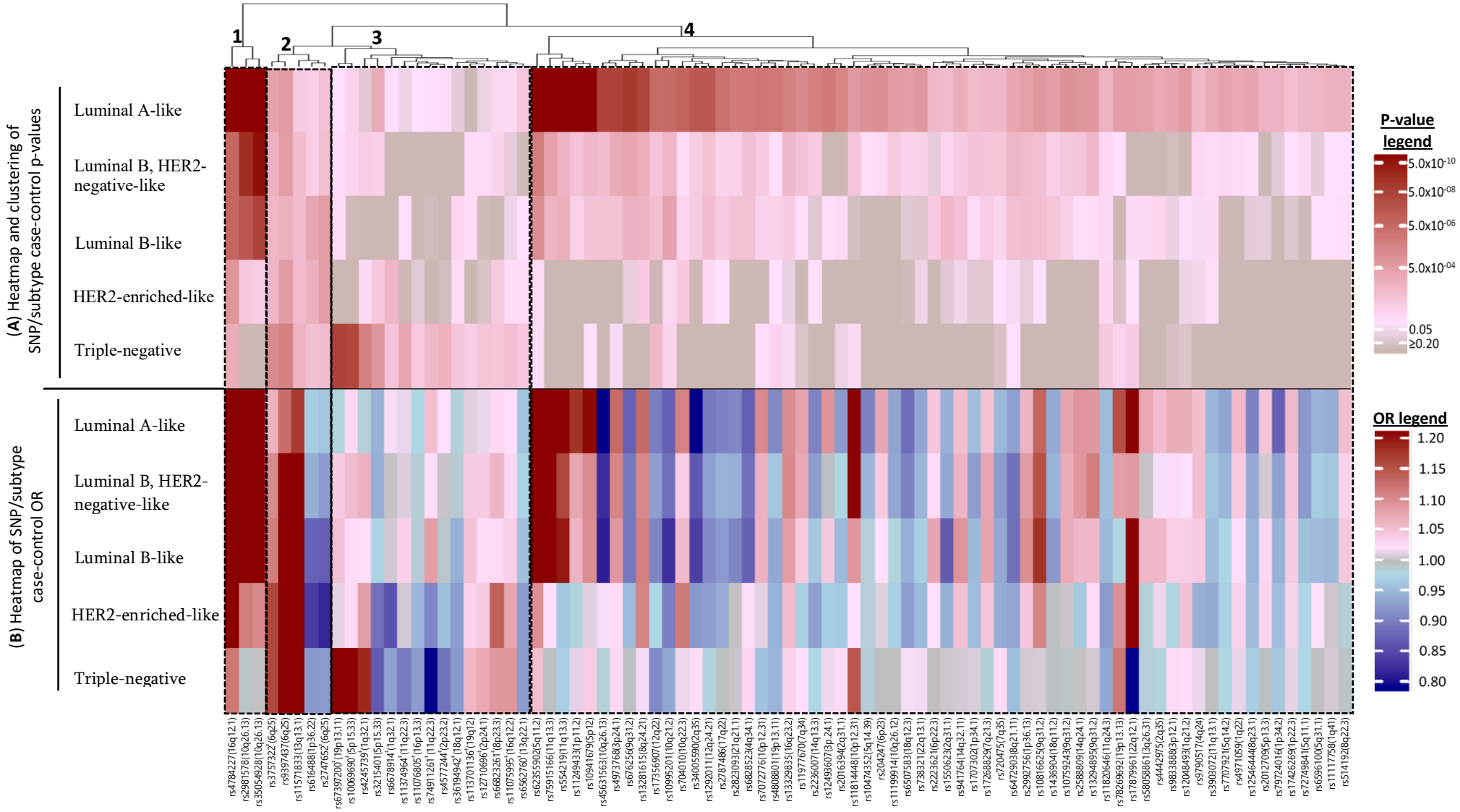


Figure 2. Heatmap and clustering of case-control P-values and odds ratios (OR) from two-stage models for intrisic-like subtypes, fitted for each of the 85 breast cancer susceptibility SNPs with evidence for global heterogeneity. Columns are labeled according to SNP and loci. (**A**) Clustering based on case-control p-value of the associations between susceptibility SNPs and breast cancer intrinsic-like subtypes[*,†]. (**B**) Heatmap of ORs between susceptibilit SNPs and intrinsic-like subtypes.

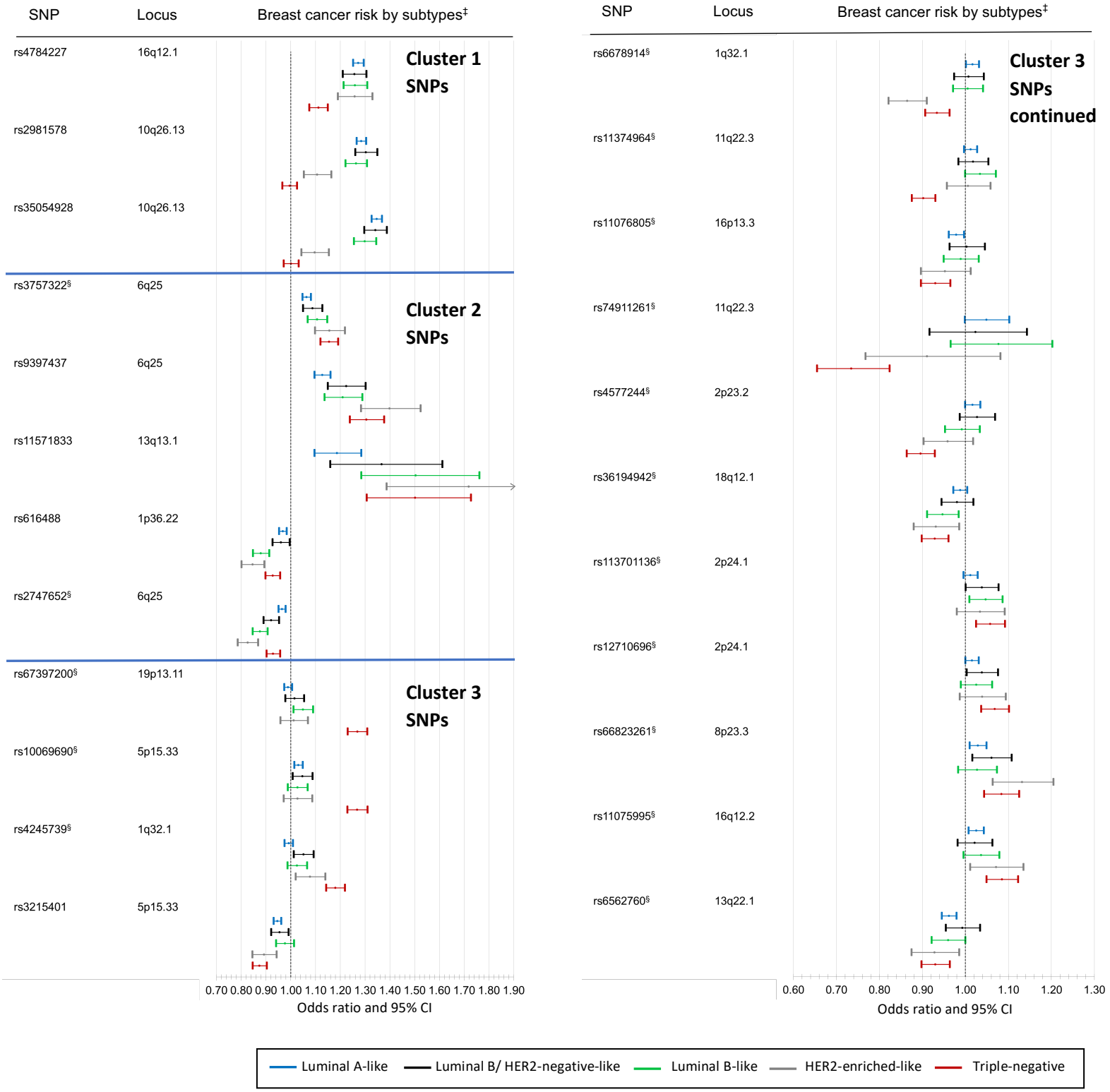


*= luminal A-like (ER+ and/or PR+, HER2-, grade 1 & 2); luminal B/HER2-negative-like (ER+ and/or PR+, HER2-, grade 3); luminal B-like (ER+ and/or PR+, HER2+); HER2-enriched-like (ER- and PR-, HER2+); TN (ER-, PR-, HER2-)

†= Clusters: 1, 3, 5: strongest associations with hormone receptor (HR)-positive subtypes; 2: Strongest associations with TN subtypes; 4: Strong associations with HR-positive and HR-negative subtypes

‡ = SNPs previously identified as primarily predisposing to ER-negative or TN disease as reported in Milne RL, et al. Nat Genet 2017;49(12):1767-1778

Figure 3. Associations* between breast cancer susceptibility SNPs with evidence of heterogeneity in the two-stage model† and intrinsic-like subtypes. SNPs presented in order as shown in figure 2, clusters 1, 2 and 3.



* Per-minor allele odds ratio (95% confidence limits).

† Two-stage model testing for heterogeneity according to estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and grade

‡ luminal A-like (ER+ and/or PR+, HER2-, grade 1 & 2); luminal B/HER2-negative-like (ER+ and/or PR+, HER2-, grade 3); luminal B-like (ER+ and/or PR+, HER2+); (4) HER2-enriched-like (ER- and PR-, HER2+), and triple-negative (ER-, PR-, HER2-)

§ SNPs previously identified as primarily predisposing to ER-negative or TN disease as reported in Milne RL, et al. Nat Genet 2017;49(12):1767-1778.